

# El impacto de subsidios al uso del transporte público sobre el desempeño de los individuos en el mercado laboral: el caso de Bogotá\*

Juan David Hernández Leal<sup>‡</sup>

## Resumen

En Bogotá, los hogares más pobres se encuentran aislados espacialmente de las oportunidades laborales. En este artículo, se evalúan los efectos de una política de subsidios al uso del Sistema Integrado de Transporte Público (SITP) sobre el desempeño de los individuos en el mercado laboral. Para realizar el análisis, se utilizan datos de la Encuesta de Movilidad para el año 2015 y de la Encuesta Multipropósito del año 2017, con los cuales se implementa un diseño de Regresión Discontinua Borrosa (RDB). Se encuentra que los hogares elegibles para recibir el subsidio tienen una probabilidad de obtenerlo 30 puntos porcentuales mayor que los no elegibles. Sin embargo, y en línea con estudios previos, no hay evidencia estadística de que el programa tenga algún efecto significativo sobre la probabilidad de participar en el mercado laboral, de estar ocupado o de estar desempleado.

**Palabras clave:** Desajuste espacial, Accesibilidad a oportunidades laborales, Oferta laboral, Subsidios al transporte público.

**Clasificación JEL:** J22, J64, R4.

---

\*Trabajo presentado para optar al título de Magíster en Economía de la Pontificia Universidad Javeriana.

<sup>†</sup>Pontificia Universidad Javeriana, email: j-hernandezl@javeriana.edu.co

<sup>‡</sup>Agradezco la dirección de la profesora Ana María Díaz Escobar, quien me guió y acompañó durante todo el proceso de realización de este documento, lo cual fue fundamental para poder llevarlo a buen término.

# 1 Introducción

La estructura espacial de las ciudades puede afectar a sus mercados laborales. Por ejemplo, en Bogotá, las personas se encuentran aisladas geográficamente de las oportunidades laborales (Díaz Escobar, 2015; Guzman y Oviedo, 2018). En la Figura 1, se puede apreciar que la densidad poblacional es mayor en las zonas alejadas del centro de negocios. A este fenómeno se le conoce en la literatura de economía urbana como desajuste espacial o “spatial mismatch” (Kain, 1968). Como consecuencia de vivir en barrios con menor acceso a fuentes de empleo, los individuos pueden tener una desventaja que se ve reflejada en una menor tasa de participación y ocupación, así como una mayor probabilidad de estar desempleados.

Figura 1: Densidad poblacional vs. Densidad de empleos en Bogotá



Fuente: Guzman y Oviedo (2018).

Ahora bien, estudiar la relación causal entre la accesibilidad a empleos y el desempeño de un individuo en el mercado laboral implica varios retos empíricos, debido al sesgo de selección que surge a raíz de dos fuentes: la primera radica en que la elección del lugar de residencia no se realiza de forma aleatoria. Esto hace potencialmente diferentes a las personas que viven en la periferia de quienes viven más cerca del centro de negocios. Por ejemplo, si los individuos que viven en la periferia tienen, en promedio, un nivel educativo menor a los que residen más cerca del centro de negocios, se puede confundir el efecto de

la desconexión espacial con el de no estar lo suficientemente calificado para conseguir un empleo. La segunda fuente se origina en la causalidad inversa entre el lugar de residencia y el desempeño en el mercado laboral. Verbigracia, una persona puede decidir mudarse al conseguir empleo con el fin de reducir el tiempo que gasta a diario desplazándose hacia su lugar de trabajo. Recíprocamente, el hecho de que viva en un barrio en la periferia puede afectar sus probabilidades de obtener un trabajo.

A finales del 2013, la Administración Distrital, motivada por el bajo número de viajes motorizados realizados por los habitantes más pobres de Bogotá, y con el objetivo de incentivar el uso del SITP, decidió otorgar un beneficio tarifario para los miembros de los hogares pertenecientes a los niveles 1 y 2 del Sistema de Selección de Beneficiarios Para Programas Sociales (SISBÉN); es decir, aquellos con puntajes de entre 0 y 40 puntos<sup>1</sup> (según la metodología SISBÉN III calculada por la Secretaría Distrital de Planeación). El Decreto 603 de 2013 estableció que se otorgaría a los beneficiarios un descuento del 40% sobre el valor del pasaje, hasta por un máximo de 21 viajes mensuales. El beneficio aplicaba solo para los individuos mayores de 16 años que no fueran ya beneficiarios de los otros subsidios otorgados por el SITP, como el de Adulto Mayor y el de Discapacitados. Además, los interesados debían solicitar la tarjeta inteligente de SISBÉN/SITP (a través de la cual los usuarios pagan el pasaje de los buses troncales (Transmilenio) y zonales (SITP urbano)) por alguno de los medios dispuestos por Transmilenio S.A. (Alcaldía Mayor de Bogotá, 2017a)<sup>2</sup>. La resolución 51 de 2014, expedida en febrero de ese año, establece los procedimientos para la implementación del beneficio, por lo que marca la fecha desde la cual este programa se encuentra vigente.

En el presente documento se evalúa si estos subsidios al transporte público tuvieron un efecto sobre el desempeño en el mercado laboral de sus beneficiarios. La hipótesis principal es que dichos subsidios, por una parte, contribuyen a reducir los costos de búsqueda de empleo, reduciendo el salario de reserva de las personas (Abebe et al., 2016); y por otro lado, mejoran la accesibilidad de los beneficiarios a oportunidades laborales de las que están

---

<sup>1</sup>“El SISBÉN, a través de un puntaje, clasifica a la población de acuerdo con sus condiciones socioeconómicas, y se utiliza para identificar de manera rápida y objetiva a la población en situación de pobreza y vulnerabilidad, con el objetivo de focalizar la inversión social y garantizar que esta sea asignada a quienes más lo necesitan. El puntaje se calcula a partir de la información reportada por el hogar en la encuesta y es un valor entre cero (0) y cien (100).

<sup>2</sup>De igual manera, se decretó que se perdería el subsidio cuando al actualizar la base del SISBÉN, el puntaje del hogar superara los 40 puntos; cuando el beneficiario figurara como dueño de un vehículo automotor; o cuando el dueño de la tarjeta le diera un uso indebido, como comercializarla o permitir que otras personas la utilizaran. Posteriormente, el decreto 442 de 2014 amplió el beneficio a 50% sobre el valor del pasaje y hasta por un máximo de 40 viajes al mes (Alcaldía Mayor de Bogotá, 2017a).

desconectados espacialmente. Así pues, se espera que el programa tenga un efecto positivo sobre la probabilidad de participar en el mercado laboral y que disminuya la probabilidad de estar desempleado.

Para probar la hipótesis, se aprovecha que, a pesar de que los potenciales beneficiarios tenían que decidir si tomaban o no el subsidio, la asignación al tratamiento se determinó por medio del puntaje de SISBÉN del hogar. Lo anterior permite implementar un diseño de Regresión Discontinua Borrosa (RDB), el cual estima el efecto del tratamiento solo para hogares con puntajes cercanos al umbral de 40, lo que contribuye a reducir o eliminar el sesgo que potencialmente afecta las estimaciones del impacto del tratamiento.

Varios autores han estudiado la implementación del SITP y del programa de subsidios “pro-pobres” en Bogotá. En especial, se destacan las investigaciones de Rodríguez et al. (2015) y Hernández y Peralta-Quiros (2016), quienes emplean un modelo probit para identificar los determinantes de la autoselección a recibir el subsidio y, de forma independiente, una metodología de diferencias en diferencias para evaluar sus efectos sobre el uso del sistema y variables del mercado laboral. Con este fin, combinan información de las encuestas del SISBÉN con datos de Transmilenio, Recaudo Bogotá y con la Gran Encuesta Integrada de Hogares (GEIH). Esto les permite calcular el efecto promedio del subsidio en las manzanas que tienen una mayor proporción de tarjetas subsidiadas y viajes subsidiados sobre el total. Las conclusiones de dicho ejercicio sugieren que en las manzanas “tratadas” los ingresos de los trabajadores informales y el uso del sistema de transporte es mayor, pero no hay ningún efecto del subsidio sobre la probabilidad de participar en el mercado laboral ni de estar desempleado.

Se considera que este trabajo puede complementar y corroborar los hallazgos previos. En particular, las contribuciones principales son: (i) Probar efectos heterogéneos del subsidio por género, es decir, evaluar si el programa afecta de forma diferencial a hombres y mujeres. Lo anterior es relevante porque se ha documentado que la brecha de género, en términos de participación laboral, incrementa a medida que se consideran localidades más alejadas del centro de negocios (Díaz Escobar, 2015). Además, Rodríguez et al. (2015) encuentran que ser mujer incrementa la probabilidad de obtener el subsidio en un 10%; y (ii) Aprovechar la discontinuidad en la asignación al tratamiento generada por el instrumento de focalización, utilizado para determinar a los potenciales beneficiarios del subsidio, para implementar una estrategia de identificación que permita darle una interpretación causal a los estimadores. Esto es factible gracias a la información de las bases de datos empleadas, que no han sido utilizadas para evaluar la política en cuestión y cuentan con información de las variables de asignación, tratamiento y resultado, de forma que son idóneas para examinar la hipótesis

planteada.

Los resultados centrales sugieren que los hogares elegibles tienen una probabilidad de que algún miembro tenga la tarjeta de SITP/SISBÉN aproximadamente 30 puntos porcentuales mayor que los no elegibles. No obstante, esto no se ve reflejado en un incremento en la tasa de participación o la tasa de ocupación de los miembros de los hogares beneficiarios, ni tampoco en una disminución de la probabilidad de estar desempleado. Lo anterior ocurre tanto para hombres como para mujeres, y se mantiene cuando se realizan una serie de pruebas de falsificación y robustez.

Para explicar los resultados, se plantea que la competencia que enfrenta el SITP con los buses del SITP provisional puede contribuir a que los beneficiarios no tengan suficientes incentivos para utilizar más los buses del sistema.

Este documento está organizado de la siguiente manera: la sección 2 revisa la literatura sobre desajuste espacial y los efectos de programas que buscan mejorar la accesibilidad a empleos. Las secciones 3 y 4 describen los datos y la metodología empleados para el análisis empírico. La sección 5 expone y discute los resultados principales, así como las pruebas de robustez y falsificación, y plantea posibles campos de acción para investigaciones futuras. Finalmente, la sección 6 concluye y expone las limitaciones e implicaciones de política pública de la investigación.

## **2 Revisión de literatura: la hipótesis del desajuste espacial y la relación entre la accesibilidad y el mercado laboral**

Esta sección se divide en dos partes: en la primera, se reseña brevemente la evolución de la hipótesis del desajuste espacial, los artículos que estudian la relación entre la accesibilidad y el desempeño de los individuos en el mercado laboral, y las investigaciones que evalúan los efectos de la implementación del Sistema Integrado de Transporte Público y los subsidios a su utilización. En la segunda, se detalla cómo el presente documento se diferencia de los estudios previos que se han esforzado por resolver preguntas similares.

La literatura de economía urbana y laboral ha estudiado de manera amplia los fenómenos del desajuste espacial, la segregación residencial e hipótesis relacionadas, así como la conexión entre la accesibilidad a oportunidades laborales, la intensidad de búsqueda de empleo y los resultados individuales en el mercado laboral. Los orígenes de la hipótesis se remontan al

final de la década de los 60s, cuando Kain (1968) planteó que la segregación residencial de los afroamericanos afectaba sus probabilidades de conseguir un empleo y sus salarios. En el curso de las siguientes 3 décadas, los investigadores se dieron la tarea de probar la existencia de una relación causal entre la desconexión espacial de las oportunidades laborales y los resultados en el mercado laboral de las minorías, encontrando evidencia a favor de la hipótesis. Empero, no se ha llegado a un consenso en cuanto a la magnitud o importancia relativa del desajuste espacial como una explicación de las disparidades raciales en el desempleo (Ihlanfeldt, 2006; Gobillon et al., 2007).

Según Díaz Escobar (2015): “hay varios mecanismos a través de los cuales la distancia a las fuentes de trabajo pueden afectar los resultados laborales: el salario de reserva, los costos de búsqueda, la eficiencia de búsqueda por fricciones de información, los pocos incentivos que tienen los trabajadores que viven lejos del centro de negocios para buscar empleo (Smith y Zenou, 2003), y los efectos negativos que puede tener la distancia sobre la productividad. Dichos mecanismos operan de la siguiente manera: (i) La distancia aumenta los costos de desplazamiento para los individuos en situación de desempleo, lo que genera un incremento en el salario de reserva (Brueckner y Zenou, 2003); (ii) Cuando los costos de búsqueda son altos, quienes buscan empleo pueden restringirse a lugares cercanos a su lugar de residencia (Gobillon et al., 2007); y (iii) Los trabajadores que viven lejos de los centros de empleo tienen menos información sobre vacantes que los que viven cerca (Gobillon et al., 2007)”.

Una cantidad considerable de estudios empíricos se han propuesto probar la relación causal entre accesibilidad y resultados en el mercado laboral. Una estrategia empleada por otros autores para solventar los retos empíricos inherentes a este problema, es restringir la muestra utilizada y utilizar variables instrumentales. Entre las investigaciones internacionales, se destacan los trabajos de Matas et al. (2010) y Jin y Paulsen (2018), quienes tratan el tema en las ciudades Barcelona y Madrid, y Chicago; respectivamente. Para el caso colombiano, se destacan los trabajos de Díaz Escobar (2015), Domínguez (2014) y Jaramillo y Rengifo (2018), que utilizan datos de las ciudades de Bogotá y Medellín. En general, los resultados sugieren que una mayor accesibilidad<sup>3</sup> a oportunidades laborales tiene un efecto positivo sobre el desempeño de las personas en el mercado laboral.

En relación a las políticas que buscan mejorar dicha accesibilidad, la teoría tradicional de búsqueda de empleo predice que cualquier disminución en los costos de búsqueda, como por

---

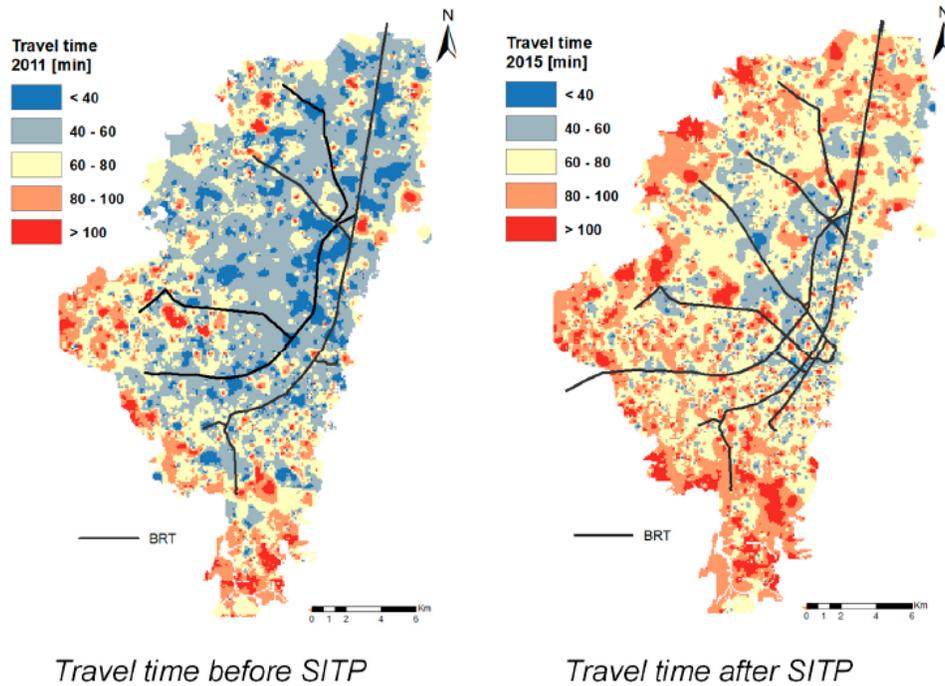
<sup>3</sup>Los artículos académicos suelen calcular la siguiente medida de accesibilidad:  $Acceso_i = \sum_j \frac{Empleo_j}{Dist_{ij}}$ , donde el numerador es el número de trabajos en la unidad geográfica  $j$  (UPZ, localidad) y el denominador es la distancia entre el área de residencia del individuo  $i$  y el lugar de destino  $j$ . En este contexto, la distancia es medida como el tiempo de viaje promedio o la métrica euclidiana (Díaz Escobar, 2015).

ejemplo un subsidio a la demanda de transporte público, debería incentivar a los individuos a buscar empleo con mayor intensidad (Pissarides, 2000). Para probarlo, algunos estudios experimentales han evaluado el efecto de reducir los costos de transporte sobre el desempeño en el mercado laboral, en ciudades es desarrollados y de países en vías de desarrollo (Phillips (2014), Abebe et al. (2016) y Franklin et al. (2015)). Las intervenciones asignan subsidios de transporte aleatoriamente y concluyen que estas intervenciones incrementan la intensidad de búsqueda de empleo de los individuos, reducen la duración del desempleo, y mejoran las probabilidades de tener un empleo permanente y formal.

