Elsa Servy
Leticia Hachuel
Gabriela Boggio
Cristina Cuesta
Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística.

ADAPTACIÓN DE LA METODOLOGÍA DE GRIZZLE, STARMER Y KOCH A MUESTRAS DE PANEL ROTATIVAS

1. INTRODUCCIÓN

El presente trabajo es de carácter metodológico pero orientado hacia el análisis de la ocupación, desocupación e inactividad del Gran Buenos Aires (GBA). Se aboca al estudio de la dinámica de la desocupación a través de las mediciones sucesivas del estado ocupacional en las ondas de la EPH (GBA), para los años 1996-1997. Incluye, el estudio de las probabilidades de transición de un estado ocupacional a otro y el de las tasas de desocupación a través del tiempo, a las que se busca modelar haciendo uso de las variables sociales, económicas y demográficas de la EPH.

Los métodos corrientes de análisis suponen que la muestra de individuos es de tipo aleatorio simple, y que los individuos son seguidos a través de todas las ocasiones bajo estudio, y éste no es el caso de la EPH.

En primer lugar, la muestra de la EPH es de tipo estratificado, con etapas múltiples de selección. Además hay que encontrar modelos longitudinales que se hagan cargo de las rotaciones de la EPH. El diseño rotativo de la EPH impone que en cada onda se renueve el 25% de los hogares, de modo que la muestra, después de cuatro ondas sea sustituida en su totalidad. Tanto el tipo complejo de la muestra como la rotación, plantean desafíos metodológico al analista.

En el marco del proyecto "Estudio del evento ocupado-desocupado utilizando mediciones repetidas en el tiempo"[1], este trabajo sólo considera análisis que toman en cuenta la rotación de la EPH, dejando para más adelante modelos más sofisticados que se hagan cargo de la complejidad del diseño de la muestra.

2. EL PROBLEMA METODOLÓGICO QUE ORIGINA LA ROTACIÓN DE LA EPH

El esquema de rotación de la encuesta impone que en cada onda salga de ella una submuestra, constituida por el 25% de los hogares, los que son reemplazados por un número equivalente de hogares elegidos en forma independiente. Por lo tanto, para las cuatro ondas de 1996 y 1997, se pueden obtener paneles independientes de diferentes duraciones, medidos en ocasiones diferentes, como lo indica el esquema siguiente:

[1] Proyecto de Investigación Científica y Tecnológica – PICT 97. Agencia Nacional de Promoción Científica y Tecnológica.

Rotación (h)	*Onda (t)*				Cant. Individuos ($N^{(h)}$)
	Mayo 96 (1)	*Octubre 96 (2)*	*Mayo 97 (3)*	*Octubre 97 (4)*	
1	x	x	x	x	1547
2	x	x	x	-	2288
3	-	x	x	x	2074
4	x	x	-	-	2778
5	-	-	x	x	2864
6	x	-	-	-	4051
7	-	-	-	x	3761

En lo que sigue, se denomina $o(h)$ al conjunto de ondas de las que participa la rotación h-ésima y $s(t)$ al conjunto de rotaciones que son medidas en la onda t-ésima. Se designa $n_i^{(h)}(t)$ y $p_i^{(h)}(t)$ al total de individuos y a la proporción de individuos de la rotación h-ésima que están en el estado i-ésimo (i=1 ocupado, i=2 desocupado, i=3 inactivo) en la onda t-ésima.

Análogamente, $n_{ij}^{(h)}(t)$ y $p_{ij}^{(h)}(t)$ tienen significado similar, pero referido a los que pasaron del estado i-ésimo en la onda (t-1)-ésima al estado j-ésimo en la onda t-ésima. Se designa con $s(t-1,t)$ al conjunto de rotaciones que permanecen en la EPH en los períodos (t-1) y t.

En lo que sigue, se analizan la estacionariedad de las probabilidades de transición y las probabilidades de desocupación mediante la adaptación de la metodología de Grizzle, Starmer y Koch (GSK) para el caso de contar con muestras se panel rotativas.

La metodología es muy flexible, y se la resume a continuación.

3. EL MÉTODO GSK

Para aplicar el método GSK los datos deben ser tales que se los pueda modelar por una multinomial o producto de multinomiales.

