

Un estudio estadístico de la investigación científica en los países de O.C.D.E.

por
BLAS CABALLERO y
DANIEL PEÑA
Departamento de Estadística, E.T.S.I.I.
Universidad Politécnica de Madrid

RESUMEN

Este trabajo define un índice de producción investigadora de los países de la OCDE aplicando análisis factorial a ocho bancos de datos de publicaciones científicas, y relaciona este índice con indicadores económicos para medir la eficiencia de la actividad científica en distintos países. Se concluye que la eficiencia española es similar a la media de la OCDE, y que su baja producción global puede ser explicada por su escasa inversión relativa en Investigación y Desarrollo.

Palabras clave: Análisis de residuos; Análisis factorial; Regresión; Banco de Datos Científicos.

1. INTRODUCCION Y RESUMEN DEL TRABAJO

La importante conexión entre ciencia, tecnología y desarrollo, ha conducido a que cada vez sea mayor el número de Estados que reúnen sistemáticamente datos estadísticos en materia de ciencia y tecnología. Sin embargo, las grandes diferencias que existen entre estas estadísticas de la ciencia de unos países a otros hacen difícil toda comparación internacional. Por estas razones, este estudio está limitado al conjunto homogéneo de países pertenecientes a la OCDE (Organización para la Cooperación y Desarrollo Económico).

Las medidas de la investigación científica se basan en general en indicadores bibliométricos (véase Elkana y otros (1978)), y el índice más frecuente de producción investigadora de un país es el número de publicaciones en revistas científicas de circulación internacional firmadas por científicos de dicho país. Aunque estas medidas adolecen de muchos defectos, (tienen en cuenta únicamente la cantidad y no la calidad de la investigación, son incapaces de captar la investigación que no se traduce en publicaciones, etc.), ofrecen una base objetiva para las comparaciones entre países y áreas científicas. La "cantidad" de publicaciones puede complementarse con un indicador de su calidad científica, basado en la frecuencia con que los trabajos generados en un país son citados por otros investigadores. (Un índice muy usado es el de Platz que consiste en el logaritmo del número de citas recibidas). Los estudios realizados (véase Jauch y Glueck, 1975) parecen indicar, sin embargo, que estos indicadores de calidad están muy correlacionados con los de cantidad, por lo que en este trabajo nos limitaremos al estudio cuantitativo de la producción investigadora.

El trabajo cubre el trienio 1982-84 y tiene los siguientes objetivos:

1. establecer un índice de producción investigadora de carácter científico-tecnológico en los países miembros de la O.C.D.E.,
2. estudiar la relación entre la producción investigadora (medida por el índice anterior de un país) y su grado de desarrollo económico,
3. utilizar los resultados anteriores para evaluar la producción científica española.

El trabajo se estructura como sigue: la sección dos está dedicada a la descripción de las ocho bases de datos de trabajos de investigación científico-tecnológica, utilizados. Dos de ellas son generales, mientras que las otras son específicas de los campos siguientes: Agricultura, Biología, Medicina, Química, Ingeniería y Física.

La sección tres investiga cómo construir un índice de actividad científico-tecnológica mediante análisis factorial aplicado a los bancos de datos anteriores. El análisis estadístico conduce a un índice agregado de actividad científico-tecnológica que es función de las ocho fuentes consultadas y ordena los países de la OCDE según la Tabla 1.

1.º Estados Unidos	6.º Canadá	11.º Suiza	16.º Austria
2.º Japón	7.º Italia	12.º Bélgica	17.º Noruega
3.º Reino Unido	8.º Australia	13.º España	18.º Yugoslavia
4.º R. F. Alemana	9.º Holanda	14.º Dinamarca	19.º Grecia
5.º Francia	10.º Suecia	15.º Finlandia	20.º Irlanda
			21.º Portugal

TABLA 1

ORDENACION DE LOS PAISES DE LA OCDE MEDIANTE UN INDICE AGREGADO DE PRODUCCION INVESTIGADORA

El índice anterior es absoluto, y no tiene en cuenta la población o riqueza del país. Para poder efectuar comparaciones más realistas, deberíamos construir un índice relativo que tenga en cuenta los factores anteriores. Este aspecto se aborda en la sección 4 que investiga la relación entre el índice general de investigación científico-tecnológica anterior y un conjunto de variables económicas y demográficas. Al relacionar la producción investigadora de cada país con el potencial investigador, medido por población y PNB, se obtiene un índice relativo de actividad con las conclusiones siguientes:

a) Países con un índice de investigación en Ciencia y Tecnología inferior al que les corresponde según su potencial: Portugal, España y Noruega.

b) Países con un índice superior al que les corresponde: Reino Unido, Yugoslavia y Holanda.

El análisis anterior no tiene en cuenta la inversión que cada país realiza para fomentar la investigación. Incluyendo en el modelo los gastos totales en I+D como medida del apoyo económico que el país proporciona a la actividad científica, se obtiene una clasificación de los países por su eficiencia investigadora. Diremos que un país tiene alta (o baja) eficiencia investigadora cuando su producción científica es alta (o baja) con relación a su potencial (medido por población y riqueza) y sus inversiones en I+D. El análisis de los residuos de la regresión entre producción investigadora y las variables anteriores conduce a la clasificación de la tabla 2.

EFICIENCIA INVESTIGADORA	PAIS
Baja	Portugal, Noruega, Bélgica, Irlanda
Media	Resto Países de la OCDE
Alta	Grecia, Holanda y Yugoslavia

TABLA 2

CLASIFICACION DE LOS PAISES DE LA OCDE SEGUN SU RELACION ENTRE LA PRODUCCION INVESTIGADORA Y SU PRODUCCION ESPERABLE SEGUN SUS RECURSOS E INVERSIONES EN I+D

Podemos concluir que España tiene un índice de investigación sustancialmente menor que el que le corresponde por su población y riqueza, confirmando la tradicional falta de interés en la investigación en nuestro país. Por otro lado, cuando incluimos la inversión realizada en Investigación y Desarrollo la eficiencia española es análoga a la media de los países de la OCDE, justificando que en España se investiga poco porque se invierte poco en Investigación.

La eficiencia de la investigación parece ser alta en Grecia, Holanda y Yugoslavia, mientras que la producción es especialmente alta en el Reino Unido, Yugoslavia y

Holanda, aunque en el caso de Yugoslavia este resultado debe tomarse con precaución al existir una heterogeneidad en la fuente de datos. Sin embargo, la alta producción y eficiencia de la investigación en Holanda sugiere un entorno especialmente apto para la investigación científica en este país.

Portugal tiene los resultados más negativos de la OCDE: es el país con menor producción investigadora en todas las bases de datos consultadas, excepto la interdisciplinar francesa, y presenta un nivel de producción muy inferior al correspondiente a su potencial investigador y una baja eficiencia investigadora.

2. LOS DATOS DE PRODUCCION CIENTIFICA

La tabla 3 indica las ocho bases de datos utilizadas, la tabla 4 la matriz de datos resultantes y la 5 sus estadísticos principales.

VARIABLE	BASE DE DATOS	AREA CIENTIFICA	LOCALIZACION
G_1	SCISEARCH	Interdisciplinar	Institute of Scientific Information. Philadelphia. PA. U.S.A.
G_2	PASCAL	Interdisciplinar	Centre National de la Reaserche Scientifique. FRANCE.
A	CAB ABSTRACTS	Agricultura	Commonwealth Agricultural Bureaux. ENGLAND.
B	BIOSIS	Biología	Biosciences Information Service. Philadelphia. PA. U.S.A.
M	EXCERPTA MEDICA	Medicina	Excerpta Medica. Amsterdam. NETHERLANDS.
Q	CHEMICAL ABSTRACTS	Química	Chemical Abstracts Service. Columbus, OH. U.S.A.
I	COMPENDEX	Ingeniería	Engineering information Inc. New York, NY. U.S.A.
F	INSPEC	Física	The Institution of Electrical Engineers. London. ENGLAND.

TABLA 3
BASES DE DATOS UTILIZADAS

Se observa que la investigación científica-tecnológica se halla muy desigualmente repartida en la OCDE, y seis países (EE.UU., Japón, el Reino Unido, la R.F. Alemana, Francia y Canadá) producen más del 70% del total de trabajos de la OCDE en cada fuente. Los Estados Unidos se erigen claramente como el país líder, correspondiéndole más de la tercera parte de la producción a nivel general y en cualquiera de las áreas estudiadas.

