



Predicción del resultado en la prueba SABER PRO para Economía a partir de la información disponible en el proceso de admisión

Jairo Alberto Cifuentes Garzón¹

Resumen

Se analiza el valor explicativo de algunas variables observables en el estudiante antes de iniciar sus estudios universitarios, con el objetivo de predecir el rendimiento en la prueba estandarizada SABER PRO para economía. Este análisis resulta de interés para seleccionar a los candidatos en los procesos de admisión por sus características tanto académicas como personales de entrada a la universidad.

Se utilizó información contenida en las pruebas SABER 11 en cuanto al puntaje de cada componente, la edad y el género del estudiante. A través de regresión cuantílica se encuentra un efecto positivo y significativo del género a favor de los hombres, el puntaje en matemáticas, en historia, en lenguaje y en idiomas sobre el resultado en la prueba SABER PRO, dicho efecto se incrementa a medida que se predicen los puntajes más altos (cuantiles superiores). El efecto asociado a la prueba de biología deja de ser significativo en los cuantiles superiores del puntaje, la edad predice mejores resultados a los estudiantes más jóvenes. Las pruebas de filosofía, física e interdisciplinar no resultan significativas.

Palabras Claves: Rendimiento académico, regresión cuantílica, admisión universitaria, SABER PRO.

¹ Trabajo presentado para optar al título de Magister en Economía de la Pontificia Universidad Javeriana.

Introducción

El Ministerio de Educación Nacional de Colombia con el ánimo de mejorar la calidad académica de los egresados de programas universitarios, instauró la presentación obligatoria de un examen estandarizado a nivel nacional conocido actualmente como SABER PRO (antes Ecaes). Dicho examen pretende evaluar las competencias necesarias para el adecuado desempeño profesional. Los programas son clasificados de acuerdo a dicha evaluación, constituyéndose en referencia para efectos de comparación. De allí que represente un reconocimiento importante obtener buenos resultados ¿qué hacer para lograrlo? ¿Qué variables sirven para predecir los mejores resultados?

Con el propósito de hacer un acercamiento a dichos cuestionamientos, se analizaron algunas condiciones y habilidades de los estudiantes antes de iniciar sus estudios universitarios y durante el proceso formativo. De igual forma se analizaron algunas características de los programas (registro de alta calidad, origen, valor de la matrícula, entre otras) para pronosticar el rendimiento académico del estudiante en la prueba SABER PRO para Economía. Esto con el fin de brindar a los programas información relevante sobre que variables incrementan la probabilidad de obtener mejores resultados, inclusive a partir del proceso de admisión de los estudiantes.

La presente propuesta formula herramientas adicionales para la selección de aquellos candidatos que por sus características de entrada permitan pronosticar un buen desempeño al finalizar. Para diseñar dichas herramientas se utilizó la información académica y personal reportada por el Instituto Colombiano para la Evaluación de la Educación ICFES, contenida en las bases de datos de la Prueba Saber 11 (antes conocida como Examen del ICFES), un examen de estado que evalúa a los estudiantes que están terminando su ciclo de Educación Media y que tiene como propósito ayudar en los procesos de selección y admisión que realizan las instituciones de educación superior.

El trabajo se divide en dos etapas, en la primera se analiza el rendimiento académico por componente en la prueba SABER 11 (Biología, matemáticas, filosofía, historia, lenguaje, física, química, interdisciplinar e idioma extranjero) como variables que predicen el rendimiento en la prueba SABER PRO. Información que es de conocimiento de los programas en sus procesos de admisión.

En una segunda etapa se consideran conjuntamente características académicas y socioeconómicas del estudiante, del entorno familiar y de la institución universitaria a la que pertenece al momento de tomar la prueba SABER PRO como predictivas

del rendimiento académico en la SABER PRO. Además se analiza el valor predictivo para cada componente que conforma la prueba: macroeconomía, microeconomía, estadística y econometría, pensamiento económico e historia económica.

En este tipo de análisis es común recurrir a mínimos cuadrados ordinarios para estudiar el efecto de un conjunto de variables independientes sobre el rendimiento académico, sin embargo el hecho de suponer que el efecto es el mismo para todos los estudiantes sin importar su nivel de desempeño (distribución de los puntajes) en situaciones de asimetría puede causar estimaciones poco confiables.

En estos casos la regresión por cuantiles puede proporcionar mayor información sobre el efecto de las variables explicativas en distintos puntos de la distribución de la variable respuesta. Trabajos similares a este como Eide & Mark (1998), Di Gresia, et al. (2009) a través de regresión por cuantiles concluyen que en general el efecto de un conjunto de variables explicativas sobre el rendimiento académico no es el mismo en todos los tramos de la distribución.

Utilizamos esta metodología con el objetivo de observar el cambio en el efecto de la predicción del puntaje especialmente en las colas superiores de la distribución, dado que el mayor interés recae sobre la capacidad de predecir los puntajes más altos a partir de la información disponible en los procesos de admisión.

Revisión bibliográfica

Trabajos pioneros en Colombia han supuesto la homogeneidad de los estudiantes en cada universidad. Estudios como Montenegro (2005), Valens (2007), Sarmiento & Sandoval (2008), analizan el rendimiento con medidas descriptivas de centralización y dispersión para cada programa.

Montenegro (2005) encontró una correlación positiva entre los puntajes por programa (promedio ponderado o simple) y el valor de la matrícula, el número de estudiantes evaluados y el pertenecer a jornada diurna, y correlación negativa si la institución es privada.

Valens (2007) a través de un análisis multinivel encuentra diferencias significativas entre los puntajes medios de las universidades y entre los obtenidos por los estudiantes que son explicadas por sus características personales y por las características de la institución a la que pertenecen.

Sarmiento & Sandoval (2008) por medio de un análisis de covarianzas (ANCOVA), encontraron que el número de estudiantes evaluados, la jornada diurna y la acreditación de la institución tienen efectos positivos sobre el puntaje por institución.

Rosefsky & Saavedra (2011) a través de un análisis de regresión por mínimos cuadrados ordinarios y un enfoque basado en propensity scores para eliminar sesgos potenciales en la selección de la muestra, encuentran que la universidad aumenta las competencias del estudiante, sin embargo dicho aumento no ocurre de igual forma para todas las universidades. Pertenecer a una universidad privada aumenta las competencias en pensamiento crítico e interpersonal (organización y gestión de un grupo de trabajo), un número bajo de estudiantes por curso son asociados con mejores resultados en pensamiento crítico. No encuentran evidencia para asociar mejores resultados al porcentaje de profesores con doctorado, profesores de tiempo completo y gastos del estudiante. Es importante analizar si estas variables que son utilizadas para establecer un ranking entre universidades y en adjudicación de certificaciones de alta calidad, representan verdaderamente un valor agregado al estudiante, y si son relevantes para exámenes tipo SABER PRO.

En la literatura extranjera se consultaron estudios que analizan variables de entrada tanto académicas como socioeconómicas, provenientes de resultados en pruebas de estado estandarizadas similares a la prueba SABER 11, pruebas de admisión, el desempeño escolar, con el fin de predecir el rendimiento académico universitario

determinado principalmente por el promedio de las calificaciones de los semestres cursados.

Geiser y Studley (2001) a través de regresiones analizan la capacidad predictiva del SAT² I y del SAT II sobre el rendimiento académico en primer año para los estudiantes de la Universidad de California, encuentran evidencia a favor del SAT II frente a SAT I en la predicción de los promedios de pregrado para estudiantes de primer año, además de ser menos afectada por condiciones de origen y socioeconómicas.

Análisis de regresión y simulación realizados por Cohn, et al. (2004) en la Universidad de Carolina del Sur, incluyen además de los puntajes en el SAT, el ranking del estudiante en secundaria encontrándolo significativo estadísticamente pero pequeño en magnitud para predecir las notas en la universidad.

Bai & Chi (2011) encontraron evidencia para concluir que el CEE³ predice significativamente los promedios de pregrado de los estudiantes de la Universidad de Tsinghua en China; las pruebas en matemáticas y lengua extranjera mostraron una correlación más fuerte. Encontraron un mejor rendimiento académico para quienes obtuvieron algún tipo de reconocimiento en la escuela secundaria, al igual que para quienes ingresaron por recomendación, y menor rendimiento para quienes ingresaron por talentos especiales o mérito deportivo.

En nuestro caso asociamos el rendimiento académico universitario con el puntaje obtenido al finalizar la carrera en la prueba de estado estandarizada SABER PRO, en cambio del promedio en las asignaturas cursadas como se hace en varios estudios descritos anteriormente; esto permite generalizar los resultados y contrastarlos entre programas de diferentes universidades.

Además, se analiza una proporción significativa de los estudiantes que presentaron la prueba SABER PRO de economía a nivel nacional en el año 2009, en cambio de analizar casos particulares para una o dos universidades como suele suceder en algunos de los estudios mencionados.

² La SAT (Scholastic Assessment Test), test de admisión estandarizado que utilizan las universidades en EE.UU., similar al saber 11 en Colombia. El SAT I mide la aptitud verbal y matemática, haciendo énfasis en la capacidad de razonamiento, mientras que la SAT II es una prueba de conocimientos específicos en las diferentes materias como matemáticas, historia, ciencias entre otras.

³ Examen nacional de ingreso a la Universidad en China.

Fuentes de información, variables y datos

Fuentes de información

El ICFES hizo públicas a partir del 2011 las bases de datos con la información suministrada en las pruebas SABER 11 en el periodo comprendido entre los años 2000 a 2011 y SABER PRO entre 2004 a 2010. Estas bases contienen información general del estudiante, de las instituciones educativas, del entorno familiar, y del desempeño académico del estudiante al momento de completar el formulario de inscripción a la respectiva prueba. El acceso a dicha información se debe solicitar al ICFES, indicando la finalidad de la consulta. El protocolo puede ser consultado en ICFES (2011).

Resulta interesante utilizar la información contenida en ambas bases. Analizar las características y particularidades de un mismo estudiante antes de iniciar la universidad (SABER 11) y al finalizar (SABER PRO). Con este propósito el ICFES identificó los estudiantes que presentaron ambas pruebas, SABER 11 en el período 2000 a 2006 (terminando el ciclo de educación media) y SABER PRO en el período 2004 a 2009 (al terminar la carrera universitaria).

Para identificar a un mismo estudiante en ambas pruebas utilizaron en primera instancia el documento de identidad, sin embargo y como para el momento de presentación de las pruebas la gran mayoría de estudiantes cambiaron de documento (de tarjeta a cédula), fue necesario utilizar otras estrategias. En la primera estrategia se identificaron aquellos estudiantes que cumplen con al menos uno de los siguientes criterios en ambas pruebas:

1. Coincidencia de documento de identidad, nombre y fecha de nacimiento.
2. Coincidencia de nombre, fecha de nacimiento y género.
3. Coincidencia de fecha de nacimiento y género con distancias entre los apellidos y los nombres inferiores a 10.
4. Coincidencia de fecha de nacimiento y género con diferencias entre los nombres y apellidos, calculadas mediante un algoritmo programado en STATA que tiene en cuenta el orden de los mismos y algunos errores ortográficos, inferiores a cierto umbral.

En la segunda estrategia y con el fin de confirmar que se tratara del mismo estudiante, a aquellos estudiantes identificados en la estrategia anterior se les confirmó la coincidencia entre el año de identificación en la prueba SABER 11 y el registrado en el formulario de inscripción para la prueba SABER PRO, al diligenciar una pregunta que indica el año en el cual presentó el examen SABER 11.

A pesar de esto no es posible identificar a la totalidad de los estudiantes, entre otras cosas porque no todos los que presentaron SABER 11 han presentado SABER PRO, y porque no todos presentaron SABER 11 en ese periodo de tiempo 2000 a 2011.

