



EL COLEGIO DE MÉXICO
CENTRO DE ESTUDIOS ECONÓMICOS

MAESTRÍA EN ECONOMÍA

TRABAJO DE INVESTIGACIÓN PARA OBTENER EL GRADO DE
MAESTRO EN ECONOMÍA

**BECAS EDUCATIVAS: LA IMPORTANCIA DE CONSIDERAR EL
PORCENTAJE QUE LA BECA REPRESENTA RESPECTO AL
GASTO DEL HOGAR PARA MAXIMIZAR SUS EFECTOS**

RAÚL ARANA DE LA ROSA

PROMOCIÓN 2014-2016

ASESOR:

DR. CARLOS CHIAPA LABASTIDA

JULIO 2016

Resumen

El objetivo de esta tesis es investigar el impacto que tienen los programas de becas sobre la retención escolar y la progresión escolar, dependiendo del porcentaje del gasto mensual que las becas representan en el hogar. El análisis presentado será desde una perspectiva alternativa a las evaluaciones de impacto, la idea consiste en basarme en la literatura existente para identificar si existe un impacto y a través de la metodología propuesta encontrar umbrales—medidos como el porcentaje del gasto mensual del hogar—donde se maximicen los efectos en las variables de interés.

La metodología utilizada es el *Generalized Propensity Score*, este método consiste en construir una función de respuesta a diversas dosis de tratamiento utilizando un conjunto de características observables de los estudiantes y de sus familias. Los resultados obtenidos presentan una forma de asignación de recursos que maximiza efectos y que no contempla el incurrir en costos adicionales a los que este tipo de programas ya incurren para obtener un contrafactual creíble.

La forma correcta de llevar a cabo este mecanismo de asignación consiste en seguir apoyando a los hogares con más carencias económicas (los cuales son justamente los que presentan elevadas tasas de abandono escolar en los niños) pero otorgándoles un monto que corresponda al 40% del gasto del hogar. De esta manera se maximizan los efectos de este tipo de programas y se puede apoyar a un mayor número de hogares.

Índice General

Contenido

I. Introducción.....	2
II. El proyecto “Schooling, Income and Health Risk,” detalles de la encuesta realizada y estadísticas descriptivas de los datos utilizados	4
A. El proyecto “Schooling, Income and Health Risk” (SIHR)	4
B. Detalles de la encuesta realizada	7
C. Estadísticas descriptivas de los datos empleados	7
III. Método econométrico	10
IV. Estimación de los resultados.....	16
V. Resultados	20
VI. Conclusiones.....	25
VII. Anexo	27
VIII. Referencias.....	34
IX. Índice de tablas	35
X. Índice de gráficas	35

I. Introducción

El objetivo de esta tesis es investigar el impacto que tienen los programas de becas sobre la retención escolar (número de estudiantes que permanecen estudiando) y la progresión escolar (número de estudiantes que logran ingresar al grado educativo inmediato superior), dependiendo del porcentaje del gasto mensual que las becas representan en el hogar. La importancia de realizar esta investigación radica en que los programas de becas, en general, otorgan montos fijos a sus beneficiarios y, no existe evidencia empírica que demuestre que de esta manera se efficienten los recursos asignados. Al considerar el porcentaje del gasto mensual que las becas representan en el hogar, se pueden encontrar montos que varíen entre cada beneficiario que sean tales que maximicen los efectos que este tipo de programas persiguen y que a su vez minimicen la cantidad de recursos destinados a cada beneficiario. Con esta forma de asignación de becas es posible maximizar los objetivos deseados con este tipo de programas y beneficiar a un número mayor de estudiantes con la misma cantidad de recursos que se destinan a becas que asignan montos fijos a los beneficiarios.

El análisis presentado será desde una perspectiva alternativa a las evaluaciones de impacto dicotómicas tradicionales: tratamiento vs. control. La idea consiste en basarme en la literatura existente para identificar si existe un impacto y, a través de la metodología propuesta, encontrar umbrales—medidos como el porcentaje del gasto mensual del hogar—donde se maximicen los efectos en las variables de interés.

Los estudios que involucran programas de transferencias de efectivo encuentran impactos estadísticamente significativos en aumentos de matrícula (Adelman y Holland, 2015), retornos a la educación, retención escolar, bienestar psicológico, reducción de maternidad temprana, postergación del matrimonio, nutrición (Baird et al., 2015), acumulación de capital, reducción de pobreza (Martínez et al., 2012) y participación de las comunidades en las escuelas (Gertler et al., 2012); efectos que pueden ser los primeros de muchos pasos para mejorar el desempeño escolar. De esta literatura se deriva el interés de analizar en este estudio los impactos sobre la retención y progresión escolar.

Dado que los programas de becas recaban información sobre aquellos estudiantes que solicitan y obtienen el apoyo y no se tiene información sobre aquellos que no reciben la beca, no es sencillo construir un contrafactual creíble. Es decir, los datos existentes solo contienen información sobre los tratados (becados) y no sobre los no tratados (no becados). Así, es complicado crear un grupo de comparación válido por la falta de datos.

Dada esta falta de datos, se propone utilizar como metodología para la medición del impacto causal de los programas el *Generalized Propensity Score*. Este método consiste en construir una función de respuesta a diversas dosis de tratamiento utilizando un conjunto de características observables de los estudiantes y de sus familias.

Los datos a utilizar en esta tesis provienen del estudio “*Schooling, Income and Health Risk*” (SIHR). Estudio que se realizó en el distrito de Zomba en Malawi y que fue diseñado para evaluar los impactos de las transferencias de dinero en diversos aspectos de las áreas de educación, matrimonio, fertilidad, salud, nutrición y comportamiento sexual de mujeres de 13 a 22 años que nunca se habían casado.

Los resultados encontrados indican que los efectos en la retención y progresión escolar se maximizan cuando la beca representa entre el 40% y el 80% del gasto mensual de los hogares, siempre y cuando los beneficiarios estén recibiendo la beca. Una vez que se ha dejado de recibir la beca, en los estudiantes para los cuales la beca representaba el 40% del gasto mensual del hogar, el efecto de retención vuelve a alcanzar su máximo en comparación con los demás estudiantes. Sin embargo, en cuanto a la progresión escolar, estos estudiantes son los que resienten más esta pérdida de ingreso.

Asimismo, los efectos encontrados en la retención y progresión escolar una vez que se dejó de asignar la beca confirman de cierta manera lo encontrado en la literatura. Una vez que se ha dejado de recibir la beca, el costo de oportunidad de mantenerse en la escuela aumenta y para los estudiantes a los que la beca representa un mayor porcentaje del gasto del hogar, ya no es posible seguir en la escuela.

Las implicaciones de estos resultados, en materia de política pública, proponen una nueva forma de asignar becas: en lugar de dar un monto de beca fijo a todos los estudiantes, se podría intentar fijar *el porcentaje que se desea que la beca represente del gasto mensual de cada hogar*. El asignar a cada beneficiario el 40% del gasto mensual de su hogar, permitirá tener los mismos efectos, que cuando se asigna un monto fijo (mayor al 40% del gasto mensual del hogar). Así, con los recursos que se ahorran, se podría beneficiar a un número mayor de estudiantes.

II. El proyecto “Schooling, Income and Health Risk,” detalles de la encuesta realizada y estadísticas descriptivas de los datos utilizados

A. El proyecto “Schooling, Income and Health Risk” (SIHR)

Los datos a utilizar en esta tesis provienen del estudio “*Schooling, Income and Health Risk*” (SIHR). Los datos me fueron compartidos por Sarah Baird, una de las investigadoras principales del estudio. El SIHR fue diseñado para evaluar los impactos de las transferencias de dinero en diversos aspectos de las áreas de educación, matrimonio, fertilidad, salud, nutrición y comportamiento sexual de mujeres de 13 a 22 años que nunca se habían casado, en el corto plazo (mientras duró la beca), en el mediano plazo (en el año posterior a cuando terminó la beca) y en el largo plazo (2 años después de terminado el tratamiento) (Baird et al., 2015). Entre estos resultados, se encuentra la retención escolar, que es una de las variables de estudio de esta tesis.¹

El SIHR es un experimento aleatorizado. Una explicación detallada de la intervención se puede encontrar en Baird et al. (2015) y en los manuales utilizados en las diversas rondas de seguimiento que tuvo el experimento. Aquí se describirá, a grandes rasgos, en qué consiste el SIHR. El énfasis de la descripción se pondrá en la manera en que se utilizarán los datos que fueron recopilados por Baird et al. para realizar el análisis sobre las variables de interés de esta tesis (retención y progresión escolar).

¹ Baird et al. (2015) encuentran que el SIHR incrementó la retención escolar en 17%. Este efecto se identifica comparando un grupo de control y otro de tratamiento que fueron determinados aleatoriamente como parte de un experimento.

El SIHR se realizó en el distrito de Zomba en Malawi. Los investigadores seleccionaron este distrito, entre otras cosas, por su alta tasa de deserción escolar. Esto hace que los datos recopilados sean útiles para realizar la investigación objeto de esta tesis. Aún más, de acuerdo con la Encuesta Integrada de Hogares Malawi-2 (2004/2005), el principal motivo reportado por los hogares, de la deserción de la escuela, eran problemas financieros. Esto es entendible si se considera que el ingreso nacional bruto (INB) per cápita de Malawi es de USD\$760. Este monto de INB per cápita es bajo aún si se le compara con el INB per cápita promedio en el África Sub-Sahariana (USD\$1,973).

Para identificar la muestra para realizar el experimento, se dividió al distrito de Zomba en 550 zonas de enumeración (AE), las cuales contenían en promedio 250 hogares. 176 AE fueron seleccionadas de entre tres estratos diferentes: la ciudad de Zomba (urbanos, 29 AE's), cerca de las zonas rurales (en un radio de 16 kilómetros de la ciudad de Zomba, 119 AE's), y lejos de zonas rurales (28 AE's). La elección de un radio de 16 kilómetros alrededor de la ciudad de Zomba fue arbitraria y se basa principalmente en la consideración de los costos de transporte.

En estos 176 AE's, cada vivienda fue visitada para obtener una lista completa de las mujeres de 13 a 22 años que nunca se habían casado. La población objetivo se dividió en dos grupos principales: las mujeres que estaban fuera de la escuela al inicio del estudio (abandonos de línea de base) y las que estaban en la escuela al inicio del estudio de línea de base (escolares). Este segundo grupo, las mujeres escolares, será la población objetivo de esta tesis. Siguiendo este procedimiento se obtuvo una muestra total de 3,796 mujeres encuestadas. De éstas, 2,907 eran escolares y 889 eran abandonos de línea de base. Así, el tamaño de muestra inicial para esta tesis es de 2,907 mujeres.

