

Pronóstico de las defunciones por medio de los modelos autorregresivos integrados de promedios móviles

Manuel Ordorica Mellado

El Colegio de México

Resumen

El objetivo del presente trabajo es realizar una proyección del número total de defunciones de México mediante la metodología de series de tiempo de Box-Jenkins, es decir, siguiendo el modelo autorregresivo de promedios móviles integrado (ARIMA). Se dispone de la información de muertes totales desde 1922 al año 2000. La idea es aplicar nuevas metodologías estadísticas para hacer pronósticos demográficos. El método de series de tiempo está muy vinculado al pasado, por lo que no considera el efecto de la estructura por edad de la población sobre el número de fallecimientos. Este hecho resulta ser una fuerte limitación del modelo. Sin embargo, puede ser útil para proyecciones demográficas de corto plazo.

Palabras clave: proyecciones de mortalidad, series de tiempo, estadísticas vitales, Box-Jenkins, México.

Abstract

Mortality projection through autoregressive integrated moving average model

In order to apply new statistical tools to make population projections, the purpose of this paper is to project total number of deaths for Mexico through Box-Jenkins methodology of time series, using autoregressive integrated moving average model (ARIMA), with total deaths data since 1922 to 2000. The Box-Jenkins method is linked to the past and do not consider the effect of the population age structure over the number of deaths. In spite of this important limitation of the model, it could be useful to short run demographic projections.

Key words: mortality projections, time series, vital statistics, Box-Jenkins, Mexico.

Introducción

Las proyecciones de la mortalidad en México se han realizado mediante el uso de modelos matemáticos como la función de Makeham, la de Gompertz, entre otras. Además se ha utilizado otro tipo de instrumento del análisis demográfico, como las tablas modelo de mortalidad que presentan distintos niveles y estructuras de las defunciones. Las hipótesis que se plantean con este tipo de métodos suponen que el país evolucionará dependiendo del

nivel y la estructura de las muertes en el que se encuentra y de acuerdo con la región del mundo a la que pertenezca. Dicha dinámica está de acuerdo con la teoría de la transición demográfica. Asimismo, se ha utilizado la función logito para este mismo propósito, proyectando los parámetros de dicho modelo. Todos estos métodos incorporan implícitamente la teoría mencionada.

Hasta la fecha son pocos los trabajos que se encuentran en la literatura de proyecciones de los componentes demográficos que utilizan modelos de series de tiempo con el enfoque de la metodología de Box y Jenkins.

Además, gran parte de las proyecciones que se realizan respecto a la mortalidad utilizan las tasas específicas de mortalidad o a la probabilidad de muerte para obtener índices de supervivencia y luego extrapolar la esperanza de vida al nacer. Sin embargo, prácticamente ningún pronóstico proyecta el número de defunciones. Esto se debe en gran medida a que el efecto de la estructura por edad sobre los números absolutos condiciona la elaboración de pronósticos del número de muertes. No obstante estas limitaciones, parecería interesante utilizar una técnica ajena a la demografía, y más cercana a la estadística, con el fin de apreciar su potencialidad. Sobre todo porque hay países que no cuentan con datos de las variables demográficas por grupos de edades y sexo, y cuando se tienen son de poca calidad. Muchos de los países que se encuentran en esta situación utilizan técnicas indirectas de estimación demográfica, con las cuales trabajan a partir de información que se recaba en censos y encuestas.

La proyección de la mortalidad mediante la utilización del sistema logito supone que el logito de la tabla de vida inicial varía linealmente en el tiempo, tendiendo hacia el logito de la tabla de mortalidad final. El matiz de esta metodología se encuentra en el ritmo de variación que se suponga para el descenso de la mortalidad. La proyección de la mortalidad mediante la función probabilidad de muerte consiste en interpolar probabilidades de fallecimiento entre las tablas de vida inicial y final, para luego construir tablas de vida e índices de supervivencia para puntos intermedios.

