



Elsa Servy
Leticia Hachuel
Gabriela Boggio
Cristina Cuesta
Natalia Giordani

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística

MODELOS LOGIT PARA EL ESTUDIO DINÁMICO DE LA DESOCUPACIÓN

1. INTRODUCCIÓN

En Argentina la medición de la ocupación y desocupación, junto a la de otras variables demográficas, económicas y sociales, se realiza principalmente a través de la Encuesta Permanente de Hogares (EPH) que el Instituto Nacional de Estadísticas y Censos (INDEC) viene llevando a cabo desde hace más de dos décadas, dos veces por año.

La muestra de la EPH es de tipo estratificado, con etapas múltiples de selección. 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 de modo que la muestra, después de cuatro ondas sea sustituida en su totalidad.

Frecuentemente se encuentran trabajos que analizan esta información mediante estudios de corte transversal donde se hace uso de los datos de cada onda en forma independiente. Para la comprensión de la dinámica del problema de la desocupación es necesario analizar la información desde un punto de vista longitudinal, esto es posible en función del esquema de rotación de la EPH.

Así, el estudio de los cambios de condición de actividad de los individuos de una onda a la siguiente puede abordarse a través de modelos que, haciendo uso de las variables sociales, económicas y demográficas, permitan estimar las probabilidades de transición. Estos modelos de transición darían respuesta a interrogantes sobre cómo cambia la probabilidad de pasar de un estado laboral a otro ante distintas características de los individuos (diferentes grados de instrucción, sexo, rama de actividad, etc.).

En el marco del proyecto "Estudio del evento ocupado-desocupado utilizando mediciones repetidas en el tiempo"¹, se analiza en esta oportunidad la desocupación en el conglomerado Gran Buenos Aires para los años 1996 y 1997 (cuatro ondas). Este fenómeno se estudia en forma independiente para cada una de las ondas a través de modelos de regresión logística múltiple: *modelos para cortes transversales*. Los cambios de condición de actividad a través del tiempo se abordan mediante modelos que requieren información de ondas consecutivas: *modelos de transición*.

Ambos modelos permiten comprender cómo influyen las características propias de los individuos y las relacionadas con la actividad laboral sobre la condición de ocupado o des-

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

ocupado. Ellos también permiten obtener estimaciones suavizadas de las probabilidades de desocupación bajo diferentes configuraciones de las covariables.

La estimación de las probabilidades de desocupación se obtiene, entonces, tanto por medio de los modelos ajustados para los datos en sus cortes transversales, como así también por los modelos de transición que aprovechan la información del individuo a través de las diferentes ondas.

2. MATERIAL Y MÉTODOS

Las estimaciones globales, tanto de las tasas de desocupación como de las probabilidades de transición, tienen sentido si las muestras son homogéneas. Esto no ocurre, en general, en la práctica. Se pueden formar subgrupos homogéneos y calcular tanto las tasas como las probabilidades de transición dentro de cada uno gracias a lo cual sería posible observar cómo influyen las variables que definen los subgrupos sobre aquéllas. Pero a medida que se refinan las clasificaciones las submuestras resultan demasiado pequeñas. Por eso es conveniente usar un método alternativo: utilizar modelos e introducir dentro del análisis aquellas características que diferencian las submuestras como covariables, en vez de dividir la muestra total.

Se modelan tanto las tasas marginales como de transición utilizando un análisis de regresión logística, donde la variable respuesta es el estado ocupacional.

En función del estudio realizado por Carola Pessino (1996), se consideran como variables explicativas a la edad del individuo, su escolaridad, rama de su actividad, tamaño de la empresa en que trabaja y tenencia de la vivienda en la que reside.

Las variables están medidas o categorizadas de la siguiente manera:

- *Estado ocupacional (Y)*: variable categórica con los siguientes niveles: 0- ocupado, 1- desocupado, 2- inactivo.
- *Edad (X₁)*: variable continua medida en años.
- *Escolaridad (X₂)*: variable ordinal con los siguientes niveles: hasta primaria incompleta - primaria completa - secundaria incompleta - secundaria completa - superior o universitaria completa - superior o universitaria completa.
- *Rama de actividad*: representada por variables dummies (X_h , $h=3, \dots, 8$) asociadas a los siguientes niveles: construcción - manufactura - comercio - transporte - servicios modernos - administración pública y defensa - servicios tradicionales. Categoría de referencia: construcción.
- *Tamaño de la empresa*: representada por variables dummies (X_h , $h=9, 10, 11$) asociadas a los siguientes niveles: una persona - de 2 a 5 personas - de 6 a 25 personas - de 26 a 500 personas - 501 o más personas. Categoría de referencia: una persona.
- *Propiedad de la vivienda*: representada por variables dummies (X_h , $h=12, 13$) asociadas a los siguientes niveles: propietario - inquilino - otro (propietario de la vivienda pero no del terreno, ocupante gratuito y ocupante con relación de dependencia). Categoría de referencia: propietario.

