

# Developments in the application of data analytics techniques to mortality modelling

Andrés VILLEGAS

CEPAR Longevity Risk Workshop  
UNSW Sydney, 27 November 2019



UNSW | AGSM  
Business School

# Acknowledgements

Work supported by

- ▶ Society of Actuaries Center of Actuarial Excellence Research Grant 2017-2020: Longevity Risk: Actuarial and Predictive Models, Retirement Product Innovation, and Risk Management Strategies
- ▶ CEPAR Australian Research Council Centre of Excellence in Population Ageing Research project number CE170100005 for 2018{2024.



**Michael Sherris**



**Jonathan Ziveyi**



**Dilan SriDaran**

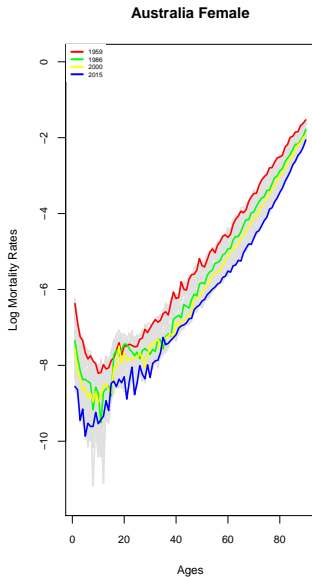
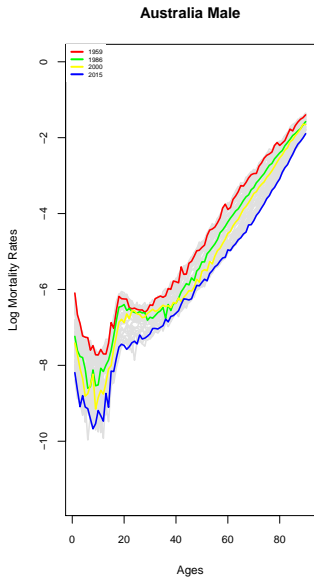


**Salvatory Kessy**

# Agenda

- ▶ Motivation
- ▶ Review of developments in mortality modelling
- ▶ Two examples
  1. A data analytics paradigm for the construction, selection, and evaluation of mortality models
  2. Mortality model combination using stacked regression ensembles
- ▶ Conclusions and outlook

# Motivation





# Motivation

- ▶ **Unexpected mortality improvements** for senior populations have **substantial economic consequences**.
- ▶ The estimated potential size of the global longevity risk market for pension liabilities is between **USD 60trn and USD 80trn (Blake et al. 2018)**.
- ▶ Each **additional year of unanticipated life expectancy** at age 65 can increase pension liabilities by 4%– 5%.
- ▶ Diverse models such as the Lee-Carter fail to fully capture the human mortality dynamics.
- ▶ A growing need for **accurately modeling and forecasting mortality rates and life expectancy**.
- ▶ Even a small improvement in the mortality forecasting accuracy can provide considerable financial savings among different entities.

# Motivation

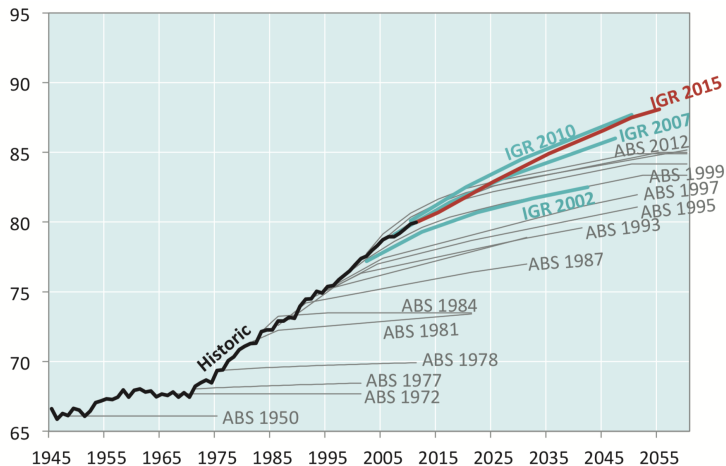


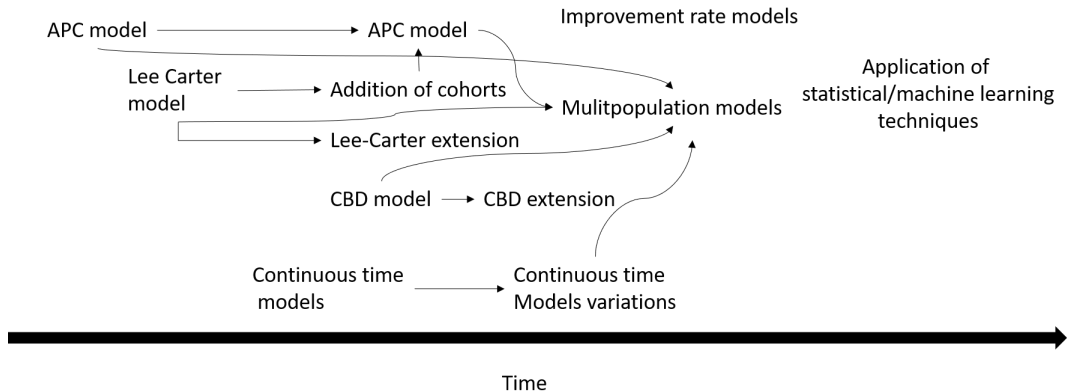
Figure 2: Actual and projected period life expectancy at birth, Australia males, 1945-2055.  
Source: CEPAR (2015).

# Mortality forecasting methodologies

A good overview of methodologies is given in the review papers Booth and Tickle (2008), Wong-Fillipow and Haberman (2004), Pitacco (2004) and in book Pitacco et al. (2009)

- ▶ Expert based
- ▶ Explanatory
  - ▶ Structural Modelling (Explanatory or Econometric).
  - ▶ Cause of death decomposition
- ▶ **Extrapolation**
  - ▶ Trend modelling

# A timeline of “recent” mortality modelling methodologies



# Extrapolation mortality models.

- ▶ Extrapolative methods are still compelling for modeling and predicting mortality rates in actuarial readership.
- ▶ Examples include
  - ▶ Parametric mortality models such as the Lee-Carter model and Cairns-Blake-Dowd (CBD).
  - ▶ Non-parametric models such as functional principal analysis (Hyndman and Shahid Ullah 2007).
  - ▶ Bayesian probabilistic mortality models (Wiśniowski et al. 2015).