En otras ocasiones se utilizan técnicas muy rudimentarias que no por eso dan malos resultados. Recordemos el famoso método de la mano alzada. La dinámica del nivel de la mortalidad de México permite aplicar este método, ubicando a la tasa específica de mortalidad para una edad en diferentes momentos en el tiempo y observar, en un gráfico semilogarítmico, cómo dicha evolución sigue un comportamiento lineal. Si esa recta la continuamos en el tiempo, obtenemos una extrapolación de dicha tasa específica de mortalidad. Y

si esto lo hacemos para todas las edades es posible obtener probabilidades de muerte y en consecuencia obtener también esperanzas de vida para distintas edades en diferentes momentos en el tiempo.

Muchos de estos métodos toman como punto de apoyo las tablas de mortalidad límite, es decir, utilizan los niveles más bajos alcanzados o que se podrían alcanzar, para estimar mediante interpolación las tablas para momentos intermedios. Se realiza una interpolación entre la tabla inicial y la final. Se supone que en el largo plazo se llegará a esos niveles límite.

Objetivo

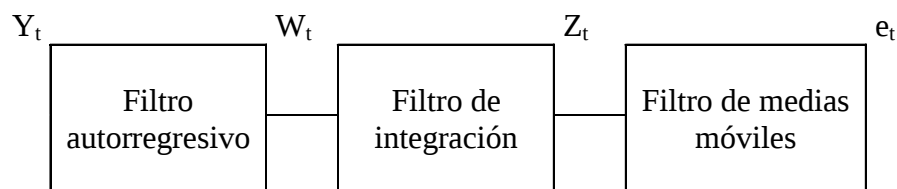
El objetivo del presente trabajo es realizar una proyección de las defunciones utilizando la técnica de Box y Jenkins. Se dispone de las defunciones totales desde 1922 hasta 2000. El método supone que la variable analizada se puede explicar en función de la misma variable en diferentes momentos en el tiempo y a partir de los errores. Es decir, se establece un modelo que vincula a la variable en estudio en un momento t con ella misma en $t - 1$, $t - 2 \dots$ y con los errores en t , $t - 1$, $t - 2 \dots$ Una ventaja de esta forma de proyectar la variable es que podemos actualizar el pronóstico cada año, una vez que dispongamos de nueva información. Así se convierte en un sistema de pronóstico que se actualiza permanentemente. Sin embargo, como ya mencionamos, una desventaja de proyectar las defunciones anuales es el efecto de la composición por edad sobre este número. Por tal motivo, a fin de evitar en la medida de lo posible dicho efecto, el periodo a cubrir en la proyección será, para el corto plazo, de sólo 10 años.

La metodología de Box- Jenkins

En 1976, Box y Jenkins publicaron su libro *Time series analysis. Forecasting and control* (Box y Jenkins, 1976) en el cual mencionan cuatro aplicaciones prácticas del pronóstico de series de tiempo: planeación económica y financiera, planeación de la producción, control de inventarios y producción, y control y optimización de procesos industriales. Como se puede apreciar, nada de esto se vincula con la demografía.

El método de Box-Jenkins consiste en extraer los movimientos predecibles de los datos observados. La serie de tiempo se descompone en varios componentes, algunas veces llamados “filtros”, precisamente porque la filosofía del método

consiste en detectar las distintas componentes usando los filtros correspondientes, hasta obtener residuales no predecibles cuyo comportamiento tiene poca influencia en el resultado final. El enfoque de Box-Jenkins hace principalmente uso de tres filtros lineales: el autorregresivo, el de integración y el de medias móviles.