2.1. Modelos para cortes transversales

Para modelar la probabilidad que tiene el i -ésimo individuo perteneciente a la población económicamente activa de estar desocupado en el tiempo t -ésimo, se ajustan modelos de regresión logística binaria para cada una de las ondas, que en este contexto constituyen *modelos para cortes transversales*. Sea Y_{it} el estado ocupacional del i -ésimo individuo en el

tiempo t-ésimo ($Y_{it}=0,1,2$) y X_{it} el vector de coordenadas $\{X_{hit}, h=1,\dots,13\}$, donde X_{hit} es el valor que toma la h-ésima variable para el individuo i-ésimo en el tiempo t.

El modelo propuesto es:

$$\log\left(\frac{P(Y_{it} = 1 / X_{it})}{P(Y_{it} = 0 / X_{it})}\right) = \beta_{0t} + \beta_{1t}^{(1)} X_{1it} + \beta_{1t}^{(2)} X_{1it}^2 + \beta_{2t} X_{2it} + \sum_{h=3}^8 \beta_{ht} X_{hit} + \sum_{h=9}^{11} \beta_{ht} X_{hit} + \sum_{h=12}^{13} \beta_{ht} X_{hit} \quad (2.1.1)$$

donde $i=1,n$; $t=1,2,3,4$ y $P(Y_{it}=1)$ es la probabilidad de que un individuo económicamente activo esté desocupado en el tiempo t. Al modelar sólo la población económicamente activa resulta $P(Y_{it}=2)=0$.

A partir de este modelo es posible estimar la chance que tiene un individuo de estar desocupado en el momento t como función de las variables explicativas. Así, un coeficiente de β puede ser interpretado como el cambio en el log-odds de desocupación por unidad de cambio en la variable X a él asociada, manteniendo constante el resto de las covariables.

Bondad de ajuste y poder predictivo

Una medida usada para comprobar la bondad del ajuste del modelo es el test de Hosmer-Lemeshow. Una limitación de esta estadística es su baja potencia para detectar algunos tipos de falta de ajuste (como no linealidad en las variables explicativas).

Como una medida de la capacidad predictiva del modelo se presenta "c": el área bajo la curva ROC (Receiver operating characteristic). En un modelo de regresión logística con un alto poder predictivo, la curva ROC crece rápidamente y por lo tanto el área bajo la curva es grande. Por el contrario, cuando el modelo tiene baja capacidad predictiva, el área es pequeña. El valor de "c" varía de 0 a 1. Un modelo que no predice mejor que el azar tiene un valor de "c" igual a 0.5 y el modelo ideal tendría un valor igual a 1. Se considera que un valor mayor de 0,7 indica que la capacidad discriminante del modelo es adecuada (Swets, 1988).

Ajustado el modelo logit (2.1.1) se obtiene fácilmente la estimación de $P(Y_{it} = 1 / X_{it})$, es decir la probabilidad de desocupación para los individuos de la población económicamente activa con valores de covariables X_{it} , frecuentemente denominada tasa de desocupación.

2.2. Modelos para datos longitudinales

Como una primera aproximación se usan sólo modelos para procesos en tiempo discreto. Lo que determina que el tiempo se considere discreto es la forma en que fueron obtenidos los valores de la variable respuesta, es decir, los tiempos de observación son tales que fuera de ellos no hay respuesta disponible. El diseño de la EPH en ondas responde a esta situación.

En este marco y conforme al supuesto de un proceso de Markov de primer orden subyacente, las probabilidades de transición pueden ser estimadas a través de un modelo logit del tipo (2.1.1), pero agregando como variable independiente el estado ocupacional del individuo en el tiempo (t-1). Para esta instancia, se consideran los tres niveles posibles de estado ocupacional: ocupado, desocupado e inactivo. Se crearon dos variables "dummies" para individualizar los estados: $Y_{1i(t-1)}$ que toma el valor 1 cuando el individuo está desocupado y 0 en caso contrario, e $Y_{2i(t-1)}$ que toma el valor 1 cuando el individuo está inactivo y 0 en caso contrario. Estas variables conforman el vector $Y_{i(t-1)}$. El modelo así modificado pasa a denominarse *modelo de transición* (Ware, Lipsitz y Speizer, 1988).