El vector de observaciones está dado por el conjunto de proporciones de la tabla o tablas cuyas casillas están en correspondencia con las de las multinomiales.

Sea $\boldsymbol{\pi}$ el vector de probabilidades de las multinomiales involucradas en el problema y $\mathbf{p}$ el vector de proporciones. La matriz de variancias y covariancias de $\mathbf{p}$, $\mathbf{V}$, es proporcional a $\left(\left(\pi_i\left(\delta_{ij}-\pi_{ij}\right)\right)\right)_{i,j}$ cuando el problema toma en cuenta una sola multinomial y es $\mathbf{V}=diag(V_1,V_2,..,V_n)$ si el número de multinomiales independientes es n>1, donde V_h h=1.2.,,,.n son las variancias de las proporciones asociadas con las diferentes multinomiales del problema. Se designa $\hat{\mathbf{V}}$ al estimador consistente de $\mathbf{V}$.

Si el modelo de interés es $\mathbf{A}\boldsymbol{\pi}=\mathbf{X}\boldsymbol{\beta}$, donde $\mathbf{A}$ y $\mathbf{X}$ son matrices de coeficientes conocidos, y las coordenadas de $\boldsymbol{\beta}$ son parámetros a estimar, el estimador de $\boldsymbol{\beta}$ se toma como

$$\hat{\beta}=\left(\mathbf{X}'\mathbf{A}'\hat{\mathbf{V}}^{-1}\mathbf{A}\mathbf{X}\right)^{-1}\left(\mathbf{A}'\hat{\mathbf{V}}\mathbf{A}\right)^{-1}\mathbf{p}\,, \tag{3.1}$$

siempre que las inversas de las matrices sean únicas.

La bondad del ajuste del modelo se prueba mediante la estadística

$$\left(\mathbf{A}\mathbf{p}-\mathbf{X}\hat{\beta}\right)'\left(\mathbf{A}\hat{\mathbf{V}}\mathbf{A}'\right)^{-1}\left(\mathbf{A}\mathbf{p}-\mathbf{X}\hat{\beta}\right) \tag{3.2}$$

y se pueden realizar tests respecto de los $\boldsymbol{\beta}$, tales como $\mathbf{C}'\boldsymbol{\beta} = 0$ mediante la estadística

$$\mathbf{C}'\left(\mathbf{X}\hat{\beta}\right)\left(\mathbf{X}'\mathbf{A}\hat{\mathbf{V}}\mathbf{A}'\mathbf{X}\right)^{-1}\mathbf{X}\hat{\beta}\,\mathbf{C} \qquad (3.3)$$

Tanto (3.2) como (3.3) tienen distribuciones χ^2. Los grados de libertad de (3.2) son w-v donde w y v son el número de filas y columnas respectivamente de $\mathbf{X}$. Los grados de libertad de (3.3) es el rango de $\mathbf{C}$ (que se supone de rango completo).

Si, en cambio, el modelo a considerar es algo más general, del tipo $\mathbf{F}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}$, siendo $\mathbf{F}(\pi)$ un vector de funciones (posiblemente no lineales) de $\boldsymbol{\pi}$, con derivadas parciales de primer y segundo orden con respecto a $\boldsymbol{\pi}$, las fórmulas (3.1), (3.2) y (3.3) son válidas sustituyendo $\mathbf{Ap}$ por $\mathbf{F}(\mathbf{p})$ y $\left(\mathrm{A}\hat{\mathbf{V}}\mathrm{A}'\right)$ por $\mathbf{S}$, donde $\mathbf{S}$ es la estimación muestral de la matriz de covariancias asintótica de $\mathbf{F}$ calculada por el método "delta".

4. EL TEST DE ESTACIONARIEDAD DE LAS MATRICES DE TRANSICIÓN PARA PANELES CON ROTACIÓN

4.1. Metodología

Se deja de lado por ahora la estructura compleja de la muestra de la EPH, y se considera que para la rotación h-ésima, h $\in$ $s(t-1,t)$, y para cada estado i (i=1,2,3), condicionalmente a los valores asumidos por { $n_i^{(h)}(t-1)$ }, las variables { $n_{ij}^{(h)}(t)$, j=1,2,3} tienen distribución multinomial ($t \in$ o(h)) con parámetros { $\pi_{ij}(t)$, $n_{ij}^{(h)}(t-1)$ }, donde $\pi_{ij}(t)$ es la probabilidad condicional de estar en el estado j-ésimo en la onda t-ésima habiendo estado en el estado i-ésimo en la onda anterior. Por lo tanto, para i=1, 2, 3,