Otro patrón que caracteriza a los estudios empíricos citados es que el efecto encontrado es de mayor magnitud para los grupos más desfavorecidos: las personas que viven lejos de los centros de negocios, las mujeres y los individuos con menor nivel educativo, son quienes más se benefician de una mayor accesibilidad (Franklin et al., 2015; Domínguez, 2014).

A pesar de que no se cuenta con evidencia experimental para Colombia, una corriente alternativa de estudios implementa simulaciones y plantea modelos teóricos para evaluar los efectos de la puesta en marcha del SITP y de los subsidios “pro-pobres” sobre la accesibilidad a oportunidades de generación de ingreso y desigualdad. Los resultados indican que el SITP es una política de transporte con el potencial de ayudar a los individuos de bajos ingresos. Se halla que tanto la estructura actual de los subsidios, así como escenarios alternativos para incrementar su cobertura, mejoran la accesibilidad y la equidad para los beneficiarios (Bocarejo et al., 2016; Guzman y Oviedo, 2018). No obstante, Guzman et al. (2018) sugiere que se han cometido errores durante la implementación del SITP que han contribuido a que la accesibilidad de las personas que viven en la periferia se haya empeorado entre 2011 y 2015 (Ver Figura 2).

Figura 2: Curvas isócronas de tiempo de viaje en transporte público antes del SITP en 2011, y después del SITP en el 2015.



Fuente: Guzman et al. (2018).

Ahora bien, es importante destacar que los documentos más parecidos al presente son los de Rodríguez et al. (2015) y Hernández y Peralta-Quiros (2016), quienes utilizan una metodología cuasiexperimental (diferencias-en-diferencias) para probar el efecto del subsidio al transporte público que se implementó en la ciudad de Bogotá. Los autores encuentran que los beneficiarios realizan un mayor número de viajes al mes, es decir que usan más el sistema de transporte público. Adicionalmente, se concluye que el subsidio únicamente tiene un efecto significativo positivo sobre el ingreso por hora de los trabajadores informales, mas no afecta a otros indicadores del mercado laboral, como la participación, la probabilidad de estar desempleado, la probabilidad de tener un trabajo formal o el número de horas trabajadas a la semana.

Dicho ejercicio empírico considera a toda la muestra de la Gran Encuesta Integrada de Hogares para evaluar los efectos causales del subsidio. A manera de contraste, la presente investigación solo considera a los ciudadanos de estratos 1 y 2 e ingresos menores a un salario mínimo mensual, que componen el subconjunto de la población con menor acceso al sistema. Este segmento de la población coincide en buena parte con aquellos que tienen un puntaje del SISBÉN cercano al corte del programa.

Además, se aprovecha la discontinuidad alrededor del umbral de 40 puntos, generada por el mecanismo de asignación del subsidio, para implementar un diseño de RDB, estrategia que no ha sido explotada anteriormente. Este método cuenta con supuestos de identificación que son fáciles de probar con los datos disponibles, y también compara individuos que tienen puntajes de SISBÉN similares, lo cual es útil para reducir el sesgo asociado a la estimación del efecto del programa.

Esto último es posible porque las fuentes de datos empleadas son distintas a las utilizadas por Rodríguez et al. (2015) y Hernández y Peralta-Quiros (2016). Además de contar con variables relacionadas con el desempeño de los individuos en el mercado laboral, características de los individuos y los hogares, las Encuestas de Movilidad de Bogotá del 2015 y la Multipropósito de Bogotá del 2017 permiten observar, para cada hogar incluido, si alguno de sus miembros tiene el subsidio, así como si el hogar cumple con los criterios de elegibilidad para recibirlo. En adición, el nivel de desagregación permite incorporar el componente de género a la investigación, pues la unidad de análisis pasa de ser la manzana al individuo.

### 3 Datos y estadísticas descriptivas

Para probar la hipótesis principal, primero se emplean datos de la Encuesta de Movilidad de Bogotá para el año 2015. Este es un corte transversal recolectado por la Secretaría Distrital de Movilidad, y fue realizada con el objetivo de estudiar los patrones de viaje de la población. El levantamiento de los datos se llevó a cabo entre el 15 de marzo y el 30 de agosto de 2015 y se realizaron 28.212 encuestas en hogares<sup>4</sup> (Alcaldía Mayor de Bogotá, 2017b).

La unidad de observación es la población mayor a 5 años de edad, residente de Bogotá y los municipios vecinos de su área de influencia. La encuesta tiene datos de la participación de los individuos en el mercado laboral y de otras variables que son consideradas por la literatura como determinantes importantes de la oferta laboral. Sin embargo, no cuenta con información de los ingresos no laborales de los individuos ni de su estado civil. Idealmente, se trabajaría con una base que también tuviera esta información y que permitiera seguir a los individuos en el tiempo, con el fin de observarlos antes y después de la implementación del subsidio. Pero no existe una fuente de información de este tipo para Colombia. Añadido a lo anterior, otra limitación de estas bases de datos es que, aunque se puede observar qué

---

<sup>4</sup>Según el DANE, el hogar censal es “una persona o grupo de personas, parientes o no, que viven (duermen) en una misma vivienda y comparten, generalmente, sus comidas” (Departamento Administrativo Nacional de Estadística - DANE, 2007).

hogares tienen al menos una tarjeta, no es posible saber qué miembros en específico del hogar tienen el beneficio para hogares de más de una persona.

Es importante notar que la encuesta le pregunta a los hogares por su puntaje de SISBÉN y si hay al menos un miembro dentro del hogar que es beneficiario del subsidio (es decir, que tiene la tarjeta SITP/SISBÉN), e indaga por la actividad principal a la que se dedica cada integrante. Por lo que cuenta con la información necesaria para estimar el efecto de los subsidios sobre los indicadores de mercado laboral de los individuos. Esto se debe a que el mecanismo de focalización para asignar los subsidios fue el puntaje de SISBÉN, de manera que se puede observar tanto qué hogares son elegibles para recibir el subsidio, cuáles efectivamente lo utilizan y cómo es el desempeño de sus miembros en el mercado laboral. Es importante señalar que la encuesta pregunta cuántos viajes realizó la persona el día anterior, los motivos del viaje y el medio de transporte en el que se movilizó.

La Tabla 1 muestra estadísticas descriptivas de la muestra que representa la población objetivo del subsidio, que corresponde a las personas residentes de Bogotá que habitan en hogares de Estratos 1 y 2, y cuyos ingresos familiares no suman más de un salario mínimo<sup>5</sup>. Cabe resaltar que el 84% de la muestra que pertenece a esta población completó a lo sumo la Secundaria, el 55% son mujeres y el 63% participa en el mercado laboral. Adicionalmente, la tasa de ocupación es del 57% y la de desempleo del 10%.

De manera similar, la Tabla 2 evalúa si hay diferencias de medias en las variables incluidas en la Tabla 1 entre beneficiarios y no beneficiarios del subsidio. Esta información es útil para evaluar de forma preliminar si los indicadores de mercado laboral de los beneficiarios del programa, así como otros determinantes tradicionales de la oferta laboral, difieren entre estos grupos. Los resultados muestran que no hay diferencias significativas entre tratados y controles en los 3 indicadores de mercado laboral considerados. Aunque, sí se encuentran diferencias en términos de nivel educativo, estrato y puntaje de SISBÉN del hogar, que son consistentes con la focalización que se hizo del subsidio a los grupos más desfavorecidos de la población urbana. Las diferencias de medias igualmente apoyan las sospechas acerca del sesgo de selección y la necesidad de utilizar una estrategia de identificación que permita reducirlo o eliminarlo.

---

<sup>5</sup>En el 2015, el salario mínimo en Colombia era de 644.350 COP.

Tabla 1: Estadísticas descriptivas - Encuesta de Movilidad

	Observaciones	Media	Desviación estándar
Mujer	3,276	0.55	0.50
Edad	3,276	36.92	14.20
Participa en el mercado laboral	3,276	0.63	0.48
Desempleo	2,071	0.10	0.31
Ocupado	3,276	0.57	0.50
Algún miembro del hogar tiene tarjeta SITP-SISBÉN	3,276	0.16	0.36
Puntaje SISBÉN del hogar	3,276	36.70	19.21
Número de personas en el hogar	3,276	3.96	1.45
Dependientes	3,276	0.53	0.50
Primaria	3,276	0.26	0.44
Secundaria	3,276	0.58	0.49
Técnico	3,276	0.10	0.30
Universitario	3,276	0.05	0.22
Posgrado	3,276	0.01	0.09
Estrato 1	3,276	0.29	0.46
Estrato 2	3,276	0.71	0.46

Elaboración propia con base en la Encuesta de Movilidad del 2015.

Tabla 2: Diferencias de medias Beneficiarios (1) vs No Beneficiarios (0)

	1	0	Dif (1-0)	P	N0	N1
Mujer	0.58	0.55	0.03	0.16	2,767	509
Edad	35.08	37.26	-2.18***	0.00	2,767	509
Participa en el mercado laboral	0.63	0.63	0.00	0.98	2,767	509
Desempleo	0.10	0.11	-0.01	0.62	1,749	322
Ocupado	0.57	0.57	0.01	0.80	2,767	509
Puntaje SISBÉN del hogar	26.58	38.56	-11.9***	0.00	2,767	509
Número de personas en el hogar	4.03	3.94	0.08	0.24	2,767	509
Dependientes	0.51	0.53	-0.02	0.32	2,767	509
Primaria	0.29	0.25	0.04*	0.06	2,767	509
Secundaria	0.57	0.59	-0.02	0.37	2,767	509
Técnico	0.10	0.10	0.00	0.80	2,767	509
Universitario	0.04	0.05	-0.02*	0.06	2,767	509
Posgrado	0.00	0.01	-0.01**	0.02	2,767	509
Estrato 1	0.34	0.29	0.05**	0.03	2,767	509
Estrato 2	0.66	0.71	-0.05**	0.03	2,767	509

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Se realiza una prueba de diferencias de medias asumiendo varianzas distintas y desconocidas.

A manera de comparación, la Tabla 3 muestra estadísticas descriptivas de la Gran Encuesta Integrada de Hogares (GEIH)<sup>6</sup> para el año 2015, que es el mismo en el que se recolectó la información de la Encuesta de Movilidad. Cabe resaltar que las muestras de ambas encuestas son distintas. Por ejemplo, los indicadores de mercado laboral son peores (menores tasas de participación y ocupación, junto con una mayor tasa de desempleo) para el caso de la Gran Encuesta. Esto puede estar relacionado con el hecho de que la proporción de hogares de Estrato 1 sea mayor en dicha muestra. La población de la GEIH también es menor (en promedio) que la de la Encuesta de Movilidad.

<sup>6</sup>La GEIH recolecta información de las condiciones de empleo de las personas, así como características generales de la población. Es la encuesta más utilizada para seguir la evolución de los indicadores del mercado laboral colombiano en el tiempo. Lastimosamente, no cuenta con el puntaje SISBÉN de los hogares ni tiene información sobre su uso de la tarjeta de SISBÉN-SITP.

Tabla 3: Estadísticas descriptivas - Gran Encuesta Integrada de Hogares

	Observaciones	Media	Desviación estándar
Mujer	646	0.59	0.49
Edad	646	31.83	11.91
Participa	646	0.51	0.50
Desocupado	326	0.33	0.47
Ocupado	646	0.34	0.47
Número de personas en el hogar	646	3.13	1.42
Dependientes	646	0.46	0.50
Ninguno	646	0.08	0.27
Bachiller	646	0.73	0.45
Técnico o tecnológico	646	0.16	0.36
Universitario	646	0.03	0.17
Posgrado	646	0.00	0.06
Estrato 1	646	0.24	0.43
Estrato 2	646	0.76	0.43
Ingresos no laborales	646	0.05	0.14
Casado/vive en pareja	646	0.35	0.48

Elaboración propia con base en la Gran Encuesta Integrada de Hogares, para el año 2015. La variable de ingresos no laborales está medida en millones de pesos.

En esa misma línea, se hizo uso de los datos de la Encuesta Multipropósito de Bogotá de 2017, que se aprovecha para evaluar los efectos del subsidio algunos años más tarde<sup>7</sup>, en un período caracterizado por el endurecimiento de las condiciones necesarias para poder adquirir el subsidio, así como una disminución de los beneficios tarifarios recibidos por los tarjetahabientes, como fue discutido en la Sección 1. La Figura 3 muestra los requisitos para adquirir el beneficio luego de la entrada en vigencia del Decreto 131 de 2017<sup>8</sup>.

<sup>7</sup>Es importante notar que no es posible seguir a los mismos hogares de la Encuesta de Movilidad de 2015 en la Encuesta Multipropósito de 2017.

<sup>8</sup>Mediante este decreto, La Alcaldía de Bogotá redujo el descuento sobre el valor total del pasaje de 50% a 25%, tanto para el componente troncal (Transmilenio), como para el componente zonal (SITP urbano) del sistema, y se disminuyó el máximo de viajes mensuales con descuento de 40 a 30. Por otra parte, se decidió que los beneficiarios ahora serían solo los hogares que tuvieran un puntaje de SISBÉN de entre 0 y 30,56 puntos.

Figura 3: Requisitos para obtener la tarjeta SITP/SISBÉN



Fuente: Transmilenio S.A. (2018).

Adicionalmente, la encuesta dispone de información del estado civil e ingresos no laborales de los individuos, y es posible evidenciar no solo en qué hogares hay personas con tarjetas, sino cuántos miembros del hogar tienen una tarjeta, a diferencia de la Encuesta de Movilidad del 2015. Un aspecto para tener en cuenta de la Encuesta Multipropósito es que el puntaje de SISBÉN con el que cuenta es el calculado directamente por el Departamento Nacional de Planeación (DNP), de manera que puede diferir del que reportan los hogares a los encuestadores en la Encuesta de Movilidad. De tal manera, no está sujeto a los errores que pueden realizar los hogares, ni a una eventual manipulación del puntaje por parte de los mismos.

Otra diferencia entre las dos encuestas es que la Multipropósito no es representativa a nivel de localidad, sino a nivel de Unidad de Planeamiento Zonal (UPZ), que es una unidad geográfica más desagregada que la primera (La UPZ puede considerarse un punto medio entre una localidad y un barrio). En adición, el objetivo de la Multipropósito no era caracterizar los patrones de movilidad de la población, sino cubrir aspectos de cobertura, calidad y gasto de los hogares en servicios públicos domiciliarios, mercado laboral y condiciones de vida a nivel de ciudad (Dirección de Metodología y Producción Estadística, 2017). Por esto, se considera que puede contribuir a evaluar la validez externa de los resultados.

Las Tablas 4 y 5 muestran las estadísticas descriptivas y diferencias de medias entre tratados y controles utilizando la Encuesta Multipropósito. Se debe notar que el número de observaciones es mayor en esta encuesta. En las estadísticas descriptivas, se destaca que los indicadores del mercado laboral de las personas de Estrato 1 y 2 con ingresos menores a un mínimo, son peores que aquellos de los de la Encuesta de Movilidad (mayor tasa de desempleo y menores tasas de participación y ocupación). La proporción de beneficiarios también es sustancialmente menor que en el caso de la Encuesta de Movilidad, lo cual puede verse explicado en parte por el endurecimiento de las condiciones para acceder al subsidio.

Tabla 4: Estadísticas descriptivas - Encuesta Multipropósito

	Observaciones	Media	Desviación estándar
Mujer	10,932	0.59	0.49
Edad	10,932	42.11	14.10
Participa en el mercado laboral	10,932	0.53	0.50
Desempleado	5,783	0.32	0.47
Ocupado	10,932	0.36	0.48
Puntaje de Sisbén del hogar	10,932	57.91	15.58
Algún miembro del hogar tiene tarjeta SITP-SISBÉN	10,932	0.02	0.15
Número de personas en el hogar	10,932	3.06	1.46
Dependientes	10,932	0.48	0.50
Primaria o menos	10,932	0.33	0.47
Secundaria o Media	10,932	0.53	0.50
Técnico o tecnológico	10,932	0.09	0.28
Universitario	10,932	0.04	0.20
Posgrado	10,932	0.01	0.07
Estrato 1	10,932	0.29	0.45
Estrato 2	10,932	0.71	0.45
Ingresos no laborales	10,932	0.03	0.12
Casado/Vive en pareja	10,932	0.48	0.50

Elaboración propia con base en la Encuesta Multipropósito del 2017. La variable de ingresos no laborales está medida en millones de pesos.

