

Estimación y Pronósticos de la Mortalidad de Rosario y la Argentina utilizando el modelo de Lee-Carter.

Lucia Andreozzi.

Cita:

Lucia Andreozzi (2009). *Estimación y Pronósticos de la Mortalidad de Rosario y la Argentina utilizando el modelo de Lee-Carter* (Tesis de Licenciatura). FACULTAD DE CIENCIAS ECONOMICAS Y ESTADISTICA ; UNIVERSIDAD NACIONAL DE ROSARIO.

Dirección estable: <https://www.aacademica.org/lucia.andreozzi/34>

ARK: <https://n2t.net/ark:/13683/preH/BRE>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

UNIVERSIDAD NACIONAL DE ROSARIO
FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA
ESCUELA DE ESTADÍSTICA

Título de tesina:

“Estimación y Pronósticos de la Mortalidad de Rosario y la Argentina utilizando el modelo de Lee-Carter”

Tesinista:

Lucía Andreozzi

Directora: Mgs. María Teresa Blaconá

Asesora: Mgs. Nora Arnesi

Agradecimientos

En primer lugar quiero agradecer profundamente a mi directora María Teresa Blaconá por su guía, dedicación y su constante presencia.

A mi asesora Nora Arnesi, por su ayuda y paciencia.

A la Lic. María Rosa Vallasciani por brindarme todos los datos necesarios.

A la Lic. Beatriz Bruera por brindarme su atención y ayuda en todos los trámites.

A Norberto Ferrari y Graciela Ferrucci por su soporte constante, por darme el empleo que posibilitó que estudie en Rosario, y darme el espacio y tiempo para desarrollar mi carrera.

A María Rut Covernton, quien halló el lugar para que viniera a vivir a Rosario.

A Ivana y Estefania, por el largo camino recorrido juntas.

A mi familia por el afecto y la ayuda constante.

INDICE

	Págs.
I.- Introducción	5
II.- Objetivos	7
II.1 Objetivo general	7
II.2 Objetivos específicos	7
III.- Metodología	8
III.1 Conceptos demográficos	8
III.2 Modelo de Lee-Carter	10
III.2.1 Consideraciones iniciales	10
III.2.2 Ajuste del modelo	12
III.2.3 Correcciones al primer ajuste	16
III.3 Pronósticos	17
III.3.1 Modelos ARIMA	17
III.3.1.1 Presentación y etapas	17
III.3.1.2 Pronósticos ARIMA	18
III.3.2 Modelos de espacio de estados	19
III.3.2.1 Modelo básico de espacio de estados	19
III.3.2.2 Pronósticos de espacio de estados	22
III.3.3 Diagnóstico y Selección	24
III.3.4 Evaluación de los pronósticos	26
III.4 Pronósticos de las tasas de mortalidad específicas por edad	26
III.5 Esperanza de vida al nacer	27
IV.- Análisis Empírico	28
IV.1 Argentina	28
IV.1.1 Los datos	28
IV.1.2 Ajuste del modelo	29
IV.1.3 Pronósticos para el índice general de mortalidad de Argentina	34
IV.1.3.1 Modelos ARIMA	34
IV.1.3.2 Modelos de espacio de estados	40
IV.1.4 Pronósticos de las tasas de mortalidad	

específicas por edad	44
IV.1.5 Esperanza de vida al nacer	47
IV.2 Rosario	49
IV.2.1 Los datos	49
IV.2.2 Ajuste del modelo	51
IV.2.3 Pronósticos para el índice general de mortalidad de Rosario	55
IV.2.3.1 Modelos ARIMA	55
IV.2.3.2 Modelos de espacio de estados	59
IV.2.4 Pronósticos de las tasas de mortalidad específicas por edad	64
IV.2.5 Esperanza de vida al nacer	65
V.- Consideraciones Finales	67
VI.- Bibliografía y software	70

Anexo

A.1 Argentina: tabla general de datos	I
A.2 Rosario: tabla general de datos	IV
A.3 Funciones de la tabla de vida abreviada y fórmulas para su cálculo según metodología ONS	VIII
A.4 Tasas de mortalidad de Argentina	IX
A.5 Tasas de mortalidad de Rosario	XII
A.6 Tablas de vida de Argentina	XV
A.7 Tablas de vida de Rosario	XXXIV

I.- Introducción

Durante el siglo 20 la esperanza de vida aumento notoriamente: el Instituto Nacional de Estadísticas y Censos (INDEC) indica que la esperanza de vida al nacer para los argentinos aumentó de 48,5 años en 1914 a 73,8 años en 2001. Por otra parte, la tendencia en las tasas de mortalidad para muchos países industrializados ha estado descendiendo en los últimos años.

Tales mejoras en la mortalidad generalmente se ven de una manera optimista: según las estadísticas se vive más que los antepasados. Pero estos cambios afectan claramente a la asignación de la tasación y de la reserva para las jubilaciones y pensiones, y representan un problema para un sistema de seguridad social que se ha planeado en base a una esperanza de vida más modesta.

El intento de encontrar una curva apropiada de la mortalidad tiene una larga historia en demografía y en las ciencias actuariales. Tradicionalmente se utiliza una curva paramétrica para ajustar las tasas anuales de mortalidad, como las propuestas por DeMoivre (1725), Gompertz (1825), Makeham (1860), Sang (1868) y Weibull (1939).

Durante los últimos diez años, se han desarrollado un gran número de nuevos enfoques para pronosticar la mortalidad usando modelos estocásticos, tales como Alho (1990, 1991), Alho y Spencer (1985, 1990), McNown y Rogers (1989, 1992), Bell y Monsell (1991), y Lee y Carter (1992). Recientemente, el modelo de Lee-Carter se ha convertido en uno de los más populares y es aplicado para realizar pronósticos de las tasas de mortalidad específicas por edad en muchos países. Este modelo es de cálculo sencillo, simple de aplicar y ha dado resultados acertados para varios países, por ejemplo, E.E.U.U. (Lee y Carter 1992), Canadá (Lee y Nault 1993), Chile (Lee y Rofman 1994), Japón (Wilmoth 1996), las siete naciones más desarrolladas económicamente (el G7) (Tuljapurkar y otros 2000) y Bélgica (Brouhns y otros 2001). Aunque en otros países los resultados no fueron tan satisfactorios, como sucedió con los estudios para datos australianos (Booth y otros 2002) y del Reino Unido (Renshaw y Haberman 2003).

El modelo de Lee y Carter y sus variantes han sido utilizadas por los actuarios para una amplia gama de propósitos, desde el pronóstico de los factores de reducción de la mortalidad (Renshaw y Haberman 2003) hasta el cálculo de las rentas de los seguros de retiro (Chia y Tsui 2003). Otros usos en el área demográfica incluyen proyecciones de población (Booth and Tickle 2003), el pronóstico de las diferencias en la mortalidad según el sexo (Lee y Carter 1992), y la proyección de los patrones de la mortalidad para “el más viejo de los viejos” (Buettner 2002).

Intrínsecamente el modelo asume que la dinámica de las tasas de mortalidad a través del tiempo está manejada sólo por el parámetro del modelo que varía en el tiempo, llamado índice de mortalidad. El pronóstico de la mortalidad se basa en la extrapolación de dicho índice a través de la selección de un modelo de series de tiempo apropiado.

Comúnmente se utilizan para pronosticar los modelos de Box-Jenkins comúnmente conocidos como procesos autorregresivos promedio móvil, ARIMA (Box y Jenkins, 1976). Para aplicar una metodología de pronóstico alternativa Lee y Carter (1996) desarrollan modelos estructurales de series de tiempo o también llamados Modelos de Espacio de Estados (MEE) (Harvey, 1989) que difieren principalmente de sus análogos ARIMA en el hecho de que introducen parámetros que varían en el tiempo en sus ecuaciones de estimación. Al utilizar ambas metodologías, Lee y Carter comparan los modelos estructurales, con sus contrapartes ARIMA para ver si mejoran su desempeño en términos de ajuste y bondad de pronóstico.

En la Argentina, a nivel oficial, el INDEC realiza proyecciones de la mortalidad las cuales se utilizan en la elaboración de proyecciones de la población por sexo y grupos quinquenales de edad a partir de una población base aplicando el Método de las Componentes (Naciones Unidas, 1956). Dicho método consiste en proyectar en forma independiente las variables determinantes de la dinámica poblacional: mortalidad, fecundidad y migración.

Según se explica en el documento “Estimaciones y proyecciones de población.Total del país. 1950-2015”, realizado por el INDEC, las últimas proyecciones de población para el período 2000-2015, se basan en una única hipótesis de evolución para la mortalidad, debido a que no se esperan variaciones importantes en esta componente. Para proyectar la mortalidad a

partir del quinquenio 2000-2005 se mantuvo una hipótesis de cambio basada en criterios de las Naciones Unidas para fijar la ganancia en años de vida hasta el quinquenio 2045-2050, lo que permite determinar los valores esperados de esperanza de vida al nacimiento para cada quinquenio de la proyección, de acuerdo a una ganancia media (Pujol, 1995).

En el presente trabajo se presenta un método alternativo al utilizado por el INDEC para proyectar la mortalidad, el modelo de Lee-Carter. Este modelo que se ha implementado en numerosos países a lo largo del mundo, se aplica en esta tesina a datos de la República Argentina y de la Ciudad de Rosario.

II.- Objetivos

II.1 Objetivo General

El presente trabajo tiene como principal objetivo, aplicar y estimar el modelo de Lee-Carter, sobre tasas de mortalidad específicas por edad, de Argentina y Rosario, para obtener un índice general de mortalidad para hombres y otro para mujeres, para luego realizar pronósticos de dicho índice mediante modelos de series de tiempo ARIMA (Box & Jenkins, 1970) y MEE (Harvey, 1989).

II.2 Objetivos específicos

-Describir y analizar los parámetros estimados a partir del modelo de Lee-Carter, para varones, mujeres y total, de Rosario y de Argentina.

-Comparar los modelos de pronóstico propuestos.

-Derivar a partir de los pronósticos del índice general de mortalidad, estimados para ambos sexos y para el total en Rosario y Argentina, los pronósticos de las tasas de mortalidad específicas por edad.

-Calcular las esperanzas de vida según género y región geográfica.

III.- Metodología

En esta tesina se estudia y estima el modelo de Lee-Carter (1992), el cual genera un índice general de mortalidad, denominado k , a partir de las tasas de mortalidad por grupos de edad. La estimación se realiza para varones, mujeres y total de la población. Se desarrollan, para pronosticar cada uno de los índices, modelos de series de tiempo ARIMA y MEE. Los mismos son comparados tanto en su bondad de ajuste como en capacidad de pronóstico. A partir de los pronósticos del índice general de mortalidad, se obtienen los pronósticos de las tasas de mortalidad y la esperanza de vida.

III.1 Conceptos demográficos

En esta sección se enuncian en forma abreviada algunos conceptos demográficos utilizados en esta tesina.

En primer lugar se presenta un concepto básico que se necesita para el desarrollo del modelo de Lee-Carter:

Tasas específicas de mortalidad según edad

El conocimiento del número de casos de algún evento en particular, en este caso muerte, presentes en una población, tiene de por sí poca utilidad para epidemiólogos o demógrafos si no se relaciona dicha frecuencia con la población de la cual proceden los casos. Esta relación se establece generalmente a través de la construcción de una tasa.

El término tasa se usa en muchos campos y su significado no es consistente en todos ellos. En demografía, las tasas se definen comúnmente como tasas de *ocurrencia/exposición* (Preston et al., 2001).

El numerador de este tipo de tasas contabiliza el número de ocurrencias de un evento de interés, mientras que el denominador combina dos factores: el número de personas en la población y la longitud del tiempo que enmarca el estudio. El denominador de una tasa de *ocurrencia/exposición* se conoce con el nombre de *personas-tiempo en riesgo* y es la suma del tiempo que cada persona permanece bajo observación y en riesgo de convertirse en un caso. El tiempo debe expresarse en unidades de medida apropiadas, por ejemplo personas-año, en cuyo caso se dice que la tasa está anualizada.

Cuando los casos se obtienen a partir del registro de estadísticas vitales no se dispone de medidas directas de personas-tiempo en riesgo. Una estimación de dicho valor, para una población en un período dado, se puede calcular como el producto entre el número total de personas que componen la población en el punto medio del período de interés y la duración total del período. Este método proporciona estimaciones adecuadas siempre que la población permanezca estable durante ese período.

Las tasas pueden calcularse para una población entera o para subgrupos específicos de la población. En el primer caso se las denomina tasas crudas o brutas y en el segundo, tasas específicas por subgrupo de interés. Por ejemplo en el estudio de la mortalidad resulta de interés el cálculo de tasas específicas por sexo y edad.

La tasa específica de mortalidad según edad se define como el cociente entre el número de defunciones acaecidas en un grupo de edad específica de la población de un área geográfica dada durante un período determinado y el correspondiente valor de personas-tiempo en riesgo en ese grupo específico de edad del área geográfica y período bajo estudio. Si las defunciones se obtienen a partir de Registros de Estadísticas Vitales, el denominador se estima a partir de los datos censales.

Esperanza de vida al nacer

La esperanza de vida al nacer es una medida resumen apta para comparar la mortalidad de diferentes poblaciones y para la misma población en el tiempo ya que no está afectada por la estructura por edades de la población.

La esperanza de vida a determinada edad es una estimación del número promedio de años que le restaría vivir a una persona si las condiciones de la mortalidad actuales permaneciesen constantes. Se calcula tomando como base las tasas de mortalidad específicas por edad. Es una medida hipotética y un buen indicador de las condiciones de salud. Para obtenerla se requiere de la elaboración de una tabla de vida que proporciona estimaciones de la esperanza de vida a diferentes edades. Sin embargo como indicador del nivel de la mortalidad se acostumbra a utilizar la esperanza de vida al nacer, es decir, a la edad 0. La esperanza de vida al nacer se define como el número promedio de años que vivirían cada integrante de una cohorte hipotética de personas que permaneciese sujeta a la mortalidad imperante en la población en estudio desde su nacimiento hasta su extinción. (CELADE, 1989)

III.2 El modelo de Lee-Carter

El modelo de Lee y Carter (1992), denotado como LC de aquí en adelante, es un modelo estadístico-demográfico que permite realizar proyecciones de las tasas de mortalidad. En términos generales, se puede definir como un modelo de carácter extrapolativo, ya que no incorpora información acerca de efectos tecnológicos o sociales en la mortalidad, esto es, no busca incorporar información externa, ni opiniones sobre posibles acontecimientos futuros. El modelo básicamente proyecta la tendencia histórica observada durante el periodo de estudio en el que se dispone de datos. El modelo LC combina un modelo demográfico de pocos supuestos con métodos estadísticos de series de tiempo, por lo cual es capaz de proveer una base probabilística a las proyecciones, así como intervalos de predicción. El cambio en el nivel de mortalidad general es representado en el modelo por el cambio de un solo índice. El método que resulta del modelo LC permite a cada tasa específica de mortalidad decrecer exponencialmente sin restricciones.

Para utilizar el modelo LC se requiere inicialmente revisar la calidad de la información disponible y sus limitaciones, ya que el único insumo es un registro histórico de tasas de mortalidad por grupos de edad. Es importante señalar que el modelo, por construcción, tiene la ventaja de que las tasas no

necesitan tener la misma periodicidad. No obstante, se requiere que los datos abarquen un periodo considerable de tiempo, una regla propuesta por Li et al. (2004) es que cubran al menos un periodo de 30 años de historia reciente.

Una vez que se han obtenido los datos y se ha validado su calidad, la etapa siguiente del método consiste en ajustar el modelo a los datos originales y evaluar su desempeño histórico. Posteriormente, a través de métodos de series de tiempo se pronostica el índice general de mortalidad y, finalmente, se pronostican las tasas de mortalidad específicas por edad.

III.2.1 Consideraciones iniciales

La premisa básica del modelo es que existe una relación lineal entre el logaritmo de las tasas específicas de mortalidad $m_{x,t}$ y dos factores explicativos: el intervalo de edad inicial x , y el tiempo, t . La ecuación que describe esto es

$$m_{x,t} = \exp(a_x + b_x k_t + e_{x,t}), \quad t=1, \dots, n \quad x=1, \dots, \omega \quad (\text{III.2.1.1})$$

o aplicando logaritmo

$$f_{x,t} = \ln(m_{x,t}) = a_x + b_x k_t + e_{x,t}, \quad t=1, \dots, n \quad x=1, \dots, \omega \quad (\text{III.2.1.2})$$

donde,

k_t : índice general de mortalidad.

$m_{x,t}$: tasa específica de mortalidad para el intervalo x y año t .

a_x : parámetro de forma, el conjunto de dichos parámetros describe el patrón de mortalidad de las tasas.

b_x : parámetro de sensibilidad. Representa el cambio en la mortalidad en el intervalo que se inicia a la edad x , frente a cambios en el índice k_t

$e_{x,t}$: error aleatorio.

ω : es el inicio del último intervalo de edad.

Con $m_{x,t}$ se representa al cociente del número de defunciones de personas cuya edad está incluida en el intervalo que comienza con la edad x ,

con respecto a la población de la misma edad a la mitad del año t . En el presente trabajo se utiliza la edad al inicio del intervalo (x) para referirse a cada uno de los grupos de edad; esta nomenclatura es la que utilizan Lee y Carter en su investigación. Si se menciona el “intervalo 5” se refiere al que comienza con la edad de cinco años.

A a_x se lo denomina parámetro de “forma”. El conjunto de las estimaciones de dicho parámetro para cada uno de los grupos etáreos describe la forma general o el patrón de las tasas de mortalidad específicas por edad. Dicho parámetro se estima como el promedio aritmético simple sobre el tiempo, de los logaritmos de las tasas específicas

El parámetro b_x , llamado de “sensibilidad” describe el cambio en la mortalidad en el intervalo x , frente a cambios en el índice k_t . El parámetro de sensibilidad representa la intensidad en el crecimiento o decrecimiento de la tasa de mortalidad, para un grupo de edad a través del tiempo.

En el modelo LC el índice estimado k_t es lineal lo cual hace que la mortalidad en cada grupo de edad cambie a su propia tasa exponencial. Conforme k_t tiende a valores negativos grandes ($k_t \rightarrow -\infty$), cada tasa específica tiende a 0. Por lo tanto en este modelo no pueden presentarse valores negativos de las tasas de mortalidad. Al hacer todas las tasas de mortalidad dependientes del parámetro k_t , el método difiere mucho de pronosticar cada tasa específica de mortalidad de manera independiente. Esto evita tener que calcular las distintas covariancias entre los errores. Además, todas las tasas individuales están restringidas a pertenecer al sistema general de mortalidad que se ajusta a los datos históricos observados.

El término $e_{x,t}$, representa el error aleatorio, con media 0 y varianza σ^2 , que refleja entre otros factores influencias históricas no capturadas por el modelo.

III.2.2 Ajuste del modelo

A partir de la expresión (III.2.1.1), se observa que hay una ecuación para cada intervalo de edad y cada tiempo, en consecuencia se está en presencia de un sistema de ecuaciones simultáneas que se tiene que resolver para

estimar los valores a_x , b_x y k_t que sean soluciones del sistema. Para ello se debe notar que un conjunto de tasas de mortalidad, con r grupos de edad distintos y observados en n momentos distintos, define un sistema de ecuaciones con $2r+n$ incógnitas correspondientes a la suma de r distintas a_x , r distintas b_x , y n distintas k_t y $r \times n$ ecuaciones. Este sistema se puede escribir matricialmente como

$$M = A + b.k, \quad (\text{III.2.2.1})$$

donde M es una matriz de dimensión $r \times n$ en la que cada elemento $M_{i,j}$, corresponde al logaritmo natural de la tasa de mortalidad para el grupo de edad i en el año j . M presenta la forma:

$$M = \begin{bmatrix} \ln(m_{11}) & \ln(m_{12}) & \dots & \ln(m_{1n}) \\ \ln(m_{21}) & \ln(m_{22}) & \dots & \ln(m_{2n}) \\ \vdots & \vdots & \vdots & \vdots \\ \ln(m_{r1}) & \ln(m_{r2}) & \dots & \ln(m_{rn}) \end{bmatrix}.$$

Por su lado, A es una matriz de tamaño $r \times n$ en la que todos los elementos que corresponden al mismo momento j son iguales, por ejemplo $a_{1j} = a_{2j} = \dots = a_{rj}$. A presenta la forma:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{r1} & a_{r2} & \dots & a_{rn} \end{bmatrix}.$$

Mientras que b es un vector de dimensión $r \times 1$ y k es un vector $1 \times n$:

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{bmatrix}, \quad k = [k_1 \quad k_2 \quad \dots \quad k_n].$$

En principio, no existe solución única para este sistema; por ejemplo, si A , b , k son una solución del sistema y p es un vector $1 \times n$ con todos sus

elementos iguales a un escalar $c \neq 0$, entonces las matrices $A - b.p$, b , $k+p$ son también una solución. Igualmente, si los arreglos A , b , k son una solución y $c \neq 0$, entonces A , $b.c$, $c^{-1}.k$ también son solución del sistema de ecuaciones.

Para obtener una solución única, es necesario agregar las siguientes dos restricciones: $\sum_{x=1}^{\omega} b_x = 1$ y $\sum_{t=1}^n k_t = 0$. Al utilizar estas restricciones, los coeficientes a_x resultan ser un promedio aritmético simple sobre el tiempo, de los logaritmos de las tasas específicas, con lo cual los parámetros b_x y k_t se determinan de forma única. Por lo tanto, los coeficientes a_x se determinan mediante la siguiente expresión

$$a_x = \frac{\sum_{t=1}^n \ln(m_{x,t})}{n} . \quad (\text{III.2.1.2})$$

Una vez determinados los valores de la matriz A , el sistema (III.2.2.1) se puede reescribir como

$$M^* = M - A = b . k. \quad (\text{III.2.1.3})$$

con las restricciones ya incluidas, este sistema cuenta con una solución única y sólo quedan por determinarse los vectores de los parámetros b y k . Aparentemente, estos vectores podrían ser estimados mediante Mínimos Cuadrados Ordinarios, sin embargo, dado que no hay una variable observable en el lado derecho de la ecuación (III.2.1.3), no es posible usar dicho método para estimar los parámetros. Se podría entonces utilizar algún método simple que brinde valores aproximados para los estimadores de los vectores de parámetros. El método de Descomposición en Valores Singulares (DVS) proporciona un ajuste exacto de Mínimos Cuadrados (M. L. Good, 1969). El método de DVS consiste en descomponer la matriz M en tres matrices, como sigue

$$M = U.S.V^T, \quad (\text{III.2.1.4})$$

donde M es la matriz original de dimensión $r \times n$, U es una matriz $r \times r$ ortonormal (es decir, sus columnas son independientes y tienen norma unitaria). La matriz S es diagonal de dimensión $r \times n$, donde cada uno de los elementos en la

diagonal corresponde a un valor singular mayor o igual que cero de la matriz original, estos valores aparecen ordenados en forma decreciente, de tal manera que los últimos elementos tendrán un valor absoluto relativamente pequeño respecto a los primeros o podrán incluso llegar a ser iguales a cero. Finalmente, V^T es la transpuesta de la matriz V , la cual es una matriz ortonormal de dimensión $n \times n$. Si se excluyen las posibles permutaciones de los renglones de U o de S , como también las combinaciones lineales de éstos, la DVS es única para una matriz M dada.

