modeling analogies in nonlife and life insurance

Arthur Charpentier[1]

[1] Université Rennes 1 - CREM & École Polytechnique

arthur.charpentier@univ-rennes1.fr

http ://blogperso.univ-rennes1.fr/arthur.charpentier/index.php/

Reserving seminar, SCOR, May 2010

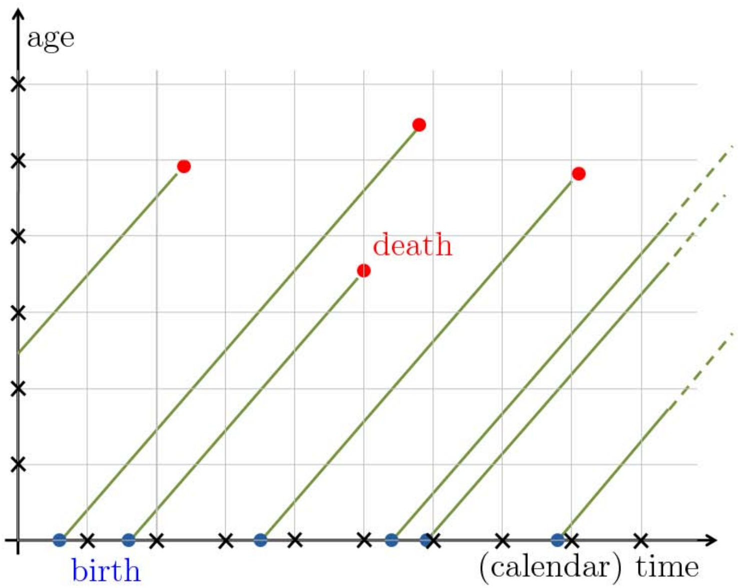
Agenda

- **Lexis diagam in life and nonlife insurance**
- **From Chain Ladder to the log Poisson model**
- **From Lee & Carter to the log Poisson model**
- **Generating scenarios and outliers detection**

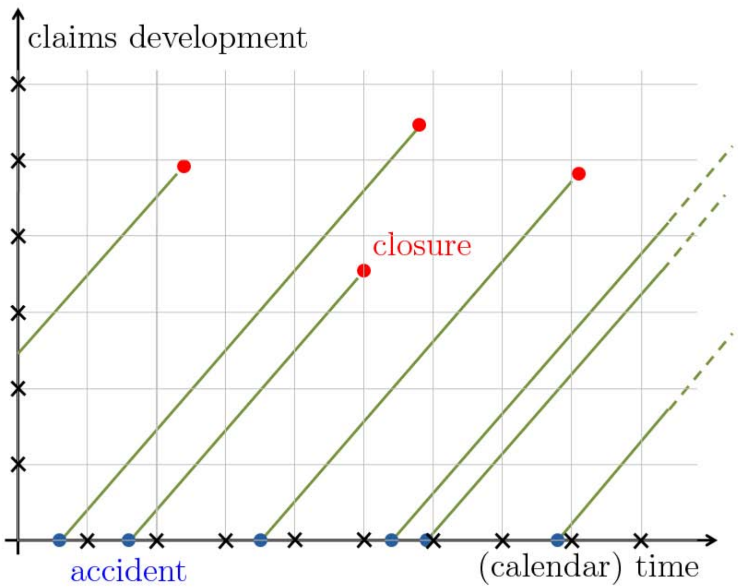
Lexis diagram in insurance

Lexis diagrams have been designed to visualize dynamics of life among several individuals, but can be used also to follow claims'life dynamics, from the occurrence until closure,

in life insurance



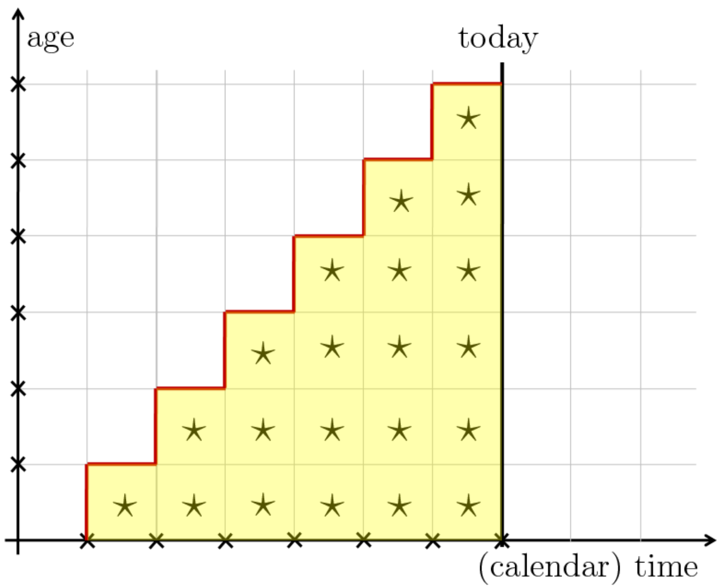
in nonlife insurance



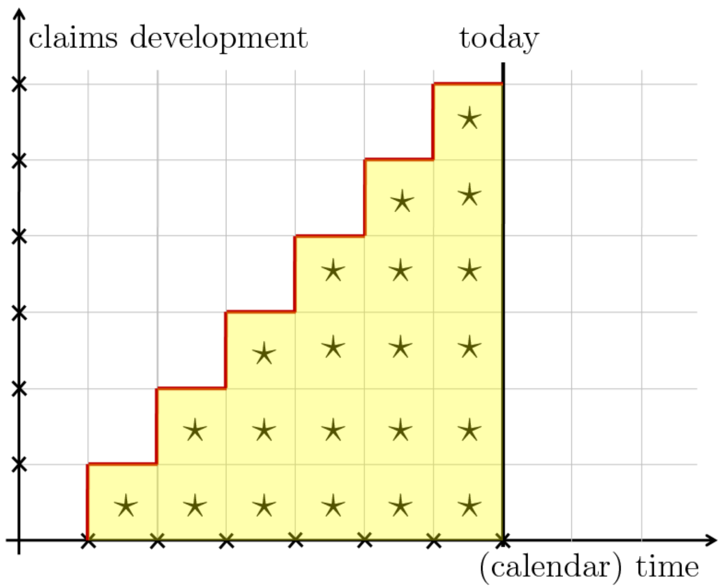
Lexis diagram in insurance

but usually we do not work on continuous time individual observations (individuals or claims) : we summarized information per year

in life insurance



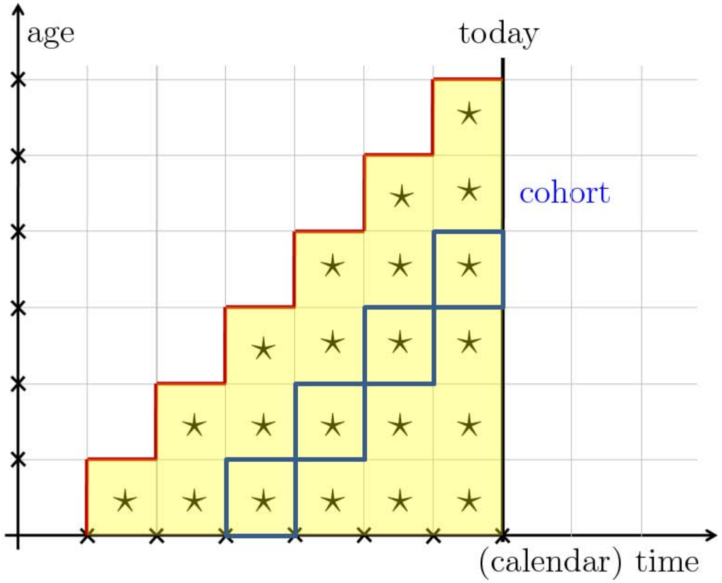
in nonlife insurance



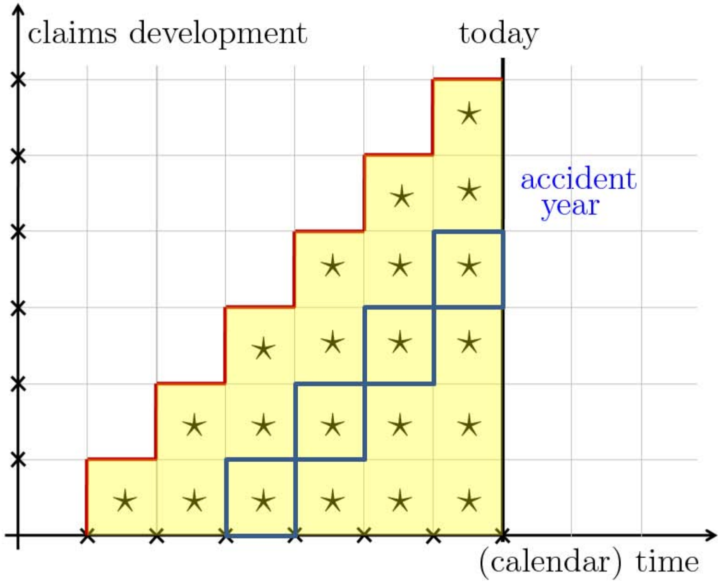
Lexis diagram in insurance

individual lives or claims can also be followed looking at diagonals,

in life insurance



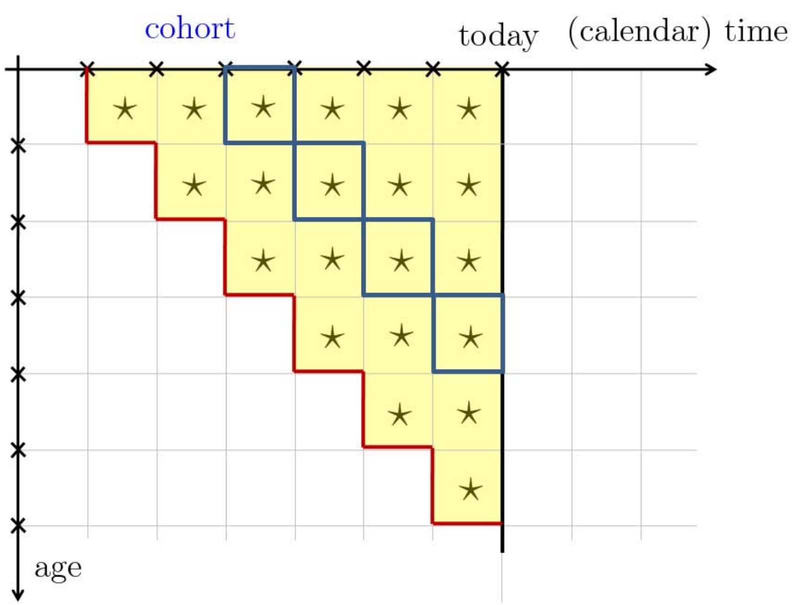
in nonlife insurance



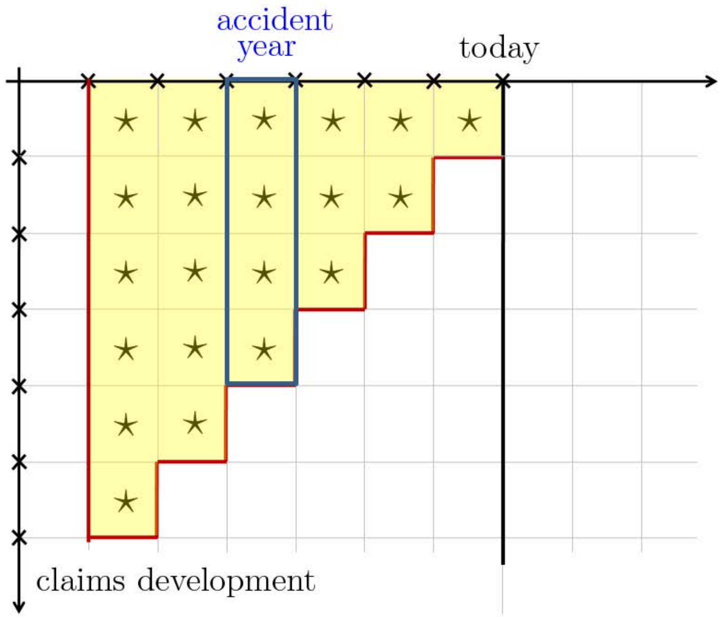
Lexis diagram in insurance

and usually, in nonlife insurance, instead of looking at (calendar) time, we follow observations per year of birth, or year of occurrence