$$p^{(h)}{}_{ij}(t) = \frac{n^{(h)}{}_{ij}(t)}{n_i^{(h)}(t-1)}, \text{ j=1, 2, 3; t: (t-1), t}\in\text{o(h) h=1,2,3,4,5} \qquad (4.1.1)$$

$$E(p^{(h)}{}_{ij}(t)) = \pi_{ij}(t) \text{ j=1, 2, 3; t: (t-1), t}\in\text{o(h) h=1,2,3,4,5} \qquad (4.1.2)$$

$$Cov(p_{ij}^{(h)}(t)p_{ij'}^{(h)}(t)) = \frac{1}{n_i^{(h)}(t-1)}\left(\pi_{ij}(t)\left(\delta_{jj'} - \pi_{ij'}(t)\right)\right) \text{ t: (t-1), t}\in\text{o(h) h=1,2,3,4,5 j, j'=1, 2, 3,} \qquad (4.1.3)$$

siendo $\delta_{jj'}$=1 si j=j', y $\delta_{jj'}$=0 si j$\neq$j'.

Sea $V_i^{(h)}(t)$ la matriz 3x3 cuyo elemento general se describe en (4.1.3). Fijado h, existe una multinomial por cada elemento del conjunto [t: (t-1),t$\in$o(h)]. Estas multinomiales, condicionalmente a {$n_i^{(h)}$(t-1), t:(t-1)$\in$o(h)} son independientes.

Dado que $\sum_{j=1}^{3}\pi_{ij}(t) = 1$ y $\sum_{j=1}^{3}p^h{}_{ij}(t) = 1$ $\forall$ i, h, t, el problema de la estacionariedad de las probabilidades de transición puede colocarse en términos de π_{i1}(t) y π_{i2}(t), y las estadísticas asociadas, en términos de $p_{i1}^{(h)}(t)$ y $p_{i2}^{(h)}(t)$ únicamente.

Sea

$$\mathbf{p}_i^{(h)'}(t) = \left(p_{i1}^{(h)}(t), p_{i2}^{(h)}(t)\right) \qquad (4.1.4)$$

$$\mathbf{p}_i{}' = \left(\mathbf{p}_i^{(1)'}(2), \mathbf{p}_i^{(1)'}(3), \mathbf{p}_i^{(1)'}(4), \mathbf{p}_i^{(2)'}(2), \mathbf{p}_i^{(2)'}(3), \mathbf{p}_i^{(3)'}(3), \mathbf{p}_i^{(3)'}(4), \mathbf{p}_i^{(4)'}(2), \mathbf{p}_i^{(5)'}(4)\right) \qquad (4.1.5)$$

y $$\boldsymbol{\pi}_i = \begin{bmatrix} \pi_{i1}(2) \\ \pi_{i2}(2) \\ \pi_{i1}(3) \\ \pi_{i2}(3) \\ \pi_{i1}(4) \\ \pi_{i2}(4) \end{bmatrix}. \qquad (4.1.6)$$

Debido a la independencia condicional de los conjuntos $\{n_{ij}^{(h)}$, j=1, 2, 3} para los valores de {t: (t-1) y $t \in o(h)$} en cada rotación y a la independencia de las submuestras asociadas a las diversas rotaciones,

$$\mathbf{V}_i = Cov(\mathbf{p}_i) = diag\left(\overset{\bullet}{\mathbf{V}}{}_i^{(h)}(t), t:(t-1), t \in o(h), h = 1,2,3,4,5 \right) \qquad (4.1.7)$$

donde $\overset{\bullet}{\mathbf{V}}{}_i^{(h)}(t)$ coincide con $\mathbf{V}_i^{(h)}(t)$ en (4.3) pero con su última fila y columna suprimidas.

