Efficient use of data for LSTM mortality forecasting

Lina Palmborg, Stockholm University

based on joint work with M. Lindholm

June 16, 2021

Efficient use of data for LSTM mortality forecasting

Combining standard, simple mortality models with forecasts by neural network models

Main contributions

- Focus on procedures for using training data efficiently
- Suggest two alternative approaches for splitting data into training data and validation data, combined with ensembling
- Show that using untransformed data for the time varying index of mortality when training the neural network is not appropriate
- Suggest a boosted version of the model which stabilises long-term predictions, and still retains improved short-term performance for populations where non-linearities are present

Mortality forecasting

- Life expectancy has increased dramatically over the last century
- Challenge for life insurance industry and social security systems
- Appropriate mortality forecasts are essential for accurate estimation of future costs in both reserving and pricing of life insurance products

Lee-Carter model

The perhaps most famous mortality forecasting model is the Lee-Carter model (Lee et al., 1992)

- $\widehat{\mu}_{x,t}$ estimate of the mortality rate for age x during calendar year t,

$$\log(\widehat{\mu}_{x,t}) = \alpha_x + \beta_x \kappa_t.$$

- To produce forecasts, estimated κ_ts are modelled as a random walk with drift

$$\widehat{\kappa}_{t+1} = \gamma + \widehat{\kappa}_t + \epsilon_{t+1},$$

where $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$ and i.i.d.

Poisson Lee-Carter model

- We observe death counts, not mortality rates
- $D_{x,t}$ number of individuals dying being of age x during calendar year t
- $r_{x,t}$ total exposure-to-risk for individuals being x years old during calendar year t
- The Poisson Lee-Carter model (Brouhns et al., 2002):

$$D_{x,t} \mid r_{x,t} \sim \text{Po}(r_{x,t}\mu_{x,t}(\boldsymbol{\theta})),$$

$$\mu_{x,t}(\boldsymbol{\theta}) := \exp\{\alpha_x + \beta_x \kappa_t\}$$

- Again, to produce forecasts, estimated κ_ts are modelled as a random walk with drift

Poisson Lee-Carter LSTM model

Linear time series models not adequate for modelling estimated κ_ts for all populations.

- Generalised Lee-Carter:

$$\widehat{\kappa}_{t+1} = f(\mathcal{F}_t; \boldsymbol{\eta}) + \epsilon_{t+1},$$

where $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$ and i.i.d., $\mathcal{F}_t = \sigma\{\widehat{\kappa}_s, s \leq t\}$, and

$$f(\mathcal{F}_t; \boldsymbol{\eta}) := \mathbb{E}[\widehat{\kappa}_{t+1} \mid \mathcal{F}_t](\boldsymbol{\eta}).$$

- $f(\mathcal{F}_t; \boldsymbol{\eta})$ is modelled as a long short-term memory (LSTM) neural network (Hochreiter et al., 1997).

Boosted model

1. Model the estimated κ_ts as a "standard" time series model, with mean-function $h(\mathcal{F}_t; \boldsymbol{\xi})$, and parameter estimate $\widehat{\boldsymbol{\xi}}$
2. Model the residual as an LSTM model

Boosted Poisson Lee-Carter LSTM model:

$$\widehat{\kappa}_{t+1} = h(\mathcal{F}_t; \widehat{\boldsymbol{\xi}}) + f(\mathcal{F}_t; \boldsymbol{\eta}) + \epsilon_{t+1},$$

where $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$ and i.i.d., $f(\mathcal{F}_t; \boldsymbol{\eta})$ is modelled as an LSTM model, and where $h(\mathcal{F}_t; \widehat{\boldsymbol{\xi}})$ acts as an $\mathcal{F}_t$-measurable (non-trainable) intercept function in the LSTM model

Long short-term memory neural networks

- Recurrent neural networks are specialised at processing sequential data
  - “Built in” concept of time
  - Standard recurrent neural networks struggle with learning long-term dependencies (vanishing gradient problem)
- LSTM is a gated recurrent neural network
  - Gates control the flow of information through the network
  - Cell state represents the long-term memory
  - Better able at learning long-term dependencies
  - Natural model class to consider for time series modelling

