SAPIENZA
UNIVERSITÀ DI ROMA

Improving mortality diagnostics and estimation through Contrast Trees

Scuola di Dottorato in Statistica

Dottorato di Ricerca in Statistica - Curriculum Scienze Attuariali (XXXV cycle)

Matteo Lizzi

ID number 1888467

Advisor

Prof.ssa Susanna Levantesi

Academic Year 2023/2024

Improving mortality diagnostics and estimation through Contrast Trees
PhD thesis. Sapienza University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: matteo.lizzi@uniroma1.it

Abstract

Contrast Trees are an iterative input space partition technique introduced by [Friedman(2020)] in order to automatically uncover regions in the input space itself where two target variables differ the most. In case inaccuracies are detected, Estimation Contrast Boosting can be used in order to reduce differences. The Distribution Contrast Boosting provides an assumption-free method of estimating the full probability distribution of an outcome variable on the same input space.

By applying for the first time such techniques in the context of mortality modeling, the aim of this work is threefold. Firstly, to test if Contrast Tree based diagnostic can be applied to mortality description and projection models, and if results obtained from this new technique are consistent with those give by some traditional indicators. Secondly, to utilize Estimation Contrast Boosting techniques for building mortality tables for small populations. Thirdly, to generalize the Italian actuarial practice for building company mortality tables, which is based on reproportioning mortality rates. Such generalization can be obtained using both Estimation Contrast Boosting and Distribution Contrast Boosting.

Contents

Chapter 1

Mortality modeling and Contrast Trees

1.1 Introduction

Since 1980, innovative approaches and developments in mortality modeling have been constantly proposed. Mortality analysis has received a considerable contribution from statistical science, building solid foundations for studying the evolution of mortality and longevity trends. Estimating these phenomena is not a trivial task; accuracy depends on the particular situation or trend, and figuring out if or when a certain method will be effective is not straightforward. Indeed, when a new mortality model appears in literature, it may take years before they it be fully evaluated. As stated by [Booth and Tickle(2008)], the accuracy of mortality estimates should be regularly tested to assess the improvement evidence. Researchers appear to be more focused on technical progress of a method rather than on the accuracy of the estimation provided, focusing on minimizing the bias.

Several approaches have been used to model mortality surfaces, determining how death rates change over time. Until the 1980s, relative simplicity and personal judgment were common features of mortality models (see [Pollard(1987)] for a detailed review on this aspect). Since then, mortality data became more easily available, which, jointly with the development of a wider array of statistical-mathematical tools, led to the creation of increasingly complex mortality models. According to [Booth and Tickle(2008)], three main paradigms for demographic modeling can be identified.
The first one (explanation) is based on structural or epidemiological models from certain causes of death: a widely known example is the connection between lung cancer and tobacco smoking. Expectation, the second paradigm, relies on expert opinion and involves varying degrees of formality. Finally, making use of the regular structures typically found in age patterns and trends over time leads to the widely adopted extrapolative paradigm. This last approach includes the more complex stochastic mortality models such as the Lee-Carter ([Lee and Carter(1992)]) and, more generally, the Generalized Age Period Cohort (GAPC) model. Despite the Lee-Carter model being recognized as one of the most influential proposals for

mortality modeling and forecasting, in recent years scholars suggested additional approaches that sparked interest in actuarial and demographic literature as in the works from [Brouhns et al.(2002)], [Renshaw and Haberman(2006)], [Cairns et al.(2006)] and [Cairns et al.(2009)].

1.2 Main frameworks for mortality modeling

Models like the Lee-Carter and its variants have been widely used, becoming a benchmark for many newly proposed methodologies, but they present several shortfalls. In this line, [Cairns et al.(2008)] tried to address the issue of what would be the best way to estimate mortality, exhibiting interesting criteria that a good mortality model should hold. They referred to good-practice guidelines such as the consistency with historical data and the long-term dynamics, biologically reasonable. Following this line of research, recent longevity literature stimulated the use of machine learning techniques in demographic research allowing the integration of stochastic models into a data-driven approach.

Machine learning techniques can support and integrate traditional mortality models in order to significantly reduce forecasting errors, both for application and research purposes. Main contributions in this area, from [Deprez et al.(2017), Levantesi and Pizzorusso(2019), Levantesi and Nigri(2020)], make use of machine learning algorithms to improve the fitting accuracy of canonical models. That is to say, the mortality surface produced by standard stochastic mortality models is corrected by adjusting mortality rates estimated by the original model. Such adjustments are obtained calibrating a machine learning estimator. As shown by these authors, machine learning better captures complex patterns that traditional models fail to identify.

The need for new tools for comparing models' performances is evident to understand mortality evolution more accurately.

1.3 Traditional diagnostic tools

In the following, we briefly mention some traditional diagnostic tools that are often used in the literature to assess the goodness-of-fit of a mortality model.

- Analysis of mortality residuals (or standardized mortality residuals) calculated as the difference between the crude estimate of mortality rate by age and year based on observed data and the corresponding estimated mortality rate using a specified mortality model. For example, [Cairns et al.(2010)] verified that they are consistent with the hypothesis of i.i.d. $N(0,1)$ and have zero correlation both across adjacent ages and across adjacent years.

- Proportion of variance explained (R^2) by the model or the parameters of the model (see, e.g., [Bongaarts(2005)])

- Model selection criteria that penalize the log-likelihood with the increase in number of parameters: Akaike Information Criterion (AIC), Schwarz-Bayes

Criterion (SBC) (or Bayes Information Criterion (BIC)) and Likelihood-ratio test (LRT) [Li et al.(2009)]. Note that in this case the evaluation of the goodness-of-fit is given on the basis of the log-likelihood.

- Qualitative model selection criteria: [Cairns et al.(2008)] provide a list of criteria that might be considered desirable in a mortality model, such as, e.g., ease of implementation, parsimony, and transparency. Relating to the fitting ability to the observed data, the model should be consistent with historical data, and parameter estimates should be robust relative to the range of data used. For example, [Djeundje et al.(2022)] consider consistency, stability, and parsimony in addition to standard goodness-of-fit indices (deviance residual, BIC, and residual patterns).

- Checking for the absence of autocorrelation in the residuals of the model by the Portmanteau test (see, e.g., [Torri(2011)]).

1.4 Contrast Trees and associated boosting techniques

Contrast Trees (CT) are an innovative approach that, leveraging tree-based machine learning techniques, allows for deeply assessing differences between two output variables, be them predictions resulting from different models models or the observed outcome data, by identifying where they most differ. Specifically, the goal of the Contrast Trees method is to uncover regions in the predictor variables space presenting very high values of some difference between two arbitrary outcome variables defined on the same space, quantified by a discrepancy measure [Friedman(2020)].

Suppose to have a system under study, with p associated variables $X_1, X_2, \ldots, X_p$. Given an observed value of the predictor variable vector $\mathbf{x} = (x_1, x_2, ..., x_p)$, the goal of the study is to estimate the unknown value of some target variable or of some property of its distribution, such as the mean or median. For each observed value $\mathbf{x}$, let y and z be the associated observed values for two outcome variables: these can be predictions for the quantity of interest, resulting from different models or methodologies, or observed values for the target variable. The data then consists of N observations of the form $\{\mathbf{x}_i; y_i, z_i\}_{i=1}^N$: every value of the predictor vector is associated with the corresponding couple of observed outcome variables. The aim of the Contrast Tree procedure is to find those values of $\mathbf{x}$ for which the respective distributions of $y|\mathbf{x}$ and $z|\mathbf{x}$, or some statistics such as mean or quantiles, are most different. Depending on the nature of y and z outcome variables, CTs can be used to uncover regions of the input space where the predictions from two models differ the most, or where the prediction of a model is most distant from the corresponding observed values for the outcome variable, thus highlighting high error regions and providing a diagnostic tool for assessing lack-of-fit.

Essentially, Contrast Trees can be used to investigate differences between any observed or predicted variable defined on a given predictor space. If differences between such variables are found to be significant fot the problem at hand, a CT-based boosting technique called Estimation Contrast Boosting can be applied in order

to reduce them. A special case of it, called Distribution Boosting, can be used to estimate the full conditional distribution of one target variable starting from data values available for the other.

Before applying CTs in the context of mortality modeling, the splitting algorithm and its associated boosting procedures will be now described in detail.

1.4.1 Contrast Trees

Following the terminology from [Hastie et al. (2009)], in the remainder of this thesis variables $X_1, X_2, \ldots, X_p$ will be called interchangeably *inputs*, *predictors* or *features*, while variables Y and Z will be referred to as *outputs*, *responses* or *outcomes*.

Discrepancy measures

Any subset X^* of the input space is associated to a discrepancy measure between the corresponding y and z values:

$$d = D(\{y_i\}_{\mathbf{x}_i \in X^*}, \{z_i\}_{\mathbf{x}_i \in X^*})$$

The choice of the discrepancy measure depends on the problem to be solved, allowing Contrast Trees to be applied to a variety of problems [Friedman(2020)]: in this sense, they are similar to loss criteria in prediction problems. However, in the context of Contrast Trees there is no requirement that they be convex or even differentiable. Moreover, they need not be expressible as a sum of terms each involving a single observation.

In the remainder of this work, the following discrepancies will be used:

1. *Mean absolute difference discrepancy*:

$$d_m^{[1]} = \frac{1}{N_m} \sum_{x_i \in R_m} |y_i - z_i| \tag{1.1}$$

   where N_m in the number of observations in region R_m.

2. *Empirical cumulative distribution discrepancy*:

$$d_m^{[2]} = \frac{1}{2N_m - 1} \sum_{i=1}^{2N_m - 1} \frac{\left| \hat{F}_y(t_{(i)}) - \hat{F}_z(t_{(i)}) \right|}{\sqrt{i \cdot (2N_m - i)}} \tag{1.2}$$

   where $t_{(i)}$ is the i^{th} value of t in sorted order, and $\hat{F}_y$ and $\hat{F}_z$ are the respective empirical cumulative distributions of y and z. It should be noted that, since $d_m^{[2]}$ is defined region-wise on certain points of input space, the empirical cumulative distribution of y and z have the same support by construction. See [Friedman(2020)] for further possible discrepancy functions.

Iterative splitting procedure

Once a suitable discrepancy function has been chosen, the algorithm for building Contrast Trees can be described as follows:

1. At the beginning of the iterative procedure, no splitting has been performed yet. Therefore, at 0-th iteration, all input space is contained within a single region.

2. Suppose to have performed M splits: so, at M-th iteration, the input space is partitioned into M disjoint regions $\{R_m\}_{m=1}^{M}$, each containing a subset $\{\mathbf{x}_i, y_i, z_i\}$ of the data. Within each region, the discrepancy measure between y and z values is calculated for each specified subset of the data $\{\mathbf{x}_i, y_i, z_i\}_{x_i \in R_m}$:

$$d_m = D(\{y_i\}_{\mathbf{x}_i \in R_m}, \{z_i\}_{\mathbf{x}_i \in R_m}) \tag{1.3}$$

3. In order to perform the next partition, at the $M+1$-th iteration each region R_m is provisionally partitioned into two sub-regions $R_m^{(l)}$ and $R_m^{(r)}$, with corresponding discrepancy $d_m^{(l)}$ and $d_m^{(r)}$. As in ordinary regression trees for numeric variables (see [Breiman et al.(1984)] for more details), splits involve one of the predictor variables, x_j and are specified by a *split point* s. The provisional sub-regions are defined by:

$$\mathbf{x} \in R_m \wedge x_j \leq s \Rightarrow \mathbf{x} \in R_m^{(l)}$$
$$\mathbf{x} \in R_m \wedge x_j > s \Rightarrow \mathbf{x} \in R_m^{(r)}$$

4. Within each region, the quality of the split into sub-regions is defined as:

$$Q_m(l, r) = \left(f_m^{(l)} \cdot f_m^{(r)}\right) \cdot \max\left(d_m^{(l)}, d_m^{(r)}\right)^{\beta} \tag{1.4}$$

   In the first factor, designed to penalize highly asymmetrical splits, $f_m^{(l)}$ and $f_m^{(r)}$ are the fraction of observations in region R_m associated with each provisional sub-region. The second factor attempts to isolate sub-regions with high discrepancy. Parameter β regulates the relative influence of the two factors but, as stated by [Friedman(2020)], results are insensitive to its value. In the following, β will be held fixed to 2.

5. Within each region R_m all possible splits are performed. The one maximizing equation 1.4, with associated sub-region discrepancies $d_m^{*(l)}$ and $d_m^{*(r)}$, is then associated with the region itself.

6. Discrepancy improvement for the m-th region is then defined as

$$I_m = \max\left(d_m^{*(l)}, d_m^{*(r)}\right) - d_m$$

   The region whose associated split maximizes I_m is then replaced by its associated sub-regions, thus partitioning the input space in $M+1$ regions.

7. The iterative splitting procedure stops when at least one of the following conditions is met:
   - $I_m < 0 \ \forall m$ i.e., discrepancy improvement is negative for all regions resulting from the iterative split procedure;
   - when the tree reaches a specified size, given by the number of disjoint regions produced;
   - when the observation count within all regions is below a specified threshold.

   Tree size is generally specified by the user and involves a trade-off between fineness and interpretability. The smaller the trees, the larger the regions (defined by simple rules and thereby easier to be interpreted). The larger the trees, the higher the potential to uncover small high discrepancy regions (defined by complex rules). Pruning strategies analogous to those in classification and regression trees based on cross-validation can also be employed to guide choice of tree size.

Lack-of-fit contrast curves

Overall results from a Contrast Tree can be summarized in a lack-of-fit contrast curve (more briefly, LOF-curve), which associates to the m-th region the following point coordinates:

$$(f_m;\ \bar{d}_m) = \left(\frac{1}{N} \sum_{d_j \geq d_m} N_j;\ \frac{\sum_{d_j \geq d_m} d_j N_j}{\sum_{d_j \geq d_m} N_j} \right)$$

where $\bar{d}_m$ is the average discrepancy across all regions with discrepancy greater or equal to d_m, weighted with the number of observations within each of these regions, N_j and f_m is the fraction of observations within those same regions.

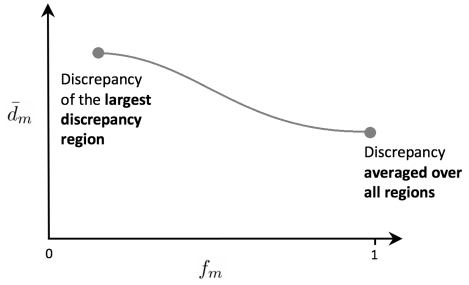


Figure 1.1. Example of a lack-of-fit contrast curve

From the above expressions, it can be deduced that the lack-of-fit curves are by construction decreasing. By way of example, we show a typical pattern of this curve in Fig. 1.1, where the leftmost point on the abscissa-axis provides the fractions of observations that fall into the region with the higher discrepancy, while the rightmost point corresponds to all the observations ($f_m = 1$). Looking at the ordinate-axis, the leftmost point on each curve represents the $\bar{d}_m$ value of the largest discrepancy region

of its corresponding tree; the rightmost point provides the $\bar{d}_m$ value across all regions. Points in between give a $\bar{d}_m$ value over the regions with the highest discrepancy that contain the corresponding fraction of observations [Friedman(2020)].

1.4.2 Estimation Contrast Boosting

As explained above, Contrast Trees can be used as a diagnostic tool to examine the lack of accuracy of predictive models. To improve the model reliability, [Friedman(2020)] proposes a contrast-boosting technique that enables models to provide more accurate predictions, by means of an iterative procedure that reduces uncovered errors and calculates an additive correction. Estimation Contrast Boosting (ECB) gradually modifies a starting value of z using an additive term, reducing its discrepancy with y. The resulting prediction for z is then adjusted accordingly. Please note that in Estimation Contrast Boosting, the two response variables are not equivalent: response y is taken as the reference, while response z is adjusted.

In this section z is supposed to be an estimate of a parameter of the full conditional distribution of a target variable given a set of predictor variables, $p_y(y|x)$, such as e.g. the distribution expected value. The algorithm can be described as follows:

1. A Contrast Tree of y versus z is estimated using discrepancy 1.1. M regions $\{R_m^{(1)}\}_1^M$ are then identified on the input space. The z values within each region are then updated with an additive term $\delta_m^{(1)}$, so that the discrepancy in that region results zero, , i.e. $d_m = 0$ in Eq. 1.3. The updated $z^{(1)}$ values are then:
$$z^{(1)} = z + \delta_m^{(1)} \quad \mathbf{x} \in R_m^{(1)}$$

2. The update predictions $z^{(1)}$ have zero discrepancy in the regions $\{R_m^{(1)}\}_1^M$, but there may be other partitions of input space producing regions where calculated discrepancy is not small. Updated $z^{(1)}$ are then contrasted again with y, in order to produce a second Contrast Tree. A second update $\delta_m^{(2)}$ is then calculated in order to have zero discrepancy in regions identified by this second tree.

3. The procedure is then iterated at most K times in total, until a stopping condition on the magnitude of δ_k updates is met.

4. Following considerations on Gradient Boosting algorithm from [Friedman(2001)], performance accuracy can be improved at the expense of computational cost by introducing a learning parameter α that modulates the update s.t.:
$$z^{(k)} = z^{(k-1)} + \alpha\delta_m^{(k)}$$
where $1 \leq k \leq K$ and $0 < \alpha \leq 1$.

Any point $\mathbf{x}$ in the input space lies within a single region $m_k(\mathbf{x})$ of each tree with its calculated update $\delta_{m_k(\mathbf{x})}^{(k)}$. Given an initial value $z(\mathbf{x})$, the estimate $\hat{z}(\mathbf{x})$ is given by:
$$\hat{z}(\mathbf{x}) = z(\mathbf{x}) + \sum_{k=1}^{K} \delta_{m_k(\mathbf{x})}^{(k)} \tag{1.5}$$

1.4.3 Distribution Contrast Boosting

Distribution Contrast Boosting (DCB) is an application of Contrast Trees that provides an assumption-free method for estimating the full probability distribution of an outcome variable given a set of predictor input variable values ([Friedman(2020)]).

Both z and y are considered observed values for random variables Y and Z generated from respective conditional distributions $p_y(y|\mathbf{x})$ and $p_z(z|\mathbf{x})$. Contrast Trees are used to identify regions where the two distributions differ the most, in order to estimate a transformation $g_{\mathbf{x}}(z)$, different for each value of $\mathbf{x}$ s.t.:

$$p_{g_{\mathbf{x}}}(g_{\mathbf{x}}(z)|\mathbf{x}) = p_y(y|\mathbf{x})$$

So, the estimated function $\hat{g}_{\mathbf{x}}(z)$ can be used to estimate the distribution of Y given a value of $\mathbf{X}$.

The algorithm can be described, with the same symbols a of previous sections, as follows:

1. A Contrast Tree of y and z with discrepancy 1.2 is estimated in order to identify M regions. In each region $R_m^{(1)}$, z values are transformed using:

   $$z^{(1)} = g_m^{(1)}(z) \quad \mathbf{x} \in R_m^{(1)}$$

   s.t. that discrepancy 1.2 is zero within the region. The transformation ensuring this is:

   $$g_m^{(1)}(z) = \hat{F}_y^{-1}\left(\hat{F}_z(z)\right)$$

   with $\hat{F}_y(y)$ being the empirical cumulative distribution of y and $\hat{F}_z(z)$ the corresponding empirical cumulative distribution of z for $\mathbf{x} \in R_m^{(1)}$. This transformation is graphically represented by the quantile-quantile plot of y versus z within $R_m^{(1)}$.

2. The distribution of the updated variable $z^{(1)}$ can be used to produce a Contrast Tree of y and $z^{(1)}$, using the same discrepancy. So, another set of regions $\{R_m^{(2)}\}_1^M$ is produced, where a new transformation $g_m^{(2)}(z)$ can be defined.

3. Such a process can be iterated K times until a stopping condition on discrepancy between y and $z^{(k)}$ is met.

4. As for Estimation Contrast Boosting, using a learning rate α to shrink the k-th transformation towards the identity increases accuracy by increasing the number of transformations:

   $$g_m'^{(k)}(z) = (1-\alpha)z + \alpha g_m^{(k)}(z)$$

The distribution $p_y(y|\mathbf{x})$ can be estimated using a sample $\{z_i\}_1^n$ drawn from the distribution of z. This will lie within $R_{m_1(\mathbf{x})}^{(1)}$ tree region for the first iteration, $R_{m_2(\mathbf{x})}^{(2)}$ for the second iteration and so on. Within each of these regions, a corresponding

transformation function $g^{(k)}_{m_k(\mathbf{x})}$ is defined. Then, a starting value z can be transformed into an estimated value $\hat{y} = \hat{g}_{\mathbf{x}}(z)$ as follows:

$$\hat{g}_{\mathbf{x}}(z) = g^{(K)}_{m_K(\mathbf{x})} \circ g^{(K-1)}_{m_{K-1}(\mathbf{x})} \circ \cdots \circ g^{(1)}_{m_1(\mathbf{x})}(z) \tag{1.6}$$

If the overall estimation of $\hat{g}_{\mathbf{x}}(z)$ can be held accurate, the distribution of the transformed sample $\{\hat{y}_i = \hat{g}_{\mathbf{x}}(z_i)\}_1^n$ can be considered an approximation to that of y, for a given value of $\mathbf{x}$: so, statistics calculated from the values of $\hat{y}$ can be selected as good estimates for the corresponding quantities for distribution $p_y(y|x)$.

