

CAPÍTULO 3

LA EVALUACIÓN DE PROGRAMAS

Después de hacer la revisión histórica de las políticas aplicadas en nuestro país, se ha visto que muchas de las propuestas y soluciones generaron otra serie de conflictos ante un país que se enfrenta a la apertura de mercados internacionales. El Programa denominado PROCAMPO, ha sido la respuesta de política más duradera, con la finalidad de generar un campo competitivo teniendo al hogar productivo como punto de partida.

El presente capítulo tiene como punto central generar un análisis sobre las evaluaciones realizadas a PROCAMPO, por lo que se hace necesario, en primer lugar, estudiar la teoría sobre las evaluaciones de impacto, básicamente la evaluación tradicional, la evaluación en base a variables instrumentales, la construcción de la frontera estocástica, el *Propensity Score*, la corrección de sesgo de Heckman, junto a la Razón Inversa de Mills. En la segunda parte se hace revisión sobre las evaluaciones de impacto realizadas a PROCAMPO y las críticas realizadas a cada una.

La Teoría de la Evaluación de Programas y Políticas Públicas

Los creadores de políticas públicas, sobre todo de aquellas que se esperan que tengan impactos socioeconómicos, buscan pruebas concretas sobre los proyectos puestos en marcha. Evaluaciones realizadas a los programas públicos en el pasado sólo otorgaban información cualitativa, la cual, carecía de elementos de observación que pudieran ser comparables con los elementos expuestos en dichos informes. El resultado de esto era otorgar ideas poco relevantes e insatisfactorias.

Una evaluación de impacto mide el funcionamiento de un programa en el logro de objetivos bien definidos contra elementos de observación comparables explícitos (contrafactuales), bajo la ausencia del programa. Un indicador de resultado observable, Y , es identificado como relevante al programa y al período de tiempo sobre el cual se esperan impactos. Se dice que es de *impacto*, debido a que busca marcar la diferencia de resultados entre aquellos que son beneficiarios de algún programa o política pública, de aquellos que no se encuentran dentro del mismo. El objetivo de una evaluación de impacto es marcar si esta diferencia es atribuible al programa evaluado (Aedo, 2005). La evaluación consiste en la inferencia de relación causal entre los estados de tratamiento y los posibles resultados que se esperan del programa.

“Ejemplos de tratamientos en el contexto económico son la inscripción en un programa de capacitación laboral, ser miembro de un sindicato, la recepción de un pago mediante transferencia de un programa social, cambios en las regulaciones para la recepción de una transferencia de un programa social, cambios en las reglas y reglamentos relativos a las transacciones financieras, los cambios en los incentivos económicos, y así sucesivamente.” (Cameron y Trivendi: 2005, p. 860). En todos estos casos existe una situación común: que el mecanismo de asignación para recibir tratamiento no es aleatorio, debido a que en la mayoría de los casos, los individuos eligen ser tratados (autoselección).

La Evaluación Tradicional

Para realizar una evaluación se requiere conocer la información. Los datos deben incluir una observación de Y_i para cada unidad i en una muestra de tamaño n . El estado de tratamiento, T_i , es observado, con $T_i = 1$ cuando la unidad i recibe el programa (es tratada) y

$T_i = 0$ cuando no. La formulación tradicional de la evaluación postula dos resultados potenciales para cada i : el valor de Y_i en el tratamiento es denotado Y_i^T mientras Y_i^C se le denomina contrafactual¹ (que no recibe tratamiento), entonces, la unidad i obtiene beneficios $G_i = Y_i^T - Y_i^C$. En la literatura, G_i se conoce como "el impacto" del efecto causal del programa para la unidad i (Wooldrige, 2001).

En base al texto de econometría de Jeffrey Wooldrige (op. cit.), una de las medidas más usadas para la medición de impacto es el efecto de tratamiento promedio sobre el tratado (ATT):

$$ATT \equiv E(G|T = 1)$$

ATT es el impacto promedio de los que reciben el programa. En la misma evaluación, también se podría estar interesado en el efecto de tratamiento promedio sobre el no tratado:

$$ATU \equiv E(G|T = 0)$$

y el efecto de tratamiento promedio (ATE):

$$ATE \equiv E(G) = ATT \Pr(T = 1) + ATU \Pr(T = 0)$$

Por lo regular se quiere conocer el impacto promedio condicional

¹ Un escenario contrafactual se refiere al supuesto de que no hubiera ocurrido o estudiado. Al referirse a un grupo de contrafactuales en la evaluación, se señala la conformación de un conjunto de población que sea comparable al grupo de personas que son tratadas por un programa o política (Ravallion, 2008). Al hablar de elementos de observación comparables, se señala que se conforma dicho grupo de comparación por medio del emparejamiento (o "matching"), que comúnmente se utiliza en las evaluaciones de impacto debido a que la participación de un programa no es un proceso aleatorio (Aedo, 2005).

$$ATT(X) \equiv E(G|X, T = 1)$$

$$ATU(X) \equiv E(G|X, T = 0)$$

$$ATE(X) \equiv E(G|X)$$

El método más común de introducir X supone que los resultados son lineales en sus parámetros y los términos de error (μ^T, μ^C), teniendo:

$$Y_i^T = X_i \beta^T + \mu_i^T \quad (i = 1, \dots, n) \quad (3.1a)$$

$$Y_i^C = X_i \beta^C + \mu_i^C \quad (i = 1, \dots, n) \quad (3.1b)$$

Se definen los parámetros β considerando:

$$E(\mu^T | X) = E(\mu^C | X) = 0$$

por lo que los impactos promedios condicionales son:

$$ATT(X) = ATE(X) + E(\mu^T - \mu^C | X, T = 1)$$

$$ATU(X) = ATE(X) + E(\mu^T - \mu^C | X, T = 0)$$

$$ATE(X) = X(\beta^T - \beta^C)$$

La literatura sobre el tema ha reconocido que el problema esencial de la evaluación de impacto es un problema de omisión de datos, dado que es físicamente imposible medir resultados para alguien en dos estados de naturaleza al mismo tiempo (participando en un programa y no participando) (Ravallion, 2008). Sólo se puede observar Y_i^T para cuando $T_i = 1$, y Y_i^C para cuando $T_i = 0$, por lo que la ganancia G_i no puede ser observable, al menos de manera directa, debido a que no se puede encontrar el caso contrario para una misma observación i . Ninguna de las ecuaciones (3.1) son un modelo que se pueda estimar.

Con la información disponible, se puede iniciar con primeras diferencias (D) partiendo de los resultados promedio entre los participantes y no participantes:

$$D(X) \equiv E[Y^T | X, T = 1] - E[Y^C | X, T = 0] \quad (3.2)$$

Esto puede ser estimado por la diferencia del promedio de las muestras o por una regresión de Mínimos Cuadrados Ordinarios (MCO) de Y sobre T . Con el modelo paramétrico se podría estimar (3.1a) para el grupo de participantes y (3.1b) sobre el resto de la muestra, teniendo:

$$Y_i^T = X_i \beta^T + \mu_i^T \quad \text{si } T_i = 1 \quad (3.3a)$$

$$Y_i^C = X_i \beta^C + \mu_i^C \quad \text{si } T_i = 0 \quad (3.3b)$$

con la estimación de la regresión sobre la muestra reunida, se pueden especificar los coeficientes arbitrarios:

$$Y_i = X_i(\beta^T - \beta^C)T_i + X_i \beta^C + \varepsilon_i \quad (i = 1, \dots, n) \quad (3.4)$$

donde

$$\varepsilon_i = T_i(\mu_i^T - \mu_i^C) + \mu_i^C \quad (3.5)$$

Un caso especial es el llamado impacto común el cual supone que el efecto causal (G_i) es análogo al efecto de tratamiento promedio (ATE) que, al mismo tiempo, es igual a la relación entre el tratamiento promedio sobre el tratado (ATT) y el tratamiento promedio del no tratado (ATU), por lo que la identidad (3.4) tendería a convertirse en:

$$Y_i = ATE * T_i + X_i \beta^C + \mu_i^C \quad (3.6)$$

y una versión menos restrictiva señala que los efectos son los mismos en los dos grupos ($\mu^T_i = \mu^C_i$), por lo que los efectos de interacción con X se mantienen.

Pero también, se tendría que suponer que la estimación de los parámetros de impacto no debe ser sesgada. Para esto, hay que considerar la diferencia de los resultados promedios tanto para los participantes como para los no participantes:

$$D(X) = ATT(X) + B^{TT}(X) \quad (3.7)$$

donde:

$$B^{TT}(X) \equiv E[Y^C|X, T = 1] - E[Y^C|X, T = 0] \quad (3.8)$$

lo cual se conoce como sesgo de selección. Mediante esto último, se puede mostrar la diferencia promedio (o el coeficiente de regresión OLS sobre T). La ecuación (3.7) otorga el efecto de tratamiento promedio sobre el tratado si el resultado contrafactual promedio no varía por el tratamiento, es decir, $B^{TT} = 0$. En términos del modelo paramétrico, se puede asumir que

$$E[\mu^C|X, T = 1] = E[\mu^C|X, T = 0] = 0 \quad (3.9)$$

lo que asegura que OLS da estimaciones consistentes de (3.6). Si esto resulta cierto, entonces se sostiene lo mismo para μ^T , por lo que OLS también dará estimaciones consistentes para (3.4). Parte de este supuesto implica que

$$E(\mu^C|X, T = t) = E(\mu^T|X, T = t) = 0 \quad (3.10)$$

para $t = 0, 1$ que se conoce como exogeneidad condicional.

El problema más común se encuentra si los creadores de una política o programa están de acuerdo en la realización de una evaluación de impacto. Las objeciones sobre este tipo de evaluaciones son claramente más persuasivas si se conoce a gente elegible a la que se le ha negado el programa, por lo que el propósito de la evaluación es conocer el “porqué” dicho proyecto no los beneficia. Sin embargo, la razón principal es que, en la práctica, existen grupos de comparación válidos, pero los recursos fiscales son inadecuados para cubrir a cada uno en sus necesidades (Ravallion, op. cit).

La asignación de un programa implica la elegibilidad, reflejando las opciones hechas por aquellos que son elegibles y la decisión administrativa de seleccionar a los participantes. Es probable que muchos de los factores que influyen en la colocación también intervengan en los resultados contrafactuales (Ravallion, op. cit). Por tanto, debe de existir una presunción general de tendencia de selección comparando resultados entre participantes y no participantes.

Es importante considerar tanto los factores observables como los no observables. Si las X 's de la información estadística capturan situaciones que no se pueden ignorar en la elegibilidad, entonces pueden ser tratadas como condicionales exógenas sobre X . Para evaluar la validez de este supuesto hay que conocer el programa específico a fondo; la exogeneidad condicional no debería ser aceptada (o rechazada) sin saber el funcionamiento del programa en la práctica, así como de la información disponible sobre el mismo (Ravallion, op. cit.).

