Compositional Data Analysis for Forecasting Mortality

Samuel Piveteau, Julien Tomas
spiveteau@scor.com

SCOR, ISFA

21 september 2018

Table of Contents

The origin of the model

Définition

The Basis Risk is the risk that the mortality level or trend differ between the population of interest and the population used to compute the mortality coefficients.

The model proposed in this presentation has two main purposes :

- Taking into account the basis risk.
- Including medical expert opinions on the mortality by cause.

Some definitions

We denote

- X, T and CoD are respectively the age at death, the year of death and the cause of death random variables.
- X is discret and finite, i.e. $X \in \{1, 2, ..., \omega\}$
- We assume that variable CoD is an exhaustive variable, i.e. $\sum_i \mathbb{P}(CoD = i) = 1$, all sources of death are identified.
- $d^i_{x,t} = \mathbb{P}(x \leq X < x+1, CoD = i | T = t)$
- $\pi^i_t = \mathbb{P}(CoD = i | T = t)$
- $\delta^i_{x,t} = \mathbb{P}(x \leq X < x+1 | CoD = i, T = t)$
- $d^i_{x,t} = \pi^i_t \delta^i_{x,t}$

Interpretation of the indicators

A Mortality Improvement can come from two ways :

- A change in the causes of death proportions π_t^i (for instance a vaccin, sanitary changes, virus extermination)
- A change in the conditional age at death $\delta_{x,t}^i$ (Improvement in the treatment of the pathology)

We propose to model the evolution of the vectors $\pi_t = \{\pi_t^1, ..., \pi_t^D\}$ and $\delta_t^i = \{\delta_{1,t}^i, .., \delta_{\omega,t}^i\}$ which are compositional data.
In consequences, the modelisation is not straightforward and requires specific tools.

A brief introduction to Compositional Data

Définition

An element is called a composition if it provides proportion of a total. More formaly Compositional Data is a vector x of which the components are positive and sum up to defined constant, $x_i > 0$ and $\sum_{i=1}^{D} x_i = C$

Aitchison (1986) proposed to tackle the constraint issues by mapping the initial data in a less constrained space.

$$clr(x) = \left(log\left(\frac{x_1}{(\prod_i^D x_i)^{\frac{1}{D}}}\right), log\left(\frac{x_2}{(\prod_i^D x_i)^{\frac{1}{D}}}\right), ..., log\left(\frac{x_D}{(\prod_i^D x_i)^{\frac{1}{D}}}\right)\right)$$

$$alr(x) = \left(log\left(\frac{x_1}{x_D}\right), log\left(\frac{x_2}{x_D}\right), ..., log\left(\frac{x_{D-1}}{x_D}\right)\right)$$

$ilr(x) = A' clr(x)$ whith A an orthonomal base of vectors.

A Compositional mortality model

We propose the model :

$$clr(\pi_t^i) = \alpha_i + \Phi(i,t) + \epsilon_{i,t}^{\pi}$$

$$clr(\delta_{x,t}^i) = \alpha_x^i + \Phi_i(x,t) + \epsilon_{x,t}^{\delta^i}$$

$$\frac{d\mathbb{E}[\pi_t^i]}{dt} = \sum_{j=1}^{D} \left(\frac{\partial \Phi(i,t)}{\partial t} - \frac{\partial \Phi(j,t)}{\partial t} \right) \mathbb{E}[\pi_t^j]\mathbb{E}[\pi_t^i], \ \forall i \in \{1, ..., D\}$$

$$\frac{d\mathbb{E}[\delta_{x,t}^i]}{dt} = \sum_{y=1}^{\omega} \left(\frac{\partial \Phi_i(x,t)}{\partial t} - \frac{\partial \Phi_i(y,t)}{\partial t} \right) \mathbb{E}[\delta_{y,t}^i]\mathbb{E}[\delta_{x,t}^i], \forall x \in \{1, ..., \omega\}$$

Oeppen Model and interpretation of the parameters

Jim Oeppen approach

The J.Oeppen (2008) model is based on the Lee-Carter(1992) model, i.e. $clr(d_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$
For the Π_t modelisation, it implies that $\Phi(i,t) = \beta_i \kappa_t$.

For two given causes of death i and j, the difference between β_i and β_j determines the sense of the transfer between the two causes.

As in the Lee-Carter model, the κ_t is assumed to be a random walk in order to be forecasted.

Asymptotic results

- When $t \to +\infty$, $\pi_t^i \to 1$ if $\beta_i > \beta_j \ \forall j \neq i$. Only one cause of death is asymptotically observed.
- In consequence :

$$lim_{t\to+\infty} d_{x,t} = lim_{t\to+\infty} \mathbb{P}(x \leq X < x+1 | T = t) = lim_{t\to+\infty} \delta_{x,t}^{i^\star}$$

where $i^\star = argmax_j \beta_j$

- $lim_{t\to+\infty} \delta_{x,t}^i = \delta_{x_i^\star}$ where $x_i^\star = argmax_y \beta_y^i$
- We note the mortality improvement : $MI_{x,t} = 1 - \frac{q_{x,t+1}}{q_{x,t}}$, we observe an increasing MI, with asymptotic convergence :

$$lim_{t\to+\infty} MI_{x,t} = 1 - exp((\beta_x^i - \beta_{x_i^\star}^i)\Delta\kappa_t^i)$$

where $i = argmax_j \beta_j$

Model modification

Mortality Improvement control

In insurance companies, a very current assumption is about the long term mortality improvement level.

