Matlab Longevity Toolbox

Eric Jondeau, Michael Rockinger, and Hristo Hristov

2015

All three authors are from University of Lausanne, Faculty of Business and Economics, CH 1012 Lausanne, Switzerland and Swiss Finance Institute. Jondeau and Rockinger are Professors of Finance and Hristov is a Master's of Science in Finance student.
Respective e-mails are: eric.jondeau@unil.ch, michael.rockinger@unil.ch, and hristo.hristov@live.fr.
Please forward all comments to michael.rockinger@unil.ch

The authors acknowledge financial support from the Amundi-Dauphine Foundation Chair in Asset Management. This work could not have been written without the data obtained from the Human Mortality Database (2015). The usual disclaimer applies.

1 Scope

This Longevity Toolbox provides a sequence of Matlab programs to estimate common mortality models. With a minimum of effort those models can be extended to investigate new venues or adapted to other data. Once parameters have been estimated, the user can forecast mortality rates by fitting time series models to the parameters.

Requirements: To run all methods, the user should have the Matlab optimizer installed. If this is not the case, many modern models such as Lee-Carter, Renshaw-Haberman and Cairns, Blake, Dowd that do not rely on the Matlab optimizer can still be estimated.

To get a quick start unpack the toolbox. Incorporate the location of the **code** folder on Matlab path using the "Add subfolders" option. Then running `mainEstimateModels.m` returns information on the estimation of many popular models.

The structure of this manual is as follows: In this first section we will describe the models provided. In the next section we describe the sample data used, a very small subset from the Human mortality database. Users of this data are strongly encouraged to register at the website of the Human Mortality Database. Upon registration users have to possibility to use for free the entire database.

We leave further technical aspects such as how to compute the maximum likelihood estimates to our working paper Jondeau, Rockinger (2015).

Basic Notations

- cohort - a group of individuals born in a certain year, sharing the same age through time.
- x - the age of an individual. In particular, we will use total mortality for both genders in the empirical section.
- m_x - central rate of mortality. It is approximately equal to the average force of mortality, averaged over the year of age.
- q_x - the probability that someone aged exactly x will die before reaching age $x + 1$.

In our empirical analysis we consider ages in the range $x_1, \cdots, x_{n_x}$. In the codes, we use $alpha$ and $omega$ to denote the youngest and oldest individual. It is assumed that the mortality rates are given over regular spaced time periods indexed by $t = 1, \cdots, T$ typically years.

Models

The models published with this toolbox are:

HP Heligman and Pollard (1980) propose parametric models useful for fitting mortality rates for the entire age spectrum. Given the very large number of parameters it is difficult to forecast this model, however it may be useful for demographic studies. For this reason we did not include it in Jondeau, Rockinger (2015) but since we tested their models, we include them here.

LC1 Lee and Carter (1992). Based on a singular value decomposition. We will consider the single factor decomposition. This model assumes Gaussian errors.

LC2 In a discussion of Lee (2000), Alho (2000) noticed a better fit if errors are assumed Poisson and if instead of fitting mortallity, one fits actual death-rates.

RH Renshaw and Haberman (2006) introduce a factor for each cohort which dramatically improves the in sample goodness of fit.

CBD1 Cairns, Blake and Dowd (2006) propose a rather different model. Instead of estimating an age specific component that gets changed over time, they use directly the given age profile. This model is a single factor model.

CBD2 Cairns, et al. (2009) propose a two factor model allowing for a cohort effect.[1]

CBD3 Cairns, et al. (2009) propose a linear and quadratic age effect in addition to a cohort effect.

The expressions of the various models are:

$$\begin{aligned}
\text{HP1:} \quad & q_x/(1-q_x) = A^{(x+B)^C} + D * \exp(-E * (\ln x - \ln F)^2) + GH^x + \eta_{x,t}, \\
\text{LC1:} \quad & \ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \eta_{x,t}, \\
\text{LC2:} \quad & \ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \varepsilon_{x,t}, \\
\text{RH:} \quad & \ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)} + \varepsilon_{x,t}, \\
\text{CBD1:} \quad & \operatorname{logit} q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \varepsilon_{x,t}, \\
\text{CBD2:} \quad & \operatorname{logit} q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \gamma_{t-x}^{(3)} + \varepsilon_{x,t}, \\
\text{CBD3:} \quad & \operatorname{logit} q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)} \left((x-\bar{x})^2 - \hat{\sigma}_x^2\right) + \gamma_{t-x}^{(4)} + \varepsilon_{x,t}.
\end{aligned}$$

