

Evaluación de la política de lectura en el marco del Programa Ilimita

por

Lenin Monak Salinas

Tesis presentada en conformidad con los requisitos
para el Máster en Economía, Finanzas y Computación

Universidad de Huelva & Universidad Internacional de Andalucía

uhu.es

un
i Universidad
Internacional
de Andalucía
A

Noviembre 2018

Evaluación de la política de lectura en el marco del Programa Ilimita

Lenin Monak Salinas

Máster en Economía, Finanzas y Computación

Supervisor/s Concepción Román Díaz y Emilio Congregado Ramirez de Aguilera
Universidad de Huelva y Universidad Internacional de Andalucía

2018

Abstract

Literacy and reading intensification programs are policies and strategies of global importance to reduce poverty and inequality. Within the framework of the Summit of Heads of State and Government of Ibero-America, the Ilimita Program was signed to promote reading and writing in the member countries. The program is evaluated through Discontinuity Regression design in three moments (2005, 2012 and 2017), in Colombia, using different models, trends and impacts. The results show that reading favors the variation of labor income and quantifies an aspect of the sociological theory of the effects of reading on the economic development of individuals.

JEL classification: B4, C31, C52, I25, J08, J24, J31.

Key words: Regresión Discontinua, habilidad lectora, ingresos laborales, Programa Ilimita.

Resumen

Los programas de alfabetización e intensificación de la lectura son políticas y estrategias de importancia global por disminuir la pobreza y la desigualdad. En el marco de la Cumbre de Jefes de Estado y de Gobierno de Iberoamérica se suscribió el Programa Ilimita para promover la lectura y la escritura en los países miembros. Se evalúa el programa a través de diseño de Regresión Discontinua en tres momentos (2005, 2012 y 2017), en Colombia, utilizando diferentes modelos, tendencia e impactos. Los resultados muestran que la lectura favorece la variación de los ingresos laborales y cuantifica un aspecto de la teoría sociológica de los efectos de la lectura en el desarrollo económico de los individuos.

Tabla de Contenidos

1. -Introducción	p.1
2.-Literatura previa	p.3
3.- Pregunta o hipótesis	p.7
4.- Estrategia econométrica y datos	p.9
5.- Resultados	p.12
6.- Conclusiones	p.19
Referencias	p.21

Lista de Tablas

Tabla 1. Variación de los ingresos mensuales por índice lector (variable de asignación) en Regresión Discontinua (2005)

Tabla 2. Variación de los ingresos mensuales por índice lector (variable de asignación) en Regresión Discontinua (2012)

Tabla 3. Variación de los ingresos mensuales por índice lector (variable de asignación) en Regresión Discontinua (2017)

Tabla 3. Características generales de los lectores ocasionales y los habituales.

Lista de figuras

- Figura 1.** Punto de corte y evaluación del impacto. Fuente: (Lee & Lemieux, 2010)
- Figura 2.** Valor esperado, extrapolación y principio de continuidad.
- Figura 3.** (2005). Ingresos laborales, índice de lectura y punto de corte. Fuente: DANE.
- Figura 4.** (2012). Ingresos laborales, índice de lectura y punto de corte. Fuente: DANE.
- Figura 5.** (2017) Ingresos laborales, índice de lectura y punto de corte. Fuente: DANE.
- Figura 6.** (2005) Curva cuantil índice lector. Fuente: DANE.
- Figura 7.** (2012) Curva cuantil índice lector. Fuente: DANE.
- Figura 8.** (2017) Curva cuantil índice lector. Fuente: DANE.
- Figura 9.** (2005) Regresión discontinua y tendencia lineal. Fuente: DANE.
- Figura 10.** (2012) Regresión discontinua y tendencia lineal. Fuente: DANE.
- Figura 11.** (2017) Regresión discontinua y tendencia lineal. Fuente: DANE.
- Figura 12.** (2005) Regresión discontinua con tendencia cuadrática. Fuente: DANE.
- Figura 13.** (2012) Regresión discontinua con tendencia cuadrática. Fuente: DANE.
- Figura 14.** (2017) Regresión discontinua con tendencia cuadrática. Fuente: DANE.

1 Introducción

En el siglo que transcurre las iniciativas de alfabetización y educación en el ámbito mundial se fortalecieron con la declaración de las Naciones Unidas “El decenio de las Naciones Unidas de la Alfabetización” (Page, 2004); con la preparación de la agenda de la educación después del 2015 en la *37 Conferencia General de la Unesco* (Unesco, 2013) y con los planteamientos del área de educación de los *Objetivos de Desarrollo Sostenible* (PNUD, 2016). Estas iniciativas exponen, comprenden y conceptualizan, entre otras, la lectura como un elemento formador de personas con capacidades, con habilidades y con actitudes que se desarrollan y usan para seguir aprendiendo y para seguir formándose.

Estas competencias se conciben como instrumentos reales para la inclusión social, para el desarrollo cultural y el desarrollo económico, siendo la falta de uso motivo de pérdida de competencias académicas y laborales. Estudios y estadísticas de la Unesco muestran la relación cercana entre pobreza y lectura deficiente o analfabetismo, como un reflejo de inequidad social, cultural y laboral (Page, 2004, págs. 37, 63, 67-84) (Unesco, 2006).

En 2003, la Cumbre de Jefes de Estado y de Gobierno de Iberoamérica suscribieron el Programa Ilimita (OEI, 2006) para emprender una acción articulada a largo plazo a favor de la promoción de la lectura y la escritura. Colombia acometió estas iniciativas años antes de la Cumbre al darse condiciones de medición con la primera encuesta de lectura (Jaramillo, Rey, Melo, & otros, 2001); de promoción de la lectura *Consejo Nacional del Libro y la Lectura* (Congreso de la República, 2002) y *Colombia crece leyendo (2001)* (Fundalectura, 2001) y normativos con los lineamientos del *Plan Nacional de Lectura y Bibliotecas* (DNP, 2003). Firmado el Programa Ilimita se integran nuevas acciones como la realización de la segunda medición de la lectura en 2005 (Jaramillo, Rey, & otros, Hábitos de lectura, asistencia a bibliotecas y consumo de libros en Colombia, 2006), se formaliza el *Plan Nacional de Lectura y Bibliotecas* (Congreso de la República, 2003), se expiden lineamientos adicionales de fomento a la lectura (Alcaldía Mayor de Bogotá, 2006), se crea el día de lectura en los parques y establecimientos carcelarios colombianos (Congreso de la República, 2006) y se organiza la Red Nacional de Bibliotecas Públicas (Congreso de la República, 2010).

Con la batería de medidas realizadas en la década de 2000, se implementaron no menos de diez planes de lectura en Colombia, del orden local, nacional, público y privado, Entre los principales están *Leer libera* (Ministerio de Cultura 2003), *Mil maneras de leer* (Ministerio de Educación y Cerlalc-Unesco 2004), *Libro al viento* (Secretaría de Cultura, Recreación y Deporte de Bogotá 2006), *Red Nacional de Escritura Creativa Relata* (Ministerio de Cultura 2006), *Leer es mi cuento* (Ministerio de Cultura y Ministerio de Educación 2010), el *Plan Dice* (Secretaría de Cultura, Recreación y Deporte de Bogotá 2010), *Leer en Familia* (Fundalectura 2012), el *Plan de Lectura, escritura y cultura digital* (Secretaría de Cultura, Recreación y Deporte de Bogotá, 2012), *Lectores Ciudadanos* (Secretaría de Cultura, Recreación y Deporte de Bogotá, 2012) y *Leer es volar* (Secretaría de Cultura, Recreación y Deporte, y la Secretaría de Educación de Bogotá 2016), entre otros.