La descomposición se realiza sobre la matriz M^* (ya que tiene solución única). Si ahora se definen $B = U$ y $K^T = S.V^T$, la ecuación (III.2.1.4) se convierte en

$$M^* = B \cdot K^T . \quad (\text{III.2.1.5})$$

Es importante señalar que B y K son matrices, a diferencia de los vectores b y k planteados originalmente. Como la DVS descompone a M^* en el producto de dos matrices, el elemento (i, j) en esta matriz resulta ser la suma del producto del renglón i de B y del renglón j de K , esto es

$$m_{i,j} = \sum_{l=1}^r B_{i,l} K_{j,l}^T . \quad (\text{III.2.1.6})$$

Por lo tanto, la descomposición produce r términos que ajustan exactamente el elemento $m_{i,j}$ de la matriz M^* . Los términos sucesivos tendrán menor valor absoluto (debido a las propiedades de las matrices S , U y V originales) por lo que, si la mayoría de los elementos en la diagonal de S son iguales o cercanos a cero, se puede lograr una buena aproximación de cada elemento, además a medida que se incrementa el número de términos usados, el grado de precisión en el ajuste mejorará sólo marginalmente. En particular, el primer término $B_{i,1} K_{j,1}^T$ constituye la mejor aproximación de una dimensión para cada elemento (i, j) . Por lo cual, en este paso se debe buscar el menor número de términos que den una aproximación razonable a la matriz M^* . Si bien un mayor número de términos proporcionan un mejor ajuste a los datos, puede que no exista una razón adicional en el modelo para incluirlos, como señalan

Lee y Carter (1992). Por ello aconsejan descomponer M^* mediante un producto de los vectores b y k . Es por ello que, al realizar la DVS sólo consideraron la aproximación de primer orden, es decir, M se aproxima como

$$M' \approx B_1 K_1^T. \quad (\text{III.2.1.7})$$

Así quedan determinados $b_1 = B$ y $k_1 = K$, con lo cual se obtiene una primera estimación de los parámetros del modelo.

III.2.3 Correcciones al primer ajuste

En general, los valores obtenidos en la primera estimación del modelo no van a proporcionar un buen ajuste a los datos observados, por lo que se pueden producir desviaciones en las proyecciones, como lo señalan Lee y Carter, y posteriormente Bell (1997). Es por ello que se propone realizar un segundo paso en la estimación de los parámetros. Este consiste en aprovechar los valores a_x y b_x del primer paso, para buscar una nueva estimación de k , de tal manera que, para una distribución de población específica, se produzca exactamente el número observado de muertes totales para el año en cuestión. Se trata de encontrar los valores k_t tales que cumplan con la condición

$$D_t = \sum_{x=0}^{\omega} N_{x,t} \exp(a_x + b_x k_t + e_{x,t}), \quad (\text{III.2.1.8})$$

donde D_t denota el número de muertes totales en el año t , $N_{x,t}$, es la población perteneciente al intervalo de edad x en el año t y ω es la edad del último grupo observado en la tabla de mortalidad.

La segunda estimación del parámetro k_t si bien se puede considerar apropiada, no es óptima y es por ello que se han propuesto diversas alternativas (véase al respecto Wilmoth, 1993, Bell, 1997, Alho, 2000, y Booth et al., 2003). En particular, estas alternativas se plantearon a partir de que se observó que el método original presentaba algunos problemas, como es la introducción de sesgo en las proyecciones. Bell (1997) muestra empíricamente el sesgo que se produce en las proyecciones de la esperanza de vida si se utiliza el método planteado por Lee y Carter para la segunda estimación de k_t , por lo que propuso como alternativa utilizar como estimador de a_x el logaritmo de la última tasa observada y no el promedio de los logaritmos de las tasas;

otra alternativa que propuso Bell es sumar el sesgo a los valores estimados, para que se vuelvan insesgados. La alternativa propuesta por Lee y Miller (2000) consiste en hacer la segunda estimación respecto a la esperanza de vida, en lugar de realizar el ajuste respecto al total de defunciones. Por otro lado, Wilmoth (1993) desarrolló un método en el que se usa la descomposición en valores singulares (DVS) con pesos para ajustar el modelo, donde cada peso es el número de muertes correspondientes a cada edad. De esta manera, sólo se necesita un paso para la estimación de los valores requeridos por el modelo.

III.3 Pronósticos

A continuación se exhiben los dos modelos que se desarrollan con el fin de obtener pronósticos del índice general de mortalidad. Se presenta una breve reseña acerca de modelos ARIMA y las etapas que comprenden el proceso de modelización. Finalmente se presenta el pronóstico ARIMA y su correspondiente intervalo de confianza. Luego se introduce a los MEE y sus pronósticos.

III.3.1 Modelos ARIMA

III.3.1.1 Presentación y etapas de la modelización

Una serie de tiempo observada y_1, y_2, \dots, y_n se puede considerar una realización particular de un proceso estocástico, es decir de un fenómeno que se desarrolla a través del tiempo según leyes probabilísticas.

La familia de modelos autorregresivos promedios móviles integrados (ARIMA), formalizada por Box-Jenkins (1976) contiene modelos de la forma:

$$(1 - B)^d \phi(B)(y_t - \mu) = \theta(B)a_t,$$

donde B es el operador de rezago ($B^k=y_{t-k}$ para $k=0, \pm 1, \pm 2, \dots$).

Este modelo se indica como ARIMA (p,d,q) donde:

d es el número de diferencias ordinarias necesarias para que el proceso sea estacionario,

μ es la media del proceso

$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ es el polinomio autorregresivos de orden p,

$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ es el polinomio promedio móvil de orden q,

Los dos polinomios deben tener sus raíces fuera del círculo unitario para que se cumplan las condiciones necesarias de estacionariedad e invertibilidad respectivamente. a_t es una serie ruido blanco. Es decir tiene media cero, variancia constante y valores no correlacionados.

La modelización es un proceso iterativo que consta de tres etapas:

i) Identificación: se realiza en base al gráfico de la serie, la función de autocorrelación muestral y la función de autocorrelación parcial muestral. A partir de ellas se determinan las transformaciones requeridas, el número de diferencias y los órdenes p y q del modelo.

ii) Estimación: Existen distintos métodos para la estimación de los parámetros del modelo identificado. En primer lugar existe el método de estimación por máxima verosimilitud exacta que se utiliza en el caso de los modelos autorregresivos. Cuando los modelos incluyen términos promedios móviles, este método no funciona satisfactoriamente, por lo que se utilizan métodos de estimación iterativos, como máxima verosimilitud condicional y no condicional. Existe también el método de mínimos cuadrados ordinarios, del mismo modo que el método de máxima verosimilitud exacta, este método no se desempeña eficazmente si se cuenta con términos promedios móviles en el modelo.

iii) Diagnóstico y Verificación: una vez estimado el modelo se debe probar si dicho modelo es adecuado verificando si se cumplen los supuestos del mismo.

III.3.1.2 Pronósticos ARIMA

Una vez que se selecciona el modelo más adecuado, se realizan las proyecciones. Cada proyección tiene una distribución probabilística y por lo tanto se pueden construir los intervalos de confianza correspondientes.

Las proyecciones con error cuadrático medio mínimo están dadas por la esperanza condicional:

$$\hat{Z}_n(h) = E(Z_{n+h} / Z_n, Z_{n-1}, \dots).$$

$\hat{Z}_n(h)$ es el pronóstico h pasos hacia delante con origen n de Z_{n+h} ,

Siendo

$$Z_{n+h} = \hat{Z}_n(h) + e_h$$

donde e_h se denomina error de pronóstico.

Para un proceso normal, es decir que los errores se distribuyen normales con media cero y variancia constante, se tiene que:

$$e_h \sim N(0, \text{Var}(e_h))$$

En consecuencia los intervalos de confianza del 95% se definen como:

$$\hat{Z}_n(h) \pm 1,96 \cdot \text{Var}(e_h)$$

III.3.2 Modelo de Espacio de Estados. (MEE)

Los modelos de espacio de estados, también conocidos como modelos estructurales, pueden ser interpretados como regresiones sobre funciones del tiempo en las cuales los parámetros varían. Esto hace que sean un vehículo natural para tratar series con cambios de nivel y/o tendencia.

III.3.2.1 Modelo básico de espacio de estados

El modelo básico de espacio de estados, que también se conoce como modelo lineal Gaussiano de espacio de estados, se puede expresar como

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim NID(0, \mathbf{H}_t) \quad (\text{III.3.2.1.1})$$

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{R}_t \boldsymbol{\eta}_t \quad \boldsymbol{\eta}_t \sim NID(0, \mathbf{Q}_t) \quad (\text{III.3.2.1.2})$$

donde \mathbf{y}_t es un vector de orden $p \times 1$ de observaciones, $\boldsymbol{\alpha}_t$ es un vector de orden $m \times 1$ no observable denominado vector de estado, las matrices $\mathbf{Z}_t(p \times m)$, $\mathbf{T}_t(m \times m)$ y $\mathbf{R}_t(m \times g)$ son conocidas y $\boldsymbol{\eta}_t$ es un vector aleatorio de orden $g \times 1$. \mathbf{R}_t se denomina matriz de selección y está formada de ceros y unos de acuerdo a si los η son determinísticos o aleatorios.

A la primera ecuación de (III.3.2.1.1) por lo general se la llama ecuación de medida y a la segunda (III.3.2.1.2), ecuación de transición. En las secciones siguientes se especificará un modelo estructural adecuado para las series de tasas de mortalidad.

Modelo estructural básico

Dada una serie de tiempo y_1, \dots, y_n , el modelo estructural básico se formula en términos de los componentes tendencia e irregular. El modelo se puede escribir

$$y_t = \mu_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad t=1, \dots, n,$$

$$\text{Nivel:} \quad \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_{1t}, \quad \eta_{1t} \sim \text{IN}(0, \sigma_{\eta_1}^2) \quad (\text{III.3.2.1.3})$$

$$\text{Pendiente:} \quad \beta_t = \beta_{t-1} + \eta_{2t}, \quad \eta_{2t} \sim \text{IN}(0, \sigma_{\eta_2}^2)$$

donde μ_t , y β_t representan la tendencia y ε_t la componente irregular. Se considera que los errores η_{1t} y η_{2t} independientes entre si y de ε_t .

Las variancias de los errores $\sigma_{\varepsilon_1}^2$, $\sigma_{\eta_1}^2$ y $\sigma_{\eta_2}^2$ son llamadas hiperparámetros del modelo y pueden ser examinadas si el correspondiente componente de nivel o pendiente son estocásticos o determinísticos. La

significación de las variancias de los componentes se estima a través de los test t (M. Kendall and J. K. Ord 1990).

La idea central de los modelos estructurales se basa en el hecho de que, en muchas aplicaciones, la tendencia conforma el aspecto más destacable de las series de tiempo. Es por ello, que la construcción del modelo está orientada a la estimación y análisis de este componente.

La especificación de los componentes μ_t y β_t se basa en el conocimiento que se tenga acerca del proceso que se analiza y en técnicas estadísticas.

Estimación de los componentes no observables

Para estimar las diferentes componentes que intervienen en el modelo se deben realizar varias etapas, que brevemente se especifican a continuación:

- a) *Filtrado*, es una operación donde se va actualizando el sistema cada vez que aparece una nueva observación y_t . Este filtrado se realiza mediante el conocido filtro de Kalman.
- b) *Iniciación del filtro*, se debe especificar cómo comienza el filtro, es decir qué valores iniciales toman las componentes.
- c) *Suavizado*, considera la estimación de los componentes tomando información de la muestra completa.

Estimación de los hiperparámetros por máxima verosimilitud

Los parámetros en los modelos de espacio de estados usualmente se denominan hiperparámetros para distinguirlos de los elementos del vector de estado, los cuales pueden ser pensados como parámetros aleatorios.

Sea el vector de hiperparámetros φ , en el caso particular del modelo (III.3.2.1.3) $\varphi = (\sigma_{\varepsilon_1}^2, \sigma_{\eta_1}^2, \sigma_{\eta_2}^2)$, el cual se estima por máxima verosimilitud. Una forma clásica de la función de verosimilitud para el caso de n variables Y_1, \dots, Y_n independientes e idénticamente distribuidas está dada a partir de la función de densidad conjunta

$$L(Y, \varphi) = \prod_{t=1}^n p(y_t) \quad (\text{III.3.2.1.4})$$

Una vez que la muestra está dada puede reinterpretarse como una función de verosimilitud. Como en series de tiempo por lo general las observaciones no son independientes, no se puede usar directamente (III.3.2.1.4) para el cálculo de la verosimilitud, pero en este caso se puede obtener utilizando propiedades de las distribuciones condicionales como

$$L(Y, \varphi) = \prod_{t=1}^n p(y_t / Y_{n-1}, \varphi), \quad (\text{III.3.2.1.5})$$

Tomando logaritmo de (III.3.2.1.5) se obtiene la descomposición del error de predicción del logaritmo de la verosimilitud y debe ser maximizada respecto de los elementos del vector φ de hiperparámetros desconocidos. También puede existir una reparametrización para reducir la dimensión de φ . Esto se puede realizar reparametrizando de forma tal que $\varphi = (\varphi, \varphi^*)$, donde φ^* tiene dimensión reducida. Se llama función de verosimilitud concentrada (Harvey, 1989) a la construida con esta nueva parametrización, porque la optimización se realiza en relación a φ^* en lugar de a φ .

Para la maximización básicamente existen tres métodos que sirven como base de la mayoría de los algoritmos desarrollados: a) Método de “scoring”, b) Método de Quase-Newton y c) Algoritmo EM. El método de optimización empleado por STAMP, que es el software utilizado en el presente análisis, se basa en el método llamado BFGS Quase-Newton (Koopman et al. 1995).

III.3.2.2 Pronósticos con modelos de espacio de estados

En el modelo compuesto por las ecuaciones (III.3.2.1.1) y (III.3.2.1.2), el filtro de Kalman conduce a_n , el estimador de error mínimo cuadrático de α_T basado en todas las observaciones de la muestra. Además se sabe que

$$a_{n+1/n} = T_{n+1} a_n, \quad (\text{III.3.2.3.1})$$

entonces la predicción una etapa hacia adelante viene dada por

$$\tilde{y}_{n+1/n} = Z_{n+1} a_{n+1/n} \quad (\text{III.3.2.3.2})$$

Ahora si se considera el problema de predicción varias etapas hacia delante, es decir realizar las predicciones de valores futuros en los momentos $n+2, n+3, \dots, n+h$, donde h es el horizonte de predicción, sustituyendo repetidamente en la ecuación de transición, en el momento $n+h$ se tiene

$$a_{n+h} = \left(\prod_{j=1}^k n_{n+j} \right) a_n + \sum_{i=1}^{h-1} \left(\prod_{j=i+1}^h n_{n+j} \right) (R_{n+j} \eta_{n+j}) + R_{n+h} \eta_{n+h} \quad h=2,3,\dots \quad (\text{III.3.2.3.3})$$

Por otro lado, se conoce que el estimador de ECM de a_{n+h} . Aplicando esperanza condicional al momento n a (III.3.2.3.3) se tiene

$$E(a_{n+h} / n) = a_{n+h/n} = \left(\prod_{j=1}^h n_{n+j} \right) a_n. \quad (\text{III.3.2.3.4})$$

La distribución condicional de a_{n+h} es Gaussiana y su matriz de variancia y covariancias $P_{n+h/n}$, se puede obtener de (III.3.2.3.3) y (III.3.2.3.4). En el caso de que sea invariante en el tiempo la expresión apropiada es

$$P_{n+h/n} = n^h P_n n^h + \sum_{j=0}^{h-1} n^j R Q R' (n^j)', \quad h=1,2,\dots \quad (\text{III.3.2.3.5})$$

El estimador ECM de y_{n+h} se puede obtener directamente de $a_{n+h/n}$, tomando esperanza condicional a la ecuación de medida en el momento $n+h$, resultando

$$E(y_{n+h} / n) = \tilde{y}_{n+h/n} = Z_{n+h} a_{n+h/n}, \quad h=1,2,\dots \quad (\text{III.3.2.3.6})$$

La matriz de ECM es

$$ECM(\tilde{y}_{n+h/n}) = Z_{n+h} P_{n+h/n} Z'_{n+h} + H_{n+h}, \quad h=1,2,\dots \quad (\text{III.3.2.3.7})$$

Se debe recordar que las matrices de ECM, $P_{n+h/n}$ no tienen en cuenta el error que surge de estimar los parámetros desconocidos del sistema de matrices n_t y otros parámetros.

III.3.3 Diagnóstico y Selección

Previo a realizar los pronósticos tanto mediante los modelos ARIMA como los MEE se debe comprobar la bondad de los modelos estimados y seleccionar el más conveniente.

Diagnóstico

a) Para comprobar el supuesto de normalidad de los errores existen varios test de hipótesis, uno de ellos, el que se utiliza en el presente trabajo, es el Test de Jarque-Bera:

Se estima:

El coeficiente de asimetría

$$SK = \hat{m}_3 / \hat{m}_2^{3/2}$$

El coeficiente de curtosis

$$K = \hat{m}_4 / \hat{m}_2^2,$$

donde

$$\hat{m}_v = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^v.$$

La hipótesis nula que se plantea es:

H_0) normalidad (y no correlación de los $\hat{\varepsilon}_t$).

$$SK^* = \left(\frac{n}{6}\right) \cdot SK \sim N(0,1) \text{ bajo } H_0$$

$$K^* = \left(\frac{n}{24}\right)^{1/2} \cdot (K - 3) \sim N(0,1) \text{ bajo } H_0$$

} Independientes entre sí

El test se calcula entonces como

$$JB = (SK^{*2} + K^{*2}) \sim \chi_2^2 \text{ bajo } H_0$$

Si se rechaza H_0 puede suceder que exista alguna observación aberrante y/o que el proceso no sea homocedástico.

b) Para comprobar el supuesto de no correlación de los residuos se calcula el test portmanteu de falta de ajuste, que utiliza las funciones de autocorrelación muestral de los residuos, planteando la hipótesis nula conjunta:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0,$$

contra la alternativa que alguno es distinto de cero.

La estadística que se plantea es:

$$Q = n \cdot (n + 2) \sum_{k=1}^K (n - k)^{-1} \cdot \hat{\rho}_k^2.$$

Esta estadística Q de Box y Ljung bajo H_0 se distribuye aproximadamente como un χ_{K-m}^2 , donde m es el número de parámetros estimados por el modelo.

Selección

Pueden existir varios modelos adecuados que pueden ser usados para representar un conjunto de datos, por ello se utilizan diversos criterios de selección de modelos basados en residuos.

El criterio de información de Akaike se define como:

$$AIC(M) = -2\ln[\text{máxima verosimilitud}] + 2M,$$

donde M es el número de parámetros del modelo. El orden óptimo del modelo se elige para el valor de M, que está en función de p y q en el caso de modelos ARIMA, de modo que AIC(M) sea mínimo. También se utiliza una extensión bayesiana del AIC llamada BIC.

III.3.4 Evaluación de los Pronósticos

Una vez estimado y seleccionado el modelo tanto con modelos ARIMA como por MEE se debe probar su eficacia para pronosticar.

Para evaluar la bondad de los pronósticos se construye el “Error porcentual absoluto medio” o “Mean absolute percentual error” (MAPE):

$$MAPE = \left(\frac{1}{n} \sum_{t=1}^n |e_t| \right) 100\%,$$

donde $e_t = \frac{\text{valor real} - \text{valor predicho un paso hacia adelante}}{\text{valor real}}$, siendo n el número

de predicciones efectuadas dentro de la muestra.

Esta es la estadística resumen más conveniente dado que toma valor absoluto y lo divide por el valor de la variable, por ello no está influenciada por la unidad de medida de la misma.

Se utiliza también para probar el desempeño predictivo de los modelos el “Error porcentual absoluto medio post muestral” o “Post sample mean absolute percentual error” (PSMAPE) que se define de la siguiente manera:

$$PSMAPE = \left(\frac{1}{H} \sum_{h=1}^H |e_{t+h}| \right) 100\%$$

donde $e_{t+h} = \frac{\text{valor real} - \text{valor predicho } h \text{ pasos hacia adelante}}{\text{valor real}}$, siendo H el número

de predicciones realizadas fuera de la muestra.

III.4 Pronósticos de las tasas de mortalidad específicas por edad

Una vez que se obtiene la serie de tiempo del índice k_t como se expresa en la sección (III.2.3) se pueden realizar pronósticos del mismo por modelos ARIMA o MEE.

A partir del índice k_t pronosticado pueden generarse los valores de las tasas de mortalidad para los años proyectados. Para ello se insertan los pronósticos k_{n+h} , en la fórmula:

$$\hat{m}_{x,n+h} = \hat{m}_{x,n} \cdot \exp \left\{ \hat{b}_x (\hat{k}_{n+h} - \hat{k}_n) \right\}, \quad h=1,2,\dots \quad x=1,2,\dots,\omega \quad (\text{III.4.1})$$

donde n es el último año para el que se dispone de información, h representa el horizonte de pronóstico y x el grupo etéreo.

Por medio de la fórmula (III.4.1) se computan los pronósticos de las tasas de mortalidad en referencia a la última tasa de mortalidad disponible.

Además Lee y Carter proponen un intervalo de confianza aproximado para los pronósticos de las tasas de mortalidad. Los límites superior e inferior se presentan en (III.4.2). Para su cálculo se emplean las estimaciones de los parámetros b_x , y los errores estándar de las proyecciones realizadas para el índice k_t .

$$\text{Límite superior: } \left\{ m_{x,t} \exp(2 b_x se_{k_t}) \right\} \quad (\text{III.4.2})$$

$$\text{Límite inferior: } \left\{ m_{x,t} \exp(-2 b_x se_{k_t}) \right\}$$

III.5 Esperanza de vida al nacer

Con el fin de obtener la esperanza de vida al nacer, a partir de las tasas de mortalidad pronosticadas, se emplea la metodología clásica, que se basa en el método desarrollado por Chiang C. L. (1984).

Para estimar la esperanza de vida al nacer se procede a la construcción de la tabla de vida abreviada. Se definen en el Anexo A.3 las diferentes funciones, y su modo de cálculo según la metodología propuesta, donde el subíndice x indica el intervalo de edad y n_x su amplitud. Además se define el factor de separación, que determina la fracción del intervalo de edad vivido por la población de la cohorte que murió en dicho intervalo. Usualmente se la llama α_x , pero aquí se utiliza f_x con el fin de distinguirlo del parámetro a del modelo de Lee-Carter.