in life insurance



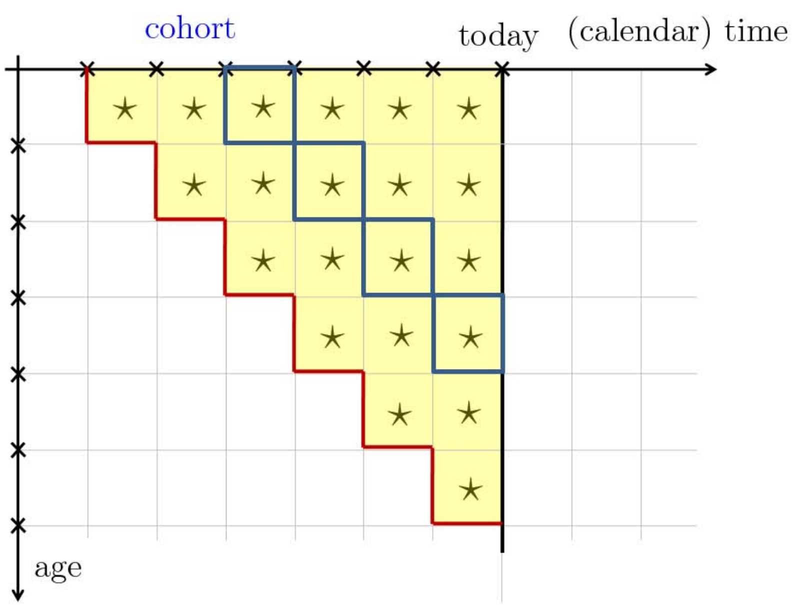
in nonlife insurance



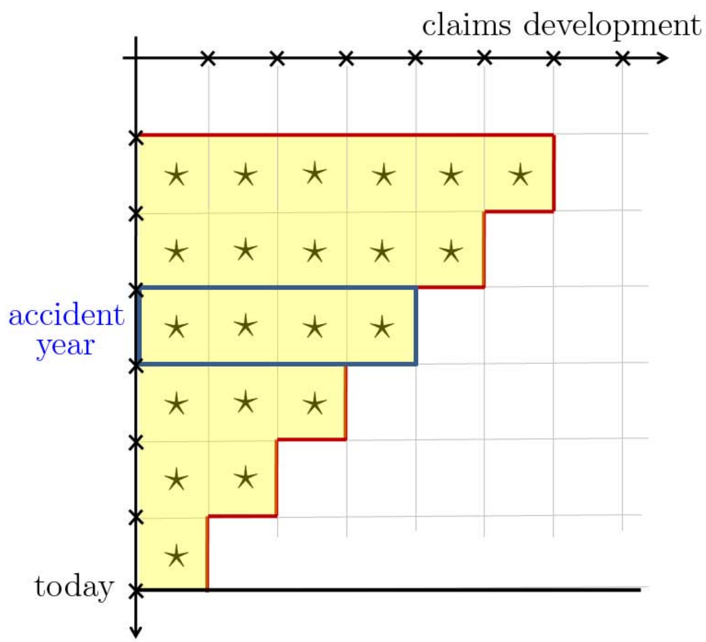
Lexis diagram in insurance

and finally, recall that in standard models in nonlife insurance, we look at the transposed triangle

in life insurance



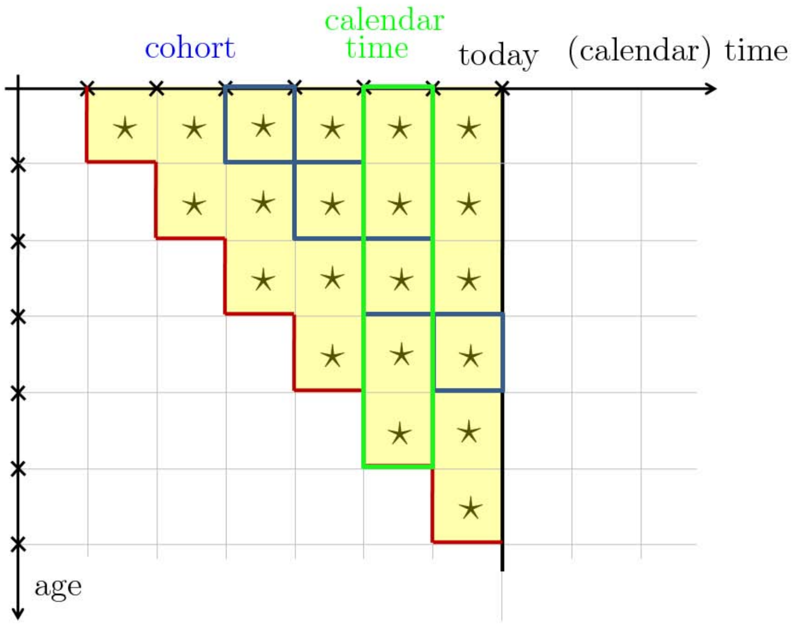
in nonlife insurance



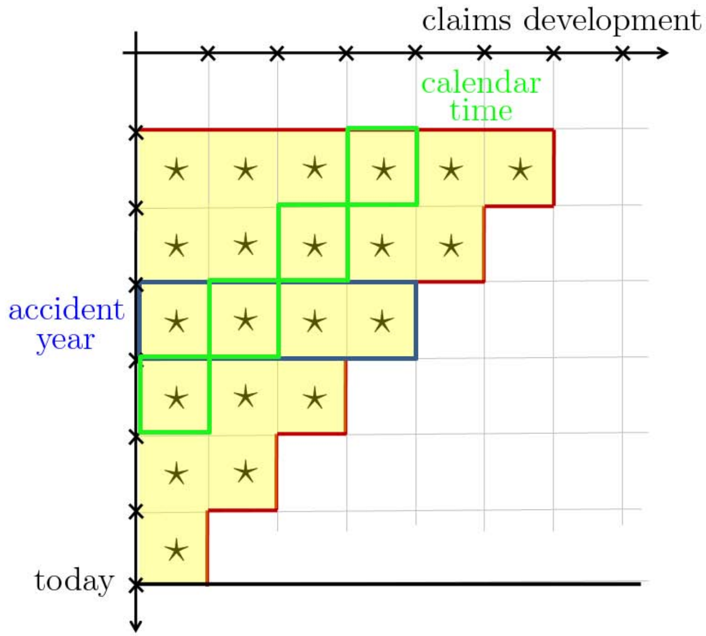
Lexis diagram in insurance

note that whatever the way we look at triangles, there are still three dimensions, year of occurrence or birth, age or development and calendar time,

in life insurance



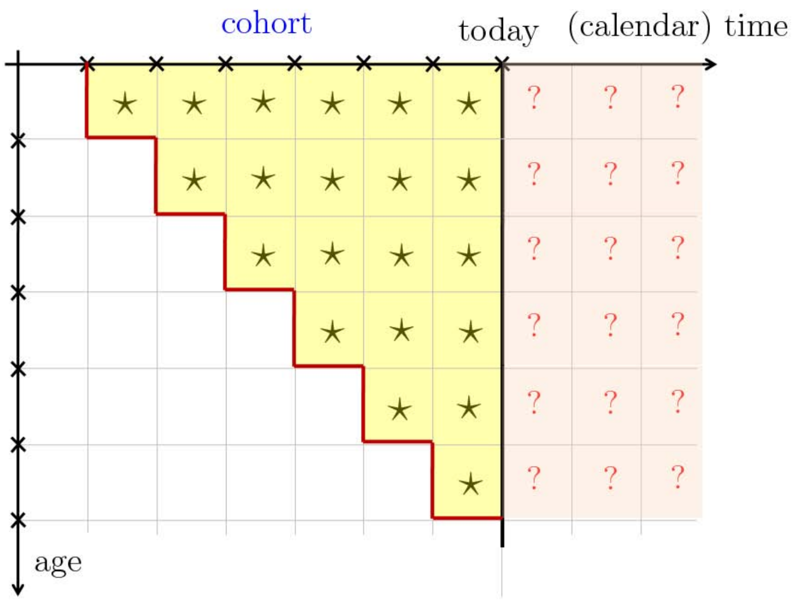
in nonlife insurance



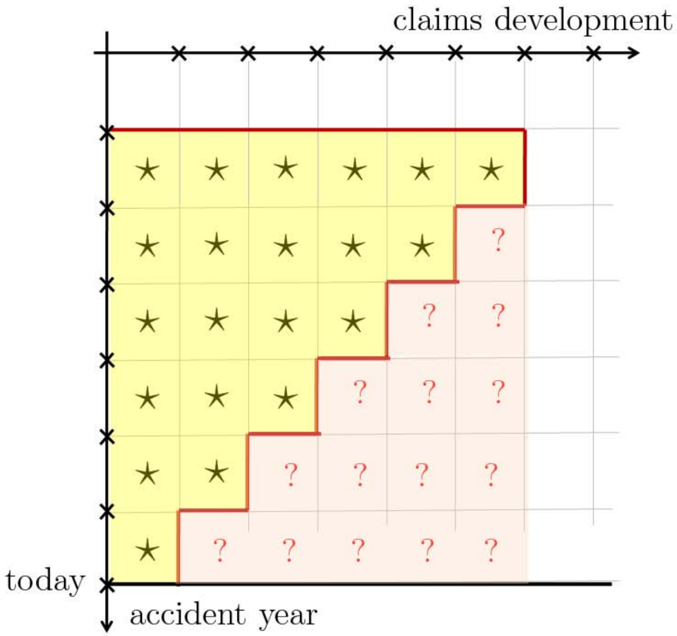
Lexis diagram in insurance

and in both cases, we want to answer a prediction question...

in life insurance



in nonlife insurance



What can be modeled in those triangles ?

In life insurance,

- $L_{i,j}$, number of survivors born year i, still alive at age j
- $D_{i,j}$, number of deaths of individuals born year i, at age j, $D_{i,j} = L_{i,j} - L_{i,j-1}$,
- $E_{i,j}$, exposure, i.e. i, still alive at age j

(if we cannot work on cohorts, exposure is needed).

