

## **UN ESTUDIO ESTADÍSTICO SOBRE LA INFLUENCIA DE LA ACTIVIDAD SOBRE LA EVOLUCIÓN DEL DESEMPLEO**

**Blaconá, María Teresa**  
**García, María del Carmen**  
**Quaglino, Marta Beatriz**  
**Ventroni, Nora**

*Instituto de Investigaciones Teóricas y Aplicadas en Estadística. Escuela de Estadística*

### **INTRODUCCION**

La Encuesta Permanente de Hogares (EPH) que realiza semestralmente, en una primera y segunda onda, el Instituto Nacional de Estadística y Censos (INDEC) se ha constituido en el instrumento básico para el relevamiento socio-laboral en el país. Entre sus propósitos se encuentra la captación de información necesaria para la construcción de indicadores básicos de la evolución de los niveles de ocupación y desocupación. La EPH se realiza en veintiocho aglomerados urbanos que representan el 70% de la población urbana del país y el 98% de la población residente en centros de más de 100000 habitantes (INDEC, 1998).

El propósito de este trabajo es analizar la evolución del desempleo a través de la tasa de desocupación, en veinticinco aglomerados urbanos de Argentina en el período 1985-1996. Los tres aglomerados que se excluyen son aquéllos para los cuales la medición se realiza desde períodos posteriores a 1985.

Para el análisis se utilizan, como proponen varios autores, modelos lineales generales de regresión para datos longitudinales considerando distribución normal y matriz de covarianza especificada. Estos modelos permiten estudiar cambios producidos en el período contemplando la correlación existente entre observaciones sucesivas de cada aglomerado.

En trabajos anteriores se abordó el presente tema bajo distintas condiciones, se consideró la tasa de desocupación a través de la segunda onda, en el período 1982-1992. Se agruparon los aglomerados en cuatro categorías de acuerdo al nivel de desocupación promedio y la tendencia de dichas series en el período considerado. Se modeló la tasa de desocupación usando como variable explicativa sólo el tiempo.

Para describir la evolución del nivel medio a través del tiempo en cada uno de los grupos, como así también el comportamiento propio de cada aglomerado, se eligió entre varios propuestos, un modelo de efectos aleatorios (Laird y Ware, 1982). La estimación de los parámetros se realizó utilizando el Filtro de Kalman, expresando el modelo bajo la forma de espacio de los estados.

En el presente trabajo se avanza sobre el análisis realizado anteriormente, incorporando los datos de la primer onda y ampliando el período a los años 1985-1996. Se incluyen además como variables explicativas la tasa de desocupación rezagada un período y la de actividad. Se ajustan modelos, tanto para la variable respuesta tasa de desocupación como incremento de la tasa, que consideran la presencia o no de grupos, los cuales fueron formados teniendo en cuenta el nivel de la tasa de desocupación a través del tiempo. Se comparan modelos que tienen la misma expresión para los valores esperados, pero suponen distintas estructuras de correlación entre las observaciones sucesivas.

Se encuentran modelos aceptables discriminando los aglomerados en grupos, tomando como variable respuesta el incremento de la tasa de desocupación, como variables explicativas el incremento de la tasa de desocupación desfásada un período y el incremento de la tasa de actividad, considerando la posibilidad de presencia de efecto aleatorio y una estructura de covarianza de independencia.

## MATERIAL Y DEFINICIONES OPERACIONALES

La tasa de desocupación se define como el cociente entre el número de desocupados de un período y la población económicamente activa (ocupados + desocupados).

La tasa de actividad, a partir de la cual se define una variable explicativa, se calcula como el cociente entre la Población Económicamente Activa y el Total de Población.

En unos pocos aglomerados falta información de algunos períodos, los cuales se estimaron suponiendo que las series para cada aglomerado se comportan como un modelo autorregresivo de orden uno.

