Congreso Ibérico
Lisboa 29-31 mayo 2008

UN ANALISIS DE LA PREDICCION DE LA MORTALIDAD APLICANDO EL MODELO LEE-CARTER

Amancio Betzuen
Universidad Pais Vasco
Amancio.betzuen@ehu.es

UN ANALISIS DE LA PREDICCION DE LA MORTALIDAD APLICANDO EL MODELO LEE-CARTER

Amancio Betzuen
Mayo 2008

Abstract

La estimación de la evolución del coste futuro de las pensiones de la Seguridad Social depende, en gran medida, de la predicción de la mortalidad futura de ciertos colectivos. Existen diferentes modelos de predicción de la mortalidad hacia el futuro. Unos más teóricos y otros más prácticos, pero todos persiguen la estimación de la evolución de un colectivo de la forma más fiable posible. En este trabajo se pretende analizar la fiabilidad del modelo Lee-Carter para la predicción de la mortalidad.

Uno de estos métodos, con un enfoque mas bien práctico es el modelo de Lee-Carter. Dicho modelo parte de la información de un conjunto de datos históricos de la mortalidad para predecir cuáles van a ser los resultados futuros.

El citado modelo coordina la incidencia de dos componentes como son: la edad de la persona (tiempo biométrico) y el tiempo físico o de calendario. El modelo requiere la utilización de una matriz de doble entrada, que se puede tratar por la metodología SVD (Singular Value Decomposition) para obtener los parámetros estimados y por el modelo ARIMA para generar la proyección de los valores en el tiempo.

El modelo se aplica a la población general española separado por sexo, para la serie de datos de 1950-2000 y para el intervalo de edades [0;99]. La predicción se extiende hasta el año 2025. La aplicación se realiza sobre la esperanza matemática de vida, tanto al nacer, como a la edad de 65 años.

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

1. INTRODUCCIÓN

Durante los últimos veinte años existe una motivación especial por predecir cuál va a ser la evolución previsible de la mortalidad futura. Hasta hace un cuarto de siglo los estudios de mortalidad se centraban básicamente en modelos estáticos, que si bien se ajustaban con gran precisión tenían el inconveniente que se quedaban anticuados en pocos años. Por este motivo, se pasó a estimar tablas de mortalidad de las denominadas dinámicas.

Existen investigadores que continuamente están proponiendo nuevos modelos para predecir la mortalidad futura. Algunos de ellos utilizaron modelos de tipo estático, en el que se analiza la variación de algunos de los parámetros más representativos, en el tiempo, para predecir su posible valor en un tiempo futuro.

Tal es el caso que utilizó el autor (Betzuen 2000) para predecir la mortalidad general, diferenciando por sexo, de la población española. En aquella ocasión se utilizó el modelo Heligman and Pollard (1980) para determinar los valores de los parámetros en cada año de calendario. La predicción se lograba al analizar la evolución de los parámetros en el tiempo. Esta labor de cálculo (ajuste) era mas bien largo y tedioso.

En este trabajo el autor pretende analizar las posibilidades del modelo Lee-Carter (1992) para predecir la mortalidad de la población general española, diferenciando por sexo.

La importancia de este tipo de análisis (investigación) creemos que está más que justificada. Es evidente que los costes de financiación de las pensiones de Seguridad Social (entre otros), dependen directamente de la evolución del colectivo del que depende y en su evolución, uno de los factores más influyentes es la mortalidad. De ahí el interés por tantos investigadores por encontrar un modelo que muestre de la manera más fiable posible cuál podría ser su posible evolución.

2. DATOS PARA EL ANÁLISIS

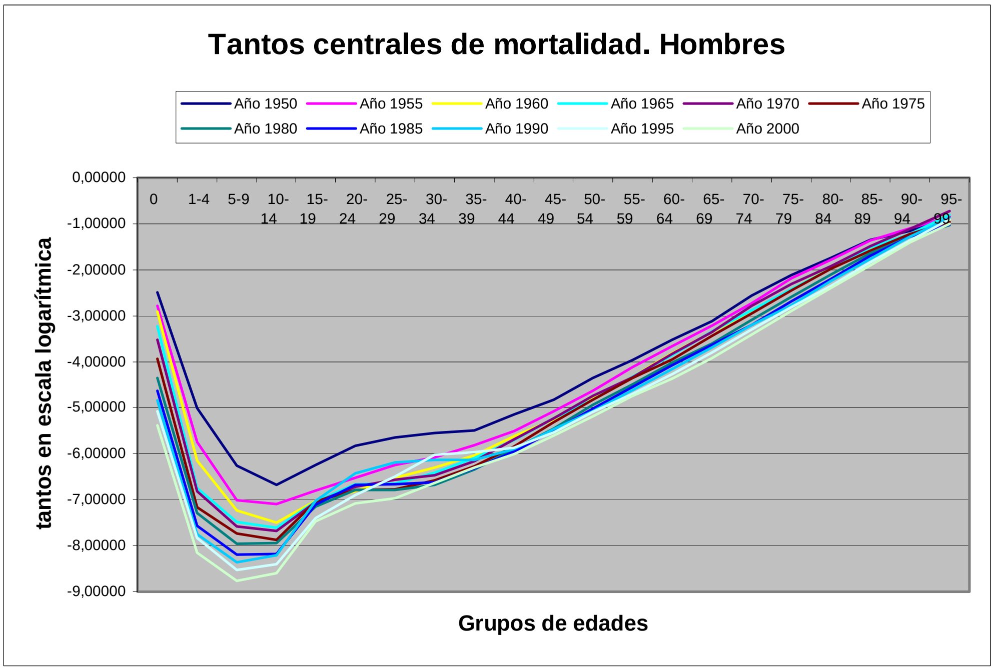
Los datos para el estudio fueron obtenidos del INE (Instituto Nacional de Estadística). Se han utilizado, en particular, valores del número de fallecimientos por edad biométrica simple y del número (valor) de los expuestos al riesgo, también por edad biométrica simple. Posteriormente y con el objeto de reducir las dimensiones de la matriz de datos se agruparon los datos por grupos de clases con edades: 0, 1-4, 5-9, 10-14,.......95-99. Un análisis de notable importancia lo constituye el seguimiento del grupo de edades a partir de los 65 años.

