

CAPITAL HUMANO Y CRECIMIENTO: NUEVAS SERIES DE ESCOLARIZACIÓN Y ALGUNOS RESULTADOS PARA LA OCDE.

.....
ÁNGEL DE LA FUENTE (*)

Instituto de Análisis Económico (CSIC)

UNA DE LAS CARACTERÍSTICAS DISTINTIVAS DE LAS «NUEVAS» TEORÍAS DEL CRECIMIENTO DESARROLLADAS EN LOS ÚLTIMOS QUINCE AÑOS HA SIDO EL ENRIQUECIMIENTO DEL CONCEPTO RELEVANTE DE CAPITAL. MIENTRAS QUE LOS

41

modelos neoclásicos tradicionales se centraban casi exclusivamente en la acumulación de capital físico (estructuras y maquinaria), las contribuciones más recientes han atribuido una importancia creciente a la acumulación de capital humano y conocimientos productivos, así como a la interacción entre estos dos factores intangibles (1).

La evidencia empírica, sin embargo, no siempre ha sido consistente con los nuevos modelos teóricos. En el caso del capital humano, en particular, algunos estudios recientes han generado resultados desalentadores. Las variables educativas

son con frecuencia no significativas o incluso entran con el signo «equivocado» en regresiones de crecimiento, especialmente cuando éstas se estiman utilizando especificaciones en diferencias o con técnicas de panel. La acumulación de resultados negativos en la literatura ha alimentado un creciente escepticismo sobre el papel de la escolarización en el proceso de crecimiento e incluso ha llevado a algunos autores (véase especialmente Pritchett, 1999) a considerar seriamente las posibles razones por las que la inversión educativa podría no contribuir al crecimiento de la productividad.

Una hipótesis alternativa, bastante extendida entre los investigadores dedicados al tema, es que estos resultados negativos podrían deberse, al menos en parte, a la mala calidad de los datos de escolarización que se han utilizado en los estudios empíricos sobre los determinantes del crecimiento. Este artículo resume algunos de los principales resultados de una serie de trabajos (De la Fuente y Doménech (D&D), 2000, 2001a, 2001b y 2002) en los que se aporta evidencia en favor de esta hipótesis.

El trabajo está organizado como sigue. En las secciones primera y segunda se repa-

res). Estas especificaciones se obtienen a partir de la función de producción, reemplazando los *stocks* de factores o sus tasas de crecimiento por aproximaciones convenientes que hacen uso de datos sobre tasas de inversión.

Existe un número importante de trabajos empíricos que han analizado la relación entre capital humano y crecimiento, con resultados contrapuestos, utilizando algunas de las especificaciones que acabo de esbozar (2). Mientras que en los primeros estudios realizados sobre el tema se obtuvieron resultados generalmente muy positivos, las conclusiones de un segundo grupo de estudios más recientes han sido bastante desalentadoras, no detectándose en muchos de ellos una correlación significativa entre la formación media de la población y el nivel de productividad (3).

La principal diferencia entre ambos grupos de trabajo tiene que ver con la utilización de técnicas econométricas que implícitamente asignan pesos diferentes a las variaciones observadas en las variables de interés a lo largo del tiempo y entre países. Mientras que en el primer grupo de estudios la estimación se realiza con datos de corte transversal (esto es, con una única observación por país que describe el comportamiento medio durante un período de varias décadas), en el segundo se utilizan varias observaciones por país tomadas sobre períodos más cortos y se emplean especificaciones en diferencias o técnicas de panel que básicamente eliminan la variación de corte transversal que existe en los datos antes de proceder a la estimación.

Aunque las técnicas de estimación utilizadas en los trabajos más recientes tienen la importante ventaja de que permiten controlar mejor por diferencias no observables entre países, también presentan algunos inconvenientes, entre los que destaca su mayor sensibilidad a la presencia de errores de medición en las variables; en particular porque éstos tienden a ser mayores en la dimensión temporal que en la transversal, pues tienden a cancelarse cuando se promedia sobre períodos largos. Esto sugiere que, como ya hemos anticipado al principio, una posible expli-

**RECUADRO 1
UN MODELO DESCRIPTIVO DE CRECIMIENTO CON CAPITAL HUMANO**

En este recuadro se desarrolla un sencillo modelo de crecimiento con capital humano. El modelo tiene dos componentes: una función de producción agregada y una relación de progreso técnico. Supondremos que la función de producción es del tipo Cobb-Douglas:

$$Y_{it} = A_{it} K_{it}^{\alpha_k} H_{it}^{\alpha_h} L_{it}^{\alpha_l} \quad [I]$$

donde Y_{it} denota el *output* agregado del país i en el período t , L_{it} es el nivel de empleo, K_{it} el *stock* de capital físico, H_{it} el *stock* medio de capital humano por trabajador, y A_{it} un índice de eficiencia técnica o PTF que resume el estado actual de la tecnología y recoge factores omitidos tales como la localización geográfica, el clima, las dotaciones de recursos naturales o la calidad de las instituciones. El coeficiente α_i (con $i = k, h, l$) mide la elasticidad del *output* con respecto a las dotaciones de los distintos factores productivos. Así, por ejemplo, un aumento del 1% en el *stock* de capital humano por trabajador aumentaría el *output* en un $\alpha_h\%$, manteniendo constantes las dotaciones de otros factores y el nivel de eficiencia técnica.