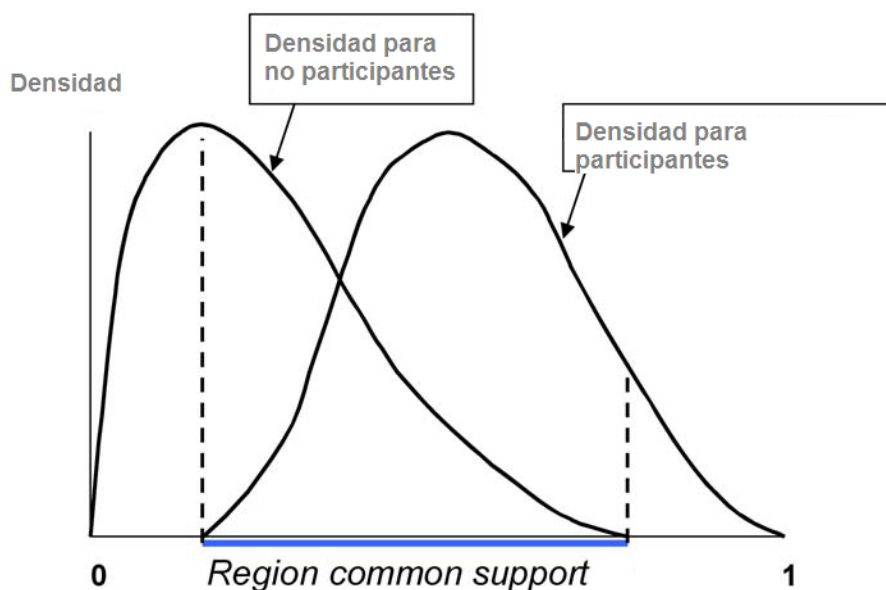
En las ecuaciones (3.3) y (3.4) las X 's son lineales en los parámetros, pero este es un supuesto que casi no se verifica ante la presencia de los programas computacionales convencionales. Pero, por otro lado, “...*hay peligros potenciales en la construcción de las estimaciones de los efectos del tratamiento. También hay diferencias sutiles de las*

interpretaciones que se derivan de las variaciones en los supuestos utilizados para la construcción de dichas medidas.” (Cameron y Trivedi: op. cit., p. 862)

Bajo el supuesto de exogeneidad condicional, se debe eliminar el sesgo de selección para asegurar que las observaciones están correctamente balanceadas entre aquellos que reciben tratamiento y los que no reciben tratamiento. Cuando la elegibilidad del programa es independiente de los resultados observados, entonces la información estadística relevante estaría balanceada entre los dos grupos de comparación, por lo que se conoce a esto como la probabilidad de participación condicional, o *Propensity Score* (Wooldrige, op. cit.).

La región de *Propensity Score* para el cual un grupo de comparación válido puede ser encontrado es conocida como región ‘*Common Support*’, como se muestra en la Figura 3.1.

FIGURA 3.1



Cuando esta región es pequeña aumenta la dificultad de identificar el efecto de tratamiento promedio. Esto es un problema en el momento de la evaluación de ciertos programas. Se puede suponer que la elegibilidad está determinada por una Prueba Proxy de

Promedios (PMT, por sus siglas en inglés) que se utiliza de manera común en el desarrollo de programas en países en vías de desarrollo. La PMT asigna una cuenta a todos los participantes potenciales en función de las características observadas que le permiten la elegibilidad al programa (Ravallion, op.cit.). En la práctica, de alguna forma, falta claridad en el uso del PMT ya que la cobertura es incompleta para realizar la prueba.

Un paso común, para estos casos, es recortar la muestra de no participantes para asegurar el *Common Support*, más allá de la ineficiencia de recolectar datos innecesarios (Wooldrige, op. cit.). El problema, para este procedimiento, se localiza cuando un subgrupo no aleatorio de participantes no es estudiado debido a que no tiene aquellas características similares (Johansen, 2006). Sin embargo, hay que realizar la comparación a partir de las características iniciales. Esto crea un posible sesgo de muestreo en la inferencia sobre el impacto, en la medida en que se deben de dejar de estudiar las unidades de tratamiento para lograr la comparabilidad (Heckman, et. al., 1998).

A pesar de este sesgo de selección, hay que considerar que el evaluador no conoce ni mide todas las variables que puedan ser relevantes. El control de las X 's por medio de la comparación (*matching*) entre unidades deja factores que no pueden ser ignorados, pero que el evaluador no puede observar a sabiendas de la existencia de grupos (o personas) que buscan entrar a un programa para la que puedan ser elegibles. Las diferencias en los promedios condicionales podrían deberse al hecho de que, en el programa, los participantes fueron seleccionados a propósito de un proceso que no se puede observar con claridad (Heckman, et. al., op. cit.). El impacto de estimación está sesgado a la cantidad dada por la ecuación (8). Por ejemplo, si el proceso de selección discrimina a los pobres, es decir,

$$E[Y^C|X, T = 1] > E[Y^C|X, T = 0] \quad (3.11)$$

entonces $D(X)$ sobreestimaré el impacto del programa. En el caso de un proceso de selección que favorezca a los pobres, se tendrían los efectos contrarios (Ravallion, op. cit.). En términos de la formulación paramétrica clásica para la evaluación, si los participantes tienen atributos que le permitan rendimientos más altos que los no participantes (con X dado), entonces los términos de error de la ecuación para los participantes (3.3a) estarán a la derecha en relación con los no participantes (3.3b). En el término de error de (3.4) no desaparecen las expectativas y OLS dará estimadores sesgados e inconsistentes. La eliminación del sesgo de selección no acaba con el problema, pero se pueden considerar ciertos sesgos considerando algún contexto específico (Heckman, et. al., op. cit.).

La formulación clásica de la evaluación supone que no hay interferencia en las unidades de comparación, lo cual permite la localización de los impactos de un programa entre los participantes directos (Ravallion, op. cit.). Observando los resultados bajo tratamiento (Y^T_i) para los participantes ($T_i = 1$) y el resultado contrafactual (Y^C_i) para los no participantes ($T_i = 0$), el grupo de comparación no es afectado por el programa.

Propiamente, sobre este tema, Edward Miguel y Michael Kremer (2003), en su estudio sobre 75 escuelas para evaluar políticas destinadas a reducir las enfermedades estomacales en Kenia, señalan que la formulación clásica puede llegar a subestimar los parámetros de los efectos de tratamiento, y que puede estar ocasionado por la aleatoriedad de niveles de unidad más grandes. Hay que considerar que, por medio de la ley de los grandes números, la ordenación aleatoria de nivel de grupo crea más variación en la densidad de tratamiento local que la realizada a nivel individual, y esta variación arbitraria puede ser usada para estimar externalidades de tipo ‘cruzado’; mientras que la ordenación aleatoria del grupo de tratamiento puede ser usada en otros ajustes con externalidades localizadas, geográficamente, o a lo largo de alguna otra dimensión, como la existencia de

vales escolares o la transmisión de la información y la difusión de tecnología. Esto último, no puede ser usado para estimar excedentes agrupados, como aquellos surgidos por efectos de precio de equilibrio generales.

Cuando se esperan externalidades en los tratamientos locales, los experimentos pueden ser diseñados para estimarlas por el tratamiento aleatorio en varios niveles. Por tanto, un diseño aleatorio subestima los beneficios del tratamiento al no hacer caso de esas externalidades (Ravallion, op. cit.). Por esa razón, en cualquier evaluación de impacto, es importante conocer los objetivos del programa, para poder identificar las variables resultado con mayor precisión. También es necesario conocer los aspectos administrativos-institucionales del programa.

Incluso para las evaluaciones no experimentales (NX), esta información es clave para diseñar encuestas que recogen de manera correcta los datos de control para el proceso de selección. El conocimiento del contexto del programa y las características de diseño también pueden ayudar en el tratamiento de selección en las variables no observables, ya que, a veces, se puede generar la identificación de posibles restricciones. *“Incluso aunque se fuera incapaz de evaluar todos los métodos no experimentales, estas pruebas sugieren que los responsables de formular la política pública estén conscientes de que las evaluaciones disponibles no experimentales de programas...pueden contener tendencias grandes y desconocidas que son resultado de errores de especificación.”* (LaLonde: 1986, p. 617). El objetivo de la investigación teórica y empírica, debería estar en el desarrollo de las características de la información disponible y de las instituciones en los contextos de evaluación propios, mismos que deben obtener estimadores no experimentales óptimos dentro de sus marcos conceptuales (Smith y Todd, 2005).

La información sobre los resultados y sus determinantes, “...incluyendo la participación en el programa, por lo regular proviene de encuestas de muestreo...” (SENPLADES, 2007), como es el caso del presente trabajo. “La unidad de observación podría ser el individuo, hogar, área geográfica o de una instalación (escuela o clínica de salud) en función del tipo de programa” (Ravallion: op. cit., p. 21). Es evidente que la recopilación de información debe abarcar el período de tiempo durante el cual se espera el impacto. El diseño de la muestra es siempre importante para la precisión de las estimaciones de impacto, al mismo tiempo, de cuánto se puede aprender de los datos obtenidos y sobre los factores determinantes de los impactos (SENPLADES, op. cit.).

Las encuestas con errores de medición pueden ser tratadas por medio de los mismos métodos utilizados para abordar el sesgo de selección. Por ejemplo, si la medición afecta a los resultados para las unidades de tratamiento y de comparación (y no están correlacionadas con las variables de control), entonces no habrá problemas para la estimación del *ATE*, al igual que no se puede saber cuantos realmente ganan por recibir beneficios debido a la existencia de una política, y cuántos son los que ‘pierden’ por no tenerlos (Carneiro, et. al., 2003).

Variables Instrumentales

Aunque directamente no es una evaluación propuesta para el presente trabajo, se tiene que mencionar debido a la importancia que reviste en la estimación del *ATE* para analizar el supuesto sobre la ignorancia del tratamiento. Los métodos *IV* para estimar el *ATE* pueden ser eficaces si se tiene un buen instrumento. El desarrollo de esta metodología da pauta a la creación de otras metodologías para las evaluaciones de impacto, por ejemplo, la metodología para la corrección del sesgo de selección de James Heckman, que se

mencionará más adelante. Para el desarrollo de la metodología presente hay que establecer la existencia de una variable denominada instrumental, el cual se denota por medio de Z , el cual puede determinar la elegibilidad del programa, independientemente de las variables X 's.

$$T_i = \gamma Z_i + \beta X_i + v_i \quad (3.12)$$

donde Z es exógena. Si los impactos tienen comportamientos homogéneos, los resultados de respuesta serán idénticos ($\mu^T_i = \mu^C_i$), por lo que el modelo expuesto en (6) sería de impacto común.

Pero no se puede suponer que la elegibilidad de un programa tiene exogeneidad condicional, ya que v_i y μ^C_i estarían correlacionados, al igual que el sesgo de selección. Joshua Angrist (1990), señala en su trabajo sobre el enrolamiento en Vietnam, que la asignación aleatoria, de ser elegible (como los concursos de números), no es ninguna garantía de que la elegibilidad es un instrumento válido para la participación, sin embargo, la asignación aleatoria proporciona una variable instrumental natural en algunos casos. Si (3.6) se sustituyera en T_i , se tendría:

$$Y_i = \pi Z_i + X_i(\beta^C + ATE) + \mu_i \quad (3.13)$$

donde $\pi = ATE\gamma$ y $\mu_i = ATEv_i + \mu^C_i$. Por lo que el estimador de impacto IV es:

$$\hat{\beta}_{IVE} = \frac{\hat{\pi}_{OLS}}{\hat{\gamma}_{OLS}} \quad (3.14)$$

El supuesto de que Z no es elemento de X sirve para identificar a π como un elemento diferente de β^C , a esto se le llama restricción por exclusión (Wooldridge, op. cit.).

Cuando una variable instrumental válida está disponible, la verdadera fuerza de un estimador de este tipo, sobre la mayoría de otros estimadores no experimentales (NX), es que es robusta ante variables no observadas que conjuntamente influyen en la elegibilidad de programas. Bajo estos supuestos, el método *IV* es menos exigente para modelar la elegibilidad de programas que el método de comparación por *Propensity Score* (Guzmán, 2005). El estimador de *IV* da una estimación constante del *ATE* en la presencia de omisiones en elegibilidad de programa, y si se tiene una variable instrumental válida, entonces se puede usar para probar la exogeneidad del *Propensity Score* o de OLS. Si los supuestos no se mantienen, los sesgos de estos estimadores pueden elevarse (Gallego y Loayza, 2002).

La mejor manera para estimar por medio del método de *IV* es a través de la aplicación de sus supuestos, por lo que sería fácil demostrar que $\gamma \neq 0$ en (3.12).

La restricción por exclusión es más difícil. Si se tiene más de una *IV* válida, entonces se tendría que realizar la prueba para demostrar la existencia o no de sobre-identificación. Sin embargo, si se tiene al menos una *IV*, entonces la restricción por exclusión es inestable dentro de los límites de la información disponible (Baker, 2000). De ninguna manera hay argumentos específicos para poder apelar la aceptación, o el rechazo, sobre la restricción por exclusión.

