

# **3. ESTIMACIÓN DEL EFECTO CAUSAL DEL USO DE ORDENADORES EN LOS RESULTADOS DE LOS ESTUDIANTES EN LA PRUEBA PISA 2012**

**STEFANO CABRAS Y JUAN D. TENA**

*Universidad Carlos III (España) y Università di Cagliari (Italia)*

## 3. ESTIMACIÓN DEL EFECTO CAUSAL DEL USO DE ORDENADORES EN LOS RESULTADOS DE LOS ESTUDIANTES EN LA PRUEBA PISA 2012<sup>1</sup>

Stefano Cabras y Juan D. Tena

Universidad Carlos III (España) y Università di Cagliari (Italia)

*“No, mother, I could not feel your church group praying for my safety. The fact that I’m home safe is not proof that it worked, that logic is Post Hoc Ergo Propter Hoc.”*

Sheldon Cooper en The Big Bang Theory, "The Electric Can Opener Fluctuation"

*“Le nez de Cléopâtre: s’il eût été plus court, toute la face de la terre aurait changé”*

Blaise Pascal

### RESUMEN

Usando la base de datos PISA 2012 hemos estimado el efecto causal del uso de ordenadores en la escuela sobre el rendimiento de los estudiantes en matemáticas para España. Dicha estimación se ha realizado mediante modelos BART basados en técnicas bayesianas de regresión no paramétrica que cuenta con importantes ventajas, comparado con los procedimientos paramétricos más habituales que han sido discutidas en el artículo. Los resultados muestran cierta evidencia del efecto positivo del uso de ordenadores sobre el

---

<sup>1</sup> Agradecemos al Ministerio de Educación, Cultura y Deporte la financiación de este trabajo. También estamos en deuda con Ismael Sanz Labrador, Ruth Martín Escanilla, Francisco Javier García Crespo y David Cervera Olivares por la amabilidad con la que han atendido nuestras dudas y requerimientos sobre la base de datos PISA 2012. Por supuesto, cualquier error en el análisis de estos datos es enteramente responsabilidad de los autores.

rendimiento escolar de los estudiantes en España. Resulta especialmente interesante el hecho de que el posible efecto positivo parece ser significativamente mayor en estudiantes que pertenecen a grupos socioeconómicos más desfavorecidos, lo que refuerza la consideración de este tipo de intervención como una herramienta para conseguir mayor equidad. También se encuentra que el efecto de esta política depende de una forma no monótona del modo en que se usan los ordenadores.

## Palabras clave

Árboles de regresión, Causalidad, ICT, Estadística bayesiana, Modelos BART.

## INTRODUCCIÓN

La educación es una fuente importante de formación de capital humano en un país y su mejora no solo está asociada a un mayor crecimiento económico, Barro (2001) y Hanushek y Kimko (2000), sino también a otros beneficios intangibles como son la salud, una menor delincuencia y un mejor comportamiento ciudadano, ver Lochner (2011) y las referencias allí mencionadas. Es por esto que resulta esencial entender cómo se produce la educación, así como poder estimar el efecto causal en dicha función de diferentes políticas públicas.

Comparado con otras posibles intervenciones gubernamentales encaminadas a mejorar los niveles educativos, la inversión en información y comunicación tecnológica (ICT a partir de ahora) en centros educativos resulta especialmente atractiva al ser este insumo observable sin ambigüedad y perfectamente divisible entre los centros de acuerdo a criterios de discrecionalidad política. Sin embargo, en principio su efecto causal sobre el desempeño de los estudiantes es ambiguo ya que, si bien puede argumentarse que el uso de ICT puede ayudar al estudiante a entender mejor el contenido de sus asignaturas, también podría entenderse que puede desplazar otras técnicas de trabajo más efectivas y también suponer una distracción de las materias de estudio por el alumno.

El objetivo de este artículo es estimar el impacto causal sobre el rendimiento escolar de la presencia y uso de ordenadores en el centro de estudios usando la información proporcionada por la base de datos PISA 2012. Un problema fundamental de este estudio, y en general de cualquier análisis de causalidad en ciencias sociales, reside en poder diferenciar entre asociación y causalidad. La solución satisfactoria de este problema solo podría ser posible en con una base de datos metafísica, donde se pudiese observar la diferencia entre la respuesta de un grupo de individuos a un determinado tratamiento y la de estos mismos individuos en idénticas circunstancias sin dicho tratamiento. Sin embargo, Dawid (2000) explica que en un modelo físico solo se puede identificar la distribución del modelo con y sin tratamiento y no la distribución conjunta sobre la que se quiere hacer inferencia. Como consecuencia de esto se concluye que incluso en experimentos aleatorios y con un gran número de datos bajo modelos tradicionales no es posible hacer inferencia sobre efectos causales sin imponer supuestos no

contrastables e incluso posiblemente ilógicos y se sugiere una forma simple de resolver este problema basado en un enfoque bayesiano de decisión analítica.

La literatura económica previa sobre el impacto de ICT en desempeño educativo de los estudiantes, que será discutida de forma más extensa en la siguiente sección, ha tratado este problema fundamentalmente de dos formas: (1) mediante la identificación de instrumentos exógenos que están correlacionados con la variable causal, pero que se asume que están incorrelacionados con el término de error del modelo o (2) diseñando experimentos en el que la asignación de ICT al estudiante sea aleatoria. Sin embargo, estas estrategias de estimación no solo no están libres de la crítica de Dawid (2000) mencionada en el párrafo anterior sino que además al usar especificaciones paramétricas obtienen resultados que son potencialmente dependientes del modelo seleccionado.

En este trabajo estimamos el impacto causal de ICT sobre el resultado de los estudiantes en PISA 2012 usando la estrategia de modelización bayesiana no paramétrica denominada Bayesian Additive Regression Trees (BART) originariamente desarrollada por Chipman et al. (2010) y aplicada a estudios de inferencia causal por Hill (2013). Este enfoque, principalmente dirigido a una estimación óptima de la superficie de respuesta, es decir el PISA score, permite la identificación casual en trabajos no experimentales sin la necesidad de estimar dos modelos, uno para la capturar la potencial endogeneidad de la variable ICT y otro para el desempeño educativo de los estudiantes. El uso de una forma funcional tan flexible tiene además importantes ventajas comparadas con las estrategias de modelización paramétricas más tradicionales. Entre ellas se puede mencionar el no requerir ningún tipo de hipótesis sobre la covariables posiblemente relacionadas con el modelo que genera la respuesta, la posibilidad de tratar con un gran número de regresores y de estudiar un gran número de efectos interactivos entre la variable de tratamiento y las otras variables en el estudio o el fácil tratamiento de datos atípicos sin la necesidad de eliminarlos previamente como normalmente se hace.

El resto de este artículo está estructurado de la siguiente forma. En la siguiente sección se ofrece una discusión más pormenorizada de la literatura previa sobre la estimación del impacto de ICT sobre el rendimiento de los alumnos. En la sección 3 describimos la metodología BART considerada en este trabajo y ya en la sección 4, se muestran y se analizan los resultados de dicha estrategia aplicados a los datos PISA 2012. Finalmente en la sección 5, se extraen las conclusiones finales del análisis.

## REVISIÓN DE LA LITERATURA

La estimación del impacto que ICT tiene sobre el desempeño de estudiantes ha atraído la atención de la investigación económica en los últimos veinte años. Una revisión muy preliminar de esta literatura puede ya encontrarse en Kirkpatrick and Cuban (1998) quienes señalan una falta de consenso sobre el efecto causal de ICT en los resultados escolares de estudios previos y explican que la mayor parte de dichas investigaciones están basadas en una

simple correlación entre ICT y el desempeño de los alumnos sin tener en cuenta la influencia de variables omitidas tales como las características de la escuela, recursos y habilidades que están potencialmente ligadas tanto a los recursos computacionales como a los resultados de los estudiantes.

