

Deepening Lee-Carter for Longevity Projections with Uncertainty Estimation

Mario Marino¹, Susanna Levantesi, Andrea Nigri

¹ Department of Statistics, Sapienza University of Rome
m.marino@uniroma1.it

16th International Longevity Risk and Capital Markets Solutions
Conference

14 August 2021

Agenda

- ▶ Introduction
- ▶ Improving the Lee-Carter mortality density forecast
- ▶ Numerical application
- ▶ Conclusions

Mortality forecasting and Neural Networks

- Predicting mortality continues to be a challenge for demographers and actuaries
- Nowadays, several (extrapolative) stochastic mortality models are available
 - Lee-Carter (LC) family
 - Cairns-Blake-Dowd (CBD) family
- Methodological advances in mortality forecasting based on Machine and Deep Learning models
 - Random forest, Gradient Boosting
 - Feed-forward, Convolutional, Recurrent Neural Networks (NN)
- Exploring NN models as predictors, the idea is to create a novel approach: the Mortality Neural Forecasting

Mortality forecasting and Neural Networks

- The present work follows and completes the study in Nigri et al. (2019)
- The approach is the following:
 - A. Consider a stochastic mortality model as reference model to fit the observed mortality surface
 - B. Forecast future mortality paths by a proper NN model
- The overall mortality model is hybrid, achieving:
 1. Ease of interpretation of age-period-cohort parameters
 2. Accuracy in estimating future mortality outcomes

Model proposal: the LC-LSTM

The LC model as reference model and the Long Short-Term Memory (LSTM) network to improve the LC mortality density forecast

Agenda

- ▶ Introduction
- ▶ Improving the Lee-Carter mortality density forecast
- ▶ Numerical application
- ▶ Conclusions

The LC-LSTM model - Formulation

► **Reference model:** LC Poisson model (Brouhns et al.(2002)).

For $x \in \mathcal{X} = \{0, 1, \dots, \omega\}$ and $t \in \mathcal{T} = \{t_0, t_1, \dots, t_n\}$, we have

$D_{x,t} \sim Poi(E_{x,t}^c m_{x,t})$ and

$$\ln m_{x,t} = \alpha_x + \beta_x k_t. \quad (1)$$

► Let $\kappa_{\mathcal{T}} = (k_{t-j})_{t \in \mathcal{T}}$ be the vector of the time lagged k_t , being $j \in \mathbb{N}$ the time lag, we consider:

$$k_t = f_{LSTM}(\kappa_{\mathcal{T}}; \mathcal{W}) + \gamma_t, \quad (2)$$

with $f_{LSTM} : \mathbb{R}^j \rightarrow \mathbb{R}$ the LSTM function and \mathcal{W} the weights.

► Over the forecasting horizon $\mathcal{T}' = \{t_n + 1, t_n + 2, \dots, t_n + s\}$, the **LC-LSTM model** expression is:

$$\ln m_{x,t} = \hat{\alpha}_x + \hat{\beta}_x (f_{LSTM}(\kappa_{\mathcal{T}'}; \mathcal{W}) + \gamma_t), \quad \forall t \in \mathcal{T}', \quad (3)$$

with $\hat{\alpha}_x$ and $\hat{\beta}_x$ the estimates of age-dependent parameters.

The LC-LSTM model - Point Predictions

- From a general perspective, the LC time-index values should be interpreted as the realization of the following process:

$$k_t = \varphi(\boldsymbol{\kappa}_{\mathcal{T}}) + \gamma_t, \quad \forall t \in \mathcal{T}, \quad (4)$$

where the unknown function $\varphi : \mathbb{R}^j \rightarrow \mathbb{R}$ maps the vector $\boldsymbol{\kappa}_{\mathcal{T}}$ to k_t over the time horizon \mathcal{T} , unless the noise component.

