

Métodos estadísticos en el análisis regional

María Pilar Martín Guzmán

1. INTRODUCCIÓN

El espectro de los métodos estadísticos utilizados en el análisis regional es amplísimo, de manera que no es fácil encontrar técnicas que, empleadas con éxito en otros campos científicos, no hayan sido experimentadas con mayor o menor fortuna en los problemas regionales. Pero algunas de ellas han resultado tan adecuadas para la modelización y análisis de este tipo de problemas, que se emplean ya de forma habitual, y han sido incorporadas al acervo de conocimientos básicos de los investigadores y especialistas de esta rama de la economía.

De entre éstas hay muchas que se han introducido en estudios regionales tras haber sido ampliamente utilizadas en otros campos de la economía. Pero hay otras para las que precisamente el análisis regional ha sido el principal motor de su desarrollo y aplicación general a la economía, como es, por ejemplo, la modelización de variables categóricas. Una buena parte de estas técnicas ha visto sus inicios en los años cincuenta y aún antes, en el mundo de la biología o la medicina, pero su incorporación masiva a la investigación económica ha comenzado en la década de los setenta. Ello podría explicarse por el hecho de que, presentando en general los fenómenos socioeconómicos una mayor complejidad de interrelaciones entre variables, y proporcionando las encuestas una información más amplia que la generada por experimentos, la cantidad de datos a manejar hizo prácticamente imposible la aplicación de técnicas estadísticas en tanto no se dispuso de instrumentos de cálculo adecuados. La introducción masiva de computadoras de alta capacidad, que tan profundamente ha modificado otras ramas de la ciencia, ha jugado también en este tema un papel fundamental.

Es difícil fijar líneas de tendencia en un campo de desbordante vitalidad creadora y rápida evolución como es en estos momentos el análisis regional

cuantitativo. Atendiendo, no obstante, a lo que puede ser en la actualidad su problemática más específica, podríamos arriesgarnos a apuntar dos direcciones:

a) Los estudios regionales suelen revestir una especial complejidad y requerir un número elevado de variables, entre las cuales hay que considerar a veces un múltiple juego de interrelaciones. Por eso resulta especialmente complicado diseñar modelos de teoría económica que reflejen adecuadamente los fenómenos a analizar. Esto explicaría la popularidad creciente de que gozan en el campo regional los denominados «modelos blandos», es decir aquellos cuya aplicación requiere nula o muy pequeña información o modelización a priori.

b) La escasez de datos regionales en la mayoría de los países, o la inadecuación de los existentes para los métodos estadísticos al uso es generalmente un problema grave. Frente a él caben dos soluciones: 1) disponer de más y mejores datos, lo que no siempre está en manos del investigador conseguir; o bien 2) adaptar los métodos existentes o crear otros nuevos que permitan utilizar de la mejor manera posible los datos disponibles. En esta línea podría encuadrarse el considerable nivel de desarrollo teórico que, a impulsos en buena medida del análisis regional está alcanzando la estadística cualitativa.

Describir, aunque sea brevemente, todas las técnicas estadísticas que se emplean hoy de forma habitual en el análisis regional, comentando la problemática específica de su aplicación en este campo, requeriría una extensión mucho mayor de la prevista en este trabajo. Es preciso seleccionar unas pocas entre ellas, lo cual no es fácil, dado su número de variedad. De entre las que metodológicamente superan la barrera de lo descriptivo nos hemos decidido por aquellas que nos han parecido de mayor interés conceptual desde el punto de vista estadístico y más en auge en las revistas y trabajos de análisis regional. Así, dedicamos los cuatro primeros apartados a estudiar los grandes grupos de métodos que constituyen hoy en día el análisis de datos, continuos o discretos, concretamente: los encaminados a la determinación de estructuras dentro de un conjunto de variables, o al análisis de relaciones entre conjuntos distintos, los empleados en la clasificación y agrupación de elementos, y los específicos en el tratamiento de variables cualitativas. Finalmente, dedicamos un apartado a los modelos econométricos multiecuacionales y su empleo en la evaluación de políticas regionales.

2. DETERMINACIÓN DE ESTRUCTURAS EN UN CONJUNTO DE VARIABLES

Anteriormente nos hemos referido a las dificultades que con frecuencia presenta la modelización teórica de las relaciones entre variables regionales. Cuando éste es el caso, la matriz de datos —cuyas filas están formadas por los

valores que toman las variables a considerar en cada una de las regiones— puede ser el único elemento de información de que dispongamos. Interesará, pues, utilizar técnicas que nos permitan «explorar» sistemáticamente la matriz de datos, intentando detectar en ella indicios de relaciones entre las variables, o bien «confirmar» empíricamente la existencia de relaciones que ya sospechábamos.

Uno de los problemas que inicialmente se nos presentan en un estudio regional para el que se carece de una modelización teórica previa es la selección de las variables a considerar. Al no tener una idea clara de cuáles son las que van a jugar un papel importante, o a aportar una mayor información, se tiende a incluir en el estudio un número grande de ellas, lo que dificulta su manejo y exploración. Entre algunas de estas variables existirán, probablemente, interrelaciones, de manera que una buena parte de la información que aporta alguna de ellas será compartida por otras varias; y esto es tanto más de esperar que ocurra cuanto mayor sea el número de variables tomadas. El análisis de los datos será mucho más fácil si de alguna manera conseguimos eliminar esta información repetida o «redundante», manteniendo básicamente la que es de interés para nuestro problema, pues ello llevará a una reducción de las dimensiones del conjunto de información, y a su transformación en otro mucho más manejable.

Las técnicas adecuadas para lograr estos objetivos son básicamente dos: el método de componentes principales, y el análisis de factores. Ambos buscan la mayor reducción de dimensionalidad de la matriz de datos compatible con el mantenimiento de una porción considerable de la información relevante, aunque ésta se valora en función de criterios diferentes, lo que hace que estas técnicas no sean aplicables indistintamente a cualquier problema. Como resultado obtendremos una visión mucho más clara de la estructura de interrelaciones entre las variables consideradas, que nos permitirá obtener conclusiones y elaborar hipótesis ad-hoc respecto a su comportamiento. Estas técnicas están siendo continuamente aplicadas a muy diversos problemas regionales, y especialmente al análisis de estructuras y determinación de tipologías en regiones.

El método de componentes principales nos proporciona un criterio objetivo para la reducción del número de variables de un problema. Tiene por objeto sintetizar la información contenida en la matriz de datos sustituyendo las variables que la forman por unas nuevas variables que, en número mucho menor, expliquen, sin embargo, una proporción suficientemente grande de la varianza inicial del problema. Estas nuevas variables se eligen de manera que no contengan información redundante, y reciben el nombre de componentes principales.

La primera componente principal se determina como la combinación lineal de las variables iniciales que explique la máxima varianza, y constituye la mejor síntesis unidimensional posible de los datos. La segunda componente prin-

cial se pretende que explique la mayor varianza posible no explicada por la primera, de manera que ambas no contengan información común: para ello se selecciona la mejor combinación lineal entre todas las que están incorrelacionadas con la primera. Se procede así hasta haber determinado un número de componentes suficiente para explicar la proporción de varianza deseada. Una de las ventajas de este método es que proporciona, junto a cada componente, la medida de su capacidad explicativa, de manera que siempre sabemos cuál es la cantidad de información que se pierde al sustituir las variables iniciales por las componentes.

