Closing-out the Algerian life tables : for more accuracy and adequacy at old-ages

Farid FLICI *

1st February 2016

(*) Permanent Researcher at Centre of Research in Applied Economics for Development (CREAD). Street Djamel Eddine El Afghani – BP° 197, Rostomia – Algiers, Algeria. Mail : farid.flici@cread.edu.dz

Abstract

The usefulness and the use of the life tables highlight the need to precision in their construction. In this sense, Closing-out of life-tables has a great importance. Given the inefficiency of mortality data beyond a certain age, particularly in developing countries, the model life tables are an unavoidable solution for the estimation of mortality at older ages. As an international standard, this tool provides approximate estimations which are not specified for a particular country ,thus the obtained results are not always satisfactory, especially when these model life tables are not used in the right way. Thereby, estimating the old ages mortality by extrapolating the observed trend on the available data is assumed to provide more relevant and coherent results. The specifics of mortality at advanced ages requires the use of appropriate models instead of conventional models.Therefore, A set of models is designed for this purpose. In the present paper, we evaluate and compare some proposed old ages mortality models to extend mortality rates beyond the age of 80 for the Algerian population. The comparison will be based on various criteria : Goodness-of-fit, predictive capacity, sex differential coherence, predicted age limit and single isexes vs both-sexes adequacy. The final results will be used to correct the historical series of life expectancy at birth for Algeria between 1977 and 2014 and will serves as a basis for old age mortality extrapolation of dynamic life tables.

Key-words: mortality rates, goodness of fit, predictive capacity, BIC, MSE, extrapolation, life expectancy, Algeria.

JEL codes : J110 , J190.

1 Introduction

The improvement of life expectancy in Algeria during the recent decades has resulted from a mortality reduction at all ages (Flici and Hammouda, 2014). As a result of this reduction, the probability to reach the age of 80 passed from 0.74 in 1977 to 0.95 in 2014, achieving 22% of reduction in the space of 37 years. In addition, The chance to reach the age of 60 marked a improvement of 15% passing from 0.82 in 1977 to 0.96 in 2014. Parallel to this, the size of the population aged 60 and over increased from 790 000 according to the General Census of Population and Housing (GCPH 1966) to more than 2.5 millions in 2008. The population aged 75 and over was estimated according to the results of 2008's Census of about 683 500 individuals. The size and the part of the old-age population should continue to increase during the forthcoming century. According to such a perspective, the old age mortality should be attentively considered as well as the mortality at younger ages.

A life table is supposed to describe the mortality level lived by a population at all ages, from birth till total extinction. For each age x, that life tables might give the corresponding risk of death. When mortality data are available and reliable, this kind of calculation can be made easy. But In practice, the quality and the availability of this data is varying with age and adult ages are usually well covered compared with lower and higher ages. The lack of the mortality data and the exposure to the death risk beyond a certain age leads to some irregularities in mortality curves. For a long time, some conventional mortality models, used in principal to graduate adult mortality curves, have been used to extend mortality to the older ages. The most practical example in this sense is the fact that the United Nations Population Division have used a Makeham-type function to extend the model life tables until the age of 85 years. For the Coale-Demeny model life tables revised in 1983, a Gompertzian function was used to extend mortality rates until 100 years old (Coale and al., 1983). All these methods were used under the assumption that mortality rates keep growing at old ages according to the same acceleration rate observed on the adults ages. As mentioned in Heligman and Pollard (1980), when childhood and young ages are surpassed , the mortality rates can be represented by an exponential function (Gompertz, 1825). Parallel to this, the deceleration of mortality rates at older ages was noticed in many works starting by Gompetrz and then Pearks (1932) (see Gavrilova and Gavrrilov, 2014). Unfortunately, assessment of such a preliminary finding needs various comparisons with a real observed data. The improvement of the population life conditions in developed countries resulted to the improvement of life expectancy at birth and the growth of centenarian population. By the end of 20^{th}century, the good data quality reashed by some developed countries provided a consistent old age population data bases allowing more flexibility in studying and modeling old age mortality. Coale and Guo (1989), in the third version of model life tables, confirmed mortality deceleration on the observation of mortality rates until the age of 100 for some countries (Netherlands, Japan, France, West Germany, Austria, Sweden, and Norway). Their idea was that even if mortality rates keep growing with age, the growth rate starts to back down starting from a certain age (80 or 85). This deceleration takes a linear trend. The same approach with some adaptation was developed one year later by Coale and Kisker (1990). This approach is still used nowadays for closing out the model life tables.

There are some observed differences between demographers and actuaries in terms of old-ages mortality modelling practices. Other methods were developed separately by actuaries, but there is some convergence regarding the final objective which is to set more accuracy to the mortality curve span at old age. Demographers focused generally on the life expectancy at birth. The effect of the Closing-out method is not very important in calculating life expectancy at birth (Buettner, 2002). The effect is more important in terms of life expectancy at 50 or 60 which coincides with the age range which actuaries are interested in.

For Algeria, the national life-tables published recently (2010-2014) by the Office of National Statistics (ONS) are closed-out at the open age group [85 and +[. For the years before 2010, the closure age was bellow. On the basis of the selected model life table, the corresponding life expectancy at the age of closure is estimated. Combined with the other parameters calculated for the youngest ages, the life expectancy at birth is concluded. The mortality rates corresponding to the oldest ages are not published in any way. Even if the chosen life expectancy is estimated and used, the adequacy of the corresponding mortality rates at old-age do not fit necessarily the existing mortality curves. A change in the reference model life tables or a wrong use of this last can have an effect on the estimated life expectancy at birth. In Flici and Hammouda (2014), we have shown how

this practice can affect the evolution of the life expectancy at birth and how it can distort all analysis about its evolution.

To estimate old age mortality summarized by the life expectancy at the closure age, we need to select an adequate model life tables among Coale-Demeny (CD) or the United Nations (UN) model life tables. Between the two, African countries prefer use the first model life tables to complete their mortality data (Ita and Banjan, 1984). Mortality of the northern African countries is generally represented the south type of CD life tables while the North type is used to describe the mortality pattern of sub-Saharan countries. However this procedure implies many inadequacy. In practice, the method of stable population is used to select the adequate model life table. As shown in (Ita and Banjan, 1984), this process leads generally to an inadequate selection compared with the use of the ordinary least squared deviations method. In addition of these two types, CD model life tables includes two other types : West and East types, and data of any African country has been considered to construct these model life tables. Here, the use of CD model life table can, in any way, guarantee a good estimation but just the best among the existent method to estimate old age mortality.

Model life tables were principally conceived to allow estimating life tables from unreliable or incomplete national mortality data. Estimating mortality at older ages for developing countries is considered from this point of view. A such estimation process can be very useful when the shape of the mortality curve at adult ages is not as regular to allow a consistent extrapolation to the older ages. However, another problem can rise if we consider this estimation process by the view of the trend of the estimated old age mortality rates and its compatibility with the adult age observed trend. Each type is based on the observation of a set of life tables having a similar mortality schemes ordered in equidistant scale in function of the corresponding life expectancy at birth. The selection process is based on the comparison of the national mortality data; at adult ages, to the mortality pattern given by the different types of model life tables. The residual life expectancy at the age of closure is directly deduced. Then, the life expectancy at birth is deduced by recurrence. In some cases, a moving average adjustment around the connection point is needed, because the two curves have basically a different evolution trends.

Our first objective here, is to reduce the fluctuations due to closing out method changes and to give more homogeneous old age mortality surface for the Algerian population. The second objective is to provide readers with a general methodology to extend mortality to the old age for the Algerian population starting from adults mortality curves. The dynamic life tables or some specific life tables which can be estimated on the basis of the Algerian mortality data should be closed out following a methodology taking into account the specificities related to Algeria. Here, we aim to make evidence of these specificity.

For this, some old age mortality models will be compared to extend mortality beyond 80 years on the basis of the observed mortality curves shape at youngest ages. The extrapolation results are generally related to the quality of the existing data and also to the shape of the mortality curve at the age range used as a basis of the extrapolation. This last can vary from a population to another and from a period to another in the same population according to the regularity of the data. Generally, starting from a certain ages (35, 40 or 45) the mortality rates grow following an exponential function. This regular trend allows to extend easily mortality to the older ages. But the final result is very related to the method and the age range chosen to calibrate the model, in the absence of observed data to be used for comparison. Some other criteria can be used to orient the

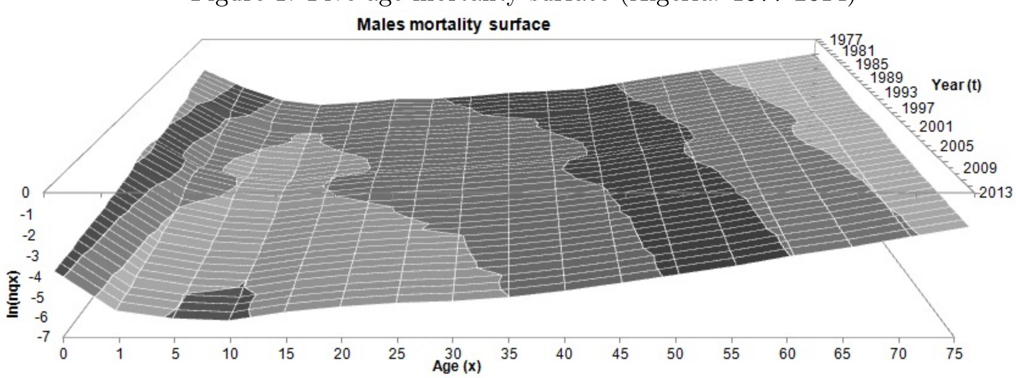
model calibration. Usually, an assumption about the ultimate surviving age is defined. The mortality rate at that age might converge to 1. The mortality curve for the Algerian population has changed many times over the observed period and the challenge is that it is so complicated to find a unique model which allows a good fitting quality all over the period ensuring some adequacy between males and females and an adequacy regarding to the year-to-year mortality rates varying. Here, we will try a set of proposed methods.

2 Data

2.1 Presentation

For the needs of the present work, we will use the Algerian life tables published by the ONS during the period 1977-2014. The missing data has been estimated in a previous work (Flici, 2015). So, our database is composed by a five age mortality rates from 0 to 75 years during 1977-2014.

Figure 1: Five-age mortality surface (Algeria: 1977-2014)



Considering the closure age, the Algerian life tables have been stopped at the open age group [85 and +[since 2010 only. For the previous years, the closure age varied between [70 and +[and [80 and +[. (see Flici and Hammouda, 2014 and Flici, 2014 for more details). for the present application, we use a completed mortality surface with the open age group [80 and +[as a common closure age for all the Algerian life table from 1977 until 2014.

