Modelling and Forecasting Mortality in Related Populations: A Comparison

I.L. Danesi[1] S. Haberman[2] P. Millossovich[2]

[1]Department of Statistics, University of Padova,

Faculty of Actuarial Science and Insurance
Cass Business School, City University, London

8^{th} Conference in Actuarial Science
and Finance on Samos — 29 May-1 June 2014

Intro

- forecasting mortality
  - ▷ mortality is a dynamic phenomenon $\rightsquigarrow$ longevity: mortality declining across most countries, non homogeneously with respect to ages
  - ▷ social cost of longevity $\rightsquigarrow$ see IMF [12] report
  - ▷ valuation of life insurance and pension liabilities
  - ▷ socio-demographic studies
  - ▷ wide literature on forecasting mortality in the last 20 years, since the pioneering paper by Lee and Carter

Looking at More than One Population

- Related populations
  - share common features
  - differ in other respects
- issues/advantages in joint mortality forecasting
  - consistency
  - exploit common patterns
  - lower sampling error in small populations
  - convergence?
- examples
  - regions of a country
  - males/females
  - smokers/non smokers
  - annuity/pension fund book vs general population
  - socio economic covariates $\leadsto$ IMD in UK
  - affluence measures $\leadsto$ pension amount, salary
  - ...

Looking at More than One Population

- Related populations
  - share common features
  - differ in other respects
- issues/advantages in joint mortality forecasting
  - consistency
  - exploit common patterns
  - lower sampling error in small populations
  - convergence?
- examples
  - regions of a country
  - males/females
  - smokers/non smokers
  - annuity/pension fund book vs general population
  - socio economic covariates $\rightsquigarrow$ IMD in UK
  - affluence measures $\rightsquigarrow$ pension amount, salary
  - ...

Looking at More than One Population

- Related populations
  - share common features
  - differ in other respects
- issues/advantages in joint mortality forecasting
  - consistency
  - exploit common patterns
  - lower sampling error in small populations
  - convergence?
- examples
  - regions of a country
  - males/females
  - smokers/non smokers
  - annuity/pension fund book vs general population
  - socio economic covariates $\rightsquigarrow$ IMD in UK
  - affluence measures $\rightsquigarrow$ pension amount, salary
  - …

Transfering Longevity Risk

- pension funds are exposed to longevity risk; typical de-risking solutions include
  - ▷ reinsurance
  - ▷ pension scheme buy-in and buy-out
  - ▷ derivative based transactions
- the last 10 years have seen, in UK and North-America, many bespoke derivative transactions (see Blake et al. [13])
  - ▷ $\rightsquigarrow$ perfect hedging
  - ▷ not transparent
  - ▷ costly
  - ▷ unattractive for other parties
- an index based (q forward, longevity swaps) transaction requires modelling the basis to understand the risk reduction $\rightsquigarrow$ multi (2?) population modelling to understand risk reduction (see Li and Hardy [11], Cairns et al. [13], Jarner and Kryger [13])

Transfering Longevity Risk

- pension funds are exposed to longevity risk; typical de-risking solutions include
  - reinsurance
  - pension scheme buy-in and buy-out
  - derivative based transactions
- the last 10 years have seen, in UK and North-America, many bespoke derivative transactions (see Blake et al. [13])
  - $\rightsquigarrow$ perfect hedging
  - not transparent
  - costly
  - unattractive for other parties
- an index based (q forward, longevity swaps) transaction requires modelling the basis to understand the risk reduction $\rightsquigarrow$ multi (2?) population modelling to understand risk reduction (see Li and Hardy [11], Cairns et al. [13], Jarner and Kryger [13])

Transfering Longevity Risk

- pension funds are exposed to longevity risk; typical de-risking solutions include
  - reinsurance
  - pension scheme buy-in and buy-out
  - derivative based transactions
- the last 10 years have seen, in UK and North-America, many bespoke derivative transactions (see Blake et al. [13])
  - $\rightsquigarrow$ perfect hedging
  - not transparent
  - costly
  - unattractive for other parties
- an index based (q forward, longevity swaps) transaction requires modelling the basis to understand the risk reduction $\rightsquigarrow$ multi (2?) population modelling to understand risk reduction (see Li and Hardy [11], Cairns et al. [13], Jarner and Kryger [13])

Different Approaches

- when modelling and forecasting mortality, several approaches are possible
  - ▷ target central death rates, normal errors (Lee-Carter [92]), log link function
  - ▷ target number of deaths as Poisson (Brouhns et al. [02]), log link function
  - ▷ target number of deaths as Binomial (CBD [06]), logit link function
  - ▷ target improvement rates, identity link function
- looking then at the (sub)populations, at least two routes can be followed
  - ▷ 'joint modelling' ⇝ all populations equal
  - ▷ 'relative (hierarchical) modelling' ⇝ one population drives the other(s)
  - ▷ model both sub and general populations ⇝ consistency?

