



Máster Universitario en Ciencias Actuariales y Financieras  
(2017-2018)

*Trabajo Fin de Máster*

Proyección biométrica de la población española  
a corto y largo plazo: Modelo ARIMA con  
restricción vs Lee-Carter dinámico en R

Autor: Tuyuan Guo

Tutores:

José Miguel Rodríguez – Pardo Castillo

Jesús Simón del Potro

Madrid 3 de julio de 2018



Esta tesis es propiedad del autor. No está permitida la reproducción total o parcial de este documento sin mencionar su fuente. El contenido de este documento es de exclusiva responsabilidad del autor, quien declara que no ha incurrido en plagio y que la totalidad de referencias a otros autores han sido expresadas en el texto.

En caso de obtener una calificación igual o superior a 8.0 Notable, autorizo la publicación de este trabajo en el centro de Documentación de la Fundación Mapfre.

- Sí, autorizo a su publicación.
- No, desestimo su publicación.

Fdo: Tuyuan Guo



## Resumen:

En el trabajo se emplea los datos de mortalidad de la población española desde 1908 hasta finales del siglo XX, para intentar predecir la tabla de mortalidad de 1,5,10 y 15 años.

El análisis de estas tablas se lleva a cabo mediante los mapas de calor. Analizando exhaustivamente los índices de mejora y la esperanza de vida por tramos. Todo ello para analizar diversos comportamientos de la mortalidad y detectar posible mejora de cohorte.

El trabajo se centra en la previsión de los modelos ARIMA, teniendo como base tres muestras distintas. En la primera muestra el trabajo se focalizará en la serie de la tasa de mortalidad desde 1943 a finales del siglo xx, la segunda muestra se focaliza en la muestra desde 1960 hasta finales del siglo xx, logrando ambos resultados realistas a corto plazo. En el tercero modelo se emplea todos los datos históricos asignando mayor peso a los datos más recientes, y se establece cierta similitud en los parámetros de las edades colindantes.

Adicionalmente se realiza los modelos Lee-Carter con parámetros evolutivos, y se combina la previsión con la de modelo ARIMA para lograr mayor precisión en la predicción a largo plazo.

Con este trabajo se pretende realizar una comparativa de distintas estrategias de previsión, teniendo en cuenta las tablas de mortalidades de distintos plazos. Todo ello para que se pueda aplicar este modelo no solo a la población española sino a todo tipo de poblaciones.



## Abstract:

This thesis tries to predict the death rate in 1, 5, 10, 15 years' time by using the mortality rate of the Spanish Population from 1908 up to the end of the 20th century.

It uses heatmaps to evaluate the mortality rate using an exhaustive analysis of the development indicators and life expectancy. All of this goes into analysing the various behaviours of the mortality rate and detecting any possible improvements that can be made.

This work revolves around the ARIMA forecasting models and uses 3 different samples as the baseline. The first sample will focus on the death rate since 1943 up until the end of the 20th century whilst the second sample will start from 1960 up to the end of the 20th century. Both this samples convey realistic results short term. The final sample will use all the available historic data but assigning more value to recent data, and it establishes a certain similitude in neighbouring years

Additionally, it uses the Lee-Carter model with developmental parameters combined with the forecast from the ARIMA model to achieve a greater precision on the long run.

All in all, this thesis revolves around the analysis of data and models from which it makes a comparison in the different prediction strategies whilst using different mortality tables in different periods. The goal is to make a model that can be applied not only to the Spanish population but to all kind of populations.

# Índice

<b>1</b>	<b><i>Introducción</i></b> .....	<b>8</b>
<b>2</b>	<b><i>Marco teórico</i></b> .....	<b>10</b>
2.1	Modelo Lee-Carter .....	10
2.1.1	Elección de número de años de muestra, la regla de 30 años. ....	12
2.1.2	Problema de posición: análisis de varianza .....	14
2.2	Modelo de ARIMA .....	15
2.2.1	ARIMA con distintos pesos .....	15
2.2.2	Regresión con peso y funciones de peso de este trabajo.....	16
<b>3</b>	<b><i>Presentación de datos</i></b> .....	<b>18</b>
<b>4</b>	<b><i>Análisis de mejora generacional cohorte</i></b> .....	<b>19</b>
4.1	Estudio de fenómeno cohorte mediante mapa de calor .....	20
4.2	enfoque alternativo: Series de esperanza de vida por tramo.....	26
4.3	Conclusión sobre la mejora generacional cohorte.....	31
<b>5</b>	<b><i>Análisis de dato y Modelación, previsión con Qx</i></b> .....	<b>33</b>
5.1	análisis y visualización de series.....	33
5.1.1	Análisis de mortalidad edad 5 – 10.....	33
5.1.2	Análisis de mortalidad edad 20 -30 .....	34
5.1.3	Análisis de mortalidad edad 40-50 .....	35
5.1.4	Análisis de mortalidad edad 60-70 .....	36
5.1.5	Análisis de mortalidad edad 80-90 .....	37
5.2	Análisis series diferenciadas .....	38
5.3	Análisis de raíz unitaria y estructura ARIMA.....	40
5.4	Conclusión y elección de estructura ARIMA .....	40
5.5	resultados.....	41
5.5.1	Resultado de previsiones 1 año.....	44
5.5.2	Resultado de previsiones 5 años .....	47
5.5.3	Resultado de previsiones 10 años .....	49
5.5.4	Resultado de previsiones 15 años .....	51
5.6	conclusiones .....	52
<b>6</b>	<b><i>Modelación y previsión con la índice mejora</i></b> .....	<b>53</b>
6.1	Visualización de series índice de mejora:.....	53
6.2	resultados.....	59
6.2.1	resultado de previsión de 1 año .....	60
6.2.2	resultado de previsión de 5 años.....	60
6.2.3	resultado de previsión de 10 años.....	61
6.2.4	resultado de previsión de 15 años.....	61
6.3	Conclusión:.....	62
<b>7</b>	<b><i>Modelo Qx con restricción y con muestra desde inicio</i></b> .....	<b>63</b>



7.1	Modelo con similitud.....	63
7.2	Modelo con toda la muestra .....	64
7.3	conclusiones .....	66
<b>8</b>	<b>Modelo Lee-Carter dinámico .....</b>	<b>68</b>
8.1	Estimación de las series $ax$ , $bx$ y $kt$ , junto con su previsión a 15 años.....	68
8.2	Resultados Lee-Carter Dinámico: .....	79
8.3	conclusión .....	81
<b>9</b>	<b>Modelo <math>Qx</math> con restricción y muestra pesada combinado con modelo Lee-Carte dinámico .....</b>	<b>82</b>
9.1	conclusión .....	85
<b>10</b>	<b> resultado de distintos modelos: .....</b>	<b>86</b>
<b>11</b>	<b> Conclusión:.....</b>	<b>88</b>
<b>12</b>	<b> Próximas líneas de investigación: .....</b>	<b>91</b>
<b>13</b>	<b> Bibliografía .....</b>	<b>92</b>

## índice de Gráficos

Gráfico 1 Mapa de calor cohorte toda muestra .....	20
Gráfico 2 Mapa de calor cohorte hasta edad 90, desde 1950 .....	23
Gráfico 3 Mapa de calor cohorte desde 1980.....	24
Gráfico 4 Mapa de calor cohorte desde 20 años y 1980.....	25
Gráfico 5 Serie diferenciada de esperanza de vida hasta 20 años.....	28
Gráfico 6 Series Q5 y Q10 .....	33
Gráfico 7 Series Q5 y Q10 desde 1960 .....	34
Gráfico 8 series Q20 y Q30 .....	34
Gráfico 9 Series Q20 y Q30 desde 1960 .....	35
Gráfico 10 Series Q40 y Q50 .....	35
Gráfico 11 Series Q40 y Q50 desde 1960 .....	36
Gráfico 12 Series Q60 y Q70 .....	36
Gráfico 13 Series Q60 y Q70 desde 1960.....	37
Gráfico 14 Series Q80 y Q90 .....	37
Gráfico 15 Series diferenciada Q20 y Q30 .....	38
Gráfico 16 Q40 y Q50 diferenciada.....	38
Gráfico 17 Series de Q20 y Q30 diferenciada desde 1960.....	39
Gráfico 18 Series Q40 y Q50 diferenciada desde 1960.....	39
Gráfico 19 Tablas de mortalidad 2014 Modelo_Qx_muestra_desde_1943 previsión 1 año .....	45
Gráfico 20 Tablas de mortalidad 2014 Modelo_Qx_muestra_desde_1960 previsión 1 año .....	45
Gráfico 21 Tablas de mortalidad 2014 previsión 1 año modelo qx .....	46
Gráfico 22 Tablas de mortalidad 2014 previsión 1 año modelo qx 25_55 .....	46
Gráfico 23 Tablas de mortalidad 2014 Modelo_Qx_muestra_desde_1943 previsión 5 año .....	47
Gráfico 24 Tablas de mortalidad 2014 Modelo_Qx_muestra_desde_1960 previsión 5 año .....	47
Gráfico 25 Tablas de mortalidad 2014 previsión 5 años modelo qx.....	48
Gráfico 26 Tablas de mortalidad 2014 previsión 5 año modelo qx 25_55 .....	48
Gráfico 27 Tablas de mortalidad 2014 Modelo_Qx_muestra_desde_1943 previsión 10 año .....	49
Gráfico 28 Tablas de mortalidad 2014 Modelo_Qx_muestra_desde_1960 previsión 10 año .....	49
Gráfico 29 Tablas de mortalidad 2014 previsión 10 años modelo qx .....	49
Gráfico 30 Tablas de mortalidad 2014 previsión 10 año modelo qx 25_55 .....	50
Gráfico 31 Tablas de mortalidad 2014 Modelo_Qx_muestra_desde_1943 previsión 15 año .....	51
Gráfico 32 Tablas de mortalidad 2014 Modelo_Qx_muestra_desde_1960 previsión 15 año .....	51
Gráfico 33 Índice mejora0 y Mejora10.....	53
Gráfico 34 Índice mejora0 y Mejora10 desde 1960.....	54
Gráfico 35 Índice mejora20 y Mejora30 desde 1960.....	55
Gráfico 36 Índice mejora40 y Mejora50 .....	56
Gráfico 37 Índice mejora40 y Mejora50 desde 1960.....	56
Gráfico 38 Índice mejora60 y Mejora70 .....	57
Gráfico 39 Índice mejora60 y Mejora70 desde 1960.....	57



Gráfico 40 Índice mejora80 y Mejora90 ..... 58

Gráfico 41 Índice mejora80 y Mejora90 desde 1960 ..... 58

Gráfico 42 Tablas de mortalidad 2014 Modelo\_mejora previsión 1 año ..... 60

Gráfico 43 Tablas de mortalidad 2014 Modelo\_mejora previsión 5 años ..... 60

Gráfico 44 Tablas de mortalidad 2014 Modelo\_mejora previsión 10 años ..... 61

Gráfico 45 Tablas de mortalidad 2014 Modelo\_Qx\_toda\_muestra\_con\_restricción de 1 a 19 años ..... 65

Gráfico 46 Tablas de mortalidad 2014 Modelo\_Qx\_toda\_muestra\_con\_restricción de 20 a 39 años ..... 65

Gráfico 47 Tablas de mortalidad 2014 Modelo\_Qx\_toda\_muestra\_con\_restricción de 40 a 59 años ..... 66

Gráfico 48 Tablas de mortalidad 2014 Modelo\_Qx\_toda\_muestra\_con\_restricción de 60 a 79 años ..... 66

Gráfico 49 serie Ax edad 0 Lee-Carter con muestra de 30 años ..... 69

Gráfico 50 serie Bx edad 0 Lee-Carter con muestra de 30 años ..... 69

Gráfico 51 serie Ax edad 10 Lee-Carter con muestra de 30 años ..... 70

Gráfico 52 serie Bx edad 10 Lee-Carter con muestra de 30 años ..... 71

Gráfico 53 serie Ax edad 30 Lee-Carter con muestra de 30 años ..... 72

Gráfico 54 serie Bx edad 30 Lee-Carter con muestra de 30 años ..... 73

Gráfico 55 serie Ax edad 50 Lee-Carter con muestra de 30 años ..... 74

Gráfico 56 serie Bx edad 50 Lee-Carter con muestra de 30 años ..... 75

Gráfico 57 serie Ax edad 70 Lee-Carter con muestra de 30 años ..... 76

Gráfico 58 serie Bx edad 70 Lee-Carter con muestra de 30 años ..... 76

Gráfico 59 serie Ax edad 90 Lee-Carter con muestra de 30 años ..... 77

Gráfico 60 serie Bx edad 90 Lee-Carter con muestra de 30 años ..... 78

Gráfico 61 Serie Kt y su previsión ..... 79

Gráfico 62 Tablas de mortalidad 2014 Modelo\_Lee-Carter dinámico de 1 a 24 años ..... 80

Gráfico 63 Tablas de mortalidad 2014 Modelo\_Lee-Carter dinámico de 25 a 65 años ..... 80

Gráfico 64 Tablas de mortalidad 2014 Modelo\_Lee-Carter dinámico de 69 a 99 años ..... 80

Gráfico 65 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter dinámico de 1 a 19 años ..... 82

Gráfico 66 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter dinámico de 20 a 39 años ..... 83

Gráfico 67 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter dinámico de 40 a 59 años ..... 83

Gráfico 68 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter dinámico de 60 a 79 años ..... 84

Gráfico 69 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter dinámico de 80 a 95 años ..... 84



## 1 Introducción

El presente trabajo, trata sobre un análisis general de los comportamientos de la mortalidad y longevidad de la sociedad española a lo largo del último siglo.

Se ha elegido a España como muestra por varias razones; la principal razón es la mayor facilidad de entender los datos que se obtienen al vivir en el país, debido a una mayor familiaridad con su pasado social y presente respecto a otros países de la Unión Europea.

El estudio pone énfasis en las posibles diferencias generacionales que se han ido dando en España a lo largo de los ciento siete años que van desde 1908 a 2014, pretendiendo lograr un profundo entendimiento del fenómeno, para realizar posteriormente proyección de las mortalidades en el futuro hasta 15 años.

El trabajo trata una muestra amplia, para poder tener una representación de sucesos tan importantes en la mortalidad como puede ser las épocas de bonanza económica, guerras civiles, dictaduras, paso de una dictadura a una democracia, Baby boom y el envejecimiento poblacional que se está dando en los países del primer mundo.

Se realiza un análisis objetivo de las generaciones pasadas y de las presentes, viendo como han ido afectando los sucesos anteriormente en la sociedad y principalmente en su mortalidad.

Se realiza un modelo ARIMA con restricción y muestra ponderada para obtener una previsión de la mortalidad en uno, en cinco, en diez años y quince años vista. Como es común en este tipo de modelo según se aleja de la fecha actual la previsión de la mortalidad se verá afectada por una disminución en su fiabilidad. Para revertir esta disminución, se procede a la modelación Lee-Carter dinámico para incrementar la precisión a largo plazo de modelo ARIMA.

Se pretende producir tablas de tarificación realistas, para algunas compañías aseguradoras del país, pudiendo ofrecer tarifas más precisas a sus clientes.

Por otro lado, se proyecta una posible herramienta para la previsión de la valoración actuarial del valor de las pensiones futuras de este país, pretendiendo hacer una revisión no solo de las cuantías a percibir por los antiguos contribuyentes, sino poder fijar y actualizar en qué momento de la vida del contribuyente es necesario implantar estas pensiones.



Finalmente, se han analizado los resultados obtenidos para dar una opinión fundada sobre las virtudes y limitaciones de modelo ARIMA y Lee-Carter para prever la mortalidad.

## 2 Marco teórico

Para la realización de este trabajo, se ha empleado principalmente dos técnicas de modelación, modelo Lee-Carter y modelo ARIMA.

En caso de modelo Lee-Carter se ha realizado más de 70 modelos de este tipo para lograr una serie temporal de parámetros y después se realiza la previsión de sus 3 parámetros para poder prever los  $q_x$ .

En caso del modelo ARIMA, se ha incorporado dos mejoras al modelo tradicional, la primera es establecer cierta similitud entre la previsión de una  $q_x$  y edades colindantes. La segunda es emplear todos los datos disponibles en la regresión asignando mayor importancia a los datos más recientes.

Al final se ejecuta un algoritmo para combinar los resultados de ambos tipos de modelos para lograr previsión de la tabla de mortalidad a largo plazo más realista.

### 2.1 Modelo Lee-Carter

Este modelo fue originariamente introducido por Lee y Carter en 1992, su ecuación es:

$$q_{xt} = e^{a_x + b_x k_t + r_{xt}}$$

Donde

$x$  representa las edades de personas

$t$  el año del registro de la mortalidad

$a_x$  es el vector de parámetros a estimar, contiene un valor por edad

$b_x$  es otro vector de parámetros a estimar

$k_t$  es otro vector de parámetros donde  $t$  representa el periodo.

$k_t$  en el modelo original es el único parámetro que no esté asociado con las edades, sino con el tiempo, se puede interpretar como el indicador del aumento de longevidad o esperanza de vida en el tiempo. Para interpretar cuantitativamente el significado de este parámetro, se debe comparar su variación con el de otro periodo, de esta forma, si en un año este parámetro  $k_t$  toma valor -1 y en el año próximo  $k_{t+1}$  toma valor -1.2, se está diciendo, la media de  $\ln(q_x)$  de todas las edades ha bajado 0.2, de una forma aproximada,



un  $q_x$  ha bajado  $e^{0.5} - 1$ , es decir 22%, una mortalidad de 0.1% bajaría a 0.82%. Así también se puede establecer la comparativa entre cualesquiera otras dos edades.

En muchos algoritmos de estimación de modelo Lee-Carter, para que exista una única solución matemática se impone que el vector de  $k_t$  tenga suma 0. Con esta restricción el parámetro en sí mismo tiene una interpretación más ya que existe una comparación implícita con la media de la muestra de estimación: Un parámetro  $k_t > 0$  indica que hay mayor mortalidad en ese año  $t$  en comparación con la media de toda la muestra de estimación, cabe señalar que esta afirmación es aproximada, ya que matemáticamente sólo es cierto que el logaritmo de  $q_x$  de ese año es mayor que la media de logaritmo de todos los  $q_x$ .

$b_x$  generalmente se interpreta como la forma de distribuir la mejora de mortalidad introducido por el parámetro  $k_x$ , es decir un parámetro  $b_{x1} > b_{x2}$  significa la reducción de edad  $x1$  es proporcionalmente mayor que la de  $x2$  cuando pasan los años (es decir cuando varia  $k_x$ ). Muchos algoritmos de estimación, para tener una única solución, no sólo imponen restricción sobre los parámetros  $k_x$ , sino también sobre  $b_x$ , generalmente la restricción consiste en que la suma de los  $b_x$  sea igual 1. Técnicamente se interpreta  $b_x$  como la elasticidad de  $q_x$  ante la variación de  $k_x$ , es decir la variación expresado en porcentaje cuando varia 1% de la mortalidad agregada introducida por  $k_x$ , que matemáticamente queda definida como:

$$b_x = \frac{\partial q_x}{\partial k_x}$$

De una forma aproximada, se puede decir que un  $b_x = 2$  significa que cuando la mortalidad media de todas las edades sube 1%, en esa edad  $x$  concreta,  $q_x$  sube 2%.

Hay diversas opiniones sobre la interpretación del  $a_x$  en el modelo original, incluso se puede llegar a opinar que este parámetro no es nada más que una especie de constante, y como sucede con los constantes de otros modelos de regresiones, la interpretación de  $a_x$  no tiene ninguna relevancia, tanto si el modelo es predictivo como si es explicativo.



A pesar de cierta falta de consenso, no existe apenas discusión sobre este tema, no se considera que la interpretación de este parámetro sea importante. Ya que generalmente la estimación de este parámetro es robusta, y en las fases de análisis o validación de parámetros los modeladores no suelen prestarle mucha atención. El presente trabajo puede ser considerado una excepción ya que como verá más adelante sí se hace la previsión de este parámetro, y sí tiene la interpretación del nivel de la previsión.

En el modelo original como consecuencia de la restricción impuesto por el algoritmo de optimización, los  $b_x$  tienen la suma igual 1, y  $k_t$  tiene media 0,  $a_x$  tienen la interpretación de la media geométrica de  $q_x$  en el periodo de muestra de estimación.

Tradicionalmente se proyecta únicamente el parámetro  $k_t$  para obtener una previsión de las tablas de mortalidad, obviando el hecho de que los otros dos parámetros son también evolutivos que cambian su valor en el tiempo. Esta hipótesis no es realista y hace que sólo pueda aprovechar una cantidad limitada de datos, se realiza un análisis a continuación detalladamente esta limitación, y su impacto sobre la elección de muestra de estimación.

### 2.1.1 Elección de número de años de muestra, la regla de 30 años.

Para la construcción del modelo, una de las decisiones más importantes a tomar es elegir qué muestra emplear. En opinión de algunos analistas, el uso de una muestra extensa, como una de más de 30 años no es adecuado para el modelo Lee-Carter, el ajuste podría no ser tan bueno como se espera, ya que a pesar de que el modelo lleva 3 vectores de parámetros, la estructura del modelo sigue siendo bastante lineal y rígida para capturar el fenómeno altamente no lineal y evolutivo. El modelo presentaría los siguientes fallos:

- El parámetro  $a_x$  como media de la misma edad de distintos años, dejaría de ser representativo y es uno de la hipótesis del modelo, en la práctica, algunos algoritmos calculan la media para estimar el valor de este parámetro.
- El parámetro  $k_t$ , en parte debido a ser el único parámetro para ajustar los niveles de unas 100 series, no es suficientemente flexible para lograr un buen ajuste de nivel en la previsión en sus distintos tramos de edad. Aunque para paliar este efecto de no linealidad, se ha aplicado la transformación logarítmica de  $q_x$ . En muchas ocasiones, dado los  $q_x$  son muy próximo a cero, la transformación logarítmica es muy similar a la serie original, y la no linealidad de las series de mortalidad sigue siendo muy importante para ser ajustada con un solo parámetro.

- En este trabajo, se considera el fallo más grave, la hipótesis de  $b_x$  única, no dinámica, se estaría asumiendo que la radio de dos índices de mejora de la mortalidad de dos edades diferentes y del mismo año es constante. Esta afirmación no es realista, especialmente en una muestra de estimación de muchos años, diversos estudios indican que la mejora de esperanza de vida produce en la primera mitad de siglo 20 es a través de la reducción de mortalidad infantil, y más tarde su principal causa es la reducción de mortalidad de personas de avanzada edad. El parámetro  $b_x$  es poco flexible para capturar estos fenómenos y otros importantes fenómenos históricos como guerras, epidemias que modifican las contribuciones a la longevidad de los distintos tramos de edades.

En este trabajo se ve claramente que los parámetros cambian de valor en los modelos con muestra de distintos años, el modelo tendría mucho peor ajuste si se hubiera construido un único modelo Lee-Carter con toda la muestra desde principio.

Lo anterior hace que el actuario si quiere aplicar el modelo Lee-Carter de forma tradicional debe limitarse en una muestra más pequeña y reciente donde no se produce un cambio de estructura de mortalidad para lograr un resultado coherente.

Para solventar este inconveniente y aprovechar lo máximo posible de potencial de los datos históricos para lograr una previsión más precisa. En este trabajo se realiza la modelación Lee-Carter de forma reiterada y dinámica para lograr unas series de parámetros a lo largo de toda la histórica de la muestra que se dispone desde 1908. Por el motivo mencionado, se ha fijado la muestra de cada estimación de 30 años en cada modelo, de esta forma se logra que los parámetros de cada estimación sean representativos. Esta modelación se realiza siguiendo estos pasos:

En primer lugar, se realiza un modelo Lee-Carter tal como plantearon originariamente los autores en 1992 con datos de los últimos 30 años.

A continuación, para crear el segundo modelo, se mueve la ventana de la muestra de estimación un periodo hacia el pasado, es decir se quita el dato del último año de la muestra del modelo anterior, y se añade el dato del año inmediatamente anterior al primer año del dato utilizado en el modelo anterior. De esta forma si el primer modelo emplea los datos de 1970 a 1999, el siguiente modelo emplearía datos de 1969 a 1998, y el



siguiente con datos de 1968 a 1997 así sucesivamente hasta llegar a emplear los datos de todos los años disponibles.

Se repite el paso anterior de forma reiterada, así se obtiene tantos vectores de parámetros como modelos que se ha realizado, siendo el tamaño de ventana temporal 30, obtenemos  $N-30$  vectores de parámetros ( $N$  es el número de años de datos que se dispone).

Una vez estimados los vectores de parámetros mencionados, se realiza previsión de los mismos mediante un modelo ARIMA.

Una vez obtenida la previsión de parámetros, se normaliza el vector  $b_x$ , multiplicando un constante para que sumen 1, el fin buscado es no introducir crecimiento del nivel de la serie con este parámetro. Los demás parámetros se han estimado bajo la hipótesis de que  $b_x$  tenga suma 1, y en este esquema, el parámetro  $b_x$  explica exclusivamente el reparto de la mejora de la mortalidad en el tiempo, no explica ni predice el cambio de nivel de la mortalidad, el parámetro  $k$  sí lo hace.

En este trabajo la predicción del parámetro  $k$  se realiza mediante un sencillo modelo ARIMA con el resultado del primer modelo Lee-Carter ya que este tiene muestra más reciente.

Con los 3 parámetros obtenido en los procesos anteriores se genera la previsión simplemente evaluando la ecuación del modelo Lee-Carter.

### 2.1.2 Problema de posición: análisis de varianza

La posición significa, el número de las personas vivas en cada tramo de edad a 30 de junio, hay estudios que opinan que este variable es relevante en el modelo Lee-Carter. Varios analistas argumentan que en caso de poca posición en algún tramo de edad avanzada puede producir problema de heterocedasticidad. Este problema no se da en este trabajo, en el análisis de dato demuestra justamente lo contrario en cuanto a la inferencia: durante muchos años, los datos en el tramo de corta edad con posición mucho más alta tienen más volatilidades que los de edad avanzada. Esta aparente paradoja viene explicada por las enormes diferencia en sus tasas de mortalidad, se procede a ilustrar con un ejemplo numérico de las tasas de mortalidad de año 2000 para edad 5 y 90:



dato año 2000	edad 5	edad 90
posición	373,003	53,944
$q_x$	0.016%	16.204%
desviación $q_x$ individual	1.26%	36.85%
desviación $q_x$	0.00207%	0.15865%
coeficiente de variación	0.1294	0.0098

Esta tabla se ha elaborado asumiendo que los  $q_x$  de la misma edad distribuye como binomial, se ve que la posición de edad 5 es 373 mil, unas 7 veces que la de edad 90, la desviación típica de una realización de un solo suceso de fallecimiento es mucho menor en caso de edad cinco: 1,26% frente a 36,85% de edad 90. La desviación de la tasa de mortalidad agregada del grupo de edad cinco también es muchísima más baja. Los anteriores suponen un factor para que el coeficiente de variación de variable de edad 5 sea más pequeño que el de edad 90. Pero la muy baja tasa de mortalidad de edad cinco es un factor para que su coeficiente de variación sea mayor, y este efecto predomina sobre el otro, da lugar a un coeficiente de variación menor que en caso de  $q_{90}$ . Es decir, la  $q_{90}$  es mucho más fiable que  $q_5$ . Coloquialmente se puede decir que, hay muchos niños, pero sus fallecimientos son unos sucesos muy raros, y su inferencia es extraordinariamente difícil. Por ello aunque las personas de edad cinco son más numerosas que las de edad 90, sus tasas de mortalidad inferidas son relativamente más volátiles y menos robustos.

En el presente trabajo no se realiza tratamiento de heterocedasticidad para Lee-Carter ya que como se explica en el ejemplo en todos los tramos hay muestra suficiente en las últimas décadas, y al estimar los modelos dinámicamente los errores de datos históricos alejada de la actualidad producen efecto limitado en la previsión.

## 2.2 Modelo de ARIMA

### 2.2.1 ARIMA con distintos pesos

En un modelo ARIMA empleado en el presente trabajo, de tipo (1,1,0), el modelo es:

$X_t = X_{t-1} + b X_{t-1} + e$ , donde  $e$  sigue una normal de media cero y varianza constante.





Dado que este tipo de modelo ARIMA sin el termino MA puede transformarse fácilmente en una regresión lineal.

$$X_t - X_{t-1} = + b X_{t-1} + e$$

Es decir, de forma  $Y = BX + e$  donde Y es la matriz de output y X la matriz de imput.

La técnica de regresión con peso consiste simplemente en multiplicar el peso en los ambos lados de la ecuación, de tal forma que se logran que los inputs de menor peso produzcan menor impacto sobre la función de verosimilitud. Dado que esta técnica de la regresión lineal clásica es ampliamente aceptada y cuyo utilidad y eficacia no es objeto de discusión desde hace muchos años, en este trabajo se omite los detalles matemáticas y técnicas (básicamente es descomponer la matrix de imput en una matrix de dato y otra de varianza-covarianza).

### 2.2.2 Regresión con peso y funciones de peso de este trabajo

Para determinar los pesos de cada dato a incorporar al modelo ARIMA, en este trabajo se ha definido una función de peso.

Esta función  $W(A, Y)$  asigna el peso de los datos en el modelo.

A es la distancia edad, es el valor absoluto de la diferencia de la edad de imput y la de output.

Y es la distancia años, es el número de años que hay entre el año de registro de imput y el de output.

Y esta función de peso se compone de dos funciones, una función de peso de edad  $W_a$ , y la otra de año de registro  $W_y$ :

$$W(A, Y) = W_a(A) * W_y(Y)$$

Y la definición de función de peso de año de registro es:

$W_y(Y) = (2^{0.5})^{(1/Cy)^{-Y}}$ : donde Cy es un constante, en este trabajo se le asigna el valor 8, la idea es el efecto de un error cuadrado producido de un dato sobre la función de verosimilitud decrece geométricamente mientras se aleja hacia el pasado, tal que un dato de hace 8 años tiene la mitad de efecto que un dato actual.

- $W_a(A) = 1$  si  $A = 0$

- $W_a(A) = Ca * (Ca/2)^{(A-1)}$  Si  $0 < A \leq 5$  , donde  $Ca = 0.895$

La idea es que entran en la regresión adicionalmente datos de los diez años colindantes, y el peso de ellos impactan sobre la función de verosimilitud aproximadamente doble que la propia serie a modelar y prever.