In life insurance,

- $C_{i,j}$, total claims payments for claims occurred year i, seen after j years,
- $Y_{i,j}$, incremental payments for claims occurred year i, $Y_{i,j} = C_{i,j} - C_{i,j-1}$,
- $N_{i,j}$, total number of claims occurred year i, seen after j years,

The log-Poisson regression model

Hachemeister (1975), Kremer (1985) and finally Mack (1991) suggested a log-Poisson regression on incremental payments, with two factors, the year of occurence and the year of development

$$Y_{i,j} \sim \mathcal{P}(\mu_{i,j}) \text{ where } \mu_{i,j} = \exp[\alpha_i + \beta_j].$$

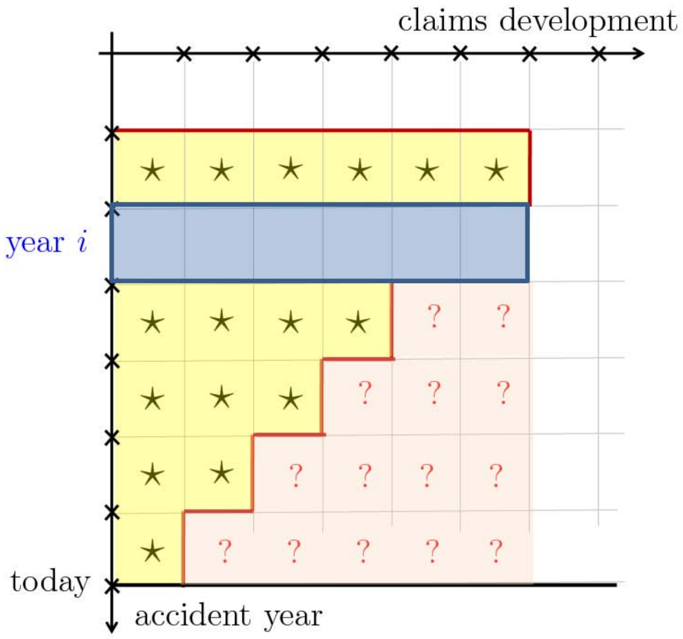
It is then *extremely* simple to calibrate the model.

The log-Poisson regression model

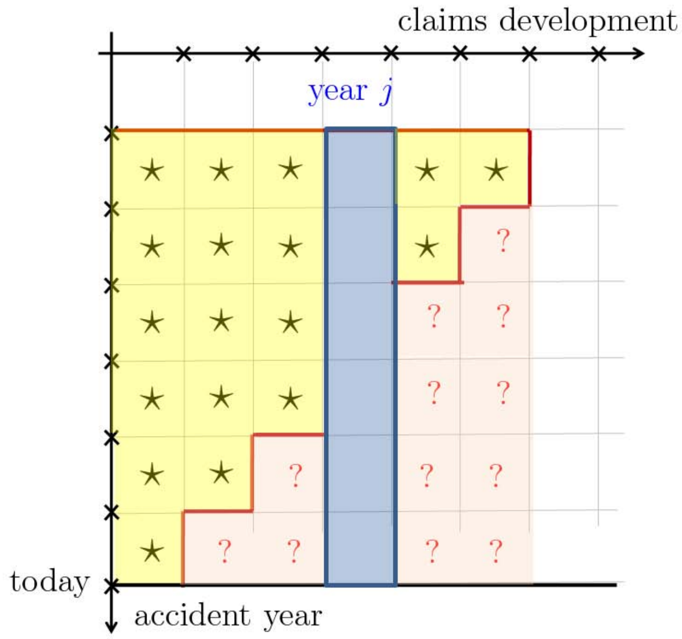
Assume that

$$Y_{i,j} \sim \mathcal{P}(\mu_{i,j}) \text{ where } \mu_{i,j} = \exp[\alpha_i + \beta_j].$$

the occurrence factor α_i



the development factor β_j



The log-Poisson regression model

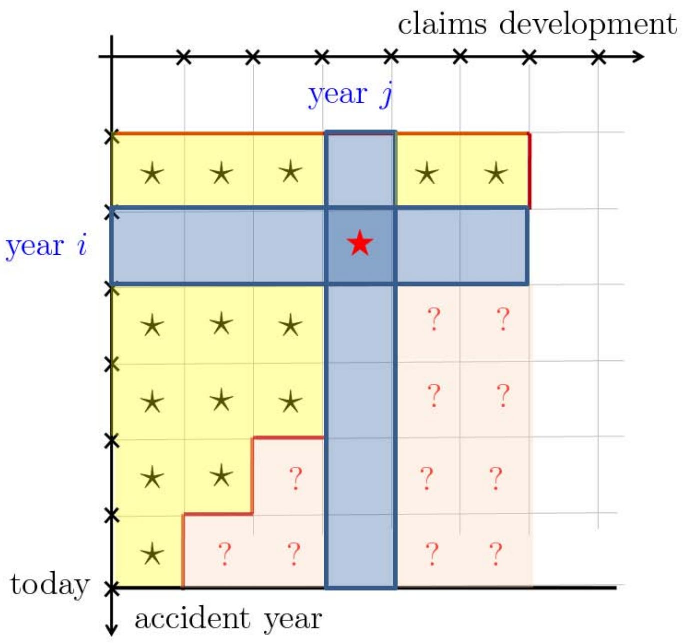
Assume that

$$Y_{i,j} \sim \mathcal{P}(\mu_{i,j}) \text{ where } \mu_{i,j} = \exp[\alpha_i + \beta_j].$$

It is then extremely simple to calibrate the model,

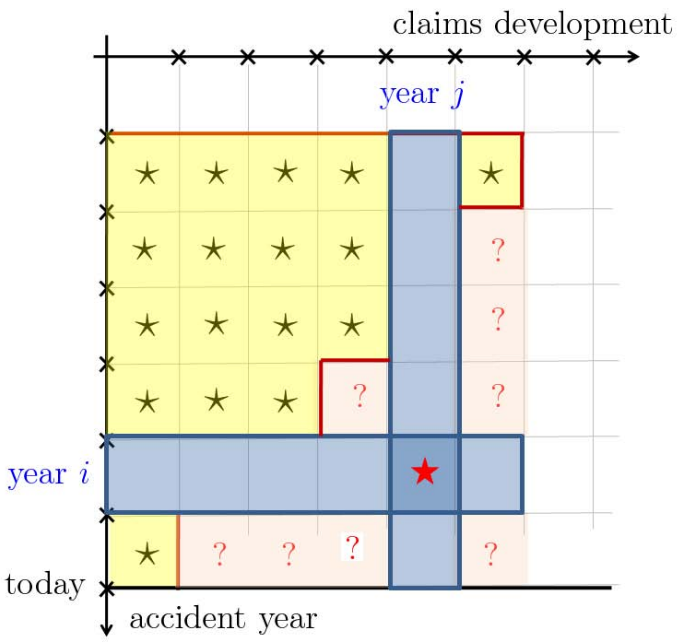
$$\widehat{Y}_{i,j} = \exp[\widehat{\alpha}_i + \widehat{\beta}_j]$$

on past observations



$$\widehat{Y}_{i,j} = \exp[\widehat{\alpha}_i + \widehat{\beta}_j]$$

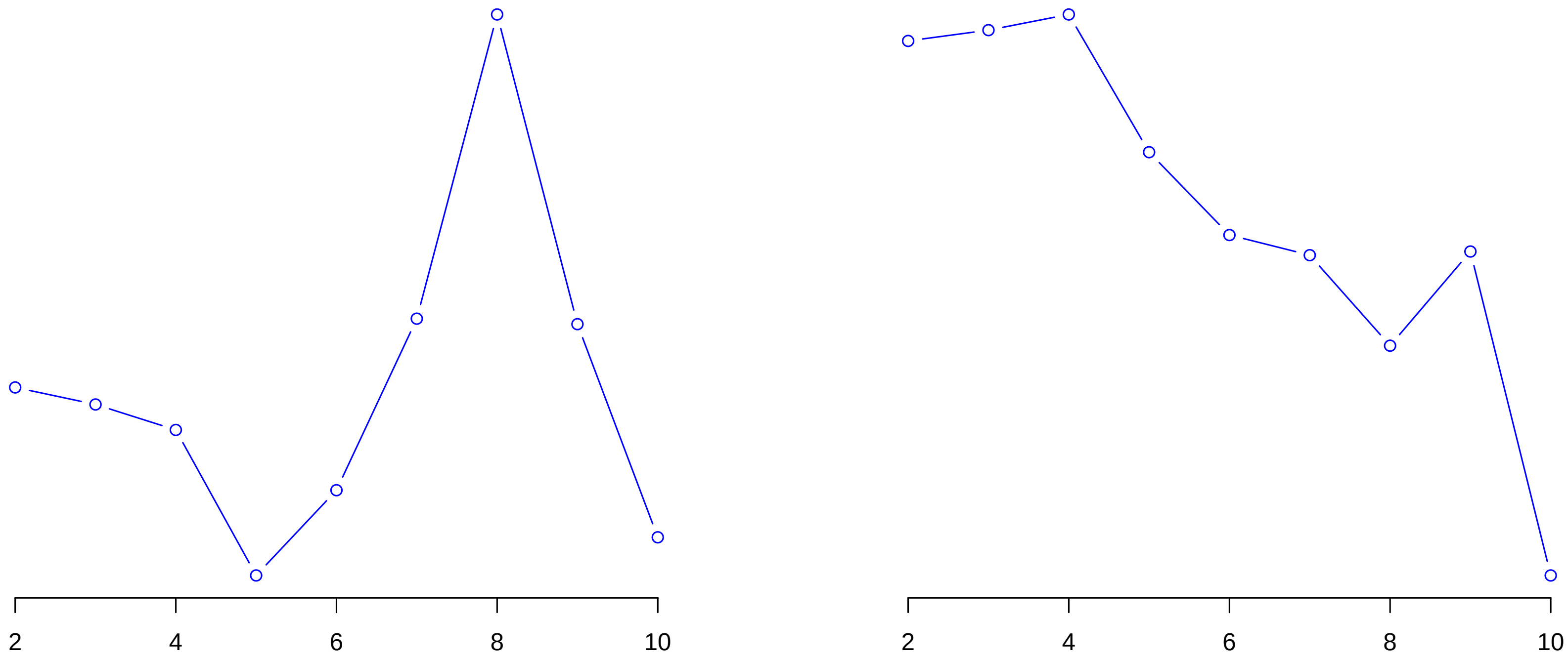
on the future



The log-Poisson regression model

Assume that

$$Y_{i,j} \sim \mathcal{P}(\mu_{i,j}) \text{ where } \mu_{i,j} = \exp[\alpha_i + \beta_j].$$



Additional remarks

Since we consider a Poisson model, then $\mathbb{E}(Y|\boldsymbol{\alpha},\boldsymbol{\beta}) = \text{Var}(Y|\boldsymbol{\alpha},\boldsymbol{\beta})$.

Further, the idea of using *only* two factors can be found in de Vylder (1978), i.e. $Y_{i,j} = r_i \cdot c_j$. But other factor based models have been considered e.g. Taylor (1977), $Y_{i,j} = d_{i+j} \cdot c_j$ where d_{i+j} denotes a calendar factor, interpreted as an inflation effect.