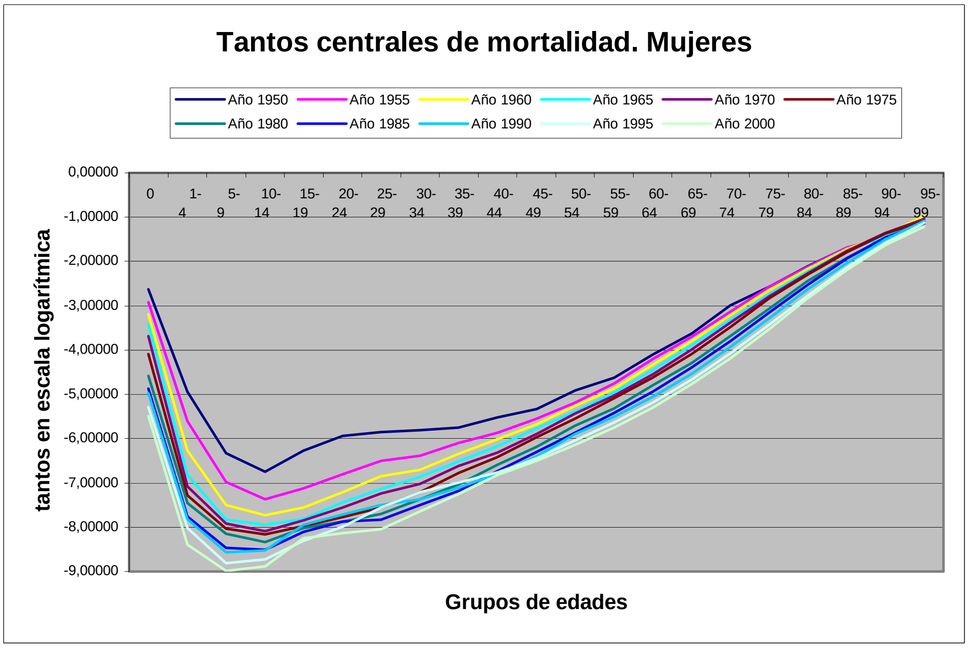
En cuanto al periodo de años de calendario se utilizaron inicialmente los periodos: 1925-2000, 1950-2000 y 1975-2000. Posteriormente seleccionamos el periodo 1950-2000 porque proporcionaba los resultados mas racionales.

El periodo 1925-2000 proporciona mayores altibajos en los valores de los parámetros y la estimación del parámetro k(t) del modelo de Lee-Carter no resultaba tan uniforme como requiere el modelo, mientras que el periodo 1975-2000 nos pareció escaso (corto) para tomarlo como una referencia histórica válida.

El horizonte temporal de predicción futura lo elegimos del 2000-2025. Entendemos que periodos más largos suponen un excesivo atrevimiento en cuanto a la predicción. Nos conformamos con lograr una buena predicción a veinticinco años. Una vez fijado este horizonte, volvimos la vista hacia atrás y fue la puntilla para decidir definitivamente como periodo histórico más adecuado el comprendido entre los años 1950 y 2000. Entendemos que el periodo de datos históricos debe ser superior al de predicción y creemos que una relación del doble es acertada.

Los datos, en forma de tantos centrales de mortalidad que se utilizan en este trabajo dan lugar a las siguientes figuras:





3. EL MODELO LEE-CARTER

Se trata de un modelo que permite predecir la mortalidad proyectando los valores de los tantos centrales de mortalidad a través de una serie de tiempos. Combina, por tanto, valores de la mortalidad biométrica con su evolución a través del tiempo. Los tantos centrales proporcionan la incidencia de la edad biométrica y sus proyecciones la incidencia del tiempo de calendario.

Para ello, construimos los tantos centrales de mortalidad de la forma:

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}} \qquad \forall \text{ x, t} \quad (\text{x, t números enteros})$$

siendo:

$m_{x,t}$: Tanto central de mortalidad para cada edad x y para cada tiempo t.

$D_{x,t}$: Número de fallecidos a la edad x y tiempo t.

