

La Real Academia Sueca de Ciencias ha decidido entregar el Premio en Economía del Banco de Suecia en memoria de Alfred Nobel a los profesores James Heckman* y Daniel Mcfadden por sus contribuciones a la Microeconometría**

La investigación micro econométrica trata del análisis empírico del comportamiento económico de agentes individuales y hogares, en lo referente a sus decisiones de oferta de trabajo, consumo, migración, categoría ocupacional que desean ocupar, etc. Los métodos micro econométricos son igualmente relevantes en estudios del comportamiento empresarial, en lo que tiene que ver con sus decisiones de producción o empleo del factor trabajo por ejemplo. En las últimas décadas, avances significativos en la investigación empírica micro econométrica se han logrado gracias a las innovaciones en los métodos micro econométricos y por la mayor disponibilidad de nuevas clases de datos. La materia prima en la investigación micro econométrica son los datos a nivel micro, donde los agentes individuales, los hogares o las empresas, son las unidades de observación. La información a nivel micro aparece en forma de datos de corte transversal, y cada vez más, en la forma de datos de panel.

Mientras el análisis de los datos a nivel micro ha permitido la evaluación de hipótesis económicas y la estimación de modelos económicos, también ha puesto de relieve nuevos problemas econométricos. Esto a su vez, ha inspirado investigación de tipo metodológico en el campo de la micro econometría, la cual puede definirse como una colección de métodos econométricos para manejar problemas de especificación, estimación y evaluación de modelos, que surgen al analizar información a nivel microeconómico. De este modo, el sello de la investigación microeconómica reciente es la cercana interacción entre el trabajo aplicado relacionado con temas económicos de importancia, con trabajo teórico acerca de problemas metodológicos.

* James Heckman, profesor de la Universidad de Chicago.

** Daniel Mcfadden, profesor de la Universidad de California en Berkely.

Esta reseña fue tomada de la información oficial de la Real Academia Sueca de Ciencias acerca del Premio en Economía del Banco de Suecia en memoria de Alfred Nobel, de los profesores James Heckman y Daniel Macfadden. Ha sido publicada con el consentimiento de Lorenette Gozzo, Text Editor Nobel Museum, Suecia, Nobel Foundation, The Royal Swedish Academy of Sciences. Traducción realizada por María Belén Freire, Phd., técnica de la Dirección de Investigaciones Económicas de la Dirección General de Estudios.

Nuevas bases de datos han sido cruciales para el desarrollo de este campo de investigación económica. Hasta fines de los años 60, la disponibilidad de datos para estudios empíricos del comportamiento económico de agentes económicos a nivel individual era muy limitado. Actualmente, hay una serie de bases de datos de tipo longitudinal que cubren a hogares como a agentes individuales, en los Estados Unidos y en la mayoría de países europeos. Un esfuerzo temprano por construir una infraestructura para investigación micro econométrica fue el Estudio de Panel de la Dinámica del Ingreso, establecido por James Morgan y otros en la Universidad de Michigan a fines de los años 60. Dicho estudio se convirtió en un estudio de frontera en el campo de la micro econometría, siendo utilizado extensamente en investigaciones aplicadas y ha servido como un modelo para la construcción de bases de datos de panel en otros países.

Es solo en la época reciente que este crecimiento sustancial en la información microeconómica acerca de individuos y hogares ha experimentado un crecimiento de igual magnitud en la información microeconómica de empresas. Como resultado de esto, las aplicaciones micro económicas han estado mayormente constituidas por estudios del comportamiento de individuos y hogares.

Las aplicaciones micro econométricas cubren un amplio espectro de especialidades económicas. Economistas de la especialidad del trabajo han utilizado técnicas micro econométricas para estudiar decisiones de oferta de trabajo, ingresos individuales, categoría educacional, movilidad del factor trabajo, y la duración de periodos de empleo y desempleo respectivamente. Los métodos micro econométricos son esenciales en estudios empíricos de finanzas públicas, por ejemplo, los efectos de imponer impuestos y políticas de bienestar en la oferta de trabajo; en investigaciones del comportamiento de los consumidores, por ejemplo el escoger entre diferentes marcas; y en economía urbana y del transporte, por ejemplo, el escoger el lugar de residencia y el modo de transporte a utilizarse. La investigación de tipo aplicado en microeconomía y organización industrial se apoya en la micro econometría en lo que tiene que ver con estudios realizados acerca de las decisiones de una empresa respecto a producción y demanda de factores de la producción. Métodos similares también son utilizados por investigadores en otras ciencias sociales.

James Heckman y Daniel McFadden han hecho contribuciones fundamentales a la micro econometría. El trabajo más influyente de Heckman tiene que ver con problemas que surgen cuando los datos son generados por un proceso de selección no aleatorio, un fenómeno común en estudios microeconómicos. Las contribuciones más importantes de McFadden tienen que ver con la teoría y métodos para el análisis de decisiones de tipo discreto, como el escoger una categoría ocupacional o el modo de transporte.