En cuanto a las diferencias de medias, nuevamente se halla que los beneficiarios tienen un menor nivel educativo que las personas sin tarjeta, y también un puntaje de SISBÉN más bajo. En adición, se observa que los beneficiarios tienen una menor probabilidad de vivir en pareja o estar casados, y su tasa de participación es menor que la de los no beneficiarios. Esta correlación negativa puede ser resultado del mecanismo de focalización del subsidio,

pero no es indicativa de un efecto causal del mismo.

Tabla 5: Diferencias de medias Beneficiarios (1) vs No Beneficiarios (0)

	1	0	Dif (1-0)	P	N0	N1
Mujer	0.58	0.59	-0.01	0.84	10,682	250
Edad	45.02	42.04	2.98***	0.00	10,682	250
Participa en el mercado laboral	0.45	0.53	-0.08***	0.01	10,682	250
Desempleado	0.28	0.32	-0.05	0.29	5,671	112
Ocupado	0.32	0.36	-0.04	0.23	10,682	250
Puntaje de Sisbén del hogar	51.02	58.07	-7.05***	0.00	10,682	250
Número de personas en el hogar	2.96	3.06	-0.10	0.31	10,682	250
Dependientes	0.59	0.48	0.11***	0.00	10,682	250
Primaria o menos	0.46	0.33	0.13***	0.00	10,682	250
Secundaria o Media	0.47	0.54	-0.07**	0.04	10,682	250
Técnico o tecnológico	0.05	0.09	-0.03**	0.02	10,682	250
Universitario	0.01	0.04	-0.03***	0.00	10,682	250
Posgrado	0.00	0.01	-0.00	0.74	10,682	250
Estrato 1	0.28	0.29	-0.02	0.57	10,682	250
Estrato 2	0.72	0.71	0.02	0.57	10,682	250
Ingresos no laborales	0.03	0.03	-0.00	0.94	10,682	250
Casado/Vive en pareja	0.38	0.48	-0.10***	0.00	10,682	250

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Se realiza una prueba de diferencias de medias asumiendo varianzas distintas y desconocidas. La variable de ingresos no labores está medida en millones de pesos.

De esta manera, por medio del uso de las dos encuestas, se garantiza que los datos analizados en esta investigación cubren tanto los períodos menos y más exigentes del subsidio, lo cual tiene las ventajas discutidas en la Sección 1.

## 4 Estrategia de identificación

A la luz de la literatura de desajuste espacial y los efectos de programas que buscan mejorar la accesibilidad a oportunidades laborales, se espera que el subsidio al transporte público mejore el desempeño de sus beneficiarios en el mercado laboral. Sin embargo, existe un reto empírico asociado con esta estimación, pues una regresión de Mínimos Cuadrados Ordinarios (MCO) de la participación laboral contra una variable dicótoma que indique si el individuo pertenece a un hogar beneficiario del programa puede resultar en un estimador sesgado del

efecto del programa, incluso controlando por otros determinantes tradicionales de la oferta laboral, como lo son el género, el estado civil, la edad, el nivel educativo y los ingresos no laborales, entre otros (Borjas y Van Ours, 2010).

El sesgo de selección surge por los motivos mencionados en la Sección 1. De hecho, las Tablas 2 y 5 presentan indicios de las diferencias entre los grupos de tratamiento y control. En adición a que el puntaje SISBÉN promedio de los beneficiarios es más bajo, se encuentra que son más jóvenes (más viejos en el caso de la Encuesta Multipropósito), tienen un menor nivel educativo (menor proporción de personas con un título universitario o de posgrado) y pertenecen a un estrato socioeconómico más bajo (esto último solo aplica para la Encuesta de Movilidad).

Nuevamente, el mecanismo de focalización usado con el objetivo de determinar la elegibilidad de los hogares para recibir la tarjeta es el puntaje del SISBÉN, que se determina de manera previa a la implementación del subsidio. Aquellos individuos en hogares con un puntaje menor o igual a 40 puntos tenían derecho a solicitar el beneficio. Naturalmente, se instrumentaría la decisión de tener la tarjeta de SITP/SISBÉN con la elegibilidad. No obstante, el problema de la metodología de Variable Instrumental ignora que, en principio, los hogares con un puntaje de SISBÉN más cercano al umbral de 40 puntos, a cada lado de este, son más comparables entre sí que aquellos hogares más alejados del mismo.

Entonces, para identificar el efecto causal del subsidio de transporte SISBÉN/SITP sobre el desempeño de los individuos en el mercado laboral, se implementa un diseño de Regresión Discontinua (RD), explotando que la focalización del programa se realizó por medio del puntaje del SISBÉN de los hogares<sup>9</sup>, según la metodología III establecida por el Departamento Nacional de Planeación (DNP). Esto genera una discontinuidad en la probabilidad de recibir el tratamiento plausiblemente exógena alrededor del umbral.

Bajo el supuesto de que no hay una manipulación precisa del puntaje alrededor de dicho umbral y de que otros determinantes de la oferta laboral no presentan una discontinuidad en el mismo, se puede interpretar un salto en alguno de los indicadores del mercado laboral considerados (participación, ocupación y desempleo) como un efecto local del subsidio de transporte para individuos en hogares que tienen un puntaje cercano a los 40 puntos.

Como no hay obediencia perfecta por parte de los individuos en cuanto a su asignación a los grupos de tratamiento y control<sup>10</sup>, el tipo de Regresión Discontinua que se debe utilizar

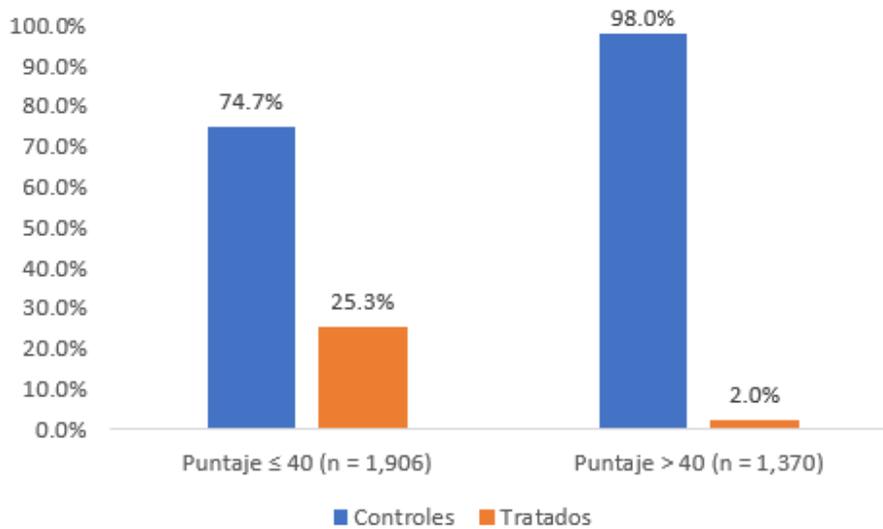
---

<sup>9</sup>A pesar de que el SISBÉN también se utiliza para focalizar otros programas sociales, no existe otro programa que pueda afectar el mercado laboral y con un umbral cercano a los 40 puntos.

<sup>10</sup>Es decir que hay hogares no elegibles que tienen la tarjeta, así como hogares elegibles que no toman el

es la Borrosa (RDB), que se puede conceptualizar como un modelo de Variable Instrumental (VI) local, es decir, una estimación por medio de VI en la que se les asignan pesos a las observaciones que disminuyen a medida que están más lejos del umbral (Wooldridge, 2010). Las Figuras 4 y 5 muestran que en ambas encuestas se observa incumplimiento de la regla de asignación a ambos lados del umbral, aunque es más común que las personas elegibles para recibir el subsidio no opten por tomarlo, en comparación con que los individuos que no son elegibles (puntaje mayor al umbral) obtengan la tarjeta.

Figura 4: Cumplimiento de la asignación al tratamiento: Encuesta de Movilidad

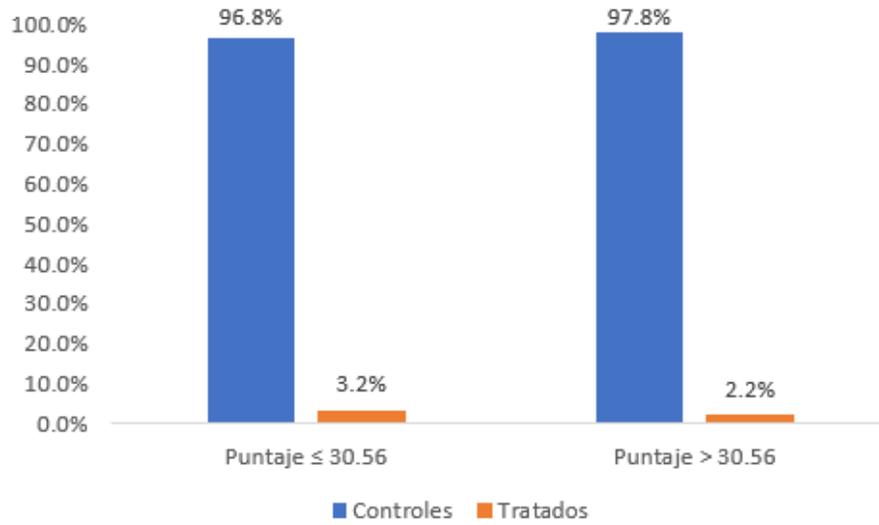


Fuente: Elaboración propia con base en la Encuesta de Movilidad del 2015.

---

subsidio. A este fenómeno se le conoce como no cumplimiento a ambos lados del umbral (Cattaneo et al., 2017).

Figura 5: Cumplimiento de la asignación al tratamiento: Encuesta Multipropósito



Fuente: Elaboración propia con base en la Encuesta Multipropósito del 2017.

Otra ventaja de los diseños de RD es que se pueden implementar un conjunto de pruebas de falsificación y robustez con el objetivo de evaluar de manera transparente el cumplimiento de los supuestos de la metodología. En específico, se implementan las siguientes pruebas: determinar si la densidad del puntaje del SISBÉN es continua en el umbral; si hay una discontinuidad en alguno de los determinantes de la participación laboral en el umbral; y qué tan sensibles son los resultados obtenidos a cambios en el tipo de kernel utilizado y a remover observaciones muy cercanas al umbral (Cattaneo et al., 2017).

Entonces, siguiendo la notación estándar de un diseño de RDB, el procedimiento de estimación se puede entender con las siguientes regresiones:

$$Y_{ik}^+ = \beta_0^+ + \sum_{j=1}^P \beta_j^+ (x_k - c)^j + \epsilon_{ik}^+ \quad (1)$$

$$Y_{ik}^- = \beta_0^- + \sum_{j=1}^P \beta_j^- (x_k - c)^j + \epsilon_{ik}^- \quad (2)$$

$$D_{ik}^+ = \alpha_0^+ + \sum_{j=1}^P \alpha_j^+ (x_k - c)^j + \nu_{ik}^+ \quad (3)$$

$$D_{ik}^- = \alpha_0^- + \sum_{j=1}^P \alpha_j^- (x_k - c)^j + \nu_{ik}^- \quad (4)$$

Donde  $Y_{ik}$  representa a las tres variables de resultado consideradas (dicótomas de participación en el mercado laboral, ocupación y desempleo del  $i$ -ésimo individuo que vive en el  $k$ -ésimo hogar).  $D_{ik}$  es el indicador de ser beneficiario del subsidio<sup>11</sup>,  $x_k$  es el puntaje de SISBÉN del  $k$ -ésimo hogar y  $c$  es el umbral que determina la elegibilidad de los individuos (40 para el 2015 y 30,56 en el 2017). Los superíndices  $+$  y  $-$  son utilizados para indicar que las regresiones se estiman por separado a la derecha e izquierda del umbral, respectivamente. Cabe resaltar que, dependiendo del orden del polinomio escogido ( $P$ ), esta metodología se conoce como regresión lineal local o regresión cuadrática local<sup>12</sup>. Los  $\beta$  y  $\alpha$  son parámetros a estimar. Por último,  $\epsilon_i$  y  $\nu_i$  son términos de perturbación.

Es importante recordar que a cada observación en las estimaciones de las ecuaciones (1) a (4) se le asigna un peso por medio de la función  $\kappa(\frac{x_i - c}{h})$ , donde  $\kappa$  es un tipo de kernel (por ejemplo, triangular o Epanechnikov) y  $h$  es un ancho de banda<sup>13</sup>, que determina hasta qué distancia del umbral se utilizan observaciones para calcular el Efecto local promedio del Tratamiento (LATE). Este último es igual al estimador de Wald en la literatura de variables instrumentales (Angrist y Pischke, 2008)<sup>14</sup>:

$$\tau_{RDB} = \frac{\hat{\beta}_0^- - \hat{\beta}_0^+}{\hat{\alpha}_0^- - \hat{\alpha}_0^+} \quad (5)$$

Intuitivamente, el estimador calcula el salto en los indicadores de mercado laboral en el umbral, y lo pondera por el salto en la probabilidad de recibir el subsidio en este mismo umbral, considerando solo observaciones con puntajes cercanos al umbral.

También se estiman regresiones de MCO y VI sin pesos para tener un punto de comparación para las estimaciones de RDB. En el modelo de VI, la especificación de la

---

<sup>11</sup>Para la construcción de esta variable, se asume que todos los miembros de un hogar que figura como beneficiario del subsidio tienen una tarjeta, excepto aquellos que tienen menos de 16 años o quienes que viven en hogares que figuran como beneficiarios de otros subsidios del SITP (Adulto mayor y Discapacitado).

<sup>12</sup>Para estas estimaciones, no se emplean polinomios locales de un orden superior, siguiendo las recomendaciones de Gelman y Imbens (2018).

<sup>13</sup>La elección del ancho de banda óptimo ha sido ampliamente estudiada por la literatura y actualmente hay una variedad de alternativas para su determinación, incluso para el caso de RDB (Arai y Ichimura, 2016; Calonico et al., 2017).

<sup>14</sup>Normalmente, el estimador se escribe restando a las constantes de la derecha del umbral las constantes de la izquierda, mas aquí se invierte la notación, porque los elegibles son quienes se encuentran a la izquierda.

primera etapa y la forma reducida son, respectivamente<sup>15</sup>:

$$D_{ik} = \gamma_0 + \gamma_1 \mathbb{1}(x_k \leq c) + \gamma_2(x_k - c) + \sum_{j=3}^J \gamma_j z_{j-2ik} + \epsilon_{ik} \quad (6)$$

$$Y_{ik} = \phi_0 + \phi_1 \mathbb{1}(x_k \leq c) + \phi_2(x_k - c) + \sum_{j=3}^J \phi_j z_{j-2ik} + \nu_{ik} \quad (7)$$

donde la notación representa lo mismo que en el modelo de RDB, a excepción de  $\mathbb{1}$  (la función indicador), la cual representa una variable dicótoma que toma el valor de 1 cuando el puntaje de SISBÉN del hogar es menor al umbral que define la asignación del tratamiento; y de las  $z_{ik}^j$ , que denotan las variables de control para el individuo  $i$  en el hogar  $k$ <sup>16</sup>. Estas incluyen determinantes tradicionales de la oferta laboral, como la edad, el nivel educativo, la presencia de dependientes en el hogar y el número de personas en el hogar<sup>17</sup>. Este conjunto también incluye variables dicótomas para cada una de las localidades (UPZs en el caso de la Encuesta Multipropósito) de la ciudad de Bogotá.

## 5 Resultados

Las Tablas 6 a 8 muestran los resultados de estimar, por medio de MCO y VI, el efecto del subsidio de transporte sobre la participación en el mercado laboral, la tasa de ocupación y la tasa de desempleo para los hogares pertenecientes a los Estratos 1 y 2 y cuyos ingresos mensuales no superen un salario mínimo. Como fue mencionado con anterioridad, estos modelos se estiman con el objetivo de tener un punto de comparación y analizar cómo cambian los resultados al implementar diferentes metodologías. En las columnas (1) y (2), se muestran los resultados de la estimación de MCO en la que las variables dependientes son los indicadores del mercado laboral mencionados, y las independientes son una variable dicótoma que indica si en el hogar hay al menos una persona que tenga la tarjeta y variables de control. No se encuentra una correlación positiva entre el subsidio y el desempeño en el mercado laboral en ninguno de los casos. De manera similar, las columnas (3) y (4) corresponden a la segunda etapa de la estimación por VI, y muestran que los resultados se mantienen al

---

<sup>15</sup>En el Apéndice C se muestran resultados de estas estimaciones para la Encuesta de Movilidad, pero con un polinomio cuadrático de la variable de asignación. Los resultados principales no cambian al realizar este procedimiento.

