

VNIVERSITAT [0%]
ID VALÈNCIA Facultat d'Economia
POSTGRÀU

**MÁSTER
EN CIENCIAS
ACTUARIALES
Y FINANCIERAS**

TRABAJO DE FIN DE MÁSTER

**"Análisis y estudio de la mortalidad en España:
Búsqueda de convergencia entre población nativa y
extranjera aplicando el modelo Lee-Carter"**

ALUMNO/A: Diego Hernán Siliato

TUTOR/A: Francisco Muñoz Murgui

DEPARTAMENTO: Economía Financiera y Actuarial

CURSO: 2024/2025

FECHA DE DEPÓSITO: 10/01/2025



RESUMEN

Este trabajo analiza la evolución de la mortalidad en España, considerando tanto a la población española como a los colectivos extranjeros residentes en el país. A través de la aplicación del modelo Lee-Carter y de técnicas de suavización, se proyectaron tasas de mortalidad a 15 años, permitiendo evaluar si las probabilidades de fallecimiento de los colectivos extranjeros convergen hacia las de la población española. El análisis incluyó la divergencia KL y la comparación de primas y SCR de mortalidad y longevidad, identificando un proceso general de convergencia, aunque con variaciones entre los colectivos. Los resultados sugieren que la integración socioeconómica, educativa y sanitaria desempeña un papel clave en este fenómeno, independientemente del punto de partida de quienes se integran. Además, los hallazgos permiten inferir que el riesgo de mortalidad y longevidad tiende a ser homogéneo en la población residente en España, con implicaciones relevantes para la evaluación de riesgos y la planificación de productos de Vida-Ahorro.

PALABRAS CLAVE: Mortalidad, riesgo de mortalidad, convergencia, modelo Lee-Carter, población española, colectivos extranjeros, suavización, proyección, divergencia KL, riesgo de longevidad, Vida-Ahorro.



RESUM

Aquest treball analitza l'evolució de la mortalitat a Espanya, considerant tant la població espanyola com els col·lectius estrangers residents al país. Mitjançant l'aplicació del model Lee-Carter i tècniques de suavització, es van projectar taxes de mortalitat a 15 anys, la qual cosa ha permès avaluar si les probabilitats de defunció dels col·lectius estrangers convergeixen amb les de la població espanyola. L'anàlisi ha inclòs la divergència KL i la comparació de primes i SCR de mortalitat i longevitat, identificant un procés general de convergència, encara que amb variacions entre els col·lectius. Els resultats suggereixen que la integració socioeconòmica, educativa i sanitària juga un paper clau en aquest fenomen, independentment del punt de partida de qui s'integra. A més, els resultats permeten inferir que el risc de mortalitat i longevitat tendeix a ser homogeni en la població resident a Espanya, amb implicacions rellevants per a l'avaluació de riscos i la planificació de productes de Vida-Estalvi.

PARAULES CLAU: Mortalitat, risc de mortalitat, convergència, model Lee-Carter, població espanyola, col·lectius estrangers, suavització, projecció, divergència KL, risc de longevitat, Vida-Estalvi.



SUMMARY

This study analyzes mortality trends in Spain, focusing on both the Spanish population and foreign resident groups. By applying the Lee-Carter model and smoothing techniques, mortality rates were projected over 15 years, enabling an assessment of whether the mortality probabilities of foreign groups converge with those of the Spanish population. The analysis included KL divergence and comparisons of premiums and mortality and longevity SCR, identifying a general convergence process with variations across groups. Results suggest that socioeconomic, educational, and healthcare integration plays a key role in this phenomenon, regardless of the starting point of those integrating. Furthermore, the findings suggest that mortality and longevity risks tend to be homogeneous within the resident population, with significant implications for risk assessment and Life-Savings product planning.

KEYWORDS: Mortality, mortality risk, convergence, Lee-Carter model, Spanish population, foreign groups, smoothing, projection, KL divergence, longevity risk, Life-Savings products.



Contenido

1.	Introducción.....	7
2.	Metodología.....	8
2.3.	El Modelo Lee – Carter	9
2.3.1.	Estructura del modelo.....	9
2.3.2.	Estimación de parámetros.....	10
2.3.3.	Residuos.....	12
2.3.4.	Predicción.....	12
2.4.	El modelo ARIMA	12
2.4.2.	El modelo ARIMA. 0, 1, 0	13
2.4.3.	Propuesta de Modelo	13
2.4.3.1.	Formulación del Modelo e interpretación	13
2.4.3.2.	Estimación de Parámetros.....	14
2.4.4.	Aplicación del Modelo	14
2.5.	KL Divergence (Divergencia de Kullback – Leibler).....	14
2.6.	Suavizado con Splines	15
2.6.1.	Definición Formal de un Spline Cúbico.....	15
	La forma general de un Spline Cúbico en un intervalo $x_i, x_i + 1$ es la siguiente:	15
3.	Software R: Paquetes Demography y Forecast	16
4.	Resultados.....	16
4.1.	Tasas de mortalidad real y suavizadas para españoles y los distintos colectivos.....	16
4.2.	Ajuste del modelo Lee – Carter y Tasas de mortalidad proyectadas para españoles y los distintos colectivos.....	19
4.3.	Divergencia Kullback-Leibler (KL).....	25
4.4.	Primas para Productos de Vida-Ahorro utilizando probabilidades de fallecimientos proyectadas.....	27
5.	Conclusión.....	30
6.	Bibliografía	32
7.	Anexos.....	33
7.1.	Principios básicos de biometría.....	33
7.1.1.	Probabilidades de supervivencia sobre un individuo.....	33
7.1.2.	Probabilidad de fallecimiento	33
7.1.3.	Extensión de los conceptos biométricos según origen del colectivo analizado.....	
	34	
7.2.	Cálculo de kt y bx utilizando SVD.....	36



Índice de gráficos

Gráfico 1. Probabilidad de fallecimiento española de nativos real y suavizada.....	17
Gráfico 2. Probabilidad de fallecimiento para el Colectivo 1 real y suavizada.....	17
Gráfico 3. Probabilidad de fallecimiento para el Colectivo 2 real y suavizada.....	18
Gráfico 4. Probabilidad de fallecimiento para el Colectivo 3 real y suavizada.....	18
Gráfico 5. Probabilidad de fallecimiento para el Colectivo 1 real y suavizada.....	18
Gráfico 6. Probabilidad de fallecimiento para el Colectivo 1 real y suavizada.....	19
Gráfico 7. Probabilidad de fallecimiento para el Colectivo 1 real y suavizada.....	19
Gráfico 8. Modelo Lee-Carter para españoles nativos y probabilidades de fallecimiento en escala logarítmica proyectadas.	20
Gráfico 9. Gráfico 8. Modelo Lee-Carter para el Colectivo 1 y probabilidades de fallecimiento en escala logarítmica proyectadas.	21
Gráfico 10. Modelo Lee-Carter para el Colectivo 2 y probabilidades de fallecimiento en escala logarítmica proyectadas.	22
Gráfico 11. Modelo Lee-Carter para el Colectivo 3 y probabilidades de fallecimiento en escala logarítmica proyectadas.	22
Gráfico 12. Modelo Lee-Carter para el Colectivo 4 y probabilidades de fallecimiento en escala logarítmica proyectadas.	23
Gráfico 13. Modelo Lee-Carter para el Colectivo 5 y probabilidades de fallecimiento en escala logarítmica proyectadas.	24
Gráfico 14. Modelo Lee-Carter para el Colectivo 6 y probabilidades de fallecimiento en escala logarítmica proyectadas.	24
Gráfico 15. Valores de KL para cada año proyectado y para cada colectivo analizado.	26
Gráfico 16. Boxplot comparativo entre valores de primas calculadas.	28

Índice de tablas

Tabla 1. PIB per cápita promedio para cada grupo. Fuente: Elaboración propia en base a datos BancoMundial.org.....	8
Tabla 2. Valores de KL para cada año proyectado y para cada colectivo analizado. Fuente: Elaboración propia.....	26
Tabla 3. Valores de primas estimadas para cada año proyectado y por grupo analizado.	27
Tabla 4. Primas, SCR mortalidad y SCR longevidad de cada subgrupo para 5, 10 y 15 años.....	29



1. Introducción.

España se ha consolidado como el país de la Unión Europea con la mayor esperanza de vida, con una media de 83,3 años, seguida de Suecia (83,1 años), Luxemburgo e Italia (ambos con 82,7 años). En las últimas décadas, la esperanza de vida al nacer ha aumentado considerablemente tanto en hombres como en mujeres (*La Vanguardia*, 2024). Sin embargo, surge la pregunta si este incremento se aplica a todas las personas residentes en España o solo a los españoles.

Según datos del Instituto Nacional de Estadística (INE), en 2022 España recibió 1.177.285 migrantes de diversas nacionalidades, lo que representa un aumento del 46 % en comparación con 2021. Este crecimiento refleja una mayor participación de extranjeros en la población total del país, lo que convierte este fenómeno en un aspecto esencial a considerar al estudiar dinámicas demográficas tan relevantes como la esperanza de vida y la mortalidad. En este contexto, se prevé que los colectivos extranjeros representen una proporción cada vez más significativa en las carteras de las compañías aseguradoras radicadas en España. Un ejemplo de ello es el caso de una aseguradora que ha incrementado en un 12 % sus ingresos gracias a los seguros dirigidos a extranjeros residentes en el país (*Cortés*, 2024).

La esperanza de vida es el indicador más utilizado para comparar la incidencia de la mortalidad en diferentes poblaciones, así como las condiciones de salud y el nivel de desarrollo. En los países occidentales, la esperanza de vida ha avanzado notablemente en el último siglo, gracias a los avances médicos y tecnológicos, la reducción de la mortalidad infantil, los cambios en los hábitos nutricionales, la mejora en las condiciones educativas y el acceso a servicios sanitarios (*García González, J. M., 2014*).

Estudios indican que la reducción de la mortalidad en el primer año de vida ha contribuido con 1,26 años al aumento de la esperanza de vida al nacer entre 1975 y 2009, totalizando una ganancia de 7,9 años. Esta contribución, que representa el 16% de las ganancias totales, es significativa en comparación con otros grupos etarios. Por ejemplo, el grupo de edad entre 45 y 64 años aportó 1,69 años, mientras que el grupo de 65 a 79 años contribuyó con 2,83 años (*Goerlich Gisbert, 2012*). Por lo tanto, es plausible que las condiciones socioeconómicas del país de origen influyan en la esperanza de vida de los migrantes. Se supone que la mayoría de los movimientos migratorios los realizan individuos que ya han superado esos primeros años de vida.

El presente Trabajo Final de Máster (TFM) tiene como objetivo analizar la evolución de la mortalidad en España, considerando tanto a la población nativa como a distintos grupos de población extranjera residente en el país. Estudios recientes (*Atance, D., 2024*) indican un claro crecimiento en los procesos de envejecimiento en todo el mundo durante los últimos 30 años, lo que resalta la importancia de investigar estas dinámicas a nivel país. El propósito central es evaluar si la mortalidad de la población extranjera tiende a converger con la de los ciudadanos españoles. De esta manera, se pretende constatar si, al integrarse y beneficiarse de los aspectos socioeconómicos, educativos y sanitarios de España, la mortalidad de los distintos colectivos tiende a asemejarse a la española nativa, independientemente de las condiciones iniciales de sus lugares de origen. De ser así, esto

proporcionará evidencia sobre si el riesgo de longevidad es homogéneo para toda la población residente.

