K-Lee-Carters

Samuel Piveteau

SCOR, ISFA

November 17, 2020

Lee-Carter model I

The Lee-Carter model (1992) is the reference stochastic mortality model due to its accuracy and ease of use.

Many extensions have been derived from it: Lee-Carter with cohorts effects, age-rotation model, state-space Lee-Carter, CoDa Lee-Carter, P-splines Lee-Carter, etc

The purpose of this presentation is to introduce another extension of the Lee-Carter model named the K-Lee-Carters.

Lee-Carter model II

Let $D_{x,t}$ be the death count variable, $E_{x,t}$ be the exposure at risk and $\mu_{x,t}$ be the force of mortality, each considered at age x and period t.

Following the standard Poisson assumption : $\mu_{x,t} = \frac{D_{x,t}}{E_{x,t}}$

$$\log(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$

with $(\alpha_x)_x$ are the age-level parameters, $(\kappa_t)_t$ are the global trend parameters, $(\beta_x)_x$ are the age-sensitivity of the log-mortality to the variation of κ_t and $(\epsilon_{x,t})_{x,t}$ are the error-terms which are assumed iid.

Lee-Carter applied on a unique population

Usually, the Lee-Carter model is applied on the whole age range of a given population.

In some applications, the results are not accurate enough for defined ages. The hypothesis is that a unique time series parameter may be not enough to drive the whole mortality.

Some authors suggested to divide the age range into several intervals and apply a LC on each subdivided population to improve the fitting.

Two sex Lee-Carter ?

A common approach when dealing with two sexes mortality consists in splitting the mortality regarding the sex before applying the LC model.

If applied on the divided bases, the accuracy is improved. The downside of this practice is the dependency loss between the trajectories of Female and Male forces of mortality.

We suggest in this presentation to give up the distinction between Female and Male for all ages and to look for an optimal split.

Causes of death

The same question arises when considering the causes of death. Several causes of death trajectories are seen as dependant because they share common risk factors.

It is complex to identify the ages for which two causes of death share common trends.

Due to this complexity, the common practice is to apply separately the Lee-Carter model on each cause of death.

Unified framework I

Mortality data can be described using two dimensions:

- periods t on which the mortality is observed,
- characteristics of the mortality time series: age, sex, localization, cause of death, etc

A mortality database is described by a matrix Y, where the rows represent the characteristics of the time series and the columns represent the observed periods.

Unified framework II

We consider the set of characteristics $\mathbb{S}$.

- $\mathbb{S} =$ Ages
- $\mathbb{S} =$ Ages $\times$ Sex
- $\mathbb{S} =$ Ages $\times$ cause of death

For each characteristic $s \in \mathbb{S}$, we associate a mortality $\mu_{s,t}$ With these notations, we can describe the mortality matrix Y as $Y = (\log(\mu_{s,t}))_{s \in \mathbb{S}, t}$.

Given a constant K, the purpose is to obtain the K groups of characteristics giving the best global fitting when applying the LC on each group separately.

K-centroids algorithm I

The K-centroids algorithm has been developed in Macqueen (1967)

The principle of this clustering algorithm is to identify similarities between individuals in a database and to group them into homogeneous classes.

We denote by $(y_s)_s$ the individuals, the groups are named $C_1, ..., C_K$, the centroids $(c_1, ..., c_K)$ and the distance $d(.,.)$. The purpose is to minimize the following Cost Function:

$$L((C_k)_{k=1,...,K}, (c_k)_{k=1,...,K}) = \sum_{k=1}^{K} \sum_{s \in C_k} d(y_s, c_k)$$