<sup>16</sup>Las variables específicas utilizadas se detallan en la Sección 5.

<sup>17</sup>En las estimaciones realizadas con la Encuesta Multipropósito, adicionalmente, se incluyen los ingresos no laborales y el estado civil del individuo.

instrumentar la variable dicótoma de tratamiento con un indicador de la elegibilidad del hogar (una variable dicótoma que es 1 si el puntaje es menor o igual que el umbral y cero de lo contrario). Aunque el signo de los coeficientes estimados cambia al comparar MCO con VI, ninguno de los coeficientes es estadísticamente distinto de cero. También cabe resaltar que la estimación usando la metodología de variable instrumental resulta más imprecisa, como es de esperarse.

En estas estimaciones, se utilizan controles considerados como determinantes tradicionales del desempeño de los individuos en el mercado laboral. Estos son: la edad; el nivel educativo; la presencia de dependientes en el hogar; y el tamaño del hogar. Los signos de estos coeficientes son consistentes con lo esperado<sup>18</sup>. Se incluyen los mismos controles para las 3 variables de resultado consideradas, pero no son presentados en las tablas por cuestiones de simplicidad. Sin embargo, están disponibles a pedido.

Tabla 6: Tasa de participación

	(1)	(2)	(3)	(4)
	OLS Hombres	OLS Mujeres	IV Hombres	IV Mujeres
Tarjeta Sisbén/SITP	0.009 (0.026)	0.043 (0.031)	-0.029 (0.121)	-0.030 (0.136)
Constante	-0.630*** (0.101)	-0.418*** (0.122)	-0.638*** (0.112)	-0.390*** (0.128)
Observaciones	1,464	1,812	1,464	1,812
R-cuadrado	0.317	0.121		
Dummies de localidad	Sí	Sí	Sí	Sí
Estadístico F			23.237	13.28

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (3) y (4) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

<sup>18</sup>La relación entre la edad y la participación tiene la forma de una u invertida; un mayor nivel educativo se encuentra asociado con una mayor participación; y la presencia de dependientes incrementa la probabilidad de participar más para hombres que para mujeres.

Tabla 7: Tasa de ocupación

	(1)	(2)	(3)	(4)
	OLS Hombres	OLS Mujeres	IV Hombres	IV Mujeres
Tarjeta Sisbén/SITP	0.035 (0.030)	0.043 (0.030)	-0.061 (0.136)	-0.098 (0.136)
Constante	-0.700*** (0.114)	-0.453*** (0.119)	-0.688*** (0.128)	-0.409*** (0.126)
Observaciones	1,464	1,812	1,464	1,812
R-cuadrado	0.281	0.114		
Dummies de localidad	Sí	Sí	Sí	Sí
Estadístico F			25.577	10.092

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (3) y (4) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad<sup>2</sup>*, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Adicionalmente, las Tablas 9 y 10 muestran los resultados de la primera etapa de la estimación por variable instrumental, es decir, de las columnas (3) y (4) de las Tablas 6, 7 y 8. Esto corresponde a una regresión de MCO en la que la variable dependiente es la participación en el programa y las independientes son la variable dicótoma que indica si el hogar es elegible o no para que sus miembros reciban el tratamiento, las variables de control empleadas en la segunda etapa, el puntaje del sisbén centrado y la interacción entre esta variable y la dicótoma que indica la elegibilidad<sup>19</sup>. Nuevamente, no se muestran los coeficientes de estas variables de control por cuestiones de presentación y simplicidad de las tablas, pero están disponibles a pedido.

<sup>19</sup>Esto último se hace siguiendo a Wooldridge (2010), Pág. 958.

Tabla 8: Tasa de desempleo

	(1)	(2)	(3)	(4)
	OLS Hombres	OLS Mujeres	IV Hombres	IV Mujeres
Tarjeta Sisbén/SITP	-0.031 (0.026)	-0.003 (0.026)	0.011 (0.107)	0.145 (0.118)
Constante	0.441*** (0.128)	0.506*** (0.134)	0.428*** (0.137)	0.489*** (0.135)
Observaciones	1,155	916	1,155	916
R-cuadrado	0.065	0.083		
Dummies de localidad	Sí	Sí	Sí	Sí
Estadístico F			3.383	2.935

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (3) y (4) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Se observa que, *ceteris paribus*, ser elegible para recibir el subsidio (tener un puntaje de SISBÉN por debajo de los 40 puntos) está positivamente asociado con que al menos algún miembro del hogar tenga la tarjeta. En particular, para los hombres, ser elegible incrementa la probabilidad de tener el subsidio en 26 puntos porcentuales, y para las mujeres el efecto es de aproximadamente 29 puntos porcentuales. Estos resultados, junto con los presentados en las tablas 6 a 8, parecen indicar que, a pesar de que las personas elegibles toman el subsidio más que las no elegibles, esto no parece afectar sus resultados en el mercado laboral.

Por último, las Tablas 9 y 10 también muestran algunos estadísticos que prueban si el instrumento en efecto está correlacionado con la participación en el tratamiento. Se muestra el P-value del estadístico F de instrumentos excluidos, el cual indica que la correlación entre el instrumento (puntaje del SISBÉN) y la variable endógena (tener la tarjeta de SITP/SISBÉN) es estadísticamente significativa. Adicionalmente, el Estadístico F de Kleibergen-Paap tiene un valor lo suficientemente alto<sup>20</sup> para concluir que dicha correlación no es débil, que, de suceder, podría llevar a estimaciones sesgadas. Por último, se muestra el P-value de la prueba

<sup>20</sup>Este valor se compara con los calculados por Stock y Yogo (2005).

de Anderson-Rubin para inferencia robusta a instrumentos débiles, cuyo valor indica que no se rechaza la hipótesis nula. En otras palabras, los coeficientes de las variables endógenas en la ecuación estructural son conjuntamente iguales a 0.

En conjunto, estos resultados preliminares no son consistentes con lo planteado en las hipótesis, pues a pesar de que ser elegible sí está positivamente correlacionado con tomar el subsidio, no parece que los miembros de los hogares con tarjeta tengan un mejor desempeño en el mercado laboral que aquellos en hogares sin tarjeta. No obstante, podría ser el caso que la estimación de variables instrumentales no esté comparando hogares tratados y controles similares entre sí, pues no se está restringiendo la estimación solo a hogares con puntajes de SISBÉN cercanos al umbral de 40 puntos.

Tabla 9: Primera etapa: Tasa de participación/Tasa de ocupación

	(1)	(2)
	Primera etapa - Hombres	Primera etapa - Mujeres
Puntaje Sisbén menor a 40 puntos	0.260*** (0.029)	0.288*** (0.026)
Constante	0.415*** (0.119)	0.374*** (0.099)
Observaciones	1,464	1,812
R-cuadrado	0.124	0.138
Dummies de localidad	Sí	Sí
F de instrumentos excluidos	77.909	120.531
P-value F	0	0
Estadístico F de KP - Identificación débil	77.909	120.531
P-value prueba de AR - Inferencia robusta a instrumentos débiles	.813	.829

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. La prueba de AR - Inferencia robusta a instrumentos débiles tiene como hipótesis nula que los coeficientes de las variables endógenas en la ecuación estructural son en conjunto iguales a 0. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 10: Primera etapa: Tasa de desempleo

	(1)	(2)
	Primera etapa - Hombres	Primera etapa - Mujeres
Puntaje Sisbén menor a 40 puntos	0.268*** (0.034)	0.327*** (0.040)
Constante	0.411*** (0.147)	0.241 (0.159)
Observaciones	1,155	916
R-cuadrado	0.127	0.156
Dummies de localidad	Sí	Sí
F de instrumentos excluidos	63.41	67.367
P-value F	0	0
Estadístico F de KP - Identificación débil	63.41	67.367
P-value prueba de AR - Inferencia robusta a instrumentos débiles	.916	.218

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. La prueba de AR - Inferencia robusta a instrumentos débiles tiene como hipótesis nula que los coeficientes de las variables endógenas en la ecuación estructural son en conjunto iguales a 0. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Teniendo en cuenta los resultados preliminares, se realiza una estimación por medio de un diseño de Regresión Discontinua Borrosa, que solo tiene en cuenta observaciones que estén dentro de un rango del umbral de 40 puntos. Dicho rango está definido por un ancho de banda seleccionado de forma óptima, y las observaciones también reciben un peso menor en la estimación mientras más alejadas del umbral se encuentren. Los resultados de estas estimaciones se presentan en las tablas 11 a 13<sup>21</sup>.

Cada columna presenta el estimador del Efecto Promedio Local del Tratamiento (LATE por sus siglas en inglés), junto con el error estándar convencional. Siguiendo a Calonico et al. (2014), se presentan intervalos para el LATE y la primera etapa en los que se realiza inferencia robusta a la escogencia de anchos de banda demasiado amplios. El kernel utilizado es el triangular<sup>22</sup>, el cual pondera a las observaciones de la siguiente forma:  $\kappa(x_i) = 1 - \left| \frac{x_i - c}{h} \right|$ . Adicionalmente, se muestra el número de observaciones empleadas en la estimación, el ancho de banda óptimo  $h$  utilizado, y el método con el que se selecciona dicho ancho de banda.

<sup>21</sup>Se realizan estas estimaciones por medio del comando *rdrobust* en Stata (Calonico et al., 2017).

<sup>22</sup>En el apéndice, se presentan resultados de la estimación de Regresión Discontinua Borrosa empleando el kernel Uniforme y el Epanechnikov, y en ningún caso se encuentra un LATE estadísticamente distinto de cero.

*mserd* se refiere a que el ancho de banda fue escogido con el objetivo de minimizar el Error Cuadrático Medio (ECM) de la estimación del LATE.

Tabla 11: Regresión Discontinua Borrosa: Tasa de participación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.377 (0.247)	-0.417 (0.269)	0.072 (0.231)	0.083 (0.254)
Intervalo robusto LATE 95%	[-1.227 .094]	[-1.111 .194]	[-.564 .653]	[-.622 .680]
Ancho de banda $h$	11.685	22.713	15.704	22.605
Tipo de ancho de banda	<i>mserd</i>	<i>mserd</i>	<i>mserd</i>	<i>mserd</i>
Número eficiente de observaciones	585	1040	986	1322
Kernel	Triangular	Triangular	Triangular	Triangular
Coficiente primera etapa	-.319	-.311	-.277	-.308
Intervalo robusto primera etapa 95%	[-.558 -.148]	[-.58 -.196]	[-.499 -.202]	[-.587 -.225]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 12: Regresión Discontinua Borrosa: Tasa de ocupación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.268 (0.254)	-0.345 (0.274)	-0.054 (0.225)	-0.097 (0.252)
Intervalo robusto LATE 95%	[-.999 .335]	[-.977 .354]	[-.71 .483]	[-.763 .525]
Ancho de banda $h$	17.428	22.58	14.005	21.485
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	863	1040	916	1287
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.258	-.312	-.303	-.319
Intervalo robusto primera etapa 95%	[-.49 -.173]	[-.581 -.194]	[-.551 -.228]	[-.606 -.236]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 13: Regresión Discontinua Borrosa: Tasa de desempleo

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.005 (0.167)	-0.069 (0.194)	0.235 (0.261)	0.255 (0.275)
Intervalo robusto LATE 95%	[-.472 .374]	[-.591 .324]	[-.372 .924]	[-.468 .852]
Ancho de banda $h$	12.764	17.931	12.046	17.002
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	488	670	406	552
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.293	-.321	-.305	-.362
Intervalo robusto primera etapa 95%	[-.541 -.125]	[-.582 -.102]	[-.594 -.131]	[-.693 -.137]

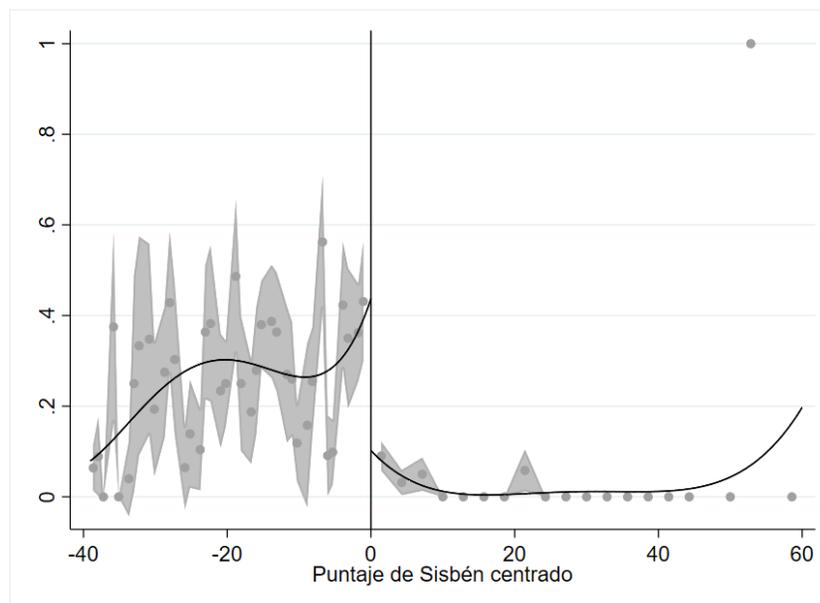
Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Nuevamente, no se encuentra ningún efecto significativo de la tarjeta de SITP/SISBÉN

sobre los indicadores del mercado laboral considerados, lo cual sugiere que los subsidios no logran mejorar el acceso de los individuos a oportunidades laborales, como sugería la hipótesis de la investigación y la literatura sobre desajuste espacial. Las Figuras 6 y 7 proveen una representación gráfica de los resultados, en la cual se evidencia un salto en la probabilidad de que algún miembro del hogar tenga la tarjeta justo en el umbral de 40 puntos, más este salto no está acompañado de ninguna discontinuidad en las variables de resultado. La mayor diferencia entre estos resultados y los obtenidos por medio de la variable instrumental sin pesos se dan en la magnitud estimada del salto en la probabilidad de que algún miembro del hogar tenga la tarjeta, que ahora oscila entre los 60 y los 20 puntos porcentuales, dependiendo del polinomio utilizado (lineal o cuadrático).

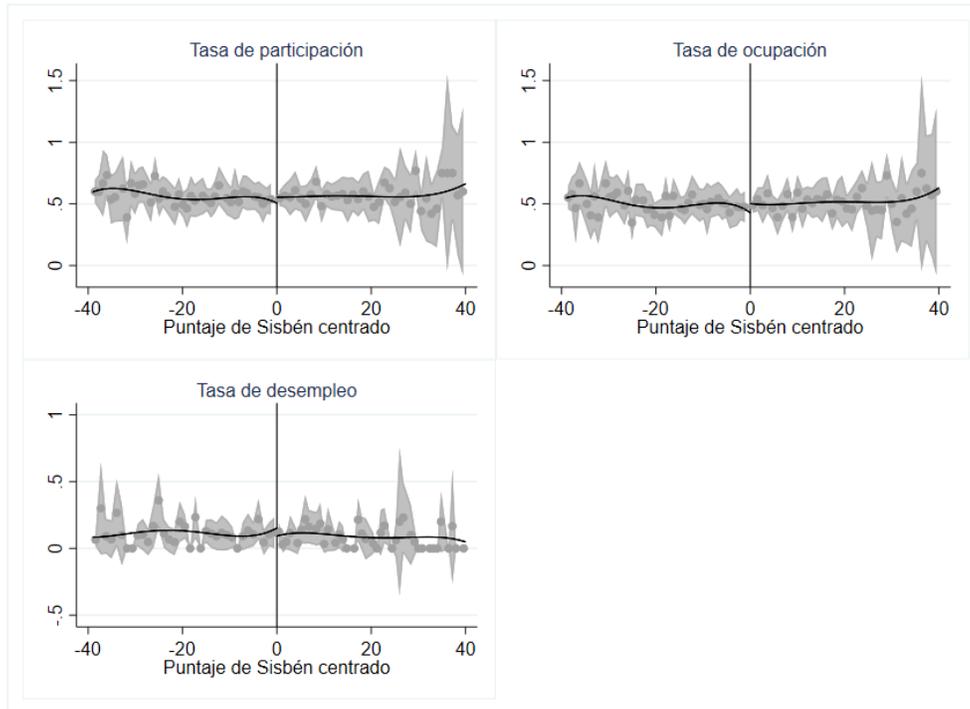
Entonces, aunque desde un punto de vista metodológico sea mejor emplear un diseño de Regresión Discontinua Borrosa para estimar el efecto de los subsidios, en la práctica se encuentra que los resultados utilizando las otras dos metodologías (MCO y VI sin pesos) conducen a las mismas conclusiones: no hay evidencia estadística de que el subsidio incremente la participación de los individuos en el mercado laboral, la tasa de ocupación o de que reduzca la probabilidad de estar desempleado.

