

## CAPÍTULO 5

### LOS MODELOS MARKOVIANOS DE VARIABLES LATENTES

JESÚS PÉREZ MAYO  
MIGUEL A. FAJARDO CALDERA  
LIDIA ANDRADES CALDITO  
NURIA CORRALES DIOS

*Grupo de Investigación en Métodos Estadísticos Aplicados a la Economía  
Departamento de Economía Aplicada y Organización de Empresas  
Universidad de Extremadura*

#### RESUMEN

*En este trabajo exponemos una clase de modelos que representan las relaciones entre un conjunto de variables categóricas a lo largo del tiempo, considerando la presencia del error de medida. En consecuencia, se hace preciso introducir la noción de variable latente. En el supuesto de que la movilidad sigue un proceso de Markov de primer orden, se presentan una serie de modelos que calculan el verdadero cambio, la heterogeneidad de la población para el cambio así como una combinación de ambos fenómenos*

#### 1. INTRODUCCIÓN

En el día a día de la investigación, es normal encontrarse con algunas variables no observadas total o parcialmente y, por tanto, es necesario poder considerar el error de medida en la especificación de los modelos estadísticos.

Este problema, el error de medida, ha dado lugar a una serie de modelos, los modelos de estructuras latentes, basados en la idea de la independencia local. Dicho concepto significa el supuesto de que las variables observadas son independientes entre sí para un valor particular de la variable no observada o latente.

Los modelos de estructuras latentes presentan distintas modelizaciones dependiendo del tipo de variables latentes y observadas que tengamos. En el caso que nos ocupa, donde todas las variables son categóricas, el modelo de

estructuras latentes se conoce como modelo de clases latentes (Lazarsfeld, 1950; Lazarsfeld y Henry, 1968; Goodman, 1974; Haberman, 1979).

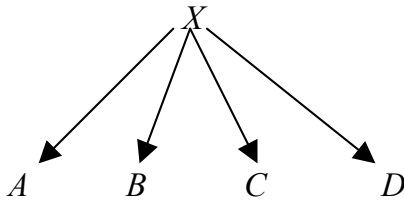
Un modelo de clases latentes, por tanto, se compone de un conjunto de variables cuyos valores se observan directamente y una variable latente no observable directamente.

## 2. EL MODELO DE CLASES LATENTES CLÁSICO

Supongamos un modelo de clases latentes con una variable latente  $X$  y cuatro indicadores o variables observadas  $A$ ,  $B$ ,  $C$  y  $D$ . Además,  $X^*$  es el número de clases latentes y  $A^*$ ,  $B^*$ ,  $C^*$  y  $D^*$ , las categorías de las variables  $A$ ,  $B$ ,  $C$  y  $D$ , respectivamente. La representación gráfica de dicho modelo aparece en la figura 1 y su ecuación básica es

$$\pi_{abcd} = \sum_{x=1}^{X^*} \pi_{xabcd}, \quad (1)$$

donde



$$\pi_{xabcd} = \pi_x \pi_{abcd|x} = \pi_x \pi_{a|x} \pi_{b|x} \pi_{c|x} \pi_{d|x}, \quad (2)$$

Figura 1. Un modelo de clases latentes

La formulación anterior del modelo de clases latentes es la clásica debida a Lazarsfeld (1950), donde  $\pi_{xabcd}$  es la probabilidad conjunta de todas las variables (manifiestas y latente),  $\pi_x$  la probabilidad de pertenecer a la clase latente  $x$  y los distintos términos  $\pi_{i|x}$ , las probabilidades condicionadas de estar en la categoría  $i$  de las respectivas variables, dada la pertenencia a la clase latente  $x$ .

Por tanto, los parámetros del modelo de clases latentes son las probabilidades condicionadas  $\pi_{a|x}$ ,  $\pi_{b|x}$ ,  $\pi_{c|x}$ ,  $\pi_{d|x}$  y las probabilidades de las clases latentes  $\pi_x$ , que estarán sometidas a las siguientes restricciones:

$$\sum_{a=1}^{A^*} \pi_{a|x} = \sum_{b=1}^{B^*} \pi_{b|x} = \sum_{c=1}^{C^*} \pi_{c|x} = \sum_{d=1}^{D^*} \pi_{d|x} = 1; \quad \sum_{x=1}^{X^*} \pi_x = 1 \quad (3)$$

De nuevo, en la ecuación (1) se manifiesta la hipótesis de independencia local, ya que la población se divide en  $X^*$  clases exhaustivas y mutuamente excluyentes por lo que la probabilidad conjunta de las variables observadas se obtiene sumando sobre las clases latentes.

La ecuación (1) se puede expresar mediante un modelo log-lineal (Haberman, 1979). El concepto de independencia local provoca que dicho modelo sea  $\{AX, BX, CX, DX\}$ , cuya expresión es

$$\log m_{abcd} = u_0 + u_x + u_a + u_b + u_c + u_d + u_{xa} + u_{xb} + u_{xc} + u_{xd} \quad (4)$$

donde  $m_{abcd} = N\pi_{abcd}$ .

La ecuación anterior, además de la media general y los términos de una variable, contiene sólo los términos de interacción entre la variable latente  $X$  y las variables manifiestas. Como las variables manifiestas son independientes entre sí dada la clase latente, no aparecen los términos de interacción entre algunas variables observadas.

Finalmente, es posible relacionar los parámetros de las ecuaciones (1) y (4) según un modelo logit (Haberman, 1979).

$$\pi_{a|x} = \frac{\exp(u_a + u_{xa})}{\sum_a \exp(u_a + u_{xa})} \quad (5)$$

## 2.1 La estimación de los modelos

A pesar de ser también un modelo log-lineal, la determinación de las estimaciones máximo-verosímiles de los parámetros de este modelo es más complicada que en el caso donde se observan todas las variables. Se utilizan distintos métodos de estimación, entre los cuales los más conocidos son el algoritmo de Newton-Raphson y el algoritmo EM (Dempster, Laird y Rubin, 1977).

El último es preferible, ya que como el algoritmo IPF, es sencillo tanto en la teoría como en el cálculo. Además, generalmente los valores iniciales elegidos aleatoriamente son suficientes para llegar a una solución. Presenta frente al Newton-Raphson el inconveniente de necesitar más iteraciones para llegar a la solución. Sin embargo, puesto que cada iteración del algoritmo EM es más rápida, este inconveniente no es tan importante.

Al considerar alguna variable latente en el modelo, podemos hablar de datos completos e incompletos. Los primeros se refieren a la distribución de frecuencias de todas las variables, observadas y no observadas, mientras que los segundos son las frecuencias de las variables observadas. No obstante, según la expresión (1), las observaciones  $n_{abcd}$  de la tabla colapsada contienen

información sobre los datos completos  $n_{xabcd}$ . Las ecuaciones (1) y (2) recogen las relaciones entre los datos completos e incompletos.

El algoritmo EM es un procedimiento iterativo y cada iteración está compuesta por dos pasos. En el paso *Esperanza* se calculan todos los valores esperados dados los valores observados y los “actuales” parámetros del modelo. En el paso *Maximización*, se maximiza la función de verosimilitud de todos los datos a partir de los valores esperados calculados en el paso anterior. Esto implica el cálculo de estimaciones actualizadas de los parámetros del modelo como si no faltaran datos, es decir, se utilizan las estimaciones  $\hat{n}_{xabcd}$  como si fueran frecuencias observadas. Para hacerlo, se utilizan los mismos procedimientos en la obtención de las estimaciones máximo-verosímiles de un modelo log-lineal normal: Newton-Raphson e IPF. Las estimaciones obtenidas se utilizan en un nuevo paso *Esperanza* para lograr nuevas estimaciones para las frecuencias de la tabla completa. Las iteraciones continúan hasta que se alcanza la convergencia.

Veremos de forma general cómo funciona este algoritmo y, más tarde, a medida que estudiemos los distintos modelos, expondremos qué expresiones surgen de cada uno de ellos.

De las expresiones (1) y (2) podemos concluir que los datos completos  $n_{xabcd}$  siguen una distribución multinomial  $M(N, \pi_{xabcd})$ . Por tanto, el logaritmo de la función de verosimilitud es

$$\sum_{xabcd} n_{xabcd} \log \pi_{xabcd} \quad (6)$$

Es posible obtener unos estadísticos suficientes para los parámetros si, en primer lugar, descomponemos la probabilidad  $\pi_{xabcd}$  mediante las ecuaciones (1) y (2).

$$\begin{aligned} \sum_{xabcd} n_{xabcd} \log \pi_{xabcd} = & \sum_x n_{x\cdots} \log \pi_x + \sum_{x,a} n_{xa\cdots} \log \pi_{a|x} + \sum_{x,b} n_{x\cdots b} \log \pi_{b|x} + \\ & + \sum_{x,c} n_{x\cdots c} \log \pi_{c|x} + \sum_{x,d} n_{x\cdots d} \log \pi_{d|x} \end{aligned} \quad (7)$$

En el paso *E*, se calculan los valores esperados de las observaciones completas dados los datos observados y los parámetros de la iteración anterior. Al seguir los datos completos una distribución multinomial, la distribución condicionada de dichos datos completos  $n_{xabcd}$  dados los datos observados  $n_{abcd}$  es una distribución multinomial  $M(n_{abcd}, \pi_{xabcd} | \pi_{abcd})$ , la cual implica que

$$E[n_{xabcd} | n_{abcd}, \pi_{xabcd}] = n_{abcd} \frac{\pi_{xabcd}}{\pi_{abcd}} = n_{abcd} \pi_{x|abcd} \quad (8)$$

En consecuencia, en el paso *E* estamos ante:

$$\begin{aligned}
E[n_{x\dots} \mid n_{abcd}, \hat{\pi}_{x|abcd}(p)] &= \sum_{a,b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p) \\
E[n_{xa\dots} \mid n_{abcd}, \hat{\pi}_{x|abcd}(p)] &= \sum_{b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p) \\
E[n_{x\cdot b\dots} \mid n_{abcd}, \hat{\pi}_{x|abcd}(p)] &= \sum_{a,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p) \\
E[n_{x\cdot c\dots} \mid n_{abcd}, \hat{\pi}_{x|abcd}(p)] &= \sum_{a,b,d} n_{abcd} \hat{\pi}_{x|abcd}(p) \\
E[n_{x\dots d} \mid n_{abcd}, \hat{\pi}_{x|abcd}(p)] &= \sum_{a,b,c} n_{abcd} \hat{\pi}_{x|abcd}(p)
\end{aligned}$$

En el paso  $M$  y dada la expresión (8), se determina la estimación del parámetro en la iteración  $p+1$ -ésima dados los estadísticos suficientes de los datos completos en la iteración  $p$ .

$$\begin{aligned}
E[n_{x\dots} \mid \hat{\pi}(p+1)] &= N \hat{\pi}_x(p+1) \\
E[n_{xa\dots} \mid \hat{\pi}(p+1)] &= N \hat{\pi}_x(p+1) \hat{\pi}_{a|x}(p+1) \\
E[n_{x\cdot b\dots} \mid \hat{\pi}(p+1)] &= N \hat{\pi}_x(p+1) \hat{\pi}_{b|x}(p+1) \\
E[n_{x\cdot c\dots} \mid \hat{\pi}(p+1)] &= N \hat{\pi}_x(p+1) \hat{\pi}_{c|x}(p+1) \\
E[n_{x\dots d} \mid \hat{\pi}(p+1)] &= N \hat{\pi}_x(p+1) \hat{\pi}_{d|x}(p+1)
\end{aligned}$$

A partir de estas expresiones, las estimaciones de las probabilidades son

$$\hat{\pi}_x(p+1) = \frac{\sum_{a,b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)}{N}, \quad (9.a)$$

$$\hat{\pi}_{a|x}(p+1) = \frac{\sum_{b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)}, \quad (9.b)$$

$$\hat{\pi}_{b|x}(p+1) = \frac{\sum_{a,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)} \quad (9.c)$$

$$\hat{\pi}_{c|x}(p+1) = \frac{\sum_{a,b,d} n_{abcd} \hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)} \quad (9.d)$$

$$\hat{\pi}_{d|x}(p+1) = \frac{\sum_{a,b,c} n_{abcd} \hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)} \quad (9.e)$$

El proceso de iteración debe continuar hasta que el incremento del logaritmo de la función de verosimilitud sea menor que un valor arbitrario muy pequeño, por ejemplo  $10^{-6}$ . Incluso si las iteraciones se repiten muchas veces, es posible encontrar un óptimo local.

