A Data Analytics Paradigm for the Construction, Selection, and Evaluation of Mortality Models

Andrés VILLEGAS
Dilan SRIDARAN, Michael SHERRIS, and Jonathan ZIVEYI

Insurance Data Science Conference,
ETH, Zurich, 14 June 2019

Acknowledgments

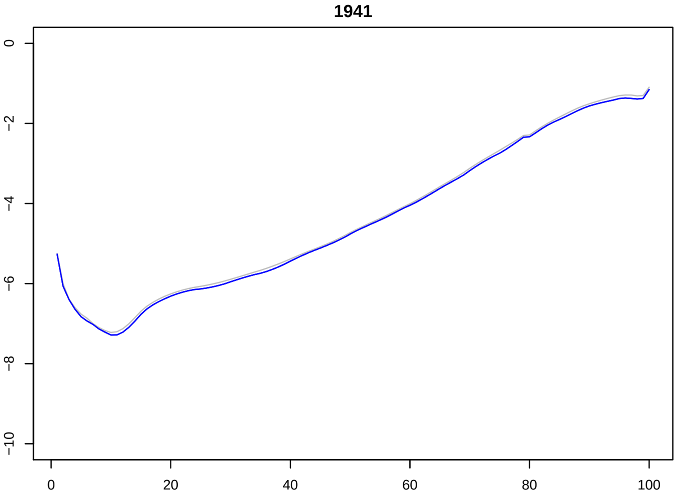
This work is supported by the Society of Actuaries Center of Actuarial Excellence 2016 Research Grant on Longevity Risk and the Australian Research Council Centre of Excellence in Population Ageing Research (CEPAR) project number CE110001029

Agenda

- Review of Mortality Models
- Review of Data Analytics Techniques
- Construction via Regularisation
- Selection via Cross Validation
- Evaluation via Out-of-Sample Forecasting
- Summary

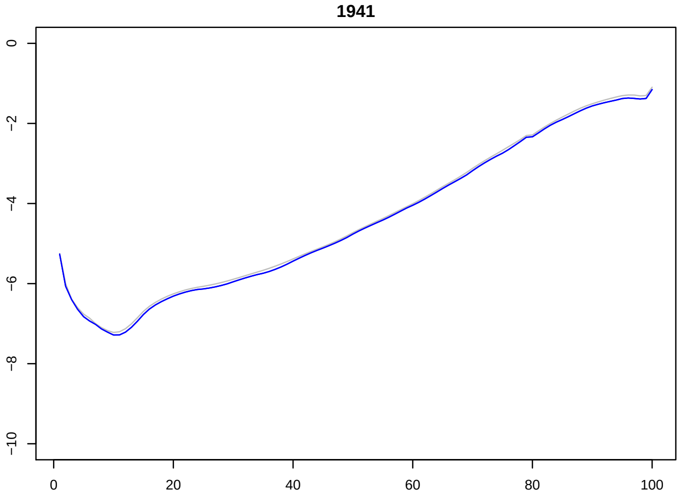
A Data Analytics Paradigm for the Construction, Selection, and Evaluation of **Mortality Models**