# Statistical Machine Learning Techniques for Mortality Modeling

Statistical learning methods are used to complement the existing standard stochastic mortality models rather than substituting them.

- ▶ Neural networks
- ▶ Clustering
- ▶ Trees
- ▶ Gaussian Processes

# Statistical Machine Learning Techniques for Mortality Modeling: Neural Networks

- ▶ Nigri et al. (2019) integrated a **recurrent neural network** with a long short-term memory architecture in Lee-Carter model.
- ▶ (Hainaut 2018) applied **neural networks** to learn the logarithm of the mortality death rates directly from the features of the mortality data.
- ▶ Richman and Wuthrich (2018) extended Lee-Carter model for multi-population modeling using **Artificial Neural Network**.

# Statistical Machine Learning Techniques for Mortality Modeling: Clustering

- ▶ Piscopo and Resta (2017) applied spectral biclustering to Italian mortality data.
- ▶ Hatzopoulos and Haberman (2011) applied fuzzy clustering to cluster the 35 European countries' populations for the period 1960-2006 into similar groups.
- ▶ Carracedo et al. (2018) applied the Spatio-temporal framework to identify the significant clusters among the 26 European countries based on time and the geographical location.



# Statistical Machine Learning Techniques for Mortality Modeling: Other Methods

- ▶ (Ludkovski, Risk, and Zail 2016) used **Gaussian Process** (GP) regression to model the mortality rates.
- ▶ Deprez, Shevchenko, and Wüthrich (2017) and Levantesi and Pizzorusso (2019) integrated **tree methods** in Lee-Carter and its extensions.

# Statistical Machine Learning Techniques for Mortality Modeling: Two detailed examples

In what follows we will discuss two examples of applying fairly standard statistical learning approaches to mortality modelling:

1. Regularisation and cross-validation
2. Ensemble learning

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

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

# Mortality Models: Constructing Traditional Mortality Models

# Mortality Models: Constructing Traditional Mortality Models

# Mortality Models: Constructing Traditional Mortality Models

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

# Mortality Models: Constructing Traditional Mortality Models

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

+



# Mortality Models: Constructing Traditional Mortality Models

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

# Mortality Models: Constructing Traditional Mortality Models

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

# Mortality Models: Constructing Traditional Mortality Models

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

# Mortality Models: Constructing Traditional Mortality Models

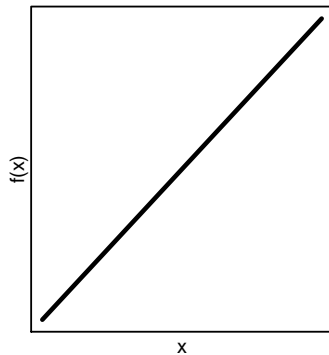
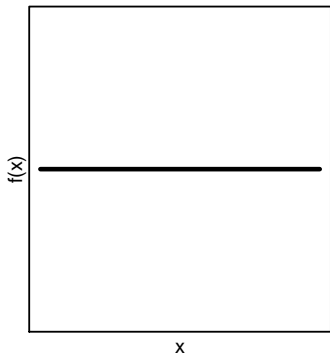
$$\ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^N f^{(i)}(x) \kappa_t^{(i)} + \gamma_{t-x}$$

# Mortality Models: Generalised Age-Period-Cohort Models

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	$\alpha_x + \kappa_t^{(1)} + (\bar{x} - x) \kappa_t^{(2)} + (\bar{x} - x)^+ \kappa_t^{(3)} + \gamma_c$	$n_a + 3n_y + n_c$

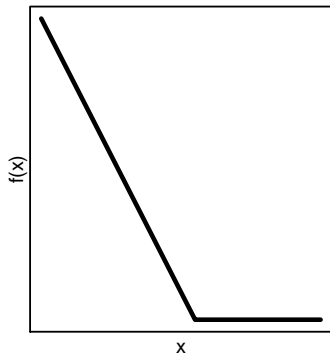
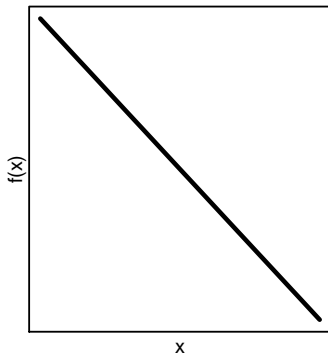
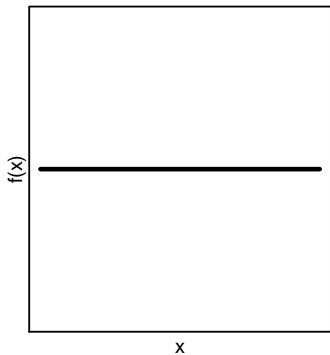
# Mortality Models: CBD Model

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



# Mortality Models: cPLAT Model

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



# Mortality Models: Key research questions

1. What model features are desired for different applications?
2. Why do we only consider a fixed set of models?
3. How can we be confident we have selected the best model?



# Mortality Models: What is the “**best**” model?

**Goodness-of-fit:** Cairns et al. (2009) and Dowd et al. (2010b)

- ▶ Desirable theoretical properties, e.g. parsimony
- ▶ BIC, likelihood ratio test, and analysis of robustness

**Forecasting:** Cairns et al. (2011) and Dowd et al. (2010a)

- ▶ Biological reasonableness, plausibility, and robustness
- ▶ Ad-hoc backtests

# Our objective

Provide a comprehensive framework to **construct**, **select**, and **evaluate** discrete-time mortality models for forecasting applications, using various **statistical learning** and predictive analytics techniques.

- ▶ **Construction** based on regularisation techniques
- ▶ **Selection** based on cross-validation techniques

# 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,

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

where  $\beta_0$  and  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  are unknown parameters.

# Data Analytics: Predictive Modelling

**Objective:** Predict response using linear combination of predictors,

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

where  $\beta_0$  and  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  are unknown parameters.

Typically, parameters are estimated by minimising OLS,

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

# Data Analytics: Predictive Modelling

**Objective:** Predict response using linear combination of predictors,

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

where  $\beta_0$  and  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  are unknown parameters.