Quantifying uncertainty in a stochastic model

The goal in claims reserving is to quantify $\mathbb{E}\left([\widehat{R}-R]^2 \middle| \mathcal{F}_{i+j}\right)$ where

$$R = \sum_{i,j,i+j>t} Y_{i,j} \text{ is the amount of reserves.}$$

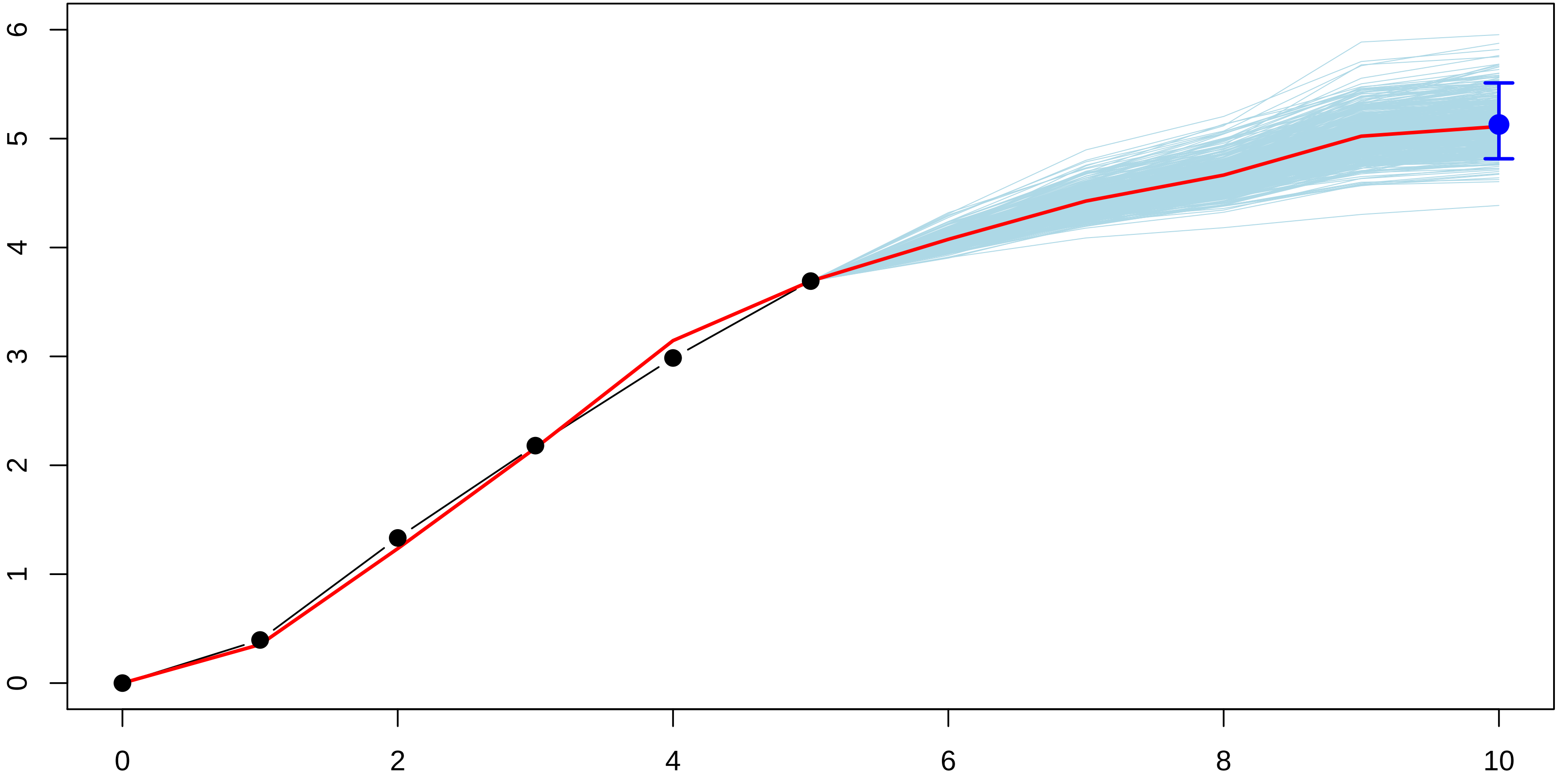
Classically, bootstrap techniques are considered, i.e. generate pseudo-triangles,

$$Y^\star_{i,j} = \widehat{Y}_{i,j} + \sqrt{\widehat{Y}_{i,j}} \cdot \widehat{\varepsilon}_{i^\star,j^\star}$$

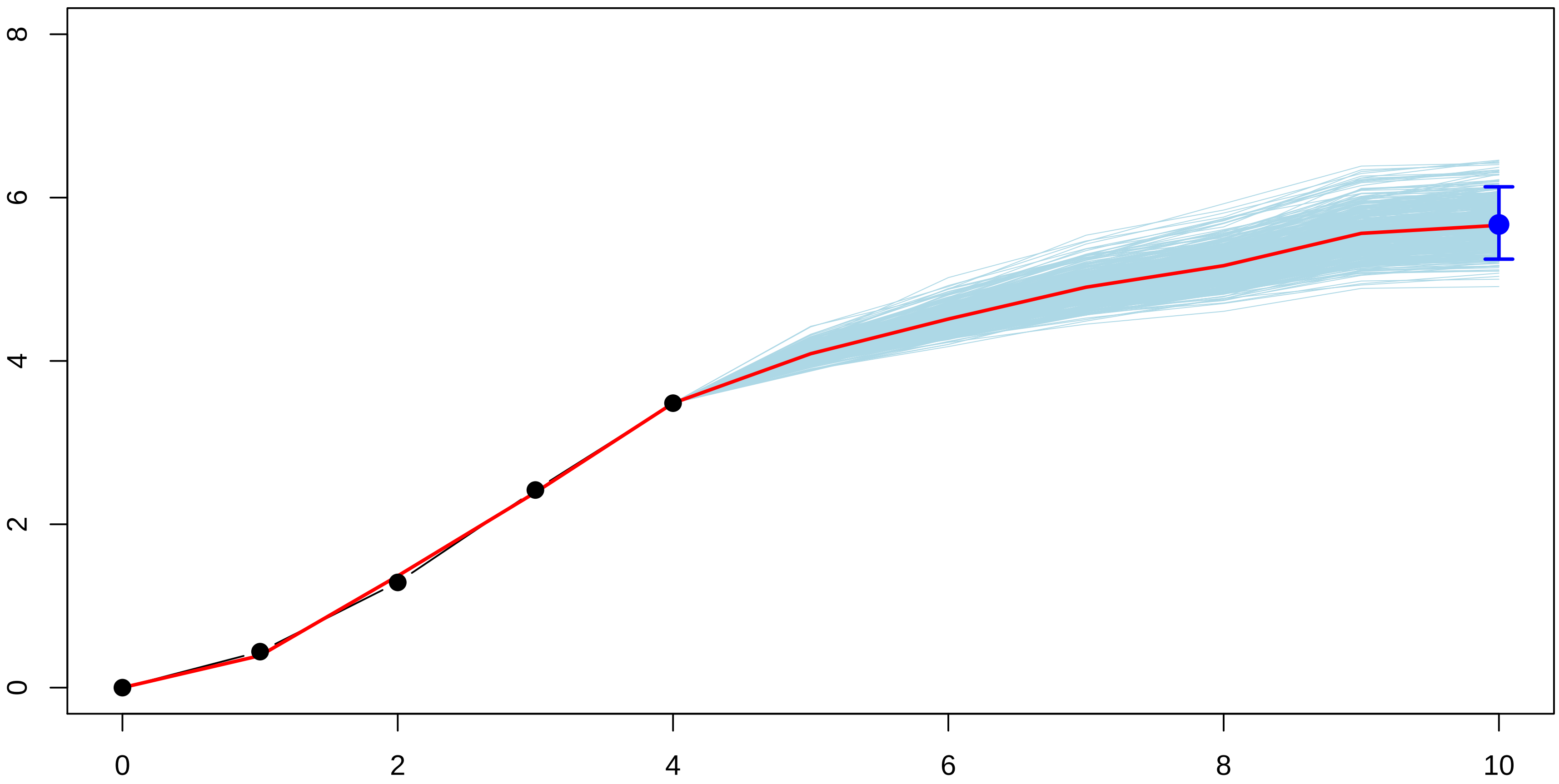
then fit a log-Poisson model $Y^\star_{i,j} \sim \mathcal{P}(\mu^\star_{i,j})$ where $\mu^\star_{i,j} = \exp[\alpha^\star_i + \beta^\star_j]$.

Then generate $\mathcal{P}(\mu^\star_{i,j})$ for future payments, i.e. $i+j>t$.

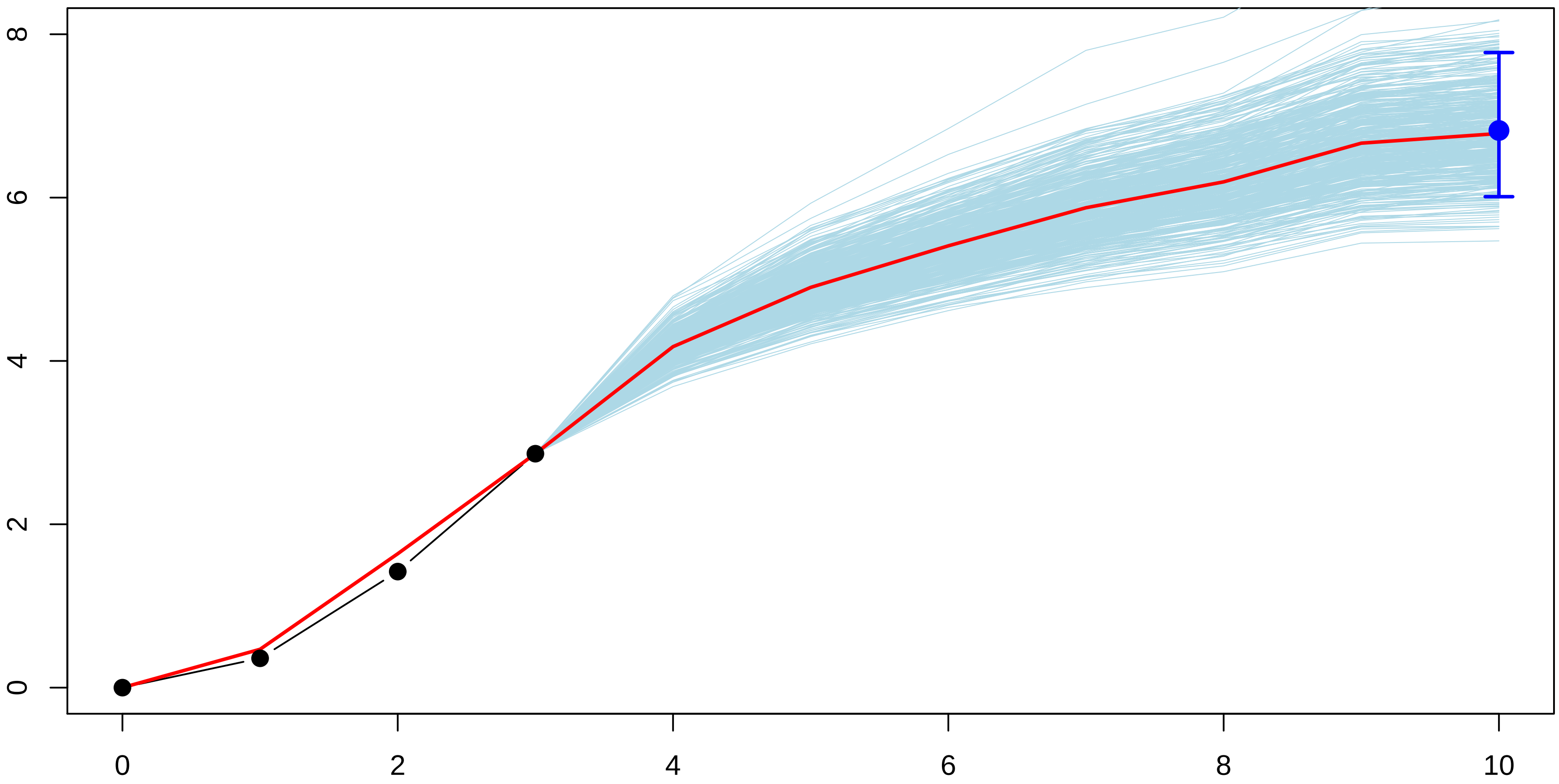
Bootstrap and GLM log-Poisson in triangles



