

Evaluación del programa de escuelas-taller y casas de oficio a partir de su efecto sobre el tiempo de búsqueda del primer empleo. El caso de Sevilla

JOSÉ MANUEL CANSINO MUÑOZ-REPISO y ANTONIO SÁNCHEZ BRAZA

Departamento de Teoría Económica y Economía Política

UNIVERSIDAD DE SEVILLA

e-mail: jmcansino@us.es; asb@us.es

RESUMEN

El artículo evalúa el efecto causal del Programa de Escuelas Taller y Casas de Oficio sobre el tiempo de búsqueda del primer empleo de los individuos. La evaluación se lleva a cabo mediante el estimador de diferencias en diferencias.

El volumen de fondos públicos y el elevado número de participantes, justifica la relevancia del programa evaluado en el contexto de las políticas de formación desarrolladas en España. El artículo contribuye a aumentar el escaso número de investigaciones de este tipo publicadas sobre España.

Los resultados obtenidos permiten a los autores concluir que, para los participantes, que son jóvenes desempleados, el tiempo de búsqueda del primer empleo se reduce en 494'57 días con respecto a los no participantes. Adicionalmente, el efecto causal del programa es mayor para los hombres que para las mujeres y aumenta con la edad a razón de 12,94 días menos de búsqueda por cada año adicional.

Palabras clave: Evaluación de políticas públicas; estimador de diferencias en diferencias; programas de formación.

Evaluation of the Training School Program's Effect on the Period to Find the First Job. The Case of Seville (Spain)

ABSTRACT

This paper evaluates the causal effect of the Spanish Training Scholl Program on the period an individual need to find his first job. The evaluation is implemented by using the differences in differences estimator –DID–.

The public funds involved and the high number of participants justified the relevance of the program evaluated in the Spanish training program's context.

Debt to the fact that no papers have been published in the literature in Spanish, this paper contributes in that way.

Results obtained let authors conclude that, for participants, the period is reduced in 494'57days comparing with controls. Additionally, the program's causal effect is higher for men than for women and increases with the age of the individual in a value equal to 12'94 days for every year.

Keywords: Public Policies Evaluation; Differences in Differences Estimator; Training Programs.

Clasificación JEL: C23, H30.

Artículo recibido en abril de 2008 y aceptado en julio de 2008.

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref. 27101.

1. INTRODUCCIÓN Y OBJETIVO¹

Los programas públicos de formación constituyen una parte importante de las políticas activas de empleo, orientándose frecuentemente a colectivos con dificultades para la incorporación al mercado laboral. La magnitud presupuestaria que han absorbido estas políticas, justifica el interés por la evaluación de sus efectos sobre sus beneficiarios. Ese interés ayuda a entender el considerable desarrollo metodológico y aplicado de la evaluación económica de los programas de formación o entrenamiento en las últimas décadas.

Ejemplos recientes de ese desarrollo, con enfoques metodológicos diversos, se encuentran en Card y Sullivan (1988) y Manski y Garfinkel (1992) para EE.UU., Bonnell, Fougère y Sérandon (1997) para Francia, Andrews, Bradley y Upward (1999) y Blundell et al. (2002) en el R.U., Bergemann, Fitzenberger y Speckesser (2002) en Alemania y Park et al. (1996) en Canadá. En España, no obstante, los trabajos evaluatorios de esta naturaleza no han gozado del mismo desarrollo. La dificultad en el acceso a microdatos explica, parcialmente, la escasez de investigaciones aplicadas pese a que alguna vía de acceso a esta información se ha señalado recientemente por Cansino y Román (2007).

La utilidad de estas investigaciones es evidente si se considera que los resultados de la evaluación cuando, como en este caso, las políticas públicas se aplican sobre poblaciones heterogéneas, proporcionan al decisor público información relevante sobre su efectividad, habilitándole para rediseñar futuras iniciativas a la luz de los resultados, en la línea de las aportaciones de Manski (2001).

El objetivo del artículo es evaluar una de las políticas activas de empleo, el Programa de Escuelas Taller y Casas de Oficio, a partir de su efecto causal sobre el *outcome* definido como la capacidad del individuo para encontrar un empleo. Para su cálculo se ha utilizado el estimador de diferencias en diferencias, en la línea de las aportaciones de Heckman, Ichimura, Smith y Todd (1998), Ashenfelter y Card (1985), Card (1990), Card y Krueger (1994 y 2000), Athey e Imbens (2002) y Abadie (2005). Su relevancia se justifica tanto por el volumen de fondos públicos absorbidos, como por el número de participantes. El ámbito geográfico de la evaluación se sitúa en la Provincia de Sevilla, y el periodo temporal entre los años 1997 a 1999. El artículo contribuye a cubrir un hueco en la literatura sobre evaluaciones aplicadas en España del efecto de programas de formación.

Para el mejor desarrollo de los objetivos propuestos, se ha estructurado este trabajo de investigación en seis apartados en los que, tras la introducción y objetivos, se define el Programa evaluado y se justifica su relevancia en el apartado 2, la base de datos BASEVAFOR se explica en el apartado 3, el estimador DID se define en el apartado siguiente en el que también se explicitan los métodos de cálculo alternativos así como la posibilidad de añadir regresores adicionales en el cálculo por

¹ Los autores agradecen el apoyo de esta investigación en el marco del Proyecto de investigación 2007/SEJ-132 "Teoría Económica y Economía Política". Asimismo, agradecen los comentarios y aportaciones de dos evaluadores anónimos.

regresión. Los resultados se obtienen y exponen en el apartado 5 mientras que las principales conclusiones se resumen en el último apartado.

2. EL PROGRAMA DE FORMACIÓN DE LAS ESCUELAS TALLER Y CASAS DE OFICIOS

2.1. Creación del programa de formación de las escuelas taller y casas de oficios

El programa de formación de las Escuelas Taller y Casas de Oficios se creó en 1985, como una medida de Fomento de Empleo Juvenil² con la finalidad de fomentar el aprendizaje de un oficio o profesión.

El programa comenzó a desarrollarse con carácter experimental y con un ámbito nacional bajo la dirección y control del Instituto Nacional de Empleo. Posteriormente, se lleva a cabo un proceso progresivo de transferencia de las competencias de su gestión a las Comunidades Autónomas correspondientes, de manera que, actualmente, es desarrollado por los Servicios Públicos Regionales de Empleo de cada Comunidad Autónoma.

En la actualidad, este programa se rige por una Orden del Ministerio de Trabajo y Asuntos Sociales³, y en el ámbito regional andaluz, por una Orden de la Consejería de Empleo y Desarrollo Tecnológico⁴.

2.2. Objetivos y beneficiarios del programa

El objetivo de este programa es insertar en el mercado de trabajo a jóvenes desempleados menores de veinticinco años, mediante la aplicación de métodos de cualificación y profesionalización que alternan conjuntamente la formación teórica y la práctica profesional a través de su participación en la realización de obras y servicios de utilidad pública o social.

Los beneficiarios de este programa deberán cumplir los siguientes requisitos:

- Haber cumplido 16 años y ser menor de 25 años.