Typically, parameters are estimated by minimising OLS,

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

**Problem:** *Interpretation and prediction accuracy.*

# Data Analytics: Regularisation

**Solution:** Minimise

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

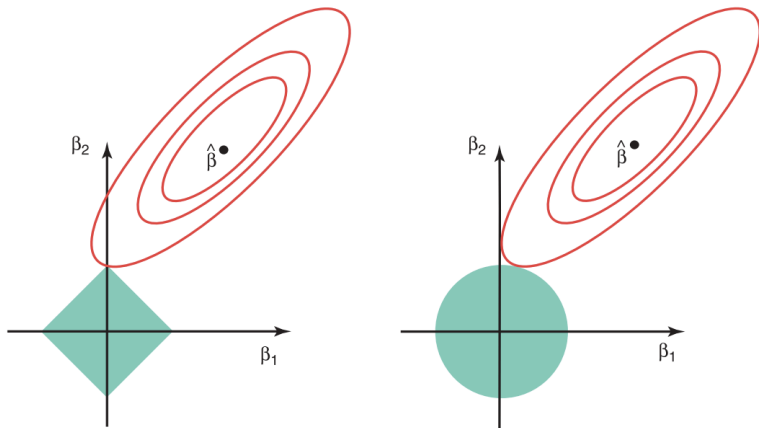
subject to  $\ell_2$ -norm constraint (ridge regression),

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

or  $\ell_1$ -norm constraint (lasso),

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

# Data Analytics: Lasso and Ridge Constraints



James et al. (2014)



# Data Analytics: Group Lasso

Consider the linear predictor structure,

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

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

# Data Analytics: Group Lasso

Consider the linear predictor structure,

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

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

The group lasso minimises the objective function,

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

where  $p_\lambda(\cdot)$  is a penalty applied to the  $\ell_2$ -norm of each group.

# Data Analytics: Group Lasso

Consider the linear predictor structure,

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

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

The group lasso minimises the objective function,

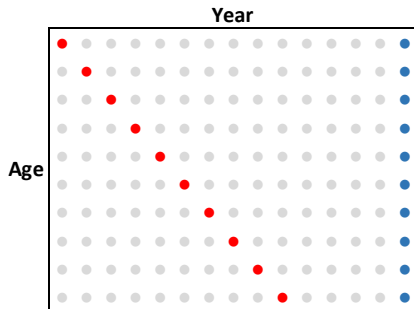
$$Q(\boldsymbol{\beta}|\mathbf{X}, \boldsymbol{\eta}) = L(\boldsymbol{\beta}|\mathbf{X}, \boldsymbol{\eta}) + \sum p_\lambda(\|\boldsymbol{\beta}_j\|), \quad (7)$$

where  $p_\lambda(\cdot)$  is a penalty applied to the  $\ell_2$ -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

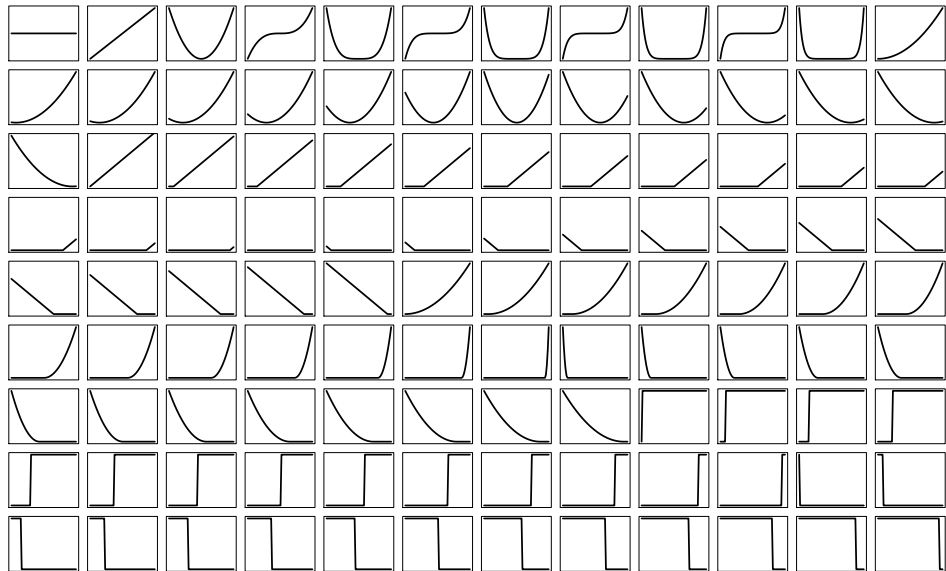
We start with a **huge** model,

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

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

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

## 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 (Currie 2016),

$$\eta = \mathbf{X}\beta = \sum_{j=0}^{B+1} \mathbf{X}_j \beta_j, \quad \mathbf{X} = [\mathbf{X}_0 : \mathbf{X}_1 : \mathbf{X}_2 : \dots : \mathbf{X}_B : \mathbf{X}_{B+1}],$$

where,

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



# Construction: Parameter Estimation

We minimise

$$Q(\beta|\mathbf{X}, \boldsymbol{\eta}) = \frac{1}{2n} \|\boldsymbol{\eta} - \mathbf{X}\beta\| + \sum_{i=1}^B p_{\lambda}(\|\boldsymbol{\kappa}^{(i)}\|) + p_{\lambda}(\|\boldsymbol{\gamma}\|). \quad (9)$$

with the minimax concave penalty (MCP)

$$p_{\lambda, \gamma}(\theta) = \begin{cases} \lambda\theta - \frac{\theta^2}{2\gamma}, & \text{if } \theta \leq \gamma\lambda, \\ \frac{1}{2}\gamma\lambda^2, & \text{if } \theta > \gamma\lambda, \end{cases} \quad (10)$$

- ▶ Fix  $\gamma = 3$
- ▶  $\lambda$  controls the trade-off between goodness-of-fit and parsimony
- ▶ Use the R package **grpreg** (Breheny and Huang 2013)

```
## StMoMo: Start fitting with grpreg
```

```
## StMoMo: Finish fitting with grpreg
```