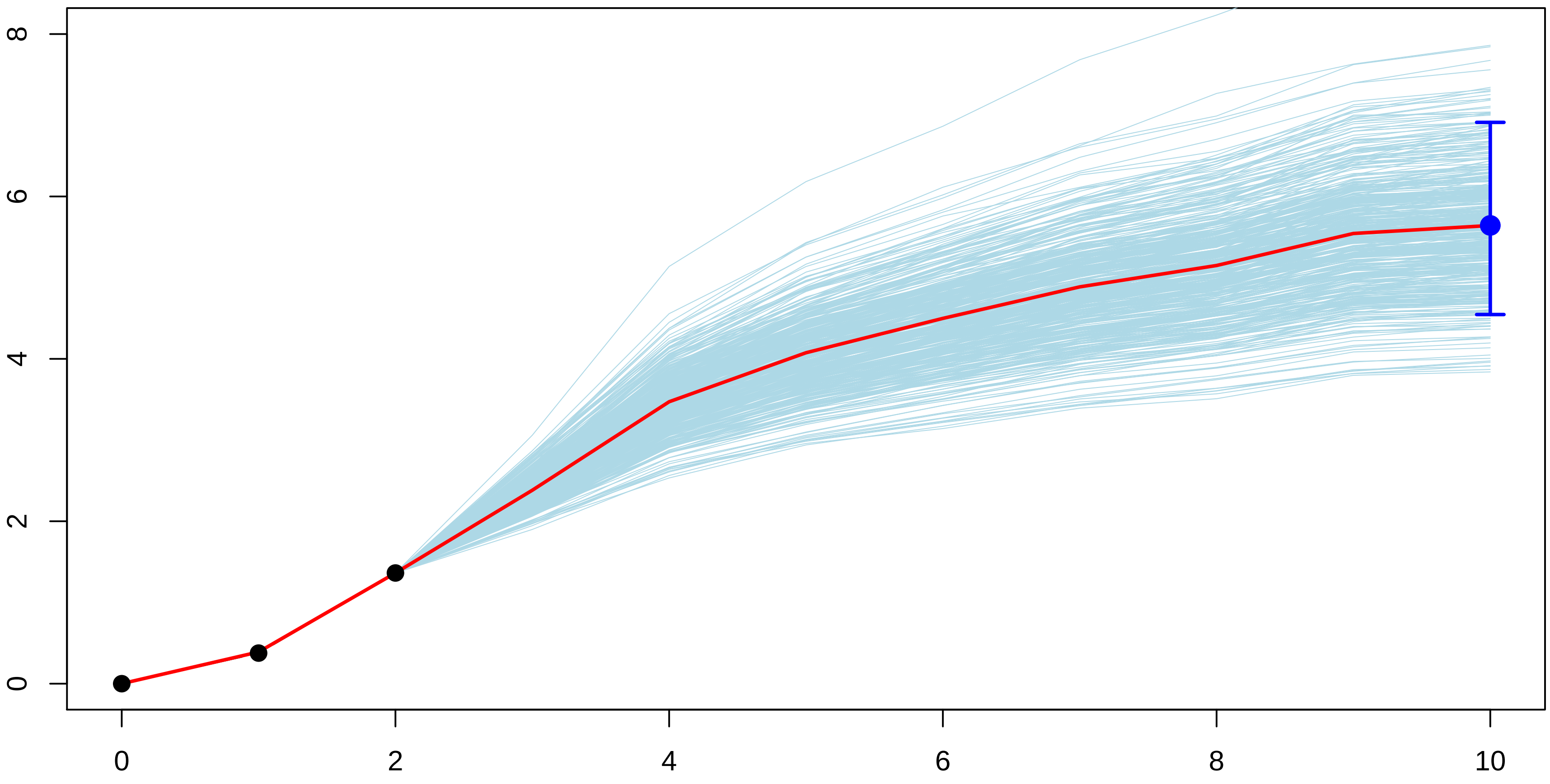
Bootstrap and GLM log-Poisson in triangles



Bootstrap and GLM log-Poisson in triangles



Bootstrap and GLM log-Poisson in triangles



Lee & Carter's approach of mortality

Dynamic models for mortality became popular following the publication of Lee & Carter (1992)'s models. The idea is that if

$$m(j,t) = \frac{\#\text{ deaths during calendar year } t \text{ aged } x \text{ last birthday}}{\text{average population during calendar year } t \text{ aged } j \text{ last birthday}}$$

$$\log m(j,t) = \alpha_j + \beta_j \gamma_t$$

- Lee & Carter (1992), $\log m(x,t) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}$,
- Renshaw & Haberman (2006), $\log m(x,t) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_{t-x}^{(3)}$,
- Currie (2006), $\log m(x,t) = \beta_x^{(1)} + \kappa_t^{(2)} + \gamma_{t-x}^{(3)}$,
- Cairns, Blake & Dowd (2006),
  $\text{logit} q(x,t) = \text{logit}(1-e^{-m(x,t)}) = \kappa_t^{(1)} + (x-\alpha)\kappa_t^{(2)}$,
- Cairns *et al.* (2007),
  $\text{logit} q(x,t) = \text{logit}(1-e^{-m(x,t)}) = \kappa_t^{(1)} + (x-\alpha)\kappa_t^{(2)} + \gamma_{t-x}^{(3)}$.

A stochastic model for mortality

Assume here that $\mathbb{E}(D|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})=\text{Var}(D|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$, thus a Poisson model can be considered. Then

$$D_{j,t}\sim\mathcal{P}(E_{j,t}\cdot\mu_{j,t})\text{ where }\mu_{j,t}=\exp[\alpha_j+\beta_j\gamma_t]$$

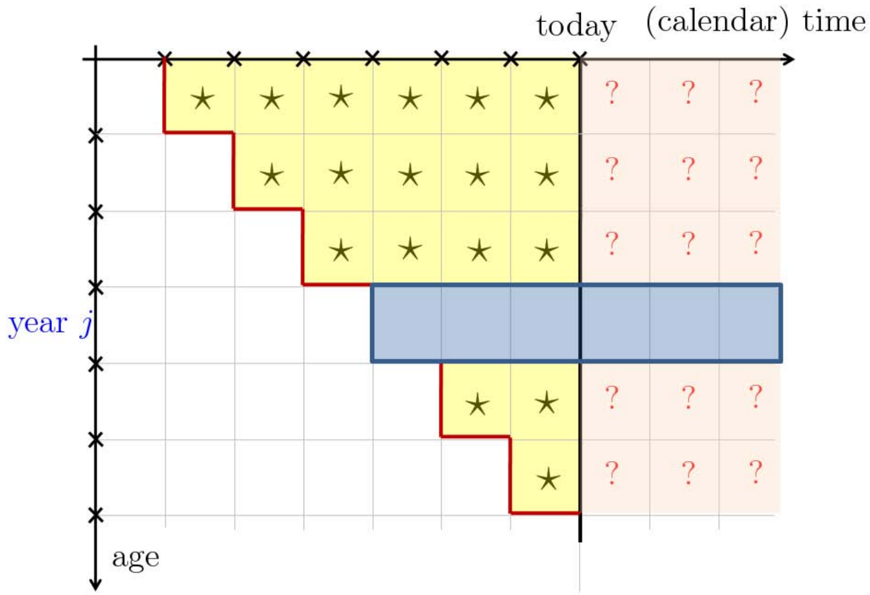
Brillinger (1986) and Brouhns, Denuit and Vermunt (2002)

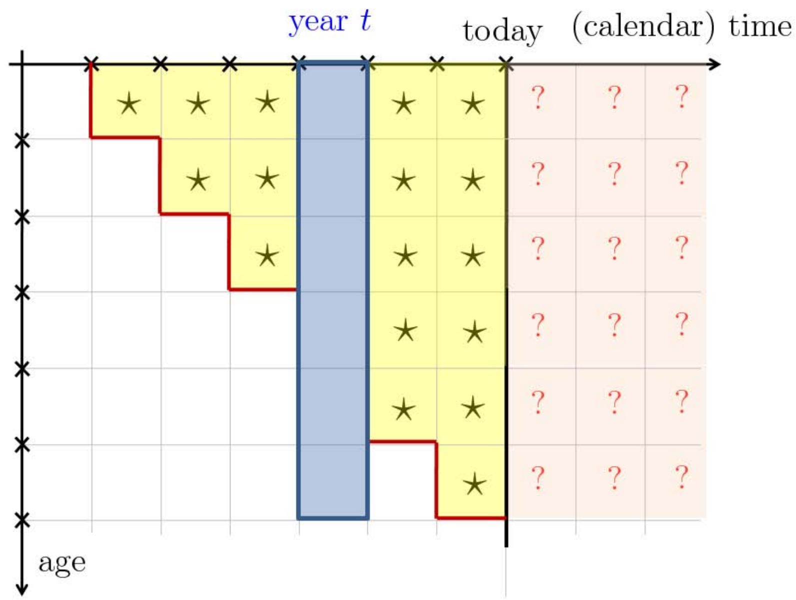
A stochastic model for mortality

Assume here that $\mathbb{E}(D|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \text{Var}(D|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$, thus a Poisson model can be considered. Then

$$D_{j,t} \sim \mathcal{P}(E_{j,t} \cdot \mu_{j,t}) \text{ where } \mu_{j,t} = \exp[\alpha_j + \beta_j \gamma_t]$$

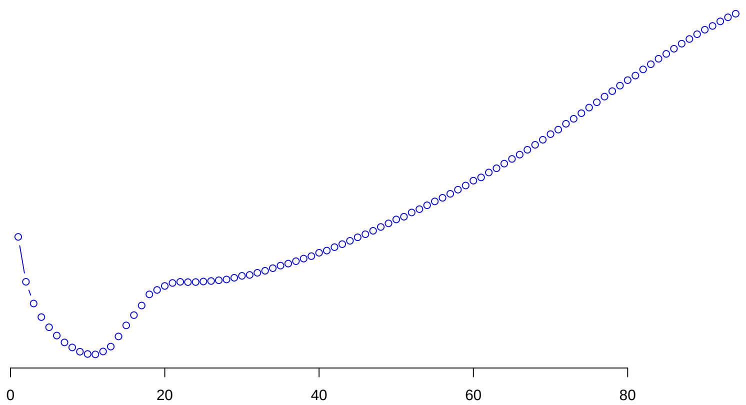
the age factors (α_j, β_j) the time factor $_t$

