

Proyección de los indicadores de mortalidad para España

Autores

Doña Ana María Debón Aucejo
Don Francisco Martínez Ruiz
Don Francisco Montes Suay
Doña Marta Moshuk

Proyección de los indicadores de mortalidad para España.

Resumen

El presente estudio expone los resultados obtenidos al aplicar el modelo Lee-Carter original y sus extensiones (con dos términos y con efecto cohorte) a las probabilidades de muerte para la población española entre los años 1991 y 2010. Estos modelos se ajustaron a los datos elaborados por el INE separadamente para varones y mujeres. Por otra parte, se calcularon los índices generales de mortalidad y sus proyecciones a futuro por medio de series temporales ARIMA entre los años 2011 y 2030.

A partir de la comparación de los modelos con medidas de bondad de ajuste se constató el modelo de Lee-Carter con efecto cohorte proporciona mejores resultados de la predicción de las probabilidades de muerte que el modelo de Lee-Carter original o el modelo con dos términos. La comparación muestra que, a pesar de la mayor complejidad de los modelos extendidos, éstos pueden explicar de mejor modo las tendencias pasadas y, por tanto, las futuras.

También se calcularon dos indicadores de mortalidad: la esperanza de vida y el índice de Gini. La mejora de ambos refleja de modo directo el incremento en el estándar de vida en la población española durante las últimas décadas.

1. Introducción

El interés en el análisis de mortalidad ha aumentado recientemente como consecuencia del gradual envejecimiento de la población que la mayor parte de las sociedades avanzadas experimentan, entre ellas la española. Este gradual pero constante envejecimiento de población tiene efectos directos sobre aspectos tan importantes como la planificación de sistemas sanitarios y de pensiones. Además, se ha experimentado una mejora en la esperanza de vida al nacer que se debe a la reducción de la mortalidad infantil y, más recientemente, a la reducción de la mortalidad en las edades más elevadas.

El interés de conocer el nivel de mortalidad en la sociedad ha hecho que desde hace décadas se haya intentado medir este fenómeno. Todos los índices utilizados para medir la mortalidad han sido desarrollados con algún propósito y todos ellos han cumplido su objetivo.

Un reciente estudio sobre los indicadores de mortalidad en España de Debón, Montes y Sala (2009) revela que, como en otros países desarrollados, la mortalidad infantil en España ha experimentado una drástica reducción, la mortalidad de personas de mediana edad se ha incrementado en la última década y la mortalidad de los grupos de edad avanzada se ha estabilizado.

Desde principios de los 90 una serie de modelos estocásticos se han desarrollado para analizar estas mejoras de mortalidad, entre ellos el modelo Lee-Carter y sus extensiones. Estos métodos han sido utilizados para predecir las tasas de mortalidad en varios países, incluyendo EEUU (Lee y Carter, 1992), Canadá (Lee y Nault, 1993), Chile (Lee y Rofman, 1994), Australia (Booth, Maindonald y Smith, 2002), Inglaterra y Gales (Renshaw y Haberman, 2003) y España (Debón, Montes y Puig, 2008). Las características más importantes del método de Lee-Carter son su sencillez y su robustez en situaciones donde las tasas logarítmicas de la mortalidad por edades específicas tienen una tendencia lineal. Aunque paulatinamente se fueron desarrollando otros métodos, el método de Lee-Carter es una referencia fija.

En diferentes trabajos se han propuesto modificaciones del método de Lee-Carter, entre ellos Booth, Maindonald y Smith (2002) y Renshaw y Haberman (2006). La exactitud del pronóstico del método de Lee-Carter y sus variantes fueron primero evaluadas por Booth et al. (2005) y además estudiadas por Booth et al. (2006). Además, han habido importantes extensiones del método de Lee-Carter, quedándose con algo de su esencia pero añadiendo rasgos estadísticos adicionales como la suavización no paramétrica o múltiples componentes principales.

En este trabajo se proponen tres modelos distintos para el análisis y predicción de mortalidad en España: Lee-Carter, Lee-Carter extendido propuesto por Booth, Maindonald y Smith (2002) y Lee-Carter con efecto cohorte propuesto por Renshaw y Haberman (2006), utilizando las tablas de mortalidad elaboradas por el INE. La información de los datos está estandarizada e incluye las siguientes variables: probabilidad de muerte y población en riesgo de morir. El análisis se extiende para ambos sexos: hombres y mujeres y el colectivo al que se aplica el estudio es la población general española en un rango de edades de 0 a 125 años, con un periodo temporal desde 1991 hasta 2010. Estos modelos tienen como objetivo final el cálculo de los indicadores de mortalidad como la esperanza de vida e índice de Gini.

El presente documento se estructura de la siguiente manera: en la Sección 2 se presentan la descripción y evaluación del modelo de Lee-Carter, modelo de Lee-Carter con dos términos y con efecto de cohorte para la predicción de la mortalidad. La Sección 3 recoge los resultados de la aplicación del modelo original de Lee-Carter y sus variantes al estudio de la mortalidad española modelando su comportamiento durante el período 1991-2010. En la Sección 4 se presentan las predicciones de las tasas de mortalidad para los años 2011-2030 obtenidos a partir del modelo elegido. La Sección 5 contiene la evolución y la proyección al futuro de la esperanza de vida y el índice de Gini para los datos de la mortalidad española de acuerdo con los modelos escogidos. El artículo finaliza con una sección de las conclusiones que se derivan de los resultados obtenidos.

2. Modelo de Lee-Carter y sus extensiones

2.1. Modelo original de Lee-Carter

El modelo de Lee-Carter se utiliza, preferentemente, para predecir la mortalidad futura dependiendo de la edad y el año de fallecimiento o nacimiento de la persona. Es por tanto un modelo bilineal en las variables edad x y año de calendario t , que supone que la tasa de mortalidad correspondiente a la edad x en el año t verifica,

$$m_{xt} = \exp(a_x + b_x k_t + \epsilon_{xt}) \quad (1)$$

o su equivalente:

$$\ln(m_{xt}) = a_x + b_x k_t + \epsilon_{xt} \quad (2)$$

donde,

m_{xt} : describe el tanto central de mortalidad correspondiente a la edad x y tiempo de calendario t .

a_x : parámetro de "forma", representa una componente que depende de la edad de la persona.

b_x : parámetro de sensibilidad, describe la "velocidad" con la que varía la mortalidad a cada edad x , cuando varía el nivel general de mortalidad.

k_t : índice general de mortalidad, representa un parámetro que depende del tiempo de calendario y refleja el nivel general de declinación de la mortalidad.

ϵ_{xt} : representa el error aleatorio, con media 0 y varianza σ_ϵ^2 , que refleja entre otros factores influencias históricas no capturadas por el modelo.

En Lee (2000) el autor remarca que no se garantiza que las estimaciones obtenidas de m_{xt} a partir de (2) no serán superior a 1, aunque este problema se puede evitar utilizando el logit de las tasas de mortalidad. Por esta razón se transforma el modelo anterior aplicando el logit a las probabilidades de muerte q_{xt} .

$$\text{logit}(q_{xt}) = \ln\left(\frac{q_{xt}}{1 - q_{xt}}\right) = a_x + b_x \cdot k_t + \varepsilon_{xt}. \quad (3)$$

Hay que tener en cuenta además que la estructura del modelo es invariante bajo las siguientes transformaciones: $(a_x, b_x, k_t) \rightarrow (a_x, \frac{b_x}{c}, ck_t)$ o $(a_x, b_x, k_t) \rightarrow (a_x + cb_x, b_x, k_t - c)$, $\forall c$, por lo que no existe la solución única para este sistema. Para evitar este problema los autores (Lee y Carter, 1992) proponen las siguientes restricciones:

$$\sum_x^w b_x = 1, \quad \sum_t^T k_t = 0. \quad (4)$$

Originalmente la estimación de los parámetros del modelo se ha estimado con el método SVD (Singular Value Decomposition). En el presente trabajo la estimación de los parámetros se lleva a cabo mediante el método de máxima verosimilitud utilizando la librería *gmm* (Turner y Firth, 2006) de R.