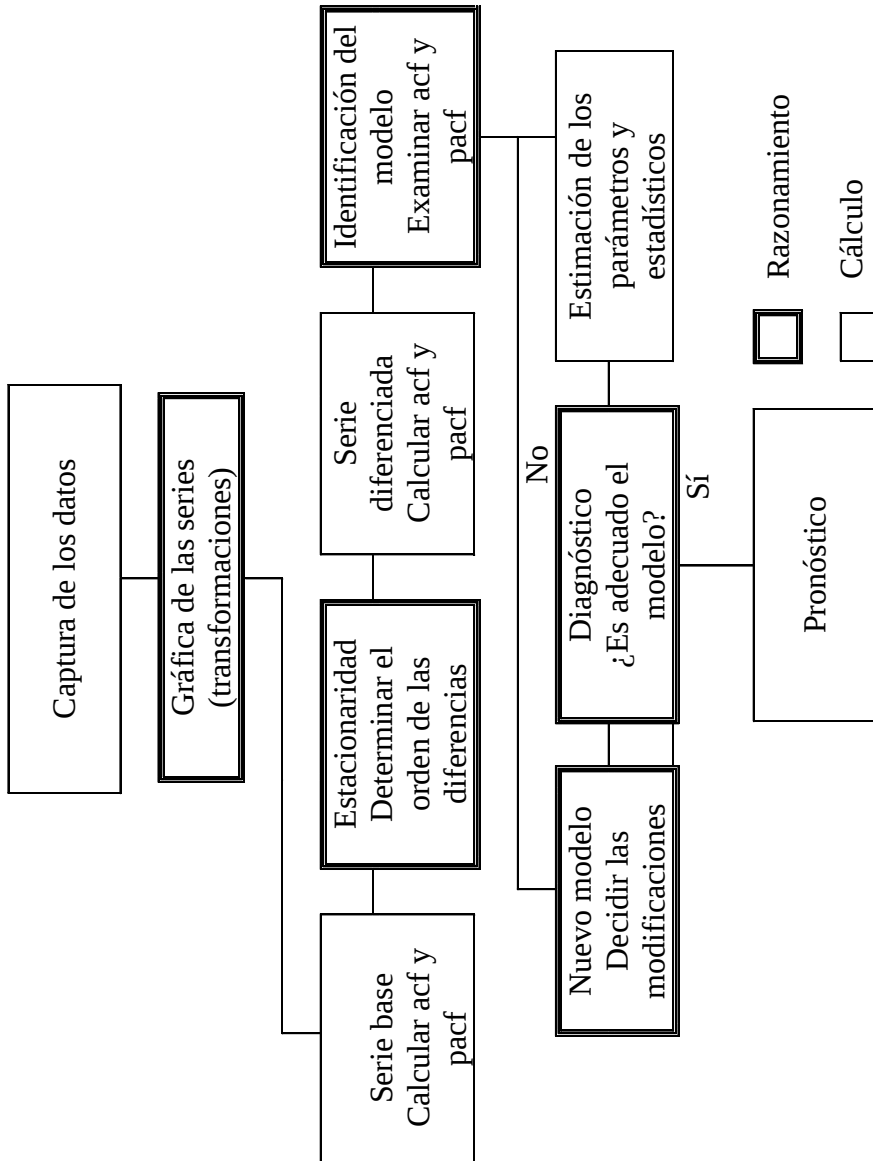


El proceso iterativo de Box-Jenkins para construir modelos lineales de series de tiempo consiste en cuatro pasos:

1. Identificación de las especificaciones.
2. Estimación de los parámetros.
3. Diagnóstico de la adecuación del modelo.
4. Pronóstico de realizaciones futuras.

En la página siguiente se presenta un diagrama funcional del método de Box-Jenkins:¹

¹ Las siglas acf representan a la función de autocorrelación y las siglas pacf representan la función de autocorrelación parcial (Vandaele, 1983: 63).



Los procedimientos empleados en la identificación son inexactos y requieren de mucha experiencia, así como del conocimiento del fenómeno en cuestión. La idea es elegir un modelo sencillo y tratar de mejorarlo. Por supuesto, no se podrá hablar de un modelo óptimo.

El modelo ARIMA multiplicativo es el más general, pues incorpora el análisis de la estacionalidad. A partir de este modelo es casi seguro que se encontrará alguno adecuado al fenómeno que se esté manejando. Con este modelo se resuelven la mayoría de los problemas. Se describe de la siguiente manera:

Modelos ARIMA multiplicativos

$$N(B) \Phi(B^s) z_t = \theta(B) \theta(B^s) e_t$$

$$\text{ó}$$

$$\text{ARIMA } (p,d,q) \times (P,D,Q)_s$$

donde

$$z_t = \left(\sum_{s=0}^D \right)^d w_t$$

Resultados

Al aplicar la técnica de Box y Jenkins se observa que el mejor modelo fue el ARIMA(10, 0, 8)(1, 0, 1)₁₆ debido a que el coeficiente de determinación es de 0.68 (cuadro 1). Se observa una estacionalidad de 16 años. Estacionalidades más pequeñas también se ajustan, pero ésta es la que presenta el coeficiente de determinación más elevado. Se probaron un gran número de modelos y éste fue el que dio un mejor ajuste. Modelos inferiores a los ARIMA(4, 0, 4)(1, 0, 1)₁₆ daban coeficientes de determinación inferiores a 0.60, por lo que no fue posible cumplir el principio de parsimonia, es decir, tener un modelo con pocos

parámetros. La ventaja fue haber utilizado el paquete JMP, que resulta ser muy amigable, como se dice en el ámbito de los programadores.

