

EVALUACIÓN DEL IMPACTO SOCIO ECONÓMICO DE LA REHABILITACIÓN DEL FERROCARRIL EN EL ECUADOR MEDIANTE EL MÉTODO DE DIFERENCIAS EN DIFERENCIAS

*Alfredo Verdezoto Noboa**

Resumen:

Este trabajo busca medir el impacto socioeconómico del Proyecto de Modernización de la Empresa de Ferrocarriles del Ecuador, en las poblaciones que viven alrededor de las estaciones ferroviarias. Para calcular dicho impacto, se aplicó un modelo de diferencias en diferencias, considerando como población de tratamiento a las personas que viven con mayor proximidad a las estaciones y como grupo de control a quienes no están dentro del área de influencia de este negocio. Los resultados muestran que la población de tratamiento del 2010 al 2012, en relación con la población de control, pudo incrementar sus ingresos gracias al turismo atraído por la rehabilitación en un promedio del 8.66%.

Palabras clave: evaluación de políticas públicas, impacto socio económico, diferencias en diferencias.

CÓDIGO JEL: C01, O10.

Abstract:

In this study, we seek to measure the socio-economic impact of the Modernization Railway Enterprise of the Ecuador Project on population who live around ñrailway stations. In order to measure this impact, we applied a difference in difference model, where the treated population are those who live near to stations and another group are those who live apart from the stations or are not in the treated group. The results show that the treated population from 2010 to 2012, in relation with control

* Economista por la Pontificia Universidad Católica del Ecuador y Máster en Desarrollo Económico por la Universidad Carlos III de Madrid, España (Beca Fundación Carolina), se desempeña como Especialista en Estudios Socioeconómicos en el ámbito de las drogas en la Secretaría Técnica de Drogas.

population, their income increased thanks to tourism with an average rate of in 8.66%.

Keys words: public policies assessment, social economic impact, difference and difference model.

JEL CODE: C01, O10.

I. INTRODUCCIÓN

Demostrar que un proyecto ferroviario ofrece más que transporte público y que podría mejorar el bienestar socio económico de las personas, es justamente el aporte y la novedad que presenta ésta investigación. Partiendo del hecho que la inversión pública en infraestructura mejora los niveles de vida de la sociedad, tomamos como referencia al proceso de rehabilitación de la línea férrea entre los años 2010 y 2012¹, pues este proyecto dinamizó la economía de las zonas por donde pasa el tren, especialmente gracias al incremento de turismo. El presente estudio tiene como objetivo principal demostrar, a través de metodologías econométricas, que efectivamente el proyecto ha tenido un impacto positivo en la población.

Es importante señalar, como antecedente, el estudio de evaluación de impacto al proyecto denominado “Alta velocidad ferroviaria e integración metropolitana en España: el caso de Ciudad Real y Puertollano” (Universidad de Castilla la Mancha, 2003). Se observó, a través de estadísticas descriptivas, cómo los trenes de pasajeros de alta velocidad influyen en los hábitos de movilidad de los usuarios y el papel de la línea férrea en el proceso de integración de estas ciudades con el área metropolitana de Madrid.

Otro caso semejante es el impacto económico en la comunidad valenciana, gracias a la construcción de un corredor ferroviario (Asociación Valenciana de Empresarios, 2011). Utilizando como metodología la matriz insumo-producto de Leontief, se analizó el aporte económico de trenes de carga y pasajeros en otros sectores productivos. En comparación a las investigaciones anteriores, en las que se enfatiza en datos descriptivos y evaluación de la demanda de los servicios ferroviarios, este trabajo se concentra en aspectos socioeconómicos, para lo cual se toma como variable de interés al ingreso, condicionada variables como la edad, nivel educativo, estado civil, zonas donde desempeñan sus funciones (norte, sur y centro), vinculación o no al negocio turístico y si recibe o no el Bono de Desarrollo Humano, el cual es una subvención que el Estado ecuatoriano da a las personas de bajos recursos.

El estudio contó con las siguientes fuentes: en el 2010 Ferrocarriles del Ecuador aplicó una encuesta con una muestra representativa de 4,148 casos; para el 2012 la Empresa nuevamente, aplicando el mismo formulario de encuesta que en el 2010, encuesta a 4,085 personas en las mismas zonas.

1 Latsany Phakdisoth and Donghun Kim, (2007) muestran un estudio realizado en Laos, donde se utilizaron datos de panel para medir el flujo de la demanda de turismo, se concluyó que la inversión en infraestructura turística e inversión en transporte, incrementó el flujo de visitantes a ese país.

Por otro lado, la fuente *secundaria* fue obtenida del Instituto Nacional de Estadísticas y Censos del Ecuador (INEC), específicamente de los datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo -ENEMDU- para el tercer trimestre del 2010. Esta investigación se la realiza a nivel nacional y coincide con la misma fecha en que se levantó la encuesta para la población de tratamiento, cubre la extensión del territorio nacional ecuatoriano y, a través de un ejercicio de identificación de los datos útiles para la presente investigación, se llegaron a filtrar 5,428 datos.

Respecto a la metodología, aplica el modelo de diferencias en diferencias, pues por la naturaleza y estructura de los datos disponibles resulta la manera más idónea de medir el impacto socioeconómico de este proyecto. Pudo haber sido posible aplicar un método de estimación del impacto mediante Variables Instrumentales. Sin embargo carecemos de una Variable Instrumental adecuada que nos ayude a medir los beneficios en términos de impacto. Por otro lado, se calculó el índice de Gini y Atkinson, útil para analizar la desigualdad en las zonas de tratados y no tratados del proyecto. Los resultados obtenidos indican que hay un impacto positivo en ingresos para los beneficiarios del proyecto de un año a otro equivalente a 8.66%.

II. BREVE HISTORIA DEL FERROCARRIL

Desde su inauguración el 25 de junio de 1908, el ferrocarril ecuatoriano fue construido con la visión de ser un símbolo de unidad nacional, convirtiéndose en eje fundamental para el desarrollo del Ecuador, siendo el responsable de la prestación del servicio de transporte masivo de pasajeros y de carga por vía férrea, abrió las fronteras comerciales de este país a través del incremento de las exportaciones y posibilitó que, por primera vez, se pudiera comercializar internamente entre la región montañosa andina y la costera. Sin embargo, ya bien entrado el siglo XX, las deficientes administraciones de la empresa de ferrocarriles y los gobiernos de turno, quienes consideraron al transporte de carga pesada por vía terrestre como más viable, provocaron un progresivo deterioro del modelo de transporte ferroviario paralelo al de un desarrollo relativo del servicio de transporte por carretera.

En 1998 se produjo un leve intento por rescatar al sistema férreo, se adquirieron equipos tractivos y remolcados nuevos (los primeros son las locomotoras y los segundos los vagones) importados de Inglaterra, mas no se tuvo el éxito esperado, pues este proceso no se realizó con una planificación a largo plazo. La tendencia estadística de la participación de los ferrocarriles como modo alternativo para transporte de pasajeros y de carga, se aprecia en la tabla siguiente:

Tabla 1: Número de Pasajeros Transportados, tonelaje arrastrado y tasas de crecimiento a nivel nacional				
Años	Pasajeros Transportados (personas)	Tasa de crecimiento (%)	Tonelaje Arrastrado (kilos)	Tasa de crecimiento (%)
1992	1,168.904	25.91	25.823	38.63
1993	2,619,306	124.08	33,936	31.42
1994	2,899,214	10.69	49,207	45.00
1995	265,528	-90.84	34,323	-30.25
1996	197,855	-25.49	13,799	-59.80
1997	168,440	-14.87	13,308	-3.56
1998	88,881	-47.83	3,122	-76.54
1999	94,029	5.79	1,978	-36.64
2000	117,484	24.94	208	-89.48
2001	77,659	-33.90	808	288.46
2002	113,112	45.65	662	-18.07
2003	70,093	-38.03	333.8	-49.56
2004	44,233	-36.69	428,76	28.51
2005	63,117	42.69	62.12	-85.51
2006	65,142	3.21	16.23	-73.87
2007	40,418	-37.95	9.31	-42.64
2008	36,577	-9.50	0.00	----

Fuente: Estadísticas de Transporte en el Ecuador – 2007, Ministerio de Transporte y Obras Públicas. Estadísticas de Ferrocarriles del Ecuador Empresa Pública

Como podemos observar en la Tabla 1, en 1993 hay un notable crecimiento del transporte de personas (124.08%), como resultado de las maquinarias nuevas adquiridas dos años atrás; no sucede lo mismo con el transporte de carga que apenas alcanza un 31.42%; en cambio, desde 1995, existe una progresiva disminución de pasajeros transportados y tonelaje de carga arrastrado, puesto que quienes se encontraban en el gobierno veían al ferrocarril como poco rentable y tuvieron compromisos políticos con sindicatos de transporte de vías que favorecían sus intereses por encima de la posible eficiencia y rentabilidad que el tren podría ofrecer. Esto hizo que el ferrocarril, en los años posteriores, tienda a deteriorarse con leves recuperaciones en 2000 y 2005, años en los que se pretendió privatizarlo para

que hagan uso de él compañías de turismo; en el 2008 los usos a este medio de transporte son prácticamente nulos, colapsando la empresa completamente: perdió sus capacidades técnicas, administrativas y no pudo seguir las tendencias ferroviarias modernas. Además, en este último año, el derecho de vía (la obligación que tienen las personas de no construir carreteras o acueductos debajo de la línea férrea) se ve amenazado por invasiones y agresiones. Este mismo problema surge con otras propiedades como son las estaciones.

En el 2010 se decide rehabilitar los ferrocarriles, a través del proyecto denominado “Modernización de la Empresa de Ferrocarriles y Rehabilitación del Sistema Ferroviario Ecuatoriano”. La obra tuvo como fin rescatar este sistema de transporte, dándole un uso turístico; no se ha considerado el transporte de personas ni de carga pesada; en este sentido el proyecto destinó su presupuesto de la siguiente manera: 92% para la obra física (consta la rehabilitación de 446 kilómetros de línea férrea que unió varias ciudades rurales y urbanas desde Quito a Guayaquil, con sus respectivas estaciones férreas que alcanzaron un número de 47); 4% para fortalecer el área de comercialización y desarrollo local, en esta última se buscaba crear emprendimientos y empleos relacionados al turismo; y un 4% para el fortalecimiento y modernización de la empresa, lo cual implicaba la rehabilitación de los espacios físicos de la empresa y la adquisición de nuevos equipos de oficinas.

III. METODOLOGÍA

3.1 Modelo de diferencias en diferencias

De acuerdo a la naturaleza de las bases de datos y a la necesidad de evaluación del proyecto se aplicará el modelo econométrico denominado diferencias en diferencias, el cual mide el cambio esperado en una variable Y (llámese ingreso, escolaridad, o cualquiera de las variables que permitan conocer el impacto de un proyecto) entre al menos dos períodos: el período posterior y el período anterior a la implementación de una política en el grupo de tratamiento, menos la diferencia esperada en Y en el grupo de control (contrafactual o contrafáctico²) durante el

2 Se entiende como contrafactual o contrafáctico a la población de control. “En términos generales, se requiere cumplir con dos objetivos cruciales al momento de evaluar un programa proyecto: el primero es lograr aislar el efecto de la intervención, y el segundo es crear un contrafactual que responde a la pregunta: ¿Qué hubiera sucedido si no se intervenía? Por lo tanto, presenta la situación sin intervención del programa. De esta manera, se logra comparar entre el grupo que recibe el tratamiento con el contrafáctico o grupo de control”, esta información ha sido tomada de los documentos Conceptuales de la Secretaría Nacional de Planificación y Desarrollo, Documentos Conceptuales, <http://www.planificacion.gob.ec/wp-content/uploads/downloads/2012/08/Metodolog%C3%ADa.pdf>.

mismo período; de esta manera eliminamos el posible sesgo de selección de las variables inobservadas en el tiempo.

