A flexible framework for selecting mortality models

Aniketh Pittea (Grant Thornton UK LLP)

Alex Diana (University of Essex)

Jackie Wong (University of Essex)

Grant Thornton

Agenda

- Background and motivation
- Bayesian model selection
- Application on DB pension scheme

Part 1: Background, motivation and a bit of history!

Background

- Our initial idea was to build a longevity model which would model longevity improvements tailored for specific products.

Background

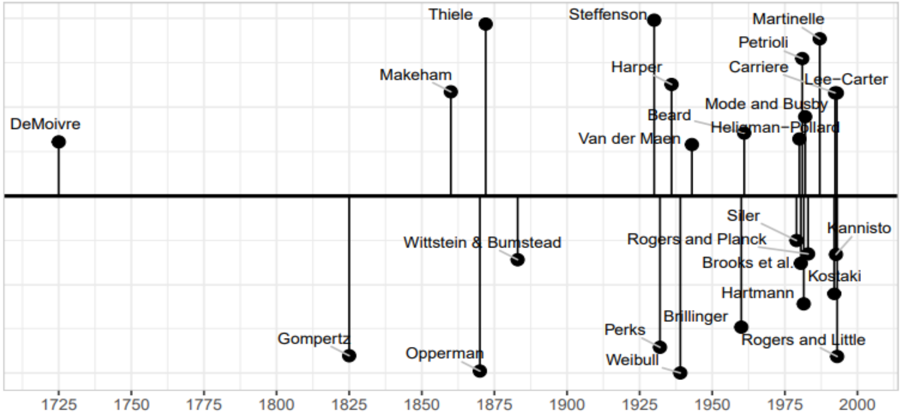
- Our initial idea was to build a longevity model which would model longevity improvements tailored for specific products.

- It quickly became difficult for us to choose a model form given the long history of longevity models and the numerous models researchers have produced.

Background

- Our initial idea was to build a longevity model which would model longevity improvements tailored for specific products.

- It quickly became difficult for us to choose a model form given the long history of longevity models and the numerous models researchers have produced.



Source: Modelling and forecasting mortality modelling, Ph.D Thesis, Marius D. Pascarui (2018).

History – the early models

- 1662 - Researcher Graunt showed that while death was uncertain, there were predictable patterns of mortality in groups and causes of death.

History – the early models

- 1662 - Researcher Graunt showed that while death was uncertain, there were predictable patterns of mortality in groups and causes of death.
- 1725 – Theoretical modelling began in DeMoivre who proposed a uniform distribution of the death model.

History – the early models

- 1662 - Researcher Graunt showed that while death was uncertain, there were predictable patterns of mortality in groups and causes of death.

- 1725 – Theoretical modelling began in DeMoivre who proposed a uniform distribution of the death model.

$$S(x) = 1 - \frac{x}{\omega}, \qquad 0 \leq x < \omega.$$

Grant Thornton

History – the early models

- 1662 - Researcher Graunt showed that while death was uncertain, there were predictable patterns of mortality in groups and causes of death.

- 1725 – Theoretical modelling began in DeMoivre who proposed a uniform distribution of the death model.

$$S(x) = 1 - \frac{x}{\omega}, \qquad 0 \leq x < \omega.$$

- 1825 – Gompertz proposed that deaths in adulthood should take an exponential form.

History – the early models

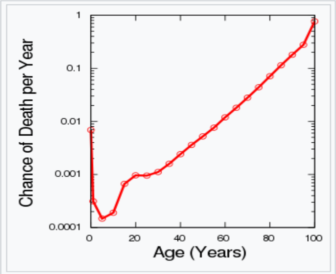
- 1662 - Researcher Graunt showed that while death was uncertain, there were predictable patterns of mortality in groups and causes of death.

- 1725 – Theoretical modelling began in DeMoivre who proposed a uniform distribution of the death model.

$$S(x) = 1 - \frac{x}{\omega}, \qquad 0 \leq x < \omega.$$

- 1825 – Gompertz proposed that deaths in adulthood should take an exponential form.

$$\left(\alpha e^{\beta x} + \lambda\right) \cdot \exp\left[-\lambda x - \frac{\alpha}{\beta}\left(e^{\beta x} - 1\right)\right]$$



GrantThornton

History – the Lee-Carter

- 1992 – the Lee-Carter proposes a breakthrough model where:
  - Mortality rates are modelled as a log bi-linear model incorporating both age and period effect;
  - The model produces stochastic forecasts.

Grant Thornton

History – the Lee-Carter

- 1992 – the Lee-Carter proposes a breakthrough model where:
  - Mortality rates are modelled as a log bi-linear model incorporating both age and period effect;
  - The model produces stochastic forecasts.

$$\ln m(x,t) = a(x) + b(x)k(t) + \varepsilon(x,t)$$

Grant Thornton

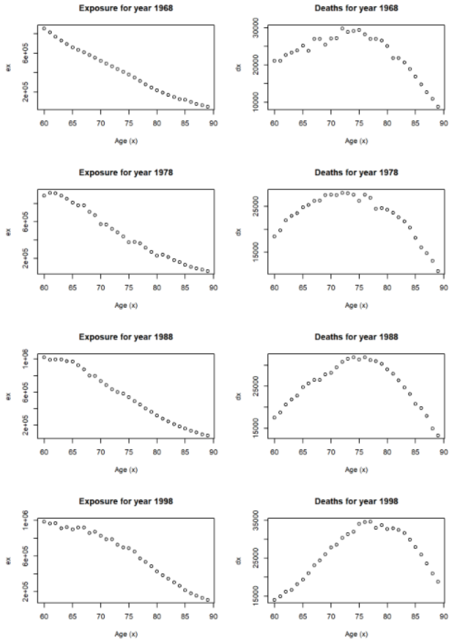
History – the Lee-Carter

- The fitting and simulation for the Lee-Carter works as follows:

History – the Lee-Carter

- The fitting and simulation for the Lee-Carter works as follows:

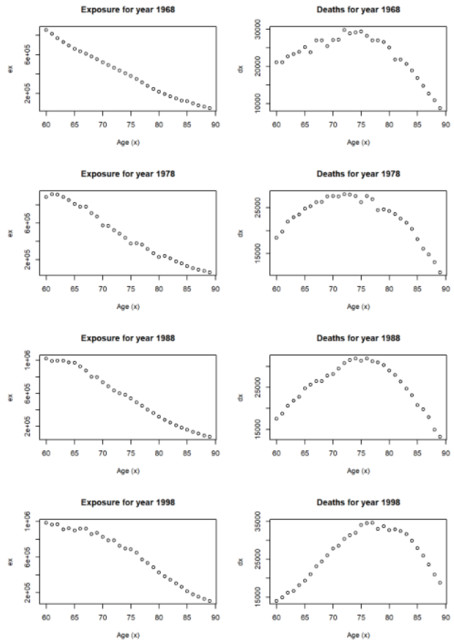
1. Collect data on deaths and exposure.



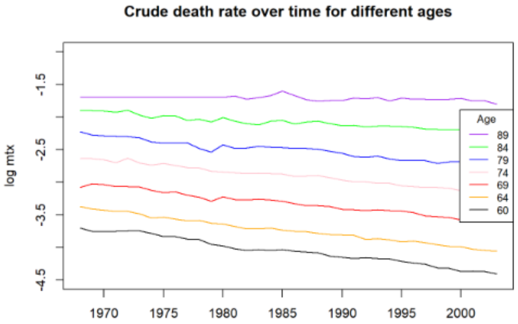
History – the Lee-Carter

- The fitting and simulation for the Lee-Carter works as follows:

1. Collect data on deaths and exposure.



2. Calculate crude death rates

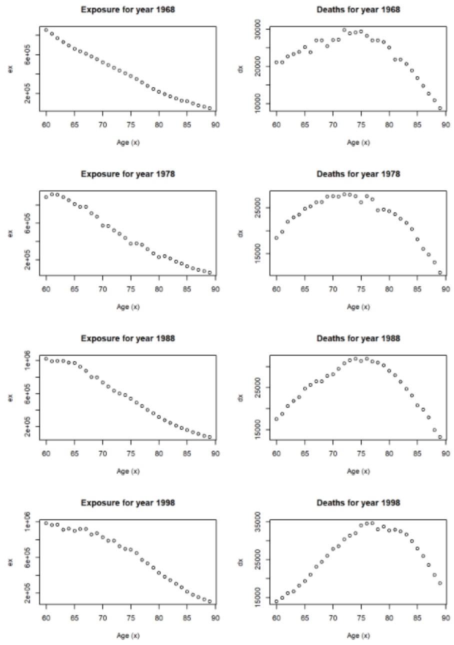


GrantThornton

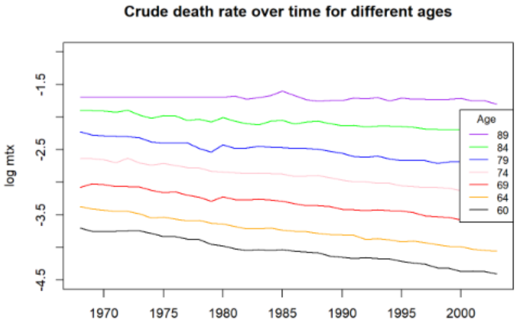
History – the Lee-Carter

- The fitting and simulation for the Lee-Carter works as follows:

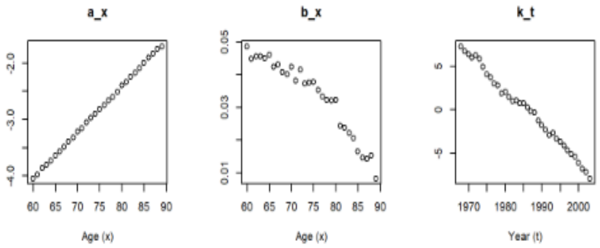
1. Collect data on deaths and exposure.