$$\log \left(\frac{P(Y_{it} = 1 / X_{it}; Y_{i(t-1)})}{P(Y_{it} = 0 / X_{it}; Y_{i(t-1)})} \right) =$$

$$= \beta_{0t} + \beta_{1t}^{(1)} X_{1it} + \beta_{1t}^{(2)} X_{1it}^2 + \beta_{2t} X_{2it} + \sum_{h=3}^8 \beta_{ht} X_{hit} + \sum_{h=9}^{11} \beta_{ht} X_{hit} + \sum_{h=12}^{13} \beta_{ht} X_{hit} + \sum_{h=1}^2 \alpha_{ht} Y_{hi(t-1)} \quad (2.2.1)$$

donde $i=1, n$; $t=2, 3, 4$; $Y_{hi(t-1)}$ son las variables dummies derivadas de la respuesta politómica y $P(Y_{it} = 1 / X_{it}; Y_{i(t-1)})$ es la probabilidad que tiene un individuo económicamente activo de estar desocupado en el tiempo t dado el estado en el tiempo $(t-1)$, para una determinada configuración de las covariables.

De acuerdo a este modelo, es posible calcular la chance que tiene un individuo de estar desocupado en el momento t en función de las variables explicativas pero también del estado ocupacional en el momento $(t-1)$ - onda anterior. El parámetro $\exp(\alpha_1)$ es la razón de odds de desocupación (versus ocupación) entre los individuos que estuvieron desocupados u ocupados en la onda anterior. El parámetro $\exp(\alpha_2)$ es la razón de odds de desocupación (versus ocupación) entre los individuos que estuvieron ocupados o inactivos en la onda anterior. El parámetro $\exp(\beta)$ puede ser interpretado como el cambio en el odds de desocupación por unidad de cambio en X manteniendo constante el resto de las covariables como así también el estado ocupacional en la onda anterior.

Con respecto a la bondad de ajuste del modelo, se emplean las mismas medidas que para el modelo (2.1.1).

A partir del ajuste del modelo logit (2.2.1) se puede obtener la estimación de la probabilidad de desocupación para los individuos de la población económicamente activa, o lo que es lo mismo, la tasa de desocupación en el tiempo t , dada una configuración de las covariables. A los efectos de simplificar la notación esa probabilidad se simboliza con $P(Y_{it} = 1)$.

Entonces,

$$P(Y_{it} = 1) = \sum_{l=0}^2 P(Y_{it} = 1 / Y_{i(t-1)} = l) \frac{n_{il}}{n_t} \quad (2.2.2)$$

donde n_t es el número de individuos que estuvieron presentes en la onda $(t-1)$ y pertenecen a la población económicamente activa en el tiempo t , y n_{il} es el grupo de tales individuos que presentaron el estado ocupacional l en $(t-1)$, con $l=0, 1, 2$ (ocupado, desocupado e inactivo respectivamente).

3. RESULTADOS

3.1. Tasas empíricas

En la siguiente Tabla se observa la distribución de los encuestados en las cuatro ondas de la EPH bajo consideración, correspondiente al estado ocupacional en sus tres niveles: ocupado, desocupado e inactivo.

Tabla 1: Distribución del estado ocupacional para las ondas de 1996 y 1997

Estado Ocupacional	Frecuencia (%)			
	Mayo 96	Octubre 96	Mayo 97	Octubre 97
Ocupado	3888 (35.3)	3736 (36.1)	3868 (37.1)	4076 (38.5)
Desocupado	870 (7.9)	882 (8.5)	807 (7.7)	686 (6.5)
Inactivo	6256 (56.8)	5730 (55.4)	5745 (55.1)	5830 (55.0)
Total	11014 (100)	10348 (100)	10420 (100)	10592 (100)

Según los datos de esta tabla, las tasas de desocupación (que toman como base las poblaciones económicamente activa) para las cuatro ondas resultan: 18.3%, 19.1%, 17.3% y 14.4%. En el Apéndice se presentan estas tasas, desagregadas según sexo, edad, nivel de educación y rama de actividad - dominios que permiten su comparación con las del estudio de Pessino (1996)-.

La consideración de los cambios de un estado ocupacional a otro en ondas consecutivas para este conjunto de individuos da lugar a las siguientes matrices de transición.

