



**“IDENTIFICANDO EFECTOS DE TRATAMIENTO AL
COMBINAR BASES DE DATOS CON FACTORES NO
OBSERVABLES”**

**Trabajo de Investigación presentado
para optar al Grado Académico de
Magíster en Economía**

Presentado por

Sr. Gonzalo Javier Rivera Gallegos

Asesor: Profesor Pablo Augusto Lavado Padilla

2015

Resumen ejecutivo

En el presente documento se considera la identificación de efectos de tratamiento cuando las variables de resultado y las de control no se encuentran en una misma base de datos. Los modelos de inferencia ecológica, en los cuales se combina información agregada de las variables de resultado con información demográfica a nivel individual, son un ejemplo común de esta clase de situaciones. En este contexto, no es posible identificar de manera puntual las distribuciones contra-factuales ni, por tanto, los efectos de tratamiento. Sin embargo, estudios recientes proveen límites inferiores y superiores para identificar de manera parcial los efectos causales. A diferencia de estos trabajos, se adoptará el supuesto de selección en no observables, es decir, la asignación no es totalmente aleatoria hasta que se controle por la heterogeneidad no observada constante en el tiempo. Se utilizan datos de panel en un modelo lineal en los factores no observables para obtener la identificación. Para ilustrar el funcionamiento de los límites, se complementa el análisis con una simulación.

Índice

| | |
|---|-----------|
| Índice de tablas..... | iv |
| Índice de anexos..... | v |
| | |
| Capítulo I. Introducción..... | 1 |
| | |
| Capítulo II. Marco teórico..... | 4 |
| | |
| Capítulo III. Identificando efectos de tratamiento..... | 8 |
| 1. Supuestos..... | 8 |
| 2. Distribuciones marginales, contra-factuales y efectos de tratamiento | 10 |
| | |
| Capítulo IV. Identificando efectos de tratamiento con información ecológica..... | 14 |
| 1. Dicotomizando el crecimiento de los resultados potenciales..... | 15 |
| 2. Trabajando directamente con el crecimiento de los resultados potenciales | 17 |
| | |
| Capítulo V. Simulación..... | 21 |
| | |
| Conclusiones y recomendaciones..... | 23 |
| Conclusiones..... | 23 |
| Recomendaciones..... | 23 |
| | |
| Bibliografía..... | 25 |
| | |
| Anexos..... | 29 |
| | |
| Nota biográfica..... | 36 |

Índice de tablas

| | | |
|----------|--------------------------------------|----|
| Tabla 1. | Rendimiento de los estimadores | 22 |
|----------|--------------------------------------|----|

Índice de anexos

| | | |
|----------|--|----|
| Anexo 1. | Funciones de distribución potenciales a partir de los datos disponibles..... | 30 |
| Anexo 2. | Desigualdad de Cambanis-Simons-Stout (CSS)..... | 32 |
| Anexo 3. | Aplicación de la desigualdad de CSS al modelo transformado..... | 33 |
| Anexo 4. | Teorema 3.2 de Fan <i>et al.</i> (2014a)..... | 34 |
| Anexo 5. | Estimadores de los límites de promedios contra-factuales..... | 35 |

Capítulo I. Introducción

Es muy común que los investigadores tengan la necesidad de combinar datos presentes en distintas fuentes de información. En efecto, una base de datos normalmente no contiene toda la información ni abarca todas las variables relevantes que el investigador desea utilizar por diversos motivos, entre ellos, uno de los más importantes, los altos costos de hacer cuestionarios extensos. No obstante, el potencial aprovechamiento de la información extraída de las distintas fuentes de información es muy grande, especialmente en el campo de la evaluación de impacto.

En este documento considero cómo identificar las distribuciones contra-factuales y los efectos del tratamiento cuando las variables de resultado y las variables de control se encuentran en distintas fuentes de información. Para ello, obtengo la identificación a través de la construcción de límites finos sobre la base de los resultados obtenidos por Fan *et al.* (2014a, 2014b), quienes trabajan bajo el supuesto de selección en observables. Relajaré este último supuesto permitiendo selección en factores no observables, de modo que la asignación al tratamiento no es aleatoria a menos que se controle tanto por la heterogeneidad observada como por la no observada.

Para lidiar con estos factores no observables utilizo datos de panel. En particular, trabajo en un contexto en el cual existe heterogeneidad no observada constante en el tiempo que afecta de forma lineal tanto a los resultados potenciales como a la participación en el programa¹. De este modo, es posible realizar una transformación al modelo (por ejemplo: primeras diferencias) que elimine el efecto de la heterogeneidad no observada y permita “volver” al supuesto de selección en observables, donde los límites desarrollados por Fan *et al.* (2014a, 2014b) son válidos.

En el campo de la evaluación de impacto, muchos estudios han investigado la identificación e inferencia del efecto del tratamiento cuando las variables de resultado y las demográficas son observadas en una única base de datos bajo el supuesto de selección en observables (ver Chernozhukov *et al.* 2013; Hirano *et al.* 2003; y Rothe 2012). En este contexto, como mencionan Fan *et al.* (2014a), las distribuciones marginales y contra-factuales de los resultados potenciales (y, por tanto, los efectos de tratamiento) están identificadas de manera puntual.

Sin embargo, el supuesto adoptado de independencia condicional en variables observables no siempre es adecuado. Existen muchos casos en los cuales dicho supuesto es violado, lo cual

¹ Ver Klevmarcken (1982) o Angrist y Krueger (1995) para la obtención de estimadores consistentes mediante el uso de variables instrumentales cuando no se observan las variables relevantes en una misma base de datos.

ocasiona que las variables de tratamiento sean endógenas debido a heterogeneidad no observada. Para lidiar con este problema, la literatura ha desarrollado distintos métodos, siendo los más comunes el uso de variables instrumentales (Heckman *et al.* 1997) y datos de panel vía la estimación de efectos fijos (Lillard y Willis 1978; Hislop 1999; Kahn 2007).

El presente trabajo se basa sobre los resultados obtenidos por Fan *et al.* (2014a), quienes adoptando el supuesto de selección en observables utilizan representaciones explícitas de la distribución marginal y contra-factual en función de la información disponible ponderando los datos con la inversa del *propensity score*; y junto con la desigualdad de Cambanis-Simons-Stout (ver Cambanis *et al.* 1976) obtienen límites finos de las distribuciones contra-factuales y de los efectos de tratamiento. Existe literatura reciente en el campo de la evaluación de impacto que ha utilizado la idea de acotar distribuciones mediante desigualdades para identificar efectos de tratamiento (ver Frank *et al.* 1987; Fan y Park (2012, 2010, 2009); Heckman *et al.* (1997); Fan y Zhu (2009), quienes adoptan el supuesto de selección en observables; y Jun *et al.* (2014), quienes utilizan datos de panel para lidiar con el supuesto de selección en no observables). Sin embargo, todos estos trabajos asumen que las variables relevantes se encuentran en una misma base de datos, situación ideal pero que no siempre ocurre en la práctica, por lo que la idea de utilizar límites para acotar distribuciones es muy útil en esta clase de situaciones.

La literatura relacionada a la combinación de distintas bases de datos en el campo de la evaluación de impacto, aun siendo pequeña, muestra un crecimiento importante en los últimos años. Ridder y Moffitt (2007) discuten el uso de las desigualdades de Frèchet-Hoeffding al combinar dos bases de datos distintas cuando no hay intersección de individuos entre ambas fuentes de información. Por otro lado, Cross y Manski (1999) derivaron límites finos para una regresión de la variable de resultado (Y) sobre dos variables de control discretas ($X; Z$) cuando las distribuciones condicionales $F(Y|X)$ y $F(Z|X)$ pueden ser identificadas a partir de bases de datos separadas. Asimismo, Fan *et al.* (2014a) adoptan el supuesto clásico de selección en observables de Rosenbaum y Rubin (1983) para identificar parcialmente el efecto de tratamiento basándose en la desigualdad de Cambanis *et al.* (1976).

El uso de distintas bases de datos puede ser aplicado a un problema bastante común en el ámbito de la evaluación de impacto conocido como “problema de inferencia ecológica”. Este problema es un caso particular donde se busca combinar una base de variables de resultado a nivel agregado con una base de variables de control a nivel de individuos con el objetivo de realizar inferencia a nivel individual. El problema de inferencia ecológica fue inicialmente desarrollado por King

(1997) y King *et al.* (2001), quienes tratan esta clase de problema cuando el objetivo principal es describir, predecir y realizar inferencia a nivel individual.

En este contexto, muchos se podrían inclinar por agregar la información individual, sin embargo, Corvalán *et al.* (2015) argumentan que la pérdida de información ocasionada por la agregación es lo suficientemente grande que impide la identificación de los efectos de tratamiento. Ante ello, Fan *et al.* (2014b) aplican los resultados de Fan *et al.* (2014a) y, adoptando el supuesto de selección en observables, proponen estimadores de los límites superiores e inferiores de los efectos de tratamiento². Como mencionan los autores, el problema de identificación en los contextos de información ecológica es análogo al problema de identificar promedios contra-factuales de las variables de resultado en un modelo de efectos de tratamiento. El método de identificación desarrollado por estos autores ha sido utilizado en Corvalán *et al.* (2015) para analizar el efecto del cambio de un régimen de voto obligatorio a uno de voto voluntario sobre la tasa de votantes efectivos en Chile.

La literatura relacionada a la combinación de bases de datos bajo el supuesto de selección en factores no observables es mucho más pequeña. Imbens y Newey (2003) utilizaron funciones de control para identificar y estimar modelos no separables bajo el supuesto de que la variable endógena y las perturbaciones del modelo son independientes condicionadas a la variable de control. De este modo, el presente estudio pretende contribuir presentando la identificación de los efectos de tratamiento a través de límites finos bajo el supuesto de selección en factores no observables cuando las variables de resultado y de control no son observadas en una misma base de datos.

El documento se organiza de la siguiente manera. El capítulo I es introductorio. El capítulo II introduce el marco teórico, así como algunos ejemplos que explican la utilidad de la combinación de bases de datos bajo el supuesto de selección en no observables. El capítulo III presenta los principales resultados de identificación en el caso de un modelo lineal en la heterogeneidad no observada, en el cual es posible realizar alguna transformación en las variables (diferencias por ejemplo) para volver al caso de selección en observables, donde los límites desarrollados por Fan *et al.* (2014a) son válidos. En el capítulo IV se aplica la identificación del efecto de tratamiento a situaciones en las que se cuenta con información ecológica. El capítulo V muestra un pequeño ejercicio de simulación. Finalmente, se presentan las conclusiones y recomendaciones.