De manera estricta no se requiere la restricción por exclusión cuando un modelo no lineal de respuesta binaria es usado para la primera etapa, en vez del modelo lineal (3.12). Entonces el impacto se identifica de la no linealidad de la regresión de primera etapa. Sin embargo, se considera preferible tener una estrategia de identificación que sea robusta en

una regresión lineal de primera etapa (Aguiar, 2008). De ningún modo esto es preocupante, cuando la identificación descansa sobre una suposición sobre la forma funcional y la distribución de un término de error (Ravallion, op. cit.).

El supuesto de impacto común en (3.6) no es una simplificación sencilla de realizar, pero es crucial al impacto promedio que se identifica usando el estimador de *IV*. Recordando el modelo general, al inicio de este capítulo, la heterogeneidad de impacto proviene de diferencias entre los factores latentes de resultados bajo tratamiento y aquellos bajo el contrafactual; formulando esto:

$$G_i = ATE + \mu_i^T - \mu_i^C \quad (3.15)$$

por lo que (3.6) sería:

$$Y_i = ATE * T_i + X_i \beta^C + [\mu_i^C + (\mu_i^T - \mu_i^C) T_i] \quad (3.16)$$

donde $[\cdot]$ es el nuevo término de error. Por lo que la covarianza de $[Z, (\mu_i^T - \mu_i^C) T_i]$ es igual a 0. Esto fallará si la selección en el programa está implicada por las diferencias del impacto de $(\mu_i^T - \mu_i^C)$, por lo que los probables beneficiarios serán atraídos a dicha política. Para interpretar el estimador *IV* como una estimación de *ATE*, se debe suponer que los agentes no conocen μ_i^T o μ_i^C , o no actúan sobre esta información (Wooldrige, op. cit.).

Con impactos heterogéneos, el estimador de *IV* identifica el impacto para un subgrupo específico de población, más si este subgrupo está dispuesto a suscribirse al programa por lo que se presentará una variación exógena atribuible a las *IV*. Este subgrupo raras veces puede ser identificado explícitamente por el método de *IV*, por lo que queda poca evidencia sobre la interpretación que se le puede otorgar al estimador de *IV*. El método de “...*IV* no estima ninguno de los efectos de tratamiento antes definidos... *IV*”

estima parámetros diferentes de acuerdo a los supuestos que se mantienen” (Heckman y Vytlacil: 2005, p. 721)

El estimador de Variables Instrumentales Locales (*LIV*) que proponen Heckman y Vytlacil (op. cit.) se relaciona con el impacto de alguien que pueda ser elegible (y resulte beneficiario) al programa. *LIV* implica una regresión no paramétrica de resultados Y sobre el *Propensity Score*, $P(X, Z)$. La pendiente de esta función de regresión es el impacto en los valores específicos de X y Z . Si hay heterogeneidad en el impacto, primero hay que revisar si hay linealidad entre Y y el *Propensity Score*. Si no existe linealidad, entonces el estimador de *LIV* sería el indicado; en caso contrario el estimador *IV* es el indicado, por lo que el coeficiente de regresión OLS de Y sobre el *Propensity Score* sería *ATE*.

Cuando se pone en práctica el método *IV*, en muchas ocasiones se otorgan resultados poco probables, por ejemplo estimadores muy grandes, o caso contrario, muy pequeños. La razón podría ser la existencia de violaciones en la restricción de exclusión. Si se pudieran hacer a un lado ciertos valores para Y por razones *a priori*, entonces se pueden establecer límites a las estimaciones de impacto. Por lo que:

$$0 \leq TT \leq E(Y^T|T=1)(\leq 1)$$

y

$$\begin{aligned} & (E[Y^T|T=1] - 1) \Pr(T=1) - E[Y^C|T=0] \Pr(T=0) \\ & \leq ATE \leq (1 - E[Y^C|T=0]) \Pr(T=0) + E[Y^T|T=1] \Pr(T=1) \quad (3.17) \end{aligned}$$

La amplitud de los límites depende de los ajustes realizados en la información. Los límites no pueden ser muy utilizados, en el caso de que los resultados resulten ser variables continuas.

Altonji, Elder y Taber (2004) (AET) proponen otra metodología al ajuste de límites. En sus dos estudios sobre los efectos sobre la educación en las escuelas católicas, señalan que el sesgo en OLS se presenta en la sobreestimación sobre el impacto real. “*Los investigadores, a menudo, abogan de manera informal por la exogeneidad de una variable explicativa mediante la examinación de la relación entre la variable y el conjunto de características observadas, o la evaluación si los puntos estimados son sensibles a la inclusión de variables de control adicionales... ahora se trata de mostrar que tales pruebas pueden ser informativas en algunas situaciones...proporcionar un camino cuantitativo para evaluar el grado de tendencia de variables omitida.*” (AET: 2005, p. 2). Al mismo tiempo AET (2005) cuestionan la metodología de la restricción por exclusión cuando se utilizan IV basadas en el pasado. OLS supone que los resultados no observados no están correlacionados con la elegibilidad del programa. AET (op. cit.) señalan que los resultados no observados tienen el mismo efecto sobre la elegibilidad que los resultados observados ($X_i\beta^C$ en (3.16)). Este último supuesto, requiere limitar el coeficiente de correlación de los términos de error de las ecuaciones y de participantes² en la función de *Score* en los resultados ($X_i\beta^C$).

La restricción por exclusión puede ser cuestionada, por ejemplo, si se considera el problema de identificar el impacto de un programa de entrenamiento laboral asignado a los salarios (Jaramillo, et. al., 2007). Partiendo de la literatura sobre economía del trabajo, uno podría usar las características del hogar como *IV's* para la participación en el programa. Este tipo de características influyen en el programa, pero es poco probable que los

² μ^C en (6) y v_i en (12) para valores dados para el coeficiente de regresión de la función de *Score* observables en la ecuación de participación (X_i en Eq. (12) con $\gamma = 0$)

empleadores observen las condiciones de vida de sus empleados, aunque pueden considerar otras variables de control observables, como la edad y la educación del trabajador individual. Sin embargo, para al menos un poco del potencial de las *IVs*, la restricción por exclusión es cuestionable cuando hay efectos de excedente relevantes en la productividad al interior de los hogares. Por ejemplo, se ha demostrado que en los países en vías de desarrollo, la presencia de una persona que sabe leer y escribir en el hogar puede ejercer efectos sobre la productividad de un familiar analfabeto (Solórzano, 2005).

AET (2004) sostienen que su estimación tiene un límite inferior al verdadero impacto cuando éste es positivo, lo que se basa en la presunción de que el término de error en la ecuación de resultados incluye por lo menos algunos de los factores que no están correlacionados con la participación. OLS también establece un límite superior. Por lo tanto, el estimador que determinan AET da una indicación útil del cómo OLS puede ser realizado ante cualquier sesgo de selección basado en no observables. Sin embargo, y a pesar de lo interesante que resulta esta propuesta, pareciera que en sus dos textos mencionados, los resultados parecen muy tendenciosos.

En el caso de Wooldrige (2001), basándose en los supuestos ya expuestos, su aplicación es de lo más sencilla. Primero se determina el *Propensity Score*, $P(X)$, se realiza la estimación, $p(X)$, y este toma el lugar de Z , para poder obtener una regresión de Y sobre el *Propensity Score*, $P(X, Z)$. El problema de esta metodología se encuentra en la presencia de heterocedasticidad, ya que si bien $Var(\mu_i|X, Z)$ es constante, no pasa lo mismo con $Var(Y|X, Z)$, porque $Var(T|X, Z)$ tampoco será constante, por lo que se tiene que aplicar una regresión robusta para evitar el problema heterocedástico.

Entre los ejemplos de estudio sobre la metodología de *IV*, se encuentran algunas evaluaciones sobre el programa PROEMPLEO aplicado en Argentina. En este trabajo de

Ravallion y Gallasso (2004), se incluyen componentes asignados de manera aleatoria, pero los resultados no fueron los esperados, ya que no había diferencias sustanciales en los ingresos (ni en la productividad) de aquellos que recibieron entrenamiento laboral y el grupo de control; y en cambio, los empleadores afectan los salarios de acuerdo a la participación de programa. De hecho, la mayor parte de los usos de estimadores de *IV* han servido de instrumentos no experimentales. En la literatura de la economía de trabajo, las regresiones de salario, por lo regular, toman a la mano de obra como variable endógena (de ahí las observaciones de los candidatos a tener salarios). Una fuente común de *IV*'s se encuentra en la modelación del problema seleccionado, en el cual se encuentran variables que influyen en los gastos y las ventajas de participación de mano de obra, pero no afectan ingresos.

En la evaluación de programas antipobreza, en países en vía de desarrollo, las tres fuentes populares de *IV* han sido:

- la colocación geográfica de programas,
- variables políticas y,
- discontinuidades creadas según el diseño de programa.

La geografía de colocación de programa ha sido usada para la identificación en varios estudios. Muchas veces implican variables sobre la naturaleza, variables topográficas o agro-climáticas, y otros rasgos que pueden ayudar la identificación geográfica. Las características políticas de áreas geográficas han sido otra fuente de instrumentos. El entendimiento de la economía política en la elegibilidad de programas puede ayudar en la identificación de impactos; un ejemplo de esto, es que en algunos estudios se considera la presencia de mujeres en los congresos (diputadas, congresistas, asambleístas, etc.) como

variable instrumental en el estudio de las políticas sobre seguros, compensaciones y mejoras salarios para los trabajadores. Las discontinuidades en el diseño de programa se localizan donde el estimador de impacto está en las cercanías de la elegibilidad o no de algún programa. En estos casos, se presentan ‘puntos de atajo’ (cut-off) en los cuales no se pueden encontrar razones inmediatas para el cual el cambio en los tamaños de la muestra podrían tener efectos independientes sobre las ganancias de un programa.

La Frontera de Producción Estocástica

Otra forma de realizar evaluación es comparar y analizar los comportamientos de los agentes tratados de los no tratados. Para el caso en cuestión, suponiendo que se quisiera realizar un pronóstico o un comportamiento promedio entre los agentes tratados y sus contrafactuales, se analizan los límites de cada grupo. Si las fronteras resultan ser diferentes, probablemente el grupo de tratamiento podría mostrar crecimiento productivo, pero por otro lado, la diferencia podría ser porque el grupo contrafactual presente retrocesos productivos. De igual manera, con este modelo se puede mostrar la diferencia de los costos de oportunidad de cada uno de los grupos.

Para entender el modelo, se puede mencionar un documento vinculado con el presente trabajo (debido al contexto de la base de datos de la presente tesis), que se dedica al estudio de la eficiencia productiva de los productores de maíz en México, realizado por Antonio Yunez-Naude, Miriam Juárez-Torres y Fernando Barceinas-Paredes (2006)³. En dicho trabajo, se desarrolla la frontera de producción, tanto para la producción de maíz

³ Los dos primeros Investigadores del Colegio de México-PRECESAM, el tercero es investigador de la Universidad Autónoma Metropolitana (UAM). Sobre el PRECESAM, la ENHRUM y la base de datos se abunda en el Capítulo 4.

como para la comercialización. La finalidad de dicho estudio es analizar de manera empírica si los agricultores pequeños y medianos tienen formas eficientes de producción.

La justificación de estudiar la frontera productiva del maíz es la misma que origina el presente trabajo: es el grano más producido en el país y con mayor tradición cultural. Lo interesante del estudio de Yunez-Naude es que desde el principio queda geográficamente identificada la producción comercial en el norte y oeste del país, mientras que la producción de subsistencia se ubica en el sureste y centro.

El modelo de dicho estudio (del cual también se realiza con algunos cambios en el presente trabajo) se basa a partir del análisis de las elasticidades por medio del modelo Cobb-Douglas, el cual parte de la función log-log:

$$\ln Y_i = \beta_0 + \sum_{n=1}^N \beta_n \ln X_{ni} + e_i \quad (3.18)$$

La identidad (3.18) parte del hecho de que no todos los hogares de todas las regiones son igual de eficientes. La base de datos, con la que cuenta, fue desarrollada por ellos mismos, y de la cual se menciona posteriormente en este trabajo.

