

EL HJ-BIPLLOT COMO BASE PARA EL ANALISIS DE LA EVOLUCION DEL INDICE DE PRODUCCION INDUSTRIAL EN LA COMUNIDAD EUROPEA EN LOS ULTIMOS AÑOS.

S. VICENTE TAVERA* M.P. GALINDO VILLARDON**

* E.U. de E. empresariales ** Facultad de Medicina

Departamento de Estadística y Matemática Aplicadas. Universidad de Salamanca

La Economía es una Ciencia que se caracteriza, esencialmente, por analizar fenómenos que están definidos en función de un elevado número de variables, por tanto la importancia de los Métodos Multivariantes en este campo es evidente.

Los Métodos Gráficos Multivariantes son métodos eminentemente descriptivos, que intentan resumir la información mediante la transcripción directa de los datos que producen un conjunto de variables; sin embargo se confunden con relativa frecuencia con los **Métodos Estadísticos Multivariantes Gráficos**; es decir, análisis de matrices de datos multidimensionales que presentan sus resultados en forma gráfica, los cuales son potentes herramientas de diagnosis: complejos procesos algebraicos asentados sobre métodos numéricos deben ser definidos antes de pasar del hiperespacio de partida a la bigeometría en la cual se representan las unidades taxonómicas. El resultado final suele ser un diagrama de dispersión sobre un sistema cartesiano.

Los Análisis de Componentes Principales, el Análisis Factorial de Correspondencias junto con el Análisis Factorial, y el Análisis de Cluster, son las Técnicas multivariantes más utilizadas en Economía, especialmente en Investigación Comercial (Marketing) y estudios socioeconómicos. (ABASCAL y GRANDE, 1989, I.N.E. 1991).

Otras técnicas, menos conocidas, son los Análisis BIPLLOT propuestos por GABRIEL en 1971 y ampliamente desarrollados en sucesivos trabajos GABRIEL 1978, GABRIEL y ODOROF 1990; TSIANCO y GABRIEL 1984 GOWER 1990. Este procedimiento multivariante proporciona representaciones de n poblaciones y p variables en un subespacio de baja dimensión, usualmente el plano o el espacio.

Ambas representaciones pueden ser superpuestas para proporcionar información adicional acerca de las relaciones entre variables y observaciones, la cual no está disponible en las representaciones separadas.

Este método, que puede ser aplicado a cualquier matriz de datos, a diferencia de lo que ocurre con el análisis Factorial de Correspondencias, ha sido ampliamente probado en otros campos de la Ciencia pero **no tenemos referencia de que haya ninguna aplicación en Economía.**

GALINDO en 1985 y 1986, propone otra alternativa de representación simultánea el HJ-BIPLLOT que es una extensión de los Biplots de GABRIEL con claras ventajas sobre ellos.

En este trabajo hemos elegido los países de la Comunidad Europea y el Índice de Producción Industrial (IPI) puesto que mediante este índice se puede expresar la evolución física del Producto Interior Bruto referido a la Industria.

Se trata de uno de los instrumentos fundamentales en el estudio de la situación actual de la Economía, en la identificación de los factores que han conducido la Economía al estado actual y en la elaboración de previsiones sobre la evolución futura a corto plazo (MOCHON & ANCOCHEA, 1981).

En Análisis de Coyuntura, por lo general, estos estudios se llevan de forma individual para los distintos países, haciendo comparación entre ellos a través de los correspondientes índices comunes.

El Análisis Multivariante nos permitirá analizar de forma conjunta todos estos países tomando como referencia todas las series de índices que deseemos incorporar en el estudio.

El objetivo planteado en este estudio es **el análisis de la evolución del Índice de Producción Industrial en la Comunidad Económica Europea, en los últimos años.**

DESCRIPCION DE LOS DATOS ANALIZADOS

Se ha considerado la matriz de 125 filas por 15 columnas, donde las filas representan los diferentes índices de producción industrial de los países de la Comunidad desde 1978 a 1989, siguiendo la clasificación según las **ramas de actividad económica**. Los datos han sido obtenidos de la revista Coyuntura Industrial correspondiente al mes de noviembre de 1990 y publicada por EUROSTAT. (Tabla 1)

En la matriz descrita, se eliminaron los índices de Luxemburgo ya que tenían muchos datos faltantes y si se pretendía incluir a dicho país, la matriz perdía representatividad. Los índices correspondientes a Portugal desde 1978 a 1984 no se incluyen en el estudio puesto que no se

dispone de información sobre los mismos.

De las 22 columnas que se podían incluir en el estudio, se eliminaron 7 de éstas, ya que no se dispone de datos suficientes para completar dicha matriz.