² Vid. la Memoria sobre la situación socioeconómica y laboral elaborada por el Consejo Económico y Social (1995, 296).

³ O. M. del Ministerio de *Trabajo* y Asuntos Sociales, de 14 de noviembre de 2001, por la que se regulan el programa de Escuelas Taller y Casas de Oficios y las Unidades de Promoción y Desarrollo y se establecen las bases reguladoras de la concesión de subvenciones públicas a dichos programas.

⁴ Orden de 8 de marzo de 2004, de la Consejería de Empleo y Desarrollo Tecnológico, por la que se regulan los Programas de Escuelas Taller, Casas de Oficios, Talleres de Empleo y Unidades de Promoción y Desarrollo y se establecen las bases reguladoras de la concesión de ayudas públicas a dichos programas.

- Estar desempleado y ser demandante de empleo inscrito en los servicios públicos de empleo.
- Cumplir los requisitos establecidos para formalizar un contrato para la formación.

La implementación del programa se realiza en dos etapas, una primera de carácter formativo de iniciación, y una segunda etapa de formación en alternancia con el trabajo y la práctica profesional.

En el caso de las Escuelas Taller, la duración de ambas etapas no será inferior a un año ni superior a dos, dividida en fases de seis meses, mientras que en el caso de las Casas de Oficios, la duración de cada una de estas etapas queda establecida en seis meses.

2.3. La relevancia del Programa evaluado

En la *tabla 1* se recoge la evolución del número de proyectos dentro del Programa de Escuelas Taller y Casas de Oficios en toda España a lo largo del periodo 1995-2003. Asimismo, en la *tabla 2* se recoge la evolución del número de participantes en estos proyectos de Escuelas Taller y Casas de Oficios a lo largo del periodo 1995-2003, alcanzándose el máximo en el año 1998 con 60450 alumnos.

TABLA 1

Escuelas Taller y Casas de Oficios: número de proyectos por año (1995-2003).

	1995	1996	1997	1998	1999	2000	2001	2002	2003
Escuelas Taller	912	949	941	1043	1021	1106	973	971	907
Casas de Oficios	360	450	347	468	458	374	356	288	194
Total	1272	1399	1288	1511	1479	1480	1329	1259	1101

Fuente: INEM.

TABLA 2

Escuelas Taller y Casas de Oficios: número de participantes por año (1995-2003).

	1995	1996	1997	1998	1999	2000	2001	2002	2003
Alumnos de Escuelas Taller	37419	39883	40143	44636	42909	44868	36888	36404	32350
Alumnos de Casas de Oficios	11138	13544	10336	15814	15211	10214	9436	7345	7065
Total	48557	53427	50479	60450	58120	55082	46324	43749	39415

Fuente: INEM.

3. LA BASE DE DATOS

La base de datos utilizada para el desarrollo del proceso de evaluación ("**BASE-VAFOR 96-03**") ha sido construida a partir de observaciones de individuos participantes en los programas de formación desarrollados por las Escuelas Taller y Casas de Oficios en la provincia de Sevilla a lo largo de la década de los años noventa⁵.

Los datos han sido suministrados para esta investigación por el Instituto Nacional de Empleo (INEM) y por el Servicio Andaluz de Empleo (SAE). Para los individuos de ambos grupos, la base de datos recoge información relacionada con la vida laboral de los individuos, incluyendo datos referentes a los episodios de desempleo, a las solicitudes de demanda de empleo de los individuos, etc. También incluye información referente a las variables relevantes que van a ser utilizadas para el desarrollo de la evaluación, entre las que se consideran edad, sexo y zona geográfica de residencia.

Otros datos como el sector económico en el que los individuos encuentran empleo, el tipo de contrato realizado o la duración del mismo, no han podido incorporarse a la **BASEVAFOR 96-03**, aun cuando su tratamiento enriquecería las conclusiones de la investigación.

De entre todos los participantes, se ha seleccionado a aquellos individuos que han participado en un programa de los concluidos a lo largo del año 1999, último año con datos desglosados disponibles cuando se inició la construcción de esta base de datos.

El número total de participantes que forman este grupo es 1528 individuos. De entre estos individuos que constituyen la población objetivo, se ha seleccionado de forma aleatoria una muestra compuesta por 150 individuos que constituyen el grupo de participantes.

De forma paralela, se ha seleccionado también un grupo de control formado por 75 individuos, con características similares a los del grupo de participantes. El grupo de control ha sido elaborado para esta evaluación por la Delegación Provincial de Sevilla del Servicio Andaluz de Empleo.

Las proporciones establecidas, el 10% de la población objetivo para el grupo de participantes y la relación 2 a 1 entre ambos grupos, se corresponde con las mismas proporciones mantenidas en otros estudios previos tomados como referencia. Las dimensiones de la base de datos son homologables a las evaluaciones similares resumidas en Deheija y Wahba (1999, p. 1056 —Tabla 1—)⁶.

⁵ El nombre "**BASEVAFOR 96-03**" corresponde a las siglas de "Base para la evaluación de programas de formación", circunscrita a los años 1996 y 2003 ya que, por un lado, incluye información referida a cursos finalizados a lo largo del año 1999, y, por tanto, iniciados en los años 1997 y 1998, incluyendo, además, información de los individuos referida al año previo, 1996. Por otro lado, la información que incluye relativa a la variable respuesta se prolonga hasta el año 2003.

⁶ La relación 2/1 nos sitúa en la línea de estudios de evaluación previos como es el del Programa JTPA. Este programa asociado a la "*Job Training Partnership Act —JTPA—*" estadounidense fue

4. EL ESTIMADOR ‘DID’ DE DIFERENCIAS EN DIFERENCIAS

4.1. Definición de la variable “ D ”

Definimos D como la variable binaria indicativa de la participación de los individuos en el programa evaluado, de manera que

- $D_i = 1$ indicará que el individuo i ha participado en el programa de formación.
- $D_i = 0$ indicará que el individuo i es un individuo del grupo de control.

4.2. Definición de la variable “ Y ”

Se define la variable Y_i como la “capacidad del individuo i -ésimo para encontrar un empleo”, de la siguiente forma:

$$Y_i = 1 - \frac{\text{Nº de días transcurridos hasta que el individuo "i" encuentra un empleo}}{\text{Tiempo total del periodo de observación considerado}} \quad (1)$$

La variable “ Y ” se comporta como un indicador del tiempo relativo invertido en encontrar un empleo. El periodo de observación considerado será de tres años (1095 días).

De esta forma, el rango de variación de Y_i queda establecido entre 0 y 1. Un valor de Y_i próximo a 1 indicará que el individuo i ha encontrado empleo de forma rápida, mientras que un valor de Y_i próximo 0 implicará que el individuo i ha necesitado un amplio periodo de tiempo para encontrar empleo.

La elección de esta variable como variable respuesta o “outcome” se justifica en que los individuos que constituyen la muestra, tanto los participantes como los individuos de control, parten de una situación de desempleo, encontrándose oficialmente inscritos como demandantes de empleo.

Dado que mayoritariamente, por su corta edad y su bajo nivel de cualificación, aún no han tenido ningún empleo previo o cuentan con una escasa experiencia laboral, resulta oportuno considerar una variable respuesta que mida la capacidad de este colectivo concreto para encontrar un empleo.

