CvmortalityMult: Cross-Validation for Multi-Population Mortality Models

by David Atance, and Ana Debón

Abstract This article presents the CvmortalityMult R package, a novel tool designed for modelling, forecasting and evaluating mortality models for several populations. The package facilitates the fitting and forecasting of multipopulation mortality models, providing accurate projections in an increasingly interconnected world characterized by minimal or no borders between countries. By incorporating different cross-validation (CV) techniques, the package allows for the assessment of the forecasting accuracy of multipopulation mortality models for specific countries or regions within a country. Through an empirical application to Spanish regions, we demonstrate the efficacy and simplicity of the CvmortalityMult R package in selecting and evaluating multipopulation mortality models. By providing accessible tools for mortality modelling, forecasting and testing, this package stands out as a valuable resource for advancing the understanding and forecasting of mortality trends across diverse populations. Its contributions extend to enhancing decision-making in critical fields such as life insurance, public health, and pension plan sustainability.

1 Introduction

Currently, 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). These models enable the joint fitting of multiple populations, regions in the same country, or both sexes simultaneously.

On the basis of the original Lee and Carter (1992) model, many researchers have developed models to fit the mortality of related populations or countries with similar socioeconomic statutes or even both sexes in the same population (Brouhns et al., 2002; Debón et al., 2011; Dowd et al., 2011; Jarner and Kryger, 2011; Li and Hardy, 2011; Russolillo et al., 2011; Villegas and Haberman, 2014; Danesi et al., 2015; Chen and Millossovich, 2018; Bégin et al., 2023). The application of these multipopulation mortality models is relevant in diverse contexts, facilitating the concurrent modelling of multiple populations. This is particularly true within the field of life insurance, where companies operate worldwide and make assumptions about future mortality trends. Therefore, using such models ensures the consistency and reliability of the products across different countries. Notably, in the European Union, where gender-neutral pricing is enforced, it is necessary to model both sexes simultaneously (Ahmadi and Li, 2014). Moreover, multipopulation mortality models can include the correlation structure among populations when projecting future trends (Antonio et al., 2017; Bozikas and Pitselis, 2020).

The rise of "big data" has shifted the focus of many problems, and its methods have become a new field complementary to statistics. Within this domain, "resampling methods" are fundamental tools; these techniques are based on repeatedly drawing samples from a dataset and refitting models to obtain additional information. Advances in computational power have increased researchers' interest in these methods, which were developed in 1990. The two most common types of "resampling methods" are bootstrap and "cross-validation" (CV) methods (James et al., 2013a).

Bootstrap is a fundamental tool in the actuarial field that has had multiple applications throughout the literature. This method has been employed for various purposes, such as prediction errors in claim insurance (England and Verrall, 1999), establishing confidence bounds for discounted reserves (Hoedemakers et al., 2003), and estimating confidence

intervals in mortality through different bootstrap versions (Brouhns et al., 2005; Koissi et al., 2006; Debón et al., 2008b; Liu and Braun, 2010; D'Amato et al., 2012). The CV method divides sample data into k folds, where $k - 1$ subsets are typically employed to train the model(s), and the remaining set is used to test the forecasting accuracy (Hastie et al., 2009). The process is iterated k times. For time series data, preserving the chronological order of the sets is necessary. This technique has been employed in other fields, such as finance, biology, and marketing. When applying CV methods to mortality modelling, these techniques must be adapted to time series analysis to use all available data for both testing and training (Tashman, 2000; Bergmeir and Benítez, 2012), which are also known as time series CVs (Hastie et al., 2009). Specifically, in each iteration, the training set must consist of observations that chronologically occurred before the test set observations corresponding to the end of each series (Hyndman and Athanasopoulos, 2021). Furthermore, in our analysis, each observation corresponds to a three-way array involving three categories: age, period and region or country. As a result, these methods have been adapted to assess the forecasting ability of multipopulation mortality models appropriately.

In the context of mortality modelling it should be noted that only recently researchers have applied CV techniques. However, most applications have focused on single-population models, see, for instance, Villegas et al. (2017); Hyndman et al. (2019); Atance et al. (2020b); Kessy et al. (2022); SriDaran et al. (2022); Lindholm and Palmborg (2022), and Barigou et al. (2023). To the best of our knowledge, the existing literature on the use of resampling methods in mortality modelling has focused on single-population mortality models, and among these studies, only one SriDaran et al. (2022) introduced a CV function designed for fitting, forecasting, and testing out-of-sample age-specific probabilities of death, with the selection of the testing period based on the mean squared error (MSE) measure.

This paper introduces the CvmortalityMult R package, which allows us to fit and forecast five multipopulation mortality models. Moreover, this package enables the application of CV methods to select the "best" multipopulation mortality model in forecasting among different scenarios. The idea is to determine which model produces the best forecasting outcomes for the period and the selected countries or regions. To achieve this goal, we implement several CV techniques following the terminology established by Tashman (2000) and Bergmeir and Benítez (2012) for evaluating time-series forecasts. Additionally, we adapt the methodology proposed by Atance et al. (2020b), which is primarily designed for single-population mortality models, to evaluate the forecasting ability of multipopulation mortality models over short, medium and long term horizons. The package incorporates multiple CV techniques to facilitate this evaluation.

Moreover, the package includes five variations of the classical Lee and Carter (1992) model to fit and forecast mortality in regions/populations that form part of a group rather than considering them individually. First, Russolillo et al. (2011) proposed adding a new multiplicative effect to represent different countries/regions within a multiplicative mortality model. Second, Debón et al. (2011) integrated the region/country effect as an additive index through an additive model. Third, Carter and Lee (1992) and Li and Lee (2005) first modelled the entire group and then incorporated a specific term for each region/country using the common-factor model. Fourth, Carter and Lee (1992) and Wilmoth and Valkonen (2001) proposed the joint-K model, which includes two specific country/region terms along with a common trend. Finally, Li and Lee (2005) and Hyndman and Ullah (2007) extended the common-factor model by incorporating two additional region/country terms with the augmented common-factor model. These five multipopulation mortality models were chosen because of their promising results compared with those of other mortality models (Debón et al., 2011; Dong et al., 2020), and their frequent usage in multipopulation modelling literature (Villegas et al., 2017). We introduce into the CvmortalityMult R package several functions that allow us to fit, forecast and evaluate those five multipopulation mortality models. However, if the functions detect only one population, they can fit the well-known Lee and Carter (1992) model.

The paper is structured as follows. Section 2 focuses on describing multipopulation mortality models. Section 3 discusses the CV methods in multipopulation mortality models.

Section 4 presents the CvmortalityMult *R* package, installation and the main functions. Section 5 presents a case study detailing the use of the package. Finally, Section 6 draws conclusions from the results in the previous section.

2 Models

In this section, we introduce five of the most important multipopulation mortality models, which serve as benchmarks to test the forecasting accuracy using CV methods. Indeed, the CvmortalityMult *R* package requires a set of crude age-specific probabilities of death for age x, period t and, in each region i, $\dot{q}_{x,t,i}$. These crude rates are directly obtained as $\dot{q}_{x,t,i} = d_{x,t,i}/E^0_{x,t,i}$, where $d_{x,t,i}$ represents the number of recorded deaths and $E^0_{x,t,i}$ denotes the initial population exposed to risk for an age x, period t and, region i. Crude mortality rates, along with other life table indicators, can be obtained using the LifeTable function from the **MortalityLaws** *R* package (Pascariu, 2022). This set of crude probabilities is then used to generate smoothed and forecasted estimates, $\hat{q}_{x,t,i}$, of the true but unknown mortality probabilities $q_{x,t,i}$. Therefore, in the context of multipopulation mortality data, "one observation ahead" corresponds to a set of data containing the probabilities of death for all ages and populations considered for the following year.

2.1 Multiplicative mortality model

Russolillo et al. (2011) proposed incorporating a multiplicative index term into the Lee and Carter (1992) model to shift the mortality for each region in the population group. That is:

$$\text{logit}\,(q_{x,t,i}) = \log\left(\frac{q_{x,t,i}}{1 - q_{x,t,i}}\right) = a_x + b_x \cdot k_t \cdot I_i + \varepsilon_{x,t,i}; \tag{1}$$

where a_x captures the general age shape of the mortality curve, k_t describes the general trend of the group of populations over time, b_x represents how each age-specific probability of death reacts to changes in the general level of mortality, and I_i represents a multiplicative index associated with mortality for each member of the group of populations considered. Both a_x and b_x are the common age-dependent and time-independent parameters for all the regions considered respectively, whereas k_t is the same for all considered regions and corresponds with the time-dependent parameter, as in the initial version of Lee and Carter (1992), I_i is a different index for each population considered.

Notably, the logit link function is employed to fit the age-specific probabilities of death. The logit transformation ensures that values of $q_{x,t,i}$ between 0 and 1 (Lee, 2000) are obtained. This transformation also maintains the historical ties to the early actuarial work of Perks (1932), as noted by Haberman and Renshaw (2011).

2.2 Additive mortality model

Debón et al. (2011) propose incorporating an additive index term into the Lee-Carter structure to modify the mortality of each region in the multipopulation model. Its expression is as follows:

$$\text{logit}\,(q_{x,t,i}) = \log\left(\frac{q_{x,t,i}}{1 - q_{x,t,i}}\right) = a_x + b_x \cdot k_t + I_i + \varepsilon_{x,t,i}. \tag{2}$$

As in the previous model, the same components a_x, b_x and k_t are shared across all the studied regions (populations). This fact is a necessary and sufficient condition to avoid divergence in the forecasting of age-specific probabilities of death of subpopulations; see Debón et al. (2011) and Ahcan et al. (2014). The number of parameters is the same as that in the model of Russolillo et al. (2011), with similar interpretations except for I_i; more details are given in Debón et al. (2011). However, the additive formulation provides a more straightforward structure than the multiplicative formulation does because it incorporates regional effects through an additive index term.

2.3 Common-factor mortality model

Carter and Lee (1992) and Li and Lee (2005) proposed modelling the mortality of different populations through a common long-term component for the whole group combined with an age-dependent specific term for each population i. This model is represented as:

$$\text{logit}\,(q_{x,t,i}) = \log\left(\frac{q_{x,t,i}}{1 - q_{x,t,i}}\right) = a_{x,i} + B_x \cdot K_t + \varepsilon_{x,t,i}, \tag{3}$$

where $a_{x,i}$ represents the baseline shape of the mortality curve for each ith specific population, while the long-term change over time across the whole mortality group is captured by $B_x \cdot K_t$. These parameters serve the same function as b_x and k_t do in previous mortality models but apply to the whole group of populations.

2.4 Joint-k mortality model

Carter and Lee (1992) and Wilmoth and Valkonen (2001) introduced a model that assumes two specific population age-dependent terms, and a common trend among the group of populations, is given by:

$$\text{logit}\,(q_{x,t,i}) = \log\left(\frac{q_{x,t,i}}{1 - q_{x,t,i}}\right) = a_{x,i} + b_{x,i} \cdot k_t + \varepsilon_{x,t,i}. \tag{4}$$

where k_t represents a common mortality trend among the different considered populations while a_{x_i} and b_{x_i} are specific for each population i. Indeed, the a_{x_i} parameter retains the same meaning as in the common factor mortality model, and b_{x_i} captures the effect of a time-varying mortality index k_t at age x for each population i.