Avoiding mortality improvement to sky rocket requires modification on the model.

$\Delta\kappa_t^i \leftarrow \Delta\kappa_t^i w(t) + \Delta\kappa_t^{i,LT}(1 - w(t))$, where :

- $\kappa^{i,LT}$ is the long term trend for the conditional age at death for cause i
- $w : \mathbb{R}_+ \rightarrow [0,1[$ is a deacreasing function such that $w(0) = 1$ and $lim_{t \rightarrow +\infty} w(t) = 0$

By doing so, the level of mortality improvement is under control.

Stochastic terms of mortality

We now analyse the terms of error $\epsilon^{\pi}_{i,t}$ and $\epsilon^{\delta^i}_{x,t}$.

VARMA process is not suitable :

- The covariance structure of the error vectors is singular due to the clr transformation.
- The number of parameters generates overfitting due to the length of this time series considered here.

We propose order 1 univariate autoregressive process :

$$\epsilon^{\delta^i}_{x,t} = \rho_x \epsilon^{\delta^i}_{x,t-1} + \eta^{\delta^i}_{x,t}$$

$$\epsilon^{\pi}_{i,t} = \rho_i \epsilon^{\pi}_{i,t-1} + \eta^{\pi}_{i,t}$$

Where the processes η^{π}_t and $\eta^{\delta^i}_t$ are White Noises.

Proportions of death

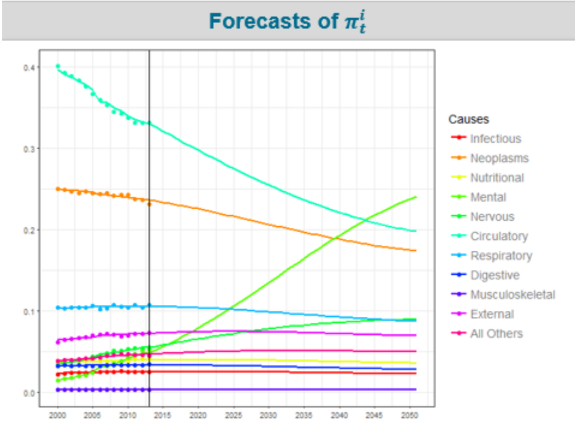

Figure – Proportions of death observed,fitted and forecasted for the male USA population from 2000 to 2013

Density forecasting

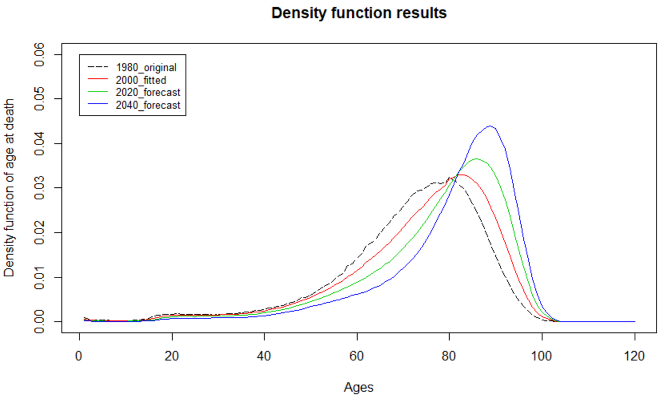

Figure – Evolution of the age at death density function of the male USA population

Mortality improvement

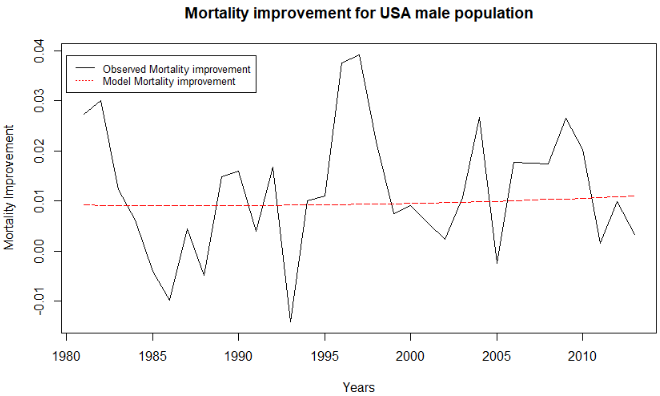

FIGURE – Evolution of the mean of mortality improvement (Experience VS Model) from 1980 to 2013 for male USA population

Stochastic error

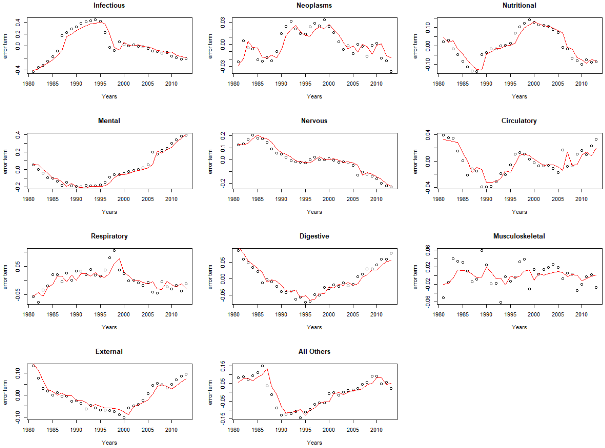

Figure – Stochastic error fitting with an autoregressive model