

APLICACIÓN DEL MODELO LEE-CARTER A URUGUAY

APPLICATION OF THE LEE-CARTER MODEL TO URUGUAY

Guillermo Magnou¹, Maria Cecilia Salles²

Palabras clave:

modelado de mortalidad, previsión de mortalidad, esperanza de vida, seguros, riesgo de longevidad

Keywords:

mortality modeling, mortality forecasting, life expectancy, insurance, longevity risk

Resumen

El modelo Lee-Carter es una de las metodologías más populares para pronosticar tasas de mortalidad. Es un modelo simple que se ha utilizado con éxito en los EE. UU. y varios países, cuyos parámetros luego se tratan como series de tiempo para producir pronósticos de mortalidad. Este trabajo describe la aplicación del modelo Lee-Carter a las tasas de mortalidad específicas por edad y género en Uruguay. Estas tasas están disponibles para el período que va de 1974 a 2020. Concluimos que el parámetro de tendencia temporal \mathcal{K}_t puede modelarse como un paseo aleatorio con deriva y utilizamos las tasas de mortalidad previstas para el período de tiempo que va de 2021 a 2050 para proyectar la esperanza de vida al nacer.

Códigos JEL: C53, J11

Abstract

The Lee-Carter model is one of the most popular methodologies for forecasting mortality rates. It is a simple model that has been used successfully in the US and several countries, whose parameters are then treated as time series to produce mortality forecasts. This paper describes the application of the Lee-Carter model to age-specific death rates by gender in Uruguay. These rates are available for the period that goes from 1974 to 2020. We concluded that the time trend parameter \mathcal{K}_t can be model like a random walk with drift and we use the mortality rates forecast for the time period that goes from 2021 to 2050 in order to project life expectancy at birth.

¹ Universidad ORT Uruguay, Facultad de Administración y Ciencias Sociales., (Uruguay).
E-mail: guillermomagnou@gmail.com ORCID: <https://orcid.org/0009-0004-2507-7781>

² Departamento de Actuaría, Banco de Seguros del Estado, (Uruguay).
E-mail: msalles@bse.com.uy

INTRODUCCIÓN

According to World Bank data, in the last 60 years, life expectancy at birth for the world population has increased from 51 years in 1960 to 72 years in 2020. This is equivalent to an increase in life expectancy of 4 months for each year. This increase in life expectancy, though a sign of social progress, poses a challenge to governments, private pension plans and insurers because of its impact on pension and health costs. Actuaries have recognized the problems caused by an aging population and rising longevity and have thus devoted significant attention to the development of statistical techniques for the modeling and projection of mortality rates.

There is a large literature on human mortality modelling. Very early on, Gompertz (1825) suggested that mortality increases exponentially with age during the adult years of life. Makeham (1860) extended the Gompertz (1825) model by adding an age-dependent component to better capture younger age mortality. These models are static and one-dimensional, and cannot be used for mortality forecasting. Subsequently, several new approaches were developed using stochastic models, one of the most influential approaches is the mortality model proposed by Lee and Carter (1992).

The Lee-Carter model has inspired numerous variants and extensions. For instance, Lee and Miller (2001), Booth et al. (2002), and Brouhns et al. (2002) have proposed alternative estimation approaches in order to improve the goodness-of-fit and the forecasting properties of the model. In particular, Brouhns et al. (2002) propose a more formal statistical approach to estimating the parameters by embedding the Lee-Carter model into a Poisson regression setting. Other authors have extended the Lee-Carter model by including additional terms, such as multiple bilinear age-period components (Renshaw and Haberman 2003; Hyndman and Ullah 2007), or a cohort effect term (Renshaw and Haberman 2006).

Our interest in the Lee-Carter model arises from its simplicity, it only has three parameters, which are easy to interpret. It must be kept in mind that most databases, like World Bank data,

group data at advanced ages, for example, over 80. This is due to the scarce and volatile nature of mortality data at these ages. For this reason, we will use the technique proposed by in Coale and Guo (1989) to extend mortality to older ages.

The paper is structured as follows: first we present how we constructed the historical data and then a description of the model that we will use. We will then present the study results for the data series and end with the conclusion.

