# Mortality projection using lifecontingencies, demography and StMoMo packages

Giorgio Alfredo Spedicato Ph.D, FCAS, FSA, CSPA C.Stat Gian Paolo Clemente, Ph.D Unicatt

#### Abstract

This paper applies mortality projection techniques (Lee Carter) to the evaluation of retirement costs. The main purpose of the paper is to show how R can be successfully used to perform life expectancy projections with practical actuarial applications for annuity insurances and social security issues. **demography** and **lifecontingencies** packages will be used. The analysis performed within this paper are mechanical and intended for didactic purposes.

Keywords: mortality projection, lifecontingencies, demography, StMoMo.

# 1. Introduction

Mortality data shows that mortality is falling at all ages with a different behavior according to different ages and country. Mortality across any ages is indeed showing continuous reduction across the world.

Prospects of longer life have led to concern over their implications for public spending on oldage support. Forecasting mortality appears indeed a key issue in different field of insurance and financial markets as pricing annuity, evaluating mortality-linked securities and quantifying longevity risk. Lee-Carter proposed in 1992 a model widely used in order to forecast mortality rates.

The Lee Cartel forecasts could be used to project a life table for each specific cohort (year) of birth on which pension annuities projection could be fit.

Our exercise will be based on Italian data downloaded from the Human Mortality Databases (HMD) via **demography** (Rob J Hyndman, Heather Booth, Leonie Tickle, and John Maindonald (2011)) package dedicated functions, that covers a year span from the second half of the XIX century until 2014. In the first part, we will use **demography** and **forecast** package in order to fit Lee - Carter model and perform 100 - years in advance extrapolations. In the second part we will work with the StMoMo package ? for the same purposes, that is a more current approach for making mortality projections.

Finally, **lifecontingencies** (Spedicato (2013)) package will be used to project the cost of a pension annuity,  $\ddot{a}_x^{(m)}$  for the cohorts of 1920, 1930, ..., 2000.

Applications have been provided by using three different data-sets regarding respectively male, female and total population. The use of several data-sets is usually motivated by both the need of a separate evaluation between males and females and of an analysis of models' performance when switching data. However, the change from female to male and a variation in the length of period of observation do not influence the quality of fit as much as the restriction of age range.

Following demographic and economic assumptions will be hold:

- x, the retirement age will be set equal to 65 regardless the cohort.
- m, the number of fractional payments per year, will be equal to 12.
- $\ddot{a}_x^{(m)}$  to be the actuarial present value of a yearly annuity of 1 monetary unit. The annuity will be evaluated assuming an interest rate of 4% and an inflation rate of 2%.

The projection has been performed using a mechanical approach, since the purpose of this paper lies in showing the procedure instead of providing sensible results.

Most of this paper is based on the examples provided in Charpentier (2012) and Charpentier and Dutang (2013) online manual.

Recently, a new R package has made demographic projection even easier, the **StMoMo** package ?. The latter part of this vignettes explains how to perform mortality projections using the **StMoMo** package.

# 2. Using Demography

#### 2.1. Intro and loading mortality data

Lee Carter original model, Lee and Carter (1992), focuses, as main forecasting methodologies, on the central mortality rates m(x,t) for age x in year t defined as the ratio between the number of deaths D(x,t), recorded during the calendar year t for people aged x, and the exposure to risk E(x,t) obtained as the average number of people living during the calendar year t.

Starting by this sample notation, Lee and Carter (1992) proposed to describe the logarithm of central mortality rates as a linear combination of parameters as expressed by Equation (1):

$$\ln m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t} \tag{1}$$

where  $a_x$  describes the general shape of mortality according to different ages and it represents the logarithm of the geometric mean of empirical mortality rates, averaged over historical years.  $e^{a_x}$  measure indeed the general shape across age of the mortality schedule.

Furthermore,  $k_t$  reproduces the underlying time trend, while a term  $b_x$  is considered in order to take into account the different effect of time t at each age.  $b_x$  is assumed to be invariant over time and it explains how rates decline rapidly or slowly in response to change in  $k_t$ .

Finally,  $\varepsilon_{x,t}$  are independent and identical distributed random variables  $N(0, \sigma^2)$  taking into account the age and time specific trends not fully captured by the model.

