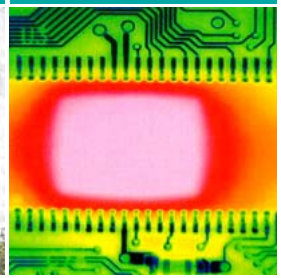
BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Computerised detection and stochastic forecast of age, period and cohort effects

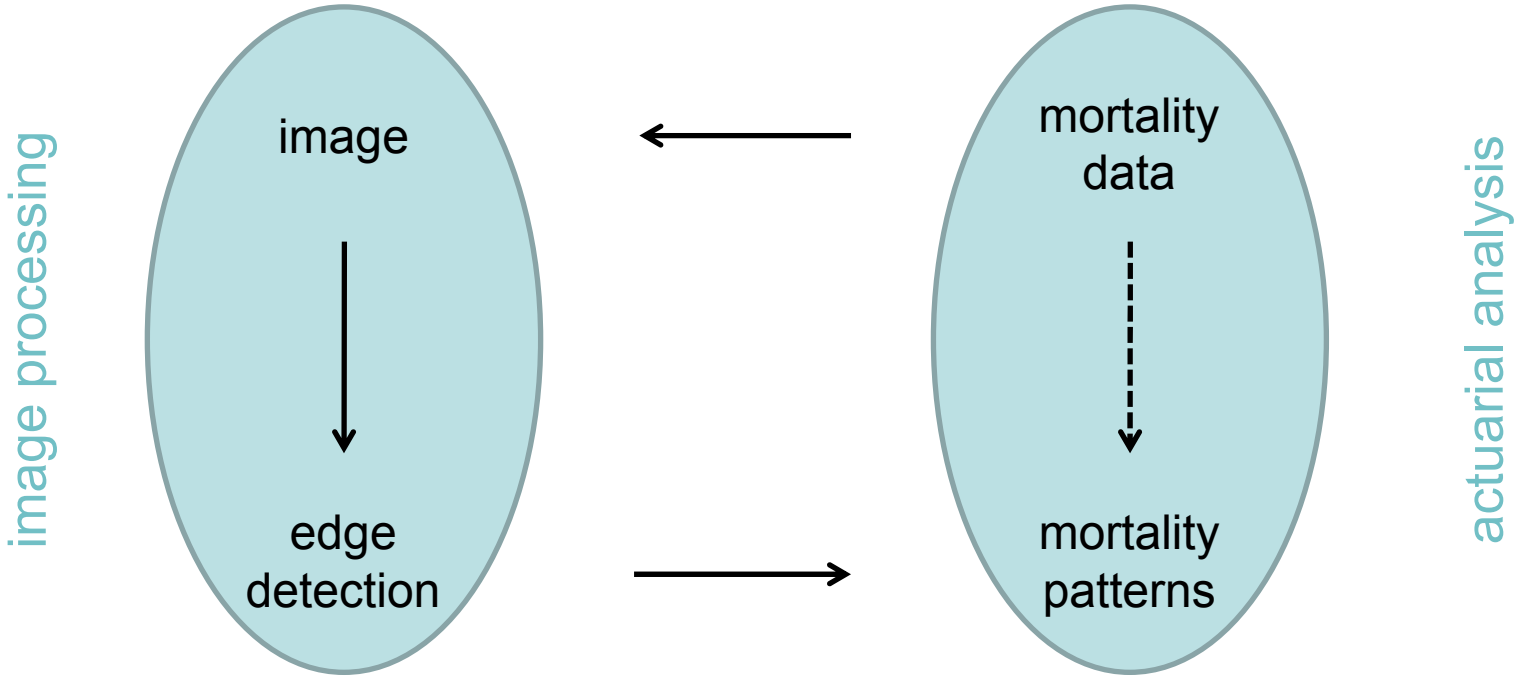
Prof. Dr. Michael Ortmann
Aktuar DAV, FIA

Jahrestagung DVfVW
21.3.2013



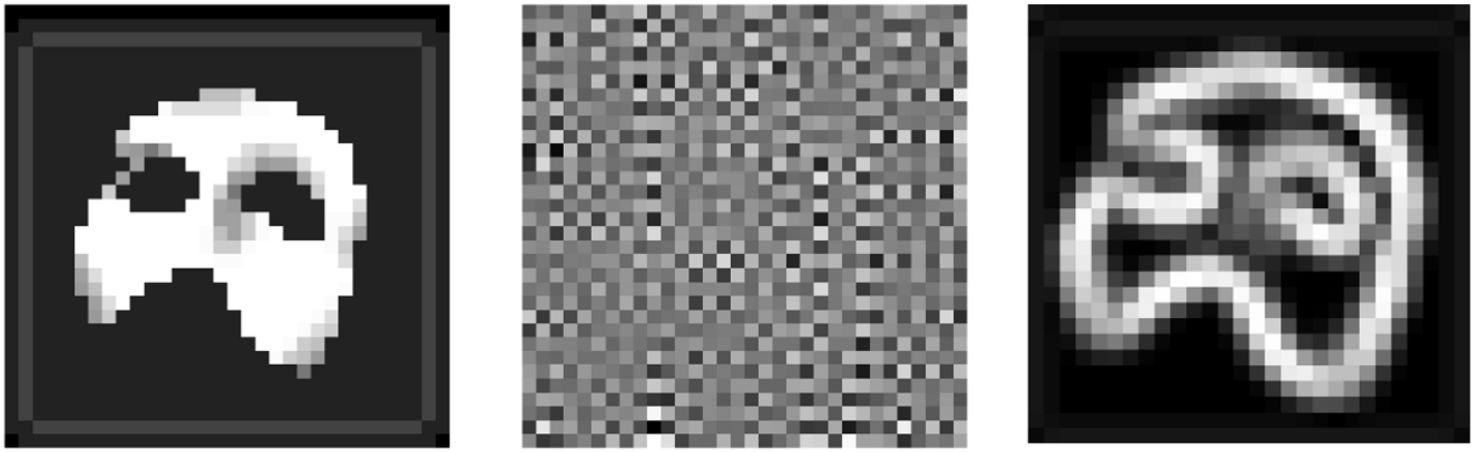
Part I – detecting mortality patterns

- to demonstrate how appropriate methods of computer vision and image procession may be applied to mortality data
- to specify a multistage detection algorithm
- to identify age, period and cohort effects



Motivation

Would you be able to see this ? The computer can !



(a) (b) (c)

(a) the input pattern

(b) the input corrupted by random noise

(c) the gradient image of (b)

see: Meer, P., Georgescu, B. (2001): Edge detection with embedded confidence, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, 1351–1365.