Chapter 2

Mortality diagnostic using CTs

Contrast Trees will now be employed in the in the context of mortality models. The response variable, unless otherwise specified, will be the central mortality rate. A common choice (see e.g. [Pitacco(2009)]) for predictor variables in dynamic mortality models, both likelihood-based and machine-learning based, is to use as predictors the age of death, the birth year (cohort) and the calendar year of death. Regions resulting from Contrast Trees will be graphically represented on the Lexis diagram. In this context, the input space generated by the Age and Year predictor variables will be called Lexis space. Given this predictor choice, partition along Lexis space can happen only in three directions: along ages (vertical split), along calendar years (horizontal split) or along cohort (diagonal split).

In the context of mortality modeling diagnostics, the main feature that distinguishes Contrast Trees from the traditional diagnostic methods mentioned in section 1.3 is the ability to automatically identify the regions in which a given model provides a high error for certain combinations of ages and calendar years. Furthermore, Contrast Trees have the advantage of being easy to interpret and can be used as a diagnostic tool to detect the inaccuracies of every kind of model, such as those whose parameters estimate are based on a likelihood function and those based on machine learning algorithms.

This chapter presents two applications of Contrast Trees to mortality models. In sections 2.1, 2.2 and 2.3 CTs are used in the context of mortality description to assess performances from both machine learning and likelihood based models within a single framework. Estimation Contrast Boosting is then applied to predictions in order to test its effect on model performances.

In sections 2.4 and 2.5 the tools provided by Contrast Trees are applied to mortality projections obtained using Generalized Age-Period-Cohort models.

2.1 Method and data

In the following sections, Contrast Trees will be employed to assess the lack-of-fit relative to four models: two pertaining to the extrapolative approach recalled in section 1.1 and two machine-learning-based. Firstly, a model calibration is carried

out for each model, producing estimated values for the central death rate. For each model, a Contrast Tree is produced, contrasting the model-predicted central death rate to the observed data. The response variable, in this case central death rate, and discrepancy is the same for all CTs. In this way, each Contrast Tree checks discrepancy between model predictions and observed data, allowing for a unified comparison.

The analyses are carried out using Italian mortality data available in the Human Mortality Database (HMD) over the period 1950-2018, relative to male population aged 0-90. Calculations are performed separately for age groups 0-29, 30-60, and 61-90 to provide further evidence of the differences in mortality that characterizes the younger ages, the adult ages, and the older ages.

A bit of care must be taken in order to avoid terminological confusion when splitting the data between training and test set : a subset S_1 of the data, containing 70% of the data is used in order to calibrate the models. The remaining 30% of the data, along with models' predictions for central death rates, is used to train and validate Contrast Trees: this fraction of the data is further split in half, to create two subsets S_2 and S_3. These are used to respectively train CTs and produce CT-related quantities, such as LOF-curves. From the perspective of the mortality model, S_1 is the training set, while $S_2 \cup S_3$ is the test set. From the perspective of the Contrast Tree, which is in itself a machine learning method, S_2 is the training set and S_3 is the test set. The training set S_1 is used solely to calibrate mortality models and obtain the parameters' estimate of each model. The estimates are then used to evaluate the out-of-sample performance. Finally, out-of-sample errors are calculated using data from the test set. The dataset is partitioned using the dissimilarity-based compound selection proposed in [Willett(1999)].

Mortality rate

The central death rates $m_{x,t}$ for each age x and year t are calculated according to the following formula:

$$m_{x,t} = \frac{d_{x,t}}{E_{x,t}} \tag{2.1}$$

Where $d_{x,t}$ is the number of deaths aged x in year t, and $E_{x,t}$ are the central exposures-to-risk aged x in year t.

2.2 Models for mortality description

Contrast Trees will now be applied to mortality models. At this stage, such mortality models are used descriptively: their goal is to describe mortality synthetically without considering rates' projections. In this sense, the data can be considered stationary, as the calendar year is treated as any other variable and non-stationarity does not constitute an issue. The purpose of the analysis is to evaluate models' quality of fit against each other in a unified framework provided by Contrast Trees. It should be underlined that, in the context of this thesis, emphasis is not on placed upon the models' accuracy. The main objective here is to analyze CTs application in the context of mortality models and not to produce state-of-the-art models. Moreover, the Contrast

Tree partition algorithm, exposed in section 1.4.1, does not make any assumption about the reliability of the values y_i or z_i or even about the way they are obtained, as long as the data can be expressed in the form $\{\mathbf{x}_i; y_i, z_i\}_{i=1}^{N}$. Should the response values differ from each other, CTs are designed specifically to highlight those differences.

In the following, four models to which the Contrast Tree methodology is applied are briefly reviewed. The first two models used to produce death rates' estimates belong to the family of generalized age-period-cohort (GAPC) that are expressed in a regression framework to be suitable for applying Contrast Trees, which requires data organized in columns. The last two are well-known machine learning techniques also used for regression tasks.

Lee-Carter model

The extension of the Lee-Carter (LC) model[Lee and Carter(1992)] proposed by [Brouhns et al.(2002)] is considered. This assumes that the number of deaths is a Poisson random variable. The Lee-Carter model under the specification of [Brouhns et al.(2002)] describes the logarithm of the central death rate at age x and time t as:

$$\log\left(m_{x,t}\right) = \alpha_x + \beta_x \kappa_t \tag{2.2}$$

The age-specific parameter α_x provides the average age profile of mortality, the age-period term $\beta_x \cdot \kappa_t$ describes the mortality trends, with κ_t the time index and β_x modifying the effect of κ_t across ages. The model is subject to the following constraints on κ_t and β_x: $\sum_t \kappa_t = 0$ and $\sum_x \beta_x = 1$. Future mortality rates are obtained by modeling the time index κ_t through an autoregressive integrated moving average (ARIMA) process. In general, a random walk with drift properly fits the data.

This model can be reformulated into a Generalized Non-linear Model (GNM) framework, as in [Villegas et al.(2018)]. The authors use a GNM and apply the maximum likelihood method to fit the model to historical data. Under this specification, the LC model can be seen as a non-linear regression model where mortality rates are the target variable, predicted using features (age and time) [Richman and Wüthrich(2021)].

Age-Period-Cohort model

The Age-Period-Cohort (APC) model has been reformulated into a Generalized Linear Model (GLM) framework in a work by [Alai and Sherris(2014)]. Their new reformulation assumes that random number of deaths follows a Poisson distribution and adopts a logarithmic link function:

$$\log\left(m_{x,t}\right) = \beta_0 + \beta_{1,x} + \beta_{2,t} + \beta_{3,t-x} \tag{2.3}$$

Where the regression coefficients $\beta_{1,x}$, $\beta_{2,t}$, $\beta_{3,t-x}$ represent respectively the age trend, the period trend and the cohort trend ($t - x$ represents the year of birth).

Gradient Boosting Machine

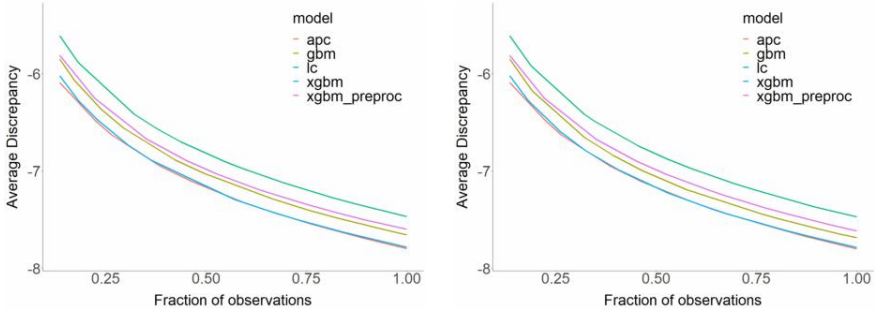
A Gradient Boosting Machine (GBM) is a tree-based algorithm proposed by [Friedman(2001)] that uses fixed-size decision trees as weak learners. The prediction is obtained by a sequential approach, where each decision tree uses the information from the previous one to improve the current fit. Given a current model fit, $F_m(\mathbf{x})$, the algorithm provides a new estimate, $F_{m+1}(\mathbf{x}) = F_m(\mathbf{x}) + h_m(\mathbf{x})$, where $h_m(\mathbf{x})$ is the weak learner fitted on the model residuals $y - F_m(\mathbf{x})$ with y target variable. Hyperparameter calibration has been performed using a manual grid search.

eXtreme Gradient Boosting Machine

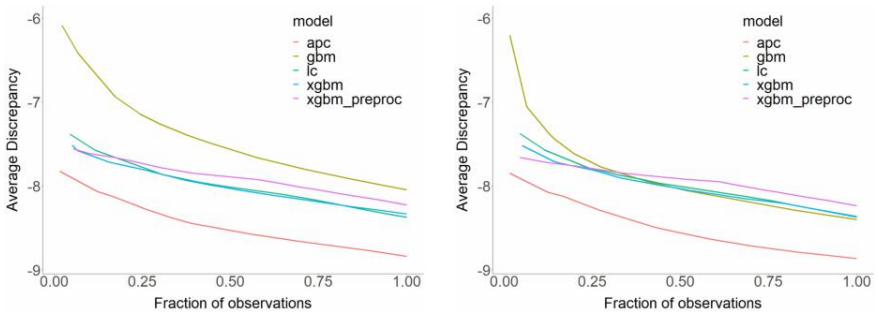
Extreme Gradient Boosting Machine (XGBM) is an efficient implementation of gradient boosting decision trees proposed by [Chen et al.(2015)], and designed to be fast to execute and highly effective. To verify if a simple data preprocessing has some meaningful effect on the quality of models, we apply XGBM to both raw and preprocessed data: the latter is obtained by centering and scaling the raw data using mean and standard deviation. Implementation was performed using R package *caret* ([Kuhn(2020)]). Hyperparameters were chosen using a grid search with a 10-fold cross-validation repeated 5 times.

2.3 Numerical results

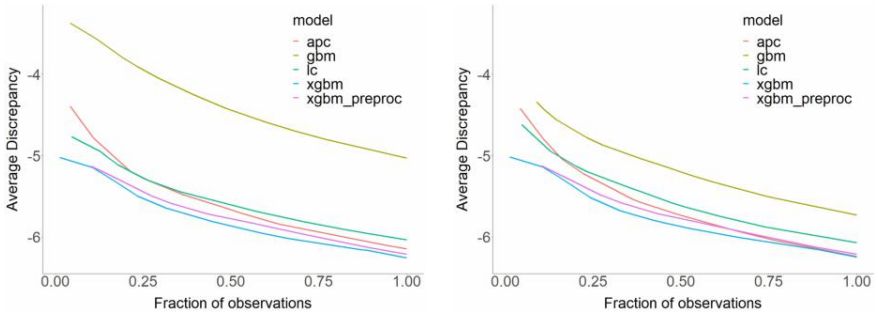
The Contrast Tree analyses have been implemented using the *conTree* R package developed by [Friedman and Narasimhan(2020)]. The minimum number of points falling in a leaf is set to 20, while maximum tree size, corresponding to the maximum number of regions, is set to 100. This choice was made to ensure the production of regions large enough to be interpreted with ease, without limiting the growth of the tree (being the Lexis space composed by 1456 observation for the described data, one can expect at most 72 leaves, therefore the stopping condition for maximum tree size is never met). A manual grid search was performed to assess how tuning these hyperparameters would affect the resulting Contrast Trees, but no noticeable effect was observed.



(a) Base model - Age 0-29 **(b)** Boosted model - Age 0-29



(c) Base model - Age 30-60 **(d)** Boosted model - Age 30-60



(e) Base model - Age 61-90 **(f)** Boosted model - Age 61-90

Figure 2.0. Lack-of-fit contrast curves in the log scale for APC, LC, GBM, XGBM and XGBM_preproc by age group. Left panels: base models; Right panels: boosted models.

Predicted values from each model have been contrasted with the corresponding observed values using discrepancy 1.1, thus producing five Contrast Trees with corresponding input space partitions. Estimation Contrast Boosting has been then performed, using observed values of the central death rate as responses and adjusting central death rates predicted from the models. The maximum number of iterations K has been set to 20, which was deemed sufficient since most of the discrepancy reduction has been seeing happening in the first 10 iterations. Results are summarized in the lack-of-fit contrast curves, shown in Fig. 2.0 for the three different ages groups analyzed. Panels (a), (c) and (e) of these figures refer to the lack-of-fit curves obtained without applying the Estimation Contrast Boosting (Base models), while panels (b), (d) and (f) refer to the lack-of-fit curves obtained after applying Estimation Contrast Boosting to the output of the models (Boosted models).

Lack-of-fit curves relative to both likelihood-based models and machine-learning valuations can be compared directly. For the 0-29 age group (Fig. 2.0, panel (a) and (b)), both APC and XGBM model have the lowest discrepancy values for each fraction of observations, providing the best fitting. The average discrepancy for this age group is higher than for the 30-60 age group. The 0-29 age group is known to be characterized by high accidental mortality, the so-called "accident hump" around age 20-25, due to accidental deaths or suicides caused by increased risk-taking behavior. Mortality at age 0-29 is therefore hard to predict, and Contrast boosting is not able

to actually reduce the average discrepancy.

For the 30-60 age group (Fig. 2.0, panel (c) and (d)), the APC model seems to best perform across all regions since the discrepancy values are consistently lower than those of the other models. For the XGBM models, we can observe that the model applied to preprocessed data (XGBM_preproc) performs better in the regions with the highest average discrepancy with respect to the model applied to raw data. From the scale of the plots, we can see that Contrast boosting reduces discrepancy across almost all regions for the GBM and LC models, where the relative effect of boosting is particularly apparent.

For the 61-90 age group (Fig. 2.0, panel (e) and (f)), the GBM model seems by far the worst performing model. Albeit the application of Contrast boosting significantly reduces the discrepancy, the GBM continues to be less accurate than the other models. It should also be noted that the effect of Contrast boosting in high-discrepancy regions for the other models is negligible, except for the APC.

Model	Age 0-29			Age 30-60			Age 61-90		
	Base	Boosted	% Δ	Base	Boosted	% Δ	Base	Boosted	% Δ
APC	0.000410	0.000409	0%	0.000145	0.000141	-3%	0.002142	0.001948	-9%
LC	0.000571	0.000568	0%	0.000231	0.000232	0%	0.002395	0.002314	-3%
GBM	0.000473	0.000459	-3%	0.000320	0.000225	-30%	0.006525	0.003238	-50%
XGBM	0.000417	0.000415	0%	0.000240	0.000233	-3%	0.001916	0.001940	1%
XGBM_p	0.000500	0.000493	-1%	0.000268	0.000265	-1%	0.002003	0.002005	0%

Table 2.1. Values of the average discrepancy $\bar{d}_m$ calculated on $m_{x,t}$ in the test set.

Table 2.1 reports the values of the average discrepancy measure across all regions for both the base and the boosted models considered in the analysis. The APC and the XGBM base models provide the lowest average discrepancy values (0.000410 and 0.000417, respectively), which remain substantially unchanged after the Contrast boosting procedure. The APC model shows the lowest value of $\bar{d}_m$ also for the age group 30-60, in line with the dynamics of the lack-of-fit curves depicted in panels (b) and (e) of Fig. 2.0. However, the lack-of-fit curves provide more structured information than the average discrepancy, in particular, regarding how and how much $\bar{d}_m$ varies across the input space. For example, for the age group 61-90 in the base model (panel (c)), we can appreciate that the main difference among models (except for GBM, which is out of range) measured by the average discrepancy is caused by the high discrepancy regions (where the fraction of observation is less than about 0.20). For ages 61-90, the GBM base model shows the worst fitting to the observed mortality data. Although Contrast boosting produces a strong improvement in the discrepancy measure, GBM remains the worst model in terms of discrepancy. Contrast boosting is very effective also for the GBM model in the age group 30-60, as it heavily lowers (-30%) the average discrepancy between observed and estimated values.

For a comparison with the average discrepancy, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are also calculated for both the base

Error	Model	Age 0-29			Age 30-60			Age 61-90		
		Base	Boosted	% Δ	Base	Boosted	% Δ	Base	Boosted	% Δ
RMSE	APC	0.002040	0.002039	0%	0.000264	0.000263	0%	0.004260	0.004139	-3%
	LC	0.003471	0.003471	0%	0.000491	0.000496	1%	0.004258	0.004363	2%
	GBM	0.001648	0.001647	0%	0.000640	0.000455	-29%	0.012248	0.005439	-56%
	XGBM	0.001517	0.001515	0%	0.000342	0.000338	-1%	0.003260	0.003278	1%
	XGBM_p	0.001939	0.001935	0%	0.000391	0.000386	-1%	0.003339	0.003345	0%

Table 2.2. Values of the RMSE calculated on $m_{x,t}$ in the test set.

Error	Model	Age 0-29			Age 30-60			Age 61-90		
		Base	Boosted	% Δ	Base	Boosted	% Δ	Base	Boosted	% Δ
MAPE	APC	14.7%	14.5%	-1%	4.5%	4.3%	-3%	3.9%	3.4%	-14%
	LC	14.2%	13.8%	-3%	7.2%	7.1%	-1%	4.9%	4.9%	0%
	GBM	23.4%	18.8%	-20%	13.0%	7.6%	-41%	18.3%	9.2%	-50%
	XGBM	15.9%	15.3%	-3%	6.9%	6.2%	-10%	3.7%	3.8%	2%
	XGBM_p	20.0%	18.2%	-9%	7.3%	7.2%	-1%	3.6%	3.6%	0%

Table 2.3. Values of the MAPE calculated on $m_{x,t}$ in the test set.

model and the boosted one. Given the N estimates of central death rates $\hat{m}_{x,t}$ and their corresponding observed values $m^{obs}_{x,t}$, RMSE and MAPE are calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{x,t}\left(\hat{m}_{x,t} - m^{obs}_{x,t}\right)^2}$$

$$\text{MAPE} = \frac{1}{N}\sum_{x,t}\frac{|m^{obs}_{x,t} - \hat{m}_{x,t}|}{m^{obs}_{x,t}}$$

Intuitively, all three measures $\bar{d}_m$, RMSE, and MAPE quantify the "distance" between the estimates and the actual observations. However, while RMSE and MAPE are commonly used error measures calculated on the overall input space without region partitioning, the average discrepancy is an innovative measure summarizing the discrepancy over all the regions identified by the Contrast Trees.

Model	Age 0-29			Age 30-60			Age 61-90		
	Base	Boosted	% Δ	Base	Boosted	% Δ	Base	Boosted	% Δ
APC	0.149906	0.148218	-1%	0.040837	0.040276	-1%	0.036633	0.035584	-3%
LC	0.151968	0.149051	-2%	0.066676	0.070489	6%	0.042757	0.039114	-9%
GBM	0.292233	0.260784	-11%	0.109899	0.052510	-52%	0.118491	0.052240	-56%
XGBM	0.195720	0.191478	-2%	0.066986	0.062600	-7%	0.036712	0.036779	0%
XGBM_p	0.207129	0.186137	-10%	0.072703	0.072571	0%	0.035729	0.035505	-1%

Table 2.4. Values of the average discrepancy $\bar{d}_m$ calculated on $\log(m_{x,t})$ in the test set.

By comparing Tables 2.2 and 2.3 showing the values of RMSE and MAPE with Table 2.1 reporting the values of the average discrepancy, a greater convergence of the error measures in the boosted models rather than in the base models can be noted . This result is intuitively straightforward since the boosted models are

Error	Model	Age 0-29			Age 30-60			Age 61-90		
		Base	Boosted	% Δ	Base	Boosted	% Δ	Base	Boosted	% Δ
RMSE	APC	0.197670	0.197502	0%	0.062659	0.062708	0%	0.049847	0.049678	0%
	LC	0.237064	0.232791	-2%	0.101946	0.106096	4%	0.060798	0.056322	-7%
	GBM	0.707561	0.694538	-2%	0.182743	0.071172	-61%	0.233815	0.081485	-65%
	XGBM	0.503457	0.496244	-1%	0.089719	0.084808	-5%	0.053659	0.053511	0%
	XGBM_p	0.306517	0.272596	-11%	0.096897	0.097073	0%	0.048344	0.048111	0%

Table 2.5. Values of the RMSE calculated on $\log(m_{x,t})$ in the test set.

Error	Model	Age 0-29			Age 30-60			Age 61-90		
		Base	Boosted	% Δ	Base	Boosted	% Δ	Base	Boosted	% Δ
MAPE	APC	2.1%	2.2%	0%	0.8%	0.8%	0%	1.4%	1.2%	1%
	LC	2.1%	2.1%	-2%	1.2%	1.3%	5%	1.6%	1.6%	1%
	GBM	3.5%	3.1%	-12%	1.9%	0.9%	-52%	4.7%	5.6%	2%
	XGBM	2.6%	2.5%	-1%	1.2%	1.1%	-4%	1.2%	1.3%	1%
	XGBM_p	2.8%	2.7%	-3%	1.2%	1.2%	0%	1.2%	1.2%	1%

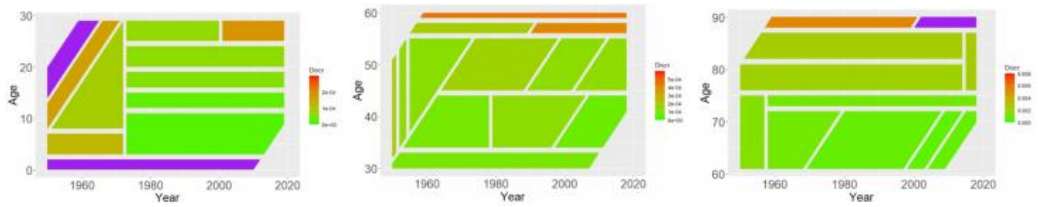
Table 2.6. Values of the MAPE calculated on $\log(m_{x,t})$ in the test set.

obtained by just reducing the discrepancy measure.