Tabla 2: Matrices de transición correspondientes a las ondas 1996 y 1997

Transición Mayo 96 - Octubre 96				
Estado ocupacional Mayo 96	Estado ocupacional Octubre 96			
	Ocupado (%)	Desocupado (%)	Inactivo (%)	Total (%)
Ocupado	2021 (87.4)	169 (7.4)	122 (5.3)	2312 (100.0)
Desocupado	177 (34.0)	229 (44.0)	115 (22.0)	521 (100.0)
Inactivo	173 (4.5)	174 (4.5)	3487 (91.0)	3834 (100.0)
Total	2371	572	3724	6667
Transición Octubre 96 - Mayo 97				
Estado ocupacional Octubre 96	Estado ocupacional Mayo 97			
	Ocupado (%)	Desocupado (%)	Inactivo (%)	Total (%)
Ocupado	1986 (86.1)	163 (7.1)	157 (6.8)	2306 (100.0)
Desocupado	199 (38.7)	195 (37.9)	120 (23.4)	514 (100.0)
Inactivo	182 (5.0)	134 (3.7)	3315 (91.3)	3631 (100.0)
Total	2367	492	3592	6451
Transición Mayo 97 - Octubre 97				
Estado ocupacional Mayo 97	Estado ocupacional Octubre 97			
	Ocupado (%)	Desocupado (%)	Inactivo (%)	Total (%)
Ocupado	2140 (88.5)	131 (5.4)	148 (6.1)	2419 (100.0)
Desocupado	196 (38.9)	176 (34.9)	132 (26.2)	504 (100.0)
Inactivo	152 (4.2)	106 (2.9)	3397 (92.9)	3655 (100.0)
Total	2488	413	3677	6578

No aparecen cambios importantes en el porcentaje de personas que pasan de un estado ocupacional a otro en las transiciones Mayo 96-Octubre 96, Octubre 96-Mayo 97, y Mayo 97-Octubre 97. Por ejemplo, el porcentaje de individuos que pasan de desocupados a ocupados es 34% en la primera transición, 38.7% en la segunda y 38.9% en la última. Sin embargo, un estudio anterior (Servy et al.; 1999) rechazó la estacionariedad de estas matrices de transición, mostrando que los cambios de la condición de actividad no se dan de la misma forma en todas las transiciones. De acuerdo a este resultado, surge la necesidad de analizar cada transición en forma independiente mediante modelos adecuados.

3.2. Modelización

Se comienza la etapa de modelación estadística utilizando un análisis de regresión logística para la estimación de las tasas de desocupación en cada onda de la EPH (corte transversal).

La variable respuesta es, entonces, el estado ocupacional y, como ya se mencionó, en función del estudio realizado por Carola Pessino (1996), se consideran como variables independientes la edad del individuo, su escolaridad, rama de su actividad, tamaño de la empresa en que trabaja y tenencia de la vivienda en la que reside.

Seguidamente, se ajustan modelos de transición.

3.2.1. Modelos para cortes transversales

En la Tabla 3 se presentan las estimaciones de los coeficientes de los modelos (2.1.1) ajustados en cada una de las ondas junto con la respectiva significación. Se informa además sobre medidas de bondad de ajuste y poder predictivo de los modelos.

Si bien de acuerdo a las estadísticas de bondad de ajuste los modelos ajustan ($p > 0.05$) en todas las ondas, su poder de predicción, descrito por la estadística "c", es bajo.

Tabla 3: Coeficientes estimados para los modelos de corte transversal

Variables	Mayo 96		Octubre 96		Mayo 97		Octubre 97	
	Coef	Prob ^(a)	Coef	Prob ^(a)	Coef	Prob ^(a)	Coef	Prob ^(a)
Constante	1.85		2.17		1.99		1.58	
Edad	-0.07	0.0000	-0.10	0.0000	-0.08	0.0000	-0.09	0.0000
Edad cuadrado	7 E-4	0.0003	1 E-3	0.0000	8 E-4	0.0002	9 E-4	0.0000
Escolaridad	-0.17	0.0000	-0.13	0.0002	-0.20	0.0000	-0.09	0.0143
Rama de actividad ^(b)		0.0000		0.0000		0.0000		0.0000
Manufactura	-0.77		-0.71		-1.05		-0.68	
Comercio	-1.08		-0.92		-0.98		-0.85	
Transporte	-1.09		-1.04		-1.03		-0.84	
Ss. modernos	-0.89		-0.63		-0.75		-1.02	
Adm. Pública	-1.60		-1.25		-1.80		-1.24	
Ss. tradicionales	-0.84		-0.70		-0.81		-0.85	
Tamaño de la empresa ^(c)		0.4232		0.0007		0.3169		0.0000
De 2 a 5 personas	-0.12		-0.37		-0.15		-0.28	
De 6 a 25 personas	-0.16		-0.37		-0.20		-0.38	
De 26 a 500 personas	-0.25		-0.37		-0.26		-0.91	
501 o más personas	-0.28		-0.91		-0.35		-0.87	
Propiedad de la vivienda ^(d)		0.0003		0.0460		0.1041		0.4066
Inquilino	-0.54		-0.25		0.24		0.12	
Otros	-0.06		0.14		-0.02		0.15	
N	4375		4106		4425		4439	
-2 log verosimilitud	3713.50		3533.77		3564.07		3200.03	
HL (gl; p)	6.26 (8; 0.6183)		4.21 (8; 0.8378)		10.34 (8; 0.2420)		10.32 (8; 0.2434)	
c	0.65		0.67		0.66		0.67	

(a) Probabilidad asociada al test condicional de razón de verosimilitud. (b) Categoría de referencia: construcción. (c) Categoría de referencia: una persona. (d) Categoría de referencia: Propietario.