2.5 Augmented common-factor mortality model

Li and Lee (2005) and Hyndman et al. (2013) introduced two population-specific terms to the common factor model. This model is expressed as:

$$\text{logit}\,(q_{x,t,i}) = \log\left(\frac{q_{x,t,i}}{1 - q_{x,t,i}}\right) = a_{x,i} + B_x \cdot K_t + b_{x,i} \cdot k_{t,i} + \varepsilon_{x,t,i}. \tag{5}$$

The first three terms correspond to the components of the common-factor mortality model (3), whereas $b_{x_i} \cdot k_{t_i}$ captures deviations in the short to medium-term changes in the age-specific probability of death for each population i to the common trend. Importantly, including the population-specific terms b_{x_i} and k_{t_i} may imply significant divergences in the mortality forecasts across different populations. Our choice of capital letters in Equations (3) and (5) follows the notation introduced by Li and Lee (2005), who use upper-case symbols to denote the common (or "group-wide") Lee–Carter factor and lower-case symbols for the population-specific effects.

All the models discussed are implemented using the software R Core Team (2022) via the **gnm** library (Turner and Firth, 2023). Details on the calibration approach can be found in Debón et al. (2011). The parameters are obtained by maximizing the model's log-likelihood, assuming a quasi-Binomial distribution of deaths in all considered models:

$$L\,(q_{x,t,i}; \hat{q}_{x,t,i}) = \sum_x \sum_t \sum_i w_{x,t,i} \left\{ q_{x,t,i} \cdot \log\,(\hat{q}_{x,t,i}) + (1 - q_{x,t,i}) \cdot \log\,(1 - \hat{q}_{x,t,i}) + \text{cte} \right\}, \tag{6}$$

where $w_{x,t,i}$ corresponds to weights assigned to each age, period, and population considered, $\hat{q}_{x,t,i}$ is derived by rearranging the terms in Equation (1):

$$\hat{q}_{x,t,i} = \frac{e^{a_x + b_x \cdot k_t \cdot I_i}}{1 + e^{a_x + b_x \cdot k_t \cdot I_i}}, \tag{7}$$

in the case of the multiplicative model. For the other models, the corresponding inverse transformations apply for (2) – (5). We maximize the log-likelihood function because it is an effective estimation method in the actuarial and demography literature for the parameter estimation process (Brouhns et al., 2002; Renshaw and Haberman, 2006; Cairns et al., 2009).

2.6 Forecasting multipopulation mortality models

To project the age-specific probabilities of death, $q_{x,t,i}$, it is essential to forecast the value of the trend parameters k_{t_n}, $k_{t_n,i}$, and K_{t_n} for all the multipopulation mortality models. These models are formulated as follows:

$$
\begin{aligned}
\text{Multiplicative} &\rightarrow \text{logit}\,(q_{x,t_n+s,i}) = a_x + b_x \cdot k_{t_n+s} \cdot I_i, \\
\text{Additive} &\rightarrow \text{logit}\,(q_{x,t_n+s,i}) = a_x + b_x \cdot k_{t_n+s} + I_i, \\
\text{Common-factor} &\rightarrow \text{logit}\,(q_{x,t_n+s,i}) = a_{x,i} + B_x \cdot K_{t_n+s}, \\
\text{Joint-k} &\rightarrow \text{logit}\,(q_{x,t_n+s,i}) = a_{x,i} + b_{x,i} \cdot k_{t_n+s}, \\
\text{Augmented common-factor} &\rightarrow \text{logit}\,(q_{x,t_n+s,i}) = a_{x,i} + B_x \cdot K_{t_n+s} + b_{x,i} \cdot k_{t_{(n+s)},i},
\end{aligned}
\tag{8}
$$

where $q_{x,t_n+s,i}$ corresponds to the forecasted age-specific probability of death for age x, period $t_n + s$ and population i, and k_{t_n+s}, $k_{t_{(n+s)},i}$, K_{t_n+s} are the projections of the trend parameters k_{t_n}, $k_{t_n,i}$, and K_{t_n} considering that t_n is the last in-sample period.

In the CvmortalityMult R package, three alternative approaches are considered to project the time series k_t, K_t and k_{t_i}, assuming they follow ARIMA (autoregressive integrated moving average model) independent processes. First, a random walk with drift (ARIMA (0,1,0) with drift) is assumed, which is a common assumption in the actuarial literature (Cairns et al., 2006; Haberman and Renshaw, 2011; Villegas et al., 2018). Second, the CvmortalityMult R package allows the user to assume the best ARIMA (p,d,q) model according to the auto.arima function in the **forecast** R- package (Hyndman and Khandakar, 2008; Hyndman et al., 2023) for each trend parameter k_t, K_t and/or k_{t_i}, as described by Debón et al. (2008b), Villegas et al. (2018) and Hunt and Blake (2020). The auto.arima function determines the best ARIMA (p,d,q) model on the basis of the outcomes according to the corrected Akaike information criterion (AICc). Third, users can specify the (p, d, q) order for each ARIMA model by setting the corresponding parameters. Across the three approaches, the user can decide whether to include different ARIMA configurations by changing the arguments in the auto.arima or Arima functions as in the **forecast** R package.

3 Cross-validation methods

CV is a tool that focuses on assessing the predictive power of models. Identifying the best model benefits insurance companies in actuarial and financial applications, such as pricing and reserving, where forecasting is arguably more relevant than explanation (SriDaran et al., 2022). Thus, this methodology is valuable for identifying which model is the most accurate forecaster for single-population mortality models (Atance et al., 2020b) and can be used for a set of related populations. This section describes the CV methods (Burman, 1989; Bergmeir and Benítez, 2012) applied to evaluate the out-of-sample accuracy of multipopulation models.

The performance of a model varies between in-sample and out-of-sample evaluations (Bartolomei and Sweet, 1989; Pant and Starbuck, 1990). Therefore, partitioning data into training and test sets is fundamental for accurately assessing the forecasting ability of models. Various possibilities exist for evaluating time series forecasts, also referred to as the CV in time series (Hastie et al., 2009). These methods differ on the basis of the forecast horizon and the method of forecasting the out-of-sample validation, also known as "last block evaluation" in individual time series analysis (Tashman, 2000; Bergmeir and Benítez, 2012). Among the different available methods, we have adapted several approaches in CvmortalityMult R package to assess the forecasting ability of multipopulation mortality

models from different perspectives. Specifically, we follow the terminology established by Tashman (2000) and Bergmeir and Benítez (2012) to evaluate time series forecasting but adapt it for multipopulation mortality models.

3.1 Fixed-origin evaluation

Fixed-origin evaluation, also known as the out-of-sample test or hold-out method (Lachenbruch and Mickey, 1968; Tashman, 2000), is one of the most commonly used methods for assessing the forecasting accuracy of mortality models (Ahcan et al., 2014; Atance et al., 2020a). In this approach, adapted for time series analysis, the dataset is chronologically divided only once into training and test sets. The model is fitted using the training set, with its final point as the fixed origin for forecasting, as shown Figure 1 for a three-way array that incorporates three dimensions: ages in rows, periods in columns, and regions in the third dimension. This fixed-origin method generates a single forecast to predict all or specific periods in the test set. The forecasting accuracy is then evaluated for different forecast horizons.

This CV approach has also been employed for assessing the forecasting ability of multi-population mortality models; see, for instance, Danesi et al. (2015), Antonio et al. (2017) and Bozikas and Pitselis (2020).
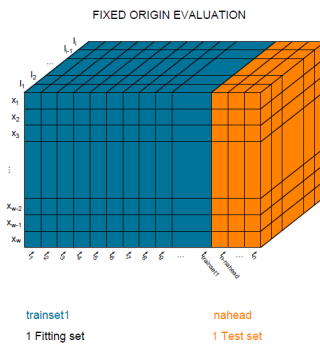

Figure 1: A schematic representation of the three-way array utilizing the fixed-origin cross-validation (CV) technique. The training and test sets are shown in blue and orange, respectively. In our package, the initial training set size is specified by the argument `trainset1`, whereas the forecast horizon is defined by `nahead`. To ensure the proper application of the fixed-origin evaluation, the sum of `trainset1` and `nahead` must equal the total number of provided periods.

3.2 Rolling-origin-recalibration evaluation

In rolling-origin-recalibration (RO-recalibration) evaluation for time series, we initiate the procedure by partitioning the sample into "k" subsets of data, while maintaining the chronological order. The first subset corresponds to the training set, and forecasts are generated with a fixed horizon to assess the model's performance. In each iteration, the model is recalibrated by incorporating all preceding information in the training set (Armstrong and Grohman, 1972; Tashman, 2000; Bergmeir and Benítez, 2012; Hyndman and Athanasopoulos, 2021). The test set periods are sequentially added to the training set to forecast the next set of periods, as shown in Figure 2. Consequently, the beginning of the evaluation shifts forward at each iteration. The model's accuracy is assessed using the average forecasting performance across the k iterations:

$$\text{RO-recalibration}_k = \frac{1}{k} \sum_{i=1}^{k} \text{Goodness of fit measure}_i. \tag{9}$$

In this type of time series CV, the specific variant of rolling-origin (RO) recalibration applied depends on the size of the initial training set and the forecast horizon of each test set. For

instance, k-fold CV (Hastie et al., 2009; James et al., 2013b; Bergmeir et al., 2018) requires the training set size and forecast horizon for each test set to be equal. Notably, to our knowledge, the application of the k-fold CV for analysing the forecasting proficiency of multipopulation mortality models has yet to be documented. We recommend that the initial training set contain more periods than the forecast horizon does to ensure reliable results. If the training and test set sizes differ, the common version RO-recalibration should be applied.

Notably, the "leave-one-out CV" (LOOCV) (Burman, 1989; Shao, 1993) is a special case of RO-recalibration, where the forecast horizon is equal to one, regardless of the initial training set size. Unlike approaches that generate two subsets of comparable size, LOOCV is a distinct approach that involves selecting and forecasting a single observation as the test set, whereas the preceding available observations constitute the training set. To implement this approach, the procedure is repeated (n - `trainset1`) times, where n denotes the total number of observations in the dataset and (`trainset1`) is the size of the initial training set.

Among the different resampling methods, this technique is widely acknowledged for assessing the predictive performance of single mortality models, as evidenced in various studies, such as those in Li and O'Hare (2019), Atance et al. (2020b), Barigou et al. (2023), and Atance and Navarro (2024). Additionally, only one preliminary work Atance and Debón (2022) applied RO-recalibration LOOCV, i.e., the prediction of multipopulation mortality models moving forward by one year for each iteration.
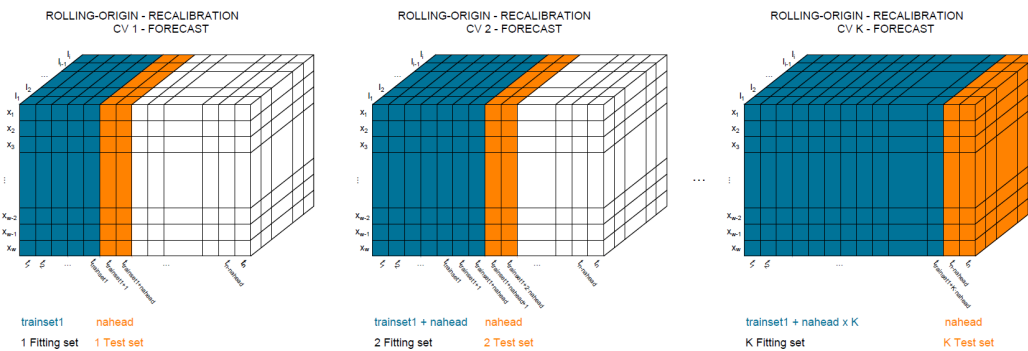

Figure 2: Schematic representation of the cross-validation method for a three-way array using the rolling-origin (RO) recalibration technique. The training, test, and omitted sets are depicted in blue, orange, and white, respectively. The initial training set size is denoted as `trainset1`, whereas `nahead` specifies both the forecast horizon and the size of each test set.