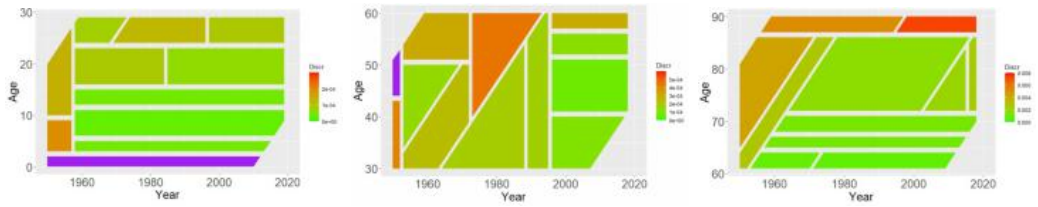
Average discrepancy, RMSE, and MAPE are also calculated for the logarithm of the central death rates (Tables 2.4-2.6). These measures assign a relatively large weight to errors at young ages, while error measures calculated on the central death rates assign a large weight to errors at older ages. Indeed, for the age group 0-29, all the errors reported in Tables 2.4-2.6 are significantly higher than those in Tables 2.1-2.3. The errors calculated on the logarithm of the central death rates highlight the ability of Contrast boosting to reduce the inaccuracy of GBM and XGBM_preproc in fitting observed mortality at ages 0-29.

The most interesting feature of the application of Contrast Trees to the field of mortality estimates is the automatic identification of the regions of the predictors' space where a given model provides high discrepancy values for certain combinations of ages-years obtained by comparing the model estimates with the observed mortality rates. These regions can be easily detected and possibly interpreted, providing a further explanation of the model performances as well as helping to assess whether a model can be reliable or not. Moreover, they are a direct result of the Contrast Tree machine learning procedure and do not depend on a subjective judgment, unlike e.g. graphical residuals' analysis for Generalized Linear Models. Fig. 2.1 and Fig. 2.2 show the heatmap of all the error regions for the base model and the boosted one, respectively. Low discrepancy regions are painted in green, while high discrepancy regions are painted in red. For the sake of plot readability, the regions presenting a discrepancy value exceeding 3e-04, 6e-04, and 0.008 for the age groups 0-29, 30-60, and 61-90, respectively, are colored in purple. Again, all quantities produced by the Contrast Trees (in this case, discrepancies and region boundaries) are directly comparable, thus allowing for an easy assessment of model performances, regardless of any of the models' characteristics.

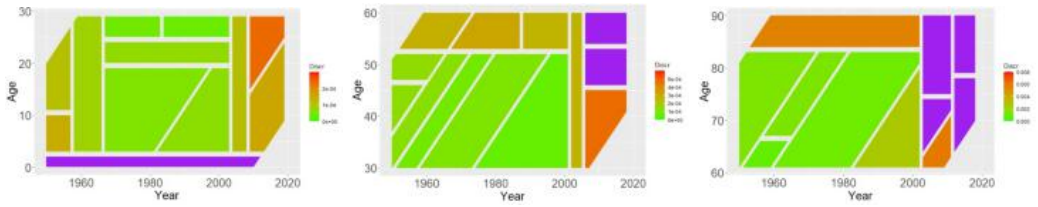
The regions' width and shape change from model to model. Some regions show remarkable mortality estimation errors in specific age groups, others in specific intervals of years, others in a specific range of cohorts. All the models considered show high discrepancy values in the first year of age (Fig. 2.1, age group 0-29, left panels), confirming the difficulty of adequately estimating the mortality of newborns. This situation remains unchanged after the application of Contrast boosting, which, in this case, seems to be not effective (Fig. 2.2, age group 0-29, left panels). For the age group 30-60 in the base model (Fig. 2.1, central panels), the two XGBM models show high discrepancy values after age 45-46, while GBM in the years 2000-2018. The LC model instead evidences high errors in estimating the mortality of cohorts born between 1920 and 1932. Considering the 61-90 age group (Fig. 2.1, right panels), it can be noticed that the GBM model continues to fail in estimating mortality rates in the years 2000-2018, while the LC model (and also APC) mortality rates in the cohorts born between 1920 and 1932. By comparing the results for the base models (Fig. 2.1) with those for the boosted ones (Fig. 2.2), a clear effect of boosting is shown on the GBM model for the 30-60 and 61-90 age groups and the XGBM for the 30-60 age group.



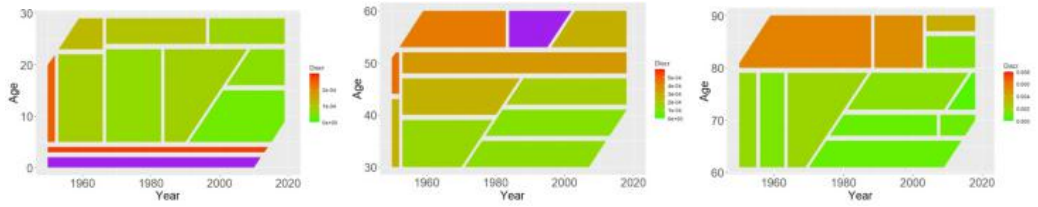
(a) APC (left: age 0-29; centre: age 30-60 right: age 61-90)



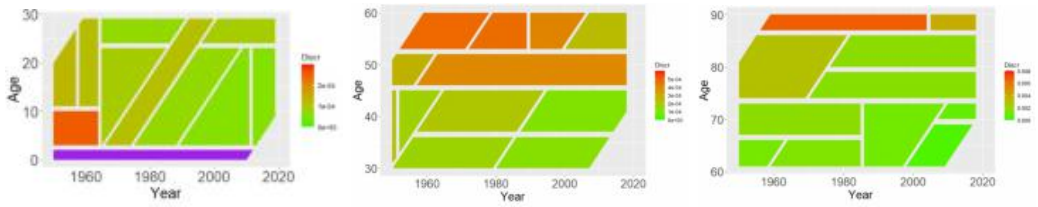
(b) LC (left: age 0-29; centre: age 30-60 right: age 61-90)



(c) GBM (left: age 0-29; centre: age 30-60 right: age 61-90)

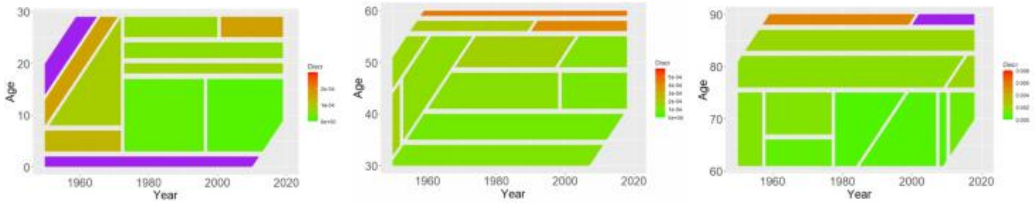


(d) XGBM (left: age 0-29; centre: age 30-60 right: age 61-90)

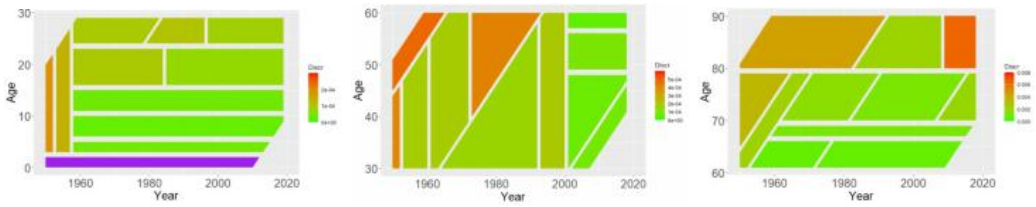


(e) XGBM_preproc (left: age 0-29; centre: age 30-60 right: age 61-90)

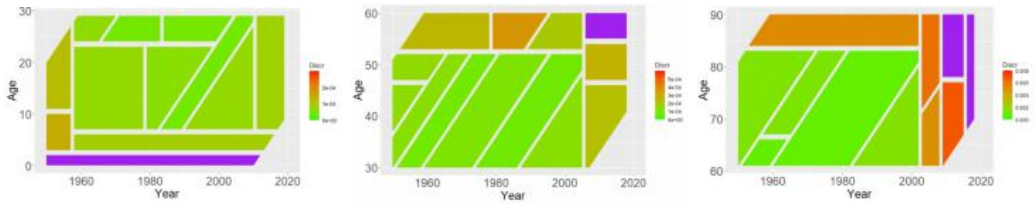
Figure 2.1. Contrast Tree regions, Base model. Years 1950-2018. Regions presenting a discrepancy value exceeding 3e-04 (age 0-29), 6e-04 (age 30-60), and 0.008 (61-90) are colored in purple.



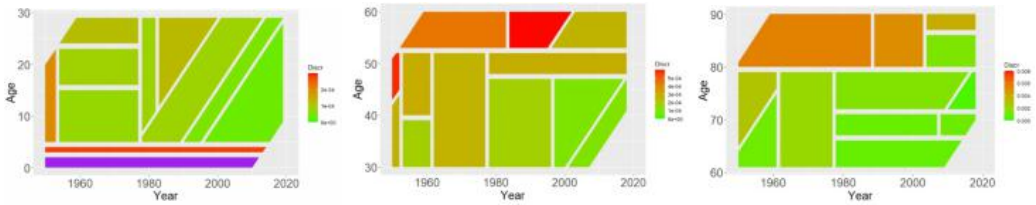
(a) APC (left: age 0-29; centre: age 30-60 right: age 61-90)



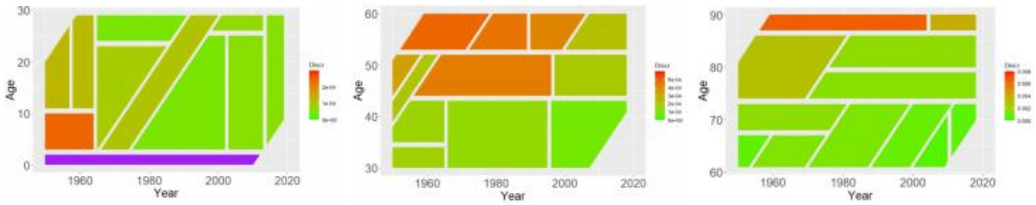
(b) LC (left: age 0-29; centre: age 30-60 right: age 61-90)



(c) GBM (left: age 0-29; centre: age 30-60 right: age 61-90)



(d) XGBM (left: age 0-29; centre: age 30-60 right: age 61-90)



(e) XGBM_preproc (left: age 0-29; centre: age 30-60 right: age 61-90)

Figure 2.2. Contrast Tree regions, Boosted model. Years 1950-2018. Regions presenting a discrepancy value exceeding 3e-04 (age 0-29), 6e-04 (age 30-60), and 0.008 (61-90) are colored in purple.

2.4 GAPC mortality models

Several works tried to find similarities between stochastic mortality models pertaining to the extrapolative paradigm in section 1.1. For example [Hunt and Blake(2015)] describe a model structure based on Age-Period-Cohort regressors that includes a majority of stochastic mortality models. A contribution from [Currie(2014)] shows that many widely used mortality projection models can be described using standard terminology for linear and non-linear models.

In a work from [Villegas et al.(2015)] Generalised Age-Period-Cohort (GAPC) stochastic mortality models are defined, similarly to Generalised Linear Models from [McCullagh and Nelder(1989)].

Let $D_{x,t}$ indicate the random number of deaths in a certain population for individuals aged x in calendar year t; $E^0_{x,t}$ and $E^C_{x,t}$ indicate respectively the corresponding initial and central numbers of exposed to risk. A GAPC mortality model is described by its four components:

1. *Random component*: it's assumed that random number of deaths $D_{x,t}$ follows a Poisson or Binomial distributions, s.t.:

$$D_{x,t} \sim \text{Pois}\left(E^C_{x,t}\mu_{x,t}\right) \quad \text{or} \quad D_{x,t} \sim \text{Bin}\left(E^0_{x,t}; q_{x,t}\right)$$

2. *Systematic component*: the effects of Age x, Period t, and Cohort $c = t - x$ variables are described through a predictor $\eta_{x,t}$:

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta^{(i)}_x \kappa^{(i)_t} + \beta^{(0)}_x \gamma_{t-x}$$

   where α_x is a function of the age describing the general shape of mortality, the N age-period terms are composed of a time index $\kappa^{(i)_t}$ and a modulating factor $\beta^{(i)}_x$ function of age and γ_{t-x}, also modulated by a function of age, describes the cohort effect.

   In the GAPC model family it is assumed that both period indexes and cohort index are stochastic processes, in order to allow for mortality projection, while the age-modulating terms can either be specified functions of age or non-parametric terms.

3. *Link function*: associates random component and systematic component as follows:

$$g\left[\mathbb{E}\left(\frac{D_{x,t}}{E_{x,t}}\right)\right] = \eta_{x,t}$$

   It is often advantageous to choose the canonical link, thus using a logarithm function in case of poissonian random component and a logit link in the case of a binomial random component.

4. *Parameter constraints*: such constraints are necessary in most of stochastic mortality models in order to ensure identifiability and therefore unique parameter estimates. In the context of GAPC mortality models, they are expressed

in terms of a constraint function that maps a vector of parameters θ into a transformed vector of parameters $\tilde{\theta}$ s.t. using θ or $\tilde{\theta}$ produces the same predictor $\eta_{x,t}$:

$$\theta = \left(\alpha_x, \gamma_{t-x}, \beta_x^{(0)}, \dots \beta_x^{(N)}, \kappa_t^{(1)}, \dots, \kappa_t^{(N)}\right) \mapsto \tilde{\theta} = \left(\tilde{\alpha}_x, \tilde{\gamma}_{t-x}, \tilde{\beta}_x^{(0)}, \dots \tilde{\beta}_x^{(N)}, \tilde{\kappa}_t^{(1)}, \dots, \tilde{\kappa}_t^{(N)}\right)$$

The parameter vector θ can be estimated via log-likelihood maximization, given, in case of Poisson death distribution, by:

$$\mathcal{L}(dx, t, \hat{d}_{x,t}) = \sum_{x,t} \omega_{x,t} \left(d_{x,t} \log \hat{d}_{x,t} - \hat{d}_{x,t} - \log d_{x,t}!\right)$$

while, in case of binomial death distribution, by:

$$\mathcal{L}(dx, t, \hat{d}_{x,t}) = \sum_{x,t} \omega_{x,t} \left[d_{x,t} \log \frac{\hat{d}_{x,t}}{E^0_{x,t}} + (E^0_{x,t} - d_{x,t}) \log \left(\frac{E^0_{x,t} - \hat{d}_{x,t}}{E^0_{x,t}} \right) + \binom{E^0_{x,t}}{d_{x,t}} \right] \tag{2.4}$$

where $\omega_{x,t}$ is a 0-1 valued weight used to include or omit data cells.

Similarly to section 2.2, six widely used stochastic mortality models will be now briefly reviewed within the framework of GAPC models, describing random and systematic components for each one.

Lee-Carter model

In the Lee-Carter (LC) projection model reformulation by [Brouhns et al.(2002)], the random number of deaths is assumed to follow a Poisson distribution, with a log link function. The predictor is:

$$\eta_{x,t} = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)}$$

As in the original Lee-Carter formulation, index $\kappa_t^{(1)}$is modeled and forecasted using ARIMA processes.

Renshaw-Haberman model

Renshaw-Haberman (RH) model is proposal for a generalization of the Lee-Carter model which include a cohort effect ([Renshaw and Haberman(2006)]). A Poisson distribution of deaths is assumed, alongside a log link function. The predictor is:

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

Mortality projections make use of ARIMA processes, assuming independence between period and cohort effect.

Age-Period-Cohort model

Age-Period-Cohort (APC) model model was introduced as a sub-model to the RH model ([Currie(2006)]), where $\beta_x^{(1)} = 1$ and $\beta_x^{(0)} = 1$. The predictor is thus:

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

Cairns-Blake-Dowd model

Cairns-Blake-Dowd (CBD) model was proposed with pre-specified age modulating parameters $\beta_x^{(1)} = 1$ and $\beta_x^{(2)} = x - \bar{x}$, with $\bar{x}$ being the average age in the data. No static age function nor cohort effect are considered. The linear predictor is then:

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

A binomial distribution for deaths and subsequently a logit link function are assumed ([Haberman and Renshaw(2009)]). The mortality projections are obtained using a bivariate random walk with drift for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ ([Cairns et al.(2006)])

M7 model

Cairns-Blake-Dowd model was extended ([Cairns et al.(2009)]) by adding a cohort effect and a quadratic age effect to the predictor:

$$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \left[(x - \bar{x})^2 - \bar{\sigma}^2\right]\kappa_t^{(3)} + \gamma_{t-x}$$

Plat model

This model ([Plat(2009)]) combines elements of CBD and LC models. A poissonian number of death with a log link function is assumed. The predictor is given by:

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

It has three age-period terms, with the third being modulated by the positive part of $x - \bar{x}$. If only older ages are of interest, it is suggested to drop the third term from the predictor, obtaining:

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

2.5 Assessment of GAPC projected mortality models

In this section, Contrast Trees will be used to confront and assess performances of mortality projections obtained from the models recalled in section 2.4. For each model, an unweighted Contrast Tree will be trained, contrasting the projected value of the central death rate for each model with its observed value, similarly to section 2.3.

Data

Again, the analysis will be performed on Italian male population data from HMD aged 30-90, without separating age groups. Data will be split into training and test set according to a temporal criterion: data from calendar years 1950-1999 will be used for mortality models calibration while data referring to 2000-2019 period will be used for model comparison. In turn, as explained in section 2.1, solely for Contrast Tree estimation and evaluation, test set data will be split in half using maximum dissimilarity approach.

Methodology

The analysis has been conducted as follows: data from time period 1950-1999 were used to calibrate six GAPC mortality models (LC, RH, APC, CBD, M7, reduced Plat), in order to obtain projected central death rates relative to the period 2000-2019. The implementation was performed using R package *StMoMo*.

These projected rates have then been compared to those observed for the same period of time, by means of RMSE, MAPE and CTs. As in section 2.1, six Contrast Trees were produced, each one contrasting the projected valued for central death rates with corresponding observed values. The analysis was carried out using discrepancy 1.1, both for linear and logarithmic scale. As before, maximum number of leaves was set to 100, while minimum number of points in a leaf was fixed to 20. A second discrepancy, based on Kullbach-Leibler divergence, was then introduced and used to build other Contrast Trees where possible.

As in previous sections, choosing the same response variable and discrepancy allows for a straightforward assessment of models' performances. While all GAPC models, as apparent from equation 2.4, belong to the likelihood-based mortality models' paradigm, they are not expressed in terms of the same response variable: CBD and M7 models assume a binomial distribution with logit link function, therefore use death probabilities $q_{x,t}$ as response. By assuming, as usual in actuarial applications, a constant force of mortality (see e.g. chapter 2 from [Pitacco(2009)]), central death rates $m_{x,t}$ can be obtained as follows:

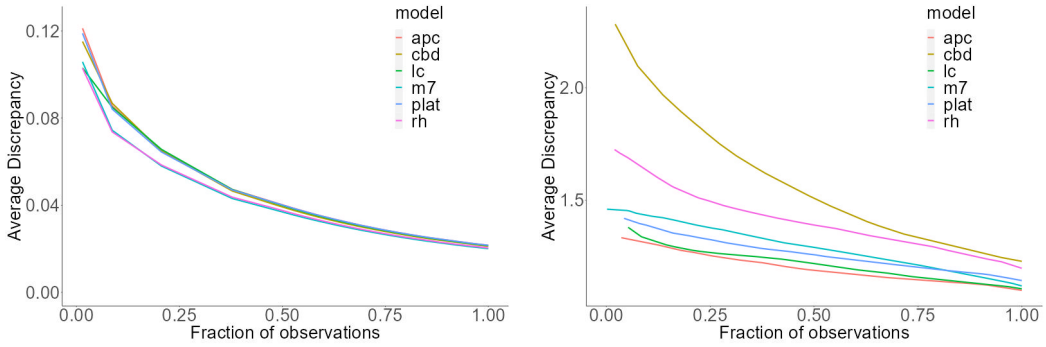
$$m_{x,t} = -\log(1 - q_{x,t})$$

Central death rates obtained from this transformation for CBD and M7 models can subsequently be employed as response variable in a Contrast Tree.

Numerical results

Figure 2.3 reports lack-of fit curves for both scales. It can be seen that differences between models linear scale are not so apparent, while applying Contrast Trees on a logarithmic scale does a better job at differentiating model performances. It's interesting to observe that CBD model seems to have bad performances both in linear and logarithmic scale in terms of RMSE and MAPE, but not in terms of average discrepancy. On the other hand, APC and LC model perform quite well on logarithmic scale but are among the least-performing models in linear scale. Conversely, Renshaw-Haberman model has the better performance in linear scale and, at the same time, the second-worst in logarithmic scale.

These observation are consistent with performance statistics reported in table 2.7. For linear scale Contrast Trees, all indicators result quite similar, without a clear indication of a best-performing model. For logarithmic scale, instead, Cairns-Blake-Dowd model is consistently pointed out as the worst-performing, while Lee-Carter and Renshaw-Haberman models, as expected from figure 2.3, seems to project central death rates closer to observed ones.



