

Un modelo con estructuras paralelas de Markov para el estudio de la relación temporal entre situación laboral y cargas familiares

ATO GARCÍA, M.*; GÓMEZ GARCÍA, J.** Y FAURA MARTÍNEZ, U.**

*Departamento de Psicología Básica y Metodología, Universidad de Murcia; **Departamento de Métodos Cuantitativos para la Economía, Universidad de Murcia; ***Departamento de Métodos Cuantitativos para la Economía, Facultad de Economía y Empresa.

Campus de Espinardo. Universidad de Murcia, 30100 MURCIA. E-mail: faura@um.es

RESUMEN

En este trabajo se aplica la metodología estadística conocida como *análisis log-lineal dirigido* (“directed log-linear modelling”), con estructuras paralelas de Markov, para analizar la estructura temporal de las relaciones entre variables a partir de la información suministrada por encuestas retrospectivas. En concreto, se analiza la relación existente entre la “Situación Familiar” y la “Situación Laboral” en función del Género y Nivel de Estudios, en España y en tres períodos de actividad laboral comprendidos entre los 25 y 45 años de edad (datos obtenidos de la Encuesta de Conciencia, Biografía y Estructura de Clase de 1991). El interés metodológico del trabajo es el empleo de los modelos *path-logit* como herramienta para comprender la verdadera estructura de la relación entre las variables, frente al uso de los modelos más clásicos, como regresión por mínimos cuadrados. Desde una perspectiva sustantiva, se verifica la existencia de asociación significativa entre Situación Laboral y Familiar en cada uno de los tres períodos estudiados.

Palabras Clave: Datos Categóricos, Análisis Log-Lineal Dirigido, Situación Familiar y Situación Laboral.

A model with structures of markov for the study of the temporary relation between labor situation and familiar loads

ABSTRACT

In this work the statistical methodology known as *directed log-linear modelling* with parallel structures of Markov, is used in order to analyze the temporary structure of the relations between variables from the information provided by retrospective surveys. In particular, the relation between the Familiar Situation

Clasificación JEL: C23, C25.

Artículo recibido en mayo de 2003 y aceptado para su publicación en octubre de 2005.

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref.: E-23310.

ISSN 1697-5731 (online) – ISSN 1133-3197 (print)

and the Labor Situation based on the Gender and Educational Level is analyzed in Spain along three periods of labor activity between 25 and 45 years old (collected data from the “Encuesta de Conciencia, Biografía y Estructura de Clase”, 1991). The main methodological interest of this work is the use of *path-logit* models as instruments to do comprehensive the true structure of the relationships between variables, instead more classical models as OLS regression. From a substantive point of view, we verify the existence of a significant association between the Familiar Situation and the Labor Situation for each one of the three analyzed periods.

Key Words: Categorical Data, Directed Log-Linear Modelling, Familiar Situation, Labor Situation.

1. INTRODUCCIÓN

A partir de mediados de los ochenta, se ha producido en España una progresiva incorporación de la mujer al mercado de trabajo constituyendo uno de los fenómenos económicos y sociales más importantes que han tenido lugar en las últimas décadas. Así, la tasa de participación de la mujer ha aumentado en un 30% en los últimos 20 años (pasando del 26,88% en el año 1978 al 37,72% en el año 1998) y si nos centramos en las mujeres de edad intermedia (24 y 44 años), ésta ha aumentado un 22% entre 1981 y 1991. Este aumento de la participación laboral de las mujeres españolas, que se debe fundamentalmente a la incorporación al mercado de trabajo de las mujeres de 25 a 45 años, ha tenido lugar en un contexto de cambio en la estructura productiva, institucional y familiar española. Son varios los factores que han causado dicho aumento, como, por ejemplo, el cambio en la composición de la demanda de trabajo hacia ocupaciones del sector servicios donde las mujeres encuentran más oportunidades de trabajo (Arellano y Bover (1995); Bover (1997)), o el incremento significativo del nivel educativo de la población femenina española, que ha llegado a superar en los últimos años al de los hombres (en el año 1978, el 3,06% de los hombres activos y el 1,90% de las mujeres activas tenían estudios universitarios; en 1998, estos porcentajes son del 14,16% y del 23,41% respectivamente). El aumento de la participación laboral femenina se ha traducido en un aumento del peso del empleo femenino, si bien existen diferencias con respecto a los varones en las condiciones de dicho empleo en lo que se refiere a la estabilidad del mismo (medido por las tasas de temporalidad) y a los salarios percibidos. Los hechos anteriores son bien conocidos y están ampliamente documentados en numerosos trabajos, véase, por ejemplo Toharia *et al.* (1998).

Otro de los aspectos interesantes de la evolución de la tasa de actividad femenina en nuestro país se refiere a las diferencias de participación laboral por grupos de edad. El aumento de las tasas de actividad femenina ha ido acompañado de un cambio en el modelo de participación laboral cada vez más similar al masculino. Toharia *et al.* (1998) han realizado un análisis desagregado por grupos de edad y observan cómo las mujeres españolas han abandonado el antiguo modelo de participación laboral unimodal, donde la participación se concentraba en las edades más jóvenes

para posteriormente abandonar el mercado laboral en el momento del matrimonio o del primer hijo. Actualmente el perfil temporal de la tasa de actividad femenina tiene forma de U invertida, con una menor participación laboral de las más jóvenes, que permanecen más tiempo como estudiantes, y de las cohortes de mayor edad. Cebrián *et al.* (1997) observan cómo en los últimos años, las mujeres casadas con hijos entre 6 y 15 años de edad han visto aumentar sus tasas de actividad hasta situarse por encima del grupo de las solteras.

En España, la familia juega un papel muy importante como forma de protección de los individuos contra situaciones precarias en el mercado de trabajo. En este sentido, las diferencias entre las tasas de paro según la situación familiar son bastante ilustrativas: en el segundo trimestre de 1998, cuando la tasa de paro general era del 18,91%, la de la persona de referencia del hogar era del 9,34% y la tasa familiar de paro (definida como la proporción de familias en las que todos los miembros están parados, tomando como población a las familias en las que hay algún miembro activo en el mercado de trabajo) era del 8,47%. Así pues, el paro parece especialmente concentrado en los individuos que constituyen las segundas fuentes de renta, véase Cebrián y Jimeno (1998).

Este trabajo se propone como objetivo estudiar las relaciones entre situación familiar (cargas familiares) y situación laboral, en función del género y del nivel de estudios en un periodo de actividad laboral comprendido entre los 25 y los 45 años¹ mediante el uso de una encuesta retrospectiva. La mayoría de los trabajos sobre estos temas, usan datos transversales, para un período de tiempo dado, empleando mínimos cuadrados ordinarios (MCO) y/o variables instrumentales. Una peculiaridad de este trabajo es la utilización de datos de panel, con medidas repetidas de los mismos sujetos, mediante una metodología estadística especialmente indicada para estudiar tales relaciones con encuestas retrospectivas, el *análisis log-lineal dirigido* (“directed log-linear modelling”), una contribución original de Goodman (1972, 1973), posteriormente desarrollada por Hagenaars y colaboradores (Hagenaars (1998), Heinen (1995), Vermunt (1996)).