A partir de las ecuaciones (9) es posible calcular las probabilidades

$$\hat{\pi}_{xabcd} \text{ y } \hat{\pi}_{abcd} = \sum_{x=1}^{X^*} \hat{\pi}_{xabcd} . \quad (10)$$

## 2.2 Identificabilidad y contraste del modelo

Los parámetros de los modelos con variables latentes no pueden determinarse de una única forma. La condición, expuesta en los modelos log-lineales, de que el número de parámetros independientes no debe exceder las frecuencias observadas no es suficiente para los modelos con variables latentes.

Según Goodman (1974), una condición suficiente para la identificabilidad local es que la matriz de información sea definida positiva. Mediante tal criterio, es posible resolver el problema de la identificabilidad al utilizar el algoritmo EM. Otra manera de contrastar si un modelo es identificable con el algoritmo EM consiste en estimar el modelo con diferentes valores iniciales. Si con esos valores iniciales distintos, el modelo proporciona el mismo valor del logaritmo de la función de verosimilitud pero distintas estimaciones de los parámetros, el modelo es no identificable.

La calidad del ajuste de un modelo log-lineal concreto puede determinarse con la comparación de las frecuencias observadas,  $n$ , con las esperadas estimadas,  $\hat{m}$ , mediante el contraste de la ji-cuadrado de Pearson y la razón de verosimilitud  $L^2$ , cuyas expresiones son las siguientes para el modelo de las tres variables consideradas en el apartado anterior.

$$\begin{aligned}
X^2 &= \sum_a \sum_b \sum_c \frac{(n_{abc} - \hat{m}_{abc})^2}{\hat{m}_{abc}} \\
L^2 &= 2 \sum_a \sum_b \sum_c n_{abc} \ln \frac{n_{abc}}{\hat{m}_{abc}}
\end{aligned} \tag{11}$$

Si el modelo es válido para la población, ambos estadísticos siguen asintóticamente una distribución ji-cuadrado. Para cada modelo el número de grados de libertad de la distribución se obtiene a partir de la expresión

$gl$  = número de celdas – número de parámetros independientes.

Si algunas frecuencias esperadas estimadas son zeros estructurales o no pueden calcularse algunos parámetros al existir zeros en algunos estadísticos suficientes, Clogg y Eliason (1987) mostraron que la diferencia anterior pasaría a ser

$gl$  = número de celdas sin zeros - número de parámetros estimables.

El estadístico  $L^2$  tiene una ventaja sobre el de Pearson porque puede descomponerse en distintas componentes referidas a diferentes efectos, submodelos o subgrupos. Esta propiedad es muy interesante cuando busquemos un modelo que se ajuste bien y, simultáneamente, sea reducido.

Llegados a este punto, conocemos las herramientas que permiten al investigador determinar en qué medida el modelo propuesto a priori se ajusta o no a los datos observados. Sin embargo, el objetivo es encontrar el mejor modelo, aquél que explica las relaciones existentes entre las variables en la población que generan los datos observados. Por tanto, los errores posibles al seleccionar un modelo se producirán cuando éste contenga más parámetros de los necesarios o se excluyan algunos parámetros que forman parte del mejor modelo.

En el proceso lógico de la modelización estadística, se parte de unas hipótesis o supuestos *a priori* que se reflejan en una formulación determinada del modelo. Dichas hipótesis naturalmente deben basarse en las ideas que el investigador tenga sobre las relaciones existentes entre las variables en la población, es decir, es conveniente utilizar los conceptos teóricos relacionados con el problema que intentemos resolver. Estadísticamente, pueden existir cientos de modelos para un solo conjunto de datos que se ajusten con la misma calidad. Si no seguimos la orientación proporcionada por el problema teórico que queremos resolver, es difícil dilucidar qué modelo elegir.

Si los supuestos de partida lleva a un único modelo log-lineal no saturado, el proceso es fácil, dado que se limitaría a la aplicación de los estadísticos mostrados en la ecuación (11). Sin embargo, como hemos expuesto

anteriormente, la dificultad comienza a la hora de descubrir cuál es el mejor modelo dentro de una gama.

Si los modelos están anidados jerárquicamente, pueden utilizarse contrastes de  $L^2$  condicionados. Dos modelos están anidados jerárquicamente cuando el modelo restringido contiene sólo un subconjunto de los efectos presentes en el modelo libre.

Entonces, el estadístico  $L^2$  de la razón de verosimilitud contrasta la significatividad de los parámetros libres del modelo libre, dado que el modelo libre es cierto para la población.

Utilizando las mismas variables categóricas  $A$ ,  $B$  y  $C$  de la expresión (11), podemos representar el estadístico condicionado  $L^2_{r|l}$ , donde el subíndice  $r$  se refiere al modelo restringido y  $l$  al libre como

$$L^2_{r|l} = L^2_r - L^2_l = 2 \sum_a \sum_b \sum_c n_{abc} \ln \frac{\hat{m}_{abc(l)}}{\hat{m}_{abc(r)}} \quad (12)$$

Los grados de libertad del contraste vienen definidos por el número de parámetros que se fijan en el modelo restringido, es decir, los grados de libertad del modelo restringido menos los del libre.

Como se puede observar en la expresión (12), el estadístico condicionado  $L^2_{r|l}$  puede calcularse como la diferencia de los estadísticos  $L^2$  no condicionados de ambos modelos. Se confirma el comentario anterior sobre los estadísticos  $L^2$  y  $X^2$ , en el que se comentaba la ventaja del primero por su capacidad de descomponerse y así es posible realizar el estadístico condicionado.

Este estadístico condicionado sigue una distribución ji-cuadrado si el modelo libre es válido y la muestra es grande y la aproximación es buena, incluso en aquellas situaciones, como las muestras pequeñas, en que el contraste no condicionado tiene problemas (Haberman, 1978).

Con el estadístico  $L^2_{r|l}$ , se contrasta la hipótesis nula de que el modelo restringido es válido para la población frente a la hipótesis alternativa del libre. Por tanto, es diferente el significado de la aceptación y el rechazo respecto del contraste del estadístico  $L^2$  sin condicionar, ya que en este último la comparación se hace con el modelo saturado, no con otro modelo no saturado.

Podríamos decir que el estadístico  $L^2_{r|l}$  contrasta la validez de las condiciones impuestas al modelo libre para obtener el restringido.

A la hora de elegir el mejor modelo, es preferible utilizar los contrastes condicionados entre dos modelos no saturados frente al test no condicionado del modelo restringido contra el modelo saturado.



A partir de la teoría de la información, es posible desarrollar otra forma de seleccionar el modelo más adecuado. El objetivo no es descubrir el modelo verdadero, sino aquél que proporciona mayor información sobre la realidad. Por un lado, las frecuencias esperadas estimadas deben ser parecidas a las observadas y, por otro, el modelo debe ser tan reducido como sea posible.

Los contrastes más conocidos basados en la teoría de la información son el *criterio de información de Akaike* (AIC) (Akaike, 1987) y el *criterio de información bayesiano* (BIC) (Raftery, 1986). El primero, penalizando al modelo según su grado de complejidad, determina hasta qué punto un modelo concreto se desvía de la realidad y su expresión es

$$AIC = -2 \log \ell + 2 npar, \quad (13)$$

donde  $\ell$  representa la función de verosimilitud y  $npar$  el número de parámetros desconocidos.

Raftery (1986) desarrolló el AIC dentro del contexto de los modelos log-lineales y propuso el BIC que puede calcularse como

$$BIC = -2 \log \ell + (\log N) npar \quad (14)$$

Cuanto menores sean los valores de los criterios, mejor será el modelo porque mayor información contendrá.

Ambos criterios pueden calcularse a partir del estadístico  $L^2$  de la siguiente forma

$$AIC = L^2 - 2 gl; \quad BIC = L^2 - (\log N) gl, \quad (15)$$

que, como se puede observar, son mucho más sencillas que las anteriores y consisten en la comparación con los respectivos criterios para el modelo saturado.

Por tanto, y refiriéndonos al BIC al ser el más adecuado en los modelos log-lineales, se podrá calcular el criterio BIC a partir del estadístico  $L^2$  no condicionado. Un valor negativo indica que el modelo es preferible al modelo saturado y además, debe elegirse aquel modelo con el menor valor. Este criterio elimina los problemas del ajuste por exceso y por defecto. Un modelo que no se ajuste bien a las frecuencias observadas tendrá un  $L^2$  elevado y, en consecuencia, incrementará el primer término de la diferencia que hará poco probable seleccionarlo. Por otra parte, si un modelo se ajusta muy bien porque posee un gran número de parámetros, al tener una cantidad muy pequeña de grados de libertad, de nuevo, el valor del criterio será muy elevado.

Asimismo, una ventaja del BIC se presenta ante las muestras muy grandes (como en la aplicación empírica que nos ocupa). Un tamaño muestral muy elevado provoca que todos los efectos del modelo saturado sean significativos y, por lo tanto, puede llevar al rechazo de los modelos no saturados si la calidad del ajuste se prueba según los estadísticos usuales  $X^2$  y  $L^2$ .

En este caso, el número de grados de libertad necesarios para calcular cualquiera de las pruebas se determina como

$$gl = C^M - 1 - \left[ (S^* - 1) - MS^*(C - 1) \right] \quad (16)$$

donde  $C$  es el número de categorías de las variables observadas y  $M$  la cantidad de variables observadas.

### 2.3 Extensiones del modelo clásico

En este apartado se contemplan aquellos modelos que superan el marco y las limitaciones del modelo clásico, es decir, varias variables observadas independientes entre sí dadas las clases de una única variable latente. Todos ellos, sin embargo, comparten la característica de ser estáticos. A lo largo del resto del capítulo, desarrollaremos sus contrapartidas en el análisis dinámico y comentaremos cómo incorporar cada una de sus particularidades dentro de un modelo dinámico con variables latentes.

#### 2.3.1 Modelos con más de una variable latente

La especificación de los modelos de clase latente con más de una variable latente se propuso por Goodman (1974) y Haberman (1979). Supongamos un modelo con dos variables latentes,  $X$  e  $Y$ , con cuatro indicadores,  $A$  y  $B$  para  $X$  y  $C$  y  $D$  para  $Y$ . Además, supongamos que ambas variables latentes están relacionadas como aparece en la figura 2.

La expresión de dicho modelo será la siguiente,

$$\pi_{xyabcd} = \pi_{xy} \pi_{a|x} \pi_{b|x} \pi_{c|y} \pi_{d|y} \quad (17)$$

que se corresponde con el modelo log-lineal  $\{XY, XA, XB, YC, YD\}$ , es decir,

$$\log m_{xyabcd} = u + u_x + u_y + u_a + u_b + u_c + u_d + u_{xy} + u_{xa} + u_{xb} + u_{yc} + u_{yd} \quad (18)$$

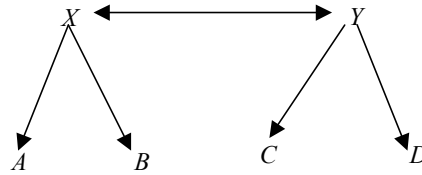


Figura 2. El modelo de clases latentes con 2 variables latentes

De nuevo, es posible imponer distintos tipos de restricciones sobre los parámetros del modelo log-lineal o las probabilidades condicionadas.

### 2.3.2 Modelos con variables exógenas

A las variables exógenas introducidas en el modelo (Hagenaars, 1990), dado que no son indicadores, no se les exige la condición de independencia local dada la variable latente.

