Review

Summary

The manuscript introduces the maximum entropy mortality model where forecasting of future mortalities can be based on extrapolation of a finite number of statistical moments for the age-at-death distribution using a multivariate time series model. As a case study, the 'England and Wales' male mortality experience for years 1960-2016 and ages 0-95 is considered. The model is shown to outperform (in the sense of out-of-sample predictive power w.r.t. ME, MAE, etc.) classic approaches such as the Lee and Carter (1992) model.

Comments and questions

The manuscript addresses a relevant issue for actuarial practice and contains an interesting empirical study.

I have the following comments and questions:

1. I suggest you shorten Section 2.1 and Section 2.2 of the manuscript considerably, hereby heightening the mathematical level of exposition. Also, Appendix 5.1 should be omitted as it can be assumed well known to your audience.

2. There appears to be a number of minor typos (e.g. p. 4 line 49, p. 7 line 34, and p. 11 line 60 – in all three cases the verb is missing a past tense 'd'). Additional proof reading would be beneficial.

3. In Section 1 you provide an interesting discussion on the use of age-at-death distributions compared to mortality rates for longevity forecasting. If I understand your arguments correctly, you conclude that extrapolation methods based on death frequencies are advantageous to methods based on mortality rates (p. 5 top). How do you reach this conclusion?

4. What do a and ω of equation (1) refer to? Is this the range of the distribution?

5. In Section 2.1 and in the beginning of Section 2.3, μ refers to the true moment, while later in Section 2.3 (p. 11 top), μ now refers to the 'observed numerical values'. What do you mean with the observed values? Are these the later forecasts of empirical moments? What about $\tilde{\mu}$ in equation (16)? Is this an empirical central moment?

6. There appears to be an error in equation (15) – compare with equation (2.7) of Mead and Papanicolaou (1984).

7. Based on which observations and predictions are you computing the out-of-sample performance? You are mentioning and the use of six different accuracy measures and the aggregation (by averaging) over all scenarios (appearing by moving the evaluation windows 1 year forward). But how do you arrive at a value for a specific accuracy

measure and scenario? What are the observation(s) and prediction(s) you compare? I suggest you add details providing a precise explanations.

8. In footnote 2, you write that the indicators evaluated in this manuscript are the life expectancies. Does this in any way relate to my comment/question 7)?

9. You do not discuss cohort effects, see e.g. Renshaw and Haberman (2006). How does the MEM model compare to extrapolation based methods for mortality rates allowing for cohort effects? Can the MEM model capture cohort effects? I strongly suggest you to elaborate on these aspects. Furthermore, including a cohort model, e.g. the extension of the Lee-Carter model from Renshaw and Haberman (2006), in the model comparison would significantly improve the practical relevance of the model comparison.

10. In Section 4 you conclude that the main advantage of the MEM model is that age-specific trends are no longer based on the assumption of constant changes in mortality (as e.g. in the Lee-Carter model). How do you reach this conclusion? Could another (main) advantage of the MEM model not be that it is able to capture cohort effects?

References (besides those cited in the manuscript)

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