		INTER.A	INTER.F	AGRIC.	BIOLO.	MEDIC.	QUIMI.	INGEN.	FISICA
EE.UU.	1	815.319	379.851	88.663	58.104	255.864	440.134	111.015	162.288
UK	2	162.103	90.332	35.158	29.802	59.630	92.725	6.409	34.349
JP	3	105.856	78.811	13.978	16.758	55.634	308.926	32.039	40.538
F	4	118.935	76.186	13.818	11.253	49.938	120.065	9.984	35.792
G	5	91.099	85.037	11.740	20.337	41.233	66.087	11.304	22.093
C	6	72.722	49.459	14.041	16.722	23.139	101.900	12.032	14.645
I	7	42.905	29.734	7.904	13.444	31.078	36.322	5.833	11.351
A	8	36.121	22.236	12.419	9.863	12.186	19.641	4.898	6.299
H	9	29.912	18.036	6.563	7.548	13.721	23.029	3.798	6.775
S	10	28.568	16.190	3.985	9.502	14.852	18.341	2.387	3.549
CH	11	26.495	14.518	3.378	3.636	11.096	19.304	2.556	5.784
E	12	16.425	11.818	3.089	3.981	7.196	15.493	1.258	2.692
B	13	17.311	11.791	3.240	4.011	8.098	11.964	1.772	3.417
D	14	14.677	555	2.635	5.667	8.368	14.266	1.197	1.999
AU	15	10.957	13.154	1.433	2.372	5.928	6.713	1.318	2.278
FI	16	11.012	6.457	2.028	4.756	5.731	6.647	1.001	1.669
N	17	9.075	5.432	1.803	3.299	4.801	5.326	912	853
Y	18	4.686	2.957	2.031	1.194	1.860	7.046	801	1.861
GR	19	3.720	2.749	692	1.293	1.518	2.415	896	1.366
IR	20	6.786	214	432	1.119	1.355	98	522	941
P	21	1.221	1.929	388	386	564	12	493	413

EE.UU. =Estados Unidos

UK = Reino Unido

JP = Japón

G = R. F. Alemana

F = Francia

C = Canadá

I = Italia

A = Australia

H = Holanda

S = Suecia

CH = Suiza

E = España

B = Bélgica

D = Dinamarca

AU = Austria

FI = Finlandia

N = Noruega

Y = Yugoslavia

GR = Grecia

IR = Irlanda

P = Portugal

TABLA 4

NUMERO DE REFERENCIAS POR PAIS Y BASE DE DATOS EN
LAS FUENTES DE LA TABLA 3

	G_1	G_2	A	B	M	Q	I	F
Media	77.424	43.688	10.925	10.717	29.228	62.688	10.113	17.188
Mediana	26.495	14.518	3.378	5.667	11.096	18.341	2.387	3.549
Media recortada	42.598	28.283	7.388	8.766	18.809	46.121	5.309	10.434
Desv. típica	174.761	82.698	19.546	13.228	55.154	111.377	24.200	35.524
C. asimetría (b_1)	4,129	3,673	3,518	2,596	3,796	2,675	4,012	3,742
C. curtosis (b_2)	17,980	14,984	13,590	8,050	15,791	7,070	16,967	15,342

TABLA 5
DESCRIPCIÓN ESTADÍSTICA DE LOS DATOS ORIGINALES

Nota: La media recortada a nivel α se define por

$$T(\alpha) = \frac{1}{n(1-2\alpha)} \sum_{i=n\alpha+1}^{n-n\alpha} x_{(i)}$$

donde $x_{(i)}$ son los valores observados ordenados. La media recortada a nivel 5% es pues el promedio del 95% de las observaciones centrales.

Los coeficientes b_1 y b_2 , de asimetría y curtosis respectivamente, se han calculado mediante:

$$b_1 = \frac{(x_i - \bar{x})^3}{n s^3} \quad b_2 = \frac{(x_i - \bar{x})^4}{n s^4} - 3$$

donde s es la desviación típica. Por tanto, estos coeficientes deberían ser cero para variables aproximadamente normales.

La tabla 5 muestra grandes diferencias entre las medias y las medianas, estando la media recortada al 5% entre ambas. Esta configuración es típica de distribuciones claramente asimétricas y con alta curtosis, aspecto confirmado por los coeficientes de asimetría y curtosis.

La tabla 6 indica el orden de los países según su producción investigadora en los bancos de datos estudiados y la 7 los coeficientes de correlación ordinal de Spearman entre estas clasificaciones. Se observa que las correlaciones son muy altas, siempre por encima de 0,9.

La asimetría y alta curtosis de los datos originales, así como su carácter de variables de "tamaño" (véase Peña (1987)) aconseja transformar las variables en logaritmos. La tabla 8 resume sus características descriptivas después de esta transformación.

		INTER.A	INTER.F	AGRIC.	BIOLO.	MEDIC.	QUIMI.	INGEN.	FISICA
EE.UU.	1	1	1	1	1	1	1	1	1
UK	2	2	2	2	2	2	5	6	4
JP	3	4	4	4	4	3	2	2	2
F	4	3	5	5	7	4	3	5	3
G	5	5	3	7	3	5	6	4	5
C	6	6	6	3	5	7	6	3	6
I	7	7	7	8	6	6	7	7	7
A	8	8	8	6	8	10	9	8	9
H	9	9	9	9	10	9	8	9	8
S	10	10	10	10	9	8	11	11	11
CH	11	11	11	11	15	11	10	10	10
E	12	13	13	13	14	14	12	14	13
B	13	12	14	12	13	13	14	12	12
D	14	14	20	14	11	12	13	15	15
AU	15	16	12	18	17	15	16	13	14
FI	16	15	15	16	17	16	17	16	17
N	17	17	16	17	16	17	18	17	20
Y	18	19	17	15	19	18	15	19	16
GR	19	20	18	19	18	19	19	18	18
IR	20	18	21	20	20	20	20	20	19
P	21	21	19	21	21	21	21	21	21

TABLA 6

ORDENACION DE LOS PAISES SEGUN LAS FUENTES DE LA TABLA 3

	INTER.A	INTER.F	AGRIC.	BIOLO.	MEDIC.	QUIMI.	INGEN.
INTER.F	.94						
AGRIC.	.97	.92					
BIOLO.	.96	.90	.94				
MEDIC.	.98	.94	.96	.96			
QUIMI.	.97	.92	.97	.92	.97		
INGEN.	.96	.95	.95	.93	.96	.97	
FISICA	.97	.95	.96	.91	.97	.98	.97

TABLA 7

COEFICIENTES DE CORRELACION DE RANGOS DE SPEARMAN DE LA ORDENACION DE LA TABLA 6

	L1	L2	L3	L4	L5	L6	L7	L8
Media	4,400	4,117	3,640	3,759	4,037	4,140	3,479	3,723
Mediana	4,423	4,162	3,529	3,753	4,045	4,260	3,378	3,550
Media recortada	4,390	4,134	3,627	3,768	4,032	4,220	3,438	3,703
desv. típica	0,642	0,779	0,602	0,532	0,639	1,050	0,614	0,661
c. asimetría (h_1)	0,253	-0,481	0,154	-0,291	0,018	-1,452	0,916	0,497
c. curtosis (h_2)	0,499	0,346	-0,094	-0,082	0,032	3,063	0,564	-0,283

TABLA 8

DESCRIPCION DE LAS VARIABLES EN LOGARITMOS

Concluimos que al transformar las variables:

a) La asimetría decrece notablemente (el coeficiente de asimetría es mucho más próximo a cero).

b) Los valores extremos tienen poca influencia: la media recortada al 5% es cercana a la media. Esto se traduce en coeficientes de curtosis próximos a cero, valor esperado para la distribución normal.

c) Al ser el logaritmo una transformación monótona, el orden de la producción científica de los países (tablas 6 y 7) no se modifica.