Early stopping

- Neural network models are in general calibrated using iterative procedures
- After how many iterations should this procedure stop?
  - Increased number of iterations $\Longrightarrow$ better fit to training data
  - Problem: might not generalise well to unseen data
  - Solution: early stopping
- Early stopping means that the procedure is stopped when the performance on an unseen data set (validation set) starts deteriorating
  - Need to split data into
    - one set used for in-sample training
    - one set used for validation (out-of-sample training)
    - one set used for testing the fitted model

Efficient use of data

Three approaches used for splitting data into training and validation set

(LO) The standard approach of withholding the last fraction of observations as validation data

(RT) Sampling observations randomly in time

(SP) Sampling individuals and randomly assigning them to subsets of the underlying population, using one subset for in-sample training, and another subset for out-of-sample validation, without splitting in the time dimension

Combined with ensembling to stabilise predictive performance

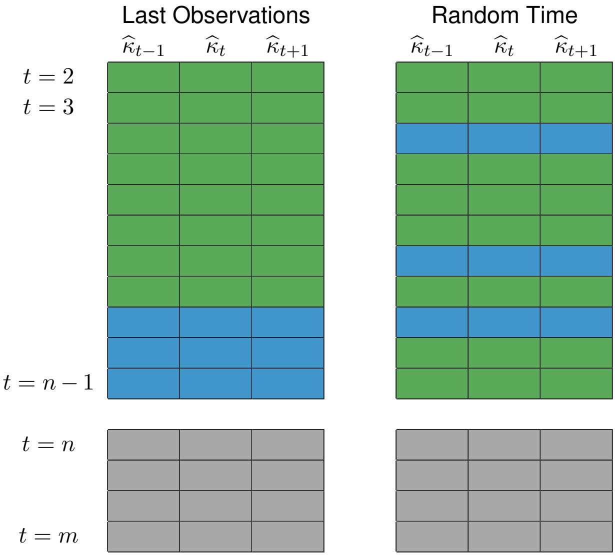
Calibration approaches

Data is structured according to

$$\begin{bmatrix} \widehat{\kappa}_1 & \widehat{\kappa}_2 & \dots & \widehat{\kappa}_p & \widehat{\kappa}_{p+1} \\ \widehat{\kappa}_2 & \widehat{\kappa}_3 & \dots & \widehat{\kappa}_{p+1} & \widehat{\kappa}_{p+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \widehat{\kappa}_{n-p} & \widehat{\kappa}_{n-p+1} & \dots & \widehat{\kappa}_{n-1} & \widehat{\kappa}_n \end{bmatrix}$$

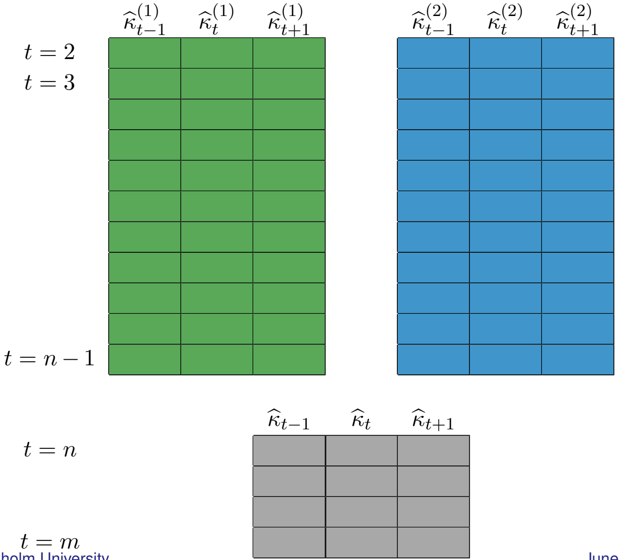
- The first p columns correspond to the input sequences
- The last column corresponds to the output that should be predicted by the model

