

UN ANÁLISIS SOBRE LAS POSIBILIDADES DE PREDICCIÓN DE LA MORTALIDAD FUTURA APLICANDO EL MODELO LEE-CARTER

Amancio Betzuen ⁽¹⁾

Resumen

Hasta finales del siglo pasado los actuarios se dedicaban a estudiar la mortalidad de colectivos de personas en función, únicamente de la edad. Se trataba de estudios de tipo estático y su aplicación práctica proporcionaba sobreestimaciones de la mortalidad con el consiguiente riesgo de longevidad en aplicaciones tales como la Seguridad Social, Planes de Pensiones, Seguros de Vida, entre otros.

Estudios de finales del siglo pasado, basándose en análisis estocásticos e introduciendo la evolución de la mortalidad a través del tiempo de calendario, proporcionaban proyecciones de la mortalidad futura en términos mucho más reales, es lo que se conoce como la evolución dinámica de la mortalidad.

Nosotros hemos estudiado la viabilidad de uno de estos métodos, quizás el más conocido como es el de Lee-Carter. En primer lugar hemos contrastado la validez del modelo utilizando un criterio tan asumible como el “back-testing”. En segundo lugar hemos elegido un intervalo factible para la validez de la proyección y finalmente lo hemos aplicado a un colectivo como es la población española, para ambos sexos, contrastando el resultado y proyectándolo hasta el año 2030.

Keywords: Lee-Carter model, Mortality forecasting, Time series, Life expectancy.

INTRODUCCION

Hasta épocas recientes los actuarios nos dedicábamos a realizar ajustes sobre la evolución de la mortalidad teniendo en cuenta, como variable, únicamente

¹ Instituto de Estudios Financiero-Actuariales. Departamento de Economía Financiera I. Universidad del País Vasco/Euskal Herriko Unibertsitatea. amancio.betzuen@ehu.es. Este artículo ha sido recibido en versión revisada el 7 de julio de 2010.

la edad y por lo tanto se trataban de ajustes estáticos. Esto es, para un determinado año de calendario. Este tipo de graduaciones se conocen como unidimensionales.

Desde hace no muchos años, los actuarios se preocupan más de medir la evolución de la mortalidad a través del paso de los años de calendario al mismo tiempo que para cada uno de los años biométricos. Así lo hicimos nosotros en nuestro artículo publicado en ANALES (1999). En este caso realizamos ajustes de la evolución de la mortalidad, para todo el horizonte de edades de una persona y para diferentes años de calendario, utilizando la fórmula de Heligman y Pollard.

Utilizando esta metodología obtuvimos valores de los parámetros, para los diferentes años de calendario y a continuación procedimos a un ajuste simple de la evolución de los valores de estos parámetros. La predicción futura de la mortalidad se obtuvo a continuación simplemente extrapolando hacia el futuro los valores de los parámetros obtenidos para el intervalo de valores históricos.

El Continuous Mortality Investigation Committee le dedicó un Working Paper (Nº 25) a las características de la metodología de proyección estocástica basada en el modelo Lee-Carter. Entre las conclusiones que se pueden sacar del estudio están las características y las implicaciones del modelo. Con este último punto conecta el trabajo que presentamos a continuación.

A nuestro juicio se trata de un modelo interesante por cuanto que proyecta los resultados históricos de mortalidad, hacia el futuro y para ello, captura la información disponible sobre la mortalidad, a lo largo de un determinado número de años de calendario, incluyendo todo el periodo de edades que puede alcanzar una persona. De esta manera proporciona unos resultados que permiten predecir la mortalidad futura para un número prudencial de años futuros de calendario.

Dicho “Working Paper” no profundiza en algunos aspectos que a nuestro juicio puede influir y, creemos que notablemente, en la predicción del modelo. Nos referimos al periodo de años histórico de calendario a considerar para obtener unos buenos resultados de los parámetros a estimar.

Por ello, hemos contrastado previamente, si el modelo proporciona buenos resultados, tomando para tal fin un periodo de calendario como [1950-1985]

y contrastamos los resultados con lo realmente acontecido durante el periodo [1986-2005] . Se trata de realizar un “back-testing” para el modelo.

Una vez detectado un aceptable tramo de años de calendario de manera que el modelo proporcione buenos resultados, lo aplicamos para la proyección hacia el futuro, tanto para todo el colectivo, como para el colectivo de personas de edades superiores a los 65 años, por ser este un colectivo importante en el conjunto de la sociedad.

El análisis se extiende para ambos sexos: hombres y mujeres y el colectivo al que se aplica el estudio es la población general española. La base de datos fue preparada a partir de los datos contenidos en “Human Mortality Database” y del INE.

Se trata de un modelo nuevo para los actuarios, aunque, otros investigadores ya han probado sus resultados, algunos con opiniones dispares. Nosotros nos proponemos probar que antes de aplicar el modelo directamente es conveniente realizar una serie de análisis previos, con el objeto de asegurarnos, en una buena medida, que el resultado que se pueda obtener sea factible.

Una vez que el periodo histórico se ha elegido con ciertas garantías de éxito, procede la elección del número de años futuros a los que parece razonable extrapolar nuestros resultados. Esto, a nuestro juicio, no debe extenderse más allá de los 20 o 25 años.

Este punto está muy relacionado con el modelo Lee-Carter el cual posibilita la elección, tanto del periodo histórico de información como del periodo de predicción. También posibilita la elección del elemento biométrico más adecuado para el ajuste biométrico. Para ello, teniendo en cuenta el criterio que hemos seguido para preparar la información, hemos considerado que el tanto central de mortalidad representaba el elemento biométrico más idóneo.

Finalmente, es necesario elegir el elemento biométrico que nos va a servir para contrastar la bondad de la predicción que pretendemos proporcionar. De entre varios de los elementos que podemos elegir, nosotros nos hemos decantado por la esperanza matemática de vida, por considerarlo uno de los elementos más significativos a la hora de contrastar la bondad de los resultados a predecir. Además es uno de los pilares en los que se apoyarán la mayoría de las previsiones demográficas, en el caso de la Seguridad Social, Planes de Pensiones, Seguros de Vida, etc.

ELECCIÓN DEL PERIODO

Como ya quedó indicado, en base al criterio utilizado para preparar la base de datos, el elemento biométrico elegido para predecir la mortalidad es, en nuestro caso, el tanto central de mortalidad. En un primer momento hemos analizado la evolución de los tantos centrales de mortalidad a lo largo del periodo de calendario 1910-2005, para diferentes grupos de edades, elegidos como significativos, como son el intervalo de 20-24 años, 40-44 años, 65-69 años y 80-85 años.

El paso siguiente consistió en elegir el periodo de años de calendario. Pues el tramo de edades en principio, lo extendemos para todo el horizonte temporal de edades. Para realizar el primer análisis hemos analizado la evolución de los tantos centrales de mortalidad, para diferentes tramos como son: [20-24], [40-44], [65-69] y [80-84], correspondientes a los años de calendario [1910,2005].

La evidencia fue clara, no deja lugar a dudas, la tendencia del tanto central de mortalidad, ha evolucionado claramente hacia la mejora, salvo algunos años de calendario puntuales. Pero sobre todo se observó una evidente cambio de tendencia después de los años 1950-1955. Este aspecto debe ser tenido en cuenta para la elección del periodo.

Si realizamos un análisis detallado de los datos citados se observa algunos detalles interesantes. Por ejemplo, por el año 1918 se da una desviación importante, posiblemente por la incidencia de una enfermedad infecciosa que se desarrolló por esa época en nuestro País.

También observamos que por los años 1936-1939 se producen asimismo alteraciones destacables. Consecuencia evidente de la guerra civil española. Estas circunstancias ocasionan alteraciones importantes en la evolución de las magnitudes biométricas del colectivo de la población española. Por tal motivo consideramos que la inclusión de este periodo de años de calendario no resulta adecuado para la estimación de los parámetros del modelo. Desvirtuando innecesariamente, a nuestro juicio, la tendencia futura.

Todo esto nos hace concluir que la estimación de los parámetros varía según que estos puntos se incluyan o no en la base de datos. Por lo tanto, si dichos puntos no fueran consecuencia de una evolución normal de los elementos que intervienen en un colectivo de personas es conveniente estudiar detalladamente la posibilidad de su exclusión de la base de datos.

