

Predicción de Performance Académica:

Estudios de Postgrado

Marcos Gallacher

I. INTRODUCCION

La calidad de un programa de post-grado depende en forma marcada del calibre de los alumnos que son seleccionados para participar en el mismo. Buenos profesores, computadoras y pisos limpios sirven de poco si la "materia prima" básica con la que se cuenta (alumnos) tienen una motivación insuficiente o condiciones intelectuales mediocres.

Este trabajo analiza el problema de la predicción de performance académica de alumnos que cursan estudios de post-grado en una *business school*. La heterogeneidad de candidatos para este tipo de programas, unido al hecho de los mismos constituyen la culminación del ciclo educativo determinan que sean muchas las variables que pueden influenciar el éxito con el cual se llevan a cabo. Analizar performance en estudios de post-grado permite además hacer inferencias sobre "calidad" de todas las etapas previas del ciclo educativo, así como también de variables personales y familiares que potencialmente afectan desempeño académico.

En la siguiente sección se presenta un breve modelo teórico de performance individual, y se mencionan algunos antecedentes existentes en la bibliografía. En la tercer sección se describe la estimación estadística realizada, y en la cuarta las conclusiones que surgen del trabajo. Un apéndice resume estadísticas básicas de la muestra empleada para el análisis.

II. PERFORMANCE ACADEMICA

El output final de un programa de postgrado de negocios es el aumento de la capacidad de *management* de aquellos que participan en el. Resulta muy difícil (en la práctica casi imposible) estimar cambios en capacidad de management ya que ésto requeriría contar con una base de datos de individuos con y sin estudios de postgrado, que incluya no sólo salarios a lo largo del tiempo, sino también medidas de capital humano, talento innato y otras variables. Un objetivo mas modesto es analizar determinantes de *performance académica* (medida como calificaciones obtenidas en los cursos) bajo el supuesto de que ésta es un *input* que afecta en forma directa la performance en el mundo laboral.

El enfoque empleado aquí es el de la **función de producción educativa**, variantes del cual son frecuentes en la literatura.¹ Concretamente, un programa de post-grado resulta en la adquisición de un conjunto de conocimientos, los cuales pueden ser representados por un vector $\mathbf{Y} = \langle Y_1, \dots, Y_n \rangle$. Estos son tanto mayores cuanto mayor es el stock de capital humano que el individuo tiene al entrar al programa (\mathbf{K}), y cuanto mayor sea el tiempo (\mathbf{E}) que el individuo dedica al estudio. En forma resumida:

$$[1](Y_1, \dots, Y_n; K, E) = 0$$

donde $\partial y_i / \partial E > 0, \partial y_i / \partial K > 0, \partial^2 y_i / \partial E \partial K > 0$. Esta "función de producción multiproducto" presenta las siguientes particularidades:

1. La mayor obtención del producto i no necesariamente está acompañada de menor obtención de producto j . Razones de no-rivalidad entre productos incluyen (a) que un producto sea insumo para otro, (b) que los Y_i requieran \mathbf{E} de distintos momentos en el tiempo, es decir que la producción no es simultánea sino secuencial.
2. Los outputs $\langle Y_1, \dots, Y_n \rangle$ generados dependen del nivel de esfuerzo \mathbf{E} invertido. \mathbf{E} puede ser vista como una función de cálculo privado de beneficios y costos. Es decir, el individuo enfrenta una restricción presupuestaria del tipo $\mathbf{T} = \mathbf{J} + \mathbf{E} + \mathbf{L}$ donde \mathbf{T} representa el tiempo total disponible, y \mathbf{J} , \mathbf{E} y \mathbf{L} el tiempo dedicado, respectivamente, al trabajo, al estudio y al ocio. El alumno enfrenta un problema estándar de optimización:

$$[2] \max U = U(L, C_1(J), C_2(E))$$

s.a.

$$L + J + E = T$$

donde C_1 y C_2 representan, respectivamente, los niveles de consumo durante y luego de terminar el postgrado. *Ceteris paribus*, \mathbf{E} disminuirá cuanto mayor sea el valor del producto marginal del tiempo en la ocupación actual, $\partial C_1 / \partial J$, ocurriendo lo contrario cuanto mayor sea el impacto percibido del aprendizaje sobre los ingresos futuros ($\partial C_2 / \partial E$). Es posible plantear la hipótesis de que \mathbf{E} será una función inversa de la edad del individuo ya que C_2 es directamente proporcional a la vida útil que resta luego de finalizar el postgrado.

