



Lee and Carter go Deep Learning

Time Series Forecasting of Mortality using Neural Networks

Recent advances in machine learning have been propelled by deep learning techniques, which are a modern approach to applying neural networks to large-scale prediction tasks. Many of these advances have been in the fields of computer vision and natural language processing, for example, the accuracy of models built to classify the 14 million images in the ImageNet database has steadily increased since 2011 (Papers with Code, 2020). Characteristically, the models used within these fields are specialized to deal with the types of data that must be processed to produce predictions. For example, when processing text data, which conveys meaning through the placement of words in a specific order, models that incorporate sequential structures are usually used.

Recently, interest in applying deep learning to actuarial topics has grown, and there is now a body of research illustrating these applications across the actuarial disciplines, including mortality forecasting. Deep learning is a promising technique for actuaries, due to the strong links between these models and the familiar technique of Generalized Linear Models (GLMs). Wüthrich (2019) discusses how neural networks can be seen as generalized GLMs, that first process the data input to the network to create new variables, which are then used in a GLM to make predictions (this is called 'representation learning'), which we illustrate in Figure 1. By deriving new features from input data, deep learning models can solve difficult problems of model specification, making these techniques promising analysing complex actuarial problems, such as multi-population mortality forecasting.

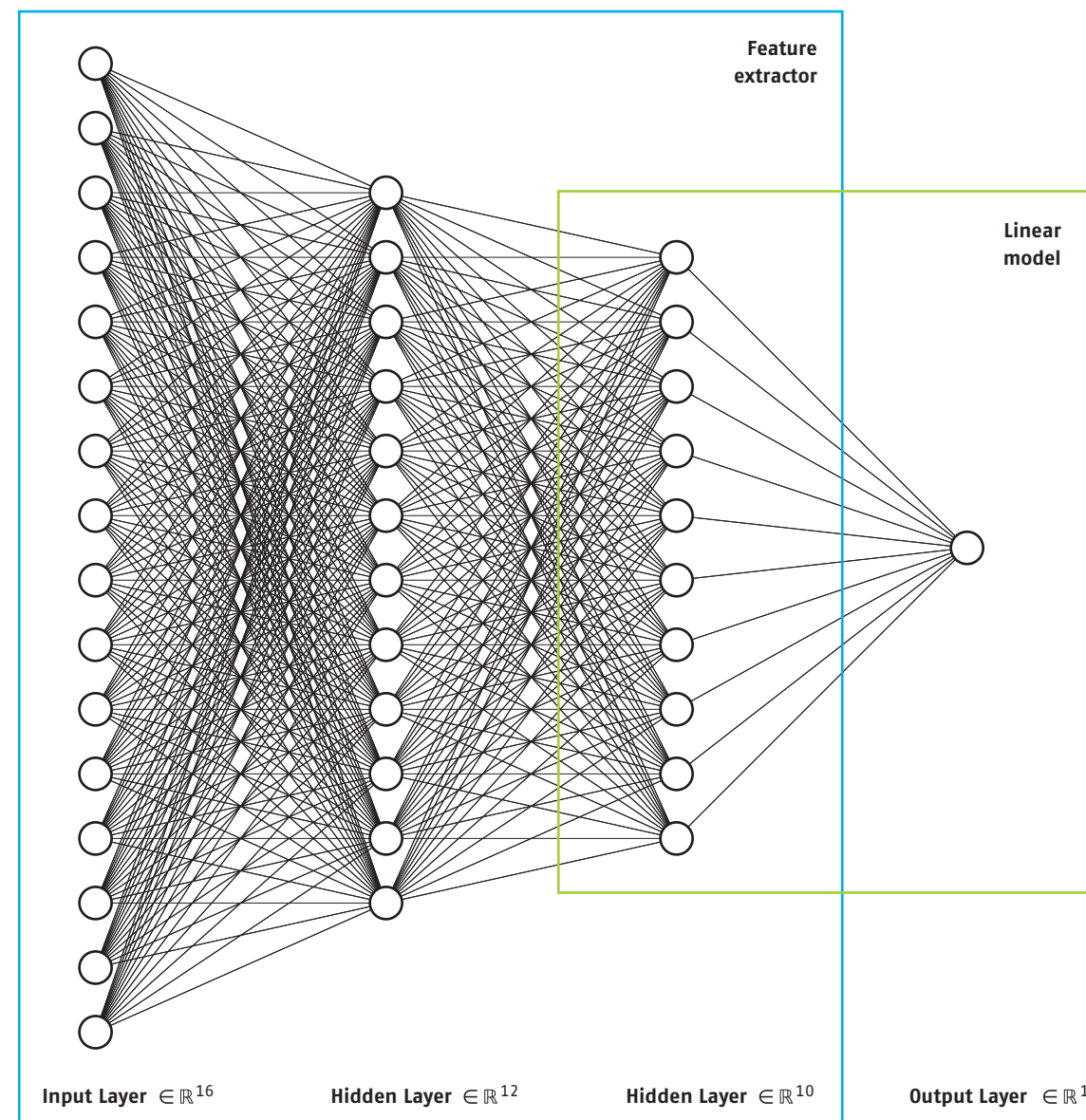


Figure 1: A diagram showing the two components of a deep neural network, within which variables are represented with circles and regression parameters with lines. The network functions in two components – a 'feature extractor' derives new variables from the data input into the network, and these feed into a linear model which makes predictions.

THE LEE-CARTER MODEL

Mortality rates and the rates at which mortality rates are expected to change over time are basic inputs into a variety of actuarial models. A starting point for setting mortality improvement assumptions is often population data, from which assumptions can be derived using mortality forecasting models. One of the most famous of these is the Lee-Carter (LC) model (Lee & Carter, 1992), which defines the force of mortality as

$$\log(\mu_{x,t}) = a_x + b_x \cdot k_t.$$