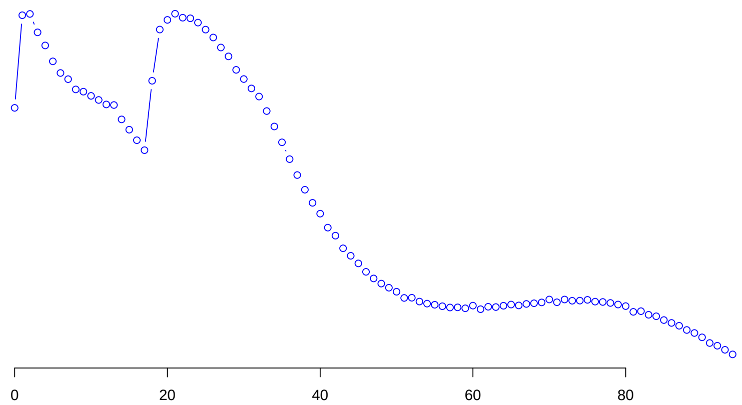




A stochastic model for mortality

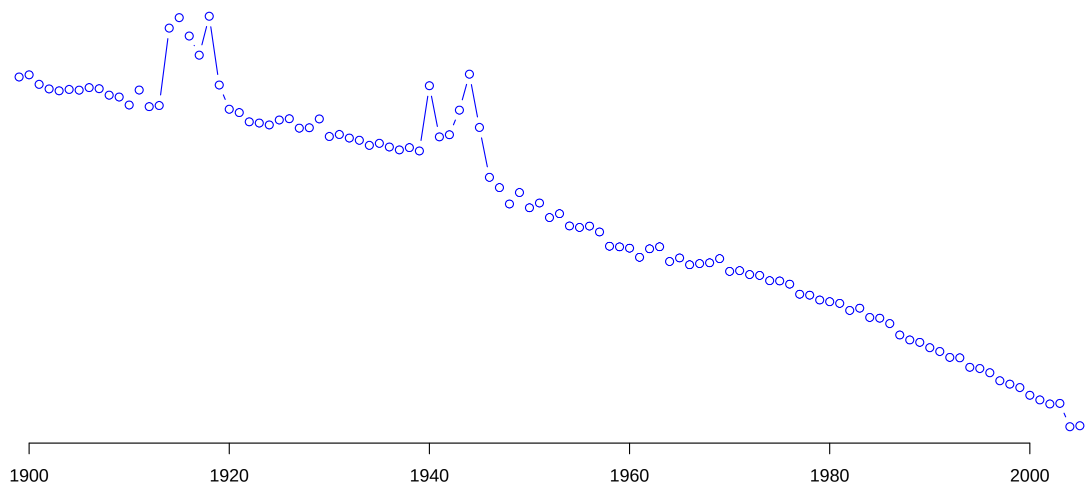
Two sets of parameters depend on the age, $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_0, \widehat{\alpha}_1, \cdots, \widehat{\alpha}_{110})$ and $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \cdots, \widehat{\beta}_{110})$.





A stochastic model for mortality

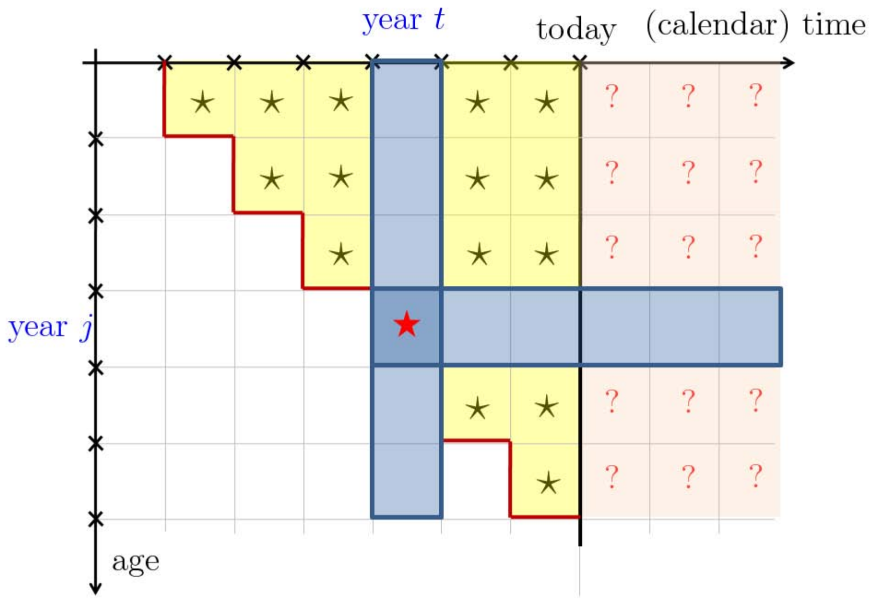
and one set of parameters depends on the time, $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_{1899}, \widehat{\gamma}_{1900}, \cdots, \widehat{\gamma}_{2005})$.



Errors and predictions

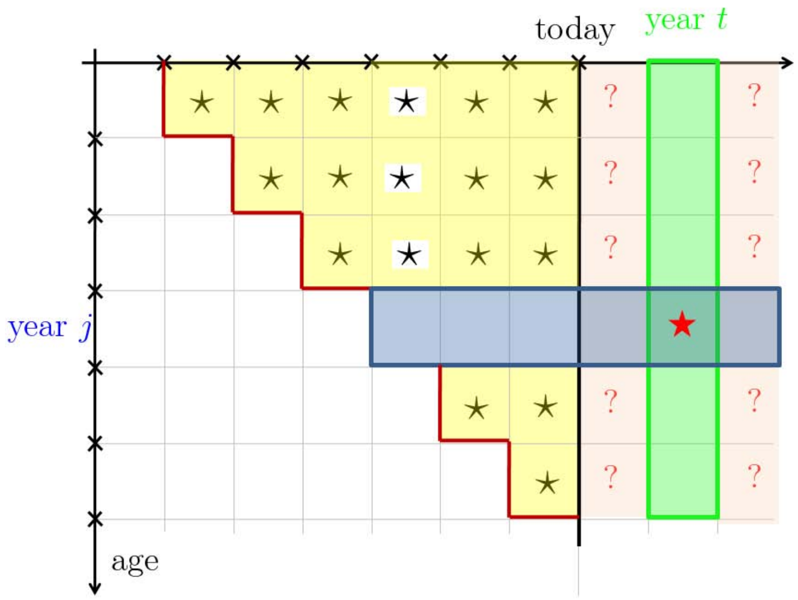
$\exp[\widehat{\alpha}_j + \widehat{\beta}_j \widehat{\gamma}_t]$

on past observations



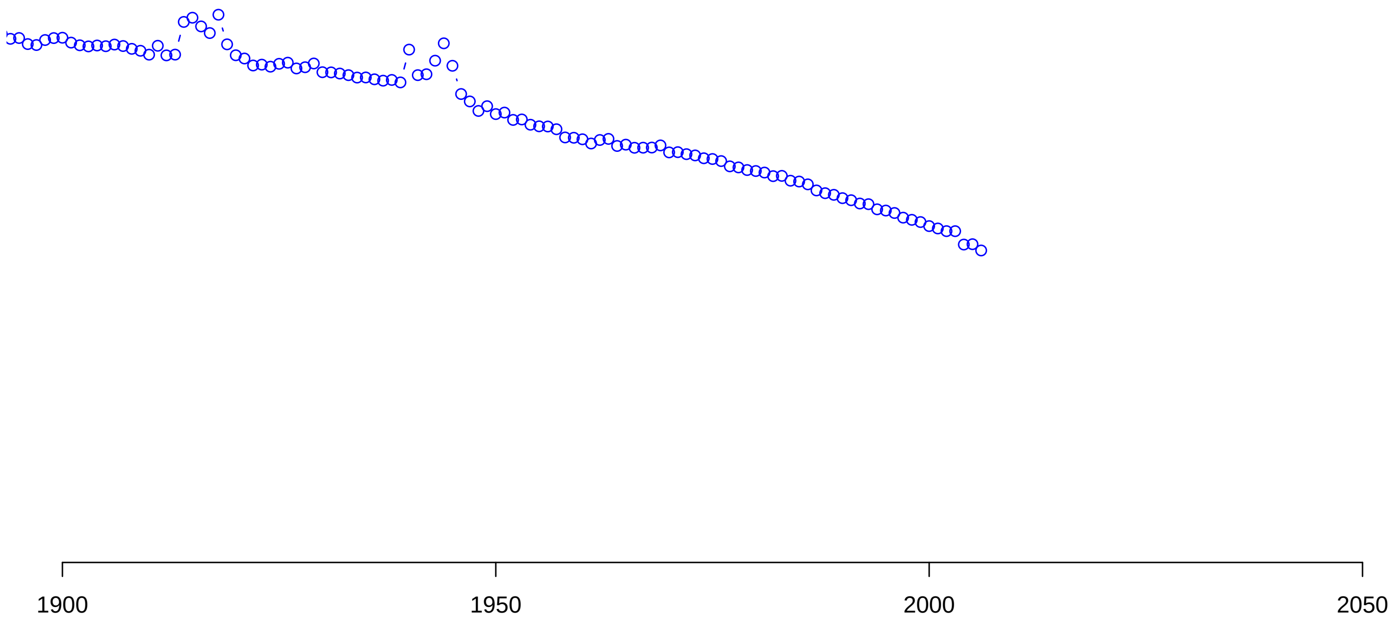
$\exp[\widehat{\alpha}_j + \widehat{\beta}_j \tilde{\gamma}_t]$

on the future



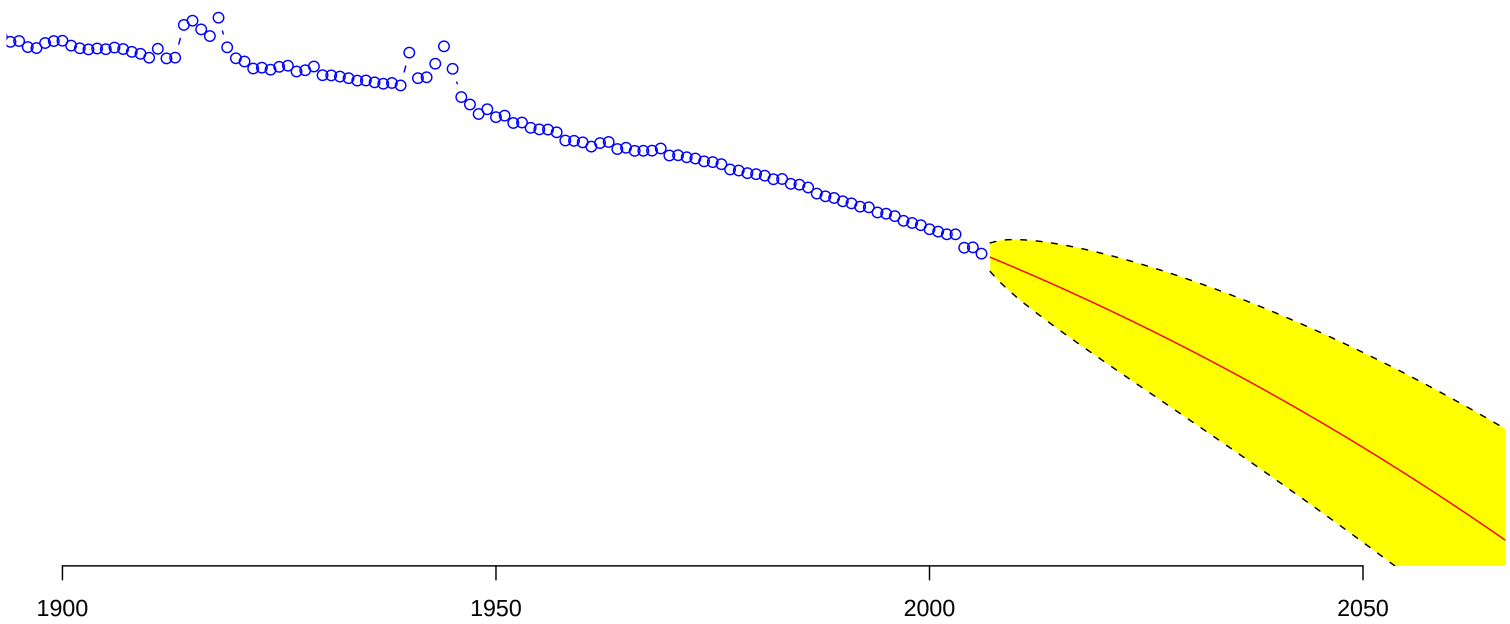
Forecasting $\widehat{\gamma}$

Based on $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_{1899}, \cdots, \widehat{\gamma}_{2005})$, we need to forecast $\boldsymbol{\gamma} = (\gamma_{2006}, \cdots, \gamma_{2050})$.