# Construction: Estimated Parameters (USA Males)

# Construction: Estimated Parameters (USA Males)

# Construction: Estimated Parameters (USA Males)

# Construction: Estimated Parameters (USA Males)

## Construction: Fitted Values (USA Males)

## Construction: Fitted Values (USA Males)

## Construction: Fitted Values (USA Males)



## Construction: Fitted Values (USA Males)

# 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: 5 Year Horizon (Training vs. Test error)

# Selection Results: Optimal Models

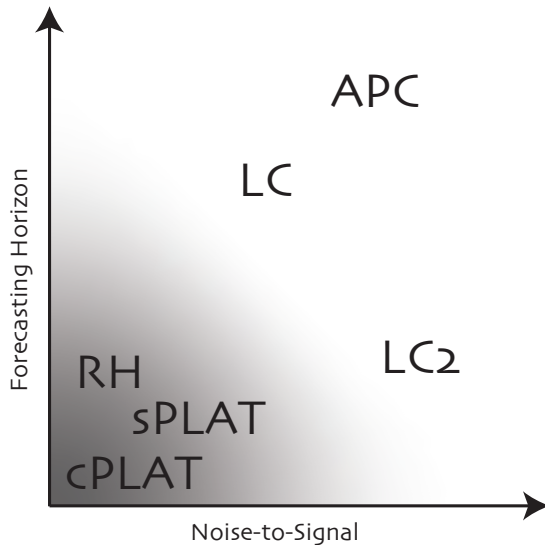
Data	1 Year Horizon	Parameters
USA (F)	$\alpha_x + \kappa_t^1 + (x - \bar{x})\kappa_t^2 + \gamma_c$	262
AUS (M)	$\alpha_x + \kappa_t^1 + (x - \bar{x})^2\kappa_t^2 + (x - 25)^+\kappa_t^3 + (25 - x)^+\kappa_t^4 + \gamma_c$	344
UK (M)	$\alpha_x + \kappa_t^1 + (x - \bar{x})\kappa_t^2 + (x - \bar{x})^2\kappa_t^3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

# Evaluation Results: Rankings for 1 Year vs 10 Year Forecasting

	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

# Key Findings



# 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 populations do not support strong cohort effects.

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



## 2. Mortality model combination using stacked regression ensembles

# Model Combination: Motivation

- ▶ Forecasting model **selection** and **combination** are the two competing approaches in the forecasting literature.
- ▶ Criteria to select a model
  - ▶ Coefficient of determination
  - ▶ Akaike Information Criterion (AIC)
  - ▶ Bayesian Information Criterion (BIC)
  - ▶ Cross-validation technique.
- ▶ These selection methods can lead to different model choices
- ▶ Difficult to identify one forecasting method that performs consistently well over all time horizons and data sets.

# Model Combination: Motivation

- ▶ Model combination could be a successful **alternative to model selection**
- ▶ Forecasts from a particular forecasting method can provide some useful information that is not contained in forecasts from other forecasting techniques.
- ▶ Combining models improves the forecasting accuracy and **reduce model uncertainty** as the modeler is not forced to select one forecasting model.
- ▶ Forecasting combination is more apparent when it is **hard to identify the best model** or the true data generating process.
- ▶ Multiple forecasting techniques usually generate varying forecasts.

# Model Combination: Motivation in a mortality context

- ▶ Comparing different mortality models does not always lead to the selection of the one best mortality model (Janssen 2018).
- ▶ It is difficult to identify the best mortality model
  - ▶ Different **fitting periods** lead to different mortality forecasts (Janssen and Kunst 2007).
  - ▶ Modelling different **life table statistics** give different mortality forecasts (Bergeron-Boucher and Kj 2019).
  - ▶ No model dominates both qualitative and quantitative criteria (Cairns et al. 2007).
  - ▶ No single mortality model performs the best in all the **countries and time horizons** (Rabbi and Mazzuco 2018).

# Model Combination: Recent advances in a mortality context

- ▶ Shang (2012) applied the Bayesian Model Averaging (BMA) to combine 14 mortality forecasting models.
- ▶ Kontis et al. (2017) applied a probabilistic BMA to combine 21 mortality models to project age-specific death rates for 35 industrialized countries.
- ▶ Shang and Haberman (2018) combined multiple mortality models using the model confidence sets technique.

# Stacking Ensemble Techniques

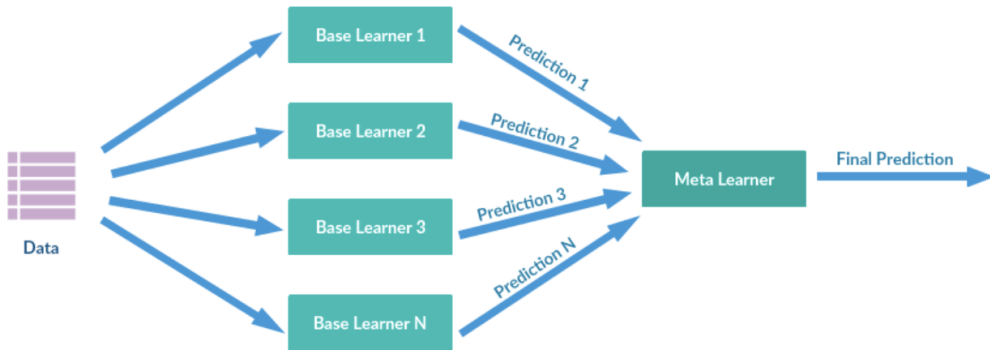
- ▶ The original idea of the stacking ensemble was first introduced by (Wolpert D. 1992).
- ▶ The stacking ensemble has been successfully applied and improved the predictive accuracy on a wide range of problems:
  1. Forecasting global energy consumption (Khairalla et al. 2018).
  2. Credit risk assessment (Doumpos and Zopounidis 2007).
  3. Planning and budgeting (Hansen and Nelson 2002).
  4. Financial time series datasets (Ma and Dai 2016).
  5. Forecasting Warfarin dosing (Ma et al. 2018).
  6. Prediction of infectious disease epidemics (Ray and Reich 2018).
- ▶ The stacking ensemble has been used by most winning teams in the data science competitions (Sill et al. 2009; Puurula, Read, and Bifet 2014).

