



I Jornadas de Análisis Input-Output. Oviedo, 22 y 23 de Septiembre de 2005

IDENTIFICACIÓN DE SECTORES CLAVE MEDIANTE ANÁLISIS DE CONGLOMERADOS DIFUSO. EL CASO DE ESPAÑA*

Bárbara Díaz

BISC, University of California at Berkeley

bdiaz@eecs.berkeley.edu

Laura Moniche

Dep. de Estructura Económica, Universidad de Málaga

lmoniche@uma.es

Antonio Morillas

Dep. de Estadística y Econometría, Universidad de Málaga

morillas@uma.es

RESUMEN

La búsqueda de los sectores clave de una economía ha sido y es uno de los temas más recurrentes del análisis input-output. Además de su liderazgo para impulsar el desarrollo, concepto demasiado amplio e impreciso, un sector puede ser clave desde una determinada perspectiva y menos o nada importante desde alguna otra diferente. También, probablemente, puede serlo para varias cuestiones a la vez, en distinto grado. En este trabajo se propone un enfoque multidimensional para clasificar a los sectores productivos de la tabla input-output española de 1995, basándose en tres grupos de variables: las relacionadas con su integración productiva, su peso específico en la economía y su dinámica económica. Además, se incorpora al análisis el nivel tecnológico, que por ser variable categórica plantea problemas metodológicos especiales. Todas estas cuestiones se abordan aplicando un análisis *cluster* robusto y difuso que arroja como resultado una clasificación de sectores ilustrativa del papel que juega cada uno de ellos en la economía española.

Palabras clave: sectores clave, cluster difuso, input-output, economía española

* Este trabajo ha sido financiado en parte por la Secretaría de Estado de Educación y Universidades de España y el Fondo Social Europeo, con la ayuda EX2003-0563.

1. INTRODUCCIÓN

La búsqueda de los sectores importantes de una economía comenzó, dentro del análisis input-output, de una forma muy elemental. Las relaciones directas de compra y venta entre sectores o la utilización del multiplicador basado en la inversa del modelo de Leontief no serían considerados en la actualidad criterios suficientes para identificar lo que se llamó en su momento sectores clave, por mucho que Hirschman (1958-61; pp. 110-114) considerase estos primeros trabajos de Chenery-Watanabe (1958) y Rasmussen (1956) como una forma de medir lo que vino en llamar eslabonamientos en su teoría del desarrollo. Es bastante probable que el debate sobre la noción de causalidad en economía (Simon, 1952; Wold, 1954) contribuyera a la idea de que el poder de arrastre de un sector, la jerarquía como criterio, era algo determinante en su relevancia económica.

Una de las primeras cuestiones controvertidas fue cuál debía ser la matriz de partida, si la clásica del modelo de Leontief o la propuesta por Ghosh (1958). Tal discusión ha quedado recogida en los trabajos de Augustinovics (1970), Cella (1984), Oosterhaven (1988) y Dietzenbacher (1997). Otros aspectos estudiados fueron los relacionados con el problema de la agregación (Hewings, 1974) o con la necesidad de normalizar y ponderar los multiplicadores (importancia relativa en el VAB, empleo o demanda final de las ramas). En Oosterhaven y Stelder (2002) y Oosterhaven (2004) se denomina multiplicador neto a los obtenidos en este último caso¹.

Otro enfoque para estudiar los sectores y/o coeficientes más importantes es el de extracción, iniciado por Strassert (1968), con la eliminación completa de una rama, fila y columna, de la tabla. Cella (1984) propondría uno menos drástico, con la posibilidad de distinguir los efectos hacia delante y hacia atrás. Otros, finalmente, utilizan matrices particionadas, como es el caso de Dietzenbacher (1992), Dietzenbacher, van der Linden y Steenge (1993), Sonis et al (1995), Dietzenbacher y van der Linden (1997) y Miller y Lahr (2000), o relacionan teoría de grafos y extracción (Aroche-Reyes, 2002).

En todos estos métodos subyace de alguna manera la idea de Hirschman sobre los eslabonamientos y la dialéctica equilibrio-desequilibrio en el crecimiento. Pero, sin duda, también, están relacionados con los conceptos de polarización, polos de

¹ Una buena y actualizada síntesis sobre sectores clave puede verse en Robles y Sanjuán (2005)

crecimiento, industrias motrices, de dominación y de asimetrías en las relaciones económicas, estudiadas por lo que podíamos llamar la escuela francesa, demasiadas veces olvidada en la literatura (Perroux, 1949 y 1973; Lantner, 1974; Huriot, 1974; Gazon, 1976).

Este interés por estudiar la jerarquía entre sectores se ha enfocado, también, posteriormente desde otras perspectivas tales como la triangulación² de la matriz interindustrial (Helmstäedter, 1957/58; Chenery y Watanabe, 1958; Aujac, 1960; Masson, 1960; Simpson y Tsukui, 1965, Korte y Oberhofer, 1970; Fukui, 1986), el análisis de la estructura causal recurrente por bloques o mediante algún procedimiento de descomposición (filtros, grafo reducido, esqueleto, puntos de articulación), generalmente próximo a la teoría de grafos (Lantner, 1972; Campbell, 1972 y 1975; Slater, 1977; Gilli, 1979; Morillas, 1981a, 1982 y 1983a; Hauknes, 1999). Los bloques resultantes, en su caso, pueden ser interpretados como *clusters* o grupos industriales.

En la mayor parte de estos trabajos se vislumbra la importancia del desarrollo de redes de empresas (sectores); es decir, el impulso del crecimiento ya no tanto desde la jerarquía de sectores impulsores autónomos, sino desde la interdependencia recursiva de conglomerados, bloques o grupos (*clusters*) de industrias. Los estudios sobre conglomerados industriales son casi tan numerosos como diversos en sus consideraciones acerca de lo que debe entenderse como tal. El trabajo pionero de Czamanski (1974) impulsado más tarde por el de Porter (1990), suscitarían un creciente interés por los *clusters* industriales desde distintas áreas de conocimiento.

El *clustering* industrial se centra, básicamente, en analizar las relaciones entre las ramas, aunque incorporen ocasionalmente otras variables (empleo, inversión, exportaciones,...) o traten aspectos tales como la innovación tecnológica (Schomookler, 1966; Verspagen, 1999; Scherer, 1982; Papaconstantino et al, 1996; Sakurai et al., 1997; Dietzenbacher y Los, 2000 y de Bresson y Hu, 1999) o el valor estratégico de las ramas en modelos de crecimiento (Los, 2001).

Las aplicaciones desarrolladas en el contexto del input-output han utilizado métodos de análisis muy diversos. Además de la ya referenciada teoría de grafos, se han empleado el llamado *método del máximo* (Peeters et al., 2001; Hoen, 2002; Roelandt et

² En Morillas (1981b), puede verse un resumen de estos métodos y su aplicación a las tablas input-output de Andalucía, Cataluña y País Vasco, con objeto de comparar sus estructuras productivas.

al., 1999; Oostenrhaven et al., 2000 y Broersma, 2001) y algunas técnicas estadísticas multivariantes.

A este respecto, una de las primeras aplicaciones del análisis multivariante al estudio de los grupos industriales a través de una TIO fue la realizada en el ya citado trabajo de Czamanski (1974). En él se realiza un análisis de componentes principales de la tabla input output de Estados Unidos del año 63, con la finalidad de agruparlos en función de sus perfiles de compras y ventas. Roepke et al. (1974) realizaron una aplicación similar sobre las tablas de Ontario de 1965. Latham (1976) y Harrigan (1982) criticaron estos métodos, por considerar que agrupaban tan solo las industrias ligadas verticalmente. En Ó hUallacháin (1984) se propone un análisis posterior de las componentes, a partir de la matriz de flujos interindustriales, para verificar la complementariedad de las industrias agrupadas, apoyando la pertinencia de estos métodos estadísticos. Feser y Bergman (2000) diferencian entre sectores fundamentales y secundarios para un grupo basándose en su peso factorial³, en una aplicación que trata de identificar los grupos industriales en el sector manufacturero de EEUU, utilizando la tabla de 1987 y los coeficientes de Czamanski. En posteriores trabajos Feser et al (2001a, 2001b) realizan un estudio de la concentración espacial del empleo en el seno de cada grupo⁴.

