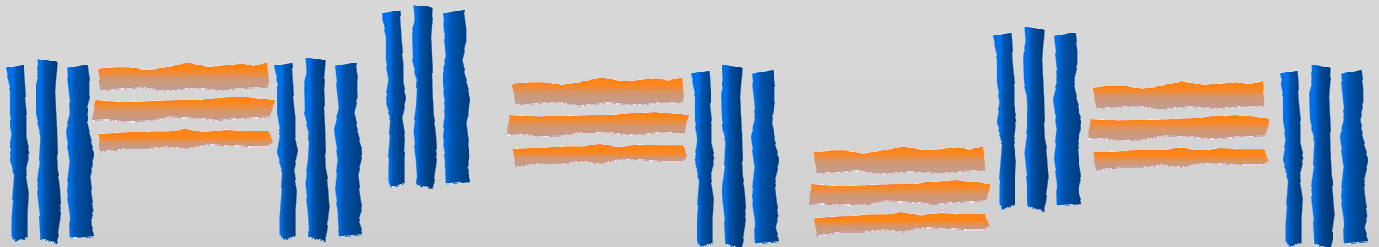


REVISTA DE
MÉTODOS CUANTITATIVOS
PARA LA ECONOMÍA Y LA EMPRESA



UNIVERSIDAD
**PABLO_D
OLAVIDE**
SEVILLA

Número 7
Junio de 2009
ISSN: 1886-516X
D.L: SE-2927-06

**REVISTA DE MÉTODOS CUANTITATIVOS
PARA LA ECONOMÍA Y LA EMPRESA**

**Journal of Quantitative Methods for
Economics and Business Administration**

Número 7. Junio de 2009.

ISSN: 1886-516X. D.L: SE-2927-06.

URL: <http://www.upo.es/RevMetCuant/>

Editores:

Dr. Eugenio M. Fedriani Martel
Universidad Pablo de Olavide
Ctra. de Utrera, Km 1 - 41013
Sevilla (Spain).
Correo-e: efedmar@upo.es

Dr. Alfredo García Hernández-Díaz
Universidad Pablo de Olavide
Ctra. de Utrera, Km 1 - 41013
Sevilla (Spain).
Correo-e: agarher@upo.es

Comité Editorial:

Dr. S. Ejaz Ahmed, University of Windsor, Ontario (Canadá)
Dr. Carlos A. Coello Coello, CINVESTAV-IPN, México D.F. (México)
Dr. Ignacio Contreras Rubio, Universidad Pablo de Olavide, Sevilla (España)
Dra. Manuela Coromaldi, University of Social Sciences UNISU, Roma (Italia)
Dr. Miguel Ángel Hinojosa Ramos, Universidad Pablo de Olavide, Sevilla (España)
Dr. Matías Irigoyen Testa, Universidad Nacional del Sur, Buenos Aires (Argentina)
Dr. M. Kazim Khan, Kent State University, Ohio (EE.UU.)
Dr. Manuel Laguna, University of Colorado at Boulder, Colorado (EE.UU.)
Dra. María Amparo León Sánchez, Universidad de Pinar del Río (Cuba)
Dr. Jesús López-Rodríguez, Universidad de A Coruña, (España)
Dr. Cecilio Mar Molinero, University of Kent, Canterbury (Reino Unido)
Dra. Ana M. Martín Caraballo, Universidad Pablo de Olavide, Sevilla (España)
Dra. M^a Carmen Melgar Hiraldo, Universidad Pablo de Olavide, Sevilla (España)
Dr. José Antonio Ordaz Sanz, Universidad Pablo de Olavide, Sevilla (España)
Dr. Andrés Rodríguez-Pose, London School of Economics, Londres (Reino Unido)
Dr. José Manuel Rueda Cantuche, IPTS-DG J.R. Centre-European Commission
Dra. Mariagrazia Squicciarini, VTT Technical Research Centre (Finlandia)
Dra. Mariangela Zoli, Università degli Studi di Roma 'Tor Vergata', Roma (Italia)

Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus

MUÑOZ ROSAS, JUAN FRANCISCO

Departamento de Métodos Cuantitativos para la Economía y la Empresa
Universidad de Granada

Correo electrónico: jfmunoz@ugr.es

ÁLVAREZ VERDEJO, ENCARNACIÓN

Departamento de Métodos Cuantitativos para la Economía y la Empresa
Universidad de Granada

Correo electrónico: encarniav@ugr.es

RESUMEN

La aparición de datos faltantes es un problema común en la mayoría de las encuestas llevadas a cabo en distintos ámbitos. Una técnica tradicional y muy conocida para el tratamiento de datos faltantes es la imputación. La mayoría de los estudios relacionados con los métodos de imputación se centran en el problema de la estimación de la media y su varianza y están basados en diseños muestrales simples tales como el muestreo aleatorio simple. En este trabajo se describen los métodos de imputación más conocidos y se plantean bajo el contexto de un diseño muestral general y para el caso de diferentes mecanismos de respuesta. Mediante estudios de simulación Monte Carlo basados en datos reales extraídos del ámbito de la economía y la empresa, analizamos las propiedades de varios métodos de imputación en la estimación de otros parámetros que también son utilizados con frecuencia en la práctica, como son las funciones de distribución y los cuantiles. Con el fin de que los métodos de imputación descritos en este trabajo se puedan implementar y usar con mayor facilidad, se proporcionan sus códigos en los lenguajes de programación R y Splus.

Palabras clave: información auxiliar; encuesta; probabilidades de inclusión; mecanismo de respuesta.

Clasificación JEL: C13; C15; C80.

2000MSC: 62D05.

Imputation methods to handle the problem of missing data: an application using R/Splus

ABSTRACT

Missing values are a common problem in many sampling surveys, and imputation is usually employed to compensate for non-response. Most imputation methods are based upon the problem of the mean estimation and its variance, and they also assume simple sampling designs such as the simple random sampling without replacement. In this paper we describe some imputation methods and define them under a general sampling design. Different response mechanisms are also discussed. Assuming some populations based upon real data extracted from the context of the economy and business, Monte Carlo simulations are carried out to analyze the properties of the various imputation methods in the estimation of parameters such as distribution functions and quantiles. The various imputation methods are implemented using the popular statistical softwares R and Splus, and codes are here presented.

Keywords: auxiliary information; survey; inclusion probabilities; response mechanism.

JEL classification: C13; C15; C80.

2000MSC: 62D05.



1. Introducción

En el mundo económico y empresarial es conocida la utilización de las encuestas como herramientas para recoger la información necesaria para llevar a cabo estudios de muy diversa índole. La calidad de los resultados obtenidos y la posibilidad de generalización de los mismos dependen de la composición de la muestra, del tipo de encuesta y de la tasa de respuesta de la misma. A su vez, también se ha constatado como un problema importante la aparición de datos faltantes en determinados *ítem* del cuestionario. En otras palabras, la aparición de valores perdidos o la falta de respuesta es un problema común presente en cualquier estudio, especialmente en el ámbito de las ciencias sociales (véase Rubin, 1996).

Una técnica tradicional y muy conocida para el tratamiento de datos faltantes es la imputación. Las técnicas de imputación se pueden clasificar, en primer lugar, en dos grandes grupos: las técnicas de imputación simples y las de imputación múltiple.

Las técnicas simples de imputación han sido una de las herramientas más conocidas y aceptadas para el tratamiento de la falta de respuesta (véase Sedransk, 1985, Kalton y Kasprzyk, 1986 y Little y Rubin, 2002). Las técnicas simples de imputación presentan algunas ventajas frente a las técnicas de imputación múltiple. Por ejemplo, las técnicas simples tienen una implantación más sencilla sin que por el contrario sufran una importante pérdida de eficiencia en comparación con las técnicas de imputación múltiple. Por último, destacamos que las técnicas simples de imputación se pueden dividir en dos categorías: aleatorias y determinísticas.

El uso de la imputación puede provocar problemas serios de subestimación de la verdadera varianza cuando la proporción de datos faltantes es apreciable (Rao y Shao 1992). En general, un método de imputación aleatorio tiene la ventaja de añadir una mayor variabilidad a través de las imputaciones que un método determinístico de imputación; es decir, las técnicas simples determinísticas de imputación generalmente subestiman más las varianzas que las técnicas simples aleatorias de imputación. Sin embargo, las técnicas determinísticas proporcionan, en general, estadísticos más precisos que las técnicas aleatorias.

La imputación múltiple fue propuesta por Rubin (1978) como una alternativa a las técnicas simples de imputación. La imputación múltiple requiere la construcción de M (≥ 2) conjuntos de datos completos, los cuales se obtienen reemplazando cada dato faltante por M valores imputados, obtenidos mediante el mismo procedimiento de imputación. Aunque la imputación múltiple es una aproximación muy potente, sufre algunas limitaciones que no debemos

pasar por alto. Por ejemplo, Fay (1991) señala que la imputación múltiple puede conducir a estimadores de la varianza inconsistentes en el caso de encuestas multietápicas estratificadas. Rao y Shao (1992) afirman que las agencias estadísticas prefieren, en general, el uso de la imputación simple, debido especialmente a las dificultades operacionales que supone el mantenimiento de conjuntos de datos múltiples. Este problema se hace aún más grave en el caso de muestras con un tamaño muestral muy elevado, como por ejemplo en el caso de muestras nacionales. Rao (1996) también destaca algunas otras desventajas de la imputación múltiple en comparación con los métodos simples de imputación.

A pesar de la diversidad y variedad de técnicas de imputación que existen en la literatura, la mayoría de las aportaciones se han centrado en el impacto de las imputaciones realizadas sobre la estimación de la media poblacional y su varianza. Por ejemplo, Bello (1993) realizó un estudio de simulación para comparar varios métodos de imputación. Bello utilizó como criterio de comparación la precisión de los distintos métodos de imputación en la estimación de la media de la variable de interés y su varianza. Sin embargo, en la práctica existen otros parámetros, tales como la función de distribución y los cuantiles, que tienen una cantidad importante de aplicaciones, pero que no han sido estudiados en el contexto de datos faltantes y su tratamiento mediante imputación. Por otra parte, la mayoría de los estudios relacionados con métodos de imputación están basados en el muestreo aleatorio simple.

El objetivo de este trabajo es plantear las técnicas más conocidas de imputación simple bajo el esquema de un diseño general y analizarlas en el problema de la estimación de numerosos parámetros, incluyendo la media poblacional, la función de distribución y los cuantiles. Este estudio también analiza los distintos métodos de imputación, tanto para mecanismos de respuesta uniforme como para mecanismos de respuesta no uniforme. Además, se proporcionan las funciones o códigos en los lenguajes de programación R y Splus de los distintos métodos de imputación descritos en este trabajo, de modo que el lector pueda implementar y usar cada uno de los métodos descritos en este trabajo.

R es un entorno de análisis y programación estadística, compatible con el popular y comercial software Splus, que está atrayendo cada vez más a un alto número de usuarios. Una gran ventaja de R frente a otros entornos es que se trata de un lenguaje gratuito y disponible en la dirección

<http://www.r-project.org> (1)

En realidad, R es un lenguaje que puede ser usado y distribuido libremente bajo los términos de Free Software Foundation's GNU General Public License en forma de código fuente, lo que implica que un gran número de personas

colaboran en su desarrollo y actualización. Aunque R es un lenguaje similar a otros lenguajes de programación muy conocidos y usados, como Fortran o C++, su nivel de ejecución, al igual que MATLAB (Merino y Vadillo, 2007), es muy superior, incluyendo numerosas operaciones con un solo comando.

Por su parte, Splus es un programa comercial distribuido por MATHSOFT Corporation, que incluye un interfaz bastante complejo y alta capacidad gráfica. A diferencia de R, Splus no es gratuito y tiene el inconveniente del coste de sus licencias.

En resumen, tanto R como Splus son entornos de programación con un lenguaje orientado a objetos y concebidos, originalmente, para ser utilizados en aplicaciones estadísticas. No obstante, en la actualidad, R y Splus son dos lenguajes poderosos y flexibles, que resultan suficientes para la resolución de la mayor parte de los problemas estadísticos habituales y sus aplicaciones en distintos ámbitos. En la página web oficial de R, dada en (1), pueden consultarse una serie de manuales actualizados y adecuados a todo tipo de usuarios. Además, podemos consultar Ihaka y Gentleman (1996), Arcos *et al.* (2004) y Arcos *et al.* (2005) como referencias más específicas. Todas las referencias anteriores podrían utilizarse para programar en Splus, puesto que en ambos lenguajes de programación, en general, se pueden utilizar los mismos comandos y funciones. No obstante, también podemos consultar Everitt (1994) y Krause y Olson (2005) como referencias más orientadas a Splus.

El presente artículo se estructura del siguiente modo. En la Sección 2 se describen los métodos simples de imputación más conocidos y usados. En la Sección 3 se llevan a cabo estudios de simulación Monte Carlo para estudiar el impacto de distintos métodos de imputación en la estimación de diferentes parámetros en el caso de un mecanismo de respuesta uniforme. Los estudios de simulación están basados en datos reales extraídos del ámbito de la Economía y la Empresa. Un mecanismo de respuesta uniforme es poco frecuente en la práctica; es decir, es bastante común encontrarse que las unidades muestrales fallan para proporcionar una respuesta con una determinada probabilidad. De este modo, en la Sección 4 se realizan algunas observaciones sobre los mecanismos de respuesta no uniforme. El comportamiento de los distintos métodos de imputación en el caso de mecanismos de respuesta no uniforme puede consultarse en la Sección 5. Este artículo también contiene dos apéndices. En el Apéndice A se incluyen los códigos en los lenguajes R y Splus de los distintos métodos de imputación descritos en este trabajo. Por su parte, en el Apéndice B mostramos, mediante un ejemplo, cómo podemos utilizar cualquiera de los mencionados métodos a partir de un vector con datos faltantes.

2. Métodos de imputación simple para un diseño muestral general

Consideremos una población finita $U = \{1, 2, \dots, N\}$, con N unidades, de la cual se extrae, mediante un determinado diseño muestral, una muestra aleatoria s_n de tamaño n . Denotaremos como π_i a la probabilidad de que la unidad i pertenezca a la muestra s_n . Esta probabilidad π_i es conocida popularmente como probabilidad de inclusión de primer orden. El peso básico del diseño asociado a la i -ésima unidad vendrá expresado por $d_i = \pi_i^{-1}$. Por otra parte, y_i es el valor de la variable de interés y para la i -ésima unidad. En este trabajo también asumiremos que existe una variable auxiliar x asociada con la variable y . La extensión al caso de varias variables auxiliares de los distintos métodos de imputación discutidos en esta sección es un problema muy simple, que está cubierto por la literatura, y de ahí que nos centremos por simplicidad en el caso de una única variable auxiliar.

En los estudios relacionados con las encuestas por muestreo se asume, en general, que todas las respuestas en la muestra s_n son conocidas. Sin embargo, esta situación puede no presentarse en la práctica; es decir, es frecuente que en las encuestas se dispongan de valores faltantes por alguna determinada razón, como por ejemplo, la negativa del encuestado a dar la información requerida, la imposibilidad de contactar con el individuo encuestado, la pérdida casual de información, etc. De este modo, suponemos que r de los n valores de la variable y son observados (*respondientes*), mientras que el resto de $m = n - r$ valores de y corresponden a datos faltantes (*no respondientes*). Las muestras $s_r = \{i \in s_n \mid y_i \text{ es observado}\}$ y $s_m = \{i \in s_n \mid y_i \text{ no es observado}\}$ denotarán, por tanto, los conjuntos de respondientes y no respondientes asociados con la variable y . Cuando $i \in s_m$, el valor y_i necesita ser imputado, mientras las técnicas de imputación no serán necesarias en la muestra s_r . La proporción de datos faltantes se denotará por $p = m/n$. Por último, $Y_\alpha = \inf\{t : F(t) \geq \alpha\}$ denotará el cuantil poblacional de orden α de la variable y , donde

$$F(t) = \frac{1}{N} \sum_{i \in U} \Delta(t - y_i)$$

es la función de distribución poblacional, $\Delta(a) = 1$ si $a \geq 0$ y $\Delta(a) = 0$ en otro caso.

El objetivo de esta sección es plantear, para un diseño muestral general, las técnicas de imputación más conocidas y utilizadas, de modo que se puedan imputar los m valores faltantes de la variable y en la muestra s_n , y poder obtener la estimación de un determinado parámetro o realizar un análisis estadístico general usando los n valores de la muestra. Por ejemplo, el estimador de la media poblacional $\bar{Y} = N^{-1} \sum_{i \in U} y_i$ basado en los n valores de la muestra

está dado por

$$\bar{y}_I = \frac{1}{\sum_{i \in s_n} d_i} \sum_{i \in s_n} d_i \tilde{y}_i, \quad (2)$$

donde $\tilde{y}_i = y_i$ si $i \in s_r$, $\tilde{y}_i = y_i^*$ si $i \in s_m$ y y_i^* es el valor imputado o donante para el dato faltante y_i .

El *método de imputación de la media* o *método de sustitución* consiste en utilizar la media muestral de los valores disponibles como donante en cada uno de los valores perdidos; es decir, los valores imputados por el método de la media están dados por $y_i^* = \bar{y}_r$, $i \in s_m$, donde

$$\bar{y}_r = \frac{1}{\sum_{i \in s_r} d_i} \sum_{i \in s_r} d_i y_i.$$

Este método es, sin duda, el más simple pero también el menos atractivo de los distintos métodos de imputación. La única ventaja de este método es que proporciona estimaciones insesgadas para la media poblacional, pero sólo en el caso de un mecanismo de respuesta uniforme. En el lado opuesto, este método de imputación distorsiona considerablemente la distribución de los datos, debido a la concentración de valores en torno a la media. Algunas consecuencias de este hecho, tal como se analiza en la Sección 3, son la presencia de sesgos muy elevados en la estimación de cuantiles y una considerable subestimación del verdadero valor de la varianza.

Una modificación del método de imputación anterior fue propuesta por Cohen (1996). Cohen propuso añadir más variabilidad a los valores imputados mediante el método de la media usando la variabilidad de los datos muestrales. Asumiendo un diseño muestral general, el método de Cohen consiste en imputar la mitad de los valores faltantes con los valores

$$\bar{y}_r + \sqrt{\frac{n+r-1}{r-1}} \hat{\sigma}_r,$$

y la otra mitad de valores faltantes con los valores

$$\bar{y}_r - \sqrt{\frac{n+r-1}{r-1}} \hat{\sigma}_r,$$

donde

$$\hat{\sigma}_r^2 = \frac{1}{\sum_{i \in s_r} d_i} \sum_{i \in s_r} d_i (y_i - \bar{y}_r)^2. \quad (3)$$

El método NNI (acrónimo de *Nearest Neighbor Imputation*) es utilizado en numerosas encuestas llevadas a cabo por el Instituto de Estadística de Canadá y

algunas agencias nacionales de Estados Unidos. Este método utiliza el criterio del valor más próximo asociado a una variable auxiliar para proporcionar los valores imputados o donantes. En el caso de que la variable auxiliar disponga de varios valores equidistantes, se presenta el problema de la presencia de varios donantes para un mismo valor faltante. En esta situación, el criterio que se sigue es elegir aleatoriamente a un donante entre los diferentes candidatos. Una revisión más detallada del método NNI, así como numerosos resultados teóricos relacionados con este método, pueden consultarse en Chen y Shao (2000).

El *método de la razón* (véase Rao, 1996) es otro método de imputación determinística popularmente conocido y usado con bastante frecuencia en numerosos estudios. El método de la razón utiliza las cantidades

$$y_i^* = \frac{\bar{x}_r}{\bar{y}_r} x_i, \quad i \in s_m \quad (4)$$

como valores imputados, donde

$$\bar{x}_r = \frac{1}{\sum_{i \in s_r} d_i} \sum_{i \in s_r} d_i x_i.$$

Este método está basado en el hecho de que los valores definidos en (4) son los mejores predictores bajo un modelo de superpoblación que presente las siguientes características

$$E(y_i) = \beta x_i, \quad V(y_i) = \sigma^2 x_i. \quad (5)$$

Cuando el modelo (5) no sostenga, una alternativa es utilizar el *método de regresión* (véase Healy y Westmacott, 1956), el cual se basa en el modelo de superpoblación

$$E(y_i) = \alpha + \beta x_i, \quad V(y_i) = \sigma^2. \quad (6)$$

Asumiendo que el modelo (6) se ajusta razonablemente bien a los datos en estudio, el método de regresión utiliza como donantes los valores

$$y_i^* = \bar{y}_r + \hat{\beta}(x_i - \bar{x}_r), \quad i \in s_m, \quad (7)$$

donde el estimador $\hat{\beta}$ en la ecuación (7) viene dado por

$$\hat{\beta} = \frac{\sum_{i \in s_r} d_i (x_i - \bar{x}_r)(y_i - \bar{y}_r)}{\sum_{i \in s_r} d_i (x_i - \bar{x}_r)^2}.$$

Notamos que los métodos de imputación anteriores se pueden clasificar como determinísticos, mientras que los métodos de imputación que describimos a continuación corresponden a métodos de imputación aleatoria.

El método de imputación conocido popularmente como *Random Hot Deck* (RHD) es uno de los métodos de imputación más utilizados en la práctica. Este método consiste en seleccionar mediante muestreo aleatorio simple con reemplazamiento m valores a partir de los r valores disponibles de la variable y . Este planteamiento, sin embargo, tan solo resulta apropiado cuando la muestra s_n es extraída bajo muestreo aleatorio simple. En el caso de un diseño muestral general, el método RHD necesita ser modificado para que los m valores seleccionados aleatoriamente tengan en cuenta el efecto del diseño muestral. De este modo, el método RHD puede usarse eficientemente en un diseño muestral general si la muestra de m valores es seleccionada con reemplazamiento y con probabilidades de selección

$$u_i = \frac{d_i}{\sum_{i \in s_r} d_i}, \quad i \in s_r.$$

Los métodos de razón y regresión descritos en (4) y (7) pueden obtener estimaciones que subestimen la verdadera varianza de la variable de interés. Por esta razón, es usual añadir pequeñas perturbaciones aleatorias a los valores predichos obtenidos en los métodos de razón y regresión. Estas perturbaciones aleatorias pueden generarse a partir de una distribución normal con media cero y varianza $\hat{\sigma}_r^2$, donde $\hat{\sigma}_r^2$ está definida en la ecuación (3).

Los métodos de imputación RHD y NNI tienen la ventaja frente al resto de métodos descritos en esta sección el utilizar como donantes a los propios valores de los respondientes; es decir, los métodos RHD y NNI utilizan valores observados para las imputaciones. Esta propiedad es especialmente atractiva en el caso de variables discretas, puesto que los donantes tomarán también valores discretos. En el lado opuesto, con los métodos de la media, Cohen, razón y regresión, los valores imputados no tienen por qué ser discretos.

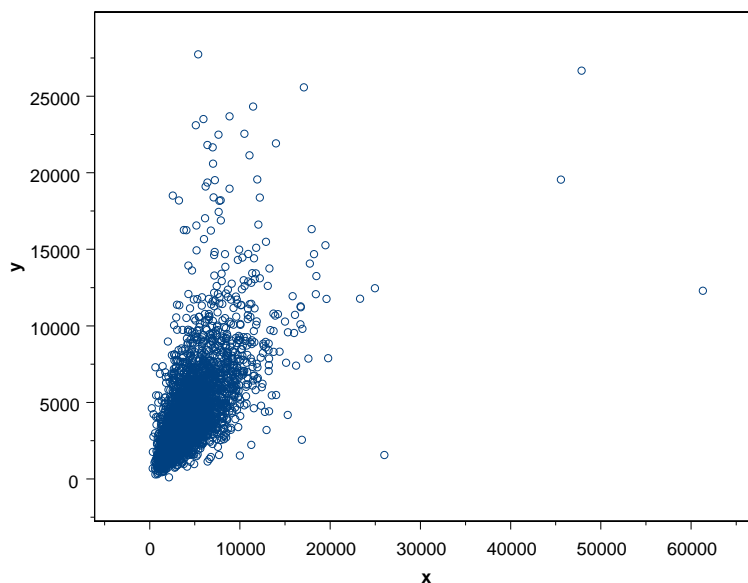
3. Comparación mediante estudios de simulación

En esta sección se comparan numéricamente los distintos métodos de imputación descritos en la sección anterior. Usaremos dos poblaciones reales para el estudio mediante simulación Monte Carlo de los distintos métodos de imputación en la estimación de varios parámetros que pueden presentarse en la práctica.

En primer lugar, los métodos de imputación se compararán con datos extraídos

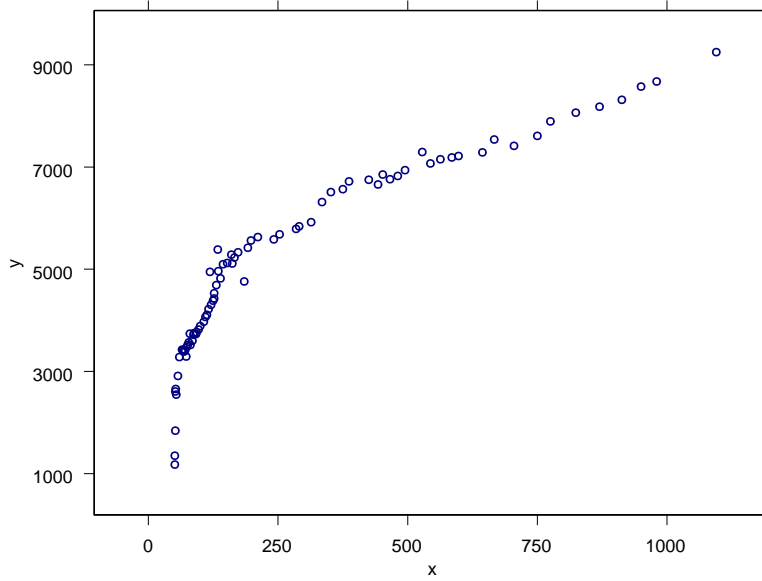
de la Encuesta Continua de Presupuestos Familiares (ECPF) del año 1997, que elabora el Instituto Nacional de Estadística (INE) de manera periódica. Esta población consta de $N = 3.114$ individuos, donde la variable de interés son los gastos familiares, mientras que la variable auxiliar son los ingresos familiares. Notamos que, debido a la naturaleza de dichas variables, la presencia de datos faltantes es bastante común en este caso; es decir, la imputación resulta ser una solución práctica a este problema. La Figura 1 muestra la nube de puntos de las variables de la población ECPF. En segundo lugar, también analizamos las distintas técnicas de imputación en la población Factories, analizada en Murthy (1967) y Kuk y Mak (1993). Para esta población, la variable auxiliar es el número de trabajadores y la variable de interés el *output* de cada factoría. La nube de puntos de las variables de la población Factories puede consultarse en la Figura 2.

Figura 1. Nube de puntos de las variables x e y de la población ECPF.



Para realizar el estudio de simulación Monte Carlo seleccionamos 1.000 muestras bajo muestreo estratificado con afijación uniforme en cada una de las dos poblaciones, siguiendo a Chambers y Dunstan (1986) y Rao *et al.* (1990). Al igual que en las referencias anteriores, se utilizó el criterio de equipartición basado en la variable auxiliar x para la formación de los estratos (véase también Silva y Skinner, 1995). En cada una de las muestras se seleccionaron de manera uniforme y aleatoria m datos de la variable y . Dichas unidades se consideraron como valores perdidos y se imputaron mediante las distintas técnicas de imputación descritas en este trabajo, de modo que a partir de los datos muestrales (incluyendo las imputaciones) se estimaron diferentes parámetros de interés. Se consideraron valores de m de forma que la propor-

Figura 2. Nube de puntos de las variables x e y de la población Factories.



ción de datos faltantes en cada una de las poblaciones fuese $p = 0,1, 0,3, 0,5$, siendo $p = m/n$. El comportamiento de las estimaciones realizadas se midió en términos de Sesgo Relativo (SR) y Error Cuadrático Médio Relativo (ECMR), donde SR y ECMR se definen como

$$SR = \frac{1}{\theta} \left[\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta) \right] \quad ; \quad ECMR = \frac{1}{\theta} \left[\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^2 \right]^{\frac{1}{2}},$$

donde θ es el parámetro de interés y $\hat{\theta}_i$ es el valor de un dado estimador $\hat{\theta}$ para la i -ésima muestra simulada.

Los distintos parámetros utilizados para evaluar el comportamiento de los métodos de imputación fueron la media poblacional, \bar{Y} , la varianza de la variable de interés, $S_y^2 = N^{-1} \sum_{i=1}^N (y_i - \bar{Y})^2$, los cuartiles poblacionales (es decir, $Y_{0,25}$, $Y_{0,5}$ y $Y_{0,75}$) y, por último, la función de distribución evaluada en los anteriores cuartiles poblacionales.

Notamos que los parámetros anteriores son bastantes frecuentes en la práctica. Por ejemplo, la media muestral es el parámetro más común en cualquier estudio mediante encuestas por muestreo. La función de distribución posee propiedades interesantes, como el hecho de describir las características más importantes de una distribución, o que otras medidas tal como los cuantiles pueden obtenerse a partir de la función de distribución. Por último, los cuantiles son también requeridos a menudo en la práctica. Por ejemplo, los cuantiles son altamente usados en muchas encuestas para obtener medidas de pobreza,

tal como la proporción de bajos ingresos (Berger y Skinner, 2003).

Los métodos de imputación involucrados en el estudio de simulación Monte Carlo son los siguientes. Asumiendo el caso de ausencia de información auxiliar, consideramos el método de imputación de la media, el método de Cohen y el método RHD. Además, con el fin de analizar si la imputación produce beneficios en términos de estimación, obtendremos estimaciones de los distintos parámetros a partir de los datos disponibles, sin usar imputaciones; es decir, consideraremos \bar{y}_r para estimar la media poblacional \bar{Y} ,

$$\hat{S}_r^2 = \frac{1}{\sum_{i \in s_r} d_i} \sum_{i \in s_r} d_i (y_i - \bar{y}_r)^2$$

para estimar S_y^2 ,

$$\hat{Y}_r(\alpha) = \inf\{t : \hat{F}_r(t) \geq \alpha\}$$

para estimar Y_α , con $\alpha = 0,25, 0,5, 0,75$ y

$$\hat{F}_r(t) = \frac{1}{\sum_{i \in s_r} d_i} \sum_{i \in s_r} d_i \Delta(t - y_i)$$

para estimar $F(t)$, con $t = Y_{0,25}, Y_{0,5}, Y_{0,75}$. A este método lo denotaremos como SI (sin usar imputaciones) en las distintas tablas de esta sección. Asumiendo información auxiliar en la etapa de estimación, consideramos el método NNI, los métodos de razón y regresión, y los métodos de razón y regresión añadiéndoles una perturbación aleatoria para corregir el problema de la subestimación en las varianzas. Estos últimos métodos se denotarán, respectivamente, como Razón.PE y Reg.PE.

Las Tablas 1 y 2 muestran los resultados obtenidos del estudio de simulación. El método de Cohen proporciona estimaciones con sesgos muy altos, que desvirtúan considerablemente las tablas, y de ahí que este método esté omitido en las tablas.

De la Tabla 1 (población ECPF) puede observarse que el método de la media produce grandes sesgos en la estimación de los distintos parámetros. Especialmente, el método de media subestima la varianza. El método de regresión también produce, en general, grandes sesgos. Este hecho quizás se deba a que la población en estudio sea más apropiada para el método de razón que para el método de regresión. Los métodos de razón y regresión que incorporan una perturbación aleatoria también proporcionan estimaciones con sesgos altos, especialmente en el caso de una alta proporción de datos faltantes. El resto de métodos de imputación obtienen estimaciones con sesgos dentro de un rango razonable de valores. En lo que respecta a la eficiencia, puede comprobarse que el método de razón es el más eficiente en la mayoría de los casos.