Mortality Models: Motivation



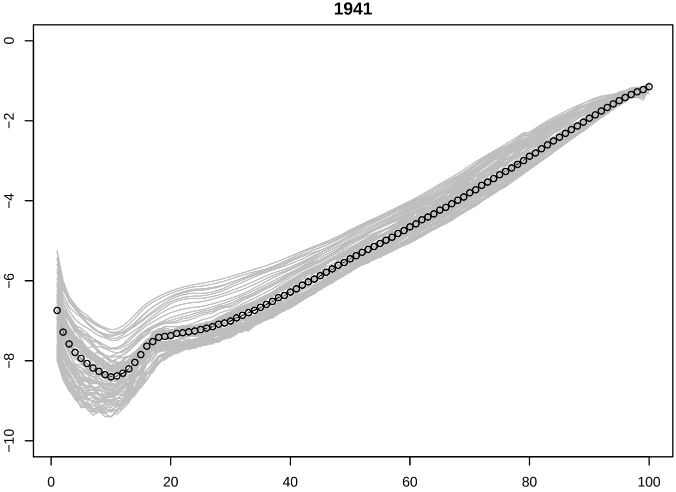
$$\ln(\mu_{x,t}) =$$

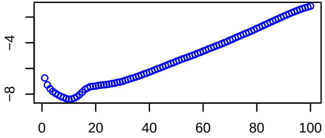
Mortality Models: Motivation



$$\ln(\mu_{x,t}) =$$

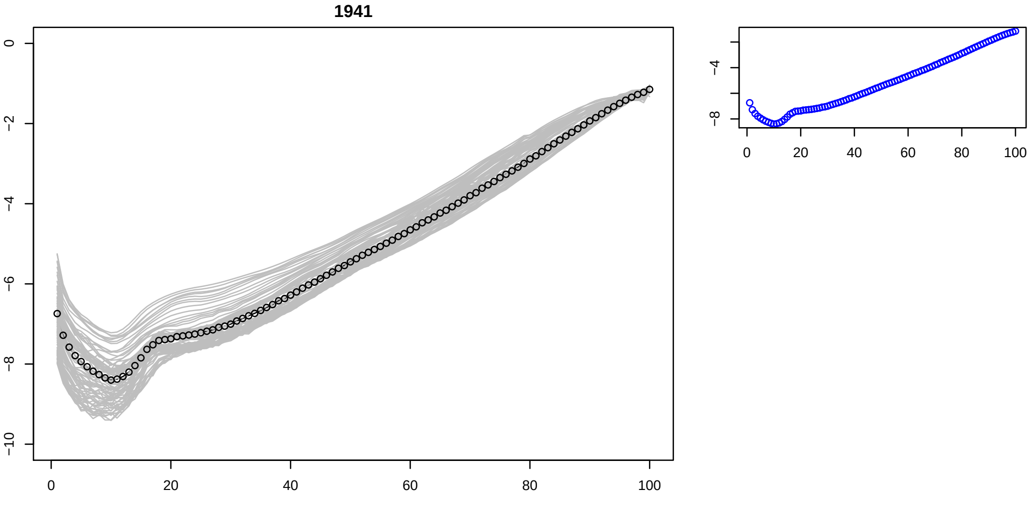
Mortality Models: Constructing Mortality Models





