Stochastic Mortality Forecasting with Smoothing and Overdispersion

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

International Mortality and Longevity Symposium, 7–9 September 2016

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

$\{y_x\}$ – 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

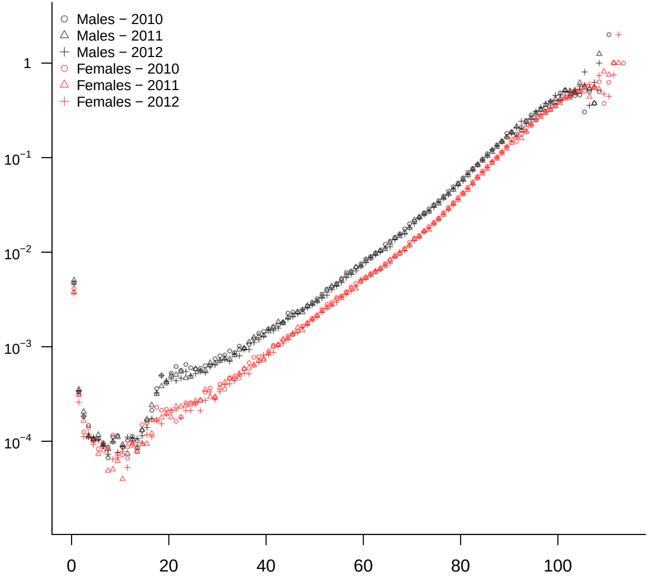
$$\tilde{m}_x = \frac{y_x}{E_x^C}.$$

This is an estimator of the underlying central mortality rate

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

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

Crude mortality rates 2010-2012



A basic smoothing model

For a large inhomogenous population, such as England and Wales, we prefer 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 (smoothing spline) model

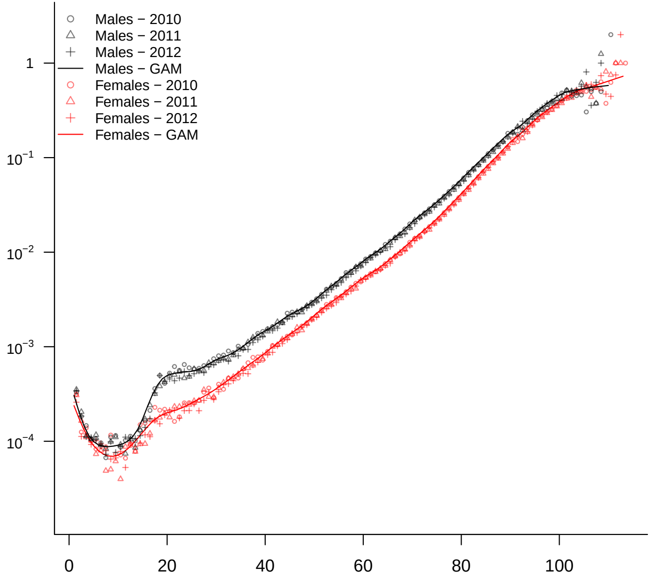
$$\log m_x = s(x;\beta)$$

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



To obtain a more robust fit at older ages, and to extrapolate the mortality function m_x beyond the range of the observed data, we 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

Therefore our proposed model across the entire range of x is

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

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).

Our proposed 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.

Bayes model averaging

Let $k = 1, \ldots, K$ index possible models $p_k(y|\theta_k)$ for observed data y

Then, a Bayesian approach updates a prior probability distribution $p(k)$ over models to a posterior distribution

$$p(k|y) \propto p(y|k)p(k)$$

where $p(y|k)$ is the *marginal likelihood*

$$p(y|k) = \int p_k(y|\theta_k)p_k(\theta_k)\mathrm{d}\theta_k.$$

Graduated estimates of m_x are then obtained as

$$\hat{m}_x = E[m_x|y] = \sum_k p(k|y)E_k(m_x|y)$$

a weighted average of the estimates under the various models

'Partial-Bayes' model-averaging

UNIVERSITY OF Southampton

A computationally efficient approach with minimal requirement for prior specification.