Finalmente observamos que durante los años de calendario, posteriores al año 1940 y hasta las proximidades del año 1955 se dan oscilaciones importantes en los valores de los tantos centrales. En consecuencia este tramo de años de calendario también lo descartamos. Por consiguiente tomaremos como periodo de datos para nuestra predicción el intervalo de años de calendario 1960-2005.

ELECCIÓN DEL MODELO

El modelo de Lee-Carter es de tipo estocástico y se utiliza preferentemente para predecir la mortalidad futura, para cada edad de la persona. Es por tanto, un modelo bilineal, como tendremos oportunidad de analizar.

Se trata de un modelo que no presenta demasiadas dificultades en su aplicación e interpretación. Captura a la vez la influencia, tanto de las mejoras por la edad de la persona como por el transcurso del tiempo de calendario. Basta para ello elegir adecuadamente la matriz de mortalidad edad-tiempo, para el periodo de análisis seleccionado. Dado su potencial de “captura” la elección del periodo y del tramo de edades son sumamente importantes.

Por otra parte este modelo facilita el valor de un índice que refleja la variación de mejora en el tiempo con el cual podremos estimar los valores de la mortalidad hacia el futuro y de esta forma construir una tabla dinámica.

Aunque el modelo permite elegir el periodo de edades, para la toma de datos y el periodo histórico de años de calendario, nosotros extendemos el periodo de edades a todo el intervalo de supervivencia de la vida de una persona y el horizonte temporal de años de calendario, de 2006 hasta 2030.

EL MODELO

Como ya hemos indicado el modelo Lee-Carter es un modelo bilineal en las variables edad (x) y año de calendario (t). Presentaron un modelo del tipo:

$$f(x,t) = \exp(a_x + b_x * k_t + \varepsilon_{x,t}) \quad (2)$$

en donde:

² Se trata de una notación similar al presentado inicialmente por Lee-Carter (1992).

- $f(x,t)$ = Describe el tanto instantáneo de mortalidad correspondiente a la edad x y tiempo de calendario t .
- a_x = Representa una componente que depende de la edad de la persona. No depende del tiempo de calendario.
- b_x = Describe la “velocidad” con la que varía la mortalidad a cada edad x , cuando varía el nivel general de mortalidad. No depende del tiempo.
- k_t = Representa un parámetro que depende del tiempo de calendario y refleja el nivel general de declinación de la mortalidad.
- $\varepsilon_{x,t}$ = Mide el riesgo de la mortalidad a la edad x y el tiempo de calendario t . Representa el término residual a la edad x y el tiempo t .

Este modelo permite extrapolar hacia el futuro la mortalidad estimada a partir de los datos históricos. Los parámetros a_x y b_x capturan la información histórica de la mortalidad por la edad de la persona y el parámetro k_t la evolución de la mortalidad histórica por el transcurso del tiempo de calendario. El modelo, combina un enfoque paramétrico con una utilización del método estadístico de series temporales.

Nosotros hemos analizado los datos que disponíamos, los cuales ya fueron indicados, y nos pareció que el elemento biométrico más adecuado, en base a la información disponible como son el número de fallecidos a cada edad simple, el número de años de exposición al riesgo y para cada año de calendario, era el tanto central de mortalidad. De modo que para la predicción hemos fijado el modelo:

$$m_{x,t} = \exp(a_x + b_x * k_t + \varepsilon_{x,t})$$

Para su tratamiento transformamos el modelo de la siguiente forma

$$\ln(m_{x,t}) = a_x + b_x * k_t + \varepsilon_{x,t} \quad (3)$$

³ Al utilizar el modelo de esta forma nos facilita enormemente los cálculos sin embargo pudieran no reflejar correctamente algunos valores biométricos como sucedería si se tomaran los tantos brutos directamente. El CMI utilizó el modelo para valores de $\ln \mu_{x,t}$.

aunque ello nos conduzca a la aparición de desviaciones, en términos absolutos, del número total de fallecidos.

Como apuntaron Lee y Carter, su modelo no queda definido ⁽⁴⁾ si no se introducen un par de condiciones como son:

$$\sum_{j=1}^w B_j^2 = 1 \quad \sum_{i=1}^n k_i = 0 \quad (1)$$

Un análisis previo nos permite observar claramente la mejora de la mortalidad, prácticamente para todos los grupos de edades. Pero también se observa que esta mejora no es uniforme ni homogénea.

Llama la atención el cambio de evolución para el grupo de edades (15-35) en torno a los años de calendario 1990-1995 como consecuencia de los accidentes de tráfico principalmente. Este cambio de tendencia afecta ligeramente a los valores de los parámetros. No obstante los hemos mantenido incorporados en la base de datos. En el colectivo de mujeres esta incidencia no ha sido significativa y por consiguiente se acomoda mejor a la predicción del modelo.

ACEPTACIÓN DEL MODELO

En primer lugar hemos querido comprobar si el ajuste de los datos al modelo es suficientemente fiable y si al mismo tiempo proporciona resultados de la proyección futura aceptables. Para ello hemos utilizado el criterio back-testing sobre el periodo [1950-1985]. Los resultados obtenidos fueron notablemente aceptables.

A continuación hemos dedicado un tiempo para analizar la influencia de la elección del periodo de datos a través del tiempo de calendario. Es evidente que un cambio en la amplitud del periodo y en la ubicación de los datos en el tiempo modifican los valores de los parámetros del modelo y como consecuencia directa la predicción futura de la mortalidad. Hemos contrastado que un periodo muy extenso de años de calendario no mejora la calidad de la proyección futura. Al igual que un periodo demasiado corto no captura información suficiente para una buena proyección futura. Finalmente después de una serie de simulaciones hemos elegido el periodo de edades de

⁴ En el sentido de que no se obtiene una solución única para el sistema de ecuaciones planteado, debido precisamente a la interrelación del producto $b_x \cdot k_t$.

cero a 110 años y un periodo [1960-2005] para años de calendario (gráficamente una superficie). A partir de este conjunto de datos hemos estimado los parámetros del modelo.

COMPUTACIÓN DEL MODELO

Se trata de obtener los parámetros a_x , b_x y k_t minimizando la matriz

$$\text{matriz } Z = f(a, b, k) = \sum_{x,t} [\ln(m_{x,t}) - \hat{a}_x - b_x * k_t]^2$$

En orden a obtener una solución única el modelo requiere la incorporación de las dos condiciones presentadas en (1).

A partir del desarrollo

$$\frac{\partial f}{\partial a_x} = \frac{\partial f}{\partial b_x} = \frac{\partial f}{\partial k_t} = 0$$

y después de operar adecuadamente en la primera ecuación y teniendo en

cuenta que $\sum_t k_t = 0$ resulta:

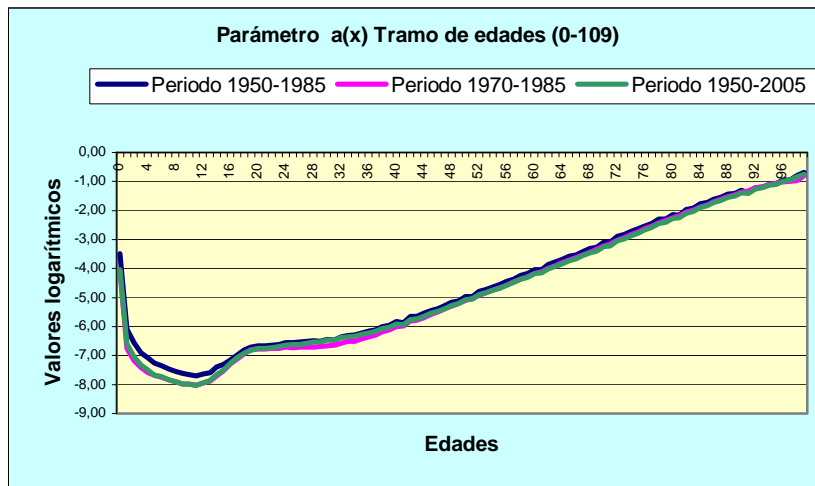
$$\hat{a}_x = \frac{1}{n} \sum_{t=t_1}^{t_n} \ln(m_{x,t}) = \ln \left[\prod_{t=t_1}^{t_n} m_{x,t}^{1/n} \right] \quad x \in [0, w)$$

De esta manera estimamos los valores del componente a_x y lo representamos por \hat{a}_x . Tal y como aparece en (2).