Clogg (1981) propuso una especificación de un modelo de este tipo similar a los modelos con múltiples indicadores y múltiples causas (MIMIC). Por ejemplo, supongamos una variable latente  $X$  con las variables  $A$ ,  $B$  y  $C$  como indicadores y dos variables exógenas  $M$  y  $N$ , relacionadas como se observa en la figura 3.

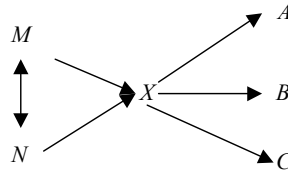


Figura 3. Un modelo de clases latentes con variables externas

En el modelo representado por la figura anterior, están relacionadas las variables exógenas, ambas influyen en la variable latente sin que exista un término de interacción entre las tres, de forma que se podría expresar como el modelo log-lineal  $\{MN, MX, NX, XA, XB, XC\}$  cuya ecuación es

$$\log m_{xabc mn} = u + u_m + u_n + u_x + u_a + u_b + u_c + u_{mn} + u_{mx} + u_{nx} + u_{xa} + u_{xb} + u_{xc} \quad (19)$$

Dicho modelo se puede expresar como probabilidades condicionadas de la siguiente manera.

$$\pi_{xabc mn} = \pi_{mn} \pi_{x|mn} \pi_{a|x} \pi_{b|x} \pi_{c|x} \quad (20)$$

Finalmente, aunque en este ejemplo las variables exógenas son explicativas, esta cualidad no es obligatoria. Es posible construir un modelo, como se verá en los modelos log-lineales causales con variables latentes, donde una variable exógena sea dependiente.

### 2.3.3 Modelos para varias subpoblaciones

En esta extensión se contempla la existencia de distintos grupos en la población. En consecuencia, la distribución de la variable latente y su relación con las variables indicadores pueden variar para cada grupo o subpoblación. Clogg y Goodman (1984, 1985) presentaron los principios para desarrollar análisis latentes simultáneos para los distintos grupos utilizando la

parametrización clásica del modelo de clases latentes. No obstante, el análisis latente simultáneo para varias subpoblaciones se puede entender como un modelo de clases latentes con variables externas puesto que implica la existencia de una variable relativa al grupo. Dicha variable, como hemos dicho antes, puede afectar a la variable latente y las probabilidades condicionadas de respuesta. Si suponemos una variable de grupo  $G$ , una latente  $X$  y tres indicadores  $A$ ,  $B$  y  $C$ , el modelo podría escribirse como

$$\pi_{xabcg} = \pi_g \pi_{x|g} \pi_{a|xg} \pi_{b|xg} \pi_{c|xg} \quad (21)$$

Este modelo se puede expresar en forma log-lineal como  $\{GXA, GXB, GXC\}$  que también refleja todos los efectos de la variable grupo que, en definitiva, es una variable exógena o covariable. Se pueden imponer distintas restricciones al modelo de la ecuación (20) para considerar distintos efectos de la variable grupo, restricciones que se reflejarían en distintos efectos de interacción para el modelo log-lineal que representa la relación entre las variables.

### 3. LOS MODELOS DINÁMICOS DE VARIABLES LATENTES

En este caso se habla de modelos dinámicos en general y no, de tablas bidimensionales y multidimensionales en particular, puesto que la introducción de una variable latente en una tabla de movilidad bidimensional es un caso muy parecido al modelo clásico de clases latentes.

Esta cuestión, la incorporación del tiempo en el análisis, no puede resolverse mediante las extensiones del modelo clásico de clases latentes arriba comentadas. La razón reside en la imposibilidad de especificar un orden causal *a priori* entre las variables. No obstante, se verá cómo pueden expresarse como unos modelos *modified path* donde una o algunas de las variables no son observadas, esto es, latentes.

En primer lugar, se expondrá la teoría relacionada con los modelos log-lineales causales con variables latentes. Hagenaars (1990) mostró cómo especificar al mismo tiempo un sistema de modelos logit para un conjunto de variables latentes y observadas ordenadas causalmente y un modelo de clases latentes para las variables latentes utilizadas en el sistema anterior. Puesto que el modelo general propuesto por él consta de una parte estructural y otra de medida análogamente al modelo LISREL (Jöreskog y Sörbom, 1988) para variables continuas, lo llamó *modelo LISREL modificado*.

El modelo representado en la figura siguiente es una versión del modelo de caminos modificado, donde la variable  $X$  se considera latente y medida a través de las variables  $D$ ,  $E$  y  $F$ . Este modelo no sólo estima la fuerza de la asociación

entre las variables, sino también incorpora alguna información *a priori* sobre el orden causal de las variables.

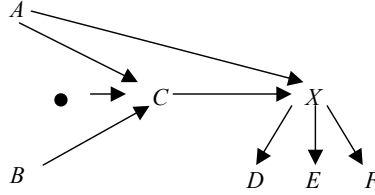


Figura 4. Un modelo LISREL modificado

En este caso concreto, la variable latente  $X$  se considera posterior a  $C$  y ésta posterior a las variables  $A$  y  $B$ . No obstante, es posible encontrar más de una variable latente. Supongamos un modelo en el que se especifican las relaciones causales entre varias variables categóricas  $A, B, C, X$  e  $Y$ , donde las dos últimas son latentes.  $A$  y  $B$  son exógenas, mientras que el resto son endógenas. Además se supone que  $C$  es posterior a  $X$  e  $Y$  posterior a  $C$ . Finalmente, las variables  $D$  y  $E$  son indicadores de la variable latente  $X$  y  $F$  y  $G$  de la variable  $Y$ .

Dicho modelo aparece reflejado en la figura siguiente.

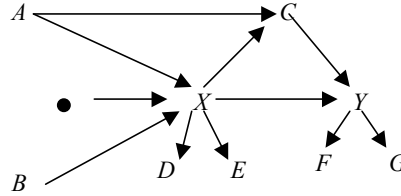


Figura 5. Un modelo LISREL modificado con más de una variable latente

La probabilidad de pertenecer a una celda determinada de la probabilidad conjunta de todas las variables se representa por  $\pi_{abxcydefg}$ . Es posible utilizar la información sobre el orden causal para descomponer la probabilidad anterior como un producto de distintas probabilidades marginales y condicionadas.

$$\pi_{abxcydefg} = \pi_{ab} \pi_{x|ab} \pi_{c|abx} \pi_{y|abxc} \pi_{defg|xy} \quad (22)$$

Son muy visibles las similitudes con un modelo *modified path*. La diferencia es la última probabilidad condicionada que recoge la relación entre las variables latentes y sus indicadores, esto es, la parte de medida del modelo. El resto de las probabilidades representan la parte estructural del modelo. Por lo tanto, de nuevo aparece esta división del modelo en dos submodelos de manera análoga al modelo LISREL.

Sin embargo, el modelo debe mejorarse ya que la expresión anterior es un modelo saturado donde una variable dependiente depende de todas las variables anteriores, incluso de todas sus interacciones.

La manera más sencilla de hacerlo es, como se vio en el capítulo anterior, utilizar la información proporcionada por el grafo del modelo, en este caso, la figura 5. Por ejemplo, en ella se puede observar que la variable  $C$  depende de las variables  $A$  y  $X$ , pero no de  $B$ . En consecuencia, la expresión (22) se podría escribir como

$$\pi_{abxcydefg} = \underbrace{\pi_{ab} \pi_{x|ab} \pi_{c|ax} \pi_{y|axc}}_{\text{estructural}} \underbrace{\pi_{d|x} \pi_{e|x} \pi_{f|y} \pi_{g|y}}_{\text{medida}} \quad (23)$$

Puesto que el modelo LISREL modificado de Hagenaars es una extensión del modelo de caminos modificado de Goodman, se puede expresar como el siguiente sistema recursivo de modelos logit.

$$\begin{aligned} \pi_{ab} &= \frac{\exp(u_a + u_b)}{\sum_{ab} \exp(u_a + u_b)}, \\ \pi_{x|ab} &= \frac{\exp(u_x + u_{ax} + u_{bx} + u_{abx})}{\sum_x \exp(u_x + u_{ax} + u_{bx} + u_{abx})}, \\ \pi_{c|ax} &= \frac{\exp(u_c + u_{ac} + u_{xc})}{\sum_c \exp(u_c + u_{ac} + u_{xc})}, \\ \pi_{y|cx} &= \frac{\exp(u_y + u_{xy} + u_{cy})}{\sum_c \exp(u_y + u_{xy} + u_{cy})}, \\ \pi_{d|x} &= \frac{\exp(u_d + u_{dx})}{\sum_d \exp(u_d + u_{dx})}, \\ \pi_{e|x} &= \frac{\exp(u_e + u_{ex})}{\sum_e \exp(u_e + u_{ex})}, \\ \pi_{f|y} &= \frac{\exp(u_f + u_{fy})}{\sum_f \exp(u_f + u_{fy})}, \\ \pi_{g|y} &= \frac{\exp(u_g + u_{gy})}{\sum_g \exp(u_g + u_{gy})}. \end{aligned}$$

Si suponemos fija la distribución marginal de las variables independientes, cada uno de los anteriores modelos logit se puede expresar como un modelo log-lineal para cada una de las distintas tablas marginales del problema.

$\log m_{ab} = u + u_a + u_b$ , para la tabla marginal  $AB$

$\log m_{abx} = \alpha_{ab} + u_{ax} + u_{bx} + u_{abx}$ , para la tabla marginal  $ABX$

$\log m_{acx} = \alpha_{ac} + u_{ax} + u_{xc}$ , para la tabla marginal  $ACX$

$\log m_{cxy} = \alpha_{cx} + u_y + u_{cy} + u_{xy}$ , para la tabla marginal  $CXY$

$\log m_{xde} = u + u_x + u_d + u_e + u_{dx} + u_{ex}$ , para la tabla marginal  $XDE$

$\log m_{yfg} = u + u_y + u_f + u_g + u_{yf} + u_{yg}$ , para la tabla marginal  $YFG$ .

Como se puede observar, tanto en las tablas como en los modelos, sólo se han considerados aquellas variables precedentes de la que depende una variable concreta. La reducción de la cantidad de las variables condicionantes hace que la estimación sea más eficiente.

Dicha reducción provoca que no sea posible calcular la razón de verosimilitud  $L^2$  del modelo completo como la suma de las respectivas razones de cada modelo logit tal y como propuso Goodman. No obstante, ya fue comentado en el capítulo anterior que este problema se subsana fácilmente calculando las frecuencias esperadas según el modelo completo y comparándolas con las observadas.

### 3.1 Algunos modelos log-lineales causales con variables latentes aplicables a las tablas de movilidad

El análisis de una tabla de movilidad presenta la peculiaridad de que se estudia la misma variable en, al menos, dos momentos de tiempo distintos. En consecuencia, los modelos que reflejan el cambio son más sencillos que los estudiados hasta ahora en este capítulo. Estudiamos la dependencia temporal, la autocorrelación, en términos econométricos.

La consideración de las variables latentes en este contexto puede deberse a distintas causas. En primer lugar, quizás la variable cuya movilidad o cambio se estudia es una variable no observable (por ejemplo, la calidad de vida o el bienestar). Por tanto, es necesario utilizar una o más variables observadas como indicadores de las latentes.

Un fenómeno que también implica la introducción de las variables latentes es el error de medida. Es decir, se analiza el cambio de una variable observada, cambio formado por un componente real y otro espurio debido a los errores de repuesta. Finalmente, las variables latentes ayudan a reconocer la heterogeneidad de la población respecto de la movilidad. En este caso, la variable latente divide la población en grupos homogéneos para el cambio, esto es, con matrices de movilidad comunes.

### 3.1.1 El modelo latente de Markov

El error de medida, razón de existencia de los modelos de clases latentes, atenúa la asociación existente entre las variables. Esto significa que la relación entre dos variables observadas sujetas a algún tipo de error de medida será generalmente más débil que la relación real. Cuando se analiza la movilidad, este hecho implica que la fuerza de las relaciones entre las categorías realmente ocupadas en dos momentos de tiempo será estimada por defecto, o, en otras palabras, la magnitud de la movilidad será estimada por defecto cuando las categorías observadas sufran algún error. Cuando se produce este hecho, las transiciones observadas son, de hecho, una mezcla de la movilidad real y cambio espurio resultante del error de medida (Van de Pol y De Leeuw, 1986; Hagenaars, 1992).