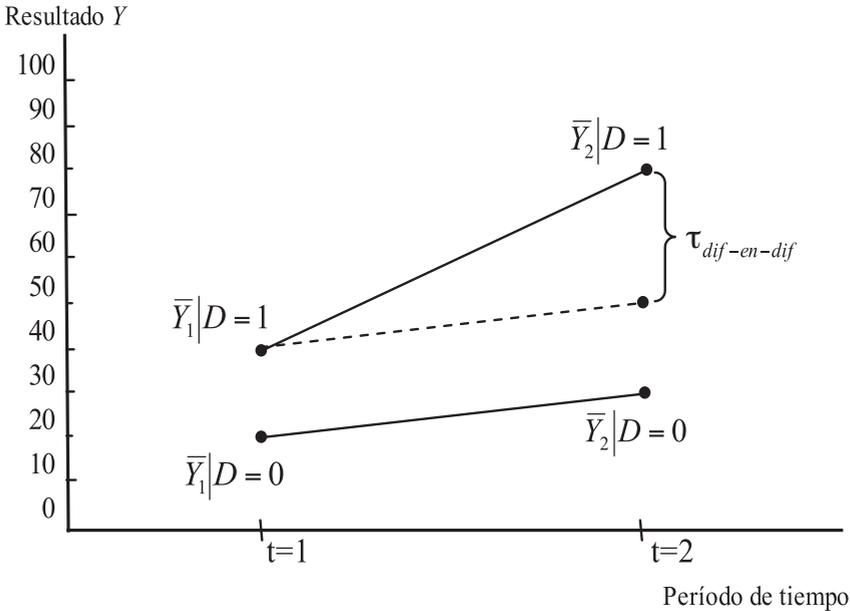
Se justifica la utilización de este modelo, pues al poseer las bases estadísticas en dos períodos, funcionan como una colección de bases de datos de corte transversal repetidas y donde cada etapa corresponde a un período diferente, es decir que no se observa a la misma persona más de una vez en el tiempo. Pudo haber sido medido el impacto con variables instrumentales pero, como se dijo anteriormente, carecemos de una variable instrumental adecuada. En el caso de este trabajo, la población de tratamiento son las personas que realizan sus actividades económicas dentro de un perímetro no mayor a 200 metros alrededor de las estaciones, siendo el contrafactual quienes están fuera de este entorno; el tiempo está comprendido entre los años 2010 al 2012; al comparar el grupo de tratamiento, el de control y el tiempo, se verificará si el proyecto de la Empresa de ferrocarriles del Ecuador ha justificado su ejecución en el incremento de los beneficios sociales.

La línea férrea fue rehabilitada en su conjunto uniendo las dos principales ciudades del Ecuador, Quito y Guayaquil. Las zonas pobladas a los alrededores de las estaciones se están viendo beneficiadas por esta obra, por ello, quienes más cerca están de las estaciones son consideradas como la población de tratamiento (no más allá de 200 metros); en este sentido, el paso del tren tiene la característica de un experimento natural, es decir que no tuvo una planificación previa de quienes serían los grupos de tratados y no tratados, sino que más bien sigue un comportamiento aleatorio, beneficiando a unos poblados y no favoreciendo a otros. A este tipo de casos se los denomina “experimento natural o cuasi experimento”, el cual es útil para aplicar el método de diferencias en diferencias, pues al existir eventos aleatorios que se construyen de manera natural antes y después del proyecto, se puede evitar lo que se conoce como el “sesgo de selección”. Para explicar de mejor manera la metodología, en el Gráfico 1 se describe el supuesto de tendencias paralelas:

Considerando por un lado a la letra *D* como una variable *dummy* que toma el valor de 1 cuando es beneficiaria del proyecto y 0 si no lo es; y por otro lado a una variable dependiente o variable de interés *Y*, la cual permite medir el impacto del proyecto (en el caso de este trabajo el logaritmo del ingreso). La línea de arriba muestra el efecto temporal de la variable ingresos para el grupo de tratados ($D=1$), inicia en el *before* (Y_1) y termina en el *after* (Y_2), la línea continua de abajo representa al grupo de control ($D=0$), de la misma manera la variable comienza en el período anterior a la ejecución del proyecto (Y_1), termina también en el punto de finalización del proyecto (Y_2), al comparar las dos líneas se dice que en ausencia del tratamiento ambos grupos tienen una evolución paralela, pero al haber proyecto las dos líneas

son diferentes y dicha diferencia es el impacto del proyecto identificado con la letra τ . Como corolario, si no existiera intervención, las dos poblaciones hubieran tenido un camino paralelo en su transición de un año a otro.

Gráfico 1: Identificación gráfica de los grupos de tratamiento y control



Fuente: Stock and Watson (2006)

El gráfico anterior puede ser expresado de acuerdo a lo señalado en la Tabla 2, en la que se estructuran los períodos y las poblaciones de tratamiento y control en cada uno de los años.

Tabla 2: Estructura del modelo de diferencias en diferencias		
Período	Tratamiento	Control
$t = 2010$	$Y_{2010} D = 1$	$Y_{2010} D = 0$
$t = 2012$	$Y_{2012} D = 1$	$Y_{2012} D = 0$

Fuente: Stock and Watson (2006)

La letra Y es la variable dependiente (para el caso de esta investigación, es el logaritmo del ingreso), que permite medir el impacto de la política pública, el sub índice nos dice el período que es 2010 para el antes y 2012 para el después. Además una variable *dummy* representada por la letra D, es 1 si pertenece al grupo de tratamiento y

0 al de control, por lo tanto, la variable dependiente está en función de su pertenencia o no a un grupo de tratamiento o control, pero en cada uno de los períodos.

Al comparar un modelo de diferencias con otro de diferencias en diferencias, el primero se usa para medir el impacto únicamente en poblaciones de tratados y no tratados; en cambio, en el segundo se adiciona al modelo anterior la variable tiempo, con lo cual se garantiza eliminar el sesgo por selección.

Por otro lado, el tiempo está dividido en el período anterior a la implementación de la mejora y posterior a la misma. Por lo tanto, la parte medular de un método de diferencias en diferencias es ubicar adecuadamente los grupos beneficiados y no beneficiados en el tiempo.

3.2 Deducción analítica del modelo

Analíticamente el modelo de diferencias en diferencias aplicado a este trabajo es como sigue:

$$\tau_{dif-en-dif} = [E(Y_1|D = 1) - E(Y_0|D = 1)] - [E(Y_1|D = 0) - E(Y_0|D = 0)] \quad (1)$$

Es decir, que tomando una variable dependiente como el caso de los Ingresos, verificamos el cambio de un período a otro, tanto para la población de tratamiento como para la de control. La ecuación 1 nos muestra la primera diferencia; en ella observamos el cambio esperado entre el período anterior y el posterior a la implementación del tratamiento en el grupo de tratamiento menos el cambio esperado entre el período anterior y el posterior a la implementación en el tratamiento en el grupo de control; las dos diferenciasson entre los años 2010 y 2012.

Con el fin de medir el impacto del proyecto, la ecuación anterior se la debe expresar en términos de una regresión multivariada, siendo el ingreso variable dependiente, y las variables que la explican los años de educación, el estado civil, edad, edad al cuadrado (para conocer los rendimientos marginales de los ingresos con un año más de educación), la actividad económica y la zona geográfica donde la desempeña. No obstante, la particularidad que tiene el modelo de diferencias en diferencias es considerar, además de las variables explicativas, dos variables *dummy*:

1. La variable *dummy* de los grupos que forman parte del experimento (sea éste aleatorio o controlado), que expresa 1 si son personas que pertenecen al grupo de tratamiento y 0 si no lo son.

2. La otra *dummy* es el tiempo, toma el valor de 1 si es el año final o ex post (en este caso 2012) y 0 si es el año inicial (para el presente trabajo 2010).

$$Y = \beta_0 + \beta_1 D_i + \beta_2 t_i + \beta_3 D_i * t_i + \beta_4 X + u_i \quad (2)$$

En la ecuación anterior se especifica un modelo de regresión multivariada, considerando las siguientes variables: Primero, la variable D_i muestra si pertenece al grupo de tratamiento ($D=1$) o si pertenece al grupo de control ($D=0$); segundo, la variable t_i es el tiempo, la cual toma el valor de 0 si pertenece al año 2010 y 1 si son del 2012; la X representa a todas las variables que permiten explicar la variable dependiente Y ; y, cuarto, la interacción de la variable tiempo y la *dummy* de los grupos $D_i * t_i$, ésta última interacción es la variable de interés que al evaluarla en su significancia estadística estamos básicamente expresando el impacto o no del proyecto³.

Al establecer falta de similitud entre poblaciones de tratamiento y control a través de los test de medias, se debe corregir la falta de semejanzas a través de iteraciones entre las variables explicativas que tienen este problema y la variable D (*dummy* que diferencia entre tratados y no tratados) con el fin de corregir posibles sesgos. En este sentido lo que hacemos es crear un nuevo modelo incluyendo estas iteraciones, de esta manera expresamos la ecuación (3) como sigue:

$$Y_{it} = \delta_0 + \delta_1 D_1 + \delta_2 T + \delta_3 T_3 * D + \delta_4 X + \delta_5 X * D + w_{it} \quad (3)$$

Las variables D , t y su interacción, así como aquella que son necesarias para explicar el modelo permanecen. Sin embargo, los estimadores de ajuste permiten evaluar de manera integral la validez del modelo; este estimador es la letra D ; se realizan iteraciones por cada variable que no sea semejante de acuerdo al test de media, corrigiendo de esta manera posibles sesgos.

IV. VARIABLES Y DATOS ÚTILES PARA LA APLICACIÓN DEL MÉTODO DE DIFERENCIAS EN DIFERENCIAS

4.1 Variables utilizadas en el modelo

Los cambios intertemporales de tratados y no tratados se verán evaluados a través del ingreso, variable que, además, nos permite comprender la oferta laboral existente en las zonas influenciadas por el proyecto (Universidad del Rosario, 2007).

3 Los detalles matemáticos que demuestran que la variable $D_i * t_i$, permite medir el impacto del modelo se encuentran en el Anexo 1 de este trabajo.

En ese sentido, se analizará los factores que determinan el ingreso en las zonas evaluadas. Al respecto Becker (1964) establece que la educación es un elemento correlacionado a la productividad e ingreso, tampoco se puede dejar de incluir al género como un elemento que explica los salarios; en este sentido, Tansel (2000) menciona la diferencia de ingresos entre hombres y mujeres y la probabilidad de pertenecer a un sector económico u otro si es mujer, también la oferta laboral se halla relacionada a un proceso de decisión del hogar. Además, se considera en estas base de datos la edad, edad al cuadrado (para medir los efectos marginales que la edad tiene sobre los ingresos), estado civil.

Por otro lado, considerando que las áreas aledañas a las estaciones del ferrocarril generan movimiento económico gracias al turismo atraído por el tren, se deben considerar las zonas geográficas (norte, sur, litoral o centro) y cómo estas definen el ingreso. Existe evidencia en la que se confirma que las características del área y la ubicación de un bien turístico en un sitio u otro juega un papel fundamental (Rémy Prud'homme, 2008, p.15). Este autor realizó un estudio en el que analiza el impacto socioeconómico del turismo en las zonas que son consideradas como patrimonio de la humanidad para la UNESCO. Entre otras, utiliza como variables explicativas la cercanía o no a un centro turístico a 50 kilómetros de la costa, si es una zona montañosa o ribereña, el hecho de disfrutar de clima agradable, etc.