Los datos provenientes de todas las rotaciones en s(t, t-1) pueden amalgamarse para lograr un estimador más preciso de $\pi_{ij}(t)$.

$$\hat{\pi}_{ij}(t) = \sum_{h \in s(t-1,t)} \frac{n_i^{(h)}(t-1) p_{ij}^{(h)}(t)}{n_i(t-1)} \quad \text{j=1, 2, 3} \quad \text{t=2,3,4,} \qquad (4.1.8)$$

con

$$n_i(t-1) = \sum_{h \in s(t-1,t)} n_i^{(h)}(t-1) \qquad (4.1.9)$$

Por lo tanto, se tiene un estimador de $\boldsymbol{\pi}_i$,

$$\hat{\boldsymbol{\pi}}_i = \mathbf{A}_i \mathbf{p}_i , \qquad (4.1.10)$$

con $$\hat{\boldsymbol{\pi}}_i' = \left(\hat{\pi}_{i1}(2) \hat{\pi}_{i2}(2) \hat{\pi}_{i1}(3) \hat{\pi}_{i2}(3) \hat{\pi}_{i1}(4) \hat{\pi}_{i2}(4) \right) \qquad (4.1.11)$$

y $\mathbf{A}_i$, una matriz de coeficientes que surgen a partir de (4.1.8).

Es decir,

$$E(\hat{\pi}_i) = E(\mathbf{A}\mathbf{p}_i) = \underset{6*18}{\mathbf{A}} \underset{18*1}{E(\mathbf{p}_i)} = \underset{6*1}{\pi_i} . \qquad (4.1.12)$$

Si se desea colocar el problema en los términos del modelo GSK, debe observarse que $E(\mathbf{p}_i)$ juega el papel de $\boldsymbol{\pi}$ de la Sección 3, que $\mathbf{X}$ es la identidad de orden 6, en tanto que $\boldsymbol{\pi}_i$ corresponde a $\boldsymbol{\beta}$.

La hipótesis de estacionariedad para i fijo (H_{0i}) se puede escribir así:

$$\begin{aligned} \pi_{i1}(2) &= \pi_{i1}(3) = \pi_{i1}(4) \\ \pi_{i2}(2) &= \pi_{i2}(3) = \pi_{i3}(4) \end{aligned} \qquad (4.1.13)$$

Con notación matricial

H_0: $\mathbf{C}\boldsymbol{\pi}_i = 0$ (4.1.14)

con

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \qquad (4.1.15)$$

Sea $\hat{V}_i$ un estimador consistente de V_i en (4.1.7)

Bajo la hipótesis (4.1.13), $Q_i=(\mathbf{C}\hat{\boldsymbol{\pi}}_i)'(\mathbf{C'AV}_i\mathbf{A'C})^{-1}\mathbf{C}\boldsymbol{\pi}_i$, (4.1.16)

se distribuye como un χ^2 con 4 grados de libertad. El test consiste en rechazar H_0 si $Q_i > \chi^2_{4;0.05}$, donde $\chi^2_{4;0.05}$ es tal que $P_r(\chi^2_4 \geq \chi^2_{4;0.05})=0.05$.

Un test conjunto, referido a los tres estados en que puede estar i, se obtiene usando $\sum_i Q_i$ y tomando como referencia la distribución χ^2 con 12 grados de libertad.

4.2. Resultados

Para las ondas años 96 y 97 se presenta a continuación el número de transiciones ocurridas en las cinco primeras rotaciones.

Rotación 1: XXXX

Ocupados (i=1)				
	j=1	J=2	j=3	
T_2	469	36	34	539
T_3	475	39	37	551
T_4	499	28	36	563

Desocupados (i=2)				
	j=1	j=2	j=3	
T_2	43	54	20	117
T_3	38	62	28	128
T_4	50	50	36	136

Inactivos (i=3)				
	j=1	j=2	j=3	
T_2	39	38	814	891
T_3	50	35	783	868
T_4	36	16	796	848