In the original version, parameters have been estimated by a two-stage process where Singular Value Decomposition (SVD) of the matrix of centered age profiles  $ln(m_{x,t}) - \hat{a}_x$ , allows a first estimation of parameters  $b_x$  and  $k_t$ .

In order to assure a unique solution for the system of equation of the model, Lee and Carter proposed the following constraints:  $\sum_{t} k_t = 0$  and  $\sum_{x} b_x = 1$ .

A second step, based on a refitting of  $\hat{k}_t$  on the number of deaths, is usually suggested in order to assure a better convergence between estimated and observed deaths. The aim is to find the  $\hat{k}_t$  such that  $D(x,t) = E(x,t)exp(\hat{a}_x + \hat{b}_x \cdot \hat{k}_t)$ .

Alternative frameworks have been proposed over the years in order to improve some drawbacks of original Lee-Carter model (in particular see ?,?, ?, ?, ?, ?, ?)

The one - year survival probability at age x during calendar year t is expressed by Equation (2). Equation (2) assumes constant force of mortality to hold between [x, x + t) and that  $\mu_x \sim m_x$ , that is the force of mortality to be approximated by the central rate of mortality<sup>1</sup>:

$$p_{x,t} = \exp\left(-\mu_{x,t}\right) \sim \exp\left(-m_{x,t}\right). \tag{2}$$

A longitudinal life table for the cohort of born in calendar year YYYY can be created selecting all  $p_{x,t}$  for which t - x = YYYY.

We will perform such exercise on Italy HMD data and by applying the original version of Lee-Carter.

R>library(demography)
R>library(forecast)
R>library(lifecontingencies)

Following code import data from the Human mortality Database and it creates a demogdata object from HMD data structure. The hmd.mx function downloads all available annual data by single years of age, but for the application we will use the already saved data.

```
R>#italyDemo<-hmd.mx(country="ITA", username="username@email.domain",
R>#password="password", label="Italy")
R>load(file="mortalityDatasets.RData")
```

Plot method is available on demogdata. Following figures report for the Italian Population the pattern of logarithm of death rates according to age and time. As detailed in demography vignette, different colors indicate different years (most recent ones in violet, earliest in red). Several behavior are shown respectively for male, female and total population.

```
R>par(mfrow=c(1,3))
R>plot(italyDemo,series="male",datatype="rate", main="Male rates")
R>plot(italyDemo,series="female",datatype="rate", main="Female rates")
R>plot(italyDemo,"total",datatype="rate", main="Total rates")
```

```
<sup>1</sup>If p_{x,t} were assumed linear between the two consecutive integer ages, we could write m_x = \frac{q_x}{1-\frac{1}{2}q_x}.
```





Italian data confirms that mortality is falling at all ages with a different behavior according to different ages

## 2.2. Fitting Lee Carter model

To fit Lee - Carter model (without going through logarithms) lca function can be used. Lee-Carter is here applied separately between male, female and total population and by considering a maximum age equal to 100.

```
R>italyLcaM<-lca(italyDemo,series="male",max.age=100)
R>italyLcaF<-lca(italyDemo,series="female",max.age=100)
R>italyLcaT<-lca(italyDemo,series="total",max.age=100)</pre>
```

**lca** returned object allows us to inspect  $a_x$ ,  $b_x$  and  $k_t$ . Figures represent the values of the estimated parameters.

```
R> par(mfrow=c(1,3))
R> plot(italyLcaT$ax, main="ax", xlab="Age",ylab="ax",type="l")
R> lines(x=italyLcaF$age, y=italyLcaF$ax, main="ax", col="red")
R> lines(x=italyLcaM$age, y=italyLcaM$ax, main="ax", col="blue")
```

```
R>
    legend("topleft" , c("Male", "Female", "Total"),
     cex=0.8,col=c("blue","red","black"),lty=1);
+
    plot(italyLcaT$bx, main="bx", xlab="Age",ylab="bx",type="l")
R>
R>
    lines(x=italyLcaF$age, y=italyLcaF$bx, main="bx", col="red")
    lines(x=italyLcaM$age, y=italyLcaM$bx, main="bx", col="blue")
R>
    legend("topright" , c("Male", "Female", "Total"),
R>
     cex=0.8,col=c("blue","red","black"),lty=1);
+
    plot(italyLcaT$kt, main="kt", xlab="Year",ylab="kt",type="l")
R>
R>
    lines(x=italyLcaF$year, y=italyLcaF$kt, main="kt", col="red")
    lines(x=italyLcaM$year, y=italyLcaM$kt, main="kt", col="blue")
R>
    legend("topright" , c("Male", "Female", "Total"),
R>
     cex=0.8,col=c("blue","red","black"),lty=1);
+
```