El problema de la endogeneidad ha producido una fuerte preocupación sobre la validez de los resultados encontrados en estudios previos y la necesidad de usar procedimientos econométricos que permitan corregir por el potencial sesgo de estimación. Algunas soluciones a este problema encontrados en investigaciones más recientes son (1) identificar una cierta política educativa que resulta generada por factores que no están correlacionados con la respuesta de interés, Angrist y Lavy (2002), Leuven et al. (2007), Machin et al. (2007) y Goolsbee y Guryan (2006); (2) diseñar un experimento que asigna de forma aleatoria a los participantes en el grupo de tratamiento o en el grupo de control de manera que la aleatoriedad elimine los efectos confundientes, Banerjee et al. (2007); y (3) usar una base de datos que contenga una extensa información sobre estudiantes y características de la escuela que no son típicamente observadas en análisis estándar para controlar por estas variables en un análisis de regresión, Fuchs and Woessman (2004) y Spiezia (2010).

La mayor parte de estos trabajos concluyen que ICT no ha tenido efecto sobre el desempeño de los alumnos. Más concretamente, Angrist y Lavy (2002) estiman el efecto de una política de dotación tecnológica, a gran escala en escuelas elementales y medias, llevada a cabo en Israel. Para estimar el efecto causal de esta política ellos controlan por características no observables de la escuela explotando el hecho de que las decisiones de financiación pública están determinadas fundamentalmente por el ranking de las escuelas solicitantes. No se encuentra evidencia de que incrementar el uso de ordenadores con fines educativos tenga alguna influencia positiva sobre la evaluación de los estudiantes. Leuven et al. (2007) evalúan el efecto sobre los resultados de los alumnos de políticas de subsidios dirigidas a escuelas para ordenadores y software en Holanda. En principio todas las escuelas con al menos un 70% de alumnos en desventaja reciben este subsidio mientras no se recibe el subsidio en caso contrario. Sin embargo, existen excepciones a esta regla lo que sirve para estimar efectos casuales estimando la efecto local de esta discontinuidad alrededor del umbral del 70% de alumnos desaventajados con diseño difuso en el que la probabilidad de recibir un fondo es una función de la elegibilidad<sup>2</sup>. Bajo esta especificación ellos encuentran que los subsidios a ordenadores no tienen un efecto positivo en los resultados de estudiantes holandeses. Goolsbee y Guryan (2006) evalúan el impacto de un programa público para subsidiar inversiones escolares en internet y comunicaciones. Ellos explican que a los distritos se les permite decidir qué escuelas incluir en sus solicitudes y que los distritos muy pobres tienen incentivos para solicitar esta ayuda mientras los colegios de mayor renta tienen incentivos para convencer a colegios más pobres en sus distritos para hacer solicitudes conjuntas lo que puede generar un sesgo en la estimación del impacto del subsidio. Para tratar con este problema consideran (1) estimación de la discontinuidad con diseño difuso alrededor del umbral que sirve para determinar las escuelas que son elegibles para recibir el subsidio, y (2)

---

<sup>2</sup> Este tipo de análisis son más conocidos utilizando la expresión inglesa “fuzzy regression discontinuity design”.

una estimación del cambio en los resultados del test que realizan los estudiantes usando como un instrumento válido el subsidio que el distrito habría recibido si hubiese sido forzado a incluir cada colegio dentro del distrito. Bajo ninguno de estos dos análisis se encuentra que el subsidio tenga un impacto significativo sobre el resultado de los estudiantes.

Sin embargo, existen trabajos que encuentran que ICT tiene un efecto positivo y significativo sobre los resultados de los estudiantes. Por ejemplo, Machin et al. (2007) evalúan el impacto de la inversión en ICT sobre los resultados de los estudiantes en Inglaterra. Para controlar por la endogeneidad potencial de la inversión en ICT consideran como variable instrumental un cambio en las reglas que gobiernan la inversión en ICT. En este caso, se encuentra evidencia de un efecto causal positivo de la inversión en ICT sobre el desempeño de los estudiantes.

Banerjee et al. (2007) también encuentran un efecto positivo de ICT usando un enfoque de estimación muy diferente a los mencionados hasta ahora. En concreto, diseñan un experimento aleatorio para estudiar el efecto causal del uso de ordenadores sobre el desempeño de estudiantes en la India. El análisis se basa en la comparación del cambio en los resultados de los estudiantes que reciben el tratamiento de enseñanza con ordenadores con un grupo de control encontrando un fuerte efecto del tratamiento en los resultados de matemáticas en el corto plazo. Sin embargo, también encuentran que este efecto positivo cae sustancialmente en el año sucesivo.

La base de datos PISA es particularmente útil para analizar el problema de la influencia de ICT en el desempeño de los estudiantes ya dispone de información relevante sobre alumnos y escuela que son típicamente variables omitidas en la mayor parte de los trabajos mencionados arriba. En Fuchs y Woessman (2004) y Spiezia (2010) se usa la base de datos PISA para estudiar el efecto de la inversión en ICT. Sin embargo, los resultados son ambiguos también en estos dos casos. Fuchs y Ludger (2000) usan información internacional de PISA 2000 encontrando una correlación positiva y significativa entre la disposición de ordenadores en la escuela y los resultados de los estudiantes pero dicha correlación se transforma en insignificante cuando otras características de la escuela se tienen en cuenta. Spiezia (2010) usa información de PISA 2006 concluyendo que ICT tiene un efecto positivo sobre los resultados en ciencias.

## METODOLOGÍA

En esta sección definimos el efecto causal de interés y la metodología de estimación. Seguiremos la notaciones y los argumentos teóricos discutidos en Hill (2011) por lo que concierne el uso del modelo estadístico para la estimación del efecto causal; en Chipman et al. (2010) para los detalles del modelo estadístico. En Leonti et al. (2011) se puede encontrar una de las pocas aplicaciones del mismo modelo a otro problema de estimación del efecto causal en la literatura de usos de plantas de medicamentos.

## Definición del efecto causal

Supongamos que tenemos  $N$  individuos participantes en el test de PISA y que las variables que se consideran a continuación están totalmente observadas o bien los datos que faltan están codificados como tales y que posiblemente haya incluso observaciones anómalas bien en la respuesta o en otras covariables. Para el individuo  $i$ ,  $i = 1, \dots, N$ , sea  $Y_i$  el valor del test de PISA o bien un *proxy* de esto como por ejemplo una extracción desde la distribución a *posteriori* predictiva de los puntos del PISA test (OECD, 2005) o incluso la media de las extracciones proporcionada, que aproxima la media a posteriori. Sea  $Z$  la variable *artificial* que indica el estado del uso de los ordenadores, es decir, la variable de intervención donde  $Z = 1$  si se tiene y se utiliza un *tablet*, un portátil o un ordenador fijo en la escuela y  $Z = 0$  en caso contrario.

Con el fin de calcular el efecto causal de  $Z$  sobre la respuesta  $Y$  en principio deberíamos conocer conjuntamente los resultados potenciales del valor del test, para el mismo individuo, bajo el uso,  $Y_i(0)$ , y el no uso de los ordenadores,  $Y_i(1)$ . Sin embargo, esto es imposible porque solo uno puede ser observado, mientras que el otro resultado no observado se le designa como resultado contrafactual y que tiene que ser estimado por medio de un modelo de regresión, como el descrito abajo. Dicho modelo se utiliza, esencialmente, en problemas de estimaciones de superficies de respuestas. En este caso, se trata de la respuesta  $Y$  al “hipotético tratamiento”  $Z$ .

