PROBABILISTIC FORECASTS OF LIFE EXPECTANCY FOR ROSARIO CITY, ARGENTINA[1]

Lucía Andreozzi*
Nora Ventroni**

Abstract. Life expectancy at birth is one of the most useful indices for measuring the overall level of mortality and estimates the level of mortality more accurately than the crude mortality rate because it is independent of the age structure of the population. It also allows comparing the levels of mortality for different populations in different historical moments. Probabilistic forecasting models generate age-specific mortality rates for future period, and from these results, it is possible to derive forecasts of life expectancy at birth and their corresponding confidence intervals. In this paper two models are applied; the precursor of probabilistic models for mortality: the Lee and Carter model (1992) and the last one proposed in the area: the functional data model developed by Hyndman and Ullah (2008). Both models predict mortality, and enable, through life tables based on mortality forecasts, to derive life expectancy at birth. It is applied to mortality data of Rosario city, in the period 1980 to 2015, so as to obtain point and interval forecasts for life expectancy at birth.

Keywords: Functional data; Lee-Carter model; Forecast interval.

1 Translated from the Spanish by María Sara Loose. Body of Translators, National University of Rosario (UNR), Argentina. Contact: departamento_traduccion@unr.edu.ar

* National University of Rosario (UNR), Argentina, and National Scientific and Technical Research Council (CONICET), Argentina.

Contact: andreozzi.lu@gmail.com

ORCID: https://orcid.org/0000-0002-1723-5725

** National University of Rosario (UNR), Argentina, and Rosario Municipality, Argentina.

Contact: nventroni@gmail.com

ORCID: https://orcid.org/0000-0001-9510-9436

1. Introduction

The General Directorate of Statistics, depending on Rosario Municipality, and the School of Statistics, Faculty of Economics and Statistics, National University of Rosario (UNR) have undertaken a project[2] that aims, among other objectives, at obtaining probabilistic forecasts for Life Expectancy at Birth (LEAB) in Rosario city. Analyzing the mortality and life expectancy of a particular city, area or country is of paramount importance to obtain some signs about a population's quality life and sanitary state; besides, studying the evolution and projection of these indicators is crucial to outline public policies.

The data used to build measures related to mortality are generally taken from two sources: statistics on mortality (Registry of Vital Records and Statistics), and statistics on populations (usually censuses).

The specific mortality rates and the crude mortality rate measure the frequency of deaths per one thousand inhabitants. The difference is that the specific one does not take the total population but a given age group or smaller age groups.

Specific rates adequately measure the level of mortality in each age group or age; however, on their own, these do not provide the overall level of mortality for the target population. Life expectancy at birth (LEAB) is one of the most accepted indices to represent overall mortality, which is usually obtained from a life table. LEAB also enables to compare mortality levels of different populations at different historic moments (Arriaga, 2014). Nevertheless, as Arriaga (2014) explains, life expectancy evaluates the level of mortality with life-years and this has an impact when measuring the change in mortality.

Projecting mortality is useful for planning at different State departments. However, any projection or forecast is somewhat uncertain. Despite the fact that demography easily describes a present situation through certain indicators, trends in births, deaths and migrations are affected by unpredictable shifts. This is one of the reasons combining demographical and statistical data produces more enlightening results.

Thus, it is worth mentioning that probabilistic forecasting methods of population are quickly gaining recognition. Their main advantage is the so-called probabilistic coherence due to forecast variables and the resulting indices (Lee & Tuljapurkar, 1994). This helped developing probabilistic forecasting methods that are being increasingly accepted (Alho, 2000) and implemented by official statistical institutes to produce their national forecasts (e.g. in Holland and the United States). Another proposal, widely spread and implemented by the United Nations, is that of Rafterty, Li, Ševcíková, Gerland and Heilig's (2012). It consists of calculating the probabilistic projection of life expectancy through Bayesian hierarchical models.

As regards probabilistic forecasting, most advances have been made for mortality. For fertility and migration, on the other hand, there are fewer methodologies proposed. A forecast that includes probability distributions enables to add the notion of uncertainty in a more explicit way through probability, therefore providing results that are more informative. Future demographic results can be considered unknown values with a probability distribution. The most concise and precise contribution to probabilistic

2 It is carried out in the framework of a Collaboration Agreement, without any financing.

forecasting methods in demography was Lee & Carter's proposal (1992). Although at present there are a large number of variants and extensions, the authors initially proposed a methodology that allowed modelling and extrapolating the tendencies observed in long-term mortality rates, and implementing such methodology to produce forecasts up to 2065.

As a summary, after applying the Lee-Carter method to data from Mexico, García Guerrero & Ordorica (2012) concluded the following:

> Finally, it is significant to highlight the virtue of the method presented, as it enables to have confidence intervals for mortality estimates. The stochastic approach admits that demographic behavior does not follow specific laws: demography –which ultimately studies human groups– is subject to randomness. Moreover, the method allows making permanent adjustments to the forecasts taking into account the number of deaths that could be avoided to reach a life expectancy goal (p. 445).