3.3 Rolling-window evaluation

"Rolling-window evaluation" is similar to RO-recalibration but maintains a constant training set size across each forecast iteration (Armstrong and Grohman, 1972; Tashman, 2000). It is also referred to as "time series CV" (TSCV) (Hart, 1994; Bergmeir and Benítez, 2012), "fixed-size rolling-window" (Swanson and White, 1997), or "fixed-size rolling sample" (Callen et al., 1996). Data are partitioned into training and test sets as in previous techniques. In each iteration, the training set incorporates the forecasted periods from the test set while discarding the earliest observations and preserving chronological order, as shown in Figure 3. The model and the forecast origin are also recalibrated at each window/iteration. The forecasting accuracy of the model is assessed using the same procedure as that in RO-recalibration, as expressed in Eq. (9).

Similar to the RO-recalibration technique, there are different variants of rolling-window evaluation. Indeed, the common CV, k-fold CV and LOOCV approaches are also variants. However, in these approaches, the training set size is constant throughout the iterations.

To our knowledge, the CvmortalityMult *R* package provides the first function for analysing the forecasting ability of multipopulation mortality models using CV techniques.
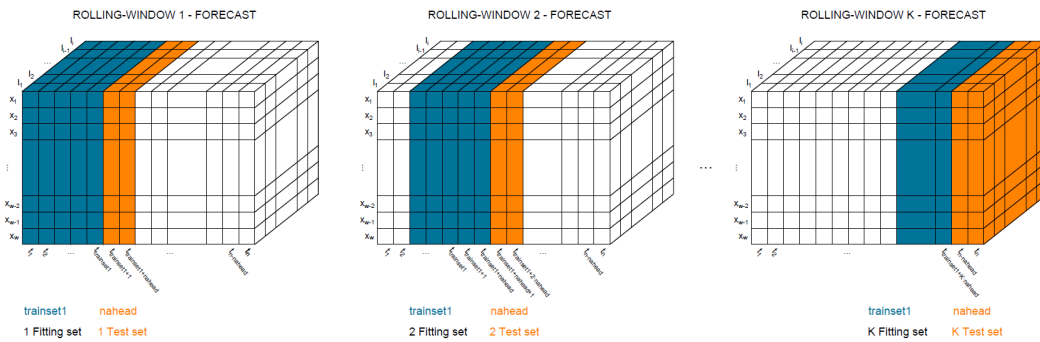
Figure 3: Schematic representation of the cross-validation method for a three-way array using the rolling-window evaluation technique. The training, test, and omitted sets are depicted in blue, orange, and white, respectively. The training set size, denoted as `trainset1`, remains consistent across iterations, starting with an initial training set. The argument `nahead` specifies both the forecast horizon and the size of each test set.

Notably, RO recalibration and rolling-window evaluation have not previously been implemented in any *R* package for assessing the forecasting accuracy of multipopulation mortality models. The CvmortalityMult *R* package addresses this gap by enabling the application of various time series CV methods.

4 The CvmortalityMult R-package

Table 1 introduces the main functions incorporated in the CvmortalityMult *R* package, along with a brief description of every function. Furthermore, these functions have been categorized into four groups: fitting, forecasting, plotting, and CV. In the following sections, we explain with several examples the procedural details and parameter structure essential for using the primary functions of the package.

Initially, the procedure involves the use of five mortality multipopulation functions to calibrate age-specific probabilities of death across various regions (populations). Notably, the CvmortalityMult *R* package allows for the fitting of a single-mortality model when the user provides data for only one population.

During the forecasting stage, the package subsequently facilitates age-specific probabilities of death projections under various $ARIMA(p, d, q)$ specifications, employing the object obtained in the preceding step. During the plotting stage, users have the opportunity to visualize the parameters obtained in the initial fitting stage, forecasts of the age-specific probability of death, and displays specific values across the Spanish regions in a geographical map. This procedure allows for a comprehensive understanding of the behaviour exhibited by each population under consideration.

Finally, during the CV stage, we introduce the function for applying different resampling methods to assess the forecasting accuracy of multipopulation mortality models. This function enables the modification of the fitting and forecasting periods, uses the functions of the preceding steps, and allows for the evaluation of the model forecasting performance using various goodness-of-fit measures. Notably, the structure of the multipopulation mortality data was aligned with a three-way array (age × time × region/population). For each probability of death, data are available for different ages, periods and regions/populations. Consequently, the applications of CV techniques need to be adapted to evaluate the projecting ability of these models effectively.

Notably, this paper does not include an example of every function argument. However, the reader is referred to the function documentation and the package vignette for a complete description of the CvmortalityMult *R* package.

Two datasets are included in the package: `SpainRegions` and `SpainNat`. These datasets originate from the Spanish National Institute (Instituto Nacional de Estadística, INE). Life

	Function name	Brief description
Fitting	fitLCmulti()	Fitting the multiplicative, additive, common-factor, joint-k or augmented common-factor multipopulation mortality models, and the single version of the Lee-Carter model.
Forecasting	forecast.fitLCmulti() or S3 method forecast()	Forecasting the multiplicative, additive, common-factor, joint-k or augmented common-factor multipopulation mortality models, and the single version of the Lee-Carter model.
Plotting	plot.fitLCmulti() the S3 method plot()	Plot the parameters for the multipopulation or single-population mortality models.
	plot.forLCmulti() the S3 method plot()	Plot the forecasting parameters for the multipopulation or single-population mortality models.
	SpainMap()	Plot the regions of Spain with the percentiles of the variable chosen by the users.
CV	multipopulation_cv()	CV techniques using the methods described in Section 2.3.
Measures of accuracy	MeasureAccuracy()	Measure for testing the accuracy of the single-population or multipopulation mortality models.

Table 1: Summary of the main functions in the CvmortalityMult *R* package.

tables and abridged lifetables were obtained with the methodology proposed by Muriel et al. (2010) on the basis of the work in Elandt-Johnson and Johnson (1980). On the one hand, the SpainRegions dataset comprises 10800 observations, encompassing 20 age groups, 30 periods (1990 - 2020), and 18 regions for both males and females in Spain, including national data (the regions in Spain are referred to as autonomous communities). On the other hand, the SpainNat dataset contains 600 observations, corresponding to national data for males and females in Spain covering 20 age groups, 30 periods (this dataset was created for the application of a single-population mortality model, and it can be obtained as a subgroup of the SpainRegions database). These datasets are structured as a data frame and include the following variables:

- ccaa: A vector of the 17 different regions of Spain, including national data. Figure 4 shows the identification of each region on the Spanish map.

- years: A vector of the period range, spanning from 1990 - 2020 for both datasets.

- ages: A vector of the age groups (children under 1 year, between 1 year and 5 years, and then by groups of 5 years, with the last group being between 90 and 94 years).

- qx_male: A vector of the age-specific probabilities of death for the male population.

- qx_female: A vector of the age-specific probabilities of death for the female population.

- lx_male: A vector of the estimated number of individual males living in each age group during each period in a specific region/population, based on an initial group of $l_0 = 100,000$ individuals aged 0 (Pitacco et al., 2009).

- lx_female: A vector of the estimated number of individual females alive in each age group during each period in a specific region/population.

- series: The sex included in both datasets "male and female population".

- label: A tag indicating the dataset type, either "Spain regions" or "Spain National population".
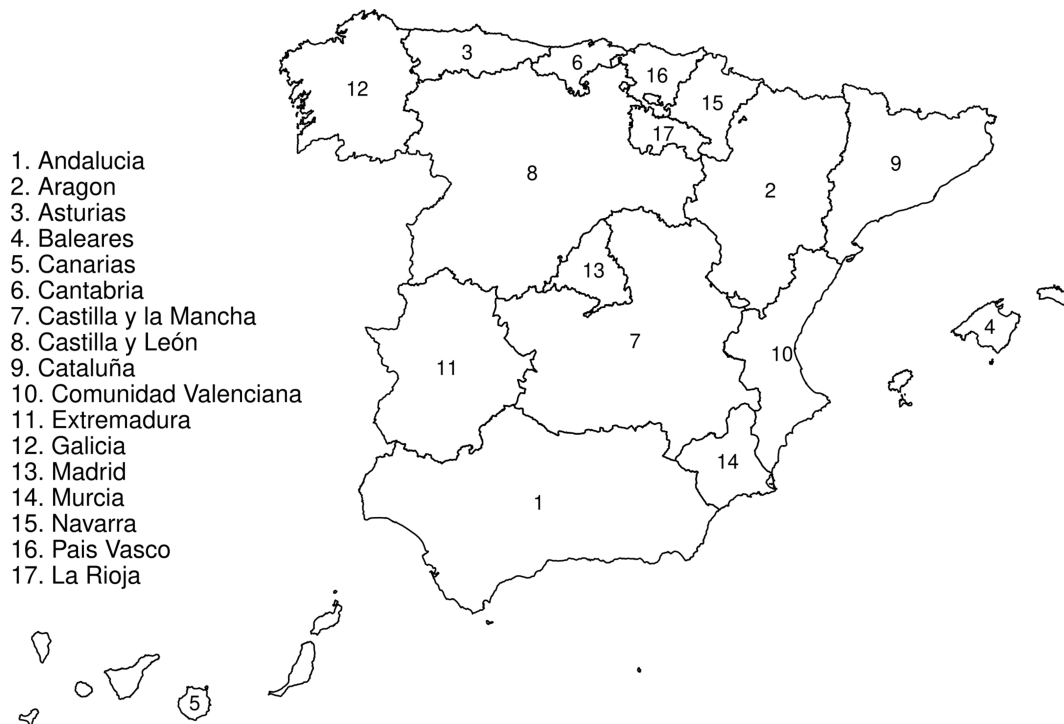

1. Andalucia
2. Aragon
3. Asturias
4. Baleares
5. Canarias
6. Cantabria
7. Castilla y la Mancha
8. Castilla y León
9. Cataluña
10. Comunidad Valenciana
11. Extremadura
12. Galicia
13. Madrid
14. Murcia
15. Navarra
16. Pais Vasco
17. La Rioja

Figure 4: Administrative structure of Spain (regions are referred to as autonomous communities).

Furthermore, we have included the dataset `regions`, which contains the geographical values of the Spanish regions. The `SpainMap` function facilitates the creation of a map displaying the Spanish regions along with the variables incorporated in this dataset.

5 Application of the CvmortalityMult R- package

5.1 Fitting

Model fitting of the age-specific probabilities of death, $q_{x,t,i}$ (at age x, period t, and i region/population), under a quasi-Binomial distribution of deaths and a logit link is performed using the **gnm** package developed by Turner and Firth (2023).

The proposed multipopulation mortality models present challenges related to parameter identifiability. The parameter solution for the considered multipopulation mortality models $(a_x, a_{x_i}, b_x, B_x, b_{x_i}, k_t, K_t, k_{t_i}, I_i)$ are not unique, as any transformation of these parameters that preserves the model structure is also a solution, highlighting inherent identifiability problems in mortality models (Enchev et al., 2017; Villegas et al., 2018). This identifiability problem is addressed as follows: by setting $k_{t_0} = 0$, $b_0 = 1$ and $I_1 = 1$ for the multiplicative model; by setting $k_{t_0} = 0$, $b_0 = 1$ and $I_1 = 0$ for the additive model (Debón et al., 2011), by setting $B_0 = 1$, and $K_{t_0} = 0$ for the common-factor model (Carter and Lee, 1992), by setting $b_{0,1} = 1$, and $k_{t_0} = 0$ for the joint-K model (Carter and Lee, 1992), and by setting $B_0 = 1$, $b_{0,1} = 1$, $K_{t_0} = 0$, and $k_{t_{0,1}} = 0$ for the augmented-common-factor model (Li and Lee, 2005).