El paso final del método Lee-Carter es encontrar un modelo apropiado para la tendencia de la mortalidad $\{\hat{k}_t\}$. Utilizando la metodología Box-Jenkins, los autores llegan a la siguiente ecuación:

$$\hat{k}_t = c + \hat{k}_{t-1} + u_t,$$

donde c es una constante y u_t un ruido blanco. Con este modelo, la predicción de \hat{k}_t varía linealmente y cada índice de mortalidad predicho varía con índice exponencial constante. La predicción de las

probabilidades de muerte para edades específicas se obtienen extrapolando k_t y fijando los a_x y b_x estimados.

El método de Lee-Carter es un método paramétrico y poco flexible. El ajuste del modelo de Lee-Carter, estimando k_t mediante la distribución por edad del total de muertes observadas tiene el inconveniente de que no es óptima. Unas de las ventajas a favor del método de Lee-Carter es que se ajusta bien a los datos históricos, es simple en uso y las predicciones a largo plazo son mejores que las obtenidas con otros métodos. Las desventajas del modelo consisten en que los parámetros a_x y b_x dependen sólo de la edad y que la predicción de futuros valores de la mortalidad se basa sólo en k_t , lo que supone admitir que no existe interacción entre la edad y el tiempo.

2.2. Modelo de Lee-Carter con dos términos

Varios autores han propuesto las modificaciones para el modelo de Lee-Carter con el objetivo de mejorar el ajuste del modelo y eliminar la tendencia que muestran los residuos. Por ejemplo: Booth, Maindonald y Smith (2002), Renshaw y Haberman (2003) proponen incluir más términos en la fórmula (3). Dicha sugerencia aplicada al $\text{logit}(q_{xt})$ proporciona el siguiente modelo:

$$\text{logit}(q_{xt}) = a_x + b_x^1 \cdot k_t^1 + b_x^2 \cdot k_t^2 + \varepsilon_{xt}. \quad (5)$$

Siguiendo el mismo desarrollo que para el modelo Lee-Carter, se deben introducir un par de condiciones para obtener una única solución.

$$\sum_x b_x^i = 1 \text{ y } \sum_t k_t^i = 0, \quad i = 1, 2.$$

Los parámetros del modelo extendido son similares a los parámetros del modelo de Lee-Carter explicados anteriormente. El ajuste es posible mediante la extensión del método anterior incorporando el segundo grado de los componentes SVD en la estructura del modelo. En nuestro caso lo hemos estimado por máxima verosimilitud.

2.3. Modelo de Lee-Carter con efecto cohorte

El método de Lee-Carter con efecto cohorte es una variante del método de Lee-Carter con un parámetro adicional. En 2006 Renshaw y Haberman introdujeron uno de los primeros modelos estocásticos para la mortalidad con un efecto de cohorte ($c = t - x$). Con esta transformación el modelo se expresa de siguiente manera:

$$\text{logit}(q_{xt}) = a_x + b_x^1 \cdot k_t + b_x^2 \cdot l_c + \varepsilon_{xt}, \quad (6)$$

donde

a_x : perfil general de la mortalidad a lo largo de la edad.

b_x^1 y b_x^2 : parámetros de sensibilidad que miden las interacciones correspondientes con la edad.

k_t : un efecto período aleatorio.

l_c : un efecto cohorte aleatorio, función del año de nacimiento $c = t - x$.

ε_{xt} : representa el error aleatorio, con media 0 y varianza σ_ε^2 .

Este modelo tiene los mismos problemas de identificación que los modelos anteriores. Por lo tanto se introducen las siguientes condiciones con el objetivo de obtener una única solución:

$$\sum_x b_x^1 = 1, \quad \sum_x b_x^2 = 1 \quad \text{y} \quad l_{t-x} = 0. \quad (7)$$

Como la realización del ajuste es problemática debido a la relación que existe entre los tres efectos principales, los autores recurren a una estrategia de ajuste en dos etapas. En una primera etapa se ajustan los valores a_x según el modelo original de Lee-Carter

$$\hat{a}_x = \frac{\sum_t \text{logit}(q_{xt})}{T}.$$

Los parámetros restantes pueden ser estimados en la segunda etapa sujetos a las restricciones de los parámetros (7). Los valores de partida se obtienen mediante el ajuste de $b_x^1 = b_x^2 = 1$ y el ajuste de k_t y l_c . También en este modelo los parámetros se obtienen mediante máxima verosimilitud.

3. Resultados de los ajustes

En el proceso de análisis preliminar de los datos se han completado los datos brutos hasta edades de 130 años utilizando una adaptación de Denuit y Goderniaux (2005) como describe Debón, Montes y Sala (2009). Para la realización de la adaptación se han usado los casos de edad menores de 85 años. A partir de estos datos reales se han estimado los valores de probabilidad de muerte y población en riesgo para edades hasta 130 años.

El conjunto de datos con el que se calcula cualquier análisis posterior está formado por los valores originales de q_{xt} para $x = \{0, 1, \dots, 85\}$, y por los \hat{q}_{xt} procedentes del ajuste del modelo para $x = \{86, \dots, 125\}$. Una vez se ha conseguido la creación de todos los datos ajustados se procede al siguiente paso de aplicación de los modelos. A continuación se elige el modelo que mejor se ajuste a los datos. Y finalmente con el modelo escogido se proporcionarán las proyecciones de las tasas de mortalidad para los 20 años siguientes, es decir, de 2011 a 2030.

3.1. Estimación de los parámetros del modelo de Lee-Carter

Cabe señalar que el modelo realiza una extrapolación hacia el futuro a partir de la tendencia de la mortalidad en el pasado. Los parámetros a_x y b_x capturan la información de la mortalidad en el pasado para la edad de la persona y el parámetro k_t la evolución de esta mortalidad en el transcurso del tiempo calendario. En total tenemos $272=126+126+20$ parámetros y dos restricciones, con lo que el modelo tiene 270 grados de libertad.

Es necesario señalar que debido a la estructura de los datos, series anuales para un mismo rango de edades, los valores de b_x presentan un patrón irregular, que se suavizó con splines cúbicos para evitar anomalías localizadas en algunos grupos de edad.

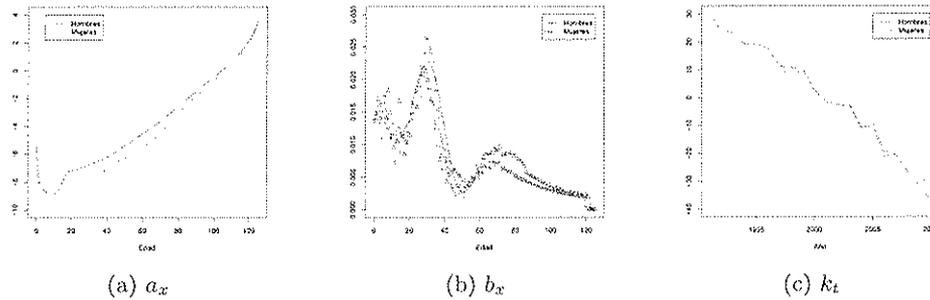
Las estimaciones para hombres y mujeres de los distintos parámetros del modelo Lee-Carter están representadas gráficamente en la Figura 1 permitiendo una interpretación sencilla.