Figura 6: Salto en la probabilidad de recibir tratamiento en el umbral



Fuente: Elaboración propia. Se ajustan polinomios de cuarto grado a cada lado del umbral para obtener una aproximación global de la función de regresión.

Figura 7: Salto en los indicadores del mercado laboral en el umbral



Fuente: Elaboración propia. Se ajustan polinomios de cuarto grado a cada lado del umbral para obtener una aproximación global de las funciones de regresión.

Si bien el subsidio representa una disminución en los costos de utilizar el sistema de transporte público, y por ende una potencial reducción de los costos de búsqueda de empleo, no parece estar incentivando a las personas a ingresar en el mercado laboral, ni tampoco tiene un efecto sobre las probabilidades de estar ocupado o desempleado. Cabe destacar que estos resultados consistentes con los encontrados por Rodríguez et al. (2015), quienes usan otros datos y otra metodología, y tampoco encuentran impacto alguno del subsidio sobre desempleo ni participación.

## 5.1 Pruebas de falsificación y robustez

Como se mencionó con anterioridad, los Diseños de Regresión Discontinua cuentan con varias pruebas que permiten evaluar la validez de los resultados obtenidos y el cumplimiento de los supuestos fundamentales de la metodología para poder interpretar los estimadores como efectos causales y asegurarse de que las estimaciones están libres de sesgo de selección.

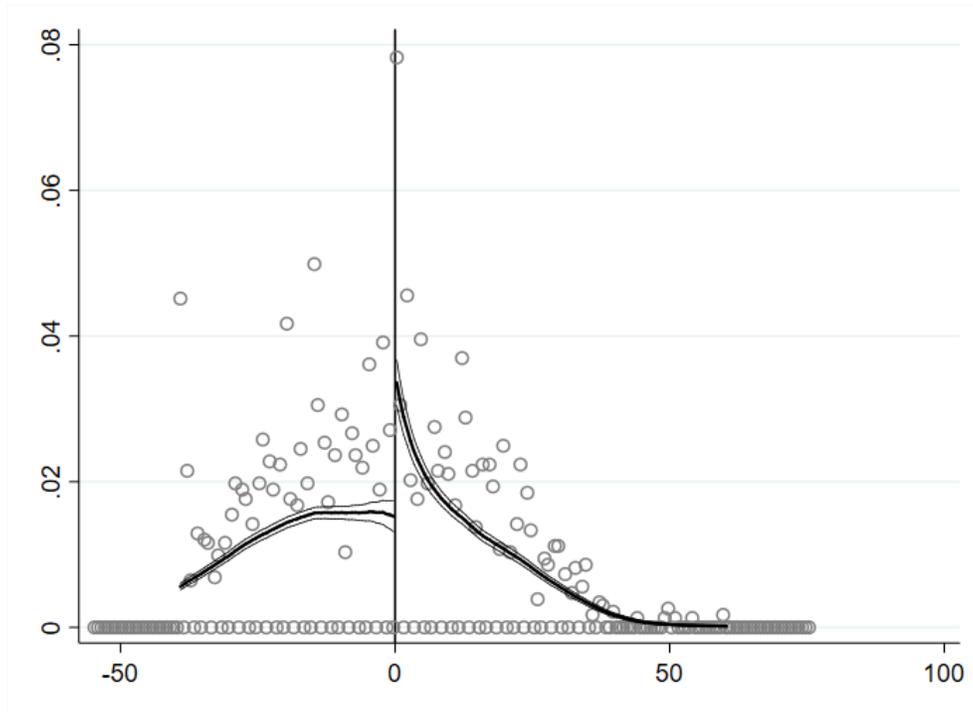
En este orden de ideas, la primera prueba implementada es evaluar la continuidad de la distribución de la variable de asignación alrededor del umbral, que determina la elegibilidad

de los hogares para recibir el subsidio. Esta prueba se conoce como de manipulación, y su objetivo es verificar que los individuos no tengan un control preciso sobre su puntaje alrededor del umbral (Lee y Lemieux, 2010).

Con este fin, se implementan las pruebas de McCrary y Cattaneo et al. La primera de estas consiste en un procedimiento de dos etapas para probar si hay evidencia de una discontinuidad en la densidad de la variable de asignación. En la primera etapa, con base en el puntaje de SISBÉN, se particiona la muestra en canecas (bins) igualmente espaciadas, y se calcula la frecuencia dentro de cada caneca. La segunda etapa trata estos conteos de frecuencia como la variable dependiente en una regresión lineal local (McCrary, 2008).

La Figura 8 muestra gráficamente los resultados de esta prueba, y en ella se observa claramente una discontinuidad en la densidad del puntaje de SISBÉN centrado justo en el umbral, que puede ser un indicio de que los individuos se seleccionan para afectar su estado de elegibilidad. No obstante, McCrary (2008) menciona que el estimador del efecto del tratamiento puede seguir siendo insesgado, incluso cuando hay una manipulación importante de la variable de asignación que cause un salto en la densidad, pues que la densidad de esta variable sea continua no siempre es una condición necesaria ni suficiente para la identificación. En particular, se espera que la prueba sea informativa cuando la manipulación es monótonica, es decir, cuando el programa incentiva a los individuos a ajustar su variable de asignación solamente en una dirección.

Figura 8: Prueba de manipulación: McCrary

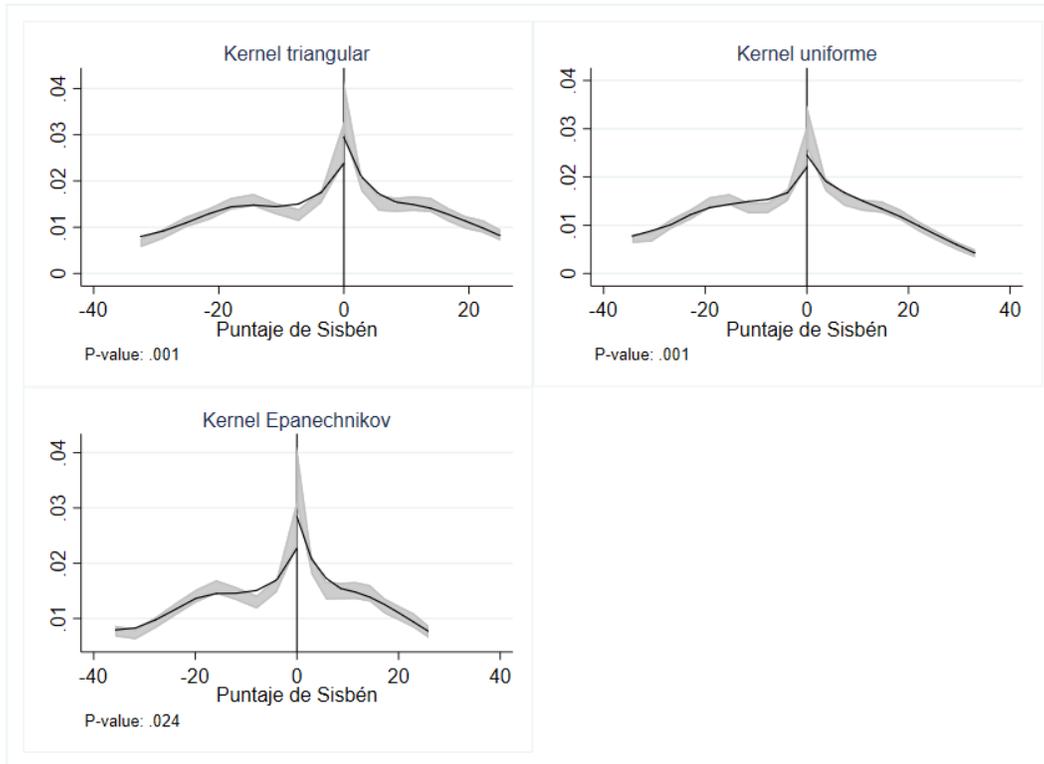


Fuente: Elaboración propia con base en McCrary (2008).

Adicionalmente, la figura 9 muestra los resultados de la prueba de Cattaneo, Janson y Ma, la cual se basa en la de McCrary, pero tiene mejores propiedades en términos de tamaño y potencia (Cattaneo et al., 2018). Para esta prueba, como en las estimaciones, se puede escoger una función kernel para ponderar las observaciones. Es posible evidenciar que, independientemente del kernel utilizado, se rechaza la hipótesis nula de que no hay una discontinuidad en el umbral, resultado consistente con el de la prueba de McCrary.

Aunque la manipulación de la variable de asignación pone en entredicho la validez de un diseño de Regresión Discontinua Nítida (RDN), puede esperarse un poco de manipulación en el caso de una Regresión Discontinua Borrosa, pues a ambos lados del umbral se observan individuos que acceden al subsidio. Además, llama la atención que la densidad del puntaje del SISBÉN es mayor del lado derecho del umbral, lo que sugiere que los individuos se están seleccionando para evitar ser elegibles para recibir el tratamiento. Esto resulta contraintuitivo, pues es de esperarse que, de haber una manipulación sistemática del puntaje del SISBÉN, los hogares buscarían obtener un puntaje menor que el verdadero, haciéndose pasar como elegibles de esta manera y accediendo a la tarjeta de forma más sencilla.

Figura 9: Prueba de manipulación: rddensity



Fuente: Elaboración propia con base en Cattaneo et al. (2018).

Sin embargo, esta evidencia indica que es necesario probar la sensibilidad de los resultados a la aparente manipulación de la variable de asignación. Con este fin, otra prueba de falsificación que se implementa es excluir las observaciones que se encuentren dentro de un rango del umbral definido arbitrariamente y con puntajes cercanos a 40. La lógica detrás de esta prueba es que, de haber manipulación, es de esperar que aquellos individuos que se encuentran más cerca al umbral sean los que tienen una mayor probabilidad de haber manipulado su puntaje. A esta técnica se le conoce como “donut-hole” (Cattaneo et al., 2017).

Se prueba excluyendo observaciones a distancias menores a 1, 2 y 3 puntos del umbral de 40 puntos, y se presentan los resultados de volver a estimar el modelo de RDB con las observaciones restantes en las Tablas 14 a 22 del Apéndice A. Los resultados principales se mantienen para la gran mayoría de las regresiones estimadas. Es decir, el intervalo robusto del LATE siempre contiene a 0 y en la primera etapa se encuentra que estar por debajo del umbral incrementa la probabilidad de tener la tarjeta. No obstante, cuando se excluyen las observaciones que están a 3 o menos puntos del umbral, en algunos ocasiones se pierde

la significancia estadística de la primera etapa, posiblemente porque el menor número de observaciones lleva a una pérdida de precisión en las estimaciones.

La segunda prueba realizada fue evaluar si hay evidencia de alguna discontinuidad en la densidad de otros determinantes del desempeño de los individuos en el mercado laboral en el umbral de 40 puntos. La lógica detrás de esta prueba es que, si los hogares carecen de la capacidad de manipular de manera precisa el puntaje de SISBÉN que reciben, no debería haber diferencias sistemáticas entre hogares con puntajes similares pero que se encuentran a lados opuestos del umbral, más allá de las encontradas en la probabilidad de recibir tratamiento y (en caso de haberlas encontrado) en las variables de resultado que podrían haberse visto afectadas por el tratamiento (Cattaneo et al., 2017).

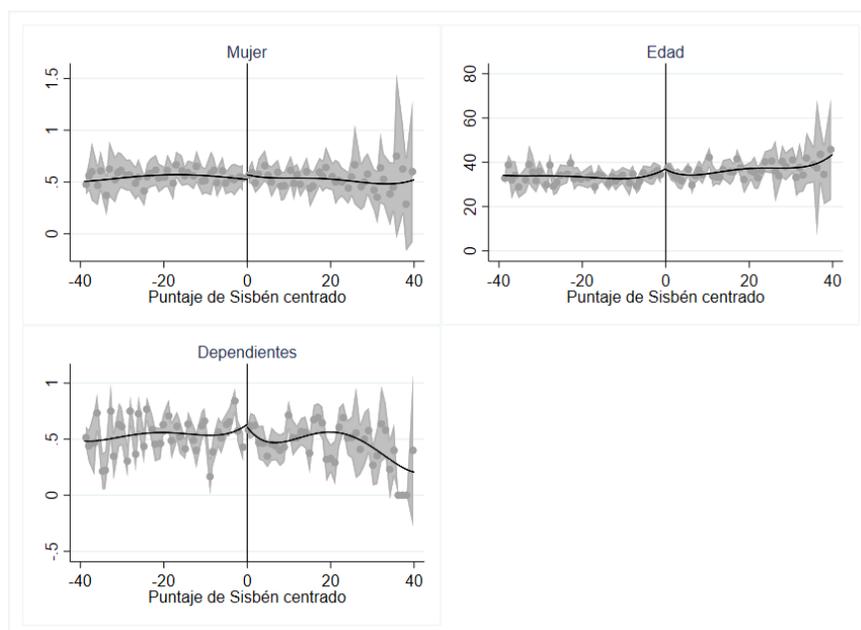
El procedimiento involucra tratar a cada una de las variables de control como variables de resultado, de manera que se estiman diseños de RDB para cada una. Es decir, se escoge un ancho de banda para cada una y se evalúa si hay un salto estadísticamente significativo en el umbral de la variable de asignación. Se muestra evidencia gráfica de que las variables de control consideradas no saltan en el umbral de manera significativa o sistemática en las Figura 10 y 11 <sup>23</sup>.

La tercera prueba implementada consiste en cambiar el kernel utilizado para estimar la RDB. A pesar de que la literatura de estadística ha mostrado que el Triangular es óptimo para estimar regresiones lineales locales en la frontera (Fan et al., 1996), también puede ser útil verificar que los resultados no dependen de la elección de un tipo de kernel en particular. Los resultados no cambian de manera significativa al utilizar los kernels Epanechnikov y Uniforme, como se puede ver en las Tablas 23 a 28 del Apéndice B.

---

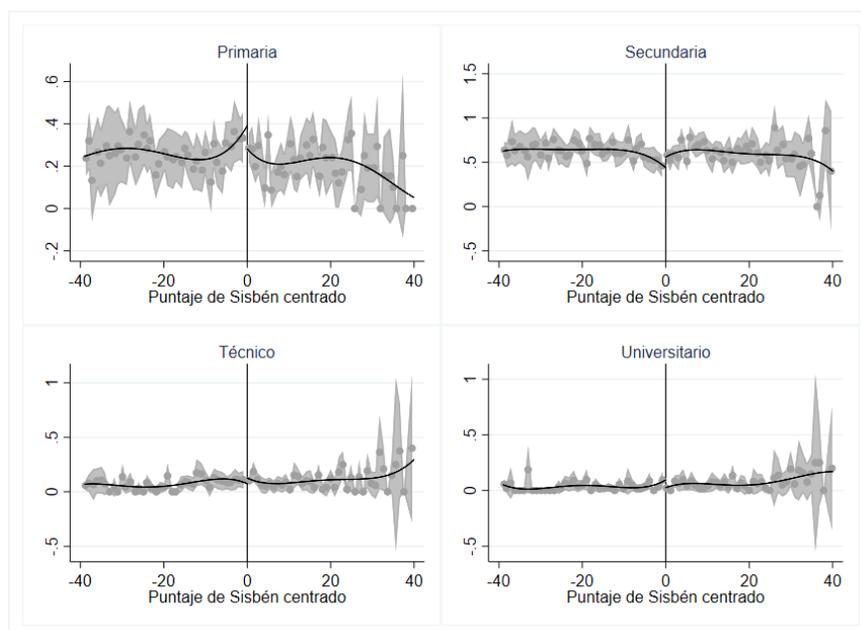
<sup>23</sup>No se presentan las tablas con los resultados formales de estas estimaciones por cuestiones de simplicidad, pero estas se encuentran disponibles a pedido.