² Además, demuestran que dichos estimadores son consistentes y asintóticamente normales.

Capítulo II. Marco teórico

El contexto en el cual trabajaré es una adaptación del modelo de resultados potenciales desarrollado por Rubin (1974). Denotaré $D_t \in (0,1)$ como la variable binaria que indica los dos posibles estados del tratamiento en cada momento en el tiempo. De esta manera, si definimos como T_0 el periodo en el cual se da el tratamiento, D_t será igual a cero para $t < T_0$ y tomará el valor de uno para los tratados a partir de $t \geq T_0$ ³. Por otro lado, denotaré Y_{dt} como la variable de resultado para cada uno de estos posibles estados $d = 0,1$ en el periodo t . Rubin considera a estos como resultados potenciales y que, en la práctica, solo es posible observar uno de ellos. El resultado observado se vincula con los resultados potenciales en cada momento del tiempo a través de la siguiente ecuación: $Y_t = D_t Y_{1t} + (1 - D_t) Y_{0t}$ para todo $t = 1, \dots, T$; donde T es el número de observaciones en el tiempo para cada individuo. Finalmente, denotaré como X_t a las variables de control (usualmente variables demográficas) en el momento t que potencialmente pueden afectar tanto D como $(Y_{1t}; Y_{0t})$.

Este marco de trabajo es muy utilizado en el campo de la evaluación de impacto, en el cual el objetivo principal es comparar ciertas características de la distribución de los resultados potenciales. Como se mencionó en la introducción, este objetivo suele ser alcanzado con el supuesto de selección en observables. No obstante, este trabajo busca relajar este supuesto y extender el análisis a un contexto en el cual no exista una asignación totalmente aleatoria hasta que no se tome en cuenta tanto la heterogeneidad observada como la no observada.

Para ello, denotaré la heterogeneidad no observada como η ; la cual asumiré que permanece constante con el tiempo. De manera general, se puede considerar a η en forma vectorial (tal como lo plantean Jun *et al.* 2014); de modo que algunos elementos puedan ser excluidos de la ecuación que determina D o de la que determina Y_{dt} para $d = 0,1$. Esta representación se formaliza de la siguiente manera:

$$Y_{dt} = g_d(X_t; \gamma; u_{dt}) \quad (1)$$

$$D_t = h(X_t; \delta; v_t) \quad (2)$$

Donde el vector η contiene a los elementos de γ (afecta únicamente a los resultados potenciales) y δ (afecta únicamente a la participación en el tratamiento). Esta representación es la de un panel estático no separable con heterogeneidad no observada particular a los resultados potenciales y a

³ Claramente, $D_t = 0$ para todos los no tratados para todo t .

la participación en el tratamiento. En el presente trabajo, adopto un modelo lineal y estático con efectos fijos comunes; es decir, el caso donde $\gamma = \delta = \eta$.

A diferencia de lo usualmente utilizado en la literatura, asumiré que las variables $(Y_t; D_t; X_t)$ no pueden ser observadas en una misma base de datos en el periodo t . En cambio, se cuenta con dos bases de datos en cada momento del tiempo: (i) aquella que contiene las variables de resultado $(Y_t; D_t)$ –que nombraré *base de resultados*– y (ii) aquella que contiene las variables de control o demográficas $(X_t; D_t)$ –que nombraré *base de control*. Asimismo, trabajo bajo el supuesto de que cada una de estas bases tiene una muestra panel, de modo que los individuos encuestados en la base de resultados y de control serán los mismos en todo momento en el tiempo; pero dichas bases no poseen individuos comunes⁴. Finalmente, asumiré que se cuenta con un panel de datos balanceado, es decir se tiene información de todas las variables relevantes para cada individuo en todos los momentos del tiempo. A continuación, presentaré algunos ejemplos en los cuales se muestre la utilidad de esta clase de estimación.

Ejemplo I: Retornos de largo plazo de la asistencia escolar (similar al retorno expuesto en Fan *et al.* 2014a). El problema de información surge en estos casos cuando la variable de resultado que se quiere utilizar es un indicador de largo plazo. Por ejemplo, el efecto de la asistencia escolar en las ganancias a lo largo de la vida de las personas. Claramente, existen factores no observables que afectan tanto la asignación del tratamiento – asistencia escolar– como la variable de resultado –los ingresos de las personas– (i.e. factores genéticos como la habilidad cognitiva y no cognitiva) lo cual ocasiona que el efecto calculado del tratamiento se encuentre sesgado. En esta situación, el uso de datos de panel relativamente largos en tiempo sería de mucha utilidad para poder controlar o “limpiar” el efecto de estos factores no observables mediante la estimación vía efectos fijos. De este modo, se puede utilizar encuestas administrativas, las cuales contienen información sobre los ingresos de las personas a lo largo del tiempo y combinarlas con otras encuestas, típicamente las encuestas de hogares⁵, para obtener efectos no sesgados del tratamiento.

Ejemplo II: Cambio en la distribución del ingreso a lo largo del tiempo (adaptado de Fan *et al.* 2014a a partir de DiNardo *et al.* 1996). DiNardo *et al.* (1996) se enfocaron en comparar el nivel de ingresos de las personas en dos años distintos, siendo la variable de tratamiento un indicador binario para cada año. En este trabajo, todas las variables se encontraban en una misma base de

⁴ Esto implica que no puede existir un individuo que se encuentre en la base de resultados y la base de control simultáneamente. Potencialmente, existe la posibilidad de que un individuo esté presente en ambas encuestas a pesar de que no se le pueda identificar. Por simplicidad, se considerará la probabilidad de que se dé esta situación como muy cercana a cero.

⁵ Muchas de ellas poseen una sección de datos de panel.

datos. Sin embargo, existen ciertas variables que podrían resultar de interés (en particular, aquellas relacionadas a factores de oferta laboral) que no se observan en la misma base de datos que los salarios. El uso de datos de panel permitiría controlar por estos factores y otros factores omitidos a la hora de calcular el efecto del tratamiento.

Ejemplo III: Efecto de fumar sobre el peso al nacer (adaptado de Jun *et al.* 2014). La decisión de una madre acerca de fumar se encuentra, en general, correlacionada con factores que afectan si lleva un estilo de vida saludable, los cuales también afectan de manera directa el peso al nacer del niño. Por tanto, no existirá una aleatorización completa a menos que se controle correctamente por estos factores que, en muchos de los casos, son prácticamente imposibles de medir u observar. Si bien los autores utilizan una única base de datos para la estimación, es plausible considerar que existen algunos factores de interés en otras bases de datos, por ejemplo, alguna medición concerniente a los hábitos de la mujer durante el embarazo en alguna encuesta de salud. Potencialmente, la variable de resultado –peso al nacer– y las variables de control pueden ser observadas en distintas fuentes de información si consideramos una encuesta que solo recoja información relacionada al niño y otra que recoja información concerniente únicamente a las características de la madre. Nuevamente, la disponibilidad de datos de panel sería de ayuda para poder lidiar con la heterogeneidad no observada presente en esta situación.

El uso de distintas fuentes de información también puede ser aplicado al problema de inferencia ecológica, es decir al caso donde se busca combinar una base a nivel agregado con otra a nivel individual. En este contexto, se entiende al tratamiento como algún evento agregado de modo que D_t denote, por ejemplo, dos zonas geográficas en el momento t . A continuación, se explica un ejemplo aplicado al ámbito electoral.

Ejemplo IV: Efecto del cambio de un régimen de voto obligatorio a voto voluntario sobre la tasa de votantes efectivos (ver Corvalán *et al.* 2015). En Chile, hasta el 2012, la participación electoral requería que las personas se registren en las listas electorales, de modo que si bien el registro era voluntario, una vez registrado, el voto era obligatorio. El Gobierno chileno emitió en enero del 2012 la Ley Electoral 20568, la cual establecía que, a partir de ese momento, el registro era automático y la votación voluntaria. Utilizando este cambio legislativo, Corvalán *et al.* (2015) comparan la proporción de personas que asistieron a votar en las elecciones municipales del 2012 y las del 2004, adoptando la metodología desarrollada por Fan *et al.* (2014b) bajo el supuesto de selección en observables. La variable D es un indicador binario de dos años distintos: $D = 0$ para el año base 2004 y $D = 1$ para el año 2012. Los autores utilizaron como base de resultados,

información agregada de las votaciones en cada año y, como base de control, una encuesta socio-económica a nivel individual⁶. En este contexto, la disponibilidad de datos de panel para ambas bases contribuye a poder controlar por la potencial heterogeneidad no observada existente que determina si el individuo decide presentarse a la votación (por ejemplo, la calidad de los candidatos presentados).

Un aspecto importante a considerar al tener datos de panel es el momento en el cual se realizan las encuestas (el *timing* de estas). Esto es importante ya que, de acuerdo a los supuestos realizados y al tipo de modelo con el que estemos trabajando, los requerimientos de información van a variar. Por ejemplo, en algunos contextos puede que no sea necesario tener datos de panel para la base de control, ya que la identificación es alcanzada únicamente con una línea de base, es decir con información en un periodo antes del tratamiento. Sin embargo, para modelos más complejos se requerirá más información o, en su defecto, supuestos más fuertes para lograr la identificación del efecto del tratamiento. En el presente documento trabajaré modelos en los que es posible eliminar la heterogeneidad no observada a través de alguna transformación en las variables (lo cual es posible, por ejemplo, en modelos lineales). Por simplicidad, asumiré la disponibilidad de una única encuesta antes del tratamiento, la cual será denotada con $t = 0 < T_0$ ⁷.

En el resto del documento adoptaré la notación utilizada por Fan *et al.* (2014a), quienes denotan $F_{A|B}(\cdot | b)$ como la función de distribución acumulada de la variable aleatoria A condicional a que $B = b$. Asimismo, denotaré como $F^{-1}(\cdot)$ a la función cuantil o inversa de la función de distribución $F(\cdot)$.