# Ensemble Techniques



- ▶ Ensemble methods use **multiple learning algorithms** to obtain **better predictive performance** than could be obtained from any of the **constituent learning algorithms**.
- ▶ Bagging and Boosting
- ▶ **Stacked regression ensemble** combines the point predictions from **diverse and accurate predictors** (Wolpert D. 1992).
- ▶ If a set of **base learners** does not contain the **true prediction function**, ensembles can give a good approximation of that function.
- ▶ Ensembles can perform better than individual base algorithms.

# Stacked Regression Ensemble





# Stacked Regression Ensemble

- ▶ Suppose that stacked regression estimate  $M$  base learners  $L_1, \dots, L_M$  which produces the mortality forecasts  $\hat{f}_1(x), \dots, \hat{f}_M(x)$ .
- ▶ Under the stacked regression setting, the mortality forecasts can be combined by:

$$\hat{f}_{\text{stacking}}(x) = \sum_{m=1}^M w_m \hat{f}_m(x).$$

- ▶ Our goal is to compute  $w_1, \dots, w_M$  by minimizing the generalization error of stacked regression ensemble:

$$\min_{w_1 \dots w_M} \sum_{i=1}^N \left( y_i - \sum_{m=1}^M w_m \hat{f}_m(x_i) \right)^2.$$

# Creating the Diversity Among the Base Learners

- ▶ Using different base learners.
- ▶ Extrapolative mortality models based on different life table statistics: death rates, death probabilities, survival probabilities, life table deaths and life expectancy at birth (Bergeron-Boucher and Kj 2019).
- ▶ Fitting same mortality model to different historical fitting periods (Janssen and Kunst 2007).

# Training the Base Learners using cross-validation

- ▶ Split the dataset  $\mathcal{D} = \{(y_n, x_n), n = 1, \dots, N\}$  into  $K$  folds.
- ▶ Train each base learners  $L_1, \dots, L_M$  on the training data set  $\mathcal{D}^K = \mathcal{D} - \mathcal{D}_K$  for each  $K^{\text{th}}$  fold.
- ▶ For each base learners  $L_1, \dots, L_M$ , predict the outcomes of the target variable  $y$  using the validation set  $\mathcal{D}_K$  in each fold based on the corresponding  $\mathcal{D}^K = \mathcal{D} - \mathcal{D}_K$  training set.
- ▶ Level-one data/metadata is given by these cross-validation predictions.

# Integration of Base Learners

- ▶ A lasso regression is used to learn the weights for combining the output of base learners while minimizing the generalization error:

$$\mathbf{w} = \min_{w_1 \dots w_M} \sum_{t=t_1}^{t_n} \sum_{x=x_1}^{x_n} \left( \ln(\mu_{x,t}) - \sum_{m=1}^M w_m \widehat{\ln(\mu_{x,t})}_m \right)^2 + \lambda \sum_{m=1}^M |w_m|.$$

- ▶ The final super learner algorithm is given by:

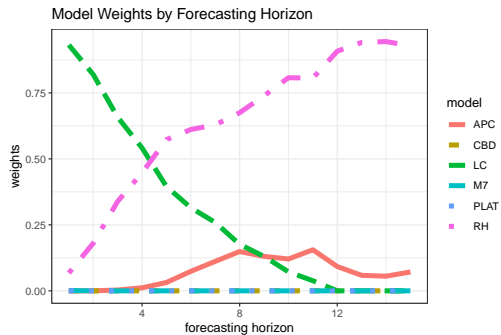
$$\hat{L}_{\text{Stacking}}(\mathcal{D}_{\text{test}}) = \sum_{m=1}^M \hat{w}_m \hat{L}_m(\mathcal{D}_{\text{test}}).$$

# Very Preliminary Results

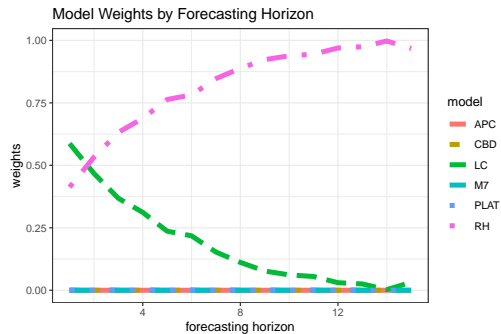
- ▶ Use US and England and Wales male mortality data from HMD.
- ▶ We train the base learners using the mortality data from 1960 to 1990 for ages 50-89 using **StMoMo** package.
- ▶ We fitted LC, RH, APC, CBD, Plat, and M7.
- ▶ The fitted models were used to predict the 1 to 15 steps ahead mortality rates.
- ▶ Lasso regression is used as the meta learner to learn the optimal weights for combining the predictions from the multiple base learners.
- ▶ Evaluation done using data for 1991-2016

# Model weights by horizon

## USA

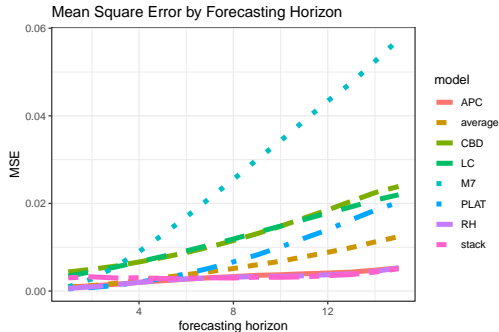


## England and Wales

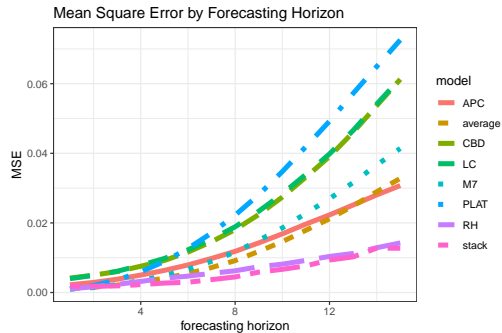


# Mean Square Error by horizon

## USA



## England and Wales



$$\text{MSE} = \frac{1}{n_a(n_y + 1)} \sum_{t=t_1}^{t_n} \sum_{x=x_1}^{x_n} \left( \widehat{\ln(\mu_{x,t})} - \ln(\mu_{x,t}) \right)^2.$$

# Summary and future work

- ▶ The stacked regression ensemble outperformed other models and simple averaging at least for longer horizons
  - ▶ We need to compare with BMA and Model Confidence set for model combination
- ▶ Without even considering more base models than Generalized APC models, it seems it is still a good idea to combine the GAPC models as they seem to give better mortality predictions.
- ▶ The optimal weights for combining the base models vary depending on the forecasting horizon, and therefore, the models that form the final ensemble prediction changes depending on the forecasting horizons.
  - ▶ Develop weights that vary with forecasting horizons.