La estructura del artículo es la siguiente. En el segundo apartado se describe con detalle el procedimiento utilizado (un modelo log-lineal con estructuras paralelas de Markov junto con un análisis log-lineal dirigido). En el tercero, se realiza un análisis de los resultados, con una interpretación de los mismos a partir de varias tablas generadas durante el proceso estadístico. Finalmente, en el último apartado se exponen las conclusiones.

1. La banda de edad comienza a los 25 años, por ser esta edad en la cual se han terminado los estudios superiores, y se termina a los 44, dado que la participación de las mujeres de más edad se ve determinada sobre todo por su situación anterior. Véase Bover (1995).

2. METODOLOGÍA

La naturaleza categórica de las variables utilizadas en este estudio nos induce a emplear alguna técnica de análisis de datos categóricos para dar respuesta a los objetivos planteados en la sección anterior. El más general de todos los procedimientos de análisis para datos categóricos es el análisis log-lineal (Bishop, Fienberg y Holland (1975), Agresti (1990), Hagenaars (1990), Christensen (1990), Ato y López (1996)).

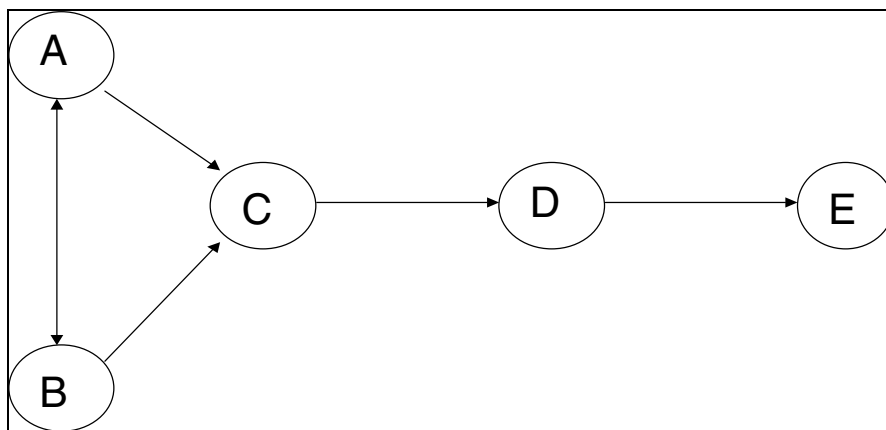
Mientras que el análisis log-lineal persigue analizar la asociación conjunta existente entre un conjunto de variables, y, por tanto, no establece supuestos a priori sobre la ordenación causal de tales variables, una extensión del análisis log-lineal que permite definir una variable de respuesta (o variable dependiente) respecto de un conjunto de variables explicativas, es el análisis logit. Goodman (1972) lo llamó también *enfoque de regresión modificado* (“modified regression approach”). La situación más típica se presenta cuando la variable de respuesta es dicotómica o binaria (modelo logit binomial), pero también se pueden utilizar variables de respuesta politómicas (modelo logit multinomial).

Una extensión del análisis logit que permite formular varios modelos para un mismo conjunto de variables, fue también desarrollada por Goodman (1973), quien en paralelismo con el *análisis de trayectorias* (“path analysis”) lo llamó *enfoque análisis de trayectorias modificado* (“modified path analysis”). Este enfoque tiene también muchas similitudes con los modelos gráficos para datos categóricos (Wermuth y Lauritzen (1983, 1990), Whittaker (1990), Smith (1995)). Recientemente, Vermunt (1996, 1997)², ha propuesto una fusión de los dos enfoques en un procedimiento refinado que llama *modelado log-lineal dirigido* (“directed log-linear modelling”). Basándose en la relación entre modelo log-lineal y modelo logit, los modelos log-lineales dirigidos definen modelos logit parsimoniosos (es decir, con restricciones posibles de componentes no significativos) para tablas de contingencia que son tan simples como es posible su representación en términos gráficos.

Para ilustrar la naturaleza y alcance de este enfoque, supóngase que se desea investigar las relaciones causales que existen entre cinco variables categóricas, que denominaremos A, B, C, D y E. La Figura 1 muestra el flujo causal que se supone entre las variables, donde las flechas entre dos variables indican la existencia de relación, la presencia de una flecha admite que se trata de una relación asimétrica (causal) y la presencia de dos flechas de una relación simétrica (no causal). En este modelo las variables A y B son variables exógenas, y, por tanto, no se especifica ningún orden de causalidad entre ellas, mientras que las variables C, D y E son variables endógenas y su orden causal establece que C es anterior a D y E es posterior a D. La figura asume que C depende de A y B, D depende de C y E depende de D.

2. Véase también Hagenaars (1998).

Figura 1. Modelo de análisis de trayectorias modificado



Siendo π_{ijklm} la probabilidad de que $A=i, B=j, C=k, D=l$ y $E=m$, en el enfoque de Goodman (1973) la información sobre el orden causal se utiliza para descomponer la probabilidad conjunta en un producto de probabilidades marginales y condicionales, de tal modo que el modelo de la Figura 1 puede formularse mediante

$$\pi_{ijklm}^{ABCDE} = \pi_{ij}^{AB} \pi_{ijk}^{C|AB} \pi_{kl}^{D|C} \pi_{lm}^{E|D} \quad (1)$$

Obsérvese la especificación del orden causal que se establece en la Ecuación (1). Por ejemplo, D depende directamente de C, e indirectamente de A y B, pero no de E, que se considera una variable causalmente posterior. Además, la estructura formada por las variables C, D y E forma un modelo de Markov de tiempo discreto de primer orden, donde C, D y E constituyen estados presentados en los momentos temporales $t, (t+1)$ y $(t+2)$, pero los estados que presenta E en el momento $(t+2)$, y D en el momento $(t+1)$, se asume que dependen sólo del estado ocupado en el momento inmediatamente anterior $(t+1)$ y t , respectivamente. El modelo de la Ecuación (1) (representado en la Figura 1), se compone de varios submodelos, a saber,

$$\pi_{ij}^{AB} = \frac{\exp(\lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB})}{\sum_{ij} \exp(\lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB})}$$

$$\pi_{ijk}^{C|AB} = \frac{\exp(\lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC})}{\sum_k \exp(\lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC})}$$

$$\pi_{kl}^{D|C} = \frac{\exp(\lambda_l^D + \lambda_{kl}^{CD})}{\sum_l \exp(\lambda_l^D + \lambda_{kl}^{CD})}$$

$$\pi_{lm}^{E|D} = \frac{\exp(\lambda_m^E + \lambda_{lm}^{DE})}{\sum_m \exp(\lambda_m^E + \lambda_{lm}^{DE})}$$

Esta representación contiene un submodelo log-lineal y tres submodelos logit. Puesto que especificar un submodelo logit para probabilidades condicionales es equivalente a especificar un submodelo log-lineal para una tabla de frecuencias en la que las variables explicativas se tratan como fijas, las ecuaciones logit anteriores son equivalentes a ecuaciones log-lineales. El conjunto completo de submodelos log-lineales es:

$$\begin{aligned} \log(m_{ij}) &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \\ \log(m_{ijk}) &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \\ \log(m_{kl}) &= \lambda + \lambda_k^C + \lambda_l^D + \lambda_{kl}^{CD} \\ \log(m_{lm}) &= \lambda + \lambda_l^D + \lambda_m^E + \lambda_{lm}^{DE} \end{aligned}$$

donde m_{ijk} , m_{kl} y m_{lm} son frecuencias esperadas de las tablas marginales ABC, CD y DE respectivamente. Obsérvese que el primer submodelo log-lineal es un subconjunto del segundo, y, por tanto, puede prescindirse de él. Los restantes submodelos log-lineales constituyen la representación básica de la Ecuación (1), (Figura 1).