⁶ Las fuentes de información que utilizaron fueron el Instituto Nacional de Estadística (INE) y la Encuesta de Caracterización Socioeconómica (Casen), la más completa encuesta socio-económica a nivel individual de Chile.

⁷ La extensión a casos en los cuales existan más de un periodo antes de tratamiento es directa y, de todos modos, se explica de manera más detallada en el siguiente capítulo.

Capítulo III. Identificando efectos de tratamiento

En esta sección presentaré la identificación de las distribuciones contra-factuales y del efecto de tratamiento (ATE y ATT) en modelos en los que es posible eliminar la heterogeneidad no observada a través de alguna transformación en las variables.

1. Supuestos

En primer lugar, presentaré el supuesto de selección tanto en factores observables como no observables. Este supuesto establece que, en cada momento del tiempo, no se tiene una asignación aleatoria hasta que se controle tanto por la heterogeneidad observada como por la no observada:

Supuesto III.1 (S0): Consideremos que $(Y_{1t}; Y_{0t}; D, X_t; \eta)$ tienen una distribución conjunta para todos los momentos en el tiempo. Se cumple que, para todo $x_t \in \mathcal{X}_t$ y $n \in \mathcal{N}$, $(Y_{1t}; Y_{0t})$ es conjuntamente independiente de D dado $X_t = x_t$ y $\eta = n$.

Un ejemplo de modelos en los cuales es posible eliminar la heterogeneidad no observada son los modelos lineales. En esta sección, adoptaré un modelo lineal en la heterogeneidad no observada, de la forma:

$$Y_{dt} = g_d(X_t) + \gamma + u_{dt} \quad (3)$$

Una práctica muy común para controlar por la heterogeneidad no observada en esta clase de modelos es utilizar los operadores de primeras diferencias o de desviaciones temporales (como hace el estimador *Within group*). Claramente, al trabajar con alguna de estas transformaciones podemos eliminar la heterogeneidad no observada en la ecuación (3).

De manera general, denotemos como $Y_d = (Y_{d0}; Y_{d1}; \dots; Y_{dT})'$ para $d = 0, 1$ los dos vectores que contienen los resultados potenciales en cada momento en el tiempo. Podemos considerar una transformación similar a la descrita por Arellano (2003) de la forma KY_d ; donde K es una matriz de orden $(T - 1) \times T$ de rango $(T - 1)$ tal que $K\iota = 0$ y ι es un vector de $T \times 1$ lleno de unos. La ortogonalidad entre la matriz K y el vector ι garantiza la eliminación de la heterogeneidad no observada, de modo que KY_d ya no dependa de η . Es importante que cualquier transformación que se realice incluya el periodo anterior al tratamiento ($t = 0$) para poder identificar el efecto del programa. Tanto la matriz de primeras diferencias como la de desviaciones con respecto a la media temporal satisfacen estos requerimientos. De este modo, los resultados potenciales KY_d ya no dependen de la heterogeneidad no observada.

Por otro lado, trabajando con solo un periodo antes del tratamiento, modelo la condición de tratamiento D a través de la siguiente ecuación:

$$D = h(X_0) + v \quad (4)$$

Donde X_0 representa las características observables antes de la implementación del tratamiento. Es decir, se modela la condición de tratamiento en función únicamente de características iniciales^{8 9}. El principal motivo es evitar que las características de los individuos utilizadas como variables de control (la heterogeneidad observada) se vean afectadas por la aplicación del tratamiento. Esta clase de modelamiento de la participación en el programa es usual en la literatura al combinar el método de emparejamiento con el de diferencias (ver Abadie 2005, Lee 2005).

La ventaja de trabajar con estas transformaciones es que ahora el nuevo modelo cumple el supuesto clásico de selección en observables. Este supuesto consiste en dos condiciones. La primera se refiere al supuesto de independencia condicional, mientras que el segundo supuesto es uno relacionado al soporte del *propensity score*. Estos supuestos se muestran a continuación (adaptados de Rosenbaum y Rubin 1983; Firpo 2007 y Fan *et al.* 2014a):

Supuesto III.1* (S1): Consideremos que $(KY_1; KY_0; D, X)$ tienen una distribución conjunta. Se cumple que, para todo $x \in \mathcal{X}$, $(KY_1; KY_0)$ es conjuntamente independiente de D dado $X = x$.

Supuesto III.2 (S2): Para todo $x_0 \in \mathcal{X}_0$ se cumple $0 < p(x_0) < 1$, donde $p(x_0)$ es el *propensity score* y se define como $p(x_0) = \Pr(D = 1 | X_0 = x_0)$.

Donde $X = (X'_1, \dots, X'_T)'$ es la matriz que contiene a todas las variables de control para todos los momentos en el tiempo. El supuesto (S1) indica que, una vez transformados los resultados potenciales de modo que ya no dependan de los factores no observables, estas transformaciones

⁸ El lector podría argumentar que el *propensity score* estaría brindando predicciones potencialmente erróneas ya que la participación en el programa depende no solo de la heterogeneidad observada, sino también de la no observada (sino no habría endogeneidad inicialmente). Sin embargo, debemos notar que el objetivo principal es la predicción de las probabilidades de participación, por tanto basta que los factores no observables tengan correlación con las características iniciales para tener una predicción más que correcta ya que de todos modos se captura la variabilidad existente en la decisión de participación. Esta condición es plausible de cumplirse en la práctica. Por ejemplo, consideremos un programa en el cual existan altos costos de inscripción. Un factor no observable sería la motivación de la persona, lo cual es de esperar que esté bastante correlacionado con las características iniciales observables utilizadas para hacer el *propensity score* (i.e ingresos, nivel educativo, etcétera). Adicionalmente, se podría argumentar que estamos trabajando con la t -ésima diferencia de la participación en el programa, por lo que la condición de tratamiento D , ya no debería depender de la heterogeneidad no observada.

⁹ Se podría modelar la participación en el tratamiento como función de la t -ésima diferencia de las características observadas en caso se asuma que estas no han sido afectadas por el tratamiento. Los resultados posteriores no se ven mayormente afectados; lo único que va a cambiar es la manera en la cual se calcula el *propensity score*.

son independientes de la participación en el tratamiento D condicional a la heterogeneidad observada. Por otro lado, la condición (S2) implica que el *propensity score* basado en las características iniciales debe ser diferente de cero y de uno tanto para los individuos tratados como los de control. Es decir, no se puede predecir de manera determinística la participación en el tratamiento¹⁰. De este modo, bajo (S1) y (S2) se obtiene el supuesto clásico de selección en observables; con la única diferencia que está aplicado al modelo transformado¹¹.

A modo de ejemplo, consideremos el caso en el cual se cuenta con dos momentos en el tiempo (antes y después del tratamiento) para la base de resultados y la base de control, y que la matriz K representa el operador de primera diferencia. En este contexto, el modelo descrito en (3) y (4) se convierte en el clásico modelo de diferencias en diferencias (DD). Es bastante conocido que para la identificación del efecto de tratamiento en este modelo es necesario imponer el supuesto de tendencias comunes. Este supuesto implica que, en ausencia del tratamiento, las variables de resultado de los individuos tratados y de control hubiesen seguido una tendencia común o paralela condicional a las variables de control (ver Abadie 2005; Lechner 2013)¹². Es fácil notar que (S1) ya abarca este supuesto. En efecto, (S1) implica que, condicional a las variables de control, las primeras diferencias del modelo son independientes de la participación en el tratamiento, lo cual incluye la independencia en media requerida por el supuesto de tendencias comunes. Por tanto, al trabajar con el modelo en diferencias, la asignación del tratamiento es aleatoria condicional a las características observables.

2. Distribuciones marginales, contra-factuales y efectos de tratamiento

En el contexto descrito en la subsección anterior, se pueden aplicar los resultados obtenidos por Fan *et al.* (2014a) para identificar las distribuciones marginales y contra-factuales, así como el efecto de tratamiento. Como mencionan Fan *et al.*, cuando no se cuenta con todas las variables en una misma base de datos, la distribución $F_{Y_{dt}|X_t,D}(y|x,D)$ no se encuentra identificada a partir de la información disponible. Por tanto, las distribuciones contra-factuales $F_{Y_{1t}}(y)$, $F_{Y_{0t}}(y)$ y $F_{Y_{0t}|D}(y|1)$, así como parámetros que son funciones de estas distribuciones (como los efectos de tratamiento) tampoco están identificadas¹³.

¹⁰ El caso en el cual exista más de un periodo antes del tratamiento no es muy distinto. Se podría obtener el *propensity score* a partir de una regresión tipo *pool* en la cual se incorporen las características de todos los periodos anteriores al tratamiento; o, de manera alternativa, se podría trabajar con un promedio de los *propensity score* en cada periodo anterior al tratamiento.

¹¹ Una idea similar ha sido trabajada por Lee (2005: capítulo 4.5).

¹² Formalmente, el supuesto de tendencias comunes puede expresarse como $E[\Delta Y_d|X = x, D = 1] = E[\Delta Y_d|X = x, D = 0]$.

¹³ Note que la distribución del resultado potencial en presencia del tratamiento condicional a que el individuo sea tratado, $F_{Y_{1t}|D}(y|1)$, sí se encuentra identificada a partir de la muestra disponible.

Para lidiar con este problema, Fan *et al.* (2014a) utilizan la desigualdad de Cambanis *et al.* (1976) para obtener límites finos de las distribuciones contra-factuales y de los efectos de tratamiento (ATE y ATT). Fan *et al.* se basan en el supuesto de selección en observables, mientras que el presente trabajo permite que la asignación del tratamiento dependa, adicionalmente, de factores no observables constantes en el tiempo.

Sin embargo, si aplicamos la transformación ortogonal descrita en la capítulo III.1, los resultados potenciales, KY_d , ya no dependerán de la heterogeneidad no observada. Por tanto, podemos adoptar el mismo procedimiento desarrollado por Fan *et al.* (2014a) aplicándolo al modelo transformado que, como se vio en el capítulo anterior, cumple con el supuesto de selección en observables. De esta manera, las distribuciones que se buscan identificar no son las relacionadas a los resultados potenciales, sino a las de sus transformaciones: $F_{KY_1}(Ky)$, $F_{KY_0}(Ky)$ y $F_{KY_0|D}(Ky|1)$.