3. CONSTRUCCION DE UN INDICE AGREGADO DE PRODUCCION CIENTIFICA

3.1. Componentes principales

Para construir un índice agregado de actividad vamos a realizar un análisis factorial de la matriz de datos expresando las variables en logaritmos. Como primera etapa del análisis y con carácter exploratorio se efectúa un análisis en componentes principales de la matriz de correlación de las variables. Los resultados principales se resumen en las tablas 9 y 10.

componente principal	valor propio	proporción varianza	proporción varianza acumulada
1	7.37	92.1	92.1
2	.24	3.0	95.1
3	.18	2.2	97.4
4	.10	1.2	98.6
5	.05	.6	99.2
6	.04	.5	99.7
7	.01	.1	99.9
8	.01	.1	100.0

TABLA 9

VALORES PROPIOS DE LA MATRIZ DE CORRELACION DE LAS VARIABLES EN LOGARITMOS

	primera componente	segunda componente	tercera componente
L1 INTER.A	.98	-.04	-.14
L2 INTER.F	.91	.30	.25
L3 AGRIC.	.98	-.02	-.01
L4 BIOLO.	.96	-.18	-.05
L5 MEDIC.	.98	-.07	-.04
L6 QUIMI.	.92	-.26	.24
L7 INGEN.	.95	.18	.12
L8 FISIC.	.97	.10	.10

TABLA 9

CORRELACION ENTRE LAS VARIABLES Y LAS PRIMERAS TRES COMPONENTES

La tabla 9 indica que los tres primeros componentes explican el 97,4% de la variabilidad y los cuatro primeros el 98,6%. Por tanto, podemos representar la mayor parte de la variabilidad observada con tres o cuatro factores. Aunque nuestro objetivo es seleccionar un indicador unidimensional global —que será el primer componente principal— vamos a analizar la información que proporcionan los tres primeros componentes respecto a la estructura de la investigación en los países de la OCDE. Según la tabla 10, despreciando coeficientes menores que 0,1 estos tres componentes pueden escribirse aproximadamente como sigue:

$$C_1 = \frac{1}{\sqrt{7,37}} [0,98 L_1 + 0,91 L_2 + 0,98 L_3 + 0,96 L_4 + 0,98 L_5 + 0,92 L_6 + 0,95 L_7 + 0,97 L_8]$$

$$C_2 = \frac{1}{\sqrt{0,242}} [0,30 L_2 - 0,18 L_4 - 0,26 L_6 + 0,18 L_7 + 0,10 L_8]$$

$$C_3 = \frac{1}{\sqrt{0,179}} [-0,14 L_1 + 0,25 L_2 + 0,24 L_6 + 0,12 L_7]$$

donde las variables L_i están estandarizadas.

La primera componente principal, C_1 , es un promedio de las ocho variables e indica el orden global científico de los países de la OCDE. La segunda, C_2 , está especialmente correlada positivamente con la base interdisciplinar francesa (PASCAL), así como con las bases de Ingeniería y Física, y negativamente con la investigación en Química y Biología. Separa por tanto aquellos países con investigación centrada en Química y Biología de aquellos orientados hacia la Física y la Ingeniería. La tercera, C_3 , contrapone la investigación definida por la base interdisciplinar francesa (PASCAL) a la americana (SCISEARCH) y la investigación en Química a la desarrollada en la Ingeniería. Los valores de C_1 , C_2 , y C_3 , se muestran en la Tabla 11.

Las figuras 1 y 2 muestran los países en el plano definido por parejas de los tres primeros componentes. Se han representado también en el gráfico las ocho variables originales, tomando como coordenadas sus correlaciones con los factores.

	C_1	C_2	C_3
EE.UU.	2.177	1.000	-1.459
UK	1.162	-0.295	-0.190
JP	1.182	0.454	-0.069
F	0.971	0.442	0.052
G	0.923	0.440	-0.053
C	0.801	0.047	0.141
I	0.525	-0.135	-0.093
A	0.316	0.055	-0.053
H	0.197	-0.077	0.133
S	0.058	-0.596	0.355
CH	-0.050	0.138	0.418
E	-0.288	-0.397	1.075
B	-0.288	-0.074	0.615
D	-0.512	-2.838	-1.326
AU	-0.517	0.478	1.126
FI	-0.514	-0.681	0.618
N	-0.688	-0.648	0.876
Y	-0.903	-0.164	1.093
GR	-1.114	-0.323	0.652
IR	-1.554	-0.358	-3.075
P	-1.945	2.885	-0.829

TABLA 11

VALORES DE LAS TRES COMPONENTES PRINCIPALES EN LOS PAÍSES DE LA OCDE

En la figura 1 la primera componente separa los países con gran producción científica (EE.UU., JP, UK) de los de baja (P, IR, GR, Y), mientras que la segunda diferencia países con investigación preferente en Ingeniería y Física (P, EE.UU., AU, F, JP) de aquellos con mayor dedicación relativa a la Química y la Biología (D, FI, N, IR, S, E). La situación de España dentro de este último grupo indica un pequeño peso relativo de la investigación tecnológica frente a la de las ciencias de la vida y la Química. Respecto a las variables, es interesante ver que la variable L2, que representa la fuente interdisciplinar francesa, se sitúa especialmente cerca de Francia, y que las variables de las fuentes de datos de Agricultura, Biología, Medicina y Química definen un cuadrante de "ciencias de la vida" que se contrapone al de las ciencias físico-tecnológicas.

La figura 2 muestra cómo la componente C_3 separa a Irlanda y, en mucha menor medida, Portugal, Dinamarca y EE.UU. del resto. Estos países están caracterizados por un peso relativo escaso de la investigación en Ciencias Químicas, lo que es especialmente cierto en el caso de Irlanda. La tabla 11 indica cómo este tercer componente está muy dominado por el bajo nivel relativo de Irlanda en la Investigación Química (véase la tabla 4).