Guerrero Guzmán & González Pérez (2007) developed a project on the application of the model to data from Mexico, and assessed the impact of the results on the social security system. Lee & Rofman (1994) applied the model in Chile, and concluded that synthetizing statistical techniques for time series and demographic models presents several advantages. In the case of Lee & Carter's, it is a simple and useful model to complete missing data. In some cases, it is even used to adjust under-reporting and forecast mortality.

Andreozzi & Blaconá's paper (2011) presents the first application of the Lee-Carter model (1992) to data from Argentina. Later, in the same line of work, they compared different estimation methods that arouse as variants of the original model (Andreozzi & Blaconá, 2012).

Before that, Hyndman & Ullah (2007) had proposed to smooth mortality through penalized regression splines by using decomposition in main components to adjust the model. They followed an approach called functional data analysis, as the basic input of the models is the total specific smoothed mortality rates.

Although the methods presented produce forecasts for mortality rates, life expectancy at birth is one of the most useful indices to measure overall mortality, as already stated. With this rate, mortality is more accurately estimated than with the crude mortality rate because it does not depend on the population's age structure.

To the data for Rosario city from 1980 to 2015, the present work applies, on the one hand, the pioneer probabilistic forecast model, the Lee-Carter model (1992) and, on the other, one of the last models proposed in the field: the one by Hyndman & Ullah (2007) for functional data. By means of these two models, point and interval forecasts of age-specific mortality rates for the period 1980-2025 are intended to achieve. Moreover, from the calculated mortality rates, forecasts for life expectancy at birth with their respective intervals are obtained.

2. Methodology

2.1 Lee-Carter Model

The Lee-Carter (2012) –LC– is a demographic statistical method for projecting mortality rates. It can be defined as an extrapolative model for univariate time series, as it does not include any data on the technological or social effects on mortality. In other words, it does not seek to add neither external data (exogenous variables) nor opinions on potential events. Basically, the model forecasts the historical trend observed during the period of study with available data. The LC model combines a demographic model with few assumptions along with statistical time-series methods. Thus, it can provide a probabilistic base for forecasts as well as prediction intervals. The change in overall mortality is represented in the model by changing only one index. The LC model allows each specific rate to decrease exponentially without restrictions.

To apply the LC model, it is necessary first to check the quality of available data and their limitations, as the only input is the historical record of mortality rates by age groups. An advantage of this model is that due to its structure, there is no need for the rates to have the same periodicity. However, data are required to cover a considerable period; a proposed rule is that they cover at least a thirty-year period of recent history.

Once the data are obtained and quality has been validated, the next stage consists on adjusting the model to the original data and evaluating historical performance.

When applying the LC model to the rates, these are decomposed into two parameters, a and b (also known as base), and an index, k_t, which represents the overall mortality. The basic premise of the model is that there is a linear relation between the logarithm of specific mortality rates $m_{x,t}$ and two explicative factors: age interval, x, and time, t. The descriptive equation is as follows:

$$m_{x,t} = exp\ (a_x+b_x k_t+e_{x,t})\ t=1,..., n,\ x=1,..., \omega$$

applying the logarithm

$$f_{x,t}=ln(m_{x,t})=a_x+b_{\ x}+k_{\ t}+e_{\ x,t}\ t=1,...,n\ x=1,...,n$$

where

$m_{x,t}$: specific mortality rate for age interval x and year t,

a_x: shape parameter, all these parameters describe the mortality pattern according to age,

b_x: sensitivity parameter. It represents the change in mortality in the interval that starts at age x, when there are changes in the kt index,

$e_{x,t}$: random error,

ω: beginning of the last age interval.

As a_x is a “shape” parameter and all the estimations of such parameter for each age group, it describes the general shape or the pattern of age-specific mortality rates. This parameter is estimated as the simple arithmetic average through the time of the logarithms of each age-specific mortality rate.

Parameter b_x (or base or basis function), called ‘sensitivity’ parameter, describes the change in mortality in the age *x* interval when there are changes in the k_t index. The sensitivity parameter represents the increasing or decreasing intensity of the mortality rate for an age group throughout time. In order to ensure a single solution, a value of 1 is imposed to the sum of b_x values; and to the sum of k_t values, 0 is imposed. This estimation is made using Singular-Value Decomposition (SVD) and, although there are other proposals to make the estimation, such as the minimal squares or the maximum likelihood estimation, by using SVD –which implies a superior computational cost– the results obtained are much better.

In the LC model, the estimated k_t index is linear. Therefore, the mortality rate for each age group changes according to its own exponential rate. Thus, in this model, there will not be mortality rates with negative values. By making all mortality rates dependent on the k_t parameter, the method is far from forecasting each specific mortality rate independently. This saves having to calculate the different covariances among errors. Moreover, all individual rates necessarily belong to the general system of mortality, which matches observed historical data.

Subsequently, the overall mortality is estimated through time series methods and, finally, age-specific mortality rates are point forecast by confidence intervals.