La esperanza de vida a la edad x , e_x , se define como

$$e_x = \frac{T_x}{l_x}$$

donde T_x son los años vividos acumulados por la cohorte en el intervalo con edad inicial x y subsiguientes. Con l_x se representa al número de personas que están con vida a la edad inicial del intervalo x de una generación inicial de l_0 nacimientos. Generalmente se define $l_0=100.000$.

IV.- Análisis empírico

El siguiente análisis comprende la aplicación de la metodología propuesta por Lee y Carter (1992) para modelar tasas de mortalidad específicas por edad. Se realiza el análisis para la República Argentina y la Ciudad de Rosario (Provincia de Santa Fe). Las tasas se modelan para el total de la población discriminada por sexo y edad, generándose un índice general de mortalidad en cada caso. Cada uno de los índices obtenidos se pronostica mediante dos metodologías, modelo ARIMA y MEE. A partir de los pronósticos que se obtienen se pueden pronosticar las tasas de mortalidad específicas por edad. Con las tasas de mortalidad para Rosario y Argentina se construyen las correspondientes tablas de vida para cada género a través de la metodología clásica, y de ellas se extrae la esperanza de vida al nacer, una medida resumen, muy informativa y representativa para la comparación de centros poblacionales. Se incluyen además las proyecciones de esperanza de vida al

nacer de los organismos oficiales disponibles a fin de comparar los resultados aquí obtenidos.

IV.1 República Argentina

IV.1.1 Los datos

Para el país se cuenta con los totales de población y defunciones por edad y sexo, desde 1979 a 2006. Si bien Lee y Carter proponen un período de 30 años como base para realizar el análisis, en la presente tesina se utilizan 28 años debido a que son los datos disponibles para la Argentina. El periodo para el que se dispone de la información para el país resulta muy cercano al propuesto por los autores.

Los datos del período 1979-2001 son proporcionados por el Sistema de Información Estadística de la Organización Mundial de la Salud. Luego se completan con datos de la Dirección de Estadística e Información de Salud del Ministerio de Salud de la Nación, de allí se dispone del período 2000-2006. Se comprueba la coincidencia de los años 2000 y 2001 tanto para el total poblacional como para las defunciones, lo que permite concatenar las series. La tabla de datos se presenta en el Anexo A.1.

Los grupos etáreos quedan determinados por los siguientes intervalos, el primero de 0 a 4 años y luego intervalos decenales, hasta el último de ellos de 75 o más años de edad, quedando conformadas 9 categorías. La implementación de nueve categorías se basa en el mayor desglose posible que permiten los datos. Esto se debe a que en el período 2002-2006 se dispone de los datos de población en categorías más amplias, que obligan a adoptar este esquema con el fin de poder calcular las correspondientes tasas específicas por edad. En particular se presenta un único intervalo de población desde los cero a los cuatro años, cuando en múltiples estudios es mas frecuente que este intervalo se halle dividido en dos, menores de un año y de uno a cuatro años.

IV.1.2 Ajuste del modelo

El modelo de Lee-Carter para el índice general de mortalidad k_t planteado es:

$$m_{x,t} = \exp(a_x + b_x k_t + e_{x,t}), \quad t=1, \dots, 28 \quad x=0, 5, 15, \dots, 75 \quad (\text{IV.1.2.1})$$

Una vez que los datos se han preparado, se estiman los parámetros a_x y b_x . Se obtiene la primera estimación del índice general de mortalidad k_t , (“cruda” ó “bruta”), y se procede a su re-estimación utilizando la metodología propuesta por Lee-Carter, teniendo en cuenta la condición indicada en la ecuación (III.2.1.8). Este procedimiento se realiza tanto para la serie de total, como para varones y mujeres.

Los parámetros a_x describen la forma general de las tasas específicas por edad, mientras el índice k_t representa la variación en el nivel de mortalidad en el tiempo t . El índice k_t describe la tendencia general de la mortalidad, cómo esta se comporta a través del tiempo. Los parámetros b_x describen el cambio en la mortalidad en el intervalo de edad inicial x , frente a un cambio en k_t . Cuando b_x es grande para cierto intervalo de edad, expresa que la tasa de mortalidad en dicho intervalo varía sustancialmente cuando el nivel general de mortalidad cambia. Esto sucede en el primer intervalo, como puede apreciarse en la Tabla IV.1.2.1.

Tabla IV.1.2.1. Estimaciones de los parámetros a_x y b_x , para las tasas de mortalidad correspondientes al total, varones y mujeres de la Argentina.

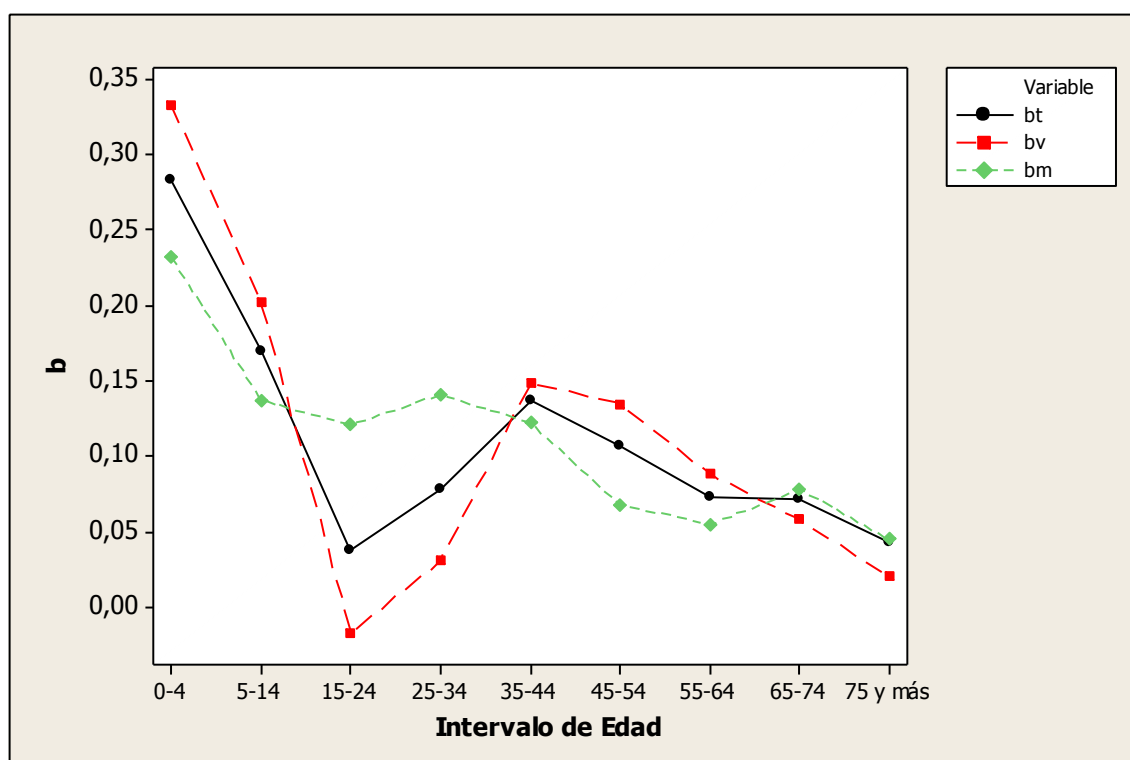
Grupo de Edad	Total		Varones		Mujeres	
	a_x	b_x	a_x	b_x	a_x	b_x
0-4	-5,24167	0,28414	-5,14386	0,33404	-5,35429	0,23262
5-14	-7,99261	0,16979	-7,83313	0,20265	-8,18974	0,13717
15-24	-7,03527	0,03720	-6,72832	-0,01741	-7,50422	0,12124
25-34	-6,66634	0,07869	-6,40873	0,03053	-7,02750	0,14107
35-44	-6,01734	0,13688	-5,78227	0,14905	-6,32116	0,12217
45-54	-5,14825	0,10667	-4,85319	0,13503	-5,55004	0,06805
55-64	-4,33273	0,07287	-4,00105	0,08816	-4,77527	0,05459
65-74	-3,55048	0,07152	-3,22951	0,05801	-3,90812	0,07838
75 y más	-2,34548	0,04223	-2,19382	0,01994	-2,45140	0,04471

Cuando b_x es pequeño, significa que las tasas de mortalidad para esa edad varían levemente cuando el nivel general de mortalidad cambia, como se observa en la tabla IV.1.2.1 para las edades avanzadas. Se destaca el valor de

b_{15} para varones, que pertenece al grupo de edad entre 15 y 24 años, ya que resulta negativo. Esto puede observarse en el Gráfico IV.1.2.1.

Este fenómeno puede deberse a algún suceso que elevó la mortalidad en algún momento dentro del período analizado, y no fue detectado, o realmente refleja un leve aumento de la mortalidad en dicho grupo etáreo. Se debe tener en cuenta que el periodo analizado incluye al año 1982, año en el cuál se desarrolló la guerra de Malvinas, éste podría ser el evento que produce el valor b_{15} para varones que resulta atípico. Es importante remarcar que el valor es muy cercano al cero, por ello puede indicar también que la tasa para ese grupo es, prácticamente constante, o que experimenta un leve ascenso.

Gráfico IV.1.2.1. Estimación de los parámetros b_x correspondientes al total, varones y mujeres de Argentina.

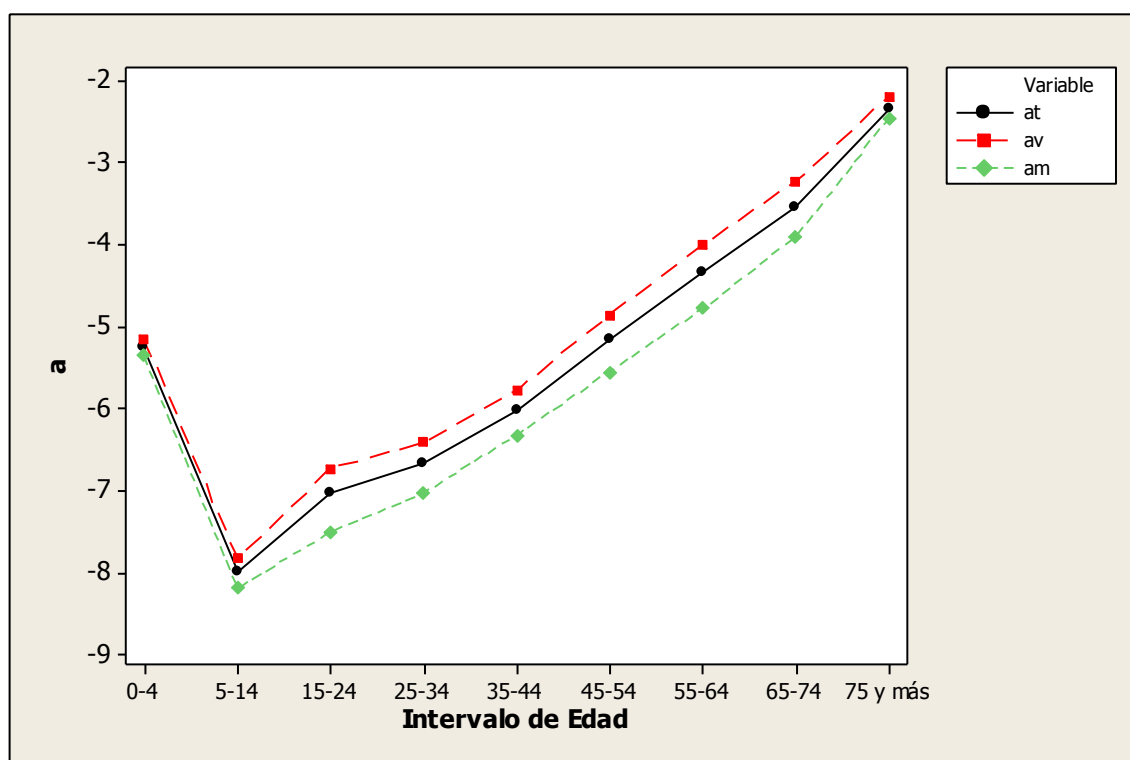


Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación.

Otra explicación posible se basa en que diversos estudios indican que, en la Argentina, han aumentado las muertes por causas violentas entre los varones de 15 a 24 años de edad. Una reciente investigación del Centro de Investigaciones Epidemiológicas de la Academia Nacional de Medicina de

Buenos Aires (Serfaty, 2003) concluye que de 1991 a 2000 ha aumentado la tasa de mortalidad por causas violentas en general, incluidos los accidentes de tránsito, otros accidentes (muertes por sumersión, ahogamiento, golpes e intoxicaciones), suicidios y homicidios, y destaca que en varones jóvenes esta tasa se ha duplicado. Esta realidad podría estar reflejada en el comportamiento que presenta el parámetro b para la categoría de 15 a 24 años. Además es notoria la diferencia que se da en los intervalos 15 a 24 y 25 a 34, entre varones y mujeres ya que los primeros presentan valores bastante más bajos que los segundos.

Gráfico IV.1.2.2. Estimación de los parámetros α_x para total, varones y mujeres de Argentina.



Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación.

Las estimaciones para total, hombres y mujeres del parámetro de “forma” (α_x), como lo denominan Lee y Carter, pueden observarse en el Gráfico IV.1.2.2. Estas estimaciones representan la forma en que la mortalidad se comporta a través de las edades, en este caso los valores para Argentina

son similares a los de la mayoría de los países: una mortalidad alta al comienzo de la vida, luego baja rápidamente hasta un mínimo en el intervalo de 5 a 14, aumenta relativamente lento hasta los 35 o 45 años, y de ahí crece más rápidamente, llegando a superar los niveles de las primeras edades (Ortega A. 1987).

Una leve diferencia entre varones y mujeres puede verse en el intervalo 5 a 14 años, que presenta un valor menor para mujeres, mientras que el intervalo siguiente, de 15 a 24 años presenta un crecimiento más fuerte en el caso de los varones.

Las primeras estimaciones del índice general de mortalidad para Argentina, calculado a partir de la ecuación IV.1.2.1, y sus re-estimaciones se presentan en la Tabla IV.1.2.2.

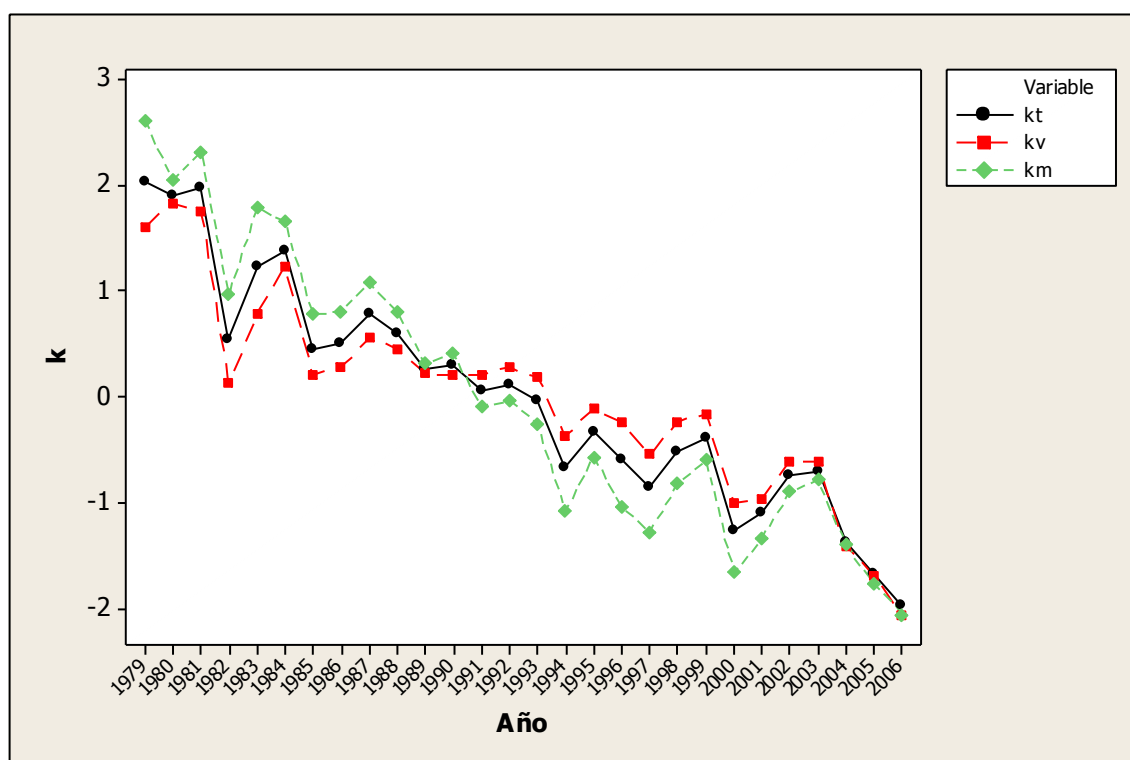
Tabla IV.1.2.2 Estimaciones del índice k, para total, varones y mujeres de la Argentina.

Año	kt(bruto)	kt	<b b="" kv(bruto)<="">	kv	km(bruto)	km
1979	2,12851	2,04060	1,72796	1,61323	2,66521	2,61227
1980	1,78021	1,89595	1,46803	1,83205	2,21339	2,05256
1981	1,93143	1,97872	1,61851	1,76066	2,36188	2,31607
1982	1,15136	0,54042	0,97213	0,13716	1,47824	0,97163
1983	1,27898	1,22637	1,11111	0,78784	1,50562	1,78910
1984	1,09233	1,38352	0,90900	1,22473	1,35949	1,65804
1985	0,35889	0,45866	0,30557	0,20613	0,52363	0,78337
1986	0,54536	0,50873	0,47089	0,27949	0,76307	0,80722
1987	0,37588	0,78205	0,35956	0,56893	0,48020	1,08565
1988	0,42945	0,59401	0,41834	0,45778	0,50643	0,79951
1989	0,35565	0,26727	0,33065	0,22066	0,43823	0,32314
1990	0,28413	0,29734	0,27822	0,21703	0,34694	0,40873
1991	0,17251	0,05749	0,21827	0,21071	0,13599	-0,09537
1992	0,14312	0,12047	0,18684	0,28522	0,00585	-0,03020
1993	-0,10043	-0,02929	-0,01559	0,18591	-0,24832	-0,25410
1994	-0,23102	-0,66949	-0,13081	-0,36490	-0,45270	-1,07628
1995	-0,37706	-0,32053	-0,33182	-0,10095	-0,47162	-0,56210
1996	-0,37417	-0,59306	-0,27146	-0,24454	-0,57391	-1,03564
1997	-0,59098	-0,85688	-0,52862	-0,54041	-0,69261	-1,27393
1998	-0,56331	-0,50946	-0,44113	-0,23550	-0,79449	-0,81906
1999	-0,72451	-0,37885	-0,65020	-0,15903	-0,88385	-0,59396
2000	-1,04603	-1,26867	-0,90406	-0,99959	-1,34160	-1,65810
2001	-1,10509	-1,08807	-1,02098	-0,96710	-1,30379	-1,33888
2002	-1,10896	-0,73392	-1,00552	-0,60263	-1,37516	-0,89049
2003	-1,01281	-0,70107	-1,00391	-0,61548	-1,08064	-0,77629
2004	-1,44899	-1,36887	-1,23332	-1,40858	-1,70731	-1,39640
2005	-1,63601	-1,66327	-1,36702	-1,69052	-1,94123	-1,75550
2006	-1,70846	-1,97018	-1,47064	-2,05829	-1,91692	-2,05097

Se denotan en esta etapa los índices correspondientes a total, varones y mujeres con la letra k acompañada de la t, v y m respectivamente. Siendo la t en formato de subíndice reservada para indicar el tiempo.

El Gráfico IV.1.2.3, permite visualizar el comportamiento de las series, que presentan una clara tendencia a disminuir a través del tiempo. Además indica la presencia de un posible valor aberrante en el año 1982, que se tendrá en cuenta al momento de modelar. En el mismo gráfico se detecta un descenso más marcado en el caso de las mujeres para la primera mitad del período analizado. El comportamiento del índice resultaría más parecido para ambos géneros desde los años 90 en adelante.

Gráfico IV.1.2.3. Estimación del índice general de mortalidad k correspondiente al total, varones y mujeres de Argentina.



Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación.

IV.1.3 Pronósticos para el índice general de mortalidad de Argentina

Para realizar los pronósticos se utilizan dos tipos de modelos: ARIMA y MEE. Al utilizar MEE para analizar series de tiempo se introduce el concepto de “parámetros que varían en el tiempo”, en contraposición al utilizar parámetros constantes en el tiempo como proponen los modelos ARIMA.

IV.1.3.1 Modelo ARIMA

Diagnóstico y selección del modelo

Se utiliza la secuencia propuesta por Box-Jenkins (1976), que consiste en identificación, estimación y diagnóstico para encontrar el proceso ARIMA que represente en forma apropiada los datos. El objetivo es hallar un mismo modelo que ajuste apropiadamente los tres índices estimados, total, varones y mujeres, pero teniendo en cuenta que el principal fin del modelo buscado es el de generar pronósticos. Se evalúan inicialmente los residuos utilizando la estadística de Box-Ljung y el test de Jarque-Bera. Luego para realizar la selección entre los distintos modelos ARIMA se emplea el criterio de Akaike, y de Schwartz. Finalmente el MAPE (mean absolute percentage error) es la medida que detecta el modelo más adecuado para pronosticar.

Los modelos planteados incluyen una variable dummy que permite representar el valor atípico del año 1982.

En la Tabla IV.1.3.1.1 se observa que los valores de probabilidad asociados resultan mayores para el ARIMA(0,1,2). Por otro lado, los valores del test Jarque-Bera indican un comportamiento normal, un proceso del error homocedástico y sin la presencia de puntos aberrantes, para ambos modelos en todos los casos.

Tabla IV.1.3.1.1 Valor de probabilidad asociado de la estadística de Box-Ljung y test de Jarque-Bera para los dos modelos ARIMA propuestos para Argentina, para total, varones y mujeres.

Modelo		Q(3)	Q(6)	Q(9)	Jarque-Bera
ARIMA(0,1,1) Con constante	Total	0,476	0,8199	0,9287	0,0132(0,993)
	Varones	0,5519	0,8204	0,9447	2,4035(0,301)
	Mujeres	0,5526	0,8578	0,9206	0,4765(0,788)
ARIMA(0,1,2) Con constante	Total	0,6731	0,822	0,9325	0,7131(0,7)
	Varones	0,8696	0,901	0,9699	3,2886(0,193)
	Mujeres	0,7175	0,908	0,9505	0,3971(0,82)

Los criterios de selección utilizados indican un mejor ajuste para el modelo ARIMA(0,1,2) en el total y varones, según AIC, y ARIMA(0,1,1) para las mujeres. Mientras que el criterio de Schwartz, indica mejor ajuste para el ARIMA(0,1,1) para mujeres, ARIMA(0,1,2) para varones, y no presenta casi diferencia para el total. Los valores de ambos criterios se presentan en la Tabla IV.1.3.1.2

Tabla IV.1.3.1.2 Criterios de Akaike y Schwartz para los dos modelos ARIMA propuestos para Argentina, para total, varones y mujeres.