- The **network approximates** $\varphi(\boldsymbol{\kappa}_{\mathcal{T}})$ according to the time-index history:

$$\hat{k}_t = \hat{f}_{LSTM}(\boldsymbol{\kappa}_{\mathcal{T}}; \hat{\mathcal{W}}) = \mathbb{E}(k_t | \boldsymbol{\kappa}_{\mathcal{T}}) \quad (5)$$

- Therefore, the **LC-LSTM model** provides the following **point predictions**:

$$\ln \hat{m}_{x,t} = \mathbb{E}(\ln m_{x,t}) = \hat{\alpha}_x + \hat{\beta}_x \hat{f}_{LSTM}(\boldsymbol{\kappa}_{\mathcal{T}'}; \hat{\mathcal{W}}), \quad \forall t \in \mathcal{T}'. \quad (6)$$

The LC-LSTM model - Uncertainty estimation

- Point predictions could be a poor information and a **measure of uncertainty**, such as **prediction intervals**, is necessary
- Exploiting the **bias-variance trade-off principle**, the total variance associated to time-index values is:

$$\sigma_{k_t}^2 = \sigma_{\hat{k}_t}^2 + \sigma_{\gamma}^2 + \mathbb{E} \left[\text{BIAS} \left(\hat{k}_t | \kappa_{T'} \right)^2 \right] \quad (7)$$

where $\sigma_{\hat{k}_t}^2$ is the NN output variance, σ_{γ}^2 is the noise variance and $\text{BIAS} \left(\hat{k}_t | \kappa_{T'} \right) = \mathbb{E} \left(\varphi \left(\kappa_{T'} \right) - \hat{k}_t | \kappa_{T'} \right)$.

The LC-LSTM - estimating $\sigma_{\hat{k}_t}^2$

- To derive $\sigma_{\hat{k}_t}^2$ the conditioned time-index distribution, $\mathbb{P}\left(\hat{k}_t | \kappa_{\mathcal{T}'}\right)$ should be known, but it is not available.
- We could extract it from the data referring to the **ensemble technique**
- Using bootstrap techniques, multiple training data samples are generated to develop an **empirical distribution**, $\hat{\mathbb{P}}\left(\hat{k}_t | \kappa_{\mathcal{T}'}\right)$
- The final estimates are then obtained aggregating, by average, the various outputs: **bootstrap aggregating or bagging**
- In bagging procedures holds that $\mathbb{E}\left[BIAS\left(\hat{k}_t | \kappa_{\mathcal{T}'}\right)^2\right] \rightarrow 0$

The LC-LSTM - estimating $\sigma_{\hat{k}_t}^2$

The bagging scheme:

- Step 1.* Over the series $\kappa_{\mathcal{T}}$, we train the LSTM model obtaining point estimates over \mathcal{T}'
- Step 2.* Generate $B \in \mathbb{N}$ samples of $\kappa_{\mathcal{T}}$ via a proper bootstrap procedure
- Step 3.* For each sample re-optimize the weights of NN defined in *Step 1*
- Step 4.* For each NN in *Step 3*, predict the associate point estimate on \mathcal{T}' , producing a **bootstrap distribution consisting of B point predictions**:

$$\hat{\mathbb{P}}\left(\hat{k}_t | \kappa_{\mathcal{T}'}\right) = \left(\hat{k}_t^{(b)} = \hat{f}_{LSTM}\left(\kappa_{\mathcal{T}}^{(b)}, \hat{\mathcal{W}}^{(b)}\right), b = 1, \dots, B\right) \quad (8)$$

The LC-LSTM - estimating $\sigma_{\hat{k}_t}^2$

Step 5. From $\hat{\mathbb{P}}\left(\hat{k}_t | \kappa_{\mathcal{T}'}\right)$ calculates estimates of interest by aggregation. Hence, the **bagged estimate of the variance** $\sigma_{\hat{k}_t}^2$ is:

$$\hat{\sigma}_{\hat{k}_t}^2 = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{f}_{LSTM} \left(\kappa_{\mathcal{T}}^{(b)}, \hat{\mathcal{W}}^{(b)} \right) - \bar{k}_t \right)^2, \quad (9)$$

where

$$\bar{k}_t = \frac{1}{B} \sum_{b=1}^B \hat{f}_{LSTM} \left(\kappa_{\mathcal{T}}^{(b)}, \hat{\mathcal{W}}^{(b)} \right)$$

is the **bagged estimate for the conditional expectation** $\mathbb{E}\left(\hat{k}_t | \kappa_{\mathcal{T}'}\right)$.