Rotación 2: XXX -

Ocupados (i=1)				
	j=1	j=2	j=3	
T_2	678	57	49	784
T_3	681	62	57	800

Desocupados (i=2)				
	j=1	j=2	j=3	
T_2	58	65	43	166
T_3	79	57	43	179

Inactivos (i=3)				
	j=1	j=2	j=3	
T_2	64	57	1217	1338
T_3	63	61	1185	1309

Rotación 3: - XXX

Ocupados (i=1)				
	j=1	j=2	j=3	
T_3	649	45	47	741
T_4	678	49	38	765

Desocupados (i=2)				
	j=1	j=2	j=3	
T_3	63	68	37	168
T_4	58	53	30	141

Inactivos (i=3)				
	j=1	j=2	j=3	
T_3	53	28	1084	1165
T_4	58	45	1065	1168

Rotación 4: XX - -

Ocupados (i=1)				
	j=1	j=2	j=3	
T_2	861	75	38	974

Desocupados (i=2)				
	j=1	j=2	j=3	
T_2	73	104	52	229

Inactivos (i=3)				
	j=1	j=2	j=3	
T_2	69	78	1428	1575

Rotación 5: - -XX

Ocupados (i=1)				
	j=1	j=2	j=3	

Desocupados (i=2)				
	j=1	j=2	j=3	

Inactivos (i=3)				
	j=1	J=2	j=3	

T_4	928	54	72	1054	T_4	86	71	64	221	T_4	56	43	1490	1589

El test (4.1.16) fue aplicado para comprobar la estacionariedad de las probabilidades de pasar a los diferentes estados ocupacionales a partir de los estados: ocupado (1), desocupado (2) e inactivo (3). Las estadísticas de estos tests tomaron los siguientes valores, junto con las probabilidades asociadas: Q_1=15.17 (p_1=0.004); Q_2=10.02 (p_2= 0.040); Q_3=17.13 (p_3=0.002).

La estacionariedad de las matrices de transición fue rechazada cualquiera fuese el estado inicial. Para realizar un test conjunto de la estacionariedad, involucrando todos los estados, se computó $Q=\sum_i Q_i$ que toma el valor 42.32 con una probabilidad asociada de 0.000, lo que conduce al rechazo de la hipótesis.

5. UN MODELO DE "MEDIDAS REPETIDAS" PARA EXPLICAR LA TASA DE DESOCUPACIÓN, A PARTIR DE UNA MUESTRA DE PANEL ROTATIVA.

Los datos de varias ondas de la EPH constituyen un ejemplo de "medidas repetidas". En lo que sigue, la variable respuesta es la tasa de desocupación y las variables explicativas, variables demográficas y socioeconómicas de la EPH. El factor vinculado a las medidas repetidas es el factor "onda", es decir, el tiempo. En este análisis sólo se toma en cuenta la población económicamente activa (PEA) de la EPH.

La técnica (GSK) que se usa en esta sección, impone que las variables explicativas sean discretas, y la muestra de tamaño grande.

En el caso considerado se está interesado en estudiar:

- el efecto de las variables socioeconómicas y demográficas sobre las tasas de desocupación,
- el efecto de las ondas sobre las tasas de desocupación y
- la interacción entre las variables explicativas y el factor "onda".

A continuación se presenta una adaptación del método GSK para su estudio, cuando las muestras de panel son rotativas.

Para fijar ideas, se considera que las variables explicativas son tres: A con n_A niveles, B con n_B niveles y C con n_C niveles. Por lo tanto, los individuos económicamente activos pueden organizarse en una tabla de dimensión n_A x n_B x n_C.

Los individuos de la PEA pueden tener dos estados ocupacionales en cada una de las ondas: ocupado (O) o desocupado (D). La sucesión de estados posibles por los que puede atravesar un individuo de la EPH depende de la rotación a la que él pertenece.

El esquema siguiente exhibe las posibles sucesiones de estados según las ondas en que se presentan para las distintas rotaciones.

Rotación (h)	*Evento e_k*	*Onda (t)*			
		1	2	3	4
	1	O	O	O	O

1	2	O	O	O	D
	3	O	O	D	O
	4	O	O	D	D
	5	O	D	O	O
	6	O	D	O	D
	7	O	D	D	O
	8	O	D	D	D
	9	D	O	O	O
	10	D	O	O	D
	11	D	O	D	O
	12	D	O	D	D
	13	D	D	O	O
	14	D	D	O	D
	15	D	D	D	O
	16	D	D	D	D
2	1	O	O	O	
	2	O	O	D	
	3	O	D	O	
	4	O	D	D	
	5	D	O	O	
	6	D	O	D	
	7	D	D	O	
	8	D	D	D	
3	1		O	O	O
	2		O	O	D
	3		O	D	O
	4		O	D	D
	5		D	O	O
	6		D	O	D
	7		D	D	O
	8		D	D	D
4	1	O	O		
	2	O	D		
	3	D	O		
	4	D	D		
5	1			O	O
	2			O	D
	3			D	O
	4			D	D
6	1	O			
	2	D			
7	1				O
	2				D