Some old age mortality models are more suitable for single ages data. For this issue, we propose to interpolate the single age mortality surface from the data that we have until now and which is structured following a five ages structure.

2.2 Single-ages mortality surface

In the literature, several methods were proposed to interpolate single-age values (Exposure, Observed deaths, Deaths rates) from five-age description (king, 1908; Sprague, 1880; Beers, 1944 and 1945). Usually, some complications can be found because of the structure of the abridged life-table. The length of the age groups is 5 years except the two first groups [0-1[and [1-5[. Some interpolation methods were proposed to deal with

unequal age-intervals principally based on the Karup-King and Lagrange interpolation methods (Crofts, 1998; Srivastava and Srivastava, 2012). For the present work, we preferred to use both Lagrangian and Karup-King methods (Shrock and al., 1993). The first method is more adequate with the lower age groups while the second is more adapted with the higher age groups. The connection point is defined in the way to minimize the gap between the two curves around this points.

The connection point is different by period and sex.

The obtained single age mortality surface for males and females is shown in figure 3.

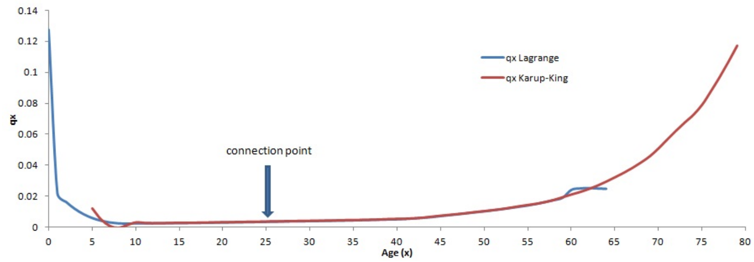
A significant part of extrapolation models of mortality at older ages are based on a single ages mortality description. For this aim, and to do not miss the opportunities that may present a model or another, we saw that it is necessary to interpolate death rates before extrapolation to the older ages.To reach this purpose, we have used and compared a set of interpolation methods: Lagrange, Karup-King and Sprague. For the interpolation of the extreme class (75-80 years), the first method requires the disposal of mortality rates up to 95 years. In the absence of suchdata, we saw that an extrapolation based on models unsuitable for old ages mortality doesn't guarantee a good extrapolation quality. The Lagrange method can be used until 65 only, and, the other two methods have the disadvantage of interpolating the quotients for only five ages groups. The age group [1- 4 years] in this case is our problematic. The breaking-out of the whole$_5q_0$ without distinction of infant and child mortality will leads to an overestimation of death quotients at 1, 2, 3 and 4 years at the expense of infant mortality rate. This last is relatively high by nature and remains a key indicator in the abridged and detailed life tables. In order To mitigate these drawbacks, the idea is to combine two formulas Lagrange (0 - x years) and karup-king (x - 80 years). X being the connection age which is fixed to optimize the quality and the regularity of the interpolated series. However, it remains preferable that the connection point corresponds to a five-year age, in order to avoid the possibility of gaps if we aim to restore the five-age mortality rates.

If we consider $q_x^*$to be the mortality rate at age (x) obtained with the Karup-king Method and $q_x^+$is the rate obtained with Lagrange for the same age (x), the connection point (k) is defined in order to minimize the Mean Absolute Errors (MAE) between the two curves around this point. This implies :

$$MinA = \frac{1}{5}\sum_{x=k-2}^{k+2} \left\| q_k^* - q_k^+ \right\|, ...k = 20 - 50,$$

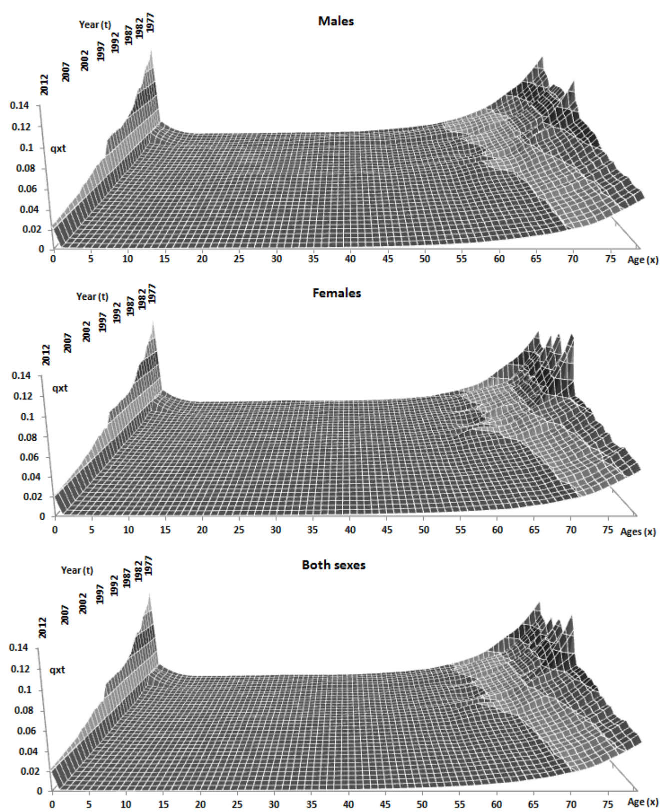
The connection point corresponds to the point where the gap is minimal. This coupling point varies from a table to another (Both sexes, male, female) and from a year to another. The following figure traces the coupling of two sets of quotients: quotients obtained by the method of Lagrange and those obtained with Karup King for male life table of 1977:

Figure 2: curves junction : Karup-King VS Lagrange



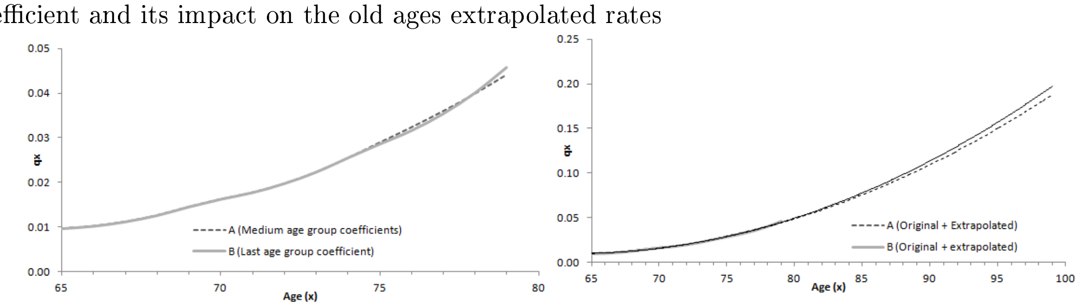
As shown in Figure 2, the connection point is 25 years. We can easily notice that the method of Karup-King gives unsatisfactory results for ages below 10 years. The same is observed with the Lagrange method for the age group of 60 and over, for all tables. Automatically, the connection age is chosen in the age range [10, 55]. For the 1977's males life table, the mean squared deviation recorded its lowest value (3.79E-11) in the neighbourhood of 25 years. Thus, from 0-25 years we used Lagrange interpolated rates while Karup-King's interpolated rates are adopted for ages of 25 years and over. The obtained single ages mortality rates are shown in the figures below:

Figure 3: Single ages mortality surfaces



Although these two formulas are used to interpolate the single ages mortality rates from the five ages rates by interpolating the observed deaths, the interpolation quality can not be of the same degree in the extremes ages as in medium ages. The reason is thatthe interpolation formula is based on a combination of the trends before and after thetargeted age group. Therefore, these interpolation methods assign coefficients to reconstruct the single age deaths number which is a fraction of five age deaths. For extreme classes, the interpolation is based solely on the basis of previous classes and the target class itself, using other coefficients that are different from those used to break out the medium classes. The adaptation of the interpolation coefficients is supposed to improve the quality of interpolation at the extreme age groups. Therefore, we would like to compare the quality that can be achieved by the two formulas in this matter. The importance of such an element is increasing if we take into account the purpose of this work that requires more accuracy regarding to mortality data for the oldest ages. in this work, we are interested in an extrapolation based on single ages data rather than five ages data. The quality of the extrapolated data depends largely on the database as well as the control data. Again, We remind that most of the available data, as published by the ONS and completed by us (Flici, 2014) , is made is extended until the five age group [75 - 79] , and only the five most recent tables are up to [80 - 84] years. Here we would like to highlight the definition of extreme class arguing that it is not common for all of the series (1977 - 2014) and is varying from [74-79] to [80-84]. Treating a class as extreme or medium is not supposed to give the same results. Interpolate $_5q_{75}$ as a common extreme age group for the whole period (1977-2014) might ensures the homogeneity of the single ages interpolated mortality rates between 75 and 79 years old. The concern is how to break out $_5q_{80}$ given that the interpolated rates will be used to orient the the extrapolation process at advanced ages. The use of the coefficients of the extreme ages groups might leads to irregularities since $_5q_{75}$ too was broken-out by using the same coefficients. Figure 4 shows this irregularity with more evidence.

Figure 4: Interpolation of the death rates of the age group [70-75[by using medium and last age group coefficient and its impact on the old ages extrapolated rates



The figure on the left shows the difference which can be created when "Medium" or "Last" age group karup-king coefficient are used to interpolate the mortality rates of the age group [70-75[. The figure on the right shows the effect of the considered coefficients on the extrapolated mortality rates beyond the age of 80.

The curve (B) shows the death rates extrapolated with considering 75-79 as a medium class and 80-84 as extreme class. The curve (A) recounts the situation ignoring the existence of the 5q80, and by considering 5q75 as a common extreme classwith the intention of approaching the case of the post-2010 period. Lateral, the 5q80 quotient is interpolated

basing on the coefficient of extreme classes of the Karup-King formula. The figure in the left shows the fluctuations in the 75-79 age caused by the procedure (A).

Similarly, it would be impossible to interpolate the 5q75 on the basis of medium coefficients for the entire area (1977-2014). The figure in the right shows the effect of the adoption of the curve (A) compared to the results obtained with the curve (B). The effect is more important when the extrapolation is done on a more distant horizon. What is certain is that; this is the solid line (poly. B) that can reconstitute the initial five-year quotient 5q80. Dashed line (poly. A) generates an overestimated quotient.