This equation states that the (log) force of mortality at age x in year t is the base mortality a_x at that age plus the rate of change of mortality b_x at that age, multiplied by a time-index k_t that applies to all ages under consideration. Like most mortality forecasting models, the LC model is fit in a two-stage process. The parameters of the model are calibrated, and then, for forecasting, the time index k_t is extrapolated.

The LC model is usually applied to forecast mortality of a single population, however, forecasts are often needed for multiple

populations simultaneously. While the LC model could be applied to each population separately, the period over which the model is fit needs to be chosen carefully so that the rates of change in mortality over time correctly reflect expectations about the future. Thus, a strong element of judgment is needed, which makes the LC model less suitable for multi-population forecasting.

MORTALITY FORECASTING USING DEEP LEARNING

Recently, several papers have applied deep neural networks to forecast mortality rates. This article focuses on the model in our recent paper (Perla et al., 2020) which applies specialized neural network architectures to model two mortality databases: the Human Mortality Database (HMD), containing mortality information for 41 countries, and the associated United States Mortality Database (USMD), providing lifetables for each state.

Our goal is to investigate whether, in common with the findings in the wider machine learning literature, neural networks specialized to process time series data can produce more accurate mortality forecasts than those produced by general neural network architectures. We also want to develop a model that is adaptable to changes in mortality rates by avoiding the need to follow a two-step calibration process. Thus, our model directly processes time series of mortality data with the goal of outputting new variables that can be used for forecasting. Finally, we wish to preserve the form of the LC model, due to the simplicity with which this model can be interpreted.

Originally published by The Actuary, August 2020. © The Institute and Faculty of Actuaries.

From left to right:
F. Perla is Professor for Financial Mathematics in the Department of Management Studies and Quantitative Methods at Parthenope University of Naples.
R. Richman is an Associate Director (R&D and Special Projects) at QED Actuaries and Consultants.
S. Scognamiglio is a Postdoctoral Research Fellow in Financial and Actuarial Mathematics in the Department of Management Studies and Quantitative Methods at Parthenope University of Naples
M.V. Wüthrich is Professor for Actuarial Science in the Department of Mathematics at ETH Zürich.