2. Calculate crude death rates



3. Estimate the parameters using Maximum Likelihood

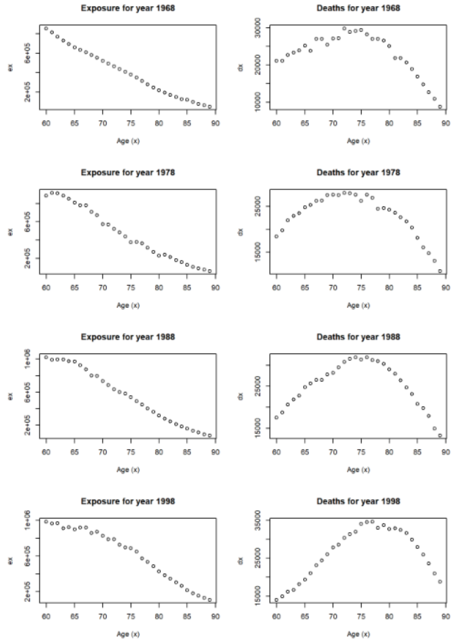


Grant Thornton

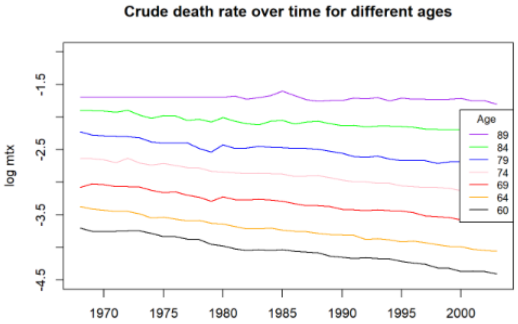
History – the Lee-Carter

- The fitting and simulation for the Lee-Carter works as follows:

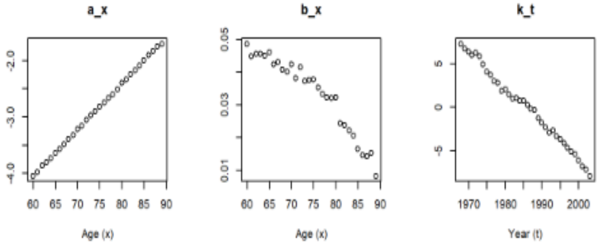
1. Collect data on deaths and exposure.



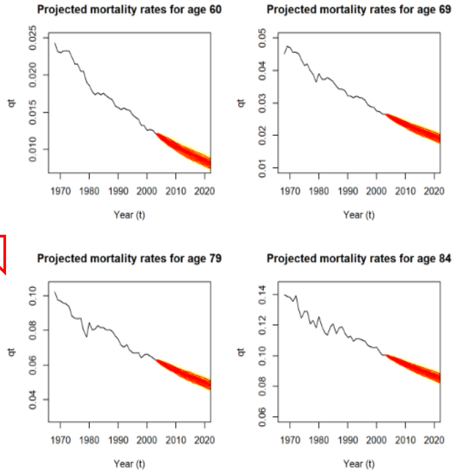
2. Calculate crude death rates



3. Estimate the parameters using Maximum Likelihood



4. Project the k_t parameter and recover projected mortality rates



History – Beyond the Lee-Carter

- Many models since then have tried to build on top of the Lee-Carter.

History – Beyond the Lee-Carter

- Many models since then have tried to build on top of the Lee-Carter.
- Cairns, Blake and Dowd (CBD) (2006) started with a similar form of the Lee-Carter but assumed age can be modelled linearly thereby reducing the number of parameters.

History – Beyond the Lee-Carter

- Many models since then have tried to build on top of the Lee-Carter.

- Cairns, Blake and Dowd (CBD) (2006) started with a similar form of the Lee-Carter but assumed age can be modelled linearly thereby reducing the number of parameters.

$$\text{logit } q(t,x) = \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}.$$

For this model simple parametric forms were assumed for $\beta_x^{(1)}$ and $\beta_x^{(2)}$:

$$\begin{aligned} \beta_x^{(1)} &= 1, \\ \text{and } \beta_x^{(2)} &= (x - \bar{x}) \end{aligned}$$

$$\text{logit } q(t,x) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}).$$

Grant Thornton

History – Beyond the Lee-Carter

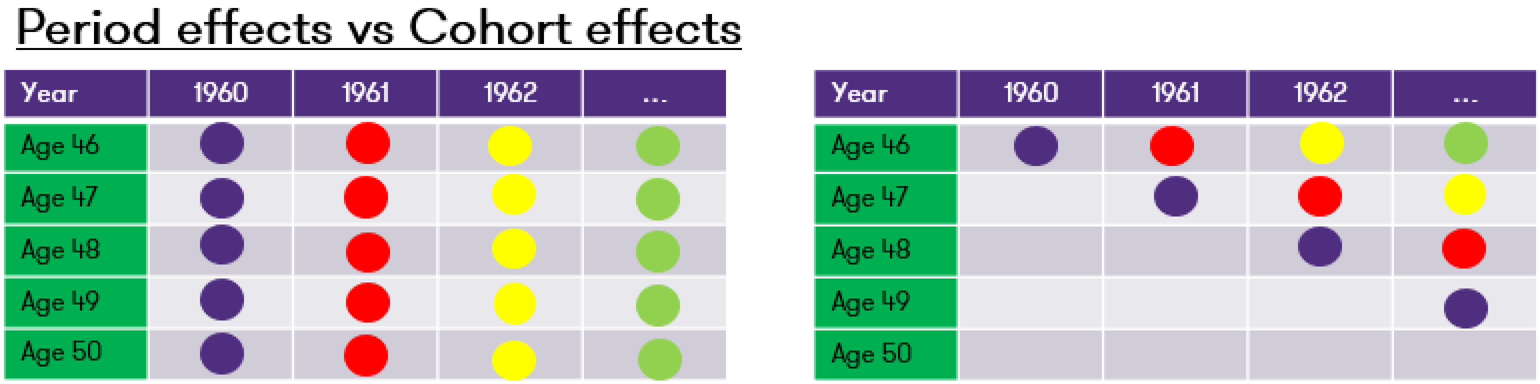
- Renshaw and Haberman (2006) added a cohort effect to the Lee-Carter to introduce an Age-Period-Cohort (APC) model.

$$\log m(t,x) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_{t-x}^{(3)}.$$

History – Beyond the Lee-Carter

- Renshaw and Haberman (2006) added a cohort effect to the Lee-Carter to introduce an Age-Period-Cohort (APC) model.

$$\log m(t,x) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_{t-x}^{(3)}.$$

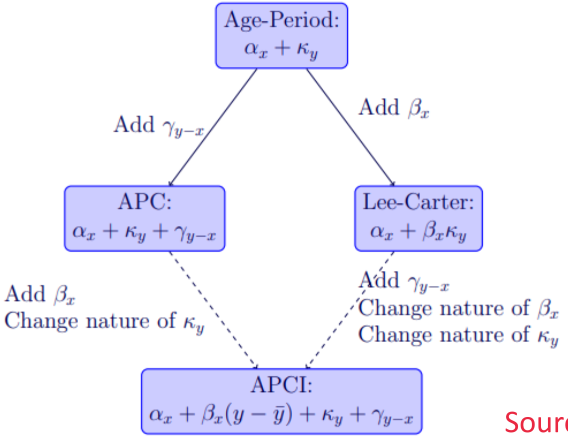


History – Beyond the Lee-Carter

- CMI (2017) introduced an APCI model – the model is a combination of the APC, CBD and Lee-Carter model.