CUADRO 1
MODELO: ESTACIONAL ARIMA(10, 0, 8)(1, 0, 1)₁₆
RESUMEN

DF	58
Sum of squared errors	1.25524e10
Variance estimate	216420286
Standard deviation	14711.2299
Akaike's 'A' information criterion	1558.22589
Schwarz's bayesian criterion	1607.9843
Rsquare	0.67613853
RSquare Adj	0.56250293
-2 Loglikelihood	1514.69176
Stable	Yes
Invertible	Yes

Fuente: cálculos propios del autor.

¿Qué significa el modelo resultante?

El modelo obtenido tiene en la variable rezagos de hasta diez periodos, y rezagos de hasta ocho periodos en los errores, además de ciclos de dieciséis años. Es importante señalar que el comportamiento de la evolución de las defunciones es oscilatorio en el periodo en el que se hace la proyección, con ciclos largos. Este hecho va en contra de lo que generalmente ocurre, es decir, de que la dinámica de los componentes del crecimiento natural difícilmente presenta perturbaciones. Sin embargo, los fallecimientos presentan fluctuaciones. El máximo de las defunciones se encuentra en el año de 1970 con 485 686 defunciones, y el mínimo se presenta en el año de 1923, con 356 574 muertes.

¿Cómo explicar esto?

Al revisar los datos se observa que entre 1922 y 1943 hay un aumento de las defunciones, y a partir de 1943 y hasta 1956 desciende el número de muertes

significativamente. La declinación en el número pudiera explicarse por los avances en la ciencia médica, cuando se empezaron a utilizar los antibióticos. Recordemos que la penicilina, descubierta por Fleming en 1929, se difunde más ampliamente por esas fechas al igual que las sulfas. En el periodo de 1921 a 1940 se registra una importante recuperación del crecimiento demográfico con una tasa de aumento anual de 1.7 por ciento. Esta velocidad de crecimiento resulta muy significativa, sobre todo entre 1921 y 1930, ya que durante la Revolución la tasa de crecimiento demográfico fue de -0.67 por ciento. En 1936 se le da un impulso al tema poblacional, toda vez que, a iniciativa del Ejecutivo, el Congreso aprueba la Ley General de Población, la cual fue renovada 11 años después. En esta ley se fomentó la protección biológica y legal de la infancia, su mejor alimentación; higiene de las habitaciones, centros de trabajo y elevación del nivel medio de subsistencia, entre otras acciones. Entre 1940 y 1960 se observó una caída muy pronunciada en los niveles de la mortalidad. En términos de la esperanza de vida al nacer, las mujeres ganaron 15.6 años, mientras que los hombres ganaron 14.8 años. Durante esta época hubo una ampliación de los servicios de salud sobre todo a partir de la fundación del Instituto Mexicano del Seguro Social en 1942. Se intensifican las inversiones en obras de infraestructura que afectan los niveles de salud, tales como la introducción de agua potable, drenaje y alcantarillado. Entre 1960 y 1980 se observa un freno en el ritmo de descenso en los niveles de la mortalidad, lo que se aprecia por las pequeñas ganancias en la esperanza de vida. De 1956 a 1970 el número absoluto de muertes se incrementa nuevamente como resultado también del aumento en la tasa de natalidad. Al haber más población hay más muertes. En este caso se observa cómo una variable interfiere sobre otra. Posteriormente vuelve a descender el número de fallecimientos de 1970 a 1986 por el efecto de la fuerte caída en los niveles de la natalidad. En 1973, el Congreso de la Unión aprobó la Ley General de Población, la que tiene entre sus objetivos centrales regular el ritmo de crecimiento demográfico; la tasa de aumento de la población descende de 3.2 en 1977 a 2.4 por ciento en 1982. Aunque con oscilaciones, se observa que entre 1986 y 1998 hay nuevamente un aumento en el número absoluto de muertes, y luego un leve descenso (gráfica 1). Las oscilaciones en periodos largos hacen que el modelo que toma en cuenta ciclos sea el que da un mejor ajuste. Se observa que las muertes pronosticadas para el decenio 2000 al 2010 oscilan alrededor de 420 mil muertes, y prácticamente se mantienen en ese número durante los 10 años con leves fluctuaciones.