El abanico de efectos de la lectura o falta de ella es complejo de medir o contar, obligando a limitar y referir a un tema específico la investigación. En este caso se pretende resolver la pregunta ¿la práctica de la lectura ayuda a mejorar los ingresos laborales?

La metodología utilizada para estimar los efectos de la Política Nacional de Lectura en Colombia es el Diseño de Regresión Discontinua -RD-, siendo la variable de asignación el índice de lectura $\{\chi\}$ y la variable de resultado los ingresos laborales mensuales $\{Y\}$. La evaluación se asignará conjuntamente a los individuos $\chi_i \lesseqgtr c$, siendo c el umbral o valor de corte. Las razones para utilizar un diseño de Regresión Discontinua radican en la veracidad de sus resultados por efecto de la reciente formalización del procedimiento y sus supuestos, especialmente, relativos de aleatorización de la variable de asignación y de la limitación o incapacidad de los agentes para manipular la variable de asignación (Lee & Lemieux, 2010, pág. 282).

Se utilizan los microdatos de los módulos de lectura de la Encuestas de Hogares de 2005 (13.972 registros), de la Encuesta de Consumo Cultural de 2012 (8.286 registros) y la Encuesta Nacional de Lectura de 2017 de Colombia (27.641 registros).

2 Literatura previa

El punto de origen de la metodología del diseño de Regresión Discontinua se ubica con el trabajo de Thistlethwaite y Campbell (Thistlethwaite & Campbell, 1960) al evaluar el impacto en la asignación de beneficios a los estudiantes con mejores resultados académicos. El estudio consistió en definir una discontinuidad c por concepto de los estudiantes con puntajes $\chi_i \geq c$ en las pruebas, recibiendo el beneficio o tratamiento, mientras que los estudiantes con puntajes por debajo $\chi_i < c$ no recibieron el beneficio. Se crea una variable dummy $D \in \{0, 1\}$, siendo $D = 1$ cuando $\chi_i \geq c$ y $D=0$ cuando $\chi_i < c$. El valor de χ son las notas obtenidas (variable de asignación). La estimación del impacto del beneficio se determinó como la diferencia en la variable de resultado $\{Y\}$ entre una puntuación individual con $D = 1$ y una puntuación individual contrafactual de $D = 0$, asumiendo que factores adicionales al beneficio recibido sobre el valor de la variable de asignación $\{\chi\}$ se mantienen constantes.

La forma simple del efecto del tratamiento es

$$Y = \alpha + D\tau + \chi\beta + \varepsilon$$

Donde D es el estatus del tratamiento siendo $D = 1$ cuando $\chi_i \geq c$ y $D=0$ cuando $\chi_i < c$; τ mide el efecto medio del tratamiento en c ATE (Average Treatment Effect); $\{\chi\}$ es la variable de asignación y ε es el término del error aleatorio.

Thistlethwaite y Campbell basaron un primer resultado ATE extremo, que sirvió de base para el desarrollo posterior, al asumir que un individuo había obtenido un puntaje $\chi = B'$ siendo $B' \geq c$ y por lo tanto recibir el beneficio y, al mismo tiempo, el individuo haber obtenido un puntaje $\chi = A''$ siendo $A'' < c$, siendo excluido del beneficio (Figura 1) (Lee & Lemieux, 2010, págs. 286-287). La estimación del impacto del beneficio es la diferencia entre $\tau = B' - A''$. Dos hechos se desprenden del análisis de Thistlethwaite y Campbell: en primer lugar, no es posible observar para un mismo individuo los valores de Y_{i1} cuando $D = 1$ y de Y_{i0} cuando $D=0$, lo que lleva a observar el efecto promedio del tratamiento ATE para cada grupo poblacional que se encuentre a la izquierda y derecha del punto de corte. En segundo lugar, la preferencia que los valores de $\chi \lesseqgtr c$ se encuentren lo más cerca del punto de corte acarrea la posibilidad de mayor sesgo en los estimadores por el menor número de datos disponibles.

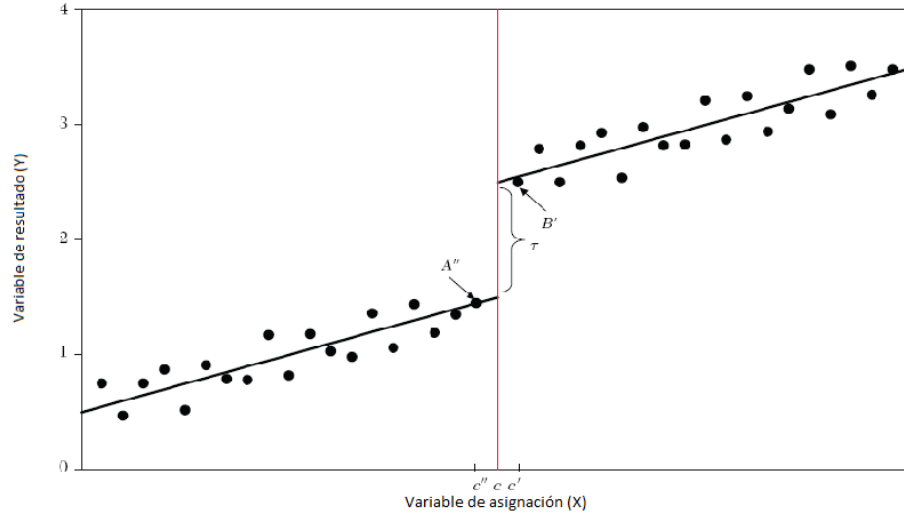


Figura 1. Punto de corte y evaluación del impacto. Fuente: (Lee & Lemieux, 2010)

En 2001 Hahn, Todd, y van der Klaauw contribuyeron con el principio de continuidad en los demás factores con respecto a $\{\chi\}$ como supuesto esencial para el ATE sin asumir la linealidad en la evaluación del efecto causal (Hahn, Petra, & Wilbert, 2001). El principio de continuidad parte en la imposibilidad de observar simultáneamente al mismo individuo con los valores de Y_{i1} cuando $D = 1$ y de Y_{i0} cuando $D=0$. El procedimiento es calcular el efecto causal promedio para las poblaciones Y_{i1} y Y_{i0} , llevando consigo las relaciones de Y_i con respecto a χ :

$$E[Y_{i1} | \chi] \text{ y } E[Y_{i0} | \chi]$$

En cada lado del punto de corte se observan los individuos que no reciben tratamiento $E[Y_{i0} | \chi]$ y los individuos que reciben tratamiento $E[Y_{i1} | \chi]$, pudiéndose extrapolar los resultados por el principio de continuidad (Figura 2) (Hahn, Petra, & Wilbert, 2001, págs. 287-289)

$$E[Y_{i1} | \chi_i = c] = \lim_{\varepsilon \rightarrow 0} E[Y_{i1} | \chi_i = c + \varepsilon]$$

$$E[Y_{i0} | \chi_i = c] = \lim_{\varepsilon \rightarrow 0} E[Y_{i0} | \chi_i = c + \varepsilon]$$