Bajo el supuesto habitual de que [I] presenta rendimientos constantes a escala en capital físico, trabajo y el *stock* total de capital humano, LH (esto es, que $\alpha_k + \alpha_l = 1$), podemos definir una función de producción *per cápita* que relaciona el *output* por trabajador ocupado con el nivel medio de formación y el *stock* de capital físico por trabajador. Sea $Q = Y/L$ del *output* por trabajador ocupado y $Z = K/L$ el *stock* de capital físico por trabajador. Dividiendo ambos lados de [I] por el empleo, L , tenemos:

$$Q = AZ^{\alpha_k} H^{\alpha_h} \quad [II]$$

La función de progreso técnico describe los determinantes de la tasa de crecimiento de la productividad total de los factores. Supondremos que el nivel de PTF de un país i viene dado por

$$A_{it} = B_t X_{it} \quad [III]$$

donde B_t denota la «frontera tecnológica» mundial (esto es, el nivel máximo de eficiencia en la producción, dado el estado actual de conocimientos científicos y tecnológicos) y $X_{it} = A_{it}/B_t$, la «brecha tecnológica» entre el país i y la frontera. Supondremos que B_t crece a una tasa constante y exógena, g , y que la tasa de crecimiento de X_{it} viene dada por:

$$\Delta x_{it} = \gamma_{io} - \lambda x_{it} + \gamma H_{it} \quad [IV]$$

donde x_{it} es el logaritmo de X_{it} y γ_{io} un efecto fijo de país que ayuda a controlar por variables omitidas, tales como la inversión en I+D. Obsérvese que esta especificación incorpora un efecto de difusión tecnológica. Si $\lambda > 0$, los países que están más cerca de la frontera tecnológica experimentarán, otras cosas iguales, menores tasas de crecimiento de la PTF. Esto hace que el nivel relativo de PTF de cada país tienda a estabilizarse con el paso del tiempo en torno a un valor de equilibrio a largo plazo que será una función creciente de su nivel educativo.

cación de los resultados adversos obtenidos en trabajos recientes tiene que ver con la mala calidad de los datos de capital humano que se han utilizado en la literatura de crecimiento. Como veremos en la sección siguiente, la mayor parte de las bases de datos internacionales sobre escolarización contienen una cantidad importante de ruido, debido a diversas inconsistencias de los da-

tos primarios utilizados para construirlas. La existencia de este ruido genera un sesgo a la baja en la estimación de los coeficientes que miden el impacto del capital humano (esto es, una tendencia a infraestimar sus valores) porque da lugar a una variabilidad espuria en el *stock* de capital humano que no se corresponde con cambios proporcionales en el nivel de productividad.

El primer paso en la construcción de estas series consistió en recopilar toda la información que pudimos encontrar sobre la distribución por nivel de estudios de la población adulta en los países de la muestra. Con este fin, hemos utilizado tanto publicaciones de organizaciones internacionales como fuentes nacionales (censos y encuestas, anuarios estadísticos y datos no publicados proporcionados por gobiernos nacionales y por la OCDE). Seguidamente hemos tratado de reconstruir un perfil razonable de la evolución del nivel medio de formación en cada país empleando toda la información disponible y un poco de sentido común. Para aquellos países en los que se dispone de series razonablemente completas, nos hemos apoyado sobre todo en fuentes nacionales. Para el resto, partimos de la estimación, aparentemente más fiable, de las existentes entre 1990 y 1995 (que en general se toma de la OCDE) y procedemos hacia atrás intentando evitar saltos poco razonables en las series que sólo pueden tener su origen en cambios en los criterios de clasificación.

La construcción de las series requirió en muchos casos juicios subjetivos de valor para elegir entre estimaciones alternativas cuando las distintas fuentes están en desacuerdo. En ocasiones, también hemos reinterpretado algunos datos tomados de compilaciones internacionales como referidos a niveles educativos ligeramente distintos de los indicados en las propias fuentes (7). Las observaciones no disponibles situadas entre datos censales se estiman por simple interpolación lineal. En el resto de los casos se recurre a extrapolaciones que utilizan, siempre que es posible, información sobre el nivel educativo de la población desagregada por grupos de edad en años cercanos al de interés.

La cantidad de información primaria disponible varía enormemente de un país a otro. El cuadro 1 muestra la fracción de las observaciones que se basan en información directa (procedente de censos o encuestas) sobre la fracción de la población con educación secundaria o superior, así como la fecha de la primera y de la última observación directa. El número potencial de observaciones por

	CUADRO 1 DISPONIBILIDAD DE DATOS PRIMARIOS					
	Educación secundaria			Educación superior		
	Fracción de obs. directas	Primera observ. directa	Última observ. directa	Fracción de obs. directas	Primera observ. directa	Última observ. directa
Alemania (Occ.)	11/24	1970	1995	17/24	1961	1995
Australia	11/24	1965	1997	11/24	1966	1997
Austria	11/24	1961	1995	7/24	1961	1995
Bélgica	13/24	1961	1995	12/24	1960	1995
Canadá	15/24	1961	1996	21/24	1960	1996
Dinamarca	9/24	1973	1994	12/24	1973	1994
España	12/21	1960	1991	12/21	1960	1991
Estados Unidos	24/24	1960	1995	24/24	1960	1995
Finlandia	16/24	1960	1995	21/24	1970	1995
Francia	12/21	1960	1989	12/21	1960	1990
Grecia	15/24	1961	1995	15/24	1961	1997
Holanda	12/24	1960	1995	12/24	1960	1995
Irlanda	15/24	1961	1998	11/24	1961	1998
Italia	15/24	1961	1999	5/8	1960	1998
Japón	8/21	1960	1990	12/21	1960	1990
Noruega	15/24	1960	1998	9/24	1960	1998
N. Zelanda	10/24	1965	1998	10/24	1965	1998
Portugal	12/21	1960	1991	8/21	1960	1991
Reino Unido	6/21	1960	1993	10/21	1960	1991
Suecia	9/24	1960	1995	9/24	1960	1995
Suiza	15/24	1960	1995	15/24	1960	1995

FUENTE: Elaboración propia.

país es generalmente 21 o 24 para cada nivel educativo, dependiendo de si la serie termina en 1990 o en 1995 (dos subniveles y un total para cada ciclo educativo por siete u ocho observaciones quinquenales). En el caso italiano, no parecen existir las carreras de ciclo corto, por lo que el número potencial de observaciones a nivel superior se reduce a ocho.

