Point and Interval Forecasts of Death Rates Using Neural Networks

Simon Schnürch, Ralf Korn
Insurance Data Science Conference, 18 June 2021

Fraunhofer
ITWM

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Good mortality models are needed in many applications.

Accurate mortality forecasts

- are important for pension systems, insurance companies, governments, ...
- are not always achieved by classical methods,
- should give an impression of the possible distribution of future mortality rates.

We propose a convolutional neural network (CNN) for mortality forecasting along with a reliable method for quantifying its prediction uncertainty.

CNNs capture two-dimensional structure in mortality rates.

- We propose to use a bootstrap ensemble of two-dimensional CNNs for forecasting death rates.
- A similar approach has been successfully investigated by Perla et al. [2021].

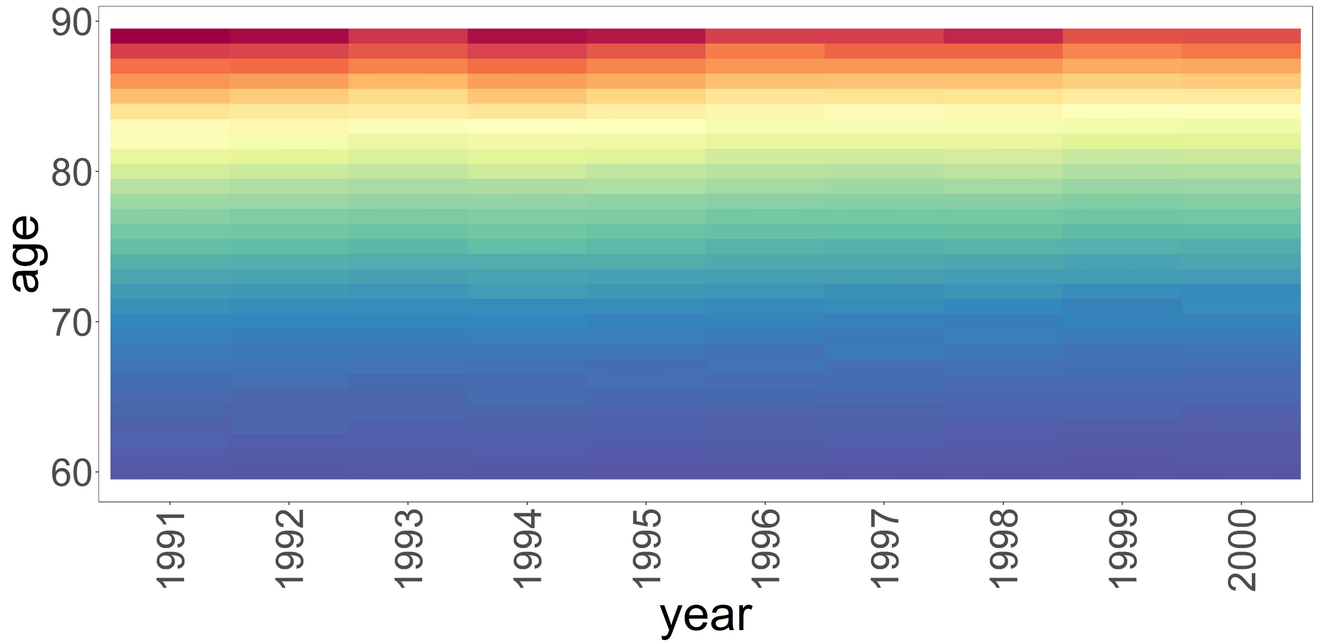

Figure: Heat map displaying the death rates of the female population of England & Wales between 1991 and 2000 for ages 60 to 89. (blue = low, red = high).

In contrast to previous neural network applications in mortality forecasting, we quantify prediction uncertainty.