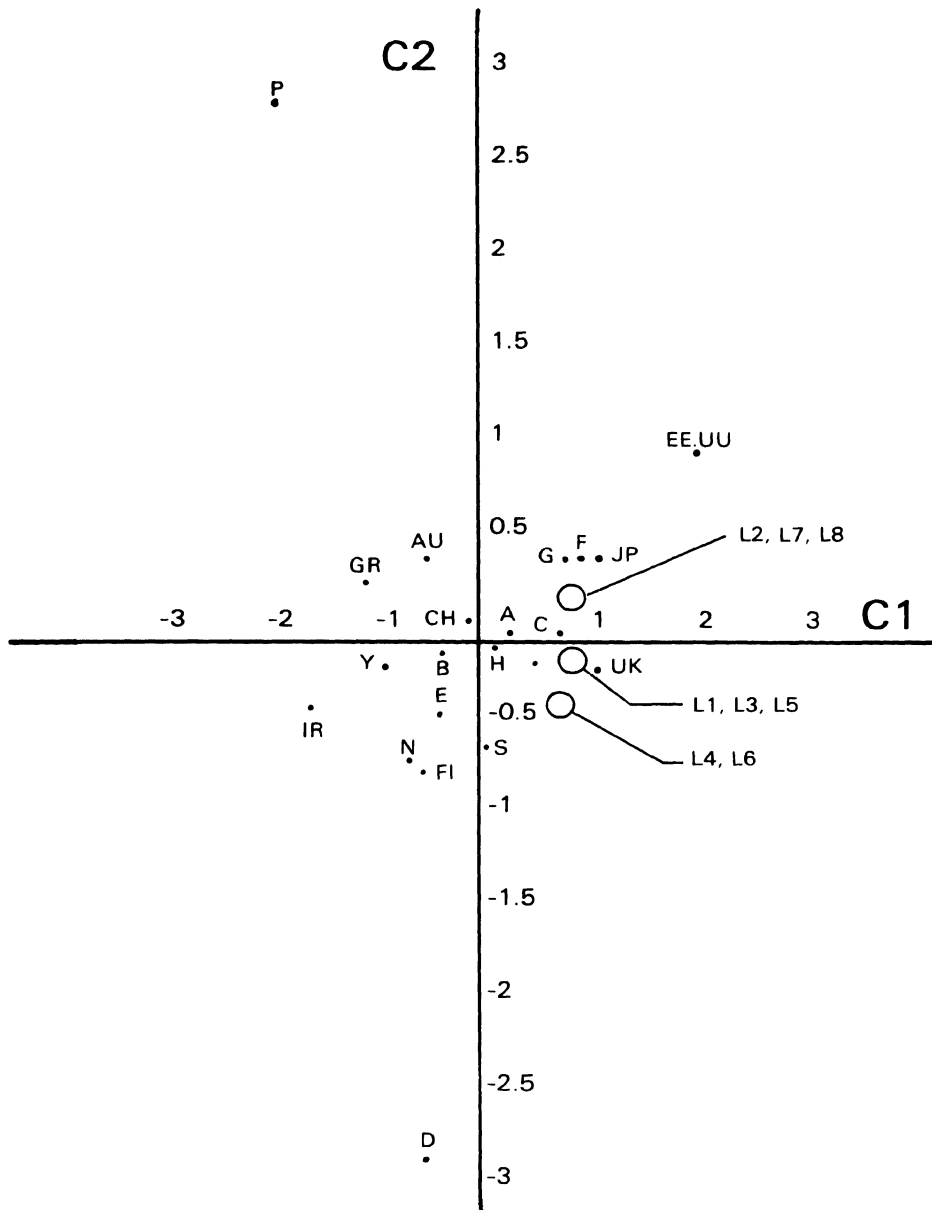


FIGURA 1

SITUACION DE LOS PAISES SEGUN LAS COMPONENTES PRINCIPALES Y CORRELACIONES CON LAS VARIABLES ORIGINALES

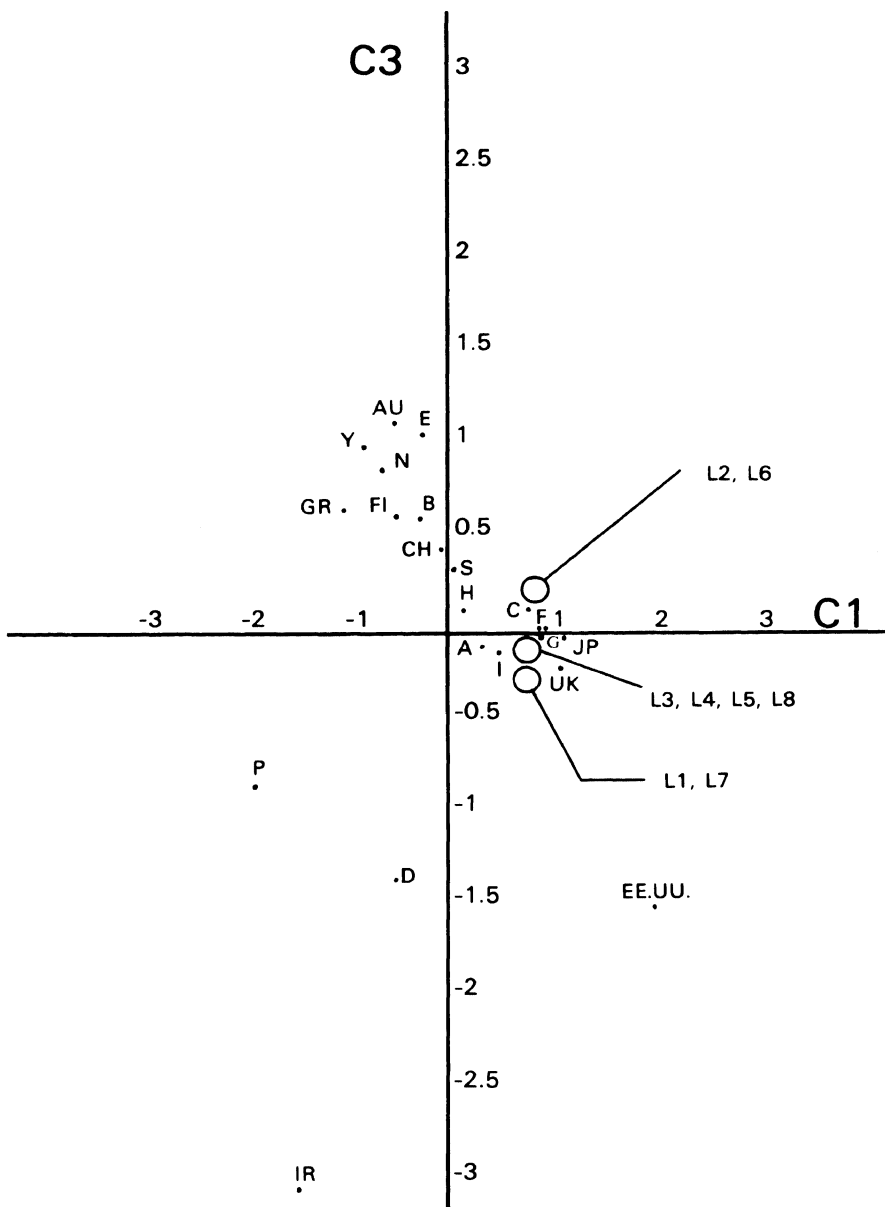


FIGURA 2
SITUACION DE LOS PAISES SEGUN LAS COMPONENTES PRINCIPALES Y
CORRELACIONES CON LAS VARIABLES ORIGINALES

Tomando el primer componente principal como un índice global de actividad, se obtiene el siguiente orden de los países de la OCDE, según su cantidad de producción científica, en los años 1982–1984:

1.º Estados Unidos	12.º Bélgica
2.º Japón	13.º España
3.º Reino Unido	14.º Dinamarca
4.º R. F. Alemana	15.º Finlandia
5.º Francia	16.º Austria
6.º Canadá	17.º Noruega
7.º Italia	18.º Yugoslavia
8.º Australia	19.º Grecia
9.º Holanda	20.º Irlanda
10.º Suecia	21.º Portugal
11.º Suiza	

En términos generales podemos afirmar que éste índice supondrá una mejora sobre las variables originales si la correlación entre este índice y cualquiera de éstas es mayor que las existentes entre una variable cualquiera y el resto. La tabla 7 indica que la mayor correlación se produce con L1 (los coeficientes de correlación de Spearman con las otras variables están en el rango 0,94–0,98) mientras que estos coeficientes entre el primer componente principal y las restantes variables son:

	L1	L2	L3	L4	L5	L6	L7	L8
INDICE	0.99	0.95	0.98	0.96	0.99	0.98	0.98	0.98

Es interesante observar que el rango es ahora 0.99–0.95, lo que mejora la correlación con L1 (base de datos interdisciplinar americana SCISEARCH). La correlación entre el INDICE y L1 es muy alta (0.99) subrayando la fiabilidad y representatividad de que goza ésta importante base de datos internacional.

3.2. *Análisis Factorial*

La aplicación de este método a los datos de las variables L_1 a L_8 en logaritmos, proporciona las contribuciones para los factores, que se resumen en orden decreciente a partir del segundo factor en la figura 3 (al primer factor le corresponde 92.1).

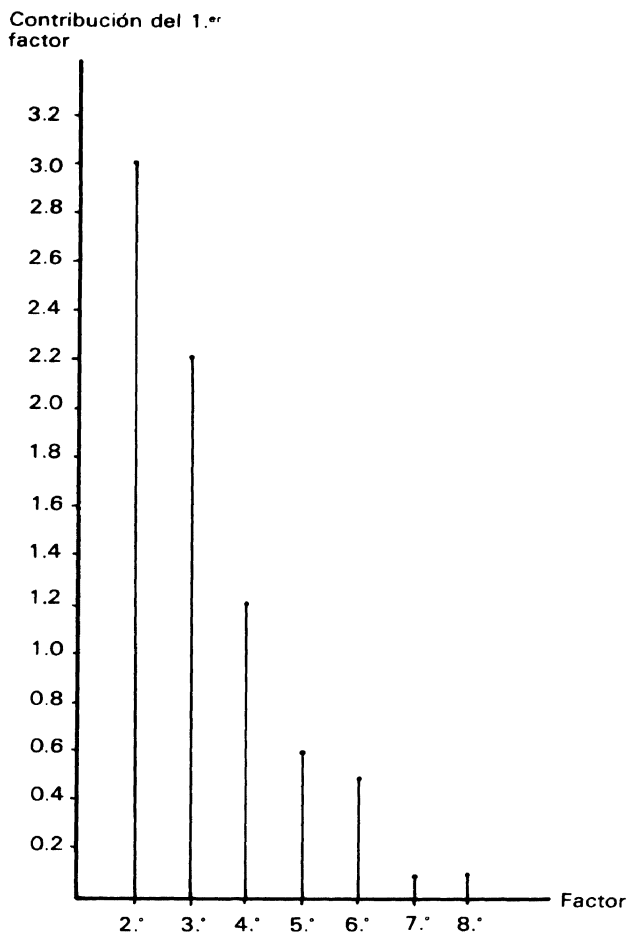


FIGURA 3
CONTRIBUCIÓN DE LOS FACTORES 2 AL 8

Los tres primeros factores cubren el 97,5% de la variabilidad explicada y la figura 3 demuestra que la contribución individual de los restantes es muy pequeña. Suponiendo que existen únicamente tres factores comunes se obtiene la matriz Δ de correlaciones de la tabla 12, que presenta también las comunalidades de cada variable y la contribución de cada factor.

	factor 1	factor 2	factor 3	comunalidades
L1 INTER.A	.98	-.04	-.15	.997
L2 INTER.F	.90	.16	.17	.864
L3 AGRIC.	.97	-.02	.05	.955
L4 BIOL.	.96	-.21	-.01	.973
L5 MEDIC.	.99	-.08	-.01	.984
L6 QUIMIC.	.90	-.13	-.12	.851
L7 INGEN.	.95	.19	-.06	.952
L8 FISIC.	.97	.14	-.07	.976
Proporción de varianza explicada por cada factor (%)	91.44	1.97	1.00	
Proporción de varianza acumulada	91.44	93.41	94.41	

TABLA 12

MATRIZ DE CORRELACIONES Y COMUNALIDADES DE CADA VARIABLE

Los dos primeros factores tienen una interpretación similar a los dos primeros componentes principales de la sección anterior. El primero es un índice general interdisciplinar, el segundo separa los países con alta investigación relativa en Ingeniería y Física de aquellos con mayor producción en Química y Biología. Sin embargo, el tercero diferencia ahora la fuente interdisciplinar americana de la francesa.