In those equations $\eta_{x,t}$ means a Gaussian distribution and $\varepsilon_{x,t}$ means that the model drives a Poisson counter describing the number of death. We have the logit transform $\operatorname{logit} q = \ln(q/(1-q))$, $\bar{x} = \frac{1}{n_a}\sum_{j=1}^{n_a} x_j$ is the average of ages considered, and $\hat{\sigma}_x^2 = \frac{1}{n_a}\sum_{j=1}^{n_a}(x_j-\bar{x})^2$ is the variance of the ages. In those various models, $\beta_x^{(1)}$ represents an average age specific mortality parameter. The parameters $\beta_x^{(2)}$ and $\beta_x^{(3)}$ are also age specific and deal with how mortality of different ages is affected by a time effect captured by $\kappa_t^{(2)}$ or by the cohort effect $\gamma_{t-x}^{(3)}$ concerning those that are born in year $t-x$. One very important assumption of models LC1, LC2 and RH1 is that the $\beta_x^{(i)}$ are constant over time.
The parameters $\kappa_t^{(i)}$capture time period related effect. Typically, this variable need to be modeled for forecasting purposes.
Model LC1 results directly from a Singular Value Decomposition (SVD) and we present only the single factor decomposition. It is trivial to extend the model to several factors. Models LC2, RH, CBD1,2,3 are estimated using the assumption that the death rate

[1]For convenience we refer to those models as CBD.

is given by a Poisson distribution. An Appendix provided in Jondeau, Rockinger (2015) specifies the updating rules for the various models. There we also indicate the identification conditions chosen.

2 Data used

The raw data files are from the Human Mortality Database. Use of this data requires registration at http://www.mortality.org. For illustration purposes, we work with a subset of US data for both genders. The Human Mortality Database contains data for many more countries as well as the full data for the US. Since we only provide an illustration our focus is on the tables of both genders.

- `bltper_1x1.txt` - Both genders life tables from 1921 to 2012.[2] We limited the data to the years 1947 to 2010. For earlier and later data, as well as other country data, please, download your own dataset. Those files need to be processed in a text editor to remove from the highest age groups 110+ the +.

Inspection of the raw file yields:

Year	Age	m_x	q_x	a_x	l_x	d_x	L_x	T_x	e_x
1947	0	0.03808	0.03689	0.15	100000	3689	96878	6668511	66.69
1947	1	0.00266	0.00266	0.50	96311	256	96183	6571633	68.23
1947	2	0.00160	0.00160	0.50	96054	154	95978	6475450	67.41
...									
1947	108	0.61790	0.47206	0.50	3	2	3	5	1.58
1947	109	0.63969	0.48467	0.50	2	1	1	3	1.54
1947	110+	0.66093	1.00000	1.51	1	1	1	1	1.51

The Matlab function `[year,age,mxt,qxt,lxt,dxt,ext] = loadData()` loads selected columns of the data.
The script file `main_Checkdata.m` prints the data to file CheckData.txt. Verify that the data is correct. Always do this.
Since it takes a little to read and process a textfile it is more convenient to transform the data once and for all and store it as a structure in matlab binary format. To do so, the program `mainCreateStruct.m` reads the Human Mortality Database data and saves it as as structure in `LifeTableS.mat`. Returns: `year = LifeTableS.year; age = LifeTableS.age; qxt = LifeTableS.qxt;`
The vector `year` contains all the years and `age` the ages.

To illustrate the impact of using period tables versus actual time varying mortality rates, we provide `main_AnnuityCost.m` and `main_AnnuityActualCost.m` that compute a life annuity, once by using mortality rates using a static actuary picture given by current mortality rates and once by using the actual rates. Convince yourself of the large difference.

3 Implementing Various Models

The script

`main_EstimateModels.m`

[2] The United States of America, Life tables (period 1x1), Total Last modified: 24-Jun-2013, MPv5 (May07).

displays a typical function call to the various models described below.
Each function call for the various estimations follows the structure

```
EstS = NameOfModel(parameters),
```

where the structure `EstS` contains information on the estimation such as the name of the model, the time the estimation took, as well as the parameter estimates. Each field can be accessed in the usual object oriented way `EstS.model`, `EstS.resids`, etc.
Each function prints and plots some results during the run. It is possible to further accelerate the speed of the functions by commenting those instructions out. We like a certain interaction with the programs.

