Mortality estimation and prediction: Models, Methods and Issues

Jon Forster
Mathematical Sciences & ESRC Centre for Population Change
University of Southampton, UK

How should pension liabilities be valued? Risk aversion and demographic uncertainty
25–26 March 2016, Royal Society, London

Joint work with Jakub Bijak, Erengul Dodd, Jason Hilton and Peter Smith

Why mortality matters

Consider an annuity paying £A now and annually to death, for an individual current age x.

The expected present value of this annuity is

$$E = A \sum_{k=0}^{\infty} v_k \, {}_k p_x$$

where v_k is the expected present value of k years hence £1 , and

$${}_k p_x = P(T > x + k | T > x) = \frac{P(T > x + k)}{P(T > x)} = \frac{\ell_{x+k}}{\ell_x}$$

where T is the random variable denoting age at death.

Knowledge of survivorship probabilities is essential.

A static summary of the distribution of age at death for a population:

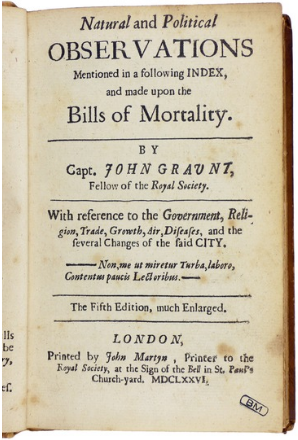
English Life Tables No 17

Period expectation of life
Based on data for England and Wales for the years 2010-2012

Age	Males							
x	m_x	q_x	l_x	d_x	L_x	T_x	μ_x	e_x
0	0.004757	0.004746	100000	475	99576.3	7896837		78.97
1	0.000306	0.000306	99525	30	99510.2	7797072	0.000369	78.34
2	0.000207	0.000207	99495	21	99484.6	7697562	0.000246	77.37
3	0.000147	0.000147	99474	14	99467.0	7598078	0.000172	76.38
4	0.000115	0.000115	99460	12	99453.9	7498612	0.000128	75.39
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
109	0.676172	0.491440	8	4	5.5	11	0.661588	1.43
110	0.701065	0.503943	4	2	2.8	5	0.685990	1.39
111	0.725677	0.516003	2	1	1.4	3	0.709972	1.34
112	0.750015	0.528125	1	1	0.7	1	0.733841	1.30

John Graunt and the first life table

UNIVERSITY OF Southampton



Natural and Political
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

BY
Capt. JOHN GRAUNT,
Fellow of the Royal Society.

With reference to the Government, Religion, Trade, Growth, Air, Diseases, and the several Changes of the said CITY.

——Non, me ut miretur Turba, laboro,
Contentus paucis Lectoribus.——

The Fifth Edition, much Enlarged.

LONDON,
Printed by John Martyn, Printer to the Royal Society, at the Sign of the Bell in St. Paul's Church-yard. MDCLXXVI.

John Graunt
(1620-1674)

“From whence it follows, that of the said 100 conceived there remains alive at six years end 64

At Sixteen years end	40	At Fifty six	6
At Twenty six	25	At Sixty six	3
At Thirty six	16	At Seventy six	1
At Fourty six	10	At Eighty	0”

$\{y_x\}$ – number of male (female) deaths in England and Wales observed aged x at last birthday, in a given time period.

$\{E_x^C\}$ – corresponding central exposed to risk for age x at last birthday

The observed (or crude) central mortality rate is

$$\tilde{m}_x = \frac{y_x}{E_x^C}.$$

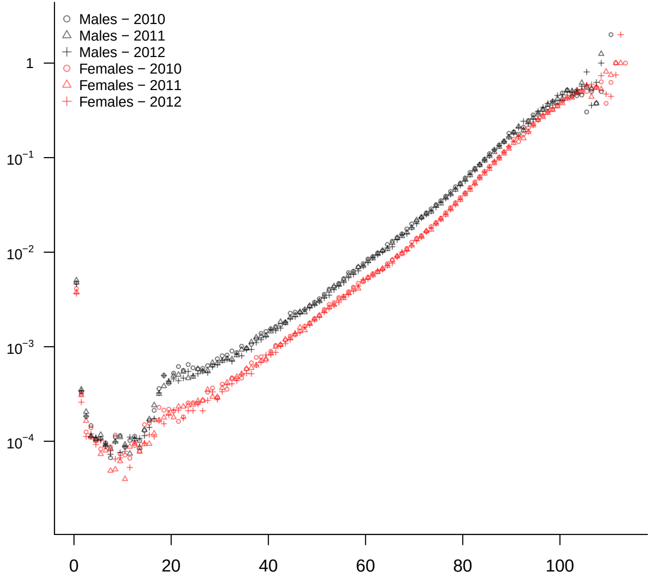
This is an estimator of the underlying central mortality rate