1. Replace $\hat{m}_x = \sum_k p(k|y) E_k(m_x|y)$ with $\hat{m}_x = \sum_k p(k|y) \hat{m}_x^{(k)}$
2. Split the data y into y_t (training) and y_v (validation), and replace marginal likelihood $p(y|k)$ with the partial marginal likelihood

$$p(y_v|k, y_t) = \int p_k(y_v|\theta_k) p_k(\theta_k|y_t) \mathrm{d}\theta_k.$$

3. Replace $p_k(\theta_k|y_t)$ above by a point mass at $\hat{\theta}'_k$, the (penalised) maximum likelihood estimate based on y_t only. Then

$$p(y_v|k, y_t) = p_k(y_v|\hat{\theta}'_k)$$

These (2 and 3) lead to partial-Bayes model probabilities

$$p(k|y) \propto p_k(y_v|\hat{\theta}'_k) p(k)$$

Practical partial-Bayes graduation

The partial-Bayes posterior model probabilities lead to partial-Bayes graduations (under a uniform prior distribution over models)

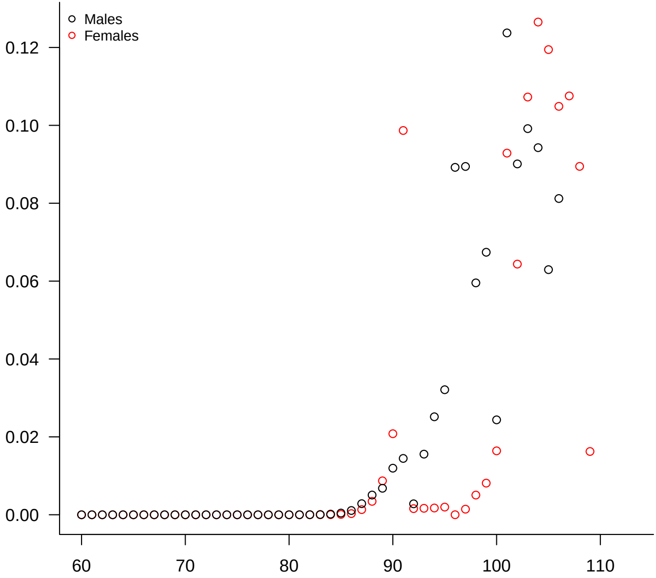
$$\hat{m}_x = \frac{\sum_k p_k(y_v|\hat{\theta}'_k)\hat{m}_x^{(k)}}{\sum_k p_k(y_v|\hat{\theta}'_k)}$$

where models are weighted on the basis of how well they predict the validation data, based on parameters estimated using the training data.

Here model index k controls log-linear/logistic extrapolation *and* threshold x_0.

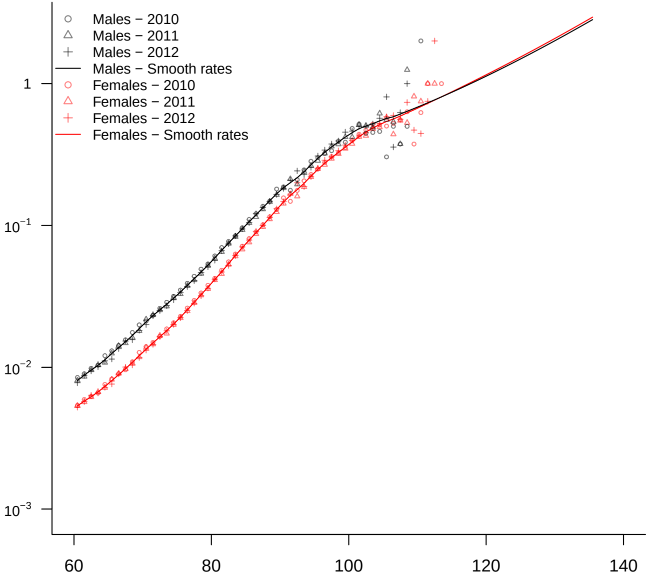
Years $\{2010, 2012\}$ are used for training and $\{2011\}$ for validation.

