Taller 3 - Modelos de mortalidad multi-poblacionales

CvmortalityMult R-package

D. Atance[1]

[1] Universidad de Alcalá, Madrid, Spain.

IV Congreso Usuarios de R 2025, Valencia, October 2025

Index

1. Resources:

1. Villegas, A. M., Kaishev, V. K., & Millossovich, P. (2018). StMoMo: An R package for stochastic mortality modeling. Journal of Statistical Software, 84, 1-38. https://www.jstatsoft.org/article/view/v084i03
2. Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, 191, 192-213. https://www.sciencedirect.com/science/article/pii/S0020025511006773
3. Hyndman, R.J. & Athanasopoulos. G. (2021) Forecasting: Principles and Practice. OTexts, Melbourne, Australia, 3rd edition edition. http://OTexts.org/fpp3/. [p2,6]
4. Atance, D. & Debón, A. (2025) CvmortalityMult: Cross-Validation for Multi-Population Mortality Models. The R Journal, 17(2), 231-258. 10.32614/RJ-2025-018

2. Single mortality models

- In the last two centuries, the developed countries have increased their results in life expectancy and their longevity results (Oeppen and Vaupel, 2002).
- Indeed, during the last 160 years, the world record in female life expectancy at birth has increased at approximately a steady pace of 3 months per year.
- This increase represents, though a sign of social progress, a challenge to governments, private pension plans, and life insurers because of its impact on pension and health costs.
- The accuracy of the prediction of age-specific probabilities of death is **the main objective of the public and private sectors, to forecast accurately the future trend of mortality.**

2. Single mortality models

- Actuaries and demographers have recognized the problems caused by an aging population and rising longevity and have thus devoted significant attention to the development of **statistical techniques for the modeling and projection of mortality rates.**
- The vast majority of the mortality models proposed in the literature are encompassed in the framework of age-period-cohort methodology.
- Among all the developed models, Lee and Carter (1992) is one the most well-known and applied methods in the demographic and actuarial fields.
- The model developed by Lee and Carter (1992) has inspired numerous variants, see for instance Renshaw and Haberman (2003, 2006); Cairns et al. (2006); Nigri et al. (2019); Perla et al. (2021); Richman and Wüthrich (2021).

2. Single mortality models

Hypothesis:

Let $D_{x,t}$ be the number of deaths in a population at age x last birthday during calendar year t. Two alternative hypotheses can be used:

- $D_{x,t} \backsim \textbf{Poisson}\left(E^c_{x,t} \cdot m_{x,t}\right)$, where $E^c_{x,t}$ is the central exposure to risk for individuals aged x during period t and $m_{x,t}$ is the central death rate at age x during period t.
- $D_{x,t} \backsim \textbf{Binomial}\left(E^0_{x,t}; q_{x,t}\right)$, where $E^0_{x,t}$ represents the initial exposure to risk of individuals aged x during year t and $q_{x,t}$ the probability of death of individuals aged x during year t.

Let's recall that one can approximate the $E^0_{x,t}$ from $E^c_{x,t}$, as follows:

$$E^0_{x,t} \approx E^0_{x,t} + (1 - a_{x,t}) \cdot D_{x,t}; \tag{1}$$

where $a_{x,t} \in (0, 1)$ provides the average period of time lived with age x for those who die with age x (last birthday) during year t.

2. Single mortality models

Recall that:

The central death rate at age x during period t, $m_{x,t}$, can be obtained with:

$$m_{x,t} = \frac{D_{x,t}}{E^{c}_{x.t}} = \frac{q_{x,t}}{1-(1-a_{x,t})\cdot q_{x,t}}, \tag{2}$$

The probability of death of individuals aged x during year t, $q_{x,t}$ is equal to:

$$q_{x,t} = \frac{D_{x,t}}{E^{0}_{x.t}} = \frac{m_{x,t}}{1+(1-a_{x,t})\cdot m_{x,t}}, \tag{3}$$

$$\operatorname{logit}(q_{x,t}) = \log\left(\frac{q_{x,t}}{1-q_{x,t}}\right), \tag{4}$$