Para este estudio, se utilizará el modelo Lee-Carter, reconocido por su simplicidad y fiabilidad en la proyección de la mortalidad futura. Una vez obtenidas las estimaciones de las probabilidades de fallecimiento futuras, se llevarán a cabo dos procedimientos para intentar obtener conclusiones congruentes. Por un lado, realizaremos la prueba de divergencia de Kullback-Leibler, comparando las probabilidades de fallecimiento de los distintos colectivos con las de la población española. Por otro, se calcularán las primas para un producto de vida-ahorro utilizando las probabilidades estimadas, con el objetivo de cuantificar monetariamente las diferencias encontradas.

2. Metodología

2.1. Fuentes de Datos

Para este estudio, utilizaremos datos proporcionados por el Instituto Nacional de Estadística (INE) de España. Los datos específicos que se emplearán incluyen:

Población residente: Número de personas que residen en territorio español con información desglosada por fecha, sexo, generación (edad a 31 de diciembre, de 0 a 90 años o más), nacionalidad (agrupación de países) y lugar de nacimiento (agrupación de países). La agrupación de países que utiliza el INE la hemos identificado con la siguiente denominación y su promedio del PBI per cápita.

Identificador	PIB per cápita promedio (US\$ a precios constantes de 2015)
Colectivo 1	33.483
Colectivo 2	2.446
Colectivo 3	10.389
Colectivo 4	18.058
Colectivo 5	40.195
Colectivo 6	12.870

Tabla 1. PIB per cápita promedio para cada grupo. Fuente: Elaboración propia en base a datos BancoMundial.org

Defunciones: Datos del Movimiento Natural de la Población (MNP), específicamente la Estadística de Defunciones (MNPdefun), que recoge los fallecimientos ocurridos en España cada año. Estos datos se obtienen a través del certificado médico de defunción/boletín estadístico de defunción, que incluye la edad y el país de origen de los fallecidos.

2.2. Procedimientos

- **Cálculo de las Probabilidades de Fallecimiento:** Los datos de fallecimientos se agruparán según las categorías de nacionalidad y lugar de nacimiento definidas

por el INE para la población residente. Esto permitirá calcular las probabilidades de fallecimiento tanto de la población española como de los distintos colectivos de población extranjera residente en territorio español.

- **Suavizado de las tasas de mortalidad:** Para abordar las fluctuaciones irregulares en las tasas de mortalidad, se utilizará un método de suavizado mediante Splines. Este enfoque permite suavizar las variaciones abruptas y proporcionar una representación más continua y precisa de las tasas de mortalidad a lo largo del tiempo
- **Aplicación del modelo Lee Carter para la obtención de las tasas futuras de fallecimiento:** Este modelo, reconocido por capturar tendencias históricas de mortalidad, proporcionará estimaciones fiables de las tasas futuras de fallecimiento. Su aplicación permitirá obtener proyecciones de la mortalidad en la población española y en los distintos colectivos de población extranjera.
- **Test de Divergencia de Kullback-Leibler:** Se utilizará para comparar las probabilidades de fallecimiento proyectadas entre los distintos colectivos de población y la población española. Esto permitirá evaluar las diferencias y observar si existe una convergencia o divergencia entre ellas.
- **Cálculo de Primas para Productos de Vida-Ahorro:** En nuestro análisis, evaluaremos las primas para un producto ficticio de vida-ahorro para una persona de 58 años y serán calculadas con las probabilidades de fallecimiento proyectadas a 15 años. Este producto se caracterizará por una duración de 5 años y un capital asegurado de 100.000 en caso de fallecimiento y 50.000 en caso de supervivencia. El tipo de descuento máximo será del 2%. Esto nos permitirá comparar las primas entre la población española y los distintos colectivos, identificando y cuantificando monetariamente las posibles diferencias.

2.3. El Modelo *Lee – Carter*.

El modelo Lee-Carter (*Lee. R. D, Carter, L. R.,1992*), es uno de los enfoques más influyentes y utilizados para la proyección de la mortalidad y la esperanza de vida. Este modelo se ha convertido en un estándar en demografía y en estudios actuariales debido a su capacidad para capturar las tendencias y patrones subyacentes en las tasas de mortalidad a lo largo del tiempo.

2.3.1. Estructura del modelo.

Se basa en la descomposición de las tasas de mortalidad en componentes que varían con la edad y el tiempo. La fórmula fundamental del modelo es la siguiente:

$$\log m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}$$

donde:

- $m_{x,t}$ es la tasa de mortalidad a la edad x en el año t .

- a_x representa el nivel promedio de mortalidad específico de la edad x .
- b_x es la sensibilidad de la mortalidad a la edad x respecto a los cambios en el índice temporal k_t .
- k_t es el índice de mortalidad en el año t , que refleja el nivel general de mortalidad en la población.
- $\epsilon_{x,t}$ es el término de error aleatorio que captura las variaciones no explicadas por el modelo.

El parámetro a_x se interpreta como el perfil de mortalidad específico de la edad, mientras que b_x indica cómo las tasas de mortalidad para una edad particular responden a cambios en el índice temporal k_t . Este último, k_t , generalmente se modela como una serie temporal para capturar las tendencias generales de mortalidad en la población.

2.3.2. Estimación de parámetros.

Para estimar los parámetros del modelo Lee-Carter, se utiliza el método de descomposición de valores singulares (SVD, por sus siglas en inglés). Este enfoque permite descomponer la matriz de tasas de mortalidad observadas en sus componentes principales. Los pasos típicos incluyen:

1. **Cálculo de a_x** : Promediando los logaritmos de las tasas de mortalidad observadas a través del tiempo para cada edad x .

$$a_x = \frac{1}{T} \sum_{t=1}^T \log m_{x,t}.$$

2. **Cálculo de k_t y b_x** : Para esta estimación se utiliza el método de la descomposición en valores singulares de una matriz A , DVS. Para este proceso se debe de conocer el cálculo de los valores y vectores propios de una matriz, este cálculo se realiza no sobre la matriz A , sino sobre la matriz producto de A con su traspuesta A^t , AA^t .

La aplicación del método DVS nos proporciona:

- Los valores propios de la matriz AA^t :

$$\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_s\}$$

Denotados por orden de mayor a menor valor.

Estos autovalores son siempre positivos o cero.

Los autovalores nulos y se asocian a un subespacio denominado núcleo.

- Los valores singulares de la matriz A se construyen a partir de los autovalores de AA^t :

$$\{\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \dots, \sigma_s = \sqrt{\lambda_s}\}.$$

- Los vectores propios de la matriz AA^t , son:

$$\{v_1, v_2, \dots, v_s\}.$$

Estos auto vectores forman una base ortonormal del espacio.

Existen unos auto vectores asociados a valores singulares positivos y otros asociados a valores singulares cero.

Se pueden construir dos familias de vectores ortogonales y que forman una base del espacio:

- $\{Av_1, Av_2, \dots, Av_s\}$: es ortogonal y $\|Av_i\| = \sqrt{\lambda_i} = \sigma_i$.
- $\left\{u_1 = \frac{Av_1}{\sigma_1}, u_2 = \frac{Av_2}{\sigma_2}, \dots, u_l = \frac{Av_l}{\sigma_l}\right\}$ es ortonormal de $\text{Col}(A)$, donde $1, \dots, l$ denotan los autovalores positivos de AA^t .
- $\{v_1, v_2, \dots, v_s\}$: es ortonormal de $\text{Ker}(A)$, donde $l + 1, \dots, s$ denotan los autovalores nulos de AA^t .
- De esta manera se tiene que:
 - $\hat{b}_x^{DVS} = u_1^x$ y $\hat{k}_t^{DVS} = \sigma_1 v_1^t$.
 - Se debe de notar que $\sum_x b_x^i = c$. Por lo tanto, para que se cumpla la condición que garantiza la unicidad se debe de realizar una reestimación de los parámetros de la forma siguiente:

$$\widehat{b}_x = \frac{\hat{b}_x^{DVS}}{c}; \quad \hat{k}_t = c \hat{k}_t^{DVS}$$

Los valores estimados pueden presentar errores en relación a los valores observados por diferentes motivos, entre ellos por utilizar el logaritmo, esto motiva que se proceda a realizar una reestimación de k_t :

- $D_t = \sum_x D_{x,t}$, la suma de las defunciones totales en cada periodo de tiempo considerado.
- $E_{x,t}$, los expuestos a cada edad en el periodo t .
- Se resuelve la ecuación:

$$\sum_x E_{x,t} = \left(\frac{e^{\check{a}_x + \sum_i k_t^i \hat{b}_x^i}}{1 + e^{\check{a}_x + \sum_i k_t^i \hat{b}_x^i}} \right);$$

- El parámetro reestimado será entonces:

$$\hat{k}_t^i = \tilde{k}_t^i - \frac{\sum_t \tilde{k}_t^i}{T} \text{ y } \hat{a}_x = \check{a}_x + \hat{b}_x^i \frac{\sum_t \tilde{k}_t^i}{T}.$$

2.3.3. Residuos.

Se pueden obtener un tipo de residuos a partir de la expresión:

$$\hat{\varepsilon}_{k,t} = \ln \left(\frac{q_{x,t}}{1-q_{x,t}} \right) - \left(\hat{a}_x + \sum_{i=1}^r \hat{b}_x^i \hat{k}_t^i \right),$$

donde r es el número de valores singulares utilizados.

Si la estimación anterior es correcta, el residuo no debe de presentar ningún tipo de patrón.

2.3.4. Predicción.

Una vez estimados los parámetros, \hat{k}_t^i se puede proyectar hacia el futuro utilizando modelos de series temporales. Esto permite prever las tasas de mortalidad futuras y, por ende, la esperanza de vida.

Un buen modelo es un $ARIMA(0,1,0)$:

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

Donde c es una constante y u_t es un ruido blanco, $N(0,1)$.

El modelo Lee-Carter ha demostrado ser robusto y flexible, permitiendo adaptaciones y extensiones. Por ejemplo, se han propuesto variaciones que incorporan covariables adicionales, que ajustan mejor para poblaciones específicas o que combinan múltiples poblaciones para mejorar la precisión de las proyecciones.

2.4. El modelo *ARIMA*.

2.4.1. Introducción al modelo *ARIMA*.

El modelo *ARIMA*, que significa Autoregressive Integrated Moving Average (Promedio Móvil Integrado Autorregresivo), es una herramienta ampliamente utilizada para el análisis y la predicción de series temporales. Este modelo combina tres componentes:

- **AR (Autorregresivo)**: Representa la dependencia entre una observación y rezagos (valores anteriores) de la misma serie.
- **I (Integrado)**: Indica el número de diferencias requeridas para hacer que una serie temporal no estacionaria se vuelva estacionaria.
- **MA (Media Móvil)**: Modela el error de predicción como una combinación lineal de errores pasados.

El modelo *ARIMA* se denota comúnmente como $ARIMA(p, d, q)$, donde:

- **p** es el orden del componente autorregresivo.
- **d** es el orden de diferenciación.
- **q** es el orden del componente de media móvil.

2.4.2. El modelo *ARIMA*. (0, 1, 0)

El modelo *ARIMA* (0,1,0) es un caso especial de los modelos *ARIMA*, donde los parámetros son:

- $p = 0$: No hay términos autorregresivos.
- $d = 1$: La serie se diferencia una vez para lograr estacionariedad.
- $q = 0$: No hay términos de media móvil.

Este modelo se puede escribir como:

$$Y_t - Y_{t-1} = \varepsilon_t$$

O, de manera equivalente:

$$Y_t = Y_{t-1} + \varepsilon_t$$

donde:

- Y_t es el valor de la serie temporal en el tiempo t .
- Y_{t-1} es el valor de la serie temporal en el tiempo $t - 1$.
- ε_t es el término de error (ruido blanco) en el tiempo t .

