Agregación de sectores en tablas Input-Output: una aplicación del análisis Cluster.

HERRI EKONOMIAZ
ECONOMÍA PÚBLICA

Bernardí Cabrer
Dulce Contreras
Eugenio Miravete

Departament d'Analisi Economica
Universitat de Valencia

Una versión algo más reducida de este trabajo se presentó en el XV Simposio de Análisis Económico (Barcelona, Diciembre de 1990). Los autores desean agradecer la información estadística facilitada por el centro L.R. Klein de la Universidad Autónoma de Madrid. Igualmente, estamos en deuda con los comentarios de un evaluador anónimo que nos han permitido mejorar esta última versión considerablemete.

Lan honetan aztertzen dugun oinarrizko arazoa, Input-Output Taulak, ekonomi eta estatistika ikuspegietatik, modurik egokienean agregatzea da. Horretarako, cluster egokiena aukeratu ahal izateko eredu bat erabili da, ekonomi berezkotasun behinenei eusten dien agregazio metodo moduan.

En este trabajo nos planteamos como problema básico la agregación de las Tablas Input-Output de la forma más adecuada desde el punto de vista económico y estadístico. Con este fin se ha aplicado un modelo que permite seleccionar el cluster más indicado como método de agregación que mantiene las propiedades económicas más relevantes.

1. INTRODUCCION

Los avances en Economía han demostrado que el desarrollo de la misma debe efectuarse con un planteamiento global, es decir fundamentando microeconómicamente los problemas macroeconómicos. El problema que surge en la unión entre ambos se conoce en la literatura como el problema de la agregación. El tratamiento habitual para la agregación de tablas se ha caracterizado por la agregación «ad hoc» de las mismas, sin dar mayor relevancia a los efectos que ésta podía producir sobre las características económicas de los sectores resultantes y su relación con las industrias que la componen.

Nuestro punto de partida es una tabla I-O de **m** industrias, donde se recoge la información estadística disponible para un momento determinado del tiempo a este nivel de desagregación. Para realizar un análisis macroeconómico sectorial, se impone habitualmente la necesidad de agregar estas industrias en un número inferior de sectores, para los que por ejemplo se disponga de otro tipo de información estadística, como evolución de precios, producción, etc. En este trabajo, se propone realizar la agregación mediante la aplicación del análisis cluster, pero no como la utilización de una mera técnica de agrupamiento, sino basándonos en un contraste que nos permite elegir la estrategia más adecuada de agregación, entendiendo por ésto, el hecho de que la agregación efectuada posea unas propiedades que implican el mantenimiento de las características de las funciones de producción a nivel industrial.

En definitiva, con el contraste que aplicamos lo que pretendemos es verificar cuál es el método de agregación más adecuado, lo que consiste en determinar si éste debe ser ponderado o no. Esta consideración no implica que estemos determinando el nivel óptimo de agregación, ya que éste debe establecerse dependiendo de la información sectorial disponible. Nuestro contraste se realiza mediante la especificación de un modelo autorregresivo que nos permite establecer la estategia más correcta a tenor de los valores que tomen los parámetros del mismo.

2. PROBLEMAS DE AGREGACION

La discusión de los fundamentos microeconómicos del análisis macroeconómico tiene ya una larga tradición en economía, y aunque quizá no nos haya proporcionado unas conclusiones definitivas al respecto, al menos ha servido para hacer constar lo complicado del tema. Creemos que es desde esta perspectiva como hay que abordar la cuestión de la agregación en las tablas Input-Output puesto que si se considera como un mero ejercicio de consolidación de industrias, éste carece de todo interés. Los argumentos en favor de la agregación versan en torno a la manejabilidad y disponibilidad de datos, y a la dificultad de trabajar con un número demasiado elevado de industrias, lo que puede inducir más bien a un análisis descriptivo del sistema económico que a facilitar una argumentación razonable de las características de la economía en cuestión porque la abundancia de información hace prácticamente imposible la propia comprensión de las mismas. Este argumento es irrelevante desde un punto de vista teórico, ya que procediendo de este modo podemos estar agrupando industrias que no producen mercancías homogéneas o sustitutivos cercanos. Importaría poco que en una categoría del modelo agregado estuviese recogido un conjunto heterogéneo de bienes si mantuviesen una proporción fija entre sí o bien si el sector que los agrupa reflejase perfectamente cuáles son las características de las industrias que incluye. Esto es, las columnas de una matriz de coeficientes recogen las características de una función de producción de cada una de las industrias, y al agruparlas, los sectores resultantes son la suma de un conjunto de industrias, sobre los que se definen de nuevo unas funciones de producción de coeficientes fijos, pero ahora a nivel sectorial. Los componentes de cada uno de estos sectores pueden decidirse arbitrariamente, o bien intentar agruparlos por el parecido de las industrias que lo componen, esto es, agrupando industrias con funciones de producción semejantes. Es evidente que si avanzamos excesivamente en el proceso de agregación, terminaremos agrupando industrias con funciones de producción «demasiado» diferentes.

Por tanto, de lo que se trata es de analizar hasta qué punto hemos de aceptar la «semejanza» entre dos industrias para que consideremos que la consolidación de las mismas en un único sector es aceptable. En el siguiente apartado analizaremos qué distorsiones produce la agregación de actividades distintas y desarrollaremos qué requisitos son necesarios para asegurar que la consolidación de varias industrias en un sector no produzca tales distorsiones.