Como se puede ver en el cuadro 1, para la mayor parte de los países disponemos de suficiente información primaria para construir series que cubren razonablemente bien el conjunto del período muestral. Los casos más problemáticos se destacan en negrita. En el caso de Italia, el principal problema es que mucha de la información disponible se refiere a la población mayor de seis años. En Dinamarca y Alemania (a nivel secundario), la primera observación direc-

ta corresponde a 1970 o a un año posterior. En estos dos casos, hemos proyectado los niveles educativos hacia atrás hasta 1960 utilizando tasas de crecimiento tomadas de OCDE (1974), pero no estamos seguros de la fiabilidad de esta extrapolación.

Tras estimar el desglose de la población por niveles de formación, hemos calculado el número medio de años de escolarización teniendo en cuenta la duración teórica de los distintos ciclos educativos en cada país. Los resultados se resumen en el cuadro 2. La última fila del cuadro muestra el promedio (no ponderado por población) del número medio de años de formación en la muestra. Esta variable aumenta un 27,3% entre 1960 y 1990 como resultado de la importante mejora en los niveles educativos de las cohortes más jóvenes registrada en prácticamente todos los países.

RECUADRO 2
EL SESGO DE ATENUACIÓN Y EL RATIO DE FIABILIDAD

El origen del sesgo de atenuación es el siguiente. Supongamos que el nivel de productividad, Q, es una función lineal del stock de capital humano, H, de forma que

$$Q = bH + u \quad [I]$$

donde u es una perturbación aleatoria. Dada esta relación, variaciones en el stock de capital humano, H, inducirán cambios en el nivel de productividad, Q, y el examen de la magnitud relativa de las variaciones de ambas variables nos permitirá estimar el valor del coeficiente b. Ahora bien, si H está medido con error, de forma que lo que observamos en realidad no es H sino una proxy ruidosa, $P = H + \varepsilon$, donde ε es un error de medición aleatorio, parte de la aparente variación del stock de capital humano (en el tiempo o entre países) se deberá al error de medición (esto es, será ruido en vez de auténtica señal). Puesto que tales variaciones lógicamente no inducen respuesta alguna en Q, esta variable parecerá menos sensible al stock de capital humano de lo que es en realidad, lo que sesgará a la baja el valor estimado de b.

El tamaño del sesgo resultante dependerá inversamente del contenido informativo de la serie, medido por su ratio de fiabilidad, r. Este ratio se define como el cociente entre la señal y la suma de señal y ruido en los datos, esto es, como:

$$r = \frac{\text{var } H}{\text{var } P} = \frac{\text{var } H}{\text{var } H + \text{var } \varepsilon} \quad [II]$$

donde var H mide la señal contenida en la serie y var ε el ruido que la distorsiona (a).

Cuando existen varias proxies ruidosas de una misma variable de interés, sus respectivos ratios de fiabilidad pueden estimarse siguiendo el procedimiento propuesto por Krueger y Lindhal (2001). Sean $P_1 = H + \varepsilon_1$ y $P_2 = H + \varepsilon_2$ dos proxies ruidosas del stock de capital humano, H. Es fácil comprobar que, si los términos de error de las dos series ε_1 y ε_2 no están correlacionados entre sí, la covarianza entre P_1 y P_2 puede utilizarse para estimar la varianza de H, que es la única magnitud desconocida en la ecuación [II]. Por tanto, bajo este supuesto, r_1 puede estimarse como:

$$\hat{r}_1 = \frac{\text{cov}(P_1, P_2)}{\text{var } P_1} \quad [III]$$

que resulta ser la fórmula del estimador de mínimos cuadrados ordinarios del coeficiente de pendiente de una regresión de P_2 sobre P_1 . Por tanto, para estimar la fiabilidad de P_1 estimamos una regresión de la forma $P_2 = c + r_1 P_1$ (b).

Obsérvese, sin embargo, que si los errores de medición de las dos series están positivamente correlacionados ($E\varepsilon_1\varepsilon_2 > 0$) como cabría esperar en muchos casos, \hat{r}_1 sobreestimaré el ratio de fiabilidad, lo que hará que tendamos a infravalorar el sesgo generado por el error de medición.

En De la Fuente y Doménech (2002) se desarrolla una extensión de este procedimiento que permite construir un estimador de mínima varianza del ratio de fiabilidad cuando existen más de dos estimaciones con error de la misma variable subyacente, manteniendo la hipótesis de que los errores de medición de las distintas series no están correlacionados entre sí. Procediendo como se indica arriba, para cada serie de años de escolarización P_j construimos una serie de estimaciones «por pares» del ratio de fiabilidad, \hat{r}_{jk} , utilizando P_j para tratar de explicar otras series educativas, P_k ; esto es, para cada j estimamos un sistema de ecuaciones de la forma:

$$P_k = c_k + r_{jk} P_j + u_k \quad \text{para } k = 1, \dots, K \quad [IV]$$

donde k denota la serie «de referencia» y varía sobre la última versión disponible de todas las bases de datos diferentes de j. La fiabilidad de la base de datos de Barro y Lee (2000), por ejemplo, se estima utilizando esta serie como variable explicativa en un conjunto de regresiones en las que la variable de referencia (la variable a explicar o dependiente) son los años medios de escolarización estimados por Kyriacou (1991), NSD (1995), D&D (2001b) y Cohen y Soto (2001). Las demás versiones de la base de datos de Barro y Lee, sin embargo, no se utilizan como referencia porque la correlación de los errores de medición dentro de una familia determinada de estimaciones es, casi con seguridad, positiva y muy elevada, lo que tendería a inflar artificialmente la estimación de la tasa de fiabilidad.