En la *tabla 3* se resumen los principales datos de estadística descriptiva de la variable Y , mientras que en el *gráfico 1* se representa el histograma con la distribución de frecuencias de esta variable.

objeto de muy influyentes estudios evaluatorios de entre los que destacan los de Heckman, Ichimura y Todd (1997, 1998) y el de Heckman, Ichimura, Smtih y Todd (1998).

TABLA 3

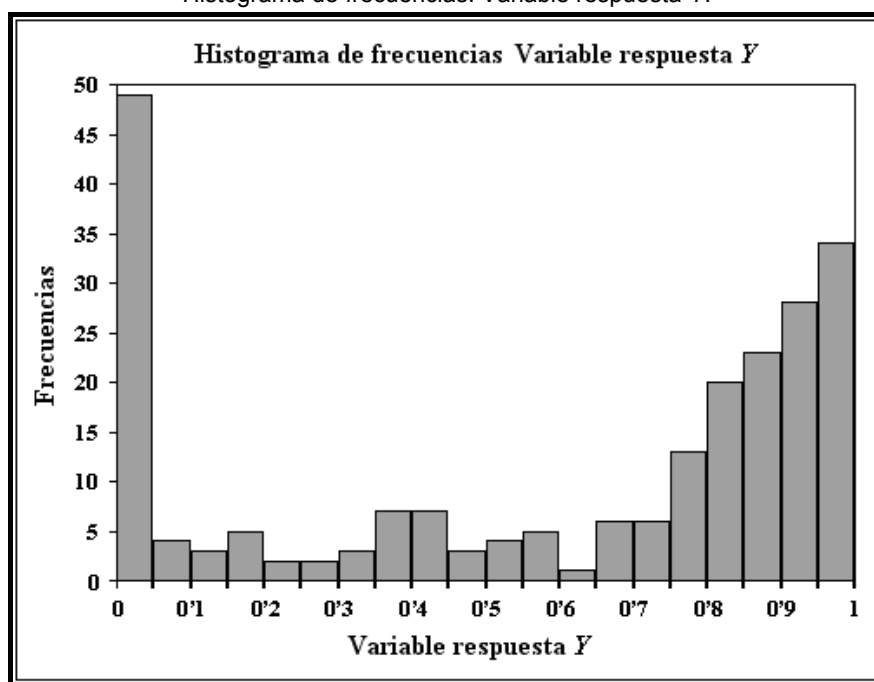
Estadística descriptiva variable respuesta Y "Capacidad de inserción laboral".

Media	0,578147	Curtosis	-1,392020
Mediana	0,773516	Coefficiente de asimetría	-0,533554
Moda	0	Mínimo	0
Desviación típica	0,380657	Máximo	1

Fuente: Elaboración propia.

GRÁFICO 1

Histograma de frecuencias. Variable respuesta Y.



Fuente: Elaboración propia.

4.3. Las variables contaminantes

La heterogeneidad de las poblaciones se sustancia en que los individuos beneficiarios y los de control habitualmente difieren en características que actúan en la evaluación como variables contaminantes o covariables. Son variables ajenas al tratamiento que pueden influir ("*contaminar*") sobre el valor de la variable respuesta Y, añadiendo sus efectos a los producidos por D. Este aspecto, crucial en ciencias

sociales, tiene implicaciones determinantes en la toma de decisiones públicas tal y como ha señalado Manski (2001).

Muchas de estas covariables o variables contaminantes podrán ser observables, y podría aislarse su efecto utilizando métodos adecuados de selección sobre variables.

Con frecuencia, no obstante, el desarrollo de la evaluación deja traslucir que los individuos difieren en características que no son observables como la habilidad, destreza, capacidad relacional, etc. En estos casos, los métodos de selección sobre observables no resultan del todo adecuados para estimar los efectos causales del programa de formación.

Sin embargo, los estimadores basados en el método de “*diferencias en diferencias*” permiten salvar esta dificultad al considerar los resultados observados de la variable respuesta Y , tanto antes de someterse al programa como después de finalizarlo, permitiendo obtener un estimador de los efectos del programa sobre los participantes ($\hat{\alpha}_{ATET}$). Se asume que las observaciones pre-programa y post-programa de la variable respuesta Y estarán igualmente contaminadas a causa de las variables diferenciadoras no observables.

4.4. El estimador ‘DID’ de diferencias en diferencias

El estimador DID utiliza datos longitudinales, por tanto, se considera T , como variable binaria indicativa del momento del tiempo en el que son observados los valores de la variable respuesta Y , que podrá ser, bien en un momento previo al programa evaluado ($T = 0$) o bien en un momento posterior al programa ($T = 1$).

Subyacente a este planteamiento está el supuesto de que los resultados promedios de los individuos participantes y de control habrían seguido una evolución paralela en caso de no existir el programa. Así, siguiendo a Athey e Imbens (2002, p. 1), la asunción subyacente de este método es que la tendencia temporal experimentada por los individuos de control se utiliza como aproximación de la tendencia temporal que habrían registrado los individuos participantes en ausencia del programa.

Los valores de la variable resultado se indexan oportunamente. Así, siguiendo la notación de Abadie (2005, pp. 7 y ss.) definimos⁷:

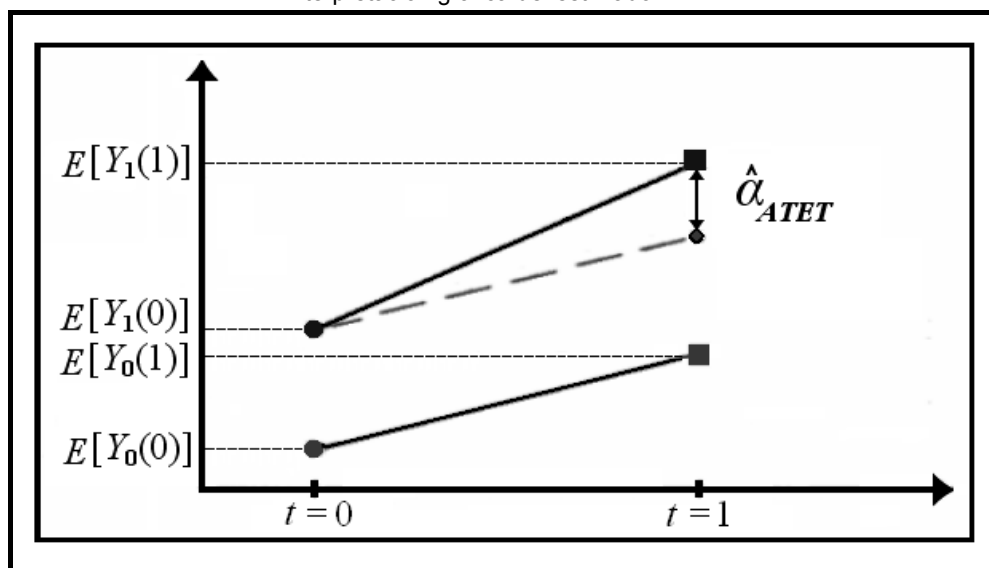
- $Y_1(1)$: Valor de la variable respuesta Y en un momento posterior al programa para un individuo del grupo de participantes ($D = 1, T = 1$).
- $Y_0(1)$: Valor de la variable respuesta Y en un momento posterior al programa para un individuo del grupo de control ($D = 0, T = 1$).
- $Y_1(0)$: Valor de la variable respuesta Y en un momento previo al programa para un individuo del grupo de participantes ($D = 1, T = 0$).

