**Revision checklist - Reviewer #1**

Paper “Enhancing diagnostic of stochastic mortality models leveraging contrast trees. An application on Italian data”

Submission ID: 51a8fa45-cb85-4c01-ad48-a73538820020

We would like to thank the anonymous reviewer for the helpful comments and valuable suggestions that contributed to improve our paper. We have revised the paper accordingly.

All of the reviewer’s comments and suggestions have been incorporated into the revised paper.

We have provided a response to each of the comments below in blue text.

**Reviewer 1**  
  
Referee:

Based on the aims and scope of the journal, an in-depth revision of the language writing, applications, and conclusions of the paper is needed before possible publication.  
As a general comment, this paper investigates a new technique, Contrast Trees, which, leveraging decision trees, provides a general approach for evaluating the quality of fit of different models by detecting the regions in the input space where models work poorly. To verify the ability of this approach, the authors consider standard stochastic mortality models and machine learning algorithms in estimating the Italian mortality rates from the Human Mortality Database. Once the low-performance regions are detected, the authors use Contrast Boosting to improve the inaccuracies of mortality estimates provided by each model. The results are discussed using graphical and numerical tools, particularly for the high-error regions.

1) First, it would be interesting to know how to implement the model in R; maybe the authors can add an Appendix. Implementation details are fascinating.

Authors: We have provided the R code used in the application in the Appendix.

2) One of the principal challenges for mortality models is robustness and residuals with good behaviour; the goodness-of-fit measures are not enough to recommend a model and test the prediction. Therefore, these tools should help improve the models and choose the best one.

In addition, the authors do state the main advantages of this "new technique" compared with other diagnostic tools.

Authors: Other diagnostic tools that can be mentioned:

* Residual analysis (Kevin Dowd, Andrew J.G. Cairns, David Blake, Guy D. Coughlan, David Epstein, Marwa Khalaf-Allah, Evaluating the goodness of fit of stochastic mortality models. Insurance: Mathematics and Economics, Volume 47, Issue 3, 2010, Pages 255-265)
* Proportion of variance explained by the model (R^2), BIC e BIC di Djeundje-Currie, 2010 🡪 solo per i modelli la cui stima è basata sulla likelihood
* Testing if the fitted errors are white noise (Torri)
* AIC, Schwarz-Bayes Criterion and Likelihood-ratio test (Astin Bulletin 39(1), 137-164. doi: 10.2143/AST.39.1.2038060) 🡪 solo per i modelli la cui stima è basata sulla likelihood

Rafforzare che il Contrast tree può essere utilizzato indifferentemte per modelli stimati con o senza likelihood, ecc. 🡪 Our goal is to improve the model’s performance to fitting historical mortality data. We aim to find the best model that fits historical mortality rates by grasping and detecting the inaccuracies of any model and boosting the predictive power.

SUSANNA

3) Finally, the author(s) should discuss the real advantage of their technique compared to the existing ones. I think the results should be revised in more detail, and conclusions or practical interpretations about them should be indicated in the Conclusions. For example, how do these diagnostic tools help to improve the model projections? How does this result help actuaries in pricing? Do the results depend on the country? Are there different results for different age ranges?

Authors:

* how do these diagnostic tools help to improve the model projections? Il nostro paper utilizza il Contrast Tree come strumento diagnostico per il fitting e non per il forecasting. Noi NON non proponiamo il CT per migliorare il forecasting.
* Prediction per imputazione di missing data. Valutiamo la performance predittiva. No estrapolazione. Fase preliminare per migliorare la qualità del dato. Utile in caso di modellazione di popolazioni assicurate che hanno dati mancanti e numeri ridotti.
* I risultati dipendono dal paese. Noi ci concentriamo sull’Italia.
* Testare differenti age ranges? Aggiungere age group 0-29 come suggerisce lui stesso al punto 7).

ANDREA-MATTEO

4) Furthermore, please, provide future manuscripts with continuous page and line numbers!

Authors:

Particular comments

5) Page 2, the age range should be justified. The challenge is to fit models with an extensive age range with ancient ages. On the other hand, if authors analyze the age groups 30-60 and 61-90 separately, they will obtain very optimistic results. Why not three age groups?

Authors: aggiungiamo age group 0-29

6) Page 3, Please explain how x+t represents the cohort.

Authors: We mean “t-x” instead of “x+t”. It was a typo that we have corrected in the new version of the paper.

7) Page 5, How is measured accuracy in terms of mxt or log(mxt)? Do results depend on those measures?

Authors: Fare anche AVECRI, MAPE e RMSE con log(mxt). Raddoppiando così i risultati.

**Revision checklist - Reviewer #2**

Paper “Enhancing diagnostic of stochastic mortality models leveraging contrast trees. An application on Italian data”

Submission ID: 51a8fa45-cb85-4c01-ad48-a73538820020

Referee:

The manuscript applies contrast boosting technique to mortality projection models. While this is a valuable task, I am not convinced that this work in its current state is substantial enough to warrant publication for the following reason.

1) The chief goal of mortality projection models is to forecast mortality rates into future. Any assessment of this type of models must go beyond goodness of fit and assess the predictive performance because adherence to past data does not necessarily translate into good prediction. Various illustrations of this can be found in Djeundje et al (2022). For model comparison in this area, goodness of fit assessment must be carried out alongside the resulting predictive performance of the models and underlying uncertainty.  
  
Reference:  
Djeundje et al. (2022) The slowdown in mortality improvement rates 2011–2017: A multi-country analysis.  European Actuarial Journal

Authors: potremmo prendere parte della risposta data al punto 3) del referee 1