A Neural Network Extension of the Lee-Carter Model to Multiple Populations

Ronald Richman and Mario Wüthrich

2 April 2019

Hosted by ACTUARIAL SOCIETY OF SOUTH AFRICA

SECTION COLLOQUIUM 2019

Agenda

- *Introduction*
- Deep Learning – Brief Overview
- Our Approach
- Discussion and Conclusion

Disclaimer

This presentation is given solely in my personal capacity and the views expressed in these slides do not necessarily represent those of the AIG Group and its subsidiaries, nor the professional organizations to which I belong (the IFoA, ASSA or IRMSA).

Any software or code referred to in the presentation is provided as is for demonstration purposes only, without any implied warranty, and is licensed under the MIT License which can be viewed on the associated GitHub repository.

Introduction

- Mortality rates and mortality improvement rates = key inputs into life insurance models
- Former usually based on experience of similar portfolios; latter often based on forecasting population mortality rates
- Foundational model for mortality forecasting is the Lee-Carter model (Lee and Carter 1992) (LC model)
- Many other approaches; within actuarial literature see Cairns, Blake and Dowd (2006) for an approach (CBD model) suited to old-age mortality (model coefficients of logistic model of q_x)

Lee-Carter Model (1)

- Mortality over time modeled using:

$$\log\left(u_{x,t}\right) = a_x + b_x k_t$$

- Mortality = *average rate + rate of change . time index*
- Latter terms = variables that must be estimated from data and then multiplied
- Could use interaction term between the variables Year and Age but this specification would require *t.x* effects to be fit compared to the *t+x* effects in the Lee-Carter model.
- => use non-linear/PCA regression (Brouhns, Denuit and Vermunt 2002; Currie 2016; Lee and Carter 1992)

Lee-Carter Model (2)

- Time index k_t estimated for years within sample…
- … so need to extrapolate k_t for out-of-sample forecasts
- Time series models of varying complexity used to forecast k_t
- Density forecasts generated using realizations of forecast time series k_t
- Some studies also consider uncertainty of:
  - Time series model parameters
  - LC Model parameters
- Two-step process – fit model and extrapolate - common to other mortality models, such as CBD model

Extending the LC Model

- Single population
  - Cohort effect (Renshaw and Haberman 2006)
  - Smoothing time series (Currie 2013)
- What about multiple populations?
- Intuition = multi-population mortality forecasting model should produce more robust forecasts
  - Common factors (similar socioeconomic circumstances, shared improvements in public health and medical technology)
  - Common trends likely captured with more statistical credibility
  - => Li and Lee (2005) recommend even if interest is in single series

Two basic models

- Augmented Common Factor (Li and Lee 2005)

$$\log(u_{x,t}) = a_x^i + b_x k_t + b_x^i k_t^i$$

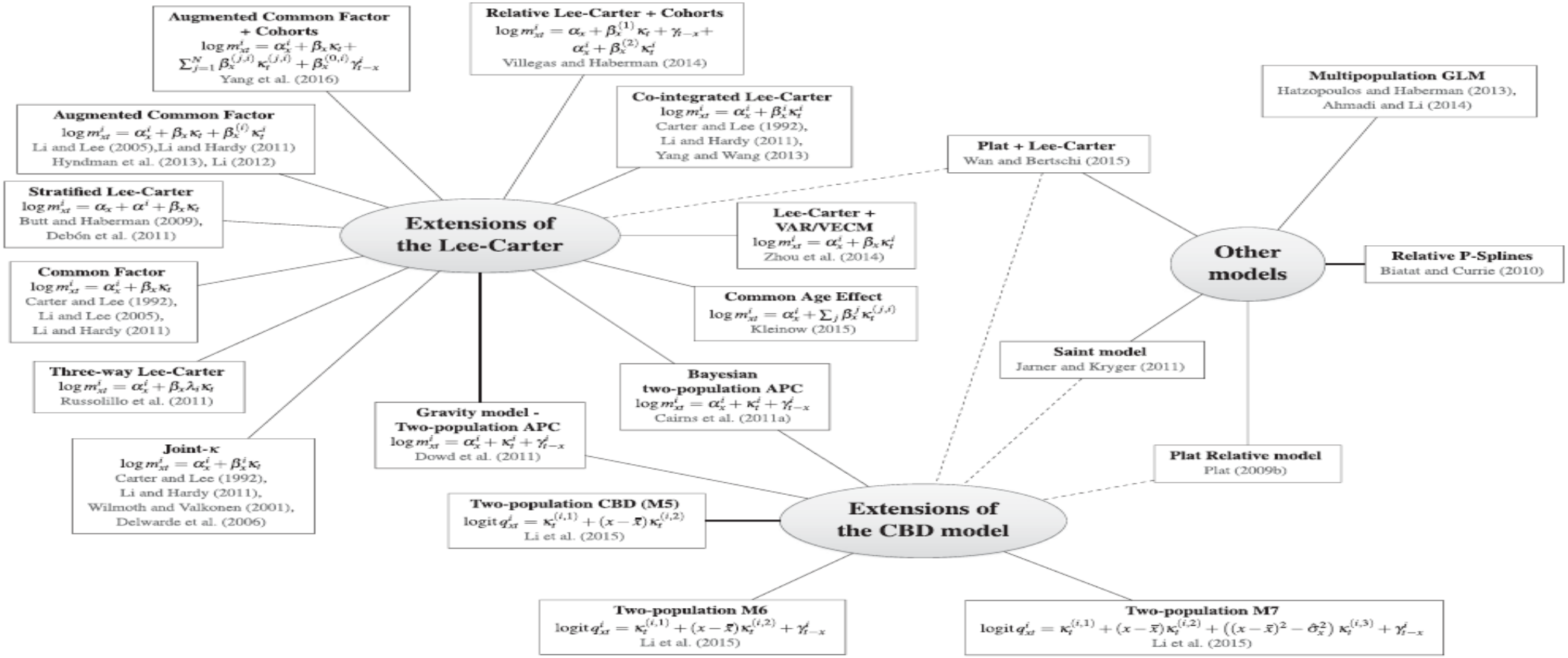
- Common Age Effect (Kleinow 2015)

$$\log(u_{x,t}) = a_x^i + b_x k_t^i$$

- Not intended for large scale mortality forecasting - generally applied on smaller sub-set of data => judgment of modeler needed
- Hard to fit (complex optimization schemes/less known statistical techniques)
- Which specification is better, when, and why?