Different Approaches

- when modelling and forecasting mortality, several approaches are possible
  - ▷ target central death rates, normal errors (Lee-Carter [92]), log link function
  - ▷ target number of deaths as Poisson (Brouhns et al. [02]), log link function
  - ▷ target number of deaths as Binomial (CBD [06]), logit link function
  - ▷ target improvement rates, identity link function
- looking then at the (sub)populations, at least two routes can be followed
  - ▷ 'joint modelling' $\rightsquigarrow$ all populations equal
  - ▷ 'relative (hierarchical) modelling' $\rightsquigarrow$ one population drives the other(s)
  - ▷ model both sub and general populations $\rightsquigarrow$ consistency?

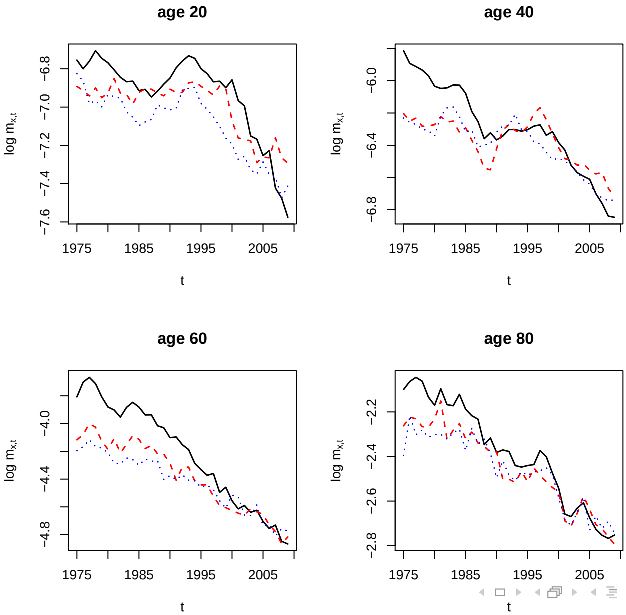
What we do

- extend (Poisson) Lee-Carter to several populations ⇒ focus on 5 specific examples
- model improvement rates for several populations and the 5 equivalent models
- focus on simple, straightforward extension of the basic Lee-Carter
- estimate and compare the 10 models on a data set of mortality data for 18 regions of Italy

What we do

- extend (Poisson) Lee-Carter to several populations ⇒ focus on 5 specific examples
- model improvement rates for several populations and the 5 equivalent models
- focus on simple, straightforward extension of the basic Lee-Carter
- estimate and compare the 10 models on a data set of mortality data for 18 regions of Italy

Mortality - 3 regions



Data

- I populations — $i = 1, \ldots, I$
- T calendar years — $t = t_0, t_0 + 1, \ldots, t_0 + T - 1$
- X age groups — $x = x_0, x_0 + 1, \ldots, x_0 + X - 1$
- for population i, year t, age group x, we have

$$d^i_{x,t} = \text{ number of deaths in } [t, t+1) \text{ aged } x \text{ last birthday,}$$

$$\mathrm{ETR}^i_{x,t} = \text{ central exposed to risk}$$

$$\Rightarrow \text{central death rate } m^i_{x,t} = \frac{d^i_{x,t}}{\mathrm{ETR}^i_{x,t}}$$

Data

- I populations — $i = 1, \ldots, I$
- T calendar years — $t = t_0, t_0 + 1, \ldots, t_0 + T - 1$
- X age groups — $x = x_0, x_0 + 1, \ldots, x_0 + X - 1$
- for population i, year t, age group x, we have

$$d^i_{x,t} = \text{ number of deaths in } [t, t+1) \text{ aged } x \text{ last birthday,}$$

$$\text{ETR}^i_{x,t} = \text{ central exposed to risk}$$

$$\Rightarrow \text{central death rate } m^i_{x,t} = \frac{d^i_{x,t}}{\text{ETR}^i_{x,t}}$$