$E_{x,t}$: Expuestos al riesgo a la edad x y tiempo t.

Si bien, el modelo Lee-Carter está desarrollado para tratar los tantos centrales $m_{x,t}$ nosotros hemos optado, como es habitual en muchos casos prácticos, por transformarlos en sus logaritmos naturales. De esta manera el modelo de Lee-Carter (1992) se puede expresar de la forma:

$$\ln(m_{x,t}) = \alpha_x + \beta_x * k_t + \varepsilon_{x,t}$$

siendo:

$m_{x,t}$: El tanto central de mortalidad para la edad x y para el año t.

α_x : Un parámetro constante (independiente del tiempo) para cada edad x. Indica el nivel de la mortalidad para cada edad específica o para cada grupo de edades.

β_x : Un parámetro que representa la medida de la variación de la mortalidad (cómo de rápido varía la mortalidad) a cada edad específica o para cada grupo de edades.

k_t : Un parámetro que varía con el tiempo.

$\varepsilon_{x,t}$: Representa el valor residual a la edad x y tiempo t.

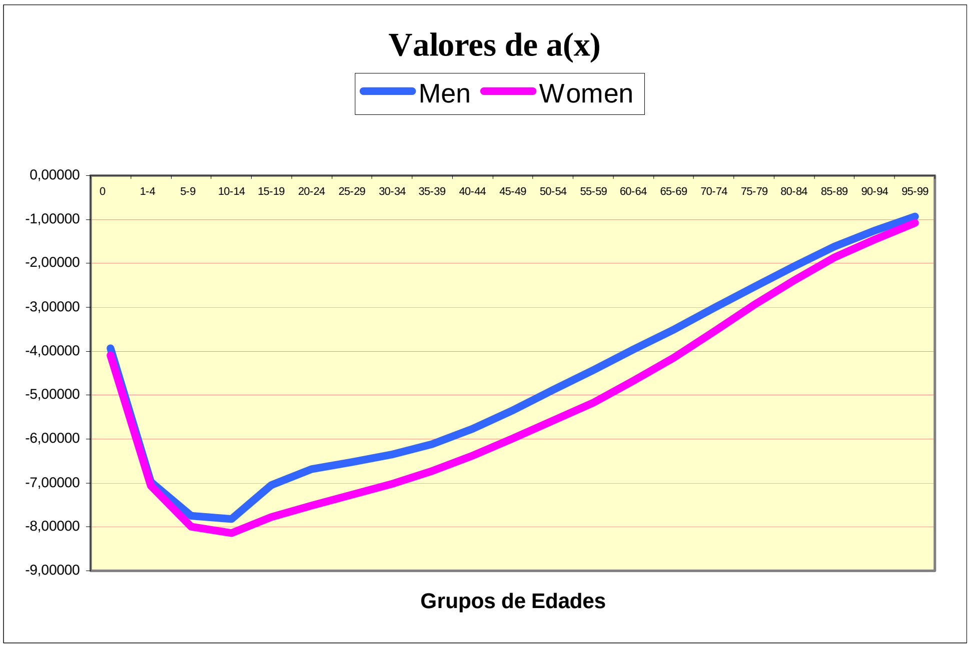
Como se puede observar se trata de un modelo que combina un método paramétrico de estimación de parámetros con un método estadístico de análisis de series de tiempo.

4. METODOLOGÍA

Es evidente que el modelo Lee-Carter no lo podemos tratar por un método de regresión ordinario. Por lo tanto, siguiendo a Lee-Carter (1992) el parámetro α_x lo estimamos como una media geométrica, para todo el intervalo [1950-2000], para cada clase de edades:

$$\alpha_x = \frac{1}{T}\sum_{t_1}^{t_n} \ln(m_{x,t}) = \ln\left[\prod_{t_1}^{t_n} m_{x,t}^{1/T}\right]$$

Los valores de α_x , $\forall$x , muestran el comportamiento de la mortalidad, para cualquier año de calendario, en función de la edad. En nuestro caso se obtiene un gráfico como el que se presenta a continuación:



Calculamos los valores de β_x . Para ello construimos una matriz del tipo:

$$Z_{x,t} = \sum_{x,t} \left[\ln\left(m_{x,t} - \hat{\alpha}_x \right) \right]$$

Para obtener los valores estimados de los parámetros β_x y k_t aplicamos a la matriz $Z_{x,t}$ el método SVD (Singular Value Decomposition). En el modelo introducimos las condiciones:

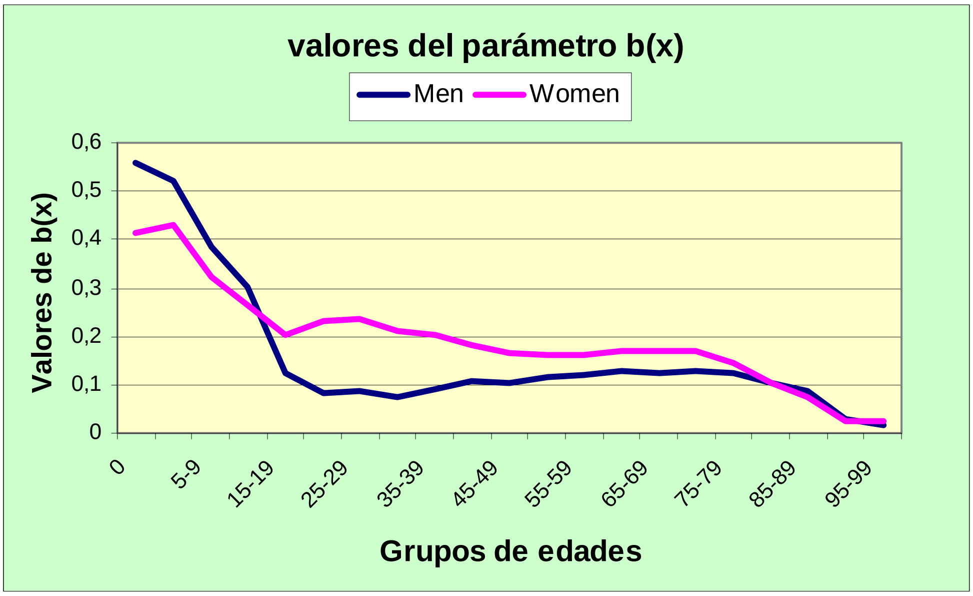
$$\sum_x \beta_x^2 = 1 \quad , \quad \sum_t k_t = 0$$

con el objeto de obtener una única solución para el sistema de ecuaciones del modelo.

Los valores que se obtienen son:

Valores de b(x)			
Men		Women	
0,558666	0,103808	0,412067	0,164578
0,520662	0,115550	0,429216	0,162974
0,385987	0,119431	0,322235	0,159370
0,302814	0,126908	0,263461	0,170619
0,123566	0,122444	0,200753	0,170008
0,083897	0,127739	0,233105	0,171440
0,087608	0,122857	0,235467	0,143231
0,076318	0,104896	0,210422	0,105207
0,092455	0,085321	0,202356	0,073054
0,106310	0,030962	0,181244	0,023887
	0,017912		0,025037

La representación gráfica correspondiente al parámetro β_x se presenta a continuación:



Tal y como se indica en el trabajo original de Lee-Carter (1992), el modelo

$$\hat{\alpha}_x + \hat{\beta}_x * \hat{k}_t = \ln\left(\hat{m}_{x,t}\right)$$

no garantiza que el número de fallecidos, con los parámetros estimados se aproxime al número de fallecidos reales. En consecuencia es necesario realizar un segundo ajuste con el objeto de garantizar el que el número de fallecidos reales y los estimados sean iguales. Para ello introducimos la condición:

$$D_t = \sum_x \left[\exp(\alpha_x + \beta_x * k_t)\mathrm{E}_{\mathrm{x,t}}\right] \quad \forall\, t$$

siendo:

D_t : Número total de fallecimientos en el año t.

$E_{x,t}$: Número de expuestos al riesgo de edad x (clase) en el año t.

Reestimando los valores de k_t, sujeto a esta restricción, se cumple con los requisitos establecidos. Para ello aplicamos un método iterativo que consiste en:

a) Comparar el número total esperado de fallecimientos con el número total real de fallecidos.

$$\sum_x \exp\left(\hat{\alpha}_x + \hat{\beta}_x * \hat{k}_t\right)\mathrm{E}_{\mathrm{x,t}} \ <,=,> \mathrm{D_t}$$

b) En los casos en los que el signo es >, tenemos que disminuir el resultado del número esperado total de fallecimientos, de manera que $\hat{k}_t^{(1)}$ deberá disminuir. Por ejemplo, deberá tomar el valor

$$\hat{k}_t^{(2)} = \hat{k}_t^{(1)}(1 \pm h) \qquad \text{según que} \qquad \hat{k}_t^{(1)} <> 0$$

Siendo h un número tan pequeño como nosotros queramos.

c) La iteración se detiene cuando consideramos que los valores esperados y los reales son suficientemente próximos.

d) En los casos en los que el signo es <, el procedimiento es al contrario.

Para proceder a esta iteración se puede desarrollar un VBA macro.

No obstante, como en todo modelo estimado, existen unas diferencias, entre los valores reales y las estimaciones que lo asociamos a los valores residuales.

$$\varepsilon_{x,t} = \ln\left(m_{x,t} - \hat{\alpha}_x - \hat{\beta}_x * \hat{k}_t^{(2)}\right)$$