En la Tabla 2 pueden consultarse los valores de SR y ECMR de los distintos

Tabla 1
Valores de $ECMR \times 100$ (y $SR \times 100$) asociados a distintos métodos de imputación en la población ECPF. Las muestras fueron seleccionadas bajo muestreo estratificado aleatorio con afijación uniforme y tamaño $n = 150$. Los 3 estratos utilizados se formaron mediante el criterio de equipartición.

p	Parámetro	Sin usar x			Usando x				
		SI	Media	RHD	NNI	Razón	Reg	Razón.PE	Reg.PE
0.1	\bar{Y}	5.7 (-0.1)	5.7 (-0.1)	5.9 (-0.3)	5.6 (-0.3)	5.6 (-0.2)	5.6 (-0.2)	5.8 (-0.3)	5.8 (-0.2)
	$Y_{0,25}$	7.4 (0.0)	9.2 (5.4)	7.7 (0.0)	7.4 (-0.2)	7.1 (0.1)	8.0 (3.3)	7.6 (-1.7)	7.6 (-0.9)
	$Y_{0,5}$	5.9 (-0.2)	10.1 (7.8)	6.3 (-0.3)	6.1 (-0.2)	5.8 (-0.1)	5.8 (0.9)	5.9 (0.4)	6.0 (0.7)
	$Y_{0,75}$	6.5 (-0.1)	8.0 (-4.9)	6.9 (-0.3)	6.6 (-0.2)	6.4 (-0.2)	6.5 (-1.2)	6.8 (1.6)	6.9 (1.6)
	$F(Y_{0,25})$	15.1 (0.0)	16.9 (-10.0)	15.8 (0.3)	15.4 (0.4)	14.7 (-0.2)	16.0 (-6.6)	14.9 (3.3)	14.7 (1.7)
	$F(Y_{0,5})$	8.7 (0.4)	12.5 (-9.6)	9.1 (0.6)	8.7 (0.5)	8.4 (0.2)	8.9 (-1.3)	8.2 (-0.4)	8.3 (-0.8)
	$F(Y_{0,75})$	5.0 (0.0)	5.6 (3.4)	5.2 (0.2)	5.0 (0.1)	4.9 (0.1)	5.0 (0.8)	5.1 (-1.2)	5.1 (-1.2)
	S_y^2	29.5 (-0.9)	28.7 (-10.8)	30.3 (-1.1)	30.2 (-1.4)	32.1 (-1.3)	29.9 (-6.4)	35.1 (7.9)	32.0 (3.0)
0.3	\bar{Y}	6.6 (0.0)	6.6 (0.0)	7.0 (-0.5)	6.6 (-0.7)	6.5 (0.0)	6.3 (0.0)	7.1 (0.0)	7.0 (0.1)
	$Y_{0,25}$	8.5 (0.0)	22.0 (20.0)	9.1 (-0.3)	8.9 (-0.3)	7.8 (0.9)	13.7 (10.1)	10.2 (-5.1)	9.5 (-2.7)
	$Y_{0,5}$	6.8 (-0.1)	19.4 (17.7)	7.5 (-0.4)	7.3 (-0.4)	6.5 (0.6)	7.2 (3.2)	7.4 (2.1)	7.9 (3.2)
	$Y_{0,75}$	7.5 (0.0)	17.0 (-15.9)	8.1 (-0.3)	7.9 (-0.4)	7.1 (0.1)	7.4 (-2.6)	9.9 (6.0)	9.7 (5.6)
	$F(Y_{0,25})$	18.0 (0.1)	32.5 (-30.0)	19.3 (0.9)	18.6 (0.8)	16.3 (-1.5)	26.6 (-19.9)	18.4 (9.0)	17.0 (4.7)
	$F(Y_{0,5})$	10.0 (0.4)	30.7 (-29.5)	10.9 (0.8)	10.6 (1.0)	9.1 (-0.5)	12.7 (-5.6)	9.2 (-2.3)	9.8 (-3.6)
	$F(Y_{0,75})$	5.6 (0.0)	10.8 (10.0)	6.1 (0.3)	5.9 (0.2)	5.5 (-0.1)	6.3 (2.2)	7.0 (-4.2)	7.1 (-4.1)
	S_y^2	35.1 (0.0)	38.8 (-30.0)	35.4 (-2.4)	34.0 (-3.6)	42.2 (-0.8)	35.6 (-16.8)	56.4 (27.7)	41.6 (11.8)
0.5	\bar{Y}	7.7 (0.0)	7.7 (0.0)	8.3 (-1.1)	7.8 (-1.0)	7.4 (0.0)	7.1 (0.1)	8.4 (0.0)	8.2 (0.1)
	$Y_{0,25}$	9.1 (0.1)	49.9 (48.4)	10.2 (-0.6)	9.6 (-0.6)	8.0 (1.0)	19.9 (15.9)	14.1 (-9.7)	11.8 (-5.0)
	$Y_{0,5}$	8.1 (0.3)	19.9 (17.7)	9.2 (-0.3)	8.4 (-0.2)	7.2 (1.1)	8.6 (4.7)	9.2 (3.9)	10.3 (5.8)
	$Y_{0,75}$	9.2 (0.3)	20.9 (-19.9)	9.9 (-0.6)	9.4 (-0.6)	7.8 (0.2)	8.5 (-3.9)	14.6 (10.4)	14.0 (9.6)
	$F(Y_{0,25})$	19.5 (-0.2)	51.0 (-50.1)	21.8 (1.6)	20.2 (1.3)	16.2 (-1.6)	37.7 (-31.8)	21.3 (14.7)	17.9 (7.4)
	$F(Y_{0,5})$	11.5 (0.1)	50.7 (-47.8)	12.9 (1.0)	12.0 (0.7)	10.1 (-1.1)	17.1 (-9.1)	10.0 (-3.9)	11.5 (-6.4)
	$F(Y_{0,75})$	6.9 (-0.1)	17.0 (16.6)	7.5 (0.4)	7.2 (0.5)	6.0 (-0.1)	7.8 (3.7)	9.5 (-6.8)	9.7 (-6.7)
	S_y^2	40.5 (-1.7)	54.7 (-50.8)	41.5 (-5.6)	42.0 (-5.7)	49.5 (0.1)	44.0 (-26.8)	79.6 (48.3)	56.1 (21.3)

Tabla 2
Valores de $ECMR \times 100$ (y $SR \times 100$) asociados a distintos métodos de imputación en la población Factories. Las muestras fueron seleccionadas bajo muestreo estratificado aleatorio con afijación uniforme y tamaño $n = 100$. ** denota cantidades superiores a 100. Los 2 estratos utilizados se formaron mediante el criterio de equipartición.

p	Parámetro	Sin usar x			Usando x				
		SI	Media	RHD	NNI	Razón	Reg	Razón.PE	Reg.PE
0.1	\bar{Y}	3.5 (-0.1)	3.5 (-0.1)	3.7 (-0.1)	3.3 (-0.1)	4.0 (0.0)	3.3 (-0.1)	4.2 (0.0)	4.0 (-0.2)
	$Y_{0,25}$	4.1 (-0.7)	5.3 (2.0)	4.3 (-0.6)	3.7 (-0.6)	5.1 (-3.6)	4.1 (1.2)	4.7 (-2.9)	4.2 (-1.0)
	$Y_{0,5}$	5.5 (-1.1)	4.0 (1.4)	5.8 (-1.1)	5.2 (-0.9)	6.4 (-2.7)	6.1 (-2.5)	6.2 (-2.5)	5.5 (-1.3)
	$Y_{0,75}$	4.0 (-0.5)	6.3 (-3.4)	4.4 (-0.6)	3.6 (-0.4)	3.8 (0.1)	4.1 (-1.0)	3.7 (0.1)	3.8 (-0.4)
	$F(Y_{0,25})$	17.7 (-0.2)	19.0 (-10.2)	18.5 (-0.2)	16.4 (-0.2)	21.6 (13.4)	19.7 (-8.3)	20.0 (10.4)	17.5 (1.6)
	$F(Y_{0,5})$	10.0 (0.3)	17.2 (-2.8)	10.5 (0.1)	9.5 (0.1)	9.7 (3.0)	9.9 (2.9)	9.7 (2.7)	9.7 (0.7)
	$F(Y_{0,75})$	5.6 (0.1)	6.1 (3.5)	5.9 (0.2)	5.2 (0.0)	5.1 (-0.6)	5.2 (1.0)	5.2 (-0.6)	5.3 (0.2)
	S_y^2	10.8 (-1.4)	14.9 (-11.3)	11.3 (-1.8)	10.1 (-1.3)	74.0 (61.3)	10.8 (-2.9)	85.5 (72.2)	15.8 (7.4)
0.3	\bar{Y}	4.1 (0.2)	4.1 (0.2)	4.3 (0.3)	3.2 (0.1)	5.5 (0.2)	3.3 (0.2)	5.7 (0.2)	3.8 (0.2)
	$Y_{0,25}$	4.9 (-0.3)	15.9 (12.9)	5.7 (-0.1)	3.9 (-0.6)	15.2 (-13.0)	5.5 (3.6)	10.4 (-8.9)	5.0 (-1.4)
	$Y_{0,5}$	6.2 (-0.5)	4.5 (1.9)	6.8 (-0.5)	4.9 (-0.6)	10.4 (-7.5)	8.3 (-5.8)	8.7 (-5.7)	5.4 (-1.1)
	$Y_{0,75}$	4.9 (-0.5)	14.2 (-12.4)	5.4 (-0.6)	3.8 (-0.3)	4.4 (1.0)	5.1 (-2.2)	4.4 (1.1)	4.2 (-0.3)
	$F(Y_{0,25})$	20.4 (-0.8)	33.7 (-30.4)	22.4 (-0.8)	17.0 (-0.6)	44.9 (41.4)	32.1 (-24.0)	36.0 (31.7)	19.1 (4.1)
	$F(Y_{0,5})$	11.6 (-0.5)	35.5 (-10.8)	12.6 (-0.9)	9.6 (-0.4)	12.3 (8.7)	12.3 (8.2)	11.3 (7.3)	9.9 (0.9)
	$F(Y_{0,75})$	6.6 (-0.2)	10.9 (9.9)	7.2 (-0.1)	5.4 (0.0)	5.3 (-1.7)	5.7 (2.5)	5.4 (-1.9)	5.6 (0.0)
	S_y^2	13.4 (-1.7)	32.6 (-31.1)	14.8 (-2.1)	11.1 (-1.6)	** (**)	13.4 (-5.7)	** (**)	31.5 (24.2)
0.5	\bar{Y}	4.9 (-0.1)	4.9 (-0.1)	5.3 (0.0)	3.4 (-0.2)	7.7 (0.6)	3.6 (0.0)	8.3 (0.7)	4.6 (0.1)
	$Y_{0,25}$	6.6 (-0.2)	35.8 (34.4)	7.9 (0.4)	4.1 (-0.9)	34.9 (-33.4)	6.5 (4.5)	25.7 (-23.1)	6.7 (-2.3)
	$Y_{0,5}$	8.0 (-1.4)	5.2 (1.5)	8.3 (-1.2)	5.4 (-1.0)	17.5 (-15.3)	11.2 (-9.6)	13.4 (-11.0)	6.1 (-1.6)
	$Y_{0,75}$	6.2 (-1.3)	23.1 (-22.7)	6.5 (-1.1)	3.9 (-0.4)	6.1 (2.8)	6.6 (-4.1)	6.1 (2.9)	4.6 (-0.5)
	$F(Y_{0,25})$	25.5 (1.1)	51.2 (-49.5)	28.4 (1.1)	18.8 (1.3)	71.8 (69.5)	45.3 (-34.5)	56.8 (53.7)	22.9 (8.1)
	$F(Y_{0,5})$	13.4 (0.2)	54.6 (-11.1)	15.1 (-0.5)	10.0 (0.0)	17.1 (14.8)	16.8 (14.0)	15.0 (12.3)	10.7 (1.6)
	$F(Y_{0,75})$	8.0 (0.2)	17.3 (16.8)	8.5 (0.2)	5.6 (0.0)	5.8 (-3.0)	6.7 (4.4)	6.0 (-3.3)	5.6 (0.4)
	S_y^2	16.1 (-2.1)	51.8 (-51.1)	17.4 (-2.5)	11.7 (-1.2)	** (**)	16.2 (-7.5)	** (**)	48.3 (41.3)

métodos de imputación para la población Factories. Puede observarse que el método de imputación de la media nuevamente obtiene estimaciones con sesgos elevados en la estimación de los distintos parámetros. En esta población, el método de razón también produce serios sesgos debido a que el modelo de superpoblación (5) no se ajusta bien a los datos en estudio. Cuando el coeficiente de correlación lineal entre y y x sea elevado y la variabilidad de la variable auxiliar sea pequeña, el método NNI tendrá un mejor comportamiento que otras técnicas de imputación. En efecto, a medida que disminuye la variabilidad de x , los valores x_i (el valor de x asociado al dato faltante y_i , con $i \in s_m$) y x_j (el valor de x en la muestra s_r que se encuentra más próximo a x_i) estarán, en general, más próximos. Si a esto unimos que la correlación lineal entre y y x es elevada, resulta evidente que el valor utilizado como donante (el valor de y en la muestra s_r asociado a x_j , es decir, y_j) reemplazará de forma más eficiente al dato faltante y_i . Este hecho explicaría que el método NNI sea más eficiente que otras técnicas de imputación en la población Factories, cuyas variables tienen una correlación lineal muy elevada. Debido a un mejor ajuste, el método de regresión generalmente es más eficiente que el método de razón.

4. Algunas indicaciones sobre el mecanismo de respuesta

Las secciones anteriores están basadas en el caso de un mecanismo de respuesta uniforme. Sin embargo, esta situación suele presentarse con poca frecuencia en la práctica. Los métodos de imputación tradicionales tratan la respuesta no uniforme formando clases de imputación y entonces asumen respuesta uniforme dentro de cada una de estas clases de imputación (Berger y Rao, 2006). De este modo, todas las unidades no tendrán la misma probabilidad de respuesta. Por ejemplo, el método RHD dentro de clases de imputación es uno de los métodos de imputación usados con mayor frecuencia para la imputación de datos faltantes en encuestas por muestreo de hogares (Brick y Kalton, 1996). Una amplia revisión sobre clases de imputación puede consultarse en Kalton (1983).

En esta sección se describen los mecanismos de respuesta no uniforme más comunes en la práctica, así como las características más importantes de cada uno de ellos. En la Sección 5 evaluaremos empíricamente los distintos métodos de imputación en la presencia de clases de imputación para el tratamiento de mecanismos de respuesta no uniforme. Esto nos ayudará a extraer algunas conclusiones importantes para el caso de mecanismos de respuesta no uniforme.

En términos generales, las clases de imputación se construyen usando una variable discreta cuyos valores son observados para todas las unidades muestrales. También, estratos o uniones de estratos son usados con frecuencia para la construcción de las clases de imputación. En el caso de estratos numerosos

y pequeños, las clases de imputación se construyen mediante estratificación a posteriori (Valliant, 1993) y/o combinación de estratos pequeños. En resumen, podemos decir que la información auxiliar también se utiliza para crear las clases de imputación y mejorar, de este modo, la estimación de parámetros cuando el mecanismo de respuesta no es uniforme.

En la práctica, son dos los mecanismos de respuesta no uniforme que pueden presentarse. Por un lado, el mecanismo de respuesta UCRM (acrónimo de *unconfounded response mechanism*) depende exclusivamente de las variables auxiliares, mientras que en el mecanismo de respuesta CRM (acrónimo de *confounded response mechanism*) la probabilidad de respuesta también depende de la variable objeto de estudio. Es importante conocer los riesgos derivados de un mecanismo de respuesta no uniforme, puesto que sesgos considerables pueden obtenerse con un mecanismo CRM. Dichos sesgos pueden reducirse si se conocen las características del mecanismo de respuesta. Desafortunadamente, el mecanismo de respuesta es desconocido en la mayoría de las aplicaciones reales. Otra opción para reducir el sesgo es el uso de una corrección en los datos imputados. En este sentido, Rancourt *et al.* (1994) discutieron varios estimadores de tipo razón (véase también Hu *et al.*, 2001) diseñados para corregir el sesgo en el caso de un mecanismo de respuesta CRM.

5. Comparación en presencia de mecanismos de respuesta no uniforme

En esta sección se comparan los distintos métodos de imputación mediante estudios de simulación Monte Carlo usando las poblaciones ECPF y Factories y considerando un número diferente de clases de imputación con el fin de poder analizar la ganancia que supone un incremento del número de clases en la estimación de los distintos parámetros. Las clases de imputación se obtuvieron mediante el criterio de equipartición, también utilizado para la formación de los estratos en la Sección 3.

Siguiendo los estudios de Rancourt *et al.* (1994), se consideraron los siguientes mecanismos de respuesta no uniforme:

- (M1) La probabilidad de no respuesta es una función decreciente de x_i dada por $\exp(-\gamma x_i)$; es decir, (M1) es un mecanismo de respuesta UCRM.
- (M2) La probabilidad de no respuesta es una función creciente de x_i dada por $1 - \exp(-\gamma x_i)$; es decir, (M2) es también un mecanismo de respuesta UCRM.
- (M3) La probabilidad de no respuesta es una función decreciente de y_i dada por $\exp(-\gamma y_i)$; es decir, (M3) es un mecanismo de respuesta CRM.
- (M4) La probabilidad de no respuesta es una función creciente de y_i dada por $1 - \exp(-\gamma y_i)$; es decir, (M4) es también un mecanismo de respuesta CRM.

La constante γ se determina de modo que la proporción media de no respuesta sea igual a los valores $p = 0,1$, $p = 0,3$ y $p = 0,5$. Los distintos métodos de imputación se compararon en términos de SR y ECMR, obteniéndose los resultados mostrados en las Tablas 3 a 6.

Asumiendo el mecanismo de respuesta (M1) y dos clases de imputación, el comportamiento de los distintos métodos de imputación puede consultarse en la Tabla 3, en la cual puede observarse que los estimadores obtenidos a partir de los datos disponibles (sin usar imputación, es decir, método SI) proporcionan sesgos muy elevados, especialmente a medida que aumenta la proporción de datos faltantes. Este hecho se debe a que la falta de respuesta es no uniforme, puesto que en el caso de respuesta uniforme se pudo comprobar que este método de estimación obtenía sesgos dentro de un rango razonable. A pesar de la introducción de clases, podemos observar cómo el método de imputación de la media también genera estimaciones muy sesgadas, especialmente cuando el parámetro de interés no es la media poblacional. En lo que respecta al sesgo del resto de estimadores, podemos destacar una importante subestimación de los métodos de razón y regresión en la estimación de la varianza, así como importantes sesgos del método de razón cuando la proporción de datos faltantes es elevada. Por último, puede observarse que la incorporación de perturbaciones en los métodos de razón y regresión solventa el problema de la subestimación en la varianza, aunque la variabilidad añadida es tan elevada que produce el efecto contrario, es decir, las estimaciones ahora sobreestiman de manera importante la varianza poblacional de los datos en estudio. Este hecho, tal como se confirma en las líneas siguientes, se debe a los *outlier* o datos anómalos presentes en la variable de interés. Desde el punto de vista de la eficiencia, podemos comprobar que ambos métodos de razón y el método de regresión son generalmente los que obtienen las estimaciones más eficientes en la mayoría de los parámetros.

En la Tabla 4 se introducen dos clases más en comparación con la Tabla 3; es decir, en la Tabla 4 se aplican los distintos métodos de imputación dentro cada una de las cuatro clases creadas en cada muestra seleccionada. Aunque la introducción de más clases de imputación obtiene estimaciones tan eficientes como las obtenidas en el caso de dos clases de imputación, puede observarse que, en general, las estimaciones en el caso de cuatro clases están menos sesgadas. Por ejemplo, el método de razón que incorpora perturbaciones aleatorias obtiene en algunos parámetros (en especial la varianza) valores de SR superiores al 5% en el caso de dos clases de imputación, mientras que en el caso de cuatro clases dichos sesgos son siempre inferiores al 5%. En este sentido, también podemos destacar que el hecho de usar un número mayor de clases hacen que los outlier estén controlados y colocados en una única clase, lo que implica que los métodos basados en perturbaciones aleatorias sean más precisos y trabajen mejor. A partir de estas indicaciones, se deduce que para la población ECPF el método de razón con perturbaciones aleatorias y cuatro clases de imputación

Tabla 3
Valores de $ECMR \times 100$ (y $SR \times 100$) asociados a distintos métodos de imputación en la población ECPF. Se han utilizado dos clases de imputación para tratar el mecanismo de respuesta (M1). Las muestras fueron seleccionadas bajo muestreo aleatorio simple y tamaño $n = 150$.

p	Parámetro	Sin usar x			Usando x				
		SI	Media	RHD	NNI	Razón	Reg	Razón.PE	Reg.PE
0.1	\bar{Y}	7.0 (4.2)	5.7 (1.7)	5.8 (1.7)	5.4 (0.2)	5.3 (-0.3)	5.3 (0.2)	5.5 (-0.3)	5.5 (0.2)
	$Y_{0,25}$	9.0 (5.6)	13.2 (11.2)	8.2 (3.5)	7.4 (0.2)	6.9 (-0.3)	7.3 (1.7)	7.2 (0.4)	7.5 (1.4)
	$Y_{0,5}$	7.3 (4.0)	6.5 (-3.0)	6.1 (1.1)	5.9 (-0.6)	5.8 (-1.1)	5.8 (-1.4)	5.9 (0.3)	5.9 (0.7)
	$Y_{0,75}$	7.3 (3.8)	6.1 (-0.3)	6.2 (1.0)	6.2 (0.1)	6.0 (-1.1)	6.0 (-0.9)	6.0 (0.4)	6.1 (0.5)
	$F(Y_{0,25})$	18.5 (-11.9)	24.3 (-20.7)	17.1 (-7.1)	15.9 (-0.4)	14.6 (0.6)	16.7 (-4.0)	14.3 (-0.6)	15.0 (-2.6)
	$F(Y_{0,5})$	10.1 (-5.6)	9.7 (3.2)	9.0 (-2.0)	8.4 (0.3)	8.4 (1.2)	8.5 (1.6)	8.3 (-0.9)	8.6 (-1.3)
	$F(Y_{0,75})$	5.5 (-2.7)	4.5 (0.2)	4.8 (-0.7)	4.7 (-0.1)	4.4 (0.8)	4.5 (0.6)	4.6 (-0.4)	4.7 (-0.4)
	S_y^2	30.5 (5.3)	27.7 (-3.5)	28.7 (1.0)	28.2 (1.6)	27.7 (0.5)	27.8 (-0.9)	29.0 (5.0)	28.7 (3.5)
0.3	\bar{Y}	10.9 (8.5)	6.7 (2.8)	7.1 (2.8)	6.3 (-0.1)	5.9 (-1.1)	6.0 (0.1)	6.3 (-1.1)	6.2 (0.0)
	$Y_{0,25}$	13.7 (11.0)	30.3 (29.2)	9.9 (4.3)	9.4 (-0.3)	8.5 (3.0)	11.1 (6.7)	8.2 (-1.6)	8.7 (0.7)
	$Y_{0,5}$	11.3 (8.9)	9.4 (-7.0)	7.2 (2.5)	6.8 (-0.2)	6.2 (1.3)	6.3 (0.4)	6.8 (1.9)	7.1 (2.6)
	$Y_{0,75}$	10.4 (7.2)	9.0 (2.9)	7.9 (1.9)	7.5 (-0.3)	7.7 (-4.5)	7.4 (-2.2)	7.4 (1.9)	7.6 (2.1)
	$F(Y_{0,25})$	25.2 (-20.0)	45.3 (-44.0)	20.7 (-9.3)	19.4 (0.3)	17.0 (-6.1)	25.3 (-15.7)	15.0 (2.8)	15.6 (-1.4)
	$F(Y_{0,5})$	15.2 (-12.1)	17.3 (6.8)	10.8 (-3.8)	9.8 (0.0)	9.9 (-2.5)	10.6 (-1.1)	9.2 (-2.8)	9.9 (-4.0)
	$F(Y_{0,75})$	8.3 (-5.9)	5.4 (-1.5)	5.8 (-1.3)	5.5 (0.3)	5.9 (3.5)	6.0 (1.8)	5.6 (-1.4)	5.8 (-1.5)
	S_y^2	36.8 (12.3)	30.1 (-15.0)	32.6 (2.5)	30.4 (-0.1)	27.7 (-9.4)	28.6 (-11.5)	31.6 (8.2)	31.1 (5.8)
0.5	\bar{Y}	13.9 (11.3)	7.8 (3.8)	8.6 (4.1)	7.1 (0.0)	6.4 (-1.6)	6.5 (0.3)	7.1 (-1.6)	7.4 (0.1)
	$Y_{0,25}$	15.5 (11.7)	35.1 (33.2)	11.9 (4.9)	11.3 (0.3)	11.8 (7.2)	15.3 (11.4)	10.8 (-4.1)	10.7 (-1.1)
	$Y_{0,5}$	14.8 (11.1)	11.9 (-9.1)	9.2 (3.4)	8.4 (0.1)	8.0 (4.0)	8.5 (2.6)	8.2 (2.4)	9.0 (3.6)
	$Y_{0,75}$	15.0 (11.8)	16.0 (12.0)	9.8 (3.5)	9.0 (0.2)	8.9 (-6.2)	9.0 (-0.8)	9.0 (3.5)	9.7 (4.0)
	$F(Y_{0,25})$	29.1 (-22.5)	62.0 (-61.3)	24.4 (-9.8)	23.9 (-0.2)	21.5 (-12.7)	34.7 (-25.8)	18.1 (6.9)	18.1 (2.3)
	$F(Y_{0,5})$	18.7 (-15.2)	25.8 (6.3)	12.9 (-5.1)	11.9 (-0.4)	13.7 (-7.3)	14.3 (-4.6)	10.5 (-3.4)	11.8 (-4.8)
	$F(Y_{0,75})$	11.0 (-8.5)	8.6 (-5.7)	7.5 (-2.9)	6.6 (-0.1)	7.4 (5.1)	7.9 (1.0)	6.8 (-2.6)	7.3 (-3.0)
	S_y^2	46.8 (18.6)	37.9 (-28.0)	40.4 (5.6)	35.9 (-1.9)	32.5 (-21.1)	35.0 (-24.2)	38.7 (12.5)	38.4 (9.7)

Tabla 4
Valores de $ECMR \times 100$ (y $SR \times 100$) asociados a distintos métodos de imputación en la población ECPF. Se han utilizado cuatro clases de imputación para tratar el mecanismo de respuesta (M1). Las muestras fueron seleccionadas bajo muestreo aleatorio simple y tamaño $n = 150$.

p	Parámetro	Sin usar x			Usando x				
		SI	Media	RHD	NNI	Razón	Reg	Razón.PE	Reg.PE
0.1	\bar{Y}	6.9 (3.8)	5.4 (0.5)	5.6 (0.5)	5.5 (-0.1)	5.4 (-0.3)	5.4 (-0.1)	5.5 (-0.3)	5.5 (0.0)
	$Y_{0,25}$	8.9 (5.6)	9.7 (6.5)	7.3 (1.6)	7.3 (0.0)	7.0 (1.2)	7.2 (1.7)	7.1 (0.9)	7.2 (1.4)
	$Y_{0,5}$	7.4 (4.4)	5.7 (-0.5)	5.8 (0.3)	5.7 (-0.2)	5.4 (-0.7)	5.5 (-0.8)	5.6 (0.6)	5.8 (0.7)
	$Y_{0,75}$	7.0 (3.5)	6.1 (-1.1)	6.1 (0.1)	6.2 (-0.1)	6.1 (-1.1)	6.1 (-1.0)	6.0 (0.1)	6.2 (0.2)
	$F(Y_{0,25})$	18.0 (-11.6)	24.8 (-15.0)	15.5 (-3.2)	15.0 (0.2)	14.7 (-2.1)	16.3 (-3.6)	14.0 (-1.3)	14.5 (-2.4)
	$F(Y_{0,5})$	10.3 (-6.1)	8.2 (0.5)	8.4 (-0.8)	8.3 (-0.1)	8.3 (0.8)	8.3 (0.8)	8.3 (-1.2)	8.4 (-1.4)
	$F(Y_{0,75})$	5.6 (-2.7)	4.6 (0.7)	4.8 (-0.2)	4.8 (0.0)	4.6 (0.7)	4.6 (0.7)	4.7 (-0.2)	4.8 (-0.2)
	S_y^2	29.4 (2.1)	27.4 (-5.1)	27.5 (-2.0)	27.6 (-1.3)	27.2 (-3.1)	27.3 (-3.6)	27.7 (0.4)	27.7 (-0.4)
0.3	\bar{Y}	11.1 (8.8)	6.0 (1.1)	6.2 (1.2)	6.2 (0.2)	5.8 (-0.3)	5.8 (0.1)	6.2 (-0.3)	6.2 (0.2)
	$Y_{0,25}$	13.8 (11.0)	14.8 (9.3)	9.0 (1.7)	9.1 (0.0)	10.7 (6.1)	10.8 (6.2)	8.5 (0.4)	8.6 (1.3)
	$Y_{0,5}$	11.7 (9.1)	7.8 (2.3)	7.1 (0.4)	7.2 (-0.2)	6.4 (0.9)	6.8 (0.7)	7.1 (1.8)	7.2 (1.8)
	$Y_{0,75}$	11.0 (7.8)	7.6 (-3.2)	7.3 (0.8)	7.4 (0.5)	7.3 (-3.4)	7.4 (-3.0)	7.1 (1.3)	7.3 (1.6)
	$F(Y_{0,25})$	24.9 (-19.8)	44.1 (-29.1)	18.6 (-3.7)	18.9 (-0.2)	20.4 (-11.6)	25.0 (-14.4)	15.4 (-0.7)	16.2 (-2.1)
	$F(Y_{0,5})$	15.5 (-12.2)	12.1 (-2.9)	10.1 (-1.1)	10.1 (0.1)	11.0 (-2.2)	11.2 (-1.7)	9.9 (-2.9)	10.2 (-3.0)
	$F(Y_{0,75})$	8.6 (-6.3)	6.1 (2.5)	5.6 (-0.5)	5.6 (-0.2)	5.7 (2.6)	5.7 (2.3)	5.5 (-0.9)	5.6 (-1.1)
	S_y^2	36.4 (13.4)	28.2 (-11.9)	30.0 (1.8)	30.1 (1.3)	27.2 (-9.6)	27.5 (-9.8)	29.6 (4.3)	29.7 (4.0)
0.5	\bar{Y}	13.9 (11.0)	7.1 (1.4)	7.7 (1.4)	7.4 (0.1)	6.5 (-0.6)	6.8 (0.1)	7.1 (-0.6)	7.4 (0.2)
	$Y_{0,25}$	15.2 (11.5)	16.8 (9.0)	10.5 (1.8)	11.1 (0.5)	15.9 (11.8)	15.0 (10.3)	9.9 (-0.1)	10.0 (0.4)
	$Y_{0,5}$	14.6 (11.3)	11.6 (6.2)	8.6 (0.9)	8.8 (-0.2)	8.5 (3.7)	9.0 (3.3)	8.4 (2.8)	8.8 (2.8)
	$Y_{0,75}$	15.0 (11.5)	10.6 (-4.5)	9.3 (1.3)	9.1 (0.4)	8.8 (-4.2)	9.3 (-3.0)	8.8 (2.3)	9.5 (2.9)
	$F(Y_{0,25})$	28.4 (-21.8)	57.3 (-37.9)	22.8 (-3.3)	22.9 (-0.3)	27.7 (-20.5)	34.8 (-23.2)	17.7 (0.8)	18.3 (0.2)
	$F(Y_{0,5})$	18.4 (-14.6)	17.8 (-7.8)	11.9 (-1.3)	12.2 (-0.2)	15.2 (-7.1)	15.0 (-5.7)	11.5 (-4.0)	11.7 (-4.0)
	$F(Y_{0,75})$	11.0 (-8.4)	8.4 (3.0)	6.8 (-1.0)	6.8 (-0.3)	7.8 (3.6)	7.5 (2.4)	6.8 (-1.7)	7.0 (-2.1)
	S_y^2	48.4 (19.2)	36.2 (-24.0)	39.3 (2.8)	37.7 (-1.1)	34.5 (-22.8)	34.6 (-21.8)	37.5 (4.9)	38.5 (6.2)

sería la técnica de imputación apropiada que estima los distintos parámetros de manera más eficiente y con sesgos, en términos absolutos, más pequeños.

Notamos que la ganancia más importante, en términos de SR y ECMR, se produce cuando pasamos de muestras sin clases de imputación a dos clases de imputación. A medida que incorporamos clases de imputación, la eficiencia de las distintas estimaciones permanece aproximadamente constante, mientras que los sesgos se van reduciendo paulatinamente. No obstante, la reducción en el sesgo se mantiene hasta el uso de cinco clases de imputación, obteniéndose incluso resultados negativos cuando el número de clases de imputación es demasiado elevado. Esto se debe a la existencia de clases con pocos valores muestrales que producen, para esa clase, estimaciones poco fiables.

Las Tablas 5 y 6 presentan los valores de SR y ECMR en la población Factories cuando el mecanismo de respuesta es (M1) y se usan, respectivamente, dos y cuatro clases de imputación. Las conclusiones que se derivan de estas tablas son similares a las que ya hemos comentado en las Tablas 3 y 4. No obstante, destacamos que si en el supuesto de respuesta uniforme el método NNI era el más eficiente en la mayoría de los casos debido a la alta correlación entre las variables, cuando el mecanismo de respuesta es no uniforme y se utilizan cuatro clases de imputación, los métodos de regresión con perturbaciones y NNI son los más eficientes de entre los distintos métodos comparados en el estudio; es decir, el método de regresión se vuelve tan eficiente como el método NNI con la introducción de clases de imputación.

Simulaciones basadas en el mecanismo de respuesta (M2) fueron también analizadas en los estudios de simulación, obteniéndose conclusiones similares a las expuestas para el mecanismo (M1), y de aquí que esta información esté omitida. Asumiendo los mecanismos de respuesta (M3) y (M4), los distintos métodos de imputación mostraron sesgos elevados debido a que el mecanismo de respuesta es de tipo CRM. En este caso, los distintos métodos de imputación necesitarían un ajuste para corregir el sesgo causado por este tipo de mecanismo no uniforme. En este sentido, en Rancourt *et al.* (1994) y Hu *et al.* (2001) se discuten algunos métodos para el problema de imputación en el caso de mecanismo de respuesta CRM. El estudio de estos métodos de imputación ajustados no es el objetivo de este trabajo y de aquí que esta información esté también omitida.