$$m_x = \frac{E[Y_x]}{E_x^C}$$

under any model for $\{Y_x\}$.

Other quantities, survival probabilities ${}_kp_x$, death probabilities ${}_kq_x \equiv 1 - {}_kp_x$ etc can be derived from m_x (by approximation/assumption).

Crude mortality rates 2010-2012



A basic smoothing model

As this is a large inhomogenous population, we propose a negative binomial model

$$Y_x \sim \mathrm{NB}\left(E_x^C m_x,\ \alpha\right)$$

where $E[Y_x] = E_x^C m_x$ and $Var[Y_x] = E_x^C m_x + (E_x^C m_x)^2/\alpha.$

Then, in a generalised additive (smooth) model

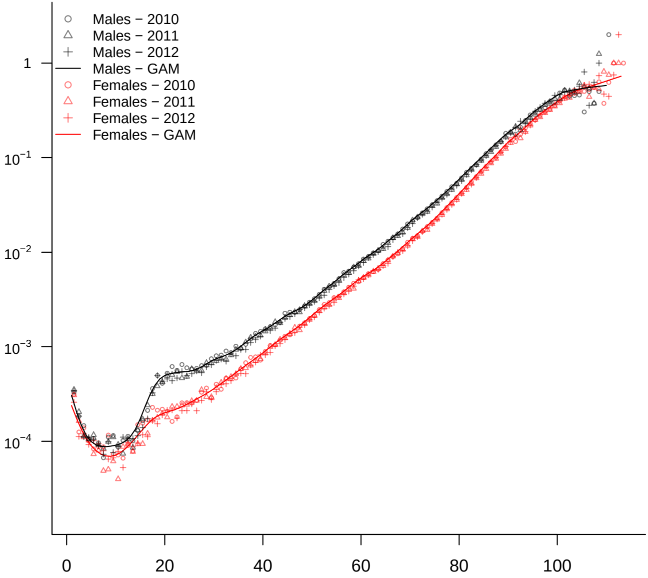
$$\log m_x = s(x; \beta)$$

where $s(x; \beta)$ is a linear (in β) function representing regression on a spline basis.

The graduated estimates $\hat{m}_x$ are obtained as

$$\hat{m}_x = \exp s(x; \hat{\beta})$$

Smooth mortality rates 2010-2012



Models for older ages and extrapolation (1)

To obtain a more robust fit at older ages, and to extrapolate the mortality function m_x beyond the range of the observed data, one might use a parametric model.

Only parsimonious models considered, as data are sparse.

The simplest obvious choice is the log-linear Gompertz model

$$\log m_x = \beta_0 + \beta_1 x, \qquad x \geq x_0$$

where x_0 is a suitable threshold

A competing extrapolation model is a logistic model (Beard, 1963)

$$m_x = \frac{\beta_2 \exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}, \qquad x \geq x_0$$

where mortality rates flatten off, converging to the limit β_2 as $x \to \infty$. Arises naturally as Gompertz with frailty.

A special case of this model, with $\beta_2 = 1$, (Thatcher et al, 1998) is used in graduating the human mortality data base (Wilmoth et al 2007).