(a) Linear **(b)** Logarithmic

Figure 2.3. Lack-of-fit curves from Contrast Trees for the GAPC models used

(a) Linear scale

Model	RMSE	MAPE	Average discrepancy
LC	0,03316	0,67198	0,02144
RH	0,02975	0,69585	0,02040
APC	0,03382	0,65997	0,02160
CBD	0,03400	0,70701	0,01633
M7	0,02974	0,67287	0,01418
Plat	0,03346	0,67461	0,02161

(b) Logarithmic scale

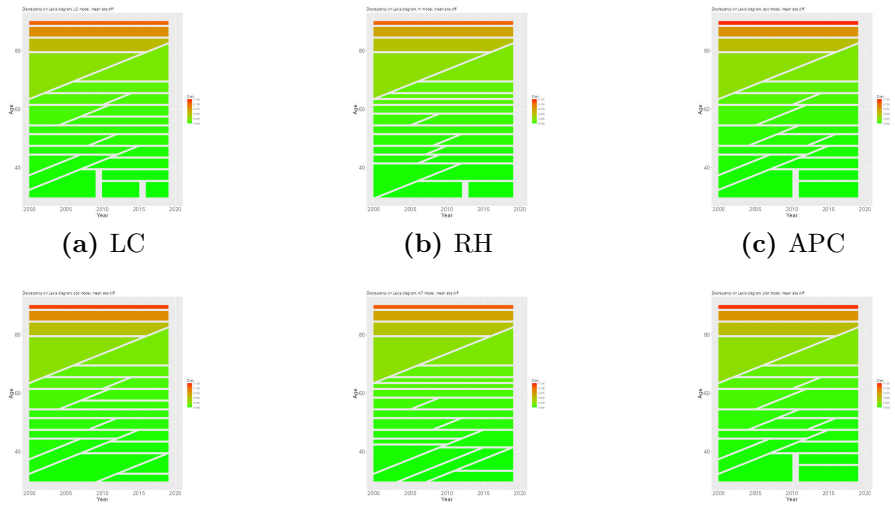
Method	RMSE	MAPE	Average discrepancy
LC	1,13610	0,26675	0,01583
RH	1,26654	0,27916	0,01422
APC	1,10150	0,26364	0,01590
CBD	1,33658	0,28919	0,01633
M7	1,17185	0,26457	0,01418
Plat	1,15157	0,27172	0,01582

Table 2.7. Performance statistics for GAPC models.

Discrepancy regions on Lexis space are reported in figures 2.4 and 2.5 for linear end logarithmic scale respectively. As can be seen in figure 2.4, all GAPC models considered produce very similar regions in linear scale, essentially agreeing on the lacking projection capability in elder ages for the GAPC used for mortality projection. In figure 2.5, the logarithm tends to penalize differences in the neighborhood of $m_x = 1$, thus indicating, quite surprisingly, high discrepancy regions for younger ages: this is especially apparent for Cairns-Blake-Dowd model. Here, it can be pointed out that, while lack-of-fit curves are useful to assess the behavior of a model throughout the whole input space, the regions originating similar x-axis coordinates are not necessarily near to each other in any sense (see e.g. regions for CBD and M7 in figure 2.5).

Discrepancy based on Kullbach-Leibler Divergence

While utilizing Contrast Trees to confront mortality models can provide a simple and unified framework, discrepancy 1.1 has no statistical meaning, simply accounting for the absolute difference between predicted central death rates and observed ones. Kullbach-Leibler divergence (also known as relative entropy) is a statistical distance that measures how a probability distribution of a random variable of interest I is different from the distribution of a reference random variable R



(a) LC **(b)** RH **(c)** APC

(d) CBD **(e)** M7 **(f)** Plat

Figure 2.4. Regions in the Lexis space for Contrast Trees in linear scale.

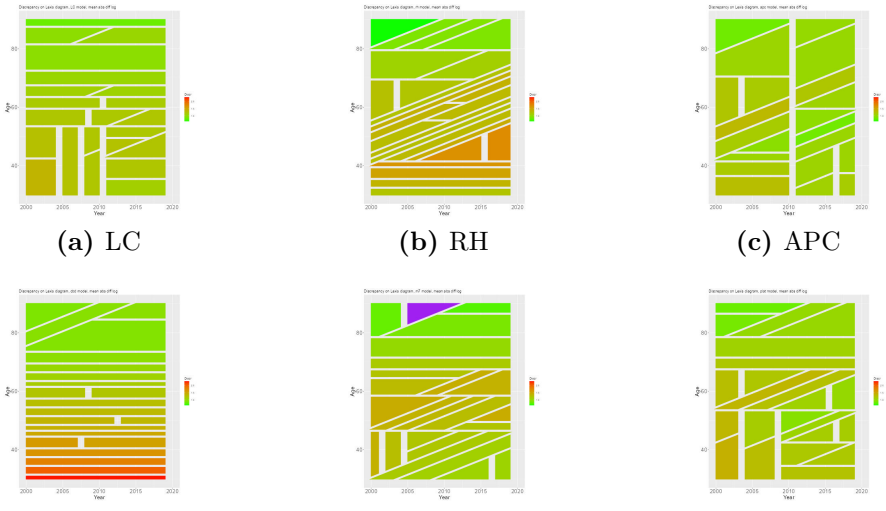
([Kullback and Leibler(1951)]). It seems then natural, for diagnostics applications, to build a discrepancy measure based on Kullback-Leibler divergence in order to build Contrast Trees based on it.

In the case of two discrete probability distributions $p^I(x)$ and $q^R(x)$ on the same sample space χ, the Kullbach-Leibler (KL) divergence is defined as:

$$D_{KL}(I||R) = \sum_{x \in \chi} p^I(x) \log \left(\frac{p^I(x)}{q^R(x)} \right) \tag{2.5}$$

The meaning of this quantity can be better understood using the language of statistical hypothesis testing. Let H_I be the hypothesis that a certain observed value x has been drawn from probability distribution $p^I(x)$ and H_R the hypothesis that it has been drawn from probability distribution $p^R(x)$. Then the logarithm of likelihood ratio, $\log \left(\frac{p^I(x)}{q^R(x)} \right)$ can be thought as the information in x for discriminating in favor of H_I against H_R. So, Kullbach-Leibler divergence represents the expected information per observation for discriminating in favor of H_I against H_R (see [Kullback(1959)] for more details).

Assuming that the number of deaths for Italian male population is drawn from a Poisson distribution with expected value $\lambda_R = E^C_{x,t} \hat{m}^{\text{HMD}}_{x,t}$, where $\hat{m}^{\text{HMD}}_{x,t}$ is the value for central death rate resulting from HMD data and that the projected central death rate for model k, $\hat{m}^{(k)}_{x,t}$, specifies a Poisson distribution with expected value



(a) LC **(b)** RH **(c)** APC

(d) CBD **(e)** M7 **(f)** Plat

Figure 2.5. Regions in the Lexis space for Contrast Trees in logarithmic scale.

$\lambda_I = E^C_{x,t}\hat{m}^{(k)}_{x,t}$, KL divergence can be calculated in closed form.

$$\begin{aligned} D_{KL}(I||R) &= \sum_{x\in\mathbb{N}} P^{\lambda_I}(x) \log\left(\frac{P^{\lambda_I}(x)}{P^{\lambda_R}(x)}\right) \\ &= \sum_{x\in\mathbb{N}} P^{\lambda_I}(x) \left[x \log\left(\frac{\lambda_I}{\lambda_R}\right) + \lambda_R - \lambda_I\right] \\ &= \lambda_I \log\left(\frac{\lambda_I}{\lambda_R}\right) + \lambda_R - \lambda_I \end{aligned}$$

So, a Kullbach-Leibler divergence based discrepancy can be defined as follows:

$$d^{[\mathrm{KL}]}_m = \frac{1}{N_m} \sum_{x_i\in R_m} D_{KL}(I||R) \tag{2.6}$$

This discrepancy is asymmetric, as the divergence it derives from, since $D_{KL}(I||R) \neq D_{KL}(R||I)$. This is not a critical point when defining a divergence, but requires attention from an interpretative point of view. When Contrast Trees were applied above for diagnostic applications y and z responses were perfectly equivalent, since discrepancy 1.1 is symmetrical, but now this property does no hold anymore. Nevertheless, in diagnostic applications, it is natural to impose that the variable representing observed values should be the reference one, while the quantity of interest should be represented by the model predictions. Moreover, as for discrepancy 1.2, support for the distribution of I and R is the same by construction.

Four Contrast Trees have been estimated using the same data and input structure as before, using discrepancy 2.6: in order to be consistent with Poisson assumption, only LC, RH, APC and Plat models have been compared.

The average discrepancy resulting for the different trees is reported in table 2.8, while the respective lack-of-fit curves are reported in figure 2.6. Renshaw-Haberman

Model	Average discrepancy
LC	281.95
RH	273.45
APC	285.32
Plat	292.11

Table 2.8. Average discrepancy for trees with KL-based discrepancy

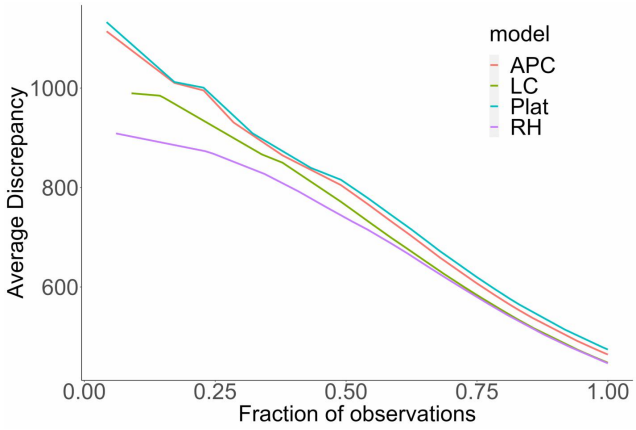
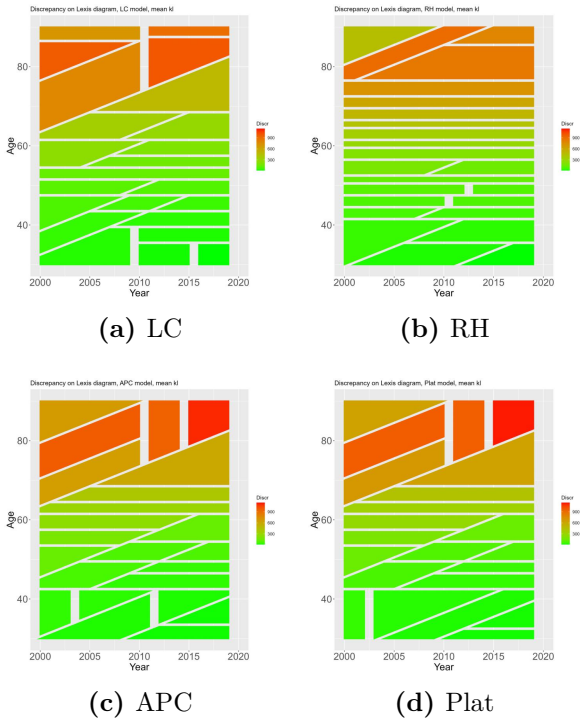


Figure 2.6. Lack-of-fit curve for trees with KL-based discrepancy

model seems to be the best performing one across all input space, in accordance with results for linear scale. Kullbach-Leibler based divergence, however, has a statistical meaning, discriminating between probability distributions, while in the previous analysis just the mean absolute differences of central death rates (or of their logarithms) were investigated.

The regions identified by the tree in the Lexis space denote a worse performance of all four models as age gets higher. This result is again in accordance with what was found in linear scale and in contrast with results for logarithmic scale.



(a) LC **(b)** RH

(c) APC **(d)** Plat

Figure 2.7. Regions in the Lexis space for Contrast Trees with Kullbach-Leibler based discrepancy.

2.6 Discussion

Evaluating, and thus eventually improving, the fit of mortality models is crucial for both demographers and actuaries. Indeed, in particular situations, common in actuarial practice, data quality can turn the mortality estimate difficult. A prime example is the case of small subpopulations where a common method such as the Lee-Carter may not guarantee a reliable estimation. In mortality modeling, the objective of diagnostic checking is to ascertain whether the model fits the historical data by obeying an underlying probabilistic hypothesis. This procedure is carried out using residuals diagnosis checking with a Gaussian or more often a Poisson assumption (see, e.g., [Renshaw and Haberman(2006)]).

Contrast Trees consist of a general method based on machine learning that can be applied to any model, expressed as a regression model, to evaluate the goodness of fit and identify the worst-performing regions in the input space. The main characteristic that discriminates this method from traditional diagnostic tools is the automatic identification of the regions in input space, as seen in section 2.3. Well-known diagnostic tools often used in the literature to assess the goodness-of-fit of a mortality model, such as BIC and AIC, require the definition of a likelihood function, which is not available for machine learning models. Therefore, Contrast Trees provide a unified approach for assessing and comparing the accuracy of traditional mortality models with machine learning algorithms.

In Contrast Trees, the detection of the regions in which a model worse performs can be considered an evolution of the standard analysis on residuals, in which the detection of the highest residuals is typically assigned to graphical analyses using heatmaps and scatter plots [Cairns et al.(2009), Villegas et al.(2018)], and of summary measures like RMSE and MAPE calculated on the overall input space. Conversely, the decision tree structure of Contrast Trees enables the quantification of discrepancy between estimates provided by a model and the actual observations in each region identified by the Contrast Tree itself.

In addition, a Contrast Tree based approach can be applied also when assessing quality of stochastic mortality models' projections: in this case local results provided by the procedure allow for a deeper understanding of the model performances. It must be noted, however, that results are quite heavily influenced both by choice of discrepancy measure and of scale (in the case presented, linear or logarithmic): thus the choice of a discrepancy fit for the problem at hand is particularly important. Moreover, the applications of Contrast Trees in the context of mortality models have been limited to Italian data: this should suggest caution when generalizing the results. The analyses could have benefitted from an extension of the data e.g. to Italian female population or to other European countries, in particular when assessing sensitivity of the results to the choice of hyperparameters.

Chapter 3

Estimation boosting in the context of mortality projection

3.1 Mortality projection for small populations

In actuarial science and demography, improving the accuracy of mortality projections is often of great importance in order to predict death rates for selected groups, such as elderly or regional populations, and to construct policyholders' death tables for an insurance company. In small populations, mortality rates tend to be characterized by high volatility, scarce availability in temporal terms and missing data.

Therefore, as found by [Booth et al.(2006)], identifying underlying trends in mortality patterns can be difficult even using long-established models, such as the Lee-Carter projection model. Furthermore, due to the short-period availability of data, projections are quite uncertain and sensitive to the fitting period: plain extrapolation of historic trends could produce questionable results and unrealistic death tables [Jarner and Kryger (2011)]. In such cases, standard mortality models may not be relevant and the resulting mortality rates could be unreliable.

The issue of mortality models for small populations has been the subject of several works, mostly based on the widely known observation that socially and economically similar populations can be jointly modeled, thus borrowing information from the larger population in order to overcome the limitations of small-populations data. Many different approaches may be taken when dealing with small demographic data, based e.g. on an optimum mixture of mortality data from other populations, such as in [Ahcan at al.(2014)]. A comparative study of mortality models for two populations has been developed by [Villegas et al.(2017)]: here the larger and smaller population are respectively named "reference" and "book" populations. Of particular interest are works from [Jarner and Kryger (2011)], [Wan and Bertschi(2015)] and [Menzietti et al.(2019)], who use a two-step method in which they initially model the mortality of the reference population and then estimate the parameter of mortality spread between the book and reference populations.

In the following sections this last approach will be adapted in order to make use of estimation boosting, starting from a standard mortality model for the book

population and then boosting such mortality with the target variable taken as the mortality of the book population. Data and methodology are described in section 3.2.1 , while numerical results are be presented and discussed in section 3.2.4.

3.2 Estimation Contrast Boosting for mortality spread modeling

3.2.1 Methodology

The estimation boosting procedure described in section 1.4.2 can be utilized to build mortality tables for small populations following the two-step approach briefly outlined in the previous section.

Let $Q^R_{X,t}$ and $Q^B_{X,t}$ be some quantity describing mortality in terms of a predictor matrix X and calendar year t, respectively for reference and book populations, available for observation years $1, \dots, T$. These quantities could be, for example, the central death rates, the force of mortality or the death probabilities, while variables age, period and cohort could be used to build matrix X. Estimation Contrast Boosting (ECB) can be used in conjunction with some projection technique, such as a well-established model like, e.g., the Lee-Carter model. The boosting procedure can be applied in order to reduce discrepancy between $Q^R_{X,t}$ and $Q^B_{X,t}$, thus estimating the value of an additive update $\delta_{X,(1,\dots,T)}$. Using notation from section 1.4.2 this means:

$$y = Q^B_{X,t} \quad z = Q^R_{X,t}$$

A shift in the target of the Estimation Contrast Boosting procedure must be emphasized: in [Friedman(2020)] the updates $\delta^{(k)}_m$ were just instrumental to reduce discrepancy between y and z, with the purpose of producing an updated estimate $\hat{z}$. Here the objective is to obtain the updates $\delta^{(k)}_m$ themselves, so that, alongside some additional assumptions, they can be used to transform the projected quantity of interest for the reference population into that relative to the book population.

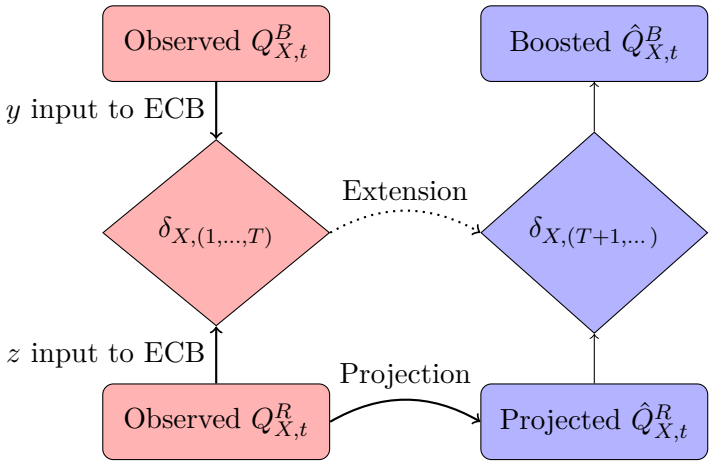


Figure 3.1. The Contrast Tree boosting process outlined in section 3.2.1. Quantities in red refer to calendar years $1, \dots, T$, while those in blue refer to calendar years $T+1, \dots$

On the other hand, the quantity of interest for the reference population $Q^R_{X,t}$ can be projected to calendar years $T+1,\dots$ using. This is feasible since the reference population is large enough to allow the use of standard mortality projection models. Thus, projected estimates $\hat{Q}^R_{X,t}$ for the quantity of interest, relative to reference population and (future) calendar years from $T+1$, can be obtained. Some assumptions on how to extend the updates $\delta_{X,(1,\dots,T)}$ to future calendar years $T+1,\dots$ must then be made. The extended updates $\delta_{X,(T+1,\dots)}$ can then be applied to projected estimates for the reference populations in order to obtain boosted estimates $\hat{Q}^B_{X,t}$ for the book population relative to calendar years $T+1,\dots$.

Note that is not requested to book population to be similar to reference population in terms of geographical, historical or socio-economical features. It can be expected, however, that the Estimation Contrast Boosting will converge faster when considering populations which share some kind of similarity.

3.2.2 Reference and book populations

Data used for numerical evaluation are taken from the Human Mortality Database. French and Italian male populations have been considered as reference populations. For each of these, three book male populations are considered, respectively Belgium, Luxembourg, Lithuania and Austria, Slovenia, Lithuania. In both cases the first two book populations can be considered close, at least in geographical terms, to the respective reference population. On the contrary, the Lithuanian population has been selected in order to investigate the behavior of the Estimation Contrast Boosting procedure when book and reference populations differ substantially. .

Country	Start Year	Notes
Austria	1947	
Belgium	1841	Starting 1 January 2010, the resident population includes asylum seekers.
Lithuania	1959	Detailed data (e.g., by single years of age) never published during the Soviet period (except for census years)
Luxembourg	1960	
Slovenia	1983	Ten Days War resulted in relatively few civilian and military deaths

Table 3.1. Data availability for book populations

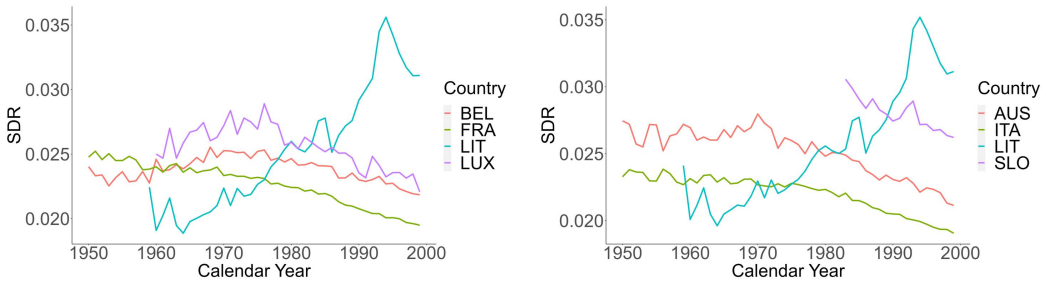
During the aforementioned periods of time, none of the considered population underwent any important change to its composition, such as one caused by a change in borders: however, it must be noted that both Lithuania and Slovenia were affected by considerable political changes in 1990-1991.