Posterior threshold probabilities (log-linear)

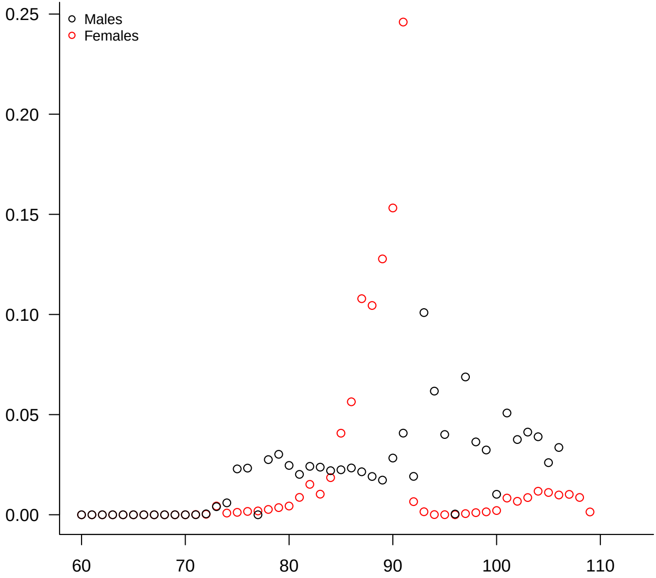


Graduation (log-linear)

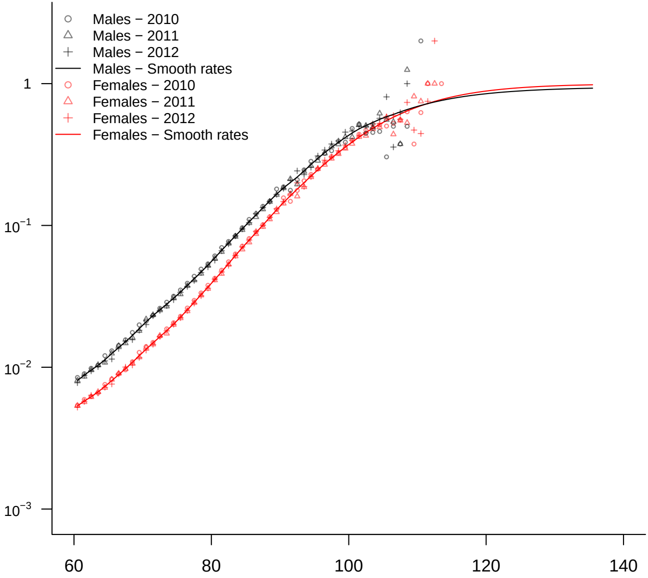
UNIVERSITY OF Southampton



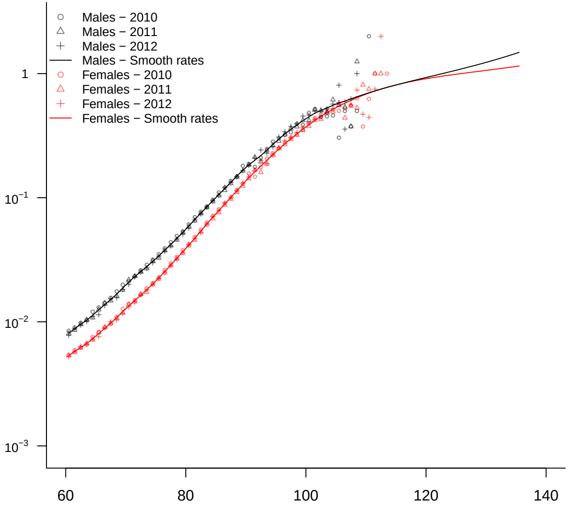
Posterior threshold probabilities (logistic)



Graduation (logistic)

