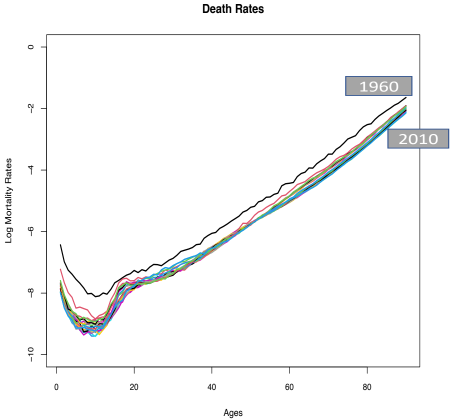
Averaging Mortality Rate Forecasts Across Fitting Periods

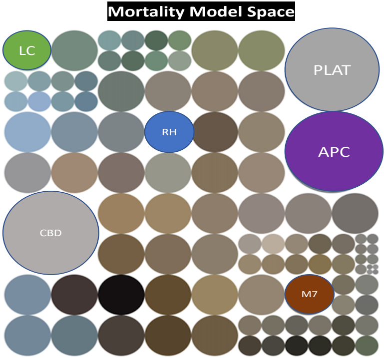
Salvatory Roman Kessy
Joint Work With: M. Sherris, A. Villegas, & J. Ziveyi.

UNIVERSITY OF NEW SOUTH WALES

AFRIC Conference, Victoria Falls

Model Selection Dilemma

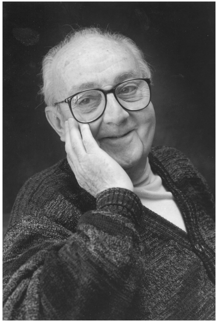




- What mortality model is likely to perform best?

Model Combination: Motivation

- George Edward Box is one of the great statistical minds of the 20th century.
- George Edward Box said **all models are wrong, but some are useful.**



Model combination is an alternative to model selection

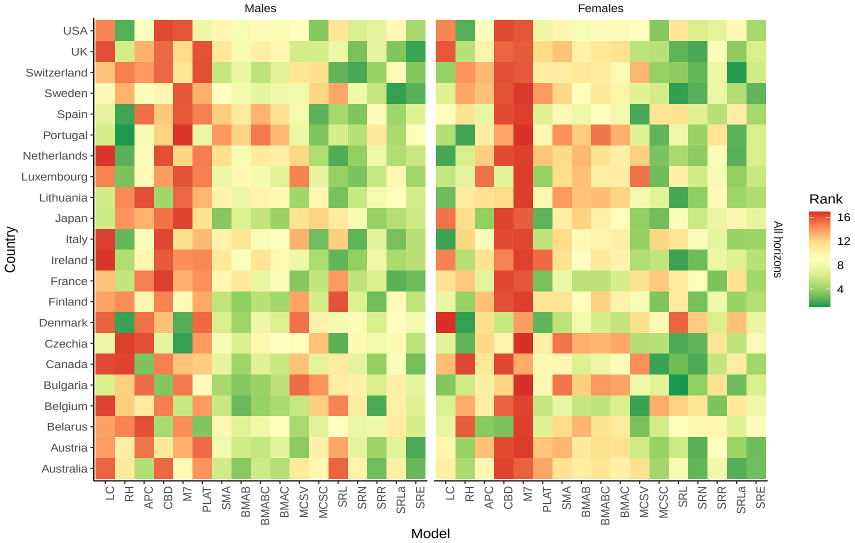
- Bayesian Model Averaging (Barigou et al. 2021).
- Model Confidence Set (Shang and Haberman 2018).
- Stacked regression ensemble (Wolpert 1992).