Este modelo es básicamente un “paseo aleatorio”, donde cada valor es simplemente el valor anterior más un error aleatorio.

2.4.3. Propuesta de Modelo

2.4.3.1. Formulación del Modelo e interpretación

Proponemos un modelo basado en el *ARIMA* (0,1,0) para modelar \hat{k}_t , que es una serie temporal con una tendencia constante modificada por un término aleatorio. El modelo se expresa como:

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

donde:

- \hat{k}_t es el valor estimado de la serie en el tiempo t .
- c es una constante que captura la tendencia promedio de la serie.
- \hat{k}_{t-1} es el valor estimado en el tiempo $t - 1$.
- u_t es el término de error en el tiempo t , asumido como ruido blanco.

Este modelo incorpora una constante c que representa la tendencia lineal en la serie temporal. A diferencia del modelo *ARIMA* (0,1,0) estándar, que asume un paseo aleatorio puro, nuestro modelo sugiere que hay una tendencia determinista (capturada por c) además del componente aleatorio.

2.4.3.2. Estimación de Parámetros

Para estimar los parámetros del modelo (c y la varianza del error u_t), se puede utilizar el método de máxima verosimilitud o métodos basados en mínimos cuadrados. Los pasos típicos incluyen:

1. **Diferenciar la Serie:** Transformar la serie original k_t en diferencias para obtener una serie estacionaria.

$$\Delta k_t = k_t - k_{t-1}$$

2. **Estimación de la Constante:** Estimar la constante c como el promedio de las diferencias:

$$\hat{c} = \frac{1}{T} \sum_{t=2}^T \Delta k_t$$

3. **Estimación del Error:** Calcular los residuos del modelo para obtener una estimación de la varianza del término de error u_t .

2.4.4. Aplicación del Modelo

Una vez estimados los parámetros, el modelo se puede utilizar para predecir valores futuros de k_t . Por ejemplo, la predicción para el próximo período sería:

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

Como puede observarse, el modelo ARIMA proporciona una herramienta útil para modelar series temporales con tendencias lineales y componentes aleatorios. Este modelo ofrece una manera simple pero efectiva de capturar tanto la tendencia determinista como la variabilidad aleatoria en los datos, haciendo que sea una opción atractiva para la proyección y análisis de series temporales en diversos campos.

2.5. KL Divergence (Divergencia de Kullback – Leibler).

La divergencia de Kullback-Leibler (*Kullback, S., Leibler, R. A., 1951*), es una medida ampliamente utilizada en Estadística y Teoría de la Información para cuantificar las diferencias entre dos distribuciones de probabilidad $P(x)$ y $Q(x)$.

En este estudio, estas distribuciones se construyen a partir de las tasas específicas de mortalidad q_x para cada edad x . La formula general del KL Divergence es:

$$D_{KL}(P||Q) = \sum_x P_{(x)} \log \left(\frac{P_{(x)}}{Q_{(x)}} \right),$$

Donde:

- $P_{(x)}$ representa la distribución de mortalidad de la población española.

- $Q_{(x)}$ representa la distribución de mortalidad de un colectivo distinto al español.
- x es la edad

Dado que las tasas de mortalidad q_x no son distribuciones de probabilidad (es decir, no suman 1 en su rango), es necesario normalizarlas antes de aplicar el divergente KL. Esto se realiza dividiendo cada q_x por la suma total de las tasas de mortalidad para todas las edades en el grupo correspondiente:

$$P_{(x)} = \frac{q_x}{\sum_{x=0}^w q_x},$$

Donde $P_{(x)}$ representa la probabilidad normalizada de morir a la edad x . El cálculo del divergente KL se realiza para cada año del periodo de análisis (2016-2022 y proyecciones futuras). Esto permite observar la evolución de la divergencia a lo largo del tiempo y evaluar si las distribuciones de mortalidad de los colectivos inmigrantes tienden a converger hacia las de la población española. Una KL Divergence decreciente a lo largo del tiempo sugiere una mayor convergencia en los patrones de mortalidad.

2.6. Suavizado con *Splines*.

El suavizado con *Splines* es una técnica matemática utilizada para ajustar curvas suaves a datos que presentan fluctuaciones o ruido. Esta técnica se utiliza principalmente cuando se requiere una representación continua y flexible de los datos, lo que permite mejorar la interpretación de las tendencias subyacentes (*Shyamalkumar, N.D., 2006*). En el contexto de este trabajo, el suavizado de las tasas de mortalidad proyectadas se realiza mediante el uso de splines cúbicos, lo que permite atenuar las fluctuaciones y obtener una representación más estable de las tendencias.

2.6.1. Definición Formal de un *Spline* Cúbico.

Un *Spline* cúbico es una función a trozos definida por polinomios de tercer grado en cada intervalo entre puntos. Dado un conjunto de puntos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ un *Spline* cúbico es una función $S(x)$ que satisface las siguientes condiciones:

- *Interpolación*: El *Spline* pasa exactamente por cada uno de los puntos de los datos. Es decir, $S(x_i) = y_i$ para cada $i = 1, 2, \dots, n$.
- *Continua y suave*: Sus primeras y segundas derivadas son continuas en los puntos de unión, lo que garantiza una transición suave entre los segmentos.
- *Condición de frontera*: Se cumplan las condiciones de frontera elegidas (como la condición de frontera natural, que implica que las segundas derivadas en los extremos sean cero).

La forma general de un *Spline* Cúbico en un intervalo $[x_i, x_{i+1}]$ es la siguiente:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$$

Donde los coeficientes a_i, b_i, c_i, d_i se determinan resolviendo un sistema de ecuaciones que proviene de las condiciones de interpolación, suavidad y las condiciones de frontera.

3. Software R: Paquetes *Demography* y *Forecast*.

El paquete *Demography* en R es una herramienta fundamental para el análisis de datos demográficos, especialmente en la modelización y proyección de tasas de mortalidad. Este paquete permite ajustar el modelo Lee-Carter, uno de los más utilizados para estudiar esta temática. Como hemos visto en el apartado 2.3, este modelo descompone la mortalidad en función de la edad y el tiempo, con un componente temporal k_t que captura la tendencia general de la mortalidad a lo largo del tiempo.

Este paquete facilita tanto el ajuste de este modelo como la visualización y evaluación de los resultados, proporcionando una base sólida para proyecciones futuras.

Una vez ajustado el modelo Lee-Carter y obtenido el componente temporal k_t , se procede a predecir las tasas de mortalidad futuras mediante modelos de series temporales. Para esta etapa, se utilizó el paquete *Forecast* de R, que permite ajustar modelos como *ARIMA* al componente k_t , extendiendo así el modelo Lee-Carter hacia años futuros. *Forecast* ofrece una herramienta eficaz para proyectar la evolución de las tasas de mortalidad, permitiendo realizar análisis sobre las tendencias futuras.

Para tener acceso al código, tablas y resultados obtenidos, puede consultar el repositorio <https://github.com/dsilvia/TFM>.

4. Resultados.

4.1. Tasas de mortalidad real y suavizadas para españoles y los distintos colectivos.

En este apartado se presentan los resultados obtenidos a partir del análisis de las tasas de mortalidad para la población española y los distintos colectivos de población extranjera residente en España. Los datos analizados comprenden las tasas de mortalidad observadas (reales) y las tasas suavizadas, calculadas con el objetivo de reducir las fluctuaciones y mejorar la interpretabilidad de las tendencias.

Las tasas de mortalidad reales presentan patrones distintivos para cada grupo, con fluctuaciones más marcadas en algunos colectivos. En los colectivos 1, 2 y 3, estas variaciones son especialmente evidentes en los rangos de edad de 0 a 40 años, mientras que en los colectivos 4, 5 y 6 las fluctuaciones se extienden a todo el rango etario, principalmente debido al tamaño muestral reducido. En contraste, para la población española, las fluctuaciones son considerablemente más tenues a lo largo de todo el rango de edades, reflejando una mayor estabilidad en las tasas de mortalidad.

Por otro lado, las tasas suavizadas, generadas mediante el uso de splines, permiten identificar con mayor claridad las tendencias subyacentes. Estas curvas reducen significativamente las fluctuaciones en las tasas de mortalidad para todos los grupos, mejorando la interpretación de los datos. Sin embargo, en el caso del Colectivo 5, si bien las curvas muestran una estabilización a lo largo del rango etario, se observan diferencias notables entre los años analizados, lo que podría reflejar variaciones interanuales en las condiciones de mortalidad de este grupo.

A continuación, se presentan los gráficos que ilustran el logaritmo de las tasas de mortalidad, tanto reales como suavizadas mediante splines, para la población española nativa y los seis colectivos analizados:

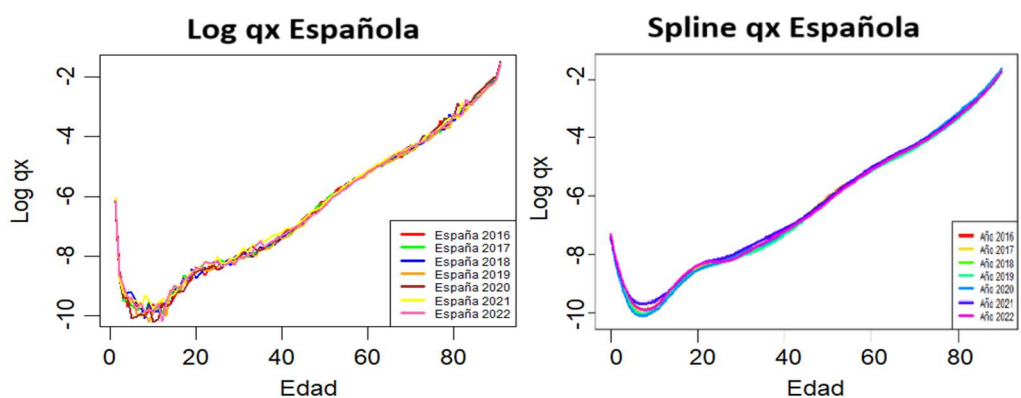


Gráfico 1. Probabilidad de fallecimiento española de nativos real y suavizada.

Fuente: Elaboración propia en base a datos INE.

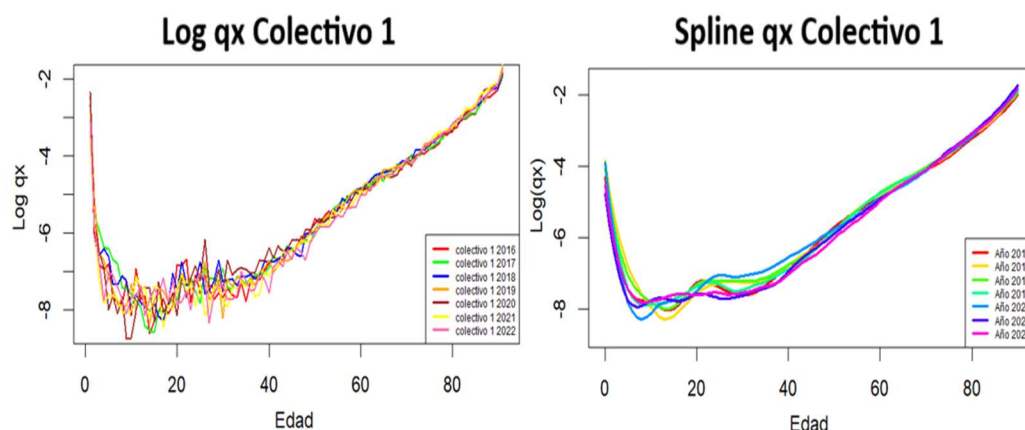


Gráfico 2. Probabilidad de fallecimiento para el Colectivo 1 real y suavizada.

