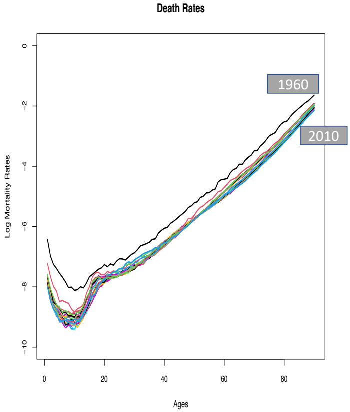
Mortality Forecasting Using Stacked Regression Ensembles

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

UNIVERSITY OF NEW SOUTH WALES

Insurance Data Science Conference: 16 June 2021

Model Selection Dilemma



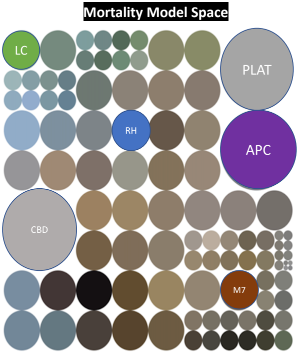


Figure 1: Model Selection Dilemma.

- What mortality model is likely to perform best?

Different Mortality Models

- Multiple mortality models capture different features of death rates such as **trends, linearity, non-linearity, curvature, and cohort effects**.

Model	**Predictor** (η_{xt})	**Parameters**
LC	$\alpha_x + \beta_x^{(1)}\kappa_t^{(1)}$	$2n_a + n_y$
RH	$\alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(0)}\gamma_c$	$3n_a + n_y + n_b$
APC	$\alpha_x + \kappa_t^{(1)} + \gamma_c$	$n_a + n_y + n_b$
CBD	$\kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)}$	$2n_y$
M7	$\kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \left((x - \bar{x})^2 - \hat{\sigma}_x^2\right)\kappa_t^{(3)} + \gamma_c$	$3n_y + n_b$
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 (GAPC) mortality models. Here, year of birth is $c = t - x$, n_a is a number of ages and n_y is a number of years. The functions $\beta_x^{(i)}, \alpha_x, \kappa_t^{(i)}$, and γ_c are age, period and cohort effects respectively with $\bar{x}$ being the mean age over the range of ages being used in the analysis, $\hat{\sigma}_x^2$ is the mean value of $(x - \bar{x})^2$.

- Better methods are needed.

Model Combination

- Simple Model Averaging (Shang 2012), Bayesian Model Averaging (Kontis et al. 2017), Model Confidence Set (Shang and Haberman 2018).

- Model combination formulation:

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

Stacking Ensemble Techniques

- Stacking ensemble **combines point predictions** from multiple models using the weights that **optimise a cross-validation criterion** (Wolpert 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. Financial time series data sets (Ma and Dai 2016).
- Most winning teams in data science competitions have been using the stacked regression ensemble (Sill et al. 2009; Puurula, Read, and Bifet 2014; Makridakis, Spiliotis, and Assimakopoulos 2019).

This Presentation is About . . .

- Propose a new approach of estimating the optimal weights for combining multiple mortality models using **stacked regression ensemble framework** (Wolpert 1992).

  1. Concurrently solve the problem of **model selection and estimation of the model combination** to improve model predictions (Sridhar, Seagrave, and Bartlett 1996).

  2. Tackle the **model list miss-specification limitation** associated with the BMA approach (Yao et al. 2017).

  3. Assigns weights to the individual mortality models by **minimising the cross-validation criterion.**

- Develops the mortality model combination that is **dependent on the forecasting horizon** (SriDaran et al. 2021; Rabbi and Mazzuco 2018).

Stacked Regression Ensemble

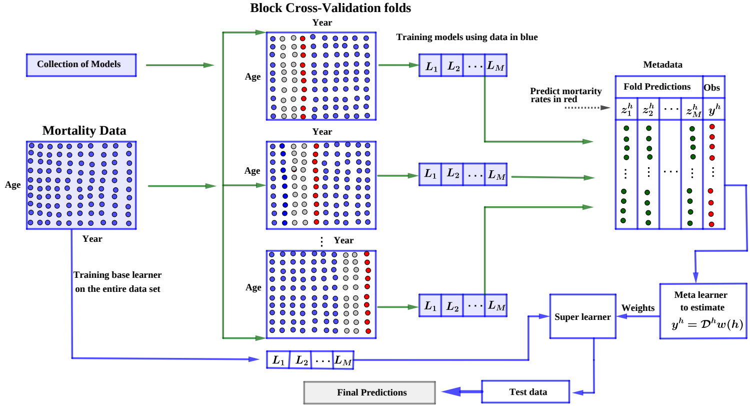


Figure 2: Stacked regression ensemble framework when forecasting three-year ahead mortality rates. The framework can be generalized for predicting mortality rates in any forecast horizon by varying the width of the testing data in red.