2. Single mortality models

- **Lee-Carter model** (Lee and Carter, 1992)
  $\Rightarrow \operatorname{logit}(q_{x,t}) = a_x + b_x^1 k_t^1 + \varepsilon_{x,t,i}.$
- **Lee-Carter model** (Renshaw and Haberman, 2003)
  $\Rightarrow \operatorname{logit}(q_{x,t}) = a_x + b_x^1 k_t^1 + b_x^2 k_t^2 + \varepsilon_{x,t,i}.$
- **CBD-models (M5, M6, M7, M8)**; M7
  $\Rightarrow \operatorname{logit}(q_{x,t}) = k_t^1 + (x - \overline{x}) \cdot k_t^2 + k_t^3 \left[(x - \overline{x})^2 - \hat{\sigma}_2\right] + \varepsilon_{x,t,i}.$
- Among others Plat (2009); Haberman and Renshaw (2012); Mitchell et al. (2013).

2. Single mortality models

Practical application in R.
Necessary files:

- 1. SinglePopulationMortalityModels.R
- Spain.RData

3. Multi-population mortality models

- The loss of clear and defined borders between states/countries is leading populations worldwide to experience a similar dynamic of mortality.
- Indeed, mortality improvements or reductions can rapidly spread to other countries, causing correlated mortality dynamics, as observed with the COVID-19 pandemic.
- Thus, multipopulation mortality models provide a valuable approach for considering mortality membership in a group rather than individually (Li and Lee, 2005).

3. Multi-population mortality models

- **Additive model** (Debón et al., 2011)
  $\Rightarrow \text{logit}\left(q_{x,t,i}\right) = a_x + b_x k_t + I_i + \varepsilon_{x,t,i}.$
- **Multiplicative model** (Russolillo et al., 2011)
  $\Rightarrow \text{logit}\left(q_{x,t,i}\right) = a_x + b_x k_t I_i + \varepsilon_{x,t,i}.$
- **Common-factor model** (Lee and Carter, 1992; Li and Lee, 2005)
  $\Rightarrow \text{logit}\left(q_{x,t,i}\right) = a_{x,i} + B_x K_t + \varepsilon_{x,t,i}.$
- **Joint-K model** (Lee and Carter, 1992; Wilmoth and Valkonen, 2001)
  $\Rightarrow \text{logit}\left(q_{x,t,i}\right) = a_{x,i} + b_{x,i} k_t + \varepsilon_{x,t,i}.$
- **Augmented common-factor model** (Li and Lee, 2005; Hyndman et al., 2013) $\Rightarrow \text{logit}\left(q_{x,t,i}\right) = a_x + B_x K_t + b_{x,i} k_{t,i} + \varepsilon_{x,t,i}.$

3. Multi-population mortality models

The multi-population mortality fitting is constructed using the next function:

```
install.packages(CvmortalityMult)
library(CvmortalityMult)
fitLCmulti(model, qxt, periods, ages, nPop, lxt = NULL)
```

- **model**: `c('additive', 'multiplicative', 'CFM', 'joint-K', 'ACFM')`.
- **qxt**: mortality rates for the different populations (matrix or data.frame).
- **periods**: vector with the period for the fitting process.
- **ages**: vector with the ages in the fitting process.
- **nPop**: number of populations to be fitted.
- **lxt**: number of persons alive aged in a life table (weights) and no need to provide it.

3. Multi-population mortality models

The multi-population mortality forecasting is constructed using the next function:

forecast.fitLCmulti(object, nahead, ktmethod = c('arimapdq', 'arima010', 'arimauser'), order = NULL)

- **object**: object created from the fitting process.
- **nahead**: number of periods to forecast.
- **ktmethod**: `c('arimapdq', 'arima010', 'arimauser'))`
- **order**: specifying the fixed components p, d, q for the ARIMA models.