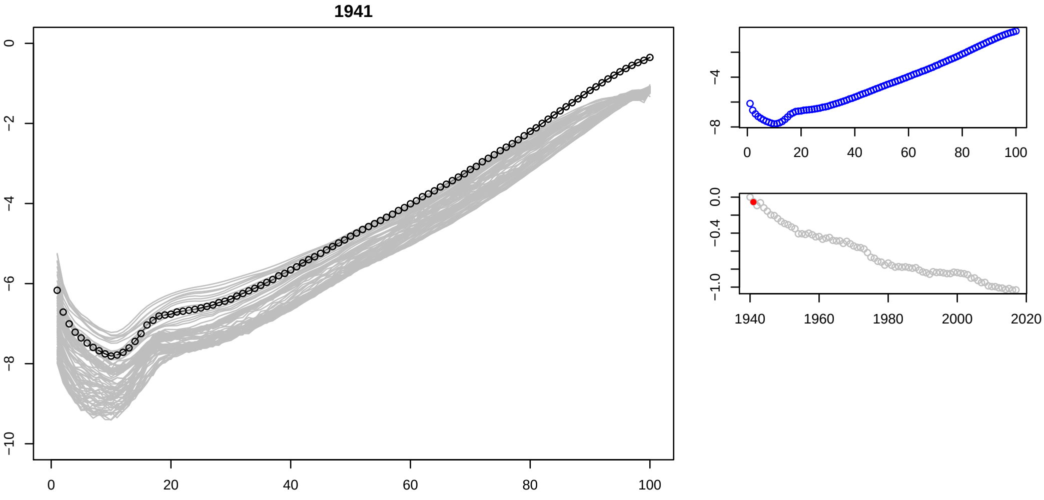
$$\ln(\mu_{x,t}) = \alpha_x$$

Mortality Models: Constructing Mortality Models



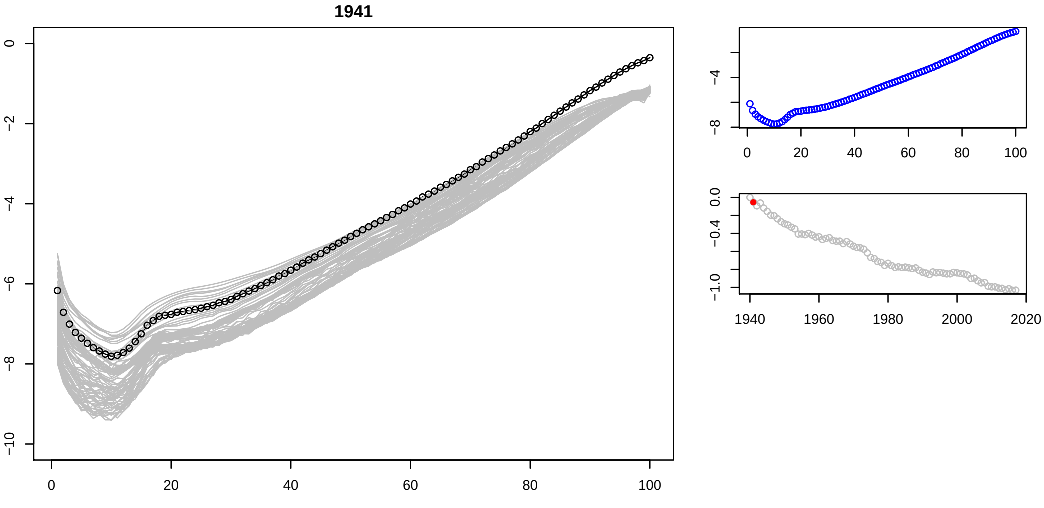
$$\ln(\mu_{x,t}) = \alpha_x$$

Mortality Models: Constructing Mortality Models



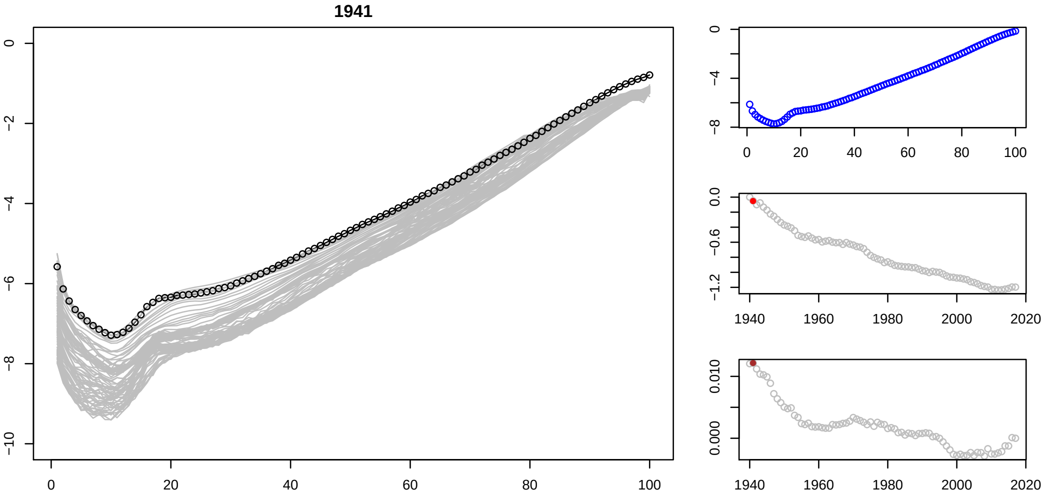
$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)}$$

Mortality Models: Constructing Mortality Models



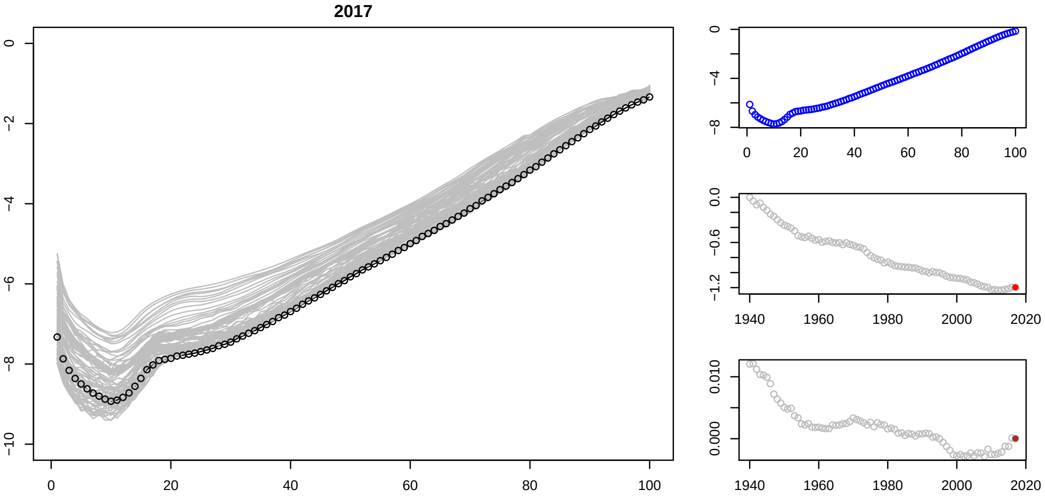
$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)}$$

Mortality Models: Constructing Mortality Models



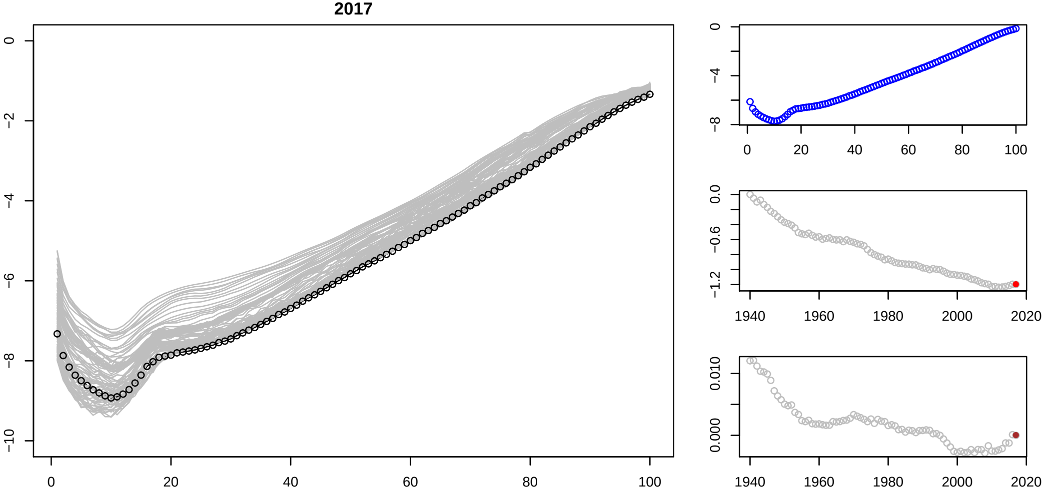
$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)}$$