El ATE en el punto de corte sería resultado de

$$\tau_{ATE} = \lim_{\varepsilon \rightarrow 0} E[Y_{i1} | \chi_i = c + \varepsilon] - \lim_{\varepsilon \rightarrow 0} E[Y_{i0} | \chi_i = c + \varepsilon]$$

$$\tau ATE = E[Y_{i1} | \chi_i = c] - E[Y_{i0} | \chi_i = c]$$

$$\tau ATE = E[Y_{i1} - Y_{i0} | \chi_i = c]$$

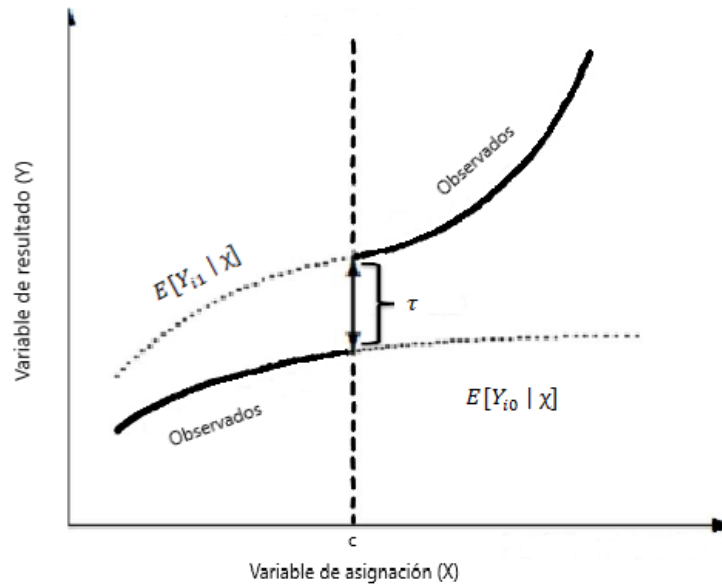


Figura 2. Valor esperado, extrapolación y principio de continuidad.

Por el supuesto de continuidad es posible utilizar el resultado promedio de los individuos que se encuentran por encima del umbral o punto de corte Y_{i1} como contrafactual de los individuos por debajo del punto de corte Y_{i0} .

Recientemente Lee y Lemieux (Lee & Lemieux, 2010, págs. 293-295) aportaron otros elementos de validez al diseño de Regresión Discontinua, al probar la similitud de las condiciones de aleatorización de un experimento aleatorio con las de uno de Regresión Discontinua. Lee y Lemieux definieron que los individuos no tienen un control preciso sobre la variable de asignación $\{\chi\}$ cuando están condicionadas al vector de todas las características predeterminadas y observables W_i y al vector de las características predeterminadas inobservables U_i , que pueden afectar el resultado $\{Y_i\}$. Estos vectores en consecuencia son variables aleatorias. La falta de control lleva a la implicación que las variaciones a la derecha del punto de corte $\chi_i \geq c$ sean aleatorias en un segmento cercano al umbral. Lee y Lemieux lo exponen con la Regla de Bayes:

$$Pr[W = w, U = u | \chi = x] = f(x | W = w, U = u) \frac{Pr[W = w, U = u]}{f(x)}$$

Siendo $f(\cdot)$ y $f(\cdot | \cdot)$ densidades condicionadas y marginales de $\{\chi\}$. Si $f(x | W = w, U = u)$ es continua en $\{\chi\}$, los valores a la derecha del punto de corte $\chi_i \geq c$ serán continuos en $\{\chi\}$. En consecuencia, las características predeterminadas observables W_i y las características predeterminadas inobservables U_i tendrán distribuciones idénticas en cada lado del punto de corte, ya sea cercanas o no al umbral. Lo destacable del trabajo de Lee y Lemieux es la positiva valoración del tratamiento al nivel de una asignación aleatoria alrededor el punto de corte por la falta de control de los individuos sobre la variable de asignación y del supuesto de continuidad del $Pr[W = w, U = u | \chi = x]$. En conclusión, la predicción del tratamiento es aleatorio localmente.

3 Pregunta o hipótesis

La relación entre educación y mercado laboral son un tópico de valía en el entendimiento del bienestar, del desarrollo y de las políticas públicas, que ocupan cada cierto tiempo especial importancia por la disrupción en los modelos de negocio, en los sistemas productivos / operativos y en las mejoras tecnológicas, entre otros. Una de las líneas de interés ha sido explorar las causas de las disparidades entre los trabajadores con el mismo nivel educativo y las diferencias salariales (Tazeen, 2008) con miras en adelantar acciones tendientes en aumentar los retornos de la educación, atender la demanda de trabajo especializado y disminuir las tasas de pobreza y/o aumentar los ingresos de los menos favorecidos (Lin, Lutter, & Ruhm, 2018).

Los estudios de la disparidades entre educación y nivel salarial contemplan como un elemento causal los atributos tangibles de los individuos como años de educación, experiencia laboral, habilidades, salud y atributos intangibles como motivación, cooperación, fiabilidad, emprendimiento, juicio, etc. Atributos que se adquieren y/o mejoran a través de la educación formal o por prácticas informales (Shomos, 2010). La teoría indica que la práctica de la lectura desde la infancia, como componente de las destrezas de alfabetización, contribuye al desarrollo de las habilidades necesarias para la vida laboral, por la mayor probabilidad de participación en el mercado laboral unido a salarios más altos. La información existente de la relación entre habilidades de alfabetización y aritmética con mercado laboral son relativamente limitada, destacándose tres áreas de influencia:

Un primer campo, son los trabajo orientados en discernir las habilidades de alfabetización como elemento necesario en el desempeño individual en el trabajo y de perspectivas salariales (Rivera-Batiz, 1990) asociadas con la destreza del lenguaje en adultos jóvenes e inmigrantes en diferencias salariales (Charette & Meng, 1994). Un segundo campo, son los estudios que abordan la alfabetización, las habilidades cognitivas y los retornos de la escolarización (Fasih, Patrinos, & Sakellariou, 2013), encontrándose rendimientos crecientes con la edad (Lin, Lutter, & Ruhm, 2018), las influencias cognitivas y no osbservadas en los ingresos (Corrado, 2013) (Green & Riddell, 2003) y la visión de los retornos desde una persectiva multidimensional en la identificación de sus principales componentes (Lee & Nagaraj, 1988). El último campo, dirige el énfasis en los efectos de la alfabetización en el desarrollo económico, la vida productiva y las mejoras en los programas de alfabetización (Wagner, 1995) (Judge, 2004); con mayores retornos

de la educación cuando hay cambios económicos rápidos (Hanushek, Schwerdt, & Wiederhold, 2017), y en la relación de los tipos de empleo y los salarios en países desarrollados (Dumont, 2006).

El carácter común, de los estudios y modelos econométricos relacionados, son los beneficios de las habilidades de lecto-escritura en la probabilidad de participación laboral, en la consecución de salarios mejor remunerados, en trabajos cualificados, en mejoras de productividad y en positivos efectos en el desarrollo y crecimiento económico.