De las 176 AE's, 88 fueron puestas en el grupo control y 88 en el tratamiento. De las 88 AE's en tratamiento, a 46 se les asignaron transferencias de efectivo condicionadas a la asistencia regular a la escuela (asistencia de al menos el 80% de días de clases).² A 27 se les

² Ninguna beneficiaria fue retirada del programa por no cumplir esta condición, simplemente se le suspendían las transferencias mientras se mejoraba la asistencia. Es decir, si posteriormente lograba alcanzar una asistencia satisfactoria, los pagos de las transferencias se reanudaban.

asignaron transferencias de efectivo no condicionadas. En las 15 restantes se trató solo a las mujeres que se encontraban en el grupo de abandonos en la línea base.

Ambos grupos de tratamiento recibieron transferencias mensuales que se ofrecieron a las beneficiarias y a sus padres. De hecho, la única diferencia entre el grupo condicionado y el no condicionado radica en la condición de asistir regularmente a la escuela durante el programa. Los montos de las transferencias a los padres variaron aleatoriamente a través de las AE's entre USD\$4, USD\$6, USD\$8 y USD\$10 por mes. Así, cada padre dentro de un AE recibió la misma oferta. Dentro de cada AE, se llevó a cabo un sorteo para determinar el importe de la transferencia a las beneficiarias del programa, que era igual a USD\$1, USD\$2, USD\$3, USD\$4, o USD\$5 por mes.

Para esta tesis, no se hará distinción entre el grupo condicionado y el no condicionado. Es decir, la unión de ambos grupos representa la población objetivo (y muestra) de esta tesis.

Tabla 1. Línea de tiempo de las rondas e intervención

	Fecha de Inicio	Fecha Final
<u>Intervención</u>		
Transferencias de efectivo	Febrero 2008	Diciembre 2009
<u>Rondas de encuestas</u>		
Encuesta de línea basal	Octubre 2007	Febrero 2008
Ronda 2 de recolección de datos	Octubre 2008	Febrero 2009
Ronda 3 de recolección de datos	Enero 2010	Julio 2010
Ronda 4 de recolección de datos*	Marzo 2012	Febrero 2013

* No se tuvo acceso a la ronda 4 de la información debido a que no estará disponible por varios meses.

La Tabla 1 detalla las líneas de tiempo relevantes. Al inicio del estudio (ronda 1), se llevaron a cabo encuestas de hogares, que fueron revisadas y se repitieron en cada ronda de seguimiento a partir de entonces. En la ronda 2, se realizó una encuesta escolar que involucró visitar todas las escuelas a las que asistían las encuestadas en 2008; así como recoger datos de biomarcadores para el VIH, HSV-2 y la sífilis en un grupo seleccionado al azar de 104 AE's. En la ronda 3, además de las encuestas de hogares, encuestas escolares, y la recopilación de datos de biomarcadores para el VIH y el HSV-2, se les aplicaron diversas preguntas a las beneficiadas

para medir su aprendizaje. Los datos utilizados en esta tesis provienen de las rondas 2 y 3 para analizar la retención y progresión escolar.

B. Detalles de la encuesta realizada

Es importante señalar la forma en la que se conformaron los hogares en el SIHR, ya que la investigación realizada se enfoca al porcentaje que la beca representa del gasto del hogar. Un hogar será definido como el conjunto de personas que comparten la gestión del domicilio en común; es decir, que aportan o dependen del presupuesto del domicilio, incluyendo personas que no estén relacionadas como familiares pero que cumplen con esta definición. A su vez, los familiares que por algún motivo no vivan en el domicilio y no dependan del presupuesto del domicilio no son considerados como parte del hogar.

Adicional a lo anterior, es importante señalar el tipo de información que fue recolectada, ya que estas serán utilizadas en la metodología propuesta.³ Esta información consta de información básica sobre cada miembro del hogar, características de la vivienda, información acerca de los diversos activos que pertenecen a cualquier integrante del hogar, consumo de comida, gastos mensuales del hogar, beneficios recibidos por programas gubernamentales y no gubernamentales, información sobre choques económicos negativos en el hogar e información acerca de la escolaridad, salud, fertilidad, matrimonio, comportamientos sexuales, conocimientos del SIDA, actividades sociales y consumo de la beneficiada.

C. Estadísticas descriptivas de los datos empleados

Los datos empleados en esta tesis corresponden al subconjunto de 988 mujeres llamado escolares que fueron asignadas a cualquiera de los grupos de tratamiento. Además, como las variables de interés de esta tesis corresponden a información recolectada en las rondas 2 y 3, la muestra se restringe aún más a las 983 mujeres beneficiarias para las cuales no hubo problemas de atrición.

³ Un listado completo de las variables a utilizar se encuentra en los anexos del documento.

Tabla 2. Estadísticas Descriptivas de la muestra objetivo en la línea basal

	Muestra Total		
	Obs.	Mean	SD
a) Características del jefe del hogar			
Edad	714	47.86	13.17
Nivel de educación en años	750	6.97	4.03
Alfabetizado	750	0.91	0.29
b) Características del cónyuge del jefe del hogar			
Edad	499	41.06	10.50
Nivel de educación en años	513	5.03	3.92
Alfabetizado	513	0.79	0.41
c) Características del hogar			
Edad media de los adultos***	744	38.72	11.04
Nivel de educación media de los adultos***	751	6.77	3.18
Proporción de adultos alfabetizados	751	0.88	0.24
Gasto (Kwacha)*	753	10,841.16	12,567.20
Monto de beca recibido (Kwacha)*	753	1,916.92	982.03
d) Estructura del hogar			
Tamaño	753	6.28	2.24
Número de adultos***	753	2.57	1.27
Número de adultos mujeres***	753	1.42	0.77
Número de adultos hombres***	753	1.15	0.91
Proporción de adultos hombres***	744	0.41	0.25
Número de niños	753	3.71	1.75
Número de niños mujeres	753	2.33	1.29
Número de niños hombres	753	1.36	1.21
Proporción de niños hombres	744	0.38	0.26
e) Características relacionadas a mujeres en edad escolar (13 a 22 años)			
Número de mujeres	753	1.63	0.79
Número de mujeres que van a la escuela	753	1.49	0.72
Proporción de mujeres que van a la escuela	751	0.92	0.20
Proporción de hogares que no envían mujeres a la escuela	751	0.02	0.14
Proporción de hogares que envían a todas las mujeres a la escuela	751	0.85	0.35
Proporción de hogares que envían menos de todas las mujeres a la escuela	751	0.13	0.33
Proporción de mujeres que se mantuvieron en la escuela (retención)**	2149	0.91	0.29
Proporción de mujeres que avanzaron de grado (progresión)**	2146	0.80	0.40

*La equivalencia utilizada en 2007 fue de USD\$1 por 140 kwachas.

**Dato obtenido con las niñas puestas en el grupo de control en la línea basal.

***Un adulto se considera alguien que tiene 18 años o más.

Tabla 3. Estadísticas Descriptivas en la línea basal de la muestra balanceada que cumple con la condición de soporte común

	Muestra Balanceada		
	Obs.	Mean	SD
a) Características del jefe del hogar			
Edad	395	48.93	13.08
Nivel de educación en años	423	6.23	3.68
Alfabetizado	423	0.90	0.31
b) Características del cónyuge del jefe del hogar			
Edad	282	42.51	10.54
Nivel de educación en años	293	3.88	3.35
Alfabetizado	293	0.72	0.45
c) Características del hogar			
Edad media de los adultos***	417	39.43	11.40
Nivel de educación media de los adultos***	421	6.11	2.92
Proporción de adultos alfabetizados	421	0.86	0.25
Gasto (Kwacha)*	423	7,130.66	6,653.68
Monto de beca recibido (Kwacha)*	423	1,928.65	1,026.23
d) Estructura del hogar			
Tamaño	423	6.17	2.14
Número de adultos***	423	2.49	1.16
Número de adultos mujeres***	423	1.42	0.78
Número de adultos hombres***	423	1.07	0.79
Proporción de adultos hombres***	417	0.40	0.24
Número de niños	423	3.68	1.76
Número de niños mujeres	423	2.32	1.31
Número de niños hombres	423	1.35	1.17
Proporción de niños hombres	418	0.34	0.26
e) Características relacionadas a mujeres en edad escolar (13 a 22 años)			
Número de mujeres	423	1.58	0.75
Número de mujeres que van a la escuela	423	1.43	0.68
Proporción de mujeres que van a la escuela	422	0.93	0.19
Proporción de hogares que no envían mujeres a la escuela	422	0.02	0.14
Proporción de hogares que envían a todas las mujeres a la escuela	422	0.87	0.33
Proporción de hogares que envían menos de todas las mujeres a la escuela	422	0.11	0.31
Proporción de mujeres que se mantuvieron en la escuela (retención)**	2149	0.91	0.29
Proporción de mujeres que avanzaron de grado (progresión)**	2146	0.80	0.40

*La equivalencia utilizada en 2007 fue de USD\$1 por 140 kwachas.

** Dato obtenido con las niñas puestas en el grupo de control en la línea basal.

***Un adulto se considera alguien que tiene 18 años o más.

Finalmente, es importante mencionar que durante las rondas 2 y 3, además de preguntarles a las niñas sobre su situación actual en cuestión de escolaridad, también se les preguntó su situación escolar en las rondas anteriores. Al hacer una comparación entre sus respuestas en la ronda actual con lo contestado en las rondas anteriores, nos quedamos con aquellas observaciones cuyas respuestas fueron congruentes durante las 3 rondas con las que se cuenta. Este procedimiento dejó un total de 753 observaciones.

La Tabla 2 presenta algunas características descriptivas acerca del jefe del hogar, la estructura del hogar y algunos datos sobre las variables de interés de esta tesis. En relación a este conjunto de características, obsérvese que tanto el jefe del hogar como su cónyuge no tienen, en promedio, terminada la educación primaria.⁴ Por otra parte, si bien las tasas de alfabetización parecen altas (entre el 80% y el 90%), esto puede ser el resultado de la forma en que se construyó esta variable: un individuo se considera alfabetizado si logró terminar al menos un año de educación a cualquier nivel escolar. Así que es muy probable que se esté sobre-estimando el porcentaje de la población que sabe leer y escribir.

El método utilizado en esta tesis contiene una condición clave para llevarse a cabo. Esta condición es la de soporte común. Más adelante se explicará con detalle en qué consiste. Lo relevante, de momento, es mencionar que una vez que se impone esta condición, se restringe la muestra a un subconjunto de individuos que son comparables entre ellos mismos. El tamaño de este subconjunto es de 423 observaciones. Con este subconjunto, se calcularon las mismas estadísticas descriptivas reportadas en la Tabla 2. La Tabla 3 reporta los resultados.