- *Estimación del parámetro a_x*

Las estimaciones para hombres y mujeres del parámetro a_x del caso estudiado se pueden observar en la Figura 1a. La componente a_x del modelo describe el comportamiento general promedio para cualquier año de calendario en función de la edad. Estas estimaciones representan la forma en que la mortalidad se comporta a través de las edades, en este caso los valores para España son similares a los de la mayoría de los países: una mortalidad alta al comienzo de la vida que disminuye rápidamente hasta un mínimo en la edad de los 10 años, aumenta lentamente hasta los 45 años y a partir de aquí crece rápidamente, llegando a superar los niveles de las primeras edades.

Una diferencia clara entre los colectivos de hombres y mujeres puede verse en el intervalo 18 a 30 años, que presenta valores más elevados para el colectivo de hombres. Esta incidencia podría ser

Figura 1: Parámetros estimados con el modelo Lee-Carter



debida a los fallecimientos por accidentes de tráfico y que algunos autores denominan como *joroba de los accidentes* (Debón, Montes y Sala, 2009).

- *Estimación del parámetro b_x*

En la Figura 1b se presentan los resultados correspondientes al parámetro b_x para todo el tramo de edades tanto de hombres como de mujeres. Recordemos que b_x describe la “velocidad” con la que varía la mortalidad a cada edad x y no depende del tiempo. Se observa en la gráfica que el parámetro b_x es positivo para todas las edades, lo que indica una disminución de la mortalidad con el tiempo independientemente de la edad porque los k_t son negativos. Los resultados muestran que en el rango de edad entre 20 y 40 años existe una fluctuación de mortalidad elevada que también aparece, aunque menos acusada en el rango de edad comprendido entre los 60 y 80 años.

- *Estimación del parámetro k_t*

La Figura 1c permite visualizar el comportamiento de las estimaciones del índice temporal de mortalidad para España, que presentan una clara tendencia a disminuir a través del tiempo, más acusadas en los hombres que en las mujeres.

Como paso previo al ajuste de la serie k_t mediante Box-Jenkins, se debe comprobar la bondad del ajuste así como el análisis de los errores.

3.2. Bondad de ajuste del modelo Lee-Carter

Una forma de analizar la bondad de ajuste del modelo es la representación gráfica de los residuos. Las Figuras 2 y 3 muestran el comportamiento de los residuos Deviance para ambos colectivos en función de edad, año de calendario y de la cohorte.

Los resultados obtenidos manifiestan que los residuos se estabilizan en el intervalo $[-2,2]$ lo que supone un buen ajuste del modelo seleccionado, de cualquier modo existen ciertas irregularidades que se deben mencionar. Se observan residuos anómalos para las cohortes con fechas de nacimiento alrededor de 1940 tanto en mujeres como en hombres. Esto puede ser consecuencia de la guerra civil española y el periodo de posguerra, cuando se dió una carencia nutricional en los niños. Este hecho también puede estar influyendo en las anomalías observadas en aquella cohorte con edades entre 55 y 85 años, de cualquier modo es muy posible que existan otros factores para explicar el fenómeno. El ajuste es muy similar para hombres y mujeres.

Figura 2: Residuos Deviance para el modelo de los hombres

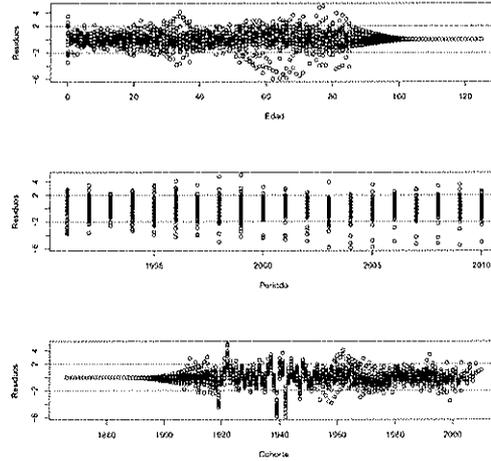
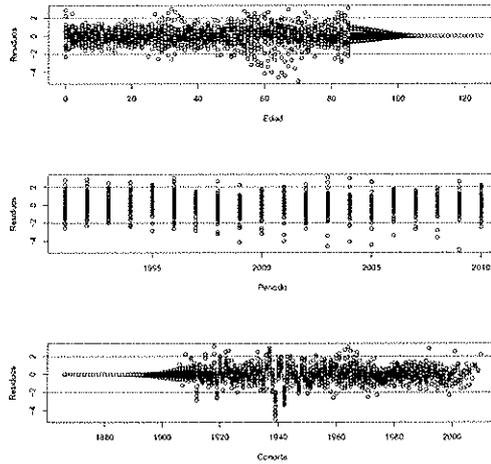


Figura 3: Residuos Deviance para el modelo de las mujeres

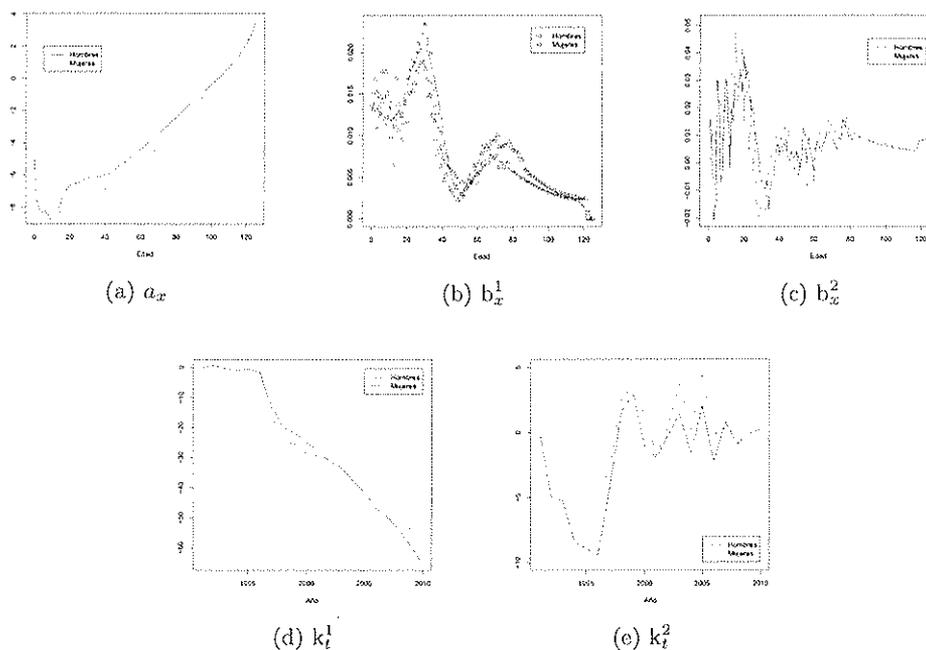


3.3. Resultados del ajuste del modelo de Lee-Carter con dos términos

Este modelo difiere básicamente del modelo anterior en que añade un nuevo término. El número de parámetros es $418=126+126+20+126+20$, con cuatro restricciones por lo que el modelo tiene 414 grados de libertad. La estimación de los parámetros se realiza siguiendo el mismo procedimiento que para el modelo de Lee-Carter original. En un segundo paso se procede a realizar el suavizado de las estimaciones de los parámetros b_x^i para evitar anomalías. Como en el caso anterior se opta por el uso de splines cúbicos para el suavizado.

En este caso las restricciones han sido $k_{t_0}^1 = 0$, $\sum_x b_x^1 = 1$, $k_{t_0}^2 = 0$ y $\sum_x b_x^2 = 1$.