La investigación se orienta, al primer tópico, en la relación de la habilidad y práctica de lectura con efectos en los ingresos en el marco de la política pública de hacer una sociedad lectora: **¿la práctica de la lectura ayuda a mejorar los ingresos laborales?**

4 Estrategia econométrica y datos

La estimación de los efectos del Programa Ilimita en Colombia empiezan considerando los supuestos y elementos básicos del diseño de Regresión Discontinua, abstraídos por Lee y Lemieux (Lee & Lemieux, 2010, págs. 283-285), como la imposibilidad de los individuos de manipular la variable de asignación; la improbable intervención en el resultado y en la discontinuidad que no se deba a la asignación al tratamiento y a la variación cerca del valor de corte o umbral. La aleatoriedad de la variable de asignación cerca del umbral permite asumir que las características previas (edad, género, nivel educativo, entre otras) antes de establecer la variable de asignación $\{\chi\}$ tienen la misma distribución estando por encima y por debajo del punto de corte.

Los modelos utilizados son paramétricos por la favorabilidad de utilizar todo el conjunto de datos, la aleatoriedad que ofrece la variable de asignación y la posibilidad de evaluar la calidad de los modelos con AIC y BIC. El primer modelo de regresión estima el efecto del tratamiento como resultado de las diferencias de los segmentos en ambos lados del umbral.

$$Y = \alpha + D\tau + f(\chi - c) + \varepsilon \quad (1.0)$$

Siendo ATE la diferencia de los interceptos de las regresiones del segmento $\chi_i \geq c$ y del segmento $\chi_i < c$, es decir $\tau = \alpha_1 - \alpha_0$. La $f(\chi - c) = f_0(\chi - c) + D[f_1(\chi - c) - f_0(\chi - c)]$.

El segundo modelo, transforma la variable de asignación $\{X\}$ en desviaciones con respecto al umbral, distinguiendo las pendientes por encima y por debajo de la discontinuidad.

$$\tilde{\chi}_i = \chi_i - c$$

$$Y_i = \alpha + D_i\tau + \beta_0(\tilde{\chi}_i) * (D_i = 0) + \beta_1(\tilde{\chi}_i) * (D_i = 1) + \varepsilon_i \quad (2.0)$$

Utilizar polinomios flexibles contribuye a evaluar el Programa reflejando la relación de la variable de resultado $\{Y\}$, y la variable de asignación $\{\chi\}$

$$Y_i = \alpha + D_i\tau + \beta f(\chi_i) + \varepsilon_i$$

$$Y_i = \alpha + D_i\tau + \beta_1\chi_i + \beta_2\chi_i^2 + \beta_3\chi_i^3 + \beta_p\chi_i^p + \varepsilon_i$$

Se profundiza la flexibilidad y los resultados con la distinción en ambos lados del umbral de $f(\chi_i)$

$$Y_i = \alpha + D_i\tau + \beta_0 f(\chi_i)D_i + \beta_1 f(\chi_i)(1 - D_i) + \varepsilon_i$$

$$Y_i = \alpha + D_i\tau + (\beta_{10}\chi_i + \beta_{20}\chi_i^2 + \dots + \beta_{p0}\chi_i^p)(D_i=0) + (\beta_{11}\chi_i + \beta_{21}\chi_i^2 + \dots + \beta_{p1}\chi_i^p)(D_i=1) + \varepsilon_i$$

El tercer modelo, usa la transformación de la variable de asignación $\{X\}$ en desviaciones con respecto al umbral para cada orden del polinomio

$$\tilde{\chi}_i = \chi_i - c$$

De donde el segmento $\chi_i < c$

$$E[Y_{i0}|\chi_i] = \alpha + \beta_{10}\tilde{\chi}_i + \beta_{20}\tilde{\chi}_i^2 + \dots + \beta_{p0}\tilde{\chi}_i^p$$

El segmento $\chi_i \geq c$

$$E[Y_{i1}|\chi_i] = \alpha + \beta_{11}\tilde{\chi}_i + \beta_{21}\tilde{\chi}_i^2 + \dots + \beta_{p1}\tilde{\chi}_i^p$$

El modelo agregado es

$$Y_i = \alpha + D_i\tau + (\tilde{\beta}_{10}\tilde{\chi}_i + \tilde{\beta}_{20}\tilde{\chi}_i^2 + \dots + \tilde{\beta}_{p0}\tilde{\chi}_i^p)(D=0) + (\tilde{\beta}_{11}\tilde{\chi}_i + \tilde{\beta}_{21}\tilde{\chi}_i^2 + \dots + \tilde{\beta}_{p1}\tilde{\chi}_i^p)(D=1) + \varepsilon \quad (3.0)$$

El ATE del Programa en el umbral o punto de corte es el valor obtenido en τ

La elección de la calidad de los modelos polinomiales utilizados resulta del menor valor obtenido entre AIC y BIC. El Akaike Information Criterion (AIC) compara la calidad de los polinomios en sus diferentes ordenes, equilibrando la ventaja que la relación entre $\{Y\}$ y $\{\chi\}$ mejora en la medida que aumenta el orden del polinomio y la desventaja de los aumentos en el orden del polinomio en saturar e incrementar los errores de los parámetros del modelo.

$$AIC = N * \ln(\hat{\sigma}^2) + 2p$$

N = tamaño de la muestra o número de observaciones; $\hat{\sigma}^2$ = error cuadrático medio de la regresión y p = número de parámetros del modelo.

El Bayesian Information Criterion (BIC) es un criterio alternativo y/o complementario al AIC para seleccionar el polinomio evaluado que mejor probabilidad tiene con los datos y número de parámetros dado. El BIC disminuye el riesgo del sobreajuste producido por el aumento del número de orden del polinomio, penalizando el mayor número de parámetros ($p * \log(N)$). El BIC tiene preferencia por modelo polinómicos parsimoniosos o simples, siendo la puntuación más baja del BIC representativa de un modelo de mejor calidad.

$$BIC = -2 * \log \mathcal{L} + p * \log(N)$$

N = Tamaño de la muestra o número de observaciones; $\mathcal{L}$ = es el valor de máxima verosimilitud del modelo; p = Número de parámetros del modelo

4.1 Fuentes de información y bases de datos

Fuente de información: la fuente oficial es el Departamento Nacional de Estadísticas de Colombia (DANE), con las encuestas: Módulo de Lectura de la Encuesta Nacional de Hogares 2005; Módulo de Lectura de la Encuesta de Consumo Cultural 2012 y Encuesta Nacional de Lectura 2017.

Población objetivo: Población lectora en diversos formatos y materiales de lectura.

Periodicidad: Encuestas de estructura: quinquenal (2005, 2012 y 2017).

Cobertura geográfica: Total nacional.

Unidad geográfica: Ciudades capitales

Registros: 2005: 13.972; 2012: 8.286; 2017: 27.641

Variables de estudio:

Logaritmo del Ingreso laboral mensual (variable de resultado)

Índice de lectura (variable de asignación)

La variable de asignación es un índice compuesto ponderado: $indice_lector = \sum_i^n \mathcal{L}_k * w_k$, donde $\mathcal{L}$ representa la lectura de k materiales y w los pesos de los k materiales.

La escala de valoración del índice lector hace referencia al número de libros promedio leídos al año, la lectura de periódicos y revistas, la lectura de otros materiales y la lectura en internet.

- Escala < 0.21 = Lector ocasional
- Escala $0.21 - 0.30$ = Lector moderado
- Escala $0.31 - 0.50$ = Lector frecuente
- Escala ≥ 0.51 = Lector intenso

Para efectos del punto de corte se toma el umbral de 0.21 para distinguir el efecto entre un lector ocasional (escala < 0.21) y un lector habitual (escala ≥ 0.21)

5 «Resultados»

Resultados gráficos

La figura 3, 4 y 5 representan el espectro de los individuos en torno a la variación de los ingresos mensuales y el índice lector, junto con la línea de corte de los lectores ocasionales versus los lectores habituales. También se observa marcas características de lectura de materiales de lectura específicos como revistas, periódicos y lectura en internet.