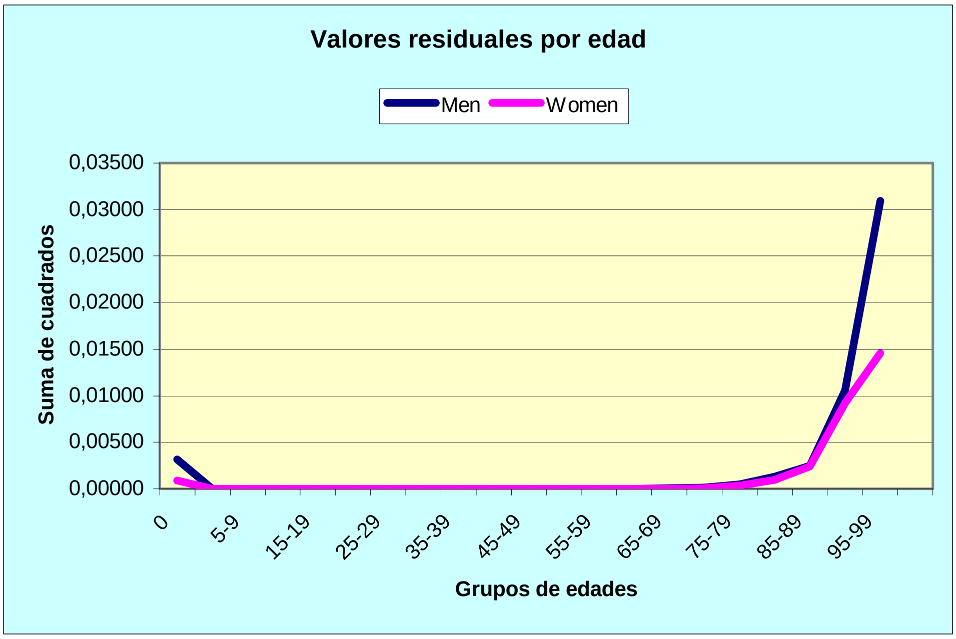
La suma de los cuadrados de los valores residuales, para cada edad, se presentan a continuación:

Valores residuales al cuadrado

Por grupos de edades

Men		Women	
0,00315	0,00000	0,00088	0,00000
0,00001	0,00001	0,00002	0,00000
0,00000	0,00001	0,00000	0,00000
0,00000	0,00002	0,00000	0,00001
0,00000	0,00007	0,00000	0,00002
0,00000	0,00017	0,00000	0,00005
0,00001	0,00046	0,00000	0,00035
0,00001	0,00129	0,00000	0,00097
0,00001	0,00249	0,00000	0,00239
0,00000	0,01060	0,00000	0,00919
	0,03097		0,01459

Gráficamente:



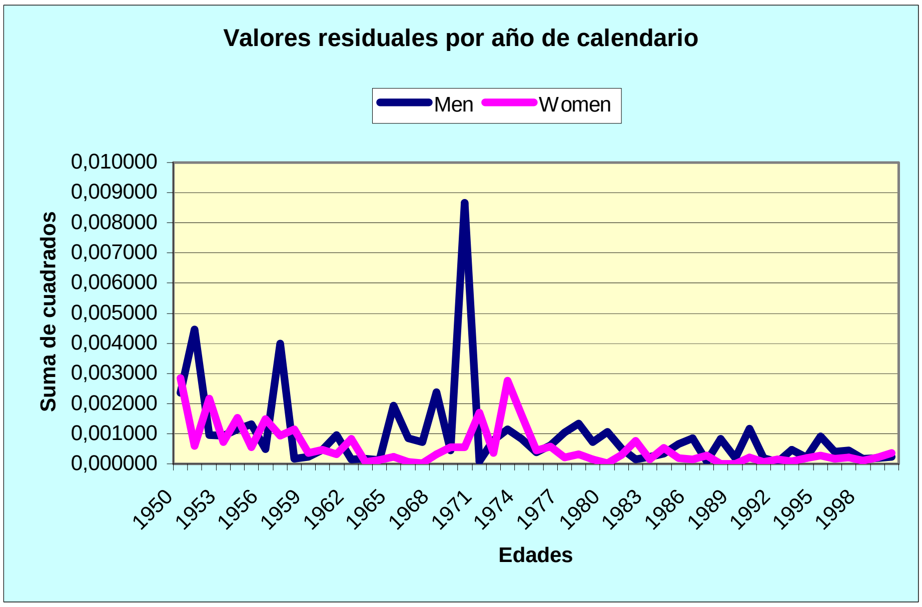
Y la suma de los cuadrados de los valores residuales, para cada año de calendario se presentan a continuación:

Suma de valores residuales al cuadrado

Por años de calendario

Men		Women	
0,002347	0,000387	0,002851	0,000441
0,004450	0,000625	0,000604	0,000564
0,000963	0,001047	0,002155	0,000212
0,000942	0,001338	0,000730	0,000316
0,001138	0,000731	0,001521	0,000154
0,001324	0,001064	0,000557	0,000021
0,000489	0,000509	0,001490	0,000302
0,003993	0,000149	0,000941	0,000763
0,000165	0,000236	0,001156	0,000146
0,000233	0,000344	0,000363	0,000538
0,000473	0,000657	0,000470	0,000192
0,000966	0,000859	0,000316	0,000155
0,000146	0,000085	0,000829	0,000278
0,000180	0,000837	0,000062	0,000006
0,000118	0,000200	0,000126	0,000005
0,001928	0,001168	0,000232	0,000207
0,000839	0,000195	0,000071	0,000067
0,000726	0,000053	0,000029	0,000155
0,002373	0,000463	0,000317	0,000084
0,000442	0,000205	0,000552	0,000201
0,008657	0,000920	0,000555	0,000280
0,000084	0,000403	0,001704	0,000168
0,000792	0,000451	0,000354	0,000208
0,001151	0,000167	0,002761	0,000105
0,000836	0,000199	0,001608	0,000210
	0,000241		0,000354

