Fast Estimation of the Renshaw-Haberman Model and Its Variants

Yiping Guo & Johnny Li

University of Waterloo

2023-09-07

Motivation

- Three main components in mortality modeling: Age, Time, **Cohort**.
  - Importance of cohort effects (particularly for UK data).
- Challenges of stochastic mortality models involving cohort effects.
  - Slow model fitting process.
  - Limited types of model fitting approaches (only likelihood-based).
- Our main objective – Develop a fast **non-likelihood-based** estimation approach.

Outline

1. Review of the Lee-Carter model and Renshaw-Haberman model (generalized APC model).
2. Main proposed methodology.
3. Numerical studies.
4. Conclusion and further research.

Review: The Lee-Carter Model

The Lee-Carter model (Lee and Carter, 1992):

$$y_{x,t} := \log(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \tag{1}$$

with $\sum_x b_x = 1$ and $\sum_t k_t = 0$,

- $m_{x,t}$: Central mortality rate for age x and time t.
- a_x: Average log mortality rate for age x.
- b_x: Age effect (sensitivity) for age x.
- k_t: Time trend.

It involves two stage:

1. Estimating a_x, b_x and k_t;
2. Time series modeling on $\{k_t\}$ and forecasting.

Two Parameter Estimation Methods for the Lee-Carter Model

1. SVD, non-likelihood-based (Lee and Carter, 1992):

$$\min_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{k})} \sum_{x,t} \left(y_{x,t} - (a_x + b_x k_t)\right)^2 . \tag{2}$$

   The solution is obtained via PCA:

$$\hat{\boldsymbol{a}} = \bar{\boldsymbol{y}}, \quad \hat{\boldsymbol{b}} = \frac{\boldsymbol{u}}{\boldsymbol{1}^T\boldsymbol{u}}, \quad \hat{\boldsymbol{k}} = (\boldsymbol{1}^T\boldsymbol{u}) \cdot (\boldsymbol{Y} - \bar{\boldsymbol{Y}})^T \boldsymbol{u}. \tag{3}$$

2. Poisson regression, likelihood-based (Brouhns et al., 2002):

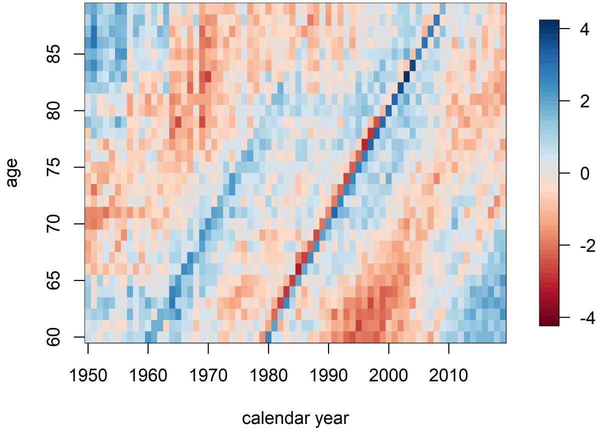
$$\max_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{k})} \sum_{x,t} \left(D_{x,t}(a_x + b_x k_t) - N_{x,t} e^{a_x + b_x k_t} \right) . \tag{4}$$

   The MLE is obtained via the iterative Newton-Raphson.

Comparison of the two methods: Interpretability? Flexibility? Computation?

Is the Lee-Carter Model Good Enough?

Residual heatmap of England and Wales data:



Cohort effects (diagonal) are significant!

Cohort Effects: The Renshaw-Haberman Model

The Renshaw-Haberman model or generalized APC model (Renshaw and Haberman, 2006):

$$y_{x,t} := \log(m_{x,t}) = a_x + b_x k_t + c_x \gamma_{t-x}, \tag{5}$$

with $\sum_x b_x = \sum_x c_x = 1$ and $\sum_t k_t = \sum_{t-x} \gamma_{t-x} = 0$.

- c_x: Age effect (sensitivity) with respect to the cohort effects.
- γ_{t-x}: Cohort effects.

Simplified variants of the model can be obtained by setting $c_x = 1/p$ or $b_x = c_x = 1/p$, where p is the number of ages.

Parameter Estimation and Slow Convergence

The parameter estimation is likelihood-based Poisson GLM framework (via iterative Newton-Raphson):

$$\max_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{k},\boldsymbol{c},\gamma)} \sum_{x,t} \left(D_{x,t}(a_x + b_x k_t + c_x \gamma_{t-x}) - N_{x,t} e^{a_x + b_x k_t + c_x \gamma_{t-x}} \right). \tag{6}$$

- Critical issue: Slow convergence rate (Cairns et al., 2009, 2011; Haberman and Renshaw, 2009, 2011).
- It hinders the use Monte-Carlo or bootstrap to examine parameter/model uncertainty.
- Some literature has worked on improving the convergence: Hunt and Villegas (2015) and Currie (2016).