Model combination formulation:

$$\ln\left(\widehat{\mu}(x, t+h)\right)_{\text{comb}} = \sum_{m=1}^{M} w_m \ln\left(\widehat{\mu}_m(x, t+h)\right).$$

Motivation: Mortality Model Combination

Average ranks of mortality models across different countries



CoMoMo Package

- Use `CoMoMo` to estimate combination weights using stacked regression ensemble, Bayesian Model Averaging (Kessy et al. 2022).
- Acess our paper

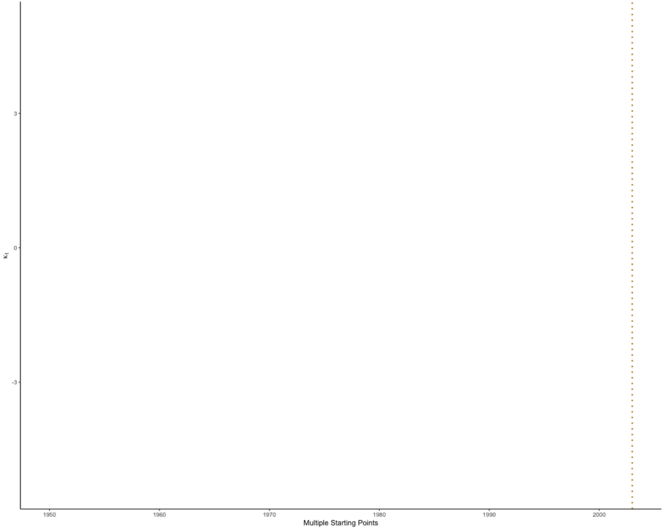
Averaging Mortality Rate Forecasts Across Fitting Periods

Fitting Period Selection Risk

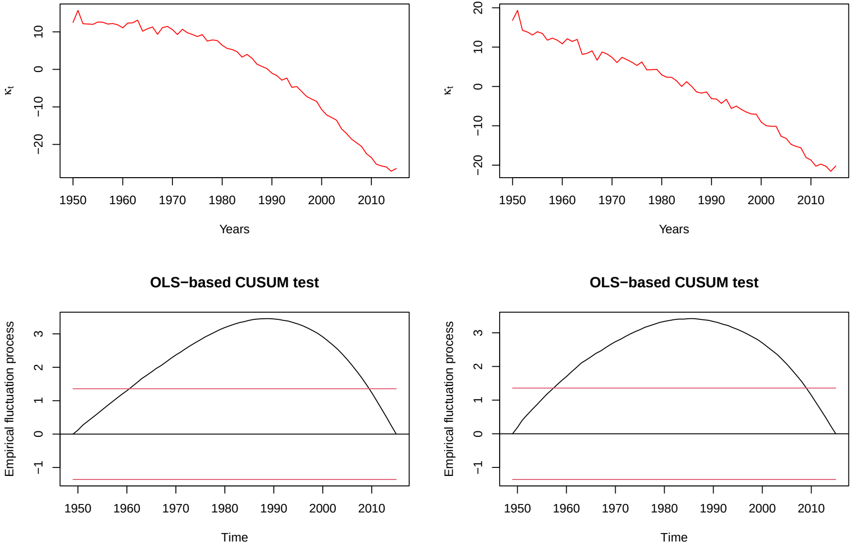
- The accuracy of mortality rate forecasts depends on the length of the fitting period used for estimating the model parameters (Booth, Maindonald, and Smith 2002).

- Choosing the wrong fitting period leads to bad decision-making.
- For example, when we underestimate the life expectancy by one year, the pension liabilities can increase by up to 5% (Bergeron-Boucher and Kjærgaard 2021).