A possible model across the entire range of x is

$$m_x = \begin{cases} \exp s(x;\beta) & x < x_0 \\ \dfrac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} & x \geq x_0 \end{cases}$$

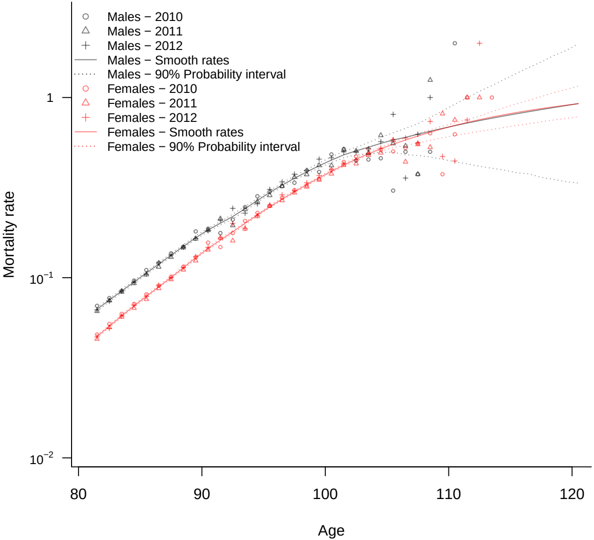
Model uncertainty

Hence, we have two possible models, log-linear and logistic both of which require the choice of a threshold age x_0 to determine the age range over which the parametric component will be fitted, and applied.

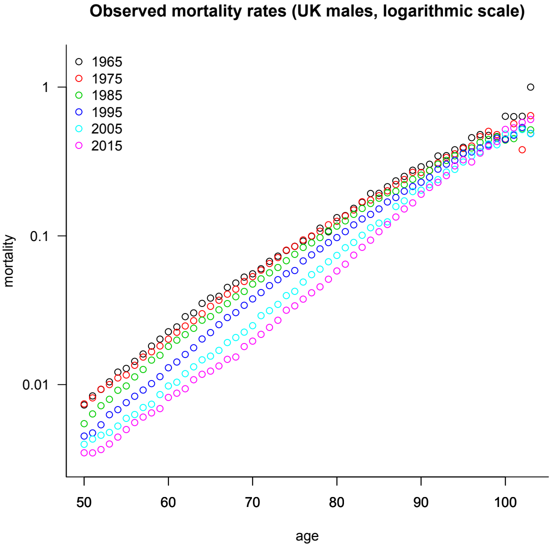
- No fundamental reason to prefer one model over the other, or to apply a particular value of x_0.
- Rather, we should base our decision on the observed data.
- Given the sparsity of the data at the highest ages, there is considerable uncertainty about this choice. Graduation should acknowledge this uncertainty.

A natural approach for incorporation of model uncertainty into estimates is a Bayesian approach.

ELT17 model-averaged graduation



Mortality improvement



$\{y_{xt}\}$ – number of observed deaths aged x at last birthday, in year t, in population of interest, for $t = 1, \ldots, T$.

$\{E^C_{xt}\}$ – corresponding central exposed to risk

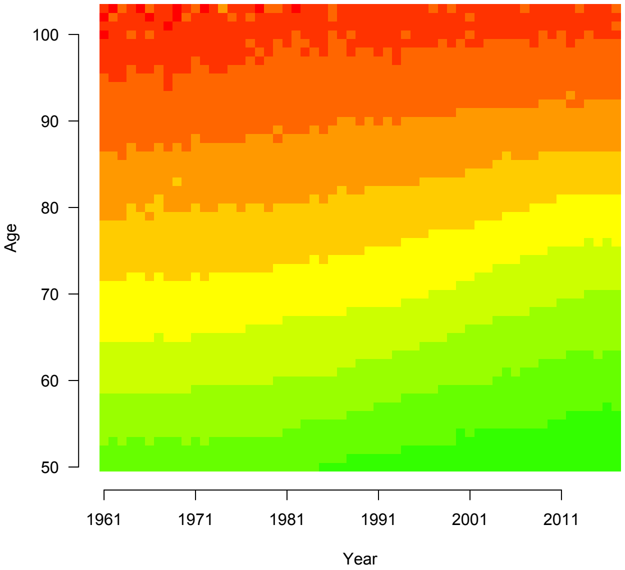
Mortality models provide a framwork for estimating the central mortality rates