Taxonomy of multi-population models

- Diagram excerpted from Villegas, Haberman, Kaishev *et al.* (2017)



Hosted by ACTUARIAL SOCIETY OF SOUTH AFRICA

SECTION COLLOQUIUM 2019

Another way?

- Explosion of interest in machine learning techniques
- For application within actuarial science (NL pricing), see Wüthrich and Buser (2018)
- Major success achieved on predictive modelling by Deep Learning in diverse fields, see LeCun, Bengio and Hinton (2015)
- Within actuarial science, review given by Richman (2018)
  - Talk tomorrow at 9:30 in ASTIN section
- Can we apply these techniques to the problem of large scale mortality forecasting?

Agenda

- Introduction
- *Deep Learning – Brief Overview*
- Our Approach
- Discussion and Conclusion

What is Machine Learning?

- To explain or predict? Shmueli (2010)
- Differences between statistical modelling (i.e. inference), and machine learning, due to distinction between tasks of predicting and explaining. Focus on predictive performance leads to:
  - Building algorithms to predict responses instead of specifying a stochastic data generating model (Breiman 2001)…
  - … favouring models with good predictive performance that are often more difficult to interpret than statistical models.
  - Accepting bias in models if this is expected to reduce the overall prediction error.
  - Quantifying predictive error (i.e. out-of-sample error) by splitting data into training, validation and testing sets, or using by cross-validation.

What is Deep Learning?

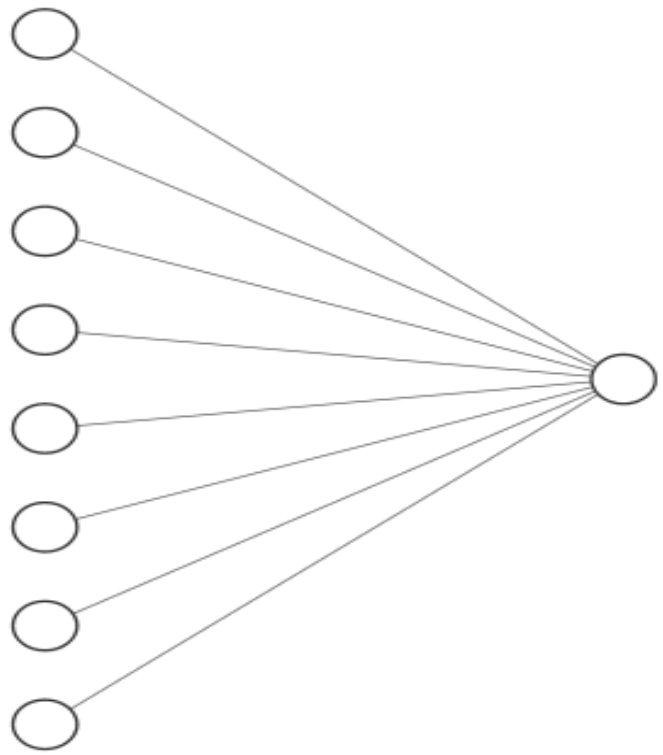
- Traditional approach to modelling relies on manual model specification
  - time consuming/tedious
  - relies on expert knowledge
  - becomes difficult with very high dimensional data
- Representation Learning = allows algorithms automatically to design the model by specifying new covariates (Bengio, Courville and Vincent 2013)
- Deep Learning = representation learning technique that constructs complex models using deep hierarchies of learned covariates
- Deep Learning relies on neural networks; see Goodfellow, Bengio and Courville (2016)

Practical Successes of Deep Learning

- Computer vision starting with AlexNet architecture of Krizhevsky, Sutskever and Hinton (2012)
- Speech recognition (Hannun, Case, Casper et al. 2014).
- Natural language processing, e.g. Google's neural translation machine (Wu, Schuster, Chen et al. 2016)
- Winning method in 2018 M4 time series forecasting competition (Makridakis, Spiliotis and Assimakopoulos 2018a).
- Analysis of GPS data (Brébisson, Simon, Auvolat et al. 2015)
- Analysis of tabular data (Guo and Berkhahn 2016) (plus other Kaggle competitions)