Como ejemplos de los modelos latentes para analizar el cambio, se presentan los dos modelos de la figura 6.

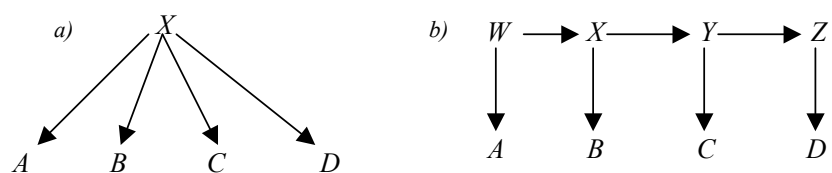


Figura 6. Modelos dinámicos de variables latentes

El modelo a) refleja aquella situación donde todo el cambio observado se debe al error de medida puesto que sólo existe una variable latente, es decir, no existe cambio latente o real. Por otro lado, en el modelo b) se observa una estructura donde se produce un cambio latente analizado mediante las relaciones entre las variables no observadas  $W$ ,  $X$ ,  $Y$  y  $Z$  medidas por las variables indicadores  $A$ ,  $B$ ,  $C$  y  $D$ . Además, tenemos que cada variable manifiesta está relacionada con una y sólo una variable latente y cada una de éstas se asocia sólo con la variable latente que recoge el estado real en el período anterior.

Este último modelo es la representación gráfica de un modelo latente de Markov (Wiggins, 1973; Poulsen, 1982) con un único indicador por ocasión. Dicho modelo presenta problemas de identificabilidad que se pueden resolver introduciendo al menos un indicador más o imponiendo restricciones adicionales a las relaciones del modelo. Algunas restricciones posibles son la estacionariedad de la matriz de transición, la independencia de los valores observados en diferentes momentos dados los latentes, la dependencia del error de medida sólo respecto del valor latente actual y no de los previos, entre otras.



Para simplificar la exposición, supongamos una única variable indicador por ocasión. Supongamos  $t$  valores consecutivos  $X_1, X_2, \dots, X_t$  de una misma variable discreta  $X$  observada con  $X^*$  categorías. Además, se supone un vector  $p$  de probabilidades iniciales y un conjunto de matrices de transición  $T_t$  para cada período. Al ser un modelo de Markov de primer orden, cada matriz de transición  $T$  recoge las probabilidades condicionadas de pertenecer a la categoría  $j$  de la variable  $X_t$  dada la pertenencia al estado  $j$  de la variable  $X_{t-1}$ .

Además, se introduce una variable discreta latente  $Y$  presente en cada uno de los períodos. No es necesario que el número  $Y^*$  de estados latentes coincida con la cantidad  $X^*$  de categorías observadas.

La estructura de las relaciones entre las distintas variables latentes y observadas para tres momentos de tiempo sigue la asociación representada en la figura 7.

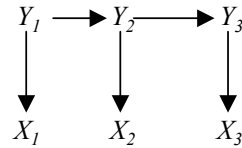


Figura 7. Un modelo latente de Markov para tres momentos de tiempo

En consecuencia, el vector que recoge la distribución de probabilidad de la variable observada  $X$ ,  $p_t$ , depende de la distribución de probabilidad de la variable latente  $Y$ , el vector  $\delta_t$ , y de la matriz  $R_t$  de las probabilidades condicionadas de respuesta o de fiabilidad  $\rho_{x_i|y_i}$ . Estas últimas son las probabilidades de pertenecer a una categoría de la variable  $X$  dado un estado latente  $Y$ . Por tanto, podemos expresar la distribución observada de probabilidad en el período  $t$  como

$$p_t = \delta_t R_t \quad (24)$$

En el diagrama siguiente, aparecen los parámetros que se estiman si se considera que las variables latentes y observadas son dicotómicas. Se puede observar que el cambio se realiza en la parte latente (estructural) mediante las matrices de transición y que las variables latentes se reflejan en las variables manifiestas a través de las matrices de fiabilidad.

Por lo tanto, los parámetros que hay que estimar a partir de los datos observados son:

- La probabilidad inicial  $\delta_{y_1}$  de pertenecer a cada una de las  $Y^*$  clases latentes en el momento inicial.
- Las probabilidades condicionadas de respuesta  $\rho_{x_i|y_i}$  de estar en cada una de las categorías  $x_i$  observadas dado el estado latente  $y_i$  en el momento  $t$ .

- Las probabilidades de transición latente  $\tau_{y_t|y_{t-1}}$  de pasar de cada clase latente  $y_{t-1}$  en el momento t-1 a la clase latente  $y_t$  en el momento t.

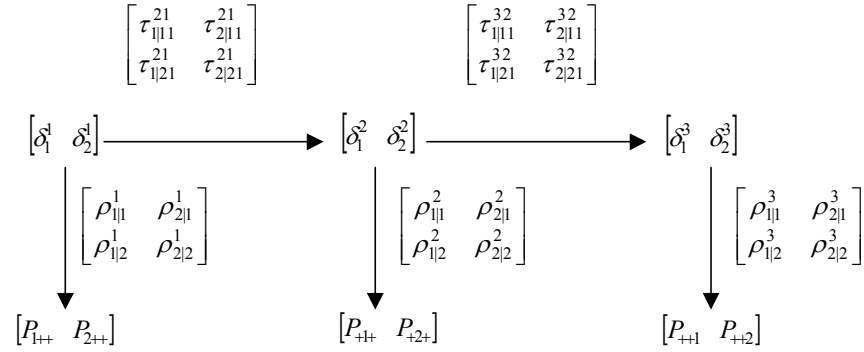


Figura 8. Relaciones y parámetros que estimar en un modelo latente de Markov dicotómico en tres periodos

La probabilidad conjunta de pertenecer a una celda determinada de la distribución según este modelo sería<sup>1</sup>:

$$\pi_{y_1 y_2 y_3 x_1 x_2 x_3} = \delta_{y_1} \rho_{x_1|y_1} \tau_{y_2|y_1} \rho_{x_2|y_2} \tau_{y_3|y_2} \rho_{x_3|y_3} \quad (25)$$

Sin embargo, sólo se conocen los datos observados  $p_{x_1 x_2 x_3}$  y, por tanto, nos enfrentamos al problema de estimar estados no observables. Es decir, sólo las probabilidades de las variables observadas tienen una contrapartida empírica para inferir estadísticamente. No obstante, colapsando sobre las variables latentes la expresión (25), podemos obtener una ecuación que relacione la probabilidad de las variables observadas con el producto de probabilidades condicionadas respecto a las latentes.

$$\pi_{x_1 x_2 x_3} = \sum_{y_1=1}^{y_1^*} \sum_{y_2=1}^{y_2^*} \sum_{y_3=1}^{y_3^*} \delta_{y_1} \rho_{x_1|y_1} \tau_{y_2|y_1} \rho_{x_2|y_2} \tau_{y_3|y_2} \rho_{x_3|y_3} \quad (26)$$

Por consiguiente, con una muestra aleatoria de N individuos en el panel, podemos suponer que el modelo latente de Markov sigue una distribución multinomial paramétrica, cuyo parámetro de probabilidad viene dado por la ecuación (26).

<sup>1</sup> Se ha utilizado la notación propuesta por los autores que han trabajado sobre estos modelos. Creemos que una letra distinta para cada tipo de probabilidad facilita la comprensión del modelo. Siguiendo la notación usada en el trabajo hasta ahora, la ecuación (25) sería

$$\pi_{y_1 y_2 y_3 x_1 x_2 x_3} = \pi_{y_1} \pi_{x_1|y_1} \pi_{y_2|y_1} \pi_{x_2|y_2} \pi_{y_3|y_2} \pi_{x_3|y_3}$$

Estimación del modelo latente de Markov

Las estimaciones máximo-verosímiles de los parámetros del modelo se obtienen mediante el algoritmo EM, expuesto anteriormente.

En primer lugar, para poder aplicar el algoritmo, buscaremos los estadísticos suficientes de los datos completos para los parámetros del modelo. Partimos del logaritmo de la función de verosimilitud según la expresión (25).

$$\begin{aligned} \sum_{y_1, y_2, y_3} \sum_{x_1, x_2, x_3} n_{y_1, y_2, y_3, x_1, x_2, x_3} \log \pi_{y_1, y_2, y_3, x_1, x_2, x_3} &= \sum_{y_1} n_{y_1, \dots} \log \delta_{y_1} + \sum_{x_1, y_1} n_{y_1, \dots, x_1} \log \rho_{x_1 | y_1} + \\ &+ \sum_{y_1, y_2} n_{y_1, y_2, \dots} \log \tau_{y_2 | y_1} + \sum_{x_2, y_2} n_{y_2, \dots, x_2} \log \rho_{x_2 | y_2} + \sum_{y_2, y_3} n_{y_2, y_3, \dots} \log \tau_{y_3 | y_2} + \\ &+ \sum_{x_3, y_3} n_{y_3, \dots, x_3} \log \rho_{x_3 | y_3} \end{aligned}$$

Como para que el modelo sea identificable se impone la restricción de que las probabilidades de respuesta sean iguales en cada momento, podemos escribir la expresión anterior como

$$\begin{aligned} \sum_{y_1, y_2, y_3} \sum_{x_1, x_2, x_3} n_{y_1, y_2, y_3, x_1, x_2, x_3} \log \pi_{y_1, y_2, y_3, x_1, x_2, x_3} &= \sum_{y_1} n_{y_1, \dots} \log \delta_{y_1} + \\ &+ \sum_{x, y} (n_{y, \dots, x} + n_{y, \dots, x} + n_{y, \dots, x}) \log \rho_{x | y} + \sum_{y_1, y_2} n_{y_1, y_2, \dots} \log \tau_{y_2 | y_1} + \sum_{y_2, y_3} n_{y_2, y_3, \dots} \log \tau_{y_3 | y_2} \quad (27) \end{aligned}$$

A partir de esta expresión es posible obtener los estadísticos suficientes para los parámetros.

Una vez hecho esto, ya puede aplicarse el algoritmo EM. En primer lugar, se expone el paso E<sup>2</sup>.

$$\begin{aligned} E[n_{y_1, \dots} | \hat{\pi}(p)] &= \sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_1 | x_1, x_2, x_3}(p) \\ E[n_{y, \dots, x} + n_{y, \dots, x} + n_{y, \dots, x} | \hat{\pi}(p)] &= \sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_1 | x_1, x_2, x_3}(p) + \\ &+ \sum_{x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_2 | x_1, x_2, x_3}(p) + \sum_{x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_3 | x_1, x_2, x_3}(p) \\ E[n_{y_1, y_2, \dots} | \hat{\pi}(p)] &= \sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_1, y_2 | x_1, x_2, x_3}(p) \\ E[n_{y_2, y_3, \dots} | \hat{\pi}(p)] &= \sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_2, y_3 | x_1, x_2, x_3}(p), \end{aligned}$$

donde

<sup>2</sup> El término  $(p)$  que aparece junto a las distintas probabilidades refleja la iteración  $p$ -ésima del algoritmo.

$$\begin{aligned}\hat{\pi}_{y_1|x_1x_2x_3}(p) &= \frac{\hat{\pi}_{y_1 \cdots x_1x_2x_3}(p)}{\hat{\pi}_{\cdots x_1x_2x_3}(p)}, \\ \hat{\pi}_{y_2|x_1x_2x_3}(p) &= \frac{\hat{\pi}_{\cdot y_2 \cdots x_1x_2x_3}(p)}{\hat{\pi}_{\cdots x_1x_2x_3}(p)}, \\ \hat{\pi}_{y_3|x_1x_2x_3}(p) &= \frac{\hat{\pi}_{\cdots y_3 \cdots x_1x_2x_3}(p)}{\hat{\pi}_{\cdots x_1x_2x_3}(p)}, \\ \hat{\pi}_{y_1y_2|x_1x_2x_3}(p) &= \frac{\hat{\pi}_{y_1y_2 \cdots x_1x_2x_3}(p)}{\hat{\pi}_{\cdots x_1x_2x_3}(p)} \text{ y} \\ \hat{\pi}_{y_2y_3|x_1x_2x_3}(p) &= \frac{\hat{\pi}_{\cdot y_2y_3 \cdots x_1x_2x_3}(p)}{\hat{\pi}_{\cdots x_1x_2x_3}(p)}\end{aligned}$$