Rey y Mattheis (2000) elaboran una metodología de consenso para combinar los resultados de varias técnicas de componentes principales y de conglomerados e incorporan, por primera vez, un cluster no jerárquico para la determinación de grupos industriales⁵. Por su parte, Lainesse y Poussart (2005) hacen una propuesta denominada híbrida, inspirada en la anterior, que añade un análisis suplementario de indicadores a fin de incluir o excluir ciertos sectores en función de su complementariedad.

Nuestra aproximación es difícil de ubicar concretamente en una línea específica de las anteriormente apuntadas. Sin embargo, desde un punto de vista conceptual, es indudable que se nutre de todas ellas. No tomamos las ligazones entre sectores como base, tal y como haría un estudio de sectores clave o, incluso, de clusters industriales.

³ Podría decirse que éste es el primer antecedente de nuestra propuesta de utilización de *clustering* borroso con objeto de obtener grupos que no sean mutuamente excluyentes

⁴ Un buen resumen de estos trabajos puede verse en Lainesse y Poussart (2005)

⁵ Las aplicaciones realizadas anteriormente de esta técnica estadística al análisis input-output (Blin y Cohen, 1977; Cabrer et al.1991; Abbott y Andrews, 1990) están más bien interesadas por los problemas de agregación.

Pero tampoco las olvidamos, puesto que la incorporación de una medida cuantitativa sintética de las mismas, como puede ser el multiplicador y su dispersión, junto con otras cualitativas, como el grado de cohesión y el índice de interrelación, van a tomar en cuenta los aspectos relacionados con la interdependencia sectorial. Tampoco hacemos consideración alguna de aspectos espaciales o de la existencia de clientes o proveedores comunes, como es propio de este tipo de análisis. Sí consideramos, sin embargo, temas tratados en la literatura como la innovación, la producción, los salarios, el empleo o la vocación exportadora. Pero, además de observarlos estáticamente (nivel), le añadiremos una perspectiva dinámica no considerada hasta ahora, que sepamos, en la literatura. Por último, buscamos una visión estratégica, cara al desarrollo, en la identificación de los sectores importantes, concibiendo como criterios de caracterización aspectos tales como su capacidad innovadora, su potencial y dinámica exportadora o la generación de rentas salariales altas. Las aportaciones de este trabajo radican no sólo en estos aspectos que acabamos de mencionar sino, como se verá más adelante, en los procedimientos metodológicos empleados, basados en un análisis cluster difuso multivariante.

2. METODOLOGÍA Y FUENTES ESTADÍSTICAS

Tradicionalmente, en los métodos cluster, solo se podían considerar variables de tipo numérico⁶, es decir, medidas, al menos, en escala de intervalos. Para hacer posible el uso de variables ordinales y nominales es preciso utilizar lo que se denominan *clusters* relacionales, en los que la información suministrada como *input* representa una medida de disimilaridad entre los datos. Este tipo de datos son frecuentes en las ciencias sociales, dónde la información existente se basa, generalmente, en la relación entre los casos.

En nuestra propuesta para la clasificación de los sectores de la economía española, además de tomar en consideración este tipo de variables, utilizamos un enfoque alternativo a lo que podría llamarse clustering tradicional, el llamado clustering difuso o

⁶ El llamado two-step clustering (Zhang et al, 1996; Chiu et al, 2001), incorporado recientemente al paquete estadístico SPSS, presenta la posibilidad del uso de variables nominales. Sin embargo, utiliza como hipótesis de partida la de normalidad para las variables numéricas y la de distribución multinomial para las nominales. Esto lo hace, en nuestra opinión, menos adecuado para trabajar con la clase de variables presentes en este tipo de estudios (los perfiles de cualquier variable asociada a los sectores de una tabla input-output son fuertemente asimétricos y nada parecidos a una distribución normal), y en general, menos robusto que los métodos que no descansan en tales hipótesis teóricas.

borroso (*fuzzy*). La realidad es que en la mayoría de las ocasiones un elemento (sector) puede pertenecer a distintos grupos, aunque con diferentes grados de posibilidad, tal y como se ha subrayado en algunas de las referencias citadas anteriormente. En este sentido, el clustering difuso flexibiliza las agrupaciones efectuadas y otorga a cada elemento un grado de pertenencia a uno u otro cluster, lo que representa mayor flexibilidad y puede enriquecer los análisis posteriores.

El análisis de conglomerados difuso tiene sus orígenes en Dunn (1973), y fue posteriormente generalizado en Bezdek (1981), que introdujo el método de *fuzzy c means*. En Kaufman y Rousseeuw (1990), se utiliza una función objetivo en la que la disimilaridad o distancia es una norma L_1 (no está elevada al cuadrado), lo que la hace más robusta a la existencia de clusters no esféricos y a la existencia de *outliers* que en el método de *fuzzy c means* (Rousseeuw, 1995, p. 285). En Hathaway, Bezdek y Hu (2000) se presenta una generalización de *fuzzy c means* para el uso de L_p normas.

Existen en la literatura versiones relacionales de *fuzzy c means*. Inicialmente, en Hathaway et al (1989), se mostró que existía un problema dual relacional del *fuzzy c means* enfocado a objetos para el caso especial en que se cumple que los datos relacionales se corresponden con las distancias euclídea entre los elementos. Su método se denomina RFCM (*relational fuzzy c means*). Mas adelante, en Hathaway y Bezdek (1994) se introduce el NERFCM (*Non euclidean relational fuzzy clustering*). Su objetivo era modificar el RFCM para que fuera efectivo para datos relacionales arbitrarios que mostraran la disimilaridad entre casos u objetos. El método *Fuzzy Analysis Clustering* (Kaufman and Rousseeuw, 1990), puede utilizarse con ambos tipos de datos de entrada, dado que las distancias de la función objetivo son distancias entre objetos y no respecto a los centros de cluster. Dave y Sen (2002) proponen una versión robusta a outliers del *Fuzzy Analysis Clustering* llamada cluster difuso relacional no euclideo (*robust-NE-FRC*).

En nuestro caso, hemos utilizado el algoritmo NERFCM (Hathaway y Bezdek, 1994). Para su inicialización se ha tomado la información sobre los clusters suministrada por el método rígido de partición alrededor de los medoides (Kaufman y Rousseeuw, 1990), que fue el utilizado en Rey y Mattheis (2000). Este último es una versión robusta del conocido K-means (MacQueen, 1967) y es menos sensible a la presencia de *outliers*. Como datos de entrada se utiliza una matriz relacional de

disimilaridades entre sectores proporcionada por el método propuesto en Kaufman and Rousseeuw (1990). También hacemos uso del método de las anchuras de las siluetas, propuesto en Rousseeuw (1987), que no solo hace más robusta la formación de los grupos, sino que ayuda a seleccionar su número idóneo, cuestión siempre importante en el análisis de conglomerados.

Recordemos, finalmente, que en el análisis *cluster* la presencia de multicolinealidad entre variables provoca que las afectadas estén más representadas en la medida de similitud utilizada. Por esta razón, dado el carácter multidimensional de la propuesta que hacemos, utilizaremos como variables las resultantes de un análisis de componentes principales previo, realizado sobre la base de datos. Esto permite, por una parte, obviar el problema de la multicolinealidad en el cálculo de la matriz de similaridad y, por otra, una mejor identificación conceptual de los clusters resultantes, al reducirse el número de variables de clasificación.