The `fitLCmulti()` facilitates the fitting of multipopulation mortality models, including the multiplicative model (Russolillo et al., 2011), additive model (Debón et al., 2011), common-factor model (CFM) (Carter and Lee, 1992), joint-K model (Carter and Lee, 1992) and augmented common-factor model (ACFM) (Li and Lee, 2005). It also supports fitting a single version of the Lee-Carter model (Lee and Carter, 1992) in the CvmortalityMult R package. The synopsis of this function is outlined below:

```
fitLCmulti(model, qxt, periods, ages, nPop, lxt = NULL)
```

The fitting function requires the following information as input:

- The model refers to the multipopulation mortality model chosen to fit the mortality rates. The available options include c('additive','multiplicative','CFM','joint-K','ACFM'). Users must select one model to fit.

- qxt is a vector or matrix containing the crude age-specific probabilities of death for every age, period, and region. The function automatically identifies the data structure (vector or matrix) that users provide.

- lxt is a vector or matrix with the estimated number of individual males alive in each age group during each period in a specific region. The function automatically identifies the data structure (vector or matrix) that users provide. If this argument is not included (NULL), the function internally estimates this value to obtain the parameters for fitting the multipopulation mortality model.

- The periods, and ages vectors reflect the period range and age range, respectively, from the dataset.

- nPop is a numeric value that indicates the number of populations/regions considered in the dataset.

Importantly, for the effective utilization of the fitting functions, the array or matrix containing the qxt and lxt (if it is included) should be organized chronologically with the primary or general population placed first. This fact is essential for the ACFM, which needs to fit first the mortality of the whole group. In the dataset labelled SpainRegions, the principal population pertains to the mortality data encompassing the entire nation of Spain and is positioned as the first entry in the dataset.

We demonstrate the application of this function by fitting the additive and multiplicative multipopulation mortality models to the SpainRegions dataset for both male and female cases. However, the other multipopulation mortality models can be applied only by modifying the model input. Indeed, in the explanation of the function fitLCmulti() in the CvmortalityMult *R* package, users can find examples of how other multipopulation mortality models (common-factor, joint-k and augmented common-factor) are fitted and forecasted for male Spain regions. Additionally, we generate a vector containing the lower age in each age group considered in the paper.

```
> SpainRegions
Mortality Data
Spain Regions for males and females
Years 1991 : 2020
Abridged Ages 0 : 90
> ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40,
+           45, 50, 55, 60, 65, 70, 75, 80, 85, 90)
```

In fact, multiplicative and additive multipopulation mortality models can be fitted using the following code:

```
> additive_Spainmales <- fitLCmulti(model= 'additive', qxt = SpainRegions$qx_male,
+        periods = c(1991:2020), ages = c(ages), nPop = 18, lxt = SpainRegions$lx_male)
> additive_Spainfemales <- fitLCmulti(model= 'additive', qxt = SpainRegions$qx_female,
+        periods = c(1991:2020), ages = c(ages), nPop = 18, lxt = SpainRegions$lx_female)
> multi_Spainmales <- fitLCmulti(model= 'multiplicative', qxt = SpainRegions$qx_male,
+        periods = c(1991:2020), ages = c(ages), nPop = 18, lxt = SpainRegions$lx_male)
> multi_Spainfemales <- fitLCmulti(model = 'multiplicative', qxt = SpainRegions$qx_female,
+        periods = c(1991:2020), ages = c(ages), nPop = 18, lxt = SpainRegions$lx_female)
```

The output from the fitting functions is an object of the class `fitLCmulti`, which provides a brief summary of the fitting process, including among other things, the following information:

- `ax`, `bx`, `kt`, and `Ii` are the estimated parameters for the multipopulation mortality models.

- `formula`, and `model` refer to the gnm formula and the fitted multipopulation mortality model, respectively.

- `data.used` includes mortality rates to fit the mortality data.

- `qxt.crude` refers to the crude values of the probabilities of death for every age, period, and region. These values are provided by the user for fitting the selected mortality model.

- `qxt.fitted`, and `logit.qxt.fitted` are the fitted values of the probabilities of death for every age, period, and region using the multipopulation mortality model on a probability or logit scale $\left(\frac{q_{x,t}}{1-q_{x,t}} \right)$.

Once we have adjusted the crude age-specific probabilities of death for different groups of ages, periods, and regions, the `plot.fitLCmulti()` function allows us to show the parameters obtained. Figures 5 and 6 provide the fitted parameters of the additive and multiplicative multipopulation mortality models, respectively, for male populations in Spain. The plots are generated using the following code:

```
> plot(additive_Spainmales)
> plot(multiplicative_Spainmales)
```

Notably, the `plot.fitLCmulti()` function generates different plots depending on the selected model. For example, if the augmented common-factor model is chosen by setting `model = 'ACFM'`, the plot function will display the estimated parameters, a_{x_i}, B_x, K_t, b_{x_i}, and k_{t_i}, for the provided populations.

Our example is the dataset of Spanish regions. We have included the `SpainMap` function in the package. This function facilitates the plotting of the I_i parameters of the regions of Spain in Figure 7. We recommend reviewing the `regions` dataset to identify the order of the regions before using the `SpainMap` function. In the context of multipopulation mortality models, the multiplicative and additive indices for the regions of Spain (with the reminder that the first population is the national dataset and will not be shown) can be obtained with the following code:

```
> SpainMap(multiplicative_Spainmales$Ii[2:18],
+        main = c("Multiplicative for males"),
+        name = c("Ii"))
> SpainMap(regionvalue = additive_Spainmales$Ii[2:18],
+        main = c("Additive for males"),
+        name = c("Ii"), bigred = FALSE)
```

Additionally, the fitting function applies to a single population. Specifically, we design this function to fit cases where mortality data are provided for only one population, and the Lee-Carter mortality model for a single population is fitted. Users can implement this by using the following *R* code:

```
> LC_SpainNatmales <- fitLCmulti(model = 'additive', qxt = SpainNat$qx_male,
+        periods = c(1991:2020), ages = c(ages), nPop = 1, lxt = SpainNat$lx_male)
> LC_SpainNatfemales <- fitLCmulti(model = 'multiplicative', qxt = SpainNat$qx_female,
+        periods = c(1991:2020), ages = c(ages), nPop = 1, lxt = SpainNat$lx_female)
```
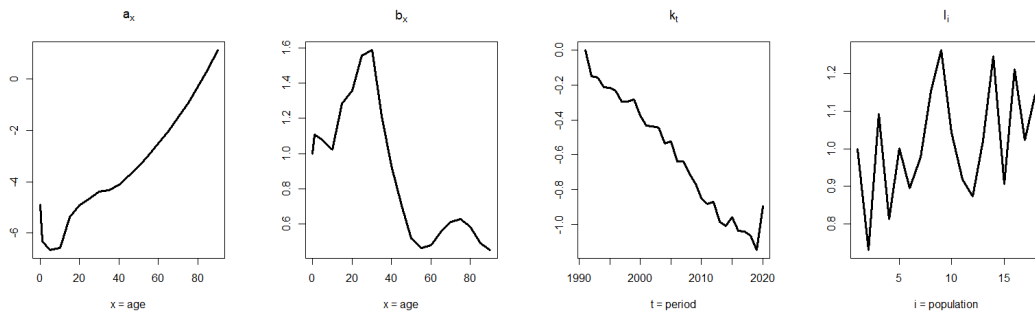
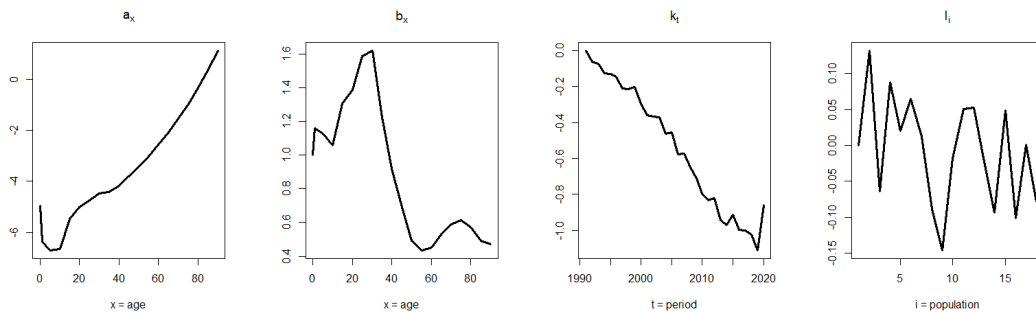**Figure 5:** Parameters for the multiplicative mortality model fitted to the male population of Spain for ages 0 – 90 and the period 1991 – 2020.

We use two of the five multipopulation models to demonstrate the operation of the function for one-single population independently of the model provided. However, there is no need to specify this argument in the `fitLCmulti` function, as it inherently fits the Lee-Carter version for a single population.

Similarly, the parameters of the Lee-Carter model for single-population mortality can be plotted using the `plot.fitLCmulti()` function. Therefore, Figure 8 can be obtained using the following code:

```
plot(LC_SpainNatmales)
```


Figure 6: Parameters for the additive mortality model fitted to the male population in Spain regions for ages 0 – 90 and the period 1991 – 2020.
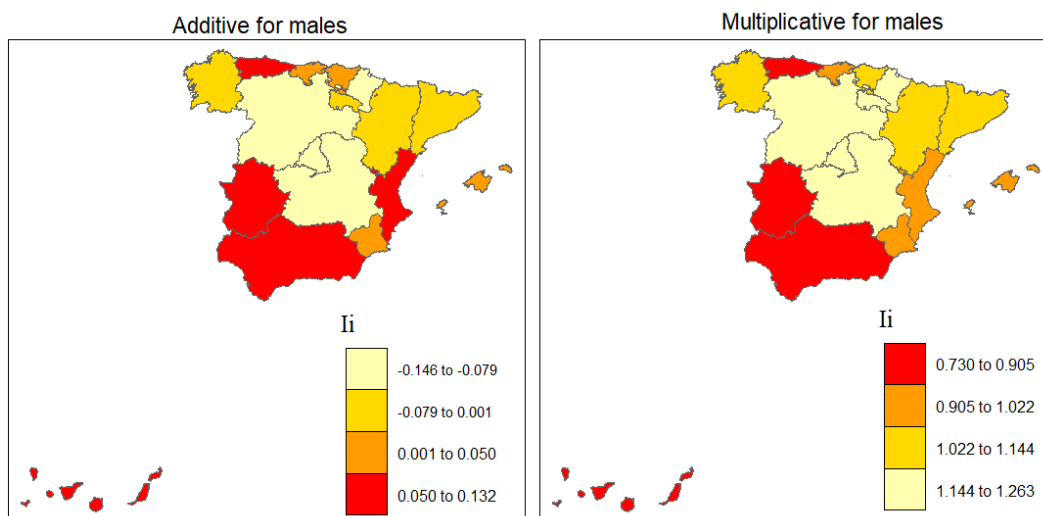

Figure 7: Geographical index, I_i, for multiplicative (left) and additive (right) multipopulation mortality models for the male populations in Spain regions for ages 0 – 90 and the period 1991 – 2020.
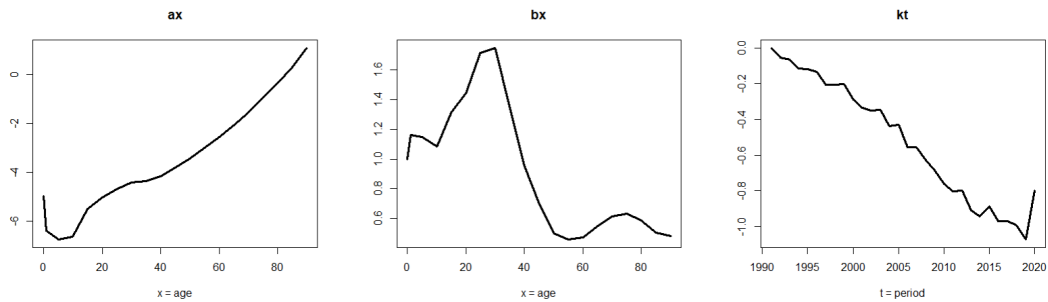
Figure 8: Parameters for the LC single-population mortality model fitted to the Spain male population for ages 0 – 90 and the period 1991 – 2020.