⁷ En relación con el desarrollo metodológico de este estimador, véanse también los trabajos de Card y Krueger (1994), Imbens, Liebman y Eissa (1998) y Angrist y Krueger (1999).

- $Y_0(0)$: Valor de la variable respuesta Y e en un momento previo al programa para un individuo del grupo de control ($D = 0, T = 0$).

En el *gráfico 2* se recoge la interpretación gráfica del estimador DID.

GRÁFICO 2
Interpretación gráfica del estimador DID.



Fuente: Elaboración propia.

Siguiendo a Heckman, Ichimura, Smith y Todd (1998, p. 1020)⁸, el cálculo del estimador DID permite calcular la diferencia de la variable respuesta de los individuos beneficiarios entre sus valores antes y después de dicho programa, y volver a calcular la diferencia con respecto a la variación producida en la variable respuesta de los individuos de control⁹.

De esta forma, el estimador del efecto promedio del programa de formación sobre los participantes ($\hat{\alpha}_{ATET}$) queda reducido a una diferencia de diferencias¹⁰.

⁸ Nos remitimos también a la exposición del estimador de diferencias en diferencias en Heckman y Robb (1985).

⁹ Como ejemplo de la aplicación de este modelo, aparte de los trabajos de Card y Krueger (1994, 2000), cabe mencionar el estudio de Card (1990) en el que el autor evalúa el efecto del fuerte incremento de la inmigración sobre el mercado laboral de Miami, aplicando para ello los fundamentos del estimador DID y utilizando como grupo de control a otras cuatro ciudades con una tendencia similar en su tasa de desempleo (Atlanta, Los Ángeles, Houston y Tampa-St. Petersburg).

¹⁰ De lo anterior se deduce la aplicación del estimador de diferencias en diferencias como un estimador que surge del análisis de políticas con datos para dos periodos de tiempo. Es decir, estimamos el efecto causal del programa utilizando datos de panel para sólo dos periodos de tiempo. Una ver-

4.5. Obtención del estimador DID mediante regresión por MCO

De manera alternativa a la especificación de Heckman, Ichimura, Smith y Todd (1998), es posible calcular el valor del estimador DID mediante regresión por MCO, en la línea de los trabajos de Ashenfelter y Card (1985) y Abadie (2005), a partir de un modelo lineal del tipo:

$$Y = \mu + \lambda D + \omega T + \alpha DT + \varepsilon \quad (2)$$

En el modelo se definen las variables que intervienen como:

- Y : variable respuesta.
- D : variable indicativa de la participación en el programa.
- T : variable indicativa del momento de observación.

Los parámetros asociados son:

- λ, ω : parámetros asociados a las variables explicativas D y T .
- μ : parámetro indicador de los efectos fijos en el modelo.
- ε : error aleatorio de media cero: $E[\varepsilon | D, T] = 0$.
- α : parámetro que determina el efecto conjunto de las variables explicativas D y T .

El parámetro α será el estimador DID del efecto promedio sobre los participantes —*ATET*—. Su obtención mediante MCO a partir de este modelo lineal sigue una fórmula similar a la propuesta por Heckman, Ichimura, Smith y Todd (1998), siendo este estimador un estimador insesgado.

La obtención del estimador de *ATET* de “*diferencias en diferencias*” por el procedimiento de regresión por MCO permite, además, la posibilidad de incluir en el modelo la consideración de variables contaminantes o covariables observables.

La inclusión de estas covariables en el modelo se realiza siguiendo la expresión:

$$Y = \mu + \lambda D + \omega T + \alpha (DT) + \beta X + \varepsilon \quad (3)$$

sión de este estimador se aplica a la combinación de cortes transversales independientes, de igual modo que otra versión derivada de dicho estimador se aplica a conjuntos de datos de panel. Para un desarrollo ampliado en relación a datos de panel nos remitimos a Arellano (2003) y Wooldridge (2002). A este respecto, como expone Wooldridge (2002, p. 637), la disponibilidad de datos de panel nos permitiría obtener un estimador consistente del efecto del programa sin necesidad de demostrar la condición de independencia (dado un vector de covariables X), y sin la necesidad de tener que recurrir a variables instrumentales, siempre y cuando el programa de formación se desarrolle a lo largo de un periodo de tiempo determinado y no esté correlacionado con covariables predeterminadas que afecten a la variable respuesta. Esta condición de independencia, denominada “*ignorability of treatment*”, es introducida por Rosenbaum y Rubin (1983), según la cual, condicionada sobre el vector X de covariables, D (variable que indica la participación en el programa) e Y (variable respuesta) son independientes.

Con $E[\varepsilon | D, X] = 0$ y donde X es la variable o vector de variables contaminantes o covariables y β el parámetro o vector de parámetros asociados a la variable o vector de variables contaminantes.

Siempre que el valor de X permanezca constante entre los periodos 0 y 1, la incorporación de covariables al modelo no limita su robustez¹¹. Así, por un lado, se controlan los efectos sobre la variable respuesta Y introducidos por las variables contaminantes, efectos debidos a características observables, y, por otro lado, mediante la realización del proceso de “diferencias en diferencias” se controlan los efectos provocados por características no observables.

De esta forma, siguiendo a Stock y Watson (2003, p. 387 y ss.) se obtiene el estimador DID con regresores adicionales, que, cumpliéndose los supuestos necesarios¹² para la regresión múltiple por MCO, será un estimador insesgado y consistente.

La incorporación de los regresores adicionales al modelo permite, por un lado, una mejora en la eficiencia del estimador, ya que se reduce la varianza del error aleatorio, y por otro, testar la aleatoriedad en la asignación de los individuos a los grupos de participantes y de control. De esta forma, incluyendo estas covariables en el modelo se estaría controlando la probabilidad de asignación de los individuos al grupo de control condicionada por estas covariables.

A partir del planteamiento del modelo del estimador de diferencias en diferencias expuesto, se han ido desarrollando otros trabajos que incorporan algunas especificidades a este modelo inicial. Tales son los casos de los trabajos¹³ de Donald y Lang (2001) y Hansen (2003), que extienden el desarrollo del modelo al tratamiento con datos de panel para un pequeño número de grupos de individuos.

4.6. Definición del vector de covariables “ $X^3 = (X_1, X_2, X_3)$ ”

Definida la variable D , X será una variable contaminante o covariable con respecto a D si, para todos y cada uno de los individuos observados, su valor se mantiene constante para cualquier valor que pueda adoptar D .