Bajo el supuesto de que los errores de medición no están correlacionados entre familias de bases de datos (esto es, que $E\varepsilon_j\varepsilon_k = 0$ para $j \neq k$ cuando j y k pertenecen a familias diferentes) todas las estimaciones por pares de r_j construidas de esta forma serán consistentes, y también lo será cualquier promedio ponderado de ellas:

$$\bar{r}_j = \sum_k \omega_k \hat{r}_{jk} \quad \text{donde } \sum_k \omega_k = 1. \quad [V]$$

Para obtener el estimador más eficiente de r_j , elegimos las ponderaciones ω_k en [V] para minimizar la varianza de \bar{r}_j . El estimador resultante, al que denominaremos \hat{r}_j , puede implementarse imponiendo un coeficiente de pendiente común en todas las ecuaciones de [IV] y estimando el sistema resultante como un SUR (sistema de ecuaciones aparentemente no relacionadas) restringido. Por tanto, llamamos a \hat{r}_j el ratio de fiabilidad SUR.

(a) Obsérvese que el denominador del último término de [II] supone implícitamente que el error de medición, ε , no está correlacionado con H. (b) Intuitivamente, la regresión de P_2 sobre P_1 nos da una idea de lo bien que P_1 explica la variable sin errores, H. Esto es así porque el error de medición en la variable dependiente (P_2 en este caso) será absorbido por la perturbación de la regresión sin que esto genere un sesgo (al contrario de lo que sucede cuando lo que se mide con error es la variable independiente). Por tanto, es casi como si estuviésemos haciendo una regresión del verdadero H sobre P_1 .

do la ecuación incluye otros regresores además del capital humano (véase D&D, 2002).

El sesgo será menor cuando los datos se utilizan en niveles o en logaritmos, pero

puede ser muy grande en especificaciones de efectos fijos o en las que utilizan diferencias tomadas sobre períodos relativamente cortos. El ratio de fiabilidad es de sólo 0,254 para las series en tasas de crecimiento quinquenales y 0,090 para di-

ferencias en niveles calculadas con la misma frecuencia.

Nuestros resultados muestran que la importancia de los errores de medición varía significativamente de unas bases de

datos a otras, aunque el *ranking* exacto de cada una de ellas depende de la transformación de los datos que se utilice. Dos de las series más usadas en la literatura de crecimiento, las construidas por Kyriacou (1991) y Barro y Lee (varios años), tienen *ratios* de fiabilidad respetables cuando se utilizan en niveles, pero, tal como observan Krueger y Lindhal (2001), contienen muy poca información cuando se diferencian. Los intentos más recientes de aumentar el contenido informativo de las series de escolarización parecen haber tenido al menos un éxito parcial, si bien el sesgo de atenuación continúa siendo potencialmente elevado incluso en estos casos. Tomando como referencia el *ratio* promedio de fiabilidad de las series de Barro y Lee (1996) (0,206), la última revisión de estas series por los mismos autores ha aumentado su contenido de información en un 21%, mientras que las estimaciones de Cohen y Soto (2001) y De la Fuente y Doménech (2001) aumentan el *ratio* estimado de fiabilidad en un 162% y 207% respectivamente.

CALIDAD DE LAS SERIES EDUCATIVAS Y ESTIMACIONES DEL IMPACTO DEL CAPITAL HUMANO

Como hemos visto en la sección anterior, el valor esperado del sesgo de atenuación es una función decreciente del *ratio* de fiabilidad de la serie empleada en la estimación. Esto sugiere que el valor estimado del coeficiente del capital humano en una regresión de crecimiento debería aumentar con la calidad de los datos educativos. En De la Fuente y Doménech (D&D, 2002) se muestra que éste es efectivamente el caso.

En este trabajo estimamos varias especificaciones alternativas de una función de producción agregada utilizando las distintas series de escolarización analizadas en la sección anterior como indicadores alternativos del *stock* de capital humano y comprobamos que tanto el tamaño como la significatividad del coeficiente de esta variable aumentan de la manera esperada con el *ratio* de fiabili-

CUADRO 3
ESTIMACIONES SUR DEL RATIO DE FIABILIDAD
MUESTRA DE 21 PAÍSES DE LA OCDE

	S_{it}	S_{it}	ΔS_{it}	Δs_{it}	$S_{it}-s_{it}$	$\Delta s_{it}-\Delta s_{it}$	Promedio
D&D (2002)	0,754	0,775	0,337	0,769	0,917	0,246	0,633
C&S (2001)	0,806	0,912	0,330	0,467	0,547	0,185	0,541
D&D (2000)	0,720	0,761	0,100	0,550	0,818	0,074	0,504
Kyr (1991)	0,723	0,600	0,024	0,065	0,111	0,026	0,258
B&L (2000)	0,707	0,603	-0,018	0,045	0,178	-0,016	0,250
B&L (1996)	0,559	0,516	-0,017	0,039	0,146	-0,007	0,206
B&L (1993)	0,526	0,436	-0,019	0,029	0,121	-0,017	0,179
NSD (1995)	0,278	0,330	-0,021	0,066	0,095	-0,115	0,106
Promedio	0,634	0,617	0,090	0,254	0,367	0,047	0,335

Notas: La estimación se realiza con todas las series expresadas en desviaciones sobre su promedio muestral en cada período.

Clave: D&D = De la Fuente y Doménech; C&S = Cohen y Soto; Kyr = Kyriacou; B&L = Barro y Lee; NSD = Nehru *et al.*

FUENTE: Elaboración propia.

dad. Seguidamente, explotamos esta correlación para construir diversas «meta-estimaciones» del parámetro de interés que corrigen el sesgo generado por el error de medición (9).