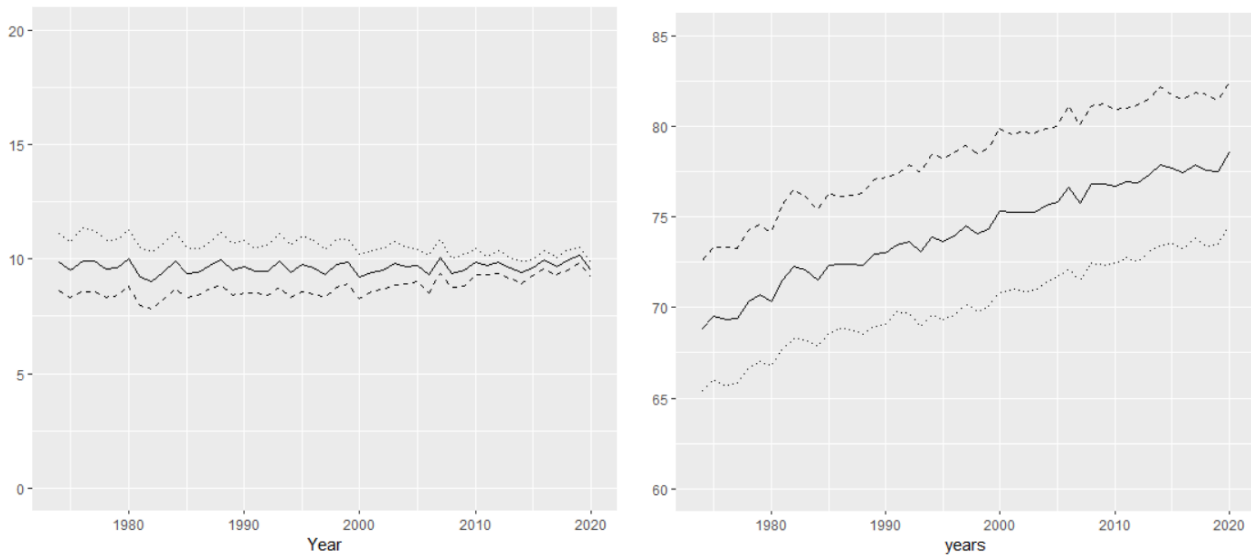
METODOLOGY

This paper combines Uruguayan mortality data from the Ministry of Public Health of Uruguay (MSP) and a dataset provided by the World Bank DataBank (WBD). For the remainder of this paper, we will refer to data from these sources as MSP data and WBD data, respectively.

The MSP receives and processes data on live births, deaths, and fetal deaths for statistical purposes. These data are widely used to calculate various indicators of mortality, birth and fertility. They therefore constitute a basic input for the development, monitoring and evaluation of different policies, plans and programs both in the health sector and in other economic and social sectors of the country. From this database we download the deaths per year for different sex and age group from 1974 to 2020.

The WBD is an online web resource that provides simple and quick access to collections of time series data. It has advanced functions for selecting and displaying data, performing customized queries, downloading data, and creating charts and maps. Users can create dynamic custom reports based on their selection of countries, indicators and years. For this paper we use the World Developed Indicators (WDI) database, which contains a series of demographic indicators for more than 250 countries, including Uruguay. From this database we download the midyear estimate population per year for different sex and age group from 1974 to 2020.

FIGURE 1
Mortality rate and Life expectancy at birth (1974-2020).



The left panel of Figure 1 shows the mortality rate, with the solid line for the total population, the dotted line for male, and the dashed line for female. It is observed that the ratio remained stable during the study period, at around 10% for the total population, and the difference between mortality according to gender decreases over time. In the right panel of Figure 1, life expectancy at birth is observed; it can be seen that it increased for males, females and therefore for the total population. In the case of female, it went from 72.6 in 1974 to 82 in 2020.

Unfortunately, the dataset extends only to the open age group 80 and over, whereas our interest extends to higher ages. Coale and Kisker (1990) showed that in populations with good data at old ages, mortality rate increases not a constant rate with age, as the Gompertz curve assumes, but rather at a linearly decreasing rate. We apply the method suggested in Coale and Guo (1989) to extend our death rate up to age group 105-109.