K-centroids algorithm II

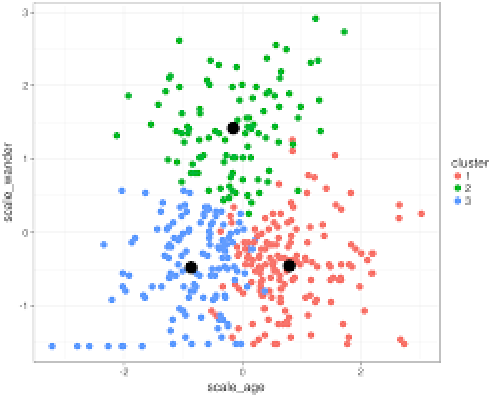


Figure: An illustration of the K-centroids algorithm with euclidean distance

K-centroids algorithm III

- Step 0: distribute the individuals into K groups $(C_1, ..., C_K)$ giving a specific rule.
- Step 1 : for each group C_k, the centroids c_k is obtained as $c_k = \mathrm{argmin}_c \sum_{s \in C_k} d(y_s, c)$.
- Step 2 : for an individual y_s, $k = \arg\min_i d(y_s, c_i)$ and y_s is assigned to the group C_k.
- Step 3 : if no change has occurred, the algorithm interrupts. Else we return to the step 1.

K-Lee-Carters Algorithm I

We propose a similar algorithm for the Lee-Carter Model. The purpose is to compute the parameters $(C_k)_k, (\beta_s)_s, (\alpha_s)_s, (\kappa_t^k)_{t,k}$ minimizing the following Cost Function:

$$L\left((C_k)_k, (\beta_s)_s, (\alpha_s)_s, (\kappa_t^k)_{t,k}\right) = \sum_{k=1}^{K} \sum_{s \in C_k} \sum_{t} (\log(\mu_{s,t}) - \alpha_s - \beta_s \kappa_t^k)^2$$

- Step 0: distribute the individuals into K groups $(C_1, ..., C_K)$ giving a specific rule

K-Lee-Carters Algorithm II

- Step 1 : for each group C_k, the model parameters are computed : $((\widehat{\alpha}_s)_{s\in C_k}, (\widehat{\beta}_s)_{s\in C_k}, (\widehat{\kappa}^k_t)_t) = \arg\min \sum_{s\in C_k} \sum_t (\log(\mu_{s,t}) - \alpha_s - \beta_s \kappa^k_t)^2$.
- Step 2 : let $\log(\mu_{s,t})$ be a time series, we compute for each κ^k, the coefficients $\widetilde{\alpha}^k_s$, $\widetilde{\beta}^k_s$ which minimize $L^k_s\left(\alpha^k_s, \beta^k_s\right) = \sum_t (\log(\mu_{s,t}) - \alpha^k_s - \beta^k_s \widehat{\kappa}^k_t)^2$.
- Step 3 : the time series associated to the characteristic s experiences a class switch for the group k which is the most suitable $C(s) = C_k$ with $k = \arg\min_i L^i_s\left(\alpha^{i,\prime}_s, \beta^{i,\prime}_s\right)$.
- Step 4 : if no change has occurred, the algorithm interrupts. Else we return to the step 1.

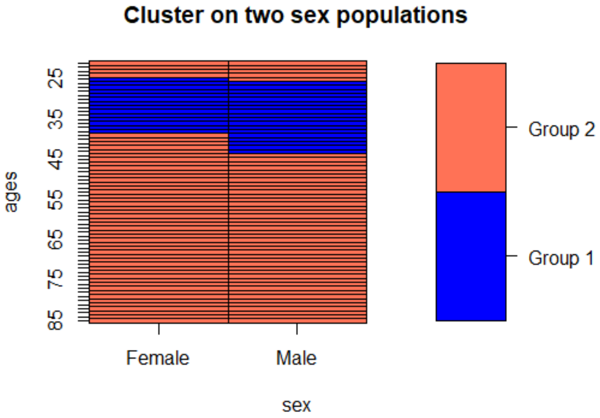
The link between K-Lee-Carters and K-centroids

The K-Lee-Carters seems to be more complex than the K-centroids due to the number of parameters.