Rara vez se puede proceder a una reducción de la dimensión sin perder información, pues para ello sería preciso que la matriz de covarianzas de los datos fuese de rango no total, es decir, que existiera una relación lineal exacta entre algunas de las variables del problema. Lo habitual es que al reducir la dimensión se pierda información, tanta más cuanto más independientes sean entre sí las variables iniciales. Pero es frecuente que un número reducido de componentes principales —tres o cuatro— explique un porcentaje muy elevado de la varianza total, permitiendo así conservar casi toda la información inicial con un número de variables mucho menor. La técnica es especialmente eficaz cuando las variables de la matriz de datos están muy relacionadas entre sí.

Este método se emplea frecuentemente para eliminar problemas de multicolinealidad entre grupos de variables. Por otra parte, es un procedimiento muy usado en la construcción de indicadores regionales: la primera variable canónica es el mejor resumen unidimensional de la información, y constituye, en general, un excelente indicador. Si la proporción de varianza acumulada por ella resulta insuficiente, lo que se hace es incorporar más información, añadiendo a la primera componente principal las sucesivas, convenientemente ponderadas en función de su capacidad explicativa —tomando, por ejemplo, como ponderación la raíz cuadrada de la varianza de cada una— hasta conseguir un indicador con un nivel de explicación suficiente¹.

Tiene, sin embargo, este método algunas limitaciones que es preciso tener en cuenta. En primer lugar, las nuevas variables componentes principales no se identifican con las variables económicas iniciales, por lo que su interpretación puede en ocasiones no ser fácil. Por otra parte, los resultados obtenidos no son invariantes respecto a cambios de escala, de manera que cuando se estén manejando variables heterogéneas no estandarizadas será preciso establecer de antemano cuál es el orden de unidades más conveniente para medir cada una de ellas.

El análisis de factores tiene también como objetivo último transformar un

¹ PETERS, W. S. y BUTLER, J. Q.: «The Construction of Regional Economic Indicators by Principal Components», *Annals of Regional Science*, IV, 1 (1970), págs. 1-14.

problema en otro de dimensión menor pero con un criterio diferente, pues no persigue conseguir la varianza máxima, sino reproducir al máximo las correlaciones. Así pues, se trata de un método esencialmente diferente del anterior, aunque en ocasiones se confundan por la semejanza de algunos de los procesos intermedios de su desarrollo.

El análisis factorial tiene por objeto descomponer las variables de la matriz de datos en suma de factores de dos tipos: los del primer tipo, en número previamente fijado, comunes a todas las variables, y los del segundo, específicos de cada una de ellas. Se trata, por tanto, de explicar por medio de los factores, lo que tienen de común las variables manejadas. Como en último término se pretende simplificar el problema, es decir, reducir al máximo su dimensión, se intenta siempre fijar el menor número posible de factores capaz de contener básicamente la información común proporcionada por los datos.

Para ello se establece el modelo

$$X = \Gamma Y + e$$

donde X es la matriz de datos de las variables iniciales, Y la de las nuevas variables o factores comunes, y e el vector de factores específicos. Los factores comunes se toman incorrelacionados, de manera que no compartan información, y los coeficientes λ_{ij} de la matriz Λ , llamados pesos, reflejan la importancia del factor j en la variable i . A

$$h_i^2 = \sum_{j=1}^m \lambda_{ij}^2$$

se le llama comunalidad, y es la medida de la cantidad de información contenida en la variable i que es compartida por el conjunto de los factores comunes, es decir, que puede ser explicada por ellos. El resto de la información queda, por tanto, en el factor específico de esa variable.

Aunque el modelo tiene la apariencia de una regresión, no puede estimarse como tal, pues los factores Y no son observables. Los únicos datos de que se dispone son los de las variables iniciales X . En estas condiciones ya se comprende que el modelo no tendrá una solución única. En realidad, para cada solución del modelo, cualquier rotación de sus factores es también una solución. Esto representa una ventaja, pues nos permite buscar, mediante rotaciones adecuadas, la composición de factores que tenga una interpretación económica más clara.

Así pues, el método consta de dos procesos que describiremos brevemente: la determinación de una solución inicial, y la rotación de la misma en busca de la orientación más favorable de los ejes. Para el primero se utilizan diversos métodos, entre los cuales los más usuales son los basados en técnicas de componentes principales o en estimaciones máximo-verosímiles. Estos últimos permiten resolver, mediante contrastes estadísticos, uno de los

problemas más delicados del análisis factorial: la determinación del número de factores independientes necesarios para explicar toda la comunalidad de las variables. Sin embargo, en los estudios regionales son mucho más usados los otros², y ello por dos razones. En primer lugar, porque no requieren la hipótesis previa de normalidad de los datos, con frecuencia arriesgada en este tipo de estudios. Pero además, porque el analista regional casi nunca está interesado en identificar todo lo que las variables tienen de común a costa de manejar un gran número de factores; en general, prefiere explicar, con un número pequeño de ellos —dos o tres— la mayor parte de la comunalidad subyacente, aunque pierda alguna información. En este sentido, las técnicas basadas en componentes principales tienen ventajas, pues van proporcionando sucesivamente factores de importancia decreciente respecto a los anteriores, de manera que los primeros de ellos contendrán la mayor capacidad explicativa que puede obtenerse para una dimensión dada. En cambio, representan el inconveniente de no ser invariantes a cambios de escala.

Determinada una solución inicial, se busca, mediante rotaciones de los ejes, la orientación más favorable para la interpretación económica de los factores resultantes que, como en el método de componentes principales, puede no ser fácil. Estas rotaciones pueden ser ortogonales u oblicuas. Sin embargo, las rotaciones oblicuas presentan respecto a las ortogonales dos inconvenientes graves: no mantienen constantes los valores de las comunalidades, y además pueden generar factores que no estén incorrelacionados. Por ello, los métodos de rotación ortogonal tipo varimax o similares son los más empleados.

3. ANÁLISIS DE RELACIONES ENTRE CONJUNTOS DE VARIABLES

En el apartado anterior vimos algunos procedimientos de síntesis de la información contenida en un conjunto de variables. Ahora vamos a ver cómo pueden establecerse relaciones entre dos o más conjuntos de variables que ayuden a analizar las vinculaciones causales existentes entre ellas y permitan, eventualmente, utilizar los valores que toman las variables de alguno de los conjuntos para predecir la situación del otro.

En el caso, especialmente sencillo, en que al menos uno de los conjuntos consta de una sola variable, el método idóneo es la regresión, que nos permite encontrar la combinación lineal de las variables del segundo conjunto —regresores— que mejor explica el comportamiento de la variable del primero. El

² JAUMOTTE, C. y PAELINCK, J.: «The Differential Economic Structures of the Belgian Provinces: a Time Varying Factor Analysis», *Regional and Urban Economics*, vol. 1, 1971, págs. 41-75; o VAN ROMPUY, P.: «A Two-step Procedure for Intertemporal Factor Analyses», *Regional Science and Urban Economics*, 9, 1979, págs. 261-274.

criterio con el que se mide la capacidad explicativa es el de correlación máxima entre la variable a explicar y la combinación lineal de regresores. Esta técnica, de fuerte arraigo en la investigación económica empírica, ha sido incorporada a la metodología del análisis regional cuantitativo desde sus comienzos, y empleada abundantemente. Con todo tiene una serie de limitaciones que pasamos a comentar.