Debido a que el impacto del proyecto está relacionado al turismo, también se especifica en el modelo si la persona encuestada está o no vinculada a este sector productivo. Al respecto podemos mencionar la trascendencia de la economía del turismo y cómo varios proyectos en países en vías de desarrollo como Marruecos que, desde 1960 busca en el turismo una fuente de ingresos y apoyo a sectores pobres, aprovechando al máximo sus potencialidades. En este sentido Ezaidi y Kabbachi (2007) describen cómo Marruecos ha publicitado el turismo de salud de sus aguas termales y cómo estos ingresos colaboran a la lucha contra la pobreza y extrema pobreza. También puede observarse el caso de Botswana, país que ha definido al turismo como un instrumento de política pública para disminuir el nivel de pobreza a través de la promoción de su cultura, la cual es difundida en sus visitantes (Jarkko Saarinen and Haretsebe Manwa, p.47, 2008). En otras zonas de Latinoamérica como en Argentina, concretamente en la Quebrada Humahuaca, también se ha fomentado el turismo comunitario con la finalidad de mejorar el nivel de vida en las zonas rurales (Korstanje, Universidad de Palermo, 2015).

No existen consensos entorno al conocimiento de cuánto el turismo aporta al PIB de los países desarrollados y en vías de desarrollo, pero es verdad que todos los

investigadores consideran que este sector productivo ha crecido a nivel mundial y que, en muchos casos, ha contribuido positivamente a combatir la pobreza⁴.

Finalmente, es importante señalar que el elemento pobreza ha sido controlado en este modelo considerando si la persona analizada recibe o no el Bono de Desarrollo Humano, el cual es una subvención que da el Estado a las personas en situación de pobreza.

4.2 Datos para la identificación de las poblaciones de tratamiento y control

En el año 2010, mientras la rehabilitación de la línea férrea entre Quito y Guayaquil alcanzaba un avance de obra del 3% y con la finalidad de medir la situación socioeconómica de las comunidades aledañas a las estaciones del tren, Ferrocarriles del Ecuador tomó una muestra estadística a 4,148 personas ubicadas en las mismas provincias donde se hallaba en ejecución el proyecto de rehabilitación del sistema ferroviario: Imbabura, Pichincha, Cotopaxi, Tungurahua, Chimborazo, Cañar y Guayas.

En ese año se encuestó sobre las condiciones de aquellas 4,148 personas: edad, género, ingreso promedio, estado civil, años de educación, si está empleado o subempleado, si recibe o no subvenciones del Estado como el Bono de Desarrollo Humano, etc. y se enfocó en las personas en edad de trabajar, así como a jefas o jefes de hogar; a través de esta encuesta se definió al grupo de tratamiento: aquellas familias y establecimientos comerciales y negocios que han sido encuestados alrededor de la línea férrea y se han beneficiado directa o indirectamente del proyecto turístico de la Empresa de Ferrocarriles. De esta manera, la distancia es importante en la conformación del contrafactual, siendo la población de control quienes se encuentran en cualquier otro sitio que no sea cercano a las estaciones, es decir aquellas que están a más de 200 metros.

Es común establecer a la distancia como un elemento diferenciador entre las poblaciones de tratados y controles. Wooldridge (2010) toma un ejemplo real

4 De acuerdo con el Department for International Development (1999), el turismo es el 2% del PIB y el 5% del valor de las exportaciones en 11 de los 12 países con mayor número de pobres (que viven con menos de 1 dólar al día), por ello buscar políticas públicas que fomenten el turismo comunitario puede ser una herramienta eficaz para el desarrollo y la lucha contra la pobreza, contribuyendo a la “reconstrucción” de las zonas rurales y urbanas debilitadas, siempre y cuando se lleve a cabo en procesos de “pro pobres”. La información ha sido tomada de la siguiente página web: <https://www.gov.uk/government/publications?page=2>.

con el que se busca medir la devaluación de los precios de las viviendas mientras más cercanas se encuentran a un incinerador de basura. Se concluye que entre los períodos de 1978 y 1982, las casas cercanas al incinerador se devaluaron en un 6%, lógicamente, en este caso, el modelo ha capturado y controlado todos los aspectos que influyen en el precio de una casa, tales como los años de su construcción, metros cuadrados, número de habitaciones y de baños, entre otros.

Para el año 2010 no se pudo contar con el contrafactual, es decir con el grupo de comparación⁵, por lo cual a este grupo se estructuró a partir de la base de datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo -ENEMDU- para el tercer trimestre del 2010. Esta investigación se la realiza a nivel nacional y coincide con la misma fecha en que se levantó la encuesta para la población de tratamiento; cubre la extensión del territorio nacional ecuatoriano, investiga a 6,876 familias conformadas, posee 84,774 observaciones y son estadísticas oficiales proporcionadas por el Instituto Nacional de Estadísticas y Censos del Ecuador -INEC-. En esta base se puede visualizar el perfil social, demográfico y económico de la población total, en edad de trabajar, ocupada, desocupada, inactiva y del subempleo en el área urbana y rural del país, a través de variables de carácter general como: sexo, edad, parentesco, nivel de instrucción, asistencia escolar, afiliación al seguro social, ingresos, entre otros. De aquí se filtraron las poblaciones que coincidan con las del tratamiento, se tomó al grupo poblacional que cumpla con semejantes características en edad, estado civil, ingresos, etc. y de este ejercicio de identificación 5,428 datos.

Por otro lado, entre noviembre y diciembre de 2012 cuando se completó el 98% de la obra (la cual incluye la totalidad de las estaciones y el 95% de la línea férrea), Ferrocarriles del Ecuador aplicó una nueva encuesta con el mismo contenido que la del 2010 a 4,085 personas y en las mismas zonas geográficas. No obstante, en esta ocasión se diferenció de quienes viven cerca de las estaciones (3,256 personas) de quienes no lo están (829 personas), siendo el primero el grupo de tratamiento y el segundo de control. Los dos grupos eran similares en sus aspectos socioculturales. Se tomaron de manera aleatoria, por lo cual se pudo evitar el sesgo de selección. En los dos períodos de tiempo fueron encuestadas a diferentes personas, por lo cual no se trata de datos de panel. Considerando las bases tanto de tratamiento y control en los dos períodos de tiempo, se poseen 13,659 observaciones.

Es importante resaltar que, pese a no poseer una base de tratados pura, que explique la situación *ex ante* en el proyecto para el año 2010 (pues se la tomó con

5 Que posean características socio culturales muy parecidas a la población de tratamiento y que sea para el año 2010,

un avance del 3%), no se considera a esto una limitante a la investigación, por las siguientes razones: primero, el avance en obra del proyecto no fue significativo; segundo, al medir el impacto a través de la variable de interés, en el caso de que fuera positivo o negativo nos indica, cuál es el mínimo o máximo nivel positivo o negativo de impacto del proyecto; y tercero, existen estudios en las que se demuestra la factibilidad de la medición de los posibles impactos que pudieran existir, aún si no se tuviera una base de datos de tratamientos antes de la ejecución de cualquier programa.⁶

V. ESTADÍSTICAS DESCRIPTIVAS

La Tabla 3 que se observa a continuación tiene dos grandes partes: en el área (A) se encuentra el análisis descriptivo de las variables que se usarán en el modelo, tanto de las poblaciones de tratamiento como las de control, para el 2010 (*ex ante*) y para el 2012 (*ex post*). Y en la parte (B) se realiza un test de medias comparando la igualdad o no de dichas variables.

En la misma Tabla 3, en el área (A), considerando la variable género para el año 2010, se observa en las poblaciones de tratamiento que el 46.72% son hombres y el 53.28% mujeres; y en el mismo año, en los controles, hay más hombres (46.26%) que mujeres (53.74%). Los porcentajes de la distribución de género para el 2012 son semejantes 51.51% para tratados y 52.33% para controles en el caso de mujeres; en cambio, los hombres tienen 48.49% en el grupo de tratamiento y 47.67% en el grupo de control, es decir la muestra está equilibrada en la participación de género.

Continuando con el análisis de la Tabla 3, observamos que las edades en las que más se concentra la población están comprendidas entre 25 y 50 años, lo cual coincide con la etapa de las personas en edad de trabajar. Este patrón de distribución coincide en los dos años, tanto en tratados como no tratados.

A nivel de educación, el grupo de tratamiento del 2010 está polarizado en personas con nivel básico y medio completo por un lado y superior por otro; en cambio, los controles para el mismo año muestran tomas de todos los niveles educativos. Esto puede deberse a que las muestras de tratamientos y controles del 2012 y tratamientos del 2010 fueron tomadas únicamente a los alrededores de la línea férrea. Estas personas se dedican en muchos casos a empleos generados por el turismo pero de

6 Nótese que inclusive podemos aplicar difference and difference cuando el estado de los tratados es inobservado, en este sentido el trabajo de Irene Botosaru y Federico Gutiérrez (2015) presentan la posibilidad de construir dicha población, pese a que ésta haya sido inobservada en el período anterior a la ejecución de la política, o lo que es lo mismo, cuando sólo se posee las observaciones de tratamientos y controles en un sólo período.

Tabla 3: Composición socio económica de las poblaciones de tratamiento y control en los años 2010 y 2012

Variable	Análisis Descriptivo (A)				Test de Medias (B)		
	2010		2012		Tratamiento	Control	
	Tratamiento (%)	Control (%)	Tratamiento (%)	Control (%)	H0: (D==1 & t==0) - (D==0 & t==0) = 0	H0: (D==1 & t==1) - (D==0 & t==1) = 0	
Genero	Hombre	46.72	46.26	48.49	46.67	$t = 1.8236$	$t = 1.902$
	Mujer	53.28	53.74	51.51	52.33	$Pr = 0.0670$	$Pr = 0.0482$
Nivel Educativo	Básico incompleto	43.63	43.2	27.4	17.73		
	Básico completo	-	2.3	-	-	$t = 4.9326$	$t = -6.3488$
	Media incompleta	-	9.14	-	-		
	Media completa	39.82	17.76	-	-		
	Superior	16.55	27.6	72.6	82.27	$Pr = 0.0000$	$Pr = 0.0000$
Estado Civil	Casado o Unión Libre	37.84	38.14	42.08	39.93	$t = 0.2015$	$t = -1.1202$
	Otro	62.16	61.86	57.92	60.07	$Pr = 0.0707$	$Pr = 0.2627$
Edades	18-24	16.47	11.07	15.05	15.04		
	25-35	29.43	22.55	31.6	29.71		
	36-50	28.56	30.01	35.52	38.14	$t = 6.9356$	$t = -5.3478$
	51-60	14.28	16.53	10.54	10.15		
	61+	11.26	19.84	7.28	6.97	$Pr = 0.0000$	$Pr = 0.0000$
BDH	Si	10.59	11.77	6.33	4.34	$t = -1.8158$	$t = 1.1612$
	No	89.41	88.23	93.67	95.66	$Pr = 0.0347$	$Pr = 0.0307$
Sector Turismo	Si	6.34	6.21	7.31	8.69	$t = 0.2699$	$t = 1.9578$
	No	93.66	93.79	92.69	91.31	$Pr = 0.3936$	$Pr = 0.0458$

Fuente: Encuestas elaboradas por FEED entre 2010-2012 y base de datos del INEC en 2010

manera informal: ventas ambulantes o comidas, generalmente poseen pocos años de educación, sucede lo contrario con las personas que manejan negocios u hoteles, pues ellos poseen más años de educación.

Se puede demostrar que las personas pobres y extremadamente pobres tienen múltiples actividades laborales, entre las que cuentan negocios ambulantes (Banerjee y Duflo, P.10, 2006). A este hecho se añade la rehabilitación de las estaciones, lo cual atrae aún más el comercio informal. Por tal motivo también, como una estrategia de controlar el factor pobreza, se consideran a las familias que reciben y que no reciben el Bono de Desarrollo Humano, el cual es una subvención que da el Estado para las personas que viven en estado de pobreza. En la muestra se observa que cada vez son menos quienes lo reciben de un año a otro, no obstante es difícil establecer

si esto obedeciera a una mejora en el nivel de vida de las familias o al retiro del apoyo del Estado de este recurso.