Figura 10: Continuidad de las variables predeterminadas alrededor del umbral



Fuente: Elaboración propia. Se ajustan polinomios de cuarto grado a cada lado del umbral para obtener una aproximación global de la función de regresión.

Figura 11: Continuidad de las variables predeterminadas alrededor del umbral



Fuente: Elaboración propia. Se ajustan polinomios de cuarto grado a cada lado del umbral para obtener una aproximación global de la función de regresión.

Por último, en el Apéndice D, se estiman los mismos modelos de las Tablas 6 a 13 con datos de la Encuesta Multipropósito de Bogotá para el año 2017, con el objetivo de darle a los resultados cierto grado de validez externa. Los datos adicionales permiten evaluar el efecto del programa algunos años después de su implementación, año en el cual los requisitos para tener la tarjeta ya eran más exigentes (el umbral se bajó de 40 a 30,56 puntos y los beneficios tarifarios fueron reducidos). Otro beneficio de esta encuesta es que posibilita el hecho de incluir dos variables de control importantes adicionales, los ingresos no laborales y el estado civil de los individuos. La limitación de la Encuesta Multipropósito radica en que el puntaje del SISBÉN con el que se dispone en este caso es uno calculado por el DNP a partir de las características de los hogares y sus miembros, a diferencia del de la Encuesta de Movilidad, que es autoreportado.

Las Tablas 34 a 41 y las Figuras B.1 a B.6 presentadas en el Apéndice D muestran que, en el caso de la Encuesta Multipropósito, tampoco se encuentra evidencia de que los beneficiarios tengan mejores resultados en el mercado laboral que los no beneficiarios. Empero, cabe resaltar que la correlación entre la elegibilidad y la participación es sustancialmente más débil para estos datos. En efecto, no parece haber una asociación estadísticamente significativa ni de gran magnitud entre encontrarse por debajo del umbral de 30,56 puntos y que algún miembro del hogar tenga la tarjeta, lo cual va en contravía de lo encontrado usando los datos de la Encuesta de Movilidad. Es posible que esto se deba a la discrepancia entre el verdadero puntaje de SISBÉN de un hogar y el calculado por la Secretaría de Planeación, el endurecimiento de las condiciones requeridas para ser elegible y la reducción de los beneficios tarifarios para los tarjetahabientes. Vale la pena destacar que no hay evidencia de manipulación de la variable de asignación (Ver Figuras B.5 y B.6).

## 5.2 Discusión

Los resultados son similares a los de Rodríguez et al. (2015), quienes evalúan los efectos de los subsidios al transporte público en Bogotá por medio de otra metodología y utilizando otras fuentes de información, y también encuentran un efecto nulo del subsidio sobre la participación laboral y la probabilidad de estar desempleado. Sin embargo, sí encuentran un efecto positivo del subsidio sobre los ingresos por hora de los trabajadores informales. ¿Por qué esta iniciativa, que es similar a otras que han sido exitosas en otras ciudades de países en vía de desarrollo (Abebe et al., 2016; Franklin et al., 2015), no parece contribuir a mejorar el desempeño de los individuos en el mercado laboral bogotano?

Se pueden mencionar dos razones por las cuales es posible que no se encuentre un efecto causal del subsidio sobre los indicadores del mercado laboral considerados.

En primer lugar, se han documentado fenómenos que afectan de manera diferencial a las mujeres y pueden incentivarlas a dejar de utilizar el transporte público. Por ejemplo, Kash (2019) expone que el 37% de las mujeres usuarias de Transmilenio incluidas en su estudio reportaron algún tipo de contacto sexual no deseado mientras usaban el sistema. La autora concluye que una de los comportamientos mediante los cuales las usuarias reaccionan a estos episodios es optar por no utilizar el sistema de transporte, lo cual va en contra de los incentivos creados por los subsidios.

En segundo lugar, la implementación del Sistema Integrado de Transporte Público en Bogotá no ha sido finalizada todavía. El SITP provisional, que consiste en una flota de buses antiguos en los que el medio de pago es el efectivo y no la tarjeta inteligente, todavía cuenta con 4,800 vehículos en circulación, los cuales cubren 29 rutas (Alcaldía de Bogotá, 2019; El Espectador, 2019). Adicionalmente, las tarifas de los buses no son sustancialmente mayores a las subsidiadas del SITP<sup>24</sup>. Si a esto se le suma que las rutas de este transporte tradicional son más conocidas por los ciudadanos (porque están en operación desde antes que las rutas del SITP) y más prácticas (porque no requieren el uso de la tarjeta, sino que el pago es en efectivo), es plausible que la competencia contribuya a explicar que las personas con tarjeta no incrementen su uso de Transmilenio y el SITP (Concejo de Bogotá, 2018).

Como mencionan Rodríguez et al. (2015), el sistema de tarifas del transporte público en Bogotá está diseñado con el objetivo de que las tarifas cobradas sean suficientes para cubrir los costos de operación, manejo, mantenimiento y reemplazo de la flota de buses. El desafío que surge a raíz de cobrar una tarifa que permite recuperar los costos es que la población de bajos ingresos puede ser excluida vía precios de usar el sistema. Las tarifas subsidiadas surgen como una iniciativa para evitar esta exclusión. Sin embargo, la evidencia proveída en esta investigación y en artículos previos parece indicar la existencia de obstáculos que no permiten el funcionamiento adecuado de los subsidios. Futuras investigaciones deberían esforzarse por explorar con más detalle dichos obstáculos.

## 6 Conclusiones

El problema del desajuste espacial en Bogotá contribuye a exacerbar las disparidades preexistentes en el desempeño en el mercado laboral de distintos grupos de individuos.

---

<sup>24</sup>Para el año 2019, mientras que las tarifas subsidiadas de los buses de los servicios troncales (Transmilenio) y zonales (Urbano, Complementario y Especial) eran de \$1,800 y \$1,650, respectivamente, las tarifas diurna y nocturna, dominical y festiva del SITP provisional/Transporte Público Colectivo eran de de \$1,750 y \$1,800 respectivamente.

Por ejemplo, la brecha de género en términos de participación laboral puede ser afectada diferencialmente por la segregación residencial, pues las mujeres se enfrentan a estructuras familiares que les imponen limitaciones espaciales (Díaz Escobar, 2015).

En dicho contexto, el transporte público es un elemento a través del cual los residentes de la periferia pueden acceder a oportunidades de todo tipo, entre ellas las laborales, que se encuentran aglomeradas en otras zonas de la ciudad. Como destaca Guzman et al. (2018), “los sistemas de transporte público pueden reconfigurar las estructuras espaciales y sociales de una ciudad”. No obstante, el esquema tarifario del Sistema Integrado de Transporte Público funciona cobrando una tarifa diseñada para cubrir los costos de operación, la cual puede no ser asequible para el segmento de la población más pobre y desconectado espacialmente. Por eso, la hipótesis de esta investigación planteó que el programa de subsidios al transporte público implementado por la Administración Distrital de Bogotá desde 2014, era una política que podía contrarrestar esta desventaja y también beneficiar de forma diferencial a los grupos más perjudicados, contribuyendo a disminuir la brecha de género en términos de desempeño en el mercado laboral.

Empero, los resultados sugieren que los subsidios al transporte público en la ciudad de Bogotá no incentivan a las personas en hogares que pertenecen a estratos 1 y 2 (y con ingresos mensuales menores o iguales a un salario mínimo) a participar en la fuerza laboral, ni tampoco incrementan su tasa de ocupación o reducen su probabilidad de estar desempleado. Lo hallado no es consistente con las predicciones teóricas de la literatura de economía urbana y de búsqueda de empleo, más sí están en línea con lo encontrado por Rodríguez et al. (2015), quienes tampoco encuentran un impacto significativo de los subsidios sobre la tasa de participación y la tasa de desempleo.

En cuanto a las limitaciones de la presente investigación, se destaca que no es posible explorar, con la información disponible, las causas de la aparente ineffectividad del subsidio. Tampoco se pueden generalizar los resultados a un grupo más grande de la población, sino que son válidos para las personas con un puntaje del SISBÉN cercano a los 40 puntos, que pertenecen a los estratos 1 y 2 y cuyos ingresos familiares no exceden un salario mínimo. Además, sería idóneo observar el momento en el que los beneficiarios reciben el subsidio, pues se podrían medir los efectos de corto y largo plazo del programa sobre la intensidad de búsqueda de empleo y las otras variables de resultado consideradas.

Los encargados de políticas públicas podrían considerar explorar los obstáculos que no permiten que los subsidios mejoren la accesibilidad a oportunidades laborales de los beneficiarios. En adición, se podría complementar la política con otras que también tengan el

objetivo de reducir las barreras espaciales y de información para acceder al mercado laboral. Por ejemplo, Phillips (2014) encuentra que incorporar subsidios al transporte público en los programas estándar de asistencia para buscar trabajo permite obtener beneficios sustanciales para los grupos más pobres de la población urbana.

Finalmente, la literatura que estudia el fenómeno de desajuste espacial y los efectos de subsidios a la demanda de transporte público en Bogotá podría beneficiarse de investigaciones que exploren la elasticidad-precio de la demanda por transporte público, y más específicamente, la elasticidad-precio cruzada de la demanda de Transporte Público Colectivo y Transmilenio/SITP, pues esto permitiría obtener indicios sobre qué tanto habría que subsidiar la tarifa para que el transporte público sea asequible para la población más pobre, y se sustituya el uso del transporte tradicional con los buses del SITP. Otra área que se podría explorar es cómo estas políticas incrementan el acceso de los hogares segregados espacialmente a otro tipo de oportunidades además de las laborales, como las educativas y las de salud.

## 7 Referencias

- Abebe, G., Caria, S., Fafchamps, M., Falco, P., Franklin, S., y Quinn, S. (2016). Curse of anonymity or tyranny of distance? the impacts of job-search support in urban ethiopia. Technical report, National Bureau of Economic Research.
- Alcaldía de Bogotá (2019). Así quedaron los ajustes a la tarifa del transporte público colectivo en bogotá. <https://bogota.gov.co/mi-ciudad/movilidad/ajustes-la-tarifa-del-transporte-publico>. Accessed: 2019-07-15.
- Alcaldía Mayor de Bogotá (2017a). Análisis del subsidio de transmilenio y sitp focalizado con base en sisbén.
- Alcaldía Mayor de Bogotá (2017b). Encuesta de movilidad 2015.
- Angrist, J. D. y Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Arai, Y. y Ichimura, H. (2016). Optimal bandwidth selection for the fuzzy regression discontinuity estimator. *Economics Letters*, 141:103–106.
- Bocarejo, J. P., Escobar, D., Hernandez, D. O., y Galarza, D. (2016). Accessibility analysis of the integrated transit system of bogotá. *International Journal of Sustainable Transportation*, 10(4):308–320.
- Borjas, G. J. y Van Ours, J. C. (2010). *Labor economics*. McGraw-Hill/Irwin Boston.
- Brueckner, J. K. y Zenou, Y. (2003). Space and unemployment: The labor-market effects of spatial mismatch. *Journal of Labor Economics*, 21(1):242–262.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., y Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2):372–404.
- Calonico, S., Cattaneo, M. D., y Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.
- Cattaneo, M. D., Idrobo, N., y Titiunik, R. (2017). A practical introduction to regression discontinuity designs. *Cambridge Elements: Quantitative and Computational Methods for Social Science-Cambridge University Press I*.
- Cattaneo, M. D., Jansson, M., y Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261.
- Concejo de Bogotá (2018). Sitp provisional continúa pese a las advertencias sobre la urgencia de su desmonte. <http://concejodebogota.gov.co/sitp-provisional-continua-pese-a-las-advertencias-sobre-la-urgencia-de/cbogota/2018-09-14/232805.php>. Accessed: 2019-07-15.
- Departamento Administrativo Nacional de Estadística - DANE (2007). Cartilla de conceptos básicos e indicadores demográficos.
- Departamento Nacional de Planeación (2017). ¿qué es el sisbén? <https://www.sisben.gov.co/sisben/paginas/que-es.aspx>. Accessed: 2019-07-21.

- Díaz Escobar, A. M. (2015). Acceso a fuentes de empleo, segregación residencial y resultados laborales de las mujeres en bogotá. Technical report, IDB Working Paper Series.
- Dirección de Metodología y Producción Estadística (2017). Colombia - encuesta multipropósito - em 2017.
- Domínguez, J. A. (2014). Job accessibility: An empirical approximation for bogotá (colombia).
- El Espectador (2019). Continuará de manera indefinida la operación del sitp provisional. <https://www.elespectador.com/noticias/bogota/continuar-de-manera-indefinida-la-operacion-del-sitp-provisional-articulo-842287>. Accessed: 2019-07-15.
- El Tiempo (2018). Así operan las mafias que revenden los pasajes del sitp y transmilenio. <https://www.eltiempo.com/bogota/mafias-de-venta-de-pasajes-en-el-sitp-y-transmilenio-257496>. Accessed: 2019-07-14.
- Fan, J., Gijbels, I., Hu, T.-C., y Huang, L.-S. (1996). A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica*, pages 113–127.
- Franklin, S. et al. (2015). Location, search costs and youth unemployment: A randomized trial of transport subsidies in ethiopia. *Centre for the Study of African Economies Working Paper WPS/2015-11*.
- Gelman, A. y Imbens, G. (2018). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, pages 1–10.
- Gobillon, L., Selod, H., y Zenou, Y. (2007). The mechanisms of spatial mismatch. *Urban studies*, 44(12):2401–2427.
- Guzman, L., Oviedo, D., y Cardona, R. (2018). Accessibility changes: Analysis of the integrated public transport system of bogotá. *Sustainability*, 10(11):3958.
- Guzman, L. A. y Oviedo, D. (2018). Accessibility, affordability and equity: Assessing ‘pro-poor’ public transport subsidies in bogotá. *Transport Policy*, 68:37–51.
- Hernández, C. R. y Peralta-Quiros, T. (2016). Balancing financial sustainability and affordability in public transport: the case of bogotá, colombia. International Transport Forum Discussion Paper.
- Holderness, C. G. et al. (2016). Problems using aggregate data to infer individual behavior: Evidence from law, finance, and ownership concentration. *Critical Finance Review*, 5(1):1–40.
- Ihlanfeldt, K. R. (2006). A primer on spatial mismatch within urban labor markets.
- Jaramillo, A. y Rengifo, C. (2018). Impacto del sistema metroplús sobre el mercado laboral de las comunas manrique y aranquez de medellín, colombia. *Lecturas de Economía*, (89):133–161.