Las 15 variables para las que había información completa son, pues, las siguientes:

- Energía.
- Industrias Extractivas.
- Industrias Manufactureras.
- Producción y primera transformación de metales.
- Industrias mineras no metálicas.
- Industria Química y fibras artificiales.
- Industrias transformadoras de metales sin mecánica de precisión.
- Fabricación de artículos de metal.
- Construcción de maquinarias.
- Construcciones eléctricas y electrónicas.
- Construcción de otros medios de transporte.
- Alimentación, bebidas y tabaco.
- Industrias de alimentación.
- Industria textil.
- Calzado y Habitualamiento.

METODOLOGIA

El análisis estadístico multivariante recoge varias técnicas de análisis de datos que tratan de describir, clasificar y clarificar la información, las cuales podrían clasificarse en técnicas probabilísticas, o inferenciales, y técnicas descriptivas, siendo éstas últimas más interesantes para conseguir el objetivo planteado.

Dentro de las técnicas descriptivas hemos seleccionado la familia de técnicas factoriales que descomponen la variabilidad global de los datos en direcciones principales de inercia, y presentan los resultados en forma de gráficos cartesianos sobre los que es posible situar las unidades taxonómicas y las variables, para poder interpretar después sus posiciones relativas en los subespacios de máxima inercia.

Estos métodos, más concretamente las técnicas de representación simultánea de datos multidimensionales, serán el armazón de este trabajo.

El Análisis Factorial de Correspondencias propuesto por BENZECRI (1973) es la técnica de representación simultánea más utilizada; sin embargo, este potente procedimiento de

ordenación está pensado para tablas de contingencia y, aunque puede ser aplicada a cualquier matriz de datos no negativos, es imprescindible que en el estudio tenga sentido contrastar la igualdad de perfiles (valores relativos a los totales marginales).

En este trabajo en el que se analizan valores de índices en países, las matrices utilizadas como INPUT no son tablas de frecuencias, por tanto los métodos BILOT de GABRIEL (op.cit), generalizados más tarde por GALINDO y CUADRAS (op.cit), se espera que proporcionen resultados mucho más interesantes.

El análisis comparativo de los resultados encontrados con ambos procedimientos (AFC y Métodos BILOT) también será ampliamente descrito, puesto que el comparar la metodología clásica, con los procedimientos de difusión más reciente, es otro de los objetivos de este trabajo.

En el trabajo se han utilizado tablas trifactoriales (**índices en instantes t**, separados por intervalos de duración constante, para los distintos **países**) siguiendo el procedimiento clásicamente utilizado por BENZECRI y sus colaboradores en el análisis de tablas de contingencia trifactoriales. (Ver, por ejemplo, BENZECRI, 1987). En este caso, **el análisis de la tabla nos muestra la evolución temporal del perfil de un índice, sobre un país, en un periodo de tiempo determinado.** (Tabla 1)

Obviamente pueden existir discrepancias entre la configuración obtenida y la secuencia real. Las discrepancias dependen de la pérdida de información al proyectar a subespacios de baja dimensión. Si por ejemplo, el subespacio de las "p" primeras dimensiones absorbe gran parte de la Inercia Total se puede suponer que lo esencial del fenómeno observado por las series cronológicas está representado.

RESULTADOS

La inspección de la matriz de datos multivariantes se lleva a cabo utilizando dos técnicas de representación simultanea, Análisis Factorial de Correspondencias y HJ-Biplot.

INDICES DE PRODUCCION

		ENERG.	I. EXTRACT.	I. MANUF.CALZ.+HABIT.
P A I S E S	B(78)	100.2	110.9
	B(79)	99.1	111.0

	B(89)	101.6	109.6
	DK(78)	100.2	79.2

	DK(89)	100.5	77.6
	D(78)	93.0	128.4

	D(89)	101.1	85.8

P(89)	137.2	110.2	

Tabla 1: Estructura de la matriz empleada en el estudio del I.P.I. por ramas de actividad en la C.E.

En ambos estudios la absorción de inercia es similar, siendo ligeramente superior en el segundo. Las dos técnicas proporcionan tres ejes principales significativos. (Tabla 2)

	Inercia absorbida		Tasa de inercia en el
	Eje 1	Eje 2	espacio Eje 3
AFC	33.34%	30.70%	9.21%
HJ-BIPLLOT	38.33 %	29.53%	11.03%
			73.32%
			78.89%

Tabla 2: *Absorción de inercia en el A.F.C. y en el HJ-BIPLLOT*

Analizando la absorción de inercia de los dos primeros ejes vemos que en el Análisis Factorial de Correspondencias la diferenciación entre los dos primeros ejes es mínima, por lo cual la interpretación de los resultados factoriales está a nivel de plano y no de ejes; este hecho está menos acentuado en el Análisis Biplot, de lo cual se deduce que en este segundo Análisis los resultados serán más claramente interpretables, en relación a cada eje.