3. AGREGACION EXACTA

Definimos $[X_{ij}]$ como la matriz cuadrada de orden **m** que representa el sistema de relaciones original. Los elementos que contempla esta matriz son **m** industrias. Por su parte, $[w_{ij}]$ describe este mismo sistema, una vez consolidado en **n** sectores. Esta última matriz es por tanto de orden **n < m.** La agregación consiste en sumar las filas y columnas de aquellos sectores de la matriz original $[X_{ij}]$, que se decida consolidar. Definimos una matriz T de dimensión (n x m) de forma que el paso de la matriz de transacciones original a la consolidada se expresa fácilmente como:

$$[w_{ij}] = T \cdot [X_{ij}] \cdot T' \qquad [1]$$

donde:

$$T = \begin{bmatrix} 1,1,...1; & 0,0,...0; & .... & 0,0,...0 \\ 0,0,...0; & 1,1,...1; & .... & 0,0,...0 \\ \multicolumn{4}{c}{....................................} \\ 0,0,...0; & 0,0,...0; & .... & 1,1,...1 \end{bmatrix} \qquad [2]$$

Esta matriz es un operador lineal que nos sirve para agregar las **n** industrias iniciales en **m** sectores. Obsérvese que en cada columna de la matriz T sólo existe un elemento distinto de cero y que en cada fila existe al menos un elemento distinto de cero. Esto ocurre porque cada industria sólo se va a consolidar en un sector, y porque toda industria debe consolidarse en alguno de ellos. El operador lineal T que se utilice en cada caso estará formado por unos y ceros dependiendo de cómo estén ordenadas las industrias en la matriz de transacciones original y de cómo se definan los sectores de la matriz consolidada. En nuestro caso, hemos supuesto además, que los elementos distintos de cero son la unidad, si bien no hay dificultad en considerar que pueden tomar valores distintos. Por tanto, se están agrupando los distintos sectores sin ponderar en absoluto la importancia que cada uno de ellos tiene sobre el total de transacciones que se realizan en la economía. Morimoto (1971) utiliza los precios de cada bien como ponderación, resultando de ello unas condiciones de agregación –en términos físicos– menos rígidas de las que vamos a obtener a continuación.

Sea $A = [a_{ij}]$ la matriz $(m \times m)$ de coeficientes técnicos originales, definidos sobre industrias, donde $a_{ij} = X_{ij} / X_j$. Si X^* representa una matriz diagonal cuyos elementos son las producciones totales de cada una de las industrias, entonces:

$$A = [a_{ij}] = [X_{ij}] \cdot [X^*]^{-1} \qquad [3]$$

Análogamente se definirían los coeficientes técnicos correspondientes al sistema consolidado y que se recogen en la matriz de coeficientes $B=[b_{ij}]$ de orden **n**. El problema de la agregación consiste en analizar en qué medida, los coeficientes de la matriz B, definidos sobre sectores, recogen las características de las funciones de producción de las industrias que se incluyen en cada sector y que representábamos por la matriz A.

Para obtener la condición que nos asegura que la agregación no produce ningún tipo de distorsión definimos Y como el vector $(m \times l)$ de demanda final y X el vector $(m \times l)$ de producción total, ambos referidos al sistema original. Los correspondientes al sistema agregado son los vectores **z**, **w**, de dimensión $(n \times l)$. Es evidente que en ambos sistemas se cumple:

$$A \cdot X + Y = X \qquad [4.a]$$

$$B \cdot w + z = w \qquad [4.b]$$

y por tanto:

$$Y = [I - A] \cdot X \qquad [5.a]$$

$$z = [I - B] \cdot w \qquad [5.b]$$

Si no se produce ningún tipo de distorsión en la agregación, la suma de la producción total o de la demanda final de todos los sectores debe ser la misma tanto si la contabilizamos en el modelo original como en el agregado. Dado que cada sector del modelo agregado es la suma de uno o varios sectores de la tabla original, entonces deberá cumplirse que (1):

$$T \cdot Y = z \qquad [6.a]$$

$$T \cdot X = w \qquad [6.b]$$

De aquí, la condición de Hatanaka se obtiene directamente:

$$T \cdot Y = T \cdot [I - A] \cdot X = [I - B] \cdot w = z$$

$$T \cdot I \cdot X - T \cdot A \cdot X = w - B \cdot w$$

de donde:

$$T \cdot A \cdot X = B \cdot w = B \cdot T \cdot X$$

y por tanto:

$$T \cdot A = B \cdot T \qquad [7]$$

Los primeros trabajos que abordan el tema [Hatanaka (1952), Kenjiro Ara (1959), Malinvaud (1954)], plantean qué condiciones deben cumplir dos industrias para que una vez agregadas en sectores, variaciones en el vector de demanda final o de producción total (evidentemente del sistema agregado) no modifiquen los coeficientes del sistema agregado, definido sobre sectores, restricciones éstas, que se conoce como condición de Hatanaka. Que se cumpla la condición de agregación de Hatanaka significa que dos o más industrias pueden agregarse en un sector si tienen una estructura de inputs homogénea (2), esto es, que la función de costes sea similar en cada industria, que es lo mismo que considerar similares sus funciones de producción dado que el precio de los bienes y de los factores de producción son, por hipótesis iguales para toda la economía. Así es como se establece la conexión entre los sectores, categorías macroeconómicas, y las industrias, categorías microeconómicas.