Mortality Models: Constructing Mortality Models



$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)}$$

Mortality Models: Constructing Mortality Models



$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_c$$

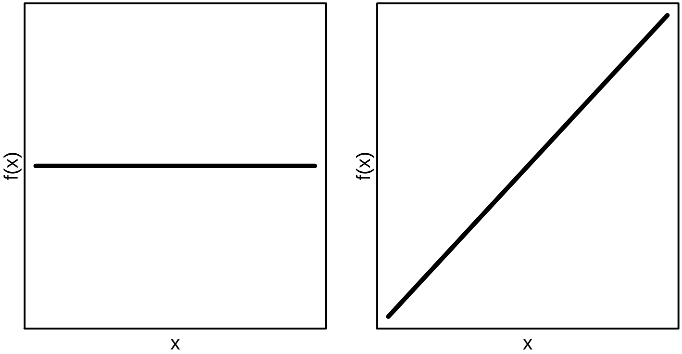
Mortality Models: Generalised Age-Period-Cohort Models

$$ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{N} f^{(i)}(x)\kappa_t^{(i)} + \gamma_c$$

Name	Form	Parameters
LC	$\alpha_x + \beta_x^{(1)}\kappa_t^{(1)}$	$2n_a + n_y$
LC2	$\alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}$	$3n_a + 2n_y$
RH	$\alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(0)}\gamma_c$	$3n_a + n_y + n_c$
APC	$\alpha_x + \kappa_t^{(1)} + \gamma_c$	$n_a + n_y + n_c$
CBD	$\kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)}$	$2n_y$
M7	$\kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + ((x - \bar{x})^2 - \sigma_x^2)\kappa_t^{(3)} + \gamma_c$	$3n_y + n_c$
sPLAT	$\alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + \gamma_c$	$n_a + 2n_y + n_c$
cPLAT	$\alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + (\bar{x} - x)^+\kappa_t^{(3)} + \gamma_c$	$n_a + 3n_y + n_c$

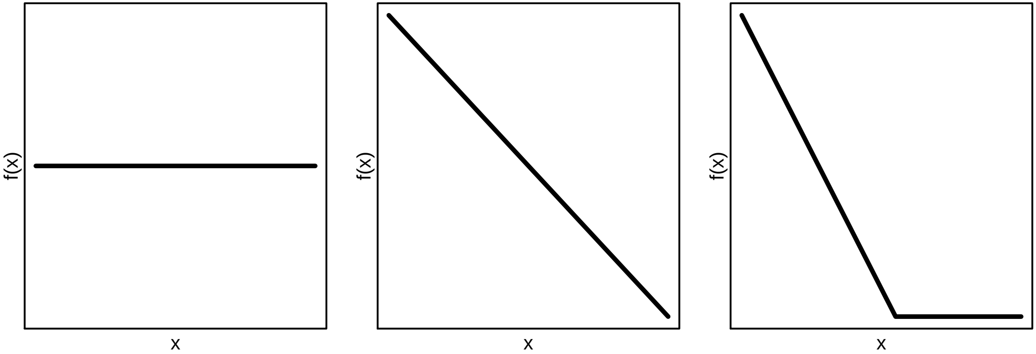
Mortality Models: CBD Model

$$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x})\,\kappa_t^{(2)}$$



Mortality Models: cPLAT Model

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\,\kappa_t^{(2)} + (\bar{x} - x)^+\,\kappa_t^{(3)} + \gamma_{t-x}$$



A **Data Analytics** Paradigm for the Construction, Comparison, and Selection of Mortality Models

Data Analytics: Predictive Modelling

Objective: Predict response using linear combination of predictors,

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_j, \tag{1}$$

where β_0 and $\beta = (\beta_1, \beta_2, ..., \beta_p)$ are unknown parameters.

Data Analytics: Predictive Modelling

Objective: Predict response using linear combination of predictors,

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_j, \tag{1}$$

where β_0 and $\beta = (\beta_1, \beta_2, ..., \beta_p)$ are unknown parameters.

Typically, parameters are estimated by minimising OLS,

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2. \tag{2}$$

Data Analytics: Predictive Modelling

Objective: Predict response using linear combination of predictors,

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_j, \tag{1}$$

where β_0 and $\beta = (\beta_1, \beta_2, ..., \beta_p)$ are unknown parameters.

Typically, parameters are estimated by minimising OLS,

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2. \tag{2}$$

Problem: *Interpretation* and *prediction accuracy*.

Data Analytics: Regularisation

Solution: Minimise

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2, \tag{3}$$

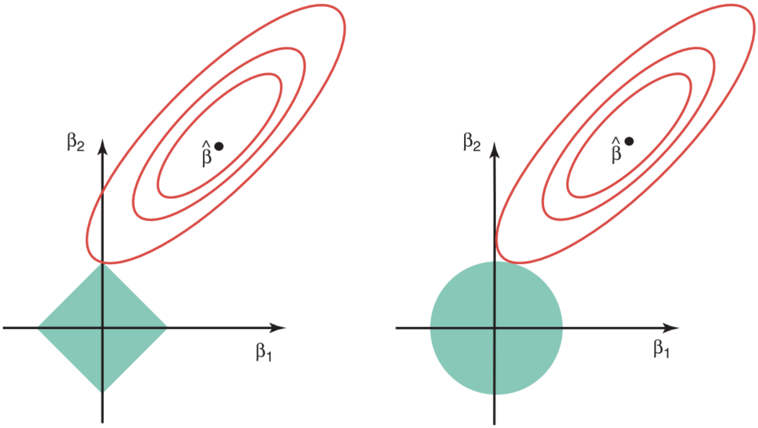
subject to ℓ_2-norm constraint (ridge regression),

$$\sum_{j=1}^{p}\beta_j^2 \leq t, \tag{4}$$

or ℓ_1-norm constraint (lasso),

$$\sum_{j=1}^{p}|\beta_j| \leq t. \tag{5}$$

Data Analytics: Lasso and Ridge Constraints



James et al. (2014)

Data Analytics: Group Lasso

Consider the linear predictor structure,

$$\boldsymbol{y} = \mathbf{X}\beta = \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j, \tag{6}$$

where $\boldsymbol{\beta}_j$ is a vector of coefficients for group j.