Cuando se comparan estos pronósticos sobre el número absoluto de defunciones con los estimados en las proyecciones de la población de México: 2000-2050, publicadas en 2002 por el Consejo Nacional de Población (Conapo), se observa cómo las calculadas por el método de series de tiempo son inferiores. El Conapo estima para el 2000, 456 886 fallecimientos; 474 041 para el 2005 y 509 756 para el 2010. Para pronósticos de más largo plazo, las diferencias son todavía más elevadas. Este aumento tiene que ver con el impacto de la composición por edad sobre el número de muertes. El método de series de tiempo no considera este fenómeno demográfico.

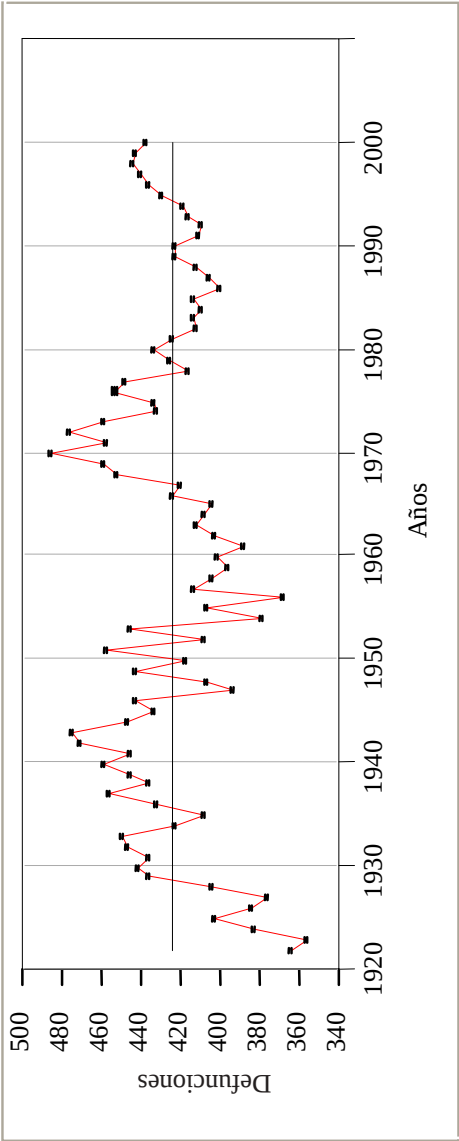
Conclusiones

El método de series de tiempo está muy vinculado al pasado, por lo que no considera el efecto de la estructura por edad sobre el número de fallecimientos. Este hecho resulta ser una fuerte limitación del modelo. Sin embargo, puede ser útil para proyecciones de corto plazo, las cuales son susceptibles de actualizarse año con año. Sería conveniente intentar un modelo más complejo, como por ejemplo, un autorregresivo integrado de promedios móviles vectorial (ARIMAV) que considere una variable adicional que tome en cuenta la estructura por edad de la población.

Finalmente, con el último modelo ajustado fue posible elevar un poco el coeficiente de determinación, pero se perdió algo en el principio de parsimonia, como lo muestran los criterios de Schwarz y de Akaike, ya que fueron muchos los parámetros estimados.

Anexo estadístico

GRÁFICA 1
E.U.M. SERIE DE TIEMPO DE LAS DEFUNCIONES POR AÑOS, 1922-2000
(EN MILES)