Este enfoque se encuadra en lo que se conoce como agregación exacta. Bajo esta perspectiva, la agregación o es correcta o no se pueden agrupar industrias de acuerdo con la matriz T diseñada. Si la condición de Hatanaka no se cumple, la única solución sería redefinir los sectores y encontrar otra matriz de agregación que no generase sesgos. La necesidad de trabajar con tablas agregadas facilita que inspirándose en este análisis se desarrollen métodos de agregación no exacta que en general exigen cierta estructura o comportamiento del vector de demanda. Si la agregación no es exacta, entonces

(1) Véase por ejemplo Kossov (1972).

(2) Ara (1959) desarrolla una condición más general, necesaria y suficiente para que $T \cdot A = B \cdot T$, que se resume en que la suma de las columnas de coeficientes de la matriz original A de las industrias que se van a consolidar en B, sumen lo mismo.

se intenta minimizar el sesgo de primer orden (3), pero la mayoría de métodos toman el esquema de agregación como dado. Estos métodos sólo proporcionan medidas aproximadas para hacer que la agregación sea lo más parecida posible a la agregación exacta, sin establecer un criterio objetivo que nos permita discernir cuál de entre dos agregaciones no exactas es preferible para el análisis.

4. APLICACION DEL ANALISIS CLUSTER

Entre los métodos de agregación no exacta, el análisis cluster agrega industrias en sectores dependiendo del grado de parecido de las características de las mismas, esto es, las de sus funciones de producción. Pero la definición de los sectores de esta agregación no está prefijado, sino que se genera de forma endógena en el propio proceso de realización del cluster.

El objetivo del análisis cluster, como es bien sabido, consiste básicamente en que dado un conjunto de individuos caracterizados por un conjunto de variables (en nuestro caso industrias caracterizadas por sus funciones de producción), se trata de clasificarlos en grupos en base a su parecido. La diferencia fundamental del análisis cluster respecto a otras técnicas de diseño clasificadoras, radica en que el análisis cluster parte de los individuos mientras que las otras técnicas necesitan partir de grupos de observaciones una vez se han establecido unos criterios previos de agrupación (4).

La forma de cuantificar el parecido entre los distintos individuos (para nosotros funciones de producción) es mediante alguna medida de distancia. Las industrias se caracterizan a través de un vector de inputs de todas las industrias que se contemplan en la T.I.O. Es de esperar que industrias pertenecientes a sectores económicos bien definidos dispongan de una función de producción más parecida que aquellas que se dedican a actividades de otra naturaleza económica. Se considera, por tanto, que las funciones de producción de dos industrias son más parecidas cuanto menor es la distancia que existe entre los correspondientes vectores de requisitos unitarios de inputs que caracterizan las funciones de producción de coeficientes fijos de estas industrias. La medida de distancia que hemos seleccionado es la Euclídea debido a su sencillez y facilidad de interpretación, así como por las propiedades que reúne ya que no es invariante al cambio de escala ni a la interdependencia de las características o variables. En la realización del cluster se van a tener en cuenta todas las variables que definen a cada una de las funciones de producción para calcular la distancia entre las distintas industrias y conjunto de industrias.

Una vez estudiada la distancia entre los individuos es necesario precisar la forma de calcular la distancia entre grupo-individuo y grupo-grupo. Una vez se han agrupado dos industrias, este grupo tiene definido también un vector de características que no coincide exáctamente con ninguno de los vectores de características de las industrias que lo forman. El problema que se plantea ahora es si a partir de este punto, cuando vamos a agrupar una industria con un grupo de industrias previamente formado, hemos de considerar de igual modo a la industria y al grupo o si debemos ponderar por el

(3) El sesgo de primer orden se define como: $w - T \cdot X = ([I-B]^{-1} \cdot T - T \cdot [I - A]^{-1}) \cdot Y$. A título informativo véase Fisher (1962), Kymn y Norsworthy (1976), Neudecker (1970) y el artículo original de Theil (1957).

(4) Sobre el análisis cluster y sus aplicaciones a este campo de la economía puede consultarse el libro de Fisher (1969)

tamaño del grupo de industrias. En esto consiste la elección de estrategia («linkage») en el agrupamiento del cluster. Existen distintos criterios de cálculo, así por ejemplo, con el «simple linkage», la distancia entre una industria y un grupo se calcula como la distancia entre la industria y el centro de gravedad del grupo, mientras que la distancia entre dos grupos es la comprendida entre sus centros de gravedad.

Dentro del conjunto de técnicas de análisis cluster existe un subconjunto caracterizado por métodos jerárquicos agregativos que tienen por objetivo agrupar clusters para formar uno nuevo, y así, sucesivamente se va efectuando esta agrupación, de forma que se sigue el procedimiento de minimizar la medida de distancia entre clusters (5). La idea básica consiste en lo siguiente: se parte de un conjunto de I-clusters, cada uno de ellos formado por un único individuo al que se le denomina nivel $N = 0$. En el siguiente nivel $N = 1$, se agrupan los dos clusters que son más parecidos. Si se continúan agrupando los distintos clusters por este procedimiento se llega al nivel $N = I - 1$, en el que sólo hay un cluster formado por todos los individuos. Cada cluster jerárquico tiene asociado un dendograma, que muestra de qué forma se van agrupando los clusters, y además, una medida de distancia para la que se agregan cada uno de los niveles.

La aplicación de los métodos jerárquicos aglomerativos es muy sencilla ya que partiendo de tantos clusters como individuos se elige una medida de similitud (que en nuestro caso es la distancia euclídea) y se van agrupando los clusters más parecidos. Una vez agrupados dos clusters es preciso recalcular las nuevas medidas de similitud entre el centro de gravedad de este grupo y el de otros grupos o individuos. Dependiendo de la estrategia utilizada («linkage»), las industrias del modelo original se agruparán de un modo u otro, resultando en sectores con distinta composición.

