

DATOS MICROECONÓMICOS, HETEROGENEIDAD Y LA EVALUACIÓN DE POLÍTICAS PÚBLICAS*

James J. Heckman**

Universidad de Chicago y American Bar Foundation

En este artículo se recogen las contribuciones de la microeconomía al conocimiento económico. En él se abordan cuatro grandes temas: (1) el desarrollo por los microeconómetras de nuevas herramientas para responder a los problemas que empezaron a surgir tras la Segunda Guerra Mundial a raíz del análisis de las nuevas fuentes de información microeconómica; (2) las contribuciones de la microeconomía a los métodos de series temporales agregadas mediante la construcción de modelos que a su vez relacionan esquemas teóricos económicos para los individuos con datos sobre el comportamiento individual; (3) la importante regularidad empírica que esta disciplina detectó, referente a la diversidad y heterogeneidad del comportamiento, con importantes consecuencias sobre la teoría económica y la práctica econométrica; (4) la sustancial contribución de la microeconomía en materia de evaluación científica de medidas de política económica pública.

Palabras clave: microeconomía, heterogeneidad de la conducta, evaluación econométrica de políticas, Conferencia Nobel.

(*) Versión revisada del discurso pronunciado por James J. Heckman en Estocolmo, el 8 de diciembre de 2000, cuando recibió, junto con Daniel McFadden, el Premio en Ciencias Económicas del Banco de Suecia, creado en memoria de Alfred Nobel. El artículo es *copyright* © de la Fundación Nobel 2000 y se publica en *RAE Revista Asturiana de Economía* gracias a la autorización de la Fundación Nobel y con el permiso del autor. La traducción ha sido realizada por Sofía García y la revisión técnica corresponde a la profesora Arielle Beyaert.

(**) Agradezco los muy útiles comentarios de Jaap Abbring, Pedro Carneiro, Lars Hansen, Steve Levitt, Costas Meghir, Robert Moffitt, Jeffrey Smith y Edward Vytlačil. La investigación ha sido realizada gracias a la ayuda 97-09-873 de la National Science Foundation, a la ayuda 40-4043-000-85-261 del National Institute of Child Health and Human Development y a las ayudas de la American Bar Foundation. Agradezco a estas tres instituciones su apoyo durante todos estos años.

En nombre de todos los economistas que analizan datos microeconómicos y que emplean la microeconometría para unir la teoría y la evidencia y para evaluar políticas de intervención de todo tipo, acepto el Premio de Ciencias Económicas que otorga el Banco de Suecia en memoria de Alfred Nobel.

El campo de la microeconometría ha surgido en los últimos 40 años para ayudar a los economistas a describir con más precisión la economía, a diseñar y evaluar políticas económicas, y a contrastar teorías económicas y estimar parámetros de modelos económicos bien planteados. Se trata de un campo científico dentro del marco de la economía que relaciona teorías sobre comportamientos individuales con datos individuales, y pueden ser objeto de estudio las empresas, las personas o las unidades familiares. La investigación en el campo de la microeconometría está basada en el análisis de datos. La existencia de nuevos tipos de datos supone la aparición de nuevos retos y oportunidades que han estimulado los importantes avances realizados en este campo y han supuesto un cambio en la percepción de la realidad económica por parte de los economistas. La investigación en esta materia también guarda una estrecha relación con las políticas económicas. Aquellas cuestiones de política económica que pueden ser estudiadas por medio de los datos han motivado gran parte de la investigación en microeconometría. El deseo de contrastar y desarrollar nuevos modelos económicos también es determinante para la investigación en este campo.

En este discurso, trataré cuatro temas esenciales de la microeconometría, campo que ha sido reconocido por el Comité del Premio Nobel por primera vez este año. El primero está relacionado con la aparición de datos nuevos y muy ricos sobre individuos y empresas tras la Segunda Guerra Mundial, lo que permitió a los economistas comprender la economía en mayor profundidad al tiempo que enfrentaba a los economistas a una serie de problemas sin solucionar que no podían tratarse correctamente con los métodos de la econometría de ecuaciones simultáneas de la Comisión Cowles. Los avances en microeconometría se han visto estimulados por los problemas empíricos que surgen al analizar datos económicos.

El segundo tema guarda una estrecha relación con el primero. La microeconometría nació de la econometría de la Comisión Cowles como respuesta a las lagunas empíricas de su teoría. La econometría de la Comisión Cowles fue concebida para datos agregados y en un principio fue aplicada a gran escala a series temporales económicas. Muchos de los modelos econométricos de la Cowles se crearon con otra aspiración distinta a la de ofrecer soluciones a problemas específicos de toma de decisiones individuales; e incluso cuando se crearon con este objetivo, la literatura que trata el problema de la agregación en econometría estableció de manera formal que la relación entre el tomador de decisiones y los datos agregados empleados para estimar los modelos no era evidente. La microeconometría ha desarrollado modelos sobre el comportamiento individual formulados con precisión y ha estimado modelos sobre datos individuales. De esta manera, la relación entre la teoría y los datos ha llegado a apreciarse con mayor claridad.

En tercer lugar, esta conferencia aborda el tema de la emergencia de un número considerable de descubrimientos empíricos que surgen de la investigación microeconométrica. El más importante es la evidencia de

que, en la vida económica, prevalecen la heterogeneidad y la diversidad. Al realizarse un estudio completo de la heterogeneidad en las respuestas, surgió una variedad de medias posibles para describir la "persona media" y se descubrió que el paradigma tradicional del "consumidor representativo" carecía de fundamento empírico. Esto supuso un cambio en la percepción por los economistas de los modelos econométricos y la evaluación de políticas económicas. Apareció pues un nuevo modelo para los fenómenos microeconómicos. En el contexto del análisis de regresión, no sólo las ordenadas en el origen, sino también los coeficientes de las pendientes, se hicieron variables, y se constató que unas y otras podrían presentar correlaciones con los regresores. Tener en cuenta la heterogeneidad y la diversidad así como sus implicaciones para la economía y la econometría es el primer objetivo de esta conferencia y el tema principal del trabajo que he realizado a lo largo de mi vida.

El cuarto tema de mi discurso aborda la importante contribución de la microeconometría en materia de evaluación científica de políticas económicas basada en modelos econométricos, la cual ha constituido desde siempre un problema central de la econometría. Cuando se hicieron evidentes las dificultades para identificar parámetros estructurales, ya fuera para datos macroeconómicos o microeconómicos, los microeconomistas, basándose en las importantes sugerencias de Marshak (1953) y Hurwicz (1962), comenzaron a plantearse si era necesario recuperar todos los parámetros de modelos estructurales para responder, de forma fundamentada, a cuestiones específicas de política. A partir de entonces, se le otorga mayor importancia a los parámetros que son específicos del problema bajo estudio, los llamados "efectos del tratamiento" (en inglés: "*treatment effects*"), que generalmente son distintos de los parámetros estructurales. Son parámetros que ofrecen respuestas a cuestiones económicas más limitadas pero que son más fáciles de identificar y de acotar. El conocer las ventajas y limitaciones de estos efectos del tratamiento y el relacionarlos con los parámetros estructurales de la literatura anterior son avances muy recientes.

1. LA MICROECONOMETRÍA: ORÍGENES Y DEFINICIÓN

La econometría es una rama de la economía que une la teoría económica y los métodos estadísticos para así poder interpretar datos económicos y diseñar y evaluar políticas sociales. La teoría económica juega un papel imprescindible en la aplicación de los métodos econométricos porque, en muchos aspectos de interpretación, los datos en sí no son suficientes, no hablan por sí solos. La econometría hace uso de la teoría económica para guiar la construcción de modelos contrafactuales y aportar cierta disciplina a la investigación empírica en economía.

La producción de una amplia base de datos que puede ser empleada para describir la economía, contrastar teorías económicas y evaluar la política económica, es uno de los elementos más importantes en el desarrollo de la economía del siglo XX. Antes de este siglo, la economía era mayoritariamente una disciplina deductiva que se basaba en observaciones anecdóticas y en la introspección para contrastar la validez de la teoría y evaluar las políticas económicas.

La útil noción teórica de Alfred Marshall de la “empresa representativa” y del “consumidor representativo” estaba fuertemente arraigada en la teoría económica cuando los economistas comenzaron a recoger y a analizar de manera sistemática datos económicos agregados. Los primeros económetras se centraron en los datos agregados para medir los ciclos económicos y para construir modelos que podían ser la base de un acercamiento empírico a la evaluación de políticas macroeconómicas. Empleando sistemas de ecuaciones lineales, estos autores desarrollaron un marco para el análisis de modelos causales y la producción de políticas contrafactuales. Por primera vez la causalidad se diferenciaba de la correlación de una forma precisa y fundamentada que podía aplicarse empíricamente.

A pesar de que constituyeran considerables aportaciones intelectuales, los resultados empíricos que se obtuvieron con estos métodos resultaron ser decepcionantes. Casi desde el comienzo, los datos agregados de series temporales se consideraban poco útiles y los modelos macroeconómicos empíricos poco efectivos para contrastar teorías y aportar consejos en cuanto a las políticas económicas a seguir (véase Morgan, 1990). A excepción de algunos casos notables, los macroeconómetras adoptaron los métodos estadísticos de series temporales para las cuales la relación entre el modelo estadístico y la teoría económica era generalmente débil.¹

Ya en 1962, Orcutt (1962) abogaba por un programa que combinara datos microeconómicos y macroeconómicos para obtener así una descripción más fiable de los fenómenos económicos y contrastar teorías económicas alternativas. En la época en que Orcutt expuso sus puntos de vista, la base de datos microeconómicos era muy reducida, los ordenadores tenían una capacidad limitada y existían numerosos problemas derivados del uso de datos microeconómicos para estimar relaciones de comportamiento que no se habían logrado entender y mucho menos resolver. Sin embargo, la propuesta de Orcutt era audaz y contribuyó a la puesta en marcha de las fuerzas motrices de la microeconomía moderna.

La microeconomía amplió la teoría de la Cowles desarrollando modelos económicos más ricos en los cuales la heterogeneidad de los agentes desempeñan un papel fundamental y las ecuaciones que se estiman guardan una relación mucho más estrecha con los datos individuales y los modelos de elección individual. El centro de la teoría económica lo constituyen los individuos y sus interacciones con los mercados y otras instituciones sociales. Los datos necesarios para poner a prueba la teoría microeconómica son a su vez microeconómicos. La literatura econométrica sobre el problema de la agregación (Theil, 1954; para una panorámica véase, por ejemplo, Green, 1964 o Fisher, 1969) puso de manifiesto la fragilidad de los datos agregados para inferir el tamaño o el signo de las

(1) Véase también, sin embargo, el importante trabajo de Fair (1976, 1994), Hansen y Sargent (1980, 1991) y Hansen y Singleton (1982), que constituye la excepción a esta regla. Heckman (2000) examina este desarrollo. Estos métodos formulan problemas de toma de decisión bien planteados para individuos al deducir su ecuación de estimación pero la aplican a datos agregados.

relaciones microeconómicas. Al final, esta literatura tuvo resultados negativos y demostró la importancia de emplear datos microeconómicos como elemento de base de una ciencia económica con fundamentos empíricos. Todo ello constituyó la razón esencial que motivó la recogida y el análisis de datos microeconómicos.

Otra de las motivaciones fue el desarrollo del Estado del bienestar moderno y una consecuente demanda de información sobre la caracterización, las causas y las soluciones a problemas sociales, así como una demanda pública de evaluación objetiva de los programas sociales dirigidos a grupos determinados. La aplicación de los principios del paradigma de la Cowles y de su ampliación por Theil (1961) originó una demanda de estimaciones estructurales basadas en datos microeconómicos. Durante el periodo optimista de los años 60 y 70, la estimación de parámetros estructurales que no varían con la política, basados en datos microeconómicos, pasaron a ser el objetivo central del análisis econométrico destinado a evaluar los efectos de políticas antiguas en un entorno nuevo y a evaluar los posibles efectos de nuevas políticas que nunca se habían puesto en práctica.

Además del interés que presentaban para el análisis político, los modelos estructurales se emplearon para poner a prueba las teorías económicas. La economía laboral en particular se vio enriquecida por la aplicación de teorías neoclásicas en el mercado de trabajo. Esta demanda se vio alimentada por la emergencia de una teoría macroeconómica basada en la microeconomía. Las magnitudes numéricas de las preferencias a nivel individual y los parámetros de la producción desempeñaron un papel crucial en los debates sobre las teorías económicas y las políticas macroeconómicas.

Otra fuente de demanda de estimaciones estructurales surgió de la necesidad de sintetizar e interpretar la avalancha de datos microeconómicos que presenció la economía desde mediados de los años 50. La aparición de encuestas microeconómicas se produjo en el momento de la introducción de los ordenadores y del desarrollo y propagación de los métodos de regresión múltiple de Theil (1961, 1971) y Goldberger (1964) y todo ello hizo posible que se produjeran rápidamente no cientos, sino miles, de regresiones. La consecuente invasión de números era difícil de interpretar o de emplear para contrastar teorías o llegar a un consenso bien fundado sobre política económica. Surgió así una demanda de modelos de pocas dimensiones que se presten a la interpretación económica y permitan resumir las crecientes montañas de datos microeconómicos, y existía un reconocimiento cada vez más generalizado de que los métodos de regresión estándar no capturaban todas las características de los datos y de que tampoco ofrecían un marco para la interpretación de datos en modelos económicos bien planteados.

Antes de pasar a hablar de los avances específicos en microeconomía, es útil considerar dos cuestiones de evaluación de política económica que difieren mucho entre sí en cuanto a los datos y a los supuestos que se requieren para dar respuesta a las preguntas que plantean. En la microeconomía de los últimos treinta años, puede describirse una evolución que va de responder a cuestiones estructurales complicadas a responder a las cuestiones relativamente más simples del efecto del tratamiento.

2. POLÍTICA ECONÓMICA, MODELOS ECONÓMICOS Y EVALUACIÓN ECONOMÉTRICA DE POLÍTICAS

A menudo se confunden dos cuestiones relativas a la evaluación de políticas económicas que son conceptualmente diferentes. La cuidadosa distinción de estas dos cuestiones representa un importante avance en microeconomía y constituye uno de los temas esenciales de esta conferencia. La primera pregunta que se plantea es la siguiente: (1) "¿Qué efecto tiene la puesta en práctica de un determinado programa sobre los participantes en el mismo y sobre los que no participan en contraposición a la puesta en práctica de un programa alternativo o a no desarrollar ningún programa?" Esto es lo que se denomina ahora el problema del "efecto del tratamiento". La segunda pregunta que surge es: (2) "¿Qué efecto puede esperarse de un nuevo programa o de un programa antiguo aplicado en un entorno diferente?" Esta segunda pregunta plantea el mismo tipo de problemas que surgen cuando se trata de estimar la demanda para un nuevo producto. Para responder a ésta, es necesaria una estimación estructural.²

Es más sencillo responder a la primera pregunta que a la segunda a pesar de que la literatura temprana sobre estos temas trató de aportar una respuesta a ambas cuestiones estimando modelos estructurales. Uno de los importantes avances en materia de evaluación de políticas económicas al que yo he contribuido, ha sido clarificar las condiciones que deben satisfacerse para responder a los dos tipos de preguntas así como a otras preguntas relacionadas con las mismas.

El objetivo de la estimación econométrica estructural consiste en aportar los ingredientes necesarios para resolver distintos problemas de toma de decisiones. Estos problemas engloban tareas tan distintas como: (a) evaluar la efectividad de una política económica existente, (b) hacer conjeturas sobre la efectividad que pueda tener una política determinada en circunstancias diferentes a las que esta política se ha aplicado anteriormente, o (c) predecir los efectos de una nueva política que nunca se había aplicado antes.³ En este discurso, consideraré únicamente aquellos problemas de decisión que surgen en el análisis de políticas económicas.

(2) Esta cuestión se examina en artículos básicos de Lancaster (1966, 1971), Quandt (1970), McFadden (1974), Domencich y McFadden (1975) y Gorman (1980) (escrito por primera vez en 1956), entre otros.

(3) Marschak (1953) insistió sobre estas características de la estimación estructural. Al estimar la demanda para nuevos productos surgen cuestiones similares. Los métodos estructurales pueden emplearse para estimar parámetros de ecuaciones de demanda en un contexto económico dado, para predecir la demanda de bienes en contextos diferentes y para predecir la demanda de un nuevo producto que no se había consumido nunca antes. Es muy importante conocer los parámetros de las funciones de demanda para contrastar las teorías alternativas de la demanda del consumidor y medir la importancia de los bienes complementarios y sustitutivos.

Los modelos estructurales tienen la ventaja añadida de poder ser empleados para contrastar la teoría económica y evaluar cuantitativamente la importancia relativa de las relaciones causales en una determinada teoría. Además, los modelos estructurales que se basan en parámetros invariables se prestan a comparaciones entre estudios empíricos. Aunque el conocimiento empírico puede acumularse dentro de marcos estructurales, para ciertas clases importantes de problemas de toma de decisiones, no es necesario conocer todos los parámetros estructurales de un modelo, y en algunos casos ni si quiera es necesario conocer ninguno de los parámetros; lo cual es una suerte porque recuperar parámetros estructurales no siempre es tarea fácil.

En la literatura reciente sobre evaluación de políticas económicas, el objetivo implícito ha sido recuperar los ingredientes de los modelos necesarios para resolver problemas de toma de decisiones más específicos. Esto puede conllevar que en algunos casos se conozcan únicamente combinaciones de parámetros estructurales o parámetros que no son estructurales en ningún sentido habitual del término. Así, la literatura moderna sobre el efecto del tratamiento en economía tiene como objetivo principal la estimación de algún parámetro de efecto del tratamiento, y no la estimación de todos y cada uno de los parámetros que es lo que persigue la econometría estructural, a pesar de que, a menudo, las preguntas concretas a las que se responde en determinados estudios no se planteen de forma clara. Estos parámetros de tratamiento están identificados en condiciones más débiles que las requeridas para abarcar todos los parámetros estructurales del modelo. Las distinciones que hace la Comisión Cowles entre variables endógenas y exógenas y las distinciones posteriores entre exogeneidad débil, fuerte y super-exogeneidad que se han venido desarrollando en la literatura sobre estimación de parámetros estructurales y simulación de políticas económicas (Engle, Hendry y Richard, 1983) son en gran parte irrelevantes para la identificación de parámetros de tratamiento de uso frecuente. Al centrarse en problemas concretos de toma de decisiones, la literatura sobre efecto del tratamiento logra su objetivo con condiciones más débiles y por lo tanto más creíbles que las que se invocan en la literatura de econometría estructural. Al mismo tiempo, los parámetros que se generan de esta manera se prestan menos a ser trasladados a contextos diferentes para estimar los efectos de la misma política en circunstancias distintas o para estimar los efectos de una nueva política, y la comparación de los efectos en unos estudios y otros resulta difícil. La literatura sobre efecto del tratamiento debe ampliarse para posibilitar tales proyecciones y comparaciones y, como era de esperar, las extensiones necesarias son versiones no paramétricas de los supuestos empleados por los econométricos de los modelos estructurales.⁴

Para acotar un poco esta discusión y mantener al mismo tiempo cierto nivel de simplicidad, consideremos el problema prototípico de deter-

(4) Este aspecto se desarrolla de forma más completa en Heckman, LaLonde y Smith (1999) y Heckman y Vytlacil (2001a, 2001d, 2002).

minar el impacto de impuestos y subsidios sobre la oferta de trabajo. Este problema promovió una literatura temprana que evaluaba el Estado del bienestar (Cain y Watts, 1973), motivó mi propia investigación y continúa siendo hoy en día un problema importante de política económica.

Aplicando la teoría tradicional de la demanda del consumidor, formulemos una ecuación de oferta de trabajo con solución interior en la que las horas de trabajo H son función de los salarios, W , y de otras variables, denominadas \mathbf{X} , entre las cuales se incluyen activos, estructura demográfica, y similares. Sea U una variable no observable por el economista. Como veremos, las variables no observables desempeñan un papel fundamental en microeconometría y hay muchos indicios de que éstas tienen una importancia empírica. La microeconometría moderna se dedica a ellas.

Para la forma más general de H , tenemos:

$$H = \phi(W, \mathbf{X}, U) \quad (1)$$

Para simplificar, supongamos que ϕ es diferenciable en todos sus argumentos. La ecuación (1) es una función causal marshaliana.⁵ Las derivadas de la función con respecto a cada argumento miden el efecto ceteris paribus de éstos sobre H . Supongamos que queremos evaluar el efecto de un cambio en un impuesto proporcional sobre la renta. Con un tipo impositivo sobre la renta igual a t , obtenemos el salario después de impuesto $W(1-t)$. Supongamos que los agentes perciben correctamente el impuesto e ignoran todo efecto de equilibrio general del impuesto. En el lenguaje de los efectos del tratamiento, el efecto del tratamiento o “efecto causal” de un cambio en el impuesto sobre la oferta de empleo definida a nivel individual es $\phi(W(1-t), \mathbf{X}, U) - \phi(W(1-t'), \mathbf{X}, U)$ para la misma persona sujeta a dos impuestos diferentes t y t' .

Una versión aditivamente separable de la función causal marshaliana (1) es

$$H = \phi(W, \mathbf{X}) + U, E(U) = 0 \quad (2)$$

Esta versión permite al analista definir los efectos ceteris paribus de W y \mathbf{X} sobre H sin necesidad de conocer el nivel de la variable U no observable y desconocida por el economista. Una versión paramétrica de (1) es:

$$H = \phi(W, \mathbf{X}, U, \theta) \quad (1')$$

donde θ es un vector de parámetros de poca dimensión que genera el ϕ de la ecuación (1). La ecuación que sigue es una versión paramétrica de (2)

$$H = \phi(W, \mathbf{X}, \theta) + U \quad (2')$$

(5) Véase Heckman (2000) o Heckman y Vytlacil (2001a, 2002) para una definición rigurosa de funciones causales marshalianas.

Los parámetros θ reducen la dimensionalidad del análisis de identificación haciendo que el problema de identificar una función de dimensión infinita se transforme en el problema de identificar un conjunto finito de parámetros. Estos parámetros desempeñan un papel fundamental en la predicción de los efectos de una política económica antigua en poblaciones diferentes, en la acumulación de evidencia empírica de distintos estudios, y en la predicción de los efectos de una nueva política. La representación de H lineal en función de los parámetros es como sigue:

$$H = \alpha'X + \beta \ln W + U \quad (3)$$

donde se adopta una especificación semi-logarítmica para representar modelos frecuentemente empleados en la literatura relativa a la oferta de trabajo (véase Killinswoth, 1983).

Como hace Marschak (1953), resulta útil distinguir tres tipos de problemas de evaluación de políticas económicas. Nos situamos en una economía en la que se establece externamente un impuesto sobre su población o sobre un subconjunto de la misma (así, el impuesto se determina independientemente de U , pero podría depender de X y de W que son variables que observamos y con respecto a las que podemos condicionar). En el caso 1, el impuesto t se ha aplicado en el pasado y queremos predecir sus efectos sobre una población que tiene la misma distribución de variables (W, X, U) que la que estaba en vigor cuando se midieron los efectos de la variación del impuesto. En el caso 2, el impuesto t ya se ha aplicado en el pasado pero queremos pronosticar los efectos que tendrá sobre una población diferente para las variables (W, X, U). En el caso 3, el impuesto nunca ha sido aplicado y queremos predecir su efecto sobre una población inicial usada para estimar (1), o bien sobre una población diferente.

Supongamos que el objetivo del análisis es determinar el efecto de los impuestos sobre la oferta media de trabajo para una población con distribución $G(W, X, U)$. En el caso 1, tenemos datos de la misma población para la cual queremos elaborar una predicción. Supongamos que observamos distintos regímenes de impuestos. Los agentes se ven sujetos a tasas de impuestos externos t_j en el régimen j , $j = 1, \dots, J$. En la muestra de cada régimen podemos identificar

$$E(H|W, X, t_j) = \int \phi(W(1-t_j), X, U) dG(U|X, W) \quad (4)$$

Para el conjunto de la población esta función será:

$$E(H|t_j) = \int \phi(W(1-t_j), X, U) dG(U, X, W) \quad (5)$$

Se supone que esta función es aplicable a la población que nos interesa. Si conocemos (4) o (5), podemos proyectar esas funciones a partir de los datos históricos para todos los periodos futuros, siempre y cuando las distribuciones conjuntas de los datos sean constantes en el tiempo. Si uno de los regímenes ya ha sido experimentado en el pasado, podremos aplicar en el futuro las lecciones que de él hayamos sacado, siempre y cuando permanezcan los mismos $\phi(\cdot)$ y $G(\cdot)$. No es necesario construir nin-

gún estado contrafactual. Para realizar el análisis de la política para el caso 1 no es necesario conocer ninguna función causal de Marshall o ningún parámetro estructural. No es necesario descomponer (4) o (5) para aislar ϕ de G .⁶

El caso 2 se parece al caso 1 excepto por una diferencia esencial. Dado que proyectamos ahora la misma política en una población diferente, es necesario separar los componentes de (4) o (5) y determinar $\phi(W(1 - t), \mathbf{X}, U)$ separado de $G(U, \mathbf{X}, W)$. El problema de la evaluación de políticas económicas se vuelve mucho más complicado. La cita de Frank Knight (1921) viene al caso: "Surge un problema de conocimiento cuando el futuro difiere del pasado mientras que aparece una posibilidad de solución a dicho problema cuando el futuro es similar al pasado" (p. 313).

Los supuestos necesarios para predecir los efectos de las medidas de política económica antiguas en un nuevo régimen requieren que extraigamos del pasado la información necesaria para determinar los componentes de (4) o de (5) en nuevas poblaciones. Detallamos estos supuestos a continuación.

a) Es necesario conocer $\phi(\cdot)$ para la nueva población. Esto podría implicar determinar ϕ en un soporte diferente del que se ha usado en la muestra inicial si la población objetivo tiene un soporte diferente que el de la población de origen. Llegados a este punto, la estimación estructural alcanza su potencial. En algunos casos nos permite extrapolar ϕ de una población de origen a una población objetivo. Sin embargo es imposible alcanzar una solución completamente no paramétrica incluso bajo el supuesto (2') de separabilidad aditiva estructural, a no ser que los soportes de la población objetivo y de la población de origen coincidan.

Debemos aplicar cierta estructura a ϕ incluso si (2') caracteriza el modelo de oferta de trabajo. La estructura paramétrica (3) es habitual en la literatura de oferta de trabajo y los modelos lineales en los parámetros son los que dominan en la investigación de econometría aplicada.⁷

b) También es necesario conocer $G(\cdot)$ para la población objetivo. En este contexto, la exogeneidad constituye un supuesto simplificador esencial.

(6) No es necesario ni siquiera que t esté externamente especificado. Si una función descriptiva de una política $t = \eta(\mathbf{X}, W, U)$ genera t y η es una relación 1 a 1 en U y t dado (\mathbf{X}, W) , entonces cada t está asociada a una sola U dado (\mathbf{X}, W) . A condición de que el objetivo del análisis sea predecir los efectos de futuros t generados por η , podemos emplear datos históricos para hacerlo. Si η no es 1 a 1 en U y t dado (\mathbf{X}, W) , entonces no es posible, en general, emplear datos históricos para predecir el efecto de la variación de t generada por ϕ sobre H media. Sin embargo, si el objetivo es predecir políticas generadas por una nueva regla (incluidas las variaciones externas de t no relacionados con U), el caso 1 ya no es relevante y es necesario hacer estimación estructural (Lucas, 1976).

(7) El supuesto de que $\phi(W, \mathbf{X})$ es analítica real tal que se pueda extender a otros dominios es otro supuesto estructural. Heckman y Singer (1984) explotan este supuesto para resolver un problema de censura en análisis de duración.

Supuesto 1. (X, W) independiente de U .

Si definimos la exogeneidad según el supuesto 1, entonces $G(U|X, W) = G(U)$.⁸ En ese caso, si suponemos que la distribución de no observables es la misma en la muestra que en la predicción o en el régimen objetivo, $G(U) = G'(U)$, donde $G'(U)$ es la distribución de no observables en la población objetivo, podemos proyectar hacia una nueva población empleando la relación

$$E(H|W, X, t_j) = \int \phi(W(1-t_j), X, U) dG(U) \quad (6)$$

a condición de poder determinar $\phi(\cdot)$ sobre el nuevo soporte de X, W, U . No obstante, si $G' \neq G$, G' deberá determinarse de alguna manera. Esto nos lleva a emplear ciertos supuestos estructurales para determinar la relación entre G y G' .

En el caso 3, que corresponde a la situación en la que no se ha introducido ningún impuesto anteriormente, es necesario conocer la población objetivo. Los impuestos producen sus efectos a través del término $W(1-t)$. Si no hay variación de los salarios en las muestras extraídas del pasado, no hay manera de identificar el efecto de los impuestos sobre la oferta de trabajo ya que por hipótesis $t = 0$ y no es posible determinar el efecto del primer argumento sobre la oferta de trabajo. El problema no hace más que empeorar si suponemos que los impuestos tienen efectos sobre la oferta de trabajo independientemente de los salarios. De esta manera, incluso si se producen variaciones en los salarios, es imposible identificar los efectos de los impuestos o proyectarlos sobre una población nueva.⁹

(8) Hay muchas definiciones de este término. El supuesto 1 a menudo se ve completado por el supuesto adicional de que la distribución de X no depende de los parámetros del modelo (por ejemplo, θ en [1'] o [2']) (véase Engle *et al.*, 1983).