Las figuras 1 y 2 recogen la proyección de las hipernubes en el subespacio de máxima inercia, con las dos técnicas arriba descritas. En la 1 según el HJ-Biplot y en la 2 según el AFC.

El estudio de las variables, en ambas técnicas, pone de manifiesto que la variable 11 (Construcción de material de transporte) tiene gran importancia en la interpretación del primer eje factorial. (Tabla 3)

En el HJ-Biplot, al analizar las Contribuciones Relativas del Factor (eje 1) al Elemento la variable 11 presenta la mayor contribución de eje 1, las variables 12 (Alimentación, Bebidas y Tabaco), 1 (Energía) y 13 (Industrias Alimenticias) tienen Contribuciones sensiblemente inferiores (490,480 y 340) frente a la primera (940).

En el Análisis Factorial de Correspondencias, le siguen en importancia a la variable 11, la variable 2 (Industrias Extractivas) con una Contribución similar, (650) frente a 670 de la primera, quedando en tercer lugar la variable 1 (Energía), pero con una Contribución algo inferior (420).

Las variables 7 (Industrias de Transformación del Metal) y 10 (Construcciones Eléctricas y Electrónicas) son las que tienen mayor interés en la interpretación del eje 2; en el HJ-Biplot en primer lugar está la variable 7 y en segundo la 10 con unas Contribuciones de 830 y 720 respectivamente, mientras que en el Análisis Factorial de Correspondencias es al contrario, con 830 y 620 respectivamente.

En el HJ-Biplot intervienen más variables sobre eje 2 que en el Análisis de Correspondencias,

ya que, aparte de las dos mencionadas anteriormente, intervienen las variables 3 (Industrias Manufactureras), 2 (Industrias Extractivas), 6 (Industrias Químicas) y 14 (Industria Textil). Estas dos últimas variables presentan también relativa importancia en el Análisis Factorial de Correspondencias.

El tercer Factor es diferente en las dos técnicas puesto que en el HJ-Biplot la variable que recibe mayor contribución es la 9 (Construcción de Maquinaria), con una Contribución de 380, y en el Análisis de Correspondencias son la 5 (Industrias Mineras no Metálicas) y 8 (Fabricación de Artículos de Metal) con 370 y 350 respectivamente.

El análisis de las filas hemos preferido hacerlo en forma individualizada para cada país. No obstante conviene hacer un pequeño comentario de la representación conjunta en las dos técnicas, las cuales, como podemos apreciar en la representación, son muy similares.

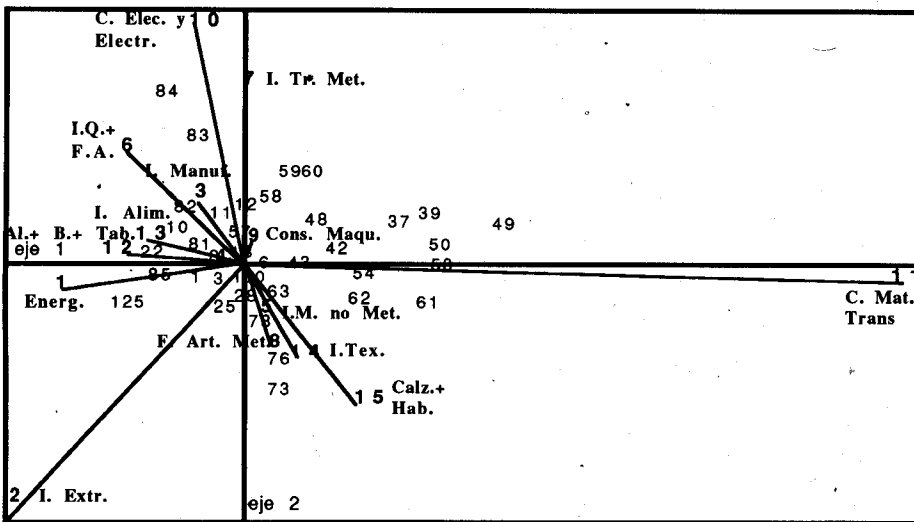


Fig. 1: Representación HJ-BIPLLOT de los Indices de Producción Industrial de los países de la C.E.E. por ramas de actividad en el plano 1-2.

