

Reflexiones sobre el trabajo de James Heckman

JENNIFER WARD-BATTS¹

Las pasadas 2 o 3 décadas, [el recientemente nominado premio Nobel de economía], James Heckman, ha hecho significativas contribuciones a la literatura econométrica, principalmente en lo relacionado con la evaluación, medición e impactos de los programas sociales. Gran parte de su trabajo ha resaltado el problema del sesgo en la selección muestral (*selection bias*). Este es un aspecto importante para muchas investigaciones empíricas, particularmente aquellas relacionadas con la evaluación de programas sociales. Por tal razón se ha visto la utilidad de contar con un conjunto de guías que permitan realizar el proceso de identificación de los modelos de selección; concepto al que se le ha dado poco tratamiento en la literatura econométrica.

Heckman y Rob, proveen un detallado conjunto de directrices que apuntan en esta dirección, en su artículo escrito en 1985², y un más accesible pero menos detallado tratamiento en su artículo de 1986³. En estos trabajos, señalan que los datos longitudinales (datos de panel) no son la panacea como a menudo se ha creído que son; primero porque la utilidad de los datos repetidos de sección cruzada ha sido subestimada; y segundo porque los datos de panel a menudo tienen observaciones basadas en un número pequeño de unidades, en relación con los datos de sección cruzada. Esto significa que, en la práctica, es posible obtener estimaciones más eficientes con este último tipo de observaciones (datos de sección cruzada), que con el primero (datos de panel).

El objetivo de estos artículos es definir algunas líneas guías, sujetas a un número mínimo de restricciones requeridas para identificar estimadores consistentes,

¹ Egresada del *Graduate Program In Sociology At The University Of Washington*. Artículo escrito en 1997 (Primavera). Email: jwbatts@u.washington.edu. Traducción hecha por Marco Baquero L., publicación realizada con la autorización de su autora.

² Heckman James and Richard Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions", In *Longitudinal Analysis of Labor Market Data*, Heckman, James J. And Burton Singer, eds. *Econometric Society Monographs series*, n.10, Cambridge University Press, pages. 156-245.

³ Heckman James and Richard Robb (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes", *Economics Research Center/NORC Discussion Paper*, March 1986.

utilizando varios tipos de datos (sección cruzada, repeticiones de datos de sección cruzada y datos longitudinales).

Heckman ha trabajado en varios campos, en donde el problema del sesgo de selección es un tópico al que le ha dado especial importancia. Su trabajo incluye tópicos como: oferta laboral de las mujeres casadas, programas de cuidado infantil, salario mínimo y la legislación anti discriminatoria, etc.

El objetivo de este artículo es triple: (i) mostrar la intuición básica que está detrás del problema de la selección muestral, (ii) porqué es importante corregirla y (iii) cuáles son los medios para hacerlo.

El problema de la selección muestral

La intuición detrás del problema es simple. Usaré un ejemplo utilizado por Roy (1951) y re-utilizado nuevamente en Madala (1983). Suponga que la gente debe elegir entre dos ocupaciones: ser cazador o pescador. Una persona escogerá la ocupación donde sea más productivo. Suponga que Y_1 e Y_2 representan el producto del cazador y pescador respectivamente. Una persona escogerá ser un cazador si $Y_1 > Y_2$.

Si observamos el producto medio de los cazadores, lo que observamos realmente, es el producto medio, dada la elección de ser cazador (es decir, dado que $Y_1 > Y_2$). Pero, si toda la población fuese forzada a elegir ser cazadores, el producto medio poblacional sería en general, diferente al producto medio observado para el grupo de personas que eligieron la opción de cazadores. Existen entonces tres posibilidades observables:

1. Los que escogen ser cazadores tienen una producción mayor al “cazador promedio” y aquellos que deciden ser pescadores tienen una producción mayor al “pescador promedio”.
2. Aquellos que escogen ser cazadores tienen mayor producción que el cazador y pescador promedios pero, son mejores cazando que pescando. De la misma forma, aquellos que eligen ser pescadores tienen una producción menor que el cazador y pescador promedios pero, de todas formas están mejor pescando que cazando.
3. La opción 3 corresponde a la opción 2, expresada en sentido contrario.

El cuarto caso no es posible. Esto es, nunca observaremos una situación en donde los cazadores están mejor que el pescador promedio y, donde el pescador esté mejor que el cazador promedio.

Por qué importa el problema de la selección muestral?

El ejemplo anterior, sin embargo, no es útil para explicar porqué es importante el problema de la selección muestral, o cómo este podría afectar a la evaluación de un programa social.