Siguiendo a Fan *et al.* (2014a) y a Firpo (2007) es posible expresar las distribuciones de los resultados potenciales transformados en función de la data, de modo que dependan de la inversa del *propensity score*¹⁴:

$$\begin{aligned} F_{KY_1}(Ky) &= E \left[\frac{D}{p(X_0)} I\{KY \leq Ky\} \right] \\ F_{KY_0}(Ky) &= E \left[\frac{1-D}{1-p(X_0)} I\{KY \leq Ky\} \right] \\ F_{KY_0|D}(Ky|1) &= \frac{1}{p_1} E \left[\frac{(1-D)p(X_0)}{1-p(X_0)} I\{KY \leq Ky\} \right] \end{aligned} \quad (5)$$

Donde $p_1 = \Pr[D = 1]$. Para entender mejor lo que capturan estas funciones de distribución, consideremos el caso en el cual solo se dispone de dos momentos en el tiempo (antes y después del tratamiento¹⁵) y, nuevamente, que la matriz K es el operador de primeras diferencias. En este caso, las funciones $F_{\Delta Y_1}(\Delta y)$, $F_{\Delta Y_0}(\Delta y)$ representan la probabilidad de que el crecimiento de la variable de resultado de un individuo tratado y no tratado sea igual a Δy , respectivamente.

¹⁴ Ver el Anexo 1 para la demostración.

¹⁵ Notemos que el caso en el que se analizan únicamente dos periodos no restringe el análisis únicamente a situaciones donde $T = 2$. Se podría pensar en un contexto en el cual hay diferentes horizontes de tiempo dependiendo de si se desea medir un efecto de corto, mediano o largo plazo; o en un contexto en el cual se cuentan con muchas variables de resultado y cada una de las cuales mida una dimensión distinta con un periodo de maduración específico.

Las expresiones descritas en la ecuación (5) no pueden ser identificadas a partir de la información disponible, por lo que, siguiendo a Fan *et al.* (2014a), es posible establecer límites finos para acotar estas distribuciones utilizando la desigualdad de Cambanis *et al.* (1976)¹⁶. Estos límites se presentan en el Teorema III.1¹⁷:

TEOREMA III.1: Para $d = 0,1$, tenemos que $F_{KY_d}^L(Ky) \leq F_{KY_d}(Ky) \leq F_{KY_d}^S(Ky)$, donde:

$$F_{KY_1}^L(Ky) = E \left[D \int_0^{F_{KY|D}(Ky|D)} F_{W|D}^{-1}(u|D) du \right]$$

$$F_{KY_1}^S(Ky) = E \left[D \int_{1-F_{KY|D}(Ky|D)}^1 F_{W|D}^{-1}(u|D) du \right]$$

$$F_{KY_0}^L(Ky) = E \left[(1-D) \int_0^{F_{KY|D}(Ky|D)} F_{V|D}^{-1}(u|D) du \right]$$

$$F_{KY_0}^S(Ky) = E \left[(1-D) \int_{1-F_{KY|D}(Ky|D)}^1 F_{V|D}^{-1}(u|D) du \right]$$

Asimismo, tenemos que $F_{KY_1|D}(Ky|1) = E[DI\{KY \leq Ky\}]/p_1$ ya está identificada, mientras que $F_{KY_0|D}(Ky|1)$ se encuentra identificado a través de: $F_{KY_0|D}^L(Ky|1) \leq F_{KY_0|D}(Ky|1) \leq F_{KY_0|D}^S(Ky|1)$, donde:

$$F_{KY_0|D}^L(Ky|1) = \frac{1}{p_1} E \left[(1-D) \int_0^{F_{KY|D}(Ky|D)} F_{V/W|D}^{-1}(u|D) du \right]$$

$$F_{KY_0|D}^S(Ky|1) = \frac{1}{p_1} E \left[(1-D) \int_{1-F_{KY|D}(Ky|D)}^1 F_{V/W|D}^{-1}(u|D) du \right]$$

Donde $W = 1/p(x_0)$ y $V = 1/(1-p(x_0))$, y se asume que las varianzas de W , V y de V/W son finitas. En este contexto, los límites presentados son finos.

Con respecto a los efectos del tratamiento, denotemos $\tau = KY_1 - KY_0$ como el efecto de tratamiento individual, de modo que el efecto promedio (ATE) y el efecto sobre los tratados (ATT) se expresen como $\mu_{ATE} = E[\tau]$ y $\mu_{ATT} = E[\tau|D = 1]$, respectivamente. Los límites (finos) de estos efectos se obtienen aplicando el Teorema 3.2 de Fan *et al.* (2014a)¹⁸ al modelo transformado en el caso en el cual $g(\tilde{Y}_d) = \tilde{Y}_d$:

¹⁶ La desigualdad de Cambanis-Simons-Stout se encuentra descrita en el Anexo 2.

¹⁷ En el Anexo 3 se explica de manera detallada la demostración.

¹⁸ Para facilitar la lectura, el Teorema 3.2 de Fan *et al.* (2014a) aplicado al modelo transformado se encuentra descrito en el Anexo 4 (considere el caso en el que $\tilde{Y} = KY$).

$$\mu_1^L - \mu_0^U \leq \mu_{ATE} \leq \mu_1^U - \mu_0^L$$

$$\frac{E[D(KY)]}{p_1} - \mu_{0|1}^U \leq \mu_{ATT} \leq \frac{E[D(KY)]}{p_1} - \mu_{0|1}^L$$

La ventaja de este resultado es que estos límites son identificables a partir de la información disponible. En efecto, $F_{KY|D}(Ky|D)$ puede ser identificado a partir de la base de resultado, mientras que las distribuciones $F_{W|D}(\cdot|D)$, $F_{V|D}(\cdot|D)$ y $F_{V/W|D}(\cdot|D)$ pueden ser obtenidas a partir de la base de control. Asimismo, estos límites son considerablemente más estrechos que los desarrollados por Manski (1990), como lo muestran Fan *et al.* (2014b).

La forma de identificación de los efectos de tratamiento permite sacar conclusiones interesantes acerca de los requerimientos de información. Bajo el supuesto de selección en factores no observables constantes en el tiempo, solo se necesita información de la base de resultados y la base de control en dos momentos en el tiempo para poder obtener límites finos del efecto de tratamiento, aplicando el modelo de diferencias en diferencias. Inclusive, si se adopta el supuesto de que la participación en el tratamiento depende únicamente de características iniciales y que se puede identificar los agentes tratados y controles antes del tratamiento¹⁹, solo se necesitaría tener datos de panel para la base de resultados, mientras que bastaría con una línea de base para la base de control para poder identificar el efecto de tratamiento. En caso se cuente con una mayor dimensión temporal, otras transformaciones más eficientes se encuentran disponibles como la de desvíos con respecto a la media temporal (estimador *Within group*).

¹⁹ Por ejemplo, en el caso en el que el tratamiento denote dos periodos de tiempo o mercados geográficos.

Capítulo IV. Identificando efectos de tratamiento con información ecológica

En este capítulo se aplican los resultados obtenidos en el capítulo III para el caso en el cual se cuenta con información ecológica, es decir en un contexto en el cual se busca combinar una base agregada –la base de resultados– con una base a nivel individual –la base de control– para realizar inferencia a nivel individual. Para ello, utilizo adaptaciones simples de los límites obtenidos por Fan *et al.* (2014b) y Corvalán *et al.* (2015) para el efecto del tratamiento. Para el caso de modelos de información ecológica, Fan *et al.* (2014b) afirman que, en estos contextos, el problema de identificación es análogo al problema de identificar promedios contra-factuales de las variables de resultado en un modelo de efectos de tratamiento, por lo que basan la identificación de los efectos causales en los límites de estos promedios contra-factuales.

Al igual que Corvalán *et al.* (2015) y Fan *et al.* (2014b), analizaré el caso en el que tanto los resultados potenciales como los observados poseen un comportamiento binario. Adoptando el supuesto de selección en observables, los efectos de tratamiento promedio y sobre los tratados (ATE y ATT; respectivamente) pueden expresarse de la siguiente forma:

$$\begin{aligned}\mu_{ATE} &= E[Y_1 - Y_0] = \Pr[Y = 1] - \Pr[Y = 0] \\ \mu_{ATT} &= E[Y_1 - Y_0 | D = 1] = \Pr[Y = 1 | D = 1] - \Pr[Y = 0 | D = 1]\end{aligned}$$

Como se puede apreciar, únicamente es necesario obtener límites finos de estos promedios (probabilidades) contra-factuales para poder identificar los efectos de tratamiento. Corvalán *et al.* (2015: Teorema I) y Fan *et al.* (2014b: sección 4) obtienen estos límites como un caso particular del Teorema 3.2 de Fan *et al.* (2014a); así como estimadores de dichos límites.

Al igual que en la sección anterior, podemos aplicar los resultados obtenidos por Corvalán *et al.* (2015) y Fan *et al.* (2014b) al modelo transformado KY_d , el cual depende únicamente de la heterogeneidad observada. Analizaré el caso particular, aunque bastante común, en el cual se cuenta únicamente con dos periodos de tiempo; por lo que la matriz K se reduce al operador de primeras diferencias²⁰. De esta manera, los efectos de tratamiento promedio y sobre los tratados (ATE y ATT, respectivamente) son:

$$\mu_{ATE} = E[\Delta Y_1 - \Delta Y_0] = E[\Delta Y_1] - E[\Delta Y_0]$$

²⁰ Modelos en los que se cuenta con más de un periodo de tiempo y/o se utiliza una transformación distinta son bastante directos.

$$\mu_{ATT} = E[\Delta Y_1 - \Delta Y_0 | D = 1] = E[\Delta Y_1 | D = 1] - E[\Delta Y_0 | D = 1]$$

A diferencia del caso analizado por Corvalán *et al.* (2015) y Fan *et al.* (2014b), los efectos de tratamiento ATE y ATT no se reducen únicamente a la diferencia de las probabilidades, ya que la primera diferencia deja de ser una variable binaria. En efecto, puede tomar los valores de 1, -1 o 0 dependiendo de si ha existido una evolución positiva, negativa o nula en los resultados potenciales, respectivamente.