Meta-learners

- Non-negative Least Square Regression (Breiman 2004; Naimi and Balzer 2018):

$$\widehat{\mathbf{w}}^*(h) = \underset{\mathbf{w}(h)}{\text{argmin}} \sum_{i=1}^{N} \left(y_i^h - \sum_{m=1}^{M} w_m(h) z_{im}^h \right)^2, \ \widehat{w}_m^*(h) \geq 0.$$

- Ridge Regression (Leblanc et al. 2016):

$$\widehat{\mathbf{w}}^*(h) = \underset{\mathbf{w}(h)}{\text{argmin}} \sum_{i=1}^{N} \left(y_i^h - \sum_{m=1}^{M} w_m(h) z_{im}^h \right)^2 + \lambda \sum_{m=1}^{M} w_m^2(h).$$

- Lasso Regression (Gunes, Wolfinger, and Tan 2017):

$$\widehat{\mathbf{w}}^*(h) = \underset{\mathbf{w}(h)}{\text{argmin}} \sum_{i=1}^{N} \left(y_i^h - \sum_{m=1}^{M} w_m(h) z_{im}^h \right)^2 + \lambda \sum_{m=1}^{M} |w_m(h)|.$$

Combination Weights for Mortality Models

- Human Mortality Database: England and Wales, Males and Females.
- Training set: 1960 to 1990, Test set: 1991 to 2015, and ages 50 – 89.

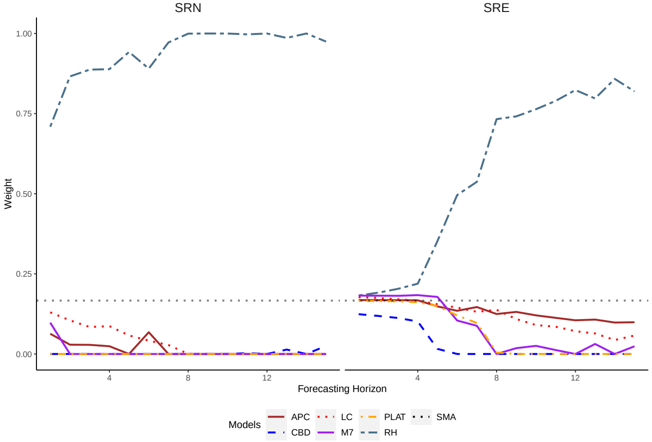


Figure 3: Horizon-specific optimal combining weights learned using different meta-learners for England and Wales males mortality data from 1960 to 1990 and ages 50 to 89.

Performance of Stacked Regression Ensemble

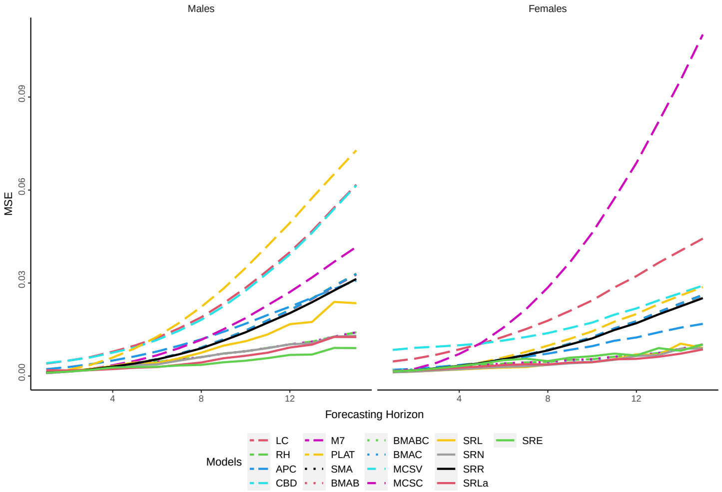


Figure 4: MSEs of the one-step-ahead to 15-step-ahead mortality rate forecasts using different mortality methods and forecast horizons for England and Wales male and female mortality data.

