AI in Actuarial Science - Two Years On

Ron Richman

SA Taxi

EAA e-Conference on
Data Science & Data Ethics

29 June 2021

Annals of Actuarial Science (2020), 1–23
doi:10.1017/S1748499520000238

REVIEW

Institute and Faculty of Actuaries

AI in actuarial science – a review of recent advances – part 1

Ronald Richman

AGENDA

- **Deep Learning in 6 slides**
- Actuarial Examples of Representation Learning
- Advances in Deep Learning
- Applying Deep Learning in Actuarial Science
- Explaining Deep Learning models
- Uncertainty estimation
- Conclusions

PRICING AS MACHINE LEARNING

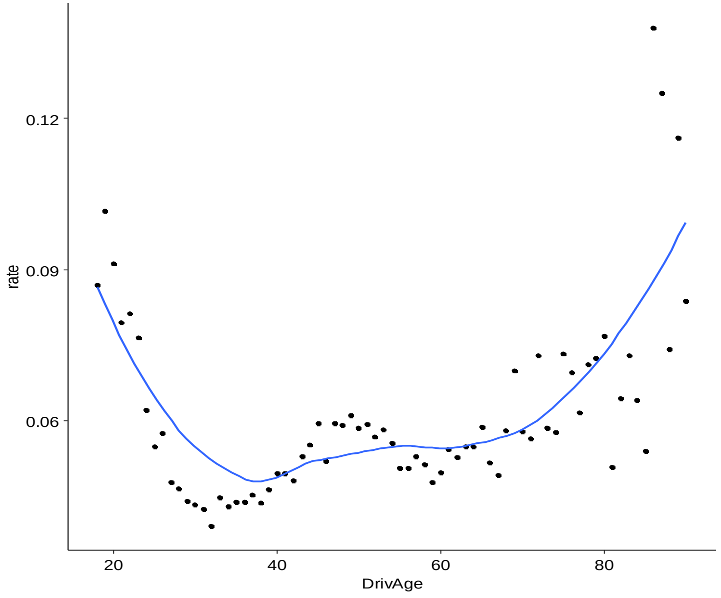
- **Supervised learning = application of machine learning to datasets that contain features and outputs with the goal of predicting the outputs from the features (Friedman, Hastie and Tibshirani 2009).**

- **Feature engineering - Suppose we realize that Claims depends on Age^2 => enlarge feature space by adding Age^2 to data. Other options – add interactions/basis functions e.g. splines**

y (*outputs*)

X (*features*)

```
> freMTPL2freq
          IDpol ClaimNb    Exposure Area VehPower VehAge DrivAge BonusMalus VehBrand  VehGas Density Region DrivAge_2
     1:       1       1 0.100000000    D        5      0      55         50      B12 Regular    1217    R82      3025
     2:       3       1 0.770000000    D        5      0      55         50      B12 Regular    1217    R82      3025
     3:       5       1 0.750000000    B        6      2      52         50      B12  Diesel      54    R22      2704
     4:      10       1 0.090000000    B        7      0      46         50      B12  Diesel      76    R72      2116
     5:      11       1 0.840000000    B        7      0      46         50      B12  Diesel      76    R72      2116
    ---
678009: 6114326       0 0.002739726    E        4      0      54         50      B12 Regular    3317    R93      2916
678010: 6114327       0 0.002739726    E        4      0      41         95      B12 Regular    9850    R11      1681
678011: 6114328       0 0.002739726    D        6      2      45         50      B12  Diesel    1323    R82      2025
678012: 6114329       0 0.002739726    B        4      0      60         50      B12 Regular      95    R26      3600
678013: 6114330       0 0.002739726    B        7      6      29         54      B12  Diesel      65    R72       841
> |
```



THE NEED FOR HUMAN INTERVENTION

- **In many domains, traditional approach to designing actuarial/machine learning systems relies on human input for model specification/ feature engineering.**

- **Three arguments against traditional approach:**

<u>Complexity</u> – which are the relevant features to extract/what is the correct model specification? Difficult with very high dimensional, unstructured data such as images or text. (Bengio 2009; Goodfellow, Bengio and Courville 2016)

<u>Expert knowledge</u> – requires suitable prior knowledge, which can take decades to build (and might not be transferable to a new domain) (LeCun, Bengio and Hinton 2015)

<u>Effort</u> – designing features is time consuming/tedious => limits scope and applicability (Bengio, Courville and Vincent 2013; Goodfellow, Bengio and Courville 2016)

- **Complexity is not only due to unstructured data. Many difficult problems of model specification arise when performing actuarial/demographic tasks at a large scale**

REPRESENTATION LEARNING

- Representation Learning = ML techniques where algorithms automatically design features that are optimal (in some sense) for a particular task

- Traditional examples are PCA (unsupervised) and PLS (supervised):

  PCA produces features that summarize directions of greatest variance in feature matrix

  PLS produces features that maximize covariance with response variable (Stone and Brooks 1990)

- Feature space then comprised of learned features which can be fed into ML/DL model

- BUT: Simple/naive RL approaches often fail when applied to high dimensional/very complex data

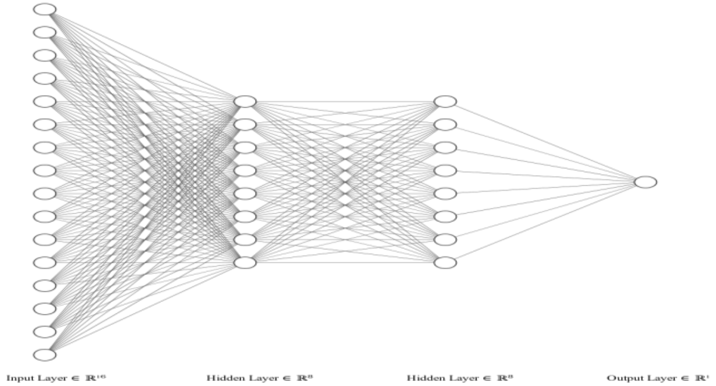
DEEP LEARNING

- Deep Learning = representation learning technique that automatically constructs hierarchies of complex features to represent abstract concepts

  Features in lower layers composed of simpler features constructed at higher layers => complex concepts can be represented automatically

- Typical example of deep learning is feed-forward neural networks, which are multi-layered machine learning models, where each layer learns a new representation of the features.

- The principle: Provide raw data to the network and let it figure out what and how to learn.



EXAMPLE: FASHION-MNIST

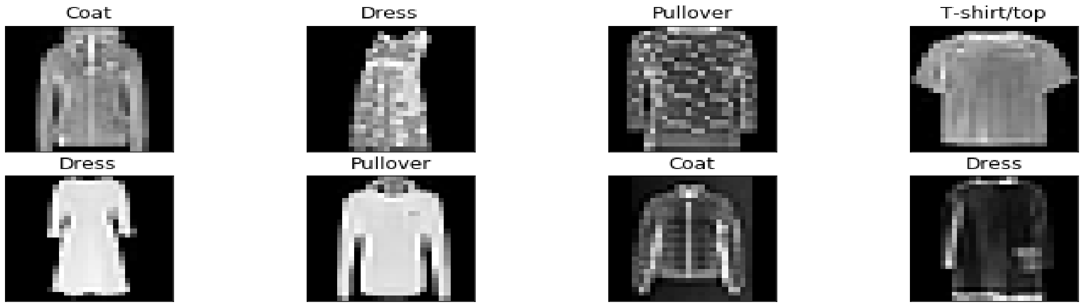
- Applied a deep autoencoder to the same data (trained in unsupervised manner)

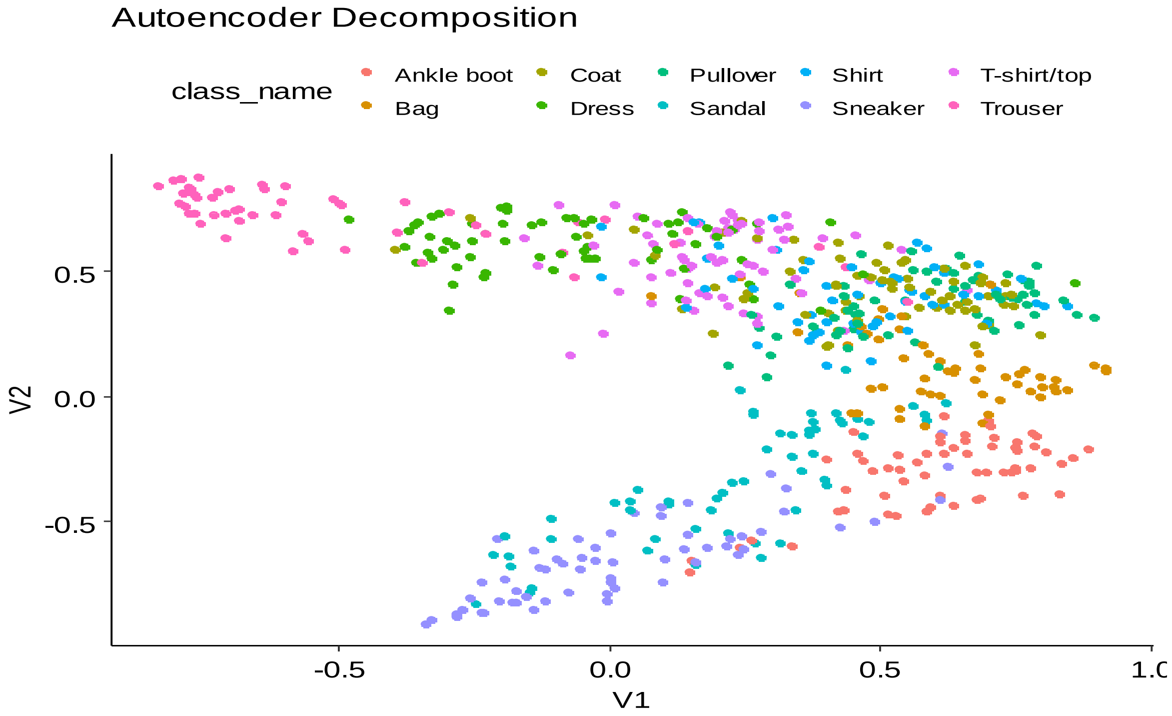
  Type of non-linear PCA

- Differences between classes shown

- Deep representation of data automatically captures meaningful differences between the images without (much) human input

- Automated feature/model specification





SPECIFYING MODELS – A NOTATION

Traditional Actuarial

$$M\left(X;T;\sum\beta_i f_i(xi);\Theta\right)=\hat{y}$$

- Linear model specification, for f_i identity (GLM), f_i spline function (GAM)
- β_i regression parameters

Machine Learning

$$M(X;T;S(A,\tilde{E});\Theta)=\hat{y}$$

- Implicit Specification of the model $\tilde{E}$ by a class of algorithms A

Deep Learning

$$M(X;\tilde{T};S(A,\tilde{E});\Theta)=\hat{y}$$

- Representation Learning: Implicit Specification of functions $\tilde{T}$ to derive features X'
- Explicit use of loss function $L(y,\hat{y})$ to measure predictive accuracy

from Richman, von Rummell, & Wüthrich (2019)