3. La magnitud de $\partial C_2 / \partial E$ depende no sólo de la vida útil (futura) de la persona, sino también del mercado en el cual ésta vende sus servicios. La tesis esbozada por T.W.Schultz hace ya

¹ Por ejemplo ver E.A.Hanushek (1987), "Educational Production Functions", en G. Psacharopoulos (editor) Economics of Education - Research and Studies.

tiempo centra atención en que los retornos al capital humano dependen del ritmo de generación de desequilibrios a los cuales está sujeta una economía.² Altos retornos a E pueden entonces obtenerse ya sea (i) permaneciendo en un sector sujeto a cambio tecnológico y/o de mercado, o (ii) cambiando el sector en el cual uno vende servicios personales, desde uno menos a otro más dinámico. Los programas de *master* contribuyen idealmente a estos dos objetivos: aumentar la productividad en el entorno en el cual el alumno se desempeña, y permitir una mayor movilidad entre entornos.³

4. La elección de E está afectada en forma marcada por la eficiencia con la cual el mercado "detecta" diferencias entre graduados promedio de aquellos que lideran su camada. En un mercado pequeño existe bajo costos de obtención de información sobre la calidad de candidatos que se ofrecen para puestos de trabajo -- *ceteris paribus* esperaríamos que E fuera mayor en un programa de *master* en economía que en uno de dirección de empresas ya que en el primero las consecuencias de invertir un bajo E son conocidas por futuros empleadores en forma más rápida que en el segundo. Los alumnos de economía "dependen" más de sus profesores que los de dirección de empresas, ya que muchos de ellos tienen interés en proseguir una carrera académica, o en completar estudios de post-grado, ambas opciones siendo favorecidas por buenas recomendaciones de ex-profesores.

5. La variable K debería ser representada como un vector, que incluye características como habilidad numérica, astucia para rendir exámenes, conocimientos específicos adquiridos durante el ciclo secundario y universitario y experiencia de trabajo. A diferencia de E , esta variable no resulta de decisiones durante sino de aquellas que han sido tomadas por el educando y sus padres previo a éste. Podemos suponer que K queda representada mediante una función:

$$[3]K = f(P, I)$$

donde P denota características personales del individuo (una amalgama de genética, entorno familiar y otras), mientras que I representa inversiones en capital humano realizadas antes de incorporarse al ciclo de post-grado.⁴ Al ser K una variable monotónicamente asociada con la cantidad/calidad de capital incorporado al individuo, $\delta K / \delta I > 0$. El problema, sin embargo, es estimar la *magnitud* de la relación existente entre esfuerzo realizado en la adolescencia/juventud con los resultados cosechados algunos años más tarde.

6. El supuesto de que mayor intensidad de K incrementa el producto marginal de E , es decir $\partial^2 y_i / \partial E \partial K > 0$, sugiere que los individuos dotados de mayor capital humano tenderán a elegir mayores

² Ver T.W.Schultz(1979), *Investing in People - The Economics of Population Quality*. University of California Press, Berkeley.

³ En muchos casos la principal motivación de encarar un MBA es emigrar de un entorno donde la productividad marginal privada (y por lo tanto la remuneración) es baja.

⁴ Esta distinción es en gran medida artificial ya que I depende de P . Es decir, existe "auto-selección" en lo referido a educación.

niveles de esfuerzo que aquellos para los cuales el nivel de K es menor. La no-observabilidad de E complica entonces la estimación econométrica del impacto de K sobre Y .

III. APLICACION EMPIRICA

III.1 Datos y Estimación

El esbozo anterior constituye la base para la estimación econométrica que presento a continuación. Como se verá, ésta sólo analiza la relación existente entre componentes del vector K y proxies de los outputs del proceso educativo (representados por calificaciones obtenidas en los cursos). Ausencia de datos y complejidad econométrica impiden desarrollar un modelo que explique la decisión de particionar el tiempo entre E por un lado, y $L + J$ por otro.

Supongo que: (i) la variable P puede ser aproximada por la edad, el sexo y la opinión que los terceros tienen del candidato, (ii) K depende del tipo de educación obtenida en el ciclo secundario y en el universitario y (iii) existen ciertas habilidades cognitivas - que si bien no ortogonales con las variables mencionadas anteriormente - son lo suficientemente distintivas como para ser consideradas en forma separada. En particular, estas variables están altamente correlacionadas con resultados obtenidos en "tests" de admisión que normalmente son tomados por aspirantes a cursos de post-grado.