Es evidente que el modelo Lee-Carter no lo podemos tratar por un método de regresión ordinario, pero el parámetro a_x lo estimamos como una media geométrica, para todo el intervalo [1960-2005], para cada clase de edades, utilizando la expresión anterior.

Los valores de a_x , $\forall x$, muestran el comportamiento de la mortalidad, para cualquier año de calendario, en función de la edad. Su valor representa un invariante con respecto al tiempo. En nuestro caso se obtiene un gráfico como el que se presenta a continuación:

Gráfico N° 1. Representación gráfica del comportamiento del parámetro a_x según la amplitud y la ubicación del periodo de calendario que consideremos.



Como ya quedó indicado, de las múltiples simulaciones de periodos que hemos realizado hemos seleccionado dos, para contrastar nuestro análisis, que fueron [1950-1985] y [1970-1985].

Por la representación gráfica de la componente \hat{a}_x del modelo, percibimos el comportamiento general promedio de la mortalidad para cada edad. Reduce por tanto una superficie que depende de x y t a una línea de valores en función de x . En el gráfico se puede observar que cuando se toma el periodo [1970-1985] los resultados se desvían más que cuando se toma el periodo [1950-1985]. También se observa que cuando se incluyen los años de calendario en torno a los años 1950-1955 también se producen superiores desviaciones en los parámetros. Estas fueron algunas de las razones que nos condujeron a elegir finalmente el tramo histórico [1960-2005].⁽⁵⁾

Como decimos, con los valores apuntados en (2) construimos la matriz Z , de la siguiente manera:

⁵ Por razones de simplificación y porque los resultados son evidentes mostramos solamente la representación gráfica.

$$Z_{x,t} = \ln(m_{x,t}) - \hat{a}_x \quad (6) \quad (2)$$

A los valores de la matriz Z aplicamos el método SVD, “Singular Value Decomposition” obteniendo las matrices ⁽⁷⁾

$$PdQ^t = SVD(Z_{x,t}) = \sum d_j * P_{xj} * Q_{tj} \quad (8)$$

Habiendo introducido las condiciones (1) con el objeto de obtener una única solución para el sistema de ecuaciones del modelo.

Del sistema anterior se obtienen los valores de \hat{b}_x y \hat{k}_t . Una vez obtenidos estos valores construimos la matriz

$$\hat{Z}_{x,t} = \hat{b}_x * \hat{k}_t$$

⁶ Esto es, $Z_{x,t} = b_x * k_t$

⁷ Esta aplicación descompone la matriz $Z_{x,t}$ en el producto de tres matrices

$$SVD(Z_{x,t}) = \sum d_j * P_{xj} * Q_{tj}$$

cuyo primer término proporciona $\hat{b}_x = P_{x1}$ y $\hat{k}_t = d_t * Q_{t1}$

Por otra parte la condición

$$\frac{\partial f}{\partial a_x} = 0 \Rightarrow \sum_t \ln(m_{x,t}) = \sum_t (a_x + b_x * k_t)$$

Nos conduce a:

$$\sum_t a_x = \sum_t \ln(m_{x,t})$$

⁸ La matriz \mathbf{P} representa la componente edad, \mathbf{d} representa los valores singulares y \mathbf{Q} representa la componente tiempo.

y por consiguiente el logaritmo natural del tanto central de fallecimiento estimado

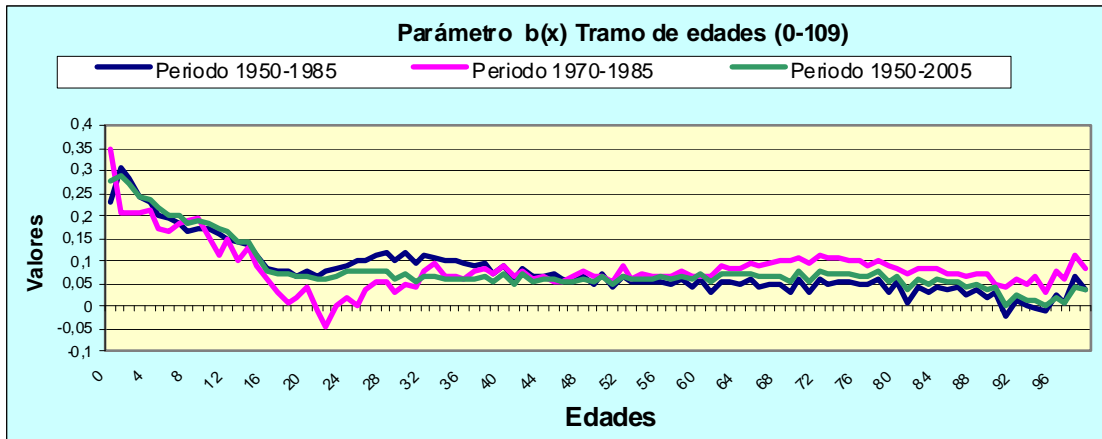
$$\ln(\hat{m}_{x,t}) = \hat{a}_x + \hat{Z}_{x,t}$$

La componente $b_x^{(9)}$ influye en la tendencia principal de $k_t^{(10)}$ modificando el resultado según que la variación a una cierta edad sea mayor o menor que la correspondiente a la tendencia principal. En principio los valores de b_x no tienen por qué mantener el mismo signo. Esto significa que las variaciones podrían ser de signo contrario.

Aunque, como apuntan algunos investigadores, esta metodología no es la óptima, para obtener las mejores estimaciones de los citados parámetros, sin embargo, hemos podido comprobar que para ciertos colectivos, proporciona buenos resultados y es más sencillo de computar. Esta representa una de sus fortalezas.

ANÁLISIS PREVIO DEL PARÁMETRO b_x

Hemos realizado un análisis previo de los diferentes resultados que se producen en la estimación del parámetro b_x . En el siguiente gráfico se presentan los resultados correspondientes a este parámetro para los tramos ya mencionados.



⁹ Recordemos que refleja la variación que experimenta el perfil en base a la edad en la medida en la que varía el parámetro k .

¹⁰ Recordemos que captura la variación en la mortalidad general a través del tiempo de calendario.

Gráfico N° 2. Representación gráfica del comportamiento del parámetro b_x según el periodo de calendario que se tome. Se representa para todo el tramo de edades y para el colectivo de hombres.

Como se puede observar conforme alteramos el periodo de datos a utilizar en el modelo la distribución de los valores del parámetro b_x cambia. En el gráfico presentamos solamente tres de las múltiples combinaciones que podemos formar. Se puede apreciar claramente que el periodo [1950-2005] proporciona una evolución más uniforme. Es evidente que el periodo [1970-1985] es demasiado corto para proporcionar información suficiente al modelo. Hemos constatado que los valores de b_x son más uniformes si el periodo [1950-2005] se reduce a [1960-2005].

ELECCIÓN DEL COLECTIVO

Dado que el modelo captura toda la información disponible en el tramo de edades y en el tramo de calendario, para el primero, en principio hemos tomado en consideración todo el conjunto de edades de vida correspondientes a una persona desde su nacimiento hasta su fallecimiento.

No obstante, a efectos prácticos también hemos tomado otro conjunto de edades superiores a los 65 años, dada su importancia en los temas relacionados con la Seguridad Social, con los Planes de Pensiones y con los Seguros de Vida, entre otros.

En cuanto al periodo de calendario y como ya se indicó anteriormente, en primer lugar queremos analizar si la elección del periodo histórico de datos tiene una influencia importante. Porque de lo que no tenemos ninguna duda es que una buena elección del periodo de información es primordial para un buen desarrollo del ajuste y de la proyección.

A partir de los datos del INE, comparados con los proporcionados por Human Mortality Database, separados por género y considerados en el sentido propuesto por Wilmoth et al (2000), se obtuvieron los valores brutos del tanto central de mortalidad para cada edad x y para cada tiempo t . Hay que tener en cuenta que algunos de los valores publicados en la base de datos (HMD), para los últimos años presentan algunas irregularidades.