¹¹ Cuando el valor de X varía entre los periodos 0 y 1, probablemente X resultará afectada por el programa, por tanto, si se incorpora al modelo sin ninguna consideración adicional afectará el resultado final del estimador del *ATE* introduciendo endogenidad en el modelo. Para solventar esta dificultad y eliminar los problemas de endogenidad incorporados por la covariable afectada por el tiempo se introducirá entonces X en el modelo de forma conjunta con la variable binaria T (indicativa del momento del tiempo en el que el individuo es observado). Esto se llevaría a cabo de acuerdo con la expresión:

$$Y = \mu + \lambda D + \omega T + \alpha (DT) + \beta_1 (TX) + \beta_0 ((1-T)X) + \varepsilon$$

¹² De acuerdo con Stock y Watson (2003, p. 159), estos supuestos son cuatro: a) La distribución condicional del error aleatorio ε_i dadas las variables explicativas $X_{1i}, X_{2i}, \dots, X_{ki}$ tiene media cero (en nuestro caso concreto, estas variables explicativas serían la variable D_i indicativa de la participación en el programa junto con las covariables incluidas en el modelo). b) todas las observaciones $i = 1, 2, \dots, n$ se distribuyen de forma independiente e idénticamente aleatoria. c) $X_{1i}, X_{2i}, \dots, X_{ki}$ y ε_i tienen cuatro momentos. d) No existe multicolinealidad perfecta.

¹³ Nos remitimos igualmente el estudio de Bertrand, Duflo y Mullainathan (2002).

A partir de la información muestral contenida en la base de datos, se consideran tres variables predeterminadas, que formarán el vector de covariables $X^3 = (X_1, X_2, X_3)$, definidas de la siguiente forma:

- X_1 : sexo. Variable cualitativa que tomará los valores $X_1 = 1$ si el individuo observado es varón y $X_1 = 2$ si es mujer.

La covariable X_1 , al tratarse de una variable cualitativa, se incluye en el modelo a partir de la construcción de variables ficticias. Para evitar la presencia de multicolinealidad perfecta, se introducen en el modelo tantas variables ficticias como categorías posibles tenga la variable cualitativa menos una. De esta forma, para considerar la covariable X_1 , se incluirá la variable ficticia X_{11} , que podrá tomar los valores 0 y 1.

$$X_{11} = \begin{cases} 1 & \text{si el individuo es hombre} \\ 0 & \text{si el individuo es mujer} \end{cases}$$

- X_2 : edad. Variable cuantitativa que recoge la edad del individuo al comienzo del periodo de observación, que variará entre los valores correspondientes a 17 y 26 años.

Para considerar la covariable X_2 , al tratarse de una variable cuantitativa, que podrá tomar valores que oscilarán entre 17 y 26 años, se incluirá como tal, sin necesidad de realizar ninguna adecuación previa.

- X_3 : zona. Variable cualitativa que muestra la zona de realización del curso o de residencia para el caso de los individuos de control.

Para definir esta covariable se ha dividido la provincia de Sevilla en cuatro zonas, siguiendo un criterio de zonificación meramente operativo: zona 1 (Sevilla ciudad), zona 2 (zona este y noreste de la provincia), zona 3 (zona sur y suroeste) y zona 4 (zona oeste y noroeste). Así, esta variable tomará los valores $X_3 = 1$ si el individuo pertenece a la zona 1, $X_3 = 2$ si pertenece a la zona 2, $X_3 = 3$ si pertenece a la zona 3 y $X_3 = 4$ si pertenece a la zona 4. Al tratarse de una variable cualitativa, se incluirá en el modelo a partir de la construcción de tres variables ficticias, X_{31} , X_{32} y X_{33} , que podrán tomar los valores 0 y 1, según se definen a continuación.

$$X_{31} = \begin{cases} 1 & \text{si el individuo pertenece a la zona 1} \\ 0 & \text{si el individuo pertenece a cualquiera de las otras tres zonas} \end{cases}$$

$$X_{32} = \begin{cases} 1 & \text{si el individuo pertenece a la zona 2} \\ 0 & \text{si el individuo pertenece a cualquiera de las otras tres zonas} \end{cases}$$

$$X_{33} = \begin{cases} 1 & \text{si el individuo pertenece a la zona 3} \\ 0 & \text{si el individuo pertenece a cualquiera de las otras tres zonas} \end{cases}$$

5. RESULTADOS EMPÍRICOS

5.1. Cálculo del estimador DID

A partir de la información muestral contenida en la base de datos y una vez contruidos los grupos de participantes y de control, se procede a la obtención del estimador DID de acuerdo con las especificaciones realizadas.

Para ello, definida la variable respuesta Y de acuerdo con la *Expresión (1)*, será preciso matizar sus posibles valores en relación con la variable T .

- $Y_{1i}(1)$ e $Y_{0i}(1)$ harán referencia a los valores de la variable respuesta Y en un momento posterior a la realización del programa ($T=1$). Este momento se fijará en momento en el que ha transcurrido el periodo de observación considerado para la construcción de la variable respuesta Y . Luego los valores de $Y_{1i}(1)$ e $Y_{0i}(1)$ corresponderán a los valores obtenidos de la variable respuesta Y de acuerdo con la aplicación de la *Expresión (1)*.
- $Y_{1i}(0)$ e $Y_{0i}(0)$ harán referencia a los valores que tomará la variable respuesta Y en un momento previo al programa, justo antes de iniciarse el mismo. En el momento $T=0$, los individuos observados se encuentran desempleados, por lo que habrá que considerar, entonces, que el valor de la variable respuesta Y , de acuerdo con su definición, será igual a cero.

De acuerdo con las especificaciones previas, en las *tablas 4 y 5* se recoge el valor del estimador DID. En la *tabla 4* se obtiene este estimador siguiendo el procedimiento de Heckman, Ichimura, Smith y Todd (1998). Por otra parte, en la *tabla 5* se muestran los resultados de la regresión realizada para la obtención del estimador por MCO siguiendo el modelo planteado.

El valor del estimador mediante regresión (*Tabla 4*) coincide con el valor obtenido según la expresión analítica (*Tabla 5*); en ambos casos es igual a 0'451659.

Este valor hace referencia al efecto causal promedio producido por el programa de formación sobre la variación experimentada en la variable respuesta Y de los individuos participantes entre los dos periodos considerados.

Es decir, los individuos participantes en el programa han experimentado una mejora adicional, por un montante de 0'451659, en la evolución de su capacidad de inserción laboral entre los momentos 0 y 1 por encima de la evolución experimentada por los individuos de control.

TABLA 4
Cálculo del estimador DID mediante expresión analítica.

Grupo de participantes	Grupo de control
$n_1 = 150$	$n_0 = 75$
1ª Diferencia: Diferencias en el Grupo de participantes	2ª Diferencia: Diferencias en el Grupo de control
$\frac{1}{n_1} \sum_{i=1}^{n_1} \{Y_i(1) - Y_i(0)\} = 0,728700$	$\frac{1}{n_0} \sum_{i=1}^{n_0} \{Y_i(1) - Y_i(0)\} = 0,277041$
DIFERENCIAS EN DIFERENCIAS	
$\hat{\alpha}_{ATET} \rightarrow DID = \frac{1}{n_1} \sum_{i=1}^{n_1} \{Y_i(1) - Y_i(0)\} - \frac{1}{n_0} \sum_{i=1}^{n_0} \{Y_i(1) - Y_i(0)\} = \mathbf{0,451689}$	

Fuente: Elaboración propia.

