

UNA REVISIÓN SOBRE LOS MÉTODOS DE ESTUDIO Y EVALUACIÓN EN LAS POLÍTICAS ACTIVAS DE EMPLEO*

F. Alfonso Arellano Espinar

WP-EC 2006-06

Correo: F. Alfonso Arellano Espinar. Departamento de Fundamentos del Análisis Económico
Universidad de Alicante. Carretera San Vicente del Raspeig s/n. Apartado de Correos 99 03080 –
Alicante. Teléfono: 965-90-32-63. Fax: 965-90-38-98. Correo Electrónico: arellano@merlin.fae.ua.es

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.

Primera Edición Junio 2006

Depósito Legal: V-2379-2006

*Los documentos de trabajo del IVIE ofrecen un avance de los resultados de las
investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a
su remisión a las revistas científicas*

* Quiero agradecer a César Alonso Borrego (Universidad Carlos III de Madrid) y a Juan José Dolado (Universidad Carlos III de Madrid) por la supervisión de este trabajo. La ayuda financiera proviene de la beca de FPU del Ministerio de Educación y Cultura AP2000-0853. También agradezco la financiación de la Dirección General de Investigación a través del Proyecto del Plan Nacional de I+D+I número BEC2003-03943. Todos los posibles errores pertenecen sólo al autor.

UNA REVISIÓN SOBRE LOS MÉTODOS DE ESTUDIO Y EVALUACIÓN EN LAS POLÍTICAS ACTIVAS DE EMPLEO

F. Alfonso Arellano Espinar

ABSTRACT

The paper provides a review of the methodology of Survival Analysis and Causal Inference in the labour market context. This literature has been used to analyse and study the effects of (active) labour market policies adopted during the last years. Determining factors of the literature are also analysed: the types of data bases, the makeup of treatment and control groups, and the outcome variables. Kaplan-Meier (non-parametric) estimates and duration models constitute the main instruments in Survival Analysis. Causal Inference focuses on Matching methods, Difference-in-Difference method and Instrumental Variables procedure to determine the causal effect of a policy.

Keywords: Causal Inference, Survival Analysis, Active labour market policies.

RESUMEN

Se ofrece en este artículo una revisión de la metodología sobre Análisis de Supervivencia e Inferencia Causal en el ámbito del mercado de trabajo. Esta literatura se ha utilizado para analizar y estudiar los efectos de las medidas laborales que se han adoptado en los últimos años en diversas economías. Se analizan los factores determinantes de esta literatura, como son la naturaleza de las bases de datos, y la constitución de los grupos de tratamiento y de control y la variable objeto de estudio. Mientras las estimaciones no paramétricas de Kaplan y Meier y los modelos de duración constituyen los principales instrumentos en el Análisis de Supervivencia, en la Inferencia Causal se utilizan los métodos de emparejamiento, de diferencias en diferencias y variables instrumentales para determinar el efecto causal de una política.

Palabras Clave: Inferencia Causal, Análisis de Supervivencia, Políticas Activas de Empleo.

Código JEL: C13, C14, C41, J18.

1. INTRODUCCIÓN

La Economía como ciencia se basa en el estudio de la asignación de recursos limitados a infinidad de objetivos. El análisis y la evaluación del impacto de las medidas de política económica en una variable objeto de estudio constituyen uno de los fundamentos de esta disciplina. El soporte instrumental para estudiar los resultados de una medida, desde el punto de vista microeconómico, es múltiple, pero al menos dos áreas de la Estadística se han encargado de esta materia: el Análisis de Supervivencia y la Inferencia Causal.

El Análisis de Supervivencia tuvo una aplicación incipiente en el campo de la Bioquímica y la Medicina, donde se pretendía estudiar el resultado de medicamentos y terapias en la duración de la vida de los pacientes afectados. Constituye una rama de la Estadística que se encarga de estudiar los aspectos dinámicos del comportamiento individual en la duración de un evento.

Lancaster (1990) presenta una panorámica excelente en el estudio del Análisis de Supervivencia. El concepto fundamental es la función de riesgo, que se define como la probabilidad instantánea de finalizar el periodo de duración de un evento en el instante del tiempo t , dado que hasta t no ha finalizado. Con este objeto como base, se han utilizado estimaciones no paramétricas centradas en la influencia de una sola variable en la duración del periodo (Kaplan y Meier, 1958), y el uso de estructuras paramétricas y semi-paramétricas que permiten la interrelación de las variables entre sí y con la duración del periodo (Cox, 1972). Sin embargo, una crítica fundamental a estos modelos en Economía es que su formato no está justificado por razones de Teoría Económica, sino por motivos puramente matemáticos.

La Inferencia Causal se puede definir como aquella parte de la Estadística que se encarga de identificar, analizar y evaluar los efectos de un tratamiento en una variable de estudio sobre un colectivo de individuos.

La metodología que subyace detrás de la Inferencia Causal se basa en la dificultad que existe cuando se pretende cuantificar la influencia de la política sobre los individuos afectados. La medida del efecto se basa en la comparación entre dos situaciones, cuando la política (tratamiento) se produce y cuando no tiene lugar dicha medida. Sea Y la variable (continua) objeto de estudio, Y_{it}^1 es el valor de la variable para el individuo i en el momento del tiempo t si el individuo ha recibido el tratamiento, y Y_{it}^0 en el caso de ausencia de tratamiento. El impacto para el individuo i en el periodo t de esta medida es $\tau_{it} = Y_{it}^1 - Y_{it}^0$. Sin embargo, esta diferencia es desconocida porque no es posible observar estos dos términos para un mismo individuo al mismo tiempo:

$$Y_{it} = D_i \cdot Y_{it}^1 - (1 - D_i) \cdot Y_{it}^0$$

donde D_i es un indicador igual a uno si el individuo i recibe el tratamiento y cero en caso contrario. Esta dificultad se conoce por el nombre de Problema Fundamental de Evaluación y fue introducida por Roy (1951), Rubin (1974), Ashenfelter (1978) y Holland (1986) entre otros.

La solución a este problema depende de tres vías de actuación básicas en un proceso de evaluación. Dos de ellas son comunes a cualquier estudio aplicado, como son la base de datos y la variable objeto de estudio. La importancia de realizar comparaciones entre grupos comparables es fundamental, como Heckman, Lalonde y Smith (1999) resaltan. La tercera vía es más propia de la evaluación de tratamientos, la constitución de dos grupos de individuos que puedan ser comparados, el *grupo de tratamiento* formado por aquellos individuos que están afectados por la política, y el *grupo de control*, donde se incluyen los individuos que no reciben ningún tipo de tratamiento.

En este artículo, se realiza una revisión de esta literatura y de la metodología utilizada para estudiar, analizar y evaluar los efectos de un tratamiento, especialmente en el ámbito de las políticas (activas) de empleo. El resto del artículo está organizado de la siguiente manera: La realidad de las políticas de empleo y las aplicaciones en el contexto del mercado laboral se presentan en la Sección 2. La Sección 3 se encarga de presentar un análisis sobre la importancia de las bases de datos empleadas para estudiar los programas. La Sección 4 incorpora un compendio sobre los parámetros de interés utilizados en Inferencia Causal. Los métodos de estimación en Inferencia Causal se analizan en la Sección 5. En la Sección 6 se introducen los fundamentos del Análisis de Supervivencia. Las conclusiones se incluyen en la Sección 7.

2. CONTEXTO ECONÓMICO Y MERCADO LABORAL

Las aplicaciones sobre la evaluación de los efectos originados por una determinada medida se han aplicado a infinidad de situaciones, como así señalan Angrist y Krueger (1999) y Heckman, Lalonde y Smith (1999). En materia laboral, se han centrado durante los últimos años en medidas de fomento del empleo, especialmente en la Unión Europea (UE) tras la Cumbre del Empleo de Luxemburgo en noviembre de 1997. Este acontecimiento supuso el punto de partida en la búsqueda de una estrategia común para coordinar y concertar las políticas nacionales en materia laboral, constituyendo las políticas activas de empleo una de las claves de la Estrategia Europea del Empleo (Dolado, Felgueroso y Jimeno, 2001).

Martin (2000) describe la clasificación que introduce la OCDE sobre las actuaciones gubernamentales en materia laboral. Se dividen en medidas activas, encargadas de coadyuvar e incentivar la incorporación del trabajador al mercado de trabajo, y las medidas pasivas, que se centran en la provisión de aseguramiento frente a la contingencia del desempleo a través de prestaciones y subsidios.

A su vez, las políticas activas de empleo se dividen en cuatro grandes grupos: (i) los servicios y administración de empleo público se encargan de promover y flexibilizar la unión entre puestos de trabajo vacantes y trabajadores; (ii) los programas de formación profesional fomentan el capital humano de los trabajadores, en especial para aquéllos que se encuentran desempleados, mejorando sus posibilidades de incorporarse activamente en el mercado de trabajo; (iii) el empleo subsidiado se centra en el fomento de puestos de trabajo, incentivando las nuevas contrataciones de las empresas y a los trabajadores desempleados en su búsqueda de empleo, o bien con la creación directa de puestos de trabajo; y por último, (iv) las medidas dirigidas a grupos de trabajadores con dificultades para conseguir un puesto de trabajo, como los colectivos con discapacidades y los trabajadores jóvenes.