2.2 Models for functional data

During the past years, many probabilistic forecast approaches have been developed (Booth, Hyndman, Tickle & de Jong, 2006), and within this category, models for functional data have become particularly relevant (Ramsay & Silverman, 2005). Such methods, which are quite recent, constitute a new framework to analyze time series, which has been adopted, among other things, to carry out forecasts for all demographic components (Hyndman & Ullah, 2007). The model for functional data is an extension of the LC model, since it decomposes mortality according to age patterns and its behavior throughout time. However, there are two differences between them: first, the model for functional data is always applied on smoothed data (functions of observed data); and second, it uses multiple bases instead of a single index, as it is the case of the LC model, which calls it k_t.

The following model is proposed to be used for $y_t(x)$ transformed observations:

$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$s_t(x) = \mu(x) + \sum_{k=1}^{k} \beta_{t,k}\, \phi_k(x) + e_t(x),$$

where $s_t(x)$ is a subjacent smooth function of x, $\varepsilon_{t,x}$ are random variables, independent and identically distributed, and the definition of σt(x) allows the variance to change according to age and time. This means that the transformed observations are the addition of the quantity to model, $s_t(x)$, a smooth function of the age and an error (first equation). The second equation shows the dynamics of $s_t(x)$ throughout time, where μ(x) is the mean of $s_t(x)$ throughout time, $\{\phi_k(x)\}$ is a group of K orthogonal basis functions calculated through a decomposition into main functional components of the matrix $[s_t(x)- \mu(x)]$ and $e_t(x)$ is the model's error (which is considered non-serially correlated). The dynamics of this process is controlled by coefficients $\{\beta_{t,k}\}$, which behave independently (ensured by the use of the main components method).

Hyndman, Booth & Yasmeen (2013) made an interesting contribution to this approach by introducing the idea of coherent forecasts into the functional data paradigm. The proposal's main idea is that the difference between forecasts for target groups must be kept constant throughout time, therefore, reproducing the existing relation in the observed data. The target groups can be geographical sub-regions or sexes, just to mention some examples.

3. Implementation in Rosario city

The basic input for both models is the numbers for population and mortality[3] by age and sex during the 1980-2015 period. In the case of mortality, the data were provided by the General Directorate of Statistics of Rosario city and, for population, by the INDEC (Argentina's National Institute of Statistics and Censuses).

Data are disaggregated into 19 groups[4]. Software R (R Development Core Team, 2008) is employed to make the necessary calculations.

To apply the statistical models (LC and FDM) the first step is to smooth the rates observed through age groups. Graphic 1 shows mortality rates for the total population, including males and females, which is the basic input for both forecast models. Curves represent a high level of mortality at the beginning of life, a notable reduction before birth, reaching the lowest levels between the ages of 5 and 15. From this point onwards, it increases up to older ages. In Argentina, as well as in Rosario, there is a relative maximum between ages 18 and 25 for males. Such high level is linked, in this case and for most populations, to traffic accidents, drugs abuse and violent deaths in general. Serfaty, Foglia, Masautis & Negri (2007) agree that in Argentina, the 10-25 age group represents 27% of the population and, in recent years the available knowledge about it has been broaden. In this sense, the authors conclude that although this group does not get clinically ill frequently, they are vulnerable to violence-related mortality: accidents, suicide and homicide. After this spike, rates show a slight decrease just to experience a sustained increase up to older ages.

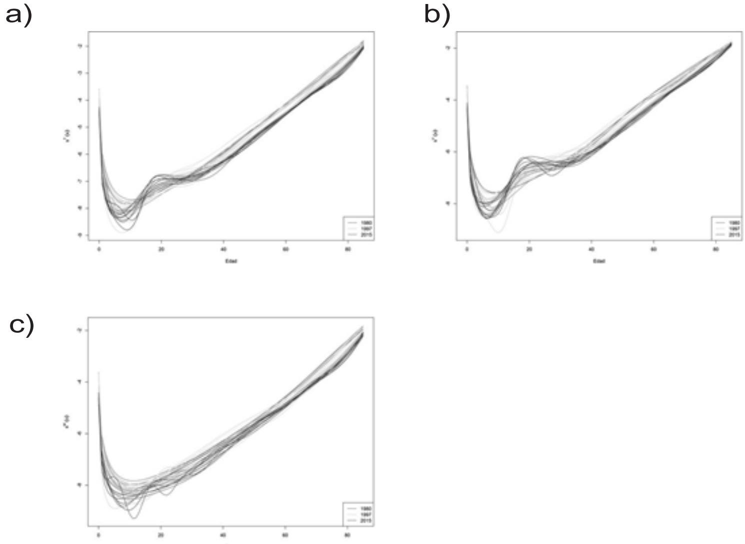
Visually, smooth curves show a marked decrease in mortality levels throughout time, visible at every age, except for the aforementioned male group, which presents what it seems to be a slighter or more fluctuating decrease. When evaluating long-term dynamics (study periods of 100 years or more), the general decrease in mortality levels is mainly related to medical improvements. Therefore, the decrease observed in this period can be linked, in part, to those causes and to others, more specific from a particular historical period in Argentina.

---

3 It is worth mentioning that for several decades Rosario city has had a high quality management of life events thanks to the coordination and articulation among different areas in charge of producing data.