3. Multi-population mortality models

Practical application in R.
Necessary files:

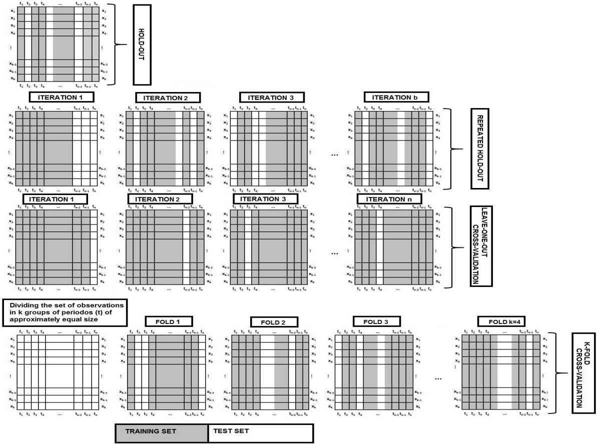
- 2.MultiPopulationMortalityModels.R
- CountryCombined.RData

4. Cross-Validation Technique

- The performance of a model varies between in-sample and out-of-sample evaluations. **Partitioning data into training and test sets, keeping chronological order**, it is fundamental to assessing the forecasting ability of the models.
- **There are Various possibilities for evaluating time series forecasts** which differ based on the forecast horizon and the method of forecasting the out-of-sample validation, "**last block evaluation**" (Tashman, 2000; Bergmeir and Benítez, 2012; Hyndman and Athanasopoulos, 2021).

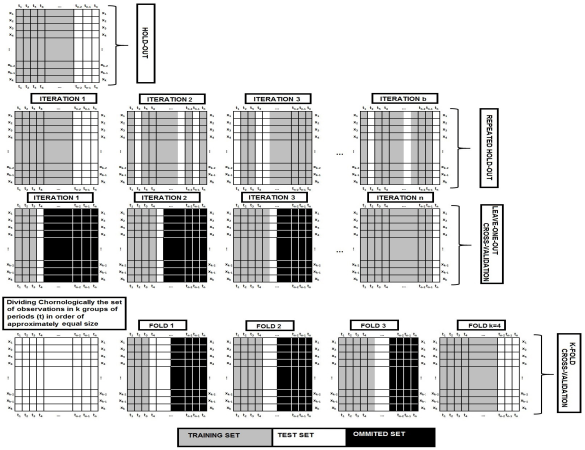
4. Cross-Validation Technique

Cross-validation of non-time series data



4. Cross-Validation Technique

Cross-validation of time series data



4. Cross-Validation Technique

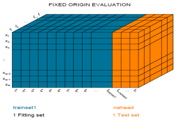
Cross-validation of time-series data following the terminology of Tashman (2000); Bergmeir and Benítez (2012).

- **Fixed-origin evaluation.** A forecast for each value present in the test set is computed using only the training set. The forecast origin is fixed to the last point in the training set. So, for each horizon only one forecast can be computed.
- **Within rolling-origin-recalibration evaluation.** Forecasts for a fixed horizon are performed by sequentially moving values from the test set to the training set, and changing the forecast origin accordingly. For each forecast, the model is recalibrated using all available data in the training set, which often means a complete retraining of the model.
- **Rolling-origin-update evaluation.** In the forecasts the values from the test set are not moved to the training set, and no model recalibration is performed. Instead, past values from the test set are used merely to update the input information of the model. **USE CAREFULLY**
- **Rolling-window evaluation.** Similar to rolling-origin evaluation, but the amount of data used for training is kept constant, so that as new data is available, old data from the beginning of the series is discarded.It is only applicable if the model is rebuilt in every window, and has merely theoretical statistical advantages, that might be noted in practice only if old values tend to disturb model generation.