Una vez conocidos o estimados los resultados potenciales, el efecto causal se puede definir como el efecto medio  $ATE = E(Y(1) - Y(0))$ , donde el valor esperado se calcula con respecto a la distribución de probabilidad para todos los individuos. El efecto causal, para cada individuo no es de interés, mientras sí puede ser el de algunos subconjuntos de individuos como, por ejemplo, el efecto causal sobre los que han recibido el tratamiento  $E(Y(1) - Y(0)|Z = 1)$  o sea los individuos que han utilizado el ordenador, en el cuyo caso, el valor esperado se calcula con respecto a la distribución de condicional de  $(Y|Z = 1)$ . Más general aún, si tenemos un conjunto de covariables  $X$  podemos calcular el efecto causal condicional a estas, es decir condicional a  $X = x$ .

En los estudios observacionales, como en el caso del test de PISA, los resultados potenciales no son, típicamente independientes de la asignación al tratamiento. Esto se conoce en la literatura como el problema de endogeneidad. En el caso del test de PISA, por ejemplo, es más probable que un alumno de familia con estudios acceda a una escuela con ordenadores y que simplemente sea el entorno familiar (y no el uso del ordenador en sí) el que determine un resultado más favorable en el test de PISA para este alumno comparado con otros que vienen de familias con menos estudios y que no utilizan el ordenador en sus escuelas. Para poder asumir que haya independencia de la asignación del tratamiento, hay que incluir todos los posibles factores confundientes representados, en este caso, por las covariables  $X$ . Concretamente, la hipótesis de *fuerte ignorabilidad* de la respuesta a la asignación al tratamiento afirma que  $Y$  es condicionalmente independiente de  $Z$  dado  $X$  y que la probabilidad de asignación al tratamiento es siempre positiva sea cual sea el valor específico de  $X$ . Para que esto se cumpla, es necesario entonces incluir en  $X$  todos los posibles factores

confundientes y por eso la matriz  $X$  suele ser de muy alta dimensionalidad y formada a partir de covariables de distinta naturaleza: cuantitativa, cualitativas y ordenables. Esta situación complica enormemente el análisis ya que requiere la utilización de modelos de regresión sofisticados para la estimación de  $Y$ .

Además, el hecho de considerar muchas covariables hace que algunos métodos clásicos como, por ejemplo, el *propensity score*, no sean de inmediata aplicación porque, al no haber podido observar tratamiento y no tratamiento para un mismo valor de  $X = x$ , se hace difícil la estimación de la puntuación que hay que asignar a cada individuo. Este hecho obliga, muchas veces al analista a considerar un conjunto de covariables de dimensionalidad menor, poniendo en duda el supuesto de fuerte ignorabilidad. Finalmente, es bien conocido que el planteamiento de modelos de regresión con muchas covariables hace imposible la búsqueda exhaustiva de todos los posibles modelos con todas las posibles interacciones. Esto, otra vez, impone al analista considerar solo efectos de interacciones entre covariables del primer o segundo orden o utilizar algoritmos de búsqueda como el *forward* o el *backward variable selection* que puedan proporcionar modelos localmente óptimos. Desgraciadamente no hay ninguna herramienta teórica, sino tan solo indicios empíricos, que nos hagan detectar el alcance de un óptimo local en vez de un óptimo global.

Debido a estos inconvenientes, y otros más que se puedan hallar en la utilización de las herramientas clásicas, en el siguiente apartado se describe el modelo BART que, no solo evita la especificación de un modelo, porque es de tipo no paramétrico y estimado por las observaciones, sino que también permite estimar con precisión satisfactoria la respuesta, y con eso también el resultado contrafactual. Esta especificación se encuentra en la clase de modelo bayesianos no paramétricos que permite de hacer inferencia condicional solo a las observaciones disponibles, sin recurrir a hipótesis de muestreo, típicas de la inferencia clásica, donde los resultados se basan en muestras jamás observadas.

## El modelo BART: verosimilitud y previas

Sean  $D$ , los datos disponibles, es decir el conjunto  $y, x, z$  observados en los  $N$  estudiantes. Indicamos con  $\pi(\cdot|\cdot)$  la distribución de probabilidad del argumento de la izquierda, condicionada al de la derecha. El objetivo del análisis es estimar la distribución de probabilidad *a posteriori* del efecto causal, es decir  $\pi(ATE|D)$ , o la distribución condicionada a algunas covariables,  $\pi(ATE|D, X = x)$ . Para estimarla utilizamos un modelo de regresión no paramétrico. Esto no quiere decir que el parámetro no exista sino simplemente que el espacio paramétrico no es un subconjunto de los reales. La novedad en este tipo de estudios de inferencia causal, además de la utilización de la inferencia bayesiana es el modelo muestral de regresión denominado BART. Como en todo los modelos bayesianos necesitamos de una función de verosimilitud sobre un conjunto de parámetros,  $\theta \in \Theta \notin \mathbb{R}$ , y una distribución previa  $\pi(\theta)$ ,  $\theta \in \Theta$ . La función de verosimilitud,  $L(y|x, z, \theta)$ , se deduce del siguiente modelo de regresión aditivo, donde la media de  $Y$  se determina a partir de la suma de modelos de estimación de la respuesta:



$$Y = \sum_{j=1}^m g(x, z; T_j, M_j) + \varepsilon, \varepsilon \sim N(0, \sigma^2) \quad (1)$$

donde  $g(x, z; T_j, M_j)$  es un árbol binario de regresión (o clasificación si  $Y$  fuera categórica) con sus variables y puntos de separación representados por  $T_j$  y sus nodos terminales representados por  $M_j$  y calculados con respecto al perfil  $x, z$  que pertenece al individuo cuya respuesta es  $Y$ . Esencialmente  $g$  es una función que asigna a cada individuo  $i$  su valor medio en el árbol  $j$ ,  $\mu_{ij} \in M_j$ . La puntuación final estimada para el individuo  $i$  correspondería al promedio de las  $m$  puntuaciones. Siendo  $T_j$  y  $M_j$  parámetros no reales, en este sentido el modelo de regresión es no paramétrico. Es bien conocido que, con el fin de minimizar el error de predicción, los árboles de clasificación tienden a crecer desmesuradamente llegando a producir *overfitting* de la respuesta y que en general es más eficiente un estimador hecho por muchos árboles pequeños que uno hecho por un solo árbol complejo. Ejemplos de este tipo de modelos aditivos son el Boosting (Shapire y Singer, 1999) o el Random Forest (Breiman, 2001).

Para conseguir esto es necesario utilizar una distribución *a priori* de regularización sobre el tamaño de los árboles,  $\pi(T, M)$  especificada en Chipman et al (2010). Dicha *a priori* de regularización hace que los árboles no crezcan mucho y que cada uno de los  $\mu_{ij}$  contribuya de manera marginal a la estimación de la respuesta. Con el fin de poder estimar la distribución *a posteriori* de  $\theta$  de forma computacionalmente factible, utilizamos una previa conjugada sobre  $\sigma^2$ , es decir una inversa-gamma que induce a una distribución condicional de  $\sigma^2$ ,  $\pi(\sigma^2 | T_1, \dots, T_m, M_1, \dots, M_m)$  que se puede expresar de forma conocida, siendo de nuevo una inversa-gamma. Como se muestra en Chipman et al (2010), los hiperparámetros de todas las distribuciones previas son especificados en relación a la muestra observada, dando lugar a las que se conocen como previas dependientes de los datos. Esta manera de proceder, no es muy ortodoxa en el sentido bayesiano, pero se halla en el conjunto de técnicas conocidas como *empirical Bayes* que son muy populares y que han sido revalorizadas desde el punto de vista teórico en un reciente trabajo Petrone et al. (2013). De todas formas, como se explica también en Hill (2011), los resultados para el análisis de los datos de PISA son robustos respecto a modificaciones de las previas en un entorno de los hiperparámetros estimados.