III. Método econométrico

Lo que se quiere estudiar es el impacto de las transferencias recibidas, como porcentaje del gasto de cada hogar, sobre la retención y progresión escolar. Esto es importante porque la mayoría de estos programas tienen como objetivo propiciar la permanencia y conclusión de estudios de niños que, por su condición económica, necesitan de este apoyo. Sin embargo, un problema en común, radica en la cobertura de estos programas. El número de solicitudes

⁴ En Malawi, la primaria consta de 8 grados.

recibidas de estudiantes que necesitan de este tipo de apoyos, para continuar con sus estudios, rebasa los montos presupuestarios destinados.

Derivado de lo anterior surge la importancia de considerar el porcentaje que la beca representa del gasto del hogar como herramienta de asignación de estos recursos. La mayoría de los programas asigna montos fijos (estrictamente positivos) a sus beneficiarios y cero recursos a los no beneficiarios. De aquí que normalmente se utilicen métodos tradicionales de evaluación de impacto, en los que la variable de tratamiento toma sólo dos valores (tratado vs. no tratado), para estudiar los efectos de estos programas. Sin embargo, al asignar un monto fijo para todos los beneficiarios, implícitamente se genera un número infinito de tratamientos (el intervalo $(0, \infty)$) cuando se considera el porcentaje que el monto de la beca representa del gasto familiar. Los métodos tradicionales de evaluación de impacto no se prestan a analizar tanta información. No obstante, el lograr explotar tanta riqueza de información puede derivar en maximizar la eficiencia del gasto en programas de becas alrededor del mundo. La metodología del *Generalized Propensity Score* (GPS), parece ser ideal para alcanzar el análisis deseado.

Con los datos del SIHR, es posible observar las variables de resultado de interés, el porcentaje que las transferencias recibidas representan del gasto del hogar y otras variables observables. El reto es, a partir de estos datos observables, construir conjuntos comparables de individuos cuyas transferencias recibidas representen distintos porcentajes del gasto del hogar y checar cómo varían las variables de resultado de interés conforme varía el porcentaje que la transferencia representa del gasto del hogar. El GPS nos permite hacer justo esto.

La descripción de la metodología del GPS presentada a continuación está basada en Flores et al., (2012) y Bia et al. (2013). A su vez, los supuestos utilizados durante esta descripción están basados en el trabajo de Hirano e Imbens (2004).

La metodología propuesta estima una función dosis-respuesta continua que relaciona cada valor de la variable de tratamiento (dosis) al resultado de una variable de interés, en este caso, relaciona el porcentaje que la beca representa del gasto del hogar con la retención y progresión escolar. Formalmente, considérese un conjunto de N observaciones, y denótese cada una

con $i = 1, \dots, N$. Para cada observación i existe un conjunto de resultados potenciales $\{Y_i(t)\}_{t \in T}$, donde T es un subconjunto de los reales. El propósito de esta metodología es estimar una función dosis-respuesta promedio, $\mu(t) = E[Y_i(t)]$.

Para cada observación i , se observa un conjunto de covariables pre-tratamiento descritas en la sección II.B y listadas en el Anexo, X_i , la dosis de tratamiento recibida (el porcentaje que la beca representa del gasto del hogar), T_i , y su correspondiente valor de resultado para la dosis de tratamiento, $Y_i = Y_i(T_i)$, retención y progresión escolar.

El supuesto fundamental de esta metodología radica en que la dosis de tratamiento a la cual son expuestos los individuos es *weakly unconfounded* dado el conjunto de covariables pre-tratamiento; es decir, $Y_i(t) \perp T_i | X_i$, para toda $t \in T$. Los datos obtenidos del SIHR conforman un conjunto de covariables pre-tratamiento y, además, la aleatorización realizada para asignar montos de beca, permite que exista esta independencia condicional entre las variables de interés (retención y progresión escolar) con el tratamiento (porcentaje que la beca representa del gasto del hogar) dadas las covariables.

Este supuesto es conocido como *weak unconfoundedness* porque solamente se requiere independencia condicional para cada resultado potencial $Y_i(t)$, en vez de independencia conjunta para todos los resultados potenciales. Este supuesto descarta cualquier selección sistemática de la asignación de la dosis de tratamiento que pudiese deberse a características no observadas que no hayan sido capturadas en las covariables observadas X_i . Es decir, que exista alguna característica X_i para la cual, se le asigne una beca que represente un determinado porcentaje del gasto del hogar. Sin embargo, la aleatorización con la que se construyó el tratamiento, impide la existencia de este tipo de fenómenos.

Bajo el supuesto de *weak unconfoundedness*, la función promedio que relaciona el porcentaje que la beca representa del gasto del hogar con la variable de interés puede ser obtenida al estimar resultados promedio en sub-conjuntos formados por las covariables X_i y las diferentes dosis de tratamiento T_i .

Bajo *weak unconfoundedness*, podemos aplicar las técnicas del GPS con tratamiento continuo. Sea $r(t, x) = f_{T|X}(t|x)$ la densidad condicional del tratamiento dadas la covariables X_i , el GPS se define como $R_i = r(T_i, X_i)$. Para utilizar este GPS se debe cumplir una condición de balanceo, es decir, entre los sub-conjuntos formados con los datos con el mismo valor de $r(t, x)$, la probabilidad de que $T = t$ no debe depender de X . Es decir, se tendrán estudiantes para los cuales el escalar calculado tiene similar magnitud dentro de cada sub-conjunto y que el porcentaje que la beca representa del gasto del hogar es distinto.

El supuesto de *weak unconfoundedness*, combinado con la propiedad de balanceo, implica que la asignación del tratamiento dado el GPS es *weakly unconfounded*. Formalmente, $f_T(t|r(t, X_i), Y_i(t)) = f_T(t|r(t, X_i))$ para toda $t \in T$.

Entonces, cualquier sesgo asociado con diferencias en la distribución de las covariables X_i a través de los sub-conjuntos formados con los datos puede ser eliminado usando el GPS. Formalmente, se puede demostrar que si la asignación del tratamiento es *weakly unconfounded* dadas las covariables pre-tratamiento X_i , entonces $\mu(t) = E[\beta(t, r(t, X_i))]$, donde $\beta(t, r) = E[Y_i(t)|r(t, X_i) = r] = E[Y_i|T_i = t, R_i = r]$. Lo cual nos indica que podemos construir una función promedio que relacione los diversos porcentajes que la beca representa del gasto del hogar con las variables de interés usando el GPS calculado.

El resultado anterior sugiere estimar la función dosis-respuesta en t usando promedios parciales, lo cual es como estimar el promedio de una función de regresión utilizando solo algunos regresores y dejando fijos otros. En este caso, el regresor que se deja fijo es la dosis del tratamiento al nivel t en $\beta(t, r(t, X_i))$.

La función dosis-respuesta puede ser estimada usando el GPS al estimar la esperanza condicional de Y dados T y R , $E[Y_i|T_i = t, R_i = r]$ y después usar promedios parciales sobre esta esperanza condicional para obtener el valor de la función dosis-respuesta en t utilizando $\beta(t, r(t, X_i))$. El reto es, proponer una función que relacione el porcentaje que la beca representa del gasto del hogar con las variables de interés y que sea metodológicamente acertada.

Es decir, existen diversos mecanismos para obtener la función dosis-respuesta usando una aproximación de promedios parciales al asumir una forma funcional de $E[Y_i|T_i = t, R_i = r]$. Sin embargo, no existen razones *ex-ante* para asumir alguna función en particular, razón por la cual se trabajará un enfoque distinto, un enfoque no paramétrico.

Para estimar la función dosis-respuesta, se pueden emplear promedios parciales basados en estimadores kernel no paramétricos para estimar $E[Y_i|T_i = t, R_i = r]$. Sea $K(u)$ una función kernel con las propiedades usuales,⁵ sea h un ancho de banda que satisface $h \rightarrow 0$ y $Nh \rightarrow \infty$ cuando $N \rightarrow \infty$, y $K_h(u) = h^{-1}K(u/h)$.

Defínase así $\beta(t, r; h, K_h)$ como el estimador no paramétrico de $\beta(t, r) = E[Y_i|T_i = t, R_i = r]$ basado en el kernel $K_h(u, v)$ y el ancho de banda h . Entonces, el estimador no paramétrico de promedios parciales de la función dosis-respuesta en t está dado por: $\hat{\mu}(t) = \frac{1}{N} \sum_{i=1}^N \hat{\beta}(t, \bar{R}_i^t; h, K_h)$ donde $\bar{R}_i^t = r(t, X_i)$.

En adición a estimar la función dosis-respuesta en un marco de promedios parciales, también se le aproxima un peso a los GPS. Intuitivamente, se le asignarán distintos pesos a los GPS dentro de cada sub-conjunto, ya que si bien los sub-conjuntos fueron formados con individuos con valores similares de GPS, a las pequeñas diferencias existentes entre estos se les asignará un peso diferente, siendo las más cercanas a la mediana del sub-conjunto las que tendrán mayor peso. Sea $w(T, X; t)$ una función que depende del tratamiento y las covariables tal que $E[w(T, X; t)|X]$ existe y es diferente de 0. Se puede mostrar que bajo *weak unconfoundedness*, la función dosis-respuesta en t es (Flores, 2005):

$$\mu(t) = E \left[\frac{w(T_i, X_i; t) Y_i}{E[w(T_i, X_i; t)|X_i]} \right].$$

⁵ Sean k_1 y k_2 kernels en φ , $a > 0$, $f: \varphi \rightarrow \mathbb{R}$, $\emptyset: \varphi \rightarrow \mathbb{R}^m$ y k_g un kernel en \mathbb{R}^m . Entonces:

1. $k(u, v) = k_1(u, v) + k_2(u, v)$
2. $k(u, v) = a * k_1(u, v)$
3. $k(u, v) = k_1(u, v) * k_2(u, v)$
4. $k(u, v) = f(u) * f(v)$
5. $k(u, v) = k_g(\emptyset(u), \emptyset(v))$

A este estimador resultante, lo denotaremos como el estimador IW de $\mu(\mathbf{t})$ y utilizaremos métodos no paramétricos para poder implementarlo.

Sea h una secuencia de números positivos que tienden a 0 cuando $N \rightarrow \infty$, $\Delta = [t - h, t + h]$ y $w(T_i, X_i; t) = \mathbf{1}(T_i \in \Delta)$. Entonces, para un h suficientemente pequeño, se puede aproximar el denominador del estimador IW como $\Pr(T_i \in \Delta | X_i) \approx 2h\bar{R}_i^t$. Con esto, el estimador IW de $\mu(\mathbf{t})$ puede ser escrito como $N^{-1} \sum_{i=1}^N \mathbf{1}(T_i \in \Delta) Y_i [2h\bar{R}_i^t]^{-1}$, donde \bar{R}_i^t es un estimador del GPS en \mathbf{t} .