El modelo estadístico fué ajustado utilizando datos de uno de los programas de post-grado de la **Universidad del CEMA**. Elegí el programa **MADE**, por ser el mas numeroso (la base de datos utilizada aquí contiene registros de 178 estudiantes). Además, a diferencia de lo que ocurre en los otros *masters*, este programa incorpora una amplia gama de profesionales: en especial ingenieros y contadores, pero también licenciados en administración de empresas, psicólogos, abogados y aún licenciados en física o geología. Es decir, la base de datos resultante brinda buena dispersión en el *background* de los candidatos, lo cual facilita de parámetros de las variables. El modelo estadístico ajustado es formulado como:

$$[3] Y_i = b_1 + b_2 EXING + b_3 SEXO + b_4 EDAD + b_5 REFER + b_6 COLPU + b_7 COLPUS + b_8 UNIPR + b_9 TECNO + b_{10} CSECO + e$$

Siendo la definición de las variables la siguiente:

Variable Dependiente:

Y_i = el i -ésimo output del proceso educativo (para $i = 1, \dots, 5$ corresponde a conocimiento general (promedio de todas

las áreas), economía, métodos cuantitativos, finanzas y management).⁵

Características Básicas del Individuo (Variable P):

SEXO = variable binaria (varones = 0)
EDAD = edad al momento de inscripción en el programa
REFER = índice de referencias obtenidas ($6 \leq \text{REFER} \leq 36$)

Stock de Capital Humano Incorporado al Alumno (Variable K):

(a) General:

EXING = nota examen de ingreso ($0 \leq \text{EXING} \leq 50$)

(b) Educación secundaria:

COLPU = variable binaria (colegio público no selectivo = 1)
COLPUS = variable binaria (colegio público selectivo = 1)

(c) Educación Universitaria:

UNIPR = variable binaria (universidad privada = 1)
TECNO = variable binaria (ingenierías = 1)
CSECO = variable binaria (ciencias económicas = 1)

La interpretación de las variables anteriores no requiere mayor explicación. Las variables **COLPU** y **COLPUS** representan, respectivamente, educación secundaria en colegio público no selectivo y colegio público selectivo, estando éstos representados por los colegios Carlos Pellegrini y Nacional Buenos Aires. La variable **UNIPR** toma valor de 1 para egresados de universidades privadas. Al respecto, vale destacar que las universidades privadas de las cuales provienen los alumnos no incluyen aquellas de "segunda generación", relativamente selectivas como lo son la Universidad de San Andrés y la Di Tella (ni tampoco de la **Universidad del CEMA**, que todavía no tiene egresados de sus licenciaturas). El en caso de **REFER** se construyó un índice utilizando la matriz **3x7** que registra opiniones de cada uno de los 2 referees provistos por el alumno.⁶ El índice de

⁵

Materias incluídas en cada output:

Economía: Análisis Económico, Macroeconomía, Teoría de la Organización Empresarial.

Métodos Cuantitativos: idem

Finanzas: Finanzas de Empresa y Mercado de Capitales

Management: Info Financiera, Marketing I y II, Producción, Recursos Humanos,

Dirección estratégica, Productividad y Calidad.

⁶ Las 3 hileras de esta matriz requieren información sobre "capacidad académica", "capacidad profesional" y "motivación". Las primeras 6 columnas clasifican al alumno en categorías desde "brillante" a "regular". La séptima columna permite que el referee conteste "sin fundamento para opinar".

referencias se calculó, para cada uno de los referees como:

$$[4]REFER = \sum_{i=1}^3 \sum_{j=1}^6 a_{ij} + \sum_{i=1}^3 a_{i7}$$

donde para $j < 7$ $a_{ij} = j$ si el casillero contiene un tilde y $a_{ij} = 0$ en caso contrario. La séptima columna de la matriz corresponde a la respuesta "sin fundamentos para opinar", la convención empleada aquí fué que para $j = 7$, $a_{ij} = 2$ si el casillero fué llenado por el opinante y $a_{ij} = 0$ en caso contrario. Es decir, cuando el referee se "reserva la opinión" asigné arbitrariamente una puntuación "promedio" (en la escala numérica igual a 2).

Por último, para el caso de que el alumno hubiera incurrido en aplazos para alguna materia, se calculó una "nota corregida" como (nota final)/2, lo cual equivale a considerar como "0" a la nota del aplazo. Este procedimiento se siguió inclusive para situaciones donde existía mas de un aplazo, lo cual puede implicar un cierto sesgo hacia arriba en las notas promedio de los alumnos muy malos.