Para verificar técnicamente la diferencia o no de las variables, en la parte (B) de la Tabla 3, levantamos un test de hipótesis para las variables de tratamiento y control. Sus resultados se encuentran expresados en la parte derecha de la tabla. Se asume como hipótesis nula que la diferencia entre variables de tratamiento y control para cada año es cero, de esta manera se establecerían que dichas variables son iguales; la *t de student* y la *p-value* nos ayudan a evaluar la veracidad de la hipótesis nula, cuando sus valores estén o no en área de aceptación y rechazo al 95% de nivel de confianza. Los resultados muestran que las variables de género son iguales para el año 2010 (*t de student* de -1.82) y para el año 2012 (*t de student* de -1.90); en el caso del nivel educativo las muestras son muy diferentes, para el año 2010 (*t de student* de 4.93) y para el 2012 (la *t* es igual a -6.3488).; En el caso del estado civil existe homogeneidad en los dos períodos de estudio en los 2 grupos, en el 2010 (*t* igual a 0.2015) y en el 2012 (*t* igual a -1.12); las edades son muy disímiles, pues en el 2010 la *t de student* es de 6.93 y en el 2012 de -5.34; la distribución del Bono de Desarrollo Humano muestra valores de igualdad pues en el 2010 la *t de student* es de -1.81 y en el 2012 de 1.16; finalmente, el sector turístico también son iguales con una *t de student* de 0.2699 en el 2010 y 1.95 en el 2012.

Por lo tanto las variables de edad y educación podrían causar sesgos al momento de realizar el análisis deregresión multivariada aplicando el modelo diferencias en diferencias, impidiendo conocer si realmente hubo impacto del proyecto. Por ello, en el apartado siguiente, también se considerará un modelo que incluyan iteraciones de la edad y educación con la variable *dummy D* (grupo de tratamiento o control).

La Tabla 4, muestra el nivel educativo de las personas que están vinculadas al sector turístico y, por ende, al proyecto de rehabilitación del ferrocarril. En este sentido, los que poseen educación superior son quienes más participan con un 42.20%, luego las personas con educación básica incompleta con un 35.93%, media completa con un 18.13%, media incompleta 2.97% y, finalmente, quienes tienen educación básica incompleta con un 0.77%. Estos resultados permitirán vincular, en lo posterior, con la distribución de ingresos.

Dado que el coeficiente que se tomará para medir el impacto es el ingreso; a continuación en la Tabla 5 se detalla cómo se lo distribuye entre tratados y no tratados en los dos años analizados, considerando las zonas en las que el proyecto se ha ejecutado:

Tabla 4: Nivel educativo de las personas que se dedican al sector turístico

Nivel Educativo	Variable	Pertenecer sector turístico
	Básico incompleto	35.93%
Básico completo	0.77%	
Media incompleta	2.97%	
Media completa	18.13%	
Superior	42.20%	
Total	100%	

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

Tabla 5: Promedio de ingresos en las poblaciones de tratamiento y control en dos periodos según la zona de influencia del proyecto

Ubicación	Variable	2010		2012	
		Tratamiento (\$)	Control (\$)	Tratamiento (\$)	Control (\$)
Ubicación	Centro	373.30	421.39	436.31	465.80
	Litoral	384.42	321.03	366.99	298.50
	Norte	372.39	185.33	715.26	425.92
	Matriz	505.48	535.20	428.39	445.78
	Sur	274.64	211.26	0	0

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

Miramos que la zona donde mayores ingresos se perciben es la norte, lo cual halla explicación en el hecho de que son sitios con alto atractivo turístico en el Ecuador. En la Tabla número 6 se pueden observar los ingresos promedios de las zonas que se dedican al sector turístico, tanto de la población de tratamiento y control en los dos años de análisis:

Efectivamente, de acuerdo a la Tabla 6 miramos que existen mayores ingresos por turismo en la zona norte, pero si buscamos observar el efecto total de los ingresos por turismo es interesante relacionar los datos anteriores de la siguiente manera: se restan los promedios de ingresos entre tratados del año 2012 y 2010 (C-A), de igual manera se restan los controles en los mismo períodos (D-B) y finalmente buscamos

el efecto total; restamos los resultados de tratados y controles en los dos períodos y obtenemos el cuadro siguiente:

Tabla 6: Promedio de ingresos por turismo en las poblaciones de tratamiento y control en las zonas de influencia del proyecto					
Ubicación	Variable	2010		2012	
		A Tratamiento (\$)	B Control (\$)	C Tratamiento (\$)	D Control (\$)
Ubicación	Centro	490.57	472.04	460.52	553.75
	Litoral	404.13	268.32	340.38	231.25
	Norte	330	180	548.76	0
	Matriz	477.32	597.46	424.35	483.42
	Sur	226.56	221.56	0	0

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

Siguiendo el método de evaluación de impacto, se considera en la Tabla 7 que el efecto total es la última columna. Se la interpreta como el cambio de ingresos promedio entre las poblaciones de tratamiento y control considerando los años 2010 y 2012. Se observa que los ingresos por turismo disminuyeron en promedio 111,75 dólares en la zona Centro; 26,66 dólares en el Litoral y 5,31 dólares en el Sur; únicamente en la zona norte miramos un efecto positivo en ingresos por turismo con 398,76 dólares⁷. Por tal razón, en el modelo econométrico se considerará como variable explicativa al ingreso, pues de acuerdo a los análisis descriptivos, a través de esta variable las poblaciones están siendo beneficiadas con el proyecto.

7 Al respecto cabe mencionar que existe una relación del crecimiento económico de los países con el aumento del flujo turístico extranjero, siempre y cuando el país pueda incrementar las facilidades de infraestructura y comunicaciones a los visitantes foráneos. Así lo muestra el estudio realizado por Samina Khalil y Mahmood Khan Kakar (2008), se concluye que en Pakistán existió un aumento del turismo internacional gracias a la intervención del Estado mediante políticas públicas que aprovecharon el crecimiento económico. Usando el concepto y métodos de cointegración y el Test de causalidad de Granger, el estudio exploró las relaciones dinámicas en el corto y en el largo plazo las condiciones de equilibrio. Los resultados fueron similares cuando se aplicó el mismo modelo para España: El crecimiento del turismo aumenta el crecimiento de la economía en el corto plazo, y la combinación de los resultados apuntaban a una causalidad bidireccional para el crecimiento económico y el crecimiento del turismo que es necesario para el desarrollo del turismo en un país con expansión económica.

Tabla 7: Efecto total de ingresos por turismo en las poblaciones de tratamiento y control en las zonas de influencia del proyecto

	Diferencias entre años		Efecto total (\$)	
	Variable	Tratados (\$)	Controles (\$)	$((D=1 \text{ y } t=1)-(D=1 \text{ y } t=0)-(D=0 \text{ y } t=1)-(D=0 \text{ y } t=0))$
Ubicación	Centro	-30.05	81.71	-111.76
	Litoral	-63.74	-37.07	-26.67
	Norte	218.76	-180	398.76
	Matriz	-52.97	-114.04	61.07
	Sur	-226.56	-221.25	-5.31

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

VI. MODELO DE DIFERENCIAS EN DIFERENCIAS APLICADO AL CASO EN PARTICULAR

6.1 Análisis de los resultados

De acuerdo a los datos descritos anteriormente y con el fin de medir el impacto entre controles y tratados de un año a otro, se ha generado una regresión multivariada, en la que se toma el logaritmo de los ingresos como variable dependiente. Además, consta de una variable edad, edad al cuadrado que permita conocer el efecto en el ingreso por el aumento de esta; una variable *dummy* de género, siendo 1 si es hombre y 0 si es mujer.

Se debe destacar en el método de diferencias en diferencias las interacciones entre las variables y el tiempo. Este es el caso del coeficiente género que es multiplicado por el tiempo. Esta mide la diferencia salarial entre hombres y mujeres de un año a otro (del 2010 al 2012); se considera además una variable *dummy* para estado civil siendo 1 si es casado o unión libre y 0 otra; una variable de años de educación; una variable *dummy* que es uno si recibe el bono de desarrollo humano y cero si no recibe, con su correspondiente interacción con el tiempo; una variable *dummy* de ubicación geográfica siendo 1 si vive en el Norte y 0 si vive en cualquier otra zona, también se la relaciona con el tiempo; finalmente, una variable *dummy* que es 1 si la persona se desarrolla en el sector turístico y 1 si trabaja en cualquier otro sector. Adicionalmente, se consideran las variables del método aquí aplicado: la letra D es una *dummy* que toma el valor de 1 si representa a la población de tratamiento y 0 si no lo es; una variable de tiempo que significa 0 si pertenece al año 2010 y 1 si es del año 2012, además de la interacción del tiempo y las poblaciones (D*t), siendo esta

la que nos ayuda a conocer si realmente existió impacto. En el caso de la Tabla 8, se establecen tres modelos: en el primero solo se consideran las variables que son condición necesaria (mas no suficiente) del modelo de diferencias en diferencias, la D , t y su interacción $D*t$, el segundo además de las variables antes mencionadas considera otras que explican los ingresos (condición necesaria y suficiente); y el tercer modelo incluye iteraciones de las variables que no son semejantes y que podrían causar sesgos en los resultados (condición necesaria y suficiente, pero ajustada).

El primer modelo solo considera las variables útiles para medir el impacto D , t y su interacción $D*t$; se observa que los parámetros no son estadísticamente significativos y la r cuadrada es muy baja, de apenas un 0.86%.

En cambio, en el segundo modelo, al incluir los coeficientes que explican el ingreso, se ve que los signos y los parámetros, además de ser individual y globalmente significativos, con una r cuadrada de 27.51%, son los esperados: por el aumento de un año en la edad, el ingreso se incrementa en un 3.38%; el parámetro de la edad al cuadrado es negativo, pues por cada año que aumenta el ingreso es marginalmente menor; en género, los hombres perciben mayor salario que las mujeres correspondientes a un 22.01%, sin embargo, esta diferencia disminuye en un 6.55%, si consideramos la interacción del tiempo en el lapso comprendido entre el 2010 y 2012; otro resultado común es encontrar que las personas casadas tienen mayores ingresos que las que no lo son. En este modelo dicha diferencia se da en un 13.65%; con relación a los años de educación, un año más implica un aumento del ingreso en 6.99%; mientras tanto hay un 74.70% de personas que no reciben el Bono de Desarrollo Humano y esta diferencia aumenta aún más entre los años 2010 y 2012 con un 33.82%.

Considerando las áreas geográficas, las personas ubicadas en la zona norte perciben más ingresos perciben con un 34.77% en relación con quienes están en otras zonas; considerando al turismo, las personas que se dedican a trabajar en este sector poseen ingresos de 5.76% en relación a quienes no lo hacen. Considerando las variables D , t y su interacción $D*t$, tenemos que la población de tratamiento supera a los ingresos percibidos a la población de control en 8.43%, aunque en el 2012 exista una disminución del 19.60% con respecto al 2010. Finalmente el proyecto tiene un impacto positivo en ingresos de 9.67%, demostrándose, de esta manera, que la creación de este proyecto ha sido una política pública adecuada.

Ahora bien, como se pudo observar en la Tabla 3, cuando comparamos las poblaciones de tratamiento y control en sus diferentes variables, pudimos ver a tra-

vés del test de medias que no existen semejanzas en las variables educación y edad, por ello se aplica un tercer modelo que sirva para corregir este problema usando iteraciones de dichas variables con la variable D, es decir con aquella que nos permite saber si es o no beneficiaria del proyecto; de esta forma ajustaremos el modelo de diferencias en diferencias.