Gráficamente:



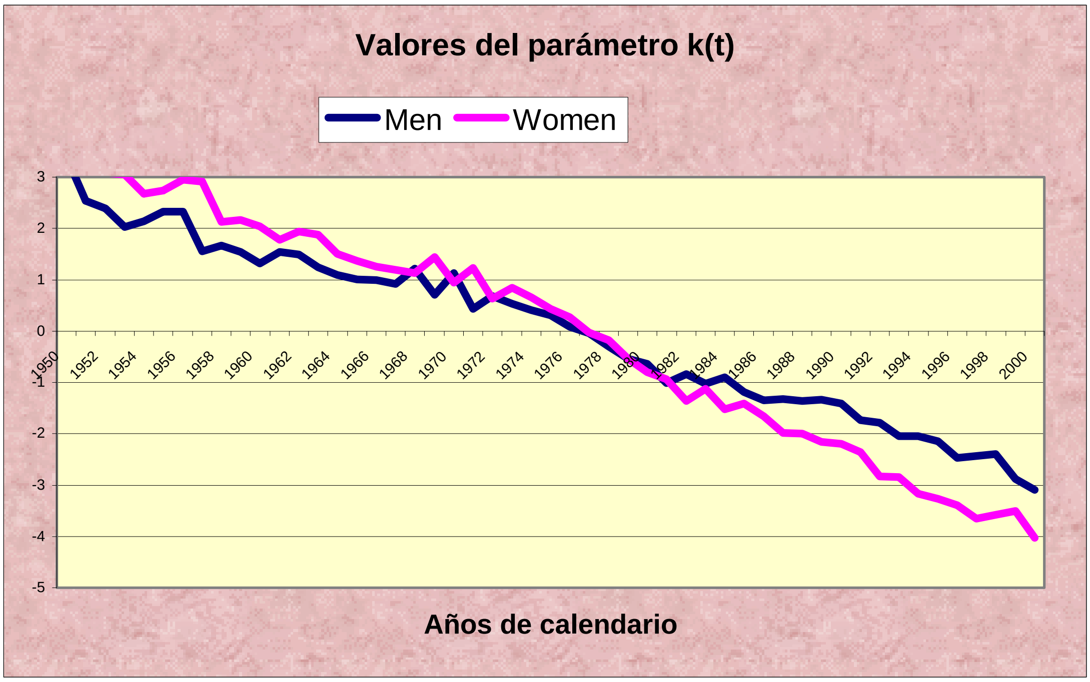
Como se puede observar, los valores residuales no son significativos.

Los valores del parámetro $\hat{k}_t^{(2)}$ resultantes son:

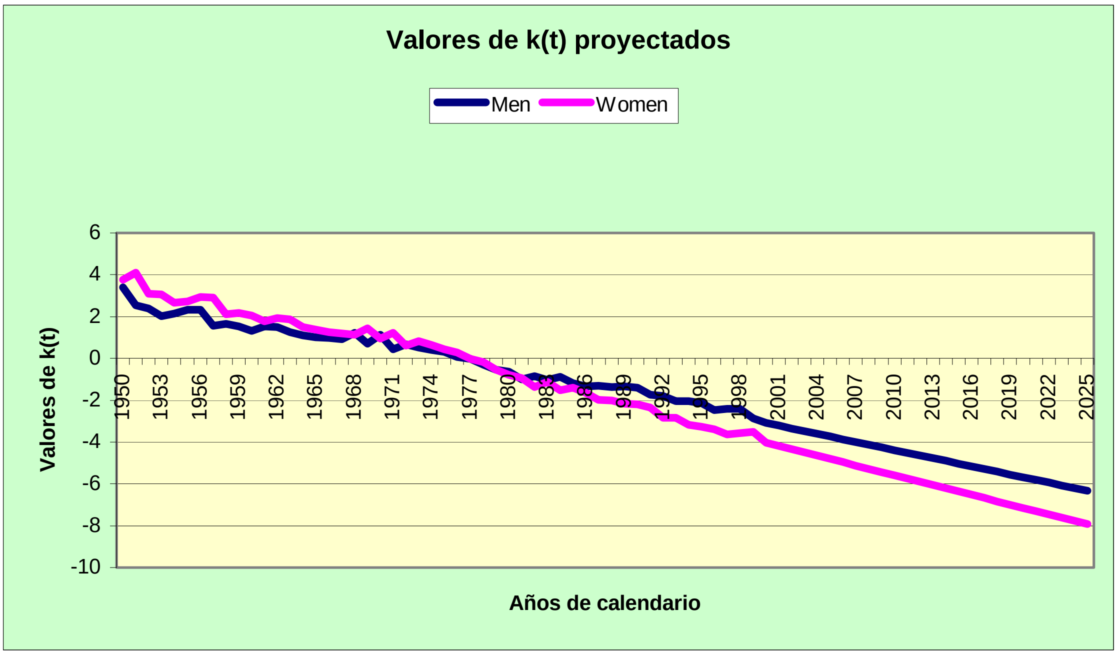
Men		Women	
3,3965	0,306	3,763	0,437
2,5384	0,08	4,108	0,273
2,393	-0,038	3,09	-0,028
2,027	-0,303	3,05	-0,175
2,134	-0,545	2,67	-0,545
2,332	-0,633	2,74	-0,805
2,328	-1,01	2,945	-0,935
1,555	-0,838	2,915	-1,365
1,667	-1,025	2,125	-1,122
1,544	-0,896	2,168	-1,528
1,317	-1,19	2,04	-1,41
1,538	-1,35	1,78	-1,658
1,4947	-1,321	1,945	-1,99
1,247	-1,36	1,88	-2
1,0978	-1,34	1,5	-2,153
1,008	-1,416	1,37	-2,2
0,993	-1,74	1,26	-2,353
0,918	-1,79	1,19	-2,83
1,223	-2,045	1,135	-2,85
0,71	-2,05	1,44	-3,17