Una vez obtenidas las estimaciones de los parámetros del modelo en el paso  $p$ -ésimo, se determinan las probabilidades estimadas que maximizan la verosimilitud en la etapa *Maximización*.

$$\begin{aligned}E[n_{y_1 \cdots} \mid N, \hat{\pi}(p+1)] &= N\delta_{y_1}(p+1), \\ E[n_{y_1 \cdots x_1 \cdots} \mid N, \hat{\pi}(p+1)] &= N\delta_{y_1}(p+1)\rho_{x|y}(p+1), \\ E[n_{\cdot y_2 \cdots x_2 \cdot} \mid N, \hat{\pi}(p+1)] &= N\rho_{x|y}(p+1) \sum_{y_1} \delta_{y_1} \tau_{y_2|y_1}(p+1), \\ E[n_{\cdots y_3 \cdots x_3} \mid N, \hat{\pi}(p+1)] &= N\rho_{x|y}(p+1) \sum_{y_1, y_2} \delta_{y_1}(p+1) \tau_{y_2|y_1}(p+1) \tau_{y_3|y_2}(p+1), \\ E[n_{y_1 \cdots x_1 \cdots} + n_{\cdot y_2 \cdots x_2 \cdot} + n_{\cdots y_3 \cdots x_3} \mid N, \hat{\pi}(p+1)] &= N\rho_{x|y}(p+1)(\delta_{y_1}(p+1) + \\ &+ \sum_{y_1} \delta_{y_1}(p+1) \tau_{y_2|y_1}(p+1) + \sum_{y_1, y_2} \delta_{y_1}(p+1) \tau_{y_2|y_1}(p+1) \tau_{y_3|y_2}(p+1)) \\ E[n_{y_1y_2 \cdots} \mid N, \hat{\pi}(p+1)] &= N\delta_{y_1}(p+1) \tau_{y_2|y_1}(p+1) \\ E[n_{\cdot y_2y_3 \cdots} \mid N, \hat{\pi}(p+1)] &= N\tau_{y_3|y_2}(p+1) \sum_{y_1} \delta_{y_1} \tau_{y_2|y_1}(p+1)\end{aligned}$$

Partiendo de las expresiones anteriores, podemos encontrar las estimaciones  $p+1$ -ésimas de los parámetros que maximizan la función de verosimilitud dadas las probabilidades de la iteración anterior.

$$\hat{\delta}_{y_1}(p+1) = \frac{\sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_1 | x_1, x_2, x_3}(p)}{N} \quad (28.a)$$

$$\hat{\rho}_{x|y}(p+1) = \frac{\sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} (\hat{\pi}_{y_1 | x_1, x_2, x_3}(p) + \hat{\pi}_{y_2 | x_1, x_2, x_3}(p) + \hat{\pi}_{y_3 | x_1, x_2, x_3}(p))}{N \rho_{x|y} \left( \delta_{y_1} + \sum_{y_1} \delta_{y_1} \tau_{y_2 | y_1} + \sum_{y_1, y_2} \delta_{y_1} \tau_{y_2 | y_1} \tau_{y_3 | y_2} \right)} \quad (28.b)$$

$$\hat{\tau}_{y_2 | y_1}(p+1) = \frac{\sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_1 y_2 | x_1, x_2, x_3}(p)}{\sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_1 | x_1, x_2, x_3}(p)} \quad (28.c)$$

$$\hat{\tau}_{y_3 | y_2}(p+1) = \frac{\sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_2 y_3 | x_1, x_2, x_3}(p)}{\sum_{y_1} \sum_{x_1, x_2, x_3} n_{x_1, x_2, x_3} \hat{\pi}_{y_1 y_2 | x_1, x_2, x_3}(p)} \quad (28.d)$$

En la ecuación (28.b) se comprueba cómo las restricciones de igualdad sobre los parámetros producen una expresión similar a una media ponderada de las probabilidades de respuesta no restringidas.

Finalmente, para contrastar la calidad del ajuste, se utilizan los estadísticos ya comentados en este trabajo. Los grados de libertad son

$$gl = (X^*)^T - S^* (X^* + S^* - 1) \quad (29)$$

### 3.1.2 El modelo mixto de Markov

La razón de ser de este modelo es la heterogeneidad de la población respecto del cambio, es decir, no toda la población tiene por qué seguir el mismo patrón de movilidad. *A priori* el investigador puede dividir la población en subgrupos según una o más variables y hacer un análisis dinámico para cada uno. Sin embargo, el modelo aquí expuesto no divide *a priori* según una variable observada, sino que considera que el proceso dinámico es una mixtura de distintos procesos dinámicos.

Su origen se remonta al trabajo de Blumen *et al.* (1955). Encontraron que los procesos de Markov predecían demasiado cambio tras muchas transiciones. En concreto, expusieron que los elementos de la diagonal principal de la matriz de

transición observada tendían a ser subestimados. Para solucionar este problema<sup>3</sup>, propusieron un modelo, llamado *Mover-stayer*, que divide la población en dos grupos. Uno de ellos, *mover*, se caracteriza por comportarse según un modelo de Markov, es decir, la probabilidad de pasar de la categoría  $i$  en el período  $t$  a la categoría  $j$  en el  $t+1$  se recoge en una usual matriz de transición; mientras que el otro, *stayer*, es un grupo de individuos estables, esto es, su matriz de transición es la matriz identidad.

Más tarde, Poulsen (1982) extiende el modelo a una mixtura de  $S^*$  grupos con patrones de cambio comunes, es decir, matrices de transición comunes. Por tanto, este modelo supone que las probabilidades de transición observadas son una mixtura de las probabilidades de un conjunto de grupos no observados, de manera que podemos decir que una variable no observada (latente) influye las probabilidades de transición. A continuación, haremos una exposición del modelo. Sea  $X$  una variable discreta politómica con  $X^*$  categorías observada durante  $T$  períodos y sea  $X_t$  la observación concreta de dicha variable en el momento  $t$ , es decir, se considera la existencia de  $T$  variables discretas politómicas  $X_t$ , donde  $i$  varía desde 1 a  $T$ , con el mismo número de categorías que miden el mismo fenómeno.

Además, se supone la existencia de una variable discreta no observada  $S$  que influye sobre las probabilidades de transición observadas entre las variables  $X_t$ . Dicha variable latente recoge la heterogeneidad de la población, como ha sido indicado anteriormente. Se conoce como modelo mixto de Markov porque supone que la transición entre las variables observadas sigue un modelo markoviano. En la figura siguiente aparecen representadas las relaciones entre tres variables observadas y una variable latente.

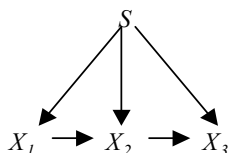


Figura 9. Un modelo mixto de Markov

Se puede comprobar cómo las variables observadas siguen un proceso de Markov de primer orden, ya que una variable en un período sólo depende de la variables en el momento inmediatamente anterior. Por esta razón, este modelo se conoce también como modelo “sin memoria”.

<sup>3</sup> Otra manera de solucionar el problema puede verse en Shorrocks (1976). Éste señala el incumplimiento del supuesto de dependencia de primer orden y propone un modelo de mayor orden para estudiar la movilidad de la renta.

Los parámetros de este modelo son los siguientes:

- la probabilidad inicial  $\pi_s$  de pertenecer a cada uno de los  $S^*$  grupos latentes.
- la probabilidad inicial  $\delta_{x_1|s}$  de estar en cada una de las categorías de la variable  $X_1$  observadas inicialmente dada la pertenencia al subgrupo latente  $s$ .
- las probabilidades de transición  $\tau_{x_t|x_{t-1}s}$  de pasar de cada categoría de la variable  $X_{t-1}$  en el momento  $t-1$  a las categorías de la variable  $X_t$  en el momento  $t$ , dada la pertenencia al grupo latente  $s$ .

Por ejemplo, consideremos una variable observada dicotómica durante tres períodos y una variable latente que divide la población en dos grupos. En la figura 12 se muestra cómo se relacionan los parámetros anteriores para producir las probabilidades observadas del ejemplo antes comentado.

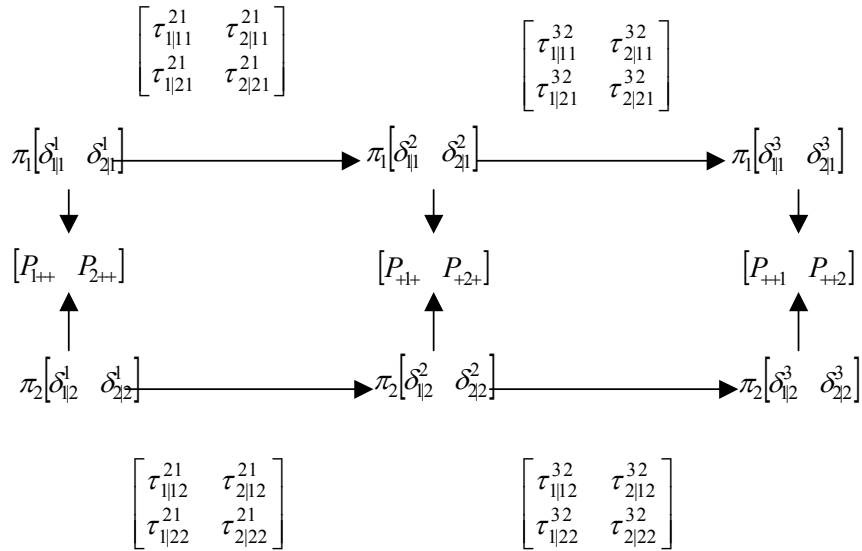


Figura 10. Relaciones entre los parámetros del modelo mixto de Markov

Dados estos parámetros, la probabilidad de pertenecer a una celda de la distribución completa es<sup>4</sup>

$$\pi_{sx_1x_2x_3} = \pi_s \delta_{x_1|s} \tau_{x_2|x_1s} \tau_{x_3|x_2s} \quad (30)$$

<sup>4</sup> Remito al lector a la nota 1 de este capítulo para la explicación de esta notación. Con la usada en el trabajo hasta ahora, la ecuación (30) sería  $\pi_{sx_1x_2x_3} = \pi_s \pi_{x_1|s} \pi_{x_2|x_1s} \pi_{x_3|x_2s}$

Tal distribución de probabilidad de los datos completos de nuevo se obtiene mediante un conjunto de probabilidades marginales y condicionadas.

A partir de la expresión anterior, se puede determinar la probabilidad conjunta de las variables observadas para cada uno de los subgrupos latentes.

$$\pi_{x_1 x_2 x_3 | s} = \delta_{x_1 | s} \tau_{x_2 | x_1 s} \tau_{x_3 | x_2 s} \quad (31)$$

Sin embargo, surge el problema de que la variable que describe la pertenencia a cada grupo no es observada. Por tanto, la proporción  $P_{x_1 x_2 x_3}$  de las variables observadas se calcula sumando sobre la variable latente la expresión (31).

$$\pi_{x_1 x_2 x_3} = \sum_s \pi_s \delta_{x_1 | s} \tau_{x_2 | x_1 s} \tau_{x_3 | x_2 s} \quad (32)$$

La ecuación (32) es una media ponderada de la ecuación (31) donde las probabilidades grupales  $\pi_s$  son los pesos o ponderaciones.

Por tanto, según la ecuación anterior, el modelo mixto de Markov es una mezcla de  $S^*$  cadenas de Markov de primer orden independientes.

Si se supone que la muestra aleatoria de  $N$  individuos, la frecuencia  $n_{x_1 x_2 x_3}$  se obtiene mediante la distribución multinomial  $M(N, \pi_{x_1 x_2 x_3})$ . En consecuencia, el modelo del que hablamos es un modelo multinomial paramétrico.