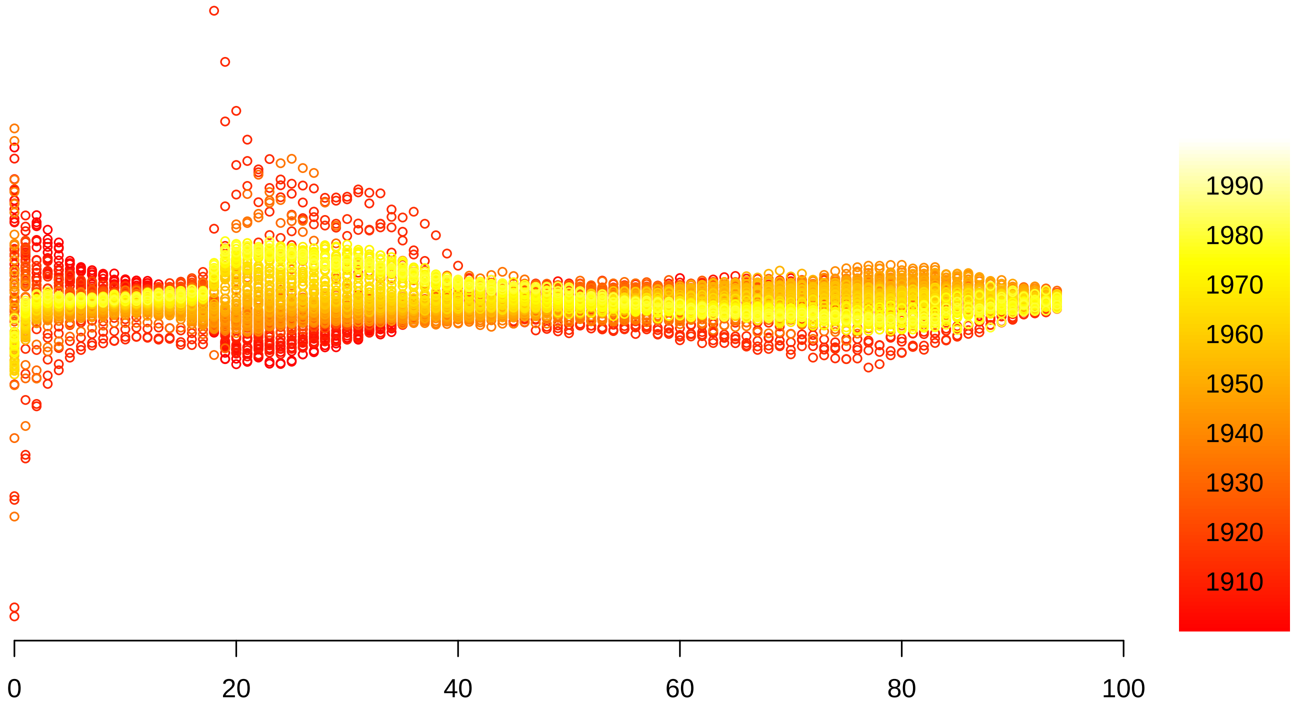
Forecasting $\widehat{\gamma}$

Classically integrated ARIMA processes are considered,



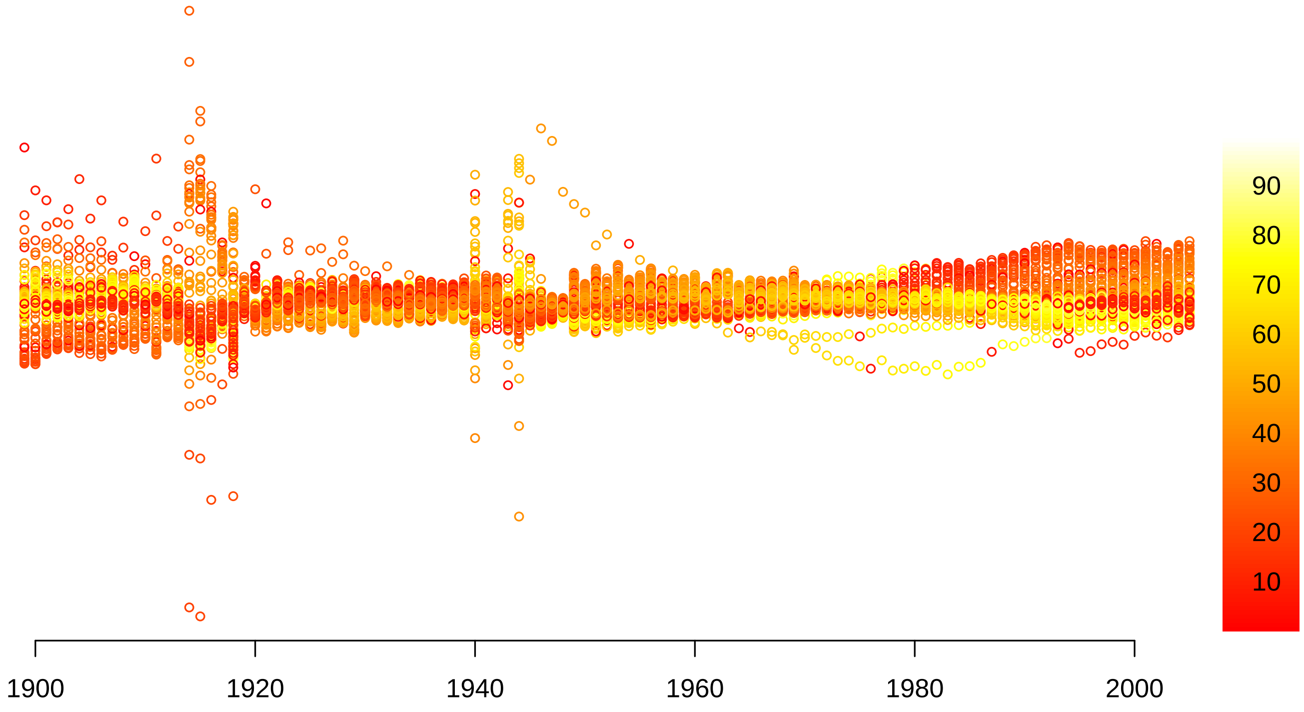
Understanding errors in stochastic models

Pearson's residuals, $\varepsilon_{j,t} = \dfrac{D_{j,t} - \widehat{D}_{j,t}}{\sqrt{\widehat{D}_{j,t}}}$, as a function of age j



Understanding errors in stochastic models

Pearson's residuals, $\varepsilon_{j,t} = \dfrac{D_{j,t} - \widehat{D}_{j,t}}{\sqrt{\widehat{D}_{j,t}}}$, as function of time t



Understanding outliers

Outliers can *simply* be understood in a univariate context. To extand it in higher dimension, Tukey (1975) defined the

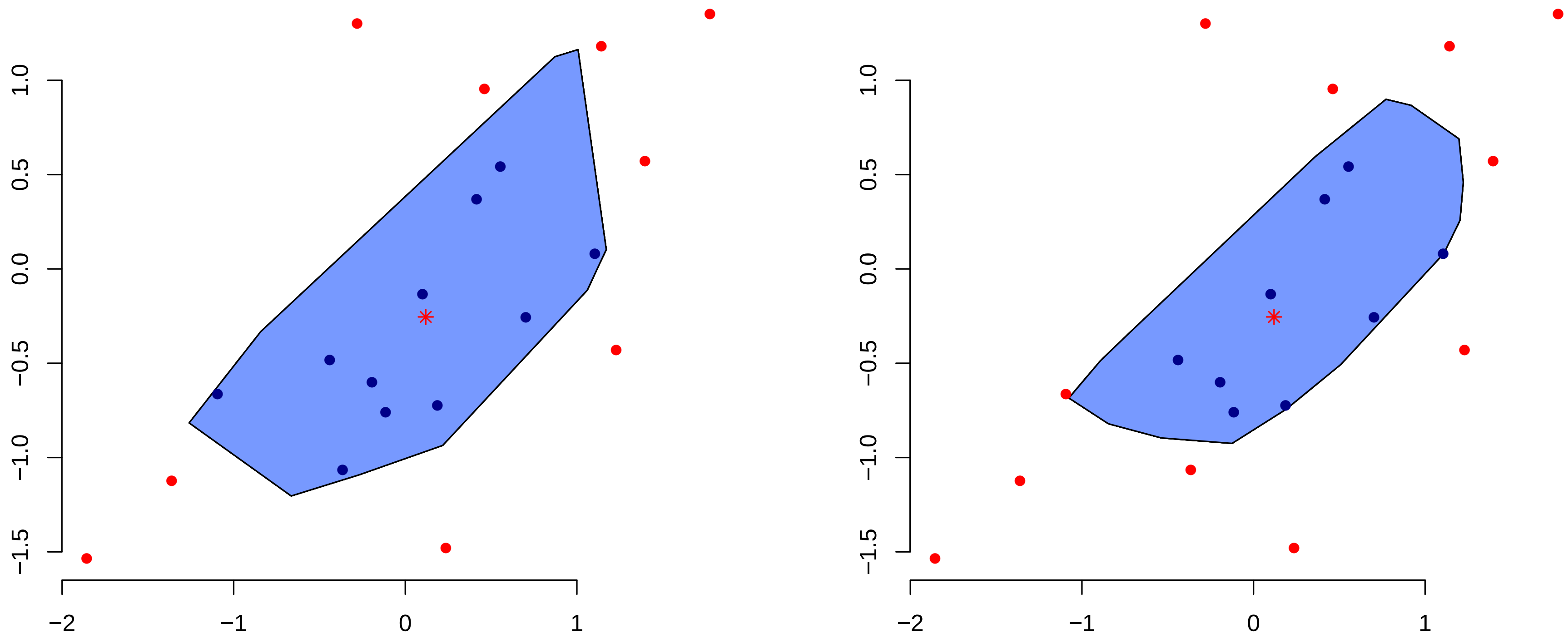
$$\text{depth}(\boldsymbol{y}) = \min_{\boldsymbol{u},\boldsymbol{u}\neq\boldsymbol{0}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{1}(\boldsymbol{X}_i \in H_{\boldsymbol{y},\boldsymbol{u}}) \right\}$$

where $H_{\boldsymbol{y},\boldsymbol{u}} = \{\boldsymbol{x} \in \mathbb{R}^d \text{ such that } \boldsymbol{u}'\boldsymbol{x} \leq \boldsymbol{u}'\boldsymbol{y}\}$ and for $\alpha > 0.5$, defined the depth set as

$$D_\alpha = \{\boldsymbol{y} \in \mathbb{R} \in \mathbb{R}^d \text{ such that depth}(\boldsymbol{y}) \geq 1 - \alpha\}.$$

The empirical version is called the bagplot function (see e.g. Rousseeuw & Ruts (1999)).