En este sentido, aún no se le han asignado pesos a los GPS estimados, pero si se define a $w(T_i, X_i; t) = K_h(T_i - t)$ es posible asignarle más peso a los individuos que se encuentren más cercanos a la dosis de tratamiento mediana \mathbf{t} . A su vez, se normalizan los pesos de los estimadores para que estos sumen 1 y se estima la función dosis-respuesta en \mathbf{t} de la siguiente manera:

$$\hat{\mu}(\mathbf{t}) = \frac{\sum_{i=1}^N \widetilde{K}_{h,X}(T_i - t) Y_i}{\sum_{i=1}^N \widetilde{K}_{h,X}(T_i - t)}$$

donde $\widetilde{K}_{h,X}(T_i - t) = K_h(T_i - t) / \bar{R}_i^t$. Se prefiere este estimador ya que evita el sesgo en las cotas y también es más fácil de trabajar con la estimación derivada. La forma explícita del estimador sería:

$$\hat{\mu}(\mathbf{t})_{IW} = \frac{D_0(\mathbf{t})S_2(\mathbf{t}) - D_1(\mathbf{t})S_1(\mathbf{t})}{S_0(\mathbf{t})S_2(\mathbf{t}) - S_1^2(\mathbf{t})}$$

$$\text{Dónde: } S_j(\mathbf{t}) = \sum_{i=1}^N \widetilde{K}_{h,X}(T_i - t) (T_i - t)^j \text{ y } D_j(\mathbf{t}) = \sum_{i=1}^N \widetilde{K}_{h,X}(T_i - t) (T_i - t)^j Y_i.$$

Se obtuvo un estimador usando un método no paramétrico de promedios parciales y estimadores IW, además, se escoge un ancho de banda basado en el proceso de Fan et al. (1996), el cual está basado en estimar los términos desconocidos de un polinomio de grado $p+3$, con p el orden del polinomio ajustado.

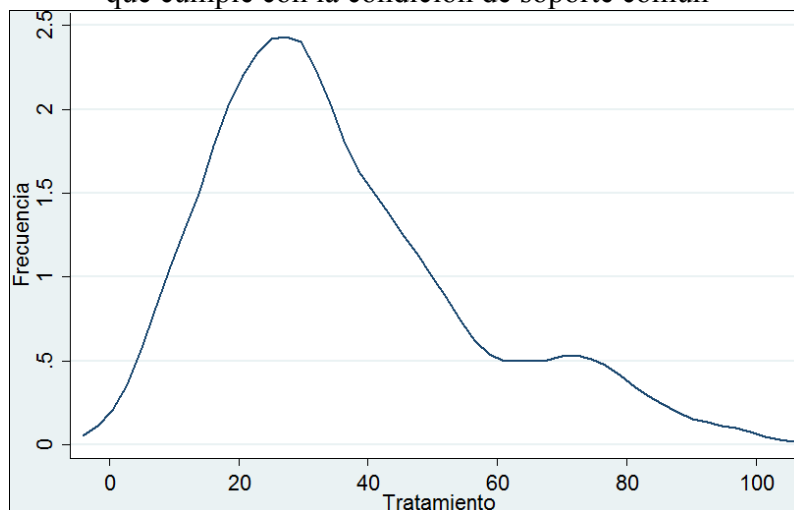
IV. Estimación de los resultados

En esta sección vamos a estimar la función dosis-respuesta promedio de las variables de interés (retención y progresión) para 3 años, donde la dosis de tratamiento corresponde al porcentaje del gasto del hogar que la beca recibida representa.

Nótese que el porcentaje que la beca representa del gasto del hogar cumple el supuesto base de *weak unconfoundedness* ya que todas las covariables (presentadas en la sección II.B) utilizadas en el cálculo del GPS son pre-tratamiento y, además, el monto de beca recibido tanto por el jefe del hogar como por la beneficiaria directa del programa fueron asignados de manera aleatoria.

Siguiendo el trabajo de Flores et al. (2012), el primer paso para poner en práctica el método del GPS consiste en modelar la distribución *condicional* del tratamiento dadas las covariables, es decir, el GPS. La Gráfica 1 muestra la distribución *no* condicional del tratamiento. De esta gráfica se desprende que si bien hay un 67% de beneficiarias para las cuales la transferencia representa menos del 40% del gasto del hogar, hay una masa positiva de hogares cuya transferencia representa diversos porcentajes del gasto del hogar.

Gráfica 1. Distribución no condicional del tratamiento de la muestra balanceada que cumple con la condición de soporte común



Si bien es importante haber visto que en efecto existen “diversas dosis de tratamiento,” como se ha indicado, la distribución que se debe modelar es la distribución condicional del tratamiento, dadas las covariables. Para hacerlo, en general, tenemos que estimar $g[E(T)]$ donde g es una función de enlace, T es el tratamiento y este tratamiento se distribuye de acuerdo a una función de distribución F . Varias especificaciones de $g[E(T)]$ pueden ser obtenidas dependiendo de la elección que se haga de la distribución F para T y de la forma funcional para g .

Así, el primer paso consiste en estimar varias especificaciones plausibles por máxima verosimilitud. Una vez hechas estas estimaciones, se debe elegir el modelo que mejor se ajusta a los datos. Para determinar qué modelo elegir, se pueden utilizar tres distintos criterios que, al final del día, apuntan al mismo modelo: la medida de la desviación (McCullagh y Nelder, 1989), el criterio de información de Akaike y el valor del registro de la función de máxima verosimilitud. Las reglas de decisión son las siguientes. Por un lado, un menor valor de la medida de desviación y del criterio de información de Akaike indican un mejor ajuste del modelo a los datos. Por otro lado, si se pone atención al valor de máxima verosimilitud, el modelo con el mayor valor de máxima verosimilitud es el que mejor se ajusta a los datos.

Las distribuciones consideradas fueron la normal, la normal inversa y la gamma. Por otro lado, las funciones de enlace que se probaron para cada distribución fueron la función identidad, la función logarítmica y las funciones potencias inversas 1 y 0.5. Se empleó el criterio de información de Akaike para elegir entre las diferentes distribuciones y las otras dos medidas para elegir entre las funciones de enlace (Hardin y Hilbe, 2007).

La Tabla 4 muestra los valores de los distintos estadísticos que permiten elegir al modelo que mejor se ajusta a los datos para las distintas combinaciones de funciones de distribución y enlace. De los resultados mostrados en esta tabla se desprende que, aunque los criterios de decisión muestran que la distribución que mejor ajusta al tratamiento, dadas las covariables, es la normal inversa con función de enlace potencia 0.5, hay que hacer mención que, en el proceso de estimación, el método utilizado (máxima verosimilitud) nunca convergió. Por esta razón, se optó

por utilizar la función de distribución y la función de enlace que tuvieran los valores más pequeños de los estadísticos de desviación y Akaike y *que convergiera*.

Tabla 4. Distribución Condicional del Tratamiento

Distribución	Criterio	Función Enlace			
		Identidad	Logarítmica	Potencia 1	Potencia 0.5
Normal	Akaike	2.1231	-0.3247	2.1231	0.5911
	Desviación	163.4612	20.5570	163.4612	37.3475
	Verosimilitud	-493.3609	287.2643	-493.3609	62.4696
Normal Inversa	Akaike	3.51 ^{e12}	-0.4629	3.51 ^{e12}	-0.6572
	Desviación	2.64 ^{e15}	994.1615	2.64 ^{e15}	907.8197
	Verosimilitud	-1.32 ^{e15}	408.2644	-1.32 ^{e15}	451.4353
Gamma	Akaike	106,579.75	0.2516	681,717.76	0.2367
	Desviación	80,255,126	201.1684	5.13 ^{e08}	189.9959
	Verosimilitud	-40,127,251	211.2864	-2.57 ^{e08}	216.8726

Así, las siguientes combinaciones de funciones de distribución y enlace también fueron descartadas por no cumplirse la condición de convergencia: normal inversa con función de enlace logarítmica, y normal con función de enlace logarítmica. La combinación óptima (restringida) de funciones de distribución y enlace fue la distribución gamma con función de enlace potencia 0.5. Una vez hecha esta selección, el siguiente paso es imponer la condición de soporte común y, posteriormente, checar el balanceo de la muestra resultante; es decir, checar qué tan comparables son las observaciones de la muestra final independientemente de la dosis de tratamiento que hayan recibido (i.e., del porcentaje del gasto del hogar que representa la transferencia que reciben).

Para encontrar la sub-muestra que cumpla la condición de soporte común se divide el rango del tratamiento en 5 quintiles, para cada quintil, se calcula el valor del GPS para todos los individuos en el nivel mediano del tratamiento en ese quintil. Los individuos que pertenecen a la región de soporte común de cada quintil son aquellos individuos que se encuentran en el conjunto:

$$CS_q = \{i: R_i^q \in [\max\{\min_{\{j:Q_j=q\}} R_j^q, \min_{\{j:Q_j \neq q\}} R_j^q\}, \min\{\max_{\{j:Q_j=q\}} R_j^q, \max_{\{j:Q_j \neq q\}} R_j^q\}]\}$$

donde R_j^q representa el GPS de cada individuo calculado con el nivel mediano del tratamiento.

Una vez obtenidos estos 5 conjuntos, nos quedamos con aquellos individuos que son comparables a través de los 5 quintiles simultáneamente. Es decir, la sub-muestra que cumple con la condición de soporte común es aquel conjunto resultante de la intersección de estos 5 conjuntos (individuos con un valor del GPS tal, que pueden pertenecer a cualquiera de los 5 quintiles).

Una vez realizado este procedimiento, la muestra original de 753 observaciones se reduce a 423 observaciones y con esta muestra final se controla la “comparabilidad” de las observaciones; es decir, se lleva a cabo la prueba de balanceo. Dado que el tratamiento es continuo en este caso, la prueba de balanceo que se aplica consiste en estimar un modelo sin restricciones que contenga los valores del GPS hasta un término cúbico y *todas* las covariables con las que se construyó el GPS. Después, se comparará este modelo sin restricciones contra un modelo restringido donde se impone que los coeficientes de las covariables sean iguales a cero. La intuición detrás de esta prueba es cerciorarse que, una vez que se controla por el GPS, no sea necesario controlar por ninguna variable observable adicional. Es decir, el GPS debe resumir toda la información (observable) relevante de todas las observaciones de la muestra y una vez que se controla por esto, se puede considerar que el que alguien haya recibido una mayor o menor dosis de tratamiento se debe al azar.