History – Beyond the Lee-Carter

- CMI (2017) introduced an APCI model – the model is a combination of the APC, CBD and Lee-Carter model.



Source: ICC Birmingham, The APCI model, a stochastic implementation, Stephen Richards (2017)

Motivation

- Build a framework which would make mortality model selection more transparent and more automated

Motivation

- Build a framework which would make mortality model selection more transparent and more automated
- Aim is to equip practitioners with a tool which would:
  - help better understand the model and parameter risks
  - make selection between models easier
  - make extending existing models easier

Grant Thornton

Part 2: Bayesian model selection

What errors arise when doing mortality modelling?

- Stochastic error: the “noise” of the process under study.

- Parameter error: error in estimating the parameters.

- Model error: error in selecting the right model.

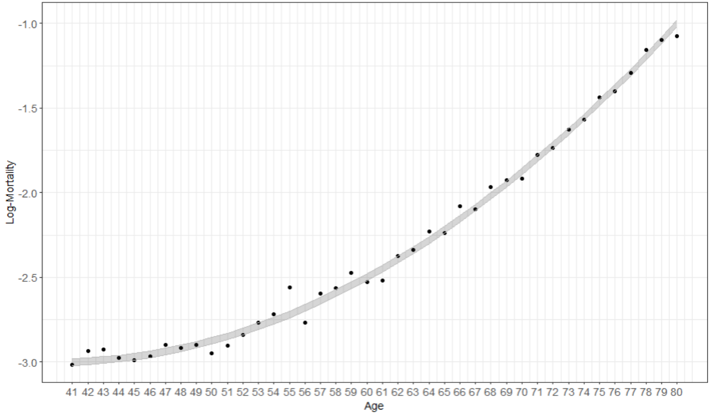
Stochastic error

This error arises even if we can perfectly estimate the parameter and the model.

For example, even if we could perfectly estimate the parameter of the model, we would not expect to perfectly predict the number of deaths for next year.

Consequences of stochastic error

What happens if we fail to estimate stochastic error correctly?

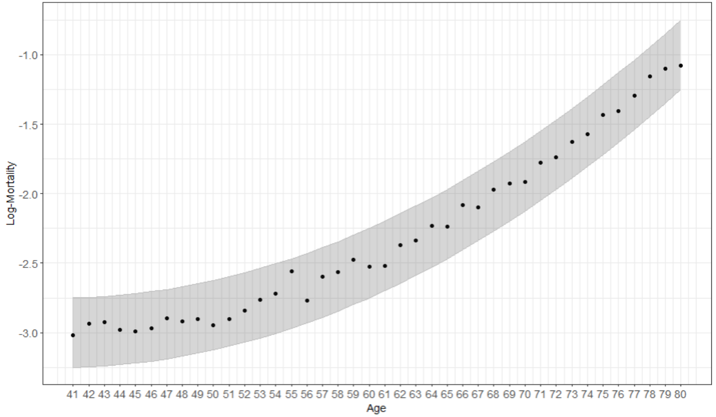


Underestimation:

- Confidence intervals for predicted deaths will be too narrow, hence we will be overconfident.
- Risk will be underestimated and we might have potential financial shortfalls.

Overestimation:

- We will put in too much capital than what is needed.



Parameter error

This error arises if we estimate the wrong set of parameters (even if we have the correct model).

Issues in failing to estimate parameter error are similar to stochastic error (but we might also get bias).

Ideally, if we had infinite amount of data, we could perfectly estimate the parameter with no uncertainty.

In practice, since we always have a limited amount of data, we account for parameter error by accurately estimating the uncertainty around our parameter estimates.

Parameter error

We can estimate uncertainty around model parameters with:

- Confidence intervals (for example using bootstrap)
- Bayesian Credible intervals.

and by propagating this error into our predictions (this is generally more straightforward in the Bayesian framework).

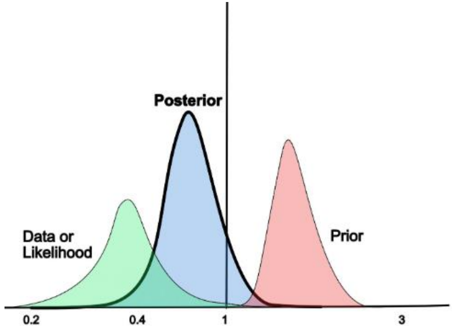
Frequentist vs Bayesian Statistics

- In the frequentist framework, there is a true parameter that is **fixed** and unknown.

- In the Bayesian framework, the parameter of interest is a random variable and therefore it has a distribution.

- The distribution we assumed on the parameter of interest **before** seeing the data is called the prior distribution

- The distribution **after** seeing the data is called the posterior distribution, which will be the distribution of interest.



- Inference in Bayesian statistics is performed via obtaining sampling from the posterior distribution.

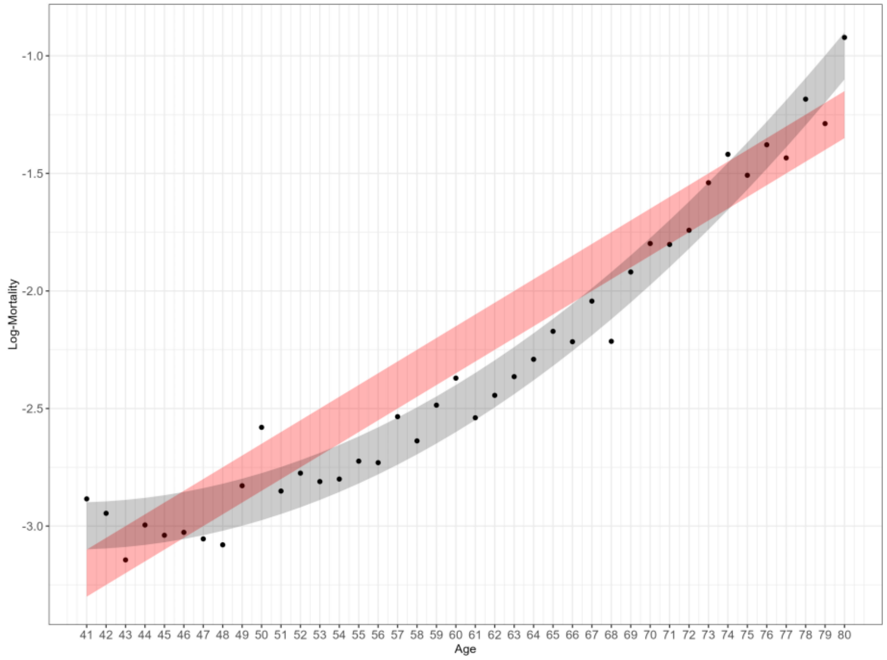
Model error

Since we select among many set of models, there is obviously the possibility that we might select the wrong model.

Like before, in practice we can only select the true model if we had infinite amount of data.

Moreover, one could argue that there is no “true” model.

“all models are wrong, but some are useful”



Grant Thornton

Approaches for model error

Traditionally, model error is mitigated using a model selection criterion such as AIC or BIC.

The idea is to fit many models and select the best one according to a metric, which usually has the form:

m = loglikelihood(goodness of fit) – penalty(number of parameters)

The downside of this approach is that we end up selecting only one model even though other models might also be informative.