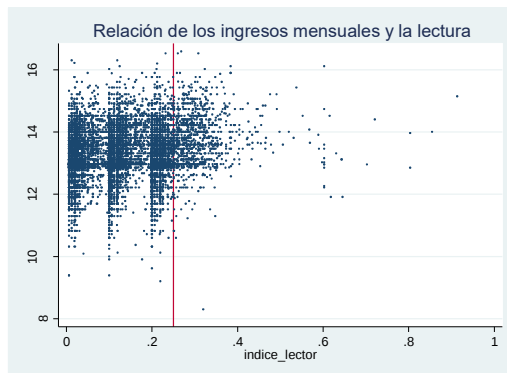


Figura 3. (2005).

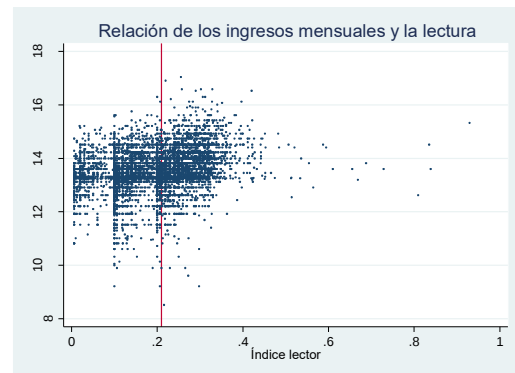


Figura 4. (2012).

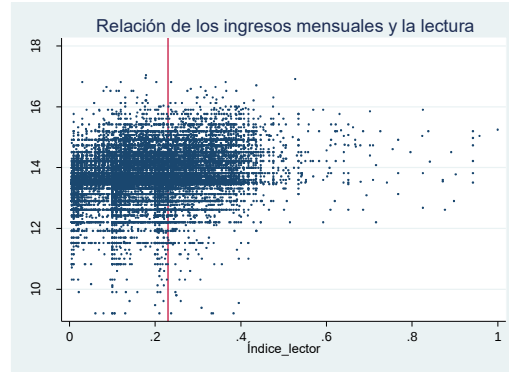


Figura 5. (2017) Ingresos laborales, índice de lectura y punto de corte. Fuente: DANE.

Los lectores ocasionales representan el 80% del total de individuos en 2005 (Figura 4), el 63% en 2012 (Figura 5) y el 64% en 2017 (Figura 6).

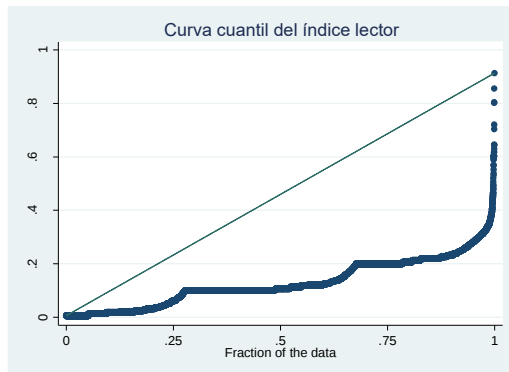


Figura 6. (2005).

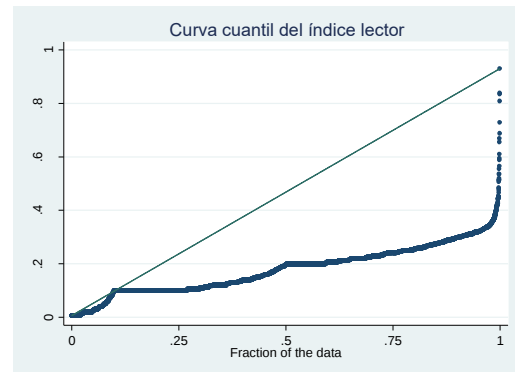


Figura 7. (2012).

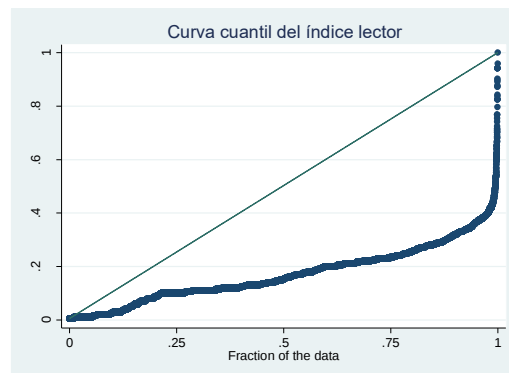


Figura 8. (2017) Curva cuantil índice lector. Fuente: DANE.

La gráfica de la regresión discontinua con tendencia lineal visualiza el efecto de la variación de los ingresos mensuales de los lectores ocasionales y los lectores habituales (Figura 7, 8 y 9))

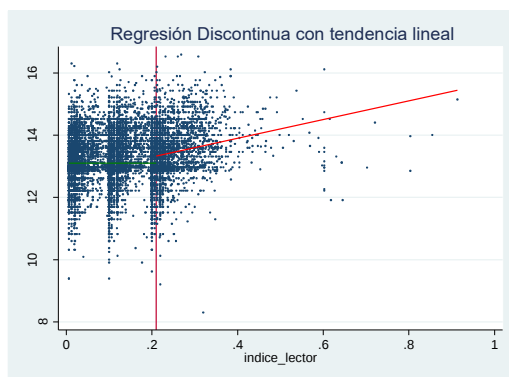


Figura 9. (2005).

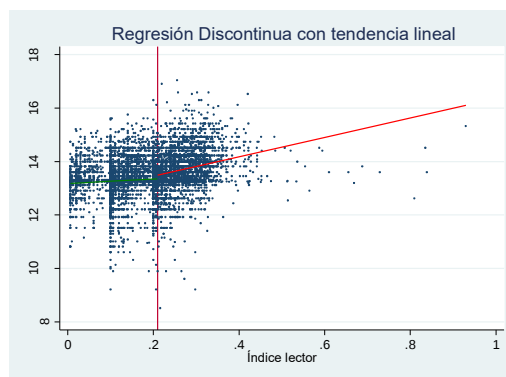


Figura 10. (2012).

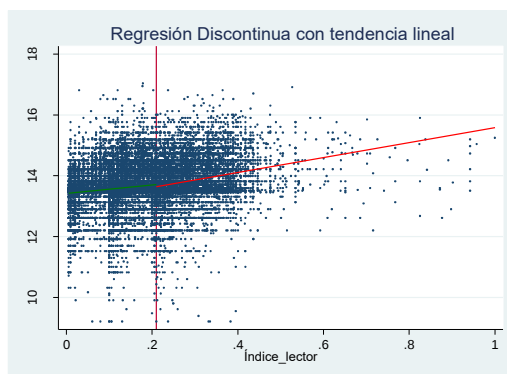


Figura 11. (2017) Regresión discontinua y tendencia lineal. Fuente: DANE.