TABLA 1: GASTO EN POLÍTICAS DE EMPLEO EN PAÍSES DE LA UE, 2001

	Austria	Bélgica	Dinamarca	Finlandia	Francia	Alemania	Grecia	P. Bajos	Portugal	España	Suecia	R. Unido
1. Servicios públicos de empleo	0,14	0,17	0,12	0,12	0,18	0,23	0,06	0,26	0,11	0,09	0,23	0,13
2. Formación Profesional	0,20	0,24	0,85	0,29	0,25	0,34	0,21	0,31	0,15	0,14	0,30	0,05
3. Medidas destinadas a jóvenes	0,03	-	0,10	0,16	0,42	0,09	0,10	0,04	0,22	0,06	0,02	0,15
4. Subvenciones a la creación de empleo	0,11	0,77	0,17	0,29	0,37	0,25	0,08	0,38	0,09	0,40	0,24	0,01
-Subvenciones al empleo	0,06	0,27	0,02	0,15	0,18	0,03	0,05	0,05	0,01	0,25	0,19	0,01
5. Medidas a favor de incapacitados	0,06	0,12	0,33	0,09	0,09	0,29	0,01	0,58	0,04	0,03	0,31	0,02
Medidas activas (1 a 5)	0,53	1,30	1,56	0,95	1,31	1,20	0,46	1,58	0,61	0,73	1,09	0,36
Medidas pasivas (*)	1,07	2,18	3,00	2,02	1,65	1,92	0,47	1,86	0,90	1,33	1,19	0,56
Total políticas de empleo	1,60	3,48	4,56	2,96	2,96	3,13	0,93	3,44	1,52	2,06	2,28	0,92
Políticas de empleo por punto tasa de paro	0,44	0,53	1,06	0,33	0,34	0,40	0,12	1,43	0,37	0,16	0,45	0,18
Políticas activas por punto de tasa de paro	0,15	0,22	0,36	0,10	0,15	0,15	0,06	0,66	0,15	0,06	0,21	0,07

(*) Incluye prestaciones por desempleo y jubilaciones anticipadas por motivos relacionados con el mercado de trabajo.
Fuente: OCDE, Perspectivas del empleo, junio 2002.

La importancia de los gastos en materia laboral para varios países de la UE se resume en la Tabla 1. Los gastos en políticas de empleo presentan una gran diversidad al medirlo en porcentaje del PIB y con respecto a su participación en la tasa de paro. Aunque el peso de las políticas pasivas de empleo en el PIB es mayor, los países nórdicos gastan grandes cantidades de recursos en políticas activas de empleo. EN cuanto a la comparación del gasto en políticas activas de empleo, los países escandinavos y del centro de Europa, junto a Grecia, dedican más recursos a los programas de Formación Profesional. En cambio, Bélgica, Holanda y España prefieren las subvenciones directas a la creación de empleo. Finalmente, Francia, Portugal y Reino Unido se centran en las políticas orientadas a mejorar las posibilidades de empleo de los jóvenes trabajadores.

Como resultado de este énfasis en el presupuesto de los estados y el interés de los políticos por averiguar la efectividad de estas medidas, la importancia del análisis de las políticas de empleo ha venido aumentando durante la última década. Las variables que son objeto de estudio en este contexto se han centrado principalmente en el salario y la duración del desempleo. Cada una de ellas se ha utilizado en función del contexto laboral de EE.UU. y la UE. La primera variable se ha estudiado más detalladamente en EE.UU., donde el mercado laboral es más flexible tanto en cantidades (empleo) como en precio (salario), y su tasa de desempleo desde finales de los años sesenta ha sido más pequeña que en la mayoría de los países europeos. Angrist (1990), Card (1990), Card y Krueger (1994), Eissa y Liebman (1996), Eberwin, Ham y Lalonde (1997) constituyen buenos ejemplos de este enfoque¹. En la UE, con una mayor rigidez en el mercado de trabajo, la duración en el desempleo constituye una variable más importante que en EE.UU. Además, debido a la variedad de legislaciones fiscales, la duración en el desempleo es una variable más homogénea entre países que los salarios. Aunque hay trabajos destacables en Europa sobre evaluación de programas de empleo en los años ochenta, como Ridder (1986) en Holanda, ha sido recientemente cuando se ha realizado un mayor número de estudios aplicados, principalmente en el centro y norte de Europa, como Bonnal, Fougère y Sérandon (1997) y Abbring, Van den Berg y Van Ours (2002) en Francia, Lechner (1999, 2000) en Alemania, Lalive y Zweimüller (2004), y Gerfin y Lechner (2002) en Suiza, Heshmati y Engström (2001), Larsson (2003), y Sianesi (2004) en Suecia y en el caso británico, Blundell y Costa-Dias (2000) y Blundell, Costa-Dias, Meghir y Van Reenen (2003).

En España, se han realizado múltiples estudios en materia laboral analizando el mercado de trabajo de los trabajadores jóvenes (Dolado, Felgueroso y Jimeno, 2000) y las políticas pasivas de empleo, en especial las prestaciones por desempleo (García y Toharia, 2000, y Bover, Arellano y Bentolila, 2002, entre otros). Sin embargo, el número de estudios sobre la eficacia de las políticas activas a nivel microeconómico es aún incipiente, centrándose en los cursos de formación para desempleados (Mato, 1999, Mato y Cueto, 2004, y Arellano, 2005a) o en las reformas laborales introducidas en los últimos años enfocadas en las subvenciones a la creación de empleo (Jimeno, Kugler y Hernanz, 2002, y Arellano, 2005b).

¹ También hay estudios destacables que analizan la duración en el desempleo en EE.UU., como Card y Sullivan (1988) y Gritz (1993).

3. BASES DE DATOS

Como indican Heckman, Lalonde y Smith (1999), la calidad de las bases de datos es decisiva en el análisis de una medida, en especial en Inferencia Causal. Las bases de datos se pueden clasificar para este propósito en dos grandes grupos, aquéllas generadas por el investigador a partir de un experimento (aleatorio) controlado denominadas bases de datos experimentales, y aquellas bases de datos en las que el investigador no ha participado en su confección (y la información ha sido recogida por un organismo competente) denominadas administrativas o sociales.

Basándose en los estudios seminales de Fisher (1951) y Cochran y Rubin (1973), Bassi (1983, 1984) y Hausman y Wise (1985) señalan que las bases de datos experimentales, elaboradas siguiendo un experimento aleatorio, facilitan el cálculo del efecto del tratamiento, evitando posibles sesgos. Sin embargo, estas bases de datos presentan ciertas dificultades en la elaboración de experimentos en el contexto económico. Resulta crucial definir apropiadamente el proceso de evaluación con objeto de analizar el parámetro de interés que define el impacto de la medida, lo cual resulta bastante costoso. En este sentido, cobra una relevancia fundamental los conceptos de validez interna y validez externa del experimento. La validez interna indica la capacidad para estimar el efecto del tratamiento en una muestra particular de forma adecuada. La validez interna falla cuando hay diferencias (al margen del tratamiento) entre individuos tratados y controlados que afectan a la variable objeto de estudio y por las que no se puede controlar. Fallos en la aleatoriedad, el desgaste de la muestra y la falta de conformidad con el protocolo experimental son razones que explican la ausencia de validez interna. La validez externa indica la capacidad de poder extrapolar los resultados de las estimaciones a otras poblaciones. La falta de representatividad de la muestra y del tratamiento son causas de la ausencia de validez externa.

Una práctica generalizada consiste en el uso de datos administrativos derivados de actuaciones políticas. La discusión entre estos dos tipos de bases de datos y las implicaciones que se derivan de ellas se plantean en Rubin (1978), Lalonde (1986), Heckman, Ichimura y Todd (1997) y Heckman, Smith y Clements (1997). Heckman y Smith (1995) ofrecen argumentos que favorecen la utilización de bases de datos sociales frente a la corriente de pensamiento anterior que sobreestimaba la utilización de bases de datos experimentales. En la actualidad, el debate que se deriva sobre los experimentos aleatorios sigue existiendo, como Smith y Todd (2005) y Dehejia (2005) ponen de manifiesto.

Se busca un grupo de control apropiado que permita ser comparable con el grupo de tratamiento para analizar el efecto de la política en el grupo de individuos que se encuentran afectados. Una dificultad añadida proviene del proceso de generación de los grupos de tratamiento y control. Una vez determinada la variable de estudio y el parámetro de interés que se pretende analizar, un factor determinante en el proceso de evaluación es la formación de un grupo de tratamiento y otro de control.

También puede haber problemas en la definición del grupo de tratamiento. La existencia de aleatoriedad en la selección de los elementos a los grupos de tratamiento y control, así como la potencial heterogeneidad en el tratamiento (el individuo puede completar o

abandonar el tratamiento una vez incluido en el grupo de tratamiento) son factores que afectan a la evaluación de la política.

Así, $D^* = 1$ identificaría a los individuos que participan en el programa siguiendo un procedimiento aleatorio y cero si no participan. En caso de participar, obtendrían un valor Y_1^* y en caso contrario, Y_0^* . Los individuos que han sido asignados de forma aleatoria al programa toman el valor $R = 1$ condicionado a que $D^* = 1$ y $R = 0$ en caso contrario, mientras $T = 1$ sería para aquéllos que participan en el tratamiento y no lo abandonan condicionado a que $R = 1$.

En la elaboración de experimentos, la aleatoriedad se incluye en la creación de los grupos de tratamiento y de control. Los experimentos con asignación aleatoria resuelven el Problema Fundamental de Evaluación a través de la construcción directa del término inobservable excluyendo a individuos del grupo de tratamiento que de otra forma hubieran sido tratados.

Sin embargo, en las bases de datos administrativas, no suelen incluir este factor. La alternativa se sustenta en las características de la base de datos que se dispone. Se plantean dos posibilidades: la aceptación de un efecto común para todos los individuos (*common effect assumption*, $\tau_i = \tau$ para todo i) o la consideración de falta de sesgo no aleatorio en las estimaciones (*no randomization bias assumption*) lo que supone que $D^* = D$ y además que $Y_1^* = Y_1$ y $Y_0^* = Y_0$.

Cuando se producen abandonos en la participación, el efecto del impacto no puede estudiarse sobre el tratamiento sino sobre la intención de ser tratado (luego $R = 1$ implica $T = 1$). Por último, es necesario que no haya otros programas a disposición del grupo de control, esto es $R = 0$ implica $T = 0$.

Tanto el parámetro de interés como el método de estimación dependen en gran parte de la disponibilidad de los datos. Angrist y Krueger (1999), Heckman, Lalonde y Smith (1999), y Blundell y Costa-Dias (2002) realizan un compendio de todas las técnicas aplicadas en Inferencia Causal, así como una revisión de los factores fundamentales sobre los que se basa.