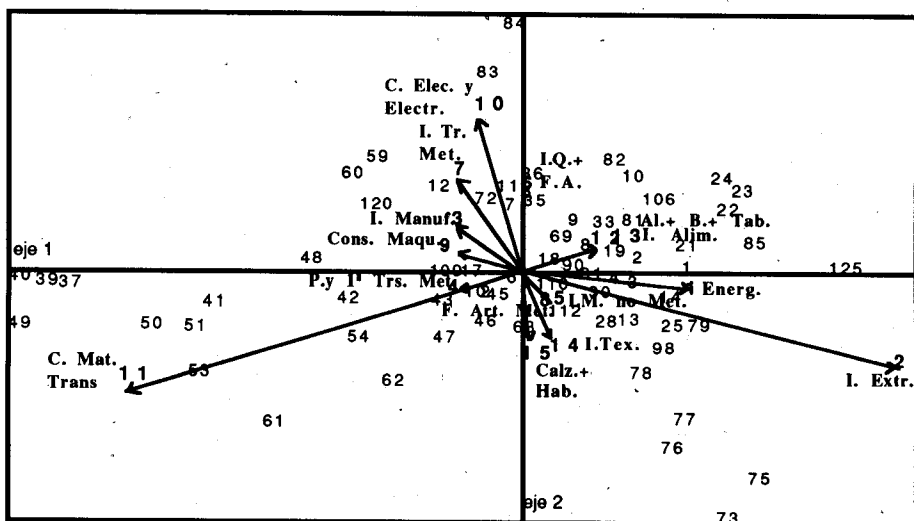


Fig. 2: A. F. C. de los Indices de Producción Industrial de los países de la C.E.E. por Ramas de Actividad en el plano 1-2.

Analizando las variables, por una parte vemos que, sobre eje 1, en ambos gráficos, la variable 11 (Construcción de material de transporte) es la que presenta mayor variabilidad, y ocupan posiciones extremas, las variables 1 (Energía), 12 (Alimentación, bebidas y tabaco) y 13 (Industrias alimenticias) respectivamente.

Sobre eje 2, la variable 10 (C. Elec. y Electrónica) ocupa posición extrema a las variables 14 (I. Textil) y 15 (Calzado y Habit.), en ambas representaciones.

El análisis de los puntos fila es también muy similar, aunque en la representación HJ-Biplot aparecen mas concentrados.

Con el fin de poder tener una mejor representación de los puntos fila en la representación HJ-Biplot hemos creído conveniente presentar los resultados referidos a la escala más extrema (Ver figura 3); así resulta más fácil la interpretación de los puntos fila.

Las variables que no tienen punta de flecha en el gráfico, como por ejemplo la 2, 11 y 10, exceden las medidas del mismo. El resultado es el siguiente:

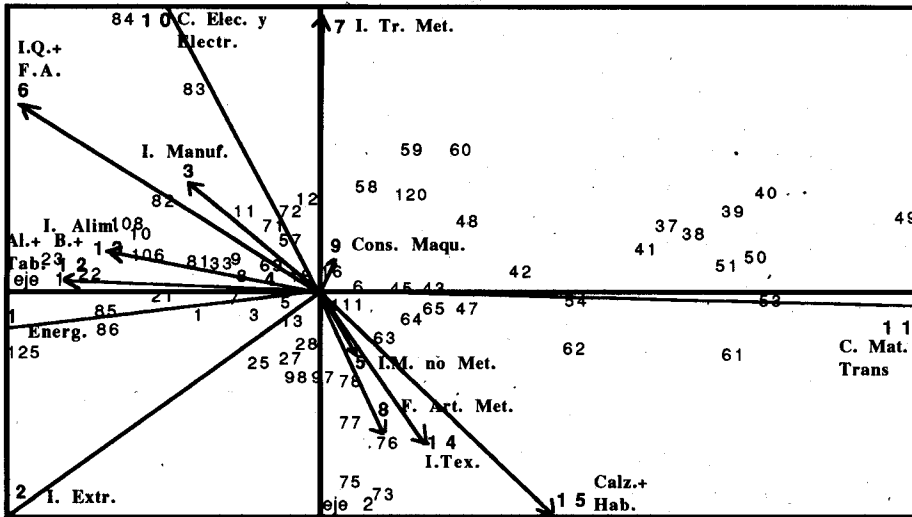


Fig. 3: Representación HJ-BIPLLOT de los Indices de Producción Industrial de los países de la C.E.E. por ramas de actividad en el plano 1-2.

HJ-BIPLLOT

CONTRIBUCIONES RELATIVAS DEL FACTOR AL ELEMENTO (C.R.F.E.)