Recall that our mortality surface is extended from 1977 to 2014 and life tables are almost extended until the age group [70 - 79]. To better orient the mortality extrapolation to older ages, we must take profit of 5q80 contained in the five recent life tables. Considering this additional data have help us to perform model selection and calibration in this sense. The concern is to ensure the homogeneity of the time series of the interpolated mortality rates of the last age category regarding to the resulted homogeneity on the extrapolated old ages mortality rates. For this, 75 - 79 mortality rates should be interpolated; for the whole period 1977-2014, using the same formula and the same coefficients. So how do we proceed? on one hand, using extreme class coefficients for all years will result to a kind of irregularity of the 75-79 interpolated mortality rates because this process implies to interpolate the 80-84 mortality rates for the years 2010-2014 with the same extreme class coefficients too. Regularity will be missed for years 2010-2014 since both 5q75 and 5q80 were broken out using the same extreme class coefficients. On the other hand, it is impossible to consider the class [75-79[as a medium class for the whole period [1977 - 2014] as in the absence of 5q80 for the period before 2010. The idea is to define a new interpolation coefficients which combine the characteristics of the medium and extreme class coefficients to be used for the whole period 1977-2014. On the basis of the observation of the interpolated rates between 75 and 79 obtained by using the medium coefficients on the data of the period 2010-2014, we define a new extreme class coefficients which lead to similar results when 5q80 are ignored. These coefficients will be applied to interpolate 75-79 mortality rates for the period 1977-2009. By this way, we ensure homogeneity of the interpolated 75-79 mortality rates for the whole period 1977-2014 while ensuring some regularity when passing from [75-79] to [80-85] interpolated mortality rates for the years 2010-2014. The regularity of this series is so important since we will use it to orient the old age mortality trend.

2.2.1 Interpolation of 5q75 by karup-king adjusted coefficients

In order to make the procedure of the 5q75 interpolation more clear and understood, we expose, in first, the Karup-King method based on 3 interpolation coefficients before spreading how to adjust the original formula to suit our case.

karup-King formula :

The Karup-King formula is designed to interpolate the number of deaths occurring between age x and age x + 5. The evolution of the survivors population at the exact age x will be subsequently interpolated. The interpolation tries to break out the five-ages mortality rate into a single ages ones which will allow to reconstruct the original five ages rates with an acceptable fidelity degree. . This relationship can be written as follows:

$$_5q_x = 1 - \prod_{s=0}^{4}(1 - q_{x+s})$$

Although this characteristic can be satisfied by different combinations (q_x, q_{x+1}, ..., q_{x+4}), However, the solutions which can ensure continuity and consistency of the evolution trend of the single ages mortality rates are limited. The interpolation methods assume that the interpolated rates grow following the same trend trend already observed on the five-ages rates. The structure of the single ages rates must join the trend footprint by five-ages rates of the age groups before and after the target age group.

For the first class, the Death of the interpolation formula is:

$$d_{x,x+i} = A_{1i}.D_{x,x+5} + A_{2i}.D_{x+5,x+10} + A_{3i}.D_{x+10,x+15} \; with \; i = 1, 2, .., 5$$

For medium classes:

$$d_{x,x+i} = B_{1i}.D_{x-5,x} + B_{2i}.D_{x,x+5} + B_{3i}.D_{x+5,x+10} \; with \; i = 1, 2, .., 5$$

For the ultimate age groups, the interpolation formula is as following:

$$d_{x,x+i} = C_{1i}.D_{x-10,x-5} + C_{2i}.D_{x-5,x} + C_{3i}.D_{x,x+5} \; with \; i = 1, 2, .., 5$$

The Coefficients A, B and C are presented in the table below:

Table 1: Karup-King interpolation coefficients

	first group N0			*Middle groupe Ni*			*Last groupe Nk*		
	N0	**N1**	**N2**	**Ni-1**	**Ni**	**Ni+1**	**Nk-2**	**Nk-1**	**Nk**
first fifth	0.344	-0.208	0.064	0.064	0.152	-0.016	-0.016	0.112	0.104
second fifth	0.248	-0.056	0.008	0.008	0.224	-0.032	-0.032	0.104	0.128
third fifth	0.176	0.048	-0.024	-0.024	0.248	-0.024	-0.024	0.048	0.176
fourth fifth	0.128	0.104	-0.032	-0.032	0.224	0.008	0.008	-0.056	0.248
last fifth	0.104	0.112	-0.016	-0.016	0.152	0.064	0.064	-0.208	0.344

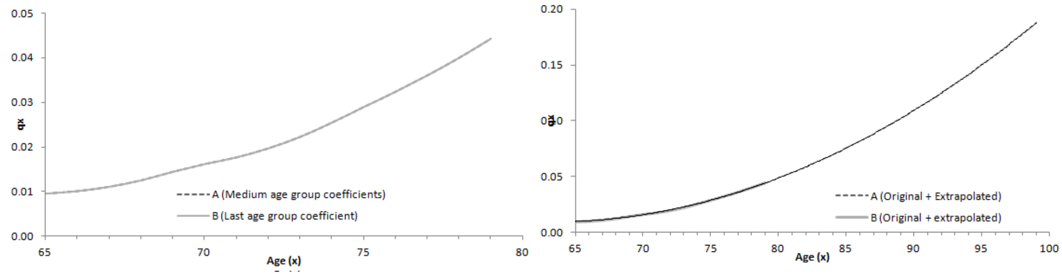
Our idea is to re-edit the coefficients used to interpolate 5q75 based on the formula of extreme classes giving the same results as those obtained by considering 5q75 as medium class (as of 2010-2014). For this, the redefinition of these coefficients will move towards the goal of achieving similar results on a set of 15 mortality tables: for males, females and both sexes for the period 2010-2014.

Table 2: Modified Karup-king last group interpolation coefficients

	Original coeficients				*Modified coeficients*			
	Nk-2	**Nk-1**	**Nk**	***Sum***	**Nk-2**	**Nk-1**	**Nk**	***Sum***
first fifth	-0.016	0.112	0.104	0.2	0.006	0.078	0.119	0.203
second fifth	-0.032	0.104	0.128	0.2	0.012	0.037	0.158	0.207
third fifth	-0.024	0.048	0.176	0.2	0.009	-0.003	0.199	0.205
fourth fifth	0.008	-0.056	0.248	0.2	-0.003	-0.039	0.240	0.198
last fifth	0.064	-0.208	0.344	0.2	-0.023	-0.074	0.284	0.187
Sum	0.000000	0.000000	1.000000	1	0.000317	-0.00057	1.000249	1

The results obtained are shown in the following graph:

Figure 5: The effect of karup-king interpolation coefficients correction



3 Old ages mortality extrapolation methods

Numerous studies have shown that mortality rates change its evolution trend beyond a certain age, usually the age of 80 or 85 years (Coale & Guo, 1989; Coale & Kikser, 1990) . According to this, the use of conventional mortality models may lead to results not reflecting the old ages mortality shape. In such a context, we use appropriate models adapted to the old ages specificity. Quashie et Denuit (2005) provide us with a fairly broad overview of the various methods designed to extrapolate mortality to the older ages.

We expose in this part the different extrapolation methods that appear to be applicable on our data.

3.1 Gompertz-Makeham model (GPZ, MKM)

For a long time, the classical mortality models have been used to extend mortality to the older ages. Gompetrz (1825) discovered that the force of mortality can be expressed as an exponential function with age :

$$\mu_x = \alpha.e^{\beta x}$$

In 1867, Makeham developed this model by adding a constant term representing the risk of death by accident which is not related to age:

$$\mu_x = c + \alpha.e^{\beta x}$$

Despite that the Gompertz model became a special case of Makeham model when $c = 0$, Makeham model has almost a better fitting quality compared to Gompertz model. Both these models were used to extend mortality beyond the age of 80. A Makeham-type function was used to extend the United Nations model life tables beyond the age of 80 (UN, 1981). The used function was expressed by the following formula:

$$\frac{{}_nq_x}{1 -_n q_x} = A + B^x$$

The sixth last estimated${}_nq_x$(x=50-75) were used to estimate the function parameters and to extend${}_nq_x$until the total extinction of survivals ($l_x = 0$).

The exposure of the accident death risk become neglectable, because old people are generally exempt from any risky activity (Pavrilov and Pavrilova, 2011). In such a case, Makeham model doesn't provide any added value compared to Gompertz model. Simply, both models lead to approximate goodness-of-fit and Gompertz model becomes preferable. For this, Gompertz model was used in 1983 to extend the Coale-Demeny model life table until the age of 100 (Coale and al, 1983).

When we extend mortality just until the age of 100, the choice of the extrapolation method is of a little importance. The shortage of old people mortality data prevented any comparison in this sense and the majority of attempts have been based on simple assumptions. Since the 80's, life expectancy has improved and the number of centenarians too (Robine and Vaupel, 2001; Buettner, 2002). The need for more sophisticated tools for old age mortality extrapolation became required. The improvement of the old people's mortality data quality permitted actuaries and demographers to evaluate the goodness-of-fit of their models and to propose adapted ones.

The availability of reliable old population mortality data allowed researchers to study the shape of the mortality curve at the older ages. Its main finding was that the old age mortality doesn't follow a Gompertzian function, but slows down slightly. According to this, Gompertz model over-estimate mortality at older ages. This phenomenon,well known as old age mortality deceleration, was observed on several populations and many of studies have tried to find a convincing explanation. A huge number of researchers assumed to relate this deceleration to a selection process concluding that when people reach a certain age, their death risk continue to grow-up but with decreasing rate (Kannisto and al., 1994; Kannisto, 1992; Coale and Guo, 1989). In their paper, Pavrilov and Pavrilova (2011) have shown that there is no old age mortality deceleration, thus the weakness of the quality of the old age mortality data and the exposure to the death risk might lead to such a consequence. In the absence of data allowing to verify these two hypothesis on the Algerian population, we keep supposing that the first assumption is more near to be verifiable.

In the same chain of ideas, Coale and Guo (1989), then Coale and Kisker (1990) achieved to formulate a model describing the old age mortality evolution. This method is still used nowadays to close-out the Coale-Demeny model life tables.

Many other models which were initially conceived to fit adult ages mortality have been tested on the old age mortality data as Perk's, Weibull and Heligman-Pollard models.