3.1 Heligman - Pollard Models

3.1.1 HP1 Model

$$\text{HP1:} \quad q_x/(1-q_x) = A^{(x+B)^C} + D * \exp(-E * (\ln x - \ln F)^2) + GH^x$$

HP1 model

- Call: EstS = HPmodel1(year,yearu,alpha,omega,qx)
- Description: Estimates the parameters of the HP1 model for a give year with Nonlinear Least Squares (NLS).
- All models in the Heligman-Polard family require the Matlab (or equivalent) optimizer (allowing for constraints).
- Inputs:
  - `year` : year for which one wishes to perform estimation
  - `yearu` : vector of all years for which qx contains data.
  - `alpha` : youngest age to be used in estimation. Typically 0.
  - `omega` : oldest age to be used in estimation. Typically 110.
  - `qx` : Matrix of mortality rates. The rows are for ages and columns are for the various years
- Outputs: `EstS` a structure with the following fields
  - `EstS.model` - name of model estimated
  - `EstS.resids` - difference between mortality and estimates
  - `EstS.paramv` - vector of parameters A,B,... as in formula
  - `EstS.exitflag` - a flag to indicate the convergence criterion of optimization. As Matlab optimizer manual indicates, only use estimates if this is 1.
  - `EstS.resids` - residuals defined as the mortality minus estimated mortality
  - `EstS.Rsquare` - a measure of goodness of fit

- Remark: The program is pretty straightforward but involves at some point a vector `scal = [ 1/1000; 1/100; 1/10; 1/10000; 10; 10; 1/100000; 1];` which is used to scale the parameters during their estimation. This is a useful tric to stabilize optimization for situations like here where the parameters are of very different magnitude. After use of scal the optimizer will use numbers of similar magnitude, say in the range 1 to 10. This dramatically stabilizes the optimization, otherwise no solution may actually be found.
- Remark: We have chosen to leave the selection of the required elements of the qx matrix in the function. By doing so it it possible to adapt the function to estimations for all years for which there is data.

3.1.2 HP1a Model

The HP1a model is obtained by a modification of the last term of the equation. This may yield a better fit for old ages.

$$\text{HP1a:} \quad q_x/(1-q_x) = A^{(x+B)^C} + D * \exp(-E * (\ln x - \ln F)^2) + GH^x/(1+GH^x)$$

HP1a model

- Call: EstS = HPmodel1a(year,yearu,alpha,omega,qx)
- Description: Estimates the parameters of the HP1 model for a give year
- Inputs and Outputs as for HPmodel1.

3.1.3 HP2 Model

The HP2 model is obtained by introducing a new parameter in the last term.

$$\text{HP2:} \quad q_x/(1-q_x) = A^{(x+B)^C} + D * \exp(-E * (\ln x - \ln F)^2) + GH^x/(1+KGH^x)$$

HP2 model

- Call: EstS = HPmodel2(year,yearu,alpha,omega,qx)
- Description: Estimates the parameters of the HP2 model for a give year
- Inputs and Outputs as for HPmodel1.

3.1.4 HP3 Model

This time again the starting vector contains 9 parameters.

$$\text{HP3:} \quad q_x/(1-q_x) = A^{(x+B)^C} + D * \exp(-E * (\ln x - \ln F)^2) + GH^x/(1+GH^{x^K})$$

HP3 model

- Call: EstS = HPmodel3(year,yearu,alpha,omega,qx)
- Description: Estimates the parameters of the HP2 model for a give year
- Inputs and Outputs as for HPmodel1.