Tabla 5. Pruebas de Balanceo

Modelo no restringido: T en función de GPS, GPS², GPS³ y X's	
Hipótesis: Los coeficientes de las (300) covariables son cero	
Log verosimilitud de la hipótesis	54.17619
Log verosimilitud de modelo no restringido	74.017228
Estadístico de prueba	39.682074
Valor- <i>p</i>	1
Número de restricciones	300
Modelo no restringido: T en función de GPS, GPS², GPS³ y X's	
Hipótesis: Los coeficientes de los (3) GPS's son cero	
Log verosimilitud de la hipótesis	62.630534
Log verosimilitud de modelo no restringido	74.017228
Estadístico de prueba	22.773388
Valor- <i>p</i>	0.00004502
Número de restricciones	3

La Tabla 5 muestra los resultados de la prueba de balanceo. Como se puede ver en la parte superior de la tabla (valor- $p = 1$), no rechazamos la hipótesis nula de que los coeficientes de las covariables del modelo no restringido (GPS , GPS^2 , GPS^3 y 297 covariables más para un total de 300 restricciones) sean iguales a cero. En cambio, como se puede observar en la parte inferior de la tabla (valor- $p \approx 0$), rechazamos la hipótesis nula de que los coeficientes de las únicas tres covariables del modelo restringido (GPS , GPS^2 , GPS^3) son iguales a cero. Esto indica que, efectivamente, el GPS está capturando de buena manera la información contenida en todas las covariables utilizadas y por lo tanto, una vez que se controla por GPS , GPS^2 , GPS^3 , se puede considerar que el que alguien haya recibido una mayor o menor dosis de tratamiento se debe al azar.

Una vez que tenemos la muestra que cumple con la condición de soporte común y ésta se encuentra balanceada, se procede a estimar las funciones dosis-respuesta. Para implementar estas estimaciones se utilizó el ado.file “drf.ado” de Bia et al. (2013), siguiendo las recomendaciones que los autores hacen sobre las especificaciones para utilizar el programa. En particular, se utilizó el método kernel (Hirano e Imbens, 2004) para la estimación de la función de dosis respuesta y se utilizó el método definido por Fan et al. (1996) para determinar el ancho de banda.

Asimismo, Hirano e Imbens (2004) afirman que los errores estándar asintóticos de la función dosis-respuesta estimada, podrían calcularse utilizando expansiones basadas en ecuaciones de estimación que utilicen los valores del GPS y los parámetros definidos para la construcción de la función dosis-respuesta. Derivado de esto, para cada una de las variables de interés, se procede a utilizar métodos de re-muestreo para estimar errores estándar que servirán para construir intervalos de confianza al 95%.

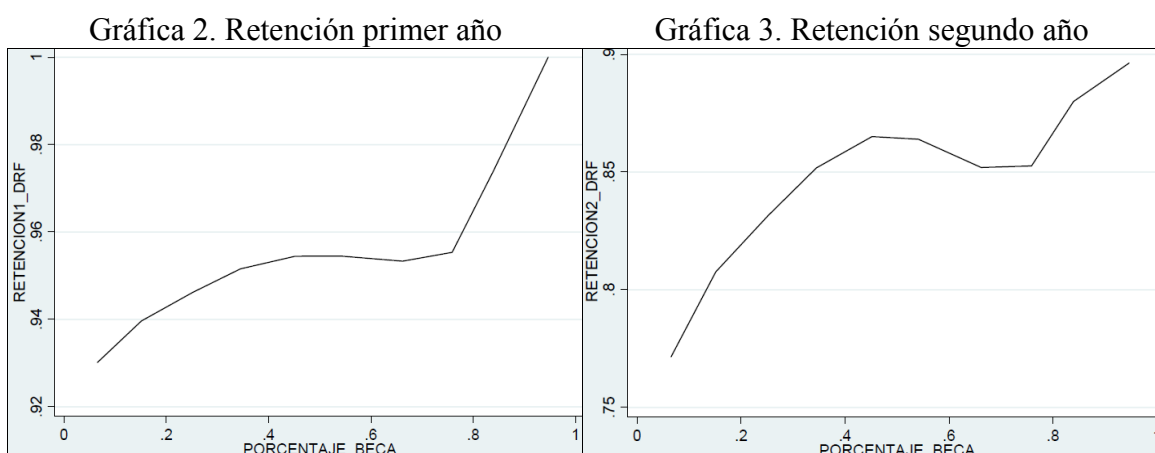
V. Resultados

Los resultados presentados, al ser no paramétricos, se presentan en gráficas. El eje de las abscisas está construido con los porcentajes que la beca representa del gasto del hogar, mientras que el eje de las ordenadas corresponde al promedio de resultados potenciales de interés. Es decir, las gráficas presentadas muestran la relación que estábamos buscando. Una forma

funcional que relacione las diversas dosis de tratamiento, porcentajes que la beca representa del gasto del hogar, con los resultados de interés (retención y progresión escolar).

Especial atención hay que prestar cuando se interpreten estas gráficas. Para poder tener una interpretación causal en el uso de esta metodología, es necesario que la interpretación sea de manera comparativa entre las diversas dosis de tratamiento. Por ejemplo, en la gráfica 3, el efecto promedio sobre la retención del primer año, para aquellas niñas cuya beca representa el 40% del gasto del hogar, es alrededor de 5 puntos porcentuales más efectivo que el efecto sobre la retención del primer año para aquellas niñas cuya beca representa el 20% del gasto del hogar.

Una vez hecha esta aclaración, la interpretación de los umbrales donde los efectos sobre las variables de interés se maximizan, son siempre desde una perspectiva comparativa entre estos umbrales y los demás puntos que representan los porcentajes que la beca representa del gasto del hogar.

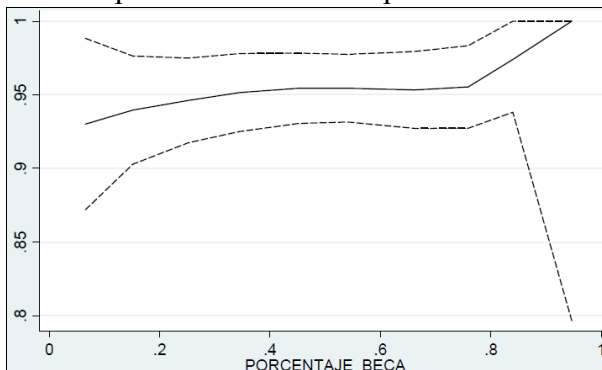


En la gráfica 2, retención escolar para el primer año, se observa que para las niñas que la beca representa un mayor porcentaje del gasto del hogar se maximizan los efectos de la retención, esto al tener una función creciente. Analizando la función hasta antes del punto donde la beca representa el 80% del gasto del hogar, observamos que la función de retención para el primer año es creciente con una retención marginal decreciente, donde los efectos se maximizan llegando al 40% del gasto del hogar y permanecen de esta manera hasta el 80% del gasto del hogar.

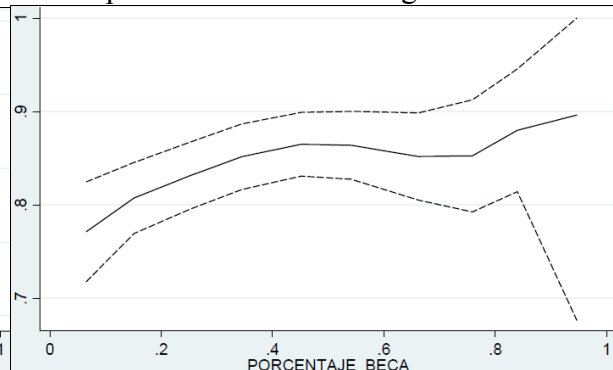
En la gráfica 3, retención para el segundo año, tenemos una función muy parecida a la retención para el primer año. Para esta función dosis-respuesta, la función es creciente con una pendiente mayor a la del primer año, a su vez, podemos identificar que cuando la beca representa entre el 40% y el 80% de la beca, se alcanza un “máximo.”

A su vez, se creó un intervalo de confianza del 95% para los efectos que tuvo la beca sobre la retención escolar. En la gráfica 4, el intervalo de confianza para la retención del primer año, se tiene una tendencia creciente, tal como en la gráfica de dosis-respuesta. Aquí, podemos observar que a partir de cuándo la beca representa más del 80% del gasto del hogar, el intervalo de confianza se expande en gran medida, lo cual, es derivado de que en esta última parte de la distribución, la condición de soporte común únicamente dejó 6 observaciones, haciendo de esta manera que la desviación estándar incremente.

Gráfica 4. Intervalo de Confianza al 95% para la retención del primer año



Gráfica 5. Intervalo de Confianza al 95% para la retención del segundo año



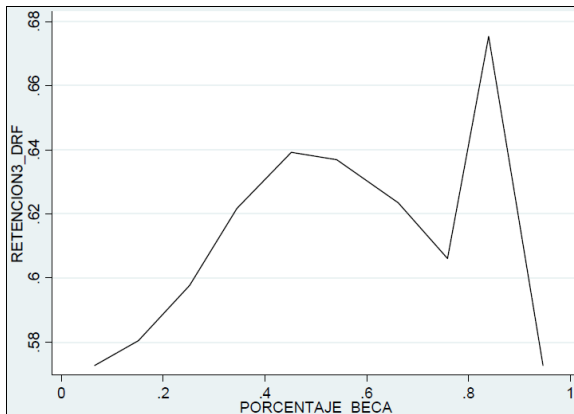
Observando la gráfica 5, referente a los intervalos de confianza para la retención del segundo año, podemos observar que cuando la beca representa más del 80% del gasto del hogar el intervalo de confianza se amplía en gran medida, esto se debe a que las desviaciones estándar existentes en esta parte de la distribución son muy grandes, reafirmando lo descrito anteriormente para las funciones dosis-respuesta del primer y segundo año.

Para el tercer año (2010) tenemos que hacer mención de que las niñas ya no recibieron la beca y, analizando la gráfica 6, podemos observar como la función se maximiza (en la parte

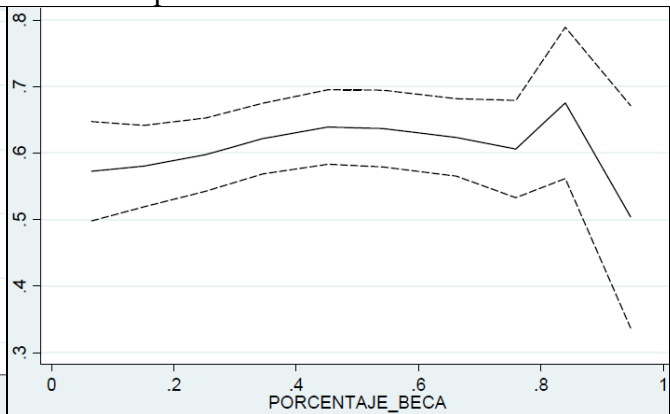
previa al 80% del gasto del hogar) cuando la beca representa el 40% del gasto del hogar, además, existe un declive a partir de este punto hasta cuando la beca representa 80% del gasto del hogar, esto puede deberse a que el costo de oportunidad de estar en la escuela creció para las niñas que se encuentran en esta parte de la distribución.