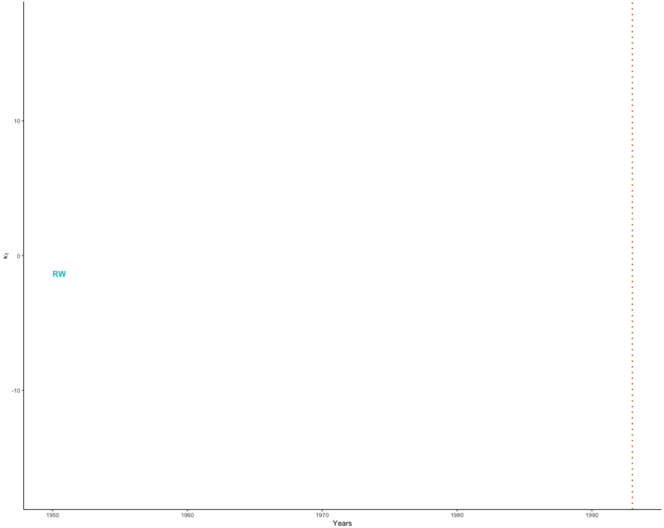
Fitting Period Selection Dilemma



Structural Changes in Mortality Patterns



Existing Methods of Selecting Optimal Fitting Period



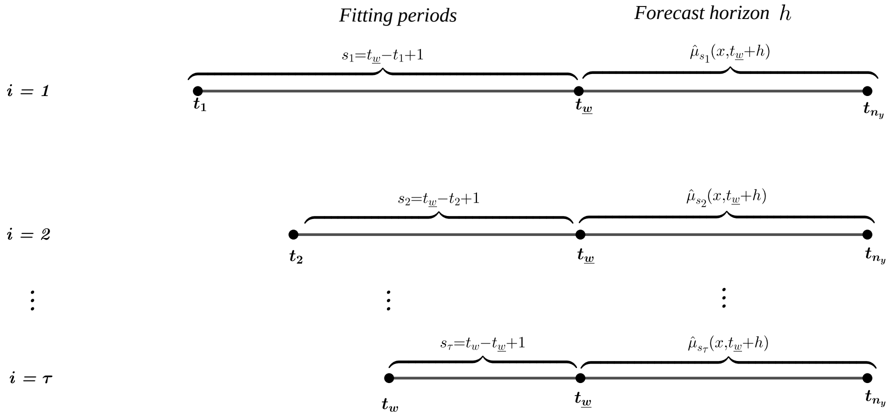
We Ask

- Do the linear mortality models generate statistically different out-of-sample mean squared errors based on multiple fitting periods selected using linear regularization methods?
- Do the mortality rate forecasts averaged from the same linear mortality model calibrated over different fitting periods selected using linear regularization methods have lower out-of-sample mean squared errors than those based on the "longest" calibration period?

Averaging Across Fitting Periods

- Presence of structural breaks and mortality time index is non-linear.

Combine Forecasts Across Fitting Periods



Averaging Across Fitting Periods

- Fit a mortality model using the fitting periods $\{t_1 : t_{\underline{w}}, t_2 : t_{\underline{w}}, \ldots, t_\omega : t_{\underline{w}}\}$.
- Define the sample size s_i as $s_i = t_{\underline{w}} - t_i + 1, \forall i = 1, \ldots, \tau$ and $\tau = \left|\{t_1 : t_{\underline{w}}, t_2 : t_{\underline{w}}, \ldots, t_\omega : t_{\underline{w}}\}\right|$.
- Use the fitted mortality model to generate the future mortality rates $\widehat{\mu}_{s_1}(x, t_{\underline{w}} + h), \ldots, \widehat{\mu}_{s_\tau}(x, t_{\underline{w}} + h)$ for age $x \in [x_1, x_{n_a}]$ at time $t_{\underline{w}} + h$, where h is the forecasting horizon $\forall h = 1, \ldots, H$.
- Model combination formulation (Pesaran and Timmermann 2007):

$$\ln\left(\widehat{\mu}(x, t_{\underline{w}} + h)\right)_{\text{comb}} = \sum_{i=1}^{\tau} w_{s_i}(h) \ln \widehat{\mu}_{s_i}(x, t_{\underline{w}} + h), \forall h = 1, \ldots, H.$$

Simple Fitting Periods Averaging

- Equal weighting multiple fitting periods is empirically robust to structural breaks of unknown break dates and sizes (Pesaran and Timmermann 2007).
- Equally weighted combined mortality rate forecast $\ln\left(\widehat{\mu}(x, t_{\underline{w}} + h)\right)_{\text{SFPA}}$ for different forecasting horizons h from τ multiple starting points is defined by the simple forecast combination rule

$$\ln\left(\widehat{\mu}(x, t_{\underline{w}} + h)\right)_{\text{SFPA}} = \sum_{i=1}^{\tau} \frac{1}{\tau} \ln \widehat{\mu}_{s_i}(x, t_{\underline{w}} + h), \forall h = 1, \dots, H.$$

Location Based Weights

- Recent mortality data contribute more to forecasting future mortality rates because they reflect the most recent dynamics of mortality evolution (Hyndman and Shang 2009).
- The combination weight should be proportional to the location $i \in \{t_1, \dots, t_{\underline{w}}\}$ such that we assign heavier weights to the out-of-sample forecasts based on the recent mortality data.
- For the forecasting horizon h, $\forall h = 1, \dots, H$,

$$\ln\left(\widehat{\mu}(x, t_{\underline{w}} + h)\right)_{\text{LOC}} = \sum_{i=1}^{\tau} \frac{i}{\sum_{i=1}^{\tau} i} \ln \widehat{\mu}_{s_i}(x, t_{\underline{w}} + h)$$