Figura 4: Parámetros estimados con el modelo Lee-Carter con dos términos



▪ *Estimación del parámetro a_x*

En la Figura 4a, correspondiente al parámetro a_x , se muestra el perfil de la mortalidad que con pocas diferencias respecto al mismo parámetro obtenido en el ajuste anterior. Se puede confirmar objetivamente que la mortalidad de las mujeres es inferior a la de los hombres. Por otro lado se observa un aumento sustancial en la mortalidad, especialmente en el colectivo de hombres, en el intervalo de edades de 18 a 30 años. Esta sobremortalidad puede explicarse, como ya se ha dicho anteriormente, por el efecto de los accidentes de tráfico.

▪ *Estimación del parámetro b_x^1*

En la Figura 4b, correspondiente al parámetro b_x^1 , se muestra la velocidad de decrecimiento de los ratios en respuesta a cambios en k_t^1 . Del presente gráfico se deduce que las estimaciones de b_x^1 y su suavizado son muy parecidos a los valores del parámetro b_x del modelo de Lee-Carter original. Se confirma la disminución de la mortalidad con el tiempo independientemente de la edad.

▪ *Estimación del parámetro k_t^1*

La Figura 4d, correspondiente al parámetro k_t^1 , muestra la tendencia de la mortalidad en hombres y mujeres. El gráfico muestra claramente la disminución de la mortalidad durante el periodo observado. Cabe destacar que esta tendencia es más pronunciada para los hombres. Este fenómeno se puede relacionar con multitud de pequeños factores tales como disminución de muertes por accidentes de tráfico, riesgos laborales, mejoras en la asistencia sanitaria, pero requeriría un análisis más detallado para poder elucidar estos parámetros.

▪ *Estimación del parámetro b_x^2*

En la Figura 4c se presentan los resultados correspondientes al segundo término b_x^2 . Este segundo término muestra sus valores más altos en las edades cercanas a los 20 años, lo que parece indicar que el efecto de añadir el segundo término actúa más en este grupo de edad. Presentan valores negativos para algunas edades tanto en hombres como en mujeres.

- *Estimación del parámetro k_t^2*

En la Figura 4e se presentan los resultados correspondientes al segundo término k_t^2 . Los resultados no presentan ninguna tendencia clara. Los valores estimados del parámetro k_t^2 decrecen para primeros cinco-seis años para ambos sexos y después no siguen ninguna pauta.

3.4. Bondad de ajuste del modelo de Lee-Carter con dos términos

En las Figuras 5 y 6 se representa el comportamiento de los residuos Deviance del modelo de Lee-Carter con dos términos para ambos sexos en función de edad, año de calendario y la cohorte. Como en el caso del modelo de Lee-Carter original los resultados obtenidos manifiestan que los residuos se estabilizan en el intervalo $[-2,2]$. También se observa una mejora del ajuste al introducir el segundo término. Aunque los residuos siguen manifestando las mismas irregularidades que el modelo de un término, estas desviaciones son menores en términos absolutos.

Figura 5: Residuos Deviance para el modelo Lee-Carter con dos términos para los hombres

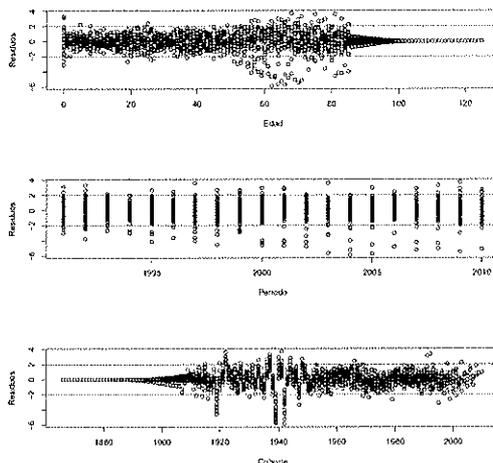
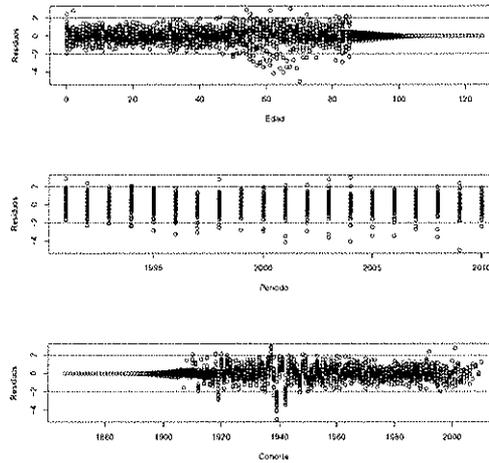


Figura 6: Residuos Deviance para el modelo Lee-Carter con dos términos para las mujeres



3.5. Estimación de los parámetros del modelo de Lee-Carter con efecto cohorte

El número de parámetros de este modelo es $543=126+126+20+126+145$, con tres restricciones por lo que el modelo tiene 540 grados de libertad.

- *Estimación del parámetro a_x*

El comportamiento del parámetro a_x se observa en la Figura 7a. Los resultados presentan un comportamiento muy similar al ya observado en los modelos anteriores. Este gráfico confirma la apreciación del desigual comportamiento en los diferentes grupos de edad para hombres y para mujeres.

- *Estimación del parámetro b_x^1*

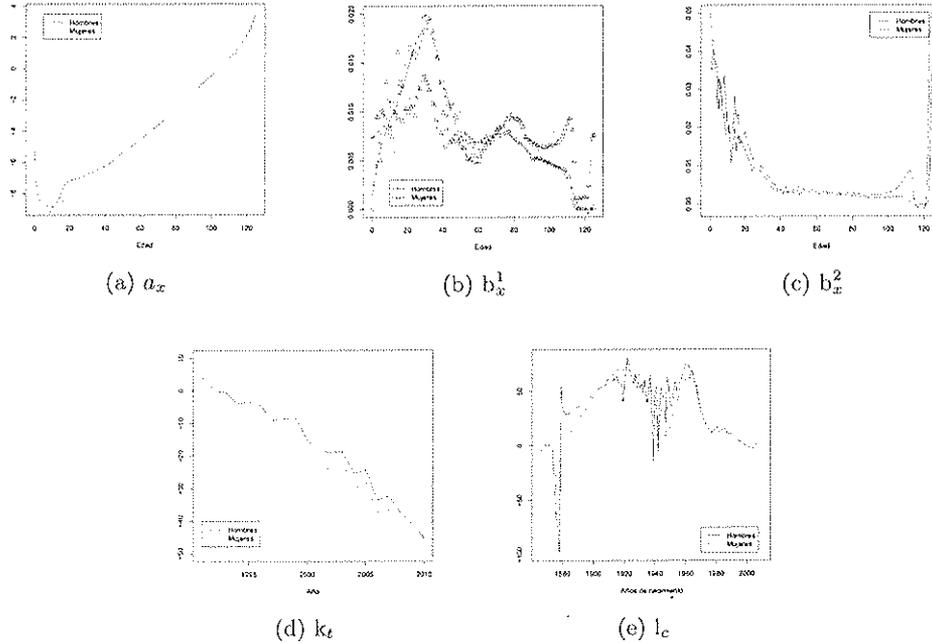
En la Figura 7b se presenta la estimación del parámetro b_x^1 y el suavizado de éste para ambos sexos. Al analizar el gráfico se aprecian algunas diferencias entre hombres y mujeres, no en su tendencia general pero sí en la velocidad del cambio en la mortalidad respecto a la edad, particularmente entre 20 y 40 años. De cualquier modo, el modelo toma valores y tendencias muy similares a los ya observados en los otros modelos. En relación a los intervalos de edades más avanzadas es de esperar un comportamiento en descenso, pero se observa en el gráfico un comportamiento con un aumento en el valor de b_x en el intervalo de 100 a 110 años en las mujeres, indicando que para este intervalo la mortalidad varía sustancialmente cuando el índice k_t cambia. Cabe ser cauto con su interpretación ya que los datos de partida son extrapolados de datos reales.