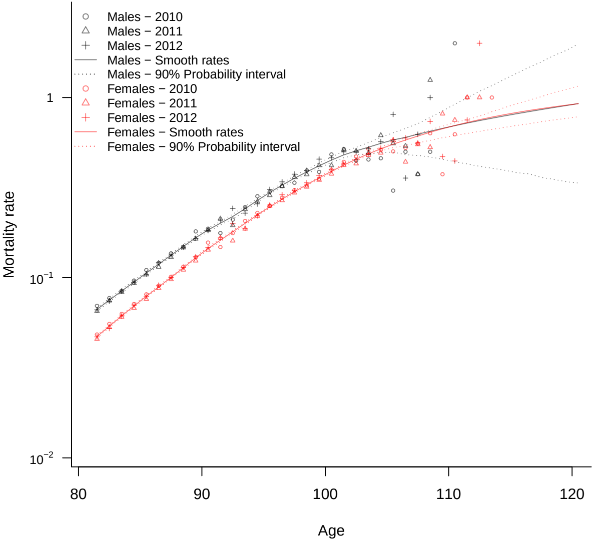


ELT17 model-averaged graduation



$P(\text{log-linear}) = 0.087$ (female) $P(\text{log-linear}) = 0.292$ (male)

Uncertainty



Benefits of this approach:

- Takes advantage of the ease with which a wide range of smooth and parametric models can routinely be fitted
- Acknowledges that in regions of sparse data there remains considerable uncertainty about the model which should be used for estimation and extrapolation.
- Computationally straightforward, but scientifically coherent approach for incorporating model uncertainty into graduation

Can be incorporated into a forecasting framework ...

Mortality projection models

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

- Generalised bilinear (e.g. Lee Carter with cohort)

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

- Generalised linear (e.g. APC with age-period interaction)

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

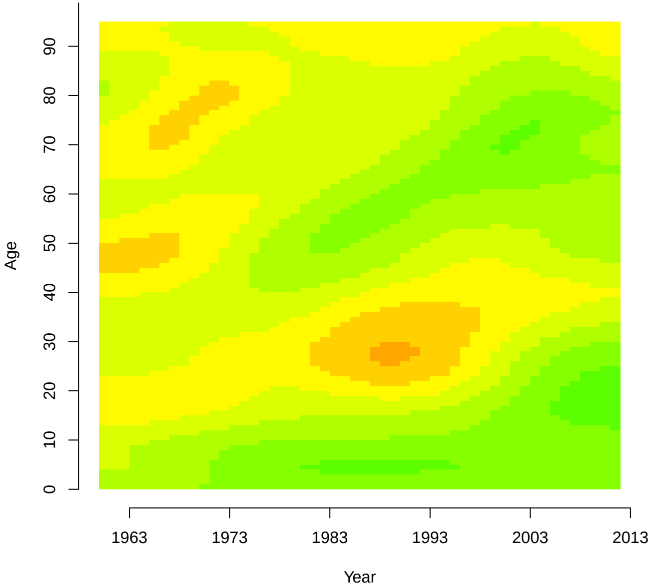
- semi-parametric

$$\log m_{xt} = s(x, t)$$

- generalised additive (GAM)

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

Smoothed UK mortality improvements



The initial proposed model specification

UNIVERSITY OF Southampton

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

$$\log \frac{m_{xt}}{m_{x\,t-1}} = \alpha_x + \kappa_t + \gamma_{t-x} \quad (1)$$

or equivalently APC model for mortality rates, with age-period interaction

$$\log m_{xt} = m_{x0} + \alpha_x t + \kappa_t + \gamma_{t-x} \quad (2)$$

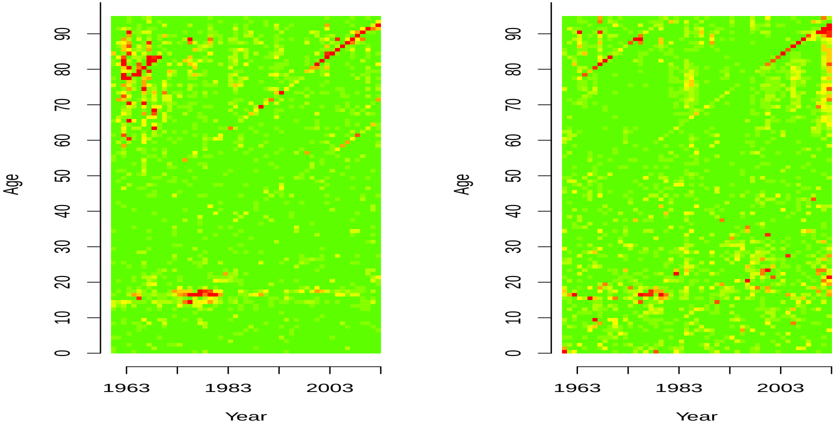
To obtain smoother estimates modify (2) to a generalised additive model (GAM):

$$\log m_{xt} = s_\mu(x) + s_\alpha(x)t + \kappa_t + s_\gamma(t-x). \quad (3)$$

where s_μ, s_α and s_γ are arbitrary smooth functions.