Para solucionar este problema, planteo dos alternativas. La primera consiste en regresar al caso binario y analizar únicamente el caso asociado al crecimiento de los resultados potenciales, de modo que se defina una nueva variable que tome el valor de uno si el crecimiento ha sido positivo (o negativo) y cero de otro modo. De esta forma, solo se necesita calcular únicamente los límites de las probabilidades contra-factuales a la Corvalán *et al.* (2015) y Fan *et al.* (2014b). La segunda alternativa, un poco más compleja, es extender los resultados de los autores al caso en el cual la variable de resultado, aun siendo discreta, puede tomar tres valores diferentes. Estas dos opciones son analizadas a continuación.

1. Dicotomizando el crecimiento de los resultados potenciales

Sin pérdida de generalidad, tomemos el caso en el cual el interés reside en analizar si la variable de resultado ha evolucionado de manera positiva producto del tratamiento. Para ello, conviene definir una variable auxiliar, \tilde{Y}_d , como:

$$\tilde{Y}_d = \begin{cases} 1, & \text{si } \Delta Y_d = 1 \\ 0, & \text{si } \Delta Y_d = 0 \text{ o } \Delta Y_1 = -1 \end{cases}$$

La variable auxiliar observada, \tilde{Y} , se relaciona con la potencial mediante la siguiente ecuación: $\tilde{Y} = D\tilde{Y}_1 + (1 - D)\tilde{Y}_0$. Los efectos de tratamiento serán:

$$\begin{aligned} \mu_{ATE} &= E[\tilde{Y}_1 - \tilde{Y}_0] = \Pr[\tilde{Y} = 1] - \Pr[\tilde{Y} = 0] \\ \mu_{ATT} &= E[\tilde{Y}_1 - \tilde{Y}_0 | D = 1] = \Pr[\tilde{Y} = 1 | D = 1] - \Pr[\tilde{Y} = 0 | D = 1] \end{aligned}$$

De esta manera, es posible identificar los efectos de tratamiento a través de límites finos para los promedios contra-factuales de la variable auxiliar del mismo modo que lo hacen Corvalán *et al.* (2015) y Fan *et al.* (2014b). Para ello, denotemos:

$$p_{00} = \Pr[\tilde{Y} = 0|D = 0] = \Pr[\Delta Y_d = 0; \Delta Y_1 = -1|D = 0]$$

$$p_{01} = \Pr[\tilde{Y} = 0|D = 1] = \Pr[\Delta Y_d = 0; \Delta Y_1 = -1|D = 1]$$

Con esto en consideración, podemos obtener los mismos límites desarrollados por Fan *et al.* (2014b) aplicando el Teorema 3.2 de Fan *et al.* (2014a).

Teorema IV.1: Asumiendo que $Var(W) < \infty$ y que $Var(V) < \infty$, se cumple que:

$$L_1^I - L_0^S \leq \mu_{ATE} \leq L_1^S - L_0^I$$

$$\frac{p_{11}}{p_1} - L_{0|1}^S \leq \mu_{ATT} \leq \frac{p_{11}}{p_1} - L_{0|1}^I$$

donde:

$$L_1^I = p_1 \int_0^{1-p_{01}} F_{W|D}^{-1}(u|1) du$$

$$L_1^S = p_1 \int_{p_{01}}^1 F_{W|D}^{-1}(u|1) du$$

$$L_0^I = (1-p_1) \int_0^{1-p_{00}} F_{V|D}^{-1}(u|0) du$$

$$L_0^S = (1-p_1) \int_{p_{00}}^1 F_{V|D}^{-1}(u|0) du$$

$$L_{0|1}^I = \frac{(1-p_1)}{p_1} \int_0^{1-p_{00}} F_{V/W|D}^{-1}(u|0) du$$

$$L_{0|1}^S = \frac{(1-p_1)}{p_1} \int_{p_{00}}^1 F_{V/W|D}^{-1}(u|0) du$$

Para el cálculo de estos límites es posible utilizar estimadores análogos a los planteados por Fan *et al.* (2014b) y Corvalán *et al.* (2015). Asumamos que la base de resultados cuenta con M regiones, mientras que la base de control cuenta con N individuos en ambos momentos en el tiempo. A partir de la base de resultado, podemos obtener estimadores de las proporciones muestrales \hat{p}_1 , \hat{p}_{11} , \hat{p}_{00} y \hat{p}_{01} tal y como proponen Corvalán *et al.* (2015):

$$\hat{p}_1 = M^{-1} \sum_{i=1}^M I\{D_i = 1\}$$

$$\hat{p}_{11} = (M\hat{p}_1)^{-1} \sum_{i=1}^M I\{\tilde{Y}_i = 1, D_i = 1\}$$

$$\hat{p}_{00} = (M\hat{p}_1)^{-1} \sum_{i=1}^M I\{\tilde{Y}_i = 0, D_i = 0\}$$

$$\hat{p}_{01} = (M\hat{p}_1)^{-1} \sum_{i=1}^M I\{\tilde{Y}_i = 0, D_i = 1\}$$

Por otro lado, los estimadores para las funciones cuantil $F_{W|D}^{-1}(u|0)$, $F_{V|D}^{-1}(u|0)$ y $F_{V/W|D}^{-1}(u|0)$ son los mismos que los desarrollados por Fan *et al.* (2014b) y Corvalán *et al.* (2015), con la diferencia de que se utilizará el *propensity score* basado únicamente en características previas al tratamiento²¹. Finalmente, se obtienen estimadores de los límites del Teorema IV.1 integrando de manera numérica en el intervalo establecido.

2. Trabajando directamente con el crecimiento de los resultados potenciales

La segunda alternativa consiste en trabajar directamente con la primera diferencia de los resultados potenciales. Por tanto, será de utilidad extender el resultado propuesto por Fan *et al.* (2014b) para el caso en el cual la variable de resultado, aun siendo discreta, puede tomar más de dos valores diferentes.

Consideremos un caso general, en el cual los resultados potenciales (denotados como \tilde{Y}_d) pueden tomar G valores distintos ($\tilde{Y}_1, \dots, \tilde{Y}_G$); ordenados de la siguiente manera: $\tilde{Y}_1 < \tilde{Y}_2 < \dots < \tilde{Y}_G$. Para obtener límites finos de los efectos de tratamiento, podemos utilizar el Teorema 3.2 de Fan *et al.* (2014a)²² en el caso particular donde $g(\tilde{Y}_d) = \tilde{Y}_d$. Para ello, notemos que:

$$F_{\tilde{Y}|D}^{-1}(u|d) = \begin{cases} \tilde{Y}_1 & u \in [0; p_{1d}[\\ \tilde{Y}_2 & u \in [p_{1d}; p_{2d}[\\ \vdots & \vdots \\ \tilde{Y}_G & u \in [p_{G-1d}; 1] \end{cases}$$

$$F_{\tilde{Y}|D}^{-1}(1-u|d) = \begin{cases} \tilde{Y}_1 & u \in [1-p_{1d}; 1[\\ \tilde{Y}_2 & u \in [1-p_{2d}; 1-p_{1d}[\\ \vdots & \vdots \\ \tilde{Y}_G & u \in [0; 1-p_{G-1d}] \end{cases}$$

Donde, para $d = 0, 1$ y $j = 1, 2, \dots, G-1$ denotamos como $p_{jd} = \Pr[Y = Y_j | D = d]$. De este modo, se obtienen los límites de los promedios contrafactuales en el Teorema IV.2.

²¹ El detalle de los estimadores de las funciones cuantil propuestos por Fan *et al.* (2014b) al caso presentado se encuentra en el Anexo 3.

²² Para facilitar la lectura, el Teorema 3.2 de Fan *et al.* (2014a) se encuentra listado en el Anexo 4.

Teorema IV.2: Asumiendo que $Var(W) < \infty$ y que $Var(V) < \infty$, se cumple que:

$$C_1^I - C_0^S \leq \mu_{ATE} \leq C_1^S - C_0^I$$

$$\frac{C_1}{p_1} - C_{0|1}^S \leq \mu_{ATT} \leq \frac{C_1}{p_1} - C_{0|1}^I$$

donde:

$$C_1 = E[D\tilde{Y}] = \sum_{i=1}^G \tilde{Y}_i p_{i1}^*$$

$$C_1^I = p_1 \left[\int_{1-p_{11}}^1 \tilde{Y}_1 F_{W|D}^{-1}(u|1) du + \int_{1-p_{21}}^{1-p_{11}} \tilde{Y}_2 F_{W|D}^{-1}(u|1) du + \dots + \int_0^{1-p_{G-11}} \tilde{Y}_G F_{W|D}^{-1}(u|1) du \right]$$

$$C_1^S = p_1 \left[\int_0^{p_{11}} \tilde{Y}_1 F_{W|D}^{-1}(u|1) du + \int_{p_{11}}^{p_{21}} \tilde{Y}_2 F_{W|D}^{-1}(u|1) du + \dots + \int_{p_{G-11}}^1 \tilde{Y}_G F_{W|D}^{-1}(u|1) du \right]$$

$$C_0^I = (1-p_1) \left[\int_{1-p_{10}}^1 \tilde{Y}_1 F_{V|D}^{-1}(u|0) du + \int_{1-p_{20}}^{1-p_{10}} \tilde{Y}_2 F_{V|D}^{-1}(u|0) du + \dots + \int_0^{1-p_{G-10}} \tilde{Y}_G F_{V|D}^{-1}(u|0) du \right]$$

$$C_0^S = (1-p_1) \left[\int_0^{p_{10}} \tilde{Y}_1 F_{V|D}^{-1}(u|0) du + \int_{p_{10}}^{p_{20}} \tilde{Y}_2 F_{V|D}^{-1}(u|0) du + \dots + \int_{p_{G-10}}^1 \tilde{Y}_G F_{V|D}^{-1}(u|0) du \right]$$

$$C_{0|1}^I = \frac{(1-p_1)}{p_1} \left[\int_{1-p_{10}}^1 \tilde{Y}_1 F_{V/W|D}^{-1}(u|0) du + \dots + \int_0^{1-p_{G-10}} \tilde{Y}_G F_{V/W|D}^{-1}(u|0) du \right]$$

$$C_{0|1}^S = \frac{(1-p_1)}{p_1} \left[\int_0^{p_{10}} \tilde{Y}_1 F_{V/W|D}^{-1}(u|0) du + \dots + \int_{p_{G-10}}^1 \tilde{Y}_G F_{V/W|D}^{-1}(u|0) du \right]$$

Donde, para $d = 0, 1$ y $j = 1, 2, \dots, G-1$ denotamos como $p_{j1}^* = \Pr[Y = Y_j; D = 1]$.