La base de datos original contiene 178 observaciones. Se descartaron todas aquellas donde faltaba información sobre alguna variable independiente. Además, se descartaron las observaciones correspondientes a alumnos que tenían mas de dos materias sin aprobar (a la fecha de Marzo de 1998). Las materias no aprobadas de alumnos incluídos en la muestra fueron calificadas con nota de 0.

III.2 Resultados

Los resultados de estimación figuran en el **Cuadro 1**. Se presenta un modelo para cada uno de los *outputs* considerados, así como para el *output global* representado por la calificación promedio obtenida en todos los cursos tomados. Las siguientes inferencias pueden hacerse a partir de los resultados logrados:

1. La performance académica resulta de difícil predicción. En efecto, para todos los modelos $R^2 \leq 0.31$. Lo anterior sugiere una de dos posibilidades: (i) el modelo excluye variables explicativas de importancia o (ii) existe cierto grado de aleatoriedad en las calificaciones. De ser cierto (i) los coeficientes de las variables incluídas sobre (sub) estimarán a los respectivos parámetros según exista correlación positiva (negativa) entre las variables excluídas y las incluídas.

2. Pese a lo anterior, los valores t resultan significativos ($p=0.10$) para un número interesante de variables. En particular:

(a) **EXING** (nota en el exámen de ingreso) presenta valores t que en todos los casos son superiores a 2.8, significativa a $p = 0.05$.

(b) **En general, las mujeres tienen mejor desempeño académico que los varones.** Lo anterior se observa a

través del promedio general, y las dimensiones economía y métodos cuantitativos. Para finanzas y management (¿materias más aplicadas?) las diferencias no son significativas

(c) Los alumnos que tienen un background en ingeniería y ciencias económicas tienen - para el output Y_1 , mejor desempeño que los provenientes de otras carreras universitarias. Los alumnos con un background en ingeniería son además mejores en métodos cuantitativos y economía.

(d) Los alumnos provenientes de carreras de ciencias económicas no son mejores en economía que los de otras especialidades, sin embargo sí lo son cuando el output considerado es Y_5 (management).

(e) Los coeficientes de la variable UNIPR sugieren que los alumnos de las universidades públicas son mejores que los provenientes de las privadas. Resulta interesante observar que para el output Y_5 (management) la diferencia entre ambos grupos no es significativa, es decir los egresados de las universidades privadas "trepan posiciones" en forma considerable en esta dimensión de desempeño.

(f) Los egresados de colegios públicos y los de colegios públicos selectivos no muestran performance diferencial con respecto a los de colegios privados.

(g) La variable EDAD resulta significativa sólo en el modelo que toma Y_5 (management) como variable dependiente. El signo del coeficiente, sin embargo, es positivo, y no negativo como se planteó en la hipótesis original. Lo anterior sugiere que en materias "prácticas" como las de *management* los profesionales de más edad tiene cierta ventaja sobre los más jóvenes.

(h) Las referencias carecen de poder predictivo. Este hecho posiblemente se deba a que el índice de referencias está elaborado en base a opiniones de individuos que no necesariamente coinciden en cuanto al criterio a adoptar para categorizar al individuo. Alguien podría opinar, asimismo, que la falta de asociación entre referencias y performance académica se debe a la falta de asociación entre cualidades necesarias para destacarse en "la calle" (mundo laboral) y el aula. Una afirmación tan temeraria, sin embargo, requeriría evidencia mucho más firme que la presentada aquí.

Dos comentarios adicionales en relación a los párrafos anteriores. En primer lugar, la falta de asociación entre tipo de educación secundaria y performance en estudios de post-grado puede deberse al hecho de que la educación universitaria de grado "empareja" - al menos parcialmente - diferencias en conocimiento y habilidad intelectual que se observan a la salida del secundario. Es decir, un egresado de un colegio público selectivo, y uno de un colegio mas modesto tendrían, luego de terminar el ciclo universitario, capacidades mas similares que las que tenían cinco o seis años antes. Este hallazgo resulta de interés, dado el actual clamor por el deterioro existente en enseñanza de grado a nivel secundario.⁷

Es también posible que la muestra de alumnos de colegios públicos no selectivos contenga una proporción relativamente mayor de alumnos "cabeza de promoción" que la de los colegios públicos selectivos. La **Figura 1** muestra las consecuencias que esto tendría sobre los parámetros estimados. La función O-A representa la respuesta - en performance académica de post-grado - de individuos de potencial mayor al promedio, mientras que la O-P la de individuos "promedio". Si la muestra que integra el banco de datos analizado contiene para los colegios públicos no selectivos alumnos del tipo "A", y para los de colegios selectivos alumnos "P", la función estimada será E - E', en lugar de O-A o O-B, que correspondería de contarse con información sobre potencial individual. Este es un típico problema de exclusión de variables.