RESULTADOS CON DISTINTAS SERIES EDUCATIVAS

Las ecuaciones que estimamos se derivan de una función de producción agregada del tipo Cobb-Douglas con rendimientos constantes a escala que incluye, además del *stock* de capital físico y el empleo, el nivel medio de formación de la mano de obra como uno de sus *inputs*. Esta ecuación se estima en niveles (con las variables expresadas en logaritmos), en niveles con efectos fijos de país y en diferencias.

También se estima una cuarta especificación en diferencias en la que se introducen efectos fijos y un proceso de difusión o *catch-up* tecnológico que permite que, otras cosas iguales, la tasa de crecimiento de la PTF sea directamente proporcional a la distancia tecnológica entre cada país y los Estados Unidos. En este caso, los efectos fijos de país capturan diferencias permanentes en niveles relativos de PTF que presumiblemente reflejarán diferencias en niveles de inversión en I+D y en otras variables omitidas (10).

Las distintas especificaciones se estiman con datos quinquenales para el período 1960-1990 con la muestra habitual de países de la OCDE, incluyendo en todos los casos efectos fijos temporales (*dummies* para los distintos subperíodos muestrales). Los valores del coeficiente que mide la elasticidad del *output* con respecto al nivel medio de formación α_s obtenidos con las distintas especificaciones y bases de datos se muestran en el cuadro 4. La última columna del cuadro muestra los valores medios de α_s y de su estadístico t para cada especificación, y las dos últimas filas los promedios de las mismas variables para cada base de datos.

El patrón de resultados que emerge según vamos cambiando la fuente de los datos de escolarización es consistente con nuestra hipótesis sobre la importancia del sesgo de medición. Para todas las bases de datos, el valor estimado de α_s es positivo y significativo para la especificación en niveles sin efectos fijos de país (primera fila del cuadro), pero el tamaño del coeficiente estimado y su nivel de significatividad aumentan apreciablemente cuando utilizamos las series con mayores *ratios* de fiabilidad (que son las que corresponden a las últimas columnas del cuadro 4). Las diferencias son aún más claras cuando la estimación se realiza con efectos fijos de país (segunda fila) o con los datos en tasas

de crecimiento (tercera y cuarta fila). Los resultados obtenidos con las series de Kyriacou, Barro y Lee y NSD expresadas en tasas de crecimiento son consistentes con los obtenidos por Kyriacou (1991), Benhabib y Spiegel (1994) y Pritchett (1999), quienes encuentran que el coeficiente del capital humano no es significativo (y a menudo tiene signo negativo) en funciones de producción estimadas en diferencias. Por otra parte, tanto nuestras series como las de Cohen y Soto generan valores positivos, relativamente grandes y precisos del coeficiente del capital humano en la mayor parte de las estimaciones.

CUADRO 4
ESTIMACIONES ALTERNATIVAS DEL COEFICIENTE DEL CAPITAL HUMANO (α_s)
EN DISTINTAS ESPECIFICACIONES DE LA FUNCIÓN DE PRODUCCIÓN CON DIFERENTES SERIES EDUCATIVAS

	NSD	KYR	B&L93	B&L96	B&L00	C&S	D&D00	D&D02	Media
Niveles	0,078 (2,02)	0,186 (2,18)	0,141 (4,49)	0,165 (4,82)	0,238 (6,19)	0,397 (7,98)	0,407 (7,76)	0,378 (6,92)	0,249 (5,30)
Efectos fijos	0,068 (0,76)	0,066 (1,86)	0,136 (3,30)	0,115 (1,80)	0,203 (3,74)	0,608 (4,49)	0,627 (3,99)	0,958 (6,51)	0,348 (3,31)
Diferencias	0,079 (0,70)	0,009 (0,15)	0,089 (2,52)	0,083 (1,47)	0,079 (1,28)	0,525 (2,57)	0,520 (2,17)	0,744 (3,10)	0,266 (1,75)
Catch-up	-0,206 (1,61)	0,014 (0,29)	0,056 (1,80)	-0,007 (0,11)	-0,019 (0,31)	0,573 (3,52)	0,587 (3,47)	0,540 (2,89)	0,192 (1,24)
Promedio	0,005 (0,47)	0,069 (1,12)	0,106 (3,03)	0,089 (2,00)	0,125 (2,73)	0,526 (4,64)	0,535 (4,35)	0,655 (4,86)	

Clave: Véanse las notas al cuadro 2.

FUENTE: Elaboración propia.

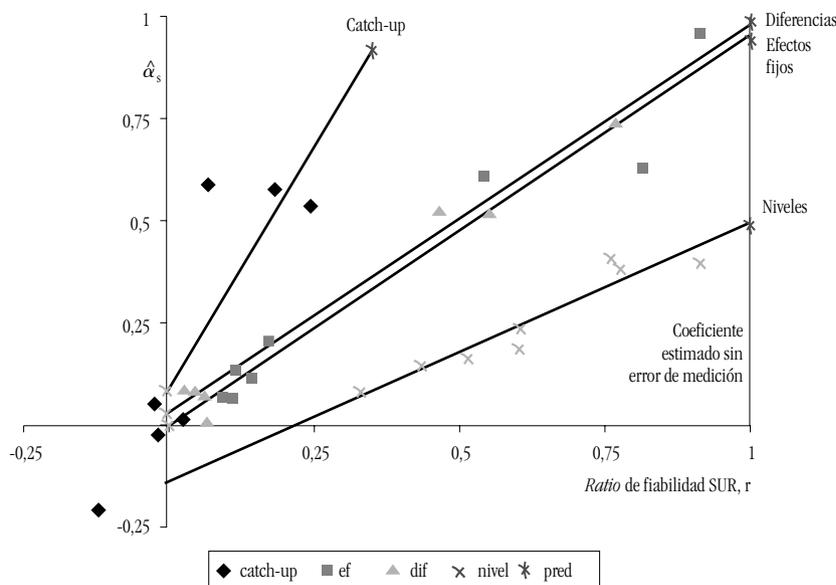
CÓMO CORREGIR EL SESGO DERIVADO DE LOS ERRORES DE MEDICIÓN

Los resultados resumidos en el cuadro 3 indican con claridad que el error de medición induce un sesgo importante a la baja en la estimación del coeficiente del capital humano en la función de producción, y que la mejora en la calidad de los datos reduce este sesgo y conduce a resultados generalmente más favorables a la hipótesis de que la inversión educativa contribuye de forma sustancial al crecimiento de la productividad.