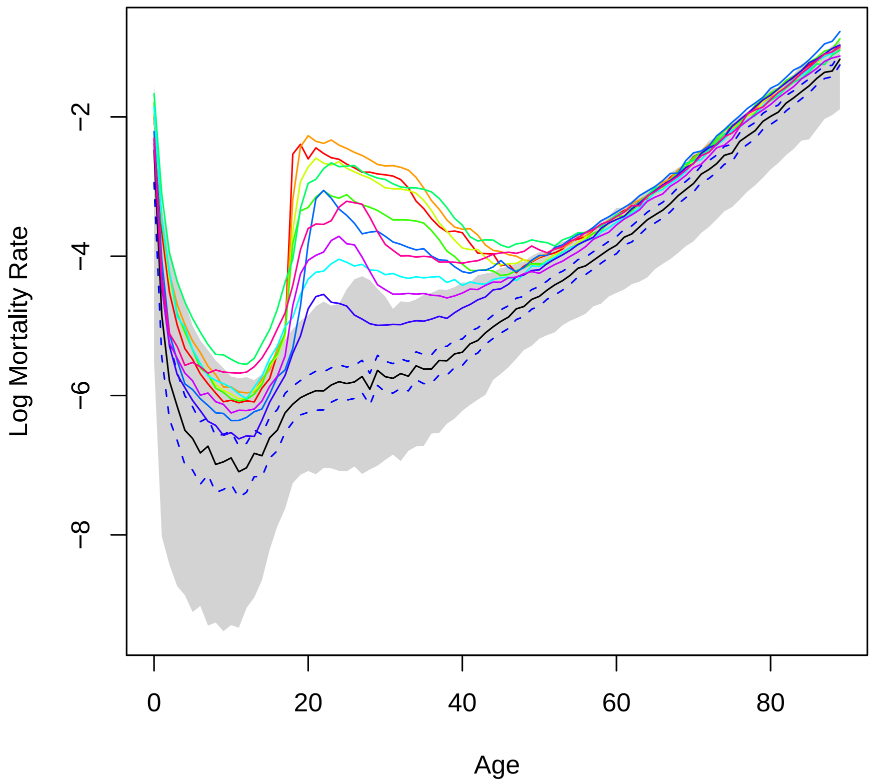
Understanding outliers

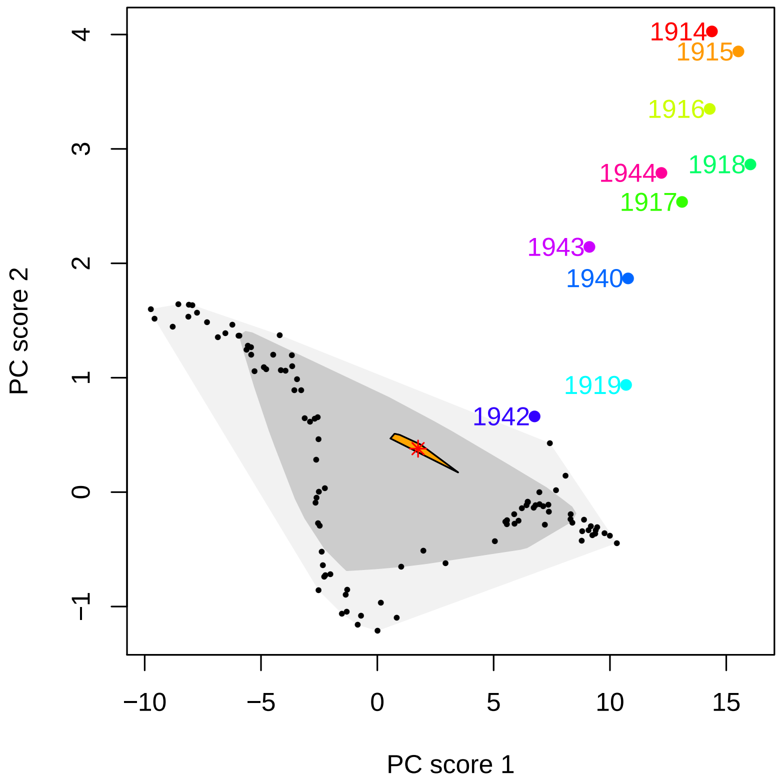


where the blue set is the empirical estimation for D_α, $\alpha = 0.5$.

Understanding outliers

It is possible to extend it to define (past) functional outliers,

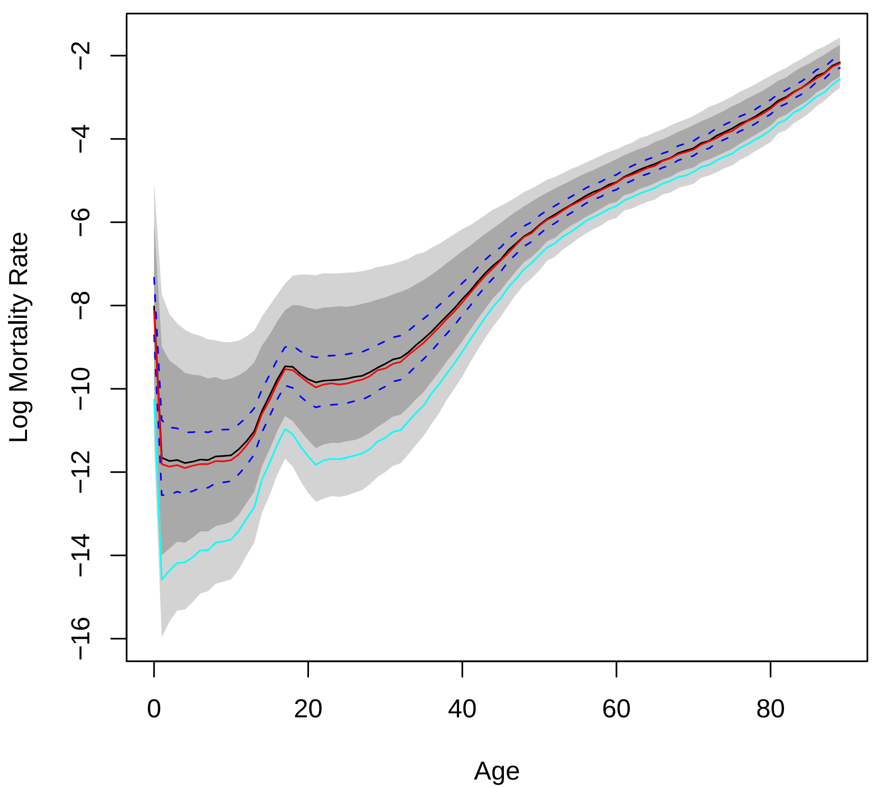


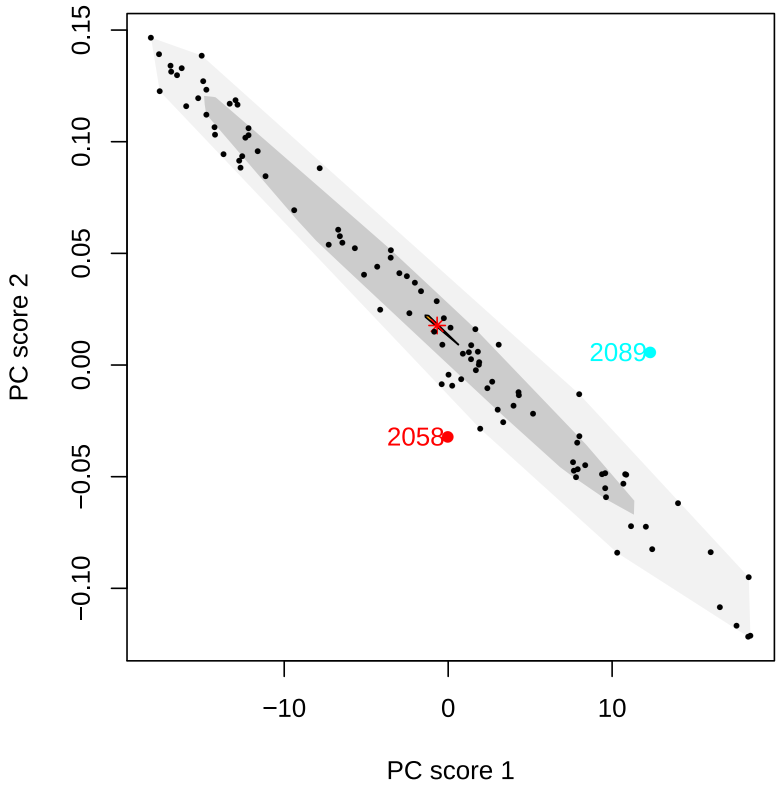


(here male log-mortality rates in France from 1899 to 2005).

Understanding outliers when generating scenarios

Based on the log-Poisson Lee & Carter model, it is possible to generate scenarios,





$\Longrightarrow$ this stochastic model does not generate extremal scenarios

References

Hachemeister, C.A. & Stanard, J. N. (1975). IBNR claims count estimation with static lag functions. *XIIth Astin Colloquium*, Portugal.

Kremer, E. (1982). IBNR-claims and the two-way model of ANOVA. *Scandinavian Actuarial Journal*, pp. 4755.

Mack, T. (1991). Which stochastic model is underlying the chain ladder method ? *Insurance : mathematics & economics*, **15**(2/3), 133-138.

De Vylder, F. (1978). Estimation of IBNR claims by least squares. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, pp. 249254.

Taylor, G. (1977). Separation and other effects from the distribution of non-life insurance claim delays. *Astin Bulletin*, **9**, 217-230.

Lee, R. D., & Carter, L. (1992), Modeling and forecasting U.S. mortality, *Journal of the American Statistical Association*, **87**, 659671.

Renshaw, A.E. & Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance : Mathematics & Economics*, **38**, 556-570.

Currie, I.D. (2006). Smoothing and forecasting mortality rates with p-splines

Cairns , A., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., & Khalaf-Allah, M. (2008), Mortality Density Forecasts : An Analysis of Six Stochastic Mortality Models. *Pensions Institute Discussion Paper PI-0801*.

Brillinger, D. R. (1986). The natural variability of vital rates and associated statistics. *Biometrics*, **42**, 693734.

Brouhns, N., Denuit, M. & Vermunt, J.K. (2002). A Poisson log-bilinear regression approach to the construction of projected life-tables. *Insurance : Mathematics and Economics*, **31**, 373-393.

Tukey, J.W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, **2**, 523-531.

Ruts, I. & Rousseeuw, P.J. (1996). Computing depth contours of bivariate point clouds, *Computational Statistics & Data Analysis*, **23**. 153168.

Rousseeuw, P.J., Ruts, I., & Tukey, J.W. (1999). The Bagplot : A Bivariate Boxplot, *Annals of Statistics*, **53**, 382387.