#### Casos particulares del modelo de Markov

El modelo expuesto en la ecuación (30) recoge dentro de él a otros modelos mediante la imposición de determinadas restricciones.

##### *El modelo Mover-stayer*

Anteriormente, fue expuesto que Blumen *et al.* propusieron este modelo que considera únicamente dos cadenas, una de ellas formada por individuos estables y la otra por los individuos que cambian de categoría.

En este caso se imponen las siguientes restricciones.

- $S^*=2$ , es decir, únicamente se consideran dos grupos en la población.
- las probabilidades de transición entre las categorías  $i$  y  $j$  para el subgrupo de individuos estables deben cumplir esta condición:  $\tau_{j|is}=1$ , si las categorías son la misma y  $\tau_{j|is}=0$ , si son categorías distintas. Esto es, la matriz de transición para el subgrupo estable se corresponde con la matriz identidad.



*El modelo “black and white”*

Este modelo se debe a Converse (1964) y sólo se aplica a variables dicotómicas. Es muy similar al anterior ya que impone las mismas restricciones con la diferencia de que se considera que los individuos móviles tienen la misma probabilidad de estar en cada una de las categorías en el período siguiente.

*El modelo “independencia-estabilidad”*

Con iguales restricciones al modelo *mover-stayer*, además, el subgrupo de los individuos móviles debe cumplir el supuesto de que la probabilidad de cambiar entre los períodos  $t$  y  $t-1$  no depende de la categoría ocupada en el momento  $t$ , es decir, se supone que todas las filas de la matriz de transición son iguales.

*Los modelos estacionarios*

Como su nombre indica, en estos modelos se exige la condición de que las matrices de transición de cada grupo son iguales para cualquier período considerado. Más tarde, se comentará qué efectos produce esta restricción a la hora de estimar el modelo.

Estimación del modelo mixto de Markov

Este modelo puede estimarse utilizando el algoritmo EM como cualquier modelo con variables latentes. En este caso, como ya ha sido comentado, la variable latente o inobservable es la pertenencia a cada subgrupo.

Por tanto, se debe maximizar el logaritmo de la función de verosimilitud sujeto a las restricciones de que los parámetros deben sumar para cada uno de sus parámetros la unidad, ya que son probabilidades.

Si consideramos sólo los datos incompletos u observados, tenemos que las frecuencias  $n_{x_1 x_2 x_3}$  siguen una distribución multinomial  $M(N, \pi_{x_1 x_2 x_3})$  y, consecuentemente, la estimación por máxima verosimilitud consiste en la maximización de la función

$$\log L = \sum_{x_1 x_2 x_3} n_{x_1 x_2 x_3} \cdot \log \pi_{x_1 x_2 x_3} \quad (33)$$

respecto de los parámetros que intervienen en  $\pi_{x_1 x_2 x_3}$  según la expresión (32), considerando que se deben cumplir las siguientes restricciones:

$$\sum_s \pi_s = 1 \quad (34.a)$$

$$\sum_{x_1} \delta_{x_1|s} = 1, \forall s = 1, \dots, S^* \quad (34.b)$$

$$\sum_{x_t} \delta_{x_t|x_{t-1}s} = 1, \forall s = 1, \dots, S^*, \forall t = 2, \dots, T. \quad (34.c)$$

Sin embargo, los parámetros y frecuencias que deseamos estimar son los referidos a los datos completos. Dichos datos  $n_{sx_1x_2x_3}$  se distribuyen según una distribución multinomial  $M(N, \pi_{sx_1x_2x_3})$ , donde  $\pi_{sx_1x_2x_3}$  se especifica según la ecuación (30).

En este caso, la expresión (33) es

$$\begin{aligned} \sum_{s, x_1, x_2, x_3} n_{sx_1x_2x_3} \log \pi_{sx_1x_2x_3} &= \sum_s n_{s\cdots} \log \pi_s + \sum_{s, x_1} n_{sx_1\cdots} \log \delta_{x_1|s} + \\ &+ \sum_{s, x_1, x_2} n_{sx_1x_2\cdots} \log \tau_{x_2|x_1s} + \sum_{s, x_2, x_3} n_{s\cdots x_2x_3} \log \tau_{x_3|x_2s} \end{aligned} \quad (35).$$

Partiendo de esta expresión, se obtienen fácilmente las frecuencias marginales como estadísticos suficientes de los parámetros del modelo y se puede comenzar a aplicar el algoritmo EM.

Según la ecuación (8), tenemos que

$$E\left[n_{sx_1x_2x_3} \mid n_{x_1x_2x_3}, \pi_{sx_1x_2x_3}\right] = n_{x_1x_2x_3} \frac{\pi_{sx_1x_2x_3}}{\pi_{x_1x_2x_3}} = n_{x_1x_2x_3} \pi_{s|x_1x_2x_3} \quad (36)$$

En la etapa  $E$  del algoritmo, se usa la expresión anterior para calcular el valor esperado del estadístico suficiente dadas las frecuencias observadas y la probabilidad conjunta estimada en la iteración anterior del algoritmo.

Dada la estimación  $p$ -ésima de los parámetros, podemos estimar los estadísticos suficientes para los datos completos como

$$\begin{aligned} E\left[n_{s\cdots} \mid n_{x_1x_2x_3}, \hat{\pi}_{sx_1x_2x_3}(p)\right] &= \sum_{x_1, x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p) \\ E\left[n_{sx_1\cdots} \mid n_{x_1x_2x_3}, \hat{\pi}_{sx_1x_2x_3}(p)\right] &= \sum_{x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p) \\ E\left[n_{sx_1x_2\cdots} \mid n_{x_1x_2x_3}, \hat{\pi}_{sx_1x_2x_3}(p)\right] &= \sum_{x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p) \\ E\left[n_{s\cdots x_2x_3} \mid n_{x_1x_2x_3}, \hat{\pi}_{sx_1x_2x_3}(p)\right] &= \sum_{x_1} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p) \end{aligned}$$

Las expresiones arriba mostradas son las correspondientes a los valores esperados de los estadísticos suficientes para las probabilidades de pertenecer a

la cadena latente  $s$ , a la categoría  $x_1$  de la variable  $X_1$  dada la pertenencia a la cadena latente  $s$ , a la categoría  $x_2$  de la variable  $X_2$  dadas la categoría  $x_1$  de la variable  $X_1$  y la cadena latente  $s$  y a la categoría  $x_3$  de la variable  $X_3$  dadas la categoría  $x_2$  de la variable  $X_2$  y la cadena latente  $s$ , respectivamente.

A continuación, en el paso  $M$  se calculan las estimaciones de los parámetros del modelo que maximizan la verosimilitud dados los estadísticos suficientes determinados en la etapa  $E$  del algoritmo.

$$E[n_{s\ldots} \mid \hat{\pi}(p+1)] = N\hat{\pi}_s(p+1)$$

$$E[n_{sx_1\ldots} \mid \hat{\pi}(p+1)] = N\hat{\pi}_s(p+1)\hat{\delta}_{x_1|s}(p+1)$$

$$E[n_{sx_1x_2\ldots} \mid \hat{\pi}(p+1)] = N\hat{\pi}_s(p+1)\hat{\delta}_{x_1|s}(p+1)\hat{\tau}_{x_2|sx_1}(p+1)$$

$$E[n_{s\ldots x_2x_3} \mid \hat{\pi}(p+1)] = N\hat{\pi}_s(p+1)\hat{\delta}_{x_2|s}(p+1)\hat{\tau}_{x_3|sx_2}(p+1),$$

donde

$$\hat{\delta}_{x_2|s}(p+1) = N\hat{\pi}_s(p+1) \sum_{x_1} \hat{\delta}_{x_1|s}(p+1) \hat{\tau}_{x_2|sx_1}(p+1).$$

En concreto, las estimaciones se extraen a partir de las siguientes ecuaciones:

$$N\hat{\pi}_s(p+1) = \sum_{x_1, x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)$$

$$N\hat{\pi}_s(p+1)\hat{\delta}_{x_1|s}(p+1) = \sum_{x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)$$

$$N\hat{\pi}_s(p+1)\hat{\delta}_{x_1|s}(p+1)\hat{\tau}_{x_2|sx_1}(p+1) = \sum_{x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)$$

$$N\hat{\pi}_s(p+1)\hat{\delta}_{x_2|s}(p+1)\hat{\tau}_{x_3|sx_2}(p+1) = \sum_{x_1} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)$$

Por tanto, las estimaciones  $p$ -ésimas de los parámetros son:

$$\begin{aligned} \hat{\pi}_s(p+1) &= \frac{\sum_{x_1, x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)}{N} \\ \hat{\delta}_{x_1|s}(p+1) &= \frac{\sum_{x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)}{N\hat{\pi}_s(p+1)} = \frac{\sum_{x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)}{\sum_{x_1, x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)} \end{aligned}$$

$$\begin{aligned}\hat{\tau}_{x_2|sx_1}(p+1) &= \frac{\sum_{x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)}{N \hat{\pi}_s(p+1) \hat{\delta}_{x_1|s}(p+1)} = \frac{\sum_{x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)}{\sum_{x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)} \\ \hat{\tau}_{x_3|sx_2}(p+1) &= \frac{\sum_{x_1} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)}{N \hat{\pi}_s(p+1) \hat{\delta}_{x_2|s}(p+1)} = \frac{\sum_{x_1} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)}{\sum_{x_1, x_2} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}(p)}\end{aligned}$$

Como ya ha sido indicado en el apartado 2.1 de este capítulo, el proceso se repite hasta que converge. En dicho punto, las estimaciones máximo-verosímiles de las probabilidades son:

$$\hat{\pi}_s = \frac{\sum_{x_1, x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}}{N} \quad (37.a)$$

$$\hat{\delta}_{x_1|s} = \frac{\sum_{x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}}{\sum_{x_1, x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}} \quad (37.b)$$

$$\hat{\tau}_{x_2|sx_1} = \frac{\sum_{x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}}{\sum_{x_2, x_3} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}} \quad (37.c)$$

$$\hat{\tau}_{x_3|sx_2} = \frac{\sum_{x_1} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}}{\sum_{x_1, x_2} n_{x_1x_2x_3} \hat{\pi}_{s|x_1x_2x_3}} \quad (37.d)$$

Una vez obtenidas las estimaciones, debe realizarse el lógico contraste de la calidad del ajuste del modelo. En este caso, los grados de libertad se determinan mediante la expresión

$$gl = (X^*)^T - 1 - [S^* - 1 + S^*(X^* - 1)(1 + X^*(T - 1))] \quad (38)$$

Finalmente, si el modelo presenta la restricción de estacionariedad, es decir, cada subgrupo posee una única matriz de transición para todos los períodos, las ecuaciones (37.c) y (37.d) se sustituyen por

$$\hat{\tau}_{x_{t+1}|sx_t} = \frac{\sum_{x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s|x_1 x_2 x_3}}{\sum_{x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s|x_1 x_2 x_3}} + \frac{\sum_{x_1} n_{x_1 x_2 x_3} \hat{\pi}_{s|x_1 x_2 x_3}}{\sum_{x_1, x_2} n_{x_1 x_2 x_3} \hat{\pi}_{s|x_1 x_2 x_3}} \quad (37.c \text{ bis})$$

ecuación a la que se llega a partir de las expresiones arriba comentadas.

### 3.1.3 El modelo mixto markoviano de clases latentes

Finalmente, el último modelo que se considera es uno que combina los dos anteriores. Fue propuesto por Langeheine y Van de Pol (1990). Dicho modelo extiende el modelo latente de Markov suponiendo que existe una mixtura de  $S^*$  grupos latentes y, por tanto, aparece un conjunto adicional de parámetros, las proporciones  $\pi_s$  de pertenencia a los subgrupos no observables. Dentro de cada subgrupo, la variable  $S$  condiciona al resto de los parámetros del modelo latente de Markov.