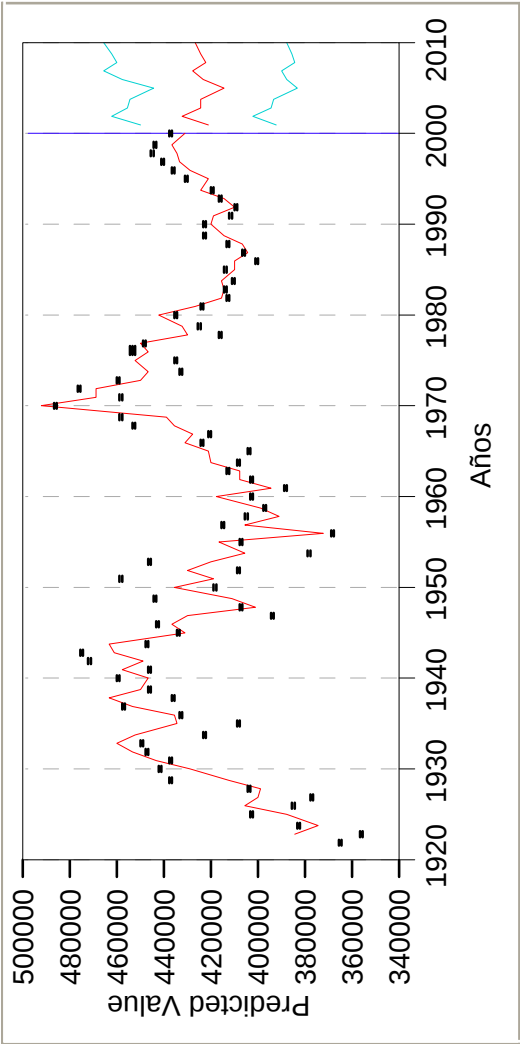
Fuente: cálculos propios del autor.

CUADRO 2
ESTIMACIÓN DE PARÁMETROS

Term	Factor	Lag	Estimate	Std error	t Ratio	Prob > t
AR 1.1	1	1	0.07776571	0.1719478	0.45	0.6528
AR 1.2	1	2	0.88949194	0.230619	3.86	0.0003
AR 1.3	1	3	0.03205953	0.120455	0.27	0.7911
AR 1.4	1	4	0.02130014	0.0563733	0.38	0.7069
AR 1.5	1	5	0.40766058	0.1666807	2.45	0.0175
AR 1.6	1	6	-0.3330903	0.1665446	-2.00	0.0502
AR 1.7	1	7	-0.4875371	0.1759247	-2.77	0.0075
AR 1.8	1	8	0.10613672	0.1765113	0.60	0.5500
AR 1.9	1	9	0.0725369	0.1168426	0.62	0.5372
AR 1.10	1	10	0.1034466	0.115787	-0.89	0.3753
AR 2.16	2	16	0.41403504	0.4512357	0.92	0.3627
MA 1.1	1	1	-0.1977483	0.0865078	-2.29	0.0259
MA 1.2	1	2	0.71030284	0.2207711	3.22	0.0021
MA 1.3	1	3	0.44753127			
MA 1.4	1	4	0.11853384			
MA 1.5	1	5	0.8824446	0.2124666	4.15	0.0001
MA 1.6	1	6	0.23919881	0.0584073	4.10	0.0001
MA 1.7	1	7	-0.8626233	0.1290754	-6.68	<0.0001
MA 1.8	1	8	-0.372373	0.1388327	-2.68	0.0095
MA 2.16	2	16	0.73329669	0.5237667	1.40	0.1668
Intercept	1	0	426 119.902	206.89992	2059.5	<0.0001
Constant Estimate			79 182.7393			