El gráfico 1 ilustra y refuerza estas conclusiones. En él he dibujado las distintas estimaciones de α_s que aparecen en el cuadro 3 como función de los *ratios* de fiabilidad (*r*) correspondientes (tomados del cuadro 2), junto con la recta de regresión estimada para cada una de las especificaciones econométricas que hemos estimado en el apartado anterior. El gráfico muestra que existe una clara correlación positiva entre estas dos variables para cada especificación y sugiere que el valor correcto de α_s es al menos 0,50 (que es la predicción para la ecuación en niveles cuando $r = 1$).

Como sugiere el gráfico 1, es posible extrapolar la relación observada en las distintas bases de datos entre el *ratio* de fiabilidad y el coeficiente estimado del capital humano con el fin de estimar el valor de α_s que se obtendría en ausencia de errores de medición. De esta forma es posible construir meta-estimaciones

GRÁFICO 1
VALOR ESTIMADO DE α_s VS RATIO DE FIABILIDAD

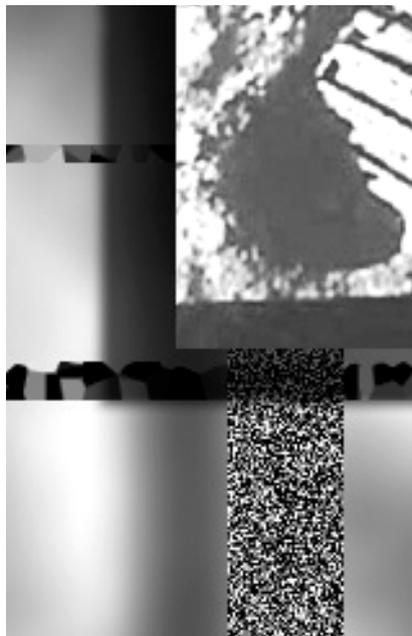


FUENTE: Estimación propia.

de este parámetro que estarán libres del sesgo de atenuación, si bien esto ha de hacerse con algo más de cuidado de lo que sugiere el gráfico (al menos cuando la ecuación de crecimiento incluye varios regresores). En De la Fuente y Doménech (2002) utilizamos un procedimiento de este tipo para obtener una serie de meta-estimaciones consistentes de α_s . Trabajando con las tres especificaciones lineales estimadas arriba (esto

es, con todas ellas, excepto el modelo de *catch-up*) y con tres supuestos diferentes sobre la naturaleza de los errores de medición (y en particular sobre su posible correlación entre bases de datos de escolarización y con el resto de las variables explicativas del modelo de crecimiento) obtenemos nueve estimaciones diferentes del valor de α_s que oscilan entre 0,587 y 2,606, con una media de 1,11.

Estos valores son significativamente más altos que los obtenidos en la literatura anterior. El menor de ellos es aproximadamente dos veces mayor que la estimación de 1/3 obtenida por Mankiw, Romer y Weil (1992). Esta estimación podría haberse tomado en su momento como un valor de consenso del parámetro de capital humano, pero generalmente se ha considerado demasiado optimista en la literatura más reciente, como consecuencia de los resultados negativos obtenidos en los últimos años. Nuestras estimaciones, por contra, apuntan a una cifra bastante más elevada y sugieren, por tanto, que la inversión en capital humano es un importante factor de crecimiento cuyo efecto sobre la productividad ha sido infravalorado en estudios anteriores debido a la mala calidad de los datos de escolarización.



CONCLUSIÓN

Las series de escolarización más utilizadas en la literatura de crecimiento contienen una cantidad importante de ruido que refleja diversas deficiencias de los datos primarios. Para intentar mejorar el contenido informativo de estos datos, hemos elaborado nuevas series educativas para una muestra de países de la OCDE utilizando información no explotada previamente y un procedimiento *ad hoc* que intenta minimizar el error generado por cambios en los criterios de clasificación. También hemos construido un indicador estadístico de la calidad de las distintas series educativas. Nuestros resultados confirman la importancia de los errores de medición y sugieren que nuestros datos constituyen una mejora significativa sobre los elaborados en trabajos previos.

El punto de partida de los estudios resumidos en este trabajo fue nuestra sospecha de que la deficiente calidad de los datos es una de las principales razones por las que diversos estudios recientes no han encontrado una relación significativa entre capital humano y crecimiento económico. Nuestros resultados, al igual que los de otros trabajos recientes de Krueger y Lindhal (2001) y Cohen y Soto (2001), apoyan esta hipótesis y su-

gieren que la contribución de la inversión educativa al crecimiento de la productividad es cuantitativamente importante. Al contrario que algunas bases de datos anteriores, nuestras series de escolarización generan resultados positivos utilizando diversas especificaciones econométricas.