4 A first age group: 0 years old, 1 to 4 years old and quinquennial age groups: 5-9, 10-14, etc. up to the final open group: "85 years and older".

Graphic 1. Smoothing of mortality rates. Rosario, 1980-2015.
(a) Total population, (b) Males, and (c) Females



Note: the graphic uses a rainbow scale, implemented by Hyndman, and adds an inscription including references to minimal, medium and higher values in a color scale to guide the reader.
Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

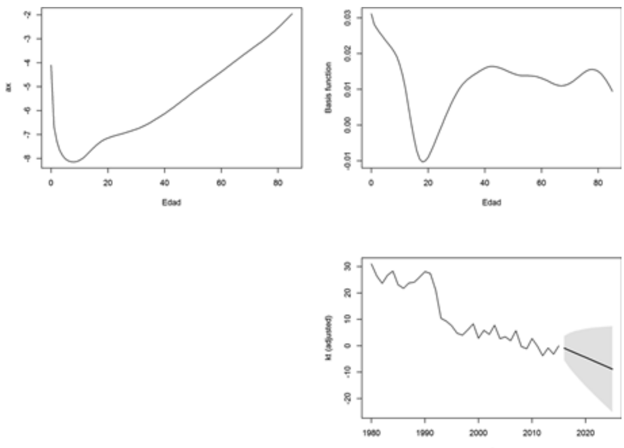
3. Lee-Carter Model for Mortality in Rosario city

By applying the Lee-Carter model, a kt overall mortality rate index is obtained, which represents the general behavior of mortality throughout time (Graphic 2). The forecast interval is represented in grey shade. Apart from the fluctuations, the graphic shows that during the period studied, mortality decreased with a similar structure for males and females. Notably, the three evaluated cases show a slight increase in mortality around 1990. It would be interesting to evaluate this fact in demographic terms, since in other cases, such as Buenos Aires, during the 1990s there was over-mortality due to HIV, which mainly affected males. For this reason, it would be adequate to analyze mortality based on its causes and even to evaluate whether there was a particular event in history that might have caused this relative maximum, or whether the fluctuation was random.

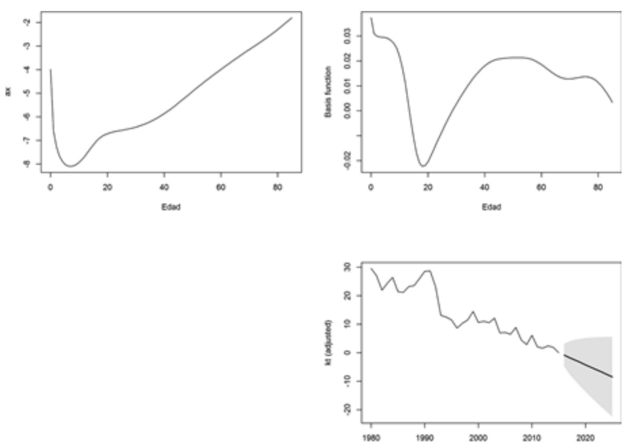
On the other hand, the b parameter shows which ages contribute, to a greater or lesser extent, to the overall decrease, which is shown by k. For males, it shows the minimum around age 18, which means this age and its adjacent values are the ones that contribute the less to the decrease. This supports the aforementioned increase of mortality in younger ages. For females, the b parameter behaves differently: initial ages, the 30s and 80s present higher values, which means that these age groups contribute the most to the decrease of female mortality. To calculate forecasts, the overall mortality is predicted through models for time series and the estimated values of the mortality rate for future years are generated, assuming that the behavior of the b parameter remains steady.

Graphic 2. Mean, basis function and overall mortality rate estimations (with 80% forecast interval) by using the Lee-Cartel model. Rosario, 1980-2015. (a) Total population, (b) Males, and (c), Females

a)



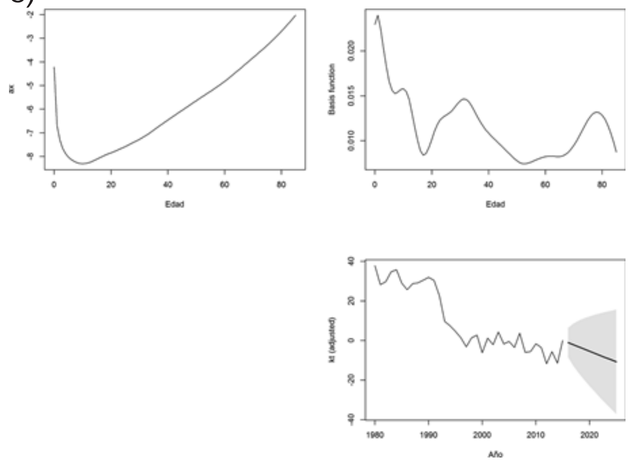
b)



(continúa...)