En primer lugar, su aplicación exige una información *a priori* de la que no siempre se dispone, pues requiere una selección previa de las variables que han de formar parte del conjunto explicativo; y esta selección tendrá que basarse en una modelación teórica o en un conocimiento empírico de su estructura de interrelaciones —que eventualmente puede estudiarse por los métodos descritos en el apartado anterior—. Además las variables seleccionadas han de satisfacer unas restricciones bastante fuertes, cuyo incumplimiento lleva consigo una reducción considerable del grado de fiabilidad de los resultados. Por otra parte, la regresión uniecuacional establece entre las variables unas relaciones asimétricas, con las que no es factible describir las relaciones de interdependencia que con frecuencia caracterizan el comportamiento de un modelo regional. Así resulta que éste queda deficientemente especificado con una ecuación de regresión, dando lugar a sesgos en las estimaciones obtenidas.

En el análisis regional se utilizan muy frecuentemente dos generalizaciones de esta técnica que resuelven en parte los problemas descritos: los modelos econométricos multiecuacionales y la correlación canónica. A la primera de ellas dedicamos un apartado posterior. Nos ocuparemos aquí de la segunda.

La correlación canónica tiene por objeto establecer relaciones entre dos conjuntos distintos de variables, determinando combinaciones lineales en cada uno de ellos de forma que las nuevas variables así construidas tengan la mayor relación posible entre ellas, es decir, que su correlación sea máxima. Así pues, esta técnica aparece como una generalización de la regresión al caso en que se pretenda explicar un conjunto de variables a partir de otro conjunto. Carece, no obstante, de algunas de las limitaciones que hemos descrito en aquélla, pues no requiere una selección previa tan estricta de las variables a utilizar y, por otra parte, tampoco exige en éstas, restricciones especiales, pudiendo, por ejemplo, incluir sin ningún tipo de inconveniente variables entre las que exista un alto grado de multicolinealidad. Así pues, la correlación canónica entra en la categoría de los modelos blandos, para los cuales no se requiere una modelización *a priori*. Por otra parte, se da en este método una simetría total, de manera que cada uno de los conjuntos de variables puede utilizarse indistintamente como predictor del otro.

La relación entre los dos conjuntos se sintetiza en forma de pares de variables, llamadas canónicas, y que se obtienen de la forma siguiente: se determina inicialmente el par de combinaciones lineales de cada uno de los conjuntos de variables de manera que su correlación sea máxima, y este par

proporciona la mejor posibilidad de predicción de cada conjunto respecto al otro. Determinado así el primer par de variables canónicas se calcula, de entre las combinaciones lineales de cada conjunto que no comparten información con las variables canónicas ya obtenidas —es decir, que estén incorreladas con ellas— el par que tenga la correlación mayor posible, y este proceso puede continuar así hasta construir tantos pares canónicos como variables iniciales haya en el conjunto más pequeño. Pero lo que suele ocurrir es que los primeros pares presentan una correlación elevada, que va disminuyendo muy rápidamente para los pares sucesivos, de manera que una buena parte de la información habrá quedado acumulada en unos pocos. Así pues, será preciso decidir cuántos pares de variables retenemos en nuestro análisis, buscando una solución de compromiso entre el deseo de simplificar el problema y la conveniencia de no perder mucha información.

Si la hipótesis de normalidad es aceptable para ambos conjuntos de variables, el test de Bartlett nos proporciona un criterio para eliminar los pares de variables canónicas cuya correlación no sea significativamente distinta de cero. Pero la reducción de dimensión resultante suele ser insuficiente, pues deja en activo pares con correlaciones pequeñas, que complican innecesariamente el problema y dificultan su interpretación. Por eso en los estudios regionales³ el criterio empleado para decidir cuántos pares de variables canónicas es conveniente mantener si se quiere conservar una cantidad suficiente de información es el coeficiente de redundancia de Stewart y Love

$$s_i = \rho_i^2 \sum_{j=1}^{p_1} \frac{r_{ij}^2}{\rho_1}$$

donde las r_{ij} nos representan la correlación entre la variable canónica i -ésima de un grupo y cada una de las variables j iniciales del mismo, ρ_i la correlación canónica entre el i -ésimo par de variables canónicas, y p_1 el número de variables del grupo. Como vemos, este coeficiente mide la capacidad media de explicación de la variable canónica i -ésima de uno de los grupos sobre el conjunto de las variables iniciales del otro. Sus valores para cada uno de los conjuntos nos dan una medida de la importancia relativa de cada par de variables canónicas en el problema, y, en base a ellos podremos decidir cuántos pares mantenemos en nuestro análisis. La mayoría de las veces se pone así de manifiesto que pares que el test de Bartlett identifica como significativos, tienen en realidad una capacidad explicativa pequeña, de manera que el coefi-

³ RANNER, A. P.: «Analysis Relations between Regions and their Surroundings: A Canonical Correlation Approach», *Regional and Urban Economics*, 4, 1974, págs. 141-162, y ARNÁIZ, G.; MARTÍN-GUZMÁN, P.; MARTÍN PLIEGO, J. y TOLEDO, I.: *Discriminación y clasificación de las regiones fiscales en España*, IEF, 1987.

ciente de redundancia facilita en general una reducción de dimensionalidad muy superior a la derivada del test de Bartlett.

El método de correlación canónica es susceptible de generalización a varios conjuntos de variables⁴.

4. CLASIFICACIÓN Y AGRUPACIÓN DE ELEMENTOS

Uno de los problemas que con más frecuencia se presentan en el análisis regional es el de la agrupación y clasificación de elementos en base a unos ciertos criterios dados por la índole del problema a estudiar: es lo que se conoce con el nombre de taxonomía regional. Su objetivo es agrupar un conjunto de unidades en clases —cuyo número y característica distintiva pueden estar previamente fijado, o no estarlo— de manera que la agrupación resultante presente la mayor homogeneidad posible dentro de cada grupo, y la mayor heterogeneidad entre grupos diferentes, medidas ambas de acuerdo con unos criterios establecidos en función de los objetivos del estudio.

Las técnicas estadísticas más usadas en los problemas de taxonomía regional son fundamentalmente el análisis discriminante y el de conglomerados. Ambas tienen de común su objetivo último: conseguir una clasificación adecuada de los elementos. Como técnicas, en cambio, son muy diferentes⁵. El análisis discriminante se clasifica sobre un esquema previo en el que están ya fijados, no solamente el número de grupos, sino también la característica esencial que los distingue, es decir, las clases a formar están perfectamente delimitadas de antemano, y el problema consiste en decidir, para cada unidad, a cuál de estas clases la asignamos. Por el contrario, en el análisis de conglomerados las clases se van diseñando y delimitando a medida que se construyen, y sus características distintivas se ponen de manifiesto como resultado de la agrupación. Puede incluso ocurrir que ni siquiera el número de grupos esté establecido de antemano, sino que se decida o acepte como un resultado más del estudio.

Hemos dicho que el análisis discriminante consiste en la asignación de un elemento a una clase, entre dos o varias delimitadas en función de una cierta característica. Si ésta fuera observable en el momento de realizar la clasificación, no sería preciso recurrir a una técnica estadística. Lo que ocurre es que lo es solamente *a posteriori*, de manera que al efectuar el estudio es necesario decidir, en base a la información disponible de ese elemento —en general, los valores que toma para una serie de variables relacionadas con la característica

⁴ KETTENRING, J. R.: «Canonical Analysis of Several Sets of Variables», *Biométrica*, 58 (1971), págs. 433-51.

⁵ RICHARDSON, D. T. y THALHEIMER, R.: «On the Use of Grouping Methods in the Analysis of Residential Housing Markets», *Regional Science and Urban Economics*, 12, 1982, págs. 285-304.

en cuestión— a cuáles de los resultados que ésta puede presentar lo asignamos⁶.