Goal:

$$\mathbb{P}\left(\hat{m}_{x,t}^{i,\text{ lower}} \leq m_{x,t}^{i} \leq \hat{m}_{x,t}^{i,\text{ upper}}\right) \geq a \text{ for some large } a \in [0,1].$$

Assumption:

$$\log m_{x,t}^{i} = \log m_{x,t}^{i,\text{ true}} + \varepsilon_{x,t}^{i}.$$

Bias-variance decomposition:

$$\mathbb{E}\left(\left(\log m_{x,t}^{i} - \log \hat{m}_{x,t}^{i}\right)^{2}\right) = \text{Bias}\left(\log \hat{m}_{x,t}^{i}\right)^{2} + \text{Var}\left(\log \hat{m}_{x,t}^{i}\right) + \text{Var}\left(\varepsilon_{x,t}^{i}\right).$$

We follow the approach proposed by Heskes [1996] for general FFNNs:

- Assume $\text{Bias}(\log \hat{m}) \equiv 0$.
- Estimate model uncertainty $\text{Var}\left(\log \hat{m}_{x,t}^{i}\right)$ based on empirical ensemble variance.
- Train an additional neural network to estimate noise variance $\text{Var}\left(\varepsilon_{x,t}^{i}\right)$.
- Under a normal assumption, this yields the interval bounds

$$\log \hat{m}_{x,t}^{i,\text{ lower|upper}} := \log \hat{m}_{x,t}^{i} \pm \Phi^{-1}\left(\frac{1+a}{2}\right) \widehat{\text{Var}}\left(\log m_{x,t}^{i} - \log \hat{m}_{x,t}^{i}\right).$$

In contrast to previous neural network applications in mortality forecasting, we quantify prediction uncertainty.

Goal:

$$\mathbb{P}\left(\hat{m}_{x,t}^{i,\,\text{lower}} \leq m_{x,t}^{i} \leq \hat{m}_{x,t}^{i,\,\text{upper}}\right) \geq a \text{ for some large } a \in [0,1].$$

Assumption:

$$\log m_{x,t}^{i} = \log m_{x,t}^{i,\,\text{true}} + \varepsilon_{x,t}^{i}.$$

Bias-variance decomposition:

$$\mathbb{E}\left(\left(\log m_{x,t}^{i} - \log \hat{m}_{x,t}^{i}\right)^2\right) = \mathrm{Bias}\left(\log \hat{m}_{x,t}^{i}\right)^2 + \mathrm{Var}\left(\log \hat{m}_{x,t}^{i}\right) + \mathrm{Var}\left(\varepsilon_{x,t}^{i}\right).$$

We follow the approach proposed by Heskes [1996] for general FFNNs:

- Assume $\mathrm{Bias}(\log \hat{m}) \equiv 0$.
- Estimate model uncertainty $\mathrm{Var}\left(\log \hat{m}_{x,t}^{i}\right)$ based on empirical ensemble variance.
- Train an additional neural network to estimate noise variance $\mathrm{Var}\left(\varepsilon_{x,t}^{i}\right)$.
- Under a normal assumption, this yields the interval bounds

$$\log \hat{m}_{x,t}^{i,\,\text{lower}|\text{upper}} := \log \hat{m}_{x,t}^{i} \pm \Phi^{-1}\left(\frac{1+a}{2}\right) \widehat{\mathrm{Var}}\left(\log m_{x,t}^{i} - \log \hat{m}_{x,t}^{i}\right).$$

Fraunhofer ITWM

In contrast to previous neural network applications in mortality forecasting, we quantify prediction uncertainty.

Goal:

$$\mathbb{P}\left(\hat{m}_{x,t}^{i,\,\text{lower}} \leq m_{x,t}^{i} \leq \hat{m}_{x,t}^{i,\,\text{upper}}\right) \geq a \text{ for some large } a \in [0,1].$$

Assumption:

$$\log m_{x,t}^{i} = \log m_{x,t}^{i,\,\text{true}} + \varepsilon_{x,t}^{i}.$$

Bias-variance decomposition:

$$\mathbb{E}\left(\left(\log m_{x,t}^{i} - \log \hat{m}_{x,t}^{i}\right)^2\right) = \text{Bias}\left(\log \hat{m}_{x,t}^{i}\right)^2 + \text{Var}\left(\log \hat{m}_{x,t}^{i}\right) + \text{Var}\left(\varepsilon_{x,t}^{i}\right).$$

We follow the approach proposed by Heskes [1996] for general FFNNs:

- Assume $\text{Bias}(\log \hat{m}) \equiv 0$.
- Estimate model uncertainty $\text{Var}\left(\log \hat{m}_{x,t}^{i}\right)$ based on empirical ensemble variance.
- Train an additional neural network to estimate noise variance $\text{Var}\left(\varepsilon_{x,t}^{i}\right)$.
- Under a normal assumption, this yields the interval bounds

$$\log \hat{m}_{x,t}^{i,\,\text{lower}|\text{upper}} := \log \hat{m}_{x,t}^{i} \pm \Phi^{-1}\left(\frac{1+a}{2}\right) \widehat{\text{Var}}\left(\log m_{x,t}^{i} - \log \hat{m}_{x,t}^{i}\right).$$

Fraunhofer ITWM

The CNN yields accurate point forecasts and reliable interval forecasts.