TABLA 5
Cálculo del estimador DID mediante regresión.

Variable Dependiente: Y (Capacidad de inserción laboral)					
Método: MCO					
Variable	Coefficiente	Valor del coeficiente	Error estándar	t-estadístico*	Probabilidad
Efecto fijo	μ	$-1,26 \cdot 10^{-16}$	$2,96 \cdot 10^{-17}$	-4,245819	0,0000
D	λ	$1,78 \cdot 10^{-16}$	$2,30 \cdot 10^{-17}$	7,707609	0,0000
T	ω	0,277041	0,039727	6,973641	0,0000
D T	α	0,451659	0,046709	9,669537	0,0000
Coef. Determ. (R^2)		0,682309	Media var. dep.		0,289074
R^2 -ajustado		0,680172	Desv. típ. var. dep.		0,395017
Error estándar regresión		0,223395	Criterio info. Akaike		-0,150902
Suma errores al cuadrado		22,25777	Criterio de Schwarz		-0,114375
Log. función verosimil.		37,95286	F-estadístico		319,2939
Estad. de Durbin-Watson		2,007492	Prob. (F-estadístico)		0,000000

* Valor del estadístico t corregido por el método de White.

Fuente: Elaboración propia.

Con respecto a la regresión mediante MCO, y de acuerdo con la información recogida en la *tabla 5*, la ecuación estimada por el modelo es la siguiente:

$$Y = -1,26 \cdot 10^{-16} + 1,78 \cdot 10^{-16} \mathbf{D} + 0,277041 \mathbf{T} + 0,451659 \mathbf{DT}$$

La significatividad individual de las variables explicativas incluidas en el modelo queda de manifiesto a partir de los valores obtenidos para el *t*-estadístico y su

probabilidad asociada. Estos valores obtenidos permiten rechazar la hipótesis nula de que el valor de cada uno de los coeficiente incluidos en el modelo, considerados individualmente, sean igual a cero ($H_0: \alpha = 0$; frente a $H_1: \alpha \neq 0$). Luego no hay indicios para rechazar la hipótesis de significatividad de las variables explicativas incluidas en el modelo. En la regresión, los valores del t -estadístico aparecen ajustados por el método de White, para corregir los posibles problemas de heterocedasticidad.

Con respecto a la bondad del ajuste, el coeficiente de determinación (R^2) alcanza un valor de 0'682309, lo que implica que la capacidad explicativa de la variables incluidas en el modelo, es del 68,2309%, con lo cual, el modelo tiene una capacidad explicativa importante. Por su parte, el R^2 ajustado es igual a 0,680172. El valor del estadístico de Durbin-Watson muy próximo a 2, casi igual a 2 (2,007492), indica que los problemas de autocorrelación de los residuos no son relevantes. Por último, también es posible testar la significatividad global de los parámetros estimados mediante el test de Wald, cuyos resultados, recogidos en la *tabla 6*, no permiten rechazar la significación conjunta de los parámetros del modelo.

TABLA 6
Test de Wald (estimación DID mediante regresión).

Hipótesis nula: $\lambda = 0$, $\omega = 0$, $\alpha = 0$			
F-estadístico	836,5499	Probabilidad	0,000000
Chi-cuadrado	2509,650	Probabilidad	0,000000

Fuente: Elaboración propia.

5.2. Cálculo del estimador DID con regresores adicionales

Adicionalmente, es posible incluir en el cálculo del estimador DID la consideración de variables contaminantes mediante la inclusión de regresores adicionales en el modelo planteado. De esta forma, es posible controlar de forma específica y por separado, por un lado, los efectos sobre la variable respuesta Y introducidos por las variables contaminantes observables, y, por otro lado, mediante la realización del proceso de “*diferencias en diferencias*” se recoge el efecto debido a características no observables. De esta forma se puede precisar y ajustar aún más el valor del efecto del programa de formación.

Las covariables que se incorporarán al modelo como regresores adicionales son las incluidas en el vector de covariables $X^3 = (X_1, X_2, X_3)$ definido. Para tener en consideración la posible existencia de multicolinealidad en el modelo, será preciso analizar previamente la matriz de correlación. La *tabla 7* recoge la matriz de correlación entre las variables explicativas incluidas en el modelo.

TABLA 7
Matriz de correlación entre las variables incluidas en el modelo de
"diferencias en diferencias" con regresores adicionales.

	<i>D</i>	<i>T</i>	<i>DT</i>	<i>X</i> ₁₁	<i>X</i> ₂	<i>X</i> ₃₁	<i>X</i> ₃₂	<i>X</i> ₃₃
<i>D</i>	1,0000	$2,1 \cdot 10^{-17}$	0,5000	0,1949	0,1688	-0,0077	0,0069	0,0279
<i>T</i>	$2,1 \cdot 10^{-17}$	1,0000	0,7071	0,0000	0,0000	$-3 \cdot 10^{-19}$	$1,6 \cdot 10^{-18}$	$1,1 \cdot 10^{-18}$
<i>DT</i>	0,5000	0,7071	1,0000	0,0974	0,0844	-0,0039	0,0034	0,0139
<i>X</i> ₁₁	0,1949	0,0000	0,0974	1,0000	0,0055	0,2182	0,0147	-0,0928
<i>X</i> ₂	0,1688	0,0000	0,0844	0,0055	1,0000	0,1437	-0,1056	0,0307
<i>X</i> ₃₁	-0,0077	$-3 \cdot 10^{-19}$	-0,0039	0,2182	0,1437	1,0000	-0,3346	-0,3240
<i>X</i> ₃₂	0,0069	$1,6 \cdot 10^{-18}$	0,0034	0,0147	-0,1056	-0,3346	1,0000	-0,4106
<i>X</i> ₃₃	0,0279	$1,1 \cdot 10^{-18}$	0,0139	-0,0928	0,0307	-0,3240	-0,4106	1,0000

Fuente: Elaboración propia.

No se aprecia correlación lineal entre las variables, al ser todos los coeficientes muy bajos y alejados de ± 1 , salvo en el caso de la relación entre las variables *D* y *T* con *DT*, casos en los que se registran valores de los coeficientes algo superiores, debido a que *DT* recoge la interacción de ambas variables. Por otro lado, el valor del determinante de la matriz de correlación es 0,1116, no demasiado próximo a cero. Todo lo expuesto indica que los problemas de multicolinealidad no van a ser relevantes en el modelo.

Una vez hechas todas las especificaciones precisas, es posible llevar a cabo la regresión por MCO incorporando los regresores adicionales. Los resultados aparecen en la *tabla 8*.

TABLA 8
Cálculo del estimador DID con regresores adicionales.