AGENDA

- Deep Learning in 6 slides
- **Actuarial Examples of Representation Learning**
- Advances in Deep Learning
- Applying Deep Learning in Actuarial Science
- Explaining Deep Learning models
- Uncertainty estimation
- Conclusions

REPRESENTATION LEARNING

- When applied to *tabular data,* DL models perform representation learning on inputs:

  - Interaction terms
  - Non-linearities

- Paradigm of representation learning extends also to new types of data:

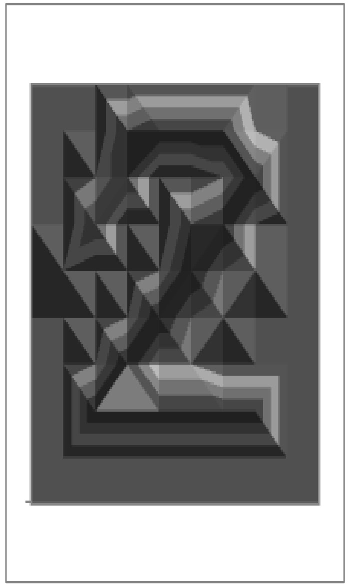
  - High dimensional
  - Unstructured

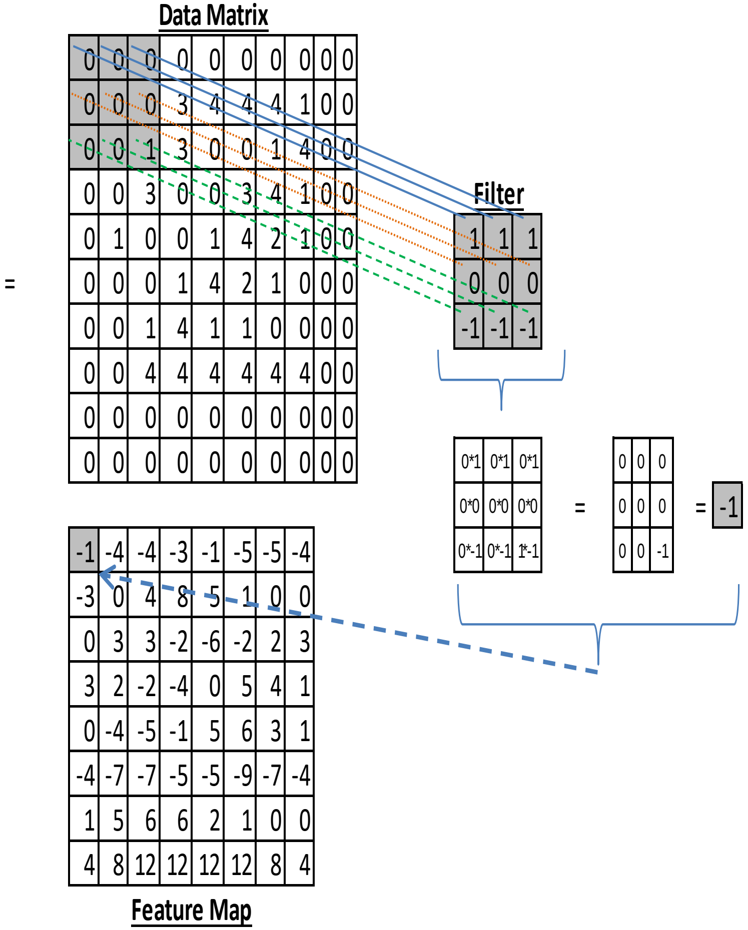
- Two recent examples – mortality forecasting + telematics

- But first, a detour: Convolutional Neural Networks

CONVOLUTIONAL NN - IMAGES

- **Prior - features in images are position invariant i.e. can recognize at any position within an image**
- **Also applies to audio/speech and text/time series data**
- **Convolutional network is locally connected and shares weights => expresses prior of position invariance**
- **Far fewer parameters than FCN**
- **Each neuron (i.e. feature map) in network derived by applying filter to input data**
- **Weights of filter learned when fitting network**





LEE-CARTER MODEL

- **Forecast mortality rates = key inputs into demographic forecasting, life insurance and pensions models**
- **Foundational model for mortality forecasting is the Lee-Carter model (Lee and Carter 1992) (LC model)**
- **Mortality over time modeled using:**

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

- **i.e. (log) mortality = average rate + rate of change . time index**
- **Relies on latent variables that must be estimated from data and then multiplied => use PCA to estimate the latent terms**

PROCESSING TIME SERIES WITH DL

- **Can we derive features for mortality forecasting directly from past mortality rates using DL?**

- **In the LC model, we have the following regression function:**

$$(t, x, i) \mapsto \log\left(u_{x,t}^{(i)}\right)$$

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

- **Rather, can we map directly from observed mortality rates of many populations to a time feature:**

$$U^{(i)} \mapsto \boldsymbol{k}_t^{(i)} \in \mathbb{R}^q$$

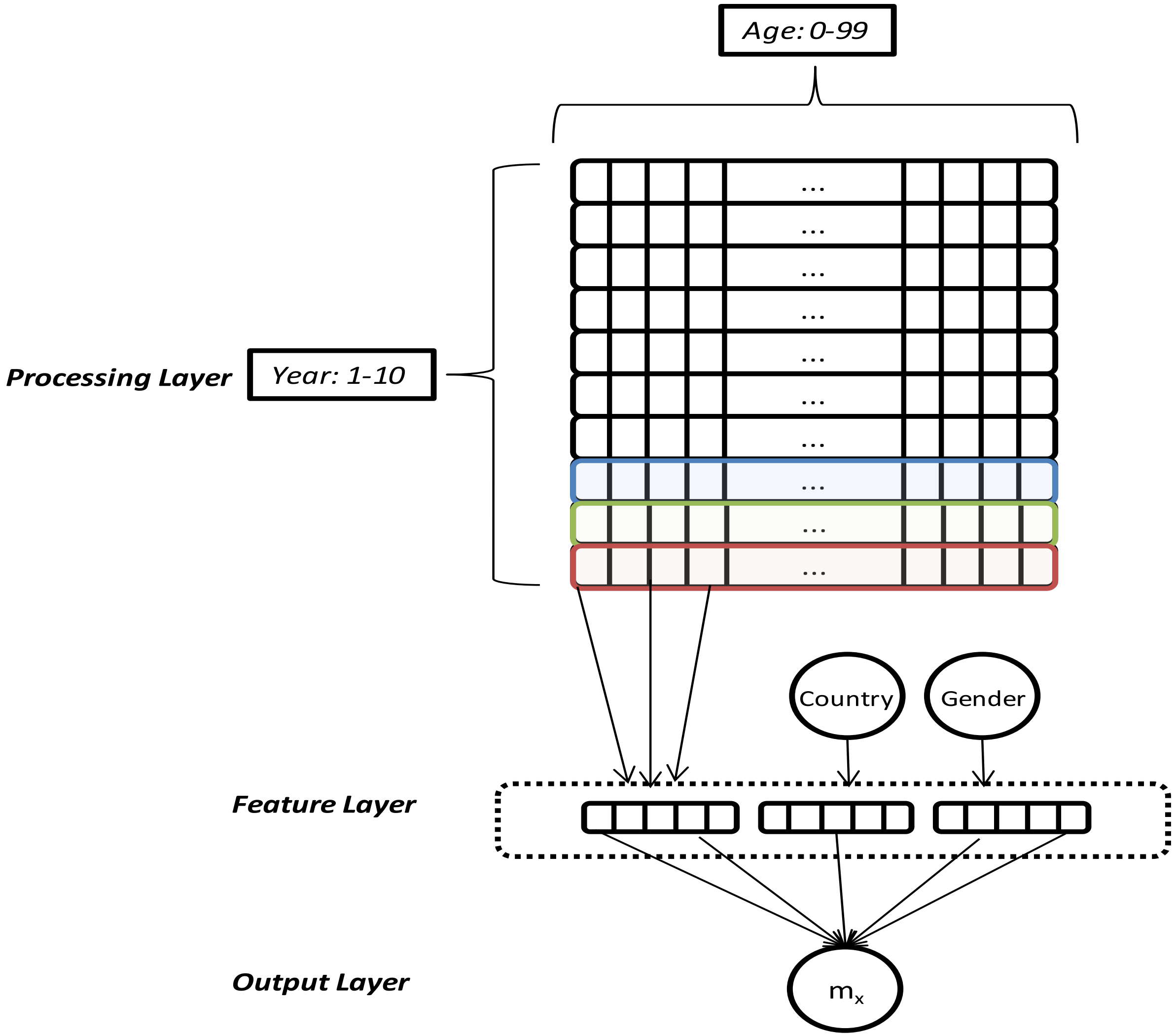
- **Addressed in Perla, Richman, Scognamiglio and Wüthrich (2020)**

LCCONV - MODEL STRUCTURE

- **Similar to LC model…**
- **… however, time variable replaced with outputs of a NN processing layer**
- **Best performance achieved with no non-linearities in the model**
- **Since features are used immediately for prediction, can be interpreted as an extended LC model:**

$$\sigma^{-1}\left(\widehat{y}^{(r,g)}_{x,t_0+T+1}\right) = w_{x,0} + \langle W^{\mathcal{R}}_x, z_{\mathcal{R}}(r)\rangle + \langle W^{\mathcal{G}}_x, z_{\mathcal{G}}(g)\rangle + \left\langle W^f_x, \boldsymbol{z}_f(U^{(i)}_{t_0})\right\rangle$$

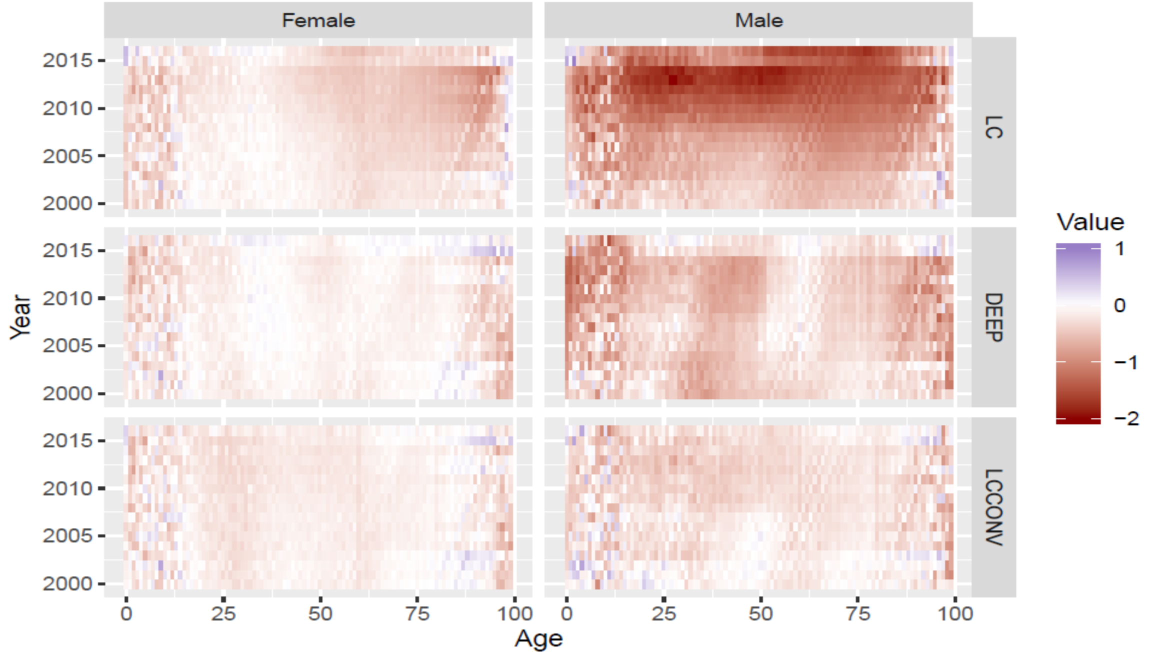
  - **First terms equivalent to ax**
  - **Last term equivalent to bx.kt**