3.2 Weibull's model (WBL)

Weibull (1951) proposed the following formula:

$$\mu_x = \alpha . x^{\beta}$$

3.3 Helligman and Pollard model (HP)

Helligman-Pollard model (1980) is the only mortality model which allows to fit mortality at all ages including infant, child and young ages.

$$\frac{q_x}{1-q_x} = A^{(x+B)^c} + D.exp(-E(ln(x)-ln(F))^2 + G.H^x$$

Each one of the three terms of the model fit a mortality at certain ages (Helligmand and Pollard, 1980). At the age of 40 or 50, the two first terms become useless and only the third terms can be kept. then it can be written as :

$$Logit[q_x] = \pi + \varphi.x \;\; with \; \pi = ln(G) \; and \; \varphi = ln[H]$$

This last formula is similar to the Gompertz mortality law. The only difference consists on the used mortality measure $q_x$and μ_x.

This method was examined by the United Nations Population Division for old age mortality extrapolation issue (Buettner, 2002).

3.4 Coale-Guo and Coale-Kisker methods (CK)

The main idea of the Coale-Guo model (1989) is that, at older ages, mortality rates keep growing-up but with an increasing growth rate. This deceleration is supposed to follow a linear trend. Firstly, the model was expressed in 5 years age groups, and adapted after to the single ages description. The growth rate of the mortality from an age to another was defined by:

$$k_x = ln[\frac{{}_5m_x}{{}_5m_{x-5}}]$$

$k_x$is supposed to increase between two consecutive ages by a constant R:

$$k_{x+5} = k_x - R$$

For two ages x and $x+5.i$, we can write:

$$k_{x+5.i} = k_x - i.R$$

The Coale-Guo model was first applied to extend central death rate until the age of 110 starting from the age of 75. When $k_{80}$is known, for $x \geq 80$ and according to the previous formula, the estimation of the death rate at certain age starting from a known death rate at the previous age can be done by the following formula:

$${}_5m_{x+5} ={}_5 m_x.exp(k_{80} - \frac{(x-80)}{5}.R)$$

and as a general formula, we can write:

$${}_5m_{80+5.i} ={}_5 m_{75}.exp(k_{80} + k_{80} - R + k_{80} - 2R + k_{80} - 3R + ... + k_{80} - i.R)$$

Which implies:

$${}_5m_{80+5.i} ={}_5 m_{75}exp(ik_{80} - \tfrac{i(i+1)}{2}R) \;\; i = 1, 2, 3, ...$$

The objective of the method was to derivate the mortality rate at the age 80, 85, 90, 95, 100 and 105 from ${}_5m_{105}$.That leads to an age closure equal to 110 years. The choice of this value was justified by the fact that the observed highest age of death is around 110 (Coale and Kisker, 1987). The last mortality rate can be found by using:

$${}_5m_{105} ={}_5 m_{75}exp(6k_{80} - 15R) \;\; i$$

Authors have arbitrarily imposed as a constraint:

$$_5m_{105} -_5 m_{75} = 0.66$$

This formula is more adapted to populations with high life expectation (Coale and Guo, 1989), that means 70 years or higher.

Coale and Kisker (1990) reformulated Coale-Guo formula and adapted it for the case of single ages mortality data. Starting from the same principal of mortality deceleration rate with constant decreasing, they achieved the following formula based on the central death rate :

$$\hat{u}_x = \hat{u}_{x-1} exp(k_{80} + s.(x-80)) \; x = 80, 81, ..., 109.$$

k_{80}**represents the growth rate of mortality at 80 years. It is defined to be the average growth rate of the age range [65-80] and can be calculated by the following relationship:**

$$k_{80} = \frac{ln(\frac{\hat{u}_{80}}{\hat{u}_{65}})}{15}$$

In some cases, the use of the Coale-Kisker formula can lead to some incoherence or to a crossover of the males and females extrapolated mortality curves. To fix this element, the authors fixed, arbitrary the rates for a relatively high age (110 years):

$$\hat{u}_{110} = \begin{cases} 1 \; for \; males \\ 0.8 \; for \; females \end{cases}$$

This leads to define S **which is equal:**

$$s = -\frac{ln(\frac{\hat{u}_{79}}{\hat{u}_{110}}) + 31k_{80}}{465}$$

s **in Coale-Kikser model is the same** R **used in Coale-Guo formula. They are calculated from the same formula.**

Finally, both of these methods are well known as Coale-Kisker model which is also called Quadratic model (Roli, 2008. Thatcher, 1999). If we start from this definition, we can understand that Coale-Kisker model consists simply to express the logarithm of death rate as a quadratic function. However, this formulation is not used or explained in the literature:

$$ln(\mu_x) = a + bx + cx^2$$

The mortality growth rate can be expressed as:

$$k_x = ln[\frac{\mu_x}{\mu_{x-1}}] = a + bx + cx^2 - a - b(x-1) - c(x-1)^2 = c(2x-1)$$

Considering a starting age of 80, we can write:

$$k_x - k_{80} = 2.c.(x-80)$$

and that leads to the same formula seen earlier :

$$\hat{u}_x = \hat{u}_{x-1} exp(k_{80} + 2.c.(x-80)) \; x = 80, 81, ..., 109.$$

Other methods have been developed later for the same issue.

3.5 Himes, Preston and Condran (HPC)

Himes, Preston and Cardan (1994) proposed an extrapolation method based on the observation of the single ages death rates at the age range 45-99. Their idea is to smooth the observed mortality curve by an 2-steps 3rd order moving average on the form:

$$g_x = \frac{\mu_{x-1}+\mu_x+\mu_{x+1}}{3}$$

Then, the smoothed death rate g_x is smoothed again by using the formula:

$$h_x = \frac{g_{x-1}+g_x+g_{x+1}}{3}$$

that leads to the following final formulation of the smoothed death rate:

$$h_x = \frac{\mu_{x-2}+2.\mu_{x-1}+3.\mu_x+2.\mu_{x+1}+\mu_{x+2}}{3}$$

For the considered age range, the death rates follow an exponential function which can be approximated to a Gompetz's model or to the third terms of Helligman and Pollard model. The authors proposed to extend mortality until the age of 115 years old by using the following formula:

$$Logit(\mu_x) = ln(\frac{\mu_x}{1-\mu_x}) = \theta + \Pi.\text{x}$$

The parameters θ and Π are estimated on the age ranges [45-79]. In the original application, many age ranges were compared : [45-79], [50-79], [55-79], [60-79] and [65-79]. The comparison of the obtained results has shown that the estimated parameters vary significantly in function of the age range used as a basis for extrapolation.

3.6 Perks, Logistic, Kannisto, Thatcher models (PRK, LOG, KST, THT)

The use of logistic functions for mortality curves graduation was first introduced by Perks (1932). He proposed the following formula to fit the mortality hazard rate with age:

$$\mu_x = \frac{\alpha.e^{\beta x}}{1+\gamma.e^{\beta x}}$$

The general form of logistic model can be written like :

$$\mu_x = \theta + \frac{\lambda.\alpha.e^{\beta x}}{1+\alpha.e^{\beta x}}$$

When Thatcher et al. (1998) tried to use this model which was used to fit old age mortality of 13 developed countries, they found that λ is very close to 1 (Thatcher, 1999). They concluded that the model can be simply written with 3 independent parameters:

$$\mu_x = \theta + \frac{\alpha.e^{\beta x}}{1+\alpha.e^{\beta x}}$$

at older age, θ become negligible as the constant accident risk in Makeham's model. Since that, the previous formula can be simplified to be (Kannisto, 1992):

$$\mu_x \approx \frac{\alpha.e^{\beta x}}{1+\alpha.e^{\beta x}}$$

in the Logit form, we found:

$$Logit(\mu_x) = ln(\alpha) + \beta.x$$

That is the same formula used independently in 1994 by Himes, Preston and Cardan. When kannisto proposed such a simplification, he was just reporting some findings without any intention to give a specific mortality model. This model was reported in (Thatcher et al., 1998) as a Kannisto model.

3.7 Denuit et Goderniaux method (DG)

This method relies on a polynomial formula of order 3 of log mortality rate:

$$ln(q_x) = a + bx + cx^2$$

In the original article, the authors have set an assumptions about the ultimate survival age which might be equal to 130 years old. In the order to respect this constraint, they imposed $q_{130} = 1$and $P_{130} = 0$.

3.8 Comparison

According to the old age mortality models presented above, there are different ways to extend mortality to the older ages. Globally, 2 mortality measures can be used: the death rate $\mu_x$and the mortality rate q_x. Pavrilov and Pavrilova (2011) explained the difference between the use of these parameters, and detailed how the choice of the indicator can affect the final result of the extrapolation process. At the older age, the mortality rate reach a high level and keep growing slowly until the value of 1 which represents the limit, while death rates can keep growing without a such a constraint. For this, $\mu_x$is more suitable for the fitting and the extrapolation use especially at the older ages. When we report to the semi-log scale, this last can easily be approximated by a straight line more than q_x. Another way to give more regular linear trend to the mortality age-specific indicators is the introduction of $Logit$ which allows passing from [0-1] to]-∞, $+\infty$[. The third element, is that there are principally two ways to extend mortality to the older ages : a transformed linear trend and transformed quadratic trend.

Among the presented models, there are only two that propose an extrapolation based on q_x: HP and DG models. The first propose an extrapolation with a linear trend of $ln(q_x)$, while the second aims to impose a quadratic form to the $Logit(q_x)$. The other models are all based on μ_x. GPZ, MKM, WBL and HPC models express the death rate in log linear form. CK model is oriented to extend $ln(\mu_x)$ with a quadratic function, while logistic models (PRK, LOG, KST, THR models) try to approximate $Logit(\mu_x)$ to a linear function. In addition, two families of models can be distinguished in function of either any closure constraint is imposed or not. GPZ, MKM, WBL, LOG, KST, THT, PRK, HP and HPC are supposed to be a behavioural extrapolative models since there is the extrapolated rates at older ages are just a result of the model calibration at younger ages. With such models, we can not impose any closure constraint. The life table is closed-out when q_x is equal to 1. For the other models (CK, DG), which are based on a quadratic

functions, although the old age mortality are extrapolated with a parametric models, the final result is significantly affected by the imposed closure constraint. In our application, we will try to compare different variants of models. for this, the models to be compared are : HP, DG, GPZ, WBL, CK and KST.