From Figures 5-8, several interesting results emerge:

- All the left panels, a_x, correspond to the average behaviour of age-specific probabilities of death across all studied periods and regions.

- The second panels, b_x, demonstrate how age-specific probabilities of death for each age group (considering all regions) respond to changes in mortality trend, as captured by k_t. Large values of b_x are observed among Spanish males between 20 and 40 years of age; therefore, there is a substantial reduction in the age-specific probabilities of death in this age group (1991 - 2020). This phenomenon is attributed to the impact of AIDS and drugs on Spanish males during the 1980s and 1990s, which led to an initial increase in age-specific probabilities of death and total deaths in these age groups, followed by a significant decline in age-specific probabilities of death due to the introduction of new therapies and medications during the 1990s and 2000s (Felipe et al., 2002; Debón et al., 2008a; Atance et al., 2020a), as can be observed in b_x for the additive and multiplicative models.

- The third panels reveal the impact of the COVID-19 pandemic on the trend parameter k_t for all considered models among males in Spain. Similar trends can be observed for females, although these trends are not shown. However, they can be generated by modifying the fitting object in the plot function. The incorporation of 2020 into the model fitting process induces an upturn in age-specific probabilities of death in the final observed period, disrupting the declining trend observed in the preceding years (1990 - 2019).

- Finally, the right panels, I_i, in the multipopulation approach depict the geographical distribution of the indices corresponding to each region. These panels allow us to discern distinct regional behaviours on the basis of the chosen multipopulation approach. To complement this presentation, we have included Figure 7, which displays the values of I_i for Spanish males in each region using the additive and multiplicative models with the regions with the highest mortality highlighted in red. Notably, the parameter I_i leads to a different interpretation in each model. In the multiplicative model, higher values indicate lower age-specific probabilities of death as the region value is multiplied by the trend parameter. Therefore, with a decreasing trend parameter, as observed in the case of Spain, higher values of I_i correspond to lower age-specific probabilities of death. Conversely, the additive model incorporates the region index to the general trend among the regions ($a_x + b_x \cdot k_t$). Consequently, lower values or the most negative index regions present lower age-specific probabilities of death.

5.2 Forecasting

The CvmortalityMult R package enables the projection of future age-specific probabilities of death using the ARIMA (p,d,q) processes. This projection applies to the trend parameters, k_t, K_t, and $k_{t,i}$ in multipopulation mortality models. Two common assumptions for

the trend parameters k_t, K_t, and $k_{t,i}$ in the actuarial and demography literature are often considered: first, a multivariate random walk with drift (ARIMA(0,1,0)) (Cairns et al., 2006, 2009; Haberman and Renshaw, 2011; Villegas et al., 2017), and second, the selection of the best ARIMA (p,d,q) process (Renshaw and Haberman, 2006; Debón et al., 2008b; Villegas et al., 2017; Atance et al., 2020a) for estimating the future values of k_t, K_t, and k_{t_i}. To estimate the future values of the trend parameters k_t, K_t, and k_{t_i}, we employ the forecast function from the **forecast** R package (Hyndman and Khandakar, 2008), allowing the projection of the future values for various types of ARIMA processes considered in our package. In the CvmortalityMult R package, users can choose different ARIMA processes, ktmethod=c('arima010','arimapdq','arimauser'). The selection process is applied for single or all trend parameters considered in each multipopulation or single-population mortality model. The ellipsis argument (...) provides users with the flexibility to include different ARIMA configurations, changing the arguments in the auto.arima or Arima functions, depending on the ktmethod provided. This functionality mirrors the behavior of the **forecast** R package for time series (Hyndman and Khandakar, 2008). Additionally, users must provide nahead, indicating the number of periods to forecast the future value of age-specific probabilities of death for each considered region. For example, the code below provides future age-specific probabilities of death for Spanish male and female regions for the next ten years (nahead = 10), using different ARIMA options in the package:

```
> fut_additive_Spainmales <- forecast(object = additive_Spainmales,
+        nahead = 10, ktmethod = 'arimapdq')
> fut_multiplicative_Spainmales <- forecast(object = multiplicative_Spainmales,
+        nahead = 10, ktmethod = 'arima010')
> fut_additive_Spainfemales <- forecast(object = additive_Spainfemales,
+        nahead = 10, ktmethod = 'arimapdq')
> fut_multiplicative_Spainfemales <- forecast(object = multiplicative_Spainfemales,
+        nahead = 10, ktmethod = 'arima010')
```

The outputs from these forecast functions are objects of the class forLCmulti, which provides a brief summary of the forecasting process, with the following information:

- ax, bx, kt, and Ii provide the estimated parameters for the multipopulation mortality models.

- arimakt provides the ARIMA (p,d,q) process considered to adjust the time series k_t, K_t, or k_{t_i} and the obtained coefficients.

- kt.fut provides the future values of k_t, K_t, or $k_{t,i}$ using the selected ARIMA (p,d,q) configurations for the nahead periods.

- kt.futintervals provides estimates of the future values of k_t, K_t, or $k_{t,i}$ for the point forecast (kt.fut). Additionally, it includes the lower and upper 80% and 95% confidence intervals, utilizing the chosen ARIMA (p,d,q) process for the specified nahead periods.

- formula and model define the gnm formula and the forecasted multipopulation mortality model, respectively.

- qxt.crude represents the crude values of the probabilities of death for every age, period, and region. These values are provided by the user for fitting the selected mortality models.

- qxt.future, and logit.qxt.future represent the future values of the probabilities of death for every, age, period and region using the chosen multipopulation mortality model in the probability or logit scale.

Once we have projected the age-specific probabilities of death from different ages, periods, and regions, the plot.forLCmulti() function allows us to show the projected values

of trend parameters k_t, K_t, and k_{t_i}. The logit death probabilities for the mean in-sample age and the out-of-sample forecast are shown for all the populations considered. Figures 9 and 10 provide these interesting results for the additive and multiplicative multipopulation mortality models, respectively, for males in Spain. The visualizations are generated with the following code:

```
> plot(fut_additive_Spainmales)
> plot(fut_multiplicative_Spainmales)
```
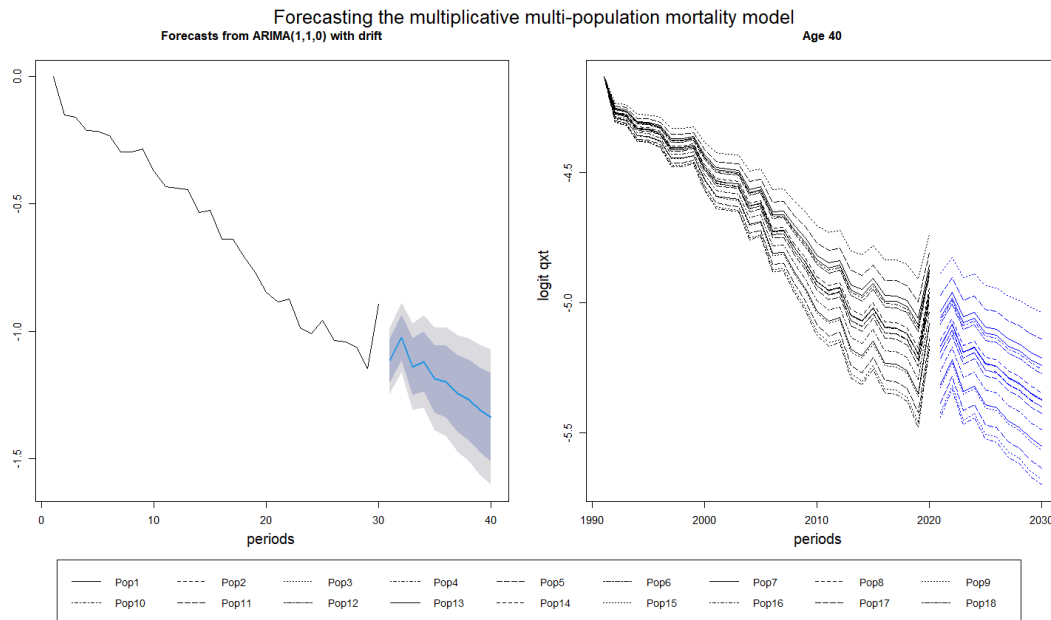

Figure 9: The left panel represents to the in-sample trend parameter k_t and its projected value. The right panel displays the actual and projected logit mortality rates using the multiplicative multipopulation mortality model for the 18 populations considered for age 40 (mean age in the populations considered) in terms of the in-sample period from 1991 – 2020 and the out-of-sample period extending, 10 years ahead.

Similarly, during the forecasting process, users can employ the same function to project future values of age-specific probabilities of death when providing data for a single population. Specifically, the function forecasts age-specific probabilities of death using the Lee-Carter model for a single population, as demonstrated below:

```
> fut_LC_Spainmales <- forecast(object = LC_SpainNatmales,
+       nahead = 10, ktmethod = 'arimapdq')
> fut_LC_Spainfemales <- forecast(object = LC_SpainNatfemales,
+       nahead = 10, ktmethod = 'arima010')
```

Equally, for the single-population mortality model, users can plot two remarkable results. For example, Figure 11 can be obtained using the following code:

```
> plot(fut_LC_Spainmales)
```