Poisson Model

- number of deaths $d^i_{x,t}$ realizations of $D^i_{x,t}$ (Brouhns et al [02])
- Cox (doubly-stochastic model): conditionally on $(m^i_{x,t})_{x,t,i}$, the number of deaths
  - are independent
  - have distribution

$$D^i_{x,t} \sim \text{Poisson}(\text{ETR}^i_{x,t}\, m^i_{x,t})$$

- we model $m^i_{x,t}$ as follows (Hyndman and Ullah [06]) assuming there are L time indices

$$\log m^i_{x,t} = \alpha^i_x + \sum_{j=1}^{L} \beta^i_{x,j}\, k_{t,j}$$

- idea:
  - number of factors L related to I
  - choose $k_{t,j}$ appropriately
  - add identifiability constraints as appropriate

Poisson Model

- number of deaths $d^i_{x,t}$ realizations of $D^i_{x,t}$ (Brouhns et al [02])
- Cox (doubly-stochastic model): conditionally on $(m^i_{x,t})_{x,t,i}$, the number of deaths
  - are independent
  - have distribution

$$D^i_{x,t} \sim \text{Poisson}(\text{ETR}^i_{x,t}\, m^i_{x,t})$$

- we model $m^i_{x,t}$ as follows (Hyndman and Ullah [06]) assuming there are L time indices

$$\log m^i_{x,t} = \alpha^i_x + \sum_{j=1}^{L} \beta^i_{x,j}\, k_{t,j}$$

- idea:
  - number of factors L related to I
  - choose $k_{t,j}$ appropriately
  - add identifiability constraints as appropriate

Poisson Models. . .

- $\mathcal{P}1$ — (Booth et al. [02], Haberman and Renshaw [03])

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1}\, k^i_{t,1} + \beta^i_{x,2}\, k^i_{t,2}$$

- $\mathcal{P}2$ — (Augmented Common Factor, Li and Lee [05], Li and Hardy [11], Hyndman et al. [13]): $k_{t,1}$ common factor, $k^i_{t,2}$ i^{th} population specific factor

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1}\, k_{t,1} + \beta^i_{x,2}\, k^i_{t,2}$$

- $\mathcal{P}3$ — (Lee and Carter [1992], Li and Hardy [11])

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x\, k^i_t$$

each group has its own factor $k^i_t \Rightarrow$ correlation arises from the modelling of $(k^i_t)_i$

Poisson Models…

- $\mathcal{P}1$ — (Booth et al. [02], Haberman and Renshaw [03])

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1}\, k^i_{t,1} + \beta^i_{x,2}\, k^i_{t,2}$$

- $\mathcal{P}2$ — (Augmented Common Factor, Li and Lee [05], Li and Hardy [11], Hyndman et al. [13]): $k_{t,1}$ common factor, $k^i_{t,2}$ i^{th} population specific factor

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1}\, k_{t,1} + \beta^i_{x,2}\, k^i_{t,2}$$

- $\mathcal{P}3$ — (Lee and Carter [1992], Li and Hardy [11])

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x\, k^i_t$$

each group has its own factor $k^i_t \Rightarrow$ correlation arises from the modelling of $(k^i_t)_i$

Poisson Models…

- $\mathcal{P}1$ — (Booth et al. [02], Haberman and Renshaw [03])

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1}\, k^i_{t,1} + \beta^i_{x,2}\, k^i_{t,2}$$

- $\mathcal{P}2$ — (Augmented Common Factor, Li and Lee [05], Li and Hardy [11], Hyndman et al. [13]): $k_{t,1}$ common factor, $k^i_{t,2}$ i^{th} population specific factor

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1}\, k_{t,1} + \beta^i_{x,2}\, k^i_{t,2}$$

- $\mathcal{P}3$ — (Lee and Carter [1992], Li and Hardy [11])

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x\, k^i_t$$

each group has its own factor $k^i_t \Rightarrow$ correlation arises from the modelling of $(k^i_t)_i$

Poisson Models…

- $\mathcal{P}4$ — $J_1, J_2, \ldots, J_g$ partition of $\{1, \ldots, I\}$ $(I > 2)$
  - ▷ some populations are 'more related' than others
  - ▷ g time indices, one for each of subgroup $\Rightarrow$ reduce number of parameters
  - ▷ if $i \in J_h$ then the time index is k_t^h

  for $i \in J_h$

$$\log m_{x,t}^i = \alpha_x^i + \beta_x^i k_t^h$$

- in our case, choose (see next slide) $J_1 = \{1, 2, 3, 4, 5, 6\}$, $J_2 = \{7, 8, 9\}$, $J_3 = \{10, 11, 12\}$, $J_4 = \{13, 18\}$, $J_5 = \{14, 15, 16, 17\}$ $\rightsquigarrow$ clustering obtained by similarity with respect to period life expectancy at birth

Poisson Models…

Figure: Italy divided in the considered 18 areas.