James J. Heckman

James Heckman nació en Chicago IL en 1944. Después de realizar sus estudios de pre- grado en Colorado College, especializándose en matemáticas, estudió economía en la Universidad de Princeton donde obtuvo su Ph.D. en 1971. Heckman ha enseñado en las universidades de Columbia, Yale, y Chicago. Desde 1995 es el Henry Schultz Profesor de Economía de Servicio Distinguido de la Universidad de Chicago.

Las numerosas contribuciones hechas por Heckman a la teoría microeconómica han sido desarrolladas en conjunto con investigaciones empíricas, especialmente en el campo de la economía laboral. Sus trabajos aplicados cubren temas como la oferta de trabajo, ingresos por trabajo, duración del desempleo, evaluación de programas acerca del mercado del trabajo, fertilidad, y discriminación.

El análisis de sesgo de selección hecho por Heckman en el campo de la investigación micro econométrica ha revolucionado las investigaciones de tipo aplicado, tanto en economía como en otras ciencias sociales. El sesgo de selección puede surgir cuando una muestra bajo estudio no representa aleatoriamente a la población subyacente. El problema que enfrenta el analista es el de obtener estimaciones de los parámetros poblacionales relevantes en medio de esta muestra sesgada. Selecciones muestrales no aleatorias pueden ser el resultado de decisiones individuales de los agentes bajo estudio (selección propia), o pueden reflejar reglas administrativas, o decisiones de parte de los estadísticos encargados de constituir la muestra.

Sesgo de selección y selección propia:

Problemas de selección son muy generalizados en investigaciones micro econométricas. Horas trabajadas y salarios se observan solo para aquellos individuos que han escogido trabajar; ingresos de migrantes se observan solo para aquellos que han escogido desplazarse; ingresos de graduados universitarios solo se observan para aquellos que han completado una educación universitaria, etc. El problema de selección puede ser analizado como un problema de falta de observaciones. Salarios y horas de trabajo no pueden ser observadas entre individuos que no trabajan, si es que hubiesen escogido trabajar; de la misma manera, los ingresos de las personas que no migraron, si hubiesen escogido migrar, no son observables para el analista; análogamente, hay una falta de información acerca de los ingresos de los trabajadores con título de bachiller de colegio, si es que hubiesen escogido seguir una carrera universitaria.

Heckman aborda el problema de selección de una forma muy ligada a la teoría económica. Su intuición clave es la de que muchas veces hay observaciones faltantes debido a decisiones conscientes (selección propia) hechas por los agentes económicos (por ejemplo la decisión de trabajar o no, migrar o no, o seguir una carrera universitaria o no). La relación entre las razones para que haya observaciones faltantes y la naturaleza de dichas observaciones toma entonces una intrigante estructura teórica. Las soluciones propuestas por Heckman a los problemas de selección pueden ser apreciadas no solo estadísticamente, pero también en términos de teoría microeconómica.

Las contribuciones de Heckman a la econometría de muestras selectivas surgieron contemporáneamente con sus estudios acerca de la oferta de trabajo a mediados de los años 70. Estos estudios fueron los pioneros de los modelos de oferta de trabajo de la "segunda generación", que se distinguen por estimar ecuaciones derivadas explícitamente de un problema de maximización de la utilidad, siendo los términos de error estocástico una parte integral del modelo, en vez de una simple añadidura. Estos modelos permitieron el análisis unificado de los factores que determinan el número de horas trabajadas, y la participación en la fuerza laboral.

Un importante ejemplo temprano de este tipo de investigación fue la contribución de Heckman (1974), que es notable por su tratamiento del problema de la selectividad inherente en todos los estudios acerca de la oferta de trabajo¹. La teoría económica tradicional ve a la participación en la fuerza laboral como el resultado de un problema de maximización de la utilidad, donde los participantes son individuos cuyos salarios de mercado exceden a sus salarios de reserva. Para obtener estimadores insesgados de los parámetros estructurales básicos, el proceso de estimación tiene que reconocer que la muestra de los participantes en la fuerza laboral no es el resultado de selección aleatoria, sino más bien de la selección propia de los individuos, resultante de la maximización de su utilidad.

Heckman (1974) presentó un modelo de la oferta de trabajo de mujeres casadas, basado en la hipótesis de la maximización de la utilidad. La muestra de mujeres que trabajan adolece del problema de selección propia en el sentido de que las horas de trabajo solo son observables para aquellas mujeres con salarios de mercado mayores a sus salarios de reserva. Heckman derivó una función de máximo verosimilitud para este problema, estimó ecuaciones para los salarios de mercado, la probabilidad de trabajar, y las horas de trabajo, y después uso los parámetros estructurales estimados para predecir la probabilidad de trabajar, las horas de trabajo, los salarios

¹ Ver también Gronau (1974) y Lewis (1974) para importantes discusiones acerca de la selección propia en el contexto de la información acerca de salarios y oferta de trabajo.

de reserva, y los salarios de mercado. Este trabajo es un ejemplo excelente de cómo la teoría microeconómica puede ser combinada con los métodos micro econométricos para clarificar un tema económico de sustancial interés.