Los estimadores de máxima verosimilitud de los parámetros log-lineales y las frecuencias esperadas de los tres submodelos log-lineales pueden ser estimados separadamente para cada submodelo aplicando cualquier programa para análisis log-lineal sobre las frecuencias marginales correspondientes (Goodman (1973), Hagenaars (1990), Vermunt (1997)). Este resultado deriva del hecho de que, cuando los submodelos que componen un modelo log-lineal dirigido son diferentes, la verosimilitud global puede ser factorizada en componentes que pueden ser maximizados de forma independiente,

$$\begin{aligned} \log L &= \sum_{ijklm} \log(\pi_{ijklm}^{ABCDE}) = \sum_{ijk} n_{ijk++} \log(\pi_{ijk}^{ABC}) \\ &\quad + \sum_{kl} n_{++kl+} \log(\pi_{kl}^{CD}) \\ &\quad + \sum_{lm} n_{+++lm} \log(\pi_{lm}^{DE}) \end{aligned}$$

donde n_{ijk++} son frecuencias de la tabla marginal ABC, n_{++kl+} de la tabla marginal CD y n_{++++m} de la tabla marginal DE. La factorización de la verosimilitud hace posible estimar los parámetros de modelos log-lineales dirigidos mediante un programa estándar para análisis log-lineal. El programa *LEM* de Vermunt (1997) es especialmente recomendable para este tipo de procedimientos por las facilidades que presenta en la especificación y la planificación de restricciones para los submodelos.

Una interesante característica de los modelos log-lineales dirigidos que se ha empleado particularmente en este trabajo concierne a la posibilidad de combinar dos o más submodelos logit en un submodelo con dos o más variables de respuesta. Esta propiedad deriva del *teorema de la colapsabilidad* (Bishop, Fienberg y Holland, 1975), que establece que si dos variables se suponen condicionalmente independientes de una tercera, la magnitud de su relación con la tercera variable puede ser estimada en cualquier tabla donde se incluya la variable condicionalmente independiente.

Además, la factorización de la verosimilitud puede también ser utilizada para probar la bondad de ajuste de los submodelos. Esta ventaja deriva del hecho de que son modelos jerárquicos. Una importante consecuencia de ello es que la desviación³ total de un modelo log-lineal dirigido puede también obtenerse sumando las desviaciones (y los grados de libertad) de cada uno de los submodelos de que se compone. Pero para ello es preciso que los submodelos sean estrictamente jerárquicos, es decir, que un submodelo con una variable causal posterior incluya todos los componentes de los submodelos con variables causales anteriores. Aunque este enfoque solo puede en esencia aplicarse a modelos recursivos, es, sin embargo, extremadamente flexible y permite formular muchos modelos interesantes de forma muy simple. Una vez especificado un modelo determinado, el interés se concentra en encontrar el modelo particular más parsimonioso que sea capaz de explicar adecuadamente los datos empíricos mediante la exclusión de los componentes no significativos. Esto se consigue utilizando restricciones apropiadas sobre los componentes posibles del modelo saturado⁴ que se deriva de la tabla marginal correspondiente. Finalmente, se ajustan los submodelos logit especificados a las tablas marginales y se obtienen sus parámetros.

3. ANÁLISIS DE LOS RESULTADOS

Los datos que se utilizan en este trabajo proceden de la Encuesta de Conciencia, Biografía y Estructura de Clase (ECBC), que se realizó por el Instituto Nacional de

3. La desviación es una medida de discrepancia para evaluar el ajuste entre los valores observados de la variable de respuesta y los valores ajustados derivados del predictor lineal. Véase Ato y López (1996).

4. Un modelo saturado o completo es el que se caracteriza por que tiene tantos parámetros como observaciones, y por tanto reproduce los datos exactamente sin ninguna simplificación. Véase Ato y López (1996).

Estadística, la Comunidad Autónoma de Madrid y el Instituto de la Mujer en 1991, utilizando el censo electoral, sobre un extenso cuestionario que recogía abundante información laboral de trabajadores, desempleados e inactivos. Se realizaron un total de 6600 encuestas, de las que 1600 se realizaron en Madrid y las otras 5000 en las restantes Comunidades Autónomas. Dos sesgos importantes que contiene esta encuesta, referentes a la sobrerrepresentación de la submuestra de Madrid y la de individuos con estudios medios y altos, se encuentran bien documentados (Carabaña *et al.* 1992).

Dos cuestiones que se registraron de forma retrospectiva son el foco del interés en este trabajo: la Situación Laboral y las Cargas Familiares a los 45, a los 35 y a los 25 años. La muestra se ha restringido a individuos de 45 años o mayores, y se han tomado además dos variables externas como covariantes, las variables G (Género) y H (Formación Académica, a 3 niveles, 1:Alta, 2:Media y 3:Baja⁵). Las variables básicas son seis: A, C y E son medidas a los 25, a los 35 y a los 45 años de la Situación Laboral (observada a dos niveles, 1: Trabaja y 2: No trabaja) y B, D y F son medidas en los mismos momentos de las Cargas Familiares (observado también a dos niveles, 1:Sin Cargas Familiares y 2:Con Cargas Familiares), entendiéndose por Cargas Familiares el estar o haber estado casado, de forma que el nivel 1 se correspondería con el estado civil soltero y el 2 con cualquier otra situación. Los datos se organizan en una tabla de contingencia $G \times H \times A \times B \times C \times D \times E \times F$, utilizando un total de 384 casillas. El tamaño muestral, utilizando todos los casos con respuesta, es de 1521 sujetos, lo que representa un *índice de esparcimiento* (*sparse index*, Agresti, 1990, p. 244) de 0,25; de las 384 casillas, solo 43 (10%) tienen frecuencias superiores a 5.

Los estimadores de máxima verosimilitud de los parámetros de los modelos tratados aquí se han obtenido con un programa estándar para modelos log-lineales que utiliza el algoritmo EM: el programa LEM (Vermunt, 1997).