## El modelo BART: estimación

Utilizando las previas especificadas arriba, es posible entonces simular muestras de la distribución *a posteriori* con un esfuerzo computacional no excesivo utilizando Markov Chain Monte Carlo (MCMC), concretamente un Metropolis Hastings dentro de Gibbs. Esto quiere decir que el algoritmo de simulación alterna pasos de Gibbs (como el necesario para simular  $\sigma^2$ ) a pasos de Metropolis Hastings cuando las condicionales, para los otros parámetros, no estén disponible en forma conocida. En particular, la distribución que se utiliza para proponer las actualizaciones de los  $T_j$  y  $M_j$  consiste en añadir/eliminar un nodo terminal, cambiar una variable de separación o su correspondiente punto de separación con ciertas probabilidades detalladas en Chipman et al. (2010). Una vez obtenida la distribución *a posteriori* de

$\theta = (T_1, \dots, T_m, M_1, \dots, M_m, \sigma^2)$ , la distribución predictiva para los puntos del PISA test en el individuo es:

$$m(Y_i \setminus x_i, z_i) = \int_{\theta \in \Theta} L(Y_i; \theta) d\pi(\theta \setminus D) \quad (2)$$

que prácticamente se estima generando valores de  $Y_i$ , según la distribución normal con la media y la varianza de cada valor  $\theta$  en la cadena MCMC y los arboles de regresión calculados en  $x_i$  y  $z_i$ . Concretamente utilizamos  $m=500$  arboles y 5000 pasos de MCMC después de un calentamiento de 1000 pasos.

De esta forma se puede estimar la distribución de cada individuo y de su correspondiente respuesta contrafactual, simplemente estimando la respuesta en  $z_i = 1$  si el estudiante no tenía ordenador en su escuela y en  $z_i = 0$  en caso contrario. Obtenidas dichas distribuciones predictivas a posteriori, se considera la diferencia de las respuestas factual y contrafactual para obtener la distribución del efecto causal individual. Al final,  $\pi(ATE \setminus D)$ , se estima como el conjunto de todas estas diferencias para todos los individuos. Si luego se requiere la estimación del efecto causal condicional, esto se obtiene simplemente considerando las diferencias en los individuos que cumple con la condición  $X = x$ .

## RESULTADOS

### Estadística descriptiva

La base de datos PISA contiene información sobre conocimientos y habilidades de estudiantes que están próximos al final de su periodo de educación obligatoria. Sirve fundamentalmente para determinar la forma en que estos estudiantes están preparados para la vida tras el periodo de educación obligatoria más que centrarse en la evaluación de conocimientos curriculares. La base de datos contiene información sobre desempeño de los estudiantes en lectura, matemáticas y ciencias además de respuestas de los estudiantes, responsables de los colegios y padres. En este artículo nos centraremos únicamente en el desempeño de los estudiantes españoles en matemáticas usando los datos de PISA 2012 dejando así para versiones posteriores un análisis más completo que englobe más países y habilidades. Esto nos deja una base de datos de 25313 observaciones de estudiantes españoles. La forma de interpretar estas observaciones es que corresponden a un muestreo de la población total de manera que cada observación lleva asociado un peso relativo a su importancia poblacional.

En lo referente a nuestra variable dependiente, el desempeño de los estudiantes en matemáticas PISA, muestra cinco valores plausibles. Cada uno de estos valores representa una observación aleatoria obtenida de la distribución a posteriori de la puntuación de los estudiantes. La media poblacional de estos valores puede obtenerse de promediar la media de estos cinco valores plausibles que a su vez ha sido obtenida de una suma ponderada de las 25313 observaciones disponibles.

La variable dependiente tiene un valor medio de 484.3 con una desviación estándar igual a 87.7. Dicha desviación estándar se ha calcula a partir de la raíz cuadrada de la combinación lineal de la varianza de imputación, que estima la desviación entre las estimaciones obtenidas con los diferentes plausibles, y la varianza muestral, que estima la desviación debida a la muestra. La varianza muestral es igual a 7697.9 mientras la varianza de imputación es 0.11 lo que sugiere, en línea con trabajos anteriores, que la mayor parte de la incertidumbre en la estimación poblacional corresponde a la variabilidad muestral y no al hecho de tomar solo uno de los cinco valores plausibles.

La variable explicativa de interés que se utilizará para realizar inferencia causal es el uso de ordenadores. Esta información ha sido obtenida mediante la construcción de una variable artificial TREAT que toma el valor 1 cuando el alumno responde “Sí, y lo uso” a al menos una de estas tres preguntas: ¿Tienes la posibilidad de utilizar en el centro donde estudias alguno de estos aparatos? Un ordenador de sobremesa (IC02Q01), un ordenador portátil o un notebook (IC02Q02) o una tableta (IC02Q03).

Resulta interesante mostrar las características de la distribución de la muestra que, por simplicidad, lo hacemos solo para el primero de los cinco valores plausibles al ser este análisis prácticamente idéntico al que se obtienen con los otros valores. Aplicando el contraste de normalidad propuesto por Royston (1991) basado en un ajuste al test de D’Agostino, Belanger y D’Agostino (1990) se rechazaría la hipótesis de normalidad incondicional a todos los niveles de significación. La hipótesis de normalidad también se rechaza si miramos la distribución de esta variable para los valores 0 y 1 del indicador TREAT definido en el párrafo anterior se encuentra que la hipótesis de normalidad también se rechaza a los niveles de significación habitual en los dos casos. La ausencia de normalidad incondicional o condicional al tratamiento no excluye que la distribución de la respuesta condicional al conjunto de variables considerada no sea normal, como sugiere el análisis de los residuos del modelo BART. Esto es debido a la posibilidad de utilizar modelos muy flexibles en la definición del predictor lineal. Las Figuras 1 y 2 muestran el histograma de la puntuación en matemáticas obtenida por los estudiantes en los dos casos.

Figura 1. Histograma muestral del primer valor plausible sobre puntuación en matemáticas para estudiantes que **no usan** ordenador en su centro escolar