(9) Si los salarios varían durante el periodo previo a la aplicación de la política, podría no ser necesario descomponer (4) en ϕ y G , ni hacer estimación estructural para estimar el efecto de los impuestos sobre la oferta de trabajo en un régimen en el que se introducen los impuestos por primera vez. Si el soporte de $W(1-t) \stackrel{\text{def}}{=} W^*$ en el régimen objetivo se encuentra en el soporte de W en el régimen histórico, los soportes de X son los mismos en los dos regímenes y las distribuciones condicionales de U dado X, W y U dado X, W^* son las mismas, entonces, el conocer (4) en el soporte de W en el régimen histórico o de origen es suficiente para determinar el efecto de los impuestos en el régimen objetivo. Más precisamente, dejando que "histórico" denote los datos pasados y que "objetivo" denote la población objetivo para la proyección, podemos escribir estos supuestos como (a) soporte(X, W)_{objetivo} \subseteq soporte(X, W)_{histórico} y (b) $G(U|X, W)$ _{objetivo} = $G(U|X, W)$ _{histórico}, donde $W^* = W(1-t)$ para variables aleatorias W definidas en el nuevo régimen y (W) _{objetivo} = W _{histórico}. En ese caso, no es necesaria ninguna estimación estructural para predecir el efecto de los impuestos sobre la oferta de trabajo en la población objetivo. Es posible realizar una evaluación de política completamente no paramétrica estimando (4) o (5) no paramétricamente (y sin descomponer $E(H|X, W)$ en los componentes $\phi(\cdot)$ y $G(\cdot)$). Bajo el supuesto a, podemos encontrar el valor de contrapartida de $W(1-t) = W^*$ en la población objetivo para cada X para insertarlo en la versión no paramétrica de (4) (o [5]). Si estas condiciones no se dan es necesario construir las funciones G y ϕ en los nuevos soportes empleando las distribuciones apropiadas. Nos adentramos en un terreno en el que es necesaria la estimación estructural, ya sea para ampliar el soporte de las funciones $\phi(\cdot)$ o para determinar $G(U|X, W)$ o ambos. Sigue siendo necesario determinar la relación entre W y X en la población objetivo.

Esta discusión se aplica con la misma fuerza tanto al análisis de datos agregados como al análisis de datos microeconómicos. El empleo de variaciones individuales en encuestas microeconómicas proporciona una nueva forma de identificar ϕ y G que no se encuentren disponibles en datos macroeconómicos. De esta manera la identificación de parámetros estructurales se hace más fácil.

La literatura sobre el efecto del tratamiento amplía el primer caso de Marschak permitiendo que el tratamiento (t) sea exógeno. Consideremos dos poblaciones, éstas pueden ser subconjuntos de una población general y nos referiremos a ellas como el grupo de tratamiento y el grupo de comparación. En una de las poblaciones el impuesto es t_j y en la otra el impuesto es t_k que incluye como caso particular la ausencia de impuesto. Si las dos poblaciones son idénticas en lo que se refiere a ϕ y G y difieren únicamente en una tasa impositiva decretada externamente, entonces es posible determinar el efecto de t_j sobre el número medio de horas trabajadas en comparación con el impuesto t_k para cualquiera de las dos poblaciones y cualesquiera valores dados de \mathbf{X} , W ; para ello, basta con comparar la media de horas en ambas poblaciones, $E(H|W, \mathbf{X}, t_j) - E(H|W, \mathbf{X}, t_k)$ sobre dominios de soporte común para W , \mathbf{X} . No es necesario conocer ϕ ni G y, por lo tanto, no hace falta ninguna estimación estructural. Además, como ya se ha anotado, hay condiciones (rigurosas) bajo las cuales este ejercicio es válido incluso si t está determinado exógenamente por una medida de política económica, siempre y cuando, esta medida sea una relación 1 a 1 en (t y U) para \mathbf{X}, W dados.

En el contexto del ejemplo de la oferta de trabajo, la literatura sobre efectos del tratamiento trata de identificar las diferencia en horas medias de trabajo en un población determinada de (\mathbf{X}, W, U) que resultarían de la aplicación de diferentes políticas (t) impuestas externamente sin descomponer la media de horas en sus componentes de ϕ o G , empleando datos de poblaciones en las que t no viene especificado externamente. Los experimentos sobre medidas de política económica (ya sean naturales o simulados) que cambian t y no cambian ϕ ni G identifican estos efectos. También pueden emplearse instrumentos que hacen variar t sin que ϕ ni G se modifiquen. Así mismo se utilizan distintos métodos para tener en cuenta diferencias observadas y no observadas en las respuestas a las políticas que no estén relacionados con la medida de política económica que está siendo evaluada. Las condiciones de identificación necesarias para estimar efectos del tratamiento son generalmente más débiles que aquéllas que se precisan para identificar ϕ y G en el sentido de que son necesarios menos supuestos para identificar los efectos del tratamiento. Al mismo tiempo, las estimaciones producidas son específicas del problema examinado y son aplicables únicamente a las poblaciones estudiadas. Los efectos del tratamiento no permiten extender ϕ a nuevos contextos ni interpretar ϕ en términos de cambios ceteris paribus ("efectos causales") para todas las variables condicionantes excepto t .

Esta dualidad entre efectos del tratamiento y ecuaciones estructurales está presente en el conjunto de la literatura y en mi propio trabajo. Volveré a tratar este tema, pero antes explicaré por qué la disposición de datos microeconómicos constituye el elemento desencadenante para el desarrollo de la microeconometría.

3. NUEVAS CARACTERÍSTICAS DE LOS DATOS MICROECONÓMICOS

La producción de datos microeconómicos a gran escala en los años 50 hizo que se revelaran patrones y características que no podían racionalizarse fácilmente empleando modelos estándares de demanda de los consumidores y de oferta de trabajo o que fuesen bien analizados por técnicas de regresiones convencionales. Había importantes aspectos de la heterogeneidad y la diversidad que pudieron ser captados y hasta entonces habían permanecido ocultos en los datos macroeconómicos. Estos descubrimientos pusieron de manifiesto las carencias de las herramientas econométricas convencionales del momento.

El estudio de datos de corte transversal revela que individuos por otra parte observacionalmente idénticos toman decisiones distintas, ganan salarios diferentes y mantienen carteras de activos que difieren en nivel y composición. Estos datos revelan lo inadecuado que resulta ser el paradigma tradicional del agente representativo.¹⁰ El cuadro 1 ofrece una muestra típica de datos de oferta de empleo. Una fracción considerable de personas no trabaja y no observamos salarios para éstas. El coeficiente de determinación R^2 (una medida de bondad de ajuste) de cualquier relación microeconómica es típicamente baja, por lo que los no observables provocan gran parte de la inestabilidad en las horas de trabajo. Según cómo se evalúen los no observables, los efectos sobre la interpretación de la evidencia serán diferentes. Por ejemplo, ¿se debe el desempleo a la atracción no observada de los trabajadores por el tiempo de ocio o a un fallo del mercado en generar ofertas de trabajo que se observan sólo si se aceptan?,¹¹ ¿ocupan todas las mujeres empleos temporales o algunas (o la mayoría) tienen una vinculación a largo plazo en el mercado laboral?¹²

El uso de estos datos presenta problemas adicionales que son mucho menos aparentes en el análisis de datos de series temporales. Los que no trabajan no tienen salario. Entonces, ¿cómo pueden estimarse los efectos de los salarios sobre la oferta de trabajo si la información sobre salarios de la que disponemos sólo es aplicable a los que trabajan?, ¿Cómo pueden interrelacionarse las distintas dimensiones de la oferta de empleo (horas de trabajo, trabajo o no trabajo, número de periodos en los que se haya trabajado) para realizar un análisis contrafactual de las medidas de política económica?

(10) Lancaster (1966, 1971), Quandt (1970), McFadden (1974) y Domencich y McFadden (1975) fueron de los primeros en poner en tela de juicio la validez empírica del paradigma del agente representativo. Véase Kirman (1992) para una valoración reciente de ese paradigma.

(11) Flinn y Heckman (1982) analizan esta cuestión y muestran la dificultad de resolverla utilizando datos sobre elecciones de mercado.

(12) Heckman y Willis (1977) y Heckman (1981a) analizan esta cuestión.

Cuadro 1
DATOS DE PARTICIPACIÓN, HORAS TRABAJADAS Y SALARIOS,
DATOS DE NLSY, 1979-94

Grupo demográfico	Porcentaje de personas que trabajan a la edad de 29 años	R ² de las regresiones	
		Total de horas trabajadas en función de la educación y de la experiencia	Logaritmo del salario en función de la educación y de la experiencia
Hombres blancos	83,5	0,12	0,10
Hombres negros	75,0	0,15	0,14
Hombres hispanos	80,0	0,11	0,10
Mujeres blancas	76,4	0,15	0,17
Mujeres negras	69,6	0,18	0,21
Mujeres hispanas	66,6	0,18	0,10

Fuente: *National Longitudinal Survey of Youth, 1979-1994*, en Carneiro *et al.* (2001).

El análisis de datos nuevos ha acarreado la aparición de nuevos problemas econométricos como (a) tener en cuenta el carácter discreto de las variables de respuesta; (b) racionalizar la elección realizada tanto en los márgenes extensivos como intensivos (modelos para elecciones discretas y para elecciones a la vez discretas y continuas) en un modelo estructural común; (c) tomar en cuenta de manera sistemática el hecho de que haya datos ausentes, como precios o salarios, *como consecuencia de* las decisiones de los individuos.

Si nos centramos únicamente en los aspectos estadísticos de la microeconomía será menos evidente distinguir sus contribuciones básicas. Después de todo, muchos estadísticos se preocuparon de algunos de estos problemas. Goodman (1968), Haberman (1974) y Bishop, Fienberg y Holland (1975) analizaron modelos para datos discretos, aunque fueron los economistas los que introdujeron el estudio de modelos con respuestas discretas y continuas determinadas conjuntamente (Heckman, 1974a, 1974c) y de modelos con carencia sistemática de datos (Gronau, 1974; Heckman, 1974a, 1974c, 1976a, 1976c, 1979).¹³ Una contribución importante de la microeconomía fue la aclaración y la ampliación de los límites de estos marcos estadísticos para estimar modelos económicos, hacer distinciones causales y resolver varias versiones del problema de la evaluación de medidas de política económica descrito en la sección 2.

Por oposición a los modelos desarrollados por los estadísticos, el tipo de modelos microeconómicos desarrollados para explotar e interpretar las nuevas fuentes de datos microeconómicos puso de manifiesto el papel de la ciencia económica y de los marcos causales en la interpretación de la evidencia empírica, en el establecimiento de rela-

(13) Véase Holt (1985) para una discusión sobre la originalidad del trabajo de los econométras al analizar modelos para datos ausentes de forma no aleatoria.

ciones causales y en la construcción de modelos contrafactuales, ya fueran salarios contrafactuales que no existían en el análisis de la oferta de trabajo femenino o estados de política contrafactuales que surgen en la evaluación de políticas sociales. La investigación en microeconomía ha demostrado que hay que tener en cuenta las diferencias manifiestas entre individuos aparentemente similares. Las diferencias en los supuestos sobre el origen de la heterogeneidad no observada tienen repercusiones importantes sobre la estimación y la interpretación económica de la evidencia empírica, sobre la evaluación de los programas que se vienen aplicando, y sobre el empleo de datos para predecir nuevas políticas y evaluar el efecto que tiene el trasladar políticas existentes a nuevos entornos.

La heterogeneidad causada por variables no medidas se convirtió en un tema importante en esta literatura debido a la evidencia de sus manifestaciones en los datos y a las nefastas consecuencias de hacer caso omiso de su existencia. El problema resultó incluso más aparente cuando se tuvo acceso a datos microeconómicos de panel y fue entonces posible observar diferencias persistentes en el tiempo para las mismas personas.

4. RESPUESTAS POTENCIALES, CONTRAFACTUALES Y SESGO DE SELECCIÓN

En un principio centré mis esfuerzos en el campo de la microeconomía en la construcción de modelos para capturar las características esenciales de datos del tipo considerado en el cuadro 1, a través de modelos teóricos bien fundados que también podían emplearse para abordar problemas de evaluación de medidas de política económica (problemas de la cuestión 2 que se definen en la sección 2). Me inspiré en el trabajo de Mincer (1962) sobre la oferta de empleo femenino y me animó la oportunidad que se me brindó de construir un marco econométrico preciso para analizar las distintas dimensiones de la oferta de empleo femenino y su relación con los salarios. Para realizar esta tarea, me centré en dos conjuntos de herramientas econométricas a las que tenía acceso y el intento de fundir esos dos conjuntos en una única herramienta de investigación desembocó a la vez en frustración y descubrimiento.

Los dos conjuntos de herramientas que se encontraban a mi disposición eran (1) la teoría de las ecuaciones simultáneas de la Comisión Cowles y (2) los modelos de elección discreta que tienen su origen en la psicología matemática y que fueron introducidos en la economía por Quandt (1956, 1970), McFadden (1974, 1981) y Domencich y McFadden (1975). Mi objetivo era unificar estas dos corrientes para crear un modelo de ecuaciones simultáneas de pocas dimensiones y de motivación económica, con variables endógenas tanto discretas como continuas que tuvieran en cuenta la ausencia sistemática de los salarios de los que no trabajan y las diferentes dimensiones de la oferta de empleo en un marco común; marco que debería ser capaz de aportar una explicación a la oferta de trabajo femenino, y que constituyera la base de un riguro-

so análisis de medidas de política económica que nunca se había desarrollado antes.

El modelo estándar de la oferta de trabajo representado en la ecuación (1), (2) o (3) no es adecuado para tratar los datos del cuadro 1. Tampoco lo es la econometría de la Cowles. En circunstancias normales, los métodos de la Cowles pueden tomar en cuenta la correlación existente entre W y U en la ecuación (3), suponiendo que se han medido los salarios de todo el mundo. Esta correlación puede venir dada por el error de medida de los salarios o porque hay variables no observables comunes en las ecuaciones de los salarios y de la oferta de empleo (por ejemplo, las personas más motivadas trabajan más y obtienen unos salarios más elevados y la motivación es un elemento no observado). Los métodos de la Cowles no indican lo que se debe hacer cuando faltan salarios, cómo tomar en cuenta a los que no trabajan, o cómo relacionar la decisión de trabajar con la de cuántas horas de trabajo prestar.

En una serie de artículos escritos entre 1972 y 1975 (Heckman, 1973, 1974a, 1974c, 1976a, 1976c, 1978a)¹⁴ desarrollé modelos índice de respuestas potenciales para unificar la econometría de la Cowles y la teoría de la elección discreta, así como para unificar la literatura sobre selección muestral, truncamientos y variables dependientes limitadas que por entonces estaba dispersa y carecía de consenso entre autores.¹⁵ También desarrollé una variedad de estimadores en dos etapas para este tipo de modelos, con el objetivo de eludir las dificultades de cómputo que conlleva su estimación por máxima verosimilitud.

Basándonos en la literatura sobre psicología matemática y elección discreta tal y como la resume y amplía McFadden (1974, 1981), definamos

$$Y_i = g(\mathbf{X}_i, U_i), \quad i=1, \dots, I \quad (7)$$

como variables aleatorias latentes que reflejan las posibles respuestas. En el contexto de la elección discreta, las Y_i son utilidades latentes asociadas a la elección i y dependen tanto de características observadas (\mathbf{X}_i) como no observadas (U_i). Esto es lo que también se conoce como modelos de función índice. Para cada elección i , el nivel de utilidad puede variar. De forma general, al igual que en el programa de la Cowles y en especial en Haavelmo (1943), la ecuación (7) podría representar cualquier respuesta posible, como salarios, horas de trabajo, etc. Las ecuaciones (1) y (7) son

(14) Los primeros artículos se publicaron en 1974 (Heckman 1974a, 1974c) pero circularon ampliamente antes de su publicación. Otros dos artículos (Heckman 1976c, 1978a) se escribieron en 1973 y llegaron, en esa época, a las manos de numerosos econométricos veteranos.

(15) Véase Heckman y MaCurdy (1985) para un desarrollo sistemático de modelos de función índice.

relaciones causales marshalianas que nos dicen cómo varían las respuestas hipotéticas de Y_i cuando los argumentos del miembro derecho se modifican manteniendo todo fijo salvo la variable que se modifica.

Dependiendo del contexto, las Y_i podrán ser observadas directamente o solamente a través de sus manifestaciones. En modelos de elección discreta, las Y_i no se observan nunca pero observamos $\text{argmax}_i \{Y_i\}$. En la clase más general de modelos que he considerado, algunas de las Y_i pueden observarse bajo ciertas condiciones.

Para examinar estos modelos desde un planteamiento más elemental, consideremos una versión con tres funciones de respuestas potenciales. La literatura analiza los modelos con muchas respuestas potenciales. Formulemos las respuestas potenciales en formas aditivamente separables como

$$\begin{aligned} Y_0 &= g_0(\mathbf{X}) + U_0 \\ Y_1 &= g_1(\mathbf{X}) + U_1 \\ Y_2 &= g_2(\mathbf{X}) + U_2 \end{aligned} \tag{8}$$

Éstas son variables latentes que podrían ser observadas sólo de forma imperfecta. En el contexto de la teoría neoclásica de la oferta de trabajo, el salario de reserva o precio de reserva establecido en cero horas de trabajo (demanda cero para el producto) desempeña un papel central. Nos informa de cuál es el precio para el cual la gente está dispuesta a trabajar la primera hora o a comprar la primera unidad de producto. Denominemos esta función de salario de reserva potencial Y_0 . Sea Y_1 la función de salario de mercado, es decir, lo que ofrece el mercado. Sin tener en cuenta los costes fijos de trabajo, un individuo trabaja ($D = 1$) si

$$Y_1 \geq Y_0 \leftrightarrow D = 1 \tag{9}$$

De lo contrario, el individuo no trabaja. Las horas de trabajo potenciales Y_2 están generadas por las mismas preferencias que producen la función de salario de reserva, de tal manera que Y_2 e Y_0 están generadas por un conjunto común de parámetros. En mi artículo de 1974c, desarrollé una clase de formas funcionales simples con las que se puede operar matemáticamente en las que

$$Y_0 = \ln R = \text{logaritmo del salario de reserva} \tag{10a}$$

$$Y_1 = \ln W = \text{salario de mercado} \tag{10b}$$

$$Y_2 = \frac{\ln W - \ln R}{\gamma}, \gamma > 0 \tag{10c}$$

y las horas de trabajo observadas se representan como

$$H = Y_2 1(\ln W \geq \ln R) = \frac{\ln W - \ln R}{\gamma} 1(\ln W \geq \ln R)$$

dónde $1(A)$ es un indicador que toma el valor 1 cuando A es cierto. Los impuestos o transferencias proporcionales t introducen otra fuente de variación en estas ecuaciones de tal manera que en el lugar de W se utili-

ce el salario neto de impuesto $W(1 - t)$. Las variables no observables U_1 y U_0 explican por qué agentes que de otro modo serían observacionalmente idénticos (con las mismas \mathbf{X}) tienen elecciones distintas.¹⁶

El innovador modelo de Roy (1951) sobre la auto-selección en el mercado de trabajo que fue redescubierta en los años 70, está estrechamente relacionado con el modelo anterior. Resulta ser una versión del modelo de las funciones índice que acabamos de presentar.¹⁷ Como indica la ecuación (8), Y_0 e Y_1 son respuestas potenciales e Y_2 es una utilidad latente:

$$\begin{aligned} Y_2 \geq 0 &\leftrightarrow D = 1 \text{ e } Y_1 \text{ observada} \\ Y_2 < 0 &\leftrightarrow D = 0 \text{ e } Y_0 \text{ observada} \end{aligned} \quad (11)$$

Así, Y observada es

$$Y = DY_1 + (1-D)Y_0 \quad (12)$$

En el modelo original de Roy, $Y_2 = Y_1 - Y_0$. En el modelo generalizado de Roy, Y_2 está especificado de forma más libre, pero puede depender de Y_1 e Y_0 .

Estos modelos de respuestas potenciales contienen varias ideas distintas. (1) Como en los análisis de la Comisión Cowles, hay una superposición hipotética de valores posibles de las Y_j en respuesta a cambios ceteris paribus de las \mathbf{X} y de las U . Estos son modelos de funciones causales marshalianas representadas normalmente por modelos estructurales de pocas dimensiones para facilitar la predicción y el análisis de medidas de política económica. (2) Por oposición a los modelos de la Comisión Cowles, pero al igual que en los modelos de elección discreta, algunas de

(16) En Heckman (1974a), presento un modelo estructural más explícito de oferta de trabajo, cuidado de niños y salarios, que desarrolla, entre otras cosas, el primer marco econométrico que permite analizar el efecto de impuestos progresivos y de mercados informales sobre la oferta de trabajo. En ese artículo, caracterizo las preferencias mediante la función de tasa marginal de sustitución, genero Y_0 e Y_2 a partir de las curvas de indiferencia del consumidor, y produzco Y_2 a partir de una solución de las condiciones de primer orden del consumidor y de las restricciones presupuestarias. En ese modelo, las variables no observables que afectan a las preferencias se traducen como variaciones de un consumidor a otro (o de un trabajador a otro) en las pendientes de las curvas de indiferencia. La caracterización de las preferencias del consumidor mediante las pendientes de las curvas de indiferencia facilita el análisis de la oferta de trabajo bajo escalas de gravamen de impuestos sobre la renta por tramos y ofrece un tipo de preferencias más flexible que el que se obtiene con especificaciones para las ecuaciones de oferta de trabajo que sean lineales simples o semilogarítmicas. El análisis de impuestos progresivos para el caso convexo aparece en un apéndice de Heckman (1974b). Por razones de limitación de espacio, el editor, T. W. Schultz, pidió una presentación condensada. El análisis formal completo se publicó más tarde en Heckman y MaCurdy (1981, 1985) y Heckman, Killingsworth y MaCurdy (1981). Hausman (1980, 1985) extiende este análisis al caso no convexo.

(17) Roy desarrolla un modelo económico de desigualdad de renta y de clasificación pero no se detiene en ninguna de las cuestiones econométricas que surgen en su modelo. Lee (1978) y Willis y Rosen (1979) son dos aplicaciones del modelo de Roy.

las variables latentes no se observan (InR, por ejemplo, no se observa, pero puede revelarse en ocasiones mediante un cuestionario). (3) Por oposición tanto a los modelos de la Comisión Cowles como a los modelos de elección discreta, algunas de las variables latentes se observan, pero únicamente como consecuencia de las elecciones, es decir, que se observan selectivamente.

De esta manera, observamos $\ln W - \ln R$ salvo por un factor multiplicativo y observamos los salarios únicamente si $\ln W \geq \ln R$ ($D = 1$). El muestreo selectivo de respuestas potenciales da lugar al problema del *sesgo de selección*. Observamos únicamente submuestras seleccionadas de las variables de población latentes. En el contexto del modelo de Roy observamos Y_0 o Y_i pero no ambas.

Si no hubiese variables no observables en el modelo, este muestreo selectivo no tendría por qué preocupar. Si nos condicionásemos respecto de X , obtendríamos estimadores insesgados o consistentes de las respuestas que faltan para aquéllos que no trabajan usando las respuestas de los que sí trabajan. Sin embargo, los datos del cuadro 1, que son habituales, revelan que las variables observables explican únicamente una pequeña proporción de la varianza de prácticamente todas las variables microeconómicas. Es necesario tomar en cuenta la heterogeneidad de las preferencias y el muestreo selectivo de no observables. En general, como consecuencia de la regla de selección (9), los salarios y las horas que observamos son muestras selectivas de las respuestas potenciales de una población más amplia.¹⁸ El tener esto en cuenta es esencial si queremos estimar relaciones estructurales (los parámetros de las funciones causales) o describir el universo de respuestas potenciales (ecuaciones tales como [10a]-[10c]). Esto dio lugar al problema del sesgo de selección. Para resolver dicho problema ha sido necesario un nuevo análisis de elección discreta y de elección mixta continua y discreta que revisó la econometría convencional de la Comisión Cowles y puso de manifiesto la inadecuación de los modelos estadísticos convencionales para hacer distinciones causales en datos discretos. La teoría de elección mixta continua y discreta puso en entredicho el paradigma ya acuñado de la Comisión Cowles relacionando la econometría de forma más directa con los procesos de elección y de toma de decisiones. En el apéndice A, hablo de estas revisiones. De forma sucinta, los modelos log-lineales utilizados por los estadísticos para modelizar datos discretos no permitían hacer distinciones *ceteris paribus* entre causalidad real y espuria, necesarias para el análisis econométrico de medidas de política económica, y se desarrollaron nuevas condiciones para la coherencia de los modelos de ecuaciones simultáneas para asegurar modelos bien definidos tanto económicamente como desde el punto de vista de la probabilidad (Heckman, 1976c, 1978a, Heckman y MaCurdy, 1986). Amemiya (1985) ofrece un resumen muy útil de los principales avances en esta literatura.

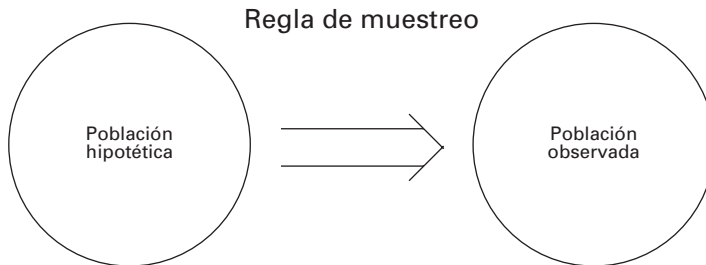
(18) Si (9) es cierto, entonces debe haber sesgo de selección al observar salarios o salarios de reserva salvo para casos degenerados (Heckman 1993). La selección en horas potenciales es consecuencia inmediata de (9) dado que $D = 1 \Leftrightarrow \ln W - \ln R \geq 0$.

5. SESGO DE SELECCIÓN Y AUSENCIA DE DATOS

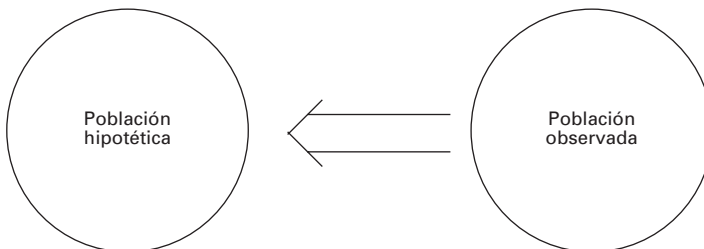
El sesgo de selección surge de la estimación de modelos estructurales con respuestas potenciales observadas parcialmente. No obstante, el problema del sesgo de selección es más general y puede surgir cuando una regla que no sea el muestreo aleatorio simple se emplea para muestrear la población subyacente objeto de estudio. La representación distorsionada de la población real en una muestra como consecuencia de una regla de muestreo constituye la esencia del problema de selección. El problema de la identificación se encuentra en la recuperación, desde una muestra observada, de las características de una población hipotética (véase gráfico 1). La población hipotética puede referirse a los salarios potenciales de todos los individuos, ya sean personas que trabajan (y para las que se observan salarios) o no, o las respuestas potenciales de cualquier problema de elección para el cual sólo se observan las elecciones que se han llegado a concretar de verdad. La distorsión de las reglas de selección puede surgir de las decisiones de los estadísticos que se ocupan de las encuestas de muestreo o de las decisiones económicas de auto-selección del tipo que hemos mencionado anteriormente, donde, como consecuencia de la auto-selección, observamos únicamente subconjuntos de una población de respuestas potenciales (por ejemplo Y_0 o Y_1 en el modelo de Roy).

Gráfico 1
RELACIÓN ENTRE POBLACIONES HIPOTÉTICAS
(CONTRAFACTUALES) Y DATOS OBSERVADOS

Proceso generador de los datos



Problema de identificación
Modelo econométrico



Una muestra aleatoria de una población nos aporta una descripción de la distribución de las características de dicha población que ofrece una enumeración completa de modelos de respuestas potenciales presentados en las secciones anteriores. Una muestra seleccionada con cualquier regla que no sea la del muestreo aleatorio nos proporciona una descripción de la distribución de las características de la población que no describe con precisión la distribución real de sus características, sea cual sea el tamaño de la muestra.

Hay dos caracterizaciones del problema de selección que presentan un interés particular. El primero, que tiene su origen en la estadística, conlleva caracterizar la regla de muestreo del gráfico 1, como la aplicación de un sistema de ponderación a distribuciones hipotéticas de la población para producir distribuciones observadas. El segundo, de origen económico, trata explícitamente el problema de selección como un problema de ausencia de datos, y en su esencia emplea variables observables para imputar valores a las no observables relevantes.

5.1. Distribuciones ponderadas

Cualquier modelo de sesgo de selección puede ser descrito en términos de distribuciones ponderadas. Sea \mathbf{Y} un vector de respuestas de interés y \mathbf{X} un vector de variables de "control" o "explicativas". La distribución de la población de (\mathbf{Y}, \mathbf{X}) es $F(\mathbf{y}, \mathbf{x})$. Para simplificar la exposición supongamos que la densidad está bien definida y la expresamos como $f(\mathbf{y}, \mathbf{x})$.

Cualquier regla de muestreo es equivalente a una función de ponderación no negativa $\omega(\mathbf{y}, \mathbf{x})$ que altera la densidad de la población. Los agentes son seleccionados para formar parte de la muestra de la población mediante una regla que es diferente, por lo general, de un muestreo aleatorio. Representemos por $(\mathbf{Y}^*, \mathbf{X}^*)$ las variables aleatorias producidas por el muestreo. La densidad de los datos muestreados $g(\mathbf{y}^*, \mathbf{x}^*)$ podría expresarse de la siguiente manera:

$$g(\mathbf{y}^*, \mathbf{x}^*) = \frac{\omega(\mathbf{y}^*, \mathbf{x}^*)f(\mathbf{y}^*, \mathbf{x}^*)}{\int \omega(\mathbf{y}^*, \mathbf{x}^*)f(\mathbf{y}^*, \mathbf{x}^*) d\mathbf{y}^* d\mathbf{x}^*} \quad (13)$$

donde el denominador de la expresión se introduce para asegurar que la integral de la función de densidad $g(\mathbf{y}^*, \mathbf{x}^*)$ valga la unidad. El muestreo aleatorio simple corresponde al caso en que $\omega(\mathbf{y}, \mathbf{x}) = 1$. Los esquemas de muestreo para los cuales $\omega(\mathbf{y}, \mathbf{x}) = 0$ para algunos valores de (\mathbf{Y}, \mathbf{X}) ocasionan problemas especiales porque no todos los valores de (\mathbf{Y}, \mathbf{X}) se muestrean.¹⁹

(19) Para las muestras en las que $\omega(\mathbf{y}, \mathbf{x}) = 0$ para una fracción no despreciable de la población, resulta útil considerar dos casos. Una *muestra truncada* es una muestra en la que no se conoce cuál es la probabilidad de observar la muestra a partir de la muestra aleatoria superior. Para tal muestra, (13) es la densidad de todos los valores \mathbf{Y} y \mathbf{X} muestreados. Una *muestra censurada* es una muestra en la que la probabilidad se conoce o puede ser estimada consistentemente.

Para muchos problemas económicos la atención se centra en $f(\mathbf{y}|\mathbf{x})$, la densidad condicional de \mathbf{Y} dado $\mathbf{X} = \mathbf{x}$. Si se seleccionan las muestras solamente según las variables \mathbf{x} ("selección según las variables exógenas"), $\omega(\mathbf{y}, \mathbf{x}) = \omega(\mathbf{x})$, y el empleo de muestras seleccionadas no constituye ningún problema para hacer inferencias válidas sobre la densidad condicional de la población.