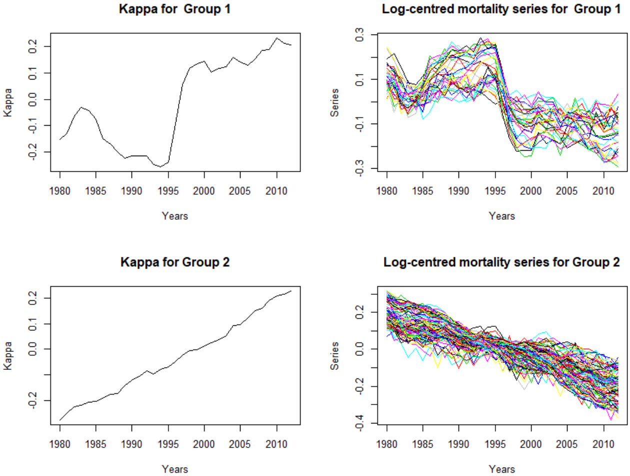
We have demonstrated that the K-Lee-Carters can be considered as a K-centroids with the distance $d(x, y) = Var(x)Var(y) - Cov(x, y)^2$.

This relation between the two algorithms allows us to use available packages (see for instance F.Lish package (2006)) and to reduce the time computation.

Two sex I


Cluster on two sex populations
ages
25
35
45
55
65
75
85
Female
Male
sex
Group 2
Group 1

Two sex II


Kappa for Group 1
Kappa
Years
Log-centred mortality series for Group 1
Series
Years
Kappa for Group 2
Kappa
Years
Log-centred mortality series for Group 2
Series
Years

Two sex III: in sample errors

Cluster	MSE	MAE	MAPE
Reference cluster	0.002425878	0.03715817	0.007663029
K-LC cluster	0.001537885	0.03022469	0.006266054

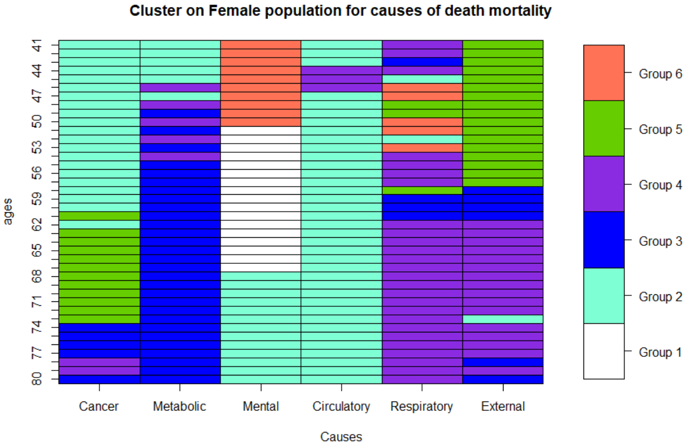
Table: In-sample errors

Two sex IV: out of sample errors

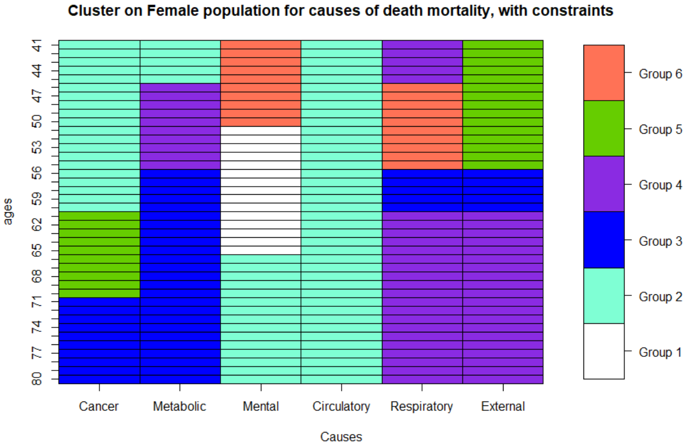
Cluster	MSE	MAE	MAPE
Reference cluster	0.0137372	0.08237318	0.0151057
K-LC cluster	0.01060344	0.07706309	0.01464475