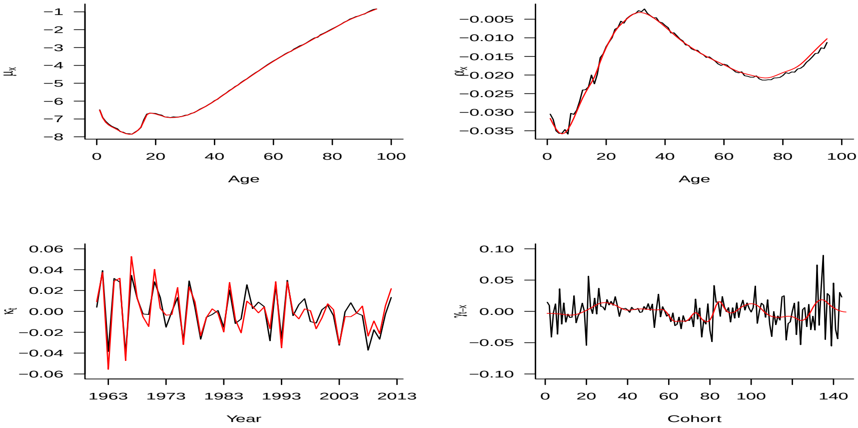
Comparison of residuals

UNIVERSITY OF Southampton



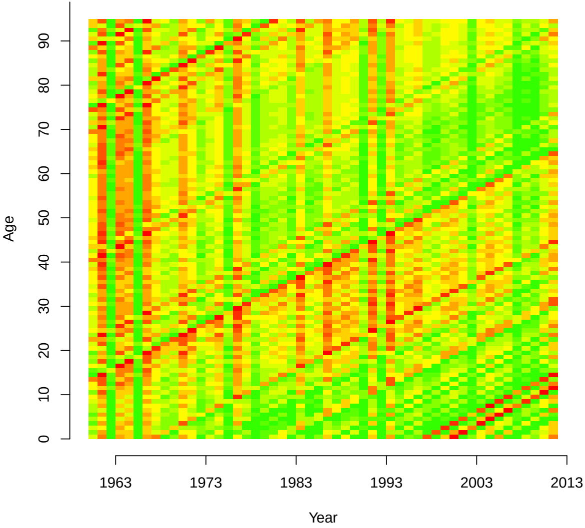
The P-spline approach allowing for overdispersion (left panel) and model (2) under the negative binomial distribution (right panel).

Parameter estimates

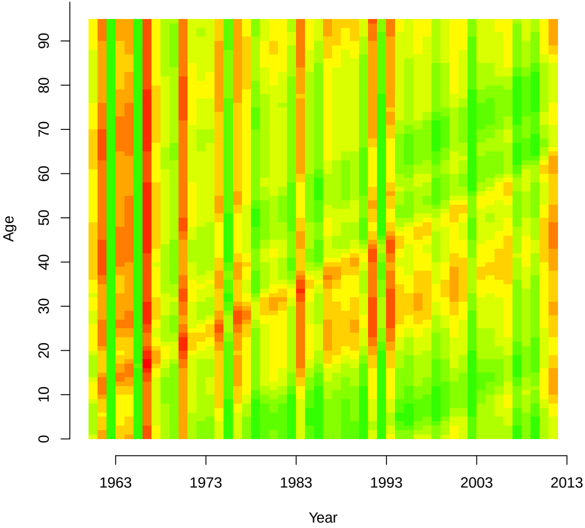


Estimates of the parameters of model (3), data for males 1961-2013, (red lines) superimposed over the corresponding estimates for model (2) (black lines).