Modelo		AIC	BIC
ARIMA(0,1,1) Con constante	Total	20,99	26,32
	Varones	26,08	31,41
	Mujeres	31,8	37,13
ARIMA(0,1,2) Con constante	Total	19,86	26,52
	Varones	24,2	30,86
	Mujeres	32,94	39,6

Teniendo en cuenta que la finalidad primordial del modelo es el pronóstico, resulta decisivo el PSMape para seleccionar el modelo definitivo. Esta medida es calculada modelando 25 observaciones, y luego comparando las últimas tres con los pronósticos obtenidos por el modelo efectuando tres predicciones fuera de la muestra. Se percibe un notorio descenso de dicha medida para las series total y mujeres del modelo ARIMA(0,1,2) con respecto a los valores obtenidos a partir del modelo ARIMA(0,1,1). Un leve descenso se observa para el caso de los varones, y un valor algo alejado de lo esperado para esta medida se obtiene en ambos modelos. Hay que destacar que los valores no son óptimos, pero son suficientemente buenos si se tiene en cuenta que las series aquí estudiadas abarcan 28 años, y presentan una gran variabilidad.

Tabla IV.1.3.1.3 PSMape para los dos modelos ARIMA propuestos para Argentina, para varones y mujeres.

Modelo		PSMAPE
ARIMA(0,1,1) Con constante	Total	0,24931823
	Varones	0,37843875
	Mujeres	0,1822726
ARIMA(0,1,2) Con constante	Total	0,12633101
	Varones	0,3568954
	Mujeres	0,14460759

Se decide trabajar con el modelo ARIMA(0,1,2), la estimación de los parámetros para los modelos se presentan a continuación. Los parámetros resultan altamente significativos en todos los casos, excepto para el parámetro MA(2) del modelo para mujeres. Los valores entre paréntesis indican el valor del desvío estandar correspondiente al parámetro que acompañan.

a) Modelo ARIMA(0,1,2) con constante, estimado para el total en Argentina.

$$(1-B)\hat{Z}_t = -0,1316 + (0,5452B + 0,4548B^2)\hat{a}_t - 0,9333I_t$$

$$(0,0092) \quad (0,2376) \quad (0,2057) \quad (0,251)$$

$$\text{donde } I_t = \begin{cases} 1 & t = 1982 \\ 0 & \text{en otro caso} \end{cases}$$

$$\sigma_{\varepsilon_T}^2 = 0,07595$$

b) Modelo ARIMA(0,1,2) con constante, estimado para varones en Argentina.

$$(1-B)\hat{Z}_t = -0,1132 + (0,4973B + 0,5027B^2)\hat{a}_t - 0,9305I_t$$

$$(0,0102) \quad (0,2336) \quad (0,2103) \quad (0,3031)$$

$$\text{donde } I_t = \begin{cases} 1 & t = 1982 \\ 0 & \text{en otro caso} \end{cases}$$

$$\sigma_{\varepsilon_V}^2 = 0,08919$$

c) Modelo ARIMA(0,1,2) con constante, estimado para mujeres en Argentina.

$$(1-B)\hat{Z}_t = -0,1615 + (0,3928B + 0,2518B^2)\hat{a}_t - 1,0211I_t$$

$$(0,028) \quad (0,2333) \quad (0,2514) \quad (0,3217)$$

$$\text{donde } I_t = \begin{cases} 1 & t = 1982 \\ 0 & \text{en otro caso} \end{cases}$$

$$\sigma_{\varepsilon_M}^2 = 0,1346$$

El hecho de que a los modelos planteados se les deba realizar una diferencia de orden uno para conseguir la necesaria estacionariedad, indica la presencia de una tendencia estocástica y la inclusión en cada modelo de una

constante, que resulta significativa, refleja que en las tres series existe además tendencia determinística.

Pronósticos ARIMA

Con el modelo seleccionado se realizan los pronósticos 5 años hacia delante. Se presentan las estimaciones con los correspondientes intervalos de pronóstico del 95% de confianza, en las Tablas IV.1.3.1.5, IV.1.3.1.4 y IV.1.3.1.6 se indica con I al límite inferior del intervalo, y con S al límite superior.

Tabla IV.1.3.1.5 Pronósticos Índice k_t para el total en Argentina.

Año	Pronóstico	I(95%)	S(95%)
2007	-1,848324	-2,442939	-1,253708
2008	-1,984486	-2,599821	-1,369152
2009	-2,120649	-2,756027	-1,485271
2010	-2,256811	-2,911619	-1,602003
2011	-2,392974	-3,066652	-1,719296

Tabla IV.1.3.1.4 Pronósticos Índice k_t para varones en Argentina.

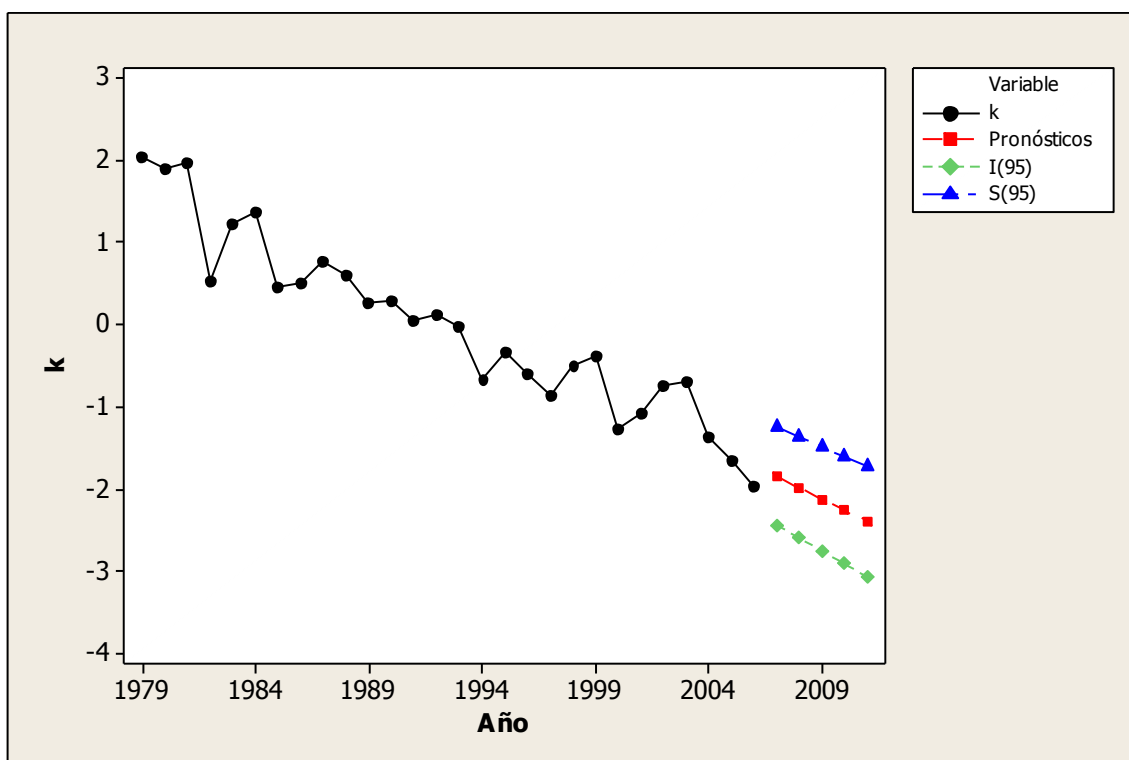
Año	Pronóstico	I(95%)	S(95%)
2007	-1,891139	-2,548422	-1,233857
2008	-2,020613	-2,727315	-1,313910
2009	-2,150086	-2,902971	-1,397201
2010	-2,279559	-3,075953	-1,483165
2011	-2,409032	-3,246678	-1,571386

Tabla IV.1.3.1.6 Pronósticos Índice k_t para mujeres en Argentina.

Año	Pronóstico	I(95%)	S(95%)
2007	-2,056216	-2,787514	-1,324919
2008	-2,217574	-3,009715	-1,425433
2009	-2,378932	-3,227565	-1,530299
2010	-2,540289	-3,441882	-1,638697
2011	-2,701647	-3,653256	-1,750038

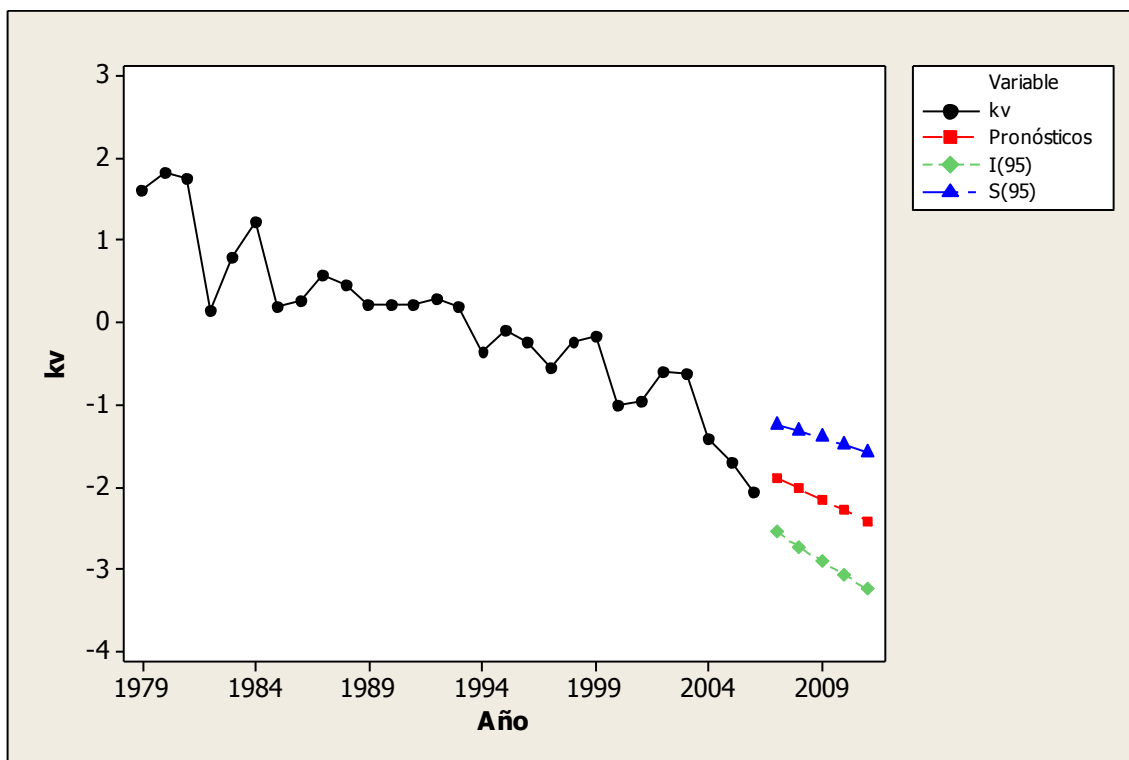
Los intervalos de pronóstico que se obtienen en los tres casos en estudio presentan una amplitud aproximadamente constante a través de las 5 proyecciones que se realizan, como puede observarse en los Gráficos IV.1.3.1.1, IV.1.3.1.2 y IV.1.3.1.3.

Gráfico IV.1.3.1.1 Proyecciones e Intervalos de Pronóstico del 95% para el índice k_t estimado para el total en Argentina.



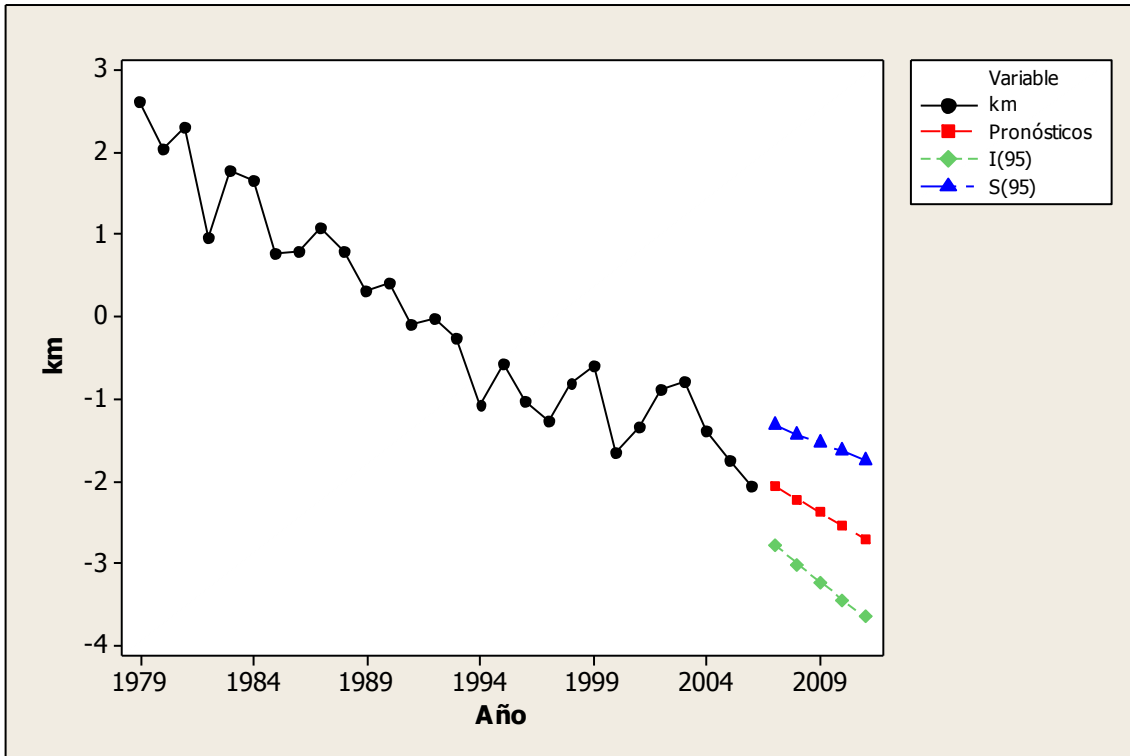
Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación.

Gráfico IV.1.3.1.2 Proyecciones e Intervalos de Pronóstico del 95% para el índice k_t estimado para varones en Argentina.



Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación.

Gráfico IV.1.3.1.3 Proyecciones e Intervalos de Pronóstico del 95% para el índice k_t estimado para mujeres en Argentina.



Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación.

IV.1.3.2 Modelo de espacio de estados

Ajuste del modelo

El modelo estructural básico formulado en términos de los componentes tendencia e irregular para el índice general de mortalidad k se puede escribir

$$y_t = \mu_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad t=1, \dots, 28,$$

$$\text{Nivel:} \quad \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_{1t} \quad (\text{IV.1.3.2.1})$$

$$\text{Pendiente:} \quad \beta_t = \beta_{t-1} + \eta_{2t}$$

donde μ_t y β_t representan la tendencia y ε_t la componente irregular.

Donde los errores η_{1t} y η_{2t} :

$$\eta_{it} \sim \text{IN}(0, \sigma_i^2), \quad i=1,2 \quad \text{independientes entre si y de } \sigma_\varepsilon^2.$$

Las estimaciones de los hiperparámetros y de las ecuaciones de estado para los MEE construidos para total, varones y mujeres, respectivamente, de la Argentina, se presentan en las Tablas IV.1.3.2.1, IV.1.3.2.2 y IV.1.3.2.3. Los hiperparámetros estimados para la pendiente resultan nulos para el total y los varones indicando comportamiento determinístico para esta componente en ambos casos. Para las mujeres dicho parámetro presenta un valor de 0,00036, que refleja la existencia de un comportamiento estocástico a diferencia de los varones y el total. El nivel presenta en los tres casos analizados un comportamiento estocástico, al igual que la componente irregular, cuya estimación resulta 0,06324 para el total, 0,06293 para los varones, y un valor de 0,08822 en el modelo estimado para las mujeres.

Tabla IV.1.3.2.1 Estimación MEE para el índice k_t estimado para el total.

	Parámetro		Estimación	
Hiperparámetros	Nivel		0,01436	
	Pendiente		0,00000	
	Componente Irregular		0,06324	
Ecuaciones de Estado	Coeficiente	Estimación	R.M.S.E	p-value
	Nivel	-1,7538	0,15959	0,0000
	Pendiente	-0,13978	0,02465	0,0000
	Outlier	-0,98044	0,28882	0,0022

Tabla IV.1.3.2.2 Estimación MEE para el índice k_t estimado para varones.

	Parámetro		Estimación	
Hiperparámetros	Nivel		0,02924	
	Pendiente		0,00000	
	Componente Irregular		0,06293	
Ecuaciones de Estado	Coeficiente	Estimación	R.M.S.E	p-value
	Nivel	-1,834	0,17887	0,0000
	Pendiente	-0,13259	0,03428	0,0007
	Outlier	-1,1428	0,30556	0,0009

Tabla IV.1.3.2.3 Estimación MEE para el índice k_t estimado para mujeres.

	Parámetro		Estimación	
Hiperparámetros	Nivel		0,02032	
	Pendiente		0,00036	
	Componente Irregular		0,08822	
Ecuaciones de Estado	Coeficiente	Estimación	R.M.S.E	p-value
	Nivel	-1,8441	0,199	0,0000
	Pendiente	-0,13551	0,58691	0,0292
	Outlier	-0,93083	0,34232	0,0115

Al analizar el cumplimiento de los supuestos acerca del comportamiento de los errores se hallan valores de probabilidad asociado para la estadística de

Box-Ljung lejanos al 10% en todos los casos, como puede verse en la Tabla IV.1.3.2.4. En dicha tabla se presentan los resultados del test de Jarque-Bera, los valores obtenidos indican que los residuos presentan un comportamiento normal.

Tabla IV.1.3.2.4 Valor de probabilidad asociado de la estadística de Box-Ljung y test de Jarque-Bera para el MEE de Argentina.

	Q(3)	Q(6)	Q(9)	Jarque-Bera
Total	0,3386	0,691	0,8423	1,1367(0,5665)
Varones	0,6007	0,8437	0,9522	2,3839(0,3036)
Mujeres	0,3961	0,7637	0,8602	0,7726(0,6796)

En la Tabla IV.1.3.2.5 se informan los valores del error medio absoluto porcentual post muestral calculado modelando 25 observaciones, y luego comparando las últimas tres con los pronósticos obtenidos por el modelo.

Tabla IV.1.3.2.5 PSMAPE para los dos MEE propuestos para Argentina, para total, varones y mujeres.

Modelo		PSMAPE
MEE	Total	0,3617
	Varones	0,4446
	Mujeres	0,3887

Estas medidas están considerablemente alejadas de lo esperado, es decir 0,10 o menos. El PSMAPE resulta una medida decisiva, ya que la finalidad primera del modelo es pronosticar.

Pronósticos obtenidos mediante la implementación de MEE

A partir del MEE que se ha ajustado se realizan las proyecciones del índice general de mortalidad, para el total y ambos géneros. Los valores obtenidos se encuentran en las Tablas IV.1.3.2.6, IV.1.3.2.7 y IV.1.3.2.8, para el total, varones y mujeres respectivamente. Los Gráficos IV.1.3.2.1, IV.1.3.2.2 y IV.1.3.2.3, permiten visualizar el comportamiento de los intervalos de pronóstico, allí se puede ver intervalos amplios, una amplitud marcada con respecto a los intervalos generados por los modelos ARIMA y que aumentan considerablemente a través del tiempo.

Tabla IV.1.3.2.6 Pronósticos Índice k para el total en Argentina.

Año	Pronóstico	I(95%)	S(95%)
2007	-1,8936	-3,267011	-0,520189
2008	-2,0334	-3,584701	-0,4821
2009	-2,1732	-3,893892	-0,452508
2010	-2,3130	-4,197172	-0,428828
2011	-2,4528	-4,495839	-0,409761

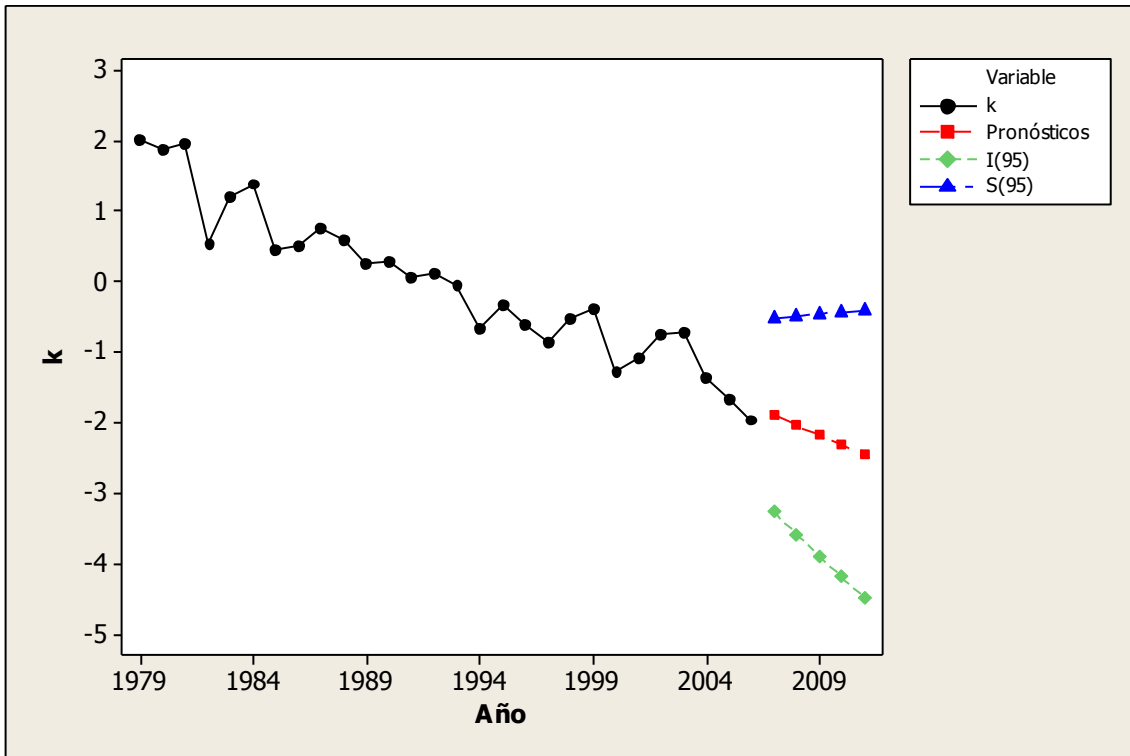
Tabla IV.1.3.2.7 Pronósticos Índice k para varones en Argentina.

Año	Pronóstico	I(95%)	S(95%)
2007	-1,9666	-3,215504	-0,717696
2008	-2,0992	-3,451328	-0,747072
2009	-2,2318	-3,685998	-0,777602
2010	-2,3644	-3,919711	-0,809089
2011	-2,4970	-4,152577	-0,841423

Tabla IV.1.3.2.8 Pronósticos Índice k para mujeres en Argentina.