A similar behavior of parameters is observed according to different data-sets. As expected the average mortality grows when age increases (see  $\hat{a}_x$  pattern).Furthermore it is clearly visible the young mortality hump for males in the age-range (20,30) due to accidental deaths.  $\hat{b}_x$  shows instead a greater value for younger ages and a greatest improvement for females in the age range (60-80). Finally, as expected,  $\hat{k}_t$  has a decreasing trend with the increment of time.

We can therefore use **forecast** package to project the future  $k_t$ s (up to 110). Projection is

```
\mathbf{6}
```

based on ARIMA extrapolation.

R>fM<-forecast(italyLcaM,h=110)
R>fF<-forecast(italyLcaF,h=110)
R>fT<-forecast(italyLcaT,h=110)</pre>

The predicted values of  $k_t$  re-scaled to zero in the last observed year (2014) are here reported.

```
R>par(mfrow=c(1,3))
R>plot(fM$kt.f,main="Male")
R>plot(fF$kt.f,main="Female",)
R>plot(fT$kt.f,main="Total")
```



Finally, it's easy to derive the full pattern of rates. Past and projected rates are here binded in the same matrix.

```
R>ratesM<-cbind(italyDemo$rate$male[1:100,],fM$rate$male[1:100,])
R>ratesF<-cbind(italyDemo$rate$female[1:100,],fF$rate$female[1:100,])
R>ratesT<-cbind(italyDemo$rate$total[1:100,],fT$rate$total[1:100,])</pre>
```

We report here the pattern of past and projected rates according to different population for people aged 65. The expected improvement is clearly visible in the Figure.

```
R>par(mfrow=c(1,1))
R>plot(seq(min(italyDemo$year),max(italyDemo$year)+110),ratesF[65,],
+ col="red",xlab="Years",ylab="Death Rates",type="l")
R>lines(seq(min(italyDemo$year),max(italyDemo$year)+110),ratesM[65,],
+ col="blue",xlab="Years",ylab="Death Rates")
R>lines(seq(min(italyDemo$year),max(italyDemo$year)+110),ratesT[65,],
+ col="black",xlab="Years",ylab="Death Rates")
R>legend("topright", c("Male","Female","Total"),
+ cex=0.8,col=c("blue","red","black"),lty=1);
```



We have applied here the original version of Lee-Carter in order to obtain a forecast of mortality rates. Alternative estimates can be derived by using **lca** function through the **demography** package (as Lee-Miller (?), Booth-Maindonald-Smith (?) and Hyndman-Ullah (?) methods). Finally Lifemetrics package allows to fit ?, ? and ? models.

#### 2.3. Perform actuarial projections

Our aim is to create a function to project life table depending by year of birth, using results

from Lee - Carter model. In particular, for ages  $0, 1, \ldots, \tau$  on which Lee-Carter model has been fit Equation 3 apply, while for extreme ages,  $\tau + 1, \ldots, \omega$  on which no data were provided, it has been assumed that on year probability decreases evenly in 20 steps.

$$\ln \hat{\mu}_{x,t} = \hat{a}_x + \hat{b}_x \hat{k}_t \hat{p}_{x,t} = \exp(-\hat{\mu}_{x,t})$$
(3)

```
R>createActuarialTable<-function(yearOfBirth,rate){</pre>
+
+
     mxcoh <- rate[1:nrow(rate),(yearOfBirth-min(italyDemo$year)+1):ncol(rate)]</pre>
     cohort.mx <- diag(mxcoh)</pre>
+
     cohort.px=exp(-cohort.mx)
+
     #get projected Px
+
     fittedPx=cohort.px #add px to table
+
           px4Completion=seq(from=cohort.px[length(fittedPx)], to=0, length=20)
           totalPx=c(fittedPx,px4Completion[2:length(px4Completion)])
           #create life table
+
           irate=1.04/1.02-1
+
+
+
           cohortLt=probs2lifetable(probs=totalPx, radix=100000,type="px",
     name=paste("Cohort",yearOfBirth))
+
           cohortAct=new("actuarialtable",x=cohortLt@x, lx=cohortLt@lx,
+
           interest=irate, name=cohortLt@name)
+
           return(cohortAct)
+
+
           }
R>
R >
```