Table: Out-of-sample error measures for 54 populations, ages 60 to 89 and years 2007 to 2016 (models trained on data up to 2006), data from Human Mortality Database [2019].

Model	MSE×10^5	MAE×10^3	MdAPE[%]	PICP[%]	MPIW
LC	5.5	4.0	5.8	74.3	0.011
FFNN	2.6	3.1	6.1	92.9	0.015
RNN	5.1	3.9	5.3	89.9	0.015
CNN	3.4	3.3	5.0	97.3	0.020

- We consider feed-forward [Richman and Wüthrich, 2019a] and recurrent [Richman and Wüthrich, 2019b] neural networks (FFNNs and RNNs) and a Lee-Carter [Lee and Carter, 1992, LC] model as benchmarks.
- The FFNN performs well with respect to squared and absolute error but rather weakly in terms of the relative error.
- The CNN is the only model exceeding the required PICP of 95%.

Error measures

Point forecasts:

$$\mathrm{MSE} := \frac{1}{N}\sum_{x,t,i}\left(\hat{m}^i_{x,t} - m^i_{x,t}\right)^2,$$

$$\mathrm{MAE} := \frac{1}{N}\sum_{x,t,i}\left|\hat{m}^i_{x,t} - m^i_{x,t}\right|,$$

$$\mathrm{MdAPE} := \mathrm{median}_{x,t,i}\left\{\frac{\left|\hat{m}^i_{x,t} - m^i_{x,t}\right|}{m^i_{x,t}}\right\}\cdot 100\%.$$

Interval forecasts:

$$\mathrm{PICP} := \frac{1}{N}\sum_{x,t,i}\mathbb{1}_{\left\{m^i_{x,t}\in\left[\hat{m}^{i,\,\mathrm{lower}}_{x,t},\,\hat{m}^{i,\,\mathrm{upper}}_{x,t}\right]\right\}},$$

$$\mathrm{MPIW} := \frac{1}{N}\sum_{x,t,i}\left(\hat{m}^{i,\,\mathrm{upper}}_{x,t} - \hat{m}^{i,\,\mathrm{lower}}_{x,t}\right).$$

Fraunhofer ITWM

There is potential for further improvement.

Possible extensions and improvement ideas for our model include

- data augmentation,
- stacking, i.e., setting up combinations of model architectures (e.g., FFNN and CNN),
- further investigating explainability of the CNN (e.g., via SHAP).

For more details, see our preprint at `ssrn.com/abstract=3796051` or contact me at `simon.schnuerch@itwm.fraunhofer.de`.

References

T. Heskes. Practical Confidence and Prediction Intervals. Proceedings of the 9th International Conference on Neural Information Processing Systems, 1996.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock (Germany). Data downloaded on July 2, 2019. URL www.mortality.org.

R. D. Lee and L. R. Carter. Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association, 87 (419):659–671, 1992. ISSN 0162-1459. doi: 10.1080/01621459.1992.10475265.

F. Perla, R. Richman, S. Scognamiglio, and M. V. Wüthrich. Time-series forecasting of mortality rates using deep learning. Scandinavian Actuarial Journal, pages 1–27, 2021. ISSN 0346-1238. doi: 10.1080/03461238.2020.1867232.

R. Richman and M. V. Wüthrich. A Neural Network Extension of the Lee-Carter Model to Multiple Populations. Annals of Actuarial Science, to appear, 2019a. ISSN 1748-4995. doi: 10.1017/S1748499519000071.

R. Richman and M. V. Wüthrich. Lee and Carter go Machine Learning: Recurrent Neural Networks. Tutorial, SSRN, 2019b. URL https://ssrn.com/abstract=3441030.

S. Schnürch and R. Korn. Point and Interval Forecasts of Death Rates Using Neural Networks. Preprint, 2021. URL https://ssrn.com/abstract=3796051.