El trabajo subsecuente de Heckman ofreció métodos computacionalmente más simples para manejar el problema de selección no aleatoria (Heckman 1976, 1979). El conocido método de corrección de Heckman, también conocido como el método de dos etapas, la lambda de Heckman, o el método Heckit² se han convertido en herramientas típicas del trabajo micro econométrico. El método puede ser descrito por medio de las siguientes dos ecuaciones:

$$w_i = x_{1i}\beta_1 + \varepsilon_{1i}, \quad (1)$$

$$e_i^* = x_{2i}\beta_2 + \varepsilon_{2i} \quad (2)$$

La ecuación (1) describe el salario de mercado del individuo, mientras que (2) es una "ecuación de participación" que describe la propensión del individuo a trabajar. Por ende, w_i es el salario de mercado observado para el individuo i si es que este trabaja, y e_i^* es una variable latente que captura la propensión al trabajo; x_{1i} y x_{2i} son vectores de variables explicativas observadas, tales como edad y educación; finalmente, ε_{1i} y ε_{2i} son errores estocásticos con media cero que representan la influencia de variables no observables que afectan a w_i y e_i^* respectivamente. Los parámetros (vectores) de interés son β_1 y β_2 .

A pesar de que la variable e_i^* no es observable, podemos definir una variable dummy $e_i = 1$ si es que $e_i^* > 0$ y $e_i = 0$ de otra manera; así observamos el salario de mercado solo si es que $e_i = 1$, es decir, si es que el individuo trabaja. Es probable que los términos de error no observables ε_{1i} y ε_{2i} estén correlacionados positivamente; individuos con salarios más altos, dado x_{1i} y x_{2i} , son presumiblemente más propensos a trabajar. Si es que esto es así, la muestra que se tiene de individuos que trabajan no va a representar adecuadamente a la población subyacente, incluso en una muestra grande. El no reconocer esta selectividad generalmente produce estimaciones inconsistentes de los parámetros en la ecuación del salario.

² El nombre Heckit fue aparentemente inventado para enfatizar las similitudes con el famoso estimador Tobit desarrollado por el acreedor del premio Nobel de economía de 1958 James Tobin.

Heckman sugirió un método simple para enfrentar el problema de selección. Nótese que la media condicional de ε_{1i} puede escribirse como:

$$E(\varepsilon_{1i} \mid e_i^* \geq 0) = E(\varepsilon_{1i} \mid \varepsilon_{2i} \geq -x_{2i}\beta_2), \quad (3)$$

y que por lo tanto

$$E(w_i \mid x_{1i}, e_i = 1) = x_{1i}\beta_1 + E(\varepsilon_{1i} \mid \varepsilon_{2i} \geq -x_{2i}\beta_2) \quad (4)$$

Por ende, la ecuación de regresión sobre la muestra seleccionada depende de ambos x_{1i} y x_{2i} . Omitiendo la media condicional de ε_{1i} sesga a los estimadores de β_1 (al menos que ε_{1i} y ε_{2i} no estén correlacionados, en cuyo caso la media condicional de ε_{1i} es cero). El sesgo de selección puede entonces considerarse como un problema típico de variable omitida. El problema es encontrar una representación empírica de la media condicional ε_{1i} e incluir esta variable en la ecuación del salario.

Bajo el supuesto de que ε_{1i} y ε_{2i} son tomadas de una distribución normal bivariada, podemos derivar la siguiente ecuación de regresión:

$$E(w_i \mid x_{1i}, e_i = 1) = x_{1i}\beta_1 + \rho\sigma_1\lambda_i \quad (5)$$

En la ecuación (5) ρ es el coeficiente de correlación entre ε_{1i} y ε_{2i} ; σ_1 es la desviación standard de ε_{1i} , y λ_i —la inversa del ratio de Mills— es igual a:

$$\lambda_i = \frac{\phi(x_{2i}\beta_2 / \sigma_2)}{\Phi(x_{2i}\beta_2 / \sigma_2)} \quad (6)$$

donde ϕ y Φ son las funciones de densidad y distribución acumulativa de una distribución normal standard y σ_2 es la desviación estándar de ε_{2i} .

Heckman mostró como estimar (5) en un proceso de dos etapas. La primera etapa involucra estimar los parámetros de (2) por el método probit, usando la muestra completa. Estas estimaciones pueden después utilizarse para computar λ_i

para cada individuo de la muestra. Una vez que λ_i está estimada, podemos estimar (5) sobre la muestra de individuos que están trabajando por medio de mínimos cuadrados ordinarios, tratando a $\rho\sigma_1$ como el coeficiente de regresión de λ_i .