Las variables edad, escolaridad y rama de actividad aportan en forma significativa a la explicación del fenómeno de la desocupación en todas las ondas. En cambio, tamaño de la

empresa y condición de propietario resultan significativas en sólo algunas de ellas (condición de propietario en mayo y octubre del 96; tamaño de la empresa en octubre 96 y 97).

Para evaluar la chance de estar desocupado ante incrementos en la edad, se elige comparar grupos etéreos que difieren en cinco años (chances que se pueden estimar a partir de los coeficientes asociados a la variable edad según un procedimiento que por razones de brevedad no fue incluido en este informe). Así, los individuos de 20 años de edad tienen una chance mayor de estar desocupados que los que tienen 25 años, en el orden de un 20 a un 30%. Si bien al comparar el riesgo de estar desocupado entre los individuos con 35 y 40 años también se observa que es mayor para los más jóvenes, el orden de importancia es menor al anteriormente mencionado (alrededor del 11%). Por el contrario, la chance de estar desocupado para un individuo de 60 años es (entre 5 y 8%) mayor que para los que tienen 55.

Con respecto al nivel de escolaridad, el valor negativo del coeficiente indica que hay mayor riesgo de desocupación en la medida que el nivel de instrucción es más bajo. A modo de ejemplo, para la onda mayo 96, la chance de estar desocupado (versus ocupado) para los individuos de un determinado nivel de escolaridad es $(1/\exp(-0.17))=1.19$ el 20% mayor que para los individuos en el nivel de escolaridad inmediato superior.

Los valores negativos asociados a la variable rama de actividad reflejan que el riesgo de desocupación para todas las ramas es menor que para la construcción. Es decir, la construcción es la que sufre la mayor desocupación, y en comparación a ella, la rama de administración pública y defensa es la que tiene menor chance de desocupación.

En relación a la variable tamaño de la empresa, en las ondas octubre 96 y 97 (donde resulta significativa), los coeficientes se presentan en el sentido esperado: a mayor tamaño de la empresa se espera menor chance de desocupación. En general, las grandes empresas son las que tienen más incentivo a eliminar el trabajo informal y, de esta manera, se reduce la probabilidad de desocupación para los trabajadores de esas empresas (Panigo, Pérez; 1997).

En las ondas mayo y octubre 96, donde el efecto de la variable propietario resulta estadísticamente significativa, la chance de desocupación para los inquilinos es menor que para los propietarios. Este hallazgo parece concordar con la teoría económica respecto a los salarios de reserva, la cual sostiene que a mayor riqueza, mayor tiempo de búsqueda (por el mayor salario de reserva) y mayor tiempo de desocupación (Panigo, Pérez; 1997).

A partir de los modelos ajustados es posible estimar las tasas de desocupación para características diferentes de los individuos. A modo de ejemplo, en la Tabla 4, se presentan las tasas de desocupación para los individuos de 30 años de edad, propietarios de la vivienda y el terreno, pertenecientes a la rama del comercio y con empresas que tienen entre dos y cinco empleados, variando la escolaridad desde primaria incompleta hasta superior o universitaria completa.

Tabla 4: Tasas de desocupación estimadas a partir de los modelos de corte transversal ¹

<i>Escolaridad</i>	<i>Mayo 96</i>	<i>Octubre 96</i>	<i>Mayo 97</i>	<i>Octubre 97</i>
Primario Incompleto	19.8	17.2	19.3	14.6
Primario Completo	17.2	15.5	16.3	13.5
Secundario Incompleto	14.8	14.0	13.7	12.5
Secundario Completo	12.8	12.5	11.5	11.6
Superior o Univ. Incom.	11.0	11.2	9.6	10.7
Superior o Univ. Comp.	9.4	10.0	8.0	9.9

¹ Individuos de 30 años, propietarios, con rama de actividad comercial y pertenecientes a empresas pequeñas (de 2 a 5 personas).

Conforme a lo esperado, las tasas de desocupación decrecen para niveles superiores de educación. Esto ocurre en todas las ondas analizadas de la EPH. Las tasas estimadas de desocupación son notoriamente menores para la última onda, resultado que concuerda con las tasas brutas correspondientes (ver Apéndice).

3.2.2. Modelos para datos longitudinales

En la Tabla 5 se presentan las estimaciones de los coeficientes del modelo (2.2.1) ajustados para las tres transiciones posibles y sus probabilidades asociadas. También se informa sobre medidas de bondad de ajuste y poder predictivo de los modelos.