Además, nuestro análisis de la variación de los resultados en función de la calidad de las series educativas revela que el valor estimado del coeficiente del capital humano en la función de producción aumenta en tamaño y en significatividad con el contenido informativo de los datos. Hemos extrapolado esta relación para estimar el valor de este parámetro que se obtendría en ausencia de errores de medición. El ejercicio sugiere que la elasticidad del *output* con respecto al *stock* de capital humano es superior a 0,50, esto es, al menos un 50% mayor que la más optimista de las estimaciones de referencia existentes en la literatura anterior.

(*) Este artículo resume los resultados de una serie de estudios que han sido realizados en colaboración con Rafael Doménech. Estos trabajos han sido financiados por el Fondo Europeo de Desarrollo Regional, la OCDE, la Fun-

dación Caixa Galicia y el Ministerio de Ciencia y Tecnología (a través de los proyectos SEC99-1189 y SEC2002-01612).

NOTAS

(1) Véanse, entre otros, Lucas (1988), Romer (1990), Azariadis y Drazen (1990), Grossman y Helpman (1991), Mankiw, Romer y Weil (1992) y Jones (1996).

(2) La sección tercera del apéndice de De la Fuente y Ciccone (2002) contiene un panorama detallado de esta literatura.

(3) Dentro del primer grupo cabe destacar los trabajos de Landau (1983), Baumol *et al.* (1989), Barro (1991) y Mankiw, Romer and Weil (1992); dentro del segundo, encontramos, entre otros, los trabajos de Kyriacou (1991), Knight *et al.* (1993), Benhabib y Spiegel (1994), Pritchett (1999), Islam (1995) y Caselli *et al.* (1996).

(4) Otras fuentes útiles son el *Anuario Demográfico* de la ONU, que ofrece datos de formación desagregados por grupos de edad y, en años recientes, el *Informe anual* de la OCDE sobre la educación en sus países miembros (*Education at a Glance*), que incluye gran cantidad de información sobre los *inputs* y los *outputs* del sistema educativo.

(5) Este estudio extiende y actualiza las series construidas en De la Fuente y Doménech (2000) para la misma muestra de países. Entre otras mejoras, la versión revisada de las series incorpora información no publicada suministrada por la OCDE y por las oficinas estadísticas nacionales de aproximadamente una docena de Estados miembros en respuesta a una petición de asistencia canalizada a través de la División de Estadísticas e Indicadores de la OCDE.

(6) Un trabajo que persigue esencialmente el mismo objetivo y utiliza una metodología similar es el de Cohen y Soto (2001). Estos autores construyen series educativas para una muestra mucho más amplia de países utilizando información censal y datos de encuestas tomados de la UNESCO, de la base de datos interna de la OCDE y de los *websites* de las agencias estadísticas nacionales.

(7) Nuestras estimaciones reflejan, por tanto, numerosas decisiones *ad hoc* que han tenido que basarse más en el sentido común que en información comprobada. Claramente, esta forma de proceder no es en principio la más deseable. Sin embargo, dada la falta de homogeneidad de la información de base, creemos que nuestro enfoque es preferible a los procedimientos aparentemente más objetivos y sistemáticos empleados por otros autores. Los resultados tienen un aspecto más razonable

que muchas de las series existentes, al menos en cuanto a su perfil temporal y, como veremos más adelante, obtienen buenos resultados en términos de un indicador estadístico de calidad.

(8) Esto es cierto siempre que se cumpla el supuesto de ausencia de correlación de los errores de medición de las distintas series entre sí y con H. Como se observa en el cuadro 3, algunas de nuestras estimaciones del *ratio* de fiabilidad están fuera de este intervalo, lo que indica alguna violación de esta hipótesis. En D&D (2002) construimos estimaciones alternativas de los *ratios* de fiabilidad que tienen en cuenta esta posibilidad y encontramos que las correcciones necesarias no cambian cualitativamente los resultados.

(9) Una meta-estimación es una estimación que no se obtiene directamente de los datos sino que se construye a partir de otras estimaciones primarias.

(10) Todas las especificaciones se obtienen a partir de la ecuación [III] del recuadro 1 utilizando los años medios de formación (S) como *proxy* para el *stock* de capital humano (H). En la última de estas especificaciones, se incorpora también una función de progreso técnico similar a la ecuación [IV] del mismo recuadro, excepto en que se omite el *stock* de capital humano. El modelo estimado, por tanto, no permite la existencia de efectos de tasa. Hemos intentado estimarlo incorporando estos efectos pero los resultados no son satisfactorios. Este problema es bastante habitual en la literatura. Véase la discusión en De la Fuente y Ciccone (2002) sobre las razones por las que puede resultar difícil separar los efectos de tasa de los de nivel.

BIBLIOGRAFÍA

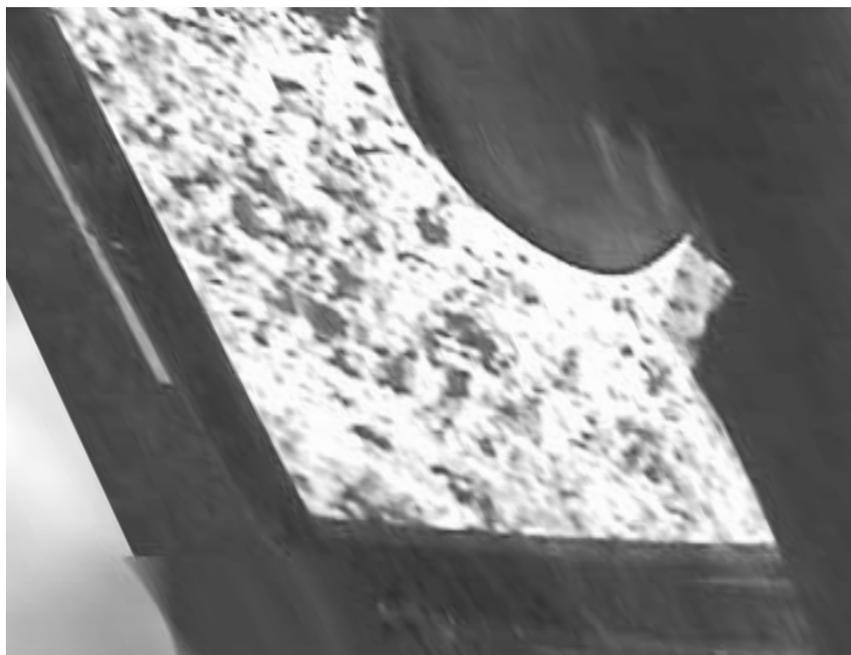
AZARIADIS, C. y Drazen, A. (1990): «Threshold Externalities in Economic Development», *Quarterly Journal of Economics*, mayo, pp. 501-26.