	EJE1	EJE2	EJE3		EJE1	EJE2	EJE3
1:	480	19	74	2:	210	530	230
3:	160	670	33	4:	36	1	200
5:	6	73	220	6:	230	440	0
7:	0	830	5	8:	13	240	160
9:	2	32	380	10:	15	720	100
11:	940	1	47	12:	490	5	0
13:	340	51	0	14:	64	420	72
		15:	110	360	150		

ANALISIS FACTORIAL DE CORRESPONDENCIAS

CONTRIBUCIONES RELATIVAS DEL FACTOR AL ELEMENTO (C.R.F.E.)

	EJE1	EJE2	EJE3		EJE1	EJE2	EJE3
1:	420	8	0	2:	650	180	150
3:	170	400	28	4:	120	46	19
5:	27	89	370	6:	0	480	45
7:	68	620	52	8:	22	91	350
9:	150	62	24	10:	15	830	50
11:	670	260	57	12:	200	110	27
13:	140	220	30	14:	20	470	48
		15:	1	290	31		

Tabla 3: C.R.F.E. para las variables en el HJ-BIPLLOT y en el A.F.C.

Analizaremos individualmente alguno de los países en la representación Biplot, comparándolos con la obtenida en el Análisis Factorial de Correspondencias. El análisis del resto de los países se haría de forma idéntica a los que presentamos.

El análisis de la calidad de representación de los puntos proyectados , permite explicar el hecho de que ciertos países planteen problemas en su estudio individualizado debido a que los puntos que representan a los mismos tienen una calidad de representación bastante baja.

Irlanda es el país mejor representado sobre eje 2, nueve de los doce puntos de su serie tienen Contribuciones Relativas de Factor al Elemento altas sobre dicho eje en las dos representaciones. (Ver figuras 4 y 5)

Los gráficos nos ponen de manifiesto la alineación de los 12 puntos sobre el eje 2. Los primeros años de la serie, en la representación Biplot, tienen relación con la variable 15 (Calzado y Habitallamiento), donde podemos apreciar como esa serie, en este país, va disminuyendo a medida que nos alejamos de los primeros años de la serie. También hemos de hacer mención, en las dos representaciones, la importancia que tienen las variables 10 y 7 (Construcciones Eléctricas y Electrónicas e Industrias de Transformación del Metal) en los últimos años de las series. Analizadas estas dos últimas series vemos la variabilidad en ambas es muy grande variando de 47.8 en 1978 a192.5 en 1989 y de 42.1 a 216.6 respectivamente. Los gráficos reflejan perfectamente esta situación para las dos variables.

En el gráfico de Correspondencias, la relación que se daba en los primeros años con la variable 15 no se aprecia de manera clara, ya que la Contribución sobre dicho eje es pobre, aunque si se puede relacionar con la variable 14 (Industria Textil), donde tiene una Contribución ligeramente más alta, variable también con cierta influencia en los primeros años de la serie en la representación Biplot.

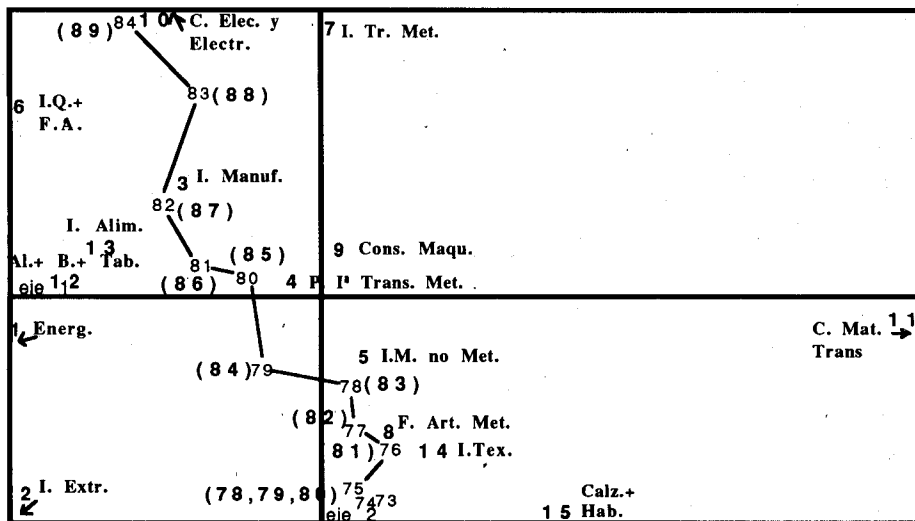


Fig 4: Representación HJ-BIPLLOT de los Indices de Producción Industrial según las Ramas de Actividad de Irlanda en el plano 1- 2