1. The Lee-Carter model

Let the random variable D_{xt} denote the number of deaths in a population at age x last birthday during calendar year t . Also let d_{xt} denote the observed

number of deaths, E_{xt}^c the central exposed to risk at age x in year t , and E_{xt}^0 the corresponding initial exposed to risk. The force of mortality and central death rates are denoted by μ_{xt} and m_{xt} , respectively, with the empirical estimate of the latter being $\hat{m}_{xt} = d_{xt}/E_{xt}^c$. Under the assumption that the force of mortality is constant over each year of age and calendar year, i.e., from age x to age $x + 1$ and year t to $t + 1$, then the force of mortality μ_{xt} and the death rate m_{xt} coincide. We assume that this is the case throughout.

All the way through this paper we assume that deaths, d_{xt} , and either central exposures, E_{xt}^c , or initial exposures, E_{xt}^0 are available in a rectangular array format comprising ages (on the rows) $x = x_1, x_2, \dots, x_k$, and calendar years (on the columns) $t = t_1, t_2, \dots, t_n$. When only central exposures are available and initial exposures are required (or vice-versa), one can approximate the initial exposures by adding half the matching reported numbers of deaths to the central exposures, i.e., $E_{xt}^0 \approx E_{xt}^c + d_{xt}/2$.

The classical Lee-Carter model (1992) is in essence a relational model

$$\ln \hat{\mu}_{xt} = a_x + b_x \mathcal{K}_t + \varepsilon_{xt}$$

where $\hat{\mu}_{xt}$ denotes the observed force of mortality at age x during year t , ε_{xt} are homoscedastic centered error terms, a_x and b_x are age-specific constants and \mathcal{K}_t is a time-varying index. To ensure identifiability of the model, Lee-Carter suggest the following set of parameter constraints:

$$\sum_x b_x = 1, \quad \sum_t \mathcal{K}_t = 0$$

It is worth mentioning that this model is not a simple regression model, since there are no observed covariates in the right-hand side. So, the authors propose to use singular value decomposition (SVD) to estimate the parameters under an ordinary least-squares (OLS).

According to Alho (2000), the main drawback of the OLS estimation via SVD is that the errors are assumed to be homoscedastic. This is related to the fact that for inference we are actually assuming that the errors are normally distributed, which is quite unrealistic. The logarithm of the observed force of mortality is much more variable at older ages than at younger ages because of the much smaller absolute number of deaths at older ages. Because the number of deaths is a counting random variable, according to Brillinger (1986), the Poisson assumption appears to be plausible.

To avoid the problems associated with the OLS method and taking into account the above, Brouhns et al. (2002) implemented the Lee-Carter model assuming a Poisson distribution of the number of deaths, $D_{xt} \sim \text{Poisson}(E_{xt}^c \mu_{xt})$ with $E(D_{xt}/E_{xt}^c) = \mu_{xt}$. Then the Lee-Carter model become

$$\ln E\left(\frac{D_{xt}}{E_{xt}^c}\right) = a_x + b_x \mathcal{K}_t$$

The vector a_x can be interpreted as an average age profile of mortality. The b_x profile tells us which rates decline rapidly and which rate decline slowly in response to change in \mathcal{K}_t . In principle b_x could be negative for some ages, indicating that mortality at those ages trends to rise when falling at other ages; in practice this does not seem to occur over the long run. When \mathcal{K}_t is linear in time, mortality at each age changes at its own

constant exponential rate. As \mathcal{K}_t goes to negative infinity, each age-specific rate goes to 0; negative death rates cannot occur in this model, which is an advantage for forecasting.

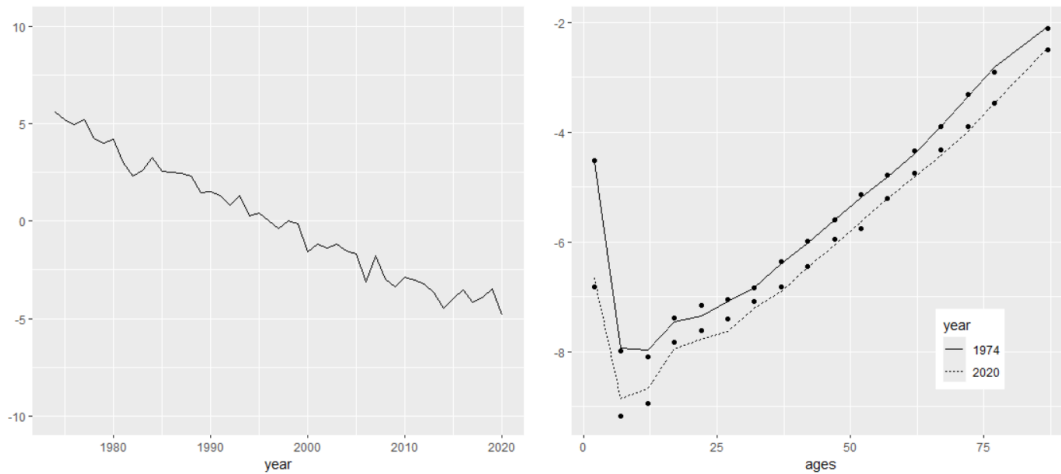
In order to project mortality, the time-varying index \mathcal{K}_t is viewed as a stochastic process. Box-Jenkins techniques (see Box. et al. 2015) are then used to modeled and forecasted \mathcal{K}_t within an ARIMA time series model.