Fuente: Elaboración propia en base a datos INE.

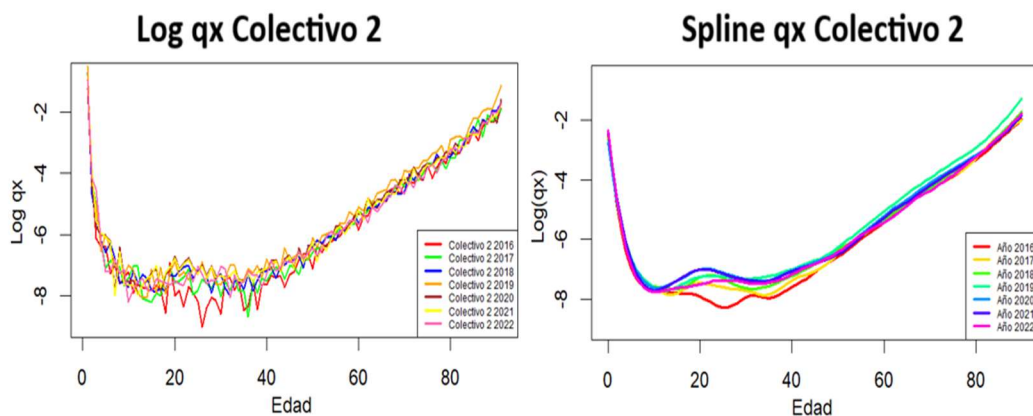


Gráfico 3. Probabilidad de fallecimiento para el Colectivo 2 real y suavizada.

Fuente: Elaboración propia en base a datos INE.

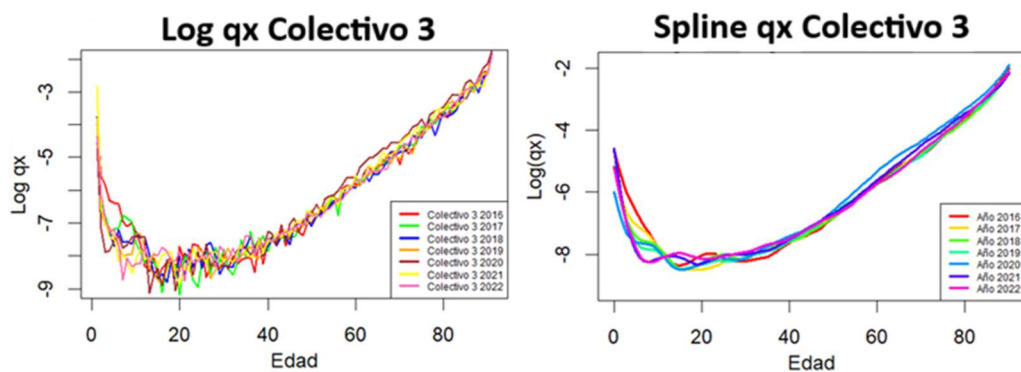


Gráfico 4. Probabilidad de fallecimiento para el Colectivo 3 real y suavizada.

Fuente: Elaboración propia en base a datos INE.

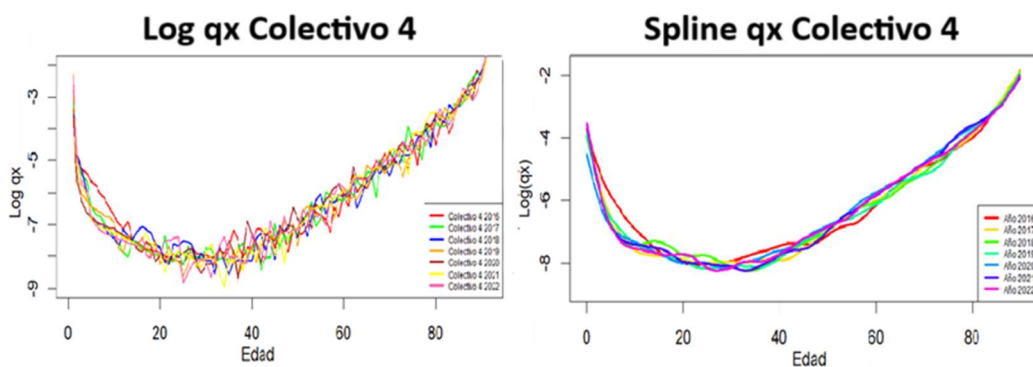


Gráfico 5. Probabilidad de fallecimiento para el Colectivo 1 real y suavizada.

Fuentes: Elaboración propia en base a datos INE.

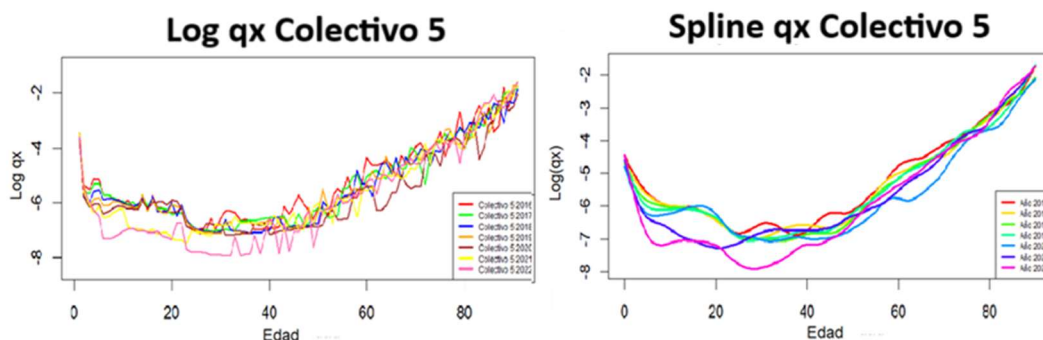


Gráfico 6. Probabilidad de fallecimiento para el Colectivo 1 real y suavizada.

Fuente: Elaboración propia en base a datos INE.

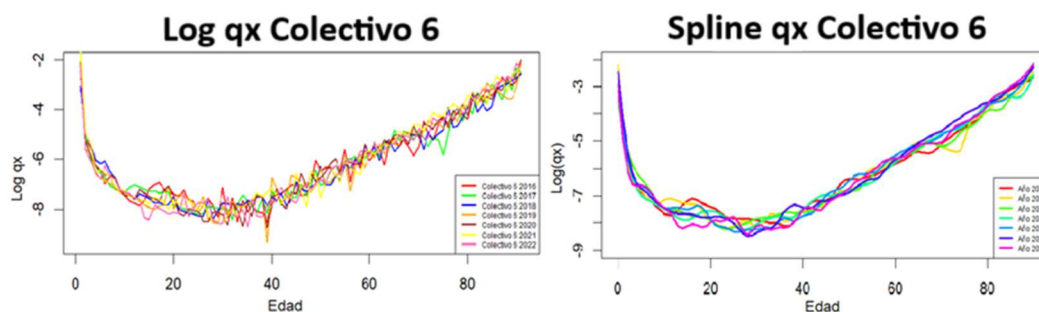


Gráfico 7. Probabilidad de fallecimiento para el Colectivo 1 real y suavizada.

Fuente: Elaboración propia en base a datos INE.

Las curvas suavizadas destacan como una herramienta fundamental para visualizar y captar tendencias, eliminando el ruido asociado a las tasas reales logradas mediante el desagregado de la población en colectivos y facilitando el análisis comparativo entre grupos.

Es importante destacar que, para los fines de este trabajo, el aspecto más relevante en este apartado no radica en si la mortalidad aumenta o disminuye en términos absolutos, sino en la estabilidad de las tasas de mortalidad calculadas a lo largo de todo el rango etario. Esta estabilidad y reducción de fluctuaciones, es especialmente relevante al evaluar el proceso de convergencia entre los distintos colectivos y la población española.

4.2. Ajuste del modelo Lee – Carter y Tasas de mortalidad proyectadas para españoles y los distintos colectivos.

En este apartado se presentan los resultados del ajuste del modelo Lee-Carter y las tasas de mortalidad proyectadas para la población española y los distintos colectivos. El modelo Lee-Carter ha sido previamente definido y ajustado utilizando datos históricos de mortalidad suavizados, permitiendo descomponer las tasas de mortalidad observadas en componentes principales, capturando tanto las tendencias a largo plazo como las fluctuaciones a corto plazo.

El objetivo principal en esta sección es predecir las tasas de mortalidad a 15 años para la población española y los distintos colectivos. Estas proyecciones permitirán analizar si las tasas de mortalidad de los colectivos convergen hacia las tasas de mortalidad de la población española.

A continuación, expondremos los 3 parámetros principales del modelo Lee – Carter, de los cuales dos están relacionados con la edad (a_x y b_x) y uno con el tiempo (k_t) y las tasas de fallecimiento proyectadas a 15 años, tanto para los españoles como para los 6 colectivos analizados. Posteriormente analizaremos e interpretaremos los resultados observados.

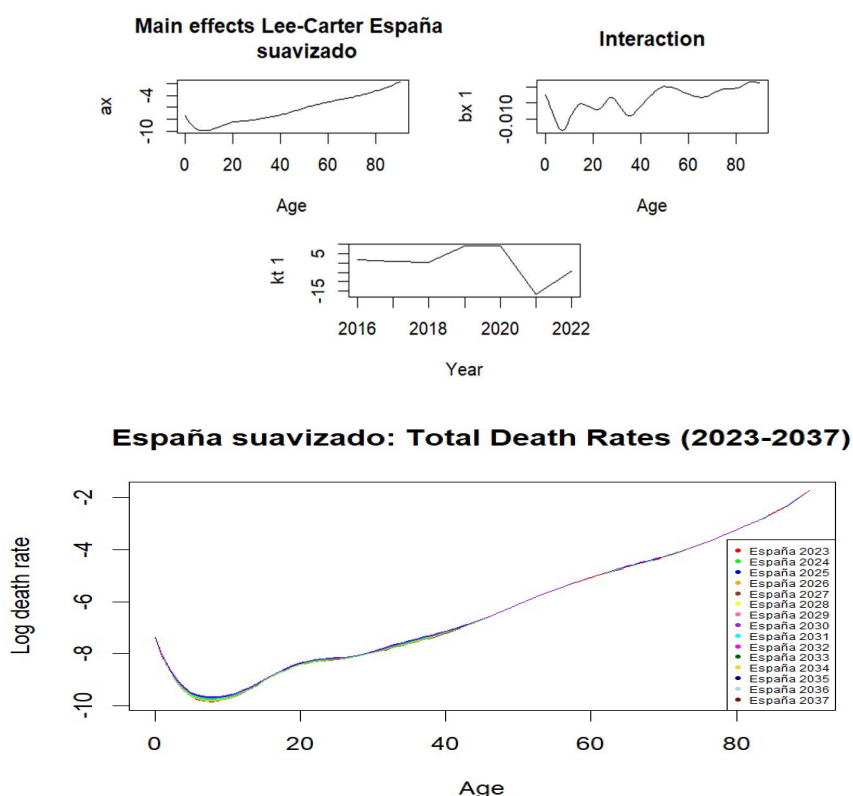


Gráfico 8. Modelo Lee-Carter para españoles nativos y probabilidades de fallecimiento en escala logarítmica proyectadas.

Fuente: Elaboración propia.