Our approach for model error

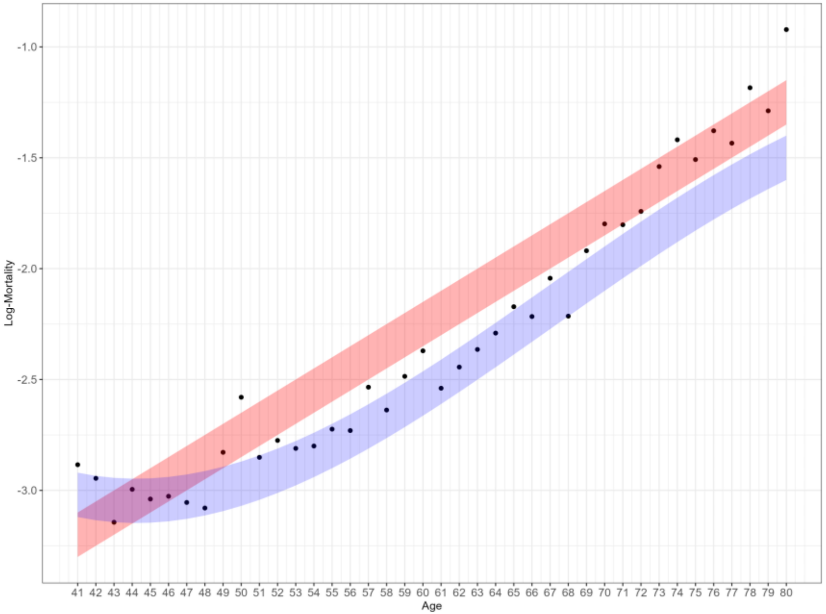
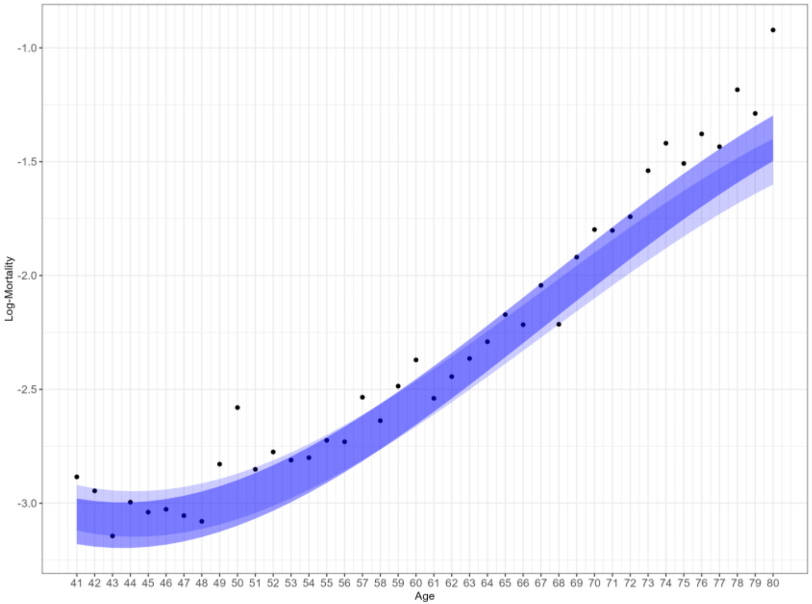
We build a unifying framework by making model selection part of the estimation procedure. In other words, we treat the choice of the model to be selected just like another parameter of the model.

To perform model selection, we will use a state-of-art inference technique in Bayesian inference, Reversible Jump Markov Chain Monte Carlo (RJMCMC), which is used to make model selection part of the inference.

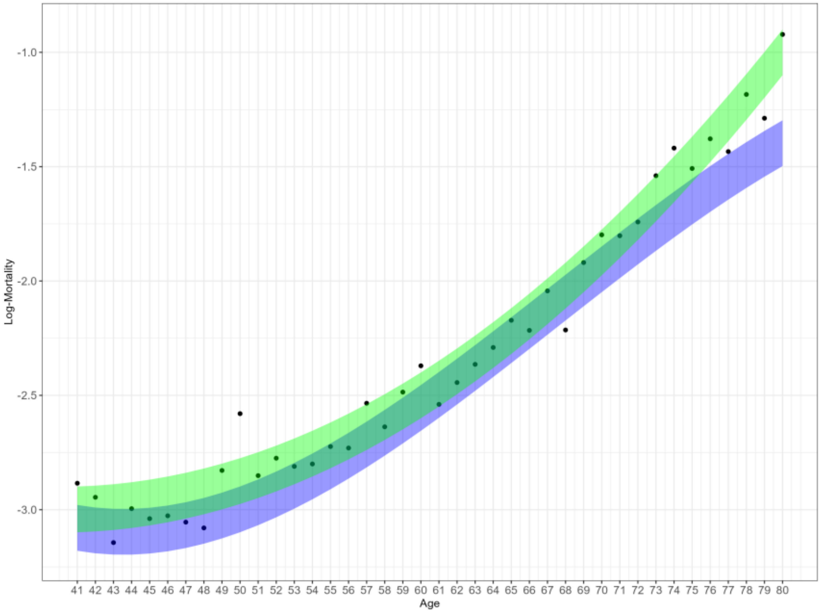
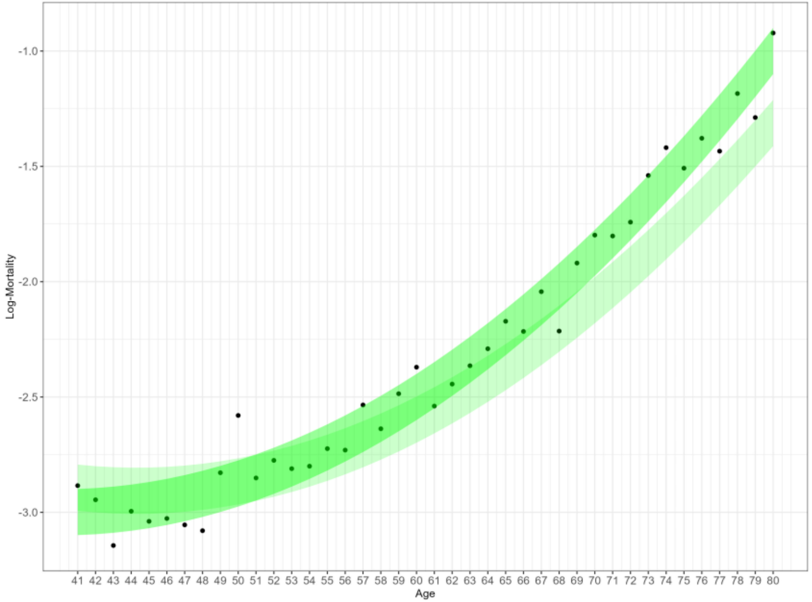
Our set of parameters is a pair (M_i, θ_i), where M_i is the current preferred model and θ_i is the current set of parameters (of model M_i), the RJMCMC algorithm is:

1) Propose a new model (M_i^*)
2) Re-estimate the parameters θ_i given the current model M_i.

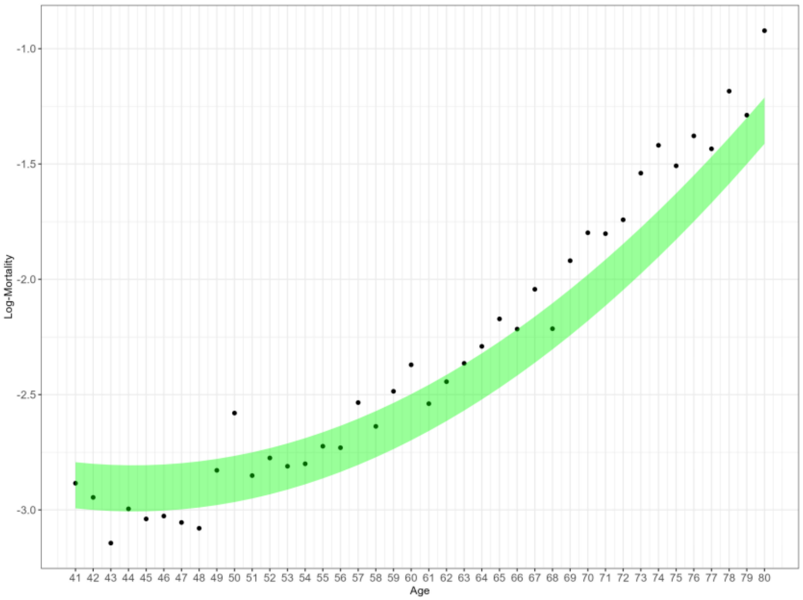
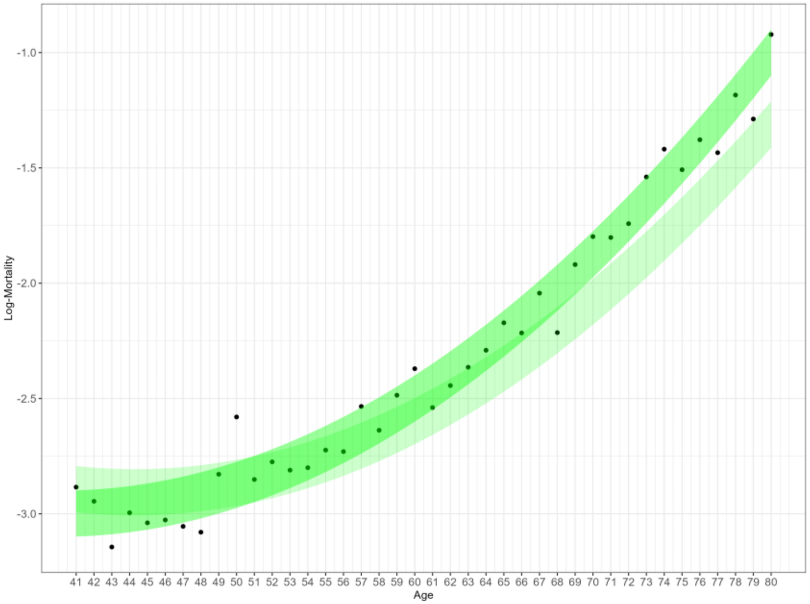
Reversible jump