$$m_{xt} = \frac{E[Y_{xt}]}{E^C_{xt}}.$$

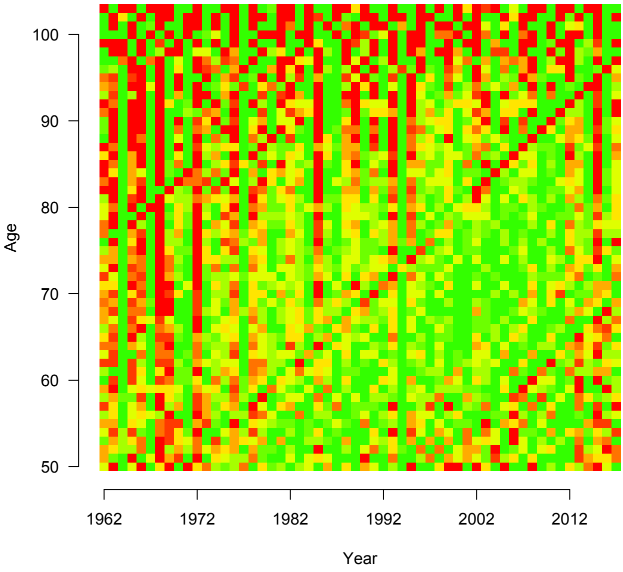
based on the data array $\{y_{xt}\}$ and computing relevant estimates for ${}_kp_x$, ${}_kq_x$ etc and $\ldots$

$\ldots$ projecting m_{xt} etc for $t = T+1, T+2, \ldots$

UK mortality rates (males, 1961-2017)



UK mortality improvements



Models for central mortality rates m_{xt} over age x and time t generally have the form:

$$g(m_{xt}) = \underset{\text{age baseline}}{f_1(x)} + \underset{\text{common period effect}}{f_2(t)} + \underset{\text{age-period interaction}}{f_3(x,t)}$$

where typically $g(m_{xt}) = \log m_{xt}$ or $g(m_{xt}) = \log(e^{m_{xt}} - 1)$

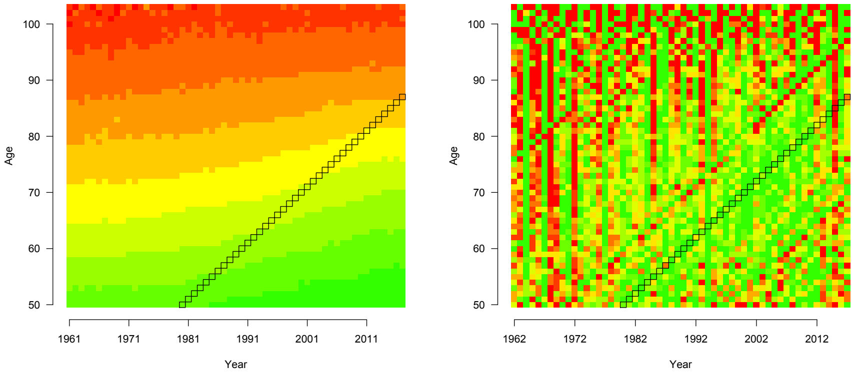
For projection, $f_2(t) + f_3(x,t)$ need to be able to be extrapolated for $t = T+1, T+2, \ldots$ (structure is required)

The most venerable model is the Lee-Carter (1992; generalised bilnear) model

$$\log m_{xt} = \alpha_x + \beta_x \kappa_t$$

[multiplicative age-period interaction, κ_t structured as $\kappa_t = \mu t + RW(0, \sigma^2)$]

Cohorts



A cohort is a subpopulation sharing a common birth-year. (1930 birth cohort identified above)

A *cohort effect* is a structured age-period interaction.

Selected mortality models

Models for central mortality rates m_{xt} over age x and time t include:

- Lee Carter with cohort (Renshaw and Haberman, 2006)

$$\log m_{xt} = \alpha_x + \beta_x \kappa_t + \gamma_{t-x}$$

- CBD generalised linear (Cairns et al, various)

$$\log(e^{m_{xt}} - 1) = \kappa_t^{(1)} + x\kappa_t^{(2)} + x^2\kappa_t^{(3)} + \gamma_{t-x}$$

- APCI generalised linear (CMI 2016, Richards et al, 2017)

$$\log m_{xt} = \alpha_x + t\beta_x + \kappa_t + \gamma_{t-x}$$

- generalised additive (GAM)

$$\log m_{xt} = s_\alpha(x) + t\, s_\beta(x) + \kappa_t + s_\gamma(t-x)$$

Cohort effect caveat

Well-understood lack of identifiability or the classic age-period-cohort (APC) model

$$\alpha_x + \beta_t + \gamma_{t-x} \quad = \quad [\alpha_x + \mu x] + [\beta_t - \mu t] + [\gamma_{t-x} + \mu(t-x)]$$

Even models which are algebraically identified may be prone to 'awkward behaviour' due to complex APC dependence (e.g. Palin, 2016, everyone!)