Las variables utilizadas⁷ corresponden a tres grandes conceptos o bloques:

- Integración económica (4 variables), que intenta recoger tanto la intensidad como la cualidad de la interdependencia sectorial. Son, concretamente, el multiplicador derivado del grafo de influencia relativa⁸, su dispersión, medida a través del coeficiente de variación, el grado de cohesión⁹ y un índice topológico de integración¹⁰, obtenido para cada uno de los sectores. Estas dos últimas variables expresan aspectos cualitativos relacionados con la posición relativa de un sector dentro de la estructura de intercambios.

⁷ Las fuentes estadísticas y la elaboración de las variables utilizadas pueden verse en el Anexo III

⁸ Este multiplicador proviene del grafo de influencia relativa, asociado a la matriz de coeficientes de distribución, y se puede interpretar en términos de elasticidades (Lantner, 1974 y 2001; Mougeot et al., 1977; Morillas, 1983a, De Mesnard, 2001). Su expresión es $\hat{x}^{-1}\Delta x = (I - D)^{-1}\hat{e}\hat{y}^{-1}\Delta y$. Al ponderarse por la tasa de exportación fuera de la estructura del grafo (\hat{e} = coeficiente de la demanda final), podría decirse que está muy próximo al multiplicador neto del que habla Oosterhaven (2002 y 2004). La matriz interindustrial ha sido privada de los autoconsumos, para reforzar el análisis de interdependencias.

⁹ El concepto de cohesión en un modelo económico fue propuesto por Rossier (1980) y aplicado por primera vez al análisis input-output en Morillas (1983a)

¹⁰ Véase, Morillas (1983b). El índice de integración para un sector concreto, por extensión del indicador global que se propuso en este trabajo, siendo e_{ij} un elemento genérico de la matriz de distancias, vendría dado por la siguiente expresión:

$$R_i = \frac{1}{2n-1} \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{1}{e_{ij}} + \frac{1}{e_{ji}} \right)$$

- Peso específico (4 variables), que sintetiza la participación relativa de cada sector en la producción nacional, en las exportaciones y en la generación de rentas salariales y empleo.

- Potencial económico (5 variables), que recoge una visión dinámica de las anteriores, a través de las tendencias observadas, más la capacidad de innovación tecnológica, de acuerdo con la información al respecto suministrada por EUROSTAT (1998a, 1998b) y los indicadores de alta tecnología del INE (véase Anexo II).

La tabla de la economía española de 1995 es la base del estudio. El hecho de que sus 70 ramas iniciales hayan quedado reducidas a 66 en este trabajo, se debe a dos motivos. El primero es la inexistencia de datos sobre tecnología para las 3 ramas primeras, correspondientes al sector primario. El segundo, la eliminación de la rama 70, hogares que emplean personal doméstico, por carecer de relaciones con el resto de sectores.

3. LAS CLAVES DE LOS SECTORES DE LA TIO ESPAÑOLA DE 1995

Como se ha comentado anteriormente, para evitar posibles incidencias de la multicolinealidad sobre la matriz de similaridad, se ha realizado la extracción de las componentes principales subyacentes en la información suministrada por las 12 variables numéricas contempladas. Se ha encontrado, finalmente, en la solución rotada, que 3 componentes con valor propio superior a 1 explican algo más del 88% de la varianza (41%, 24% y 23%, respectivamente). Tomando en consideración sólo las cargas factoriales más importantes, la primera de esas componentes estaría formada por las variables relativas al multiplicador, a su medida de dispersión y la producción de los sectores. La denominaremos en lo sucesivo *Peso económico*. La segunda, contempla las dos medidas cualitativas, basadas en la teoría de grafos, ya comentadas: el índice de cohesión y el de integración productiva. La llamaremos *Integración*. Por último, la tercera, formada por las exportaciones y su tendencia, la vamos a catalogar como *Base exportadora*.

Con estas tres variables numéricas más la tecnología (ordinal) se ha llevado a cabo el análisis de conglomerados difuso¹¹ expuesto anteriormente. La mejor opción

¹¹ En realidad, el resultado de aplicar el algoritmo NERFCM es el que se recoge en el Anexo I. No hay asignación directa de cada elemento a un grupo concreto. Se dan sólo los valores de la función de

proporcionada por el gráfico de siluetas en el método de partición alrededor de los medoides, que posteriormente será utilizada como entrada en el algoritmo NERFCM, es la de 3 grupos (véase la ilustración 1), con 32, 23 y 11 sectores, respectivamente.

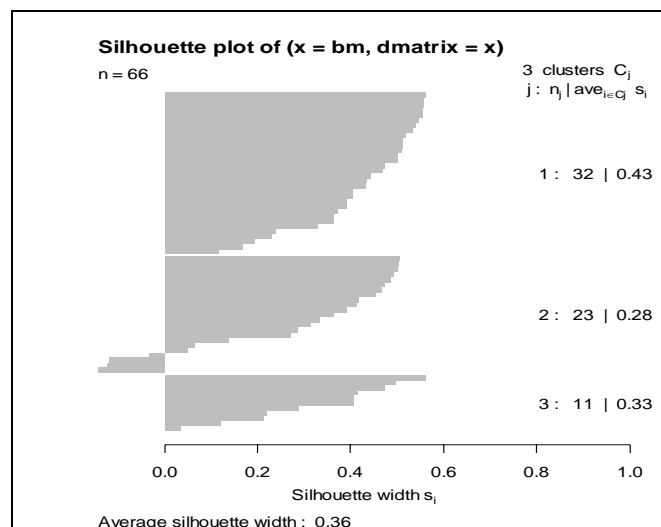


Ilustración 1: Siluetas para la mejor clasificación

El grupo con una silueta media más baja es el 2 (0,28), debido a la clasificación probablemente errónea (a la izquierda del cero del eje de abscisas) de cuatro sectores: los número 25, 1, 22, y 45, todos ellos con un valor de pertenencia al grupo 1 muy próximo al que presentan en el grupo 2. De cualquier forma, la anchura global de las siluetas ya nos está informando de que la estructura de clasificación no está claramente delimitada. Hay mucha borrosidad, lo que suscita la conveniencia de una agrupación difusa.

Los estadísticos que arrojaría el análisis para una clasificación rígida que tomara los valores más altos de la función de pertenencia arrojados por el procedimiento NERFCM para asignar los sectores a uno u otro grupo, se presentan en la tabla 1. Como puede observarse, el primer cluster tiene la media más alta en *Peso económico* y en *Integración*, siendo, sin embargo, el más bajo en cuanto a *Base exportadora*. Todos sus sectores tienen un nivel tecnológico bajo. Podría decirse, por tanto, que estos 32 sectores conforman la base local de la economía española. Entre ellos están los de más relevancia económica y los que presentan una mayor interdependencia. Esta es la clave para interpretar el papel que juegan los sectores agrupados en el cluster 1. La amplia

pertenencia a cada uno de ellos. Lo que hacemos aquí es una primera aproximación “rígida” en función de los correspondientes máximos. También el estudio de las siluetas es un proceso independiente.

dispersión que presenta la primera componente y el valor del máximo están indicando la posible presencia de valores atípicos (*outliers*) en esta variable.

Tabla 1: Centroides de los clusters

	<i>Estadísticos</i>	<i>PESO ECO</i>	<i>INTEGRA</i>	<i>BASE EXP</i>	<i>TECNO</i>
Cluster 1 N=32	<i>Media</i>	,1300528	,6478550	-,2797109	-
	<i>Desv. típica</i>	1,22042902	,39891890	,61633728	-
	<i>Mínimo</i>	-1,06797	,03722	-1,08783	1
	<i>Máximo</i>	5,33522	1,52491	1,52311	1
Cluster 2 N=23	<i>Media</i>	-,0787448	-1,0532104	-,1774113	-
	<i>Desv. típica</i>	,68528397	,82422585	,62557149	-
	<i>Mínimo</i>	-,82332	-2,24326	-,70716	1
	<i>Máximo</i>	1,57307	-,05051	1,37482	1
Cluster 3 N=11	<i>Media</i>	-,2136864	,3175027	1,1846545	-
	<i>Desv. típica</i>	,85495511	,71193397	1,61088763	-
	<i>Mínimo</i>	-,97006	-1,13500	-,52199	2
	<i>Máximo</i>	1,63377	1,56280	5,40637	3

El cluster 2 se distingue, fundamentalmente, por ser el menos integrado, teniendo un peso económico y una base exportadora inferiores a la media (por debajo de cero en la componente correspondiente). Su nivel tecnológico también es bajo. La clave para interpretar a este grupo de sectores es su falta de integración y escasa relevancia como impulsores de la economía, tanto por el pobre impulso económico que son capaces de generar como por su poca capacidad exportadora.