Data and definitions

- mortality rates

$$m(x,t)$$

- incremental mortality development factors

$$IF(x,t) = \frac{m(x,t)}{m(x,t-1)} - 1$$

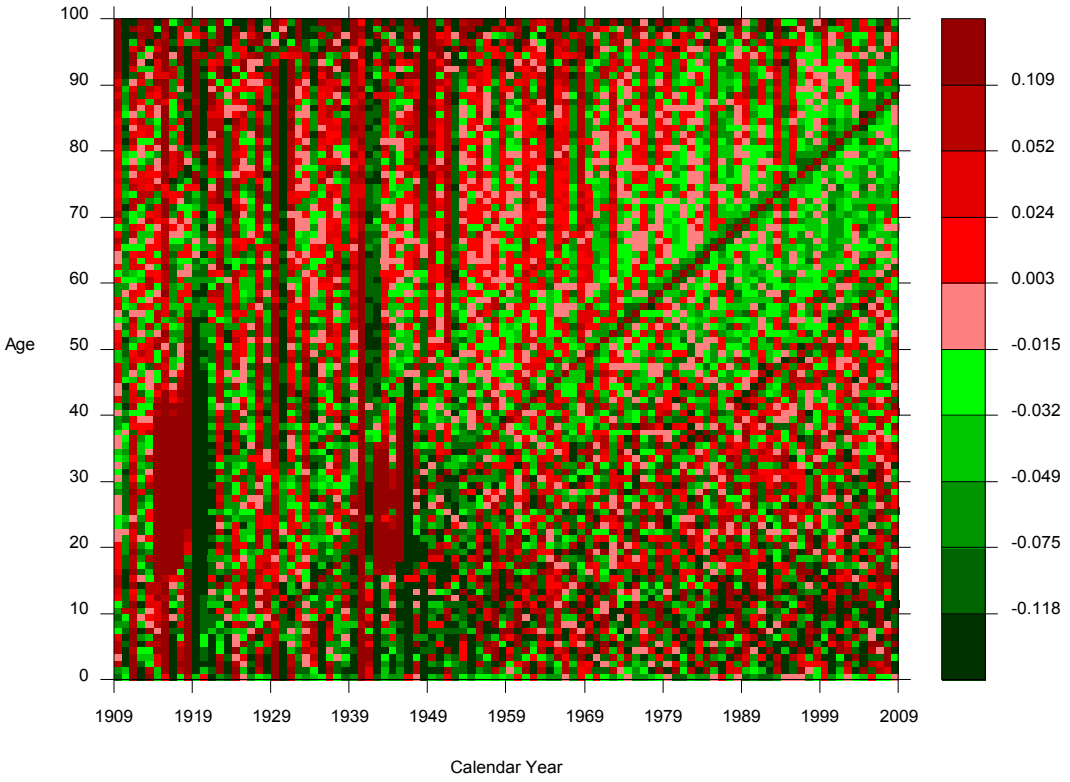
- overall mortality

$$m(x,t) = m(x,0) \cdot \prod_{t>0} \left(1 + IF(x,t)\right)$$

Data for England and Wales

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction

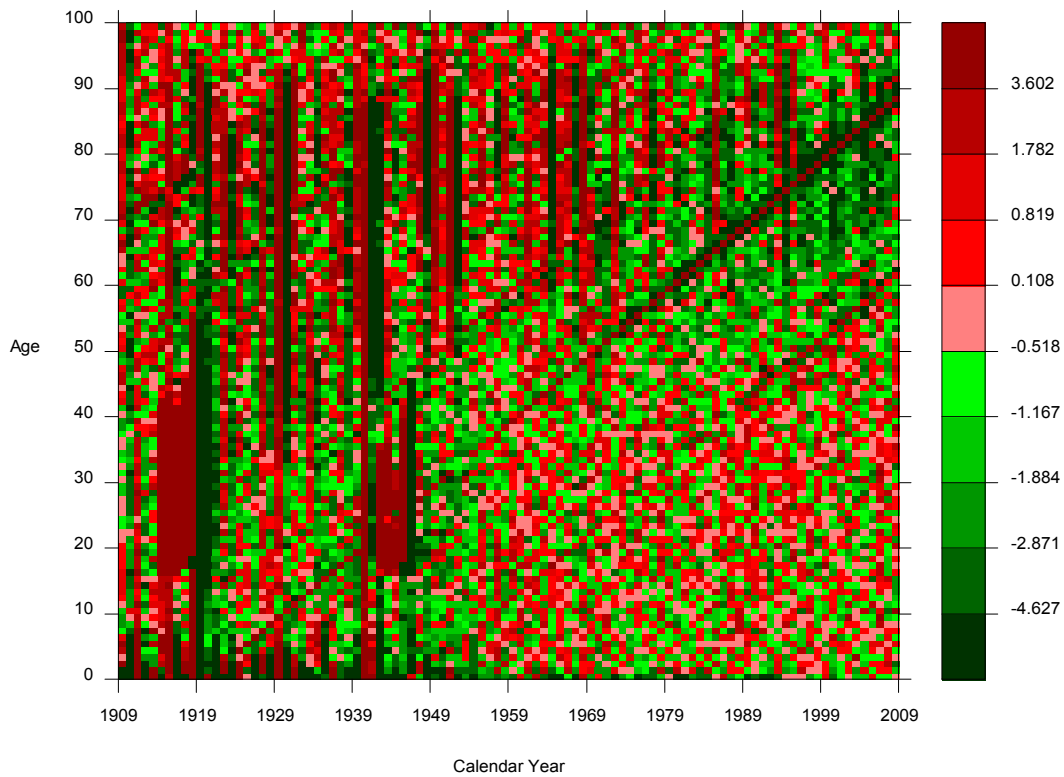


- age effects – horizontal lines
- period effects – vertical lines
- cohort effects – diagonal lines

Incremental mortality development factors

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction

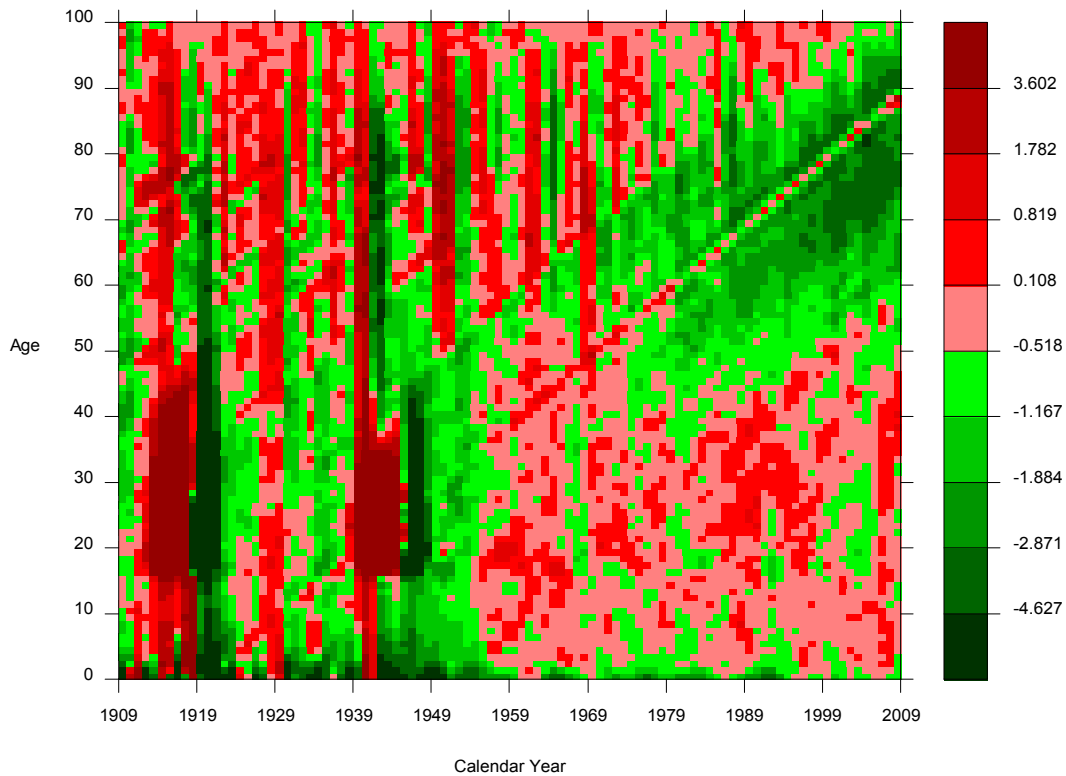


$$\overline{IF}(x,t) = IF(x,t) \cdot \sqrt{\frac{l(x,t-1) \cdot m(x,t-1)}{1 - m(x,t-1)}}$$

Smoothing

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction



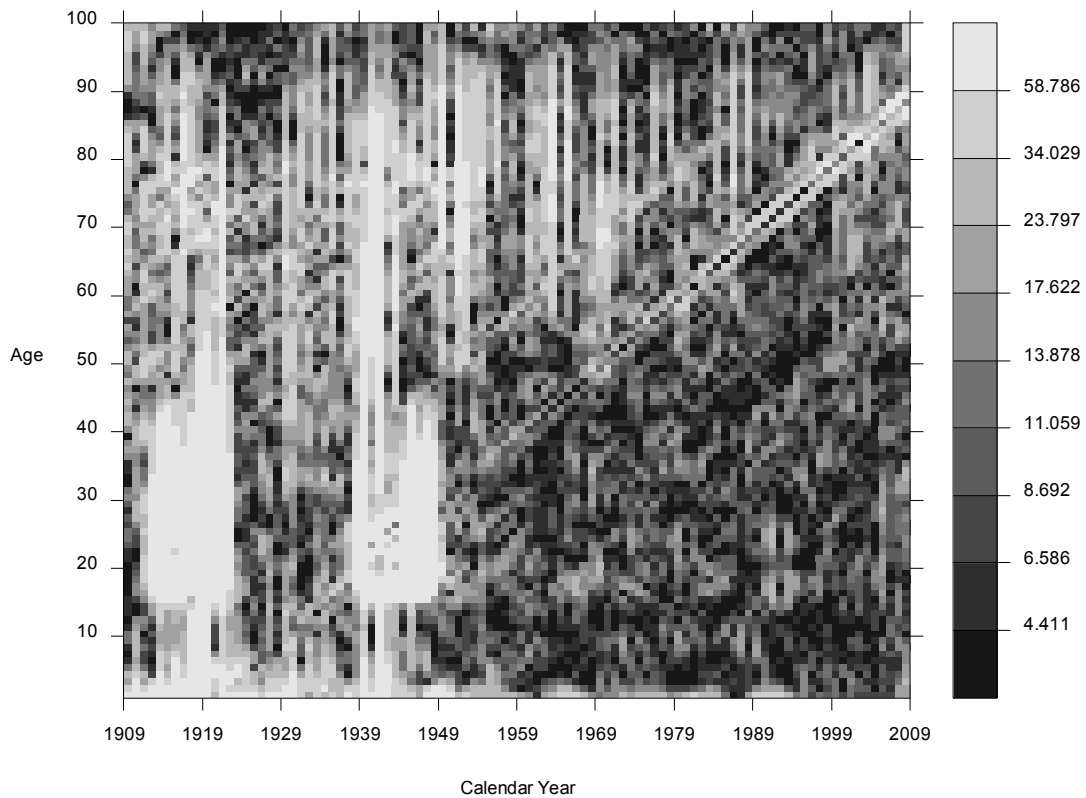
$$IF_{\sigma,n}(x,t) = \sum_{i=-n}^{n} \sum_{j=-n}^{n} \overline{IF}(x+i, t+j) \cdot g_{\sigma}(i,j)$$

$$g_{\sigma}(x,t) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+t^2}{2\sigma^2} \right)$$