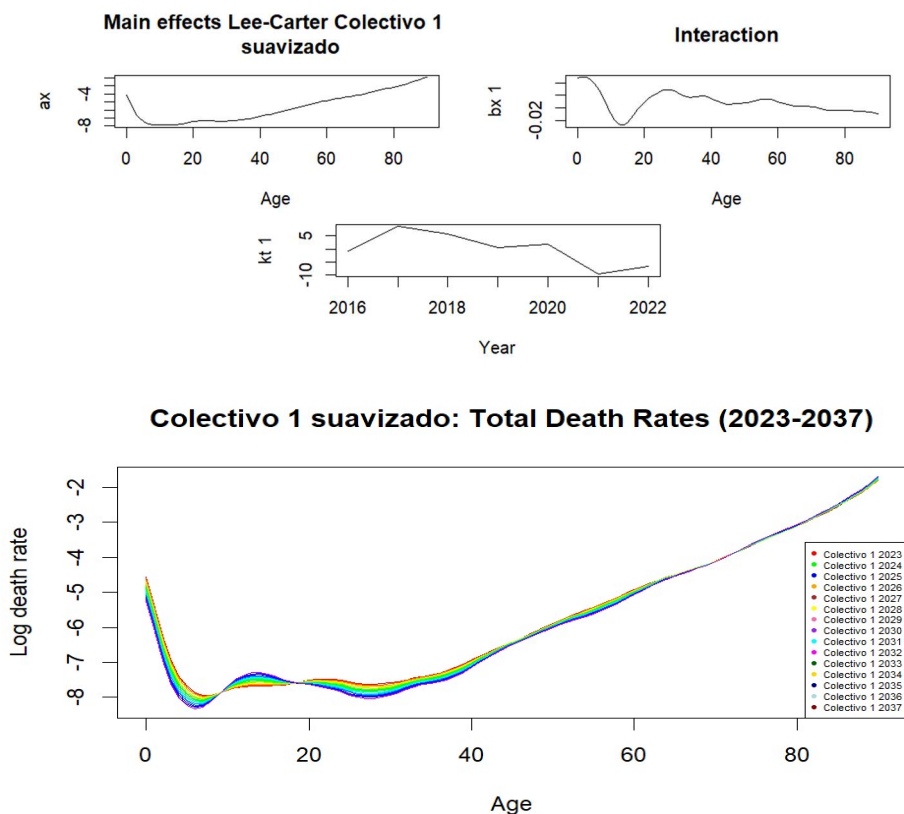
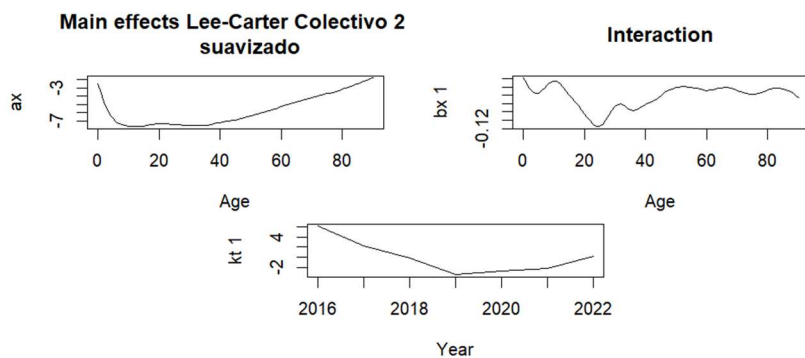


Gráfico 9. Gráfico 8. Modelo Lee-Carter para el Colectivo 1 y probabilidades de fallecimiento en escala logarítmica proyectadas.

Fuente: Elaboración propia.



Colectivo 2 suavizado: Total Death Rates (2023-2037)

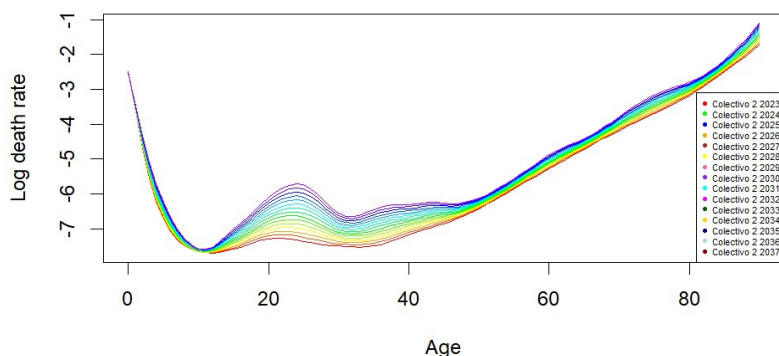
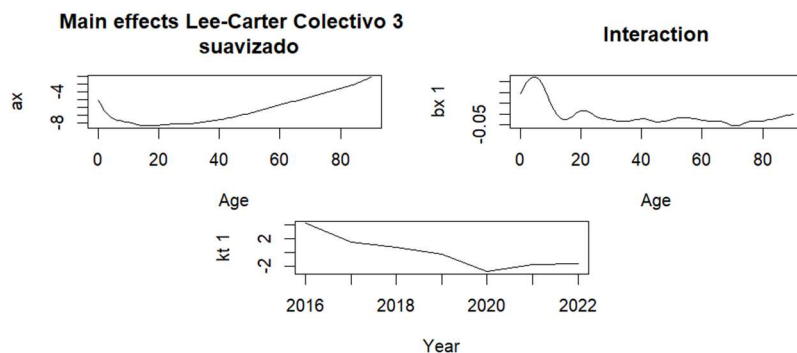


Gráfico 10. Modelo Lee-Carter para el Colectivo 2 y probabilidades de fallecimiento en escala logarítmica proyectadas.

Fuente: Elaboración propia.



Colectivo 3 suavizado: Total Death Rates (2023-2037)

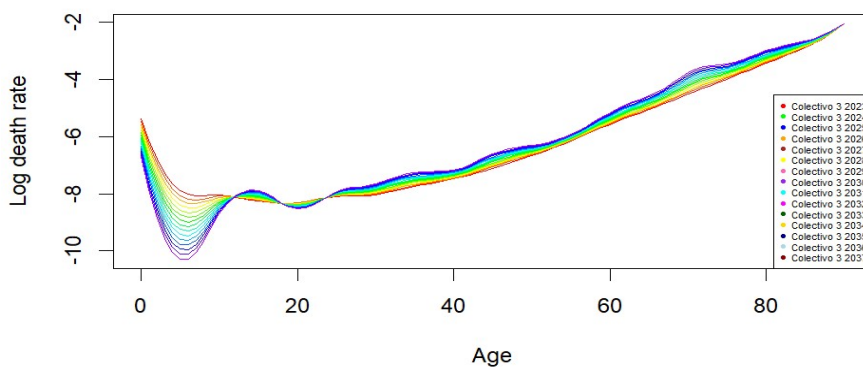


Gráfico 11. Modelo Lee-Carter para el Colectivo 3 y probabilidades de fallecimiento en escala logarítmica proyectadas.

Fuente: Elaboración propia.

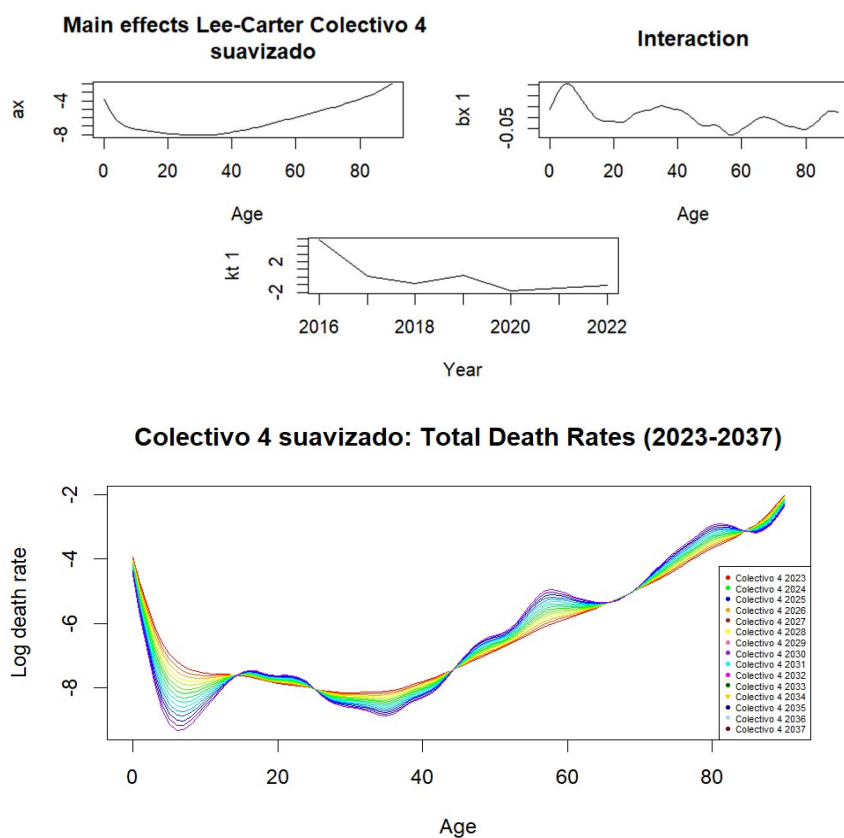
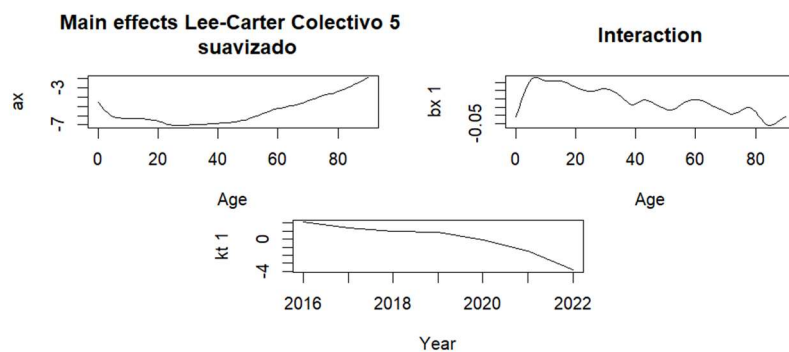


Gráfico 12. Modelo Lee-Carter para el Colectivo 4 y probabilidades de fallecimiento en escala logarítmica proyectadas.

Fuente: Elaboración propia.



Colectivo 5 suavizado: Total Death Rates (2023-2037)

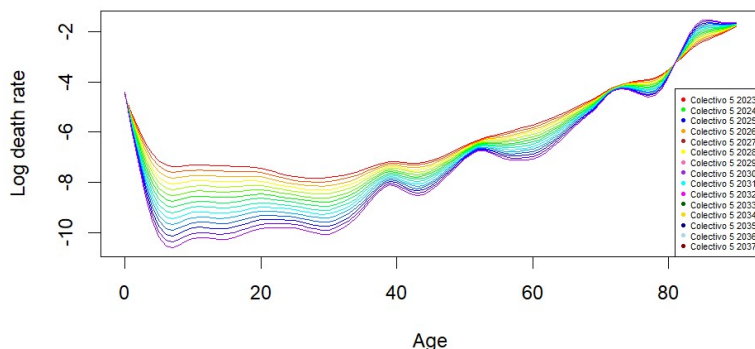
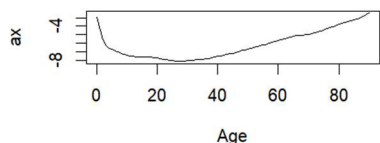


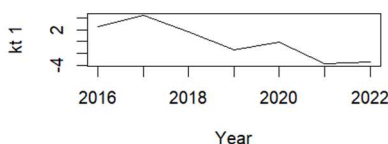
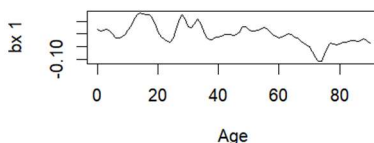
Gráfico 13. Modelo Lee-Carter para el Colectivo 5 y probabilidades de fallecimiento en escala logarítmica proyectadas.