La gráfica 7, que muestra los intervalos de confianza para la retención del tercer año, nos confirma la existencia de mayores desviaciones estándar en la última parte de la distribución.

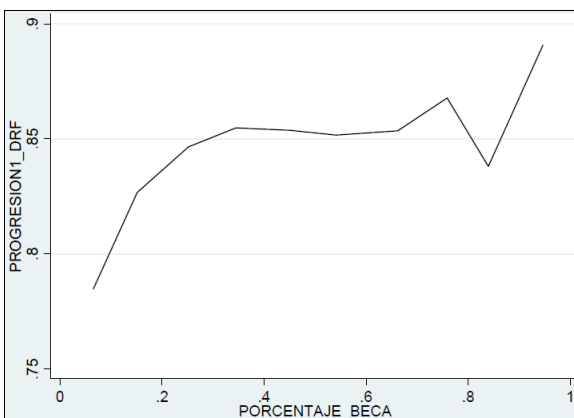
Gráfica 6. Retención tercer año



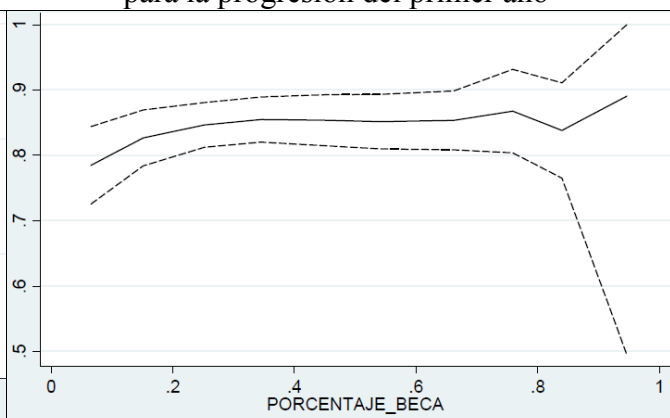
Gráfica 7. Intervalo de Confianza al 95% para la retención del tercer año



Gráfica 8. Progresión primer año



Gráfica 9. Intervalo de Confianza al 95% para la progresión del primer año

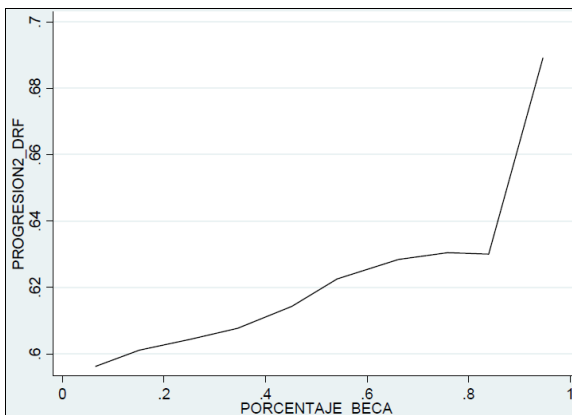


En cuanto a la progresión del primer año de los estudios de las niñas, el avance de un año a otro, podemos observar un efecto parecido al de la retención, es decir, en la gráfica 8 podemos

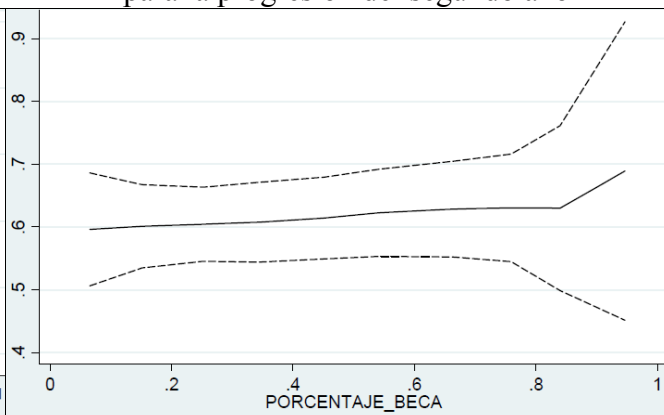
observar una función creciente cuyo máximo se encuentra cerca de cuando beca representa un 40% del gasto del hogar y se mantiene en este punto hasta cuando la beca representa el 80% del gasto del hogar. A su vez, la gráfica 9 referente al intervalo de confianza para la progresión del primer año, nos vuelve a mostrar la gran desviación estándar existente en la última parte de la distribución.

En la gráfica 10 se tiene la función dosis-respuesta para la progresión del segundo año de la beca, en ella podemos observar una función creciente, pero con una pendiente más pequeña, siendo su crecimiento bastante lento. En la gráfica 11, el intervalo de confianza vuelve a presentar la pequeña pendiente que se mostró en la función dosis-respuesta y nuevamente vuelve a mostrar la existencia de una gran desviación estándar al final de la distribución.

Gráfica 10. Progresión segundo año

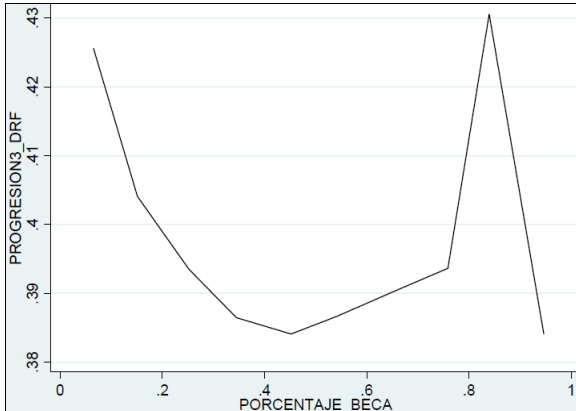


Gráfica 11. Intervalo de Confianza al 95% para la progresión del segundo año

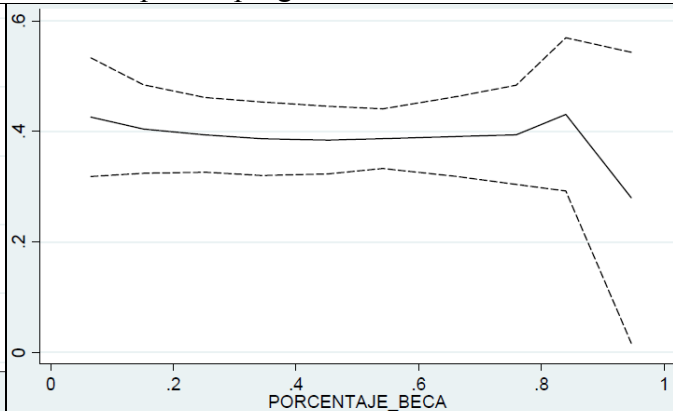


A diferencia de las demás funciones, la función de dosis-respuesta para la progresión durante el tercer año, mostrada en la gráfica 12, es convexa hasta antes de que la función represente más del 80% del gasto del hogar, esto puede deberse a la desmotivación por seguir estudiando ya que para este año ya no contaron con la beca las niñas. Una vez más, la gráfica 13 que muestra el intervalo de confianza para la progresión del tercer año, vuelve a mostrar como el pequeño número de observaciones, que la condición de soporte común dejó, en esta última parte de la distribución de dosis de tratamiento, hace que el intervalo de confianza se vuelva más grande.

Gráfica 12. Progresión tercer año



Gráfica 13. Intervalo de Confianza al 95% para la progresión del tercer año



VI. Conclusiones

El trabajo presentado en esta tesis permite realizar una evaluación de impacto alternativa a aquellas en la que se construye un contrafactual a partir de un único tratamiento. Con el enfoque presentado es posible realizar una evaluación a los programas sin incurrir en costos adicionales a los que usualmente se presentan en este tipo de programas.

La literatura existente sobre programas que implican transferencias de efectivo a sus beneficiados permite concluir que sí existe un impacto en la retención y progresión escolar de los beneficiados, como contribución a este hecho se encontraron umbrales en los cuales estos efectos se maximizan.

La importancia de considerar el porcentaje que la beca representa del gasto del hogar radica en que los efectos de estos programas se maximizan, en la población considerada, cuando la beca representa entre el 40% y 80% del gasto del hogar. Si bien se tienen registros de datos donde el porcentaje es mayor al 80%, estos datos no son del todo fiables dado el pequeño número de observaciones que la condición de soporte común dejó en esta última parte de la distribución, lo cual, puede observarse en los grandes saltos que se muestran en los intervalos de confianza construidos.

La mayoría de los programas de becas existentes otorgan montos fijos de efectivo a sus beneficiados. Cabe resaltar que no existe evidencia empírica que muestre que esta forma de asignarse los recursos maximice los efectos de este tipo de programas. Aquí encontramos rangos para los cuales el efecto de estos programas se maximiza. Se encontró que los efectos se maximizan cuando la beca representa entre el 40% y el 80% del gasto del hogar.

Por esta razón es que una forma de asignación de estos recursos, que maximiza los efectos, sería otorgar a los beneficiados un monto que represente el 40% del gasto del hogar, ya que no existe una diferencia en su efecto en comparación de asignar un monto mayor a este.

Esta forma de asignación depende del gasto del hogar y la forma correcta de llevar a cabo este mecanismo de asignación consiste en seguir apoyando a los hogares con más carencias económicas (los cuales son justamente los que presentan elevadas tasas de abandono escolar en los niños) pero otorgándoles un monto que corresponda al 40% del gasto del hogar. De esta manera se maximizan los efectos de este tipo de programas y se puede apoyar a un mayor número de hogares.

En relación a la validez externa del presente trabajo, cabe recalcar que los aspectos culturales y económicos son muy diferentes entre países, por esta razón el porcentaje donde se maximizan los efectos depende totalmente de cada país o región. Sin embargo, aquí se ha presentado una forma de asignación de recursos que maximiza efectos y que no contempla el incurrir en costos adicionales a los que este tipo de programas ya incurren para obtener un contrafactual creíble. Por lo cual, se podrían estimar los umbrales pertinentes de otros programas de becas en otras regiones del mundo siguiendo este trabajo.

En relación a la validez interna, el problema presentado en la última parte de la distribución de las gráficas dosis-respuesta es derivado de que la condición de soporte común dejó un número pequeño de observaciones. Sin embargo, el acceso a un número mayor de observaciones debería resolver dicho problema y encontrar una función dosis-respuesta más suave en todos sus puntos, dejando de observar los saltos tanto en las funciones dosis-respuesta como en los intervalos de confianza.