Single Layer NN = Linear Regression

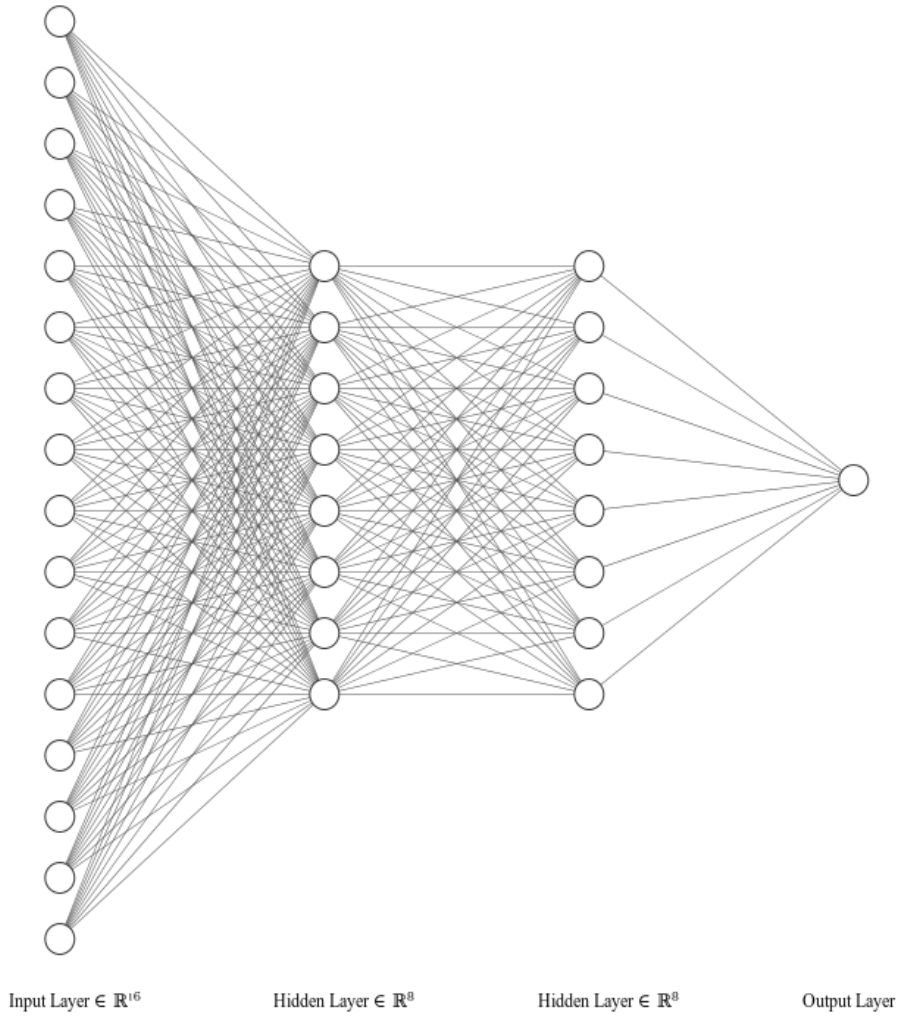
- Single layer neural network
- Circles = variables
- Lines = connections between inputs and outputs
- Input layer holds the variables that are input to the network…
- … multiplied by weights (coefficients) to get to result
- Single layer neural network is a linear regression!



Input Layer $\in \mathbb{R}^8$

Deep Feedforward Net

- Deep = multiple layers
- Feedforward = data travels from left to right
- Fully connected network = all neurons in layer connected to all neurons in previous layer
- More complicated representations of input data learned in hidden layers
- Subsequent layers represent regressions on the variables in hidden layers



Embedding layers

- Several specialized types of neural networks depending on purpose
- Key principle - Use architecture that expresses useful priors about the data => major performance gains
- Embedding layer learns dense vector transformation of sparse input vectors and clusters similar categories together; see Section 3.3 in Richman (2018)

	Actuary	Accountant	Quant	Statistician	Economist	Underwriter
Actuary	1	0	0	0	0	0
Accountant	0	1	0	0	0	0
Quant	0	0	1	0	0	0
Statistician	0	0	0	1	0	0
Economist	0	0	0	0	1	0
Underwriter	0	0	0	0	0	1

	Finance	Math	Stastistics	Liabilities
Actuary	0.5	0.25	0.5	0.5
Accountant	0.5	0	0	0
Quant	0.75	0.25	0.25	0
Statistician	0	0.5	0.85	0
Economist	0.5	0.25	0.5	0
Underwriter	0	0.1	0.05	0.75

Agenda

- Introduction
- Deep Learning – Brief Overview
- *Our Approach*
- Discussion and Conclusion

Extending LC – two perspectives

- Lee Carter model = regression model using features derived from data using PCA
  - CAE + ACF = regression models with features derived at a regional level
- Perspective 1: Use a neural network to model the regression problem and let it decide on the feature set
- Lee Carter model has a neural network formulation; see Richman and Wüthrich (2018)

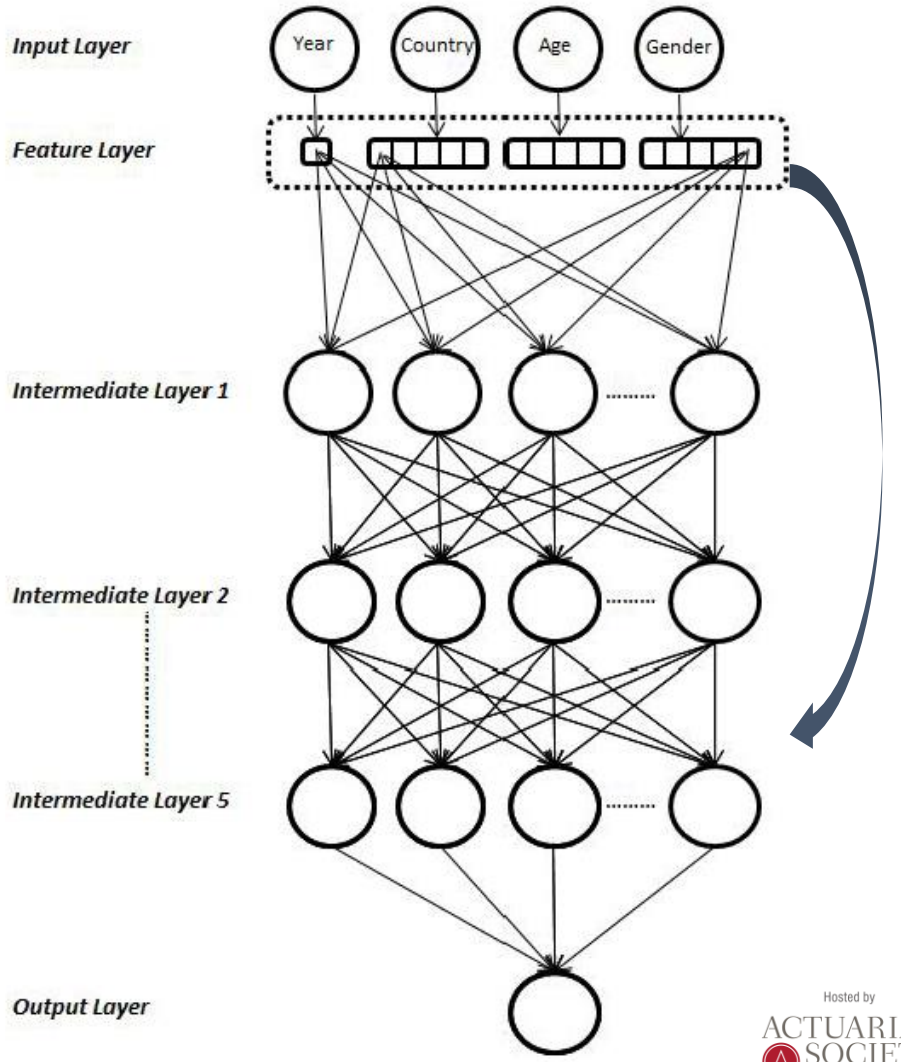
$$\log\left(u_{x,t}\right) = g(x) + h(x)i(t),$$