A partir de los datos anteriormente mencionados se eligieron adecuadamente los valores de $D_{x,t}$ y $E_{x,t}$. En donde el numerador representa el número de fallecidos de edad x en el año de calendario t y el denominador el número

de años de exposición al riesgo de edad x en el tiempo de calendario t . Se han utilizado, en particular, valores del número de fallecimientos por edad biométrica simple y del número (valor) de los expuestos al riesgo, también por edad biométrica simple.

BACK-TESTING PREVIO

Con el objeto de contrastar la predicción que proporciona el modelo, hemos elegido un periodo histórico previo a otro, del cual ya disponemos de información histórica. Por ejemplo, podemos introducir en el modelo datos correspondientes al periodo 1950-1985, obtener la predicción que proporcione el modelo para el periodo 1986-2005 y contrastar con la obtenida en la realidad.

En una primera etapa hemos contrastado los valores reales de los tantos centrales de mortalidad con los valores estimados. Hemos constatado que el modelo no ajusta adecuadamente los tantos centrales para las edades más jóvenes. Sobre todo el primer año de edad. También se desvían los resultados para las edades más altas.

Esta evidencia la obtuvimos como sumas de los valores residuales entre los tantos centrales de mortalidad reales y los estimados, para el periodo [1950-1985]. Los hemos tomado en valor absoluto. Son los correspondientes a.

$$\varepsilon_{x,t} = \ln(m_{x,t}) - \left[\hat{a}_x + \hat{b}_x * \hat{k}_t \right]$$

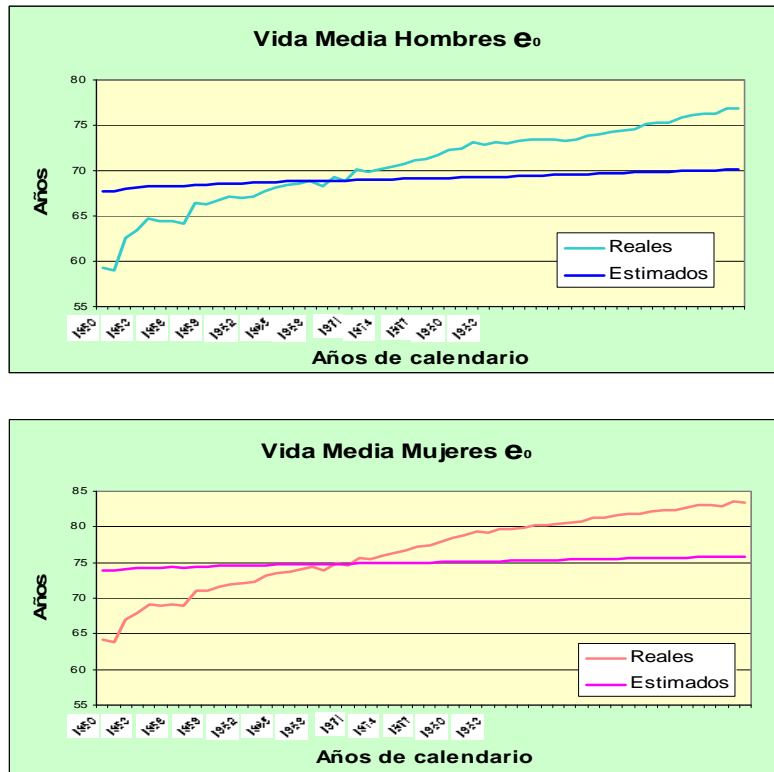
Los resultados obtenidos corresponden a la predicción con el valor inicial de $k(t)$, que lo representamos por $k_1(t)$. Como ya quedó indicado este valor inicial del parámetro $k(t)$ no proporciona un número similar de fallecimientos, a los reales, en cada año de calendario. Para ello es necesario y suficiente proceder a un reajuste del parámetro introduciendo esta nueva condición.

Después de realizado el reajuste de los valores del parámetro $k(t)$ en el sentido señalado por Lee-Carter (1992) los resultados mejoraron claramente.

Una vez comprobado que el modelo proporciona unos valores aceptables de los tantos centrales de mortalidad, sobre todo en los tramos de edades [10-90], pasamos a una segunda etapa. En esta hemos elegido el elemento biométrico vida media en el origen de edades para chequear la predicción del modelo. Ahora no se trata de un ajuste de los datos sino de comprobar si el

modelo predice bien, cómo va a evolucionar la mortalidad hacia el futuro. Los resultados que obtuvimos, mostrados gráficamente, para la esperanza matemática de vida en el origen, fueron:

Gráfico N° 3. Representación gráfica de los valores correspondientes a la vida media en el origen de edades, aplicando el parámetro $k_1(t)$. En la parte superior para el colectivo de hombres y en el inferior para el de mujeres.



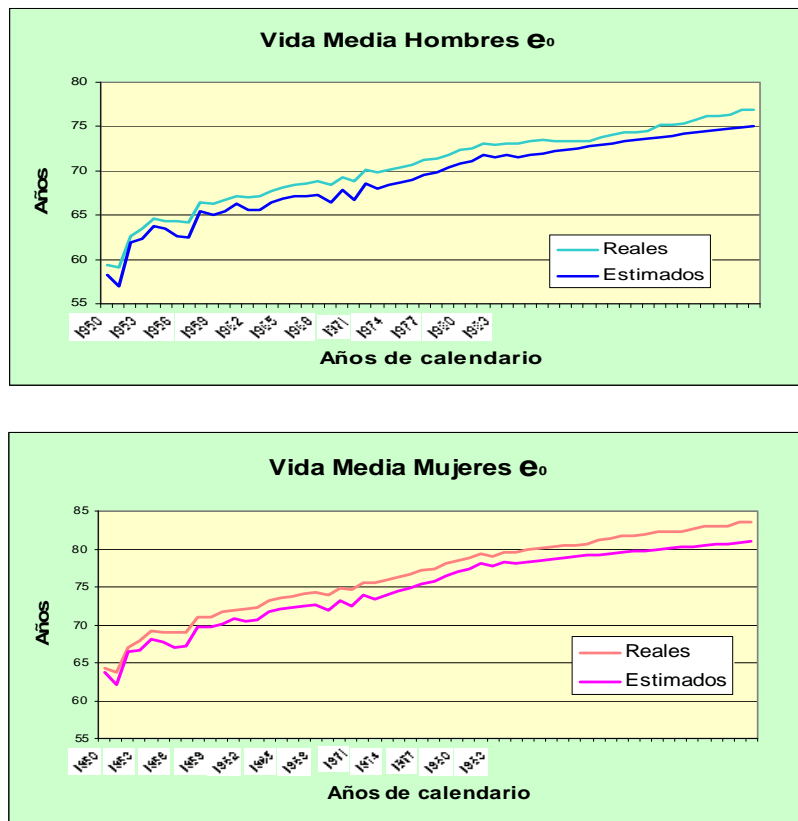
En este gráfico, los valores correspondientes al periodo [1950-1985] son los históricos y los correspondientes al periodo [1986-2005] corresponden al back-testing.

Lee-Carter en su trabajo (1992) sugirió la posibilidad de reajustar el valor del parámetro $k(t)$, de manera que el número total de fallecidos en cada año de calendario coincidiera con el número total real de fallecimientos. Mediante un método iterativo de cálculo se recalculan los nuevos valores de

$k(t)$, obteniéndose un nuevo valor $k_2(t)$.⁽¹¹⁾. Veamos si esta actuación mejora los resultados.

Lo primero que apreciamos es que la predicción se desvía de la realidad, para el tramo de años de calendario [1986-2005]. Evidentemente el modelo regulariza y suaviza la evolución pero sobreestima la evolución de la mortalidad. Ahora bien, estos resultados se obtuvieron para el parámetro estimado $k_1(t)$. Veamos lo que sucede si utilizamos en el modelo el parámetro $k_2(t)$. Gráficamente se obtienen los siguientes resultados:

Gráfico N° 4. Representación gráfica de los valores correspondientes a la vida media en el origen de edades, aplicando el parámetro $k_2(t)$. En la parte superior para el colectivo de hombres y en el inferior para el de mujeres.