La mayoría de los aglomerados en estudio presenta tendencia creciente de la tasa de desocupación, pudiendo visualizarse tres comportamientos distintos a través del tiempo:

- aglomerados con un crecimiento muy pronunciado en su tasa, llegando en el último período a una tasa promedio de 19.14% (Bahía Blanca, Partidos del conurbano bonaerense, Gran Córdoba, Gran La Plata, Gran Rosario, Gran Tucumán y Tafí Viejo, Salta, Santa Fe y Santo Tomé, Gráfico 1);
- aglomerados con crecimiento más moderado que el grupo anterior, alcanzando un nivel promedio de 12.34% (Capital Federal, Comodoro Rivadavia, Corrientes, Gran Catamarca, Gran Resistencia, Gran San Juan, Jujuy y Palpalá, La Rioja, Neuquén y Plottier, Paraná, Santiago del Estero y La Banda, Gráfico 2) y
- aglomerados con comportamiento más estable (Formosa, Gran Mendoza, Posadas, San Luis y El Chorrillo, Santa Rosa y Toay, Río Gallegos, Gráfico 3).

## MODELOS PARA DATOS LONGITUDINALES

Se proponen varios modelos para el análisis de los datos de desocupación a través del tiempo, teniendo en cuenta distintas estructura de dependencia entre las observaciones sucesivas. La expresión general del modelo para el aglomerado  $i$ -ésimo es:

$$Y_i = X_i\beta + Z_i\gamma_i + e_i, \quad i=1, 2, \dots, 25$$

$Y_i$ : vector de  $n_i$  observaciones para el  $i$ -ésimo aglomerado cuyas componentes son  $\{y_{it}, t=1 \dots n_i\}$

$X_i$ : matriz de diseño conocida de dimensión  $n_i \times k$ , con  $k$ =número de parámetros de efectos fijos y elemento genérico  $x_{p,q}^{(i)}$ ;

$\beta$ : vector de parámetros de efectos fijos desconocido, de dimensión  $k$ ;

$Z_i$ : matriz de diseño conocida para los efectos aleatorios de dimensión  $n_i \times h$ , con  $h$ =número de parámetros de efectos aleatorios, y elemento genérico  $z_{p,q}^{(i)}$ ;

$\gamma_i$ : vector de efectos aleatorios que describen las propiedades intrínsecas del  $i$ -ésimo aglomerado con  $E(\gamma_i) = 0$  y matriz de variancia y covariancias  $G$  de dimensión  $h$ ;

$e_i$ : vector de errores de dimensión  $n_i$ , con  $E(e_i) = 0$  y matriz de variancias y covariancias  $W_i$ .

$\gamma_i$  y  $e_i$  son independientes.

$$Y_i \sim MVN(X_i\beta, \Sigma_i), \text{ donde } \Sigma_i = Z_iGZ_i' + W_i$$

GRAFICO 1: Tasas de Desocupación Grupo a (en %)

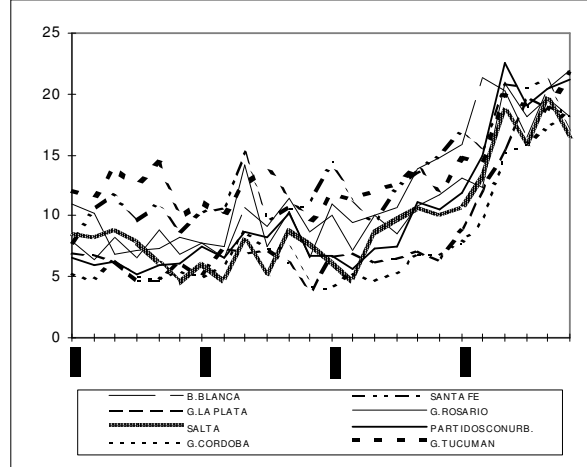


GRAFICO 2: Tasas de Desocupación Grupo b (en %)