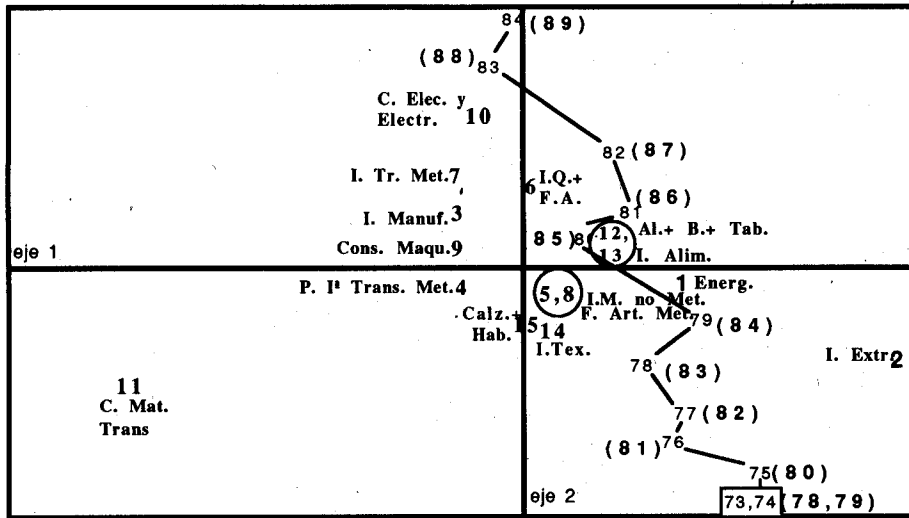


Fig. 5: Representación simultánea de A F C de los Indices de Producción Industrial por Ramas de Actividad de Irlanda en el plano 1-2.

Uno de los países que mejor queda representado en las dos proyecciones es **Grecia**. El análisis de las Contribuciones Relativas del Factor al Elemento en el HJ-Biplot permite apreciar que la mayoría de los puntos correspondientes a este país tienen altas contribuciones de eje 1 y en menor importancia sobre eje 3. En el Análisis Factorial de Correspondencias ocurre lo mismo pero la importancia del eje 3 es menor. (Ver figuras 6 y 7)

En la primera representación podemos apreciar la importancia que tiene la variable 11 (Construcción de Material de Transporte). La representación sigue perfectamente la serie en sus 12 valores, puesto que vemos como existe un aumento desde el año 78 al 81 con índices que van de un 141.3 a 172.3, disminuyendo la serie hasta 1985 para posteriormente aumentar hasta el 88 y disminuir ligeramente en el 89. La variable 15 (Calzado y Habitualmente) tiene cierta importancia en los años 78,79 y 80 donde toma los valores más altos.

Las conclusiones que se obtienen en la representación mediante la técnica del Análisis Factorial de Correspondencias son similares manteniéndose la misma tendencia en la serie de la variable 11, excepto el año 1989.

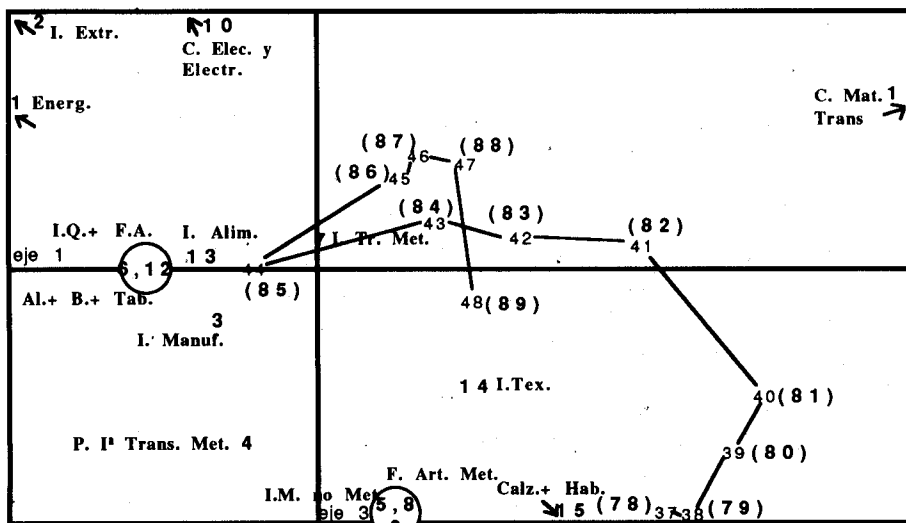


Fig. 6: Representación HJ-BIPLOT de los Indices de Producción Industrial según las Ramas de Actividad de Grecia en el plano 1-3.