- Exponentially decaying weights (Hyndman and Shang 2009)

$$\ln\left(\widehat{\mu}(x, t_{\underline{w}} + h)\right)_{\text{EXP}} = \sum_{i=1}^{\tau} \lambda(1-\lambda)^{s_i} \ln \widehat{\mu}_{s_i}(x, t_{\underline{w}} + h), \forall h = 1, \dots, H.$$

Averaging across Windows and Models

- To account for model uncertainty, multiple structural changes, and the impact of the fitting period.
- Let the h-year-ahead mortality rate forecasts from mortality models $m = 1, \ldots, M$ be $\widehat{\mu}_{s_i}^{1}(x, t_{\underline{w}} + h), \ldots, \widehat{\mu}_{s_i}^{M}(x, t_{\underline{w}} + h)$ estimated over the sample window $s_i, \forall i = 1, \ldots, \tau$ for age $x \in [x_1, x_{n_a}]$ at time $t_{\underline{w}} + h$.
- Combine mortality rate forecasts from across mortality models and different fitting periods $\ln\left(\widehat{\mu}(x, t_{n_y} + h)\right)_{\text{AveAve}}$ using

$$\ln\left(\widehat{\mu}(x, t_{\underline{w}} + h)\right)_{\text{AveAve}} = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{\tau} w_{s_i}(h) \ln \widehat{\mu}_{s_i}^{m}(x, t_{\underline{w}} + h), \forall h = 1, \ldots, H$$

Generate Metadata Using Cross-Validation

- Generate cross-validation predictions using the block-cross validation (SriDaran et al. 2021).
- Produce the cross-validation predictions from multiple starting years, $\{t_1 : t_{\underline{w}}, t_2 : t_{\underline{w}}, \dots, t_\omega : t_{\underline{w}}\}$ using a given mortality model.
- Train the mortality model on the window $t_1 : t_{\underline{w}^*}$ to generate cross-validation predictions, $\boldsymbol{z}_{h,s_1}\ \forall h = 1, 2, \dots, H$.
- Train the mortality model on the window $t_2 : t_{\underline{w}^*} + 1$ to generate cross-validation predictions, $\boldsymbol{z}_{h,s_2}\ \forall h = 1, 2, \dots, H$.
- Repeat this procedure until the calibration window is long enough to generate cross-validation predictions for the longest horizon which is $H = 15$.

Generate Metadata Using Cross-Validation

- When $h = 1, \ldots, H$, let $\boldsymbol{z}_{h=h,s_i}$ for $i = 1, 2, \ldots, \tau$ forms the matrix

$$\mathcal{M}_{h=1} = [\boldsymbol{z}_{h=1,s_1}, \boldsymbol{z}_{h=1,s_2}, \ldots, \boldsymbol{z}_{h=1,s_\tau}].$$

- Then for $h = 1, \ldots, H$, stack $\mathcal{M}_{h=1}, \mathcal{M}_{h=2}, \ldots, \mathcal{M}_{h=H}$ in row order on top of each other to form a new matrix $\mathcal{M}^+$ given as

$$\mathcal{M}^+ = \begin{bmatrix} \mathcal{M}_1 \\ \vdots \\ \mathcal{M}_H \end{bmatrix}.$$

- The matrix $\mathcal{M}^+$ along with the observed mortality rates $\boldsymbol{y}$ forms the metadata, and we need to solve the following linear regression

$$\boldsymbol{y} = \mathcal{M}^+ \mathbf{w},$$

  where $\mathbf{w} = [w_1, \ldots, w_\tau]'$ and we estimate $\mathbf{w}$ using regularized linear regression methods.