We can therefore calculate the APV of  $\ddot{a}_{65}^{(12)}$  for the selected cohorts. Values have been derived separately between males and females and by using directly the total population.

```
getAnnuityAPV<-function(yearOfBirth,rate) {</pre>
R>
+
                   actuarialTable<-createActuarialTable(yearOfBirth,rate)
                   out=axn(actuarialTable,x=65,m=12)
+
                   return(out)
+
           }
+
R>rate<-ratesM
R>for(i in seq(1920,2000,by=10)) {
                   cat("For cohort ",i, "of males the e0 is",
+
                   round(exn(createActuarialTable(i,rate)),2),
+
                   " and the APV is :",round(getAnnuityAPV(i,rate),2),"\n")
+
+
           }
+
For cohort 1920 of males the eO is 50.76 and the APV is : 4.24
For cohort 1930 of males the eO is 60.3 and the APV is : 5.13
For cohort 1940 of males the e0 is 64.76 and the APV is : 5.96
```

```
For cohort 1950 of males the e0 is 72.13 and the APV is : 6.3
For cohort 1960 of males the e0 is 75.51 and the APV is : 6.63
For cohort 1970 of males the e0 is 78.56 and the APV is : 6.95
For cohort 1980 of males the e0 is 81.31 and the APV is : 7.26
For cohort 1990 of males the e0 is 83.14 and the APV is : 7.56
For cohort 2000 of males the e0 is 84.53 and the APV is : 7.86
R>rate<-ratesF
R>for(i in seq(1920,2000,by=10)) {
             cat("For cohort ",i, "of females the e0 at birth is",
          round(exn(createActuarialTable(i,rate)),2),
+
+
           " and the APV is :",round(getAnnuityAPV(i,rate),2),"\n")
+
          7
+
For cohort 1920 of females the e0 at birth is 59.51 and the APV is : 6.54
For cohort 1930 of females the e0 at birth is 67.15 and the APV is : 7.37
For cohort 1940 of females the e0 at birth is 71.26 and the APV is : 8
For cohort 1950 of females the e0 at birth is 78.41 and the APV is : 8.52
For cohort 1960 of females the e0 at birth is 82.16 and the APV is : 8.91
For cohort 1970 of females the e0 at birth is 84.97 and the APV is : 9.28
For cohort 1980 of females the e0 at birth is 87.33 and the APV is : 9.62
For cohort 1990 of females the e0 at birth is 88.69 and the APV is : 9.95
For cohort 2000 of females the e0 at birth is 89.79 and the APV is : 10.27
R>rate<-ratesT
R>for(i in seq(1920,2000,by=10)) {
       cat("For cohort ",i, "of total population the e0 is",
+
                   round(exn(createActuarialTable(i,rate)),2),
+
                   " and the APV is :",round(getAnnuityAPV(i,rate),2),"\n")
+
+
           }
+
For cohort 1920 of total population the e0 is 55.06 and the APV is : 5.51
For cohort 1930 of total population the e0 is 63.71 and the APV is : 6.34
For cohort 1940 of total population the e0 is 68.04 and the APV is : 7.09
For cohort 1950 of total population the e0 is 75.37
                                                     and the APV is : 7.54
For cohort 1960 of total population the e0 is 78.95
                                                     and the APV is : 7.92
For cohort 1970 of total population the e0 is 81.92 and the APV is : 8.29
For cohort 1980 of total population the e0 is 84.54
                                                     and the APV is : 8.65
For cohort 1990 of total population the e0 is 86.19
                                                     and the APV is : 8.99
For cohort 2000 of total population the e0 is 87.49
                                                     and the APV is : 9.31
```

# 3. Using the StMoMo package

The StMoMo package provides many useful functions to manage mortality data.