Con dicho modelo se puede señalar si alguno de los dos grupos de estudio utiliza sus factores de manera eficiente o ineficiente, en comparación al grupo de control. Con este mismo modelo se podría afirmar si los factores utilizados, utilizando de ejemplo al grupo contrafactual, son ociosos en comparación del grupo de tratamiento.

El problema de este modelo radica en que si bien se puede mostrar la ineficiencia de alguno de los grupos, no se puede mostrar si la situación parte de la presencia de factores ociosos o de una diferencia tecnológica entre los productores con tratamiento y los productores sin tratar. Otra dificultad que puede presentar el modelo radica en que, si se

demuestra ineficiencia de alguno de los grupos de estudio, pero al mismo tiempo se muestra que los recursos son utilizados de forma eficiente, entonces habría que estudiar la finalidad del proceso productivo (comercialización o subsistencia).

Las diferencias de los costos de oportunidad entre los productores de ambos grupos no señalan la particularidad de la región, sino más bien, se convierte en un elemento más de la diferencia entre ser tratado y no ser tratado.

La importancia del documento de Yunez-Naude radica en el manejo de la información en base a la medición de elasticidades, lo cual permite analizar los cambios de las variables en términos porcentuales. Al concebir que cada uno de los hogares tienen características diferentes, un análisis en términos propios de nivel sólo analiza el cambio en términos absolutos, y probablemente las diferencias entre la producción de los hogares con y sin programa genere poca confiabilidad, mientras que el análisis porcentual parte, cómo es sabido, de los términos relativos de la producción, el autoconsumo y el gasto.

Método del *Propensity Score*

Un método bastante utilizado es el que proponen Rosenbaum y Rubin (1983) (R&R). El concepto de *Propensity Score* es útil cuando la asignación para ser tratado “...*depende estocásticamente en un vector de variables observables x , como en los datos de observación o cuando el tratamiento se dirige a alguna población definida por ciertas características observables (como la edad, sexo o nivel socioeconómico)*...” (Cameron y Trivedi: op. cit., p. 864) y no a que dicha asignación se hiciera de manera aleatoria. Se realiza bajo el supuesto sobre la ignorancia del tratamiento para la estimación del *ATE*. “*El supuesto es equivalente a la asignación del tratamiento que hace caso omiso de los*

resultados” (Cameron y Trivedi: op. cit.. p. 863). La aplicación del *Propensity Score* incluye:

- 1) el cotejo (*matching*) de la muestra sobre el *Propensity Score*, que es una generalización de la discriminación comparativa,
- 2) el ajuste multivariante según la subclasificación sobre el *Propensity Score* donde las mismas subclases son usadas para estimar efectos de tratamiento para todas las variables de resultado y en todas las subpoblaciones, y
- 3) la representación visual del ajuste de la covarianza multivariada por medio de un gráfico bidimensional

El desarrollo del *Propensity Score* parte de cinco teoremas:

- i) El *Propensity Score* se debe considerar como una cuenta de equilibrio.
- ii) El *Propensity Score*, genera una cuenta más precisa, ya que al interior de las X 's se puede encontrar una cuenta (*score*) de equilibrio mejor definida.
- iii) Bajo la condición fuerte de ignorancia de tratamiento (del cual se hablará más adelante), dadas las X 's, entonces existe esta condición dado cualquier *Score* equilibrado.
- iv) A cualquier valor de un *Score* equilibrado, la diferencia entre el grupo de tratamiento y el grupo de control es un estimador insesgado de *ATE* al interior de los valores del *Score*, sí y sólo sí el grupo de tratamiento tiene la condición señalada en iii). Bajo dicha condición, el par comparable, las subclasificaciones y las covarianzas ajustadas en un *Score* balanceado pueden producir estimadores insesgados de efectos de tratamiento

- v) A través de la muestra la estimación del *Score* balanceado puede producir un balance muestral en las X 's. (R&R, op. cit.)

En las ecuaciones (3.3) μ_i^T y μ_i^C no pueden ser observados, pero *ATE* y *ATT* pueden ser estimados modelando:

$$P(X_i) \equiv P(T=1|X_i) \quad (3.19)$$

que es la probabilidad del tratamiento dadas las covariables. La función $P(X_i)$, es la probabilidad de respuesta al tratamiento, el cual es propiamente el *Propensity Score*. *ATE* y *ATT* puede escribirse en función del *Propensity Score* (R&R, op. cit.). Bajo el supuesto del promedio condicional:

$$0 < P(X_i) < 1 \quad (3.20)$$

para todo (X_i) , entonces:

$$ATE = E([T - P(X_i)]Y_i / \{P(X_i)[1 - P(X_i)]\}) \quad (3.21)$$

y

$$ATT = E\{[T - P(X_i)]Y_i / [1 - P(X_i)]\} / P(T=1) \quad (3.22)$$

considerando (3.1), (3.3) y (3.4) en (21) se tiene:

$$[T - P(X_i)]Y_i = [T - P(X_i)](1 - T)Y^C + TY^T = TY^T - P(X_i)(1 - T)Y^C - P(X_i)TY^T \quad (3.23)$$

retomando el supuesto de promedio condicional:

$$T\beta^T(X_i) - P(X_i)(1 - T)\beta^C(X_i) - P(X_i)T\beta^T(X_i)$$

donde

$$\beta(X) \equiv E(Y|X)$$

por lo que si se considera la expectativa condicional en las X 's:

$$P(X_i)\beta^T(X_i) - P(X_i)[1 - P(X_i)]\beta^C(X_i) - [P(X_i)]^2\beta^T(X_i) = P(X_i)[1 - P(X_i)][\beta^T(X_i) - \beta^C(X_i)]$$

De esa manera el valor esperado para ATE es el que ya se había señalado con anterioridad:

$$ATE(X) = [\beta^T(X_i) - \beta^C(X_i)] = X_i(\beta^T - \beta^C) \quad (3.24)$$

De igual forma, para la ecuación (3.21) se tiene:

$$E = \{[T - P(X_i)]Y_i/[1 - P(X_i)]|X\} = P(X_i)[\beta^T(X_i) - \beta^C(X_i)]$$

por lo que bajo el supuesto del promedio condicional:

$$E = \{P(X_i)[\beta^T(X_i) - \beta^C(X_i)]\} = E\{T[\beta^T(X_i) - \beta^C(X_i)]\} = E[T(Y^T - Y^C)]$$

pero

$$E[T(Y^T - Y^C)] = P(T=1)E[T(Y^T - Y^C)|T=1] + P(T=0)E[T(Y^T - Y^C)|T=0]$$

$$= P(T=1)E(Y^T - Y^C|T=1)$$

por lo que el lado derecho de (3.21) es

$$ATT = \{P(T=1)E(Y^T - Y^C|T=1)\}/P(T=1) \quad (3.25)$$

R&R (op. cit.) llaman a (3.21) y (3.22) como la condición fuerte de ignorancia de tratamiento. Pero ni ATE ni ATT tienen parámetros plenamente especificados bajo esta condición, debido a que la probabilidad de respuesta $E[Y^T|X, T = 1]$ está sujeta a la dirección que tome el análisis de identificación. La ecuación (3.25) muestra que ATT todavía es identificado si $P(X)=0$ para alguna X_i , pero esta causalidad tiene poco valor práctico pues se excluyen las unidades que no tienen ninguna posibilidad de ser tratados (Ham, et. al., 2005). En la estimación ATE (o ATT incluso puede ocurrir), se excluye $P(X)=1$ por lo que no se puede estimar ninguno de los dos debido a la probable inclusión de variables demográficas y de población que son tratadas y condicionadas en las X 's.

De hecho para estimar ATE y ATT , se necesita un estimador de $P(\cdot)$. R&R (op. cit.) sugieren un modelo *logit*, pero Wooldrige (op. cit.) sugiere un modelo *probit*, donde las X 's y funciones de estas se incluyen en el modelo. En este caso, no hay ningún peligro de que $p(X)^4$ sea igual a cero o a uno porque los valores estimados por logit (o probit) se encuentran dentro de este intervalo, pero esta restricción de la forma funcional puede ocultar ciertos problemas al interior de la población. De esa manera, si $p(X) \equiv F(X; \gamma)$, entonces $p(X)$ es un buen estimador, donde γ se obtiene en la primera etapa de la estimación de respuesta binaria T en las X 's, por lo que el estimador consistente de ATE es:

$$ate = N^{-1} \sum [T - p(X_i)] Y_i / \{p(X_i)[1 - p(X_i)]\} \quad (3.26)$$

Este estimador puede ser muy convincente pero puede presentar varianzas asintóticas muy estrechas entre los estimadores que están basados en el supuesto de que las X 's, T y (Y^T, Y^C) son independientes y la condición propuesta en (3.20). Un estimador

⁴ Las letras minúsculas implica que se habla en términos de valor estimado

dentro de la evaluación de programa se puede obtener por medio de una regresión OLS que incluye el *Propensity Score* estimado, $p(X)$, como un regresor adicional, es decir:

$$Y = \beta_0 + \beta_1 T + \beta_2 p(X) + u \quad (3.27)$$

donde T estima el efecto de tratamiento, mientras que el *Propensity Score* estimado es la función de control. El *Propensity Score* estimado contiene toda la información relevante en el proceso de estimación del efecto de tratamiento.

Otro aspecto que se puede suponer es que:

$$E(Y_i^T - Y_i^C | X) = \beta^T(X_i) - \beta^C(X_i)$$

no está correlacionado con:

$$\text{Var}(T|X) = P(X_i)[1 - P(X_i)]$$

Si el estimador paramétrico $p(\cdot)$ es consistente y es normal en términos asintóticos, entonces el coeficiente estimado en T es consistente y normal en términos asintóticos para el *ATE*. De hecho, el cálculo de $P(X_i)$ es válido para la condición de tratamiento como la de los no tratados, debido a la existencia de valores coincidentes entre las sub-muestras tratadas y no tratadas (Cameron y Trivendi, op. cit.), tal y como ocurre en la Figura 3.1.

Nuevamente R&R (op. cit.) proponen una versión más general al modelo de regresión (3.27):

$$Y = \beta_0 + \beta_1 T + \beta_2 p(X) + \beta_3 T(p(X) - \zeta_p) + u \quad (3.28)$$

donde ζ_p es el valor promedio de $p(\cdot)$.

Se puede probar que:

$$\begin{aligned}
E[T|Y^T, Y^C, P(X)] &= E[E(T|Y^T_i, Y^C_i, X_i)|Y^T_i, Y^C_i, P(X)] \\
&= E[P(X)|Y^T, Y^C, P(X)] = P(X) \quad (3.29)
\end{aligned}$$

por lo que se puede obtener $E[Y|T, P(X)]$. Formulando:

$$Y = Y^C + (\xi^T - \xi^C)T + T(\mu^T - \mu^C)$$

con el supuesto de que T y (Y^T, Y^C) son independientes dado $P(X)$, entonces se tendría:

$$\begin{aligned}
E[Y|T, P(X)] &= E[Y^C|P(X)] + (\xi^T - \xi^C)T + T\{E[\mu^T|P(X)] - E[\mu^C|P(X)]\} \\
&= \beta_0 + \beta_1 P(X) + (\xi^T - \xi^C)T + \beta_2 T(P(X) - \xi_p) \quad (3.30)
\end{aligned}$$

donde $\xi_p \equiv E[P(X)]$.

La hipótesis de linealidad para $E[Y^C|P(X)]$ y $E[Y^T|P(X)]$ es demasiado restrictiva en muchas aplicaciones. Como $P(X)$ es limitada entre cero y uno, $E[Y^C|P(X)]$ y $E[Y^T|P(X)]$ son limitadas en virtud de la linealidad (Wooldridge, 2001). En el caso de que Y es binario, la linealidad de estos valores esperados también sería discutible, pero podría ser una aproximación razonable. Por supuesto, es una simple cuestión de reemplazar p_i con un polinomio de orden bajo en sí mismo, lo que reduce los términos antes de construir su interacción con T_i .

Todas las estimaciones de *ATE* basadas en el *Propensity Score*, implican regresiones o fórmulas que son similares a regresiones en el sentido de, que en esta medición, se incluye el promedio muestral.