5.3 Cross-Validation

In this section, we present the CV function developed in the CvmortalityMult *R* package to evaluate the forecasting ability of multipopulation mortality models with different CV methods. Thus, the CV time series function uses the next synopsis:

```
multipopulation_cv(qxt, model = c('multiplicative', 'additive', 'CFM', 'joint-K', 'ACFM'),
+       periods, ages, nPop, lxt = NULL,
```
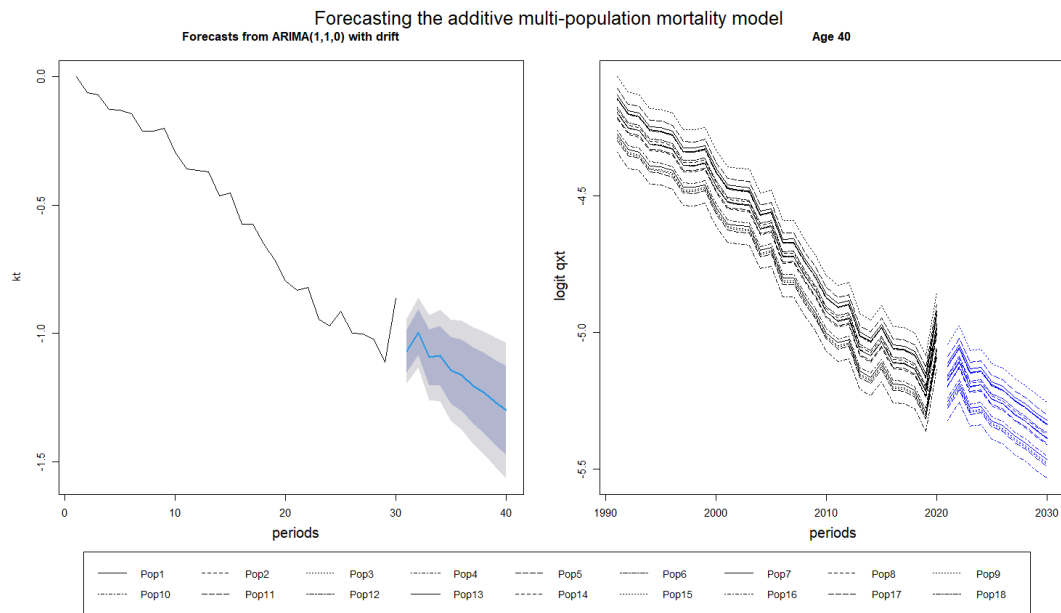
Figure 10: The left panel represents the in-sample trend parameter k_t and its projected value. The right panel displays the actual and projected logit mortality rates using the additive multipopulation mortality model for the 18 populations considered for age 40 (mean age in the populations considered) in terms of the in-sample period from 1991 – 2020 and the out-of-sample period extending, 10 years ahead.
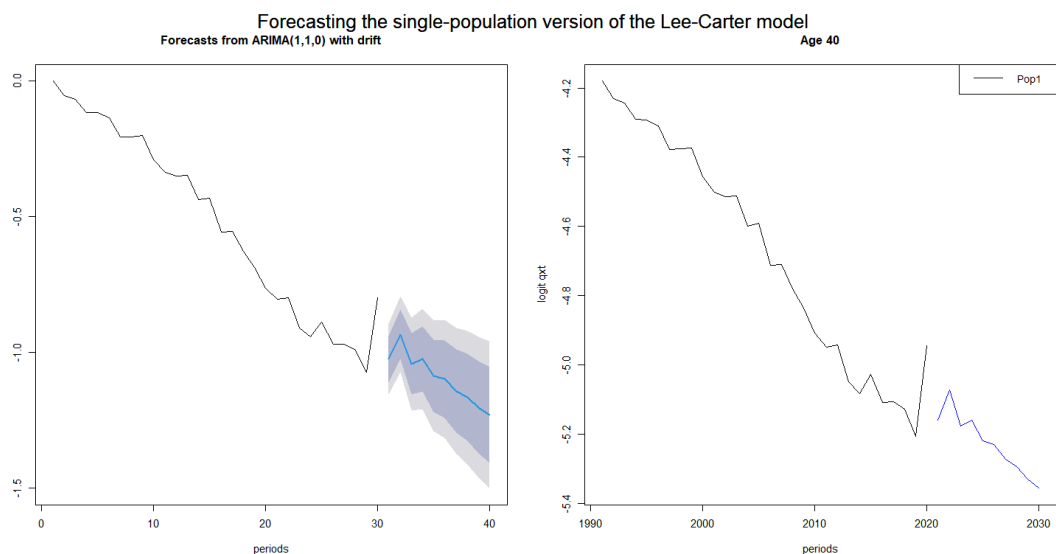

Figure 11: The left panel represents the in-sample trend parameter k_t and its projected value. The right panel displays the actual and projected logit mortality rates using the Lee-Carter model for a single population for the only population considered for age 40 (mean age in the populations considered) in terms of the in-sample period from 1991 – 2020 and the out-of-sample period extending, 10 years ahead.

```
+          nahead, trainset1, fixed_train_origin = TRUE,
+          ktmethod = c('arimapdq', 'arima010', 'arimauser'), order = NULL,
+          measures = c('SSE', 'MSE', 'MAE', 'MAPE', 'All'))
```

This CV function requires the following information as input:

- `qxt`, `lxt`, `periods`, `ages`, `nPop`, `ktmethod` and `order` should match the corresponding values used as in the fitting and forecasting functions for the multipopulation and single mortality models as inputs.

- `model = c('multiplicative','additive','CFM','joint-K','ACFM')` specifies the multipopulation mortality model that users wish to assess for forecasting ability using the specific resampling technique. Users can apply the multiplicative, additive, common-factor, joint-K, and augmented common-factor multipopulation models and the single version of the Lee-Carter model presented in this paper separately.

- `measures = c('SSE','MSE','MAE','MAPE','All')` denotes the adjustment measure that users wish to employ for testing the forecasting ability of the model using the specific resampling technique. If `measures = c('All')`, all the measures will be provided by the function. Additionally, each accuracy measure has a dedicated help function to clarify the underlying equations. Users can access this help function in the CvmortalityMult *R* package using the following code: ?MeasureAccuracy, where users can select the specific measure of accuracy for testing the age-specific mortality rates (SSE, MSE, MAE or MAPE).

- `trainset1` specifies the number of chronological periods to consider as the initial training set. This value must be greater than 2 to meet the minimum time series size (Hyndman and Khandakar, 2008). Additionally, we recommend that this value be greater than `nahead` to maintain consistency among the forecasts in every iteration (Tashman, 2000).

- `nahead` is the number of periods to project ahead in each iteration and the size of each test set among the selected CV techniques. Moreover, it should be noted that the `multipopulation_cv()` function aims to maintain a uniform length for all the testing sets (iterations). However, the last test set may have fewer periods to align with the total number of periods provided by the user as (`periods`).

- `fixed_train_origin = c(TRUE,FALSE,'add_remove1')` is a logical variable that specifies whether the starting point of the initial training set remains fixed throughout the CV process. This option allows users to maintain a constant starting point where the model will be fitted in every iteration or allow it to shift, thereby determining whether a rolling window evaluation is applied. By default, the package sets `fixed_train_origin = TRUE`, meaning that the first period in the training set remains fixed across all iterations and the model recalibrations of the CV method. However, users can opt to allow the training set starting point to shift by setting `fixed_train_origin = FALSE` or `fixed_train_origin = 'add_remove1'`, thereby implementing a rolling-window evaluation while keeping the training set size constant in each iteration, defined by the user by `nahead` argument. When `fixed_train_origin = FALSE`, in every iteration, the user-defined `nahead` periods are removed from the beginning of the training set, while the next `nahead` periods are added to the training set from the previous test set. Consequently, the number of projected periods (`nahead`) also determines how many periods are added or deleted in every iteration. In contrast, when `fixed_train_origin = 'add_remove1'`, the training set size is also fixed across iterations. However, only one new period is added and one period is removed in each forecast. This process allows users to evaluate the forecasting accuracy of `nahead`-step-ahead projections using more test samples. Users may specify `fixed_train_origin = 'add_remove1'`, and any value of `nahead` greater than or equal to one, provided it is consistent with the number of periods available in the dataset. Notably, when `nahead = 1`, the configuration `fixed_train_origin = 'add_remove1'` yields results equivalent to those obtained using `fixed_train_origin = FALSE`, which means in a LOOCV method keeping the same size of the training set (`trainset1`) across iterations.

The sizes of `nahead` and `trainset1` can be determined by the temporal correlation within the values of the analysed series by using "blocks" of data rather than choosing data randomly (Racine, 2000; Bergmeir and Benítez, 2012).

With this function, the user can apply different CV methods for multipopulation models depending on three main inputs that must be provided: `nahead`, `trainset1`, and `fixed_train_origin`. Indeed, the following CV techniques can be applied:

1. **Fixed-origin evaluation** is implemented by setting the arguments so that nahead + trainset1 = periods while keeping the default value of fixed_train_origin = TRUE.

2. **RO-recalibration evaluation** requires that trainset1 > 2 and that fixed_train_origin = TRUE remain at its default value, regardless of the value assigned to nahead. Specifically, when nahead = 1, leave-one-out CV (LOOCV) is applied. When nahead = trainset1, k-fold CV (CV) is performed. For all other values, a standard time series CV approach is used while keeping the origin of the first training set fixed in all possible options.

3. **Rolling-window evaluation** requires setting fixed_train_origin = FALSE or fixed_train_origin = 'add_remove1', independently of the values assigned to nahead and trainset1. As in the previous CV technique, if nahead = 1, a LOOCV approach with a rolling window of 1 is applied, which remains equivalent whether fixed_train_origin is set to FALSE or 'add_remove1'. When nahead > 1 and fixed_train_origin = FALSE, the training set is updated by incorporating and discarding nahead periods in each iteration. Conversely, when fixed_train_origin = 'add_remove1', the training set updates by adding and removing only one observation per iteration while forecasting nahead periods.

We present the results for RO-recalibration using the standard CV approach for male Spanish regions. The main input parameters are set as follows: trainset1 = 8, nahead = 5 and fixed_train_origin = TRUE (default value). This procedure is applied to the five multipopulation mortality models included in the package. To replicate these results, the user can use the following code:

```
> SpainRegions
> ages <- c(0, 1, 5, 10, 15, 20, 25, 30, 35, 40,
+          45, 50, 55, 60, 65, 70, 75, 80, 85, 90)

> cv_SM_multi <- multipopulation_cv(qxt = SpainRegions$qx_male,
+         model = c('multiplicative'), #see options below
+         periods = c(1991:2020), ages = c(ages), nPop = 18, lxt = SpainRegions$lx_male,
+         trainset1 = 8, nahead = 5, ktmethod = c('arimapdq'), measures = c("MSE"))
```

While we executed to female Spanish regions an RO-recalibration was performed using trainset1 = 10, nahead = 1 and fixed_train_origin = TRUE (default value). This configuration corresponds to a LOOCV approach, which fixes the origin in the first training, using the following code:

```
> loocv_SF_multi <- multipopulation_cv(qxt = SpainRegions$qx_female,
+         model = c('multiplicative'), #see options below
+         periods =  c(1991:2020), ages = c(ages), nPop = 18, lxt = SpainRegions$lx_female,
+         trainset1 = 10, nahead = 1, ktmethod = c('arimapdq'), measures = c("MSE"))
```

Available values for model = 'additive','multiplicative','CFM','joint-K', and 'ACFM'. Changing the string assigned to model is sufficient to switch to the corresponding specification. The output from the CV function is an object of the MultiCv class, which provides a brief summary of the CV method employed, including the following information:

- ax, bx, Ii, kt.fitted, kt.future, and kt.arima correspond to the same outputs as those in the adjustment and forecast functions. However, since the adjustment process has been repeated several times (depending on the process), each of these outputs is a list of the iterations executed, denoted as follows: loop-h from period-1 to period-2, where "h" denotes the corresponding iteration.

- meas_ages, meas_periodsfut, meas_pop, and meas_total represent the accuracy measures provided by the resampling technique, each emphasizing different aspects of

the forecasting ability. Indeed, the objective of the CvmortalityMult *R* package is to provide a tool for evaluating the forecasting accuracy of multipopulation models from various ages, namely, across different age groups, future periods, regions considered, or a global measure spanning all ages, future periods, and regions considered. This function allows users the flexibility to choose the specific viewpoint they wish to prioritize in the decision-making process regarding forecasting capabilities.

- `model`, and `CV_method` designate the multipopulation mortality model and the CV-method that users wish to apply for testing the forecasting ability, respectively. Users can apply both, the multiplicative and additive models presented in this paper separately.

Figures 12-13 present the results of the CV techniques for the five multipopulation mortality models applied across ages, periods and regions, respectively. Additionally, we have included Figures 14 and 15 with the MSE measure throughout different regions of Spain only for the multiplicative and additive multipopulation mortality models. The result of the other models are available upon request to the authors and can be found in the reproduction file. These plots can be reproduced using the R script entitled `CvmortalityMult_reproduction.R`. From Figures 13-15, we note the following points:

- The ACFM and joint-K models present lower forecasting results when the MSE measure is used across the ages and future periods considered, using both CV time series techniques.

- The five considered models yield similar results for the age range 0 - 60, whereas for the last section of the mortality curve (60 - 90), the ACFM and joint-K model perform better in terms of the forecasting results.