(continúa...)

c)



Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

Taking mortality rates forecasts as starting point, it is possible to obtain life expectancy forecasts at birth by using life tables. This case shows values up to 2024 (Chart 1). Although the values obtained are the true output of these models and the focus is on them when designing and planning public policies, the graphic showing the results can help evaluate whether the results are consistent with observed data. Thus, Graphic 3 shows series representing life expectancy at birth for the available period (1980-2015), which are called “observed expectancies”, and the lines representing their forecasts, “forecast expectancies”, for these three groups: total population, males and females. The three groups show a slight increasing trend in the lines representing forecasts. They reach 78.06 years of life expectancy for total population, 74.49 years for males, and 81.08 for females. A way of evaluating the quality of forecasts is through interval width (the wider the interval, the less informative the forecast). If the width of the intervals obtained by LC was to be estimated, the result would be 4.49 years. Therefore, it would be interesting to compare this data with those obtained for the following models:

Chart 1. LC forecasts of life expectancy at birth and 80% forecast intervals. Rosario, 1980-2015. Total population, Males and Females

Year	Total	Male	Female
2016	76.60(75.66-77.50)	73.24(72.41-73.92)	79.63(78.51-80.64)
2017	76.66(75.53-77.99)	73.39(72.38-74.28)	79.79(78.15-81.39)
2018	76.93(75.55-78.32)	73.53(72.24-74.60)	79.95(78.02-81.88)
2019	77.09(75.45-78.55)	73.67(72.27-75.01)	80.11(77.83-82.35)
2020	77.25(75.51-79.02)	73.81(72.27-75.26)	80.27(77.63-82.72)
2021	77.41(75.37-79.57)	73.95(72.22-75.53)	80.43(77.53-83.48)

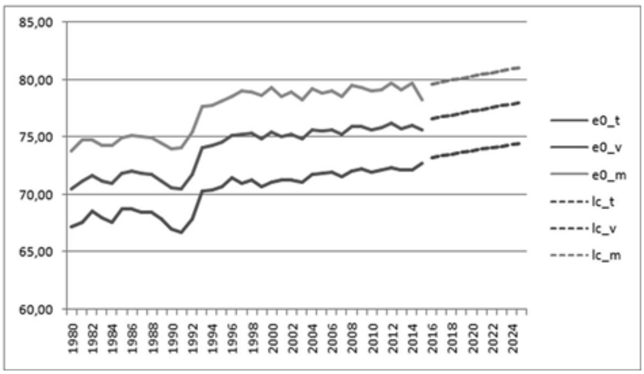
(continúa...)

(continúa...)

2022	77.57(75.29-79.65)	74.09(72.27-75.63)	80.60(77.51-83.73)
2023	77.74(75.25-79.88)	74.22(72.26-75.92)	80.76(77.20-84.24)
2024	77.90(75.11-80.03)	74.35(72.46-76.06)	80.92(77.09-84.34)
2025	78.06(75.34-80.67)	74.49(72.33-76.32)	81.08(77.37-84.82)

Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

Graphic 3. Life expectancy at birth observed (1980-2015) and forecast through the LC model (2016-2025). Total population, Males and Females



Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

3.1 Functional data model of mortality in Rosario city

Applying the functional data model (FDM) to mortality rates allows decomposing rates through coefficients, bases and mean. This decomposition is applied to total population, males and females.

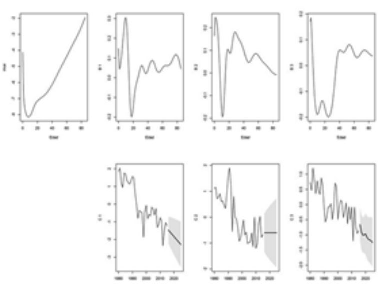
Graphic 4 shows the mean, bases (their 80% forecast intervals), estimated functional coefficients for total population, males and females. The sub graphics (one for the total and another for each sex) are to be interpreted by column. The first column shows the average behavior of mortality throughout the ages. Then, from the second column, the upper row shows bases and the lower row, coefficients. Therefore, each column has a pair base-coefficient that must be interpreted together. For example, in the case of mortality for total population, the first coefficient shows a decrease in mortality throughout time (second column, lower row). However, for a correct interpretation, this behavior must be paired with the corresponding base (lower row), as the base shows at which ages and degree of intensity the decrease appears, in the same way b and k were related in the LC model.

In this case, the estimated mean represents the average profile of mortality throughout life (obtained as an average of the functions throughout time). An analysis of the pair base-coefficient for males shows that ages 10 to 14 contribute to the decrease with fluctuations of the first coefficient (according to the corresponding base), while ages closer to 18 contribute the less. The second pair base-coefficient could represent (according to the base) the differential value in mortality for initial ages versus age 10, as there is a rather high value for the first group and a rather low one for the second group. Moreover, the corresponding coefficient would show that this represents a decrease with fluctuations throughout time. Beyond the third pair base-coefficient, it is hard to obtain a demographic relation of the results.

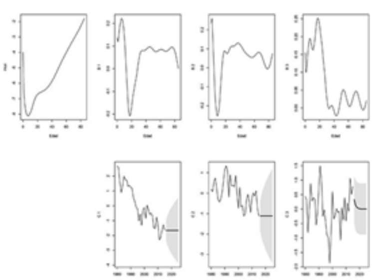
In the case of females, the first pair shows that females between 20 and 40 years are the ones that contribute the most to the decrease in mortality expressed by the coefficient. In turn, the second pair base-coefficient can be related to the decrease shown in initial ages (less than 5) and the older ones (80 and more) since the base presents high values for both ages.