4. PARÁMETROS DE INTERÉS EN INFERENCIA CAUSAL

La imposibilidad de identificar el efecto individual del tratamiento ha inducido al estudio de los momentos poblacionales, más concretamente en la media. Según Heckman, Smith y Clements (1997), el uso de los momentos de primer orden para evaluar el efecto de los programas se sustenta en una relación directa entre la variable agregada y el bienestar y en la limitada importancia de los aspectos distributivos o su resolución a través de transferencias a partir de una función de bienestar social. Estos supuestos son restrictivos, pero la distribución de los impactos es difícil de obtener, pues no es posible normalmente derivar una distribución conjunta a partir de las distribuciones marginales de los individuos tratados y no tratados. Además, la identificación de otros momentos poblacionales requiere condiciones más complejas que las establecidas para las medias condicionales. Un caso especial es el supuesto de efecto común para todos los individuos, donde la distribución del impacto es degenerada y se concentra en la media del impacto. Abadie (2001) y Heckman, Smith y Clements (1997) tratan el estudio de distribuciones de la variable objeto de estudio. Presentan otros ejemplos de parámetros de interés, como pueden ser la proporción de individuos tratados que se beneficia del programa (que se define por la expresión $\Pr(Y_1 > Y_0 | D = 1)$) o diversos cuantiles de la función de distribución del efecto ($\inf_{Y_1 - Y_0} \{Y_1 - Y_0 : F(Y_1 - Y_0 | D = 1) > q\}$, donde q representa el valor del cuantil).

Al margen de estas disquisiciones, los diferentes condicionantes que presentan las bases de datos han originado la creación de múltiples definiciones de parámetros de interés sobre el efecto medio del tratamiento. Los parámetros más conocidos son: (i) el efecto medio del tratamiento, (ii) el efecto medio del tratamiento en los tratados, (iii) el efecto medio del tratamiento local y (iv) el efecto medio del tratamiento marginal.

El efecto medio del tratamiento (*Average Treatment Effect* o ATE) se define como la diferencia de dos valores esperados:

$$\tau^{ATE} = E(Y_1 - Y_0) = E(Y_1) - E(Y_0)$$

Este parámetro ofrece el efecto del tratamiento sobre una muestra aleatoria de individuos considerando un colectivo de observaciones elegidas de forma aleatoria para los grupos de tratamiento y control, con objeto de averiguar el impacto medio del programa si fuera obligatorio para toda la población.

El efecto medio del tratamiento en los tratados (*Average Treatment effect on the Treated* o ATT) determina el valor medio del tratamiento para las personas que recibieron el tratamiento en comparación con los no tratados en el caso hipotético de que ellos también hubieran recibido el tratamiento:

$$\tau^{ATT} = E(Y_1 - Y_0 | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 1)$$

Otro parámetro de interés es el efecto medio del tratamiento local (*Local Average Treatment Effect* o LATE):

$$\tau^{LATE} = \frac{E(Y|Z=z) - E(Y|Z=z')}{\Pr(D=1|Z=z) - \Pr(D=1|Z=z')}$$

Representa el impacto medio de las personas que cambian su condición de participación como resultado de un cambio en el instrumento Z . Cada instrumento define un valor del parámetro diferente.

El efecto medio del tratamiento marginal (*Marginal Average Treatment Effect* o MATE) analiza el impacto sobre los individuos de un cambio marginal de la participación en la política. Sea B una variable de política y que causa el cambio en la participación que toma dos valores b y b' . El MATE se define como

$$\tau^{MATE} = E(\tau | D=1 \text{ si } B=b' \text{ y } D=0 \text{ si } B=b)$$

Estos dos últimos efectos son similares, aunque la diferencia clave es que la variable de política (B) no tiene por qué ser un instrumento (Z).

Sin embargo, los momentos poblacionales no resuelven por sí mismos en ocasiones el Problema Fundamental de Evaluación. Es necesaria una serie de condiciones que permitan la identificación de estos parámetros. Así, un factor importante para identificar el efecto del tratamiento de forma apropiada consiste en tener en cuenta los efectos de Equilibrio General (*General Equilibrium effects*). El supuesto de que los individuos que no forman parte del grupo de tratamiento y de control no se encuentran influidos por el tratamiento (*Stable Unit Treatment Value Assumption* o *SUTVA*) suele ser utilizado como alternativa a este problema, aunque el carácter plausible de esta condición depende del programa estudiado y de los datos disponibles.

Otra condición de identificación es que el impacto es independiente del valor de la variable objeto de estudio en el caso del grupo de control, $Y_0 \perp \tau$. Con este supuesto se logra adoptar una forma funcional para τ .

En el caso del ATT, no se puede observar el segundo término de la ecuación. Si se sustituye por el término observable, se obtiene:

$$\begin{aligned} E(Y|D=1) - E(Y|D=0) &= E(Y_1|D=1) - E(Y_0|D=0) = \\ &= E(Y_1 - Y_0|D=1) - [E(Y_0|D=1) - E(Y_0|D=0)] \end{aligned} \quad (1)$$

donde el primer término de la última igualdad en la expresión (1) es el ATT y el segundo término entre corchetes es un sesgo en la estimación. La solución consistiría en aceptar la condición de que el sesgo es nulo, $E(Y_0|D=1) = E(Y_0|D=0)$. Esto es cierto si se satisface la condición de que la toma del tratamiento es independiente de los valores potenciales de Y :

$$Y_0, Y_1 \perp D$$

La aleatoriedad en la asignación a los grupos de tratamiento y de control permite la satisfacción de esta condición.

5. MÉTODOS DE ESTIMACIÓN EN INFERENCIA CAUSAL

Los métodos de estimación utilizados en Inferencia Causal se encuadran en tres grandes grupos: (i) los métodos de emparejamiento (*matching methods*), (ii) el método de diferencias en diferencias (*differences-in-differences method*) y (iii) el método de variables instrumentales (*instrumental variables method*) junto al modelo de Heckman. Aunque la necesidad de establecer una serie de supuestos de identificación del efecto (medio) del tratamiento es común a todos ellos, cada uno requiere unas condiciones diferentes.

La elección de los métodos depende de la estructura de las bases de datos, pues no hay un método que a priori prevalezca sobre los demás. La validez interna del estudio queda afectada por la combinación y aplicación de cada uno de los procedimientos de evaluación de esta sección, ya que hay problemas comunes a estos métodos derivados de la mecánica de los datos, bien por el proceso de selección del grupo de tratamiento, bien por la extensión del tratamiento, así como por la existencia de auto-selección o de un efecto sustitución. En general, las soluciones no suelen ser comunes y dependen de la aceptación de supuestos restrictivos o de las propias características de la base de datos.

El sesgo de selección es universal a cualquier base de datos y se produce por la falta de datos sobre aquellos factores que afectan a la participación en el tratamiento y a la variable objeto de estudio. Lalonde (1986), usando un grupo de tratamiento experimental y sus estimaciones como referencia, los comparó con aquéllos derivados de una base de datos social, y obtuvo que las estimaciones de la base de datos social eran sustancialmente diferentes de aquéllas obtenidas de los datos experimentales. Sin embargo, este estudio mostró diversas insuficiencias que limitan sus conclusiones, como la posibilidad de que hubiera sesgo de selección. La forma más apropiada de resolverlo consiste en la obtención de nuevos datos. Así, Heckman, Ichimura y Todd (1995) demuestran que la existencia de una gran cantidad de información en la base de datos, da lugar a resultados muy parecidos entre datos experimentales y no experimentales. Dado que cada método trata el problema de selección de forma diferente, la utilización de estos métodos para un mismo parámetro de interés puede generar resultados diferentes en muestras no experimentales.

Selección a partir de características observables: métodos de emparejamiento

Los métodos de emparejamiento han sido ampliamente utilizados en Inferencia Causal. Lalonde (1986) comparó los resultados de un proceso de evaluación aleatorio de un programa de empleo subsidiado con aquéllos derivados de una base de datos no aleatoria y técnicas econométricas tradicionales. Lalonde descubrió que los estimadores tradicionales eran incapaces de replicar los resultados de la base de datos experimental. Por el contrario, Dehejia y Wabha (1999) aplicando los métodos de emparejamiento con la probabilidad de participación mostraron que estos estimadores sí eran capaces de replicar los resultados experimentales.

Los métodos de emparejamiento se centran en buscar y comparar uno o varios elementos del grupo de control semejantes a cada elemento del grupo de tratamiento. De esta forma, las diferencias en la variable objeto de estudio se deben exclusivamente al tratamiento. Un requisito fundamental de este método es el supuesto de que la adopción del tratamiento es independiente de los valores potenciales de la variable objeto de estudio, dadas las características observables X :

$$(Y_0, Y_1 \perp D) | X$$

Esta condición, conocida como Supuesto de Independencia Condicional (*Conditional Independence Assumption* o *CIA*), no implica que Y sea independiente del tratamiento o que no exista efecto del tratamiento. Este supuesto se cumple si el vector X incluye todas las variables que afectan a la participación y a la variable objeto de estudio. Por tanto, este método requiere mucha información en la base de datos, como así señalan Heckman, Ichimura y Todd (1997) y Lechner (2002). Esta condición es similar a la utilizada en los métodos de regresión lineal, sin embargo el método de emparejamiento no exige una forma lineal del modelo. Hasta la fecha, esta condición sólo se ha podido contrastar con datos experimentales.

El método de emparejamiento más simple compara a individuos con idénticas características observables. Este método funciona cuando el vector X es discreto. Así, calcula la diferencia media en cada posible valor del vector X y posteriormente toma la media ponderada de las diferencias en función de la proporción de individuos tratados en cada valor:

$$\sum_{k=1}^C \frac{n_{1k}}{\sum_{k=1}^C n_{1k}} \left[\sum_{i \in k \cap \{D_i=1\}} \frac{Y_{1i}}{n_{1k}} - \sum_{j \in k \cap \{D_j=0\}} \frac{Y_{0j}}{n_{0k}} \right]$$

donde el vector X toma C valores posibles, hay n_{1k} observaciones del grupo de tratamiento y n_{0k} observaciones del grupo de control en el valor k del vector X ($k=1, \dots, C$).