CONVOLUTIONAL NEURAL NETWORKS

Here, we focus on the convolutional neural network (CNN) presented in our paper. A CNN works by processing directly matrices of data that are input into the network, which could represent images or time series. We present a toy example of how this works in Figure 2. Data processing is accomplished by multiplying the data matrix with a ‘filter’, which is a smaller matrix comprised of parameters which are calibrated when fitting the model. Each filter is applied to the entire input data matrix, resulting in a processed matrix called a ‘feature map’. By calibrating the parameters of the filter in a suitable manner, CNNs can derive feature maps that represent important characteristics of the input data. See the caption of Figure 2 for more detail.

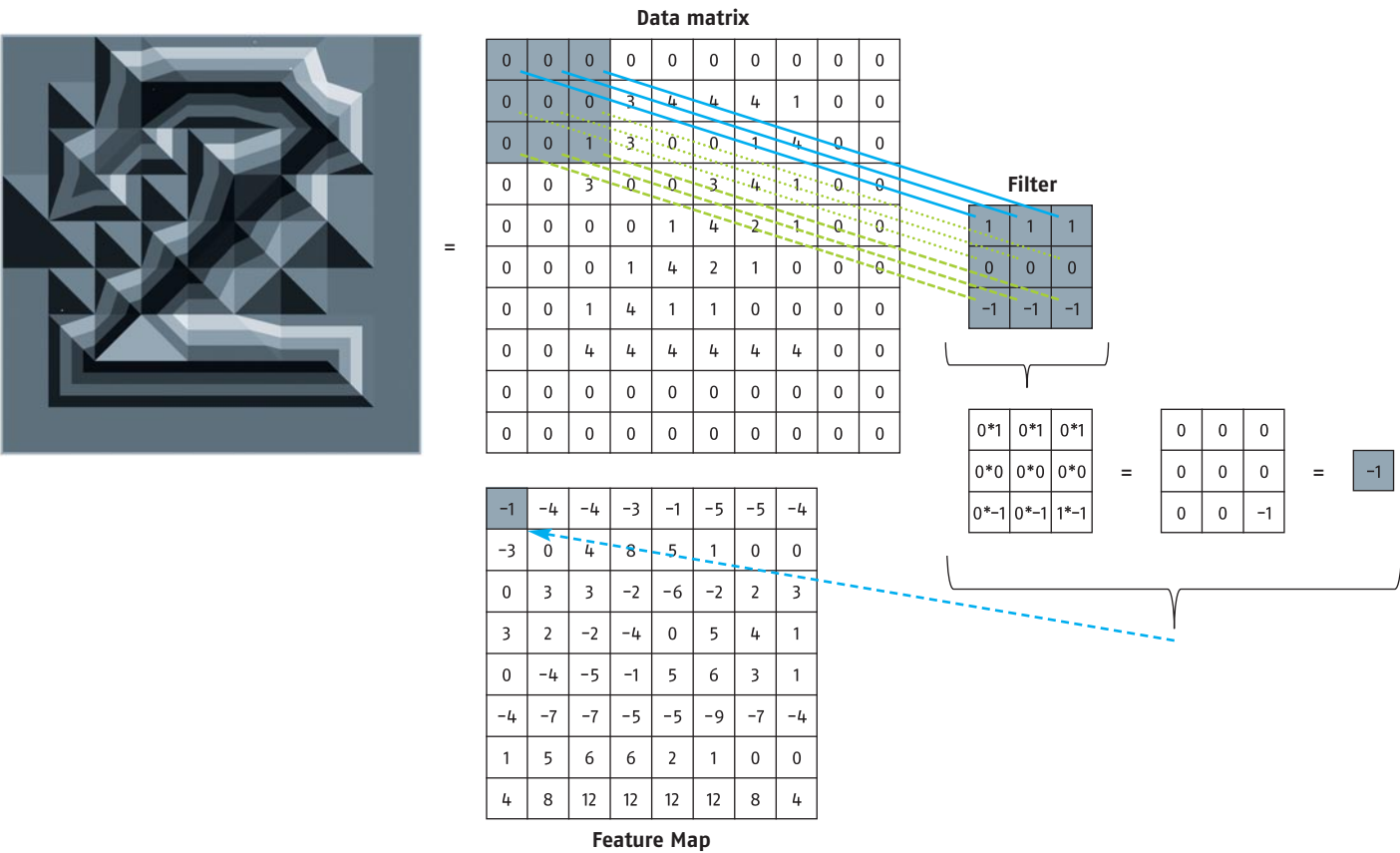


Figure 2: A toy example of a convolutional neural network. An image of a ‘2’ is represented by pixel intensities in the data matrix. The network consists of filters, which are multiplied with the data matrix to produce a new matrix called a ‘feature map’. Here, the filter is an edge-detector i.e. when the values of the feature map are large, a horizontal edge has been identified. This can be seen in the bottom row of the feature map, where large values correspond to the base of the ‘2’. Diagram excerpted from Richman (2018).