Gradient estimation

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction



$$H_x = \begin{pmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{pmatrix} \quad \text{and} \quad H_t = \begin{pmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{pmatrix}$$

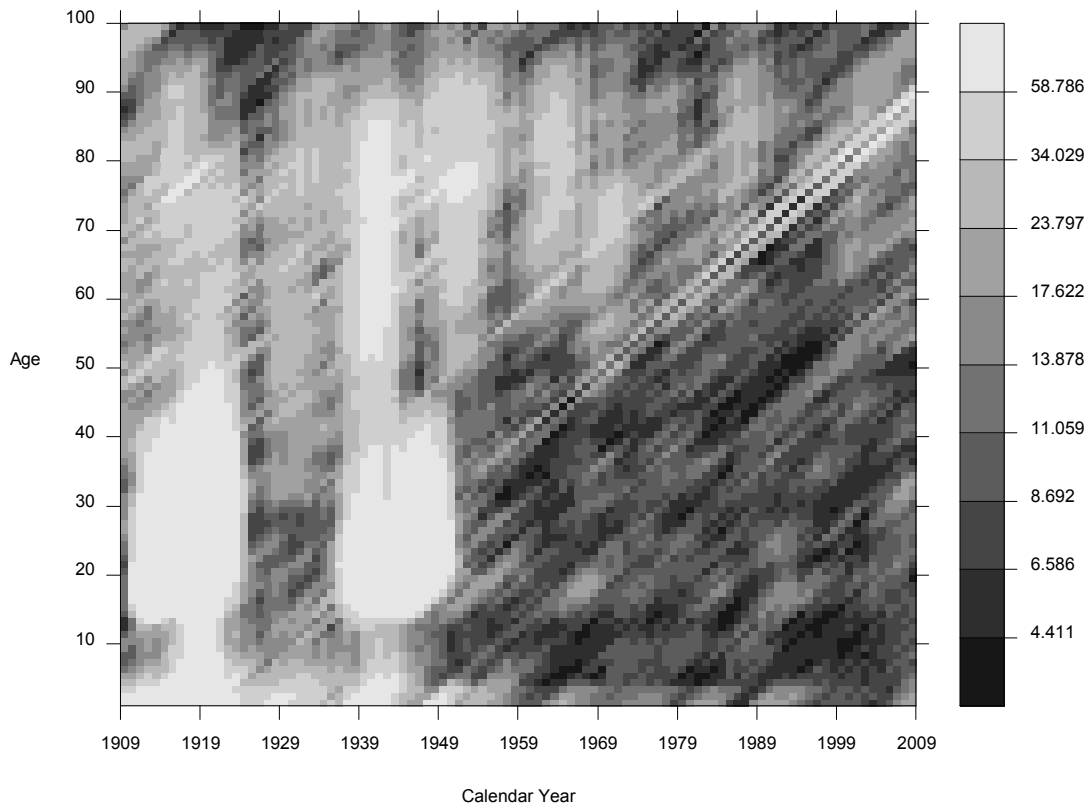
$$D_x = IF_{\sigma,n} * H_x$$
$$D_t = IF_{\sigma,n} * H_t$$

$$D(x_0,t_0) = \sqrt{\left(D_x(x_0,t_0)\right)^2 + \left(D_t(x_0,t_0)\right)^2}$$

Separating and tracking - cohort effects

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction

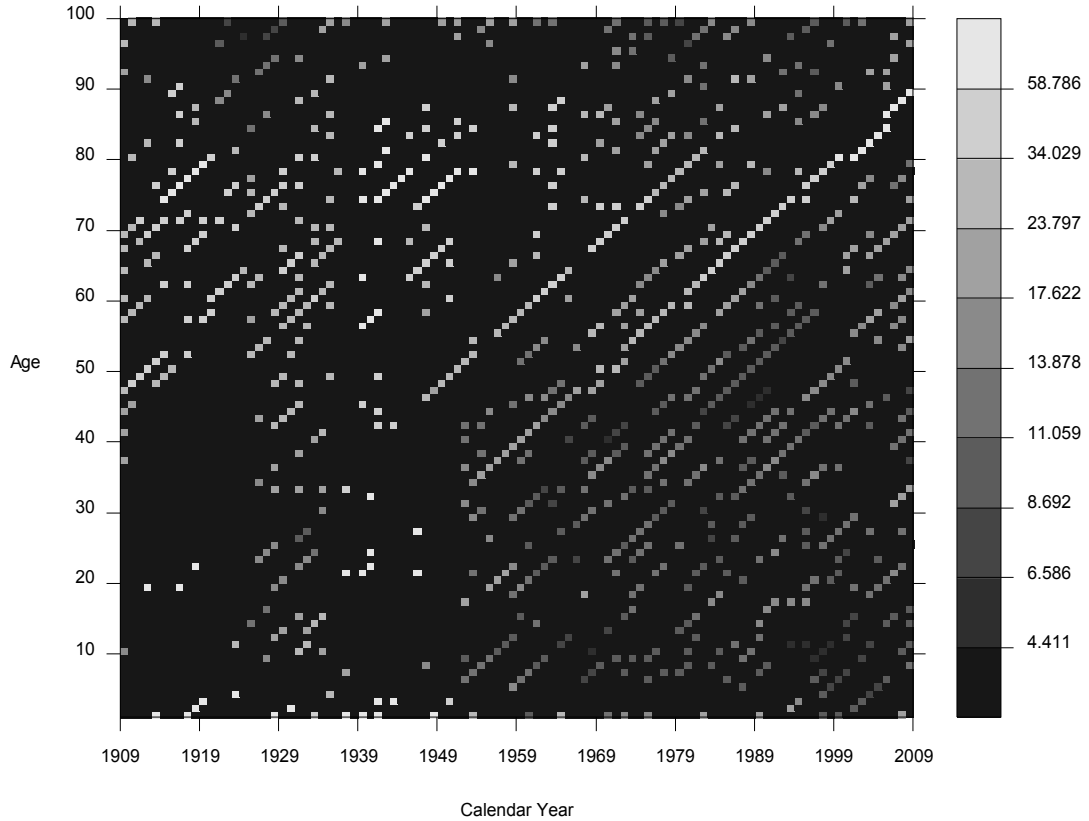


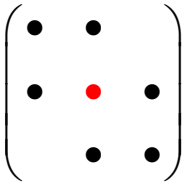
$$W_C = \frac{1}{n}\begin{pmatrix} 0 & & 1 \\ & \mathrm{N} & \\ 1 & & 0 \end{pmatrix} \qquad T_C = D * W_C$$