FORECAST RESULTS

- **The CNN model (LCCONV) achieves better performance versus the LC model on 75/76 populations in the HMD**
- **Unadjusted model also generalized well – beat LC on 101/102 populations in the USMD**
- **LCCONV beats the LCNN model in an extra 8 populations and achieves a substantially lower out-of-sample MSE**
- **Residual plot shows that model is substantially better for males, whereas the performance is similar for females**

model	test_loss	ensemble MSE	# populations
LCCONV	2.27	2.24	75/76
LCCONV_tanh	2.62	2.58	61/76
LCCONV_relu	3.26	3.10	57/76
LCLSTM1	2.86	2.54	69/76
LCLSTM1_tanh	3.32	3.03	58/76
LCLSTM1_relu	3.33	3.25	52/76
LCLSTM2	2.43	2.32	74/76
LCLSTM2_tanh	2.36	2.27	75/76
LCLSTM2_relu	3.44	3.11	56/76
DEEP	2.83	2.53	67/76

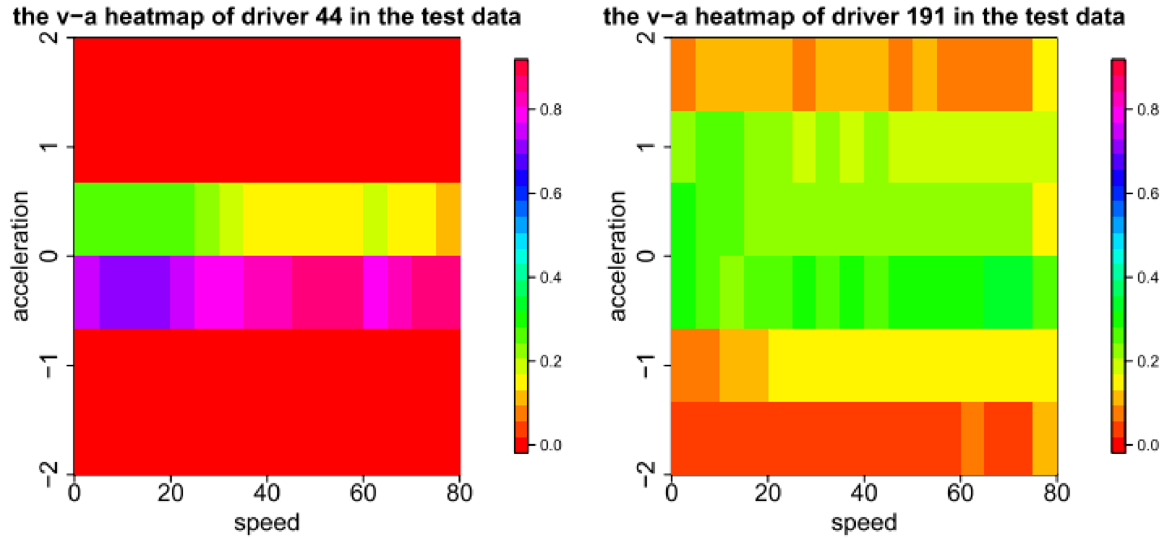


TELEMATICS ANALYSIS

- **Non-life pricing often performed using GLMs**
- **Traditional covariates relate to *policyholder, driver and* vehicle characteristics**
- **Recently enhanced through deriving features from telematics data**
- **Features of telematics data:**
  - **High dimensional**
  - **Sampled at high frequency**
  - **Incorporate physical measurements (location, seed, acceleration)**
- **Simple approaches include considering number/rate of unwanted events (see Guillen, Nielsen, & Pérez-Marín (2021))**
- **Other approaches – summarize data in feature matrices for further analysis**

V-A HEATMAPS

- **Velocity-Acceleration heatmaps due to Wüthrich (2017)**
- **Form 2S density plot of velocity and acceleration based on telematics data**



- **Can be analyzed using traditional methods e.g. k-means or PCA**
- **Recent work (Gao, Wang, & Wüthrich, 2021) analyzes heatmaps directly using FCN and CNN**

BOOSTING PRICING MODEL WITH TELEMATICS

- **Combine traditional actuarial covariates with telematics data using boosting:**
  - **First model – GLM using actuarial covariates**
  - **Second model – once fit, add neural network component to improve calibrations**

$$Y_i \overset{\text{ind.}}{\sim} \text{Poisson}(e_i \widehat{\lambda}(\boldsymbol{x}_i) \rho^{\text{dnn}}(\boldsymbol{Z}_i)),$$

$$Y_i \overset{\text{ind.}}{\sim} \text{Poisson}(e_i \widehat{\lambda}(\boldsymbol{x}_i) \rho^{\text{cnn}}(\boldsymbol{Z}_i)),$$

- **Results show that adding covariates learned from heatmaps decreases test set error by ~10%:**

Error	Homogeneous (2.6)	GLM (2.4)	dnn Listing 1	cnn Listing 2	dnn + glm (4.1)	cnn + glm (4.2)
Learning error	1.0717	1.0205	1.0376	1.0415	0.9982	0.9992
Test error	1.1703	1.1230	1.1035	1.1075	1.0655	1.0690
Reduction in test error		0.0473	0.0668	0.0628	0.1048	0.1013

AGENDA

- Deep Learning in 6 slides
- Actuarial Examples of Representation Learning
- **Advances in Deep Learning**
- Applying Deep Learning in Actuarial Science
- Explaining Deep Learning models
- Uncertainty estimation
- Conclusions

NEURAL NETWORK ARCHITECTURES

- **Most state of the art deep learning results use specialized architectures:**
  - **Convolutional neural networks**
  - **Recurrent neural networks**
  - **Embeddings**
- **Older ideas enhanced by modern approach to neural networks:**
  - **More powerful computing (GPUs/TPUs)**
  - **Larger datasets (ImageNet/NLP corpora)**
  - **Advances in methodology: dropout, batchnorm, ReLu**
- **Led to state of the art advances on problems across many areas of machine learning:**
  - **Computer vision**
  - **Natural language processing**
  - **Speech recognition**

ATTENTION IS ALL YOU NEED

- **Newer approach proposed in 2017 relies on *attention mechanisms***
- **One of most cited papers in machine learning (>23k):**

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

- **Proposed in context of machine translation**
- **Extended to other NLP tasks, computer vision and more recently, tabular data**

ATTENTION

- **Most actuarial models rely on fixed relationship between covariates (features) and outcomes; more flexible models (varying coefficient models) in statistical literature allow coefficients of models to vary with time (or other covariates)**

- **Generalized approach to allow for varying relationships between covariates and outcomes is *attention* (example from Xu et al., 2015)**

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background.

A little girl sitting on a bed with a teddy bear.

A group of people sitting on a boat in the water.

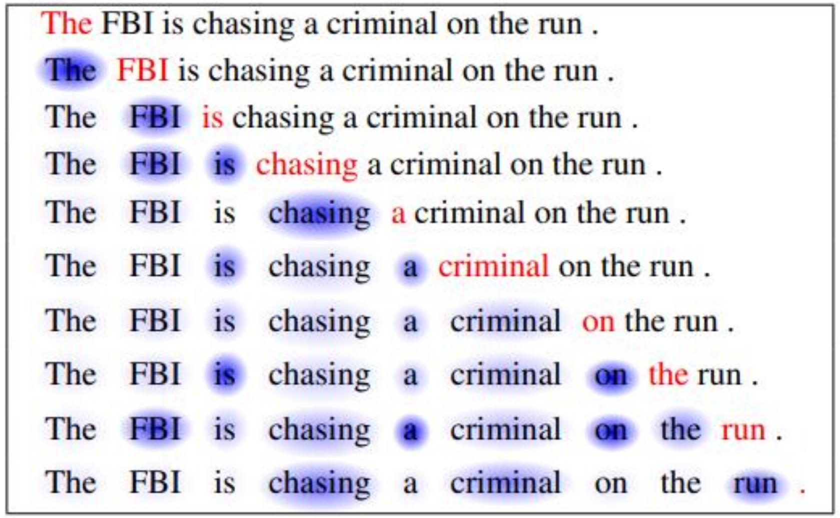
A giraffe standing in a forest with trees in the background.

INTUITIONS OF SELF-ATTENTION

- **When building models, relationships between covariates and outcomes may depend on *context.***

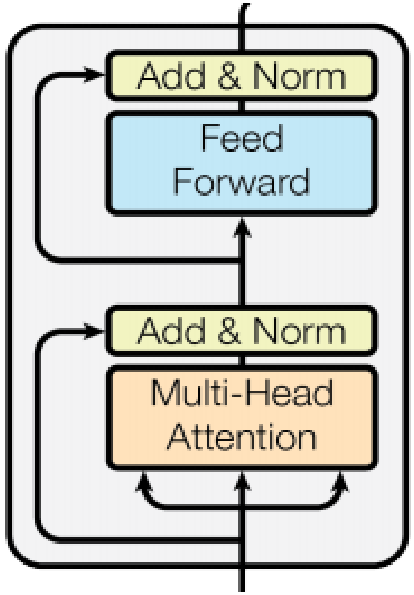
  Famous example in non-life insurance – young drivers and male drivers often have increased frequency, but *young male* drivers may experience even higher frequency => context allow for via interaction effect

- **Automated method for building context into models: *self-attention* i.e. apply attention over inputs; sequence example (Cheng, Dong, & Lapata, 2016)**



- Transformer models apply self-attention to augment model covariates depending on context (Vaswani et al., 2017):

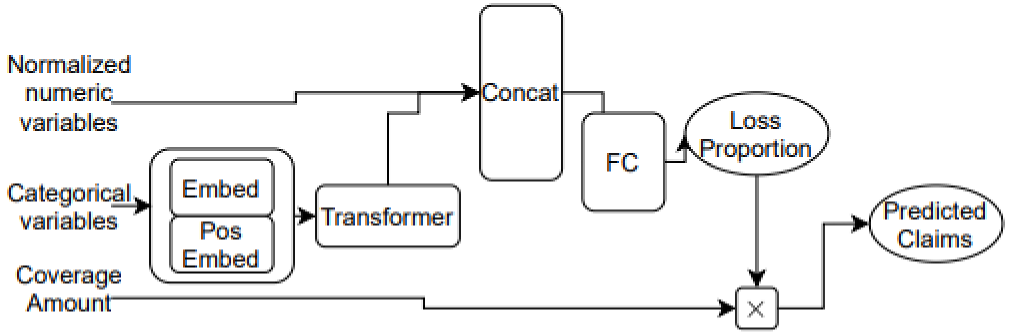
  1. *Inputs fed to multiple self-attention components*
  2. *Attention results added to original input and then centred/scaled*
  3. *Then fed into feed forward network*
  4. *Attention and feedforward results added and then centred/scaled*