### 3.1. Loading data

We will recycle previous example's data to initialize mortality data. The interested reader will find further details on the **StMoMo** package's vignette.

```
R>#loading StMoMo package
R>require(StMoMo)
R>#get data
R>#get data
R>##use total death, mid year exposures
R>ita.StMoMoData<-StMoMoData(data=italyDemo, series = "total",type="central")
R>#bring to initial year
R>ita.StMoMoData.Ini<-central2initial(ita.StMoMoData)
R>#assume max age 103
R>ages.fit = 0:103
R>#generate weight matrix
R>wxt <- genWeightMat(ages = ages.fit, years = ita.StMoMoData.Ini$years,clip = 3)</pre>
```

#### 3.2. Fitting models

The code below fits the Lee Carter models as well as the Reinshaw Haberman (RH) and the Cairn Blake Down (CBD) one ? ones.

```
R>#fitting the models
R>## LC
R>LC <- lc(link = "logit")
R>LCfit <- fit(LC, data = ita.StMoMoData.Ini, ages.fit = ages.fit, wxt = wxt)
R>## CBD
R>CBD <- cbd()
R>CBDfit <- fit(CBD, data = ita.StMoMoData.Ini, ages.fit = ages.fit, wxt = wxt)</pre>
```

Some are long to be fit (especially the RH one):

```
R>## RH
R>RH <- rh(link = "logit", cohortAgeFun = "1")
R>RHfit <- fit(RH, data = ita.StMoMoData.Ini, ages.fit = ages.fit, wxt = wxt,start.ax =
+ start.bx = LCfit$bx, start.kt = LCfit$kt)</pre>
```

```
R>load(file = "StMoMoModels.RData")
```

Plotting fitted models is immediate as well: The figure below shows LC fits:

R>plot(LCfit)



The figure below shows CDB fits:



The figure below shows RH fits:

R>plot(RHfit)

## 3.3. Projecting and simulating mortality models

The following code projects individual fits for forthcoming 50 years:

R>horizon=50
R>LCfor <- forecast(LCfit, h = horizon)
R>#assuming arima with drift
R>#RHfor <- forecast(RHfit, h = horizon, gc.order = c(1,1,0))
R>CBDfor <- forecast(CBDfit, h = horizon)</pre>

Also it is possible to simulate mortality models' projections:

```
R>#simulazione & proiezione
R>##simulazione
R>nsims=100
```

```
R>#RHsim <- simulate(RHfit, nsim = nsims, h = 50, gc.order = c(1, 1, 0))
R>LCsim <- simulate(LCfit, nsim = nsims, h = 50)
R>CBDsim <- simulate(CBDfit, nsim = nsims, h = 50)
```

See for example 20 year projected LC trajectories:

```
R>##LC
R>plot(LCfit$years, LCfit$kt[1, ], xlim = range(LCfit$years, LCsim$kt.s$years), ylim = ran
R>matlines(LCsim$kt.s$years, LCsim$kt.s$sim[1, , 1:20], type = "1", lty = 1)
R>plot(LCfit$years, LCfit$kt[1, ], xlim = range(LCfit$years, LCsim$kt.s$years),
+ ylim = range(LCfit$kt, LCsim$kt.s$sim[1, , 1:20]), type = "1",
+ xlab = "year", ylab = "kt", main = "LC model simulations")
R>matlines(LCsim$kt.s$years, LCsim$kt.s$sim[1, , 1:20], type = "1",
+ lty = 1)
```



# LC model simulations

#### 3.4. Obtaining life tables

The **StMoMo** package provides a useful function to extract mortality rates from model objects' fit and forecasts, the **extractCohort** one. In order to obtain a projected life table for a given