- Perspective 2: use a more general step function formulation to specify the multi-population model

$$g(x) = \begin{cases} a_1 & \text{for } x = 1, \\ a_2 & \text{for } x = 2, \\ \vdots & \\ a_\omega & \text{for } x = \omega, \end{cases}$$

Deep neural network

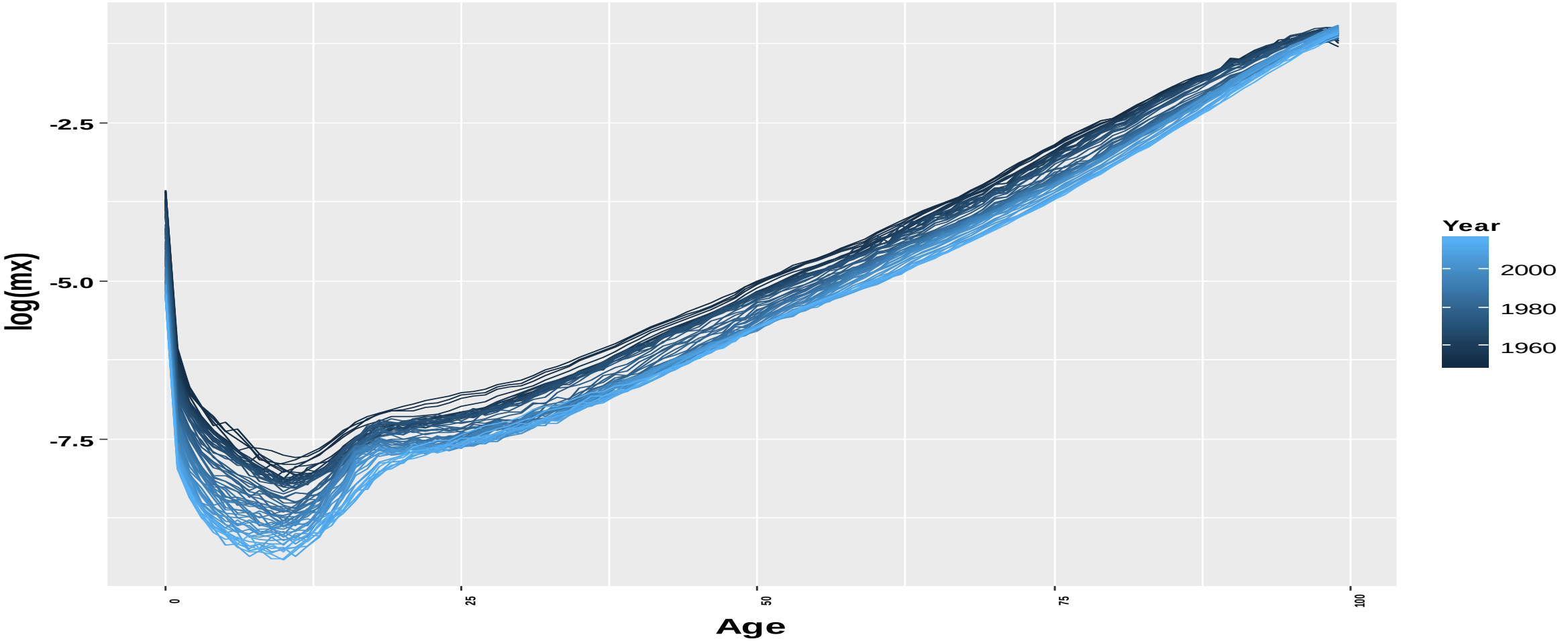
- Categorical inputs to network defined using embedding layers = vector valued step functions of parameters calibrated from input data
- Year input is numerical
- Intermediate layers combine the inputs into new features (128 nodes per layer) using non-linear transformations
- Deep networks hard to optimize => add a skip connection (He, Zhang, Ren *et al.* 2016)



Data from HMD

- Mortality data sourced from Human Mortality Database (HMD)
- Covers mortality rates for ~41 countries, for both genders, from 1950-2016
- Divide data into training and test sets:
  - Training set = observations at ages 0-99 occurring in the years before 2000
  - Test set = observations in the years 2000-2016
- Countries in the HMD that have at least ten years of data before year 2000 (excludes Germany, Croatia and Chile)
- 38 of the 41 countries used = aim to forecast 76 distinct sets of mortality rates