La representación gráfica del efecto por transformación en desviaciones con respecto al punto de corte y polinómica de orden 2 conserva el efecto en los ingresos mensuales de los lectores habituales con respecto a los ocasionales (Figura 10, 11 y 12))

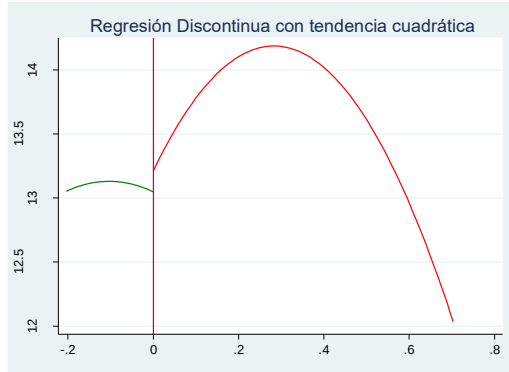


Figura 12. (2005).

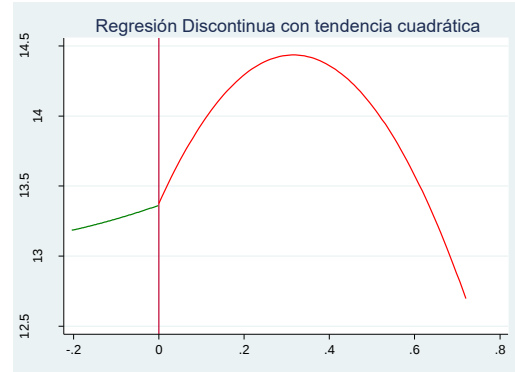


Figura 13. (2012).

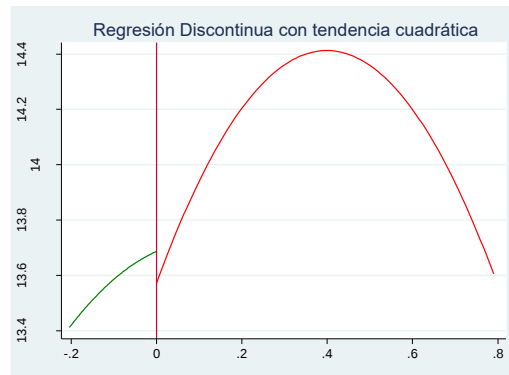


Figura 14. (2017) Regresión discontinua con tendencia cuadrática. Fuente: DANE.

Resultados econométricos

Los modelos diseñados de Regresión Discontinua, para cada año, difieren de acuerdo la calidad del polinomio. Para el año 2005 el polinomio de cuarto orden (modelo 5) arroja el menor valor AIC y BIC con un efecto del tratamiento en la variación de los ingresos mensuales del 70.5% con una varianza de los ingresos laborales explicada por el índice de lectura del 10.1% (Tabla 1). Para el año 2012, el polinomio de cuarto orden (modelo 5) ofrece el menor valor AIC y BIC con un efecto del tratamiento del 45.7% (Tabla 2) explicado por la variable de asignación en 14.9%, finalmente, en 2017, el modelo del polinomio de orden seis (modelo 7) ofrece el menor AIC y BIC con un efecto del tratamiento en la variación de los ingresos mensuales del 15,6% , explicado por la variable de asignación en 8.9%.

Tabla 1. Variación de los ingresos mensuales por índice lector (variable de asignación) en Regresión Discontinua (2005)

	(1) Y	(2) Y	(3) Y	(4) Y	(5) Y	(6) Y	(7) Y	(8) Y
treated	0.275*** (0.0228)	0.223*** (0.0231)	0.166*** (0.0282)	0.141*** (0.0332)	0.705*** (0.0421)	0.634*** (0.0578)	0.00640 (0.0733)	-0.143 (0.0883)
iltreatment	0.635*** (0.1000)							
iltreatment0		0.0382 (0.111)	-1.585*** (0.383)	-6.176*** (1.363)	-74.42*** (3.147)	-65.82*** (6.418)	44.10*** (9.835)	84.91*** (14.16)
iltreatment1		3.013*** (0.222)	6.888*** (0.425)	10.96*** (0.717)	10.24*** (1.252)	11.50*** (1.828)	12.36*** (2.713)	7.002 (3.681)
iltreatme~02			-7.652*** (1.729)	-66.79*** (16.95)	-1458.2*** (60.42)	-1186.4*** (186.6)	3799.3*** (387.1)	6434.3*** (763.0)
iltreatme~12			-12.18*** (1.140)	-42.38*** (4.441)	-33.90*** (13.03)	-55.41* (26.23)	-75.09 (53.13)	89.40 (93.09)
iltreatme~03				-190.2*** (54.23)	-10037.2*** (414.5)	-6744.5** (2179.4)	85126.6*** (6626.8)	154594.1*** (18560.8)
iltreatme~13				42.76*** (6.078)	15.58 (39.78)	136.3 (133.8)	296.1 (398.8)	-1592.5 (964.1)
iltreatme~04					-22709.7*** (948.2)	-5743.9 (11065.1)	797567.4*** (55861.3)	1699386.8*** (231914.8)
iltreatme~14					24.41 (35.32)	-231.8 (273.4)	-781.0 (1320.2)	9114.8 (4785.4)
iltreatme~05						31287.4 (20330.6)	3352192.3*** (227322.6)	9475579.5*** (1545199.4)
iltreatme~15						178.8 (189.2)	1005.1 (1952.8)	-24453.6* (11994.0)
iltreatme~06							5222972.7*** (356111.9)	26086910.3*** (5219824.4)
iltreatme~16							-449.2 (1056.7)	30801.1* (14564.7)
iltreatme~07								28155115.8*** (7027571.2)
iltreatme~17								-14606.4* (6789.6)
_cons	13.17*** (0.0132)	13.10*** (0.0143)	13.05*** (0.0190)	13.00*** (0.0236)	12.44*** (0.0327)	12.50*** (0.0497)	13.12*** (0.0650)	13.31*** (0.0798)
N	13972	13972	13972	13972	13972	13972	13972	13972
R-sq	0.041	0.051	0.060	0.064	0.101	0.101	0.115	0.116
adj. R-sq	0.041	0.051	0.060	0.064	0.101	0.101	0.114	0.115
AIC	31407.3	31266.7	31137.5	31079.8	30520.7	30521.4	30311.6	30294.9
BIC	31430.0	31296.8	31182.7	31140.1	30596.2	30612.0	30417.2	30415.6
rmse	0.744	0.741	0.737	0.736	0.721	0.721	0.716	0.715

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Tabla 2. Variación de los ingresos mensuales por índice lector (variable de asignación) en Regresión Discontinua (2012)