Sin embargo, cuando el número de variables y los posibles valores de cada variable crecen, existe un problema de dimensión (*dimensionality problem*). La alternativa consiste en la utilización de métodos de emparejamiento no exactos que utilicen una distancia métrica que dependa de X . El emparejamiento se basa en la distancia más que en las características observables. Una distancia utilizada es la probabilidad de participación en el grupo de tratamiento condicionada al vector de características observables X (*propensity score*):

$$P(X) = \Pr(D=1|X)$$

Rosenbaum y Rubin (1983) muestran la importancia de esta distancia para estimar el efecto medio del tratamiento. Para ello utilizan el concepto de probabilidad de equilibrio, que se define como una función $b(X)$ tal que la distribución de X condicionada a dicha función es la misma para el grupo de tratamiento y de control:

$X \perp D \mid b(X)$. Si la base de datos satisface la CIA entonces también lo satisface con $P(X)$.

De esta forma, el proceso de estimación de esta probabilidad se puede realizar de forma paramétrica (a través de un modelo de elección discreta) o no paramétrica (calculando la probabilidad de participación para intervalos de valores de X). La versión paramétrica presenta tres virtudes: (i) la imposición de una forma funcional es menos problemática cuando la variable dependiente está acotada; (ii) las diferencias entre la forma paramétrica y la utilización de experimentos de Monte Carlo son pequeñas; y (iii) la utilización de este mecanismo para analizar contrastes de equilibrio y comprobar que la especificación de la probabilidad de participación es suficientemente flexible.

En los métodos de emparejamiento existe un problema de soporte común (*common support problem*) cuando no existe ningún individuo del grupo de control que presente características similares a las del individuo tratado.

Para el efecto medio del tratamiento en los tratados, la condición de soporte común es formalmente:

$$\Pr(D = 1|X) < 1 \text{ para todo } X$$

Por otra parte, para el efecto medio del tratamiento la condición formal es:

$$0 < \Pr(D = 1|X) < 1 \text{ para todo } X$$

La reducción de observaciones de los grupos de tratamiento y de control no constituye una solución apropiada a este problema. Heckman, Ichimura y Todd (1997) demuestran que aplicando los métodos no experimentales a grupos de individuos que constituyen el soporte común, el parámetro de interés estimado puede diferir de aquél estimado a partir de un experimento. Aunque la forma de aplicación es sencilla, hay graves inconvenientes, pues se pueden eliminar observaciones que potencialmente constituyen parejas adecuadas para otros casos y la eliminación de individuos tratados cambia la definición del parámetro que se pretende estimar.

Una de las nuevas líneas de actuación que trata de solventar el problema anterior consiste en flexibilizar el concepto de tratamiento, bien considerando la aplicación de varias medidas a la vez o el estudio de varios tratamientos para un mismo colectivo de forma consecutiva. Los métodos de emparejamiento permiten ajustarse a estas nuevas modalidades de tratamiento, como el tratamiento dinámico y el tratamiento múltiple. El emparejamiento dinámico (Lechner y Miquel, 2002) se origina cuando hay dos tratamientos que se producen en serie, y los participantes del primero eligen si quieren participar en el segundo. Suele ocurrir con los programas de formación y empleo. En este caso es necesario reinterpretar el parámetro de interés y el efecto del CIA.

El emparejamiento con tratamiento múltiple se utiliza cuando se ofrecen varias políticas en un mismo periodo de tiempo. Aparte de definir claramente cuál es el grupo de comparación, es indispensable utilizar una versión más fuerte del CIA para poder identificar el efecto del tratamiento (Imbens, 2000, Lechner, 2001, y Frölich, 2004). Considerando $M+1$ estados diferentes, $D = \{0, 1, \dots, M\}$, $M > 1$, donde $D=0$ identifica la

ausencia de tratamiento, para cada individuo i sólo se puede observar uno de los valores del conjunto $\{Y_{i0}, Y_{i1}, \dots, Y_{iM}\}$. En este caso, el efecto medio del tratamiento, y el efecto medio del tratamiento en los tratados (dados dos estados m y m') son respectivamente:

$$\tau_{m,m'}^{ATE} = E(Y_m - Y_{m'}) = E(Y_m) - E(Y_{m'})$$

$$\tau_{m,m'}^{ATT} = E(Y_m - Y_{m'} | D = m) = E(Y_m | D = m) - E(Y_{m'} | D = m)$$

El nuevo CIA se define en función de los valores de la variable objeto de estudio como $Y_0, Y_1, \dots, Y_M \perp D | X$ o condicionada al menos a $P(X)$. Esta condición requiere que se observen todas las variables que influyen en la selección a los diferentes tratamientos y a la variable objeto de estudio. Además se necesita que todos los individuos puedan estar incluidos en cualquiera de los $M+1$ estados. Por tanto, $P_j(X) = \Pr(D = j | X) \in (0, 1), \forall j = 0, \dots, M$.

Lechner (2001) define cada uno de estos efectos como

$$\begin{aligned} \tau_{m,m'}^{ATE} &= \sum_{j=0}^M \left\{ E_{P_{m|m,j}(X)} \left[E(Y_m | P_{m|m,j}(X), D = m) \right] | D = j \right\} - \\ &- E_{P_{m|m',j}(X)} \left\{ E \left[(Y_{m'} | P_{m|m',j}(X), D = m') | D = j \right] P(D = j) \right\} \\ \tau_{m,m'}^{ATT} &= E(Y_m | D = m) - E_{P_{m|m,m'}(X)} \left[E(Y_{m'} | P_{m|m,m'}(X), D = m') | D = m \right] \end{aligned}$$

donde $P_{m|m,m'}(X) = \frac{P_{m'}(X)}{P_{m'}(X) + P_m(X)}$.

Gerfin y Lechner (2002) establecen una guía de actuación para estimar estos parámetros. Lo ilustran con un estudio que realizan sobre las diferentes políticas activas de empleo que llevaron a cabo en Suiza al final de los años noventa utilizando un método de emparejamiento para múltiples políticas de empleo, con objeto de analizar el efecto sobre la probabilidad de empleo del trabajador desempleado. Las medidas se centraron en tres actuaciones: (i) programas de empleo al margen del mercado de trabajo, (ii) empleos temporales subsidiados que eran solicitados por el mercado de trabajo, y (iii) cursos de formación. Los resultados sugieren que la primera actuación tiene efectos negativos, la segunda muestra efectos positivos y los cursos de formación presentan resultados ambiguos.

La variedad de los métodos de emparejamiento radica en la multiplicidad de los criterios de selección entre observaciones tratadas y no tratadas. Asintóticamente, los métodos de emparejamiento comparan parejas exactas y proporcionan el mismo resultado. En muestras finitas, existen diferencias y la elección de cada uno de los métodos depende de los datos. A priori, no hay ningún método que prevalezca sobre los demás, ya que cada uno prima de forma distinta calidad frente a cantidad de parejas. Los métodos de emparejamiento más utilizados son el método del vecino más cercano,

el método radial, el método del núcleo y el método de estratificación (Heckman, Ichimura y Todd, 1997, y Becker e Ichino, 2002).

En el método del vecino más cercano (*nearest neighbor matching*), para evaluar el ATT cada individuo del grupo de tratamiento se empareja a un individuo (o individuos) del grupo de control más cercano. La elección del número de observaciones usadas en el grupo de control depende de la base de datos disponible. Un mayor número de individuos no tratados aumenta el sesgo de la estimación al introducir observaciones con una peor calidad en el emparejamiento, pero decrece su varianza porque se introduce más información para construir la contrapartida al individuo tratado.

Esta unión se puede realizar con reemplazo, utilizando una misma observación del grupo de control para unir a varias observaciones del grupo de tratamiento. El emparejamiento sin reemplazo está influido por el orden de emparejamiento, y puede constituir peores parejas de observaciones si el número de individuos del grupo de control es pequeño en comparación con el grupo de tratamiento. En conclusión, el emparejamiento con reemplazo mantiene un mayor control sobre el sesgo a costa de aumentar la varianza, mientras que en el emparejamiento sin reemplazo se prima la varianza frente al sesgo de estimación.

Sea $S(i)$ el conjunto de observaciones del grupo de control que se emparejan a la observación i del grupo de tratamiento. Para el método del vecino más cercano, $S(i)$ viene definido por la siguiente expresión:

$$S(i) = \left\{ j \mid \min_j \|p_i - p_j\| \right\}$$

Este método calcularía el valor medio de las diferencias de cada unión:

$$\hat{\tau} = \frac{1}{N_1} \sum_{i \in T} Y_{1i} - \frac{1}{N_1} \sum_{i \in T} \sum_{j \in S(i)} \omega_{ij} Y_{0j}$$

donde $\omega_{ij} = \frac{1}{N_{1i}}$ si $j \in S(i)$ y cero en caso contrario, y N_{1i} es el número de observaciones del grupo de control emparejadas con la observación i del grupo de tratamiento ($i \in T$).

El emparejamiento entre observaciones del grupo de tratamiento y de control está garantizado independientemente de la calidad de la unión con el método del vecino más cercano. Para solventar problemas derivados de la falta de calidad de los emparejamientos, se plantean otros métodos. El método radial introduce una distancia máxima permitida (r) entre el individuo tratado y no tratado. Si no se cumple la condición para ninguna observación del grupo de control no se considera dicha observación del grupo de tratamiento. En este caso, la definición del conjunto $S(i)$ viene dada por la expresión:

$$S(i) = \left\{ j \mid \|p_i - p_j\| < r \right\}$$

En cambio, el método de estratificación divide la muestra de n_1 individuos tratados y n_0 individuos no tratados en bloques ($q= 1, \dots, Q$), de tal forma que las variables se encuentran equilibradas entre ambos grupos. De esa forma, la asignación al tratamiento se puede considerar aleatoria. Se calcula la media de la variable objeto de estudio para los n_{1q} individuos tratados y los n_{0q} individuos del grupo de control en cada bloque q :

$$\hat{\tau}_q = \sum_{i \in q} \frac{Y_{1i}}{n_{1q}} - \sum_{j \in q} \frac{Y_{0j}}{n_{0q}}$$

El efecto medio del tratamiento se calcula utilizando una media de los bloques ponderada por su número de observaciones tratadas:

$$\hat{\tau} = \sum_{q=1}^Q \frac{n_{1q}}{n_1} \hat{\tau}_q$$

Estos dos últimos métodos presentan dos inconvenientes, el efecto desgaste y la arbitrariedad en la elección de los elementos de corte. El efecto desgaste tiene su origen en la eliminación de una gran cantidad de observaciones que afecten al parámetro de interés. La elección del límite de bloques Q y del límite r debería estar fundamentada a priori y el resultado puede ser sensible a la elección.