Data Analytics: Group Lasso

Consider the linear predictor structure,

$$\boldsymbol{y} = \mathbf{X}\beta = \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j, \tag{6}$$

where $\boldsymbol{\beta}_j$ is a vector of coefficients for group j.

The group lasso minimises the objective function,

$$Q(\boldsymbol{\beta}|\mathbf{X}, \boldsymbol{\eta}) = L(\boldsymbol{\beta}|\mathbf{X}, \boldsymbol{\eta}) + \sum p_\lambda(\|\boldsymbol{\beta_j}\|), \tag{7}$$

where $p_\lambda(\cdot)$ is a penalty applied to the ℓ_2-norm of each group.

Data Analytics: Group Lasso

Consider the linear predictor structure,

$$\boldsymbol{y} = \mathbf{X}\beta = \sum_{j=1}^{J} \mathbf{X}_j \boldsymbol{\beta}_j, \tag{6}$$

where $\boldsymbol{\beta}_j$ is a vector of coefficients for group j.

The group lasso minimises the objective function,

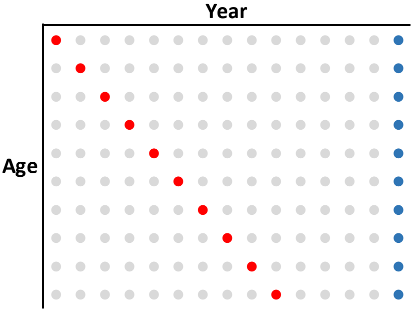
$$Q(\boldsymbol{\beta}|\mathbf{X}, \boldsymbol{\eta}) = L(\boldsymbol{\beta}|\mathbf{X}, \boldsymbol{\eta}) + \sum p_\lambda(\|\boldsymbol{\beta}_{\boldsymbol{j}}\|), \tag{7}$$

where $p_\lambda(\cdot)$ is a penalty applied to the ℓ_2-norm of each group.

If group j is selected, $\beta_{jk} \neq 0$ for all k; else $\beta_{jk} = 0$ for all k.

A Data Analytics Paradigm for the Construction, Comparison, and Selection of Mortality Models

Construction: Data



Data Source: Human Mortality Database (30 Countries, 2 Genders)

Years: 1960 - 2015

Ages: 20 - 89

Construction: “Formalised” Model-Building Framework

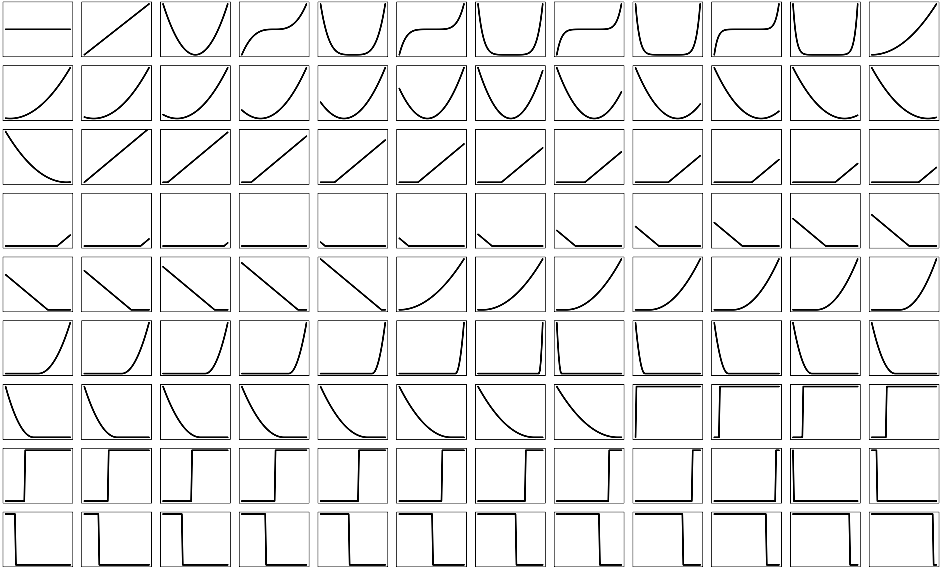
We start with a **huge** model,

$$\ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{B} f^{(i)}(x)\kappa_t^{(i)} + \gamma_{t-x},$$

where the suite of basis functions $(f^{(i)}(x))$ included are:

$$f^{(i)}(x) = \begin{cases} 1, & \text{Unit} \\ (x-\bar{x})^n & \text{Polynomial} \\ (x-n)^+ & \text{Call} \\ (n-x)^+ & \text{Put} \\ 1_{x<n} & \text{Below} \\ 1_{x>n} & \text{Above} \end{cases} . \tag{8}$$

Construction: Our Model



Construction: GLM Representation

$$\eta_{x,t} = \ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{B} f^{(i)}(x)\kappa_t^{(i)} + \gamma_c,$$