El detalle metodológico y el cruce de códigos pueden ser consultados en las bases de datos del ICFES, carpeta “Tablas Genéricas” que se encuentra en una tabla llamada “SB11-SBPRO-CRUCES”. Para mayor información consultar ICFES (2011).

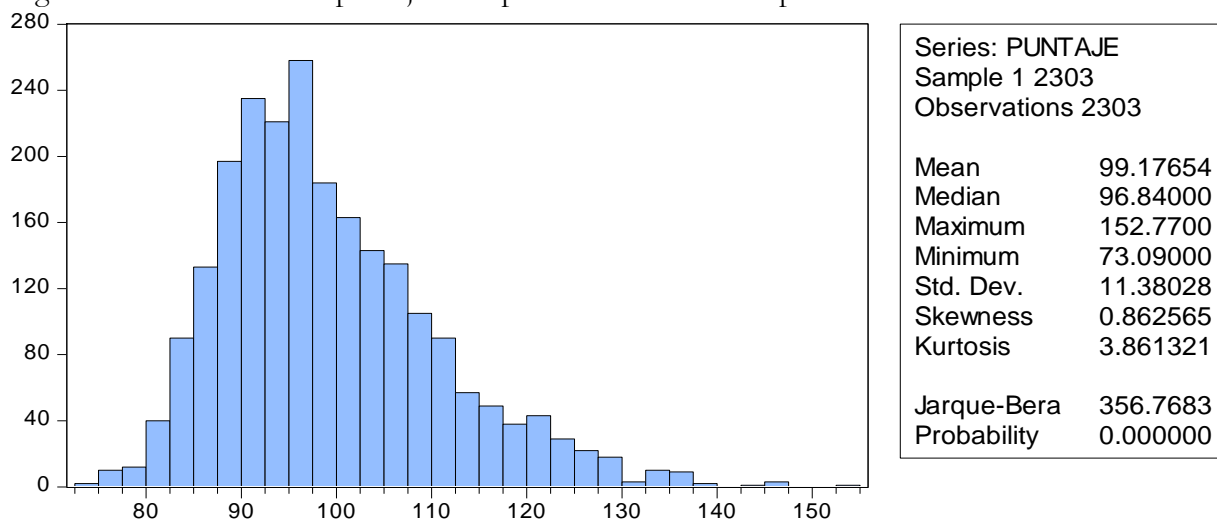
Variables y datos

El punto de partida lo constituyen 3491 estudiantes que presentaron la prueba SABER PRO en economía en el año 2009. Con ayuda de la tabla descrita anteriormente se identificaron 2303 que coinciden tanto en la base SABER PRO como en alguna de las bases de las pruebas SABER11.

Una vez depuradas las bases, se construye una sola base que incluye conjuntamente información académica y personal de cada estudiante en ambas pruebas. A continuación presentamos las variables incluidas y sus principales características descriptivas para estos 2303 estudiantes.

La variable de interés corresponde al puntaje general (puntaje) obtenido en la prueba SABER PRO 2009 para economía, esta es la variable que deseamos pronosticar. En la figura 1 encontramos las principales características descriptivas del puntaje para la muestra analizada.

Figura 1. Distribución del puntaje en la prueba SABER PRO para los estudiantes de la muestra.



Se observa un promedio muestral de 99.17 con una desviación estándar de 11.38. La variable se distribuye en el intervalo de 73.09 a 152.77 puntos, la mitad de los estudiantes presentan puntajes inferiores a 96.84, se observa un sesgo a la derecha en la distribución.

El examen SABER PRO para economía está compuesto por cuatro componentes: macroeconomía (macroeconomía), microeconomía (microeconomía), estadística y econometría (econometría), y pensamiento económico e historia económica (pensamiento). En la tabla 1 presentamos las principales características descriptivas

Tabla 1. Estadística descriptiva por componente en la prueba SABER PRO.

	ECONOMETRIA	MICROECONOMIA	MACROECONOMIA	PENSAMIENTO
Mean	9.879114	10.01824	9.822058	9.986496
Median	9.800000	9.900000	9.600000	10.10000
Maximum	14.80000	15.60000	16.60000	13.90000
Minimum	5.500000	7.100000	6.600000	0.000000
Std. Dev.	1.071085	1.067047	1.175872	1.222542
Skewness	0.507410	0.985162	0.840729	-0.243080
Kurtosis	3.700138	4.982920	4.532248	4.682983
Jarque-Bera	145.8619	749.8326	496.5925	294.4754
Probability	0.000000	0.000000	0.000000	0.000000
Sum	22751.60	23072.00	22620.20	22998.90
Sum Sq. Dev.	2640.905	2621.034	3182.919	3440.590
Observations	2303	2303	2303	2303

Como predictivas del rendimiento académico (puntaje) se analizaron en principio los resultados académicos en la prueba SABER 11. Estos resultados incluyen el puntaje obtenido en cada núcleo común (biología, matemáticas, filosofía, historia, lenguaje, física, química, interdisciplinar e idioma). El componente interdisciplinar indaga al estudiante sobre una de las siguientes problemáticas: medio ambiente, violencia y sociedad o medios de comunicación. Además se considera la edad (edads11) al momento de tomar la prueba y una variable dummy para el género, definida de la siguiente forma

$$\text{género} = \begin{cases} 1 & \text{hombre} \\ 0 & \text{mujer} \end{cases}$$

En la Tabla 2 encontramos las principales características descriptivas, para estas variables.

Tabla 2. Estadística descriptiva por algunas variables en la prueba SABER 11.

	GENERO	EDADS11	BIOLOGIA	MATEMATICAS	QUIMICA	FILOSOFIA
Mean	0.449414	22.96787	50.88088	46.44697	48.98987	48.92119
Median	0.000000	23.00000	50.79000	45.00000	48.00000	48.19000
Maximum	1.000000	46.00000	87.49000	92.78000	87.10000	88.00000
Minimum	0.000000	19.00000	30.00000	12.00000	26.00000	24.00000
Std. Dev.	0.497542	1.933250	6.655011	8.310155	7.024463	7.291517
Skewness	0.203388	2.010027	0.526450	1.213246	0.777716	0.816801
Kurtosis	1.041367	18.28494	4.033039	6.281516	4.585407	5.168446
Jarque-Bera	383.9975	23969.45	208.7830	1598.302	473.3510	707.2913
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Sum	1035.000	52895.00	117178.7	106967.4	112823.7	112665.5
Sum Sq. Dev.	569.8567	8603.622	101953.7	158973.1	113587.8	122388.6
Observations	2303	2303	2303	2303	2303	2303

	HISTORIA	LENGUAJE	IDIOMA	FISICA	INTERDISCIPLINAR
Mean	48.66690	56.46123	47.71647	49.02852	54.35230
Median	48.14000	56.00000	45.00000	48.54000	55.00000
Maximum	66.72000	104.6000	101.0000	85.06000	90.20000
Minimum	29.00000	33.00000	20.54000	21.11000	0.000000
Std. Dev.	5.802366	8.822539	10.98309	7.649228	9.351956
Skewness	0.015815	0.634627	1.216857	0.533807	-0.542307
Kurtosis	2.843591	4.574863	4.558150	4.632311	4.630652
Jarque-Bera	2.443523	392.5844	801.3282	365.0486	368.0396
Probability	0.294711	0.000000	0.000000	0.000000	0.000000

Sum	112079.9	130030.2	109891.0	112912.7	125173.4
Sum Sq. Dev.	77502.48	179181.2	277686.4	134691.6	201330.8
Observations	2303	2303	2303	2303	2303

Esta información es de conocimiento de las universidades en sus procesos de admisión. Resulta de interés conocer su valor predictivo para pronosticar el rendimiento académico en la universidad, esto con el fin de seleccionar los mejores candidatos. No se utilizan otras variables de interés como el rendimiento académico en la secundaria, la condición socioeconómica porque no se encuentra información completa al alcance.

Con el fin de hacer un acercamiento al valor agregado por el hecho de pertenecer a una universidad en particular se consideran variables dummies, para algunas universidades, definidas de la siguiente forma

$$u_{andes} = \begin{cases} 1 & \text{Pertenece a la Universidad de los Andes} \\ 0 & \text{No Pertenece a la Universidad de los Andes} \end{cases}$$

$$u_{javeriana} = \begin{cases} 1 & \text{Pertenece a la Universidad Javeriana} \\ 0 & \text{No Pertenece a la Universidad Javeriana} \end{cases}$$

$$u_{unal} = \begin{cases} 1 & \text{Pertenece a la Universidad Nacional de Colombia – Bogotá} \\ 0 & \text{No Pertenece a la Universidad Nacional de Colombia – Bogotá} \end{cases}$$

$$u_{rosario} = \begin{cases} 1 & \text{Pertenece a la Universidad del Rosario} \\ 0 & \text{No Pertenece a la Universidad del Rosario} \end{cases}$$

$$u_{norte} = \begin{cases} 1 & \text{Pertenece a la Universidad del Norte} \\ 0 & \text{No Pertenece a la Universidad del Norte} \end{cases}$$

$$u_{externado} = \begin{cases} 1 & \text{Pertenece a la Universidad Externado de Colombia} \\ 0 & \text{No Pertenece a la Universidad Externado de Coombia} \end{cases}$$

Para analizar conjuntamente las características académicas y socioeconómicas del estudiante, del entorno familiar y de la institución universitaria a la que pertenece el estudiante, se utilizaron algunas variables al momento de completar el formulario de inscripción a las pruebas SABER PRO, disponibles en las bases descritas anteriormente.

- Edad (edadspro): Edad del estudiante al tomar el examen SABER PRO.

- Estrato (estrato): Estrato socioeconómico de la vivienda donde reside el estudiante. Al momento de completar una pregunta del formulario de inscripción al examen SABER PRO, que indaga sobre el estrato socioeconómico de la vivienda según el recibo de energía eléctrica.

- Ingreso familiar (ingresofam): Ingreso familiar en salarios mínimos mensuales (SM) al momento de tomar esta prueba, según el ICFES en la siguiente escala:

1	Menos de 1 SM	5	Entre 5 y 7 SM
2	Entre 1 y 2 SM	6	Entre 7 y 10 SM
3	Entre 2 y 3 SM	7	Más de 10 SM
4	Entre 3 y 5 SM		

- Nivel educación de la madre (edumadre): Corresponde al último nivel educativo aprobado por la madre del estudiante, según el ICFES en la siguiente escala:

0	Ninguna	6	Técnica/tecnológica
1	Primaria incompleto	7	Profesional incompleto
2	Primaria completo	8	Profesional completo.
3	Secundaria incumplido	9	Posgrado
4	Secundaria completo		
5	Técnica/tecnológica incompleto		

En la misma escala se encuentra el nivel educativo del padre.

- Valor de la matrícula en la universidad (vmatricula): Valor anual de la matrícula del estudiante en el año anterior a la prueba (2008), según el ICFES en la siguiente escala:

0	No paga	3	Entre 1 y 3 millones
1	Menos de 1/2 millón	4	Entre 3 y 5 millones
2	Entre 1/2 y 1 millón	5	Más de 5 millones

- Origen de la Universidad (origenuni): Una variable dummy, definida de la siguiente forma:

$$origenuni = \begin{cases} 1 & \text{No oficial} \\ 0 & \text{Oficial} \end{cases}$$

- Conexión a internet (*internet*): Una variable dummy que caracteriza aquellos estudiantes cuyo hogar habitual cuenta con conexión a Internet.

$$internet = \begin{cases} 1 & \text{Si} \\ 0 & \text{No} \end{cases}$$

- Registro de alta calidad (*altacalidad*): El Concejo Nacional de acreditación (CNA) es el organismo encargado en revisar el proceso de acreditación de los programas, los orienta, organiza, fiscaliza, y da fe de su calidad. Si los programas cumplen con los criterios establecidos son certificados por el ministerio de educación previa recomendación del CNA. Esta variable vale uno si el programa es acreditado de alta calidad y cero en el caso contrario (Fuente: CNA consejo nacional de acreditación)

$$altacalidad = \begin{cases} 1 & \text{Si} \\ 0 & \text{No} \end{cases}$$

En la Tabla 3 se muestran las principales características descriptivas para estas variables.