In order to have at least a qualitative approach for evaluate the similarity between the mortality of the considered two groups of populations, and following [Menzietti et al.(2019)], direct standardization is performed. The standard population age structure is given, respectively, by the two reference populations (see chapter

4 in [Keyfitz and Caswell(2005)] for more details). Standardized Death Rates (SDR) for country I, ages ranging from a_s to a_e, relative to reference population R are defined as:

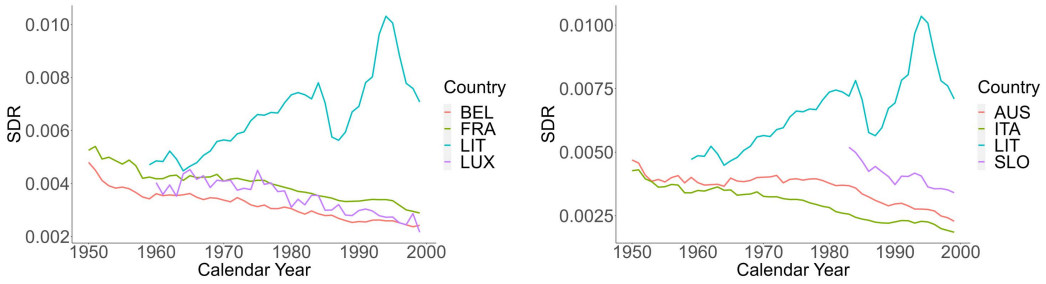
$$\text{SDR}^I_{t,a_s,a_e} = \frac{\sum_{x=a_s}^{a_e} E^R_{x,t} m^I_{x,t}}{\sum_{x=a_s}^{a_e} E^R_{x,t}}$$

The calculation of such standardized rates are performed on the time period 1950-1999 in two steps: initially for all the ages taken into account and then separately for age groups 30-50, 50-70 and 70-90: results are represented in figures from 3.2 to 3.5.



(a) FRA **(b)** ITA

Figure 3.2. SDR for ages 30-90

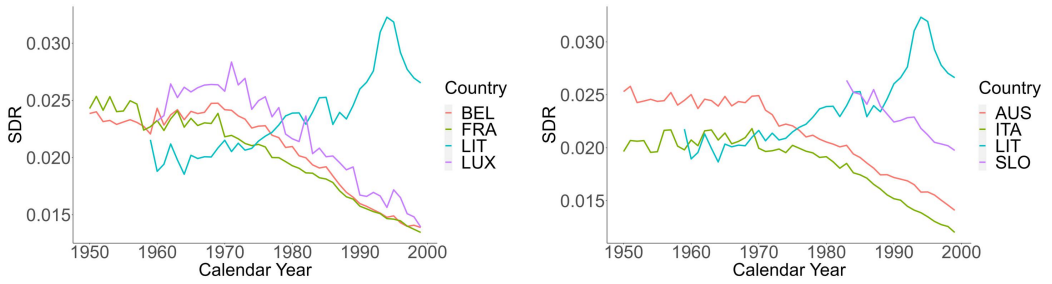


(a) FRA **(b)** ITA

Figure 3.3. SDR for ages 30-50

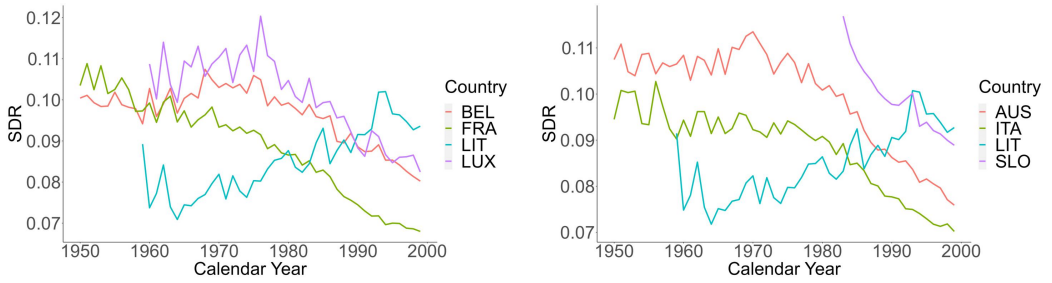
It can be observed that French and Belgian mortality are overall quite close and present the same trend, as can be seen in both separate and joint age groups. Luxembourger mortality presents a more irregular pattern in respect to French one, but it can be said that the trend can be considered generally similar, especially from 1960, when mortality seems to decrease for all the three countries. Both Belgian and Luxembourger mortality tend to be higher than the French one. On the other hand, Lithuanian mortality exhibits a completely different pattern, being smaller the the french until year 1975 due to ages from 50 to 90, then steadily increasing until year 1990 and eventually decreasing again.

As for the second population group, Austria and Italy show the same similarities as France and Belgium, while Slovenian trend seems a bit irregular. It can be seen that Slovenian data is available only from a much later date, while mortality pattern



(a) FRA **(b)** ITA

Figure 3.4. SDR for ages 50-70



(a) FRA **(b)** ITA

Figure 3.5. SDR for ages 70-90

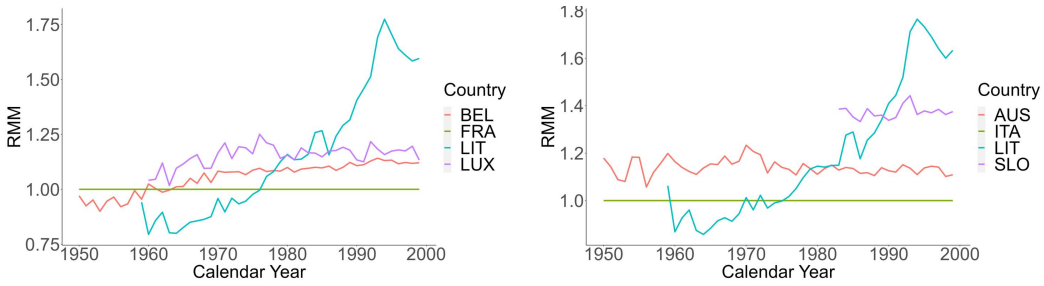
from Lithuanian population is quite different from the Italian one.

To more clearly asses the closeness of the mortality of a population to that of a reference one, [Menzietti et al.(2019)] define a Relative Measure of Mortality (RMM) as follows:

$$\text{RMM}^{I}_{t,a_s,a_e} = \frac{\sum_{x=a_s}^{a_e} E^{R}_{x,t} m^{I}_{x,t}}{\sum_{x=a_s}^{a_e} E^{R}_{x,t} m^{R}_{x,t}}$$

The nearer RRM value is to 1, the closer the mortality of the book population is to that of the reference population.

In both population groups, we can observe from figure 3.6 that the mortality of the reference populations tends to be the lowest within the group themselves. The only difference seems to be that Italian mortality evolves similarly to that of Austrian and Slovenian male populations, while French mortality seems to be improving faster than that of Belgium and Luxembourg. Lithuanian relative mortality dynamics behaves in both cases differently than that of the other book populations: while initially lower than that of the reference populations, it goes higher and higher from the mid-sixties to the mid-nineties, and then decreases again. This suggests to use caution when using data from the entire period 1950-1999 for estimate the boosting updates $\delta_{X,(T+1,\dots)}$ for Lithuanian book population, since this may produce a biased estimate of the updates.



(a) FRA **(b)** ITA

Figure 3.6. Relative Mortality measure for ages 30-90

3.2.3 Extension of ECB updates

The methodology for update extension from $\delta_{X,(1,\dots,T)}$ to $\delta_{X,(T+1,\dots)}$ hasn't been specified. Three possible approaches are investigated:

1. a first, verbatim, application of the Estimation Contrast Boosting algorithm on all three Age, Period and Cohort variables. The updates are taken to be the same from period 1950-1999 to period 2000-2019, s.t.

$$\delta_{(A,P,C);(1950-1999)} = \delta_{(A,P,C);(2000-2019)} \tag{3.1}$$

   ECB-based procedure using this update extension will be referred to as APC-based boosting or APC boosting. Such APC approach may lead to poor results, due to the lack of a mechanism to account for extending the updates, who have been estimated on the whole period 1950-1999, to years following 2000. In fact, one could say that such an approach makes use of a tree-like algorithm to model a phenomenon that is much likely non-stationary;

2. extending just the update from the last calendar year used for estimation, i.e. 1999.

$$\delta_{(A,1999,C);(1950-1999)} = \delta_{(A,P,C);(2000-2019)} \tag{3.2}$$

   This essentially means to assume that the updates referred to the last calendar year are extended to projection years, without any projection procedure. ECB-based procedure using this update extension will be referred to as 1999 correction. This approach essentially assumes that ECB updates do not depend on time (in the sense that they are the same as in the last year used for calibrating them), thus avoiding non-stationarity issue, and could be reasonable when projecting mortality over a short time horizon;

3. In order to produce updates that can be extended more straightforwardly to projection years, the estimation boosting can take into account a reduced form of the input design matrix, thus being applied just to age and cohort variables, or just to age or cohort.

$$\delta_{(A,C);(1950-1999)} = \delta_{(A,C);(2000-2019)} \tag{3.3}$$

$$\delta_{(A);(1950-1999)} = \delta_{(A);(2000-2019)} \tag{3.4}$$

$$\delta_{(C);(1950-1999)} = \delta_{(C);(2000-2019)} \tag{3.5}$$

This is a workaround to deal with non-stationarity issues: it assumes that the time-dependance of ECB updates can be estimated in terms of just Age or Cohort variables, hence removing the necessity for ECB update projection. ECB-based procedures using these update extensions will be referred to as, respectively Age-Cohort-based, Age-based or Cohort-based boosting (or, more briefly AC, a or C boosting).

3.2.4 Numerical results

The Estimation Contrast Boosting procedure was calibrated as follows. The quantity of interest to be projected and boosted is the central death rate s.t.:

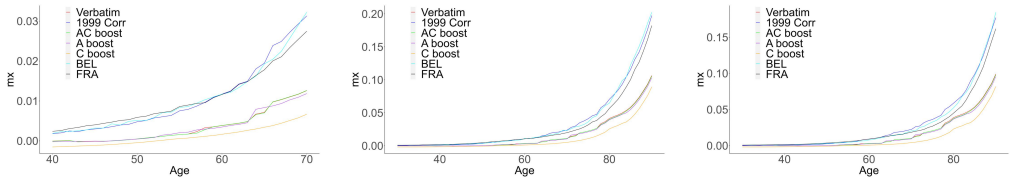
$$Q_{X,t} \leftarrow m_{x,t}$$

Time period from 1950 to 1999 was chosen to estimate parameters for reference populations' mortality projections, over ages from 30 to 90. Estimation Contrast Boosting updates are estimated, when possible, to the same period (see table 3.1 for availability of calendar year data for book populations), and using book population central exposed-to-risk as weights. Central death rates were projected using Lee-Carter mortality projection model, adopting a random walk with drift to model the random process of κ_t coefficients, as suggested by [Tuljapurkar et al.(2000)] for countries within the G7 group, as is the case for both reference populations taken into account. Mortality projection was applied to calendar years from 2000 to 2019 and ECB updates were extended to the same period.

The methodology summarized in figure 3.1 has been replicated for each of the ECB extensions described in equations from 3.1 to 3.5 and for each reference-book populations couple.

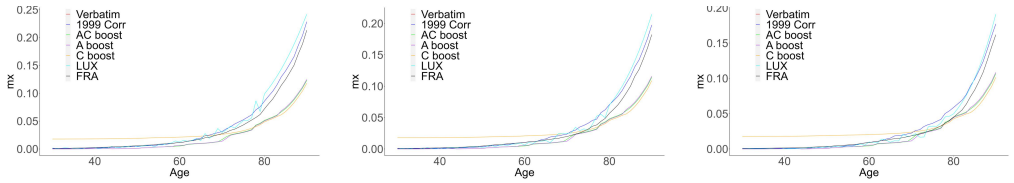
The boosted values $\hat{m}^B_{x,t}$ obtained for each book population are represented in figures from 3.7 to 3.12, at different future calendar years 2000, 2010, 2019. In each plot, mortality profiles $(x, m_{x,t})$ for reference population and those obtained from each of extensions in section 3.2.3 are represented. The mortality profile for book populations is also reported, in order to more easily understand how well the updated mortality rates approximate those of the book population: it must be recalled, however, that mortality rates for book populations in projection years were not used in the boosting update procedure and therefore constitute out-of-sample data.

It can be seen from figures 3.7 and 3.8 that the only update extension seemingly bringing French mortality near to Belgian or Luxembourger one is the 1999 correction. All other updates, except Cohort-based boosting, produce an updated book mortality population lower than that of the reference population, which is clearly in contrast



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

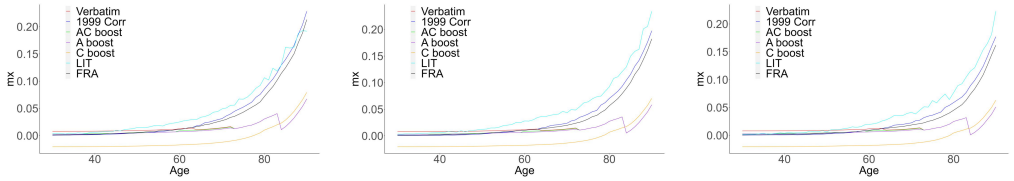
Figure 3.7. Updated $m_{x,t}$ for Belgian population, starting from French mortality.



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 3.8. Updated $m_{x,t}$ for Luxembourgish population, starting from French mortality.

to what figure 3.6 suggested. Contrast boosting based on update 3.5 produces an even lower updated book population mortality and must therefore be regarded as an unreliable result.

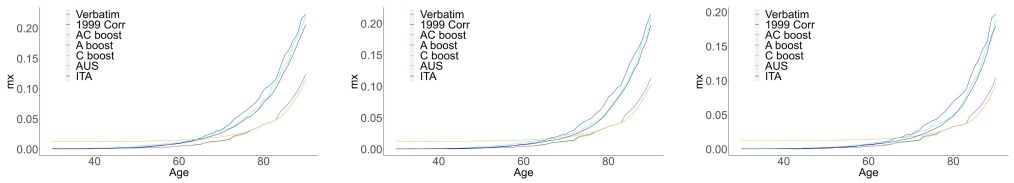


(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 3.9. Updated $m_{x,t}$ for Lithuanian population, starting from French mortality.

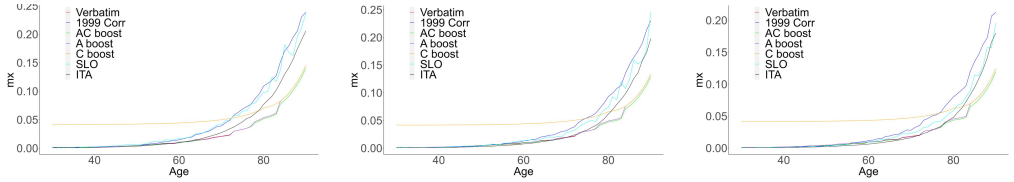
For the Lithuanian book population in figure 3.9, similar considerations hold: it should be noted the odd behavior for $m_{x,t}$ produced by Age-based boosting: this could come from the peculiar Lithuanian mortality history in years from 1950 to 1970, highlighted in figures from 3.3 to 3.6.

Similar observations can be repeated for Italian reference population group. From figure 3.10 and 3.11 it can be seen that Cohort-based boosting tends to overestimate mortality for younger ages. Another point of attention is the domain of updated $m_{x,t}$: as can be seen for Cohort-based boosting in figure 3.12, updated mortality rates for book population can be negative. In this case, use of Estimation Contrast Boosting technique of section 3.2.1 for Lithuanian data must be discarded.



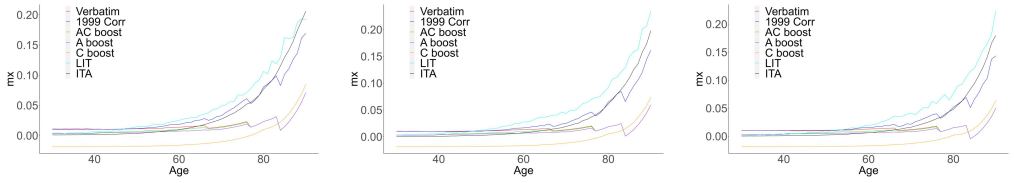
(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 3.10. Updated $m_{x,t}$ for Austrian population, starting from Italian mortality.



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 3.11. Updated $m_{x,t}$ for Slovenian population, starting from Italian mortality.

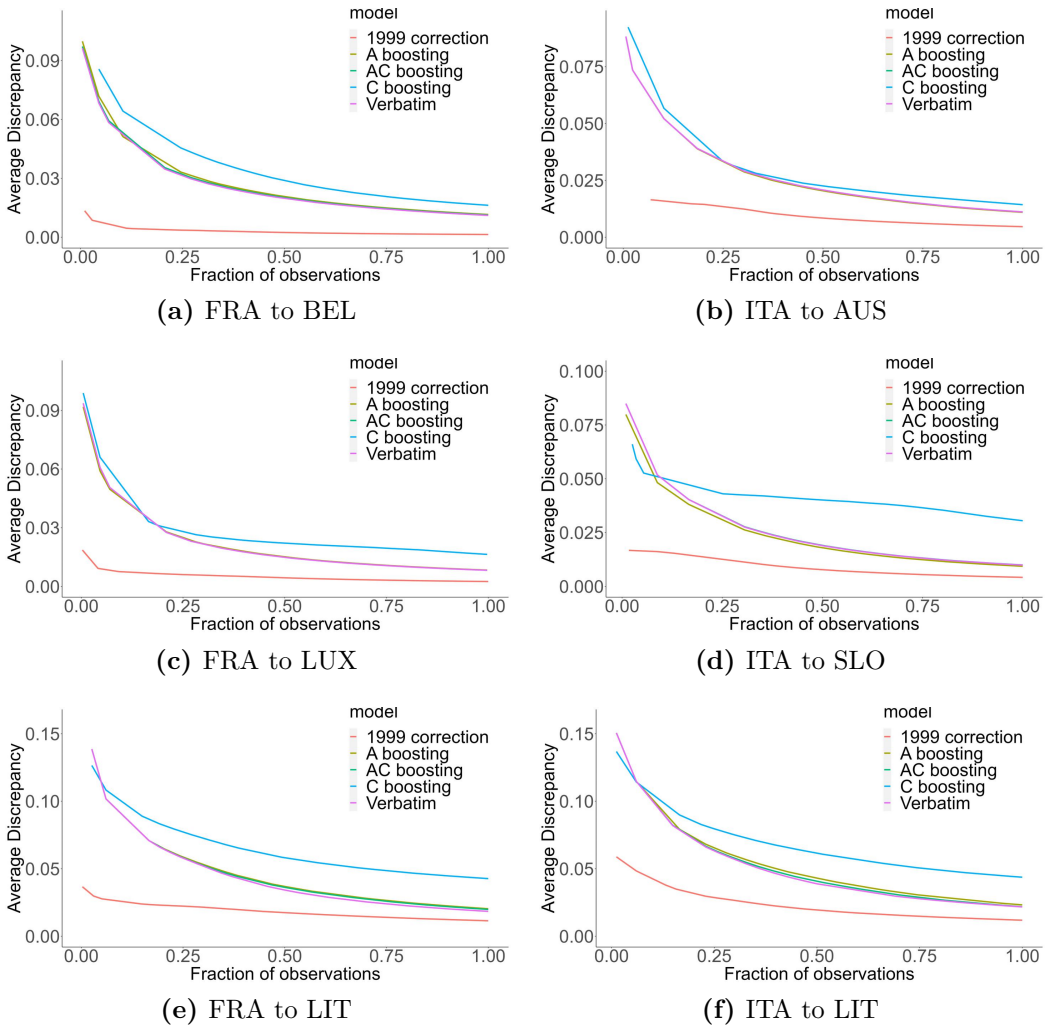


(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 3.12. Updated $m_{x,t}$ for Lithuanian population, starting from Italian mortality.

Result assessment using CTs

In order to investigate the performances of methodology outlined in section 3.2.1, estimates $\hat{m}^B_{X,t}$ are compared to values $m^B_{X,t}$ available from the HMD for all book populations. As in section 2.3, performance assessment will be carried out using Contrast Trees for diagnostics, contrasting predicted $\hat{m}_{x,t}$ with observed $m_{x,t}$ for each book population, using mean absolute difference discrepancy as in eq. 1.1 and weights given from the central exposure to risk for book populations. Each tree had maximum size set to 100 and minimum number of points in a leaf fixed to 20. Training and test set were obtained splitting in half data relative to the projection time interval using maximum dissimilarity approach. Results from CTs have also been compared to classical MSEP and MAPE prediction error estimators, also weighted using central exposure to risk for book populations.



(a) FRA to BEL **(b)** ITA to AUS

(c) FRA to LUX **(d)** ITA to SLO

(e) FRA to LIT **(f)** ITA to LIT

Figure 3.13. Lack-of-fit curves from Contrast Trees for the two populations groups

Lack-of-fit curves resulting from different boosting approaches for each couple

reference-book population are reported in figure 3.13.

Three kinds of behavior can be distinguished for all six book populations:

1. As immediately apparent, the best boosting approach for both French and Italian population groups seems to be the extension from the last year used for projection. It can also be seen that the performance of such correction remains consistently good across all input space: this suggests that using a subset of the time period from 1959 to 1999 may yield better results compared to using the whole period as input for the boosting procedure.

2. APC boosting, AC boosting and A boosting yield a very similar LOF curve: this is due to the fact that, for models including the age variable, the iterative tree partition operates mainly using age variable, thus leading to similar updates and similar boosting performances.

3. Estimation boosting based on the cohort variable exhibits a peculiar behavior. It can however be concluded, on the basis of the LOF curve, that other boosting approach lead to better results, except in some of the worst performing regions. It can be noted that from the curves in figure 3.13 the unacceptability of some results (e.g. negative $m_{x,t}$ for C boosting from Italian to Lithuanian mortality) can't be inferred.