3.1. Primera fase: análisis log-lineal

Con el objeto de obtener una perspectiva global de las asociaciones de la tabla de contingencia $2 \times 3 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$ resultante, en una primera fase interesé realizar un análisis de los mejores subconjuntos de interacciones de un orden determinado, comenzando por el subconjunto de interacciones de orden superior y descendiendo hasta el de orden inferior. Los modelos con interacciones de hasta tercer orden no pudieron ser ajustados debido a problemas de convergencia causados por la especial estructura de esparcimiento que los datos presentaban. El modelo con todas las interacciones de segundo orden (Modelo M1) presentó un ajuste perfecto⁶,

5. La Formación Académica se ha categorizado en tres niveles: "Alta": estudios universitarios y similares; "Media": Bachiller, Formación Profesional II y similares; "Baja": estudios inferiores a los que definen la Formación Media.

$L^2(263)=24,75$; $P=1,0000$, pero la matriz de información no era de rango completo y precisó más de 10000 iteraciones EM para alcanzar el criterio de convergencia establecido por el programa LEM ($10 \cdot e^{-6}$).

El modelo con todas las interacciones de primer orden (Modelo M2) también se ajustó de forma óptima, $L^2(339)=273,72$; $P=0,9961$ y precisó tan solo 25 iteraciones EM obteniendo una matriz de información de rango completo. Sin embargo, presentaba muchos parámetros no significativos y excluía componentes interactivos de segundo orden relevantes, particularmente las interacciones GHA, GHB, GAB, GCD y GEF. La incorporación de tales interacciones condujo a un nuevo modelo (Modelo M3), que también obtuvo un ajuste óptimo, $L^2(334)=163,49$, $P=1,0000$. La diferencia entre ambos modelos resultó estadísticamente significativa: $L^2(M3|M2)=110,23$, con 5 grados de libertad, lo que revela que el modelo más complejo (M3) es un modelo más conveniente. Y finalmente, la exclusión de los componentes asociativos de primer orden no relevantes, y jerárquicamente no involucrados en las interacciones de segundo orden anteriormente incorporadas (asociaciones AD, AE, AF, BC, BE, BF, CF, DE, HE y HF), condujo a un modelo final (Modelo M4) que obtuvo asimismo un ajuste perfecto, $L^2(346)=176,85$, $P=1,0000$. La diferencia con respecto al modelo M3 no es estadísticamente significativa: $L^2(M4|M3)=13,36$, con 12 grados de libertad, y, por tanto, seleccionamos el modelo M4 por ser más parsimonioso.

La Tabla 1 presenta un resumen de la bondad de ajuste de los modelos log-lineales utilizados en la primera fase. El estadístico BIC (*Bayesian Information Criteria*⁷) más bajo favorece al Modelo M4 respecto a los restantes modelos sometidos a prueba en esta fase. El ajuste de todos los modelos se obtuvo con el programa LEM, utilizando codificación ficticia (*dummy coding*) para identificación de los modelos, fijando como nivel de referencia la primera categoría de la variable, que se hace igual a cero para facilitar la interpretación de las restantes categorías y reducir el número de parámetros necesario.

Tabla 1. Análisis log-lineal primera fase

Modelo	g.l.	L^2	P	BIC
(M1) Interactivo de segundo orden completo	263	24,75	1,000	-1902,3
(M2) Interactivo de primer orden completo	339	273,72	0,996	-2210,0
(M3) Modelo (2) + [GAB,HAB,GCD,GEF]	334	163,49	1,000	-2273,9
Diferencia (M3 M2)	5	110,23	0,000	
(M4) Modelo (3) – [AD,AE,AF,BC,BE,BF,CF,HE,HF]	346	176,851	1,000	-2358,3
Diferencia (M4 M3)	12	13,36	0,343	
(M5) Modelo GHA,GAB,GCD,GEF,AC,BD,CE,DF	348	202,25	1,000	-2347,6

6. L^2 es la razón de verosimilitud y P la probabilidad asociada. El criterio utilizado para clasificar la bondad del ajuste es: perfecto si $P=1$, óptimo si $P>0,8$ y aceptable si $0,1 < P < 0,8$. Véase Ato y López (1996).

7. Véase Raftery (1995).

3.2. Segunda fase: análisis log-lineal dirigido

Un detenido examen de las características del Modelo M4 (GHA, GHB, GAB, GCD, GEF, HD, HE, AC, CE, BD, DF) desde una perspectiva substantiva ha permitido llegar a varias conclusiones de interés para simplificarlo. En primer lugar, el relevante papel que juega la covariante Género (G) en el modelo, por la presencia de asociaciones con las variables registradas en el mismo momento temporal A-B, C-D y E-F (asociaciones GAB, GCD y GEF). Las tres asociaciones de segundo orden incluyen las asociaciones de primer orden entre las variables básicas del mismo momento temporal (asociaciones AB, CD y EF). Este hecho refleja que las diferencias de género son importantes para comprender la relación entre situación laboral y cargas familiares durante la primera mitad de la vida laboral.

En segundo lugar, la asociación que se encuentra entre Formación Académica (H) y las variables registradas en el primer momento temporal A-B (asociación HAB), pero no en los momentos temporales subsiguientes. La formación académica parece significativa a la hora de incorporarse por primera vez al mundo laboral y también es importante para comprender las cargas familiares a los 25 años, pero carece de interés en momentos posteriores.

En tercer lugar, la asociación que existe entre las dos covariantes (G y H) con el trabajo a los 25 años (asociación GHA) por un lado, y con las cargas familiares (asociación GHB) por otro.

Y finalmente, la existencia de dos estructuras temporales perfectamente diferenciadas, que involucran los tres registros temporales, de la situación laboral (asociaciones AC y CE) y de las cargas familiares (asociaciones BD y DF). Este es el núcleo central que se revela crucial para comprender la asociación entre variables. La inexistencia de asociación entre momentos con más de una demora temporal en ambos casos sugiere emplear para cada una de tales estructuras un modelo de Markov de primer orden. Sin embargo, mientras que las asociaciones entre formación académica y situación laboral (HA) y cargas familiares (HB) a los 25 años muestran la importancia que la formación tiene en ese punto temporal, no parecen jugar un papel relevante desde una perspectiva teórica las asociaciones HD y HE, sin que lo sean las asociaciones HC y HF, ya que corresponden a momentos temporales diferentes.