Fuente: Elaboración propia

Main effects Lee-Carter Colectivo 6 suavizado



Interaction



Colectivo 6 suavizado: Total Death Rates (2023-2037)

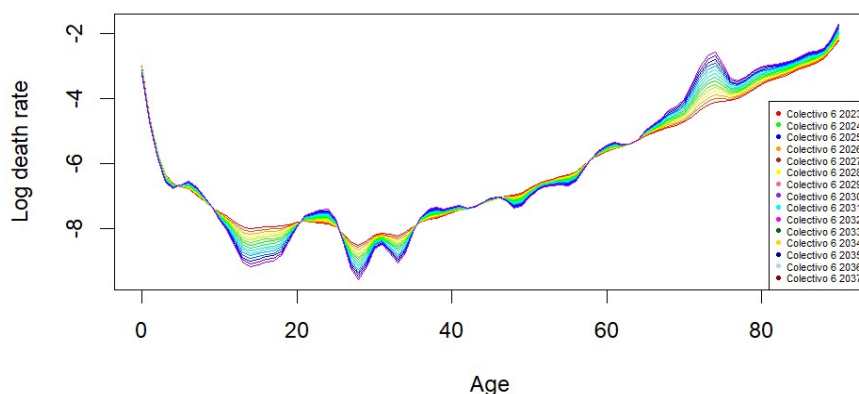


Gráfico 14. Modelo Lee-Carter para el Colectivo 6 y probabilidades de fallecimiento en escala logarítmica proyectadas.

Fuente: Elaboración propia.

En primer lugar, podemos analizar los efectos de la mortalidad exclusivos para cada una de las edades. Para ello, utilizamos el coeficiente a_x proporcionado por el modelo Lee-Carter. Como se aprecia en los gráficos, las curvas de mortalidad en escala logarítmica se presentan suavizadas, tanto para los españoles como para los distintos colectivos analizados.

Estas tasas están influenciadas por otros factores en consonancia con la ecuación del modelo. En este sentido, podemos observar el comportamiento de los b_x , que representan las constantes asociadas a la edad, revelando la sensibilidad de la tasa de mortalidad a los cambios en el tiempo. Para los españoles y los colectivos 1, 2, 3 y 4, los b_x presentan fluctuaciones hasta los 40 años, seguidas de cierta estabilización; sin embargo, sus valores permanecen cercanos a 0, evidenciando que no afectan significativamente al modelo. Por otro lado, en los colectivos 5 y 6, el componente b_x muestra fluctuaciones a lo largo de todo el rango etario, aunque también con valores muy próximos a 0.

Por último, analizamos el componente k_t , que refleja cómo cambian las tasas de mortalidad en todas las edades a medida que transcurre el tiempo. Un valor decreciente de k_t indica una mejora general en la mortalidad, lo cual puede observarse en cada uno de los subgrupos analizados.

Utilizando el modelo Lee-Carter, se proyectaron las q_x a 15 años, generando estimaciones futuras de las tasas de mortalidad para cada grupo analizado. Estas proyecciones serán fundamentales en los próximos apartados, donde se estudiará si las tasas de mortalidad de los distintos colectivos convergen hacia las correspondientes a la población española, permitiendo así evaluar posibles procesos de convergencia a largo plazo.

4.3. Divergencia Kullback-Leibler (KL).

En esta sección, presentamos los resultados del análisis de la divergencia Kullback-Leibler (KL) entre las tasas de mortalidad calculadas y proyectadas para los diferentes colectivos y las tasas de mortalidad esperadas basadas en la distribución de mortalidad española. La divergencia KL es una medida que cuantifica la diferencia entre dos distribuciones de probabilidad.

Definimos dos distribuciones de probabilidad clave:

- (P): Representa las tasas de mortalidad estimadas para los diferentes colectivos (Colectivo 1 al Colectivo 6). Estos colectivos son los subgrupos específicos de la población bajo estudio.
- (Q): Representa las tasas de mortalidad estimadas para la población española nativa. Esta distribución sirve como nuestra referencia o línea base, reflejando los patrones generales de mortalidad de los españoles.

Al comparar estas dos distribuciones utilizando la divergencia KL, evaluamos el grado de convergencia o divergencia entre los colectivos y la población española. Un valor de divergencia KL cercano a 0 indica una fuerte similitud, mientras que valores mayores

sugieren diferencias significativas. Por lo tanto, nuestro análisis tiene como objetivo comprender cuán estrechamente se alinean los colectivos con los patrones de mortalidad observados en España.

A continuación, presentamos los resultados obtenidos en nuestro análisis, acompañados de una tabla con los datos relevantes y un gráfico que ilustra las tendencias observadas. Estos elementos proporcionan una visión completa de la convergencia o divergencia de los colectivos de mortalidad con respecto a la distribución de mortalidad española.

Año Proyectado	Colectivo_1	Colectivo_2	Colectivo_3	Colectivo_4	Colectivo_5	Colectivo_6
1	0,0305694555	0,3054402097	0,0187232971	0,1329630417	0,0597011235	0,2869528698
2	0,0282949471	0,2951476330	0,0164519845	0,1213464428	0,0614325175	0,2721700980
3	0,0261889755	0,2852935426	0,0149430973	0,1115504971	0,0656438579	0,2587282096
4	0,0242433692	0,2758742722	0,0141012199	0,1035215340	0,0719322119	0,2466403624
5	0,0224503824	0,2668867059	0,0138472628	0,0972114250	0,0799632026	0,2359235457
6	0,0208026753	0,2583283195	0,0141157929	0,0925761679	0,0894592840	0,2265985631
7	0,0192932947	0,2501972244	0,0148527763	0,0895746059	0,1001897761	0,2186899771
8	0,0179156553	0,2424922144	0,0160136737	0,0881672847	0,1119625075	0,2122260109
9	0,0166635216	0,2352128148	0,0175618387	0,0883154498	0,1246168779	0,2072384041
10	0,0155309910	0,2283593352	0,0194671729	0,0899801854	0,1380181481	0,2037622147
11	0,0145124771	0,2219329263	0,0217049989	0,0931216923	0,1520527694	0,2018355657
12	0,0136026940	0,2159356389	0,0242551172	0,0976987044	0,1666245826	0,2014993297
13	0,0127966409	0,2103704881	0,0271010178	0,1036680374	0,1816517352	0,2027967483
14	0,0120895880	0,2052415210	0,0302292231	0,1109842660	0,1970641872	0,2057729839
15	0,0114770622	0,2005538889	0,0336287389	0,1195995222	0,2128016937	0,2104746006

Tabla 2. Valores de KL para cada año proyectado y para cada colectivo analizado. Fuente: Elaboración propia.

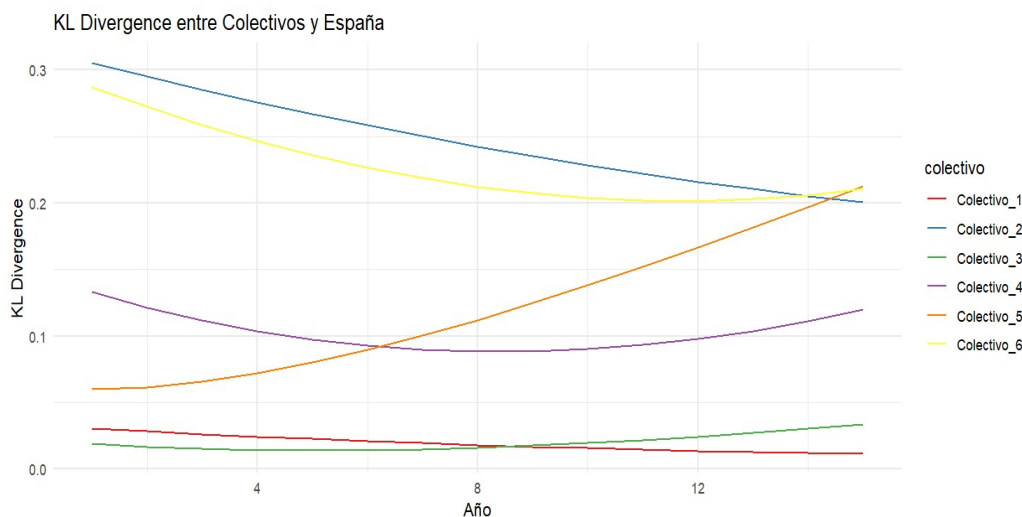


Gráfico 15. Valores de KL para cada año proyectado y para cada colectivo analizado.

Fuente: Elaboración propia.

En general, los resultados indican que los distintos colectivos convergen hacia la mortalidad española, evidenciada por valores de divergencia KL cercanos a 0. Específicamente, los colectivos 1 y 3 exhiben los valores mínimos de divergencia. Sin



embargo, dentro del mismo contexto convergente, se observa que el colectivo 5 presenta una tendencia creciente, a diferencia de los otros colectivos, lo que sugiere una posible divergencia a futuro. Por otro lado, los colectivos 2 y 6 presentan una mayor divergencia con valores de KL más elevados, aunque con una tendencia a la baja. Estos resultados reflejan un proceso generalizado de convergencia, con pequeñas diferencias que destacan ciertos matices entre los colectivos en el tiempo.

4.4. Primas para Productos de Vida-Ahorro utilizando probabilidades de fallecimientos proyectadas.

En esta sección se presentan los resultados del cálculo de las primas para un producto de Vida-Ahorro, descrito previamente en el apartado 2.2 del presente trabajo.

Estos cálculos se basan en las probabilidades de fallecimiento proyectadas mediante el modelo Lee-Carter. Las primas han sido estimadas utilizando las q_x proyectadas a 15 años, tanto para cada colectivo como para la población española, permitiendo analizar las diferencias e impactos en los costos asociados a este producto entre los distintos grupos.

El análisis se centra en evaluar las variaciones en las primas calculadas, identificando cómo los riesgos específicos de cada grupo influyen en los costos del producto. Asimismo, dado que el objetivo principal es determinar la existencia de convergencia en las tasas de mortalidad, este análisis examinará si dichas tendencias también se reflejan en las primas calculadas.

A continuación, se presenta una tabla que muestra las primas calculadas para cada grupo de estudio a lo largo de un horizonte temporal de 15 años. La tabla incluye los valores estimados para cada año proyectado, permitiendo una comparación detallada entre los diferentes colectivos y la población española:

Año Proyectado	España	Colectivo_1	Colectivo_2	Colectivo_3	Colectivo_4	Colectivo_5	Colectivo_6
1	47.104,52	47.439,99	46.842,62	46.441,84	46.139,26	46.278,03	46.335,61
2	47.108,78	47.421,83	46.884,68	46.477,91	46.176,22	46.191,77	46.344,40
3	47.113,06	47.403,83	46.927,87	46.515,10	46.214,99	46.112,99	46.353,29
4	47.117,34	47.385,99	46.972,21	46.553,44	46.255,66	46.041,04	46.362,28
5	47.121,64	47.368,31	47.017,74	46.592,98	46.298,33	45.975,34	46.371,39
6	47.125,94	47.350,78	47.064,48	46.633,75	46.343,11	45.915,35	46.380,61
7	47.130,26	47.333,41	47.112,47	46.675,78	46.390,11	45.860,58	46.389,95
8	47.134,58	47.316,19	47.161,74	46.719,11	46.439,45	45.810,57	46.399,39
9	47.138,92	47.299,12	47.212,32	46.763,79	46.491,26	45.764,92	46.408,96
10	47.143,26	47.282,20	47.264,24	46.809,85	46.545,67	45.723,24	46.418,64
11	47.147,62	47.265,43	47.317,55	46.857,34	46.602,82	45.685,19	46.428,44
12	47.151,98	47.248,81	47.372,27	46.906,30	46.662,85	45.650,46	46.438,36
13	47.156,36	47.232,34	47.428,44	46.956,77	46.725,93	45.618,76	46.448,40
14	47.160,75	47.216,01	47.486,10	47.008,81	46.792,22	45.589,82	46.458,56
15	47.165,14	47.199,82	47.545,29	47.062,45	46.861,89	45.563,40	46.468,85

Tabla 3. Valores de primas estimadas para cada año proyectado y por grupo analizado.