Finalmente, el tercer cluster tiene como característica principal la de ser, con diferencia, el más exportador y tener los mayores niveles tecnológicos. Su integración es superior a la media, pero su peso económico en la economía es el menos relevante. La clave correcta para interpretarlos, a nuestro entender, es que se trata de sectores exportadores de media y alta tecnología. Podría decirse que son los auténticos sectores clave desde la perspectiva de un crecimiento basado en la demanda externa y la innovación; es decir, de un escenario de desarrollo bastante deseable de la economía española. Puede observarse, también, la posible presencia de outliers en la tercera componente.

Un resumen del peso de cada grupo en la economía española, que aporta alguna pincelada más sobre los grupos formados, puede verse en la tabla 2.

El grupo 1, con más del 64% de la producción, es el que aporta más empleo y más salarios, siendo también importante su volumen de exportaciones. Pero todo ello en relación menos que proporcional al valor de su producción y siendo el porcentaje de

salarios inferior en 10 puntos al del empleo. Es decir, se trata de los sectores con salarios medios más bajos, como puede verse en la última columna.

Tabla 2: Participación de los clusters en la economía española

	N	% PRODUCCIÓN	% EXPORTACIONES	% EMPLEO	% SALARIOS	SALARIO MEDIO *
Cluster 1	32	64.2%	35.3%	55.9%	45.9%	95%
Cluster 2	23	22.8%	17.2%	37.6%	35.7%	102%
Cluster 3	11	13.0%	47.5%	6.5%	18.3%	110%
TOTAL RAMAS	66	100%	100%	100%	100%	100%

(*) Porcentaje del salario medio del cluster con respecto al salario medio total

El grupo 2, sin embargo, exceptuando las exportaciones, presenta unas cifras de empleo y de salarios más que proporcional al tamaño de su producción. Su productividad (aparente) media sería la más baja y su salario medio se sitúa por encima del anterior.

Por último, el grupo 3 representa tan sólo un 13% de la producción nacional, pero aporta más del 47% de las exportaciones. Doce puntos más que el grupo 1, cuyo número de sectores es tres veces superior. Además, su productividad y sus salarios medios son con diferencia los más elevados. Su media es un 10% superior a al salario medio total de la economía española.

En resumen, aún antes de pasar a comentar con más detalle las clasificaciones y los sectores asignados a cada grupo, podemos asegurar que, tanto desde la perspectiva de las exportaciones y de la tecnología, como de la productividad y de la generación de rentas salariales elevadas, el grupo 3 es con diferencia el de mejor comportamiento. Estamos en presencia, por tanto, de los sectores más importantes de la economía española para una estrategia de crecimiento basada en la innovación y la competitividad.

Pero la partición obtenida mediante el cluster difuso aporta más información de la que se acaba de dar mediante la clasificación de los sectores en el grupo al que pertenecen con un mayor grado de credibilidad. La flexibilidad en las clasificaciones permite afinar en el significado real de la asignación de un sector a un determinado grupo y caracterizarlo con mayor conocimiento de causa. En la tabla 3, con ánimo de simplificar la información sobre las funciones de pertenencia, se ha considerado que un valor igual o superior a 0,5 en un grupo dado es indicativo de que la clasificación del sector correspondiente es prioritaria, aunque no exclusiva, en ese grupo. En esta situación se encuentran 24 sectores del grupo 1, 15 del grupo 2 y 9 del grupo 3. Aún

aceptando este umbral, quiere esto decir que los 18 sectores restantes, mas del 27% de los sectores de la tabla pueden ser considerados como clasificables simultáneamente desde diferentes perspectivas. Este resultado tan razonable no se habría puesto en evidencia de haberse utilizado cualquier otra técnica de agrupación no difusa¹².

Tabla 3: Grupos y valores de la función de pertenencia

CLUSTER	Pertenencia ≥ 0.5			Pertenencia < 0.5			
	Ramas			Rama	Pertenencia	Cluster alternativo	Pertenencia
Cluster 1 Base local: - Peso económico - Integrados - Escasa base exportadora	5	18	43	58	0.4993	Cluster 2	0.3769
	6	19	46	48	0.4744	Cluster 2	0.4175
	7	21	51	23	0.4619	Cluster 2	0.3897
	8	27	52	42	0.4475	Cluster 2	0.4366
	14	35	55	4	0.4462	Cluster 2	0.4242
	15	38	56	44	0.4445	Cluster 2	0.3021
	16	39	57	41	0.4357	Cluster 2	0.3361
17	40	59	37	0.3786	Cluster 2	0.3739	
Cluster 2 Servicios, Consumo final: - Poco integrados - Escaso peso económico - Poca base exportadora	2	50		61	0.4965	Cluster 1	0.3106
	3	60		1	0.4781	Cluster 1	0.4037
	9	62		22	0.4743	Cluster 1	0.4090
	10	63		11	0.4727	Cluster 1	0.3276
	12	64		24	0.4614	Cluster 1	0.3723
	13	65		25	0.4522	Cluster 1	0.4322
36	66		26	0.4093	Cluster 1	0.3623	
49			45	0.3993	Cluster 1	0.3708	
Cluster 3 Base exportadora: - Tecnológico, exportador - Poco peso económico - Integración media alta	28	34		20	0.4600	Cluster 1	0.2942
	29	47				Cluster 2	0.2458
	30	53		33	0.3900	Cluster 2	0.3233
	31	54				Cluster 1	0.2867
	32						

Puede observarse que los cuatro sectores del grupo 2 a los que aludimos como mal clasificados en el gráfico de siluetas (1, 22, 25 y 45) tienen pertenencias muy próximas al grupo 1. Se trata de los sectores *Antracita, hulla, lignito y turba, Cemento, cal y yeso, Minerales no metálicos y Transporte aéreo y espacial*, que ocupan los lugares 42 a 45 en la componente de integración. Especialmente difusa es la clasificación de los sectores *Transporte aéreo y espacial* (45) y *Minerales no metálicos* (25), con pertenencias muy similares en los otros grupos. El primero de ellos presenta un buen comportamiento en la componente exportadora y el segundo en la de peso económico, especialmente por su tendencia reciente de crecimiento y su alto multiplicador.

¹² Sólo la graduación de la intensidad de las correlaciones proporcionada por el método de Czamanski o de las cargas factoriales aplicado en Feser y Bergman (2000), como se comentó anteriormente, podrían acercarse ligeramente a este resultado.

En el grupo 1, se han detectado otros 8 sectores de difícil clasificación, especialmente lo son los números 42, 4 y 37; es decir, *Transporte por ferrocarril*, *Minerales no metálicos ni energéticos* y *Construcción*. Los dos primeros no tienen un perfil claro en ninguna de las componentes, mientras que el tercero, siendo claramente un elemento del grupo 1, por ser el de mayor peso económico de todos los sectores, presenta serios problemas de clasificación por ocupar el lugar 40 en el ranking de integración.

Finalmente, en el grupo 3 hay dos sectores que podrían pertenecer también al 1 y al 2. Se trata de *Productos químicos* (2) y *Vehículos* (33), que, con tecnología intermedia, tienen una amplia componente exportadora, pero que, simultáneamente, ocupan los lugares 7 y 3, respectivamente, en *Peso económico*. Además, el primero es el mejor en *Integración* mientras que el segundo está muy poco integrado.