About this Project

To the best of our knowledge, all the existing estimation methods are likelihood-based.

Our objectives:

1. Propose a **non-likelihood-based** (PCA-based) estimating method for the Renshaw-Haberman model and its variants.
2. Accelerate the parameter estimation process.

Non-Likelihood Parameter Estimation: Key Idea

The non-likelihood-based approach to fit the Lee-Carter model is via SVD or PCA:

$$\min_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{k})} \sum_{x,t} \left(y_{x,t} - (a_x + b_x k_t)\right)^2 . \tag{7}$$

Similarly, we aim at **directly minimizing the L^2 error** of the estimated log mortality rates under the RH framework:

$$\min_{(\boldsymbol{a},\boldsymbol{b},\boldsymbol{k},\boldsymbol{c},\gamma)} \sum_{x,t} \left(y_{x,t} - (a_x + b_x k_t + c_x \gamma_{t-x})\right)^2 . \tag{8}$$

No close-form solution exists, due to the non-orthogonality.

Main Algorithm: Alternating Minimization

- The minimization of (8) is via an iterative scheme:
  1. Set initial values of $\boldsymbol{\theta} := (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{k}, \boldsymbol{c}, \gamma)$.
  2. Fixing $\boldsymbol{b}$, $\boldsymbol{k}$, $\boldsymbol{c}$ and γ, update $\boldsymbol{a}$:

$$\min_{\boldsymbol{a}} \sum_{x,t} [\underbrace{(y_{x,t} - b_x k_t - c_x \gamma_{t-x})}_{\text{given}} - a_x]^2. \quad \textbf{(Easy)} \qquad (9)$$

  3. Fixing $\boldsymbol{a}$, $\boldsymbol{c}$ and γ, update $\boldsymbol{b}$ and $\boldsymbol{k}$:

$$\min_{(\boldsymbol{b},\boldsymbol{k})} \sum_{x,t} [\underbrace{(y_{x,t} - a_x - c_x \gamma_{t-x})}_{\text{given}} - b_x k_t]^2. \quad \textbf{(Easy)} \qquad (10)$$

  4. Fixing $\boldsymbol{a}_x$, $\boldsymbol{b}$ and $\boldsymbol{k}$, update $\boldsymbol{c}$ and γ:

$$\min_{(\boldsymbol{c},\gamma)} \sum_{x,t} [\underbrace{(y_{x,t} - a_x - b_x k_t)}_{\text{given}} - c_x \gamma_{t-x}]^2. \quad \textbf{(Difficult)} \qquad (11)$$

  5. If the objective function has not converged, go back to Step 2.

Parameter Estimation: Challenges in the Optimization

Step 4 is much more challenging and has no explicit solution:

$$\min_{(\boldsymbol{c},\gamma)} \sum_{x,t} [\underbrace{(y_{x,t} - a_x - b_x k_t)}_{\text{given}} - c_x \gamma_{t-x}]^2.$$

But it turns out that Step 4 is equivalent to solving a **PCA with missing values**.

Cohort-effect Estimation: Why PCA with Missing Values?

Denote the age range and time range by $(x_1, \cdots, x_p)$ and $(t_1, \cdots, t_n)$. Letting $z_{x,t} := y_{x,t} - a_x - b_x k_t$, we can transform the age-time matrix to the new age-cohort matrix[1]

$$\begin{bmatrix} \times & \times & \cdots & \cdots & \times & z_{1,1} & \cdots & z_{1,n-1} & z_{1,n} \\ \times & \times & \cdots & \cdots & z_{2,1} & z_{2,2} & \cdots & z_{2,n} & \times \\ \vdots & \vdots & & \iddots & & & \iddots & \vdots & \vdots \\ \vdots & \vdots & \iddots & & & \iddots & & \vdots & \vdots \\ \times & z_{p-1,1} & \cdots & z_{p-1,n-1} & z_{p-1,n} & \cdots & \cdots & \times & \times \\ z_{p,1} & z_{p,2} & \cdots & z_{p,n} & \times & \cdots & \cdots & \times & \times \end{bmatrix}$$

[1]This is guaranteed by the one-to-one transformation between (x, t) and $(x, t - x)$. Also, $\times$ represents missing values.

Cohort-effect Estimation: PCA with Missing Values - Formulation and Algorithms

- Letting $s = t - x$, the PCA with missing values problem can be formulated as:

$$\min_{(\boldsymbol{a},\boldsymbol{c},\gamma)} \sum_{(x,t-x)\in\mathcal{O}} [z_{x,t-x} - c_x \gamma_{t-x}]^2, \tag{12}$$

where $\mathcal{O}$ is the set of the indices of the observed values.

- Different approaches exist to solve PCA with missing values:
  1. **Small-scale: Iterative SVD (focus of this project);**
  2. Large-Scale and sparse: matrix completion with nuclear norm regularization (Mazumder et al., 2010), etc.