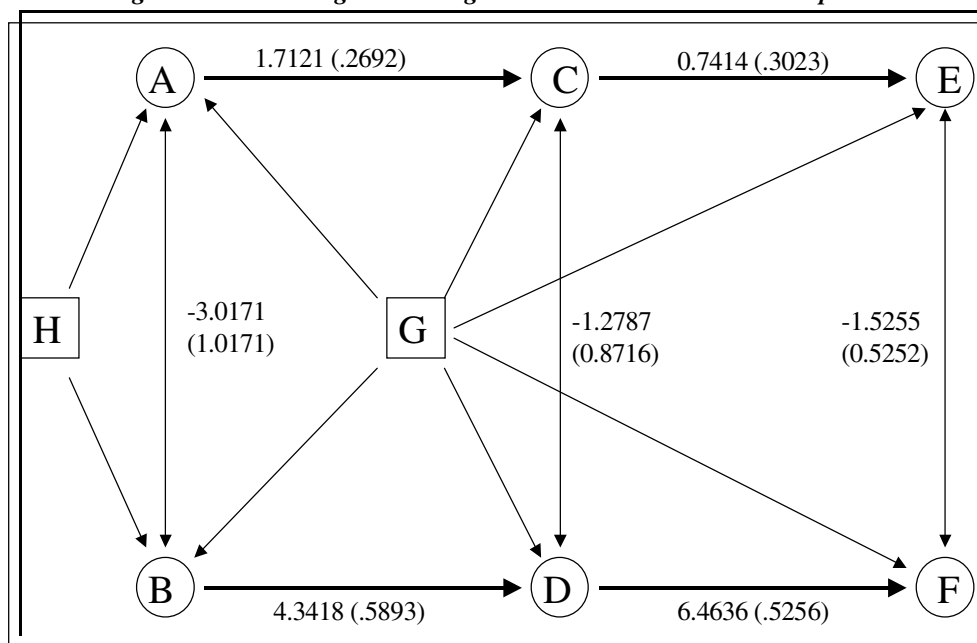
Por tanto, parece adecuado descartar las asociaciones HD y HE. El modelo log-lineal resultante (Modelo M5, véase Tabla 1) se ajusta también de forma óptima, $L^2(348)=202,25$, $P=1,0000$, obteniendo un criterio de información bayesiano relativamente bajo ($BIC=-2347,6$). Aunque la diferencia respecto del modelo M4 es estadísticamente significativa ($L^2(M5|M4)=25,40$, con 5 grados de libertad), la decisión de seleccionar el Modelo M5 se adopta teniendo en cuenta que este modelo es más parsimonioso.

Las estructuras de Markov paralelas y el efecto de las covariantes conducen directamente a postular un modelo causal, tomando información a priori sobre la ordena-

ción causal de las variables. El modelo log-lineal dirigido resultante se compone de una serie de tres submodelos logit para tres diferentes tablas marginales. El flujo adecuado para representar este modelo (Figura 2), utilizando notación marginal con LEM, es el siguiente:

$$\begin{aligned}
 &GH \\
 &AB \mid GH \{GAB \text{ HAB}\} \\
 &CD \mid GAB \{GCD \text{ AC BD}\} \\
 &EF \mid GCD \{GEF \text{ CE DF}\}
 \end{aligned}$$

Figura 2. Modelo log-lineal dirigido con estructuras de Markov paralelas.



H (formación académica) y G (género) son covariantes. A, C y E forman una estructura de Markov para la evaluación (retrospectiva) de la situación laboral y B, D y F forman otra estructura de Markov para la evaluación (retrospectiva) de las cargas familiares. El valor numérico sobre las flechas indica el coeficiente de regresión path y entre paréntesis su error típico correspondiente.

El Modelo M5 incluye un sistema de cuatro submodelos que se estiman de forma independiente. Siendo $\pi_{abcdefgh}$ la probabilidad conjunta de que $G=g, H=h, A=i, B=j, C=k, D=l, E=m$ y $F=n$, se descompone en el siguiente producto de probabilidades marginales y condicionales,

$$\pi_{ghijklmn} = \pi_{gh}^{GH} \pi_{ghij}^{AB|GH} \pi_{gijkl}^{CD|GAB} \pi_{gklmn}^{EF|GCD}$$

En esta formulación se indica que ciertas variables dependen de variables que le preceden en el tiempo, y no de variables registradas con posterioridad. Así, las variables C y D dependen de las variables G,H, A y B, pero no de las variables E y F. Se trata por tanto de un modelo log-lineal dirigido. Además, algunos de los submodelos de que se compone presentan restricciones de ciertos elementos que se suponen no significativos.

En el primer submodelo, la asociación marginal entre las covariantes (asociación GH) no tiene ninguna restricción, y, por tanto, se trata de un submodelo saturado. La parametrización que le corresponde es la siguiente,

$$\pi_{gh}^{GH} = \frac{\exp(\lambda_g^G + \lambda_h^H + \lambda_{gh}^{GH})}{\sum_{gh} \exp(\lambda_g^G + \lambda_h^H + \lambda_{gh}^{GH})}$$

En segundo lugar, el submodelo AB|GH utiliza como variable dependiente la asociación AB y como predictores la asociación GH, que en la formulación logit se considera un componente fijo. De todos los términos posibles, este submodelo restringe el interés a las asociaciones GHA, GHB y GAB, igualando a cero HAB y la asociación conjunta GHAB. Su parametrización como modelo logit se define

$$\pi_{ghab}^{AB|GH} = \frac{\exp(\lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} + \lambda_{ghi}^{GHA} + \lambda_{ghj}^{GHB} + \lambda_{gij}^{GAB})}{\sum_{ij} \exp(\lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} + \lambda_{ghi}^{GHA} + \lambda_{ghj}^{GHB} + \lambda_{gij}^{GAB})}$$

En tercer lugar, el submodelo CD|GAB utiliza como variable dependiente la asociación CD y como predictor la asociación GAB, que en la formulación logit es un componente fijo. Los términos de interés son GCD, AC y BD, restringiendo a cero todos los demás. Su parametrización como modelo logit es

$$\pi_{ijkl}^{CD|GAB} = \frac{\exp(\lambda_k^C + \lambda_l^D + \lambda_{kl}^{CD} + \lambda_{gkl}^{GCD} + \lambda_{ik}^{AC} + \lambda_{jl}^{BD})}{\sum_{kl} \exp(\lambda_k^C + \lambda_l^D + \lambda_{kl}^{CD} + \lambda_{gkl}^{GCD} + \lambda_{ik}^{AC} + \lambda_{jl}^{BD})}$$

Y finalmente, el submodelo EF|GCD utiliza como variable dependiente la asociación EF y como predictor la asociación GCD, que es un componente fijo. El interés se localiza en los términos CEF, CE y DF, restringiendo a cero los demás términos. Su parametrización es

$$\pi_{gklmn}^{EF|GCD} = \frac{\exp(\lambda_m^E + \lambda_n^F + \lambda_{mn}^{EF} + \lambda_{gmn}^{GEF} + \lambda_{km}^{CE} + \lambda_{ln}^{DF})}{\sum_{mn} \exp(\lambda_m^E + \lambda_n^F + \lambda_{mn}^{EF} + \lambda_{gmn}^{GEF} + \lambda_{km}^{CE} + \lambda_{ln}^{DF})}$$

Es importante observar que los dos últimos submodelos han sido restringidos fijando en cero los componentes que no son significativos, para ser consecuente con la

notación del programa LEM. Procediendo de acuerdo con el enfoque original de Goodman (1973), la desviación y los grados de libertad residuales correspondientes a cada uno de los submodelos deben ser definidos como refleja la Tabla 2. El primer submodelo requiere ser especificado sobre la tabla marginal GHAB y produce un ajuste óptimo: $L^2(4)=1,14$; $P=0,8874$. El segundo submodelo debe especificarse sobre la tabla marginal GHABCD y produce también un ajuste aceptable: $L^2(64)=76,78$; $P=0,1314$. En el flujo de LEM este modelo se ha definido sobre una tabla marginal diferente (GABCD) al no encontrarse ninguna asociación con H estadísticamente significativa. Y finalmente, el tercer submodelo debe especificarse sobre la tabla marginal GHABCDEF y produce así mismo un ajuste perfecto: $L^2(280)=124,38$; $P=1,0000$. También este submodelo se ha reducido considerablemente al definirse sobre una tabla marginal más parsimoniosa (GCDEF) en el flujo de LEM. Especificando las tablas marginales de cada submodelo de esta manera, la suma de las desviaciones de los tres submodelos es aproximadamente igual a la desviación global del modelo M5.