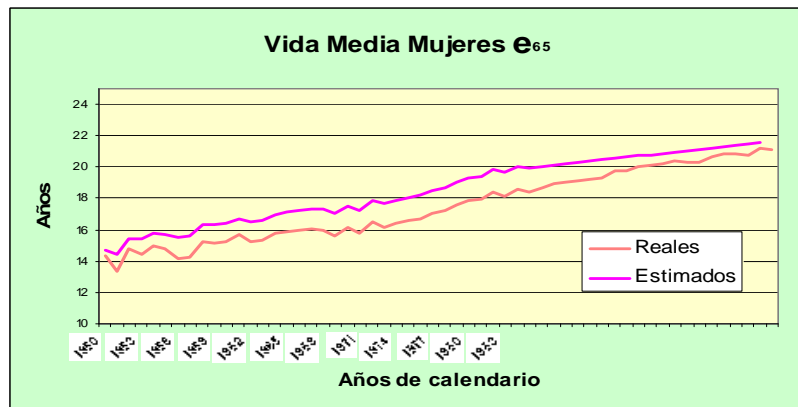
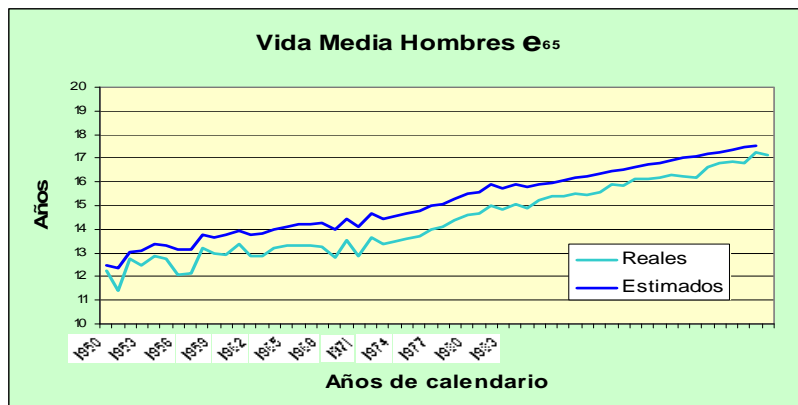


¹¹ Se obtiene el mismo resultado mediante el método de la prueba y el error.

De esta manera se logra que los valores estimados sigan el curso de los valores reales, tanto en el periodo histórico como en el periodo de predicción (en este caso forma parte del back-testing). Durante este periodo lógicamente la predicción sigue una línea suavizada pero casi paralela a los valores de la vida media, en el origen, reales.

Se puede concluir, por tanto, que los resultados que se obtienen con el nuevo valor de $k(t)$ son manifiestamente superiores en la calidad de la predicción.

Gráfico N° 5. Representación gráfica de los valores correspondientes a la vida media a los 65 años, aplicando el parámetro $k_2(t)$. Pero tomando todo el tramo de edades de la supervivencia de una persona. En la parte superior para el colectivo de hombres y en el inferior para el de mujeres.



El gráfico anterior muestra bien a las claras que la tendencia que predice el modelo y la tendencia que resultó en la realidad tienen un comportamiento muy similar. Evidentemente, esta tendencia, la que predice el modelo, cambiará conforme cambiamos la amplitud del periodo de calendario. Esto significa que un estudio de mayor profundidad implicaría el cálculo de la optimización del intervalo de años, lo cual no es el objetivo de este trabajo. En cualquier caso hemos constatado que periodos de tiempo muy cortos, como por ejemplo 10 años, no mejora los resultados de la predicción. Tras diferentes simulaciones hemos obtenido los mejores resultados tomando periodos de amplitud entre 20 y 30 años.

APLICACIÓN DEL MODELO

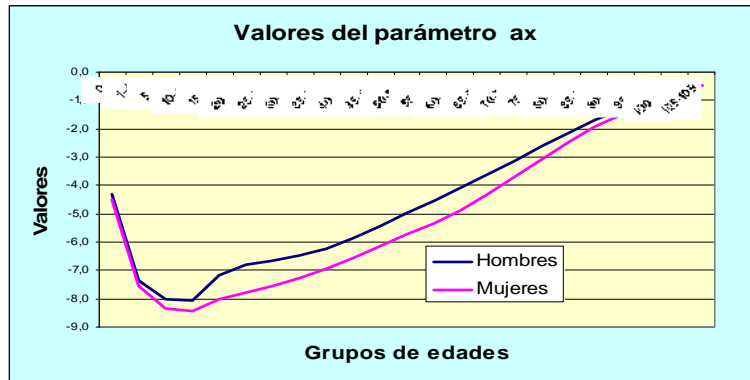
Por todo lo anterior, hemos elegido para nuestra predicción final el intervalo de datos de calendario [1960-2005]. Con la elección de la fecha inicial salvamos las pequeñas irregularidades de los años 50 y la fecha final porque corresponde a los datos más recientes que hemos podido preparar. La amplitud del periodo es suficiente para lo que pretendemos en este trabajo y para la extensión de la predicción a realizar. En cuanto al intervalo de edades, en principio elegimos todo el horizonte de la vida de una persona, aun habiendo observado que para los primeros años el modelo no ajusta adecuadamente los datos. El horizonte temporal de la predicción la extendemos al periodo [2006-2030]. Creemos que es un periodo aceptable según los años de datos que hemos preparado. El elemento biométrico de la predicción es la vida media, como ya quedó justificado. En principio, en el origen, aunque también analizaremos la vida media a los 65 años por la importancia que tiene en la sociedad este resultado.

Una vez preparado el escenario de datos procedemos a la aplicación del modelo en el sentido que hemos dejado de manifiesto en los apartados anteriores.

1) Estimación del parámetro a_x .

Los resultados obtenidos se presentan en forma gráfica a continuación:

Gráfico N° 6. Valores correspondientes al parámetro a_x correspondiente a la predicción del modelo para todo el horizonte de edades de una persona y para el periodo 1960-2005. Tanto para el colectivo de hombres como de mujeres.



Se puede observar que el parámetro a_x captura la incidencia, en el colectivo de hombres, del suceso de fallecimientos por accidentes de tráfico en las edades de los 20 años y superiores. En el colectivo de mujeres la evolución es más suave. El valor de este parámetro permanece constante para todo el periodo de tiempo de calendario contemplado en la predicción.

Los valores correspondientes al parámetro a_x se presentan en la siguiente tabla.

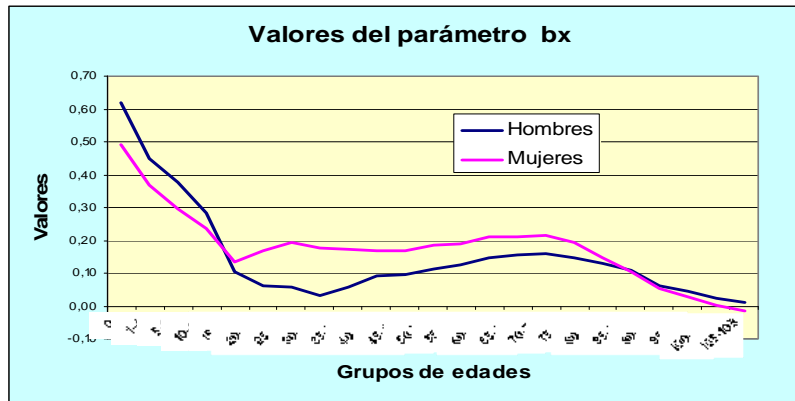
Tabla N° 1

Valores de $a(x)$			
Hombres		Mujeres	
-4,293038	-4,978056	-4,518852	-5,739589
-7,349885	-4,529899	-7,537912	-5,331954
-8,017493	-4,068892	-8,355775	-4,855585
-8,044666	-3,608383	-8,430866	-4,331754
-7,156149	-3,124289	-8,003833	-3,725142
-6,782076	-2,637825	-7,780618	-3,096757
-6,655538	-2,161135	-7,546964	-2,492422
-6,471937	-1,710355	-7,267660	-1,939440
-6,218509	-1,317933	-6,945073	-1,479284
-5,858065	-0,966207	-6,562024	-1,060384
-5,427975	-0,680625	-6,149333	-0,722150
	-0,458783		-0,465202

II) Estimación del parámetro b_x .