Stacked Regression Ensemble in Different Countries

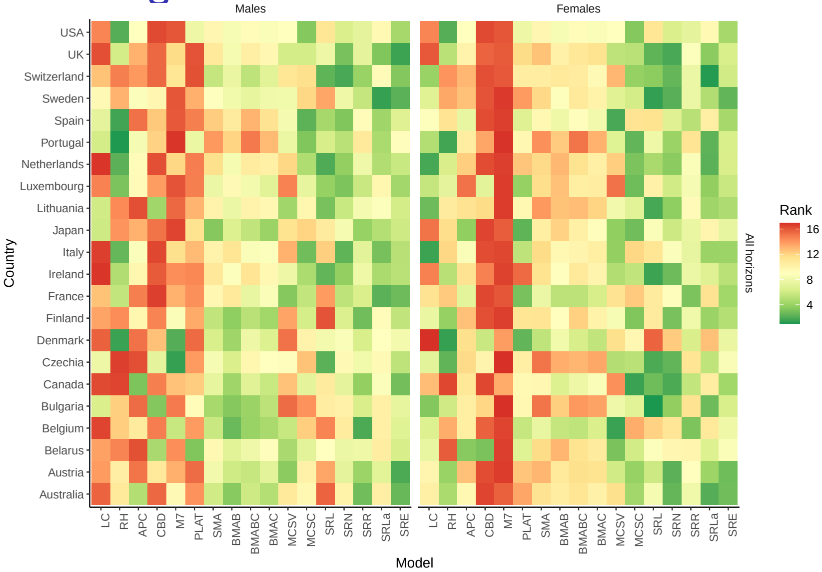


Figure 5: Heat maps showing the average ranks of mortality models across different countries for males and females.

Conclusion

- Using 44 populations from the Human Mortality Database, stacking mortality models increases predictive accuracy.
- Stacked regression (SR) achieved an average accuracy of $13\% - 49\%$ and $20\% - 90\%$ over the individual mortality models for males and females.
- SR also achieved better predictive accuracy than other model combination methods.
- The weights for combining the individual mortality models vary depending on the meta-learner, forecasting horizon, country, and gender.
- Estimating weights or choosing the individual mortality models via cross-validation proves to be a crucial step.
- Our results confirm the superiority of SR over the individual and other model combination methods in forecasting the mortality rates.

CoMoMo R Package

1. Install the CoMoMo package

```
devtools::install_github("amvillegas/StMoMo",
                         ref = "GroupLasso", force = TRUE)
devtools::install_github("kessysalvatory/CoMoMo")
```

2. Download the mortality data

```
library(demography); library(StMoMo)
MorData <- hmd.mx(country = 'GBRTENW', username = username,
                  password = password)
DataStMoMo <- StMoMoData(MorData, "male")
agesFit <- 50:89; yearsFit <- 1960:1990
nAg <- length(agesFit); nYr <- length(yearsFit)
```

3. Define the mortality models

```
LC <- lc(); APC <- apc(); CBD <- cbd(link = "log")
M7 <- m7(link = "log"); PLAT <- plat()
RH <- rh(approxConst = TRUE)
models <- list("LC" = LC, "RH" = RH, "APC" = APC,
               "CBD" = CBD, "M7" = M7, "PLAT" = PLAT)
```

CoMoMo R Package

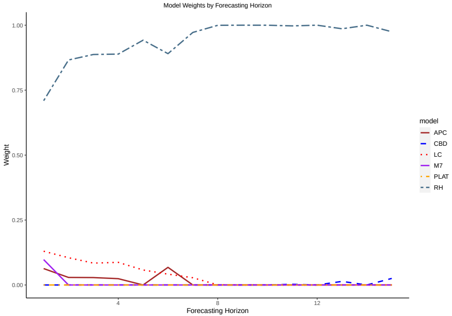
4. Generate the metadata

```
library(CoMoMo)
metaData <- stackMetadata(models, data = DataStMoMo,
            ages.fit = agesFit, years.fit = yearsFit, h = 15)
```

5. Compute the weights

```
stack_nnls_weight <- stack(metaData, metalearner = "nnls")
```

```
plot(stack_nnls_weight)
```



CoMoMo R Package

- Bayesian Model Averaging (BMA)

```
bma_weight_cv <- bma(models, data = DataStMoMo,
                ages.fit = agesFit, years.fit = yearsFit,
                h = 15, method = "cv")
```

- Model Confidence Set (MCS)

```
mcs_weight_cv <- mcs(models, data = DataStMoMo,
                 ages.fit = agesFit, years.fit = yearsFit,
                  h = 15,  method = "cv")
```

6. Fit the mortality models

```
modelFits <- fitCoMoMo(models, data = DataStMoMo,
                ages.fit = agesFit, years.fit = yearsFit)
```

7. Combine the fitted mortality models and combination weights.

```
modcomb <- CoMoMo(modelFits, weight = stack_nnls_weight)
```

CoMoMo R Package

8. Forecast the mortality rates

```
mortalityForecast <- forecast(modcomb, h = 15)
```

```
## # A tibble: 600 x 5
##    ages  year     h model     rate
##   <int> <dbl> <dbl> <chr>    <dbl>
## 1    50  1991     1 comb   0.00483
## 2    50  1992     2 comb   0.00484
## 3    50  1993     3 comb   0.00484
## 4    50  1994     4 comb   0.00483
## 5    50  1995     5 comb   0.00485
## 6    50  1996     6 comb   0.00483
## # ... with 594 more rows
```

Scan to Acess Our Working Paper

References

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

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.

Gunes, Funda, Russ Wolfinger, and Pei-Yi Tan. 2017. "Stacked Ensemble Models for Improved Prediction Accuracy." *Sas*, 1–19.

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.

Leblanc, Michael, Robert Tibshirani, Michael Leblanc, and Robert