Sea n_{abc} el número de personas en la celda que corresponde a la a-ésima categoría de A, b-ésima de B y c-ésima de C, K(h) el número de eventos (o sucesiones) que corresponden a la h-ésima rotación y $\left\{n_{k,abc}^{(h)}, k=1,...,K(h)\right\}$ las frecuencias de los eventos de la h-ésima rotación,

$$n_{abc}^{(h)} = \sum_{k=1}^{K(h)} n_{k,abc}^{h} \text{ , y}$$

$$p_{k,abc}^{(h)} = \frac{n_{k,abc}^{(h)}}{n_{abc}^{(h)}} \text{ k=1,...,K(h).} \tag{5.1}$$

Se considera que las frecuencias $\left\{n_{k,abc}^{(h)}, k=1,...,K(h)\right\}$ siguen una ley multinomial con parámetros $\left\{n_{abc}^{(h)}, \pi_{k,abc}^{(h)} k=1,...,K(h)\right\}$. Para diferentes valores de h, las multinomiales son independientes.

Se designa $\mathbf{p}_{abc}^{(h)}$ al vector de coordenadas (5.1), de dimensión $K(h)$x1.

Sea $K_t^{(h)}(D)$ el conjunto de eventos $e_{k,abc}^{(h)}$ de la h-ésima rotación, tales que el evento (o la sucesión) posea letra D en la onda t-ésima, y $D_{t,abc}^{(h)}$ el evento "estar desempleado en el tiempo t, pertenecer a la h-ésima rotación y a la PEA que tiene características abc". Este evento tiene probabilidad nula si la rotación h-ésima no interviene el tiempo t-ésimo.

La frecuencia relativa de $D_{t,abc}^{(h)}$, es:

$$p\left(D_{t,abc}^{(h)}\right)=\sum_{k\in K_t^{(h)}(D)} p_{k,abc}^{(h)} \quad h \in s(t),\ t=1,2,3,4. \tag{5.2}$$

Sea $D_{t,abc}$ el evento "estar desocupado en el tiempo t y pertenecer a la PEA con característica abc", y sea $N_{t,abc} = \sum_{h\in s(t)} n_{abc}^{(h)}$.

La frecuencia relativa de $D_{t,abc}$ es

$$p\left(D_{t,abc}\right)=\sum_{h\in s(t)} p\left(D_{t,abc}^{(h)}\right)\frac{n_{abc}^{(h)}}{N_{t,abc}}=\sum_{h\in s(t)}\left(\sum_{k\in K_t^{(h)}(D)} p_{k,abc}^{h}\right)\frac{n_{abc}^{(h)}}{N_{t,abc}} \quad t=1,2,3,4. \tag{5.3}$$

Si $\mathbf{p}\left(D_{abc}\right)$ es el vector de coordenadas $\left\{p\left(D_{t,abc}\right), t=1,2,3,4\right\}$ y $\mathbf{p}_{abc}$ es el vector cuyas coordenadas son $\left\{p_{k,abc}^{(h)}, k=1,...,K(h), h=1,2,3,4,5,6,7\right\}$, (5.3) muestra que,

$$\mathbf{p}\left(D_{abc}\right)=\mathbf{A}\mathbf{p}_{abc} \tag{5.4}$$

donde los coeficientes de $\mathbf{A}$ surgen de (5.2) y (5.3). La dimensión de $\mathbf{A}$ es 4x44.