De acuerdo a los resultados observados en la última columna de la Tabla 7, vemos que los signos y los parámetros son los esperados, además de ser individual y globalmente significativos, con una r cuadrada de 27.54%. Miramos los siguientes resultados: con el incremento de un año en la edad, el ingreso también aumenta en 3.61%; el parámetro de la edad al cuadrado es negativo, pues por cada año que aumenta el ingreso es marginalmente decreciente; en relación al género, los hombres perciben mayor salario que las mujeres correspondientes a un 22.17%; otro dato lógico es el de las personas casadas con mayores ingresos con respecto a las que no lo son; en este modelo dicha diferencia se da en un 13.41%; la variable educación nos dice que por cada año de estudios los ingresos se incrementan en 6.97%, además se añade la iteración de educación con la D y su resultado es -0.023%; la variable Bono de Desarrollo Humano, nos habla que un 74.99% de personas no poseen esta subvención y la iteración con el tiempo nos habla que de un año a otro el bono aumentó en un 34.63%.

En relación a la ubicación de los beneficiarios del proyecto, las personas que se encuentran en la zona norte son las únicas que perciben ingresos positivos con un 34.72% en relación con quienes están en otras zonas; en este modelo la variable turismo nos dice que hay un incremento en ingresos de quienes están en este sector equivalente a 5.72%.

Considerando el parámetro D tenemos que la población de tratamiento supera a los ingresos promedio percibidos por la población de control en 18.45%, aunque de acuerdo al parámetro t del tiempo en el 2012 exista una disminución en ingresos del 18.91% con respecto al 2010. Finalmente, el proyecto, considerando la variable de interés y que permite medir su relevancia, esta es la iteración de D^*t , observamos que tiene un impacto positivo en ingresos en un promedio de 8.66%, la cual no difiere demasiado de la obtenida en el modelo anterior. Sin embargo la r cuadrada mejoró y los estadísticos *t de student* también disminuyeron. Por lo tanto, el efecto definitivo del proyecto nos dice que por cada dólar las poblaciones beneficiadas incrementan sus ingresos en 8.66%, esto entre los años 2010 y 2012.

Tabla 8: Funciones del logaritmo de ingresos

Variable	Modelo 1	Modelo 2	Modelo 3
Edad		0.0338 (0.00249)	0.0361 (0.0026)
Edad^2		-0.31 (-0.00002)	-0.0003 (0.00002)
Género		0.220105 (0.01546)	0.2217 (0.01666)
Género*t		-0.0655 (0.03242)	-0.0652 (0.02989)
Estado Civil		0.1365 (0.01453)	0.1341 (0.0145)
Años de educación		0.06992 (0.00164)	0.0697 (0.0016)
Años de educación*D			-0.0023 (0.0009)
Bono de Desarrollo Humano		-0.7470 (0.028)	-0.7499 (0.027)
Bono de Desarrollo Humano*t		0.3382 (0.0728)	0.3463 (0.058)
Ubicación Norte		0.3477 (0.0318)	0.3472 (0.0309)
Ubicación Centro		-0.0989 (0.0142)	-0.0998 (0.0145)
Ubicación Sur		-0.22631 (0.04141)	-0.2271 (0.0414)
Turismo		0.05763 (0.02703)	0.0572 (0.027)
D	0.1107 (0.0189)	0.08435 (0.0150)	0.1845 (0.04181)
T	0.0639 (0.0343)	-0.1960 (0.0377)	-0.18917 (0.03432)
D*t	0.0446 (0.0405)	0.09678 (0.0380)	0.0866 (0.03530)
Constante	5.56750 (0.0124)	40.354 (0.0563)	3.964292 (0.06419)
Observaciones	13654	13654	13654
r^2	0.0086	0.2751	0.2754
F	39.66	383.45	324.00
p Valor	0.00000	0.00000	0.0000

VII. ANÁLISIS DE LA DISTRIBUCIÓN DE LOS INGRESOS EN TÉRMINOS DE IMPACTO

A continuación, usando las mismas bases de datos con las que se generó el modelo de diferencias en diferencias calculado en el apartado anterior, desarrollaremos un análisis de la distribución de los ingresos en términos de impacto, para lo cual se calcula el índice de Gini y Atkinson para los grupos de beneficiados y no beneficiados del proyecto entre los años 2010 y 2012; las dobles diferencias entre ellos da como resultado indicadores de distribución de la renta, es decir nos muestra cómo el proyecto pudo ayudar a mejorar la equidad en las zonas de estudio.

7.1 Índice de Gini

El índice de Gini de la expresión (4) está en función de la distribución de los ingresos para cada una de ellas, sus resultados son útiles a la hora de evaluar la distribución de los ingresos; la manera que este se calcula es a través de la relación de los porcentajes acumulados de ingresos y de poblaciones:

$$G = \frac{1}{2 * n^2 * \mu} \sum_{i=1}^n \sum_{j=1}^n |X_i - X_j| \quad (4)$$

Este índice, que es siempre un número entre 0 y 1, nos dice que mientras más se acerca a la unidad el grupo de análisis es más inequitativo.

7.1.1 Cálculo del Índice de Gini en términos de impacto

En el siguiente apartado se calcula el índice de Gini y Atkinson, teniendo como base la agrupación de los ingresos según lo que muestra la Tabla 9:

La tabla siguiente se ha obtenido de acuerdo a los criterios generales de clasificación de pobreza y extrema pobreza en el Ecuador, según lo establecido por el Instituto Nacional de Estadísticas y Censos del Ecuador (INEC)⁸, La primera fila es de las personas extremadamente pobres y quienes viven con menos de 32.28 dólares; la segunda fila considera pobres a quienes tienen ingresos comprendidos entre 32.28 y 57.29 dólares; el tercer margen es de quienes poseen rentas medias comprendidas entre 340 y 1,500 dólares, siendo los 340 dólares el mínimo vital en

8 Instituto Nacional de Estadísticas y Censos del Ecuador, Medidas de Pobreza y Extrema Pobreza por Ingresos, Resumen Ejecutivo; Quito, 2014.

el Ecuador y, en el último margen están las personas con renta alta que ganan entre 1,500 y 7,000 dólares mensuales.

Tabla 9: Clasificación de las rentas para el cálculo de los indicadores de distribución de la renta

	Niveles de renta (\$ mensuales)	
	Extremadamente pobres	0
Pobres	32.28	57.29
Media baja	57.29	340
Media alta	340	1,500
Alta	1,500	7,000

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

Luego se calcula el índice de Gini expresado en términos de la evaluación de impacto; la primera diferencia es para poblaciones de tratamiento pero en diferentes años; la segunda diferencia corresponde a las poblaciones de control también en los períodos comprendidos entre el 2012 y 2010:

$$G_{efecto} = [G_{2012}^T - G_{2010}^T] - [G_{2012}^C - G_{2010}^C] \quad (5)$$

Al aplicar la fórmula (5) obtenemos los resultados expuestos en la Tabla número 10 El índice de Gini para las poblaciones de tratamiento en el 2010 es de 34.53%, para el mismo año la población de control es del 36.29%; por otro lado en el 2012 la población de tratamiento es del 27.69% y la de control del 31.36%.

Tabla 10: Medición del índice de Gini como efecto de la política pública

	2010	2012	Gini en términos de impacto
Tratamiento	34.53%	27.69%	-6.85%
Control	36.29%	31.36%	-4.93%
Efecto total entre las dos poblaciones y el tiempo			-1.91%

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

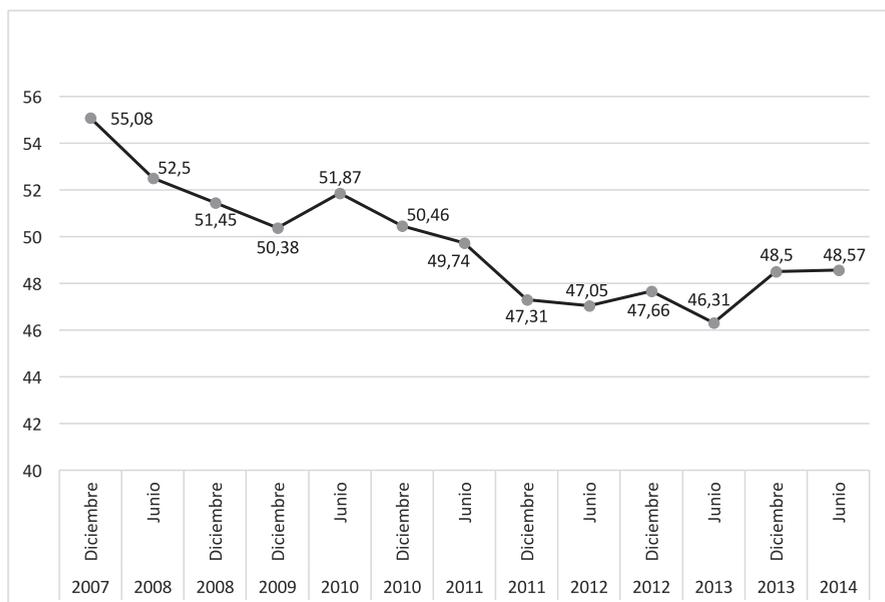
Aplicando la fórmula (5) obtenemos los resultados expuestos en la Tabla 10. En ella realizamos la primera diferencia restando el Gini del 2012 para los tratados que corresponde a 27.69% menos 34.53% del 2010. Su resultado es de -6.85%, lo cual significa que la inequidad bajó entre la población de tratamiento de un año a otro en este porcentaje; la segunda diferencia corresponde al grupo de los controles

de un año a otro y se lo calcula restando 31.36% menos 36.29%, su resultado es de -4.93%, lo que significa que de un año a otro la inequidad bajó en este porcentaje en la población de control. El efecto total -1.91%, que está al final de la Tabla 9, es la resta de -6.85% a -4.93%. En primera instancia se podría concluir que, en términos de impacto, el proyecto hizo que disminuya la inequidad en la distribución de las rentas en 1.91%.

Gracias al Gráfico 2 observamos la evolución del índice de Gini nivel nacional en Ecuador desde el 2007 hasta junio del 2014: existe una tendencia a la baja. En el 2007 este era del 55.08% y en diciembre del 2010 (año en el cual se tomó la primera muestra para realizar este trabajo) era de 51.87%.

Al comparar la Tabla 10 con el Gráfico 2 del INEC, vemos que en esta tabla la población de tratamiento (Gini de 34.53%) y de control (Gini de 36.29%) tienen ponderaciones menores al Gini a nivel nacional que para diciembre de 2010 es de 50.46%.

Gráfico 2: Índice de Gini de Ecuador a nivel nacional por años



Fuente: Instituto Nacional de Estadísticas y Censos del Ecuador, INEC

Además, en el Gráfico 2 vemos que el índice de Gini para diciembre del 2012 es de 47.05% comparándolo nuevamente con nuestros resultados obtenidos en la

Tabla 10, tratados (27.69%) y controles (31.36%), vemos que estos últimos son bajos y continúan siendo menores que el Gini a nivel nacional⁹, incluso han disminuido en relación al 2010. Este hecho se explica por dos razones: uno, las poblaciones de tratamiento y control siguen la tendencia nacional de disminución de la desigualdad y dos, efectivamente, el proyecto tuvo un impacto positivo en términos de mejora de los niveles de distribución de la renta.

7.1.2 Análisis del índice de Gini de acuerdo a la distribución porcentual de ingresos entre las poblaciones

A continuación, tomamos las distribuciones porcentuales acumuladas de ingresos y de poblaciones, tanto en el 2010 como en el 2012 para las poblaciones de tratamiento y control, con la finalidad de conocer por qué disminuyó la desigualdad en término de impacto y cuáles fueron los grupos de ingresos que más influyeron en dicha disminución.