Hacer muestreo tanto para \mathbf{y} como para \mathbf{x} se denomina *muestreo estratificado general*, y se pueden caracterizar varios esquemas de muestreo distintos según la estructura que se le da a las ponderaciones (Heckman, 1987).

No es posible recuperar la verdadera densidad de $f(\mathbf{y}, \mathbf{x})$ de una muestra de datos si no conocemos la regla de ponderación. Por otra parte, si conocemos la ponderación $\omega(\mathbf{y}^*, \mathbf{x}^*)$ y el soporte de (\mathbf{y}, \mathbf{x}) y si $\omega(\mathbf{y}, \mathbf{x})$ es no nulo, entonces siempre podemos recuperar $f(\mathbf{y}, \mathbf{x})$ porque

$$\frac{g(\mathbf{y}^*, \mathbf{x}^*)}{\omega(\mathbf{y}^*, \mathbf{x}^*)} = \frac{f(\mathbf{y}^*, \mathbf{x}^*)}{\int \omega(\mathbf{y}^*, \mathbf{x}^*) f(\mathbf{y}^*, \mathbf{x}^*) d\mathbf{y}^* d\mathbf{x}^*} \quad (14)$$

por hipótesis, conocemos tanto el numerador como el denominador del miembro izquierdo, y sabemos que $\int f(\mathbf{y}^*, \mathbf{x}^*) d\mathbf{y}^* d\mathbf{x}^* = 1$, por lo tanto es posible determinar $\int \omega(\mathbf{y}^*, \mathbf{x}^*) f(\mathbf{y}^*, \mathbf{x}^*) d\mathbf{y}^* d\mathbf{x}^*$. Es fundamentalmente más fácil corregir el sesgo que resulte de planes de muestreo con ponderaciones no negativas conocidas o ponderaciones que pueden ser estimadas por separado del modelo completo que corregir una selección para la cual las ponderaciones no se conocen y deben ser estimadas conjuntamente con el modelo.²⁰ El muestreo con sesgo de elección, el muestreo con sesgo de longitud y el muestreo con sesgo de tamaño son ejemplos de lo primero; el muestreo que surge de la selección en el modelo de las ecuaciones (10a)-(10c) o en el modelo de Roy generalizado es un ejemplo de lo segundo.²¹

La exigencia de que se conozca (a) el soporte de (\mathbf{y}, \mathbf{x}) y de que (b) $\omega(\mathbf{y}, \mathbf{x})$ sea no nulo no es inocua. En muchos problemas económicos de impor-

(20) Desde mediados de los años 50 se han estudiado selecciones con ponderaciones conocidas bajo el nombre de estimaciones de Horvitz-Thompson. Rao (1965,1985) resume esta investigación en estadística. Manski y Lerman (1977), Cosslett (1981) y Manski y McFadden (1981) tuvieron contribuciones importantes a la literatura económica sobre muestreo basado en la elección. El muestreo basado en criterios de longitud es analíticamente equivalente al muestreo basado en criterios de elección y ha sido estudiado desde finales del siglo XIX por actuarios daneses (véase Sheps y Menken, 1973; Trivedi y Baker, 1983). Heckman y Singer (1985) amplían el análisis clásico de muestreo basado en criterios de longitud realizado en análisis de duración considerando modelos con no observables que varían entre periodos de actividad y con variables que se modifican con el tiempo. En su caso más general, no pueden emplear métodos simples de ponderación con pesos determinados independientemente del modelo.

(21) Lewbel (2001) presenta un análisis interesante de un modelo de selección basado en un modelo de tipo Roy en el que las ponderaciones pueden construirse independientemente del conjunto del modelo para recuperar las distribuciones marginales de respuestas, pero no las distribuciones completas de respuestas y de reglas de selección.

tancia la exigencia (*b*) no se cumple: la regla de muestreo no permite la observación de ciertos valores de (\mathbf{y}, \mathbf{x}) , y por lo tanto es imposible determinar la distribución de (\mathbf{Y}, \mathbf{X}) para esos valores sin recurrir a otros supuestos. Si no se conocen ni el soporte ni la ponderación es imposible, sin recurrir a fuertes supuestos, determinar si el hecho de que falten datos para algunos valores de (\mathbf{y}, \mathbf{x}) se debe al plan de muestreo o a que la densidad de la población no tiene soporte para esos valores. Empleando estos marcos, Heckman (1987) analiza una variedad de estrategias de muestreo que presentan interés en economía, mostrando los supuestos que hacen sobre la ponderación y el modelo para resolver el problema de inferencia y remontar de la población observada a la población hipotética.

El gráfico 2 ilustra de forma simple el problema que surge de que $\omega(\mathbf{x}, \mathbf{y}) = 0$. En el gráfico 2a represento una distribución truncada de \mathbf{Y} con datos ausentes para los valores de \mathbf{Y} por debajo de *c*. Por debajo de *c* cualquier forma es posible para la verdadera densidad de la hiperpoblación. El gráfico 2b muestra una versión de regresión del mismo problema para una función de oferta de trabajo *H* expresada en términos de salarios *W*. Podemos ajustar la regresión a la muestra pero, ¿qué se puede hacer para proyectarla sobre nuevas muestras o sobre la población hipotética?

5.2. Una representación en forma de modelo de regresión del problema de selección cuando hay selección sobre variables no observables

En los trabajos de Gronau (1974), Heckman (1976a, 1976c, 1978a, 1979) y Lewis (1974) se encuentra el origen de una versión del modelo de regresión del problema de selección cuando las ponderaciones $\omega(\mathbf{y}, \mathbf{x})$ no pueden ser estimadas independientemente del modelo. Nace del modelo de Roy empleando (8) y suponiendo que (U_0, U_1, U_2) son independientes de \mathbf{X}, \mathbf{Z} . Está estrechamente relacionada con la caracterización de Lester Telser de sesgos de ecuaciones simultáneas en un sistema convencional de la Comisión Cowles.²² Uso \mathbf{Z} para representar las variables que afectan a las elecciones mientras que las \mathbf{X} son las variables que afectan a las respuestas. Es posible que \mathbf{X} y \mathbf{Z} tengan variables en común. Observamos \mathbf{Y} (véase ecuación [12]). Entonces:

$$E(\mathbf{Y}|\mathbf{X}, \mathbf{Z}, D = 1) = E(\mathbf{Y}_1|\mathbf{X}, \mathbf{Z}, D = 1) = \mu_1(\mathbf{X}) + E(U_1|\mathbf{X}, \mathbf{Z}, D=1) \quad (15a)$$

y

$$E(\mathbf{Y}|\mathbf{X}, \mathbf{Z}, D = 0) = E(\mathbf{Y}_0|\mathbf{X}, \mathbf{Z}, D = 0) = \mu_0(\mathbf{X}) + E(U_0|\mathbf{X}, \mathbf{Z}, D=1) \quad (15b)$$

Las medias condicionales de U_0 y U_1 son las "funciones de control" o funciones de sesgo tal como se introducen y definen en Heckman (1980a) y Heckman y Robb (1985, 1986). Las respuestas observadas medias (las variables del miembro izquierdo) están generadas por la media de las respuestas potenciales más un sesgo.

(22) Véase el sistema de ecuaciones (A1) en el apéndice A. Véase Telser (1964) y las discusiones en Heckman (1976c, 1978a, 2000).

Gráfico 2a
DATOS PARA $Y < c$ AUSENTES. DOS PENDIENTES POSIBLES
PARA LA DENSIDAD POR DEBAJO DE c

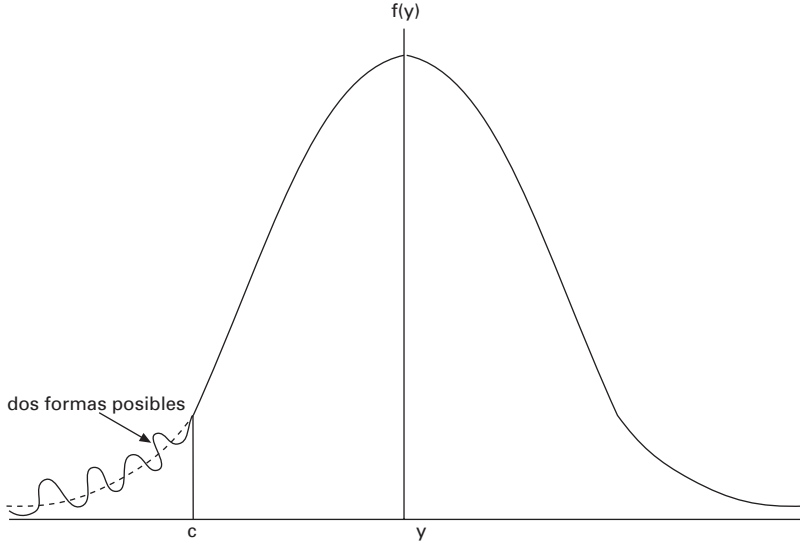
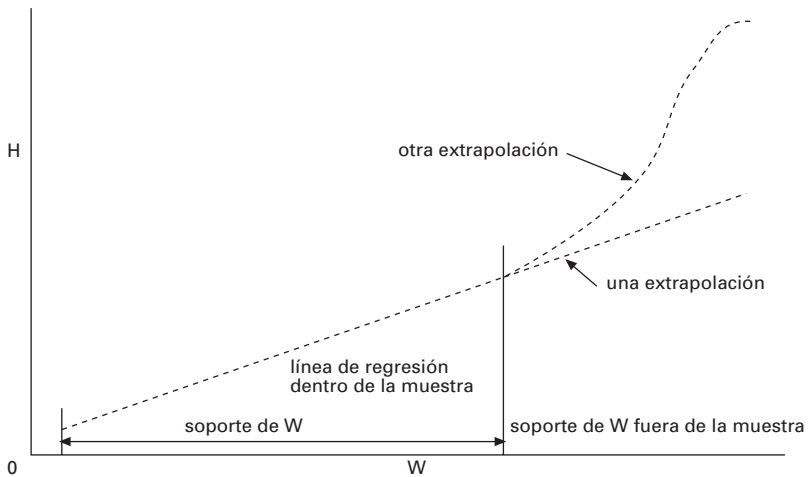


Gráfico 2b
EL PROBLEMA DE EXTRAPOLAR FUERA DE LA MUESTRA



Definamos $P(z) = \Pr(D = 1 | Z = z)$. Como consecuencia de la regla de decisión (11), demostremos en Heckman (1980a) que bajo condiciones generales, siempre podemos formular estas expresiones como

$$E(Y|X, Z, D = 1) = \mu_1(X) + K_1(P(Z)) \quad (16a)$$

y

$$E(Y|X, Z, D = 0) = \mu_0(X) + K_0(P(Z)) \quad (16b)$$

donde $K_1(P(Z))$ y $K_0(P(Z))$ son funciones de control y dependen de Z sólo a través de P . Las formas funcionales de las variables K dependen de supuestos específicos de distribución. Para un catálogo de ejemplos, véase Heckman y MaCurdy (1985).

El valor de P está relacionado con la magnitud del sesgo de selección. A medida que las muestras se hacen más representativas, $P(z) \rightarrow 1$ y $K_1(P) \rightarrow 0$. Véase el gráfico 3 que representa la función de control $K_1(P(z))$ en función de P . A medida que $P \rightarrow 1$ la muestra es cada vez más representativa puesto que la probabilidad de que cualquier individuo esté incluido en la muestra es la misma (y $P = 1$). La función de sesgo decrece con P . Podemos calcular la media poblacional de Y_1 en muestras con poca selección (P alto). En general, las regresiones sobre muestras seleccionadas presentan sesgo para $\mu_1(X)$. Combinamos el sesgo de selección con la función que nos interesa. Si hay variables en Z que no están en X , las regresiones sobre las muestras seleccionadas indicarían que estas variables "pertencen" a la regresión. La representación de (16a) y (16b) constituye la base de toda una literatura econométrica sobre sesgos de selección en funciones de selección.²³ La idea clave en toda esta literatura es el control del efecto de P sobre las relaciones ajustadas.²⁴

Las funciones de control relacionan los datos ausentes (las variables U_0 y U_1) con los observables. Bajo una serie de supuestos, es posible construir estas funciones, que incluirán parámetros desconocidos, e identificar los $\mu_0(X)$, $\mu_1(X)$ y parámetros desconocidos a partir de un análisis de regresión, a la vez que se controla el sesgo de selección (véase Heckman, 1976a; Heckman y Robb 1985, 1986; Heckman y Vytlacil, 2002).

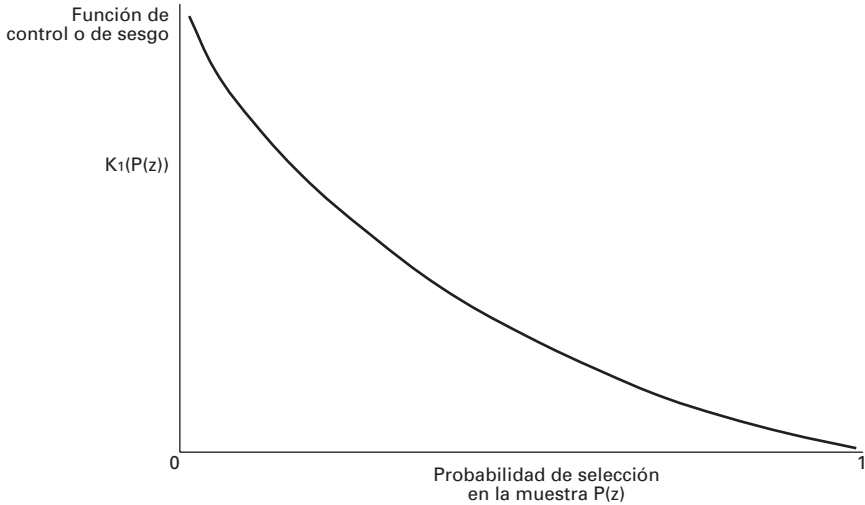
En la literatura temprana, las formas funcionales específicas de (15) y (16) han sido derivadas suponiendo que las U tenían una distribución conjunta normal.

Supuesto 2. $(U_0, U_1, U_2) \sim N(0, \Sigma)$

(23) Heckman, Ichimura, Smith y Todd (1998) presentan métodos para contrastar la adecuación de esta expresión en un contexto semiparamétrico.

(24) Heckman (1980a) propone una expansión en serie de las funciones de K_1 y K_0 en términos de polinomios de P y sugiere que para contrastar la ausencia de selección podemos basarnos en un contraste de si el conjunto completo de polinomios es estadísticamente significativo en una ecuación de respuestas. Andrews (1991) y Newey (1994) ofrecen un análisis más general.

Gráfico 3
FUNCIÓN DE CONTROL O SESGO DE SELECCIÓN
COMO FUNCIÓN DE $P(z)$



Supuesto 3. (U_0, U_1, U_2) independientes de (\mathbf{X}, \mathbf{Z}).

El supuesto 2 combinado con el supuesto 3 proporciona formas funcionales precisas para \mathbf{K}_1 y \mathbf{K}_0 . Para muestras censuradas, se desarrolló un procedimiento de estimación en dos etapas: (1) estimar $P(\mathbf{Z})$ a partir de los datos relativos a la decisión de trabajar y (2) usar $P(\mathbf{Z})$ estimada para formar $\mathbf{K}_1(P(\mathbf{Z}))$ y $\mathbf{K}_0(P(\mathbf{Z}))$ que incluyen parámetros desconocidos. (16a) y (16b) pueden entonces estimarse por regresión. Esto da lugar a una cómoda expresión lineal en los parámetros cuando $\mu_1(\mathbf{X}) = \mathbf{X}\beta_1$ y $\mu_0(\mathbf{X}) = \mathbf{X}\beta_0$.²⁵ Se desarrolló un procedimiento de regresión directo (en una sola etapa) para muestras truncadas (véase Heckman y Robb 1985, 1986). Las ecuaciones (16a) y (16b) se convirtieron en la base de toda una literatura, que generalizó y amplió los primeros modelos y que sigue viva hoy en día.

5.3. Resultados empíricos de estos modelos y sus consecuencias para la economía

El marco de regresión es útil para investigar los fenómenos microeconómicos de muestras seleccionadas en el caso general de selección

(25) En Heckman (1979) y Newey y McFadden (1994) se ofrecen correcciones para emplear $P(\mathbf{Z})$ estimado en la primera etapa. Los supuestos 2 y 3 también se emplearon para estimar el modelo por máxima verosimilitud como en Heckman (1974a, 1974c). La literatura temprana no ofrecía explicaciones claras sobre las fuentes de identificación, sobre la necesidad de las restricciones de exclusión y sobre el papel de la normalidad.

que el modelo de Roy cubre. Por lo general, no existen ponderaciones simples que puedan estimarse por separado del modelo completo y que permitan resolver el problema de selección en el modelo de Roy. Se han aplicado versiones de este modelo a problemas económicos además de a la investigación sobre oferta de trabajo y salarios.

La toma de conciencia de la importancia potencial de la selección está determinando la manera en que se interpretan datos económicos y sociales y se evalúan la efectividad de las políticas sociales. Consideremos, por ejemplo, la cuestión importante de si se ha producido alguna mejora en el estatus económico de los afro-americanos. Como lo muestra el gráfico 4a, la proporción del salario mediano del varón negro con respecto al varón blanco aumentó en Estados Unidos durante el periodo 1940-1980 y se estabilizó más adelante (véase la curva negra en el gráfico 4a). Esta estadística se cita ampliamente como justificación del conjunto completo de las políticas sociales desarrolladas durante esas décadas. En el mismo periodo, los negros se estaban retirando de la fuerza de trabajo ($P(Z)$ disminuye en ese tramo), y por lo tanto de las estadísticas utilizadas para medir los salarios, y lo estaban haciendo en mucha mayor medida que los blancos (véase gráfico 4b). Si tenemos en cuenta el sesgo muestral provocado por la retirada de los trabajadores negros de bajo salario del mercado de trabajo, el progreso económico de los hombres negros respecto al de los hombres blancos resulta prácticamente nulo y esto pone en entredicho las evaluaciones optimistas que se hacían al respecto.²⁶

Enfocar el problema de esta manera tiene repercusiones mucho más generalizadas. Afecta a la forma en que analizamos la desigualdad y sus efectos sobre el empleo y el bienestar de otras formas de organizar el mercado laboral. En discusiones europeas, a menudo se comparan desfavorablemente los mercados de trabajo de los Estados Unidos, con fuertes desigualdades y salarios bajos, con los mercados europeos con pocas desigualdades y salarios altos.

Estas comparaciones fallan por razones similares a las que surgen en las discusiones sobre la diferencia de salarios entre negros y blancos en Estados Unidos. En Europa, los parados y los que no trabajan no se contabilizan en la medición de salarios que se emplea para evaluar el comportamiento del mercado laboral. De esta manera, la desigualdad de salarios queda infravalorada y el nivel de los salarios sobrevalorado para el conjunto de la población, por el hecho de computar sólo a los que trabajan. Un artículo reciente de Blundell, Reed y Stoker (1999) señala la importancia

(26) La corrección de selección concreta empleada para producir los números que se utilizan en este gráfico tiene por objeto usar salarios medianos de trabajadores suponiendo que los trabajadores de salarios bajos son los que abandonan y que los abandonos representan menos del 50 por ciento del conjunto de la población. Butler y Heckman (1977) fueron los primeros en plantear esta cuestión. La investigación posterior de Brown (1984), Juhn (1997), Chandra (2000) y Heckman, Lyons y Todd (2000) confirma la importancia de tener en cuenta los abandonos al analizar las diferencias de salarios entre negros y blancos. La investigación sobre esta importante cuestión es muy activa.

Gráfico 4a
PROPORCIÓN DEL SALARIO MEDIANO DEL VARÓN NEGRO CON
RESPECTO AL VARÓN BLANCO, 1940-90

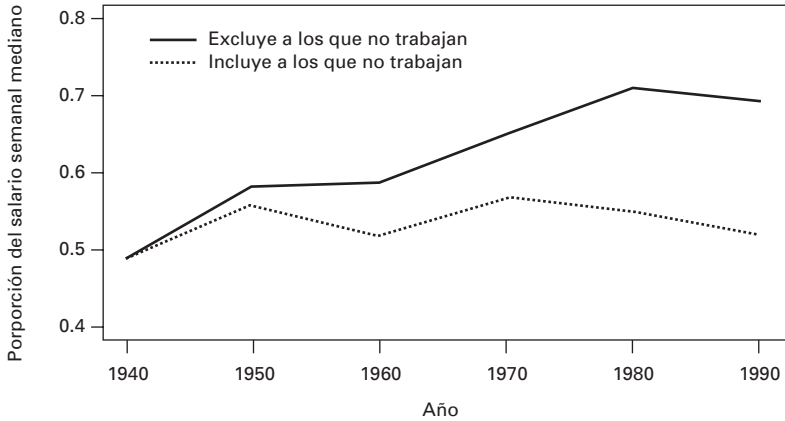
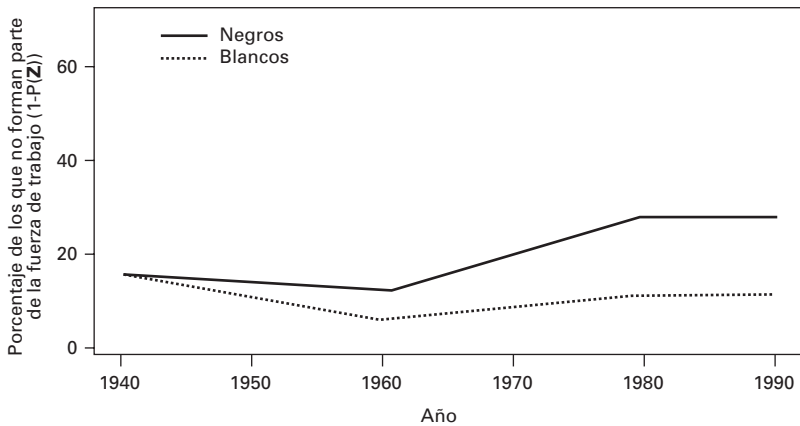


Gráfico 4b
PORCENTAJE DE VARONES QUE NO FORMAN PARTE
DE LA FUERZA DE TRABAJO, 1940-90



Fuente: Heckman y Todd (1999).

del problema de selección en el contexto inglés. Los datos ingleses revelan un incremento en los salarios reales de los trabajadores en el periodo 1978-94 (véase la curva superior en el gráfico 5a). Al mismo tiempo la proporción de personas que trabajan ha disminuido (véase gráfico 5b), y si tenemos en cuenta los abandonos, el nivel y la tasa de crecimiento de los salarios reales se ven reducidos. Por otra parte, el incremento observado en los salarios reales podría ser consecuencia de una mejora en los recursos de mano de obra especializada y de un aumento en la remuneración

de la misma (por ejemplo, $\mu_1(\mathbf{X})$) o de mejoras en el sector ajeno al mercado, lo cual cambia la media condicional de las variables no observables en la ecuación de los salarios eliminando del mercado de trabajo a los trabajadores que tienen salarios potencialmente bajos. El ajuste para tomar en cuenta la selección (las dos curvas inferiores en el gráfico 5a) reduce considerablemente el incremento estimado de los salarios.

Gráfico 5a
PREDICIONES DE SALARIO DE UN MODELO MICROECONÓMICO,
SALARIO AGREGADO Y CORRECCIONES CON NUEVA
BASE EN 1978

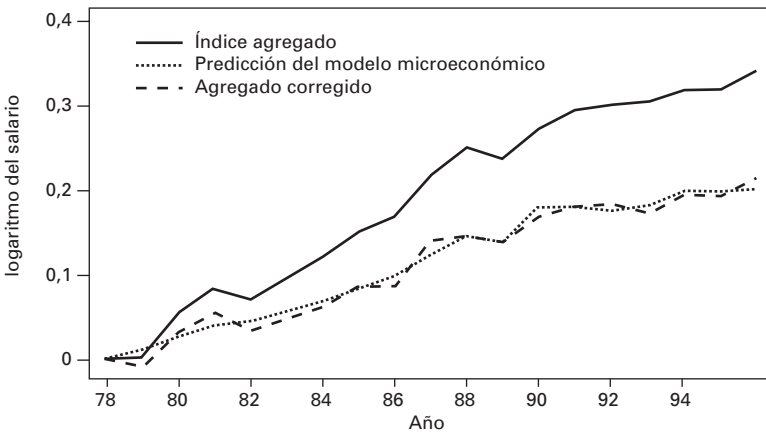
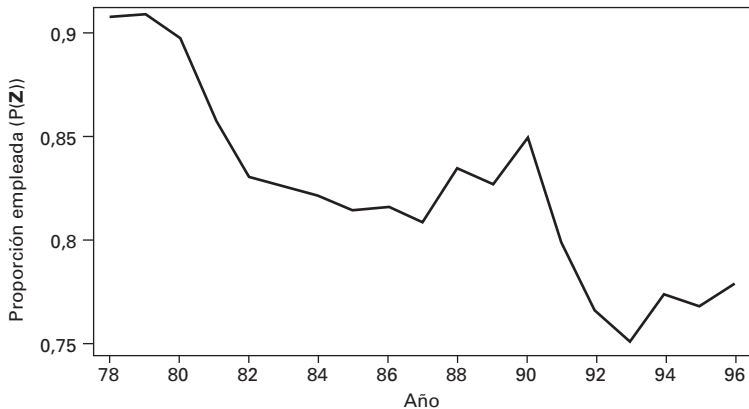


Gráfico 5b
SALARIOS Y PARTICIPACIÓN EN EL MERCADO DE TRABAJO,
VARONES BRITÁNICOS



Fuente: Blundell *et al.* (1999).

El tener en cuenta la selección también afecta a las medidas de la variabilidad de los salarios en el ciclo (Bils, 1985). Los individuos con salarios bajos abandonan la fuerza de trabajo en las recesiones (y por tanto no se encuentran representados en las estadísticas empleadas para medir los salarios de los trabajadores) y vuelven a ella en época de auge. Al cambiar la composición, se compensa parcialmente la variabilidad de las medidas de los salarios. De esta manera, las medidas de los salarios parecen mostrar una "variabilidad insuficiente" a lo largo del ciclo económico. Cuando se estima un modelo de Roy de auto-selección con varios sectores del mercado, el argumento se hace más sutil. A lo largo del ciclo, no solamente hay entradas y salidas en la fuerza de trabajo, sino que también se dan migraciones de los trabajadores de un sector a otro dentro de la fuerza de trabajo. Por lo tanto, las medidas de los salarios no representan solamente el precio de los servicios del trabajo. Además del efecto de selección estándar, las medidas de los salarios incluyen el efecto de aplicar ponderaciones sobre las distintas combinaciones de aptitudes usadas por los trabajadores a lo largo del ciclo (Heckman y Sedlacek, 1985).

Si tenemos en cuenta el margen extensivo así como el intensivo, esto tiene repercusiones sobre nuestra percepción del comportamiento del mercado laboral. Consideremos la ecuación (2). Observamos únicamente las horas de trabajo de los trabajadores. Para centrarnos en el problema de selección, hagamos el falso supuesto de que se han observado los salarios de todos los individuos e ignoremos cualquier endogeneidad en los salarios (de tal manera que W es independiente de U). Usemos $D = 1$ para representar el trabajo. La oferta de trabajo observada condicionada a W y a \mathbf{X} es

$$E(H|W, \mathbf{X}, D = 1) = H(W, \mathbf{X}) + E(U|W, \mathbf{X}, D = 1) \quad (17)$$

donde el parámetro de la oferta de trabajo marshaliana (o parámetro causal) de los salarios es $\partial H/\partial W$, es decir el cambio en la oferta de trabajo debido a un cambio en los salarios. Si tenemos en cuenta los efectos renta, podemos construir a partir de aquí una función de oferta de trabajo con utilidad constante que nos permita realizar un análisis de bienestar para, por ejemplo, calcular el excedente del consumidor. Pero en una función de oferta de trabajo ajustada a una muestra seleccionada de trabajadores, se pueden identificar dos efectos de los salarios: un efecto marshaliano y un efecto de composición o de selección debido a las entradas y salidas de la fuerza de trabajo (Heckman, 1978c). De esta manera

$$\frac{\partial E(H|W, \mathbf{X}, D = 1)}{\partial W} = \frac{\partial H(W, \mathbf{X})}{\partial W} + \frac{\partial E(U|W, \mathbf{X}, D = 1)}{\partial W} \quad (18)$$

El segundo término es un efecto de selección, o de composición, que surge del cambio en la composición de las variables no observables debido a la entrada o a la salida de individuos de la fuerza de trabajo inducida por el cambio de los salarios. No se trata de un cambio ceteris paribus que corresponda a parámetros de la teoría clásica del consumidor. La ecuación (18) no nos dice en qué medida un trabajador dado cambiaría su oferta de trabajo al cambiar los salarios. Sin embargo, sí que nos informa de lo que un cambio de salario dentro de la muestra

produciría sobre la oferta de trabajo media. Esto contesta a la pregunta de evaluación del tipo de la pregunta 1 de Marschack. Bajo condiciones adecuadas para el soporte de W , la ecuación (18) se puede usar para estimar los efectos dentro de la muestra de los impuestos sobre la oferta de trabajo.²⁷

Las elasticidades de la oferta agregada de trabajo incluyen efectos de la entrada y salida de individuos de la fuerza de trabajo así como los efectos del desplazamiento a lo largo de la curva marshalliana de oferta de trabajo. Esta observación simple ha tenido efectos sustanciales sobre la especificación, la estimación, y la interpretación de la oferta de trabajo en los modelos macroeconómicos.

A principios de los años 80, apareció toda una literatura en macroeconomía que afirmaba que las elasticidades agregadas de la oferta de trabajo eran demasiado pequeñas y los movimientos de los salarios demasiado importantes como para que el modelo neoclásico del mercado de trabajo permitiese explicar los datos de series temporales de Estados Unidos. Ya he expuesto por qué la variación de la medida de los salarios infravalora la variación real del precio del trabajo. En Heckman (1984) señalo incluso que la literatura en macroeconomía se centraba exclusivamente en el componente de la oferta de trabajo con solución interior, el primer término del miembro derecho de (18), e ignoraba el efecto de selección provocado por la entrada y salida de trabajadores en/de la fuerza de trabajo. Dado que la mitad de la oferta de trabajo agregada se encuentra en el margen extensivo (Coleman 1984), donde la elasticidad de la oferta de trabajo es mayor, los cálculos estándar de los años 80 estaban infravalorando la verdadera elasticidad de la oferta de trabajo agregada y subestimaban también la capacidad del modelo neoclásico de oferta de trabajo para dar cuenta de las fluctuaciones en la oferta de trabajo agregada. Cuando se comenzó a tomar en cuenta la elección en los márgenes extensivos, cambió la percepción y la forma de modelizar el mercado laboral de los macroeconomistas (véase, por ejemplo, Hansen, 1985; Rogerson, 1988).