En resumen, se ha constatado que el mecanismo de respuesta es un factor muy importante en el problema de la imputación para el tratamiento de datos faltantes. Si bien la no utilización de imputación puede producir estimaciones eficientes y poco sesgadas en un mecanismo de respuesta uniforme, esta metodología resulta poco apropiada en el caso de mecanismos de respuesta no uniforme. Por otro lado, hemos comprobado que el uso de clases de imputación proporcionan estimaciones más eficientes y menos sesgadas, espe-

Tabla 5
Valores de $ECMR \times 100$ (y $SR \times 100$) asociados a distintos métodos de imputación en la población Factories. Se han utilizado dos clases de imputación para tratar el mecanismo de respuesta (M1). Las muestras fueron seleccionadas bajo muestreo aleatorio simple y tamaño $n = 100$.

p	Parámetro	Sin usar x			Usando x				
		SI	Media	RHD	NNI	Razón	Reg	Razón.PE	Reg.PE
0.1	\bar{Y}	5.1 (3.9)	3.7 (1.9)	3.8 (1.9)	3.1 (0.0)	3.4 (-1.5)	3.1 (0.5)	3.4 (-1.5)	3.3 (0.5)
	$Y_{0,25}$	6.8 (4.2)	9.3 (7.0)	5.4 (2.3)	3.5 (-1.2)	4.5 (-2.9)	3.7 (-1.2)	4.1 (-2.3)	4.0 (0.2)
	$Y_{0,5}$	5.5 (3.3)	5.3 (-1.8)	4.6 (1.2)	4.9 (-1.2)	5.1 (-1.4)	5.2 (-1.6)	4.9 (-1.1)	4.6 (-0.5)
	$Y_{0,75}$	3.3 (1.6)	3.6 (-0.7)	3.5 (-0.6)	3.6 (-0.7)	3.6 (-0.6)	3.6 (-0.7)	3.5 (-0.6)	3.5 (-0.6)
	$F(Y_{0,25})$	24.8 (-19.0)	30.7 (-27.1)	23.0 (-15.0)	15.9 (-0.4)	16.3 (7.1)	17.1 (0.8)	15.5 (4.6)	17.3 (-5.7)
	$F(Y_{0,5})$	14.0 (-10.1)	8.8 (0.9)	11.6 (-5.7)	8.9 (-0.1)	8.8 (0.4)	8.8 (0.8)	8.9 (-0.2)	9.3 (-1.3)
	$F(Y_{0,75})$	6.6 (-3.5)	5.0 (0.1)	5.0 (0.1)	5.0 (0.1)	5.0 (0.1)	5.0 (0.1)	5.0 (0.0)	5.0 (0.1)
	S_y^2	11.2 (-4.4)	14.8 (-11.1)	12.7 (-7.2)	10.3 (0.1)	14.6 (10.6)	11.2 (-5.0)	18.8 (14.6)	11.1 (-1.1)
0.3	\bar{Y}	11.4 (10.7)	5.1 (3.9)	5.2 (3.9)	3.0 (0.2)	4.9 (-3.7)	3.2 (0.8)	5.1 (-3.7)	3.4 (0.8)
	$Y_{0,25}$	17.8 (15.0)	21.1 (20.6)	10.2 (6.5)	3.6 (-1.2)	11.7 (-10.0)	3.8 (-0.7)	9.4 (-8.0)	5.0 (0.2)
	$Y_{0,5}$	15.1 (13.3)	10.4 (-8.6)	6.2 (4.1)	4.8 (-0.8)	6.7 (-3.7)	7.2 (-4.7)	5.6 (-1.8)	4.7 (0.1)
	$Y_{0,75}$	6.0 (5.5)	3.6 (-0.3)	3.1 (-0.2)	3.3 (-0.4)	3.2 (0.3)	3.1 (-0.3)	3.1 (0.6)	2.9 (0.2)
	$F(Y_{0,25})$	40.6 (-37.4)	57.3 (-56.2)	31.9 (-24.6)	16.5 (-0.3)	30.9 (27.4)	21.3 (-0.9)	26.4 (22.1)	18.3 (-4.3)
	$F(Y_{0,5})$	29.8 (-28.0)	11.7 (9.0)	17.7 (-13.2)	8.9 (-0.8)	8.2 (3.2)	9.5 (5.4)	7.9 (1.0)	8.9 (-2.1)
	$F(Y_{0,75})$	15.4 (-13.7)	5.2 (-0.5)	5.1 (-0.4)	5.0 (-0.1)	5.4 (-1.5)	5.1 (-0.4)	5.7 (-1.9)	5.3 (-1.1)
	S_y^2	12.6 (-5.1)	24.9 (-22.9)	16.8 (-11.1)	10.9 (0.3)	34.6 (32.7)	13.4 (-8.8)	47.1 (44.5)	13.6 (3.3)
0.5	\bar{Y}	15.9 (15.3)	5.6 (4.1)	6.0 (4.2)	3.2 (0.1)	6.5 (-5.0)	3.3 (0.5)	6.6 (-4.9)	3.6 (0.5)
	$Y_{0,25}$	24.9 (21.6)	21.3 (20.4)	12.3 (6.8)	3.9 (-1.3)	21.1 (-19.0)	4.7 (-0.7)	17.8 (-15.7)	6.0 (-0.9)
	$Y_{0,5}$	23.8 (22.0)	12.9 (-12.1)	7.4 (3.8)	5.6 (-1.5)	10.8 (-8.6)	10.1 (-8.3)	8.4 (-5.6)	5.3 (-1.0)
	$Y_{0,75}$	8.9 (8.4)	5.3 (0.8)	3.9 (0.3)	3.6 (-0.6)	4.2 (0.6)	3.2 (-0.1)	3.9 (1.3)	3.3 (0.7)
	$F(Y_{0,25})$	46.9 (-43.1)	72.3 (-71.2)	34.9 (-22.1)	19.7 (0.2)	45.6 (41.9)	27.7 (-1.5)	39.4 (35.8)	19.9 (-0.3)
	$F(Y_{0,5})$	38.3 (-36.5)	22.0 (20.6)	20.3 (-13.3)	9.8 (-0.1)	12.1 (9.3)	13.6 (10.8)	9.8 (5.8)	8.9 (0.1)
	$F(Y_{0,75})$	24.5 (-23.2)	5.8 (-2.4)	5.5 (-1.4)	5.0 (-0.2)	5.6 (-2.1)	5.3 (-1.2)	6.2 (-3.1)	5.7 (-2.2)
	S_y^2	15.0 (-0.6)	29.0 (-26.3)	18.3 (-7.7)	12.9 (-0.3)	56.2 (54.1)	15.2 (-9.4)	76.6 (73.5)	19.5 (9.7)

Tabla 6
Valores de $ECMR \times 100$ (y $SR \times 100$) asociados a distintos métodos de imputación en la población Factories. Se han utilizado cuatro clases de imputación para tratar el mecanismo de respuesta (M1). Las muestras fueron seleccionadas bajo muestreo aleatorio simple y tamaño $n = 100$.

p	Parámetro	Sin usar x				Usando x			
		SI	Media	RHD	NNI	Razón	Reg	Razón.PE	Reg.PE
0.1	\bar{Y}	5.1 (3.8)	3.2 (0.6)	3.2 (0.6)	3.1 (0.0)	3.1 (-0.4)	3.1 (0.0)	3.1 (-0.4)	3.1 (0.0)
	$Y_{0,25}$	6.4 (3.8)	4.3 (-2.9)	3.7 (0.1)	3.4 (-1.2)	3.8 (-2.0)	3.8 (-1.8)	3.7 (-1.5)	3.6 (-1.0)
	$Y_{0,5}$	5.5 (3.3)	5.0 (-0.9)	4.8 (-1.0)	4.8 (-1.1)	4.8 (-1.3)	4.7 (-1.2)	4.8 (-1.0)	4.7 (-0.9)
	$Y_{0,75}$	3.5 (1.7)	3.9 (-0.6)	3.8 (-0.6)	3.9 (-0.6)	3.9 (-0.6)	3.9 (-0.6)	3.8 (-0.6)	3.8 (-0.6)
	$F(Y_{0,25})$	23.4 (-18.1)	20.4 (7.5)	18.1 (-6.8)	14.9 (0.1)	14.7 (3.7)	14.7 (2.8)	14.5 (1.3)	14.9 (-0.9)
	$F(Y_{0,5})$	13.9 (-10.0)	8.9 (-0.8)	8.9 (-0.4)	8.9 (-0.1)	8.8 (0.2)	8.9 (0.2)	8.8 (-0.3)	9.0 (-0.6)
	$F(Y_{0,75})$	6.9 (-3.8)	5.2 (-0.1)	5.2 (-0.1)	5.2 (-0.1)	5.2 (-0.1)	5.2 (-0.1)	5.2 (-0.2)	5.2 (-0.1)
	S_y^2	11.1 (-3.7)	11.4 (-4.8)	11.6 (-3.0)	10.3 (0.5)	10.3 (2.6)	10.4 (-0.8)	11.6 (4.3)	11.0 (1.1)
	0.3	\bar{Y}	11.3 (10.7)	3.4 (1.2)	3.5 (1.2)	3.1 (0.1)	3.3 (-1.1)	3.1 (0.2)	3.4 (-1.1)
$Y_{0,25}$		17.8 (15.0)	5.8 (-4.0)	5.0 (0.7)	3.6 (-1.1)	5.7 (-4.5)	4.8 (-2.5)	5.2 (-3.6)	4.4 (-0.9)
$Y_{0,5}$		15.0 (13.1)	6.3 (1.5)	5.3 (-0.1)	4.8 (-0.9)	5.5 (-2.7)	4.2 (-1.0)	5.4 (-1.8)	4.4 (0.0)
$Y_{0,75}$		6.0 (5.5)	3.7 (-0.4)	3.5 (-0.4)	3.5 (-0.5)	3.4 (0.0)	3.5 (-0.5)	3.3 (0.1)	3.4 (-0.3)
$F(Y_{0,25})$		40.2 (-37.1)	43.8 (18.2)	23.5 (-9.5)	16.2 (-0.5)	20.0 (14.1)	16.1 (6.2)	17.4 (9.6)	14.6 (-0.3)
$F(Y_{0,5})$		29.8 (-28.0)	10.4 (-5.0)	9.5 (-1.9)	9.1 (-0.7)	8.9 (3.2)	9.1 (0.2)	8.4 (1.1)	9.3 (-2.0)
$F(Y_{0,75})$		15.5 (-13.9)	5.1 (-0.6)	5.1 (-0.4)	5.0 (-0.2)	5.2 (-1.1)	5.0 (-0.3)	5.3 (-1.2)	5.0 (-0.4)
S_y^2		12.5 (-4.5)	13.6 (-8.2)	13.4 (-3.8)	11.2 (0.5)	13.2 (8.5)	10.8 (-1.6)	17.8 (13.3)	13.1 (3.0)
0.5		\bar{Y}	15.8 (15.1)	3.6 (1.0)	3.7 (1.0)	3.3 (0.0)	4.0 (-1.7)	3.2 (0.0)	4.1 (-1.8)
	$Y_{0,25}$	25.1 (22.0)	7.1 (-4.3)	6.2 (0.8)	3.9 (-1.3)	8.2 (-6.7)	6.0 (-3.4)	7.8 (-5.9)	5.5 (-1.4)
	$Y_{0,5}$	23.7 (21.9)	8.2 (3.2)	5.8 (-0.2)	5.6 (-1.6)	7.8 (-5.4)	4.1 (-1.1)	7.3 (-4.3)	4.7 (-0.1)
	$Y_{0,75}$	8.7 (8.2)	4.7 (-0.2)	3.9 (-0.4)	3.9 (-0.8)	4.0 (0.2)	3.8 (-0.7)	3.8 (0.3)	3.4 (-0.5)
	$F(Y_{0,25})$	47.8 (-44.1)	61.3 (21.8)	28.6 (-8.4)	18.9 (0.0)	28.7 (22.5)	18.6 (9.1)	24.6 (17.5)	16.2 (1.6)
	$F(Y_{0,5})$	38.4 (-36.6)	11.6 (-7.2)	10.1 (-1.6)	10.3 (0.1)	12.2 (7.8)	10.5 (1.0)	10.3 (5.0)	9.8 (-1.2)
	$F(Y_{0,75})$	24.2 (-22.7)	5.7 (-2.0)	5.5 (-0.5)	5.2 (0.0)	5.4 (-1.5)	5.2 (-0.5)	5.5 (-1.7)	5.2 (-0.4)
	S_y^2	15.2 (-1.3)	16.2 (-9.3)	16.1 (-2.7)	13.2 (-0.8)	19.1 (15.2)	12.5 (-2.6)	26.3 (22.2)	16.2 (4.4)

cialmente en el caso de mecanismos de respuesta no uniforme. Con el fin de reducir en la medida de lo posible los sesgos, recomendamos el uso de cuatro o más clases de imputación, si bien debemos de controlar, en cualquier caso, que no existen clases con pocas unidades, puesto que esto podría producir resultados poco satisfactorios. Los métodos NNI, razón con perturbaciones y regresión con perturbaciones se han mostrado como los más eficientes en el caso de mecanismos de respuesta no uniforme, aunque debemos realizar un análisis previo para utilizar de entre ellos el más apropiado. Por ejemplo, el método NNI resulta apropiado cuando la correlación entre las variables es muy elevada. Para la elección entre los métodos de razón y regresión tendremos que estudiar el modelo ajustado entre las variables y analizar cuál de los dos métodos se adapta mejor a dicho modelo.

APÉNDICE A. Implementación de métodos de imputación mediante R/Splus

En este apéndice se proporcionan las funciones o códigos en el entorno de los lenguajes de programación estadística R y Splus para la implementación de los métodos de imputación descritos en este trabajo. Notamos que la idea de proporcionar códigos en R/Splus fue también seguida en Wu (2005) para la implementación del reciente método de verosimilitud empírica en el contexto del muestreo en poblaciones finitas.

En primer lugar, describimos la función `SEPARA.muestras` que separa la muestra inicial s_n en las muestras s_r y s_m . Además, esta función devuelve otros objetos, tales como el valor de m , necesarios para el uso del resto de funciones descritas en este apéndice. Notamos que los lenguajes de programación R y Splus utilizan el comando “NA” para indicar que un determinado valor no está disponible. La función `SEPARA.muestras` dispone de tres argumentos, los cuales se detallan a continuación:

1. `muestray`: valores de la variable y en la muestra s_n que tiene datos faltantes.
2. `muestrax`: valores de la variable x en la muestra s_n .
3. `Pi`: probabilidades de inclusión de primer orden asociadas a las unidades de la muestra s_n .

```
SEPARA.muestras <- function(muestray, muestrax, Pi)
{
  POS.faltantes <- is.na(muestray)
  POS.disponibles <- !POS.faltantes
  datosy.r <- muestray[POS.disponibles]
  datosx.r <- muestrax[POS.disponibles]
  Pi.r <- Pi[POS.disponibles]
  datosx.m <- muestrax[POS.faltantes]
  m <- length(datosx.m)
  list(datosy.r=datosy.r, datosx.r=datosx.r, Pi.r=Pi.r,
       m=m, datosx.m=datosx.m, POS.faltantes=POS.faltantes)
}
```

A continuación se describe la función `METODO.media` que permite implementar el método de imputación de la media. La función `METODO.media` da como salida los m donantes que sustituirán los m datos faltantes en la variable y . Esta función dispone de los siguientes argumentos, los cuales se pueden obtener a partir de la función `SEPARA.muestras`:

1. `datosy.r`: valores de la variable y en la muestra s_r .
2. `Pi.r`: probabilidades de inclusión de primer orden asociadas a las unidades de la muestra s_r .
3. `m`: número de datos faltantes en la variable y .

```
METODO.media <- function(datosy.r, Pi.r, m)
{
  Pesos      <- 1/Pi.r
  N.est      <- sum(Pesos)
  Media      <- (1/N.est)*sum(Pesos*datosy.r)
  DONANTES.media <- rep(Media,m)
  DONANTES.media
}
```

El siguiente método de imputación que hemos implementado en R es el método de Cohen. Los argumentos de esta función son los mismos que los descritos para la función `METODO.media`.

```
METODO.cohen <- function(datosy.r, Pi.r, m)
{
  r      <- length(Pi.r)
  n      <- r + m
  Pesos  <- 1/Pi.r
  N.est  <- sum(Pesos)
  Media  <- (1/N.est)*sum(Pesos*datosy.r)
  Dr     <- sqrt((1/N.est)*sum(Pesos*(datosy.r-Media)^2))
  m1     <- round(m/2)
  m2     <- m - m1
  Raiz   <- sqrt(n+r+1)/sqrt(r-1)
  DONANTES.cohen <- c(rep(Media+Raiz*Dr,m1),rep(Media-Raiz*Dr,m2))
  DONANTES.cohen
}
```

La función `METODO.NNI` da como salida los m valores imputados mediante el método de imputación NNI. Los argumentos requeridos en esta función son los siguientes:

1. `datosy.r`: valores de la variable y en la muestra s_r .
2. `datosx.r`: valores de la variable x en la muestra s_r .
3. `datosx.m`: valores de la variable x en la muestra s_m .
4. `m`: número de datos faltantes en la variable y .

```
METODO.NNI <- function(datosy.r, datosx.r, datosx.m, m)
{
  DONANTES.NNI <- c()
  for (j in 1:m)
  {
    Diferencias <- abs(datosx.m[j] - datosx.r)
    Dif.min     <- min(Diferencias)
    POS.min     <- Dif.min==Diferencias
    DONANTES    <- datosy.r[POS.min]
    Num.T       <- sum(POS.min)
    if (Num.T==1) DONANTES.NNI <- c(DONANTES.NNI, DONANTES)
    else        DONANTES.NNI <- c(DONANTES.NNI, sample(DONANTES,1))
  }
}
```

```
DONANTES.NNI
}
```

Imputaciones mediante los métodos de la razón y regresión pueden realizarse, respectivamente, mediante las funciones `METODO.razon` y `METODO.regresion`. Por otra parte, con las funciones `METODO.razon.aleatorio` y `METODO.regresion.aleatorio` se obtienen los donantes mediante los métodos de razón y regresión que añaden perturbaciones aleatorias. Los argumentos de estas funciones ya han sido descritos en funciones anteriores.

```
METODO.razon <- function(datosy.r, datosx.r, Pi.r, datosx.m)
{
  Pesos      <- 1/Pi.r
  N.est      <- sum(Pesos)
  MediaY     <- (1/N.est)*sum(Pesos*datosy.r)
  MediaX     <- (1/N.est)*sum(Pesos*datosx.r)
  DONANTES.razon <- (MediaY/MediaX)*datosx.m
  DONANTES.razon
}

METODO.regresion <- function(datosy.r, datosx.r, Pi.r, datosx.m)
{
  Pesos      <- 1/Pi.r
  N.est      <- sum(Pesos)
  MediaY     <- (1/N.est)*sum(Pesos*datosy.r)
  MediaX     <- (1/N.est)*sum(Pesos*datosx.r)
  Beta.est   <- sum(Pesos*(datosx.r-MediaX)*(datosy.r-MediaY))/sum(Pesos*(datosx.r-MediaX)^2)
  DONANTES.reg <- (MediaY/MediaX)*datosx.m
  DONANTES.reg
}

METODO.razon.aleatorio <- function(datosy.r, datosx.r, Pi.r, datosx.m)
{
  Pesos      <- 1/Pi.r
  N.est      <- sum(Pesos)
  Media      <- (1/N.est)*sum(Pesos*datosy.r)
  Desviacion <- sqrt((1/N.est)*sum(Pesos*(datosy.r-Media)^2))
  DONANTES.razonA <- METODO.razon(datosy.r, datosx.r, Pi.r, datosx.m) + rnorm(m,0,Desviacion)
  DONANTES.razonA
}

METODO.regresion.aleatorio <- function(datosy.r, datosx.r, Pi.r, datosx.m)
{
  Pesos      <- 1/Pi.r
  N.est      <- sum(Pesos)
  Media      <- (1/N.est)*sum(Pesos*datosy.r)
  Desviacion <- sqrt((1/N.est)*sum(Pesos*(datosy.r-Media)^2))
  DONANTES.regA <- METODO.regresion(datosy.r, datosx.r, Pi.r, datosx.m) + rnorm(m,0,Desviacion)
  DONANTES.regA
}
```

La función `METODO.RHD` implementa el método de imputación RHD. Los argumentos de esta función coinciden con los argumentos de las funciones `METODO.media` y `METODO.cohen`.

```
METODO.RHD <- function(datosy.r, Pi.r, m)
{
  Pesos      <- 1/Pi.r
  Prob       <- Pesos/sum(Pesos)
  DONANTES.RHD <- sample(datosy.r, m,replace=T, prob=Prob)
  DONANTES.RHD
}
```

Por último, describimos la función `REALIZA.imputacion`, que devuelve todos los valores de la variable y en la muestra s_n ; es decir, esta función asigna los valores devueltos por cada una de las funciones de imputación descritas a las posiciones donde se ha producido la no respuesta.

```
REALIZA.imputacion <- function(muestray, DONANTES, POS.faltantes)
{
muestray[POS.faltantes] <- DONANTES
muestray
}
```

APÉNDICE B. Ejemplo para el uso de las funciones descritas

En este apéndice se presenta un ejemplo que describe cómo utilizar las funciones descritas en el Apéndice A, dadas una muestra con unidades faltantes y las probabilidades de inclusión de cada una de las unidades de la mencionada muestra. Por simplicidad, la muestra utilizada en este ejemplo tiene tamaño $n = 10$ y ha sido seleccionada mediante muestreo aleatorio simple de la población *Factories* descrita en la Sección 3. Los datos seleccionados fueron los siguientes:

Muestra de y : 7152, 5630, 6752, 6660, 4762, 3821, 7416, 5562, 6567, 5286.

Muestra de x : 563, 211, 425, 443, 185, 97, 705, 198, 375, 160.

Además, a partir de las $n = 10$ unidades que componen la muestra, se seleccionaron aleatoriamente $m = 3$ unidades, las cuales jugarán el papel de datos faltantes. Las unidades seleccionadas como datos faltantes fueron las que ocupan las posiciones 2, 4 y 7; es decir, los valores 5630, 6660 y 7416 serán tratados como datos faltantes. Las instrucciones que podemos seguir para imputar estos datos faltantes mediante el método NNI, por ejemplo, son las siguientes:

```
## Introducimos los datos muestrales de ambas variables, incluyendo los datos faltantes:
muestray <- c(7152, NA, 6752, NA, 4762, 3821, NA, 5562, 6567, 5286)
muestrax <- c(563, 211, 425, 443, 185, 97, 705, 198, 375, 160)

## Introducimos las probabilidades de inclusión:
Pi <- rep(1/10,10)

## Separamos la muestra s_n y obtenemos el resto de información necesaria:
SALIDA <- SEPARA.muestras(muestray, muestrax, Pi)

## Aplicamos el método NNI para obtener los donantes:
DONANTES.NNI <- METODO.NNI(SALIDA$datosy.r, SALIDA$datosx.r, SALIDA$datosx.m, SALIDA$m)

## Reemplazamos los datos faltantes por los donantes obtenidos:
REALIZA.imputacion(muestray, DONANTES.NNI, SALIDA$POS.faltantes)
```

Los valores devueltos por la última instrucción son

7152, 5562, 6752, 6752, 4762, 3821, 7152, 5562, 6567, 5286.

Estas cantidades corresponden a los valores muestrales de y después de sustituir los datos faltantes por sus correspondientes imputaciones.

Todas las funciones descritas en los Apéndices A y B están diseñadas para el caso de un diseño muestral general sin clases de imputación. En la presencia de clases de imputación, el uso de las funciones es bastante simple. En este caso tendremos que utilizar las mencionadas funciones dentro de cada una de las clases de imputación. No obstante, se puede solicitar información a los autores sobre las funciones descritas, u otras que sean requeridas, mediante correo electrónico.

Referencias

- [1] Arcos, A., Gámiz, M.L., González, A., Martínez, M.D. y Rueda, M.M. (2004). *Programación en R. Aplicaciones al muestreo*. Ed. Los autores. ISBN: 84-609-3077-7. Depósito legal: GR-1880-2004.
- [2] Arcos, A., Gámiz, M.L., González, A., Martínez, M.D., Muñoz, J.F., Román, Y. y Rueda, M.M. (2005). *Estadística Computacional con SPSS y R*. Ed. Los autores. ISBN: 84-689-5347-4. Depósito legal: GR-2110-2005.
- [3] Bello, A.L. (1993). Choosing among imputation techniques for incomplete multivariate data: a simulation study. *Communication in Statistics*, **22** 823–877.
- [4] Berger, Y.G. y Rao, J.N.K. (2006). Adjusted jackknife for imputation under unequal probability sampling without replacement. *Journal of the Royal Statistical Society, Series B*, **68** 531–547.
- [5] Berger, Y.G. y Skinner, C.J. (2003). Variance estimation for a low income proportion. *Journal of the Royal Statistical Society, Series B*, **52** 457–468.
- [6] Brick, J.M. y Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, **5** 215–238.
- [7] Chambers, R.L. y Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, **73** 597–604.
- [8] Chen, J. y Shao, J. (2000). Nearest neighbor imputation for survey data. *Journal of Official Statistics*, **16** 113–131.
- [9] Cohen, M.P. (1996). A new approach to imputation. *American Statistical Association Proceeding of the Section on Survey Research Methods* 293–298.
- [10] Everitt, B.S. (1994). *A handbook of Statistical Analysis using S-Plus*. Chapman and Hall, New York.
- [11] Fay, R.E. (1991). A design-based perspective on missing data variance. In Proc. Seventh Annual Res. Conf., Washington, D.C.: U.S. Bureau of the Census. 429–440.
- [12] Hu, M., Salvucci, S. y Lee, R. (2001). *A Study of Imputation Algorithms*. Working Paper No. 200117. Washington DC: U.S. Department of Education, National Center for Education Statistics, 2001. 27 Stata Statistical Software.

- [13] Healy, M.J.R. y Westmacott, M. (1956). Missing values in experiments analysed on automatic computers. *Applied Statistics*, **5** 203–206.
- [14] Ihaka, R. y Gentleman, R. (1996). R: a Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, **5** 299–314.
- [15] Kalton, G. (1983). *Compensating for missing data*. Ann Arbor: Institute for Social Research, University of Michigan.
- [16] Kalton, G. y Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology* **12** 1–16.
- [17] Krause, A. y Olson, M. (2005). *The basic of S-Plus. Fourth Edition*. Springer.
- [18] Kuk, A.Y.C. y Mak, T.K. (1989). Median estimation in the presence of auxiliary information. *Journal of the Royal Statistical Society, Series B*, **51** 261–269.
- [19] Little, R.J.A. y Rubin, D.B. (2002). *Statistical analysis with missing data. 2nd edition*. New York: John Wiley & Sons, Inc.
- [20] Merino, M. y Vadillo, F. (2007). Matemática financiera con MATLAB®. *Métodos Cuantitativos para la Economía y la Empresa*, **4** 35–55.
- [21] Murthy, M.N. (1967). *Sampling theory and method*. Calcutta: Statistical Publishing Society.
- [22] Rancourt, E., Lee, H. y Särndal, C.E. (1994). Bias correction for survey estimates from data with ratio imputed values for confounded nonresponse. *Survey Methodology*, **20** 137–147.
- [23] Rao, J.N.K. (1996). On variance estimation with imputed survey data (with discussion). *Journal of the American Statistical Association*, **91** 499–520.
- [24] Rao, J.N.K., Kovar, J.G. y Mantel, H.J. (1990). On estimating distribution function and quantiles from survey data using auxiliary information. *Biometrika*, **77** 365–375.
- [25] Rao, J.N.K. y Shao, J. (1992). Jackknife Variance Estimation With Survey Data Under Hot-Deck Imputation. *Biometrika*, **79** 811–822.
- [26] Rubin, D.B. (1978). Multiple imputations in sample surveys. A phenomenological bayesian approach to nonresponse. Proceedings of the Survey Research Methods Section, American Statistical Association. 20–34.
- [27] Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, **91** 473–489.
- [28] Särndal, C.E., Swensson, B. y Wretman, J.H. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.
- [29] Sedransk, J. (1985). The objective and practice of imputation. In *Proc. First Annual Res. Conf.*, Washington, D.C.: Bureau of the Census. 445–452.
- [30] Silva P.L.D. y Skinner C.J. (1995). Estimating distribution function with auxiliary information using poststratification. *Journal of Official Statistics*, **11** 277–294.
- [31] Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, **88** 89–96.
- [32] Wu, C. (2005). Algorithms and R codes for the pseudo empirical likelihood methods in survey sampling. *Survey Methodology*, **31** 239–243.

Localización empresarial en Aragón: una aplicación empírica de la ayuda a la decisión multicriterio tipo ELECTRE I y III. Robustez de los resultados obtenidos

GARCÍA CEBRIÁN, LUCÍA ISABEL

Departamento de Economía y Dirección de Empresas

Universidad de Zaragoza

Correo electrónico: lgarcia@unizar.es

MUÑOZ PORCAR, ANTONIO

Departamento de Economía y Dirección de Empresas

Universidad de Zaragoza

Correo electrónico: amunoz@unizar.es

RESUMEN

La decisión de la localización es una de las opciones estratégicas más importantes que tienen que tomar las empresas, con influencia directa sobre su competitividad. Por este motivo, se han desarrollado numerosas metodologías que ayudan en la toma de estas decisiones. En este trabajo, se realiza una aplicación empírica de una metodología de ayuda a la decisión multicriterio; específicamente, se ha optado por el método ELECTRE en sus versiones I y III, con el fin de seleccionar el mejor emplazamiento para una empresa. Para esta aplicación, se utilizan como alternativas de localización las Comarcas de Aragón, como factores de localización se utilizan 8 criterios y como peso o importancia de esos factores los obtenidos mediante una encuesta. Además, se aporta evidencia sobre la consistencia de los resultados obtenidos. Como se aprecia en el trabajo, aunque la tipología de los problemas a los que van destinados ambos métodos son distintos, los resultados obtenidos pueden ser considerados como consistentes.

Palabras clave: localización empresarial; métodos ELECTRE; comarcas aragonesas.

Clasificación JEL: C49; R39.

2000MSC: 90B50; 91B06.

Industrial Location Decisions in Aragon: an Empirical Application of a Multi-criterion Methodology of Aid to the Decision using ELECTRE Methods I and III. Robustness of the Obtained Results

ABSTRACT

The location decision is one of the most important strategic decisions that have to take the firms, with direct influence on their competitiveness. For this reason, many methodologies have been developed to help in the decision making. In this paper, we make an empirical application of a multi-criterion methodology of aid to the decision; specifically, ELECTRE method has been chosen in its versions I and III, in order to select the best place for a firm. The alternatives are counties in Aragon (Spain), and we use eight criteria. The weight has been obtained by surveys. In addition, evidence is contributed about the consistency of the obtained results. Though the typology of the problems to which both methods are destined are different, the obtained results can be considered consistent.

Keywords: industrial location, ELECTRE methods, counties in Aragón (Spain).

JEL classification: C49; R39.

2000MSC: 90B50; 91B06.



1. INTRODUCCIÓN

Las últimas décadas han supuesto, para el entorno económico en el que operan las empresas, importantes cambios y transformaciones que han provocado variaciones sustanciales en su estructura productiva. Estos cambios se producen a un ritmo tan acelerado que los modelos de respuesta estratégica a los desafíos competitivos del pasado inmediato no constituyen modelos válidos a la hora de encaminar el futuro de los negocios. Este nuevo escenario económico está propiciando la aparición de nuevas pautas de comportamiento de las industrias a la hora de afrontar sus decisiones de inversión y localización (Ravelo, Mesa y otros, 2002). Esta decisión de localización presenta unas características que la hacen merecedora de especial atención: la escasa frecuencia con la que se toma, la complejidad de esa decisión y la subjetividad, especialmente en las pymes. Estas empresas de reducida dimensión suelen tomar estas decisiones sin valorar factores tradicionales de localización, sustentando esa decisión en factores mucho más subjetivos, como la proximidad al domicilio. Además, la localización es una de las decisiones estratégicas más importantes para cualquier empresa, ya que afectará a sus beneficios y costes a largo plazo, y resulta complicada y costosa de cambiar (Heizer y Render, 2001).

El componente espacial ha estado presente en la literatura económica desde sus orígenes. La teoría clásica del comercio entre naciones establecía que el comercio internacional entre países se determina según la abundancia relativa de factores de producción que los países posean. Porter (1990) amplía este concepto clásico añadiendo que los países difieren no solo en la dotación inicial de factores, sino también en sus capacidades para crear, mejorar y sustentar la innovación y el desarrollo tecnológico que se requiere para aumentar la productividad en determinadas industrias. Adam Smith (1776) introduce la idea de ventaja absoluta, por la que un país o región con costes bajos puede dominar el mercado exportando a otros.

La teoría de la localización ha experimentado un constante proceso de evolución a lo largo de los últimos años. Originariamente, Weber (1909), que fue considerado padre de esta teoría, basaba sus principios en los costes de transporte; más adelante siguieron y ampliaron sus teorías autores como Lösch (1940), Palander (1935) y Hoover (1948). La evolución de la teoría de la localización está íntimamente ligada a los grandes cambios estructurales que ha experimentado la actividad industrial, tales como las modificaciones en el tamaño de las empresas, la incorporación de nuevas tecnologías y la extensión y ampliación de los mercados. Poco tienen en común los factores que consideraban las pequeñas empresas de la pasada revolución industrial, centrados en el coste del transporte, la obtención de materias

primas y la optimización de la energía, frente a decisiones actuales que se centran en la apertura de nuevos mercados, especialmente en los países asiáticos, aunque es cierto que todavía en la actualidad sigue habiendo una fuerte presencia del factor coste.

Como se ha indicado, una de las características más importantes de la decisión de localización es la complejidad de esa decisión. Esa complejidad se percibe, entre otros, por el hecho de que esa decisión se toma en función de varios criterios u objetivos y no en base a uno único. Por este motivo, investigadores de diversas áreas de conocimiento han desarrollado en las últimas décadas un paradigma de decisión alternativo al paradigma de decisión clásico monocriterio: la decisión multicriterio. Dentro de este nuevo paradigma multicriterio de ayuda a la toma de decisiones, la Metodología ELECTRE ha alcanzado especial relevancia y ha sido utilizada ampliamente por investigadores y profesionales. Esta metodología está basada en el concepto de superación, que indica el grado de dominación de una alternativa sobre otra. Para su aplicación requiere de un conjunto de alternativas de elección, que en este caso son localizaciones potenciales de una empresa, y un conjunto de atributos o criterios medibles sobre cada localización. El desarrollo de esta metodología ELECTRE ha supuesto la aparición de diversas versiones, siendo la más utilizada la versión III. Esta versión sirve para la resolución de tipos de problemas de ordenación de alternativas, desde la mejor hasta la peor. También la versión I de esta metodología ha sido utilizada en abundantes investigaciones y aplicaciones prácticas, aunque su objetivo no es resolver problemas de ordenación sino problemas de selección de un subconjunto de alternativas de entre un conjunto dado.

Este trabajo aborda un problema de localización empresarial. Vamos a suponer que una empresa se enfrenta a la decisión de localizar una nueva instalación, fruto tanto de nueva implantación como de una ampliación de sus necesidades. Esa decisión se va a tomar en función de una serie de criterios que han sido seleccionados de una revisión de la literatura. Para el desarrollo de esta metodología es necesario, además de unos criterios sobre los que sustentar la decisión, un conjunto de alternativas que van a ser comparadas en base a esos criterios. Las alternativas de localización son una nueva división administrativa que se ha implantado en la Comunidad Autónoma de Aragón: las comarcas.

La estructura del trabajo es la que se indica a continuación: en el siguiente apartado se explica la metodología que se va a seguir en el trabajo, presentando los modelos ELECTRE I y ELECTRE III, utilizados en el análisis; a continuación se ha realizado un análisis empírico dividido en dos etapas; en la primera se han obtenido, mediante la utilización de encuestas,

qué factores son determinantes en las decisiones de localización y en la segunda se han aplicado los modelos ELECTRE I y III a los datos referidos a las Comarcas Aragonesas; en el último apartado se presentan las conclusiones.

2. LOS MODELOS ELECTRE I Y ELECTRE III COMO MÉTODOS DE ELECCIÓN MULTICRITERIO DISCRETO

La necesidad de tomar decisiones sobre la base de información estructurada y objetiva impera, entre otros, en el contexto de las Ciencias Sociales y, más concretamente, en la Administración de Empresas, en cuyo ámbito aparecen frecuentemente problemas en los que la selección de soluciones posibles debe realizarse en presencia de criterios múltiples y de diversa naturaleza. La introducción de problemas en los que se han de considerar varios atributos en su resolución ha impulsado el desarrollo de una disciplina científica conocida como Decisión Multicriterio, que arranca alrededor de 1950. Los métodos de toma de decisión multicriterio son unas poderosas herramientas que ayudan a generar consenso en contextos complejos de decisión. Se pueden aplicar estas técnicas a casos en los que sea necesaria la confluencia de intereses y puntos de vista de diferentes grupos o personas. También tienen la ventaja de posibilitar el análisis de problemas teniendo en cuenta factores que de otra forma se escaparían al decisor. Una de las ramas de esta disciplina científica, la Programación Multicriterio Discreta, consiste en el estudio de los casos en los que el número de soluciones posibles es finito y, habitualmente, se tiene en cuenta como criterio de decisión la optimización de los objetivos propuestos para la empresa. Puesto que en las decisiones de localización de las empresas es normal considerar un conjunto de atributos de las alternativas consideradas, convendría analizar el problema que se nos plantea dentro de la metodología proporcionada por la Decisión Multicriterio. En concreto, se ha optado por emplear el método “Elimination and Choice Translating Algorithm” (ELECTRE), desarrollado por Roy (1968).

Para decidir la elección de una alternativa o conjunto de alternativas de localización que pueden considerarse satisfactorias¹, este método se basa en una “relación de superación” entre cada par de alternativas, que posteriormente hay que aceptar o rechazar. Para verificar que una alternativa de localización a_j supera a otra a_k , el método utiliza dos principios:

- principio de concordancia, que requiere que en la mayoría de los criterios se pueda afirmar que efectivamente la alternativa a_j supera a la alternativa a_k ;

¹ Es lo que Barba-Romero y Pomerol (1997) definen como solución o soluciones eficientes, no dominadas u óptimas en el sentido de Pareto.

- principio de no discordancia, que requiere que, respecto a los criterios para los que la alternativa a_j no supera a la alternativa a_k , la diferencia entre ambas alternativas no sea muy grande.