Cohort-effect Estimation: Iterative SVD Algorithm

Iterative SVD algorithm (also known as hard-imputation):

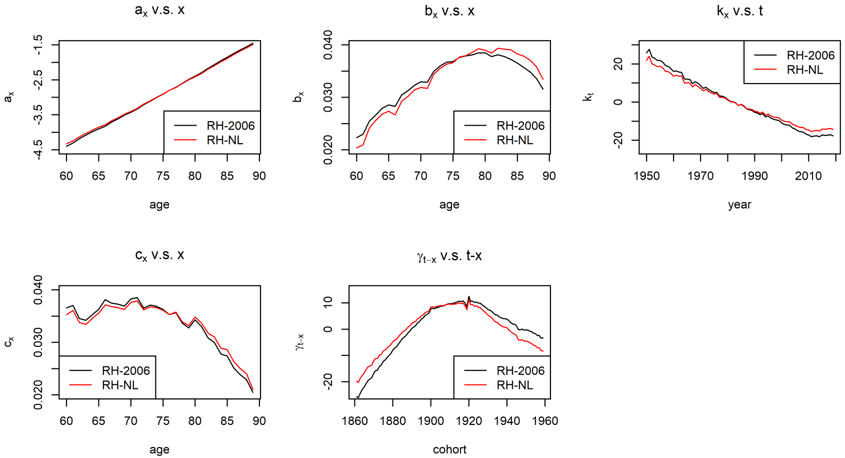
1. Initialization: Impute the missing values and initialize the approximate complete matrix $\boldsymbol{Z}_c$.
2. Approximate PCA: Implement PCA via SVD to the approximate complete matrix $\boldsymbol{Z}_c$.
3. Imputation: Update the missing values by using the corresponding PCA reconstructions in Step 2.
4. Repeat Steps 2 and 3 until the convergence criterion is satisfied.

We proved the convergence of the iterative SVD algorithm within our context.

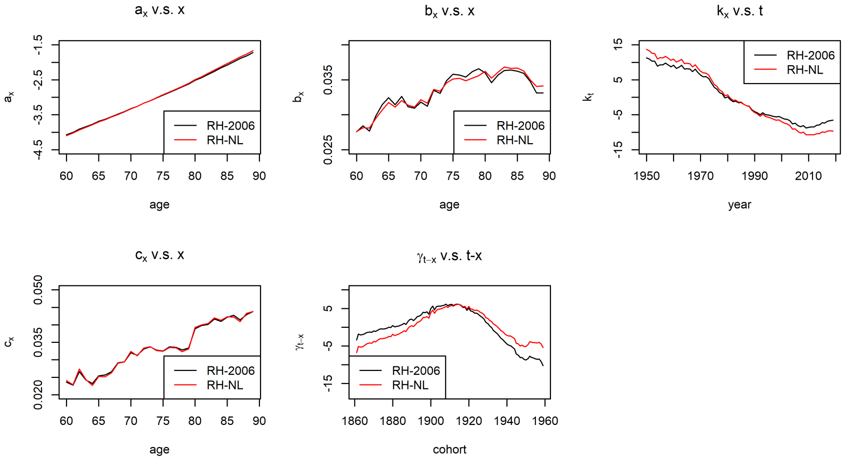
Experimental Design

- In this talk, we compare two methods:
  - *RH-2006*: Standard iterative Newton-Raphson method (Renshaw and Haberman, 2006).
  - *RH-NL*: Proposed non-likelihood-based method.
- Data set: England & Wales (EW) and US male mortality rates, from 1950 to 2019, age from 60-89.
- We compare the fitted L^2 errors, log-likelihoods and computation time.

Plots: RH Model, EW Data



Plots: RH Model, US Data



Numerical Results

RH Model	Data	*RH-2006*	*RH-NL*
L^2 error	EW	0.578	0.565
	US	0.472	0.465
Log-likelihood	EW	−12843	−12890
	US	−17736	−17828
Time (sec)	EW	336.68	38.72
	US	228.72	10.44

We have chosen the same tolerance levels (relative changes of the objective function) to make the comparison fair.

Conclusion and Future Research

- We proposed a non-likelihood-based estimating method for the Renshaw-Haberman model.
- The proposed method is significantly faster than the traditional likelihood-based method, while producing satisfactory estimation results.
- Possible future research:
    - Further accelerate the algorithm by incorporating additional identification constraints.
    - Extension to multi-component structure, with proper regularization.

Reference I

Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, 31(3), 373-393.

Cairns, A. J., 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), 1-35.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48(3), 355-367.

Reference II

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

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

Haberman, S., & Renshaw, A. (2011). A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1), 35-55.

Hunt, A., & Villegas, A. M. (2015). Robustness and convergence in the Lee–Carter model with cohort effects. *Insurance: Mathematics and Economics*, 64, 186-202.

Reference III

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

Mazumder, R., Hastie, T., & Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, 11, 2287-2322.

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.