Además, el análisis cluster que llevemos a cabo puede poseer las propiedades de compatibilidad y combinatoriedad [Sneath y Sokal, 1973]. Al hablar de compatibilidad y combinatoriedad hacemos referencia a las siguientes condiciones:

1. *Combinatoriedad (C.1):* Sea d^2_{ik} la distancia entre los clusters (i) y (k). Si estos son los que tienen la menor distancia, se obtiene un nuevo cluster (ik), con tamaño $n_i + n_k$ (suma de los individuos de cada grupo). El siguiente paso es agregar un nuevo grupo (ikh) con los grupos anteriores (ik) y (h), que son los que exhiben la menor distancia (mayor similitud). Si $d^2_{(ik)h}$ puede expresarse estrictamente con d^2_{ik}, d^2_{ih}, d^2_{kh}, entonces esta estrategia posee esta propiedad. En el caso que analizamos, (distancia euclídea) esta condición se expresa como:

$$d^2_{(ik)h} = \alpha_i\, d^2_{ih} + \alpha_k\, d^2_{kh} + \beta\, d^2_{ik} + \tau \mid d^2_{ih} - d^2_{kh} \mid \qquad [8]$$

donde α_i, α_k, β y τ son los parámetros que definen la estrategia utilizada.

2. *Compatibilidad (C.2):* Esta propiedad se cumple cuando en [8] $\tau = 0$, y además:

(5) Véase Escudero (1977).

$$\alpha_i + \alpha_k + \beta = 1 \qquad [9]$$

que significa que las distancias entre diferentes grupos tienen las mismas características. Es decir, los valores de las distancias se van haciendo mayores uniformemente según el grado de agregación. Esto implica que todas tienen la misma dimensionalidad, y que pueden ser estudiadas por un mismo modelo [Escudero (1977)].

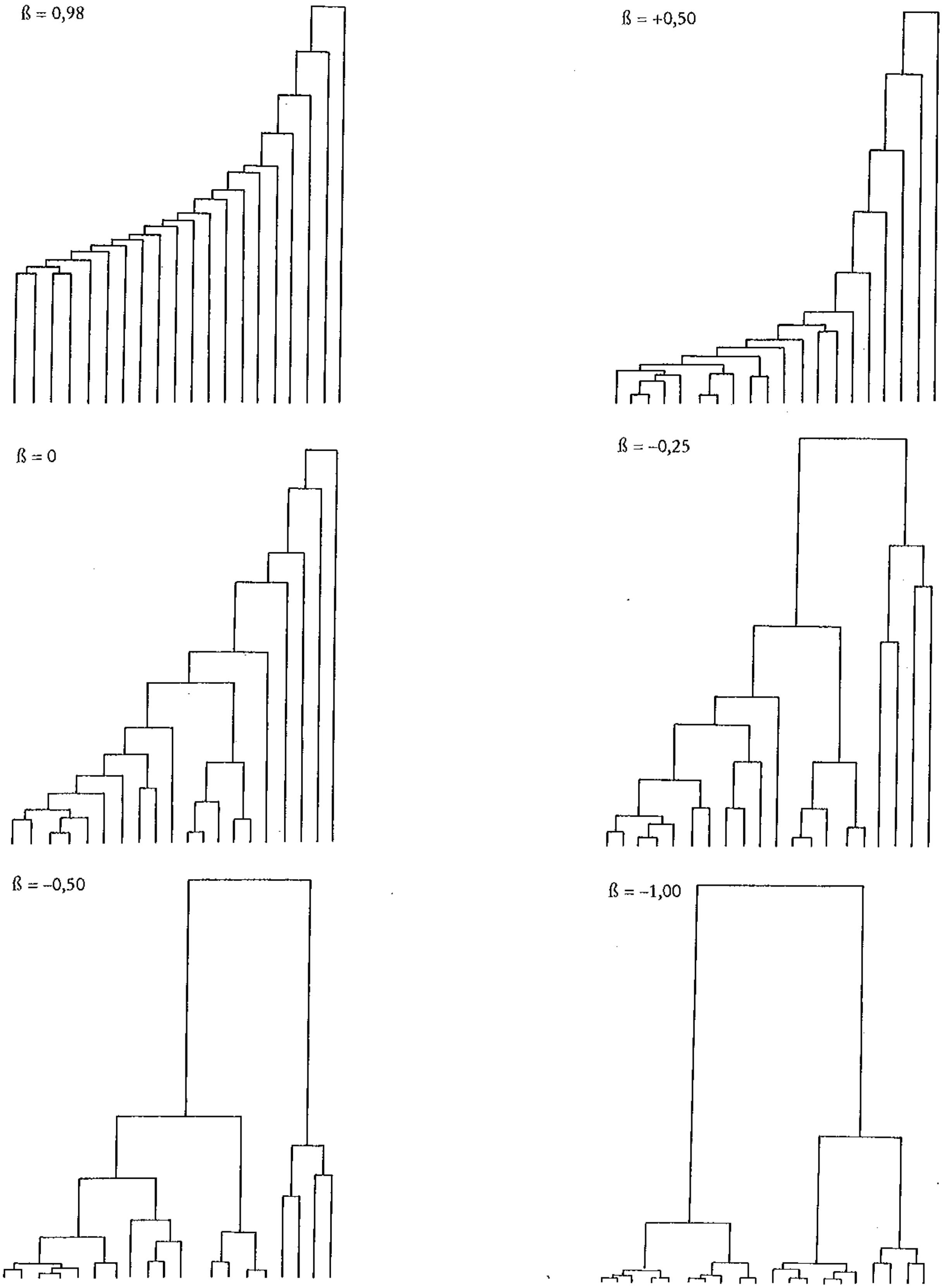
En el Cuadro 1, se puede comparar de forma heurística las diferentes estrategias que se pueden utilizar con la distancia euclídea, según los valores que se obtengan de los parámetros.