APPLYING TRANSFORMERS FOR ACTUARIAL MODELLING

- **Transformer model recently applied in Kuo & Richman (2021) to model flood loss severity**

- **Model tries to predict damage ratio (proportion of exposure damaged) using several continuous and categorical covariates**



- **Best results achieved using Transformer based model**

Model	RMSE	MAE
Model 4: MLP with multidimensional embeddings	60,973	37,574
Model 5: Simple Attention	60,601	36,739
Model 6: TabNet	62,938	38,386
Model 7: TabTransformer	59,900	36,343
Linear regression, predictions capped below at 0.01	60,879	38,431

EMBEDDING LAYER – CATEGORICAL DATA

- **One hot encoding expresses the prior that categories are orthogonal => similar observations not categorized into groups**
- **Embedding layer prior – related categories should cluster together**
- **Learns dense vector transformation of sparse input vectors and clusters similar categories together**

	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

CONTEXTUAL EMBEDDINGS

- **Embeddings for categorical covariates + Self-attention for context = *contextual embeddings***
- **Example shown for flood zone embeddings (left) colored according to house design variable (crawl space)**

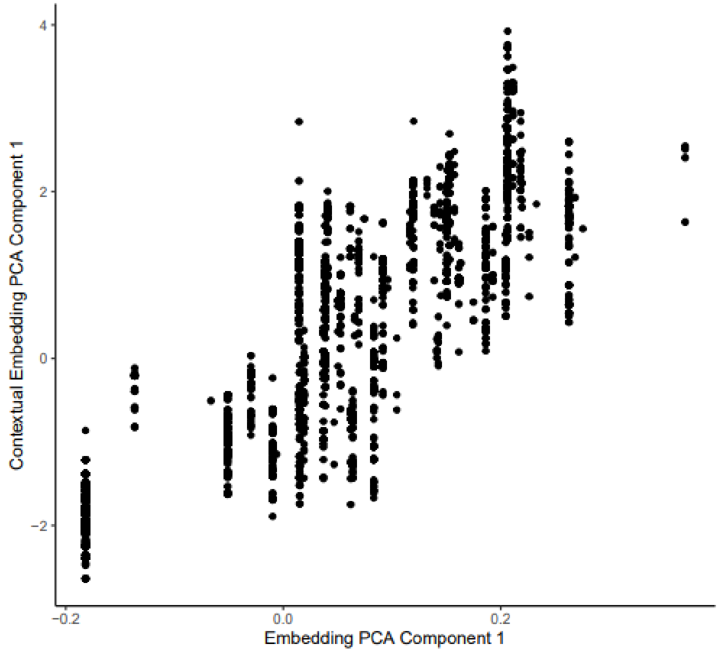
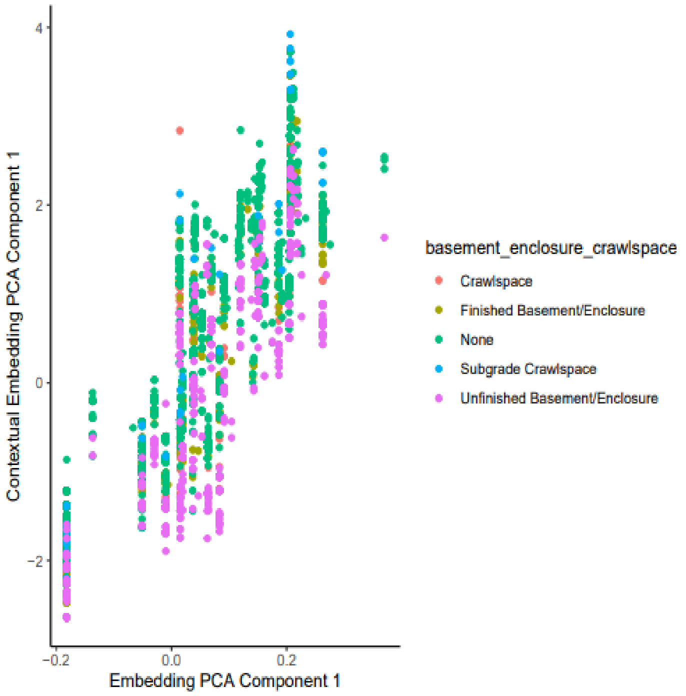


Figure 8: Embedding versus Contextual Embedding, Flood Zone, first PCA components

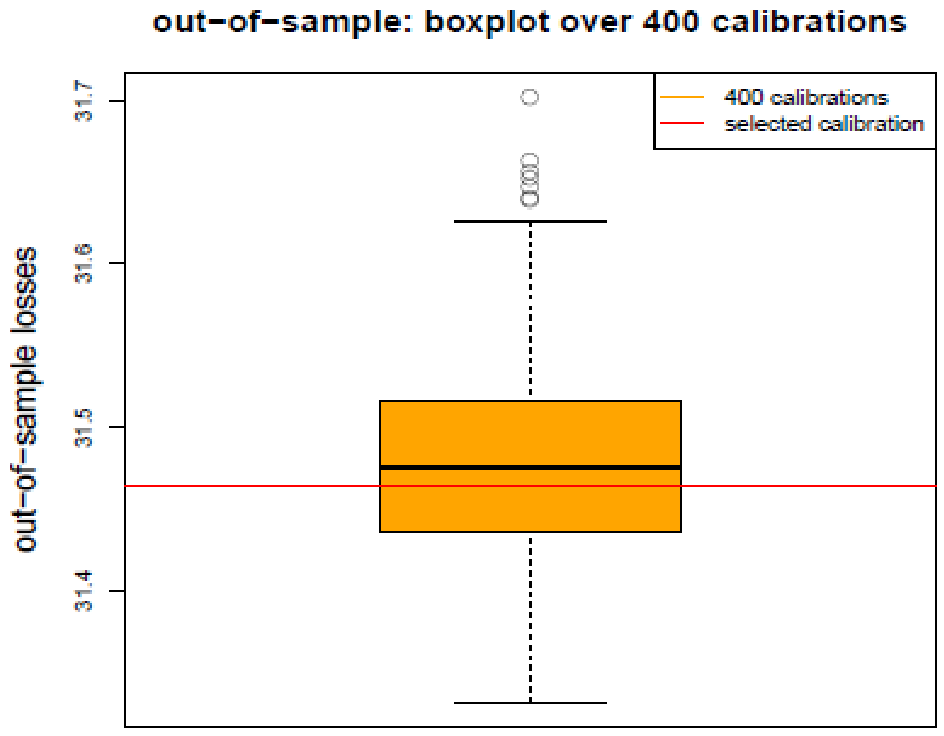


AGENDA

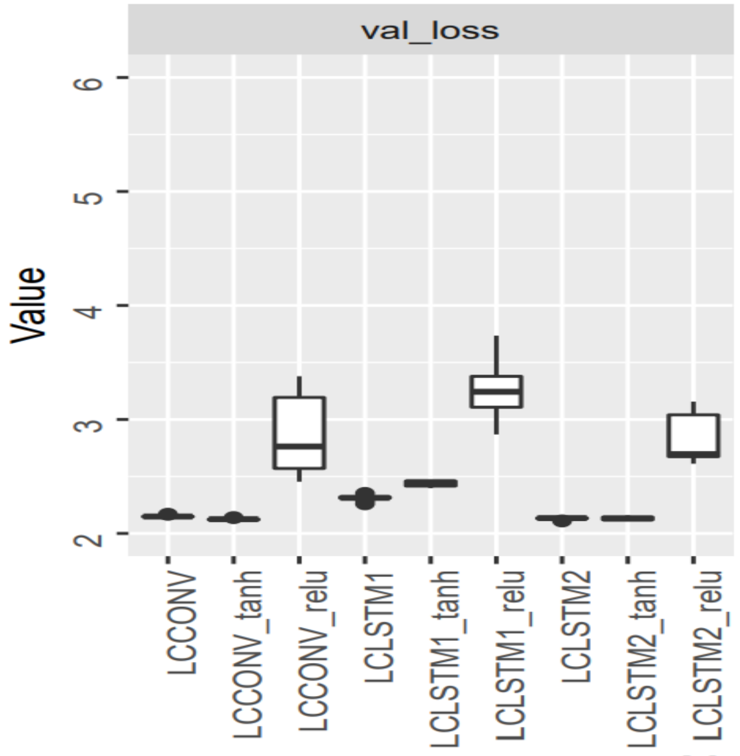
- Deep Learning in 6 slides
- Actuarial Examples of Representation Learning
- Advances in Deep Learning
- **Applying Deep Learning in Actuarial Science**
- Explaining Deep Learning models
- Uncertainty estimation
- Conclusions

RANDOM TRAINING, RANDOM OUTCOMES

- Neural network training incorporates random processes
- Fundamental source of randomness: random initialization of starting parameters and not training to convergence (early stopping)
- Other sources:
  - random ordering of batches fed to network
  - dropout = randomly switch off parts of the network to regularize
  - within vision models – random data augmentation
- Leads to robust models on the one hand…
- … and models that depend on the random seed (i.e. are not reproducible) on the other



Neural networks fit to French MTPL dataset
Richman and Wüthrich (2020)



Neural networks fit to HMD dataset
Perla, Richman, Scognamiglio and Wüthrich (2020)

NAGGING PREDICTORS

- Aggregating is a statistical technique that helps to reduce noise and uncertainty in predictors and is justified theoretically using the law of large numbers.

- An i.i.d. sequence of predictors is not always available thus, Breiman (1996) combined bootstrapping and aggregating, called bagging.

- Combine networks and aggregating to receive the *nagging* predictor i.e. use multiple network predictors for aggregation (Richman & Wüthrich, 2020)

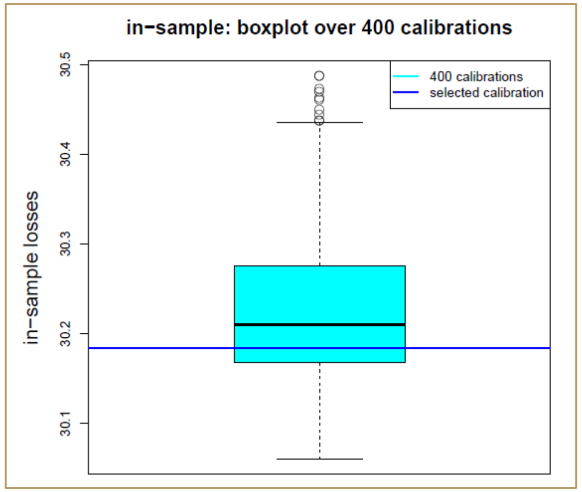
- => Same situation as Breiman (1996) after having received the bootstrap samples

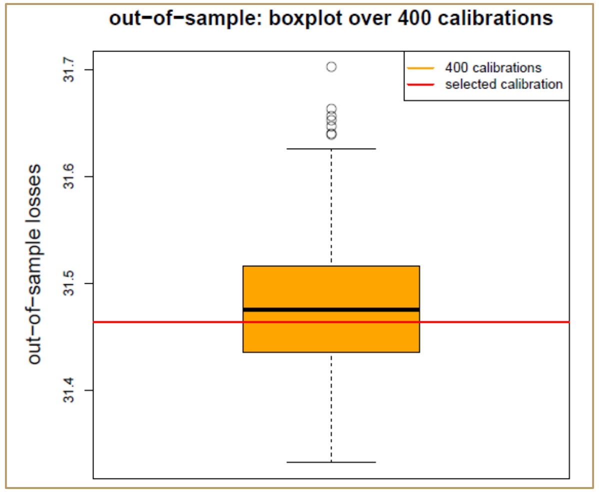
- Leads to more stable results and enhanced predictive performance.