Y quizás  $W(A, Y) = W_a(A) * W_y(Y)$  no es la alternativa optima, la distancia euclídea podría tener más sentido económico antes una posible sinergia de los dos variables (edad, año), pero por prudencia y primar la robustez, en este trabajo se procede de forma más conservador.

La motivación para implementar estas técnicas es en parte la manifiesta heterocedasticidad que se presenta en los datos. El fenómeno de la mortalidad es altamente evolutivo, en alguna edad la mortalidad experimenta una reducción de nivel muy importante. Por este motivo, los datos recientes tienen mucho menor variación que los antiguos. La técnica más habitual, realizar la transformación logarítmica no es idónea en este trabajo, ya que en general los niveles de la serie son muy próximos a cero y el variable transformado es muy similar al original. Otro intento del presente trabajo como se verá más adelante es la transformación de la serie a índice de mejora, sin embargo, el ajuste final ha sido peor que el modelo de los  $q_x$  sin transformación. En parte buscando resolver a esta heterocedasticidad en este trabajo el último modelo ARIMA usa muestra con peso, del contrario los datos más recientes tienen escaso peso en el algoritmo de optimización para encontrar el valor de parámetro del modelo.



### 3 Presentación de datos

Para este trabajo se ha empleado información obtenida de la base de datos de la página: <http://www.mortality.org> de The Human Mortality. Dicha base de datos fue creada originariamente por la universidad de California y el Instituto Max Planck Max Planck de Investigación Demográfica en Rostock de Alemania. Actualmente esta base de datos está siendo completada y actualizada por científicos de todo el mundo y periodistas del mundo académico.

Para poder llevar a cabo este trabajo, se ha usado información desagregada de la base de datos anteriormente mencionada. Los datos presentados son todos los ciudadanos españoles comprendidos entre los años 1908 y 2014, teniendo en cuenta un intervalo de edad de 0 a 110 años. La información recogida por fallecimiento con dicho trabajo se obtiene el número de fallecidos entre el número poblacional de dicho año.

Se ha encontrado como principal inconveniente, la calidad, cantidad y veracidad de los datos anteriores al año 1943, como ya se observará cuando se analice el modelo.



## 4 Análisis de mejora generacional cohorte

En este primer apartado se analizan las mejoras generacionales de cohorte mediante un mapa de calor y la esperanza de vida por tramo, presentando finalmente una conclusión de los mismos.

## 4.1 Estudio de fenómeno cohorte mediante mapa de calor

Para realizar el estudio del fenómeno cohorte, se procede a crear un mapa de calor para detectar gráficamente las mejoras de la cohorte siendo la mejor técnica a la hora de detectar este fenómeno mediante las líneas de color oscuro que aparecen en la gráfica.

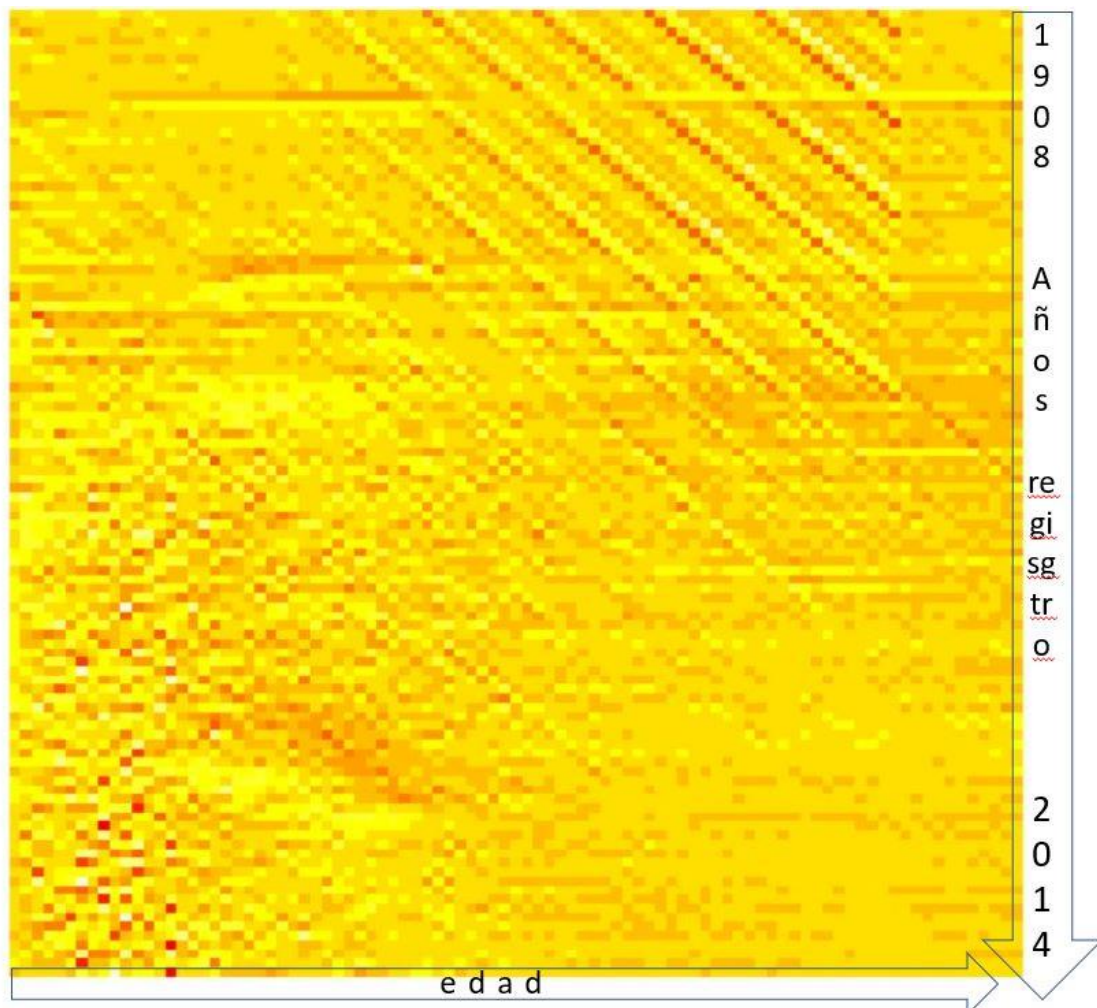
Definición de la mejora:

Tal como lo indicó Ruiz Chou(2017), la mejora  $M(\text{año}, \text{edad}) =$

$$M(a,e) = 1 - \frac{Q_e^a}{Q_e^{a-1}}$$

- $Q_e^a$  es la tasa de mortalidad de edad  $e$  en el año de registro  $a$
- $Q_e^{a-1}$  es la tasa de mortalidad de edad  $e$  en el año de registro  $a - 1$

Gráfico 1 Mapa de calor cohorte toda muestra



Elaboración propia con dato de Human Metabolome Database (HMDB)



Podemos observar que efectivamente se da el fenómeno de cohorte, especialmente en la primera mitad del pasado siglo. Esto se debe principalmente a los grandes avances sanitarios que acontecen en ese periodo de tiempo que mejora radicalmente la esperanza de vida.

En la gráfica que se estudia, las líneas diagonales observadas en el cuadrante superior derecho deben ser tenidas en cuenta con reservas, debido a que a la hora de afirmar la existencia de mejora de cohorte, se observan una serie de anomalías que se observan en los años 60,70 y 80. Apareciendo anomalías repetitivas en la tasa de mortalidad, en las edades múltiplo de 10 partiendo de los 60 años. Es por ello que en los 60, 70 y 80 la tasa de mortalidad es claramente más alta que los años colindantes a estas fechas, apareciendo un patrón repetitivo en cada década.

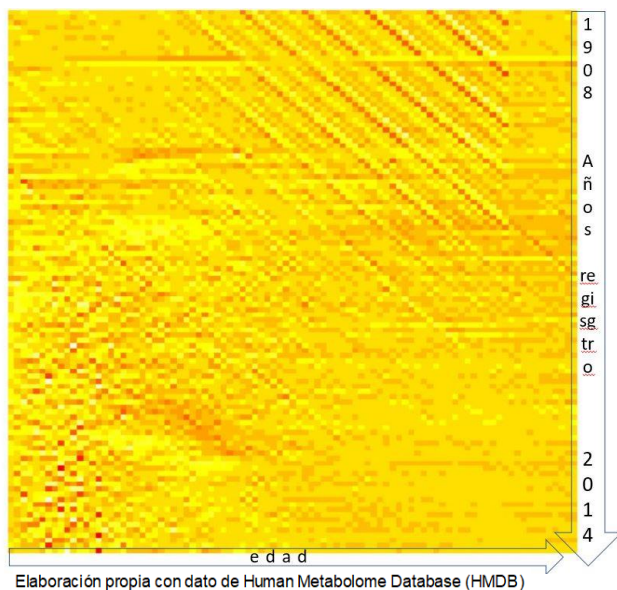
Dado la carencia de medios de la época, a la hora de la recopilación de datos, se intuye que no se registra con la edad exacta las muertes, pudiéndose dar un redondeo en los múltiplos de 10 a la hora de la recopilación de los datos.





Es por ello se puede afirmar que es uno de los factores principales por el cual se puede dar la aparición de las líneas de mejora de cohorte. También al producirse esta anomalía en los últimos años de vida de los individuos de esa época estas anomalías aparecen en el cuadrante superior derecho de la gráfica.

Para corroborar este fenómeno se vuelve a ver la gráfica 1



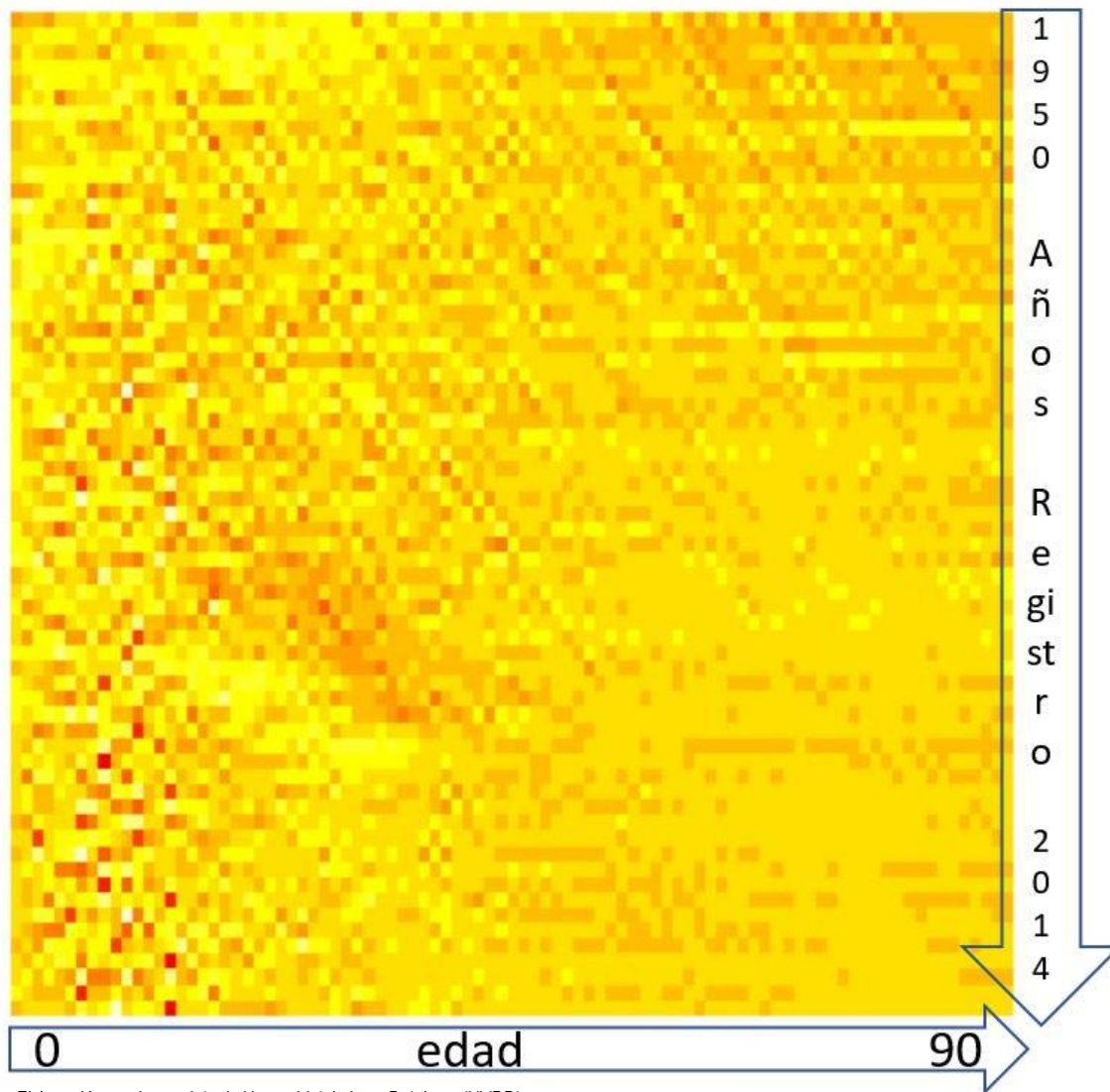




Podría existir un problema a la hora de escalar los datos, debido a larga línea temporal elegida, donde no se observa un comportamiento de cohorte significativo.

Es por ello el estudio se ha centrado en la segunda mitad del siglo pasado, donde se podría pensar que existe un problema de heterocedasticidad en la gráfica1.

*Gráfico 2 Mapa de calor cohorte hasta edad 90, desde 1950*



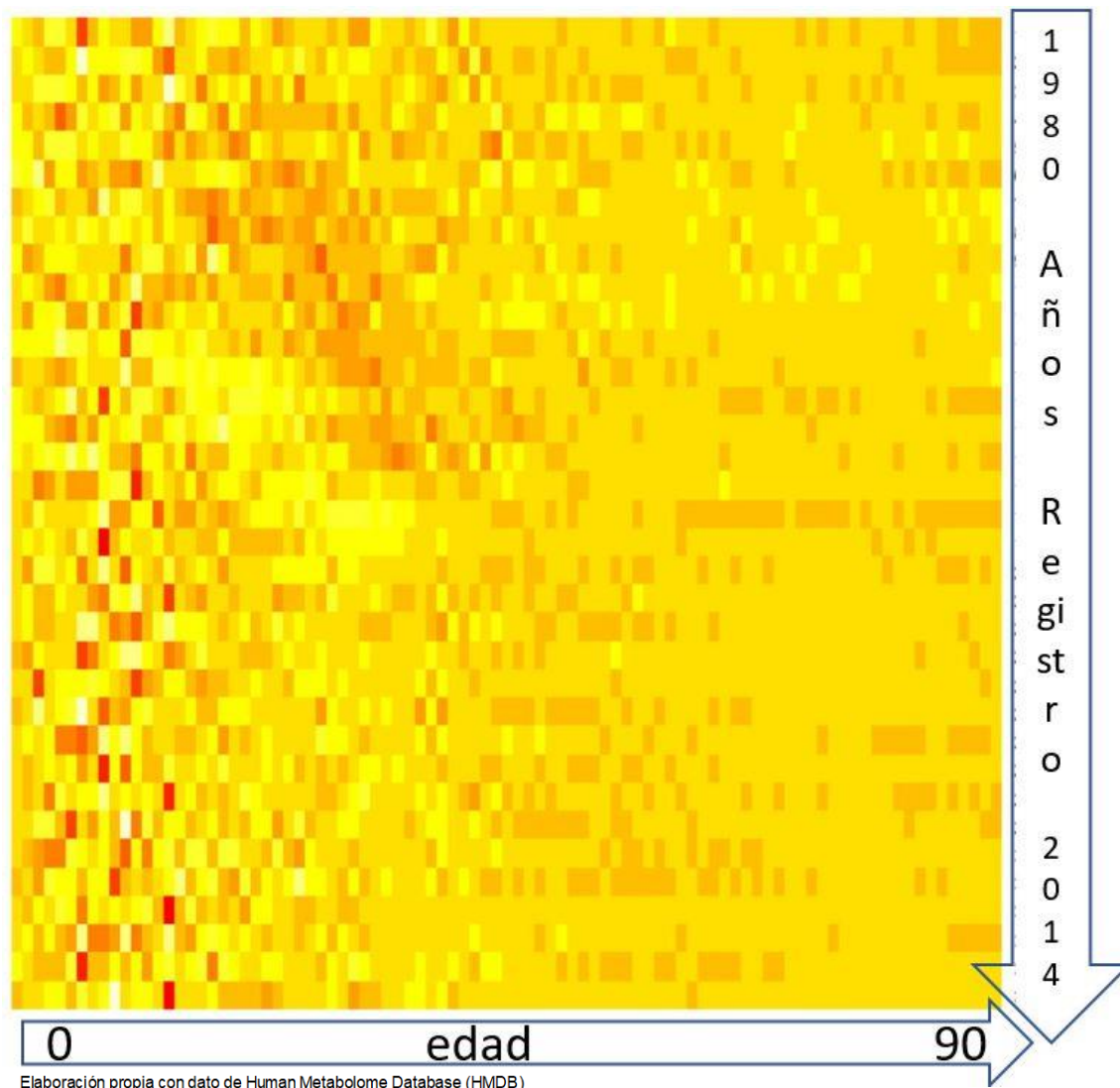
Comparando ambas gráficas de cohorte, se puede observar en la siguiente gráfica 2 una disminución de la aparición de la línea de cohorte. Esto es debido a una disminución de los grandes avances médicos y sanitarios que aparecieron a principios de siglo XX.

En esta gráfica se observa como en la parte superior derecha de la gráfica se mitiga los fenómenos que se ha mencionado en la gráfica 2.



Uno de los factores por los que se puede dar dicha mitigación en las líneas de cohorte que aparecen en la parte superior derecha de la gráfica, es debido a que, a partir de finales de la década de los cuarenta en España, la recogida de la información de las defunciones se lleva a cabo con registros con un mayor control. Esto es debido a que no solo se tiene más cuidado en la recogida de la fecha de defunción sino también en la fecha de nacimiento de los individuos.

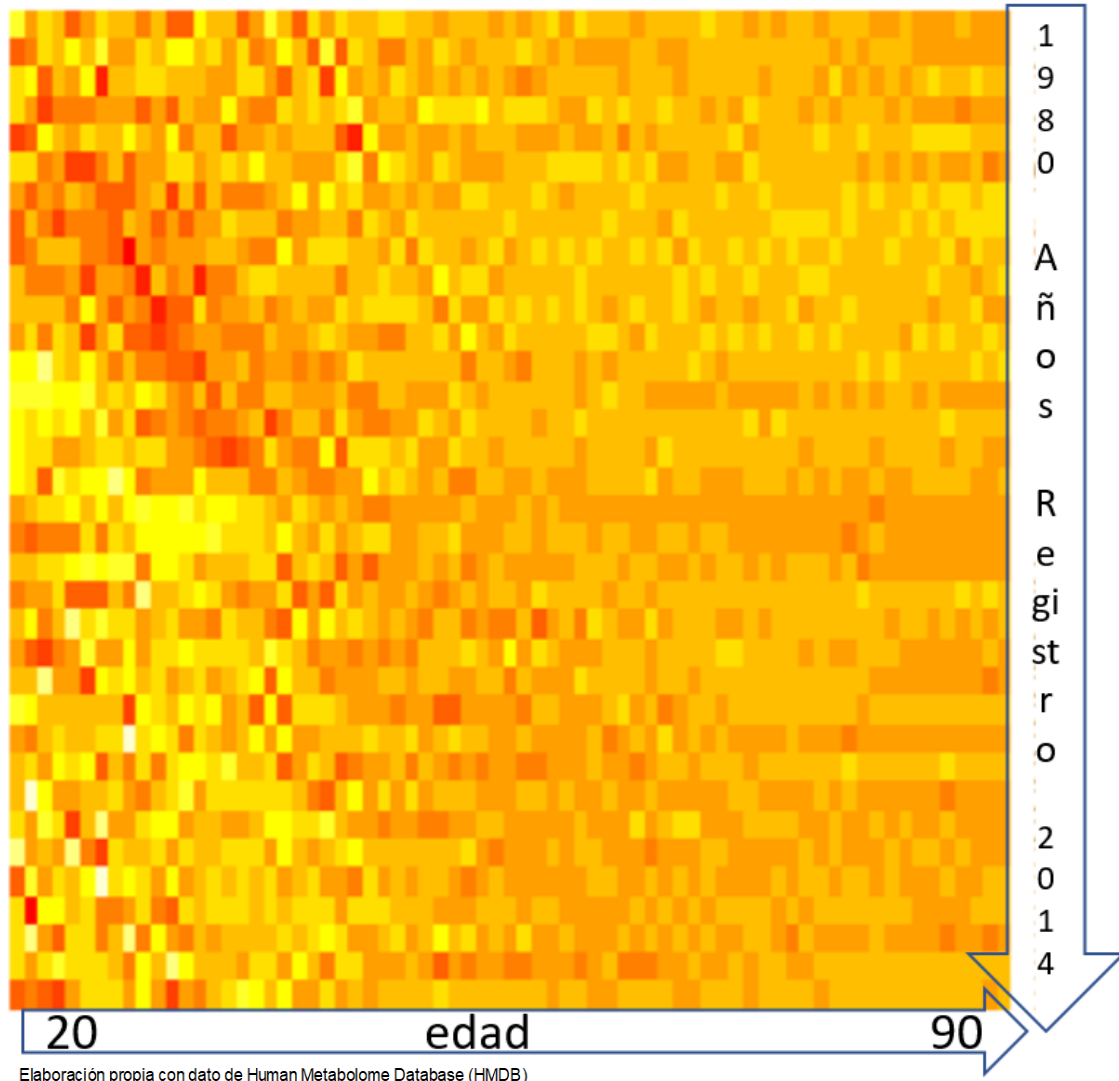
Gráfico 3 Mapa de calor cohorte desde 1980



En este gráfico se observa, una mejora en el contraste de los primeros años de vida del individuo. Uno de los tramos con mayor índice de fallecimiento entre los individuos, se solía dar en el primer año de vida, sobre todo entre los varones.

En esta grafica podemos ver como hay una mejora en la mortalidad de los individuos durante los primeros años de vida, principalmente debido a los avances médicos.

*Gráfico 4 Mapa de calor cohorte desde 20 años y 1980*



En el gráfico cuatro, el estudio se centra en las edades comprendidas entre los 20 y 90 años. Esto es debido a la intención de reflejar como se estabiliza la mortalidad, una vez que pasan los primeros años de vida de la población que se estudia.

Es por ello el efecto de corte es cada vez menos representativo al tener la mortalidad del país un comportamiento más progresivo y constante, dándose un menor número de saltos cualitativos de mejora.

Esto es debido a que en los años 1980 en el país de estudio; España, no acontece ninguna catástrofe naturales ni guerras que puedan desvirtuar los datos de la muestra.

## 4.2 enfoque alternativo: Series de esperanza de vida por tramo.

En las anteriores graficas se han usado la edad, año de registro y mejoras de esperanza de vida de los individuos como datos principales, para poder tener una información lo más completa posible.

Una vez presentados y analizados los mapas de calor en las anteriores gráficas habiendo tenido como puntos principales las mejoras en la esperanza de vida, año de registro y edad. Es hora de centrarse en la serie temporal de la esperanza de vida del individuo en sus diferentes tramos que se analizan. Estos datos son más claros y sencillos a la hora de explicar el fenómeno de cohorte. Con ello se obtendrá una información más rica y concisa que en el caso anterior. Para ello se procede a una desagregación de la información, con representación en varias gráficas.

Con ello se pretende agrupar la información del estudio en grupos generacionales de 20 años.

La definición estadística de esperanza de vida por tramo para el grupo de personas nacidos en el mismo año es:

$$\sum_{i=1}^{fin-ini} i P_{ini}$$

Donde:

- ini es la edad inferior del tramo
- fin es la edad superior del tramo

Con esta información se pretende obtener la esperanza de vida acotada anualmente dentro de cada uno de los grupos de muestra que comprenden 20 años.

Con esta táctica de procesamiento de la información se pretende hacer un estudio de las anomalías poblacionales durante un año. De esta forma es más sencillo localizar los fenómenos de cohorte y los comportamientos anómalos de la muestra.

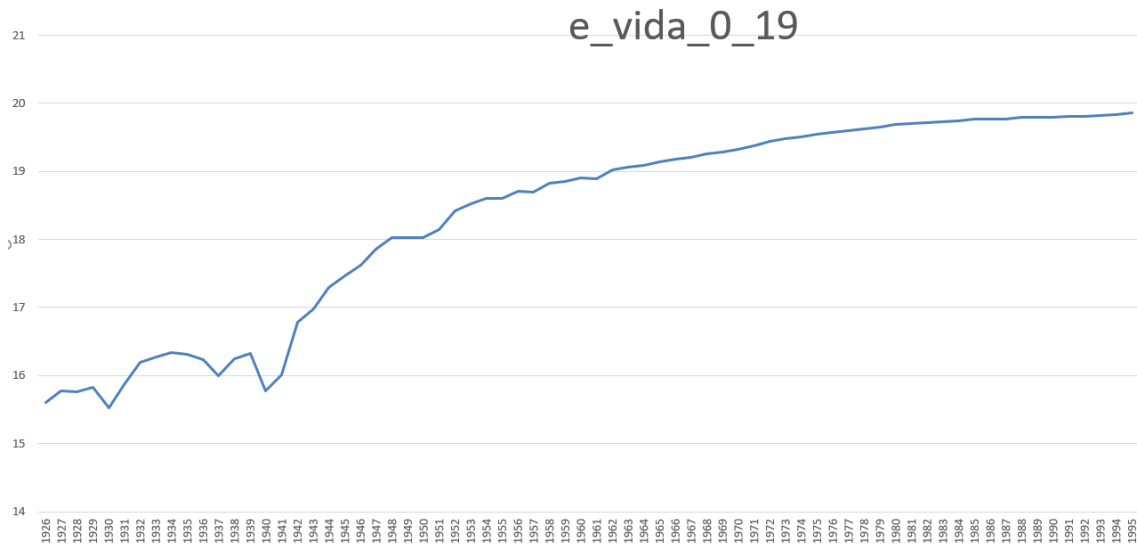
Las principales anomalías que se pueden llegar a dar son las guerras, pandemias y problemas estructurales.



Las principales ventajas que presenta este método son: la robustez ante los errores mencionados de datos y la sinterización de la información.

Se obtienen las series de esperanza de vida.

Esperanza de vida de tramo de edad entre 0 y 20 años:



Elaboración propia con dato de Human Metabolome Database (HMDB)

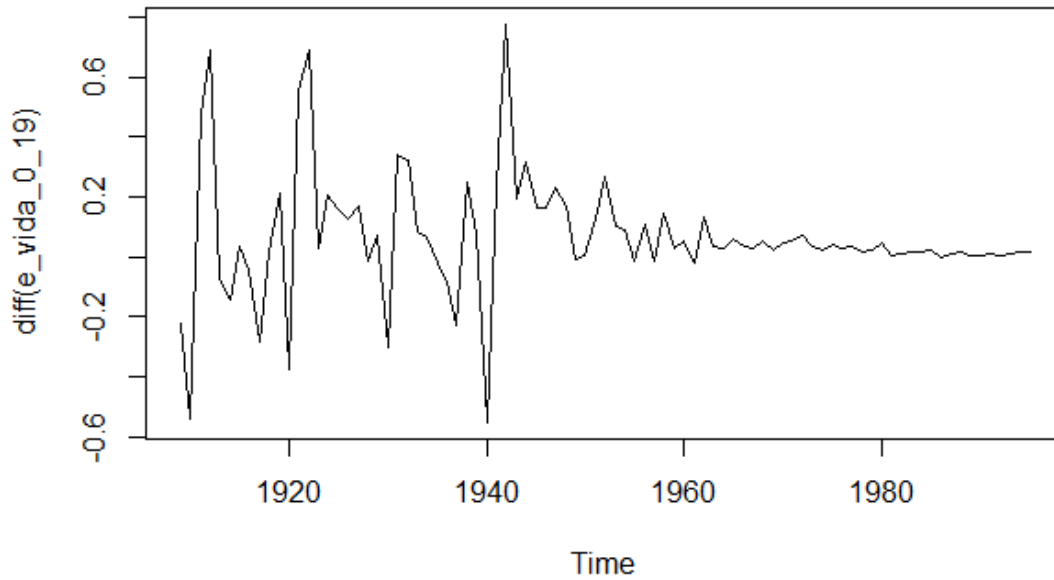
Como se puede observar en esta gráfica, existe una espectacular mejora de esperanza de vida en el tramo comprendido entre 0 y 20 años, especialmente entre 1940 y 1960, sin embargo, en los últimos años las mejoras observables son menos reseñables que en periodos anteriores. Pudiéndose observar como poco a poco la esperanza de vida para una generación concreta experimenta un crecimiento continuado hasta que se estabiliza cercano al máximo.

Dado que uno de los objetivos es la previsión de la tabla de mortalidad, se procede a proyectar con el modelo ARIMA de esta serie.

A la vista de gráfica, se puede afirmar que existe al menos una raíz unitaria, y lo confirma el test DFA.