CUADRO 1

LINKAGE	α_i	α_k	β	τ	C.1	C.2
Simple Linkage ...	1/2	1/2	0	–1/2	SI	SI
Complete Linkage	1/2	1/2	0	1/2	SI	SI
Centroid	$n_i/(n_i+n_k)$	$n_k/(n_i+n_k)$	$-\alpha_i\,\alpha_k$	0	SI	NO
Median	1/2	1/2	$-\alpha_i\,\alpha_k = -1/4$	0	SI	NO
Group Average ..	$n_i/(n_i+n_k)$	$n_k/(n_i+n_k)$	0	0	SI	SI
Weighted Arithmetic Average	1/2	1/2	0	0	SI	SI
Ward	$\dfrac{(n_h+n_i+n_k)n_i}{n_h(n_i+n_k)^2}$	$\dfrac{(n_h+n_i+n_k)n_k}{n_h(n_i+n_k)^2}$	$-\alpha_i\,\alpha_k$	0	SI	NO

Al aplicar el análisis cluster se utilizan distintas estrategias, pero no se proporciona habitualmente ninguna razón que justifique el uso de una en concreto. Si examinamos la Figura 1, observando los dendogramas, podemos verificar cómo los mismos individuos son agrupados de diferente forma dependiendo de la estrategia utilizada. Además, los clusters pueden poseer o no las propiedades mencionadas, y esto depende de la estrategia que se haya utilizado. Este punto es lo que convierte la elección de la estrategia a utilizar en un elemento crucial del análisis cluster aplicado a la agregación de Tablas Input-Output, puesto que en ellas hay industrias que son absolutamente distintas y que por tanto no pueden ser agregadas en el mismo grupo en cualquier etapa del cluster, sólo en las últimas agrupaciones, cuando se unen los sectores más dispares. En el análisis que efectuamos, exigimos que el cluster satisfaga la propiedad de combinatoriedad, pero no la de compatibilidad, ya que en este caso, en cada etapa del cluster una o más industrias estarían agrupadas en el mismo grupo, lo que quiere decir que aunque inicialmente las industrias fueran distintas, se agruparían en un único sector

FIGURA 1
Agrupación según diferentes estrategias



Fuente: Escudero, L.F. (1977).

homogéneo, y por tanto toda la economía podría ser explicada por el mismo modelo, es decir, una función de coste o una función de producción agregada. De este modo se obtendría una agregación perfectamente homogénea, lo que no es compatible con el análisis multisectorial que aborda el modelo input-output, que precisamente trata de recoger las diferencias entre las distintas ramas de actividad económica.

Por lo tanto, y este es un punto importante, debemos utilizar aquella estrategia que genere una agrupación de industrias de modo que tenga una justificación económica en el sentido que aquí se ha expuesto, esto es, que se agrupen los sectores técnicamente más parecidos, pero que la forma en que se agrupen no venga explicada por un único modelo, es decir, una única función de producción a nivel agregado que estaría detrás de las de cada uno de los sectores. En la Figura 1 se contempla cómo la agregación tiene lugar de este modo para clusters con un valor de $\beta < 0$.

Observando el Cuadro 1, se comprueba que hay tres estrategias que cumplen este requisito: Centroid (unweighted pair group centroid), Median (weighted pair group centroid), y el método de Ward [Anderberg (1973)]. Estas estrategias cumplen la propiedad de combinatoriedad, pero no la de compatibilidad. Vamos a estudiar cuál de ellas es la más adecuada para un caso particular, y propondremos para ello un método para discriminar entre estas estrategias.

5. ALGUNAS CONDICIONES PARA ELEGIR LA ESTRATEGIA DEL CLUSTER

El modelo que estimamos se basa en la ecuación [8], que se refiere al nivel de agregación N-1. La ecuación [10] representa la misma propiedad de combinatoriedad para el nivel de agregación N.

$$d^2_{(ikh)j} = \alpha_{ik}\, d^2_{(ik)j} + \alpha_h\, d^2_{hj} + \beta d^2_{(ik)h} + \tau \,|\, d^2_{(ik)j} - d^2_{hj} \,| \qquad [10]$$

Utilizando el análisis paramétrico del Cuadro 1, podemos afirmar que se debe cumplir que:

$$\tau = 0 \qquad [11.a]$$

$$-\alpha_{ik}\,\alpha_h = -\alpha_i\,\alpha_k = \beta < 0 \qquad [11.b]$$

Para el Método de Ward, la condición [11.b] sólo se cumple cuando N es suficientemente grande. Por lo tanto, en este caso particular, el test que proponemos será asintóticamente equivalente a los que apliquemos sobre los clusters realizados con los otros dos métodos. Por tanto:

$$d^2_{(ikh)j} - d^2_{(ik)h} = \alpha_{ik}\, d^2_{(ik)j} + \alpha_h\, d^2_{hj} - \alpha_i\, d^2_{ih} - \alpha_k\, d^2_{kh} + \beta \cdot [d^2_{(ik)h} - d^2_{ik}] \qquad [12]$$

Definiendo Y_N como la distancia en el nivel de agregación N y utilizando las ecuaciones [11.a] y [11.b], podemos especificar la ecuación [12] como:

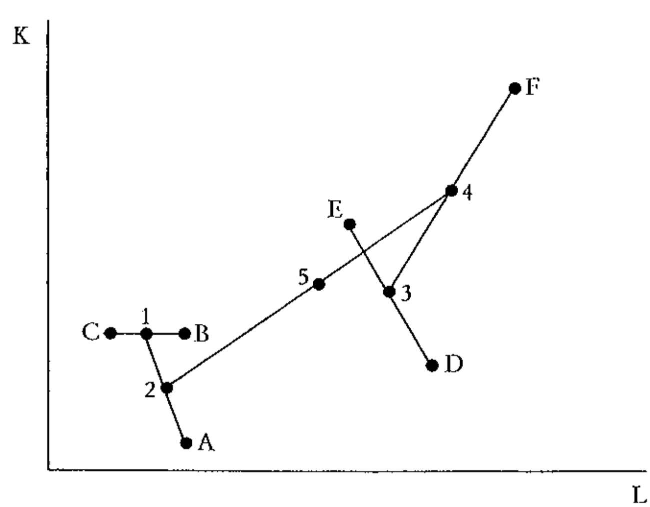
$$[Y_N - Y_{N-1}] = \delta + \beta \cdot [Y_{N-1} - Y_{N-2}] \qquad [13]$$

Habitualmente, en la literatura sobre el tema, no existe una justificación racional para elegir una estrategia u otra para realizar el análisis cluster en la agregación de industrias en sectores (6). Las condiciones que ahora presentamos, basadas en las propiedades que cumple cada estrategia, nos permiten discriminar en algunos casos entre las que pueden utilizarse y las que no. Podemos encontrarnos con tres posibles situaciones:

a) Si obtenemos un valor positivo de β con alguna de las estrategias aplicadas, esto implica que dicha estrategia no puede ser empleada correctamente.
b) Si por el contrario, los valores estimados de β son negativos, entonces el método del Centroide o el de Ward pueden emplearse, si bien, para utilizar el método de la Mediana, es necesario que β tome un valor próximo a (–1/4) (debemos poder aceptar la hipótesis nula de que $\beta = -1/4$).
c) Si el valor estimado de β nos permite utilizar más de una estrategia (con un valor negativo distinto de (–1/4) podríamos utilizar tanto el método del Centroide como el de Ward), entonces proponemos que en tal caso se utilice un criterio de información para discriminar entre ellas.

A partir de la Figura 2 vamos a intentar exponer brevemente de forma intuitiva la construcción de este contraste. Supongamos que disponemos de

FIGURA 2



(6) Véase, por ejemplo, Blin y Cohen (1977).

un conjunto de seis industrias, caracterizadas por sus necesidades de capital y trabajo por unidad de output. Estas industrias, y sus características se representan por los vectores: A (5,1); B (5,5); C (2,5); D (14,4); E (11,9); F (17,14). Si debemos agrupar estas industrias, nuestro método lo realizaría del siguiente modo: en primer lugar se agruparían las industrias B y C puesto que sus vectores son los que presentan una menor distancia ($d_1 = 3$), dando lugar al sector 1, que siguiendo la estrategia de la mediana vendría caracterizado por unas necesidades de capital y trabajo que son la media de B y C. Si el proceso continúa, a este grupo de industrias se le agregaría la industria A dado que la distancia entre A y 1 es la menor de todas las que pueden establecerse ($d_2 = 4.272$). En un nivel de agregación posterior se agruparían las industrias D y E ($d_3 = 5.83$). Si detuviésemos la agregación en este nivel, habríamos reducido las seis industrias originales a tres sectores: uno, formado por las industrias A, B, y C, otro, formado por las industrias D y E, y otro formado exclusivamente por la industria F. Dado que el análisis cluster no determina cuándo hay que detener este proceso de agregación, si se prosigue, se agrupará en el siguiente nivel, la industria F con el sector formado por D y E, representado por el vector 3 ($d_4 = 8.746$). En la etapa final de agregación, se agruparían los dos grupos de industrias representados por los vectores 2 y 4 ($d_5 = 12.75$), formando un cluster que agruparía a todas las industrias originales.

Obsérvese que utilizando un cluster jerárquico, cada vez que se agregan industrias se obtiene una medida de distancia. Estas medidas están ordenadas de acuerdo con el nivel de agregación en que se realiza cada agrupación. Utilizando estas distancias, estimaríamos el modelo autorregresivo dado por la ecuación [13] para comprobar, en el caso de nuestro ejemplo, si se puede aceptar que $\beta = -0.25$ de modo que el método de la mediana se pueda admitir como adecuado.

Nuestro modo de proceder sería el siguiente: se agruparían las industrias originales utilizando un cluster jerárquico y la distancia euclídea como medida de similitud, pero haciendo uso en cada caso de una estrategia distinta. Con cada una de las series de distancias obtenidas, se estimaría la ecuación [13], comprobando por el valor estimado de β si cada una de las estrategias es aceptable o no (7).