4 Old age mortality extrapolation results

4.1 Some methodological orientations

To compare mortality models, we use conventionally a set of quantitative and qualitative criteria. The principal ones are the goodness-of-fit and the predictive capacity. Goodness-of-fit aims to evaluate the fidelity of the model of the data used for its calibration. A comparison between observed (or original) and fitted values is necessary in such a case. The selection process must lead us to select a model which have more capacity to predict nearly the real trend of the extrapolated variable beyond the last age for which data is available. It is well known among statisticians that a good fitting quality doesn't guarantee necessarily a good predictive capacity. In order to make a sense for this evaluation, we need some data to confront the values extrapolated by the model against the observed ones. we remind that according to the data which we have until now, the single age death probabilities are extended in almost case until the age of 79. Only for the 5 last years, these mortality rates were extended until the age of 84. Here, we have to arrange this data in order to allow a good evaluation of the goodness-of-fit and the predictive capacity of the selected model. Using a lower ages to calibrate the models can lead to a good fitting results in spite of its being useless for old age mortality extrapolation issue. In other words, the more the upper limit of the age range used to calibrate the model is highest (near to the extreme available specific-age-mortality rate), the more the calibrated model is suitable for extrapolation issue. Moreover, to evaluate the predictive capacity of a model, some data (ages) must be used for comparison. The More the used age range is longer,the more the evaluation takes a sense. However, given the shortage of the data for age beyond 75 and 80 (for the 5 last years) we must better manage and combine the available data for both issues : goodness-of-fit and and predictive capacity. According to all these elements, we judged the most suitable to use is the age of 74 as a maximum age for model calibration and the age range [75-79] to evaluate the predictive capacity of all models. Since the mortality rates for the age range [80-84] are available but only for 5 years, it will be advantageous to consider it in the predictive capacity evaluation. To avoid the problem of the weight generated by the number of years with available data for each age range, we can use a weighted evaluation criterion which takes the errors observed at [75-79] (38 years) in addition to those observed at [80-84] (5 years) by an equivalent weight. It is clear that the use of data until 84 years does not allow to predict the exact old age mortality trend, but it allows at least to predict the starting-up of this trend. For this, it is necessary to exploit all the available data.

As a fitting and evaluating criteria, the Mean Squared Error (MSE) seems to be suitable both with our data and objective. We remind that the six first selected models are based on different significant quantities: $ln(\mu_x)$, $logit(\mu_x)$ and $logit(q_x)$. To ensure the comparability of the models for the fitting quality and the predictive capacity evaluations, the MSE is used for all models. It must be expressed in a unified mortality measure that may be $ln(q_x)$ which is used only for comparison between models. The goodness-of-fit can be evaluated on the age range [x-74] by:

$$MSE_{[x-74]} = \frac{1}{t*x}\sum_{x=X}^{74}\sum_{t=1977}^{2014}[ln(\hat{q}_{xt}) - ln(q_{xt})]^2$$

For the predictive capacity evaluation, this criteria is estimated in two steps: In first, we calculate the MSE corresponding for each age range [75-79] and [80-84] by the respective corresponding formula:

$$MSE_{[75-79]} = \frac{1}{5*38}\sum_{x=75}^{79}\sum_{t=1977}^{2014}[ln(\hat{q}_{xt}) - ln(q_{xt})]^2$$

and

$$MSE_{[80-84]} = \frac{1}{5*5}\sum_{x=80}^{84}\sum_{t=2010}^{2014}[ln(\hat{q}_{xt}) - ln(q_{xt})]^2$$

Then, the weighted MSE corresponding to all the age range [75-84] is simply calculated by :

$$WMSE_{[75-84]} = \frac{1}{2}(MSE_{[75-79]} + MSE_{[80-84]})$$

In first, the six models will be extrapolated to the ages beyond 75 without imposing any closure constraint. We remind that when some extrapolation models were first proposed, they were based on the age of 80 years as the age to start extrapolation. Here, all the used models are adapted to give an extrapolation starting from the age of 75. The extrapolation process will be applied independently for males, females and both sexes population.

In addition to the evaluation of goodness-of-fit and the predictive capacity, some other qualities are required to qualify any model to be a good one. The extrapolation's results must be coherent regarding to some classical finding in the mortality modelling field that we will present as following:

1. Mortality rates keep increasing with age. That allows to write :$q_{x+1} > q_x$; $\forall x \geq 75$. The mortality deceleration which can be observed starting from a certain age (Coale and Guo, 1989; Coale and Kisker, 1990) does not imply a decrease in mortality rates with age but a decrease of the growth rate. We introduce this element, which appears to be evident, because of polynomial functions (DG and CK methods for example) can lead under certain conditions to a reversal in the mortality rates beyond a certain age. This propriety will be only checked-out for DG and CK models.

2. Since extrapolation is made independently for males, females and both sexes populations, the extrapolated mortality rates for both sexes population should be closer to the weighted average of male and female rates. According to the observation of the structure of the Algerian population aged 70 and +, we can observe that males and females have approximately the same weight.

That allows to write : for a fixed year (t) : $q_x^{both} \approx \frac{(q_x^m + q_x^f)}{2}; \forall x \geq 75$. In the case of a change in these sex distribution at very advanced ages in any way, the mortality rate for both sexes population q_x^{both} must be situated between the male and female rates, and can not be outside in any case . That implies : $q_x^b \in [q_x^f - q_x^m]$;

3. Male mortality rates are higher than female mortality rates:$q_x^m \geq q_x^f, \forall x \geq 75$. This constraint is checked only for the years which do not have a female over mortality for ages below 75 years. Since the extrapolated old age mortality is widely determined by the trends observed at the prior ages, an observed female over mortality at adult ages will lead to a similar effect at old ages.

4. The last element which can give a clear judgement about the goodness of the extrapolation is the age limit predicted by the model. Theoretically, the age limit (w) is reached when the mortality rate attains the value 1: $q_w = 1$. An old age mortality model can be described as a good extrapolative model when it leads to an age-limit coherent compared to the maximum surviving age observed in reality. For the Algerian population, we remind that, according to the MICS IV (Multi Indicators Clustery Survey) results, a maximal surviving age of 112 years old was observed for females and 110 for males (Flici et Hammouda, 2015). That does not represent an estimation of the maximal surviving age for the Algerian population, but just a minimal of the interval where the ultimate age can be situated.

4.2 Model selection

4.2.1 Goodness-of-fit and predictive capacity

In almost cases, goodness-of-fit is widely related to the age range used to calibrate a model. Usually, the quality is higher as the length is shorter, but at the expenses of the robustness of the model. When a model is estimated on long length data, the variation bands of the predicted series are narrow. So, it is so important to define a fitting criteria which combine goodness of fit and robustness. The use of the Bayesian Information Criterion (BIC) is more suitable for such purpose since it considers in addition to the distance between fitted and original values, the number of parameters and the number of observations. Although the BIC was firstly proposed for Likelihood Estimations issues (Schwarz, 1978), the formula was adapted later to suit the Least Squares Errors (LSE) estimation method (Burnham and Anderson, 1998; Hansen, 2007). The adapted formula can be written as:

$$BIC = n.ln(\frac{1}{n}SSE) + k.ln(n)$$

with n representing the number of observations, and k the number of parameters in the model.

BIC is used to evaluate both the goodness-of-fit and predictive capacity of the models. For each model, these qualities are related to age range used to calibrate the model, and it is not evident to define a common age range which ensure the best quality for all models. Also, there is practically no "best model" in all situations. It is very important to reduce the initial set of candidate models to a sub-set leading to approximate quality. According to this, each model is calibrated on various age ranges : [40-74], [45-74], [50-74], [55-74] and [60-74] and then extrapolated beyond the age 75. The models are ranked according the two criteria. Results are presented in the table bellow:

When we try to compare a set of models with different age ranges, it is not evident to find a model which is the best with any age range. Similarly, we can not find a specific age range which guarantee the best result for every model. That became more

Table 3: old age mortality models evaluation and ranking

Age range for calibration	40-74		45-74		50-74		55-74		60-74		Models Ranking	
Age range for evaluation	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	Goodness of fit	Predictive capacity
Model												
DG	-173.44	*-50.68*	-163.82	*-50.80*	-137.50	*-53.83*	-107.75	*-45.01*	-81.20	*-46.14*	2	1
HP	-173.02	*-29.85*	-165.40	*-33.75*	-138.12	*-35.92*	-109.86	*-37.36*	-89.85	*-36.63*	4	4
GPZ	-174.04	*-31.39*	-165.92	*-35.51*	-138.49	*-37.75*	-110.06	*-39.06*	-89.98	*-37.80*	3	3
WBL	-166.11	*-16.02*	-141.67	*-16.02*	-125.87	*-20.25*	-105.19	*-23.93*	-86.76	*-26.98*	6	6
CK	-178.28	*-46.95*	-157.36	*-45.61*	-131.87	*-45.91*	-102.75	*-46.90*	-75.63	*-48.42*	1	2
KST	-171.95	*-28.73*	-164.81	*-32.11*	-137.72	*-34.23*	-109.63	*-35.77*	-89.70	*-35.47*	5	5
Females												
Age range for calibration	40-74		45-74		50-74		55-74		60-74		Models Ranking	
Age range for evaluation	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	Goodness of fit	Predictive capacity
Model												
DG	-152.12	*-24.75*	-139.67	*-23.93*	-114.48	*-32.64*	-95.36	*-29.48*	-82.05	*-31.07*	2	2
HP	-144.06	*-13.61*	-140.44	*-16.52*	-121.59	*-20.07*	-97.99	*-23.05*	-84.36	*-26.29*	4	4
GPZ	-144.81	*-14.15*	-141.07	*-17.13*	-121.98	*-20.76*	-98.20	*-23.83*	-84.44	*-27.01*	3	3
WBL	-115.92	*-3.48*	-119.29	*-7.02*	-110.25	*-11.04*	-93.31	*-15.09*	-81.85	*-19.41*	6	6
CK	-152.28	*-22.83*	-139.06	*-22.13*	-115.29	*-33.39*	-94.99	*-27.90*	-69.77	*-27.62*	1	1
KST	-143.31	*-13.05*	-139.78	*-15.97*	-121.19	*-19.34*	-97.77	*-22.35*	-84.28	*-25.58*	5	5
Both sexes												
Age range for calibration	40-74		45-74		50-74		55-74		60-74		Models Ranking	
Age range for evaluation	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	[x-74]	[75-84]	Goodness of fit	Predictive capacity
Model												
DG	-169.38	*-36.08*	-150.69	*-35.17*	-124.04	*-35.67*	-96.69	*-36.69*	-71.77	*-37.53*	1	2
HP	-160.24	*-20.57*	-155.85	*-23.87*	-132.04	*-26.93*	-105.61	*-29.41*	-90.70	*-31.24*	4	4
GPZ	-161.23	*-21.51*	-156.64	*-24.90*	-132.54	*-28.05*	-105.89	*-30.53*	-90.85	*-32.21*	3	3
WBL	-126.52	*-7.25*	-130.85	*-11.24*	-118.59	*-15.25*	-99.95	*-19.18*	-87.13	*-23.00*	6	6
CK	-165.80	*-30.63*	-161.33	*-38.73*	-133.65	*-40.35*	-103.78	*-40.48*	-76.63	*-39.99*	2	1
KST	-159.24	*-19.71*	-155.02	*-22.96*	-131.50	*-25.92*	-105.31	*-28.33*	-90.53	*-30.30*	5	5