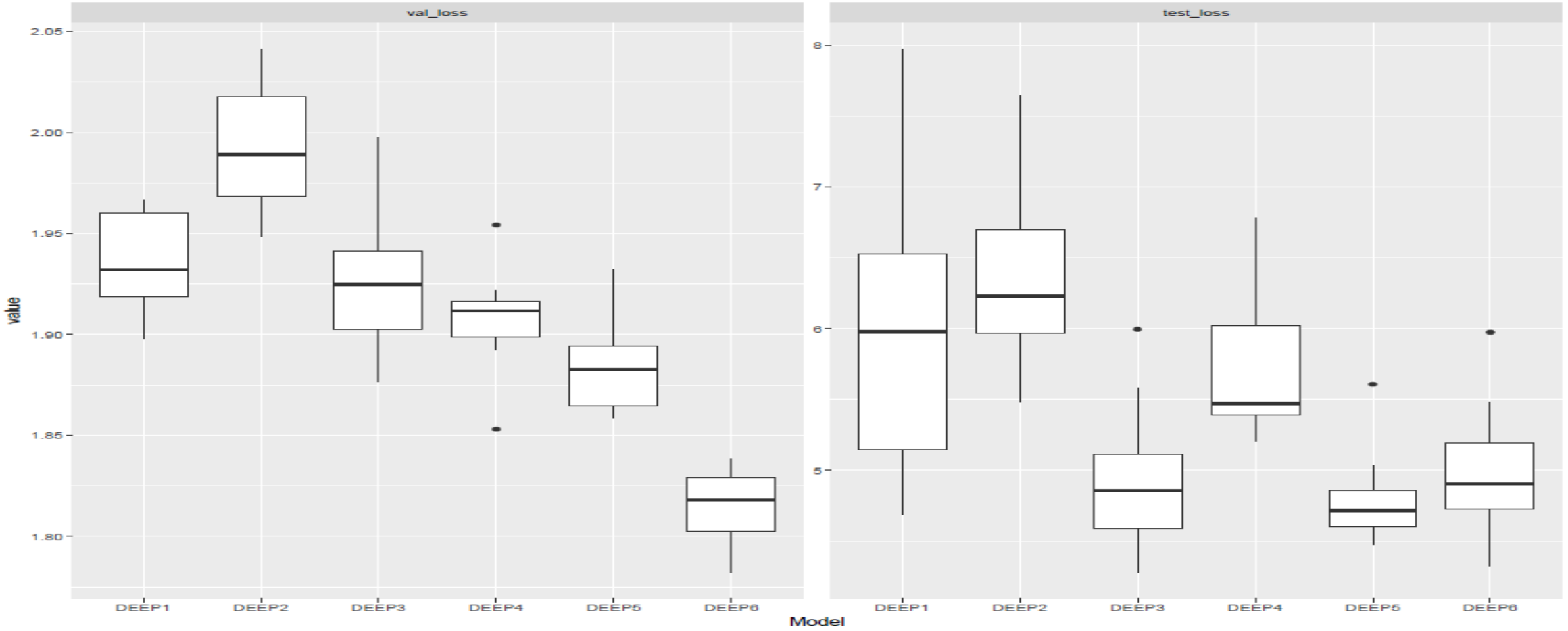
Female Mortality, USA, 1950-2016



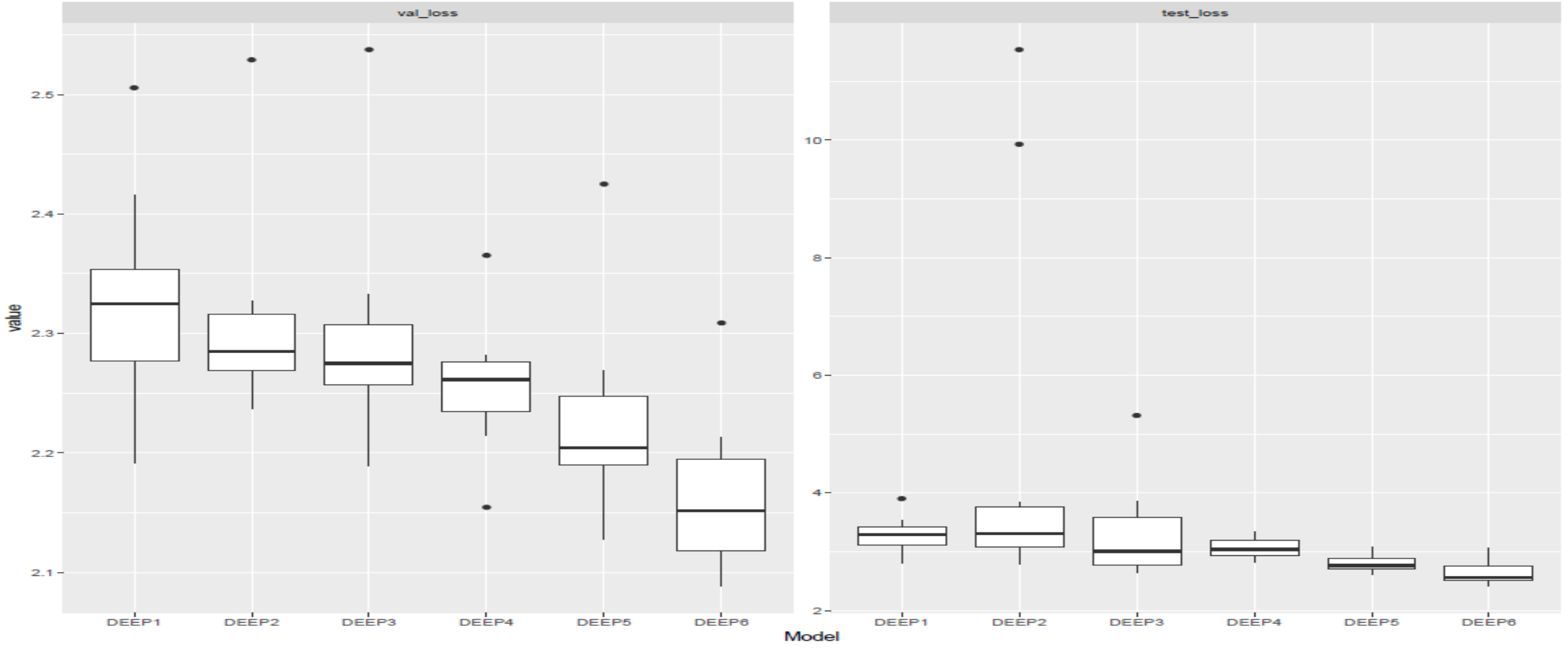
Testing the models

- Test criterion = smallest MSE on out-of-sample mortality forecasts
  - MSE is natural choice
  - Optimization form of PCA/SVD uses MSE (Efron and Hastie 2016)
  - Maximises likelihood of Gaussian model
- Chose best model of each class for the tests:
  - Lee Carter fit with SVD
  - Lee Carter calibrated to regional mortality
  - Augmented Common Factor model
  - Common Age Effects model
  - Best of Deep Neural Networks
- Best deep network determined on forecasts in years 1990-1999