En análisis factorial ocurre a veces que la estructura subyacente aparece más claramente de manifiesto rotando los factores. En nuestro caso, las rotaciones ortogonales no condujeron a cambios significativos en la interpretación. Aplicando una rotación oblicua "oblimin" (véase Harman (1980)) se obtiene la matriz de cargas de la tabla 13, que es similar a la 12 y conduce a una interpretación análoga.

Podemos concluir que la producción investigadora de los países de la OCDE puede explicarse por tres factores. El primero es un indicador de tamaño, el segundo es de "forma", y diferencia aquellos países con investigación centrada en la Física y la Tecnología de los centrados en Biología y Química. El tercer factor responde a "quien" recoge los datos y diferencia la fuente interdisciplinar de datos francesa de la americana.

El primer factor define un índice de producción científica que ordena a los países de la OCDE en el orden que presentamos en la tabla 1.

	factor 1	factor 2	factor 3
L1 INTER.A	1.00	.03	-.16
L2 INTER.F	.87	.08	.22
L3 AGRIC.	.97	-.04	.04
L4 BIOLO.	.98	-.18	-.07
L5 MEDIC.	.99	-.06	-.04
L6 QUIMIC.	.90	-.17	.07
L7 INGEN.	.94	.21	.00
L8 FISIC.	.97	.16	-.02

TABLA 13
MATRIZ DE CORRELACIONES TRAS GIRO OBLICUO

4. EVALUACION DEL POTENCIAL Y LA EFICIENCIA INVESTIGADORA DE LOS PAISES DE LA OCDE

4.1. Metodología

En la sección anterior hemos obtenido un indicador global de la producción investigadora en los países de la OCDE. Conceptualmente esta variable debe depender del *potencial* investigador del país, medido por su población y su riqueza, de los *recursos* dedicados a la investigación, y de la *eficiencia* en la utilización de dichos recursos (que será a su vez función de la cultura y tradición investigadora del país, de la organización de la investigación, etc.). Por tanto, llamando y a la variable producción investigadora:

$$y = f(P_1, \dots, P_h; R_1, \dots, R_j; E_1, \dots, E_k) \quad (4.1.)$$

donde las variables explicativas P_i miden el potencial, las R_j los recursos destinados a la investigación y los E_k la eficiencia investigadora.

Este modelo no puede estimarse empíricamente ya que, aunque se dispone de medidas del potencial investigador (Población, PNB o Renta per cápita), las medidas de los recursos dedicados a la investigación son escasas y las medidas de eficacia inexistentes. La tabla 14 resume las variables explicativas disponibles y la tabla 15 sus valores numéricos para los países de la OCDE.

Supongamos que estimamos un modelo del tipo.

$$\hat{y} = f(P_1, \dots, P_h; R_1, \dots, R_j) \quad (4.2.)$$

entonces los residuos de la regresión contendrán información sobre la eficiencia investigadora y el resto de las variables no incluidas. Si suponemos que el efecto principal no incluido en (4.2.) es la eficiencia, aquellos países con grandes residuos positivos indicarán alta eficiencia y aquellos con residuos más negativos, menor eficiencia relativa.

Esta hipótesis puede someterse además a contrastación empírica como sigue: supongamos que ajustamos el modelo (4.2.) con datos de distintos años y que en todos los casos el país X aparece con un residuo positivo alto. Esto indica que este residuo extremo esta recogiendo un componente sistemático debido a una variable omitida, y podemos deducir el valor relativo de dicha variable omitida en ese país.

Para aclarar esta idea, supongamos que se relaciona el rendimiento académico de un grupo de estudiantes con su formación previa (medida por x_1) y las horas de estudio (medidas por x_2), según el modelo.

$$y = f(x_1, x_2) + u$$

donde $u \sim N(0, \sigma^2)$. Si ajustamos este modelo a los datos y comprobamos que los residuos son normales, esto *no* indica que carezcan de información. Pueden en este caso reflejar la inteligencia de los estudiantes que, también, se distribuye normalmente. Para contrastar si los residuos recogen efectos sistemáticos o no, podríamos estimar el modelo con los datos de los mismos estudiantes en diferentes asignaturas. Si en todos los casos el residuo del estudiante 1 es muy alto parece razonable concluir que el residuo contiene el efecto de la inteligencia y que dicho estudiante tiene un CI superior a la media.

Vamos a aplicar esta metodología a los datos anteriores para clasificar a los países en función de su actividad.

VAR	DEFINICION	UNIDAD MEDIDA	AÑO
PNB	Producto Nacional Bruto	billones dolares USA	1979
RN	Renta per cápita	miles dólares USA	1979
GID	Gasto Nacional Bruto en I+D	millones dólares USA	1979
PER	Personal Total empleado en trabajos de I+D	miles de personas	1979
GES	Gastos de la "Educación Superior" en I+D	millones dólares USA	1979
GEO	Gastos en I+D realizados en el sector de las Empresas Comerciales	millones dólares USA	1979
POB	Población	millones de personas	1979

Fuente: la O.C.D.E., según informe "Science and Technology Indicators. Resources Devoted to R & D" (Paris 1984).

TABLA 14
VARIABLES EXPLICATIVAS PRESELECCIONADAS

		PNB	RN	GID	PER	GES	GEC	POB
EE.UU.	1	2376.8	10.8	56560	1334.0	7883	38150	220.6
UK	2	398.1	7.1	7961	310.0	697	5108	55.9
JP	3	867.5	7.5	18189	601.2	3086	10575	115.9
F	4	522.1	8.5	12531	363.2	1635	8659	61.4
G	5	440.3	8.2	7964	230.8	1000	4740	53.5
C	6	223.0	9.4	2438	59.0	538	1030	23.7
I	7	362.9	6.4	3086	94.6	420	1797	56.9
A	8	128.2	8.9	1183	40.7	263	277	14.4
H	9	105.7	7.5	2098	53.8	392	1081	14.0
S	10	81.6	9.8	1608	35.5	333	1067	8.3
CH	11	61.2	9.6	1469	36.6	229	1104	6.4
E	12	198.1	5.3	521	28.5	24	300	37.1
B	13	78.8	8.0	1074	32.1	185	747	9.9
D	14	42.9	8.4	417	15.3	78	213	5.1
AU	15	56.0	7.5	345	15.4	100	291	7.5
FI	16	37.6	7.9	406	16.0	58	222	4.8
N	17	38.4	9.4	524	14.7	124	259	4.1
Y	18	68.1	3.1	586	51.3	95	310	22.2
GR	19	40.6	4.3	75	4.3	8	2	9.5
IR	20	15.6	4.6	116	6.2	17	43	3.4
P	21	31.8	3.2	91	7.6	15	12	9.8

TABLA 15
VALOR, POR PAIS, DE LAS VARIABLES DE LA TABLA 14

NOTA a la tabla: Los datos en negrita en la tabla 15 hacen referencia a 1978 en lugar de a 1979. El paso de monedas nacionales a dólares USA se ha realizado según las "purchasing power parities" y no según el cambio corriente, excepto para Yugoslavia.

	PNB	RN	GID	PER	GES	GEC
RN	.38					
GID	.99	.40				
PER	.98	.38	.99			
GES	.99	.41	.99	.98		
GEC	.99	.40	.99	.98	.99	
POB	.98	.30	.96	.98	.96	.95

TABLA 16
CORRELACIONES ENTRE LAS VARIABLES DE LA TABLA 15

La variable dependiente que utilizamos es el Índice Agregado de Producción Científica obtenido en el apartado 3, al que hemos sumado una constante de valor 2 para trabajar con valores positivos. Interpretaremos esta variable como logarítmica a efectos del cálculo de elasticidades.