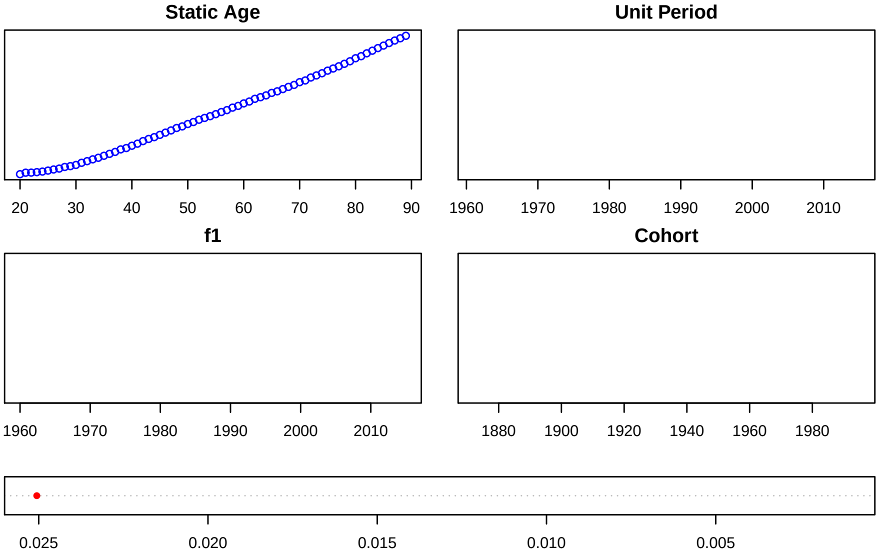
can be expressed as a GLM,

$$\eta = \mathbf{X}\beta = \sum_{j=0}^{B+1} \mathbf{X}_j \boldsymbol{\beta}_j, \qquad \mathbf{X} = [\mathbf{X_0} : \mathbf{X_1} : \mathbf{X_2} : \ldots : \mathbf{X_B} : \mathbf{X_{B+1}}],$$

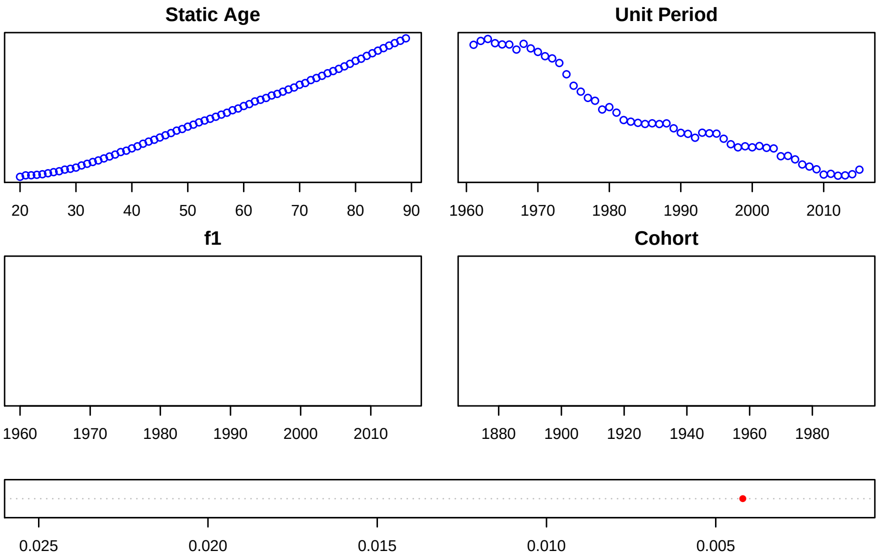
where,

$$\boldsymbol{\beta} = \{\boldsymbol{\beta_i}\}_{i=0}^{B+1}, \quad \boldsymbol{\beta_0} = \{\alpha_x\}_{x=1}^{n_x}, \quad \boldsymbol{\beta_i} = \{\kappa_t^{(i)}\}_{t=1}^{n_t}, \quad \boldsymbol{\beta_{B+1}} = \{\gamma_c\}_{c=1}^{n_c}.$$