En definitiva, estos comentarios pueden extenderse al resto de sectores de difícil clasificación, poniéndose de relieve la visión poliédrica que muestra la realidad y el adecuado tratamiento que le otorga el *clustering* difuso. Parece evidente, por tanto, que este método tiene más sentido, ya que flexibiliza la clasificación rígida de los sectores que podría alcanzarse mediante un análisis de conglomerados tradicional. En la tabla 4 se presentan los que tienen una mayor posibilidad de pertenecer al grupo 3, que podrían ser considerados como los auténticos sectores clave de la economía española cara a un desarrollo con futuro.

Tabla 4: Sectores con mayor pertenencia al cluster 3

Rama	Nombre	Función de pertenencia		
		Cluster 1	Cluster 2	Cluster 3
29	Maquinaria de oficina y equipo informático	0,1648	0,1568	0,6784
53	Actividades informáticas	0,1669	0,1555	0,6776
47	Telecomunicaciones	0,1867	0,1736	0,6398
54	Investigación y desarrollo	0,1865	0,1769	0,6366
31	Material electrónico	0,1822	0,1963	0,6215
32	Instrumentos médicos y de precisión	0,2075	0,1800	0,6125
34	Otro material de transporte	0,2133	0,2001	0,5866
30	Maquinaria y material eléctrico	0,2247	0,2039	0,5714
28	Maquinaria y equipo mecánico	0,2524	0,2146	0,5330
20	Productos químicos	0,2942	0,2458	0,4600
33	Vehículos de motor	0,2867	0,3233	0,3900

La relación completa de sectores con sus correspondientes funciones de pertenencia puede verse en el Anexo I. Los sectores más claramente pertenecientes al grupo 1 son

los manufactureros (madera, corcho, vestido, edición, etc.) y algunos servicios recreativos y relacionados con el transporte, especialmente por su alto nivel de integración. Curiosamente, entre los sectores con mayor valor de la función de pertenencia en el grupo 1, los dos que tienen una menor posibilidad de pertenecer a este grupo son *Construcción* (37) y *Hostelería* (41). Sin embargo, ocupan el primer y segundo lugar en peso económico. La razón no puede ser otra más que el hecho de que son dos claros casos atípicos (*outliers*), tanto en la primera componente como en las variables originales que la forman.

En el grupo 2, están mejor clasificados sectores orientados hacia la demanda final, especialmente al consumo privado (industrias de alimentación, básicamente) y al público (sectores de servicios no de mercado). Entre los que tienen un valor de pertenencia más bajo están los ya comentados como mal clasificados en el gráfico de las siluetas más *Alimentación para animales* (11) y *Metalurgia* (26). Estos dos últimos ocupan, respectivamente, los lugares 4 y 6 de entre los de mayor peso económico y están entre los de mejor comportamiento exportador. Son, como en el caso anterior, valores atípicos en su grupo, en relación con las variables de exportación y de peso económico.

4. CONCLUSIONES

Los procedimientos para identificar los sectores o bloques de sectores importantes de una economía mediante el análisis input-output, a pesar del tiempo transcurrido desde sus inicios, continúan, mayoritariamente, centrándose en una u otra visión de la matriz interindustrial. Sin embargo, un sector importante no puede basarse únicamente en ese tipo de consideraciones. Ciertamente, en la literatura sobre grupos industriales se han ido introduciendo variables ajenas a dicha matriz, relacionadas con indicadores relevantes para evaluar el potencial económico del mismo (tecnología, empleo, salarios, etc.). Nos parece, sin embargo, que no se ha puesto demasiado énfasis en este tipo de enfoque multivariante. En este trabajo se ha adoptado este criterio como base para construir la información objeto de estudio, tomando variables pertenecientes a una serie de ámbitos que tienen que ver con la intensidad y estructura de las relaciones entre sectores, su fortaleza económica, incluido su nivel tecnológico, y su dinámica más reciente.

Por otro lado, como ponen de relieve Rey y Mattheis (2000) no se han aplicado otros métodos de *clustering* más que el jerárquico. Precisamente, la primera aplicación de un análisis de clusters no jerárquico se hizo en este trabajo. Lo que aquí aportamos es el uso por primera vez, que sepamos, de un algoritmo que, además de robusto, clasifica de forma borrosa o difusa, sin rigideces, permitiendo flexibilizar la interpretación de los grupos e informando de interesantes matices, como se ha puesto de manifiesto.

Aunque tal forma de clasificar no establece grupos rígidos, seleccionando los valores más altos de las funciones de pertenencia se ha asignado cada sector a los 3 grupos encontrados como solución más idónea. Tales grupos se han caracterizado adecuadamente. El primero, puede tildarse como el de sectores de base local (con mayores pesos económicos y más integrados). El segundo, formado por sectores con poca integración, orientados a la demanda interna, básicamente. El tercero, por último, se identifica por su orientación a la demanda externa y por su mayor nivel tecnológico. Una serie de consideraciones adicionales nos ha llevado a concluir que se trata del grupo de sectores más importantes para desarrollar una estrategia positiva de crecimiento de la economía española.

Desde el punto de vista metodológico, aparte de mejorar y ampliar la base de datos a otros aspectos relevantes, hay algunas cuestiones que convendría explorar en el futuro. Sabido es que en la literatura se hace referencia a que este tipo de análisis suele acabar en la definición de unos pocos *megaclusters*. Por otro lado, hemos detectado que la gran mayoría de los sectores con dificultades de clasificación son valores atípicos (*outliers*) en las distribuciones univariantes consideradas.

Cualquiera que haya trabajado con una tabla input-output sabe por experiencia que casi cualquier medida definida sobre las filas o columnas de la matriz interindustrial va a producir este tipo de resultados. Algo similar puede ocurrir con la información sectorial ajena a la tabla que se pudiera utilizar, pues los comportamientos de los diversos sectores suelen ser muy asimétricos. Extraña, pues, que en la literatura no se haya puesto en relación la aparición de estos *megaclusters* con la existencia de dichos valores atípicos. Más aún, sus efectos no afectan solo al análisis de conglomerados, sino que, sin duda, su presencia ha condicionado de alguna manera los resultados de todas las aplicaciones que se han realizado, desde el trabajo de Czamanski, con base en la extracción de componentes principales. A este nuestro, también.

La presencia de *outliers*, en definitiva, puede ser un problema realmente importante, tanto para el cálculo de estas componentes, por su incidencia sobre la matriz de correlaciones, como para la definición de los *clusters*. Siendo, sin duda, un problema serio, sin embargo, no ha sido explorado nunca, que sepamos, en este contexto del análisis input-output. Sin duda su tratamiento no es fácil, ya que no se puede prescindir de ellos en el análisis, pero es una línea de trabajo tan necesaria como prometedora.