Graphic 4. Mean, basis functions (with 80% forecast intervals) and coefficient estimations using the Functional Data Model. Rosario, 1980-2015. (a) Total population, (b) Males and (c) Females.

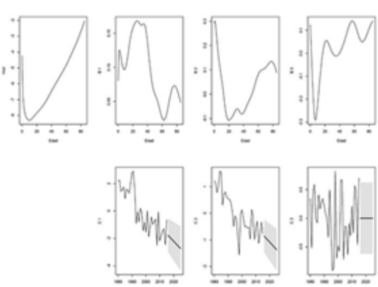
a)



b)



c)



Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

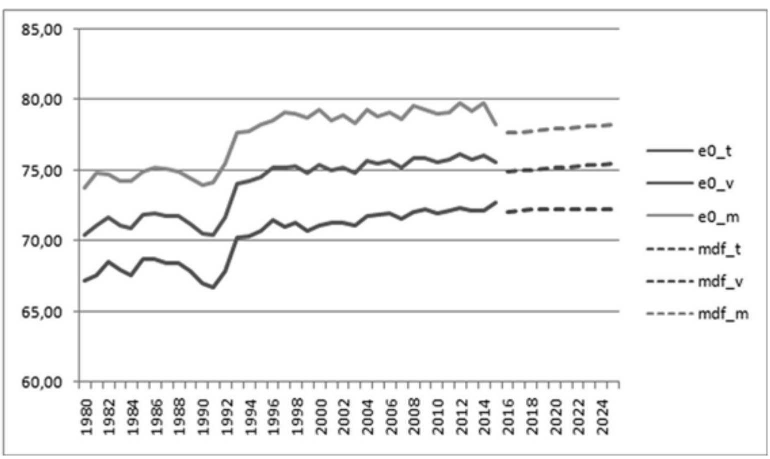
As it is the case for the LC model, life expectancies at birth are calculated from mortality rates for future years by sex, with their corresponding forecasts intervals by means of the FDM. Chart 2 shows life expectancies forecasts for 2016-2025, where the average amplitude for these intervals is 4.46 years. Moreover, Graphic 5 shows a clear divergence in the behavior of forecasts: female life expectancy increases over time while male life expectancy is stagnant. These results do not seem to agree with the behavior of the same measure for different regions and cities of Argentina (Andreozzi & Blaconá, 2012).

Chart 2. FDM forecasts of life expectancy at birth and 80% forecast intervals. Rosario, 1980-2015. Total population, Males and Females

Year	Total	Male	Female
2016	76.34(75.27-77.46)	73.05(71.91-74.24)	80.83(79.26-82.65)
2017	76.55(75.36-77.84)	73.09(71.60-74.46)	80.97(79.29-82.86)
2018	76.69(75.37-77.98)	73.10(71.37-74.70)	81.14(79.31-82.95)
2019	76.76(75.29-78.15)	73.10(70.94-74.80)	81.31(79.50-83.11)
2020	76.82(75.46-78.38)	73.10(70.85-75.01)	81.47(79.64-83.52)
2021	76.96(75.41-78.41)	73.10(70.69-75.18)	81.64(79.73-83.61)
2022	77.07(75.33-78.61)	73.10(70.53-75.30)	81.81(79.78-83.92)
2023	77.16(75.38-78.80)	73.10(70.12-75.55)	81.98(80.02-84.16)
2024	77.25(75.36-79.03)	73.10(70.25-75.64)	82.15(80.20-84.24)
2025	77.36(75.41-79.04)	73.10(70.01-75.75)	82.32(80.15-84.48)

Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

Graphic 5. Life expectancy at birth observed (1980-2015) and forecast through FDM (2016-2025). Total population, Males and Females



Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

To conclude, FDM is applied with constraints that allows obtaining coherent estimated forecasts for both sexes. The proposal's main idea is that the difference between the forecasts for target groups must remain constant throughout time, reproducing the relation present in the observed data. The target groups can be geographical sub regions or sexes (to provide some examples) for which new values and their corresponding forecast intervals can be found. In other words, the coherent forecast method intends to ensure that forecasts for related populations keep some structural relation based on historical data and theoretical considerations. For example, observations have shown that male mortality is higher than female mortality at all ages. While the available evidence holds both the biological/genetic and the socio/cultural/environmental/behavioral hypotheses, Kalben (2000) concludes that the determining factor is biological. It is expected that the sex-differences in mortality will persist in the future within observed limits.