Tabla 3. Estadística descriptiva para algunas variables en la prueba SABER PRO.

	EDADSPRO	ESTRATO	INGRESOFAM	EDUMADRE
Mean	22.96787	3.225358	3.789405	4.970039
Median	23.00000	3.000000	4.000000	4.000000
Maximum	46.00000	6.000000	7.000000	9.000000
Minimum	19.00000	1.000000	1.000000	0.000000
Std. Dev.	1.933250	1.179767	1.674003	2.459307
Skewness	2.010027	0.552623	0.499616	0.167921
Kurtosis	18.28494	3.055477	2.380376	1.884795
Jarque-Bera	23969.45	117.5150	132.6528	130.1648
Probability	0.000000	0.000000	0.000000	0.000000
Sum	52895.00	7428.000	8727.000	11446.00
Sum Sq. Dev.	8603.622	3204.039	6450.861	13922.93
Observations	2303	2303	2303	2303

	ORIGENUNI	INTERNET	ALTACALIDAD	VMATRICULA
Mean	0.553191	0.735562	0.414242	2.775510
Median	1.000000	1.000000	0.000000	3.000000
Maximum	1.000000	1.000000	1.000000	5.000000
Minimum	0.000000	0.000000	0.000000	0.000000
Std. Dev.	0.497271	0.441129	0.492698	1.406967
Skewness	-0.213980	-1.068229	0.348191	0.119502
Kurtosis	1.045788	2.141113	1.121237	1.936951
Jarque-Bera	384.0345	508.7845	385.2438	113.9213
Probability	0.000000	0.000000	0.000000	0.000000
Sum	1274.000	1694.000	954.0000	6392.000
Sum Sq. Dev.	569.2340	447.9574	558.8129	4556.939
Observations	2303	2303	2303	2303

Finalmente se consideran variables dummy para analizar la fuente de los recursos con los cuales el estudiante pagó la matrícula en la universidad el año anterior (2008).

$$matriculaest = \begin{cases} 1 & \text{El estudiante pagó su matrícula} \\ 0 & \text{Otra fuente de pagó} \end{cases}$$

$$matriculapad = \begin{cases} 1 & \text{Los padres pagaron la matrícula} \\ 0 & \text{Otra fuente de pagó} \end{cases}$$

$$matriculabeca = \begin{cases} 1 & \text{El estudiante tiene una beca que cubre su matrícula} \\ 0 & \text{Otra fuente de pagó} \end{cases}$$

$$matriculacred = \begin{cases} 1 & \text{El estudiante pago a través de un crédito educativo} \\ 0 & \text{Otra fuente de pagó} \end{cases}$$

Otras variables de interés para las cuales no se tiene suficiente información en el presente estudio son: el ranking en el colegio, el promedio acumulado en los estudios universitarios, el número de cursos perdidos, prácticas pedagógicas particulares, la ubicación geográfica, posibles discapacidades físicas y estados de

salud, la motivación del estudiante, la ansiedad, la disposición frente a la prueba, el conocimiento de la estructura de la misma, el carácter del colegio (público o privado), el número de estudiantes por curso, la jornada, entre otras. En futuros estudios y a medida que la información sea disponible se pueden incluir para estudiar su valor predictivo.

Metodología, resultados encontrados y discusión

Metodología

La propuesta no implica causalidad entre las variables. Se quiere brindar una herramienta para predecir los resultados en la prueba SABER PRO a través de la información contenida en las bases descritas anteriormente, en ningún momento se infiere causalidad en los resultados.

En secciones anteriores se citaron algunos autores que indagan sobre variables relevantes para predecir un buen desempeño académico en la universidad, se da cuenta de la utilización masiva de regresiones por mínimos cuadrados. Sin embargo el hecho de suponer que la media condicional explica el conjunto de datos en situaciones de asimetría en la variable respuesta, en presencia de datos atípicos y bajo el supuesto de normalidad, puede causar estimaciones poco confiables. En estas situaciones el efecto de las variables explicativas puede cambiar conformé cambie la distribución de la variable respuesta.

En adición a mínimos cuadrados ordinarios, se utilizará regresión cuantílica. Esta metodología permite hacer estimaciones con respecto a cualquier cuantil de la distribución de la variable. Fue propuesta por Koenker y Basset (1978), permite excluir supuestos importantes incluidos en los métodos tradicionales con respecto a la aleatoriedad de la relación

$$y_i = x_i\beta + u_i$$

Mientras que en mínimos cuadrados se supone $u_i \sim N(0, \sigma^2)$ la regresión cuantílica no necesita de dichos supuestos, clasificando al método como semiparamétrico. Su construcción es análoga al de mínimos cuadrados, partimos de una variable

$$Y \sim F(y) \text{ continua y monótona}$$

El τ - ésimo cuantil de Y es el valor de la variable $Q_y(\tau)$ que satisface

$$F(Q_y(\tau)) = \tau$$

un número de la distribución tal que la probabilidad de que ocurran valores menores es τ .

En mínimos cuadrados asumimos que el valor u_i condicionado a las observaciones es cero $E(u_i|x_i) = 0$, luego la media condicional a y_i con respecto a las observaciones es

$$E(y_i|x_i) = \beta x_i$$

En regresión cuantílica en cambio se asume que el τ - ésimo cuantil del error con respecto a las variables explicativas es cero $Q_\tau(u_{i,\tau}|x_i) = 0$, entonces

$$Q_\tau(y_i|x_i) = \beta_\tau x_i$$

En el caso de mínimos cuadrados se buscan valores de β que solucionen el problema de optimización

$$\min \sum_{i=1} (y_i - \beta x_i)^2$$

En regresión cuantílica no se minimiza la suma de errores al cuadrado, en su lugar se minimiza la ponderación de la suma de los errores absolutos, con pesos asimétricos:

$$\min \sum_{y_i \geq \beta_\tau x_i} \tau |y_i - \beta_\tau x_i| + \min \sum_{y_i < \beta_\tau x_i} (1 - \tau) |y_i - \beta_\tau x_i|$$

que es equivalente al problema de optimización

$$\min \sum_{i=1} \rho_\tau(y_i - \beta_\tau x_i)$$

Donde ρ_τ es una función de chequeo definida como

$$\rho_\tau(r) = \tau P^+(r) + (1 - \tau) P^-(r)$$

$P^+(r) = \max\{0, r\}$, $P^-(r) = \max\{0, -r\}$, permite ponderar asimétricamente cada observación dependiendo de su ubicación en la distribución de la variable dependiente Y . Este problema de minimización puede ser resuelto a través de programación lineal Koenker (2005).

Para cada cuantil τ resulta un vector de coeficientes de regresión permitiendo estimar el efecto de las variables explicativas en diferentes partes de la distribución. No significa esto que solo se utilice la información de cada cuantil por separado, se siguen utilizando todos los datos de la muestra convenientemente ponderados.

Esta metodología representa para nuestro análisis una ventaja importante entre otras cosas al permitir predecir puntajes con respecto a los cuantiles superiores de la distribución, que incluyen los mejores resultados en la SABER PRO. Sin embargo, se contrasta con las tradicionales regresiones por mínimos cuadrados utilizando errores estándar robustos de White para corregir heterocedasticidad.

Resultados encontrados y discusión

En la primera etapa se analizaron algunas características del estudiante antes de iniciar sus estudios universitarios; la edad, el género, y los resultados por componente en la prueba SABER 11 como variables que permiten predecir el puntaje en la prueba SABER PRO. En la tabla 4 se presentan los resultados encontrados a través de la regresión por mínimos cuadrados con errores estándar robustos de White para corregir heterocedasticidad.

Tabla 4. Regresión por mínimos cuadrados con errores robustos.

Dependent Variable: PUNTAJE				
Method: Least Squares				
Date: 10/22/12 Time: 15:46				
Sample: 1 2303				
Included observations: 2303				
White heteroskedasticity-consistent standard errors & covariance				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36.80657	2.968921	12.39729	0.0000
GENERO	3.555556	0.337639	10.53063	0.0000
EDADS11	-0.469854	0.093174	-5.042733	0.0000
BIOLOGIA	0.181203	0.035600	5.089988	0.0000
MATEMATICAS	0.194908	0.027447	7.101346	0.0000
FILOSOFIA	0.065283	0.025680	2.542126	0.0111
HISTORIA	0.274769	0.033113	8.297838	0.0000
LENGUAJE	0.200275	0.025725	7.785312	0.0000
IDIOMA	0.115040	0.021623	5.320319	0.0000
FISICA	0.061738	0.026380	2.340289	0.0194
QUIMICA	0.290886	0.033839	8.596193	0.0000
INTERDISCIPLINAR	0.048768	0.017345	2.811659	0.0050
R-squared	0.567670	Mean dependent var	99.17654	
Adjusted R-squared	0.565594	S.D. dependent var	11.38028	
S.E. of regression	7.500681	Akaike info criterion	6.873062	
Sum squared resid	128892.1	Schwarz criterion	6.902981	
Log likelihood	-7902.330	Hannan-Quinn criter.	6.883969	
F-statistic	273.4719	Durbin-Watson stat	1.761486	
Prob(F-statistic)	0.000000			

Todos los coeficientes resultan significativos al 5%. El coeficiente asociado a la variable dummy género resulta significativo y positivo, lo que predice mejores resultados en la prueba para los hombres. Por componentes se observan coeficientes positivos, con mayor magnitud para los asociados a química e historia. Un punto adicional en estos componentes aumenta de mayor forma el valor pronosticado en comparación con los demás. La edad arroja un coeficiente con signo negativo, lo cual predice mejores resultados en la SABER PRO para estudiantes más jóvenes a la hora de tomar la SABER 11.

A través de regresión cuantílica, queremos encontrar coeficientes de regresión para predecir el puntaje no solo con respecto a la media sino con respecto a cualquier

cuantil. Nos interesa predecir los mejores resultados al tomar por ejemplo el cuantil 0.9 que abarca el 10% de los mejores resultados, encontramos

Tabla 4. Regresión cuantílica (cuantil 0.9)

Dependent Variable: PUNTAJE
Method: Quantile Regression (tau = 0.9)
Date: 10/22/12 Time: 15:28
Sample: 1 2303
Included observations: 2303
Huber Sandwich Standard Errors & Covariance
Sparsity method: Kernel (Epanechnikov) using residuals
Bandwidth method: Hall-Sheather, bw=0.0262
Estimation successfully identifies unique optimal solution

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	38.34824	4.080306	9.398374	0.0000
GENERO	6.479576	0.667479	9.707539	0.0000
EDADS11	-0.518462	0.096994	-5.345270	0.0000
BIOLOGIA	0.088236	0.048665	1.813133	0.0699
MATEMATICAS	0.230103	0.054524	4.220205	0.0000
FILOSOFIA	0.032965	0.045937	0.717614	0.4731
HISTORIA	0.329970	0.060344	5.468110	0.0000
LENGUAJE	0.291547	0.047978	6.076671	0.0000
IDIOMA	0.143818	0.045869	3.135422	0.0017
FISICA	0.073468	0.043389	1.693237	0.0905
QUIMICA	0.363030	0.055306	6.564046	0.0000
INTERDISCIPLINAR	0.029439	0.027717	1.062113	0.2883
Pseudo R-squared	0.398358	Mean dependent var	99.17654	
Adjusted R-squared	0.395469	S.D. dependent var	11.38028	
S.E. of regression	12.16931	Objective	3274.759	
Quantile dependent var	114.6800	Restr. objective	5443.037	
Sparsity	42.22108	Quasi-LR statistic	1141.230	
Prob(Quasi-LR stat)	0.000000			

No resultan significativos los coeficientes para biología, filosofía, física e interdisciplinaria. Los demás coeficientes resultan significativos y con los mismos signos encontrados a través de mínimos cuadrados ordinarios. Los coeficientes asociados a química e historia siguen presentando un mayor efecto en la predicción con respecto a los demás. En la tabla 5 encontramos los coeficientes para cada decil de la distribución del puntaje.