Aplicamos este test a la Tabla Input-Output de la Economía Valenciana de 1980, formada por 50 sectores [ver Apéndice 1]. Llevamos a cabo seis cluster jerárquicos (utilizando la distancia euclídea) sobre la Inversa de

(7) Dada la no estacionariedad en varianza que presentan los datos que utilizamos tomamos los logaritmos de los mismos. Esta situación, se presenta siempre que utilizamos métodos jerárquicos agregativos, ya que cada vez que se agrega un individuo, la medida de la distancia de este individuo será mayor que la de los agregados previamente, que eran más parecidos entre sí. Esto nos lleva a trabajar de forma general con las diferencias de los logaritmos de las distancias:

$$[\log Y_N - \log Y_{N-1}] = \delta + \beta \cdot [\log Y_{N-1} - \log Y_{N-2}] \qquad [14]$$

CUADRO 2

LINKAGE	MATRIX	Valor Teórico	Valor Estimado	Error Std.
CENTROID ..	$(I-A)^{-1}$	$\beta < 0$	–0,034	0,235
CENTROID ..	A	$\beta < 0$	–0,021	0,151
MEDIAN	$(I-A)^{-1}$	$\beta = -1/4$	–0,434	0,306
MEDIAN	A	$\beta = -1/4$	–0,312	0,144
WARD	$(I-A)^{-1}$	$\beta < 0$	0,554	0,020
WARD	A	$\beta < 0$	0,637	0,032

Leontief y sobre la Matriz de Coeficientes Técnicos. La utilización de la Inversa de Leontief está justificada por dos razones fundamentalmente: en primer lugar, se asociarían industrias que se caracterizan por los requisitos unitarios directos e indirectos de toda clase de inputs, y en segundo lugar, se contemplan estas industrias como sectores verticalmente integrados sobre los que se definen las funciones de producción. Utilizamos el método de Centroide, Mediana y de Ward. Para elegir el método adecuado, estimamos la ecuación [14] para las seis medidas de similaridad propuestas. Observando los resultados obtenidos, y que presentamos en el Cuadro 2, podemos establecer que el único método correcto que se puede aplicar en este caso es el de la Mediana. Utilizando esta estrategia, agregamos las industrias por medio del análisis cluster y el resultado para la matriz de transacciones de la Economía Valenciana de 1980, se presenta en el dendograma del apéndice 2.

6. CONCLUSIONES

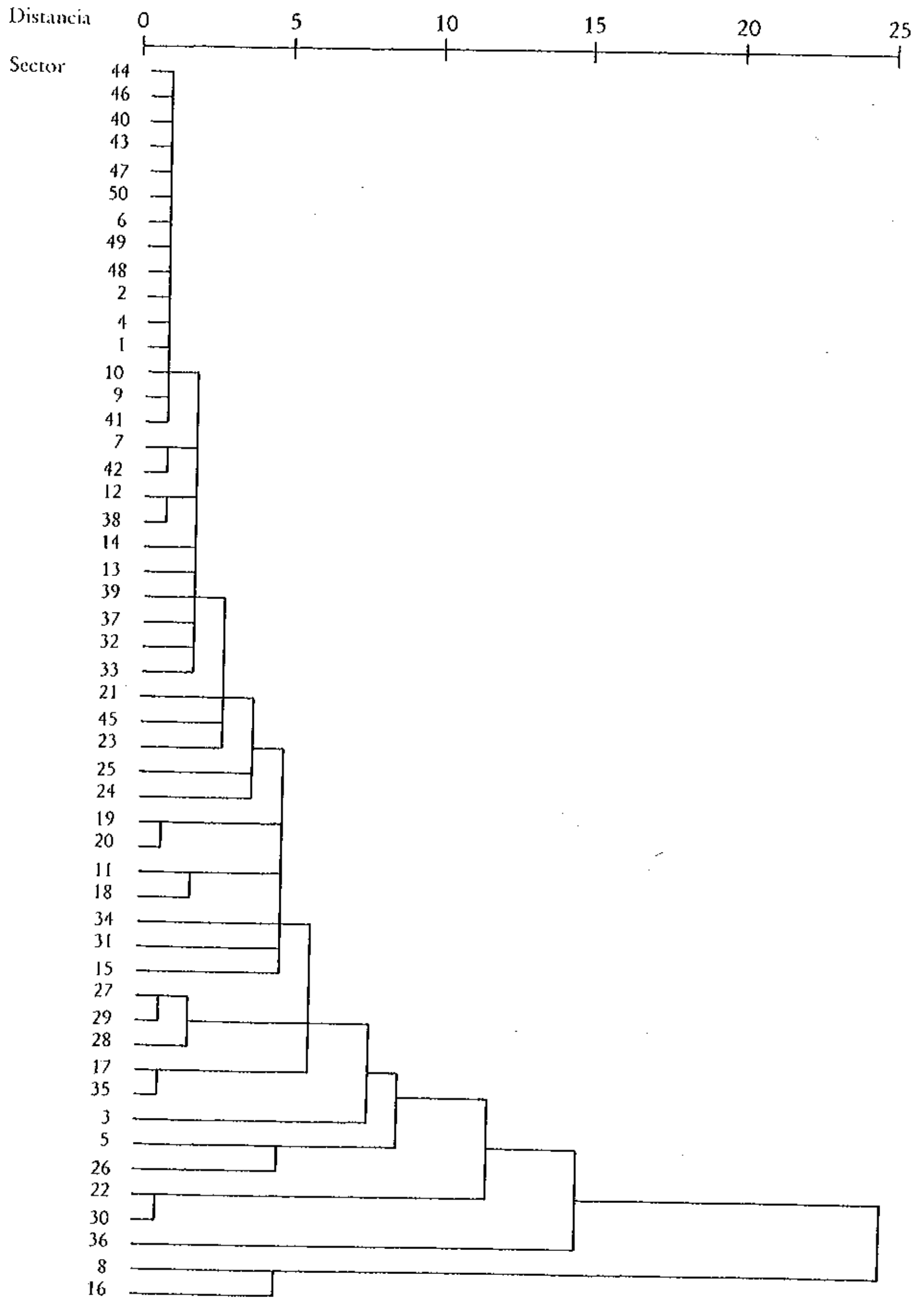
Esta técnica nos permite determinar qué tipo de estrategia ha de utilizarse al llevar a cabo la agregación de industrias en sectores, utilizando el análisis cluster. Esto nos posibilita discriminar entre las distintas agrupaciones que se obtendrían realizando clusters de distinto tipo. La agrupación obtenida mediante esta técnica puede servir como modelo de referencia para juzgar si la definición de sectores, con los que se trabaja habitualmente, es adecuada o no.