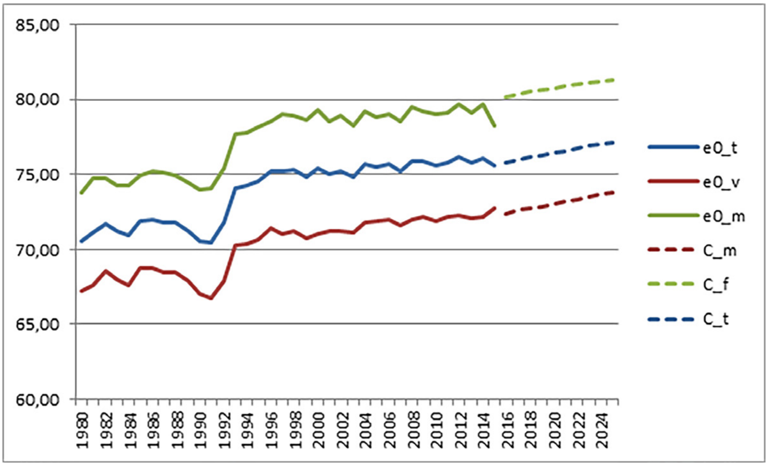
Thus, forecasts of life expectancy at birth are obtained and these account for the aforementioned difference between sexes. These are shown in Chart 3 together with their 80% forecast intervals. The average amplitude for these intervals is 2.18 years, being the lowest of the three measures obtained. In turn, Graphic 6 shows an increasing behavior in the lines representing life expectancies for the three groups. This methodology forecasts a life expectancy at birth of 77.17 years for the total population, 73.82 for males and 81.36 for females for 2025.

Chart 3. Coherent FDM forecasts of life expectancy at birth and 80% forecast intervals. Rosario, 1980-2015. Total population, Males and Females

Year	Total	Male	Female
2016	75.82(74.88-76.53)	72.41(71.51-73.25)	80.14(79.20-81.06)
2017	76.03(75.12-76.93)	72.61(71.70-73.44)	80.37(79.45-81.35)
2018	76.17(75.20-77.03)	72.73(71.68-73.65)	80.53(79.66-81.51)
2019	76.31(75.37-77.19)	72.88(71.83-73.87)	80.66(79.63-81.79)
2020	76.46(75.39-77.37)	73.04(71.88-74.20)	80.78(79.67-81.98)
2021	76.60(75.51-77.57)	73.20(72.14;74.25)	80.89(79.78-81.95)
2022	76.74(75.72-77.95)	73.35(72.30-74.48)	81.01(79.82-82.33)
2023	76.89(75.80-78.10)	73.51(72.35-74.82)	81.13(79.96-82.33)
2024	77.03(75.99-78.10)	73.66(72.47-74.81)	81.24(79.98-82.47)
2025	77.17(75.96-78.43)	73.82(72.61-75.16)	81.36(80.11-82.68)

Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

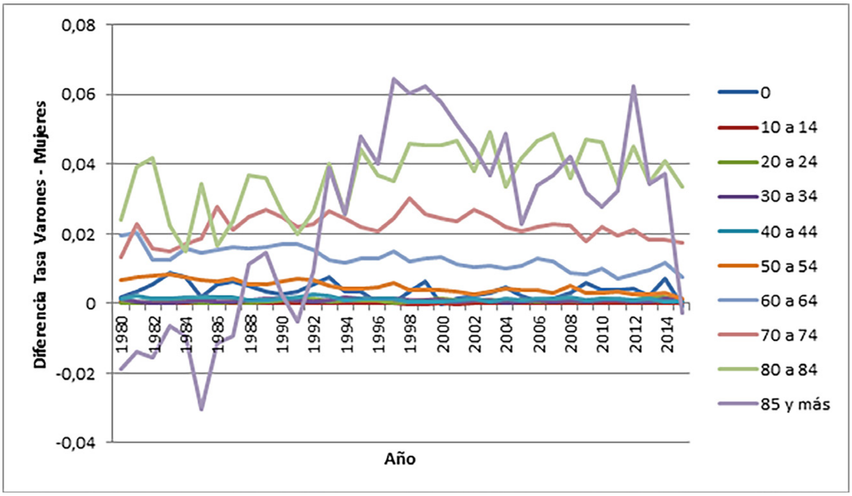
Graphic 6. Life expectancy at birth observed (1980-2015) and forecast through coherent FDM (2016-2025). Total population, Males and Females



Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

To conclude, Graphic 7 shows the differences between male and female rates for the period 1980-2015: for most ages, the difference between the observed rates for both sexes remains constant over time, except for the last age interval analyzed. However, at present, this interval is wide, as it covers ages that display different behaviors due to an increase in life expectancy. Thus, it would be appropriate to rely on more disaggregated data in future. If the analysis of the last open interval is not taken into account, it is clear that the remaining ages show constant differences between sexes, reinforcing the usage of this model for coherence.

Graphic 7. Difference in observed male and female mortality rates (selected ages). Rosario, 1980-2015.



Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

Moreover, goodness of forecast measures can be calculated, which assess the performance of the models when estimating future values. In order to do so, the modeling has to exclude a set of five values (2011 to 2015) and then calculate the mean absolute percentage error (MAPE) (Chart 4). The coherent forecasts model has lower values, which implies that this model is better for forecasting.