Tabla 5. Regresión cuantílica para diferentes cuantiles del puntaje.

Quantile Process Estimates
Equation: SABER11
Specification: PUNTAJE C GENERO EDADS11 BIOLOGIA MATEMATICAS FILOSOFIA
HISTORIA LENGUAJE IDIOMA FISICA QUIMICA INTERDISCIPLINAR

	Quantile	Coefficient	Std. Error	t-Statistic	Prob.
C	0.100	36.01846	4.799687	7.504335	0.0000
	0.200	36.98792	4.187482	8.832973	0.0000
	0.300	38.36611	3.491735	10.98769	0.0000

	0.400	36.09386	3.573294	10.10100	0.0000
	0.500	33.78494	3.814286	8.857473	0.0000
	0.600	31.52581	3.389343	9.301452	0.0000
	0.700	30.88615	3.186532	9.692716	0.0000
	0.800	34.98849	3.562821	9.820447	0.0000
	0.900	38.34824	4.080306	9.398374	0.0000
<hr/>					
GENERO	0.100	2.470615	0.481131	5.135014	0.0000
	0.200	2.377229	0.420621	5.651711	0.0000
	0.300	2.441716	0.382646	6.381133	0.0000
	0.400	2.620479	0.394702	6.639127	0.0000
	0.500	2.870223	0.403925	7.105838	0.0000
	0.600	3.230584	0.413570	7.811455	0.0000
	0.700	3.768664	0.452607	8.326561	0.0000
	0.800	4.545150	0.549585	8.270157	0.0000
	0.900	6.479576	0.667479	9.707539	0.0000
<hr/>					
EDADS11	0.100	-0.375109	0.148135	-2.532210	0.0114
	0.200	-0.354919	0.130744	-2.714599	0.0067
	0.300	-0.387378	0.111219	-3.483013	0.0005
	0.400	-0.362998	0.110807	-3.275946	0.0011
	0.500	-0.309728	0.117377	-2.638747	0.0084
	0.600	-0.307166	0.108605	-2.828291	0.0047
	0.700	-0.316923	0.098907	-3.204267	0.0014
	0.800	-0.430183	0.084505	-5.090619	0.0000
	0.900	-0.518462	0.096994	-5.345270	0.0000
<hr/>					
BIOLOGIA	0.100	0.197379	0.042969	4.593488	0.0000
	0.200	0.200901	0.045351	4.429958	0.0000
	0.300	0.235910	0.039328	5.998455	0.0000
	0.400	0.218191	0.042110	5.181481	0.0000
	0.500	0.200576	0.043295	4.632808	0.0000
	0.600	0.224047	0.044676	5.014962	0.0000
	0.700	0.205587	0.046161	4.453676	0.0000
	0.800	0.138456	0.051436	2.691805	0.0072
	0.900	0.088236	0.048665	1.813133	0.0699
<hr/>					
MATEMATICAS	0.100	0.152429	0.032974	4.622693	0.0000
	0.200	0.165199	0.032318	5.111675	0.0000
	0.300	0.162546	0.027591	5.891312	0.0000
	0.400	0.187400	0.029064	6.447904	0.0000
	0.500	0.209866	0.031686	6.623306	0.0000
	0.600	0.212408	0.031423	6.759718	0.0000
	0.700	0.207090	0.035322	5.862888	0.0000
	0.800	0.212534	0.046401	4.580399	0.0000
	0.900	0.230103	0.054524	4.220205	0.0000
<hr/>					
FILOSOFIA	0.100	0.116419	0.041653	2.794993	0.0052
	0.200	0.092724	0.035165	2.636802	0.0084
	0.300	0.056796	0.029322	1.936998	0.0529
	0.400	0.056945	0.030928	1.841203	0.0657
	0.500	0.056022	0.031939	1.754012	0.0796
	0.600	0.057264	0.032508	1.761526	0.0783
	0.700	0.044712	0.035246	1.268559	0.2047
	0.800	0.058467	0.031192	1.874428	0.0610
	0.900	0.032965	0.045937	0.717614	0.4731
<hr/>					
HISTORIA	0.100	0.213541	0.039389	5.421334	0.0000
	0.200	0.179237	0.040662	4.407925	0.0000
	0.300	0.225512	0.037436	6.023945	0.0000
	0.400	0.229199	0.038289	5.985962	0.0000

	0.500	0.263708	0.039042	6.754544	0.0000
	0.600	0.267167	0.041372	6.457650	0.0000
	0.700	0.319854	0.040346	7.927883	0.0000
	0.800	0.332368	0.052331	6.351311	0.0000
	0.900	0.329970	0.060344	5.468110	0.0000
LENGUAJE	0.100	0.171945	0.031055	5.536835	0.0000
	0.200	0.176475	0.031662	5.573647	0.0000
	0.300	0.158995	0.028984	5.485599	0.0000
	0.400	0.175980	0.026237	6.707235	0.0000
	0.500	0.186844	0.029054	6.430820	0.0000
	0.600	0.218501	0.033104	6.600491	0.0000
	0.700	0.237957	0.033401	7.124156	0.0000
	0.800	0.251114	0.037960	6.615273	0.0000
	0.900	0.291547	0.047978	6.076671	0.0000
IDIOMA	0.100	0.057272	0.021366	2.680514	0.0074
	0.200	0.101898	0.028617	3.560787	0.0004
	0.300	0.115768	0.024713	4.684570	0.0000
	0.400	0.109150	0.024113	4.526678	0.0000
	0.500	0.113656	0.027997	4.059508	0.0001
	0.600	0.127009	0.028192	4.505136	0.0000
	0.700	0.119493	0.028084	4.254901	0.0000
	0.800	0.112283	0.032968	3.405836	0.0007
	0.900	0.143818	0.045869	3.135422	0.0017
FISICA	0.100	0.083854	0.036546	2.294463	0.0219
	0.200	0.087971	0.037597	2.339838	0.0194
	0.300	0.035810	0.030865	1.160224	0.2461
	0.400	0.051746	0.028891	1.791064	0.0734
	0.500	0.033543	0.031081	1.079189	0.2806
	0.600	0.034237	0.031030	1.103332	0.2700
	0.700	0.051753	0.032530	1.590943	0.1118
	0.800	0.105678	0.037313	2.832234	0.0047
	0.900	0.073468	0.043389	1.693237	0.0905
QUIMICA	0.100	0.192044	0.043610	4.403695	0.0000
	0.200	0.206890	0.042494	4.868678	0.0000
	0.300	0.236731	0.040025	5.914617	0.0000
	0.400	0.263176	0.039861	6.602297	0.0000
	0.500	0.289871	0.040601	7.139450	0.0000
	0.600	0.300923	0.040140	7.496843	0.0000
	0.700	0.307582	0.044097	6.975166	0.0000
	0.800	0.324158	0.047564	6.815134	0.0000
	0.900	0.363030	0.055306	6.564046	0.0000
INTERDISCIPLINAR	0.100	0.048472	0.026142	1.854220	0.0638
	0.200	0.048352	0.020099	2.405747	0.0162
	0.300	0.062615	0.017517	3.574591	0.0004
	0.400	0.065136	0.018286	3.562132	0.0004
	0.500	0.059395	0.020200	2.940413	0.0033
	0.600	0.048298	0.021848	2.210663	0.0272
	0.700	0.048202	0.025152	1.916400	0.0554
	0.800	0.024283	0.025860	0.939011	0.3478
	0.900	0.029439	0.027717	1.062113	0.2883

El coeficiente asociado al género es positivo y significativo para todos los cuantiles de la distribución del puntaje. En la figura 2 se puede observar un incremento en su

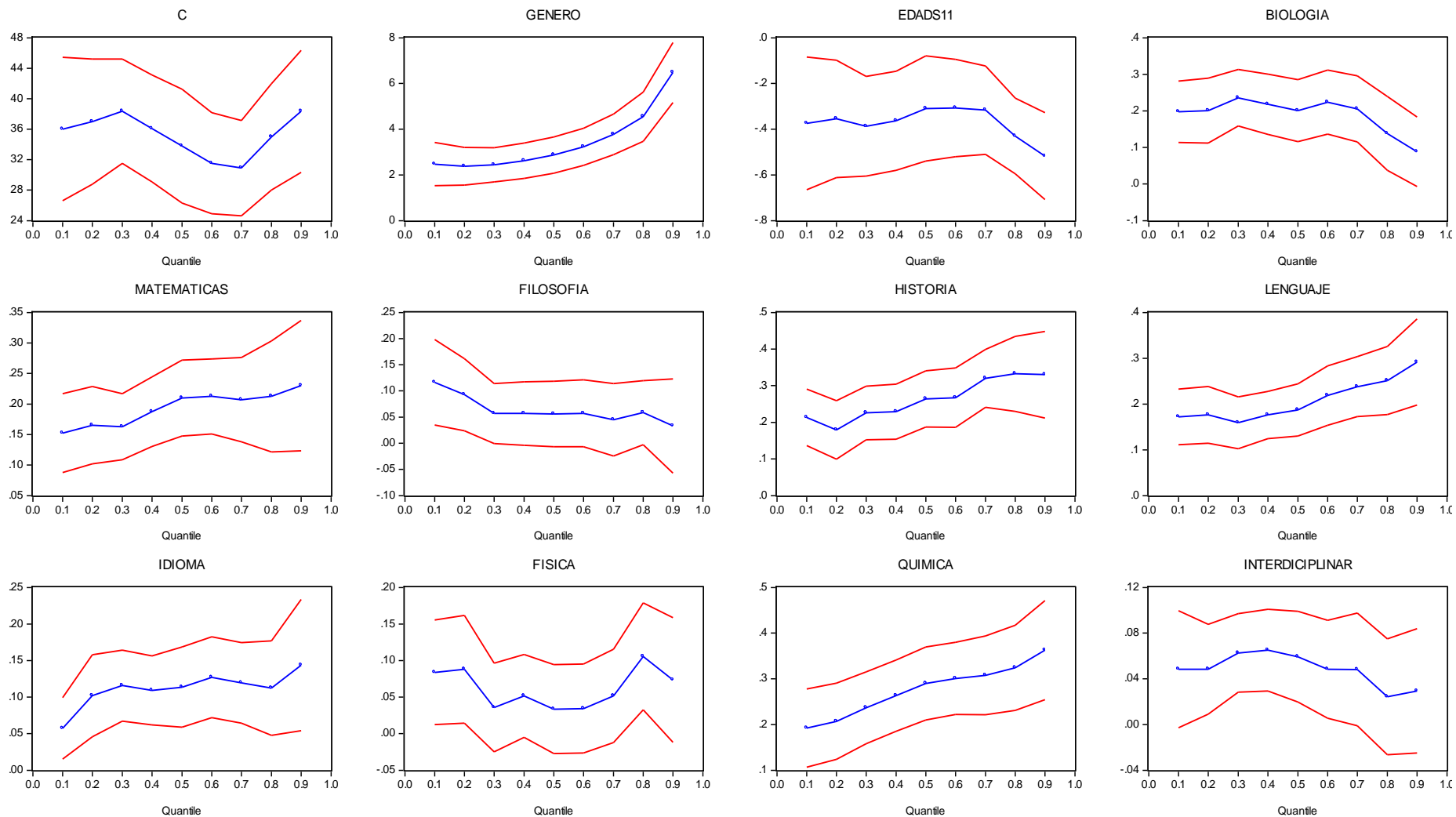
valor a medida que se predice con respecto a cuantiles superiores (puntajes más altos), es decir su efecto es mayor para predecir puntajes altos.