A lo largo del tiempo se han ido proponiendo varias versiones del método ELECTRE y en este trabajo se van a utilizar las versiones I y III. La razón por la que se van a realizar los cálculos con ambas versiones es porque cada una de ellas aplica supuestos diferentes para los cálculos a realizar y para el establecimiento de los límites a considerar para la aceptación de la relación de superación entre alternativas; así, en la versión I se fijan unos umbrales de concordancia y discordancia que marcan los límites para aceptar la relación de superación, mientras que en la versión III la reflexión gira en torno al grado de credibilidad que se le otorga a la relación de superación. Además, los objetivos perseguidos son distintos: la versión I busca un conjunto de soluciones que superan a las demás, mientras que la versión III busca una ordenación de todas las alternativas, desde la mejor a la peor. De esta forma se puede analizar la coherencia y robustez entre los resultados obtenidos por ambas versiones, pudiendo analizar las eventuales discrepancias en virtud de los diferentes supuestos aplicados en los cálculos. Por lo tanto, aunque la filosofía de la que parten es la misma (una ayuda multicriterio en el proceso de toma de decisiones), tanto los objetivos como el algoritmo utilizado son distintos: mientras que el ELECTRE I proporciona como solución un subconjunto tan restringido como sea posible de alternativas no superadas, el ELECTRE III proporciona una ordenación de todas las alternativas. El algoritmo utilizado también varía, ya que en la versión III se introduce una graduación (fuerte o débil) en la relación de superación que ambos métodos utilizan. Por otra parte, las versiones primeras del método ELECTRE no se consideran superadas por las posteriores, como muestra el hecho de la aplicación de la versión I en trabajos recientes como los de Almeida (2005), Nowak (2005), Vasiloglou (2004) y Morais y otros (2006). A continuación se exponen de forma sucinta las bases teóricas de ambas metodologías.

2.1. El método ELECTRE I

Vamos a suponer que una empresa se enfrenta a la elección de una inversión entre dos alternativas a_j y a_k . Supongamos que esa empresa considera n factores relevantes de localización o criterios cada uno de ellos y además esa empresa le otorga una importancia o ponderación de w_i a cada uno de ellos. Por último, vamos a considerar que el valor numérico que la alternativa j tiene respecto al criterio i se expresa por E_{ij} .

Si se comparan los valores de las alternativas anteriores respecto de uno de los criterios considerados (criterio i), solo se pueden producir una de las tres situaciones siguientes, según el método ELECTRE I:

Si $E_{ij} > E_{ik} \Leftrightarrow a_j$ es preferida a a_k .

Si $E_{ij} = E_{ik} \Leftrightarrow a_j$ es indiferente con a_k .

Si $E_{ij} < E_{ik} \Leftrightarrow a_k$ es preferida a a_j .

Este método utiliza dos índices que sintetizan la información de las comparaciones por parejas de los valores E_{ij} y E_{ik} . Estos dos índices sirven para afirmar que una alternativa a_j es mejor que otra alternativa a_k cuando las valoraciones en la mayoría de los criterios son mejores y que, en aquellos que no es mejor, no lo es de forma notoria. El índice de concordancia (IC_{jk}) entre esas dos localizaciones, a_j y a_k , se calcula mediante la siguiente expresión:

$$IC_{jk} = \frac{\sum_{\forall i, E_{ij} \geq E_{ik}} w_i}{\sum_{i=1}^n w_i} \quad (1)$$

El numerador de este cociente es la suma de las ponderaciones w_i asignadas a los factores para los que la localización a_j tiene una puntuación que es mayor o igual a la de la localización a_k , y su denominador es la suma total de las ponderaciones asignadas al conjunto de los factores. Este índice toma valores comprendidos entre 0 y 1. Cuanto mayor es el valor del índice, mejor es la localización a_j frente a la localización a_k . Si el índice toma el valor 1, esto significa que la alternativa a_j es mejor que la alternativa a_k en todos los criterios considerados. Por el contrario, si toma el valor 0 indica que no es mejor en ninguno de ellos.

El índice de discordancia para ese par de alternativas a_j y a_k (ID_{jk}) se calcula, para aquellos factores en los cuales la localización a_j no domina a la a_k , como el mayor de los cocientes entre la diferencia en la puntuación para un factor i entre a_j y a_k y la amplitud total de la escala considerada para ese factor (TE_i), de acuerdo con la siguiente expresión:

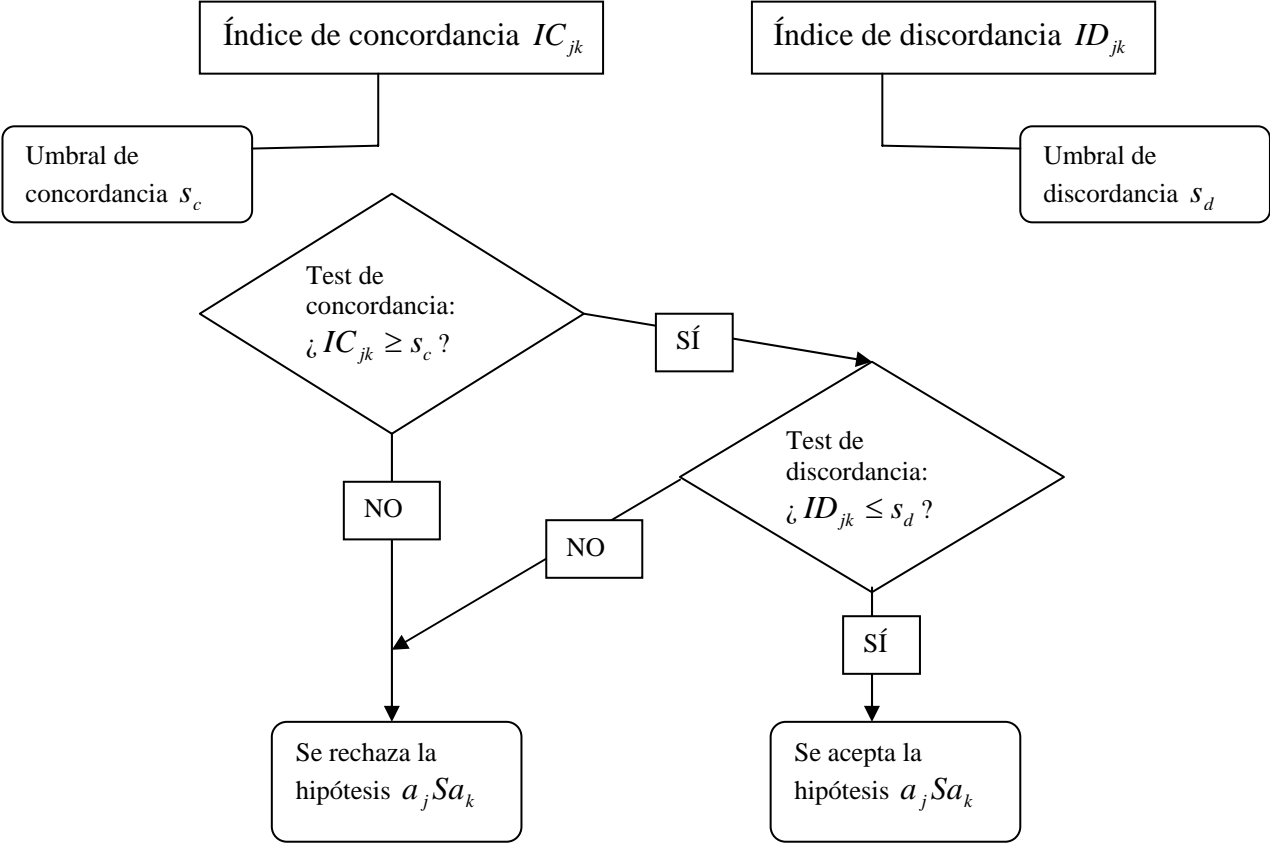
$$ID_{jk} = \text{Max}_i (\forall i, E_{ij} \leq E_{ik}) \left\{ \frac{E_{ik} - E_{ij}}{TE_i} \right\} \quad (2)$$

El índice de discordancia toma valores comprendidos entre 0 y 1, resultando el valor 0 cuando la localización a_j no es inferior en factor alguno a la localización a_k (lo que implica que el índice de concordancia IC_{jk} ha de ser 1) y siendo 1 cuando, en al menos un factor, la

divergencia a favor de la localización k es la totalidad de la escala; cuanto más grande sea la infravaloración de la localización a_j respecto la a_k en al menos un factor, mayor valor tomará el índice de discordancia.

Una vez calculados los índices de concordancia y discordancia, para poder afirmar que una alternativa supera a otra, hay que determinar los valores de estos índices a partir de los cuales se admite la hipótesis de superación, ya que si se utilizan los valores extremos (1 y 0), estaríamos suponiendo que solo se podría afirmar esa superación en los casos en los que una alternativa esté mejor valorada que la otra en todos los criterios de comparación considerados. Para ello, se introducen unos umbrales de concordancia y discordancia, denotados por s_c y s_d , respectivamente. El umbral de concordancia expresa el mínimo de concordancia requerido para que la proposición “ a_j supera a a_k ” no sea rechazada. El umbral de discordancia expresa el máximo de discordancia tolerado para que la hipótesis “ a_j supera a a_k ” no sea rechazada. En consecuencia, una vez introducidos los umbrales de concordancia y discordancia por el decisor, se define la relación de superación S como: $a_j S a_k$ si y solamente si $IC_{jk} \geq s_c$ y $ID_{jk} \leq s_d$. Los pasos a seguir teniendo en cuenta lo anterior aparecen resumidos en el Gráfico 1.

Gráfico 1. Secuencia de aplicación del método ELECTRE I.



Una vez calculadas las relaciones de superación entre cada par de alternativas, se pueden separar las alternativas en dos grupos: aquellas alternativas que no se ven superadas por ninguna otra y aquellas alternativas que son superadas por al menos una de ellas. El primer grupo de alternativas recibe el nombre de núcleo y se denota por N .

Si el conjunto de alternativas A es finito y la relación de superación cumple la propiedad transitiva, el núcleo estaría formado por el conjunto de soluciones eficientes del problema multicriterio en el sentido dado por Barba-Romero y Pomerol (1997). Las alternativas eficientes son óptimas en el sentido de Pareto, ya que no están dominadas estrictamente por ninguna otra; dicho de otro modo, no se puede encontrar otra alternativa que sea mejor o igual en todos los criterios y estrictamente mejor para alguno de ellos. En los casos en los que no se cumple la propiedad transitiva en la relación de superación, el núcleo no existe o, si existe, puede no ser único y no se puede hablar de alternativas eficientes. Se consigue clasificar en dos grupos las alternativas de localización: por una parte se tiene las pertenecientes al núcleo que no son superadas por ninguna alternativa, y por otra, las de fuera del núcleo, que son superadas al menos por una, pertenezca o no al núcleo.

2.2. El método ELECTRE III

La modelización de las preferencias del decisor en la versión ELECTRE III es menos rígida, ya que se toma como punto de partida el siguiente argumento: si la diferencia entre las valoraciones de las alternativas a_j y a_k es muy pequeña, ¿el decisor continuará prefiriendo una de ellas?; ¿es esa pequeña diferencia razón suficiente para hacer más preferida una que la otra? El método ELECTRE III consigue esta flexibilización con la introducción de tres nuevos umbrales, definidos para cada uno de los criterios considerados:

- a) el umbral de preferencia (p), que sería la magnitud en que exigimos que la valoración de la alternativa a_j sea mayor que la de la alternativa a_k para poder hablar de preferencia fuerte de la primera respecto a la segunda.
- b) el umbral de indiferencia (q), que sería la magnitud en que permitimos que la valoración de la alternativa a_k sea menor que la de la alternativa a_j para seguir siendo indiferentes.
- c) el umbral de veto (v), que sería la magnitud de la diferencia ($a_j - a_k$) a partir de la cual nunca se aceptará que la alternativa k supere a la alternativa j , aunque la supere en el resto de criterios.

Estos tres umbrales son valores no negativos y además ordenados ($p, q, v \geq 0$; $q \leq p \leq v$).

Ahora la comparación no se realiza únicamente a partir de la valoración de cada alternativa respecto a los criterios definidos en el desarrollo del modelo, sino que tiene en cuenta los umbrales de preferencia p , de indiferencia q , y de veto v , lo que supone la introducción de lo que esta metodología denomina pseudo-criterios². Por lo tanto, con la introducción de estos umbrales la modelización de las preferencias será³:

Si $-q_i \leq E_{ij} - E_{ik} \leq q_i \Leftrightarrow a_j$ es indiferente con a_k .

Si $q_i < (E_{ij} - E_{ik}) \leq p_i \Leftrightarrow a_j$ es preferido de forma débil sobre a_k .

Si $p_i < (E_{ij} - E_{ik}) \leq v_i \Leftrightarrow a_j$ es preferido fuertemente a a_k .

Si $(E_{ij} - E_{ik}) \geq v_i \Leftrightarrow$ la posibilidad de superación de a_k sobre a_j es vetada.

Si $-p_i < (E_{ij} - E_{ik}) \leq -q_i \Leftrightarrow a_k$ es preferido de forma débil sobre a_j .

Si $-v_i < (E_{ij} - E_{ik}) \leq -p_i \Leftrightarrow a_k$ es preferido fuertemente a a_j .

Si $(E_{ij} - E_{ik}) \geq -v_i \Leftrightarrow$ la posibilidad de superación de a_j sobre a_k es vetada.

Esta versión del método ELECTRE utiliza también dos índices: el índice de concordancia y el de discordancia. A diferencia de la versión I, el índice de concordancia es calculado obteniendo primero un índice de concordancia por criterio y luego un índice de concordancia global. El índice de concordancia por criterio afirma en qué medida la alternativa a_j es al menos tan buena como la alternativa a_k para el factor i . Se denota por $c_i(a_j, a_k)$ y se define como:

$$\begin{aligned} c_i(a_j, a_k) &= 1 \Leftrightarrow (E_{ik} - E_{ij}) \leq q_i \\ 0 < c_i(a_j, a_k) < 1 &\Leftrightarrow q_i < (E_{ik} - E_{ij}) \leq p_i \quad * \\ c_i(a_j, a_k) &= 0 \Leftrightarrow p_i < (E_{ik} - E_{ij}) \end{aligned}$$

* El valor final del índice se calcula mediante interpolación lineal.

El índice de concordancia global se obtiene a partir de la siguiente expresión:

$$IC_{jk} = \frac{\sum_{i=1}^m w_i * c_i(a_j, a_k)}{\sum_{i=1}^m w_i}$$

² Un criterio verdadero es un pseudo-criterio con umbrales de preferencia p , de indiferencia q y de veto v , iguales a cero.

Para el cálculo del índice de concordancia, en la versión III del método ELECTRE se suman los pesos de los criterios, excluyendo aquéllos en los que su índice de concordancia sea 0. Además, en aquellos criterios cuyo índice de concordancia esté comprendido entre 0 y 1, el peso del criterio se minora en el importe del propio índice. El índice de discordancia para cada criterio i [$d_i(a_j, a_k)$] se define como:

$$\begin{aligned} d_i(a_j, a_k) &= 1 \Leftrightarrow v_i < E_{i_k} - E_{i_j} \\ 0 < d_i(a_j, a_k) < 1 &\Leftrightarrow p_i < (E_{i_k} - E_{i_j}) \leq v_i \quad * \\ d_i(a_j, a_k) &= 0 \Leftrightarrow (E_{i_k} - E_{i_j}) \leq p_i \end{aligned}$$

* El valor final del índice se calcula mediante interpolación lineal.

Para cada par de alternativas de (a_j, a_k) ahora existe una medida de concordancia y otra de discordancia. En la etapa final de los cálculos se combinan ambas medidas para generar una medida del grado de superación, esto es, una matriz de credibilidad que valora la fortaleza de la afirmación “la alternativa a_j supera a la alternativa a_k ”. Previamente hay que definir, para cada par de alternativas, el subconjunto \bar{F} de criterios que tiene como elementos aquéllos para los cuales el índice de discordancia por criterio es superior al índice de concordancia global:

$$\bar{F} = \left\{ i / i \in F, d_i(a_j, a_k) > C_{jk} \right\}$$

El grado de credibilidad (δ_{jk}) para cada par de alternativas (a_j, a_k) se define como:

$$\delta_{jk} = \begin{cases} C_{jk} & \text{si } \bar{F} = \text{conjunto vacío} \\ C_{jk} * \prod_{i \in \bar{F}} \frac{1 - d_i(a_j, a_k)}{1 - C_{jk}} & \text{en otros casos} \end{cases}$$

De la definición del grado de credibilidad se obtiene que éste será el mismo que el valor del índice de concordancia global si ningún valor de los índices de discordancia por criterio es mayor que él. Si para algún criterio ese valor del índice de discordancia por criterio es mayor que el índice de concordancia global, el grado de credibilidad será el índice de concordancia global disminuido por la discordancia. De igual forma, el grado de credibilidad de cualquier par de alternativas será 0 si, para algún criterio, el valor de la discordancia fuera 1.

³ Siempre que el criterio sea a maximizar; si el criterio es a minimizar, habría que invertir el orden de la diferencia.

El siguiente paso en la versión III del método ELECTRE es ordenar las alternativas a partir de los resultados obtenidos para el grado de credibilidad, llevando a cabo un proceso de destilación descendente y otro ascendente⁴. De ambos procesos se obtienen dos ordenaciones intermedias, que finalmente se aúnan en una ordenación final.

Este método proporciona, por lo tanto, una ordenación final de las alternativas que es una combinación de las obtenidas mediante las destilaciones ascendente y descendente, de tal forma que cuando en ambas una alternativa supera a otras, así lo refleja la ordenación final, mientras que si entre las dos ordenaciones hay discrepancia respecto a qué alternativa supera a otra, la ordenación final las considera “incomparables”. Este hecho provoca que la ordenación final sea parcial.

3. ANÁLISIS EMPÍRICO DE LA LOCALIZACIÓN INDUSTRIAL EN LAS COMARCAS ARAGONESAS

Para cumplir el objetivo propuesto, un paso previo tiene que ser determinar cuáles son los factores o criterios determinantes en la localización empresarial por parte de las empresas, además de la importancia o peso de esos criterios considerados importantes.

3.1. Factores determinantes de la localización industrial y su importancia por parte de los empresarios aragoneses e instituciones

El estudio y análisis de los factores de localización conlleva el estudio de la jerarquización de esos factores por parte del agente decisor. La literatura ha estudiado con interés ese proceso de jerarquización o de ordenación de los factores, pudiendo agrupar los trabajos, siguiendo a Aydalot (1985), en dos grupos:

- a) El grupo que utiliza datos secundarios y, mediante el tratamiento de estos datos con técnicas estadísticas, pretende inferir la importancia de esos factores.
- b) El grupo que utiliza datos primarios mediante la utilización de encuestas encaminadas a ordenar la incidencia de esos factores.

Este trabajo se enmarca en el segundo grupo. Se realizó una encuesta con dos objetivos: identificar los factores relevantes en las decisiones de localización y obtener una ponderación o importancia de esos factores.

Las organizaciones a las que se les envió el cuestionario eran de dos tipos: empresas

⁴ Para un mayor detalle del proceso véase Maystre y otros (1994).

que se habían instalado o aumentado sus instalaciones en los polígonos industriales de las Comarcas Aragonesas y a instituciones públicas que fomentan la actividad industrial. Además, se envió el cuestionario a los agentes sociales presentes en la Comunidad Autónoma de Aragón.

La encuesta fue enviada en el primer semestre de 2004, por correo electrónico, a 40 empresas instaladas en la Comunidad Autónoma Aragonesa. Se envió también a las siguientes asociaciones empresariales: Confederación Regional de Empresarios de Aragón (CREA), Confederación de Empresarios de Zaragoza (CEZ), Confederación Empresarial Turolense (CET), Confederación Empresarial Oscense (CEOS) y la Confederación de la Pequeña y Mediana Empresa, tanto regional (Cepyme Aragón) como de cada una de las tres provincias aragonesas. Respecto a los Organismos oficiales, se envió al Gobierno de Aragón (en concreto, a los Departamentos de Economía, Obras Públicas e Industria y al Instituto Aragonés de Fomento), al Instituto del Suelo y la Vivienda de Aragón y a los Ayuntamientos de todas las cabeceras de comarca, a través de sus Agentes de Desarrollo Local. En lo que se refiere a promotores industriales, se contactó tanto con privados (Empesarium) como con públicos (PLAZA y SEPES). Se puede considerar que la cantidad de respuestas recibidas es significativa, puesto que respondieron 19 de las 40 empresas contactadas, el 50% de los Ayuntamientos y el 100% de las asociaciones de empresarios y de los organismos pertenecientes al Gobierno de Aragón.⁵

Las cuestiones sobre las que se preguntó a estas organizaciones están basadas en las propuestas por Ulgado (1995), excepto en aquellos factores de localización que no tienen variación entre comarcas (nivel de sindicación, disponibilidad de aeropuerto, coste del capital y su disponibilidad, nivel impositivo, clima, subsidios familiares, ayudas públicas). En la Tabla 1 se muestran los ocho factores que ya estaban fijados en la encuesta que se envió, junto con la fuente estadística utilizada.

⁵ Lu (2003) utiliza 91 respuestas sobre 485 envíos, Fryxell, Butler y Choi (2004) utilizan una muestra con 67 respuestas sobre 130 envíos y Oum y Park (2004) utilizan una de 83 respuestas sobre 818 envíos.

Tabla 1. Factores de localización, variables de medida y referencias bibliográficas.

FACTOR	VARIABLE DE MEDIDA	REFERENCIAS	FUENTE
Coste del suelo industrial	Precio por metro cuadrado (medido en euros).	Stonebraker y Leong 1994; Hekman 1992.	Instituto Aragonés de Fomento (1), promotores públicos y privados.
Cercanía a la capital de la provincia	Kilómetros de distancia hasta la capital de la provincia.	Crozet, Mayer y Muccielli 2003; Arauzo 2005.	Ministerio de Fomento (2).
Accesibilidad a redes de transporte	Distancia a la carretera nacional, autovía, puerto y aeropuerto más próximos.	Barrios, Gorg y Strobl 2006; Arauzo 2005.	Ministerio de Fomento (2).
Infraestructura	Comunicaciones (redes telefónicas y fibra óptica), equipamientos (pavimentación, iluminación redes antiincendios), agua (acceso a agua potable, acometidas y depósitos), energía (gas y electricidad), gestión medioambiental (redes de saneamientos, tratamiento de aguas residuales, canon por utilización).	Baaij, Van den Bosch y Volberda 2004.	Instituto Aragonés de Fomento (1).
Tamaño del polígono	Superficie total del polígono (medida en m ²).	Woodward 1992.	Instituto Aragonés de Fomento (1), Ayuntamientos.
Grado de ocupación del polígono	Porcentaje de ocupación del polígono y posibilidad o no de ampliación.	Woodward 1992.	Instituto Aragonés de Fomento (1), Ayuntamientos.
Tamaño de la población	Número de habitantes.	Guimaraes, Figueiredo y Woodward 2000. Barrios, Gorg y Strobl 2006.	Instituto Nacional de Estadística (3).
Categoría de la población	Densidad de población, peso del sector industrial, crecimiento vegetativo, cualificación de mano de obra, infraestructura del municipio (nº de farmacias, nº bancos, nº de bibliotecas, nº de establecimientos hosteleros).	Baaij, Van den Bosch y Volberda 2004.	Instituto Nacional de Estadística (3), Instituto Aragonés de Estadística (4), Anuario de “El País” 2003.

(1) www.iaf.es

(3) www.ine.es

(2) www.mfom.es

(4) www.aragob.es

En los cuestionarios enviados se preguntaba por la importancia que se otorgaba a los 8 criterios de elección considerados, ya que la aplicación de los dos métodos ELECTRE requieren de la ponderación de los criterios. Una vez obtenidas las respuestas, y con el fin de facilitar la interpretación de los resultados, se ha calculó una media aritmética de los valores asignados a cada factor, cuyo resultado aparece en la Tabla 2.

Tabla 2. Ponderación de los factores de localización.

FACTOR	PONDERACIÓN
Coste del suelo industrial	7,67
Cercanía a la capital de provincia	5,53
Accesibilidad a redes de transporte	6,47
Infraestructura	8,11
Tamaño del polígono	7,27
Grado de ocupación del polígono	4,73
Tamaño de la población	5,27
Categoría de la población	5,73

Para la aplicación de los modelos ELECTRE en este trabajo, es necesaria la normalización previa de todos los datos a utilizar, ya que en este trabajo se consideran factores medidos por diferentes variables que pueden estar expresadas en unidades o escalas de medida no comparables. En este caso se ha realizado una transformación sobre una escala de tal forma que, una vez realizada la normalización, todas las alternativas respecto de cada una de las variables pasan a tener un valor comprendido entre 0 y 10. Más en concreto, la normalización ha consistido en el cálculo de un valor que es:

- a) si el criterio es a maximizar:

$$10 - \frac{(\text{valor a normalizar} - \text{máximo de la serie}) * 10}{(\text{Mínimo} - \text{Máximo}) \text{ de la serie}}$$

- b) si el criterio es a minimizar:

$$10 - \frac{(\text{valor a normalizar} - \text{mínimo de la serie}) * 10}{(\text{Máximo} - \text{Mínimo}) \text{ de la serie}}$$

De esta forma, el valor más próximo a diez reflejaría el mejor valor y el más alejado a diez sería el peor valor. En los casos en que la medición de un factor de localización se hace únicamente a través de una variable, el valor normalizado de ésta es el que se asigna al factor. Por otra parte, la valoración dada a los factores medidos a través de varias variables es la

media aritmética de su valoración normalizada. La Tabla 3 presenta los datos obtenidos por las alternativas respecto de los criterios utilizados, una vez que han sido normalizados.

Tabla 3. Valores alcanzados por las alternativas respecto a los criterios, una vez normalizados.

	Precio venta	Localización	Comunicación	Equipamientos	% ocupación	Tamaño	Tamaño población	Categoría población
ALBARRACÍN	9,60	3,32	5,23	7,22	2,38	3,75	2,50	5,51
ALTO GALLEGO	8,64	3,88	7,70	8,26	6,43	3,75	6,00	5,73
ANDORRA	9,83	5,83	5,88	8,40	3,99	5,25	4,50	4,07
ARANDA	9,07	6,26	6,67	7,40	9,14	3,67	5,00	5,26
BAJO ARAGÓN	9,80	5,99	5,93	7,67	7,76	4,88	5,13	4,56
BAJO CINCA	9,20	4,88	7,83	7,13	6,66	4,75	6,25	4,52
BAJO MARTÍN	9,90	7,20	7,29	8,15	3,29	6,00	4,50	4,09
BARBASTRO	9,71	5,56	7,03	7,42	7,77	5,33	4,00	6,37
BELCHITE	9,86	7,37	2,65	7,08	10,00	1,00	1,00	3,57
BORJA	9,48	7,02	7,88	7,87	3,32	5,00	4,50	4,08
CALAMOCHA	9,84	6,47	8,19	6,18	6,27	4,67	3,67	4,26
CALATAYUD	9,42	6,31	8,14	7,33	8,13	5,00	5,75	4,79
CARIÑENA	9,50	8,80	8,15	6,91	8,78	3,50	4,50	4,61
CASPE	9,51	5,17	8,12	6,34	5,31	3,40	4,80	3,98
CINCA MEDIO	9,42	4,17	8,03	8,09	9,90	8,00	7,00	6,11
CINCO VILLAS	9,69	6,73	7,27	7,19	7,22	5,50	5,25	4,71
CUENCAS MINERAS	9,95	6,50	6,19	6,97	6,75	4,60	4,60	4,00
DAROCA	9,62	7,80	6,61	4,49	4,86	4,00	4,00	4,55
GÚDAR-JAVALAMBRE	9,77	2,05	7,47	7,53	9,53	4,00	3,25	4,81
HUESCA	9,35	8,16	8,51	7,28	6,78	5,00	5,00	5,78
JALON MEDIO	9,30	8,36	8,31	8,43	5,97	7,40	4,80	4,31
LA LITERA	9,54	6,40	6,84	7,42	3,75	5,50	4,00	4,10
MATARRAÑA	9,70	3,03	5,43	7,64	8,14	3,33	3,67	3,70
MONCAYO	8,96	5,12	8,06	7,92	6,84	7,50	7,00	5,56
MONEGROS	9,77	7,26	7,69	6,99	6,28	4,60	4,80	4,90
RIBAGORZA	9,10	2,23	5,58	8,20	9,40	5,00	5,00	4,22
RIBERA ALTA DEL EBRO	8,31	8,54	8,63	6,69	7,37	5,00	4,00	4,29
RIBERA BAJA DEL EBRO	9,90	8,16	8,10	7,90	6,67	4,67	4,67	3,86
SOBRARBE	7,54	0,78	3,94	6,52	8,78	3,50	3,00	5,02
TERUEL	9,75	4,99	7,66	7,55	5,64	5,60	4,60	5,29
ZARAGOZA	6,38	9,44	8,71	7,06	8,14	5,60	4,55	6,02

3.2. Resultados obtenidos

En este trabajo, la información relativa a los factores de localización considerados se ha obtenido de los polígonos industriales existentes en Aragón. Sin embargo, las alternativas de localización consideradas son las comarcas y, por lo tanto, en aquéllas en las que existe más de un polígono industrial los datos correspondientes a la comarca son el promedio lineal de los polígonos existentes en ella⁶.

⁶ Para la decisión final de una empresa referente a qué polígono industrial elegir para su localización, tendría dos

3.2.1. Resultados del modelo ELECTRE I

En principio, el método ELECTRE I considera exclusivamente aquellas alternativas en las que se produzca la concordancia total, es decir, aquellas en las que $s_c=1$ y $s_d=0$. Tal como se muestra en la Tabla 4, este alto nivel de exigencia lleva a que de las 31 comarcas aragonesas consideradas, solo salgan del núcleo las comarcas Albarracín y Sobrarbe, con lo que el núcleo estaría formado por 29 alternativas.

Tabla 4. Alternativas pertenecientes al núcleo para distintos niveles de umbrales de concordancia y discordancia.

Umbrales	Alternativas del núcleo
$S_c=1$ $S_d=0$	Alto Gállego, Andorra, Aranda, Bajo Aragón, Bajo Cinca, Bajo Martín, Barbastro, Belchite, Borja, Calamocha, Calatayud, Cariñena, Caspe, Cinca Medio, Cinco Villas, Cuencas Mineras, Daroca, Gudar-Javalambre, Huesca, Jalón Medio, La litera, Matarraña, Moncayo, Monegros, Ribagorza, Ribera Alta del Ebro, Ribera Baja del Ebro, Teruel, Zaragoza.
$S_c=0,9$ $S_d=0,1$	Alto Gállego, Andorra, Aranda, Bajo Aragón, Bajo Cinca, Bajo Martín, Barbastro, Belchite, Borja, Calamocha, Calatayud, Cariñena, Caspe, Cinca Medio, Cinco Villas, Cuencas Mineras, Daroca, Gudar-Javalambre, Huesca, Jalón Medio, La litera, Matarraña, Moncayo, Monegros, Ribagorza, Ribera Alta del Ebro, Ribera Baja del Ebro, Teruel, Zaragoza.
$S_c=0,8$ $S_d=0,2$	Aranda, Bajo Aragón, Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Cinco Villas, Huesca, Jalón Medio, Moncayo, Ribera Baja del Ebro, Zaragoza.
$S_c=0,7$ $S_d=0,3$	Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Huesca, Jalón Medio, Zaragoza.

Para intentar aumentar las relaciones de dominio y reducir el número de alternativas que contiene el núcleo, se pueden disminuir los niveles de exigencia en cuanto a la concordancia y la discordancia (Ravelo Mesa y otros (2002) y Barba-Romero y Pomerol (1997)). Conforme vamos disminuyendo los umbrales de concordancia y discordancia, van disminuyendo las alternativas del núcleo y se van haciendo más robustas las relaciones que se establecen. En el caso de que se fijen unos umbrales de $s_c=0,7$ y $s_d=0,3$,⁷ las alternativas incluidas en el núcleo han disminuido y está formado por Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Huesca, Jalón Medio, Zaragoza.

formas alternativas de actuar: o bien, una vez elegida la comarca a partir de los cálculos que se van a presentar en este trabajo, repite unos cálculos similares para los polígonos industriales de esa comarca, o bien aplica el método ELECTRE directamente a los polígonos industriales.

⁷ Estos niveles son los mismos que fijan Ravelo Mesa y otros (2002) y Barba-Romero y Pomerol (1997). Maestre y otros (1994) los fija en 0.65 y 0.35 para el umbral de concordancia y el de discordancia, respectivamente. El valor de los umbrales puede estar entre 0 y 1, pero valores inferiores a 0,5 para el índice de concordancia carecen de significación.

3.2.2. Resultados del modelo ELECTRE III

Para la aplicación de esta metodología, hay que fijar el valor de los umbrales que van a ser utilizados. La fijación de los umbrales de indiferencia, de preferencia estricta y de veto es siempre subjetiva y según Maystre (1994) no son el resultado de un proceso de búsqueda de datos exactos, sino más bien la oportunidad de reflejar en ellos lo que hay de arbitrario en los datos.⁸ En los datos que se han utilizado, hay que tener en cuenta que previamente han sido normalizados en una escala de 1 a 10, por lo tanto, los umbrales tienen que ser los mismos para todos los criterios. Siguiendo a Diop (1988), el umbral de indiferencia se ha fijado en 1 y el de preferencia estricta en 2 y no se ha introducido umbral de veto, ya que por la normalización de los valores de las alternativas respecto de los criterios considerados la diferencia entre estos valores son muy pequeñas y la introducción de valores altos de este parámetro carece de sentido en este caso, ya que en algunos criterios el umbral de preferencia estricta está jugando este papel. La Tabla 5 muestra los resultados obtenidos en la ordenación final de las alternativas consideradas en este trabajo.

Tabla 5. Ordenación final de las alternativas utilizando el método ELECTRE III.

Posición	Alternativa	Posición	Alternativa
1	Cinca Medio	12	Aranda
2	Jalón Medio	13	Alto Gállego Calamocha
3	Huesca	14	Bajo Aragón
4	Moncayo	15	Cuencas Mineras La Litera
5	Cariñena	16	Gúdar
6	Calatayud	17	Andorra
7	Ribera Baja Zaragoza	18	Caspe
8	Cinco villas Ribera Alta	19	Ribagorza
9	Barbastro Monegros	20	Daroca
10	Bajo Cinca Bajo Martín	21	Belchite Matarraña
11	Borja Teruel	22	Albarracín
		23	Sobrarbe

A la vista de los resultados obtenidos, si el objetivo que persigue el decisor es invertir en la mejor alternativa, aquélla será en Cinca Medio, que es la que aparece en primer lugar, seguida del Jalón Medio, Huesca y el Moncayo.

⁸ Rogers y Buen (1998), realizan una aproximación alternativa para fijar estos valores de forma más objetiva en los casos de impacto ambiental del ruido.