The comparison between these results and classical MSEP and MAPE indicators, reported alongside average discrepancy in table 3.2, brings similar conclusions. From the RMSE perspective, the 1999 correction seems almost always the best approach to extend $\delta_{(APC;1950-1999)}$ to calendar years 2000-2019: interestingly enough, this doesn't hold for MAPE in the case of Lithuanian book population, both for French and Italian reference population. For Italian reference population the value of RMSE is particularly low. It can also be noted that MAPE value for Cohort boosting is much greater than that relative to other boosting approaches.

When comparing MAPE between Italian and French reference populations, it appears that the estimation boosting procedure yields better results when using Italian mortality as a starting point. This may be due to the fact that, as pointed out in figure 3.6, the relative levels of mortality for Italy and its corresponding book population remains steady for the 1950-1999 period, contrary to those of France.

Moreover, as expected from the lack-of-fit curves, APC-based boosting, Age-Cohort-based boosting and Age-based boosting all tend to lead to similar results in terms of RMSE, MAPE and average discrepancy.

Regions in the Lexis space uncovered by the diagnostic Contrast Trees are reported in figures from 3.14 to 3.19 to investigate where performance drops. The scale is the same for all boosting approaches and populations, in order to allow a direct comparison. It can be said that higher age regions in general are affected by the worst performance. For Cohort boosting, average discrepancy is greater across all the Lexis space, while 1999 correction manages to keep average discrepancy low also for elderly ages.

(a) FRA to BEL

Boosting approach	RMSE	MAPE	Average discrepancy
APC boosting	0,0180	0,8069	0,0082
1999 correction	0,0024	0,1433	0,0011
AC boosting	0,0183	0,8081	0,0083
A boosting	0,0188	0,8312	0,0084
C boosting	0,0239	1,3908	0,0116

(b) ITA to AUS

Boosting approach	RMSE	MAPE	Average discrepancy
APC boosting	0,0168	0,4119	0,0059
1999 correction	0,0060	0,2869	0,0023
AC boosting	0,0168	0,4140	0,0059
A boosting	0,0167	0,4119	0,0058
C boosting	0,0192	4,3572	0,0108

(c) FRA to LUX

Boosting approach	RMSE	MAPE	Average discrepancy
APC boosting	0,0159	0,5836	0,0062
1999 correction	0,0040	0,2783	0,0019
AC boosting	0,0159	0,5844	0,0062
A boosting	0,0157	0,5790	0,0062
C boosting	0,0208	8,0705	0,0154

(d) ITA to SLO

Boosting approach	RMSE	MAPE	Average discrepancy
APC boosting	0,0157	0,2958	0,0045
1999 correction	0,0063	0,2119	0,0022
AC boosting	0,0157	0,3046	0,0046
A boosting	0,0148	0,2915	0,0043
C boosting	0,0348	13,6714	0,0346

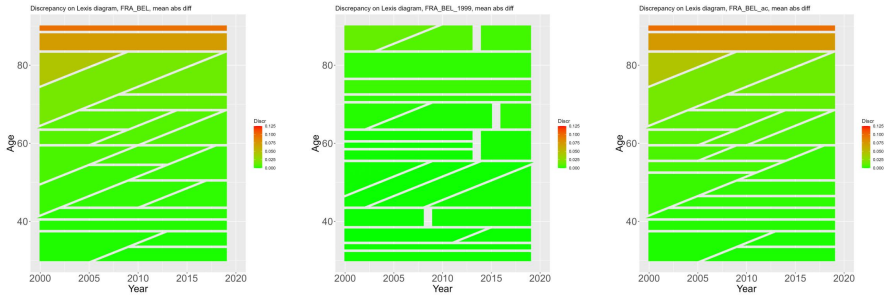
(e) FRA to LIT

Boosting approach	RMSE	MAPE	Average discrepancy
APC boosting	0,0295	0,6234	0,0134
1999 correction	0,0132	0,6487	0,0103
AC boosting	0,0301	0,5612	0,0146
A boosting	0,0304	0,6008	0,0152
C boosting	0,0456	3,2876	0,0382

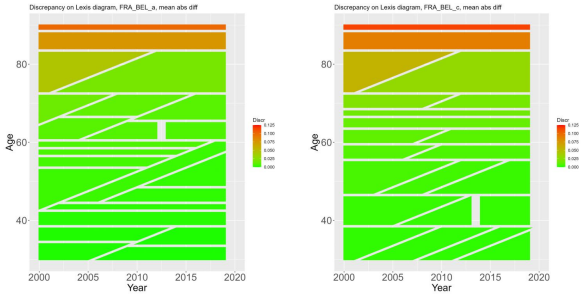
(f) ITA to LIT

Boosting approach	RMSE	MAPE	Average discrepancy
APC boosting	0,0002	0,7844	0,0118
1999 correction	0,0146	0,6808	0,0076
AC boosting	0,0307	0,4785	0,0120
A boosting	0,0316	0,5478	0,0133
C boosting	0,0446	3,0950	0,0352

Table 3.2. Performance statistics for different boosting approaches.

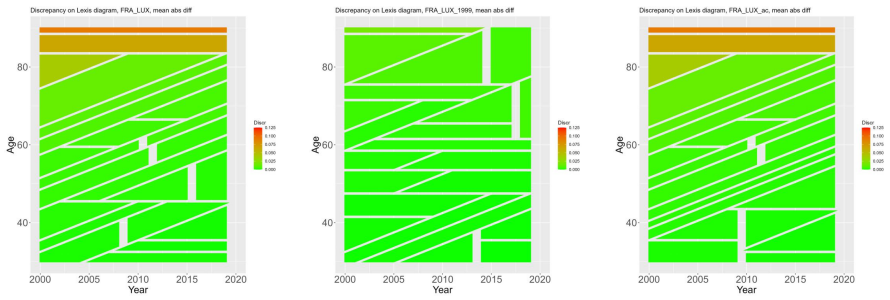


(a) APC boosting (b) 1999 correction (c) AC boosting

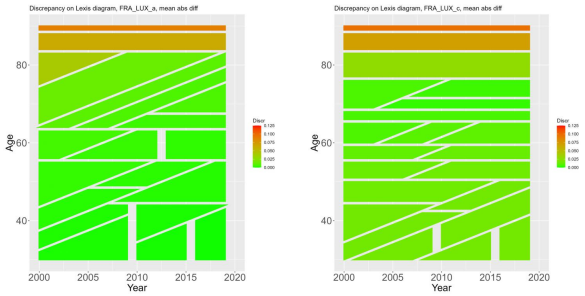


(d) A boosting (e) C boosting

Figure 3.14. Regions in the Lexis space for boosting from FRA to BEL.

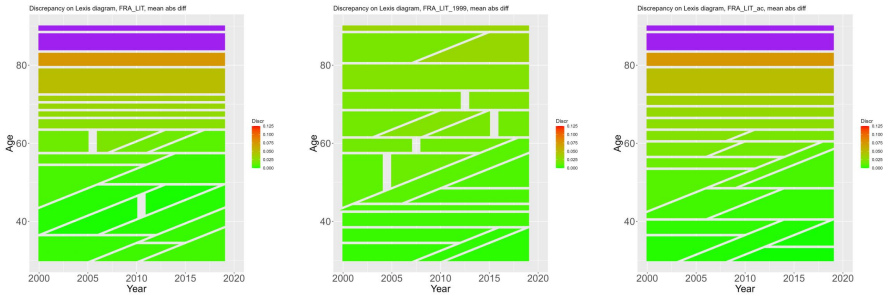


(a) APC boosting (b) 1999 correction (c) AC boosting

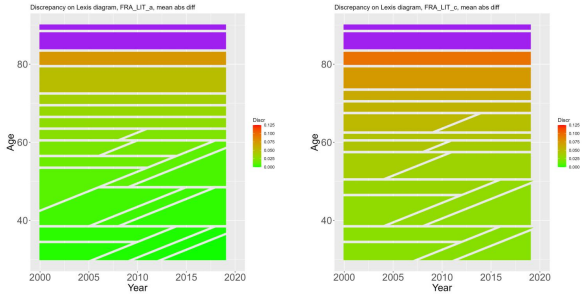


(d) A boosting (e) C boosting

Figure 3.15. Regions in the Lexis space for boosting from FRA to LUX.

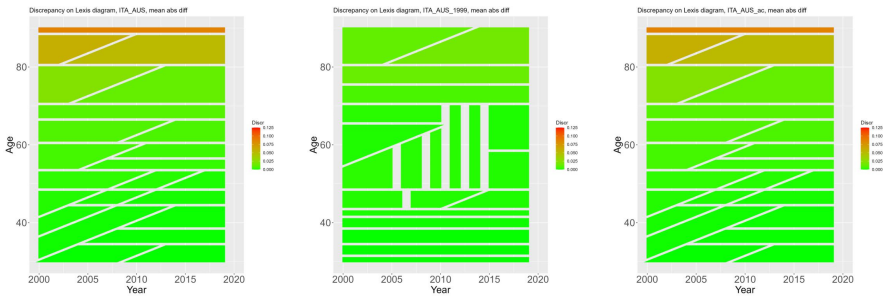


(a) APC boosting **(b)** 1999 correction **(c)** AC boosting

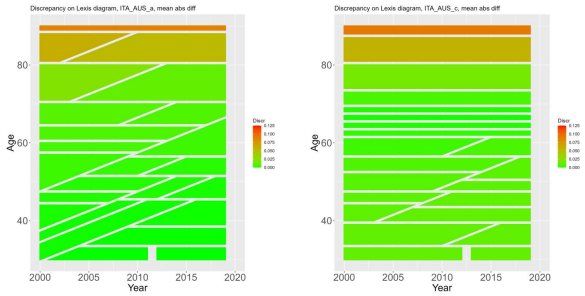


(d) A boosting **(e)** C boosting

Figure 3.16. Regions in the Lexis space for boosting from FRA to LIT.



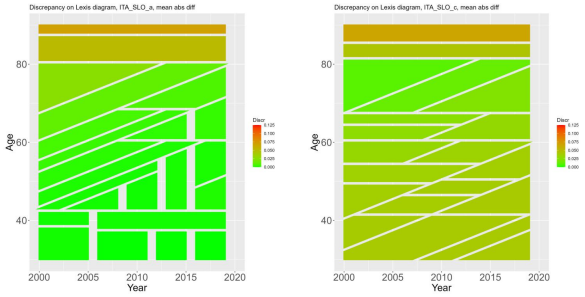
(a) APC boosting **(b)** 1999 correction **(c)** AC boosting



(d) A boosting **(e)** C boosting

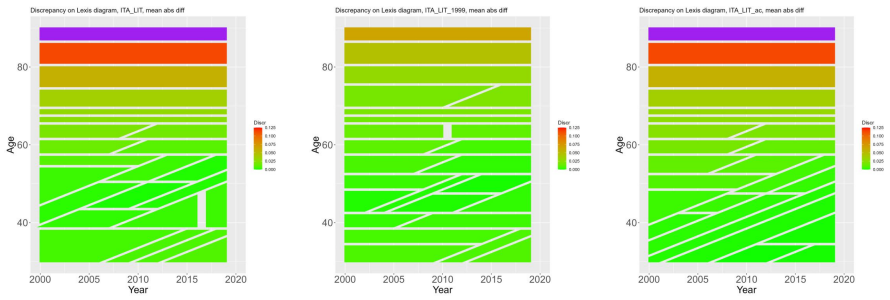
Figure 3.17. Regions in the Lexis space for boosting from ITA to AUS.

(a) APC boosting (b) 1999 correction (c) AC boosting

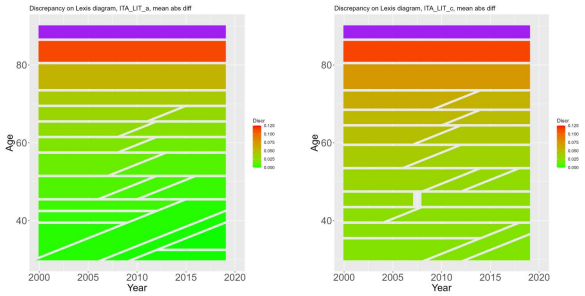


(d) A boosting (e) C boosting

Figure 3.18. Regions in the Lexis space for boosting from ITA to SLO.



(a) APC boosting (b) 1999 correction (c) AC boosting



(d) A boosting (e) C boosting

Figure 3.19. Regions in the Lexis space for boosting from ITA to LIT.

3.3 Discussion

In this chapter, a general methodology for mortality spread modeling through Estimation Contrast Boosting has been proposed. One of the possible choices for this methodology application has been applied to two population groups, each constituted by one reference population and three book populations. Results show that particular care must be taken when examining results, since the boosting procedure does not necessarily yield acceptable boosted estimates for book populations $\hat{m}^B_{x,t}$, since equation 1.5 doesn't impose any bound on the updated values $\hat{z}(\mathbf{x})$.

Another point worthy of attention is the choice of extension for Estimation Contrast Boosting updates $\delta_{\mathbf{X},t}$, since the definition of a procedure to project such updates , which are obtained using a tree-based algorithm, is not trivial. While A and AC boosting feature ease of projection, the fact that the extension of 1999 correction still yields better results suggests that using not all data available from calendar years 1950-1999, but a more recent subset, could maybe lead to better performing ECB mortality spread modeling.

It can also be concluded that, since the boosted estimate from APC, AC, and A boosting procedure tend to overlap, the prevailing variable in determining the partitions of input space used to calculate the updates is variable age: therefore, when applying the procedure of section 3.2.1 to the populations analyzed, an age-based Estimation Contrast Boosting could be adopted for the sake of variable parsimony.

Chapter 4

Adapting mortality tables using CTs

4.1 Mortality modeling in Italian insurance companies

In Italian actuarial practice, the assessment of mortality in life insurance companies is quite impacted from ISVAP Regulation N.22 of 4th April 2008 and its successive modifications[1]. Article 23-bis, comma 9 of the regulation states that "The Insurance company conducting life business presents to IVASS the comparison between technical bases, different from interest rates, used for the calculation of technical provisions, and the results of direct experience". This happens trough the filling of table 1/1 of Module 41 contained in the Additional Information to Financial Statements (*"Informazioni aggiuntive al bilancio d'esercizio"*).

The module is organized separately by risk type (longevity or mortality), product typology and age range: these features define the risk classes to be considered. The underlying idea, common in actuarial mathematics, is to perform the assessment separately for groups defined by known risk factors, so that individual composing these groups can be considered homogeneous from the point of view of probabilistic evaluation. The comparison between direct experience and technical bases requested by the Regulation consists in reporting the expected number of deaths in the company portfolio, the expected sum of benefit to be paid, the actual number of deaths and the actual paid sum of benefits. Expected number of deaths $\tilde{d}_x^C$ and expected sum of benefit to be paid $\tilde{B}_x^C$ are defined as follows:

$$\tilde{d}_x^C = q_x^{(1)} * n_x^C \quad \tilde{B}_x^C = q_x^{(1)} * S_x^C$$

where n_x^C is the number of insureds of age x in risk cluster C at the beginning of the year, S_x^C is the total insured sum for insureds of age x in cluster C at the beginning of the year and $q_x^{(1)}$ is the probability of death between ages x and $x+1$, usually dependent on sex. The table $\{q_x^{(1)}\}_0^\omega$ constitutes part of the prudential technical

[1] Provvedimento ISVAP of 29 January 2010 N. 2771, Provvedimento ISVAP of 17 November 2010 N. 28452, Provvedimento IVASS of 6 December 2016 N. 53, Provvedimento IVASS of 14 February 2018 N. 68.

basis used for pricing purposes or, in Italian actuarial terminology, "Base tecnica del primo ordine" (First-order technical basis). The time period to which all quantities refer is the solar year of the Financial Statements in consideration.

While not strictly required by Solvency II directive, in Italian actuarial practice a very similar approach is adopted in order to produce the mortality hypotheses to be used in the calculation of best estimate liabilities and, more generally, to estimate company mortality tables. In this case, given a certain risk cluster C usually described by sex, risk type and product type, an adaptation coefficient α^C is calculated as:

$$\alpha_d^C = \frac{\sum_x q_x^{(2)} n_x^C}{\sum_x d_x^C} \tag{4.1}$$

or, using the sum of benefits as a reference:

$$\alpha_B^C = \frac{\sum_x q_x^{(2)} S_x^C}{\sum_x B_x^C} \tag{4.2}$$

where d_x^C is the observed number of deaths for age x in cluster C, B_x^C is the sum of benefits paid for insureds deceased at age x in cluster C and $q_x^{(2)}$ is the probability of death, distinct by sex, from some reference mortality table, usually a national table of some sort (e.g. SIM2011 table or one ISTAT tables for people residing in Italy). The table $\{q_x^{(2)}\}_0^\omega$ is a demographic realistic technical base that can be updated over time ("Base tecnica del secondo ordine" - Second order technical basis).

The adapted death rates $q_{d,x}^C$ or $q_{B,x}^C$ which will constitute the company life table are then calculated as follows:

$$q_{d,x}^C = \alpha_d^C q_x^{(2)} \quad q_{B,x}^C = \alpha_B^C q_x^{(2)} \tag{4.3}$$

In conclusion, this procedure scales death rates in order to reproduce the total number of deaths or the total benefits paid. Data used for the calculation include observations from the last calendar years. The number of calendar years used is usually around ten, in order to have a sufficiently stable output without having to resort to data from much earlier calendar years, which is often not available.

While this method is quite simple in its implementation, it has the disadvantage of applying the same adaptation coefficient to all ages, thus creating a mortality profile which has the same shape as the reference one, while the company underwriting process may affect each age group differently.

4.2 A proposal for extension using CTs

The methodology from section 3.2.1 may be modified as follows in order to produce an extended version of the adaptation procedure explained in section 4.1.

Suppose to fix a risk cluster C, for example Male insureds for Term Life Insurance. In order to mimic the process outlined in the previous section, the quantity of interest

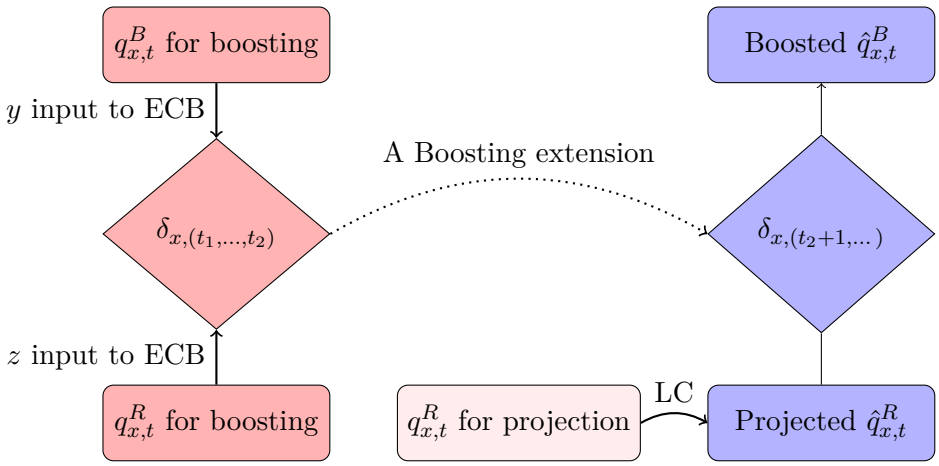


Figure 4.1. The Estimation Contrast Boosting process outlined in section 3.2.1 adapted to the generalized procedure outlined in section 4.1. Cells in red refer to calendar years $t_1, \dots, t_2$; cells in pink refer to calendar years $t_0, \dots, t_2$, with $t_0 \leq t_1$, while cells in blue refer to years $t_2 + 1, \dots$.

will be the probability of death $q_{x,t}$ for reference (national) and book (company) populations, s.t.:

$$Q^B_{X,t} \leftarrow q^B_{x,t}; \quad Q^R_{X,t} \leftarrow q^R_{x,t}$$

Age boosting, similarly as equation 3.4, will be adopted as the boosting approach for update extension, in order to apply corrections $\delta_{x,(t_1,\dots,t_2)} = \delta_{x,(t_2,\dots)}$ that depend exclusively on age but are not necessarily held fixed for all values of x: such updates will be estimated using data from calendar years $t_1, \dots, t_2$. In order to obtain more reliable results, mortality projection for reference population could make use of a data series longer than that used for estimation boosting, since it's not necessary that reference population data for mortality projection and update extension to refer to the same time span. Calendar years $t_0, \dots, t_2$, with $t_0 \leq t_1$, will be used to calibrate mortality projections models.

Being the reference population a national one, a classical projection method, such the Lee-Carter model, may be applied to chosen data. Since such methods may apply to central death rate $m_{x,t}$, constant force of mortality will be assumed (see [Pitacco(2009)] for more details), so that:

$$q_{x,t} = 1 - e^{-m_{x,t}} \tag{4.4}$$

The process is illustrated in figure 4.1.

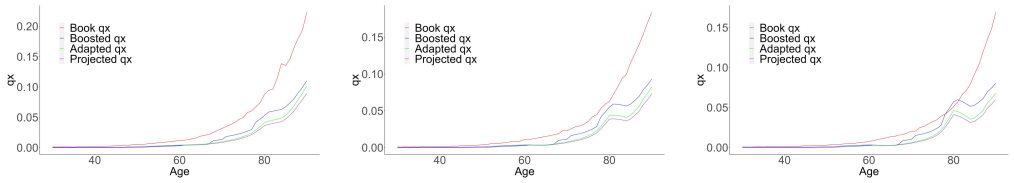
4.3 Numerical results

In this section results from the procedure outlined in section 4.2 will be presented. Company data would be required in order to asses the feasibility of this new extension: since no such data were available during the writing of this work, the same

data described in section 3.2.2 will be used. The results of the procedure will then be compared to those obtained by applying the adaptation procedure for the number of deaths as in equation 4.1.