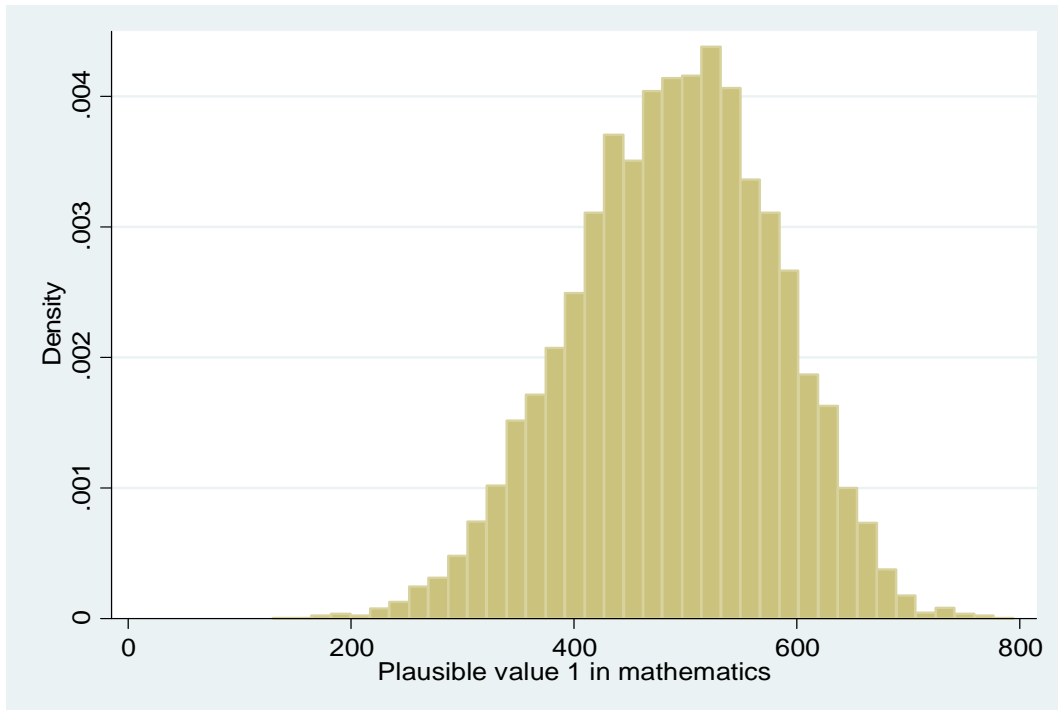
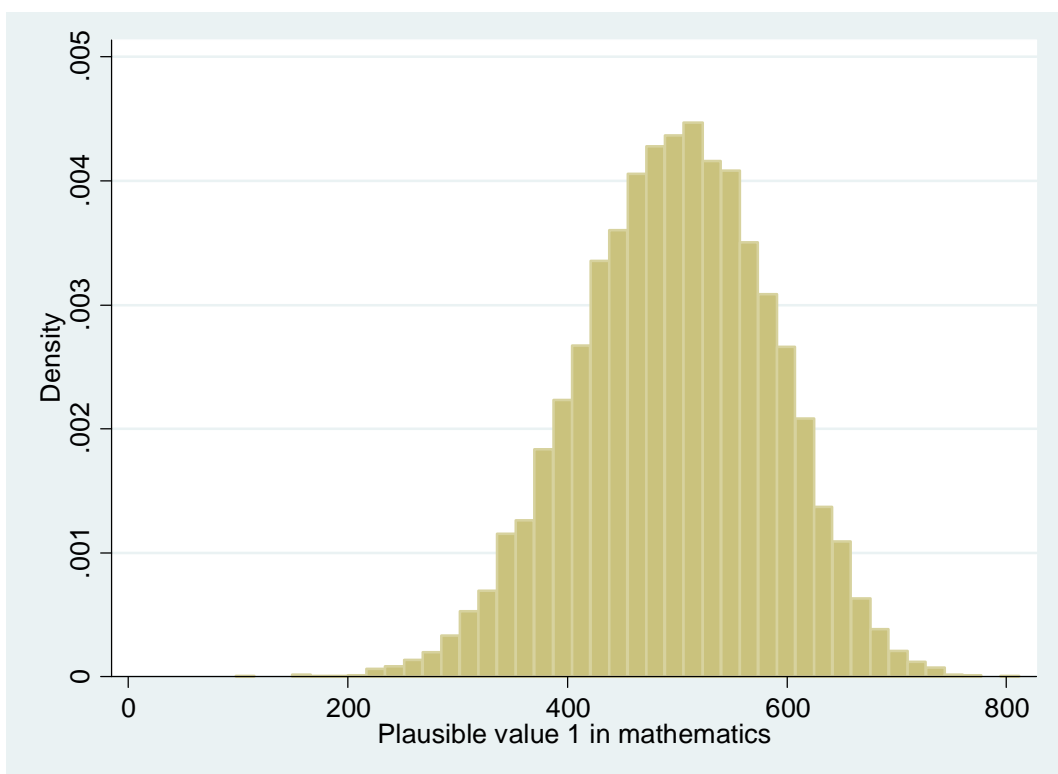


Figura 2. Histograma muestral del primer valor plausible sobre puntuación en matemáticas para estudiantes que **usan** ordenador en su centro escolar



Las variables explicativas adicionales, o covariables, contienen información sobre cómo son sus características previas o innatas, la posición socioeconómica de las familias a las que pertenecen o características del centro escolar en el que estudian. Mientras que las características previas del estudiante pueden observarse fácilmente de sus respuestas a los cuestionarios, los índices fundamentales relativos a la posición socioeconómica de sus familias a las características de la escuela a la que acuden pueden construirse directamente de la información en los cuestionarios, ver por ejemplo OECD (2009). De manera más concreta, las variables consideradas son las siguientes:

### ***Características previas del estudiante***

RELATIVE\_AGE: Indicador de si el estudiante ha nacido en los primeros seis meses del año.

GENDER: Género.

EARLY\_EDUCATION: Indicador de si el estudiante ha asistido a la escuela infantil no obligatoria en el pasado.

### ***Características socioeconómicas de la familia a la que pertenece el estudiante***

ESCS: Índice de status económico, social y cultural.

FAMSTRUC: Estructura familiar.

HEDRES: Recursos educativos en casa.

HISCED: Nivel educativo más alto de los padres.

HISEI: Nivel ocupacional más alto de los padres.

HOMEPOS: Posesiones en casa.

IMMIG: Estatus de inmigrante.

TIMEINT: Tiempo de uso del ordenador (en minutos).

WEALTH: Bienestar.

WEALTH2: Bienestar al cuadrado.

### ***Características del colegio***

CLSIZE: Tamaño de la clase.

SCMATEDU: Calidad de los recursos educativos de la escuela.

STRATIO: Ratio estudiantes por profesor.

SMRATIO: Ratio estudiantes por profesor en matemáticas.

SCHLTYPE: Indicador de si el colegio es público o privado.

Muchas de estas variables contienen valores no observados que aparecen en la base tomando valores extremos. Como explicaremos más adelante, esto no es un problema en nuestra metodología de estimación ya que las observaciones extremas terminan en nodos pequeños y el modelo automáticamente encuentra nodos que contienen principalmente outliers y estos no invalidan el cálculo de la respuesta media para las observaciones no extremas. Sin embargo, con el fin de mostrar alguna información inicial de estadística descriptiva, que mostramos en el siguiente cuadro, hemos eliminado observaciones con valores extremos y poco creíbles. Claramente, disponer de un ordenador en la escuela determina el valor esperado de muchas de las características de los estudiantes y del colegio y esto es un indicador de la necesidad de controlar por la influencia de todas estas variables en un análisis causal.

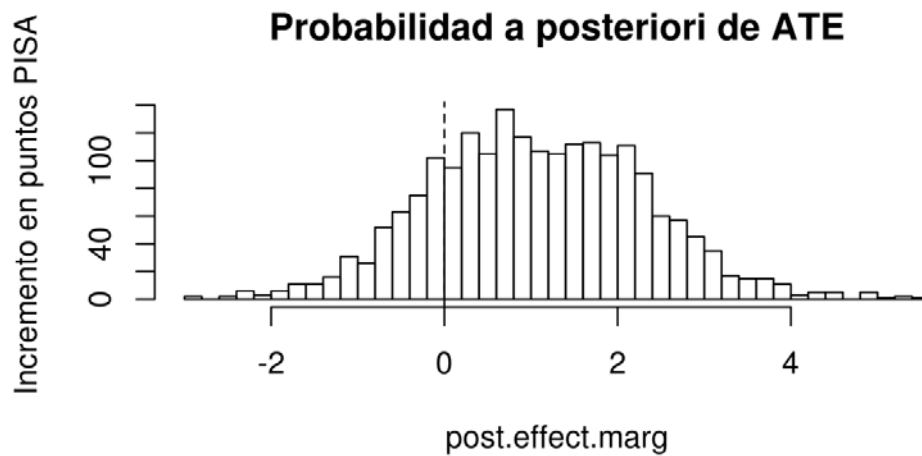
Cuadro 1. Estadística descriptiva para las covariables consideradas en el modelo. Datos muestrales sin ponderar