NAGGING RESULTS (1)

- **Applied nagging to French MTPL data and fit 400 networks**

	In-Sample Loss on $\mathcal{D}$		Out-of-Sample Loss on $\mathcal{T}$	
(a) homogeneous model	32.935		33.861	
(b) generalized linear model	31.267		32.171	
(c) boosting regression model	30.132		31.468	
(d) network regression model (seed $j = 1$)	30.184		31.464	
(e) average over 400 network calibrations	30.230	(0.089)	31.480	(0.061)
(f) nagging predictor for $M = 400$	30.060		31.272	





NAGGING RESULTS (2)

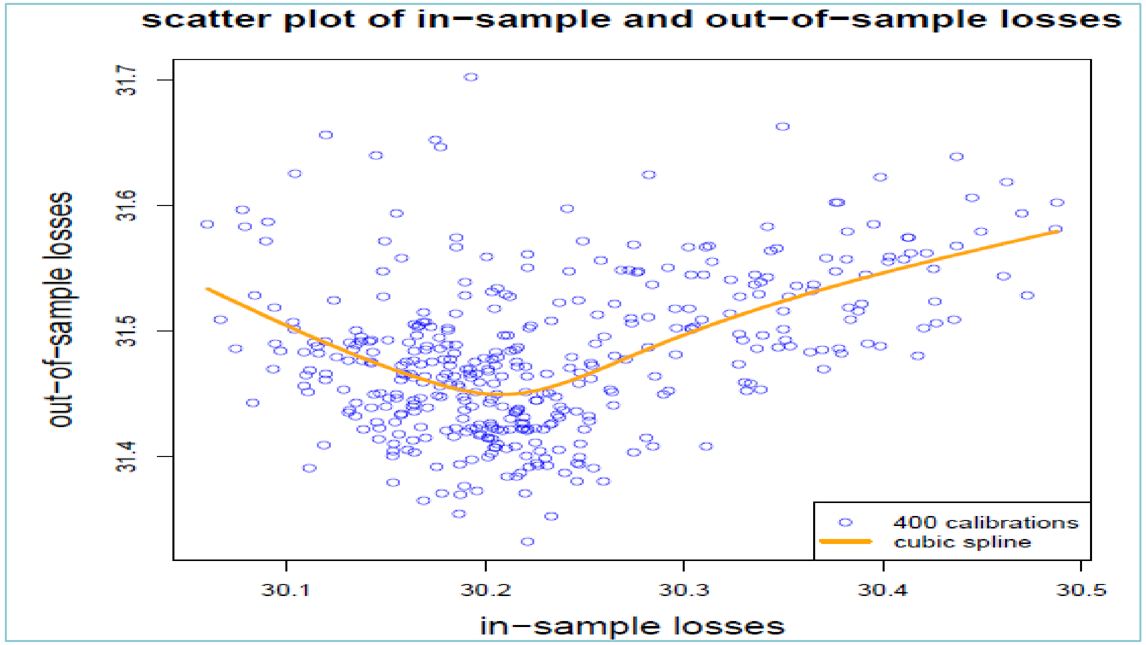
- **Shape of training losses versus testing losses => don't underfit or overfit the training data**

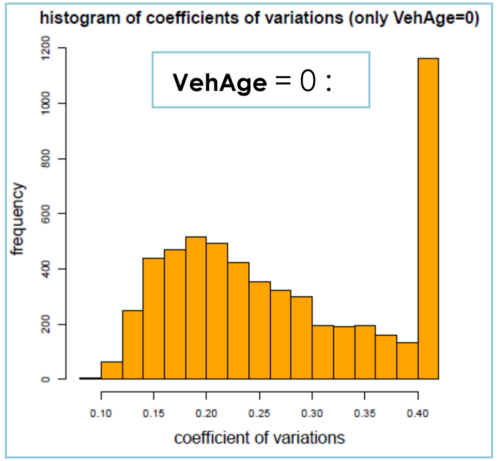
- **Diagnostic for model convergence for individual observations = CoV of predictions**

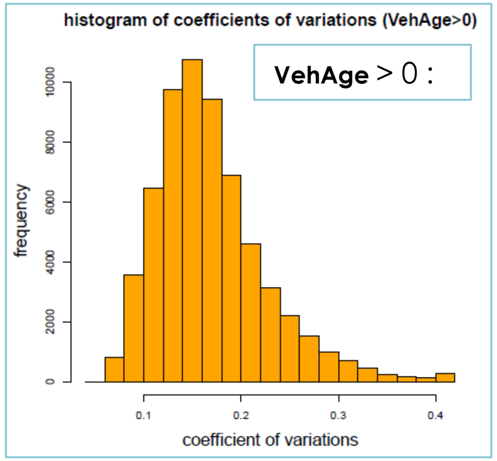
$$\widehat{\mathrm{CoV}}_t = \frac{\widehat{\sigma}_t}{\bar{\bar{\mu}}_t^{(M)}} = \frac{\sqrt{\frac{1}{M-1}\sum_{j=1}^{M}\left(\widehat{\mu}_t^{(j)} - \bar{\bar{\mu}}_t^{(M)}\right)^2}}{\bar{\bar{\mu}}_t^{(M)}}, \qquad (15)$$

- **Allows for identification of observations that are harder to fit**

- **Within French MTPL data, observations with vehicle age 0 appear to have different properties**







SOURCES OF RANDOMNESS

- **How is stability of neural networks affected by different choices of architecture, regularization and training procedure?**

- **Investigated in Richman (2021) in context of mortality forecasting**

- **Following components tested:**
  - **dimension of the intermediate layers**
  - **dimension of the embedding and convolutional layers**
  - **activation function of the intermediate layers**
  - **application of batch normalization**
  - **depth of the network**
  - **drop-out rates**
  - **size of batches**
  - **learning rate, restarts and optimizer**

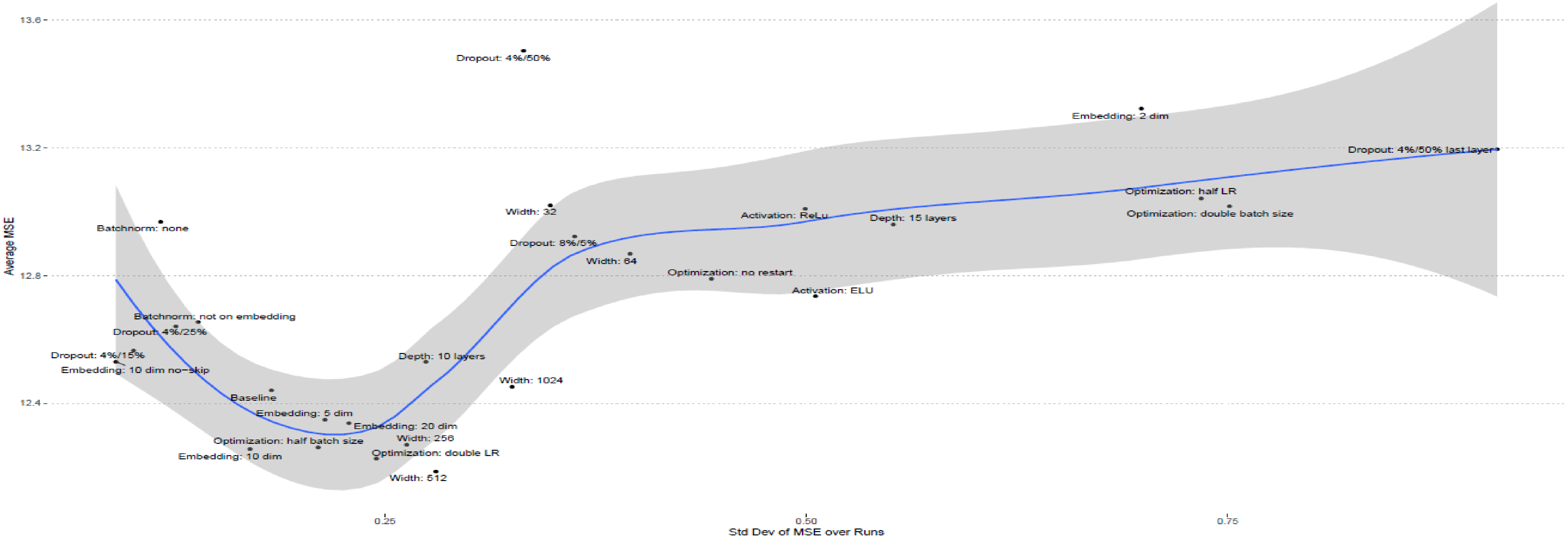
	Model	type	Average MSE	Median MSE	Best Performance
1	LC_SVD	National	5.55	2.48	5
2	LC_SVD	Sub-National	22.08	1.48	20
3	DEEP	National	2.38	1.31	71
4	DEEP	Sub-National	20.49	0.78	216

	Description	Average MSE	Std Dev of MSE
1	Baseline	12.55	0.18
2	Width: 256	12.47	0.26
3	Width: 512	12.43	0.28
4	Width: 1024	12.67	0.33
5	Width: 32	13.16	0.35
6	Width: 64	13.01	0.40

	Description	MSE	Best Performance over Populations
1	Width: 512	12.18	296
2	Width: 256	12.27	288
3	Baseline	12.44	287
4	Width: 1024	12.45	279
5	Width: 64	12.87	247
6	Width: 32	13.02	235

THE BIRDSEYE VIEW

- **Plot shows relationship between randomness of outcomes and nagging predictor performance => some variability is good, but too little/too much is bad**



BIAS OF NETWORKS

- **One view on non-life pricing = finding good base rate predictions for portfolio + set of relativities to allow pricing to vary with risk**

- **If using GLM => portfolio base rates reproduced by model i.e. the 'balance property' is preserved**

- **Neural networks and other ML algorithms do *not have this property* so must correct for this, see Wüthrich (2019) and Denuit, Charpentier, & Trufin (2021)**

- **Within life insurance, experience analysis assesses *bias* of predictions using AvE metrics (and only more rarely do we consider predictive accuracy)**

- **See Rossouw & Richman (2019) for discussion of bias regularization in a life reinsurance context**

AGENDA

- Deep Learning in 6 slides
- Actuarial Examples of Representation Learning
- Advances in Deep Learning
- Applying Deep Learning in Actuarial Science
- **Explaining Deep Learning models**
- Uncertainty estimation
- Conclusions