DEFINING THE MODEL

The CNN we apply for mortality forecasting works in a similar manner: we populate a matrix with mortality rates at ages 0–99 observed over ten years for each population and gender. This matrix is processed by multiplying the observed values of mortality rates with filters that span across the entire age range of the matrix and extend over three years, as shown in the top part of Figure 3. The filters derive a feature map that feeds into the rest of the model.

We also provide the model with variables representing the country being analysed and the gender of the population. To encode these variables, we applied a technique which maps categorical variables to low dimensional vectors called embeddings, in other words, each level of the category is mapped to a vector containing several new parameters, specifically, a 5-dimensional embedding layer, shown in the middle part of Figure 3.

Finally, we use the feature map and the embeddings directly in a GLM to forecast mortality rates in the next year, in other words, no other model components process the features before they enter the GLM. This is represented in the last part of Figure 3, which shows the direct connection of the output of the network to the feature layer.

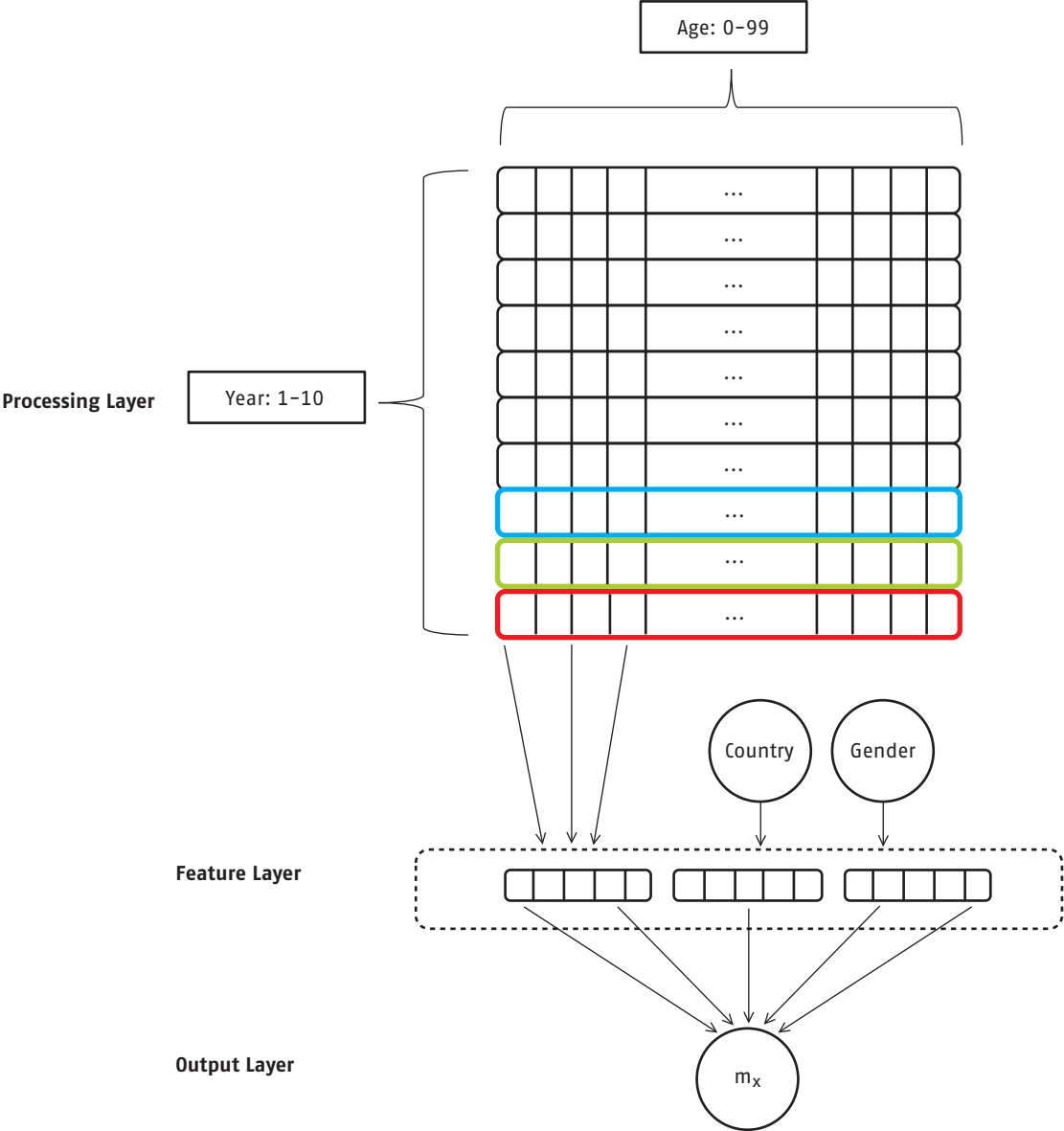


Figure 3: Diagrammatic representation of the convolutional mortality forecasting model, comprised of three parts. A processing layer derives new features from matrices of mortality data. These features are combined with a vector representation of the categorical data in the feature layer. Finally, mortality predictions are made in the output layer, which is a GLM.

RESULTS

We calibrated this model to the mortality experience in the HMD in the years 1950–1999 and tested the out of sample forecasting performance of the model on the experience in the years 2000–2016. Our benchmarks, against which the model was tested, were the original LC model, as well as the deep learning model from Richman & Wüthrich (2019), which is constructed without a processing layer geared towards time series data. We found that the out of sample forecasts were more accurate than the LC model 75 out of 76 times, and significantly outperformed the deep learning model. Residuals from the models are shown in Figure 4, indicating that while both deep learning models have better forecasting performance than the LC model, the CNN model fits the data for males significantly better than any other model. In the paper, we also show that the CNN model works well on the data in the USMD without any modifications.