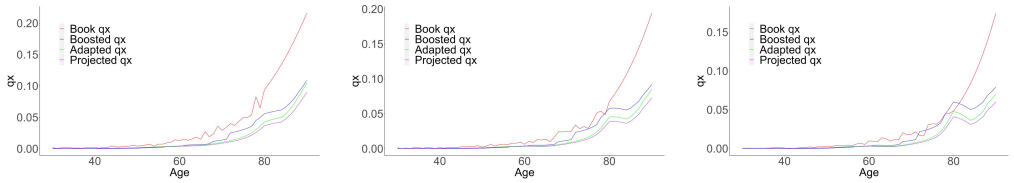
The time period for calibrating mortality projection is chosen to be from 1980 to 1999, while the calendar years used to calculate Estimation Contrast Boosting boosting updates and death rates adaptation are from 1990 to 1999. Estimation Contrast Boosting has been performed using mean absolute difference discrepancy 1.1, using the number of individuals in book populations as weights. Central mortality rates, and subsequently death rates, have been projected to period 2000-2019 and compared to observed result using, as usual by now, Contrast Trees, RMSE and MAPE.

Figures from 4.2 to 4.7 reports observed values for $q_{x,t}$ relative to the book population, projected $q_{x,t}$ for the reference population, adapted $q_{x,t}$ obtained from the rescaling procedure described in section 4.1 and updated $q_{x,t}$ from the boosting procedure outlined in section 4.2. For the sake of presentation simplicity, just results for calendar years 2000, 2010 and 2019 are reported.



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

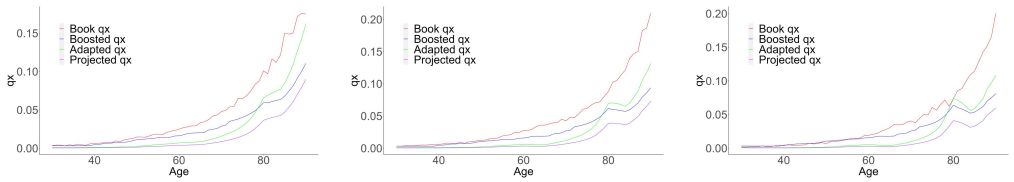
Figure 4.2. Updated and adapted $q_{x,t}$ for Belgian population, starting from French mortality.



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

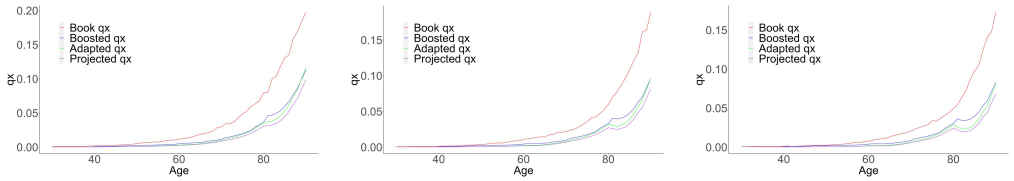
Figure 4.3. Updated and adapted $q_{x,t}$ for Luxembourgish population, starting from French mortality.

It is clear, from each figure, the effect of q_x adaptation: the death rate curve for reference population is simply rescaled and the output curve will never cross the initial death rate projected curve. In this case, being $\alpha_d^C > 1$ for every book population, this means that adapted book mortality will be higher than reference population mortality for every age. On the contrary, the effect of estimation boosting is not always proportional: this can be seen e.g. in figure 4.16a where the boosting



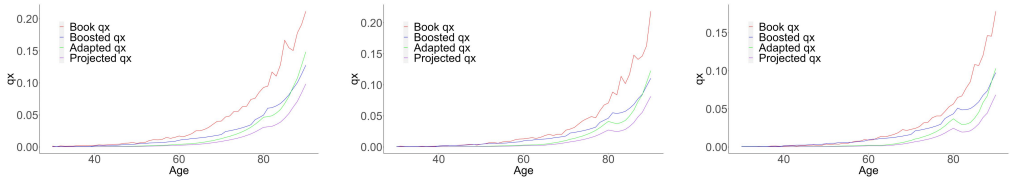
(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.4. Updated and adapted $q_{x,t}$ for Lithuanian population, starting from French mortality.



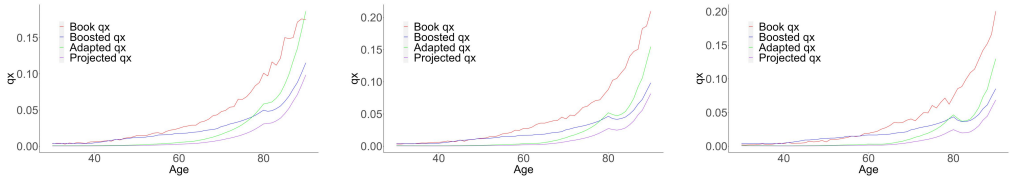
(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.5. Updated and adapted $q_{x,t}$ for Austrian population, starting from Italian mortality.



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.6. Updated and adapted $q_{x,t}$ for Slovenian population, starting from Italian mortality.



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.7. Updated and adapted $q_{x,t}$ for Lithuanian population, starting from Italian mortality.

effect in ages 80-90 is clearly lower than that in ages 70-80. The non-proportionality of the affect can also be seen in figure 4.6b where the curves for adapted and boosted q_x intersect.

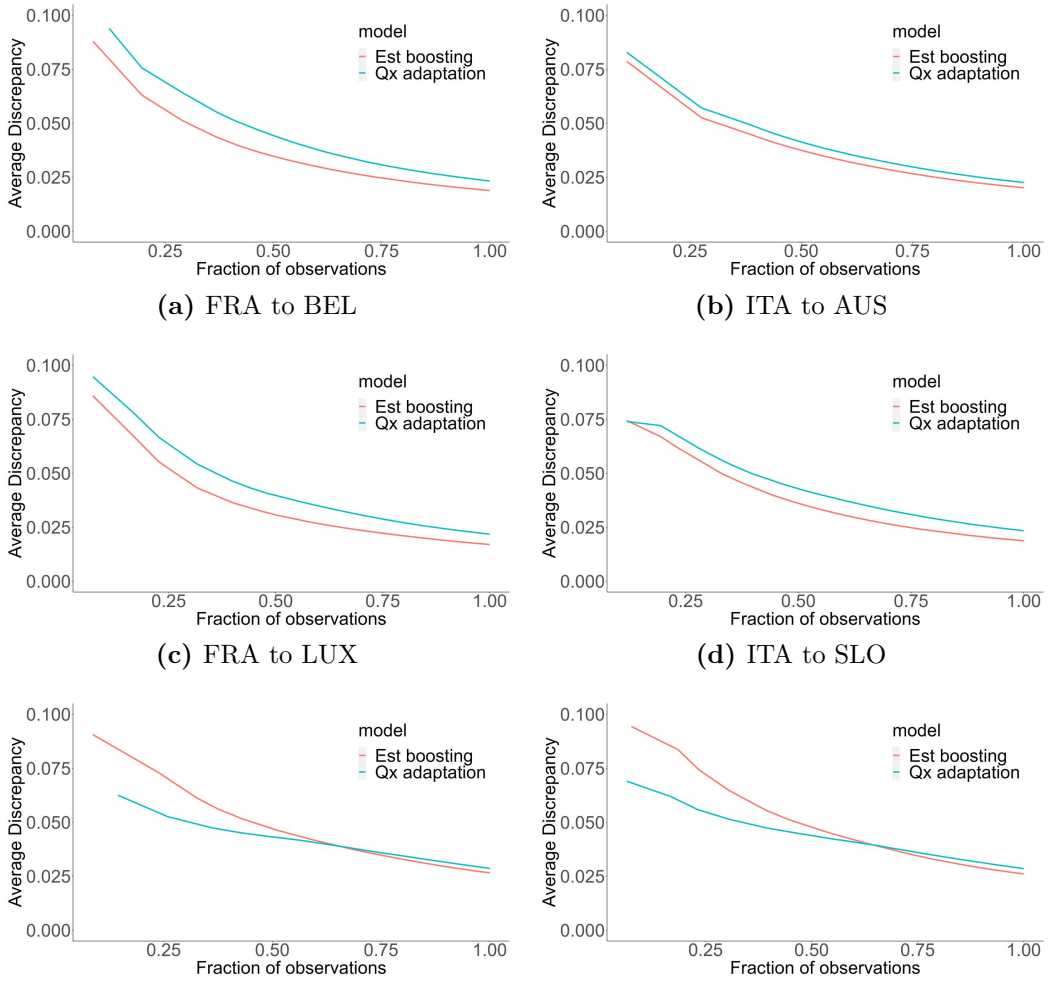
It can be observed that, while output death rates are higher than those resulting from reference population projection, all procedures tend to underestimate book death rates, especially at elder ages. Two patterns can be discerned:

- In the case of Belgian, Luxembourger, Austrian and Slovenian book populations, q_x from boosting procedure seem to approach book population observed death rates more closely across almost all age groups, with some exceptions at elder ages.
- For Lithuanian population, both in case of French and Italian reference populations, q_x resulting from boosting methodology consistently seem a better approximation to observed ones until age 75-80, also with a reduced underestimation effect. Then the effect of rescaling gets stronger, resulting in a higher adapted q_x for elder ages.

As in the previous chapter, results from adaptation and boosting methods are compared using Contrast Trees. Each set of estimated death rates is contrasted with corresponding observed values using discrepancy 1.1. Maximum tree size is set to 100, minimum number of objects in a tree is set to 20, weights are equal to the number of individuals in book populations. Training and test set for CT procedure were obtained splitting in half data referring to time period 2000-2019 with maximum dissimilarity approach.
Lack-of-fit curves are presented in figure 4.8. For book populations different from Lithuania, the estimation boosting method for q_x seems to consistently yield better results across all input space. The situation is reversed for Lithuania: while in low-performing (elder age) regions, discrepancy is much lower for adapted mortality rates, the difference between the two methodologies reduces until, for average discrepancy around 0.4, the trend reverses, thus resulting in an average discrepancy that supports the estimation boosting approach. This suggests that for Lithuanian book population, the process should be split in two: adaptation for older ages and estimation boosting otherwise.

Performance indicators reported in table 4.1 confirm the previous observations. RMSE and average discrepancy for all book populations suggest the q_x adaptation boosting as the best-performing methodology. Here the usefulness of the lack-of-fit curve can be appreciated: by looking at the average values in the tables, the effect of different performances between elder ages and all other age classes for Lithuania would not be discernible.

The regions identified by Contrast Trees for q_x resulting from the two methodologies are reported in figure 4.9. For all book populations except Lithuania, both adaptation and boosting have far worse performances for elder ages. While this also hold for Lithuania, the effect for adaptation is scattered among all ages greater than 60, instead of being noticeable at age 75-80.



(a) FRA to BEL

(b) ITA to AUS

(c) FRA to LUX

(d) ITA to SLO

(e) FRA to LIT

(f) ITA to LIT

Figure 4.8. Lack-of-fit curves from Contrast Trees for the two populations groups

(a) FRA to BEL

Method	RMSE	MAPE	Average discrepancy
Adaptation	0,0156	0,6146	0,0137
Boosting	0,0125	0,8590	0,0116

(b) ITA to AUS

Method	RMSE	MAPE	Average discrepancy
Adaptation	0,0149	0,6699	0,0117
Boosting	0,0132	0,4888	0,0101

(c) FRA to LUX

Method	RMSE	MAPE	Average discrepancy
Adaptation	0,0135	0,5655	0,0150
Boosting	0,0104	0,7544	0,0117

(d) ITA to SLO

Method	RMSE	MAPE	Average discrepancy
Adaptation	0,0148	0,6682	0,0163
Boosting	0,0109	0,3010	0,0112

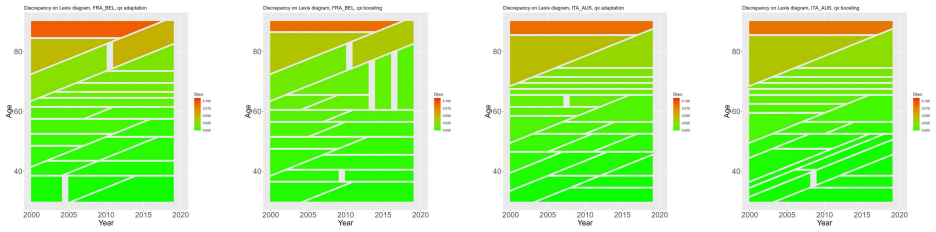
(e) FRA to LIT

Method	RMSE	MAPE	Average discrepancy
Adaptation	0,0178	0,7455	0,0175
Boosting	0,0150	0,2880	0,0146

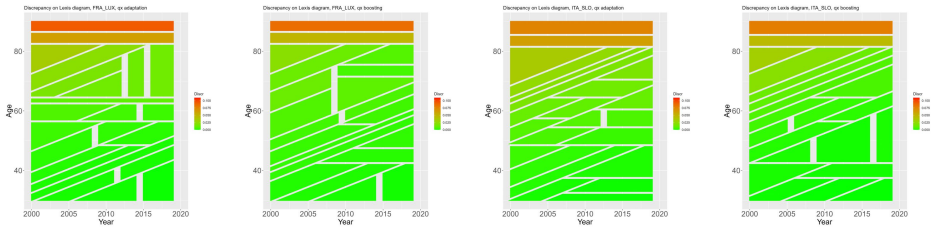
(f) ITA to LIT

Method	RMSE	MAPE	Average discrepancy
Adaptation	0,0191	0,7901	0,0191
Boosting	0,0173	0,2773	0,0164

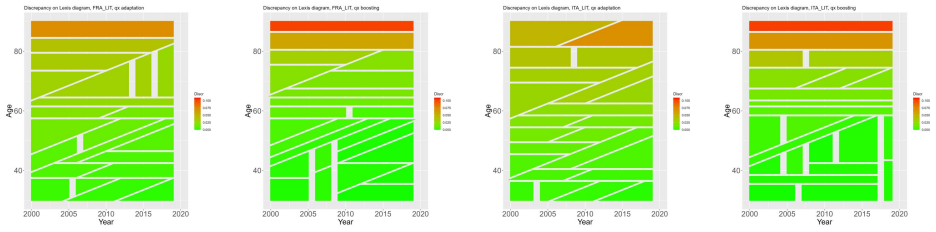
Table 4.1. Performance statistics for the two methods.



(a) Adaptation, FRA to BEL **(b)** Boosting, FRA to BEL **(c)** Adaptation, ITA to AUS **(d)** Boosting, ITA to AUS



(e) Adaptation, FRA to LUX **(f)** Boosting, FRA to LUX **(g)** Adaptation, ITA to SLO **(h)** Boosting, ITA to SLO



(i) Adaptation, FRA to LIT **(j)** Boosting, FRA to LIT **(k)** Adaptation, ITA to LIT **(l)** Boosting, ITA to LIT

Figure 4.9. Discrepancy regions on Lexis space

4.4 Distribution Contrast Boosting applied to mortality projection

Now a variant of the methodology explained in section 4.2 will be described and tested. The variant will make use of Distribution Contrast Boosting (DCB) of section 1.4.3, while data structure will remain the same as section 3.2.2.

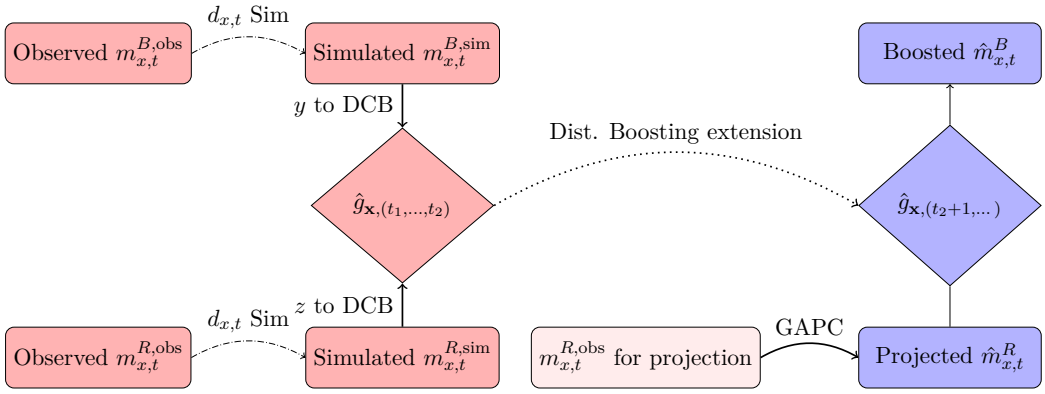


Figure 4.10. The Estimation Contrast Boosting process outlined in section 3.2.1 adapted to the generalized procedure outlined in section 4.4. Cells in red refer to calendar years $t_1, \dots, t_2$; cells in pink refer to calendar years $t_0, \dots, t_2$, with $t_0 \leq t_1$, while cells in blue refer to years $t_2 + 1, \dots$.

For every data cell in Age-Period-Cohort input space, and in accordance with distributions for GAPC models, a given distribution for number of deaths will be assumed, with appropriate parameter values depending on cell. Starting from data for both reference and book populations, the distributional assumption will be utilized in order to create simulated data both for book and reference populations: these will to be used as input for some Distribution Contrast Boosting procedure, obtaining Distribution Boosting transformation $\hat{g}_{\mathbf{x},(t_1,\dots,t_2)}$ as in equation 1.6.

On the other hand, mortality projection will be performed by means of a Generalized Age-Period-Cohort model chosen in order to be consistent with distributional hypotheses: projected mortality rates will then be transformed accordingly to an extension $\hat{g}_{\mathbf{x},(t_2+1,\dots)}$ of Distribution Boosting transformation, in order to obtain boosted mortality rates. The procedure is summarized in figure 4.10.

For each data cell, 150 simulated rows will be created, having same $\mathbf{x}$ value as the starting row, while the simulated value for central death rate $m^{\text{sim}}_{x,t}$ is obtained from simulating death numbers from a Poisson distribution with parameter $E^{C}_{x,t} m^{\text{obs}}_{x,t}$, where $m^{\text{obs}}_{x,t}$ is the observed value for central death rate from HMD and $E^{C}_{x,t}$ is the corresponding value of central exposed to risk. Based on the considerations in sections 3.2.4 and 4.2, the Distribution Boosting approach based on Age variable will be considered, in order to extend the distribution boosting transformation in the following way, similar to extensions 3.3 and 3.4:

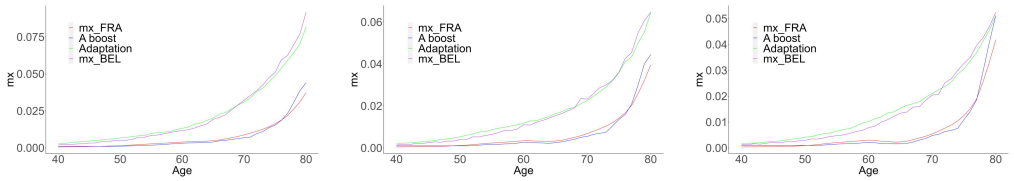
$$\hat{g}_{A;(1990-1999)} = \hat{g}_{A;(2000-2019)} \tag{4.5}$$

DCB-based procedure using this extension will be referred to as Age Distribution Boosting or A distribution boosting.

Data from calendar years 1990-1999 are used for both Poisson simulation and distribution boosting. Central death rates $\hat{m}^R_{x,t}$ will be obtained from a GAPC Lee-Carter model, whose death distribution assumption is the same used for data simulation, calibrated on calendar years 1980-1999. While the methodology can be used to obtain the whole distribution of central death rates for book populations in calendar years 2000-2019, in this procedure only the median will be considered, due to computational limitations. It can be recalled that, for Poisson distributions having mean parameter large enough, the median is a good approximation of the average (see e.g. [Choi(1994]).

In order to assess performances of this whole procedure, boosted $\hat{m}^B_{x,t}$ will be compared to those obtained from the adaptation procedure described in section 4.1, whose resulting mortality rates have been converted to central death rates using formula 4.4. Calendar years from 1990 to 1999 are used in order to estimate adaptation coefficients.

In order to visualize the effect of Distribution Contrast Boosting, in figures from 4.11 to 4.16 the estimated values of $\hat{m}^B_{x,t}$ are reported for calendar years 2000, 2012, 2019 and for each of the six book population. For the sake of figure readability, only ages from 40 to 80 are reported. it can be seen that, for each book population, mortality rates calculated from the adaptation procedure seem to better approximate actual $m_{x,t}$ for the book populations. This effect can be seen, e.g. in figures 4.11c, 4.12a and 4.15b. It's also apparent that the effect of Distribution Contrast Boosting is not proportional.

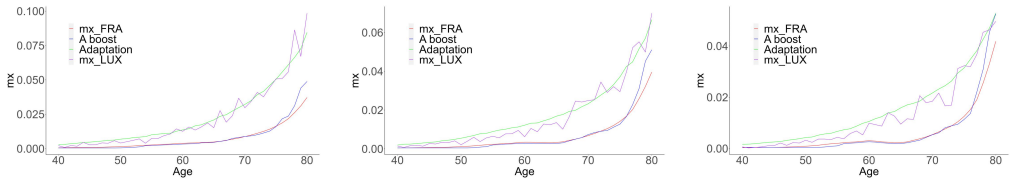


(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.11. $\hat{m}_{x,t}$ adapted and obtained from Age Distribution Boosting for Belgian population, starting from French mortality.