Non-maxima suppression - cohort effects

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction

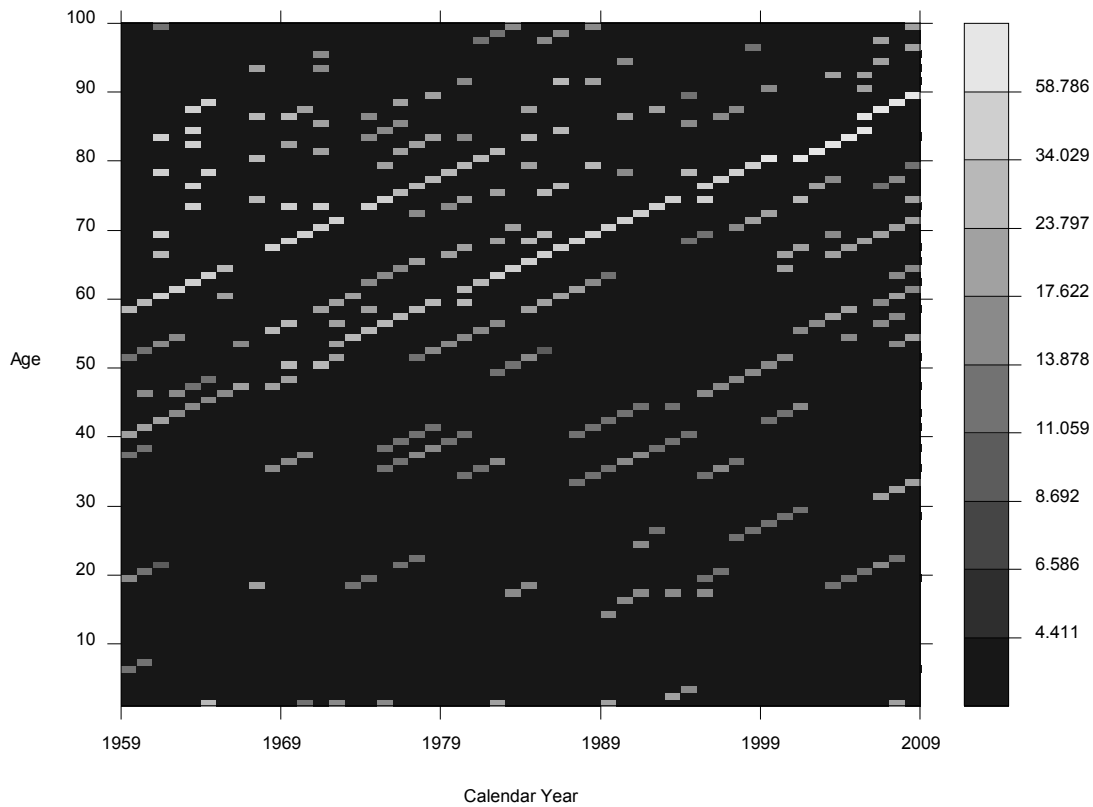




Hysteresis thresholding - cohort effects

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction

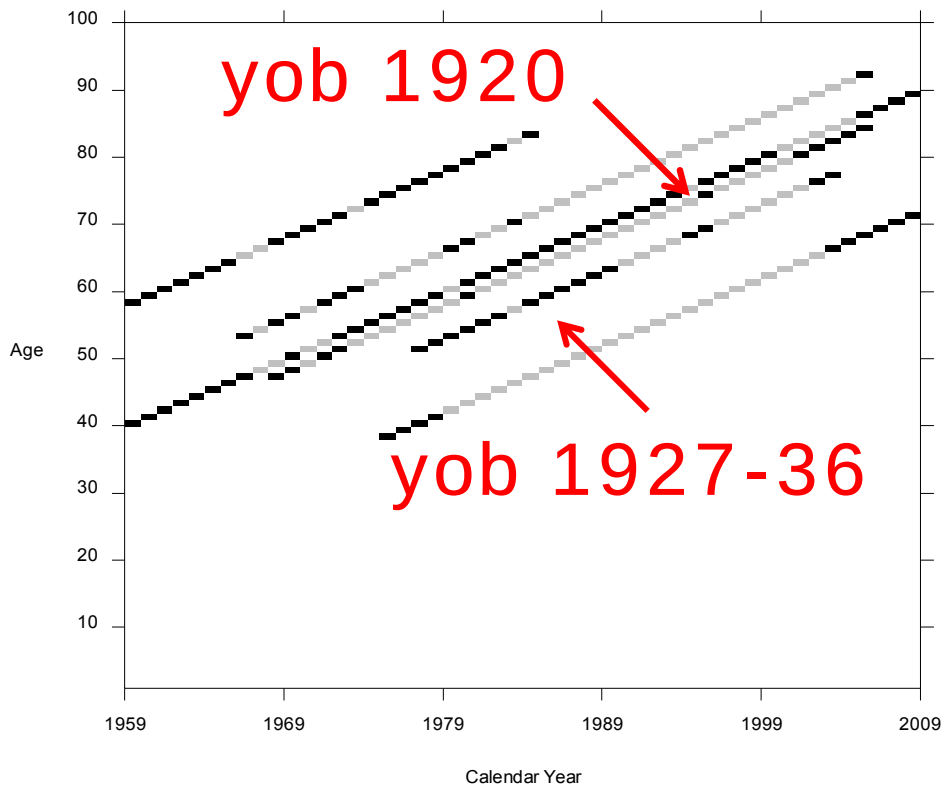


data points whose value is higher than t_u are retained, while the ones with a value lower than t_l are discarded, i.e. their value is set to zero.

Feature extraction - cohort effects

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction



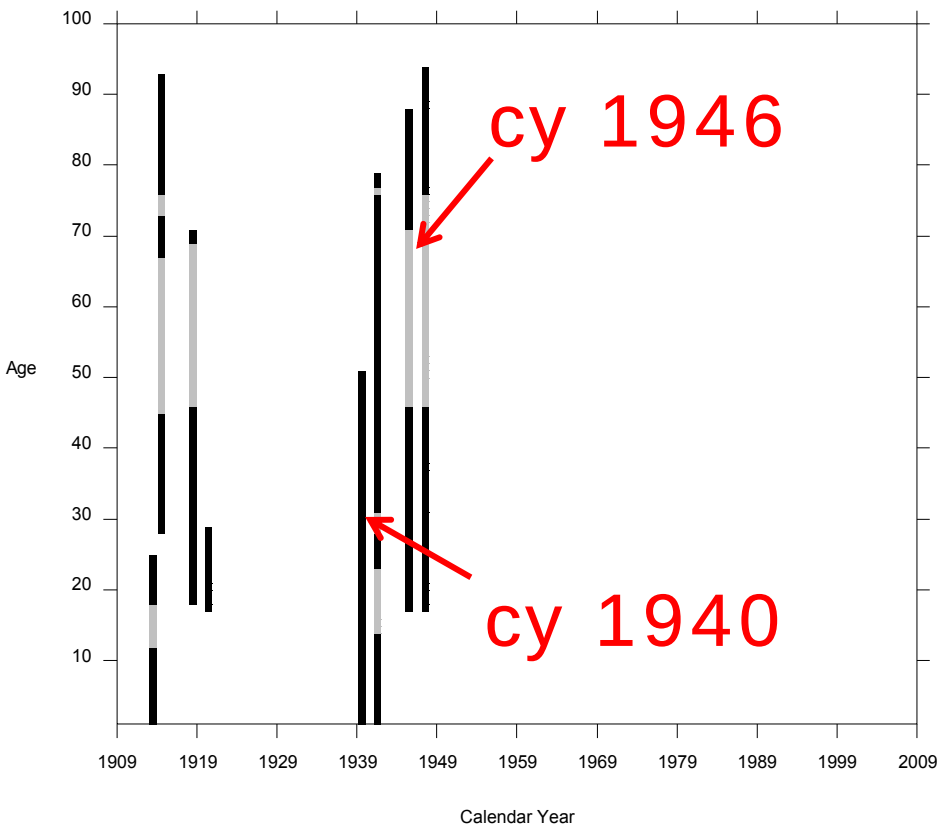
feature extraction based on idea of Hough transformation:

apply ranking of importance in combination with contour filling

Feature extraction - period effects

Canny algorithm

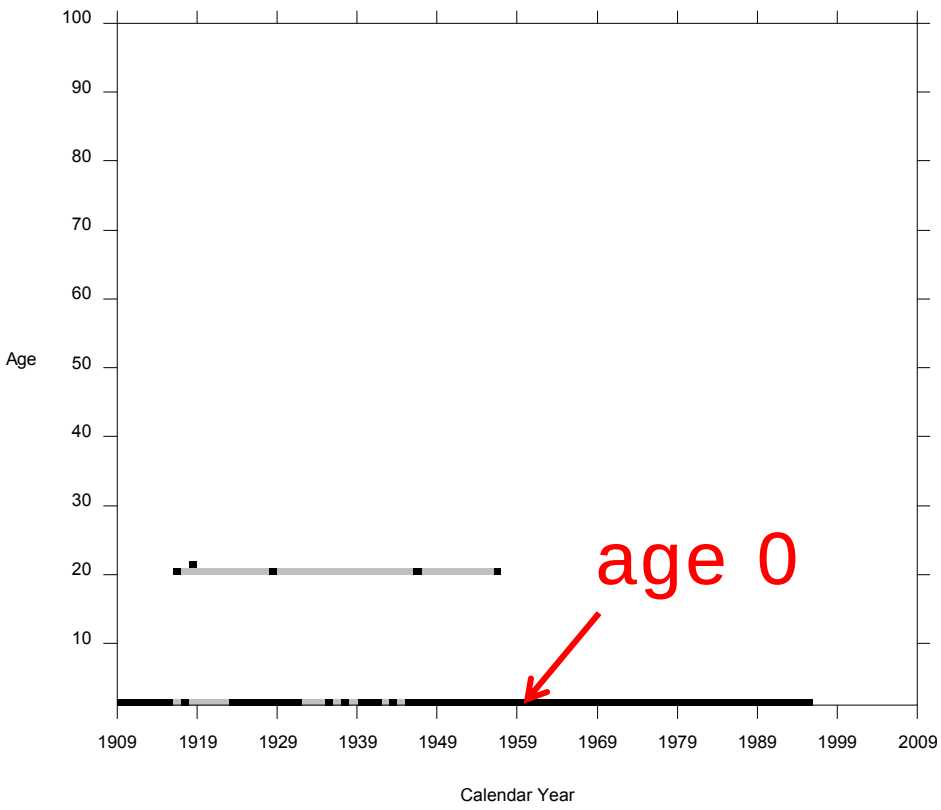
- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction



Feature extraction - age effects

Canny algorithm

- data
- standardising
- smoothing
- gradient estimation
- separating & tracking
- non-maxima suppression
- hysteresis thresholding
- feature extraction



Summary Part I

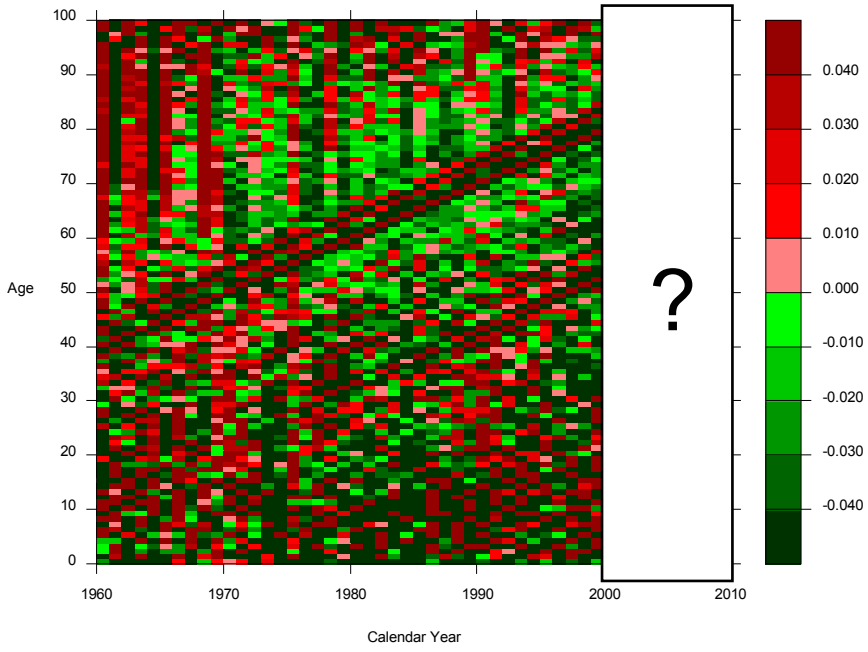
- techniques already established in computer vision and image processing may be applied to analyse mortality data
- a suitably adapted version of the Canny operator may serve to more objectively detect age, period, and cohort effects
- the methodology may provide quantitative assistance in actuarial analyses at large

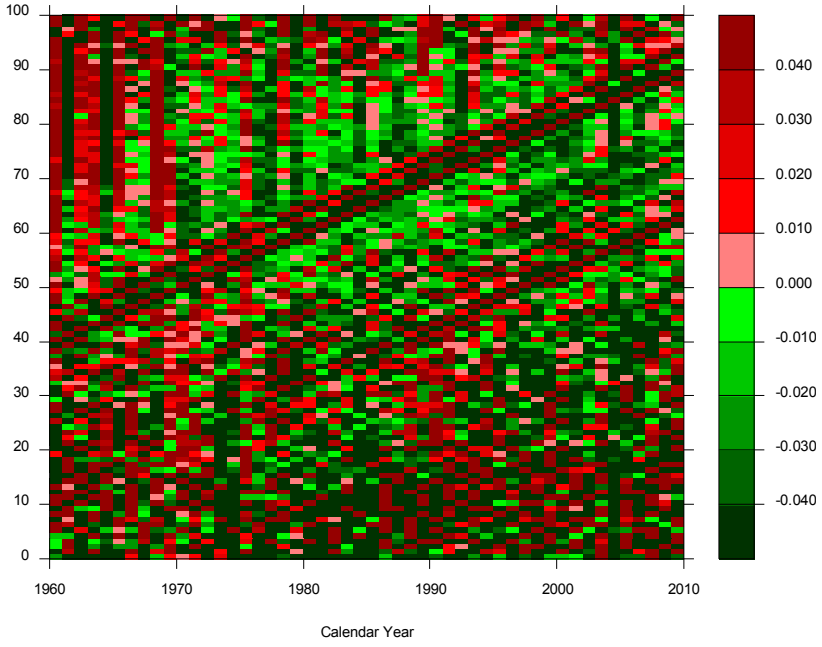
Part II – forecasting mortality patterns

- to apply a stochastic mortality forecast model
- to separate age, period, and cohort effects from the data
- to forecast mortality patterns and residual mortality separately

Motivation

How would you have forecasted German mortality from 2000 - 2009? See what DAV2004RM assumes!

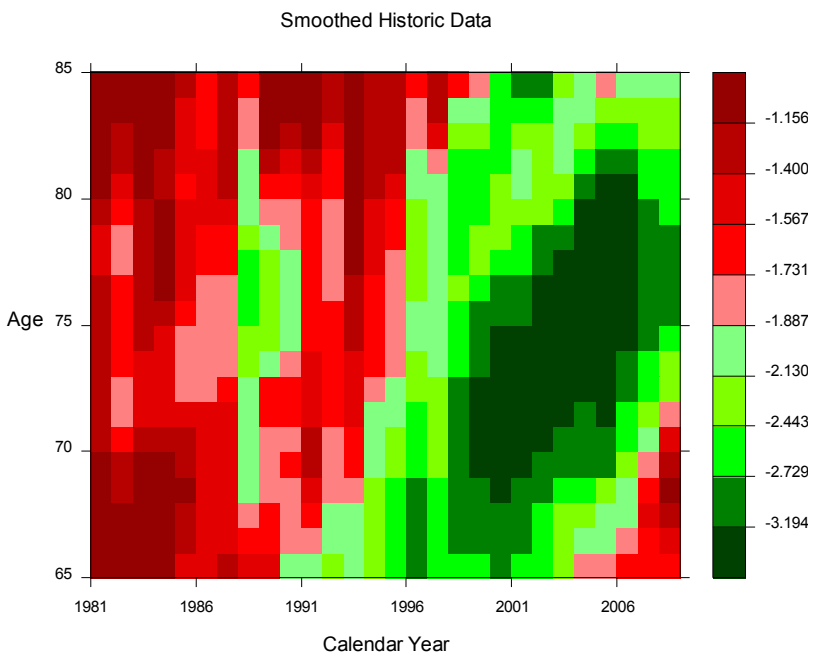


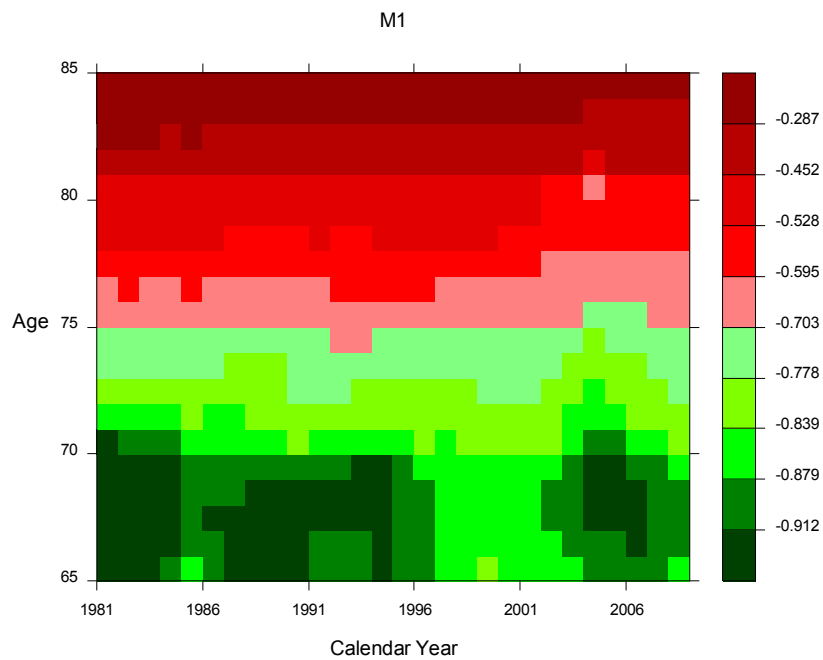


- improvement rates for males in West Germany from 1960 to 1999 in combination with implicit mortality improvements contained in DAV2004RM for 2000 to 2009 on the left
- and historic improvement rates from 1960 to 2009 on the right