The LC-LSTM - estimating σ_γ^2

- Mortality dynamic incorporates an intrinsic randomness not explained by the network: **the noise** γ
- A NN appropriately trained catches the key input-output data schemes, skimming noisy examples (avoid overfitting)
- Let \mathcal{T} be the training set interval, we deal with the series

$$(\gamma_t)_{t \in \mathcal{T}} = (k_t - \hat{k}_t)_{t \in \mathcal{T}}$$

as **a proxy of the unwrapped noise by NN**

- It helps to evaluate the estimates $\hat{\sigma}_\gamma^2$ as the time-index residual uncertainty over \mathcal{T} , spreading the random error over the forecast horizon \mathcal{T}' through a random walk representation

Agenda

- ▶ Introduction
- ▶ Improving the Lee-Carter mortality density forecast
- ▶ **Numerical application**
- ▶ Conclusions

Numerical application

- Countries: Australia, Spain and Japan. Data from HMD, both genders
- Ages: $\mathcal{X} = \{0, \dots, 99\}$
- Calendar Years: Two periods to check model robustness
 $\mathcal{T} = \{1950, \dots, 2018\}$ and $\mathcal{T} = \{1960, \dots, 2018\}$
- Lag $J = 1$, so that $\kappa_{\mathcal{T}} = (k_{t-1})_{t \in \mathcal{T}}$ and
 $k_t = f_{LSTM}(k_{t-1}; \mathcal{W}) + \gamma_t$
- For the bagging scheme: bootstrap from Koissi et al.(2006), with $B = 1000$
- Benchmark: LC Poisson model (Brounhs et al.(2002)), selecting the best ARIMA(p,d,q) model.

Numerical application - Learning process

- Setting $T = 2000$ as forecasting year, the series κ_T is splitted in:

$$\text{TRAINING SET: } \mathcal{TR} = (k_t | k_{t-1})_{t=t_0, \dots, T} \quad (10)$$

$$\text{TESTING SET: } \mathcal{TS} = (k_t | k_{t-1})_{t=T+1, \dots, t_n},$$

where $t_0 = \{1950, 1960\}$.

- To validate the model we divide the \mathcal{TR} set into a sub-training set and in a validation set, considering the splitting rule 80% – 20%:

$$\text{SUB-TRAINING SET: } \mathcal{TR}^{\text{sub}} = (k_t | k_{t-1})_{t=t_0, \dots, T^{\text{sub}}} \quad (11)$$

$$\text{VALIDATION SET: } \mathcal{VS} = (k_t | k_{t-1})_{t=T^{\text{sub}}+1, \dots, T}$$

- Tuning by grid search

Numerical application - Performance metrics

- Accuracy metrics for point predictions

$$RMSE = \sqrt{\frac{\sum_{t=t_n+1}^{t_n+s} (\ln m_{x,t} - \ln \hat{m}_{x,t})^2}{s-1}}$$

- Quality metrics for prediction interval

Prediction Interval Coverage Probability

$$PICP = \frac{1}{s-1} \sum_{t=t_n+1}^{t_n+s} \mathbf{1}_{\{\ln m_{x,t} \in [\ln \hat{m}_{x,t}^L, \ln \hat{m}_{x,t}^U]\}},$$

Mean Prediction Interval Width

$$MPIW = \frac{1}{s-1} \sum_{t=t_n+1}^{t_n+s} \ln \hat{m}_{x,t}^U - \ln \hat{m}_{x,t}^L,$$

with $\mathbf{1}_{\{\cdot\}} = 1$ if $\ln m_{x,t} \in [\ln \hat{m}_{x,t}^L, \ln \hat{m}_{x,t}^U]$, and $\mathbf{1}_{\{\cdot\}} = 0$ otherwise.