	COMPUTER=1					COMPUTER=0				
	Nº observaciones	Media	Desviación Estándar	Mínimo	Máximo	Nº observaciones	Media	Desviación Estándar	Mínimo	Máximo
RELATIVE_AGE	17850	0.50	0.50	0	1	7463	0.51	0.50	0	1
GENDER	17850	1.50	0.50	1	2	7463	1.50	0.50	1	2
EARLY_EDUCATION	17850	0.06	0.24	0	1	7463	0.06	0.25	0	1
ESCS	17812	-0.12	0.99	-3.92	2.73	7309	-0.09	1.03	-5.3	2.55
FAMSTRUC	16909	1.91	0.31	1	3	6888	1.90	0.32	1	3
HEDRES	17765	0.10	0.87	-3.93	1.12	7303	-0.02	0.91	-3.93	1.12
HISCED	17599	4.39	1.63	0	6	7219	4.46	1.66	0	6
HISEI	17658	47.58	21.42	11.01	88.96	7248	48.88	21.95	11.01	88.96
HOMEPOS	17784	0.13	0.81	-6.48	3.76	7320	0.07	0.86	-6.65	3.76
IMMIG	17227	1.18	0.57	1	3	7041	1.22	0.61	1	3
TIMEINT	17721	55.22	40.45	0	206	6596	51.57	41.67	0	206
WEALTH	17781	-0.01	0.79	-5.32	2.91	7316	-0.04	0.83	-5.32	2.91
CLSIZ	17850	33.54	24.30	13	99	7463	35.05	25.67	13	99
SCMATEDU	17648	0.11	0.91	-3.59	1.98	7361	0.04	0.90	-3.59	1.98
STRATIO	16700	11.91	6.76	1.11	139	6945	12.06	7.55	1.11	139
SMRATIO	16547	108.02	110.55	2	1820	6917	106.80	90.02	2	1820
SCHLTYPE	17309	2.58	0.58	0	8	7263	2.60	0.60	1	3

## Análisis y discusión de resultados con el modelo BART

A continuación, describimos los resultados del análisis para la base de datos descrita arriba utilizando como variable respuesta el primer valor plausible en la prueba de matemáticas PV1MAT. El resultado principal, es decir el efecto causal del uso de los ordenadores en el test de PISA para matemáticas es descrito en la Figura 3 que muestra la aproximación, por medio de extracciones MCMC, de la distribución a posteriori  $\pi(ATE \setminus D)$ .

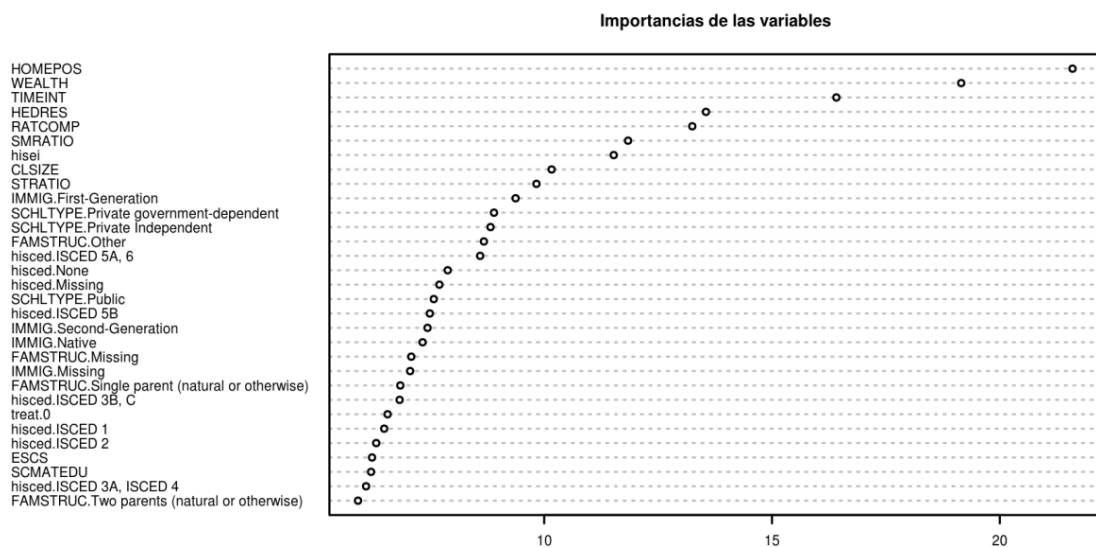
Figura 3. Extracciones de la distribución a posteriori



Según la Figura 3, se puede ver que hay un efecto tendencialmente positivo en el incremento de puntos, aunque hay cierta probabilidad de que no exista o sea incluso negativo. En particular, la probabilidad de un efecto positivo es 0.8, lo que equivale a decir que es 4 veces más probable de que el efecto sea positivo con respecto a nulo o negativo. En media el efecto es de 1 punto y el intervalo de credibilidad del efecto al 95% es (-1.2 , 3.4) lo que quiere decir que hay evidencias para efectos muy positivos o quizás negativos.

Las variables más importantes para explicar la respuesta se miden, de modo muy simple, contando el número de veces que estas variables aparecen en los árboles de clasificación. La Figura 4 ilustra este conteo.

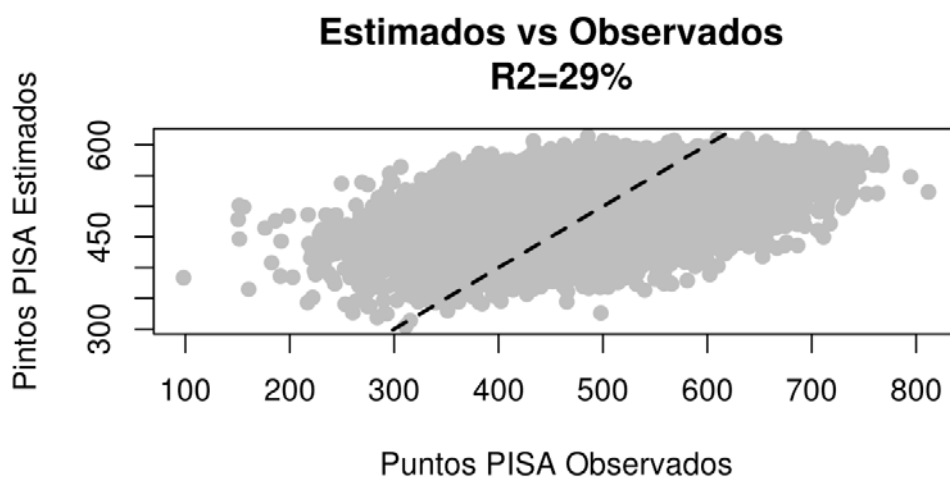
Figura 4. Variables más importantes según el número de veces que aparecen en los árboles



Las variables más importantes son HOMEPOS, TIMEINT, WEALTH y HEDRES. Sin embargo la variable de tratamiento (TREAT) no está entre las más importantes y esto explica porque el efecto causal del uso de los ordenadores no sea excesivamente elevado.

El modelo BART estimado con las covariables indicada en la Figura 4 explica alrededor del 29% de la variabilidad de la respuesta según detallado en la Figura 5. Resulta interesante comparar este procedimiento con una regresión paramétrica lineal mediante mínimos cuadrados ponderados de la variable de interés PV1MATH con respecto a TREAT y todas las demás variables de control ya mencionadas. La regresión resultante tiene un valor del estadístico  $R^2$  igual a 0.22 que tiene un valor sustancialmente más bajo que el obtenido mediante modelos BART. En la estimación paramétrica también puede concluirse que la variable TREAT tiene un efecto positivo pero no significativo a los niveles de significación habitual sobre PV1MATH. En concreto, el valor del coeficiente estimado es igual a 1.9 con un p-valor de 0.13.

Figura 5. Puntos Pisa observados y estimados ( $R^2=0.29$ )



Esencialmente resulta imposible, con las covariables disponibles explicar un gran porcentaje de la variabilidad de la respuesta. Para obtener porcentajes más elevados, tenemos que considerar el efecto de cada escuela mediante la variable SCHOOLID, pero al hacerlo el efecto causal del uso de los ordenadores tiende a distribuirse prácticamente alrededor de 0. Esto se interpreta como que la heterogeneidad en las escuelas es tal que esconde el efecto del uso de los ordenadores lo que quizá se podría obviar utilizando una base de datos más grande.