Tabla 2. Descomposición de la desviación del modelo M5

Submodelo	Tabla marginal	Desviación	g.l.	Probabilidad
ABIGH {GHA GAB HAB}	GHAB	1,13	4	0,8874
CDIGAB {GCD AC BD}	GHABCD	76,77	64	0,1314
EFIGCD {GEF CE DF}	GHABCDEF	124,36	280	1,0000
Total		202,26	348	1,0000

3.3. Tercera fase: interpretación del modelo final

La Tabla 3 presenta los parámetros del modelo final (M5) que se han estimado mediante el programa LEM (Vermunt, 1997).

Todos ellos se han obtenido utilizando codificación ficticia, tomando como referencia el primer nivel o categoría de cada una de las variables. Los coeficientes *lambda* para el efecto marginal de una variable son logaritmos de razones de probabilidad (*odds*) que comparan el segundo (o tercer) nivel con el primer nivel de la variable categórica manteniendo constantes todas las demás variables. Por ejemplo, representa la disminución, en unidades logarítmicas, de no trabajar respecto de trabajar a los 25 años. Es decir, trabajar a los 25 años se estima que tiene una probabilidad $1/\exp(-4,5894)=1/0,0102 \approx 98$ veces superior que no trabajar. Los coeficientes *lambda* para un efecto marginal de la asociación entre dos variables se interpretan en términos de logaritmos de razones de productos cruzados (*odds ratios*). Así, la asociación entre sexo y situación laboral a los 25 años (asociación GA) es $\lambda_{22}^{GA} = 2,5523$. Para los hombres, la razón entre los que trabajan y los que no trabajan, se estima que es $\exp(2,5523) \approx 12,8$ veces superior que para las mujeres.

Tabla 3. Parámetros del modelo M5

<i>Efecto</i>	<i>Lambda</i>	<i>Std.Error</i>	<i>Z</i>	<i>Exp(lambda)</i>	<i>Wald</i>	<i>g.l.</i>	<i>Prob</i>
Tabla GHAB							
A 2	-4,589	0,5031	-9,123	0,0102	83,23	1	0,000
B 2	-0,6682	0,0876	-7,626	0,5126	58,16	1	0,000
GA 22	2,5523	0,5564	4,587	12,8371	21,04	1	0,000
HA 22	1,9957	0,6097	3,273	7,3573			
HA 32	3,4420	0,5217	6,598	31,2492	53,97	2	0,000
GB 22	0,7901	0,1589	4,973	2,2036	24,73	1	0,000
HB 22	-0,2789	0,1945	-1,434	0,7566			
HB 32	-0,8171	0,1829	-4,467	0,4417	20,17	2	0,000
AB 22	-3,0171	1,0142	-2,975	0,0489	8,85	1	0,003
GHA 222	-1,4118	0,7206	-1,959	0,2437			
GHA 232	-2,7048	0,5681	-4,615	0,0669	23,64	2	0,000
GHB 222	-1,0258	0,4016	-2,555	0,3585			
GHB 232	-0,4101	0,3003	-1,366	0,6636	7,24	2	0,027
GAB 222	4,0518	1,0442	3,880	57,5007	15,06	1	0,000
Tabla GABCD							
C 2	-4,4698	0,7206	-6,203	0,0115	38,48	1	0,000
D 2	1,7091	0,0988	17,295	5,5239	299,11	1	0,000
GC 22	-0,0166	1,0106	-0,016	0,9836	0,00	1	0,987
AC 22	1,7121	0,2692	6,359	5,5407	40,44	1	0,000
GD 22	-1,7876	0,1659	-10,777	0,1674	116,14	1	0,000
BD 22	4,3418	0,5893	7,368	76,8440	54,29	1	0,000
CD 22	-1,2787	0,8716	-1,467	0,2789	2,15	1	0,142
GCD 222	3,7713	1,1365	3,318	43,4355	11,01	1	0,001
Tabla GCDEF							
E 2	-2,5503	0,4648	-5,487	0,0781	30,10	1	0,000
F 2	-0,0959	0,1815	-0,529	0,9085	0,28	1	0,597
GE 22	-0,8951	0,7485	-1,196	0,4086	1,43	1	0,232
CE 22	0,7414	0,3023	2,452	2,0990	6,01	1	0,014
GF 22	-1,5913	0,2956	-5,383	0,2037	28,98	1	0,000
DF 22	6,4636	0,5256	12,298	641,3899	151,23	1	0,000
EF 22	-1,5255	0,5252	-2,905	0,2175	8,44	1	0,004
GEF 222	3,4357	0,8022	4,283	31,0526	18,34	1	0,000

Los parámetros de mayor interés son los que conciernen a la estructura paralela de los dos modelos de Markov que se muestra en la Figura 2. Los parámetros *lambda* y sus errores típicos para los efectos AC, CE, BD y DF y para las asociaciones AB, CD y EF se muestran también en la Figura 2.

3.3.1 Análisis de los modelos de Markov de primer orden

Un modelo de Markov de primer orden con las asociaciones AC y CE muestra el cambio en la situación laboral experimentado durante el primer ciclo de vida laboral activa. El cambio es positivo en ambos casos, aunque de diferente magnitud. Así, a los 35 años la razón entre los que trabajan y los que no trabajan es $\exp(1,7121)=5,5$ veces superior que a los 25 años. Sin embargo, a los 45 años la razón entre los que trabajan y los que no trabajan es $\exp(0,7414)=2,1$ veces superior que a los 35 años. Este moderado descenso sugiere un cierto patrón de estabilidad laboral durante este ciclo.

Similarmente, otro modelo de Markov de primer orden con las asociaciones BD y DF muestra el cambio en la variable carga familiar experimentada durante el primer ciclo de vida laboral activa. A los 35 años, la razón entre no tener cargas familiares y tener cargas familiares es $\exp(4,3418)=77$ veces superior que a los 25 años. Pero a los 45 años, esta razón es $\exp(6,4636)=641$ veces superior que a los 35 años. Estos datos demuestran el importante cambio que se produce durante este ciclo en esta variable.