- *Estimación del parámetro k_t*

La Figura 7d, correspondiente al parámetro k_t , muestra la tendencia de la mortalidad en hombres y mujeres y muestra valores y tendencias similares a los modelos anteriormente explicados.

- *Estimación del parámetro b_x^2*

Figura 7: Parámetros estimados con el modelo Lee-Carter con efecto cohorte



En la Figura 7c, que representa el comportamiento del segundo término b_x^2 , se observan valores más altos en el rango de edades de 0 a 20 años para ambos sexos, estabilizándose aproximadamente en 0 para edades más avanzada. Esto indica que el efecto de añadir el segundo término actúa más en este grupo de edad, no siendo tan significativo para edades entre 40 y 100 años. El acusado crecimiento que se observa en las edades finales es debido al efecto de completar todas las tablas de periodo hasta 130 años, edad que necesariamente ha de tener una probabilidad de muerte igual a 1. Por esta razón las probabilidades para estas últimas edades han de crecer para alcanzar dicho valor.

▪ *Estimación del parámetro l_c*

Respecto al término con efecto cohorte que está representado en la Figura 7e se observan fluctuaciones muy marcadas en la cohorte cuya fecha de nacimiento se sitúa alrededor de 1940, tanto en mujeres como en hombres. Estas fluctuaciones se pueden explicar debido a los efectos de guerra civil española y la época de postguerra. También es destacable la pronunciada declinación en l_c en el periodo entre 1960 y 2010 donde se aprecia la rápida mejora en la mortalidad para la cohorte nacida en este periodo que continua hasta la actualidad aunque a un ritmo más pausado.

En el gráfico se pueden ver los distintos aspectos históricos que influyeron negativamente en la población:

- Años 1918-1920 de las epidemias de la gripe “española” y hundimiento de la economía española a causa de la 1a guerra mundial.
- Años 1936-1939 guerra civil española.

- Años 1940-1950 periodo de postguerra y autarquismo económico, imposibilitando la correcta alimentación y asistencia sanitaria.
- Posiblemente las irregularidades entre 1975 y 1990 se puedan asociar a los fenómenos de accidentes de automóvil y al consumo de drogas.

3.6. Bondad de ajuste del modelo de Lee-Carter con efecto cohorte

Como en los casos anteriores la representación gráfica de los residuos Deviance para ambos colectivos en función de edad, año de calendario y la cohorte permitirá evaluar la bondad de ajuste del modelo. Los resultados se muestran en las Figuras 8 y 9. A partir de las gráficas de los residuos del modelo propuesto se deduce que este modelo captura toda la presencia de los factores incluidos en el modelo aunque se manifiesta mejor en el colectivo de las mujeres. Cabe destacar que no hay datos atípicos ni sobreinfluyentes, lo que indicaría la adecuación del modelo.

Figura 8: Residuos Deviance para el modelo Lee-Carter con efecto cohorte para los hombres

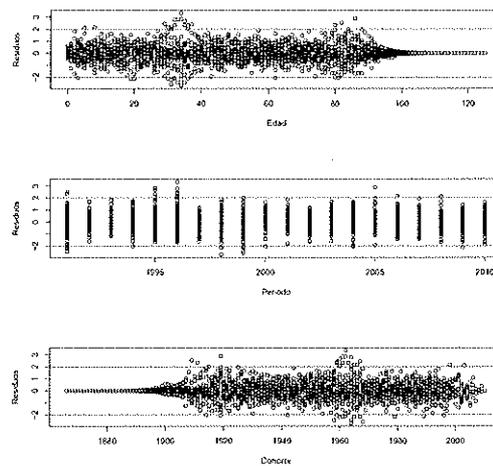
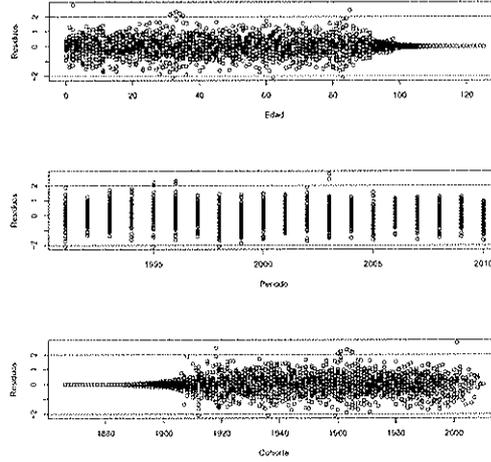


Figura 9: Residuos Deviance para el modelo Lee-Carter con efecto cohorte para las mujeres



3.7. Comparación de los modelos

El análisis de los residuos permite deducir que los tres modelos ajustan razonablemente los datos de la mortalidad española. Para determinar cual de ellos muestra un mejor comportamiento se compararon sus estimaciones con los datos observados mediante las siguientes medidas de bondad ajuste.

- Deviance, que mide la distancia entre los valores q_{xt} observados y los valores \hat{q}_{xt} estimados por el modelo mediante la expresión,

$$D(\hat{q}_{xt}) = 2\log L(q_{xt}) - 2\log L(\hat{q}_{xt}).$$

- Error cuadrático medio (ECM), definido como

$$ECM(\hat{q}_{xt}) = \sqrt{\sum_x \frac{(q_{xt} - \hat{q}_{xt})^2}{T}}, \quad t = 1, \dots, T.$$

- Error porcentual absoluto medio (EPAM), cuya expresión

$$EPAM(\hat{q}_{xt}) = \frac{\sum_x \frac{(q_{xt} - \hat{q}_{xt})}{q_{xt}}}{T}, \quad t = 1, \dots, T.$$

Los correspondientes valores se muestran en la Tabla 1. En general, se observa un grado de ajuste ligeramente superior para las mujeres, lo que parece ser debido a que la población masculina ha sufrido más bruscamente los efectos de guerras, drogas o accidentes de automóvil.

Tabla 1: Estadísticos de bondad de ajuste de los modelos estimados

Modelo	Deviance		ECM		EPAM	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Lee-Carter	2840,13	1532,25	0,0149	0,0082	5,84	6,63
Lee-Carter (dos términos)	2277,77	1282,72	0,0127	0,0076	5,16	6,09
Lee-Carter (efecto cohorte)	1041,52	782,48	0,4001	0,4333	4,93	6,00

Utilizando el contraste de la diferencia de Deviance respecto del modelo de Lee-Carter más sencillo, que se distribuye como una χ^2 con 144 grados de libertad ($144=414-270$) en el caso del modelo de Lee-Carter con dos términos y 270 grados de libertad ($270=540-270$) para el modelo de Lee-Carter con efecto cohorte, se rechaza la hipótesis nula de equivalencia de los modelos. Según esto, el modelo de Lee-Carter con efecto cohorte es el que ofrece mejor ajuste global. El estadístico EPAM corrobora este hecho. La explicación puede ser debida a la introducción del efecto cohorte. El método de Lee-Carter con dos términos también ofrece un buen ajuste, puesto que presenta un error cuadrático medio menor tanto para hombres como para mujeres. De acuerdo con todo ello, sería aconsejable utilizar los modelos extendidos para las predicciones.

4. Predicciones

Las predicciones a futuro se realizaron usando la secuencia propuesta por Box y Jenkins (1970), 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 índices estimados, para varones y mujeres, pero teniendo en cuenta que el principal fin del modelo buscado es el de generar pronósticos. El cálculo se realizó usando las funciones *auto.arima* y *forecast*, de la librería *forecast* (Hyndman, 2008) de R, que proporcionan de manera automática el ajuste del correspondiente modelo ARIMA y la predicción de futuros valores.