Tabla 11: Porcentaje acumulado de la renta y los ingresos para el año 2010 de la población de tratamiento	
Distribución de la Renta de tratados 2010	
Porcentaje acumulado de la Población	Porcentaje acumulado de la Renta
0	0
0.675	0.020
1.905	0.120
60.468	21.280
97.926	83.964
100	100

Fuente: Encuestas elaboradas por FEEP en 2010

9 Los resultados de este trabajo son consistentes con la tendencia histórica en varias sociedades, considerando al índice de Gini entre 0 y 1, siendo 1 alta y 0 baja desigualdad, “con un coeficiente de Gini de 0.19, la distribución de los ingresos del trabajo percibida en los países escandinavos de los años 1970-1980 no distaba mucho de la igualdad absoluta. A la inversa, con un coeficiente de Gini de 0.85 el reparto de los capitales observado en Europa en la bella Época no estaba lejos de la desigualdad absoluta”; con esta aclaración definimos las razones cuando se define un índice de Gini de ingresos por trabajo es alto y en otras cuando es bajo. Siguiendo al mismo autor, a la luz de la experiencia en su trabajo, advierte que “el coeficiente de Gini varía aproximadamente entre 0.2 y 0.4 para las distribuciones de los ingresos del trabajo observadas en la práctica en diferentes sociedades, entre 0.6 y 0.0 para las de la propiedad del capital y entre 0.3 y 0.5 para la desigualdad del ingreso total” (Pickety, 2014, p. 280).

Para el año 2010, usando la base de datos de la Empresa de Ferrocarriles para la población de tratamiento, miramos en la Tabla 11 cómo los bajos porcentajes de la población (primera columna) perciben también bajos porcentajes de renta (segunda columna); en cambio, los porcentajes más altos de la primera también perciben altos porcentajes de la segunda; esa es la razón por la cual el índice de Gini es bajo (34.53%), evidenciando un equilibrio en la distribución de la renta.

En el mismo año 2010, pero usando la base de datos obtenida del INEC para los controles, en la Tabla 12 hacemos el mismo análisis comparativo anterior de la población acumulada versus porcentaje acumulado de la renta:

Tabla 12: Porcentaje acumulado de la renta y los ingresos para el año 2010 de la población de control	
Distribución de la Renta de controles 2010	
Porcentaje acumulado de la Población	Porcentaje acumulado de la Renta
0	0
1.142	0.031
10.096	0.704
62.509	18.168
95.707	69.399
100	100.0

Fuente: Base de datos de 2010 del INEC

Se observa en la Tabla 12 que el porcentaje acumulado de la renta y población no poseen una distribución equilibrada del ingreso: por ejemplo, el 95.7% de la población percibe el 69.39% de la renta; y si comparamos a este mismo nivel en la Tabla 10, el 97.92% de la población posee el 83.96% de los ingresos. Con ello se explica por qué en la población de control para el 2010 el índice de Gini sea de 36.29% y es mayor que el obtenido con los tratados en este mismo período de tiempo. Por ende, la población de tratamiento posee mejor distribución en el ingreso que la de control en el 2010.

Ahora hacemos el mismo análisis para las poblaciones tratadas del 2012. En la Tabla número 13 miramos los siguientes resultados:

En el caso del grupo de tratamiento para este año tenemos un índice de Gini igual a 27.69%, muy por debajo de los anteriores cálculos, esto se debe a que los

Tabla 13: Porcentaje acumulado de la renta y los ingresos para el año 2012 de la población de tratamiento

Distribución de la Renta de tratados 2012	
Porcentaje acumulado de la Población	Porcentaje acumulado de la Renta
0	0
3.250	0.086
5.469	0.250
44.000	12.842
99.563	96.941
100	100

Fuente: Encuestas elaboradas por FEEP en 2012

porcentajes acumulados de poblaciones y rentas están distribuidas de manera muy pareja, el 99.56% de la población posee el 96.94% del ingreso, es decir durante el período 2010 y 2012 existe una disminución de la inequidad en la distribución de la renta en los grupos de tratados.

Finalmente, en la Tabla 14 analizamos la población de control para el año 2012:

Tabla 14: Porcentaje acumulado de la renta y los ingresos para el año 2012 de la población de control

Distribución de la Renta de controles 2012	
Porcentaje acumulado de la Población	Porcentaje acumulado de la Renta
0	0
3.619	0.099
7.238	0.374
48.130	14.172
99.156	93.905
100	100

Fuente: Encuestas elaboradas por FEEP en 2012

En este caso el índice de Gini es de 31.36%, posee una distribución de sus ingresos muy semejante que explica que el índice de Gini sea bajo. Por lo tanto, se concluye que en términos de impacto existe mejora en la distribución de los ingresos, pero se debe saber quiénes están siendo más beneficiados con esta mejora, si las personas con mayores o menores rentas. No obstante, para poder visualizar esto, se

debe usar un indicador que permita diferenciar entre rentas altas y bajas, para esto usaremos el índice de Atkinson.

7.2 Índice normativo de Atkinson

La familia de índices de desigualdad de Atkinson (A_α) constituye el enfoque normativo a la desigualdad de la renta. En 1970, Atkinson argumentó que tras la utilización de los índices objetivos tradicionales de desigualdad se ocultaban necesariamente juicios de valor que sería mejor desvelar directamente en términos de una función de bienestar social que incorporará explícitamente las propiedades normativas que se considerasen deseables. Así, Atkinson propone una familia de índices relativos de desigualdad cuyas expresiones son:

$$A_\alpha = 1 - \prod_{i=1}^n \left(\frac{Y_i}{\mu} \right)^{\frac{1}{\alpha}}; \alpha=1 \quad (6)$$

$$A_\alpha = 1 - \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i}{\mu} \right)^{1-\alpha} \right]^{\frac{1}{1-\alpha}}; \alpha \geq 1, \alpha \neq 1 \quad (7)$$

Las expresiones (6) y (7), tienen la característica de contener al parámetro α , susceptible de ser interpretado como un "coeficiente de aversión a la desigualdad", la letra Y representa el ingreso, n la población y μ la media de ingresos. En la ecuación (6) vemos el índice de Atkinson para alfa igual a 1 y la ecuación (7) se la usa para un valor de alfa igual a 0.5.

Este índice mide la sensibilidad ante transferencias a distintos niveles de renta, al aumentar α crece el peso que se otorga a las transferencias en la cola baja de la distribución (personas con menores rentas) y decrece el peso de las realizadas en la cola alta de la misma (personas con mayores rentas), encontrándose los valores más utilizados entre 0.5 y 1. Los valores adoptados por estos índices se pueden entender como la proporción de la renta a la que se estaría dispuesto a renunciar a cambio de la igualdad distributiva. A mayor valor del índice, mayor el grado de la desigualdad, tomando valores en el rango alfa [0.1].

La Tabla 15 se muestra los resultados del índice de Atkinson, obtenidos de las bases de datos para grupos de tratamiento y control para los años 2010 (tratados tomada de la base de datos de FEED y control base de datos del INEC) y 2012 (base de datos de FEED para tratados y controles):

En la primera fila hacemos uso de la fórmula (6) para un valor de alfa igual a 1 (se considera el peso a la cola más baja o de menores ingresos), y en la segunda

fila se aplica la fórmula (7) con un valor alfa igual a 0.5 (el mayor peso es para la cola más alta o de quienes poseen más renta). Al comparar los dos resultados vemos que la población de tratamiento en el 2010, con un alfa igual a 1 tiene un Atkinson del 22.89%, mientras que para los tratados en ese mismo año es del 19.19%, esto implica que las poblaciones de tratamiento con ingresos más bajos son más inequitativas que los tratados con mayores ingresos; este mismo patrón de desigualdad se repite para controles del 2010 y controles y tratados del 2012; también estos valores se pueden expresar diciendo que las poblaciones con menores ingresos son más inequitativas que las poseedoras de mayores rentas.

Tabla 15: Medición del índice de Atkinson para poblaciones de tratamiento y control en el 2010 y 2012

	2010		2012	
	Tratamiento	Control	Tratamiento	Control
Atkinson $\alpha=1$	22.89%	39.01%	30.78%	34.51%
Atkinson $\alpha=0.5$	19.19%	21.43%	14.12%	16.46%

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

Además, se observa que los grupos de tratados son más equitativos comparados con los controles, tanto en el año 2010 como en el 2012 y para un alfa igual a 1 y 0.5. Por ejemplo, la población de tratamiento con un alfa igual a 1 del 2010 tiene un índice de Atkinson igual a 22.89%, en cambio para el mismo año el grupo de control es de 39.01%.

Para analizar el efecto del impacto aplicamos la fórmula (8) para alfa 1 y 0.5, de esta manera observaremos el efecto total de un período a otro para la población de tratamiento y control:

$$A_{efecto} = [A_{2012}^T - A_{2010}^T] - [A_{2012}^C - A_{2010}^C] \tag{8}$$

La fórmula (8) explica el Atkinson en términos de impacto. La primera diferencia es el Atkinson de tratados entre el 2012 y 2010; la segunda diferencia corresponde al Atkinson para controles con los mismos períodos de tiempo.

La Tabla 16 nos indica el índice de Atkinson con un valor de alfa igual a 1, tanto para la población de tratamiento y control; tenemos los siguientes resultados:

Se debe interpretar el efecto total del Atkinson de acuerdo a lo establecido en la Tabla 16, para ello nuevamente aplicamos la fórmula (8). La primera diferencia es la resta de los Atkinson de los tratados entre períodos, es decir 22.89% menos 30.78% su resultado es 7.89%; luego restamos el Atkinson de los controles entre los mismo períodos 39.01% menos 34.51%, cuyo resultado es -4.50%. El impacto final es la resta de las dos diferencias anteriores, es decir 7.89% menos -4.74%, el resultado dice que hubo un incremento en la desigualdad para las poblaciones más pobres en 12.39%.

Tabla 16: Medición del índice de Atkinson para alfa igual a 1			
	2010	2012	Variación anual del Atkinson
Tratamiento	22.89%	30.78%	7.89%
Control	39.01%	34.51%	-4.50%
Efecto total entre las dos poblaciones y el tiempo			12.39%

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

Ahora consideraremos tratados y controles en los dos períodos, pero con un alfa de 0.5, es decir aplicando un mayor peso a los ingresos más altos:

Tabla 17: Medición del índice de Atkinson para alfa igual a 0.5			
	2010	2012	Variación anual del Atkinson
Tratamiento	19.19%	14.12%	-5.08%
Control	21.43%	16.46%	-4.97%
Efecto total entre las dos poblaciones y el tiempo			-0.10%

Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

En la primera fila de la Tabla 17 observamos la población de tratamiento para el año 2010 con un Atkinson de 19.19%, éste se resta del 14.12% para el mismo grupo pero en el año 2012 y su resultado es una disminución igual a 5.08%; luego la segunda diferencia consiste en la resta de los controles entre el 2012 y 2010, es decir 21.43%, menos 16.46%, siendo el resultado una disminución de la desigualdad equivalente a -4.97%. El efecto total, la resta de las dos diferencias -5.08%

menos -4.97% tiene como resultado -0.10%, el cual es interpretado como una disminución de la inequidad entre tratados y no tratados, pero para las personas con mayores ingresos.

En consecuencia, el proyecto ha permitido incrementar el bienestar de las personas con rentas más altas; en cambio los grupos con ingresos bajos no mejoraron la distribución de sus ingresos. Este hecho corrobora lo observado en las estadísticas descriptivas anteriores, donde mostramos a las personas con mayor nivel educativo dedicadas al sector turístico (Tabla 4), además se sabe que mientras mayor es el nivel educativo mayores son los ingresos.

VIII. CONCLUSIONES Y RECOMENDACIONES

La principal contribución de este estudio es definir cómo, a través de metodologías econométricas muy difundidas, como es el caso del modelo de diferencias en diferencias, podemos evaluar una política pública de mejora al Bienestar Social en las poblaciones aledañas a las estaciones del tren. Parecería que este tipo de proyectos no tienen impacto social, pues al rehabilitar un ferrocarril la primera idea que aparece es su obra física (sus rieles, vagones y locomotoras). Sin embargo, la evidencia que aquí se expone demuestra todo lo contrario.