## Conclusion and outlook

# Conclusion outlook

- ▶ There is a thriving literature on the application of statistical/machine learning techniques to mortality modelling
  - ▶ Use of new techniques
  - ▶ Data analytic mindset to approach more classical techniques
- ▶ We have shown two examples with very promising results
  - ▶ Techniques will be made available in open-source code via **StMoMo**
- ▶ Important to have the problem in mind and adapt techniques rather than applying off-the-shelf algorithms

# Conclusion outlook

- ▶ Space for application into more complex problems beyond modelling single populations and all-cause mortality
  - ▶ Multipopulation modelling
  - ▶ Cause of death modelling
- ▶ Application of statistical/machine learning techniques to individual level data
  - ▶ Quantify mortality heterogeneity
  - ▶ Impact of socio-economic factors
  - ▶ Impact of risk factors

# Questions?

---

Email: [a.villegas@unsw.edu.au](mailto:a.villegas@unsw.edu.au)

---

# Reserve slides

# Implementation in StMoMo

```
# install.packages("devtools")
# devtools::install_github("amvillegas/StMoMo", ref = "GroupLasso")
library('demography')
library('StMoMo')
DNKdata <- hmd.mx(country = "DNK", username = "a.villegas@unsw.edu.au",
                  password = password, label = "Denmark")
DNKStMoMo <- StMoMoData(DNKdata, series = "male")
DNKStMoMo
```

```
## Mortality data for Denmark
##      Series:  male
##      Years:  1835 - 2016
##      Ages:   0 - 110
##      Exposure:  central
```

# Implementation in StMoMo - Define model

```
#Create big model
strikes <- seq(25,85,5)
bModel <- StMoMo(link = "log-Gaussian", staticAgeFun = TRUE,
                 periodAgeFun = c("1", genPoly(1:5), genCall(strikes),
                                   genPut(strikes)),
                 cohortAgeFun = "1")
bModel
```

```
## Gaussian model with predictor:  $\log m[x,t] = a[x] + k1[t] +$   
 $f2[x] k2[t] + f3[x] k3[t] + f4[x] k4[t] + f5[x] k5[t] + f6[x]$   
 $k6[t] + f7[x] k7[t] + f8[x] k8[t] + f9[x] k9[t] + f10[x] k10[t] +$   
 $f11[x] k11[t] + f12[x] k12[t] + f13[x] k13[t] + f14[x] k14[t] +$   
 $f15[x] k15[t] + f16[x] k16[t] + f17[x] k17[t] + f18[x] k18[t] +$   
 $f19[x] k19[t] + f20[x] k20[t] + f21[x] k21[t] + f22[x] k22[t] +$   
 $f23[x] k23[t] + f24[x] k24[t] + f25[x] k25[t] + f26[x] k26[t] +$   
 $f27[x] k27[t] + f28[x] k28[t] + f29[x] k29[t] + f30[x] k30[t] +$   
 $f31[x] k31[t] + f32[x] k32[t] + g[t-x]$ 
```

# Implementation in StMoMo - Fit model with grouped penalised regularisation

```
#Penalise all term
```

```
bMgrpfit <- grpfit(bModel, nlambdas = 50, data = DNKStMoMo, ages.fit = 20:89,  
                  years.fit = 1960:2016)
```

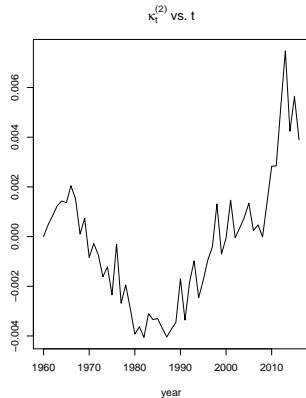
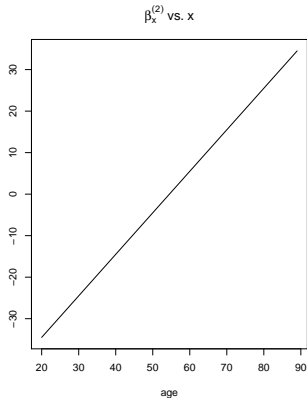
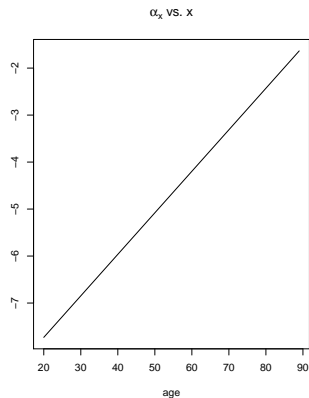
```
## StMoMo: Start fitting with grpreg
```

```
## StMoMo: Finish fitting with grpreg
```



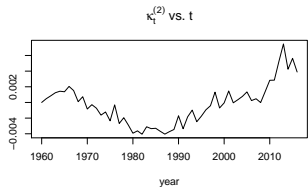
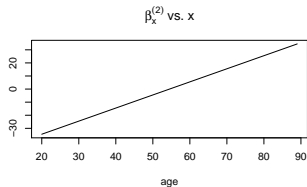
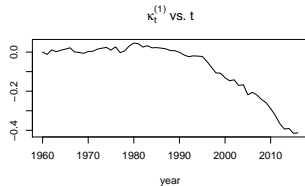
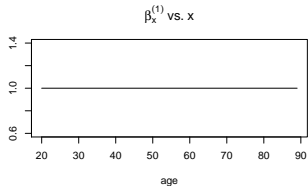
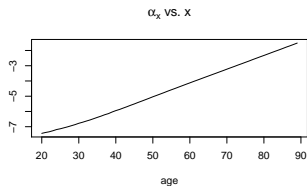
# Implementation in StMoMo - Parameter paths

```
plot(extractStMoMo(bMgrpfit, 10), nCol = 3)
```