4.2. *Relación entre la Producción Investigadora y el Potencial Investigador de un país.*

La regresión que obtenemos al relacionar el Índice Agregado de Producción Investigadora con las variables PNB y RN, que pretenden medir el Potencial Investigador de un país es:

$$I = -2,89 + 1,43 L (PNB) + 2,25 L (RN) \tag{4.3.}$$

(-11,31) (16,43) (7,26)

$$R^2 = 0,97 \qquad s = 0,20 \qquad R^{-2} = 0,966$$

donde entre paréntesis indicaremos siempre los estadísticos *t*.

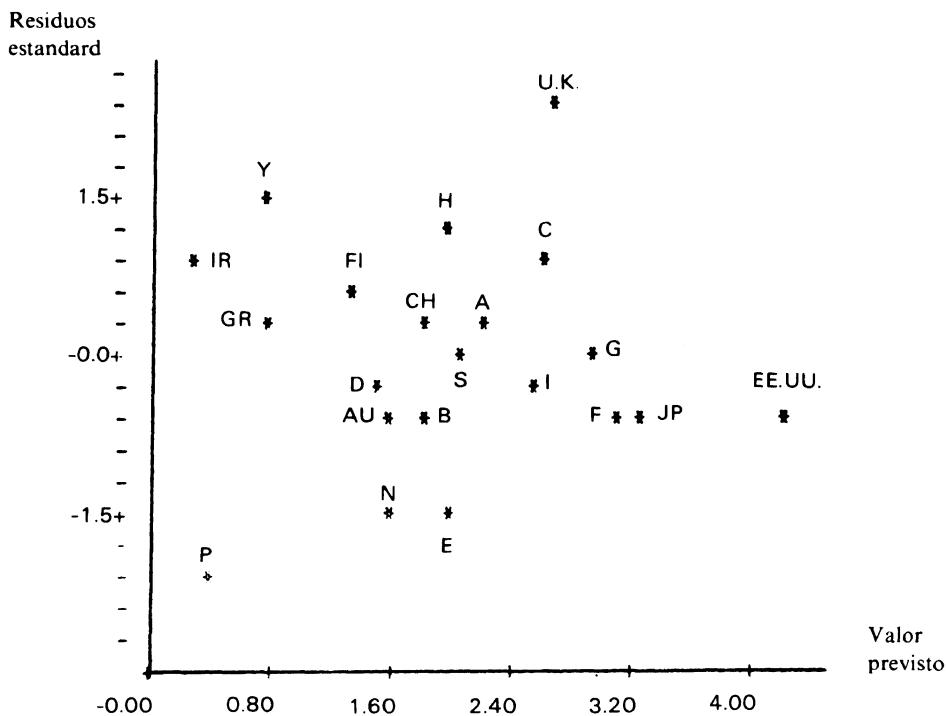


FIGURA 3
RESIDUOS FRENTE A VALORES PREVISTOS PARA EL MODELO (4.3.)

La figura 3 presenta un gráfico de los residuos (estandarizados por su desviación típica), en función de los valores previstos. Se observa que algunos países como Reino

Unido y Yugoslavia tienen residuos muy altos, indicando una producción por encima de su potencial mientras que otros, como España y Portugal, tienen residuos negativos, indicando una producción bastante por debajo de su potencial.

El análisis factorial ha puesto de manifiesto que la tercera componente separaba la base de datos interdisciplinar francesa frente a la anglosajona. Por otro lado, la figura 3 muestra que todos los países de habla inglesa, excepto EE.UU., tiene residuos positivos. Para contrastar la presencia de un sesgo a favor de los países de habla inglesa en las bases de datos, introduciremos en el modelo una variable ficticia, que llamamos "SESGO", que tomará el valor 1 si el país es de habla inglesa (Estados Unidos, Reino Unido, Canadá, Australia e Irlanda) y 0 en caso contrario.

La ecuación estimada resultante es:

$$I = -2,82 + 1,39 L(\text{PNB}) + 2,20 L(\text{RN}) + 0,19 \text{SESGO} \quad (4.4.)$$

(-11,89) (17,11) (7,67) (2,05)

$$R^2 = 0,97 \quad S = 0,18 \quad R^{-2} = 0,967$$

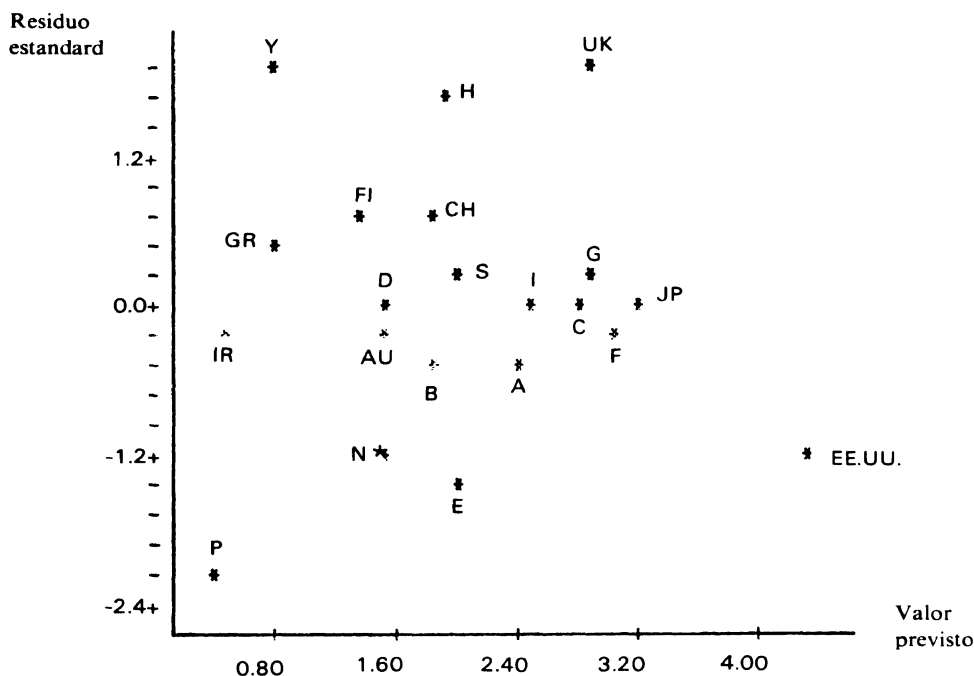


FIGURA 4
RESIDUOS FRENTE A VALORES PREVISTOS PARA EL MODELO (4.4.)

Así pues las elasticidades parciales que obtenemos de la producción científica frente al PNB y la RN son 1,39 y 2,20 respectivamente. La elasticidad total del Índice frente al PNB es 1,68 mientras que la total frente a la RN es de 4,30.

El modelo estima un sesgo positivo del 19% para los países de habla inglesa.

La figura 4 presenta los residuos de esta regresión que delimitan tres grupos de países:

a) Países con un nivel de investigación inferior a su potencial (residuos inferiores al 1,5): Portugal y España.

b) Países con un nivel de investigación superior a su potencial (residuos superiores a 1,5): Yugoslavia, Reino Unido y Holanda.

c) Países con nivel similar a su potencial (residuos entre $\pm 1,5$) resto de la OCDE.

Según el modelo conceptual de la sección 4.1. los residuos de esta regresión recogen conjuntamente el efecto de los recursos dedicados a la investigación y de la eficacia investigadora. Para comprobar que los residuos contienen información sistemática hemos repetido la regresión anterior con datos de distintos años y hemos comprobado que la ordenación anterior se mantiene constante.

4.3. *Relación entre la Producción Investigadora con respecto al Potencial Investigador y los gastos realizados en I+D.*

Primeramente intentamos explicar nuestra variable dependiente (que como hemos dicho al comienzo del apartado 4, es el Índice Agregado de Producción Científica más una constante de valor 2 para trabajar con valores positivos) con las variables definidas en la tabla 1.4. El modelo que obtuvimos presentaba heterocedasticidad a simple vista, que se confirmaba al realizar un contraste de razón de verosimilitudes. También, la relación entre el indicador y las variables explicativas parecía ser lineal y los gráficos de las relaciones parciales sugerían transformar los regresores mediante el logaritmo para obtener linealidad.