El método del núcleo constituye un método intermedio a los anteriormente expuestos. Permite controlar por calidad sin eliminar observaciones utilizando funciones de densidad. Este método utiliza como ponderación en el emparejamiento el cociente

$\frac{G_{ij}}{\sum_{k \in \{D_k=0\}} G_{ik}}$, donde la función $G(\cdot)$ (denominada función central o *kernel function*) es tal

que $G_{ik} = G\left(\frac{P(X_i) - P(X_k)}{a_n}\right)$, donde a_n es un parámetro de amplitud de banda. Las

funciones más utilizadas para $P(\cdot)$ son la función de distribución normal o la función de Epanechnikov.

Estimación de diferencias en diferencias

En ocasiones, hay razones para creer que los individuos tratados y no tratados se diferencian en características inobservables que están asociadas a la variable objeto de estudio, aunque se controle por las diferencias existentes en las características observables. En ese caso, las observaciones de los grupos de tratamiento y control no son directamente comparables.

La estimación de diferencias en diferencias requiere bases de datos con información anterior y posterior a la aplicación del tratamiento. Un ejemplo de este método es el artículo de Card (1990) que evalúa el efecto de la inmigración de miles de cubanos

procedentes del puerto de Mariel, entre mayo y septiembre de 1980, sobre el mercado de trabajo de Estados Unidos. Card analiza la situación de los salarios antes y después de este acontecimiento en Miami y en otras ciudades parecidas que no estuvieron afectadas. Los resultados sugieren que no hubo ninguna influencia de este hecho en el mercado laboral de Miami, debido a su capacidad de ajuste a movimientos migratorios que se habían producido con anterioridad.

Cuando este tipo de información está disponible, es posible comparar el grupo de tratamiento antes y después de la aplicación del proceso. Sin embargo, una simple comparación de la variable objeto de estudio antes y después suele estar contaminada por tendencias temporales o el efecto de eventos que se producen entre ambos periodos. La estimación de diferencias en diferencias (*Differences-in-Differences* o DID *estimation*) trata de corregir este problema.

Sea t el momento posterior al tratamiento y t' el momento anterior al tratamiento. El Problema Fundamental de Evaluación sigue existiendo, porque no es posible calcular la diferencia $Y_{1it} - Y_{0it}$ para un individuo i . Considerando como parámetro de interés el ATT, un supuesto de identificación fundamental para la estimación DID es que la evolución en media de la variable dependiente para individuos tratados y no tratados es paralela en ausencia del tratamiento:

$$E(Y_{0t} - Y_{0t'} | D = 1) = E(Y_{0t} - Y_{0t'} | D = 0)$$

Considerando esta condición, el ATT está identificado y es igual a:

$$E(Y_{1t} - Y_{0t} | D = 1) = [E(Y_t | D = 1) - E(Y_t | D = 0)] - [E(Y_{t'} | D = 1) - E(Y_{t'} | D = 0)]$$

Este mismo estimador se puede obtener usando técnicas de regresión. Sea T una variable binaria que toma valor 1 para el momento posterior al tratamiento y cero en caso contrario. El modelo lineal sin variables se define como:

$$Y = \beta_0 + \alpha D + \delta T + \varphi(D \cdot T) + U$$

donde $E(U | D, T) = 0$. Entonces,

$$E(Y_{1t} - Y_{0t} | D = 1) = [E(Y | T = 1, D = 1) - E(Y | T = 1, D = 0)] - [E(Y | T = 0, D = 1) - E(Y | T = 0, D = 0)] = \alpha$$

Con un modelo lineal y un vector X de variables que no varían con el tiempo, éstas no afectan la estimación de α en muestras grandes. En caso contrario, es posible que estén influidas por el tratamiento, generando endogeneidad. Este proceso de estimación de α se complica cuando Y es discreta y no se plantea un modelo de probabilidad lineal. Si Y es binaria, en el caso de que se considere un modelo de elección binaria:

$$\begin{aligned}
E(Y_{1t} - Y_{0t} | D=1) &= \\
&= [E(Y|T=1, D=1) - E(Y|T=1, D=0)] - [E(Y|T=0, D=1) - E(Y|T=0, D=0)] = \\
&= ME(Y|D, T=1) - ME(Y|D, T=0) = \\
&= [F(\beta_0 + \alpha + \delta + \varphi) - F(\beta_0 + \delta)] - [F(\beta_0 + \alpha) - F(\beta_0)]
\end{aligned}$$

donde $F(\cdot)$ indica la función de distribución seleccionada y $ME(\cdot)$ indica el efecto marginal de la variable.

Analizado de otra forma, el sesgo de selección invariante en el tiempo B se define como:

$$B = E(Y_{0t} | X, D=1) - E(Y_{0t} | X, D=0) = E(Y_{0t'} | X, D=1) - E(Y_{0t'} | X, D=0)$$

La idea principal que subyace en las estrategias de identificación de los efectos fijos es el uso de información sobre individuos en diferentes momentos del tiempo. De esta forma se controlan aquellas características inobservables que no varían con el tiempo y que están relacionadas con otras variables, en especial la variable objeto de estudio. Con datos de panel se puede eliminar la parte invariante del sesgo construyendo estimaciones en los periodos t y t' y tomando diferencias.

$$\tau_t^M = \tau^{TT} + B \text{ y } \tau_{t'}^M = B, \text{ luego } \tau_t^M - \tau_{t'}^M = \tau^{TT}$$

La aplicación de estos métodos se encuentra limitada por la disponibilidad de datos. Así, Heckman, Ichimura y Todd (1997) determinan que la combinación del método de emparejamiento con el método de DID normalmente tiene un mejor comportamiento entre los diferentes estimadores propuestos, en especial cuando hay características invariantes con el tiempo que están omitidas, lo cual es una fuente de sesgo importante.

Por supuesto, los métodos de DID presentan riesgos en su utilización. Así, Abadie (2005) señala la importancia de las tendencias temporales en la variable objeto de estudio o el efecto de otros acontecimientos ajenos al tratamiento que ocurren entre ambos momentos. Otro problema es el planteado por Ashenfelter (1978) a raíz de un estudio sobre salarios, donde las funciones que generan los salarios de los grupos de tratamiento y de control pueden diferir en el periodo anterior al tratamiento, lo cual distorsiona el efecto del tratamiento. Este fenómeno es conocido como caída de Ashenfelter (*Ashenfelter's dip*).

Selección sobre características inobservables: Métodos de variables instrumentales y Modelo de Heckman

En determinadas situaciones, los analistas consideran que las variables observadas no pueden explicar toda la dependencia entre la selección del tratamiento y los valores potenciales de la variable objeto de estudio, o no se dispone de datos de corte longitudinal. Si además hay un problema de endogeneidad, el método de Variables

Instrumentales (VI) constituye una alternativa que los anteriores métodos no logran solventar.

Un ejemplo didáctico de este método lo presenta Angrist (1990), que estudia el efecto de haber sido soldado veterano de la guerra de Vietnam en los ingresos que se obtienen como civiles para hombres de raza blanca. Angrist argumenta que el indicador de veteranía está correlacionado con componentes inobservables de la ecuación de ingresos o del propio individuo. Por ello utiliza una variable correlacionada con el indicador de veteranía que sea ortogonal al término de error, como es la variable binaria de llamada a filas, que se realizó de forma aleatoria a través de una lotería. Usando esta variable como instrumento, la conclusión principal es que los años de servicio militar sólo sustituyen parcialmente a la experiencia laboral, ya que los ingresos de los veteranos son menores en un 15% a los individuos que no son veteranos.

Un instrumento se define como aquella variable que afecta a la probabilidad de participar en el tratamiento y sólo influye en las variables objeto de estudio a través del tratamiento. Por tanto, el instrumento está correlacionado con la variable endógena (el indicador de tratamiento) pero no debe estar correlacionado con la parte inobservable de la ecuación de la variable objeto de estudio. Si se dispone de un instrumento que induzca la selección del tratamiento de forma exógena, el método de VI proporciona una estrategia de evaluación adecuada. Normalmente los instrumentos se derivan de no linealidades o de discontinuidades entre el regresor de interés y la variable objeto de estudio. Relacionado con el método de VI, Angrist y Krueger (1999) presentan los modelos de regresión con discontinuidad (*regression-discontinuity designs*). Estos modelos usan un modelo paramétrico o semi-paramétrico para controlar por tendencias graduales, infiriendo causalidad cuando la variable de interés cambia de forma abrupta por razones no justificadas.

El planteamiento econométrico que presenta el método de VI parte de un modelo lineal:

$$Y = \beta_0 + \alpha D + \beta_X X + U \quad (2)$$

Donde α mide el efecto común del tratamiento (esto es, $\alpha = Y_{1i} - Y_{0i}$ para todo individuo i).

Sea además una ecuación de participación:

$$IN = Z\gamma + \eta \quad (3)$$

Donde $D=1$ si $IN > 0$ y cero en caso contrario, y Z , que es una variable instrumental binaria, es independiente de η .

Se produce un sesgo de selección cuando existe una relación entre D y U condicional a X :

$$E(U|X, D) \neq 0 \quad (4)$$

Si existe una relación entre U y Z se dice que existe una selección sobre características observables.

Dadas estas referencias, las dos condiciones necesarias para definir un instrumento (Z) en este contexto son la ausencia de correlación con el término de error del modelo lineal (esto es, $E(U|X, Z) = E(U|X)$), y que dados dos valores de Z , 0 y 1, se cumpla que la probabilidad de ser tratado depende de una forma no trivial de X y Z (luego $\Pr(D=1|X, Z=1) \neq \Pr(D=1|X, Z=0)$). Una asignación aleatoria satisface ambas condiciones.