Al igual que la exposición de este modelo, se supone un único indicador por ocasión para que la presentación sea más simple. Supongamos  $t$  observaciones consecutivas de una misma variable discreta  $X$  con  $X^*$  categorías.

En consecuencia, existe un vector  $p$  de proporciones iniciales a cada categoría y una matriz o matrices (dependiendo de la estacionariedad o no del modelo) de transición que miden la probabilidad de pasar de una categoría a otra entre la ocasión  $t$  y la ocasión  $t+1$ .

Sin embargo, dado que se tiene en cuenta la existencia del error de medida, se introduce una variable categórica latente  $Y$  en el modelo. El número de categorías de dicha variable,  $Y^*$ , no tiene por qué ser igual al de categorías observadas,  $X^*$ . En consecuencia, suponemos que las matrices observadas de transición con un reflejo de las matrices de transición reales o latente con cierto grado de imperfección.

Por último, se supone que existe una variable latente no observada  $S$ . De esta manera, se expone la consideración de la heterogeneidad de la población respecto del cambio real. Es decir, este modelo presenta una variable latente que condiciona la relación existe entre un conjunto de variables también latentes.

En la siguiente figura se muestra la estructura de las relaciones existe en un modelo de este tipo para tres ocasiones.

Dada esta estructura, el vector de proporciones observadas  $p_t$  depende de la distribución de probabilidad de la variable latente  $Y_t$ ,  $\delta_t$ , y las probabilidades de respuesta  $\rho_{x|y}$  en cada subgrupo, esto es, los parámetros anteriores están condicionados por la variable  $S$ .

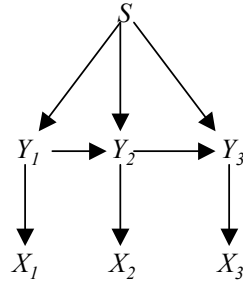


Figura 11. Un modelo mixto markoviano de clases latentes para tres ocasiones

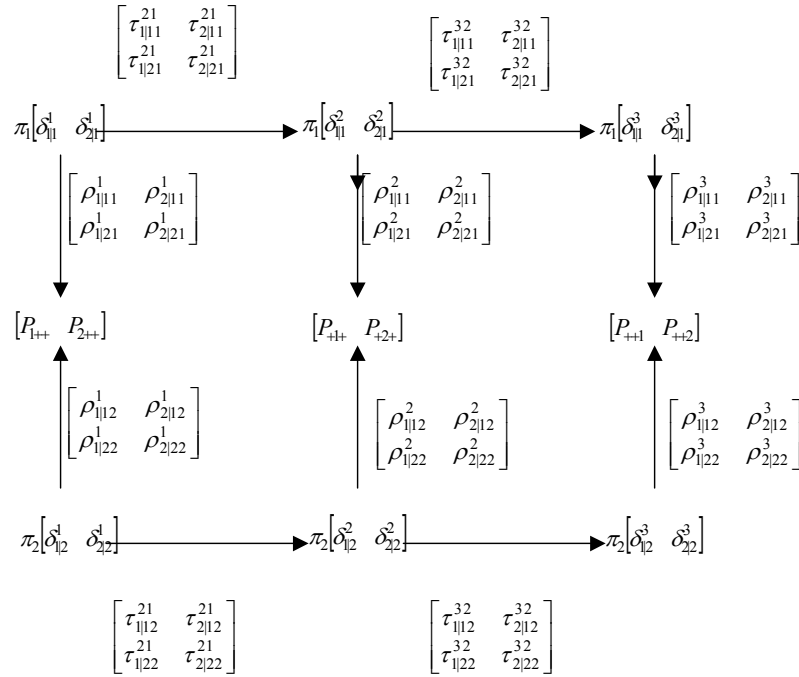


Figura 12. Parámetros y sus relaciones del modelo markoviano mixto de clases latentes

En consecuencia, los parámetros de este modelo son los siguientes:

- la probabilidad inicial  $\pi_s$  de pertenecer a cada uno de los  $S^*$  grupos latentes.

- la probabilidad inicial  $\delta_{y_1|s}$  de estar en cada una de las categorías de la variable latente  $Y_1$  observadas inicialmente dada la pertenencia al subgrupo latente  $s$ .
- las probabilidades condicionadas de respuesta  $\rho_{x_t|y_t,s}$  de estar en cada una de las categorías  $x_t$  observadas dado el estado latente  $y_t$  en el momento  $t$  y el subgrupo latente  $s$ .
- las probabilidades de transición latente  $\tau_{y_t|y_{t-1},s}$  de pasar de cada clase latente  $y_{t-1}$  en el momento  $t-1$  a la clase latente  $y_t$  en el momento  $t$  dado el subgrupo latente  $s$ .

Si se considera un modelo con tres ocasiones y dos subgrupos donde tanto la variable latente como la observada son dicotómicas, la representación gráfica de las asociaciones de las variables y sus parámetros sería la mostrada en la figura 14.

No se consideran todas las distribuciones de probabilidad latentes ya que dadas la distribución de probabilidad inicial y las matrices de transición latentes, se puede obtener la distribución en cada momento.

La probabilidad conjunta de pertenecer a cada celda de la distribución de probabilidad viene dada por la expresión

$$\pi_{sy_1y_2y_3x_1x_2x_3} = \pi_s \delta_{y_1|s} \rho_{x_1|y_1,s} \tau_{y_2|y_1,s} \rho_{x_2|y_2,s} \tau_{y_3|y_2,s} \rho_{x_3|y_3,s} \quad (39)$$

Continuando la línea ya expuesta en los apartados anteriores, sólo se conocen las frecuencias incompletas. En consecuencia, se colapsa la ecuación [39] para relacionar la probabilidad de las variables observadas con el producto de probabilidades condicionadas respecto de las variables latentes.

$$\pi_{x_1x_2x_3} = \sum_{s=1}^{S^*} \sum_{y_1=1}^{Y_1^*} \sum_{y_2=1}^{Y_2^*} \sum_{y_3=1}^{Y_3^*} \pi_s \delta_{y_1|s} \rho_{x_1|y_1,s} \tau_{y_2|y_1,s} \rho_{x_2|y_2,s} \tau_{y_3|y_2,s} \rho_{x_3|y_3,s} \quad (40)$$

Los modelos anteriores son casos especiales de éste. Si en la ecuación (39),  $S^*$  es igual a la unidad, el modelo se convierte en el modelo latente de Markov y si suponemos que la matriz de fiabilidad es una matriz diagonal, es decir, las probabilidades de respuesta valen la unidad cuando coinciden la categoría observada con la latente y 0 cuando son distintas, la ecuación (39) se corresponde con la del modelo mixto de Markov.

#### Estimación del modelo latente mixto de Markov

De nuevo, al existir variables latentes en el modelo, la estimación por máxima verosimilitud de los parámetros se realiza mediante el algoritmo EM.

Dada la ecuación [39], el logaritmo de la función de verosimilitud es

$$\begin{aligned} \sum_{s, y_1, y_2, y_3, x_1, x_2, x_3} n_{x_1 x_2 x_3} \log \pi_{s, y_1 y_2 y_3 x_1 x_2 x_3} &= \sum_s n_{s, \dots} \log \pi_s + \sum_{s, y_1} n_{s y_1, \dots} \log \delta_{y_1 | s} + \\ &+ \sum_{s, y_1, x_1} n_{s y_1 \dots x_1} \log \rho_{x_1 | y_1 s} + \sum_{s, y_1, y_2} n_{s y_1 y_2, \dots} \log \tau_{y_2 | y_1 s} + \sum_{s, y_2, x_2} n_{s \dots y_2 \dots x_2} \log \rho_{x_2 | y_2 s} + \\ &+ \sum_{s, y_2, y_3} n_{s \dots y_2 y_3 \dots} \log \tau_{y_3 | y_2 s} + \sum_{s, y_3, x_3} n_{s \dots y_3 \dots x_3} \log \rho_{x_3 | y_3 s} \end{aligned} \quad (41)$$

En la primera etapa, *Esperanza*, se calculan los valores esperados de los datos completos dados las probabilidades del modelo, cuyas expresiones en la iteración  $p$ -ésima son las siguientes.

$$\begin{aligned} E[n_{s, \dots} | \hat{\pi}(p)] &= \sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s | x_1 x_2 x_3}(p) \\ E[n_{s y_1, \dots} | \hat{\pi}(p)] &= \sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s y_1 | x_1 x_2 x_3}(p) \\ E[n_{s y_1 \dots x_1} | \hat{\pi}(p)] &= \sum_{x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s y_1 | x_1 x_2 x_3}(p) \\ E[n_{s y_1 y_2, \dots} | \hat{\pi}(p)] &= \sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s y_1 y_2 | x_1 x_2 x_3}(p) \\ E[n_{s \dots y_2 \dots x_2} | \hat{\pi}(p)] &= \sum_{x_1, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s y_2 | x_1 x_2 x_3}(p) \\ E[n_{s \dots y_2 y_3, \dots} | \hat{\pi}(p)] &= \sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s y_2 y_3 | x_1 x_2 x_3}(p) \\ E[n_{s \dots y_3 \dots x_3} | \hat{\pi}(p)] &= \sum_{x_1, x_2} n_{x_1 x_2 x_3} \hat{\pi}_{s y_3 | x_1 x_2 x_3}(p) \end{aligned}$$

donde

$$\begin{aligned} \hat{\pi}_{s | x_1 x_2 x_3}(p) &= \frac{\hat{\pi}_{s \dots x_1 x_2 x_3}(p)}{\hat{\pi}_{\dots x_1 x_2 x_3}(p)}, \\ \hat{\pi}_{s y_1 | x_1 x_2 x_3}(p) &= \frac{\hat{\pi}_{s y_1 \dots x_1 x_2 x_3}(p)}{\hat{\pi}_{\dots x_1 x_2 x_3}(p)} \\ \hat{\pi}_{s y_2 | x_1 x_2 x_3}(p) &= \frac{\hat{\pi}_{s \dots y_2 \dots x_1 x_2 x_3}(p)}{\hat{\pi}_{\dots x_1 x_2 x_3}(p)}, \end{aligned}$$



$$\begin{aligned}\hat{\pi}_{sy_3|x_1x_2x_3}(p) &= \frac{\hat{\pi}_{s \cdot y_3 x_1 x_2 x_3}(p)}{\hat{\pi}_{\dots x_1 x_2 x_3}(p)}, \\ \hat{\pi}_{sy_1y_2|x_1x_2x_3}(p) &= \frac{\hat{\pi}_{s y_1 y_2 \cdot x_1 x_2 x_3}(p)}{\hat{\pi}_{\dots x_1 x_2 x_3}(p)} \text{ y} \\ \hat{\pi}_{sy_2y_3|x_1x_2x_3}(p) &= \frac{\hat{\pi}_{s \cdot y_2 y_3 x_1 x_2 x_3}(p)}{\hat{\pi}_{\dots x_1 x_2 x_3}(p)}\end{aligned}$$

Continuando con el algoritmo, en la etapa *Maximización*, se determinan las estimaciones de los parámetros que maximizan la verosimilitud dados los valores esperados del paso anterior.