3.2 LC Models

3.2.1 LC1 Model

The Lee-Carter model 1 can be expressed in the following equation

$$\text{LC1:} \quad \ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \varepsilon_{x,t}$$

LC1 Model

- Call: `EstS = LCmodel1(yearStart,yearEnd,yearv,alpha,omega,agev,qx)`
- Description: Estimation of the parameters of the LC1 model using singular value decomposition. This model requires several years of data.
- Inputs:
  - `yearStart` - Year of start of the estimation of the parameters
  - `yearEnd` - Year of end of the estimation of the parameters
  - `yearv` - vector of all possible years. Must contain yearStart and yearEnd
  - `alpha` - Minimum age involved in estimation
  - `omega` - Maximum age allowed by the model
  - `agev` - vector of all possible ages
  - `qx` - a matrix with as many rows as there are ages (in agev) and as many columns as there are years (in yearv)
- Outputs:
  - `EstS.model` - name of model estimated
  - `EstS.alpha` - youngest age involved
  - `EstS.omega` - oldest age could be 85 or 90
  - `EstS.yearStart` - first year of data to be used in estimation

– `EstS.yearEnd` - last year for the estimation

– `EstS.nbYrEst` - returns the number of years over which the estimation was done

– `EstS.nbAges` - number of ages for which the estimation was made

– `EstS.tEst` - time required for the estimation

– `EstS.axhat` - $\beta_x^{(1)}$ parameter

– `EstS.bxhat` - $\beta_x^{(2)}$ parameter

– `EstS.kthat` - $\kappa_t^{(2)}$ parameter

– `EstS.resids` - residuals defined as difference between actual mortality and estimated mortality for the block of years and ages considered

■ Remark: The program can be easily extended to a second and even a third factor via the singular value decomposition.

The following figures have been generated with the following set of parameters: `yearStart = 1947, yearEnd = 2010, alpha = 50, omega = 90, gender = 'both'.`

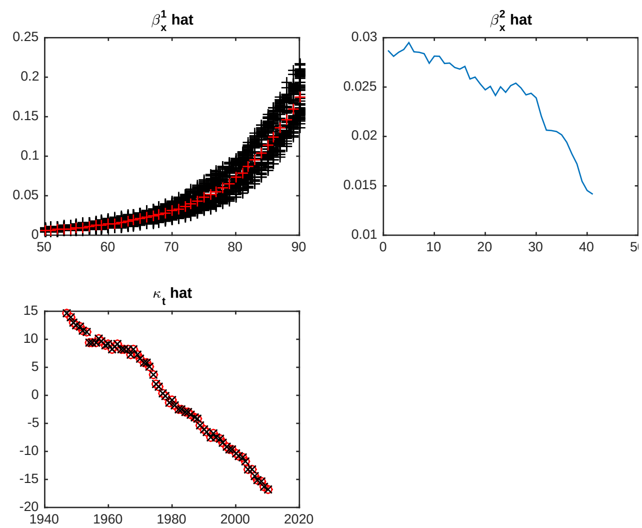

Fig. Various parameter estimates (in red) and in black the actual observations of mortality rates (in $\hat{\beta}_x^1$)

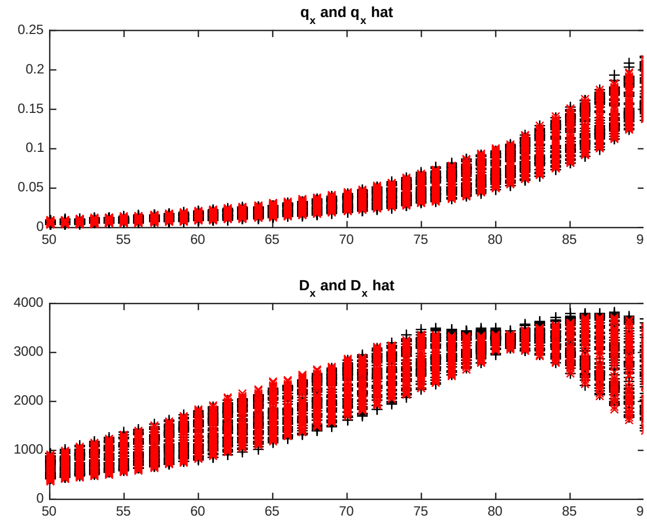

Fig. Mortality rate and death counts, actual and estimated for the LC1 model

3.2.2 LC2 Model

The Lee-Carter model 2 can be expressed with the following equation

$$\text{LC2:} \quad \ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \varepsilon_{x,t}.$$

Estimation of this model follows the description given in McCullagh, Nelder (1989) and reminded in the appendix of Jondeau, Rockinger (2015). It involves maximizing the likelihood in an iterative manner along different gradient directions.