Es posible aplicar los resultados del Teorema IV.2 al modelo de primeras diferencias, en el cual la variable puede tomar únicamente tres valores: -1, 0 o 1. Para ello, tomemos $\tilde{Y}_d = \Delta Y_d$ y definamos la función cuantil de la primera diferencia como:

$$F_{\Delta Y|D}^{-1}(u|d) = \begin{cases} -1, & u \in [0; p_{\Delta-1d}[\\ 0, & u \in [p_{\Delta-1d}; p_{\Delta 0d}[\\ 1, & u \in [p_{\Delta 0d}; 1] \end{cases}$$

Con esto en consideración, los límites de los efectos de tratamiento se definen en el Teorema IV.3.

Teorema IV.3: Asumiendo que $Var(W) < \infty$ y que $Var(V) < \infty$, se cumple que:

$$C_1^I - C_0^S \leq \mu_{ATE} \leq C_1^S - C_0^I$$

$$\frac{C_1}{p_1} - C_{0|1}^S \leq \mu_{ATT} \leq \frac{C_1}{p_1} - C_{0|1}^I$$

donde:

$$C_1 = E[D\Delta Y] = p_{\Delta 11}^* - p_{\Delta -11}^*$$

$$C_1^I = p_1 \left[\int_0^{1-p_{\Delta 01}} F_{W|D}^{-1}(u|1) du - \int_{1-p_{\Delta -11}}^1 F_{W|D}^{-1}(u|1) du \right]$$

$$C_1^S = p_1 \left[\int_{p_{\Delta 01}}^1 F_{W|D}^{-1}(u|1) du - \int_0^{p_{\Delta -11}} F_{W|D}^{-1}(u|1) du \right]$$

$$C_0^I = (1-p_1) \left[\int_0^{1-p_{\Delta 00}} F_{V|D}^{-1}(u|0) du - \int_{1-p_{\Delta -10}}^1 F_{V|D}^{-1}(u|0) du \right]$$

$$C_0^S = (1-p_1) \left[\int_{p_{\Delta 00}}^1 F_{V|D}^{-1}(u|0) du - \int_0^{p_{\Delta -10}} F_{V|D}^{-1}(u|0) du \right]$$

$$C_{0|1}^I = \frac{(1-p_1)}{p_1} \left[\int_0^{1-p_{\Delta 00}} F_{V/W|D}^{-1}(u|0) du - \int_{1-p_{\Delta -10}}^1 F_{V/W|D}^{-1}(u|0) du \right]$$

$$C_{0|1}^S = \frac{(1-p_1)}{p_1} \left[\int_{p_{\Delta 00}}^1 F_{V/W|D}^{-1}(u|0) du - \int_0^{p_{\Delta -10}} F_{V/W|D}^{-1}(u|0) du \right]$$

Donde, para $d = 0,1$, se define $p_{\Delta 11}^* = \Pr[\Delta Y = 1; D = 1]$, $p_{\Delta -11}^* = \Pr[\Delta Y = -1; D = 1]$, $p_{\Delta 0d} = \Pr[\Delta Y = 0|D = d]$ y $p_{\Delta -1d} = \Pr[\Delta Y = -1|D = d]$.

Para obtener estimadores de estos límites, adoptaremos los sugeridos por Corvalán *et al.* (2015) y Fan *et al.* (2014b), con la única diferencia de que ahora son aplicados a la primera diferencia del modelo. Nuevamente, asumamos que la base de resultados cuenta con M regiones, mientras que la base de control cuenta con N individuos tanto antes como después del tratamiento. A partir de la base de resultado, podemos utilizar los siguientes estimadores de las proporciones muestrales \hat{p}_1 , $\hat{p}_{\Delta 11}^*$, $\hat{p}_{\Delta -11}^*$, $\hat{p}_{\Delta -10}$, $\hat{p}_{\Delta -11}$, $\hat{p}_{\Delta 00}$ y $\hat{p}_{\Delta 01}$:

$$\hat{p}_1 = \frac{1}{M} \sum_{i=1}^M I\{D_i = 1\}$$

$$\hat{p}_{\Delta 11}^* = \frac{1}{M} \sum_{i=1}^M I\{\Delta Y_i = 1, D_i = 1\}$$

$$\begin{aligned}\hat{p}_{\Delta-11}^* &= \frac{1}{M} \sum_{i=1}^M I\{\Delta Y_i = -1, D_i = 1\} \\ \hat{p}_{\Delta 01} &= \frac{1}{M\hat{p}_1} \sum_{i=1}^M I\{\Delta Y_i = 0, D_i = 1\} \\ \hat{p}_{\Delta-11} &= \frac{1}{M\hat{p}_1} \sum_{i=1}^M I\{\Delta Y_i = -1, D_i = 1\} \\ \hat{p}_{\Delta 00} &= \frac{1}{M(1-\hat{p}_1)} \sum_{i=1}^M I\{\Delta Y_i = 0, D_i = 0\} \\ \hat{p}_{\Delta-10} &= \frac{1}{M(1-\hat{p}_1)} \sum_{i=1}^M I\{\Delta Y_i = -1, D_i = 0\}\end{aligned}$$

Los estimadores de las funciones cuantil son los mismos que los planteados en la alternativa uno y se encuentran detallados en el Anexo 3, basándose en el *propensity score* obtenido a partir de las características previas al tratamiento. Finalmente, se integra de manera numérica para obtener los límites planteados en el Teorema IV.3.

Capítulo V. Simulación

Para contrastar el rendimiento de los límites, realizo un ejercicio de simulación bastante simple. Consideremos el siguiente modelo de dos periodos ($t = 0,1$):

$$\begin{aligned} Y_{1t}^* &= \alpha_1 X_t + \gamma + u_{1t}; & Y_{0t}^* &= \alpha_0 X_t + \gamma + u_{0t} \\ Y_{1t} &= I\{\alpha_1 X_t + \gamma + u_{1t} \geq 0\}; & Y_{0t} &= I\{\alpha_0 X_t + \gamma + u_{0t} \geq 0\} \\ D &= I\{\delta X_0 - \varepsilon \geq 0\} \end{aligned}$$

Donde $(X_1; X_0; u_{11}; u_{10}; u_{01}; u_{00}; \gamma; \varepsilon) \sim N(0; I_8)$. Este modelo considera la participación en el tratamiento únicamente en función de las características iniciales. Asimismo, como en el caso de inferencia ecológica, se analizará el caso en el cual los resultados potenciales (y, por tanto, los observados) dependen de manera lineal de la heterogeneidad no observada constante en el tiempo.

El objetivo es comparar los límites de los efectos de tratamiento ATE y ATT planteados en el capítulo IV obtenidos bajo el supuesto de selección en factores no observables con los límites desarrollados por Fan *et al.* (2014a, 2014b); quienes se basan en la selección en observables. Asimismo, se comparan dichos resultados con el simple estimador de diferencias definido como:

$$\frac{\sum_{i=1}^M Y_i D_i}{\sum_{i=1}^M D_i} - \frac{\sum_{i=1}^M Y_i (1 - D_i)}{\sum_{i=1}^M (1 - D_i)}$$

Recordemos que este estimador es consistente bajo el supuesto de que la muestra es totalmente aleatoria, es decir que los resultados potenciales son independientes a la asignación en el tratamiento. Inclusive, bajo este supuesto, los efectos ATE y ATT son iguales a dicho estimador.

En términos de la simulación, se debería esperar que tanto el estimador de diferencias como los límites de Fan *et al.* (2014a, 2014b) se encuentren sesgados ante la presencia de factores no observables, a la vez que dicho sesgo se corregiría al aplicar los mismos límites pero aplicados al modelo transformado. Para este ejercicio, se eliminarán la heterogeneidad no observada aplicando el método de diferencias, por lo que se aplicarán los límites desarrollados en el Teorema IV.3²³.

²³ Para integrar numéricamente se utilizó la regla de Simpson al tratarse de funciones discretas.

Los resultados de una simulación de Monte Carlo se muestran en la Tabla 1. Los verdaderos valores de los efectos de tratamiento ATE y ATT fueron computados directamente a partir de los datos simulados.

Tabla 1: Rendimiento de los estimadores

| | Verdadero Valor | Asignación Aleatoria | Observables | | Inobservables | |
|------------|-----------------|----------------------|---------------------|--------------------|---------------------|--------------------|
| | | | Inferior | Superior | Inferior | Superior |
| ATE | -0,018 | 0,0064 (0,0216) | -0,0172 (0,0200) | 0,0383 (0,0195) | -0,0524 (0,0139) | 0,0487 (0,0154) |
| ATT | 0,023 | 0,0064 (0,0216) | -0,0242 (0,0201) | 0,1311 (0,0240) | -0,0834 (0,0315) | 0,0901 (0,0296) |

Nota: Los valores reales fueron obtenidos a través de simulaciones. El *propensity score* fue obtenido a través de un modelo *probit*. Se reportan los promedios de 100 repeticiones. El tamaño de muestra fue de 2.000 observaciones (1.000 por periodo). Entre paréntesis se encuentra la desviación estándar (de las repeticiones).

Valores: $\delta = 0,5$, $\alpha_1 = 3$, $\alpha_0 = -3$

Fuente: Elaboración propia 2015.

Ante la presencia de factores no observables constantes en el tiempo, tanto el estimador de diferencias como los límites desarrollados por Fan *et al.* (2014a, 2014b) resultan estar sesgados. Analizando el estimador de diferencias se concluye que el efecto del tratamiento es muy cercano a cero, subestimando tanto el ATE como el ATT. Por otro lado, los límites desarrollados por Fan *et al.* (2014a) bajo selección en observables sobrestiman el efecto del tratamiento si se incluye heterogeneidad no observada constante en el tiempo. Si bien en el caso del ATT, el verdadero valor se encuentra incluido dentro del intervalo (a pesar de que esté muy cercano a la cota inferior); podemos ver que en el caso del ATE el verdadero valor ya no pudo ser captado por dichos límites. Este problema desaparece si es que se utilizan los límites desarrollados en el Teorema IV.3. Notemos que, tanto para el ATT como para el ATE, el valor medio del intervalo es muy cercano al verdadero valor.