El paralelismo de los dos modelos queda definido por la existencia de las asociaciones significativas entre situación laboral y cargas familiares en los tres periodos temporales (asociaciones AB, CD y EF, respectivamente). A los 25 años, para los individuos que trabajan la razón entre no tener y tener cargas familiares es $1/\exp(-3,0171)=1/0,0489=20,44$ veces menor que para los que no trabajan. Esta misma relación parece estabilizarse posteriormente, siendo a los 35 años $1/\exp(-1,2787)=1/0,2784=4$ veces menor y a los 45 años $1/\exp(-1,5255)=1/0,2175=4,6$ veces menor.

3.3.2 Análisis de las probabilidades condicionales GH (Tabla 4)

Para el nivel de estudios (H) y el sexo (G), constatamos que en la muestra utilizada el nivel de estudios más alto corresponde a los hombres, aunque hay una mayor probabilidad de hombres con nivel bajo. Mientras las diferencias en probabilidad entre hombre y mujer, con nivel medio, es casi de 4 (0,1137/0,0316) veces, para los demás niveles se mantiene prácticamente igual, por encima de 2.

Tabla 4. Probabilidades de la asociación GH

	Nivel de estudios		
	Bajo	Medio	Alto
Hombres	0,3846	0,1137	0,2124
Mujeres	0,1742	0,0316	0,0835

3.3.3 Análisis de las probabilidades condicionales AB|GH (Tabla 5)

A partir de las probabilidades condicionadas AB dado GH, se observa que las mujeres con nivel de estudios bajo tienen menos probabilidad de trabajar a los 25

años que los hombres a esa misma edad, 0,9932 (0,6566+0,3366) frente a 0,7962 (0,3739+0,4223). La misma situación se da para el nivel medio (0,9480 frente a 0,7500), mientras que en el nivel alto se produce una importante disminución en los hombres, llegándose a igualar con las mujeres que se siguen manteniendo (0,7925 frente a 0,7165). Por otro lado, si no trabajan, los hombres tienen más probabilidad de no tener cargas familiares conforme aumenta su nivel cultural, mientras que en las mujeres esta pauta es independiente del nivel cultural.

Un resultado interesante observado en esta tabla es el que se refiere a la situación “no trabajo y tener cargas familiares” para hombres y mujeres. Para hombres, esta situación tiene probabilidades insignificantes para todos los niveles de estudios (0,0020; 0,0010; 0,0023), mientras que en mujeres esas probabilidades son más importantes (0,1550; 0,1158; 0,1367). Es muy improbable que existan hombres que no trabajen y no tengan cargas familiares, pero no en mujeres, y esta pauta es independiente del nivel de estudios.

Tabla 5. Probabilidades condicionales AB|GH

Situación Laboral 25	Cargas Familiares 25	Nivel de estudios		
		Bajo	Medio	Alto
<i>Hombres</i>				
Trabaja	Sin cargas familiares	0,6566	0,6831	0,6462
Trabaja	Con cargas familiares	0,3366	0,2649	0,1463
No trabaja	Sin cargas familiares	0,0067	0,0511	0,2052
No trabaja	Con cargas familiares	0,002	0,0010	0,0023
<i>Mujeres</i>				
Trabaja	Sin cargas familiares	0,3739	0,5741	0,5383
Trabaja	Con cargas familiares	0,4223	0,1759	0,1782
No trabaja	Sin cargas familiares	0,0488	0,1342	0,1467
No trabaja	Con cargas familiares	0,1550	0,1158	0,1367

3.3.4 Análisis de las probabilidades condicionales CD|GAB (Tabla 6)

Entre los hombres de 35 años hay una alta probabilidad de tener trabajo y tener cargas familiares independientemente de que tuviesen o no trabajo a los 25. Para las mujeres la situación es diferente, pues sólo las que tenían trabajo a los 25 años y tenían cargas familiares, tienen una alta probabilidad de permanecer en esa misma situación (0,8963). La mujer que trabajaba a los 25 años y estaba soltera (sin cargas familiares) seguirá trabajando a los 35 años, pero tiene la misma posibilidad de continuar que de cambiar su situación familiar (de estado civil) (0,4851; 0,4484). Las que no trabajaban a los 25 años y estaban solteras pueden cambiar de situación familiar y laboral a los 35 años, mientras que si tenían cargas familiares tendrán similar probabilidad de tener o no trabajo a los 35 años.

Tabla 6. Probabilidades condicionales CDIGAB

Situación Laboral 35	Situación Laboral y Cargas Familiares 25		Situación Laboral y Cargas Familiares 35		Situación Laboral y Cargas Familiares 35	
	Cargas Familiares	Trabaja Sin cargas familiares	Trabaja Con cargas familiares	No trabaja Sin cargas familiares	No trabaja Con cargas familiares	No trabaja Con cargas familiares
<i>Hombres</i>						
Trabaja	Sin cargas familiares	0,1526	0,0023	0,1496	0,0023	
Trabaja	Con cargas familiares	0,8430	0,9450	0,8263	0,9802	
No trabaja	Sin cargas familiares	0,0017	0,0000	0,0095	0,0001	
No trabaja	Con cargas familiares	0,0027	0,0032	0,0146	0,0173	
<i>Mujeres</i>						
Trabaja	Sin cargas familiares	0,4851	0,0122	0,3725	0,0080	
Trabaja	Con cargas familiares	0,4484	0,8693	0,3444	0,5651	
No trabaja	Sin cargas familiares	0,0055	0,0001	0,0232	0,0005	
No trabaja	Con cargas familiares	0,0611	0,1184	0,2599	0,4264	

3.3.5 Análisis de las probabilidades condicionales E | GCD (Tabla 7)

La relación que se da en los hombres durante el periodo de 35 a 45 años es similar a la que se daba en la situación de paso de 25 y 35 años, en cuanto que independientemente de que se tuviese o no trabajo a los 35 años y de la situación familiar, se tiene una alta probabilidad de tener trabajo a los 45 años, aunque con algunas matizaciones respecto del periodo anterior, pues el soltero (trabaje o no trabaje) tiene igual probabilidad de cambiar de situación familiar.

En las mujeres se produce un importante cambio respecto a la situación anterior, pues las que estaban solteras a los 35 años (tuviesen o no trabajo a los 35 años) tenderán a seguir estando solteras a los 45, y las que trabajaban y tenían cargas familiares seguirán trabajando y en la misma situación. Si trabajaba a los 35 años, se sigue trabajando a los 45 años, y no se cambia respecto a las cargas familiares. Sin embargo, si no se trabajaba a los 35 años, se mantiene la situación familiar, pero con bastantes probabilidades de tener empleo a los 45 años.