- Jin, J. y Paulsen, K. (2018). Does accessibility matter? understanding the effect of job accessibility on labour market outcomes. *Urban studies*, 55(1):91–115.
- Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The quarterly journal of economics*, 82(2):175–197.
- Kash, G. (2019). Always on the defensive: the effects of transit sexual assault on travel behavior and experience in colombia and bolivia. *Journal of Transport & Health*, 13:234–246.
- Lee, D. S. y Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355.
- Matas, A., Raymond, J.-L., y Roig, J.-L. (2010). Job accessibility and female employment probability: the cases of barcelona and madrid. *Urban Studies*, 47(4):769–787.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714.
- Phillips, D. C. (2014). Getting to work: Experimental evidence on job search and transportation costs. *Labour Economics*, 29:72–82.
- Pissarides, C. A. (2000). *Equilibrium unemployment theory*. MIT press.
- Publimetro (2018). Así funciona el negocio de la reventa de pasajes en transmilenio y el sitp. <https://www.publimetro.co/co/noticias/2018/08/01/reventa-de-pasajes-transmilenio-y-sitp.html>. Accessed: 2019-07-15.
- RCN Radio (2018). Reventa de pasajes del sitp, otro problema para transmilenio. <https://www.rcnradio.com/bogota/reventa-de-pasajes-del-sitp-otro-problema-para-transmilenio-0>. Accessed: 2019-07-15.
- Rodriguez, C., Gallego, J. M., Martinez, D., Montoya, S., y Peralta-Quiros, T. (2015). Examining the implementation and labor market outcomes of targeted transit subsidies: Sisben subsidy for bogota’s urban poor. *Transportation Research Record Paper*, pages 16–4349.
- Smith, T. E. y Zenou, Y. (2003). Spatial mismatch, search effort, and urban spatial structure. *Journal of Urban Economics*, 54(1):129–156.
- Stock, J. H. y Yogo, M. (2005). Testing for weak instruments in linear iv regression, in dwk andrews and jh stock, eds., *identification and inference for econometric models: Essays in honor of thomas j. rothenberg*. cambridge: Cambridge university press.
- Transmilenio S.A. (2018). Beneficios de transporte para personas sisbenizadas. [https://www.sitp.gov.co/publicaciones/40906/beneficios\\_de\\_transporte\\_para\\_personas\\_sisbenizadas/](https://www.sitp.gov.co/publicaciones/40906/beneficios_de_transporte_para_personas_sisbenizadas/). Accessed: 2019-07-19.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

# Apéndice A Pruebas de sensibilidad de los resultados a observaciones cercanas al umbral

## Apéndice A.1 Excluyendo observaciones a una distancia menor a 1 punto del umbral

Tabla 14: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de participación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.415 (0.280)	-0.446 (0.342)	0.143 (0.240)	0.230 (0.268)
Intervalo robusto LATE 95%	[-1.111 .183]	[-1.18 .273]	[-.406 .768]	[-.342 .812]
Ancho de banda $h$	7.779	11.547	10.188	13.469
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	353	516	516	763
Kernel	Triangular	Triangular	Triangular	Triangular
Coficiente primera etapa	-.408	-.439	-.397	-.483
Intervalo robusto primera etapa	[-.658 -.239]	[-.709 -.222]	[-.624 -.292]	[-.713 -.319]

1 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 15: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de ocupación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.476* (0.274)	-0.446 (0.375)	-0.053 (0.235)	-0.024 (0.258)
Intervalo robusto LATE 95%	[-1.166 .104]	[-1.302 .331]	[-.636 .507]	[-.596 .521]
Ancho de banda $h$	9.700	10.619	10.67	14.22
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	430	475	475	826
Kernel	Triangular	Triangular	Triangular	Triangular
Coficiente primera etapa	-.38	-.437	-.392	-.479
Intervalo robusto primera etapa 95%	[-.611 -.234]	[-.72 -.193]	[-.613 -.292]	[-.704 -.322]

1 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 16: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de desempleo

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	0.140 (0.173)	0.045 (0.300)	0.358 (0.253)	0.356 (0.279)
Intervalo robusto LATE 95%	[-.216 .551]	[-.609 .714]	[-.11 1.013]	[-.24 .97]
Ancho de banda $h$	7.958	10.537	11.164	10.774
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	270	362	362	307
Kernel	Triangular	Triangular	Triangular	Triangular
Coficiente primera etapa	-.373	-.337	-.364	-.535
Intervalo robusto primera etapa 95%	[-.618 -.172]	[-.642 -.045]	[-.59 -.176]	[-.897 -.234]

1 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

## Apéndice A.2 Excluyendo observaciones a una distancia menor a 2 puntos del umbral

Tabla 17: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de participación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.237 (0.283)	-0.458 (0.454)	-0.022 (0.466)	0.005 (0.486)
Intervalo robusto LATE 95%	[-.978 .406]	[-1.61 .527]	[-1.144 1.146]	[-1.113 1.114]
Ancho de banda $h$	12.509	10.287	10.347	17.56
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	516	428	428	937
Kernel	Triangular	Triangular	Triangular	Triangular
Coficiente primera etapa	-.356	-.579	-.257	-.3
Intervalo robusto primera etapa 95%	[-.614 -.183]	[-1.118 -.123]	[-.492 -.067]	[-.560 -.086]

2 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 18: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de ocupación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.450 (0.312)	-0.519 (0.480)	-0.182 (0.498)	-0.153 (0.489)
Intervalo robusto LATE 95%	[-1.298 .252]	[-1.691 .573]	[-1.455 1.008]	[-1.375 .922]
Ancho de banda h	12.589	10.299	9.749	16.797
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	516	428	428	876
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.355	-.577	-.255	-.313
Intervalo robusto primera etapa 95%	[-.621 -.187]	[-1.116 -.125]	[-.51 -.062]	[-.603 -.095]

2 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 19: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de desempleo

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.101 (0.253)	0.126 (0.203)	0.227 (0.769)	0.477 (0.859)
Intervalo robusto LATE 95%	[-.566 .52]	[-.316 .621]	[-1.442 2.267]	[-1.255 2.566]
Ancho de banda h	5.165	10.595	7.939	16.214
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	172	329	329	442
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.415	-.648	-.206	-.214
Intervalo robusto primera etapa 95%	[-.98 .031]	[-1.252 -.17]	[-.577 .169]	[-.598 .123]

2 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

## Apéndice A.3 Excluyendo observaciones a una distancia menor a 3 puntos del umbral

Tabla 20: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de participación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.150 (0.449)	-0.015 (0.476)	-0.512 (0.626)	-0.524 (0.660)
Intervalo robusto LATE 95%	[-1.229 .906]	[-1.155 1.062]	[-2.168 .99]	[-2.064 .894]
Ancho de banda h	10.911	18.045	10.845	17.544
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	346	700	700	845
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.352	-.432	-.281	-.364
Intervalo robusto primera etapa 95%	[-.688 -.054]	[-.89 -.08]	[-.586 -.011]	[-.707 -.027]

3 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 21: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de ocupación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.473 (0.526)	-0.596 (0.564)	-0.708 (0.772)	-0.522 (0.649)
Intervalo robusto LATE 95%	[-1.88 .832]	[-1.912 .761]	[-2.793 1.16]	[-2.112 .810]
Ancho de banda h	11.006	16.849	9.827	17.274
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	387	630	630	845
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.351	-.494	-.275	-.377
Intervalo robusto primera etapa 95%	[-.758 -.079]	[-1.039 -.145]	[-.664 .015]	[-.735 -.036]

3 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 22: Sensibilidad de los resultados a observaciones cercanas al umbral: Tasa de desempleo

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	0.244 (0.218)	0.372 (0.279)	-0.337 (0.663)	0.754 (2.563)
Intervalo robusto LATE 95%	[-.188 .784]	[-.255 1.036]	[-1.847 1.467]	[-5.214 6.726]
Ancho de banda $h$	6.889	13.874	7.536	13.214
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	140	381	381	310
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-1.084	-.897	-.36	-.155
Intervalo robusto primera etapa 95%	[-1.885 -.544]	[-1.739 -.354]	[-1.075 .302]	[-1.019 .61]

3 del umbral

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

## Apéndice B Pruebas de sensibilidad de los resultados a la elección de la función kernel

### Apéndice B.1 Kernel Epanechnikov

Tabla 23: Regresión Discontinua Borrosa: Tasa de participación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.411 (0.256)	-0.401 (0.262)	0.036 (0.225)	0.060 (0.253)
Intervalo robusto LATE 95%	[-1.261 .076]	[-1.062 .203]	[-.574 .589]	[-.634 .671]
Ancho de banda $h$	10.247	20.859	13.013	20.454
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	544	982	853	1250
Kernel	Epanechnikov	Epanechnikov	Epanechnikov	Epanechnikov
Coefficiente primera etapa	-.325	-.325	-.307	-.319
Intervalo robusto primera etapa 95%	[-.560 -.145]	[-.592 -.202]	[-.546 -.229]	[-.62 -.244]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 24: Regresión Discontinua Borrosa: Tasa de ocupación

	(1) Hombres - Lineal	(2) Hombres - Cuadrático	(3) Mujeres - Lineal	(4) Mujeres - Cuadrático
LATE	-0.331 (0.247)	-0.343 (0.277)	-0.063 (0.225)	-0.112 (0.248)
Intervalo robusto LATE 95%	[-1.02 .261]	[-.964 .367]	[-.718 .485]	[-.706 .599]
Ancho de banda h	11.502	21.845	12.735	18.614
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	585	1012	792	1138
Kernel	Epanechnikov	Epanechnikov	Epanechnikov	Epanechnikov
Coefficiente primera etapa	-.32	-.307	-.308	-.346
Intervalo robusto primera etapa 95%	[-.568 -.186]	[-.579 -.202]	[-.554 -.224]	[-.659 -.254]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 25: Regresión Discontinua Borrosa: Tasa de desempleo

	(1) Hombres - Lineal	(2) Hombres - Cuadrático	(3) Mujeres - Lineal	(4) Mujeres - Cuadrático
LATE	0.032 (0.161)	-0.040 (0.182)	0.290 (0.283)	0.289 (0.313)
Intervalo robusto LATE 95%	[-.388 .416]	[-.512 .345]	[-.369 1.038]	[-.529 1.005]
Ancho de banda h	12.018	18.537	10.58	14.056
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	488	699	352	463
Kernel	Epanechnikov	Epanechnikov	Epanechnikov	Epanechnikov
Coefficiente primera etapa	-.303	-.325	-.296	-.353
Intervalo robusto primera etapa 95%	[-.559 -.147]	[-.59 -.126]	[-.578 -.1]	[-.700 -.087]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

## Apéndice B.2 Kernel Uniforme

Tabla 26: Regresión Discontinua Borrosa: Tasa de participación

	(1) Hombres - Lineal	(2) Hombres - Cuadrático	(3) Mujeres - Lineal	(4) Mujeres - Cuadrático
LATE	-0.265 (0.244)	-0.327 (0.278)	0.049 (0.242)	0.132 (0.288)
Intervalo robusto LATE 95%	[-.98 .205]	[-1.116 .27]	[-.547 .666]	[-.658 .847]
Ancho de banda $h$	13.049	23.342	13.724	21.33
Tipo de ancho de banda	<i>mserd</i>	<i>mserd</i>	<i>mserd</i>	<i>mserd</i>
Número eficiente de observaciones	683	1081	853	1287
Kernel	Uniform	Uniform	Uniform	Uniform
Coefficiente primera etapa	-.262	-.265	-.263	-.261
Intervalo robusto primera etapa 95%	[-.432 -.136]	[-.519 -.164]	[-.43 -.158]	[-.498 -.15]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 27: Regresión Discontinua Borrosa: Tasa de ocupación

	(1) Hombres - Lineal	(2) Hombres - Cuadrático	(3) Mujeres - Lineal	(4) Mujeres - Cuadrático
LATE	-0.187 (0.276)	-0.377 (0.298)	-0.034 (0.241)	0.077 (0.281)
Intervalo robusto LATE 95%	[-.927 .544]	[-1.104 .346]	[-.796 .47]	[-.904 .517]
Ancho de banda $h$	14.176	22.895	13.359	25.503
Tipo de ancho de banda	<i>mserd</i>	<i>mserd</i>	<i>mserd</i>	<i>mserd</i>
Número eficiente de observaciones	726	1040	853	1448
Kernel	Uniform	Uniform	Uniform	Uniform
Coefficiente primera etapa	-.235	-.273	-.263	-.244
Intervalo robusto primera etapa 95%	[-.448 -.13]	[-.518 -.163]	[-.446 -.16]	[-.416 -.109]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 28: Regresión Discontinua Borrosa: Tasa de desempleo

	(1) Hombres - Lineal	(2) Hombres - Cuadrático	(3) Mujeres - Lineal	(4) Mujeres - Cuadrático
LATE	0.049 (0.172)	0.032 (0.226)	-0.042 (0.217)	0.308 (0.276)
Intervalo robusto LATE 95%	[-.439 .386]	[-.52 .535]	[-.566 .48]	[-.202 1.026]
Ancho de banda $h$	12.699	21.963	17.727	19.432
Tipo de ancho de banda	<i>mserd</i>	<i>mserd</i>	<i>mserd</i>	<i>mserd</i>
Número eficiente de observaciones	488	781	552	590
Kernel	Uniform	Uniform	Uniform	Uniform
Coefficiente primera etapa	-.274	-.242	-.248	-.317
Intervalo robusto primera etapa 95%	[-.498 -.131]	[-.461 -.066]	[-.36 -.046]	[-.58 -.139]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

## Apéndice C Estimación de Variable Instrumental con polinomio cuadrático de la variable de asignación - Encuesta de Movilidad de 2015

Tabla 29: Tasa de participación

	(1) IV Hombres	(2) IV Mujeres
Tarjeta de Sisbén/SITP	-0.023 (0.116)	-0.046 (0.133)
Constante	-0.646*** (0.111)	-0.368*** (0.128)
Observaciones	1,464	1,812
Dummies de localidad	Sí	Sí
Estadístico F	22.482	12.652

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (1) y (2) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Estado civil*, *Ingresos no laborales*, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 30: Tasa de ocupación

	(1)	(2)
	IV Hombres	IV Mujeres
Tarjeta de Sisbén/SITP	-0.060 (0.132)	-0.105 (0.133)
Constante	-0.690*** (0.127)	-0.399*** (0.126)
Observaciones	1,464	1,812
Dummies de localidad	Sí	Sí
Estadístico F	24.726	9.76

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (1) y (2) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Estado civil*, *Ingresos no laborales*, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 31: Tasa de desempleo

	(1)	(2)
	IV Hombres	IV Mujeres
Tarjeta de Sisbén/SITP	0.016 (0.105)	0.138 (0.113)
Constante	0.421*** (0.136)	0.493*** (0.136)
Observaciones	1,155	916
Dummies de localidad	Sí	Sí
Estadístico F	3.3	2.845

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (1) y (2) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Estado civil*, *Ingresos no laborales*, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 32: Primera etapa: Tasa de participación/ Tasa de ocupación

	(1)	(2)
	Primera etapa - Hombres	Primera etapa - Mujeres
Puntaje Sisbén menor a 40 puntos	0.269*** (0.030)	0.295*** (0.026)
Constante	0.385*** (0.121)	0.348*** (0.101)
Observaciones	1,464	1,812
R-cuadrado	0.127	0.140
Dummies de localidad	Sí	Sí
F de instrumentos excluidos	81.014	128.621
p-value F	0	0
Estadístico F de KP - Identificación débil	81.014	128.621
p-value prueba de AR - Inferencia robusta a instrumentos débiles	.843	.733

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. La prueba de AR - Inferencia robusta a instrumentos débiles tiene como hipótesis nula que los coeficientes de las variables endógenas en la ecuación estructural son en conjunto iguales a 0. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 33: Primera etapa: Tasa de desempleo

	(1)	(2)
	Primera etapa - Hombres	Primera etapa - Mujeres
Puntaje Sisbén menor a 40 puntos	0.274*** (0.034)	0.336*** (0.039)
Constante	0.388*** (0.149)	0.232 (0.161)
Observaciones	1,155	916
R-cuadrado	0.129	0.157
Dummies de localidad	Sí	Sí
F de instrumentos excluidos	64.83	72.942
p-value F	0	0
Estadístico F de KP - Identificación débil	64.83	72.942
p-value prueba de AR - Inferencia robusta a instrumentos débiles	.883	.226

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. La prueba de AR - Inferencia robusta a instrumentos débiles tiene como hipótesis nula que los coeficientes de las variables endógenas en la ecuación estructural son en conjunto iguales a 0. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

# Apéndice D Resultados con la Encuesta Multipropósito de Bogotá de 2017

## Apéndice D.1 Estimaciones

Tabla 34: Tasa de participación

	(1)	(2)	(3)	(4)
	OLS Hombres	OLS Mujeres	IV Hombres	IV Mujeres
Tarjeta Sisbén/SITP	-0.124** (0.050)	-0.017 (0.039)	-2.430 (1.804)	2.445 (2.181)
Constante	0.231** (0.098)	0.189** (0.088)	0.295** (0.120)	0.083 (0.142)
Observaciones	4,521	6,411	4,521	6,411
R-cuadrado	0.157	0.194		
Dummies de UPZ	Sí	Sí	Sí	Sí
Estadístico F			64.182	215.243

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (3) y (4) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Estado civil*, *Ingresos no laborales*, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 35: Tasa de ocupación

	(1)	(2)	(3)	(4)
	OLS Hombres	OLS Mujeres	IV Hombres	IV Mujeres
Tarjeta SITP/Sisbén	-0.040 (0.049)	-0.010 (0.038)	-2.385 (1.970)	4.172 (2.853)
Constante	-0.091 (0.102)	-0.304*** (0.078)	-0.037 (0.128)	-0.473*** (0.180)
Observaciones	4,521	6,411	4,521	6,411
R-cuadrado	0.131	0.129		
Dummies de UPZ	Sí	Sí	Sí	Sí
Estadístico F			118.673	98.135

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (3) y (4) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Estado civil*, *Ingresos no laborales*, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 36: Tasa de desempleo

	(1)	(2)	(3)	(4)
	OLS Hombres	OLS Mujeres	IV Hombres	IV Mujeres
Tarjeta Sisbén/SITP	-0.036 (0.063)	-0.027 (0.061)	1.185 (2.179)	-6.314 (4.080)
Constante	0.651*** (0.115)	1.231*** (0.142)	0.657*** (0.119)	1.425*** (0.380)
Observaciones	3,214	2,569	3,214	2,569
R-cuadrado	0.141	0.172		
Dummies de UPZ	Sí	Sí	Sí	Sí
Estadístico F			12019.378	257.993

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. Las columnas (3) y (4) muestran los resultados de la segunda etapa de la estimación de MCO en dos etapas. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Estado civil*, *Ingresos no laborales*, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 37: Primera etapa: Tasa de participación/ Tasa de ocupación

	(1)	(2)
	Primera etapa - Hombres	Primera etapa - Mujeres
Puntaje Sisbén menor a 40 puntos	0.035*	0.027*
	(0.020)	(0.016)
Constante	-0.006	0.019
	(0.031)	(0.032)
Observaciones	4,521	6,411
R-cuadrado	0.026	0.030
Dummies de UPZ	Sí	Sí
F de instrumentos excluidos	3.097	3.013
P-value F	.079	.083
Estadístico F de KP - Identificación débil	3.097	3.013
P-value prueba de AR - Inferencia robusta a instrumentos débiles	.061	.105

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. La prueba de AR - Inferencia robusta a instrumentos débiles tiene como hipótesis nula que los coeficientes de las variables endógenas en la ecuación estructural son en conjunto iguales a 0. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 38: Primera etapa: Tasa de desempleo

	(1)	(2)
	Primera etapa - Hombres	Primera etapa - Mujeres
Puntaje Sisbén menor a 40 puntos	0.028	0.028*
	(0.020)	(0.015)
Constante	-0.024	0.012
	(0.030)	(0.055)
Observaciones	3,214	2,569
R-cuadrado	0.029	0.049
Dummies de UPZ	Sí	Sí
F de instrumentos excluidos	1.98	3.661
P-value F	.16	.056
Estadístico F de KP - Identificación débil	1.98	3.661
P-value prueba de AR - Inferencia robusta a instrumentos débiles	.560	.013

\*, \*\*, y \*\*\* denotan significancia estadística a niveles del 10%, 5 %, y 1% , respectivamente. Errores estándar robustos a heteroscedasticidad en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. La prueba de AR - Inferencia robusta a instrumentos débiles tiene como hipótesis nula que los coeficientes de las variables endógenas en la ecuación estructural son en conjunto iguales a 0. Todos los modelos estimados incluyen las siguientes variables de control: *Edad*, *Edad*<sup>2</sup>, *Nivel educativo*, *Puntaje del SISBÉN centrado en 40*, *Presencia de dependientes en el hogar* y *Número de personas en el hogar*.