Para ello necesitamos disponer de un conocimiento previo del comportamiento de los elementos de cada una de las clases, que generalmente tendremos en la forma de muestras de individuos que se han decantado ya por cada uno de los resultados de la característica, y de los que conozcamos los valores que han tomado las variables a considerar. Partiendo de esta información previa, el análisis discriminante nos proporciona un «retrato-robot» del elemento típico de cada una de las clases, basado en las variables conocidas, y que nos conduce a criterios racionales de asignación de cada nuevo elemento a alguna de ellas.

El «retrato-robot» citado suele consistir en una o varias funciones lineales de las variables conocidas que se llaman funciones lineales discriminantes y que resumen la información aportada por el nuevo elemento en uno o varios números. Estableciendo intervalos adecuados, tendremos un criterio de asignación de elemento a alguna de las clases de acuerdo con los valores de estos números. Si las poblaciones correspondientes a cada clase son muy distintas entre sí, concretamente, si sus matrices de covarianzas son muy diferentes, puede ofrecer mejores resultados el empleo de funciones cuadráticas⁷.

Bajo ciertas condiciones —normalidad de las poblaciones de cada clase— la función lineal discriminante nos proporciona los mejores criterios de asignación posibles. En los demás casos, podemos confiar en que nos dé una clasificación bastante buena. Con todo, es importante establecer una medida de su bondad, o validación de los resultados, empleando un procedimiento de tipo jackknife o similar⁸.

El análisis de conglomerados es una técnica mucho más simple desde el punto de vista estadístico, y para la cual no es preciso disponer de una muestra previa, pues ahora carece de sentido intentar una caracterización de clases que aún no han sido diseñadas. Para aplicarla basta con la información aportada por los propios elementos a clasificar. Si estableciéramos una ordenación de las técnicas multivariantes en función de su mayor o menor «dureza» —o cantidad de información *a priori* requerida para su aplicación— el análisis de conglomerados estaría situado en el extremo inferior del espectro.

⁶ ANCOT, J. P. y PAELINCK, J.: «A Discriminant Analysis Approach to Regional Threshold Problems: A Progress Report», *Paper of the Regional Science Association*, vol. 42, 1979, págs. 139-151, y O'FARRELL, P.N. y CROUCHLEY, R.: «The Locational Pattern of New Manufacturing Establishments: An Application of Discriminant Analysis», *Regional Studies*, vol. 13, 1979, págs. 39-59.

⁷ SMITH, C. A. B.: «Some Examples of Discrimination», *Ann. Eugen.*, 13, 1947, págs. 272-282.

⁸ Vid. MILLER, R. G.: «The Jackknife: a Review», *Biométrica*, 61, 1974.

Los elementos a agrupar vendrán caracterizados por los valores que tomen en ellos unas ciertas variables, en general en número bastante elevado, y ocurrirá seguramente que elementos muy parecidos en algunas variables sean muy diferentes respecto a otras. Será preciso sintetizar, por tanto, esta información multivariante, estableciendo una definición de distancia, que asigne a cada par de individuos del conjunto a clasificar un número no negativo — positivo siempre que ambos sean distintos—, y que cumplan además algunas otras propiedades de entre las que la más frecuentemente exigida es la triangular. Con ella se forma la matriz de distancias, cuadrada, simétrica y de dimensión igual al número de elementos.

Existen diversas definiciones de distancias⁹, y la elección de una de ellas es una decisión importante, pues puede condicionar los resultados obtenidos. Es frecuente utilizar la distancia euclídea, por su mayor sencillez de cálculo; pero tiene el inconveniente de no ser invariante a cambios de escala, con los problemas que esto trae consigo y que hemos visto en apartados anteriores. Por eso a veces se emplea la distancia de Mahalanobis, que sí lo es.

Cuando las variables que caracterizan a cada elemento en el estudio son de tipo cualitativo, el instrumento de agrupación distancia se sustituye por el de semejanza entre dos individuos, que puede definirse de muy variadas formas (valorando, por ejemplo, las coincidencias entre ellos, o solamente las coincidencias positivas), en función de cuál sea el objetivo del análisis, pero que siempre será un número positivo y simétrico. Con objeto de poder aplicar las técnicas de agrupación, que suelen estar diseñadas para matrices de distancias, se pueden convertir las semejanzas en distancias mediante la transformación de Gower

$$d_{ij} = (s_{ij} + s_{ji} - 2s_{ij})^{1/2}$$

Los elementos se clasifican de manera que vayan agrupándose en cada clase los que, de acuerdo con la matriz de distancias elaborada, estén «más próximos» entre sí. Para ello pueden usarse diversos algoritmos, que se dividen en dos tipos: jerárquicos y no jerárquicos.

Los jerárquicos se caracterizan por un proceso iterativo en el que en cada etapa se agrupan conglomerados de unidades de la etapa anterior formando un árbol, cuya representación gráfica recibe el nombre de dendograma. Precizando un poco más, diremos que un método de agrupación es jerárquico cuando, al pasar de una etapa en que tenemos $g + 1$ clases a la etapa siguiente, en que éstas quedan agrupadas en g , ocurre siempre que $g - 1$ clases permanecen invariantes y la otra resulta de la fusión de las dos restantes. Por tanto, en un proceso jerárquico no puede ocurrir que un elemento

⁹ DURAN, B. S. y ODELL, P. L.: *Cluster Analysis: A Survey*, Springer-Verlag, 1974.

que en una etapa del proceso quede encuadrado en un determinado grupo, se desgaje de él en una etapa posterior para formar parte de otro.

En cada etapa del proceso jerárquico, las clases que se agrupan son aquellas que distan menos entre sí. Si la matriz de distancias cumple la propiedad ultramétrica, esta norma da lugar a una única forma de agrupación¹⁰. Pero si, como es lo frecuente, es simplemente enclídea, las distancias entre clases no quedan determinadas de forma única a partir de ella, pudiendo definirse de muy diversas maneras. Las más usuales son la del vecino (*single linkage*) en la que la distancia entre dos clases es la de sus elementos más próximos, la del vecino más alejado (*complete linkage*), en la que se toma como distancia la de sus elementos más alejados, y la distancia media, que toma la media de todas las distancias entre elementos de una de las clases respecto a los de la otra.

Cada una de estas definiciones puede dar lugar a estructuras de agrupación diferentes. Así, la del vecino más alejado tiende a proporcionar mayor número de conglomerados y más compactos, en tanto que la del vecino más próximo propende a establecer mayores conexiones entre los elementos, extendiéndose en uniones sucesivas, y creando menos conglomerados. Ambos métodos tienen el inconveniente de ser muy sensibles a la presencia de *outliers*; por eso se prefiere en algunos casos el de la distancia media, mucho menos sensible a este problema, y que proporciona un tipo de conglomerado intermedio entre los otros dos.

En los métodos jerárquicos es frecuente que el número de clases a formar no esté previamente fijado, sino que se decida a la vista de la agrupación resultante. Por el contrario, en los no jerárquicos suele establecerse de antemano. En ellos cada paso en la iteración puede descomponer un conglomerado ya formado y redistribuir sus elementos entre otros¹¹. El criterio de formación de los mismos suele consistir en la optimización de algún tipo de medida de las distancias, y puede definirse de muy variadas formas en función del problema que se aborda¹².

Uno de los más usados es el llamado de las *K*-medias, en el que una división inicial en *K* grupos se va modificando iterativamente con el criterio de que

¹⁰ GNANADESIKAN, R.: *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, 1977.