A través de la estadística de Hosmer y Lemeshow se puede observar el buen ajuste de los dos primeros modelos no así el correspondiente a la tercera transición. Esto se contrapone de alguna forma con el alto poder predictivo del mismo medido por la estadística "c". Sin embargo, diversos autores opinan que la mencionada estadística de bondad de ajuste suele ser inestable. Por estas razones se toman con cautela los hallazgos correspondientes a la tercera transición.

Tabla 5: Coeficientes estimados para los modelos de transición

Variables	Octubre 96		Mayo 97		Octubre 97	
	Coef	Prob ^(a)	Coef	Prob ^(a)	Coef	Prob ^(a)
Constante	-0.52		-1.02		-1.55	
Edad	-0.03	0.0554	9 E-3	0.2800	-0.02	0.6564
Edad cuadrado	2 E-3	0.4537	-2 E-4	0.5411	2 E-4	0.5414
Escolaridad	-0.11	0.0162	-0.15	0.0012	-0.03	0.5706
Rama de actividad ^(b)		0.0074		0.0000		0.0122
Manufactura	-0.74		-1.14		-0.24	
Comercio	-0.88		-0.96		-0.46	
Transporte	-0.86		-0.79		-0.67	
Ss. modernos	-0.69		-0.78		-0.79	
Adm. Pública	-0.80		-2.17		-1.02	
Ss. tradicionales	-0.74		-0.94		-0.73	
Tamaño de la empresa ^(c)		0.1615		0.4412		0.0554
De 2 a 5 personas	-0.27		-0.05		-0.46	
De 6 a 25 personas	-0.02		-0.10		-0.34	
De 26 a 500 personas	0.19		0.15		-0.57	
501 o más personas	-0.19		0.43		-0.55	
Propiedad de la vivienda ^(d)		0.3630		0.0866		0.0549
Inquilino	0.09		0.36		0.38	
Otros	0.25		0.22		0.31	
Estado ocupacional anterior ^(e)		0.0000		0.0000		0.0000
Desocupado	2.54		2.26		2.53	
Inactivo	2.25		1.99		2.09	
N	2629		2714		2696	
-2 log verosimilitud	1886.60		1869.18		1594.44	
HL (g; p)	10.91 (8; 0.2071)		6.18 (8; 0.6268)		18.48 (8; 0.0179)	
c	0.81		0.80		0.82	

(a) Probabilidad asociada al test condicional de razón de verosimilitud. (b) Categoría de referencia: construcción. (c) Categoría de referencia: una persona. (d) Categoría de referencia: propietario. (e) Categoría de referencia: ocupado.

La supuesta correlación de los estados ocupacionales del individuo en momentos consecutivos, se evidencia en la alta significación de los coeficientes correspondientes al estado en la onda anterior.

De acuerdo a los modelos ajustados, los cambios de condición de actividad de una onda a la siguiente quedan explicados por la rama de actividad en todas las transiciones y también por el nivel de instrucción en las dos primeras.

En base a estos modelos se aprecia que la rama de construcción es la más perjudicada en términos de la desocupación. Con respecto al nivel de instrucción, se confirma el efecto negativo de la baja escolaridad.

El modelo (2.2.1) se puede usar para estimar las probabilidades de transición desde un estado ocupacional a la desocupación en la onda siguiente para diferentes configuraciones de edad, educación, rama de actividad, etc.

A modo de ejemplo se calculan las probabilidades de transición a la desocupación para los individuos de 30 años, propietarios de la vivienda y el terreno, pertenecientes a la rama del comercio y con trabajo en empresas que tienen entre dos y cinco empleados, variando el nivel de escolaridad desde primario incompleto hasta superior o universitario completo.

Tabla 6: Probabilidades de transición a la desocupación en las tres últimas ondas según el nivel de educación ¹

Escolaridad	Octubre 96			Mayo 97			Octubre 97		
	Estado anterior			Estado anterior			Estado anterior		
	Ocupa	Desoc	Inactiv	Ocupa	Desoc	Inactiv	Ocupa	Desoc	Inactiv
Hasta primaria incomp.	0.0290	0.2075	0.1769	0.0377	0.2090	0.1789	0.0222	0.1766	0.1323
Primaria completa	0.0261	0.1953	0.1651	0.0329	0.1927	0.1633	0.0216	0.1736	0.1297
Secundaria incompleta	0.0235	0.1832	0.1537	0.0287	0.1768	0.1484	0.0211	0.1706	0.1270
Secundaria completa	0.0211	0.1713	0.1427	0.0249	0.1613	0.1341	0.0205	0.1675	0.1244
Superior o univ. Incomp.	0.0190	0.1597	0.1321	0.0217	0.1465	0.1207	0.0199	0.1646	0.1219
Superior o univ. compl.	0.0171	0.1485	0.1220	0.0188	0.1323	0.1081	0.0194	0.1616	0.1193