How do other stochastic models forecast E & W mortality?

weighted and smoothed improvement rates for ages 65 to 84 from 1981 to 2008
historic realisations versus predicted outcome of Lee-Carter model (M1)



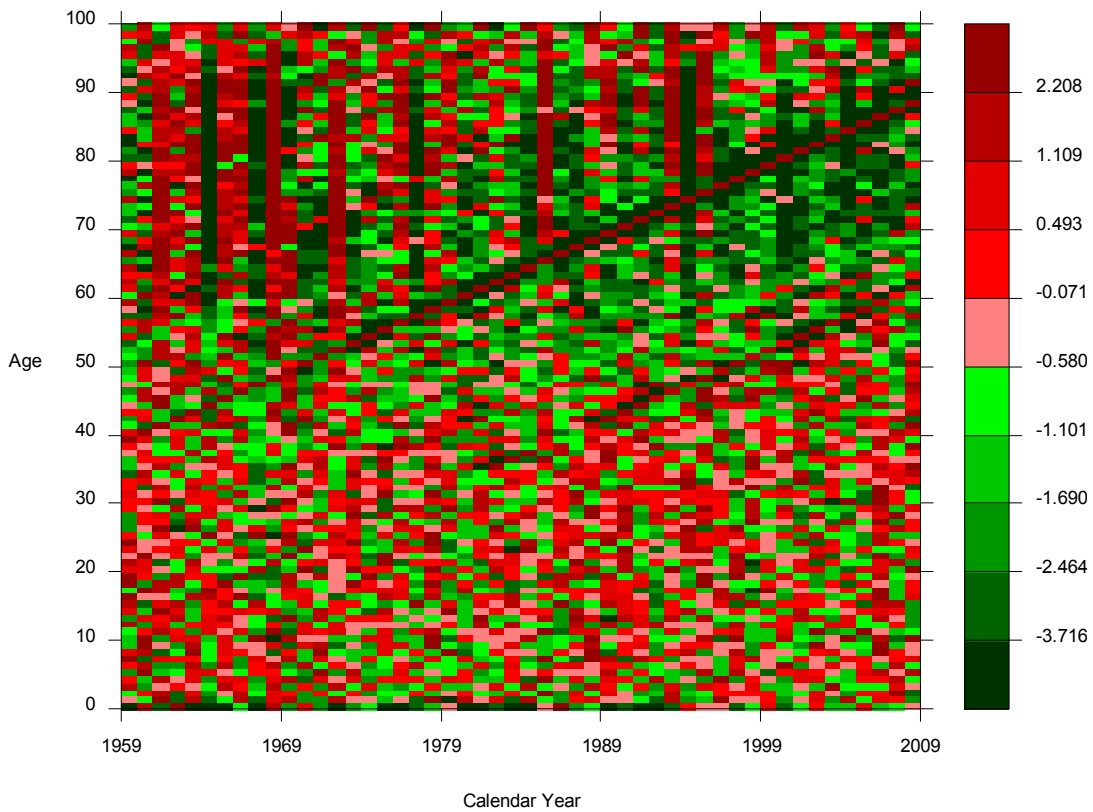


$$m(x,t) = a_x + b_x k_t$$

Data for England and Wales

Separation

- data
- mortality patterns
- residual mortality

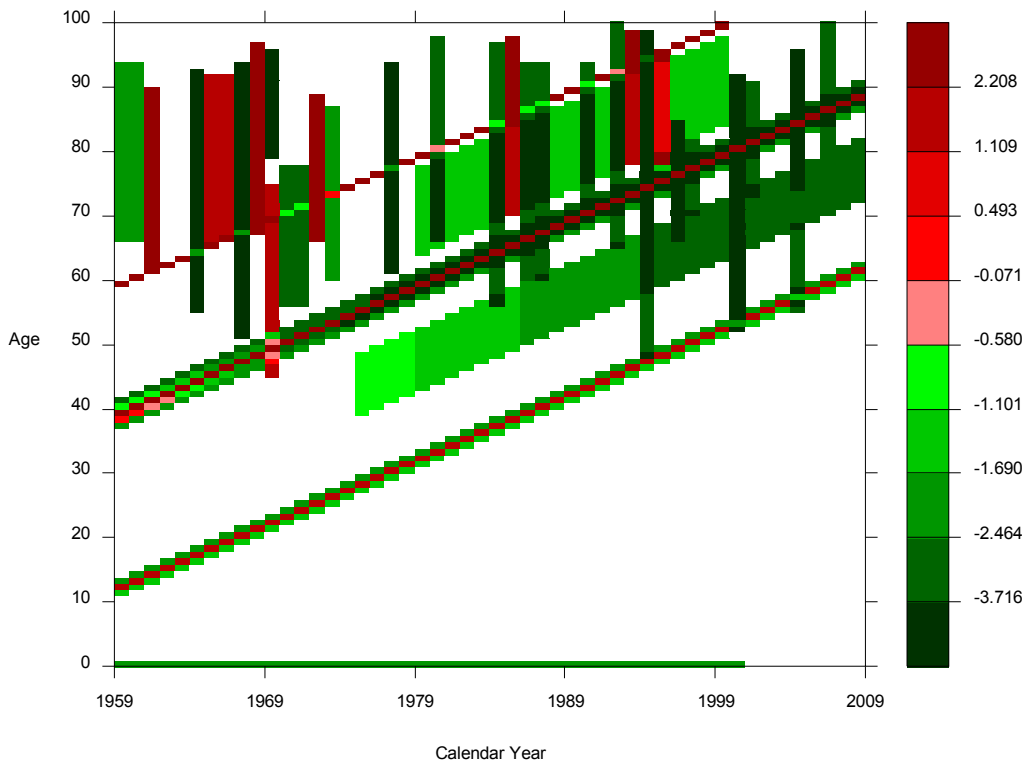


- separation and forecast is based on weighted incremental mortality development factors

Identifying age, period, and cohort effects

Separation

- data
- mortality patterns
- residual mortality

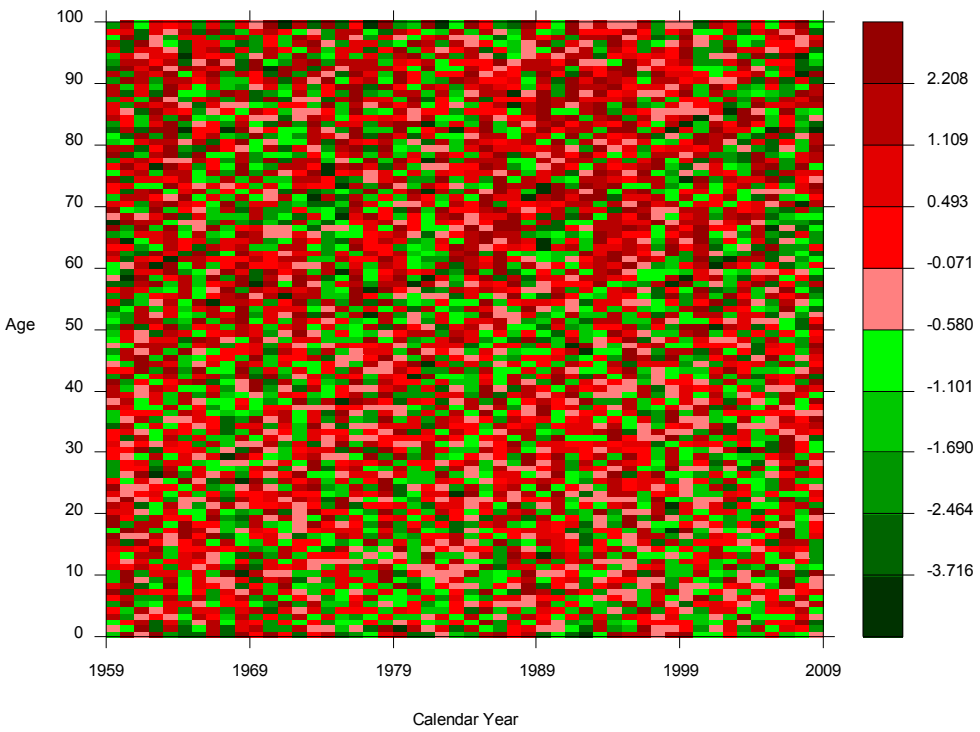


- mortality patterns are identified by the detection algorithm
- it is possible to apply a judgemental approach in addition
- goodness of fit can be measured