La función *auto.arima* se aplicó a los parámetros \hat{k}_t^i del modelo de Lee-Carter con dos términos y a los parámetros k_t y l_c del modelo Lee-Carter con efecto cohorte. Los resultados del diagnóstico del ajuste de los modelos se muestran en las Tablas 2 y 3.

Tabla 2: Criterios de Akaike y Schwartz para los modelos ARIMA para el modelo Lee-Carter con dos términos

Índice	Hombres			Mujeres		
	Modelo	AIC	BIC	Modelo	AIC	BIC
k_t^1	ARIMA(0,1,1)	95,49	98,33	ARIMA(0,1,0)	84,42	86,31
k_t^2	ARIMA(0,1,0)	97,91	98,85	ARIMA(1,0,0)	96,62	98,61

Tabla 3: Criterios de Akaike y Schwartz para los modelos ARIMA para el modelo Lee-Carter con efecto cohorte

Índice	Hombres			Mujeres		
	Modelo	AIC	BIC	Modelo	AIC	BIC
k_t	ARIMA(1,1,0)	95,16	98,00	ARIMA(1,1,0)	98,57	101,40
l_c	ARIMA(1,1,1)	1224,55	1233,46	ARIMA(1,1,2)	1268,04	1279,92

Los resultados de las predicciones obtenidos para los hombres y las mujeres se muestran en las Figuras 10 y 11. Se observa que la tendencia de mortalidad es decreciente en ambos casos. Hay que señalar que los intervalos de predicción presentan una gran amplitud que aumenta a medida que nos alejamos en el tiempo. Esta amplitud se aprecia más en el caso de los hombres para el modelo de Lee-Carter con dos términos.

Figura 10: Proyecciones e intervalos de predicción del 95 % para el índice k_t^1 del modelo Lee-Carter extendido

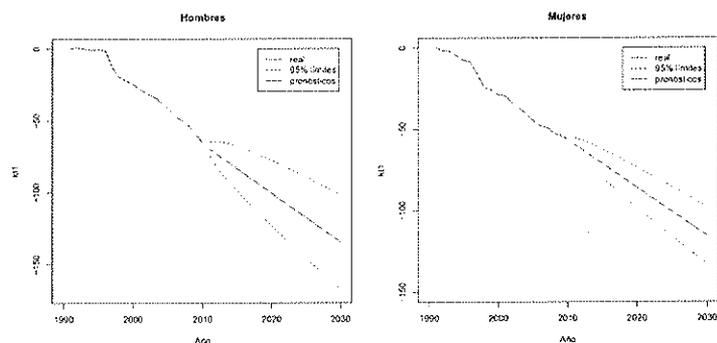
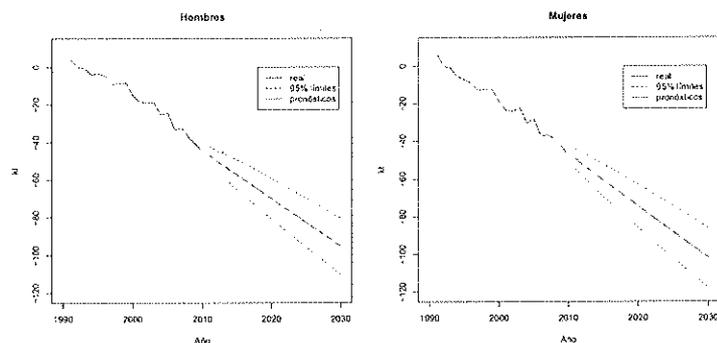


Figura 11: Proyecciones e intervalos de predicción del 95 % para el índice k_t del modelo Lee-Carter cohorte



A partir de los pronósticos generados por el modelo ARIMA y las series temporales ajustadas para el período 1991-2010 se calcularon los pronósticos de probabilidades de muerte q_{xt} para el período 2011-2030. Las Figuras 12 y 13 muestran la superficie de las proyecciones de q_{xt} obtenidas por los dos modelos para hombres y mujeres.

Figura 12: Proyecciones para los ratios de mortalidad (Lee-Carter extendido)

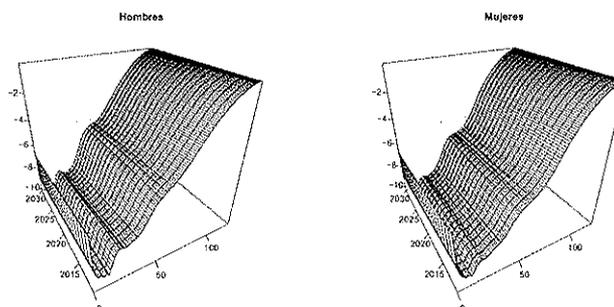
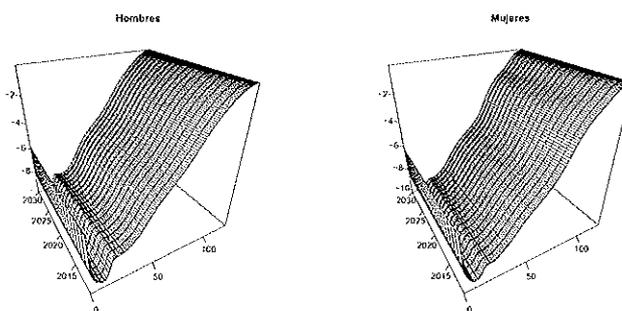


Figura 13: Proyecciones para los ratios de mortalidad (Lee-Carter cohorte)



El cálculo de los intervalos de confianza para las probabilidades de muerte proyectadas se realizó con la técnica propuesta por los autores (Lee y Carter, 1992). El método consiste en obtener los intervalos de confianza teniendo en cuenta sólo el error en la predicción de los índices de mortalidad k_t^i para el modelo con dos términos, y de los índices k_t y l_c para el modelo con efecto cohorte calculados por el modelo ARIMA. Para obtener los extremos respectivos de sus intervalos de confianza se sustituyen los valores del intervalo de confianza según la serie ARIMA en la fórmula de q_{xt} . Con esto se asume que la principal fuente de incertidumbre para el modelo Lee-Carter con dos términos son k_t^i y para el modelo con efecto cohorte son k_t y l_c . Esto permite obviar los errores a a_x y b_x .

Las Figuras 14 y 15 visualizan el ajuste de ambos modelos para el período 1991-2010 y la predicción de las probabilidades de muerte de 2011-2030 con ellos obtenidas. Junto a dichos valores se han representado también los intervalos de confianza de $\text{logit}(q_{xt})$ para las predicciones a partir del año 2011. Analizando los gráficos se observa que los dos modelos presentan tendencias decrecientes razonables más pronunciadas en las edades cortas y medianas y más estables a partir de los 50 años, tanto en hombres como en mujeres. Esto demuestra que la mortalidad ha mejorado durante los últimos años. A destacar que los intervalos de confianza generados a partir de los modelos ARIMA para las mujeres presentan mayor amplitud para el modelo Lee-Carter con efecto cohorte, lo que se podría explicar por la influencia de dicho efecto.