EXPLAINING DEEP LEARNING MODELS

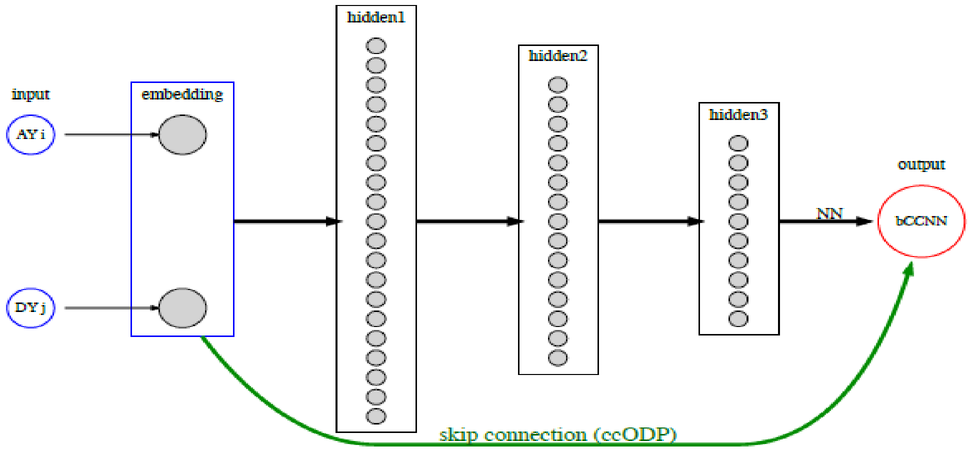
Simulatability	Being able to comprehend the model as a whole	• Develop expert knowledge on model architecture and heuristics to assess models • Issue pertains to some traditional models as well
Decomposability	All components of the model can be inspected and make sense	• Inspect learned representations in last layer of the model • Manual intervention and inclusion of prior knowledge is more difficult
Algorithmic Transparency	Transparency of the learning algorithm and corresponding techniques	• Many techniques to mitigate risks of instability and consistency of DLM still to be developed • Ensembling many models and use of toy models to test outcomes of DLM

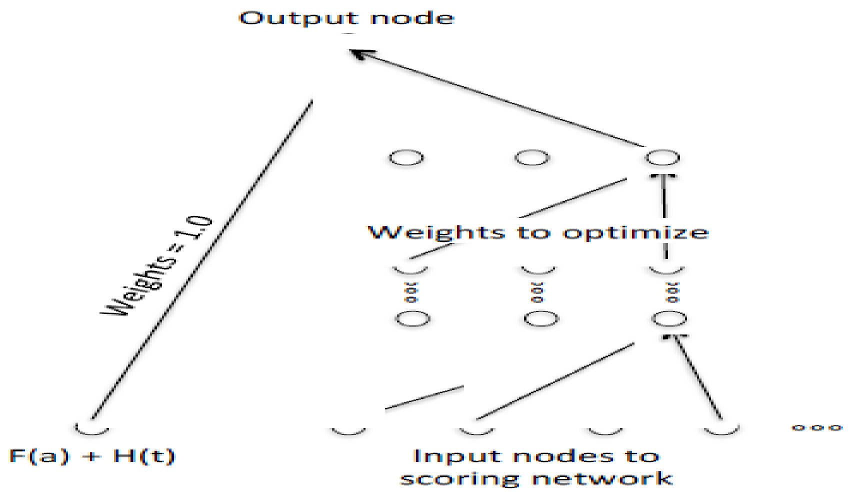
Increased exposure to model risk for ML/DL models and additional controls required

1: Lipton (2016) defines framework for model interpretability to assess two basic questions: Transparency or “How does the model work?” and Post-hoc Interpretability or “What else can the model tell me?”

COMBINED ACTUARIAL NEURAL NET (CANN)

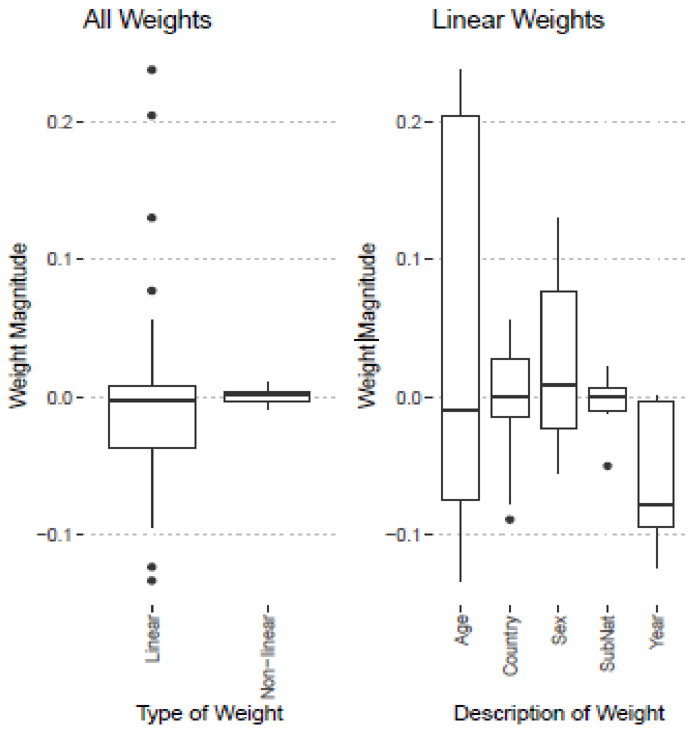
- Combine a traditional actuarial model together with a neural net (Wüthrich and Merz 2018). Implemented so far for pricing and reserving (Gabrielli 2019; Gabrielli, Richman and Wuthrich 2018)
  - Traditional model (calibrated with MLE) directly connected with output of network using skip connection
  - Model output then enhanced by model structure learned by neural net to explain residuals
  - Easy to interpret (and fast to calibrate)
- Shifts the interpretability problem – delta from GLM
- See Breeden and Leonova (2019) who use a similar proposal to incorporate prior economic information into a credit model

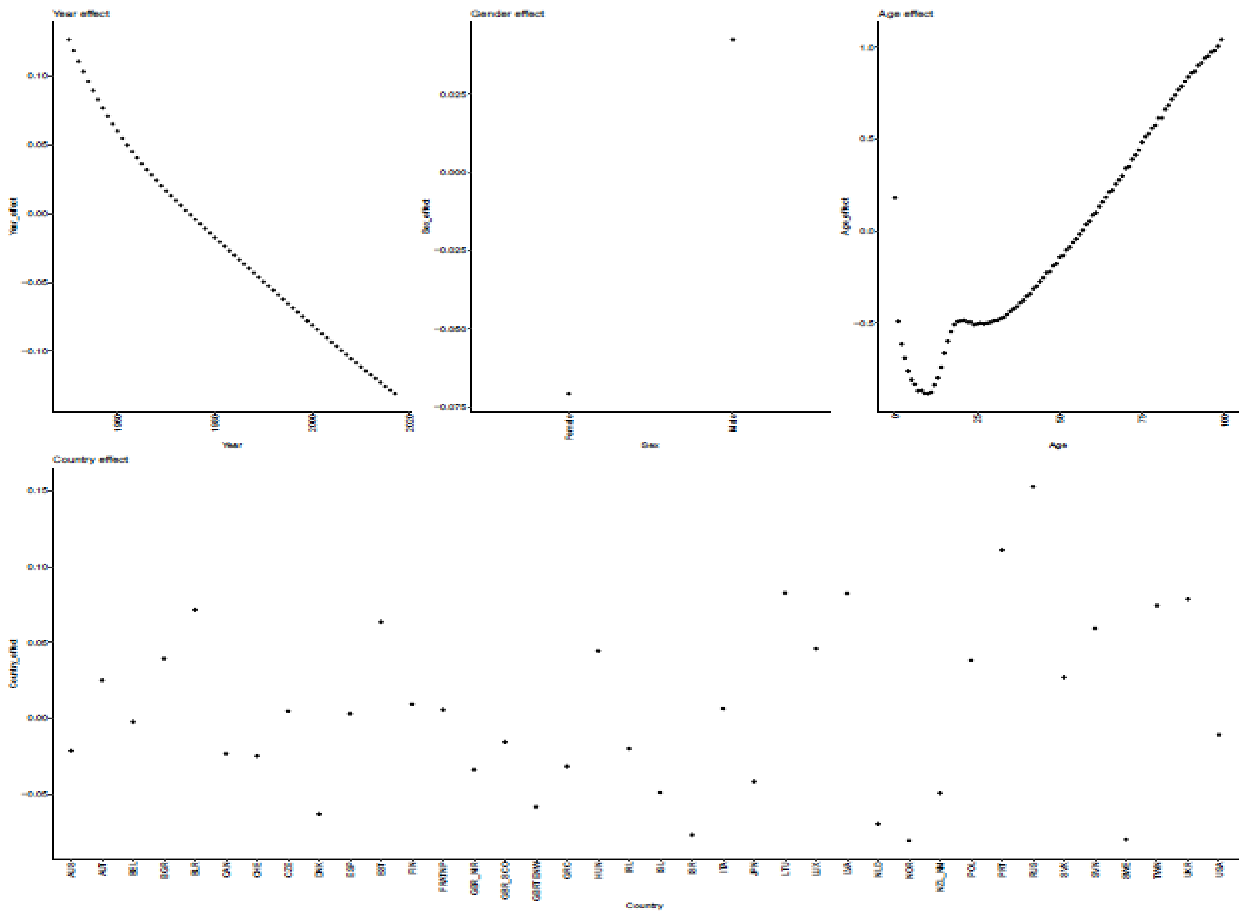




CANN FOR MORTALITY FORECASTING

- **Here we show linear effects from the model and weight magnitude of non-linear component**
- **Analysis in Richman (2021)**