Determining residual mortality

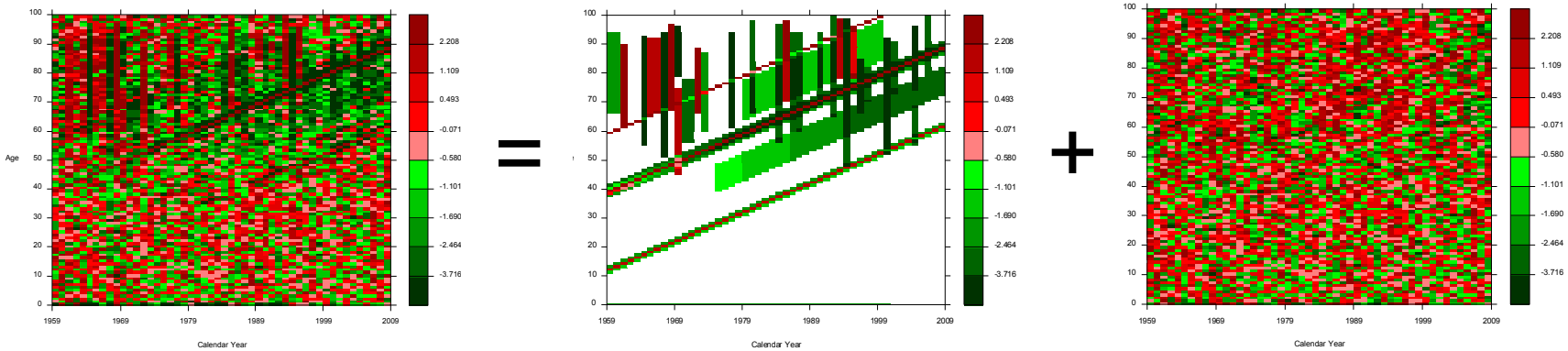
Separation

- data
- mortality patterns
- residual mortality



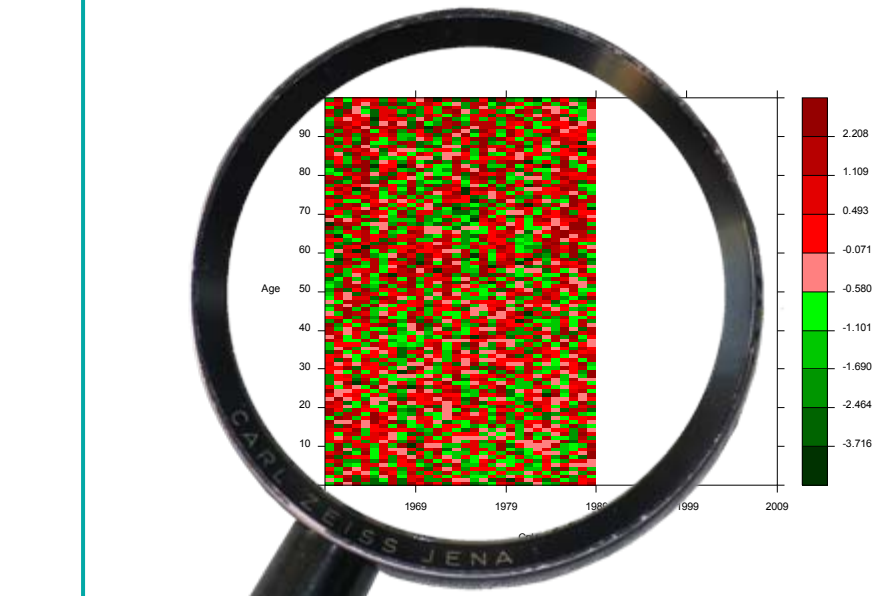
- residuals are obtained by subtraction
- ideally these should be standard normal distributed

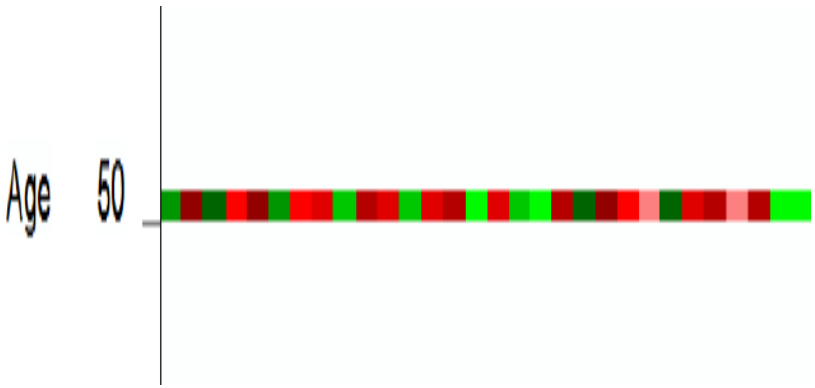
Separation



- the stochastic forecast is based on incremental mortality development factors.
- improvement rates are separated into two parts referring to mortality patterns and residuals respectively.
- statistical tests are performed to measure goodness of fit.
- either part is forecasted separately.

Forecasting residual mortality



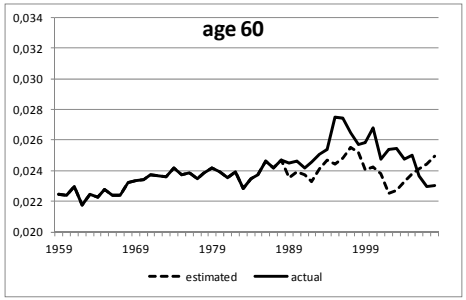


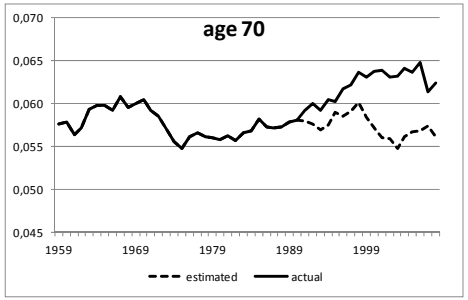
- observation period: 1959 – 1988
- prediction period: 1989 - 2008
- block bootstrap method for each age

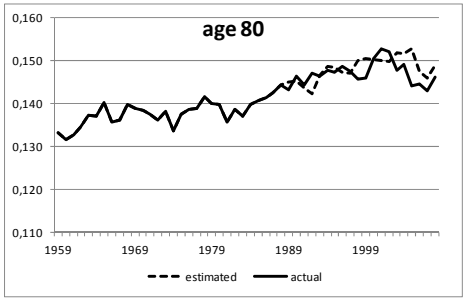
Backtesting residual mortality

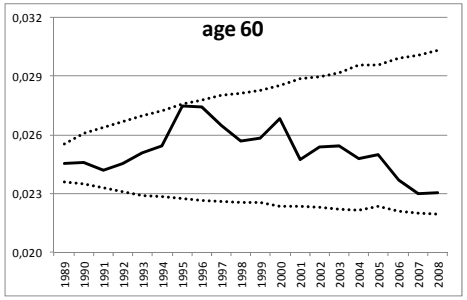
actual mortality versus estimated across time for selected ages

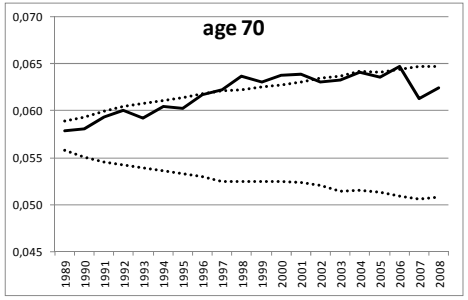
historic data from 1959 - 1988 ; forecast period 1989 - 2008