En cuanto a la matriz de covariancias, se tiene que para la h-ésima rotación,

$$V\left(\mathbf{p}_{abc}^{(h)}\right)=\left(\frac{\pi_{k,abc}^{(h)}\left(\delta_{kk'}-\pi_{k',abc}^{(h)}\right)}{n_{abc}^{(h)}}\right)_{kk'=1,...,K(h)}. \tag{5.5}$$

Como las submuestras de las diversas rotaciones son independientes, así se toman las multinomiales que las modelan, de modo que la matriz de covariancias de $\mathbf{p}_{abc}$ es diagonal en bloques:

$$V\left(\mathbf{p}_{abc}\right)=diag\left(V\left(\mathbf{p}_{abc}^{(1)}\right),V\left(\mathbf{p}_{abc}^{(2)}\right),...,V\left(\mathbf{p}_{abc}^{(7)}\right)\right) \tag{5.6}$$

A partir de (5.4)

$$\mathbf{V}_{abc}=V\left(\mathbf{p}(D_{abc})\right)=\mathbf{A}V(\mathbf{p}_{abc})\mathbf{A}' \quad a=1,...,n_a;\ b=1,...,n_b;\ c=1,...,n_c \tag{5.7}$$

Fórmulas análogas a las 5.1-5.6 pueden obtenerse si en vez de trabajar con el fenómeno de la desocupación se trata con el de la ocupación.

Para fijar ideas se considera que $n_a=n_b=n_c=2$.

Sea $\mathbf{p}'(D)=\left(\mathbf{p}'(D_{000})\mathbf{p}'(D_{001})\mathbf{p}'(D_{010})\mathbf{p}'(D_{011})\mathbf{p}'(D_{100})\mathbf{p}'(D_{101})\mathbf{p}'(D_{110})\mathbf{p}'(D_{111})\right)$.

El método de GSK puede ser utilizado sobre el vector $\mathbf{p}(D)$, con 8x4=32 coordenadas. Desde que para diferentes valores de a, b, c, las rotaciones son independientes,

$$V(\mathbf{p}'(D)) = diag\left(\mathbf{V}_{000}\mathbf{V}_{001}\mathbf{V}_{010}\mathbf{V}_{011}\mathbf{V}_{100}\mathbf{V}_{101}\mathbf{V}_{110}\mathbf{V}_{111}\right). \tag{5.8}$$

Interesa ajustar a $\mathbf{p}(D)$ un modelo logit, tal que la variable a explicar es,

$$\mathbf{F}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = L\,, \tag{5.9}$$

donde $\pi = E(\mathbf{p}(\mathrm{D}))$, y $\mathbf{F}$ debe interpretarse como el vector cuyas coordenadas son la transformación logit de cada una de las coordenadas de $\boldsymbol{\pi}$. El modelo es,

$$\mathbf{L} = \mathbf{X}\mathbf{b}\,, \tag{5.10}$$

y $\mathbf{L}$, $\mathbf{X}$ y $\mathbf{b}$ son tales que

$$\begin{aligned} l_{t,abc} &= \mu + \alpha_a + \beta_b + (\alpha\beta)_{ab} + \gamma_c + (\alpha\gamma)_{ac} + (\beta\gamma)_{bc} + (\alpha\beta\gamma)_{abc} + \tau_t + \\ &+ (\alpha\tau)_{at} + (\beta\tau)_{bt} + (\alpha\beta\tau)_{abt} + (\gamma\tau)_{ct} + (\alpha\gamma\tau)_{act} + (\beta\gamma\tau)_{bct} + (\alpha\beta\gamma\tau)_{abct} \end{aligned}. \tag{5.11}$$

Dado el orden establecido en la disposición de las coordenadas de $\mathbf{p}(D)$, las filas líderes de $\mathbf{X}'$ (relacionadas con α, β, γ y τ), adoptando una reparametrización conveniente, son:

$$\begin{array}{lcccccccccccccccccccccccccccccccc}
\mu: & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\alpha: & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\beta: & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\gamma: & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
\tau_1: & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 \\
\tau_2: & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\
\tau_3: & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1
\end{array} \tag{5.12}$$

Las interacciones pueden obtenerse como productos de las filas anteriores.

Las estimaciones de los parámetros del modelo se estiman así:

$$\hat{\mathbf{b}} = \left(\mathbf{X}'\hat{\mathbf{V}}(\mathbf{L})\mathbf{X}\right)^{-1}\mathbf{X}'\hat{\mathbf{V}}(\mathbf{L})^{-1}\mathbf{L}\,, \tag{5.13}$$

donde $\hat{\mathbf{V}}(\mathbf{L})$ es la estimación de la matriz de covariancia (asintótica) de $\mathbf{L}$ obtenida mediante el método "delta".