$$\begin{aligned}E[n_{s \cdot \dots} | \hat{\pi}(p+1)] &= N \hat{\pi}_s(p+1) \\ E[n_{s y_1 \cdot \dots} | \hat{\pi}(p+1)] &= N \hat{\pi}_s(p+1) \hat{\delta}_{y_1|s}(p+1) \\ E[n_{s y_1 \cdot \cdot x_1 \cdot} | \hat{\pi}(p+1)] &= N \hat{\pi}_s(p+1) \hat{\delta}_{y_1|s}(p+1) \hat{\rho}_{x_1|s y_1}(p+1) \\ E[n_{s y_1 y_2 \cdot \dots} | \hat{\pi}(p+1)] &= N \hat{\pi}_s(p+1) \hat{\delta}_{y_1|s}(p+1) \hat{\tau}_{y_2|s y_1}(p+1) \\ E[n_{s \cdot y_2 \cdot \cdot x_2 \cdot} | \hat{\pi}(p+1)] &= N \hat{\pi}_s(p+1) \hat{\delta}_{y_2|s}(p+1) \hat{\rho}_{x_2|s y_2}(p+1) \\ E[n_{s \cdot y_2 y_3 \cdot \dots} | \hat{\pi}(p+1)] &= N \hat{\pi}_s(p+1) \hat{\delta}_{y_2|s}(p+1) \hat{\tau}_{y_3|s y_2}(p+1) \\ E[n_{s \cdot \cdot y_3 \cdot \cdot x_3 \cdot} | \hat{\pi}(p+1)] &= N \hat{\pi}_s(p+1) \hat{\delta}_{y_3|s}(p+1) \hat{\rho}_{x_3|s y_3}(p+1)\end{aligned}$$

donde

$$\begin{aligned}\hat{\delta}_{y_2|s} &= N \hat{\pi}_s \sum_{y_1} \delta_{y_1|s} \tau_{y_2|s y_1}(p+1) \\ \text{y } \hat{\delta}_{y_3|s} &= N \hat{\pi}_s \sum_{y_1, y_2} \hat{\delta}_{y_1|s}(p+1) \hat{\tau}_{y_2|s y_1}(p+1) \hat{\tau}_{y_3|s y_2}(p+1)\end{aligned}$$

Igualando las expresiones de ambas etapas del algoritmo, se obtienen unas expresiones a partir de las cuales despejar las estimaciones  $p+1$ -ésimas de las probabilidades, expresiones que aparecen a continuación.

$$\begin{aligned}\hat{\pi}_s(p+1) &= \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s|x_1 x_2 x_3}(p)}{N}, \\ \hat{\delta}_{y_1|s}(p+1) &= \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s y_1|x_1 x_2 x_3}(p)}{N \hat{\pi}_s(p+1)} = \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s y_1|x_1 x_2 x_3}(p)}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s|x_1 x_2 x_3}(p)},\end{aligned}$$

$$\begin{aligned}
\hat{\rho}_{x_1|sy_1}(p+1) &= \frac{\sum_{x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1|x_1 x_2 x_3}(p)}{N \hat{\pi}_s(p+1) \hat{\delta}_{y_1|s}(p+1)} = \frac{\sum_{x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1|x_1 x_2 x_3}(p)}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1|x_1 x_2 x_3}(p)}, \\
\hat{\tau}_{y_2|sy_1}(p+1) &= \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1 y_2|x_1 x_2 x_3}(p)}{N \hat{\pi}_s \hat{\delta}_{y_1|s}(p+1)} = \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1 y_2|x_1 x_2 x_3}(p)}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1|x_1 x_2 x_3}(p)}, \\
\hat{\rho}_{x_2|sy_2}(p+1) &= \frac{\sum_{x_1, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2|x_1 x_2 x_3}(p)}{N \hat{\pi}_s(p+1) \hat{\delta}_{y_2|s}(p+1)} = \frac{\sum_{x_1, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2|x_1 x_2 x_3}(p)}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2|x_1 x_2 x_3}(p)}, \\
\hat{\tau}_{y_3|sy_2}(p+1) &= \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2 y_3|x_1 x_2 x_3}(p)}{N \hat{\pi}_s(p+1) \hat{\delta}_{y_2|s}(p+1)} = \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2 y_3|x_1 x_2 x_3}(p)}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2|x_1 x_2 x_3}(p)}, \\
\hat{\rho}_{x_3|sy_3}(p+1) &= \frac{\sum_{x_1, x_2} n_{x_1 x_2 x_3} \hat{\pi}_{sy_3|x_1 x_2 x_3}(p)}{N \hat{\pi}_s(p+1) \hat{\delta}_{y_3|s}(p+1)} = \frac{\sum_{x_1, x_2} n_{x_1 x_2 x_3} \hat{\pi}_{sy_3|x_1 x_2 x_3}(p)}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_3|x_1 x_2 x_3}(p)}.
\end{aligned}$$

Este proceso se repite hasta que se alcance la convergencia, es decir, hasta que el incremento de la verosimilitud sea inferior a un valor arbitrario muy pequeño. En este caso, las estimaciones máximo-verosímiles de los parámetros son las mostradas a continuación.

$$\hat{\pi}_s = \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s|x_1 x_2 x_3}}{N} \quad (42.a)$$

$$\hat{\delta}_{y_1|s} = \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1|x_1 x_2 x_3}}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{s|x_1 x_2 x_3}} \quad (42.b)$$

$$\hat{\rho}_{x_1|sy_1} = \frac{\sum_{x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1|x_1 x_2 x_3}}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1|x_1 x_2 x_3}} \quad (42.c)$$

$$\hat{\tau}_{y_2|sy_1} = \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1 y_2 | x_1 x_2 x_3}}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_1 | x_1 x_2 x_3}} \quad (42.d)$$

$$\hat{\rho}_{x_2|sy_2} = \frac{\sum_{x_1, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2 | x_1 x_2 x_3}}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2 | x_1 x_2 x_3}} \quad (42.e)$$

$$\hat{\tau}_{y_3|sy_2} = \frac{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2 y_3 | x_1 x_2 x_3}}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_2 | x_1 x_2 x_3}} \quad (42.f)$$

$$\hat{\rho}_{x_3|sy_3} = \frac{\sum_{x_1, x_2} n_{x_1 x_2 x_3} \hat{\pi}_{sy_3 | x_1 x_2 x_3}}{\sum_{x_1, x_2, x_3} n_{x_1 x_2 x_3} \hat{\pi}_{sy_3 | x_1 x_2 x_3}} \quad (42.g)$$

Estas probabilidades se refieren al modelo no estacionario para las transiciones y la fiabilidad, es decir, las probabilidades de transición y de respuesta no son constantes a lo largo del tiempo. En el caso estacionario, los parámetros  $\hat{\tau}_{y_t|sy_{t-1}}$  y  $\hat{\rho}_{x_t|sy_t}$  serían las medias ponderadas de las ecuaciones (42.d) y (42.f) y (42.c), (42.e) y (42.g), respectivamente.

Finalmente, este modelo se puede contrastar utilizando los conocidos contrastes de calidad del ajuste, donde los grados de libertad vienen dados por el número de parámetros independientes del modelo.

#### 4. CONCLUSIONES

Para terminar y, a modo de resumen, podemos afirmar que los modelos aquí presentados permiten reflejar con gran exactitud los cambios que sufre una variable categórica a lo largo del tiempo. Mediante la incorporación de variables latentes es posible superar problemas como los cambios espurios, la heterogeneidad de la población respecto de la movilidad o la imposibilidad de medir el cambio de una variable no observable.

En el campo en que trabajamos, la economía, la aplicación de estas técnicas, así como cualquier otra que considere como variables de interés un conjunto de variables discretas, es incipiente. No obstante, podemos citar trabajos como el

de Van de Pol y Langeheine (1989) para el desempleo, Ghellini *et al* (1996) para la pobreza, Pérez Mayo y Fajardo (1998) para el riesgo bursátil o Pérez Mayo y Fajardo (en prensa) aplicado al análisis de la movilidad de la renta.

Creemos que aún resta mucho trabajo por realizar, no sólo en el ámbito empírico, sino también en el teórico.

## BIBLIOGRAFÍA

- AKAIKE, H. (1987) Factor analysis and AIC. *Psychometrika*, 52, 317-332.
- BARTHOLOMEW, D.J. (1987) *Latent variables models and factor analysis*. Londres: Griffin.
- BLUMEN, I., KOGAN, M., y McCarthy, P.J. (1955) *The industrial mobility of labor as a probability process*. Ithaca, N.Y.: Cornell University Press.
- CLOGG, C.C. (1981) New developments in latent structure analysis. En: Jackson, D.J. y Borgotta, E.F. (eds) *Factor analysis and measurement in sociological research*, 215-246. Beverly Hills: Sage.
- CLOGG, C.C., ELIASON, S.R. (1987). Some common problems in log-linear analysis. *Sociological Methods and Research*, 16, 8-14.
- CLOGG, C.C., GOODMAN, L.A. (1984) Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79, 762-771.
- CLOGG, C.C., GOODMAN, L.A. (1985) Simultaneous latent structure analysis in several groups. En: Tuma, N.B. (ed) *Sociological Methodology 1985*, 81-110. San Francisco: Jossey-Bass
- CONVERSE, P.E. (1964) The nature of belief systems in mass publics. En: Apter, D.E. (ed) *Ideology and discontent*, 206-261. Nueva York: Free Press.
- DEMPSTER, A.P., LAIRD, N.M., RUBIN, D.B. (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39:1-38.
- GHELLINI, G., PANNUZZI, N., TARQUINI, S. (1995) *A latent Markov model for poverty analysis: the case of the GSOEP*. Documento de trabajo CEPS 11, Differdange (Luxemburgo).
- GOODMAN, L.A. (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61: 215-231.
- HABERMAN, S.J. (1978) *Analysis of qualitative data, Vol. 1, Introduction topics*. Nueva York: Academic Press.
- HABERMAN, S.J. (1979) *Analysis of qualitative data, Vol. 2, New developments*. Nueva York: Academic Press.
- HAGENAARS, J.A. (1990) *Categorical longitudinal data. Log-linear Panel, Trend, and Cohort Analysis*. Londres: Sage Publications

- HAGENAARS, J.A. (1992) Exemplifying longitudinal log-linear analysis with latent variables. En: Van der Heijden, P.G.M., Jansen, W., Francis, B., Seeber, G.U.H. (eds.), *Statistical modelling*, 105-120, Amsterdam: Elsevier Science Publishers.
- HEINEN, A. (1993) *Discrete latent variables models*. Tilburg University (Netherlands): Tilburg University Press.
- JORESÖÖG K.G., SÖRBOM, D. (1988) *Lisrel 7: A guide to the program and applications*.
- LANGEHEINE, R., VAN DE POL, F. (1994) Discrete-Time Mixed Markov latent class models. En: Dale, A., Davies, R.B. (ed) *Analyzing social and political change*. Londres: Sage Publications.
- LAZARSFELD, P.F. (1950). The logical and mathematical foundation of latent structure analysis. En: Stouffer, S.A. et al. (eds.), *Measurement and prediction*, 362-472, Princeton: Princeton University Press.
- LAZARSFELD, P.F., HENRY, N.W. (1968) *Latent structure analysis*. Boston: Houghton Mifflin.
- PÉREZ MAYO, J. FAJARDO CALDERA, M.A. (1998) Análisis dinámico del riesgo en el mercado bursátil. En: Actas de la XII Reunión nacional ASEPELT. Córdoba
- PÉREZ MAYO, J. FAJARDO CALDERA, M.A. (en prensa) Un modelo para corregir el error de medida en el análisis de la movilidad de la renta. Madrid: Instituto de Estudios Fiscales.
- POULSEN, C.A. (1982) *Latent structure analysis with choice modelling*. Aarhus School of Business Administration and Economics, Aarhus (Denmark)
- RAFTERY, A.E. (1986) Choosing models for cross-classifications. *American Sociological Review* 51:145-146.
- VAN DE POL, F., DE LEEUW, J. (1986) A latent Markov model to correct for measurement error. *Sociological Methods and Research* 15: 188-141.
- VAN DE POL, F., LANGEHEINE, R. (1989) Mover-stayer models, mixed Markov models and the EM algorithm; with an application to labour market data from the Netherlands Socio Economic Panel. En: Coppi, R., Bolasco, S. (eds.) *Multiway data analysis*, 485-495. Amsterdam: North Holland.
- VAN DE POL, F., LANGEHEINE, R. (1990) Mixed Markov latent class models. En: Clogg, C.C. (ed.) *Sociological Methodology 1990*. Oxford: Basil Blackwell.
- VERMUNT, J.K., LANGEHEINE, R., BÖCKENHOLT, U. (1995) *Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates*. WORC Paper 95.06.013/7. Tilburg University (Holanda).
- VERMUNT, J.K. (1997) *Log-linear Models for Event Histories*. Londres: Sage Publications.
- WIGGINS, L.M. (1973) *Panel analysis*. Amsterdam: Elsevier.