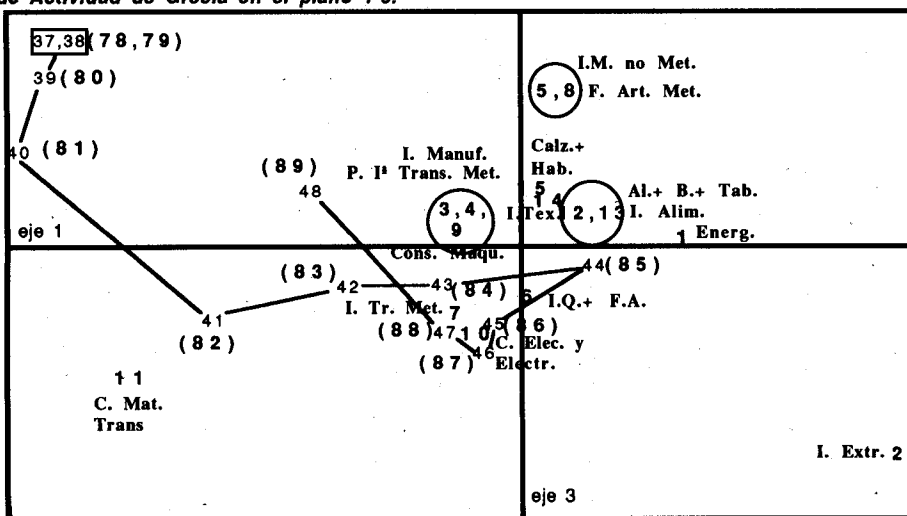


Fig. 7: Representación simultanea de Análisis Factorial de Correspondencias de los Indices de Producción Industrial por Ramas de Actividad de Grecia en el plano 1-3.

Como en el análisis A.F.C. la variable 15 apenas interviene, no se detecta su importancia en los primeros años de la serie, cosa que si ocurría en el estudio del Biplot. La explicación de este hecho está en que esta técnica trabaja sobre valores relativos que hace que se suavice las diferencias, de aquí que no detecte la importancia de esta variable.

Al igual que Grecia, **España** es otro de los países que mejor queda representado; éste en el plano principal. Analizando las Contribuciones Relativas del Factor al Elemento, en los dos métodos vemos que la mayoría de los puntos fila tienen altas Contribuciones sobre eje 1, ésto hace que su representación quede alrededor de dicho eje. (Ver figuras 8 y 9)

En ambos casos, en mejor medida en el Biplot, la serie de la variable 11 (Construcción de Material de Transporte) se ve perfectamente reflejada en los gráficos; se aprecia una disminución en la serie desde 1978 (211.4) a 1979 (181.9), aumentando ligeramente hasta 1983 (196.5), para disminuir progresivamente hasta 1985, aumentando de nuevo hasta 1989 (152.1).

También es interesante apreciar la relación que tienen los años 1986 a 1989 con las variables 7 y 10 (Industrias de Transformación del Metal y Construcciones Eléctricas y Electrónicas), donde en esos años dicho país toma valores altos. La variable 3 (Industrias Manufactureras) tiene cierta importancia desde 1986 a 1989; efectivamente se aprecia un incremento apreciable en dicho periodo.

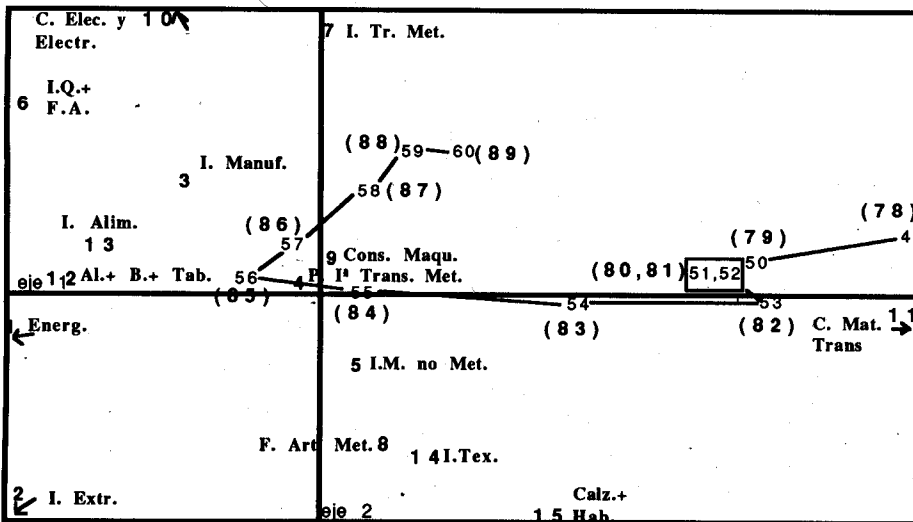


Fig. 8: Representación HJ-BIPLLOT de los Indices de Producción Industrial según las Ramas de Actividad de España en el plano 1-2.