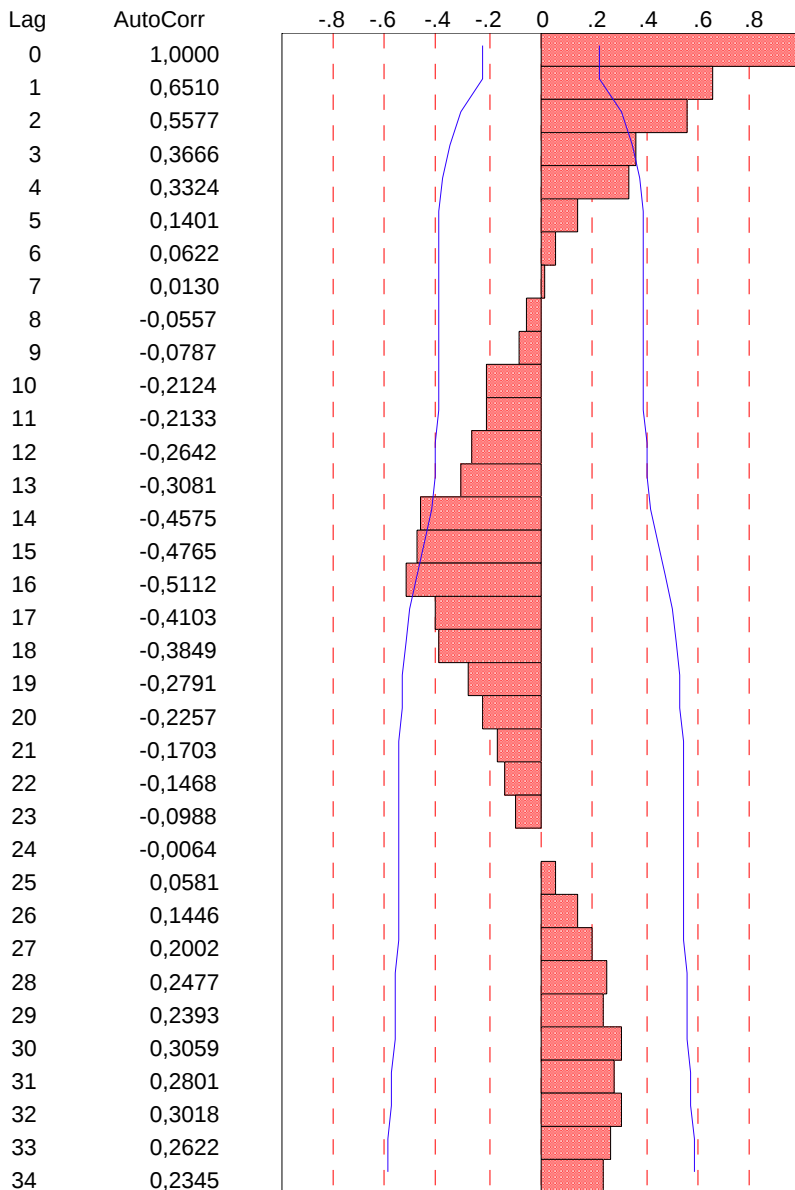
Fuente: cálculos propios del autor.

GRÁFICA 2
PRONÓSTICO



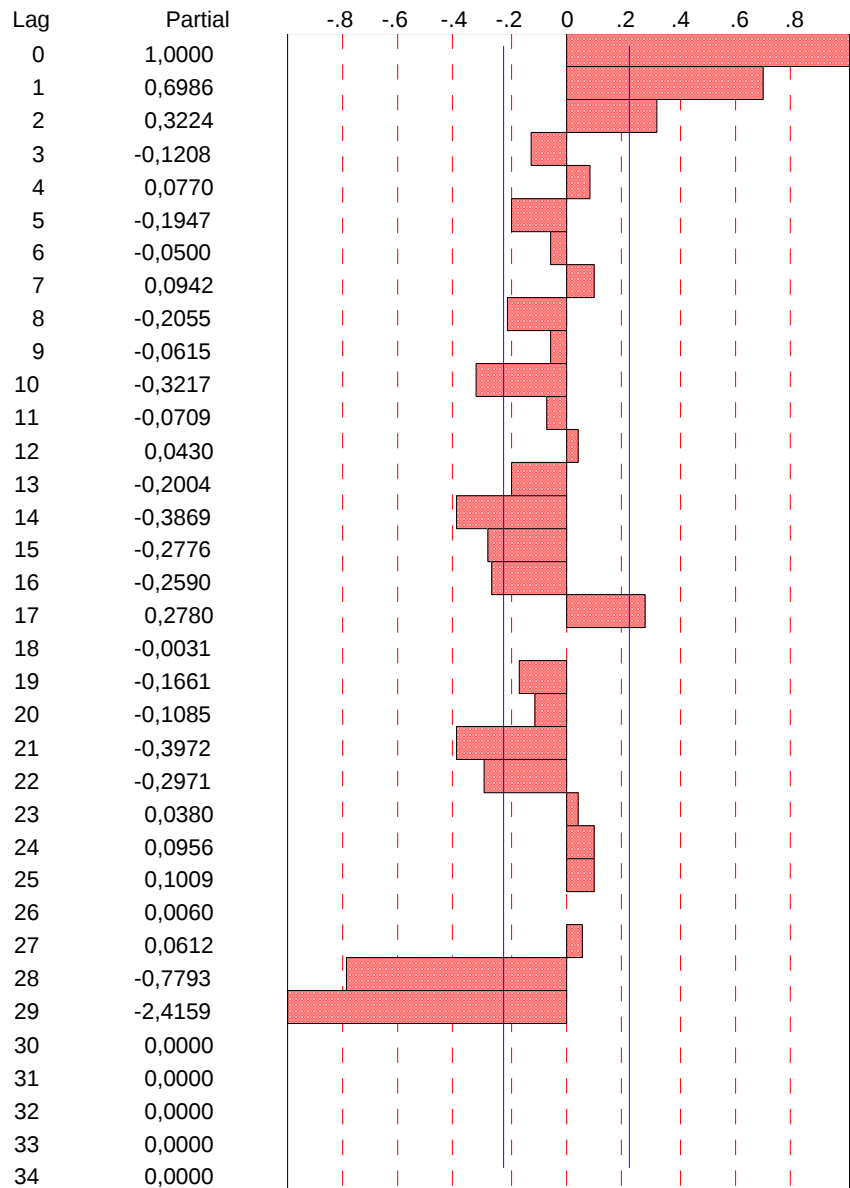
Fuente: cálculos propios del autor.