$$\hat{\mathbf{V}}(\mathbf{L}) = \mathbf{H}\hat{\mathbf{V}}(\mathbf{P}(\mathbf{D}))\mathbf{H}' \quad \text{con} \quad \mathbf{H} = \left[\frac{\partial\mathbf{F}(\pi)}{\partial\pi}\Big/\pi = \mathbf{P}(\mathbf{D})\right]. \tag{5.14}$$

La bondad del ajuste del modelo se puede probar mediante la estadística,

$$\mathbf{L}'\hat{\mathbf{V}}(\mathbf{L})\mathbf{L} - \hat{\mathbf{b}}'(\mathbf{X}'\hat{\mathbf{V}}(\mathbf{L})\mathbf{X})\hat{\mathbf{b}}\,. \tag{5.15}$$

Los tests respecto a los efectos marginales de los factores A, B, C, así como los del tiempo y sus interacciones se pueden obtener mediante estadísticas del tipo:

$$\mathbf{b}'\mathbf{C}'\left[\mathbf{C}(\mathbf{X}'\hat{\mathbf{V}}(\mathbf{L})\mathbf{X})^{-1}\mathbf{C}'\right]^{-1}\mathbf{C}\mathbf{b}\,, \tag{5.16}$$

refiriéndolas a variables χ^2, cuyos grados de libertad dependen del rango de $\mathbf{C}$ o bien como diferencias de las estadísticas de bondad de ajuste de modelos convenientes.

Los resultados de los tests pueden acomodarse en una tabla ANOVA, que para este caso, con tres factores de dos niveles cada uno y cuatro tiempos, tendría el aspecto que sigue:

Tabla de ANOVA

Fuente de variación	*Grados de libertad*
Interceptación	1
Factor A	1
Factor B	1
Interacción AxB	1
Factor C	1
Interacción AxC	1
Interacción BxC	1
Interacción AxBxC	1
Ondas	3
Interacción AxOndas	3
Interacción BxOndas	3
Interacción AxBxOndas	3
Interacción CxOndas	3
Interacción AxCxOndas	3
Interacción BxCxOndas	3
Interacción AxBxCxOndas (Residual)	3

A partir de éste, otros análisis más específicos pueden realizarse sin dificultades.

6. DISCUSIÓN E INVESTIGACIÓN FUTURA

Las secciones anteriores muestran, una vez más, la flexibilidad del método de Grizzle, Starmer y Koch. Este método tiene, sin embargo, algunas debilidades. Requiere que la muestra sea lo suficientemente grande como para que, en cada casilla originada por el cruce de las variables haya un número razonable de individuos. Esto obliga, en general, a considerar pocas variables explicativas con categorías muy amplias.

En lo inmediato se aplicará el método GSK, adaptado para el problema de las medidas repetidas, a los datos de cuatro ondas de la EPH. Para ello debe desarrollarse un software específico, que está en vías de conclusión.

Tampoco es posible introducir variables continuas en el contexto del GSK. Se planea hacer uso del método GEE (General Estimating Equation) para introducir variables como edad, ingreso, etc.

REFERENCIAS

ANDERSON, T. W. & GOODMAN, L. A., 1957. *Statistical Inference about Markov Chains.* Annals of Mathematical Statistics 28: 89-100.

BISHOP, Y. M. M; FIENBERG, S. E & HOLLAND, P. W., 1980. *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press.

DIGGLE, P. J.; LIANG K. & ZEGER S. L., 1994. *Analysis of Longitudinal Data*. Oxford University Press.

GRIZZLE, J. E.; STARMER, C. F. & KOCH, G. G., 1969. *Analysis of Categorical Data Linear Models.* Biometrics 25: 489-504.

KOCH, G. G.; FREEMAN, D. H. & FREEMAN, J. L., 1971. *Strategies in the Multivariate Analysis of Data from Complex Surveys.* International Statistical Review, 43: 59-78.

PESSINO, C, 1996. *La anatomía del desempleo.* Desarrollo económico, número especial 36: 223-266.

WARE, J. H.; LIPSITZ, S. & SPEIZER, F. E., 1988. *Issues in the Analysis of Repeated Categorical Outcomes.* Statistics in Medicine, 7: 95-107.

AGRADECIMIENTOS

Se agradece al INDEC la provisión de las bases de datos y en especial a la Estadística Clyde Charre de Trabuchi por su asesoramiento en relación a la interpretación de las mismas