VII. Anexo

Tabla 6. Variables utilizadas

Sexo del jefe del hogar
Edad del jefe del hogar
Nivel más alto de educación del jefe del hogar
Grado más alto de educación del jefe del hogar dentro del nivel más alto alcanzado
Dummy sobre si actualmente está estudiando el jefe del hogar
Estatus marital del jefe del hogar
Sexo del cónyuge del jefe del hogar
Edad del cónyuge del jefe del hogar
Nivel más alto de educación del cónyuge del jefe del hogar
Grado más alto de educación del cónyuge del jefe del hogar dentro del nivel más alto alcanzado
Dummy sobre si actualmente está estudiando el cónyuge del jefe del hogar
Estatus marital del cónyuge del jefe del hogar
Sexo del cónyuge del jefe del hogar
Edad del cónyuge del jefe del hogar
Nivel más alto de educación del cónyuge del jefe del hogar
Grado más alto de educación del cónyuge del jefe del hogar dentro del nivel más alto alcanzado
Dummy sobre si actualmente está estudiando el cónyuge del jefe del hogar
Estatus marital del cónyuge del jefe del hogar
Sexo del primer hijo del jefe del hogar
Edad del primer hijo del jefe del hogar
Nivel más alto de educación del primer hijo del jefe del hogar
Grado más alto de educación del primer hijo del jefe del hogar dentro del nivel más alto alcanzado
Dummy sobre si actualmente está estudiando el primer hijo del jefe del hogar
Estatus marital del primer hijo del jefe del hogar
Sexo del segundo hijo del jefe del hogar
Edad del segundo hijo del jefe del hogar
Nivel más alto de educación del segundo hijo del jefe del hogar
Grado más alto de educación del segundo hijo del jefe del hogar dentro del nivel más alto alcanzado
Dummy sobre si actualmente está estudiando el segundo hijo del jefe del hogar
Estatus marital del segundo hijo del jefe del hogar
Propiedad de la vivienda (dueños, en compra, renta, viven gratis, empleados de la casa)
Tipo de vivienda
Tipo de material principal de las paredes de la vivienda
Tipo de material principal del techo de la vivienda
Tipo de material principal del suelo de la vivienda
Número de cuartos de la vivienda sin baños, cocheras o almacenes
Fuente principal de combustible para cocinar de la vivienda
Dummy sobre si hay electricidad en la vivienda
Dummy sobre si la vivienda tiene acceso a una línea telefónica
Dummy sobre si la vivienda tiene acceso a un teléfono celular
Fuente principal de agua para beber en el mes pasado de la vivienda
Tipo de inodoros que la vivienda tiene
Dummy sobre si el hogar tiene mortero
Dummy sobre si el hogar tiene camas

Dummy sobre si el hogar tiene colchones
Dummy sobre si el hogar tiene mesa para comer
Dummy sobre si el hogar tiene mesa para café
Dummy sobre si el hogar tiene sillas
Dummy sobre si el hogar tiene conjuntos de sillones
Dummy sobre si el hogar tiene alacena, cajones o guardaropa
Dummy sobre si el hogar tiene linterna
Dummy sobre si el hogar tiene reloj
Dummy sobre si el hogar tiene plancha
Dummy sobre si el hogar tiene radio
Dummy sobre si el hogar tiene reproductor de CD's
Dummy sobre si el hogar tiene televisión
Dummy sobre si el hogar tiene línea telefónica
Dummy sobre si el hogar tiene teléfono celular
Dummy sobre si el hogar tiene bicicleta
Dummy sobre si el hogar tiene carro
Dummy sobre si el hogar tiene tambor para fabricar cerveza
Dummy sobre si el hogar tiene panga (donde se alimenta a ganado)
Dummy sobre si el hogar tiene azadón
Dummy sobre si el hogar tiene hacha
Dummy sobre si el hogar tiene hoz
Dummy sobre si el hogar tiene paneles solares
Dummy sobre el consumo (en los últimos 7 días) de harina normal
Dummy sobre el consumo (en los últimos 7 días) de harina fina
Dummy sobre el consumo (en los últimos 7 días) de maíz forrajero
Dummy sobre el consumo (en los últimos 7 días) de frijoles
Dummy sobre el consumo (en los últimos 7 días) de guandul
Dummy sobre el consumo (en los últimos 7 días) de maní
Dummy sobre el consumo (en los últimos 7 días) de nkhwani
Dummy sobre el consumo (en los últimos 7 días) de thanaposi
Dummy sobre el consumo (en los últimos 7 días) de hojas verdes salvajes recogidas
Dummy sobre el consumo (en los últimos 7 días) de pescado seco
Dummy sobre el consumo (en los últimos 7 días) de pescado fresco
Dummy sobre el consumo (en los últimos 7 días) de pollo
Dummy sobre el consumo (en los últimos 7 días) de cabra
Dummy sobre el consumo (en los últimos 7 días) de huevos
Dummy sobre el consumo (en los últimos 7 días) de res
Dummy sobre el consumo (en los últimos 7 días) de azúcar
Dummy sobre el consumo (en los últimos 7 días) de azúcar canela
Dummy sobre el consumo (en los últimos 7 días) de aceite para cocinar
Porcentaje del gasto del hogar gastado en alimentos dentro del hogar en el mes pasado
Porcentaje del gasto del hogar gastado en alimentos fuera del hogar en el mes pasado
Porcentaje del gasto del hogar gastado en bebidas no alcohólicas fuera del hogar en el mes pasado
Porcentaje del gasto del hogar gastado en productos escolares en el mes pasado
Porcentaje del gasto del hogar gastado en fuentes de combustible en el mes pasado
Porcentaje del gasto del hogar gastado en transporte en el mes pasado
Porcentaje del gasto del hogar gastado en ropa en el mes pasado
Porcentaje del gasto del hogar gastado en joyería o relojes en el mes pasado

Porcentaje del gasto del hogar gastado en paraguas, bolsas o billeteras en el mes pasado
 Porcentaje del gasto del hogar gastado en teléfono en el mes pasado
 Porcentaje del gasto del hogar gastado en cigarros en el mes pasado
 Porcentaje del gasto del hogar gastado en cortes de pelo en el mes pasado
 Porcentaje del gasto del hogar gastado en periódicos o revistas en el mes pasado
 Porcentaje del gasto del hogar gastado en pilas en el mes pasado
 Porcentaje del gasto del hogar gastado en jabones, pastas y afines en el mes pasado
 Porcentaje del gasto del hogar gastado en productos personales como cremas en el mes pasado
 Porcentaje del gasto del hogar gastado en renta en el mes pasado
 Porcentaje del gasto del hogar gastado en salarios de trabajadores en el mes pasado
 Gasto total en el mes pasado
 Dummy sobre si fueron beneficiados por el programa comida gratis o distribución de maíz
 Dummy sobre si fueron beneficiados por el programa trabajo por comida (PWP)
 Dummy sobre si fueron beneficiados por el programa trabajo por insumos agrícolas
 Dummy sobre si fueron beneficiados por el programa nutrición objetivo
 Dummy sobre si fueron beneficiados por el programa alimento suplementario para niños desnutridos
 Dummy sobre si fueron beneficiados por el programa bono fertilizado
 Beneficiario del programa bono fertilizado en 2005
 Beneficiario del programa bono fertilizado en 2006
 Beneficiario del programa bono fertilizado en 2007
 Monto del bono recibido
 Dummy sobre si fueron beneficiados por el programa distribución de insumos agrícolas gratuita
 Dummy sobre si fueron beneficiados por algún programa escolar para secundaria
 Dummy sobre si fueron beneficiados por algún programa escolar superior
 Dummy sobre si fueron beneficiados con transferencias de dinero por otra organización
 Dummy sobre si fueron beneficiados por algún otro programa
 Si fueron afectados por alguna inundación o sequía en los últimos 3 años
 Si fueron afectados por alguna enfermedad o peste en los cultivos en los últimos 3 años
 Si el ganado murió o fue robado en los últimos 3 años
 Si algún negocio familiar fallo en los últimos 3 años
 Si alguien perdió su salario o trabajo en los últimos 3 años
 Si se irrumpieron ingresos del hogar externos en los últimos 3 años
 Si se afectaron por la caída en precios de cultivos en los últimos 3 años
 Si se afectaron por el aumento en precios de comida en los últimos 3 años
 Si algún miembro del hogar se enfermó o accidente de manera grave en los últimos 3 años
 Si hubo algún nacimiento en el hogar en los últimos 3 años
 Si hubo alguna muerte de algún miembro en los últimos 3 años
 Si murió algún trabajador del hogar en los últimos 3 años
 Si murió algún otro miembro familiar en los últimos 3 años
 Si alguien se divorció en los últimos 3 años
 Si sufrieron algún robo en los últimos 3 años
 Si hubo algún daño fuerte a la vivienda en los últimos 3 años
 Si el padre natural de la beneficiada vive en el hogar
 Número de hermanas mayores a 15 años que la beneficiada tiene
 Si la beneficiada habla Chichewa
 Si la beneficiada habla Chitumbuka
 Si la beneficiada habla Chiyao

Si la beneficiada habla Ingles
Si la beneficiada habla Chingoni
Si la beneficiada habla Chitonga
Si la beneficiada habla Chisena
Si la beneficiada habla Chisenga
A que tribu pertenece la beneficiada
Religión de la beneficiada
Si la beneficiada puede leer en Chichewa
Si la beneficiada puede escribir en Chichewa
Si la beneficiada puede leer en inglés
Si la beneficiada puede escribir en Ingles
Si la beneficiada puede leer en cualquier otro idioma
Si la beneficiada puede escribir en cualquier otro idioma
Si alguna vez ha asistido a la escuela
Nivel más alto de educación de la beneficiada
Grado más alto alcanzado dentro del nivel de alcanzado por la beneficiada
Diploma escolar más alto adquirido por la beneficiada
Edad cuando se atendió por primera vez a la escuela la beneficiada
Si actualmente la beneficiada asiste a la escuela
Si la escuela a la que asiste aún sigue en clases
Tipo de escuela a la que asiste
Si la escuela es de asistencia diaria u otro tipo
Como llega a la escuela
Cuanto tiempo le toma llegar a la escuela
Cuántos días la escuela estuvo abierta en las últimas 2 semanas
Cuántos días asistió en las últimas 2 semanas
Cuál es la probabilidad de que la beneficiada termine este ciclo escolar en opinión de ella
Cuál es la probabilidad de que ella termine la secundaria en su opinión
Si durante el último año la beneficiada dejo de asistir temporalmente a la escuela por más de 2 semanas
Cuanto se gastó en el último año en tutores extra
Cuanto se gastó en el último año en libros y otros materiales
Cuanto se gastó en el último año en uniformes
Cuanto se gastó en el último año en inscripciones escolares
Cuanto se gastó en el último año en contribuciones para la escuela
Cuanto se gastó en el último año en la asociación de padres
Cuanto se gastó en el último año en otros gastos relacionados a los anteriores
Tipo de trabajo al que la beneficiada se dedicó en el último año
Actividad principal en la que dedicó su tiempo durante el último año
Dinero que ganó la beneficiada por su trabajo en el último año
Si el dinero ganado por la beneficiada esta ahorrado de alguna manera
Monto aproximado de lo ahorrado
Opinión personal de la salud de la beneficiada
Como compara su salud con el resto de la gente de su villa
Factibilidad para barrer la vivienda
Factibilidad para caminar 5 kilómetros
Si ha sufrido de alguna enfermedad o lesión en las últimas 2 semanas
Que tanto fuma la beneficiada