Fuente: Elaboración propia

Para complementar el análisis utilizaremos el Boxplot que es una herramienta eficaz para lograr comparar la distribución de primas entre España y los distintos colectivos, ya que proporciona una representación clara de la variabilidad y las tendencias centrales de los

datos. En este caso, se analizan los montos promedio de primas calculados a 15 años, utilizando las tasas proyectadas para los diferentes grupos estudiados.

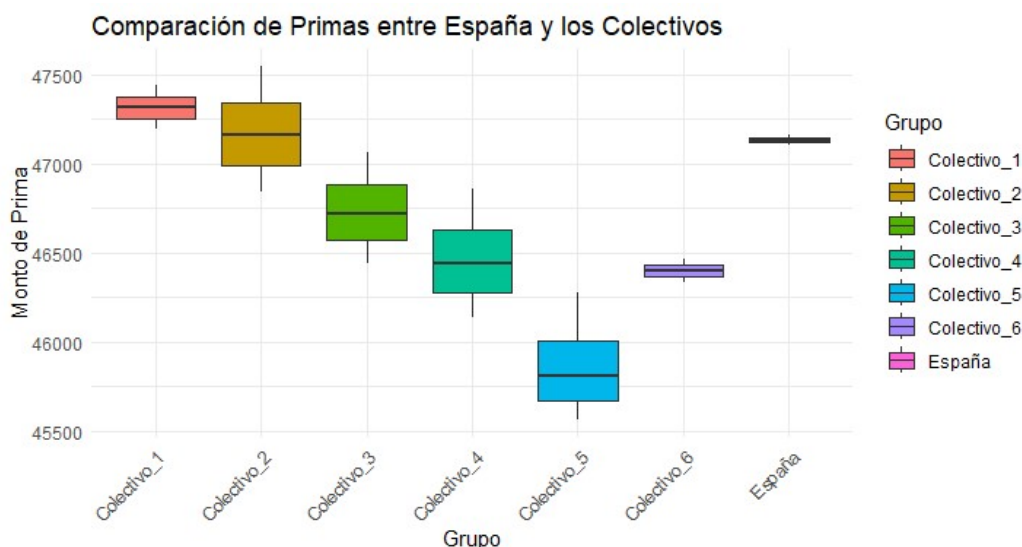


Gráfico 16. Boxplot comparativo entre valores de primas calculadas.

Fuente: Elaboración propia.

Las primas correspondientes a la población española presentan una menor variabilidad, lo que indica estabilidad en los valores calculados. En términos generales, las primas de los colectivos muestran diferentes niveles en relación con las de España, lo que refleja las particularidades de los riesgos proyectados para cada grupo.

Los colectivos 1 y 2 presentan primas promedio cercanas a las de España, con diferencias marginales. Esto sugiere que los riesgos de mortalidad proyectados para estos grupos convergen hacia los de la población española. En contraste, el colectivo 5 muestra primas promedio considerablemente más bajas que las de España, acompañadas de una mayor dispersión, lo que podría indicar cierta divergencia. Por otro lado, el colectivo 6 presenta primas promedio también inferiores, aunque con menor variabilidad, lo que refleja estabilidad en sus valores, pero cierta distancia respecto a España.

En la mayoría de los colectivos se observan diferencias en las primas promedio con respecto a España, lo que puede estar influido por las particularidades de las tasas de mortalidad proyectadas para cada grupo. Sin embargo, estas diferencias tienden a ser más pronunciadas en algunos colectivos (como el 5 y el 6) y más sutiles en otros (como el 1 y el 2).

En el análisis anterior se observó las diferencias entre los resultados obtenidos en los cálculos de primas utilizando las q_x españolas proyectadas y las de los distintos colectivos también proyectadas. No obstante, para que esta información sea más útil en nuestro estudio de convergencia, es de suma importancia analizar si esas diferencias se acentúan en el tiempo o disminuyen. Para ello, analizaremos la tabla presentada a continuación, muestra las diferencias en primas, SCR de mortalidad y SCR de longevidad entre cada

colectivo y la población española, calculadas utilizando las q_x proyectadas a 5, 10 y 15 años.

Estas diferencias permiten observar la convergencia de las q_x de los colectivos con las de los españoles, analizando el comportamiento a lo largo del horizonte temporal.

Grupo	Prima 5 Años	Prima 10 Años	Prima 15 años	SCR Mortalidad 5 Años	SCR Mortalidad 10 Años	SCR Mortalidad 15 años	SCR Longevidad 5 Años	SCR Longevidad 10 Años	SCR Longevidad 15 años
España	49.292,84 €	47.143,26 €	47.165,14 €	535,56 €	252,78 €	255,73 €	-721,78 €	-338,68 €	-342,66 €
Colectivo 1	47.368,31 €	47.282,20 €	47.199,82 €	283,21 €	271,85 €	260,96 €	-379,68 €	-364,36 €	-349,69 €
Diferencia	-1.924,53 €	138,94 €	34,68 €	-252,35 €	19,07 €	5,22 €	342,10 €	-25,68 €	-7,03 €
Colectivo 2	47.017,74 €	47.264,24 €	47.545,29 €	236,36 €	269,39 €	306,87 €	-316,57 €	-361,05 €	-411,59 €
Diferencia	-2.275,10 €	120,98 €	380,15 €	-299,19 €	16,61 €	51,13 €	405,21 €	-22,37 €	-68,93 €
Colectivo 3	46.592,98 €	46.809,85 €	47.062,45 €	179,04 €	208,52 €	242,74 €	-239,54 €	-279,13 €	-325,15 €
Diferencia	-2.699,86 €	-333,41 €	-102,69 €	-356,51 €	-44,26 €	-13,00 €	482,25 €	59,55 €	17,50 €
Colectivo 4	46.298,33 €	46.545,67 €	46.861,89 €	138,16 €	170,88 €	212,38 €	-184,69 €	-228,59 €	-284,33 €
Diferencia	-2.994,51 €	-597,59 €	-303,25 €	-397,40 €	-81,90 €	-43,35 €	537,09 €	110,09 €	58,33 €
Colectivo 5	45.975,34 €	45.723,24 €	45.563,40 €	94,79 €	60,29 €	38,32 €	-126,61 €	-80,48 €	-51,13 €
Diferencia	-3.317,50 €	-1.420,02 €	-1.601,74 €	-440,76 €	-192,49 €	-217,42 €	595,17 €	258,20 €	291,53 €
Colectivo 6	46.371,39 €	46.418,64 €	46.468,85 €	148,47 €	154,71 €	161,34 €	-198,51 €	-206,89 €	-215,79 €
Diferencia	-2.921,45 €	-724,62 €	-696,29 €	-387,09 €	-98,07 €	-94,39 €	523,27 €	131,79 €	126,87 €

Tabla 4. Primas, SCR mortalidad y SCR longevidad de cada subgrupo para 5, 10 y 15 años.

Fuente: Elaboración propia

En cuanto a las primas, se observa que las diferencias son más pronunciadas en el corto plazo, es decir, a 5 años. Por ejemplo, el Colectivo 5 presenta una diferencia de -3.317,50 € en este horizonte. Sin embargo, a medida que se extiende el plazo a 10 y 15 años, las diferencias tienden a reducirse de forma notable. Este comportamiento es evidente en el caso del Colectivo 1, donde la diferencia inicial de -1.924,53 € a 5 años se reduce a 138,94 € a 10 años y, finalmente, a 34,68 € a 15 años. En el Colectivo 3, la diferencia, que es de -2.699,86 € a 5 años, también se minimiza con el tiempo, llegando a -333,41 € a 10 años y a -102,69 € a 15 años. Este patrón sugiere una convergencia progresiva de las primas de los colectivos hacia las de la población española.

En el caso del SCR de mortalidad, se aprecia una tendencia similar. Las diferencias iniciales a 5 años son más significativas, como ocurre en el Colectivo 1, que presenta una diferencia de -252,35 €. Sin embargo, esta diferencia se reduce considerablemente a 19,07 € a 10 años y a 5,22 € a 15 años. Un comportamiento análogo se observa en otros colectivos, como el Colectivo 2, donde las diferencias también disminuyen sustancialmente con el aumento del horizonte temporal. Esto refleja cómo las probabilidades de fallecimiento proyectadas para los colectivos tienden a alinearse más estrechamente con las de la población española en periodos más largos.

Por otro lado, en el SCR de longevidad, también se evidencia una reducción significativa de las diferencias a lo largo del tiempo. Por ejemplo, el Colectivo 4 muestra una diferencia de 537,09 € a 5 años, que disminuye a 110,09 € a 10 años y a 58,33 € a 15 años. En el Colectivo 6, las diferencias siguen una trayectoria similar, pasando de 523,27 € a 5 años, a 131,79 € a 10 años y finalmente a 126,87 € a 15 años. Este comportamiento destaca la



convergencia en las probabilidades de longevidad de los diferentes colectivos con respecto a la población española.

En términos generales, tanto las primas como los SCR de mortalidad y longevidad muestran una clara tendencia de reducción de las diferencias con el tiempo. Esto indica un proceso de convergencia progresiva entre las q_x proyectadas de los colectivos y las de la población española, especialmente en horizontes temporales más largos. Estos resultados son consistentes con la hipótesis planteada sobre la convergencia de las tasas de mortalidad entre los distintos colectivos y la población española nativa.

5. Conclusión.

El presente Trabajo ha analizado la evolución de la mortalidad en España, considerando tanto a la población española nativa como a distintos colectivos de población extranjera residente. El principal objetivo fue determinar si las tasas de mortalidad de los colectivos extranjeros convergen hacia las de la población española nativa, evaluando el impacto de la integración socioeconómica, educativa y sanitaria en este proceso. A lo largo del trabajo, se utilizaron herramientas avanzadas, como el modelo Lee-Carter, curvas suavizadas y la divergencia KL, para analizar y proyectar las tasas de mortalidad y examinar su comportamiento en distintos horizontes temporales.

El uso de las tasas de mortalidad suavizadas permitió una reducción del ruido asociado a las fluctuaciones de las tasas reales para los colectivos generadas por el desagregado poblacional, facilitando así una evaluación más precisa de las tendencias. Además, el comportamiento de los coeficientes del modelo Lee-Carter a_x , b_x y k_t , reflejó la estabilidad y los cambios en la mortalidad a lo largo del tiempo y las edades. Los valores de b_x cercanos a cero en la mayoría de los colectivos indicaron una baja sensibilidad a variaciones temporales, mientras que la disminución de los k_t confirmó una mejora general en la mortalidad en todos los grupos analizados.