Las estimaciones del *Propensity Score* también son usadas de muy diferentes maneras en los textos sobre los efectos de tratamiento. Varios estimadores *matching* han sido propuestos y, sus distribuciones asintóticas, son viables en la mayoría de los casos. El

matching aproximado propuesto por R&R (op. cit.) está motivado por el ensayo en el que se supone la selección de un *Propensity Score*, $P(X)$, aleatorio de la población de estudio. Entonces, se seleccionan a dos agentes de la población que comparten el *Propensity Score* clasificado, donde un agente recibe el tratamiento y el otro no. Bajo el supuesto de que las X 's, T y (Y^T, Y^C) son independientes, la diferencia esperada en los resultados observados para estos agentes es

$$E[Y|T=1, P(X)] - E[Y|T=0, P(X)] = E[Y^T - Y^C|P(X)] \quad (3.31)$$

por lo que *ATE* es condicional de $P(X)$.

Por expectativas iteradas, promediando a través de la distribución del *Propensity Score* se tiene que:

$$ATE = (Y_i^T - Y_i^C)$$

Una estrategia de valoración requiere la estimación del *Propensity Score*. Primero está el cálculo de las diferencias de respuesta para pares comparados sobre la base del *Propensity Score* estimado, y luego se determina el promedio sobre todos esos pares. Obtener una predicción de probabilidades idénticas no es un trabajo sencillo, en vez de eso, se tienen que agrupar en *células* o localidades para que de esa forma se puedan realizar estos promedios. Agentes con similares *Propensity Score*, son considerados un *match*. Un problema práctico en la comparación del *Propensity Score* es que es difícil de encontrar agentes tratados y no tratados con *Propensity Score* similares.

Entre los primeros estudios que analizan este tipo de estrategia, se encuentra el ya mencionado artículo de Robert J. LaLonde (op. cit.), el cual explica el problema de las estimaciones en la evaluación de programas de entrenamiento laboral. Las estimaciones

econométricas, señala el investigador, a menudo se diferencian considerablemente de los resultados experimentales. Además, incluso cuando las estimaciones econométricas pasan pruebas de especificación convencionales, estas todavía fallan en reproducir los resultados experimentalmente determinados. Con los procedimientos no experimentales, señala LaLonde, se pueden producir estimaciones grandes y positivas para cierto conjunto de participantes, mientras que pueden ser pequeños y negativos para otro conjunto de participantes con otras características (esto puede ser por género, nivel socio-económico u otra característica en la que se puedan dividir el conjunto de participantes a un programa en específico), por lo que estos resultados pueden ser reproducidos de mejor manera en el primer grupo que en el segundo. Pero en el caso de trabajar con información longitudinal, se reduce la posibilidad de tener errores de especificación a través de la información de sección cruzada. De hecho, la única forma de saber si estos métodos son buenos es realizando el *matching*.

Hay muchas oportunidades de usar el método experimental de evaluar el potencial del sesgo de selección en la evaluación de programas sociales y en otras áreas de investigación econométrica (LaLonde, op. cit.).

Petra Smith y Jeffrey Todd (2005) señalan que el *matching* en el *Propensity Score* es bueno para analizar las diferencias de las covarianzas entre el grupo de tratamiento y el grupo de comparación, por lo que es necesario realizar esta prueba de balanceo para los propósitos requeridos. Después de aplicar una serie de estimaciones a través de información de sección cruzada, se llega a la conclusión de que el *Propensity Score Matching* (PSM) es una herramienta econométrica muy útil, aunque no representa una solución general a los problemas de evaluación. De acuerdo al trabajo de LaLonde (op. cit.), refiriéndose a estos autores, señalan que sus pruebas no pueden ser interpretadas como la exposición de la

superioridad del *matching* sobre los estimadores econométricos tradicionales. Los estimadores se diferencian en sus supuestos de identificación, y los supuestos base de un estimador dado a veces se mantienen con la información y en otros casos fallan.

En algunos contextos, en particular de aquellos con buena información en variables relacionadas con la participación y resultados, el *matching* puede ser la mejor opción. El método de *matching* no funciona bien cuando se eliminan fuentes de sesgo, una tarea para la cual la información no fue diseñada. Las conclusiones positivas en cuanto a otros estimadores *matching*, destacan la importancia de escoger un método no experimental compatible con los rasgos de los datos y las instituciones en un contexto dado. “*El conocimiento en cuanto a las características de los estimadores a partir de los datos disponibles y el contexto institucional...debería ser usado en el diseño y la evaluación de la investigación de evaluación*” (Smith y Todd: op. cit., p. 352). Los autores citados señalan que la imposición del *Common Support* puede afectar las tendencias estimadas en algunos casos, pero no en todos los casos.

Muchas de las propiedades del *matching*, aún no se entienden del todo, en específico las propiedades asintóticas. Los estimadores de los métodos PSM, en muchas ocasiones, no llegan a ser mejores que aquellos que son producto de las regresiones tradicionales, y llegan a ser acertados en el momento en el que se eliminan los malos resultados de estimadores no experimentales (Glazerman, Levy and Myers, 2003). En la práctica, los errores estándar se obtienen por el método de eliminación de heterocedasticidad, a pesar de que este método no es evidente en todos los casos. Los estimadores para el *matching*, en la determinación de *ATE*, son utilizados en la investigación de evaluación a pesar del hecho que sus propiedades muestrales no han sido establecidas en muchos casos.

La ausencia de resultados formales en esta área, puede estar provocado por el hecho de que extensiones estándar asintóticas no se aplican a estimadores dentro del área *common support*. Las propiedades formales de muestras grandes resultan ser bastante pobres. Se puede demostrar que los estimadores *matching* pueden incluir un término de tendencia condicional que no desaparece en muestras de gran tamaño, por lo que no serán muy consistentes en el momento de realizar simulaciones basadas en proyectos existentes en las evaluaciones no experimentales. Una ejecución sencilla de este estimador, donde el sesgo está ajustado en base a la regresión lineal, parece funcionar mejor en comparación a los estimadores del *matching* sin el ajuste de sesgo, como a estimadores en base a la regresión en términos de ese sesgo y el error al cuadrado promedio. Esto provoca la obtención de buenos intervalos de confianza entre el 90% y 95%, por lo que es un buen estimador (Abadie and Imbens, 2004)

La metodología de corrección de sesgo de Heckman

En sus trabajos sobre los salarios, James Heckman introdujo el concepto de "sesgo de selección" en la econometría moderna. De acuerdo a Heckman, en estos estudios, los investigadores sobre economía laboral, por lo regular, calculaban los promedios de los salarios, pero en dichas mediciones, no estaban considerados los individuos que en ese instante dejan de laborar, por lo que desechaban dicha información y después deducían promedios usando las observaciones restantes, en otras palabras, no se estaba considerando al desempleo friccional (Hernández y Méndez, 2005). Heckman demostró que procesos no aleatorios pueden conducir a sesgos de selección, y por tanto, a tener conclusiones erróneas sobre programas y políticas públicas. Por esa razón, se debe tener más cuidado al momento de generar la evaluación.

En muchos casos, señalan Cameron y Trivedi (op. cit.), los resultados dependen de que la selección para participar en algún programa o actividad se deba a un proceso de auto-selección, o en todo caso, al generar la selección de la muestra, se caiga en el error del sobremuestreo. En ambos casos, existen problemas muy relacionados, por lo que dichos autores le llaman modelos de selección de la muestra.

El sesgo proveniente de la selección de una muestra aleatoria sirve para estimar las conductas relacionadas al error de especificación. El método de estimación de dos etapas sirve para corregir este tipo de sesgos. La corrección es de fácil aplicación y tiene una base firme en la teoría estadística. La corrección de Heckman implica un supuesto de normalidad y proporciona una prueba para examinar el sesgo de selección, con lo cual se puede formular un modelo para su corrección (Puhani, 2000).

En el modelo se quiere buscar a los determinantes de una variable dependiente Y , pero sólo se tiene acceso a un número limitado de observaciones. De esta población, se toma una muestra (la cual no llega a ser aleatoria), y que al momento de estimar este conjunto de la población, se puede encontrar la existencia del sesgo. Por esa razón, la corrección de sesgo se realiza en dos etapas, y de ahí su nombre a este método. Primero se formula un modelo para determinar la probabilidad de obtención de un programa o proyecto. Heckman (al igual que Wooldrige (op. cit.)), como ya se mencionó antes, prefiere la especificación de un modelo probit de la forma:

$$P(X_i) \equiv P(T=1|Z) = \Phi(Z\gamma) \quad (3.32)$$

donde T sigue siendo igual 1 en caso tratamiento, pero es igual a cero en otro caso, Z es el vector de variables explicatorias para el modelo probit, mientras que γ es el vector para parámetros que son desconocidos. Por otro lado, $\Phi(\cdot)$ es la función de distribución normal

acumulada. La estimación del modelo (3.32) genera resultados que pueden ser utilizados para predecir las probabilidades de cada observación. En la segunda etapa, se corrige la autoselección incorporando una transformación de estas probabilidades pronosticadas como una variable adicional explicativa. La ecuación puede ser especificada:

$$Y = X\beta + u \quad (3.33)$$

donde Y es la variable dependiente que puede ser observada en base a la condición de si es tratado o no. La esperanza condicional de (3.33) sería

$$E[Y|X, T = 1] = X\beta + E[u|X, T = 1]$$

que bajo el supuesto de que los términos de error tienen una distribución normal conjunta:

$$E[Y|X, T = 1] = X\beta + \rho\sigma_u\lambda(Z\gamma) \quad (3.34)$$

donde ρ es la correlación entre los factores determinantes no observados de la propensión a ser tratados y los factores no observables para obtener las ganancias esperadas bajo tratamiento, σ_u es la desviación estándar de los errores, λ es la Razón Inversa de Mills (de la cual se hablará más adelante) evaluada en $Z\gamma$. La ecuación (3.34) puede expresar que la selección de la muestra puede ser vista como una forma de sesgo de variables omitidas, condicionadas tanto a X como a λ , simulando que la selección de la muestra se realizó de forma aleatoria. La ecuación (3.34) puede ser estimada sustituyendo γ con la estimación probit de la primera etapa, elaborando el término λ , e incluir ésta en el modelo como una variable explicativa. Suponiendo que la desviación estándar de los errores es estrictamente positiva, el coeficiente de λ será cero sí y sólo sí $\rho = 0$, por lo que la hipótesis nula de que λ

= 0 es equivalente a probar la selectividad de la muestra (Heckman, Tobias y Vytlačil, 2000).

Heckman, Tobias y Vytlačil (op. cit.), partiendo de la idea de que el problema del "tratamiento" es semejante al problema del sesgo de selección, proponen un método para la estimación de un *ATE* válido para contextos gaussianos. Aunque dicha propuesta no está exenta de limitaciones, su sencillez la convierte en un buen punto de partida para el análisis del efecto de cualquier tratamiento, de tal forma que (3.34) dependerá de la condición de que si *Y* recibe o no recibe tratamiento, por lo que la ganancia sería $G_i = Y_i^T - Y_i^C$. De igual manera, según estos autores, y de acuerdo con los experimentos realizados, a medida que aumenta la muestra, será más difícil distinguir la selección de sesgo.

Sin embargo, comparar entre diferentes modelos, la selección de sesgo puede ser lo menos importante ya que las estimaciones de parámetro de tratamiento serán similares ($ATE = ATT$) en ausencia de selección de sesgo, y controlar la autoselección no puede ser importante para la evaluación del programa dado.

Así, cuando la selectividad no es una característica importante de la información, las estimaciones de parámetros de tratamiento a través de estos modelos serán muy parecidas, y por lo tanto, el problema del modelo de selección se vuelve algo secundario (Ravallion, op. cit.).

En un texto anterior Heckman con Hidehiko Ichimura, Jeffrey Smith y Petra Todd (1998) (HIST), se realiza una serie de ensayos experimentales basados en 1) el método del *matching*, 2) el modelo de comprobación de sesgo de selección, tomando al sesgo como una función de la participación de un programa o política, y 3) el método de diferencia en diferencias basados en programas de adiestramiento de trabajo. En dicho trabajo, se descompone el sesgo bajo tres componentes: a) diferenciando los apoyos de políticas a

partir de los parámetros estimados entre participantes y no participantes; b) diferenciando el comportamiento de la distribución de los grupos en la región de *Common Support*, y; c) la selección de sesgo, el cual se define a partir de los valores comunes que tienen los parámetros estimados tanto en el grupo de participantes como en el grupo de comparación. Las dos primeras partes se pueden dejar de observar en el momento en que se realiza el *matching* en base a las características comunes en ambos grupos; sólo se mantiene la selección de sesgo.