- Concerning the forecasting periods, the ACFM and joint-K models demonstrate better forecasting results in the medium term, as captured by RO-recalibration CV. However, while the five models exhibit comparable results when evaluating short-term forecasting ability, RO-recalibration LOOCV with the multiplicative model yields the worst result.

- The MSE measures for the different regions of Spain considered for each CV method are shown in Figures 14 and 15. The multipopulation mortality model produces different results depending on the region. Specifically, the multiplicative model yields the best forecasting results for Galicia, Pais Vasco, Cataluña, and Comunidad Valencia. In contrast, depending on the CV model and population considered, the additive model produces superior outcomes for Galicia, Asturias, Pais Vasco, Cataluña, Comunidad Valencia, and Andalucia.

Notably, the CV function allows the computation of global measures of forecasting ability considering all ages, future periods, and regions, as shown in Table 2. The results indicate that the ACFM and joint-K model for standard CV, whereas the CFM and ACFM for LOOCV demonstrate better forecasting ability when all available forecasting information is used. This finding is consistent with the observations mentioned above.

Therefore, the CvmortalityMult *R* package displays the forecasting ability of the multipopulation mortality models in various ways, allowing the user to determine the most suitable model for their specific objectives. We present two alternatives to assess the forecasting ability of the models but there are other additional CV time series techniques that can be implemented with the `multipopulation_cv()` function modifying `nahead`, `trainset1` and `fixed_train_origin`.

6   Summary and discussion

Accurately forecasting age-specific probabilities of death is essential for dealing with life-contingent risk, ensuring solvency within the European (re)insurance industry, and address-

CV method	Multipopulation mortality model				
Common CV males	Multiplicative	Additive	CFM	joint-K	ACFM
MSE	0.000160	0.000142	0.000142	0.000132	0.000124
LOOCV females	Multiplicative	Additive	CFM	joint-K	ACFM
MSE	0.000089	0.000066	0.000061	0.000062	0.000059

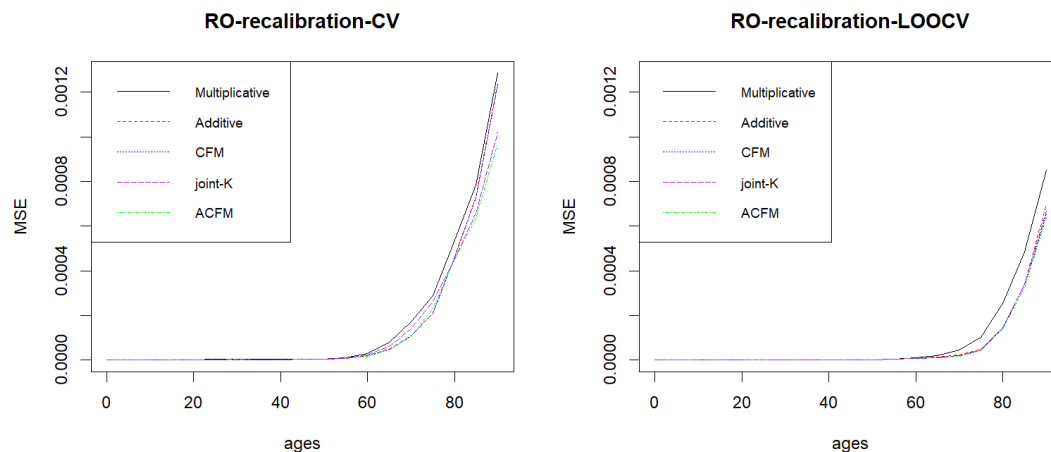**Table 2:** Summary of the MSE global measure of forecasting ability.

Figure 12: Plot visualizing the MSE over the group of ages considered in regions of Spain for males and females, applying CV time series techniques for five
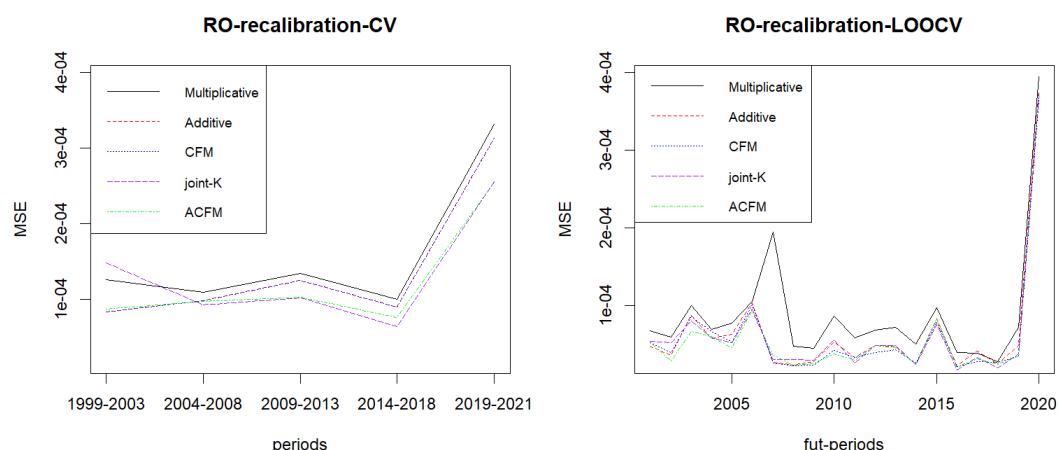multipopulation mortality models.

Figure 13: Plot visualizing the MSE over the test sets of future periods considered in regions of Spain for males and females, applying two CV time series techniques for five
multipopulation mortality models.

ing the sustainability of public pension system plans, among other purposes. Multipopulation mortality models offer a valuable approach to forecasting age-specific probabilities of death. These models allow the incorporation of data from regions in the same country or group of countries with similar characteristics, transcending borders in a globalized context, where national and international movements occur daily. Moreover, these models are recommended to enrich mortality data with observations from different regions in the same country or group of countries sharing similar characteristics. The CvmortalityMult *R* package facilitates access to five of these multipopulation mortality models, providing an *R* interface to the functions necessary for model fitting and forecasting simply.

Furthermore, comparing various models can be challenging when the most suitable model is selected. Indeed, CV methods offer a valuable tool for evaluating the forecasting
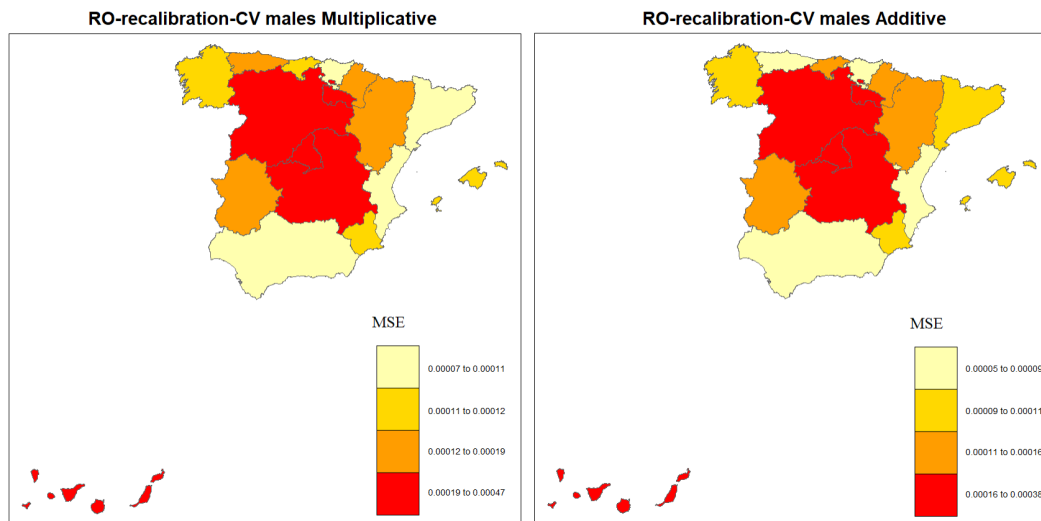
Figure 14: Plot visualizing the MSE in regions of Spain for males, applying RO-recalibration CV for the multiplicative and additive
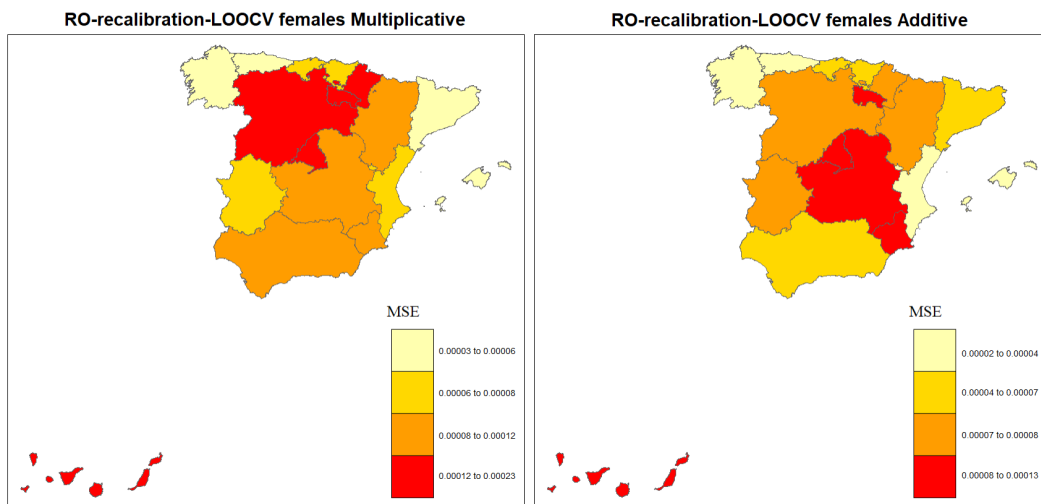multipopulation mortality models.

Figure 15: Plot visualizing the MSE in regions of Spain for males, applying RO-recalibration LOOCV for the multiplicative and additive multipopulation mortality models.

ability of models. The CvmortalityMult *R* package allows the application of several CV time series techniques, for assessing the forecasting ability of multiple populations over short, medium and long term horizon. To the best of our knowledge, the CvmortalityMult *R* package is the first to apply these methods to multipopulation mortality models, especially for three-way array data. Users only need to provide multipopulation mortality data, specify the number of periods to be used as the training or testing set, and decide whether it remains fixed at the origin of the first training set. Then, the CvmortalityMult *R* package computes various measures highlighting different forecasting ability aspects. Consequently, users can prioritize selecting the most appropriate multipopulation mortality model on the basis of their specific requirements and perspectives.