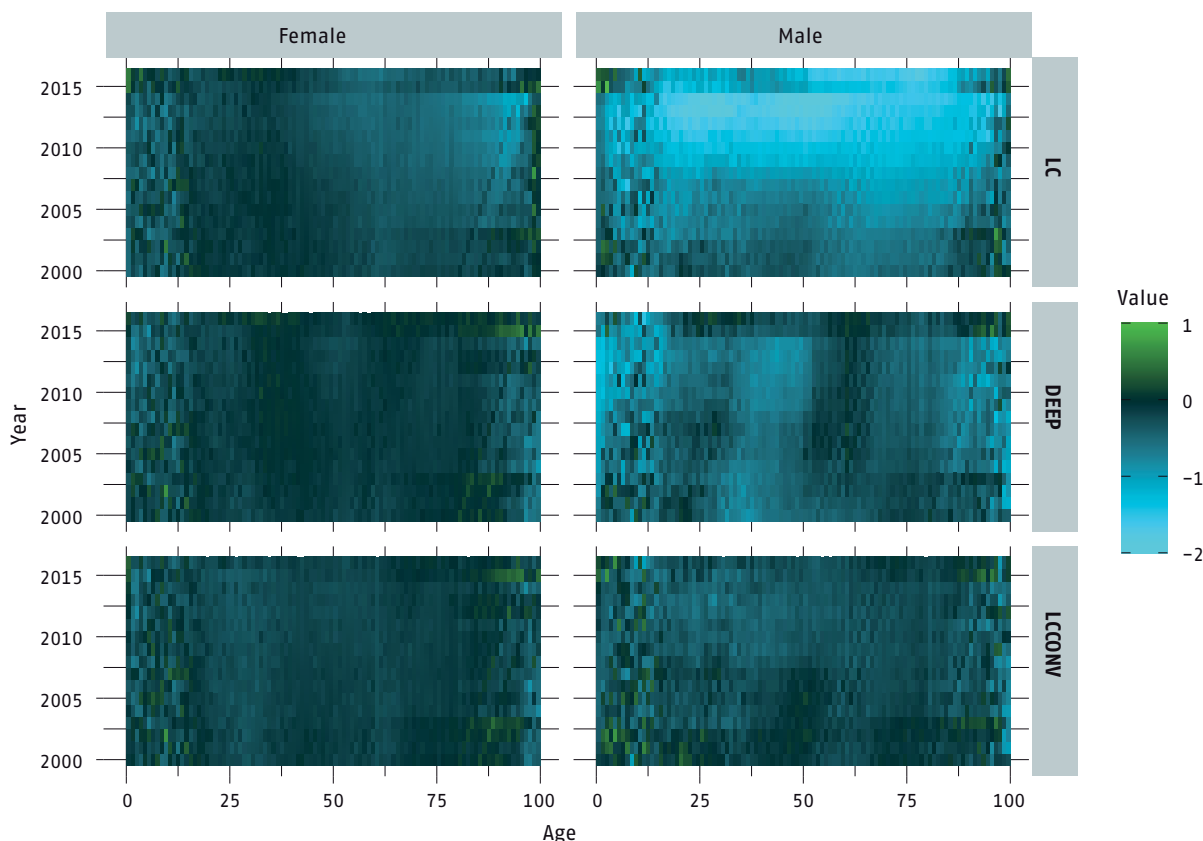


Figure 4: Out of sample residuals produced by using the LC and two deep neural network models to forecast mortality rates for the populations in the HMD.

INTERPRETATION WITHIN THE LEE-CARTER PARADIGM

Deep learning has been criticised as often being difficult to interpret. We can provide an intuitive explanation of how the convolutional model works in the framework of the LC paradigm for mortality forecasting. As mentioned above, the three sets of features derived with the neural network – which are features relating to population, gender and those derived using the convolutional network – are used directly in a GLM to forecast mortality. We show this mathematically using simplified notation in the following equation:

$$\log(\mu_{x,t}^{p,g}) = \widehat{a}_x^p + \widehat{a}_x^g + b_x \cdot \widehat{k}_t,$$

which states that the neural network predicts mortality based on new variables that have been estimated from the data, represented as variables with a 'hat'. The first two of these (\widehat{a}_x^p and \widehat{a}_x^g) play the role of estimating the average mortality for the population p and gender g under consideration, respectively, and in combination are equivalent to the a_x term in the Lee-Carter model. The third of these variables is a time index derived directly from the mortality data, which is equivalent to the k_t term in the LC model. This time index is calibrated each time new data is fed to the network, meaning to say, we have eliminated the two-stage procedure mentioned earlier, of fitting the model and then producing forecasts through extrapolation.