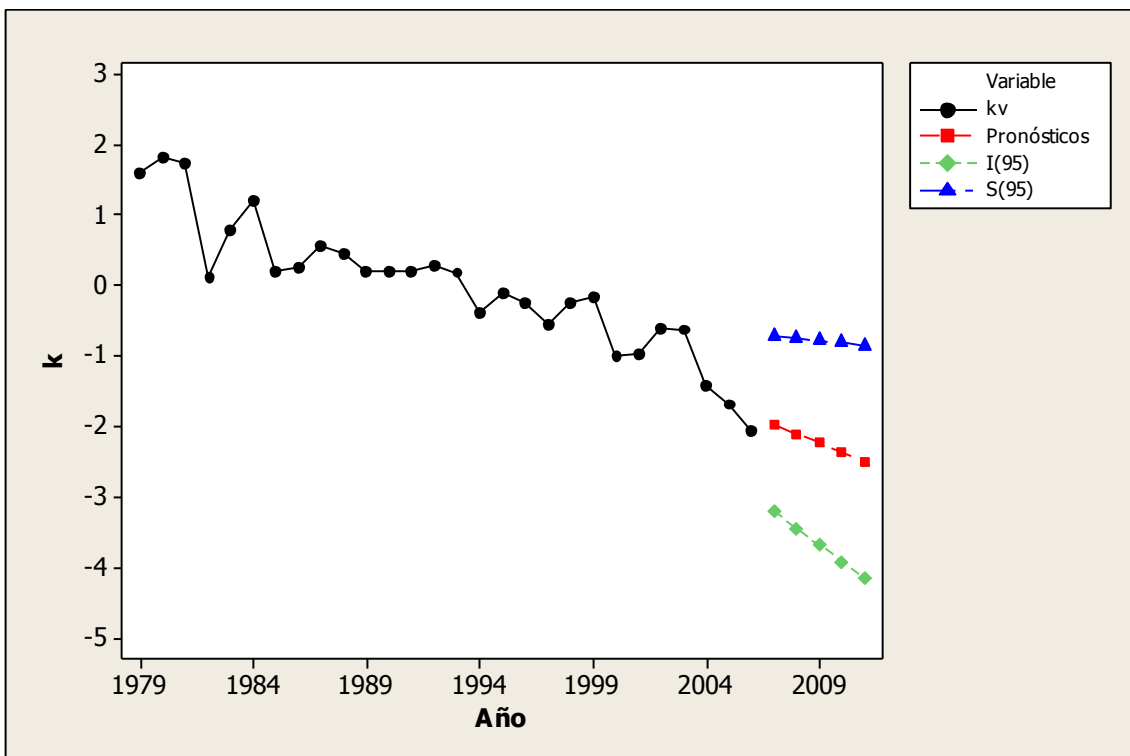
Año	Pronóstico	I(95%)	S(95%)
2007	-1,9796	-3,516548	-0,442652
2008	-2,1151	-3,831219	-0,398981
2009	-2,2506	-4,159991	-0,341209
2010	-2,3861	-4,501553	-0,270647
2011	-2,5216	-4,854872	-0,188328

Gráfico IV.1.3.2.1 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para el total en Argentina.



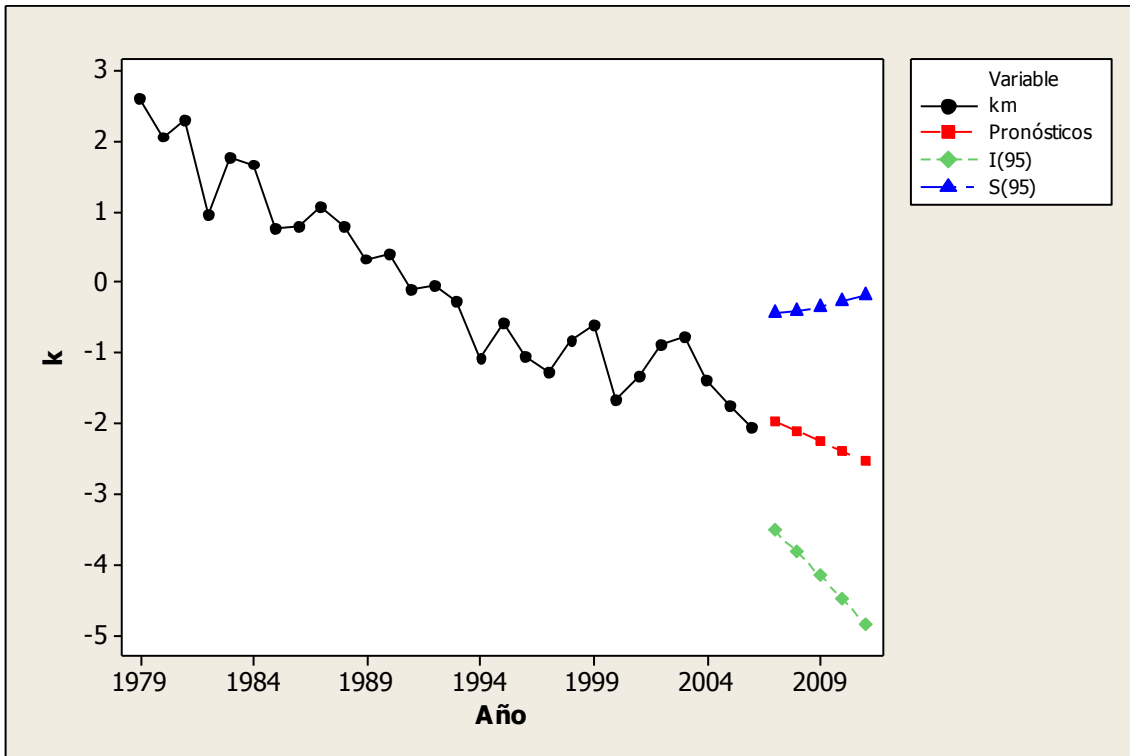
Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación

Gráfico IV.1.3.2.2 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para varones en Argentina.



Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación

Gráfico IV.1.3.2.3 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para mujeres en Argentina.



Fuente: Elaboración propia en base a datos de la OMS y Ministerio de Salud de la Nación

IV.1.4 Pronósticos de las tasas de mortalidad específicas por edad

A partir de los pronósticos generados por los modelos ARIMA y MEE se calculan los pronósticos de tasas de mortalidad. Para ello se insertan los pronósticos k_{2006+h} , en la fórmula:

$$\hat{m}_{x,2006+h} = \hat{m}_{x,2006} \cdot \exp\left\{\hat{b}_x \left(\hat{k}_{2006+h} - \hat{k}_{2006}\right)\right\}, \quad h=1,2,\dots,5 \quad x=0,5,\dots,75$$

El cálculo se realiza en referencia al último año para el que se dispone de información, que es el año 2006. El horizonte de pronóstico es representado por h y x indica la edad al inicio del intervalo para el cual se calcula la tasa.

Así se computan los pronósticos de las tasas de mortalidad en alineación con la última tasa de mortalidad disponible. Las Tablas IV.1.4.1 y IV.1.4.3 presentan las tasas de mortalidad específicas por edad calculadas a partir de los pronósticos ARIMA del índice k_t y los pronósticos realizados a través de MEE de dicho índice, respectivamente. Se destaca una clara coincidencia de los valores estimados a partir de las dos metodologías aquí empleadas; se puede ver en las tablas que los valores estimados de las tasas

difieren en su mayoría a partir de la tercera cifra decimal. Además se calculan las estimaciones por intervalos de confianza del 95% para cada una de las tasas, las cuales se presentan en las tablas IV.1.4.2 y IV.1.4.4. Las Tablas realizadas para ambos géneros se encuentran en el Anexo A.4.

Tabla IV.1.4.1 Tasas de mortalidad específicas por edad para el total de Argentina, estimadas a partir de pronósticos ARIMA en el período 2007-2011.

	2007	2008	2009	2010	2011
0-4	0,003279	0,003155	0,003035	0,002920	0,002809
5-14	0,000294	0,000287	0,000281	0,000274	0,000268
16-24	0,000816	0,000812	0,000808	0,000804	0,000800
25-34	0,001044	0,001032	0,001021	0,001011	0,001000
35-44	0,001946	0,001910	0,001874	0,001840	0,001806
45-54	0,004771	0,004702	0,004634	0,004568	0,004502
55-64	0,011453	0,011340	0,011228	0,011117	0,011007
65-74	0,025009	0,024766	0,024526	0,024289	0,024053
74 y más	0,088449	0,087942	0,087438	0,086936	0,086438

Tabla IV.1.4.2 Estimación por intervalos del confianza del 95% para las tasas de mortalidad específicas por edad para el total de Argentina, estimadas a partir de pronósticos ARIMA en el período 2007-2011.

	0-4		5-14		15-24	
2007	0,00275974	0,00389596	0,00026522	0,0003259	0,00079779	0,00083463
2008	0,00263947	0,00377122	0,00025798	0,00031929	0,00079325	0,00083119
2009	0,00252437	0,00364892	0,00025171	0,0003137	0,00078875	0,00082772
2010	0,00241507	0,00353049	0,00024461	0,00030692	0,00078426	0,00082423
2011	0,00231059	0,00341492	0,00023848	0,00030118	0,0007798	0,00082072
	25-34		35-44		45-54	
2007	0,00099533	0,00109506	0,00179091	0,00211452	0,00447199	0,00518416
2008	0,00098225	0,00108427	0,0017527	0,00208142	0,00439739	0,00512399
2009	0,00097022	0,00107444	0,00171486	0,00204791	0,00432435	0,00506405
2010	0,00095922	0,00106558	0,00167918	0,00201622	0,00425375	0,00500549
2011	0,00094734	0,00105558	0,00164381	0,00198419	0,00418369	0,00494619
	55-64		65-74		75 y más	
2007	0,01095765	0,01197074	0,02394695	0,02611815	0,08621145	0,09074462
2008	0,01083284	0,01187091	0,02367844	0,02590351	0,08564079	0,09030505
2009	0,01070987	0,0117712	0,0234147	0,02569004	0,08507646	0,08986509
2010	0,01058868	0,01167168	0,02315558	0,02547789	0,08451722	0,089424
2011	0,01046921	0,01157241	0,02289904	0,02526511	0,08396478	0,08898407

Si bien los valores que se obtienen por ambos modelos son parecidos, los pronósticos que se generan a partir de modelos ARIMA se pueden considerar más precisos porque los intervalos de confianza para el índice k_t en dichos modelos resultan más estrechos.

Se destaca en todos los casos que los intervalos de confianza generados a partir de modelos ARIMA resultan más estrechos que los

obtenidos a partir de MEE. Este resultado es de esperarse ya que en la construcción de dicho intervalo se utiliza el desvío estándar de los pronósticos de k_t que es justamente el que resulta más pequeño para las proyecciones que se calculan en base a modelos ARIMA.

Tabla IV.1.4.3 Tasas de mortalidad específicas por edad para el total de Argentina, estimadas a partir de pronósticos MEE en el período 2007-2011.

	2007	2008	2009	2010	2011
0-4	0,003237	0,003111	0,002990	0,002873	0,002762
5-14	0,000292	0,000285	0,000278	0,000272	0,000265
16-24	0,000815	0,000811	0,000807	0,000802	0,000798
25-34	0,001040	0,001029	0,001017	0,001006	0,000995
35-44	0,001934	0,001897	0,001861	0,001826	0,001791
45-54	0,004748	0,004678	0,004608	0,004540	0,004473
55-64	0,011415	0,011299	0,011185	0,011071	0,010959
65-74	0,024928	0,024680	0,024434	0,024191	0,023951
74 y más	0,088280	0,087760	0,087244	0,086730	0,086220

Tabla IV.1.4.4 Estimación por intervalos del confianza del 95% para las tasas de mortalidad específicas por edad para el total de Argentina, estimadas a partir de pronósticos MEE en el período 2007-2011.

	0-4		5-14		15-24	
2007	0,00217372	0,00482038	0,00023017	0,00037044	0,0007736	0,00085862
2008	0,00198409	0,00487796	0,00021783	0,00037288	0,00076462	0,00086019
2009	0,00181553	0,00492423	0,00020634	0,00037456	0,00075597	0,00086147
2010	0,00166373	0,00496122	0,00019624	0,000377	0,00074664	0,00086146
2011	0,00152745	0,00499437	0,000186	0,00037755	0,00073845	0,00086235
	25-34		35-44		45-54	
2007	0,00093141	0,00116125	0,00159642	0,00234297	0,00408872	0,00575203
2008	0,00090849	0,0011655	0,00152745	0,00235596	0,00395119	0,0058098
2009	0,00088576	0,00116768	0,00146342	0,00236659	0,00382096	0,00585988
2010	0,00086475	0,00117032	0,00140349	0,00237571	0,00369818	0,00590675
2011	0,00084446	0,00117238	0,00134637	0,00238246	0,00358114	0,00595016
	55-64		65-74		75 y más	
2007	0,01030682	0,01264234	0,02255058	0,02755606	0,08320694	0,09366236
2008	0,01006802	0,01268049	0,02203826	0,02763841	0,08208518	0,09382714
2009	0,00984169	0,01271166	0,02155053	0,02770329	0,08100906	0,09395882
2010	0,00962369	0,01273598	0,02108316	0,02775696	0,07996647	0,09406559
2011	0,00941446	0,01275694	0,02063338	0,02780206	0,07895388	0,09415482

IV.1.5 Esperanza de vida al nacer

Al momento de aplicar la metodología propuesta por la ONS para el cálculo de esperanza de vida surge un inconveniente. Tanto en la metodología que se utiliza aquí, como en la mayoría de las existentes para tablas de vida, se define un intervalo para menores de un año y otro para niños de uno a cuatro años. Dados los datos disponibles, como se detalla en el punto IV.1.1, en el presente trabajo dichas categorías aparecen colapsadas.

Es necesario contar con un factor de separación para cada categoría. La metodología clásica propone un factor de 0,5 para mayores de un año, que indica que las muertes se distribuyen uniformemente a lo largo del intervalo y un factor de 0,1 para los menores de un año.

En este trabajo dado que el primer grupo de edad es de 0 a 4 se utiliza la metodología clásica con tres posibles factores de separación (0,1, 0,3 y 0,5) y se comparan los resultados obtenidos en la esperanza de vida, que es el indicador de interés. Las diferencias observadas son inferiores a los 10 días y por lo tanto se decide el valor del factor de separación que genera un resultado más conservador para la esperanza de vida, este valor resulta ser el 0,1. Las tablas de vida completas y para cada factor de separación se presentan en el Anexo A.6.

Tabla IV.1.5.1 Estimación de la esperanza de vida al nacer para el total en Argentina, obtenidas a partir de los pronósticos de las tasas de mortalidad basados en modelos ARIMA y MEE.

Año/Pronósticos	Esperanza de vida al nacer	
	ARIMA	MEE
2007	75,67	75,73
2008	75,85	75,92
2009	76,02	76,09
2010	76,19	76,27
2011	76,37	76,44

Tabla IV.1.5.2 Estimación de la esperanza de vida al nacer para varones en Argentina, obtenidas a partir de los pronósticos de las tasas de mortalidad basados en modelos ARIMA y MEE.

Año/Pronósticos	Esperanza de vida al nacer	
	ARIMA	MEE
2007	71,91	72,01
2008	72,08	72,18
2009	72,24	72,34
2010	72,40	74,50
2011	72,55	72,66

Tabla IV.1.5.2 Estimación de la esperanza de vida al nacer para mujeres en Argentina, obtenidas a partir de los pronósticos de las tasas de mortalidad basados en modelos ARIMA y MEE.

Año/Pronósticos	Esperanza de vida al nacer	
	ARIMA	MEE
2007	79,53	79,44
2008	79,72	79,60
2009	79,91	79,76
2010	80,09	79,92
2011	80,28	80,07

Se observan en la Tabla IV.1.5.1 las estimaciones de la esperanza de vida para el total, se destaca la pequeña diferencia entre uno y otro método de pronóstico. Las diferencias son de entre veinte días y un mes, indicando que tanto el uso de modelos ARIMA, como MEE conducen prácticamente a los mismos resultados. En la Tabla IV.1.5.2 se presentan las estimaciones para los varones, en este caso las diferencias son de un mes y una semana para todos los años pronosticados, nuevamente los resultados difieren poco entre los dos métodos. Es en las estimaciones realizadas para las mujeres donde se presentan diferencias más notorias. Las diferencias en este caso oscilan entre los dos meses y los dos meses y medio.

En la Tabla IV.1.5.3 se presentan las estimaciones oficiales que brinda el INDEC. Lamentablemente no se poseen estimaciones de la esperanza de vida al nacer para cada año, sino por quinquenios. Lo que si es posible analizar es que los pronósticos de esperanza de vida obtenidos a partir del modelo de Lee-Carter, utilizando ambas metodologías, superan a las estimaciones del INDEC. Es importante tener en cuenta que fenómenos como la mortalidad tienden a estabilizarse por procesos en general intrínsecos (biológicos, etc) y no sólo a seguir la tendencia observada.

Tabla IV.1.5.3 Esperanza de vida al nacer proyectada por el INDEC para la Argentina.

Quinquenio	Esperanza de vida al nacer		
	Total	Varones	Mujeres
2005-2010	75,24	71,56	79,06
2010-2015	76,13	72,45	79,95

IV.2 Ciudad de Rosario

IV.2.1 Los Datos

Para la ciudad de Rosario se dispone de los totales de población por grupo de edad y sexo (excepto para el año 1991), así como también las defunciones por edad y sexo. Los datos que se analizan se obtienen de los Anuarios Estadísticos y/o Informes de la Dirección General de Estadística de la Municipalidad de Rosario.

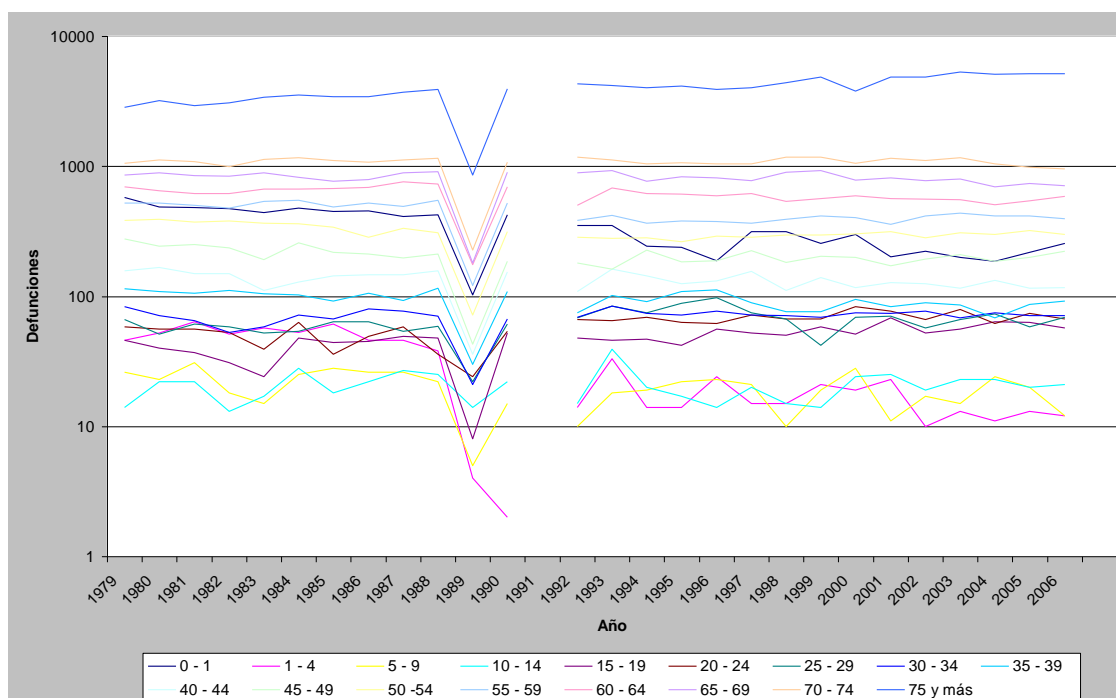
Las series comprenden el período 1979-2006, con excepción del año 1991. Se presentan tanto total poblacional como defunciones, para mujeres y varones registrados en la ciudad de Rosario. La información se presenta desglosada en 17 intervalos de edad, menores de un año, luego intervalos de cinco años, y por último un intervalo para mayores de 75 años. Los datos se presentan en el Anexo A.2.

Además de presentarse datos faltantes correspondientes a las defunciones del año 1991, se detecta que el año 1989 presenta posibles errores de registro, esto puede verse en el Gráfico IV.2.1.1, ya que los datos se alejan consistentemente del resto, para cada una de las series sexo-edad, presentando una caída abrupta en las defunciones registradas en dicho año que no tiene ninguna explicación aparente. Por ello se procede a estimar ambos datos, para cada una de las series, es decir para cada sexo en cada categoría, conformando un total de 34 series de tiempo. Se estima el modelo ARIMA que mejor ajusta los datos en el período 1979-1988, se pronostica el 1989, se agrega el año 1990, se actualiza el modelo, y se pronostica el año 1991. Aunque se cuenta con los años 1990 y 1992, estos igualmente se pronostican para informar el error de pronóstico. Una medida de bondad de los pronósticos se calcula a partir del promedio de los errores. Para los varones el promedio de errores cometidos al pronosticar 1990 es 1,45, mientras que se obtiene un promedio de error de 0,39 cuando se pronostica 1992. Esta misma

medida resulta similar en ambos casos para las mujeres, se obtiene un promedio de errores de 1,44 y 0,56, para 1990 y 1992 respectivamente. Tanto en varones como en mujeres se observan errores pequeños.

Con la finalidad de comparar Rosario y la Argentina se decide colapsar categorías, quedando determinada la conformación de los grupos por intervalos compuestos por un primer grupo de 0 a 4 años, luego grupos decenales, y finalmente un intervalo para personas de 75 o más años de edad.

Gráfico IV.2.1.1 Defunciones para el total de residentes en Rosario, registradas durante el período 1979-2006, según grupo etáreo.



Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

Luego se detecta que en el año 2000, la serie de defunciones de varones, en la categoría “75 y más” presenta un valor considerablemente alejado e los valores que se registran en dicha categoría en los otro años bajo estudio. No se halla ningún hecho que explique el valor, aunque se sospecha puede tener origen en que la categoría en cuestión idealmente debería estar subdividida en más intervalos, por ello se decide estimarlo. Se construye un modelo ARIMA, que considera el período 1979-1999, se pronostica el valor correspondiente al año 2000, y 2001, este último se estima con el fin de obtener una medida del error. La diferencia entre el valor observado y el pronosticado para el año 2001 es cero.

Conviene destacar dos puntos, en primer lugar, los datos de Rosario parecen tener algunos posibles errores de registro además de la falta de observaciones para un año, y en segundo lugar se sospecha que la población obtenida en el censo 2001 y publicada por el INDEC, subestima considerablemente la población real, según lo informa la Dirección Municipal de Estadística de la Ciudad de Rosario . Esto genera una sobreestimación de las tasa de mortalidad para los años 2001 en adelante, por ello es posible que se ocasione un desbalance en el cálculo del índice de Lee-Carter.