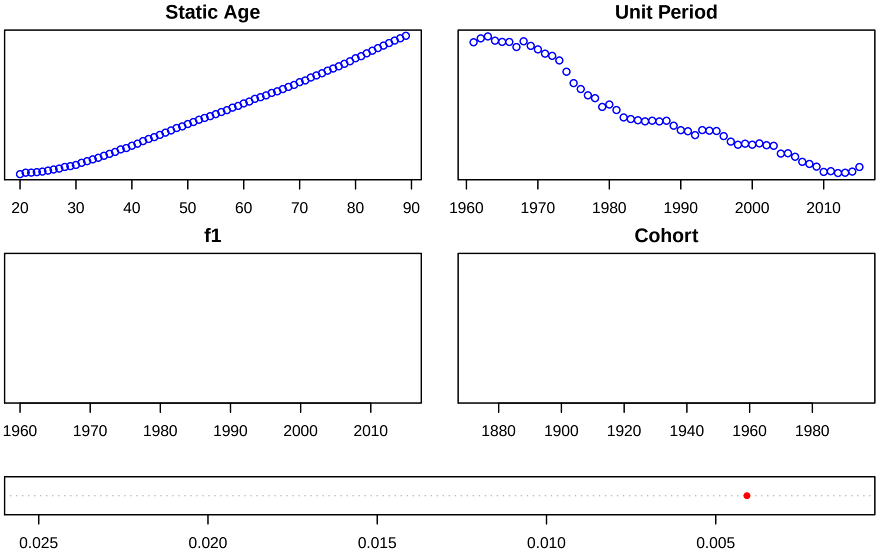
Construction: Estimated Parameters



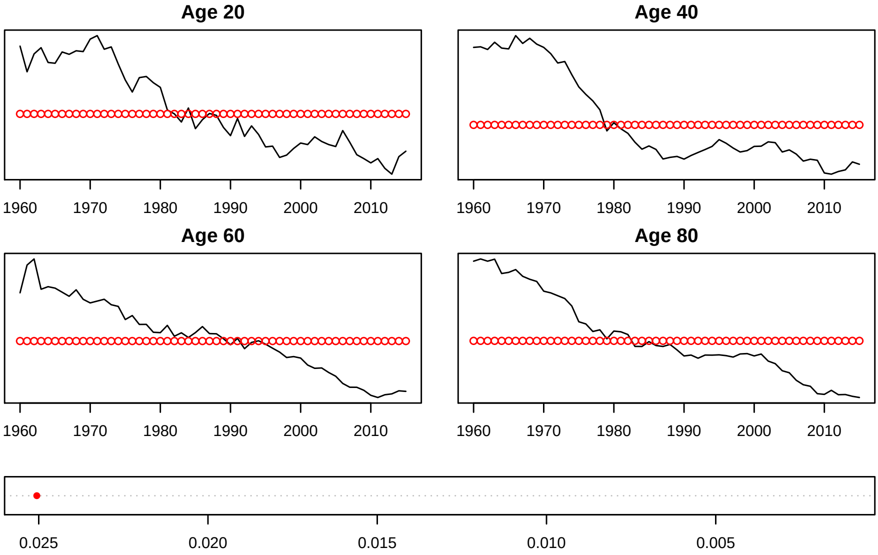
Construction: Estimated Parameters



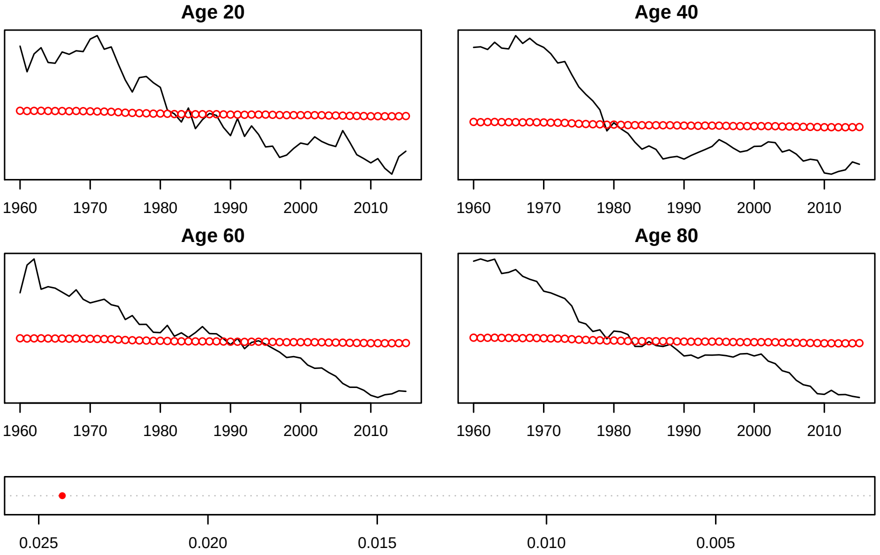
Construction: Estimated Parameters



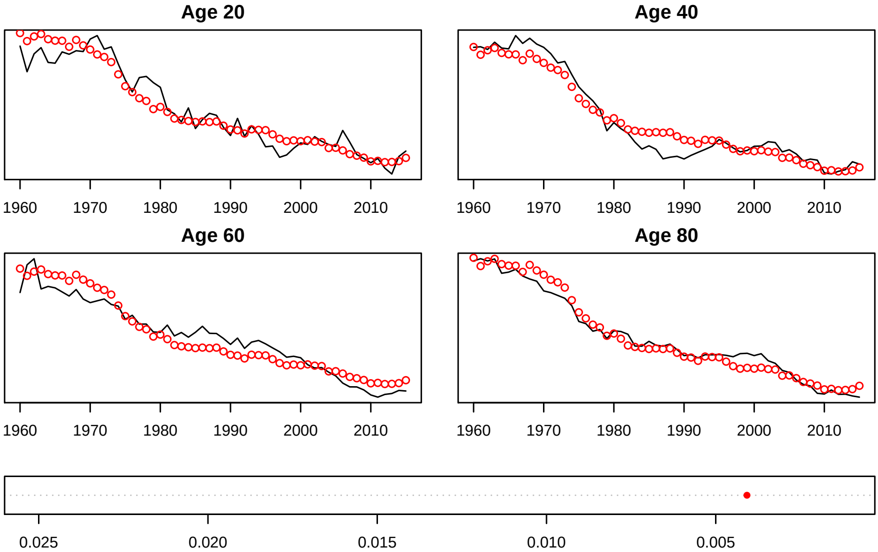
Construction: Fitted Values



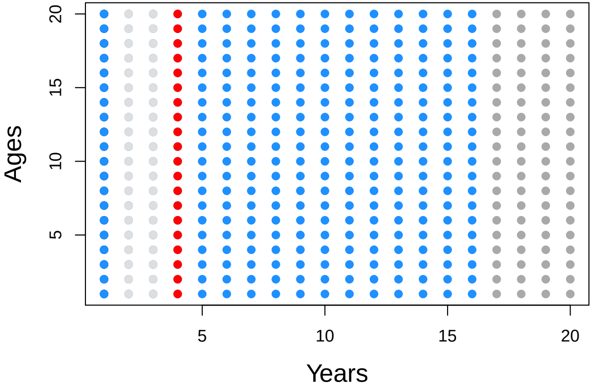
Construction: Fitted Values



Construction: Fitted Values



Selection and Evaluation: Cross Validation Framework

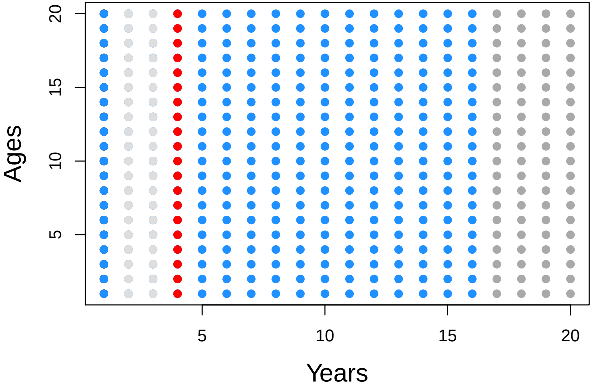


Training/Test Set: 1960 - 1990

Validation Set: 1991 - 2015

Test Set Width: Forecasting Horizon

Selection and Evaluation: Cross Validation Framework



Training/Test Set: 1960 - 1990

Validation Set: 1991 - 2015

Test Set Width: Forecasting Horizon

Selection Results: Optimal Models

Data	1 Year Horizon	Parameters
USA (F)	$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_c$	262
AUS (M)	$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})^2\kappa_t^{(2)} + (x - 25)^+\kappa_t^{(3)} + (25 - x)^+\kappa_t^{(4)} + \gamma_c$	344
UK (M)	$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (x - \bar{x})^2\kappa_t^{(3)} + \gamma_c$	303
CAN (F)	$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (25 - x)^+\kappa_t^{(3)} + \gamma_c$	303

Data	10 Year Horizon	Parameters
USA (F)	$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_c$	262
AUS (M)	$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})^2\kappa_t^{(2)} + \gamma_c$	262
UK (M)	$\alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (x - \bar{x})^2\kappa_t^{(3)} + \gamma_c$	303
CAN (F)	$\alpha_x + \kappa_t^{(1)}$	111

Evaluating Results: By Forecasting Horizon

	1 Year	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9