Los avances empíricos en la literatura de la oferta de trabajo confirmaron estas conclusiones. Anteriormente, la evidencia había puesto en entredicho la validez empírica del modelo (10). Los costes fijos del trabajo hacen que sea poco probable que el índice de horas de trabajo esté tan estrechamente unido a la decisión de participación en el mercado de trabajo como sugiere el modelo. Cuando los trabajadores se incorporan al mercado laboral, suelen trabajar muchas horas, y no pocas horas, como sugiere la ecuación (10) si las U están distribuidas normalmente. Heckman (1976a, 1980b) y Cogan (1981) propusieron un modelo más general con coste fijos en el cual las ecuaciones de participación y de horas de trabajo están menos estrechamente relacionados. Esto da lugar a una elasticidad aún mayor para el segundo término de la ecuación (18). La evi-

(27) La condición necesaria es el supuesto a en la nota 9 de la sección 2.

dencia también puso en entredicho la validez del supuesto de normalidad, especialmente para los datos relativos a las horas de trabajo. Las distribuciones de las horas trabajadas en muchos países han revelado que presentan picos en el número estándar de horas trabajadas.²⁸ Esto ha llevado a desarrollos destinados a relajar el supuesto de normalidad usado en los primeros modelos.

5.4. Identificación

Gran parte de la literatura econométrica sobre el problema de selección combina discusiones sobre la identificación (el remontar de las poblaciones generadas por las reglas de selección a la población de origen) con discusiones sobre la estimación para resolver los problemas planteados para pasar de las muestras observadas a las poblaciones hipotéticas.²⁹ Resulta analíticamente útil distinguir entre las condiciones necesarias para identificar el modelo de selección a partir de datos ideales, y los numerosos e importantes problemas prácticos que surgen para estimar el modelo. Es esencial comprender las fuentes de identificación de un modelo para entender hasta qué punto lo que estamos extrayendo de un modelo empírico es consecuencia de lo que hemos incorporado en él.

En 1985, un congreso en el *Educational Testing Service* (ETS) reunió a economistas y estadísticos y permitió contrastes interesantes de puntos de vista sobre la construcción de modelos causales y de modelos de selección (véase Wainer, 1986, reeditado en 2000).³⁰ En esa conferencia, Holland (1986) empleó la ley de expectativas iteradas para expresar la distribución condicionada a \mathbf{X} de una respuesta que denominaremos Y_1 , de la manera siguiente:

$$F(Y_1|\mathbf{X}) = F(Y_1|\mathbf{X}, D = 1) \Pr(D = 1|\mathbf{X}) + F(Y_1|\mathbf{X}, D = 0) \Pr(D = 0|\mathbf{X}) \quad (19)$$

Del análisis de (11) y (12), sabemos que observamos Y_1 solamente si $D = 1$. En una muestra censurada, podemos identificar $F(Y_1|\mathbf{X}, D = 1)$, $\Pr(D = 1|\mathbf{X})$, y por lo tanto $\Pr(D = 0|\mathbf{X})$. No observamos Y_1 cuando $D = 0$, por lo que no identificamos $F(Y_1|\mathbf{X})$. En trabajos independientes, Smith y Welch (1986) hicieron una descomposición similar de las medias condicionales (reemplazando F por E).

Holland cuestionó que se pudiera identificar $F(Y_1|\mathbf{X})$ y comparó brevemente los modelos de selección con otros enfoques. Smith y Welch

(28) Véanse los artículos en la publicación especial sobre impuestos y oferta de trabajo en países industrializados del *Journal of Human Resources* (vol. 25, verano 1990).

(29) Véase Heckman (2000) para un definición precisa de la identificación.

(30) El intercambio de pareceres que mantuve con Tukey y que quedó registrado en ese volumen pone de manifiesto las divergencias entre estadísticos y econométricos en cuanto al valor que se le otorga a discutir la identificación y a hacer distinciones causales.

(1986) y algunos de los autores que se encontraban en el congreso del ETS discutieron sobre cómo acotar $F(Y_1|\mathbf{X})$ (o $E(Y_1|\mathbf{X})$) aplicando cotas a los componentes ausentes $F(Y_1|\mathbf{X}, D=0)$ ($E(Y_1|\mathbf{X}, D=0)$ respectivamente).³¹ Un precedente claro de esta idea fue el trabajo de Peterson (1976), que desarrolló cotas no paramétricas para el modelo de riesgos en competencia (*competing risk model*) de análisis de duración y que es matemáticamente idéntico al modelo de Roy de las ecuaciones (11) y (12).³² En el apéndice B examino algunos de los avances recientes en esta literatura.

El supuesto de normalidad que se usaba frecuentemente en la literatura temprana fue puesto entonces en entredicho. Arabmazar y Schmidt (1981) y Goldberger (1983) presentaron unos análisis de Monte Carlo poniendo de manifiesto sesgos sustanciales para modelos con respuestas continuas cuando se había supuesto normalidad, siendo así que el modelo en realidad no era normal. La evidencia empírica está más dividida. El supuesto de normalidad no es malo si se trata de analizar modelos de auto-selección para el logaritmo del salario una vez que se toma en cuenta el truncamiento y la auto-selección.³³ Véanse los gráficos 6a y 6b de Heckman y Sedlacek (1985) y el trabajo relacionado con ello de Blundell *et al.* (1999). Estos estudios demuestran que cuando se tiene en cuenta el sesgo de selección, un modelo normal se ajusta a las distribuciones de salarios de forma bastante satisfactoria. Olsen (1980) y Lee (1982) ofrecen ampliaciones precoces no normales, aunque paramétricas, del primitivo enfoque normal de Roy. Heckman y MaCurdy (1985) presentan una síntesis de esta literatura. Heckman (1980a) presenta un primer estimador no paramétrico de la función de control recurriendo a una expansión en series de P .

Heckman y Honoré (1990) consideran la identificación del modelo de Roy bajo una serie de condiciones. Establecen que, bajo normalidad, el modelo está identificado incluso si no hay regresores, y por lo tanto no hay restricciones de exclusión. También establecen que un modelo está identificado (salvo por los subíndices) incluso si uno observa únicamente Y , pero no sabe si se trata de Y_1 o de Y_0 . El supuesto inicial de normalidad empleado en los modelos de selección estaba basado en potentes supuestos sobre la forma funcional.³⁴

(31) Smith y Welch emplean su análisis para acotar los efectos que tienen los abandonos sobre las diferencias de salarios entre negros y blancos de las que se trata en la subsección 5.3.

(32) El modelo de riesgos en competencia sustituye $\max(Y_0, Y_1)$ por $\min(Y_0, Y_1)$.

(33) La normalidad de las variables latentes resulta ser un supuesto aceptable para modelos de elección discreta salvo bajo circunstancias extremas (Todd 1996).

(34) Potentes pero contrastables. El modelo está sobreidentificado. Véase, por ejemplo, Bera, Jarque y Lee (1984) para los contrastes sobre supuestos de distribución en una clase de modelos de variable dependiente limitada.

Gráfico 6 DISTRIBUCIÓN DEL LOGARITMO DE LOS SALARIOS OBSERVADOS VS. DISTRIBUCIÓN DEL LOGARITMO DE LOS SALARIOS PREDICHOS POR EL MODELO GENERALIZADO DE ROY

Gráfico 6a
Sector no manufacturero

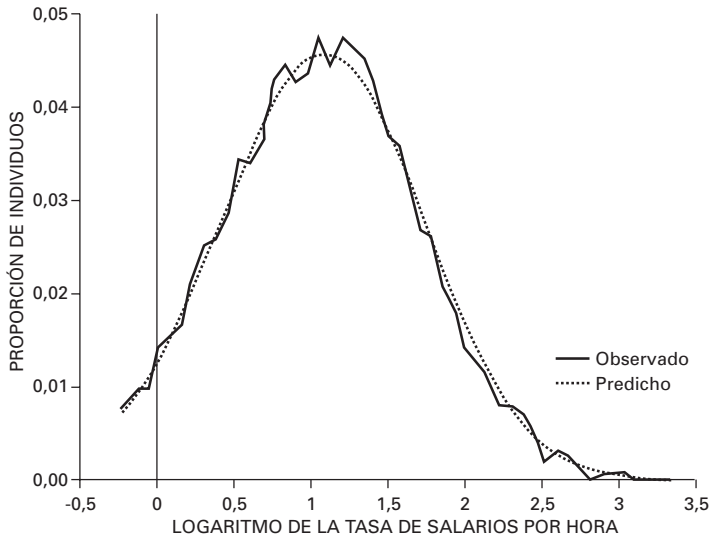
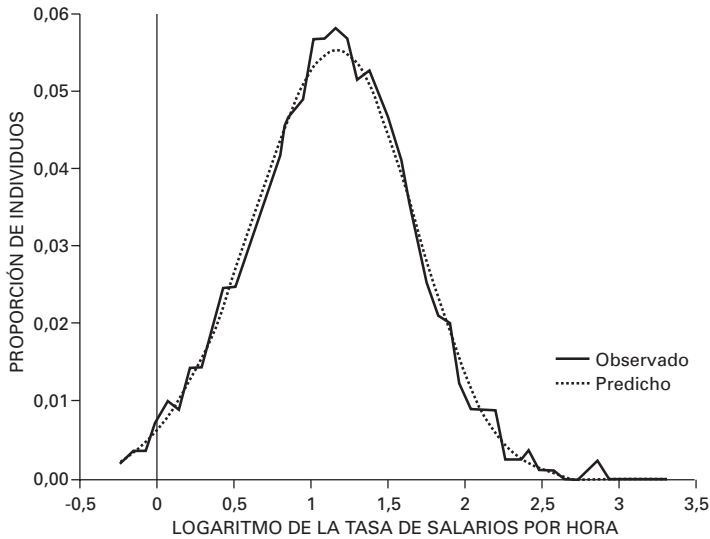


Gráfico 6b
Sector manufacturero



Fuente: Heckman y Sedlacek (1985).

Heckman y Honoré desarrollan un modelo de Roy no paramétrico y establecen las condiciones para las cuales la variación de los regresores a lo largo del tiempo o entre individuos puede identificar el modelo de forma no paramétrica. Se pueden sustituir los supuestos de distribución por diferentes tipos de variación en los datos para identificar la versión de Roy del modelo de selección. Heckman y Smith (1998) amplían esta línea de análisis al modelo de Roy generalizado. La regla de decisión (12) desempeña un papel esencial para asegurar la identificación del modelo de selección. En un caso más general en que Y_2 podría depender de Y_1 - Y_0 así como de otras variables no observables, es posible identificar sólo una parte del modelo completo de selección, incluso en presencia de variaciones importantes de los regresores entre individuos o a lo largo del tiempo. Cuando los modelos no están identificados, sigue siendo posible acotar los parámetros cruciales, y se ha desarrollado toda una literatura alrededor de esta idea. En el apéndice B se examina esta literatura. Heckman, Ichimura, Smith y Todd (1998), entre otros, analizan la estimación semiparamétrica de modelos de selección (véanse también Robinson, 1988; Ahn y Powell, 1993).

6. MICRODINÁMICA Y DATOS DE PANEL: HETEROGENEIDAD VERSUS DEPENDENCIA DEL ESTADO Y OFERTA DE TRABAJO DE CICLO VITAL

Los primeros datos microeconómicos eran de corte transversal. Así, los primeros trabajos que se realizaron sobre elección discreta, variables dependientes limitadas y modelos con variables endógenas mixtas continuas y discretas, tenían una naturaleza de corte transversal pero se centraban exclusivamente en explicar la variación entre individuos en un momento dado. Esto dio lugar a múltiples interpretaciones de las fuentes de los no observables en (7) y (8). Los modelos de utilidad aleatoria, que se introdujeron en la literatura sobre elección discreta, los interpretaban como shocks en las preferencias temporalmente independientes (McFadden, 1974), especialmente cuando se consideraba la elección discreta. Otras interpretaciones eran que se debían a (a) variaciones sistemáticas en las preferencias no observadas que eran estables en el tiempo y (b) características omitidas de las elecciones y de los agentes que pueden o no ser estables a lo largo del tiempo.³⁵

La aparición de datos de panel en la economía del trabajo, que se debió en gran parte a la contribución de Morgan y de su grupo en el Institute for Survey Research de la Universidad de Michigan, permitió explorar estas fuentes de variación de manera más sistemática (véase Stafford, 2001). Se trató de un avance importante especialmente para el estudio de la oferta de trabajo femenina.

Mincer (1962) empleó una versión implícita del modelo de utilidad aleatoria para argumentar que los datos de participación en la fuerza de tra-

(35) Heckman y Snyder (1997) estudian la evolución histórica de estas ideas.

bajo de corte transversal podían ser utilizados para estimar los efectos renta y sustitución de Hicks-Slutsky. Su idea era que H en la ecuación (1) medía la parte de sus vidas que los individuos dedicaban a trabajar, de modo que el tiempo de ocio es perfectamente sustituible a lo largo del tiempo, el momento de la oferta de trabajo es irrelevante y podría ser determinado lanzando una moneda al aire. Por lo tanto, una regresión de la tasa de participación a la fuerza de trabajo sobre W identificaría un efecto de salarios de Hicks-Slutsky.³⁶ Ben-Porath (1973) supuso, en su lugar, que los shocks eran características permanentes y estables de los individuos e interpretó la participación en la fuerza de trabajo como una solución de esquina, demostrando que la regresión de las tasas de participación en la fuerza de trabajo sobre los salarios, identificaría los parámetros de la distribución del gusto por el trabajo y no el efecto de sustitución de Hicks-Slutsky (es decir que definiría los parámetros de $\Pr(D = 1 | X)$ de la ecuación [9]).

Éste es un aspecto importante también para entender los datos de empleo y desempleo. Se constata con regularidad empírica en el análisis de datos de desempleo que aquéllos que han sido desempleados en el pasado o han trabajado en el pasado tienen mayores posibilidades de mantener su condición (de desempleado o de trabajador) en el futuro que de cambiarla. ¿Se debe esto a un efecto causal de la condición de cada uno o se trata de una manifestación de una característica estable (por ejemplo, algunas personas son más vagas que otras y las observables son persistentes)? Hay una teoría en macroeconomía que se construyó sobre la premisa de que la promoción del trabajo mediante políticas macroeconómicas aportaría mayores niveles de empleo (Phelps, 1972). La distinción entre efectos verdaderos y espurios viene a ser la distinción entre dependencia del estado verdadera y espuria.

En una serie de artículos, desarrollé modelos econométricos empleando datos de panel para investigar estos aspectos. Una parte de los estudios se basa en el modelo de las ecuaciones (10a)-(10c) pero las sitúa en un contexto de ciclo vital. Mi trabajo sobre la oferta de trabajo en el ciclo vital (Heckman, 1974b, 1976b) demostraba que funciones de demanda con utilidad marginal constante de la riqueza constituían el concepto relevante para analizar la evolución de la oferta de trabajo a lo largo del ciclo vital en entornos de certidumbre perfecta o que tienen mercados de demanda contingente. Basándonos en este trabajo y empleando el trabajo de Heckman (1974b) y la tesis de investigación de MaCurdy (1978, 1981), MaCurdy y yo (Heckman y MaCurdy, 1980) formulamos y estimamos una versión de ciclo vital del modelo de las ecuaciones (10) que interpretaba una de las no observables clave del modelo, como la utilidad marginal de la riqueza, λ . En los supuestos económicos, establecimos que λ es una variable no observable estable o un efecto fijo derivado de la teoría económica. Los modelos que desarrollamos extendían, por primera

(36) Heckman (1978c) ofrece un análisis formal y presenta la relación que guarda con el modelo de utilidad aleatoria estudiado por McFadden (1974, 1981).

vez, modelos para variables dependientes limitadas, datos sistemáticamente ausentes, y variables endógenas mixtas continuas y discretas a un marco de datos de panel.³⁷ La evidencia que aportamos y el trabajo relacionado que realicé con Willis (Heckman y Willis, 1977) sugerían que procedía hacer una síntesis de los puntos de vista de Ben-Porath y Mincer, y que no era apropiada una especificación de utilidad aleatoria pura. Este marco de trabajo ha sido ampliado en importantes artículos de Altug y Miller (1990, 1998) para incluir el capital humano y la incertidumbre.

En trabajos relacionados, generalicé los modelos estáticos de corte transversal de elección discreta a un marco dinámico y empleé esta generalización para tratar el problema de la heterogeneidad versus la dependencia del estado. Este problema fundamental puede entenderse de forma más simple considerando los siguientes esquemas de urnas (Heckman, 1981a).

En el primer esquema hay un número I de individuos que poseen urnas con el mismo contenido de bolas rojas y negras. En un número T de extracciones independientes, el individuo i saca una bola y luego la devuelve a su urna. Si saca una bola roja en la extracción t , el individuo i vive el suceso (siendo el suceso estar empleado o desempleado, por ejemplo). Si saca la bola negra, el individuo i no vive el suceso. Este modelo corresponde a un modelo simple de Bernoulli y captura la idea esencial que subyace en el proceso de elección en el trabajo de McFadden (1974) sobre elección discreta. Con los datos generados por este esquema de urnas, no se observaría la regularidad empírica de que una persona que ha vivido el suceso en el pasado tiene más probabilidades de experimentarlo en el futuro. Independientemente de su historial de sucesos, todos los individuos tienen la misma probabilidad de vivir el suceso.

El segundo esquema de urnas genera datos que darían lugar a un efecto de los sucesos pasados sobre los sucesos presentes debido únicamente a la heterogeneidad. En este modelo los individuos tienen distintas urnas que difieren en su contenido de bolas rojas y negras. Como en el primer modelo, el muestreo se hace restituyendo las bolas. Sin embargo, por oposición al primer modelo, la información relativa a la experiencia pasada de un individuo proporciona información útil para localizar la posición del individuo en la distribución poblacional de la composición de las urnas.

El registro de las experiencias pasadas de una persona puede emplearse para estimar la composición de la urna de un individuo determinado. La probabilidad condicional de que el individuo i experimente el suceso en el momento t es función de su experiencia pasada del suceso. El contenido de cada urna no se ve afectado por los resultados reales y es, de hecho, constante. No existe una verdadera dependencia del estado.

(37) Browning, Deaton y Irish (1985) adaptan esta idea a datos de sección cruzada repetida empleando métodos estándares para el análisis de cohortes sintéticos.

El tercer esquema de urnas genera datos caracterizados por una verdadera dependencia del estado. En este modelo, los individuos comienzan con urnas de contenido idéntico. En cada extracción, el contenido de la urna cambia *como consecuencia del resultado de la extracción*. Por ejemplo, si una persona saca una bola roja y vive el suceso, se añaden nuevas bolas rojas a la urna. Los resultados posteriores están influenciados por los resultados anteriores porque el conjunto de elección establecido para las extracciones posteriores ha sido alterado como consecuencia de haber experimentado el suceso.

Puede construirse una variante del tercer esquema que corresponde a un modelo de renovación. En este esquema, se añaden nuevas bolas rojas a la urna de un individuo en extracciones sucesivas de bolas rojas hasta que se extraiga una bola negra, y entonces todas las bolas rojas que habían sido añadidas cuando se había extraído una bola roja en las extracciones más recientes se quitan de la urna. La composición de la urna es entonces idéntica a la que tenía antes de que se extrajera de la misma la primera bola roja. Un modelo que corresponde a costes fijos para la entrada en la fuerza de trabajo es una variante del esquema de reposición en el que se añaden nuevas bolas a la urna de un individuo solamente en la primera extracción de una bola roja en cualquier tanda de extracciones de bolas rojas.

El elemento crucial que diferencia al tercer esquema del segundo es que el contenido de la urna (el conjunto de elección) se ve alterado como consecuencia de la experiencia previa. El punto clave no es que el conjunto de elección cambie de unas extracciones a otras, sino que cambia de una forma que depende de los resultados anteriores del proceso de elección. Para aclarar este punto, resulta útil considerar un cuarto esquema de urnas que corresponde a modelos con tipos de heterogeneidad más variables que introduzco de forma más concreta más adelante.

En este modelo, cada individuo comienza con urnas idénticas, exactamente como en el primer esquema de urnas. Después de cada extracción, pero independientemente del resultado de ésta, el contenido de la urna de cada individuo cambia al quitar una proporción de bolas de forma aleatoria y sustituir las bolas que han sido quitadas por un grupo de bolas que ha sido seleccionado de forma aleatoria de una urna mayor (que contiene un gran número de bolas de ambos colores). Si las urnas individuales no se rellenan por completo tras cada extracción, la información sobre los resultados de extracciones anteriores es útil para predecir los resultados de extracciones futuras, aunque la información que aportan extracciones anteriores disminuye con el aumento del tiempo transcurrido desde que se realizaron. Como en el segundo y tercer modelo de urnas, los resultados anteriores nos ofrecen información sobre el contenido de cada urna. Por oposición al segundo modelo, el cuarto modelo es un esquema en el que la información se deprecia ya que el contenido de las urnas cambia de forma aleatoria. Contrariamente al tercer modelo, el contenido de la urna no cambia como consecuencia de un resultado del proceso de elección.

En la literatura sobre la participación femenina en la fuerza de trabajo, los modelos de homogeneidad extrema (que corresponden al modelo de

urna 1) y los de extrema heterogeneidad (que corresponden al modelo de urna 2 en el que las urnas o son completamente negras o completamente rojas) son ambas consistentes con la evidencia de corte transversal. Ése es el contraste entre Mincer (1962) y Ben-Porath (1973). Heckman y Willis (1977) estiman un modelo de heterogeneidad en las probabilidades de participación femenina en la fuerza de trabajo que constituye un análogo de tipo probit al modelo de urna 2.

El modelo de urna 3 presenta un interés especial. Concuere con la teoría del capital humano y otros modelos que insisten sobre el impacto de la experiencia laboral anterior en las elecciones de trabajo actuales. La inversión de capital humano adquirida por la formación que proporciona la propia experiencia de trabajo puede generar dependencia del estado estructural. Los costes fijos acarreados por aquéllos que se unen a la fuerza de trabajo pueden también generar dependencia del estado estructural como en un proceso de renovación. También puede tener este efecto el capital humano específico del trabajo por turnos. Este modelo de urna es también consistente con modelos de elección psicológica en los cuales, como consecuencia de un estímulo de trabajo, las preferencias de las mujeres se ven modificadas de tal manera que la actividad de la fuerza de trabajo aumenta (Atkinson, Bower y Crothers, 1965) o con modelos económicos de formación de hábito.

Los datos de panel pueden emplearse para discriminar entre estos modelos. Por ejemplo, el segundo modelo de urnas implica que la probabilidad de que una mujer participe en la fuerza de trabajo no cambia según su experiencia laboral. Una implicación del tercer modelo en el caso general es que las probabilidades de participación cambian con la experiencia laboral. Para diferenciar estos dos modelos, existe un método que emplea historiales individuales suficientemente largos de la fuerza de trabajo para estimar la probabilidad de participación en diferentes subintervalos del ciclo de vida. Si las probabilidades estimadas para una mujer dada no cambian en diferentes momentos del ciclo de vida, no existe ninguna evidencia de dependencia del estado estructural.

Heckman (1981b, 1981c) desarrolla una clase de procesos estocásticos discretos que generalizan el modelo de elección discreta de McFadden (1974) a un contexto dinámico. Esta clase es lo suficientemente general como para poder contrastar los cuatro esquemas de urnas y presentar un marco para elecciones discretas dinámicas.³⁸ Heckman y Singer (1986) presentan una generalización explícita del modelo de McFadden en la que unos shocks de Weibull ocurren a intervalos de tiempo de una Poisson. Dagsvik (1994) ofrece una generalización a un contexto temporal continuo.

(38) Heckman y Willis (1973, 1976) presentan un prototipo para este tipo de modelos en su análisis de dinámica de la fertilidad. Lillard y Willis (1978) aplican este modelo al análisis de dinámica de los ingresos.

Heckman (1978d) demuestra que es posible emplear contraste de rachas para distinguir entre heterogeneidad y dependencia del estado. La intuición detrás de este contraste es que, si el supuesto de heterogeneidad pura es apropiado, lo importante para predecir la probabilidad actual de cierto resultado no es el orden en que los sucesos acontecen sino solamente el número total de experiencias pasadas de los mismos. Si la dependencia del estado es relevante, entonces el momento en que acontecieron los sucesos pasados también lo es. Esta idea es la base de un trabajo reciente de Honoré y Kyriazidou (2000) para analizar modelos de datos discretos con variables dependientes retardadas. También lo emplean Chiapporu y Heckman (2000) para distinguir la selección adversa de modelos de riesgo moral. Heckman y Borjas (1980) presentan un estudio relacionado para análisis de duración en tiempo continuo y lo aplican a la importante cuestión de política económica de si el desempleo en el pasado causa desempleo en el futuro o si el desempleo pasado es solamente un síntoma de que una persona sea propensa al desempleo. Para el caso de hombres jóvenes en Estados Unidos, llegaron a la conclusión de que la segunda opción es la más probable. La experiencia pasada de desempleo no tiene efectos durables sobre el empleo en el futuro. Heckman (1981a) analiza el empleo femenino y concluye que para las mujeres casadas de más edad (de entre 45 y 59 años de edad), el haber estado empleada en el pasado aumenta la probabilidad de ser empleada en el futuro incluso teniendo en cuenta los efectos de las variables no observables. Layard, Nickell y Jackman (1991) resumen la reciente evidencia relativa a una verdadera dependencia del estado extraída de los historiales de desempleo.

El cuadro 2 (extraído de Heckman, 1981a) revela que la heterogeneidad y la persistencia temporal son características de los datos sobre la participación de las mujeres en la fuerza de trabajo. En el cuadro queda registrado el historial de empleo de las mujeres en periodos de tres años; 1 indica que se ha trabajado en ese año y 0 refleja que no se ha trabajado. Si el modelo clásico de utilidad aleatoria caracterizase los datos, todas las filas con el mismo número de 1 y de 0, independientemente del orden en el que aparecen, deberían ser aproximadamente iguales, y claramente no lo son, especialmente en el caso de las mujeres jóvenes. Las que trabajan tienden a seguir trabajando, y las que no trabajan tienden a seguir desempleadas. Esto es cierto incluso después de condicionar en función de las variables no observables. Las variables no observables persistentes en el tiempo son una característica esencial de los datos microeconómicos que afecta a nuestra estimación de los modelos y nuestra interpretación de los datos.

Los primeros trabajos sobre este tema trataban la heterogeneidad persistente empleando distribuciones paramétricas para las no observables.³⁹

(39) Para una visión panorámica de modelos de mixing en demografía, véase Sheps y Menken (1973). Heckman y Willis (1973, 1976, 1977) estiman modelos de fertilidad y de oferta de trabajo con modelos paramétricos de mixing. Lancaster (1979, 1990) y Lancaster y Nickell (1980) aplican estos modelos para analizar los periodos de desempleo.

Una de las cuestiones importantes consistía en saber si era posible distinguir la heterogeneidad de la dependencia del estado sin emplear una estructura paramétrica. Heckman (1978c), empleando un análisis de contrastes de rachas, demostró que era posible contrastar las dos fuentes de dependencia. La pregunta que quedaba en el aire era si era posible medir las contribuciones relativas como lo había hecho Heckman (1981b,1981c) utilizando estructuras paramétricas.

Cuadro 2
PATRÓN DE RACHAS EN LOS DATOS

Patrón de rachas				Patrón de rachas			
1968	1969	1970	Observaciones	1971	1972	1973	Observaciones
A. Mujeres de entre 45 y 59 años en 1968							
0	0	0	87	0	0	0	96
0	0	1	5	0	0	1	5
0	1	0	5	0	1	0	4
1	0	0	4	1	0	0	8
1	1	0	8	1	1	0	5
0	1	1	10	0	1	1	2
1	0	1	1	1	0	1	2
1	1	1	78	1	1	1	76
B. Mujeres de entre 30 y 44 en 1968							
0	0	0	126	0	0	0	133
0	0	1	16	0	0	1	13
0	1	0	4	0	1	0	5
1	0	0	12	1	0	0	16
1	1	0	24	1	1	0	8
0	1	1	20	0	1	1	19
1	0	1	5	1	0	1	8
1	1	1	125	1	1	1	130

Fuente: Heckman (1981a).

Nota: 1 corresponde a trabajar en el año, 0 corresponde a no trabajar.

Elbers y Ridder (1982) demostraron que era posible, desde el punto de vista teórico, en el contexto de un modelo de riesgo proporcional en análisis de duración, pero no ofrecieron un algoritmo de estimación. Heckman y Singer (1984) amplían su análisis para desarrollar un método de estimación consistente. Cameron y Heckman (1998, 2001) y Hansen, Heckman y Vytlačil (2000) generalizan y extienden ese análisis a respuestas en tiempo discreto con variables de respuesta a la vez discretas y continuas. Un importante descubrimiento empírico del trabajo de Heckman y Singer, que ha sido copiado en numerosos estudios posteriores, es que la distribución de no observables puede aproximarse con mixturas finitas de baja

dimensión o "tipos". Esto ha resultado útil para el análisis de elección discreta dinámica (véase Eckstein y Wolpin, 1999).

El campo de la elección discreta dinámica ha progresado muchísimo y es un área de investigación muy activa. Flinn y Heckman (1982) presentan el primer modelo estructural riguroso de elección discreta dinámica en el contexto de un modelo de búsqueda para el desempleo. Investigan la identificación no paramétrica del modelo de búsqueda y presentan una serie de teoremas de identificación que también se han revelado útiles en el contexto de la identificación de modelos de subastas (véanse, por ejemplo, Laffont, Ossard y Vuong, 1995; Donald y Paarsch, 1996; Hong, 1998). Trabajos posteriores de Pakes (1986), Rust (1987), Eckstein y Wolpin (1989, 1999), Keane y Wolpin (1997, 1999) y otros, sobre elección dinámica discreta se han centrado más particularmente en los métodos computacionales para modelos paramétricos.⁴⁰ Véase Rust (1996) para una revisión reciente de ese enfoque.