1. Empirical Results

In this section, we will show the main results of the implementation of the Lee-Carter model on historical mortality data from Uruguay. The data set was made up of 6 matrices, each one has 47 columns, representing a year, from 1974 to 2020, while the 17 rows represent the age groups, from 0-4 years to 80 years or more. Three matrices represent the total deaths, male and female. The rest represent the population at mid-year.

Parameter estimates of the Lee-Carter model can be obtained by maximizing the log-likelihood. In the mortality literature, maximization of the log-likelihood is typically performed using the Newton-Raphson iterative procedure tailored for each model (see, e.g., Brouhns et al. (2002) and Cairns et al (2009)). Nonetheless, as discussed by Currie (2016), many stochastic mortality models, like Lee-Carter, are examples of generalized linear models or generalized non-linear model, which facilitates their fitting using standard statistical software. We will use the R software through the StMoMo package. The results for the female population are presented below.

Figure 2
Lee-Carter \mathcal{K} for 1974 to 2020 and log mortality rate fit for 1974 and 2020.



The left panel of Figure 2 shows the Lee-Carter estimates of \mathcal{K}_t . As shown, \mathcal{K}_t decreases approximately linearly between 1974 and 2020, which is consistent with the pattern of change in life expectancy shown in the right panel of Figure 1. The right panel of Figure 2 shows how well the model fits Uruguayan mortality in 1974 and 2020, the two extremes of the range, showing the familiar shape of mortality by age and larger relative declines at younger ages. The estimates of a_x and b_x are given in Table 1. These can be used with forecasts of \mathcal{K}_t to construct age specific death rates.

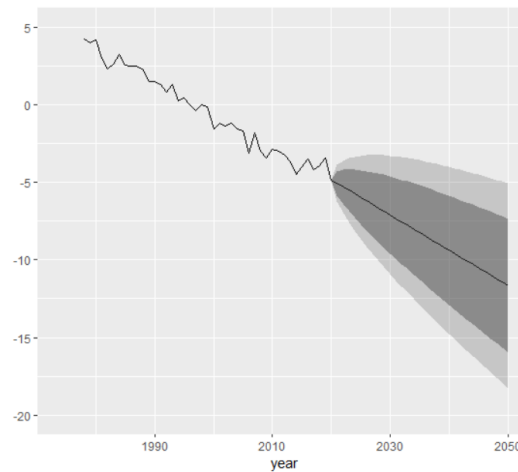
TABLE 1
Lee-Carter fitted values for a and b .

Age group	a_x	b_x
0-04	-5,67335	0,20253
05-09	-8,42645	0,08783
10-14	-8,35273	0,06818
15-19	-7,71631	0,04761
20-24	-7,57490	0,04083
25-29	-7,38086	0,05148
30-34	-7,03324	0,03452
35-39	-6,66295	0,04931
40-44	-6,26092	0,04203
45-49	-5,84729	0,04514
50-54	-5,42471	0,04305
55-59	-5,03306	0,03752
60-64	-4,61638	0,04035
65-69	-4,18129	0,05022
70-74	-3,69166	0,05979
75-79	-3,17022	0,06217
80+	-2,28404	0,03744