IV. IMPLICANCIAS

La performance en un medio relativamente "artificial" como lo son las aulas debería ser en principio de mas facil modelización que aquella en un el "mundo real", donde los *outputs* relevantes son muchos e invariablemente de mas difícil cuantificación. Es decir, si resulta difícil predecir **adquisición de conocimiento** con mas razón va a ser difícil predecir **potencial gerencial**, que depende no sólo de conocimiento sino de perseverancia, astucia, capacidad de negociación, de liderazgo y de otros factores. La predicción, sin embargo, debe ser hecha: ya sea para racionar vacantes limitadas, o aún cuando éstas sean abundantes, para impedir ingreso de individuos de bajo potencial a un programa que quiere mantener su prestigio.

⁷ Ver, por ejemplo, Guillermo Jaim Etcheverry, "La carpa de los padres". La Nación, 17 de febrero de 1998.

Este estudio sugiere que existe un vector de variables asociadas con performance: en especial calificaciones en exámenes de ingreso, pero también sexo, tipo de universidad y especialización profesional de grado. También encuentro que el impacto de especialización de grado y tipo de universidad depende de si el *output* considerado es un área tradicional de conocimiento científico (en especial economía y estadística), o un área "soft" que requiere menor capacidad de abstracción y conocimiento formal. En particular, especialización profesional en ingeniería, como así también estudios en la universidad pública son factores asociados a egresados superiores en cuanto a capacidad analítica, pero no necesariamente en cuanto a *management*. Como fué advertido las universidades privadas consideradas no incluyen aquellas de "segunda generación" que típicamente tienen planes de estudio con mayor contenido teórico y cuantitativo.

Posiblemente existan retornos asociados a mejorar entrevistas de admisión a una *business-school*. Por ejemplo, la elaboración de un "índice subjetivo de potencial gerencial" del entrevistado podría contribuir a una selección donde no sólo se toman en cuenta las habilidades lógico/cognitivas sino aquellas mas "soft" que muchas veces son críticas en el lugar de trabajo. Los candidatos podrían entonces rankearse por un índice que pondera tanto "intelecto" como "potencial de liderazgo". Sería interesante conocer experiencias que en este sentido tengan programas de MBA del exterior.

El hallazgo de ausencia de asociación entre tipo de colegio secundario y performance en estudios de post-grado resulta preocupante para aquellos que - como quien ésto escribe - deciden hoy la estrategia de educación que mas favorecerá a sus hijos y a su propio bolsillo. Pero esa es otra historia.

APENDICE

Se emplearon como datos los registros de los alumnos del programa **MADE** de la **Universidad del CEMA** que se graduaban en Mayo de 1998. De los 178 registros originales, se eliminaron 52 por estar la información de éstos incompleta. La estimación se realizó entonces con 126 observaciones. Las **Tablas A1** y **A2** transcriben, respectivamente, estadísticas básicas de **variables continuas** y de **variables** discretas del banco de datos empleado:

Tabla A1: Variables Continuas (n = 126)

Variable	Media	DS	Min	Max
Y₁	2.5	0.47	1.5	3.5
Y₂	2.4	0.68	0.8	4.0
Y₃	2.1	1.14	0.0	4.0
Y₄	2.2	0.90	0.5	4.0
Y₅	2.6	0.42	1.7	3.7
Edad	28	3.2	23	39
Ex.Ingreso	39	3.6	33	46
Referencias	28	3.4	20	36

Tabla A2: Variables Discretas (n = 126)
(número de observaciones por clase)

Variable	D = 0	D = 1
COLPUN	85	41
COLPUS	113	13

<i>UNIV</i>	69	57
<i>TECNO</i>	74	52
<i>ECO</i>	69	57

Los datos de la tabla **A1** no requieren mayor explicación. Con respecto a la tabla **A2** podemos ver que: (i) el 42 por ciento de los alumnos provienen de colegios públicos ($[41+13]/126$), (ii) de éstos 29 por ciento provienen de colegios públicos selectivos ($13/[41+13]$), (iii) 45 por ciento por ciento de universidades públicas ($57/126$), (iv) 41 por ciento de carreras tecnológicas ($52/126$) y (v) 45 por ciento por ciento de carreras vinculadas a las ciencias económicas ($57/126$).