Thus, the seemingly complex model presented in this article can be interpreted in terms that are familiar to actuaries working in mortality forecasting.

CONCLUSION

We introduce a new mortality forecasting model in this article which uses a neural network to process mortality data directly to produce forecasts. The model has a simple interpretation within the framework of the original LC model and produces forecasts on the HMD with high out of sample accuracy. The model generalises well indicating that it is robust for the purpose of population mortality forecasting. Further investigation could focus on applying the model to insurance portfolio data, which often span shorter time periods than population data, and consideration of how uncertainty intervals might be generated. ■

References

- Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659–671. <https://doi.org/10.1080/01621459.1992.10475265>
- Papers with Code. (2020). ImageNet Leaderboard. Retrieved June 15, 2020, from Papers with Code website: <https://paperswithcode.com/sota/image-classification-on-imagenet>
- Perla, F., Richman, R., Scognamiglio, S., & Wüthrich, M. V. (2021). Time-Series Forecasting of Mortality Rates using Deep Learning. To appear *Scandinavian Actuarial Journal*. <https://doi.org/10.1080/03461238.2020.1867232>
- Richman, R. (2018). AI in Actuarial Science. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3218082>
- Richman, R., & Wüthrich, M. V. (2019). A neural network extension of the Lee-Carter model to multiple populations. *Annals of Actuarial Science*. <https://doi.org/10.1017/S1748499519000071>
- Wüthrich, M. V. (2019). From Generalized Linear Models to Neural Networks, and Back. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3491790>