	(1) Y	(2) Y	(3) Y	(4) Y	(5) Y	(6) Y	(7) Y	(8) Y
treated	0.148*** (0.0266)	0.135*** (0.0265)	0.0127 (0.0325)	0.0591 (0.0397)	0.457*** (0.0481)	0.407*** (0.0646)	0.255** (0.0798)	0.187 (0.0961)
iltreatment	1.923*** (0.142)							
iltreatment0		0.890*** (0.179)	1.066* (0.530)	-9.685*** (1.712)	-55.64*** (3.432)	-45.62*** (7.120)	-13.24 (10.94)	1.759 (16.04)
iltreatment1		3.619*** (0.229)	6.722*** (0.390)	10.10*** (0.712)	8.291*** (1.131)	7.130*** (1.696)	5.119* (2.425)	5.418 (3.360)
iltreatme~02			0.951 (2.689)	-141.7*** (21.77)	-1186.3*** (71.25)	-853.0*** (219.5)	721.2 (459.9)	1707.9 (898.3)
iltreatme~12			-10.64*** (1.084)	-36.34*** (4.664)	-14.96 (11.46)	4.458 (24.04)	49.95 (46.01)	40.86 (84.42)
iltreatme~03				-472.1*** (71.51)	-8640.7*** (536.0)	-4419.8 (2683.2)	26654.7** (8417.6)	53483.3* (22607.6)
iltreatme~13				34.77*** (6.141)	-36.18 (35.37)	-144.6 (123.2)	-518.8 (345.5)	-413.8 (887.5)
iltreatme~04					-20356.5*** (1324.0)	2227.1 (14129.1)	291006.2*** (75484.1)	651324.2* (291730.6)
iltreatme~14					64.71* (31.78)	289.1 (246.3)	1591.9 (1150.5)	1039.6 (4449.9)
iltreatme~05						43017.9 (26795.2)	1299273.5*** (323688.6)	3829638.7 (2005232.3)
iltreatme~15						-151.5 (164.9)	-2117.4 (1703.7)	-702.5 (11142.3)
iltreatme~06							2060446.7*** (529078.6)	10961560.5 (6981409.7)
iltreatme~16							1061.9 (916.0)	-656.3 (13403.6)
iltreatme~07								12371362.4 (9675324.3)
iltreatme~17								791.1 (6156.3)
_cons	13.45*** (0.0159)	13.36*** (0.0185)	13.36*** (0.0236)	13.24*** (0.0300)	12.87*** (0.0381)	12.93*** (0.0554)	13.11*** (0.0707)	13.17*** (0.0872)
N	8286	8286	8286	8286	8286	8286	8286	8286
R-sq	0.098	0.107	0.117	0.125	0.150	0.150	0.152	0.152
adj. R-sq	0.097	0.107	0.117	0.125	0.149	0.149	0.151	0.151
AIC	17965.8	17880.3	17788.3	17716.9	17483.5	17484.1	17471.6	17473.9
BIC	17986.8	17908.4	17830.4	17773.1	17553.8	17568.4	17569.9	17586.3
rmse	0.715	0.712	0.708	0.704	0.695	0.694	0.694	0.694

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Tabla 3. Variación de los ingresos mensuales por índice lector (variable de asignación) en Regresión Discontinua (2017)

	(1) Y	(2) Y	(3) Y	(4) Y	(5) Y	(6) Y	(7) Y	(8) Y
treated	-0.0819*** (0.0144)	-0.0688*** (0.0145)	-0.115*** (0.0200)	-0.123*** (0.0253)	0.239*** (0.0309)	0.328*** (0.0375)	0.156*** (0.0458)	0.0425 (0.0556)
iltreatment	1.915*** (0.0639)							
iltreatment0		1.387*** (0.0896)	0.684* (0.334)	-0.917 (0.928)	-37.10*** (1.876)	-52.58*** (3.424)	-19.80*** (5.833)	4.344 (9.211)
iltreatment1		2.458*** (0.0910)	4.215*** (0.169)	5.626*** (0.324)	6.052*** (0.532)	7.642*** (0.779)	7.291*** (1.114)	8.580*** (1.496)
iltreatment-02			-3.238* (1.483)	-23.16* (10.87)	-795.4*** (36.51)	-1311.4*** (102.2)	217.4 (242.9)	1708.1*** (502.7)
iltreatment-12			-5.283*** (0.428)	-14.31*** (1.819)	-18.70*** (4.719)	-42.93*** (9.885)	-35.71 (19.12)	-70.28* (32.90)
iltreatment-03				-64.28 (34.75)	-5852.1*** (263.7)	-12470.8*** (1252.9)	16511.7*** (4361.1)	55470.3*** (12300.7)
iltreatment-13				11.23*** (2.199)	24.43 (13.30)	149.2** (46.67)	96.00 (129.5)	443.1 (298.4)
iltreatment-04					-14003.5*** (632.6)	-49909.0*** (6674.8)	210221.5*** (38083.7)	721373.7*** (155639.6)
iltreatment-14					-11.14 (11.07)	-254.1** (87.83)	-85.68 (392.4)	-1693.8 (1305.8)
iltreatment-05						-69291.9*** (12823.4)	1031937.5*** (159247.6)	4570144.8*** (1056667.6)
iltreatment-15						155.6** (55.78)	-78.77 (535.0)	3604.4 (2902.3)
iltreatment-06							1770061.2*** (255136.9)	14102069.3*** (3649769.1)
iltreatment-16							117.8 (267.5)	-3925.8 (3143.2)
iltreatment-07								17039039.7*** (5030525.9)
iltreatment-17								1696.5 (1313.9)
_cons	13.77*** (0.00843)	13.71*** (0.0107)	13.69*** (0.0164)	13.66*** (0.0215)	13.29*** (0.0271)	13.18*** (0.0337)	13.36*** (0.0421)	13.46*** (0.0522)
N	27641	27641	27641	27641	27641	27641	27641	27641
R-sq	0.062	0.064	0.069	0.070	0.087	0.088	0.089	0.090
adj. R-sq	0.062	0.064	0.069	0.070	0.086	0.087	0.089	0.089
AIC	58170.0	58101.7	57948.8	57923.2	57440.3	57407.4	57363.1	57353.9
BIC	58194.6	58134.6	57998.1	57989.1	57522.6	57506.1	57478.2	57485.5
rmse	0.693	0.692	0.690	0.690	0.684	0.683	0.683	0.683

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

6 «Conclusiones»

El espectro de aplicación del diseño de Regresión Discontinua avanza a terrenos alternativos al área educativa o laboral, encontrándose, además, desarrollos y evaluaciones diferentes al interés en resultados de logros o pruebas (Lee & Lemieux, 2010, págs. 338-343). El trabajo desarrollado desea ampliar el campo de aplicación con tres aspectos. Utilizar como regla de discontinuidad un índice categorizado, fortaleciendo el supuesto de aleatorización con el hecho de no existir un premio en el punto de corte. Segundo, ofrecer una magnitud o valor, a la teoría sociológica, del efecto de la lectura en el bienestar económico a través de los ingresos laborales, sin desconocer las condiciones sociales y culturales de la práctica lectora. Finalmente, analizar la evolución de la política de lectura en tres momentos en 17 años.

La primera conclusión, son los efectos positivos de la cobertura de la política en la creación de hábitos de lectura en la población. En 2005, el 64,8% de la población era lectora, de los cuales el 80% la ejercía de manera ocasional. Tan solo el 20% restante practicaban la lectura de manera habitual. En 2012, la tasa de participación lectora aumenta al 86,6% de la población, de los cuales los lectores ocasionales disminuyen a 63%, mientras que los lectores habituales aumentan a 37%. En 2017, la población lectora aumenta a 94,9% de la población, manteniéndose prácticamente constante la participación de los lectores ocasionales (64%) y habituales (36%).

La lectura es una actividad diversa en prácticas, medios, soportes, géneros, formatos, entre otros. Esta diversidad se impulsó con la disrupción tecnológica del internet, de los lectores digitales, de los formatos electrónicos, de las prácticas fragmentarias y de solapamiento y de la visión amplia al librodeterminismo (Igarza & Monak, 2014, págs. 27 - 33). Prueba de la disrupción es la relación entre la participación de la población lectora y la cantidad de libros leídos en promedio al año. En 2005, en promedio se leyeron 4,3 libros al año para la población de 18 y más años. En 2012, el promedio disminuyó a 3,9 libros, a pesar de haber aumentado la participación de lectores en la población. En 2017, el promedio de libros aumentó a 4,9. Las recientes mediciones contemplan elementos adicionales a la lectura de libros para cualificar a un lector de un no lector, como la lectura en internet de blogs, páginas web, periódicos y revistas digitales, etc. Pese a ello, los libros siguen siendo un referente del lector y de las evaluaciones de políticas en torno a la lectura.