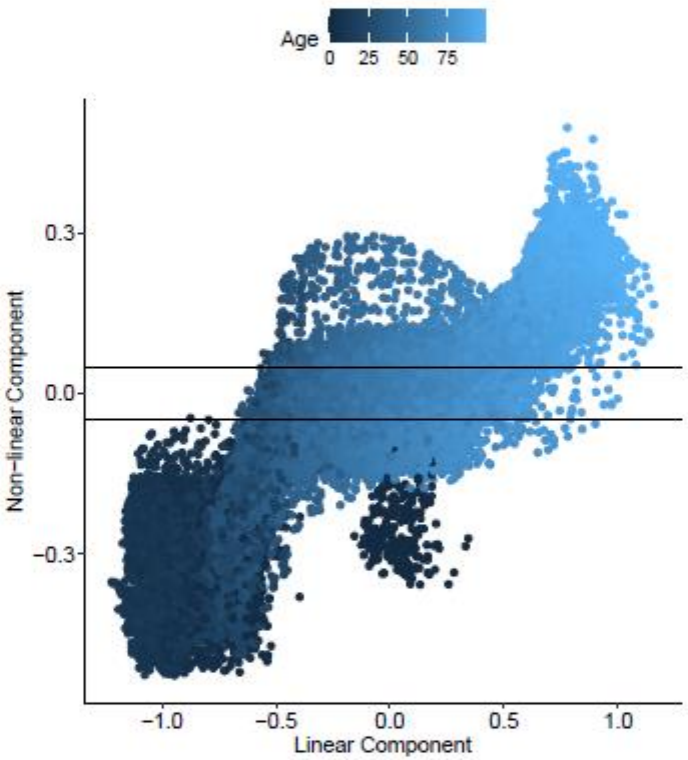


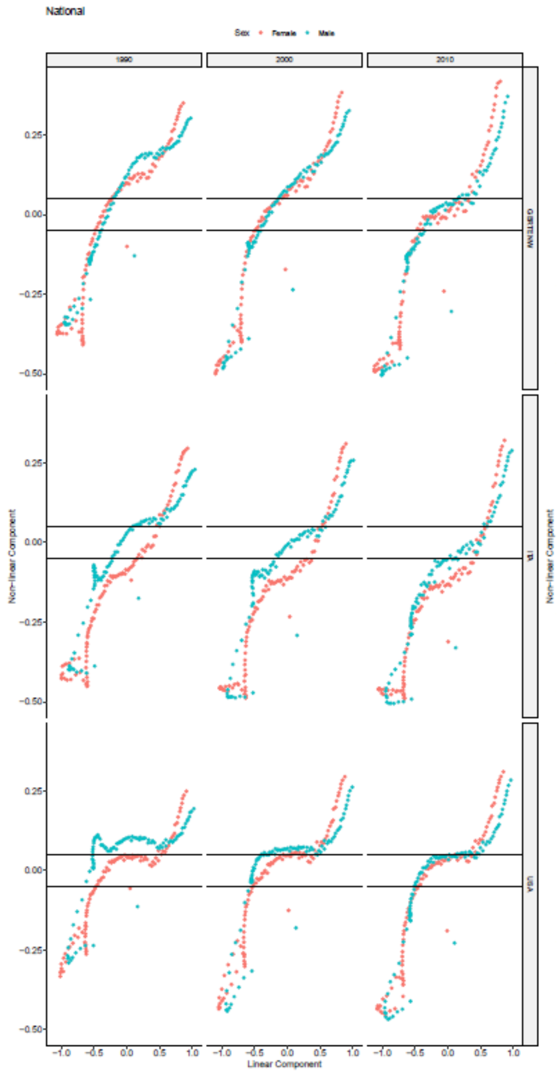


INVESTIGATING NETWORK NON-LINEARITIES

- Use the CANN model to highlight major differences from predictions of traditional model i.e. isolate the network output => model diagnostic

- Complex relationship between linear and non-linear component as function of Year, Country, Gender and Age variables





- Hard to disentangle relationship between inputs and outputs in a deep learning model =>
- Can we use the flexibility/function approximation capability of neural networks to fit specific variable combinations?
- Explainable neural networks (XNNs) and Neural Additive Models (NAM) of Vaughan *et al.* (2018) and Agarwal *et al.* (2020)

$$\hat{y_i} = \mu + \gamma_1 f_1(x_{i,1}) + \gamma_2 f_2(x_{i,2}) + \cdots + \gamma_P f_P(x_{i,P})$$

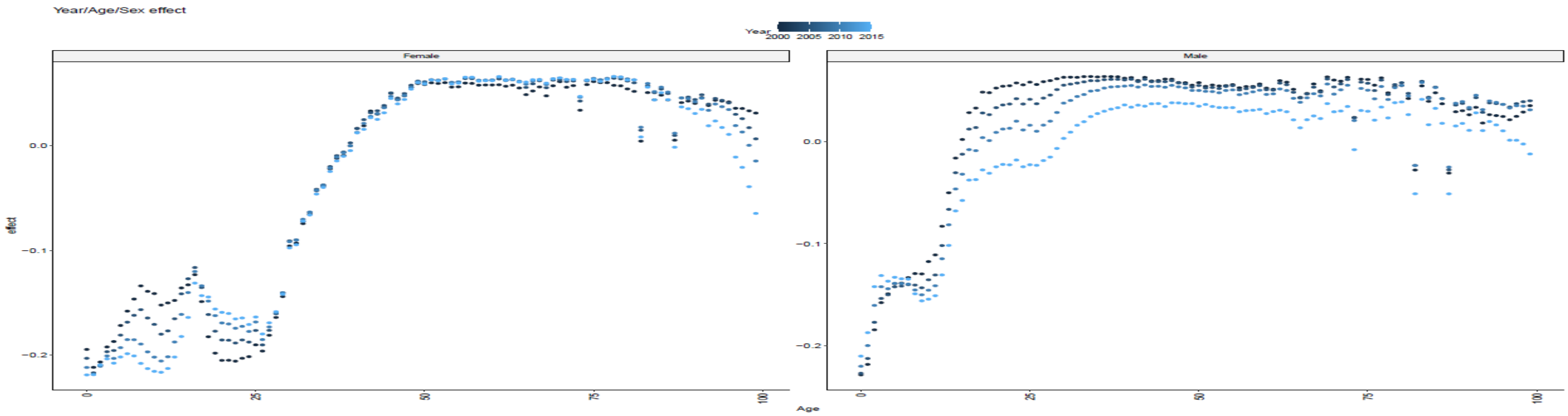
- Combined Actuarial eXplainable Neural Network (CAXNN)
- See Richman (2021) for extensions of XNNs and applications within mortality forecasting

CAXNN RESULTS

- **CAXNN Model reproduces nagging predictor in an explainable manner**

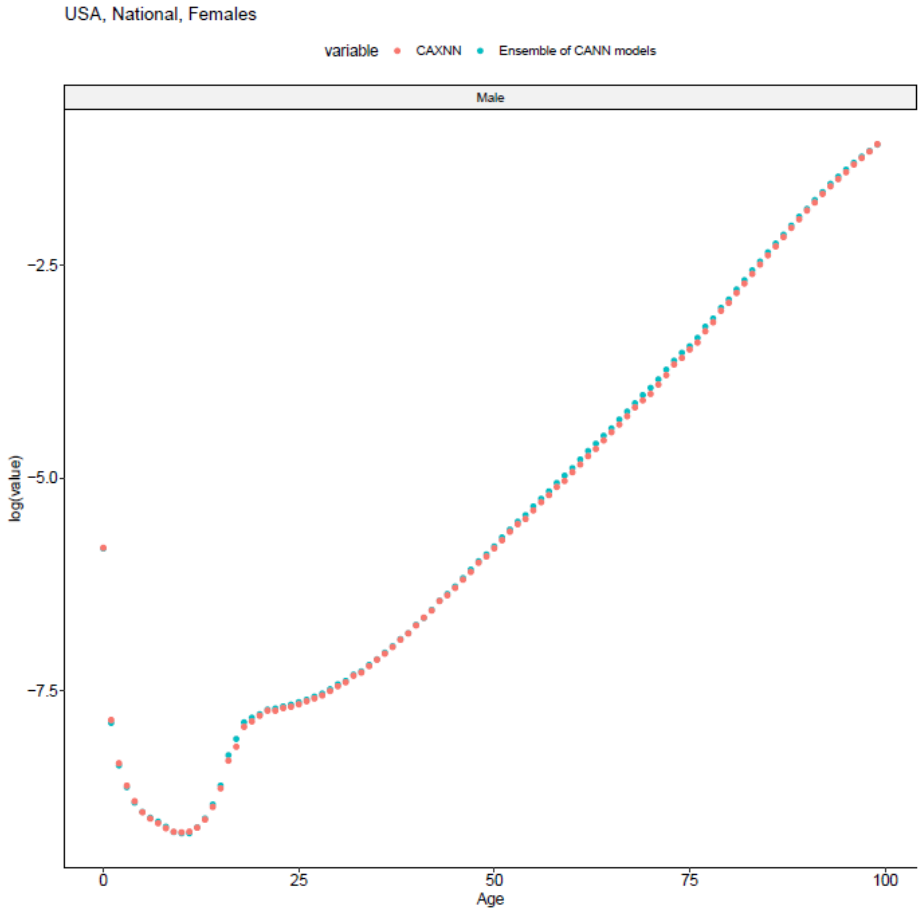
	Model	type	Average MSE	Median MSE	Best Performance
1	LC_SVD	National	5.55	2.48	35
2	LC_SVD	Sub-National	22.08	1.48	43
3	Linear	National	4.28	3.07	41
4	Linear	Sub-National	20.94	0.95	193

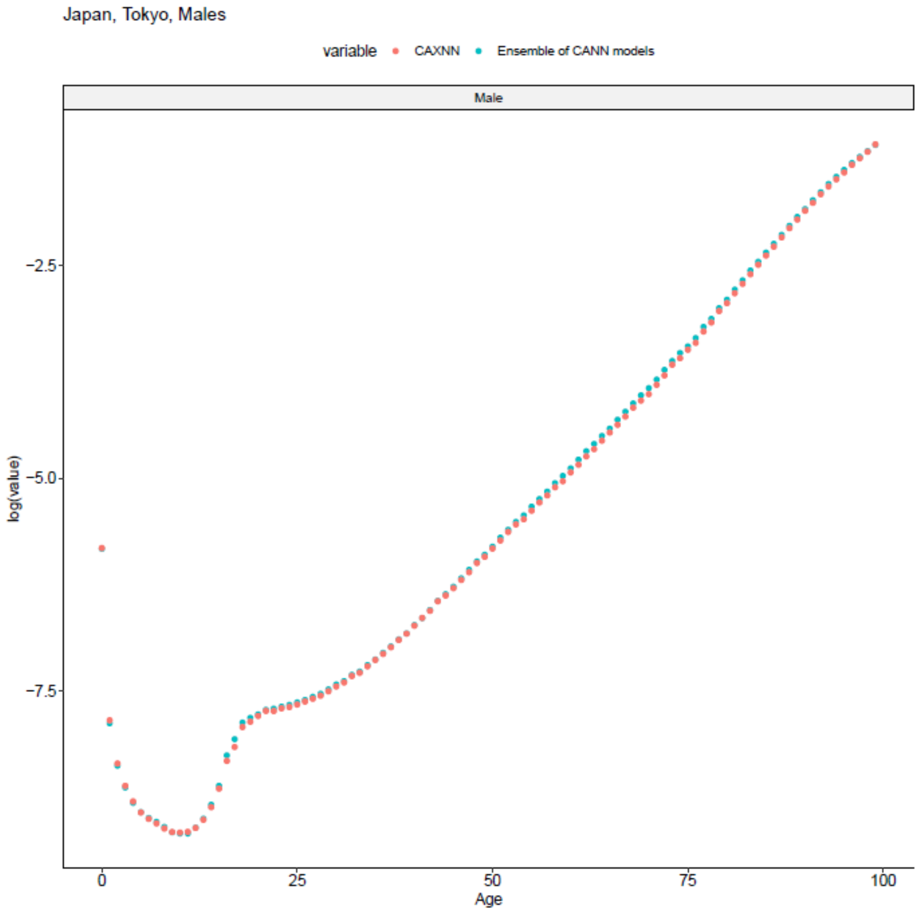
	Model	type	Average MSE	Median MSE	Best Performance
1	LC_SVD	National	5.55	2.48	9
2	LC_SVD	Sub-National	22.08	1.48	21
3	DEEP	National	2.67	1.46	67
4	DEEP	Sub-National	20.60	0.90	215



CAXNN APPROXIMATION OUT-OF-SAMPLE

- **CAXNN Model reproduces nagging predictor in an explainable manner**





AGENDA

- Deep Learning in 6 slides
- Actuarial Examples of Representation Learning
- Advances in Deep Learning
- Applying Deep Learning in Actuarial Science
- Explaining Deep Learning models
- **Uncertainty estimation**
- Conclusions