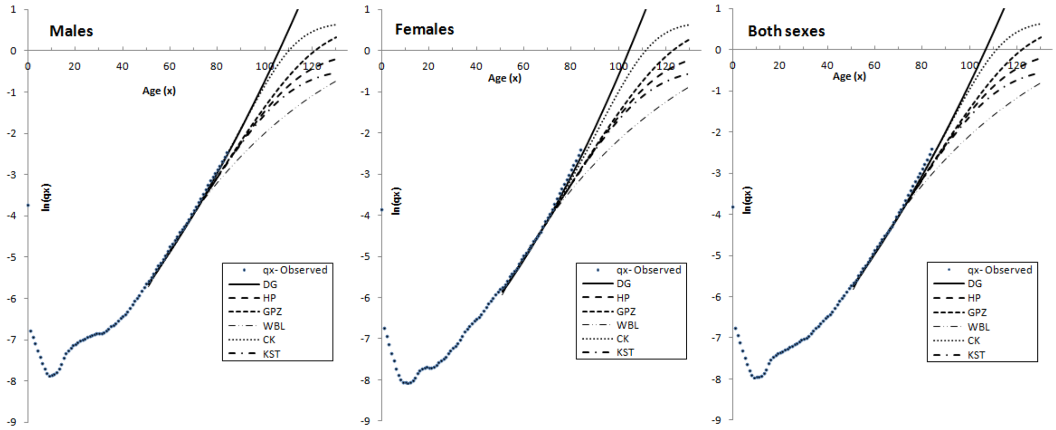
Every model is calibrated on 5 different age ranges. BIC is calculated for [x-74] and [75-84] as an indicators of goodness of fit and the predictive capacity. The best age range which leads to the best models regarding to the GOF is marked in grey. The best age range for each model regarding to PC is marked in blue.

complex when we try to combine goodness-of-fit and predictive capacity. But from the observation of the results given in the tables behind, we can conclude some evidence about the best models and the best age ranges according to the goodness of fit and the predictive capacity criteria.

According to the ranking of the six models and to the two selected criteria obtained on the three populations, we observe some differences in ranking which changes in function of the criteria and of the age range but some evidences appear clearly: The three best models for fitting and predicting are DG and CK in the first and the second positions followed by GPZ in the third position. HP, GPZ and WBL came in 4th, 5th and 6th positions. we observe that, every model leads to its best quality in a specific age range. Nevertheless, it is necessary to select a common age range for all models to reduce irregularities which can appear at very old ages. Here, we propose to use the age range [50-74] to fit all models. This age range is the only range which allows to keep the final ranking (Goodness-of-fit and predictive capacity) obtained on the three populations.

The figure bellow shows the mortality rates extrapolated to old ages with the six models:

Figure 6: Extrapolation models comparison

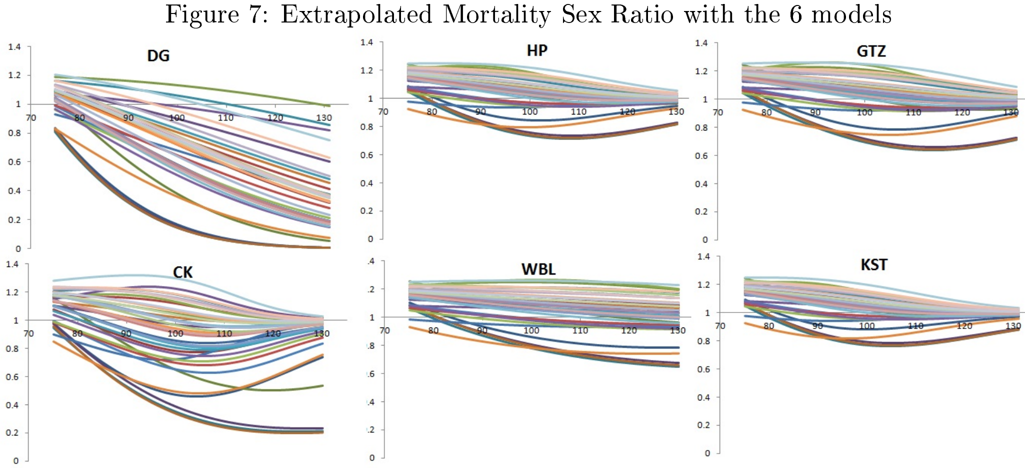


We observe that quadratic models (DG and CK) provide a near qualities regarding to the goodness-of-fit and the predictive capacity compared with the linear transformed model. An evaluation based only on the goodness of fit and the predictive capacity criteria is not sufficient to select the best models. The complementary criteria that we will use for the evaluation and comparison issue is the coherence between male and female expected mortality and also the coherence between single sexes and both sexes extrapolations. Also, the expected surviving age limit is used to evaluate the extrapolation results.

4.2.2 Expected males females sex ratio

The observation of the male female sex ratio calculated on the extrapolated mortality rates beyond the age of 75 for the period 1977-2014 are shown separately for each of the six compared models. Results are presented in the figure below:

Figure 7: Extrapolated Mortality Sex Ratio with the 6 models

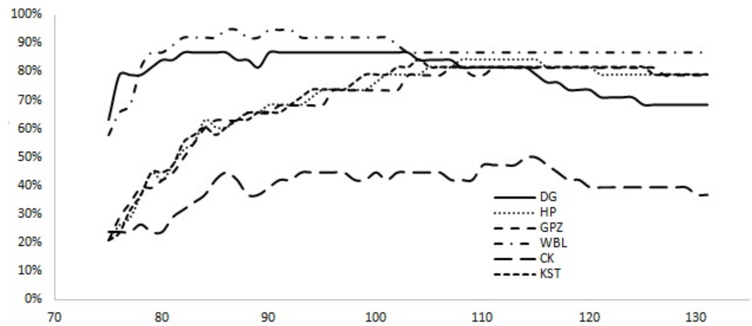


At old ages and very old ages, when mortality rates grow to the value 1, male and female mortality rates tend to converge to each other. In such a case, the male female sex ratio must converge in the 1. KST, HP and GPZ have led to more formal results compared to the three other models.

4.2.3 Coherence between unisex and both-sexes expected mortality rates

To evaluate the quality of each model in terms of coherence between single sexes and both-sexes mortality estimations, we calculate the part of cases where both-sexes expected mortality rate is situated out of the interval between males and females mortality rates.The more this failure rate is lower, the more the extrapolation is coherent. results are presented in the figure bellow:

Figure 8: Failure rate between unisex and both-sexes expected mortality rates

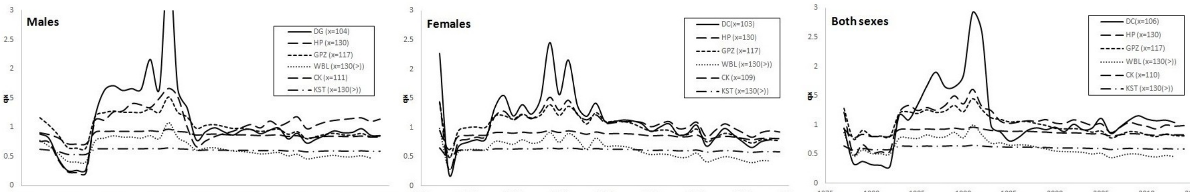


The DG and WBL models led to the most failed result when we compare single sexes and both-sexes expected mortality rates. The other models give an acceptable failure rate under 20% increasing beyond the age of 100 around 70%. KST model give the less important failure ratio.

4.2.4 Expected age limit

The surviving age limit is attained when the death's probability is equal to 1. One way to evaluate the coherence of the old age extrapolated mortality rates is to compare the different models according to their expected age limit. Surely, the expected age limit obtained with a given model can not be constant over time, but slightly increasing with time. To compare the six models presented above, we will try to present the expected mortality rate at a high age which should be near to the age limit predicted by each model. The obtained result on the period 1977-2014 is shown in the figure bellow:

Figure 9: comparison of some high ages q_x predicted by the models



Other than the turbulence that mark the evolution of the mortality rates (nearly) at the age limit, some models lead to incoherent age limit. By the age of 130, HP, WBL and KST models give a mortality rate under 1. The other models lead to more coherent results. The age limit expected with DG model is equal to 104, 103 and 106 in average respectively for males, females and both sexes populations. 106 is the age limit expected for the year 2014 for the three populations. For CK model, the expected age limit is situated between 109 and 111 years old. In adverse, GPZ model leads to higher age limit equal 117 in average and between 121 and 122 years for 2014.

According to the maximum surviving age observed for the Algerian population until now, GPZ model seems to be the most coherent model according to the expected age limit. Resulting from the MICS survey, the maximum surviving age was equal to 112 for females and 110 for males (Flici and Hammouda, 2015). This survey is far to provide an accurate estimation of the age limit of Algerian population, since it is not exhaustive compared with the civil stats registration data, but it allows to fix the lowest bound of the age limit estimation. Also, we must consider the future evolution of the age limit since longevity keeps increasing. The age limit is defined to be the age which can not be surpassed by any human being in a certain geographic area. Denuit and Goderniaux (2005) supposed this age to be equal to 130 for the developed countries. In the same chain of ideas, we can suppose that this age limit can be lower in developing countries and we can consider 120 as a reasonable limit for the Algerian population.

4.2.5 Results discussion

We have seen through this comparison, that quadratic models gave a good quality regarding to goodness-of-fit and predictive capacity compared with the linear transformed models. GPZ model makes exception of this rule. The study of the expected age limit predicted by the different models confirmed this finding. The comparison based on other criteria has shown a different judgement. The models ranking overturned in almost case, when we consider the coherence between single sexes and both-sexes estimation and also the sex differential expected mortality. GTZ model keep a good score in all the situations.