Variable Dependiente: Y (Capacidad de inserción laboral)					
Método: MCO					
Variable	Coeficiente	Valor del coeficiente	Error estándar	t-estadístico*	Probabilidad
Efecto fijo	μ	-0,241568	0,108743	-2,221458	0,0268
D	λ	-0,021243	0,008819	-2,408794	0,0164
T	ω	0,277041	0,039102	7,085034	0,0000
D T	α	0,451659	0,046065	9,804869	0,0000
X₁₁	β_{11}	0,064382	0,023084	2,788982	0,0055
X₂	β_2	0,011823	0,005137	2,301793	0,0218
X₃₁	β_{31}	-0,020468	0,033567	-0,609779	0,5423
X₃₂	β_{32}	-0,014091	0,030144	-0,467465	0,6404
X₃₃	β_{33}	-0,037950	0,030254	-1,254369	0,2104
Coef. Determ. (R^2)		0,693073	Media var. dep.		0,289074
R^2 -ajustado		0,687506	Desv. típ. var. dep.		0,395017
Error estándar regresión		0,220819	Criterio info. Akaike		-0,163149
Suma errores al cuadrado		21,50363	Criterio de Schwarz		-0,080964
Log. función verosimil.		45,70849	F-estadístico		124,4782
Estad. de Durbin-Watson		2,022655	Prob. (F-estadístico)		0,000000

* Valor del estadístico t corregido por el método de White.

Fuente: Elaboración propia.

A partir de los valores obtenidos para el t -estadístico y su probabilidad asociada aparecen como significativas todas las variables salvo las variables ficticias construidas para incluir la covariable zona, X_{31} , X_{32} y X_{33} . Para contrastar la no significatividad de estas tres variables se aplica el test de Wald, cuyos resultados, recogidos en la *tabla 9*, no permiten rechazar la hipótesis nula $H_0: \beta_{31} = \beta_{32} = \beta_{33} = 0$.

TABLA 9
Test de Wald (estimación DID con regresores adicionales:
significatividad covariables zona X_{31} , X_{32} y X_{33}).

Hipótesis nula: $\beta_{31} = 0$, $\beta_{32} = 0$, $\beta_{33} = 0$			
F-estadístico	0,567509	Probabilidad	0,636673
Chi-cuadrado	1,702527	Probabilidad	0,636372

Fuente: Elaboración propia.

Se eliminan, por tanto, las variables ficticias construidas para incluir la covariable zona (X_{31} , X_{32} y X_{33}) y se lleva a cabo de nuevo la regresión por MCO sin incluir estas tres variables. Los resultados aparecen en la *tabla 10*.

TABLA 10
Cálculo del estimador DID con regresores adicionales.

Variable Dependiente: Y (Capacidad de inserción laboral)					
Método: MCO					
Variable	Coefficiente	Valor del coeficiente	Error estándar	t-estadístico*	Probabilidad
Efecto fijo	μ	-0,250496	0,101040	-2,479181	0,0135
D	λ	-0,021743	0,008282	-2,625271	0,0090
T	ω	0,277041	0,039326	7,044792	0,0000
D T	α	0,451659	0,046176	9,781213	0,0000
X₁₁	β_{11}	0,063843	0,022369	2,854019	0,0045
X₂	β_2	0,011348	0,004976	2,280709	0,0230
Coef. Determin. (R^2)		0,691892	Media var. dep.		0,289074
R^2 -ajustado		0,688423	Desv. típ. var. dep.		0,395017
Error estándar regresión		0,220495	Criterio info. Akaike		-0,172642
Suma errores al cuadrado		21,58637	Criterio de Schwarz		-0,117852
Log. función verosimil.		44,84440	F-estadístico		199,4110
Estad. de Durbin-Watson		2,001027	Prob. (F-estadístico)		0,000000

* Valor del estadístico t corregido por el método de White.

Fuente: Elaboración propia.

Como en las estimaciones anteriores, los valores del *t*-estadístico aparecen ajustados por el método de White, para corregir los posibles problemas de heterocedasticidad.

El valor del estimador DID sigue siendo en este caso 0,451659, pero ahora, mediante la inclusión de los regresores se obtiene también por separado el efecto de las covariables identificadas. De acuerdo con las especificaciones del modelo, como el valor de las covariables introducidas permanece constante entre los periodos 0 y 1, la incorporación de éstas al modelo no afecta al valor obtenido del parámetro asociado a la variable *T* (ω) ni al resultado final del estimador del *ATET* (α). En caso contrario, si los valores de estos dos coeficientes se vieran afectados, se estaría evidenciando la introducción de endogenidad en el modelo con la incorporación de los regresores adicionales.

En resumen, mediante la introducción de los regresores adicionales, se controlan los efectos sobre la variable respuesta provocados por las variables contaminantes identificadas, y, mediante la realización del proceso de “*diferencias en diferencias*”, se controlan los efectos provocados por características no observables.

Los valores estimados de los parámetros que acompañan a las covariables recogen los efectos de éstas sobre la variación experimentada por la variable respuesta *Y* entre los momentos 0 y 1. De esta forma:

- El coeficiente β_{11} , asociado a X_{11} , recoge el efecto producido por la covariable sexo.

Teniendo en cuenta que es una variable binaria muestra que, cuando el individuo considerado es hombre ($X_{11} = 1$), la variación experimentada por la variable respuesta Y entre los momentos 0 y 1 se ve incrementada en 0,064382. En caso de que el individuo sea mujer ($X_{11} = 0$), categoría tomada como base, este efecto no se añade.

Por lo tanto, se evidencia cómo en el caso de los hombres, éstos presentan una mejor evolución de su capacidad de inserción laboral. Concretamente, el incremento en su capacidad de inserción laboral se ve mejorado en 0,064382.

- El coeficiente β_2 , asociado a X_2 , recoge el efecto producido por la covariable edad.

Al ser ésta una variable cuantitativa y dado que el valor estimado para el coeficiente β_2 es de 0,011823, este efecto se irá incrementado en función de los distintos valores que adopte la variable edad, añadiéndose 0,011823 a la variación de la variable respuesta Y por cada año adicional recogido por la variable X_2 .

De esta forma, dentro del grupo de individuos considerado, que tendrán entre 17 y 25 años, se evidencia cómo éstos van a presentar una mayor variación en su capacidad de inserción laboral cuanto mayor sea su edad.

Con respecto a la bondad del ajuste, el coeficiente de determinación (R^2) es 0,693073, por lo que, el modelo tiene una capacidad explicativa importante. Por su parte, el R^2 ajustado es 0,687506. El valor del estadístico de Durbin-Watson muy próximo a 2 (2,022655), indica que los problemas de autocorrelación de los residuos no son relevantes. Comparándolo con la estimación sin regresores adicionales, este modelo mejora la bondad del ajuste ya que registra unos menores valores de los criterios de información de Akaike y Schwarz y un mayor valor del R^2 ajustado y de la función de verosimilitud.

Finalmente, también es posible testar la significatividad global de los parámetros estimados mediante el test de Wald, cuyos resultados, recogidos en la *tabla 11*, no permiten rechazar de la significación conjunta de los parámetros del modelo.

TABLA 11

Test de Wald (estimación DID con regresores adicionales).

Hipótesis nula: $\lambda = 0$, $\omega = 0$, $\alpha = 0$, $\beta_{11} = 0$, $\beta_2 = 0$			
F-estadístico	225,4452	Probabilidad	0,000000
Chi-cuadrado	1127,226	Probabilidad	0,000000

Fuente: Elaboración propia.