El signo del sesgo de selección depende de la correlación entre los errores de las ecuaciones del salario y la participación laboral (ρ) y la correlación entre λ_i y las variables en la ecuación del salario (x_{1i}). Ya que λ_i es una función decreciente de la probabilidad de selectividad en la muestra, el coeficiente β de las variables x_{1i} que probablemente incrementan tanto a los salarios como a la participación laboral, tales como la educación, estará sesgado hacia abajo si es que la técnica Heckit no es utilizada (dado que ρ es > 0).

El trabajo de Heckman en esta área ha generado muchas aplicaciones empíricas en economía así como en otras ciencias sociales y en estadística aplicada acerca de datos faltantes no aleatorios. Un ejemplo temprano de aplicaciones empíricas es el trabajo de Lee (1978), que examinó los efectos sobre los salarios relativos de la participación sindical en los Estados Unidos, reconociendo que el pertenecer o no a un sindicato no es una cuestión aleatoria pero más bien el producto de la selección propia. Otro es el trabajo de Willis y Rosen (1979), que investigaron los premios salariales asociados a la educación superior, reconociendo que la decisión en materia educacional es endógena y depende de las ganancias percibidas a la educación. Posteriormente, Heckman y Guilherme Sedlacek (1985) presentaron un modelo empírico de la alocaación sectorial de trabajadores en el mercado laboral de los Estados Unidos, siguiendo la línea de Roy (1951) en un modelo de distribución del ingreso: individuos que maximizan su utilidad pueden trabajar en varios sectores, pero en un solo sector al mismo tiempo. El análisis incluye un análisis de como la selección propia impacta a la desigualdad en términos de salarios.

El trabajo de Heckman también ha generado una vasta literatura acerca de métodos econométricos. El modelo original ha sido extendido en un número de vías por Heckman y otros³. Estos esfuerzos, típicamente dirigidos hacia eliminar el supuesto restrictivo de la normalidad bivariada, han involucrado el uso de métodos semi-paramétricos; véase por ejemplo, Heckman y Robb (1985b), Heckman (1990), Mansk (1989), Newey, Powell and Walker (1990), y Lee (1994 a, b).

³ Véase Vella (1998) para un recuento de la literatura econométrica que siguió al trabajo pionero de Heckman.

Evaluación de programas activos del mercado de trabajo

Programas activos del mercado de trabajo, tales como entrenamiento, asistencia en la búsqueda de empleo, y subsidios al empleo, se han ido generalizando de modo creciente en la mayoría de países de la OECD. Dichos programas están generalmente dirigidos hacia individuos que están desempleados o que tienen bajo nivel de ingresos o habilidades.

El problema clásico de la evaluación de un programa es determinar como la participación en un programa específico afecta los resultados individuales, tales como ingresos o empleo, comparado a la no participación en dicho programa. La dificultad consiste en caracterizar a la situación que no se dio en la realidad, esta es, responder la pregunta: ¿Qué hubiese sucedido si es que el individuo no hubiese participado en el programa?. Ya que es imposible observar al individuo como participante y no participante, es necesario utilizar información de los resultados de los no participantes para este propósito. Dado que la alocaón de los individuos a los programas no es aleatoria por lo general, el grupo de participantes se convierte en una muestra seleccionada con características observadas y otras no observadas que pueden diferir de las de la población entera.

James Heckman es el investigador más importante del mundo en el tema de la evaluación econométrica de políticas económicas. En esta área, así como en sus estudios acerca de la oferta de trabajo, Heckman se basa en un enfoque estructural basado en la teoría microeconómica para guiar la especificación del modelo y para interpretar los resultados empíricos. Los ingredientes esenciales de la evaluación de políticas es doble: (i) un modelo de participación en programas y (ii) un modelo de resultados de los programas. La investigación de Heckman acerca de evaluación de programas puede ser vista como una continuación natural a su trabajo previo acerca de modelos de selección. En trabajo conjunto con otros, Heckman ha presentado una serie de nuevos resultados concernientes a la identificación y la estimación de los efectos de programas sociales (Heckman y Robb 1985 a, b) y, más recientemente, ha acrecentado nuestro entendimiento acerca de los pros y contras de usar datos experimentales versus datos no experimentales en la evaluación de programas (Heckman y Smith, 1995; Heckman, Ichimura y Todd, 1997); Heckman, Smith y Clements, 1997). Este último grupo de estudios también ha incentivado avances teóricos en el uso de métodos de "matching" como estimadores de evaluación econométrica. En general, Heckman no emerge como un gran simpatizante del enfoque experimental, arguyendo que los experimentos sociales son válidos solo bajo ciertos supuestos de tipo estadístico y también supuestos particulares acerca del comportamiento de los individuos. Una lección central es que no hay métodos universalmente válidos para evaluar programas. Que método funciona mejor depende del tema que se esté tratando como de los modelos económicos que determinan la participación y los resultados. Heckman también ha

presentado importantes resultados empíricos acerca de los efectos de varios programas del mercado laboral. Las conclusiones acerca de dichos efectos son en una gran parte de casos algo pesimistas. Un recuento enfocado en las investigaciones hechas para el caso de los Estados Unidos (Heckman, LaLonde y Smith, 1999) concluye que los programas frecuentemente tienen efectos positivos muy pequeños y a veces efectos negativos para los participantes y no pasan pruebas convencionales de costo beneficio. Por otro lado, si parece haber una gran heterogeneidad en los resultados de los programas entre participantes y entre tipos de programas.