7. EFECTOS DEL TRATAMIENTO

La identificación y la estimación de modelos econométricos estructurales son tareas que retan en cierto modo al econométra, especialmente las de modelos dinámicos con incertidumbre. Investigar este tema tiene un valor incuestionable. Pero la dificultad de la tarea también es indiscutible. Los descubrimientos empíricos que demuestran la importancia de la heterogeneidad en los datos económicos han hecho que la tarea se vuelva aún más difícil.

La econometría de la Cowles se centraba en un modelo lineal en los parámetros para la persona i ,

$$Y_i = \mathbf{X}_i \beta + U_i \quad (20a)$$

donde $E(U_i) = 0$, y analizaba los problemas que surgieron cuando $E(U_i | \mathbf{X}_i) \neq 0$. La heterogeneidad entre individuos se modelizaba como la heterogeneidad en las ordenadas en el origen. La evidencia empírica de toda la investigación en microeconometría de los últimos treinta años defiende una visión más amplia de la heterogeneidad, que afecta tanto a las pendientes como a las ordenadas en el origen:

$$Y_i = \mathbf{X}_i \beta_i + U_i, \quad E(U_i) = 0 \quad (20b)$$

donde las pendientes varían de un individuo a otro. Sea $\tilde{\beta} = E(\beta_i)$ y $v_i = \beta_i - \tilde{\beta}$. Gran parte de la evidencia reciente de diferentes áreas de la microeconometría resalta la correlación entre \mathbf{X}_i y β_i además de que $E(U_i | \mathbf{X}_i) \neq 0$ y $E(v_i | \mathbf{X}_i) \neq 0$ (Carneiro, Heckman y Vytlacil, 2001; Heckman y Vytlacil, 2001c,

(40) La investigación relacionada con este tema de Athey y Haile (2000) y otros autores sobre subastas consideran más a fondo el problema de la identificación no paramétrica.

2001d). Es lo que se llama el “modelo de coeficiente aleatorio correlacionado”. Surge de manera natural en el análisis de los rendimientos económicos de la escolarización. Si la escolarización es una variable X y la Y representa el logaritmo de las ganancias, el componente de β asociado a la escolarización es una tasa de rendimiento que puede variar de un individuo a otro y está plausiblemente correlacionada con el nivel de escolarización. Los métodos tradicionales de variables instrumentales para β fallan cuando los agentes actúan sobre v_i al seleccionar X_i (Heckman y Robb, 1985, 1986; Heckman, 1997; Heckman y Vytlačil, 1998). Los modelos como (20b) no son más que la punta del iceberg empírico.⁴¹ La mayoría de los modelos estructurales tienen una forma general no lineal para las ecuaciones de estimación que, a menudo, sólo puede ser definida recursivamente (véase, por ejemplo, Rust, 1996).⁴² Esto representa una dificultad añadida para la estimación de ecuaciones estructurales.

Habida cuenta de la complejidad que conlleva el estimar modelos estructurales cuando exista heterogeneidad en las pendientes y en las ordenadas en el origen, no es sorprendente que los economistas empíricos hayan buscado métodos más sencillos para responder a preguntas muy concretas en lugar de abordar todo el conjunto de preguntas que pueden contestarse con métodos de ecuaciones estructurales. La literatura sobre efectos del tratamiento investiga un tipo de intervenciones con cobertura parcial, de tal manera que hay un “grupo de tratamiento” y un “grupo de control”. No resulta útil para la evaluación de la intervención que se aplica de forma universal en una economía, a no ser que existan datos sobre diferentes economías que han vivido diferentes intervenciones y que las economías se encuentren aisladas la una de la otra. Esa literatura elude los problemas de equilibrio general al suponer que las respuestas de los que no participan (miembros de los grupos de control) son las mismas que las que experimentarían si no hubiese habido ninguna intervención (Heckman y Smith, 1998).⁴³

La literatura sobre el efecto del tratamiento trata el problema de la evaluación de políticas económicas de la misma manera que los bioestadísticos tratan el problema de evaluar un medicamento. Las respuestas de los agentes que se ven expuestos a una política se comparan con las respuestas de aquéllos que no lo están. La analogía resulta más bien forzada en el contexto de la evaluación de muchas políticas sociales ya que, en

(41) Carneiro *et al.* (2001), Dustmann y Meghir (2001) y Heckman y Vytlačil (2001c, 2001d) presentan estimaciones basadas en modelos de coeficiente aleatorio correlacionados.

(42) Los modelos de coeficiente aleatorio con X_i independiente de β_i fueron introducidos en la econometría por Rubin (1950). Los modelos de coeficiente aleatorio de tipo probit con X_i independiente de β_i en índices lineales aparecen en Thurstone (1927) y fueron introducidos en econometría por Domencich y McFadden (1975). Ichimura y Thompson (1998) ofrecen una reciente contribución aclaratoria de esta literatura.

(43) El libro de Campbell y Stanley (1963) es una referencia clásica de estadística de la educación. Entre las tempranas contribuciones de economistas se encuentran los artículos de Goldberger (1972), Cain (1975) y Barnow, Cain y Goldberger (1980).

una economía moderna, las respuestas de los agentes están relacionados por los mercados y otras formas de interacción social. Es lo que da lugar a la diferenciación entre los que se ven “directamente” afectados y los que se ven afectados “de forma indirecta”. Así, aquéllos que siguen estudios universitarios con la ayuda de un programa de becas de estudio se ven directamente afectados. El resto de la sociedad se ve afectada indirectamente por el coste de los impuestos destinados a financiar las becas y por los efectos de una expansión en el capital educativo sobre los precios del trabajo de una mano de obra formada y del de una mano de obra sin formación.⁴⁴

En las siguientes líneas se exponen, de forma simple, las diferencias esenciales entre la literatura sobre el efecto del tratamiento y la literatura sobre ecuaciones estructurales. Consideremos los modelos de respuestas potenciales que se dan en (7). Se trata de funciones no lineales, y las U_i pueden ser vectores de variables no observables relacionadas directamente con la respuesta. Las U_i pueden ser estocásticamente dependientes de las X_i . El modelo de coeficiente aleatorio de (20a) es un caso especial de (7) cuando interpretamos las U_i como vectores. Consideremos un modelo con dos respuestas potenciales (Y_0, Y_1). Escribamos las respuestas como funciones causales marshalianas

$$Y_0 = g_0(\mathbf{X}, U_0) \quad (21a)$$

y

$$Y_1 = g_1(\mathbf{X}, U_1) \quad (21b)$$

Más concretamente, pensemos en (Y_0, Y_1) como las ganancias potenciales de una persona que haya realizado estudios de nivel de secundaria o estudios universitarios, respectivamente.

El enfoque estructural trata de determinar g_0 y g_1 , recurriendo generalmente a la separabilidad aditiva; por ejemplo, en el caso de que (U_0, U_1) sean escalares,

$$Y_0 = g_0(\mathbf{X}) + U_0 \quad (22a)$$

y

$$Y_1 = g_1(\mathbf{X}) + U_1 \quad (22b)$$

La determinación de estas funciones y el mecanismo económico que selecciona cuál de los componentes de (Y_0, Y_1) se observa permite al analista responder a la gama completa de cuestiones contrafactuales de polí-

(44) Véase Heckman y Smith (1998). Heckman, Lochner y Taber (1998a, 1998b, 1998c, 1999) demuestran lo importantes que pueden llegar a ser estos efectos “indirectos”. Lewis (1963) aborda estos efectos de equilibrio general en el contexto de la evaluación del impacto del sindicalismo sobre la economía e introduce la noción de efectos directos e indirectos.

tica económica que se consideran en la sección 2 bajo las restricciones de las condiciones de soporte discutidas en dicha sección.

La literatura sobre efecto del tratamiento se centra en un tipo de cuestiones más restringido y responde a estas cuestiones bajo condiciones más débiles que las que son necesarias para identificar y estimar ecuaciones estructurales. En el contexto de las ecuaciones (21a) y (21b) es normal considerar la intervención como el paso de una persona del estado 0 al estado 1: $D = 1$ si una persona se encuentra en 1 (por ejemplo, una persona con estudios universitarios) y $D = 0$ si no (por ejemplo, una persona con estudios de nivel de secundaria).

Hay dos efectos del tratamiento que reciben especial atención en la literatura actual. Se trata por una parte del efecto del tratamiento medio (ATE, del inglés "*average treatment effect*") para $\Delta = Y_1 - Y_0$, que es el efecto de escoger un individuo de forma aleatoria para recibir el tratamiento

$$E(\Delta|\mathbf{X}=\mathbf{x})=E(Y_1-Y_0|\mathbf{X}=\mathbf{x})=ATE(\mathbf{x})$$

y, por otra parte, del efecto del tratamiento sobre los tratados, TT (del inglés "*treatment on the treated*"), que se refiere al efecto del tratamiento sobre los que están realmente tratados

$$E(\Delta|\mathbf{X}=\mathbf{x}, D=1)=E(Y_1-Y_0|\mathbf{X}=\mathbf{x}, D=1)=TT(\mathbf{x})$$

En términos del modelo estructural de las ecuaciones (21a) y (21b),

$$ATE(\mathbf{x})=E(Y_1-Y_0|\mathbf{X}=\mathbf{x})=E[g_1(\mathbf{X}, U_1)-g_0(\mathbf{X}, U_0)|\mathbf{X}=\mathbf{x}]$$

y

$$TT(\mathbf{x})=E(Y_1-Y_0|\mathbf{X}=\mathbf{x}, D=1)=E[g_1(\mathbf{X}, U_1)-g_0(\mathbf{X}, U_0)|\mathbf{X}=\mathbf{x}, D=1]$$

En términos de una representación aditivamente separable que nos resulta más familiar,

$$ATE(\mathbf{x})=g_1(\mathbf{x})-g_0(\mathbf{x})+E[U_1-U_0|\mathbf{X}=\mathbf{x}]$$

y

$$TT(\mathbf{x})=g_1(\mathbf{x})-g_0(\mathbf{x})+E[U_1-U_0|\mathbf{X}=\mathbf{x}, D=1]$$

Para la definición de estos parámetros no es necesario que las \mathbf{X} sean exógenas ($E(U_1 - U_0|\mathbf{X} = \mathbf{x}) \neq 0$ o $E(U_1 - U_0|\mathbf{X} = \mathbf{x}, D = 1) \neq 0$) ni que las funciones estructurales g_1 y g_0 estén identificadas. Evidentemente, cuando se conocen las funciones estructurales y la dependencia entre (U_0, U_1) y \mathbf{X} , se pueden identificar los parámetros de tratamiento y el enfoque estructural es más general.

Si escribimos los parámetros del tratamiento en el marco de un modelo de coeficiente aleatorio correlacionado como (20b), podremos sacar conclusiones interesantes. Definamos la Y observada como

$$Y=(1-D)Y_0+DY_1 \tag{23}$$

(Y puede ser discreta, continua o mezcla de discreta y continua). Definamos los residuos de las esperanzas condicionales de (Y_0, Y_1) como

$$\begin{aligned}\epsilon_0 &= Y_0 - E(Y_0 | \mathbf{X} = \mathbf{x}) = Y_0 - E(g_0(\mathbf{X}, U_0) | \mathbf{X} = \mathbf{x}) \\ \epsilon_1 &= Y_1 - E(Y_1 | \mathbf{X} = \mathbf{x}) = Y_1 - E(g_1(\mathbf{X}, U_0) | \mathbf{X} = \mathbf{x})\end{aligned}$$

En el caso aditivamente separable,

$$\begin{aligned}\epsilon_0 &= Y_0 - g_0(\mathbf{x}) - E(U_0 | \mathbf{X} = \mathbf{x}) \\ \epsilon_1 &= Y_1 - g_1(\mathbf{x}) - E(U_1 | \mathbf{X} = \mathbf{x})\end{aligned}$$

Incluso en este caso, sólo si los errores (U_0, U_1) son de media cero independiente de \mathbf{X} , coincidirá (U_0, U_1) con (ϵ_0, ϵ_1). En la literatura sobre el efecto del tratamiento, no es necesaria la independencia de la media para (U_0, U_1). Para simplificar la notación, definamos $E(Y_1 | \mathbf{X} = \mathbf{x}) = \mu_1(\mathbf{x})$ y $E(Y_0 | \mathbf{X} = \mathbf{x}) = \mu_0(\mathbf{x})$. Nótese que $\mu_1(\mathbf{x})$ y $\mu_0(\mathbf{x})$ no son funciones causales marshallianas excepto cuando $E(U_0 | \mathbf{X}) = 0$ y $E(U_1 | \mathbf{X}) = 0$.

Con esta notación,

$$\begin{aligned}\text{ATE}(\mathbf{x}) &= \mu_1(\mathbf{x}) - \mu_0(\mathbf{x}) \\ \text{TT}(\mathbf{x}) &= \mu_1(\mathbf{x}) - \mu_0(\mathbf{x}) + E[\epsilon_1 - \epsilon_0 | \mathbf{X} = \mathbf{x}, D = 1]\end{aligned}$$

Podemos escribir (23) como

$$Y = \mu_0(\mathbf{X}) + D[\mu_1(\mathbf{X}) - \mu_0(\mathbf{X}) + \epsilon_1 - \epsilon_0] + \epsilon_0 \quad (24)$$

Se trata de un modelo no paramétrico de coeficiente aleatorio de la forma (20b), donde el coeficiente aleatorio corresponde a D y D puede estar correlacionado con ϵ_0 y $\epsilon_1 - \epsilon_0$. Nótese que el coeficiente de D es $\Delta = Y_1 - Y_0 = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x}) + \epsilon_1 - \epsilon_0$, el efecto del tratamiento a nivel individual. Tiene como media $\text{ATE}(\mathbf{x}) = E(Y_1 - Y_0 | \mathbf{X} = \mathbf{x}) = \bar{\Delta}(\mathbf{x})$. Para simplificar la notación, en lo sucesivo mantengo implícita la dependencia respecto de \mathbf{X} .

Podemos escribir (24) de dos maneras diferentes. En primer lugar, en términos de ATE, obtenemos

$$Y = \mu_0 + (\mu_1 - \mu_0)D + [\epsilon_0 + D(\epsilon_1 - \epsilon_0)] = \mu_0 + (\text{ATE})D + [\epsilon_0 + D(\epsilon_1 - \epsilon_0)]$$

En términos de TT obtenemos

$$\begin{aligned}Y &= \mu_0 + [\bar{\Delta} + E(\epsilon_1 - \epsilon_0 | D = 1)]D + \{\epsilon_0 + [(\epsilon_1 - \epsilon_0) - E(\epsilon_1 - \epsilon_0 | D = 1)]D\} \\ &= \mu_0 + (\text{TT})D + \{\epsilon_0 + [(\epsilon_1 - \epsilon_0) - E(\epsilon_1 - \epsilon_0 | D = 1)]D\}\end{aligned}$$

Ahora el problema econométrico de identificar estos efectos del tratamiento se centra en el problema de que D puede estar correlacionado con los términos de error que se encuentran entre corchetes.

Obsérvese que pueden definirse distintas medias sobre distintos conjuntos de variables condicionantes. Así pues, existen diferentes respues-

tas medias que responden a su vez a preguntas distintas. No existe un único agente medio o "representativo" que sirva para contestar a todas las preguntas relativas a la evaluación de medidas de política económica, excepto en casos especiales que se indican más adelante. En presencia de selección sobre variables no observables idiosincráticas, no hay un único "efecto" que describa un programa o una intervención. Es posible definir varios efectos del tratamiento que dependen de los conjuntos de variables condicionantes utilizadas para definir "el" efecto. La elección aleatoria de personas para incluirlas en el programa y la comparación de sus respuestas medias con las de otras personas seleccionadas al azar y excluidas del programa produce el ATE. La elección aleatoria de personas para incluirlas en el programa y la comparación de sus respuestas medias con las del mismo tipo de personas a las que no se les permite acceder al programa definen el parámetro TT. Cuando la conformidad es total, éste es el parámetro implícito de los recientes experimentos sociales que niegan el acceso a solicitantes que de otra manera podrían ser aceptados.

Resulta útil distinguir tres casos de la ecuación (24) y relacionarlos con modelos econométricos más convencionales. El primer caso surge cuando las respuestas al tratamiento son las mismas para todos (dado $\mathbf{X} = \mathbf{x}$).

Caso 1. $\epsilon_1 = \epsilon_0$, y así Δ es una constante.

En ese caso, $E(\Delta|D = 1) = E(\Delta) = \Delta$. Un solo agente representativo resume los resultados medios del programa. Hay un solo "efecto" medio. El problema del sesgo de selección viene a ser el problema de la Cowles de que D podría estar correlacionado con ϵ_0 .⁴⁵

El segundo caso surge cuando las respuestas al tratamiento varían de unas personas a otras, pero las decisiones de adoptar el tratamiento no se basan en esas respuestas variables.

Caso 2. $\epsilon_1 \neq \epsilon_0$, y así Δ varía de unas personas a otras; pero la media de $\epsilon_1 - \epsilon_0$ es independiente de D de tal manera que $E(\Delta|D = 1) = E(\Delta)$.

En este caso, los rendimientos de participar en la actividad que se evalúa varían ex post pero son los mismos, en promedio, para todas aquellas personas que tengan los mismos valores de \mathbf{X} . Como en el caso de un coeficiente común, este caso se presta a una descripción de la intervención que se está evaluando a través de un agente representativo. Una vez más, el problema econométrico viene a ser el problema de que D puede estar correlacionado con ϵ_0 .

El tercer caso surge cuando las respuestas al tratamiento varían de unas personas a otras, y las decisiones de adoptar el tratamiento se basan en esa variación.

(45) Nótese que $U_1 = U_0$ no implica que $\epsilon_1 = \epsilon_0$.

Caso 3. $\epsilon_1 \neq \epsilon_0$, y así Δ varía de unas personas a otras; pero la media de $\epsilon_1 - \epsilon_0$ no es independiente de D , de tal manera que $E(\Delta|D = 1) \neq E(\Delta)$.

El participar o no en el tratamiento depende, por lo menos en parte, de ganancias no observadas. Por lo tanto, el problema econométrico conlleva no solamente tener en cuenta la correlación entre ϵ_0 y D , sino también entre Δ y D y entre (Δ, D) y ϵ_0 .

La heterogeneidad en las respuestas al tratamiento con la que los agentes reaccionan viola el supuesto de parámetro común de la econometría de la Cowles. No hay un único efecto del tratamiento sino más bien una variedad de efectos en función de las variables condicionantes.⁴⁶ Esto se aleja radicalmente del supuesto de parámetros estructurales constantes, basados en g_0 y g_1 , y que no varían con las políticas, lo cual constituye el sello característico de la econometría de la Cowles. La unidad y simplicidad con que la literatura estructural produce parámetros que pueden ser transferidos y comparados en distintos entornos económicos parece haberse perdido en la literatura sobre efectos del tratamiento.

Para extraer de los datos estos parámetros de evaluación, son necesarios menos supuestos que los que requiere la recuperación de parámetros estructurales. Sin embargo supone problemas no triviales que nos remiten al problema de selección para estimar los parámetros estructurales, aunque toma una forma distinta. Si nos centramos en las medias, podemos estimar consistentemente a partir de datos observados $E(Y|X = x, D = 1) = E(Y_1|X = x, D = 1)$ (por ejemplo, los ingresos de personas con un nivel de formación de estudios universitarios) y $E(Y|X = x, D = 0) = E(Y_0|X = x, D = 0)$ (por ejemplo, los ingresos de personas con una formación de estudios de secundaria).

Resulta problemático emplear los ingresos de personas con un nivel de estudios de secundaria como aproximación de los ingresos que habrían tenido las personas que han realizado estudios universitarios si se hubieran detenido en el nivel de estudios de secundaria. Si comparamos los ingresos medios de ambos grupos obtenemos

$$\begin{aligned}
 E(Y_1|X=x, D=1) - E(Y_0|X=x, D=0) &= \underbrace{E(Y_1 - Y_0|X=x, D=1)}_{TT(X)} \\
 &+ \underbrace{E(Y_0|X=x, D=1) - E(Y_0|X=x, D=0)}_{\text{Sesgo de selección}}
 \end{aligned}$$

El segundo término es la diferencia entre lo que ganarían los que han realizado estudios universitarios si se hubieran detenido en secundaria y

(46) Este aspecto lo reconocieron de forma temprana Heckman y Robb (1985, 1986) y Björklund y Moffitt (1987).

los ingresos que tendrían los que tienen un nivel de estudios de secundaria. Surgiría otra definición del sesgo de selección si escogiésemos ATE como nuestro parámetro.

Si no hubiese no observables, o si diera la casualidad de que el condicionar en función de X eliminase las diferencias de medias en los no observables, como lo suponen los estadísticos que abogan por el método del matching, entonces el término correspondiente al sesgo de selección desaparecería. Sin embargo, la mala calidad del ajuste de las ecuaciones sobre datos microeconómicos sugiere que el supuesto de inexistencia de no observables no es aceptable. El confiar en el matching es un acto de fe.⁴⁷ Se han propuesto diversos enfoques diferentes para eliminar estos sesgos de selección. En el trabajo que realicé en colaboración con Robb (Heckman y Robb, 1985, 1986) desarrollamos distintos estimadores para distintos modelos económicos, distintas estrategias de muestreo, y distintos parámetros de tratamiento. Consideramos distintos supuestos de identificación y empleamos la economía y la estadística para justificar las elecciones de los estimadores.

Si diferentes entornos se caracterizan por diferentes dependencias de (X, U_0, U_1) , los efectos del tratamiento estimados en un entorno no son transferibles a otros entornos.⁴⁸ Las variaciones de X no son cambios *ceteris paribus* y no responden a cuestiones estructurales. La literatura sobre efectos del tratamiento elude estos problemas centrándose en la estimación de un efecto o de un conjunto limitado de efectos que se aplican a un entorno determinado y no pueden ser aplicados a otros entornos.⁴⁹ Este enfoque evita la mayoría de los problemas de estimación estructural con el inconveniente, sin embargo, de producir estimaciones que no responden a preguntas estructurales. No solamente no son en general comparables las estimaciones de estos parámetros entre estudios basados en muestras diferentes (lo que implica que es difícil acumular conocimiento con los estudios), sino que además es difícil relacionar los diferentes efectos del tratamiento estimados sobre una misma muestra.

(47) El modelo clásico de sesgo de habilidad en las ecuaciones de ingresos que relacionan la escolarización y los ingresos supone que el sesgo de selección en el rendimiento de la escolarización es positivo incluso si condicionamos a X y de esta manera es inconsistente con el matching. El modelo de Roy es inconsistente con el matching (véase Heckman y Vytlačil, 2002).

(48) Ésta es una versión en términos del efecto del tratamiento de la cuestión planteada originalmente por Haavelmo (1943).

(49) Heckman (1992) hace esta afirmación en el contexto de la estimación de parámetros determinados por asignaciones aleatorias al tratamiento. Véase Heckman y Vytlačil (2001b, 2001e) para una discusión más extensa. Marschak (1953) y Hurwicz (1962) abordan problemas de decisiones económicas que no plantean un conocimiento completo de parámetros estructurales.

7.1. El empleo del efecto marginal del tratamiento para unificar la literatura sobre efectos del tratamiento

En trabajos conjuntos recientes (Heckman y Vytlacil, 1999, 2000a, 2000b, 2001a, 2001b, 2001c, 2001d, 2002), se unifica la literatura sobre efectos del tratamiento empleando un parámetro de tratamiento económicamente interpretable: el efecto marginal del tratamiento (MTE, del inglés "*marginal treatment effect*"). Relacionamos la literatura sobre efecto del tratamiento con la literatura más convencional sobre ecuaciones estructurales para así aprovechar algunas de las ventajas que aporta un enfoque estructural.

El MTE es el efecto medio del programa para aquéllos que se encuentran en el límite de participar en él, para valores dados de observables y condicionando en función de las no observables en la ecuación de participación del programa.⁵⁰

Presentamos condiciones bajo las cuales es posible representar todos los parámetros convencionales de tratamiento como medias ponderadas del MTE, donde parámetros diferentes corresponden a ponderaciones distintas (véase Heckman y Vytlacil, 1999, 2000a, 2001c, 2001d, 2002). Bajo las mismas condiciones, organizamos la literatura sobre evaluación econométrica clasificando los estimadores bajo el criterio de si adoptan el supuesto de que el MTE depende de las no observables en la ecuación que determina la participación en el programa o no. Las variables instrumentales lineales ordinarias están caracterizadas como medias ponderadas del MTE, donde las ponderaciones difieren generalmente de las que se emplean para definir los parámetros de tratamiento convencionales (Heckman y Vytlacil, 2000a, 2000b, 2001b, 2001d, 2001e). Más específicamente, nuestro trabajo empieza con (21a) y (21b) y relaciona la literatura sobre elección discreta postulando la existencia de una variable latente $D^* = \mu_D(\mathbf{Z}) - U_D$ de tal manera que $D = 1$ si $D^* \geq 0$; $D = 0$ si no. Así, el mecanismo de elección en el modelo de Roy y las generalizaciones de este modelo se mantienen, aunque en un planteamiento semiparamétrico. Suponemos que (a) $\mu_D(\mathbf{Z})$ es una variable aleatoria no degenerada condicionada a \mathbf{X} ; (b) U_D es absolutamente continuo con respecto a la medida de Lebesgue; (c) (U_1, U_D) y (U_0, U_D) son independientes de \mathbf{Z} condicionando a \mathbf{X} ; (d) Y_1 y Y_0 tienen primeros momentos finitos y, (e) $1 > \Pr(D = 1 | \mathbf{X} = \mathbf{x}) > 0$ para cada $\mathbf{x} \in \text{Supp}(\mathbf{X})$. Los supuestos a y c son supuestos de "variable instrumental" de que hay al menos un instrumento que determina la participación en el programa pero no las respuestas. El supuesto b es un supuesto técnico principalmente de conveniencia. El supuesto d garantiza que los parámetros de interés estarán bien definidos. El supuesto e es el supuesto de que existe en la población tanto un grupo de tratamiento

(50) Este parámetro fue introducido en la literatura de evaluación por Björklund y Moffitt (1987). Se trata de la forma límite del parámetro del efecto del tratamiento medio local (LATE, del inglés "*local average treatment effect*"). El parámetro LATE fue introducido en Imbens y Angrist (1994), y la forma límite del LATE la incorporaron Heckman (1997) y Angrist, Graddy y Imbens (2000) (véase también Heckman y Smith, 1998).

como un grupo de control para cada \mathbf{X} . Estas condiciones imponen restricciones que son contrastables con los datos (Heckman y Vytlacil, 2000a). El vector \mathbf{X} no tiene por qué ser exógeno siempre y cuando estemos evaluando programas que ya están en práctica y no proyectando un programa a nuevas poblaciones.⁵¹ Sin pérdida de generalidad incluimos los elementos de \mathbf{X} en \mathbf{Z} . Definimos $P(\mathbf{z})$ como la probabilidad de recibir el tratamiento condicionado a $\mathbf{Z} = \mathbf{z}$: $P(\mathbf{z}) \equiv \Pr(D = 1 | \mathbf{Z} = \mathbf{z}) = F_{U_D | \mathbf{x}}(\mu_D(\mathbf{z}) | \mathbf{x})$, dónde $F_{U_D | \mathbf{x}}(\cdot | \mathbf{x})$ representa la distribución de U_D condicionada a $\mathbf{X} = \mathbf{x}$. Sin pérdida de generalidad, imponemos la normalización de que $U_D \sim \text{Unif}[0, 1]$ de tal manera que $\mu_D(\mathbf{z}) = P(\mathbf{z})$. Vytlacil (2002) demuestra bajo las condiciones *a-e* que el modelo de selección es equivalente al modelo LATE de Imbens y Angrist (1994).

El efecto medio del tratamiento sobre los que se encuentran en el límite de participación del programa al nivel $U_D = u_D$ es el MTE: $\Delta^{\text{MTE}}(\mathbf{x}, u_D) \equiv E(\Delta | \mathbf{X} = \mathbf{x}, U_D = u_D)$. Representa la base de unificación de los estimadores y de los parámetros de tratamiento. Desarrollamos el método de variables instrumentales locales (LIV, del inglés "*local instrumental variables*") para estimar ese parámetro. Puede estimarse con la derivada de $E(\mathbf{Y} | \mathbf{X} = \mathbf{x}, P = p)$ con respecto a p como recalco más adelante.

Establecemos que bajo condiciones *a-e* todos los parámetros de tratamiento de la población empleados en la literatura sobre evaluación de políticas son versiones ponderadas de la MTE. Así, para el parámetro de tratamiento j ,

$$\text{parámetro}_j(\mathbf{x}) = \int_0^1 \Delta^{\text{MTE}}(\mathbf{x}, u_D) \omega_j(\mathbf{x}, u_D) du_D \tag{25}$$

Más específicamente, tenemos

$$\text{ATE}(\mathbf{x}) = \int_0^1 \text{MTE}(\mathbf{x}, u_D) du_D$$

y

$$\text{TT}(\mathbf{x}) = \int_0^1 \text{MTE}(\mathbf{x}, u_D) g_x(u_D) du_D$$

La función $g_x(u)$ se define en Heckman y Vytlacil (1999, 2000a).

El parámetro LATE de Imbens y Angrist (1994) es

$$\text{LATE}(\mathbf{x}, P(\mathbf{z}), P(\mathbf{z}')) = \left[\int_{P(\mathbf{z}')}^{P(\mathbf{z})} \text{MTE}(\mathbf{X}=\mathbf{x}, U_D=u_D) du_D \right] \times \frac{1}{P(\mathbf{z}) - P(\mathbf{z}')}$$

El límite en probabilidad de los estimadores mediante variables instrumentales de los efectos del tratamiento (así como el de otros estima-

(51) Mientras que no hay exigencias de exogeneidad referentes a \mathbf{X} , una condición contrafactual de "no retroalimentación" que corresponde a los supuestos clásicos de no causalidad en econometría estructural es necesaria para que el condicionar en función de \mathbf{X} no esconda los efectos de D . Denotemos por X_d el valor de \mathbf{X} cuando D toma valor d . Entonces, una condición suficiente que garantiza la no-retroalimentación de D hacia \mathbf{X} es $X_1 = X_0$ casi siempre. Heckman y Vytlacil (2001d, 2002) analizan el papel de los supuestos de exogeneidad para proyectar efectos estimados del tratamiento en nuevos entornos.

dores) puede también expresarse como medias ponderadas del MTE.⁵² Así, condicionado a $\mathbf{X} = \mathbf{x}$, el límite en probabilidad del estimador k es

$$\text{plim estimador}_k(\mathbf{x}) = \int_0^1 \Delta^{\text{MTE}}(\mathbf{x}, u_D) \omega_k(\mathbf{x}, u_D) du_D \quad (26)$$

Se puede demostrar que las ponderaciones que corresponden a los parámetros de tratamiento convencionales integran sobre la unidad, como también lo hacen las ponderaciones que corresponden a gran parte de los estimadores, incluyendo el de variables instrumentales que emplea $P(\mathbf{Z})$ como instrumento. Por lo general, los límites en probabilidad de los distintos estimadores ponderan MTE de una forma diferente a como lo hacen los parámetros.⁵³ Nótese que los parámetros y los estimadores que tienen la forma que se propone más arriba coinciden si las respuestas no varían de unos individuos a otros (dado $\mathbf{X} = \mathbf{x}$) o, cuando varían, si los agentes no participan en el programa en función de esta variación. Por lo general y en la mayoría de los estudios de la literatura, los estimadores y los parámetros difieren.