6. CONCLUSIONES

Con un valor de 0,451659 para el estimador DID, puede concluirse una evaluación favorable de la eficacia del programa de Escuelas Taller y Casas de Oficio, en el ámbito geográfico y temporal investigado.

Con este resultado, un individuo participante emplea, en promedio, 297,07 días en encontrar un empleo, esto es, 494,57 días menos que un no participante.

Los resultados de la evaluación permiten llegar a conclusiones adicionales cuando se incluyen las variables contaminantes de las que contiene información la BASEVAFOR 96-03. Así, si el individuo participante es varón, el periodo invertido en encontrar un empleo se reduce en 70,49 días en comparación con el tiempo que tendría que invertir si el participante fuese mujer.

La edad del participante también reduce, progresivamente, el tiempo necesario hasta encontrar un empleo. De esta forma, por cada año adicional de edad del participante, el periodo de tiempo necesario para encontrar empleo se reduce en 12,94 días.

Consecuencia de lo anterior es que la máxima efectividad del programa evaluado se alcanza en los individuos participantes, varones y con 25 años de edad.

La covariable que toma en cuenta el lugar de residencia, no resulta significativa, dados los valores tomados por el *t*-estadístico y su probabilidad asociada.

Un desarrollo futuro de esta evaluación debería beneficiarse de una ampliación de la información contenida en la BASEVAFOR 96-03, que permitiera considerar un mayor número de covariables que hiciesen referencia, por ejemplo, al sector económico en el que los individuos encuentran empleo o la duración de éste.

REFERENCIAS BIBLIOGRÁFICAS

- ABADIE, A. (2005): "Semiparametric Difference-In-Differences Estimators", *Review of Economic Studies*, 72 (1): 1-19.
- ANDREWS, M.; BRADLEY, S. y UPWARD, R. (1999): "Estimating Youth Training Wage Differentials During and After Training", *Oxford Economic Papers*, 51 (3): 517-544.
- ANGRIST, J.D. y KRUEGER, A.B. (1999): "Empirical Strategies in Labor Economics", en O. Ashenfelter y D. Card, D. (eds.), *Handbook of Labor Economics*, vol. 3 (1), capítulo 23: 1277-1366, North-Holland, Amsterdam, Ed.: Elsevier Science.
- ARELLANO, M. (2003): *Panel Data Econometrics*, Oxford (etc.), Ed.: Oxford University Press.
- ASHENFELTER, O. y CARD, D. (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs", *Review of Economics and Statistics*, 67 (4): 648-660.
- ATHEY, S. e IMBENS, G.W. (2002): "Identification and Inference in Nonlinear Difference-in-Differences Models". *NBER Technical Working Paper n° 280*, septiembre 2002.
- BERGEMANN, A.; FITZENBERGER, B. y SPECKESSER, S. (2005): *Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences*, Discussion Papers n° 1848 from Institute for the Study of Labor (IZA), Noviembre, 2005.

- BERTRAND, M.; DUFLO, E. y MULLAINATHAN, S. (2002): "How Much Should We Trust Differences-in-Differences Estimates?". *NBER Working Paper n° 8841*, Marzo 2002.
- BLUNDELL, R.; COSTA, M.; MEGHIR, C. y VAN REENEN, J. (2004): "Evaluating the Employment Impact of a Mandatory Job Search Program", *Journal of the European Economic Association*, 2 (4): 569-606.
- BONNALL, L.; FOUGÈRE, D. y SÉRANDON, A. (1997): "Evaluating the Impact of French Employment Policies on Individual Labour Market Histories", *Review of Economic Studies*, 64 (4): 683-713.
- CANSINO, J.M. y ROMÁN, R. (2007): "Evaluación de políticas públicas sobre poblaciones heterogéneas ¿pueden los órganos de control externo contribuir a su avance?". *Revista Española de Control Externo*, Vol. IX (25): 107-129.
- CARD, D. (1990): "The Impact of the Mariel Boatlift on the Miami Labor Market", *Industrial and Labor Relations Review*, vol. 43 (2), 245-257.
- CARD, D. y KRUEGER, A. (2000): "Minimum Wages An Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply", *American Economic Review*, 90 (5): 1397-1420.
- CARD, D. y KRUEGER, A.B. (1994): "Minimum wages and employment: A case study of the fast food industry", *American Economic Review*, 84 (4): 772-793.
- CARD, D. y SULLIVAN, D. (1988): "Measuring the Effects of Subsidized Training Programs on Movements In and Out of Employment", *Econometrica*, 56 (3): 497-530.
- CONSEJO ECONÓMICO Y SOCIAL (1995): *España... economía, trabajo y sociedad: memoria sobre la situación socioeconómica y laboral*, Madrid, Ed.: Consejo Económico y Social.
- DEHEJIA, R.H. y WAHBA, S. (1999): "Causal Effects in Non-Experimental Studies: Re-evaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94 (448): 1053-1062.
- DONALD, S. y LANG, K. (2001): "Inference with Differences-in-Differences and Other Panel Data". *Working Paper*, Marzo 2001. Universidad de Boston, Department of Economics.
- HANSEN, C. (2003): "Generalized Least Squares Estimation in Differences-in-Differences and Other Panel Models", *Working Paper* MITDepartment of Economics, Marzo 2003.
- HECKMAN, J.J.; ICHIMURA, H.; SMITH, J. y TODD, P.E. (1998): "Characterizing Selection Bias Using Experimental Data", *Econometrica*, 66 (5): 1017-1098.
- HECKMAN, J.J.; ICHIMURA, H. y TODD, P.E. (1998): "Matching As an Econometric Evaluation Estimator", *Review of Economics Studies*, 65 (2): 261-294.
- HECKMAN, J.J.; ICHIMURA, H. y TODD, P.E. (1997): "Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *Review of Economics Studies*, 64 (4): 605-654.
- HECKMAN, J.J. y ROBB, R. (1985): "Alternative Methods for Evaluating the Impact of Interventions", en J. Heckman y B. Singer (eds.), *Longitudinal Analysis of Labour Market*: 156-245, New York, Ed.: Cambridge University.
- IMBENS, G.W.; LIEBMAN, J.B. y EISSA, N. (1997): "The Econometrics of Difference in Difference". *Working Paper*, Universidad de Harvard.
- MANSKI, C.F. (2001): "Designing programs for heterogeneous populations: The value of covariate information", *American Economic Review*, 91 (2): 103-106.
- MANSKI, C. y GARFINKEL, I. (1992): *Evaluating Welfare and Training Programs*, Cambridge, Massachusetts, Ed.: Harvard University Press.
- PARK, N.; POWER, B.; RIDDELL, W.C. y WONG, G. (1996): "An assessment of the Impact of Government-Sponsored Training", *Canadian Journal of Economics*, 29 (Special Issue: Part I): S93-S98.
- ROSENBAUM, P.R. y RUBIN, D.B. (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70 (1): 41-55.

STOCK, J.M. y WATSON, M.W (2003): *Introduction to Econometrics*, Boston, MA. Ed.: Addospm Wesley.

WOOLDRIDGE, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Massachussets. Ed.: MIT Press.