Metadatda

	$\widehat{\mu}_{1950}$	$\widehat{\mu}_{1951}$	$\widehat{\mu}_{1952}$	$\widehat{\mu}_{1953}$	...	$\widehat{\mu}_{1970}$	Actual
1	-4.91	-4.90	-4.87	-4.75	...	-4.91	-4.93
2	-4.87	-4.91	-4.85	-4.73	...	-4.90	-4.93
3	-4.86	-4.92	-4.85	-4.73	...	-4.90	-4.89
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1197	-1.48	-1.50	-1.49	-1.38	...	-1.48	-1.49
1198	-1.50	-1.46	-1.46	-1.38	...	-1.44	-1.44
1199	-1.50	-1.46	-1.47	-1.36	...	-1.42	-1.52

$$\Rightarrow \boxed{Y = Z\mathbf{w}}$$

Regularization Methods to Estimate Weights

- Use regularization methods such as non-negative least regression to handle correlated forecasts in metatadata (Breiman 2004).
- Based on Equation (20), detetermine the weighting coefficients $\widehat{w_{s_i}} \;\; \forall i = 1, 2, \ldots, \tau$ as the minimizers of

$$\widehat{\mathbf{w}}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{j=1}^{M^*} \left(y_j - \sum_{i=1}^{\tau} w_{s_i} z_{s_i} \right)^2 .$$

- In lasso regression, we estimate the weights as

$$\widehat{\mathbf{w}}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{j=1}^{M^*} \left(y_j - \sum_{i=1}^{\tau} w_{s_i} z_{s_i} \right)^2 + \lambda \sum_{m=1}^{\tau} |w_{s_i}|.$$

- Transform the weights estimated using the regularization methods such that the recent mortality data weigh more than the past data using

$$w_{s_i}^* = w_{s_i} (1 - w_{s_i})^{s_i}, \forall i = 1, \ldots, \tau.$$

Empirical Results

- Generalised Age-Period-Cohort mortality models (Villegas, Kaishev, and Millossovich 2018).

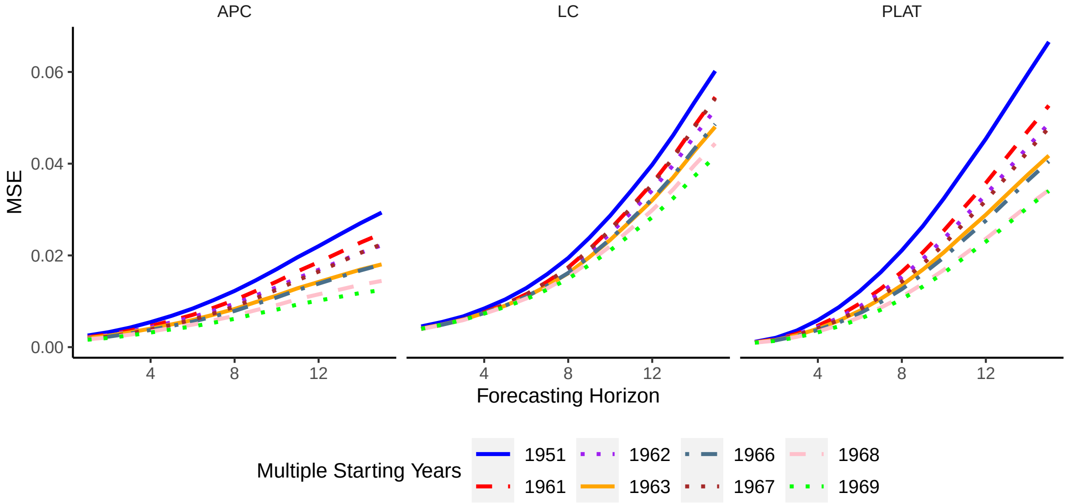
Model	**Predictor** (η_{xt})	**Parameters**
LC	$\alpha_x + \beta_x^{(1)}\kappa_t^{(1)}$	$2n_a + n_y$
APC	$\alpha_x + \kappa_t^{(1)} + \gamma_c$	$n_a + n_y + n_b$
PLAT	$\alpha_x + \kappa_t^{(1)} + (\bar{x} - x)\kappa_t^{(2)} + (\bar{x} - x)^+\kappa_t^{(3)} + \gamma_c$	$n_a + 3n_y + n_b$

Table 1: Generalized age-period-cohort mortality models.

- Human Mortality Database: England and Wales male mortality data (University of California Berkeley and the Max Planck Institute for Demographic Research 2020).
- Training set: 1950 to 1990, Test set: 1991 to 2015, and ages $50 - 89$.