Es importante señalar que el nivel de agregación que se establezca como definitivo no es endógeno a este modelo, sino que depende del criterio del investigador. Llevando el análisis cluster hasta el nivel máximo de agregación, se agruparían evidentemente todas las industrias en un único sector, por dispares que fuesen entre sí. Sin embargo, esta característica queda perfectamente reflejada en el método que aplicamos, puesto que se proporciona una medida de disparidad de la agregación realizada a cada nivel. Así, por ejemplo, en nuestro caso, las industrias que se agregan en una primera etapa (las que muestran menores distancias) son aquéllas que agrupan sectores intensivos en mano de obra, básicamente las agrícolas y las de servicios. Comprobamos que al mismo nivel de agregación, se agrupan las industrias del sector textil, las del sector automovilístico, las de la industria química, las derivadas de la ganadería, las industrias relacionadas con la construcción, las dependientes del petróleo, etc. Observamos cómo para etapas posteriores, se va produciendo una agregación paulatina de industrias más dispares, y en general más intensivas en capital. El último grupo de industrias que se agregan son las del sector energético. Puede comprobarse la racionalidad de este proceso, que nos permite determinar el nivel de agregación más adecuado en función de la similitud de las industrias que agregamos. Una vez hecho esto, queda determinado cuáles van a ser las industrias que configuran los sectores finales.

APENDICE 1

1. Cítricos
2. Hortofrutícola
3. Vid y vino
4. Otros productos agrícolas
5. Ganadería
6. Silvicultura y otros
7. Pesca
8. Energía eléctrica
9. Agua y gas
10. Extracción de minerales
11. Siderurgia
12. Cemento y material de construcción
13. Cerámica
14. Azulejos
15. Vidrio
16. Petroquímica
17. Otros productos químicos
18. Productos metálicos
19. Fab. y rep. maquinaria
20. Vehículos y motores
21. Construcción otros materiales de transporte
22. Carne, leche y aceite
23. Conservas
24. Chocolate, turrón y otros
25. Bebidas
26. Otras ind. alimentarias
27. Fibras, hilo y tejidos
28. Confección
29. Textil hogar
30. Cuero, piel (exc. calzado)
31. Zapato de cuero
32. Madera (exc. muebles)
33. Muebles de madera
34. Papel y artes gráficas
35. Caucho y plástico
36. Ind. diversas exc. juguetes
37. Juguetes
38. Construcción y Obras Pub.
39. Recuperación y reparación
40. Comercio
41. Hostelería
42. Transporte y anexos
43. Comunicaciones
44. Inst. de crédito y seguros
45. Alquileres
46. Serv. a las empresas
47. Enseñanza e investigación
48. Sanidad
49. Administración Pública
50. Servicio doméstico y otros

Apendice 2
Dendograma utilizando el método de la mediana. A



REFERENCIAS

ANDERBERG, M.R (1973): *«Cluster Analysis for Applications»*. Academic Press, London.

ARA, K. (1959): «The Aggregation Problem in Input-Output Analysis». *Econometrica,* 27, 257-262.

BLIN, J.M. and COHEN, C. (1977): «Technological Similarity and Aggregation in Input-Output Systems: A Cluster-Analytic Approach». *The Review of Economics and Statistics,* 59, 82-91.

ESCUDERO, L.F. (1977): *«Reconocimiento de Patrones»*. Paraninfo, Madrid.

FISHER, W.D. (1962): «Optimal Aggregation in Multi-Equation Prediction Models». *Econometrica,* 30, 744-769.

FISHER, W.D. (1969): *«Clustering and Aggregation in Economics»*. John Hopkins Press, Baltimore.

HATANAKA, M. (1952): «Note on Consolidation within a Leontief System».*Econometrica,* 20, 301-303.

KOSSOV, V. (1972): «The Theory of Aggregation in Input-Output Models», in CARTER, A.P. and BRODY, A. (eds.): *«Contributions to Input-Output Analysis»*. North-Holland Publishing, Amsterdam.

KYMN, K.O. and NORSWORTHY, J.R. (1976): «A Review of Industry Aggregation in Input-Output Models». *The American Economist,* Spring, 5-10.

MALINVAUD, E. (1954): «Aggregation Problems in Input-Output Model» in BARNA, T. : *«The Structural Interdependence of the Economy»*. John Wiley and Sons, New York.

MORIMOTO, Y. (1971): «A Note on Weighted Aggregation in Input-Output Analysis». *International Economic Review,* 12, 138-143.

NEUDECKER, H. (1970): «Aggregation in Input-Output Analysis: An Extension of Fisher's Method». *Econometrica,* 38, 921-926.

SNEATH P.H.A. and SOKAL, R.R. (1973): *«Numerical Taxonomy»*. W.H. Freeman Press, San Francisco.

THEIL, H. (1957): «Linear Aggregation in Input-Output Analysis». *Econometrica,* 25, 111, 122.