Siguiendo los procesos estándar de la modelación ARIMA, se procede a la primera diferencia de la serie, y se procede la visualización de la serie diferenciada:

Gráfico 5 Serie diferenciada de esperanza de vida hasta 20 años



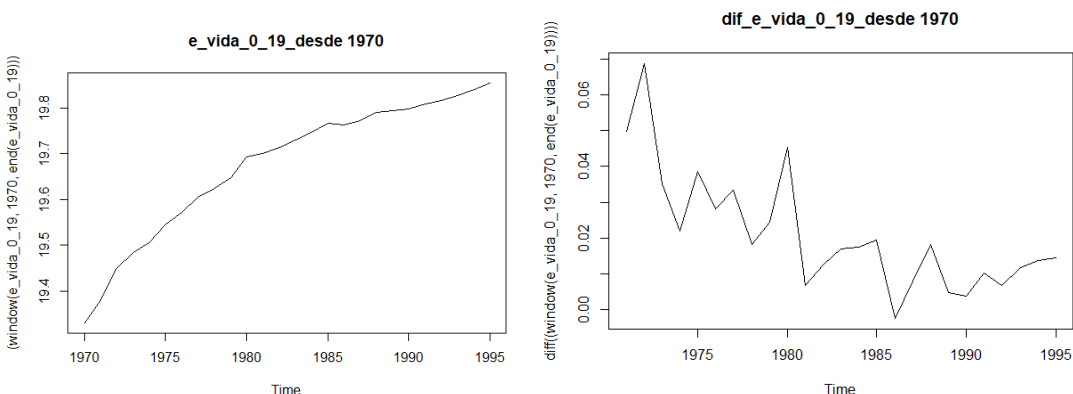
La serie presenta un claro punto de ruptura en los años 50, y sus comportamientos antes y después de este punto son notablemente diferentes, de esta forma, se ve que la estructura ARIMA de la serie original, la serie desde origen hasta 1970 y la serie desde 1970 hasta final son muy distintos entre ellos.

```
> auto.arima(e_vida_0_19)
Series: e_vida_0_19
ARIMA(0,1,2) with drift
Coefficients:
    ma1    ma2  drift
    0.2536 -0.3407 0.0684
s.e.  0.0978  0.0909  0.0186

> auto.arima(window(e_vida_0_19, 1915, 1970))
Series: window(e_vida_0_19, 1915, 1970)
ARIMA(1,1,1) with drift
Coefficients:
    ar1    ma1  drift
    -0.5137  0.7674  0.0907
s.e.  0.1971  0.1334  0.0326

> auto.arima(window(e_vida_0_19, 1970, end(e_vida_0_19)))
Series: window(e_vida_0_19, 1970, end(e_vida_0_19))
ARIMA(0,2,1)
Coefficients:
    ma1
    -0.5843
s.e.  0.1443
```

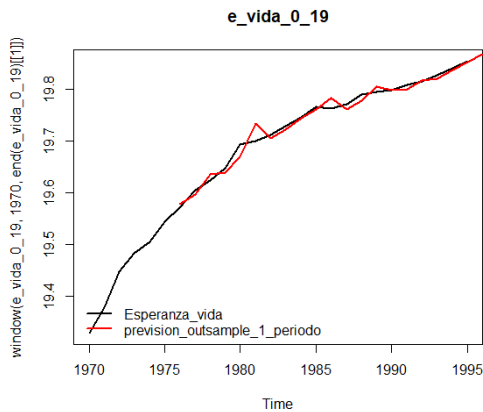
Se visualice la serie diferenciada desde 1970, se observa que existe crecimiento positivo y cambiante, si bien en los últimos años el incremento se sitúa en torno de 0.015



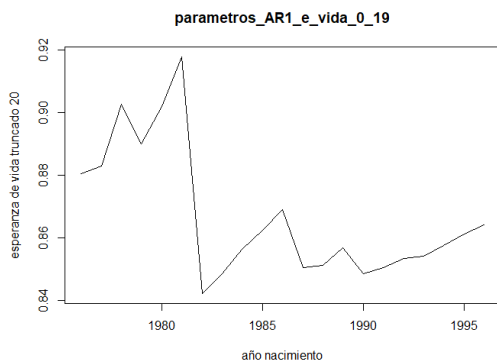


Se concluye que una estrategia de modelo ARIMA sin intervención es muy difícil de capturar la estructura de esta serie ni prever con precisión, y por ello se procede con prudencia en busca de robustez en el modelo, y se opta por un modelo sencillo de ARIMA(1,1,0) para la previsión.

La previsión de un periodo se quedaría:



También se ve la inestabilidad del modelo, uno de los indicadores es la serie de parámetros AR1 de los modelos con datos de distintos cortes finales, sin embargo, esta serie a partir de las 1985 no presenta un cambio de nivel importante, con eso no se descarta que es factible una proyección a medio-largo plazo.

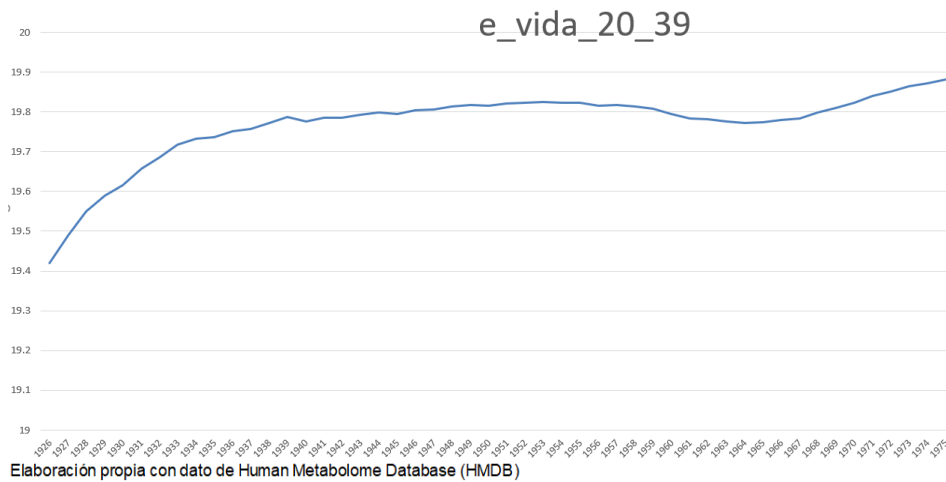


En el análisis anterior se observa una estructura compleja, pero susceptible de proyectar en el futuro, al menos en corto-medio plazo.

En las siguientes graficas de esperanza de vida de otros tramos se ve la estructura similar con la misma posibilidad de proyección.

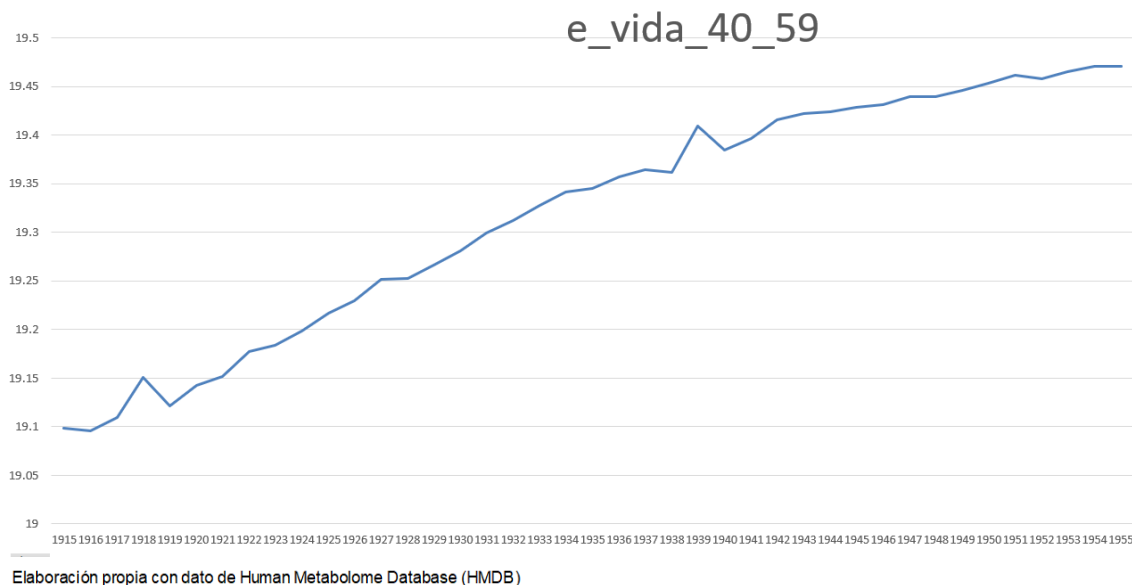


Se visualiza también la esperanza de vida de tramo de edad entre 20 y 40 años:



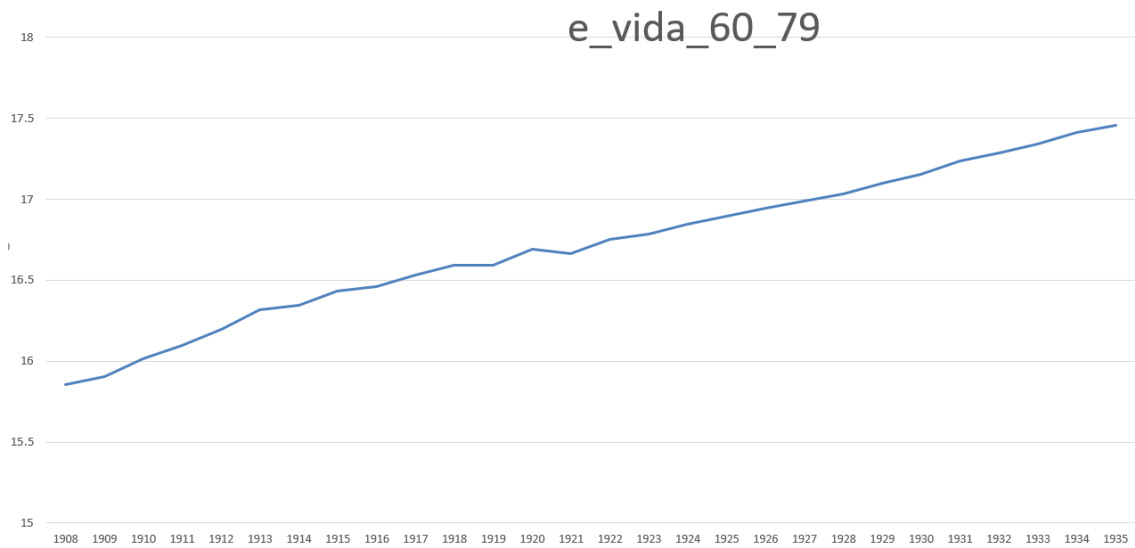
Se puede observar que también existe mejora muy importante durante 1925 – 1940. Son la población que ha vivido la guerra durante su niñez o adolescencia, y sus esperanzas de vida de este tramo no parece afectado por la guerra civil. Después de los 1940, la mejora es mucho menor, incluso con descenso durante los años 60 - 70. Posiblemente debido a los nuevos factores de fallecimientos como VIH, accidente de tráfico.

Esperanza de vida de tramo de edad entre 40 y 60 años:



La mejora parece más continua, lineal, si bien presenta atípicos que contribuye a la aparición de mejora de cohorte.

Esperanza de vida de tramo de edad entre 60 y 80 años:



Elaboración propia con dato de Human Metabolome Database (HMDB)

Esta serie parece fácil de predecir ya que experimenta un crecimiento lineal y bastante constante, de tal manera que, con la técnica adecuada, es factible su proyección de estos comportamientos de la mortalidad.

Esperanza de vida de tramo de edad entre 80 y 100 años:



Elaboración propia con dato de Human Metabolome Database (HMDB)

### 4.3 Conclusión sobre la mejora generacional cohorte.

Tanto los mapas de calor como la esperanza de vida por tramo, señala que cada vez la series son más suaves, y el efecto de mejora cohorte es cada vez menos evidente, especialmente en los últimos años donde se centra nuestro interés de la previsión de la





tabla de mortalidad. Parece que el cambio de comportamiento de la mortalidad es suave y una proyección de la tabla de mortalidad parece factible en vista de esta.

## 5 Análisis de dato y Modelación, previsión con $Q_x$

Una vez realizado los primeros análisis de datos, se procede a su modelación y previsión de las mortalidades.

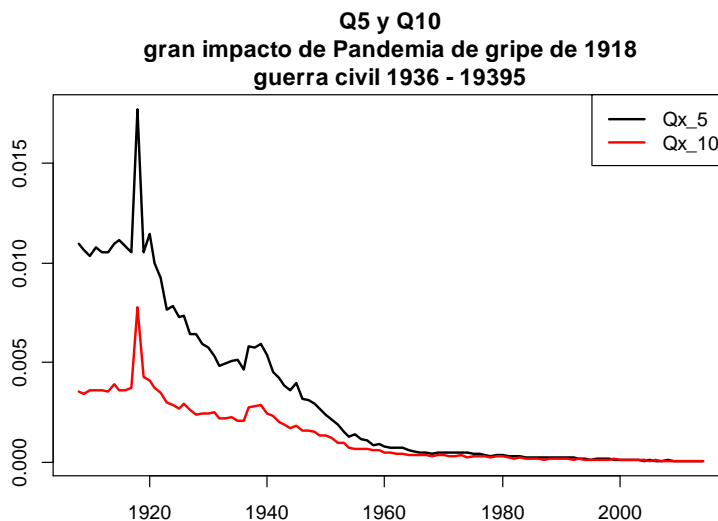
### 5.1 análisis y visualización de series

Un enfoque bastante directo de modelar es modelar  $q_x$  sin realizar transformación alguna y proyectar la mortalidad con un modelo ARIMA, para este fin, se procede a visualizar y analizar las series de  $q_x$ , y dado que el nivel de la serie antes de 1960 es mucho mayor que después, también se genera y analiza gráficamente por separado la serie a partir de 1960.

#### 5.1.1 Análisis de mortalidad edad 5 – 10

Para la previsión de serie con modelo ARIMA, primero se procede a analizar gráficamente las series:

Gráfico 6 Series  $Q_5$  y  $Q_{10}$

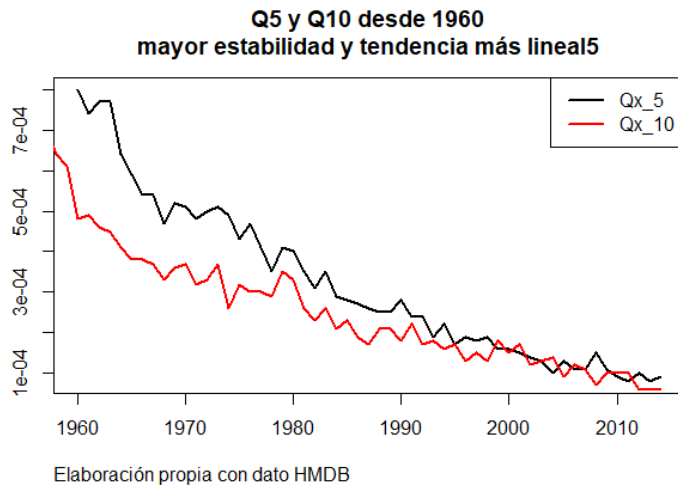


Elaboración propia con dato de Human Metabolome Database (HMDB)

Destaca el enorme impacto de la pandemia de gripe de 1918 y la guerra civil de 1936 – 1939, y debido al problema de escalar que produce estos dos fenómenos, no se observa con nitidez los comportamientos de las series en otros años. Para poder estudiar los comportamientos de la serie de la segunda mitad de siglo XX, se procede a visualizar la serie desde 1960.



Gráfico 7 Series Q5 y Q10 desde 1960



Se ve en los últimos años de esta serie, hay mayor estabilidad y los cambios de la serie son mucho más suaves sin salto cuantitativo, sin embargo, no es suficientemente lineal para poder realizar una proyección de forma sencilla. Aunque el cambio de pendiente es suave, indica que la previsión a corto plazo es de gran facilidad y precisión siendo incógnita el comportamiento de la previsión a largo plazo.

### 5.1.2 Análisis de mortalidad edad 20 -30

Gráfico 8 series Q20 y Q30

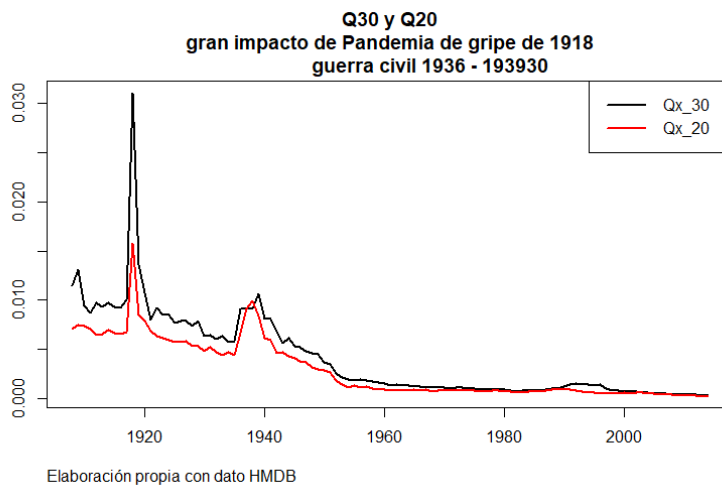
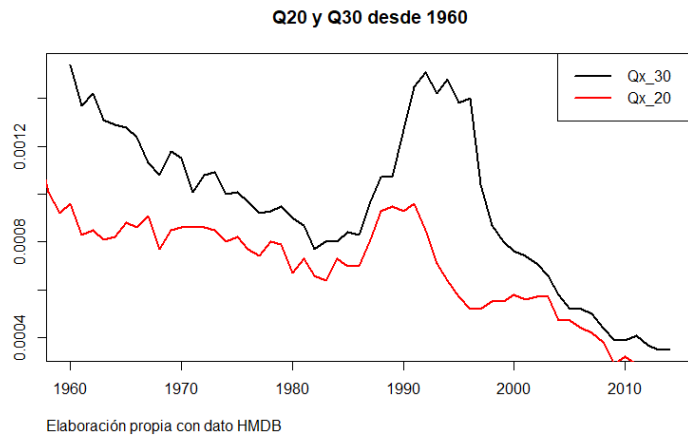




Gráfico 9 Series Q20 y Q30 desde 1960



Los repuntes de años 1988 – 1995 posiblemente son debido a VIH u otro factor de fallecimiento como accidente de tráfico. Se puede observar los mismos comportamientos algunas las series de edad 40 y 50.

### 5.1.3 Análisis de mortalidad edad 40-50

Gráfico 10 Series Q40 y Q50

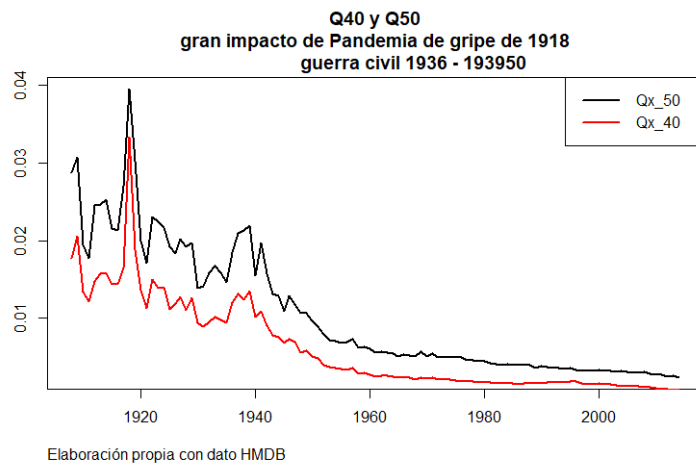
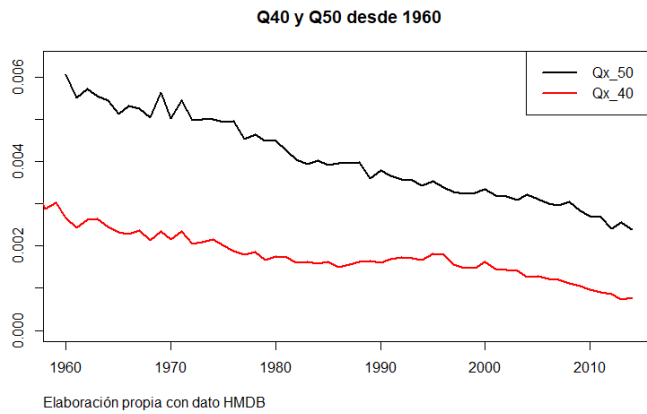


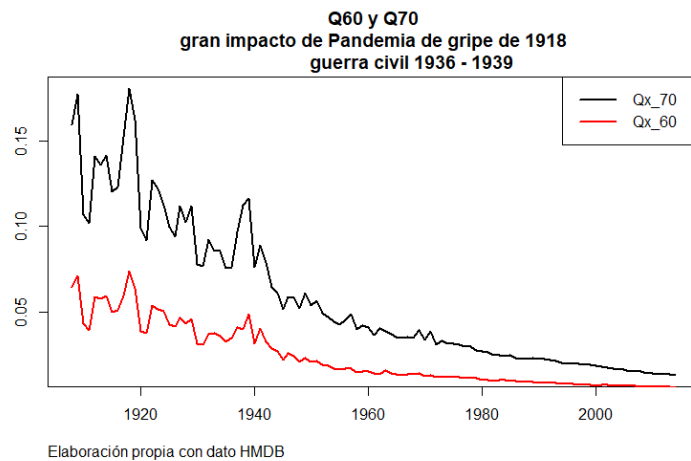
Gráfico 11 Series Q40 y Q50 desde 1960



Se puede observar cierta convergencia entre las tasas de mortalidad de  $q_{40}$  y  $q_{50}$  en el sentido de que con el paso de tiempo, la diferencia de mortalidad de ambas edades son menores, se trata del efecto resultante de la mejora a mayor velocidad en la serie de Q50 en comparación con Q40.

#### 5.1.4 Análisis de mortalidad edad 60-70

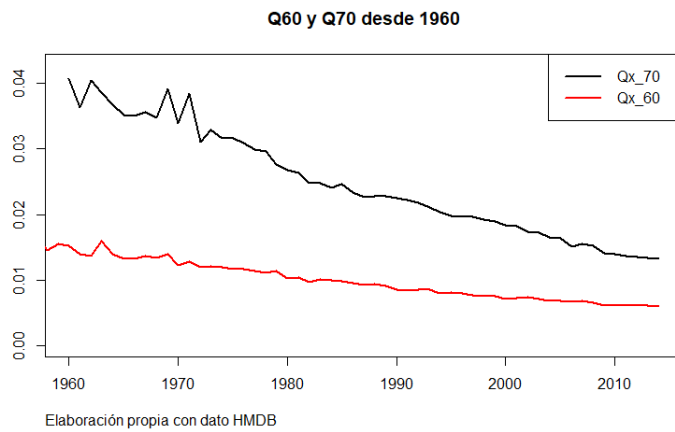
Gráfico 12 Series Q60 y Q70



Los efectos de guerra civil se impactan menos en estas series de mortalidad de la población de edad avanzada, pero sigue siendo muy importante.



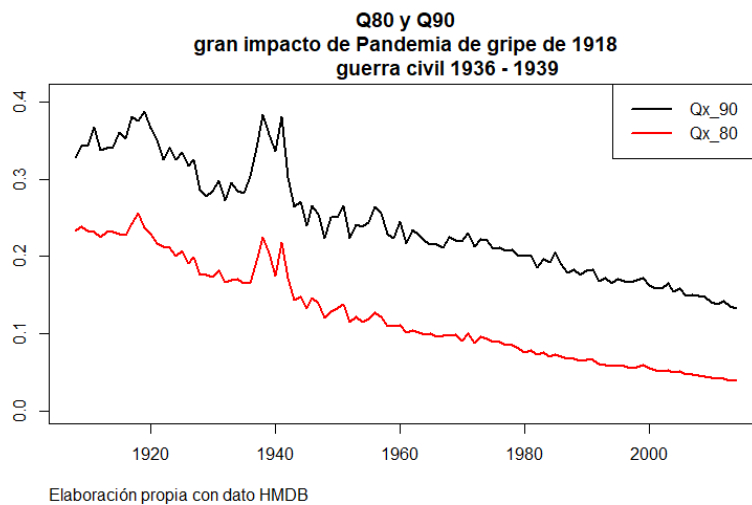
Gráfico 13 Series Q60 y Q70 desde 1960



Se puede observar la misma convergencia entre las dos series, efecto resultante de la mayor velocidad en mejora en la serie de Q70.

### 5.1.5 Análisis de mortalidad edad 80-90

Gráfico 14 Series Q80 y Q90



Sorprendentemente las dos series son casi paralelas, la explicación podría venir de que una vez se haya alcanzado edad avanzada, y superada cierto umbral, los efectos de muchas enfermedades sobre la mortalidad ya no varía con el aumento de la edad. Y el impacto de pandemia de gripe tiene un efecto retardado.

La guerra civil genera impacto muy importante sobre la mortalidad de estos dos grupos, como sucede en otros grupos de la población.

## 5.2 Análisis series diferenciadas

Una vez vistos las gráficas anteriores, se procede siguiendo el procedimiento habitual de modelo ARIMA; la siguiente decisión que se afronta es cuantas veces se debe diferenciar la serie.

Se procede a aplicar una diferencia a las series, ya que conceptualmente la serie tiene que tener al menos una raíz unitaria, se visualiza las gráficas de las series diferenciada una vez:

Gráfico 15 Series diferenciada Q20 y Q30

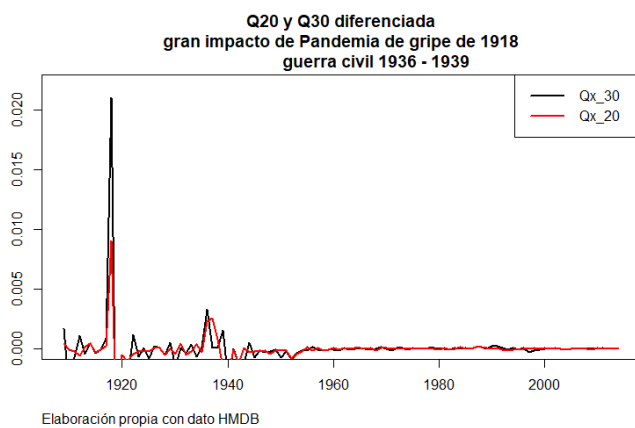
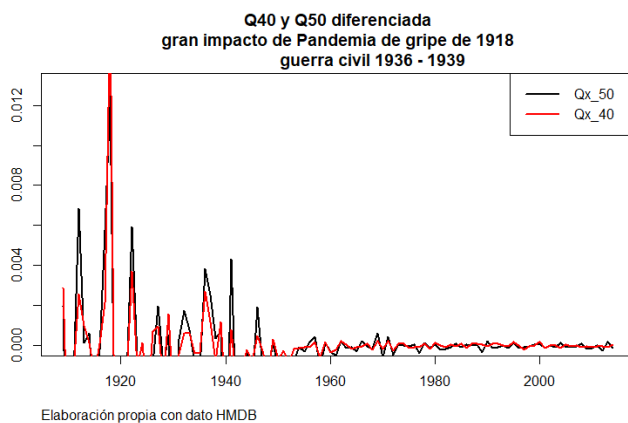


Gráfico 16 Q40 y Q50 diferenciada



A la vista de estas gráficas de series diferenciadas, se confirma la sospecha de que el impacto de epidemia de gripe de 1918 y guerra civil española es extraordinario no solo en la serie en nivel sino también en la diferenciada, dado esta anomalía, una estrategia prudente sería excluir los datos antes de 1942 de la muestra de estimación para que los parámetros sean robustos, conceptualmente podemos decir que esos dos sucesos son



extremadamente imposibles de reproducirse en nuestro periodo de previsión, estos dos sucesos tienen sus comportamientos completamente diferentes a la estructura que se desea inferir, consecuentemente no aportaría mejora importante para la previsión.

Se surge otra duda, de que, si se debe incluir los datos para la estimación desde 1942 hasta 1960, ya que por una parte se sabe que la mortalidad en el periodo de previsión (al principio de siglo 21) tienen comportamiento bastante diferente que el del periodo comprendido entre 1942 y 1960, y es un motivo para excluir en la estimación. Por la otra parte, si se estima el modelo con datos desde 1960, en la estimación sólo se dispondrá unos 40 datos, y sería una muestra bastante pequeña para capturar sus distintos comportamientos en el tiempo. De tal forma se procede a construir dos modelos, uno con datos desde 1943 y otro desde 1960 para su posterior comparación de sus resultados.

Gráfico 17 Series de Q20 y Q30 diferenciada desde 1960

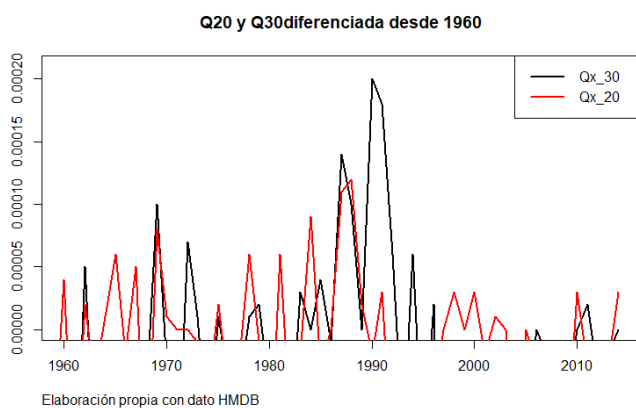
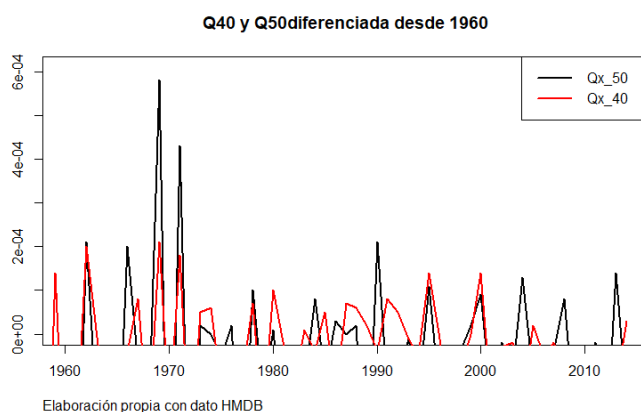


Gráfico 18 Series Q40 y Q50 diferenciada desde 1960



En esta grafica sí observamos el comportamiento más homogéneo de los datos desde 1960.



### 5.3 Análisis de raíz unitaria y estructura ARIMA

Para determinar la estructura ARIMA, se procede al análisis de número de raíces unitarias y estructura ARIMA mediante distintos test estadísticos.