Los valores resultantes, que se presentan gráficamente, fueron los siguientes:

Gráfico N° 7. Representación gráfica del comportamiento del parámetro b_x , correspondiente a la predicción del modelo para todo el horizonte de edades de una persona y para el periodo 1960-2005. Tanto para el colectivo de hombres como de mujeres.



Los valores del parámetro b_x muestra una fuerte reducción para las edades [15-30] años, en el colectivo de hombres, consecuencia de la “loma” que presentan los datos para algunos años de calendario. Para los últimos años los resultados se deben tomar con cierta precaución dado que los mismos son menos fiables ya en el origen, al tomar los datos históricos.

Los valores correspondientes al parámetro b_x se presentan en la siguiente tabla.

Tabla N° 2

Valores de $b(x)$			
Hombres		Mujeres	
0,617296	0,112942	0,490089	0,184932
0,447926	0,126872	0,367629	0,190787
0,377624	0,148336	0,296503	0,209326
0,282174	0,153620	0,236774	0,212575
0,104789	0,158623	0,132368	0,216507
0,062039	0,147571	0,170193	0,191739
0,056211	0,129577	0,194700	0,147708
0,033682	0,107152	0,176505	0,104196
0,056688	0,061345	0,173893	0,051439
0,090697	0,044996	0,166308	0,026460
0,096926	0,024245	0,166710	0,001447
	0,009316		-0,013441

III) Estimación del parámetro k_1 .

Los valores correspondientes al parámetro $k_1(t)$ siguen una evolución bastante regular y decreciente como corresponde a la predicción que buscamos. No obstante como ya hemos podido comprobar, a lo largo de este trabajo, el valor del parámetro $k(t)$ puede ser mejorado utilizando los valores del parámetro $k_2(t)$. Para proceder a esta iteración se puede desarrollar una VBA macro.

Los valores correspondientes al parámetro $k_2(t)$ se presentan en la siguiente tabla.

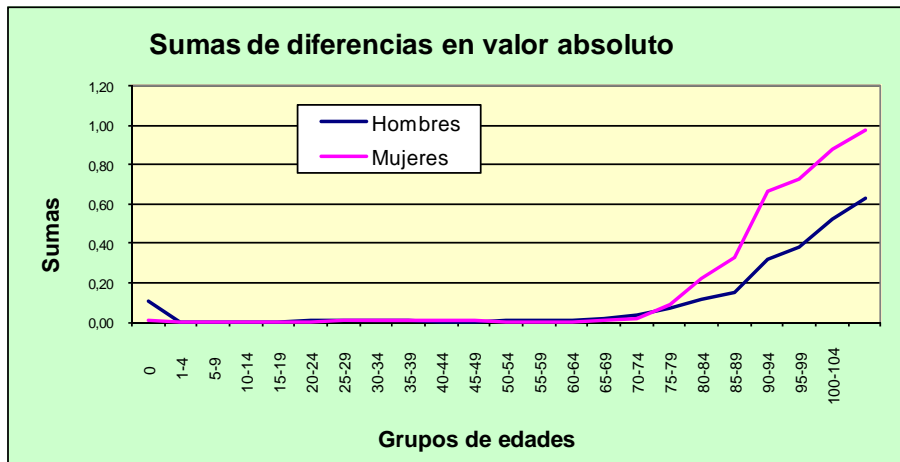
Tabla N° 3

k2(t)							
Hombres				Mujeres			
Año	Valor	Año	Valor	Año	Valor	Año	Valor
1960	2,013	2006	-2,527	1960	2,539	2006	-2,827
1961	1,773	2007	-2,623	1961	2,317	2007	-2,923
1962	1,956	2008	-2,720	1962	2,453	2008	-3,020
1963	1,942	2009	-2,816	1963	2,400	2009	-3,116
1964	1,708	2010	-2,913	1964	2,076	2010	-3,213
1965	1,570	2011	-3,010	1965	1,967	2011	-3,310
1966	1,495	2012	-3,106	1966	1,875	2012	-3,406
1967	1,480	2013	-3,203	1967	1,815	2013	-3,503
1968	1,420	2014	-3,299	1968	1,765	2014	-3,599
1969	1,710	2015	-3,396	1969	2,015	2015	-3,696
1970	1,234	2016	-3,492	1970	1,609	2016	-3,792
1971	1,609	2017	-3,589	1971	1,833	2017	-3,889
1972	0,990	2018	-3,686	1972	1,339	2018	-3,986
1973	1,200	2019	-3,782	1973	1,515	2019	-4,082
1974	1,072	2020	-3,879	1974	1,360	2020	-4,179
1975	0,980	2021	-3,975	1975	1,180	2021	-4,275
1976	0,861	2022	-4,072	1976	1,043	2022	-4,372
1977	0,633	2023	-4,169	1977	0,799	2023	-4,469
1978	0,557	2024	-4,265	1978	0,681	2024	-4,565

1979	0,325	2025	-4,362	1979	0,380	2025	-4,662
1980	0,095	2026	-4,458	1980	0,170	2026	-4,758
1981	0,024	2027	-4,555	1981	0,064	2027	-4,855
1982	-0,312	2028	-4,652	1982	-0,274	2028	-4,952
1983	-0,147	2029	-4,748	1983	-0,084	2029	-5,048
1984	-0,305	2030	-4,845	1984	-0,305	2030	-5,145
1985	-0,195			1985	-0,310		
1986	-0,450			1986	-0,510		
1987	-0,596			1987	-0,764		
1988	-0,568			1988	-0,774		
1989	-0,583			1989	-0,888		
1990	-0,565			1990	-0,923		
1991	-0,634			1991	-1,040		
1992	-0,913			1992	-1,408		
1993	-0,959			1993	-1,419		
1994	-1,170			1994	-1,662		
1995	-1,168			1995	-1,732		
1996	-1,252			1996	-1,817		
1997	-1,501			1997	-2,021		
1998	-1,466			1998	-1,969		
1999	-1,432			1999	-1,908		
2000	-1,848			2000	-2,305		
2001	-2,042			2001	-2,549		
2002	-2,103			2002	-2,542		
2003	-2,057			2003	-2,344		
2004	-2,475			2004	-2,876		
2005	-2,430			2005	-2,730		

Se aprecia con claridad el tramo de la predicción de la tendencia donde la línea es prácticamente recta. Vemos también cómo la tendencia sigue algo similar a lo que hubiéramos obtenido al aplicar una regresión lineal a los valores de los parámetros $k_2(t)$ durante el periodo histórico.

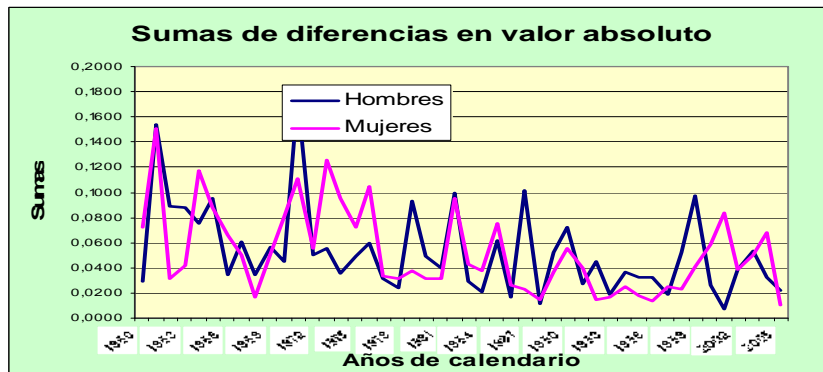
Gráfico N° 8. Representación gráfica de la suma de los valores residuales tomados en valor absoluto para grupos de edades, durante el proceso de predicción. Tanto para hombres como para mujeres aplicando los valores del parámetro $k_2(t)$, por grupos de edades.



Con el objeto de seguir los pasos marcados en la primera parte del trabajo, procedimos a contrastar la bondad del ajuste de los tantos centrales de mortalidad a través de la suma de los valores residuales, en valor absoluto, para el tiempo de calendario [1960-2005], tanto para el colectivo de hombres como de mujeres.

Los resultados obtenidos son más que aceptables entre los 5 y los 75 años y bastante aceptables hasta los 90 años.