Poisson Models...

- $\mathcal{P}5$ — (Joint *K* Model, Carter and Lee [92], Li and Hardy [11], Wilmoth and Valkonen [01], Delwarde et al. [06]) a single time index driving all the rates $\Rightarrow$ perfect correlation

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x k_t$$

- models are nested: $\mathcal{P}5 \subset \mathcal{P}4 \subset \mathcal{P}3 \subset \mathcal{P}2 \subset \mathcal{P}1$

Improvement Rates

- use mortality improvement rates rather than rates ⇒ slope vs level
- used recently by
  - ▷ Willets [04]
  - ▷ Richards et al. [05]
  - ▷ Baxter [07]
  - ▷ Haberman and Renshaw [12,13]
  - ▷ Mitchell et al [13]
- idea
  - ▷ detrend the data
  - ▷ model each rate in terms of the previous year's one
  - ▷ provides an alternative route: given m's, transform, model, estimate and forecast the improvement rates, transform back

Improvement Rates

- use mortality improvement rates rather than rates ⇒ slope vs level
- used recently by
  - ▷ Willets [04]
  - ▷ Richards et al. [05]
  - ▷ Baxter [07]
  - ▷ Haberman and Renshaw [12,13]
  - ▷ Mitchell et al [13]
- idea
  - ▷ detrend the data
  - ▷ model each rate in terms of the previous year's one
  - ▷ provides an alternative route: given m's, transform, model, estimate and forecast the improvement rates, transform back

Improvement Rates

- use mortality improvement rates rather than rates ⇒ slope vs level
- used recently by
  - Willets [04]
  - Richards et al. [05]
  - Baxter [07]
  - Haberman and Renshaw [12,13]
  - Mitchell et al [13]
- idea
  - detrend the data
  - model each rate in terms of the previous year's one
  - provides an alternative route: given m's, transform, model, estimate and forecast the improvement rates, transform back

Improvement Rates

- suppress the population index i here; define for $t = t_0 + 1, \ldots, t_0 + T - 1$ ($T - 1$ calendar years) the relative improvement rates (Haberman and Renshaw [12])

$$z_{x,t} = \frac{m_{x,t-1} - m_{x,t}}{\frac{1}{2}\left(m_{x,t-1} + m_{x,t}\right)} = 2\frac{1 - \frac{m_{x,t}}{m_{x,t-1}}}{1 + \frac{m_{x,t}}{m_{x,t-1}}}$$

- note that
  - given z, recover m

$$m_{x,t} = m_{x,t-1}\frac{2 - z_{x,t}}{2 + z_{x,t}}$$

  - z is the (discrete version of the) time derivative of m

$$z_{x,t} \approx \frac{1}{m_{x,t}}\frac{\partial m_{x,t}}{\partial t}$$

Improvement Rates

- suppress the population index i here; define for $t = t_0 + 1, \ldots, t_0 + T - 1$ ($T - 1$ calendar years) the relative improvement rates (Haberman and Renshaw [12])

$$z_{x,t} = \frac{m_{x,t-1} - m_{x,t}}{\frac{1}{2}(m_{x,t-1} + m_{x,t})} = 2\frac{1 - \frac{m_{x,t}}{m_{x,t-1}}}{1 + \frac{m_{x,t}}{m_{x,t-1}}}$$