GRÁFICA 3
FUNCIÓN DE AUTOCORRELACIÓN



Fuente: cálculos propios del autor.

GRÁFICA 4
FUNCIÓN DE AUTOCORRELACIÓN PARCIAL



Fuente: cálculos propios del autor.

CUADRO 3
ESTADOS UNIDOS MEXICANOS: DEFUNCIONES TOTALES, 1922-2000

Años	Defunciones	Años	Defunciones	Años	Defunciones	Años	Defunciones
1922	364 832	1942	471 600	1962	403 046	1982	412 345
1923	356 574	1943	474 950	1963	412 834	1983	413 345
1924	383 129	1944	447 198	1964	408 275	1984	410 550
1925	402 700	1945	433 694	1965	404 164	1985	414 003
1926	384 850	1946	442 935	1966	424 141	1986	400 079
1927	377 046	1947	394 285	1967	420 298	1987	406 102
1928	404 299	1948	407 708	1968	452 910	1988	412 987
1929	437 303	1949	443 559	1969	458 886	1989	423 304
1930	441 712	1950	418 430	1970	485 686	1990	422 803
1931	437 038	1951	458 238	1971	458 323	1991	411 131
1932	447 532	1951	408 823	1972	476 206	1992	409 814
1933	449 394	1953	446 127	1973	458 915	1993	416 335
1934	422 802	1954	378 752	1974	433 104	1994	419 074
1935	408 471	1955	407 522	1975	434 499	1995	430 278
1936	432 762	1956	368 740	1976	453 939	1996	436 321
1937	456 940	1957	414 545	1977	448 523	1997	440 437
1938	436 476	1958	404 529	1978	416 290	1998	444 665
1939	446 216	1959	396 924	1979	425 503	1999	443 950
1940	458 906	1960	402 545	1980	434 465	2000	437 667
1941	446 361	1961	388 855	1981	424 274		

Fuente: Dirección General de Estadística.

Bibliografía

- BOX, George y Gwilym M. Jenkins, 1976, *Time series analysis. Forecasting and control*, Holden Day.
- CHATFIELD, Christopher, 1992, *The analysis of time series: an introduction*, Chapman and Hall, Nueva York.
- CONSEJO NACIONAL DE POBLACIÓN, 2002, *Proyecciones de la población de México: 2000-2050*, México.
- DIEBOLD, Francis, 1999, *Elementos de pronósticos*, International Thomson Editores.
- GUERRERO, Víctor, 1991, *Análisis estadístico de series de tiempo económicas*, Colección CBI, Universidad Autónoma Metropolitana, México.
- HAMILTON, James, 1994, *Time series analysis*, Princeton University Press, Princeton, Nueva Jersey.
- MAKRIDAKIS, Spyros, 1983, *Forecasting: methods and applications*, John Wiley and Sons, Nueva York.
- VANDAELE, Walter, 1983, *Applied time series and Box-Jenkins models*, Academic Press.