Por otra parte, los resultados obtenidos mediante la divergencia KL evidenciaron un proceso general de convergencia entre los colectivos extranjeros y la población española. En términos cuantitativos, los colectivos 1 y 3 presentaron los valores más bajos de divergencia, lo que indica una alta similitud con los patrones de mortalidad de España. No obstante, el Colectivo 5 mostró una tendencia ligeramente divergente a futuro, aunque con valores que se mantienen cercanos a 0, mientras que los colectivos 2 y 6, a pesar de presentar valores más altos de divergencia, evidenciaron una trayectoria descendente. Este análisis subraya que, aunque existen matices y diferencias entre los colectivos, el patrón general apunta a un acercamiento progresivo hacia los niveles de mortalidad de la población española nativa.

En cuanto a las primas y los SCR de mortalidad y longevidad, se observó una clara tendencia de reducción de las diferencias con el paso del tiempo. A 5 años, las diferencias eran más marcadas; sin embargo, a 10 y 15 años, estas se redujeron de forma significativa. Este comportamiento respalda la hipótesis de convergencia, ya que las tasas proyectadas de los colectivos se alinearon progresivamente con las de la población española, mostrando una tendencia homogénea a largo plazo.



En conclusión, los resultados obtenidos confirman que las tasas de mortalidad de los colectivos extranjeros residentes en España tienden a converger con las de la población española nativa a medida que aumenta el horizonte temporal. Este proceso refleja los efectos positivos de la integración socioeconómica, educativa y sanitaria en el país, contribuyendo a la homogeneización del riesgo de longevidad para toda la población residente.



6. Bibliografía

- Atance, D. C. (2024). Convergence and divergence in mortality: A global study from 1990 to 2030. *PLoS ONE*, 19(1), e0295842.
<https://doi.org/10.1371/journal.pone.0295842>.
- Booth, H., & Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. *Annals of actuarial science*, 3(1-2):3-43.
- Carter, R. D. & Lee and L. R. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 659-671.
- Cortés, A. C. (11 de julio de 2024). *Generali aumenta un 12% los ingresos en el seguro para extranjeros residentes en España*. Obtenido de El Economista:
<https://www.eleconomista.es/banca-finanzas/noticias/12904711/07/24/generali-aumenta-un-12-los-ingresos-en-el-seguro-para-extranjeros-residentes-en-espana.html>
- Debón Aucejo, A., Montes, F., & Puig, F. (2008). Modelling and forecasting mortality in Spain. . *European Journal of Operational Research.*, 189(3):624-637.
- García González, J. M. (2014). ¿Por qué vivimos más? Descomposición por causa de la esperanza de vida española de 1980 a 2009. *Revista Española de Investigaciones Sociológicas*, 148: 39-60.
- García González, J. M. (2018). Evaluación de la calidad de las estadísticas españolas de mortalidad a edades avanzadas (1910-2014). *Boletín de la Asociación de Geógrafos Españoles*, 76, 455-478.
- Goerlich Gisbert, F. (2012). Esperanza de vida y causas de muerte: Un análisis de descomposición. (1975 - 2009). *Fundación BBVA*, Documento de trabajo 13.
- Guan, T. (10 de 7 de 2014). Multiple-Decrement Compositional Forecasting with the Lee-Carter Model. SIMON FRASER UNIVERSITY.
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 79 - 86.
- Macías, Y., & M., S. (4ª época, 24, 2018). APLICACIÓN DE LOS MODELOS LEE-CARTER Y RENSHAW-HABERMAN EN LOS SEGUROS DE VIDA Y MIXTOS. *Anales del Instituto de Actuarios Españoles*, , 53-78.
- Shyamalkumar, N. D. (2006). Analysis of Mortality Data using Smoothing Spline Poisson Regression. *Dept. of Stat. & Act. Sci., The University of Iowa*.
- Tabeau, E. (2001). A Review of Demographic Forecasting Models for Mortality. *European Studies of Population* , 1-32.
- Vanguardia, L. (3 de 5 de 2024). España vuelve a liderar la esperanza de vida en la Unión Europea, con 84 años de media. Madrid, España.

7. Anexos.

7.1. Principios básicos de biometría

En este apartado se detallarán los principales conceptos biométricos que sirven de base para la realización del presente trabajo. El objetivo es proporcionar al lector de un sustento teórico que facilite la comprensión de los procedimientos y resultados obtenidos a lo largo del estudio.

Para comenzar, definiremos tres conceptos que serán necesario para la determinación de los posteriores, estos son.

l_x : Número de individuos vivos a la edad x .

l_ω : Número de individuos vivos a la edad ω , siendo ω la edad límite del colectivo. Por definición, al final del año no hay supervivientes de esa edad, $l_\omega = 0$.

d_x : Número de individuos fallecidos a la edad x (durante el intervalo $x, x + 1$). Es la diferencia entre el número de individuos vivos a la edad x y el número de individuos vivos a la edad $x + 1$.

Relacionando los conceptos detallados previamente, obtenemos:

$$d_x = l_x - l_{x+1}, \text{ donde además, } l_x = \sum_{t=x}^{\omega-1} d_t ; l_0 = \sum_{t=0}^{\omega-1} d_t$$

7.1.1. Probabilidades de supervivencia sobre un individuo.

- Probabilidad de que un individuo de edad x alcance la edad $x+1$, o sobreviva 1 año más:

$$p_x = \frac{l_{x+1}}{l_x}$$

- Probabilidad de vivir a la edad límite, o de que un individuo de edad $(\omega - 1)$ alcance la edad ω :

$$p_{\omega-1} = \frac{l_\omega}{l_{\omega-1}} = 0$$

- Probabilidad de que un individuo de edad x alcance la edad $x + n$, esto es, que viva n años más:

$$p_x = \frac{l_{x+n}}{l_x}$$

7.1.2. Probabilidad de fallecimiento

- Probabilidad de que un individuo de edad x no alcance con vida la edad $x + 1$:

$$q_x = \frac{d_x}{l_x} = \frac{l_x - l_{x+1}}{l_x} = 1 - p_x$$

Donde la relación entre p_x y q_x es

$$q_x + p_x = \frac{d_x}{l_x} + \frac{l_{x+1}}{l_x} = \frac{l_x - l_{x+1} + l_{x+1}}{l_x} = 1$$

- Tanto de mortalidad diferido o probabilidad de que un individuo de edad x fallezca exactamente a la edad $x + n - 1$ años:

$${}_{n-1}/q_x = \frac{d_{x+n-1}}{l_x} = \frac{l_{x+n-1} - l_{x+n}}{l_x}$$

La expresión más utilizada para esta probabilidad, es equivalente a decir que un individuo de edad x sobreviva $n - 1$ años y fallezca en el año n :

$${}_{n-1}/q_x = \frac{d_{x+n-1}}{l_x} = \frac{l_{x+n-1}}{l_x} \frac{d_{x+n-1}}{l_{x+n-1}} = {}_{n-1}/p_x q_{x+n-1}$$

- Tanto de mortalidad temporal o Probabilidad de que un individuo de edad x fallezca en el transcurso de los n próximos años:

$${}_nq_x = 1 - {}_np_x = \frac{l_x - l_{x+n}}{l_x} = 1 - {}_np_x$$

7.1.3. Extensión de los conceptos biométricos según origen del colectivo analizado.

Se considerará el conjunto $\Omega = \{i_1, i_2, i_3, \dots, i_N\}$ como el conjunto de lugares de origen de los colectivos definidos como variables de estudio. Este conjunto representa los distintos orígenes geográficos o demográficos que se utilizarán para analizar las características y comportamientos de los colectivos bajo análisis. Es importante destacar que, para el propósito de este trabajo, se asume que existe independencia entre los elementos de Ω , lo cual se formaliza como ($\text{corr}\{i_1, i_2, i_3, \dots, i_N\} = 0$).

Esta suposición de independencia implica que las variables que representan los diferentes lugares de origen no están correlacionadas entre sí. Esta hipótesis es clave para simplificar el modelo y asegurar que las conclusiones derivadas del análisis se mantengan válidas bajo el supuesto de que los colectivos procedentes de diferentes lugares de origen no presentan una relación directa entre ellos.

Redefiniendo los conceptos tendremos que:

$$l_x = l_x^1 + l_x^2 + l_x^3 + \dots + l_x^N =$$

$$= \sum_{i=1}^N l_x^i$$

Donde l_x^i es el número de individuos vivos a la edad x del colectivo cuyo origen es i .

$$d_x = d_x^1 + d_x^2 + d_x^3 + \dots + d_x^N =$$

$$= \sum_{i=1}^N d_x^i$$

Donde d_x^i es el número de individuos fallecidos a la edad x del colectivo cuyo origen es i .

Asimismo, resulta relevante detallar las probabilidades de supervivencia y mortalidad asociadas a cada colectivo en estudio, ya que estas métricas son fundamentales para comprender la dinámica poblacional y los riesgos a los que se enfrenta cada grupo.

Estas probabilidades permiten evaluar el comportamiento de la mortalidad en función del origen geográfico, y las condiciones socioeconómicas a las que estuvieron expuesto.

- Probabilidad de que un individuo de edad x y perteneciente al colectivo cuyo origen es i alcance la edad $x+1$, o sobreviva 1 año más:

$$p_x^i = \frac{l_{x+1}^i}{l_x^i}$$

- Probabilidad de que un individuo de edad x y perteneciente al colectivo cuyo origen es i alcance la edad $x+n$, esto es, que viva n años más:

$${}_n p_x^i = \frac{l_{x+n}^i}{l_x^i}$$

- Probabilidad de que un individuo de edad x y perteneciente al colectivo cuyo origen es i no alcance con vida la edad $x + 1$:

$$q_x^i = \frac{d_x^i}{l_x^i} = \frac{l_x^i - l_{x+1}^i}{l_x^i} = 1 - p_x^i$$

- Tanto de mortalidad temporal o Probabilidad de que un individuo de edad x y perteneciente al colectivo cuyo origen es i fallezca en el transcurso de los n próximos años:

$${}_n q_x^i = 1 - {}_n p_x^i = \frac{l_x^i - l_{x+n}^i}{l_x^i} = 1 - {}_n p_x^i$$

7.2. Cálculo de k_t y b_x utilizando SVD.

A. Preparación de los datos

Primero, necesitas las tasas de mortalidad $m_{x,t}$ para diferentes edades x y diferentes años t . Estas tasas se transforman tomando el logaritmo:

$$\log m_{x,t}$$

B. Cálculo de a_x

Calcula el valor promedio de $\log m_{x,t}$ a través del tiempo para cada edad x :

$$a_x = \frac{1}{T} \sum_{t=1}^T \log m_{x,t}$$

Aquí, T es el número total de años en tu conjunto de datos. Esto te da una serie de valores a_x que representan el perfil promedio de mortalidad específico de la edad.

C. Centrando los datos

Resta a_x de cada $\log m_{x,t}$ para centrar los datos. Esto se hace para cada edad x :

$$\log m_{x,t} - a_x$$

Vamos a llamar a la matriz resultante M . Cada elemento $M_{x,t}$ de esta matriz representa:

$$M_{x,t} = \log m_{x,t} - a_x$$

D. Aplicación de SVD

Aplica la descomposición en valores singulares (SVD) a la matriz M :

$$M = U \Sigma V^t$$

En esta descomposición:

- U es una matriz cuyas columnas son los vectores singulares izquierdos.
- Σ es una matriz diagonal cuyos elementos son los valores singulares.
- V^t es la transpuesta de la matriz V , que contiene los vectores singulares derechos.

E. Extracción de k_t y b_x .

Selecciona el primer vector singular de U (correspondiente al mayor valor singular) como b_x :

$$b_x = U[:, 1]$$



Selecciona el primer vector singular de V (correspondiente al mayor valor singular) como k_t :

$$k_t = V[:, 1] \cdot \Sigma[1, 1]$$

Multiplica el primer valor singular (el elemento diagonal de Σ) por el primer vector singular de V para obtener k_t .