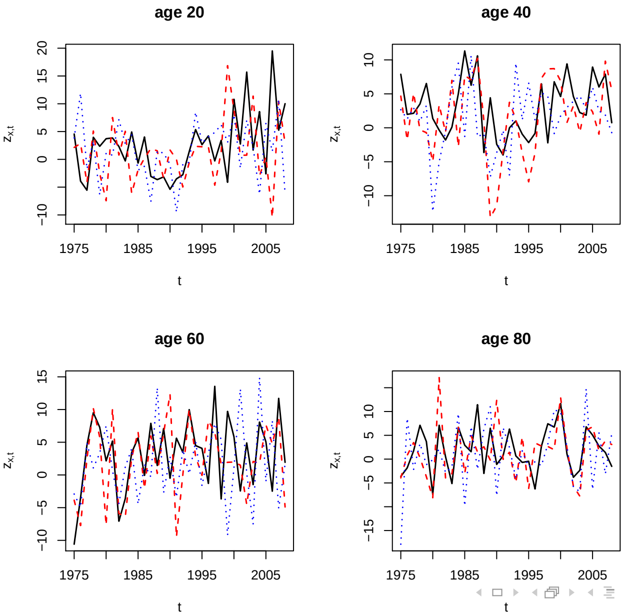
- note that
  - given z, recover m

$$m_{x,t} = m_{x,t-1}\frac{2 - z_{x,t}}{2 + z_{x,t}}$$

  - z is the (discrete version of the) time derivative of m

$$z_{x,t} \approx \frac{1}{m_{x,t}} \frac{\partial m_{x,t}}{\partial t}$$

. . . Improvement Rates



. . . Improvement Rates

- Assume $z^i_{x,t}$ are realizations of iid rv $Z^i_{x,t}$ with

$$Z^i_{x,t} \sim N(\eta^i_{x,t}, \sigma_i)$$

- similarly to death rates

$$\eta^i_{x,t} = \sum_{j=1}^{L} \beta^i_{x,j} k_{t,j}$$

- note that there is no ‘time-average’ α^i_x term

…Improvement Rates

- $\mathcal{I}1$

$$\eta^i_{x,t} = \beta^i_{x,1}\, k^i_{t,1} + \beta^i_{x,2}\, k^i_{t,2}$$

- $\mathcal{I}2$

$$\eta^i_{x,t} = \beta^i_{x,1}\, k_{t,1} + \beta^i_{x,2}\, k^i_{t,2}$$

- $\mathcal{I}3$

$$\eta^i_{x,t} = \beta^i_x\, k^i_t$$

- $\mathcal{I}4$ — $J_1, J_2, \ldots, J_g$ partition of $\{1, \ldots, I\}$ $(I > 2)$; for $i \in J_h$

$$\eta^i_{x,t} = \beta^i_x\, k^h_t$$

- $\mathcal{I}5$

$$\eta^i_{x,t} = \beta^i_x\, k_t$$

- models are nested: $\mathcal{I}5 \subset \mathcal{I}4 \subset \mathcal{I}3 \subset \mathcal{I}2 \subset \mathcal{I}1$

Application

- use mortality data from $I = 18$ Italian regions
- ages $20 - 89$ ($X = 70$)
- years 1974-2008, use 1974-1999 ($T = 26$) for estimation, 2000-2008 (9 yr) for forecasting
- maximize likelihood
  - $\mathcal{P}$ models:

$$l = K + \sum_{x,t,i} \left(d^i_{x,t} \log m^i_{x,t} - \mathrm{ETR}^i_{x,t}\, m^i_{x,t} \right)$$

  - $\mathcal{I}$ models:

$$l = -\frac{1}{2} \sum_{x,t,i} \left(\log(2\pi\sigma^2_i) + \frac{(z^i_{x,t} - \eta^i_{x,t})^2}{\sigma^2_i} \right)$$

Application

- use mortality data from $I = 18$ Italian regions
- ages $20 - 89$ ($X = 70$)
- years 1974-2008, use 1974-1999 ($T = 26$) for estimation, 2000-2008 (9 yr) for forecasting
- maximize likelihood
  - ▷ $\mathcal{P}$ models:

$$l = K + \sum_{x,t,i} \left(d^i_{x,t} \log m^i_{x,t} - \mathrm{ETR}^i_{x,t}\, m^i_{x,t} \right)$$

  - ▷ $\mathcal{I}$ models:

$$l = -\frac{1}{2} \sum_{x,t,i} \left(\log(2\pi\sigma_i^2) + \frac{(z^i_{x,t} - \eta^i_{x,t})^2}{\sigma_i^2} \right)$$

Comparison - Goodness of Fit

- $AIC = 2(np - l^*)$, $AIC^c = AIC + \frac{2np(np+1)}{nd-np-1}$, $BIC = np\log(nd) - 2l^*$ (l^* = maximized likelihood, np = number of estimable parameters, nd = number of data)