3.2.3. Análisis de Robustez

El análisis de robustez consiste en hacer variar los parámetros que se han utilizado en los algoritmos en torno a su valor inicial y si los nuevos resultados obtenidos no son modificados de manera importante respecto a los primitivos, la recomendación que se haga será robusta.

Los parámetros susceptibles de variación (debido bien a la incertidumbre de los datos o a la subjetividad) varían en función del modelo utilizado. Así, en el ELECTRE I, los parámetros a verificar podrían ser las amplitudes de las escalas de los criterios, los pesos de los criterios y los umbrales de concordancia y discordancia. En el ELECTRE III, los parámetros a contrastar son las variaciones en los pesos y las variaciones en los umbrales de indiferencia, preferencia estricta y veto.

Para analizar el grado de robustez de los resultados obtenidos de la aplicación del ELECTRE III, en este trabajo se han modificado dos parámetros: los valores de los umbrales de indiferencia y de preferencia estricta⁹ y los valores de los pesos asignados a los criterios. En el caso del umbral de indiferencia el rango de variación ha sido desde 0.5 a 1.2 y el de preferencia estricta desde 2.2 hasta 1.5. Los resultados obtenidos, junto con los iniciales, se muestran en la Tabla 6, en la que se aprecia que los resultados son prácticamente los mismos.

A la vista de los resultados, se podría afirmar que el resultado obtenido resulta robusto y, por lo tanto, se podría realizar una recomendación fiable al decisor. Las posiciones extremas de la ordenación siempre son ocupadas por las mismas alternativas (Cinca Medio, Jalón Medio y Moncayo en la parte alta y Sobrarbe, Albarracín, Belchite y el Matarraña en la parte baja). Aunque se producen variaciones en las posiciones intermedias de la ordenación, estas diferencias son menores y solo suponen subir una posición, bajar una o formar parte de un grupo de alternativas incomparables, como por ejemplo Cariñena, o la comarca del Moncayo.

Continuando con el análisis de robustez del método ELECTRE III, también se han modificado los pesos asignados a los criterios. Partiendo de los resultados obtenidos, se han calculado los rangos de variación de los pesos asignados a los criterios para dentro de los cuales la ordenación de las alternativas seguiría siendo la misma. Dichos rangos aparecen en la Tabla 7.

⁹ El de veto no se modifica, ya que no se ha fijado ningún valor en el resultado original.

Tabla 6. Análisis de robustez, con variaciones en los umbrales de preferencia, indiferencia y veto, en el Electre III.

$q=1$ $p=2$ $v=0$		$q=0,9$ $p=1,9$ $v=0$		$q=0,8$ $p=1,8$ $v=0$		$q=0,5$ $p=1,5$ $v=0$	
Posición	Alternativa	Posición	Alternativa	Posición	Alternativa	Posición	Alternativa
1	Cinca Medio	1	Cinca Medio	1	Cinca Medio	1	Cinca Medio
2	Jalon Medio	2	Jalon Medio	2	Jalon Medio	2	Jalon Medio
3	Huesca	3	Huesca	2	Huesca	2	Huesca
4	Moncayo	4	Moncayo	3	Moncayo	3	Moncayo
5	Cariñena	5	Calatayud	4	Calatayud	3	Calatayud
6	Calatayud	6	Cariñena	5	Cariñena	4	Ribera Baja
7	Ribera Baja	7	Ribera Baja	5	Ribera Baja	4	Zaragoza
	Zaragoza	8	Zaragoza	5	Zaragoza	5	Cariñena
8	Cinco villas	9	Cinco Villas	6	Barbastro	6	Cinco Villas
	Ribera Alta		Ribera alta	7	Bajo Martín	7	Bajo Martin
9	Barbastro	10	Barbastro	7	Cinco Villas	8	Barbastro
	Monegros		Monegros	8	Monearos	8	Monearos
10	Bajo Cinca	11	Bajo Martin	8	Bajo Cinca	9	Ribera Alta
	Bajo Martín	12	Bajo Cinca	9	Ribera Alta	9	Alto Gallego
11	Borja		Borja	10	Alto gállego	9	Borja
	Teruel	13	Alto gállego	10	Borja	10	Bajo Cinca
12	Aranda		Aranda	11	Teruel	10	Teruel
13	Alto Gállego	14	Teruel	12	Aranda	11	Aranda
	Calamocha	15	Calamocha	13	Calamocha	12	Andorra
14	Bajo Aragón	16	Bajo Aragón	13	Cuencas Mine	12	Bajo Aragón
15	Cuencas Mi.	17	Cuencas Mi	14	Bajo Aragón	12	Calamocha
	La Litera		La Litera	14	Litera	12	Cuencas Mine
16	Gúdar	18	Gudar	15	Gudar	12	Litera
17	Andorra	19	Andorra	16	Andorra	13	Gudar
18	Caspe	20	Caspe	17	Ribagorza	13	Ribagorza
19	Ribagorza	21	Ribagorza	18	Caspe	14	Belchite
20	Daroca	22	Daroca	18	Daroca	14	Caspe
21	Belchite	23	Albarracin	19	Albarracin	15	Daroca
	Matarraña		Belchite	19	Belchite	16	Albarracin
22	Albarracin		Matarraña	20	Matarraña	17	Matarraña
23	Sobrarbe	24	Sobrarbe	21	Sobrarbe	18	Sobrarbe

$q=1,1$ $p=2,1$ $v=0$		$q=1,2$ $p=2,2$ $v=0$	
Posición	Alternativa	Posición	Alternativa
1	Cinca Medio	1	Cinca Medio
2	Jalon Medio	2	Jalon Medio
3	Huesca	3	Moncayo
4	Moncayo	4	Cariñena
5	Cariñena	5	Huesca
6	Calatayud	6	Calatayud
	Monegros	6	Zaragoza
7	Ribera Baja	7	Cinco Villas
	Zaragoza	7	Ribera Alta
8	Cinco Villas	7	Ribera Baja
	Ribera Alta		
9	Bajo Cinca	8	Barbastro
	Barbastro	8	Monegros
10	Bajo Martín	9	Aranda
11	Aranda	9	Bajo Cinca
	Teruel	9	Bajo Martín
12	Borja	10	Borja
13	Alto Gállego	10	Teruel
14	Calamocha	11	Alto Gállego
15	Bajo Aragón	11	Calamocha
	Cuencas Mine		
16	Gudar	12	Cuencas Mine
	La Litera	12	Gúdar
17	Andorra	13	Bajo Aragón
	Caspe	13	Litera
18	Daroca	14	Andorra
	Ribagorza	14	
19	Belchite	15	Caspe
	Matarraña	15	
20	Albarracin	16	Ribagorza
21	Sobrarbe	17	Matarraña
		17	
			Daroca
		18	Albarracin
			Belchite
		19	Sobrarbe

Tabla 7. Rango de variación de los pesos que proporcionan la ordenación obtenida previamente.

Criterio	Rango de variación
Coste del suelo industrial	(7.64, 7.87)
Cercanía a la capital de provincia	(5.42, 5.55)
Accesibilidad a redes de transporte	(6.10, 6.49)
Infraestructura	(7.9, 8.32)
Tamaño del polígono	(7.16, 7.40)
Grado de ocupación del polígono	(4.70, 4.76)
Tamaño de la población	(5.26, 5.41)
Categoría de la población	(5.68, 6.00)

Como se puede apreciar, el margen de variación en los pesos es amplio, especialmente en los criterios de infraestructura, accesibilidad a redes de transporte y grado de ocupación del polígono, en los que el rango de variación es superior a 0.3. En la Tabla 8 se muestran de nuevo los resultados obtenidos con los datos originarios y los obtenidos tras el análisis de robustez tras la aplicación del ELECTRE I en la que, por un lado, se han modificado los niveles de los umbrales de concordancia y discordancia y, por otro, se han variado los pesos de los criterios utilizando los rangos presentados en la Tabla 6 con el fin de facilitar los cálculos realizados.

Se podría concluir que el resultado es robusto, pues al modificar los valores de las ponderaciones, de las 31 comarcas analizadas, solo tres ven modificada su pertenencia al núcleo respecto al resultado inicial: la comarca del Moncayo pertenece al núcleo con las nuevas ponderaciones en el caso de aplicar los valores de los umbrales menos estrictos y la de la Ribera Baja del Ebro y la del Matarraña no pertenecen al núcleo con pesos iguales al mínimo del rango si se toman $s_c=0,8$ y $s_d= 0,2$ y $s_c=0,9$ y $s_d= 0,1$ respectivamente, siendo que en el resultado inicial y con pesos igual al máximo sí pertenecen al núcleo con esos valores de los umbrales.

Para poder comparar las soluciones obtenidas con estos dos métodos hay que indicar con absoluta claridad que los objetivos que persiguen son distintos: el método I se utiliza para seleccionar un conjunto de alternativas de entre el conjunto total que no se ven “superadas” por ninguna otra, mientras que la versión III persigue realizar una ordenación de las alternativas desde la mejor hasta la peor.

Tabla 8. Análisis de robustez, con variación de pesos, en el ELECTRE I.

Con pesos igual al mínimo del rango		Resultado inicial		Con pesos igual al máximo del rango	
$S_c=0,7$ $S_d=0,3$	Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Huesca, Jalón Medio, Moncayo Zaragoza	$S_c=0,7$ $S_d=0,3$	Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Huesca, Jalón Medio, Zaragoza	$S_c=0,7$ $S_d=0,3$	Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Huesca, Jalón Medio, Moncayo, Zaragoza
$S_c=0,8$ $S_d=0,2$	Aranda, Bajo Aragón, Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Cinco Villas, Huesca, Jalón Medio, Moncayo, Zaragoza.	$S_c=0,8$ $S_d=0,2$	Aranda, Bajo Aragón, Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Cinco Villas, Huesca, Jalón Medio, Moncayo, Ribera Baja del Ebro, Zaragoza.	$S_c=0,8$ $S_d=0,2$	Aranda, Bajo Aragón, Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Cinco Villas, Huesca, Jalón Medio, Moncayo, Ribera Baja del Ebro, Zaragoza.
$S_c=0,9$ $S_d=0,1$	Alto Gállego, Andorra, Aranda, Bajo Aragón, Bajo Cinca, Bajo Martín, Barbastro, Belchite, Borja, Calamocha, Calatayud, Cariñena, Caspe, Cinca Medio, Cinco Villas, Cuencas Mineras, Daroca, Gudar-Javalambre, Huesca, Jalón Medio, La litera, Moncayo, Monegros, Ribagorza, Ribera Alta del Ebro, Ribera Baja del Ebro, Teruel, Zaragoza.	$S_c=0,9$ $S_d=0,1$	Alto Gállego, Andorra, Aranda, Bajo Aragón, Bajo Cinca, Bajo Martín, Barbastro, Belchite, Borja, Calamocha, Calatayud, Cariñena, Caspe, Cinca Medio, Cinco Villas, Cuencas Mineras, Daroca, Gudar-Javalambre, Huesca, Jalón Medio, La litera, Matarraña, Moncayo, Monegros, Ribagorza, Ribera Alta del Ebro, Ribera Baja del Ebro, Teruel, Zaragoza.	$S_c=0,9$ $S_d=0,1$	Alto Gállego, Andorra, Aranda, Bajo Aragón, Bajo Cinca, Bajo Martín, Barbastro, Belchite, Borja, Calamocha, Calatayud, Cariñena, Caspe, Cinca Medio, Cinco Villas, Cuencas Mineras, Daroca, Gudar-Javalambre, Huesca, Jalón Medio, La litera, Matarraña, Moncayo, Monegros, Ribagorza, Ribera Alta del Ebro, Ribera Baja del Ebro, Teruel, Zaragoza.
$S_c=1$ $S_d=0$	Alto Gállego, Andorra, Aranda, Bajo Aragón, Bajo Cinca, Bajo Martín, Barbastro, Belchite, Borja, Calamocha, Calatayud, Cariñena, Caspe, Cinca Medio, Cinco Villas, Cuencas Mineras, Daroca, Gudar-Javalambre, Huesca, Jalón Medio, La litera, Matarraña, Moncayo, Monegros, Ribagorza, Ribera Alta del Ebro, Ribera Baja del Ebro, Teruel, Zaragoza.	$S_c=1$ $S_d=0$	Alto Gállego, Andorra, Aranda, Bajo Aragón, Bajo Cinca, Bajo Martín, Barbastro, Belchite, Borja, Calamocha, Calatayud, Cariñena, Caspe, Cinca Medio, Cinco Villas, Cuencas Mineras, Daroca, Gudar-Javalambre, Huesca, Jalón Medio, La litera, Matarraña, Moncayo, Monegros, Ribagorza, Ribera Alta del Ebro, Ribera Baja del Ebro, Teruel, Zaragoza.	$S_c=1$ $S_d=0$	Alto Gállego, Andorra, Aranda, Bajo Aragón, Bajo Cinca, Bajo Martín, Barbastro, Belchite, Borja, Calamocha, Calatayud, Cariñena, Caspe, Cinca Medio, Cinco Villas, Cuencas Mineras, Daroca, Gudar-Javalambre, Huesca, Jalón Medio, La litera, Matarraña, Moncayo, Monegros, Ribagorza, Ribera Alta del Ebro, Ribera Baja del Ebro, Teruel, Zaragoza.

Una vez asumido este principio, observando los resultados obtenidos, se puede apreciar que los dos métodos son consistentes. Con el nivel más flexible de los umbrales de concordancia y discordancia en el método I, las alternativas en el núcleo son: Barbastro, Belchite, Calatayud, Cariñena, Cinca Medio, Huesca, Jalón Medio y Zaragoza, que prácticamente coinciden con las 7 primeras posiciones en el método III (Barbastro aparece en novena posición y Belchite en la número 21 en el método ELECTRE III y Moncayo no aparece en el núcleo del método I para todos los valores de las ponderaciones). Además, si analizamos las últimas posiciones que nos da el método III (Sobrarbe y Albarracín), coinciden con las únicas alternativas que no aparecen en el núcleo del método I, para unos niveles de umbrales de concordancia y discordancia de 1 y 0, respectivamente.

4. CONCLUSIONES

La elección del lugar en el que una empresa se localiza, es una cuestión muy compleja. Este trabajo se ha realizado con un doble objetivo: ayudar en la toma de esa decisión cuando las alternativas de localización a las que se enfrenta la empresa son las comarcas de Aragón; y, como segundo objetivo, aportar evidencia de la coherencia en los resultados obtenidos con la utilización de dos metodologías de ayuda, que sirven para resolver problemas no similares.

Para la aplicación de estas metodologías se han utilizado criterios relevantes de localización y su importancia se ha obtenido de encuestas a empresas e instituciones y se han tomado como posibles alternativas de localización las comarcas aragonesas. Respecto a la importancia que los agentes han realizado sobre los factores de localización, si bien en los resultados obtenidos se ha encontrado que los factores de localización clásicos, como son los referidos a los costes (especialmente el coste del suelo industrial), continúan teniendo importancia, otros factores van adquiriendo importancia. Así factores como la accesibilidad a redes de transporte y la infraestructura del polígono tienen una ponderación muy alta.

Con las salvedades que son propias de cada uno de los métodos expuestos (la versión I sirve para extraer un subconjunto de alternativas del conjunto total, mientras que la versión III ordena todas las alternativas, desde la mejor hasta la peor), el trabajo también nos permite extraer la conclusión de que los resultados alcanzados son muy parecidos. Las alternativas que están en el núcleo de la versión I son las que aparecen en los primeros lugares de la ordenación que se obtiene con la versión III. Además, las alternativas que primero salen del núcleo en la versión I, son las que ocupan las últimas posiciones de la versión III. Por lo tanto, las conclusiones que se obtienen de las dos versiones son similares, y los resultados robustos.

Los resultados podrían utilizarse para orientar a las autoridades de la Comunidad Autónoma de Aragón en la toma de decisiones que afecten al fomento de la inversión en su territorio. Estos resultados han mostrado diferencias en el atractivo de las comarcas de la Comunidad Autónoma. El grupo formado por las comarcas de Cinca Medio, Jalón Medio, Huesca y el Moncayo aparecen en las dos versiones y con los criterios utilizados en el trabajo siempre en las primeras posiciones o en el núcleo si utilizamos la versión 1. Además, el análisis de robustez calculado, nos indica que esas posiciones siempre son las mismas. Del mismo modo, hay un grupo que siempre ocupa la posición final de la tabla o salen del núcleo: en concreto se trata de las comarcas de Sobrarbe, Albarracín y Matarraña. Tres de los resultados obtenidos merecen especial atención puesto que corroboran que los factores detectados como relevantes han influido en la inversión realizada en determinadas zonas:

En primer lugar, los resultados que ha obtenido el área metropolitana de Zaragoza, nos permiten afirmar que la situación estratégica de la ciudad le permite superar su mala valoración en precio de venta o equipamientos. Siempre aparece en el núcleo y en la posición 7ª de la clasificación según el ELECTRE III. La reflexión que se puede obtener de esto es que algunas empresas si que han podido modificar su ubicación atraídos por precios del suelo más bajos o mejores equipamientos.

En segundo lugar, hay que destacar la posición que han obtenido las comarcas de la provincia de Teruel. En el núcleo más flexible, obtenido tras la utilización del método ELECTRE I, no aparece ninguna de ellas, mientras que en los resultados proporcionados por la versión ELECTRE III, la primera que aparece lo hace en la undécima posición y seis de las ocho comarcas que hay en la provincia aparecen entre las doce últimas posiciones. La falta de infraestructuras de transporte en estas comarcas no es la única explicación para esta situación, ya que la comarca mejor valorada (Cinca Medio) no tiene ningún kilómetro de autovía o de autopista en su territorio. La diferencia viene por factores tales como la diferencia en la categoría de la población, en la dotación de los servicios públicos y, especialmente, en la dotación de infraestructuras de los polígonos.

Para concluir, la significativa variación de la comarca de Belchite podría explicarse por la valoración que tiene esta alternativa respecto a los criterios y al juego de los pesos, ya que en cinco de los factores considerados es la peor alternativa, pero el peso de estos criterios es más bajo, mientras que aparece bien valorada en criterios cuyo peso es elevado (precio del suelo, localización y equipamientos). Este hecho resulta importante en el método I, pero no tanto en el III. Hay que recordar que, en el método I, en el momento en que la valoración de

una alternativa respecto a un criterio es mayor que la de otra alternativa, aquélla supera a ésta, aunque la diferencia en la valoración sea muy pequeña. Por el contrario, en la versión III se deja un margen en esa superación para afirmar con rotundidad que una alternativa supera a otra. Ese margen lo determinan precisamente los umbrales de preferencia e indiferencia.

Agradecimiento

Los autores agradecen a Eduardo Barranco su ayuda en la fase de recogida de información.

REFERENCIAS BIBLIOGRÁFICAS

- Almeida, A. (2005) “Multicriteria Modelling of Repair Contract Based on Utility an ELECTRE I Method with Dependability and Service Quality Criteria”, *Annals of Operations Research*, Vol.138: 113–126.
- Arauzo, J.M. (2005) “Determinants of industrial locations: an application for catalan municipalities”. *Papers in Regional Science*, Vol.84,1: 1–144.
- Aydalot, P. (1985) *Economie Regional et Urbaine*. Economica, Paris.
- Barba-Romero, S.; Pomerol, J.C. (1997): *Decisiones multicriterio: fundamentos teóricos y utilización práctica*. Servicio de publicaciones de la Universidad de Alcalá, Madrid.
- Baaij, M.; Van den Bosch, F.; Volberda, H. (2004) “The international relocation of corporate centres: are corporate centres sticky?”. *European Management Journal*, Vol.22, n°2: 141–149.
- Barrios, S.; Strobl, E.; Görg, H. (2006) “Multinational’s location choice, agglomeration economies and public incentives”. *International Regional Science Review*, 29,1: 81–107.
- Crozet, M.; Mayer, T.; Mucchielli, J.L. (2003) “How do firms agglomerate? A study of FDI in France”. *Regional Science & Urban Economics*, Vol.34: 27–54.
- Cuadrado, J.R.; Aureoles, J. (1989) “La localización industrial en España”, *Estudios de la Fundación Fondo para la Investigación Económica y Social*, Madrid.
- Diop, O. (1988) “Contribution à l’étude de la gestion des déchets solides de Dakar: Analyse systématique et aide à la décision”, These de doctorat n° 757, DGE-EPFL, Lausanne.
- Fryxell, G.; Butler, J.; Choi, A. (2004) “Successful localization programs in China: an important element in strategy implementation”. *Journal of World Business*, Vol.39: 268–282.
- Guimaraes, P.; Figueiredo, O.; Woodward, D. (2000) “Agglomeration and Location of Foreign Direct Investment in Portugal”. *Journal of Urban Economics*, Vol.47:115–135.
- Heizer, J.; Render, B. (2001) *Dirección de la producción. Decisiones estratégicas*, Pearson Education, Madrid, 6ª edición.
- Hekman, J.S. (1992) “What are businesses looking for?” *Federal Reserve Bank of Atlanta Economic Review*, Vol.67(June): 6–19.
- Hoover, E.M. (1948) *The Location of Economic Activity*. McGraw-Hill, New York.

- Lösch, A. (1940) *Die Räumliche Ordnung der Wirtschaft* (traducción al inglés: “The Economics of Location” por Woglom, W. en 1954, Yale University Press, New Haven; traducción al español: “Teoría Económica Espacial”, en 1957, Ateneo, Buenos Aires).
- Lu, C.S. (2003) “Market segment evaluation and international distribution centers”. *Transportation Research Part E*, Vol.39: 49–60.
- Maystre, L.Y.; Pictet, J.; Simons, J. (1994) “Méthodes Multicritères ELECTRE”, *Collection Gérer l'Environnement*, Presses Polytechniques et Universitaires. Romandes, Lausanne.
- Morais, D.; Almeida, A. (2006) “Water supply system decision making using multicriteria analysis” *Water SA*, Vol.32: 229–236.
- Nowak, M. (2005) “INSDECM-an interactive procedure for stochastic multicriteria decision problems”, *European Journal of Operational Research* Vol.175: 1413–1430.
- Oum, T.; Park J. (2004) “Multinational firms’ location preference for regional distribution centres: focus on the Northeast Asian region”. *Transportation Research Part E*, Vol.40: 101–121.
- Porter, M.E. (1990) *The Competitive Advantage of Nations*. Free Press, New York.
- Palander, T. (1935) “Beiträge zur Standortstheorie”. Uppsala, Almqvist et Wiksells Boktryckeri.
- Ravelo, T.; González, G.; Moreno, M.C.; Sedeño, A. (2002) *La Localización Industrial en Canarias. Una aproximación Multicriterio*, Fundación FYDE-Caja Canarias.
- Rogers, M.; Bruen, M. (1998) “Choosing realistic values of indifference, preference and veto thresholds for use with environmental criteria within ELECTRE”. *European Journal of Operational Research*, Vol.107: 542–551.
- Roy, B. (1968) “Classement et Choix en présence de Points de Vue Multiples, la Méthode ELECTRE”, *R.I.R.O.*, Vol.2, n°8: 57–75.
- Smith, A. (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford University Press, Glasgow, 1976 (traducción al español: “La naturaleza y causas de la riqueza de las naciones”, por Collado, J.C; Mira-Perceval, A. Oikos Tau, Barcelona, 1988).
- Stonebraker, P.W.; Leong, G.K. (1994) *Operations strategy: focusing competitive excellence*, Allyn & Bacon, Boston.
- Ulgado, F.M. (1996) “Location characteristics of manufacturing investments in the United States: a comparison of American and foreign-based firms”. *Management International Review*, Vol.36.
- Vasiloglou, V. (2004) “New Tool for Landfill Location”. *Waste Management & Research*, Vol.22, n° 6: 427–439.
- Weber, A. (1909) *Über den Standort der Industrien* (traducción al inglés: “Theory of Location of Industries”, University of Chicago Press, Chicago, 1929).
- Woodward, D. (1992) “Location Determinants of Japanese Manufacturing Start-Up in the United States”, *Southern Economic Journal*, Vol.58: 690–708.

Evaluación de la eficiencia de una cartera de asegurados en el sector del automóvil

SEGOVIA GONZÁLEZ, M. MANUELA

Departamento de Economía, Métodos Cuantitativos e Historia Económica

Universidad Pablo de Olavide

Correo electrónico: mmseggon@upo.es

CONTRERAS RUBIO, IGNACIO

Departamento de Economía, Métodos Cuantitativos e Historia Económica

Universidad Pablo de Olavide

Correo electrónico: iconrub@upo.es

MAR MOLINERO, CECILIO

Kent Business School

University of Kent (United Kingdom)

Correo electrónico: C.Mar-Molinero@kent.ac.uk

RESUMEN

La situación actual del sector del seguro del automóvil exige a las compañías aseguradoras diseñar productos cada vez más ajustados al tipo de cliente al que se dirige. En el presente trabajo analizamos los 80.000 registros ofrecidos por una entidad aseguradora con una importante cuota en el mercado español. Tras identificar una serie de perfiles en los clientes asegurados, se estudia la eficiencia de cada uno de ellos mediante la aplicación de Análisis Envoltante de Datos (DEA) y estadística multivariante. Los resultados obtenidos permiten identificar aquellos clientes que mejor rendimiento ofrecen a la compañía. Mediante la aplicación de test estadísticos paramétricos y no paramétricos, se determinan, además, las similitudes y diferencias existentes entre grupos de clientes que comparten una determinada característica.

Palabras clave: eficiencia; seguro de automóvil; análisis envoltante de datos; estadística multivariante.

Clasificación JEL: C44; D81; G22.

2000MSC: 62H25; 62J05.

Efficiency Assessment in the Automobile Insurance Sector

ABSTRACT

Insurance companies have to take risk and cost into account when pricing car insurance policies in order to design new products to cover the risk of private use of cars. In this paper we use data from 80,000 car insurance policies in order to assess, once risk and cost have been taken into account, the policies that generate the highest returns for the company. After different profiles of clients are identified, including among the characteristics the regional location of the clients, we use Data Envelopment Analysis and multivariate statistics. The results obtained not only permit us to identify the best groups of clients for the company but also to determine similarities and differences among profiles of clients.

Keywords: efficiency; car insurance; data envelopment analysis; multivariate statistics.

JEL classification: C44; D81; G22.

2000MSC: 62H25; 62J05.



1. INTRODUCCIÓN

En los últimos años se han ido produciendo importantes cambios en el mercado asegurador del automóvil. En nueve o diez años, se ha pasado de una situación en la que este sector operaba con márgenes técnicos negativos, a una situación en la que se ha desatado una lucha intensa, por parte de las compañías, para mejorar el posicionamiento en el mercado e incrementar su cuota de negocio en este ramo. Este cambio tan brusco se puede atribuir principalmente a dos causas. En primer lugar, la publicación de los baremos de lesionados ha permitido aumentar la estabilidad y control técnico del ramo. En segundo lugar, la aparición de un nuevo canal de distribución, la venta por teléfono e internet. No obstante, el comportamiento del sector del seguro del automóvil es el de un sector totalmente maduro. Sus características principales son un bajo crecimiento, una excesiva oferta, poca innovación de productos, productos estandarizados con muy poca diferenciación (salvo en el precio) y una fuerte competencia.

Las razones expuestas anteriormente, unidas a esta fase de madurez que atraviesa el sector, ha originado que se estén practicando, por parte de algunas compañías, políticas muy agresivas. Dichas prácticas no están justificadas técnicamente y su único fin es adquirir cuota en el mercado, aun a costa de ofrecer prestaciones que suponen una nula rentabilidad para la compañía que efectúa dicha oferta. Todo esto no ha favorecido el control y buen desarrollo del negocio asegurador del automóvil. Esto ha hecho que la mayoría de las compañías hayan optado por mantener y elaborar productos novedosos, con amplias coberturas para el cliente y precios competitivos. Con esta estrategia pretenden lograr en todo momento una política comercial, que mantenga cotas altas de competitividad en sus productos y huir de las variaciones y grandes oscilaciones que se observan en el mercado en relación con las políticas de precio de algunas compañías de seguros.

En el ramo del seguro de automóvil operan un gran número de aseguradoras que ofrecen productos muy similares. Por ello, las normas de contratación de cada compañía es una de las actividades de mayor trascendencia en cuanto a los resultados de su negocio. Dichas normas deben buscar el equilibrio técnico-comercial, de tal forma que, por un lado, sirvan de filtro para los riesgos que mayor siniestralidad presentan a priori y, por otro lado, permitan mantener lo más abierto posible el mercado potencial asegurador. Existen estudios en los que se analiza el comportamiento del asegurado ante la incertidumbre que puede provocar la existencia de riesgos adicionales. Éstos pueden venir dados por la posibilidad de que el asegurador no cumpla con el pago de la indemnización acordada en caso de siniestro, o bien por otro tipo de riesgos secundarios que afectan a la riqueza final del individuo. Esta incertidumbre puede provocar en el cliente la elección de una mayor cobertura para protegerse así indirectamente. Los primeros que abordaron este tipo de cuestiones fueron Mayers y Smith (1983) y Doherty y Schlesinger (1983). Asimismo, otras cuestiones similares fueron tratadas por Doherty y Schlesinger (1990) y Mahul (2000), entre otros. Por otra parte, hay trabajos que muestran de forma empírica cómo

el asegurado cuando está satisfecho con el trato recibido y con el servicio que se le ha dado, posee una mayor predisposición a renovar su póliza con la actual aseguradora (Pujol y Bolancé, 2004). No obstante, cuando el asegurado decide contratar o renovar su póliza, el precio o prima del seguro es un factor determinante, aunque no es el único. Las tarifas del ramo de automóviles están estructuradas en función de una serie de variables y los riesgos de cada producto tienen sus particularidades en cuanto a la tarificación. En este sentido destacamos los trabajos de Boj *et al.* (2004) y Guillén *et al.* (2005), en los que podemos encontrar análisis estadístico multivariante aplicado a la selección de factores de riesgo en la tarificación y el cálculo actuarial. Las compañías de seguros, utilizando su propia experiencia y las estadísticas suministradas por la patronal de entidades aseguradoras y reaseguradoras española (UNESPA), son capaces de identificar, *a priori*, la posible siniestralidad de grupos homogéneos de asegurados. Para ello, tienen en cuenta las características que son más relevantes a la hora de tener una mayor o menor siniestralidad. Dichas categorías son: el grupo de tarificación del vehículo, la zona de circulación, el uso del vehículo, las circunstancias personales del conductor y las bonificaciones y penalizaciones. Siendo dichas características las que tendremos en cuenta para definir los diferentes tipos de clientes con los que realizaremos nuestro trabajo empírico. La información de que disponemos nos ha sido facilitada por una compañía del sector del automóvil que alcanza importantes cuotas en el mercado asegurador español.

En el presente trabajo se busca estudiar la eficiencia de las pólizas vendidas en el sector del automóvil para uso privado teniendo en cuenta las características que determinan el nivel de riesgo, y los ingresos que de ellas se obtienen, a través del Análisis Envolvente de Datos (DEA).

De manera sintética, el DEA es una técnica no paramétrica de análisis de eficiencia de unidades de producción que a partir de múltiples *inputs* es capaz de producir múltiples *outputs* utilizando una estructura productiva homogénea. Denotamos cada unidad como UOA (*unit of assessment*) en vez del término tradicional DMU (*decision making unit*), siguiendo la opinión de Sarrico *et al.* (1997), ya que un perfil de asegurado no toma ninguna decisión. En este trabajo no se usa el concepto de DMU en el sentido tradicional de una unidad de gestión que toma sus propias decisiones. Por ello, utilizamos las siglas UOA, que extienden el concepto a unidad en evaluación. Para una introducción al análisis DEA véase, por ejemplo, Thanassoulis (2001) o Cooper *et al.* (2000).

Un elemento determinante en los resultados obtenidos a partir del DEA es la selección de los *inputs* y *outputs* que serán incluidos en el modelo. Cada combinación de *inputs* y *outputs* que se considere dará lugar a unos coeficientes de eficiencia distintos y, consecuentemente, a una ordenación diferente de las unidades de producción. Es más, una UOA puede ser eficiente o no dependiendo de cuales sean las variables que se incluyan en el modelo. Para salvar este inconveniente, proponemos aplicar la metodología propuesta en Serrano-Cinca *et al.* (2005) y

Serrano-Cinca y Mar-Molinero (2004). En ambos trabajos, se propone la utilización conjunta del análisis DEA y del Análisis de Componentes Principales (PCA) con el objetivo de guiar al decisor en la selección del modelo DEA. Asimismo, la metodología aplicada permitirá identificar las fortalezas y debilidades de una UOA en función de la forma en la que dicha unidad logra su nivel de eficiencia. Cada tipo de cliente es tratado como una UOA y lo que se pretende analizar es la eficiencia de cada unidad desde el punto de vista del rendimiento obtenido por la empresa aseguradora.

Para tal fin, propondremos dos *inputs* y dos *outputs* con los que evaluar nuestras unidades de producción. Como *inputs* incluimos en el análisis aquellas variables que suponen un coste para la compañía y que sintetizamos a través del número de siniestros total y el coste medio derivado de ellos. Como *outputs* consideramos los ingresos que generan dichos clientes a la compañía y que medimos a través de los valores de prima bruta y prima neta. En todos los casos, consideramos los valores por cada 100 clientes, de manera que eliminamos el efecto que pueda tener el que en cada grupo de clientes el número de observaciones sea distinto. La metodología que aplicamos consistirá en estimar todos los posibles modelos, obtenidos como combinación de las variables propuestas, y visualizar los resultados por medio del PCA. Los resultados obtenidos del PCA nos permitirán relacionar los modelos y las UOA, en tanto identificará similitudes y discordancias entre los diferentes casos.

El resto del trabajo se estructura de la siguiente manera. En la Sección 2 se presentan los datos con los que se analizan los diferentes tipos de clientes considerados mediante la aplicación de DEA y estadística multivariante. En la Sección 3 discutimos los resultados obtenidos y se obtienen algunos resultados adicionales. La Sección 4 está dedicada a conclusiones. Por último, al final del trabajo se recogen las referencias bibliográficas utilizadas en este trabajo.

2. ANÁLISIS DE EFICIENCIA DE LA CARTERA DE CLIENTES

En la presente sección analizamos el comportamiento de diferentes tipos de clientes para la entidad aseguradora mediante la combinación de análisis DEA y técnicas de análisis multivariante.

2.1 Datos

En primer lugar merece destacarse que resulta complejo, dadas las restricciones a que obliga la legislación, obtener información acerca de la cartera de clientes de una compañía aseguradora. No obstante, una compañía española con una amplia implantación a nivel nacional nos ha cedido los datos correspondientes al ejercicio del año 2002, sin incluir aquellos valores que eliminarían su carácter anónimo.

Para realizar nuestro estudio nos restringiremos a los individuos asegurados en dicho período que conduzcan un vehículo de tipo turismo y cuyo uso sea particular. El motivo por el que realizamos estas restricciones es que pretendemos realizar un estudio lo mas homogéneo posible. Si considerásemos otro tipo de vehículos, como es el caso de camiones, motos o taxis, se podrían distorsionar los resultados. Piénsese, por ejemplo, como los profesionales que trabajan con su vehículo están más expuestos al riesgo.