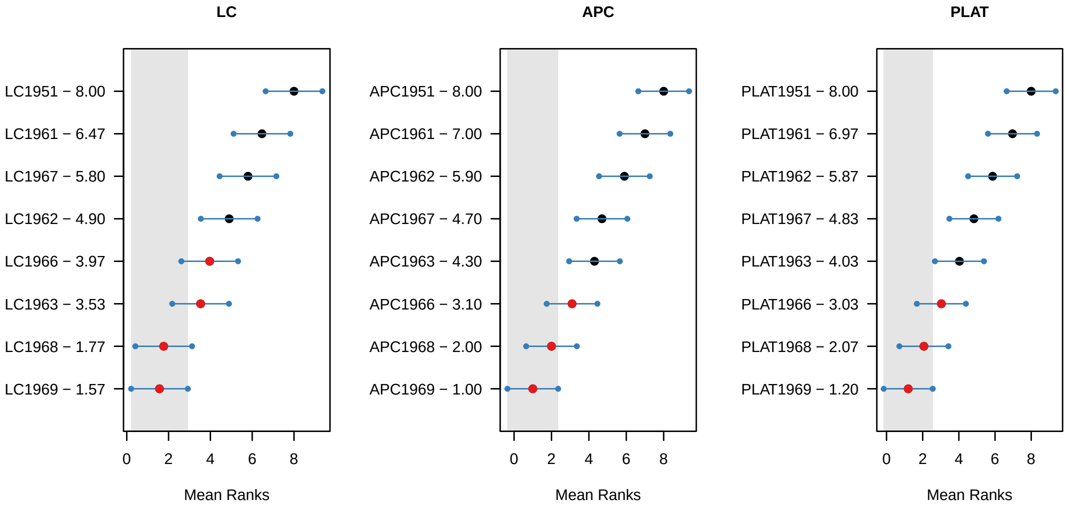
Forecast Instability From Multiple Starting Periods

MSE of the one-step-ahead to 15-step-ahead mortality rate forecasts using different mortality models fitted to multiple fitting periods