Table: Out-sample errors

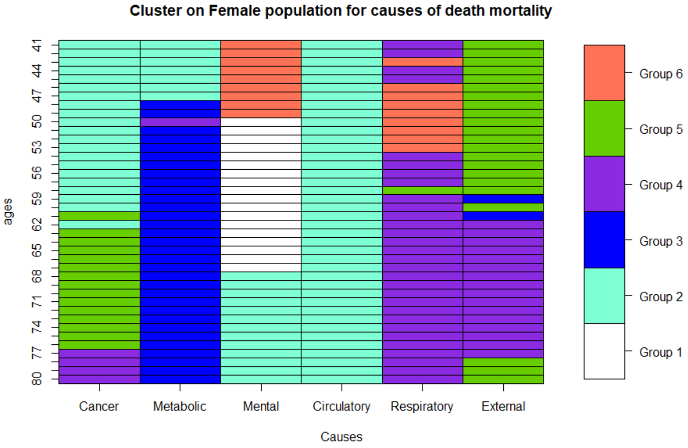
Causes of death I: standard approach



Causes of Death II: constrained approach



Causes of Death III: penalized approach



Causes of death IV: trends by group

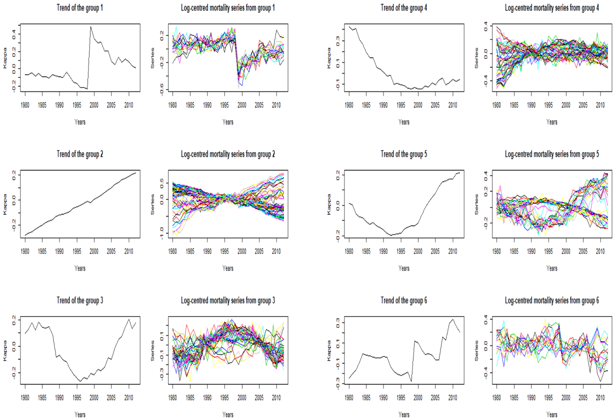


Figure: Trends for each group

Causes of death V: in sample errors

Cluster	MSE	MAE	MAPE
Reference cluster	0.005571466	0.05283739	0.007000738
K-LC	0.003887775	0.0465221	0.006256402
K-LC with constraints	0.00401991	0.04733094	0.006379771
K-LC penalized	0.00398385	0.04723434	0.006371326

Table: Errors in the calibration sample

Causes of death VI: out of sample errors

Cluster	MSE	MAE	MAPE
Reference cluster	0.0169789	0.1029326	0.01419595
K-LC cluster	0.01310745	0.08807986	0.01228526
K-LC with constraints	0.01287824	0.08643274	0.01208211
K-LC penalized	0.01248059	0.08665662	0.01217052

Table: Out-sample errors

Discussion

- The method allows to find the optimal groups regarding the Lee-Carter model. In the paper, I showed that the forecast obtained are also better when considering K-LC groups than for arbitrary groups.
- The article explores some constraint specification allowing to reduce the volatility in the clustering.
- It can be applied in conjunction with the BIC criterion to find how many groups are required.
- The algorithm can be applied on a large basis, but requires to be programmed in a low-level language.

Bibliography

- John Bongaarts. Trends in senescent life expectancy. Population Studies, 63(3):203 213, 2009. PMID: 19851933.
- Ronald Lee. The lee-carter method for forecasting mortality, with various extensions and applications. North American Actuarial Journal, 4, 01 2000.
- Ronald D. Lee and Lawrence R. Carter. Modeling and forecasting u. s. mortality. Journal of the American Statistical Association, 87(419):659671, 1992.
- Friedrich Leisch. A toolbox for k-centroids cluster analysis. Computational Statistics and Data Analysis, 51(2):526544, 2006.

Bibliography

- Friedrich Leisch and Bettina Grün. Extending standard cluster algorithms to allow for group constraints. 02 2020.
- Nan Li, Ronald Lee, and Patrick Gerland. Extending the lee-carter method to model the rotation of age patterns of mortality decline for long-term projections. Demography, 50(6):20372051, 2013.
- A.E. Renshaw and Steven Haberman. A cohort-based extension to the lee-carter model for mortality reduction factors. Insurance: Mathematics and Economics, 38:556570, 06 2006.