Después de llevar a cabo las restricciones anteriores y realizar una exhaustiva depuración de los datos nos quedamos con más de 80.000 asegurados. Sobre estos valores, realizamos una estratificación de la muestra según el siguiente detalle: sexo, edad, zona geográfica y tipo de turismo.

- La variable sexo se justifica por los previsibles diferentes comportamientos entre los hombres y las mujeres a la hora de conducir.
- La edad del conductor actúa como un elemento importante a considerar en los resultados del estudio. Distinguimos entre los individuos menores o iguales a 30 años y los individuos mayores de 30 años.
- La zona geográfica, ya que la climatología, la densidad de población y las infraestructuras actúan como componentes importantes en los resultados obtenidos. Se considerarán cuatro zonas distintas: la zona Norte, Centro, Mediterránea y Sur.
- El tipo de turismo, distinguiendo entre gama baja, media y alta. Un coche de gran tamaño y potencia puede provocar, a iguales circunstancias de accidente, mayores daños que uno pequeño. De la misma manera, también serán mayores los gastos de reparación de los daños propios del vehículo.

Los técnicos de la propia compañía aseguradora nos indicaron que esta estratificación es la que habitualmente realizan la mayoría de las compañías. En particular, cada una de ellas, al lanzar al mercado un producto, consideran además otras características. Esto es debido a que cuanto más se ajuste a la realidad de los clientes mayor éxito tendrá el nuevo producto, en el sentido de que aportará mayores beneficios económicos para la compañía y de satisfacción para el cliente.

La distribución de los asegurados en cuanto al tipo de turismo se presentan con unos porcentajes del 12,48%, 40,38% y 47,14%, para los vehículos de gama alta, media y baja, respectivamente. En cuanto al sexo tenemos que la proporción de hombres es muy superior a la de mujeres encontrando un 82,73% de hombres frente a un 17,27% de mujeres. La zona Sur es la que presenta un mayor porcentaje de asegurados con un 52,22%, seguida de la zona Mediterránea con un 18,31%, la zona Norte con un 17,67% y por último, la zona Centro con un 11,80%. Si consideramos la edad del conductor, nos encontramos con un menor porcentaje de

conductores jóvenes un 8,33% de asegurados menores o iguales a 30 años frente a un 91,67% de asegurados mayores de 30 años. Esta última característica se presenta de manera frecuente en la mayoría de las compañías del sector debido a las numerosas dificultades y a las grandes tarifas que aplican a la hora de asegurar a los individuos con poca edad y experiencia.

En la Tabla 1 se muestran los distintos perfiles o tipos de clientes que se obtienen al considerar la estratificación anterior.

		<i>Alta (A)</i>		<i>Intermedia (I)</i>		<i>Baja (B)</i>	
		<i>Mujer (F)</i>	<i>Hombre (M)</i>	<i>Mujer (F)</i>	<i>Hombre (M)</i>	<i>Mujer (F)</i>	<i>Hombre (M)</i>
<i>Norte (N)</i>	>30 (+)	AFN+	AMN+	IFN+	IMN+	BFC+	BMN+
	≤30 (-)	-	AMN-	IFN-	IMN-	BFC-	BMN-
<i>Centro (C)</i>	>30 (+)	AFC+	AMC+	IFC+	IMC+	BFC+	BMC+
	≤30 (-)	-	AMC-	IFC-	IMC-	BFC-	BMC-
<i>Mediterránea (M)</i>	>30 (+)	AFM+	AMM+	IFM+	IMM+	BFM+	BMM+
	≤30 (-)	-	AMM-	IFM-	IMM-	BFM-	BMM-
<i>Sur (S)</i>	>30 (+)	AFS+	AMS+	IFS+	IMS+	BFS+	BMS+
	≤30 (-)	AFS-	AMS-	IFS-	IMS-	BFS-	BMS-

Tabla 1. Perfiles de clientes.

Es importante destacar que en lugar de trabajar con 48 perfiles, que sería el resultado de aplicar la estratificación anterior, lo haremos con 45. Esto es debido a que existen tres perfiles en los que no disponemos de suficiente tamaño muestral para que la representación de ellos sea significativa. Los perfiles no considerados en nuestro estudio son los correspondientes a las mujeres con coche de gama alta, menores de 30 años de las zonas Norte, Centro y Mediterránea y que en la tabla se denotan con “-“.

2.2 Análisis DEA

Nuestro objetivo principal será proporcionar a la compañía el comportamiento de cada uno de los perfiles seleccionados. Todo el estudio se desarrollará desde el punto de vista de la entidad aseguradora, pretendiendo comparar la eficiencia de cada uno de los tipos de cliente recogidos en la Tabla 1.

Para tal fin aplicamos DEA, ya que permite estudiar la eficiencia relativa de cada una de las unidades, esto es, una unidad será o no eficiente en función de los resultados obtenidos por el resto, realizando comparaciones múltiples a través de modelos basados en la programación lineal. Un elemento determinante del resultado que se obtendrá con la aplicación del DEA es la combinación de *inputs* y *outputs* que se considere finalmente en el modelo.

En todos los trabajos existentes en la literatura en los que se aplica el DEA al sector asegurador, el objetivo que se persigue es analizar el comportamiento de las diferentes

compañías aseguradoras dentro de la industria del seguro. Algunos trabajos realizados en este campo son, entre otros, Cummins y Rubio-Misas (1998), Fecher *et al.* (1993), Fukuyama (1997), Tone y Sahoo (2005) o Kao y Hwang (2008). Ninguno de los trabajos citados analiza el problema desde el punto de vista del cliente, en todos ellos se estudia la eficiencia de las diferentes aseguradoras. Aun así, en ellos se considera como *outputs* de los modelos el número de primas o los ingresos derivados de las mismas, que pueden considerarse, sin duda, un elemento determinante en este tipo de análisis.

En el presente trabajo definimos a la unidad evaluada como un tipo de riesgo para la compañía. La información relevante para la aseguradora sobre el mismo puede sintetizarse en dos grupos de variables: aquellas que suponen un coste, de manera directa o indirecta y que, en cualquier caso, valores bajos de estas variables identifican a los buenos riesgos. Por otro lado, han de tenerse en cuenta aquellos conceptos que suponen un ingreso para la compañía. En este segundo caso valores altos identifican a los mejores riesgos. Los primeros conceptos se asocian con los *inputs* del modelo mientras que los segundos se identifican con los *outputs*.

Para los primeros hemos considerado dos variables, una estimación de la probabilidad del siniestro y el coste medio de los mismos. Para los segundos consideramos las primas ingresadas a partir de cada tipo de cliente. De manera sintética, las variables consideradas son las siguientes:

Input A: número de siniestros por cada 100 asegurados.

Input B: coste medio de los siniestros por cada 100 asegurados (euros).

Output 1: ingresos por prima bruta por cada 100 asegurados (euros).

Output 2: ingresos por prima neta por cada 100 asegurados (euros).

Los ingresos de prima neta, *output 2*, se obtienen como ingresos de prima bruta menos las bonificaciones asignadas por la compañía. De esta forma tendremos en cuenta en nuestro modelo el comportamiento de los asegurados en años anteriores a través de las bonificaciones asignadas a cada cliente. Esta variable captura de alguna forma la calidad y fidelidad de los clientes en tanto las compañías ofrecen este tipo de bonificaciones con el objetivo de retener a sus mejores asegurados. Algunos trabajos relacionando las primas con la experiencia pasada de los individuos en cuanto a la siniestralidad son, entre otros, Lemaire (1985), Henriet y Rochet (1986) y Dionne y Vanasse (1992).

De manera resumida, en la Tabla 2 se muestra la información de *inputs* y *outputs* recogidos en los modelos. En todos los casos consideramos los valores por cada 100 asegurados para evitar el efecto que pueda tener el diferente número de asegurados en cada categoría.

	<i>Inputs</i>		<i>Outputs</i>	
	A	B	1	2
<i>Máx</i>	76,92	51.009,10	142.989,31	99.631,13
<i>Mín</i>	21,28	1.125,38	48.812,08	30.016,87
<i>Media</i>	37,49	17.774,16	88.362,13	58.729,66
<i>Desv. Std.</i>	9,70	8.743,68	26.381,49	18.732,99

Tabla 2. Valores de *inputs* y *outputs*.

Para la evaluación de las unidades, hemos aplicado el modelo con rendimientos a escala constante y orientación al output, modelo CCR-O. La orientación al output se justifica porque la empresa aseguradora no puede controlar ni el riesgo del siniestro ni su costo medio, aunque sí que puede controlar los ingresos que los clientes generan a través de la política de tarifas.

La manera tradicional de evaluar las unidades consiste en resolver el modelo que incluye simultáneamente las cuatro variables, *inputs* A y B y *outputs* 1 y 2. La metodología que aplicaremos propone resolver todos los modelos que surgen de combinar las cuatro variables descritas. Con dos *inputs* y dos *outputs*, pueden obtenerse 9 modelos diferentes. La notación utilizada para los modelos es la que aparece en la Tabla 3.

<i>Modelo</i>	<i>Inputs</i>	<i>Outputs</i>
A1	A	1
AB1	A,B	1
B1	B	1
A2	A	2
AB2	A,B	2
B2	B	2
A12	A	1,2
AB12	A,B	1,2
B12	B	1,2

Tabla 3. Modelos DEA.

La razón principal para resolver todos los posibles modelos es doble. Por un lado, como se dijo, la eficiencia de una unidad dependerá de la combinación de *inputs* y *outputs* que se considere, por lo que tiene sentido evaluar a cada UOA en función de todos los posibles modelos que puedan construirse. Por otro, estamos interesados en estudiar las fortalezas o debilidades de cada unidad con respecto a cada variable y esta información puede obtenerse estudiando los coeficientes correspondientes a los distintos modelos. En la Tabla 4 se representan los valores de eficiencia de las 45 unidades de producción en los 9 modelos considerados. Este valor aparece expresado en porcentaje.

Un primer análisis de los resultados obtenidos permite comprobar cómo la eficiencia obtenida por una UOA particular depende de cómo se especifique el modelo, tal y como puede verse en la Tabla 4.

Considerando el modelo general, modelo AB12, solo dos unidades son eficientes: IFN- y AFS-. Todas coinciden con ser mujeres, con coches de gama media o alta y jóvenes (menores de 30). Estos dos grupos de clientes son los que mejor comportamiento tendrían para la compañía si consideramos todas las variables explicativas. Podemos ver cómo este buen comportamiento se debe, no obstante, a distintos factores en cada caso.

Para AFS-, el coeficiente de eficiencia de 100 en el modelo general contrasta con menores puntuaciones obtenidas en los modelos en los que el único *input* considerado es el A, número de siniestros por cada 100 asegurados. Esto indica que el buen resultado obtenido se debe al comportamiento de este tipo de clientes en las restantes variables o, lo que es lo mismo, a pesar de mostrar unos malos registros en cuanto al número de accidentes, el buen comportamiento en el *input* restante, así como en la cuantía de las primas, son suficientes para compensarlo y hacer que la unidad sea considerada eficiente.

En cuanto a la unidad IFN-, se puede ver como todos sus coeficientes son 100 excepto en aquellos modelos en los que el único *input* es el coste de los siniestros (B1, B2 y B12). En este caso, la eficiencia de la unidad se explica por un buen comportamiento en el *input* A, así como por una buena combinación en las variables que determinan la eficiencia global.

En la siguiente sección se realiza un análisis multivariante sobre los coeficientes DEA en el que se identifican las principales características de los valores de la Tabla 4. En dicha tabla nos encontramos con un gran volumen de información acerca de las unidades de producción eficientes, aunque también nos podemos encontrar con información redundante ya que algunos modelos DEA pueden ser equivalentes (Serrano-Cinca y Mar-Molinero, 2004). Para lograr todo lo anterior, se aplicarán técnicas de análisis multivariante, tales como, el análisis factorial y análisis de regresión múltiple.

2.3 Análisis factorial

El análisis factorial es una técnica de reducción de datos estándar que elimina la información redundante, resalta la información más significativa del conjunto de datos y ayuda a visualizar las principales relaciones existentes entre cada una de las observaciones (Dunteman, 1989).

En los valores recogidos en la Tabla 4, los diferentes modelos DEA pueden considerarse como variables y los perfiles definidos anteriormente como observaciones de nuestro estudio, de forma que disponemos de 45 observaciones y 9 variables. Puede ocurrir que con un conjunto más pequeño de datos podamos explicar las relaciones existentes entre cada uno de los perfiles considerados. Para tal fin, aplicamos la técnica del análisis factorial.

<i>UOA</i>	<i>A1</i>	<i>AB1</i>	<i>B1</i>	<i>A2</i>	<i>AB2</i>	<i>B2</i>	<i>A12</i>	<i>AB12</i>	<i>B12</i>
AFN+	60	60	31	54	54	25	60	60	31
AFC+	34	37	24	33	35	21	34	37	24
AFM+	43	43	15	42	42	13	43	43	15
AFS+	34	34	17	34	34	15	34	34	17
AFS-	67	100	100	75	100	100	75	100	100
AMN+	58	58	24	51	51	19	58	58	24
AMN-	86	86	39	85	85	35	86	86	39
AMC+	44	44	25	41	41	21	44	44	25
AMC-	34	34	10	36	36	9	36	36	10
AMM+	47	47	20	44	44	17	47	47	20
AMM-	47	60	45	51	61	44	51	61	45
AMS+	40	41	25	39	40	22	40	41	25
AMS-	53	53	25	57	57	24	57	57	25
IFN+	48	48	24	42	42	19	48	48	24
IFN-	100	100	59	100	100	53	100	100	59
IFC+	36	45	34	32	38	28	36	45	34
IFC-	61	63	39	61	64	35	61	64	39
IFM+	38	38	16	35	35	13	38	38	16
IFM-	51	51	21	52	52	19	52	52	21
IFS+	27	31	21	26	28	18	27	31	21
IFS-	39	48	35	42	49	34	42	49	35
IMN+	51	51	25	42	42	18	51	51	25
IMN-	55	55	25	55	55	22	55	55	25
IMC+	40	46	31	36	40	24	40	46	31
IMC-	35	35	15	37	37	14	37	37	15
IMM+	44	44	24	38	38	19	44	44	24
IMM-	34	34	16	36	36	15	36	36	16
IMS+	27	31	21	25	27	17	27	31	21
IMS-	31	31	17	34	34	17	34	34	17
BFC+	39	43	28	34	36	21	39	43	28
BFC-	51	63	46	49	58	40	51	63	46
BFC+	29	38	30	26	32	24	29	38	30
BFC-	47	59	44	51	60	43	51	60	44
BFM+	37	41	26	33	35	21	37	41	26
BFM-	35	38	24	37	39	23	37	39	24
BFS+	24	29	20	22	25	17	24	29	20
BFS-	25	33	26	28	33	25	28	33	26
BMN+	48	48	26	40	40	19	48	48	26
BMN-	40	40	23	40	40	20	40	40	23
BMC+	38	44	29	33	37	23	38	44	29
BMC-	28	37	30	28	35	27	28	37	30
BMM+	40	42	26	35	36	20	40	42	26
BMM-	25	26	16	28	29	16	28	29	16
BMS+	27	31	21	24	27	17	27	31	21
BMS-	26	29	19	28	31	18	28	31	19

Tabla 4. Coeficientes de eficiencia DEA (%).

Para estudiar el número de factores que se deben incluir en el modelo, se empezó por realizar un análisis de componentes principales. Los valores propios obtenidos aparecen en la Tabla 5. Es importante apuntar que el mínimo valor para la extracción de autovalores ha sido tomado como 0,8, siguiendo la recomendación de Joliffe (1972). Se obtienen dos autovalores que son mayores al límite expuesto. La primera componente aporta, con gran diferencia, la mayor parte de información, explicando un 86,03% de la variabilidad total. Considerando las

dos primeras componentes se explica un 99,17%, por lo que muestran una adecuada representación de los datos. Por esta razón, el modelo factorial contiene tan solo dos factores.

<i>Componente</i>	<i>Autovalor</i>	<i>% de varianza</i>	<i>Acumulado %</i>
PC1	7,74	86,03	86,03
PC2	1,18	13,14	99,17

Tabla 5. Resultados del análisis de componentes principales.

En la Tabla 6 exponemos las correlaciones existentes entre cada uno de los modelos y los dos factores. Los resultados se muestran ordenados respecto a las correlaciones con el primer factor. Podemos observar que el primer factor está fuertemente correlacionado y de forma positiva con todos los modelos. Dicho factor representa una “medida global de eficiencia”, tomando el máximo valor cuando consideramos el modelo completo AB12, seguido de los modelos con dos *inputs* (AB1, AB2). Para interpretar el segundo factor tenemos que tener en cuenta tanto el valor de las correlaciones como el signo de las mismas. Encontramos que dicho factor está correlacionado de forma positiva con los modelos B1, B2 y B12 y de forma negativa con los modelos A1, A2 y A12. La correlación positiva ocurre con aquellos modelos cuyo único *input* es B y la correlación negativa con aquellos modelos cuyo único *input* es A. Puede concluirse que la segunda componente discrimina entre modelos cuya eficiencia se concentra en el coste del siniestro y aquellos que la derivan del número de siniestros. Dicho factor puede denotarse como “coste del siniestro *versus* número de siniestros”.

	<i>Factor</i>	
	<i>F1</i>	<i>F2</i>
AB12	0,9938	-0,0900
AB1	0,9923	-0,0825
AB2	0,9872	-0,0872
A2	0,9343	-0,3335
A12	0,9293	-0,3651
A1	0,9016	-0,4205
B1	0,8712	0,4875
B12	0,8712	0,4875
B2	0,8536	0,5131

Tabla 6. Análisis factorial, matriz de correlaciones.

Las puntuaciones de cada perfil respecto a cada factor ofrecen una medida de eficiencia de cada uno de los conceptos anteriores, eficiencia global y eficiencia coste frente al número de siniestros. La Figura 1 muestra la representación de las puntuaciones en los dos factores (F1 y F2). Los perfiles con mayores puntuaciones en el primer factor son aquellos con mayor eficiencia global, destacando los perfiles AFS- y IFN-. Por otro lado en el extremo izquierdo tenemos a IFS+ y AFC+, que es un tipo de cliente que muestra una baja eficiencia en todos los modelos considerados. Es importante destacar que la eficiencia se incrementa según nos

desplazamos de izquierda a derecha en la Figura 1, lo que confirma la interpretación del primer factor como una medida de eficiencia global.

El segundo factor se identifica como el valor que distingue entre la eficiencia debida al coste y al número de los siniestros. En la parte superior derecha de la Figura 1 destaca AFS- con un coeficiente de eficiencia igual a 100 salvo en los modelos que contienen únicamente el *input* A. Por el contrario, en la parte inferior derecha podemos ver los perfiles AMN- y IFN- que ven reducida su eficiencia cuando el único *input* es B.

Pueden estudiarse las relaciones entre cada modelo y el valor obtenido en las puntuaciones factoriales. Para ello, se ha realizado un análisis de regresión en el que como variable dependiente hemos tomado la eficiencia de cada uno de los modelos y como variables independientes las puntuaciones de los dos primeros factores (Schiffman *et al.*, 1981). De esta forma, podremos constatar que representar los distintos modelos de eficiencia en términos de sus puntuaciones factoriales resulta adecuado y estaremos reduciendo la dimensionalidad de nuestro problema. Asimismo, podremos obtener modelos similares de eficiencia, puesto que el ángulo entre cualesquiera dos vectores está relacionado con la correlación existente entre la eficiencia generada por los dos modelos en cuestión. En la Tabla 7 mostramos el coeficiente de determinación ajustado y los coeficientes de regresión junto con los estadísticos de contraste para ver si son significativos o no. En nuestro caso todos los valores de R^2 son altos, todos son superiores a 0,98.

	<i>Directional cosines</i>		<i>F</i>	<i>Aj R²</i>
	$\varphi 1$	$\varphi 2$		
A1	0,902	-0,420	1996,172	0,989
	57,263**	-26,707**		
AB1	0,992	-0,083	2454,588	0,991
	69,825**	-5,806**		
B1	0,871	0,487	6338,496	0,997
	98,257**	54,978**		
A2	0,934	-0,334	1304,407	0,983
	48,104**	-17,172**		
AB2	0,987	-0,087	1162,574	0,981
	48,033**	-4,243**		
B2	0,854	0,513	2536,086	0,991
	61,041**	36,691**		
A12	0,929	-0,365	6854,379	0,997
	108,978**	-42,808**		
AB12	0,871	0,487	6338,496	0,997
	98,257**	54,978**		
B12	0,994	-0,090	4915,243	0,996
	98,745**	-8,940**		

**Significativo al nivel 0,01.

Tabla 7. Análisis ProFit. Resultados de la regresión lineal.

En la Figura 1 representamos igualmente los vectores normalizados obtenidos al realizar la regresión. Pueden diferenciarse tres grupos de modelos. En la gráfica, los modelos se representan con vectores normalizados y los perfiles de clientes mediante las etiquetas correspondientes.

Un primer grupo de modelos es el formado por A1, A2 y A12, en los que el único *input* es A y aproxima la eficiencia con respecto al número de siniestros. Un segundo grupo formado por B1, B2 y B12, en el que el único *input* es B, permite ver la eficiencia con respecto al coste medio del siniestro. Por último, un tercer grupo formado por los restantes modelos mide la eficiencia global de la unidad. De cada uno de estos grupos se ha seleccionado un modelo para indicar la dirección de mejora de cada una de las eficiencias mencionadas anteriormente, respectivamente, A12, B12 y AB12.

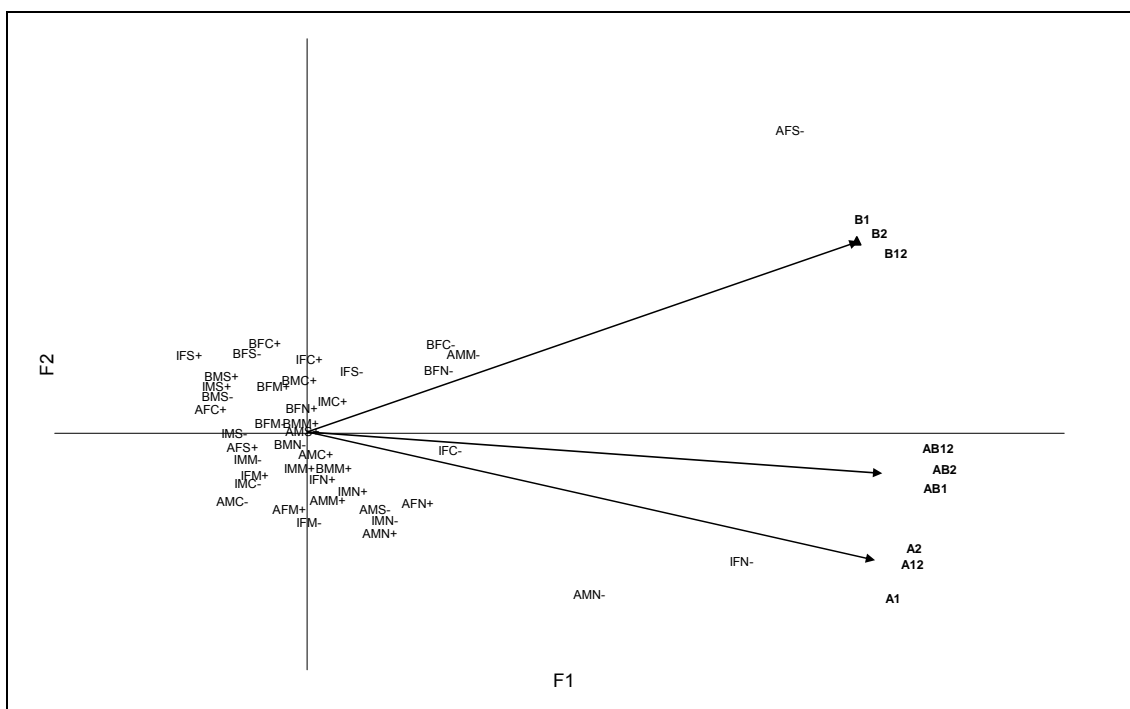


Figura 1. Gráfico de cargas factoriales, F1 vs. F2.

3. DISCUSIÓN DE LOS RESULTADOS

A partir de los valores presentados en la Sección 2 pueden obtenerse algunas conclusiones sobre el comportamiento de los perfiles de clientes considerados.

En las Figuras 2 y 3 se representan, respectivamente, los valores de las 45 unidades con respecto a los factores F1 y F2, diferenciando los asegurados por sexo y edad. Puede apreciarse como los clientes femeninos (Figura 2) y menores de 30 (Figura 3) tienen un mejor comportamiento para la compañía, esto es, obtienen mayores valores de eficiencia.

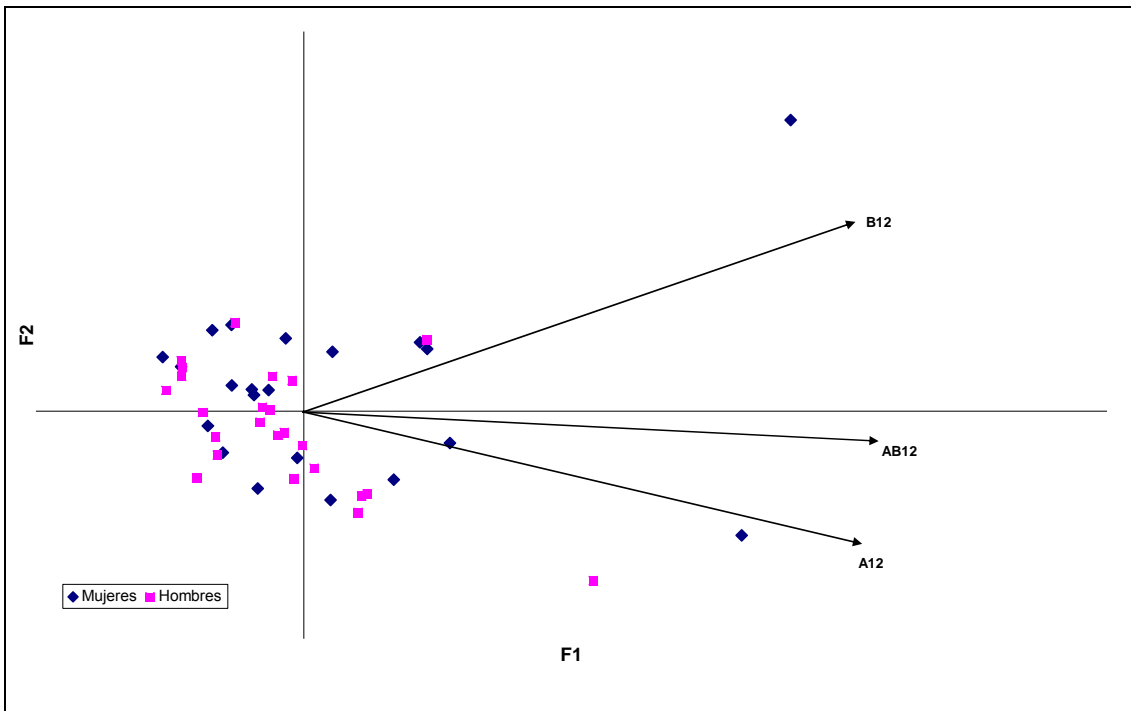


Figura 2. F1 vs. F2. Los perfiles se identifican por el sexo de los clientes.

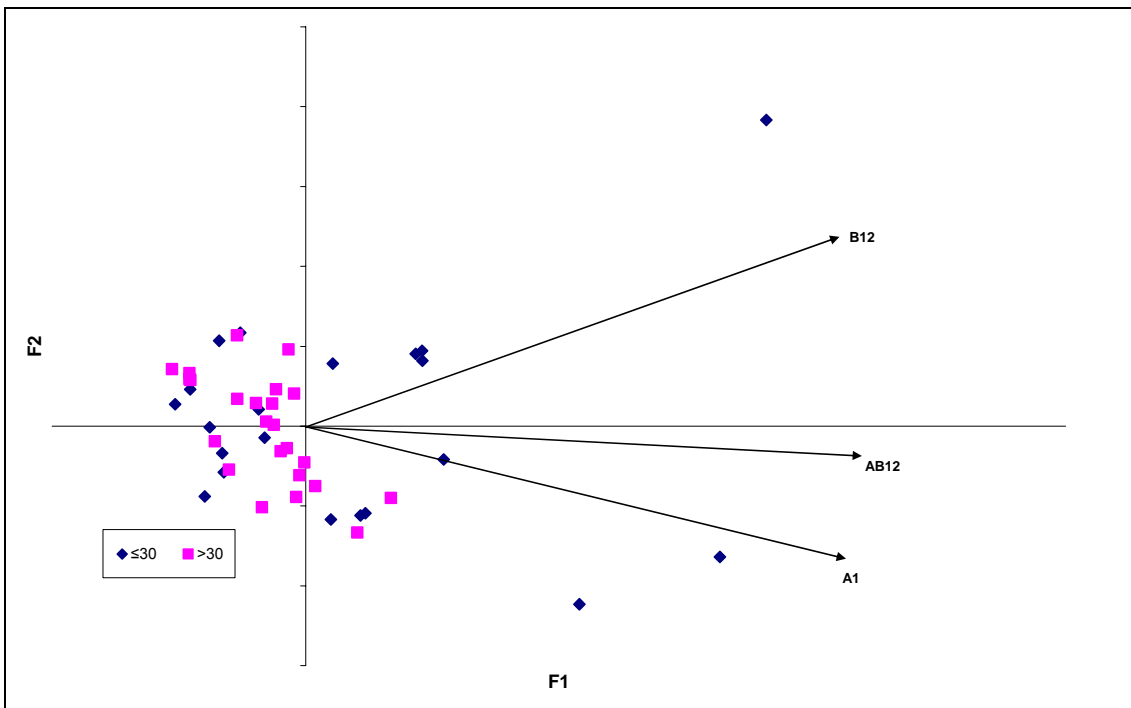


Figura 3. F1 vs. F2. Los perfiles se diferencian por edad.

Con respecto a las zonas de circulación de los vehículos, se comprueba que existen diferencias significativas tanto si se considera una medida de eficiencia global como si se considera la eficiencia respecto al número de siniestros. Estas diferencias, que se aprecian en la Figura 4, se corroboran a partir de los resultados obtenidos con el test de Kruskal-Wallis (p-valor 0,006 y 0,003, respectivamente). Con el test de Dunn de comparaciones múltiples, comprobamos que existen diferencias significativas con ambas medidas de eficiencia entre la zona Norte y las zonas Sur, Mediterránea y Centro, siendo más eficientes los conductores de la zona Norte.

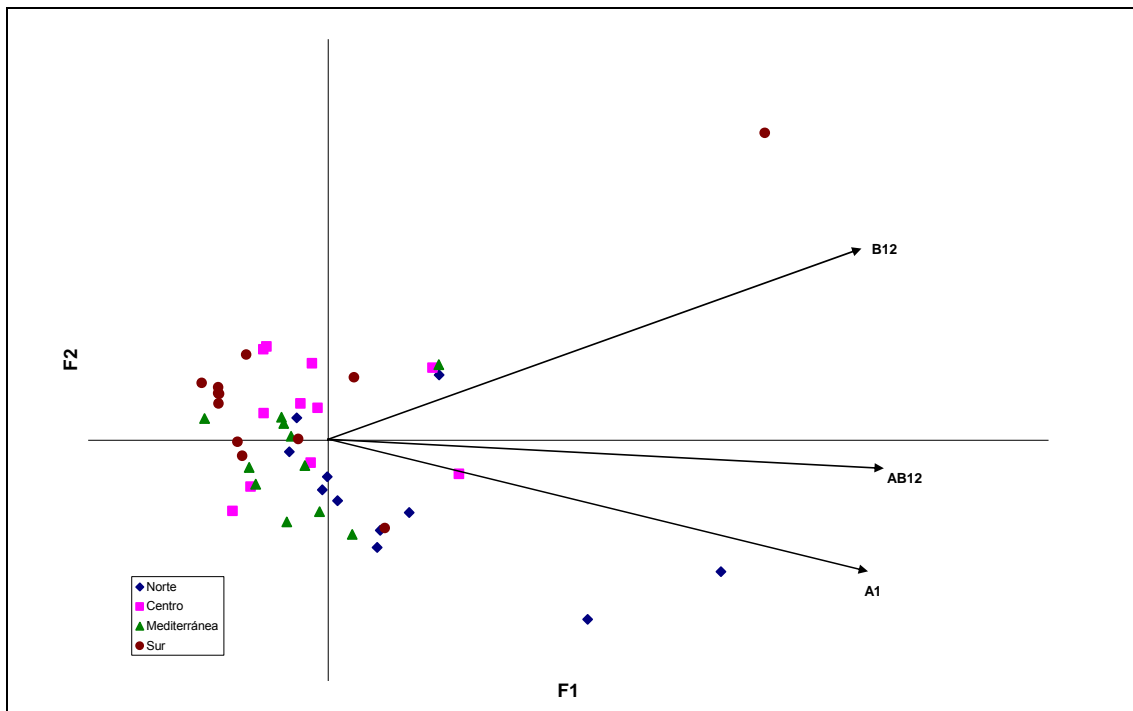


Figura 4. F1 vs. F2. Los perfiles se identifican por la zona geográfica.

En cuanto a la gama del vehículo, también se detectan diferencias en el modelo A12 a través del test de Kruskal-Wallis (p-valor 0,017). Los resultados del test de Dunn de comparaciones múltiples muestran que existen diferencias significativas entre los coches de gama alta y baja, siendo superior en los primeros. La eficiencia en cuanto al número de siniestros es mayor en los coches de gama alta.

Junto a los anteriores estudios globales para detectar diferencias entre los diferentes tipos de clientes, es posible realizar estudios parciales. Para ello, consideramos una variable y segmentamos la información con respecto a cualquier otra. Los resultados obtenidos, así como el test utilizado en cada caso, se recogen en la Tabla 8. En cada caso, pueden considerarse las tres medidas de eficiencia utilizadas hasta ahora, eficiencia global, eficiencia respecto al número de siniestros y eficiencia respecto al coste y que representamos, respectivamente, por los modelos AB12, A12 y B12.