7.2. Estimar el MTE y comprender las variables instrumentales

Consideremos el caso de aplicar el método de variables instrumentales para estimar los parámetros de efectos del tratamiento estándares. Los economistas han venido utilizando variables instrumentales (IV) durante más de 70 años.⁵⁴ La intuición que se esconde tras el método IV se entiende fácilmente. Mimetiza la variación experimental utilizando la variación IV. En el modelo del caso 1 de coeficiente de regresión correlacionado, el método IV es una solución al problema de que D está correlacionado con el término de error ϵ_0 . Los métodos IV estándares fallan en el planteamiento más general considerado en el caso 3 (Heckman, 1997).

Consideremos el empleo de $P(\mathbf{Z})$ como instrumento para D . Supongamos que queremos estimar el parámetro TT en el caso 3. Suprimamos la dependencia explícita con respecto a \mathbf{X} , para simplificar la notación. Bajo los supuestos de Heckman y Vytacil (2000a), deducimos de (24) que el parámetro TT puede escribirse como

$$\begin{aligned} \text{TT} &= E(\Delta | \mathbf{Z} = \mathbf{z}, D = 1) = E(\Delta | P(\mathbf{z}) \geq U_D) = \mu_1 - \mu_0 + E(\epsilon_{1-\epsilon_0} | P(\mathbf{z}) \geq U_D) \\ &= \bar{\Delta} + E(\epsilon_{1-\epsilon_0} | P(\mathbf{z}) \geq U_D) \end{aligned} \quad (27)$$

(52) La expresión (26) se asemeja a expresiones que aparecen en Yitzhaki (1996) y Angrist *et al.* (2000).

(53) Bajo las condiciones *a-e*, la diferencia máxima posible entre dos parámetros de políticas representables bajo la forma de (25), o entre estimadores y parámetros puede escribirse como el producto de la diferencia entre los valores más grandes y más pequeños posibles del MTE y una medida de la distancia entre las dos ponderaciones.

(54) Véase la revisión histórica en Morgan (1990). Wright (1928) fue el primero en utilizarla.

En los casos 1 y 2, el término final en las dos últimas expresiones es cero. Si empleamos (24), podemos escribir

$$E(Y|Z=z) = \mu_0 + [\bar{\Delta} + E(\epsilon_{1-\epsilon_0} | P(z) \geq U_D)] P(z) \tag{28}$$

Si $E(\epsilon_{1-\epsilon_0} | P(z) \geq U_D) = 0$, como en el caso 1 o el caso 2, entonces podemos aplicar la lógica de las variables instrumentales para expresar dos valores de z, z' de tal manera que $P(z) \neq P(z')$,

$$E(Y|Z=z) = \mu_0 + [\bar{\Delta} + E(\epsilon_{1-\epsilon_0} | P(z) \geq U_D)] P(z)$$

$$E(Y|Z=z') = \mu_0 + [\bar{\Delta} + E(\epsilon_{1-\epsilon_0} | P(z') \geq U_D)] P(z')$$

restarle la ecuación inferior a la superior y dividir por $P(z) - P(z')$ para obtener

$$\begin{aligned} & \frac{E(Y|Z=z) - E(Y|Z=z')}{P(z) - P(z')} = \\ & = \bar{\Delta} + \frac{E(\epsilon_{1-\epsilon_0} | P(z) \geq U_D) P(z) - E(\epsilon_{1-\epsilon_0} | P(z') \geq U_D) P(z')}{P(z) - P(z')} \end{aligned} \tag{29}$$

Si $E(\epsilon_{1-\epsilon_0} | P(z) \geq U_D) = 0$, el método IV identifica “el” efecto del tratamiento $\bar{\Delta}$, y todos los parámetros de tratamiento medio son los mismos. De forma general, el método identifica únicamente la combinación de parámetros en (29).

En los casos 1 y 2, $P(Z)$ entra en (28) de manera lineal como la variable que multiplica el término que se encuentra entre corchetes. En el caso 3, $P(Z)$ entra en la expresión en dos sitios. Así, el contraste de si el caso 1 o el caso 2 es válido viene a ser el contraste de la linealidad de (28) en términos de $P(Z)$. Esto también es un contraste de la validez del método IV convencional para identificar $\bar{\Delta}$. Véase el gráfico 7 que muestra dos casos: (a) un caso en el que $E(Y|P=p)$ es lineal en p , de tal manera que los casos 1 y 2 son aplicables y (b) el caso general 3 (que corresponde a la línea discontinua), donde $E(Y|P=p)$ es una función no lineal de P .

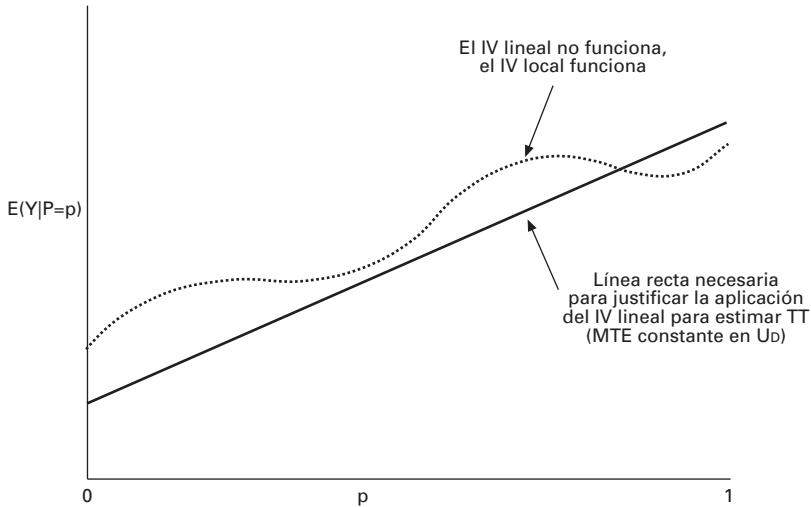
Incluso en el caso de que el contraste diera resultados negativos, de tal manera que el método IV lineal no identificase $\bar{\Delta}$, podemos extraer información importante de (28). Para mostrarlo, empleemos la propiedad de separabilidad para escribir (28) como:

$$E(Y|P(Z)=P(z)) = \mu_0 + \bar{\Delta} P(z) + \int_{-\infty}^{\infty} \int_0^{P(z)} (\epsilon_{1-\epsilon_0}) f(\epsilon_{1-\epsilon_0} | U_D = u_D) du_D d(\epsilon_{1-\epsilon_0})$$

Diferenciemos con respecto de $P(Z)$ para obtener

$$\begin{aligned} & \frac{\partial E(Y|P(Z)=P(z))}{\partial P(z)} = \\ & \bar{\Delta} + \int_{-\infty}^{\infty} (\epsilon_{1-\epsilon_0}) f(\epsilon_{1-\epsilon_0} | U_D = P(z)) d(\epsilon_{1-\epsilon_0}) = \text{MTE} \end{aligned}$$

Gráfico 7
CUÁNDO FALLA EL MÉTODO IV CONVENCIONAL
Y CUÁNDO FUNCIONA



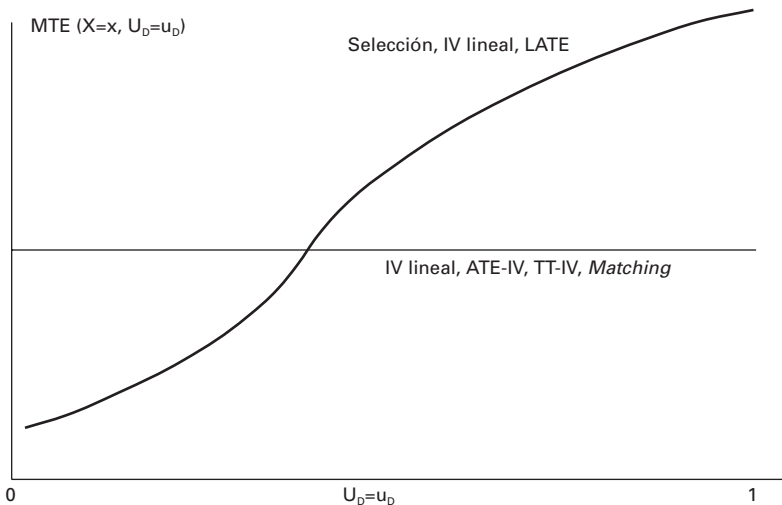
Al conocer el MTE, podemos recuperar todos los parámetros de tratamiento integrando el MTE así identificado mediante las relaciones que se aportan en la subsección 7.1. Esto es el método de variables instrumentales locales (LIV, del inglés "*local instrumental variables*") introducido en esta literatura por Heckman y Vytlacil (1999, 2000a, 2001d).

Cuando el soporte de $P(\mathbf{Z})$ es el intervalo unitario completo, podemos estimar los parámetros de tratamiento estimando MTE y empleando la ecuación (25) para recuperar el parámetro de interés. También desarrollamos métodos más generales de estimación de parámetros que no requieren que el soporte de $P(\mathbf{Z})$ sea el intervalo unitario completo (por ejemplo, permitimos que \mathbf{Z} sea discreta) y que no requieren tampoco la estimación de una derivada de una esperanza condicional. Podemos reemplazar los MTE por los LATE y las integrales por sumas. No obstante, estos métodos siguen requiriendo que se cumplan condiciones para el soporte de $P(\mathbf{Z})$, que dependen del parámetro que nos interesa. Cuando estas condiciones no se dan, desarrollamos cotas estrictas sobre los parámetros de tratamiento que explotan toda la información contenida en el modelo y en los datos de los que disponemos (Heckman y Vytlacil, 2000a, 2001b).

Con la función MTE, podemos clasificar todos los estimadores econométricos de la literatura sobre evaluación de políticas según que permitan o no evaluar la participación en el programa sobre la base de beneficios no observables (véase el gráfico 8). Los estimadores de *matching* y los estimadores IV convencionales suponen que no hay selección en

función de los beneficios.⁵⁵ Por lo tanto, el MTE es plano para este tipo de estimadores. Los modelos de selección, el LATE y el LIV toman en consideración la selección por no observables por lo que son consistentes con el modelo de Roy. En este caso, el MTE es una función no trivial de u_d . La evidencia extraída de la literatura microeconómica sobre el MTE y que se reproduce en el cuadro 3, sugiere que el MTE varía de un estudio de fenómenos económicos a otro, por lo que los métodos de *matching* y IV convencionales no identifican ningún parámetro de tratamiento de interés.⁵⁶ Cuando el soporte de $P(\mathbf{Z})$ es solamente parcial, se pueden construir cotas estrechas, simples y fácilmente aplicables para todos los parámetros de tratamiento (véase Heckman y Vytlacil, 2000a, 2001b).

Gráfico 8
CONDICIONES SOBRE EL MTE PARA JUSTIFICAR ESTRATEGIAS DE ESTIMACIÓN ALTERNATIVAS



(55) Con el término de "matching", me refiero a los estimadores que explotan todas las condiciones de Rosenbaum-Rubin (1983): (Y_0, Y_1) independiente de $D|\mathbf{X}$. Como se apunta en Heckman, Ichimura y Todd (1997) y Heckman, Ichimura, Smith y Todd (1998), para estimar TT, nos bastan las condiciones más débiles de Y_0 independiente de $D|\mathbf{X}$ o $E(Y_0|\mathbf{X}, D=1) = E(Y_0|\mathbf{X})$, lo cual permite ganancias de selección $(Y_1 - Y_0)$.

(56) En este cuadro $Y_2 = \mu_2(\mathbf{Z}) + U_2$. La mayoría de los estimadores del cuadro derivan de modelos estructurales basados en supuestos de normalidad descritos por los supuestos 2 y 3 de la sección V empleando los modelos de índice definidos en la sección IV, especialmente en la ecuación (8)-(11), y de esta manera están sujetos a la crítica de que se producen con métodos que dependen de formas funcionales. Véase Heckman, Tobias y Vytlacil (2000) para la definición del parámetro MTE para el modelo normal clásico. Muestran que si $\sigma_{12} \neq \sigma_{02}$, MTE es una función no constante de u_D . Carneiro *et al.* (2001) emplean métodos semiparamétricos para demostrar la no constancia del MTE.

Cuadro 3
EVIDENCIA DE SELECCIÓN EN NO OBSERVABLES

Estudio	Método	Resultado
Sindicalismo		
Lee (1978)	Modelo de selección normal ($H_0: \sigma_{12} = \sigma_{02}$)	$\sigma_{12} \neq \sigma_{02}$ No rechaza
Farber (1983)	Modelo de selección normal ($H_0: \sigma_{12} = \sigma_{02}$)	$\sigma_{12} = \sigma_{02}$ No rechaza
Duncan y Leigh (1985)	Modelo de selección normal ($H_0: \sigma_{12} = \sigma_{02}$)	$\sigma_{12} = \sigma_{02}$ No rechaza
Robinson (1989)	Modelo de selección normal $(\mu_1 - \mu_0)_{IV} = (\mu_1 - \mu_0)_{normal}$	$\sigma_{12} \neq \sigma_{02}$ No rechaza
Nivel de estudios (universitarios vs. de secundaria)		
Willis y Rosen (1979)	Modelo de selección normal ($H_0: \sigma_{12} = \sigma_{02}$)	$\sigma_{12} \neq \sigma_{02}$ Rechaza
Heckman, Tobias y Vytlačil (2000)	Modelo de selección normal ($H_0: \sigma_{12} = \sigma_{02}$)	$\sigma_{12} \neq \sigma_{02}$ Rechaza
Formación para el puesto de trabajo		
Björklund y Moffitt (1987)	Modelo de selección normal ($H_0: \sigma_{12} = \sigma_{02}$)	$\sigma_{12} \neq \sigma_{02}$ Rechaza
Heckman, Ichimura, Smith y Todd (1998, suppl.)	$E(U_1 - U_0 D=1, \mathbf{Z}, \mathbf{X}) = E(U_1 - U_0 D=1, \mathbf{X})$	Rechazar la selección en los no observables
Elección sectorial		
Heckman y Sedlacek (1990)	Modelo de selección normal ($H_0: \sigma_{12} = \sigma_{02}$)	$\sigma_{12} \neq \sigma_{02}$ Rechaza
Migración		
Pessino (1991)	Modelo de selección normal ($H_0: \sigma_{12} = \sigma_{02}$)	$\sigma_{12} \neq \sigma_{02}$ Rechaza
Tunali (2000)	$H_0: E(U_1 - U_0 D=1) = 0$ (estimado utilizando selección robusta)	No puede rechazar

NOTA: $Y = DY_1 + (1-D)Y_0$; $Y_1 = \mu_1(\mathbf{x}) + U_1$; $Y_0 = \mu_0(\mathbf{x}) + U_0$; \mathbf{Z} independiente de (U_0, U_1) , \mathbf{Z} no independiente de D ; $D = 1$ ($Y_2 = \mu_2(\mathbf{Z}) + U_2 \geq 0$) donde $\mu_2(\mathbf{Z}) + U_2$ es el índice que determina la selección como "1" o "0". Hipótesis: No hay selección en los no observables, $H_0: E(U_1 - U_0 | D=1, \mathbf{Z}, \mathbf{X}) = M(\mathbf{X})$ (en un modelo normal, $\sigma_{12} = \sigma_{02}$).

7.3. Parámetros de tratamiento que son relevantes para las medidas de política económica

Los parámetros de tratamiento convencionales se justifican de forma intuitiva. Su relación con el análisis de coste-beneficio y con los marcos económicos interpretables es opaca. Heckman y Smith (1998) desarrollan

la relación entre estos parámetros y los que corresponden al análisis coste-beneficio. En ocasiones, los parámetros tradicionales responden a preguntas interesantes relativas a la política y en otras no lo hacen.

En Heckman y Vytlačil (2001a, 2001c, 2001d, 2002) se desarrolla un enfoque más directo para definir los parámetros económicos de tratamiento que consiste en plantear una pregunta de política económica o un problema de toma de decisión de interés y derivar y estimar el parámetro que responde a esta pregunta o problema. Por lo general, al adoptar esta postura, no se producen los parámetros de tratamiento convencionales.

Consideremos un tipo de políticas que afectan a P , la probabilidad de participar en un programa, pero que no afecta directamente al MTE. Un ejemplo en economía de la educación serían las políticas que cambian los gastos de inscripción o las distancias hasta el colegio, pero que no inciden directamente sobre los rendimientos brutos de la escolarización. Definamos P como la probabilidad que sirva de referencia y P^* como la probabilidad producida bajo un régimen de política económica alternativo.

Para simplificar, comparemos las políticas usando el criterio de Bentham y consideremos el efecto de las políticas sobre la utilidad media de los individuos con un nivel dado de $\mathbf{X} = \mathbf{x}$. Para la utilidad V , suponiendo que $E(V(Y_1))$ y $E(V(Y_0))$ existen y son finitos, tenemos que

$$\begin{aligned}
 & E(V(Y)|\text{se aplica la política, } \mathbf{X}=\mathbf{x}) \\
 & -E(V(Y)|\text{se está en la situación de referencia, } \mathbf{X}=\mathbf{x}) \\
 & = \int_0^1 \Delta_V^{\text{MTE}}(\mathbf{x}, u_D) \omega^*(\mathbf{x}, u_D) du_D \tag{30}
 \end{aligned}$$

donde las ponderaciones de políticas son

$$\omega^*(\mathbf{x}, u_D) = F_{P|\mathbf{X}}(u_D|\mathbf{x}) - F_{P^*|\mathbf{X}}(u_D|\mathbf{x})$$

siendo $F_{P|\mathbf{X}}(\cdot|\mathbf{x})$ la distribución de P condicionada a $\mathbf{X} = \mathbf{x}$,⁵⁷ y

$$\Delta_V^{\text{MTE}}(\mathbf{x}, u_D) = E(V(Y_1) - V(Y_0)|\mathbf{X}=\mathbf{x}, U_D=u_D) = E(\Delta_V|\mathbf{X}=\mathbf{x}, U_D=u_D)$$

donde $\Delta_V = V(Y_1) - V(Y_0)$. Cuando V es la función identidad comparamos resultados medios como en el análisis convencional de coste - beneficio.

(57) Si seguimos condicionando implícitamente en función de \mathbf{X} , tenemos

$$\begin{aligned}
 E(V(Y)|\text{situación de referencia}) &= \int_0^1 E(V(Y)|P(\mathbf{Z})=p) dF_P(p) \\
 &= \int_0^1 \left[\int_0^1 \mathbf{1}_{[0, p]}(u) E(V(Y_1)|U=u) + \mathbf{1}_{[p, 1]}(u) E(V(Y_0)|U=u) du \right] dF_P \\
 &= \int_0^1 \{ [1 - F_P(u)] E(V(Y_1)|U=u) + F_P(u) E(V(Y_0)|U=u) \} du
 \end{aligned}$$

donde $\mathbf{1}_A(u)$ es una función indicadora del acontecimiento $u \in A$. Si comparamos de esta manera la situación de referencia con el nuevo régimen, obtenemos

$$E_{P^*}(V(Y)) - E_P(V(Y)) = \int_0^1 E(\Delta_V|U=u) [F_P(u) - F_{P^*}(u)] du$$

El parámetro de política económica es una media ponderada del MTE como se define anteriormente. En lugar de esperar que los parámetros de tratamiento convencional respondan a preguntas económicas interesantes, un enfoque mejor es el de estimar Δ_V^{MTE} y ponderarlo teniendo en cuenta la manera en que la política económica cambia la distribución de P .

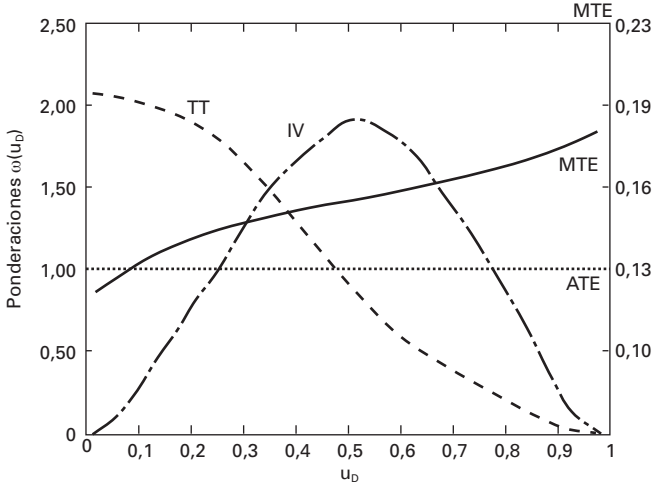
Un enfoque alternativo en materia de evaluación de políticas económicas consiste en producir una variable instrumental ponderada por la política basada en una elección específica de $\omega(\mathbf{x}, u_D)$ que captura el efecto provocado por el cambio de política. Si elegimos las ponderaciones para el estimador $\omega_k(\mathbf{x}, u_D)$ en (2) de tal manera que coincidan con las ponderaciones para el cambio de política económica, $\omega^*(\mathbf{x}, u_D)$ en (3), podemos producir un estimador a la medida del cambio político que nos interesa. Heckman y Vytlacil (2001c, 2001d, 2002) establecen que la variable instrumental relevante para la política económica es $[f_{P^*}(P)/f_P(P)] - 1$, donde f_{P^*} y f_P son las densidades de P^* y P , respectivamente. Podemos determinar la distribución de P y P^* independientemente de los demás ingredientes necesarios para formar el parámetro de tratamiento relevante.

El gráfico 9, extraído de la investigación de Carneiro *et al.* (2001), representa el MTE estimado como función de u_D relativo a los rendimientos de la formación universitaria para una muestra de hombres blancos en Estados Unidos a finales de los años ochenta y a principios de los noventa. Es una función creciente de u_D , sugiriendo que los rendimientos monetarios alcanzan el nivel más alto para aquellas personas con menos probabilidad de ir a la universidad. También se encuentran representadas las funciones de ponderación del MTE que están implícitas en la definición del TT y del ATE y en el uso del método IV lineal convencional con $P(\mathbf{Z})$ como instrumento para estimar "el" efecto de la escolarización sobre la educación. Las ponderaciones para los distintos parámetros de tratamiento difieren las unas de las otras y son distintas a las del IV. El hecho de que el MTE se incremente implica que los métodos convencionales de matching y de IV lineal no identifican ni el TT ni el ATE en estos datos.

El gráfico 10, extraído también de Carneiro *et al.* (2001), muestra la ponderación de política del MTE para tres políticas económicas definidas en la leyenda del gráfico.⁵⁸ El efecto del tratamiento producido por el método IV pondera el MTE acercándolo a la ponderación necesaria para evaluar la política III, pero se encuentra muy lejos de lo requerido para evaluar las políticas I y II. La proximidad de las ponderaciones IV y de las ponderaciones de la política III es fortuita. Los experimentos sociales que excluyen a la gente del programa de forma aleatoria en el momento en el que solicitan el derecho de participación y logran una aceptación, estiman TT bajo conformidad total (véase Heckman, LaLonde y Smith, 1999; Heckman

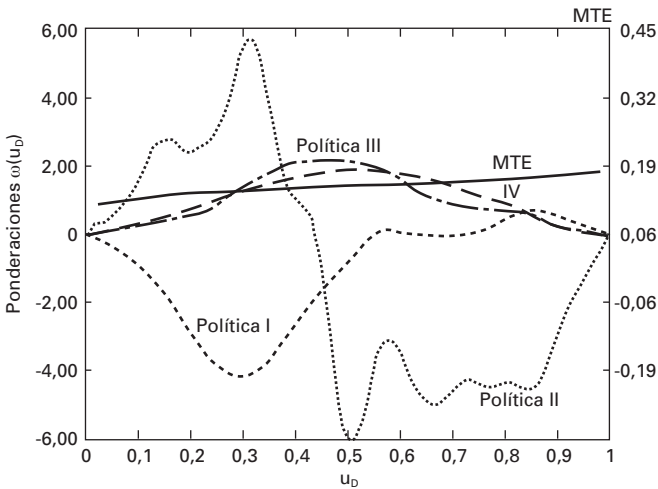
(58) Estas ponderaciones de políticas se normalizan mediante $D\bar{P}$, lo que es la proporción de personas en la población que se ven inducidas a cambiar su condición de escolarización por la intervención. Esto hace que las ponderaciones de políticas sean comparables a las ponderaciones IV.

Gráfico 9
EFFECTO DE TRATAMIENTO MARGINAL VS. PONDERACIONES ATE, IV Y TT



Fuente: Carneiro *et al.* (2001).

Gráfico 10
EFFECTO DEL TRATAMIENTO MARGINAL VS. PONDERACIONES DE POLÍTICAS Y DE IV: MODELO DE REFERENCIA SIN INTERACCIONES



Fuente: Carneiro *et al.* (2001). Nota.- Política I: llevar a la gente hacia los extremos de la distribución de la enseñanza. Política II: llevar a la gente a la media de la distribución de la enseñanza. Política III: disminución de 1.000 dólares US en la distribución de la enseñanza.

y Vytlačil, 2002). Dada la forma de las ponderaciones de TT, tales experimentos no evalúan de forma precisa el efecto sobre las respuestas medias de ninguna de las tres políticas que se consideran en el gráfico 10.

Si el MTE es plano, hay un solo "efecto" para todas las políticas y el método IV estima este efecto. Por lo general y en la mayoría de los estudios revisados en la panorámica de Heckman y Vytlačil (2002) así como en el cuadro 3, se asocian distintas ponderaciones a distintas políticas, y solamente de forma accidental el método IV lineal podría identificar la respuesta adecuada de política económica.

Cuando el soporte de P^* no está contenido en el soporte de P , de tal manera que la intervención política que se estudia extiende P^* fuera de los datos históricos, resulta necesario establecer más supuestos para realizar un análisis de política económica fundamentado. Si se hacen supuestos paramétricos sobre $P(\mathbf{Z})$, si la probabilidad viene determinada por los datos históricos y si la intervención que se estudia cambia la distribución de \mathbf{Z} de forma conocida, entonces es inmediato determinar la distribución de $P^*(\mathbf{z})$, incluyendo el nuevo soporte. Sin embargo, el MTE se identifica sólo no paramétricamente en el soporte de P (véase Heckman y Vytlačil, 2000a, 2002).

7.4. Extrapolar a nuevas poblaciones

Si lo que queremos es aplicar el MTE estimado a nuevas poblaciones en las que la dependencia entre \mathbf{X} y (U_0, U_1, U_D) es diferente de la de la muestra utilizada para estimarlo, o en las que el soporte de (\mathbf{X}, U_D) es diferente del que se emplea en la muestra de estimación, es necesario hacer el mismo tipo de supuestos de independencia y de soporte que los que se han examinado anteriormente en la sección 2. Esto ha sido reconocido por los econométricos desde la época de Haavelmo (1943). La literatura sobre el efecto del tratamiento evita los supuestos estructurales eludiendo las preguntas planteadas por la econometría estructural.

7.5. Comparación de tres enfoques

El cuadro 4 compara los tres enfoques de evaluación de política económica de los que he hablado en esta conferencia, poniendo de manifiesto las ventajas y las limitaciones de cada uno de ellos: el enfoque estructural, el enfoque convencional del efecto del tratamiento y el enfoque del efecto del tratamiento basado en el MTE que se ha desarrollado recientemente.

El enfoque basado en el MTE comparte con el enfoque estructural la interpretabilidad de los parámetros. Al igual que el enfoque estructural, trata un conjunto de preguntas de evaluación de política económica. El parámetro es menos comparable y se presta a ser extrapolado de unos entornos a otros en menor medida que los parámetros estructurales, a no ser que se establezcan supuestos específicos sobre la relación entre observables y no observables, tanto en las muestras de estimación como en las muestras objetivo. Sin embargo, es comparable entre poblaciones

Cuadro 4

	Enfoque econométrico estructural	Enfoque de efecto del tratamiento	Enfoque basado en el MTE
Interpretabilidad	Parámetros económicos bien planteados y comparaciones de bienestar	Relación con la economía y comparaciones de bienestar: poco claro	Interpretable en términos de la disposición a pagar; las medias ponderadas del MTE responden a preguntas económicas bien planteadas
Tipo de preguntas tratadas	Responde a muchas preguntas contrafactuales	Se centra en un efecto del tratamiento o en una pequeña gama de efectos	Con condiciones sobre el soporte, genera todos los parámetros de tratamiento
Extrapolación a nuevos entornos	Aporta ingredientes para la extrapolación	Evalúa un programa dado en un entorno determinado	Puede ser extrapolado parcialmente; se extrapola a nuevos entornos con distribuciones diferentes de la probabilidad de participación si las diferencias surgen exclusivamente por diferencias en las distribuciones de Z
Comparabilidad de unos estudios con otros	Parámetros que no varían con las políticas: comparables de unos estudios con otros	No comparables en general	Parcialmente comparables; comparables de unos entornos a otros que difieren en las distribuciones de la probabilidad de participación si las diferencias surgen exclusivamente por diferencias en las distribuciones de Z
Problemas econométricos clave	Exogeneidad; super-exogeneidad y sesgo de selección	Sesgo de selección	Sesgo de selección
Tipos de políticas que pueden ser evaluadas	Programas de aplicación parcial o universal dependiendo de la variación en los datos (precios/dotaciones)	Programas de aplicación parcial (grupos de tratamiento y de control)	Programas de aplicación parcial (grupos de tratamiento y de control)
Extensión a evaluación de equilibrio general	Necesidad de relacionar con datos de series temporales; parámetros compatibles con la teoría del equilibrio general	Difícil debido a que la relación con la economía no está especificada con precisión	Puede relacionarse con modelos no paramétricos de equilibrio general

con distribuciones diferentes de P , y los resultados de una población pueden aplicarse a otra población a condición de que la dependencia P - X esté controlada. El coste de estas ventajas del enfoque estructural es que hay un mayor número de problemas de tipo econométrico a los que hay que aportar una solución. El enfoque convencional del tratamiento produce parámetros que no pueden relacionarse con modelos económicos bien planteados y que, por lo tanto, no proporcionan los cimientos para un análisis de equilibrio general con motivación empírica. El MTE estima las preferencias de los agentes que se estudian y proporciona la base para una integración con modelos económicos bien planteados.