The less performance resulted with the quadratic models (CK and DG) can be explained by the nature of the polynomial functions which can lead to some plausible values when it is used to extrapolate or to extend mortality rates for a huge age ranges. Although they guarantee a better goodness of fit and predictive capacity compared with the linear transformed models, they generate some incoherence regarding the sex differential or single sexes vs both-sexes extrapolated mortality rates. To avoid this disadvantage, these models were usually used under some constraints about the age limit (Denuit and Goderniaux, 2005) or about a fixed mortality level at certain high age (Coale and Kisker, 1990). Imposing such a construct may avoid a crossover of the extrapolated mortality curves considered in the time dimension as in the sex differential one (Buettener, 2002).

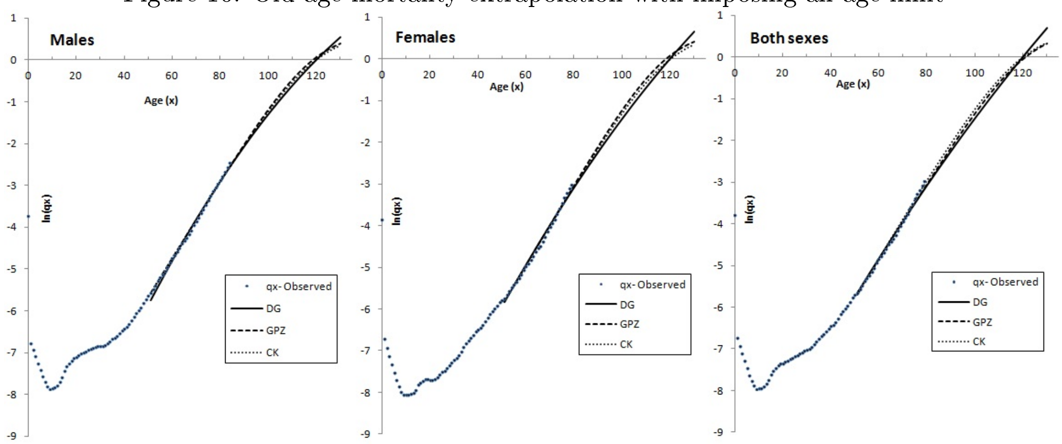
In the rest of the present work, we keep working with 3 models: DG, CK model with 120 as an assumed age limit, and GPZ model as the best linear transformed model.

4.3 Model calibration under age limit assumption

Until now, we have seen that the quadratic models (DG, CK) and GPZ model are near to describe the old age mortality of the Algerian population than the other models. How-

ever, some weaknesses were observed. The trend of mortality rates beyond 80 obtained with the models are almost incoherent especially those obtained with the quadratic models. The introduction of an age limit assumption should reduce the observed incoherences (Buettner, 2002). For this, DG and CK models are re-estimated by imposing 120 years as an age limit. This additive condition have to impair the fitting quality of these models. To recover such a lack in fitting quality and to keep the 2 quadratic models in the same fitting performance as GPZ model, we reduce the length of the age range used to calibrate the two models. we observe that when DG and CK are calibrated on the age range 60-79, they give approximately the same fitting quality as GPZ model calibrated on the age range [50-79]. The evaluation of the three models after imposing the new constraints are given in the figure bellow:

Figure 10: Old age mortality extrapolation with imposing an age limit



We observe that the three models lead approximately to the same results regarding to the extrapolated trend of mortality rates beyond 80 until 120 years. To confirm the goodness of these results, we must check out their coherence, principally regarding the male vs female and single sexes vs both sexes differential mortality.

Figure 11: Male Female Mortality ratio under age limit constraint

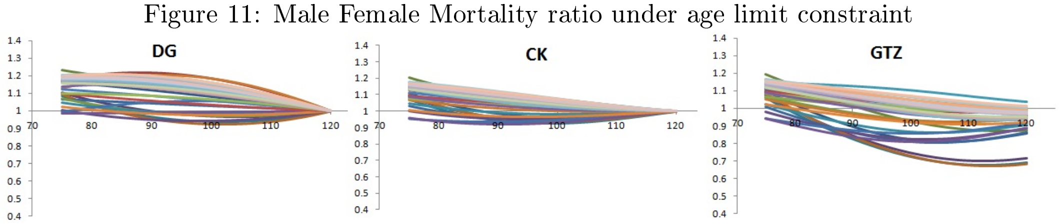
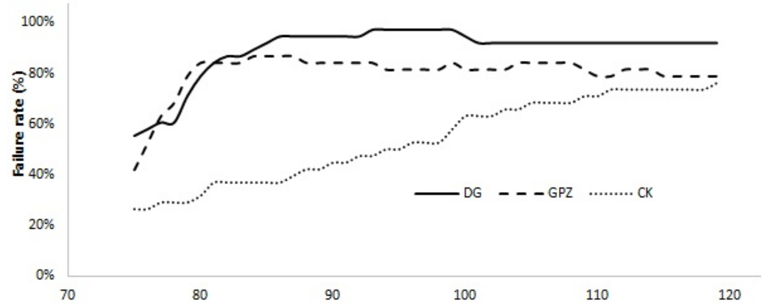


Figure 12: Unisex vs Both sexes extrapolation under age limit constraint



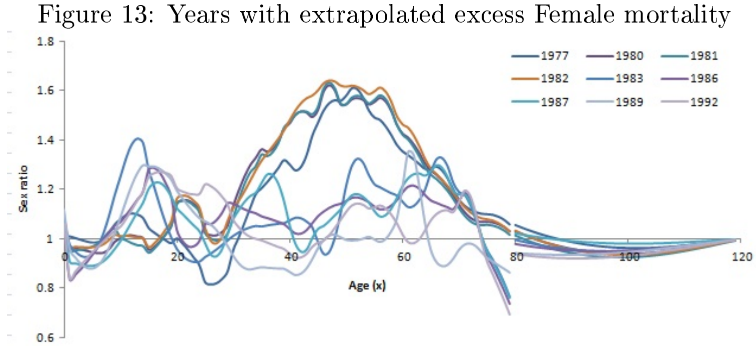
According to these two last criteria, CK seems to be better than DG and GPZ. The expected male females mortality ratio converge to the value 1 more rapidly in CK model than with DG model. For the coherence between singles sexes and both sexes projection, the increasing trend of the failure ratio can be explained by the fact that at near age limit, all mortality rates converge to the unit value. Furthermore, the difference between males, females and both sexes mortality rates become as smaller as the age limit approach. However, it is necessary to improve the quality of the extrapolation again by imposing some additive constraints about the comparative evolution of males, females and both sexes extrapolated rates.

4.4 Coherence constraints

Here, we impose two additional constraints; The first one aims to keep the females mortality under the males mortality while the second one will try to keep the both sexes expected mortality in between males and females mortality.

For some years of the period before 1994, we could not obtain a sex ratio greater than 1 for the extrapolation age range [80-120]. These years have marked a slight females mortality excess during the age range used for the extrapolation. Also, a heavy decrease in sex ratio is observed by the late of the age range used for extrapolation. For the period post 1994, males mortality keeps increasing over the females mortality until the age limit.

Figure 13: Years with extrapolated excess Female mortality



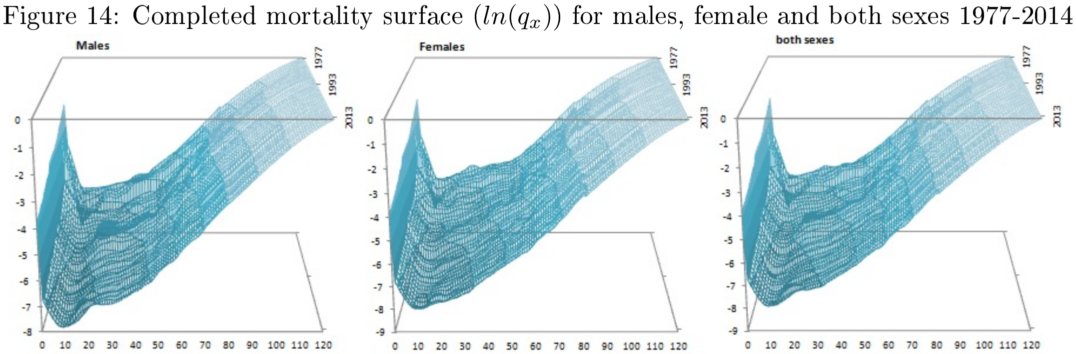
For the second constraint about the coherence between both sexes and single sexes extrapolations, we imposed that the both sexes extrapolated mortality rates vary in between male and female extrapolated mortality rates which were estimated in the previous part. This constraint was fully respected without affecting significantly the goodness of fit.

5 Final results

After imposing all the necessary constraints on the extrapolation process, we can conclude that the three selected models: GPZ, CK and CG give approximately the same quality regarding the considered evaluation criteria. Therefore, it is very difficult to decide whether one model is more appropriate compared to the others. The quadratic models allow more flexibility in old age mortality extrapolation, beside they ensure a good fitting quality on the age range used as a basis for the extrapolation. When the age limit constraint is imposed, the extended mortality surface shows a high regularity. GPZ model have shown a good performance regarding all selection criteria. However, the use of unfitted adult age mortality surface for its calibration, have led to unstable expected age limit series. This disadvantage does not appear in the case of quadratic models by imposing a common age limit constraints for all years. Consequently, GPZ model can be a perfect one to extrapolate old age mortality on a well fitted mortality surface. As a result and giving that CK model provide better quality compared to DG model on the basis of the other selection criteria, the decision is to adopt the CK model to extrapolate old age mortality for the Algerian mortality surface.

5.1 Complete mortality surface

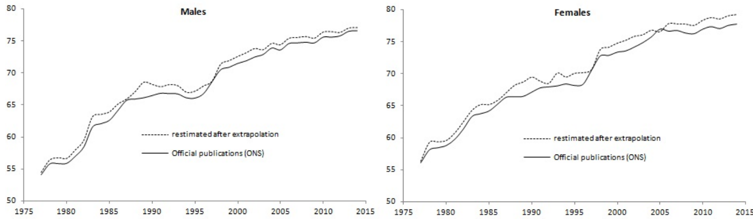
Figure 14: Completed mortality surface ($ln(q_x)$) for males, female and both sexes 1977-2014



5.2 Corrected life expectancy at birth:

As we have already pointed out, the life expectancy at birth is largely related to how the life tables are closed out. We will illustrate the impact of the closure of the life tables on the evolution of life expectancy at different ages. Life expectancy following the conventional notation, is calculated from a five age mortality patterns. In actuarial calculations, a single ages pattern can be used and that leads to a slight difference. To avoid any methodological effect when comparing the official life expectancy and thus re-estimated according to the extrapolation results obtained in the present work, we use five ages mortality rates to re-estimate life expectancy.