1,135	-2,15	0,95	-3,27
0,433	-2,465	1,225	-3,387
0,679	-2,428	0,628	-3,65
0,527	-2,4	0,843	-3,58
0,405	-2,88	0,653	-3,51
	-3,09		-4,03

Representación gráfica



Como se puede observar la evolución del parámetro k(t) es bastante regular y su tendencia decreciente y muy regular. Este comportamiento permite estimar su evolución de forma satisfactoria, como se puede apreciar en el gráfico adjunto, para periodos futuros (a partir del año 2000). Se corresponde con una declinación lineal y una varianza relativamente constante, lo cual facilita las cosas para una predicción futura.

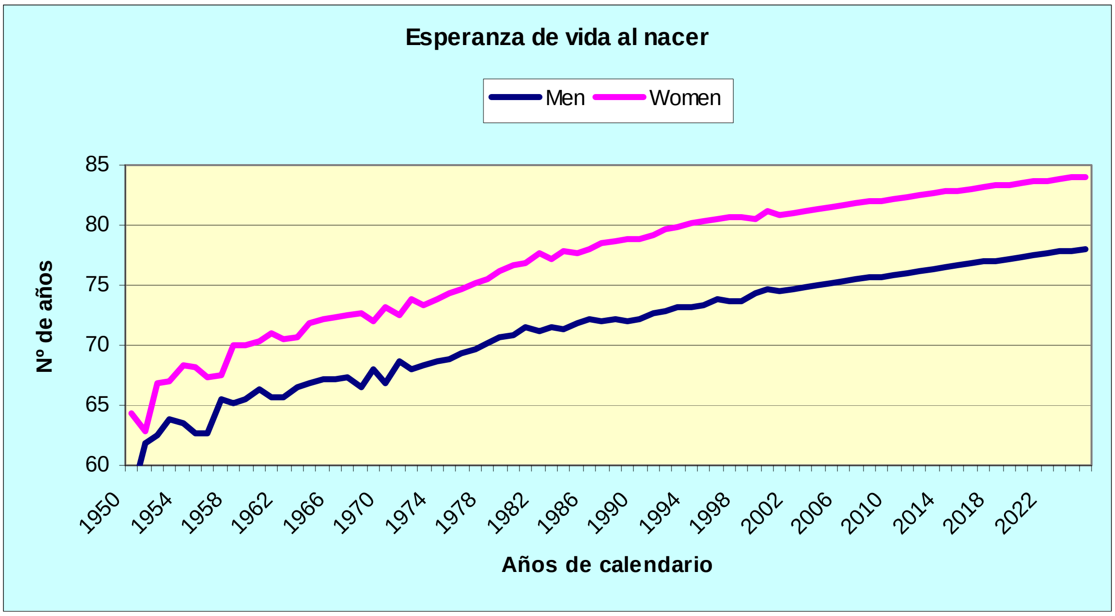


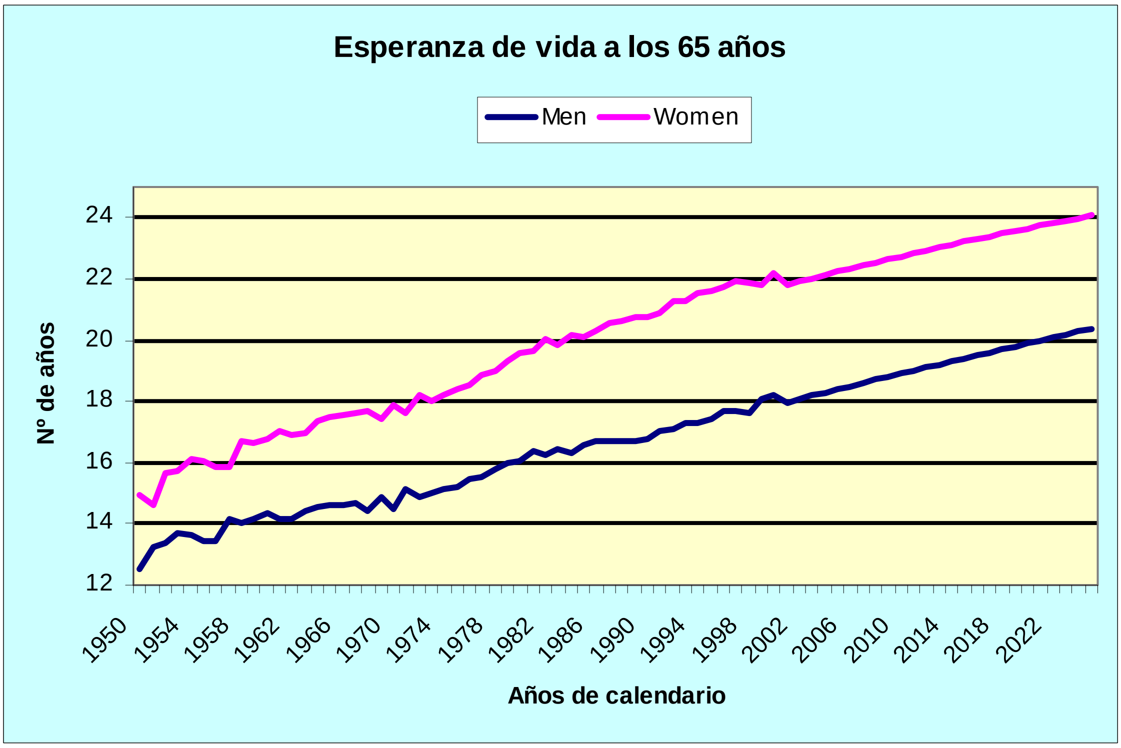
5. PREDICCIÓN DE LAS ESPERANZAS DE VIDA FUTURAS

En estas condiciones hemos procedido a la estimación de las esperanzas matemáticas de vida futuras; a partir del año 2000 y hasta el año 2025 y a partir del momento del nacimiento y a partir de los 65 años. Los resultados obtenidos se presentan en los siguientes cuadros:

	al nacimiento		a los 65 años	
Año	Men	Women	Men	Women
1950	57,3	64,3	12,5	15,0
1955	62,7	68,1	13,4	16,1
1960	66,3	70,3	14,3	16,8
1965	67,2	72,1	14,6	17,5
1970	66,8	73,1	14,5	17,9
1975	68,9	74,3	15,2	18,4
1980	70,8	76,7	16,1	19,6
1985	71,8	77,7	16,6	20,1
1990	72,2	78,9	16,8	20,8
1995	73,3	80,3	17,4	21,6
2000	74,7	81,2	18,2	22,2
2005	75,1	81,5	18,4	22,2
2010	75,9	82,2	18,9	22,7
2015	76,6	82,9	19,4	23,2
2020	77,4	83,5	19,9	23,7
2025	78,0	84,1	20,4	24,1

Las representaciones gráficas correspondientes a estas esperanzas de vida incluyendo el periodo histórico y las estimaciones futuras se presentan a continuación:





6. CONCLUSIONES

Aunque existen diferentes versiones posteriores al modelo Lee-Carter (Wilmoth 1993) nosotros hemos preferido utilizar el inicial, en una primera aproximación. A lo largo del análisis hemos podido comprobar que al menos para nuestros datos, los resultados han sido aceptables.

También cabe la posibilidad de aplicar el método sobre los tantos centrales de mortalidad directamente y no sobre los valores transformados mediante el logaritmo, e incluso sobre los tantos de mortalidad bruta, mas bien que sobre los tantos centrales de mortalidad. Evitando, en algunos casos el recálculo del parámetro k(t).