Tabla 1 Estructura arima de los modelos Qx

edad	Estructura ARIMA		P_VALUE test ADF		P_VALUE test PP		n de raíces unitaria test KPSS		n de raíces unitaria test PP	
	muestra 1943	muestra 1960	muestra 1943	muestra 1960	muestra 1943	muestra 1960	muestra 1943	muestra 1960	muestra 1943	muestra 1960
0	3000120	2100120	59.40%	36.16%	76.02%	97.98%	2	2	0	1
5	0200120	2200110	1.00%	1.00%	83.28%	50.50%	2	1	0	1
10	1100120	0100110	1.00%	6.19%	82.07%	1.00%	2	1	0	0
20	0100120	0000110	1.00%	21.52%	68.81%	57.60%	2	1	0	1
30	3200120	1100110	2.14%	53.66%	80.91%	77.63%	2	1	0	1
40	0200120	0100110	1.00%	79.10%	74.39%	38.61%	2	1	0	1
50	2200120	2100110	1.00%	53.57%	70.14%	1.00%	2	1	0	0
60	2200120	0100110	1.00%	80.33%	32.96%	1.00%	2	1	0	0
70	0100110	1000110	41.27%	48.78%	1.00%	1.00%	1	1	1	0
80	0100110	2000110	27.28%	70.62%	1.00%	1.00%	1	1	1	0
90	4000110	0100110	15.30%	23.63%	1.00%	1.00%	1	1	1	0
100	0100110	0100110	84.77%	63.82%	7.50%	1.00%	1	1	0	0

### 5.4 Conclusión y elección de estructura ARIMA

En vista de análisis anterior, a pesar de que se ha eliminado de la muestra los valores atípicos (pandemia y guerra civil), la estructura ARIMA de las series siguen siendo bastantes inestables tanto la serie desde 1943 como la serie desde 1960.

Se da este fenómeno porque las causas de la evolución de la mortalidad son muy numerosas y cambiantes en el tiempo, existen distintos factores sociales–sanitarios como comportamiento en la dictadura, tracción a la democracia, despegue económico e incluso la crisis económica acompañado de las nuevas enfermedad (VIH, cáncer etc), es extremadamente difícil de encontrar una única estructura ARIMA que encaje en todas estas 202 series. De esta forma, adoptando una actitud cautelosa, desde punto de vista de lograr mayor estabilidad y mayor robustez en los modelos, se procede a modelar con la estructura ARIMA(1,2,0), es decir con un parámetros ar1 y doble diferencia sobre la serie original.

## 5.5 resultados

Una vez estimado el modelo, llegamos a estos parámetros:

edades	previsión 1 año		previsión 5 años		previsión 10 años		previsión 15 años	
	muestra 1943	muestra 1960	muestra 1943	muestra 1960	muestra 1943	muestra 1960	muestra 1943	muestra 1960
0	-0.493	-0.940	-0.492	-0.937	-0.492	-0.932	-0.491	-0.926
1	-0.341	-0.954	-0.341	-0.951	-0.341	-0.948	-0.340	-0.952
2	-0.792	-0.602	-0.790	-0.604	-0.790	-0.628	-0.786	-0.566
3	-0.804	-0.700	-0.801	-0.698	-0.796	-0.689	-0.792	-0.692
4	-0.575	-0.622	-0.575	-0.626	-0.575	-0.635	-0.574	-0.634
5	-0.778	-0.582	-0.778	-0.595	-0.777	-0.578	-0.776	-0.577
6	-0.600	-0.604	-0.595	-0.586	-0.596	-0.589	-0.604	-0.625
7	-0.750	-0.883	-0.751	-0.886	-0.751	-0.889	-0.748	-0.888
8	-0.631	-0.658	-0.631	-0.659	-0.617	-0.626	-0.617	-0.630
9	-0.727	-0.339	-0.725	-0.296	-0.725	-0.299	-0.725	-0.278
10	-0.747	-0.666	-0.749	-0.672	-0.749	-0.669	-0.747	-0.652
11	-0.907	-0.711	-0.907	-0.713	-0.909	-0.727	-0.907	-0.696
12	-0.523	-0.471	-0.526	-0.488	-0.527	-0.492	-0.521	-0.458
13	-0.559	-0.637	-0.561	-0.641	-0.565	-0.647	-0.572	-0.660
14	-0.746	-0.782	-0.753	-0.796	-0.758	-0.806	-0.744	-0.791
15	-0.600	-0.668	-0.599	-0.669	-0.596	-0.658	-0.596	-0.665
16	-0.445	-0.542	-0.432	-0.504	-0.426	-0.488	-0.422	-0.475
17	-0.637	-0.593	-0.642	-0.609	-0.649	-0.635	-0.651	-0.640
18	-0.554	-0.671	-0.537	-0.651	-0.541	-0.658	-0.543	-0.665
19	-0.670	-0.630	-0.675	-0.643	-0.670	-0.621	-0.670	-0.617
20	-0.439	-0.546	-0.435	-0.537	-0.432	-0.532	-0.431	-0.530
21	-0.511	-0.721	-0.505	-0.714	-0.504	-0.728	-0.504	-0.740
22	-0.494	-0.624	-0.494	-0.624	-0.495	-0.633	-0.493	-0.635
23	-0.243	-0.344	-0.239	-0.334	-0.234	-0.323	-0.237	-0.333
24	-0.426	-0.744	-0.426	-0.759	-0.424	-0.756	-0.414	-0.754
25	-0.607	-0.655	-0.606	-0.654	-0.607	-0.660	-0.605	-0.657
26	-0.632	-0.595	-0.629	-0.578	-0.627	-0.568	-0.621	-0.521
27	-0.590	-0.371	-0.590	-0.376	-0.589	-0.372	-0.589	-0.365
28	-0.665	-0.624	-0.663	-0.615	-0.661	-0.607	-0.660	-0.604
29	-0.682	-0.709	-0.681	-0.707	-0.681	-0.708	-0.683	-0.712
30	-0.790	-0.482	-0.790	-0.485	-0.790	-0.489	-0.793	-0.505

Se puede observar que, al cambiar el periodo de estimación, los parámetros sí cambian de valor.

Para comparar el resultado, se ha generado otro tipo de previsión llamada ultimo  $Q_x$ , que es el enfoque tradicional, donde se asigna el  $q_x$  del último período como la previsión.

La previsión a un año del ultimo  $Q_x$  de 2014 es el  $q_x$  real registrada de 2013, que coincide con el último dato de la muestra de estimación de otros modelos.

La previsión a 10 años del ultimo  $Q_x$  de 2014 es  $q_x$  real registrada de 2004. La idea es comparar el resultado del modelo con la previsión más fácil de obtener e implementada en muchas compañías de seguro actualmente.

Una vez estimado el modelo, se llega a los siguientes resultados de MAPE.

La MAPE es el estadístico empleado para medir el ajuste del modelo, significa el error medio en porcentaje, en este caso se calcula de esta forma

MAPE es la media de todos  $(r - p) / r$  donde:

- $r$  significa la tasa de mortalidad
- $p$  significa la tasa de mortalidad prevista por el modelo

De esta forma, un MAPE de 0.61% indica que cuando en el periodo de previsión hay 10,000 fallecidos, la previsión prevé de media 10,061 fallecidos

<b>MAPE (0- 65años)</b>	<b>ultimo qx</b>	<b>model Qx 1943</b>	<b>model Qx 1960</b>
1 año	-4.54%	0.61%	1.48%
5 años	-28.28%	11.83%	11.54%
10 años	-63.98%	12.34%	15.16%
15 años	-93.28%	16.45%	18.18%

Dado que los comportamientos de la mortalidad en distintos tramos de edad son diferentes, y para orientarse a la valoración del seguro de vida TAR, algo muy común en España, se procede a elaborar este estadístico para el tramo de edad de 25 - 55

<b>MAPE (25- 55años)</b>	<b>ultimo qx</b>	<b>model Qx 1943</b>	<b>model Qx 1960</b>
1 año	-3.88%	1.36%	1.73%
5 años	-28.85%	8.85%	9.30%
10 años	-65.88%	2.17%	6.77%
15 años	-94.85%	64.49%	53.42%

Se puede observar que el ajuste del modelo a un año es bastante aceptable, su error sitúa en alrededor de 1.5%, se ve que el ajuste del modelo con muestra desde 1943 es ligeramente mejor que el modelo de 1960.



La otra utilidad de la previsión de la tabla de mortalidad es el cálculo actuarial de las rentas u otros seguros de vida a largo plazo, para ellos se procede a calcular las MAPE para edades de 56 a 95.

<b>MAPE (56 -95 años)</b>	<b>ultimo qx</b>	<b>model Qx 1943</b>	<b>model Qx 1960</b>
1 año	-1.37%	-0.37%	-0.81%
5 años	-10.43%	4.22%	4.57%
10 años	-21.72%	13.63%	11.12%
15 años	-38.60%	-51.20%	-50.57%

Se puede observar un comportamiento similar destacando el muy bueno ajuste la previsión de un año. Conforme que se aumenta el plazo de previsión, la proyección se empeora. En base a este resultado, se puede afirmar que este modelo ARIMA no es apropiado para la proyección a largo plazo.

También se procede a elaborar otro estadístico para medir el ajuste, errores en nivel:

su definición es simplemente la media de la tasa de mortalidad real menos la prevista. De esta forma un error en nivel de -0.03% quiere decir el error en la predicción de número de fallecidos equivale a 0.03% de la total población, es decir la ratio de estos dos siguientes:

- Diferencia entre la tarifa pura según el modelo y la indemnización real final materializada
- Suma asegurada

Igual que en el caso de MAPE, también se calcula los errores en nivel para los 3 tramos de edades:

<b>error (0-65años)</b>	<b>ultimo qx</b>	<b>model Qx 1943</b>	<b>model Qx 1960</b>
1 año	0.00%	0.00%	0.00%
5 años	-0.02%	0.01%	0.01%
10 años	-0.04%	0.02%	0.02%
15 años	-0.07%	-0.03%	-0.03%

<b>error (25-55 años)</b>	<b>ultimo qx</b>	<b>model Qx 1943</b>	<b>model Qx 1960</b>
1 año	-0.004%	0.001%	0.001%
5 años	-0.025%	0.007%	0.007%
10 años	-0.052%	-0.014%	-0.012%
15 años	-0.070%	-0.010%	-0.015%

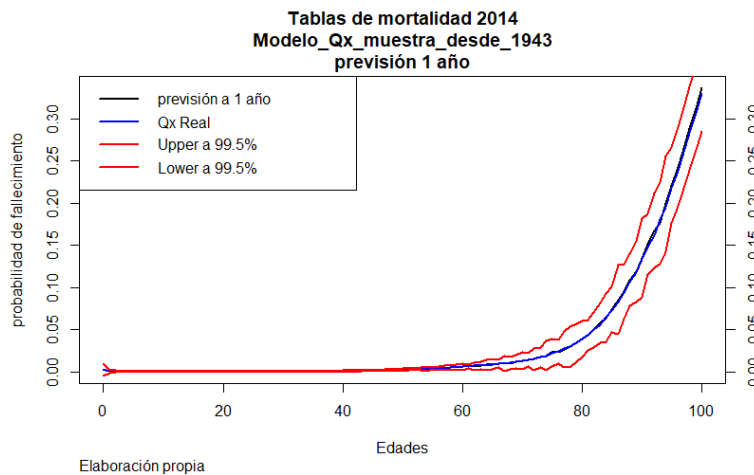
<b>error (25-55 años)</b>	<b>ultimo qx</b>	<b>model Qx 1943</b>	<b>model Qx 1960</b>
1 año	-0.06%	-0.06%	-0.08%
5 años	-0.52%	0.13%	0.12%
10 años	-0.99%	0.50%	0.39%
15 años	-1.69%	-3.21%	-3.20%

Se puede interpretar esta ratio como dotación adicional expresado sobre la suma asegurada, es más fácil de interpretar estos estadísticos comparando con los de otros modelos; existe una mejora notable de los dos modelos ARIMA frente a la previsión último Qx.

### 5.5.1 Resultado de previsiones 1 año

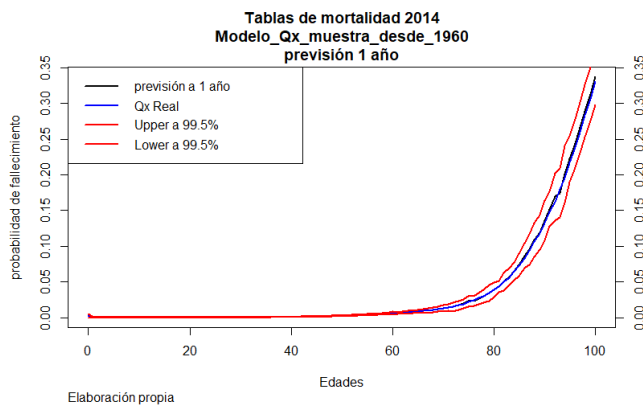
Se visualiza los resultados de la previsión gráficamente junto con las bandas de intervalo de confianza a 99,5% de nivel de confianza, se elige este nivel ya que es el que exige la normativa SolvenciaII.

Gráfico 19 Tablas de mortalidad 2014 Modelo\_Qx\_muestra\_desde\_1943 previsión 1 año



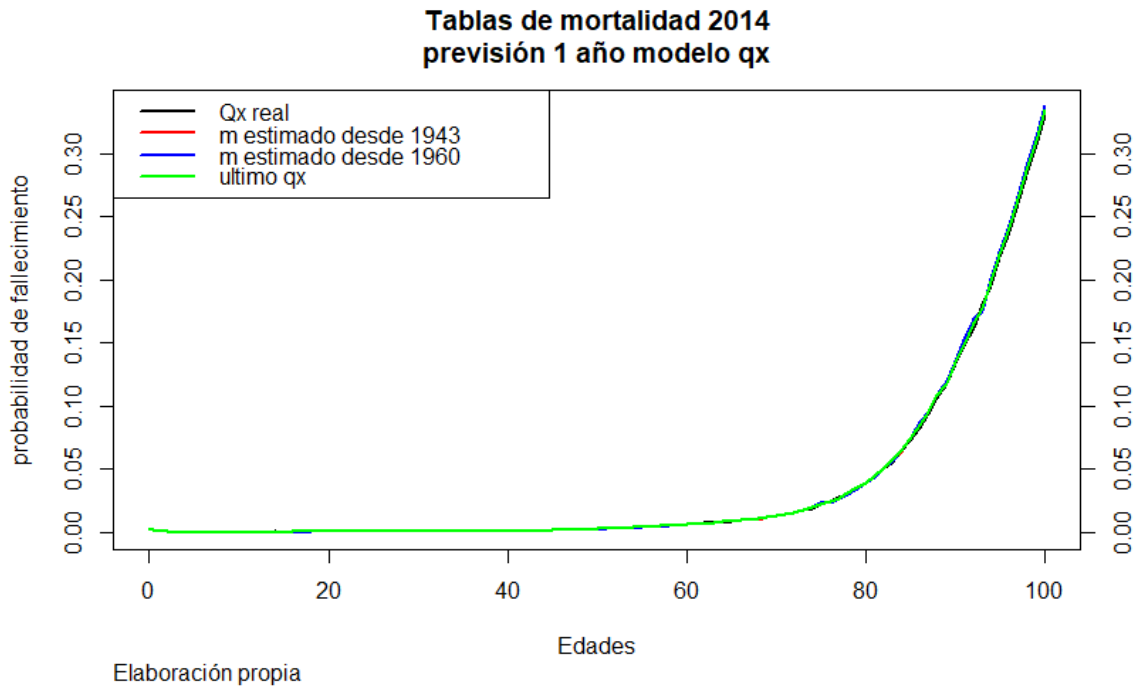
Se puede ver que la predicción tiene una forma similar a la real, y la serie real está localizada entre las dos bandas. Esto indica que este modelo podría servir como modelo interno para el cálculo de capital económico a efecto de SolvenciaII. Sin embargo las bandas de nivel de confianza a 99,5%, como se puede observar en la gráfica tienen su intervalo muy amplio haciendo que la dotación del capital económica sea mayor que lo exigido por la formula estándar y supondría un coste adicional para la empresa aseguradora.

Gráfico 20 Tablas de mortalidad 2014 Modelo\_Qx\_muestra\_desde\_1960 previsión 1 año



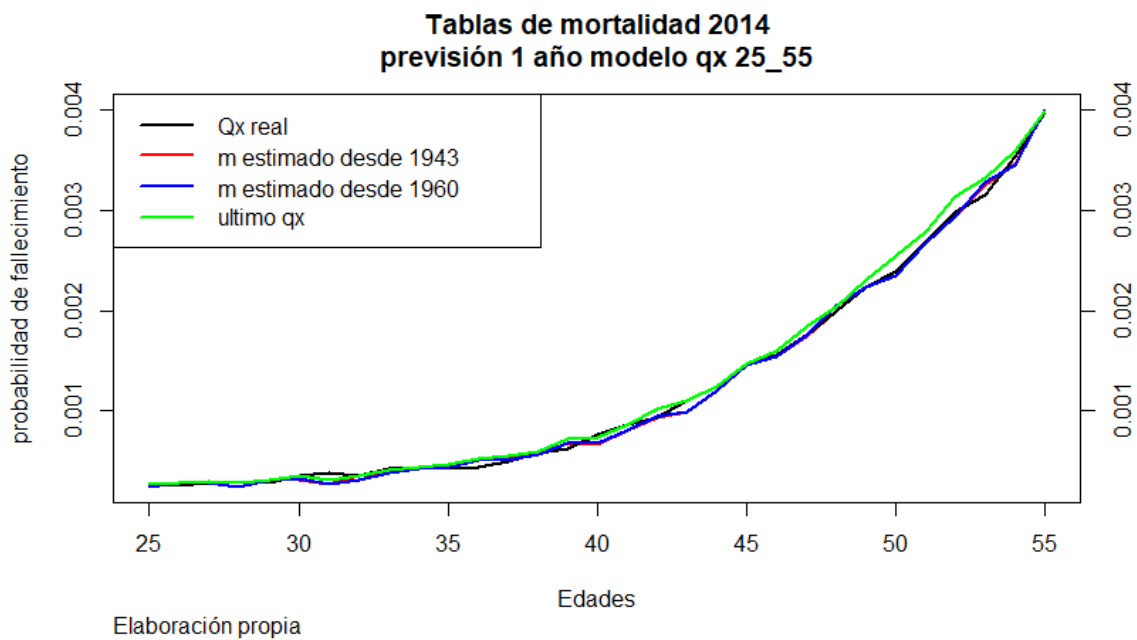
A continuación, se procede a visualizar y comparar gráficamente los modelos ARIMA con las dos muestras diferentes empleadas en el trabajo junto con el modelo ultimo Qx.

Gráfico 21 Tablas de mortalidad 2014 previsión 1 año modelo qx



Se puede observar un ajuste casi perfecto, con el fin de aislar las distorsiones inducido por los posibles problemas de escalar, se procede a graficar el tramo de edad desde 25 a 55 para lograr un análisis más detallada de este tramo.

Gráfico 22 Tablas de mortalidad 2014 previsión 1 año modelo qx 25\_55



En esta gráfica, Se puede afirmar que los dos modelos dan previsiones similares, ambas son de gran precisión, y mejor que ultimo qx.

### 5.5.2 Resultado de previsiones 5 años

A continuación, se procede también a visualizar los resultados de previsión a 5 años:

Gráfico 23 Tablas de mortalidad 2014 Modelo\_Qx\_muestra\_desde\_1943 previsión 5 año

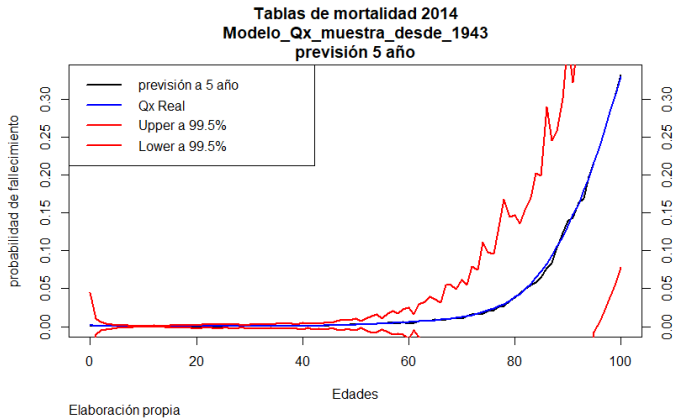
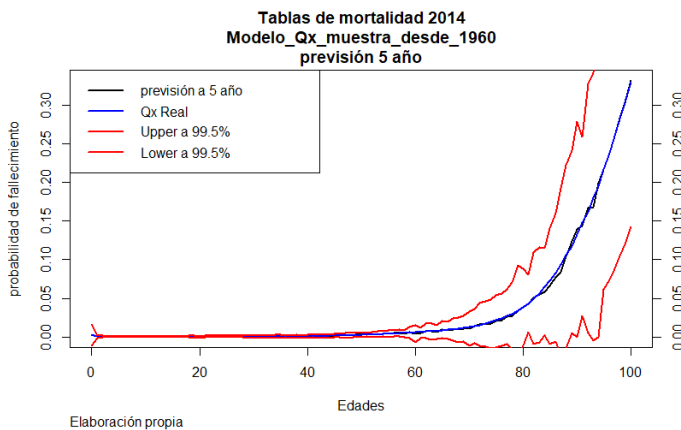


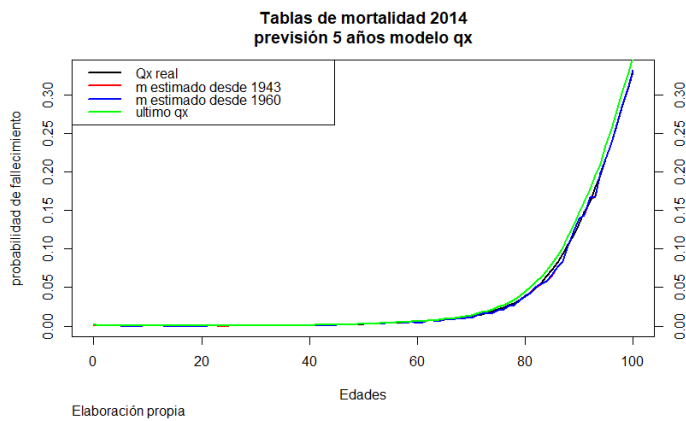
Gráfico 24 Tablas de mortalidad 2014 Modelo\_Qx\_muestra\_desde\_1960 previsión 5 año



En estas tablas agregadas, se ve que los resultados tienen buenos ajustes, aunque las bandas de confianza de 99,5% son excesivamente abiertas para tener una utilidad en la práctica, ya que independientemente de si puede servir como modelo interno, llevaría a un cálculo de capital económico excesivamente elevada.

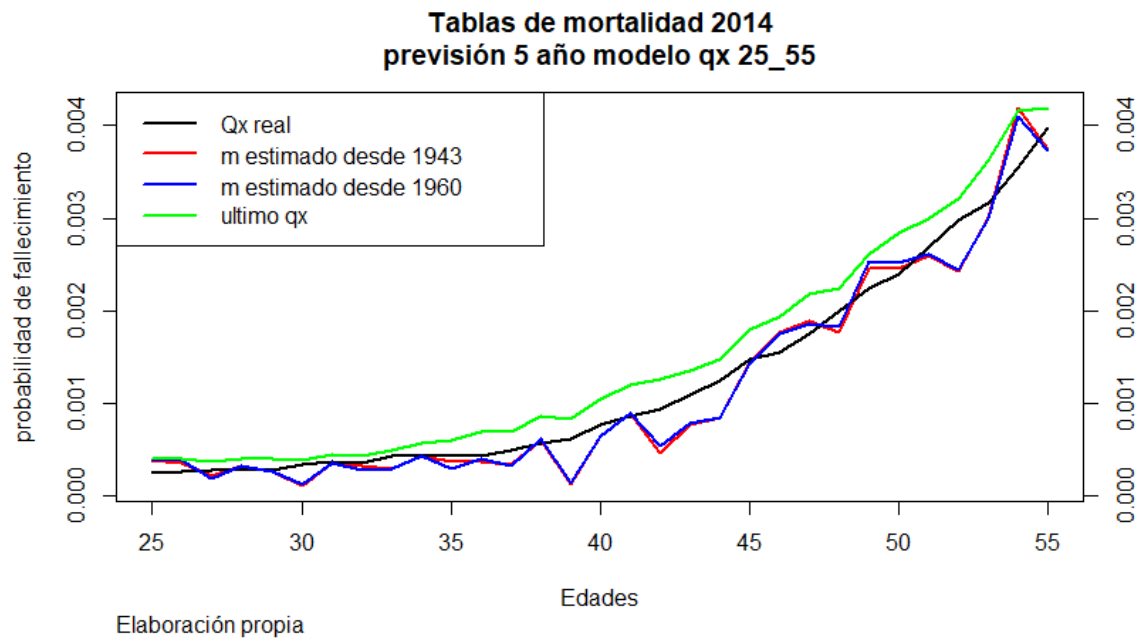


Gráfico 25 Tablas de mortalidad 2014 previsión 5 años modelo qx



En esta gráfica se ve las previsiones tienen un comportamiento parecido a las de un año.

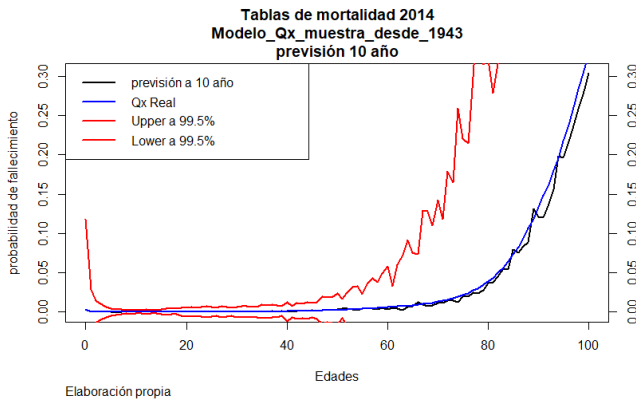
Gráfico 26 Tablas de mortalidad 2014 previsión 5 años modelo qx 25\_55



Sin embargo, cuando se visualiza esta grafica con resultados de tramo de edad 25 a 55, donde concentra mayor interés para el seguro TAR, se ve que el ajuste presenta un deterioro notable, y se observa una volatilidad excesiva. Como punto positivo, se puede afirmar que los niveles de las predicciones son más cercanos a la real que el modelo ultimo Qx.

### 5.5.3 Resultado de previsiones 10 años

Gráfico 27 Tablas de mortalidad 2014 Modelo\_Qx\_muestra\_desde\_1943 previsión 10 año



Adicionalmente se ve que a partir de edad 60 el ajuste es peor y la volatilidad es aún mayor, haciendo necesario otro enfoque de modelación para la proyección a largo plazo de las tablas de mortalidad.

Gráfico 28 Tablas de mortalidad 2014 Modelo\_Qx\_muestra\_desde\_1960 previsión 10 año

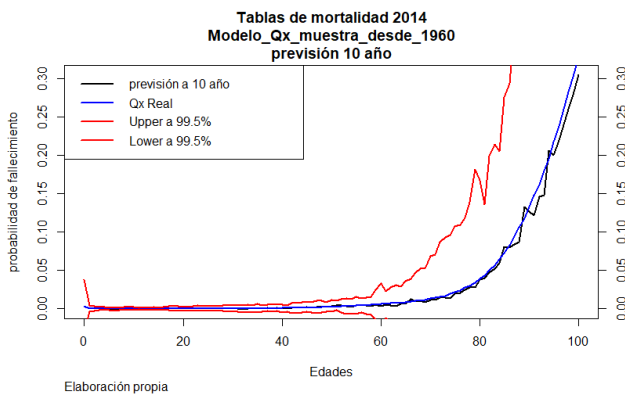
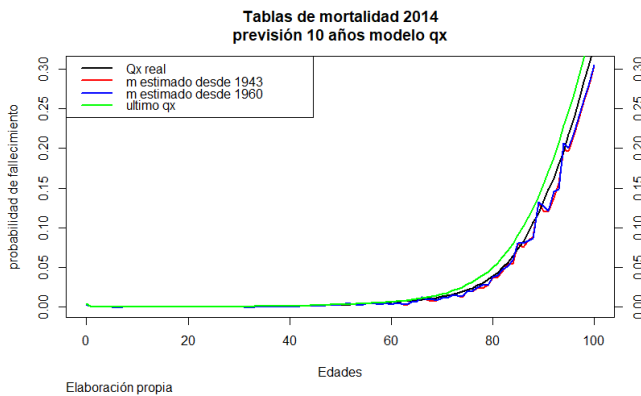
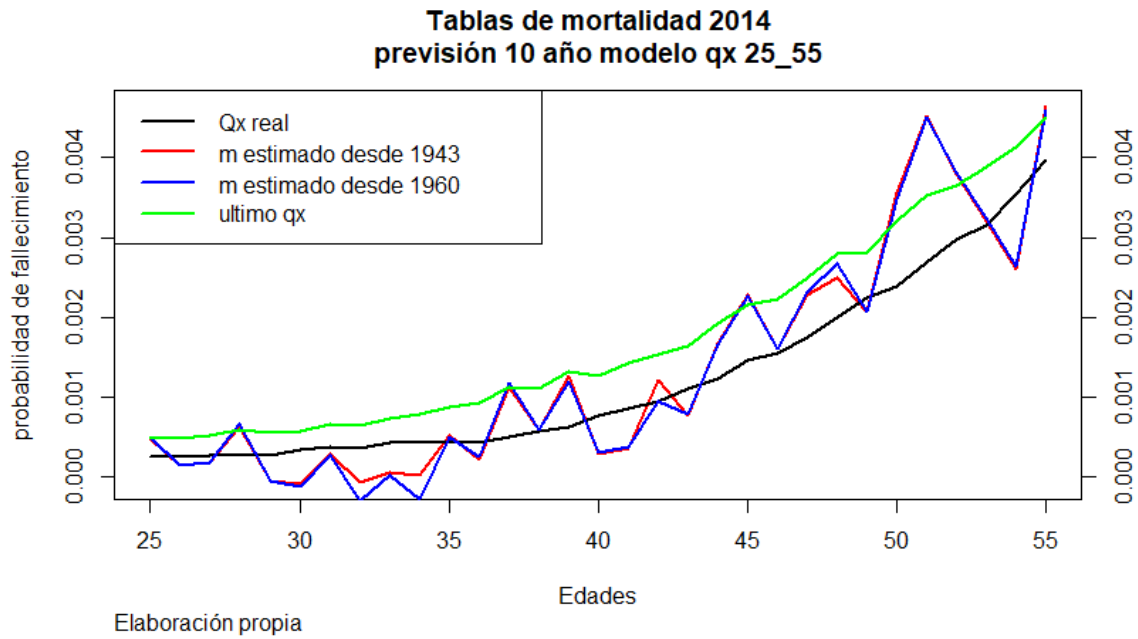


Gráfico 29 Tablas de mortalidad 2014 previsión 10 años modelo qx





Se ve que las previsiones de 10 años de los dos modelos son más precisas que ultimo qx, aunque existen volatilidad que hace la previsión no realista, e impide su implementación en la práctica.