Los coeficientes asociados a la edad son significativos y de signo negativo para todos los deciles, se puede decir que son similares en magnitud, de manera que el efecto sobre la predicción no es afectado por la distribución del puntaje.

Los coeficientes asociados a matemáticas, historia, lenguaje, idioma y química son todos significativos y positivos independientemente del cuantil, se observa en todos ellos una tendencia creciente (figura 2) a medida que se predice con respecto a cuantiles superiores. Su efecto en la predicción es más relevante a medida que aumenta los puntajes en la SABER PRO. El caso de biología es diferente, decrece en magnitud a medida que aumenta el cuantil, hasta dejar de ser significativo en el cuantil 0.9.

Los coeficientes asociados a la prueba interdisciplinar son significativos en los cuantiles intermedios de la distribución, presentan signo positivo pero cercano a cero, lo que no implica un efecto importante en la predicción. Los coeficientes asociados a filosofía y física en su gran mayoría no son significativos.

Figura 2. Tendencia de los coeficientes estimados a través de regresión cuantílica.
 Quantile Process Estimates (95% CI)



A las estimaciones anteriores se les incluyen variables dummies para distinguir el efecto en la predicción dependiendo de la universidad donde se cursó el programa. En la tabla 6 se observa los resultados a través de mínimos cuadrados robustos.

Tabla 6. Regresión por mínimos cuadrados, agregando dummies por universidad.

Dependent Variable: PUNTAJE
Method: Least Squares
Date: 11/14/12 Time: 15:18
Sample: 1 2303
Included observations: 2303
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	44.64851	3.114531	14.33555	0.0000
GENERO	3.321146	0.332008	10.00322	0.0000
EDADS11	-0.457898	0.090791	-5.043455	0.0000
BIOLOGIA	0.161618	0.035096	4.605038	0.0000
MATEMATICAS	0.172834	0.027939	6.186099	0.0000
FILOSOFIA	0.063929	0.025894	2.468848	0.0136
HISTORIA	0.259934	0.032300	8.047599	0.0000
LENGUAJE	0.173457	0.025462	6.812302	0.0000
IDIOMA	0.061648	0.022593	2.728604	0.0064
FISICA	0.051869	0.025928	2.000518	0.0456
QUIMICA	0.259800	0.033202	7.824875	0.0000
INTERDISCIPLINAR	0.046120	0.016671	2.766570	0.0057
UNAL	5.032164	0.778510	6.463836	0.0000
UANDES	4.574147	0.831257	5.502690	0.0000
UJAVERIANA	3.645256	1.249381	2.917651	0.0036
UNORTE	9.179503	2.114188	4.341857	0.0000
UROSARIO	7.985463	1.843703	4.331208	0.0000
UEXTERNADO	6.728442	1.333813	5.044516	0.0000
R-squared	0.589781	Mean dependent var		99.17654
Adjusted R-squared	0.586729	S.D. dependent var		11.38028
S.E. of regression	7.315942	Akaike info criterion		6.825774
Sum squared resid	122300.1	Schwarz criterion		6.870652
Log likelihood	-7841.879	Hannan-Quinn criter.		6.842135
F-statistic	193.2468	Durbin-Watson stat		1.800370
Prob(F-statistic)	0.000000			

No hay modificación en el signo y la significancia de las variables originales. Los coeficientes asociados a cada universidad resultan positivos y significativos, con diferencias considerables en magnitud. Por ejemplo, la Universidad del Norte adiciona 9 puntos en la predicción del puntaje en la prueba SABER PRO (que tiene una media muestral de 99.1765). En la tabla 7 encontramos los resultados de las regresiones cuantílicas.

Tabla 7. Regresión cuantílica para diferentes cuantiles del puntaje, agregando dummies por Universidad.

Quantile Process Estimates

Equation: SABER11UNIV

Specification: PUNTAJE C GENERO EDADS11 BIOLOGIA MATEMATICAS FILOSOFIA
HISTORIA LENGUAJE IDIOMA FISICA QUIMICA INTERDISCIPLINAR UNAL
UANDES UJAVERIANA UNORTE UROSARIO UEXTERNADO

	Quantile	Coefficient	Std. Error	t-Statistic	Prob.
UNAL	0.100	5.194016	1.267445	4.098019	0.0000
	0.200	5.787069	0.962257	6.014061	0.0000
	0.300	5.964707	1.048691	5.687764	0.0000
	0.400	6.310916	0.781236	8.078121	0.0000
	0.500	6.247258	0.786283	7.945304	0.0000
	0.600	5.577098	0.814994	6.843118	0.0000
	0.700	4.848390	0.861818	5.625769	0.0000
	0.800	5.020144	1.127325	4.453148	0.0000
UANDES	0.100	5.880430	1.429023	4.114999	0.0000
	0.200	2.833244	1.073545	2.639149	0.0084
	0.300	4.910217	1.120725	4.381288	0.0000
	0.400	5.163106	1.211643	4.261244	0.0000
	0.500	5.522176	0.932452	5.922211	0.0000
	0.600	5.851035	0.991098	5.903587	0.0000
	0.700	6.353212	1.052716	6.035070	0.0000
	0.800	4.968244	1.036287	4.794275	0.0000
UJAVERIANA	0.800	4.987699	1.340103	3.721877	0.0002
	0.900	4.519200	1.828695	2.471271	0.0135
	0.100	1.005793	1.726997	0.582394	0.5604
	0.200	3.833728	2.149952	1.783170	0.0747
	0.300	4.949063	1.088787	4.545485	0.0000
	0.400	3.586019	1.044053	3.434711	0.0006
	0.500	3.965655	1.190259	3.331758	0.0009
	0.600	4.375438	1.428358	3.063265	0.0022
UNORTE	0.700	3.017827	1.297403	2.326052	0.0201
	0.800	4.835238	2.218662	2.179349	0.0294
	0.900	4.470539	1.505329	2.969809	0.0030
	0.100	8.531647	1.470805	5.800665	0.0000
	0.200	6.633462	2.012563	3.296027	0.0010
	0.300	9.333739	2.344692	3.980796	0.0001
	0.400	8.302847	2.654893	3.127375	0.0018
	0.500	9.693981	2.310479	4.195659	0.0000
UROSARIO	0.600	8.668956	2.080596	4.166574	0.0000
	0.700	7.018666	1.708603	4.107838	0.0000
	0.800	6.497599	3.399931	1.911097	0.0561
	0.900	4.403208	2.240820	1.964998	0.0495
	0.100	4.590533	1.549100	2.963354	0.0031
	0.200	8.910024	4.261883	2.090631	0.0367
	0.300	8.709896	1.578699	5.517134	0.0000
	0.400	8.593778	1.616502	5.316281	0.0000
UEXTERNADO	0.500	8.323170	1.835454	4.534667	0.0000
	0.600	8.771668	2.758722	3.179613	0.0015
	0.700	9.748644	2.304955	4.229429	0.0000
	0.800	8.142324	1.838548	4.428670	0.0000
	0.900	12.10738	2.652541	4.564447	0.0000
	0.100	5.254908	1.319753	3.981737	0.0001

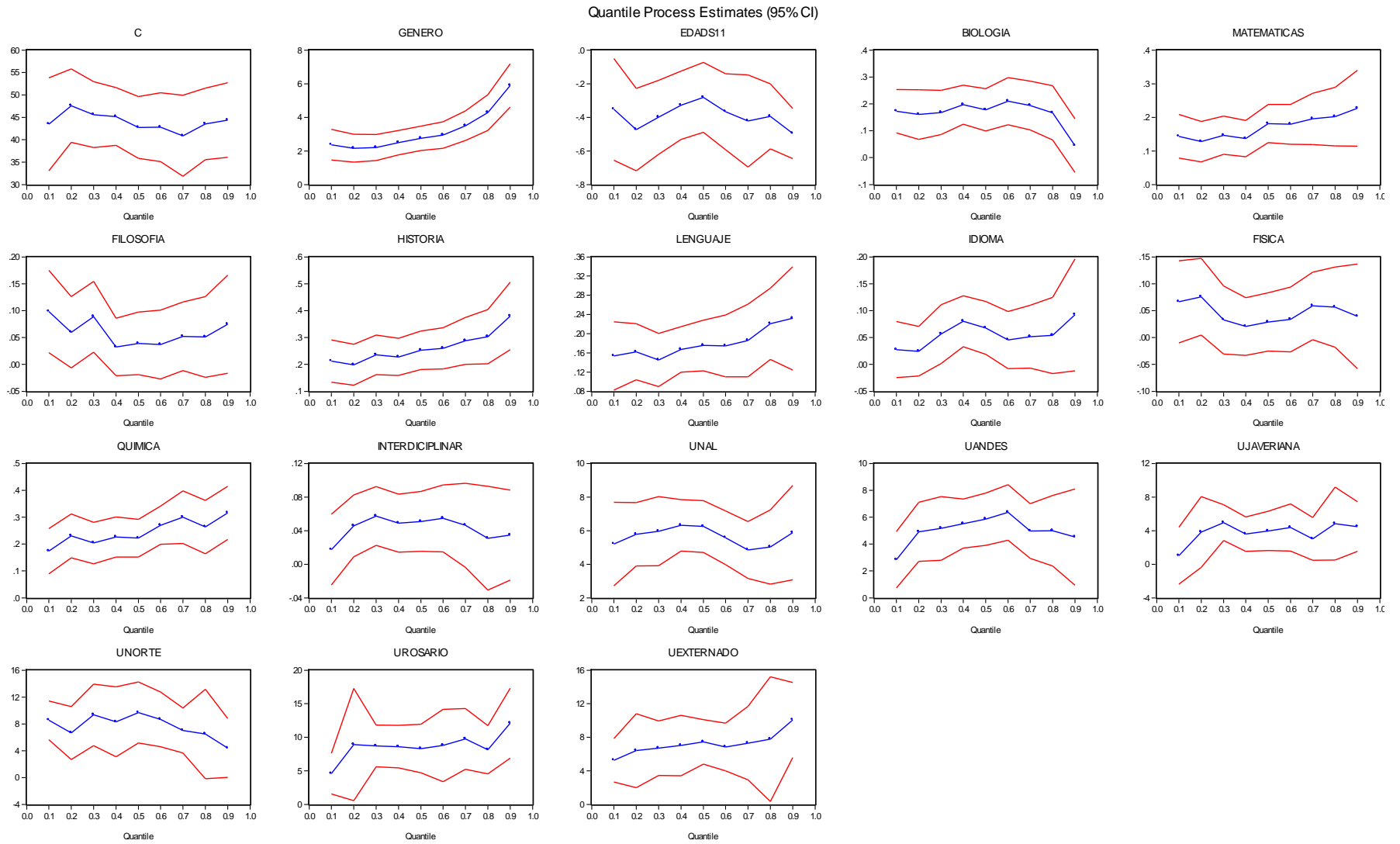
0.200	6.402461	2.253764	2.840786	0.0045
0.300	6.692384	1.651837	4.051481	0.0001
0.400	7.019363	1.834790	3.825703	0.0001
0.500	7.450331	1.355193	5.497618	0.0000
0.600	6.841332	1.453157	4.707909	0.0000
0.700	7.300923	2.238171	3.262005	0.0011
0.800	7.765005	3.790070	2.048776	0.0406
0.900	10.05577	2.281460	4.407603	0.0000

En la tabla anterior se observa que los coeficientes asociados a las universidades Rosario y Externado son todos significativos y positivos independientemente del cuantil. Además, presentan una tendencia creciente a medida que se predice con respecto a cuantiles superiores (figura 3), su efecto en la predicción es más relevante para puntajes altos de la distribución.