Modelos de duración

El análisis de datos de duración tiene una larga tradición en ingeniería y en investigación biomédica. Más recientemente, se ha adentrado en la investigación económica y en otro tipo de ciencias sociales, donde los modelos de duración han sido aplicados a una variedad de problemas. Los modelos de duración son ahora herramientas típicas cuando se estudia lo que dura un período de desempleo, eventos demográficos (matrimonio, fertilidad, mortalidad y migración), eventos políticos (por ejemplo, como la ocurrencia de crisis gubernamentales depende del tiempo transcurrido desde la última elección), y algunas relaciones industriales (por ejemplo la duración de huelgas). Estos modelos también se usan en investigación acerca de los consumidores, para estudiar el patrón de tiempo de la adquisición de productos, así como en investigación macroeconómica para examinar temas como la duración de los ciclos económicos.

En su trabajo acerca del análisis econométrico de la duración, Heckman ha estado particularmente preocupado de los efectos de la heterogeneidad no observada, es decir, diferencias individuales en variables no observadas que pueden influenciar la duración del desempleo o del empleo. Como la heterogeneidad no observada en el contexto de datos de duración introduce problemas específicos de selección, el trabajo de Heckman en esta área se coordina excelentemente con su agenda de investigación global acerca de la selectividad en las muestras.

Esto puede ejemplificarse por los estudios acerca de cómo la tasa de salida del desempleo hacia una situación de empleo evoluciona durante un período de desempleo. Un problema aquí es que los individuos con prospectos relativamente bajos de encontrar un empleo parecen “sobrevivir” como desempleados hasta un grado más alto que los individuos con características más favorables. Por ende, la “calidad” del stock de desempleados en cada punto del tiempo es el resultado de un proceso de selección en parte impulsado por factores no observables por el analista. La sobre representación de individuos que están más propensos a ser desempleados por períodos de tiempo más largos puede fácilmente llevarnos a la conclusión de una dependencia de duración negativa, es decir, que la tasa de salida al empleo

declina durante un período de desempleo. Sin embargo, lo que aparece como dependencia negativa de la duración puede ser simplemente un efecto de selección.

En trabajo conjunto con Burton Singer, Heckman trata el problema de tratar la heterogeneidad no observada sin imponer supuestos restrictivos acerca de la distribución de las variables que no se observan. Heckman y Singer (1984³) proponen un estimador no paramétrico que actualmente se utiliza mucho en trabajos aplicados en economía y en demografía.

Otras contribuciones sustanciales de Heckman a la literatura de la duración incluyen resultados de identificación para una clase de modelos de duración (Heckman y Singer, 1984b y Heckman y Honoré, 1989) y el tratamiento más avanzado de temas relacionados a la identificación (Heckman, 1991 y Heckman y Taber, 1994). Heckman también ha escrito trabajos empíricos acerca de la duración del desempleo y la fertilidad.

Daniel L. McFadden

Daniel L. McFadden nació en Raleigh, NC en 1937. Obtuvo su título de pregrado de la Universidad de Minnesota, con especialidad en física. McFadden recibió su Ph.D. en economía de la Universidad de Minnesota en 1962. Sus trabajos académicos incluyen plazas de profesor en las universidades de Pittsburg, Yale, MIT, y la Universidad de California en Berkeley. Desde 1990 es el E. Morris Cox Professor de economía en Berkeley.

La contribución más importante e influyente de McFadden ha sido el desarrollo de la teoría económica y la metodología econométrica para las elecciones de tipo discreto, es decir, la elección entre un grupo finito de alternativas. También ha hecho contribuciones significativas en otros campos de la economía, incluyendo teoría de la producción y economía ambiental. El trabajo de McFadden puede caracterizarse mejor por su habilidad para combinar el desarrollo de la teoría y de la metodología con aplicaciones a problemas empíricos sustanciales.

Análisis de elecciones discretas

Problemas de elección discreta aparecen frecuentemente en economía como en otras ciencias sociales. Considere por ejemplo, el modelar un fenómeno tal como la participación individual en la fuerza laboral, decisiones de tipo ocupacional o de lugar de asentamiento, o acerca del modo de transporte a utilizarse. En todos estos casos las observaciones a ser explicadas son discretas (o cualitativas) y no pueden ser representadas por variables continuas. La teoría tradicional de la demanda así

como métodos econométricos tradicionales, dirigidos a explicar variaciones en variables continuas, son generalmente no apropiadas para analizar el comportamiento referente a decisiones de tipo discreto.