En dicha investigación, con el método de comprobación de sesgo de selección, se mostró la presencia de sesgos más pequeños en el primer componente en comparación a los otros dos. El hecho está en que el método *matching* reduce pero no elimina el sesgo (definido de forma convencional), esto ocurre debido a que esta metodología parte del supuesto de que el sesgo es igual a cero, y en algunas pruebas se puede provocar su incremento. Las pruebas señaladas en dicho artículo, dicen los autores, demuestran el rechazo al supuesto de no sesgo. HIST (op. cit.) señalan que a pesar de los problemas encontrados, el sistema de diferencia en diferencia permite encontrar estimadores más consistentes, pero se requiere una observación en lapsos de tiempo en el cual, a lo largo de este periodo, el sesgo de selección sea el mismo antes y después de entrar al programa y en los mismos intervalos simétricos.

En la descomposición del sesgo en los componentes mencionados, se determina si un estimador de una dimensión pequeña se debe a una combinación fortuita de compensar los sesgos, o si cada uno de los componentes del sesgo tiene poco peso. Fuentes de sesgo debido a las fallas en la región *Common Support*, las discrepancias en el peso y la comparación entre los participantes, dependen del plan de muestreo utilizado para recoger datos para el grupo de comparación y, por lo tanto, son susceptibles de variar a través de

evaluaciones. Los factores de generación de auto-selección son más probables que sean similares a los efectos propios de las evaluaciones (HIST, op. cit.). Los autores concluyen que el conocimiento de los tres componentes facilita la generalización de las pruebas de un estudio a otros entornos, y es más informativa acerca de las fuentes de los sesgos que las medidas tradicionales que se utilizan para resumir esos sesgos. Su descomposición demuestra que, en la información que manejan, la selección de sesgo es más grande en relación con los impactos del programa experimental que se estima, pero es pequeño en relación con la medida convencional de sesgo.

En los dos textos mencionados, Heckman demuestra los beneficios de los ensayos aleatorios. Mientras que el sesgo se reduce utilizando métodos no experimentales por medio del *Common Support* y las ponderaciones, de todos modos no llega a eliminarse. Los experimentos resuelven el problema del *Common Support*, debido a que los grupos de control y de tratamiento, por lo regular, se encuentran dentro de la misma localidad, situación que no ocurriría si las observaciones están dispersas en diferentes localidades. Al encontrarse en la misma localidad, la distribución se puede balancear por medio de las características similares entre ambos grupos (de control y de tratamiento), y pueden ser estimados por medio de un modelo de probabilidades.

Así, la ecuación (3.34) se le conoce como el procedimiento de Heckman de dos etapas. Cabe recalcar que si se estimara dicha ecuación por Mínimos Cuadrados Ordinarios, o incluso si se considera la existencia de heterocedasticidad, los errores estándar reportados estarían equivocados (Cameron y Trivedi, op. cit.), por lo que al utilizar dicha regresión el estimador β resultaría bastante consistente por “*las siguientes razones: (1) Es fácil de aplicar; (2) el enfoque es aplicable a una amplia gama de modelos de selección...; (3) el estimador requiere supuestos más débiles sobre la distribución que el de los de la*

normalidad conjunta de los errores; y (4) estos supuestos de distribución pueden debilitarse aún más para permitir la estimación semiparamétrica” (Cameron y Trivedi: op. cit., pp. 550-551)

Razón Inversa de Mills

En estadística, la Razón Inversa de Mills (IMR), es el cociente de la función de densidad de probabilidad sobre la función de distribución acumulada. La utilización de la IMR está motivada por el supuesto de la distribución normal truncada. Si X es un vector aleatorio donde sus miembros tienen distribución normal y varianza constante (σ^2), entonces se puede mostrar que:

$$E(x|x > \alpha) = \mu + \sigma \left[\frac{\varphi((\alpha - \mu)/\sigma)}{1 - \Phi((\alpha - \mu)/\sigma)} \right] \quad (3.35)$$

y

$$E(x|x < \alpha) = \mu + \sigma \left[\frac{-\varphi((\alpha - \mu)/\sigma)}{\Phi((\alpha - \mu)/\sigma)} \right], \quad (3.36)$$

donde α es una constante, $\varphi(\cdot)$ sigue la función de densidad normal y $\Phi(\cdot)$ la función de distribución normal acumulada. El término entre los corchetes [], es la IMR. Para ambos casos es visible que ambos términos no son negativos de manera simultánea, “...aunque sí ser ambos positivos o uno de cada signo. En el primer caso (3.35), en que ambas esperanzas condicionales superan a la esperanza no condicional en toda la población, se está ante un ordenamiento no jerárquico y ambos grupos serían estándares en la nomenclatura de Heckman. Si uno de los grupos no es estándar, (3.36) esto podría ser

considerado un ordenamiento jerárquico inferior y tendría sentido decir que capta a los menos beneficiados dentro de la población” (Rodríguez, 2004).

La IMR se plantea en los análisis de regresión para considerar la existencia del sesgo de selección. Si una variable dependiente es censurada, esto causa una concentración de observaciones en valores cero. Una variable dependiente se dice que es censurada cuando no siempre se muestra una observación cuyo resultado pueda ser positivo⁵. Este problema fue dado a conocer por James Tobin en 1958 quien señaló que si esto no se considera en la estimación OLS, se producirán parámetros estimados sesgados (Heckman, Tobias y Vytlačil, op. cit.). Con variables dependientes censuradas, hay violación en los supuestos de Gauss-Markov para mínimos cuadrados, puesto que la correlación entre las variables independientes y el término de error es diferente de cero. Por dicha razón, James Heckman, en diversos documentos, propone la utilización de la IMR para tener en cuenta la existencia del sesgo de selección.

En la primera etapa se corre una regresión para observar un resultado (que se espera positivo) de la variable dependiente que se modela por medio de la regresión probit. Los parámetros estimados son utilizados para calcular la IMR, la cual es incluida como una variable explicatoria adicional en la estimación de OLS. En la segunda etapa se estima la ecuación por Mínimos Cuadrados Ordinarios, incluyendo como regresor a la IMR; si se rechaza la hipótesis de que su coeficiente es cero puede afirmarse que efectivamente existe sesgo de selección.

Por lo regular, los modelos con errores normalmente distribuidos casi no se pueden identificar, por lo que al realizar la ecuación (3.34) el término de la IMR es

⁵ Esta situación es propia de los modelos de selección de la muestra como señalan Cameron y Trivedi (2009)

aproximadamente lineal a lo largo de su amplia gama de argumentos. Pero esto puede llevar a un problema de multicolinealidad. La multicolinealidad puede detectarse de manera sencilla al estimar el modelo con un subconjunto de las variables que tienen menos probabilidades de ser colineales, para el caso, serían los regresores β y $\lambda(Z\gamma)$ de la ecuación (3.34). “El problema es menos grave cuanto mayor es la variación de $(Z\gamma)$ a través de las observaciones, es decir, un modelo probit puede discriminar entre los participantes y no participantes.” (Cameron y Trivedi: op. cit., p. 551)

Aquí cabe abrirse una pregunta ¿porqué es importante el concepto del sesgo de selección? La respuesta inmediata la otorga Wooldrige al señalar que una ecuación como la (34) prueba que al utilizar mínimos cuadrados ordinarios, cuando $Y > 0$, no se muestra consistencia en los parámetros estimados, por lo que la IMR es una variable omitida que por lo general está correlacionada con el resto de las variables explicatorias (Wooldrige, 2009).

Para el caso propio de la tesis, en el proceso de la evaluación de impacto, este concepto es importante debido a que en los programas sociales existe un subconjunto de población con respuesta potencial a dicho proyecto público (Heckman, Tobias y Vytlačil, op. cit.), el cual puede afirmar la existencia de sesgo de autoselección, en este caso.

Evaluaciones a PROCAMPO y críticas

En los últimos años, se han desarrollado una serie de estudios sobre los programas sociales aplicados en México, en especial sobre Oportunidades (antes PROGRESA) y PROCAMPO.

Iniciando precisamente con el tema de análisis, se encuentra un estudio sobre el comercio internacional relacionado con el maíz a partir del PROCAMPO. En dicho estudio

de 2001, José Alberto García-Salazar, estudia las elasticidades de la producción de maíz en el Norte, Centro y Sur del país para los dos ciclos anuales: Primavera-Verano y Otoño-Invierno, tomando en cuenta la producción, consumo y precio del maíz y del frijol. También considera factores como las lluvias y la disponibilidad de agua en ciertas regiones. El problema de este estudio radica en que el periodo de análisis es de 1970 a 1996, cuando PROCAMPO inicia en 1994. Por esa razón, la conclusión resultante de este estudio es que la relación entre la producción de maíz y los pagos de PROCAMPO es inelástica, por lo que no representa ser un programa efectivo para aumentar la producción de maíz. A pesar de estos resultados, y de acuerdo con las variables sobre producción y consumo, el manejo de mínimos cuadrados de dos etapas resulta ser adecuado.

Otro trabajo es el de Marta Ruiz-Arranz con Benjamin Davis, Marco Stampini, Paul Winters y Sudhanshu Handa (2002), en el que se analiza el impacto de los programas Oportunidades y PROCAMPO en la seguridad alimentaria. Se trata de analizar los aspectos de la variedad de la producción rural con el consumo de calorías. El modelo a primera vista parece ser muy bueno, pero de forma detallada se basa sólo en el consumo de alimentos, al igual que el consumo de calorías, y en el modelo de consumo se implica al PROCAMPO, cuando este tiene como fin el incrementar el ingreso de los productores rurales. La medición realizada a través del método de Variables Instrumentales (*IV*), no queda clara si se debe efectivamente a la posibilidad de encontrar problemas de correlación serial, o que las variables explicativas resultan ser estocásticas. Los resultados de los R^2 , ya sea por OLS o por *IV*, raramente resultan ser mayores a 0.5. Los aciertos de este texto se encuentran en el análisis del consumo proporcional de bienes alimenticios por hogar, lo cual requiere una sólida base de datos.

En ambos casos, tal vez hubiese sido interesante estudiar la variable de precios de los principales productos agrícolas que se destinan al consumo humano con sus respectivas productividades.

En términos de Evaluación de Impacto, utilizando una metodología similar, se encuentran las que realiza el Grupo de Economistas y Asociados (GEA), por cada periodo de tiempo. Se aplica la metodología del PSM, pero la búsqueda del contrafactual no queda del todo clara. Por ejemplo, en la evaluación del periodo Primavera-Verano 2006, de su encuesta obtuvieron 3,127 observaciones de los cuales 2,826 eran beneficiarios del programa PROCAMPO, mientras que el resto no. Sin embargo, el modelo se corre con 3,043 observaciones, sin argumentar si son datos que se pierden al correr el modelo econométrico, ya que se menciona que toda la muestra tiene las mismas características en términos del tamaños del predio, régimen hídrico, tenencia de la tierra, posesión de la tierra (propia o rentada), edad del jefe de hogar, género del jefe de hogar, escolaridad promedio y tipo de cultivo.

Al fin de cuentas, es dudable que las 3,127 observaciones tengan el mismo género, o fluctúen en un rango de edad similar y/o tengan rangos cercanos del nivel escolar, aun teniendo como características similares el resto de las variables, es difícil encontrar este tipo de similitudes en los hogares rurales. Sin embargo, no hay que descartar su metodo de medición, ya que esta puede ser utilizada cuando la información recabada pueda tener problemas desde el proceso de recopilación de datos.