Los coeficientes asociados a las Universidades Nacional y Andes son significativos y presentan signo positivo a lo largo de la distribución, su efecto es similar a lo largo de la distribución del puntaje (figura 3).

Los coeficientes asociados a la universidad del Norte son decrecientes en magnitud y pierden su fuerza de significancia a medida que se predice con respecto a tramos superiores en el puntaje. Los coeficientes asociados a la universidad Javeriana no son significativos para varios tramos de la distribución.

Figura 3. Tendencia de los coeficientes estimados a través de regresión cuantílica, incluidas dummies por Universidad.



Queremos analizar si las estimaciones dependen del componente específico de la prueba, o si por el contrario son consistentes con los resultados obtenidos para el puntaje general. En la tabla 8 encontramos las estimaciones para el puntaje de cada componente. Elegimos el cuantil 0.9, del 10% de los mejores puntajes de la muestra.

Los coeficientes asociados a género, química, historia y lenguaje son significativos y positivos para todos los componentes de la prueba, el mismo resultado se obtuvo con respecto al puntaje general. Por su parte los coeficientes asociados a filosofía, física e interdisciplinar no son significativos para la predicción de los puntajes por componente, siendo este resultado consistente con los resultados obtenidos con respecto al puntaje general.

En estimaciones anteriores encontramos que los coeficientes asociados a las variables edad, matemáticas e idiomas son significativos con respecto al puntaje general. Por componentes algunos de ellos dejan de ser significativos, el coeficiente asociado a la edad deja de ser significativo al 2.94 para el componente de macroeconomía, el coeficiente asociado a matemáticas no es significativo para componente de pensamiento económico, el asociado a idioma no es significativo ni para pensamiento económico, ni para macroeconomía.

Los coeficientes asociados a la biología no son significativos, excepto el coeficiente para predecir el puntaje del componente de macroeconomía donde su signo es positivo.

COMPONENTE DE LA PRUEBA SABER PRO

VARIABLES	MACROECONOMIA	MICROECONOMIA	PENSAMIENTO	ECONOMETRIA
GENERO	C/P	C/P	C/P	C/P
EDADS11	C	C/P	C/P	C/P
BIOLOGIA	P	C/P	C/P	C/P
MATEMATICAS	C/P	C/P	C	C/P
FILOSOFIA	C/P	C/P	C/P	C/P
HISTORIA	C/P	C/P	C/P	C/P
LENGUAJE	C/P	C/P	C/P	C/P
IDIOMA	C	C/P	C	C/P
FISICA	C/P	C/P	C/P	C/P
QUIMICA	C/P	C/P	C/P	C/P
INTERDISCIPLINAR	C/P	C/P	C/P	C/P

	SIGNIFICATIVAS Y SIGNO POSITIVO
	SIGNIFICATIVAS Y SIGNO NEGATIVO
	NO SIGNIFICATIVA
C/P	EN EL COMPONENTE ESPECIFICO Y EN LA PRUEBA
C	EN EL COMPONENTE

Tabla 8. Regresión cuantílica (cuantil 0.9) para cada componente de la prueba SABER PRO como variable dependiente.

Dependent Variable: MACROECONOMIA					Dependent Variable: MICROECONOMIA				
Method: Quantile Regression (tau = 0.9)					Method: Quantile Regression (tau = 0.9)				
Date: 11/14/12 Time: 08:09					Date: 11/14/12 Time: 08:10				
Sample: 1 2303					Sample: 1 2303				
Included observations: 2303					Included observations: 2303				
Huber Sandwich Standard Errors & Covariance					Huber Sandwich Standard Errors & Covariance				
Sparsity method: Kernel (Epanechnikov) using residuals					Sparsity method: Kernel (Epanechnikov) using residuals				
Bandwidth method: Hall-Sheather, bw=0.0262					Bandwidth method: Hall-Sheather, bw=0.0262				
Estimation successfully identifies unique optimal solution					Estimation successfully identifies unique optimal solution				
Variable	Coefficient	Std. Error	t-Statistic	Prob.	Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.457000	0.446983	9.971292	0.0000	C	4.549489	0.386737	11.76378	0.0000
GENERO	0.619175	0.072471	8.543706	0.0000	GENERO	0.395812	0.066324	5.967826	0.0000
EDADS11	-0.052467	0.012166	-4.312453	0.0000	EDADS11	-0.036034	0.007406	-4.865543	0.0000
BIOLOGIA	0.020286	0.007122	2.848331	0.0044	BIOLOGIA	0.006205	0.007846	0.790946	0.4291
MATEMATICAS	0.023086	0.004282	5.391236	0.0000	MATEMATICAS	0.029583	0.005296	5.585863	0.0000
FILOSOFIA	0.005187	0.005719	0.906906	0.3646	FILOSOFIA	0.004947	0.004688	1.055095	0.2915
HISTORIA	0.031462	0.006819	4.613906	0.0000	HISTORIA	0.022811	0.006569	3.472424	0.0005
LENGUAJE	0.016539	0.005137	3.219308	0.0013	LENGUAJE	0.022571	0.004499	5.016647	0.0000
IDIOMA	0.009469	0.005163	1.833986	0.0668	IDIOMA	0.017258	0.004375	3.944427	0.0001
FISICA	0.003345	0.004903	0.682141	0.4952	FISICA	0.013513	0.006141	2.200570	0.0279
QUIMICA	0.035455	0.008015	4.423473	0.0000	QUIMICA	0.026352	0.006062	4.347306	0.0000
INTERDISCIPLINAR	0.004152	0.002606	1.593370	0.1112	INTERDISCIPLINAR	0.000142	0.003007	0.047293	0.9623
Pseudo R-squared	0.324764	Mean dependent var	9.822058		Pseudo R-squared	0.366510	Mean dependent var	10.01824	
Adjusted R-squared	0.321522	S.D. dependent var	1.175872		Adjusted R-squared	0.363468	S.D. dependent var	1.067047	
S.E. of regression	1.439636	Objective	374.1281		S.E. of regression	1.258363	Objective	323.5360	
Quantile dependent var	11.30000	Restr. objective	554.0700		Quantile dependent var	11.40000	Restr. objective	510.7200	
Sparsity	4.687732	Quasi-LR statistic	853.0158		Sparsity	4.416399	Quasi-LR statistic	941.8633	
Prob(Quasi-LR stat)	0.000000				Prob(Quasi-LR stat)	0.000000			

Dependent Variable: PENSAMIENTO
Method: Quantile Regression (tau = 0.9)
Date: 11/14/12 Time: 08:12
Sample: 1 2303
Included observations: 2303
Huber Sandwich Standard Errors & Covariance
Sparsity method: Kernel (Epanechnikov) using residuals
Bandwidth method: Hall-Sheather, bw=0.0262
Estimation successfully identifies unique optimal solution

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.019352	0.542390	9.254135	0.0000
GENERO	0.432599	0.068060	6.356164	0.0000
EDADS11	-0.034285	0.010813	-3.170549	0.0015
BIOLOGIA	0.003385	0.006407	0.528329	0.5973
MATEMATICAS	0.007086	0.005585	1.268650	0.2047
FILOSOFIA	0.006497	0.005365	1.211178	0.2260
HISTORIA	0.052168	0.006768	7.708338	0.0000
LENGUAJE	0.024191	0.005396	4.482953	0.0000
IDIOMA	0.006520	0.003919	1.663697	0.0963
FISICA	0.000955	0.004711	0.202715	0.8394
QUIMICA	0.026184	0.006186	4.233055	0.0000
INTERDISCIPLINAR	0.006261	0.004162	1.504493	0.1326

Pseudo R-squared	0.240367	Mean dependent var	9.986496
Adjusted R-squared	0.236720	S.D. dependent var	1.222542
S.E. of regression	1.490833	Objective	365.6037
Quantile dependent var	11.60000	Restr. objective	481.2900
Sparsity	4.686006	Quasi-LR statistic	548.6135
Prob(Quasi-LR stat)	0.000000		

Dependent Variable: ECONOMETRIA
Method: Quantile Regression (tau = 0.9)
Date: 11/14/12 Time: 08:14
Sample: 1 2303
Included observations: 2303
Huber Sandwich Standard Errors & Covariance
Sparsity method: Kernel (Epanechnikov) using residuals
Bandwidth method: Hall-Sheather, bw=0.0262
Estimation successfully identifies unique optimal solution

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.436997	0.904719	7.114912	0.0000
GENERO	0.368400	0.078703	4.680889	0.0000
EDADS11	-0.054425	0.024972	-2.179460	0.0294
BIOLOGIA	0.003319	0.008537	0.388719	0.6975
MATEMATICAS	0.014459	0.005138	2.813806	0.0049
FILOSOFIA	0.003917	0.004190	0.934950	0.3499
HISTORIA	0.020559	0.007644	2.689516	0.0072
LENGUAJE	0.020276	0.005654	3.586038	0.0003
IDIOMA	0.017704	0.005352	3.307932	0.0010
FISICA	0.010221	0.006091	1.677873	0.0935
QUIMICA	0.021950	0.007399	2.966695	0.0030
INTERDISCIPLINAR	0.001072	0.004287	0.250024	0.8026

Pseudo R-squared	0.259409	Mean dependent var	9.879114
Adjusted R-squared	0.255853	S.D. dependent var	1.071085
S.E. of regression	1.434563	Objective	362.8598
Quantile dependent var	11.40000	Restr. objective	489.9600
Sparsity	5.396974	Quasi-LR statistic	523.3393
Prob(Quasi-LR stat)	0.000000		

Para corroborar parte de la literatura en cuanto al valor predictivo de otras variables de interés como el género, la edad, el estrato, el ingreso familiar, la educación de los padres, el origen de la universidad, la acreditación de alta calidad y el valor de la matrícula, se estimó la regresión mostrada en la tabla 9. Los coeficientes son significativos para todas las variables mencionadas.

Tabla 9. Regresión por mínimos cuadrados, variables de SABER PRO.
 Dependent Variable: PUNTAJE
 Method: Least Squares
 Date: 11/13/12 Time: 11:00
 Sample: 1 2303
 Included observations: 2303
 White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	96.93592	2.639502	36.72508	0.0000
GENERO	4.629421	0.422402	10.95976	0.0000
EDADSPRO	-0.595007	0.154089	-3.861443	0.0001
ESTRATO	1.262874	0.215088	5.871425	0.0000
EDUCACIONMADRE	0.485391	0.092223	5.263230	0.0000
ORIGENUNI	-1.136775	0.439845	-2.584488	0.0098
INTERNET	1.321755	0.474538	2.785350	0.0054
ALTACALIDAD	8.145905	0.450483	18.08259	0.0000
R-squared	0.280662	Mean dependent var	99.17654	
Adjusted R-squared	0.278468	S.D. dependent var	11.38028	
S.E. of regression	9.666756	Akaike info criterion	7.378730	
Sum squared resid	214459.0	Schwarz criterion	7.398676	
Log likelihood	-8488.608	Hannan-Quinn criter.	7.386002	
F-statistic	127.9188	Durbin-Watson stat	1.578222	
Prob(F-statistic)	0.000000			

Los coeficientes asociados a las variables género, estrato, educación de la madre, conexión a internet y registró de alta calidad presentan signo positivo. Por ejemplo, pertenecer a una universidad con registro de alta calidad adiciona 8 puntos en la predicción del puntaje en la SABER PRO. Los resultados a través de regresión cuantílica se muestran en la tabla 10.