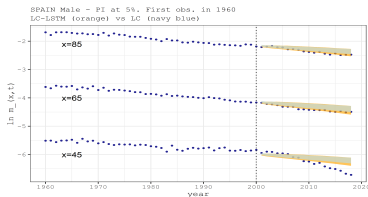
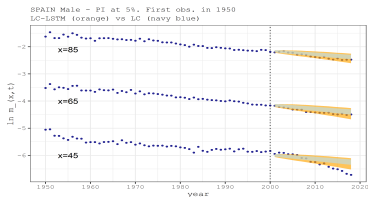
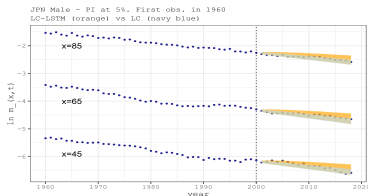
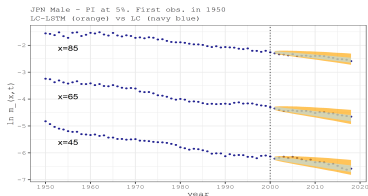
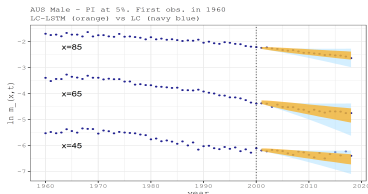
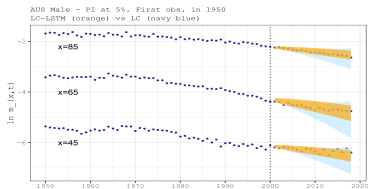


Figure 1: MALE PI ($\alpha = 5\%$). Forecasting period: 2001-2018. Training period: 1950-2000 (left), 1960-2000 (right).

Numerical application - Long Term Forecasts

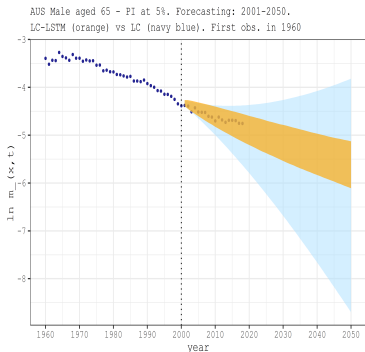
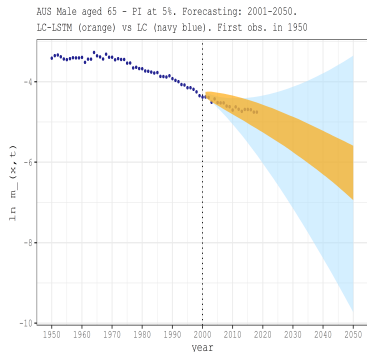


Figure 2: AUSTRALIA MALE ($\alpha = 5\%$). Age 65. Forecasting period: 2001-2050. Training period: 1950-2000 (left), 1960-2000 (right).

Agenda

- ▶ Introduction
- ▶ Improving the Lee-Carter mortality density forecast
- ▶ Numerical application
- ▶ **Conclusions**

Conclusions

- The LC-LSTM seems to be an effective improvement of the canonical LC model predictive ability, in terms of **both point and interval predictions**
- The proposed model reflects important features, also in the long-run, as:
 - **Robustness** w.r.t. to the training period
 - **Biologically consistency**
 - **Plausibility in uncertainty levels**
- *The LC-LSTM model poses a compromise between the interpretation of the mortality phenomenon and high precision in anticipating its future realizations*

Thanks for your attention!

Contact: m.marino@uniroma1.it

Check out the paper and references therein:

Marino, M., Levantesi, S., Nigri, A. (2021). *Deepening Lee-Carter for longevity projections with uncertainty estimation.*

<https://arxiv.org/abs/2103.10535>