7.6. Evidencia extraída de la literatura sobre efectos del tratamiento

A pesar de sus limitaciones, la literatura sobre efectos del tratamiento ha producido un número importante de estudios destacables sobre medidas de política económica que han cambiado nuestra forma de entender la eficacia de las mismas. Esta literatura es demasiado amplia como para resumirla en esta conferencia.

Esbozaré brevemente algunos de los resultados empíricos más relevantes sobre evaluación de políticas del mercado de trabajo que extraigo de mi estudio panorámico de este campo que escribí con LaLonde y Smith (Heckman, LaLonde y Smith, 1999). De unos países a otros y a lo largo del tiempo, la mayoría de las políticas activas del mercado de trabajo son ineficaces para promover incrementos salariales y el empleo a largo plazo. La aplicación de métodos empíricos básicos indica que la comparación de respuestas de los participantes antes y después de la aplicación de un programa, que es práctica común en muchos planes de evaluación, sobreestima enormemente los beneficios de participar en el programa de formación. Esto ocurre porque los participantes en un programa de formación suelen tener problemas de empleo justo antes de entrar en el programa y el programa en sí constituye un procedimiento de búsqueda de trabajo.⁵⁹ Gran parte de las mejoras en los resultados después del programa habrían tenido lugar en ausencia de cualquier programa de formación.

La aplicación de principios básicos (como el utilizar un grupo de control de no participantes en programas de formación que sea comparable en localización geográfica y en historial de empleo o entregar cuestionarios a participantes escogidos al azar) tiende a eliminar el sesgo de selección en la evaluación de programas sociales. Sin embargo, esto no elimi-

(59) Ashenfelter (1978) demostró que los participantes de un programa sufren una disminución en sus ingresos antes de inscribirse en el mismo. Heckman (1978b) y Heckman y Robb (1985) presentan el primer modelo económico para este fenómeno utilizando un modelo basado en la maximización del valor presente de la renta como criterio de participación en el programa. Heckman, Ichimura, Smith y Todd (1998), Heckman y Smith (1998) y Heckman, LaLonde y Smith (1999) demuestran que es la dinámica de empleo, y no la dinámica de ingresos, lo que afecta a la inscripción en los programas. Heckman, LaLonde y Smith (1999) presentan un modelo de búsqueda que extiende el análisis de Heckman y Robb (1985).

na todo el sesgo y el tomar en cuenta el efecto de selección sobre variables no observables es importante para obtener estimaciones precisas del parámetro denominado "tratamiento sobre los tratados" que es el comúnmente utilizado⁶⁰ (véase Heckman, Ichimura, Smith y Todd 1996, 1998; Heckman, LaLonde y Smith 1999).

8. UNIENDO LA MACRO Y LA MICROECONOMETRÍA: EVALUACIÓN DE POLÍTICA ECONÓMICA EN UN MARCO DE EQUILIBRIO GENERAL

La evidencia procedente de los datos microeconómicos ya ha tenido un efecto importante sobre el desarrollo de la teoría macroeconómica, que poco a poco va abandonando el paradigma del agente representativo. Ya he hablado de cómo el tomar en cuenta la elección en los márgenes extensivos ha modificado las discusiones macroeconómicas sobre el mercado de trabajo. Se podrían dar muchos otros ejemplos de evidencia resultante de datos microeconómicos que han incidido sobre la teoría macroeconómica (véase la panorámica de Browning, Hansen y Heckman, 1999).

Los métodos microeconómicos han sido desarrollados para evaluar una multitud de programas sociales implementados por el Estado del bienestar moderno y para interpretar las relaciones económicas empíricas. Pero el campo de aplicación de la microeconomía tomada de manera aislada es necesariamente limitado. Muchos programas, tales como los de becas de estudio de los que se ha hablado en la sección 7, tienen un carácter nacional y se prestan a tener efectos de equilibrio general. El crecimiento del número de personas formadas reducirá probablemente el rendimiento del trabajo realizado por personas con formación. La reducción de los impuestos sobre el trabajo hace que aumente la oferta de trabajo y disminuyan los salarios reales. Los métodos de equilibrio parcial no pueden ir más allá de esto para evaluar el impacto completo de los programas públicos a gran escala. La metodología del efecto del tratamiento resulta totalmente inefectiva cuando se trata de analizar programas de aplicación universal a no ser que sean economías enteras que puedan servir de grupo de tratamiento y de grupo de control.

(60) Un estudio influyente de LaLonde (1986) pone en duda la capacidad de los métodos econométricos no experimentales para evaluar programas sociales. Los análisis posteriores de Heckman, Ichimura, Smith y Todd (1996, 1998) demostraron que los descubrimientos de LaLonde estaban generados por la comparación de personas incomparables. Sus participantes en los programas viven en mercados de trabajo diferentes, responden a cuestionarios distintos y tienen valores de X diferentes de los de su grupo de comparación. El hacer que los participantes sean comparables a personas que no participan en ningún programa elimina la mayoría, pero no la totalidad del sesgo de LaLonde. Véanse los importantes artículos de Smith y Todd (2000, 2001), que vuelven a analizar los datos de LaLonde y demuestran que los métodos de "matching" no eliminan los componentes empíricamente importantes del sesgo de selección de LaLonde. Para problemas con experimentos sociales, véase Heckman (1992), Heckman y Smith (1993, 1995), Manski (1996), Heckman, LaLonde y Smith (1999) y Heckman y Vytlacil (2001a).

Es necesaria una síntesis de los enfoques macroeconómicos y microeconómicos para analizar las políticas aplicadas a nivel nacional con impactos de equilibrio general y para interpretar relaciones de precios de equilibrio tales como las funciones de ingresos o las ecuaciones de precios de activos. La variación de corte transversal no puede identificar los efectos de precios y de tipos de interés que son los mismos para todos. Sin embargo, es esencial, para investigar todos los efectos de políticas nacionales sobre la formación de la población, el tener en cuenta el efecto de retroalimentación de los mercados de capitales (costes del préstamo) sobre las decisiones de inversión en capital humano y físico. Además, la variación de corte transversal en los salarios, en los precios o en los tipos de interés tiene pocas probabilidades de ser exógena y esto da lugar a otras cuestiones econométricas.

La síntesis necesaria del equilibrio general macroeconómico y de la microeconomía que propone Orcutt (1962) no ha hecho más que empezar, pero los primeros resultados de este programa de investigación son prometedores. El objetivo de esta línea de investigación es desarrollar un equilibrio general con base empírica que mejore la calibración como fuente de estimación para los parámetros de modelos de equilibrio general y que aporte la base teórica y empírica rigurosa para la evaluación de programas sociales a gran escala como las becas de estudios que alteran los precios y las reformas de la seguridad social con alcance universal.

Heckman, Lochner y Taber (1998a, 1998b, 1998c, 1999) presentan un prototipo de síntesis de datos microeconómicos y datos macroeconómicos. En ese trabajo generalizamos el marco de trabajo de Auerbach y Kotlikoff (1987) para formular y estimar un modelo de formación, en un marco de equilibrio general dinámico y de predicción perfecta, que generaliza el modelo de Roy en un planteamiento completamente dinámico con acumulación endógena de formación. Los datos macro y microeconómicos se combinan para determinar los parámetros del modelo que recoge la heterogeneidad y la auto-selección en el mercado de trabajo y que explica la creciente desigualdad salarial en Estados Unidos en las dos últimas décadas. El análisis de políticas económicas basado en este modelo indica que el no tomar en cuenta los efectos de equilibrio general tal como se suele hacer en la literatura sobre el efecto del tratamiento y en gran parte de la literatura microeconométrica de equilibrio parcial, sobreestima en gran medida los efectos de las reducciones de becas sobre el número de ingresos en la universidad. Es importante, tanto por el contenido como por la teoría, tener en cuenta los efectos de equilibrio general.⁶¹ El reto en esta literatura está en el desarrollo de relaciones estructurales empíricamente creíbles basadas en datos microeconómicos que pueden correlacionarse con los agregados macroeconómicos.

(61) Lee (2000) presenta una ampliación de este modelo para elección relativa a la ocupación en vez de elección relativa a la educación. Los estudios de Calmfors (1994) y Davidson y Woodbury (1993) demuestran la importancia de tener en cuenta el efecto de desplazamiento al evaluar distintas políticas activas del mercado de trabajo. Véase la panorámica de Heckman, LaLonde y Smith (1999).

9. RESUMEN Y CONCLUSIONES

En los últimos cincuenta años, la economía se ha visto enriquecida por la gran cantidad de nuevas fuentes de datos microeconómicos. Estos datos han hecho que los economistas tomen conciencia de la diversidad y la heterogeneidad de la vida económica. Han permitido a los economistas entender de forma más completa un conjunto amplio de problemas sociales y evaluar programas sociales diseñados para resolver estos problemas. Aquéllos que comenzaron la recolección de datos microeconómicos merecen nuestro agradecimiento sincero.

Estos datos pusieron en tela de juicio los métodos econométricos tradicionales. Ciertos problemas que parecen sin importancia al examinar medias agregadas se transforman en problemas cruciales al analizar datos microeconómicos. Fueron estos problemas, así como el interés por evaluar políticas económicas, lo que motivó la recopilación sistemática de datos microeconómicos, los cuales dieron lugar a la microeconomía moderna. Esta disciplina une la economía y la estadística para producir resúmenes interpretables de datos microeconómicos, para contrastar teorías sobre el individuo usando datos de individuos, y para construir contrafactuales económicos.

El campo de la microeconomía está en plena expansión. Se han realizado importantes progresos en comprender cuáles son las fuentes de identificación de los modelos, en relajar formas funcionales arbitrarias y supuestos sobre la distribución, y en diseñar y analizar nuevas encuestas. Se está desarrollando un enfoque más robusto para la evaluación de políticas económicas que hará del análisis de éstas una actividad más objetiva y de mayor rigor empírico.

Esta disciplina seguirá desarrollándose si se renueva para enfrentarse a los nuevos retos que surjan de los nuevos problemas abordados en economía. Si sólo se limita a buscar refinamientos de los modelos originales que crearon el campo, morirá.

Entre los retos importantes en esta disciplina, se incluye el desarrollo de una teoría del equilibrio general basada en los datos microeconómicos para contrastar la teoría y evaluar los impactos de políticas aplicadas a gran escala. También se incluye entre estos retos el desarrollo de esquemas econométricos de análisis de coste-beneficio empíricamente creíbles para evaluar políticas microeconómicas que acercan la literatura sobre evaluación de programas a la teoría económica. Tengo el convencimiento de que los microeconomistas sabrán enfrentarse con éxito a estos y otros retos y de que en los próximos años obtendrán innovaciones de la investigación en esta disciplina que expondrán desde esta tribuna.

Apéndice 1. Condiciones de coherencia y causalidad

Consideremos un modelo prototípico de ecuaciones simultáneas para dos variables endógenas (Y_1 , Y_2) expresadas como función de variables exógenas X :

$$\begin{aligned} \alpha_{11}Y_1 + \alpha_{12}Y_2 &= \mathbf{X}\beta_1 + U_1 \\ \alpha_{21}Y_1 + \alpha_{22}Y_2 &= \mathbf{X}\beta_2 + U_2 \end{aligned} \tag{A1}$$

y

$$E(U_1, U_2 | \mathbf{X}) = (0, 0)$$

El grupo de la Cowles desarrolló una teoría muy elaborada para identificar y estimar este tipo de modelos cuando Y_1 e Y_2 son variables continuas del tipo que aparece en ecuaciones convencionales de equilibrio de mercado o de demanda con solución interior.

¿Hasta qué punto esta teoría se presta a ser trasladada a los planteamientos en los que (Y_1, Y_2) son variables discretas o variables mixtas continuas y discretas? Estas preguntas se formularon en artículos de Amemiya (1973, 1974), Heckman (1976c, 1978a), Gouriéroux, Laffont y Monfort (1980) y Schmidt (1981) en diferentes planteamientos. Nótese, en primer lugar, que si Y_1 y Y_2 son discretas, (U_1, U_2) no puede ser continuo. En particular, los supuestos convencionales de normalidad para (U_1, U_2) no son apropiados. Al redefinir (A1) como un modelo en el que Y_1 y Y_2 son variables latentes con $D_1 = 1(Y_1 \geq 0)$ y $D_2 = 1(Y_2 \geq 0)$ preservamos el paradigma de la Cowles.⁶² Sin embargo, el contenido del modelo en cuanto a ecuaciones de comportamiento no siempre está claro y conlleva relaciones entre variables latentes que son difíciles de justificar con una teoría precisa. Más que un modelo con una justificación económica firme, éste es una copia por analogía de las ecuaciones simultáneas convencionales.

Un modelo con una fundamentación conductual más clara contiene variables ficticias que desplazan las ecuaciones determinantes de los índices latentes.⁶³ Por consiguiente, las variables (Y_1, Y_2) son latentes y

$$\begin{aligned} Y_1 &= \gamma_{11}D_1 + \gamma_{12}D_2 + \mathbf{X}\beta_1 + U_1 \\ Y_2 &= \gamma_{21}D_1 + \gamma_{22}D_2 + \mathbf{X}\beta_2 + U_2 \end{aligned} \tag{A2a}$$

donde

$$D_1 = 1(Y_1 \geq 0), D_2 = 1(Y_2 \geq 0) \tag{A2b}$$

Este modelo tiene la característica extraña de que Y_i causa D_i pero D_i también causa Y_i . Puesto que las dos están relacionadas mecánicamente a partir de (A2b), parece más natural establecer $\gamma_{11} = \gamma_{22} = 0$. Pero incluso en este caso, surgen problemas de lógica: $D_1 = 0$ puede coexistir con $Y_1 \geq 0$. Esto produce modelos con probabilidades negativas o probabilidades superiores a uno. La aplicación mecánica de los métodos de la Cowles fracasa. Descartar estas patologías para un (U_1, U_2) general requiere que

$$\gamma_{12}\gamma_{21} = 0 \text{ (condición de coherencia)} \tag{A3}$$

(62) Este modelo fue desarrollado por Heckman (1973, 1976c, 1978a) y Mallar (1977).

(63) La versión Tobit de este modelo se desarrolló en Heckman (1973, 1976c, 1978a) y fue aplicada por mi alumno Lawrence Olson (Nelson y Olson, 1978).

Esta condición parece descartar una verdadera simultaneidad y hace que los modelos adquieran una forma recursiva. De hecho, elimina los modelos económicos malos. En cada aplicación, la condición de coherencia tiene una interpretación económica clara (véase, por ejemplo, Heckman, 1976c, 1978a; Ransom, 1987; Blundell y Smith, 1994). Heckman y MaCurdy (1985) presentan una discusión exhaustiva de la condición de coherencia.

El modelo (A2a) y (A2b) nos permite hacer las distinciones entre causalidad espuria y causalidad verdadera, que son un elemento central de la econometría de la Cowles, pero con tipos más generales de variables endógenas. Si $\gamma_{11} = \gamma_{12} = \gamma_{22} = 0$, pero $\gamma_{21} \neq 0$, existe un verdadero efecto de D_1 sobre D_2 si $\gamma_{21} \neq 0$, pero aparece un efecto espurio si D_1 no es independiente de U_2 . Un modelo a la Cowles reformulado adecuadamente puede todavía hacer las distinciones cruciales de la Cowles entre asociaciones causales y asociaciones estadísticas.⁶⁴

Este tipo de análisis también ilustra los beneficios que aporta el enfoque econométrico que une la estadística y la economía. Los modelos logarítmicos lineales para analizar datos discretos, desarrollados por estadísticos (véase, por ejemplo, Goodman 1970; Bishop *et al.*, 1975) no pueden distinguir el efecto de γ_{21} del de U_2 no independiente de D_1 , y por lo tanto no pueden utilizarse para hacer distinciones causales (Heckman, 1978a). Las extensiones con datos de panel de estos modelos se pueden usar para distinguir si el que un acontecimiento haya ocurrido en el pasado afecta causalmente a la probabilidad de que se produzcan eventos futuros o simplemente figura como componente no medido, es decir, estos modelos pueden examinar cuestiones de causalidad en un planteamiento dinámico, tema que abordamos en la sección 6.

Apéndice 2. Análisis de acotamiento y de sensibilidad

Si partimos de la ecuación (19) o de su versión para las medias condicionales, los artículos de Smith y Welch (1986), Holland (1986) y Glynn, Laird y Rubin (1986) caracterizan el problema de selección de forma más general sin la estructura de índice, y ofrecen métodos bayesianos y clásicos para llevar a cabo un análisis de sensibilidad para los efectos que producen supuestos de identificación diferentes sobre inferencias relativas a la media de la población.

La selección de observables resuelve el problema de selección suponiendo que Y_1 es independiente de $D|X$ de tal manera que $F(Y_1|X, D=1)=F(Y_1|X)$. Este es el supuesto en el que se apoyan los modelos de matching. Es incon-

(64) Heckman (1976c) estima las ecuaciones (A2a) y (A2b) bajo la condición de coherencia. El término Y_1 es un índice de sentimiento a favor de aprobar una ley de antidiscriminación e Y_2 es una medida del impacto que tendría sobre variables tales como salarios o empleo.

sistente con el modelo de Roy de auto-selección (Heckman y Vytlacil, 2001d, 2002).

En la literatura se han propuesto distintos enfoques para acotar esta distribución, o los momentos de la misma; todos ellos a partir de los hallazgos de Holland (1986) y Peterson (1976). Para ilustrar estas ideas con el planteamiento más simple posible, sea $g(Y_1|\mathbf{X}, D = 1)$ la densidad de las respuestas (por ejemplo, salarios) para personas que trabajan ($D = 1$ corresponde a trabajar). Supongamos que son muestras censuradas. Falta $g(Y|\mathbf{X}, D = 0)$ (es decir, la densidad de los salarios de los que no trabajan).

Con el fin de estimar $E(Y_1|\mathbf{X})$, Smith y Welch (1986) utilizan la ley de expectativas iteradas para obtener

$$E(Y_1|\mathbf{X}) = E(Y_1|\mathbf{X}, D=1) \Pr(D=1|\mathbf{X}) + E(Y_1|\mathbf{X}, D=0) \Pr(D=0|\mathbf{X})$$

Para estimar el miembro de la izquierda de esta expresión, es necesario obtener información sobre el componente ausente $E(Y_1|\mathbf{X}, D = 0)$. Smith y Welch proponen y ponen en práctica acotaciones sobre $E(Y_1|\mathbf{X}, D = 0)$, por ejemplo, $Y_L \leq E(Y_1|\mathbf{X}, D = 0, Z) \leq Y^U$ donde Y^U es una cota superior y Y_L una cota inferior.⁶⁵ Utilizando esta información construyen las cotas

$$\begin{aligned} E(Y_1|\mathbf{X}, D=1)\Pr(D=1|\mathbf{X}) + Y_L\Pr(D=0|\mathbf{X}) &\leq E(Y_1|\mathbf{X}) \\ &\leq E(Y_1|\mathbf{X}, D=1) \Pr(D=1|\mathbf{X}) + Y^U\Pr(D=0|\mathbf{X}) \end{aligned}$$

Llevando a cabo un análisis de sensibilidad, producen una serie de valores para $E(Y|\mathbf{X})$ que dependen explícitamente de los valores supuestos para $E(Y|\mathbf{X}, D = 0)$. Trabajos posteriores de Manski (1989, 1990, 1994, 1995), Robins (1989) y Horowitz y Manski (1995) desarrollan este tipo de análisis de manera más sistemática para diferentes modelos.

Glynn *et al.* (1986) presentan un análisis de sensibilidad para las distribuciones empleando métodos bayesianos bajo una serie de supuestos distintos sobre $F(Y_1|\mathbf{X}, D = 0)$ con el objetivo de determinar los posibles valores de $F(Y|\mathbf{X})$. Holland (1986) propone un análisis de sensibilidad más clásico que hace variar los posibles parámetros de los modelos. Rosenbaum (1995) discute distintos análisis de sensibilidad.

El objetivo de estos análisis de cotas y de los análisis de sensibilidad bayesianos y clásicos es separar claramente lo que se conoce de lo que se conjetura sobre los datos y explorar la sensibilidad de las estimaciones obtenidas a los supuestos empleados para conseguirlos. Manski (1990, 1994) y Heckman y Vytlacil (2000a, 2000b, 2001b) demuestran cuáles son las restricciones adicionales que conlleva el empleo de modelos de índice para producir acotaciones sobre las respuestas.

(65) En su problema hay bandas plausibles de salarios que pueden ganar los que abandonan.

Gran parte del análisis teórico presentado en la literatura reciente es no paramétrico, aunque, en la práctica, una parte importante de la experiencia en estadística y econometría demuestra que la estimación no paramétrica de grandes dimensiones no es factible con los tamaños habituales de las muestras de las que se dispone en la econometría de corte transversal. Debemos imponer alguna forma de estructura si queremos obtener una estimación no paramétrica fiable. No obstante, unas versiones paramétricas factibles de estos métodos corren el riesgo de imponer una estructura paramétrica falsa.⁶⁶

REFERENCIAS BIBLIOGRÁFICAS

- Ahn, H. y Powell, J. L. (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism", *Journal of Econometrics*, vol. 58, julio, pp. 3-29.
- Altug, S. y Miller, R. A. (1990): "Household Choices in Equilibrium", *Econometrica*, vol. 58, mayo, pp. 543-570.
- Altug, S. y Miller, R. A. (1998): "The Effect of Work Experience on Female Wages and Labour Supply", *Review of Economic Studies*, vol. 65, enero, pp. 45-85.
- Amemiya, T. (1973): "Regression Analysis When the Dependent Variable Is Truncated Normal", *Econometrica*, vol. 41, noviembre, pp. 997-1016.
- Amemiya, T. (1974): "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal", *Econometrica*, vol. 42, noviembre, pp. 999-1012.
- Amemiya, T. (1985): *Advanced Econometrics*, Harvard University Press, Cambridge.
- Andrews, Donald W. K. (1991): "Asymptotic Normality of Series Estimators for Non-parametric and Semiparametric Regression Models", *Econometrica*, vol. 59, marzo, pp. 307-345.
- Angrist, J. D.; Graddy, K. y Imbens, G. W. (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish", *Review of Economic Studies*, vol. 67, julio, pp. 499-527.

(66) Los métodos utilizados en la literatura sobre cotas dependen fuertemente de la elección de las variables condicionantes X . En principio, todas las elecciones posibles de variables condicionantes deberían explorarse, especialmente al calcular las cotas para todos los modelos posibles, pero en la práctica nunca se hace. Si se exploraran, la banda de estimaciones producida por las cotas sería mucho mayor de lo que ya reflejan las amplias cotas que recoge esta literatura.

- Arabmazar, A. y Schmidt, P. (1981): "Further Evidence on the Robustness of the Tobit Estimator to heteroskedasticity", *Journal of Econometrics*, vol. 17, noviembre, pp. 253-258.
- Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings", *Review of Economics and Statistics*, vol. 60, febrero, pp. 47-57.
- Athey, S. y Haile, P. (2000): "Identification of Standard Auction Models", manuscrito, Massachusetts Inst. Tech., Cambridge.
- Atkinson, R. C.; Bower, G. H. y Crothers, E. J. (1965): *An Introduction to Mathematical Learning Theory*, Wiley, Nueva York.
- Auerbach, A. J. y Kotlikoff, L. J. (1987): *Dynamic Fiscal Policy*, Cambridge University Press, Cambridge.
- Barnow, B. S.; Cain, G. G. y Goldberger, A.S. (1980): "Issues in the Analysis of Selectivity Bias", en Stromsdorfer, E. W. y Farkas, G. (eds.), *Evaluation Studies: Review Annual*, vol. 5, Sage Publications, Beverly Hills.
- Ben-Porath, Y. (1973): "Labor-Force Participation Rates and the Supply of Labor", *Journal of Political Economy*, vol. 81, mayo-junio, pp. 697-704.
- Bera, A. K.; Jarque, C. M. y Lee, L.-F. (1984): "Testing the Normality Assumption in Limited Dependent Variable Models", *International Economic Review*, vol. 25, octubre, pp. 563-578.
- Bils, M. J. (1985): "Real Wages over the Business Cycle: Evidence from Panel Data", *Journal of Political Economy*, vol. 93, agosto, pp. 666-689.
- Bishop, Y. M.; Fienberg, S. E. y Holland, P. W. (1975): *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge.
- Björklund, A. y Moffitt, R. (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models", *Review of Economics and Statistics*, vol. 69, febrero, pp. 42-49.
- Blundell, R.; Reed, H. y Stoker, T. (1999): "Interpreting Movements in Aggregate Wages: The Role of Labor Market Participation", manuscrito, Universidad de California, Center Labor Econ., Berkeley.
- Blundell, R. y Smith, R. J. (1994): "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables", *Journal of Econometrics*, vol. 64, septiembre-octubre, pp. 355-373.
- Brown, C. (1984): "Black-White Earnings Ratios since the Civil Rights Act of 1964: The Importance of Labor Market Dropouts", *Quarterly Journal of Economics*, vol. 99, febrero, pp. 31-44.
- Browning, M. J.; Deaton, A. e Irish, M. (1985): "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle", *Econometrica*, vol. 53, mayo, pp. 503-543.

- Browning, M. J.; Hansen, L. P. y Heckman, J. J. (1999): "Micro Data and General Equilibrium Models", en Taylor, John B. y Woodford, M. (eds.), *Handbook of Macroeconomics*, vol. 1A, North-Holland, Amsterdam, pp. 543-633.
- Butler, R. y Heckman, J. J. (1977): "The Government's Impact on the Labor Market Status of Black Americans: A Critical Review", en Block, F. E. et al. (eds.), *Equal Rights and Industrial Relations*, Indus. Relations Res. Assoc., Madison, Wisconsin.
- Cain, G. G. (1975): "Regression and Selection Models to Improve Nonexperimental Comparisons", Bennett, C. A. y Lumsdaine, A. A. (eds.), *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*, Academic Press, Nueva York.
- Cain, G. G. y Watts, H. (1973): *Income Maintenance and Labor Supply*, Markham, Chicago.
- Calmfors, L. (1994): "Active Labour Market Policy and Unemployment—a Framework for the Analysis of Crucial Design Features", *OECD Economic Studies*, nº 22, primavera, pp. 7-47.
- Cameron, S. V. y Heckman, J. J. (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy*, vol. 106, abril, pp. 262-333.
- Cameron, S. V. y Heckman, J. J. (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males", *Journal of Political Economy*, vol. 109, junio, pp. 455-499.
- Campbell, D. T. y Stanley, J. C. (1963): *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago.
- Carneiro, P.; Heckman, J. J. y Vytlacil, E. J. (2001): "Estimating the Return to Education When It Varies among Individuals", trabajo presentado como *Economic Journal Lecture* en el encuentro de la Royal Economic Society, Durham, Inglaterra, abril de 2001 (presentado también como *Review of Economics and Statistics Lecture*, Universidad de Harvard, abril de 2001).
- Chandra, A. (2000): "Labor-Market Dropouts and the Racial Wage Gap: 1940-1990", *American Economic Review*, vol. 90, mayo, pp. 333-338.
- Chiappori, P.-A. y Heckman, J. J. (2000): "Testing Moral Hazard vs. Adverse Selection in Insurance Markets", manuscrito, Universidad de Chicago, Chicago.
- Cogan, J. F. (1981): "Fixed Costs and Labor Supply", *Econometrica*, vol. 49, julio, pp. 945-963.
- Coleman, Th. (1984): "Essays on Aggregate Labor Market Business Cycle Fluctuations", tesis doctoral, Universidad de Chicago, Chicago.

- Cosslett, S. R. (1981): "Maximum Likelihood Estimator for Choice-Based Samples", *Econometrica*, vol. 49, septiembre, pp. 1289-1316.
- Dagsvik, J. K. (1994): "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes", *Econometrica*, vol. 62, septiembre, pp. 1179-1205.
- Davidson, C. y Woodbury, S. A. (1993): "The Displacement Effect of Reemployment Bonus Programs", *Journal of Labor Economics*, vol. 11, octubre, pp. 575-605.
- Domencich, Th. A. y McFadden, D. (1975): *Urban Travel Demand: A Behavioral Analysis*, North-Holland, Amsterdam.
- Donald, S. G. y Paarsch, H. J. (1996): "Identification, Estimation, and Testing in Parametric Empirical Models of Auctions within the Independent Private Values Paradigm Source", *Econometric Theory*, vol. 12, agosto, pp. 517-567.
- Duncan, G. M. y Leigh, D. E. (1985): "The Endogeneity of Union Status: An Empirical Test", *Journal of Labor Economics*, vol. 3, julio, pp. 385-402.
- Dustmann, C. y Meghir, C. (2001): "Wages, Experience and Seniority", Working Paper nº 01/01, enero, Inst. Fiscal Studies, Londres.
- Eckstein, Z. y Wolpin, K. I. (1989): "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience", *Review of Economic Studies*, vol. 56, julio, pp. 375-390.
- Eckstein, Z. y Wolpin, K. I. (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities", *Econometrica*, vol. 67, noviembre, pp. 1295-1339.
- Elbers, C. y Ridder, G. (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model", *Review of Economic Studies*, vol. 49, julio, pp. 403-409.
- Engle, R. F.; Hendry, D. F. y Richard, J.-F. (1983): "Exogeneity", *Econometrica*, vol. 51, marzo, pp. 277-304.
- Fair, R. C. (1974-76): *A Model of Macroeconomic Activity*, 2 vols., Ballinger, Cambridge.
- Fair, R. C. (1994): *Testing Macroeconometric Models*, Harvard University Press, Cambridge.
- Farber, H. S. (1983): "Worker Preferences for Union Representation", en Reid, J. D. Jr. (ed.), *Research in Labor Economics*, suppl. 2, *New Approaches to Labor Unions*, JAI Press, Greenwich.
- Fisher, W. D. (1969): *Clustering and Aggregation in Economics*, Johns Hopkins Press, Baltimore.