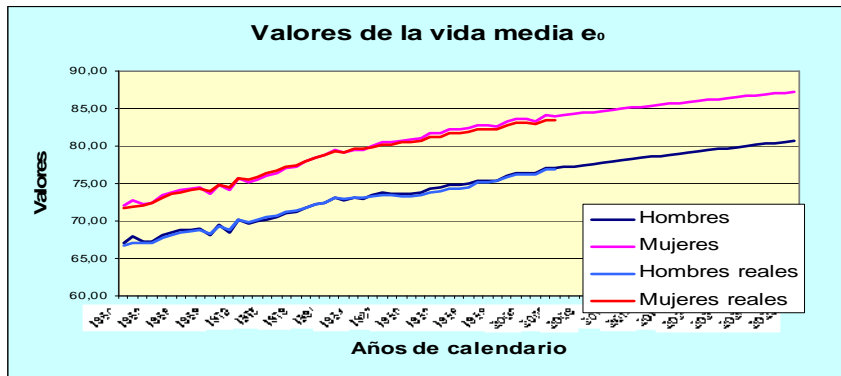
Gráfico N° 9. Representación gráfica de la suma de los valores residuales tomados en valor absoluto para el periodo 1960-2005. Tanto para hombres como para mujeres aplicando los valores del parámetro $k_2(t)$, por años de calendario.



Si mostramos la comparación, por años de calendario, el resultado sigue siendo aceptable, tanto para el colectivo de hombres como el de mujeres. Se observa que el resultado mejora conforme se avanza en el tiempo de calendario.

Si realizamos el contraste a través de la vida media en el origen, en vez de en base a los tantos centrales de mortalidad, lo que por otra parte es el objetivo último de este trabajo, vemos que la tendencia entre los valores reales y los estimados está muy próxima para todo el horizonte histórico [1960,2005]. Se observa, a su vez, que la tendencia, según este modelo (y también se aprecia visualmente) es de mejora continuada pero menos que proporcionalmente. Lo cual significa que se prevé que la mejora continuará pero atenuándose.

Gráfico N° 10. Representación gráfica de los valores de la vida media en el origen de edades, para el periodo 1960-2005 y para el periodo de predicción 2006-2030. Tanto para el colectivo de hombres como para el de mujeres.



Los resultados obtenidos se presentan en la siguiente tabla:

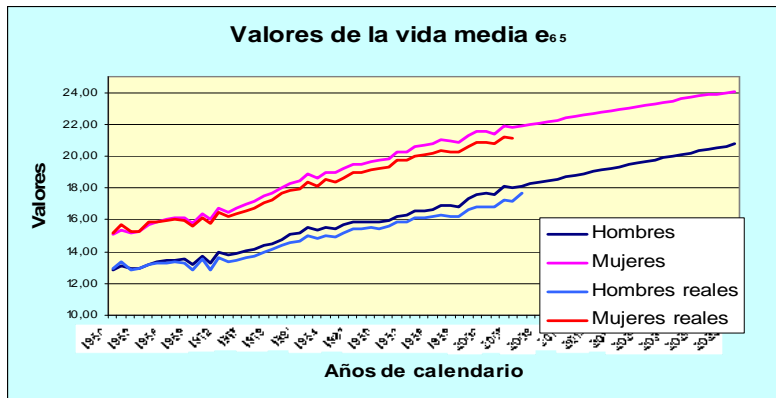
Tabla N° 4

e0							
Hombres				Mujeres			
Años	Valores	Años	Valores	Años	Valores	Años	Valores
1960	67,0347	2006	77,1656	1960	72,0016	2006	84,1071
1961	67,8486	2007	77,3231	1961	72,7280	2007	84,2552
1962	67,2333	2008	77,4796	1962	72,2869	2008	84,4015
1963	67,2815	2009	77,6350	1963	72,4604	2009	84,5462
1964	68,0572	2010	77,7894	1964	73,4844	2010	84,6892
1965	68,4897	2011	77,9428	1965	73,8155	2011	84,8305
1966	68,7177	2012	78,0953	1966	74,0901	2012	84,9702
1967	68,7627	2013	78,2469	1967	74,2668	2013	85,1083
1968	68,9410	2014	78,3975	1968	74,4126	2014	85,2449
1969	68,0508	2015	78,5473	1969	73,6705	2015	85,3798
1970	69,4754	2016	78,6962	1970	74,8597	2016	85,5133
1971	68,3693	2017	78,8442	1971	74,2140	2017	85,6452
1972	70,1390	2018	78,9914	1972	75,6064	2018	85,7756
1973	69,5703	2019	79,1377	1973	75,1234	2019	85,9045
1974	69,9204	2020	79,2832	1974	75,5495	2020	86,0319
1975	70,1653	2021	79,4280	1975	76,0310	2021	86,1579
1976	70,4743	2022	79,5719	1976	76,3884	2022	86,2825
1977	71,0435	2023	79,7151	1977	77,0063	2023	86,4057
1978	71,2271	2024	79,8575	1978	77,2970	2024	86,5274
1979	71,7706	2025	79,9991	1979	78,0158	2025	86,6478
1980	72,2863	2026	80,1400	1980	78,4991	2026	86,7668
1981	72,4413	2027	80,2801	1981	78,7376	2027	86,8845
1982	73,1508	2028	80,4195	1982	79,4749	2028	87,0009
1983	72,8071	2029	80,5582	1983	79,0647	2029	87,1159
1984	73,1364	2030	80,6961	1984	79,5408	2030	87,2296
1985	72,9080			1985	79,5515		
1986	73,4318			1986	79,9699		
1987	73,7232			1987	80,4855		
1988	73,6678			1988	80,5054		
1989	73,6969			1989	80,7304		

1990	73,6618	1990	80,7995
1991	73,7977	1991	81,0265
1992	74,3369	1992	81,7185
1993	74,4240	1993	81,7387
1994	74,8176	1994	82,1777
1995	74,8139	1995	82,3016
1996	74,9672	1996	82,4514
1997	75,4174	1997	82,8022
1998	75,3549	1998	82,7134
1999	75,2940	1999	82,6086
2000	76,0258	2000	83,2770
2001	76,3580	2001	83,6719
2002	76,4613	2002	83,6607
2003	76,3834	2003	83,3409
2004	77,0815	2004	84,1831
2005	77,0069	2005	83,9573

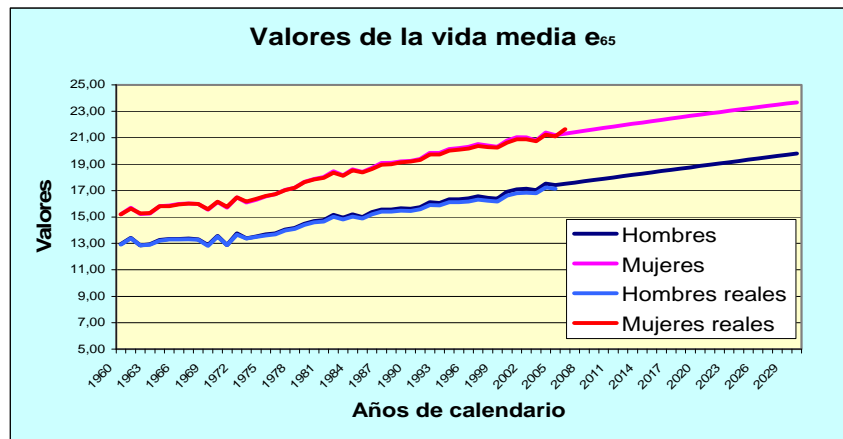
Veamos ahora los resultados cuando nos fijamos en la vida media a los 65 años. En este caso los resultados no son tan satisfactorios. Ello es debido a que para estimar esta magnitud biométrica hemos considerado todo el colectivo de edades de las personas.

Gráfico N° 11. Representación gráfica de los valores de la vida media a los 65 años, para el periodo 1960-2005 y para el periodo de predicción 2006-2030, cuando se toma para los cálculos todo el horizonte de edades de una persona. Tanto para el colectivo de hombres como para el de mujeres.