Attribution of linear effect (by constraint) may even be a benefit for forecasting?

The model from here ...

Age-period-cohort (APC) GAM for mortality improvements

$$\log \frac{m_{xt}}{m_{x\,t-1}} = s_\alpha(x) + \kappa_t + s_\gamma(t-x)$$

or 'equivalently' APCI GAM for mortality rates

$$\log m_{xt} = s_\mu(x) + s_\alpha(x)t + \kappa_t + s_\gamma(t-x).$$

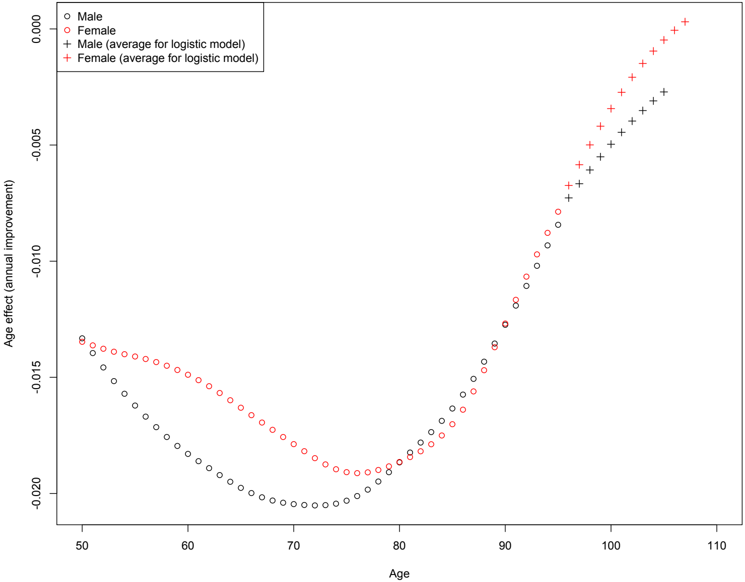
where s_μ, s_α and s_γ are arbitrary smooth functions.

For the highest ages x, use parametric model

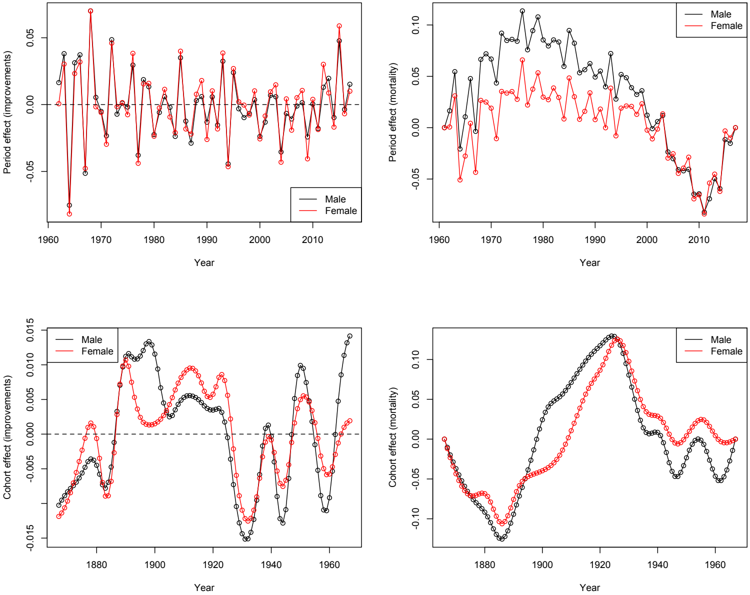
$$\log \frac{m_{xt}}{\beta - m_{xt}} = \mu_0 + \mu_1 x + (\alpha_0 + \alpha_1 x)t + \kappa_t + s_\gamma(t-x) \qquad x > x_0$$

where κ_t, $s_\gamma(t-x)$ are estimates obtained from fitting the APC GAM to the main body of the data $(0 < x \le x_0)$.