Con esta estructura se deriva el Estimador por VI:

$$E(Y|X, Z=1) = \beta_0 + \alpha \Pr(D=1|X, Z=1) + \beta_X X + E(U|X)$$

$$E(Y|X, Z=0) = \beta_0 + \alpha \Pr(D=1|X, Z=0) + \beta_X X + E(U|X)$$

Calculando la diferencia de ambas expresiones y despejando α se tiene el estimador del LATE:

$$\alpha = \frac{E(Y|X, Z=1) - E(Y|X, Z=0)}{\Pr(D=1|X, Z=1) - \Pr(D=1|X, Z=0)}$$

En caso de que exista heterogeneidad en el efecto del tratamiento, α_i , la estructura anterior sería insuficiente para identificar el parámetro de interés². El valor esperado de Y condicional a X y Z depende de α_i , y la diferencia de los dos términos daría lugar a una expresión que no permite generar un estimador para uno de los parámetros de interés.

Siguiendo los resultados de Imbens y Angrist (1994), se redefine la variable D en función de los dos valores de Z para cada individuo i , esto es $D_i(1)$ y $D_i(0)$. Dada la definición de las variables, se presentan cuatro posibles casos:

Caso 1: $D_i(1) > D_i(0)$, luego $D_i(1) = 1$ y $D_i(0) = 0$.

Caso 2: $D_i(1) = D_i(0) = 1$.

Caso 3: $D_i(1) = D_i(0) = 0$.

Caso 4: $D_i(1) < D_i(0)$, luego $D_i(1)=0$ y $D_i(0)=1$.

Se considera un modelo como el anterior (expresiones (2) a (4)) sin vector de variables X , con las mismas condiciones que en el caso del efecto común. Al añadir un supuesto de monotonicidad, no se produce nunca el Caso 4. Esta última condición no es contrastable, pues no es posible identificar para un individuo i en qué caso se encuentra, al observarse sólo uno de los términos de cada relación.

Con estos supuestos, Imbens y Angrist (1994) demuestran que:

² Otras referencias interesantes sobre el efecto de la heterogeneidad sobre las estimaciones del parámetro de interés en los modelos de VI son Heckman y Vytlacil (2000, 2001).

$$E(Y_1 - Y_0 | D(1) - D(0)) = \frac{E(Y|Z=1) - E(Y|Z=0)}{\Pr(D=1|Z=1) - \Pr(D=1|Z=0)} =$$

$$= \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)} = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)}$$

Abadie (2001) propone una generalización de este procedimiento cuando se dispone de un vector de variables X .

Como ocurre con los anteriores métodos, existen ventajas e inconvenientes del método de VI. Angrist y Krueger (2001) comentan que este método consigue superar los problemas con variables omitidas y medidas de error, pero hay dificultades como la obtención de una variable instrumental que esté correlacionada con las variables omitidas o con el término de perturbación o que esté débilmente correlacionada con el regresor endógeno.

En cuanto al Modelo de Heckman en su versión de efectos comunes, considérese las expresiones (2) a (4) con sus correspondientes definiciones. Se considera que η y U son independientes de X y Z , y que $E(U|X, D, Z) = E(U|X, D)$. Reescribiendo la expresión (2):

$$Y = \beta_0 + \alpha D + \beta_X X + E(U|X, D, Z) + U^*, \text{ donde } U^* = U - E(U|X, D, Z) \quad (5)$$

Considerando que $(U, \eta) \sim N(0, \Sigma)$, donde $\Sigma = \begin{pmatrix} \sigma_U^2 & \sigma_{U\eta} \\ \sigma_{\eta U} & \sigma_\eta^2 \end{pmatrix}$, se puede dar una forma funcional a la esperanza condicionada de U que depende de las funciones de distribución y densidad normales (Φ y ϕ , respectivamente):

$$E(U|X, D=1, Z) = \lambda = \frac{\phi(-Z\gamma)}{1 - \Phi(-Z\gamma)}$$

$$E(U|X, D=0, Z) = \tilde{\lambda} = \frac{\phi(-Z\gamma)}{\Phi(-Z\gamma)}$$

En este caso, la expresión (5) quedaría como:

$$Y = \beta_0 + \alpha D + \beta_X X + \frac{\sigma_{U\eta}}{\sigma_\eta} \lambda D + \frac{\sigma_{U\eta}}{\sigma_\eta} \tilde{\lambda} (1 - D) + U^*$$

Los valores de λ y $\tilde{\lambda}$ se pueden estimar a partir de un modelo probit, y las estimaciones de $\frac{\sigma_{U\eta}}{\sigma_\eta}$ se derivan de las estimaciones del modelo lineal, donde el denominador se normaliza a 1. El signo de este término indica el signo de la correlación entre las dos ecuaciones, que normalmente tiene gran interés.

El modelo de Heckman no requiere un instrumento y proporciona estructura adicional, al obtener una estimación de la correlación entre inobservables en las dos ecuaciones. Se diferencia de los métodos de emparejamiento en que busca la variación en la participación para corregir la selección sobre características inobservables.

6. ANÁLISIS DE SUPERVIVENCIA

Toda la literatura de Inferencia Causal se fundamenta en una estructura inicialmente no paramétrica. Sin embargo, en el estudio de la duración del desempleo existe una literatura complementaria denominada Análisis de Supervivencia, relacionada con métodos epidemiológicos y biológicos.

Esta metodología se ha asociado al estudio de los efectos de las medidas laborales y se ha desarrollado más en Europa, cuyo instrumento principal son los denominados modelos de duración. El éxito de esta literatura se centra, entre otras razones, en su fácil tratamiento matemático. Dado que Van den Berg (2000) realiza una excelente revisión de esta literatura, esta sección sólo pretende introducir los conceptos y modelos básicos que permitan iniciar su estudio.

Si bien la metodología de la Inferencia Causal considera el tratamiento como la pieza de estudio fundamental, evaluando y centrándose en su efecto causal, el Análisis de Supervivencia interpreta el tratamiento como un regresor más que influye en la variable objeto de estudio. Desde el punto de vista empírico, el Análisis de Supervivencia se centra en parámetros estructurales que determinan el comportamiento del individuo. Sin embargo, estudios como el de Abbring y Van den Berg (2003) buscan relanzar la conexión entre ambas áreas, al analizar la especificación e identificación de modelos de duración causales.

El Análisis de Supervivencia se inicia con una aproximación inicial no paramétrica como son las estimaciones de Kaplan y Meier (1958) de funciones de supervivencia³. Sea T una variable aleatoria continua que representa la duración de un estado, donde t representa una realización de T . Tiene una función de densidad $f(t)$ y una función de distribución $F(t)$, cuyo complemento es la función de supervivencia $S(t)=1-F(t)$. Una caracterización alternativa de la distribución de T es la denominada función de riesgo (*hazard function*), o tasa instantánea de abandono de la situación de desempleo por unidad de tiempo en el instante temporal t , definida como:

$$h(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t)}{dt}$$

El numerador indica la probabilidad condicional de que salga del estado inicial en el intervalo temporal $(t, t+dt)$ dado que hasta t no ha sucedido. El denominador define el tamaño del intervalo. Otra posible interpretación que ofrece Lancaster (1990) de esta expresión es que $h(t)dt$ es la probabilidad de salida del estado en el intervalo de tiempo dt después de t condicional a que hasta t se ha mantenido en el estado. La anterior expresión se puede escribir de la siguiente manera:

³ Esta sección se centra en los modelos de duración en tiempo continuo en un solo periodo con dos posibles estados. Para más información sobre los modelos en tiempo discreto, véase Meyer (1995); con múltiples periodos, Kalbfleisch y Prentice (1980) y Lindeboom y Kerkhofs (2000); y en el caso de múltiples estados, Han y Hausman (1990) entre otros.

$$h(t) = \frac{f(t)}{S(t)}$$

Dados estos resultados, es posible estimar $h(t)$ a través de una estimación de $F(t)$ o de $S(t)$. En este caso, se pretende estimar la función de supervivencia a través de la fórmula propuesta por Kaplan y Meier (1958):

$$S(t) = \prod_{j|t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Es una estimación por máxima verosimilitud no paramétrica, donde n_j representa el número de individuos que sobrevive en el instante del tiempo t_j y d_j el número de fallos en t_j . Por tanto, se estima la probabilidad de que los individuos no abandonen la situación de desempleo al menos antes de un instante del tiempo t .

Las estimaciones de funciones de supervivencia están sujetas al sesgo que genera la abundancia de valores de censura, en especial si dicha censura no es independiente, como señala Miller (1981). La introducción de regresores también puede complicar la utilización de estos procedimientos. Si el vector de variables X es invariante en el tiempo, sólo habría que condicionar las anteriores expresiones por este vector, empezando por la función de riesgo:

$$h(t|X) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t, X)}{dt}$$

En el caso de que el vector X esté formado por características que varían con el tiempo, sería necesario suponer que estas variables son constantes en el intervalo de observación para poder definir la función de riesgo condicional:

$$h(t|X(t)) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt | T \geq t, X(t + dt))}{dt}$$

A esta dificultad se añade la relación con la variable dependiente. Lancaster (1990) y Kalbfleisch y Prentice (1980) comentan las diferentes versiones de exogeneidad que permiten estimar apropiadamente los modelos de duración.

Con las estimaciones de Kaplan y Meier sólo se puede estudiar el efecto individual de cada variable en la probabilidad de salir del desempleo. La alternativa consiste en modelos que permitan analizar las interrelaciones entre variables, como los modelos AFT (*Accelerated Failure Time models*). Siguen una estructura paramétrica similar a los modelos de regresión lineal y, en consecuencia, muestran las mismas ventajas e inconvenientes de dichos modelos. Aunque estos modelos no se centran en los parámetros de la función de riesgo de cada individuo como los parámetros de interés, esta aproximación puede ser útil si se está interesado en el signo y la significatividad del efecto de las variables y en la duración del individuo en la muestra de riesgo (Van den Berg, 2001).

Sean T una variable aleatoria continua y positiva que indica la duración de un periodo, X un vector de variables explicativas observables invariantes en el tiempo de tamaño finito, y b un vector de coeficientes de regresión. Como Van den Berg (2001) y Van Ours (2001) suponen, no hay heterogeneidad inobservable entre los individuos que pueda afectar a las funciones de riesgo, ni anticipación en las decisiones de los individuos, y todos los individuos son independientes entre sí.