	Update model	Update parameters
Iteration 1		

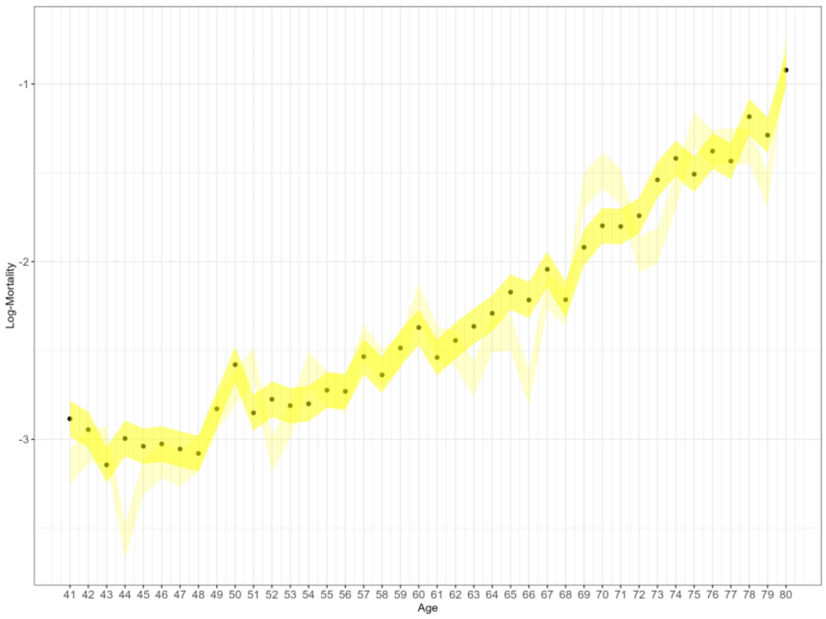
Reversible jump

	Update model	Update parameters
Iteration 2		

Reversible jump

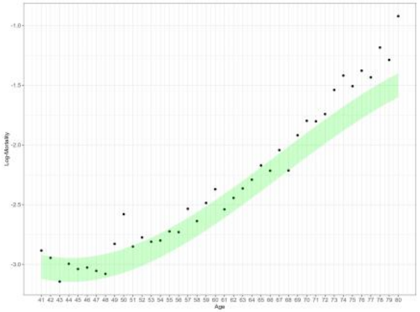
	Update model	Update parameters
Iteration 3		

Reversible jump

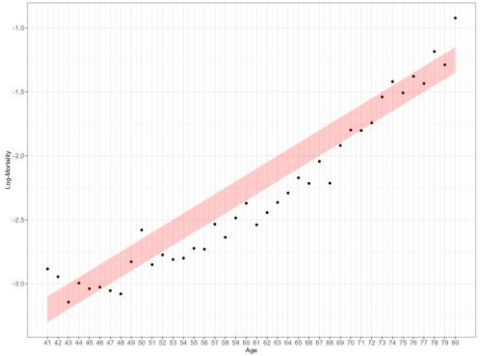
	Update model	Update parameters
Iteration 4		

Reversible jump

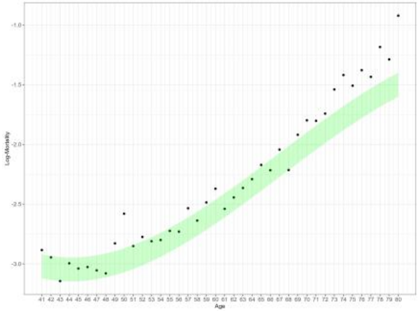
What happens in each iteration?



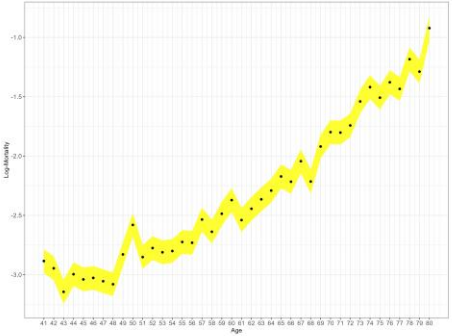
The framework can either revert the model to a simpler model



Or stay in the same model



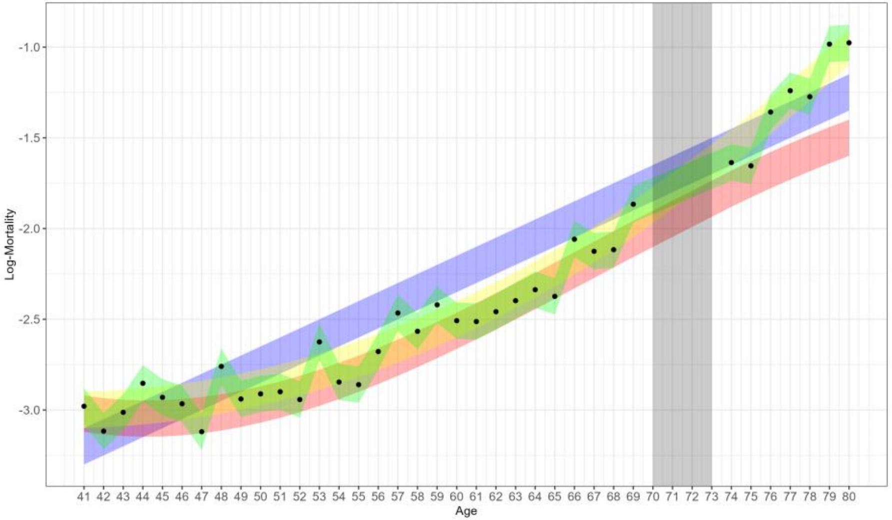
Or expand to a larger model



Grant Thornton

Reversible jump

What happens over many iterations?



- The framework will ‘favour’ certain models by spending more iterations on better models and fewer iterations on less desirable models

- This is different from AIC/BIC model selection which would select only one model

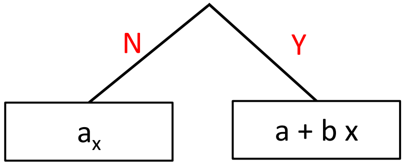
Model building framework

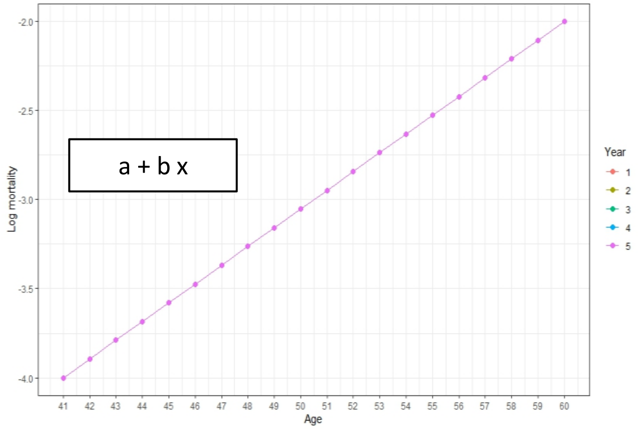
To use the RJMCMC procedure, we need to define a framework for building mortality models.

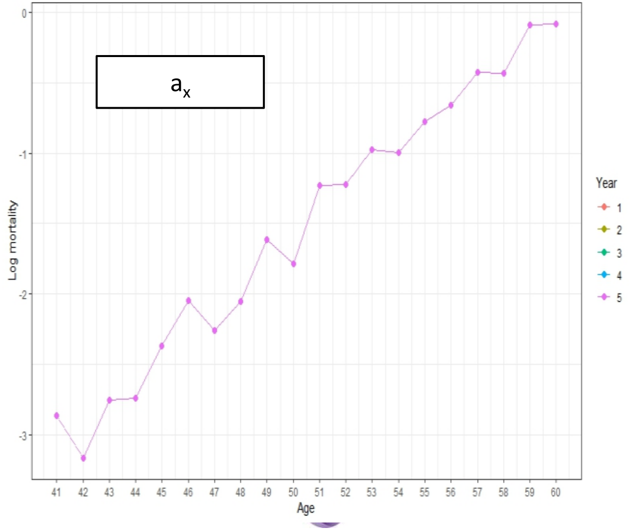
In particular, we need to associate each model with a choice of parameters (and next perform inference on those parameters).

Ideally, the framework would encompass as many models as possible, in order to model the data as close as possible.

Is the **age** effect linear?