Mortality improvement estimates



Period and Cohort estimates



The forecast

$$\log m_{xt} = \hat{s}_\mu(x) + \hat{s}_\alpha(x)t + \hat{\kappa}_t + \hat{s}_\gamma(t-x) \qquad \text{for } t = T+1, \ldots$$

requires us to forecast

- $\{\kappa_t, t = T+1, \ldots\}$
- $s_\gamma(t-x)$ for $t - x > T - x_{\text{min}}$

In practice this is done by

- κ_t: random walk dynamics
- $s_\gamma(t-x)$: extrapolation of GAM smooth

Potential for forecasts over long horizons to be expert-moderated, e.g.

$$\hat{s}_\alpha(x) \to \alpha_x^{\text{exp}} \qquad \hat{s}_\gamma(t-x) \to 0$$

over some intermediate time horizon

Issues

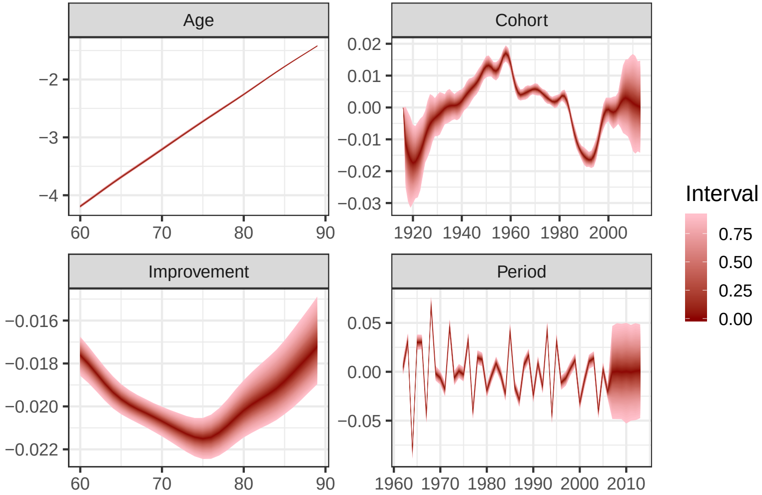
- Modelling assumptions
- Uncertainty
- Prior/expert opinion and Bayesian methods
- Series length and moderated forecasts
- Cohorts
- Joint modelling and borrowing strength
- Recent experience and random walks

Modelling assumptions

- Smoothness
- 'Error' distribution:
  - Poisson
  - negative binomial
  - quasi-Poisson
  - lognormal
- Cohort assumptions
- Sparse regions
- Joint v. hierarchical fitting

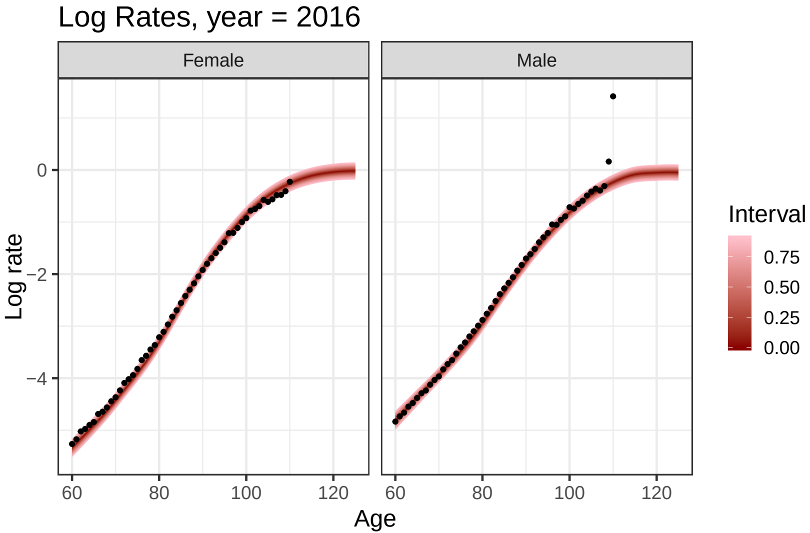
Uncertainty

A Bayesian approach allows coherent quantification of uncertainty encompassing all aspects (males 60+, data up to 2006)