7 Acknowledgements

References

A. Ahcan, D. Medved, A. Olivieri, and E. Pitacco. Forecasting mortality for small populations by mixing mortality data. *Insurance: Mathematics and Economics*, 54:12–27, 2014. [p3, 6]

S. S. Ahmadi and J. S.-H. Li. Coherent mortality forecasting with generalized linear models: A modified time–transformation approach. *Insurance: Mathematics and Economics*, 59: 194–221, 2014. [p1]

K. Antonio, S. Devriendt, W. de Boer, R. de Vries, A. De Waegenaere, H.-K. Kan, E. Kromme, W. Ouburg, T. Schulteis, E. Slagter, M. van der Winden, C. van Iersel, and M. Vellekoop. Producing the Dutch and Belgian mortality projections: a stochastic multi–population standard. *European Actuarial Journal*, 7:297–336, 2017. [p1, 6]

J. S. Armstrong and M. C. Grohman. A comparative study of methods for long–range market forecasting. *Management Science*, 19(2):211–221, 1972. [p6, 7]

D. Atance and A. Debón. *Contributions to Risk Analysis: RISK 2022*, chapter Two multi-population mortality models: A comparison of the forecasting accuracy with resampling methods, pages 51–58. Fundación MAPFRE, 2022. [p7]

D. Atance and E. Navarro. A simplified model for measuring longevity risk for life insurance products. *Financial Innovation*, 1:61, 2024. [p7]

D. Atance, A. Balbás, and E. Navarro. Constructing dynamic life tables with a single–factor model. *Decisions in Economics and Finance*, 43(2):787–825, 2020a. [p6, 14, 15]

D. Atance, A. Debón, and E. Navarro. A comparison of forecasting mortality models using resampling methods. *Mathematics*, 8(9):1550, 2020b. [p2, 5, 7]

K. Barigou, P.-O. Goffard, S. Loisel, and Y. Salhi. Bayesian model averaging for mortality forecasting using leave–future–out validation. *International Journal of Forecasting*, 39(2): 674–690, 2023. [p2, 7]

S. M. Bartolomei and A. L. Sweet. A note on a comparison of exponential smoothing methods for forecasting seasonal series. *International Journal of Forecasting*, 5(1):111–116, 1989. [p5]

C. Bergmeir and J. M. Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012. [p2, 5, 6, 7, 18]

C. Bergmeir, R. J. Hyndman, and B. Koo. A note on the validity of cross–validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83, 2018. [p7]

A. Bozikas and G. Pitselis. Incorporating crossed classification credibility into the Lee–Carter model for multi-population mortality data. *Insurance: Mathematics and Economics*, 93:353–368, 2020. [p1, 6]

N. Brouhns, M. Denuit, and J. K. Vermunt. A Poisson log–bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393, 2002. [p1, 5]

N. Brouhns, M. Denuit, and I. Van Keilegom. Bootstrapping the Poisson log–bilinear model for mortality forecasting. *Scandinavian Actuarial Journal*, 2005(3):212–224, 2005. [p2]

P. Burman. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514, 1989. [p5, 7]

J.-F. Bégin, B. Sanders, and X. Xu. Modeling and forecasting subnational mortality in the presence of aggregated data. *North American Actuarial Journal*, 0(0):1–27, 2023. doi: 10.1080/10920277.2023.2231996. URL https://doi.org/10.1080/10920277.2023.2231996. [p1]

A. Cairns, D. Blake, and K. Dowd. A two–factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718, 2006. [p5, 15]

A. J. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich. 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, 2009. [p5, 15]

J. L. Callen, C. C. Kwan, P. C. Yip, and Y. Yuan. Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting*, 12(4):475–482, 1996. [p7]

L. R. Carter and R. D. Lee. Modeling and forecasting US sex differentials in mortality. *International Journal of forecasting*, 8(3):393–411, 1992. [p2, 4, 10]

R. Y. Chen and P. Millossovich. Sex–specific mortality forecasting for UK countries: a coherent approach. *European Actuarial Journal*, 8(1):69–95, 2018. [p1]

V. D'Amato, S. Haberman, G. Piscopo, and M. Russolillo. Modelling dependent data for longevity projections. *Insurance: Mathematics and Economics*, 51(3):694–701, 2012. [p2]

I. L. Danesi, S. Haberman, and P. Millossovich. Forecasting mortality in subpopulations using Lee–Carter type models: a comparison. *Insurance: Mathematics and Economics*, 62:151–161, 2015. [p1, 6]

A. Debón, F. Montes, J. Mateu, E. Porcu, and M. Bevilacqua. Modelling residuals dependence in dynamic life tables: A geostatistical approach. *Computational Statistics & Data Analysis*, 52(6):3128–3147, 2008a. [p14]

A. Debón, F. Montes, and F. Puig. Modelling and forecasting mortality in Spain. *European Journal of Operational Research*, 189(3):624–637, 2008b. [p2, 5, 15]

A. Debón, F. Montes, and F. Martínez-Ruiz. Statistical methods to compare mortality for a group with non–divergent populations: an application to Spanish regions. *European Actuarial Journal*, 1(2):291–308, 2011. [p1, 2, 3, 4, 10]

Y. Dong, F. Huang, H. Yu, and S. Haberman. Multi–population mortality forecasting using tensor decomposition. *Scandinavian Actuarial Journal*, 2020(8):754–775, 2020. [p2]

K. Dowd, A. J. Cairns, D. Blake, G. D. Coughlan, and M. Khalaf-Allah. A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, 15(2):334–356, 2011. [p1]

R. Elandt-Johnson and N. Johnson. *Survival Models and Data Analysis*. Wiley, New York, 1980. [p9]

V. Enchev, T. Kleinow, and A. J. Cairns. Multi–population mortality models: fitting, forecasting and comparisons. *Scandinavian Actuarial Journal*, 2017(4):319–342, 2017. [p10]

P. England and R. Verrall. Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics*, 25(3):281–293, 1999. [p1]

A. Felipe, M. Guillén, and A. Perez-Marin. Recent mortality trends in the Spanish population. *British Actuarial Journal*, 8(4):757–786, 2002. [p14]

S. Haberman and A. Renshaw. A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1):35–55, 2011. [p3, 5, 15]

J. D. Hart. Automated kernel smoothing of dependent data by using time series cross–validation. *Journal of the Royal Statistical Society Series B: (Methodological)*, 56(3):529–542, 1994. [p7]

T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer Series: New York, NY, USA, 2009. [p2, 5, 7]

T. Hoedemakers, J. Beirlant, M. J. Goovaerts, and J. Dhaene. Confidence bounds for discounted loss reserves. *Insurance: Mathematics and Economics*, 33(2):297–316, 2003. [p1]

A. Hunt and D. Blake. Identifiability in age/period/cohort mortality models. *Annals of Actuarial Science*, 14(2):500–536, 2020. [p5]

R. J. Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Melbourne, Australia, 3rd edition edition, 2021. URL http://OTexts.org/fpp3/. [p2, 6]

R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22, 2008. doi: 10.18637/jss.v027.i03. [p5, 15, 18]

R. J. Hyndman and M. S. Ullah. Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956, 2007. [p2]

R. J. Hyndman, H. Booth, and F. Yasmeen. Coherent mortality forecasting: the product–ratio method with functional time series models. *Demography*, 50:261–283, 2013. [p4]

R. J. Hyndman, H. Booth, L. Tickle, and J. Maindonald. *demography: Forecasting mortality, fertility, migration and population data*, 2019. URL https://CRAN.R-project.org/package=demography. R package version 1.22. [p2]

R. J. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen. *forecast: Forecasting functions for time series and linear models*, 2023. URL https://pkg.robjhyndman.com/forecast/. R package version 8.21.1. [p5]

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning: with applications in R*, volume 1. Springer: New York, NY, USA, 2013a. [p1]

G. James, D. Witten, T. Hastie, R. Tibshirani, et al. *An introduction to statistical learning with application in R*, volume 112. Springer Series: New York, NY USA, 2013b. [p7]

S. F. Jarner and E. M. Kryger. Modelling adult mortality in small populations: the SAINT model. *ASTIN Bulletin*, 41(2):377–418, 2011. [p1]

S. R. Kessy, M. Sherris, A. M. Villegas, and J. Ziveyi. Mortality forecasting using stacked regression ensembles. *Scandinavian Actuarial Journal*, 2022(7):591–626, 2022. [p2]

M.-C. Koissi, A. F. Shapiro, and G. Högnäs. Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, 38(1):1–20, 2006. [p2]

P. A. Lachenbruch and M. R. Mickey. Estimation of error rates in discriminant analysis. *Technometrics*, 10(1):1–11, 1968. [p6]

R. D. Lee. The Lee–Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, 4(1):80–91, 2000. [p3]

R. D. Lee and L. R. Carter. Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419):659–671, 1992. [p1, 2, 3, 10]

H. Li and C. O'Hare. Mortality forecasting: How far back should we look in time? *Risks*, 7 (1):22, 2019. [p7]

J. S.-H. Li and M. R. Hardy. Measuring basis risk in longevity hedges. *North American Actuarial Journal*, 15(2):177–200, 2011. [p1]

N. Li and R. D. Lee. Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. *Demography*, 42(3):575–594, 2005. [p1, 2, 4, 10]

M. Lindholm and L. Palmborg. Efficient use of data for LSTM mortality forecasting. *European Actuarial Journal*, 12:749–778, 2022. [p2]

X. Liu and W. J. Braun. Investigating mortality uncertainty using the block bootstrap. *Journal of Probability and Statistics*, 2010:1–15, 2010. [p2]

S. Muriel, M. Cantalapiedra, and F. López. Towards advanced methods for computing life tables. Technical report, Instituto Nacional de Estadística, 2010. URL www.ine.es. [p9]

P. N. Pant and W. H. Starbuck. Innocents in the forest: Forecasting and research methods. *Journal of Management*, 16(2):433–460, 1990. [p5]

M. D. Pascariu. *MortalityLaws: Parametric Mortality Models, Life Tables and HMD*, 2022. URL https://CRAN.R-project.org/package=MortalityLaws. R package version 1.9.4. [p3]

W. Perks. On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63(1):12–57, 1932. [p3]

E. Pitacco, M. Denuit, S. Haberman, and A. Olivieri. *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press, Oxford, 2009. [p9]

R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL https://www.R-project.org/. [p4]

J. Racine. Consistent cross–validatory model–selection for dependent data: hv–block cross–validation. *Journal of Econometrics*, 99(1):39–61, 2000. [p18]

A. E. Renshaw and S. Haberman. A cohort–based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570, 2006. [p5, 15]

M. Russolillo, G. Giordano, and S. Haberman. Extending the Lee–Carter model: a three–way decomposition. *Scandinavian Actuarial Journal*, 2011(2):96–117, 2011. [p1, 2, 3, 10]

J. Shao. Linear model selection by cross–validation. *Journal of the American statistical Association*, 88(422):486–494, 1993. [p7]

D. SriDaran, M. Sherris, A. M. Villegas, and J. Ziveyi. A group regularisation approach for constructing generalised age–period–cohort mortality projection models. *ASTIN Bulletin*, 52(1):247–289, 2022. [p2, 5]

N. R. Swanson and H. White. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, 13(4):439–461, 1997. [p7]

L. J. Tashman. Out–of–sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437–450, 2000. [p2, 5, 6, 7, 18]

H. Turner and D. Firth. *gnm: Generalized Nonlinear Models*, 2023. URL https://CRAN.R-project.org/package=gnm. R package version 1.1-5. [p4, 10]

A. M. Villegas and S. Haberman. On the modeling and forecasting of socioeconomic mortality differentials: an application to deprivation and mortality in England. *North American Actuarial Journal*, 18(1):168–193, 2014. [p1]

A. M. Villegas, S. Haberman, V. K. Kaishev, and P. Millossovich. A comparative study of two–population models for the assessment of basis risk in longevity hedges. *ASTIN Bulletin*, 47(3):631–679, 2017. [p2, 15]

A. M. Villegas, V. K. Kaishev, and P. Millossovich. StMoMo: An R package for stochastic mortality modeling. *Journal of Statistical Software*, 84(3):1–38, 2018. doi: 10.18637/jss.v084.i03. [p5, 10]

J. Wilmoth and T. Valkonen. 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*, 2001. [p2, 4]

David Atance
Universidad de Alcalá
Departmento de Economía y Dirección de Empresas
Facultad de Ciencias Económicas, Empresariales y Turismo
Plaza San Diego s/n, 6020, Alcalá de Henares
Spain
ORCiD: 0000-0001-5860-0584
david.atance@uah.es

Ana Debón
Universitat Politècnica de València
Centro de Gestión de la Calidad y del Cambio
Facultad de Administración y Dirección de Empresas
Camino de Vera, s/n, 46022, Valencia
Spain
ORCiD: 0000-0002-5116-289X
andeau@eio.upv.es