### 5.5.4 Resultado de previsiones 15 años

Se procede a generar las gráficas de previsión a 15 años

Gráfico 31 Tablas de mortalidad 2014 Modelo\_Qx\_muestra\_desde\_1943 previsión 15 año

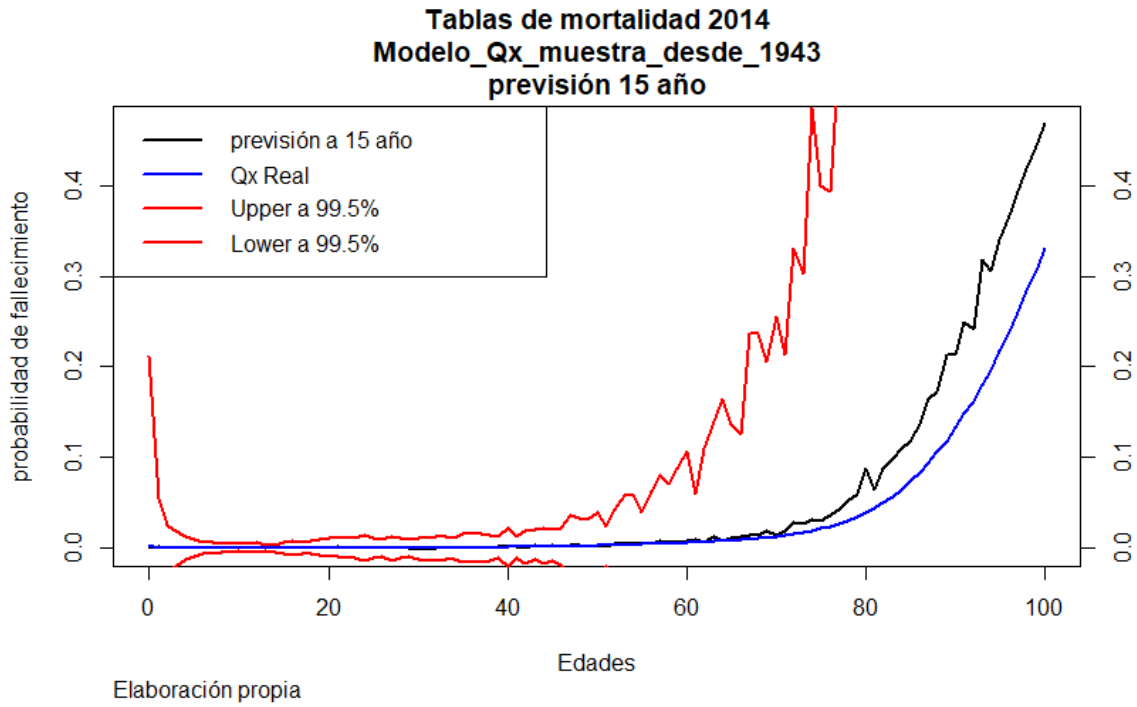
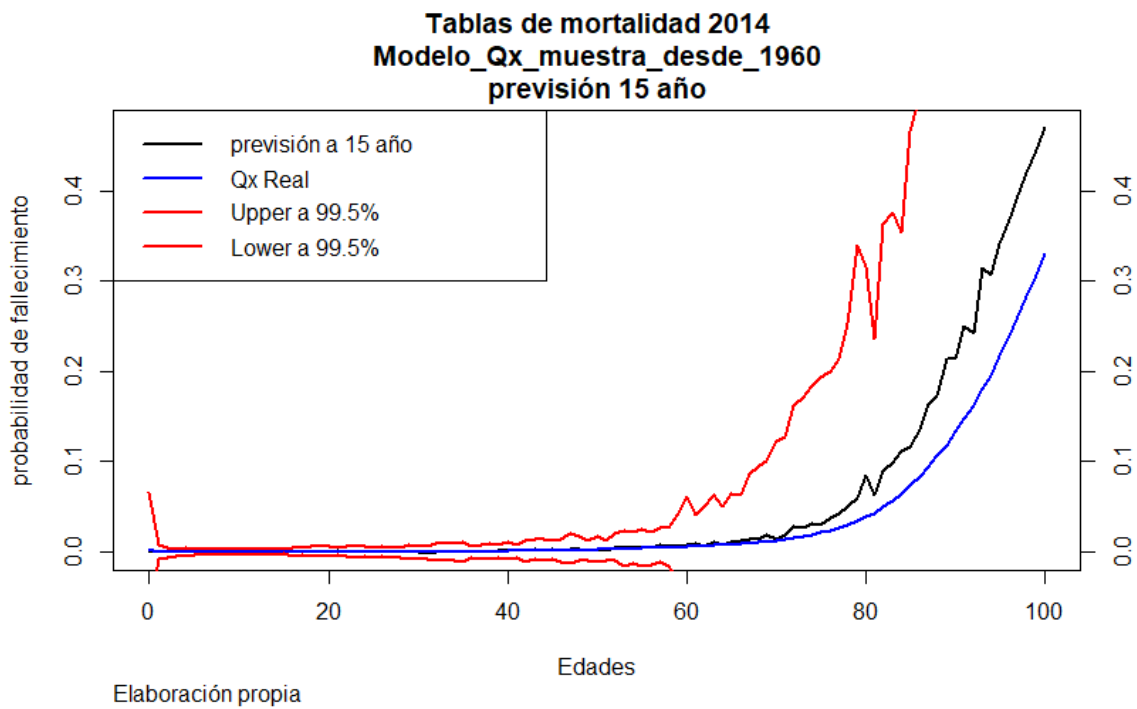
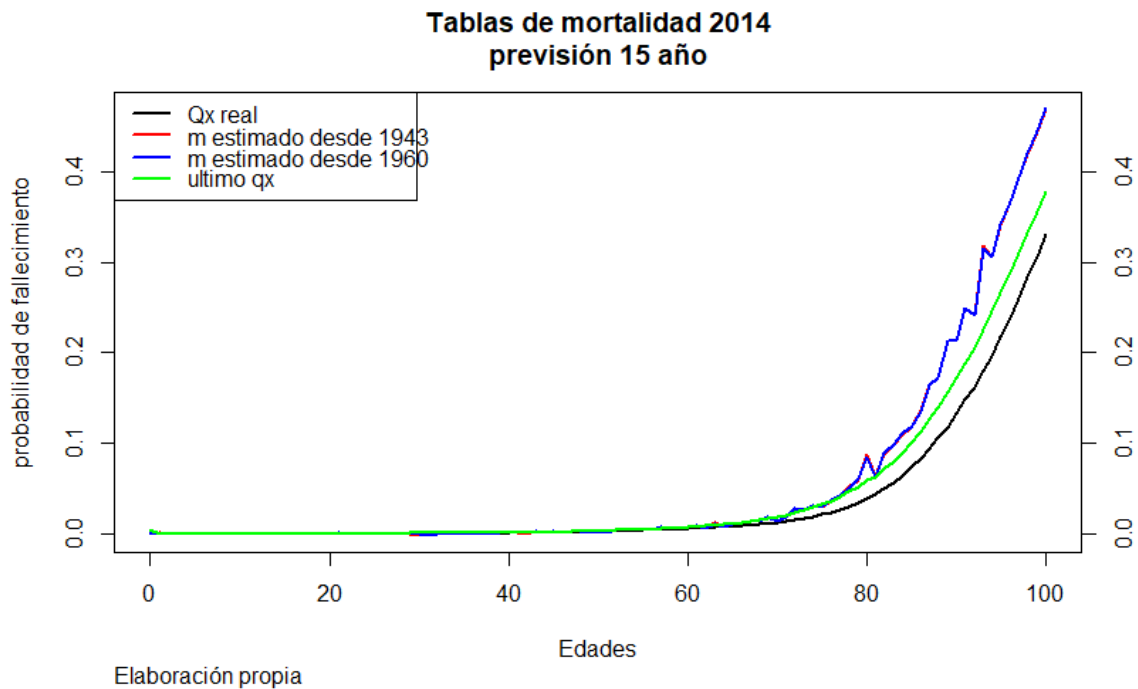


Gráfico 32 Tablas de mortalidad 2014 Modelo\_Qx\_muestra\_desde\_1960 previsión 15 año



Se puede observar que como sucede en las previsiones de 5, 10 años, el ajuste en el tramo de edades avanzada es peor, y las bandas de confianza tienen intervalo excesivo para tener una utilidad práctica.



En esta gráfica, se puede afirmar que el modelo falla completamente en la previsión a 15 años, llegando a ser incluso peor que el modelo último Qx.

## 5.6 conclusiones

se puede concluir los siguientes puntos de los dos modelos ARIMA, uno con muestra desde 1943 y otro desde 1960:

- tienen resultados muy similares, el modelo con muestra desde 1943 tiene una previsión ligeramente más acertada.
- los dos son válidos como modelo interno bajo SolvenciaII, pero son excesivamente conservadores especialmente en las edades mayor que los 65, y llevaría a una considerable sobreestimación de capital económica.
- Los ajustes a corto plazo son bastante preciso, mucho mejor que ultimo Qx, pero cuando se aleja el horizonte temporal de estimación, desde 10 años, el modelo falla en su previsión presentando los desajustes en nivel y una alta volatilidad.

## 6 Modelación y previsión con la índice mejora

En lugar de modelar el  $q_x$  directamente como se ha hecho anteriormente, a continuación, se procede a modelar el índice mejora, este variable es propuesto por Rui Zhou para la elaboración de mapa de calor analizado anteriormente. De esta forma se puede comparar estos dos enfoques de modelación.

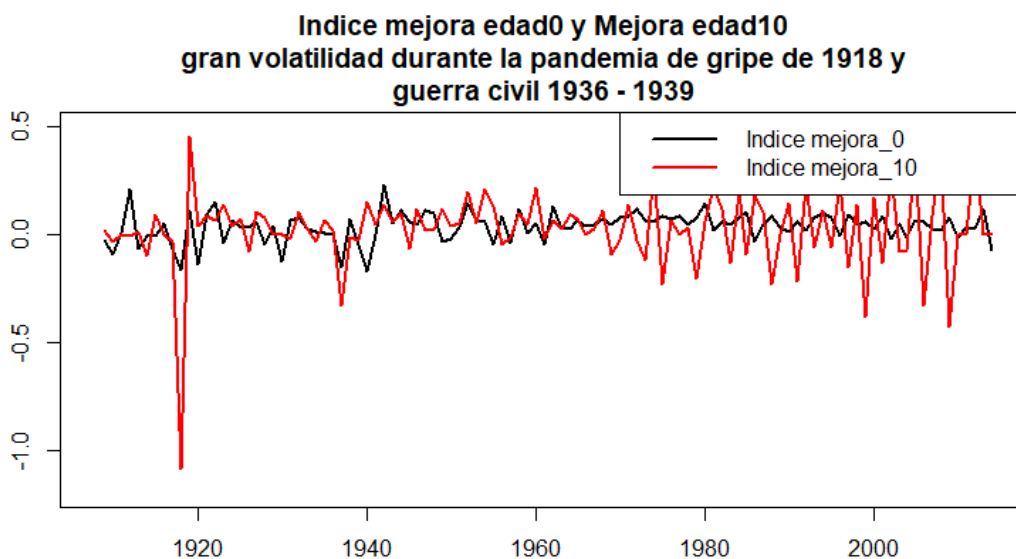
Cabe destacar que este tipo de modelo aporta una ventaja muy importante con respecto al modelo anteriormente realizado; Debido a su construcción de variables, el problema de heterocedasticidad es muchísimo menor.

Con el fin de primar la robustez del modelo, se procede a modelar esta serie con el esquema ARIMA (1,1,0), es decir un solo parámetro AR1 aplicando una sola diferencia.

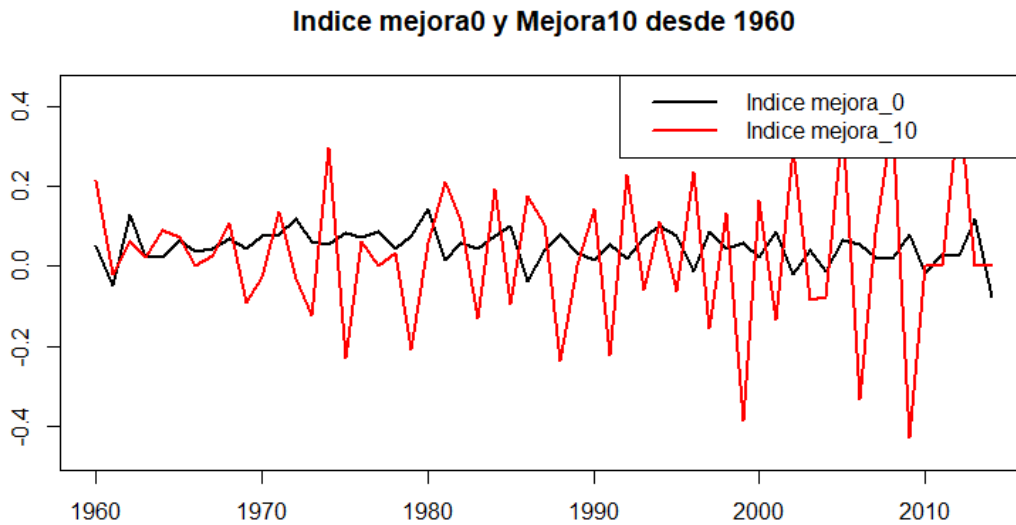
### 6.1 Visualización de series índice de mejora:

A continuación, se visualiza gráficamente estos índices de mejora:

Gráfico 33 Índice mejora0 y Mejora10

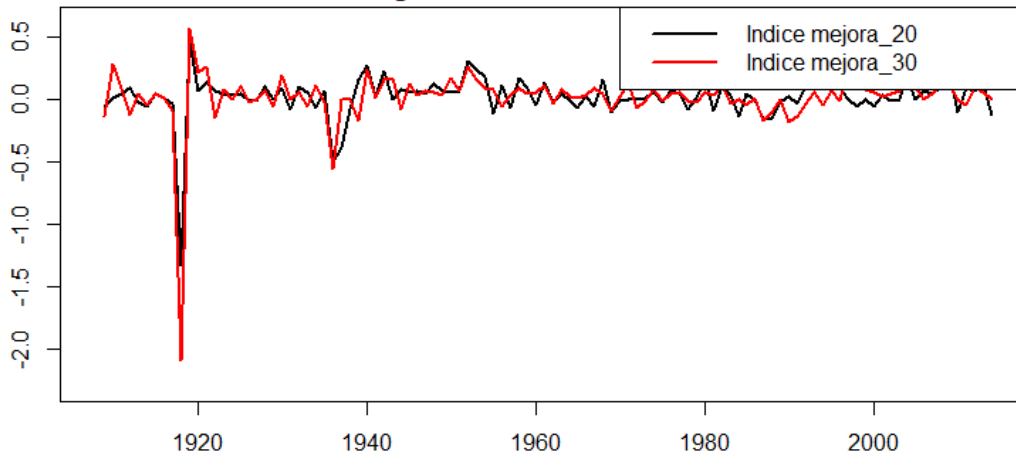


Se aprecia la enorme variación de la serie durante la pandemia de gripe, después se experimenta un periodo de unas décadas de estabilidad, y a continuación surgen periodo de mayor volatilidad marcado por cambio de comportamiento de mortalidad con los VIH, accidente de tráfico, nuevas enfermedades como cáncer, diabetes etc.



En esta grafica de Índice de mejora de 0 y 10 años, sorprendentemente la serie de edad 0 tiene menor volatilidad, se puede interpretar que este grupo tiene mortalidad alta al principio, y el margen de mejora es mayor; la mejora sanitaria ha sido continua y progresiva de forma que no se experimenta variación de magnitud grande con salto significativo. Mientras la serie de mejora de edad 10, al ser una edad de mortalidad muy baja, una pequeña variación en el nivel termina generando una variación considerable en su término porcentual.

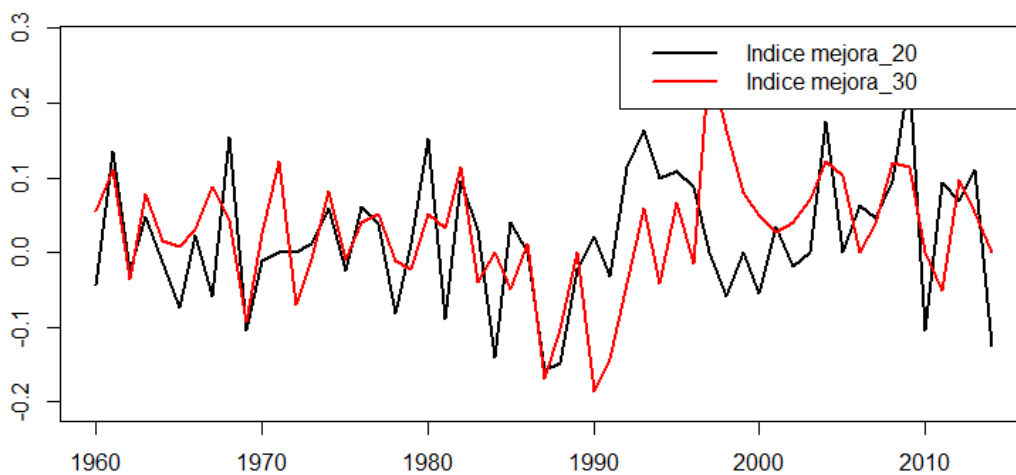
**Indice mejora edad20 y Mejora edad30  
gran volatilidad durante la pandemia de gripe de 1918 y  
guerra civil 1936 - 1939**



Elaboración propia con dato HMDB

Gráfico 35 *Indice mejora20 y Mejora30 desde 1960*

**Indice mejora20 y Mejora30 desde 1960**



Elaboración propia con dato HMDB

En estas graficas se puede ver que los comportamientos de edades 20 y 30 son similares, el primer gran impacto que sufren ambas series es la epidemia de gripe.

El segundo es la guerra civil, ha sido enorme para este grupo de personas, como cabe esperar, este grupo de personas son los combatientes durante la guerra. Una vez terminada de la guerra, la serie es en general positiva, es decir se da la mejora de la longevidad, aunque de forma discontinua.



Gráfico 36 Índice mejora40 y Mejora50

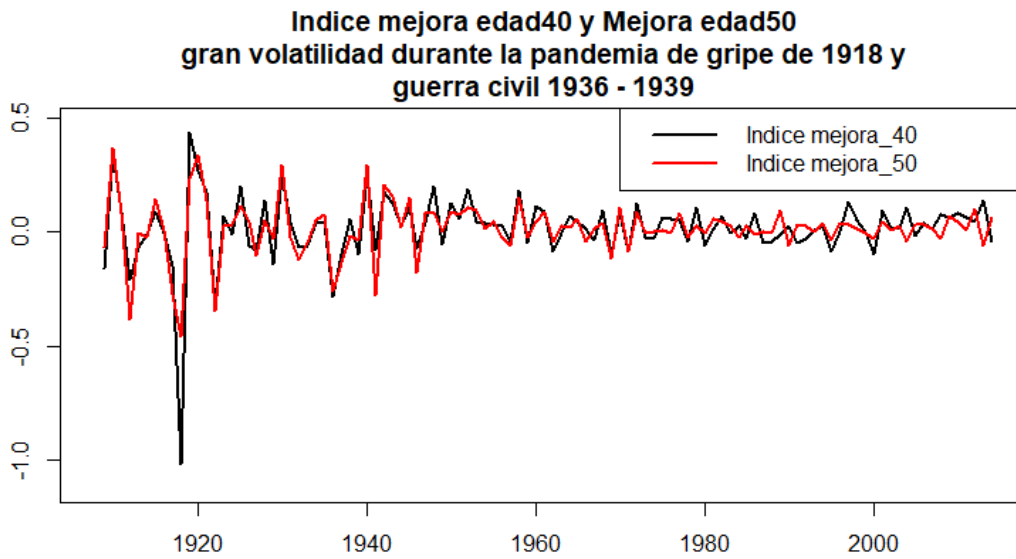
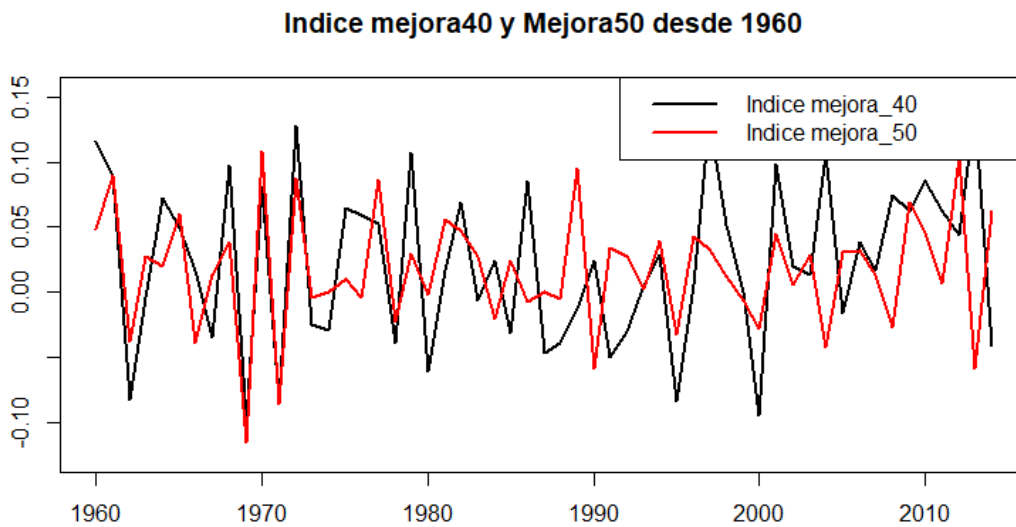


Gráfico 37 Índice mejora40 y Mejora50 desde 1960



Las series de índice de mejora de las edades 40 y 50 tienen algunos comportamientos similares a los tramos de edades de 20 y 30. Sorprendentemente la guerra civil no ha impactado de forma importante a estas series, se pone manifiesto la gran diferencia sanitario-social que hay entre la primera mitad del siglo 20 y la actualidad. Mientras actualmente ese grupo de personas con 40-50 años sigue siendo trabajadores activos de pleno rendimiento, durante los años 40, la época de la guerra, dado las pésimas

condiciones de vida, este grupo de personas podrían ser considerados cercanos a la fase de vejez y su papel en los combates de la guerra civil es más limitado.

Gráfico 38 Índice mejora60 y Mejora70

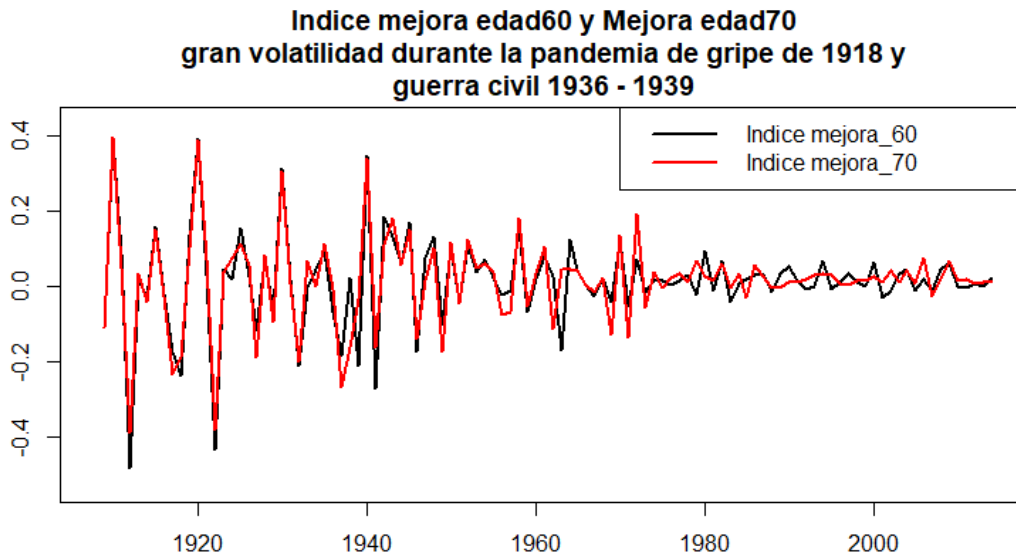
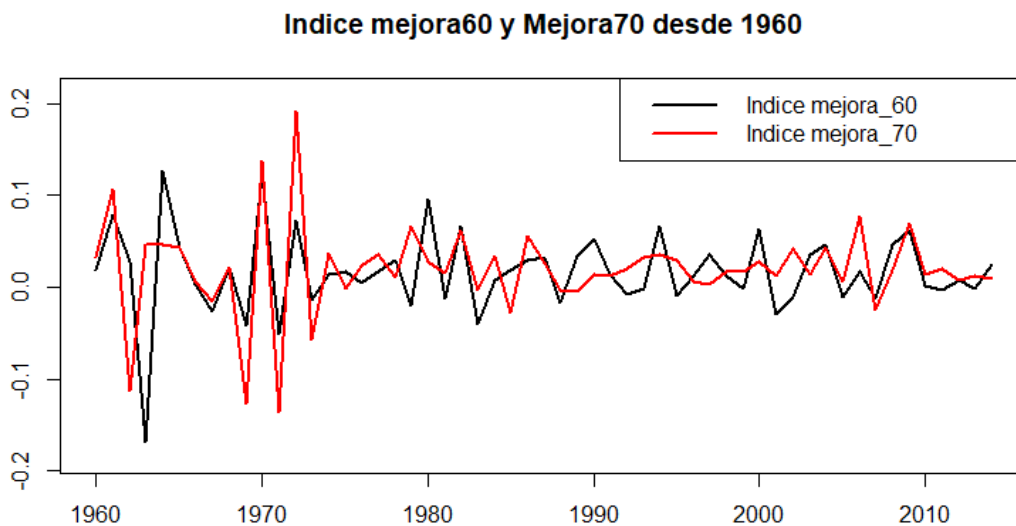


Gráfico 39 Índice mejora60 y Mejora70 desde 1960



en este grupo de personas de avanzadas edades, se puede observar que el impacto de las guerras civil tiene efecto limitado. Una vez concluida la guerra, como indican números estudios sobre la mortalidad-longevidad, la mejora en condiciones de vida y atenciones sanitarias han experimentado mejora continua convirtiéndose en el grupo que más contribuye al aumento de la esperanza de vida.

Gráfico 40 Índice mejora80 y Mejora90

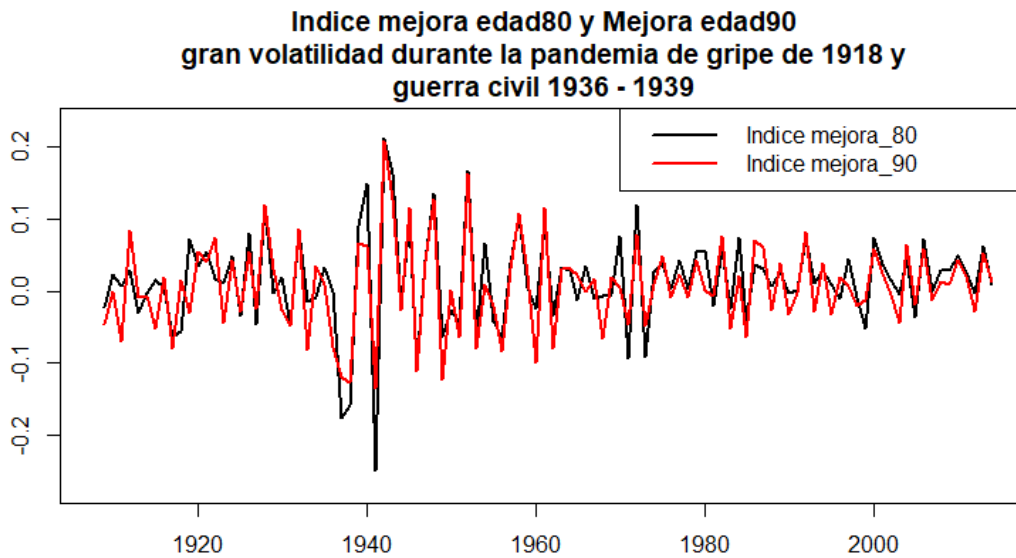
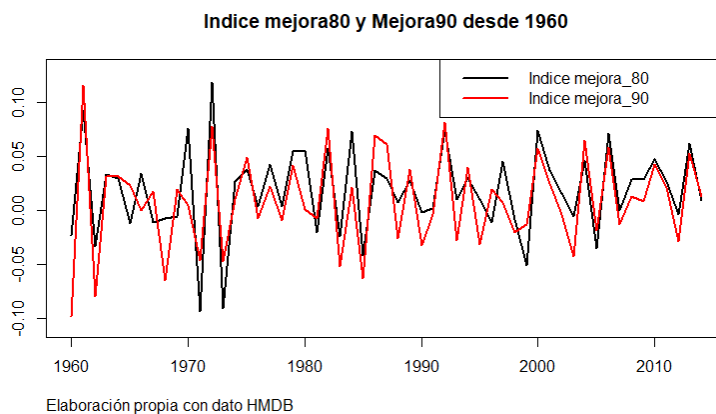


Gráfico 41 Índice mejora80 y Mejora90 desde 1960



De este grupo de personas con edad muy avanzadas, se observa el curioso fenómeno de que la epidemia de gripe no ha generado un gran impacto sobre ellos, por una parte, puede deberse al que las enfermedades producen efectos muy diferentes para las personas adultas y de muy avanzadas edades. Por otra puede deberse a que este grupo de personas eran muy escasos en la época de los años 50, además la recogida de datos carece de rigor, y la interpretación de estas graficas deben proceder con cierta cautela.