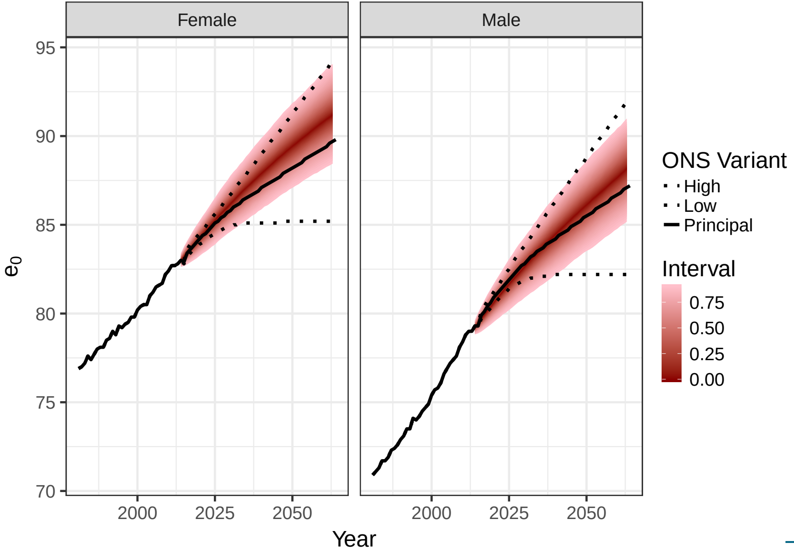
Forecast uncertainty

Fit on data up to 2006, 10 year projection.



Forecast uncertainty (life expectancy

Fit on data up to 2006, 10 year projection.



Prior/expert opinion and Bayesian methods

- Smoothness (regularisation)
- Parameter values (restrictions?)
- Parameter 'sharing' (male/female or other splits)
- Expert moderation of forecasts (to follow)
- Borrowing of strength (to follow)

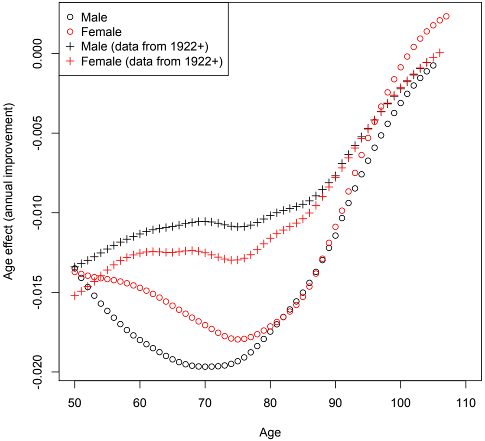
In all these cases Bayes methods combine uncertainties into a single posterior (predictive) distribution for inference.

Series length and moderated forecasts

UNIVERSITY OF Southampton

By necessity observed data series are always shorter than we would like.

For a given observed series length, what is a reasonable range for extrapolation?



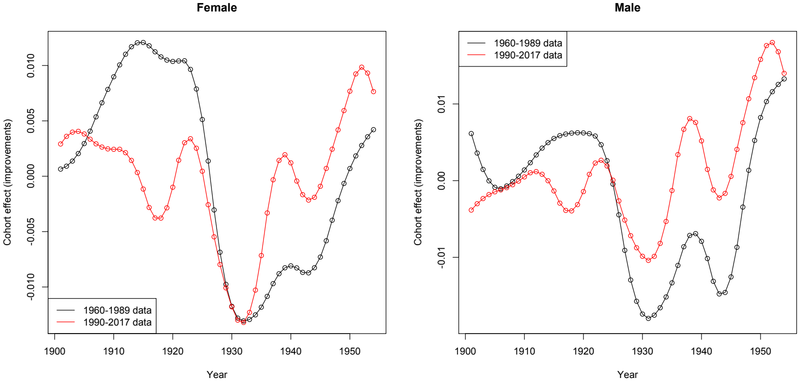
Potential for forecasts over long horizons to be expert-moderated.

Cohorts

Cohort-identifying assumption can be sensitive to range of data used for fitting.

How strong is cohort-effect persistence through the life-cycle?