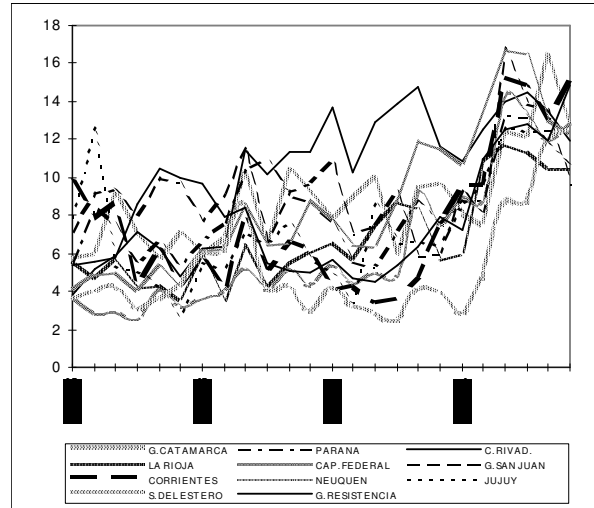
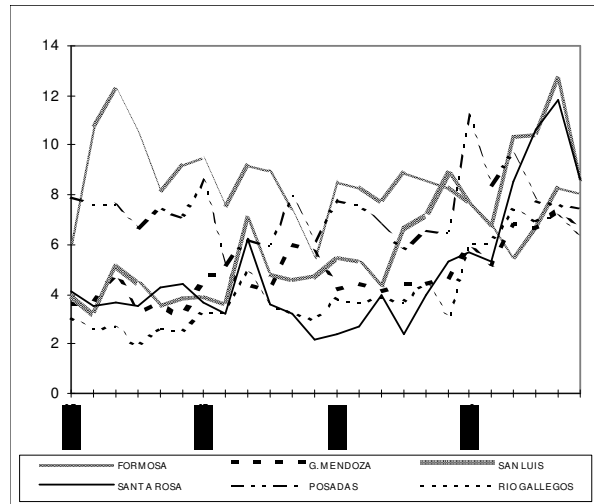


GRAFICO 3: Tasas de Desocupación Grupo c (en %)



Se pueden introducir distintos supuestos sobre la matriz de covariancia. Una estructura de covariancia razonable es la que con pocos parámetros permite captar las probables covariancias entre las observaciones cercanas en el tiempo y puede ser estimada con los datos disponibles. Es necesario también que la estructura de covariancia supuesta sea compatible con los valores esperados del modelo seleccionado.

En el presente trabajo se definen matrices de covariancia que consideran distintas estructuras de dependencia estocástica entre las observaciones de un mismo aglomerado, ellas son:

i) Los errores son independientes entre observaciones sucesivas y no se consideran efectos aleatorios del aglomerado:  $\Sigma_i = W_i = \sigma^2 I$

ii) La correlación entre los errores de observaciones sucesivas es constante y no se consideran efectos aleatorios:

$$\Sigma_i = W_i = \begin{pmatrix} \sigma^2 + \chi & \chi & \chi & \dots & \dots & \chi \\ \chi & \sigma^2 + \chi & \chi & \dots & \dots & \chi \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \chi & \dots & \dots & \dots & \sigma^2 + \chi & \chi \\ \chi & \dots & \dots & \dots & \chi & \sigma^2 + \chi \end{pmatrix}$$

iii) Los errores se comportan como un proceso AR(1) y no se consideran efectos aleatorios<sup>1</sup>:

$$R_i = \begin{pmatrix} 1 & \rho & \dots & \dots & \rho^{\eta-2} & \rho^{\eta-1} \\ \rho & 1 & \dots & \dots & \rho^{\eta-3} & \rho^{\eta-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{\eta-2} & \rho^{\eta-3} & \dots & \dots & 1 & \rho \\ \rho^{\eta-1} & \rho^{\eta-2} & \dots & \dots & \rho & 1 \end{pmatrix}$$

iv) Existe un efecto aleatorio propio del aglomerado con matriz de covariancias G no especificada y  $W_i$  puede representar un proceso autorregresivo de orden uno como en el caso iii) o tener una estructura arbitraria.