4. CONCLUSIONES

En este estudio se ensayan procedimientos metodológicos recientemente propuestos para el análisis temporal de variables categóricas. En concreto, se utiliza un modelo de Markov articulado en un modelado log-lineal dirigido que estudia la estructura temporal de la relación entre situación laboral y cargas familiares. El análisis log-lineal y los procesos de Markov son modelos de análisis de datos categóricos que se

Tabla 7. Probabilidades condicionales EFIGCD

Situación Laboral y Cargas Familiares 35					
Situación Laboral 45	Cargas Familiares 45	Trabaja Sin cargas familiares	Trabaja Con cargas familiares	No trabaja Sin cargas familiares	No trabaja Con cargas familiares
<i>Hombres</i>					
Trabaja	Sin cargas familiares	0,4995	0,0017	0,4751	0,0017
Trabaja	Con cargas familiares	0,4538	0,9815	0,4317	0,9637
No trabaja	Sin Cargas Familiares	0,0390	0,0001	0,0778	0,0003
No trabaja	Con cargas familiares	0,0077	0,0167	0,0514	0,0343
<i>Mujeres</i>					
Trabaja	Sin cargas familiares	0,7957	0,0069	0,7587	0,0058
Trabaja	Con cargas familiares	0,1472	0,8169	0,1385	0,6844
No trabaja	Sin cargas familiares	0,0254	0,0002	0,0501	0,0004
No trabaja	Con cargas familiares	0,0317	0,1760	0,0626	0,3095

han usado en el pasado de forma separada en varias áreas de investigación, tales como la sociología, la psicología y la biología. La virtud del enfoque que se emplea en este trabajo es la fusión de ambos tipos de análisis para la evaluación del cambio longitudinal y transversal.

El resultado fundamental de este estudio es la existencia de una asociación significativa entre situación laboral y cargas familiares en cada uno de los tres periodos estudiados, mediante la utilización de los modelos *path logit* que nos permiten analizar la estructura de la relación entre ambas variables.

Aplicada esta metodología a los datos de la Encuesta de Conciencia, Biografía y estructura de Clase de 1991, se desprenden las siguientes conclusiones:

- El nivel de estudios sólo influye significativamente sobre la situación laboral al comienzo de la vida laboral, a los 25 años, y no posteriormente (a los 35 y 45 años). Dicha asociación es más intensa en el caso de los hombres.
- La probabilidad de tener trabajo es mayor en los hombres que en las mujeres, independientemente de la variable cargas familiares y del momento estudiado, aunque a los 25 años, estas diferencias son mínimas si tienen un nivel de estudios alto.
- A los 25 años, a medida que aumenta el nivel educativo, se incrementa la probabilidad de que el varón sea soltero, siempre que no trabaje, mientras que en la mujer esta situación no depende del nivel de estudios.
- A los 35 años, la probabilidad de cambio tanto en las cargas familiares como en la situación laboral, es mayor en el hombre que en la mujer.
- A los 45 años, es poco probable que la mujer cambie su situación familiar respecto de la que tenía a los 35, mientras que en el hombre esta probabilidad es mayor.

Los resultados de este trabajo reflejan las potencialidades del enfoque empleado y revelan interesantes perspectivas interpretativas de la metodología propuesta. Sin embargo, sería recomendable comparar en trabajos posteriores los resultados obtenidos con los de metodologías alternativas, tales como los modelos de ecuaciones estructurales y los modelos de clase latente.

REFERENCIAS BIBLIOGRÁFICAS

- Agresti, A. (1990): *Categorical Data Analysis*; New York; Wiley.
- Arellano, M. y Bover, O. (1995): "Female labour force participation in the 1980s: The case of Spain", *Investigaciones Económicas*, nº 72, pp.171-194.
- Ato, M. y López, J.J. (1996): *Análisis Estadístico para Datos Categóricos*; Madrid; Síntesis.
- Bishop, Y.V.V.; Fienberg, S.E. y Holland, P.W. (1975): *Discrete Multivariate Analysis. Theory and Practice*; Cambridge, MA; The MIT Press.
- Bover, O. (1997): "Cambios en la composición del empleo y la actividad laboral femenina", *Papeles de Economía Española*, pp. 38-51.
- Carabaña, J.; González, J.J., Serrano, A.; Vallejos, A.; Veredas, S. y Riviere, J. (1992): Informe Técnico de la Encuesta de Estructura, Conciencia y Biografía de Clase. Comunidad de Madrid, Consejería de Economía.
- Cebrián, I y Jimeno, J.F. (1998): "Situación familiar y situación laboral. Determinantes de la actividad laboral de cónyuges e hijos", *Papeles de Economía Española* nº 77, pp.158-170.
- Cebrián, I.; Moreno, G. y Toharia, L. (1997): "Las transiciones de las mujeres casadas en España 1987-1996", *Información Comercial Española* nº 760, pp. 129-143.
- Christensen, R. (1990): *Log-linear models*, New York, NY, Springer-Verlag.
- Goodman, L. (1972): "A general model for the analysis of surveys". *American Journal of Sociology*, 78, pp.1135-1191.
- Goodman, L. (1973): "The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach". *Biometrika*, 60, pp.179-192.
- Hagenaars, J. (1990): *Categorical Longitudinal Data: Log-linear Panel, Trend and Cohort Analysis*, Newbury Park, CA: Sage.
- Hagenaars, J. (1998): "Categorical causal modeling: latent class analysis and directed log-linear models with latent variables". *Sociological Methods and Research*, 26(4), pp.436-486.
- Heinen, T. (1995): "Causal modelling with directed log-linear models and latent variables". Forcina, G.M. Marchetti, R. Hatzinger y G. Galmacci (Eds). *Statistical Modelling: Proceedings of the 11th International Workshop on Statistical Modelling* (pp. 70-79): Orvieto, IT: Graphos.

- Raftery, A. (1995): Bayesian model selection in social research. P.V. Marsden, ed.: *Sociological Methodology 1995*, vol.25, pp.111-163.
- Smith, D. (1995): *Introduction to Graphical Modelling*, New York, NY: Springer-Verlag.
- Toharia, L.; Albert, C.; Cebrián, I.; García Serrano, C.; García Serrano, I.; Malo, M.A.; Moreno, G. y Villagómez, E. (1998): *El Mercado de Trabajo en España*, McGraw-Hill, Madrid.
- Vermunt, J. (1996): *Log-linear Event History Analysis.: A General Approach with Missing Data, Latent Variables and Unobserved Heterogeneity*, Tilburg, The Netherlands: Tilburg University Press.
- Vermunt, J. (1997): *LEM: Log-linear and Event History Analysis with Missing Data Using the EM Algorithm*, 2nd version. Tilburg, The Netherlands.
- Wermuth, N. y Lauritzen, S.L. (1983): "Graphical and recursive models for contingency tables", *Biometrika*, 70, pp.537-552.
- Wermuth, N. y Lauritzen, S.L. (1990): "On substantive research hypothesis, conditional independence graphs and graphical chain models (with discussion)". *Journal of Royal Statistical Society*, B52, pp.21-72.
- Whittaker, J. (1990): *Graphical Models in Applied Multivariate Statistics*, Chichester, UK, Wiley.