## 6.2 resultados

Una vez realizado el modelo, se procede a comparar el resultado de MAPE y error en nivel entre los 4 modelos realizados hasta ahora:

MAPE (25-55años)	ultimo qx	model Mejora 1943	model Mejora 1960	model Qx 1943	model Qx 1960
1 año	-3.88%	0.94%	0.95%	1.36%	1.73%
5 años	-28.85%	-0.01%	0.36%	8.85%	9.30%
10 años	-65.88%	-23.41%	-22.48%	2.17%	6.77%
15 años	-94.85%	-24.90%	-26.52%	64.49%	53.42%

MAPE (0-65años)	ultimo qx	model Mejora 1943	model Mejora 1960	model Qx 1943	model Qx 1960
1 año	-4.54%	-1.12%	-1.06%	0.61%	1.48%
5 años	-28.28%	-5.72%	-5.50%	11.83%	11.54%
10 años	-63.98%	-42.56%	-42.38%	12.34%	15.16%
15 años	-93.28%	-127.21%	-124.51%	16.45%	18.18%

MAPE (56-95 años)	ultimo qx	model Qx 1943	model Qx 1960	model Qx 1943	model Qx 1960
1 año	-1.37%	-0.79%	-0.60%	-0.37%	-0.81%
5 años	-10.43%	2.90%	2.62%	4.22%	4.57%
10 años	-21.72%	5.47%	6.26%	13.63%	11.12%
15 años	-38.60%	-54.53%	-55.16%	-51.20%	-50.57%

Se puede ver en este caso, el hecho de haber empleado distintas muestras en este modelo de índice de mejora no produce diferencia notoria en MAPE entre ellos. La transformación hace que la serie, especialmente la submuestra comprendido entre de años 1940 y 1960 tengan un comportamiento similar al resto, por ello añadir este periodo de dato no impacta de forma notable sobre el modelo y la previsión, al menos no tan importante como sucede en caso de modelo Qx.

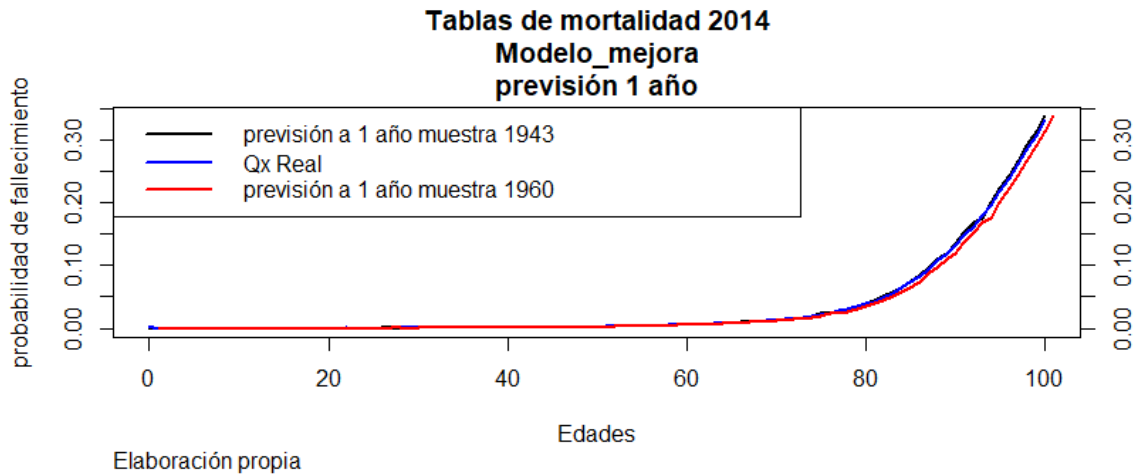
Los resultados como se puede ver son ligeramente peor que los modelos Qx.



### 6.2.1 resultado de previsión de 1 año

A continuación, se procede a visualizar gráficamente los resultados del modelo Índice de mejora.

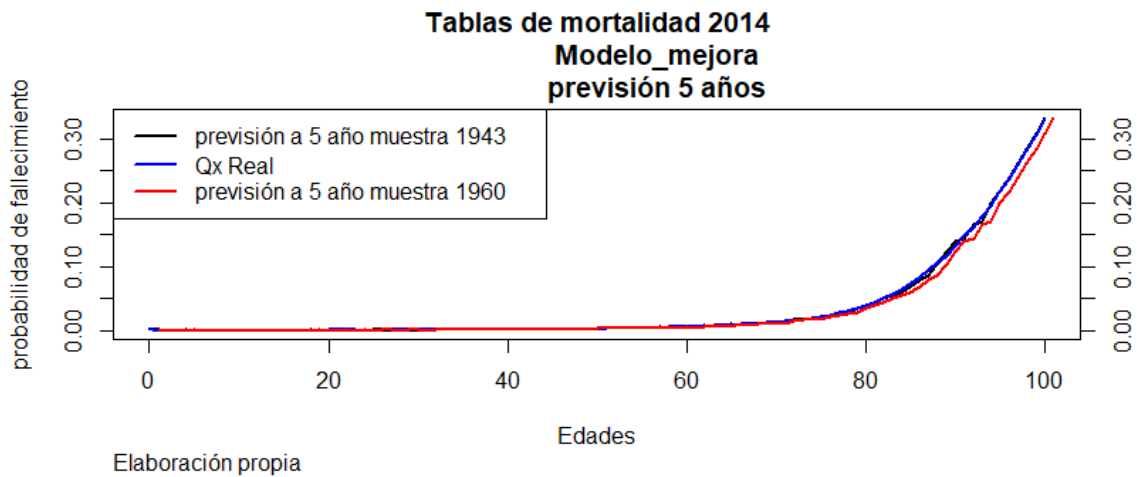
Gráfico 42 Tablas de mortalidad 2014 Modelo\_mejora previsión 1 año



Se puede ver que el ajuste del modelo es muy preciso, aunque en las edades a partir de la edad 70 sí se observa cierto desajuste. Los modelos con muestra desde 1943 o 1960 dan previsiones a un año muy similares.

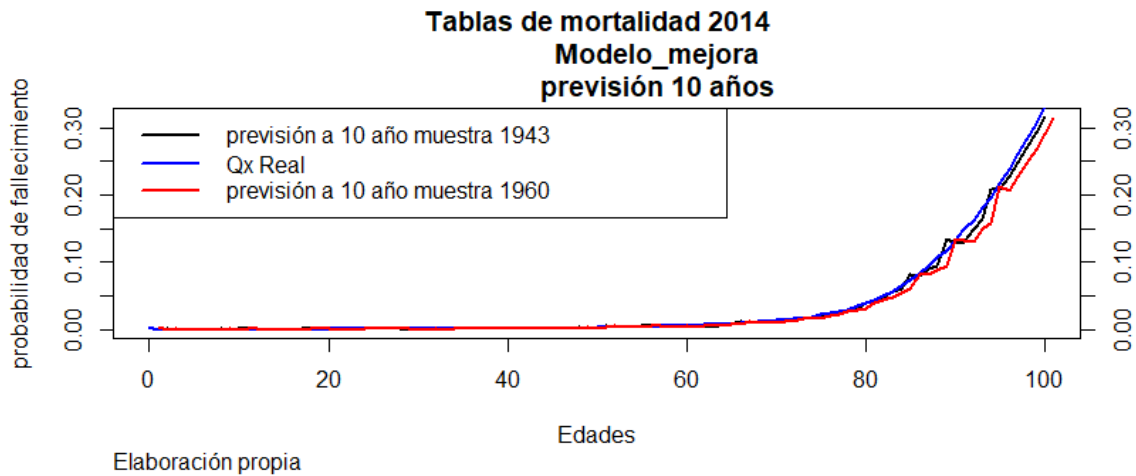
### 6.2.2 resultado de previsión de 5 años

Gráfico 43 Tablas de mortalidad 2014 Modelo\_mejora previsión 5 años



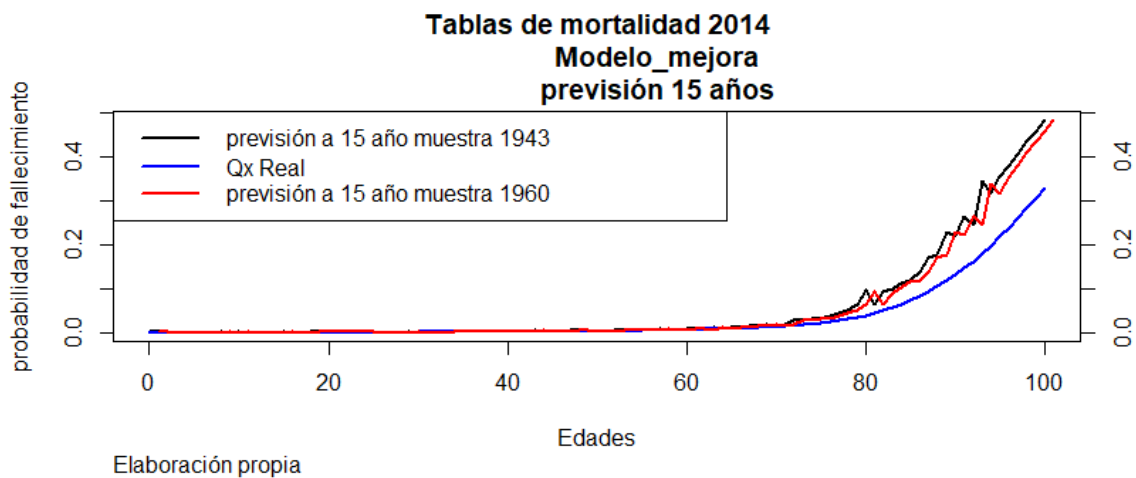
### 6.2.3 resultado de previsión de 10 años

Gráfico 44 Tablas de mortalidad 2014 Modelo\_mejora previsión 10 años



Se puede ver las previsiones de 5 y 10 años como cabe esperar son peor que la de 1 año, y adicionalmente hay algunos valores atípicos en las previsiones en las edades entre 80 y 100.

### 6.2.4 resultado de previsión de 15 años



Se puede ver en este caso, los dos modelos de mejora de distintas muestras (desde 1943 ó 1960) tienen resultados similares entre ellos. Como sucede en los modelos de Qx, en la previsión a corto plazo tienen una gran precisión, y cuando se aleja el horizonte de previsión a 10 o 15 años, los ajustes se deterioran, existe un problema de nivel y adicionalmente se observa el problema de la volatilidad donde las dos edades colindantes pueden tener la mortalidad muy distintas.

### 6.3 Conclusión:

En general se puede decir de los modelos y previsiones basados en el índice de mejora tienen los resultados ligeramente peor que los modelos  $Q_x$ , y las conclusiones de los modelos  $Q_x$  también son aplicables en este caso; los resultados de estimar muestras desde 1943 o desde 1960 no tienen un impacto importante, a corto plazo la previsión es realista y a mayor plazo la previsión falla.

Se ve como sucede en casos de series  $Q_x$ , hay comportamientos atípicos durante la epidemia de 1918 y guerra civil, si bien en el caso de índice de mejora, en lugar de cambio de nivel, se manifiesta de forma de un aumento espectacular de volatilidad, ya que, de forma implícita, durante el proceso de transformar la serie en índice de mejora, se ha aplicado una diferencia.

Los modelos de índice de mejora en esencia son parecidos al modelo con transformación logarítmica, ya que debido a que las tasas de mortalidad tienen valores muy próximos a cero, la transformación logarítmica produce un efecto muy limitado para ellos.

Por otra parte, el índice de mejora es básicamente una ratio, y es similar a la diferencia de datos con transformación logarítmica, precisamente en la modelación de  $Q_x$  se aplica una diferencia más que en la modelación índice de mejora.

Por este motivo, no se vuelve a realizar otro modelo con transformación logarítmica que en condiciones normales son muy común en el contexto de modelación ARIMA.

## 7 Modelo $Q_x$ con restricción y con muestra desde inicio

### 7.1 Modelo con similitud

En el análisis de los 4 modelos anteriores, se puede ver unos de los fallos es la de volatilidad. Como se puede ver en la previsión de 10 años de tramo de edad de 25 a 55 hay una gran volatilidad. Un actuario o analista no daría por válidas las previsiones mencionadas ya que no puede asumir que unos  $q_x$  puedan ser doble que los de las edades colindantes. El motivo de que suceda esto es que en modelos ARIMA no hay ninguna restricción entre ellos, y las pequeñas volatilidades que se da en la mortalidad de una edad concreta una vez proyectada en el futuro, se amplifica produciendo el resultado poco realista descrito anteriormente.

De esta forma, sería conveniente que se imponga alguna restricción en los parámetros, para que un parámetro de una edad no sea muy distinto de el parámetro de las edades contigua.

Ante escasez de paquete que incorpora restricciones en el modelo ARIMA en R y otras aplicaciones estadísticas disponibles actualmente, se procede una forma aproximada recurriendo a algunas técnicas de la modelación de regresión lineal clásica previamente realizando ciertas transformaciones al modelo ARIMA. Se describe a continuación el algoritmo empleado en el presente trabajo para lograr la restricción y similitud entre los parámetros de edad contiguas.

En cada modelo ARIMA de una edad  $x$ , se introduce los inputs de edades colindantes (hasta  $x-5$  y  $x+5$ ) en la regresión:

Los variables de  $x+1$  o  $x-1$  pondera en este trabajo 0.895.

$x+2$  o  $x-2$  pondera  $0.895 * (0.895/2)$

$x+2$  o  $x-2$  pondera  $0.895 * (0.895/2) * (0.895/2)$

$x+5$  o  $x-5$  pondera  $0.895 * (0.895/2) ^{(5-1)}$

la idea es a la hora de impactar sobre el valor de función de verosimilitud y consecuentemente determinar el valor de parámetro, las 5 edades inmediatamente menor que  $x$  tiene el mismo peso que la edad  $x$ , y los 5 mayores también.



## 7.2 Modelo con toda la muestra

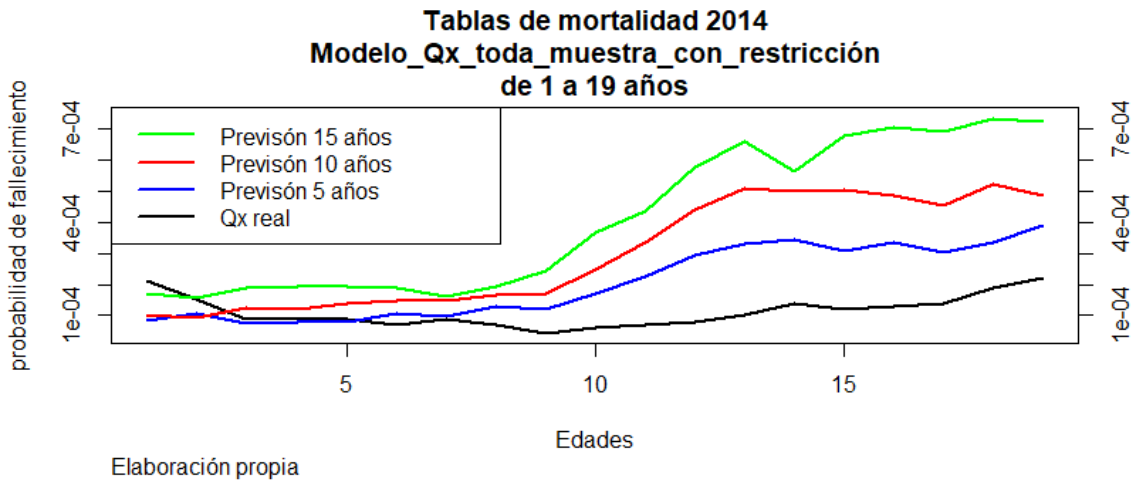
Otra decisión polémica que se ha enfrentado en este trabajo es elegir la muestra óptima para el modelo, ya que en la literatura no hay ningún estudio que proporcione una regla cerrada y consensuada. En este trabajo hasta este punto se ha probado con muestra de estimación desde 1943 y 1960. Para aprovechar la información de toda la muestra, se debe estimar con todos los datos desde su inicio, debido al cambio de estructura progresiva que presentan en las series, el modelo debe asignar a los últimos datos mayor peso. Para lograr este fin se ha procedido de esta forma:

- Los datos durante epidemia y guerra civil no entran en la regresión ya que se trata de fenómenos excepcionales que no se espera que se reproduzca en el periodo de previsión a realizar en el trabajo.
- El peso del último dato disponible es 1, y cuando se aleja un año hacia el pasado, el peso de dato se ve reducido al multiplicarse a  $(2^{0.5})^{(-1/15)}$ , de esta forma el peso del penúltimo dato es  $(2^{0.5})^{(-1/15)}$ , y de hace 2 años es  $((2^{0.5})^{(-1/15)})^2$ , así sucesivamente hasta que se asigne el peso a toda la muestra siendo el de hace n años  $((2^{0.5})^{(-1/15)})^n$
- La idea es el dato de hace 15 años pesa la mitad del dato actual

Cabe señalar que en caso de edad 0, no se añade de la regresión ninguna edad inferior, no se emplea ninguna técnica para corregir esta aparente excepción anómala, ya que no es ninguna incoherencia conceptual; la edad 0 presenta un comportamiento particular, la similitud con las edades 1,2,3 son menores, y en este caso el peso de edad 0 para impactar sobre el parámetro suyo es 0.5 frente a 0.33 en otros casos.

Se visualiza los resultados:

Gráfico 45 Tablas de mortalidad 2014 Modelo\_Qx\_toda\_muestra\_con\_restricción de 1 a 19 años



En esta grafica donde se visualiza las previsiones de distintos plazos con la serie real, se puede observar que mientras se aleja el horizonte de previsión se tiende a incrementar la sobreestimación de la previsión, este sucede debido a que el modelo ARIMA reproduce fielmente la estructura de los últimos años, y tanto las gráficas de las series históricas como numeroso estudios académicos indican que existen rendimiento decreciente en la mejora de la mortalidad infantil, y la contribución de este grupo de población al aumento de esperanza de vida es menor en las últimas décadas.

Gráfico 46 Tablas de mortalidad 2014 Modelo\_Qx\_toda\_muestra\_con\_restricción de 20 a 39 años

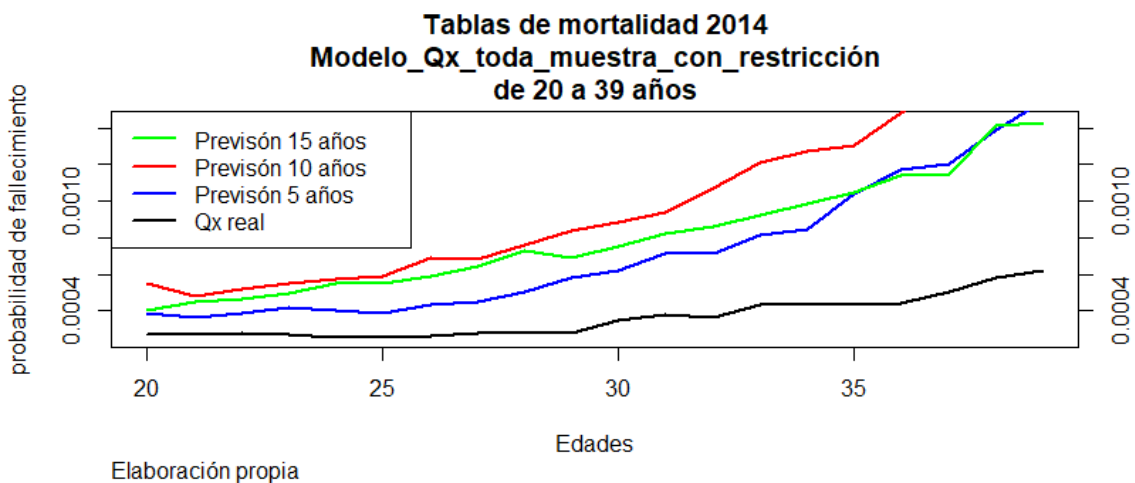




Gráfico 47 Tablas de mortalidad 2014 Modelo\_Qx\_toda\_muestra\_con\_restricción de 40 a 59 años

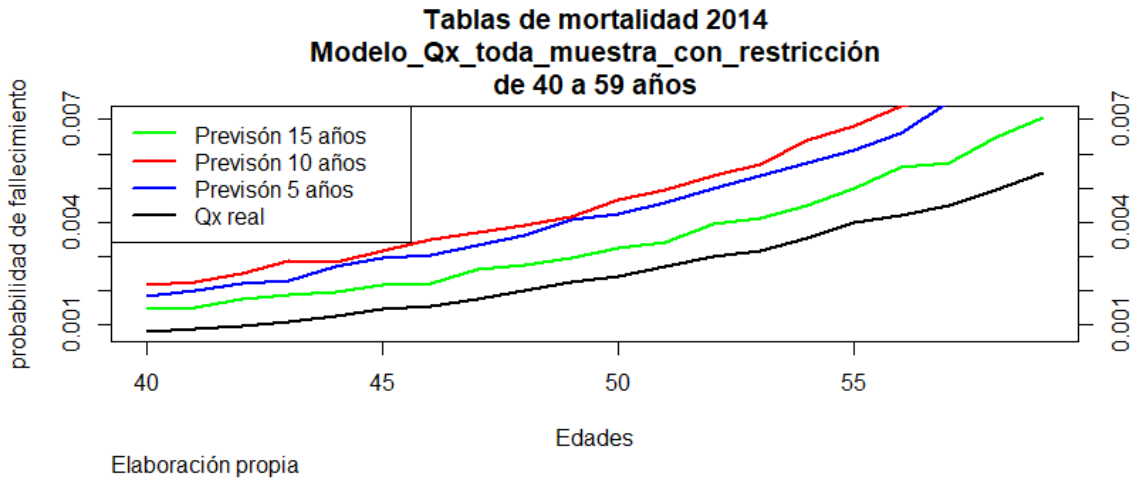
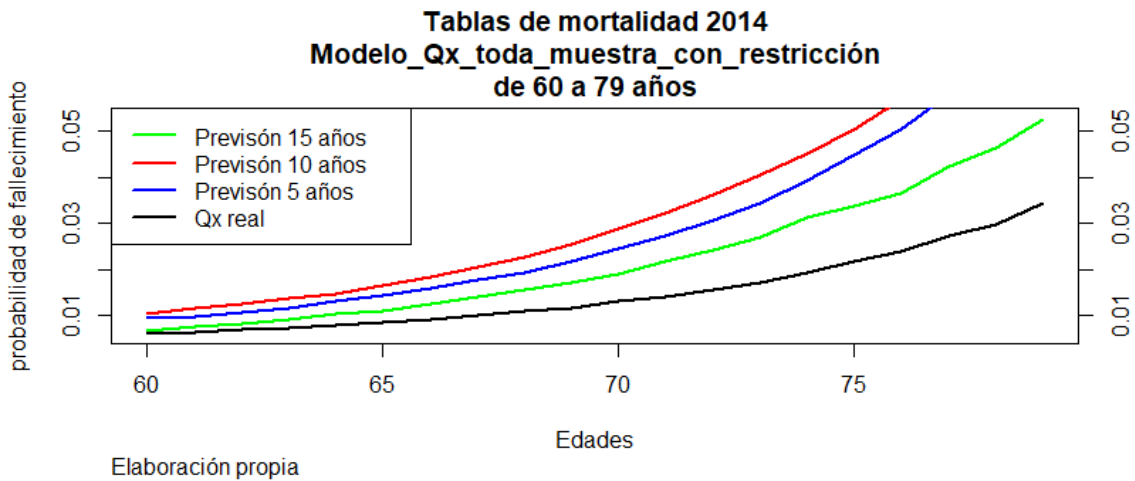


Gráfico 48 Tablas de mortalidad 2014 Modelo\_Qx\_toda\_muestra\_con\_restricción de 60 a 79 años



\* se ha aplicado una combinación de otra proyección de modelo ARIMA, dado que no afecta al resultado final, se omite este detalle que será explicado en el proceso de combinación con LeeCarter en adelante.

En estas graficas se ve que el algoritmo que impone similitud ha dado el resultado esperado resolviendo el problema de la excesiva volatilidad, todas las series de previsión tanto de corto como de largo plazo son suaves en el sentido de que la previsión de una edad no se diferencia sustancialmente de sus edades colindantes.

En vista de estos resultados, con este algoritmo se da por resuelto el problema de la volatilidad quedando pendiente de resolver el otro problema de nivel.

### 7.3 conclusiones

Tanto modelos con muestra desde 1943 como desde 1960, se presenta el problema de volatilidad, para su resolución, se implementa con éxito la estimación con restricción. Se logra resultado progresivo sin salto significativo, pero ninguno de los 3 modelos ha sido capaz de proyectar a más de 5 años de forma realista. El modelo ARIMA tiene estructura



demasiada sencilla para inferir y predecir el cambio de nivel, en este caso la estructura del comportamiento de mortalidad viene definido por varias décadas, y el modelo ARIMA solo es capaz de reproducir la estructura de los últimos pocos años, y es una herramienta poco eficaz para predecir el cambio de nivel de la tabla de mortalidad.

## 8 Modelo Lee-Carter dinámico

Para lograr una predicción con nivel correcto se va a preceder el modelo de Lee-Carter dinámico descrito en el apartado marco teórico.

El tamaño de la ventana de muestra elegido es de 30 años, con la ventaja de que, entre este espacio temporal relativamente reducida, el impacto de cambio de estructura es limitado. En este trabajo concreto se ha tenido en cuenta de que la proyección llega a 15 años, y se espera que la estabilidad de parámetros inferida de un periodo de 30 años sea suficientemente robusta para la proyección de 15 años.

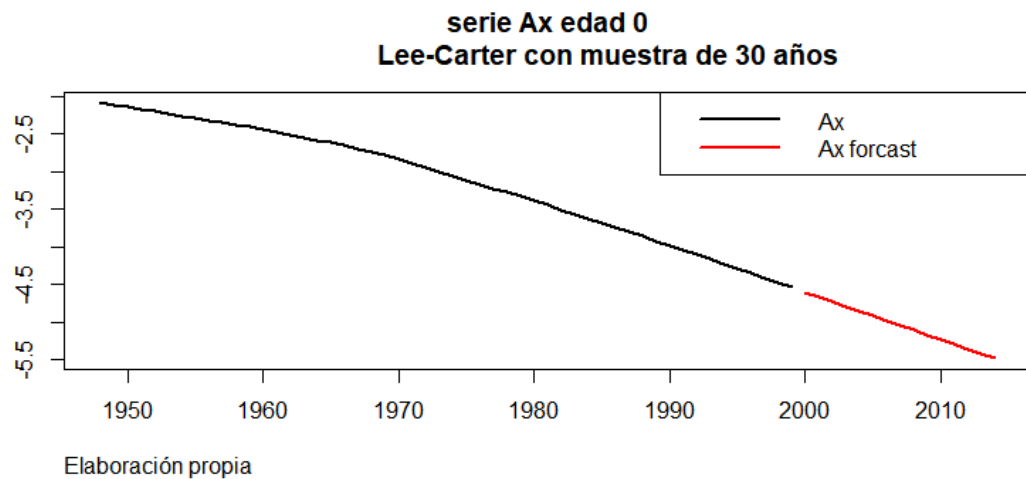
### 8.1 Estimación de las series $a_x$ , $b_x$ y $k_t$ , junto con su previsión a 15 años

Como se explica en el marco teórico del presente trabajo, este modelo estima 3 vectores de parámetros,  $a_x$ ,  $b_x$  y  $k_t$ . Para inferir lo más fiel posible la estructura de la mortalidad a lo largo de todos los años, se procede a estimar el modelo dinámicamente, que en este caso consiste en abrir una ventana de tiempo de 30 años, se estima el modelo Lee-Carter, se mueve dicha ventana un año hacia pasado, se vuelve a estimar, así sucesivamente hasta final de la serie. (se omite los detalles técnicos que han sido analizado en el apartado de marco teórico)

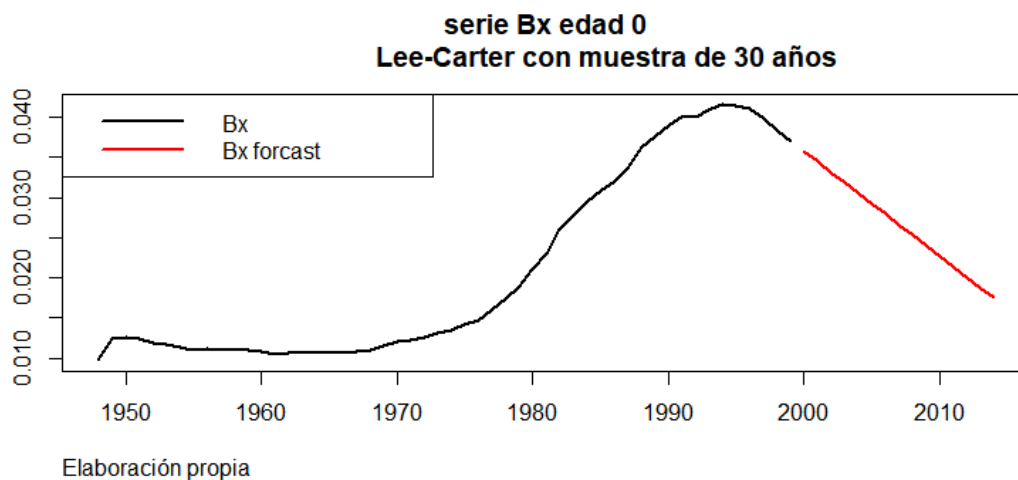
Por cada edad, se obtiene una serie de vector de  $a_x$ ,  $b_x$ .

Para poder generar previsión de tablas de mortalidad, se procede a proyectar esas series de parámetros con un modelo ARIMA sencillo (1,2,0) por el motivo similar en los casos de modelo Qx y índice de mejora buscando la misma robustez.

Se visualiza algunos de ellos junto con su previsión a 15 años:

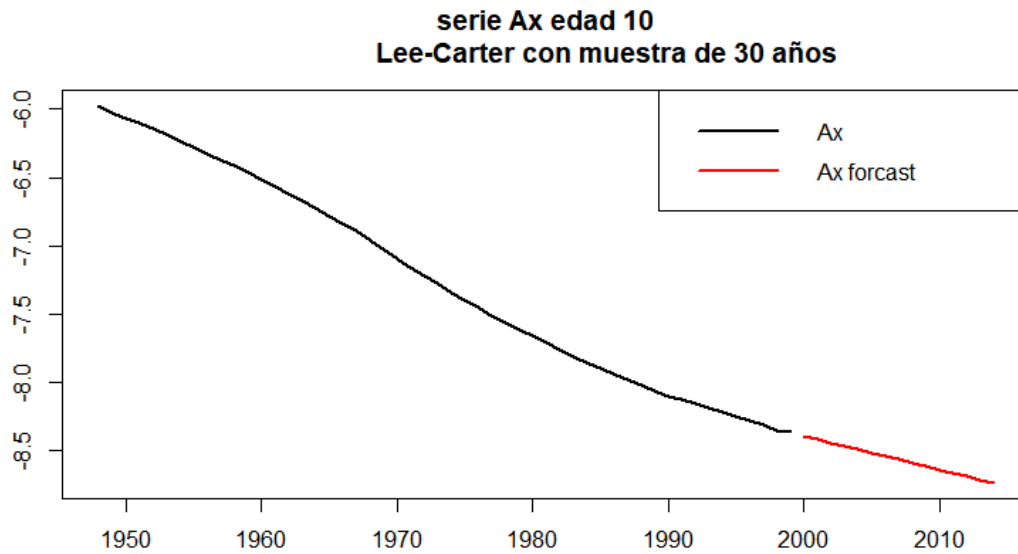
Gráfico 49 serie  $A_x$  edad 0 Lee-Carter con muestra de 30 años

Se puede decir que este parámetro  $a_x$  que es una aproximación de la media móvil de la mortalidad con la transformación logarítmica, como se puede observar en la gráfica, la serie es bastante lineal, con un cambio de pendiente en alrededor de 1968. Y dado su suavidad, sí es factible la previsión de 15 años.