### ANÁLISIS DE LA INFORMACIÓN

Para evitar la dependencia con el tiempo de la variable Tasa de Desocupación ( $Des_{it}$ ,  $i=1\dots 25$ ,  $t=1\dots 23$ ) se realiza la primera diferencia y se trabaja con la variable Incremento ( $Y_{i,t} = \Delta Des_{i,t} = Des_{i,t} - Des_{i,t-1}$ ), la cual refleja los cambios que se producen en dicha tasa entre períodos de medición consecutiva, y que tiene comportamiento estacionario para todos los aglomerados.

Como variables explicativas del modelo, se utilizan el incremento de la Tasa de Actividad ( $X^{(1)}_{2,t} = \Delta Act_{i,t}$ ) y el incremento de la tasa de desocupación desfasada un período ( $X^{(1)}_{1,t} = \Delta Des_{i,t-1}$ ) La elección de la primera como variable regresora se debe a que permite explicar los cambios de la tasa de desocupación y está registrada para todos los aglomerados por la misma fuente en el período considerado. En total se dispone de  $N = 575$  observaciones con  $n_i = 23$  respuestas por aglomerado.

En el trabajo Blaconá y otros (1996), se encontró un modelo dinámico<sup>2</sup> que ajustaba aceptablemente las tasas de desempleo de los aglomerados. El mismo consideraba los tres grupos enunciados en la sección 2 y usaba como variable explicativa el incremento de la tasa de actividad. La matriz de covariancias correspondía a un proceso autorregresivo de primer orden.

En el presente estudio, tomando en cuenta los resultados del trabajo mencionado, se prueba estadísticamente que resulta más conveniente un modelo que considere solamente dos grupos, uniendo los llamados grupos a) y b) de la sección anterior. A partir de este resultado se plantean

<sup>1</sup> Por conveniencia de notación se escribe la matriz de correlaciones  $R_i$  en lugar de  $W_i$ .

<sup>2</sup> Para la estimación de los modelos dinámicos, se toman como valores iniciales la primera observación de cada aglomerado, realizando la estimación con los restantes  $n_i - 1$  datos por aglomerado.

modelos con el mismo valor esperado, teniendo en cuenta diferentes estructuras de la matriz de covariancias y la existencia o ausencia de efectos aleatorios propios de los aglomerados.

Se estima un modelo que contempla una estructura autorregresiva de primer orden entre las observaciones sucesivas de cada aglomerado (Modelo iii), utilizando el Filtro de Kalman (Jones (1995)). La estimación máximo verosímil del coeficiente autorregresivo es  $-0.05$ , no significativamente distinto de cero. La variancia del efecto aleatorio resulta no significativa.

En base a este resultado se reestima el modelo, considerando las estructuras de covariancia de independencia (Modelo i) y simetría compuesta (Modelo ii), utilizando el procedimiento MIXED de SAS.

En ambos casos la estimación del valor esperado y los valores de los criterios usuales para la bondad del ajuste son muy similares. Teniendo en cuenta la simplicidad en la interpretación, se opta por el modelo de independencia. El modelo estimado resulta:

$$\left\{ \begin{array}{l} \Delta Des_{it}^{(1)} = 0.462876 - 0.281634 \Delta Des_{it-1}^{(1)} + 0.702819 \Delta Act_{it}^{(1)} \\ \quad (0.0001) \quad (0.0001) \quad (0.0001) \\ \Delta Des_{it}^{(2)} = 0.168998 - 0.321082 \Delta Des_{it-1}^{(2)} + 0.218330 \Delta Act_{it}^{(2)} \\ \quad (0.2948) \quad (0.0043) \quad (0.1431) \end{array} \right.$$

Los valores entre paréntesis representan las probabilidades asociadas de la estadística  $t$ , para evaluar si los coeficientes son significativamente distintos de cero.

El Grupo 1 está formado por diecinueve aglomerados, que constituyen la mayoría de la población urbana económicamente activa del país y en los cuales la tasa de desocupación presenta un importante incremento a través del período en estudio. En este grupo se encuentra que el cambio de la tasa de desocupación, está inversamente relacionado con el cambio ocurrido en el período anterior y en forma directa con el cambio de la tasa de actividad en el mismo período.