Conclusiones y recomendaciones

Conclusiones

La necesidad de combinar diferentes fuentes de información para modelar efectos causales es muy común en las ciencias sociales. Algunos potenciales usos podrían ser analizar retornos educativos a largo plazo, modelar el comportamiento electoral de algún país (como Corvalán *et al.* 2015), comparar el efecto de ciertas políticas regionales, los efectos de guerras tanto internas como externas sobre indicadores de salud, entre otros. Adoptando el supuesto de selección en observables, es posible identificar efectos de tratamiento utilizando los resultados de Fan *et al.* (2014a, 2014b). Sin embargo, muchas veces este supuesto no es el más apropiado. Existen muchos casos en los cuales la selección en observables es violada, lo cual ocasiona que las variables de tratamiento sean endógenas debido a la presencia de heterogeneidad no observada y sesgo de selección. En estos casos, se necesita una forma alternativa de identificación.

En el presente documento considero la identificación de las distribuciones contra-factuales y los efectos de tratamiento cuando las variables de resultado y las de control se encuentran en diferentes bases de datos bajo el supuesto de selección en no observables. Para ello, me baso en los resultados desarrollados por Fan *et al.* (2014a, 2014b), quienes obtienen límites finos para las distribuciones contra-factuales y los parámetros de interés (ATE y ATT) asumiendo selección en observables. Trabajando con un modelo de heterogeneidad no observada constante en el tiempo que afecte de forma lineal a los resultados potenciales, muestro que es posible realizar un procedimiento similar al propuesto por Fan *et al.* (2014a, 2014b) utilizando una transformación ortogonal (por ejemplo: primeras diferencias) en el modelo, de modo que podamos “regresar” al supuesto de selección en observables. Como caso particular, considero el problema de inferencia ecológica, en el cual información a nivel agregado e individual deben ser combinadas para realizar inferencia a nivel individual.

Recomendaciones

El siguiente paso consiste en buscar maneras alternativas de identificar los efectos de tratamiento en modelos basados en otro tipo de heterogeneidad no observada; por ejemplo, modelos dinámicos o que incluyan factores no observables constantes entre individuos pero variantes en el tiempo. Asimismo, se podría considerar flexibilizar la forma funcional de la heterogeneidad no observada. La fuente de identificación de la presente investigación es la relación lineal entre el efecto no observado y los resultados potenciales. No obstante, este supuesto de linealidad puede ser muy restrictivo en ciertas situaciones, por lo que el desarrollo de nuevos métodos de

identificación robustos a diferentes formas funcionales de la heterogeneidad no observada sería de mucha utilidad.

Una posibilidad es utilizar límites alternativos disponibles en la literatura como los desarrollados por Manski (1990) o los límites de Frèchet-Hoeffding. No obstante, los intervalos obtenidos a partir de dichos límites suelen ser poco informativos, por lo que sería de mucha utilidad obtener formas alternativas de identificación. De todos modos, esto resulta bastante desafiante, por lo que se plantea como investigación futura.

Bibliografía

Abadie, A. (2005). “Semiparametric Difference-in-Difference Estimators”. *Review of Economic Studies* 72, 1-19.

Angrist, J. D. y Krueger A. B. (1995). “Split-Sample Instrumental Variable Estimates of the Return to Schooling”. *Journal of Business & Economic Statistics*, 13(2), 225-235.

Arellano, M. (2003). *Panel Data Econometrics*. 2^{da} ed. España: Centro de Estudios Monetarios y Financieros (CEMFI).

Cambanis, S.; Simons, G.; y Stout, W. (1976). “Inequalities for $E_k(X, Y)$ When the Marginals Are Fixed”. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 36, 285-294. [812, 816].

Chernozhukov, V.; Fernández-Val, I.; y Melly, B. (2013). “Inference on Counterfactual Distributions”. *Econometrica*, 81 (6), 2205-2268.

Corvalán, A.; Melo, E.; Sherman, R.; y Shum, M. (2015). *Bounding Causal Effects in Ecological Inference Problems*. Working Paper. California Institute of Technology.

Cross, P. J., y Manski, C. F. (1999). “Regressions, Short and Long”. *Econometrica*, 70 (1), 357-368.

DiNardo, J.; Fortin, N.; y Lemieux, T. (1996). “Labor Market Institutions and the Distribution of Wages 1973-1992: A Semiparametric Approach”. *Econometrica*, 64, 1001-1044.

Fan, Y.; Sherman, R.; y Shum, M. (2014a). “Identifying Treatment Effects Under Data Combination”. *Econometrica*, 82 (2), 811-822.

Fan, Y.; Sherman, R.; y Shum, M. (2014b). *Estimation and Inference in an ecological inference model*. Working Paper. California Institute of Technology.

Fan, Y.; y Park, S. (2012). “Confidence Intervals for the Quantile of Treatment Effects in Randomized Experiments”. *Journal of Econometrics*, 167, 330-344.

Fan, Y.; y Park, S. (2010). “Sharp Bounds on the Distribution of Treatment Effects and Their Statistical Inference”. *Econometric Theory*, 26, 931-951.

Fan, Y.; y Park, S. (2009). *Partial Identification of the Distribution of Treatment Effects and Confidence Sets. Advances in Econometrics: Nonparametric Econometric Methods*. Bingley, U.K.: Emerald Group.

Fan, Y.; y Zhu, D. (2009). *Partial Identification and Confidence Sets for Functionals of the Joint Distribution of Potential Outcomes*. Working Paper. Department of Economics, Vanderbilt University.

Firpo, S. (2007). “Efficient Semiparametric Estimation of Quantile Treatment Effects”. *Econometrica* 75, 259-276.

Frank, M. J.; Nelsen, R. B.; y Schweizer, B. (1987), “Best-Possible Bounds on the Distribution of a Sum—a Problem of Kolmogorov”. *Probability Theory and Related Fields* 74, 199-211.

Heckman, J.; Smith, J.; y Clements, N. (1997). “Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts”. *Review of Economic Studies*, 64, 487–535.

Hirano, K.; Imbens, G. W.; y Ridder, G. (2003): “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score”. *Econometrica*, 71, 1161-1189.

Hislop, D. R. (1999). “State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Woman”. *Econometrica*, 67 (6), 1255-1294.

Imbens, G. W.; y Wooldridge, J. (2009). “Recent Developments in the Econometrics of Program Evaluation”. *Journal of Economic Literature*, 47 (1), 5-86.

Imbens, G.; y Newey, W. (2003). *Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity*. Manuscrito. UC Berkeley: Departamento de Economía.

Jun S. J.; Lee, Y.; y Shin, Y. (2014). *Treatment Effects with Unobserved Heterogeneity: A Set Identification Approach*. Pennsylvania: Pennsylvania State University.

- Kahn, L. M. (2007). “The Impact of Employment Protection Mandates on Demographic Temporary Employment Patterns: International Microeconomic Evidence”. *The Economic Journal*, 117 (521): F333-F56.
- King, G.; Rosen, O.; y Tanner, M. (2001). *Ecological Inference: New Methodological Strategies*. Cambridge: Cambridge University Press.
- King, G. (1997). *A Solution to the Ecological Inference Problem*. Princeton: Princeton University Press.
- Klevmarcken, N. A. (1982). *Missing Variables and Two-Stage Least-squares Estimation From More Than One Dataset*. En 1981 Proceedings of the American Statistical Association, Business and Economic Statistics Section, 156–161.
- Lechner, M. (2013). *Treatment Effects and Panel Data*. Discussion Paper. Department of Economics, Universität St.Gallen.
- Lee, M. J. (2005). *Micro-Econometrics for Policy, Program, and Treatment Effects*. Oxford: Oxford University Press.
- Lillard, L. A.; y Willis, R. J. (1978). “Dynamic Aspects of Earning Mobility”. *Econometrica*, 46 (5), 985-1012.
- Manski, C. F. (1990). “Non-parametric Bounds on Treatment Effects”. *American Economic Review*, Papers and Proceedings 80, 319-323.
- Ridder, G.; y Moffitt R. (2007). “Econometrics of Data Combination”. *The Handbook of Econometrics*, Vol 6B, Chapter 75. Nueva York: North-Holland.
- Rosenbaum, P.; y Rubin, D. (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects”. *Biometrika*, vol. 70, N° 1, 4155.
- Rothe, C. (2012). “Partial Distributional Policy Effects”. *Econometrica*, 80, 2269-2301.

Rubin, D. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". *Journal of Educational Psychology*, 66, 688-701.

Williamson, R. C. and Downs T. (1990). "Probabilistic Arithmetic I: Numerical Methods for Calculating Convolutions and Dependency Bounds". *International Journal of Approximate Reasoning* 4, 89-158.

Anexos

Anexo 1. Funciones de distribución potenciales a partir de los datos disponibles²⁴.

Empecemos de la definición de las funciones de distribución de la ecuación (5):

$$\begin{aligned} F_{KY_1}(Ky) &= \Pr[KY_1 \leq Ky] \\ F_{KY_0}(Ky) &= \Pr[KY_0 \leq Ky] \\ F_{KY_0|D}(Ky|1) &= \Pr[KY_0 \leq Ky|D = 1] \end{aligned} \quad (5')$$

Recordemos que se asumía la presencia de un único periodo anterior al tratamiento (denotado por $t = 0$) y que $Y_1 = (Y_{10}, Y_{11}, \dots, Y_{1T})'$ e $Y_0 = (Y_{00}, Y_{01}, \dots, Y_{0T})'$ son los vectores de $T \times 1$ que contienen a los resultados potenciales en todos los momentos en el tiempo, mientras que $X = (X_1', \dots, X_T)'$ es la matriz que contiene a las variables de control para todo el horizonte temporal. La variable KY_d representa la transformación de los resultados potenciales que no dependan de la heterogeneidad no observada.