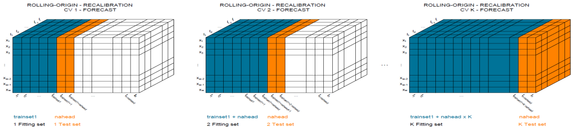
4. Cross-Validation Technique

Cross-validation of panel data

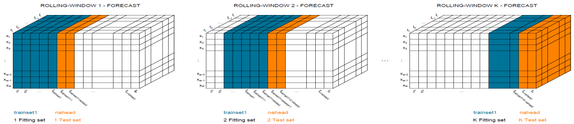
- **Fixed-origin evaluation.**



- **Rolling-origin-recalibration evaluation**



- **Rolling-window evaluation**



4. Cross-Validation Technique

The cv methods are implemented in **CvmortalityMult** R-package with the next function:

multipopulation_cv (qxt, model, periods, ages, nPop, lxt=NULL, ktmethod, order = NULL, nahead, trainset1, fixed_train_origin = TRUE, measures)

- fitting object — **qxt**, **ages**, **nPop**, **periods**, **lxt**.
- forecasting object — **ktmethod**, **order**.
- **model** — `c('additive', 'multiplicative', 'CFM', 'joint-K', 'ACFM')`.
- **CV tehcnique applied** — **nahead**, **trainset1**, and **fixed_train_origin** = `c(TRUE, FALSE, 'add_remove1')`.
- **Measures of Forecasting accuracy** — `measures = c('SSE', 'MSE', 'MAE', 'MAPE', 'All')`.

4. Cross-Validation Technique

Practical application in R.
Necessary files:

- 2.MultiPopulationMortalityModels.R
- CountryCombined.RData

5. Conclusions

- We present an **R-package** for fitting different multi-population mortality models.
- Using the library makes it possible to **test the forecasting accuracy of multi-population mortality models using different cross-validation techniques**.
- The package also allows users to **evaluate which model performs best in the short and long term**, distinguishing between **male and female populations**, and assessing performance across **different age groups, years, and regions**.

5. References

Bergmeir, C. and Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213.

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two–factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Debón, A., Montes, F., and Martínez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non–divergent populations: an application to Spanish regions. *European Actuarial Journal*, 1(2):291–308.

Haberman, S. and Renshaw, A. (2012). Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and economics*, 50(3):309–333.

Hyndman, R. and Athanasopoulos, G. (2021). *Forecasting. Principles and Practice*. 3rd ed. Melbourne, Australia: OTexts.

Hyndman, R. J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: the product–ratio method with functional time series models. *Demography*, 50:261–283.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of forecasting*, 8(3):393–411.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. *Demography*, 42(3):575–594.

Mitchell, D., Brockett, P., Mendoza-Arriaga, R., and Muthuraman, K. (2013). Modeling and forecasting mortality rates. *Insurance: Mathematics and economics*, 52(2):275–285.

Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., and Perla, F. (2019). A deep learning integrated Lee–Carter model. *Risks*, 7(1):33.

Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy.

Perla, F., Richman, R., Scognamiglio, S., and Wüthrich, M. V. (2021). Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*, 2021(7):572–598.

Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3):393–404.

Renshaw, A. E. and Haberman, S. (2003). Lee–Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.

Renshaw, A. E. and Haberman, S. (2006). A cohort–based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and economics*, 38(3):556–570.

Richman, R. and Wüthrich, M. V. (2021). A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*, 15(2):346–366.

Russolillo, M., Giordano, G., and Haberman, S. (2011). Extending the Lee–Carter model: a three–way decomposition. *Scandinavian Actuarial Journal*, 2011(2):96–117.

Tashman, L. J. (2000). Out–of–sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437–450.

Wilmoth, J. and Valkonen, T. (2001). A parametric representation of mortality differentials over age and time. In *Fifth Seminar of the EAPS Working Group on Differentials in Health, Morbidity and Mortality in Europe, Pontignano, Italy*.

Thanks for your attention