Potential for borrowing of strength for forecasting unobserved cohorts?



A much smaller (but helpful) literature on joint modelling of two populations or modelling of a portfolio and its population, e.g.

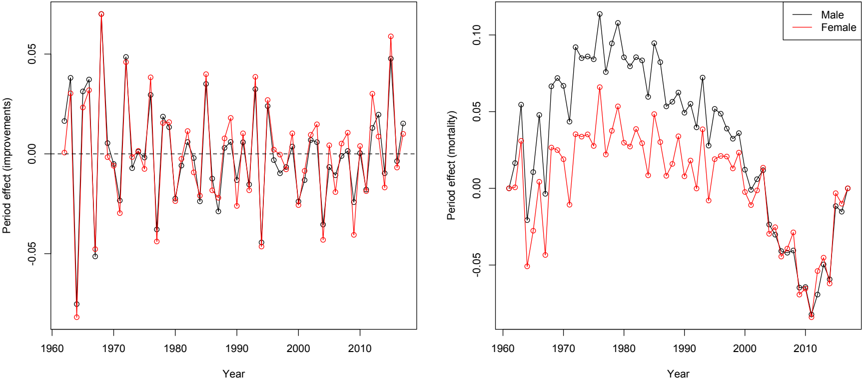
- Li (2012)
- Villegas and Haberman (2014)

and including Bayesian multi-population approaches, e.g.

- Cairns et al (2011)
- van Berkum et al (2017)

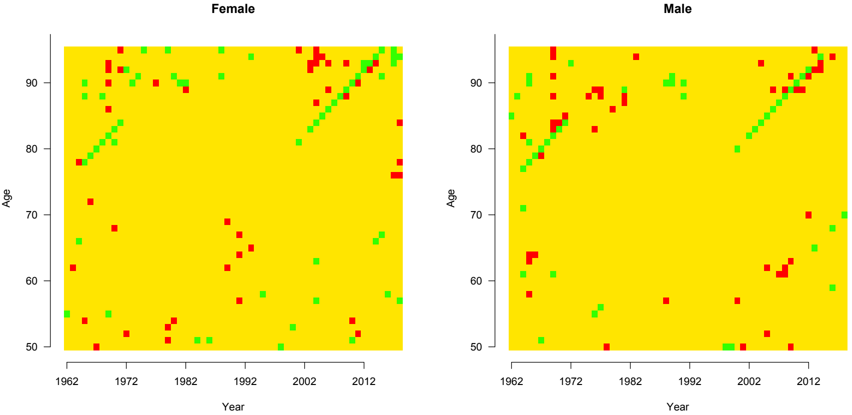
Joint modelling and expert opinion?

Recent experience

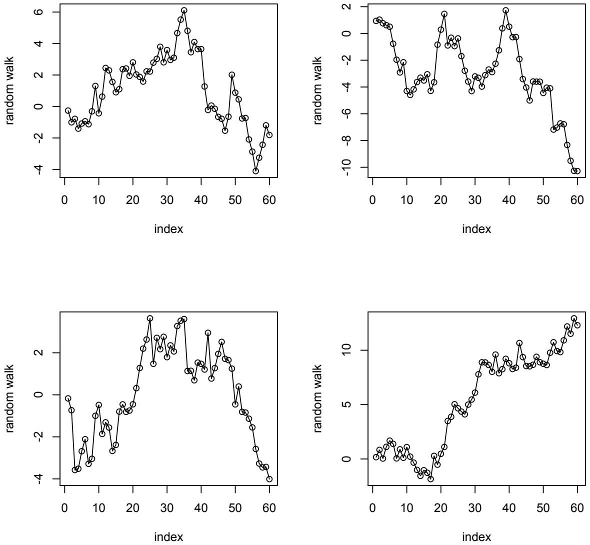


2015 experience stands out, but is not an outlier

Residuals


Female
Male
Age
Year
1962
1972
1982
1992
2002
2012
50
60
70
80
90

Random walks (simulated)



Discussion

- Modelling assumptions
- Uncertainty
- Prior/expert opinion and Bayesian methods
- Series length and moderated forecasts
- Cohorts
- Joint modelling and borrowing strength
- Recent experience and random walks