El sesgo de selección puede ser importante para la evaluación de programas sociales cuando los beneficiarios del programa pueden estar simultáneamente en dos “estados”. Por ejemplo, considere el problema de una madre soltera que debe decidir sobre si aplicar o no a un seguro de desempleo. Ella cuantificará las dos opciones (participar o no participar) y escogerá aquella opción en donde obtenga el mayor beneficio. Algunas mujeres, por varias razones, elegirán no inscribirse en el seguro de desempleo. Si, adicionalmente al seguro de desempleo, el Gobierno crea un programa de entrenamiento laboral (donde también pueden participar las mujeres inscritas en el programa de seguro de desempleo), las mujeres que en un principio decidieron no aplicar al seguro de desempleo, podrían ahora decidir enrolarse, ya que con el programa de entrenamiento laboral los beneficios netos cambiaron, y ahora conviene inscribirse en el programa.

Al considerar los efectos de la aplicación de un programa de entrenamiento laboral, no debe omitirse el problema de la selección muestral. De igual manera, cuando evaluamos el éxito de un programa social debemos considerar que muchas de las aplicantes al programa de entrenamiento laboral no tienen el mismo perfil que las inscritas al programa de seguro de desempleo. Por el contrario, las primeras tendrían características que les permitiría tener mejores oportunidades de éxito respecto a las aplicantes al seguro de desempleo.

Por lo tanto, al comparar los resultados “después” de aplicado el programa de entrenamiento laboral, respecto de los resultados “antes” del programa; lo que se está comparando en esencia son: naranjas de “antes” (madres que no optaron por el seguro de desempleo), con manzanas y naranjas “después” (madres con seguro de desempleo y madres sin seguro de desempleo). Cuando lo que se quiere realmente comparar son naranjas con naranjas; y para lo cual las técnicas de corrección de selección muestral, nos permiten separar las manzanas.

El ejemplo anteriormente utilizado es quizás un poco más complejo, que el utilizado por Heckman y Rob, pero permite responder otras preguntas. Específicamente, en Heckman y Rob (1985 y 1986) se plantean dos preguntas que interesan responder:

1. Cuál es el efecto “promedio” del programa de entrenamiento laboral sobre los ingresos de los participantes, si se supone que los participantes fueron elegidos aleatoriamente para participar en dicho programa?
2. Considerando únicamente al grupo de gente que ha sido seleccionada para participar en dicho programa, cuál es el efecto “promedio” del programa sobre sus ingresos? En otras palabras, cómo son afectados los ingresos del grupo entrenado, respecto a los ingresos que habrían tenido, en caso de no participar en dicho programa de entrenamiento?

Heckman y Rob señalan que la última pregunta es la más interesante, debido a su relevancia para aplicaciones de política.

Como corregir el problema del sesgo de selección?

El sesgo de selección puede ser corregido usando técnicas de estimación por máxima verosimilitud, pero este procedimiento fue históricamente un mecanismo demasiado complicado. En su lugar, un procedimiento alternativo, desarrollado por Heckman en 1979, es el método usualmente utilizado.

A pesar de que en la actualidad el tiempo de computadora es barato, el método de Heckman no ha perdido popularidad. El método consiste en un proceso en dos etapas. En la primera etapa los parámetros de la regla de decisión (selección) son estimados, y en la segunda etapa el modelo de regresión es estimado. El procedimiento es a menudo descrito de forma más específica (i.e., en la primera etapa se usa un modelo probit para estimar la regla de selección, y en la segunda etapa, se estima el modelo de regresión utilizando mínimos cuadrados ordinarios), pero las variaciones en este procedimiento en dos etapas son simplemente variaciones de Heckman.

En los artículos de Heckman y Rob, se sugieren una variedad de métodos. Estos incluyen técnicas de variables instrumentales, que últimamente han estado ganando protagonismo en la literatura. En este ensayo, no se pretende hacer un sumario de las diferentes técnicas sugeridas por los autores.

Para terminar, y a modo de conclusión, vale resaltar que los datos de sección cruzada puede llegar a ser información muy poderosa para estimar modelos en donde se presentan problemas de selección. Este tipo de datos requieren diferentes, pero no necesariamente más fuertes, supuestos que los utilizados en aplicaciones con datos longitudinales. Los datos de sección cruzada son mucho menos costosos de obtener, y no presentan problemas de *attrition* (aunque pueden presentar problemas de no conformidad).

Por último, cualquiera que realice investigación empírica, debería comprender que si impone supuestos, a fin de realizar estimaciones, sus resultados serán tan buenos como los supuestos hechos.

Bibliografía

- Greene, William H. (1993), "Econometric Analysis", second edition., Macmillan Publishing Co.
- Heckman, James (1974), "Shadow Prices, Market Wages, and Labor Supply", *Econometrica*, Vol.42, pp.679-694.
- Heckman, James (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, Vol.47, pp.153-161.
- Heckman, James (1990), "Varieties of Selection Bias", *American Economic Review*, pp.313-318.
- Heckman James and Richard Robb (1985), "Alternative Methods for Evaluating the Impact of Interventions", In *Longitudinal Analysis of Labor Market Data*, Heckman, James J. And Burton Singer, eds. Econometric Society Monographs series, n.10, Cambridge University Press, pages, 156-245.
- Heckman James and Richard Robb (1986), "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes", *Economics Research*.