BIBLIOGRAFÍA

- Abbott, T.A., Andrews, S.H. (1990): *The Classification of Manufacturing Industries : An Input Based Clustering of Activity*. Discussion Papers, U.S. Census Bureau. CES 90-7. Disponible en <http://www.ces.census.gov/paper.php?paper=100140&PHPSESSID=838fec06e9652892337cde662d45ed5a>
- Aroche-Reyes (2002): "Structural Transformations and Important Coefficients in the North American Economies". *Economic Systems Research*, Vol. 14, N° 3.
- Augustinovics, M. (1970): "Methods of international and intertemporal comparison of structure", *Contributions to Input-Output Analysis*, North Holland Publishing Company, Amsterdam-London, pp. 249-269.
- Aujac, H. (1960): "La Hierarchie des Industries dans un Tableau des Echanges Interindustriels", *Revue Economique*, 2, pp. 169-238.
- Bezdek, J.C. (1981): *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New York.
- Blin J., Cohen C. (1977): "Technological Similarity and Aggregation in Input-Output Systems". *The Review of Economics and Statistics*, vol. LIX no1, p.82-91.
- Broersma, L. (2001): "The Role of Services in Innovative Clusters". *Paper within the framework of the Research Programme Structural Information; Provision on Innovation in Services (SIID) for the Ministry of Economic Affairs, Directorate for General Technology Policy*. Disponible en: http://www.ggdc.net/pub/SIID_papers/innovative_clusters.pdf
- Cabrer, B., Contreras, D., Miravete, E.J. (1991): Aggregation in Input-Output Tables: How to Select the Best Cluster Linkage. *Economic Systems Research* 3 :99-110.
- Campbell (1972): "Growth pole theory, digraph analysis and the interindustry relation-ships". *Tijdschrift Voor Economie en Social Geografie*. 62, pp. 69-87.
- Campbell (1975): "Application of graph theoretic analysis to interindustry relation-ships". *Regional Science and Urban Economics*. 5, pp. 91-106.
- Cella, G. (1984): "The input-Output Measurement of Interindustry Linkages". *Oxford Bulletin of Economics and Statistics*, Vol. 46, n° 1, pp. 73-83.
- Chenery, H. B. y Watanabe, T. (1958): "International comparison of the structure of production", *Econometrica*, vol. XXVI, n° 26, pp. 487-521.
- Chiu, T., Fang, D., Chen, J., Wang, Y., Jeris, C. (2001): „A robust and scalable clustering algorithm for mixed type attributes in large database environment". *Proceedings of the seventh ACM SIGMOD international conference on knowledge discovery and data mining*, 263.
- Czamanski, S. (1974): *Study of Clustering of Industries*, Dalhousie University, Halifax.
- Dave, R.J., Sen, S. (2002): "Robust Fuzzy Clustering of Relational Data". *IEEE Transactions on Fuzzy Systems*, Vol, 10, num.6, 713-727.
- DeBresson, C. y X. Hu (1999): "Identifying Clusters of Innovative Activity: A New Approach and a Toolbox", in: OECD, *Boosting Innovation: The Cluster Approach*, Paris: OECD, 27-60.
- De Mesnard, L. (2001): "On Boolean Topological Methods of Structural Analysis", en Lahr, M.L. y Dietzenbacher, E. (eds.): *Input-Output Analysis: Frontiers and Extensions*, Palgrave, New York, cap. 3.
- Dietzenbacher, E. (1992): "The measurement of interindustry linkages: key sectors in the Netherlands". *Economic modeling*, 9, 419-437.

- Dietzenbacher, E. (1997). "In vindication of the Ghosh model: a reinterpretations as a price model", *Journal of Regional Science*, vol. 37, n° 4, pp. 629-651.
- Dietzenbacher, E. y Van der Linden, J. (1997): "Sectoral and spatial linkages in the EC production structure" *Journal Regional Science*, 37: 2, pp. 235-257.
- Dietzenbacher, E. y Los, B. (2000): "Analyzing R&D Multipliers". *Papers of the XIII International Conference on Input-Output Techniques*. University of Macerata, Italy, August 21-25th.
- Dietzenbacher, E., Van der Linden, J. Y Steenge, A.E. (1993): "The Regional extraction method: Ec Input-Output comparisons". *Economic Systems Research*, 5.
- Dunn, J. C. (1973): "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". *Journal of Cybernetics* 3(3), 32-57.
- EUROSTAT (1998a): *Ressources humaines en Haute Technologie. Serie 'Statistiques en Bref'*.
- EUROSTAT (1998b): *Employment in high technology manufacturing sectors at the regional level*
- Feser, E.J., Bergman, E.M. (2000): "National Industry Cluster Templates: a Framework for Applied Regional Cluster Analysis". *Regional Studies*, 34(1), 1-19.
- Feser, E.J., Koo, J., Renski, H. y Sweeney, S.H. (2001a): *Incorporating spatial analysis in applied industry cluster studies*. Working paper. Disponible en:<http://www.planning.unc.edu/pdf/FeserPUBS/EDQ%20Revised.pdf>
- Feser, E.J., Sweeney, S.H. y Renski, H.C. (2001b): *A Descriptive Analysis of Discrete U.S. Industrial Complexes*. Prepared for the Journal of Regional Science. Disponible en:<http://www.planning.unc.edu/pdf/FeserPUBS/Complexes.pdf>
- Fukui, Y. (1986): "A More Powerful Method for Triangularizing Input-Output Matrices and the Similarity of Production Structures", *Econometrica*, Vol. 54, No. 6, pp. 1425-1433.
- Gazon, J. (1976): *Transmisión de l'Influence Economique. Une Approche Structurele*. Sirey, Paris.
- Gilli, M (1979): *Etude et Analyse des Structures Causales dans les Modèles Economiques*. Ed. Peter Lang, Berna.
- Gosh, A. (1958): "Input-output approach in an allocation system", *Economica*, 25, pp. 58-64.
- Harrigan, F. (1982): The Relationship Between Industrial and Geographical Linkages. *Journal of Regional Science* 22 : 19-31.
- Hathaway, R.J., Bezdek, J.C. y Hu, Y. (2000): "Generalized Fuzzy c-Means Clustering Strategies Using Lp Norm Distances", *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, october 2000
- Hathaway, R.J. y Bezdek, J.C. (1994): NERF c-means: "Non- Euclidean Relational Fuzzy Clustering", *Pattern Recognition* 27, 3, 429-437.
- Hathaway, R.J., Davenport, J. W. y Bezdek, J.C. (1989): "Relational duals of the c-means algorithms", *Pattern Recognition*, 22, 205-212.
- Hauknes, J. (1999): "Norwegian Input-Output Clusters and Innovation Patterns", en Roelandt, T. den Hertog, P. (eds). *Boosting Innovation: The Cluster Approach*. Organisation for Economic Cooperation and Development, Paris.
- Helmstaedter, E. (1957/58): "Structure of Production and Growth", *Jahrbuches für Nationalökonomie and Statistik*, vol. 169.
- Hewings,, G.J.D. (1974): "The effect of aggregation on the empirical identification of key sectors in a regional economy: a partial evaluation of alternative techniques", *Environment and Planning*, n°6.
- Hirschman, A. O. (1958): *The Strategy of Economic Development*,. Primera versión en español: Fondo de Cultura Económica, 1961.
- Hoen, A. (2002): "Identifying Linkages with a Cluster-based Methodology". *Economic Systems Research*, Vol 14, n° 2. pp. 131-146.
- Huriot, J.M. (1974): *Dependence et Hiérarchie dans une Structure Interindustrielle*. Ed. Sirey. París.
- INE (2000): Tablas input-output de la Economía Española de 1995. www.ine.es
- INE. Contabilidad Nacional. Base 1995. www.ine.es
- INE. INEbase en la metodología expuesta en Investigación y desarrollo tecnológico. Indicadores de alta tecnología. Disponible en:<http://www.ine.es/inebase/cgi/um?M=%2Ft14%2Fp197&O=inebase&N=&L=>
- Kaufman, L., Rousseeuw, P.J. (1990): *Finding Groups in Data*. Wiley, New York.
- Korte, B. y Oberhofer, W. (1970). "Triangularizing Input-Output Matrices and the Structure of Production", *European Economic Review*, summer, pp. 482-511.
- Lainesse, L., y Poussart, B. (2005): « Methode de reperege des filieres industrielles sur le territoire quebecois dasee sur les tableaux d'entrees-sorties» , *L'economie du savoir. Institut de la Statistique du Quebec*. Disponible en http://www.stat.gouv.qc.ca/publications/savoir/methode_reperage_pdf.htm
- Lantner, R. (1972): "L'Analyse de la Dominance Economique", *Revue d'Economie Politique*, n° 2, pp. 216-283.