14

cohort of birth, we need to blend fitted rates from historical data to projection, all from the same cohort.

```
R>#plotting historical fitted rates, until 2014
R>chosen_cohort=1950
R>plot(0:64, extractCohort(fitted(LCfit, type = "rates"), cohort = chosen_cohort),
+ type = "1", log = "y", xlab = "age", ylab = "q(x)",
+ main = "Cohort 1950 mortality rate", xlim = c(0,103), ylim = c(0.0005, 0.07))
R>#adding fitted projections
R>lines(65:103, extractCohort(LCfor$rates, cohort = chosen_cohort), lty = 2, lwd=2, col="r
R>chosen_cohort = 1950
R>plot(0:64, extractCohort(fitted(LCfit, type = "rates"), cohort = chosen_cohort),
+ type = "1", log = "y", xlab = "age", ylab = "q(x)", main = "Cohort 1950 mortality r
+ xlim = c(0, 103), ylim = c(5e-04, 0.07))
R>lines(65:103, extractCohort(LCfor$rates, cohort = chosen_cohort),
```

```
+ lty = 2, lwd = 2, col = "red")
```





So, the whole mortality rates,  $m_x$ , can be obtained as follows:

Now we use lifecontingencies functions to transform mortality rates to mortality probabilities

```
R>lc_qx_1950<-mx2qx(lc_rates_1950)
R>#rh_qx_1950<-mx2qx(rh_rates_1950)
R>cbd_qx_1950<-mx2qx(cbd_rates_1950)</pre>
```

And to get lifetable objects:

```
R>lc_lifetable_1950<-probs2lifetable(probs=lc_qx_1950,type = "qx",
+ name = paste("LC","1950","lt",sep="_"))
R>#rh_lifetable_1950<-probs2lifetable(probs=rh_qx_1950,type = "qx",name = paste("RH","1950
R>cbd_lifetable_1950<-probs2lifetable(probs=cbd_qx_1950,type = "qx",
+ name = paste("CDB","1950","lt",sep="_"))
```

Then, we can easily obtain expected life time at x = 65:

```
R>exn(lc_lifetable_1950,x=65)
```

[1] 21.43934

R>#exn(rh\_lifetable\_1950,x=65)
R>exn(cbd\_lifetable\_1950,x=65)

[1] 19.2421

Finally, after having created actuarial tables:

```
R>lc_acttbl_1950<-new("actuarialtable",x=lc_lifetable_1950@x,lx=lc_lifetable_1950@lx, inte
R>#rh_acttbl_1950<-new("actuarialtable",x=rh_lifetable_1950@x,lx=rh_lifetable_1950@lx, int
R>cdb_acttbl_1950<-new("actuarialtable",x=cbd_lifetable_1950@x,lx=cbd_lifetable_1950@lx, int</pre>
```

```
16
```

We can compute the APV of an annuity under the three mortality assumptions:

R>axn(actuarialtable = lc\_acttbl\_1950,x=65)

[1] 18.76793

R>#axn(actuarialtable = rh\_acttbl\_1950,x=65)
R>axn(actuarialtable = cdb\_acttbl\_1950,x=65)

[1] 17.2352

#### Acknowledgments

The authors are deeply indebited with Prof. Rob J. Hyndman for his suggestions on these vignettes and to all people that contributed to **lifecontingencies** package.

# References

- Charpentier A (2012). "Actuarial Science with R 2: Life Insurance and Mortality Tables." http://freakonometrics.blog.free.fr/index.php?post/2012/04/04/ Life-insurance,-with-R,-Meielisalp. Accessed: 04/11/2012.
- Charpentier A, Dutang C (2013). Actuariat Avec R. http://cran.r-project.org/doc/ contrib/Charpentier\_Dutang\_actuariat\_avec\_R.pdf.
- Lee R, Carter L (1992). "Modeling and Forecasting U.S. Mortality." Journal of the American Statistical Association, 87(419), 659–675. doi:10.2307/2290201.
- Rob J Hyndman, Heather Booth, Leonie Tickle, John Maindonald (2011). *demography: Forecasting Mortality, Fertility, Migration and Population Data.* R package version 1.09-1, URL http://CRAN.R-project.org/package=demography.
- Spedicato GA (2013). "The lifecontingencies Package: Performing Financial and Actuarial Mathematics Calculations in R." *Journal of Statistical Software*, **55**(10), 1–36. URL http://www.jstatsoft.org/v55/i10/.

#### Affiliation:

Gian Paolo Clemente Catholic University of Milan Department of Mathematics, Finance and Econometrics gianpaolo.clemente@unicatt.it