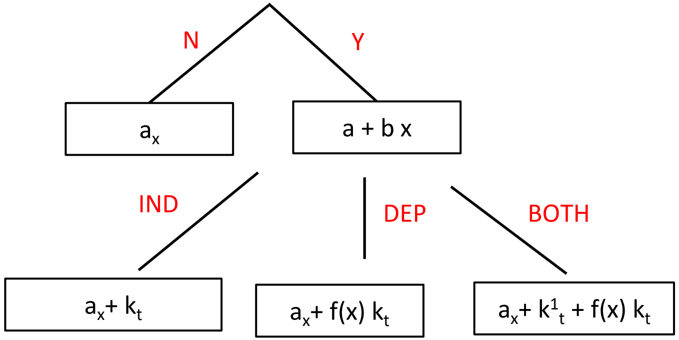


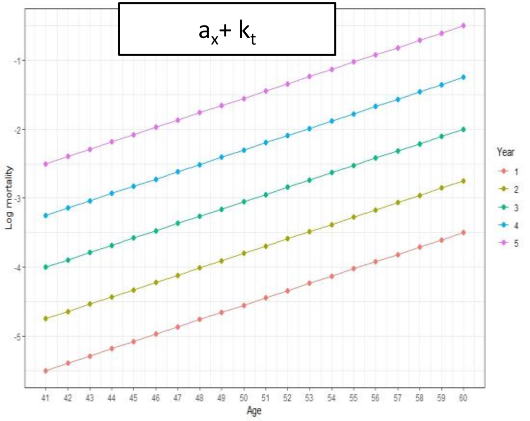


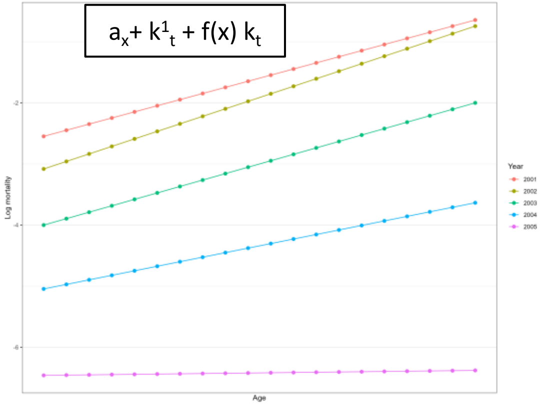


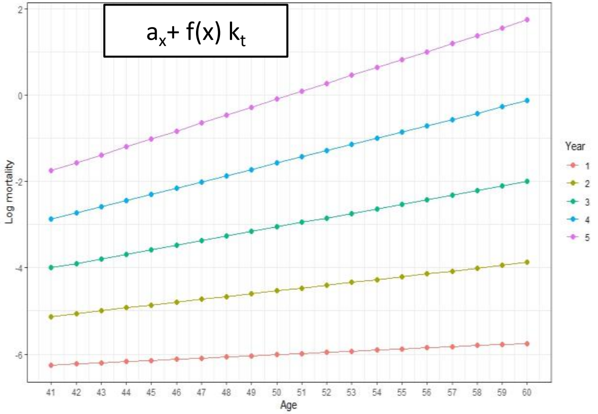
Is the **age** effect linear?

Is there a **period** effect independent and/or an effect interacting with age?





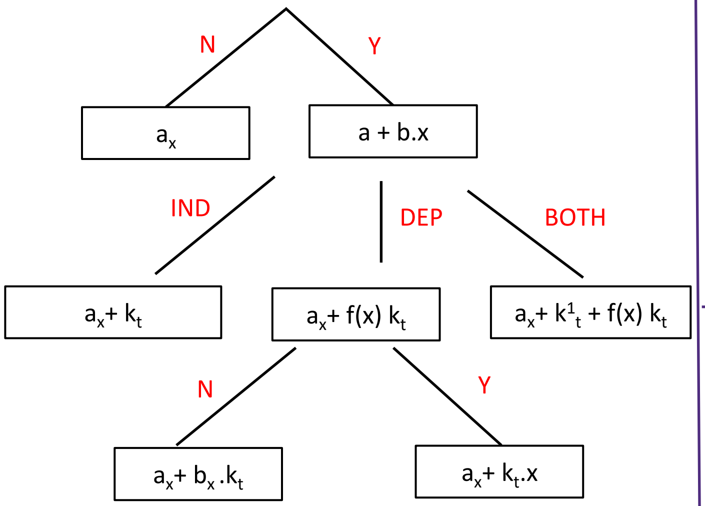


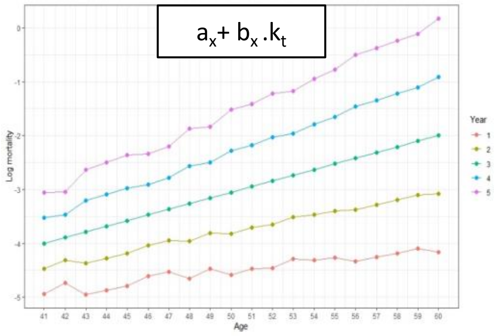


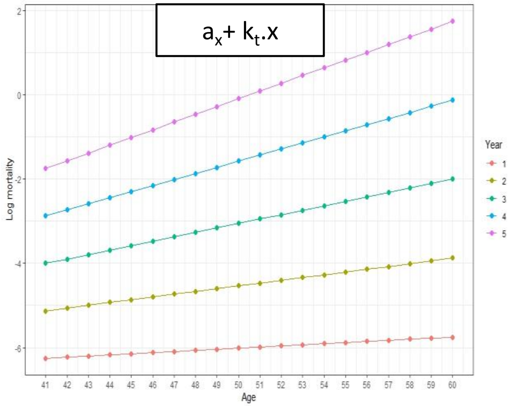
Is the **age** effect linear?

Is there a **period** effect independent and/or an effect interacting with age?

Is the **age-period** effect linear with age?







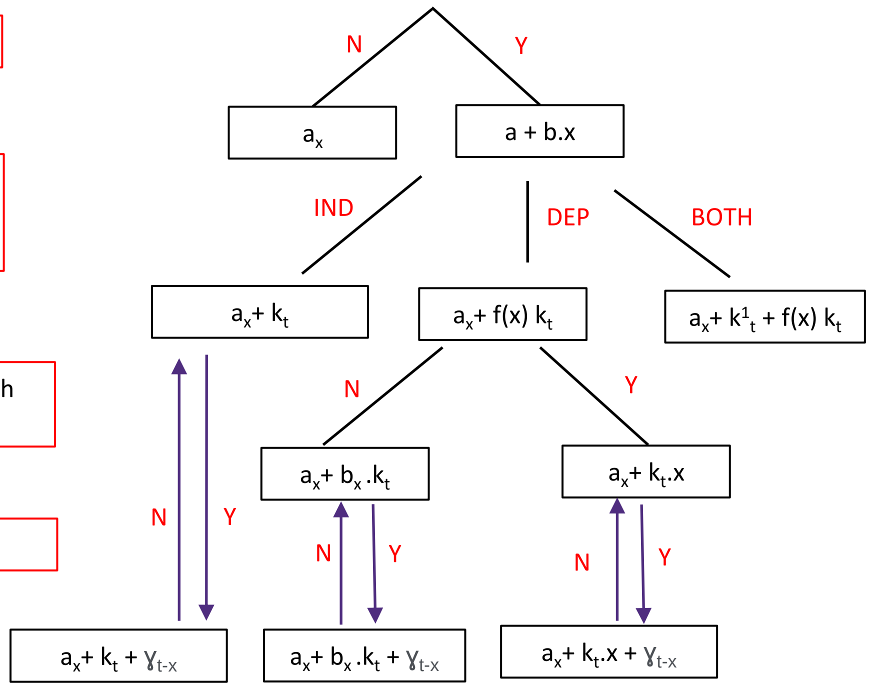
GrantThornton

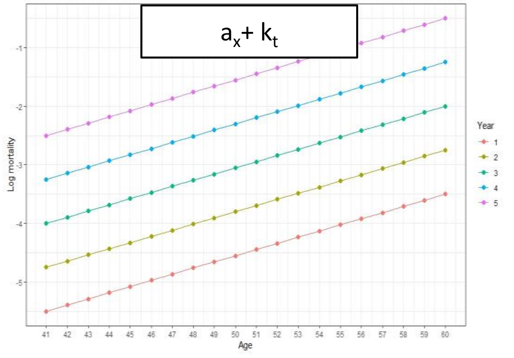
Is the **age** effect linear?

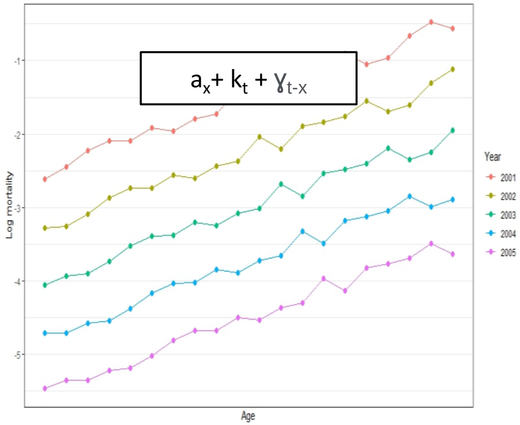
Is there a **period** effect independent and/or an effect interacting with age?