- Lantner, R. (1974): *Théorie de la Dominance Economique*. Paris. Dunod.
- Lantner, R. (2001): "Influence Graph Theory Applied to Structural Analysis", en Lahr, M.L. y Dietzenbacher, E. (eds.): *Input-Output Analysis: Frontiers and Extensions*, Palgrave, New York, cap. 15.
- Latham, W.R. (1976): "Needless Complexity in the Identification of Industrial Complexes". *Journal of Regional Science* 16 : 45-55.
- Los, B. (2001): "Identification of Strategic Industries: A Dynamic Perspective". *41st European Regional Science Association Congress* (Zagreb, August 2001). Disponible en: <http://www.ub.rug.nl/eldoc/ccso/200202/200202.pdf>
- MacQueen J. B. (1967): "Some Methods for Classification and Analysis of Multivariate Observations", *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*", Berkeley, University of California Press, 1:281-297.
- Masson, D. (1960): "Méthode de Triangulation du Tableau Européen des Echanges Interindustriels", *Revue Economique*, 2, pp. 239-257.
- Miller, R.E. y Lahr, M.L. (2000): "A taxonomy of Extractions", *Regional Science Perspectives in Economic Analysis; A Festschrift in Memory of Benjamin H. Stevens*. Elsevier Publishers, Amsterdam, pp. 407-441.
- Morillas, A. (1981a): *La teoría de grafos y el análisis estructural en los modelos económicos*. Tesis Doctoral. Universidad de Málaga.
- Morillas, A. (1981b): "Análisis de la especialización y desarticulación de una economía regional mediante la triangulación de la matriz interindustrial", *Estudios de Economía e Historia*. nº 1. 1981.
- Morillas, A. (1982): "Análisis estructural de modelos económicos. Aportaciones de la teoría de grafos." *Estadística Española*. Jul.-Sep., nº. 96, pp. 31-50.
- Morillas, A. (1983a): *La teoría de grafos en el análisis input-output*. Secretariado de Publicaciones. Universidad de Málaga.
- Morillas, A. (1983b): "Indicadores topológicos de las características estructurales de una tabla input-output. Aplicación a la economía andaluza". *Investigaciones Económicas*, nº 20, pp. 103-118.
- Mougeot, M., Duru, G. y Auray, J.P. (1977): *La Structure Productive Française*, Economica, Paris.
- Ó hUallachain, B. (1984): "The Identification of Industrial Complexes". *Annals of the Association of American Geographers* 74 :420-436.
- Oosterhaven, J. y Stelder, D. (2002): "Net Multipliers Avoid Exaggerating impacts: With a bi-regional illustration for the Dutch transportation sector". *Journal of Regional Science*, 3/42: 533-543.
- Oosterhaven, J. (1988): "On the plausibility of the supply-driven input-output model", *Journal of Regional Science*, vol. 28, pp. 203-217.
- Oosterhaven, J. (2004): "On the Definition of Key Sectors and the Stability of Net versus Gross Multipliers", University of Groningen, Research Institute SOM, number 04C01
- Oosterhaven, J., Eding, G. J. y Stelder, T. M. (2000): "Clusters, Linkages and Regional Spillovers: Methodology and Policy Implications for the Two Dutch Mainports and the Rural North", SOM Research Report 99c69, University of Groningen.
- Papaconstantinou, G. N. Sakurai and A Wyckow, (1996): "Embodied Technology Diffusion: An Empirical Analysis for 10 OECD Countries", *STI Working Papers* 1996/1, OECD, Paris.
- Peeters, L., Tiri, M. y Berwert, A. (2001): *Techno-economic Clusters in Flanders and Seitzerland: an Input-Output -Analysis*. Center for Science and Technology Studies. CEST 2001/9.
- Perroux, F. (1949): "Economic Spaces: Theory and Application". *Quarterly Journal of Economics*, 64, pp. 89-104.
- Perroux, F. (1973): *Pouvoir et Economie*, Bordas, Paris.
- Porter, M.E. (1990) *The Competitive Advantage of Nations*. Harvard Business Review, 68, 2, pp. 77-93.
- Rasmussen, N. P. (1956). *Studies in Intersectoral Relations*, North Holland Publishing Company, Amsterdam.
- Rey, S. y Mattheis, D. (2000) *Identifying Regional Industrial Clusters in California*. Volume I : Conceptual Design; Volume II : Methods Handbook; Volume III : Technical Documentation of Statewide Industry Clusters. Prepared for the California Employment Development Department. San Diego State University. Disponible en: <http://irsr.sdsu.edu/%7Eserge/node7.html>
- Robles Teigeiro, L. y Sanjuán, J. (2005): "Sectores claves. Coeficientes grandes y coeficientes importantes". I Jornadas de Análisis Input-Output. Oviedo 22 y 23. Septiembre 2005.
- Roelandt, T. J. A. den Hertog, P., van Sinderen y van den Hove. (1999): "Cluster Analiysis and Cluster Policy in The Netherlands", OECD, *Boosting Innovation: The Cluster Approach*, Paris. OECD, pp. 315-338.
- Roepke, H., Adams D., Wiseman, R. (1974): A New Approach to the Identification of Industrial Complexes Using Input-Output Data". *Journal of Regional Science* 14 : 15-29.

- Rossier, E. (1980): *Economie Structural*, Económica, Paris.
- Rousseeuw, P.J.(1987): Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* 20, 53-65.
- Rousseeuw, P.J. (1995): "Discussion: Fuzzy Clustering at the Intersection". *Technometrics*. Vol. 37, num. 3, 283-285.
- Sakurai, N. G. Papaconstantinou, E. Ioannidis, (1997): "Impact of R&D and Technology Diffusion on Productivity Growth: Empirical Evidence for 10 OECD Countries", *Economic Systems Research*, 9 (1), 81-110.
- Scherer, F.M. (1982): "Interindustry Technology Flows and Productivity Growth", *Review of Economics and Statistics*, 64, 627-634.
- Schmookler, J. (1966): *Invention and Economic Growth*, Cambridge MA: Harvard University Press.
- Simon, H.A. (1952): "On the Definition of the Causal Relation", *Journal of Philosophy*, vol. 49, pp. 517-528.
- Simpson, D. y Tsukui, J. (1965): "The Fundamental Structure of Input-Output Tables, An International Approach", *Review of Economics and Statistics*, 47, pp. 434-446.
- Slater, P. B. (1977): "The Determination of Functionally Integrated Industries in the United State Using a 1967 Interindustry Flow Table". *Empirical Economics*, 2, pp. 1-19.
- Sonis, M., Guilhot, J.J.M., Gewings, G.J.D. y Martins, E.B. (1995): "Linkages, Key Sectors, and Structural Change: some new perspective". *The Developing Economies*, XXXIII-3. pp. 233-270.
- Strassert, G. (1968): "Zur bestimmung strategischer sektoren Zur bestimmung strategischer sektoren", *Zur bestimmung strategischer sektoren*, 182
- Verspagen, B. (1999): "The Role of Large Multinationals in the Dutch Technology Infrastructure: A Patent Citation Analysis", mimeographed, University of Maastricht/MERIT.
- Wold, H. (1954): "Causality and Econometrics", *Econometrica*, vol. 22, pp. 162-177.
- Zhang, T., Ramakrishnon, R., Livny, M. (1996): BIRCH: "An efficient data clustering method for very large databases". *Proceedings of the ACM SIGMOD Conference on Management of Data*, 103-114, Montreal, Canada.