Que tanto alcohol toma la beneficiada
Si alguna vez ha estado embarazada la beneficiada
Si alguna vez ha recibido alguna propuesta de matrimonio
Si tiene deseos de casarse en algún punto de su vida
Grado que la beneficiada desea estudiar antes de casarse
Grado que la beneficiada desea que su esposo tenga antes de casarse
Si espera que su familia reciba dinero o regalos por parte de su esposo
Si espera estar casada en 3 años
Edad a la que le gustaría casarse
Si espera tener bienes comunes con su futuro esposo
Si le molestaría que su esposo tuviera otra esposa
Si le preocupa que su esposo tenga otra mujer que no sea ella
Si le gustaría tener hijos en el futuro
Cuanto tiempo pretende esperar para tener hijos después de casarse
Si cree que es apropiado que una esposa deje a su esposo si este no provee financieramente a ella y sus hijos
Si cree que es apropiado que una esposa deje a su esposo si este la golpea frecuentemente
Si cree que es apropiado que una esposa deje a su esposo si este no es fértil
Si cree que es apropiado que una esposa deje a su esposo si este está infectado de VIH
Si cree que es apropiado que una esposa deje a su esposo si este le prohíbe tener planeación familiar
Si cree que es apropiado que una esposa deje a su esposo si este no sigue una religión
Si cree que es apropiado que una esposa se rehusé a tener sexo sin protección si cree que ella tiene VIH
Si cree que es apropiado que una esposa se rehusé a tener sexo sin protección si no quiere embarazarse
Si cree que es apropiado que un esposo no coma su comida si se le negó sexo
Si cree que es apropiado que un esposo tenga otra pareja si se le negó sexo
Si cree que es apropiado que un esposo la obligue a tener sexo si se le negó sexo
Si cree que es apropiado que un esposo no provea el hogar si se le negó sexo
Propuestas de amoríos recibidas en los últimos 3 meses
Si ha aceptado alguna propuesta de amorío en lo últimos 3 meses
Edad en comparación de la última persona que le propuso un amorío
Número de personas con las cuales ha tenido sexo
Si alguna vez ha tenido sexo con algún novio
Si su mejor amiga ha tenido sexo en el último año
Número de personas con las que ha tenido sexo su mejor amiga
Preocupación de la beneficiada por contraer SIDA
Tipo de persona de la cual estaría más preocupada por infectarse
Número de personas de las que sospecha murieron por SIDA
Número de personas que sospecha murieron por SIDA en el último año
Número de personas que sospecha tienen SIDA
Probabilidad de que ella tenga SIDA
Probabilidad de que ella se contagie de SIDA
Porcentaje de gente que cree que tienen SIDA
Porcentaje de gente que cree tendrán SIDA en 5 años
Probabilidad que cree que se contagiaría de SIDA si tiene sexo con una persona infectada solo una vez

Mejor forma de protegerse de contraer SIDA según la beneficiada
Si alguna vez ha escuchado sobre pruebas voluntarias de SIDA
Si alguna vez ha escuchado de la medicina que se le da a los infectados en el hospital
Si cree que esta medicina puede curar el SIDA
Si conoce a alguna persona que actualmente este siendo tratada
Si alguna vez se ha realizado una prueba de SIDA
Número de personas que conoce que se han realizado esta prueba
Número de estas personas que le han dicho sus resultados
Si su pareja regular se ha realizado una prueba de SIDA
Si cree que una mujer embarazada le puede transmitir SIDA a su hijo
Si cree que una mujer le puede transmitir SIDA a su hija por lactancia
Si cree que puede enfermarse de SIDA si tiene sexo con alguien que luce sano
Si cree que el SIDA tiene cura
Si está de acuerdo con que la gente de su villa se siente cómoda con gente que tiene SIDA
Si le compraría verduras a un vendedor que tiene SIDA
Número de veces a las que ha asistido a un funeral en los últimos 30 días
Número de veces a las que ha asistido a una presentación de drama en los últimos 30 días
Número de veces a las que ha asistido a un bar en los últimos 30 días
Número de veces a las que ha asistido a bailar en los últimos 30 días
Número de veces a las que ha asistido a un mercado en los últimos 30 días
Número de veces a las que ha asistido a una boda en el último año
Número de veces a las que ha asistido a una presentación sobre planeación familiar en el último año
Número de veces a las que ha asistido a un evento político en el último año
Si pertenece a algún grupo de agricultores
Si pertenece a algún grupo de salud
Si pertenece a algún grupo de SIDA
Si pertenece a algún grupo financiero
Si existe alguna organización comunitaria en se villa que asista a orfanatos
Si existe alguna organización comunitaria en se villa que asista a personas que están enfermas en casa
Si existe alguna organización comunitaria en se villa que asista con desarrollo económico
Número de veces que ha ido a comer o beber con sus amigos en el último mes
Número de llamadas recibidas en el último mes
Forma en la que conseguiría kwachas para viajar
Que tanto confía en la gente de su comunidad
Sensación de seguridad ante el crimen en el hogar
Sensación de seguridad ante el crimen cuando sale
Sensación acerca de cómo reaccionaría la demás gente de su comunidad si ella necesita ayuda
Si siempre debe estar alerta de la gente de la villa
Si en el último mes compró o alguien le compro jabón con esencia
Si en el último mes compró o alguien le compro un celular
Si en el último mes compró o alguien le compro tiempo aire
Si en el último mes compró o alguien le compro zapatos de mujer
Si en el último mes compró o alguien le compro cremas para la piel
Si en el último mes compró o alguien le compro maquillaje
Si en el último mes compró o alguien le compro refrescos o botanas
Si en el último mes compró o alguien le compro alcohol
Si en el último mes compró o alguien le compro condones

Si en el último mes compró o alguien le compro transporte aparte del de la escuela

Si en el último mes alguien le dio dinero

Habilidad de la beneficiaria al responder en Chichewa según el entrevistador

Si hubo alguna persona presente durante la entrevista que comprometiera la confidencialidad

Confianza de la veracidad de la información por parte del entrevistador

VIII. Referencias

- Adelman, Melissa A., and Peter A. Holland.** 2015. "Increasing Access by Waiving Tuition: Evidence from Haiti" *World Bank Policy Research Working Paper* (7175).
- Baird, Sarah, Ephraim Chirwa, Craig McIntosh, and Berk Ozler.** 2015. "What Happens Once the Intervention Ends? the Medium-Term Impacts of a Cash Transfer Programme in Malawi" *3ie Grantee Final Report*.
- Bia, Michela, Alfonso Flores-Lagunes, and Alessandra Mattei.** 2013. *A Stata package for the application of semiparametric estimators of dose-response functions*.
- Blimpo, Moussa, David Evans, and Nathalie Lahire.** 2015. "Parental Human Capital and Effective School Management: Evidence from the Gambia" *World Bank Policy Research Working Paper* (7238).
- Carneiro, Pedro, Oswald Koussihouèdé, Nathalie Lahire, Costas Meghir, and Corina Mommaerts.** 2015. *Decentralizing education resources: school grants in Senegal*.
- Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman.** 2011. *School inputs, household substitution, and test scores*.
- Fan, Jianqing, Irène Gijbels, Tien-Chung Hu, and Li-Shan Huang.** 1996. "A Study of Variable Bandwidth Selection for Local Polynomial Regression" *Statistica Sinica*: 113-127.
- Flores, Carlos A.** 2005. "Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment". *Unpublished doctoral dissertation, University of California, Berkeley*.
- Flores, Carlos A., Alfonso Flores-Lagunes, Arturo Gonzalez, and Todd C. Neumann.** 2012. "Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps" *Review of Economics and Statistics*, 94(1): 153-171.
- Gertler, Paul J., Harry A. Patrinos, and Marta Rubio-Codina.** 2012. "Empowering Parents to Improve Education: Evidence from Rural Mexico" *Journal of Development Economics*, 99(1): 68-79.
- Gertler, Paul J., Sebastian W. Martinez, and Marta Rubio-Codina.** 2012. "Investing Cash Transfers to Raise Long-Term Living Standards" *American Economic Journal: Applied Economics*: 164-192.
- Hardin, James W., Joseph M. Hilbe, and Joseph Hilbe.** 2007. *Generalized Linear Models and Extensions*: Stata Press.
- Hirano, Keisuke, and Guido W. Imbens.** 2004. "The Propensity Score with Continuous Treatments" *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164: 73-84.
- Kremer, M., C. Brannen, and R. Glennerster.** 2013. "The Challenge of Education and Learning in the Developing World" *Science (New York, N.Y.)*, 340(6130): 297-300.
- McCullagh, Peter, and John A. Nelder.** 1989. *Generalized Linear Models*. Vol. 37: CRC press.
- McEwan, Patrick J.** 2015. "Improving Learning in Primary Schools of Developing Countries A Meta-Analysis of Randomized Experiments" *Review of Educational Research*, 85(3): 353-394.

IX. Índice de tablas

Tabla 1. Línea de tiempo de las rondas e intervención.....	6
Tabla 2. Estadísticas Descriptivas de la muestra objetivo en la línea basal	8
Tabla 3. Estadísticas Descriptivas en la línea basal de la muestra balanceada que cumple con la condición de soporte común	9
Tabla 4. Distribución Condicional del Tratamiento	18
Tabla 5. Pruebas de Balanceo	19
Tabla 6. Variables utilizadas.....	27

X. Índice de gráficas

Gráfica 1. Distribución no condicional del tratamiento de la muestra balanceada que cumple con la condición de soporte común	16
Gráfica 2. Retención primer año	21
Gráfica 3. Retención segundo año	21
Gráfica 4. Intervalo de confianza para la retención del primer año	22
Gráfica 5. Intervalo de confianza para la retención del segundo año	22
Gráfica 6. Retención tercer año	23
Gráfica 7. Intervalo de confianza para la retención del tercer año	23
Gráfica 8. Progresión primer año	23
Gráfica 9. Intervalo de confianza para la progresión del primer año	23
Gráfica10. Progresión segundo año	24
Gráfica 11. Intervalo de confianza para la progresión del segundo año	24
Gráfica 12. Progresión tercer año	25
Gráfica 13. Intervalo de confianza para la progresión del tercer año	25