Is the **age-period** effect linear with age?

Is there a **cohort** effect?







Grant Thornton

Model building framework

To view each model as a parameter, we introduce a variable for each of the previous 4 choices:

- Whether the age effect is not linear or not.
- Whether the period effect is independent/dependent on age.
- Whether the age-dependent period effect is linear with age or not.
- Whether there is a cohort effect or not.

Model building framework

Model	δ_1	δ_2	δ_3	δ_4
$a_x + k_t$	1	1	1	1
$a + b(x - \bar{x}) + k_t$	2	1	1	1
$a_x + k_t + k_t b(x - \bar{x})$	1	2	1	1
$k_t^1 + (k_t^1 + b)(x - \bar{x})$	2	2	1	1
$a_x + k_t^1 b_x$ (Lee Carter)	1	2	2	1
$a + b(x - \bar{x}) + k_t b_x$	2	2	2	1
$a_x + k_t^1 + k_t^2(x - \bar{x})$	1	3	1	1
$k_t^1 + k_t^2(x - \bar{x})$ (CBD)	2	3	1	1
$a_x + k_t^1 + k_t^2 b_x$	1	3	2	1
$k_t^1 + b(x - \bar{x}) + k_t^2 b_x$	2	3	2	1
$a_x + k_t + \gamma_{t-x}$ (APC)	1	1	1	2
$a + b(x - \bar{x}) + k_t + \gamma_{t-x}$	2	1	1	2
$a_x + k_t + k_t b(x - \bar{x}) + \gamma_{t-x}$	1	2	1	2
$k_t^1 + (k_t^1 + b)(x - \bar{x}) + \gamma_{t-x}$	2	2	1	2
$a_x + k_t b_x + \gamma_{t-x}$	1	2	2	2
$a + b(x - \bar{x}) + k_t b_x + \gamma_{t-x}$	2	2	2	2
$a_x + k_t^1 + k_t^2(x - \bar{x}) + \gamma_{t-x}$ (APCI)	1	3	1	2
$k_t^1 + k_t^2(x - \bar{x}) + \gamma_{t-x}$	2	3	1	2
$a_x + k_t^1 + k_t^2 b_x + \gamma_{t-x}$	1	3	2	2
$k_t^1 + b(x - \bar{x}) + k_t^2 b_x + \gamma_{t-x}$	2	3	2	2

Example: CBD

$$\kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x})$$

- The age effect is linear.	$a + bx$
- There is an independent period and one interacting with age.	$a + bx + k_t^{(1)} + k_t^{(2)} f(x)$
- The age-period effect is linear with age.	$a + bx + k_t^{(1)} + k_t^{(2)} x$
- There is no cohort effect.	$a + bx + k_t^{(1)} + k_t^{(2)} x$

Example: APC

$$\beta_x^{(1)} + \kappa_t^{(2)} + \gamma_{t-x}^{(3)}$$

- The age effect is nonlinear.	$\beta_x^{(1)}$
- There is a period effect independent of age.	$\beta_x^{(1)} + k_t^{(2)}$
-	
- There is a cohort effect.	$\beta_x^{(1)} + k_t^{(2)} + \gamma_{t-x}^{(3)}$

Example: Lee-Carter

$$\beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)}$$

- The age effect is nonlinear.	$\beta_x^{(1)}$
- There is a period effect dependent of age.	$\beta_x^{(1)} + f(x) k_t^{(2)}$
- The age-period effect is nonlinear with age.	$\beta_x^{(1)} + \beta_x^{(2)} k_t^{(2)}$
- There is no cohort effect.	$\beta_x^{(1)} + \beta_x^{(2)} k_t^{(2)}$

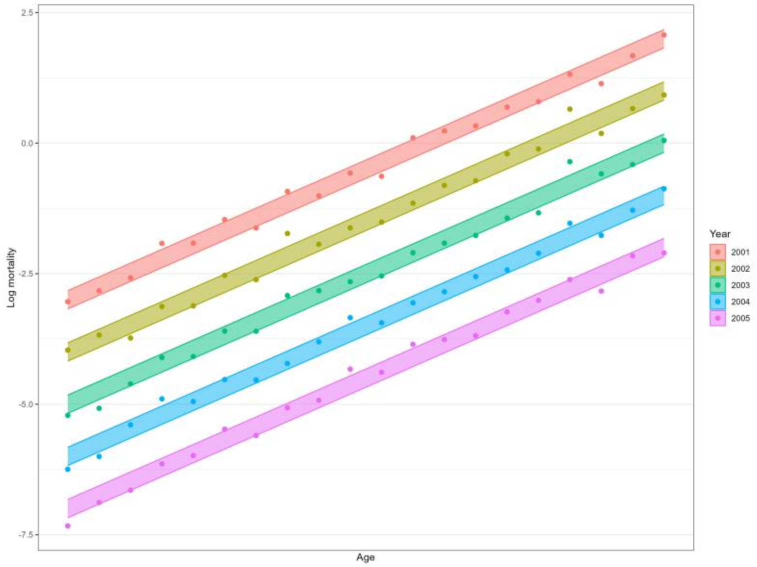
Inference algorithm

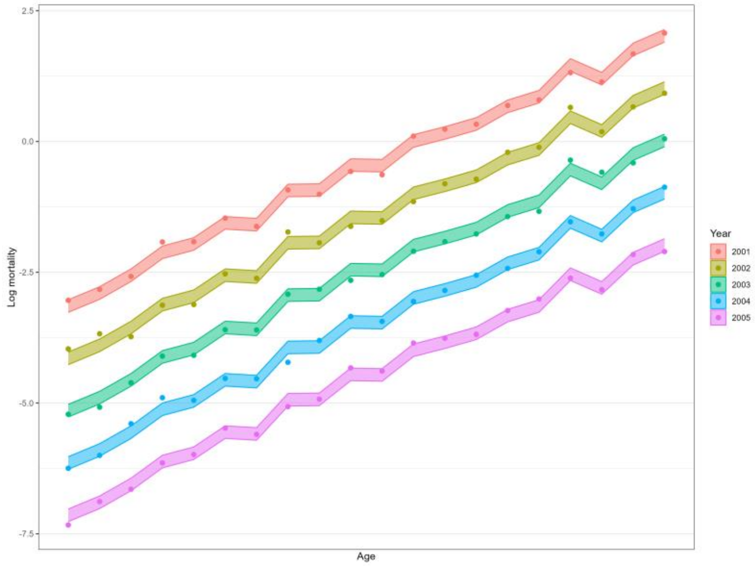
In each iteration, the algorithm will iterate between the possible steps:

- Update the set of parameters θ_1 of the age effect.
- Update the model for the age effect, δ_1.
- Update the set of parameters θ_2 of the period effect.
- Update the model for the period effect, δ_2.
- Update the set of parameters θ_3 of the age-period effect.
- Update the model for the age-period effect, δ_3.
- Update the set of parameters θ_4 of the cohort effect.
- Update the model for the cohort effect, δ_4.

Algorithm example

The algorithm might propose to switch from a linear age effect to a nonlinear age effect