- Flinn, C. J. y Heckman, J. J. (1982): "New Methods for Analyzing Structural Models of Labor Force Dynamics", *Journal of Econometrics*, vol. 18, enero, pp. 115-168.
- Glynn, R. J.; Laird, N. M. y Rubin, D. B. (1986): "Selection Modeling versus Mixture Modeling with Nonignorable Nonresponse", en Wainer, H. (ed.), *Drawing Inferences from Self-Selected Samples*, Springer-Verlag, Nueva York.
- Goldberger, A. S. (1964): *Econometric Theory*, Wiley, 1964.
- Goldberger, A. S. (1972): "Selection Bias in Evaluating Treatment Effects", Discussion Paper nº 123-172, Universidad de Wisconsin, Inst. Res. Poverty, Madison.
- Goldberger, A. S. (1983): "Abnormal Selection Bias", en Karlin, S.; Amemiya, T. y Goodman, L. A. (eds.), *Studies in Econometrics, Time Series, and Multivariate Statistics: In Honor of Theodore W. Anderson*, Academic Press, Nueva York.
- Goodman, L. A. (1968): "The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables with or without Missing Entries", *Journal of American Statistical Association*, vol. 63, diciembre, pp. 1091-1131.
- Goodman, L. A. (1970): "The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications", *Journal of American Statistical Association*, vol. 65, marzo, pp. 226-256.
- Gorman, W. M. (1980): "A Possible Procedure for Analysing Quality Differentials in the Egg Market", *Review of Economic Studies*, vol. 47, octubre, pp. 843-856.
- Gourieroux, C.; Laffont, J. J. y Monfort, A. (1980): "Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes", *Econometrica*, vol. 48 abril, pp. 675-695.
- Green, H. A. John (1964): *Aggregation in Economic Analysis: An Introductory Survey*, Princeton University Press, Princeton.
- Gronau, R. (1974): "Wage Comparisons—a Selectivity Bias", *Journal of Political Economy*, vol. 82, noviembre-diciembre, pp. 1119-1143.
- Haavelmo, T. (1943): "The Statistical Implications of a System of Simultaneous Equations", *Econometrica*, vol. 11, enero, pp. 1-12.
- Haberman, S. J. (1974): *The Analysis of Frequency Data*, University of Chicago Press, Chicago.
- Hansen, G. D. (1985): "Indivisible Labor and the Business Cycle", *Journal of Monetary Economics*, vol. 16, noviembre, pp. 309-327.

- Hansen, K.; Heckman, J. J. y Vytlačil, E. J. (2000): "Dynamic Counterfactuals and Dynamic Treatment Effects", manuscrito, Universidad de Chicago, Chicago.
- Hansen, L. P. y Sargent, T. J. (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models", *Journal of Economics, Dynamics and Control*, vol. 2, febrero, pp. 7-46.
- Hansen, L. P. y Sargent, T. J. (1991): "Two Difficulties in Interpreting Vector Autoregressions", en Hansen, L. P. y Sargent, T. J. (eds.), *Rational Expectations Econometrics*, Westview, Boulder, pp. 77-119.
- Hansen, L. P. y Singleton, K. J. (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica*, vol. 50, septiembre, pp. 1269-1286.
- Hausman, J. A. (1980): "The Effect of Wages, Taxes, and Fixed Costs on Women's Labor Force Participation", *Journal of Public Economics*, vol. 14, octubre, pp. 161-194.
- Hausman, J. A. (1985): "Taxes and Labor Supply", Auerbach, A. J. y Feldstein, M. (eds.), *Handbook of Public Economics*, vol. 1, North-Holland, Nueva York, pp. 213-263.
- Heckman, J. J. (1973): "Dummy Endogenous Variables", manuscrito, Universidad de Columbia, Nueva York.
- Heckman, J. J. (1974a): "Effects of Child-Care Programs on Women's Work Effort", *Journal of Political Economy*, vol. 82, n° 2, pt. 2, marzo-abril, s136-s143 (reimpreso en Schultz, Th. W. (ed.), *Economics of the Family: Marriage, Children, and Human Capital*, University of Chicago Press (para el NBER), Chicago).
- Heckman, J. J. (1974b): "Life Cycle Consumption and Labor Supply: An Explanation of the Relationship between Income and Consumption over the Life Cycle", *American Economic Review*, vol. 64, marzo, pp. 188-194.
- Heckman, J. J. (1974c): "Shadow Prices, Market Wages, and Labor Supply", *Econometrica*, vol. 42, julio, pp. 679-694.
- Heckman, J. J. (1976a): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models", *Ann. Econ. and Soc. Measurement*, vol. 5, otoño, pp. 475-492.
- Heckman, J. J. (1976b): "A Life-Cycle Model of Earnings, Learning, and Consumption", *Journal of Political Economy*, vol. 84, n° 4, pt. 2, agosto, s11-s44.
- Heckman, J. J. (1976c): "Simultaneous Equation Models with Continuous and Discrete Endogenous Variables and Structural Shifts", en Goldfeld, S. M. y Quandt, R. E. (eds.), *Studies in Nonlinear Estimation*, Ballinger, Cambridge.

- Heckman, J. J. (1978a): "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, vol. 46, julio, pp. 931-959.
- Heckman, J. J. (1978b): "Longitudinal Studies in Labor Economics: A Methodological Review", trabajo presentado en la *Social Science Research Council Conference*, septiembre, Mt. Kisco, Nueva York.
- Heckman, J. J. (1978c): "A Partial Survey of Recent Research on the Labor Supply of Women", *American Economic Review*, vol. 68, mayo, pp. 200-207.
- Heckman, J. J. (1978d): "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence", *Ann. INSEE*, nºs 30-31, abril-septiembre, pp. 227-269.
- Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, vol. 47, enero, pp. 153-161.
- Heckman, J. J. (1980a): "Addendum to 'Sample Selection Bias as a Specification Error'", en Stromsdorfer, Ernst W. y Farkas, G. (eds.), *Evaluation Studies: Review Annual*, vol. 5, Sage Publications, Beverly Hills, California.
- Heckman, J. J. (1980b): "Sample Selection Bias as a Specification Error with an Application to the Estimation of Labor Supply Functions", en Smith, J.P. (ed.), *Female Labor Supply: Theory and Estimation*, Princeton University Press, Princeton, Nueva Jersey.
- Heckman, J. J. (1981a): "Heterogeneity and State Dependence", en Rosen, S. (ed.), *Studies in Labor Markets*, University Chicago Press (para el NBER), Chicago.
- Heckman, J. J. (1981b): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence", en Manski, C. F. y McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.
- Heckman, J. J. (1981c): "Statistical Models for Discrete Panel Data", en Manski, C. F. y McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.
- Heckman, J.J. (1984): "Comments on the Ashenfelter and Kydland Papers" ("Macroeconomic Analyses and Microeconomic Analyses of Labor Supply" y "Labor-Force Heterogeneity and the Business Cycle"), *Carnegie-Rochester Conf. Ser. Public Policy*, vol. 21, otoño, pp. 209-224.
- Heckman, J. J. (1987): "Self-Selection", en Eatwell, J.; Milgate, M. y Newman, P. (eds.), *The New Palgrave: A Dictionary of Economics*, Stockton, Nueva York.

- Heckman, J. J. (1992): "Randomization and Social Policy Evaluation", en Manski, C. F. y Garfinkel, I. (eds.), *Evaluating Welfare and Training Programs*, Harvard University Press, Cambridge.
- Heckman, J. J. (1993): "What Has Been Learned about Labor Supply in the Past Twenty Years?", *American Economic Review*, vol. 83, pp. 116-121.
- Heckman, J. J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations", *Journal of Human Resources*, vol. 32, verano, pp. 441-462.
- Heckman, J. J. (2000): "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective", *Quarterly Journal of Economics*, vol. 115, febrero, pp. 45-97.
- Heckman, J. J. y Borjas, G. J. (1980): "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence", *Economica*, vol. 47, agosto, pp. 247-283.
- Heckman, J. J. y Honoré, B. E. (1990): "The Empirical Content of the Roy Model", *Econometrica*, vol. 58, septiembre, pp. 1121-1149.
- Heckman, J. J.; Ichimura, H.; Smith, J. y Todd, P. E. (1996): "Sources of Selection Bias in Evaluating Social Programs: An Interpretation of Conventional Measures and Evidence on the Effectiveness of Matching as a Program Evaluation Method", *Proc. Nat. Acad. Sci.*, vol. 93, noviembre, pp. 13416-13420.
- Heckman, J. J.; Ichimura, H.; Smith, J. y Todd, P. E. (1998): "Characterizing Selection Bias Using Experimental Data", *Econometrica*, vol. 66, septiembre, pp. 1017-1098.
- 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 Economic Studies*, vol. 64, octubre, pp. 605-654.
- Heckman, J. J.; Killingsworth, M. R. y MaCurdy, T. E. (1981): "Recent Theoretical and Empirical Studies of Labor Supply: A Partial Survey", en Hornstein, Z.; Grice, J. y Webb, A. (eds.), *The Economics of the Labour Market*, Her Majesty's Stationery Office, Londres.
- Heckman, J. J.; LaLonde, R. J. y Smith, J. (1999): "The Economics and Econometrics of Active Labour Market Programs", en Ashenfelter, O. y Card, D. (eds), *Handbook of Labour Economics*, vol. 3A, North Holland, Amsterdam, pp. 1865-2097.
- Heckman, J. J.; Lochner, L. y Taber, C. (1998a): "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents", *Review of Economics and Dynamics*, nº 1, enero, pp. 1-58.

- Heckman, J. J.; Lochner, L. y Taber, C. (1998b): "General-Equilibrium Treatment Effects: A Study of Tuition Policy", *American Economic Review*, vol. 88, mayo, pp. 381-386.
- Heckman, J. J.; Lochner, L. y Taber, C. (1998c): "Tax Policy and Human-Capital Formation", *American Economic Review*, vol. 88, mayo, pp. 293-297.
- Heckman, J. J.; Lochner, L. y Taber, C. (1999): "General-Equilibrium Cost-Benefit Analysis of Education and Tax Policies", en Ranis, G. y Raut, L. K. (eds.), *Trade, Growth and Development: Essays in Honor of T. N. Srinivasan*, Elsevier Science, Amsterdam, pp. 291-349.
- Heckman, J. J.; Lyons, T. M.; and Todd, P. E. (2000): "Understanding Black-White Wage Differentials, 1960-1990", *American Economic Review*, vol. 90, mayo, pp. 344-349.
- Heckman, J. J. y MaCurdy, T. E. (1980): "A Life Cycle Model of Female Labour Supply", *Review of Economics and Statistics*, vol. 47, enero, pp. 47-74.
- Heckman, J. J. y MaCurdy, T. E. (1981): "New Methods for Estimating Labor Supply Functions", en Ehrenberg, R. G. (eds.), *Research in Labor Economics*, vol. 4, JAI Press, Greenwich.
- Heckman, J. J. y MaCurdy, T. E. (1985): "A Simultaneous Equations Linear Probability Model", *Canadian Journal of Economics*, vol. 18, febrero, pp. 28-37.
- Heckman, J. J. y MaCurdy, T. E. (1986): "Labor Econometrics", en Griliches, Z. y Intriligator, M. D. (eds.), *Handbook of Econometrics*, vol. 3, North-Holland, Amsterdam, pp. 1197-1977.
- Heckman, J. J. y Robb, R., Jr. (1985): "Alternative Methods for Evaluating the Impact of Interventions", en Heckman, J. J. y Singer, B. (eds.), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press (para la Econometric Soc.), Nueva York, pp. 156-245.
- Heckman, J. J. y Robb, R., Jr. (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes", en Wainer, H. (ed.), *Drawing Inferences from Self-Selected Samples*, Springer-Verlag, Nueva York.
- Heckman, J. J. y Sedlacek, G. L. (1985) "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market", *Journal of Political Economy*, vol. 93, diciembre, pp. 1077-1125.
- Heckman, J. J. y Sedlacek, G. L. (1990): "Self-Selection and the Distribution of Hourly Wages", *Journal of Labor Economics*, vol. 8, nº 1, pt. 2, enero, s329-s363.
- Heckman, J. J. y Singer, B. (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, vol. 52, marzo, pp. 271-320.

- Heckman, J. J. y Singer, B. (1985): "Social Science Duration Analysis", en Heckman, J. J. y Singer, B. (eds.), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press (para la Econometric Society), Cambridge, pp. 39-110.
- Heckman, J. J. y Singer, B. (1986): "Econometric Analysis of Longitudinal Data", en Griliches, Z. y Intriligator, M. D. (eds.), *Handbook of Econometrics*, vol. 3, North-Holland, Amsterdam, pp. 1689-1783.
- Heckman, J. J. y Smith, J. A. (1993): "Assessing the Case for Randomized Evaluation of Social Programs", en Jensen, K. y Madsen, P. K. (eds.), *Measuring Labour Market Measures: Evaluating the Effects of Active Labour Market Policy Initiatives*, Ministerio de Trabajo, Copenhagen, pp. 35-95.
- Heckman, J. J. y Smith, J. A. (1995): "Assessing the Case for Social Experiments", *Journal of Economic Perspectives*, vol. 9, primavera, pp. 85-110.
- Heckman, J. J. y Smith, J. A. (1998): "Evaluating the Welfare State", en Strom, S. (ed.), *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, Cambridge University Press (para la Econometric Society), Nueva York, pp. 241-318.
- Heckman, J. J. y Snyder, J. M., Jr. (1997): "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators", *Rand Journal of Economics*, vol. 28, número especial, s142-s189.
- Heckman, J. J.; Tobias, J. y Vytlačil, E. J. (2000): "Simple Estimators for Alternate Treatment Parameters in a Latent Variable Framework with an Application to Estimating the Returns to Schooling", NBER, Working Paper n° 7950, Cambridge.
- Heckman, J. J. y Todd, P. E. (1999): "Understanding Black-White Wage Differentials over the Last Fifty Years", manuscrito, Universidad de Chicago, Chicago.
- Heckman, J. J. y Vytlačil, E. J. (1998): "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return Is Correlated with Schooling", *Journal of Human Resources*, vol. 33, otoño, pp. 974-987.
- Heckman, J. J. y Vytlačil, E. J. (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects", *Proc. Nat. Acad. Sci.*, vol. 96, abril, pp. 4730-4734.
- Heckman, J. J. y Vytlačil, E. J. (2000a): "Local Instrumental Variables", en Hsiao, Ch.; Morimune, K. y Powell, J. (eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, Cambridge University Press, Cambridge.

- Heckman, J. J. y Vytlacil, E. J. (2000b): "The Relationship between Treatment Parameters within a Latent Variable Framework", *Economic Letters*, vol. 66, enero, pp. 33-39.
- Heckman, J. J. y Vytlacil, E. J. (2001a): "Causal Parameters, Treatment Effects and Randomization", manuscrito, Universidad de Chicago, Chicago.
- Heckman, J. J. y Vytlacil, E. J. (2001b): "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effects", en Lechner, M. y Pfeiffer, F. (eds.), *Econometric Evaluations of Active Labor Market Policies in Europe*, Physica Verlag, Nueva York.
- Heckman, J. J. y Vytlacil, E. J. (2001c): "Policy Relevant Treatment Effects", *American Economic Review*, vol. 91, mayo, pp. 107-111.
- Heckman, J. J. y Vytlacil, E. J. (2001d): "Structural Equations, Treatment Effects and Econometric Policy Evaluation", Fisher Schultz Lecture presentada en el encuentro de la Econometric Society, Seattle.
- Heckman, J. J. y Vytlacil, E. J. (2002): "Econometric Evaluation of Social Programs", en Heckman, J. J. y Leamer, E. (eds.), *Handbook of Econometrics*, vol. 6, North-Holland, Amsterdam (en prensa).
- Heckman, J. J. y Willis, R. J. (1973): "Contraception Strategies and Fertility Outcomes: Estimates of a Stochastic Model of Reproduction", en *Papers and Proceedings of the American Statistical Association, Social Statistics Section*, American Stat. Assoc., Washington.
- Heckman, J. J. y Willis, R. J. (1976): "Estimation of a Stochastic Model of Reproduction: An Econometric Approach", en Terleckyj, N. E. (ed.), *Household Production and Consumption*, Studies in Income and Wealth, nº 40 (para el NBER), Columbia Univ. Press, Nueva York.
- Heckman, J. J. y Willis, R. J. (1977): "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women", *Journal of Political Economy*, vol. 85, febrero, pp. 27-58.
- Holland, P. W. (1986): "A Comment on Remarks by Rubin and Hartigan", en Wainer, H. (ed.), *Drawing Inferences from Self-Selected Samples*, Springer-Verlag, Nueva York.
- Holt, D. (1985): Review of *Planning and Analysis of Observational Studies* por Cochran, William E. (eds.), *Journal of American Statistics Association*, vol. 80, pp. 772-773.
- Hong, H. (1998): "Equilibrium and Econometric Models of Asymmetric Auctions", tesis doctoral, Universidad de Stanford.
- Honoré, B. E. y Kyriazidou, E. (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica*, vol. 68, julio, pp. 839-874.

- Horowitz, J. L. y Manski, C. F. (1995): "Identification and Robustness with Contaminated and Corrupted Data", *Econometrica*, vol. 63, marzo, pp. 281-302.
- Hurwicz, L. (1962): "On the Structural Form of Interdependent Systems", en Nagel, E.; Suppes, P. y Tarski, A. (eds.), *Logic, Methodology and Philosophy of Science*, California Stanford University Press, Stanford.
- Ichimura, H. y Thompson, T. S. (1998): "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution", *Journal of Econometrics*, vol. 86, octubre, pp. 269-295.
- Imbens, G. W. y Angrist, J. D. (1994): "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, vol. 62, marzo, pp. 467-475.
- Juhn, C. (1997): "Labor Market Dropouts, Selection Bias, and Trends in Black and White Wages", manuscrito, Universidad de Houston, Houston.
- Keane, M. P. y Wolpin, K. I. (1997): "The Career Decisions of Young Men", *Journal of Political Economy*, vol. 105, junio, pp. 473-522.
- Keane, M. P. y Wolpin, K. I. (1999): "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment", manuscrito, Universidad de Pennsylvania, Philadelphia.
- Killingsworth, M. R. (1983): *Labor Supply*, Cambridge Univ. Press, Cambridge.
- Kirman, A. P. (1992): "Whom or What Does the Representative Individual Represent?", *Journal of Economic Perspectives*, vol. 6, primavera, pp. 117-136.
- Knight, F. H. (1921): *Risk, Uncertainty and Profit*, Houghton Mifflin, Nueva York.
- Laffont, J. J.; Ossard, H. y Vuong, Q. (1995): "Econometrics of First-Price Auctions", *Econometrica*, vol. 63, julio, pp. 953-980.
- LaLonde, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, vol. 76, septiembre, pp. 604-620.
- Lancaster, K. J. (1966): "A New Approach to Consumer Theory", *Journal of Political Economy*, vol. 74, abril, pp. 132-157.
- Lancaster, K. J. (1971): *Consumer Demand: A New Approach*, Columbia University Press, Nueva York.
- Lancaster, T. (1979): "Econometric Methods for the Duration of Unemployment", *Econometrica*, vol. 47, julio, pp. 939-956.
- Lancaster, T. (1990): *The Econometric Analysis of Transition Data*, Cambridge Univ. Press (para la Econometric Soc.), Cambridge.

- Lancaster, T. y Nickell, S. J. (1980): "The Analysis of Re-employment Probabilities for the Unemployed", *Journal of Royal Statistics Society*, ser. A, vol. 143, part. 2, pp. 141-152.
- Layard, R.; Nickell, S. J. y Jackman, R. (1991): *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford University Press, Oxford.
- Lee, D. (2000): "An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice", Working Paper, Universidad de Pennsylvania, Filadelfia.
- Lee, L. F. (1978): "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables", *International Economics Review*, vol. 19, junio, pp. 415-433.
- Lee, L. F. (1982): "Some Approaches to the Correction of Selectivity Bias", *Review of Economic Studies*, vol. 49, julio, pp. 355-372.
- Lewbel, A. (2001): "Two Stage Least Squares Estimation of Endogenous Sample Selection Models", manuscrito, abril, Boston College, Boston.
- Lewis, H. G. (1963): *Unionism and Relative Wages in the United States: An Empirical Inquiry*, University of Chicago Press, Chicago.
- Lewis, H. G. (1974): "Comments on Selectivity Biases in Wage Comparisons", *Journal of Political Economy*, vol. 82, noviembre/diciembre, pp. 1145-1155.
- Lillard, L. A. y Willis, R. J. (1978): "Dynamic Aspects of Earning Mobility", *Econometrica*, vol. 46, septiembre, pp. 985-1012.
- Lucas, R. E. Jr. (1976): "Econometric Policy Evaluation: A Critique", en Brunner, K. y Meltzer, A. H. (eds.), *The Phillips Curve and Labor Markets*, Rochester Conference Series on Public Policy, vol. 1, North-Holland, Amsterdam.
- MaCurdy, T. E. (1978): "Two Essays on the Life Cycle", tesis doctoral, Universidad de Chicago.
- MaCurdy, T. E. (1981): "An Empirical Model of Labor Supply in a Life-Cycle Setting", *Journal of Political Economy*, vol. 89, diciembre, pp. 1059-1185.
- Mallar, C. D. (1977): "The Estimation of Simultaneous Probability Models", *Econometrica*, vol. 45, octubre, pp. 1717-1722.
- Manski, C. F. (1989): "Anatomy of the Selection Problem", *Journal of Human Resources*, vol. 24, verano, pp. 343-360.
- Manski, C. F. (1990): "Nonparametric Bounds on Treatment Effects", *American Economic Review*, vol. 80, mayo, pp. 319-323.

- Manski, C. F. (1994): "The Selection Problem", en Sims, C. A. (ed.), *Advances in Econometrics: Sixth World Congress*, vol. 1, Cambridge University Press, Cambridge.
- Manski, C. F. (1995): *Identification Problems in the Social Sciences*, Harvard Univ. Press, Cambridge.
- Manski, C. F. (1996): "Learning about Treatment Effects from Experiments with Random Assignment of Treatments", *Journal of Human Resources*, vol. 31, pp. 709-733.
- Manski, C. F. y Lerman, S. R. (1977): "The Estimation of Choice Probabilities from Choice Based Samples", *Econometrica*, vol. 45, noviembre, pp. 1977-1988.
- Manski, C. F. y McFadden, D. (1981): "Alternative Estimators and Sample Designs for Discrete Choice Analysis", en Manski, Ch. F. y McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.
- Marschak, J. (1953): "Economic Measurements for Policy and Prediction", en Hood, W. C. y Koopmans, T.C. (eds.), *Studies in Econometric Method*, Cowles Commission for Research in Economics, monografía n° 14, Wiley, Nueva York.
- McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior", en Zarembka, P. (ed.), *Frontiers in Econometrics*, Academic Press, Nueva York.
- McFadden, D. (1981): "Econometric Models of Probabilistic Choice", en Manski, Ch. F. y McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.
- Mincer, J. (1962): "Labor Force Participation of Married Women", en Lewis, H. Gregg (ed.), *Aspects of Labor Economics*, Princeton Univ. Press (para el NBER), Princeton.
- Morgan, M. S. (1990): *The History of Econometric Ideas*, Cambridge University Press, Cambridge.
- Nelson, F. y Olson, L. (1978): "Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables", *International Economic Review*, vol. 19, octubre, pp. 695-709.
- Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators", *Econometrica*, vol. 62, noviembre, pp. 1349-1382.
- Newey, W. K. y McFadden, D. (1994): "Large Sample Estimation and Hypothesis Testing", en Engle, R. F. y McFadden, D. (eds.), *Handbook of Econometrics*, vol. 4, North-Holland, Nueva York.
- Olsen, R. J. (1980): "A Least Squares Correction for Selectivity Bias", *Econometrica*, vol. 48, noviembre, pp. 1815-1820.

- Orcutt, G. H. (1962): "Microanalytic Models of the United States Economy: Need and Development", *American Economic Review*, vol. 52, mayo, pp. 229-240.
- Pakes, A. S. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks", *Econometrica*, vol. 54, julio, pp. 755-784.
- Pessino, C. (1991): "Sequential Migration Theory and Evidence from Peru", *Journal of Development Economics*, vol. 36, julio, pp. 55-87.
- Peterson, A. V. (1976): "Bounds for a Joint Distribution Function with Fixed Sub-distribution Functions: Application to Competing Risks", *Proceedings of National Academic Science*, vol. 73, enero, pp. 11-13.
- Phelps, E. S. (1972): *Inflation Policy and Unemployment Theory: The Cost-Benefit Approach to Monetary Planning*, Norton, Nueva York.
- Quandt, R. E. (1956): "A Probabilistic Theory of Consumer Behavior", *Quarterly Journal of Economics*, vol. 70, noviembre, pp. 507-536.
- Quandt, R. E. (1970): *The Demand for Travel: Theory and Management*, Heath, Lexington.
- Ransom, M. R. (1987): "A Comment on Consumer Demand Systems with Binding Non-negativity Constraints", *Journal of Econometrics*, vol. 34, marzo, pp. 355-359.
- Rao, C. R. (1965): "On Discrete Distributions Arising Out of Methods of Ascertainment", en Ganapati Patil, G. P. (ed.), *Classical and Contagious Discrete Distributions*, Pergamon, Calcuta.
- Rao, C. R. (1985): "Weighted Distributions Arising Out of Methods of Ascertainment: What Population Does a Sample Represent?", en Atkinson, A. C. y Fienberg, S. E. (eds.), *A Celebration of Statistics: The ISI Centenary Volume*, Springer-Verlag, Berlin.
- Robins, J. (1989): "The Analysis of Randomized and Non-randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies", en Freeman, H. E. y Mulley, A. G. (eds.), *Health Services Research Methodology: A Focus on AIDS*, Public Health Service, Washington.
- Robinson, C. (1989): "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models", *Journal of Political Economy*, vol. 97, junio, pp. 639-667.
- Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression", *Econometrica*, vol. 56, julio, pp. 931-954.
- Rogerson, R. (1988): "Indivisible Labor, Lotteries and Equilibrium", *Journal of Monetary Economics*, vol. 21, enero, pp. 3-16.

- Rosenbaum, P. R. (1995): "Sensitivity to Hidden Bias", en *Observational Studies*, Leipzig, Springer-Verlag.
- Rosenbaum, P. R. y Rubin, D. B. (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, vol. 70, abril, pp. 41-55.
- Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, vol. 3, junio, pp. 135-146.
- Rubin, H. (1950): "Note on Random Coefficients", en Koopmans, T. C. (ed.), *Statistical Inference in Dynamic Economic Models*, Cowles Commission for Research in Economics, Monograph n° 10, Wiley, Nueva York.
- Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher", *Econometrica*, vol. 55, septiembre, pp. 999-1033.
- Rust, J. (1996): "Numerical Dynamic Programming in Economics", en Amman, H. M.; Kendrick, D. A. y Rust, J. (eds.), *Handbook of Computational Economics*, vol. 1, Elsevier Science, Nueva York.
- Schmidt, P. (1981): "Constraints on the Parameters in Simultaneous Tobit and Probit Models", en Manski, Ch. F. y McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.
- Sheps, M. C. y Menken, J. A. (1973): *Mathematical Models of Conception and Birth*, University of Chicago Press, Chicago.
- Smith, J. P. y Welch, F. (1986): "Closing the Gap: Forty Years of Economic Progress for Blacks", manuscrito, California Rand Corporation, Santa Monica.
- Smith, J. A. y Todd, P. E. (2000): "Does Matching Overcome LaLonde's Critique of Nonexperimental Evaluation Methods?", manuscrito, University Western Ontario, Londres.
- Smith, J. A. y Todd, P. E. (2001): "Reconciling Conflicting Evidence on the Performance of Propensity Score Matching Methods", *American Economic Review*, vol. 91, mayo, pp. 112-118.
- Stafford, F. (2001): "How the PSID Came to Be: A Life Course and Genealogical Panel of Families", Working Paper, Ann Arbor, Survey Res. Center, Universidad de Michigan.
- Telser, L. G. (1964): "Iterative Estimation of a Set of Linear Regression Equations", *Journal of American Statistical Association*, vol. 59, septiembre, pp. 845-862.
- Theil, H. (1954): *Linear Aggregation of Economic Relation*, North-Holland, Amsterdam.

- Theil, H. (1961): *Economic Forecasts and Policy*, 2ª edición, North-Holland, Amsterdam.
- Theil, H. (1971): *Principles of Econometrics*, Wiley, Nueva York.
- Thurstone, L. (1927): "A Law of Comparative Judgement", *Psychophysical Review*, vol. 34, pp. 273-286.
- Todd, P. E. (1996): "Essays on Empirical Methods for Evaluating the Impact of Policy Interventions in Education and Training", tesis doctoral, Universidad de Chicago.
- Tunali, I. (2000): "Rationality of Migration", *International Economic Review*, vol. 41, noviembre, pp. 893-920.
- Trivedi, P. K. y Baker, G. M. (1983): "Unemployment in Australia: Duration and Recurrent Spells", *Economic Record*, vol. 59, junio, pp. 132-148.
- Vytlačil, E. J. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result", *Econometrica*, vol. 70, nº 1, marzo, pp. 331-341.
- Wainer, H. (ed.) (1986): *Drawing Inferences from Self-Selected Samples*, Springer-Verlag, Nueva York (nueva edición en Mahwah (ed.), N.J., Erlbaum Assoc., 2000).
- Willis, R. J. y Rosen, S. (1979): "Education and Self-Selection", *Journal of Political Economy*, vol. 87, nº 5, pt. 2, octubre, pp. 7-36.
- Wright, P. G. (1928): *The Tariff on Animal and Vegetable Oil*, Macmillan, Nueva York.
- Yitzhaki, S. (1996): "On Using Linear Regressions in Welfare Economics", *Journal of Business and Economic Statistics*, vol. 14, pp. 478-486.

ABSTRACT

This paper summarizes the contributions of microeconometrics to economic knowledge. Four main themes are developed. (1) Microeconometricians developed new tools to respond to econometric problems raised by the analysis of the new sources of micro data produced after the Second World War. (2) Microeconometrics improved on aggregate time-series methods by building models that linked economic models for individuals to data on individual behavior. (3) An important empirical regularity detected by the field is the diversity and heterogeneity of behavior. This heterogeneity has profound consequences for economic theory and for econometric practice. (4) Microeconometrics has contributed substantially to the scientific evaluation of public policy.

Key words: microeconometrics, heterogeneity of behavior, econometric policy evaluation, Nobel Lecture.