IV.2.2 Ajuste del modelo

Se realiza la estimación de los parámetros y el índice general de mortalidad que comprenden el modelo de Lee-Carter, presentado en el punto III.2.1, para cada uno de los casos analizados. A continuación se analiza el comportamiento de las estimaciones de los parámetros de forma y sensibilidad, (a y b). Luego se grafica y describe el índice k obtenido a partir del modelo para el total, los varones y las mujeres de la ciudad de Rosario.

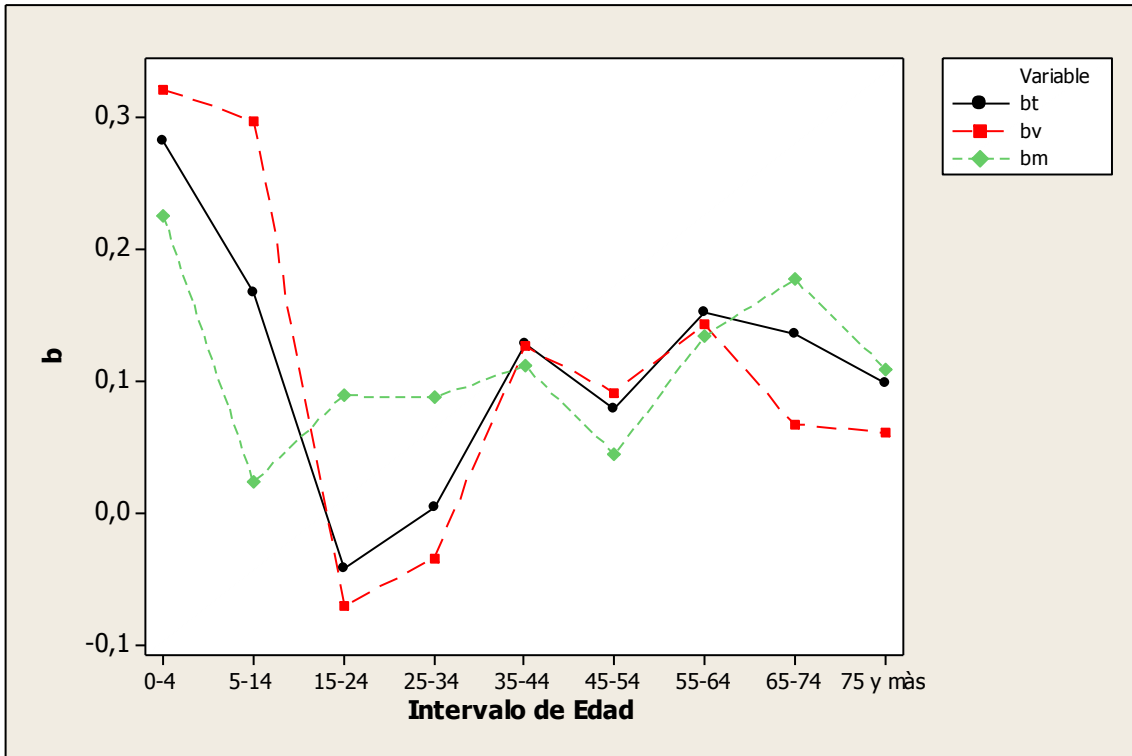
Tabla IV.2.2.1 Estimaciones de los parámetros a_x y b_x , para las tasas de mortalidad correspondientes al total, varones y mujeres de Rosario.

Grupo de Edad	Total		Varones		Mujeres	
	a_x	b_x	a_x	b_x	a_x	b_x
0-4	-5,40812	0,281102	-5,38273	0,320747	-5,44764	0,225011
5-14	-8,25947	0,167064	-8,12795	0,297017	-8,44352	0,023392
15-24	-7,20086	-0,042490	-6,82644	-0,070770	-7,79132	0,088744
25-34	-6,80580	0,004482	-6,48633	-0,034407	-7,26656	0,086930
35-44	-6,16046	0,128207	-5,91451	0,125871	-6,46676	0,111536
45-54	-5,25806	0,077975	-4,93277	0,091288	-5,68206	0,044298
55-64	-4,35048	0,151235	-3,97090	0,142866	-4,82167	0,134096
65-74	-3,46225	0,135075	-3,13879	0,066849	-3,80289	0,177006
75 y mas	-2,18980	0,097348	-2,06974	0,060540	-2,26523	0,108988

Al analizar la estimación del parámetro b_x (Gráfico IV.2.2.1), llamado de sensibilidad, se presentan marcadas diferencias entre varones y mujeres, si bien en el intervalo de 0 a 4 años se presenta el mayor valor positivo en ambos casos, este resulta mayor en los varones, indicando que para este intervalo la mortalidad varía sustancialmente cuando el índice k cambia. El intervalo de 5

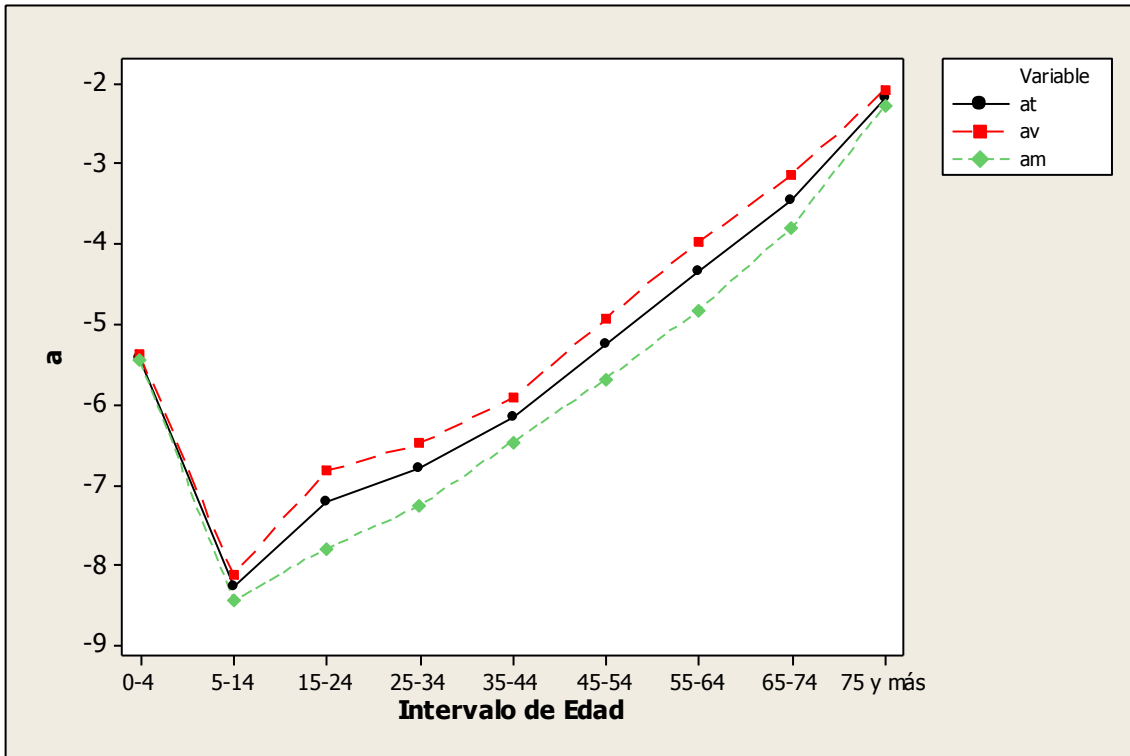
a 14 años, presenta un valor alto para los varones y muy cercano al cero para las mujeres, indicando comportamientos opuestos para las tasas de este grupo

Gráfico IV.2.2.1 Estimación de los parámetros b_x correspondientes a varones y mujeres de Rosario.



Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

Gráfico IV.2.2.2 Estimación de los parámetros a_x correspondientes a varones y mujeres de Rosario.



Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

frente a cambios en k . En lo que respecta a los intervalos 15 a 24 y 25 a 34, los varones presentan valores negativos, que estarían indicando una tendencia a aumentar la mortalidad, mientras que las mujeres presentan valores positivos cercanos al 0,1 que denotarían una leve baja frente a cambios en k .

En relación a los intervalos de edades más avanzadas es de esperar un comportamiento en descenso, pero se ve en el Gráfico IV.2.2.1 un comportamiento fluctuante, con un aumento en el valor de b en el intervalo 55-64 en varones y 65-74 en mujeres. Aquí es importante destacar que las series para Rosario presentan gran variabilidad y este hecho podría estar reflejándose en ciertos puntos que se alejan del comportamiento esperado.

Tabla IV.2.2.2 Estimaciones del índice k , para total, varones y mujeres de Rosario.

Año	kt(bruto)	kt	kv(bruto)	kv	km(bruto)	km
1979	2,15906	2,64462	1,92143	2,98638	2,41402	2,57459
1980	1,54488	1,43545	1,60661	1,83031	1,67469	1,27317
1981	1,47773	0,90516	1,52324	1,42947	1,34457	0,49538
1982	0,82705	0,70177	0,67893	0,71518	1,21551	0,72059
1983	0,83285	1,12932	0,74114	1,31486	0,59752	1,07488
1984	1,31659	1,36764	1,2093	1,42897	1,34527	1,46667
1985	1,00584	0,8724	1,04074	0,6728	0,95158	1,09816
1986	0,95238	0,80883	1,03469	0,64781	1,09904	0,98942
1987	0,94458	1,18962	1,2005	1,06797	1,05603	1,40849
1988	0,99526	1,35391	0,85094	1,25958	1,10982	1,57955

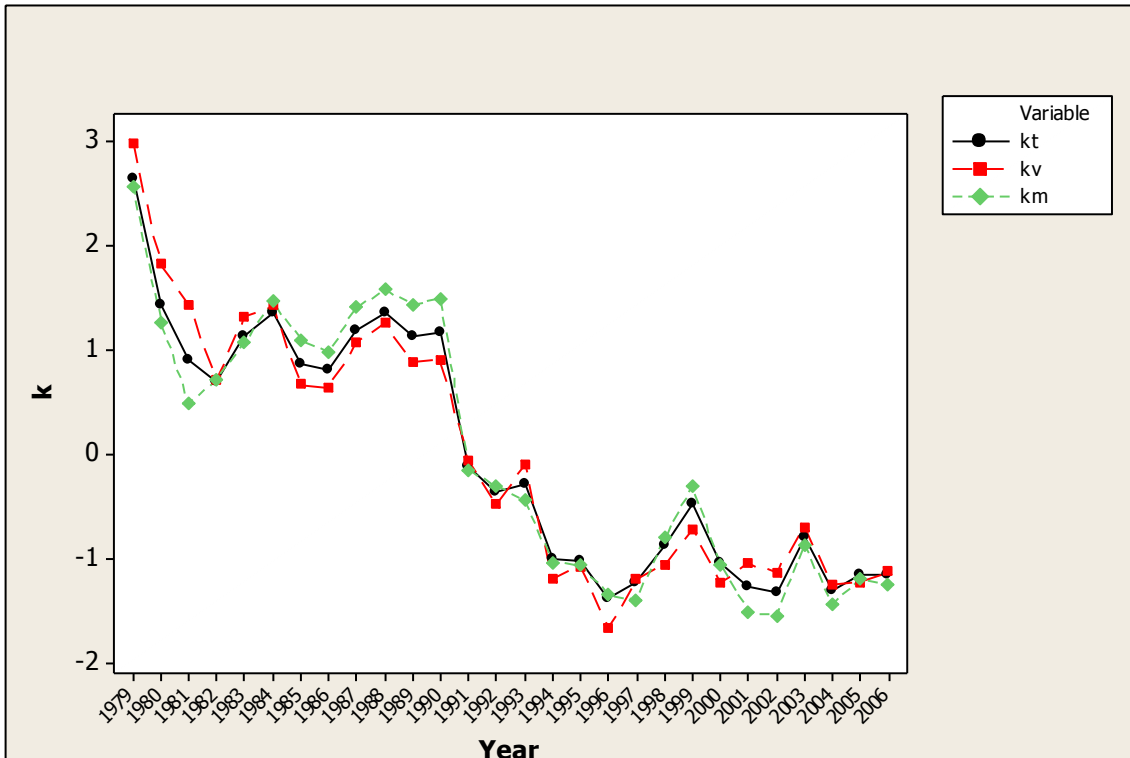
1989	0,68201	1,1338	0,76425	0,89425	0,72298	1,43817
1990	0,40443	1,17404	0,40593	0,91141	0,61473	1,49866
1991	0,36707	-0,11575	0,56892	-0,04854	0,276	-0,15126
1992	-0,98419	-0,35588	-1,25538	-0,47354	-0,59158	-0,3007
1993	0,1796	-0,28166	0,4072	-0,10186	-0,62003	-0,43351
1994	-1,08592	-1,00535	-0,91423	-1,18863	-1,23556	-1,03062
1995	-1,16639	-1,01215	-1,17211	-1,06469	-1,31352	-1,0579
1996	-1,57738	-1,38101	-1,76854	-1,65559	-1,38672	-1,34341
1997	-0,83572	-1,23078	-0,89824	-1,19249	-0,84178	-1,38836
1998	-1,4376	-0,8623	-1,86075	-1,05187	-0,89544	-0,79416
1999	-1,33849	-0,47786	-1,81677	-0,71647	-0,9077	-0,30022
2000	-0,91499	-1,02872	-0,44133	-1,21912	-1,41228	-1,05382
2001	-0,80205	-1,26871	-1,12861	-1,04299	-0,52139	-1,50831
2002	-0,68539	-1,31579	-0,52049	-1,12432	-0,93886	-1,54172
2003	-0,7349	-0,78399	-0,80514	-0,69022	-0,62289	-0,85785
2004	-0,83147	-1,29632	-0,30746	-1,24558	-1,5001	-1,42933
2005	-0,69321	-1,15294	-0,34469	-1,22569	-1,09448	-1,1847
2006	-0,60165	-1,14738	-0,72007	-1,11737	-0,53943	-1,24187

En el Gráfico IV.2.2.2 puede verse una estructura similar en el parámetro α_x , en ambos grupos, una leve diferencia se detecta en el intervalo de 15 a 24 años, ya que en el caso de los varones, el valor del parámetro “de forma” presenta una suba claramente más marcada con respecto al valor correspondiente al mismo intervalo, para las mujeres.

La Tabla IV.2.2.2 presenta las estimaciones y re-estimaciones del índice general de mortalidad para el total y ambos géneros de Rosario.

Al representar el índice k para varones y mujeres, Gráfico IV.2.2.3, se detecta un gran cambio entre el período previo y posterior al año 1991, que se podría deber a los datos que brinda el Censo Nacional de Población y Vivienda de 1991. Este cambio genera una gran variabilidad en el índice estimado en los dos géneros.

Gráfico IV.2.2.3 Estimación del índice general de mortalidad k_t correspondiente a varones y mujeres de Rosario.



Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

Otro aspecto importante es que a partir del año 1993 no se detecta una franca tendencia a descender como sucede generalmente en las diferentes regiones en las que se ha implementado este modelo en los últimos tiempos. Este suceso podría deberse a que se han utilizado datos de población que se obtienen a partir del Censo, que en caso del año 2001 se cree tiene un valor de población total que subestima a la real.

IV.2.3 Pronósticos para el índice general de mortalidad de Rosario

IV.2.3.1 Modelo ARIMA

Diagnóstico y selección del modelo

En la Tabla IV.2.3.1.1 se presentan los valores de las estadísticas utilizadas para analizar los supuestos de los modelos planteados. En general ambos modelos presentan valores de probabilidad asociados del test de Portmanteu que indican un proceso del error ruido blanco. Asimismo el Test de

Jarque-Bera revela un comportamiento normal, un proceso homocedástico y ausencia de outliers.

Tabla IV.2.3.1.1 Valor de probabilidad asociado de la estadística de Box-Ljung y test de Jarque-Bera para los dos modelos ARIMA propuestos para Rosario, para varones y mujeres.

Modelo		Q(3)	Q(6)	Q(9)	Jarque-Bera
ARIMA(0,1,0) Con constante	Total	0,8506	0,9845	0,7548	3,5943(0,166)
	Varones	0,6028	0,9075	0,7765	1,8301(0,40)
	Mujeres	0,06851	0,2446	0,4303	0,5184(0,771)
ARIMA(1,1,0) Con constante	Total	0,8718	0,9875	0,788	3,9276(0,14)
	Varones	0,9247	0,9925	0,7398	1,9699(0,374)
	Mujeres	0,0818	0,2757	0,527	0,7845(0,676)

Se plantean múltiples modelos, los dos que presentan menor AIC y relativamente bajo BIC son ARIMA(0,1,0) con pendiente y ARIMA(1,1,0) con pendiente. Los restantes modelos planteados reflejan un aumento en ambas medidas. Los resultados se presentan en la Tabla IV.2.3.1.2.

Tabla IV.2.3.1.2 Criterios de Akaike y Schwartz para los dos modelos ARIMA propuestos para Rosario, para varones y mujeres.

Modelo		AIC	BIC
ARIMA(0,1,0) Con constante	Total	38,08	40,74
	Varones	42,62	45,28
	Mujeres	44,43	47,09
ARIMA(1,1,0) Con constante	Total	40,04	44,04
	Varones	43,45	47,45
	Mujeres	41,6	45,6

Dado que la finalidad del análisis es pronosticar el índice k se decide optar por el modelo que presente una medida menor del PSMape, medida que se calcula modelando 25 observaciones y comparando los pronósticos obtenidos con las tres observaciones que se dejaron fuera de la muestra. La Tabla IV.2.3.1.3 indica que el modelo ARIMA(0,1,0) con constante presenta menores valores de PSMape para el total y mujeres, siendo levemente más alto en los varones con respecto al otro modelo propuesto. Por ello se decide utilizar el modelo ARIMA(0,1,0) con constante para continuar el análisis.

Tabla IV.2.3.1.3 PSMape para modelos ARIMA propuestos para Rosario, para varones y mujeres.

Modelo		PSMAPE
ARIMA(0,1,0)	Total	0,1380054

Con constante	Varones	0,17960015
	Mujeres	0,12349454
ARIMA(1,1,0) Con constante	Total	0,17395719
	Varones	0,15302413
	Mujeres	0,22503146

A continuación se presentan las estimaciones de los modelos, acompañando a cada estimación de los parámetros se encuentra entre paréntesis el correspondiente desvío standard. Las constantes estimadas en cada caso resultan significativas al 5%. Los modelos presentan, tanto para el total como para varones y mujeres, valores altos en sus variancias estimadas. Esto es de esperar, ya que al observar el comportamiento del índice k se destaca la presencia de una marcada variabilidad.

a) Modelo ARIMA(0,1,0) con constante, estimado para el total de Rosario.

$$(1 - B)\hat{Z}_t = -0,1404$$

(0,0875)

$$\sigma_T^2 = 0,2068$$

b) Modelo ARIMA(0,1,0) con constante, estimado para varones de Rosario.

$$(1 - B)\hat{Z}_t = -0,152$$

(0,0952)

$$\sigma_V^2 = 0,2447$$

c) Modelo ARIMA(0,1,0) con constante, estimado para mujeres de Rosario.

$$(1 - B)\hat{Z}_t = -0,1727$$

(0,0985)

$$\sigma_M^2 = 0,2617$$

Para los tres casos las series se comportan como un camino aleatorio con constante, reflejando la existencia de tendencia estocástica y determinística.

Pronósticos ARIMA

A partir del modelo elegido se realizan los pronósticos para el total, varones y mujeres, acompañados de su correspondiente intervalo del 95% de

confianza. Los intervalos obtenidos en los tres casos estudiados presentan una amplitud similar, que se incrementa a medida que aumenta el horizonte de pronóstico. En consistencia con la situación actual las mujeres presentan valores del índice k menores que los correspondientes a los varones.

Tabla IV.2.3.1.4 Pronósticos Índice k para el total de Rosario.

Año	Pronóstico	I(95%)	S(95%)
2007	-1,287824	-2,179226	-0,39642222
2008	-1,428269	-2,688901	-0,16763582
2009	-1,568713	-3,112666	-0,02475946
2010	-1,709157	-3,491961	0,07364659
2011	-1,849602	-3,842837	0,14363366

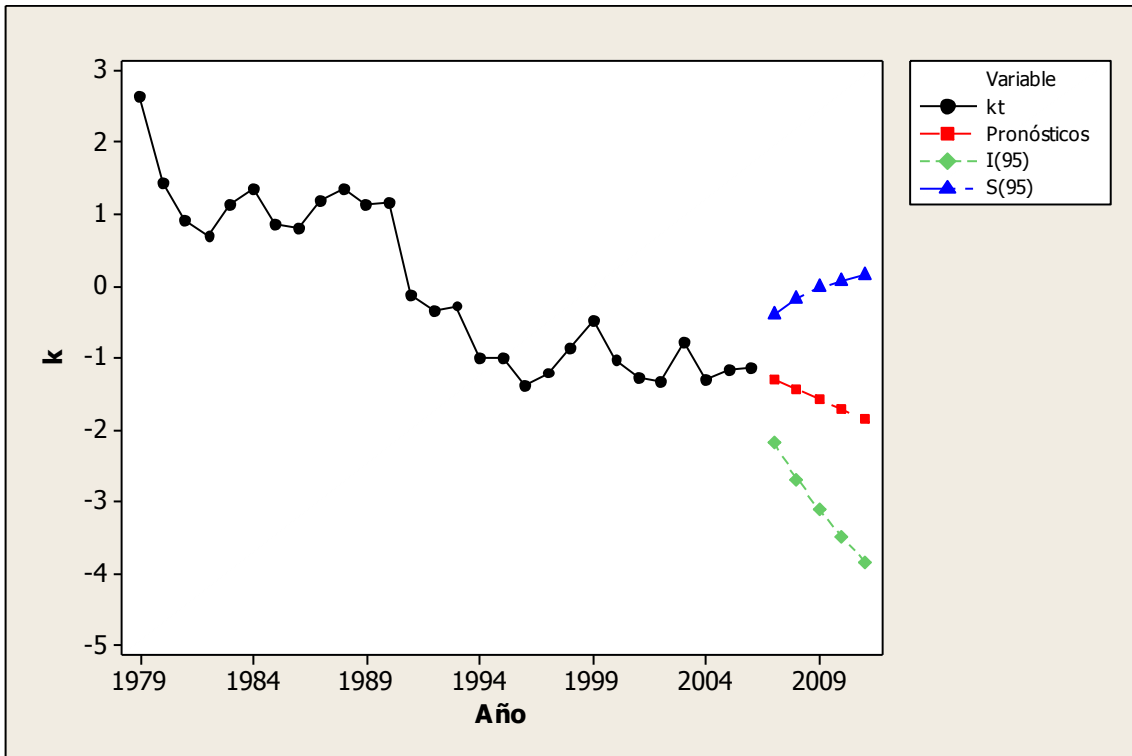
Tabla IV.2.3.1.5 Pronósticos Índice k para varones de Rosario.