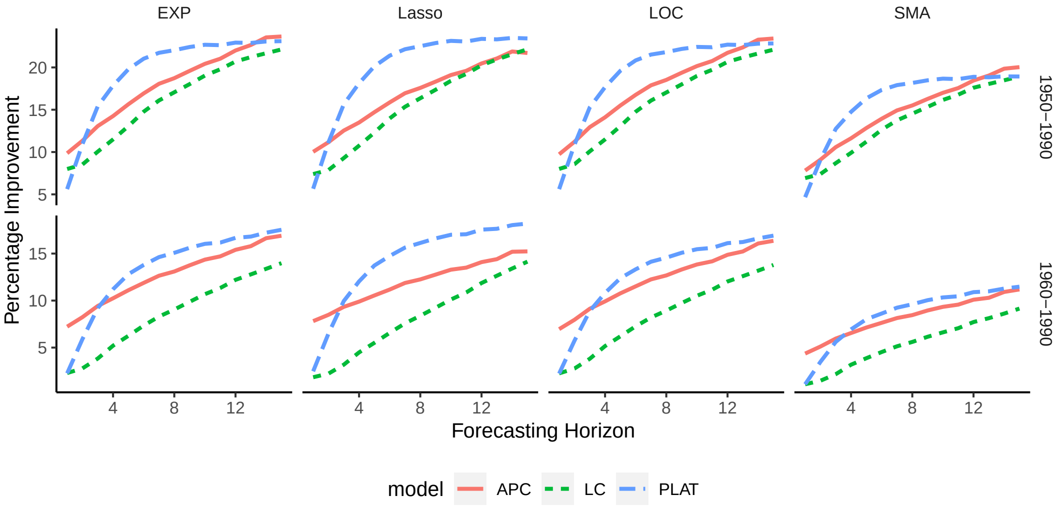
Statistical Test of Forecast Errors

Multiple comparisons with the best style plot for the mortality models



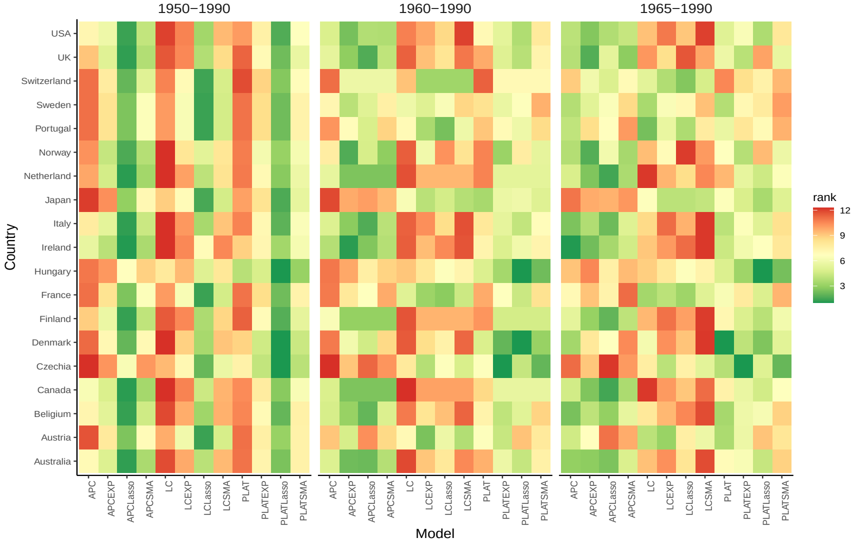
Percentage Improvement to 19 Populations

Percentage improvement relative to different fitting periods



Average Ranks to 19 Populations

Average ranks of mortality models across different countries



Conclusion and Future Research

- Mortality models fitted to different base periods generate out-of-sample mortality rate forecasts that differ statistically significantly.
- Empirical experiments on data from 19 male populations reveal averaging forecasts from different base periods further improves out-of-sample precision in the presence of structural breaks and when the mortality time index is non-linear.
- Future research could include the prediction of the occurrence of possible future structural changes while forecasting the mortality rates.
- Develop model combination approaches that generate prediction intervals from averaging multiple fitting periods.


Q&A

References I

Barigou, Karim, Pierre-Olivier Goffard, Stéphane Loisel, and Yahia Salhi. 2021. "Bayesian Model Averaging for Mortality Forecasting Using Leave-Future-Out Validation." *Https://Hal.archives-Ouvertes.fr/Hal-03175212/Document*, 31.

Bergeron-Boucher, Marie-Pier, and Søren Kjærgaard. 2021. "Mortality Forecasting at Age 65 and Above: An Age-Specific Evaluation of the Lee-Carter Model." *Scandinavian Actuarial Journal*, May, 1–16. https://doi.org/10.1080/03461238.2021.1928542.

Booth, Heather, John Maindonald, and Len Smith. 2002. "Applying Lee-Carter Under Conditions of Variable Mortality Decline." *Population Studies* 56 (3): 325–36.

Breiman, Leo. 2004. "Stacked Regressions." *Machine Learning* 24 (1): 49–64. https://doi.org/10.1007/bf00117832.

References II

Hyndman, Rob, and Han Lin Shang. 2009. "Forecasting Functional Time Series." *Journal of the Korean Statistical Society* 38 (3): 199–211. https://doi.org/10.1016/j.jkss.2009.06.002.

Kessy, Salvatory R, Michael Sherris, Andrés M Villegas, and Jonathan Ziveyi. 2022. "Mortality Forecasting Using Stacked Regression Ensembles." *Scandinavian Actuarial Journal*, November, 591–626. https://doi.org/10.1080/03461238.2021.1999316.

Pesaran, Hashem M., and Allan Timmermann. 2007. "Selection of Estimation Window in the Presence of Breaks." *Journal of Econometrics*, 28.

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.

References III

SriDaran, Dilan, Michael Sherris, Andrés Villegas, and Jonathan Ziveyi. 2021. "A Group Regularisation Approach for Constructing Generalised Age-Period-Cohort Mortality Projection Models." *Forthcoming in ASTIN Bulletin Journal*.

University of California Berkeley and the Max Planck Institute for Demographic Research. 2020. "Human Mortality Database." *Https://Www.mortality.org/*.

Villegas, Andrés M., Vladimir K. Kaishev, and Pietro Millossovich. 2018. "StMoMo: An R Package for Stochastic Mortality Modelling." *Journal of Statistical Software 84 (3).*, no. 1992. https://doi.org/https://doi.org/10.18637/jss.v084.i03.

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