After adjusting the Lee-Carter model, we are now ready to move to the problem of forecasting. The first step is to find an appropriate ARIMA time series model for the mortality index \mathcal{K}_t . We found that a random walk with drift describes \mathcal{K}_t well. So, our forecast model with standard error in parentheses, is as follows:

$$\mathcal{K}_t = -0,2274_{(0,0895)} + \mathcal{K}_{t-1} + \varepsilon_t$$

where $\sigma_{\mathcal{K}}^2 = 0,3768$ and $R^2 = 0.987$. The constant term, $-0,2274$ indicates the average annual change in \mathcal{K} , which drives the forecasts of long-run change in mortality. Over 30-year horizon, we will forecast a decline in \mathcal{K} of 30 times 0,2274, or 6,822. The standard error ($\sigma_{\mathcal{K}}$) indicates the uncertainty associated with a one-year forecast, as the forecast horizon increases, the standard error grows with the horizon's square root. Figure 3 plot the past values of \mathcal{K} along with the forecasts based on the time series and the associated 80% and 95% confidence intervals.

FIGURE 3Lee-Carter \mathcal{K} forecast to 2050 with 80% and 95% confidence band.**TABLE 3**

Forecasts the Female Age-Specific death rates per 100.000 at five-year intervals (2025-2050).

Age group	2025	2030	2035	2040	2045	2050
0-04	103	81	65	51	41	32
05-09	13	12	11	10	9	8
10-14	16	15	13	12	12	11
15-19	34	32	30	29	27	26
20-24	40	38	37	35	33	32
25-29	46	43	41	38	36	34
30-34	72	69	66	64	61	59
35-39	95	90	85	80	76	72
40-44	149	142	135	129	123	117
45-49	221	210	199	189	180	171
50-54	341	325	309	294	280	267
55-59	521	499	478	458	439	421
60-64	777	742	709	677	647	618
65-69	1.132	1.069	1.010	954	901	851
70-74	1.745	1.630	1.523	1.423	1.329	1.242
75-79	2.897	2.700	2.515	2.344	2.184	2.035
80-84	4.812	4.471	4.155	3.861	3.588	3.334
85-89	7.991	7.406	6.863	6.360	5.894	5.463
90-94	13.271	12.266	11.336	10.478	9.684	8.950
95-99	22.039	20.315	18.725	17.260	15.909	14.664
100-104	36.602	33.646	30.930	28.432	26.137	24.026
105-109	60.785	55.726	51.089	46.837	42.939	39.366

Table 3 contains forecasts of all the five years death rates for 2025 and every five years thereafter through 2050. Infant mortality rates are forecast to fall to less than one per thousand. From the forecasts of death rates, it is straight forward to calculate life tables and life expectancy at birth. Table 4 contains forecasts of l_x (proportions surviving from birth to exact age x) for five-year age group as in Table 3. In 2020, around 34.6% of

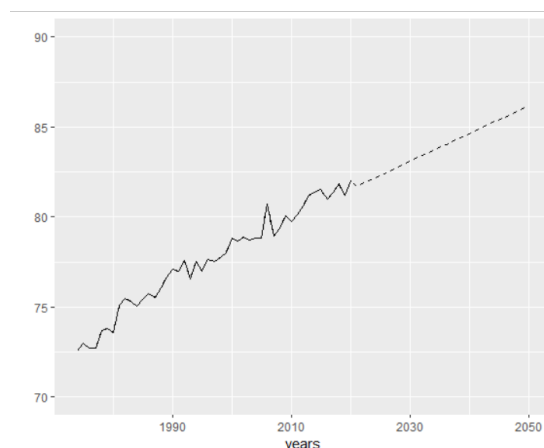
births survive to age 90. We estimate that by 2030, the value will increase to 37.4%, then by 2040 it will rise to 42,6% and by 2050 47,6% will survive to age 90. Figure 5 shows life expectancy at birth for women from 1974 to 2020 and its forecast until 2050. An increase in life expectancy at birth of a little more than four years is forecast for the next 30 years, that is, life expectancy at birth will go from 82 years in 2020 to 86.2 years in 2050.

TABLE 4

Forecasts of number surviving to exact ages out of 100.000 births at five-year intervals 2025-2050, for female.