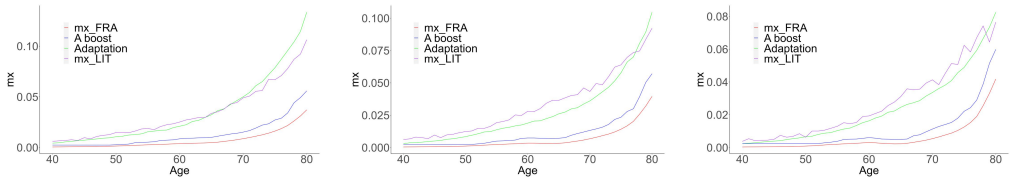
As usual by now, diagnostics were performed by means of Contrast Trees. As before, trees were built using discrepancy 1.1, with maximum tree size set to 100, minimum number of objects in a tree set to 20 and weights equal to the number of individuals in book populations. Training and test set for CT procedure were obtained splitting in half data relative to calendar years from 2000 to 2019 using maximum dissimilarity approach.

Figure 4.17 report the lack-of-fit curves for both boosting and adaptation procedures. As suggested by figures from 4.11 to 4.16, the $m_{x,t}$ adaptation procedure



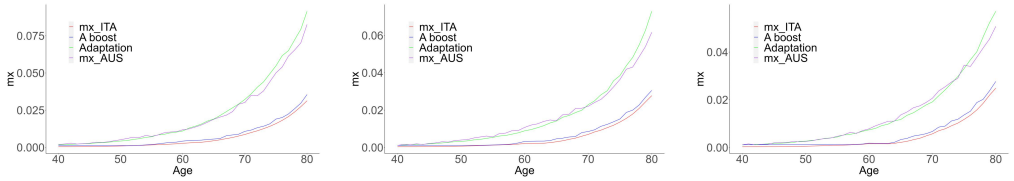
(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.12. $\hat{m}_{x,t}$ adapted and obtained from Age Distribution Boosting for Luxembourger population, starting from French mortality.



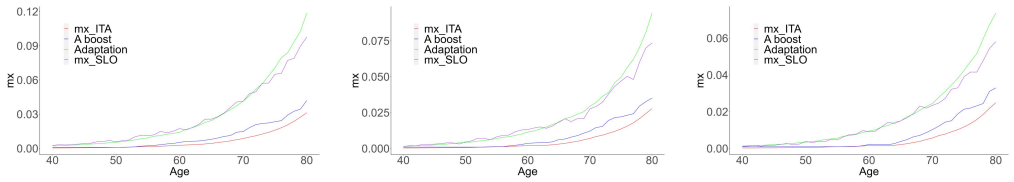
(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.13. $\hat{m}_{x,t}$ adapted and obtained from Age Distribution Boosting for Lithuanian population, starting from French mortality.



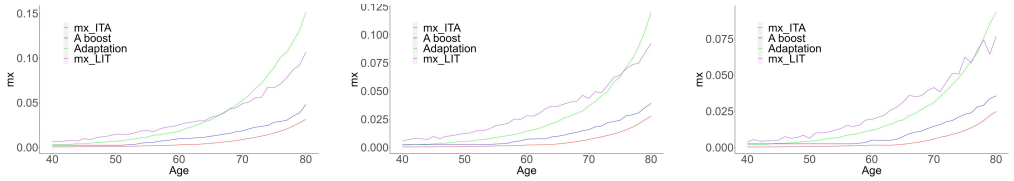
(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.14. $\hat{m}_{x,t}$ adapted and obtained from Age Distribution Boosting for Austrian population, starting from Italian mortality.



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.15. $\hat{m}_{x,t}$ adapted and obtained from Age Distribution Boosting for Slovenian population, starting from Italian mortality.



(a) Year 2000 **(b)** Year 2010 **(c)** Year 2019

Figure 4.16. $\hat{m}_{x,t}$ adapted and obtained from Age Distribution Boosting for Lithuanian population, starting from Italian mortality.

(a) FRA to BEL

Method	RMSE	MAPE	Average discrepancy
A boosting	0,0176	2,0924	0,0153
Mx adaptation	0,0019	0,1727	0,0016

(b) ITA to AUS

Method	RMSE	MAPE	Average discrepancy
A boosting	0,0291	7,962	0,0216
Mx adaptation	0,0041	0,11	0,0028

(c) FRA to LUX

Method	RMSE	MAPE	Average discrepancy
A boosting	0,0172	2,7919	0,0152
Mx adaptation	0,0032	0,2992	0,0027

(d) ITA to SLO

Method	RMSE	MAPE	Average discrepancy
A boosting	0,0337	10,1777	0,025
Mx adaptation	0,0105	0,1642	0,0072

(e) FRA to LIT

Method	RMSE	MAPE	Average discrepancy
A boosting	0,0251	2,2556	0,0236
Mx adaptation	0,0137	0,5278	0,0125

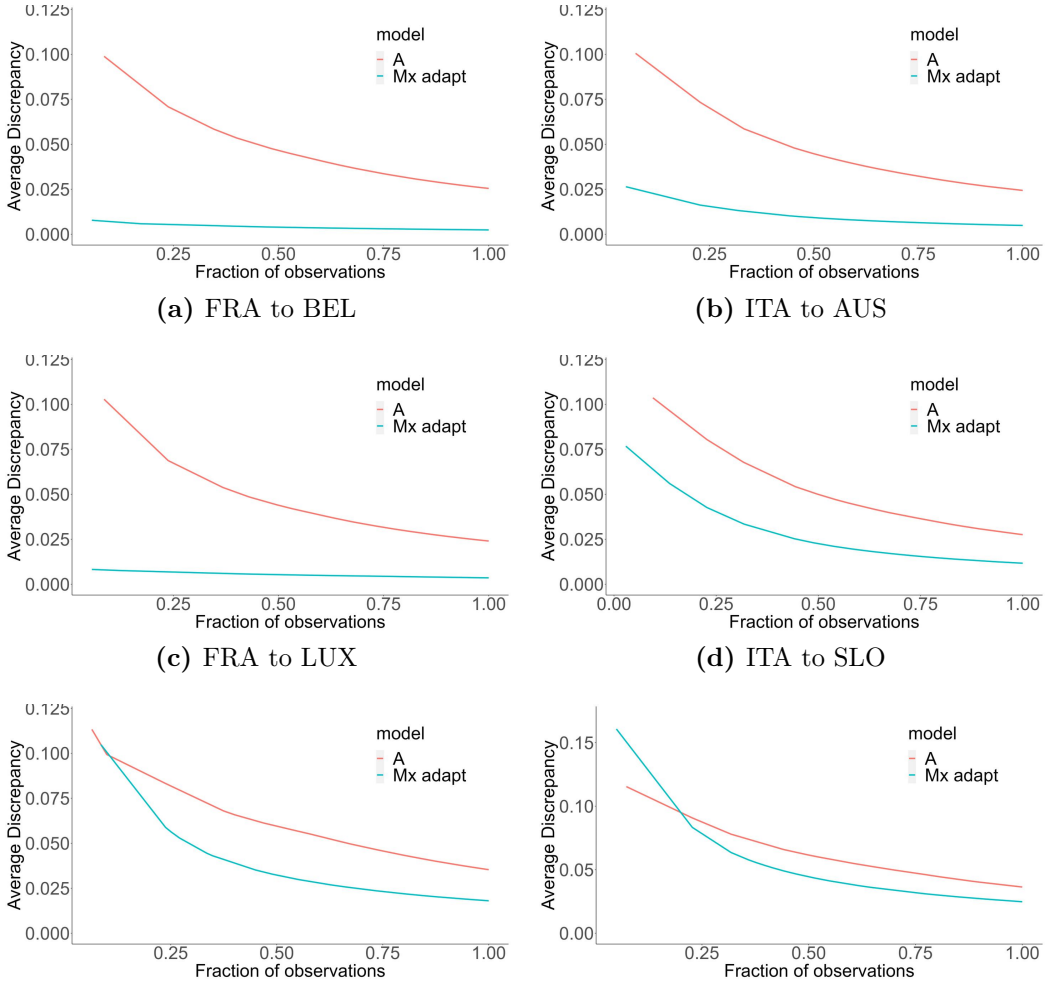
(f) ITA to LIT

Method	RMSE	MAPE	Average discrepancy
A boosting	0,0415	19,1179	0,0346
Mx adaptation	0,0203	1,1076	0,0174

Table 4.2. Performance statistics for the two methods.

performs consistently better across all input space. The only exception happens when adapting Lithuanian mortality: in this case the A boosting procedure produces better results for high-discrepancy (in this case, for elder ages) regions.

Finally, results in table 4.2, who reports the usual performance indicators, are consistent with graphical analysis: in all cases and for all three indicators, the adoption of $m_{x,t}$ adaptation procedure is suggested.



(a) FRA to BEL **(b)** ITA to AUS

(c) FRA to LUX **(d)** ITA to SLO

(e) FRA to LIT **(f)** ITA to LIT

Figure 4.17. Lack-of-fit curves from Contrast Trees for the two methodologies

4.5 Discussion

In actuarial practice, the estimation of company mortality tables is of particular importance, in order to correctly evaluate mortality (and longevity) risk. In the Italian actuarial context, a death rate adaptation procedure inspired by authority regulations to adapt mortality tables is often adopted. While being very easily implementable, the downside of this procedure lies in the assumptions that mortality at all ages is affected in the same (multiplicative) way.

In order to allow for more detailed corrections, Estimation Contrast Boosting can be utilized within the context of a more general procedure. A wide array of choices is available in terms of both mortality projection models and update extensions, though, for the sake of projection simplicity, Lee-Carter model and extensions 3.4 and 3.3 have been chosen to conduct the analysis. Estimation of future mortality rates for book populations via Estimation Boosting seems to consistently provide more accurate result in respect to simpler mortality adaptation. Moreover, lack-of-fit curves can be used to asses which methodology yields better results in a certain region of the input space. While these results may appear promising, caution is advised since no company data were available for testing the procedure.

Application of a Distribution Contrast Boosting method for building mortality tables for small populations, while feasible, does not lead to an improvement in prediction accuracy compared to the traditional mortality adaptation method. The analysis presented in this work, however, is affected by several limitations. First of all, due to computational constraints, the simulation was limited to 150 rows for each data point. Moreover, and more importantly, the key feature of Distribution Contrast Boosting, which is its capability to provide an estimate for the whole distribution of the quantity of interest, was not taken advantage of. Also, it would have been interesting to extend the results from methodology of section 4.4 using different extensions, such as a.g. an Age-Cohort extension. Again, hardware constraints did not allow for more in-depth analysis.

Conclusions

Leveraging [Friedman(2020)], who introduces Contrast Trees to estimate the full conditional probability distribution without any parametric assumptions, an application to mortality description models is proposed, with particular regard to the intersection of machine learning and mortality modeling fields. In this sense, this novel application fills the gap between mortality modeling and model diagnostics, providing a unified framework for assessing nontraditional machine learning models and extrapolative likelihood based ones. Contrast Trees can be applied to both purely descriptive models and stochastic mortality projection models, providing a new set of diagnostic instruments, that can be utilized to assess the behavior of models in automatically produced subsets of the input space. In particular, in the context of modelization of mortality models on Age-Period-Cohort input space, such subsets are easily representable on the Lexis diagram, allowing for ease of interpretation and association with points in lack-of-fit curves.

The Estimation Contrast Boosting procedure, which in the original work of [Friedman(2020)] was geared towards discrepancy reduction between a quantity of interest and some reference quantity, is then applied to the problem of mortality modeling for small populations. Book population mortality tables are produced using updates estimated by the boosting procedure to modify the mortality rates of a reference, much larger, population, whose mortality can be projected using well-attested models. This allows for a great flexibility in the choice of discrepancy, update extension mechanism and mortality projection model. While some care should be taken in assessing results from the estimation boosting procedure and in selecting the mechanism to extend the updates to projection years, this methodology can be used to produce mortality tables for small populations.

Finally, taking into account Italian actuarial practice, Estimation Contrast Boosting has been used to generalize the mortality adaptation technique for estimating a realistic demographic technical basis often used within insurance companies. This new procedure seems to lead to better estimation results compared to those obtained with the traditional techniques and also allows to overcome the limitation of the latter to apply the same adaptation coefficient to all ages. A similar methodology, using Contrast Distribution Boosting to estimate a function for calculating book population mortality , can be applied but, at the moment, does not seem to allow for an improvement in mortality prediction.

List of acronyms

APC Age-Period-Cohort (could refer to a mortality projection model or to a predictor structure). 13, 23, 38

CBD Cairns-Blake-Dowd (mortality projection model). 24

CT Contrast Tree. 3

DCB Distribution Contrast Boosting. 8, 58

ECB Estimation Contrast Boosting. 7, 34, 38

GAPC Generalised Age-Period-Cohort (mortality projection models). 22

GBM Gradient Boosting Machine. 14

GLM Generalized Linear Model. 13

HMD Human Mortality Database. 12, 24

KL Kullbach-Leibler (divergence). 27

LC Lee-Carter (mortality projection model). 13, 23

LOF Lack-Of-Fit (contrast curve). 6

MAPE Mean Absolute Percentage Error. 16

RH Renshaw-Haberman (mortality projection model). 23

RMM Relative Measure of Mortality. 37

RMSE Root Mean Square Error. 16

SDR Standardized Death Rates. 36

XGBM eXtreme Gradient Boosting Machine. 14

Bibliography

[Ahcan at al.(2014)] Ahcan, A., Medved, D., Olivieri, A., Pitacco, E. (2014). Forecasting mortality for small populations by mixing mortality data. Insurance: Mathematics and Economics, 54, 12-27.

[Alai and Sherris(2014)] Alai, D.H., Sherris, M. (2014). Rethinking age-period-cohort mortality trend models. Scandinavian Actuarial Journal, 3: 208-227.

[Bongaarts(2005)] Bongaarts, J. (2005). Long-Range Trends in Adult Mortality: Models and Projection Methods. Demography, 42(1): 23-49.

[Booth and Tickle(2008)] Booth, H., Tickle, L. (2008). Mortality modelling and forecasting: A review of methods. Annals of Actuarial Science, 3(1-2): 3-43. DOI: 10.1017/S1748499500000440.

[Booth et al.(2006)] Booth, H., Hyndman, R. J., Tickle, L., De Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. Demographic research, 15, 289-310

[Breiman et al.(1984)] Breiman, L., Friedman, J., Olshen, R. A., Stone, C. J. (1984). Classification and regression trees Chapman & Hall. New York.

[Brouhns et al.(2002)] Brouhns, N., Denuit, M., Vermunt, J. (2002). A Poisson log-bilinear approach to the construction of projected life tables, Insurance: Mathematics and Economics, 31: 373-393.

[Chen et al.(2015)] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.

[Choi(1994] Choi, K. P. (1994). On the medians of gamma distributions and an equation of Ramanujan. Proceedings of the American Mathematical Society, 121(1), 245-251.

[Cairns et al.(2006)] Cairns, A.J.G., Blake, D., Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. Journal of Risk and Insurance, 73: 687-718.

[Cairns et al.(2008)] Cairns, A.J.G., Blake, D., Dowd, K. (2008). Modelling and management of mortality risk: a review. Scandinavian Actuarial Journal, 73 (2-3): 79-113.

[Cairns et al.(2009)] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North American Actuarial Journal, 13: 1-35.

[Cairns et al.(2010)] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Khalaf-Allah, M. (2010). Evaluating the goodness of fit of stochastic mortality models. Insurance: Mathematics and Economics, 47(3): 255-265.

[Deprez et al.(2017)] Deprez, P., Shevchenko, P.V., Wúthrich, M.V (2017). Machine learning techniques for mortality modeling. European Actuarial Journal, 7: 337-352. https://doi.org/10.1007/s13385-017-0152-4.

[Currie(2006)] Currie, I. D. (2006). Smoothing and forecasting mortality rates with P-splines. Talk given at the Institute of Actuaries.

[Currie(2014)] Currie, I. D. (2016). On fitting generalized linear and non-linear models of mortality. Scandinavian Actuarial Journal, 2016(4), 356-383.

[Djeundje et al.(2022)] Djeundje, V.B., Haberman, S., Bajekal, M. et al. (2022). The slowdown in mortality improvement rates 2011-2017: a multi-country analysis. European Actuarial Journal. DOI: 10.1007/s13385-022-00318-0

[Friedman(2001)] Friedman, J.H. (2001). Greedy function approximation: A Gradient Boosting Machine. Annals of Statistics 29: 1189-1232.

[Friedman(2020)] Friedman, J.H. (2020). Contrast trees and distribution boosting. Proceedings of the National Academy of Sciences, 117 (35): 21175-21184. DOI: 10.1073/pnas.1921562117

[Friedman and Narasimhan(2020)] Friedman, J.H., Narasimhan, B. (2020). conTree: Contrast Trees and Distribution Boosting. R package version 0.2-8.

[Haberman and Renshaw(2009)] Haberman, S., Renshaw, A. (2009). On age-period-cohort parametric mortality rate projections. Insurance: Mathematics and Economics, 45(2), 255-270.

[Hastie et al. (2009)] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: Springer.

[Hunt and Blake(2015)] Hunt, A., Blake, D. P. (2015). On the structure and classification of mortality models.

[Jarner and Kryger (2011)] Jarner, S. F., Kryger, E. M. (2011). Modelling adult mortality in small populations: The SAINT model. ASTIN Bulletin: The Journal of the IAA, 41(2), 377-418.

[Keyfitz and Caswell(2005)] Keyfitz, N., Caswell, H. (2005). Applied mathematical demography (Vol. 47). New York: Springer.

[Kuhn(2020)] Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... & Team, R. C. (2020). Package "'caret"'. The R Journal, 223(7).

[Kullback(1959)] Kullback, S. (1959). Information Theory and Statistics. Pages 5-6. Wiley, New York.

[Kullback and Leibler(1951)] Kullback, S., Leibler, R. A. (1951). On information and sufficiency. The annals of mathematical statistics, 22(1), 79-86.

[Lee and Carter(1992)] Lee, R.D., Carter, L.R. (1992). Modeling and forecasting US mortality. Journal of the American statistical association, 87 (419): 659-671.

[Levantesi and Nigri(2020)] Levantesi, S., Nigri, A. (2020). A random forest algorithm to improve the Lee-Carter mortality forecasting: impact on q-forward. Soft Computing, 24: 8553-8567. DOI: 10.1007/s00500-019-04427-z

[Levantesi and Pizzorusso(2019)] Levantesi S., Pizzorusso, V. (2019). Application of Machine Learning to Mortality Modeling and Forecasting. Risks, 7(1), 26. ISSN: 2227-9091. DOI:10.3390/risk7010026

[Li et al.(2009)] Li, J. S.H., Hardy, M. R., Tan, K. S. (2009). Uncertainty in mortality forecasting: an extension to the classical Lee-Carter approach. Astin Bulletin 39(1), 137-164.

[McCullagh and Nelder(1989)] McCullagh, P., Nelder, J., 1989. Generalized Linear Models, 2nd Edition. Chapman & Hal, London.

[Menzietti et al.(2019)] Menzietti, M., Morabito, M. F., Stranges, M. (2019). Mortality projections for small populations: An application to the Maltese elderly. Risks, 7(2), 35.

[Nigri et al.(2022)] Nigri, A., Barbi, E., Levantesi, S. The relay for human longevity: country-specific contributions to the increase of the best-practice life expectancy. Quality & Quantity, 56, 4061-4073. https://doi.org/10.1007/s11135-021-01298-1

[Pitacco(2009)] Pitacco, E. (2009). Modelling longevity dynamics for pensions and annuity business. Oxford University Press.

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

[Pollard(1987)] Pollard, J.H. (1987). Projection of age-specific mortality rates. In: Population Bulletin of the United Nations 21/22: 55-69.

[Renshaw and Haberman(2006)] Renshaw, A.E. , 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 and Wüthrich(2021)] Richman, R. and Wüthrich, M. (2021). A neural network extension of the Lee-Carter model to multiple populations. Annals of Actuarial Science, 15: 346-366.

[Scognamiglio and Marino(2022)] Scognamiglio, S., Marino, M. Backtesting stochastic mortality models by prediction interval-based metrics. Quality & Quantity. https://doi.org/10.1007/s11135-022-01537-z

[Torri(2011)] Torri, T. (2011). Building blocks for a mortality index: an international context. Eur. Actuar. J. 1 (Suppl 1): S127-S141

[Tuljapurkar et al.(2000)] Tuljapurkar, S., Li, N., Boe, C. (2000). A universal pattern of mortality decline in the G7 countries. Nature, 405(6788), 789-792

[Villegas et al.(2015)] Villegas, A., Kaishev, V. K., Millossovich, P. (2015, December). StMoMo: An R package for stochastic mortality modelling. In 7th Australasian Actuarial Education and Research Symposium.

[Villegas et al.(2017)] Villegas, A. M., Haberman, S., Kaishev, V. K., Millossovich, P. (2017). A comparative study of two-population models for the assessment of basis risk in longevity hedges. ASTIN Bulletin: The Journal of the IAA, 47(3), 631-679.

[Villegas et al.(2018)] Villegas, A.M., Kaishev, V. and Millossovich, P. (2018). StMoMo: An R Package for Stochastic Mortality Modelling. Journal of Statistical Software, 84 (3): 1-38.

[Wan and Bertschi(2015)] Wan, C., Bertschi, L. (2015). Swiss coherent mortality model as a basis for developing longevity de-risking solutions for Swiss pension funds: A practical approach. Insurance: Mathematics and Economics, 63, 66-75.

[Willett(1999)] Willett, P. (1999). Dissimilarity-based algorithms for selecting structurally diverse sets of compounds. Journal of Computational Biology, 6 (3-4): 447-457.