Choosing the best NN – 1990-1999



Performance – 2000-2016

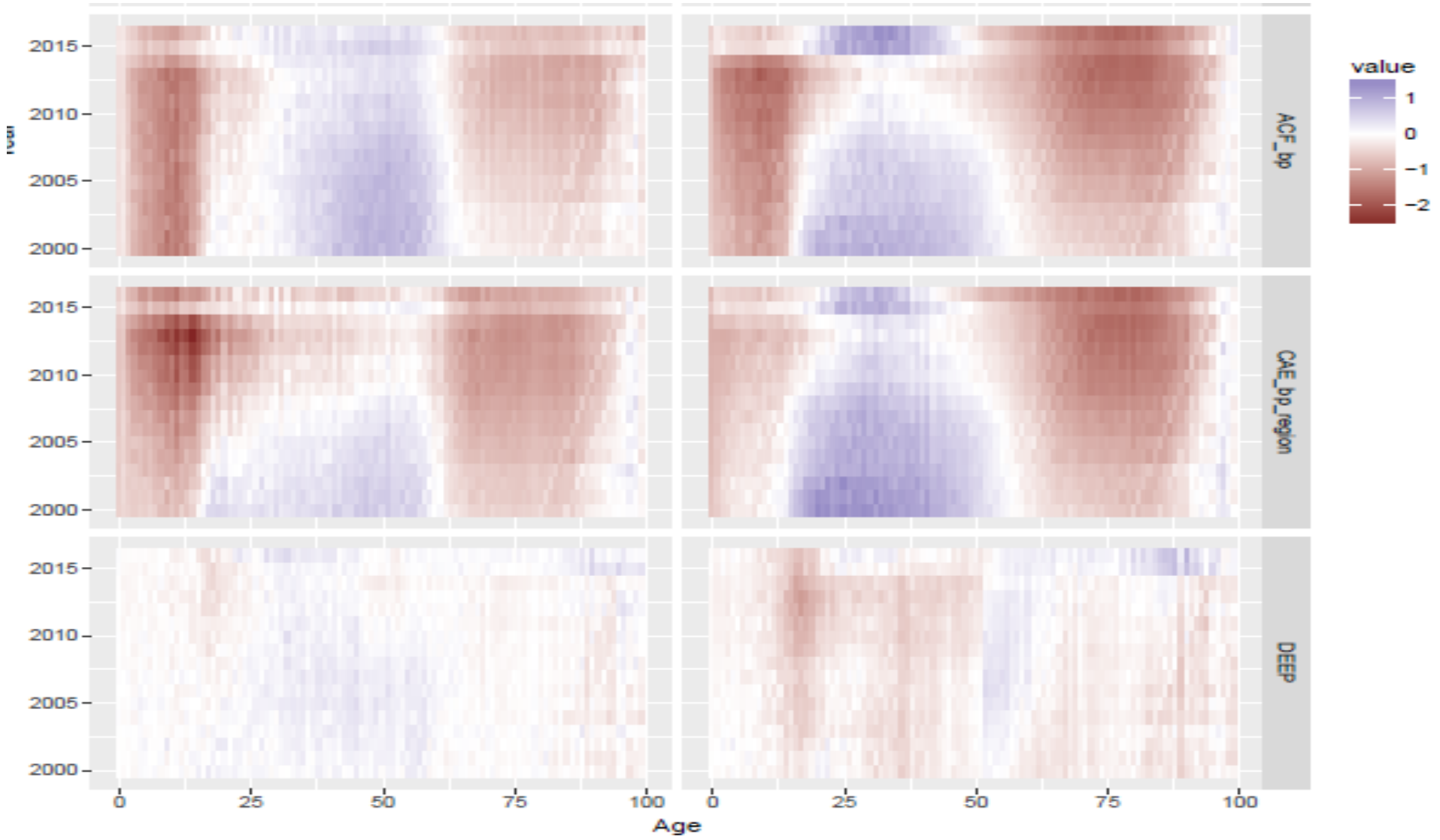


Results

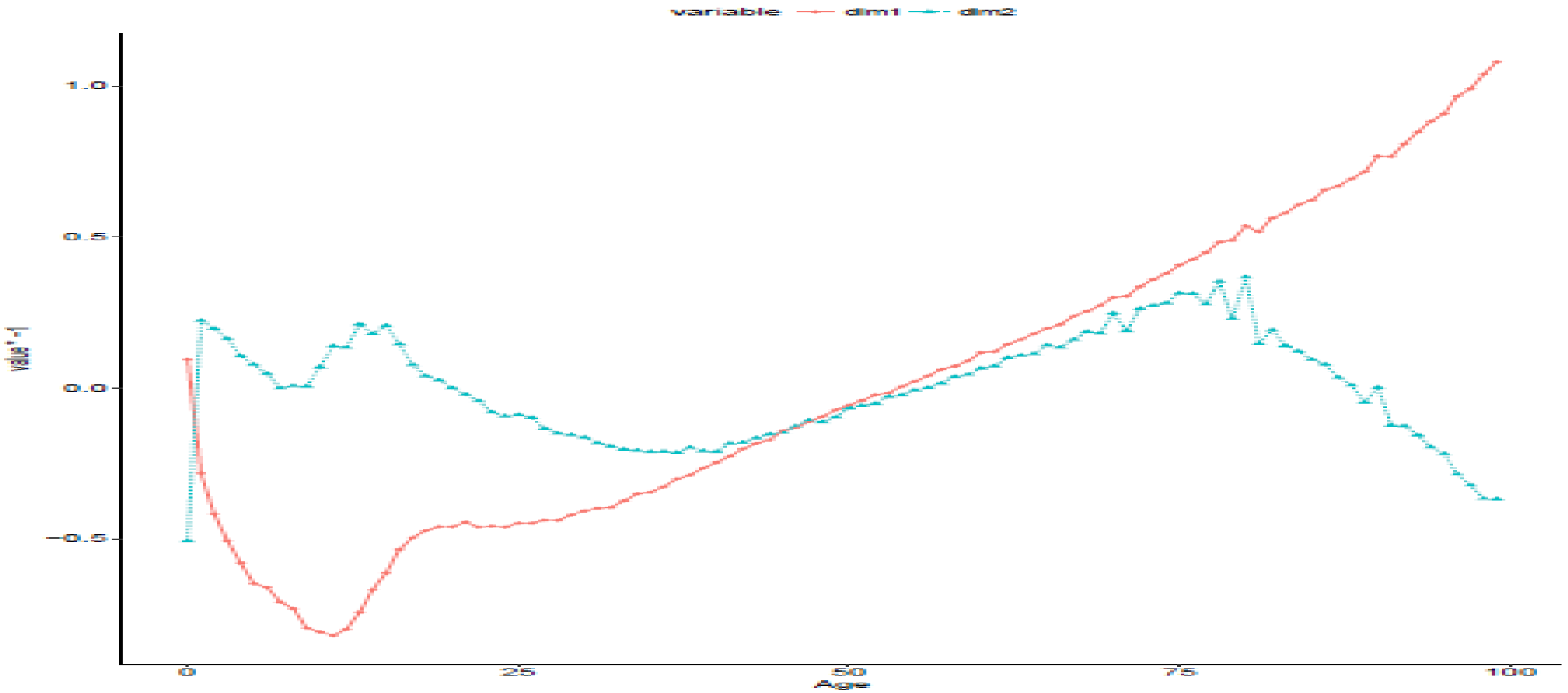
- Results of comparing the models
- Best performing model is deep neural network…
- …produces the best out-of-time forecasts 51 out of 76 times
- for purposes of large scale mortality forecasting, deep neural networks dramatically outperform traditional single and multi-population forecasting models