También se puede aplicar la técnica de los mínimos cuadrados, ponderados o no o la técnica de máxima verosimilitud, pero nosotros hemos preferido la técnica SVD, que nos ofreció resultados aceptables.

Con todo ello hemos obtenido unas estimaciones de las esperanzas de vida hacia el futuro que nos parecen razonables, quizás mas bien prudentes, pero pensamos que en ningún caso se pueden catalogar como infravaloradas.

Finalmente, pensamos que la estimación de los parámetros, por esta metodología, son aceptables como hemos podido comprobar en el cálculo de los valores residuales. Se trata de una metodología que combina los valores biométricos con los valores a través del tiempo de calendario y por consiguiente consigue combinar la incidencia de ambas variables y la construcción de una tabla de mortalidad futura es, inmediata y sencilla. Pues basta con obtener los valores de $m_{x,t}$ con los parámetros estimados y a continuación los de $q_{x,t}$ con lo cual para una l_0=100.000 (raiz de la tabla) se construye una tabla de mortalidad para cualquier fecha futura.

REFERENCIAS

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.

CARTER, L. and PRSKAWETZ, A. (2001). “examining Structural Shifts in Mortality Using the Lee-Carter Methods”. Working Paper. March 2001. Max Planck Institute for Demographic Research. Rostock.

GOOD, I.J. (1969). “Some Applications of the Singular Decomposition of a Matrix”. Tecnometrics, 11.

HELIGMAN, L and POLLARD, J.M. (1980). “The Age Pattern of Mortality”. Journal of the Institute of Actuaries, 107.

KENDALL, Maurice and KEITH, J. (1990). “Time Series”. Edward Arnold.Suffolk.

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. (2000). “The Lee-Carter Method for forecasting mortality, with various extensions and applications”. North American Actuarial Journal,4.

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, J.R. (2005). “Some methodological issues in mortality projection, based on an analysis of the U.S. social security system”. Genus, LXI.