Age group	2025	2030	2035	2040	2045	2050
0	100.000	100.000	100.000	100.000	100.000	100.000
5	99.489	99.594	99.677	99.744	99.796	99.838
10	99.425	99.536	99.624	99.696	99.753	99.799
15	99.347	99.463	99.557	99.634	99.695	99.746
20	99.181	99.306	99.408	99.492	99.561	99.618
25	98.981	99.115	99.226	99.318	99.395	99.459
30	98.755	98.901	99.024	99.127	99.215	99.290
35	98.401	98.560	98.696	98.811	98.911	98.997
40	97.934	98.118	98.277	98.415	98.535	98.642
45	97.209	97.426	97.616	97.783	97.933	98.066
50	96.143	96.410	96.649	96.863	97.057	97.233
55	94.519	94.859	95.167	95.449	95.707	95.945
60	92.088	92.520	92.917	93.285	93.627	93.946
65	88.578	89.148	89.681	90.179	90.647	91.088
70	83.702	84.506	85.264	85.978	86.654	87.293
75	76.705	77.889	79.009	80.072	81.079	82.036
80	66.344	68.039	69.660	71.207	72.684	74.093
85	52.096	54.357	56.550	58.671	60.718	62.692
90	34.747	37.374	39.986	42.572	45.122	47.626
95	17.434	19.832	22.326	24.899	27.532	30.210
100	5.047	6.473	8.088	9.888	11.864	14.002
105	224	558	1.034	1.672	2.487	3.493
110	0	0	0	0	0	0

Figure 4
Actual life expectancy at birth (1974-2020) and forecasts (2021-2050).



DISCUSSION

The world population is experiencing an aging process, which is reflected in the increase in life expectancy at birth. This increase in life expectancy, while a sign of social progress, poses several challenges for actuaries. For annuities and defined benefit pension plans, increasing longevity can lead to low-frequency, high-severity losses. These challenges require close and continuous monitoring to ensure the stability of the entities they insure, which is why their adequate measurement is critical in determining the accounting, financial and solvency situation of the insurance entities.

There is a large literature on human mortality modelling. The Lee-Carter Model is one of the most popular methodologies for forecasting mortality rates. The original Lee-Carter method was used to US data. In addition, the model has been used successfully in other countries, including Canada (Lee & Nault, 1993), Japan (Wilmoth, 1996), Chile (Lee & Rofman, 1994), Belgium (Brouhns & Denuit, 2001) and Argentina (Andreozzi et. al., 2011) among others.

As already mentioned, the Lee-Carter model is not a simple regression model, since no covariates are observed on the right side. To ensure the identifiability of the model, Lee-Carter propose a set of restrictions for the parameters. To achieve a better fit of the model parameters, we followed the suggestions made by Brouhns et al. (2002) and we assume a Poisson distribution for

the number of deaths. This is a better choice than the normal distribution for two reasons: the number of deaths is a counting random variable and the logarithm of the observed force of mortality is much more variable at older ages than at younger ages.

An important aspect of Lee-Carter methodology is that the time factor \mathcal{K}_t is intrinsically viewed as a stochastic process. Box-Jenkins techniques are then used to estimate and forecast \mathcal{K}_t within an ARIMA times series model. These forecasts in turn yield projected age-specific mortality rates and life expectancies. It should be noted that the Lee-Carter method does not attempt to incorporate assumptions about advances in medical science or specific environmental changes; no information other than previous history is taken into account. This means that this approach is unable to forecast sudden improvements in mortality due to the discovery of new medical treatments. Similarly, future deteriorations caused by epidemics, like COVID-19, cannot enter the model.

CONCLUSION

This paper describes the application of the Lee-Carter model to age-specific death rates by gender in Uruguay. These rates are available for the period that goes from 1974 to 2020. In this period, we observed that the mortality rate remained relatively stable at around 10%. At the same time, we note that the composition by age groups varies considerably in some ages, for

example, a decrease in infants. This variation in the mortality rate causes an increase in life expectancy at birth of approximately 10 years, in the period 1974-2020.

We estimated the Lee-Carter parameters by maximum log-likelihood. The vector a_x can be interpreted as an average age profile of mortality, the vector tracks mortality changes over time, and the vector b_x determines how much each age group changes when \mathcal{K}_t changes. We found that the time trend parameter \mathcal{K}_t , is essentially lineal in all the period. We use standard statistical methods to model and forecast the index of mortality as a random walk with drift, which implies that each age group's mortality continues to decline at its own age-specific exponential rate. From the forecasts of rates, we construct forecasts of life expectancy. We anticipate that it will rise by about 4.2 years to 86.2 in the year 2050.