¹¹ FISCHER, M.: «Regional Taxonomy: a Comparison of Some Hierarchic and Non-Hierarchic Strategies», *Regional Science and Urban Economics*, vol. 10, 1980, págs. 503-537.

¹² Vid. ERTL, H.; FISCHER, M. y WOHLSCHLAGL, H.: «A Methodological Approach for Large Regional Taxonomic Problems: Spatial Patterns of Population Development in Austria», *Papers of the Regional Science Association*, vol. 44, 1980, págs. 119-135; KEANE, M.: «A functional Distance Approach to Regionalisation», *Regional Studies*, vol. 12, 1978, págs. 379-386; MILLER, A.: «A Taxonomy of Nebraska County Economics: An Application of the Optimization Approach to Identifying a System of Regions», *Journal of Regional Science*, vol. 16, 1976, págs. 225-236; y HARTIGAN, J. A.: *Clustering Algorithms*, Wiley, 1975.

cada elemento se una a la clase a cuya media o centro de gravedad esté más próximo, y que se emplea mucho en modelos de ubicación de centros de servicios y asignación de individuos a ellos.

Es frecuente que en un proceso de agrupación regional los métodos taxonómicos descritos se combinen con técnicas multivariantes de síntesis de la información con objeto de hacerlos más fácilmente aplicables. Por ejemplo, el proceso puede en ocasiones simplificarse si casi toda la información disponible se sintetiza en unas pocas componentes principales.

5. MODELIZACIÓN DE VARIABLES CUALITATIVAS

No todas las variables que juegan un papel fundamental en el análisis socioeconómico son medibles. Existen algunas que son esencialmente cualitativas, y cuya inclusión en un modelo con tratamiento cuantitativo requiere la asignación de valores numéricos a las diferentes opciones o categorías que estas variables ofrecen. Esto es especialmente cierto en el campo de la economía regional, pues muchos de sus modelos tienen como objetivo principal el estudio de la elección que un agente económico lleva a cabo entre unas cuantas alternativas.

Una variable cualitativa o categórica tiene en términos cuantitativos una representación en forma de variable discreta que toma un número pequeño de valores, con frecuencia sólo dos. La inclusión de variables de este tipo suele afectar en mayor o menor grado a la validez o ámbito de aplicación de las técnicas estadísticas tradicionales, frecuentemente diseñadas sobre hipótesis de normalidad. Así, los métodos del análisis multivariante pierden su capacidad inferencial cuando la inclusión de variables discretas nos aleja de la distribución normal, limitando con ello sus posibilidades al aspecto descriptivo. En el caso de la regresión, es posible incluir variables discretas o dicótomas como regresores en forma de *dummies*, aunque a costa de una reducción de los grados de libertad. Pero si es la variable respuesta la cualitativa, se presentan problemas graves que obligan a redefinir el modelo en otros términos.

Precisamente para abordar este tipo de cuestiones surge en la estadística económica una nueva rama que tiene en estos momentos un fuerte potencial de crecimiento. Aunque las técnicas básicas de la modelización estadística cualitativa estaban siendo utilizadas por biólogos en los años 50¹³, su actual desarrollo está siendo en buena medida promocionado por economistas, que se ven abocados a aplicarlas en los modelos de elección. Y en este impulso ha tenido justamente el análisis regional un papel protagonista, ya que una buena

¹³ Vid. BERKSON, J.: «Why I Prefer Logits to Probits», *Biometrics*, dec. 1951, págs. 327-359.

parte de los trabajos pioneros en esta línea ha consistido en el estudio de problemas típicamente incluidos en el campo regional¹⁴.

Los modelos estadísticos discretos abarcan una serie de técnicas que se agrupan bajo el nombre genérico de modelo lineal generalizado. Pueden distinguirse en ellas dos grandes bloques: las que tienen por objeto la elaboración de modelos destinados a explicar el comportamiento de variables económicas discretas, que son los *logit*, *probit* y sus derivados, y aquellas cuyo objetivo consiste en analizar determinadas relaciones entre variables discretas a partir de tablas de contingencia, o modelos log-lineales.

Hemos indicado ya que la inclusión de variables discretas como explicativas en un modelo de regresión no presenta problemas graves. No es éste el caso, en cambio, si la variable a explicar sólo puede tomar unos pocos valores, pues esta situación trae consigo dos inconvenientes. En primer lugar, la hipótesis de normalidad de los residuos no es admisible, lo que implica una serie de limitaciones fuertes en el proceso de estimación y en la fiabilidad de los resultados. Pero lo más grave es que con el modelo de regresión las probabilidades de cada una de las opciones —es decir, de cada uno de los valores que puede tomar la variable respuesta— pueden estar fuera el intervalo (0,1), de manera que el planteamiento carece de sentido, y la regresión ordinaria no es aplicable a este tipo de problemas.

Los modelos *logit* y *probit* son precisamente los métodos adecuados para resolver estos problemas. Su fundamento consiste en modificar la estructura lineal de la regresión ordinaria introduciendo funciones adecuadas de la combinación de regresores que «fuercen» a las probabilidades correspondientes a mantenerse dentro del intervalo (0,1). Más concretamente, supongamos que en una elección con dos alternativas se establece un modelo lineal para sus probabilidades

$$Y_i = \eta_i + \varepsilon_i$$

donde $\eta_i = x_i' \beta$ y x_i son las variables observables mediante las cuales se pretende explicar la elección. No existe, evidentemente, ninguna limitación a los valores que pueda tomar η_i . Una forma de acortarlos en el intervalo (0,1), es definir un nuevo modelo

$$y_i = g(\eta_i)$$

de manera que la función g sea una función monótona no decreciente y que

¹⁴ Vid. DOMENCICH, T. y McFADDEN, D.: *Urban Travel Demand*, North-Holland, 1975; QUIGLEY, J. M.: «Housing Demand in the Sshort Run: An Analysis of Polytomons Choice», *Explorations in Economic Research*, 3, 1976, págs. 76-102; y McFADDEN, D.: «Modelling the Choice of Residential Location in Spatial Interaction Theory and Planning Models», en Karlquist, Lundquist, Snickars y Weibull eds., *Spatial Interaction Theory and Planning Models*, North-Holland, 1978.

sólo pueda tomar valores en este intervalo. Cualquier función de distribución de probabilidad reúne estas propiedades y, por consiguiente, sirve para el caso. Así pues, nuestro modelo queda en la forma

$$Y_i = F(x_i; \beta) + \varepsilon_i$$

donde F es una función de distribución¹⁵.

En un modelo de elección, el individuo pasa de una decisión a otra cuando la utilidad, reflejada en la combinación de variables explicativas, supera un cierto umbral, distinto para cada individuo. Puesto que es razonable suponer que estos umbrales se distribuyan normalmente, parece lógico tomar como F la función de distribución normal. En este caso tenemos definido el modelo *probit*.

Si se sustituye F por la función logística, es decir, se toma

$$y_i = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}}$$

se obtiene el modelo *logit*. No existe ninguna justificación teórica para esta elección. Simplemente se ha seleccionado esta función por ser muy parecida a la normal y más sencilla de calcular. De hecho, en el caso de elección binaria el modelo *logit* es casi siempre una buena aproximación al *probit*, y mucho más sencillo de cómputo, por lo que es el que con más frecuencia se usa.