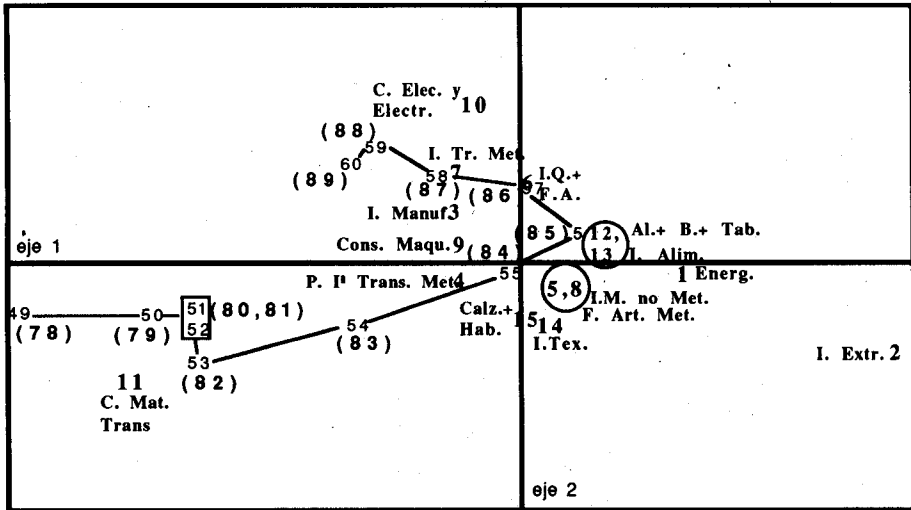


Fig. 9: Representación simultánea de Análisis Factorial de Correspondencias de los Índices de Producción Industrial por Ramas de Actividad de España en el plano 1-2.

CONCLUSIONES

El método HJ-Biplot se ha revelado como una herramienta multivariante sumamente útil en el análisis de datos económicos para describir simultáneamente una estructura y su evolución en el tiempo. Su aplicación, puede ser muy interesante en Análisis de Coyuntura, especialmente en la etapa descriptiva, ya que nos permite estudiar conjuntamente un elevado número de indicadores coyunturales referidos a diferentes países.

Este método tiene claras ventajas en relación al Análisis Factorial de Correspondencias especialmente debido a las limitaciones de aplicabilidad de ésta última técnica.

BIBLIOGRAFIA

- ABASCAL, E. y GRANDE, I. (1989): *Métodos Multivariantes para la Investigación Comercial. Teoría, aplicaciones y programación BASIC*. Ariel Economía.
- BENZECRI, J.P. (1973/82): *L'Analyse des données*. Tomo II. L'Analyse des correspondences. Dunod, Paris. 4ª Ed
- BENZECRI, J.P. (1987): 'L'Analyse des Series Chronologiques Decalées. Exemple de L'Histoire Monétaire de la France de 1910 à 1945'. *Les Cahiers de L'Analyse des Données*, Vol. XII. Nº 4: 219-309.
- GABRIEL K.R. (1971): 'The Biplot-graphic display of matrices with application to Principal Component Analysis'. *Biometrika* 58: 453-467
- GABRIEL, K.R. (1978): 'Least Squares Approximation of Matrices by Additive and Multiplicative Models'. *J. R. Statist. Soc. Boll*, 40 (2): 186-196.
- GABRIEL, K.R. y ODOROFF, CH.L. (1990): 'Biplots in Biomedical Research'. *Statistic in Medicine*, Vol. 9, 469-485.
- GALINDO, M.P. (1986): 'Una alternativa de representación simultanea: HJ-Biplot'. *Qüestió* . nº1:13-23
- GALINDO, M.P. y CUADRAS, C.M. (1986): 'Una Extensión del Método Biplot y su Relación con Otras Técnicas'. *Public. de Bioestadística y Biomatemática*. Nº 17. Univ. de Barcelona.
- GOWER, J.C. (1990): 'Three-Dimensional Biplots'. *Biometrika*, 77: 773-785.
- I.N.E. (1991): *Indicadores Sociales*. Monografía Técnica.
- MOCHON, F. ; ANCOCHEA, G. (1981): *El Análisis de la Coyuntura. Una Metodología*. Ed. Pirámide.
- TSIANCO, M.C. y GABRIEL, K.R. (1984): 'Modeling Temperature Data: An illustration of the Use of Biplots in Nonlinear Modeling'. *Journal of Climate and Applied Meteorology*, 23: 787-799.