Tabla 10. Regresión cuantílica para diferentes cuantiles del puntaje, con variables SABER PRO.

Quantile Process Estimates

Equation: SABERPRO

Specification: PUNTAJE C GENERO EDADSPRO ESTRATO EDUCACIONMADRE

ORIGENUNI INTERNET ALTACALIDAD

	Quantile	Coefficient	Std. Error	t-Statistic	Prob.
C	0.100	90.82017	6.726111	13.50263	0.0000
	0.200	90.49423	3.115740	29.04422	0.0000
	0.300	95.44539	4.182803	22.81853	0.0000
	0.400	95.12110	2.592411	36.69214	0.0000
	0.500	95.33750	5.374833	17.73776	0.0000
	0.600	96.60356	5.017610	19.25290	0.0000
	0.700	96.16313	1.871407	51.38547	0.0000
	0.800	100.2268	2.232517	44.89409	0.0000
	0.900	108.2432	2.185279	49.53290	0.0000
GENERO	0.100	2.833707	0.558058	5.077800	0.0000
	0.200	2.888000	0.484904	5.955816	0.0000
	0.300	3.624605	0.471692	7.684256	0.0000
	0.400	3.708621	0.489232	7.580499	0.0000
	0.500	4.147500	0.500310	8.289861	0.0000
	0.600	4.774237	0.501153	9.526501	0.0000
	0.700	5.253988	0.552858	9.503320	0.0000
	0.800	5.538182	0.660770	8.381410	0.0000
	0.900	6.214178	0.845443	7.350204	0.0000
EDADSPRO	0.100	-0.520259	0.382839	-1.358949	0.1743
	0.200	-0.403154	0.176295	-2.286813	0.0223
	0.300	-0.621316	0.247817	-2.507156	0.0122
	0.400	-0.585103	0.152404	-3.839172	0.0001
	0.500	-0.534643	0.329310	-1.623524	0.1046
	0.600	-0.534576	0.310595	-1.721139	0.0854
	0.700	-0.396074	0.098827	-4.007736	0.0001
	0.800	-0.485455	0.114630	-4.234976	0.0000
	0.900	-0.745335	0.094189	-7.913149	0.0000
ESTRATO	0.100	0.089741	0.299813	0.299325	0.7647
	0.200	0.711000	0.252174	2.819482	0.0049
	0.300	0.620921	0.258332	2.403581	0.0163
	0.400	1.031586	0.291266	3.541732	0.0004
	0.500	1.407500	0.257396	5.468222	0.0000
	0.600	1.549831	0.231865	6.684194	0.0000
	0.700	1.599387	0.264126	6.055395	0.0000
	0.800	1.494545	0.313287	4.770530	0.0000
	0.900	1.084077	0.472101	2.296281	0.0217
EDUCACIONMADRE	0.100	0.236638	0.128633	1.839637	0.0660
	0.200	0.327692	0.101520	3.227862	0.0013
	0.300	0.445132	0.105173	4.232395	0.0000
	0.400	0.517724	0.100721	5.140157	0.0000
	0.500	0.526786	0.108687	4.846798	0.0000
	0.600	0.492203	0.108010	4.556998	0.0000
	0.700	0.452638	0.118131	3.831646	0.0001
	0.800	0.382273	0.157592	2.425708	0.0154
	0.900	0.741359	0.228349	3.246600	0.0012
ORIGENUNI	0.100	-0.691552	0.545140	-1.268576	0.2047
	0.200	-1.416615	0.509215	-2.781959	0.0054
	0.300	-1.365789	0.469882	-2.906665	0.0037

	0.400	-1.523103	0.491706	-3.097589	0.0020
	0.500	-1.639286	0.536175	-3.057371	0.0023
	0.600	-1.299492	0.556704	-2.334261	0.0197
	0.700	-1.495276	0.583305	-2.563454	0.0104
	0.800	-0.702727	0.701599	-1.001609	0.3166
	0.900	-0.590974	1.004344	-0.588417	0.5563
INTERNET	0.100	1.528190	0.611732	2.498136	0.0126
	0.200	1.083462	0.549811	1.970606	0.0489
	0.300	1.365395	0.505120	2.703110	0.0069
	0.400	1.548000	0.523438	2.957370	0.0031
	0.500	0.916071	0.580356	1.578465	0.1146
	0.600	0.865085	0.567164	1.525280	0.1273
	0.700	1.301350	0.594594	2.188635	0.0287
	0.800	1.711818	0.703185	2.434377	0.0150
	0.900	1.761075	1.258687	1.399136	0.1619
ALTACALIDAD	0.100	5.171379	0.598702	8.637652	0.0000
	0.200	5.534846	0.511183	10.82752	0.0000
	0.300	6.481842	0.476651	13.59871	0.0000
	0.400	7.256276	0.535842	13.54182	0.0000
	0.500	8.125357	0.589687	13.77909	0.0000
	0.600	8.982373	0.570707	15.73904	0.0000
	0.700	9.383497	0.614999	15.25775	0.0000
	0.800	10.64136	0.827080	12.86618	0.0000
	0.900	11.06787	0.952681	11.61760	0.0000

Se observa que los coeficientes asociados a las variables género y alta calidad son todos positivos y significativos e incrementan su magnitud a medida que aumenta el cuantil, es decir su efecto es mayor para la predicción de puntajes altos (Figura 4).

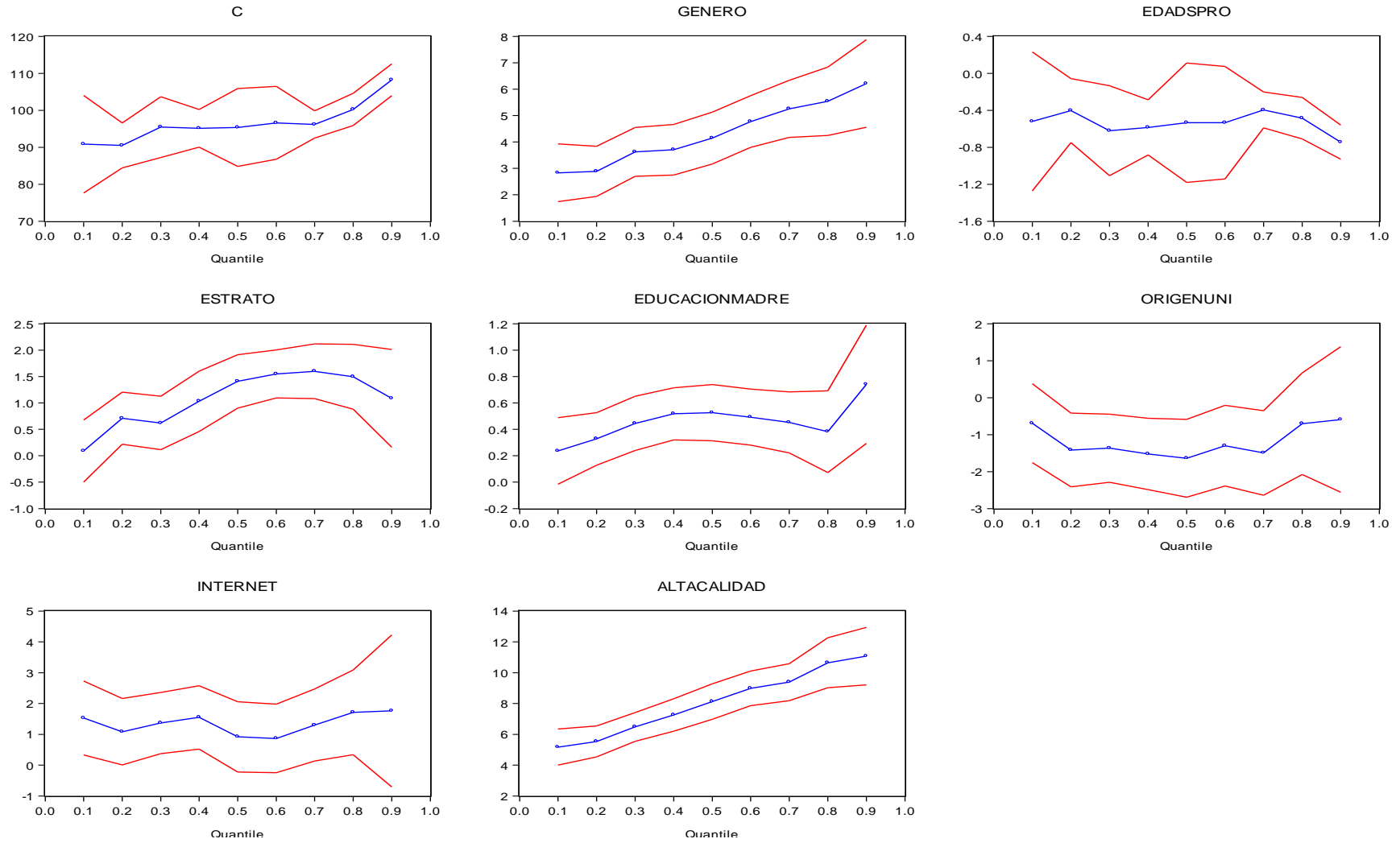
El coeficiente asociado al nivel educativo de la madre es positivo y significativo para todos los cuantiles excepto para el cuantil 0.1, incrementa su efecto para cuantiles superiores del puntaje (figura 4).

Los coeficientes asociados al estrato y al origen de la universidad no son significativos en las colas de la distribución. En los cuantiles intermedios se observa un efecto positivo para el estrato y su magnitud toma forma de u invertida, es decir incrementa su efecto en la predicción de puntajes intermedios de la distribución. En el caso de los coeficientes asociados al origen de la universidad son todos negativos y similares a lo largo de la distribución del puntaje, asocia mejores resultados a estudiantes del sector oficial; sin embargo, su efecto pronosticado es de tan solo 1 punto. El efecto asociado a la edad a la hora de tomar la prueba SABER PRO es negativo, similar al encontrado para la edad en la SABER 11, sin embargo deja de ser significativo en el cuantil 0.1, 0.5 y 0.6, su efecto en magnitud es similar a lo largo de la distribución.

El coeficiente asociado a la variable internet presenta signo positivo a lo largo de la distribución, sin embargo solo es significativo en los cuantiles 0.3 y 0.4.

Figura 4. Tendencia de los coeficientes estimados a través de regresión cuantílica, con variables de SABER PRO.

Quantile Process Estimates (95% CI)



Como era de esperar las variables estrato, valor de matrícula e ingreso familiar están correlacionadas, no se deben utilizar en una misma estimación. En la regresión de la tabla 10, se sustituyó la variable valor de la matrícula por estrato. Los coeficientes asociados a esta nueva variable son todos positivos y significativos, independientemente del cuantil, como se observa en la tabla 11. Los coeficientes asociados a las demás variables son similares a los encontrados en la estimación con estrato.

Tabla 11. Regresión cuantílica con variables de SABER PRO sustituyendo valor de la matrícula por estrato.