Ambos modelos son generalizables al caso de elección múltiple. Con la hipótesis de que los residuos tienen una distribución conjunta normal multivariante se obtiene el modelo *multi-probit*. Si se supone que se distribuyen independiente e idénticamente según la distribución de Weibull, tenemos el *multi-logit*, en el que

$$y_{r,1} = \frac{e^{x_i \beta_r}}{\sum_{s=1}^R e^{x_i \beta_s}} \quad r = 1, \dots, R$$

siendo R el número de alternativas. Desgraciadamente en este caso los resultados que se obtengan de la modelización *multi-logit* o de la *multi-probit* pueden ser muy diferentes. De hecho, la hipótesis de independencia de los residuos es muy fuerte, y limita la validez del modelo *logit* multivariante al caso en que se satisfaga el llamado principio de independencia de alternativas irrelevantes, que implica que la proporción existente entre las probabilidades de cada par de alternativas se mantiene constante al introducir alternativas nuevas, de manera que éstas afectan en forma equitativa a las anteriores. Ya se

¹⁵ AMEMIYA, T.: «Qualitative Response Models: A Survey», *Journal of Economic Literature*, XIX, 1981, págs. 1483-1536.

comprende que muchas veces esto no ocurre. En estos casos, el modelo *multi-logit* no es aplicable, y es preciso recurrir al *multi-probit*.

Pero la estimación de un modelo *probit* multivariante suele presentar una grandísima complejidad de cálculo. Para evitar su uso en los casos en que no se da la independencia de alternativas irrelevantes, se ha recurrido a diversas soluciones, de las que las más frecuentes son los modelos *dogit* y los *logit* anidados. Los primeros son una combinación de las dos versiones anteriores¹⁶, de manera que son aplicables a los casos en que al menos algunos pares de alternativas satisfacen el principio de independencia mencionado. Los *logit* anidados consisten en transformar un modelo de alternativas múltiples en una secuencia de elecciones binarias, en los que la elección en cada nivel está condicionada a los resultados de elecciones anteriores, y en los que las probabilidades se estiman por separado, y condicionalmente a los procesos previos¹⁷.

Una interesante generalización de los modelos *logit*, muy adecuada para explotar mejor la información aportada por datos panel, son los modelos *beta-logísticos*, que permiten analizar la heterogeneidad del comportamiento individual mediante la modelización de la distribución de probabilidad de cada una de las alternativas¹⁸.

Otra línea actual de extensión de estos modelos es su dinamización, de manera que puedan reflejar la influencia que tenga en un decisor su experiencia previa en procesos similares. Para ello, se define una función de utilidad que tenga en cuenta explícitamente el carácter intertemporal del proceso de decisión, incluyendo en ella variables que reflejen el resultado de elecciones anteriores¹⁹.

Los modelos logísticos se han venido aplicando a procesos de clasificación, en ocasiones con ventaja respecto al análisis discriminante. En efecto, cuando algunas de las variables son discretas, los resultados obtenidos con este procedimiento suelen ser algo deficientes, no tanto respecto a la proporción total de aciertos cuanto al sesgo de los errores, mucho más frecuentes en una clase que en la otra. En estos casos ocurre a menudo que el modelo *logit* produce una clasificación mejor²⁰.

¹⁶ GAUDRY, M. J. y DAGENAIS, M. G.: «The dogit model», *Transportation Research B*, 13 B, págs. 105-111.

¹⁷ SOBEL, K.: «Travel demand Forecasting with the Nested Multinomial Logit Model», *Transportation Research Record*, 775, págs. 48-55.

¹⁸ DUNN, R. y WRIGLEY, N.: «Beta-logistic Models of Urban Shopping Center Choice», *Geographical Analysis*, 17, 2, 1985, págs. 95-113.

¹⁹ Vid. FISCHER, M. y NIJAKAMP, P.: «From Static Towards Dynamic Discrete Choice Modelling», *Regional Science and Urban Economics*, 17, 1987, págs. 3-27.

²⁰ EFRON, B.: «The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis», *Journal of the American Statistical Association*, 70, 1975, págs. 892-898; y McFADDEN, D.:

Así como los modelos anteriormente descritos pueden aplicarse indistintamente con variables explicativas continuas o discretas, los modelos *log-lineales* actúan solamente sobre variables discretas, en las cuales no es necesario, por otra parte, establecer una distinción entre variables explicativas y variable respuesta. Los modelos *log-lineales* son en realidad una generalización de los análisis de independencia en variables discretas. Su objetivo consiste en detectar estructuras subyacentes en una tabla de contingencia, fundamentalmente de independencia o de equiprobabilidad²¹. Para ello se ajustan sobre los datos de la tabla tomados en logaritmos una serie de modelos lineales, en general dispuestos de forma jerárquica, empezando por el más general o saturado, es decir, aquel que considera todas las interacciones posibles entre las variables y continuando con todas las posibilidades de eliminación de parámetros que puedan irse anidando dentro de los anteriores, hasta llegar al modelo más simple posible, con un solo parámetro. De todos estos modelos, aquel que presente un mejor ajuste a los datos será el seleccionado. La ausencia, en él de un determinado parámetro de interacción implica independencia múltiple o condicional entre las variables correspondientes. Si falta el parámetro asociado con una de las variables, ello supone que sus categorías correspondientes son equiprobables de manera que esta variable es redundante y podemos prescindir de ella. Si el modelo mejor es el más simple de todos, podemos deducir una estructura de equiprobabilidad para todas las variables implicadas²².

6. LOS MODELOS ECONÓMICOS REGIONALES

Cuando se intenta reflejar los efectos de las políticas regionales en un modelo, es preciso tener en cuenta el complejo juego de interrelaciones entre los diversos tipos de variables que intervienen, de manera que en ocasiones las que son objetivos en un determinado bloque de relaciones económicas resultan actuar como instrumento en otro. En este contexto los modelos de una sola ecuación sólo pueden reflejar un planteamiento parcial excesivamente simplificado del problema, en el que se ignoren muchas de estas interrelaciones. La especificación que resulte será, por tanto, deficiente, y ello redundará en detrimento de la calidad de las estimaciones. Por todo esto, los modelos regionales se plantean mayoritariamente en la actualidad en forma multiecuacional y, generalmente, con bastantes ecuaciones.

«A Comment on Discriminant Analysis versus Logit Analysis», *Annals of Economic and Social Measurement*, 5, 4, 1976, págs. 511-519.

²¹ Vid. EVERITT, B. S.: *The Analysis of Contingency Tables*, Chapman and Hall, 1977.

²² Vid. UPTON, G. J. G.: «Log-linear Models, Screening, and Regional Industrial Surveys», *Regional Studies*, 15, k1981, págs. 33-45.

Esta tendencia a los macromodelos contribuye a poner más de manifiesto los problemas que apuntamos en el apartado de introducción como características del análisis regional. La falta de una modelización teórica previa que explique el comportamiento e influencias recíprocas de las variables deberá ser suplida con hipótesis *ad-hoc*, a veces un tanto arriesgadas. Por otra parte, la escasez de datos regionales, que obliga en ocasiones a utilizar otros que no son los más adecuados, dificulta el proceso de especificación; y no olvidemos que una especificación correcta es fundamental, ya que condiciona la calidad de los resultados, predicciones y evaluaciones que se obtengan del modelo. Pese a todo ello, esta metodología ha sido muy abundantemente empleada en los últimos años²³.

La construcción de estos modelos pasa por una serie de etapas:

- a) la modelización o información previa nos conduce a elaborar unas hipótesis;
- b) en base a ellas, se especifican las ecuaciones;
- c) se estiman sus parámetros por algún procedimiento estadístico;
- d) se contrastan los resultados obtenidos;
- e) si la etapa anterior da resultados aceptables, el modelo se utiliza con fines predictivos o de evaluación de políticas. En caso contrario, se revisa la especificación.