En el caso de las edades avanzadas se han representado las edades comprendidas entre 70 y

Figura 14: Predicciones de q_x para algunas edades (Lee-Carter con dos términos)

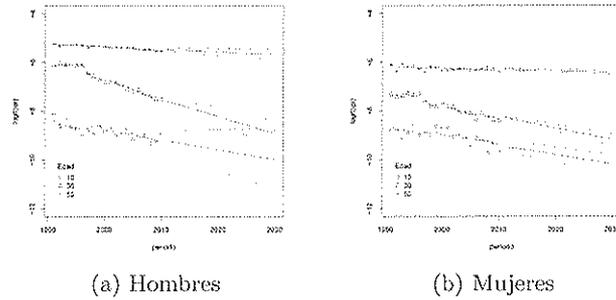
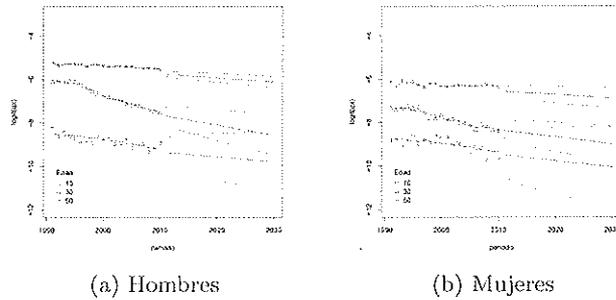


Figura 15: Predicciones de q_x para algunas edades (Lee-Carter con efecto cohorte)



90 años, ambos inclusive. Se observa en las Figuras 16 y 17, que todas ellas presentan tendencias futuras poco pronunciadas. La desaceleración en la probabilidad de muerte a medida que transcurren los años puede explicarse porque de manera natural se produce una selección de los individuos más saludables en edades elevadas (Debón, Montes y Puig, 2008).

Figura 16: Predicciones de q_x para edades elevadas(Lee-Carter con dos términos)

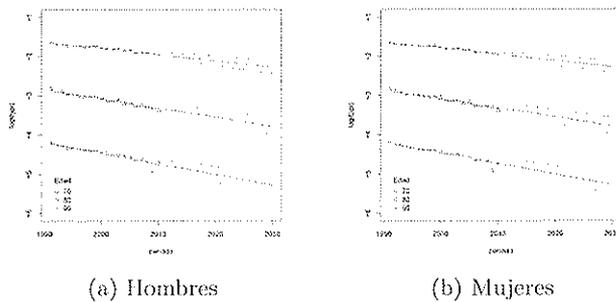
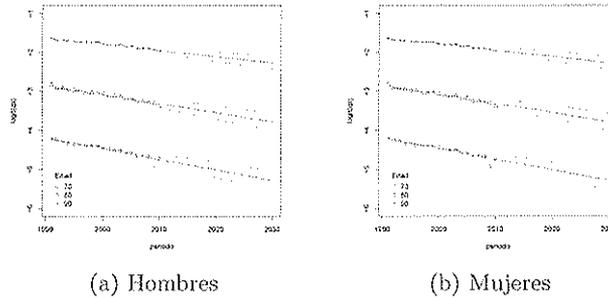


Figura 17: Predicciones de q_x para edades elevadas(Lee-Carter con efecto cohorte)



5. Indicadores de mortalidad

5.1. Esperanza de vida

La esperanza de vida es uno de los indicadores más conocidos y utilizados para conocer la evolución de la mortalidad a lo largo del tiempo. Su valor ha experimentado un crecimiento espectacular durante el último siglo. Cabe recordar que la esperanza de vida a la edad x se interpreta como el número promedio de años que esperan vivir las personas que alcanzan con vida la edad exacta x .

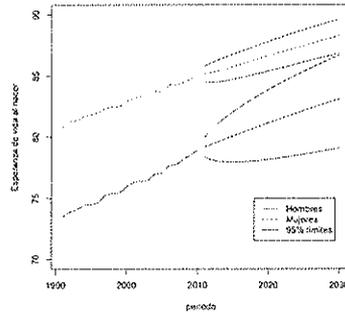
5.2. Coeficiente de Gini

Aplicado a la mortalidad, el coeficiente de Gini permite medir las desigualdades en la edad de la muerte y se basa en la curva de Lorenz, que en este contexto es la representación gráfica de la proporción acumulada de años vividos por la población (ordenadas) frente a la proporción acumulada de población (abscisas). Cuando el coeficiente de Gini se aproxima a cero indica que todos los individuos mueren aproximadamente a la misma edad, mientras que si se aproxima a 1 indica que hay grandes diferencias en la edad de muerte, una gran cantidad de individuos mueren muy precozmente y muy pocos consiguen sobrevivir más que la media.

5.3. Evolución y predicción de esperanza de vida al nacer e índice de Gini

En la Figura 18 se muestra la esperanza de vida para hombres y mujeres, estimada en el periodo 1991-2010 y proyectada para el periodo 2011-2030, así como los intervalos de confianza de las últimas. De los resultados que se muestran se desprende que la esperanza de vida se incrementa a lo largo de tiempo tanto en los hombres como en las mujeres. Es interesante destacar la diferencia entre las amplitudes de los intervalos de confianza entre hombres y mujeres. También se percibe la existencia de una mayor longevidad en las mujeres.

Figura 18: Evolución de la esperanza de vida al nacer (Lee-Carter con dos parámetros)



La Figura 19 muestra la evolución del coeficiente de Gini estimado a lo largo del periodo 1991-2010 y proyectado para el periodo 2011-2030. Lo que más destaca en este gráfico es la asimetría en los intervalos de confianza lo que es debido a la utilización de residuos logit. Según Renshaw y Haberman (2008) el problema puede superarse parcialmente mediante residuos Deviance ya que estos residuos permiten el cumplimiento de la hipótesis de la distribución inicial de la medición de la mortalidad. Por otro lado, se observa que el índice de Gini es mayor en los hombres que en las mujeres, lo que muestra una menor variación en la edad de muerte en las mujeres que en los hombres. Se aprecia que el índice en ambos casos es relativamente bajo lo que puede ser debido a la reducción de la mortalidad infantil.

Figura 19: Evolución del índice de Gini (Lee-Carter con dos parámetros)

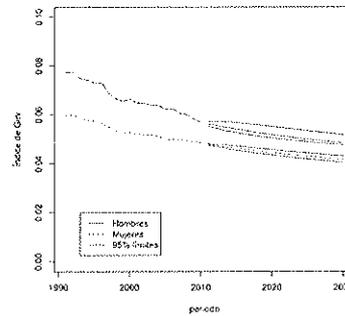


Tabla 4: Predicciones de indicadores de mortalidad para el periodo 2011-2030

Año	Lee-Carter con dos parámetros				Lee-Carter con efecto cohorte			
	Esperanza de vida		Índice de Gini		Esperanza de vida		Índice de Gini	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
2011	79,18	85,15	0,0562	0,0479	80,53	86,32	0,0564	0,0493
2012	79,41	85,27	0,0556	0,0474	80,82	86,59	0,0563	0,0487
2013	79,64	85,42	0,0551	0,0470	81,04	86,71	0,0562	0,0488
2014	79,86	85,59	0,0545	0,0466	81,29	86,92	0,0560	0,0485
2015	80,08	85,76	0,0540	0,0462	81,52	87,10	0,0558	0,0483
2016	80,30	85,93	0,0535	0,0459	81,76	87,27	0,0556	0,0482
2017	80,51	86,10	0,0531	0,0455	81,98	87,46	0,0554	0,0480
2018	80,73	86,28	0,0526	0,0452	82,21	87,64	0,0552	0,0478
2019	80,93	86,45	0,0522	0,0448	82,44	87,81	0,0549	0,0477
2020	81,14	86,63	0,0518	0,0445	82,66	88,00	0,0547	0,0475
2021	81,35	86,80	0,0514	0,0442	82,88	88,17	0,0545	0,0474
2022	81,55	86,97	0,0510	0,0439	83,10	88,35	0,0544	0,0472
2023	81,75	87,14	0,0506	0,0436	83,31	88,53	0,0542	0,0471
2024	81,95	87,31	0,0503	0,0433	83,52	88,70	0,0540	0,0470
2025	82,15	87,47	0,0499	0,0430	83,73	88,88	0,0538	0,0469
2026	82,34	87,64	0,0496	0,0427	83,94	89,06	0,0536	0,0467
2027	82,53	87,80	0,0493	0,0424	84,15	89,23	0,0534	0,0466
2028	82,72	87,96	0,0490	0,0422	84,35	89,40	0,0533	0,0465
2029	82,91	88,13	0,0487	0,0419	84,55	89,58	0,0531	0,0464
2030	83,10	88,28	0,0484	0,0416	84,75	89,75	0,0529	0,0463

En la Tabla 4 se representan los indicadores de mortalidad proyectados para el periodo 2011-2030. En comparación con estudios similares llevados a cabo utilizando datos españoles, nuestros resultados sobre la esperanza de vida son ligeramente superiores a los obtenidos por Debón, Martínez-Ruiz y Montes (2011). Respecto al índice de Gini también se observa una diferencia ligera en los resultados, lo que puede ser debido al factor edad, ya que este estudio se ha realizado con el rango de edad de 0 hasta 125 años lo que podría influir en la igualdad de muerte.