Año	Pronóstico	I(95%)	S(95%)
2007	-1,269361	-2,238951	-0,29977202
2008	-1,421352	-2,792558	-0,05014558
2009	-1,573342	-3,252720	0,10603541
2010	-1,725333	-3,664512	0,21384548
2011	-1,877324	-4,045391	0,29074386

Tabla IV.2.3.1.6 Pronósticos Índice k para mujeres de Rosario.

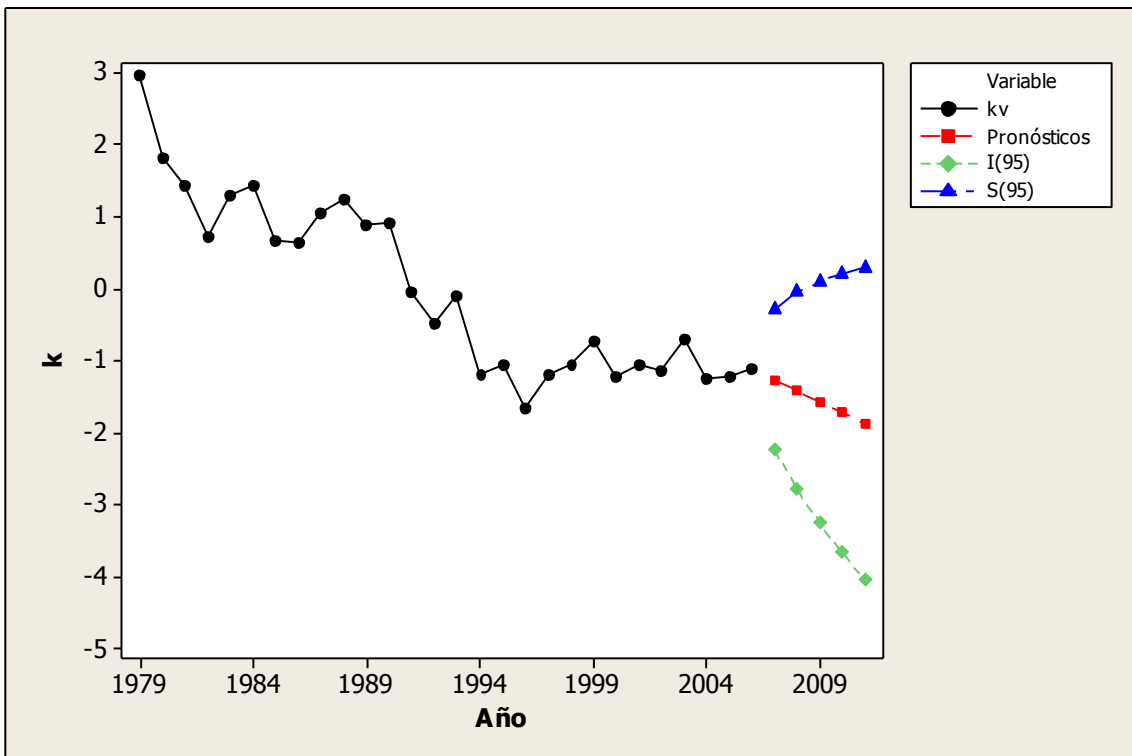
Año	Pronóstico	I(95%)	S(95%)
2007	-1,383217	-2,443273	-0,32316057
2008	-1,524567	-3,023713	-0,02542128
2009	-1,665918	-3,501989	0,17015370
2010	-1,807268	-3,927380	0,31284453
2011	-1,948618	-4,318976	0,42173949

Gráfico IV.2.3.1.1 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para el total de Rosario.



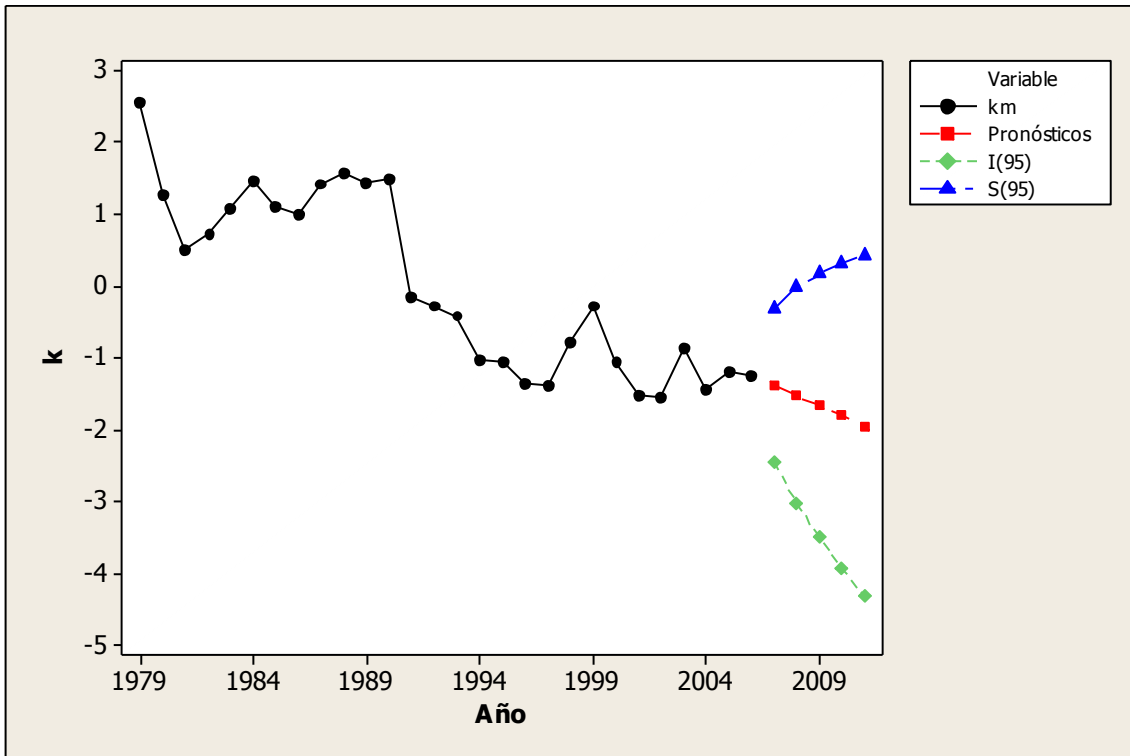
Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

Gráfico IV.2.3.1.2 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para varones de Rosario.



Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

Gráfico IV.2.3.1.2 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para mujeres de Rosario.



Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

IV.2.3.2 Modelo de espacio de estados

Las estimaciones de las ecuaciones de estado y los correspondientes hiperparámetros se presentan en las Tablas IV.2.3.2.1, IV.2.3.2.2 y IV.2.3.2.3 para el índice k de total, varones y mujeres respectivamente de la ciudad de Rosario. Se destacan las estimaciones de los hiperparámetros para pendiente y componente irregular, que resultan nulas para el total y las mujeres indicando el comportamiento determinístico de los mencionados componentes. En el modelo estimado para el índice k calculado para los varones se observa que tanto la pendiente como la componente irregular presentan comportamiento estocástico, expresado por los hiperparámetros 0,00122 y 0,04631. El hiperparámetro correspondiente al nivel presenta un valor alto en los tres casos; 0,2148 para el total, 0,15162 para varones y 0,30378 para las mujeres. Los tres modelos presentan una componente de nivel con comportamiento estocástico. Dado que la estimación de la pendiente no resulta significativa en ninguno de los tres casos se decide estimar los modelos sin dicho parámetro, pero se presentan problemas de convergencia en el algoritmo de estimación. Por ello se decide trabajar con los modelos expuestos en las Tablas IV.2.3.2.1,

IV.2.3.2.2 y IV.2.3.2.3 dado que se pretende utilizar el mismo modelo para el total, varones y mujeres, y los mencionados resultados son los que mejor aproximan un modelo común.

Ajuste del modelo

Tabla IV.2.3.2.1 Estimación MEE para k total.

	Parámetro		Estimación	
Hiperparámetros	Nivel		0,2148	
	Pendiente		0,0000	
	Componente Irregular		0,0000	
Ecuaciones de Estado	Coeficiente	Estimación	R.M.S.E	p-value
	Nivel	-1,1474	0,0000	
	Pendiente	-0,14044	0,08919	0,1274

Tabla IV.2.3.2.2 Estimación MEE para k varones.

	Parámetro		Estimación	
Hiperparámetros	Nivel		0,15162	
	Pendiente		0,00122	
	Componente Irregular		0,04631	
Ecuaciones de Estado	Coeficiente	Estimación	R.M.S.E	p-value
	Nivel	-1,1587	0,19487	0,0000
	Pendiente	-0,09284	0,12149	0,4516

Tabla IV.2.3.2.3 Estimación MEE para k mujeres.

	Parámetro		Estimación	
Hiperparámetros	Nivel		0,30378	
	Pendiente		0,00000	
	Componente Irregular		0,00000	
Ecuaciones de Estado	Coeficiente	Estimación	R.M.S.E	p-value
	Nivel	-1,2419	0,00000	
	Pendiente	-0,14135	0,10607	0,1942

La evaluación de supuestos se realiza a través de la estadística de Box-Ljung y el test de Jarque-Bera (Tabla IV.2.3.2.4). Los valores de probabilidad asociada calculados para 3, 6 y 9 rezagos indican que el proceso del error sigue un proceso ruido blanco (errores no correlacionados).

El test de Jarque-Bera arroja valores disímiles. En el caso del modelo estimado para el total, no se rechaza la hipótesis del test, pero presenta un valor cercano al 0,10. En cuanto a los varones, la estadística calculada para los errores presenta un valor de 0,80, lo que indica que los errores se distribuyen normalmente y no están correlacionados. La estadística calculada para las mujeres posee un valor de probabilidad asociada de 0,0368, lo que revela que el proceso del error en este caso puede presentar puntos aberrantes o ser

heterocedástico. En un estudio más exhaustivo no se detectan outliers para el índice k_t estimado para mujeres, en cambio la serie presenta una marcada asimetría.

Tabla IV.2.3.2.4 Valor de probabilidad asociado de la estadística de Box-Ljung y test de Jarque-Bera para el MEE de Rosario.

	Q(3)	Q(6)	Q(9)	Jarque-Bera
Total	0,9855	0,9976	0,8427	3,9696(0,1374)
Varones	0,9783	0,9973	0,8489	0,4449(0,8000)
Mujeres	0,9417	0,9941	0,8722	6,6012(0,0368)

La Tabla IV.2.3.2.5 exhibe los valores de los errores medios absolutos porcentuales post muestrales calculados realizando tres predicciones dejando los tres últimos valores fuera del modelo. Un valor relativamente mayor se observa para los varones, mientras que para el total y las mujeres los valores son cercanos al 0,10.

Tabla IV.2.3.2.5 PSMAPE para los dos modelos ARIMA propuestos para Rosario, para varones y mujeres.

Modelo		PSMAPE
MEE	Total	0,13799507
	Varones	0,26463157
	Mujeres	0,1234811

Pronósticos con MEE

Tabla IV.2.3.2.6 Pronósticos Índice k_t para el total de Rosario, con sus correspondientes intervalos del 95% de confianza.

Año	Pronóstico	I(95%)	S(95%)
2007	-1,2878	-2,2128612	-0,3627388
2008	-1,4283	-2,7597084	-0,0968916
2009	-1,5687	-3,2271932	0,0897932
2010	-1,7092	-3,6559308	0,2375308
2011	-1,8496	-4,060872	0,361672

Tabla IV.2.3.2.7 Pronósticos Índice k_t para varones de Rosario, con sus correspondientes intervalos del 95% de confianza.

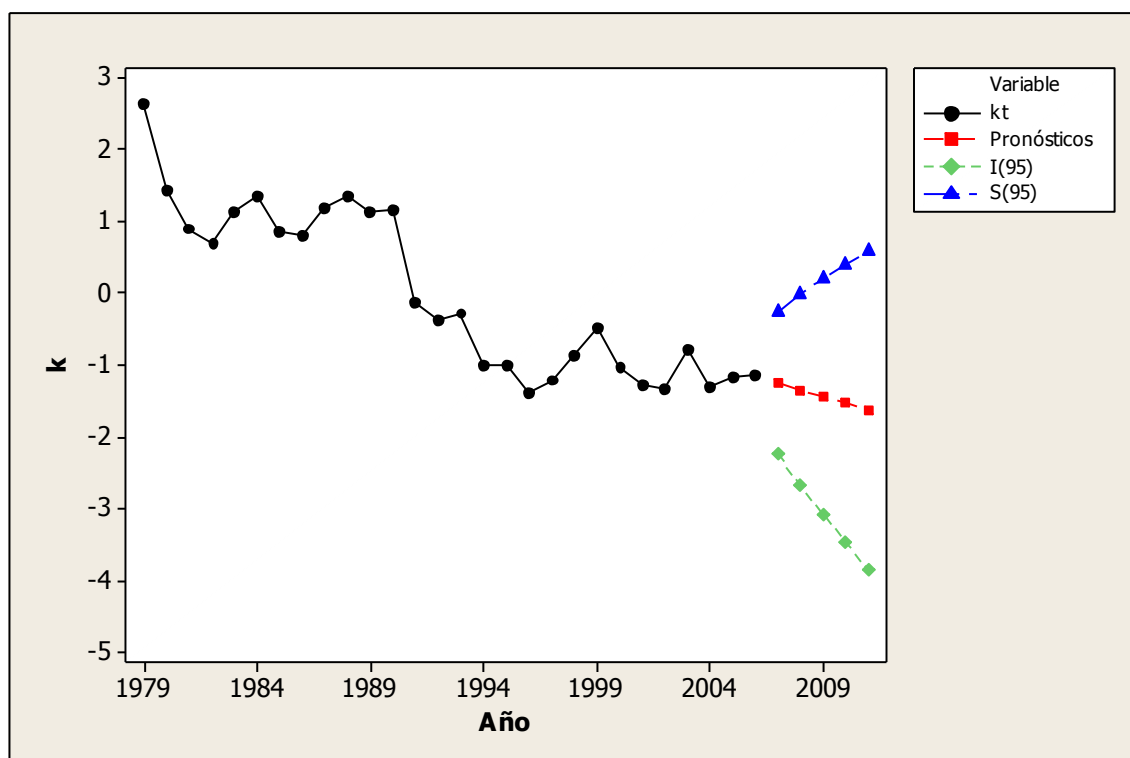
Año	Pronóstico	I(95%)	S(95%)
2007	-1,2516	-2,2455552	-0,2576448
2008	-1,3444	-2,6748872	-0,0139128
2009	-1,4373	-3,0743508	0,1997508
2010	-1,5301	-3,4605236	0,4003236
2011	-1,6230	-3,840936	0,594936

Tabla IV.2.3.2.8 Pronósticos Índice k_t para mujeres de Rosario, con sus correspondientes intervalos del 95% de confianza.

Año	Pronóstico	I(95%)	S(95%)
2007	-1,3832	-2,4832892	-0,2831108
2008	-1,5246	-3,1079076	0,0587076
2009	-1,6659	-3,638248	0,306448
2010	-1,8073	-4,122256	0,507656
2011	-1,9486	-4,578332	0,681132

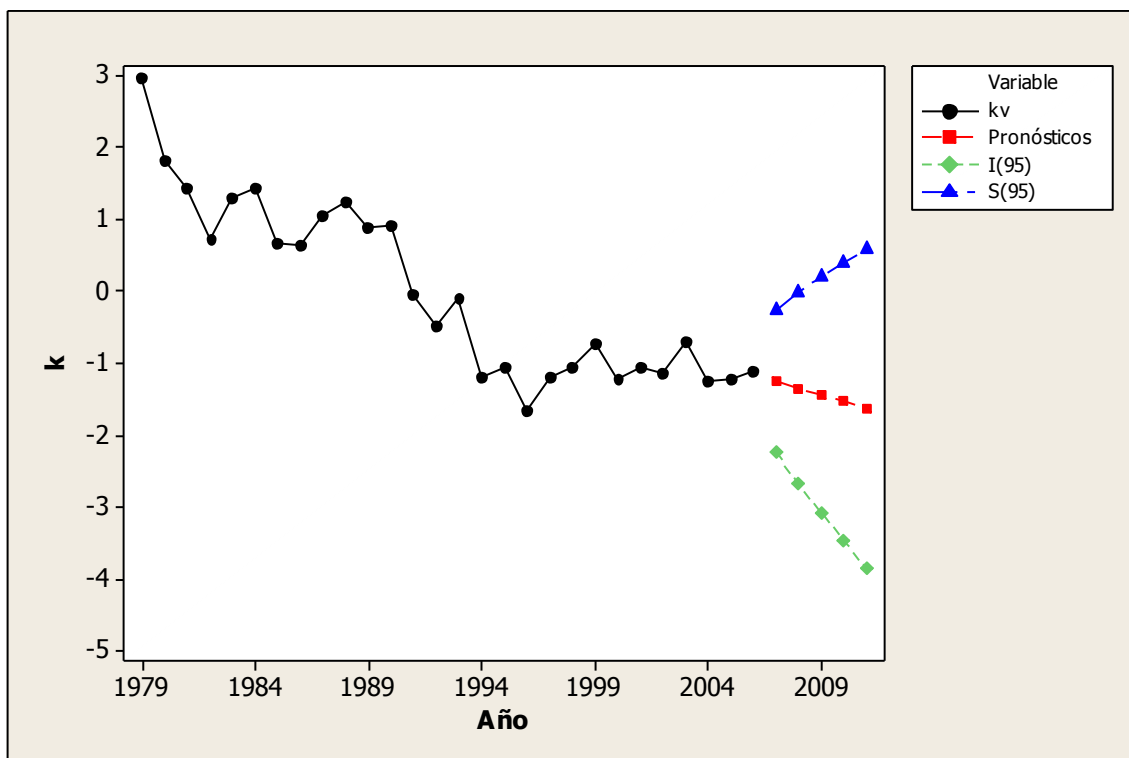
Con el modelo estimado para cada caso se generan cinco pronósticos acompañados de los intervalos del 95% de confianza. Puede verse en el Gráfico IV.2.3.2.3 una notoria amplitud en los intervalos estimados para los pronósticos del índice k de las mujeres, además se observa como dicho índice presenta un descenso notorio a través del tiempo. Mientras que los Gráficos IV.2.3.2.1 y IV.2.3.2.2 que representan el comportamiento del índice k para el total y varones muestran una amplitud importante pero no tan marcada como en el caso de las mujeres, asimismo los pronósticos presentan descenso en el tiempo pero mas leve que en el caso de las mujeres.

Gráfico IV.2.3.2.1 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para el total en Rosario.



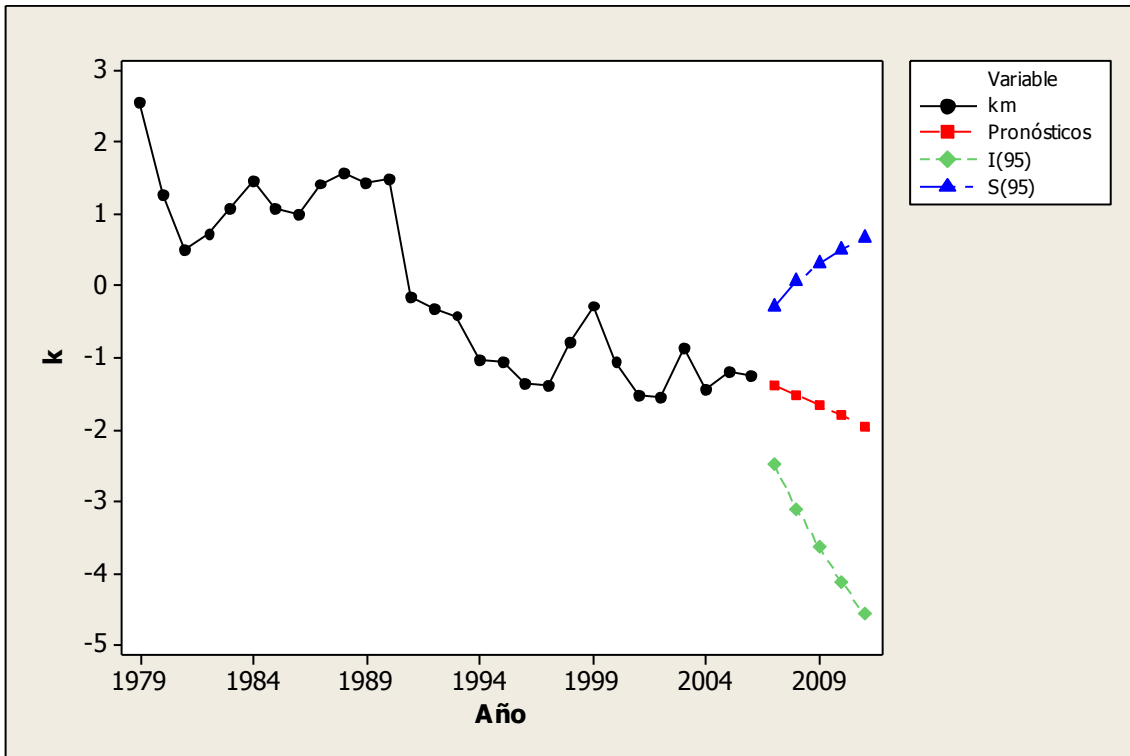
Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

Gráfico IV.2.3.2.2 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para varones en Rosario.



Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

Gráfico IV.2.3.2.3 Proyecciones e Intervalos de Pronóstico del 95% del índice k_t estimado para mujeres en Rosario.



Fuente: Elaboración propia en base a datos de la Dirección Municipal de Estadística.

IV.2.4 Pronósticos de las tasas de mortalidad específicas por edad

Con los pronósticos de cada índice se calculan las tasas de mortalidad específicas por edad, basándose en la ecuación III.3.1. Las Tablas IV.2.4.1 y IV.2.4.2 presentan los pronósticos de las tasas de mortalidad para el total obtenidas a partir de los pronósticos calculados con modelos ARIMA y MEE respectivamente. Ambas metodologías implementadas arrojan resultados ampliamente coincidentes, que difieren por lo general a partir de la tercera cifra decimal. Aunque en las tasas para el total y las mujeres es lógico que se de la coincidencia, ya que los pronósticos en base a los cuales se obtienen las tablas resultan coincidentes.

Tabla IV.2.4.1 Tasas de mortalidad específicas por edad para el total de Rosario, estimadas a partir de pronósticos ARIMA en el período 2007-2011.

	2007	2008	2009	2010	2011
0-4	0,003968	0,003815	0,003667	0,003525	0,003388
5-14	0,000215	0,000210	0,000205	0,000201	0,000196
16-24	0,000758	0,000762	0,000767	0,000772	0,000776
25-34	0,001133	0,001132	0,001131	0,001130	0,001130
35-44	0,001877	0,001844	0,001811	0,001779	0,001747
45-54	0,005055	0,005000	0,004946	0,004892	0,004839
55-64	0,012200	0,011944	0,011693	0,011447	0,011206
65-74	0,024277	0,023820	0,023373	0,022933	0,022503

74 y más	0,096374	0,095065	0,093775	0,092501	0,091245
-----------------	----------	----------	----------	----------	----------

Tabla IV.2.4.2 Estimaciones por intervalos de confianza del 95% de las tasas de mortalidad específicas por edad para el total de Rosario, estimadas a partir de pronósticos ARIMA en el período 2007-2011.