Finally, a possible extension of this research is to consider more complex models, for example, those referred to in the literature as stochastic mortality models. This includes the extensions of the Lee-Carter proposed in Renshaw and Haberman (2003, 2006), the original CBD model, and the extended CBD models of Cairns et al. (2009). In addition, all the model structures considered in Haberman and Renshaw (2011), Lovász (2011) and van Berkum et al. (2014), as well as the models of Plat (2009), Aro and Pennanen (2011), O'Hare and Li (2012), Borger et al. (2013) and Alai and Sherris (2014).

REFERENCES

- Alho, J. (2000). "The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications". *North American Actuarial Journal*, 4(1), 91-93.
- Andreozzi, L., Blaconá, M., Arnesi, N. (2011). "The Lee Carter Method for Estimating and Forecasting Mortality: An Application for Argentina". *The 31st International Symposium on Forecasting. International Institute of Forecasters, Praga*.
- Booth H, Maindonald J, Smith L (2002). "Applying Lee-Carter under Conditions of Variable Mortality Decline." *Population Studies*, 56(3), 325-336.
- Box, G, Jenkins, G, Reinsel, G., & Ljung, G. (2015). "Time series analysis: forecasting and control". *John Wiley & Sons*.
- Brillinger, D. (1986). "The natural variability of vital rates and associated statistics." *Biometrics* 42, 693-734.
- Brouhns, N., Denuit, M. (2002). "Risque de longévité et rentes viagères. II. Tables de mortalité prospectives pour la population belge". *Belgian Actuarial Bulletin*, 2(1), 49-63.
- Brouhns N, Denuit M, Vermunt J (2002). "A Poisson Log-Bilinear Regression Approach to the Construction of Projected Lifetables." *Insurance: Mathematics and Economics*, 31(3), 373-393.
- Cairns A, Blake D, Dowd K (2006). "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration." *Journal of Risk and Insurance*, 73(4), 687-718.
- Cairns A, Blake D, Dowd K, Coughlan G, Epstein D, Ong A, Balevich I (2009). "A Quantitative Comparison of Stochastic Mortality Models Using Data from England and Wales and the United States." *North American Actuarial Journal*, 13(1), 1-35.

Coale, A., Guo, G. (1989). "Revised Regional Model Life Tables at Very Low Levels of Mortality," *Population Index*, 55, 613-643.

Coale, A., Kisker, E. (1990). "Defects in data on old age mortality in the United States: New procedures for calculating approximately accurate mortality schedules and life tables at the highest ages". *Asian and Pacific Population Forum*, vol. 4. pp. 1-31.

Currie I (2016). "On Fitting Generalized Linear and non-linear Models of Mortality." *Scandinavian Actuarial Journal*, (4), 356-383.

Hyndman RJ, Ullah S (2007). "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach." *Computational Statistics & Data Analysis*, 51(10), 4942-4956.

Gompertz, B., (1825). "On the nature of the function expressive of the law of human mortality and on the mode of determining the value of life contingencies". *Phil. Trans. R. Soc. A* 115, 513-585.

Lee R, Carter L (1992). "Modeling and Forecasting U.S. Mortality." *Journal of the American Statistical Association*, 87(419), 659-671.

Lee R, Miller T (2001). "Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality." *Demography*, 38(4), 537-549.

Lee, R., Nault, F., (1993). "Modeling and forecasting provincial mortality in Canada". *Paper presented at the World Congress of International Union for the Scientific Study of Population, Montreal*.

Lee, R., Rofman, R., (1994). "Modeling and forecasting mortality in Chile". *Natas* 22 (59), 182-313.

Makeham, W. (1860). "On the law of mortality and construction of annuity tables". *J. Inst. Actuar.* 8 (6), 301-310.

Renshaw A, Haberman S (2003). "Lee-Carter Mortality Forecasting with Age-specific Enhancement." *Insurance: Mathematics and Economics*, 33(2), 255-272.

Renshaw A, Haberman S (2006). "A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors." *Insurance: Mathematics and Economics*, 38(3), 556-570.

Villegas, A., Kaishev, V., Pietro, M. (2018). "An R Package for Stochastic Mortality Modeling". *Journal of Statistical Software*, 84 (3), pp 1-38.

Wilmoth, J. (1996). "Mortality projections for Japan: A comparison of four methods". *Health and mortality among elderly populations*, 266-287.