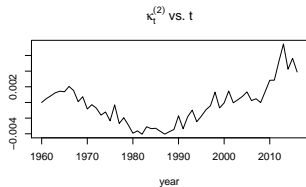
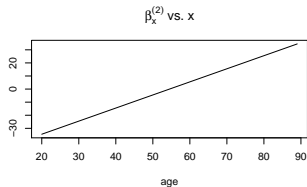
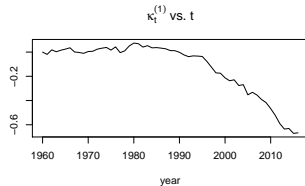
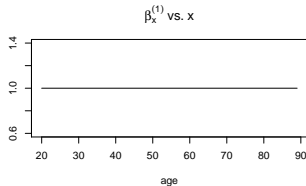
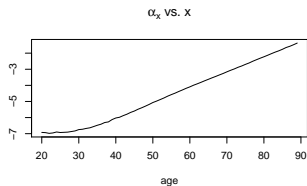
# Implementation in StMoMo - Parameter paths

```
plot(extractStMoMo(bMgrpfit, 15), nCol = 3)
```



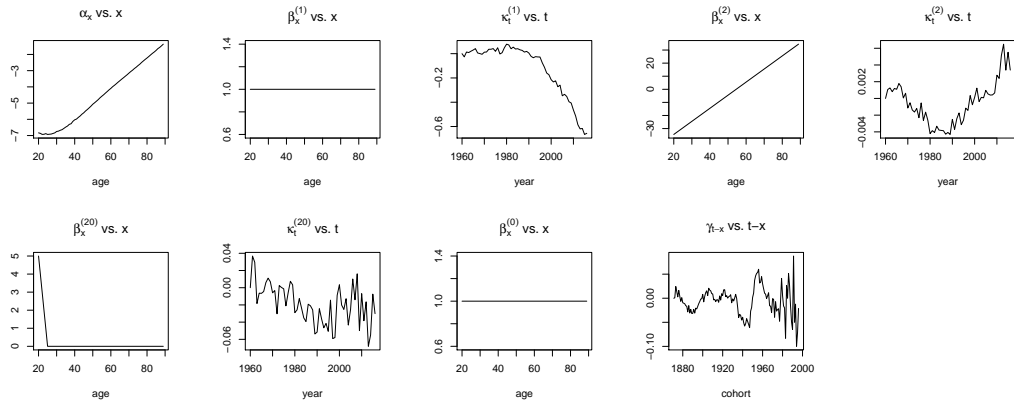
# Implementation in StMoMo - Parameter paths

```
plot(extractStMoMo(bMgrpfit, 20), nCol = 3)
```



# Implementation in StMoMo - Parameter paths

```
plot(extractStMoMo(bMgrpfit, 25), nCol = 5)
```



```
## StMoMo: Start fitting with grpreg  
## StMoMo: Finish fitting with grpreg
```

## Construction: Estimated Parameters (Denmark Males)

## Construction: Fitted Values (Denmark Males)

## Construction: Estimated Parameters (Denmark Females)

## Construction: Fitted Values (Denmark Females)



# Construction: Estimated Parameters (UK Males)

## Construction: Fitted Values (UK Males)

# References I

Bergeron-Boucher, Marie-Pier, and Søren Kj. 2019. "What to Forecast? An Exploratory Study of the Implication of Using Different Indicators to Forecast Mortality," 15.

Blake, David P., Andrew J. G. Cairns, Kevin Dowd, and Amy R. Kessler. 2018. "Still Living with Mortality: The Longevity Risk Transfer Market After One Decade." SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3271283>.

Booth, Heather, and Leonie Tickle. 2008. "Mortality modelling and forecasting: A review of methods." Annals of Actuarial Science 1 (2): 3–43.

Breheny, Patrick, and Jian Huang. 2013. "Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors." Statistics and Computing 25 (2): 173–87. <https://doi.org/10.1007/s11222-013-9424-2>.

Cairns, Andrew J. G., David Blake, Kevin Dowd, Guy D. Coughlan, David Epstein, and Marwa Khalaf-Allah. 2011. "Mortality density forecasts: An analysis of six stochastic mortality models." Insurance: Mathematics and Economics 48 (3): 355–67.

## References II

Cairns, Andrew J. G., David Blake, Kevin Dowd, Guy D. Coughlan, D. Epstein, A. Ong, and I. Balevich. 2009. "A quantitative comparison of stochastic mortality models using data from England and Wales and the United States." North American Actuarial Journal 13 (1): 1–35.

Cairns, Andrew J. G., David P. Blake, Kevin Dowd, Guy Coughlan, and David Epstein. 2007. "A Quantitative Comparison of Stochastic Mortality Models Using Data from England & Wales and the United States." SSRN Electronic Journal.  
<https://doi.org/10.2139/ssrn.1340389>.

Carracedo, Patricia, Ana Debón, Adina Iftimi, and Francisco Montes. 2018. "Detecting Spatio-Temporal Mortality Clusters of European Countries by Sex and Age." International Journal for Equity in Health 17 (1): 38. <https://doi.org/10.1186/s12939-018-0750-z>.

Currie, Iain D. 2016. "On fitting generalized linear and non-linear models of mortality." Scandinavian Actuarial Journal, no. 4: 356–83.

## References III

Deprez, Philippe, Pavel V. Shevchenko, and Mario V. Wüthrich. 2017. "Machine Learning Techniques for Mortality Modeling." European Actuarial Journal 7 (2): 337–52.

<https://doi.org/10.1007/s13385-017-0152-4>.

Doumpos, Michael, and Constantin Zopounidis. 2007. "Model Combination for Credit Risk Assessment: A Stacked Generalization Approach." Annals of Operations Research 151 (1): 289–306. <https://doi.org/10.1007/s10479-006-0120-x>.

Dowd, Kevin, Andrew J. G. Cairns, David Blake, Guy D. Coughlan, David Epstein, and Marwa Khalaf-Allah. 2010a. "Backtesting stochastic mortality models: An ex-post evaluation of multi-period-ahead density forecasts." North American Actuarial Journal 14 (3): 281–98.

———. 2010b. "Evaluating the goodness of fit of stochastic mortality models." Insurance: Mathematics and Economics 47 (3): 255–65.

<https://doi.org/10.1016/j.insmatheco.2010.06.006>.