<i>Variable</i>	<i>Segmentación</i>	<i>Modelo</i>	<i>Test (p-valor)</i>	<i>Más eficiente / Menos eficiente</i>
Sexo				
	Edad ≤30	B12	U (0.009)	Mann-Whitney Mujeres/Hombres
Edad				
	Mujeres	AB12	U (0.009)	Mann-Whitney ≤30 / >30
		A12	U (0.018)	Mann-Whitney ≤30 / >30
		B12	U (0.006)	Mann-Whitney ≤30 / >30
Área Geográfica				
	Hombres	AB12	Kruskal-Wallis (0,048)	Norte/Sur
		A12	Kruskal-Wallis (0,030)	Norte/Sur
	Gama intermedia	A12	Kruskal-Wallis (0,046)	Norte/Sur
	Gama baja	AB12	Kruskal-Wallis (0,034)	Centro/Medit., Norte/Sur
		A12	Kruskal-Wallis (0,019)	Norte/Sur, Centro/Sur
		B12	Kruskal-Wallis (0,029)	Centro/Sur, Centro/Medit.
	Edad >30	AB12	ANOVA (0,0005)	Norte/Sur, Norte/Centro, Norte/Medit., Centro/Sur, Medit./Sur
		A12	ANOVA (0,0005)	Norte/Centro, Norte/Sur, Medit./Sur
		B12	ANOVA (0,003)	Centro/Medit., Centro/Sur
Gama				
	Hombres	AB12	Kruskal-Wallis (0,049)	Alta/Baja
		A12	Kruskal-Wallis (0,028)	Alta/Baja
	Edad ≤30	A12	Kruskal-Wallis (0,032)	Alta/Baja
	Zona Sur	AB12	Kruskal-Wallis (0,030)	Alta/Baja
		A12	Kruskal-Wallis (0,037)	Alta/Baja

Tabla 8. Resultados de los análisis parciales.

En los siguientes análisis trabajaremos a un nivel de confianza del 95%. En primer lugar, comparamos la eficiencia entre hombres y mujeres. Para ello, segmentamos la información con respecto a las variables edad, zona geográfica y gama del coche. Si consideramos por separado los dos grupos de edad, obtenemos que en el grupo de menores de 30 años existen diferencias significativas en cuanto al comportamiento de hombres y mujeres si lo evaluamos por medio del modelo B12, siendo las mujeres más eficientes para la compañía. Este resultado se obtiene a partir del test no paramétrico de la U de Mann-Whitney, en tanto no se verifican condiciones de normalidad, obteniendo un p-valor igual a 0,009. Para los individuos mayores de 30 años no se han observado diferencias significativas con respecto al sexo, al igual que si tomamos las distintas zonas de circulación o las distintas gamas del vehículo.

En segundo lugar, comparamos la eficiencia con respecto a la edad. Al segmentar por la variable sexo, observamos que en cuanto a los hombres no existen diferencias significativas. Sí aparecen diferencias en el grupo de mujeres, siendo más eficientes las menores de 30 años.

En tercer lugar, comparamos la eficiencia con respecto a la zona geográfica en la que reside el conductor. Segmentando por la variable sexo, obtenemos que en los hombres existen diferencias significativas en cuanto a la zona de circulación, tanto en la eficiencia global como

en cuanto al número de siniestros (Kruskall-Wallis, 0,048 y 0,030, respectivamente), entre las zonas Norte y Sur. Aparecen diferencias de igual forma cuando segmentamos por la gama del vehículo. En los vehículos de gama baja aparecen diferencias significativas en los tres modelos de eficiencia. En cuanto a los coches de gama intermedia, las diferencias las encontramos entre las zonas Norte y Sur según el modelo A12. Segmentando por la variable edad, confirmamos que existen diferencias significativas para el grupo de conductores mayores de 30 años en los tres casos considerados.

En cuarto lugar compararemos la eficiencia con respecto a los distintos tipos de vehículos. Si segmentamos por la variable sexo o zona geográfica, encontramos que en los hombres y en la zona Sur existen diferencias significativas en cuanto a la eficiencia debida a los modelos AB12 y A12. En cuanto a la edad, las diferencias se encuentran respecto del modelo A12, para el grupo de conductores menores de 30 años. En todos los casos, la eficiencia de la gama alta es superior a la gama baja.

4. CONCLUSIONES

En el presente trabajo hemos estudiado el comportamiento de los diferentes tipos de asegurados que integran la cartera de clientes de una compañía aseguradora. A partir de los datos proporcionados por una aseguradora española con una elevada cuota de mercado, hemos analizado el comportamiento de cada tipo de cliente en términos de eficiencia. Se ha llevado a cabo aplicando conjuntamente el análisis envolvente de datos y técnicas de estadística multivariante.

Uno de los principales determinantes del análisis DEA es la elección de los *inputs* y *outputs* que forman el modelo. La elección de una u otra variable determina el resultado que se obtiene del análisis. Para salvar dicha dificultad se han aplicado todos los posibles modelos que se tienen con los *inputs* y *outputs* seleccionados. La gran cantidad de información obtenida se estudia y sintetiza con técnicas de estadísticas multivariante como el análisis factorial y análisis de regresión.

El primer resultado que se obtiene es la existencia de tres indicadores de eficiencia, “eficiencia global”, “eficiencia en el número” y “eficiencia en el coste de los siniestros”. En todos los casos, la medida de eficiencia hace referencia al resultado obtenido por la compañía para ese tipo de cliente.

En términos generales, se identifican los comportamientos más interesantes para la empresa en asegurados jóvenes, mujeres, ubicados en la zona Norte y que conducen vehículos de gama alta. Además de las comparaciones directas entre las variables, se han obtenido resultados parciales segmentando la información según las diferentes variables consideradas en

el estudio, lo que permite hacer comparaciones entre cada grupo de asegurados considerando por separado las tres medidas de eficiencia.

Los resultados obtenidos del análisis permitirían a la empresa aseguradora optimizar la composición de su cartera de clientes. Cada unidad puede considerarse como un tipo de riesgo para la compañía, el trabajo identifica aquellos más interesantes, clientes con altos valores de eficiencia, en los que los valores de las primas son suficientemente altos como para compensar los valores de siniestralidad y coste derivado de los mismos. De esta manera, la aseguradora puede optar por concentrar sus esfuerzos en atraer a este tipo de clientes incluso ajustando el valor de las primas.

De igual modo, se identifican a los clientes menos eficientes para la compañía. En este caso, el valor de las primas que se ingresa en comparación con la siniestralidad y coste derivado no resulta rentable para la compañía en términos relativos, esto es, la rentabilidad que obtiene la empresa aseguradora es menor que la que obtiene de clientes con mayores coeficientes de eficiencia. En este caso, la empresa puede reconsiderar el papel de este tipo de clientes o ajustar al alza sus primas para igualar su rendimiento al de clientes más eficientes.

BIBLIOGRAFÍA

- BOJ, E., CLARAMUNT, M.M. Y FORTIANA, J. (2004): *Análisis multivariante aplicado a la selección de factores de riesgo en la tarificación*. Cuadernos de la Fundación MAPRE estudios. Instituto de Ciencias del Seguro. Madrid.
- COOPER, W.W., SEIFORD, L.M. Y TONE K. (2000): *Data Envelopment Analysis: a comprehensive text with models, applications, references*. Kluwer Academic Publisher, Boston.
- CUMMINS, J.D. Y RUBIO-MISAS, M. (1998): "M. Efficiency in the Spanish insurance industry". *Working paper, The Wharton School, University of Pennsylvania*.
- DIONNE, G. Y VANASSE, C. (1992): "Automobile insurance ratemaking in the presence of asymmetrical information". *Journal of Applied Econometrics*, 7 (2); pp.149–166.
- DOHERTY, N.A. Y SCHLESINGER, H. (1983): "Optimal insurance in incomplete markets". *Journal of Political Economy*, 91 (6); pp.1045–1054.
- DOHERTY, N.A. Y SCHLESINGER, H. (1990): "Rational insurance purchasing: consideration of contract non-performance". *Quarterly Journal of Economics*, 105; pp.143–153.
- DUNTEMAN, G.H. (1989): "Principal components analysis". *Series: Quantitative Applications in the Social Sciences*, nº 69. Series Editor: Michael S. Lewis-Beck, University of Towa.
- FECHER, F., KESSLER D., PERELMAN S. Y PESTIEAU P. (1993): "Productive performance of the French insurance industry". *Journal of Productivity Analysis*, 4; pp.77–93.

- FUKUYAMA, H. (1997): "Investigating productive efficiency and productivity changes of Japanese life insurance companies". *Pacific-Basin Finance Journal*, 5; pp.481–509.
- GUILLEN, M., AYUSO, M., BERMÚDEZ, L. Y MORILLO, I. (2005): *El Seguro de automóviles: estado actual y perspectiva de la técnica actuarial*. Fundación MAPFRE estudios. Instituto de Ciencias del Seguro. Madrid.
- HENRIET, D. Y ROCHET, J.C. (1986): "La logique des systèmes Bonus-Malus en assurance automobile: une approche théorique". *Annales d'Économie et de Statistique*, 1; pp.133–152.
- JOLIFFE, I.T. (1972): "Discarding variables in Principal Components Analysis". *Applied Statistics*, 21; pp.160–173.
- KAO, C. Y HWANG, S.N. (2008). "Efficiency decomposition in two-stage data envelopment analysis: an application to non-life insurance companies in Taiwan". *European Journal of Operational Research*, 185(1); pp.418–429.
- LEMAIRE, J. (1985): *Bonus-Malus systems in automobile insurance actuarial models*. Kluwer Academic Publishers, Boston.
- MAYERS, D. Y SMITH, C.W. (1983): "The interdependence of individual portfolio decisions and the demand for insurance". *Journal of Political Economy*, 91; pp.304–311.
- MAHUL, O. (2000): "Optimal insurance design with random initial wealth". *Economics Letters*, 69; pp.353–358.
- PUJOL, M. Y BOLANÉ, C. (2004): *La matriz valor fidelidad en el análisis de los asegurados en el ramo del automóvil*. Fundación MAPFRE estudios. Instituto de Ciencias del Seguro. Madrid.
- SARRICO, C.S., HOGAN, S.M., DYSON, R.G. Y ATHANASSOPOULOS, A.D. (1997): "Data envelopment analysis and university selection". *Journal of the Operational Research Society*, 48; pp.1163–1177.
- SCHIFFMAN, J.F., REYNOLDS, M.L. Y YOUNG, F.W. (1981): *Introduction to Multidimensional Scaling: Theory, Methods and Applications*. Academic Press: London.
- SERRANO-CINCA, C., FUERTES CALLEN, Y. Y MAR-MOLINERO, C. (2005): "Measuring DEA efficiency in Internet companies". *Decision Support Systems*, 38; pp.557–573.
- SERRANO-CINCA, C. Y MAR-MOLINERO, C. (2004): "Selecting DEA specifications and ranking units via PCA". *Journal of the Operational Research Society*, 55; pp.521–528.
- THANASSOULIS, E. (2001): *Introduction to the theory and application of data envelopment analysis: A foundation text with integrated software*, Kluwer Academic Publishers, Boston.
- TONE, K. Y SAHOO, B.K. (2005): "Evaluating cost efficiency and returns to scale in the Life Insurance Corporation of India using data envelopment analysis". *Socio-Economic Planning Sciences*, 39; pp.261–285.

Bifurcaciones de codimensión 2 en un modelo dinámico del mercado potencial y actual: aplicación al mercado cervecero español

VELASCO MORENTE, FRANCISCO
Departamento de Economía Aplicada I
Universidad de Sevilla
Correo electrónico: velasco@us.es

NADAL MORALES, PURIFICACIÓN
Departamento de Economía Aplicada I
Universidad de Sevilla
Correo electrónico: pnadal@us.es

GONZÁLEZ ABRIL, LUIS
Departamento de Economía Aplicada I
Universidad de Sevilla
Correo electrónico: luisgon@us.es

ORTEGA RAMÍREZ, JUAN ANTONIO
Departamento de Lenguajes y Sistemas Informáticos
Universidad de Sevilla
Correo electrónico: ortega@lsi.us.es

RESUMEN

En este trabajo, a partir de la continuación de las bifurcaciones de codimensión 1 en un modelo dinámico discreto aplicado al mercado actual y potencial de las organizaciones, se establecen las condiciones necesarias para la existencia de bifurcaciones de codimensión 2. Como caso práctico, se estudia el caso del mercado cervecero español, considerando como referencia la marca “San Miguel”, analizando su evolución e interpretación económica.

Palabras clave: bifurcaciones de codimensión 2; mercado cervecero español.
Clasificación JEL: C61; C62; M39.
2000MSC: 37C; 37G; 37N40.

Bifurcations of Codimension 2 in a Dynamical Model of Current and Potential Market: An Application to the Spanish Beer Market

ABSTRACT

In this work we establish the necessary conditions for the existence of bifurcations of codimension 2 in a discrete dynamic model applied to the current and potential market of the organizations, starting from the continuation of the codimension 1 bifurcations. The case of the market Spanish brewer is studied and we analyse its evolution and interpret the economical results obtained.

Keywords: codimension 2 bifurcations; Spanish brewer market.

JEL classification: C61; C62; M39.

2000MSC: 37C; 37G; 37N40.



1. INTRODUCCIÓN

En Landa y Velasco (2004) y Velasco *et al.* (2007), se exponen cómo las organizaciones mercantiles pueden mantener sus cuotas de mercado en equilibrio durante determinados periodos de tiempo. Esto no significa que la solución de equilibrio tenga que durar de manera permanente, ya que la estrategia de cada organización competitiva, en un mercado de referencia y para un producto cualquiera de ese mercado, ha de ir encaminada a conseguir la cuota de mercado más elevada posible; pero no es menos cierto que cualquier comportamiento competitivo encaminado a conseguir ese objetivo será respondido por las organizaciones competidoras que defenderán sus respectivas posiciones de mercado. Un entorno agresivo, un esfuerzo de marketing mal orientado u otras causas, hacen que la posición conseguida por cada organización competitiva no permanezca, necesariamente, en equilibrio estable.

Podemos ver en Gandolfo (1997), Lorenz (1997), Vílchez *et al.* (2002), Velasco *et al.* (2002), Vílchez *et al.* (2003), Landa y Velasco (2004), Vílchez *et al.* (2004), Bosi *et al.* (2005), Haunsmied *et al.* (2005), He XZ y Westerhoff FH (2005), Li MQ (2005), Wagener (2005), Wei-Bing-Zhang (2006), Magnitskii y Siderov (2006) y Chang *et al.* (2007), entre otros, ejemplos económicos en los que se estudian bifurcaciones en modelos económicos. En el trabajo de Velasco *et al.* (2007) se hace un estudio de la estabilidad del sistema y de las bifurcaciones de codimensión 1 en un modelo económico discreto y se analiza bajo qué condiciones, si se realizan pequeñas variaciones en los parámetros del modelo, éstas van a implicar que el comportamiento del mismo sea cualitativamente diferente en un entorno del punto fijo. En particular, se estudió la estabilidad del modelo referido al mercado cervecero español y se encontraron las bifurcaciones de codimensión 1 asociadas a cada uno de los valores de los parámetros del modelo.

En este artículo estudiamos las bifurcaciones de codimensión 2, que vienen caracterizadas por dos condiciones independientes que involucran simultáneamente dos parámetros. El interés de esta búsqueda radica en que la aparición de estas bifurcaciones lleva asociada un comportamiento para ambos mercados mucho más complejo que el descrito hasta el momento por el espacio paramétrico de cada bifurcación de codimensión 1. Para ello, hemos de partir de los puntos en los que existen bifurcaciones de codimensión 1, realizar la continuación de dichos puntos obligando a que cumplan las condiciones de no degeneración de la correspondiente bifurcación y comprobar para qué valores de los parámetros no se cumple alguna de las condiciones impuestas. Se ha realizado un estudio exhaustivo de todas las posibles combinaciones entre dos parámetros a fin de obtener todas y cada una de las bifurcaciones de codimensión 2 en el sistema dinámico discreto. Los posibles comportamientos de la dinámica de un sistema en un entorno de una bifurcación de codimensión 2 se comentan en el Anexo.

El resto del artículo está estructurado como sigue: en la segunda sección se comenta brevemente el modelo dinámico discreto propuesto en Velasco *et al.* (2007). En la tercera sección se exponen los conceptos necesarios que se van a utilizar a lo largo del trabajo. En la cuarta sección, realizamos la continuación de los puntos de bifurcación de codimensión 1 para los valores de los parámetros relevantes y se obtienen las principales bifurcaciones de codimensión 2 en el modelo presentado. Se finaliza con una sección de conclusiones y posibles trabajos futuros.

2. EL MODELO

Si se representan por $x_t = x(t)$ e $y_t = y(t)$ al número de compradores potenciales y al número de clientes actuales de una marca determinada en el instante t , respectivamente, el incremento del número actual de clientes, $\Delta y(t) = y(t+1) - y(t)$, viene dado por aquellas personas que son atraídas al producto, debido al efecto que produce el contacto de las personas que componen el mercado actual con aquellas que constituyen el mercado potencial, es decir, $a(t) x(t) y(t)$, donde $a(t)$ es la llamada *tasa de contacto global*. Feitchinger (1992) supone que esta tasa de contacto global es proporcional al número de clientes actuales del período de tiempo correspondiente, es decir, $a(t) = a y(t)$. Además, a ese poder de atracción habrá que sustraerle el número de clientes actuales que dejan el mercado de la organización para pasar al mercado potencial, $b y(t)$, y los que dejan de ser clientes actuales por muerte o emigración, $e y(t)$; ambos se consideran proporcionales al número de clientes actuales, donde b es la proporción en la que los clientes actuales cambian a un marca rival, pasando por tanto a engrosar el número de clientes potenciales y e es la tasa por desaparición de los clientes en general. Consideremos, también el esfuerzo de publicidad, c , que realiza la organización, es decir, la publicidad tratará de atraer hacia el mercado actual elementos que pertenecen al mercado potencial, donde hemos de hacer notar que una mala actuación en publicidad puede hacer que este valor tenga efectos negativos, con lo que no podemos despreciar el hecho de que sea negativo. De forma análoga el incremento del número de clientes potenciales $x(t)$ tiene la misma estructura pero teniendo en cuenta que los signos (+ y -) cambian su sentido. Se ha de tener en cuenta también que en este caso hemos de añadir una tasa de entrada, que viene dada por el número de personas que se incorporan al mercado potencial del producto y que denotamos por k , debido a un incremento² en su poder adquisitivo. Nuestro sistema dinámico lo podemos expresar por tanto como sigue:

$$\begin{cases} x_{t+1} = x_t + k - ax_t y_t^2 + by_t - cx_t \\ y_{t+1} = y_t + ax_t y_t^2 - by_t - ey_t + cx_t \end{cases}, t \in \mathbb{R}, a, b, k, e, c \geq 0 \quad (1)$$

En Velasco *et al.* (2007) se explican con más detalle las ecuaciones de este sistema dinámico, así como la posibilidad de que en algunos casos los valores de los parámetros c , k y b sean negativos y se exponen, además, tanto las condiciones de estabilidad del modelo como las condiciones para la obtención de las bifurcaciones de codimensión 1. A lo largo del artículo quedará constancia de los signos de los parámetros.

3. BIFURCACIONES DE CODIMENSIÓN 2 DE LOS PUNTOS FIJOS DE UN SISTEMA DINÁMICO DISCRETO

Sea un sistema dinámico discreto bi-paramétrico (Wiggins (1990), Guckenheimer (1993), Kuznetsov (1998)):

$$x \rightarrow f(x, \alpha) \quad (2)$$

donde $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}^2$ y f es suficientemente suave. Supongamos que para el valor del parámetro $\alpha = \alpha^0$, el sistema (1) tiene un punto fijo o de equilibrio en $x = x^0$, para el que se satisface alguna condición de bifurcación Fold, Flip o Neimark-Sacker. Entonces

² Si suponemos una marca de alto poder adquisitivo, hablaremos de incremento. Ahora bien, si hemos bajado en nuestro poder adquisitivo, hablamos de decremento.

la ecuación de punto fijo, $f(x, \alpha) - x = \theta$, donde θ representa el vector nulo, y la correspondiente condición de bifurcación $\psi(x, \alpha) = 0$, que se le impone a los multiplicadores de la matriz Jacobiana $J = f_x$ evaluada en (x, α) , define una curva Γ en el espacio \mathbb{R}^{n+2} . Cada punto $(x^0, \alpha^0) \in \Gamma$ corresponde a un punto de bifurcación x^0 del sistema (2) que satisface la condición de bifurcación relevante para el valor del parámetro α^0 . La curva de bifurcación B sobre el plano paramétrico es la proyección estándar $\pi : (x, \alpha) \rightarrow \alpha$, que proyecta Γ sobre la curva de bifurcación $B = \pi(\Gamma)$ en el plano paramétrico.

Si el sistema (2) tiene en $\alpha^0 = (\alpha_1^0, \alpha_2^0)$ un equilibrio $x = x^0$, con un multiplicador $\mu = 1$, entonces el sistema de ecuaciones

$$\begin{cases} f(x, \alpha) - x = \theta, x \in \mathbb{R}^n \\ \det(f_x(x, \alpha) - I) = 0 \end{cases}$$

define genéricamente³ una variedad unidimensional suave $\Gamma \subset \mathbb{R}^{n+2}$ que pasa por el punto $(x^0, \alpha_1^0, \alpha_2^0)$. Cada punto $(x, \alpha) \in \Gamma$ define un punto fijo x del sistema (2) con un multiplicador $\mu = 1$ para el valor del parámetro α . La proyección estándar traslada Γ sobre la curva $B = \pi(\Gamma)$ en el plano paramétrico y es sobre esta curva donde se cumple la bifurcación Fold.

Análogamente, la bifurcación Flip viene definida genéricamente por el sistema:

$$\begin{cases} f(x, \alpha) - x = \theta, x \in \mathbb{R}^n \\ \det(f_x(x, \alpha) + I) = 0 \end{cases}$$

con un multiplicador $\mu = -1$.

Para el caso de la bifurcación de Neimark-Sacker, hemos de suponer un sistema dinámico plano $n = 2$, tal como:

$$x \rightarrow f(x, \alpha), \quad x = (x_1, x_2) \in \mathbb{R}^2, \quad \alpha = (\alpha_1, \alpha_2) \in \mathbb{R}^2$$

donde suponemos que en $\alpha^0 = (\alpha_1^0, \alpha_2^0)$ hay un punto fijo $x^0 = (x_1^0, x_2^0)$ con un par de multiplicadores no reales sobre el círculo unidad: $\mu_{1,2} = e^{\pm i \theta_0}$, $0 < \theta_0 < \pi$. Se considera entonces el sistema de tres ecuaciones escalares en \mathbb{R}^4 con coordenadas $(x_1, x_2, \alpha_1, \alpha_2)$ siguiente:

$$\begin{cases} f(x, \alpha) - x = \theta, x \in \mathbb{R}^2 \\ \det(f_x(x, \alpha)) - 1 = 0 \end{cases} \quad (3)$$

La bifurcación de Neimark-Sacker viene definida genéricamente por el sistema (3) en el punto de equilibrio x^0 en el valor paramétrico α^0 .

Si en la segunda ecuación de (3) tenemos dos puntos multiplicadores reales $\mu_1 = \tau$, $\mu_2 = \frac{1}{\tau}$, con $\tau > 1$, entonces al punto de equilibrio se le denomina silla neutral.

En general, si variamos los parámetros (α_1, α_2) simultáneamente, entonces se pueden obtener bifurcaciones de codimensión 2 sobre la curva Γ o sobre B , ya que podemos tener los siguientes casos: i) pueden aproximarse otros multiplicadores al círculo unidad haciendo variar la dimensión de la variedad centro, ii) pueden dejar de cumplirse las condiciones de no-degeneración para la bifurcación de codimensión 1.

³ El término genérico indica que el rango de la matriz jacobiana es máximo, es decir, $n+1$.

Si se considera la curva de bifurcación Fold B_T , entonces un punto cualquiera sobre esta curva corresponde a un punto fijo con un multiplicador simple $\mu = 1$ y ningún otro multiplicador sobre el círculo unidad. La restricción del sistema dinámico (2) sobre la variedad centro es unidimensional y tiene la forma $\xi \rightarrow \xi + a \xi^2 + O(\xi^3)$. Además, en un punto de bifurcación Fold no degenerado se tiene que $a \neq 0$. De forma similar, si se considera la curva de bifurcación Flip B_F , el multiplicador es $\mu = -1$ y la restricción del sistema dinámico (2) sobre la variedad centro en un punto Flip no degenerado tiene la forma $\xi \rightarrow -\xi + b \xi^3 + O(\xi^4)$, donde $b \neq 0$. Ahora bien, si seguimos una curva de bifurcación Neimark-Sacker B_{NS} , entonces un punto cualquiera sobre dicha curva es un punto fijo con un par simple de multiplicadores no reales sobre el círculo unidad $\mu_{1,2} = e^{\pm i \theta_0}$, siendo éstos los únicos multiplicadores sobre el círculo unidad. En este caso la variedad centro bidimensional y el sistema (2) sobre dicha variedad centro viene dada en notación compleja como $z \rightarrow z e^{i \theta_0} (1 + d_1 |z|^2) + O(|z|^4)$, donde $d_1 \in \mathbb{R}$. En este caso, las condiciones de no-degeneración son de dos tipos: i) ausencias de resonancias fuertes, es decir $e^{i q \theta_0} \neq 1$, $q = 1, 2, 3, 4$, ii) no-degeneración cúbica, es decir $c = \text{Re}(d_1) \neq 0$.

En general, los puntos de degeneración que se pueden obtener en los sistemas genéricos en tiempo discreto con dos parámetros, al moverse en las curvas de codimensión 1, son once (Kuznetsov, 1998)⁴:

1) $\mu_1 = 1$, $a = 0$ (bifurcación cúspide), con forma normal $\eta \rightarrow \eta + \beta_1 + \beta_2 \eta + s \eta^3 + O(\eta^4)$ y $s = \text{sign}(f_{xxx}(0,0)) = \pm 1$, con condiciones de no-degeneración:

- i) $f_{xxx}(0,0) \neq 0$;
- ii) $(f_{\alpha_1} f_{x\alpha_2} - f_{\alpha_2} f_{x\alpha_1})(0,0) \neq 0$.

2) $\mu_1 = -1$, $b = 0$ (bifurcación Flip generalizada), con forma normal

$\eta \rightarrow -(1 + \beta_1)\eta + \beta_2 \eta^3 + s \eta^5 + O(\eta^6)$, $s = \text{sign}(D_1(0)) = \pm 1$, con condiciones de no degeneración:

i) $D_1(0) = \left(\frac{1}{15} f_{x^5} + \frac{1}{2} f_x f_{x^4} - [f_x]^4 \right) (0,0) \neq 0$;

ii) La aplicación $\alpha \rightarrow (\mu(\alpha) + 1, B(\alpha))$ es regular en $\alpha = 0$, donde $\mu(0) = 1$ y

$B(0) = \frac{1}{4} [f_{xx}(0,0)]^2 + \frac{1}{6} f_{xxx}(0,0) \neq 0$.

3) $\mu_{1,2} = e^{\pm i \theta_0}$, $c = 0$ (bifurcación de Chenciner);

$$\begin{cases} \rho \rightarrow \rho (1 + \beta_1 + \beta_2 \rho^2 + L_2(\beta) \rho^4) + \rho^6 R(\rho, \beta), \\ \varphi \rightarrow \varphi + \theta(\beta) + \rho^2 Q(\rho, \beta) \end{cases} \quad (\text{Forma normal})$$

Condiciones de no-degeneración:

- i) $e^{i q \theta_0} \neq 1$, $q = 1, 2, \dots, 6$;

⁴ Al realizar los diversos cambios de sistema de referencia para cada bifurcación, surgen unos parámetros β_i , que son funciones de los parámetros del modelo y que caracterizan la bifurcación en particular. En el caso que nos ocupa los casos 9), 10) y 11) no son válidos ya que nuestro modelo es bidimensional.

ii) la aplicación $\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \rightarrow \begin{pmatrix} r(\alpha) - 1 \\ \text{Re}(d_1(\alpha)) \end{pmatrix}$ es regular en $\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$;

iii) $L_2(0) = \text{Re}(d_1(0)) \neq 0$.

4) $\mu_1 = \mu_2 = 1$ (Resonancia 1:1);

5) $\mu_1 = \mu_2 = -1$ (Resonancia 1:2);

6) $\mu_{1,2} = e^{\pm i \theta_0}$, $\theta_0 = \frac{2\pi}{3}$ (Resonancia 1:3);

7) $\mu_{1,2} = e^{\pm i \theta_0}$, $\theta_0 = \frac{\pi}{2}$ (Resonancia 1:4);

8) $\mu_1 = 1$, $\mu_2 = -1$;

9) $\mu_1 = 1$, $\mu_{2,3} = e^{\pm i \theta_0}$;

10) $\mu_1 = -1$, $\mu_{2,3} = e^{\pm i \theta_0}$;

11) $\mu_{1,2} = e^{\pm i \theta_0}$, $\mu_{3,4} = e^{\pm i \theta_1}$.

Las bifurcaciones obtenidas están caracterizadas por dos condiciones independientes (bifurcación de codimensión 2) y no hay otras bifurcaciones de codimensión 2 en los sistemas dinámicos discretos genéricos (Kuznetsov, 1998).

4. ESTUDIO DE BIFURCACIONES LOCALES DE CODIMENSIÓN 2 EN EL MERCADO CERVECERO ESPAÑOL

Al ir recorriendo alguna de las curvas de bifurcación de codimensión 1, obtenida al proyectar sobre el plano paramétrico, el sistema formado por la ecuación del punto de equilibrio y las condiciones de bifurcación impuestas sobre los multiplicadores de la matriz Jacobiana, se pueden obtener valores de los parámetros del sistema que den lugar, en un entorno de los mismos, a nuevos comportamientos del modelo en estudio, bien porque exista un número mayor de multiplicadores que se aproximen al círculo unidad, o bien porque se dejen de cumplir algunas de las condiciones de no degeneración, que nos indica que ocurre una bifurcación de codimensión 2.

Basándonos en ello, nos planteamos completar el análisis realizado en Velasco *et al.* (2007) del comportamiento de los mercados actuales y potenciales del mercado cervecero español, tomando como referencia la marca San Miguel, estudiando la aparición de bifurcaciones de codimensión 2 en este sistema. El interés de esta búsqueda radica en que la aparición de estas bifurcaciones lleva asociada un comportamiento para ambos mercados mucho más complejo que el descrito por las bifurcaciones de codimensión 1 (ver Anexo).

De Velasco *et al.* (2007) se sigue que los multiplicadores de la matriz jacobiana del sistema (1) particularizada en el punto de equilibrio vienen dados por:

$$\mu_1 = \frac{-A + \sqrt{(A+2)^2 - 4B}}{2}, \quad \mu_2 = \frac{-A - \sqrt{(A+2)^2 - 4B}}{2}$$

donde $A = b + c + e + \frac{ak^2}{e^2} - \frac{2a(b+e)k^2}{ce^2 + ak^2} - 2$, $B = ce + \frac{ak^2}{e}$, y las regiones de estabilidad

e inestabilidad puede observarse en la Figura 1. Además, se obtuvo que $x^* = 12,766677$, $y^* = 8,327809$ es un punto de equilibrio⁵; asimismo este punto fijo es un foco estable, ya que verifica la desigualdad $4B > (A+2)^2$, $B < A+2$, pues los

⁵ Los valores de población vienen expresados en tantos de millón.

valores correspondientes de A y B , dentro del mercado cervecero español, que vienen dados en Velasco *et al.* (2007) son: $A = -1.991$, $B = 0.0001$. Los valores de los parámetros vienen dados en la tabla que exponemos a continuación:

a	b	c	k	e
0.0034	0.0001	0.0001	0.0790	0.3616

Varias son las ideas que nos hacen pensar que el mercado cervecero español puede tender a alcanzar alguno de esos tipos de bifurcaciones de codimensión 2:

1. Si observamos el plano paramétrico (Figura 1), vemos que la situación del punto (A, B) con los valores obtenidos está muy próxima al punto de corte de la recta $A = -2$ con la parábola $(A + 2)^2 - 4B = 0$, que es la condición necesaria para la existencia de una *resonancia fuerte 1:1*; es decir, los dos multiplicadores tienen el valor unitario.
2. Los multiplicadores de la matriz jacobiana en el punto crítico tienen su parte real muy próxima al valor uno y su parte imaginaria muy cercana al valor cero. Sabemos que una de las condiciones para la existencia de una *resonancia fuerte 1:1* es que la matriz jacobiana en el punto crítico tenga un multiplicador real igual a uno con multiplicidad dos.

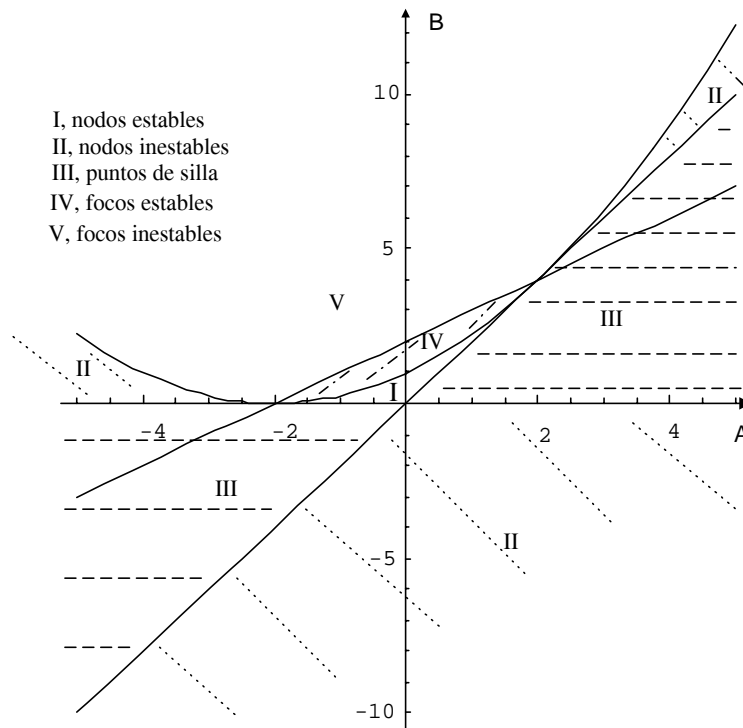


Figura 1. Representación gráfica de las regiones de estabilidad e inestabilidad del sistema dinámico (1).

3. El coeficiente $a(0)$ de la forma normal asociada que encontramos en los puntos críticos de algunas de las bifurcaciones de Neimark-Sacker halladas, tiene un valor muy próximo a cero⁶; por lo cual, pequeñas variaciones en los valores de los parámetros pueden tener como resultado la anulación de dicho coeficiente. Sabemos que una de las condiciones de no-degeneración para este tipo de bifurcación es que dicho coeficiente no sea nulo.

⁶ Ver Velasco *et al.* (2007), concretamente las páginas 436, 438, 440, 442 y 444.

4.1. Influencia conjunta del esfuerzo de publicidad (c) y el paso a una marca rival (b)

Comenzamos nuestro análisis partiendo del punto de bifurcación de Neimark-Sacker que encontramos en el mercado cervecero español para el valor crítico del esfuerzo de publicidad⁷ $c=0.019101486$ y realizamos la continuación de dicho punto, utilizando para ello el software CONTENT, donde realizamos la simulación numérica activando dos parámetros conjuntamente (el esfuerzo de publicidad y cada uno de los restantes) y obtenemos en el punto crítico: los valores de los multiplicadores, los valores de los parámetros y los valores de ambos mercados. De esta forma conseguimos que, al proyectar sobre el plano paramétrico, obtengamos la gráfica de continuación para los puntos de Neimark-Sacker⁸.