El proyecto de rehabilitación del tren cobra relevancia al ser acompañado de políticas de apoyo a la creación de empleos vinculados al turismo. Pero para verificar su trascendencia debe medirse su impacto, a través del modelo de diferencias en diferencias que, si bien es un método muy conocido, es útil para verificar el impacto positivo en ingresos del grupo focal (o de tratamiento) en relación al grupo de control. Comparando el primer grupo con el segundo, se establece que han mejorado sus ingresos desde el año 2010 al 2012 en un 8.66%, éste es un resultado estadísticamente significativo e indica que proyectos de infraestructura que fomentan el turismo, tienen efectos positivos en las poblaciones beneficiadas.

Al comprobarse, a través de los tests de medias, que las variables educación y edad no eran completamente homogéneas (supuesto muy importante en evaluación de impacto), se ha corregido esta deficiencia mediante iteraciones de la variable D con las variables que tenían este problema; el resultado fue favorable porque el modelo se ajustó, las desviaciones estándar disminuyeron y la r cuadrada aumentó.

Otro aporte que presenta esta investigación es el de abordar los impactos de las políticas públicas en relación a la mejora de la distribución de los ingresos en los grupos de tratamiento. En ese sentido se ha calculado el índice de Gini en términos

de impacto, éste muestra que el proyecto de rehabilitación de la línea férrea entre Quito y Guayaquil, tuvo un efecto positivo en la distribución de los ingresos entre población de tratados y no tratados, puesto que la desigualdad disminuyó desde el 2010 al 2012 en -1.91%. Este hecho se debe a que los porcentajes acumulados de poblaciones y rentas están distribuidas de manera muy pareja.

Además, al aplicar el índice de Atkinson en la distribución de ingresos para las poblaciones más pobres, los resultados sugieren que este grupo, en términos de impacto, entre los años 2010 y 2012, disminuyeron el grado de equidad en 12.39%, es decir el proyecto no ayudó a las colas más bajas de las rentas a ser más equitativas. Por el contrario, cuando se realiza el mismo análisis, pero para personas con rentas más altas, el índice de Atkinson nos dice que la distribución de los ingresos mejoraron en 0.10%, con esto evidenciamos que el proyecto ha permitido incrementar la mejora de la distribución de los ingresos pero a las poblaciones con ingresos superiores.

Siendo una debilidad de este trabajo no poseer un *before* puro, es decir al no tener datos de corte transversal tomados antes de la iniciación del proyecto, sino cuando el mismo avanzó en un 3%, resulta importante probar con un método alternativo que incluya un modelo aleatorio como base de datos *before* y que sea anterior a la ejecución del proyecto, esto es factible de acuerdo a recientes investigaciones realizadas por Botosaru y Gutiérrez (2015), donde a través del Método Generalizado de Momentos se ha logrado medir el impacto de políticas públicas sin poseer una base *ex ante*, los resultados serían de utilidad para compararlos con los obtenidos en esta investigación. Adicionalmente, es recomendable, que en futuros análisis se considere la actualización de la base de datos de corte transversal, se la incluya como un nuevo período en las ya existentes y se analice la evolución del impacto al largo plazo.

Al articular un proyecto de rehabilitación del tren a la mejora de las actividades turísticas en zonas rurales, se ha definido una política pública favorable, pues ha permitido mejorar los ingresos de las personas que poseen negocios y emprendimientos a los alrededores de las estaciones y, de cierta manera, ha colaborado en la mejora de la distribución de los ingresos. Por lo tanto, este trabajo investigativo recomienda replicar políticas públicas semejantes en otras rutas y lugares donde se pueda sacar provecho de los entornos turísticos, dando énfasis en el apoyo a personas con escasos recursos económicos y con menor grado de educación.

BIBLIOGRAFÍA

- Asociación Valenciana de Empresarios. (2011). *Estudio del impacto económico de las inversiones del corredor ferroviario mediterráneo en la comunidad valenciana*.
- Altés, C. (2006). *El turismo en América Latina y el Caribe y la experiencia del BID*. Inter-American Development Bank.
- Banerjee, A. V., & Duflo, E. (2007). *The economic lives of the poor*. *The journal of economic perspectives*, 21(1), 161-165.
- Bernal, R., Peña, X., Josse, C., Cuesta, F., Navarro, G., Barrena, V., ... & Saito, J. (2011). *Guía práctica para la evaluación de impacto* (No. 361.25 B517). Universidad de Los Andes, Bogotá (Colombia).
- Botosaru, I., & Gutierrez, F. H. (2014). *Difference-in-Differences When Treatment Status is Observed in Only One Period*. Available at SSRN 2377483.
- Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2011). *Impact Evaluation in Practice; La Evaluación de Impacto en la Práctica*. World Bank Publications.
- Guataquí, J. C., García, A. F., & Rodríguez, M. (2009). *Estimaciones de los determinantes de los ingresos laborales en Colombia con consideraciones diferenciales para asalariados y cuenta propia*. Documentos de Trabajo, 70, 1-22.
- Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on impact evaluation: quantitative methods and practices*. World Bank Publications.
- Medina, H. (2001). *Consideraciones sobre el índice de Gini para medir la concentración del ingreso*. CEPAL.
- Phakdisoth, L., & Kim, D. (2007). *The determinants of inbound tourism in Laos*. ASEAN Economic Bulletin, 24(2), 235-237.
- Piketty, T., & Cazenave, E. (2014). *El capital en el siglo XXI*. Fondo de Cultura Económica.

- Prud'homme, R., Gravari-Barbas, M., Jacquot, S., Talandier, M., Nicot, B. H., & Odzirlik, B. (2008). *Les impacts socio-économiques de l'inscription d'un site sur la liste du patrimoine mondial: trois études*. Unpublished report prepared for the UNESCO World Heritage Committee. Available from the author of this study.
- Ravallion, M. (2005). *Evaluating anti-poverty programs*. World Bank Policy Research Working Paper Series.
- Saarinen, J., & Manwa, H. (2008). *Tourism as a Socio-Cultural Encounter: Host-Guest Relations in Tourism Development in Botswana*. Botswana Notes and Records, 43-53.
- Stock, J. H., & Watson, M. W. (2003). *Introduction to econometrics* (Vol. 104). Boston: Addison Wesley.
- Ureña, J. M., Menéndez, J. M., Guirao, B., Escobedo, F., Rodríguez, F., Coronado, J. M., ... & Martínez, Á. (2005). *Alta velocidad ferroviaria e integración metropolitana en España: el caso de Ciudad Real y Puertollano*. EURE (Santiago), 31(92), 87-104.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Nelson Education.

ANEXOS

Anexo 1: Deducción Analítica del estimador de Impacto del Programa por Mínimo Cuadrados Ordinarios en el Modelo de Diferencias en Diferencias

Tenemos el estimador de Impacto en primera diferencia del Proyecto denominado D_i , ésta es una variable dummy que toma el valor de 1 si es población de tratamiento (población focal) y cero si no lo es, puede ser estimada a través del parámetro β_1 en la siguiente ecuación:

$$Y_i = \beta_0 + \beta_1 D_i + u_i \quad (a)$$

Despejamos el término de error y luego resolvemos el problema de minimización con el fin de conocer los estimadores de MCO β_0 y β_1 .

$$\hat{u}_i = Y_i - \hat{\beta}_0 + \hat{\beta}_1 D_i$$

$$\text{Min}_{\beta_0, \beta_1} \sum_{i=1}^N \hat{u}_i^2 = \text{Min}_{\beta_0, \beta_1} \sum_{i=1}^N (Y_i - \hat{\beta}_0 + \hat{\beta}_1 D_i)^2$$

Adicionalmente definimos N como el tamaño total de la muestra, N_T es el número total de individuos tratados y N_C es el número de individuos de control, en consecuencia tenemos:

$$N = N_T + N_C$$

$$N_T = \sum_{i=1}^N D_i$$

$$N_C = \sum_{i=1}^N Y_i(1 - D_i)$$

Por otra parte, definimos \bar{Y} como el promedio muestral de la variable de resultado; \bar{Y}_T es el promedio de la variable de resultado en el grupo de los tratados y \bar{Y}_C es el promedio de la variable de resultado en el grupo de control:

$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$$

$$\begin{aligned}\bar{Y}_T &= \frac{\sum_{i=1}^N Y_i D_i}{N_T} \\ \bar{Y}_C &= \frac{\sum_{i=1}^N Y_i (1 - D_i)}{N_C} \\ \bar{Y} &= \frac{N_T}{N} \bar{Y}_T + \frac{N_C}{N} \bar{Y}_C \quad (b)\end{aligned}$$

La condición de primer orden para el intercepto β_0 se calcula como sigue:

$$\begin{aligned}\frac{\partial u_i}{\partial \beta_0} &= \sum_{i=1}^N 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 D_i)^2 (-1) = 0 \\ \sum_{i=1}^N Y_i - \sum_{i=1}^N \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^N D_i &= 0 \\ \frac{\sum_{i=1}^N Y_i - N \hat{\beta}_0 - \hat{\beta}_1 N_T}{N} &= 0 \\ \bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \frac{N_T}{N} &= 0 \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \frac{N_T}{N} \quad (c)\end{aligned}$$

Las condición de primer orden para la pendiente β_1 se calcula como sigue:

$$\begin{aligned}\frac{\partial u_i}{\partial \beta_1} &= \sum_{i=1}^N 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 D_i)^2 (-D_i) = 0 \\ \sum_{i=1}^N Y_i D_i - \hat{\beta}_0 \sum_{i=1}^N D_i - \hat{\beta}_1 \sum_{i=1}^N D_i^2 &= 0 \\ \sum_{i=1}^N Y_i D_i - \hat{\beta}_0 N_T - \hat{\beta}_1 N_T &= 0 \quad (d)\end{aligned}$$

Reemplazando (c) en (d) tenemos:

$$\begin{aligned}\sum_{i=1}^N Y_i D_i - \left(\bar{Y} - \hat{\beta}_1 \frac{N_T}{N} \right) N_T - \hat{\beta}_1 N_T &= 0 \\ \frac{\sum_{i=1}^N Y_i D_i}{N_T} - \bar{Y} \frac{N_T}{N_T} + \frac{\hat{\beta}_1 \frac{N_T}{N}}{N_T} - \frac{\hat{\beta}_1 N_T}{N_T} &= 0\end{aligned}$$

$$\bar{Y}_T - \bar{Y} = \hat{\beta}_1 \left(\frac{N - N_T}{N} \right)$$

$$\bar{Y}_T - \bar{Y} = \hat{\beta}_1 \frac{N_C}{N} \quad (e)$$

Reemplazando (b) en (e) tenemos:

$$\bar{Y}_T - \left(\frac{N_t}{N} \bar{Y}_T + \frac{N_c}{N} \bar{Y}_C \right) = \hat{\beta}_1 \frac{N_C}{N}$$

$$\frac{N \bar{Y}_T - N_t \bar{Y}_T}{N} - \frac{N_c}{N} \bar{Y}_C = \hat{\beta}_1 \frac{N_C}{N}$$

$$\frac{N_c}{N} \bar{Y}_T - \frac{N_c}{N} \bar{Y}_C = \hat{\beta}_1 \frac{N_C}{N}$$

$$\bar{Y}_T - \bar{Y}_C = \hat{\beta}_1 \quad (f)$$

La ecuación (f) nos muestra la pendiente de la ecuación (a), en este caso es la diferencia promedio entre población de tratamiento y control en el período ex ante, es decir es la línea de base o tiempo t=0. Por lo tanto, también se la puede expresar de la siguiente manera:

$$E(Y_0|D = 1) - E(Y_0|D = 0) = \hat{\beta}_1 \quad (g)$$

La ecuación (g) está expresando en términos de tiempo y de tipo de población la ecuación (f), es decir $E(Y_0|D = 1)$ es el valor esperado de la variable dependiente Y_0 en el tiempo t=0, pero en la población de tratamiento; y $E(Y_0|D = 0)$ es nuevamente el valor esperado de la variable Y_0 en el tiempo t=0, pero en la población de control.