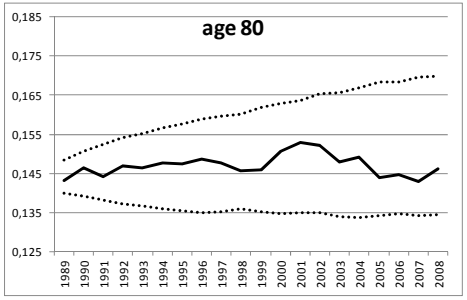










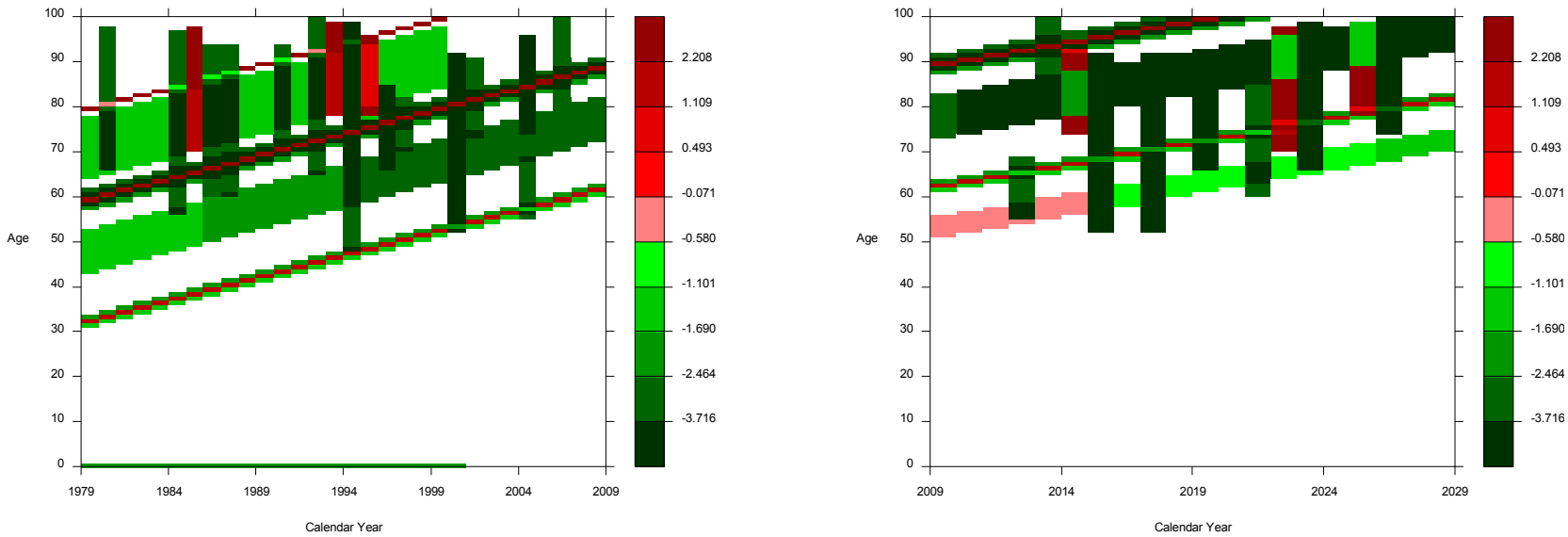


(top) sample path (dashed line) versus actual (solid line)

(bottom) 95%-prediction intervals based on 1.000 bootstrap simulations.

Overall, the percentage of estimates that are smaller than actual mortality is 0.45.

Forecasting mortality patterns



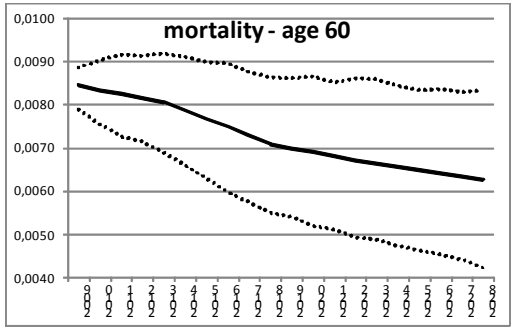
(a) historic mortality patterns (weighted improvement rates) for ages 0−99 during 1979−2008 on the left hand side

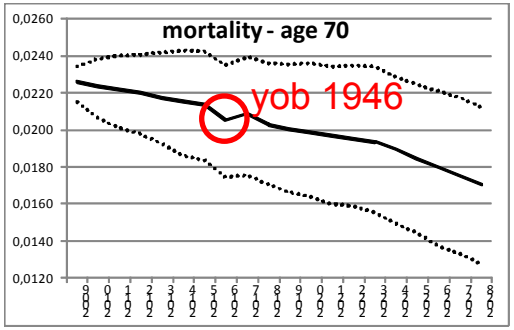
(b) a random sample of simulated mortality patterns for 2009−2028 on the right.

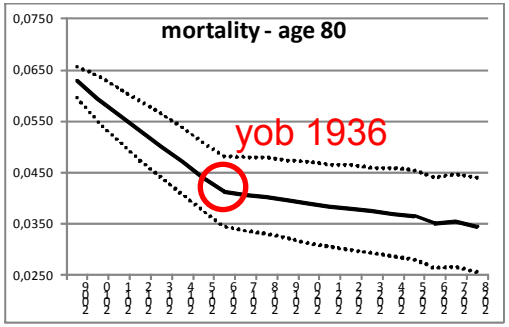
Evaluating mortality forecasts

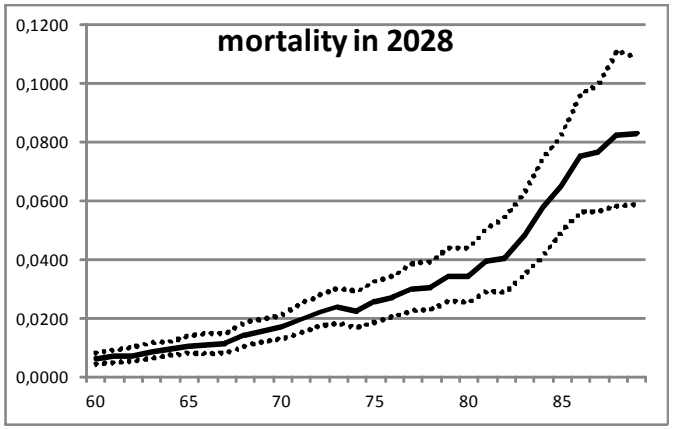
forecasted mean mortality rates across time for selected ages

historic data from 1979 - 2008 ; forecast period 2009 - 2028







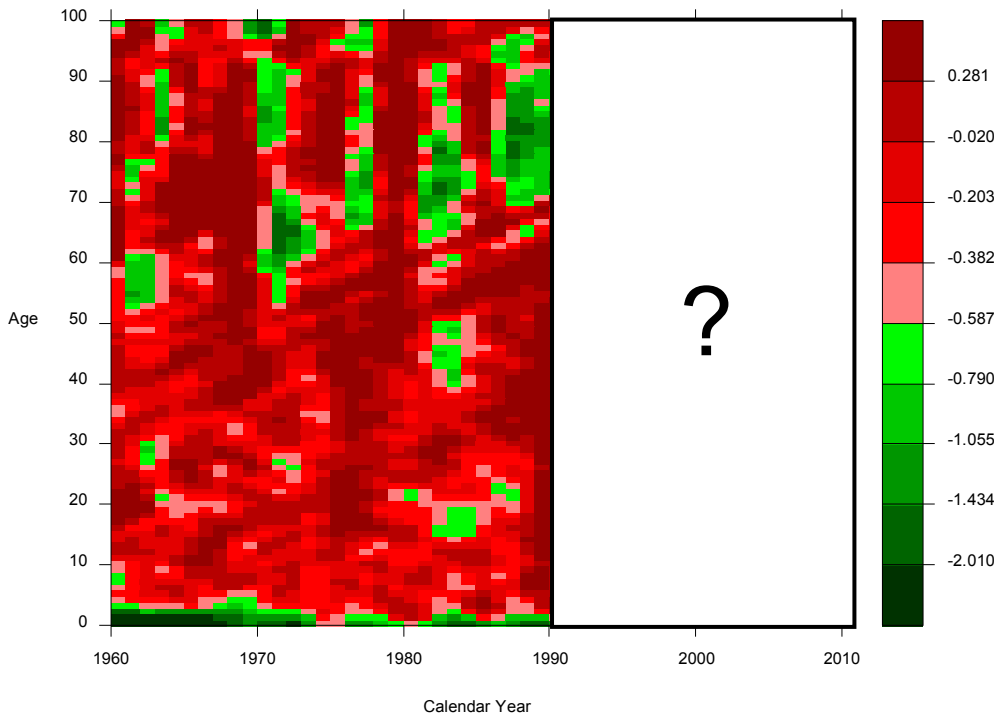


(a) mean forecasted mortality rates (solid line) and 95%-prediction intervals based on 1.000 bootstrap simulations.

(b) mean forecasted mortality rates for ages 60−89

Specific scenarios

How would you have forecasted this data set?
See what happened in actual fact!



weighted and smoothed improvement rates
for males in East Germany from 1960 to 1989 and subsequent realisations

Summary Part II

- Age, period, and cohort effects are separated from the data.
- Mortality patterns and residual mortality are forecasted separately.
- The model allows for incorporating mortality shock scenarios.

Thank you for your attention !

References

Ortmann K. M.: Computerised Detection of Age, Period, and Cohort Effects, available at SSRN:

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1947210

Ortmann K. M.: Stochastic Forecasts of Mortality Effects, available at SSRN:

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1947211