¹ Individuos de 30 años, propietarios, con rama de actividad comercial y pertenecientes a empresas pequeñas (de 2 a 5 personas)

A partir de la Tabla 6, se detecta que para una persona del grupo considerado, la probabilidad de pasar a la desocupación de una onda a la siguiente disminuye en la medida que el nivel de educación es más alto. Si bien esta tendencia se observa cualquiera sea el estado ocupacional anterior, es notorio que las probabilidades mencionadas son mucho menores cuando el estado ocupacional inmediato anterior era "ocupado". Desde que se están considerando individuos de 30 años, la transición a la desocupación por parte de los inactivos parece deberse a la decisión de pasar a la PEA de personas que en el período anterior no buscaban trabajo.

La Tabla 7 muestra las tasas de desocupación que se pueden obtener a partir de la probabilidades de transición de la Tabla 6, según fórmula (2.2.2).

Tabla 7: Tasas de desocupación estimadas a partir de los modelos de transición ¹

Educación	Octubre 96	Mayo 97	Octubre 97
Primario Incompleto	16.1	17.4	11.7
Primario Completo	14.9	15.7	11.5
Secundario Incompleto	13.8	14.1	11.2
Secundario Completo	12.7	12.6	11.0
Superior o Univ. Incompleto	11.7	11.3	10.8
Superior o Univ. Completo	10.8	10.0	10.5

¹ Individuos de 30 años, propietarios, con rama de actividad comercial y pertenecientes a empresas pequeñas (de 2 a 5 personas).



Estas tasas, presentan el mismo comportamiento observado a partir de las estimadas por los modelos de corte transversal (Tabla 4), aunque con un rango de variación algo menor. La escasa variabilidad de las tasas entre niveles de educación para octubre 97 deben tomarse con reserva, en tanto el ajuste del modelo de transición correspondiente no fue satisfactorio.

Cabe destacar que comparadas con las obtenidas a partir de los modelos de corte transversal, estas tasas parecen más apropiadas para describir un fenómeno dinámico ya que utilizan información de la historia ocupacional individual. La comprobación estadística de esa suposición está sujeta a estudio.

4. DISCUSIÓN

Este trabajo pretende mostrar los resultados que se alcanzan al estudiar la desocupación a partir de los datos de la EPH, desde dos enfoques metodológicos diferentes: el de corte transversal y el de datos longitudinales.

En el primer enfoque, los factores que sostenidamente (a través de las diferentes ondas) influyen sobre la desocupación son edad, escolaridad y rama de actividad del individuo. En el segundo enfoque, que tiene en cuenta la historia ocupacional del individuo, pierden importancia sus características personales, pero sigue pesando la rama de actividad conjuntamente con el estado ocupacional previo. Estos resultados son consecuencia de las covariables seleccionadas. Sería pertinente continuar la búsqueda de mejores ajustes a partir de otros conjuntos de covariables, también provistos por la EPH.

Un problema para la modelización de las tasas de los diferentes estados ocupacionales es el hecho de que las mismas covariables no se miden en los diferentes estados. Las preguntas que en la EPH se realizan a los ocupados y desocupados difieren de las realizadas a los inactivos. Si se desea usar el mismo modelo para los tres estados ocupacionales deben incluirse como covariables sólo las que emergen de preguntas comunes como sexo, edad, escolaridad, etc., y no las referidas a las características de la actividad laboral. Sin embargo, la elección de un conjunto de covariables definidas para todos los individuos (incluidos los inactivos) para el ajuste de los modelos de transición, permitiría obtener estimaciones más directas de las matrices de transición (evitando el uso de ponderaciones en la estimación de las probabilidades).

Si bien se espera que las estimaciones de las probabilidades de desocupación por medio de los modelos para datos longitudinales sean más precisas que las obtenidas de los modelos para cortes transversales, se está estudiando la forma de compararlas. Las curvas ROC constituyen una adecuada herramienta para ello, pero las asociadas a los modelos de transición requieren una construcción especial, no provista por los software estándares.

Además, en el marco de este proyecto, se está trabajando en forma paralela, en la adaptación del método de Grizzle, Starmer y Koch (1969) para tomar en cuenta el esquema rotativo de la EPH y en la consideración del diseño muestral complejo de la misma.

REFERENCIAS

- AGRESTI, A., 1990. *Categorical Data Analysis*. John Wiley & Sons.
- 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.