LC2 Model

- Call: `EstS = LCmodel2(yearStart,yearEnd,yearv,alpha,omega,agev,qx)`
- Description: Estimation of the parameters of the LC2 model under the assumption that the death count generated by the mortality rate follows a Poisson distribution.
- Inputs: same as for LCmodel1.
- Outputs: same as for LCmodel1. In addition the program returns
  - `Est.dist` - the sum of absolute value of difference of consecutive vectors produced by the iteration involved in the ML estimation. This number should be smaller than 10^{-10}.
  - `Est.iterCtr` - a counter that gets incremented in each iteration. This number should be smaller than 10'000 for a correct convergence. In the code, if necessary, this number can be increased.
- Remark: The accuracy of the estimation, that is the maximum absolute value of the parameter vector between two iterations, is set to 10^{-10}. It can be reduced but the estimation is very fast anyhow.

3.3 Renshaw Haberman Model

This section covers the model defined by Renshaw and Haberman (2006).

$$\text{RH:} \quad \ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_{t-x}^{(3)} + \varepsilon_{x,t}.$$

This model gets estimated under the assumption of Poisson distributed errors.

RH model

- Call: `EstS = RHmodel(yearStart,yearEnd,yearv,alpha,omega,agev,qx)`
- Description: Estimates the parameters of the RH model under assumption that ML is Poisson driven.
- Inputs: same as for LCmodel1.
- Outputs:
  - `EstS.model` - name of model estimated
  - `EstS.alpha` - youngest age involved
  - `EstS.omega` - oldest age could be 85 or 90
  - `EstS.yearStart` - first year of data to be used in estimation
  - `EstS.yearEnd` - last year for the estimation
  - `EstS.nbYrEst` - returns the number of years over which the estimation was done
  - `EstS.nbAges` - number of ages for which the estimation was made
  - `EstS.tEst` - time required for the estimation
  - `Est.dist` - the sum of absolute value of difference of consecutive vectors produced by the iteration involved in the ML estimation. This number should be smaller than 10^{-6}
  - `Est.iterCtr` - a counter that gets incremented in each iteration. This number should be smaller than 10'000 for a correct convergence. In the code, if necessary, this number can be increased.
  - `EstS.axhat1` : $\beta_x^{(1)}$ parameter
  - `EstS.bxhat1` : $\beta_x^{(2)}$ parameter
  - `EstS.kthat1` : $\kappa_t^{(2)}$ parameter
  - `EstS.bxhat2` : $\beta_x^{(3)}$ parameter
  - `EstS.gzhat` : $\gamma_{t-x}^{(3)}$ parameter
  - `EstS.resids` - residuals defined as difference between actual mortality and estimated mortality for the block of years and ages considered

  Remark: The reconstruction of the actual mortality rates is a bit more complex here since each cohort evolves along the diagonal of the age-year matrix. This needs to be implemented carefully. One should use the function

  ```
  qxt = getRHqx(EstS)
  ```

  for this purpose.

3.4 Cairns-Blake-Dowd et al. Models

This section covers the models as defined by Cairns-Blake-Dowd (2006) and Cairns, Blake, Dowd, Coughlan, Epstein, Ong, Balevich(2009). All these models get again estimated via ML under the assumption that the model is Poisson.

3.4.1 CBD1 Model

$$\text{CBD1:} \quad \operatorname{logit} q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \varepsilon_{x,t}.$$

CBD1 Model

- Call: `EstS = CBDmodel1(yearStart,yearEnd,yearv,alpha,omega,agev,qx)`
- Description: Estimates the 2 parameters of the CBD1 model.
- Inputs: same as for RHmodel.
- Outputs:
  - `EstS.model` - name of model estimated
  - `EstS.alpha` - youngest age involved
  - `EstS.omega` - oldest age could be 85 or 90
  - `EstS.yearStart` - first year of data to be used in estimation
  - `EstS.yearEnd` - last year for the estimation
  - `EstS.nbYrEst` - returns the number of years over which the estimation was done
  - `EstS.nbAges` - number of ages for which the estimation was made
  - `EstS.tEst` - time required for the estimation
  - `EstS.dist` - the sum of absolute value of difference of consecutive vectors produced by the iteration involved in the ML estimation. This number should be smaller than 10^{-6}
  - `EstS.iterCtr` - a counter that gets incremented in each iteration. This number should be smaller than 10'000 for a correct convergence. In the code, if necessary, this number can be increased.
  - `EstS.k1` : $\kappa_t^{(1)}$ parameter estimates
  - `EstS.k2` : $\kappa_t^{(2)}$ parameter estimates
  - `EstS.resids` - residuals defined as difference between actual mortality and estimated mortality for the block of years and ages considered