De todas ellas, la de especificación es la que en los modelos regionales reviste características propias, ya que puede llevarse a cabo en función de supuestos de base diferentes.

El más sencillo es el unirregional, en el que se considera solamente la región y se analizan las relaciones de las variables dentro de ella. En realidad, estos modelos siguen básicamente las pautas de los nacionales. Pero tienen el inconveniente de no poder reflejar las estructuras de interrelación de la región con su encuadre nacional —ni, por consiguiente, los efectos derivados de la aplicación de políticas adoptadas por unidades de decisión independientes en cada nivel, y que pueden eventualmente superponerse o entrar en conflicto— ni tampoco las interregionales. Además, pueden fácilmente resultar inconsistentes con los nacionales. Por esta razón, los modelos regionales tienden a adoptar orientaciones regional-nacionales o interregionales²⁴. En los primeros, se supone que las relaciones interregionales se transmiten a las regiones en cuestión a través del nivel nacional, con lo que las interrelaciones entre ellas no

²³ Para una amplia relación y descripción de macromodelos realizados en diversos países, véase el cap. VIII del libro de RODRÍGUEZ SÁIZ, J.; MARTÍN PLIEGO, J.; PAREJO GÁMIR, J. A. y ALMAQUERA GÓMEZ, A.: *Política Económica Regional*, Alianza Editorial, 1987.

²⁴ NIJKAMP, P.; RIETVELD, P. y SNICKARS, F.: «Regional and Multiregional Economic Models: a Survey», en Nijkamp, ed. *Handbook of Regional and Urban Economics*, North-Holland, 1986.

cuentan. En los segundos, en cambio aparecen representadas todas las regiones, y analizadas además sus interacciones mutuas. Aunque esta última especificación es la más completa, tiene el inconveniente de requerir para su estimación una mayor cantidad de datos; pues, mientras en un modelo regional-nacional con R regiones el número de interacciones a considerar es $2R$, en el interregional sería R^2 .

Por otra parte, las vinculaciones entre los niveles nacional y regional pueden enfocarse de dos maneras distintas, dando así lugar a los llamados modelos *bottom-up* y *top-down*²⁵. Los primeros se justifican en la creencia de que la política regional no solamente va a afectar a la redistribución económica, sino que además va a incidir en las variables objetivo nacionales, y que esta incidencia debe poder ser detectada por el modelo. Consiguientemente, se especifica éste sobre variables regionales, y las nacionales se determinan por agregación de aquéllas; ello permite analizar los efectos de las políticas regionales a ambos niveles. Por el contrario los modelos *top-down* se basan en magnitudes nacionales, de las cuales las correspondientes regionales se obtienen como un subproducto, de manera que no permiten analizar efectos de realimentación entre ambos niveles, sino que se limitan a estudiar las reacciones regionales a dinámicas nacionales. Su estructura es, de hecho, la de un modelo nacional del que últimamente se derivan resultados regionales.

Los modelos *bottom-up* son, evidentemente más completos que los *top-down* y permiten análisis más detallados de los procesos de interrelación entre políticas regionales y magnitudes nacionales. Pero los *top-down* tienen en cambio la ventaja de requerir menor cantidad de datos. Por otra parte, son fáciles de compatibilizar con los nacionales. En realidad, la mayoría de los modelos construidos no pertenecen estrictamente a ninguno de estos grupos, pues es frecuente que incluso los elaborados con criterios *bottom-up* fijen algunas de las variables a nivel nacional. Tal es, por ejemplo, el caso del famoso modelo francés Regina²⁶.

Los modelos regionales permiten evaluar los efectos de políticas alternativas. Pero este proceso puede ser especialmente complicado en el caso en que los objetivos regionales entren en conflicto con los nacionales —por ejemplo, equidad frente a crecimiento— o de que como consecuencia de la aplicación de determinadas políticas aparezcan efectos externos negativos. El problema puede resolverse con método de busca de compromisos entre objetivos en conflicto, o de decisión multiobjetivo. Es frecuente que se defina como criterio de valoración una función de las variables objetivo en cuyas pondera-

²⁵ KLEIN, L. y CLICKMAN, N. J.: «Econometric Model Building at Regional Level», *Regional Science and Urban Economics*, 7, 1977, págs. 3-23.

²⁶ COURBIS, R.: «The REGINA Model: A Regional-national Model for French Planning», *Regional Science and Urban Economics*, 9, 1979, págs. 117-139.

ciones se refleje una estructura de preferencias. En este caso, el problema se transforma en el de establecer dicha estructura, o fijar los valores de las ponderaciones²⁷.

Un problema asociado al anterior es el de la determinación de los valores que habría que dar a los instrumentos para obtener resultados óptimos. Su solución requiere el uso de las técnicas específicas de optimización. En particular, cuando el modelo toma en consideración los posibles retardos en la actuación de las políticas, adquiriendo una estructura dinámica, el método que permite obtener las trayectorias de políticas óptimas es el control óptimo.

La bondad de una política o conjunto de políticas en un modelo dinámico suele valorarse midiendo la discrepancia de la trayectoria resultante de las variables objetivo al aplicar esa política respecto a la otra trayectoria, que se establece como la deseable. En general, se toma como medida una función cuadrática de dichas desviaciones. El problema consistirá, pues, en determinar las trayectorias de los instrumentos que hagan mínima esta función, a las cuales se les llama trayectorias de control. Si éstas se fijan al comienzo de la planificación, con independencia de la información que posteriormente pueda recogerse sobre el estado del sistema, tendremos políticas en bucle abierto. Si, por el contrario, la información que se va recibiendo acerca de la evolución del proceso modifica estas políticas, tendremos una regla de control *feedback*²⁸.

En un modelo econométrico lineal con función objetivo cuadrática, el algoritmo de Chow²⁹, basado en las técnicas de programación dinámica de Bellman, permite obtener trayectorias óptimas de control lineales en las variables de estado con realimentación secuencial en cada período. Este interesante resultado se ha venido aplicando durante los últimos años a diversos modelos de planificación regional, y Tan y Bennet³⁰ lo han generalizado al caso en que las variables objetivo o instrumento estén sujetas a restricciones, lo cual permite eliminar uno de los principales inconvenientes de la función objetivo cuadrática: que las desviaciones respecto a la trayectoria deseada en los dos sentidos estén igualmente penalizadas.

²⁷ RIETVELD, P.: *Multiple objective decision methods and regional planning*, North-Holland, 1980.

²⁸ MARTÍN-GUZMÁN, M. P.: *Evaluación de políticas con modelos econométricos*, INE, 1987.

²⁹ CHOW, G.: *Analysis and Control of Dynamic Systems*, Wiley, 1975.

³⁰ TAN, K. C. y BENNETT, R. J.: *Optimal Control of Spatial Systems*, Allen and Unwin, 1984; y BENNETT, R. J. y HORDIK, L.: «Regional Economic and Dynamic Models», en Nijkamp, ed., *Handbook of Regional and Urban Economics*, North-Holland, 1986.