El Grupo 2 está formado por seis aglomerados con tasas de desocupación más estables en el período de estudio. Para este grupo existe relación inversa con el cambio de la tasa en el período anterior (igual que en el Grupo 1), pero no existe relación con el incremento de la tasa de actividad en el mismo período.

## CONSIDERACIONES FINALES

La utilización de modelos para datos longitudinales permitió analizar un conjunto de series de tasas de desocupación, en el período 1985-1996, de aglomerados urbanos de Argentina, en los cuales el Instituto Nacional de Estadística y Censos releva la Encuesta Permanente de Hogares.

Mediante dichos modelos es posible obtener estimaciones de comportamientos promedio, teniendo en cuenta la correlación existente entre observaciones consecutivas.

Se encuentra un modelo dinámico aceptable, que explica la respuesta media que existe en el incremento de la tasa de desocupación y usa como covariable el incremento de la tasa de actividad.

La respuesta promedio difiere si el grupo de aglomerados presenta crecimiento en la tasa de desocupación en el período considerado respecto del grupo cuyas tasas de desocupación permanecen estables.

En el primer grupo, formado por diecinueve aglomerados urbanos que representan la mayoría de la población activa del país, existe una relación inversa con el incremento de la tasa de desocupación del período anterior y directo con el incremento de la tasa de actividad. En el segundo grupo existe la misma relación con el incremento de la tasa de desocupación desfasada un período, sin embargo, no existe relación significativa con el incremento de la tasa de actividad.

Los aglomerados que más se asemejan al comportamiento medio de los Grupos 1 y 2 son: Capital Federal y Mendoza, respectivamente.



La relación inversa en ambos grupos entre los incrementos de la tasa de desocupación de dos períodos consecutivos, estaría reflejando un efecto estacional en las mediciones, dado que éstas se realizan en los meses de mayo y octubre de cada año.

Las relaciones encontradas a través de este modelo estadístico, podrían ayudar a interpretar enfoques económicos sobre los movimientos de la fuerza de trabajo en un país, teniendo en cuenta las características propias de cada región.

## **BIBLIOGRAFÍA**

- BOX, G.E.P. and JENKINS, G.M. (1976). *Time series analysis, forecasting and control*. Revised Edition, Holden-Day, San Francisco.
- COOX, N.R. and WARE, J.H. (1983). "Design and analysis methods for longitudinal research". *Annals Review Public Health*, **4**, 1-23.
- DE JONG, P. (1988). "The likelihood for a state space model". *Biometrika*, **75**, 165-169.
- (1991). "Stable algorithms for the state space model". *Journal Time Series Analysis*, **12**, 143-157.
- (1991). "The diffuse Kalman filter". *Ann.Statist.*, **19**, 1073-1083.
- DIGGLE, P.J. (1988). "An approach to the analysis of repeated measurements data". *Biometrics*, **45**, 959-971.
- DIGGLE, P.J., LIANG, K.Y. and ZEGUER, S.L. (1996). *Analysis of Longitudinal Data*, Oxford Science Publications.
- HAND, D. and CROWDER, M. (1996). *Practical Longitudinal Data Analysis*. Chapman & Hall.
- INDEC (1998). *Documento de trabajo: Propuesta de Reformulación de la Encuesta Permanente de Hogares*. Argentina.
- JONES, R.H. (1993). *Longitudinal data with serial correlation: A state-space approach*. Chapman and Hall.
- LINDSEY, J.K. (1994). *Models for Repeated Measurements*. Oxford Science Publications.
- SERVY, E., BLACONA, M.T., CUESTA, C., GARCIA, M.del C., QUAGLINO, M. (1994). *Análisis de datos longitudinales*. Informe de trabajo presentado a INDEC.
- TIMM, N. and MIECZKOWSKI, T. (1997). *General Linear Models: Theory and applications using SAS Software*, Cary, NC: SAS Institute Inc., 619 pp.