Suponiendo que el modelo sea aceptable, es decir, ignorando la heterogeneidad de cada una de las escuelas podríamos estimar de manera muy fácil el efecto causal del uso de los ordenadores condicionalmente a las covariables de interés más usadas en el análisis. El cálculo de todos estos efectos interactivos sería muy complejo bajo una modelización paramétrica, pero resulta muy sencillo mediante modelos BART. Las siguientes Figuras 6 y 7 sugieren cierta tendencia en la asociación del efecto causal con algunas covariables o perfiles de los estudiantes. Estas distribuciones se determinan limitando el análisis anterior sobre las salidas de la cadena MCMC exclusivamente a los individuos en los cuales la variable condicionante asume ciertos valores como el indicado en los BoxPlots o en el eje horizontal de los gráficos en las Figuras 6 y 7.



Figura 6. Distribuciones condicionadas del efecto causal (a)

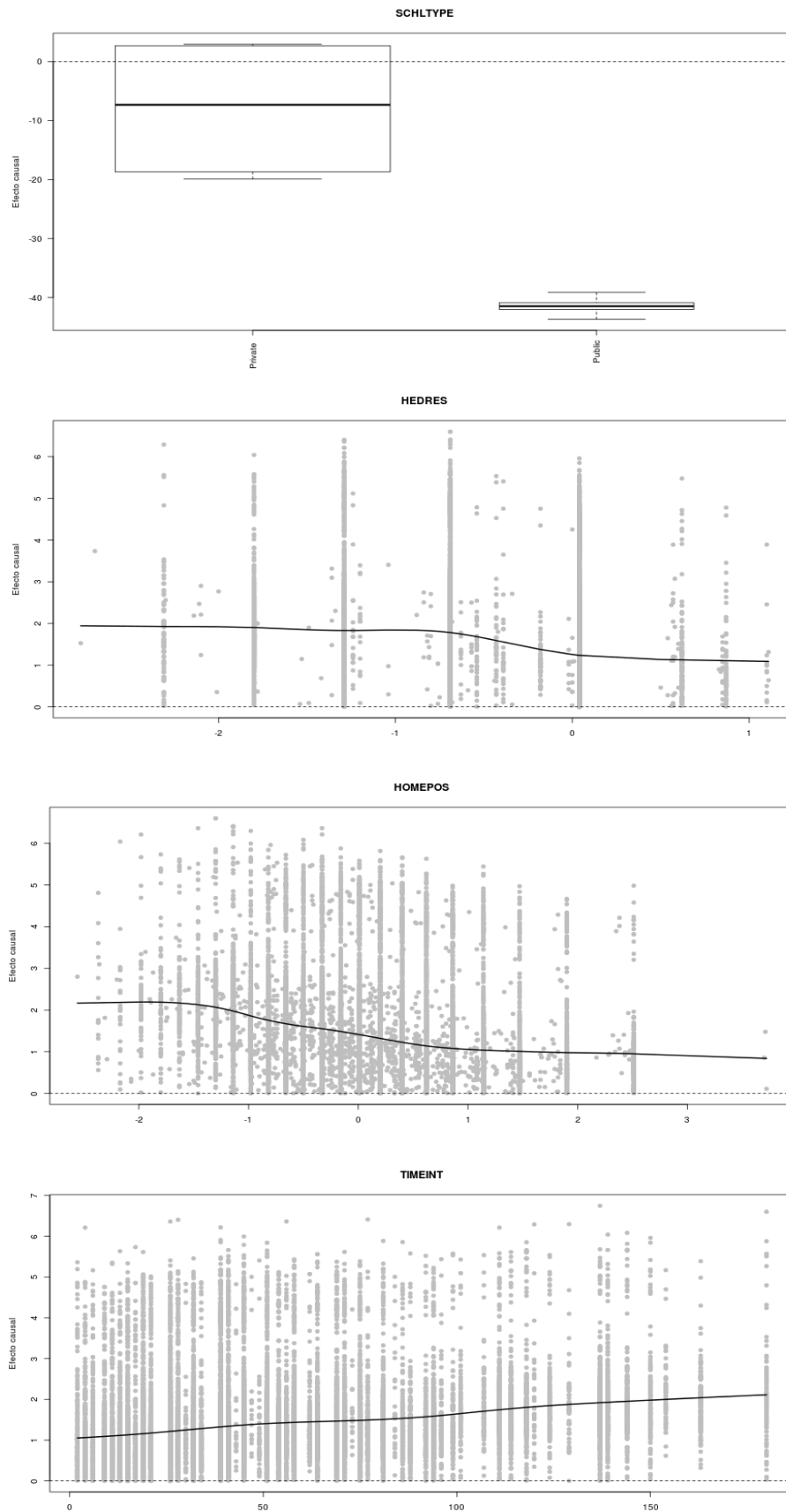
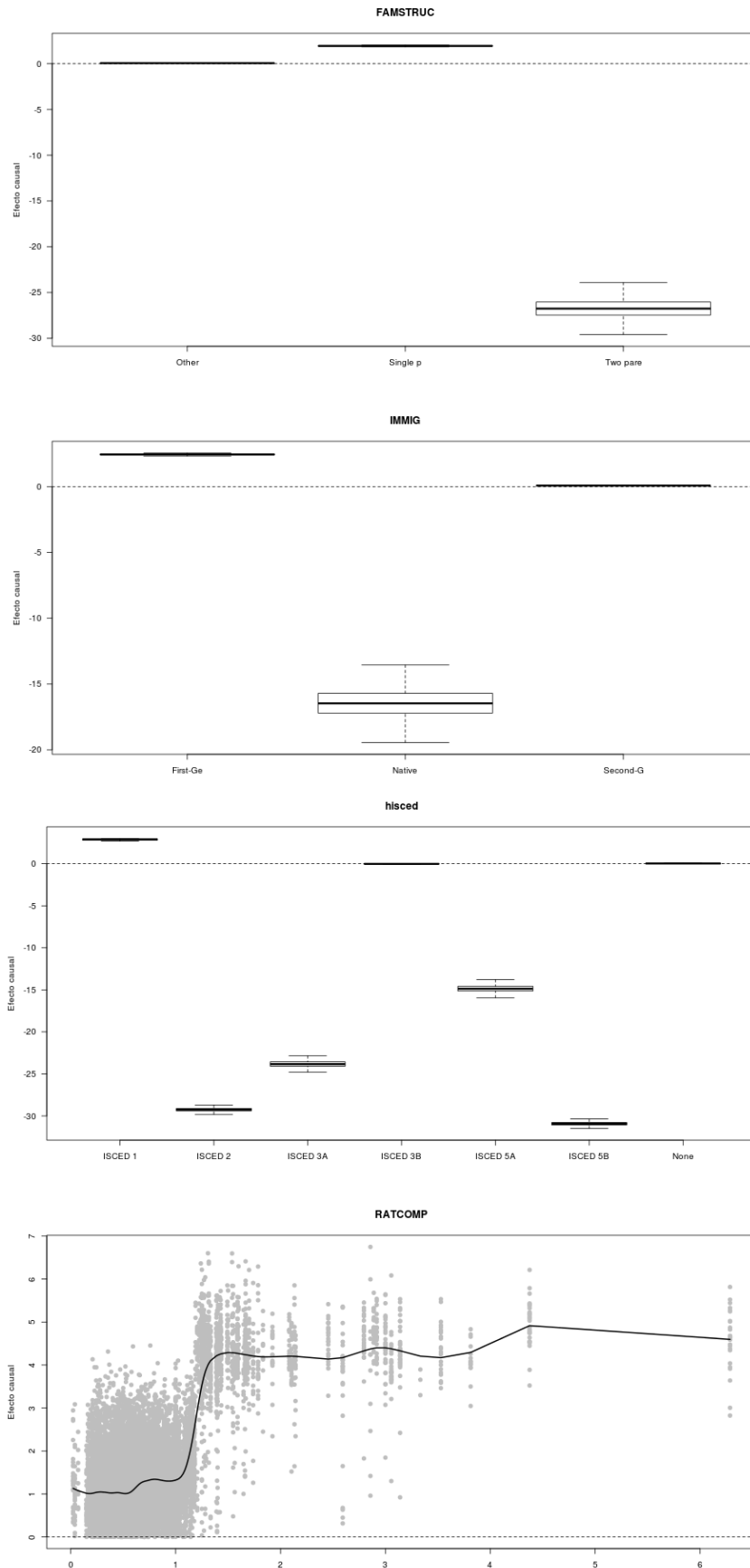