Problemas de elección cualitativa fueron inicialmente tratados en trabajos de sicometría, biometría y en cierto grado, también de econometría. Contribuciones tempranas de particular importancia fueron Thurstone (1927) y Luce (1959), que formularon modelos de elección discreta probabilística. De acuerdo a la interpretación psicológica, el comportamiento de elección individual es intrínsecamente probabilístico.

Al contrario, el enfoque económico y econométrico desarrollado por McFadden trata las decisiones individuales como deterministas. Se enfoca en la falta de información del analista, tal como la información imperfecta acerca de las características de las alternativas y de los individuos bajo estudio. Mientras los psicólogos han estado generalmente preocupados de las elecciones individuales per se, los economistas generalmente han estado más interesados en resultados agregados, tal como la fracción de la población que elige una alternativa particular⁴.

El modelo condicional logit

La contribución más importante de McFadden es la integración de la teoría económica con metodología econométrica para el análisis de elecciones discretas. Su trabajo fundamental, titulado "Análisis Condicional Logit del Comportamiento de Elecciones Cualitativas" (McFadden, 1974⁵), y casos de estudio contemporáneos de tipo empírico cambiaron de una manera fundamental la forma en la que los analistas piensan acerca del análisis econométrico del comportamiento individual. El análisis de elecciones de tipo discreto se desarrolló rápidamente hasta convertirse en uno de los campos predominantes de la econometría moderna.

El enfoque de McFadden puede resumirse de la siguiente manera. Supongamos que cada individuo en una población tiene un rango finito de alternativas y escoge una alternativa que maximice su utilidad. Los datos disponibles al analista se asumen son generados por sacar observaciones aleatorias de un individuo de la población y anotando un vector a de los atributos del individuo (tales como edad, sexo, ingreso, etc.), el rango Γ de alternativas disponibles al individuo (por ejemplo, viajar en bus, carro, metro, etc.), la elección i hecha por el individuo del rango de alternativas. Asumimos que la utilidad del individuo por escoger i es de forma aditiva tal que $u(i, a) = v(i, a) + e(i, a, w)$ donde $v(i, a)$ es común a todos los

⁴ Ver Ben-Akiva y Lerman (1985) y Anderson, de Palma y Thisse (1992) para recuentos y discusiones acerca de modelos de elección discreta.

individuos con atributos observables a , mientras que $e(i, a, w)$ es particular al individuo w . Ambos términos que contribuyen a la utilidad son determinísticos, el primero refleja gustos “representativos” de la población y el segundo refleja variaciones idiosincrásicas en el gusto. Tratando el término de la utilidad nó observada $e(i, a, w)$ como realizaciones de variables aleatorias $\varepsilon(i, a)$, y dejando que $P(i | a, \Gamma)$ denote la probabilidad de elección condicional de que el individuo aleatoriamente seleccionado escoja la alternativa i elemento de Γ , dados sus atributos observables a y el rango de alternativas Γ , se obtiene:

$$P(i | a, \Gamma) = \Pr[v(i, a) + \varepsilon(i, a) \geq v(j, a) + \varepsilon(j, a) \forall j \in \Gamma] \quad (7)$$

Esto se conoce como el modelo aditivo de utilidad aleatoria, de las elecciones discretas. El lado derecho de (7) es la probabilidad de que un individuo seleccionado aleatoriamente de una población tenga una función de utilidad que haga que i sea la elección que maximice la utilidad, dados los atributos individuales a y el rango de elección Γ .

Si es que el vector aleatorio $(\varepsilon(i, a))_{(i \in \Gamma)}$ tiene una función de distribución acumulativa conjunta F , y si escribimos $\Gamma = \{1, \dots, I\}$, el lado derecho de la ecuación (7) puede escribirse como una integral en la derivada parcial i , F_i , de F :

$$P(i | a, \Gamma) = \int_{-\infty}^{\infty} F_i[x + v(i, a) - v(1, a), \dots, x + v(i, a) - v(I, a)] dx. \quad (8)$$

En particular, si es que las variables aleatorias $\varepsilon(i, a)$ se distribuyen independientemente con una función de distribución acumulativa $\exp[-e^{-\sigma x}]$ para $\sigma > 0$ (la distribución de valor extremo Tipo I, o Gumbel), su distribución conjunta F es:

$$F(x_1, \dots, x_I) = \exp\left[-\sum_{j \in \Gamma} e^{-\sigma x_j}\right] \quad (9)$$

y las probabilidades de elección de la ecuación (7) se reducen a la forma analíticamente conveniente logit:

$$P(i | a, \Gamma) = \frac{\exp \sigma v(i, a)}{\sum_{j \in \Gamma} \exp \sigma v(j, a)} \quad (10)$$

El parámetro $\sigma > 0$ es inversamente proporcional a la desviación estándar de los términos de utilidad rándomica $\varepsilon(i, a)$. En el límite cuando $\sigma \rightarrow \infty$, las probabilidades de elección $P(i, a, \Gamma)$ en (10) asignan toda la masa de probabilidad a las alternativas con máxima utilidad “representativa” $v(i, a)$, y obtenemos el modelo microeconómico tradicional, determinísticos de maximización de la utilidad.