Tabla 39: Regresión Discontinua Borrosa: Tasa de participación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-0.719 (4.116)	0.064 (2.288)	2.239 (3.049)	0.516 (2.372)
Intervalo robusto LATE 95%	[-10.648 8.093]	[-5.13 4.753]	[-4.959 9.153]	[-4.503 5.317]
Ancho de banda h	4.285	10.426	7.611	8.927
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	250	592	612	735
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.031	-.054	.04	.052
Intervalo robusto primera etapa 95%	[-.106 .027]	[-.141 .007]	[-.035 .12]	[-.035 .144]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

Tabla 40: Regresión Discontinua Borrosa: Tasa de ocupación

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	-9.157 (13.448)	-2.125 (2.932)	2.850 (3.361)	0.757 (2.196)
Intervalo robusto LATE 95%	[-46.207 11.55]	[-8.474 4.095]	[-4.802 10.446]	[-3.932 5.169]
Ancho de banda h	3.677	10.67	9.623	9.308
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	201	607	803	764
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.023	-.051	.033	.052
Intervalo robusto primera etapa 95%	[-.074 .047]	[-.138 .01]	[-.037 .098]	[-.034 .142]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

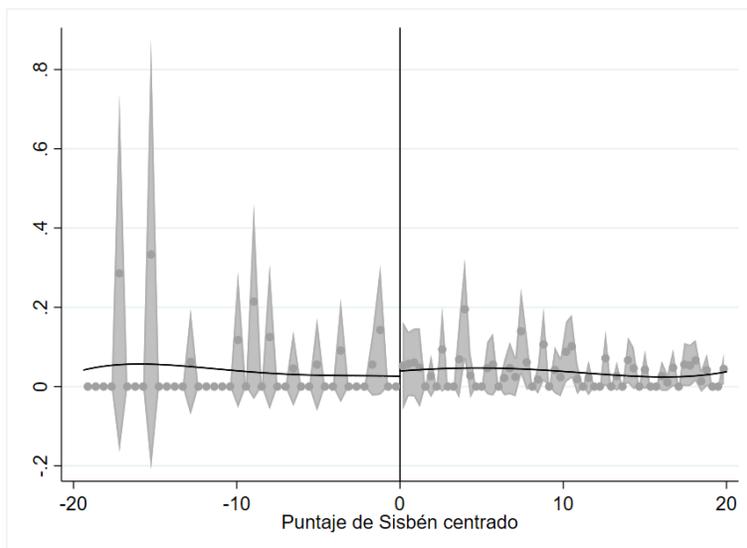
Tabla 41: Regresión Discontinua Borrosa: Tasa de desempleo

	(1)	(2)	(3)	(4)
	Hombres - Lineal	Hombres - Cuadrático	Mujeres - Lineal	Mujeres - Cuadrático
LATE	10.234 (10.213)	6.524 (5.827)	-0.691 (3.074)	-0.355 (3.299)
Intervalo robusto LATE 95%	[-10.132 43.296]	[-5.796 18.448]	[-7.511 7.119]	[-7.229 6.646]
Ancho de banda $h$	2.946	7.881	5.725	7.711
Tipo de ancho de banda	mserd	mserd	mserd	mserd
Número eficiente de observaciones	119	306	185	252
Kernel	Triangular	Triangular	Triangular	Triangular
Coefficiente primera etapa	-.025	-.041	.055	.063
Intervalo robusto primera etapa 95%	[-.05 .031]	[-.104 .007]	[-.029 .153]	[-.024 .16]

Errores estándar en paréntesis. *Observaciones* hace referencia al número de individuos incluidos en la estimación. LATE se refiere a Efecto Promedio Local del Tratamiento. El número eficiente de observaciones es el que resulta de utilizar para la estimación solo aquellas que están dentro del rango definido por el ancho de banda. *Robusto* se refiere a una corrección del estimador puntual y los errores estándar para controlar por la escogencia de anchos de banda demasiado amplios. El tipo de ancho de banda *mserd* se refiere a una forma de seleccionar el ancho de banda para la estimación que minimiza el Error Cuadrático Medio (ECM) y escoge un  $h$  común para ambos lados del umbral.

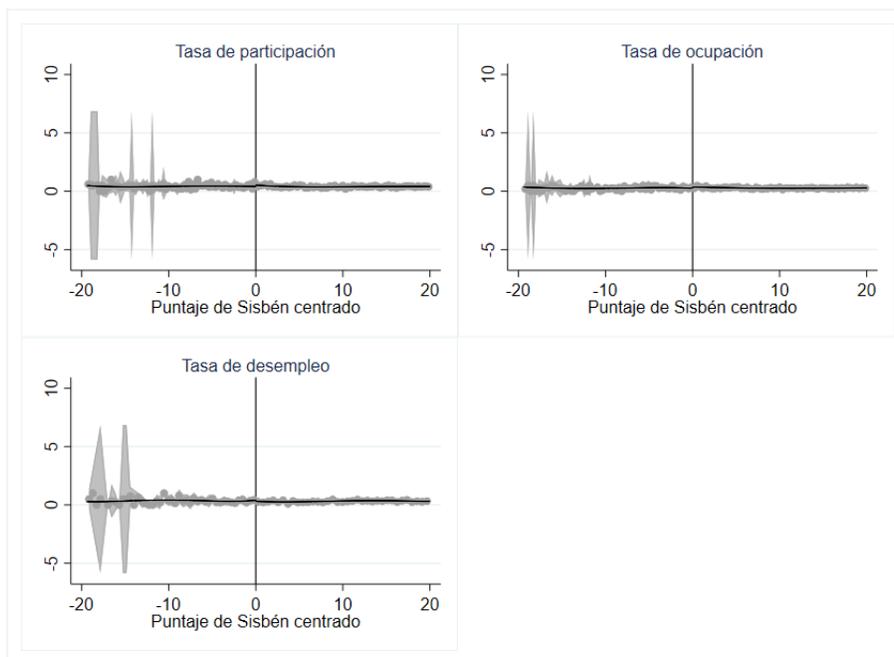
## Apéndice D.2 Gráficas de resultados

Figura B.1: Salto en la probabilidad de recibir tratamiento en el umbral



Fuente: Elaboración propia. Se ajustan polinomios de cuarto grado a cada lado del umbral para obtener una aproximación global de la función de regresión.

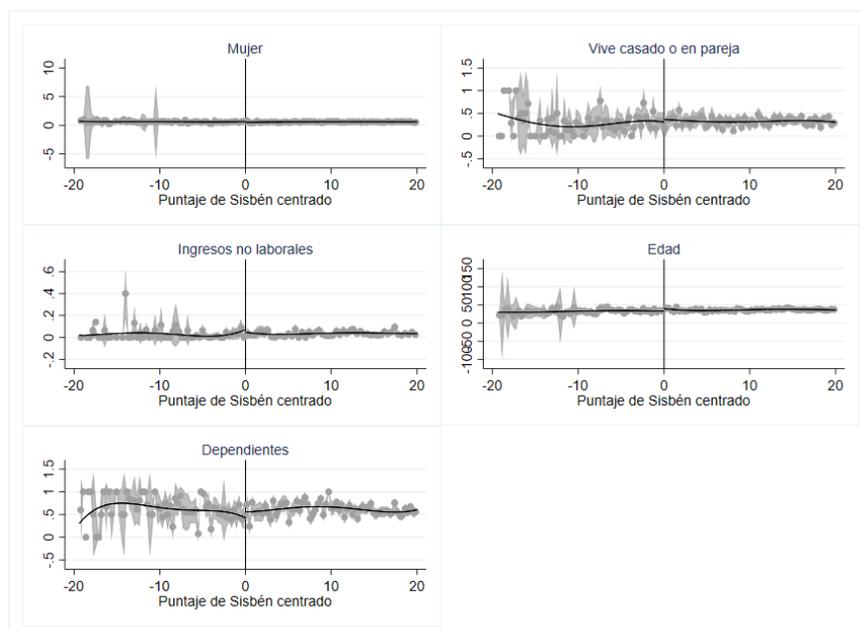
Figura B.2: Salto en los indicadores del mercado laboral en el umbral



Fuente: Elaboración propia. Se ajustan polinomios de cuarto grado a cada lado del umbral para obtener una aproximación global de las funciones de regresión.

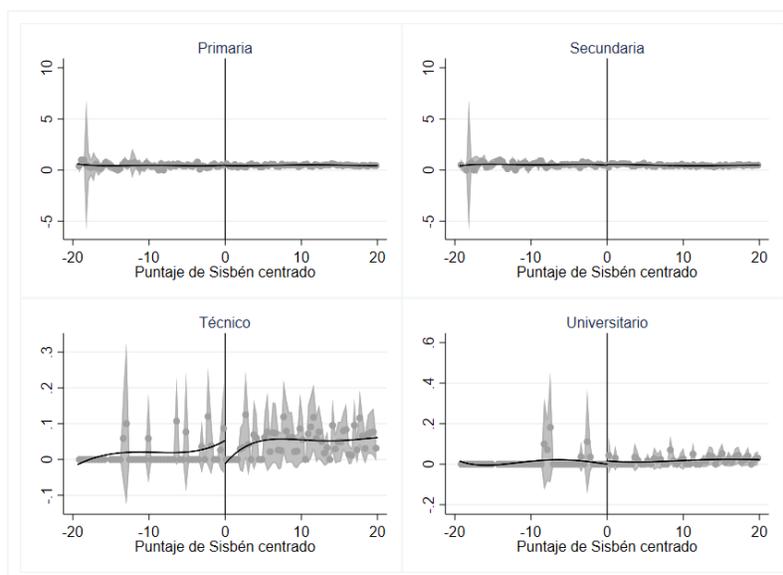
## Apéndice D.3 Gráficas de pruebas de sensibilidad y robustez

Figura B.3: Continuidad de las variables predeterminadas alrededor del umbral



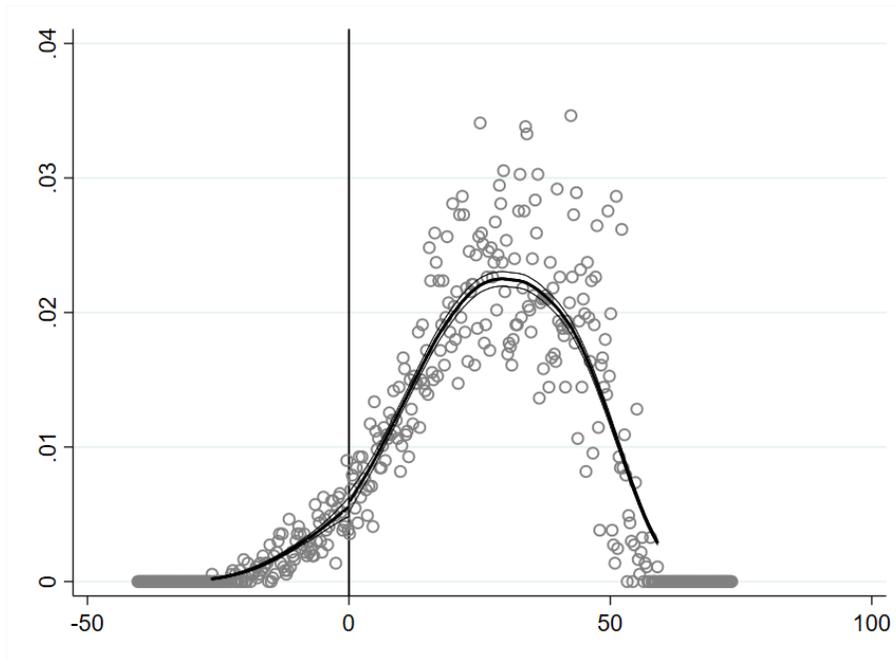
Fuente: Elaboración propia. Se ajustan polinomios de cuarto grado a cada lado del umbral para obtener una aproximación global de la función de regresión.

Figura B.4: Continuidad de las variables predeterminadas alrededor del umbral



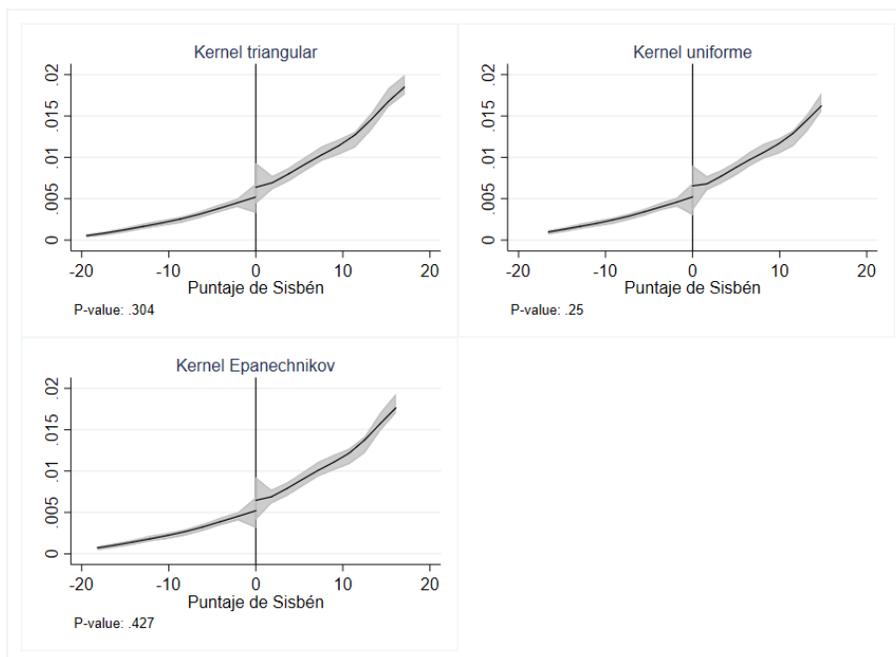
Fuente: Elaboración propia. Se ajustan polinomios de cuarto grado a cada lado del umbral para obtener una aproximación global de la función de regresión.

Figura B.5: Prueba de manipulación: McCrary



Fuente: Elaboración propia con base en McCrary (2008).

Figura B.6: Prueba de manipulación: rddensity



Fuente: Elaboración propia con base en Cattaneo et al. (2018).