3.4.2 CBD2 Model

$$\text{CBD2:} \quad \operatorname{logit} q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)} + \varepsilon_{x,t}.$$

CBD2 Model

- Call: `EstS = CBDmodel2(yearStart,yearEnd,yearv,alpha,omega,agev,qx)`
- Description: Estimates the 3 parameters of the CBD2 model.
- Inputs: Same as for CBDmodel1.
- Outputs: Similar to the ones of CBDmodel1 except that parameter estimates are:
  - `EstS.k1` : $\kappa_t^{(1)}$ parameter estimates
  - `EstS.k2` : $\kappa_t^{(2)}$ parameter estimates
  - `EstS.g3` : $\gamma_{t-x}^{(3)}$ parameter estimates

3.4.3 CBD3 Model

$$\text{CBD3:} \quad \text{logit}\, q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}\left((x-\bar{x})^2 - \hat{\sigma}_x^2\right) + \gamma_{t-x}^{(4)} + \varepsilon_{x,t}.$$

CBD3 Model

- Call: `EstS = CBDmodel3(yearStart,yearEnd,yearv,alpha,omega,agev,qx)`
- Description: Estimates the 4 parameters of the CBD3 model.
- Inputs: Same as for CBD1model1.
- Outputs: Similar to the ones of CBDmodel1 except that parameter estimates are:
  - `EstS.k1` : $\kappa_t^{(1)}$ parameter estimates
  - `EstS.k2` : $\kappa_t^{(2)}$ parameter estimates
  - `EstS.k3` : $\kappa_t^{(3)}$ parameter estimates
  - `EstS.g4` : $\gamma_{t-x}^{(4)}$ parameter estimates

4 License and final comments

License: This toolbox is distributed under a GNU General Public License as published by the Free Software Foundation, version 3. This means essentially that this program is distributed without any warranty. The user remains responsible for the correctness of the estimations. You may also distribute and modify all the stuff in here as long it remains free. Further information may be found in the file `License.txt`.

Bugs: Even though we spent quite some time on debugging we cannot guarantee that all functions are error free. We welcome bug reports and useful feedback about the toolbox. Concerning the former, it should include the command that produced the errors, a description of the data used (a zipped .MAT file with the data) and the version of MATLAB. All codes have been successfully run under MATLAB version 2015a and 2012b.

Citation: Credit may be given to this toolbox as: Jondeau, E., M. Rockinger, and H. Hristov, 2015, Matlab Longevity Toolbox, University Lausanne, IBF.

References

[1] Alho, J.M., 2000. Discussion of Lee 2000. North American Actuarial Journal 4, 91–93.

[2] Jondeau, E. and M. Rockinger, 2015, Backtesting Longevity Models: An International Perspective, HEC Lausanne Working paper. Also available on SSRN.

[3] Cairns, A.J.G., Blake, D., Dowd, K., 2006. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. Journal of Risk and Insurance 73(4), 687–718.

[4] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., Balevich, I., 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North American Actuarial Journal 13(1), 1–35.

[5] Carter, L.R., Lee, R.D., 1992. Modeling and forecasting US sex differentials in mortality. International Journal of Forecasting 8, 393–411.

[6] Heligman, L., Pollard J.H., 1980. The age pattern of mortality, Journal of the Institute of Actuaries, 49–80.

[7] Human Mortality Database, 2015. University of California, Berkeley (USA), and Max Plank Institute for Demographic Research (Germany). URL Available at www.mortality.org or www.humanmortality.de.

[8] Lee, R.D., Carter, L., 1992. Modelling and forecasting the time series of US mortality. Journal of the American Statistical Association 87(419), 659–671.

[9] Lee, R.D., 2000. The Lee-Carter method of forecasting mortality, with various extensions and applications. North American Actuarial Journal 4, 80–93.

[10] McCullagh, P., Nelder, J.A., 1989. Generalized linear models, 2nd Ed.. Monographs on Statistics and Applied Probability 37, Chapman and Hall.

[11] Renshaw, A., Haberman, S., 2006. A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insurance: Mathematics and Economics 38, 556–570.