Hay que considerar otros dos aspectos importantes de las evaluaciones del GEA. En primer término, son evaluaciones en el cual tratan a todos los hogares por igual, y como se ha visto en el capítulo anterior, no todos los productos agrarios que entran en el programa han mostrado crecimiento, por lo que hubiese sido más conveniente realizar dicha

evaluación de acuerdo a cada bien producido, puesto que esto hubiese implicado si los incrementos y/o caídas de la producción de cada uno de los bienes se deben al grupo de beneficiarios, o al grupo de no-beneficiarios, o en todo caso, a otra serie de factores. El hecho de tratar a todos los productores de PROCAMPO por igual, y sin distinción de bien producido, sólo genera la sospecha de que dicha evaluación sólo se dirige al plano informativo, dejando de lado aspectos cualitativos que la misma evaluación genera. Como prueba, en la evaluación realizada en el 2005 dice, posterior a la evaluación econométrica “...*De manera evidente a la luz de estos resultados PROCAMPO cumple con el objetivo fundamental de incrementar los ingresos de los productores, no sólo por el subsidio directo que otorga, sino por la vía de generar un valor mayor de la producción de los beneficiarios en comparación con un estado en el que no lo recibieran...*” (GEA: 2005, p. 187). En otras palabras, la evaluación es bastante general sin especificar productos ni estudiar los efectos paralelos que un programa de apoyos como PROCAMPO conlleva.

En segundo lugar, tratan de establecer la metodología de James Heckman a través de la herramienta econométrica “*treatreg*”, el cual sí demuestra un avance para el desarrollo de este proceso de evaluación. En la presente tesis, se utiliza directamente la herramienta “*heckman*”, del mismo programa STATA, aunque no se deja de lado el efecto de tratamiento que realmente mide el comando “*treatreg*”.

Una de las evaluaciones más completas que se han realizado a PROCAMPO, es la que realiza en 2004, la misma Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA)⁶, puesto que analiza diversos indicadores, y no sólo la

⁶ Por ley, las instituciones públicas relacionadas con el programa no pueden realizar las evaluaciones correspondientes. A lo que SAGARPA, a través de ASERCA, están facultados, es en la supervisión de las

producción, sino que también analiza los aspectos relacionados con el acceso a insumos y bienes de capital. Pero al igual que la evaluación del GEA, la evaluación revisa a los productores beneficiarios del programa en lo general, por lo que se pierden también las particularidades por cada bien producido.

La evaluación presentada por SAGARPA (donde no hay participación GEA) incluye el aspecto referente a que muchos de los recursos son utilizados para el consumo doméstico. *“De acuerdo con los datos de la encuesta, los productores beneficiarios del Programa destinan un 29.5% del apoyo, a los diferentes conceptos que integran el consumo doméstico...De los recursos del PROCAMPO que los beneficiarios entrevistados destinan al consumo doméstico, el principal rubro lo ocupa el gasto en alimentos...En segundo término, los productores de todos los estratos manifestaron gastar el apoyo PROCAMPO en educación y en vestido...”* (SAGARPA: 2004, p. 89). Este hecho implica precisamente que no sólo se analice la evaluación sobre la producción, sino que también se implique esta evaluación sobre los gastos en bienes básicos, sobretodo en alimentos, de igual forma, queda establecido que si los subsidios de PROCAMPO pueden ser utilizados de la forma que más convenga a los productores, entonces también se mantiene la idea que la producción final también puede ser utilizada de la forma que le corresponda al productor, por lo que la evaluación realizada en el presente trabajo también se vincula al autoconsumo doméstico.

Cabe señalar que las evaluaciones presentadas por SAGARPA, a través de ASERCA, tienen características muy similares a la comentada, sin embargo, se ven muy

encuestas que hacen las entidades contratadas por los organismos involucrados, así como en la publicación de las mismas como evaluaciones externas, pero con los nombres de las entidades gubernamentales.

claras las diferencias existentes entre cada una de ellas. Para empezar, los tamaños de muestra difieren bastante año con año, situación que reconoce el organismo en la evaluación publicada en 2008, por ejemplo, en 2002 se encuestaron a 45,525 personas, pero para 2003 y 2004 se evaluaron a 3,478 y 4,101; ya se comentó sobre el documento de 2006. De igual forma, las evaluaciones mostradas en su página oficial de internet, para el año 2005 se muestra una evaluación realizada por GEA.

De las evaluaciones de SAGARPA, cabe mencionarse, el trabajo mostrado en el año 2003 resulta de lo más interesante, en primer lugar porque se estratifica por tamaño de predios cultivados, mostrando que efectivamente existen cambios en el nivel de ingreso, además de que aminoran los efectos negativos que se sufren en diversas zonas rurales (por ejemplo, la migración).

Otra evaluación que llama la atención es la que se publica en 2004 (que ya se mencionó más arriba), pues a diferencia de la realizada en el año anterior, en lugar de mostrar las mejoras de acuerdo a la jerarquización por tamaño de predio, el estudio se realiza por cultivo, mostrando que es el maíz el producto que mayormente ha crecido gracias al PROCAMPO, por encima de los demás cultivos beneficiados por el programa.

Las dos evaluaciones, cabe recalcar, resultan ser bastante completas, pues ambas abarcan cuestiones sobre consumo productivo, acercamiento al mercado, capitalización, protección de recursos naturales, etc., mostrando mejoras en esos rubros, aunque señalando las preocupaciones sobre ciertos aspectos que pueden afectar dichos avances. Otros indicadores utilizados en ambas evaluaciones se encuentran: superficie apoyada, recursos presupuestales, número de productores beneficiados, predios apoyados, agricultura por contrato. De igual forma, se consideran situaciones operativas como expedientes, entrega de apoyos, credenciales, cumplimiento de requisitos, etc. La diferencia en este sentido

radica en que la segunda considera aspectos como los beneficios a ciertos grupos sociales como son indígenas, mujeres, adultos mayores y grupos en vulnerabilidad.

Evaluaciones posteriores publicadas por los organismos públicos involucrados al programa, a pesar de que mencionan resultados positivos debido al programa, y de que la metodología mejora (pasando de primeras diferencias a otras metodologías como la R&R), pasan a ser estudios muy generalizados sin considerar lo publicado en 2003 o en 2004.

A pesar de que dichas evaluaciones buscan objetivos similares, la forma de recaudar información implica analizarlos por vías diferentes. Sin embargo, en posteriores evaluaciones, la finalidad se vuelve bastante general, ejemplo de ello es lo que se dice en la evaluación publicada en 2007 que dice: *“Evaluar mediante trabajo de gabinete y con apoyo de información proporcionada por el Programa de Apoyos y Servicios a la Comercialización Agropecuaria (ASERCA) la consistencia del Programa de Apoyos Directos al Campo (PROCAMPO) en cuanto a su diseño, planeación estratégica, cobertura y focalización, operación, percepción de la población objetivo y resultados. Lo anterior con el fin de que las recomendaciones de este análisis puedan retroalimentar el diseño y la gestión del Programa”* (ASERCA, 2007). Con esto se deja de hacer evaluaciones con el detalle realizado en años anteriores, siguiendo la propuesta metodológica-cualitativa realizada por GEA.

En términos generales, la secretaria ha tenido, a lo largo de sus evaluaciones, dos conclusiones básicas, la primera de ellas consiste en que PROCAMPO es el mecanismo más eficaz para transferir los apoyos monetarios a los beneficiarios. En la segunda conclusión del organismo, aunque reconoce que no se ha llegado al proceso de capitalización del campo, por lo menos ha aminorado las fallas del mercado rural.

Tal vez, por eso mismo, las evaluaciones presentadas por la secretaría a través de ASERCA, no se han librado de las críticas, en el sentido de que hay problemas de diseño (Duran, et. al.: 2007), y obviamente cabría pensarse que también en los resultados. Cabe señalar que para las evaluaciones publicadas en 2002 y 2003 se contrataron los servicios del Colegio de Ingenieros Agrónomos de México A.C (CIAMAC), mientras que para la publicación de 2004, se contó con los servicios de Universidad Autónoma de Chapingo (UACH). Para las evaluaciones posteriores a las de GEA no se menciona qué organismo o quién realiza los resultados publicados. Esto lleva por consecuencia dudas en los resultados posteriores a 2006, en el sentido mencionado.

A título personal, independientemente si los resultados fueron hechos a conveniencia de SAGARPA, al no considerar las mismas características, de un año a otro, y a pesar de que los objetivos son similares, se está evaluando bajo condiciones diferentes, por lo cual los resultados no podrían ser comparables. Por un lado, y coincidiendo en parte con Durán, et. al. (op. cit.), el estudio realizado por el CIAMAC sólo queda en plano de auditoría basados en la percepción de los beneficiarios en términos de sus ingresos y otros aspectos socioeconómicos como predios apoyados y migración; por otro lado, la aplicación de la UACH implicó un mayor número de aspectos, por lo que el manejo de la información cuantitativa fue diferente, y las conclusiones estuvieron mejor sustentadas.

Las evaluaciones realizadas por GEA, a pesar de ser bastante generales y, como ya se ha mencionado, utilizan las herramientas instrumentales basadas en Heckman, lo cual permite revisar los resultados desde otra perspectiva diferente con respecto a las anteriores. Las evaluaciones posteriores a GEA cayeron en el plano de la auditoría, pero a diferencia de la realizada por el CIAMAC, sólo establecen resultados en el aspecto de los ingresos percibidos por los beneficiarios del programa, sin dejar bien claro el proceso de medición.

Otra investigación es la realizada por George A. Dyer (2007), publicada por el Centro de Estudios para el Desarrollo Rural Sustentable y Soberanía Alimentaria (CEDRSSA). El objetivo principal fue revisar los efectos indirectos de las políticas de transferencia en el sector, es decir, no sólo aplica el PROCAMPO, sino que también lo vincula con Oportunidades. Utilizando la información de ENHRUM y la metodología de evaluación de Heckman, señala que “... *tanto la cobertura de PROCAMPO como de los montos transferidos son menores para los productores de autoconsumo que para otros grupos. Nuevamente, dicha distribución no tiene fundamento en los objetivos del programa. Tanto PROCAMPO como Oportunidades propician la inversión productiva en grupos sin acceso al crédito. No obstante, la proporción de recursos que se destinan a la inversión es reducida. Posiblemente menos del 2.5% de los recursos del PROCAMPO tradicional se invierte en la producción rural...*” (Dyer: op. cit., p. 43)

Las transferencias de PROCAMPO varían de acuerdo al tamaño del predio. La evaluación realizada por Dyer (op. cit.) señala que, los poseedores de más de 5 hectáreas tienen un impacto mayor sobre el nivel de ingreso, en comparación de aquellos que tienen menos de esa cantidad. Además, si por un lado el programa ayuda en 2.7% en el incremento del recursos monetarios, por otro lado, sólo el 10% de estos se dedica a la inversión de actividades productivas.

Si bien los resultados otorgados por G. Dyer son muy interesantes, lo que realiza es una evaluación conjunta de PROCAMPO con Oportunidades, por lo que los impactos registrados se pierden entre uno y el otro programa. Al igual que el GEA, la evaluación realizada se basa en todas las observaciones, sin distinguir el tipo de cultivo, situación que lo deja en el plano meramente cualitativo, destacando el papel del maíz en los beneficiarios de ambas políticas, aunque propone un replanteamiento de los mismos.

El que aporta mucho a la crítica, y en especial a estas dos evaluaciones mencionadas⁷, es Pedro Durán (et. al., op. cit.) quien establece la duda sobre las conclusiones optimistas a esta política al campo, y la discusión empieza debido al acercamiento de ASERCA en la aprobación de los institutos y publicación oficial de los resultados. Este investigador establece su desconfianza con respecto a las evaluaciones mencionadas, pues señala que “... *La amplia gama de indicadores y de objetivos ha dado lugar a que las evaluaciones arrojen, al parecer, expectativas favorables en torno a la continuidad del programa...*” (Durán, et. al.: op. cit., p. 8).

José María Larrú (2003), afirma que si el programa es ambicioso, las evaluaciones quedan en el plano de lo general y menor es el impacto real del programa, debido a que, y como señala en un texto posterior: “*Muchos de los contenidos que en los informes de evaluación aparecen como juicio del impacto, en realidad son sólo impresiones subjetivas o —en el mejor de los casos— reconocimiento de que no hay información suficiente para poder realmente pronunciarse sobre el impacto de una intervención. Los principales problemas para juzgar el impacto son la atribución y la agregación*” (Larrú y Lugris: 2008, p. 6).