Gráfico 50 serie  $B_x$  edad 0 Lee-Carter con muestra de 30 años

En esta serie se observa que durante los años 1970 hay un cambio de comportamiento, ese cambio coincide en el tiempo con la finalización de la guerra civil. Después la serie empieza a crecer hasta llegar su nivel máximo en los años 1995.

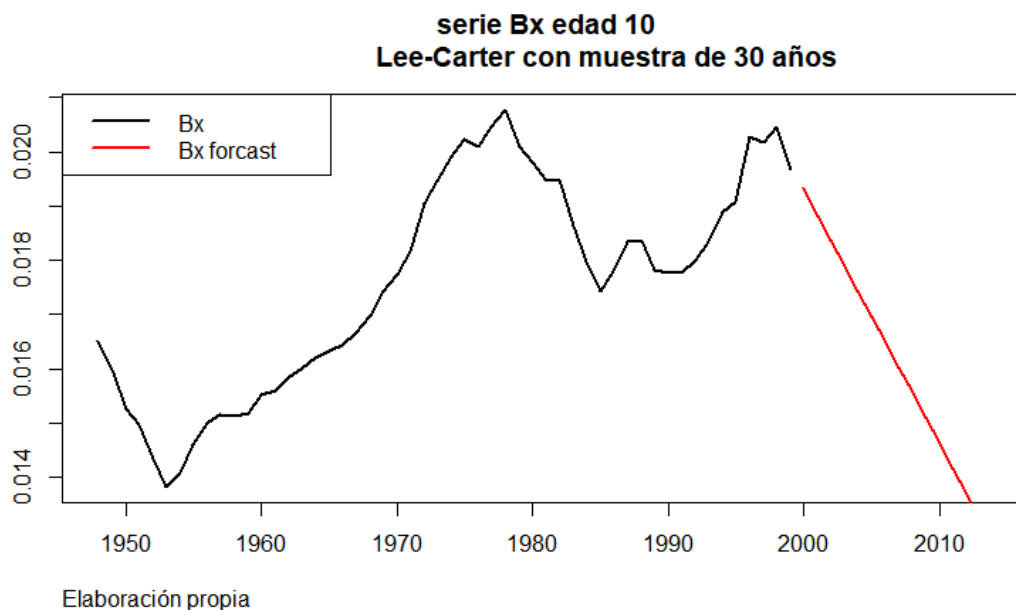
De la previsión se puede ver una bajada continua de la serie, algo extraordinario que en principio se asume que es correcto en el trabajo ya que, la bajada continua de la tasa de mortalidad de los recién nacidos sí muestra un efecto de rendimiento decreciente, es decir su segunda derivada con respecto al tiempo es negativo tal como indica la gráfica.



Elaboración propia

De esta gráfica se puede decir que existe cierta similitud al comportamiento a la edad 0, se ve cierto cambio de pendiente en los años 70, y en general la serie es bastante lineal, y la pendiente tiende a reducirse conforme transcurre el tiempo.

La bajada continuada de la previsión durante los 15 años sí es algo razonable, ya que el fenómeno de reducción de la tasa de mortalidad de forma apenas interrumpida en los últimos 45 años son esperable. (se habla de últimos 45 años no sólo de 15 años debido a que el modelo emplea media de los últimos 30 años, y el modelo de hace 15 años lleva la información de hace 45.)



En esta serie también se ve la clara evolución del parámetro, esta evolución indica que no se debe introducir toda la muestra en el mismo modelo. Como sucede en la gráfica de edad 0, de la previsión se puede ver una bajada continua de la serie, algo no fácil de encajar en los estudios actuales sobre la mortalidad. El motivo está en las características del modelo ARIMA, ya que el modelo reproduce la estructura de los últimos años sin tener en cuenta los comportamientos del dato más alejado en el tiempo. Antes esta anómalo observado en la gráfica, se procede a modelación sin emplear ninguna técnica específica para tratar esta incidencia por estos dos motivos.

El primero, en este trabajo dado que en el modelo existen otros parámetros para poder reducir el efecto de esta bajada aparentemente excesiva.

El segundo, un rendimiento decreciente en la bajada de tasa de mortalidad infantil que sí podría darse, su segunda derivada en la tasa de mortalidad con respecto al tiempo sí debe ser negativa, algo bastante coherente con los estudios de longevidad que argumenta que las principales de contribución en la mejora de la esperanza de vida ya no provienen de los de tramo de edad baja, sino de tramo de edad avanzada.

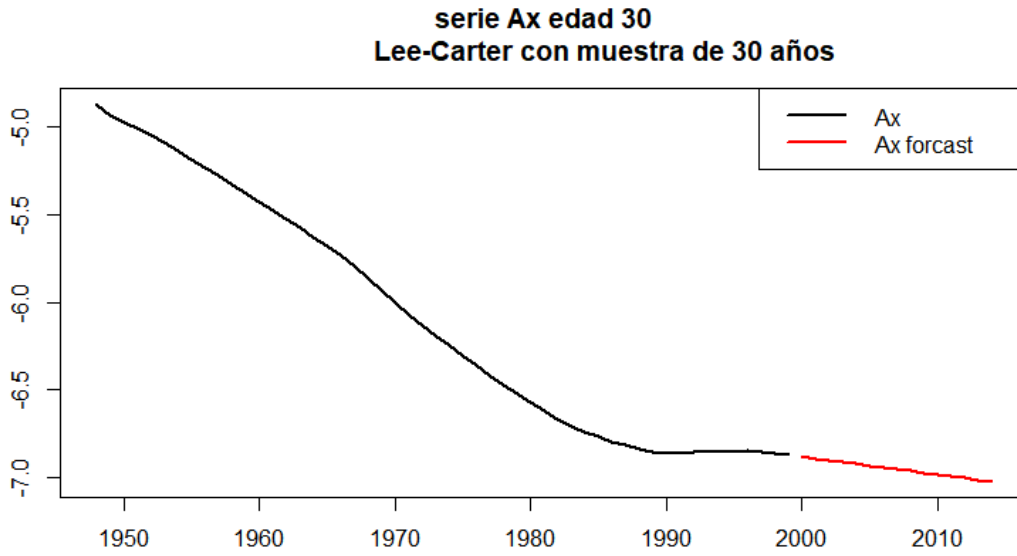
Durante los años 1985 y 1995, la bajada de este parámetro podría justificarse por el comportamiento de la mortalidad asociada con la aparición de VIH y otros factores de mortalidades como accidentes de tráfico en las personas de edad adulta, este produce gran





efecto en las edades adultas, y el algoritmo de la optimización puede modificar sobre la serie de parámetros B de persona de menor edad también.

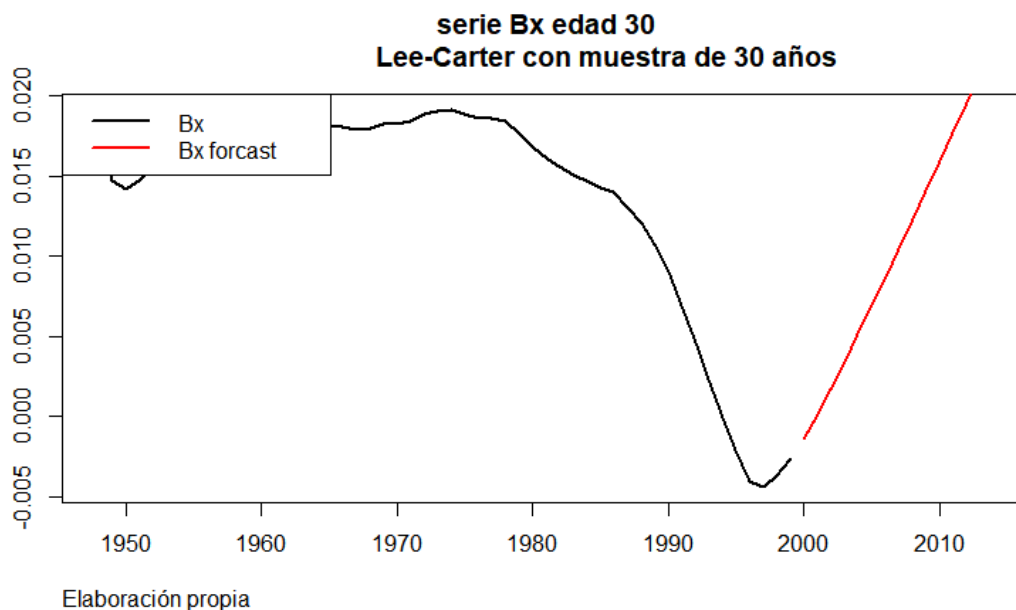
Gráfico 53 serie Ax edad 30 Lee-Carter con muestra de 30 años



Elaboración propia

De análisis de esta gráfica, se puede afirmar que es otra prueba contundente de que existe evolución en parámetros para justificar la implementación de un modelo Lee-Carter dinámico, es más, en este tramo el comportamiento de subida en el año 1995 es una anomalía que debe ser estudiado en un futuro (como verá, hay un desajuste en este tramo de edad en previsión a largo plazo).

Gráfico 54 serie Bx edad 30 Lee-Carter con muestra de 30 años



Se puede ver un enorme cambio de nivel hace que el modelo dinámico sea necesario, la bajada en los años 1995 puede deberse a la aparición de nuevos factores de mortalidad como VIH y otros como accidente de tráfico en los años anteriores.

Sin embargo, el posible fenómeno de VIH no será sostenible en el tiempo, tampoco la mejora que experimenta años después del este episodio, sin embargo, el modelo ha reproducido este fenómeno en su previsión y ha generado un aumento de nivel espectacular, en la gráfica siguiente donde muestra el ajuste del modelo final implementado se puede observar que conforme que se aleja en el tiempo la previsión, la sobreestimación es cada vez mayor. Este anómalo detectado durante este análisis de la proyección de parámetros también se ha manifestado en la esperanza de vida en el tramo de 20-39 años. Sin lugar de duda, ese es uno de los puntos de futuras mejoras importante, durante el presente trabajo debido a la orientación de la proyección se centra en el cálculo actuarial de la renta de las edades avanzadas, se emite este detalle técnico.

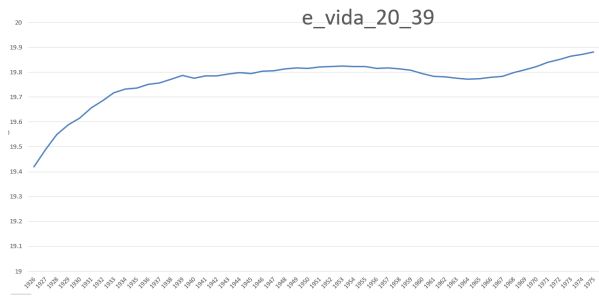
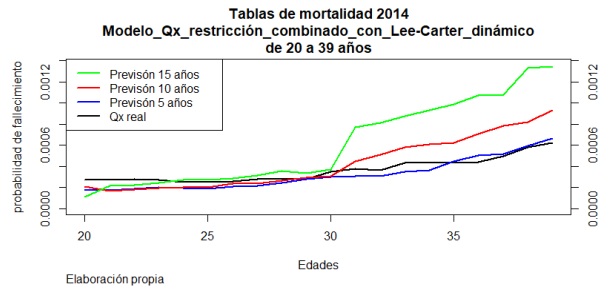
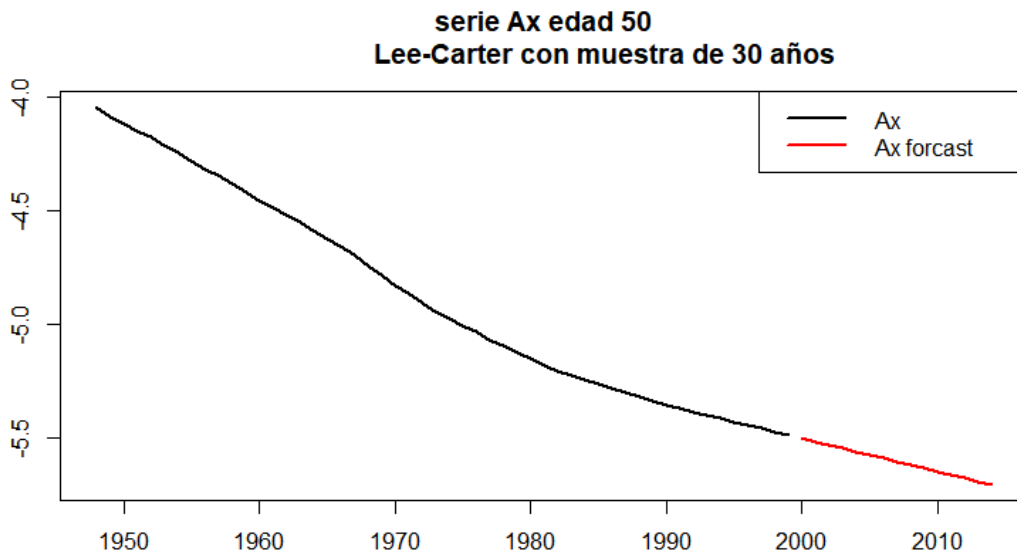
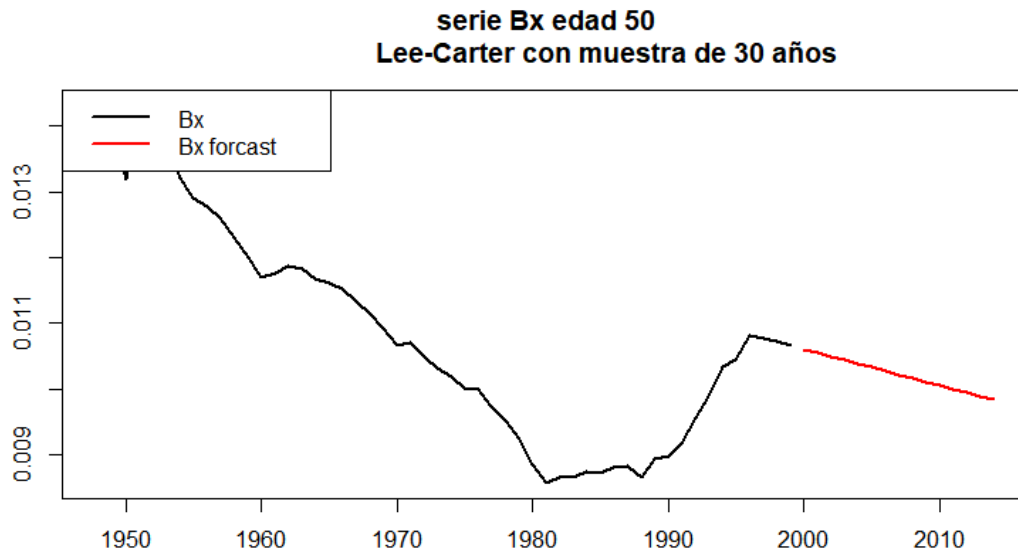


Gráfico 55 serie Ax edad 50 Lee-Carter con muestra de 30 años



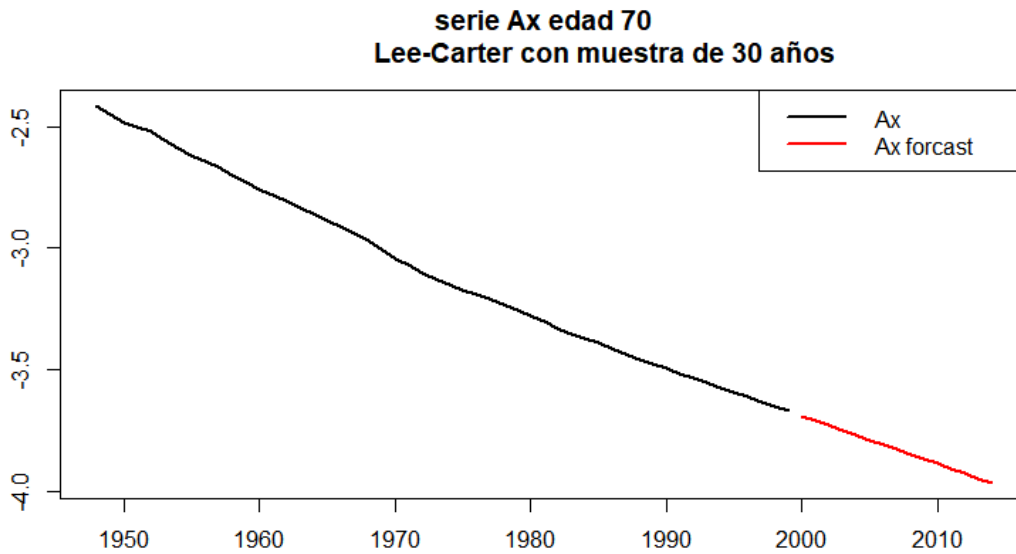
Se puede ver el comportamiento similar a otras series, la previsión también parece realista y razonable. La estructura parece más fácil de proyectarse en el futuro, se espera previsión de este tramo sea mejor que los tramos anteriores analizados. (y el resultado final confirma que es verdad).



Elaboración propia

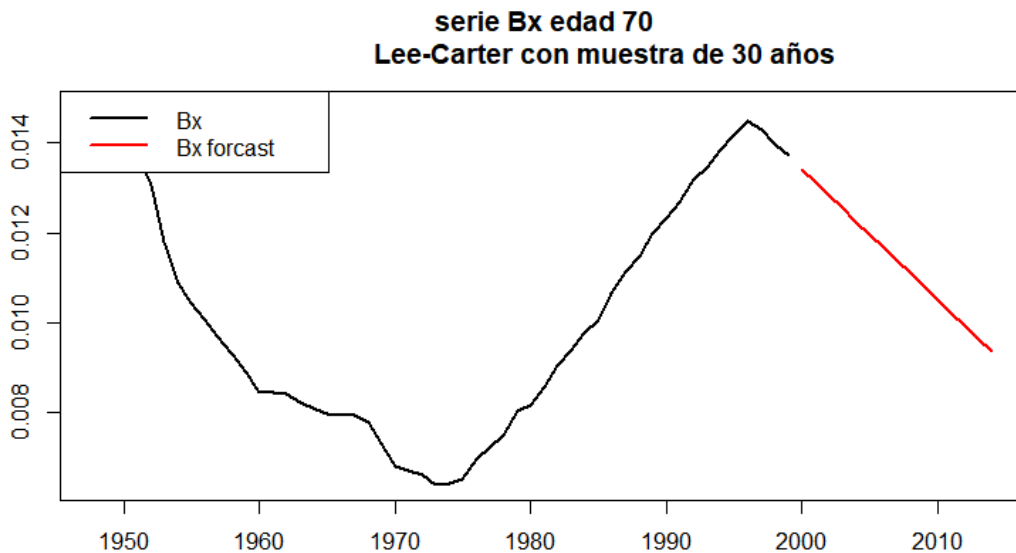
Se puede ver un comportamiento similar a otras series  $b_x$ , se puede afirmar que existe correlación bastante fuerte entre los parámetros de distintas edades.

La serie se cambia dos veces de su tendencia, la previsión no parece muy fácil de acertar, sin embargo, el problema es menor de lo que sugiere la gráfica a simple vista, ya que la serie no cambia mucho de nivel, la tendencia bajista que indica la previsión, aunque errara, sólo está produciendo la serie a la baja apenas 10% de sus niveles previos a la previsión.

Gráfico 57 serie  $A_x$  edad 70 Lee-Carter con muestra de 30 años

Elaboración propia

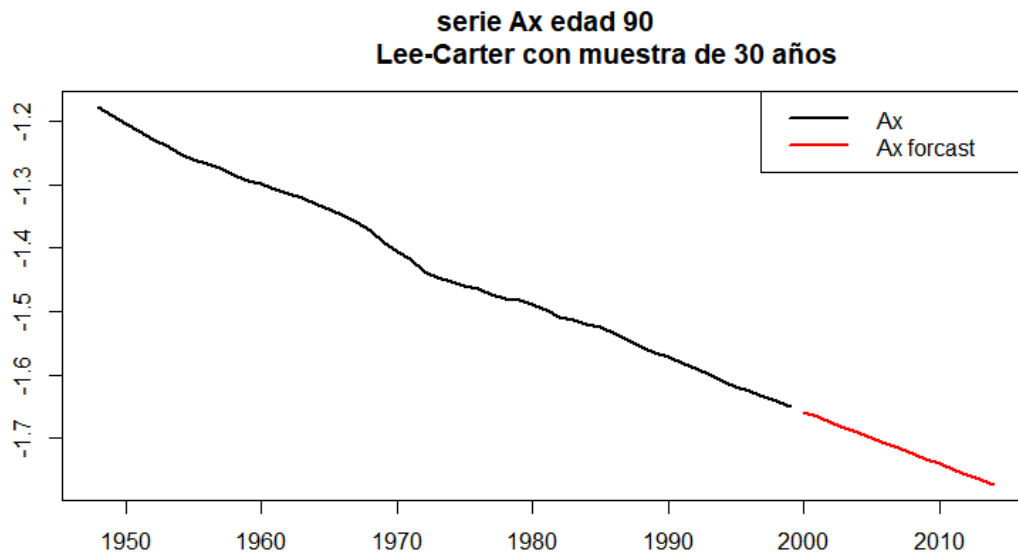
La serie es muy lineal, la previsión de ésta sería fácil y fiable, es la consecuencia directa de la bajada de la mortalidad lineal de este tramo de población.

Gráfico 58 serie  $B_x$  edad 70 Lee-Carter con muestra de 30 años

Elaboración propia

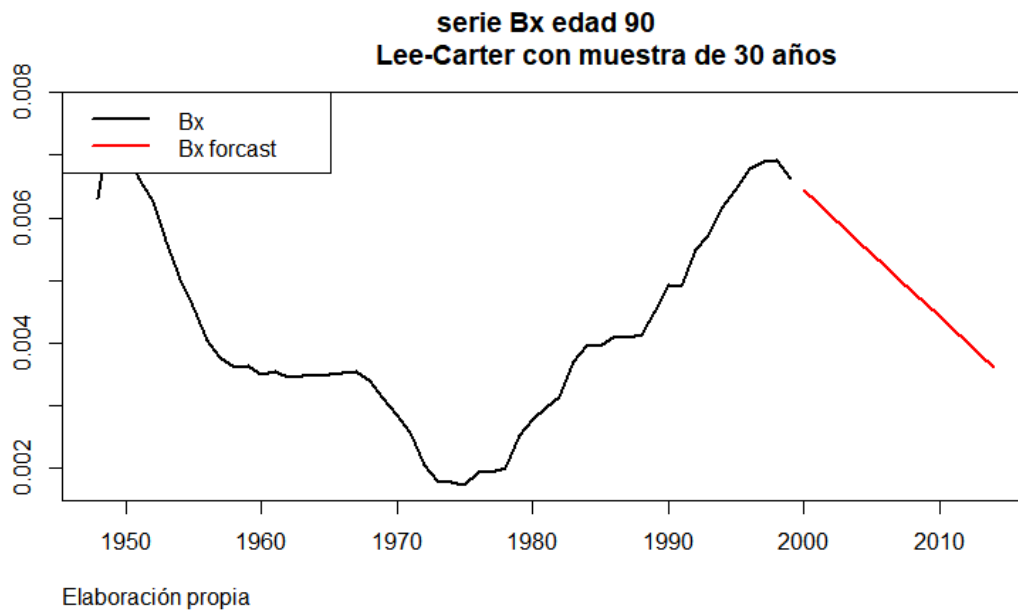
El comportamiento de esta serie como sucede en otras  $b_x$ , en comparación con las  $a_x$  es más complejo de interpretar, y la previsión es menos fiable. Sin embargo, se puede decir que la banda de variación de la serie es pequeña y la previsión de este parámetro no tiene mucho error en su porcentaje.

Gráfico 59 serie Ax edad 90 Lee-Carter con muestra de 30 años



Elaboración propia

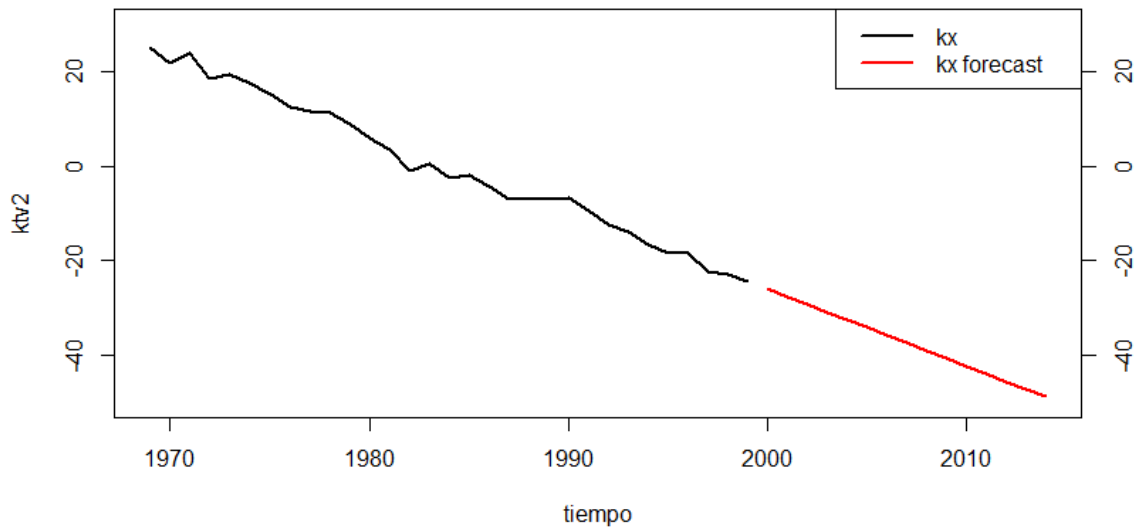
Como se anticipaba en la gráfica de esperanza de vida del presente trabajo, la previsión del nivel representado por este parámetro es bastante realista y fiable, dado la enorme estabilidad que se manifiesta la serie.



En esta grafica se puede ver que la serie es suave, y la previsión recoge la estructura bajista de los dos últimos datos.

En general la previsión de los parámetros  $a_x$  parece más fácil y realista, ya que este parámetro tiene el papel similar a la constante en un modelo de regresión lineal (sea clásica o generalizada), y en estos modelos siempre es más fácil inferir nivel que la variación.

Aunque debido a que las series son latentes que por definición no es posible compararse con una serie empírica, aunque se ha referido sucesos en el pasado.

Serie parámetro  $k_t$  y su previsión a 15 años

Elaboración propia

La previsión de la serie de  $k_t$  también es aceptable, y sin apenas cambio de tendencia, es algo muy coherente con el hecho de que la esperanza de vida sigue en incremento continuo desde hace varias décadas en España, y la previsión indica la misma tendencia. Esta visión del comportamiento de la mortalidad es apoyada por numerosos estudios sobre la población española.

Cabe mencionar que la suma de  $B_x$  en la previsión a 15 años es 1.050877, una desviación aceptable, en la previsión final, se ha normalizado para que suman 1.