Quantile Process Estimates
Equation: SABERPRO
Specification: PUNTAJE C GENERO EDADSPRO VMATRICULA EDUCACIONMADRE ORIGENUNI INTERNET ALTACALIDAD

	Quantile	Coefficient	Std. Error	t-Statistic	Prob.
VMATRICULA	0.100	1.987143	0.388790	5.111098	0.0000
	0.200	2.319434	0.352111	6.587224	0.0000
	0.300	2.453729	0.344354	7.125600	0.0000
	0.400	2.805000	0.328127	8.548515	0.0000
	0.500	2.759524	0.348468	7.919027	0.0000
	0.600	2.908425	0.355372	8.184172	0.0000
	0.700	3.114667	0.408454	7.625509	0.0000
	0.800	3.072955	0.488585	6.289494	0.0000
	0.900	2.588571	0.746316	3.468467	0.0005

Al sustituir la variable estrato por ingreso familiar, se encuentran coeficientes positivos y significativos, independientemente del cuantil tabla 12. Los coeficientes asociados a las demás variables son similares a los encontrados en la estimación con estrato.

Tabla 12. Regresión cuantílica con variables de SABER PRO sustituyendo ingreso familiar por estrato.

Quantile Process Estimates
Equation: SABERPRO
Specification: PUNTAJE C GENERO EDADSPRO INGRESOFAM EDUCACIONMADRE ORIGENUNI INTERNET ALTACALIDAD

	Quantile	Coefficient	Std. Error	t-Statistic	Prob.
INGRESOFAM	0.100	1.137887	0.220172	5.168178	0.0000
	0.200	1.409199	0.196309	7.178460	0.0000
	0.300	1.577500	0.193439	8.155014	0.0000
	0.400	1.726111	0.194777	8.861972	0.0000
	0.500	1.802381	0.180801	9.968846	0.0000
	0.600	1.880526	0.182962	10.27825	0.0000
	0.700	2.161538	0.213308	10.13341	0.0000
	0.800	2.175063	0.250254	8.691413	0.0000
	0.900	1.723529	0.435593	3.956740	0.0001

Con el fin de analizar el efecto sobre la estimación del puntaje de la forma de pago de la matrícula (crédito, beca, recursos propios, recursos de los padres) se adicionan variables dummy a la estimación mostrada en la tabla 10, los resultados se encuentran en la tabla 13.

Tabla 13. Regresión cuantílica con variables de SABER PRO, y dummies para la forma de pagó.

Quantile Process Estimates
Equation: SABERPROMATRICULA
Specification: PUNTAJE C GENERO EDADSPRO ESTRATO EDUCACIONMADRE
ORIGENUNI INTERNET ALTACALIDAD MATRICULABECA
MATRICULACRED
MATRICULAEST MATRICULAPAD

	Quantile	Coefficient	Std. Error	t-Statistic	Prob.
MATRICULABECA	0.100	1.947333	0.945793	2.058943	0.0396
	0.200	1.632603	0.888261	1.837976	0.0662
	0.300	1.942013	0.898889	2.160459	0.0308
	0.400	2.030000	1.045510	1.941637	0.0523
	0.500	3.158889	0.938693	3.365199	0.0008
	0.600	3.087273	0.966801	3.193286	0.0014
	0.700	5.123208	1.557999	3.288326	0.0010
	0.800	4.925714	1.302997	3.780295	0.0002
	0.900	7.623617	1.785622	4.269446	0.0000
MATRICULACRED	0.100	-0.680000	0.837974	-0.811481	0.4172
	0.200	-0.165333	0.719873	-0.229670	0.8184
	0.300	-0.160658	0.669145	-0.240094	0.8103
	0.400	-0.599423	0.743854	-0.805834	0.4204
	0.500	-0.186667	0.958340	-0.194781	0.8456
	0.600	0.925455	0.735205	1.258771	0.2082
	0.700	0.617744	0.712853	0.866581	0.3863
	0.800	-0.366286	0.762914	-0.480114	0.6312
	0.900	-1.080745	1.048327	-1.030923	0.3027
MATRICULAEST	0.100	0.018667	0.899510	0.020752	0.9834
	0.200	-0.129143	0.663078	-0.194763	0.8456
	0.300	-1.028067	0.628583	-1.635531	0.1021
	0.400	-1.340962	0.708436	-1.892849	0.0585
	0.500	-0.542222	0.795666	-0.681470	0.4956
	0.600	-0.604545	0.743175	-0.813464	0.4160
	0.700	-1.379599	0.671427	-2.054726	0.0400
	0.800	-1.997143	0.764864	-2.611109	0.0091
	0.900	-2.955745	1.115356	-2.650046	0.0081
MATRICULAPAD	0.100	0.814667	1.023237	0.796166	0.4260
	0.200	1.150159	0.750060	1.533423	0.1253
	0.300	0.779584	0.687235	1.134378	0.2568
	0.400	0.303269	0.747520	0.405700	0.6850
	0.500	1.018889	0.872281	1.168074	0.2429
	0.600	1.322727	0.744866	1.775793	0.0759
	0.700	1.277093	0.721362	1.770391	0.0768
	0.800	0.241429	0.797243	0.302829	0.7620
	0.900	-0.192340	1.051599	-0.182903	0.8549

Los coeficientes asociados al pago de matrícula a través de beca resultan significativos solo para los cuantiles superiores y de signo positivo con tendencia a incrementar su valor a medida que se predice con respecto a cuantiles superiores. En el cuantil 0.9 por ejemplo, se predice un aumento en el puntaje SABER PRO de algo más de 7 puntos para estudiantes que poseen beca para el pago de su matrícula. En cuanto a los coeficientes asociados al pago de la matrícula con recursos del estudiante son significativos solo para los cuantiles 0.8 y 0.9, presenta signo negativo lo cual presume puntajes mas bajos para estudiantes en esta condición.

Los coeficientes asociados al pago a través de crédito universitario y a través de recursos de los padres no son significativos para ningún a cuantil en la distribución del puntaje.

Conclusiones generales

En la predicción del puntaje en el prueba SABER PRO a través de algunas variables observables en el estudiante antes de iniciar sus estudios universitarios a partir de la información disponible de la prueba SABER 11, se encuentra un efecto positivo y significativo para las variables género, matemáticas, historia, lenguaje, idioma y química, tanto por mínimos cuadrados como por regresión cuantílica. A través de esta segunda metodología se observa que el efecto es más relevante en la predicción de los puntajes en los cuantiles superiores (puntajes más altos). El caso de biología es diferente, su efecto decrece en magnitud a medida que aumenta el cuantil, hasta dejar de ser significativo en los cuantiles superiores.

Los coeficientes asociados a la edad son significativos y de signo negativo para todos los cuantiles y para la media, lo cual predice mejores resultados a estudiantes de menor edad al momento de tomar la prueba SABER 11. Los coeficientes asociados a la prueba interdisciplinar son estadísticamente significativos sólo en los cuantiles intermedios de la distribución, aunque no son económicamente significativos porque en magnitud son cercanos a cero. Los coeficientes asociados a filosofía y física en su gran mayoría no son significativos.

Al considerar variables dummies por universidad se encuentran efectos positivos y significativos para las universidades analizadas (Andes, Javeriana, Nacional, externado, Rosario y del Norte), con diferencias considerables en magnitud (Tabla 7). Por ejemplo, la Universidad del Norte adiciona 9 puntos en la predicción del puntaje en la prueba SABER PRO, al utilizar regresión con respecto a la media.

Se estimó cada componente de la prueba SABER PRO para economía por separado como variable dependiente, se encuentra un efecto positivo y significativo asociado a las variables género, química, historia y lenguaje para todos los componentes. Por su parte los coeficientes asociados a filosofía, física e interdisciplinar no son significativos para la predicción de los puntajes de ningún componente, siendo este resultado consistente con los resultados obtenidos con respecto al puntaje general.

El efecto asociado a las variables edad, matemática e idiomas son significativos con respecto al puntaje general, algunos dejan de ser significativos por componente. La edad no es significativa para el componente de macroeconomía, el coeficiente asociado a matemáticas no es significativo para componente de pensamiento económico, el de idioma no lo es ni para pensamiento económico, ni para macroeconomía. El coeficiente asociado a biología es significativo solo para el componente de macroeconomía donde su signo es positivo.

Otras variables de interés como hacer parte de programas con certificados de alta calidad, aumentos en el nivel educativo de los padres, incrementos en el ingreso familiar, permiten predecir mejores resultados en la prueba SABER PRO e incrementan su efecto en la medida que se predicen puntajes en los cuantiles superiores.

Los coeficientes asociados al origen de la universidad son todos negativos y pequeños en magnitud, aumentos en el estrato y en el valor de la matrícula predicen mejores resultados en la prueba SABER PRO. El efecto asociado a la variable conexión a internet solo es significativo en los cuantiles intermedios.

Finalmente se analizó el efecto de la forma de pago de la matrícula. El pago a través de beca resulta significativo en la cola superior de la distribución, se predice un aumento en el puntaje SABER PRO de algo más de 7 puntos para estudiantes que poseen beca para el pago de su matrícula. El efecto asociado al pago de la matrícula con recursos del estudiante es significativo solo para los cuantiles 0.8 y 0.9, presenta signo negativo lo cual presume puntajes más bajos para estudiantes en esta condición. Los coeficientes asociados al pago a través de crédito universitario y a través de recursos de los padres no son significativos para ningún cuantil en la distribución del puntaje.

Las anteriores conclusiones se derivan exclusivamente de la muestra analizada. Resulta interesante validar los resultados con otras muestras para otros años.

Bibliografía.

Bai, C. & Chi, W. (2011). “Determinants of undergraduate GPAs in China: college entrance examination scores, high school achievement, and admission route,” se encuentra en: <http://mpra.ub.uni-muenchen.de/32797/>, School of Economics and Management, Tsinghua University, China.

Cohn, E., Cohn, S., Balch, D. y Bradley, J. (2004). “Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank,” *Economics of Education Review*, 23, pp. 577–586.

Eide, Eric, Mark H. Showalter (1998). “The Effect of School Quality on Student Performance: A Quantile Regression Approach” *Economics Letters*, 58 (3), pp 345-350.

Geiser, S. with R. Studley. (2002). “UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California.” *Educational Assessment*, vol. 8, no. 1, pp. 1-26.

ICFES. (2011) “Guía de acceso a bases de datos” Versión septiembre de 2011, disponible en: <http://www.icfes.gov.co/investigacion/acceso-a-bases-de-datos>.

Koenker, R. & Bassett, G. W. (1978), “Regression Quantiles”, *Econometrica* 46, pp. 33–50.

Koenker, R. (2005), “Quantile Regression”, *Econometric Society Monographs*, Cambridge University Press.

López, H. Mora, H. (2006), “Cálculo de la regresión cuantílica por medio del método ACCPM”, *Revista Colombiana de Estadística*, Volumen 30 No. 1. pp. 53 a 68. Junio 2007.

Montenegro, A. (2005). “Los ECAES de Economía”, *Documentos de Economía*, 20, Universidad Javeriana.

Sarmiento, J., & Sandoval, L., “Análisis descriptivo de los resultados de los ECAES en economía (2004- 2006)”, *Investigación y Reflexión*, vol. XVI, núm. 2, diciembre, 2008, pp. 79-104, Universidad Militar Nueva Granada, Colombia.

Sosa Escudero, Walter, Paula Inés Giovagnoli y Alberto Porto (2009). “The effects of individual characteristics on the distribution of college performance”. *Económica*, Vol. LV: 99-130.

Rosefsky, A. & Saavedra, J. (2011). “Do colleges cultivate critical thinking, problem solving, writing and interpersonal skills?”, *Economics of Education Review*, 30, pp. 1516– 1526.

Valens, M. (2007). “Calidad de la Educación Superior en Colombia: Un análisis multinivel con base en el ECAES de Economía 2004”, *Documentos de Trabajo*, Cidse, 99, Universidad del Valle.