Chart 4. Mean absolute percentage error (2011-2015)

	LC	FDM	Coherent FDM
Total	6.43	0.35	0.39
Females	6.17	2.00	1.63
Males	0.22	0.46	0.45

Source: produced by the authors, based on data provided by the General Directorate of Statistics, Rosario Municipality.

4. Final considerations

The ultimate aim of this work is to obtain estimations of the life expectancy at birth in Rosario city for the period 2013-2025. This is the reason the selection of the best result comprises statistical and demographic criteria. The coherent FDM is the model that presents the best results, as it has the narrowest forecasts intervals and trajectories that accounts for the differences between sexes. Moreover, it is also suitable from a demographic viewpoint, compared with the FDM, which shows stagnation in the case of males. In addition, it shows the best quality of forecasts measured by MAPE.

It is important to highlight that the coherent FDM model is a modern methodology and that several official statistics agencies use it worldwide to create forecasts of different demographic components and measures. However, it is worth mentioning that this model is still being developed and that its application in different regions depends on several factors, such as base input quality and accuracy: vital statistical and census data. In this regard, it is important to continue studying the results through the analysis of base-coefficient pair interpretations related to mortality by cause and historical events, to enhance the results obtained here.

It should be pointed out that functional models can also be used to analyze and forecast fertility and migration. These results, along with the mortality results, enable probabilistic population projections. In addition, it enables forecasts of measure derived from the rates, such as the total fertility rate, which is obtained from fertility rates or forecasts of the dependency ratio, determined from the population by ages, among others. Moreover, each derived measure is estimated and point forecast by confidence intervals.

These results are particularly important when outlining public policies since the intervals provide data on the inferior and superior limits related to point estimation, so there is a degree of uncertainty. It is quite possible for official statistics agencies to implement these methods since the only input used would be the data collected by the city, and since calculations are made through the free software called R. Thus, this article constitutes a proposal to use a modern methodology that has been successfully adopted by statistics agencies of several countries worldwide.

Bibliographical References

Alho, J. M. (2000). A statistical look at Modeen's forecast of the population of Finland in1934. *Yearbook of Population Research in Finland, 36*, 107-120.

Andreozzi, L. & Blaconá, M. T. (2011). The Lee Carter method, for estimating and fore-casting mortality: an application for Argentina. *Presented atthe ISF 2011 Prague - Proceedings*. http://www.forecasters.org/submissions/>ANDREOZZILUCIAISF2011.pdf

Arriaga, E. (2014). *Análisis demográfico de la mortalidad. Centro de investigaciones y estudios sobre Cultura y Sociedad*. Córdoba, Argentina: CONICET & UNC.

Booth, H., Hyndman, R., Tickle, L., & de Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research, 15*(9), 289-310.

Blaconá, M. T & Andreozzi, L. (2012). Comparación de métodos de estimación del modelo de Lee y Carter. *Estadística, 64*(182 y 183), 57-84.

García Guerrero, V. M. & Ordorica Mellado, M. (2012). Proyección estocástica de la mortalidad mexicana por medio del método de Lee-Carter. *Estudios Demográficos y Urbanos, 27*(2), 409-448.

Guerrero Guzmán, V. M. & González Pérez, C. (2007). Pronósticos estadísticos de mortalidad y su impacto sobre el sistema de pensiones de México. *Prize-winningwork, Premio de Pensiones 2007 granted by Comisión Nacional del Sistema de Ahorro para el Retiro, 2007*. http://www.consar.gob.mx/pre-mio_ pensiones/premio_2007.shtml

Hyndman, R. J. with contributions of Booth, H., Tickle, L. & Maindonald, J. (2017). *Demography: Forecasting Mortality, Fertility, Migration and Population Data*. Rpackage version 1.19. https://CRAN.R-project.org/package=demography

Hyndman, R. & Ullah. H. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis, 51*, 4942-4956.

Hyndman, R. & Booth, H. (2008). Stochastic population forecast using functional data models formortality, fertility and migration. *International Journal of Forecasting*, 24, 323-342.

Hyndman, R., Booth, H. & Yasmeen, F. (2013). Coherent Mortality Forecasting: The Product-Ratio Method WithFunctional Time Series Models. *Demography, 50*(1), 261-283.

Kalben, B. (2000). Why men die younger. *North American Actuarial Journal, 4*(4), 83-111.

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

Lee, R. D. & R. Rofman(1994). Modelación y Proyección de la Mortalidad en Chile. *Notas de Población, XXII*, 59,182-213.

Lee, R. D. & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association, 89*, 1175-1189.

Raftery, A. E. , Li ,N. , Sevˇc´ıkov´a, H. , Gerland, P. & Heilig, G. K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences, 109*,13915–13921

Ramsay, J. O. & Silverman, B. W. (2005). *Functional data analysis*. 2nd ed. New York, USA: Springer.

R Core Team (2014). *R: A language and environment for statistical computing*. Viena, Austria: R Foundation for Statistical Computing. http://www.R-project. org/

Serfaty, E., Foglia, L., Masautis, A. & Negri, G. (2007). Mortalidad por causas violentasen adolescentes y jóvenes de 10- 24 años. *Revista Vertex, 40*, 25-30.