Calibration approaches



Calibration approaches

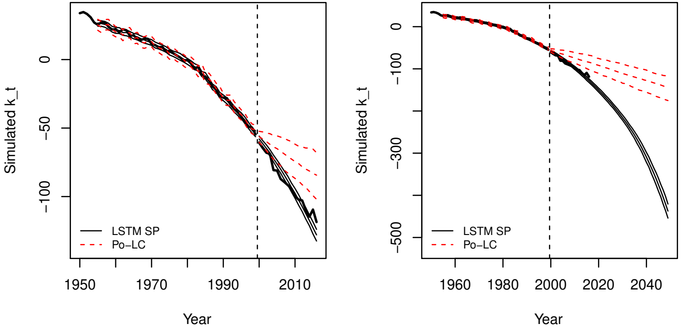
Split Population



Italian men, κ_t in-sample and out-of-sample

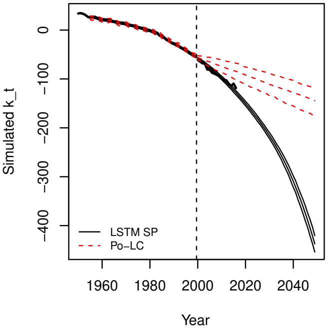
Non-boosted model

Figure: Left: κ_t short-term prediction. Right: κ_t long-term prediction.



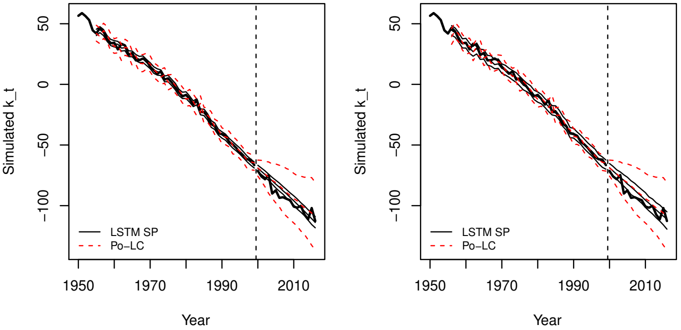
Italian men, κ_t long-term prediction

Figure: Left: Non-boosted model. Right: Boosted model.



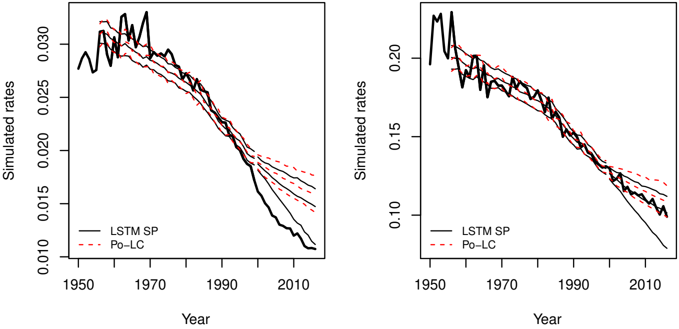
Italian women, κ_t in-sample and out-of-sample

Figure: Left: Non-boosted model. Right: Boosted model.



Italian men, mortality rates with boosted model

Figure: Left: age 65. Right: age 85.



Log-likelihood out-of-sample for boosted model.

	Po-LC	LO	RT	SP
ITA male	-105 612	-91 232	-78 431	**-78 139**
ITA female	-23 112	-19 012	-17 614	**-16 986**
SWE male	-16 183	-13 580	-14 062	**-12 947**
SWE female	-8 019	-9 303	**-7 616**	-8 170
USA male	-224 054	**-218 720**	-220 277	-236 334
USA female	-78 182	-95 775	-85 316	**-82 625**

Main references

- Preprint available at SSRN:
  https://papers.ssrn.com/abstract=3805843
- R. D. Lee and L. R. Carter (1992), Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419):659–671.
- N. Brouhns, M. Denuit, J. K. Vermunt (2002), A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insur Math Econ* 31, 373-393.
- S. Hochreiter and J. Schmidhuber (1997), Long short-term memory. *Neural computation*, 9(8):1735–1780.