	AIC		AIC^c		BIC	
1st	$\mathcal{P}2$	$\mathcal{I}1$	$\mathcal{P}2$	$\mathcal{I}1$	$\mathcal{P}5$	$\mathcal{I}2$
2nd	$\mathcal{P}1$	$\mathcal{I}2$	$\mathcal{P}4$	$\mathcal{I}2$	$\mathcal{P}4$	$\mathcal{I}3$
3rd	$\mathcal{P}4$	$\mathcal{I}5$	$\mathcal{P}3$	$\mathcal{I}3$	$\mathcal{P}2$	$\mathcal{I}5$
4th	$\mathcal{P}3$	$\mathcal{I}3$	$\mathcal{P}1$	$\mathcal{I}4$	$\mathcal{P}3$	$\mathcal{I}4$
5th	$\mathcal{P}5$	$\mathcal{I}4$	$\mathcal{P}5$	$\mathcal{I}5$	$\mathcal{P}1$	$\mathcal{I}1$

Forecast

- if $(k^i_{t,j})$ in $\mathcal{P}$ is modelled using a VARIMA process, then the corresponding $(k^i_{t,j})$ in $\mathcal{I}$ should be modelled using a VARMA process
- in the application, we use
  - a multivariate random walk with drift for $\mathcal{P}$, where appropriate
  - a VAR(1) for $\mathcal{I}$, where appropriate

Forecast

- if $(k^i_{t,j})$ in $\mathcal{P}$ is modelled using a VARIMA process, then the corresponding $(k^i_{t,j})$ in $\mathcal{I}$ should be modelled using a VARMA process
- in the application, we use
  - ▷ a multivariate random walk with drift for $\mathcal{P}$, where appropriate
  - ▷ a VAR(1) for $\mathcal{I}$, where appropriate

Comparison - Forecast

- Compare truncated expected cohort residual lifetimes (MAPE in % across all regions)

	$\mathcal{P}1$	$\mathcal{P}2$	$\mathcal{P}3$	$\mathcal{P}4$	$\mathcal{P}5$
$e^{\text{cohort}}_{60:\overline{9}\rceil}$	0.06	0.11	0.12	0.18	0.50
$e^{\text{cohort}}_{70:\overline{9}\rceil}$	0.11	0.20	0.20	0.21	0.88
$e^{\text{cohort}}_{80:\overline{9}\rceil}$	0.29	0.43	0.38	0.45	1.46

	$\mathcal{I}1$	$\mathcal{I}2$	$\mathcal{I}3$	$\mathcal{I}4$	$\mathcal{I}5$
$e^{\text{cohort}}_{60:\overline{9}\rceil}$	0.21	0.24	0.21	0.16	0.19
$e^{\text{cohort}}_{70:\overline{9}\rceil}$	0.39	0.24	0.39	0.27	0.26
$e^{\text{cohort}}_{80:\overline{9}\rceil}$	0.64	0.75	0.72	0.74	0.72

Comparison - Forecast

- Compare truncated expected cohort residual lifetimes (MAPE in % across all regions)

	$\mathcal{P}1$	$\mathcal{P}2$	$\mathcal{P}3$	$\mathcal{P}4$	$\mathcal{P}5$
$e^{\text{cohort}}_{60:\overline{9}\rceil}$	0.06	0.11	0.12	0.18	0.50
$e^{\text{cohort}}_{70:\overline{9}\rceil}$	0.11	0.20	0.20	0.21	0.88
$e^{\text{cohort}}_{80:\overline{9}\rceil}$	0.29	0.43	0.38	0.45	1.46

	$\mathcal{I}1$	$\mathcal{I}2$	$\mathcal{I}3$	$\mathcal{I}4$	$\mathcal{I}5$
$e^{\text{cohort}}_{60:\overline{9}\rceil}$	0.21	0.24	0.21	0.16	0.19
$e^{\text{cohort}}_{70:\overline{9}\rceil}$	0.39	0.24	0.39	0.27	0.26
$e^{\text{cohort}}_{80:\overline{9}\rceil}$	0.64	0.75	0.72	0.74	0.72

Conclusion

- in terms of goodness of fit,
  - when modelling death rates, more elaborate models seems to be preferable
  - however, when more weight is put on the number of parameters, the ranking is reverted
  - when modelling improvement rates, more complex models are at advantage
- in terms of out of sample forecast
  - when targeting death rates, again more elaborate models provide better forecast
  - when targeting improvement rates, simpler models are as good as elaborate ones

Conclusion

- in terms of goodness of fit,
  - when modelling death rates, more elaborate models seems to be preferable
  - however, when more weight is put on the number of parameters, the ranking is reverted
  - when modelling improvement rates, more complex models are at advantage
- in terms of out of sample forecast
  - when targeting death rates, again more elaborate models provide better forecast
  - when targeting improvement rates, simpler models are as good as elaborate ones

Thanks for Your attention!