Fitted mortality improvements for model (1)



Fitted mortality improvements for model (3)



Model specification for old ages

For the highest ages x, use parametric models:

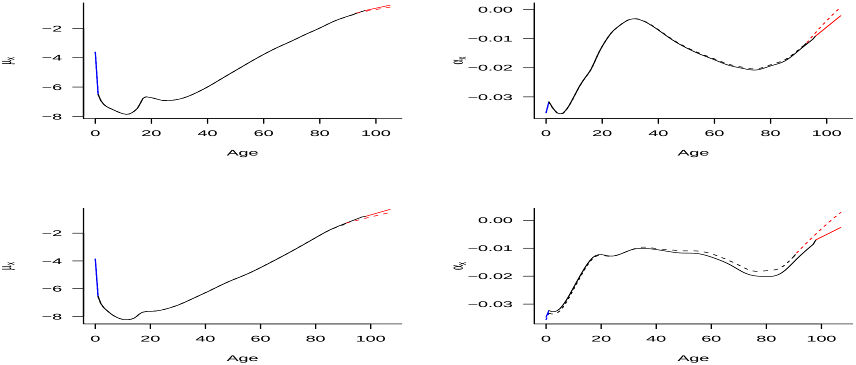
$$\log m_{xt} = \mu_0 + \mu_1 x + (\alpha_0 + \alpha_1 x)t + \kappa_t + s_\gamma(t - x) \qquad x > x_0 \qquad (4)$$

or

$$\log \frac{m_{xt}}{\beta - m_{xt}} = \mu_0 + \mu_1 x + (\alpha_0 + \alpha_1 x)t + \kappa_t + s_\gamma(t - x) \qquad x > x_0 \quad (5)$$

where κ_t, $s_\gamma(t - x)$ are estimates obtained from fitting (3) to the main body of the data $(0 < x \leq x_0)$.

Parameter estimates



Estimates of the parameters of models (3), (4) and (5), 1961-2013, for males (upper panels; x_0=95 for log-linear model (solid line) and x_0=92 for logistic model (dashed line)) and females (lower panels; x_0=97 for log-linear model (solid line) and x_0=90 for logistic model (dashed line)).

Model specification for infants

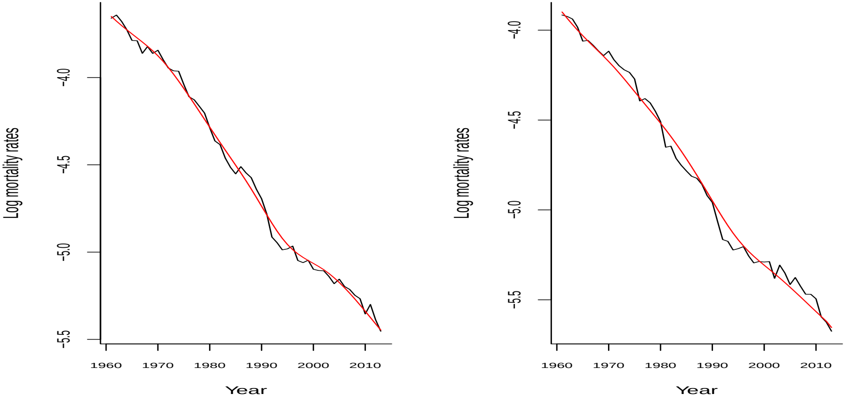
UNIVERSITY OF Southampton

For infants (age 0) we use:

$$\log \mu_{0t} = \mu_0 + \alpha_0 t + s_\gamma(t - x) \qquad (6)$$

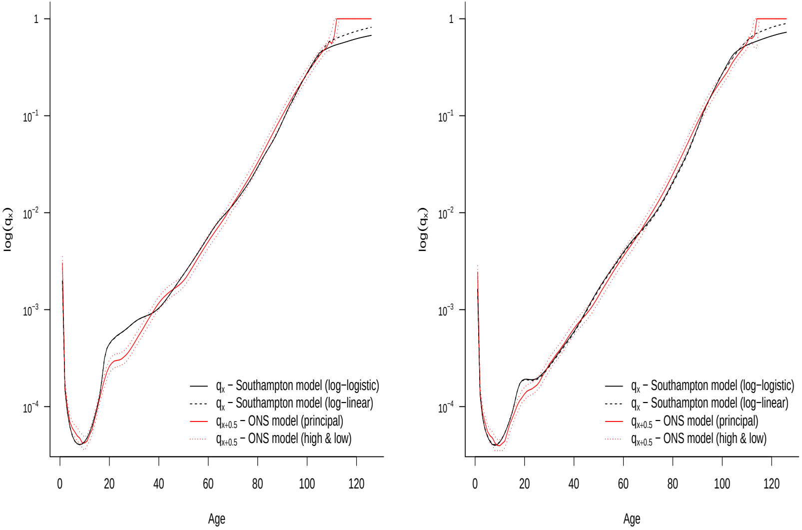
where $s_\gamma(t - x)$ are estimates obtained from fitting model (3) to the main body of the data $(0 < x \leq x_0)$.

Estimates of infant mortality rates



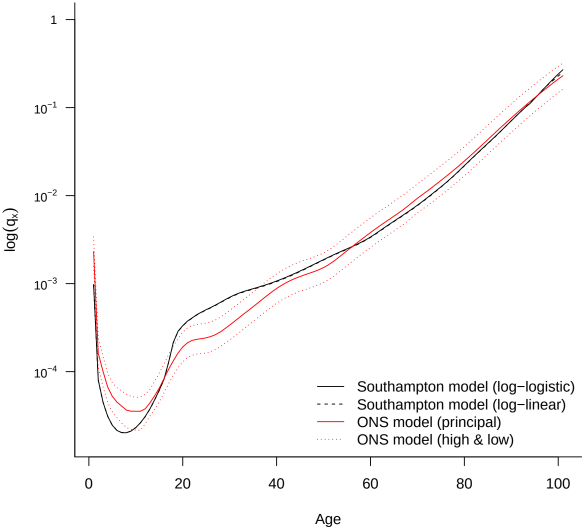
Estimates of infant mortality rates, 1961-2013, for males (left panel) and females (right panel) using model (6; red lines), compared with observed rates (black lines).

2035 Projections Males and Females



(a) Males (b) Females

2055 Projections (Males)

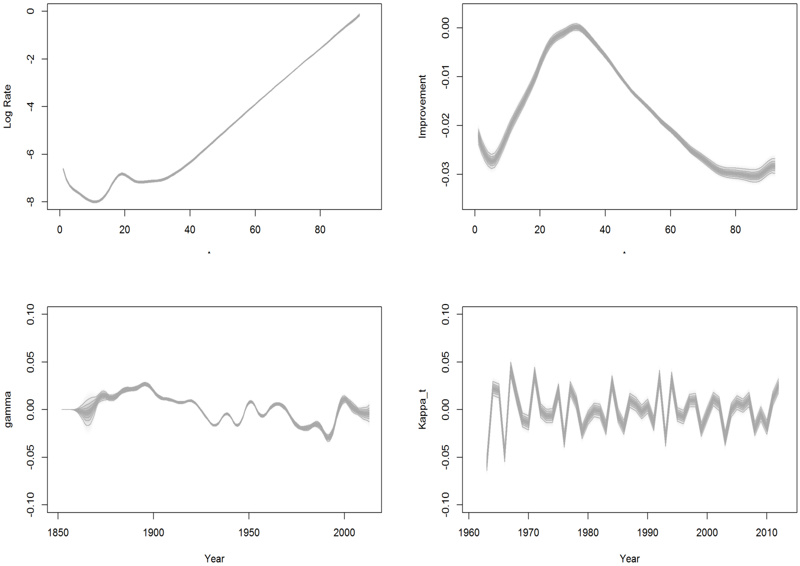


Further developments

UNIVERSITY OF Southampton

- Completely integrated estimation
- Bayesian approach with expert opinion and full uncertainty quantification
- Time-varying old-age threshold and/or mortality asymptote
- Models by causes of death

Posterior distributions



Posterior distribution (infant rates)

UNIVERSITY OF Southampton