	0-4		5-14		15-24	
2007	0,00382621	0,00411504	0,0002104	0,0002197	0,00075384	0,00076218
2008	0,00367402	0,00396139	0,00020535	0,00021475	0,00075768	0,00076635
2009	0,00352716	0,00381239	0,00020032	0,00020979	0,00076251	0,00077152
2010	0,00338654	0,00366912	0,00019627	0,00020584	0,00076734	0,00077669
2011	0,00325117	0,00353059	0,00019126	0,00020086	0,00077118	0,00078085
	25-34		35-44		45-54	
2007	0,00113234	0,00113366	0,00184611	0,00190841	0,00500423	0,00513959
2008	0,00113132	0,00113268	0,0018126	0,00187594	0,00494805	0,00508661
2009	0,0011303	0,0011317	0,00177917	0,0018434	0,00489294	0,00503449
2010	0,00112928	0,00113072	0,00174678	0,00181181	0,00483793	0,00498223
2011	0,00112926	0,00113074	0,00171446	0,00178016	0,00478398	0,00493084
	55-64		65-74		75 y más	
2007	0,01196349	0,01244119	0,02385621	0,02470521	0,0951672	0,09759611
2008	0,01170447	0,01218844	0,02339288	0,02425492	0,09383338	0,09631278
2009	0,01145094	0,01194018	0,02294037	0,02381379	0,09252079	0,09504622
2010	0,01120286	0,01169646	0,02249566	0,02337884	0,09122625	0,09379356
2011	0,01096019	0,01145732	0,02206161	0,02295322	0,08995158	0,09255701

Tabla IV.2.4.3 Tasas de mortalidad específicas por edad para el total de Rosario, estimadas a partir de pronósticos EE en el período 2007-2011.

	2007	2008	2009	2010	2011
0-4	0,003968	0,003814	0,003667	0,003525	0,003388
5-14	0,000215	0,000210	0,000205	0,000201	0,000196
16-24	0,000758	0,000762	0,000767	0,000772	0,000776
25-34	0,001133	0,001132	0,001131	0,001130	0,001130
35-44	0,001877	0,001844	0,001811	0,001779	0,001747
45-54	0,005055	0,005000	0,004946	0,004892	0,004839
55-64	0,012200	0,011943	0,011693	0,011447	0,011206
65-74	0,024277	0,023820	0,023373	0,022933	0,022503
74 y más	0,096374	0,095065	0,093775	0,092501	0,091245

Tabla IV.2.4.4 Estimaciones por intervalos de confianza del 95% de las tasas de mortalidad específicas por edad para el total de Rosario, estimadas a partir de pronósticos EE en el período 2007-2011.

	0-4		5-14		15-24	
2007	0,00364814	0,0043159	0,00020452	0,00022601	0,00074843	0,00076769
2008	0,00346859	0,00419381	0,00019848	0,00022219	0,00075114	0,00077301
2009	0,00330051	0,00407418	0,00019256	0,00021824	0,00075489	0,00077931
2010	0,00314112	0,00395579	0,00018769	0,00021526	0,00075866	0,00078557
2011	0,00298983	0,00383919	0,00018196	0,00021112	0,00076147	0,0007908
	25-34		35-44		45-54	
2007	0,00113148	0,00113452	0,00180641	0,00195035	0,00493851	0,00525253
2008	0,00113029	0,00113371	0,00176586	0,00192559	0,00487005	0,00522124
2009	0,0011291	0,0011329	0,00172608	0,00190009	0,00480363	0,00518932

2010	0,00112792	0,00113208	0,00168786	0,00187506	0,00473801	0,00515614
2011	0,00112775	0,00113225	0,00165017	0,00184951	0,00467406	0,00512294
	55-64		65-74		75 y más	
2007	0,01166064	0,01276431	0,0233161	0,0252775	0,09360942	0,09922022
2008	0,01134835	0,01256881	0,02275784	0,02493174	0,09199053	0,09824223
2009	0,011049	0,01237453	0,02221982	0,02458603	0,09041709	0,09725762
2010	0,01075849	0,01217957	0,02169697	0,02423944	0,08888025	0,09626925
2011	0,01047704	0,01198567	0,02119093	0,02389631	0,08737874	0,09528233

Se presentan además las estimaciones por intervalos de confianza del 95%, en las tablas IV.2.4.2 y IV.2.4.4, en ellas se destaca que los intervalos obtenidos para las tasas proyectadas a partir de modelos ARIMA resultan más estrechos que los obtenidos por MEE. Las tablas que contienen las estimaciones puntuales y por intervalos para cada género se presentan en el Anexo A.5.

IV.2.5 Esperanza de vida al nacer

Con las tasas de mortalidad obtenidas en el punto IV.2.4 se construyen tablas de vida aplicando la metodología propuesta por la ONS. Se calculan las tablas utilizando tres posibles factores de separación. Dichos procedimientos difieren tan solo en días y se decide incluir en el presente trabajo los resultados más conservadores. Las tablas de vida completas y para cada factor de separación se presentan en el Anexo A.7.

En los diferentes sitios en los que se ha implementado, el modelo de Lee-Carter ha generado esperanzas de vida mayores a las calculadas por métodos tradicionales introduciendo la idea de que dichos métodos “subestiman” la verdadera esperanza de vida. Debido a ello se informa la esperanza de vida que resulta menor de entre las tres tablas correspondientes a los diferentes factores de separación.

Tabla IV.2.5.1 Estimación de la esperanza de vida al nacer para el total de Rosario, obtenidas a partir de los pronósticos de las tasas de mortalidad basados en modelos ARIMA y MEE.

Año/Pronósticos	Esperanza de vida al nacer	
	ARIMA	MEE
2007	74,77	74,77
2008	75,03	75,03
2009	75,29	75,29

2010	75,55	75,55
2011	75,82	75,82

Tabla IV.2.5.2 Estimación de la esperanza de vida al nacer para varones de Rosario, obtenidas a partir de los pronósticos de las tasas de mortalidad basados en modelos ARIMA y MEE.

Año/Pronósticos	Esperanza de vida al nacer	
	ARIMA	MEE
2007	71,03	71,01
2008	71,25	71,14
2009	71,46	71,27
2010	71,66	71,40
2011	71,87	71,52

Tabla IV.2.5.2 Estimación de la esperanza de vida al nacer para mujeres de rosario, obtenidas a partir de los pronósticos de las tasas de mortalidad basados en modelos ARIMA y MEE.

Año/Pronósticos	Esperanza de vida al nacer	
	ARIMA	MEE
2007	78,05	78,05
2008	78,33	78,33
2009	78,61	78,61
2010	78,90	78,90
2011	79,18	79,18

Las estimaciones de la esperanza de vida al nacer realizadas utilizando modelos ARIMA y MEE coinciden en el caso las mujeres y el total. Este resultado es consecuencia de la igualdad en los pronósticos realizados con ambas metodologías. En cuanto a la esperanza de vida al nacer estimada para los varones se detecta una diferencia de entre cuatro y veinte días entre los resultados obtenidos por uno y otro método de pronóstico. Siendo esta diferencia pequeña en relación al cambio de un año a otro.

V.- Consideraciones Finales

En la presente tesina se aplica el modelo de Lee-Carter a las tasas de mortalidad específicas por edad y discriminadas por género, de la ciudad de Rosario y de la República Argentina. El periodo analizado abarca desde el año 1979 hasta el año 2006, el cuál si bien es algo menor al recomendado por Lee y Carter, se considera que resulta apropiado para aplicar dicha metodología. A partir del análisis propuesto por Lee y Carter se obtiene un índice general de

mortalidad para cada género, y los parámetros de forma y sensibilidad. Mediante modelos ARIMA y MEE se pronostica el índice para los años 2007 a 2011. Los puntos destacados de esta primera etapa del trabajo son:

- Los datos de Argentina se hallan completos para todo el período. Tanto las defunciones como la población se presentan en nueve categorías de edad, mientras que las de Rosario son más. Por ello se decide unir categorías en esta última para contar con igual número en ambos casos. En relación a las defunciones se observa en una primera inspección visual una menor variabilidad en los datos de la Argentina que en los de Rosario. Estos últimos presentan además falta de información en el año 1991 y un posible error de registro para el año 1989.

- Los parámetros de forma estimados por el modelo (a_x) son nueve, uno para cada categoría de edad. El comportamiento es similar para Argentina y Rosario, además es parecido al observado en la mayoría de las poblaciones donde se ha utilizado este tipo de análisis.

- En relación al parámetro b_x , llamado “de sensibilidad”, se observa que en la Argentina las estimaciones resultan más suaves que las obtenidas para Rosario en ambos géneros. Las estimaciones de Argentina aparentan reflejar un comportamiento “promedio” o “suavizado”, hecho que es lógico ya que el país abarca regiones que poseen diferentes conductas en cuanto a la mortalidad y lo que el índice refleja es un comportamiento general para todo el país. El desempeño del parámetro b_x estimado resulta similar para Argentina y Rosario para las primeras categorías de edad y se destaca una diferencia en las últimas tres categorías, es decir “55-64”, “65-74” y “75 y más”. En estas tres últimas categorías se observan para la Argentina valores pequeños y en decrecimiento, mientras que en Rosario los valores no decrecen. Lo observado en Argentina estaría más de acuerdo con los resultados obtenidos en otros países que aplicaron la misma metodología, como en Estados Unidos, (Lee y Carter, 1992), o Chile (Lee y Rofman, 1994).

- El índice general de mortalidad (k_t), describe a través del tiempo una serie que para la Argentina presenta poca variabilidad y aparece dominada por una clara tendencia decreciente. Al analizar la misma serie para Rosario se observa un comportamiento muy llamativo, con caídas marcadas al comienzo

(1979-1981) y a la mitad de período (1990-1992), presentando un nivel aparentemente constante entre dichas caídas, y desde el 1993 al 2006.

- Se selecciona el modelo ARIMA que mejor ajusta los datos para ambos géneros y el total, para la Argentina el modelo ARIMA(0,1,2) es considerado como el que mejor describe el comportamiento del índice k_t . Para la ciudad de Rosario se opta por el modelo del camino aleatorio, ARIMA(0,1,0).

- Los MEE obtenidos presentan las siguientes características:

En los modelos estimados para la Argentina los hiperparámetros estimados para la pendiente resultan nulos para el total y los varones indicando comportamiento determinístico para este componente en ambos casos. Para las mujeres dicho parámetro presenta un valor distinto de cero, que refleja la existencia de un comportamiento estocástico a diferencia de los varones y el total. El nivel presenta en los tres casos analizados un comportamiento estocástico, al igual que el componente irregular.

En el caso de Rosario las estimaciones de los hiperparámetros para la pendiente y componente irregular resultan nulas para el total y las mujeres indicando un comportamiento determinístico. Al intentar ajustar un modelo sin pendiente, se presentan problemas de convergencia. En el caso de las mujeres se detecta falta de normalidad debida a la presencia de un comportamiento asimétrico, que afecta también a la serie estimada para el total. En el modelo estimado para el índice k_t calculado para los varones se observa que tanto la pendiente como la componente irregular presentan comportamiento estocástico. El hiperparámetro correspondiente al nivel presenta un valor distinto de cero en los tres casos. Lo que indica que el componente de nivel es estocástico.

- En relación a la construcción de intervalos de pronóstico, los mismos resultan para la Argentina más estrechos que los de Rosario y por ende más informativos. Si se compara entre los dos modelos de pronóstico utilizados, en Argentina los intervalos correspondientes a los pronósticos ARIMA resultan claramente más estrechos que los obtenidos por MEE. En el caso de Rosario los intervalos de pronóstico son amplios y se ensanchan con el tiempo, además no se detecta diferencia para los obtenidos mediante uno u otro modelo.

En una segunda etapa, utilizando los pronósticos del índice k_t obtenidos por ambos modelos, se pronostican las tasas de mortalidad específicas por

edad. Se emplean además para el cálculo de las tasas las estimaciones de los parámetros de forma y sensibilidad (a_x y b_x). A partir de dichas tasas se calculan las esperanzas de vida al nacer utilizando la metodología propuesta por la ONS (Office for National Statistics, del reino Unido). En cuanto estos resultados se destaca:

- Las tasas de mortalidad específicas por edad obtenidas con las estimaciones de los modelos ARIMA o MEE difieren para Argentina a partir del tercer o cuarto decimal, mientras que para Rosario resultan idénticas para el total y las mujeres, y difieren levemente para los varones. Esta situación es consecuencia de los pronósticos muy parecidos que se obtienen para ambos casos.

- Por último se realizan las estimaciones de una medida resumen muy útil: la esperanza de vida al nacer. Para la Argentina los valores difieren levemente entre los generados a partir de modelos ARIMA y los obtenidos por MEE. Las diferencias son de un mes a dos meses aproximadamente. En Rosario las estimaciones coinciden para el total y las mujeres. Las esperanzas de vida calculadas para los varones presentan diferencias de veinte días o menos entre uno y otro método utilizado.

- Se destaca que en ambos géneros las esperanzas de vida obtenidas para Rosario resultan aproximadamente un año menor que las estimaciones de esperanza de vida al nacer para la Argentina.

Es importante remarcar que para ambos modelos de pronóstico utilizados las estimaciones de las tasas y las esperanzas de vida son similares. Es indistinto el uso de modelos ARIMA o de MEE a la hora de pronosticar el índice general de mortalidad, si bien el primero posee intervalos de confianza más precisos.

Como no se calculan pronósticos de la esperanza de vida al nacer para la ciudad de Rosario mediante el método tradicional en forma oficial, no se pueden realizar comparaciones con los resultados que se obtienen en esta tesina.

En la República Argentina como en todos los centros urbanos donde se ha aplicado el modelo de Lee-Carter, las esperanzas de vida pronosticadas por dicho modelo resultan mayores a las generadas por los métodos empleados por organismos oficiales de estadística. Este hecho probablemente se deba a

que la mortalidad tiende a estabilizarse y no solo a seguir el comportamiento observado en el pasado.

Finalmente se destaca el uso del modelo de Lee-Carter como una metodología con sustento probabilístico. Además permite incorporar diferentes métodos para realizar pronósticos, en este caso modelos ARIMA y MEE. El modelo de Lee-Carter es práctico, de sencilla aplicación y genera múltiples medidas y resultados que permiten describir la mortalidad actual como así también la mortalidad en el futuro. En esta tesina su aplicación brinda buenos resultados para el caso de los datos de Argentina, coincidentes con los obtenidos por mismo método en otros países. Sin embargo la aplicación para la ciudad de Rosario presenta resultados más erráticos que pueden deberse a la calidad de los datos.

Por último como se dijo anteriormente el modelo de LC se basa exclusivamente en la información histórica de datos de mortalidad y población, por ello es indispensable contar con información confiable por períodos considerables de tiempo, mostrando la importancia que tiene para un país, región o ciudad obtener los datos en forma eficiente, mantener los registros actualizados y aplicar la metodología en forma reiterada cada vez que se disponga de nuevos datos.

VI.- Referencia bibliográfica y Software

Referencia bibliográfica

Alho, J.M. (1990). "*Stochastic methods in population forecasting*". International Journal of Forecasting 6, 521-530.

Alho, J.M. (1991). "Effect of aggregation on the estimation of trend in mortality". Mathematical Population Studies 3, 53-67.

Alho, Juha M. (2000). "A statistical look at Modeen's forecast of the population of Finland in 1934." Yearbook of Population Research in Finland, Vol. 36, 2000, 107-20 pp.

Alho, J.M.; Spencer B.D. (1985). "Uncertain Population Forecasting" Journal of the American Statistical Association 80 (1985): 306-314.

Alho, J.M.; Spencer B.D. (1990). "Error Models for Official Mortality Forecasts." Journal of the American Statistical Association Vol. 85, No. 411. American Statistical Association.

Blaconá, M. T., Abril, J. C. (2000) "Modelo estructural de espacio de estado para la demanda diaria promedio de energía eléctrica en la República Argentina."

Bell W R. (1997). "Comparing and assessing time series methods for forecasting age specific fertility and mortality rates." Journal of Official Statistics, 13 (3): 279–303.

Bell, W.R., Monsell, B., (1991). "Using principal components in time series modelling and forecasting of age-specific mortality rates." Proceedings of the American Statistical Association, Social Statistics Section, pp. 154-159.

Booth, H.; Maindonald, J.; Smith, L. 2002. "Applying Lee-Carter under conditions of variable mortality decline". *Population Studies* 56, 325-336.

Booth, H.; Hyndman, R.J.; Tickle, L.; de Jong, P. 2006. "Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions". *Demographic Research* 15, 289-310.

Booth, H; Maindonald, J. and Smith, L. (2002). "Age-Time Interactions in Mortality Projection: Applying Lee-Carter to Australia."

Booth, H. & Tickle, L. (2003). *"The future aged: New projections of Australia's elderly population."* Australasian Journal on Ageing, 22(4), pp. 196-202.

Brouhns N., Denuit M. (2001), *"Tables de mortalite prospectives pour la population belge"*, Discussion Paper 0138, Institut de Statistique, Univ. Catholique de Louvain.

Buettner, T. (2002). *"Approaches and Experiences in Projecting Mortality Patterns for the Oldest Old."* International Symposium "Living to 100 and Beyond: Survival at Advanced Ages" (Lake Buena Vista, Florida, January 17-18, 2002)

Carter, L. (1996). *"Forecasting U.S. Mortality: A comparison of Box-Jenkins ARIMA and Structural Time Series Models."* The Sociological Quarterly 37, No 1:127-44.

Chia, N.C. and A.K.C. Tsui (2003). *"Life annuities of compulsory savings and income adequacy of the elderly in Singapore."* Journal of Pension Economics and Finance, March 2003

Chiang, C. L. (1978). *"Life Table and Mortality Analysis"*. World Health Organisation".

Chiang, C. L. (1984). *"The Life Table and its Applications"*. Malabar (FL), Robert E. Krieger Publ. Co.

DeMoivre. (1725). *"Annuities upon Lives"*

Departamento de Asuntos Económicos y Sociales. Oficina de Estadística de las Naciones Unidas. Naciones Unidas. (1955). *"Manual de Métodos de Estadísticas Vitales"*. Estudios Metodológicos. Serie f, Nro.7. Nueva York, EE.UU.

Departamento de Asuntos Económicos y Sociales. Oficina de Estadística de las Naciones Unidas. Naciones Unidas. (1974) *"Principios y Recomendaciones para un Sistema de Estadísticas Vitales"*. Informes Estadísticos. Serie M, Nro.19, Rev.1. Nueva York, EE.UU.

García, V., Primante, D.A. (1989). *"Mortalidad. Apuntes de clase"*. CELADE, San José. Costa Rica.

Giroso, F., King G. (2008). "Demographic forecasting". Princeton University Press.

Gompertz B. (1825). "On the nature of the function expressive of the Law of human mortality".

González Pérez C.Y., Guerrero Guzmán V. M. (2007) *"Pronósticos estadísticos de mortalidad y su impacto sobre el Sistema de Pensiones de México."*

Haberman S., Russolillo M. (2005) *"Lee Carter Mortality Forecasting: Application to the Italian Population."* Actuarial Research Paper No. 167 ISBN 1 901615-93-6

Koissi M.C., Shapiro A. F., Högnäs G. (2005) *"Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval."* Insurance: Mathematics and Economics 38 (2006) 1–20

Lee, R. D., Carter, L. (1992). *"Modeling and Forecasting the Time Series of U.S. Mortality."* Journal of the American Statistical Association 87:659-71.

Lee R D, Miller T. (2001). "Evaluating the performance of the Lee-Carter method for forecasting mortality." Demography, 38 (4): 537–549.

Lee, R.; Miller, T. (2000). *"Evaluating the performance of Lee Carter mortality forecasts."* University of California, Berkeley.

Lee, R. D. and Nault, F. (1993). *"Modeling and Forecasting Provincial Mortality in Canada."* Presented at the World Congress of the International Union for the scientific Study of Population, Montreal.

Lee, R. D. and Rofman, R. (1994). *"Modeling and Forecasting Mortality in Chile."* *Notas* 22, No. 59:182-213.

Lee, R.D.; Tuljapurkar, S. (2000). *"Population forecasting for fiscal planning: issues and innovations"*, pp. 7-57 in Auerbach, A.; Lee, R.D. 2000. *Demography and Fiscal Policy*. Cambridge, Cambridge University Press.

McNown, R. F., and Rogers, A. (1989), *"Forecasting Mortality: A Parameterized Time Series Approach."* *Demography*, 26, 645-660.

McNown, R. F., and Rogers, A. (1992). *"Forecasting Cause-Specific Mortality Using Time Series Methods."* *International Journal of Forecasting* 8 (1992) 413-423.

Makeham W. (1860). *"On the Law of mortality and the construction of annuity tables"*. *Journal of the Institute of Actuaries and Assurance Magazine* 8.

Ministerio de Salud. Dirección de Estadística e Información de Salud. Programa Nacional de Estadísticas de Salud. (2000). *"Modelos de Formularios e Instructivos del Sistema de Estadísticas Vitales"*. Buenos Aires, Argentina.

Ortega A. (1987), *"Tablas de mortalidad"*, CELADE, Serie E N° 1004, San José, Costa Rica.

Renshaw, A.; Haberman, S. 2003. *"Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections"*. *Applied Statistics* 52, 119-137.

Serfaty, E (2003) *“Mortalidad por causas violentas en adolescentes y jóvenes de 10 a 24 años. Argentina 1991 a 2000”*. Centro de Investigaciones Epidemiológicas de la Academia Nacional de Medicina de Buenos Aires.

Toson, B., Baker, A. (2003) *“Life expectancy at birth: methodological options for small populations”*. Office for National Statistics.
http://www.statistics.gov.uk/methods_quality/publications.asp

Tuljapurkar, S.; Li, N.; Boe, C. 2000. *“A universal pattern of mortality decline in the G7 countries”*. *Nature* 40, 789-792.

Wang, J. Z.,(2007) *“Fitting and Forecasting Mortality for Sweden: Applying the Lee-Carter Model”*. Mathematical Statistics Stockholm University.

Wei, W. W. S. (1990) *“Time series analysis, Univariate and Multivariate Methods”*. Addison-Wesley Publishing Company, Inc.

Wilmoth, J. R. (1993). *“Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality change.”* Technical Report, Department of Demography”, University of California, Berkeley.

Wilmoth, John R. (1996). *“Mortality Projections for Japan: A comparison of Four Methods.”* Health and Mortality among Elderly Population, Eds Graziella Caselli and Alan Lopez. New York: Oxford University Press.

Software

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Stamp (2000): Structural Time Series Analyser, Modeller and Predictor,
London: Timberlake Consultants Press. Koopman S.J., Harvey, A.C., Doornik,
J.A. and Shephard, N.

MINITAB® Release 14.12.0. (1972 – 2004) Minitab Inc.