ANEXO I: Funciones de pertenencia a los clusters

Rama	Nombre	Cluster 1	Cluster 2	Cluster 3
1	Antracita, hulla, lignito y turba	0,4037	0,4781	0,1182
2	Petróleo crudo y Gas natural	0,2655	0,6019	0,1325
3	Minerales metálicos	0,2965	0,5274	0,1760
4	Minerales no metálicos ni energéticos	0,4462	0,4242	0,1296
5	Coque, refino de petróleo y nuclear	0,5292	0,3300	0,1409
6	Producción y distribución de electricidad	0,5287	0,3053	0,1661
7	Serv. de distribución de gas y agua	0,5590	0,3263	0,1147
8	Serv. de recogida y depuración de agua	0,5898	0,2872	0,1230
9	Carne	0,2985	0,5829	0,1186
10	Lácteos	0,2654	0,6260	0,1085
11	Alimentos para animales y otros	0,3276	0,4727	0,1997
12	Bebidas	0,2937	0,5825	0,1238
13	Tabaco	0,2874	0,5419	0,1707
14	Productos textiles	0,5861	0,2690	0,1448
15	Vestido y piel	0,6469	0,2398	0,1133
16	Cuero y calzado	0,5244	0,3170	0,1586
17	Madera y corcho	0,6588	0,2419	0,0993
18	Papel y cartón	0,6313	0,2450	0,1236
19	Edición y grabado	0,6462	0,2463	0,1075
20	Químicos	0,2942	0,2458	0,4600
21	Caucho y plástico	0,5718	0,2671	0,1611
22	Cemento, cal y yeso	0,4090	0,4743	0,1168
23	Vidrio	0,4619	0,3897	0,1485
24	Cerámica	0,3723	0,4614	0,1663
25	Minerales no metálicos	0,4322	0,4522	0,1157
26	Metalurgia	0,3623	0,4093	0,2283
27	Metálicos	0,5724	0,2816	0,1459
28	Maq. y equipo mecánico	0,2524	0,2146	0,5330
29	Maq. de oficina y equipo informático	0,1648	0,1568	0,6784
30	Maq. y material eléctrico	0,2247	0,2039	0,5714
31	Material electrónico	0,1822	0,1963	0,6215
32	Instrs. médicos, de precisión, óptica y relojería	0,2075	0,1800	0,6125
33	Vehículos	0,2867	0,3233	0,3900
34	Otro material de transporte	0,2133	0,2001	0,5866
35	Muebles y otros manufacturados n.c.o.p.	0,6374	0,2423	0,1203
36	Reciclaje	0,2630	0,6102	0,1268
37	Construcción	0,3786	0,3739	0,2475
38	Comercio y reparación de vehículos	0,5871	0,3132	0,0997
39	Comercio al por mayor e intermediarios	0,5003	0,3150	0,1846
40	Comercio al por menor	0,5019	0,3152	0,1829
41	Hostelería	0,4357	0,3361	0,2282
42	Transporte por ferrocarril	0,4475	0,4366	0,1158
43	Otro transporte terrestre	0,6225	0,2648	0,1126
44	Transporte marítimo	0,4445	0,3021	0,2534
45	Transporte aéreo y espacial	0,3708	0,3993	0,2299
46	Anexos al transporte	0,6356	0,2465	0,1179
47	Telecomunicaciones	0,1865	0,1769	0,6366
48	Intermediación financiera	0,4744	0,4175	0,1081
49	Seguros y planes de pensiones	0,3378	0,5608	0,1015
50	Auxiliares a la intermediación financiera	0,2992	0,5879	0,1129
51	Servicios inmobiliarios	0,5215	0,3301	0,1484
52	Servicios de alquiler	0,5460	0,3407	0,1133
53	Servicios de informática	0,1669	0,1555	0,6776
54	Servicios de I+D	0,1867	0,1736	0,6398
55	Otros servicios empresariales	0,5158	0,3260	0,1583
56	Serv. de educación de mercado	0,5760	0,2837	0,1403
57	Serv. sanitarios y sociales de mercado	0,5077	0,3112	0,1811
58	Serv. de saneamiento público de mercado	0,4993	0,3769	0,1237
59	Serv. recreativos y otros de mercado	0,6299	0,2622	0,1079
60	Otros serv. personales	0,2846	0,6068	0,1086
61	A.A.P.P., defensa y SS	0,3106	0,4965	0,1928
62	Serv. de educación de no mercado	0,2913	0,5441	0,1645
63	Serv. sanitarios y sociales de no mercado	0,3011	0,5220	0,1768
64	Saneamiento público de no mercado	0,2907	0,5392	0,1700
65	Sindicatos y otras asociaciones	0,2868	0,5437	0,1695
66	Serv. recreativos y otros de no mercado	0,2841	0,5536	0,1623

ANEXO II
Sectores de Alta y Media-Alta Tecnología.
Lista utilizada por el INE

CNAE-93	Sectores
Sectores manufactureros de tecnología alta	
244	Industria farmacéutica
30	Maquinaria de oficina y material informático
321	Componentes electrónicos
32-32.1	Aparatos de radio, TV y comunicaciones
33	Instrumentos médicos, de precisión, óptica y relojería
35.3	Construcción aeronáutica y espacial
Sectores manufactureros de tecnología media-alta	
24-24.4	Industria química excepto industria farmacéutica
29	Maquinaria y equipos
31	Maquinaria y aparatos eléctricos
34	Industria automóvil
35-35.3	Construcción naval, ferroviaria, de motocicletas y bicicletas y de otro material de transporte
Servicios de alta tecnología o de punta	
64	Correos y telecomunicaciones
72	Actividades informáticas
73	Investigación y desarrollo

FUENTE: INE. INEbase, documento metodológico pág. 5.

ANEXO III: Bloques de variables

	VARIABLE	DEFINICIÓN	COMENTARIO
Integración Económica	MULTIPLIC	<i>Multiplicador</i>	<i>Suma por columnas de la inversa de Ghosh, multiplicador de Augustinovic (1970).</i>
	CVINVERS	<i>Inversa del coeficiente de variación del los multiplicadores</i>	<i>Se caracteriza a cada rama según la inversa de la dispersión, en la inversa de Ghosh, representada por el coeficiente de variación.</i>
	COHESIO	<i>Grados de cohesión</i>	<i>Nº de veces que un sector sirve de camino</i>
	Ri	<i>Indicador de integración productiva total</i>	<i>Se ha calculado el índice de integración oferta-demanda para cada rama.</i>
Peso Específico	PRODUC	<i>Producción</i>	<i>Se ha utilizado la producción interior a precios básicos procedente de la tabla simétrica española.</i>
	EXPORT	<i>Exportaciones totales</i>	<i>La fuente es la tabla simétrica española para 1995</i>
	EMPLEO	<i>Empleo equivalente total</i>	<i>Al disponer solamente de esta cifra por ramas de actividad no homogéneas (tabla de destino a precios básicos), se le ha aplicado el coeficiente corrector que presenta el vector de remuneración de asalariados, por ramas de actividad.</i>
	SALARIOS	<i>Remuneración de asalariados media</i>	<i>Se estima mediante los datos de empleo asalariado, también corregidos, (procedente de la tabla de destino a precios básicos) y los datos de remuneración de asalariados (procedente de la tabla simétrica)</i>

Potencial Económico	TENPROD	<i>Tendencia en la producción</i>	<i>Es la tasa de variación real de la producción para cada rama para el periodo 1995-2000 (fuente: la Contabilidad Nacional del INE base 95).</i>
	TENEMPL	<i>Tendencia en el empleo</i>	<i>Es la tasa de variación del empleo equivalente de cada rama, con respecto al periodo inicial, desde el 1995 al 2000 (se ha utilizado el deflactor implícito del VABpb por ramas que presenta la Contabilidad Nacional del INE).</i>
	TENSALAR	<i>Tendencia en el salario medio</i>	<i>Es la tasa de variación real del salario medio para cada rama para el periodo 1995-2000 (se ha utilizado la variación del IPC durante ese periodo (14,2) para deflactar los salarios del año 2000)</i>
	TENEXPO	<i>Tendencia en las exportaciones</i>	<i>Es la tasa de variación real de las exportaciones para cada rama para el periodo 1995-2000, para deflactar se aplica el mismo procedimiento que para la producción</i>
	TECNO	<i>Capacidad de innovación tecnológica</i>	<i>Clasificación de sectores según su nivel tecnológico (alta, media o baja tecnología)</i>