En efecto si realizamos de nuevo todos los cálculos pero considerando ahora, solamente el colectivo total de personas con edades superiores a los 65 años, los resultados mejoran considerablemente. Los valores residuales son ahora despreciables y los valores de la esperanza matemática de vida resultan muy aceptables. Estos últimos resultados se presentan, de forma gráfica y tanto para el colectivo de hombres como de mujeres, en el siguiente gráfico:

Gráfico N° 12. Representación gráfica de los valores de la vida media a los 65 años, para el periodo 1960-2005 y para el periodo de predicción 2006-2030, cuando se toma para los cálculos al colectivo de personas con edades superiores a los 65 años. Tanto para el colectivo de hombres como para el de mujeres.



Una vez que los valores estimados de los parámetros gozan de nuestra confianza, la obtención de los valores correspondientes a los tantos centrales de mortalidad y por ende los correspondientes a los tantos anuales de mortalidad y demás magnitudes biométricas son muy simples para los actuarios, e incluso para aquellas personas con escasos niveles de conocimientos biométricos.

Los resultados presentados en forma de tabla son los siguientes:

Tabla N° 5

E65							
Hombres				Mujeres			
Años	Valores	Años	Valores	Años	Valores	Años	Valores
1960	12,9502	2006	17,5071	1960	15,2036	2006	21,3014
1961	13,4262	2007	17,6045	1961	15,6982	2007	21,4108
1962	12,8613	2008	17,7017	1962	15,2246	2008	21,5192
1963	12,9365	2009	17,7988	1963	15,2521	2009	21,6267
1964	13,2553	2010	17,8958	1964	15,8180	2010	21,7332
1965	13,3281	2011	17,9927	1965	15,8532	2011	21,8388
1966	13,3347	2012	18,0895	1966	15,9809	2012	21,9435
1967	13,3645	2013	18,1860	1967	16,0287	2013	22,0472
1968	13,3115	2014	18,2825	1968	15,9825	2014	22,1500
1969	12,8711	2015	18,3788	1969	15,5518	2015	22,2518
1970	13,5772	2016	18,4749	1970	16,1385	2016	22,3527
1971	12,8711	2017	18,5709	1971	15,7219	2017	22,4527
1972	13,7509	2018	18,6666	1972	16,4755	2018	22,5517
1973	13,3877	2019	18,7622	1973	16,0924	2019	22,6499
1974	13,5306	2020	18,8577	1974	16,3050	2020	22,7471
1975	13,6873	2021	18,9529	1975	16,5605	2021	22,8434
1976	13,7499	2022	19,0479	1976	16,7227	2022	22,9388
1977	14,0531	2023	19,1428	1977	17,0537	2023	23,0332
1978	14,1779	2024	19,2374	1978	17,2236	2024	23,1268
1979	14,4963	2025	19,3318	1979	17,6576	2025	23,2195
1980	14,6833	2026	19,4260	1980	17,8840	2026	23,3113
1981	14,7668	2027	19,5200	1981	18,0158	2027	23,4022
1982	15,1580	2028	19,6137	1982	18,4568	2028	23,4922
1983	14,9460	2029	19,7073	1983	18,1559	2029	23,5813
1984	15,1991	2030	19,8005	1984	18,6057	2030	23,6695
1985	15,0075			1985	18,4116		
1986	15,3670			1986	18,7131		
1987	15,5729			1987	19,0887		
1988	15,5661			1988	19,0944		
1989	15,6588			1989	19,2189		
1990	15,6141			1990	19,2527		
1991	15,7481			1991	19,3984		

1992	16,1125	1992	19,8529
1993	16,0712	1993	19,8475
1994	16,3326	1994	20,1437
1995	16,3360	1995	20,2130
1996	16,4048	1996	20,3083
1997	16,5698	1997	20,5049
1998	16,4392	1998	20,3951
1999	16,3979	1999	20,3136
2000	16,9097	2000	20,7809
2001	17,0914	2001	21,0516
2002	17,1360	2002	21,0240
2003	17,0400	2003	20,7912
2004	17,5326	2004	21,3853
2005	17,4097	2005	21,1911

CONCLUSIONES

A la vista de los gráficos presentados anteriormente se observa que el modelo proporciona unos buenos resultados de la predicción de la esperanza matemática de vida al nacimiento, cuando se elige adecuadamente el colectivo que formará parte de la base de datos a partir de la cual se va a realizar la predicción de la mortalidad hacia el futuro, según el modelo Lee-Carter.

Se observa también que una regresión lineal sería una buena estimación hacia el futuro. No obstante el modelo predice que la evolución de la vida media será creciente pero a un ritmo ligeramente inferior al proporcional.

El modelo no proporciona los mismos resultados en cuanto a la esperanza matemática de vida a partir de los 65 años, cuando se toma en consideración todo el grupo de edades desde la edad inicial hasta la edad límite superior biométrico.

Con el objeto de mejorar la predicción de este colectivo de personas hemos creado otra base de datos incluyendo en el modelo, para los mismos periodos de tiempo de calendario, información correspondiente a las personas con edades iguales o superiores a los 65 años. Como ya quedó indicado, esto afecta claramente a los valores de los parámetros. Pero procediendo mediante la misma metodología que para el colectivo total, hemos obtenido los nuevos valores de los parámetros. Evidentemente los valores residuales

son ahora muy inferiores a cuando tomamos todo el colectivo de edades y en cuanto a la predicción de la esperanza matemática de vida los resultados que hemos obtenido son muy satisfactorios.

En cualquier caso, nosotros creemos que para obtener buenos resultados con este modelo, se requiere estudiar con detalle, previamente, la información disponible e incluso realizar una serie de simulaciones previas sobre el comportamiento del modelo. Es importante elegir con acierto el intervalo de tiempo de calendario histórico. Un estudio en profundidad nos llevaría a una aplicación de tipo de programación matemática, pero pensamos que tampoco es necesario tal nivel de precisión teniendo en cuenta el escenario en el que se van a aplicar los resultados. También es conveniente tener en cuenta la amplitud del intervalo de edades a incluir en el modelo. Cuando se trata de un colectivo de personas y lo que se trata es de predecir la mortalidad futura, es preciso limitarse al conjunto de edades para los que se va aplicar el producto de seguro. Pues como hemos visto se gana de forma importante en la calidad de la predicción.

Creemos que una de las fortalezas del modelo Lee-Carter es que una vez que hemos estimado los valores de los parámetros (vectores) \hat{a}_x , \hat{b}_x y \hat{k}_t para su predicción, es suficiente con analizar la evolución del índice \hat{k}_t . En sus orígenes, Lee-Carter utilizaron para su predicción un modelo de series de tiempo univariante estandar como es el ARIMA(0,1,0) ⁽¹²⁾.

REFERENCIAS BIBLIOGRAFICAS

BETZUEN, Amancio (1999). “Una predicción de los tantos de mortalidad general” ANALES. Tercera época. Nº 5. Madrid.

BETZUEN, Amancio (2000). “Una estimación de la tendencia de la mortalidad abreviada futura a través de la evolución de los parámetros”. V Congreso Nacional y III Hispano-Italiano. Universidad del país Vasco. Tomo I. Bilbao.

BETZUEN, Amancio (2008). “Un analisis de la prediccion de la mortalidad aplicando el modelo Lee-Carter “. I Congreso Ibérico de Actuarios. Lisboa.

BETZUEN, Amancio (2009). “La evolución y prediccion de la mortalidad en la Peninsula Ibérica según el modelo Lee-Carter”. II Congreso Ibérico de Actuarios. Bilbao.

¹² Para otro conjunto de valores, otro tipo de modelo ARIMA resultaría más adecuado que el anterior. De ahí la necesidad del análisis.

CARTER, L.R. and Prskawetz, A. (2001) “Examining Structural Shift in Mortality Using the Lee-Carter Method”.

LEE, R.D., CARTER, L.R. (1992) “Modelling and Forecasting U.S. Mortality”. Journal of the American Statistical Association 87. 659-671.

LEE, R.D. and MILLER Timothy. (2000) “Evaluating the Performance of Lee-Carter Mortality Forecasts”.

RENSHAW A. and HABERMAN S. (2003) “Lee-Carter Mortality Forecasting with Age Specific Enhancement”. Insurance: Mathematics & Economics. Vol. 33.

WILMOTH J (1993). “Computational methods for filling and extrapolating the Lee-Carter model of mortality change”. Department of Demography. University of California. Berkeley.

WOHLFART, Peter (2006). “Mortality Predictions for Longevity Analysis and Annuity Valuations”.