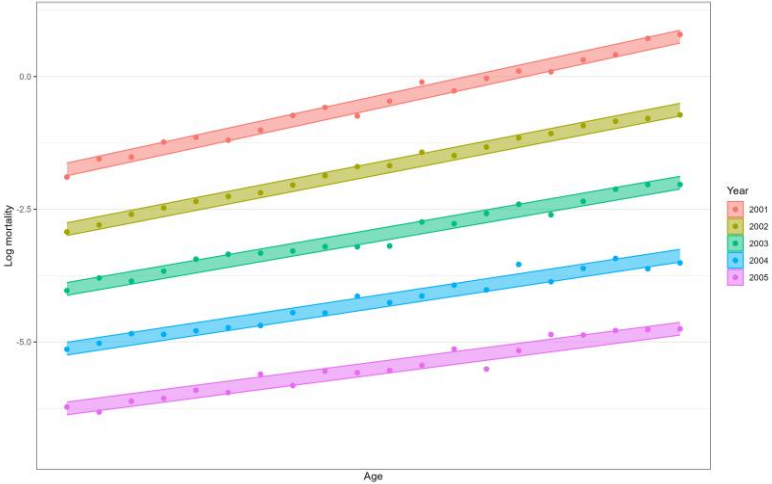


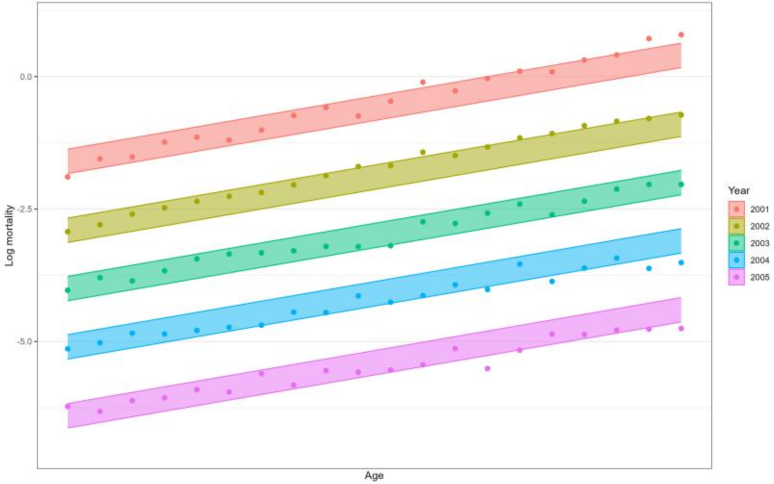


Grant Thornton

Algorithm example

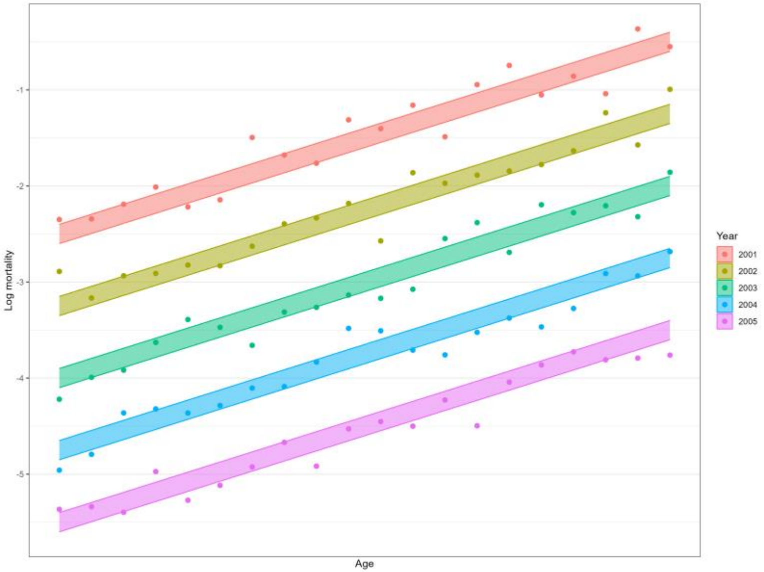
The algorithm might choose between a period effect dependent on age or not.

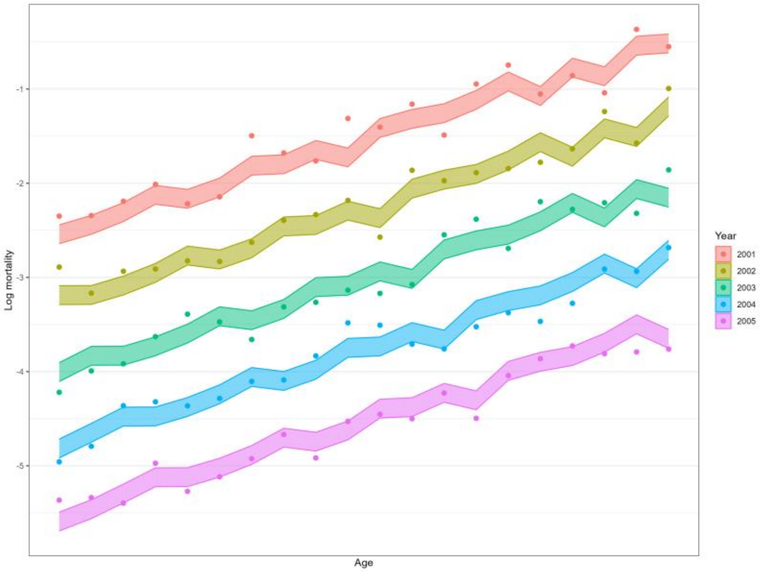




Algorithm example

The algorithm might propose to add a cohort effect or remove it.





Part 3 – Application on DB Pension Schemes

Application on DB pension schemes

We discuss how the profit emergence of a DB pension scheme changes based on the choice of the longevity model

We define profit emergence as:

Profit = Reserves (t-1) – Reserves (t) - actual cashflows (t) + investment returns(t)

Example 1 - Projected mortality rates > Actual	
Reserves (0)	£100
Reserves (1)	-£90
Actual cashflows (1)	-£12
Investment returns (1)	£0
Surplus/Deficit	**-£2**

Example 2 - Projected mortality rates < Actual	
Reserves (0)	£102
Reserves (1)	-£90
Actual cashflows (1)	-£10
Investment returns (1)	£0
Surplus/Deficit	**£2**

Grant Thornton

Application on DB pension schemes

We model the liabilities of a large DB pension scheme as of 2014.

We compare the projected cashflows under using the CMI 2014 to the *actual cashflows** for years 2015-2021 and calculate the profit emergence each year.

We carry sensitivities on the results assuming we had a model which was:

i) 10% more accurate than the CMI 2014

ii) 50% more accurate than the CMI 2014

Grant Thornton

Application on DB pension schemes

Membership Profile 2014

Active	Number	167,545
	Average pensionable salary	£42,729
	Average age	43.8
	Average past service	12.5
Deferred Members	Number	110,430
	Average deferred pension	£2,373
	Average age	45.1
Pensioners	Number	70,380
(including dependents)	Average pension	£17,079
	Average age	71.1

Cashflows

Total contribution: 22% of salary until they retire.

A lump sum is 3 times the salary is paid to the spouse on death.

Deferred pension increases in line with RPI

Annual pension = Pension salary x Pensionable service x Accrual rate

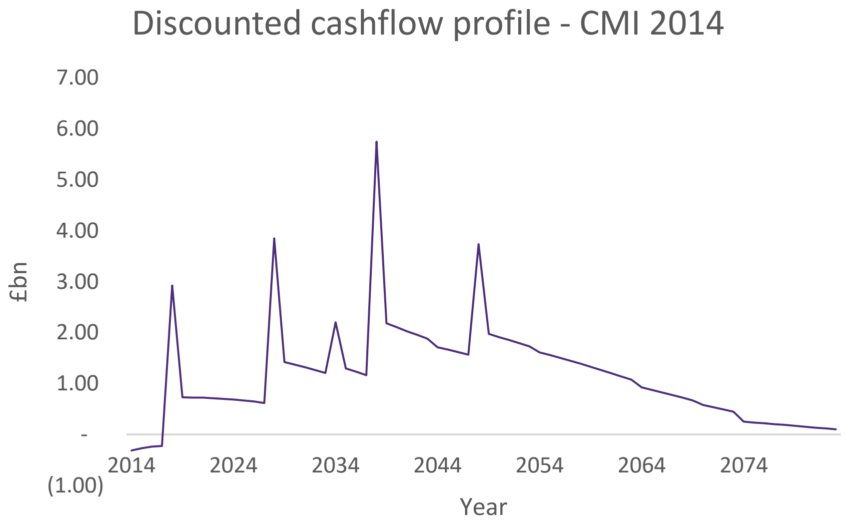
Accrual rate: 1/80th final salary benefit for service to April 1, 2016 but falls to 1/75th after that.

Half the pension paid to the spouse on death

Grant Thornton

Application on DB pension schemes

Cashflow profile produced using the CMI 2014



	CMI 2014			
Year	**Reserve (£bn)**	**Reserve release (£m)**	**Actual Cashflow (£m)**	**Surplus emergence (£m)**
2013	81.96			
2014	82.28	311	311	- 0.00
2015	82.55	272	272	0.19
2016	82.79	241	241	0.13
2017	83.02	226	227	0.35
2018	80.09	- 2,924	- 2,922	1.34
2019	79.36	- 727	- 726	1.15
2020	78.64	- 725	- 723	2.22
2021	77.92	- 720	- 717	2.93
Surplus/Deficit		**- 4,046**	**- 4,037**	**8.31**

Application on DB pension schemes

Surplus emergence between 2014-2021 (£bn)

Projection	CMI 2014	Projection model 10% more accurate	Projection model 50% more accurate
Reserve	81.965	81.959	81.934
Surplus emergence	0.0083	0.0075	0.0042

Grant Thornton

Conclusion

In this presentation, we have discussed:

-Longevity models, their history and what they try to capture.

-Different approaches to longevity model selection

-Our approach to longevity model selection using a Bayesian Reversible Jump

-The impact of choosing different longevity models on surplus emergence of a DB pension scheme.

Grant Thornton

Thank you

grantthornton.co.uk