En primer lugar, podemos aplicar la ley de expectativas iteradas para condicionar todo a las variables de control en todo momento en el tiempo.

$$F_{KY_1}(Ky) = \mathbb{E}[\Pr[KY_1 \leq Ky|X]]$$

Por el supuesto (S1) sabemos que la transformación KY_1 es independiente al tratamiento D condicional a las variables de control X , por lo que la función de distribución es:

$$F_{KY_1}(Ky) = \mathbb{E}[\Pr[KY_1 \leq Ky|X, D = 1]]$$

Utilizando la relación entre resultados potenciales y observados, $Y_t = DY_{1t} + (1 - D)Y_{0t}$ para $t = 0, \dots, T$ y la definición de probabilidad $\Pr[A] = E[I\{A\}]$,

$$F_{KY_1}(Ky) = \mathbb{E}[\Pr[KY \leq Ky|X, D = 1]] = \mathbb{E}[\mathbb{E}(DI\{KY \leq Ky\}|X, D = 1)]$$

Donde $Y = (Y_0, Y_1, \dots, Y_T)'$ es el vector de $T \times 1$ que contiene a los resultados observados en el tiempo. Luego, usamos que $\mathbb{E}[S|X] = p(X)\mathbb{E}(S|X, D = 1) + (1 - p(X))\mathbb{E}(S|X, D = 0)$, donde S es una variable aleatoria. Considerando que el *propensity score* se basa en características iniciales no afectadas por el tratamiento y que solo contamos con una encuesta antes del tratamiento, la expresión anterior quedaría expresada únicamente en términos de $p(X_0)$,

$$F_{KY_1}(Ky) = \mathbb{E} \left[\frac{1}{p(X_0)} \mathbb{E}(DI\{KY \leq Ky\}|X) \right]$$

²⁴ Adaptado de Firpo (2007).

Finalmente, usando la ley de expectativas iteradas una vez más, llegamos al resultado deseado.

$$F_{KY_1}(Ky) = \mathbb{E} \left[\frac{D}{p(X_0)} I\{KY \leq Ky\} \right]$$

Se pueden obtener los resultados análogos para $F_{KY_0}(Ky)$ y $F_{KY_0|D}(Ky|1)$ utilizando el mismo procedimiento.

Anexo 2. Desigualdad de Cambanis-Simons-Stout (CSS)

Lema: Sean R y S dos variables aleatorias con distribuciones marginales fijas y conocidas, F_R y F_S ; respectivamente. Bajo el supuesto de que tanto R como S tienen varianzas finitas, se cumple que:

$$\int_0^1 F_R^{-1}(1-u)F_S^{-1}(u)du \leq E(RS) \leq \int_0^1 F_R^{-1}(u)F_S^{-1}(u)du$$

Los límites son finitos y finos.

Anexo 3. Aplicación de la desigualdad de CSS al modelo transformado²⁵

Partamos de los resultados mostrados en la ecuación (5). Si denotamos $W = 1/p(X_0)$ y $V = 1/[1 - p(X_0)]$ y aplicamos la Ley de Expectativas Iteradas:

$$\begin{aligned} F_{KY_1}(Ky) &= E[DE(I\{KY \leq Ky\}W|D)] \\ F_{KY_0}(Ky) &= E[(1-D)E(I\{KY \leq Ky\}V|D)] \\ F_{KY_0|D}(Ky|1) &= \frac{1}{p_1}E[(1-D)E(I\{KY \leq Ky\}V/W|D)] \end{aligned} \quad (5'')$$

Cada expresión posee la esperanza condicional del producto de dos variables aleatorias ($I\{KY \leq Ky\}$ con W , V y V/W , respectivamente) por lo que se puede aplicar de manera directa el Lema III para obtener un resultado muy similar al mostrado en el Teorema III.1:

Se cumple para $d = 0,1$ que $B_{Kd}^I \leq F_{KY_d}(Ky) \leq B_{Kd}^S$, donde:

$$\begin{aligned} B_{K1}^I &= E \left[D \int_0^1 F_{IKY|D}^{-1}(1-u|D) F_{W|D}^{-1}(u|D) du \right] \\ B_{K1}^S &= E \left[D \int_0^1 F_{IKY|D}^{-1}(u|D) F_{W|D}^{-1}(u|D) du \right] \\ B_{K0}^I &= E \left[(1-D) \int_0^1 F_{IKY|D}^{-1}(1-u|D) F_{V|D}^{-1}(u|D) du \right] \\ B_{K0}^S &= E \left[(1-D) \int_0^1 F_{IKY|D}^{-1}(u|D) F_{V|D}^{-1}(u|D) du \right] \end{aligned}$$

Asimismo, $F_{KY_1|D}(Ky|1) = E[DI_{KY}]/p_1$ se encuentra identificado, mientras que $F_{KY_0|D}(Ky|1)$ se encuentra identificado a través de $B_{K0|D}^I \leq F_{KY_0|D}(Ky|1) \leq B_{K0|D}^S$, donde:

$$\begin{aligned} B_{K0|D}^I &= \frac{1}{p_1} E \left[(1-D) \int_0^1 F_{IKY|D}^{-1}(1-u|D) F_{V/W|D}^{-1}(u|D) du \right] \\ B_{K0|D}^S &= \frac{1}{p_1} E \left[(1-D) \int_0^1 F_{IKY|D}^{-1}(u|D) F_{V/W|D}^{-1}(u|D) du \right] \end{aligned}$$

Donde $I_{KY} = I\{KY \leq Ky\}$. Para obtener el Teorema III.1, es preciso notar que:

$$F_{IKY|D}^{-1}(u|D) = \begin{cases} 0, & \text{para } u \in [0, 1 - F_{KY|D}(Ky|D)] \\ 1, & \text{para } u \in [1 - F_{KY|D}(Ky|D), 1] \end{cases}$$

Finalmente, reemplazando esta igualdad en el resultado anterior, se obtienen los límites descritos para $F_{KY_1}(Ky)$, $F_{KY_0}(Ky)$ y $F_{KY_0|D}(Ky|1)$.

²⁵ Adaptado de Fan *et al.* (2014a).

Anexo 4. Teorema 3.2 de Fan *et al.* (2014a)

(i) Sea $\mu_d(g) \equiv E[g(\tilde{Y}_d)]$. Entonces, $\mu_d^L(g) \leq \mu_d(g) \leq \mu_d^U(g)$, para $d = 0, 1$, y:

$$\begin{aligned}\mu_1^L(g) &= E \left[D \int_0^1 F_{g(\tilde{Y})|D}^{-1}(1-u|D) F_{W|D}^{-1}(u|D) du \right] \\ \mu_1^U(g) &= E \left[D \int_0^1 F_{g(\tilde{Y})|D}^{-1}(u|D) F_{W|D}^{-1}(u|D) du \right] \\ \mu_0^L(g) &= E \left[(1-D) \int_0^1 F_{g(\tilde{Y})|D}^{-1}(1-u|D) F_{V|D}^{-1}(u|D) du \right] \\ \mu_0^U(g) &= E \left[(1-D) \int_0^1 F_{g(\tilde{Y})|D}^{-1}(u|D) F_{V|D}^{-1}(u|D) du \right]\end{aligned}$$

Sin información adicional, estos límites son finos.

(ii) Sea $\mu_{d|1}(g) \equiv E[g(\tilde{Y}_d)|D = 1]$. Entonces, $\mu_{1|1}(g)$ está identificada: $\mu_{1|1}(g) = E[Dg(\tilde{Y})]/p_1$ y $\mu_{0|1}^L(g) \leq \mu_{0|1}(g) \leq \mu_{0|1}^U(g)$, donde:

$$\begin{aligned}\mu_{0|1}^L(g) &= \frac{1}{p_1} E \left[(1-D) \int_0^1 F_{g(\tilde{Y})|D}^{-1}(1-u|D) F_{V/W|D}^{-1}(u|D) du \right] \\ \mu_{0|1}^U(g) &= \frac{1}{p_1} E \left[(1-D) \int_0^1 F_{g(\tilde{Y})|D}^{-1}(u|D) F_{V/W|D}^{-1}(u|D) du \right]\end{aligned}$$

Sin información adicional, estos límites son finos.

Anexo 5. Estimadores de los límites de promedios contra-factuales

Para obtener estimados consistentes del *propensity score* se puede utilizar la base de control. Bajo el supuesto de que la participación en el programa depende únicamente de características iniciales, el estimador consistente, $\hat{p}(x_0)$, estará basado sobre la base de control antes del tratamiento²⁶.

Fan *et al.* (2014b) definen, en primer lugar, la función cuantil estimada del *propensity score* condicional a $D = d$ como $\hat{F}_{p(x_0)|D}^{-1}(u|d) = \inf\{a: \hat{F}_{p(x_0)|D}(a|d) > u\}$ donde $\hat{F}_{p(x_0)|D}(a|d)$ es la función de distribución acumulada estimada del *propensity score* dado $D = d$. Para $d = 0,1$; esta función estimada se puede expresar como:

$$\hat{F}_{p(x_0)|D}(a|d) = \frac{\sum_{i=1}^N \{\hat{p}(x_0) \leq a, D_i = d\}}{N\hat{p}_d}$$

Utilizando el estimado de la función cuantil del *propensity score*, Fan *et al.* (2014b) obtienen los siguientes estimadores de las demás funciones cuantil:

$$\hat{F}_{W|D}^{-1}(u|d) = \frac{1}{\hat{F}_{p(x_0)|D}(1-u|d)}$$

$$\hat{F}_{V|D}^{-1}(u|d) = \frac{1}{1 - \hat{F}_{p(x_0)|D}(u|d)}$$

$$\hat{F}_{V/W|D}^{-1}(u|d) = \frac{\hat{F}_{p(x_0)|D}(u|d)}{1 - \hat{F}_{p(x_0)|D}(u|d)}$$

²⁶ La extensión al caso en el cual la participación en el tratamiento dependa de la t-ésima diferencia de las características observadas es directa. En este caso, se debería obtener el *propensity score* basado en dicha diferencia y, luego, se realiza el promedio ponderado (por población) de los *propensity scores*. Finalmente, se obtienen las funciones cuantil de un modo análogo al presentado por Corvalán *et al.* (2015) y Fan *et al.* (2014b).

Nota biográfica

Gonzalo Javier Rivera Gallegos

Nació en Lima, el 19 de diciembre de 1991. Economista, bachiller de la Universidad del Pacífico. Magíster en Economía de la Universidad del Pacífico.

Cuenta con cuatro años de experiencia en investigación. Ha participado en consultorías para el Ministerio de Salud, el Ministerio de Inclusión Social y el Banco Mundial. Actualmente, desempeña el cargo de Asistente de Investigación en el Centro de Investigación de la Universidad del Pacífico (CIUP), laborando con el PhD. Pablo Lavado.