	Model	Average MSE	Median MSE	Best Performance
1	LC_SVD	5.50	2.48	7
2	LC_ACF_region	3.46	2.50	10
3	ACF_BP	6.12	3.00	4
4	CAE_BP	5.59	3.46	4
5	DEEP	2.68	1.38	51

Implicit Cohort Effects



Age Embedding



Agenda

- Introduction
- Deep Learning – Brief Overview
- Our Approach
- *Discussion and Conclusion*

Discussion

- Deep neural nets have enormous potential to solve model specification problems … once a suitable deep architecture has been found
- Skip connections make a big difference since model only needs to learn residuals (He, Zhang, Ren et al. 2016)
  - See recent work by Gabrielli, Richman and Wüthrich (2018) and Schelldorfer and Wüthrich (2019)
- Against conventional wisdom: tanh was better than ReLU on this tabular data set
- Embeddings are a powerful way to understand and extend traditional statistical models

Conclusion

- One important comment we received stated that although neural network methods are a black box, their superiority in out-of-sample forecasting is clearly demonstrated.
- How can we give key stakeholders (including regulators) comfort around deep neural networks?
  - Interpretability - LIME (Ribeiro, Singh and Guestrin 2016)
  - Design the model for interpretability - Combined Actuarial Neural Net (CANN)– (Wüthrich and Merz 2018)
- Future research to consider:
  - Ensembling of models (56/76 by ensembling ReLU+ tanh)
  - Uncertainty bounds

References

Bengio, Y., A. Courville and P. Vincent. 2013. "Representation learning: A review and new perspectives", IEEE transactions on pattern analysis and machine intelligence 35(8):1798-1828.

Breiman, L. 2001. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)", Statistical Science 16(3):199-231.

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

Cairns, A.J.G., D. Blake and K. Dowd. 2006. "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration", Journal of Risk & Insurance 73(4):687-718.

Currie, I.D. 2013. "Smoothing constrained generalized linear models with an application to the Lee-Carter model", Statistical Modelling 13(1):69-93.

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

De Brébisson, A., É. Simon, A. Auvolat, P. Vincent et al. 2015. "Artificial neural networks applied to taxi destination prediction", arXiv arXiv:1508.00021

Efron, B. and T. Hastie. 2016. Computer Age Statistical Inference. Cambridge University Press.

Goodfellow, I., Y. Bengio and A. Courville. 2016. Deep Learning. MIT Press.

Guo, C. and F. Berkhahn. 2016. "Entity embeddings of categorical variables", arXiv arXiv:1604.06737

Hannun, A., C. Case, J. Casper, B. Catanzaro et al. 2014. "Deep speech: Scaling up end-to-end speech recognition", arXiv arXiv:1412.5567

He, K., X. Zhang, S. Ren and J. Sun. 2016. "Deep residual learning for image recognition," Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition. 770-778.

Ioffe, S. and C. Szegedy. 2015. "Batch normalization: Accelerating deep network training by reducing internal covariate shift", arXiv preprint arXiv:1502.03167

Kleinow, T. 2015. "A common age effect model for the mortality of multiple populations", Insurance: Mathematics and Economics 63:147-152.

Krizhevsky, A., I. Sutskever and G. Hinton. 2012. "Imagenet classification with deep convolutional neural networks," Paper presented at Advances in Neural Information Processing Systems. 1097-1105.

LeCun, Y., Y. Bengio and G. Hinton. 2015. "Deep Learning", Nature 521(7553):436.

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

Li, N. and R. Lee. 2005. "Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method", Demography 42(3):575-594.

Makridakis, S., E. Spiliotis and V. Assimakopoulos. 2018. "The M4 Competition: Results, findings, conclusion and way forward", International Journal of Forecasting

Renshaw, A.E. and S. Haberman. 2006. "A cohort-based extension to the Lee–Carter model for mortality reduction factors", Insurance: Mathematics and Economics 38(3):556-570.

Ribeiro, M.T., S. Singh and C. Guestrin. 2016. "Why should I trust you?: Explaining the predictions of any classifier," Paper presented at Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM. 1135-1144.