BARRO, R. (1991): «Economic Growth in a Cross Section of Countries», *Quarterly Journal of Economics* CVI (2), pp. 407-43.

BARRO, R. y LEE, J.-W. (1993): «International Comparisons of Educational Attainment», *Journal of Monetary Economics* 32, pp. 363-94.

BARRO, R. y LEE, J.-W. (1996): «International Measures of Schooling Years and Schooling Quality», *American Economic Review* 86 (2), *Papers and Proceedings*, pp. 218-23.

BARRO, R. y LEE, J.-W. (2000): «International data on educational attainment, updates and implications», NBER Working Paper n° 7911 (<http://www.nber.org/papers/w7911>).



BAUMOL, W., BATEY BLACKMAN, S. A. y WOLF, E. (1989): *Productivity and American Leadership: the Long View*, MIT Press.

BEHRMAN, J. y ROSENZWEIG, M. (1994): «Caveat emptor: cross-country data on education and the labor force», *Journal of Development Economics* 44, pp. 147-71.

BENHABIB, J. y SPIEGEL, M. (1994): «The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data», *Journal of Monetary Economics* 34, pp. 143-73.

CASELLI, F., ESQUIVEL, G. y LEFORT, F. (1996): «Reopening the convergence debate: a new look at cross-country growth empirics», *Journal of Economic Growth* 1 (3), pp. 363-89.

COHEN, D. y SOTO, M. (2001): «Growth and human capital: good data, good results», CEPR Discussion Paper n° 3025.

DE LA FUENTE, A. y CICCONE, A. (D&C, 2002): «Human capital and growth in a global and knowledge-based economy», Report for the European Commission, DG for Employment and Social Affairs (<http://ideas.repec.org/e/pde52.html#papers>).

DE LA FUENTE, A. y DOMÉNECH, R. (D&D, 2000): «Human capital in growth regressions: how much difference does data quality make?», OECD Economics Department Working Paper n° 262, París.

DE LA FUENTE, A. y DOMÉNECH, R. (D&D, 2001a): «Schooling data, technological diffusion and the neoclassical model», *American Economic Review* 91 (2), *Papers and Proceedings*, pp. 323-27.

DE LA FUENTE, A. y DOMÉNECH, R. (D&D, 2001b): «Educational attainment in the OECD, 1960-90», CEPR Discussion Paper n° 3390.

DE LA FUENTE, A. y DOMÉNECH, R. (D&D, 2002): «Human capital in growth regressions: how much difference does data quality make? An update and further results», CEPR Discussion Paper n° 3587 (<http://ideas.repec.org/e/pde52.html#papers>).

GROSSMAN, G. y HELPMAN, E. (1991): *Innovation and Growth in the Global Economy*, MIT Press, Cambridge, Mass.

ISLAM, N. (1995): «Growth empirics: a panel data approach», *Quarterly Journal of Economics* 110, pp. 1127-70.

JONES, C. (1996): «Human capital, ideas and economic growth», mimeo, Stanford University.

KNIGHT, M., LOAYZA, N. y VILLANUEVA, D. (1993): «Testing the neoclassical theory of economic growth: A panel data approach», *IMF Staff Papers* 40 (3), pp. 512-41.

KRUEGER, A. y LINDAHL, M. (2001): «Education for growth: why and for whom?», *Journal of Economic Literature* XXXIX, pp. 1101-36.

KYRIACOU, G. (1991): «Level and Growth Effects of Human Capital, A Cross-Country Study of the Convergence Hypothesis», mimeo, NYU.

LANDAU, D. (1983): «Government Expenditure and Economic Growth: a Cross-Country Study», *Southern Economic Journal*, enero, pp. 783-92.

LAU, L., JAMISON, D., y LOUAT, F. (1991): «Education and Productivity in Developing Countries: an Aggregate Production Function Approach», Report n° WPS 612, the World Bank.

- LAU, L., BHALLA, S. y LOUAT, F. (1991): «Human and physical capital stock in developing countries: construction of data and trends», Draft mimeo, *World Development Report*, World Bank.
- LUCAS, R. (1988): «On the Mechanics of Economic Development», *Journal of Monetary Economics* 22, pp. 3-42.
- MANKIW, G., ROMER, D. y WEIL, D. (1992): «A Contribution to the Empirics of Economic Growth», *Quarterly Journal of Economics*, pp. 407-37.
- NEHRU, V., SWANSON, E. y DUBEY, A. (1995): «A New Database on Human Capital Stocks in Developing and Industrial Countries: Sources, Methodology and Results», *Journal of Development Economics*, 46, pp. 379-401.
- OECD (1974): *Annuaire des Statistiques de l'Enseignement*, vol. I, París.
- PRITCHETT, L. (1999): «Where has all the education gone?», mimeo, World Bank.
- ROMER, P. (1990): «Endogenous Technical Change», *Journal of Political Economy*, pp. S71-S102.
- STEEDMAN, H. (1996): «Measuring the quality of educational outputs: a note», Center for Economic Performance, Discussion Paper nº 302, LSE.