Los modelos lineales están conectados con los modelos de riesgo a través de los modelos AFT. Sea $Y = \ln T$, que está relacionado con X por medio del siguiente modelo lineal:

$$Y = X'b + z \quad (6)$$

donde z es un término inobservable con función de densidad $f(\cdot)$. El tipo de distribución del término inobservable determina el modelo de regresión. Kalbfleisch y Prentice (1980) y Lancaster (1990) presentan más información sobre las distribuciones consideradas. Si se utiliza la función exponencial en la expresión (6) se obtiene la siguiente expresión:

$$T = \exp(X'b)T'$$

donde $T' = \exp(z)$ tiene una función de riesgo $h_0(t')$ independiente de b . Por tanto, se define la función de riesgo de T a partir de una función de riesgo de referencia (*baseline hazard function*) $h_0(\cdot)$:

$$h(t|X) = h_0(t \exp(-X'b)) \exp(-X'b)$$

Este modelo especifica que el efecto de las variables es multiplicativo en t pero no en la función de riesgo, como ocurre con los modelos PH. El efecto de las variables consiste en alterar (acelerar o no) el tiempo de fallo de una persona.

Sean $f(\cdot)$ y $S(\cdot)$ la función de densidad y la función de supervivencia apropiadas al modelo de regresión paramétrico deseado respectivamente. Se pretende estimar el vector de parámetros v . Para ello, se maximiza el logaritmo de la función de verosimilitud:

$$\ln L = \sum_{j=1}^U \ln \left[f(t_j, v | t_{0j}) \right] + \sum_{j=U+1}^N \ln \left[S(t_j, v | t_{0j}) \right]$$

donde los individuos que abandonan el desempleo en t contribuyen a la función de verosimilitud con el valor de la función de densidad en t condicional al momento de entrada t_0 , mientras que la observación censurada (sólo se conoce que está desempleado hasta t) contribuye con la probabilidad de sobrevivir más allá de t condicional a la entrada en t_0 . De un total de N individuos mutuamente independientes, un número U de ellos se encuentran no censurados, considerándose la censura no informativa. Miller (1981) y Kalbfleisch y Prentice (1980) discuten las condiciones para los datos censurados con objeto de conseguir estimaciones válidas.

La estructura semi-paramétrica clásica de los modelos de duración procede del modelo de Cox (1972) o PH (*Proportional Hazard models*). Sea $h(t|X)$ la función de riesgo de un individuo con características definidas por un vector de dimensión finita X en el instante del tiempo t . El modelo de riesgos proporcionales propuesto por Cox (1972) se especifica de la siguiente forma:

$$h(t|X) = h_0(t) \exp(X'b)$$

donde $h_0(t)$ es una función de riesgo de referencia arbitraria para la variable continua T y la función exponencial constituye la parte sistemática del modelo⁴. Las variables incluidas en el vector X actúan multiplicativamente en la función de riesgo en este modelo, a diferencia de los modelos AFT. Este modelo proporciona estimaciones del vector de coeficientes b , pero no proporciona una estimación directa de $h_0(t)$. Por tanto, resulta complicado comparar las estimaciones de funciones de riesgo entre los modelos de Cox estimados y los modelos AFT.

El principal supuesto de este modelo es la proporcionalidad en la variable temporal entre la función de riesgo y la función de riesgo de referencia. Este supuesto es evaluable usando un contraste de riesgos proporcionales basado en la generalización realizada por Grambsch y Therneau (1994).

Con objeto de capturar la heterogeneidad inobservable surgen los modelos de Cox mixtos o MPH (*Mixed Proportional Hazard models*) que incorporan una variable aleatoria u a la estructura de la función de riesgo:

$$h(t|X, u) = h_0(t) \exp(X'b)u \quad (7)$$

Otro supuesto que subyace en los modelos MPH es el último término de la expresión (7), que es una función de variables explicativas inobservables. Se supone que es invariante en el tiempo, independiente del vector de características observables X y sigue una función de distribución $V(\cdot)$. El supuesto de la distribución depende principalmente de argumentos de teoría económica, como en los modelos de búsqueda, pero las formas más utilizadas han sido distribuciones discretas y distribuciones Gamma. En el caso de las distribuciones discretas, Heckman y Singer (1984) muestran que la estimación por Máxima Verosimilitud de V es una distribución discreta si se cumplen ciertos criterios de regularidad. Abbring y Van den Berg (2001) justifican la elección de una función de la familia de distribuciones Gamma.

⁴ Sobre los condicionantes de los modelos PH y Mixed Proportional Hazard (MPH) así como los supuestos de identificación de estos modelos, véase Van den Berg (2001), Heckman (1991) y Heckman y Taber (1994).

7. CONCLUSIONES

Este artículo pretende introducir una metodología que se basa en la misma filosofía que el modelo causal de Rubin (Holland, 1986) de transformar análisis de efectos (causales) en un modelo estadístico a partir de sus aplicaciones al campo de las políticas activas de empleo.

Sin embargo, todas estas técnicas dependen de forma crucial de las bases de datos disponibles, que permiten la utilización y justificación razonada de los supuestos necesarios para poder interpretar los resultados obtenidos de forma apropiada. En ciertas ocasiones, resulta difícil derivar conclusiones apropiadas dada la naturaleza limitada de las bases de datos.

Pero el cumplimiento de una serie de condiciones por parte de las bases de datos es sólo una condición necesaria para el proceso de análisis y evaluación. La determinación del efecto de un tratamiento (en este caso una política activa de empleo) se basa en la comparación entre un conjunto de observaciones afectado por la medida (grupo de tratamiento) y otro no tratado (grupo de control). La constitución de grupos comparables así como una identificación clara del tratamiento que se pretende evaluar son también piezas fundamentales en el atribulado proceso de evaluación. Este último aspecto se relaciona con la definición y existencia de una variable objeto de estudio apropiada que nos permita derivar conclusiones relevantes desde un punto de vista de política económica dentro del contexto de medidas activas de empleo.

La selección de los métodos de Inferencia Causal y Análisis de Supervivencia está marcada por estos condicionantes. Respecto a la Inferencia Causal, la aplicación de los métodos de emparejamiento se fundamenta en la plena disponibilidad de la información relevante sobre las observaciones. De esta forma, controlando por todos los factores posibles, las diferencias vienen motivadas por el tratamiento. En el caso de información antes y después del tratamiento, el método de diferencias en diferencias permite evaluar el efecto controlando por otros factores producidos en ese periodo de estudio que pueden afectar a la evaluación. Sin embargo, en ocasiones la condición de información completa es muy restrictiva. La existencia de componentes inobservables puede generar problemas de endogeneidad que los anteriores métodos no solucionan. La utilización del método de variables instrumentales pretende solventar estas dificultades añadidas en el proceso de evaluación.

El Análisis de Supervivencia constituye una vía alternativa de estudio, en especial cuando se pretende estudiar la duración en el desempleo. En este caso, el tratamiento constituye una variable independiente más en el modelo. Se pretende combinar la sencillez en los cálculos con la flexibilidad en la estructura del modelo, sin perder su conexión con la teoría económica.

Los recursos destinados a políticas activas de empleo en España son elevados en términos absolutos, pero el esfuerzo es algo menor que el de los países núcleo de la UE y muy inferior al de los países nórdicos. A partir de 1997, con el “proceso de Luxemburgo”, se ha producido una notable mejora que se ha traducido en un aumento de esta partida. El núcleo del gasto se concentra en subvenciones, cuyo coste de

oportunidad probablemente sea elevado. Todo ello se refleja en un reducido esfuerzo presupuestario por trabajador parado, la medida más apropiada de la intensidad de la política activa de empleo, que nos sitúa a la cola de la UE.

Sin embargo, de ello no debe inferirse una llamada a incrementar el gasto en políticas activas, sin evaluar antes concienzudamente los programas actualmente operativos. En un mundo globalizado, donde aumenta la demanda y disminuye la oferta de protección social, la provisión de estas políticas se encarece, por lo que es imprescindible proceder a una evaluación rigurosa de las mismas para seleccionar las mejores. Las evaluaciones llevadas a cabo hasta hace poco en España presentaban serias contradicciones respecto a las propias recomendaciones efectuadas.

Las políticas activas de empleo se financian generalmente con dinero público y el contribuyente está en su derecho de conocer la rentabilidad social de sus impuestos. Por tanto, las autoridades no deberían aprobar una sola de estas medidas, sin experimentar con ellas. Los servicios públicos de empleo están en condiciones de construir grupos de tratamiento y control adecuados, e incluso de impulsar transitoriamente acceso aleatorio a determinados programas. Esta forma de proceder debe convertirse en la norma y no en la excepción. En resumen, este artículo pretende constituir un inicio no sólo en la metodología utilizada para el estudio de las políticas activas de empleo, sino en la concienciación de los políticos y de otros agentes económicos de la utilización de procesos experimentales que permitan analizar y evaluar de forma apropiada los futuros programas de empleo.