Figure 15: Re-estimated life expectancy after extrapolation compared to the official statistic



The first finding we an carry out from figure 15 is that the life expectancy as we estimated it by the present work is slightly greater than the values included in ONS official publications. The average gap in this sense is around 0.9 year for males and 1.2 year for females. As mentioned in Flici and Hammouda (2014), some change points observed on the series of life expectancy at birth find an explanation (or confirmation) in the results obtained here. In 2005 for example, the change point observed on the female curve is significantly due to the closing-out procedure.

6 Conclusion

We have seen along this paper why it is important to close-out life tables on the basis of mortality rates extrapolation rather than use model life tables as an external reference. The first advantage of the first approach is to allow extend the age mortality pattern beyond the usual closure age and as near as possible to the surviving age limit. These details is generally needed in actuarial calculations and population forecasts. Because of the information lack and unreliability, such detail is usually not included in official life tables which are closed out earlier. Only the residual life expectancy at the closure age is published to summarize the mortality age pattern for the ages beyond. In adverse, the quality of the estimations resulted from the use of the model life tables to close-out national life tables is very related to the adequate model life table selection process. Since that, a wrong use of model life tables may lead to unrealistic estimated life expectancy at the closure age. Also, an eventual change in the selection methodology may lead to an apparent level change in the series of the life expectancy at birth. We have seen in the introduction of this work some evidences concerning the wrong use of model life tables in the case of some African countries (Ita and Banjan, 1984).

We have shown also in a previous work (Flici and Hammouda, 2014) that some changes in the life expectancy evolution series in Algeria during the past half century are purely due to methodological effects. In this sense, the closing-out methodology led to some imperfections and irregularities regarding the resulted life expectancy at birth. For this, we proposed in the present work to try an other approach to close-out the Algerian official life tables ensuring more accuracy, adequacy and regularity. Our approach was to estimate the old age mortality by the extrapolation of the observed mortality trend at adult age by using appropriate models. A set of old age mortality models were presented and compared for this issue : Gompertz (1825), Weibull (1952), Kannisto (1992), Helligman and Pollard (1980), Coale and Kisker (1989) and Denuit and Goderniaux (2005). The model evaluation and selection were based on a set of criteria : goodness of fit, predictive capacity, predicted age limit, coherence between male and female mortality ratio and also

the coherence between single sexes and both sexes mortality extrapolations. After a first comparison, three models were evaluated as suitable to our data : CK, DG and GPZ. To improve the quality of the resulted extrapolation under the three models, we modified the age range used for model calibration and imposed the age of 120 years old as an age limit constraint for the quadratic models (CK and DG). In addition to this, some other constraints were imposed to ensure coherence in males vs females, single vs both sexes extrapolations. In final, we concluded that the three models lead approximately to a similar results. Because of the irregularities in the resulted series of the expected age limit, the GPZ model was excluded from the comparison. Among the two quadratic models, CK model marked better quality than DG model regarding to the male vs female and single vs both sexes extrapolations. Once the old age mortality rates were extended until 120 with the CK model, life expectancy at birth was re-estimated. The comparison of the evolution of the resulted series of life expectancy at birth with the national statistics showed that our method leads globally to a gain of about 0.5 years in average on the whole period 1977-2014. Also, the obtained series shows more regularity in terms of the evolution trends.

In final, we would like to highlight how it is so important to close out the Algerian life table by extrapolating the observed mortality trend at adult ages rather than the use of model life tables as an external reference. This last approach allows to reduce regularities in the mortality indicators time evolution series which is supposed to suit perfectly an pertinent analysis of mortality natural evolution. Also, the presented methodology provides lecturer a way to extrapolate mortality at the older ages in all situation especially in dynamic life tables construction.

References

[1] Beers, Beers, H. S. 1944. six-term formulas for routine actuarial interpolation. The record of the American Institute of Actuaries, 33, part (II) (68): 245-260.

[2] Beers, H. S. 1945. Modified-interpolation formula that minimize fourth differences. The record of the American Institute of Actuaries, 34, part. 1, (69): 14-20. (June)

[3] Burnham, K.P. and D.R. Anderson. 1998. Model Selection and Inference. Springer-Verlag, NY.

[4] Buettner T. "Approaches and Experiences in Projecting Mortality Patterns for the Oldest-Old". North American Actuarial Journal. Volume 6. Issue 3. 2002.

[5] Calot G., J.-P. Sardon. Méthodologie relative au calcul des indicateurs démographiques d'Eurostat. Rapport détaillé de l'Observatoire Démographique Européen. Population et conditions sociales 3/2003/E/no 26. Working paper et études. 2004.

[6] Coale A. and Guo G. 1989. Revised Regional Model Life Tables at Very Low Levels of Mortality. Population Index, Vol. 55, No. 4, pp. 613-643.

[7] Coale, Ansley J., Demeny, Paul, and Vaughan, Barbara. 1983. Regional Model Life Tables and Stable Populations (Second Edition). New York/London: Academic Press.

[8] Coale, A J. and Kisker, E.E. 1987. 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." Forthcoming in an NCHS series.

[9] Coale, A.J, and Kisker, E.E. 1990. Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages, Asian and Pacific Population Forum 4:1-31

[10] Crofts G., 1998. A Karup-King Formula with Unequal Intervals". Actuarial Research Clearing House. Vol. 1: 299-303.

[11] Denuit, M. et Goderniaux, A. 2005. Closing and projecting life tables using log-linear models. Bulletin de l'Association Suisse des Actuaries, 1, 29-49.

[12] Flici, F. 2014. Estimation of the missing data in the Algerian mortality surface by using an Age-time-segmented Lee Carter model. Conference paper. Stochastic Modeling and Data Analysis Conference. Lisbon, Portugal (June).

[13] Flici, F. and Hammouda, N.E. 2014. Analysis of half century of motality changes in Algeria : 1962-2012. Conference paper. Middle East Economic Association Conference. Telemcen, Algeria (June).

[14] Flici, F. et Hammouda, N.E. 2015. Analyse de la mortalité en Algérie à travers les résultats de l'enquête MICS IV : Mortalité générale par age et sexe. Présentation atelier de restiution. Alger (Novembre).

[15] Hansen, B. 2007. Notes and comments least squares model averaging. Econometrica, 75(4), 1175-1189.

[16] Heligman, L., and Pollard, J. 1980. The Age Pattern of Mortality. Journal of the Institute of Actuaries 107, 437—55.

[17] Himes, C. L., Preston, S. H., & Condran, G. A. 1994. A relational model of mortality at older ages in low mortality countries. Population Studies, 48(2), 269-291.

[18] Ita, I. K. and Ranjan, K. S. 1984. The problem of choosing model life tables for African countries, Genius, Vol. 40, N° 1/2.

[19] Gavrilov, L. A., and Gavrilova, N. S. 2011. Mortality measurement at advanced ages: a study of the Social Security Administration Death Master File. North American actuarial journal, 15(3), 432-447.

[20] Gavrilova, N. S. and Gavrilov, L. A. 2014. Mortality Trajectories at Extreme Old Ages: A Comparative Study of Different Data Sources on U.S. Old-Age Mortality. Living 100 Monogr.: https://www.soa.org/Library/Monographs/Life/Living-To-100/2014/mono-li14-3a-gavrilova.pdf.

[21] Gompertz, B. 1825. On the Nature of the Function Expressive of the Law of Human Mortality and on a New Mode of Determining Life Contingencies. Royal Society of London, Philosophical Transactions, Series A 115, 513—85.

[22] Kannisto, V. 1994. Development of Oldest-Old Mortality, 1950-1990: Evidence from 28 Developed Countries. Odense University Press, Odense, ISBN: 87 7838 015 4.

[23] Kannisto, V. 1992. Workshop Old Age Mortality, Odense, June.

[24] Kannisto, V. Lauresten, J., Thatcher, A. R., and Vaupel, J.W. 1994. Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. Population and development review. 20, n°. 4 (December)., 793-810.

[25] King, G. 1908. On the Construction of Mortality Tables from Census Returns and Records of Deaths. Journal of the Institute of Actuaries. 42, 3, 225-294.(JULY).

[26] Makeham,W.M.1860. On the law of mortality and the construction of annuity tables. Journal of the Institute of Actuaries 8:301-310.

[27] Meslé, F, et Vallin, J. 2002. Comment améliorer la précision des tables de mortalité aux grands âges ? Le cas de la France. Population, 57e année, n°4-5: 603-631.

[28] Perks, W. 1932. On some experiments on the graduation of mortality statistics, Journal of the Institute of Actuaries 63:12-40.

[29] United Nations. 1981. Model Life Tables for Developing Countries. United Nations Publication, Sales No. E.1981.XIII.7.

[30] Robine, J.-M. and Vaupel, J. W. 2001. Supercentenarians: slower ageing individuals or senile elderly? Experimental Gerontology, Vol. 36, pp. 915-930.

[31] Roli, G. 2008. An adaptive procedure for estimating and comparing the old-age mortality in a long historical perspective: Emilia-Romagna, 1871-2001. Quaderni di Dipartimento Serie Ricerche 2008, n.8 ISSN 1973-9346. http://amsacta.unibo.it/2529/1/Quaderni_2008_Roli_AdaptiveProcedure.pdf

[32] Shrock, H.S., Siegel, J.S. and Associates. 1993. Interpolation: Selected General Methods. Readings in Population Research Methodology. Vol. 1: Basic Tools. Social Development Center / UN Population Fund, Chicago, pp. 5-48–5-72.

[33] Sprague T. B. 1880. Explanation ov a New Formula for Interpolation. Journal of the Institute of Actuaries and Assurance Magazine, 22, 4 , 270-285 (JULY).

[34] Schwarz, G. 1978. Estimating the dimension of a model. The annals of statistics, 6(2), 461-464.

[35] Thatcher, A. R. 1999. The Long-Term Pattern of Adult Mortality and the Highest Attained Age. Journal of the Royal Statistical Society. Series A (Statistics in Society), Vol. 162, No. 1, pp. 5-43.

[36] Thatcher, A. R., Kannisto, V. and Vaupel, J. W. 1998 .The Force of Mortality at Ages 80 to 120. Odense: Odense University Press.

[37] Weibull, W. A. 1951. A statistical distribution function of wide applicability, J Appl. Mech.18:293-297.

[38] Yi, Z. and Vaupel, J. W. 2003. Oldest-Old Mortality in China. Demographric Research. 8, 7: pp 215-244 (April) www.demographic-research.org/Volumes/Vol8/7/ DOI: 10.4054/DemRes.2003.8.7