Ahora bien, para establecer el período t=1, seguimos todos los pasos anteriores (desde el paso a hasta el f), utilizando un modelo MCO con poblaciones de tratamiento y control, se deduce la ecuación a la que denominaremos Y:

$$E(Y_1|D = 1) - E(Y_1|D = 0) = \hat{\gamma}_1 \quad (h)$$

Por ende el modelo de diferencias en diferencias resulta de la diferencia entre la ecuación (g) y (h):

$$[E(Y_0|D = 1) - E(Y_0|D = 0)] - [E(Y_1|D = 1) - E(Y_1|D = 0)] = \hat{\beta}_1 - \hat{\gamma}_1 \quad (i)$$

Reordenando la ecuación anterior tenemos:

$$[E(Y_1|D = 1) - E(Y_0|D = 1)] - [E(Y_1|D = 0) - E(Y_0|D = 0)] = \alpha_3 \quad (j)$$

La ecuación (j) representa el modelo de evaluación de impacto y muestra dos diferencias: la de poblaciones de tratamiento y control en períodos ex ante y ex post, estando representada dicha diferencia por el parámetro α_3 en la siguiente ecuación:

$$Y = \alpha_0 + \alpha_1 D_i + \alpha_2 t_i + \alpha_3 D_i * t_i + w_i (k)$$

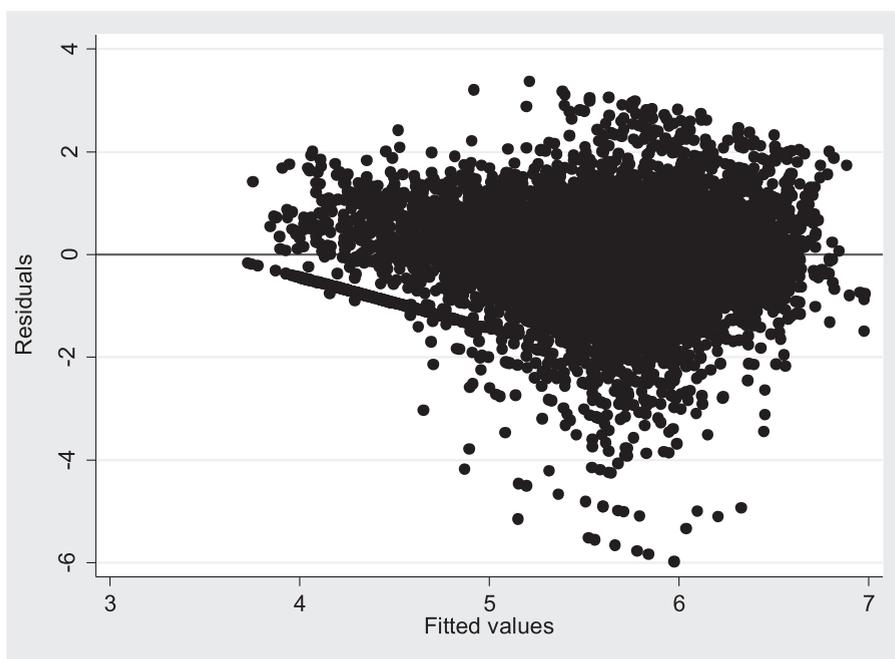
La ecuación (K) representa el modelo de diferencias en diferencias siendo el parámetro α_3 la variable de interés y que ha sido utilizada para medir el efecto total de la política pública en dos períodos de tiempo, comparando la población de tratamiento y control.

Anexo 2: Prueba de multicolinealidad para el modelo 3		
Variable	VIF (a)	1/VIF (b)
Edad	37.53	0.026646
Edad^2	34.46	0.029021
D	9.56	0.1046
Años de educación*D	8.95	0.111775
T	5.44	0.183871
Dxt	4.99	0.20059
Género*t	2.44	0.410256
Género	1.53	0.654088
Bono de Desarrollo Humano	1.48	0.675935
Bono de Desarrollo Humano*t	1.33	0.749746
Años de educación	1.29	0.775191
Norte	1.19	0.842002
Centro	1.16	0.863691
Estado Civil	1.11	0.900426
Sur	1.08	0.924474
Turismo	1	0.998505
Promedio VIF	7.16	

Pese a que en presencia de multicolinealidad los estimadores pueden ser Mejores Estimadores Lineales Insesgados (es decir óptimos y eficientes), los errores estándar de los coeficientes estimados tienden a ser sobreestimados, facilitando que se acepten a uno o más regresores. En este sentido, en la tabla anterior, a través del factor inflador de varianza, se mide la velocidad con la cual se incrementan las varianzas y las covarianzas de los estimadores y si esta velocidad está sobre estimando los errores estándar, puede analizarse.

En la columna (a) del cuadro anterior mirados los VIF de cada una de las variables del modelo 3, las cuales no pasan de 39, considerándose, por lo tanto, que las variables independientes son combinación lineal de otras variables independientes. Alternativamente, en la columna (b), se observa el índice de Tolerancia (1/VIF), siendo aceptado un índice de tolerancia de hasta 0.33; miramos que ninguna de estas variables sobre pasa dicho número, por lo cual se concluye que no hay multicolinealidad.

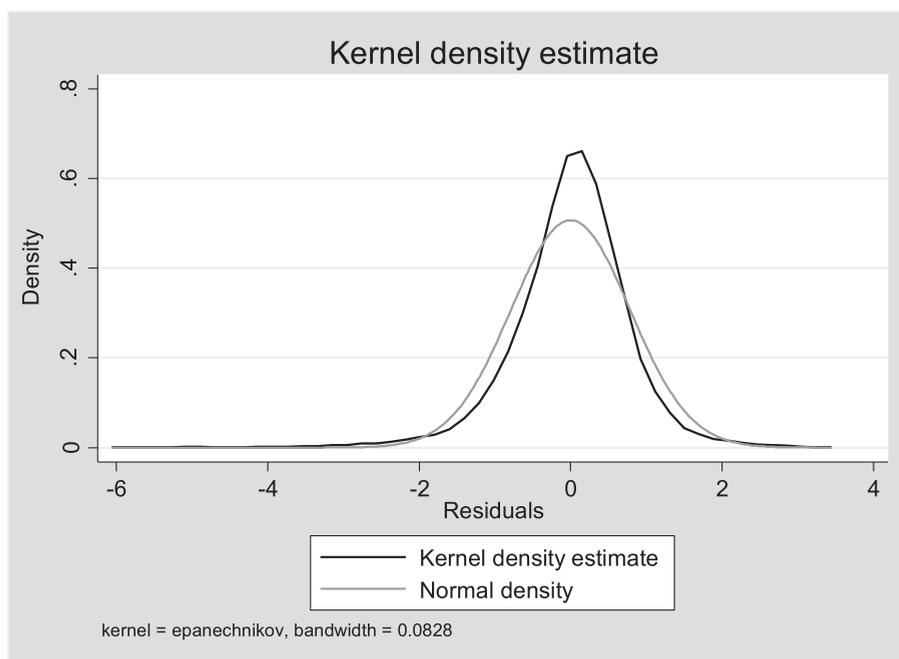
Anexo 3: Prueba para detectar heterocedasticidad en el modelo 3



Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

El gráfico anterior hace una relación de los residuos con los valores ajustados de la variable dependiente; al relacionarlos, lo que se pretende es verificar un posible patrón de comportamiento, que en caso de existir comprobaría la presencia de heterocedasticidad. Al no haber tal patrón definido se evidencia homocedasticidad, considerándose que la varianza de los residuales es constante.

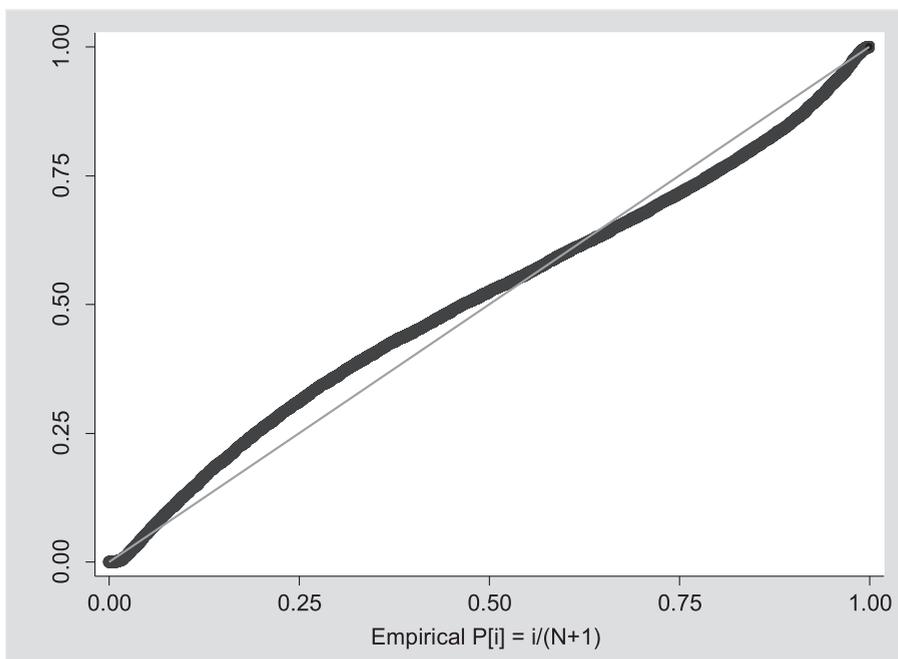
Anexo 4: Prueba kdensity para detectar normalidad en los errores del modelo 3



Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

El gráfico anterior compara la función de densidad de los residuales con una función de densidad normal, se puede observar un leve apuntalamiento y asimetría en los residuales. Sin embargo, existe un ajuste favorable a que los residuos tiendan a tener un comportamiento normal, lo cual favorece a un insesgamiento de los parámetros observados en el modelo 3.

Anexo 5: Prueba para detectar normalidad en los errores a través de cuantiles para el modelo 3



Fuente: Encuestas elaboradas por FEEP entre 2010-2012 y base de datos del INEC en 2010

Esta comprobación gráfica de normalidad contrasta cuantiles de los residuos (eje de las abscisas) contra cuantiles de una distribución normal (en el eje de las ordenadas) cuanto más cerca esté de la línea continua de 45 grados, más cerca están los residuos de seguir un comportamiento normal. En este caso observamos que los cuantiles de los residuos no están muy distantes de asemejarse a los cuantiles de la distribución normal, comprobándose que efectivamente existe la posibilidad que los residuos tengan una distribución normal.

Anexo 6: Criterio de Akeike y Bayesiano para los tres modelos

Modelo (a)	Observaciones (b)	Grados de libertad (c)	Criterio de Akeike (AIC) (d)	Criterio Bayesiano (BIC) (e)
1	13,654	4	36,488.1	36,518.19
2	13,654	16	32,238.32	32,358.67
3	13,654	17	32,233.51	32,361.38

El cuadro anterior muestra en la columna (a) los tres modelos usados en esta investigación, en la (b) el número de observaciones usadas, en la (c) los grados de libertad y en las columnas (d) y (e) los criterios de Akeike y Bayesiano, respectivamente. Los dos criterios nos permiten establecer la pérdida de información que existe en cada uno de los modelos, por lo tanto, mientras menos información se pierda es mejor el modelo. Considerando que los dos criterios no testean bajo un rango establecido de confianza la fiabilidad de los modelos, sino que simplemente deja entrever cuál modelo aporta más información para explicar la variable dependiente, es evidente que el tercero es el mejor de los tres modelos establecidos en esta investigación.