Para hacer que el modelo logit resultante sea tractable debido a objetivos de predicción, usualmente se asume que los términos de utilidad representativa $v(i, a)$ dependen de características conocidas de las alternativas y de la población en alguna forma analíticamente tractable. Por ejemplo, en el caso de la elección del modo de transporte, estas características podrían ser tiempo de viaje, costos del viaje, etc. Los vectores de parámetros asociados pueden después ser estimados por el método de máximo verosimilitud.

McFadden llamo a su innovación el modelo condicional logit. A pesar de que modelos multinomiales logit han existido por algún tiempo (Theil, 1969, Quandt, 1970), la derivación del modelo por parte de McFadden, basada en una teoría económica del comportamiento de elección de la población, fue completamente nueva. Su contribución fue inmediatamente reconocida como un trabajo paradigmático, y abrió el camino para estimación estadística y aplicaciones.

Desarrollo subsecuente del análisis de las elecciones discretas

El atractivo del modelo multinomial logit recae en su combinación de sólidos fundamentos microeconómicos y su simplicidad computacional. Dicha simplicidad se deriva del supuesto de independencia estadística de los términos de utilidad aleatoria, un supuesto que implica independencia de alternativas irrelevantes (IIA). La relación entre las probabilidades de escoger entre cualesquier dos alternativas es independiente de las propiedades de todas las demás alternativas, como puede notarse al utilizar (10) y construir la relación entre cualesquiera dos alternativas. Por ejemplo, un rango de opciones extenso no afecta la probabilidad de escoger entre dos opciones. Bajo el supuesto de independencia de alternativas irrelevantes, los parámetros relevantes a dos alternativas pueden ser consistentemente estimados al utilizar datos solo sobre esos individuos que han escogido esas dos alternativas. Así, la propiedad IIA permite la estimación del modelo multinomial logit sobre la base de muestras basadas en la elección, que se obtienen con mucha mayor facilidad que muestras poblacionales, con la compensación respectiva por el surgimiento del problema del sesgo muestral; véase Manski y Lerman (1977) y Manski y McFadden (1981).

Sin embargo, como McFadden ha enfatizado, el supuesto de IIA es restrictivo en muchas aplicaciones. Por ejemplo, no es probable que el ratio de probabilidades de cualesquier dos alternativas sea invariante a la introducción de una nueva alternativa que sea un sustituto cercano a la alternativa existente⁵. Este problema ejemplifica la dificultad más general de estimar parámetros poblacionales de muestras con selección propia. Hausman y McFadden (1984) desarrollaron un procedimiento para probar la validez del supuesto de IIA basados en la idea de comparar estimaciones de una muestra con selección propia y una muestra con el rango de elección completo; las dos estimaciones no deberían cambiar sustancialmente si es que el supuesto de IIA es válido. Otras pruebas de especificación fueron desarrolladas en McFadden (1987).

McFadden también ha mostrado como relajar el supuesto de IIA a través de su desarrollo de los modelos nested logit multinomial modelos de valor extremo generalizados (GEV). El modelo nested logit, introducido por Ben-Akiva (1973) y McFadden (1978), relaja el supuesto de IIA permitiendo cierta dependencia estadística entre las alternativas. En este modelo las decisiones individuales pueden ser interpretadas como teniendo una estructura jerárquica. Consideremos por ejemplo, la elección conjunta de lugar de destino y modo de transporte. Una formulación posible de un modelo logit nested para este problema de decisión podría consistir en asumir que para cada destino el individuo selecciona el modo preferido de transporte, y, tomando esto en cuenta, escoge su destino.

El modelo GEV, desarrollado por McFadden (1978, 1981), es más general y analíticamente elegante. Para derivarlo, nótese que si generalizamos la ecuación (9) a:

$$F(x_1, \dots, x_I) = \exp[-G(e^{-\alpha_1}, \dots, e^{-\alpha_I})] \quad (11)$$

para alguna función linealmente homogénea G , la ecuación (7) se convierte en

$$P(i | a, \Gamma) = \frac{G_i(e^{-\sigma v(i,a)}, \dots, e^{-\sigma v(I,a)}) \exp \sigma v(i,a)}{G(e^{-\sigma v(1,a)}, \dots, e^{-\sigma v(I,a)})} \quad (12)$$

donde G_i es la derivada parcial de G con respecto al elemento i ésimo. Utilizando la fórmula de Euler, el denominador puede escribirse como la suma sobre las derivadas parciales de G , y obtenemos:

⁵ Gerard Debreu (1960), que obtuvo el premio Nobel de economía en 1983, fue el primero que notó esto, en el contexto del modelo de elección probabilística de Luce (1959)

$$P(i | a, \Gamma) = \frac{\exp \sigma v(i, a) + \ln G_i(e^{-\sigma v(i, a)}, \dots, e^{-\sigma v(i, a)})}{\sum_j \exp[\sigma v(j, a) + \ln G_j(e^{-\sigma v(j, a)}, \dots, e^{-\sigma v(j, a)})]} \quad (13)$$

En otras palabras, las probabilidades de elección aun tienen una forma logit. Sin embargo, ahora la probabilidad de elección de una alternativa i depende no solo de sus propios atributos observables, vía $v(i, a)$, sino también de los atributos observados de otras alternativas, tal que IIA no tiene que mantenerse. El modelo logit tradicional se obtiene como un caso especial cuando la función G es la suma de sus argumentos (entonces todas sus derivadas parciales son iguales a la unidad, y el segundo terminó en el numerador y el denominador de (13) desaparece), y el modelo nested logit se obtiene como un caso especial cuando G es una función CES.