- de VRIES, S; FIDLER, V; KUIPERS, W; HUNINK, M, 1998. Fitting Multistate Transition Models with Autoregressive Logistic Regression: Supervised Exercise in Intermittent Claudication. *Medical Decision Making* 18: 52-60.
- 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.
- HOSMER, D. W. & LEMESHOW, S., 1989. *Applied Logistic Regression*. John Wiley & Sons. New York.
- PANIGO, D. & PÉREZ, E, 1997. Perfil de los desocupados de la provincia de Buenos Aires. Un análisis cualitativo y probabilístico. *Anales – Asociación Argentina de Economía Política*. Tomo IV.
- PESSINO, C., 1996. La anatomía del desempleo. *Desarrollo económico*, número especial 36: 223-266.
- SERVY, E.; HACHUEL, L.; BOGGIO, G.; CUESTA, C; 1999. *Modelos de transición para el estudio del evento ocupado-desocupado en el Gran Buenos Aires*. Actas CLATSE IV. Mendoza.
- SWETS, JA, 1988. Measuring the accuracy of diagnostic system. *Science*, 240:1285-1293.
- 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.

APENDICE

Tasas de desocupación por edad y sexo, GBA.

Varones				
Edad (años)	Mayo 96	Octubre 96	Mayo 97	Octubre 97
15-19	40.5	43.5	40.4	33.1
20-34	16.3	16.8	13.1	12.1
35-49	11.8	12.4	10.4	8.3
50-64	16.3	14.3	13.3	12.1
Total	16.7	16.8	14.4	12.3

Mujeres				
Edad (años)	Mayo 96	Octubre 96	Mayo 97	Octubre 97
15-19	42.8	52.8	46.1	43.9
20-34	21.3	22.3	23.1	19.7
35-49	16.7	17.2	17.6	13.9
50-64	16.6	18.1	15.9	11.8
Total	20.3	22.2	21.4	17.4

Fuente: EPH, INDEC.

Tasas de desocupación por nivel de educación, GBA.

<i>Educación</i>	<i>Mayo 96</i>	<i>Octubre 96</i>	<i>Mayo 97</i>	<i>Octubre 97</i>
Nunca asistió	21.2	18.7	21.3	6.8
Primario incompleto	20.5	21.7	23.6	18.4
Primario completo	20.5	21.1	20.0	16.6
Secundario incompleto	22.6	22.9	18.2	15.7
Secundario completo	15.8	16.3	16.9	13.6
Superior o universitario incompleto	16.5	20.1	15.2	14.5
Superior o universitario completo	8.3	8.2	7.0	5.9

Fuente: EPH, INDEC.

Tasas de desocupación para nuevos ingresantes y con ocupación previa, GBA.

	<i>Mayo 96</i>	<i>Octubre 96</i>	<i>Mayo 97</i>	<i>Octubre 97</i>
Tasa general	18.0	18.8	17.1	14.3
Con ocupación previa	15.8	15.9	14.7	12.4
Nuevos ingresantes	2.2	2.9	2.3	1.7
Se desconoce	0.1	0.0	0.1	0.2

Fuente: EPH, INDEC.

Tasas de desocupación por rama de actividad, GBA.

<i>Rama de actividad</i>	<i>Mayo 96</i>	<i>Octubre 96</i>	<i>Mayo 97</i>	<i>Octubre 97</i>
Actividades primarias	6.8	27.1	15.2	6.2
Industrias alimentos, bebida y tabaco	15.0	15.0	14.2	13.5
Industrias textiles, confecciones y calzado	23.1	18.0	19.4	15.3
Ind. Prod. quím., ref. petróleo y comb. nucl.	13.7	15.5	11.6	10.6
Ind. Prod. metálicos, maquinarias y equipos	18.2	15.4	6.8	11.6
Otras industrias manufactureras	16.2	20.9	14.1	16.3
Suministro de electricidad, gas y agua	13.2	8.1	0.0	14.3
Construcción	34.4	32.1	31.6	26.9
Comercio al por mayor	12.9	8.5	11.4	10.1
Comercio al por menor	13.8	15.3	15.8	14.6
Restaurantes y hoteles	17.8	20.2	14.5	12.9
Transporte	12.3	9.9	12.4	10.3
Ss. conexos de transporte y comunicaciones	11.5	13.0	13.3	11.2
Intermediación financiera	10.7	12.5	7.6	7.1
Activ. de alquiler, inmobiliarias y empres.	13.3	16.4	14.3	8.3
Administración pública y defensa	6.5	8.1	4.2	4.4
Enseñanza	9.8	10.3	8.9	4.9
Servicios sociales y de salud	10.4	6.9	12.4	8.2
Otras activ. de ss., comunitarias y sociales	12.7	13.9	10.2	11.1
Servicios de reparación	22.5	23.7	15.4	18.4
Hogares privados con servicio doméstico	21.8	24.3	24.0	17.7
Otros servicios personales	13.4	22.2	14.4	10.2
Sin especificar	19.3	11.3	20.4	20.7

Fuente: EPH, INDEC.