Al variar conjuntamente ambos parámetros, encontramos dos puntos de bifurcación cuyos datos recogemos en la tabla siguiente:

Tipo bif.	Valor c	Valor b	Cl. Pot.	Cl. Act.
DN (Chenciner)	-0.015689665	-0.021536786	52078520	83278461
R1	-0.060613800	-0.078177474	$1.237359 \cdot 10^{-6}$	88278461

y cuya gráfica de continuación puede apreciarse en la Figura 2. A continuación, vamos a realizar el análisis de los datos obtenidos, tanto para estas bifurcaciones como para las que obtenemos posteriormente, bajo dos enfoques: el teórico-analítico y el económico-práctico.

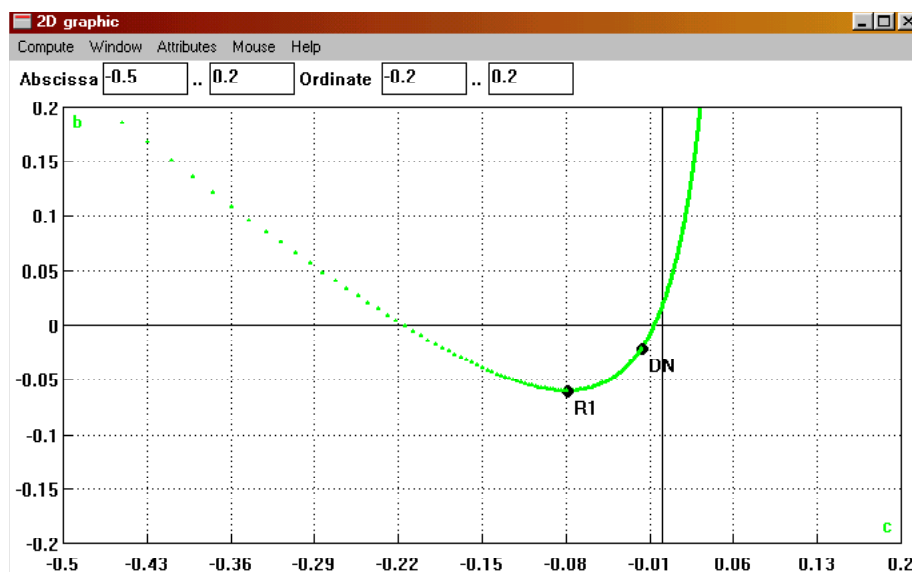


Figura 2. Bifurcaciones de codimensión 2 con el esfuerzo de publicidad y el paso a una marca rival.

* Respecto a la bifurcación de Chenciner (DN) encontrada, debemos realizar las siguientes consideraciones:

- Esta bifurcación se presenta para los valores de los parámetros $c = -0.015689665$ y $b = -0.021536786$ que, hasta este momento, no habíamos tenido en consideración, ya

⁷ Ver página 437 en Velasco *et al.* (2007) para los detalles de la obtención del punto de bifurcación N-S.

⁸ Al ir construyendo la continuación de la curva de Neimark-Sacker, la simulación gráfica se para en cada punto de bifurcación de codimensión 2 que encuentra, dándole unas iniciales que corresponden al tipo de bifurcación. Así, tenemos que:

- DN representa una bifurcación Neimark-Sacker degenerada o una bifurcación de Chenciner.
- R1, R2, R3 y R4 representan, respectivamente, las bifurcaciones: *Resonancia 1:1*, *Resonancia 1:2*, *Resonancia 1:3* y *Resonancia 1:4*.

que, aunque en valor absoluto son menores que uno, toman valores negativos. Ahora bien, podemos aplicar el modelo a valores de los parámetros no convencionales, si observamos el efecto que ello representa en el sistema de ecuaciones que define la relación dinámica entre el mercado potencial y el actual. Así, si sustituimos este valor del esfuerzo de publicidad en las ecuaciones mencionadas, vemos que dicho esfuerzo contribuye a detraer personas del mercado actual del producto para aumentar el mercado potencial de dicho producto. Este hecho suele ocurrir cuando el esfuerzo de publicidad está mal orientado y no es bien acogido por los clientes del producto en estudio. En cambio, al sustituir el paso a una marca rival, el efecto es el contrario, ya que ahora se restarían personas del mercado potencial que pasarían a formar parte del mercado actual.

- Este comportamiento se refleja claramente sobre los valores de los clientes actuales y potenciales encontrados, ya que, como los valores de estos parámetros no influyen sobre el equilibrio alcanzado por el mercado actual, éste permanece inalterable; en cambio, el mercado potencial pierde un gran número de individuos, lo que lleva implícito que la población consumidora sufra un importante retroceso.

- Por otro lado, sabemos que el comportamiento del sistema después de la aparición de este tipo de bifurcación de codimensión 2 puede llegar a ser muy complicado, pudiendo tender el mercado a ciclos estables o inestables que aparecen y desaparecen y que pueden tener una alta periodicidad⁹.

- Desde el punto de vista analítico debemos indicar que este tipo de bifurcación no se puede detectar si se consideran, tan solo, los multiplicadores de la matriz jacobiana en el punto crítico, ya que ambos son complejos conjugados de módulo unitario. El problema añadido es que se infringe una de las condiciones de no-degeneración, porque en este caso se anula el coeficiente $a(0)$ de la forma normal asociada.

* Respecto a la bifurcación resonancia 1:1 (R1)¹⁰:

- Los valores críticos de los parámetros: $c = -0.0606138$, $b = -0.078177474$ vuelven a ser negativos, pero se conservan en módulo menor que uno.

- Esta bifurcación solo presenta un interés teórico, pues se observa un hundimiento del mercado potencial que queda fuera de una estructura razonable de un sistema comercial.

- Bajo el enfoque analítico podemos añadir que, como era de esperar, los multiplicadores de la matriz jacobiana son ambos reales e iguales a uno: $\mu_1 = 1$; $\mu_2 = 1$.

Además, el sistema restringido a la variedad centro bidimensional en el valor crítico de los parámetros, se puede transformar en el sistema dinámico en tiempo continuo:

$$\begin{cases} \xi_1' = \xi_1 + \xi_2 \\ \xi_2' = \xi_2 + 0.0387791\xi_1\xi_2 + O(\|\xi\|^3) \end{cases}$$

donde $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$.

La influencia entre el esfuerzo de publicidad y cada uno de los parámetros restantes no ha producido ningún tipo de bifurcación de codimensión 2.

En la siguiente subsección pasamos a estudiar la influencia entre la tasa de entrada y el resto de los parámetros. Las posibilidades en este caso se describen a continuación.

⁹ Ver Anexo, donde se comenta la aparición y desaparición de ciclos en la bifurcación DN.

¹⁰ Ver Anexo, donde se comenta el comportamiento de esta bifurcación.

4.2. Influencia conjunta de la tasa de entrada (k) y la tasa de salida (e)

Ambas tasas tienen influencia no solo en el equilibrio del mercado potencial sino también en el equilibrio del mercado actual. En este caso, los puntos críticos de bifurcaciones de codimensión 2 se detallan en la siguiente tabla¹¹:

Tipo bif.	Valor k	Valor e	Cl. Pot.	Cl. Act
R4	0.931981	0.715682	14303370	49620045
R3	1.190810	0.777457	13118120	58362867
R2	1.409835	0.816451	12117189	65797323

y que pueden observarse en la Figura 3. Analizamos ahora cada una de las bifurcaciones encontradas:

* *Resonancia fuerte 1:4* (R4) respecto a la cual debemos realizar las siguientes indicaciones:

- Esta bifurcación de codimensión 2 aparece en los valores críticos de la tasa de entrada y de salida $k=0.931981$, $e=0.715682$ que, aunque permanecen dentro del intervalo de definición de los parámetros de ajuste, implican un aumento muy significativo de ambas, ya que se traduce en un incremento de nueve décimas en la tasa de entrada y de siete décimas en la tasa de salida del mercado.

- Ahora bien, esta bifurcación solo tiene un interés teórico debido a que, al ser la tasa de entrada en el mercado mayor que la tasa de salida del mismo (la razón entre ambas es 1.3), se produce un incremento excesivo de la población consumidora que solo tiene sentido si se produce un aumento no esperado de la población adulta.

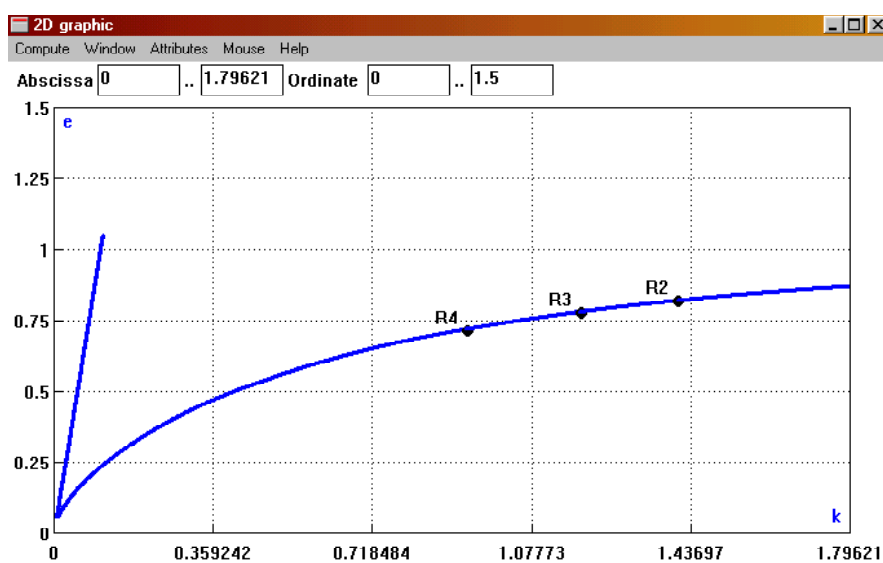


Figura 3. Bifurcaciones de codimensión 2 con la tasa de entrada y de salida del mercado.

- El comportamiento del sistema en un entorno de una bifurcación de este tipo es muy complejo, pudiendo aparecer, a su vez, distintos tipos de bifurcaciones al variar los parámetros. La descripción exacta del comportamiento del sistema para este caso requeriría la computación numérica de algunas curvas de bifurcación que lleva aparejada este tipo de resonancia (ver Anexo).

- Analíticamente podemos añadir que los multiplicadores de la matriz jacobiana en el punto crítico son: $\mu_1 = e^{i \pi/2}$; $\mu_2 = e^{-i \pi/2}$, con lo cual queda reflejado que se infringe la

¹¹ Los valores de los clientes actuales y potenciales se obtienen utilizando los valores de equilibrio obtenidos en la ecuación (15) y el valor de la población (17) en Velasco *et al.* (2007).

condición de no-degeneración $e^{iq\theta} \neq 1$; $q = 1, 2, 3, 4$, ya que $\theta = \pi/2$.

* *Resonancia 1:3 (R3)*

- Esta bifurcación se produce para $k=1.1908108$, $e=0.77745704$, donde la tasa de entrada presenta, de nuevo, un valor no convencional al ser casi dos décimas mayor que uno. Dos son las razones que nos impulsan a tenerla en consideración: la primera es precisamente su proximidad al intervalo exigido para el ajuste de los parámetros; la segunda es que su valor se puede traducir en que se ha producido un incremento no controlado de la población adulta.

- Al igual que ocurre con el caso anterior, esta bifurcación solo presenta un interés analítico, ya que el aumento de población se traduce en un aumento exagerado de la población consumidora al volver a ser la tasa de entrada en el mercado superior a la tasa de salida del mismo.

- Sabemos que para este tipo de bifurcación se tiene una bifurcación de Neimark-Sacker que implica que una curva cerrada e invariante rodea al equilibrio; es decir, en principio se tendría un comportamiento cíclico. Ahora bien, para los valores de los parámetros cercanos a los del punto crítico, la aplicación asociada a la forma normal presenta un ciclo de periodo tres que corresponde a tres nuevos puntos fijos silla. Si se consideraran los términos de orden superior se puede hablar, incluso, de que se producen estructuras homoclínicas que desembocan en comportamientos muy complejos (ver Anexo).

- Por último, podemos indicar que los multiplicadores de la matriz jacobiana en el punto fijo son $\mu_1 = e^{i 2\pi/3}$; $\mu_2 = e^{-i 2\pi/3}$, con lo cual queda reflejado que se infringe la condición de no-degeneración $e^{iq\theta} \neq 1$; $q = 1, 2, 3, 4$, ya que $\theta = 2\pi/3$.

* *Resonancia 1:2 (R2)*

- Los valores críticos de los parámetros son $k=1.4098356$, $e=0.81645175$, lo que nos indica que la tasa de salida ha sufrido un aumento muy considerable pero se mantiene dentro del intervalo de definición del ajuste al mercado cervecero español. No ocurre lo mismo con la tasa de entrada en el mercado, la cual vuelve a aumentar y además se mantiene superior a la tasa de salida del mismo (aumenta incluso la razón entre ambas) llevándonos, de nuevo, a un incremento incontrolado de la población consumidora.

- Analíticamente podemos añadir que el comportamiento que puede presentar el sistema en un entorno del punto crítico nos lleva a la aparición de periodo doble, algo que podíamos esperar, pues aparece el multiplicador -1 con multiplicidad dos. Al considerar los términos de orden superior, pueden aparecer comportamientos de difícil descripción, ya que su estructura implica la existencia de ciclos de gran periodo que aparecen y desaparecen (ver Anexo).

Debemos hacer notar que hemos realizado las simulaciones numéricas para obtener las curvas de puntos de Neimark-Sacker viendo qué influencia tienen sobre el sistema la variación simultánea de dos parámetros de ajuste. Hemos encontrado diversos tipos de bifurcaciones de codimensión 2 cuyo estudio no hemos reflejado, ya que se producen para valores de los parámetros lejos de los valores exigidos para el ajuste del modelo al sistema cervecero español. Aunque, indudablemente, presentan un interés teórico, no aportan nada nuevo, al no encontrar ninguna bifurcación cúspide.

Reflejamos en la Figura 4 una proyección de todas las bifurcaciones encontradas partiendo del punto crítico $c = 0.019101486$ sobre los ejes formados por el parámetro c y la variable x . Se pueden ver otras proyecciones sin más que variar los ejes.

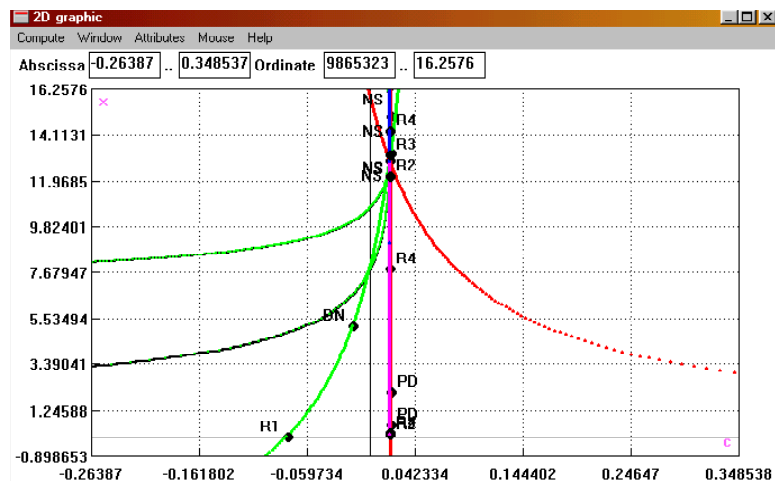


Figura 4. Proyección de las bifurcaciones de codimensión 2 sobre la variable x y el parámetro c .

4.3. Curva de bifurcación de duplicación del periodo

A fin de completar el análisis del mercado cervecero español, hemos buscado las nuevas bifurcaciones de codimensión 2 que pueden aparecer cuando recorremos la curva de bifurcación para las bifurcaciones de duplicación del periodo que hemos encontrado. Aunque sabemos que inicialmente no tienen relevancia económica, nos parece interesante no solo reflejar todas las situaciones posibles con las que deben enfrentarse los mercados en estudio sino también analizar, desde un aspecto más teórico, los comportamientos más singulares a los que pueden tender los mercados antes mencionados.

Para lograr este objetivo, partiendo del valor crítico del esfuerzo de publicidad¹² $c=1.8417928$, hemos realizado las simulaciones numéricas y las simulaciones gráficas variando, nuevamente, los parámetros de dos en dos y estudiando como repercutían sobre los mercados estas variaciones. Las diversas simulaciones gráficas¹³ aparecen en la Figura 5 y la Figura 6. Esta última es una ampliación de la figura anterior donde pueden apreciarse las bifurcaciones encontradas con mayor claridad para los valores $(c, x) \in [1.341, 2.009] \times [-1.02, 1.97]$.

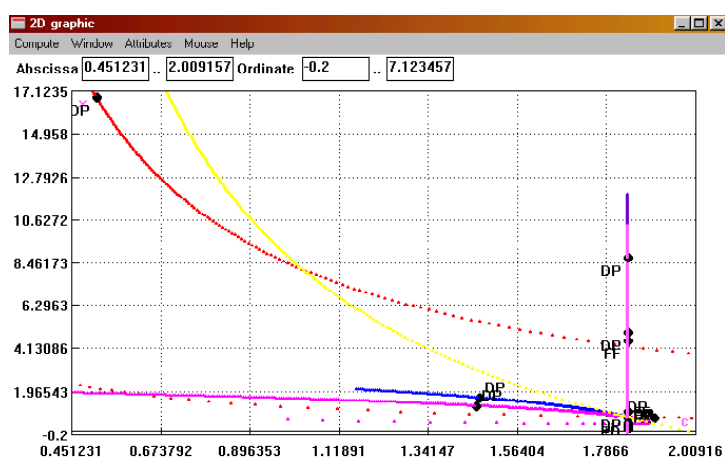


Figura 5. Bifurcaciones de codimensión 2 con ejes la variable x y el parámetro c .

¹² Velasco *et al.* (2007), página 436.

¹³ Hemos representado la variable x , que representa a los clientes potenciales del mercado cervecero, frente a las variaciones del parámetro c , que representa el esfuerzo de publicidad adicional que se realiza sobre el mercado.

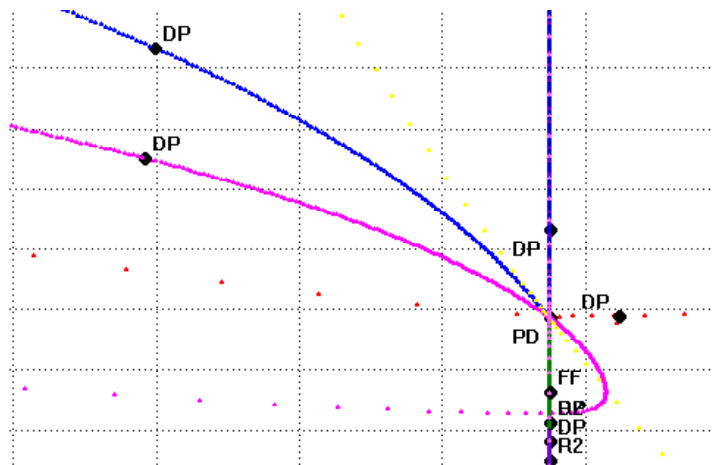


Figura 6. Ampliación de las bifurcaciones de codimensión 2, $(c, x) \in [1.341, 2.009] \times [-1.02, 1.97]$.

5. CONCLUSIONES Y TRABAJO FUTURO

Una de las conclusiones más sobresalientes de este trabajo es que se han encontrado bifurcaciones Flip degeneradas (denotadas por DP), las cuales están asociadas a la aparición de un multiplicador real igual a -1 , pero que vulneran la condición de no degeneración, al anularse el coeficiente del término cúbico de la forma normal a una bifurcación Flip no degenerada. El comportamiento del sistema en un entorno del punto crítico para este tipo de bifurcación depende de la región del plano paramétrico en la que estemos situados, pudiendo incluso coexistir dos ciclos de periodo dos. Se puede apreciar gráficamente que, salvo en una de las bifurcaciones de este tipo que hemos encontrado, se produce el hundimiento del mercado potencial. Hay dos razones que nos impulsan a no considerar ni siquiera esta posibilidad: la primera es que esta bifurcación proviene de una situación que no está incluida en la lógica de un sistema de mercados; y la segunda se debe a que, de nuevo, los valores de los parámetros implicados (la tasa de contacto y el esfuerzo de publicidad) no pertenecen al intervalo exigido para el ajuste al mercado cervecero español.

A parte del estudio analítico realizado al modelo se ha de destacar la bifurcación de Chenciner (DP) para los valores $c = -0.015689665$; $b = -0.02153676786$, en que el mercado puede tender a ciclos estables o inestables que aparecen y desaparecen, con las repercusiones añadidas acerca de la variabilidad del número de individuos en los mercados. Como se comenta en la cuarta sección, esto puede muy bien ser debido a una mala orientación en el esfuerzo de publicidad, haciendo que ella sea contraproducente para los objetivos de aumentar el número de clientes de la marca. Ha habido ocasiones en el que por motivos de sensibilización de la población, lo que en unos años podría ser una buena publicidad, acaba siendo una aberración unos años después. No hay más que ver la evolución que ha habido en la publicidad, no solo en lo relativo a productos de alcohol, sino en todos los ámbitos. No hay más que recordar la publicidad reciente acerca del alcohol y de los accidentes de tráfico. Es por ello por lo que hay que tener muy en cuenta la sensibilidad de la población, con la forma de introducir en nuestro caso el alcohol, aunque sea de baja graduación, es que no solo se ha de tomar con moderación sino, lo que es mucho más importante, con responsabilidad.

Dentro del modelo aparecen, además, singularidades de codimensión 2 no estudiadas con anterioridad, que son la suma de una bifurcación Fold y otra Flip (denotadas por FF). Estas singularidades están asociadas a la aparición de dos multiplicadores reales, uno igual a 1 y otro igual a -1 .

Indicar, por último, que aparecen también resonancias fuertes que no aportan nada nuevo al estudio ya realizado.

Como posibles trabajos futuros, estamos trabajando en un nuevo modelo que describa algunos de los comportamientos de los consumidores, tal como el descontento existente con dicha marca y la repercusión en los clientes actuales y potenciales.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente subvencionado por las ayudas PAI-2008/00000607, PAI-2007/00000633 y por el proyecto de excelencia P06-TIC-02141 de la Junta de Andalucía. Así mismo, ha sido subvencionado por el proyecto del Ministerio de Educación y Ciencia TSI2006-13390-C02-02. Los autores desean agradecer los comentarios de los revisores que hacen que este trabajo tenga una mejor lectura.

REFERENCIAS

BISCHI, G.I., DIECI, R., RODANO, G., SALTARI, E. (2001). Multiple attractors and global bifurcations in a Kaldor-type business cycle model. *Journal of Evolutionary Economics*, 11 (5): 527–554.

BRAUN, M. (1990). *Ecuaciones diferenciales y sus aplicaciones*, Grupo editorial Americana, México D.F.

CHANG, J.J., CHANG, W.Y., LAI, C.C., WANG, P. (2007). Equilibrium dynamics in an endogenous growth model of money and banking, *Journal of Money Credit and Banking*, 39, 7: 1683–1710.

DONG, J., ZHANG, D., NAGURNEY, A.(1996). A projected dynamical systems model of general financial equilibrium with stability analysis. *Mathematical and computer Modelling*, 24 (2): 35–44.

FEICHTINGER, G. (1992). Hopf bifurcation in an advertising diffusion model, *Journal of Economic Behavior and Organization* 17, 401–411.

GANDOLFO, G. (1997). *Economic Dynamics*, third ed., Springer-Verlag, Berlin.

GOVAERTS, W.J. (2000). *Numerical Methods for Bifurcations of dynamical equilibria*, SIAM, Philadelphia.

GUCKENHEIMER, J., HOLMES, P. (1993). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, fourth ed., Springer-Verlag, New York.

GUCKENHEIMER, J., MYERS, M., STURMFELS, B. (1997). Computing Hopf Bifurcations I, *Siam J. Numer. Anal.*, 34 (1): 1–21.

JOOSTER, R. (1996). Deterministic evolutionary dynamics: A unifying approach. *Journal of Evolutionary Economics*, 6 (3): 313–324.

KUZNETSOV, Y.A. (1998). *Elements of Applied Bifurcation Theory*, Applied Mathematical Sciences, second ed., Springer, New York.

- KUZNETSOV, Y.A., LEVITIN, V.V. (2004). CONTENT. A multiplatform environment for analysing dynamical systems, <http://www.math.uu.nl/people/Kuznet/CONTENT>
- LANDA, J., VELASCO, F. (2004). Análisis dinámico del mercado actual y potencial de las organizaciones. *Revista Europea de Dirección y Economía de la Empresa*, 13 (1): 131–140.
- MITRA, T., NISHIMURA, K. (2005). Intertemporal complementarity and optimality study of a two-dimensional dynamical system, *International Economic Review*, 46 (1): 93–131.
- NADAL, P. (2004). Bifurcaciones en modelos dinámicos discretos: Una aplicación al mercado cervecero, Tesis doctoral, Universidad de Sevilla.
- NAGURNEY, A., ZHANG, D. (1996). On the stability of an adjustment process for spatial price equilibrium modelled as a projected dynamical system, *Journal of Economics Dynamics & Control*, 20 (1-3): 43–62.
- ORTEGA, J.A., MARTÍNEZ, R., TORRES, J., TORO, M., GONZÁLEZ, L., VELASCO, F., ANGULO, C. (2005). Metodología semicualitativa para razonar sobre sistemas dinámicos. *Computación y sistemas. Revista Iberoamericana de Computación*. Vol. 8. Núm. 3: 231–251.
- PERRINGS, C. (1998). Resilience in the dynamics of economy-environment systems, *Environmental & Resource Economics*, 11 (3-4): 503–520.
- TSE, N.S.F., ROBB, F.F. (1994). Dynamical Systems Theory applied to Management Accounting-Chaos in cost behaviour in a standard costing system setting. *Transactions of the Institute of Measurement and Control*, 16 (5): 269–279.
- VELASCO, F., BEGINES, F., NADAL, P., CHAMIZO, C., VÍLCHEZ, M. (2002). Continuación de los equilibrios de un sistema dinámico económico con bifurcaciones de codimensión 1 y 2, *Computación y Sistemas* 5 (3): 169–179.
- VELASCO, F., NADAL, P., GONZALEZ, L., VILCHEZ, M.L. (2007). Estudio de la estabilidad y de las bifurcaciones de los equilibrios de un sistema dinámico. Aplicación al mercado cervecero español. *Estudios de Economía Aplicada*, 25 (1)(on line): 419–452.
- VÍLCHEZ, M., VELASCO, F., GARCÍA, J. (2002). Bifurcaciones transcricas y ciclos límite en un modelo dinámico de competición entre dos especies. Una aplicación a la pesquería de *engraulis encrasicholus* de la región suratlántica española, *Estudios de Economía Aplicada*, 20 (3): 651–677.
- VÍLCHEZ, M., VELASCO, F., GONZÁLEZ L., ORTEGA, J. (2003). Bifurcaciones de Hopf: Análisis cualitativo y aplicación a un modelo bioeconómico de pesquerías, *Computación y Sistemas*, 6 (4): 273–283.
- VÍLCHEZ, M., VELASCO, F., HERRERO, I. (2004). An optimal control problem with Hopf bifurcations and application to the striped venus fishery in the gulf of Cadiz, *Fisheries Research*, 67 (3): 295–306.
- WIGGINS, S. (1990). *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer-Verlag.

ANEXO

Bifurcación de Chenciner o de Neimark-Sacker degenerada(DN)

Dado el sistema discreto $x \rightarrow f(x, \alpha)$, $x \in \mathbb{R}^2$, $\alpha \in \mathbb{R}^2$, donde la función f es suave, con un punto fijo en $x=0$, en el que se tiene una bifurcación de Neimark-Sacker ($\mu_{1,2} = e^{\pm i \theta_0}$). Este sistema se puede pasar a coordenadas polares y se obtiene un sistema que depende de unos nuevos parámetros que son funciones de los parámetros del modelo inicial. Al realizar el diagrama de bifurcación correspondiente, aparecen tres regiones paramétricas delimitadas por una recta de bifurcación de Neimark-Sacker y una curva de bifurcación. En la primera región el *map* presenta un único punto fijo que es estable. En la segunda se tiene un punto fijo inestable que tiende hacia un círculo invariante estable. En la tercera, el punto fijo es estable y tiene dos círculos invariantes, uno inestable más cercano al punto fijo y otro estable más alejado. Así, al pasar en una determinada dirección la recta de bifurcación, el punto fijo pasa de estable a inestable y los dos círculos invariantes colisionan desapareciendo el inestable. Análogamente, al atravesar en una determinada dirección la curva de bifurcación, el punto fijo permanece estable, pero la colisión de los dos círculos invariantes hace que surja un círculo invariante silla. La situación se complica cerca de la curva de bifurcación, ya que en un entorno de la misma pueden darse variedades invariantes estables e inestables de intersecciones de puntos de silla que dan lugar a una estructura homoclínica.

Se puede ver Kuznetsov (1998) para un estudio más extenso de las bifurcaciones de codimensión 2 para un *map*.

Resonancias fuertes

Al tratar con las resonancias fuertes se suele aproximar el sistema dinámico discreto en el punto de equilibrio por un sistema dinámico continuo determinado. Esto permite predecir bifurcaciones globales de curvas invariantes cerradas que ocurren en los sistemas dinámicos discretos cerca de bifurcaciones homoclínicas o heteroclínicas del sistema dinámico continuo aproximado. Aunque la estructura exacta de la bifurcación del sistema discreto es diferente que la del sistema dinámico aproximado, su uso suministra información que es disponible solo para el análisis del sistema discreto.

Bifurcación R 1:1

Sea el sistema dinámico discreto $x \rightarrow f(x, \alpha)$, $x \in \mathbb{R}^2$, $\alpha \in \mathbb{R}^2$, con un punto fijo en $x = \theta$, para $\alpha = \theta$ y con multiplicador unitario doble $\mu_{1,2} = 1$. Este sistema puede ser representado por un sistema dinámico continuo cuyo comportamiento se describe por la bifurcación de Bogdanov-Takens. El diagrama de bifurcación describe las bifurcaciones del *map* aproximadamente, los equilibrios corresponden a puntos fijos, mientras que los ciclos límite se interpretan como curvas cerradas invariantes. Sobreviven del diagrama correspondiente al sistema continuo dos curvas de bifurcación, una Fold y otra Neimark-Sacker que se encuentran tangencialmente en un punto de codimensión dos, y para unos determinados valores de los parámetros existe una curva invariante cerrada que se bifurca desde el punto fijo estable. Existen otros tipos de comportamientos más complejos que dan lugar a tangencias homoclínicas en dos curvas de bifurcación suaves.

Bifurcación R 1:2

Sea el sistema dinámico discreto $x \rightarrow f(x, \alpha)$, $x \in \mathbb{R}^2$, $\alpha \in \mathbb{R}^2$, con un punto fijo en $x = \theta$, para $\alpha = \theta$ y con multiplicador unitario doble $\mu_{1,2} = -1$. En el diagrama de

bifurcación del sistema dinámico continuo asociado aparecen puntos de equilibrio que se bifurcan del trivial mediante una bifurcación *pitchfork*. Así mismo, aparecen bifurcaciones Hopf que hacen que el equilibrio pase de estable a inestable apareciendo ciclos límites. Se construye un *map* aproximante para el cual la bifurcación *pitchfork* conserva su sentido, mientras que las bifurcaciones Hopf se transforman en bifurcaciones Neimark-Sacker con curvas invariantes cerradas que conservan la estabilidad de los ciclos límites. Como el *map* aproxima la segunda iteración del *map* original, resulta que la bifurcación *pitchfork* se transforma en una de doble periodo. Como en los casos de las bifurcaciones de codimensión dos descritas previamente, existen un conjunto de bifurcaciones más complejas. Las colisiones entre las variedades invariantes sillar y las curvas invariantes cerradas se sustituyen por infinitas series de bifurcaciones en las que se envuelven estructuras homoclínicas.

Bifurcación R 1:3

Sea el sistema dinámico discreto $x \rightarrow f(x, \alpha)$, $x \in \mathbb{R}^2$, $\alpha \in \mathbb{R}^2$, con un punto fijo en $x = \theta$, para $\alpha = \theta$ y con multiplicadores simples $\mu_{1,2} = e^{\pm i\theta_0}$ para $\theta_0 = 2\pi/3$. El sistema dinámico continuo asociado presenta un punto de equilibrio trivial y tres puntos de equilibrio no triviales. Aparecen una línea de bifurcación de Hopf y una curva de bifurcación sobre la cual el sistema tiene un ciclo heteroclínico similar a un triángulo y que conecta los tres puntos de equilibrio no triviales. El *map* original siempre tiene un punto fijo trivial que sufre una bifurcación de Neimark-Sacker. Además, cerca del punto de codimensión dos presenta un ciclo silla de periodo tres que corresponde a los tres puntos fijos (sillas) no triviales, que a su vez corresponden con los tres equilibrios (sillas) ya mencionados. Así mismo, se producen cerca de la resonancia R1:3 comportamientos más complejos, ya que las variedades estables e inestables se intersectan formando una estructura homoclínica.

Bifurcación R 1:4

Sea el sistema dinámico discreto $x \rightarrow f(x, \alpha)$, $x \in \mathbb{R}^2$, $\alpha \in \mathbb{R}^2$, con un punto fijo en $x = \theta$, para $\alpha = \theta$ y con multiplicadores simples $\mu_{1,2} = \pm i$. El sistema dinámico continuo asociado presenta ocho regiones en las cuales hay comportamientos tan diversos tales como lazos homoclínicos, ciclos heteroclínicos cuadrados, ciclos heteroclínicos en forma de trébol, conexiones heteroclínicas nodo-silla u órbitas heteroclínicas nodo-silla degeneradas.

Volumen 7 (junio de 2009)

Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus

Imputation methods to handle the problem of missing data: an application using R/Splus

Muñoz Rosas, Juan Francisco; Alvarez Verdejo, Encarnación

Páginas 3–30

Localización empresarial en Aragón: una aplicación empírica de la ayuda a la decisión multicriterio tipo ELECTRE I y III. Robustez de los resultados obtenidos

Industrial Location Decisions in Aragon: an Empirical Application of a Multi-criterion Methodology of Aid to the Decision using ELECTRE Methods I and III. Robustness of the Obtained Results

García Cebrián, Lucía Isabel; Muñoz Porcar, Antonio

Páginas 31–56

Evaluación de la eficiencia de una cartera de asegurados en el sector del automóvil

Efficiency Assessment in the Automobile Insurance Sector

Segovia González, M. Manuela; Contreras Rubio, Ignacio; Mar Molinero, Cecilio

Páginas 57–76

Bifurcaciones de codimensión 2 en un modelo dinámico del mercado potencial y actual: aplicación al mercado cervecero español

Bifurcations of Codimension 2 in a Dynamical Model of Current and Potential Market: An Application to the Spanish Beer Market

Velasco Morente, Francisco; Nadal Morales, Purificación; González Abril, Luis; Ortega Ramírez, Juan Antonio

Páginas 77–94