Figura 7. Distribuciones condicionadas del efecto causal (b)



Los resultados encontrados en esta figuras sugieren en primer lugar que la el efecto de variable de tratamiento es condicional al número de ordenadores en la clase y a su tiempo de uso. En concreto, la interacción de la variable de tratamiento con el índice del número de ordenadores en la escuela por el número de estudiantes (RATCOMP) tiene un efecto umbral que parece sugerir que para que la inversión en ordenadores sea efectiva debe superarse un número mínimo de ordenadores por clase. El efecto interactivo del tiempo de uso del ordenador (TIMEINT) con la variable de tratamiento también tiende a ser positivo aunque con una pendiente no muy pronunciada. Esto puede deberse bien a que nuestra variable de tratamiento ya incluye información sobre el uso del ordenador por lo que el efecto marginal de TIMEINT es pequeño.

Resulta especialmente interesante ver como la variable de tratamiento interactúa con otras variables que son indicativas de la posición socioeconómica del estudiante. Un resultado muy remarcable desde el punto de vista político es que, en general, la variable de tratamiento tiende a tener un mayor efecto positivo en estudiantes que pertenecen a entornos más desfavorecidos. Esto puede observarse en la interacción del tratamiento con la posición social (HOMEPOS) y los recursos educativos en la familia del estudiante (HEDRES) y en un mayor efecto en los inmigrantes de primera y segunda generación que en los nativos (INMIGRATE). Sin embargo, el nivel educativo de los padres indicados por HISCED no tiene una interacción clara con la variable de tratamiento.

## CONCLUSIONES

Usando la base de datos PISA 2012 hemos estimado el efecto causal del uso de ordenadores en la escuela sobre el rendimiento de los estudiantes en matemáticas para España. Dicha estimación se ha realizado mediante el uso de modelos BART basado en técnicas bayesianas de regresión no paramétrica que cuenta con importantes ventajas comparado con los procedimientos paramétricos más habituales. En concreto, para tratar con la endogeneidad no existe la necesidad de especificar dos modelos uno para la variable de tratamiento, el uso de ordenadores, y otro para la variable de resultados del estudiante. Otras ventajas de este procedimiento residen en menor necesidad de supuestos en la especificación del modelo, la posibilidad de analizar con un elevado número de variables así como los efectos interactivos de cada una de ellas con nuestra variable de decisión y la facilidad en el tratamiento de datos no observados y atípicos.

Se encuentra una moderada evidencia sobre el efecto positivo del uso de ordenadores en el rendimiento escolar de los estudiantes en España. Resulta muy interesante el hecho de que este efecto positivo parece ser significativamente mayor en estudiantes que pertenecen a grupos socioeconómicos más desfavorecidos lo que refuerza la consideración de este tipo de intervención como una herramienta para conseguir mayor equidad. También se encuentra que el efecto de esta política depende de una forma no monótona del modo en que se usan los ordenadores y de la proporción de alumnos por profesor.

Este trabajo constituye una primera aproximación del análisis del efecto causal de los ordenadores en el rendimiento escolar. El hecho de que los resultados no hayan sido claramente significativos puede deberse a que efectivamente la variable no tenga un impacto importante sobre el rendimiento escolar. Para encontrar resultados más concluyentes habrá que repetir este mismo análisis con la base de datos internacional que cuenta con mayor número de observaciones. En una versión posterior de este artículo estudiaremos el efecto causal de los ordenadores considerando los datos PISA 2012 para un grupo representativo de países.

## REFERENCIAS

- Angrist, J., y V. Lavy (2002). "New Evidence on Classroom Computers and Pupil Learning". *The Economic Journal*, 112, 735-765.
- Banerjee, A., S. Cole, E. Duflo y L. Linden (2008). "Remedying Education: Evidence from Two Randomized Experiments in India", *Quarterly Journal of Economics*, 122(3), 1235-1264.
- Barro, R. (2001). "Human capital and growth". *American Economic Review*, 91(2), 12–17.
- Breiman, L. (2001). "Random forests". *Machine Learning*, 45, 5-32.
- Chipman H.A., George E.I., Mculloch R. (2010). "BART: Bayesian Additive Regression Trees", *The Annals of Applied Statistics*, 4(1), 266–298.
- D'Agostino, R.B., A.J. Belanger y R.B. D' Agostino, Jr. 1990. "A suggestion for using powerful and informative tests of normality". *American Statistician* 44: 316-321.
- Dawid, A. P. (2000). "Causal Inference Without Counterfactuals". *Journal of the American Statistical Association*. 95 (450), 407-424.
- Fuchs, T. y L. Woessman (2004). "Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at Schools", *Brussels Economic Review*, 47, 359-389.
- Goolsbee, A. y J. Guryan (2006). "The Impact of Internet Subsidies in Public Schools", *The Review of Economics and Statistics*, 88(2), 336–347.
- Hill, L.H. (2013). "Bayesian Nonparametric Modeling for Causal Inference", *Journal of Computational and Graphical Statistics*, 20(1), 217-240.
- Hanushek, E., & Kimko, D. (2000). "Schooling, labor force quality and economic growth". *American Economic Review*, 90(5), 1184–1208.
- Kirkpatrick, H. y L. Cuban (1998). "Computers Make Kids Smarter – Right?", *TECHNOS Quarterly for Education and Technology*, 7(2), 1-11.
- Leonti, M., S. Cabras, C. Weckerle, M. Solinas, and C. L. (2010). "The causal dependence of present plant knowledge on herbals—contemporary medicinal plant use in campania (italy) compared to matthioli (1568)", *Journal of Ethnopharmacology* 130 (2), 379–391.
- Leuven, E., M. Lindahl, H. Oosterbeek y D. Webbink (2007). "The effect of extra funding for disadvantaged pupils on achievement", *The Review of Economics and Statistics*, 89(4), 721–736.

Lochner, L. (2011). "Non-Production Benefits of Education: Crime, Health, and Good Citizenship," NBER Working Papers 16722, National Bureau of Economic Research, Inc.

Machin, S., S. McNally y O. Silva (2007). "New Technology in Schools: Is There a Payoff?," *Economic Journal*, 117(522), 1145-1167.

OECD, PISA 2009 Technical Report, OECD, Paris.

Petrone, S., Rousseau, J., y Scriocciolo, S. (2013). "Bayes and empirical Bayes: Do they merge?," *Biometrika* (to appear).

Royston, P. (1991). "Comment on sg3.4. and an improved D'Agostino test". Stata Technical Bulletin 3: 23-24. Reprinted in Stata Technical Bulletin Reprints, 1, 110-112. College Station, Tx: Stata Press.

Schapire, R.E. y Singer, Y. (1999). "Improved boosting algorithms using confidence-rated predictions", *Machine Learning*, 37, 297-336.

Spiezia, V. (2010). "Does Computer Use Increase Educational Achievements? Student-level Evidence from PISA," *OECD Journal: Economic Studies*, 1, 1-22.