BIBLIOGRAFÍA

- AMEMIYA, T.: «Qualitative Response Models: A Survey», *Journal of Economic Literature*, XIX, 1981, págs. 1483-1536.
- ANCOT, J. y PAELINCK, J.: «A Discriminant Analysis Approach to Regional Trres old Problems: A Progress Report», *Papers of the Regional Science Association*, 42, 1979, págs. 239-151.
- ARNÁIZ, G.; MARTÍN-GUZMÁN, P.; MARTÍN PLIEGO, J. y TOLEDO, I.: *Discriminación y clasificación de las regiones fiscales en España*, IEF, 1987.
- BENNETT, R. J. y HORDIK, L.: «Regional Economic and Dynamic Models», en Nijkamp, ed. *Handbook of Regional and Urban Economics*, North-Holland, 1986.
- BERKSON, J.: «Why I Prefer Logits to Probits», *Biometrics*, 1951, págs. 327-359.
- CHOW, G.: *Analysis and Control of Dynamic Systems*, Wiley, 1975.
- COURBIS, R.: «The REGINA Model: A Regional-national Model for French Planning», *Regional Science and Urban Economics*, 9, 1979, págs. 117-139.
- DOMENCICH, T. y McFADDEN, D.: *Urban Travel Demand*, North-Holland, 1975.
- DUNN, R. y WRIGLEY, N.: «Beta-logistic Models of Urban Shopping Center Choice», *Geographical Analysis*, 17, 2, 1985, págs. 95-113.
- DURAN, B. y ODELL, P.: *Cluster Analysis: A Survey*, Ssprinter Verlag, 1974.
- EFRON, B.: «The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis», *Journal of the American Statistical Association*, 70, 1975, págs. 892-898.
- ERTL, H.; FISCHER, M. y WOHLSCHAGL, H.: «A Methodological Approach for Large Regionala Taxonomic Problems: Spatial Patterns of Population Development in Austria», *Papers of the Regional Science Association*, 44, 1980, págs. 119-135.
- EVERITT, B. S.: *The Analysis of Contingency Tables*, Chapman and Hall, 1977.
- FISCHER, M.: «Regional Taxonomy: A Comparison of Some Hierarchic and Non-Hierarchic Strategies», *Regional Science and Urban Economics*, 10, 1980, págs. 503-537.
- FISCHER, M. y NIJKAMP, P.: «From Static Towards Dynamic Discrete Choice Modelling», *Regional Science and Urban Economics*, 17, 1987, págs. 3-27.
- GAUDRY, M. J. y DAGENAIS, M. G.: «The Dogit Model», *Transportation Research B*, 13 B, págs. 105-111.
- GLICKMAN, N. J.: «Using Empirical Methods for Regional Policy Analysis» en Abegov, Andersson y Snickers, eds. *Regional Development Modelling: Theory and Practice*, North-Holland 1982.
- GNANADESIKAN, R.: *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, 1977.
- HARTIGAN, J. A.: *Clustering Algorithms*, Wiley, 1975.
- HORDIJK, L.: «Problems in Estimating Econometric Relations in Space», *Papers of the Regional Science Association*, 42, 1979, págs. 99-115.
- JAUAMOTTE, C. y PAELINCK, J.: «The Differential Economic Structures of the Belgian Provinces: A Time-varying Factor Analysis», *Regional and Urban Economics* 1, 1, 1971, págs. 41-75.
- JEFFREY, D. y HUBBARD, N. J.: «Foreign Tourism, the Hotel Industry and Regional Economic Performance», *Regional Studies*, 22, 4, 1988, págs. 319-329.
- KEANE, M.: «A Functional Distance Approach to Regionalisation», *Regional Studies*, 12, 1978, págs. 379-386.
- KETTENRING, J.: «Canonical Analysis of Several Sets of Variables», *Biométrica*, 58, 1971, págs. 433-451.
- KLEIN, L. y CLICKMAN, N. J.: «Econometric Model Building at Regional Level», *Regional Science and Urban Economics*, 7, 1977, págs. 3-23.
- LIAW, K. L. y LEDENT, J.: «Nested Logit Models and Maximum Quasi-likelihood Method: A Flexible Methodology for Analysing Interregional Migration Patterns», *REgional Science and Urban Economics*, 17, 1987, págs. 67-88.
- McFADDEN, D.: «A Comment on Discriminant Analysis versus Logit Analysis», *Annals of Economic and Social Measurement*, 5, 4, 1976, págs. 511-519.
- «Modelling the Choice of Residential Location in Spatial Interaction Theory and Planning Mo-

- dels», en Karqvist, Lundqvist, Snickars y Weibull, eds., *Spatial Interaction Theory and Planning Models*, North-Holland, 1978.
- MANSKI, C. y SALOMON, I.: «The Demand for Teleshopping: An Application of Discrete Choice Models», *Regional Science and Urban Economics*, 17, 1987, págs. 109-121.
- MARTÍN-GUZMÁN, P.: *Evaluación de políticas con modelos econométricos*, INE, 1987.
- MILLER, A.: «A Taxonomy of Nebraska County Economies: An Application of the Optimization Approach to Identifying a System of Regions», *Journal of Regional Science*, 16, 1976, págs. 225-236.
- MILLER, R.: «The Jackknife: A Review», *Biométrica*, 61, 1974.
- NIJKAMP, P.; RIETVELD, P. y SNICKARS, F.: «Regional and Multiregional Economic Models: A Survey», en Nijkamp, ed., *Handbook of Regional and Urban Economics*, North-Holland, 1986.
- O'FARRELL, P. y CROUCHLEY, R.: «The Locational Pattern of New Manufacturing Establishments: An Application of Discriminant Analysis», *Regional Studies*, 13, 1979, págs. 39-59.
- PETERS, W. S. y BUTLER, J. Q.: «The Construction of Regional Economic Indicators by Principal Components», *Annals of Regional Science*, 4, 1, 1970, págs. 1-14.
- PINDER, D.: *Regional Economic Development and Policy*, Allen & Unwin, 1983.
- QUIGLEY, J. M.: «Housing Demand in the Short Run: An Analysis of Politomous Choice», *Explorations in Economic Research*, 3, 1976, págs. 76-102.
- RANNER, A. P.: «Analysis Relations Between Regions and their Surroundings: A Canonical Correlation Approach», *Regional and Urban Economics*, 4, 1974, págs. 141-162.
- RICHARDSON, D. y THALHEIMER, R.: «On the Use of Grouping Methods in the Analysis of Residential Housing Markets», *Regional Science and Urban Economics*, 12, 1982, págs. 285-304.
- RIETVELD, P.: *Multiple Objective Decision Methods and Regional Planning*, North-Holland, 1980.
- RODRÍGUEZ SÁIZ, L.; MARTÍN PLIEGO, J.; PAREJO GÁMIR, J. A. y ALMAGUERA GÓMEZ, A.: *Política Económica Regional*, Alianza Editorial, 1987.
- SMITH, C.: «Some Examples of Discrimination», *Ann. Eugen*, 13, 1947, págs. 272-282.
- SNICKARS, F.: «Interregional Linkages in Multiregional Economic Models», en Issaev, Nijkamp, Rietveld y Snickars, eds., *Multiregional Economic Modelling: Practice and Prospect*, North-Holland, 1982.
- SOBEL, K.: «Travel Demand Forecasting with the Nested Multinomial Logit Model», *Transportation Research Record*, 775, 1980, págs. 48-55.
- TAN, K. C. y BENNET, R. J.: *Optimal Control of Spatial Systems*, Allen & Unwin, 1984.
- UPTON, G.: «Log-linear Models, Screening and Regional Industrial Surveys», *Regional Studies*, 15, 1981, págs. 33-45.
- VAN ROMPUY, P.: «A Two-step Procedure for Intertemporal Factor Analysis», *Regional Science and Urban Economics*, 9, 1979, págs. 261-274.