## References IV

- Hainaut, Donatien. 2018. "A Neural-Network Analyzer for Mortality Forecast." ASTIN Bulletin 48 (2): 481–508. <https://doi.org/10.1017/asb.2017.45>.
- Hansen, James V, and Ray D Nelson. 2002. "Data Mining of Time Series Using Stacked Generalizers." Neurocomputing 43 (1-4): 173–84. [https://doi.org/10.1016/S0925-2312\(00\)00364-7](https://doi.org/10.1016/S0925-2312(00)00364-7).
- Hatzopoulos, P., and S. Haberman. 2011. "A Dynamic Parameterization Modeling for the Age–Period–Cohort Mortality." Insurance: Mathematics and Economics 49 (2): 155–74. <https://doi.org/10.1016/j.insmatheco.2011.02.007>.
- Hyndman, Rob J., and Md. Shahid Ullah. 2007. "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach." Computational Statistics & Data Analysis 51 (10): 4942–56. <https://doi.org/10.1016/j.csda.2006.07.028>.
- Janssen, Fanny. 2018. "Advances in Mortality Forecasting: Introduction." Genus 74 (1). <https://doi.org/10.1186/s41118-018-0045-7>.

## References V

Janssen, Fanny, and Anton Kunst. 2007. "The Choice Among Past Trends as a Basis for the Prediction of Future Trends in Old-Age Mortality." Population Studies 61 (3): 315–26. <https://doi.org/10.1080/00324720701571632>.

Khairalla, Mergani A, Xu Ning, Nashat T AL-Jallad, and Musaab O El-Faroug. 2018. "Short-Term Forecasting for Energy Consumption Through Stacking Heterogeneous Ensemble Learning Model." Energies 11 (6). <https://doi.org/10.3390/en11061605>.

Kontis, Vasilis, James E. Bennett, Colin D. Mathers, Guangquan Li, Kyle Foreman, and Majid Ezzati. 2017. "Future Life Expectancy in 35 Industrialised Countries: Projections with a Bayesian Model Ensemble." The Lancet 389 (10076): 1323–35. [https://doi.org/10.1016/S0140-6736\(16\)32381-9](https://doi.org/10.1016/S0140-6736(16)32381-9).

Levantesi, Susanna, and Virginia Pizzorusso. 2019. "Application of Machine Learning to Mortality Modeling and Forecasting." Risks 7 (1): 26–26. <https://doi.org/10.3390/risks7010026>.

## References VI

- Ludkovski, Michael, James Risk, and Howard Zail. 2016. "Gaussian Process Models for Mortality Rates and Improvement Factors." Ssrn. <https://doi.org/10.2139/ssrn.2831831>.
- Ma, Zhiyuan, Ping Wang, Zehui Gao, Ruobing Wang, and Koroush Khalighi. 2018. "Ensemble of Machine Learning Algorithms Using the Stacked Generalization Approach to Estimate the Warfarin Dose." PLoS ONE 13 (10): 1–12. <https://doi.org/10.1371/journal.pone.0205872>.
- Ma, Zhongchen, and Qun Dai. 2016. "Selected an Stacking ELMs for Time Series Prediction." Neural Processing Letters 44 (3): 831–56. <https://doi.org/10.1007/s11063-016-9499-9>.
- Nigri, Andrea, Susanna Levantesi, Mario Marino, Salvatore Scognamiglio, and Francesca Perla. 2019. "A Deep Learning Integrated Lee–Carter Model." Risks 7 (1): 1–16. <https://doi.org/10.3390/risks7010033>.
- Piscopo, Gabriella, and Marina Resta. 2017. "Applying Spectral Biclustering to Mortality Data." Risks 5 (2): 24–24. <https://doi.org/10.3390/risks5020024>.



## References VII

Pitacco, Ermanno. 2004. "Survival models in a dynamic context: a survey." Insurance: Mathematics and Economics 35 (April): 279–98.  
<https://doi.org/10.1016/j.insmatheco.2004.04.001>.

Pitacco, Ermanno, Michel Denuit, Steven Haberman, and Annamaria Olivieri. 2009. Modelling longevity dynamics for pensions and annuity business. Oxford: Oxford University Press.

Puurula, Antti, Jesse Read, and Albert Bifet. 2014. "Kaggle LSHTC4 Winning Solution," May. <http://arxiv.org/abs/1405.0546>.

Rabbi, Ahabab Mohammad Fazle, and Stefano Mazzucco. 2018. "Mortality and Life Expectancy Forecast for (Comparatively) High Mortality Countries." Genus 74 (1): 18.  
<https://doi.org/10.1186/s41118-018-0042-x>.

Ray, Evan L., and Nicholas G. Reich. 2018. "Prediction of Infectious Disease Epidemics via Weighted Density Ensembles." PLOS Computational Biology 14 (2): e1005910.  
<https://doi.org/10.1371/journal.pcbi.1005910>.

## References VIII

Richman, Ronald, and Mario V. Wuthrich. 2018. "A Neural Network Extension of the Lee-Carter Model to Multiple Populations." SSRN Electronic Journal, 1–21.  
<https://doi.org/10.2139/ssrn.3270877>.

Shang, Han Lin. 2012. "Point and Interval Forecasts of Age-Specific Life Expectancies: A Model Averaging Approach." Demographic Research 27: 593–644.  
<https://doi.org/10.4054/DemRes.2012.27.21>.

Shang, Han Lin, and Steven Haberman. 2018. "Model Confidence Sets and Forecast Combination: An Application to Age-Specific Mortality." Genus 74 (1).  
<https://doi.org/10.1186/s41118-018-0043-9>.

Sill, Joseph, Gabor Takacs, Lester Mackey, and David Lin. 2009. "Feature-Weighted Linear Stacking," November. <http://arxiv.org/abs/0911.0460>.

Wiśniowski, Arkadiusz, Peter W. F. Smith, Jakub Bijak, James Raymer, and Jonathan J. Forster. 2015. "Bayesian Population Forecasting: Extending the Lee-Carter Method." Demography 52 (3): 1035–59. <https://doi.org/10.1007/s13524-015-0389-y>.

# References IX

Wolpert D., H. 1992. "Stacked Generalization." Neural Networks 5: 241–59.

Wong-Fupuy, C., and Steven Haberman. 2004. "Projecting mortality trends: recent developments in the United Kingdom and the United States." North American Actuarial Journal 8 (2): 56–83.