La tabla 4 describe las características lectoras de los grupos poblacionales obtenidos por encima y por debajo del punto de corte es decir entre un lector ocasional y un lector habitual.

Tabla 3. Características generales de los lectores ocasionales y los habituales.

Tipo de lector	Promedio libros leídos al año	Frecuencia de lectura y materiales de lectura
Lector ocasional (índice < 0.21) ¹	2005 3,3	Una o dos veces por semana de periódicos y/o revistas.
	2012: 2,3	
	2017: 2,9	
Lector habitual (índice ≥ 0.21)	2005: 6,9	Todos los días y/o varias veces a la semana de libros, periódicos y revistas.
	2012: 5,3	
	2017: 6,9	
Total	2005: 4,3	
	2012: 3,9	
	2017: 4,9	

Los resultados econométricos destacan en calidad el polinomio de cuarto grado en los años 2005 y 2012. En 2017, el modelo señalado corresponde al polinomio de grado seis. Para el año 2005, un lector habitual recibía 70.5% más de ingresos mensuales que un lector ocasional. En 2012, la diferencia es 45.7% y en 2017 de 15.6% con un porcentaje de varianza de los ingresos laborales explicada por el índice de lectura en un rango de 10.1% a 15.6%. La diferencia porcentual de ingresos entre un lector ocasional y un lector habitual decrece a lo largo del período de análisis, induciendo a indagar a futuro elementos adicionales de explicación.

¹ En 2017 la Enlec amplió la representatividad de las ciudades capitales con sus áreas metropolitanas a las 32 capitales de Colombia. Un efecto de comparabilidad con las encuestas de lectura de 2005 (13 ciudades capitales) y 2012 (ciudades capitales representativas por regiones) es el impacto de los antiguos territorios nacionales. Se encontró una mayor participación de individuos con altos ingresos mensuales y bajos índices de lectura.

7 Referencias

- Alcaldía Mayor de Bogotá. (21 de abril de 2006). Lineamientos de política pública de fomento a la lectura para el período 2006-2016. *Decreto 133 de 2006*. Bogotá, Colombia.
- Charette, M., & Meng, R. (1994). Explaining language proficiency. Objective versus self-assessed measures of literacy. *Economic Letters*, 44, 313-321.
- Congreso de la República. (7 de marzo de 2002). Consejo Nacional del Libro y la Lectura. *Decreto 267 de 2002*. Bogotá, Colombia: Diario Oficial de Colombia.
- Congreso de la República. (26 de junio de 2003). Plan Nacional de Desarrollo 2003-2006 (Ley 812 de 2003). *Plan Nacional de Desarrollo 2003-2006*. Bogotá, Colombia: Diario Oficial de Colombia.
- Congreso de la República. (25 de julio de 2006). Día de lectura en los parques y establecimientos carcelarios colombianos. *Ley 1034 de 2006*. Bogotá, Colombia: Diario Oficial de Colombia.
- Congreso de la República. (15 de enero de 2010). Red Nacional de Bibliotecas Públicas. *Ley 1379 de 2010*. Bogotá, Colombia: Diario Oficial de Colombia.
- Corrado, A. (2013). Earnings persistence and schooling returns. *Economics Letters*, 482-484.
- DNP. (21 de abril de 2003). Lineamientos del Plan Nacional de Lectura y Bibliotecas. *Conpes 3222 de 2003*. (DNP, Ed.) Bogotá, Colombia. Recuperado el 10 de septiembre de 2018, de <https://www.dnp.gov.co/CONPES/documentos-conpes/Paginas/documentos-conpes.aspx#Default={%22k%22:%22ConpesNumero:3222 OR Title:3222%22}>
- Dumont, M. (2006). The reliability—or lack thereof—of data on skills. *Economics Letters*, 348-353.
- Fasih, T., Patrinos, H. A., & Sakellariou, C. (November de 2013). Functional Literacy, heterogeneity and the returns to schooling. *Policy Research Working Paper(WPS6697)*. The World Bank.

- Fundalectura. (10 de julio de 2001). *Fundalectura*. Recuperado el 13 de septiembre de 2018, de <http://fundalectura.org/colombia-crece-leyendo/>
- Green, D., & Riddell, W. (Abril de 2003). Literacy and Earnings: An Investigation of the Interaction of Cognitive and Unobserved Skills in Earnings Generation. *Labour Economics*, 10(2), 165-164.
- Hahn, J., Petra, T., & Wilbert, K. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 201-209.
- Hanushek, E. A., Schwerdt, G., & Wiederhold, S. (2017). Coping with change: International differences in the returns to skills. *Economics Letters*, 15-19.
- Igarza, R., & Monak, L. (2014). *Metodología común para explorar y medir el comportamiento lector. El encuentro con lo digital*. Bogotá, Colombia: Cerlalc - Unesco.
- Jaramillo, B., Rey, G., & otros. (2006). *Hábitos de lectura, asistencia a bibliotecas y consumo de libros en Colombia*. Bogotá: Fundalectura.
- Jaramillo, B., Rey, G., Melo, J., & otros. (2001). *Hábitos de lectura y consumo de libros en Colombia*. Bogotá, D.C.: Fundalectura.
- Judge, K. (15 de Diciembre de 2004). *The Treasury New Zealand* . Obtenido de Publications: <https://treasury.govt.nz/publications/wp/adult-literacy-and-economic-growth-wp-04-24-html#section-8>
- Lee, D. S., & Lemieux, T. (junio de 2010). Regressions Discontinuity Designs in Economics. *Journal of Economic Literature*, 282.
- Lee, K. -H., & Nagaraj, S. (1988). Earnings and the principal components of job performance. *Economic Letters*, 95-97.
- Lin, D., Lutter, R., & Ruhm, C. J. (2018). Cognitive performance and labour market outcomes. *Labour Economics*, 121-135.
- OEI. (2006). *Conferencias y Foros Iberoamericanos*. OEI.

Page, E. J. (2004). *El Decenio de las Naciones Unidas de la Alfabetización: Los comienzos (2003-2004)*. Paris, Francia: Unesco.

PNUD. (1 de enero de 2016). *Objetivos de Desarrollo Sostenible*. Recuperado el 12 de septiembre de 2018, de www.undp.org/content/undp/es/home/sustainable-development-goals.html

Rivera-Batiz, F. L. (1990). Literacy skills and the wages of young black and white males in the U.S. *Economic Letters*, 377-382.

Shomos, A. (2010). *Links Between Literacy and Numeracy Skills and Labour Market Outcomes*. Melbourne: Productivity Commission Staff Working Paper.

Tazeen, F. (2008). *Linking Education Policy to Labor Market Outcomes*. Washington: The World Bank.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment. *Journal of Educational Psychology*, 309-3017.

Unesco. (2006). *Educación para todos, La alfabetización un factor vital*. Unesco.

Unesco. (2013). *La Educación después de 2015*. 37 C/56. Paris: Unesco.

Wagner, D. A. (1995). Literacy and development: rationales, myths, innovations, and future directions. *International Journal Educational Development*, 341-362.