Además, con la experiencia de la regresión realizada en 4.1., añadimos al conjunto de las variables explicativas la variable SESGO. Las variables que miden la asignación de recursos a I+D resultan ser altamente colineales y el mejor modelo obtenido es:

$$I = -2,68 + 0,93 L(\text{PNB}) + 1,67 L(\text{RN}) + 0,40 L(\text{GID}) + 0,20 \text{SESGO}$$

$$\begin{matrix} (-12,49) & (4,75) & (5,15) & (2,53) & (2,43) \end{matrix}$$

$$R^2 = 0,98 \quad S = 0,158 \quad R^{-2} = 0,975 \quad (4.5.)$$

Los residuos de este modelo se presentan en las figuras 5 y 6. La figura 5 indica falta de linealidad en la relación y la 6 sugiere que la renta nacional puede ser responsable de ésta.

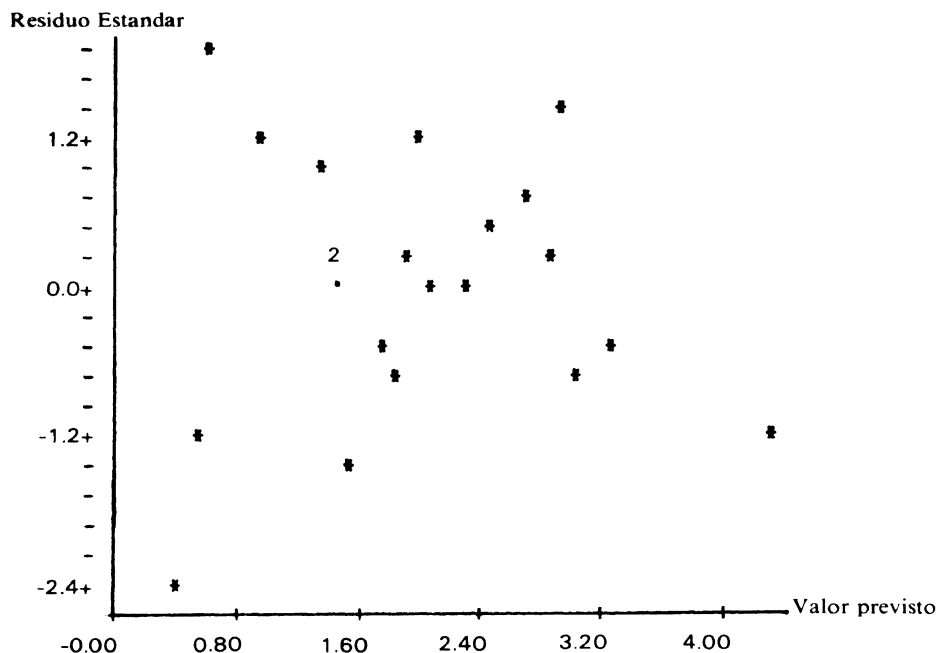


FIGURA 5

RESIDUOS FRENTE A VALORES PREVISTOS PARA EL MODELO (4.5.)

Ambos gráficos sugieren también heterocedasticidad. (La variabilidad parece ser mayor en los países con poca producción investigadora que en los de alta). Para solucionar la falta de linealidad incorporamos al modelo la variable RN^2 , tratando de captar la curvatura puesta de manifiesto en la figura 6. El modelo resultante es:

$$\begin{aligned}
 I = & -3,2 - 0,005 RN^2 + 0,88 L (PNB) + 2,62 L (RN) + 0,44 L (GID) \\
 & (-8,12) \quad (-1,53) \quad (4,65) \quad (3,76) \quad (2,86) \\
 & + 0,23 SESGO \quad R^2 = 0,98 \quad R^{-2} = 0,978 \\
 & (2,78)
 \end{aligned}
 \tag{4.6.}$$

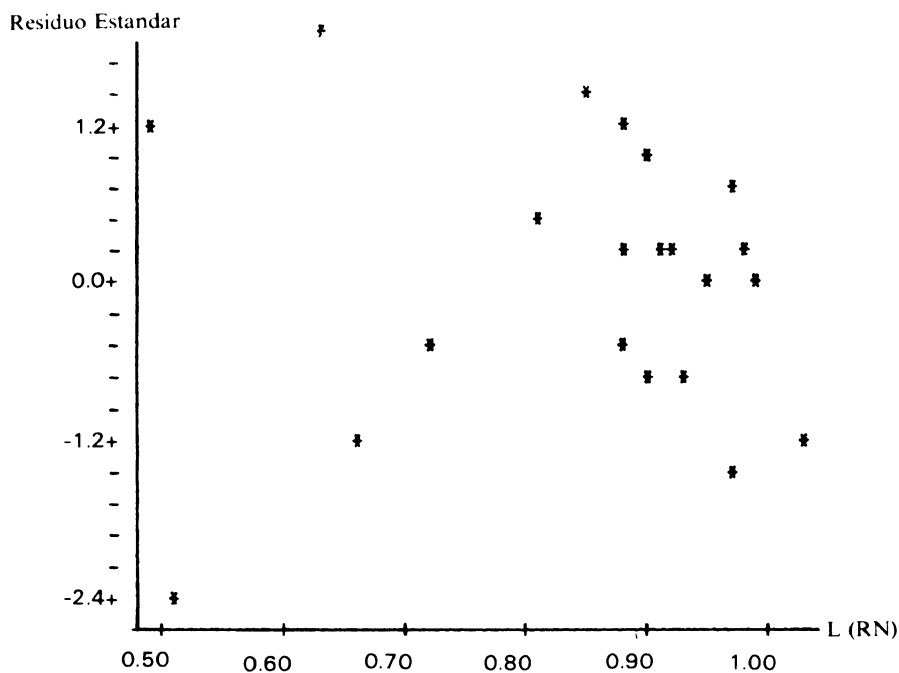


FIGURA 6
RESIDUOS DEL MODELO (4.5) FRENTE A LA VARIABLE L(RN)

Al dibujar la curva $I = f(RN)$ se aprecia que para valores bajos de RN tiene poca pendiente (la producción investigadora aumenta poco respecto al aumento de renta), aumenta en la zona de rentas medias bajas de la OCDE para, finalmente, disminuir de modo asintótico. Esta relación sugiere que aumentos en la renta per cápita de los países más pobres no produce efectos notables en la cantidad de investigación producida. En la segunda zona de la curva, rentas medias bajas, es donde se produce el despegue de la actividad investigadora que tiende a una saturación al aumentar la renta.

Este modelo estima un sesgo positivo del 23% para los países de habla inglesa y la elasticidad de la producción científica respecto al PNB, según el modelo (4.6.) es 0,88 y respecto a los gastos en I+D es 2,62.

La figura 7 presenta los residuos del modelo (4.6.) e indica que, aunque la falta de linealidad se ha corregido, la heterocedasticidad es claramente visible. Para corregirla existen, en principio, dos caminos principales (véase Peña (1987)): el primero es transformar la variable respuesta, el segundo utilizar mínimos cuadrados generalizados. La aplicación de mínimos cuadrados generalizados cuando hay heterocedasticidad, que es en definitiva estimar por máxima verosimilitud los parámetros suponiendo esta hipó-

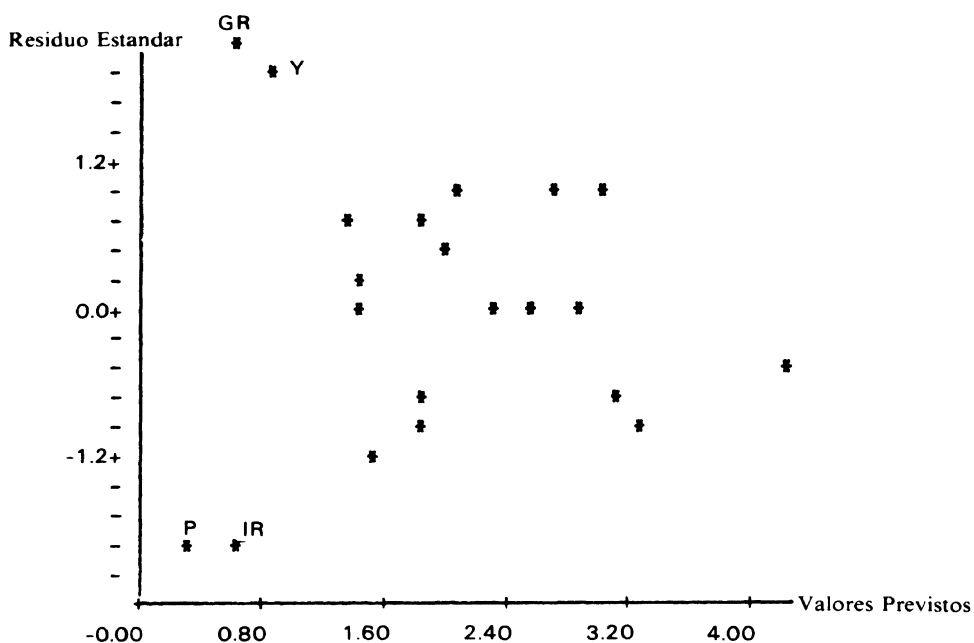


FIGURA 7

RESIDUOS FRENTE A VALORES PREVISTOS PARA EL MODELO (4.6.)

tesis, – se reduce a minimizar una suma cuadrática ponderada de los residuos con coeficientes de ponderación las inversas de las varianzas. Para estimar las varianzas de los datos hemos acudido al procedimiento siguiente:

1.º) Hacer dos grupos, el primero incluye países de RN muy baja: Portugal, Irlanda, Grecia y Yugoslavia. El segundo los restantes.