1	GL	27	10	16	5	1	0	1	0	0
2	LC2	21	14	6	7	7	4	1	0	0
3	cPLAT	4	13	3	6	6	17	9	2	0
4	LC	3	6	6	9	3	10	10	8	5
6	APC	1	10	12	15	14	4	2	2	0
8	sPLAT	0	1	7	7	15	12	13	5	0

	10 Years	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9
1	GL	15	17	9	11	5	3	0	0	0
2	APC	15	11	13	15	2	3	1	0	0
3	LC	13	10	11	15	10	1	0	0	0
4	LC2	7	15	22	11	4	1	0	0	0
7	sPLAT	0	0	1	0	4	19	21	10	5
8	cPLAT	0	0	0	1	3	5	25	22	4

Evaluating Results: By Population Size

Size (m)	LC	LC2	CBD	M7	APC	RH	sPLAT	cPLAT	GL
0.00-2.50	5.1	1.2	7.2	7.4	3.4	5.3	6.5	5.9	3.0
2.50-3.75	3.8	2.3	7.7	8.2	3.6	4.5	6.5	5.9	2.5
3.75-5.00	5.1	2.4	8.3	7.3	3.5	5.2	5.8	5.7	1.7
5.00-10.00	4.1	2.6	9.0	7.4	3.5	5.3	5.8	5.1	2.0
10.00-25.00	6.6	3.4	8.7	5.7	4.9	6.0	4.5	3.4	1.8
$>$25.00	7.0	3.9	8.8	5.9	4.8	6.7	4.2	1.9	1.8

Key Takeaways

- Complex models perform well for short-term forecasting;
- Simpler models perform well for long-term forecasting;
- Complex models perform well on large populations;
- Complex models perform poorly on small populations;
- Simpler models perform relatively consistently across population size;
- Complex models produce very volatile outputs;
- Small, less developed populations do not exhibit strong cohort effects.

No pre-defined model will perform well for all contexts.

Future work and limitations

- **Estimation**: Gaussian assumption
  - Extension to a Poisson setting
- **Selection**: Based exclusively on MSE (point estimates)
  - Interval based-measures, life-span disparity
- **Construction**: Focus on single population mortality rates
  - Multiple populations, improvement rates
- Integration of the new techinques into the R Package **StMoMo**

Questions?

Email: a.villegas@unsw.edu.au