6. Conclusiones

- El comportamiento del parámetro estimado a_x revela que la mortalidad de las mujeres es inferior a la de los hombres. Se observa un aumento en la mortalidad especialmente en el colectivo de hombres en el intervalo de edades de 18 a 30 años. Esta incidencia puede explicarse por el efecto de los accidentes de tráfico mayor en hombres.
- En relación al parámetro b_x se confirma la disminución de la mortalidad con el tiempo independientemente de la edad.
- El índice general de mortalidad k_t presenta una clara tendencia a disminuir a través del tiempo, más acusada en hombres que en mujeres.

A partir de la comparación de los modelos presentada en la Sección 3.7 se constata que el modelo de Lee-Carter con efecto cohorte presenta una mejor bondad del ajuste que el modelo de Lee-Carter original o el modelo con dos términos. 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, haciendo intervenir tanto el año de muerte como el de nacimiento, a diferencia de los otros dos en los que sólo interviene el año de muerte. De esta manera el modelo permite predecir la mortalidad futura para un número prudencial de años de calendario. El modelo es sensible a hechos históricos influyentes en un corto periodo de tiempo como la guerra civil, posguerra y epidemias, o a fenómenos de largo alcance temporal como la mejora sostenida en sanidad y alimentación a lo largo del siglo XX.

El cálculo y la predicción de los índices de esperanza de vida y de Gini permite afirmar:

- Las probabilidades de muerte en España han mejorado durante los últimos años tanto en hombres como en mujeres, esta tendencia se puede extrapolar a los siguientes años.
- Se percibe la existencia de una mayor longevidad en las mujeres que en los hombres, mostrando la esperanza de vida mayor y menor índice de Gini, aunque la diferencia tiende a decrecer. Las predicciones muestran que continuará en el futuro cercano.

Referencias

- Betzuen, A. (2010). Un análisis sobre las posibilidades de predicción de la mortalidad futura aplicando el modelo Lee-Carter. *Anales del Institución de Actuarios Españoles*, 16:111–140.
- Betzuen, A., Felipe, A., y Guillén, M. (1997). Modelos de tablas de mortalidad en España y situación actual. *Anales del Institución de Actuarios Españoles*, 3:79–104.
- Booth, H., Maindonald, J., y Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Demographic Research*, 15:325–336.
- Booth, H. and Hyndman, R., Tickle, L., y de Jong, P. (2006). Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research*, 15:289–310.
- Box, G. y Jenkins, G. (1970). Time series analysis: Forecasting and control. *San Francisco: Holden-Day*.
- Cairns, A., Blake, D., y Dowd, K. (2008). Modelling and Management of Mortality Risk: A Review. *Scandinavian Actuarial Journal*, 2–3:79–113.
- Chan, W.-S., Li, S.-H., y Cheung, S.-H. (2008). Testing deterministic versus stochastic trends in the Lee-Carter mortality indexes and its implications for projecting mortality improvements at advanced ages. *Living to 100 and Beyond Symposium, Orlando*.
- Debón, A. (2003). *Graduación de tablas de mortalidad. Aplicaciones actuariales*. Tesis doctoral, Departamento de Estadística e Investigación Operativa. Universitat de València., València.
- Debón, A., Martínez-Ruiz, F., y Montes, F. (2010). A geostatistical approach for dynamic life table: The effect of mortality on remaining lifetime and annuities. *Insurance: Mathematics & Economics*.
- Debón, A., Martínez-Ruiz, F., y Montes, F. (2011). Temporal evolution of some mortality indicators. application to spanish data. *Living to 100 Symposium. Society of Actuaries*.
- Debón, A., Montes, F., y Puig, F. (2008). Modelling and forecasting mortality in Spain. *European Journal of Operation Research*, 3(189):624–637.

- Debón, A., Montes, F., y Sala, R. (2009). *Tablas de mortalidad dinámicas. Una aplicación a la hipoteca inversa en España*. Publicaciones de la Universitat de València, València.
- Denuit, M. y Goderniaux, A. (2005). Closing and projecting life tables using log-linear models. *Bulletin of the Swiss Association of Actuaries*.
- Giroi, F. y King, G. (2007). Understanding the Lee-Carter mortality forecasting method. <http://gking.harvard.edu/files/abs/lc-abs.shtml>.
- Goerlich, F. y Pinilla, R. (2006). Esperanza de vida en España a lo largo del siglo XX. Technical report, Fundación BBVA., Bilbao.
- Hyndman, R. J. (2008). *Forecast: Forecasting functions for time series. R package version 1.11*.
- IMF. (2012). Global financial stability report. The financial impact of longevity risk. chapter 4. International Monetary Fund.
- INE. (1946). Tablas de mortalidad de la población española. Años 1930-31. Instituto Nacional de Estadística (INE). Madrid.
- INE. (1997). Evolución de la población de España entre los censos de 1981 y 1991. Instituto Nacional de Estadística (INE). Madrid.
- INE. (2009). Tablas de mortalidad. Metodología. Instituto Nacional de Estadística (INE). Madrid.
- Land, K., Yang, Y., y Yi, Z. (2005). *The Handbook of Population: Part III. Mathematical demography*, capítulo 22. Kluwer Academic/Plenum Publishers.
- Lee, R. y Carter, L. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association.*, 87(419):659–671.
- Lee, R. y Nault, F. (1993). Modeling and forecasting provincial mortality in Canada. Paper presented at the World Congress of the International Union for Scientific Study of Population.
- Lee, R. y Rofman, R. (1994). Modelación y proyección de la mortalidad en Chile. *Notas de población.*, 22(59):182–213.
- Llorca, J., Prieto, M., Alvarez, C., y Delgado-Rodríguez, M. (1998). Age differential mortality in Spain, 1900-1991. *Journal of Epidemiology and Community Health.*, 4(52):259–261.
- Llorca, J., Prieto, M., y Delgado-Rodríguez, M. (2000). Medición de las desigualdades en la edad de muerte: cálculo del índice de Gini a partir de las tablas de mortalidad. *Revista española de Salud Pública.*, 74(1):5–12.
- Palacios, H. (2009). *Introducción al cálculo actuarial*. Madrid.
- Renshaw, A. y Haberman, S. (2003). Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. *Journal of the Royal Statistical Society C.*, 1(52):119–137.
- Renshaw, A. y Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics & Economics.*, 3(38):556–570.
- Renshaw, A. y Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling. *Insurance: Mathematics & Economics.*, (42):797–816.

- Shkolnikov, V., Andreev, E., y Begun, A. (2003). Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples. *Demographic Research.*, 8(11):305-358.
- Turner, H. y Firth, D. (2006). *Generalized nonlinear models in R: An overview of gnm package. R package version 1.0-1.*