Richman, R. 2018. "AI in actuarial science", Available at SSRN: https://ssrn.com/abstract=3218082

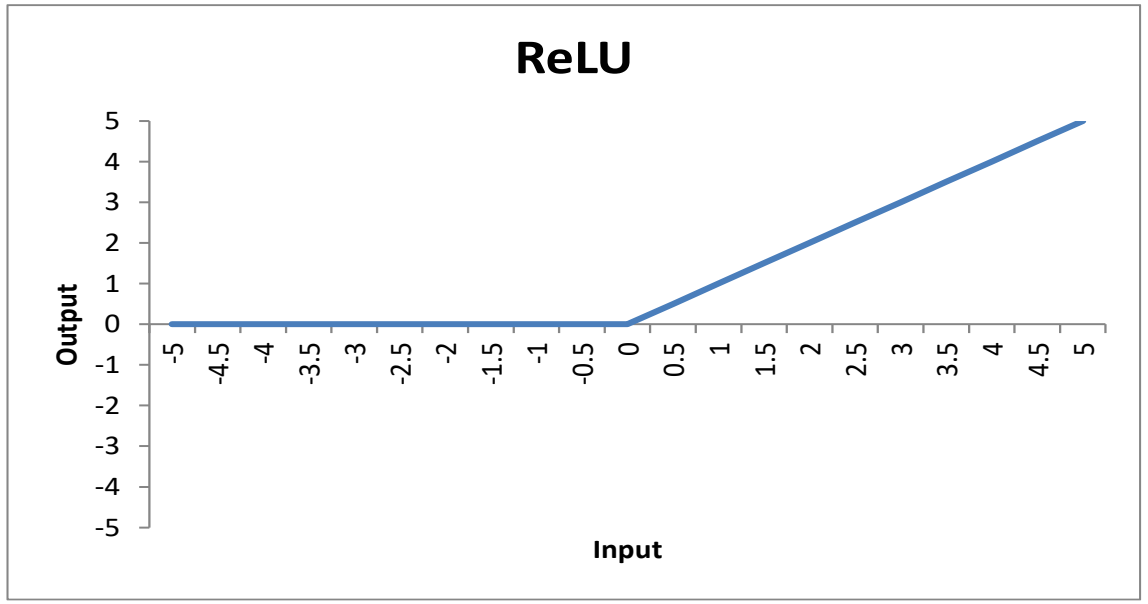
Richman, R. and M. Wüthrich. 2018. "A Neural Network Extension of the Lee-Carter Model to Multiple Populations", Available at SSRN: https://ssrn.com/abstract=3270877

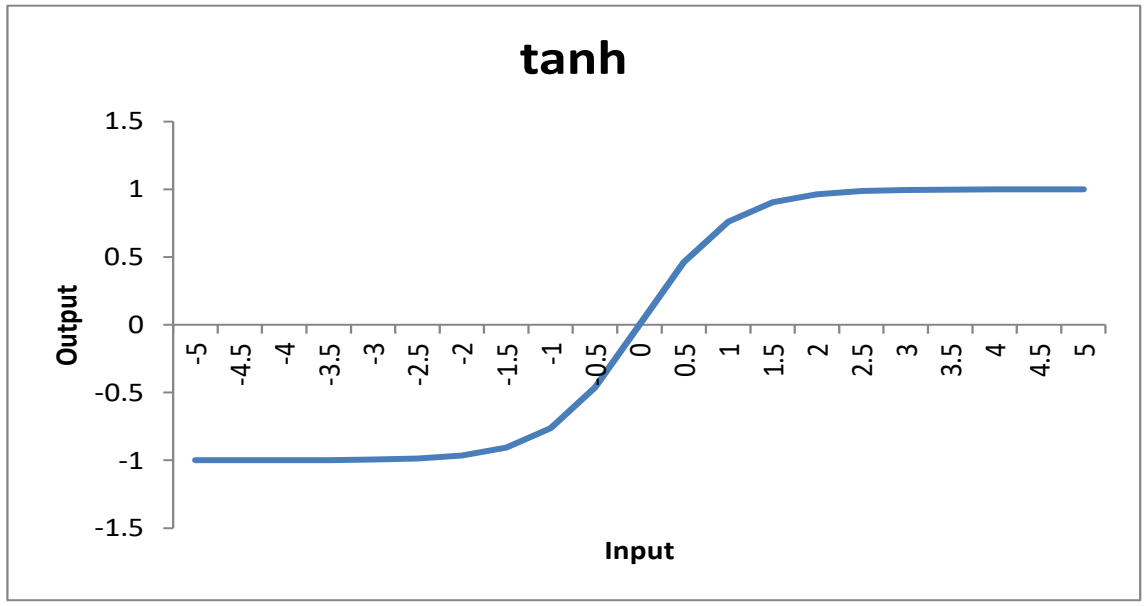
Shmueli, G. 2010. "To explain or to predict?", Statistical Science:289-310.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever et al. 2014. "Dropout: a simple way to prevent neural networks from overfitting", The Journal of Machine Learning Research 15(1):1929-1958.

Technical Details

- 5 dimensional embeddings
- Regularization - Dropout = 5%; see Srivastava, Hinton, Krizhevsky *et al.* (2014)
- Design - Batchnorm; see Ioffe and Szegedy (2015)
- Tried several combinations:
  - *Non-linear function* - ReLu vs tanh
  - *Depth* - 2 layers vs 5 layers
  - *Design* - no skip connection vs skip connection





SECTION COLLOQUIUM 2019

Get involved

- Insurance Data Science conference - 14 June 2019
- ETH Zurich
- https://insurancedatascience.org/
- Amazing line-up of papers, presentations and speakers!

- Kasa.ai – launching soon, led by Kevin Kuo of Rstudio
- An open research group encouraging innovation in insurance analytics
- Some interesting projects planned

Thanks for listening - Any questions?

Paper: www.ssrn.com/abstract=3270877
Contact: ron@ronaldrichman.co.za

Hosted by ACTUARIAL SOCIETY OF SOUTH AFRICA

SECTION COLLOQUIUM 2019

IAA AAI SECTION COLLOQUIUM 2019

THE MODERN ACTUARY - CHALLENGE • INFLUENCE • LEAD

ASTIN • IAAHS • IAALS • IACA • PBSS

2 - 5 April
2019
Cape Town
South Africa
CTICC

www.colloquium2019.org.za

Hosted by
ACTUARIAL SOCIETY OF SOUTH AFRICA