Este es prácticamente el argumento que utiliza Durán, et. al. (op. cit.), para criticar las evaluaciones realizadas y los resultados “subjetivos” a los que llegan, puesto que las evaluaciones, según el investigador “... *han tendido a confirmar la eficiencia operativa de PROCAMPO y a justificar su continuidad, usando para ello diversos indicadores ... dando*

⁷ A título personal, dentro del contexto de las evaluaciones, las realizadas por el Colegio de Ingenieros Agrónomos de México A.C y Universidad Autónoma de Chapingo, resultan ser las más completas en término de los objetivos particulares a lograr y la estructura metodológica, por lo que tienen suma atención en este trabajo, así también las críticas realizadas por Pedro Durán y Mauricio Merino.

lugar a que las evaluaciones arrojen, al parecer, expectativas favorables en torno a la continuidad del programa, (pero que) en concreto, las evaluaciones, tal y como fueron hechas, no aportan evidencia concluyente respecto al impacto multiplicador en la actividad productiva y la calidad de vida de los beneficiarios...” (Durán, et. al.: op. cit., p. 13)

A pesar de que el objetivo de la crítica de Durán es llegar a una propuesta de metaevaluación, ésta sólo implica resultados meramente analíticos sin llegar a una proposición concreta. A pesar del buen sustento de su análisis, e ideas bastante claras, las propuestas quedan sólo en un plano de lo positivo sin ofrecer ideas para la acción cuantitativa. Dentro de sus exhortos, se puede subrayar la idea de realizar una evaluación manteniendo la estratificación de la UACH, pero dividiendo el estudio por tipo de cultivo, y acercándose a Entidades Federativas y productos representativos como maíz y frijol.

Lo que sí se puede señalar abiertamente es que los indicadores que propone ASERCA, y en los que se basan las evaluaciones, no permite una valoración eficaz sobre los objetivos propios del programa, en términos de la capacidad de capitalización de las unidades productivas. Otro aspecto a consideración, es que las conclusiones juzgan que las mejoras en la calidad de vida se deben a la existencia del programa, por lo que hay sospecha de que hay factores que no se están considerando y que son los que están dando los efectos positivos.

Aún sin ser una evaluación, pero que cabe la pena mencionar, es el trabajo de Mauricio Merino, investigador del Centro de Investigación y Docencia Económicas, A. C. (CIDE), publicado en 2009, el cual conjunta una serie de investigaciones anteriores sobre PROCAMPO. En su análisis, no cuestiona los resultados de las evaluaciones, sino los resultados de la política, pues no se han compensado las asimetrías con respecto a los

vecinos del norte, además señala que “... *Dentro de los mayores obstáculos que enfrenta la agricultura mexicana está su baja competitividad, entendida como la capacidad de un sector de mantenerse o expandirse. A pesar del dinero público invertido, México es altamente dependiente de las importaciones agroalimentarias. En varios países –Japón, Estados Unidos y los miembros de la Unión Europea, por ejemplo–, la soberanía alimentaria es un tema de seguridad nacional. En nuestro país, en cambio, la balanza comercial agroalimentaria refleja un deterioro creciente en el valor de los flujos comerciales; 1995 fue el último año en el que México registró superávit en el rubro, y 2007 fue, al menos desde 1990, el año más deficitario ...*” (Merino: 2009, p. 25), por lo que se ha perdido competitividad.

Si los resultados de las evaluaciones resultan favorables, señala Merino, es porque PROCAMPO se convirtió en una medida de estabilidad política y económica para el campo; o en otras palabras, este programa es una renta estatal del cual dependen más de dos millones y medio de campesinos. Otra razón, es porque la obtención de la información llega a ser poco precisa (Merino, op. cit.). La distribución de los recursos ha abierto la pregunta sobre la capitalización del campo “... *contrastando los estados que más apoyos han recibido de PROCAMPO... con aquellos que más beneficiarios tienen registrados, se observa una clara fragmentación regional... los datos hacen suponer que la política de apoyo al campo ha seguido una doble estrategia. Por un lado, los estados ricos del norte-occidente deberían incrementar su competitividad, producción y productividad; por otro, los del sur-centro, aquejados por una marcada pobreza, deberían reducirla...*” (Merino: op. cit., p. 30).

El investigador del CIDE señala que, si PROCAMPO ha continuado, es porque existe el compromiso de las instituciones públicas con los beneficiarios, y otra serie de

compromisos que los gobiernos federal y estatales no han querido romper, generando conflicto entre estos y con los organismos sociales (Merino, op. cit.)

Pero las críticas a las evaluaciones oficiales no sólo provienen del interior. En la página de la Organización de Naciones Unidas para la Agricultura y la Alimentación (FAO), se señala que los análisis oficiales realizados a PROCAMPO carecen de seriedad pues su uso es bastante limitado. Además señalan que la base de datos que se utiliza antes de 2001 es la misma que se utiliza para evaluar Oportunidades (Davis, 2004), por lo que sus conclusiones muestran una percepción limitada.

La evaluación realizada por Benjamin Davis⁸ (op. cit.) señala que los traspasos monetarios del PROCAMPO generan consecuencias indirectas entre los beneficiarios, pues se detectan efectos multiplicadores de entre 1.5 a 2.6 pesos por cada peso transferido. Así, por cada actividad productiva del hogar, el ingreso se ve multiplicado por dicha cantidad por cada peso de la transferencia, reflejando el costo de oportunidad del ingreso marginal no realizado por su restricción presupuestaria, y que se ve limitado por dicho proceso.

El investigador del texto mencionado señala que “... *las transferencias del PROCAMPO crearon un efecto multiplicador positivo en el ganado, puesto que cada peso de las transferencias generó 0.28 pesos de ingreso por concepto de ganado. De la misma forma, las transferencias del PROCAMPO crearon un efecto multiplicador positivo en la agricultura, ya que cada peso generó 0.33 pesos de ingreso agrícola. La disponibilidad de asistencia técnica y la propiedad de tierras de regadío aumentan el multiplicador de ingreso agrícola, mientras que la disponibilidad de crédito lo disminuye. Por ende, los*

⁸ Con respecto a la metodología de Benjamin Davis (op. cit.) realmente no queda del todo clara, pues no señala como determina su muestra y qué método econométrico es el que utiliza para obtener los resultados aquí señalados.

hogares que controlan más tierras de regadío y tienen acceso a la asistencia técnica y sin tener acceso al crédito, aprovechan al máximo este efecto multiplicador...” (Davis, op. cit.)

En torno a las críticas, la construcción de un contrafactual correcto (o grupo espejo como le llaman algunos autores) es un desafío técnico, que incluso las evaluaciones realizadas por Ravallion y Heckman no han podido superar; situación que ha sido reconocida por los mismos autores. Por un lado, si bien las críticas de Durán, Merino y de Benjamin Davis, tienen sustentos válidos, por otro lado, se tendría que considerar la postura de la SAGARPA (aunque es obvio que su defensa sobre los resultados positivos de los análisis publicados se mantiene para la continuidad del programa), por lo que abre la pregunta ¿qué miden las evaluaciones “críticas” y qué miden las “oficiales”? Queda claro que las publicaciones de SAGARPA van en torno al objetivo principal de PROCAMPO que es el incremento del ingreso de los beneficiarios, mientras que las conclusiones de los críticos están dirigidas a los efectos colaterales que conlleva la política y allí la disyuntiva entre unos y otros.

En una publicación reciente, José Alberto Zarazúa (et. al., 2011), mencionando sobre una evaluación publicada en 2008 por la FAO, en la cual el organismo destaca “... *que se ha cumplido con el objetivo principal de incrementar el nivel de ingreso de sus beneficiarios, a pesar de que el apoyo recibido del Programa no se destina en su totalidad a las actividades productivas. De hecho, la participación de PROCAMPO en el ingreso derivado de la actividad agrícola se ha incrementado, a nivel nacional; de 3.45% en el año 2005 a 20% en 2007, porque los beneficiarios reciben ahora mayores ingresos de otras actividades, que de la agrícola...*” (Zarazúa, et. al.: op. cit., p. 92).

El mismo estudio del investigador del Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional-IPN, tiene como fin revisar los efectos colaterales del PROCAMPO, tomando como muestra a los productores del Estado de México (se consideraron 200 encuestas), estratificando por hectárea. Sin embargo, los resultados no pueden considerarse de impacto, pues se utilizó un proceso comparativo acuerdo a la jerarquización señalada. La conclusión reafirma que, en lo que se refiere a la mejora del ingreso, el resultado es positivo, pero a lo referente a la capitalización desde el punto de vista del conocimiento productivo y adopción de innovaciones tecnológicas y de organización, no se ha cumplido este aspecto, y menos en aquellos que poseen pocas hectáreas. De igual forma, no se ha logrado la reconversión productiva de hortalizas, frutas y flores hacia la producción de granos.

En esto último, también coincide la Organización para la Cooperación y el Desarrollo Económico (OCDE), pues “... *no es tan redistributivo como programa de combate contra la pobreza.*” (OCDE, 2006). Si bien el trabajo-resumen de la OCDE no es propiamente una evaluación al programa, cabe señalar que sus conclusiones sobre el cambio de dirección al programa, según el organismo, están basadas en la revisión de las publicaciones oficiales como de las críticas a los mismos.

Actualmente (2011), el Banco Interamericano de Desarrollo (BID) está implementando una encuesta propia para poder evaluar al PROCAMPO. El planteamiento que hace el organismo es que, si bien el objetivo de las transferencias es que no genere desviaciones en el mercado, no se ha demostrado la existencia de esa neutralidad en programas similares, pues se pueden encontrar cambios, ya sea en las actividades productivas o reconversiones de las mismas, al igual que en el tamaño de los predios u otros aspectos (BID, 2010).

La metodología de esta evaluación está basada en el “Modelo de Hogares Agrícolas” (al igual que la realizada por la ENHRUM), debido a que, por un lado, las decisiones de consumo son inseparables de las de producción, y por otro lado, el hogar se comporta como unidad de producción. Para plantear este tipo de aspectos, entonces se tienen supuestos basados en restricciones financieras como acceso al crédito y la ausencia de ahorro, la existencia de autoconsumo y los tamaños de los predios no cambian. (BID, op. cit.).

La idea básica está basada en el concepto de maximización de la utilidad del hogar, pues bajo la restricción presupuestaria, estaría diciendo que al no existir limitaciones financieras, el consumo de los miembros del hogar estaría basado menos en autoconsumo y más al acceso al mercado. Así, mientras existan más limitaciones financieras, tanto las Relaciones Marginales de Sustitución, tanto en el consumo como en la producción, son más elevados que los costos relativos, o en otros términos, el acceso al crédito, al ahorro y la posesión de efectivo tienen las mismas limitaciones.

De lo último, el BID (op. cit.) considera que PROCAMPO funciona para amortiguar las limitantes financieras de los productores rurales considerados, y por otro lado, puede provocar incremento en la demanda de bienes de consumo y de insumos. Sobre lo último se considera que la compra de insumos está precedido por un incremento de la producción, aunque como ya se ha señalado, la utilización de este ingreso es discrecional (consumo, inversión agrícola, inversión no agrícola, etc.).

Por tanto, la hipótesis del organismo para la evaluación que busca efectuar a partir de la encuesta que está realizando, se encuentra en que los impactos del programa se focalizan en las inversiones productivas, más que en el gasto del consumo, pues PROCAMPO estaría asociado con la ausencia crediticia de los favorecidos. Así que los

efectos estarían vinculados con las inversiones agrícolas, no agrícolas y la compra de insumos, por lo que existirían cambios en la competitividad de los beneficiarios.

Por último, de acuerdo al planteamiento del organismo internacional, el mecanismo de evaluación está basado en el *Propensity Score Matching* (PSM), combinado con primeras diferencias. El documento mencionado promete publicar un estudio preliminar en Octubre de 2011 y la evaluación final dos años después.