Otra generalización útil del modelo logit multinomial es el llamado *modelo logit mixto*. Esta generalización se obtiene agregando comportamiento de elección sobre diversas subpoblaciones, donde todos los individuos en cada subpoblación tienen los mismos atributos observables, y el comportamiento de elección de cada subpoblación se modela como se delineó mas arriba. Como se demuestra en McFadden y Train (1998), cualesquier modelo bien comportado de utilidad aleatoria puede ser aproximado con cualquier grado de perfección, a un modelo de logit mixto. Aplicaciones de este modelo usualmente requieren métodos de simulación de Monte Carlo.

Otro enfoque es el *modelo multinomial probit* con errores correlacionados. Sin embargo, como con el modelo logit mixto, las dificultades computacionales surgen cuando este modelo es aplicado a problemas con mas de unas pocas alternativas, básicamente porque el cálculo de las probabilidades de elección involucran la evaluación de integrales múltiples. Lerman y Manski (1981) introdujeron la idea de calcular las probabilidades de elección por medio de métodos de simulación de Monte Carlo donde se escogen observaciones aleatoriamente de una distribución normal multivariada. McFadden (1989) desarrollo esta idea aun más al proponer un método de estimación conocido como el *método de los momentos simulados*. El artículo de McFadden desarrollo las propiedades estadísticas básicas de este método y una vasta literatura ha surgido desde entonces en esta área.

Un paso natural en la evolución del análisis de las elecciones individuales es el desarrollo de modelos y métodos que expliquen elecciones discretas y continuas al mismo tiempo. El artículo de Dubin y McFadden (1984) es notable como un buen ejemplo de cómo integrar una contribución metodológica general con estudio empírico de utilidad práctica (la elección de los hogares entre aparatos electrónicos es de tipo discreto y su elección de consumo eléctrico es de tipo continuo).

Como ya se enfatizó, la mayor parte del trabajo de McFadden está caracterizado por una íntima relación entre la teoría económica, la metodología econométrica y los estudios empíricos aplicados. Aplicaciones empíricas tempranas a la demanda urbana de transporte se encuentran en McFadden (1974b) y Domencich y McFadden (1975). Durante los ochentas y noventas, McFadden estuvo involucrado en trabajo empírico relativo a la demanda de energía residencial (Cowing y McFadden, 1984), la demanda de servicios telefónicos (McFadden, Train y Ben-Akiva, 1987) y la demanda por residencias entre los ancianos (McFadden, 1994^a).

Otras contribuciones

McFadden ha hecho importantes contribuciones en varios otros campos. En los inicios de los 60 y los 70, trabajó intensamente en el análisis teórico y empírico de la producción. La mayor parte de estas contribuciones no fueron publicadas hasta que aparecieron en una colección de dos volúmenes de artículos editada por Fuss y McFadden (1978). El trabajo de McFadden en esta área fue muy influyente y estableció el principio de la dualidad entre las funciones de costos y de ganancias y de producción como una herramienta básica para el análisis empírico de la producción.

En otros trabajos importantes, en colaboración con Peter Diamond, McFadden ha explotado la dualidad entre la función de gasto y las funciones de demanda que maximizan la utilidad para explorar problemas tales como el peso muerto de los impuestos y los impuestos óptimos sobre bienes (Diamond y McFadden, 1974). Este artículo estableció la dualidad como una herramienta indispensable para ser utilizada en el campo de la economía pública moderna.

En los años noventa, McFadden ha contribuido a la economía ambiental, estudiando la voluntad para pagar por los recursos naturales. McFadden (1994b) examina en detalle las propiedades del método de valuación contingente para estimar el llamado valor de existencia de los recursos naturales y desarrolla nuevas técnicas econométricas. Con Jerry Hausman y Gregory Leonard, el desarrollo un modelo de elección discreta empírico para evaluar las pérdidas de bienestar derivadas del daño a los recursos naturales (Hausman, Leonard y McFadden, 1995). El modelo fue aplicado a la demanda recreacional en Alaska y, en particular, para estimar la pérdida de bienestar de los Alaskianos causada por el derrame de petróleo del tanquero Exxon Valdez en 1989. El trabajo de McFadden en este campo es otro ejemplo de su habilidad magistral para integrar la teoría económica, la metodología econométrica y las aplicaciones empíricas de gran interés.