## 8.2 Resultados Lee-Carter Dinámico:

Una vez realizadas las previsiones de parámetros, se llega a la previsión de las tablas de mortalidad de 5,10,15 años:



Gráfico 62 Tablas de mortalidad 2014 Modelo\_Lee-Carter dinámico de 1 a 24 años

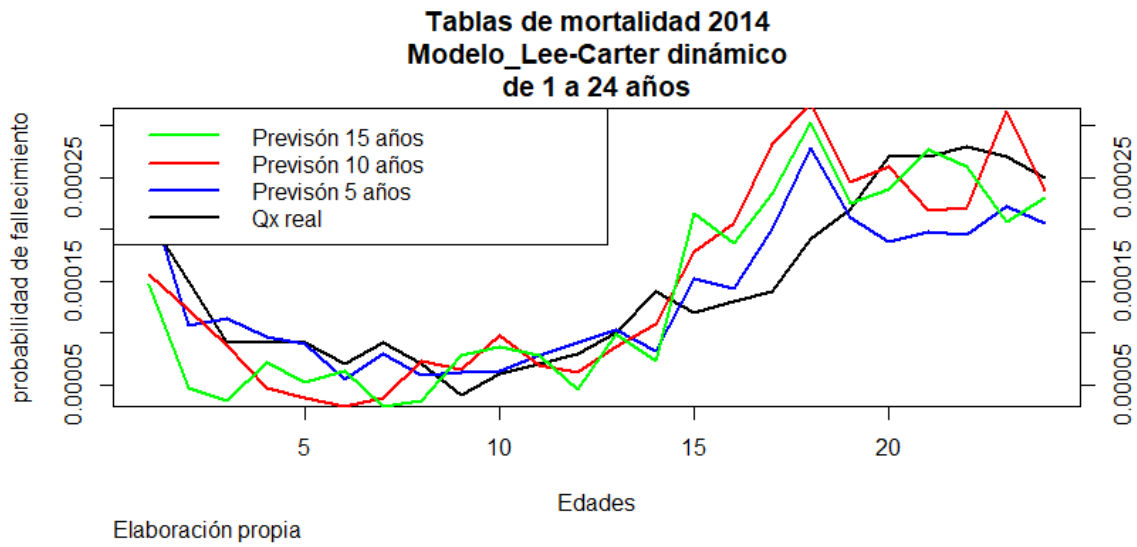


Gráfico 63 Tablas de mortalidad 2014 Modelo\_Lee-Carter dinámico de 25 a 65 años

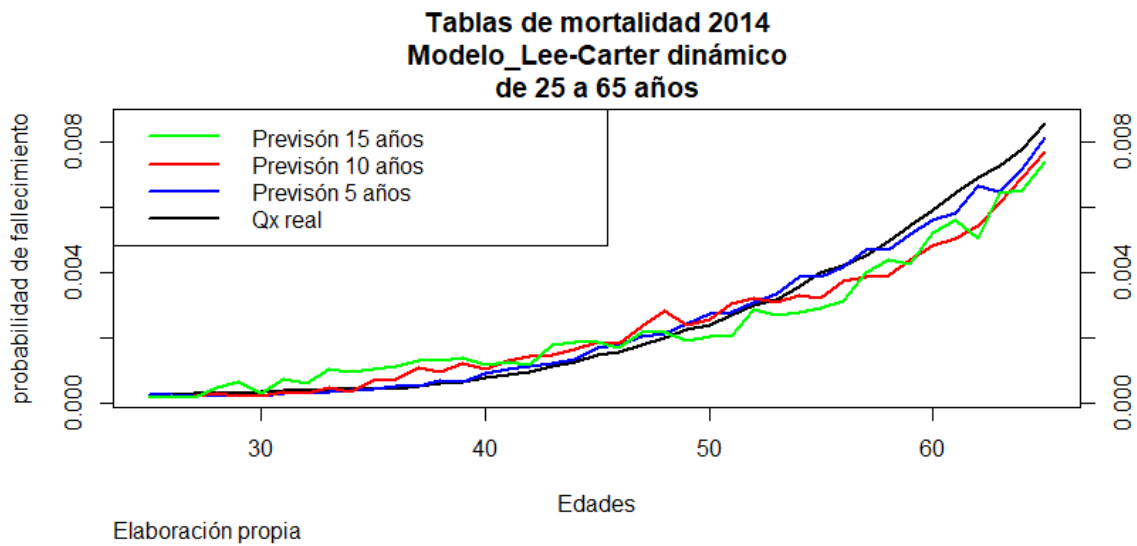
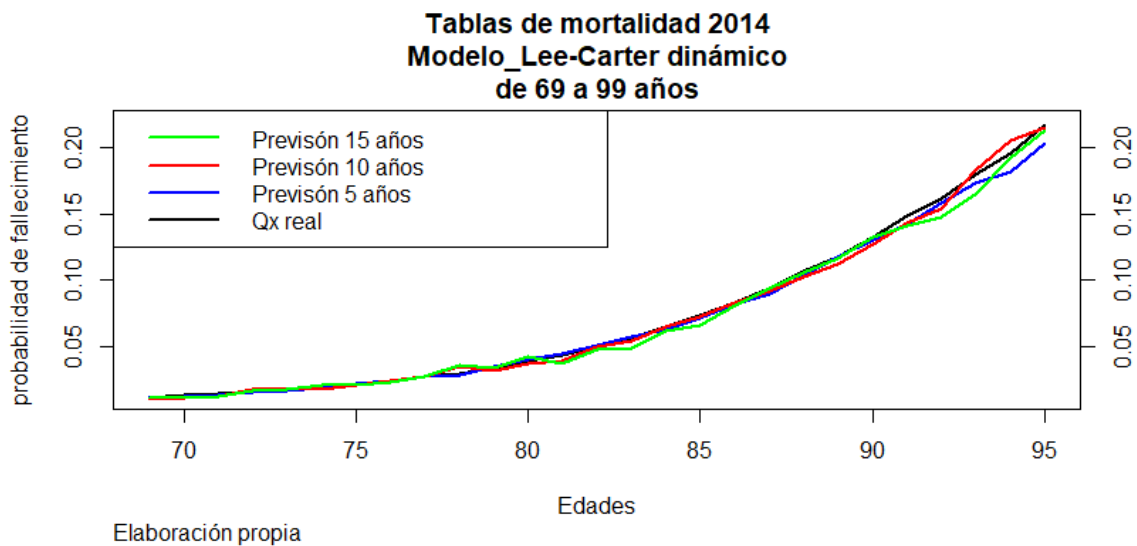


Gráfico 64 Tablas de mortalidad 2014 Modelo\_Lee-Carter dinámico de 69 a 99 años



### 8.3 conclusión

En la previsión que genera el modelo Lee-Carter dinámico, curiosamente se ve que, el ajuste es mejor en tramo de edad muy avanzada. Este hecho podría deber a que en la estimación de parámetro mejora  $K$  en el modelo Lee-Carter, se pone especialmente énfasis en lograr el buen ajuste para este grupo de personas de edad avanzada. Ya que, en este tramo de edad las mortalidades son mayores, y tanto su variación en el tiempo como los errores también son de magnitudes mayor, y su impacto sobre el algoritmo de optimización, de su capacidad para modificar el valor de parámetros también es mayor. Durante el proceso de estimación de los parámetros, se ha “priorizado al ajuste de este tramo”. Otro motivo es al tener nivel más alto, proporcionalmente los errores no se ven resultados gráficamente.

Es algo que se podría haber evitado recurriendo a alguna técnica de modelación, no se realiza ningún tratamiento especial para este respecto en este trabajo, ya que este tramo es donde concentran mayor interés para la proyección a largo plazo para el estudio de las rentas actuariales, mientras para grupo de edad menores el mayor interés incide en la tarificación de TAR, es decir predicción de mortalidad a corto plazo.

En general se puede decir que el modelo Lee-Carter genera previsión a largo plazo más realista que el modelo ARIMA. Como inconveniente, se observa que existe volatilidad elevada en la previsión, similar a que se observa en la primera fase de modelación del modelo ARIMA.

## 9 Modelo Qx con restricción y muestra pesada combinado con modelo Lee-Carter dinámico

A continuación, se procede a generar el último modelo del presente trabajo para las previsiones de mayor de un año.

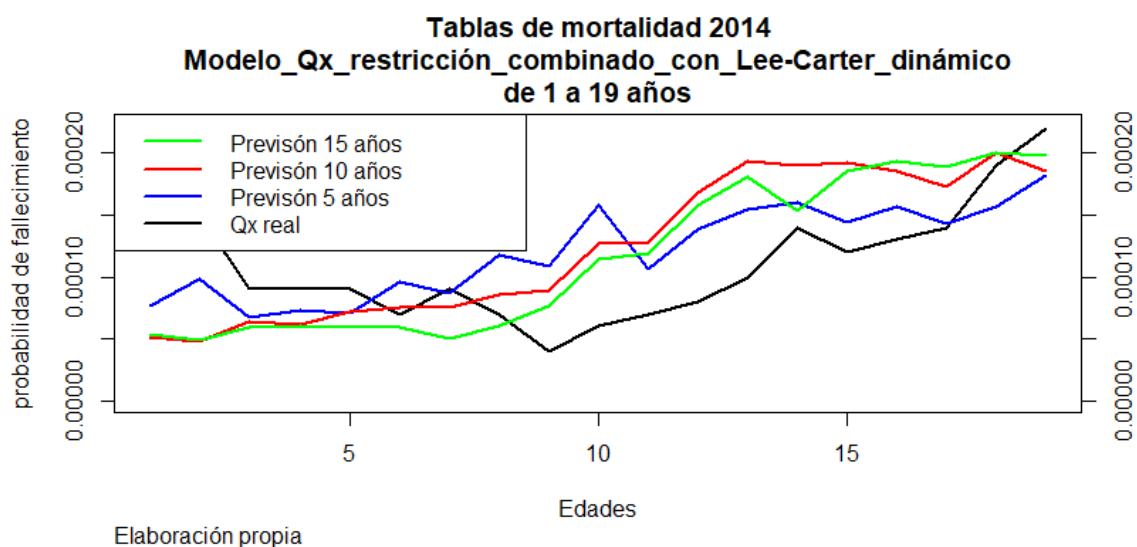
Para ello se combina los dos modelos, donde ARIMA con la restricción y con toda la muestra predice la forma, y Lee-Carter dinámico el nivel.

Para lograr la combinación se multiplica la previsión de cada grupo de edad de 10 años del modelo ARIMA por un constante, de tal manera que el valor de  ${}_{10}q_x$  de cada grupo del modelo ARIMA coincide con el que da el modelo Lee-Carter. Es decir, la probabilidad de una persona de 10 o 20 o 30 de sobrevivir en próximos 10 años según la predicción del modelo Lee-Carter dinámico y la predicción final del presente trabajo sean la misma.

Se halla este valor con el método secante, dado la sencillez del algoritmo, se omite los detalles.

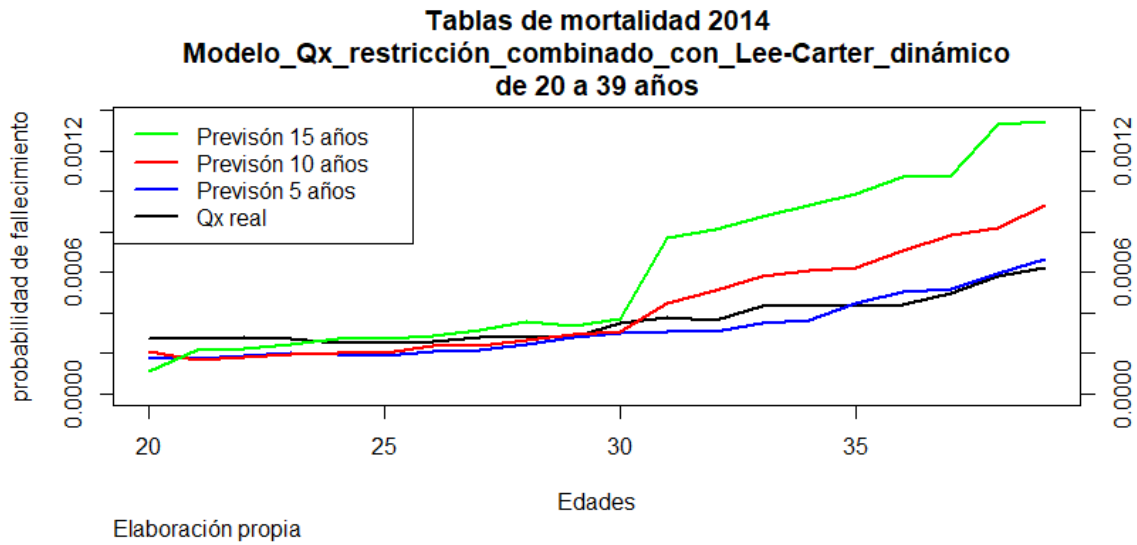
Se visualiza a continuación los resultados de la predicción:

Gráfico 65 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter\_dinámico de 1 a 19 años



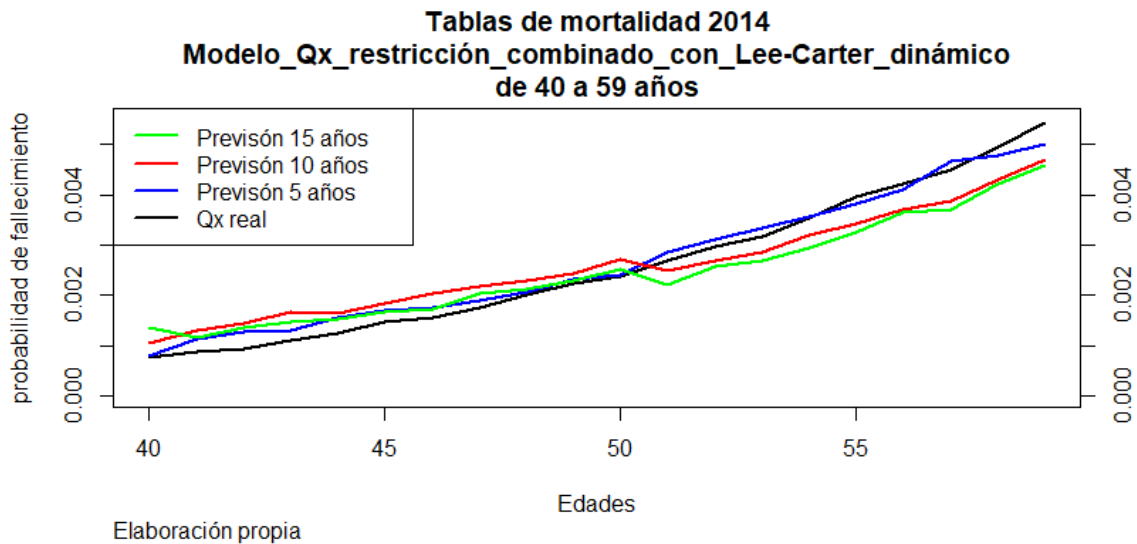
Se puede observar que los resultados del último modelo del presente trabajo no acaban de acertar en las predicciones de largo plazo para las mortalidades de edad menor de 20 años.

Gráfico 66 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter\_dinámico de 20 a 39 años



En este tramo de edad adulta, las previsiones mejoran bastantes con respecto a la previsión de tramo de edad más joven, aunque sigue sin ser de alta precisión especialmente la previsión más lejana en el tiempo, la de 15 años.

Gráfico 67 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter\_dinámico de 40 a 59 años



En tramo de edad comprendido entre 40 y 60, la mejora previsión con respecto a los modelos anteriores son muy importantes, adicionalmente, como se puede observar, las previsiones ya no presentan apenas el problema de alta volatilidad de otros modelos anteriores.

Gráfico 68 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter\_dinámico de 60 a 79 años

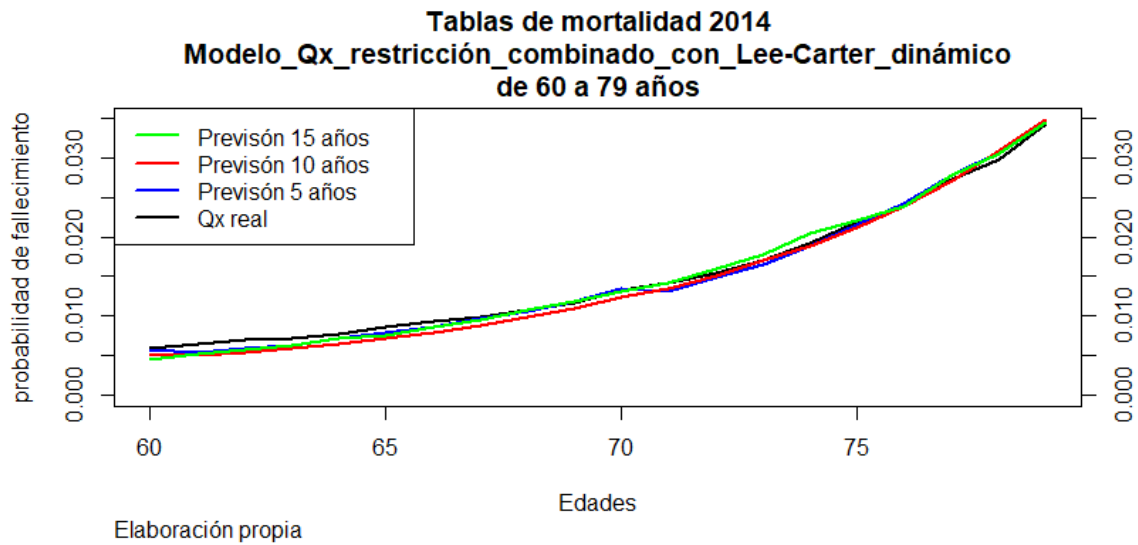
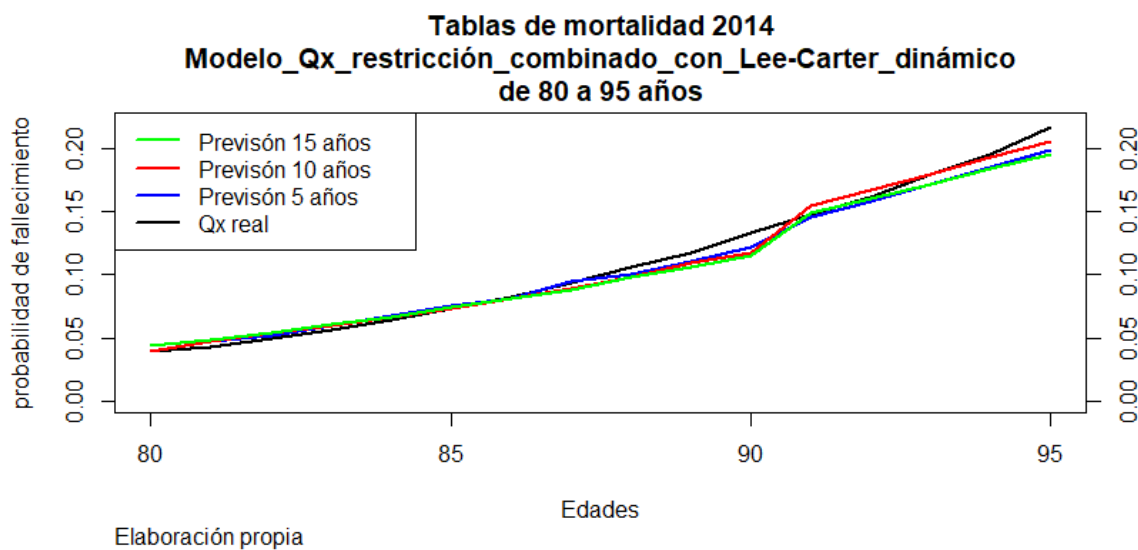


Gráfico 69 Tablas de mortalidad 2014 Modelo\_Qx\_restricción\_combinado\_con\_Lee-Carter\_dinámico de 80 a 95 años



En el tramo de 60 a 80 años, el tramo de mayor interés para la valoración de renta actual en el mercado de seguro español, los resultados son bastante realistas, y tanto la forma de las series como sus niveles son de gran precisión.

La validez de este resultado es apoyada por los análisis de mortalidades de este tramo de edad donde se observa una tendencia lineal y progresiva presentando una gran estabilidad tanto en las series  $q_x$  individuales como la agregada en forma de esperanza de vida por



tramo. En el análisis de Lee-Carter que se realiza anteriormente, se observa que, para este tramo de edad avanzada, las evoluciones de los 3 parámetros son especialmente robustas, encajan adecuadamente en numerosos estudios sobre la mortalidad. Por estos dos, se puede afirmar que este acierto no es fruto de la aleatoriedad y sino es debido al enfoque de modelación especialmente idóneo.

Se calcula los MAPES para las edades avanzadas.

Se ve que los resultados son muchísimos más preciso que si sólo se usara la última tabla disponible.

MAPE (56 -95 años)	model Qx lc final	ultimo qx
5 años	-2.54%	-28.85%
10 años	-5.33%	-65.88%
15 años	-3.98%	-94.85%

## 9.1 conclusión

los modelos de ARIMA con restricción y con todas muestras combinado con modelo Lee-Carter dinámico generan previsión a largo plazo con mucho más precisión que los modelos ARIMA, aunque el tramo de edad joven, alrededor de edad 37 tiene el ajuste de peor calidad, debido a que en los modelos Lee-Carter con un solo parámetro de mejora k no ha sido capaz de capturar todos los fenómenos de nuevos comportamientos de la mortalidad como VIH y accidentes de tráfico especialmente para el grupo de población adulta.

En general antes de las edades 50 se sobreestima el modelo la mortalidad, y a partir de esa edad el modelo infraestima, es una clara señal de que en la proyección, el principal parámetro para recoger la variación del nivel es el K de modelo Lee-Carter no captura toda la variación de la mortalidades de grupos de distintas edades, y claramente se trata de un punto por mejorar.

## 10 resultado de distintos modelos:

Se resumen en la siguiente tabla los resultados MAPE de los distintos enfoques del presenta.

La MAPE es el estadístico empleado para medir el ajuste del modelo, significa el error medio en porcentaje, en este caso se calcula de esta forma

$MAPE = (r - p) / r$  donde:

- r significa la tasa de mortalidad
- p significa la tasa de mortalidad prevista por el modelo

De esta forma, un MAPE de 0.94% indica que cuando en el periodo de previsión hay 10,000 fallecidos, la previsión prevé de media 10,094 fallecidos

Para el grupo de población adulto joven, perfil del cliente de seguro TAR:

MAPE (25-55años)	ultimo qx	model Qx lc final	model Qx 1943	model Qx 1960
1 año	-3.88%		1.36%	1.73%
5 años	-28.85%	<b>1.18%</b>	8.85%	9.30%
10 años	-65.88%	<b>20.17%</b>	2.17%	6.77%
15 años	-94.85%	<b>43.58%</b>	64.49%	53.42%

Los resultados de perfil de personas mayores de 56 a 95, grupo de población que es público objetivo de la renta actuarial.

MAPE (56 -95 años)	model Qx lc final	ultimo qx	model Qx 1943	model Qx 1960
1 año		-1.37%	-0.79%	-0.60%
5 años	<b>-2.54%</b>	-10.43%	2.90%	2.62%
10 años	<b>-5.33%</b>	-21.72%	5.47%	6.26%
15 años	<b>-3.98%</b>	-38.60%	-54.53%	-55.16%



Sintetizando los análisis de los distintos modelos que se ha realizado, se puede llegar a estas principales conclusiones:

1. El modelo ARIMA prevé con mucha precisión a corto plazo, pero falla a largo.
2. los problemas de volatilidad y sesgo en nivel son problemas que no se resuelve modificando la muestra estimación.
3. El modelo ARIMA con restricción resuelve el problema de volatilidad, pero no el sesgo en ajuste del nivel.
4. Modelo Lee-Carter dinámico predice mejor el nivel que el modelo ARIMA a largo plazo, pero no resuelve el problema de volatilidad
5. La combinación de los dos modelos da una previsión bastante realista incluso en largo plazo, especialmente para edad avanzadas, justamente el tramo que se interesa para el cálculo la renta actuarial.



## 11 Conclusión:

En un primer momento para detectar las mejoras de cohorte y debido a la gran muestra que se ha elegido, se han usado los mapas de calor. Con ello se ha buscado usar esta técnica para poder detectar mejoras de cohortes en toda la muestra en general y en segundo estudio más pormenorizado se analizaron en submuestras más pequeña para ver estas anomalías más detalladamente. Con esto se han detectado mejoras significativas en la primera mitad del siglo XX y mejoras no tan representativas en la segunda mitad.

Con estos datos se pueden afirmar en relación con la mortalidad, la existencia de saltos más significativos en esa primera etapa de siglo, viéndose una progresión en la mejora de la mortalidad más uniforme y menos escalonada en la segunda parte de siglo.

Una vez que con el mapa de calor se ha detectado esa diferencia entre la primera parte del siglo XX y la segunda parte del siglo XX, el trabajo se ha centrado en la etapa donde la mejora de la mortalidad ha sido más continuada. De esta forma se ha aislado la muestra de la Guerra Civil Española y de los avances médicos de estas primeras décadas. Para ello se ha utilizado el método de ARIMA, analizando en un primer momento la muestra desde 1943 y 1960 hasta final del siglo XX.

Se realiza el modelo ARIMA, con los datos desde 1943 hasta final del siglo XX y nuevamente desde 1960 hasta final del siglo XX. Para poder comparar ambos resultados y analizar el MAPE (el error medio en porcentaje) entre ambos análisis.

De esta forma se puede proceder a verificar en detalle ambos análisis, y confirmar la idoneidad del modelo ARIMA. Dicha técnica concretamente se ha utilizado en el trabajo para la proyección de la tabla de mortalidad de población española a un año. Al utilizar este método en el trabajo, el resultado que aporta es un error medio inferior al 1%. Este dato se obtiene del análisis en las edades comprendidas entre los 25 y 55 años en las diferentes muestras de 1943 y 1960.

Las edades comprendidas entre los 25 y 55 años, son muy interesantes para las empresas aseguradoras, concretamente en los seguros de vida a corto plazo. Para poder tener un cálculo preciso de la prima técnica de sus pólizas TAR a la hora de generar una previsión del modelo a corto plazo, se puede afirmar que sería una previsión realista y con aplicaciones prácticas.



Al analizar ambos modelos ARIMA de 1943 y 1960 en el trabajo, se observa proyecciones muy precisas para el corto plazo en el primer año. Pero al querer realizar previsiones en los siguientes años los resultados que aparecen con este método son muy volátiles y de esta forma las previsiones en el largo plazo carecen de veracidad.

Para intentar corregir esta peculiaridad del modelo ARIMA, en la modelación de una edad concreta se ha procedido a incluir no solo de esa edad mencionada, sino que se han tenido en cuentas las mortalidades de las edades colindantes a la edad analizada en un rango de 5 años por encima y 5 años por debajo de la fecha principal. Con ello se genera regresiones parecidas y tener unas previsiones sin saltos tan significativos como en el primer método de ARIMA.

Se procedió a la revisión de las previsiones obtenidas con el método de ARIMA, para proceder a la creación de un nuevo modelo con restricciones. Se pretende con este nuevo modelo compara ambos modelos para poder usar todos los datos potenciales.

Esto se ha llevado a cabo realizado una ponderación del peso de los datos en toda la muestra. Con ello se obtiene unos resultados menos volátiles y progresivos.

Al implementar el modelo de ARIMA con restricción aparece una previsión con sesgo. Para la resolución de dicho problema. En el trabajo se ha procedido a combinación de los resultados del ARIMA con restricciones con el modelo Lee-Carter Dinámico. Bajo la combinación de ambos modelos se obtiene un aumento notable de precisión en la previsión de la tasa de mortalidad. Con ello se obtiene un MAPE (error medio en porcentaje) de 2.5% para previsiones de 5 años vista y 4% para previsiones de 15 años vista.

Basando en los diagnósticos de los modelos realizados en este trabajo, se resalta que el método ARIMA es una técnica idónea para obtener la proyección realista de la tabla de mortalidad a un año. Es importante resaltar que dicho modelo a medio plazo (5 años) es necesario proceder a la combinación con el modelo Lee Carter para que tenga una mínima fiabilidad. Esto es debido a que el modelo ARIMA no tiene la capacidad para capturar estructuras más complejas y por ello no es recomendable su uso para la proyección a mayor plazo.



El modelo Lee-Carter, especialmente en su versión dinámica combinado con el ARIMA genera previsión a 5 años bastante precisa para prácticamente todas las edades de la población demostrando su utilidad en la práctica.

Para previsiones de más de 5 años, es recomendable proceder con mucha precaución, a pesar de los datos que en este trabajo se han logrado proyecciones realistas en las tablas de mortalidad a 15 años y en tramos de edad avanzada. No se ha encontrado ningún otro estudio que sugiera que esta estabilidad de comportamiento biométrico es sostenible en el tiempo ni es aplicable en otra población.



## 12 Próximas líneas de investigación:

Hacer tramos en la estimación de Lee-Carter, es decir crear modelos por tramos de edad, corregiría el problema de los niveles de algunas edades entre 25 y 55 en previsión de largo plazo. Y corregiría el problema de sobreestimación en las edades joven y infraestimación en las edades de vejez.

Corregir el efecto de bajada de esperanza de vida en los años 1960 de edad 20-39, eliminar este efecto de la proyección de parámetros  $a_x$  y  $b_x$ .

Para recoger mejor el efecto de rectangularización, analizar y introducir unos parámetros tal como el ratio de mejora (ratios de parámetros B de dos edades distintas en modelo Lee-Carter) para medir mejor el reparto de mejora en el tiempo introducido de parámetro K (de modelo Lee-Carter)

Mejorar la combinación: la combinación multiplicando por un solo número es muy simplicista, introduce cierto salto en la frontera.

Mejorar la previsión de parámetros, lo ideal sería estimar todos los modelos incluido el proceso de combinación en un algoritmo que tiene en cuenta y optimiza todos los parámetros, se sabe por regla general que el ajuste es mejor.

## 13 Bibliografía

- ❖ García González, J.M (2015). La transformación de la longevidad en España de 1910 a 2009. *Centro de investigaciones Sociológicas(CIS)*.
- ❖ Herce, J.A. El impacto del envejecimiento de la población en España. *Cuadernos de Información Económica (FUNCAS)*.
- ❖ Blanes Llorans, A & Cabré PLA, A. Inercias e incertidumbres en el futuro demográfico de España. *Revista del ministerio de empleo y seguridad social*.
- ❖ Betzuen Zalbidegoitia, A & Betzuen Álvarez, A.J(2015). La posibilidad de un índice de longevidad para la unión europea bajo las directrices de solvencia II. *Anales del Instituto de Actuarios Españoles, 3a época, 21, 2015/73-106*
- ❖ Herrera, L.E(2015). Cobertura de la dependencia en España prestaciones económicas y prestaciones de servicios. *Tesis doctoral en Ciencias de la Enfermería de la Universidad Europea*.
- ❖ Gómez Redondo, R(1997). La mortalidad de la España actual. *Departamento de Sociología II, UNED*.
- ❖ Ramiro Fariñas, D & Sanz Gimeno, A(1999). Cambios estructurales en la mortalidad infantil y juvenil española 1860 – 1990. *Boletín de la Asociación de Demografía Histórica, XVII,I,1999*.
- ❖ Zoido Naranjo, F & Arroyo Pérez,A. La población de España. *Departamento de Geografía Física y Análisis Geográfico Regional de la Facultad de Geografía e Historia & Departamento de Estadística e Investigación Operativa de la Facultad de Matemáticas de la Universidad de Sevilla*.
- ❖ F & Arroyo Pérez,A. Fecundidad. *Departamento de Estadística e Investigación Operativa de la Facultad de Matemáticas de la Universidad de Sevilla*.
- ❖ Hernández Rodríguez, J.A. Evolución futura de la población. *Departamento de Estadística e Investigación Operativa de la Facultad de Matemáticas de la Universidad de Sevilla*.
- ❖ Viciano Fernandez, F. Mortalidad. *Departamento de Ciencias Socio-Sanitarias, Facultad de Medicina de la Universidad de Sevilla*.
- ❖ Andreozzi, L. Estimación y pronósticos de la mortalidad de Argentina utilizando el modelo de Lee-Carter. *Tesis de mortalidad/modelo de Lee-Carter/modelo ARIMA/Argentina/modelos de espacios de estados/esperanza de vida de la Universidad Nacional de Rosario*.



- ❖ Ochoa Molina, C.A(2015). El modelo Lee-Carter para estimar y pronosticar mortalidad: Una aplicación para Colombia. *Facultad de Ciencias, escuela de ciencias, escuela de Estadística de la Universidad Nacional de Colombia.*