2.º) Calcular las varianzas de los residuos de cada grupo. Se obtiene una varianza de $(2,266)^2$ para el primero y $(0,743)^2$ para el segundo.

3.º) Utilizar las inversas de estas varianzas como ponderaciones para minimizar la suma de cuadrados.

La tabla 17 resume el resultado de esta estimación comparándola con la estimación por mínimos cuadrados ordinarios (modelo (4.6.).

	β_0	RN^2	L (PNB)	L (RN)	L (GID)	SESGO
Regresión ordinaria	-3'19 (-8'12)	-0'0056 (-1'53)	0'88 (4'65)	2'63 (3'76)	0'447 (2'86)	0'228 (2'78)
Regresión mínimos cuadrados generalizados	-3'18 (-5'64)	-0'0072 (-2'19)	0'68 (4'20)	2'73 (3'15)	0'577 (4'41)	0'334 (5'09)

TABLA 17

La elasticidad de la producción científico-tecnológica frente al PNB es de 0'88 en el primer caso y 0'68 en el segundo, mientras que frente a los Gastos de Investigación y Desarrollo es de 0'447 y 0'577 respectivamente. Las diferencias entre ambos modelos se manifiestan especialmente en el coeficiente del sesgo: según la regresión ordinaria el hecho de que un país sea de habla inglesa supone un aumento del 22'8% en el índice indicador de ciencia y tecnología, mientras que según la regresión por mínimos cuadrados generalizados supone un 33'4%

A pesar de estas diferencias la clasificación de los países por sus residuos es similar en ambas regresiones, con lo que se obtiene la siguiente clasificación de los países de la O.C.D.E. (esta clasificación se mantiene al estimar el modelo con datos de tres años consecutivos distintos):

a) Países con baja eficiencia investigadora al tener un índice de investigación en ciencia y tecnología inferior a lo que les correspondería según sus recursos y lo que invierten en I+D: Portugal, Noruega, Bélgica e Irlanda.

b) Países con alta eficiencia investigadora por tener su índice de investigación en ciencia y tecnología superior al que le corresponde según sus recursos y lo que invierten en I+D: Holanda, Yugoslavia y Grecia.

España se encuentra en el grupo de países con índice de eficacia en investigación en ciencia y tecnología medio. La razón de su baja producción relativa se obtiene estudiando la distribución de gastos en I+D en los países de la OCDE con respecto al PNB y a la RN. La ecuación resultante es:

$$L (GID) = -0'35 + 1'13 L (PNB) + 1'30 L (RN) \quad (4.7.)$$

$$(-1'17) \quad (11'10) \quad (3'53)$$

$$R^2 = 0'92 \quad S = 0'23 \quad R^{-2} = 0,91$$

El análisis de la parte no explicada por la regresión muestra que España es el país con menor residuo, es decir, es el que gasta menos en I+D de la O.C.D.E. teniendo en cuenta su potencial medio por su PNB y su RN. La ecuación (4.7.) indica, por ejemplo,

que España debería haber gastado 1.788 millones de dólares USA en I+D durante 1979, esto es, un 343% más de lo que gastó.

5. CONCLUSIONES

La actividad investigadora de los países de la OCDE puede interpretarse en términos de tres factores: el primero está relacionado con el potencial del país y sus gastos en I+D (que explican el 98% de su variabilidad); el segundo diferencia países con orientación investigadora hacia las ciencias Biológicas y Químicas de aquellos más centrados en Física e Ingeniería; el tercero depende de la selección de fuentes escogidas para medir la producción investigadora.

La variable más importante para explicar la cantidad de actividad realizada es la renta nacional siendo la elasticidad investigación – renta, decreciente con la renta, pero mayor que la unidad. La segunda es el PNB, con elasticidad menor que uno, y la tercera los gastos en I+D con elasticidad del orden de 0,5. Existe un sesgo en los bancos de datos a favor de los países de habla inglesa que puede estimarse entre el 20 y el 30 %.

Entre los países de la OCDE la eficiencia investigadora es especialmente alta en Grecia, Holanda y Yugoslavia y especialmente baja en Portugal. El caso de Yugoslavia debe considerarse con cierta precaución ya que el factor de conversión monetario utilizado ha tenido que ser distinto por la falta de datos – que en el resto de los países.

España destaca por su escasa producción investigadora, que se explica por ser el país de la OCDE con menor inversión relativa en I+D. La eficiencia de la investigación en España teniendo en cuenta los fondos invertidos es, sin embargo, análoga a la media de los países de la OCDE.

Agradecimientos

Los autores agradecen a Antonio Aznar sus comentarios que han contribuido a mejorar la presentación de este trabajo.

BIBLIOGRAFIA

- AIDA MENDEZ (oct. 1986): Los indicadores bibliométricos de la ciencia y su utilidad en la Política Científica. *Política Científica*.
- BMPD STATISTICAL SOFTWARE: University of California Press. USA (1983).
- CONOVER, W. I. (1980): *Practical Nonparametric Statistics*. John Wiley & Sons. USA.

- DILLON, W. y GOLDSTEIN, M. (1984): *Multivariate Analysis: Methods and Applications*. John Wiley & Sons. USA.
- ELKANA, T. et al (1978): *Toward a Metric of Science. The advent of Science Indicators*. John Wiley & Sons. USA.
- FUINCA (1985): *Bases de Datos del mundo*. Consejo Superior de Investigaciones Científicas. España.
- HARMAN, H. (1980): *Análisis Factorial Moderno*. Saltés.
- HERRERA, A. (1968): La Ciencia en el desarrollo de América Latina. *Estudios Internacionales*. Año 2.º, n.º 1. Argentina.
- JAUCH, L. R. y GLUECK, W. F. (1975): Evaluation of University Professors Research performance. *Management Science*, 22, 1, 66–95.
- JOHNSON, R y WICHERN, D. (1982): *Applied Multivariate Statistical Analysis*. Prentice–Hall Inc. USA.
- LOPEZ PIÑERO, J. M. et al (1984): *La Creciente aportación Española a la Ciencia*. Ministerio de Sanidad y Consumo. España.
- MORRISON, D. (1976): *Multivariate Statistical Methods*. McGraw–Hill. USA.
- NORMAN, H. NIE et al (1975): *SPSS Statistical Package for the Social Sciences*. McGraw–Hill. USA.
- O.C.D.E. (1984): *Science and Technology Indicators. Resources Devoted to R & D*. Paris.
- PEÑA, D (1987): Observaciones influyentes en modelos econométricos. *Investigaciones Económicas*, XI, 1, 3–24.
- PEÑA, D (1987): *Estadística: Modelos y Métodos. 2. Modelos Lineales y Series Temporales*. Alianza Universidad Textos.
- PRICE, D. J. S. (1978): *Toward a Model for Science Indicators*.
- RANIS, G y BERANEK, W. (1978): *Science, Technology and Economic Development*. Praeger Publishers, USA.
- RYAN, B., JOINER, B y RYAN, T. (1985): *Minitab. Handbook*. Duxbury Press. USA.
- SABATO, J. A. (1970): Ciencia, Tecnología y Desarrollo. *Confirmado*. Año V. n.º 241. Argentina.
- UNESCO (1984): *Estadísticas relativas a la ciencia y la tecnología*.

SUMMARY

AN STATISTICAL ANALYSIS OF RESEARCH PRODUCTION IN THE OCDE COUNTRIES

This work defines an index of research production in the OCDE countries using factor analysis applied to eight scientific data banks. The relationship between this index and economic indicators is studied as a mean of defining research efficiency in different countries. It is shown that the research efficiency in Spain is similar to other OCDE countries and that is low research production is can be explained by its small investment on that activities.

Key words: Residual Analysis; Factor Analysis, Regresión; Scientific Data Banc.

AMS, 1980. Subject classification: 62J05; 62H25; 62P20.