BIBLIOGRAFÍA

- Abbring, J. H. y G. J. Van den Berg (2001), The unobserved heterogeneity distribution in duration analysis, Working paper, Free University, Amsterdam.
- Abbring, J. H. y G. J. Van den Berg (2003), The Nonparametric Identification of Treatment Effects in Duration Models, *Econometrica*, 71: 1491-1517.
- Abbring, J. H., G. J. Van den Berg y J. C. Van Ours (2002), The anatomy of unemployment dynamics, *European Economic Review*, 46: 1785-1824.
- Abadie, A. (2001), Semiparametric Instrumental variable estimation of treatment response models, John F. Kennedy School of Government, Harvard University.
- Abadie, A. (2005), Semiparametric difference-in-differences estimators, *Review of Economic Studies*, 72: 1-19.
- Angrist, J. D. (1990), Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records, *American Economic Review*, 80: 313-335.
- Angrist, J. D. y A. B. Krueger (1999), Empirical strategies in Labor Economics, in Ashenfelter A. and D. Card (eds.) *Handbook of Labor Economics*, vol. 3, Elsevier, Amsterdam.
- Angrist, J. D. y A. B. Krueger (2001), Instrumental variables and the search of identification: from supply and demand to natural experiments, *Journal of Economic Perspectives*, 15 (4): 69-85.
- Arellano, A. (2005a), Do Training Programmes get the unemployed back to work?: A look at the Spanish experience, Working Paper 05-25, Universidad Carlos III de Madrid.
- Arellano, A. (2005b), Evaluating the effects of labour market reforms "at the margin" on unemployment and employment stability: the Spanish case, Working Paper 05-12, Universidad Carlos III de Madrid.
- Ashenfelter, O. (1978), Estimating the effect of training programs on earnings, *Review of Economics and Statistics*, 60: 47-57.
- Bassi, L. (1983), The effect of CETA on the post-program earnings of participants, *Journal of Human Resources*, 18: 539-556.
- Bassi, L. (1984), Estimating the effects of training programs with non-random selection, *Review of Economics and Statistics*, 66: 36-43.
- Becker, S. O. y A. Ichino (2002), Estimation of average treatment effects based on propensity scores, *The Stata Journal*, 2 (4): 358-377.
- Blundell, R. y M. Costa-Dias (2000), Evaluation methods for non-experimental data, *Fiscal Studies*, 21: 427-468.
- Blundell, R. y M. Costa-Dias (2002), Alternative approaches to evaluation in empirical microeconomics, *Portuguese Economic Journal*, 1: 91-115.
- Blundell, R., M. Costa-Dias, C. Meghir y J. Van Reenen (2003), Evaluating the Employment Impact of a Mandatory Job Search Programme, *CEPR Discussion Papers*, 3786.
- Bonnal, L., D. Fougère y A. Sérandon (1997), Evaluating the impact of French employment policies on individual labor market histories, *Review of Economic Studies*, 64: 683-713.
- Bover, O., M. Arellano y S. Bentolila (2002), Unemployment duration, benefit duration and the business cycle, *The Economic Journal*, 112: 223-265.

- Card, D. E. (1990), The impact of the Mariel boatlift on the Miami labor market, *Industrial and Labor Relations Review*, 43: 245-257.
- Card, D. E. y A. B. Krueger (1994), Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania, *American Economic Review*, 84: 772-784.
- Card, D. y D. Sullivan (1988), Measuring the effect of subsidized training programs on movements in and out of employment, *Econometrica*, 56: 497-530.
- Cochran, W. y D. Rubin (1973), Controlling bias in observational studies, *Sankhya*, 35: 417-446.
- Cox, D. R. (1972), Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society, Series B*, 34: 187-220.
- Dehejia, R. H. (2005), Practical propensity score matching: a reply to Smith and Todd, *Journal of Econometrics*, 125: 355-364.
- Dehejia, R. H. y S. Wahba (1999), Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs, *Journal of the American Statistical Association*, 94: 1053-1062.
- Dolado, J. J., F. Felgueroso y J. F. Jimeno (2001), Mercado de trabajo y políticas de empleo en el inicio del siglo XXI, *Moneda y Crédito*, 212: 211-257.
- Eberwein, C., J. C. Ham y R. J. Lalonde (1997), The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: evidence from experimental data, *Review of Economic Studies*, 64: 655-682.
- Eissa, N. y J. B. Liebman (1996), Labour supply response in the earned income tax credit, *The Quarterly Journal of Economics*, 111: 605-637.
- Fisher, R. (1951), The design of experiments, 6th edition, Oliver and Boyd, London.
- Frölich, M. (2004), Programme evaluation with multiple treatments, *Journal of Economic Surveys*, 18: 181-224.
- García, I. y L. Toharia (2000), Prestaciones por desempleo y búsqueda de empleo, *Revista de Economía Aplicada*, 8: 5-33.
- Gerfin, M. y M. Lechner (2002), A microeconomic evaluation of the active labour market policy in Switzerland, *The Economic Journal*, 112: 854-893.
- Gritz, R. M. (1993), The impact of training on the frequency and duration of employment, *Journal of Econometrics*, 57: 21-51.
- Han, A. y J. A. Hausman (1990), Flexible parametric estimation of duration and competing risk models, *Journal of Applied Econometrics*, 5: 1-28.
- Hausman, J. A. y D. A. Wise (1985), Social experimentation, *NBER*. University of Chicago Press, Chicago.
- Heckman J. J. (1991), Identifying the hand of the past: distinguishing state dependence from heterogeneity, *American Economic Review*, 81 (suplement): 71-79.
- Heckman, J. J., H. Ichimura y P. Todd (1995), Non-parametric characterization of selection bias using experimental data: A study of adult males in JTPA, unpublished paper, University of Chicago.
- Heckman, J. J., H. Ichimura y P. Todd (1997), Matching as an econometric evaluation estimator: evidence from evaluating a job training program, *Review of Economic Studies*, 64: 605-654.

- Heckman, J. J., R. J. Lalonde y J. A. Smith (1999), The economics and econometrics of active labor market programs, in Ashenfelter A. and D. Card (eds.) *Handbook of Labor Economics*, vol. 3, Elsevier, Amsterdam.
- Heckman J. J. y B. Singer (1984), A method for minimizing the impact of distributional assumptions in econometric models for duration data, *Econometrica*, 52: 271-320.
- Heckman J. J. y J. A. Smith (1995), Assessing the case for social experiments, *Journal of Economic Perspectives*, 9: 85-110.
- Heckman, J. J., J. A. Smith y N. Clements (1997), Making the most out of program evaluations and social experiments: accounting for heterogeneity in program impacts, *Review of Economic Studies*, 64: 487-536.
- Heckman J. J. y C. R. Taber (1994), Econometric mixture models and more general models for unobservables in duration analysis, *Statistical Methods in Medical Research*, 3: 279-302.
- Heckman, J. J. y E. J. Vytlacil (2000), Instrumental variables, selection models, and tight bounds on the average treatment effect, in Lechner, M. and F. Pfeiffer (eds.) *Econometric Evaluation of Active Labour Market Policies*, Heidelberg, Physica-Verlag.
- Heckman, J. J. y E. J. Vytlacil (2001), Policy-relevant treatment effects, *American Economic Review*, 91: 107-11.
- Heshmati, A. y L. Engström (2001), Estimating the effects of vocational rehabilitation programs in Sweden, in Lechner, M. y F. Pfeiffer (eds.) *Econometric Evaluation of Active Labour Market Policies*, Heidelberg, Physica-Verlag.
- Holland, P. W. (1986), Statistics and Causal Inference, *Journal of the American Statistical Association*, 81: 945-970.
- Imbens, G. W. (2000), The role of the propensity score in estimating dose-response functions, *Biometrika*, 87: 706-710.
- Imbens, G. W. y J. D. Angrist (1994), Identification and estimation of local average treatment effects, *Econometrica*, 62: 467-475.
- Jimeno, J. F., A. Kugler y V. Hernanz (2002), Employment consequences of restrictive permanent contracts: evidence from Spanish labor market reforms, *IZA Discussion Paper*, 657.
- Kalbfleisch, J. D. y R. L. Prentice (1980), *The Statistical Analysis of Failure Time Data*, Nueva York, John Wiley & Sons.
- Kaplan, E. L. y P. Meier (1958), Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, 53: 457-481.
- Lalive, R. y J. Zweimüller (2004), Benefit entitlement and unemployment duration: the role of policy endogeneity, *Journal of Public Economics*, 88: 2587-2616.
- Lalonde, R. (1986), Evaluating the econometric evaluations of training programs with experimental data, *American Economic Review*, 76: 604-620.
- Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press.
- Larsson, L. (2003), Evaluation of Swedish youth labor market programs, *Journal of Human Resources*, 38.
- Lechner, M. (1999), Earnings and employment effects of continuous off-the-job training in East Germany after unification, *Journal of Business and Economic Statistics*, 17: 74-90.

- Lechner, M. (2000), An evaluation of public sector sponsored continuous vocational training programs in East Germany, *Journal of Human Resources*, 35: 347-375.
- Lechner, M. (2001), Identification and estimation of causal effects of multiple treatments under the conditional independence assumption, in Lechner, M. and F. Pfeiffer (eds.) *Econometric Evaluation of Active Labour Market Policies*, Heidelberg, Physica-Verlag.
- Lechner, M. (2002), Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods, *Journal of the Royal Statistics Society: Series A*, 165: 59-82.
- Lindeboom, M. y M. Kerkhofs (2000), Multistate models for clustered duration data: an application to workplace effects on individual sickness absenteeism, *Review of Economics and Statistics*, 82: 668-684.
- Martin, J. P. (2000), What works among active labour market policies: evidence from OECD countries' experiences, *OECD Economic Studies*, 30, 2000/1.
- Mato, F. J. (1999), La estimación de los efectos sobre el empleo de los programas de formación de desempleados: un ejercicio experimental, Universidad de Oviedo, mimeo.
- Mato, F. J. y B. Cueto (2004), "Evaluación de políticas de mercado de trabajo: un análisis aplicado de las subvenciones al autoempleo", *ICE - Revista de Economía*, 813.
- Meyer, B. D. (1995), Semiparametric estimation of hazard models, Working paper, Northwestern University, Evanston.
- Miller, R. G. (1981), *Survival Analysis*, Nueva York, John Wiley & Sons.
- Ridder, G. (1986), An event history approach to the evaluation of training, recruitment and employment programmes, *Journal of Applied Econometrics*, 1: 109-126.
- Rosenbaum, P. R. y D. B. Rubin (1983), The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70: 41-55.
- Roy, A. D. (1951), Some thoughts on the distribution of earnings, *Oxford Economic Papers*, 3: 135-46.
- Rubin, D. B. (1974), Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, 66: 688-701.
- Rubin, D. B. (1978), Bayesian inference for causal effects: the role of randomization, *The Annals of Statistics*, 6: 34-58.
- Sianesi, B. (2004), An evaluation of the Swedish system of active labour market programmes in the 1990s, *Review of Economic and Statistics*, 86: 133-155.
- Smith, J. A. y P. Todd (2005), Does matching overcome LaLonde's critique of nonexperimental estimators?, *Journal of Econometrics*, 125: 305-353.
- Van den Berg, G. J. (2001), Duration models: specification, identification, and multiple durations, in: Heckman, James J., and Leamer, Edward (eds.) *Handbook of Econometrics*, vol. 5, North-Holland.
- Van Ours, J. C. (2001), Do active labor market policies help unemployed workers to find and keep regular jobs?, in Lechner, M. and F. Pfeiffer (eds.) *Econometric Evaluation of Active Labour Market Policies*, Heidelberg, Physica-Verlag.