UNCERTAINTY ESTIMATION

- **Ability to quantify extent of uncertainty in predictions is key to many actuarial tasks; however, focus of deep learning literature is on best estimate**

- **Several approaches proposed in DL literature:**

  - **Use of dropout as an approximation of model uncertainty (Gal 2016; Kendall and Gal 2017)**
  - **Quantile regression to derive prediction bounds (Smyl 2018)**
  - **Use neural networks for GAMLSS regression**

- **Not immediately obvious how to reconcile to traditional actuarial framework (often relies on bootstrapping)**

- **Seemingly, framework of Kendall and Gal (2017) for computer vision correlates with traditional actuarial understanding (model and parameter risk = epistemic uncertainty; process risk = aleatoric uncertainty)**

RECENT WORK

- Gabrielli, Richman and Wüthrich (2019) apply bootstrap to the multi-LoB ODP NN model – found that decreased bias almost to zero but increased RMSEP versus separate ODP models
  - Bootstrap only feasible due to fast calibration of CANN models
- More recently, Schnürch & Korn (2021) apply bootstrapping to neural network-based mortality forecasting models and find intervals thus produced are well calibrated for some of the models
- Also see Marino & Levantesi (2020)

TWO SIMPLE ML METHODS

- In Richman (2021) we have applied methods from the ML/DL literature
- Quantile regression (using pinball loss) shown to produce very well calibrated prediction intervals in M4 Forecasting competition

$$L(y_i, \hat{y}_i, \tau) = \begin{cases} \tau \ (y_i - \hat{y}_i) & \text{if } y_i - \hat{y}_i \ \geq \ 0 \\ (\tau - 1) \ (y_i - \hat{y}_i) & \text{if } y_i - \hat{y}_i < \ 0 \end{cases},$$

- Deep ensembles use heteroskedastic Gaussian regression and multiple training runs to derive prediction intervals (Lakshminarayanan *et al.* 2017):

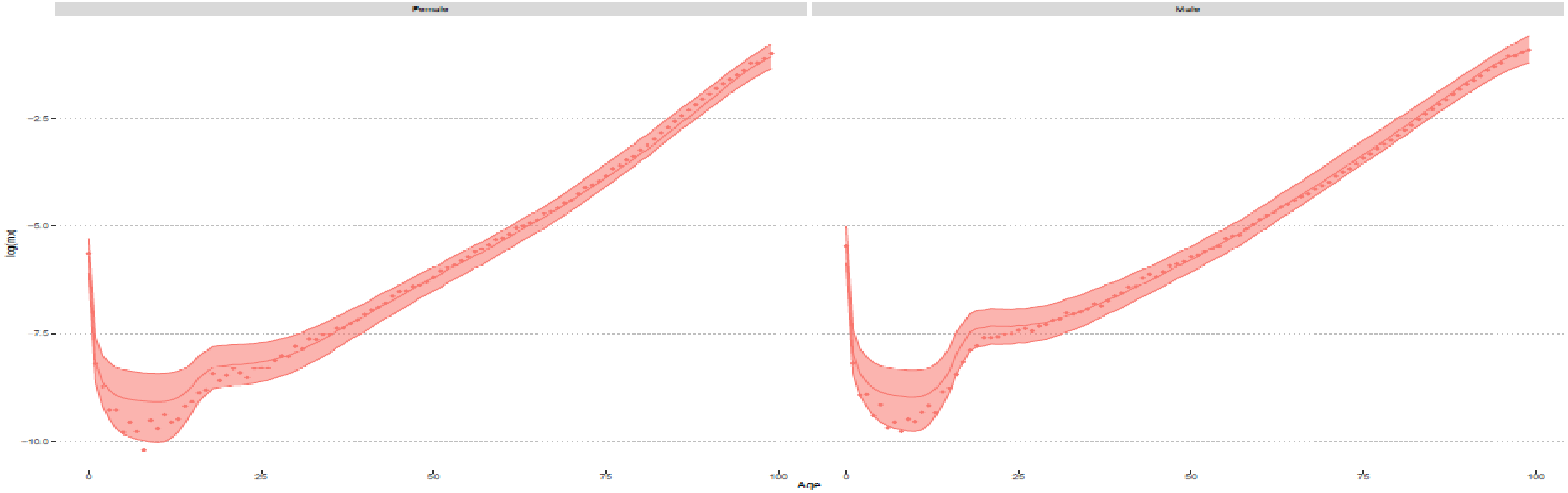
$$L(\hat{y}_i) = \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2} + \frac{\log(\sigma_i^2)}{2}.$$

- Note that DE have Bayesian interpretation (Wilson & Izmailov, 2020)

TWO SIMPLE ML METHODS

- **Achieve excellent empirical coverage out of sample i.e. well calibrated**

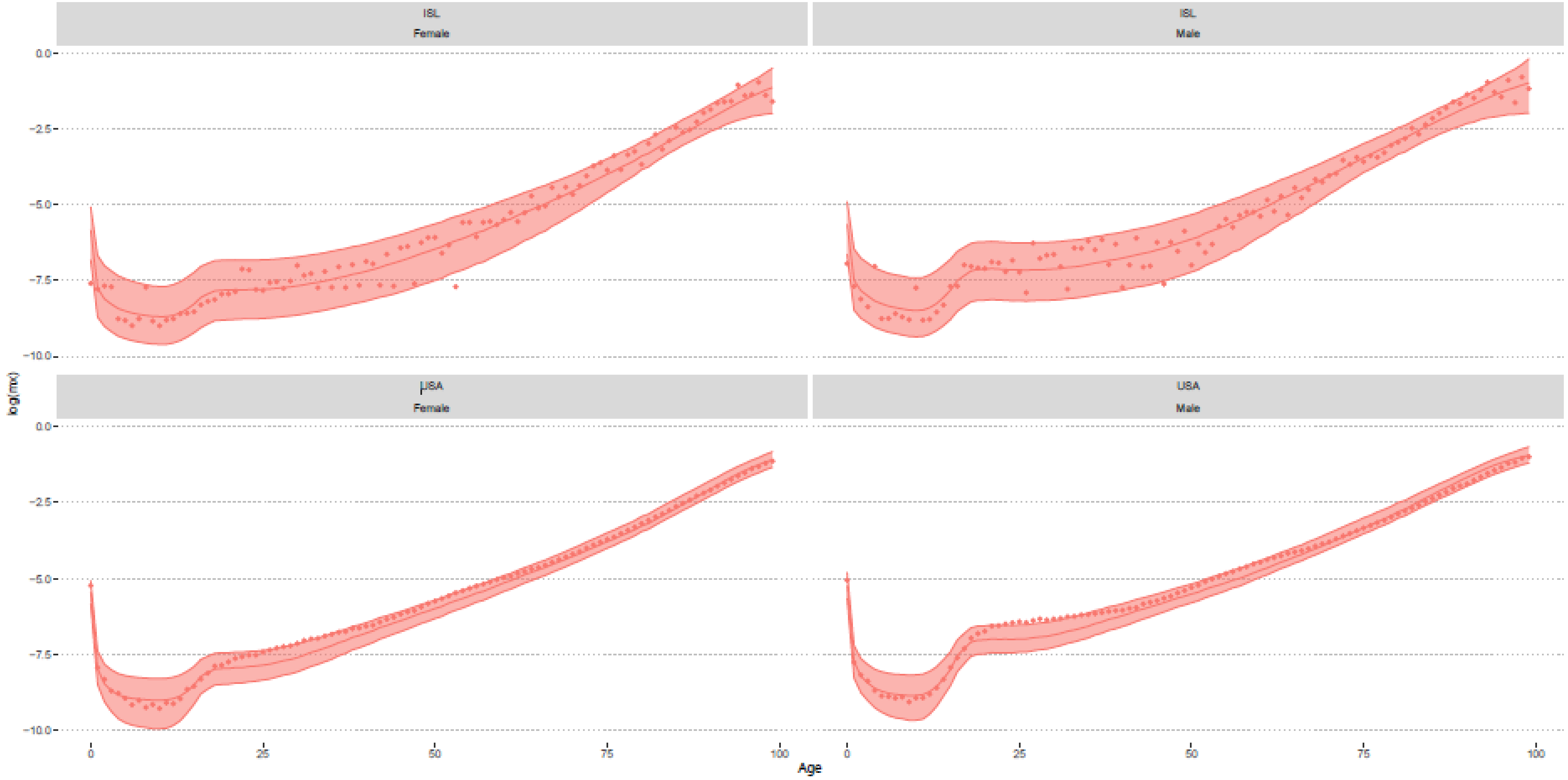
	Description	Data Exceeding 97.5%	Data Exceeding 2.5%	Coverage	Delta
1	Pinball Loss, ReLu branches	0.022	0.026	0.048	0.002
2	Pinball Loss, tanh branches	0.021	0.026	0.048	0.002
3	Deep Ensemble	0.026	0.029	0.055	0.005
4	Deep Ensemble, ReLu branches	0.012	0.030	0.043	0.007
5	Pinball Loss	0.017	0.025	0.042	0.008
6	Deep Ensemble, tanh branches	0.014	0.027	0.041	0.009



- **Perhaps too narrow due to parameter error that was not evaluated**

QUANTILE REGRESSION – INTUITIVE RESULTS

- **Results accord with intuition**



AGENDA

- Deep Learning in 6 slides
- Actuarial Examples of Representation Learning
- Advances in Deep Learning
- Applying Deep Learning in Actuarial Science
- Explaining Deep Learning models
- Uncertainty estimation
- **Conclusions**

WHAT DID WE NOT COVER?

- **Outside of actuarial science:**
  - **Applying transformers for computer vision (Dosovitskiy *et al.* 2020)**
  - **The "self-supervised revolution" (LeCun & Misra, 2021)**
  - **DL versus GBDT (Kadra *et al.* 2021)**
- **Within actuarial science:**
  - **Last layer analysis (Richman, von Rummell, & Wüthrich, 2019)**
  - **Marginal attribution by condition on quantiles (Merz, Richman, Tsanakas, & Wüthrich, 2021)**
  - **Discrimination free pricing (Lindholm, Richman, Tsanakas, & Wüthrich, 2020)**

CONLCUSIONS

- **Deep learning:**
  - opens new possibilities for actuarial modelling by solving difficult model specification problems, especially those involving large scale modelling problems
  - allows new types of high frequency data to be analysed
  - enhances the predictive power of models built by actuaries
- **Recent work has expanded the toolkit of actuarial data science by:**
  - applying representation learning directly on novel data sources
  - applying new DL methods
  - showing how DL models can be made explainable/interpretable
- **More work is needed on uncertainty estimation**

european actuarial academy

LEARN MORE

- Reading club to go through new book by Mario Wüthrich and Michael Merz
  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3822407

STATISTICAL FOUNDATIONS OF ACTUARIAL LEARNING
AND ITS APPLICATIONS

MARIO V. WÜTHRICH
RISKLAB SWITZERLAND
DEPARTMENT OF MATHEMATICS
ETH ZURICH

MICHAEL MERZ
FACULTY OF BUSINESS ADMINISTRATION
HBS – HAMBURG BUSINESS SCHOOL
UNIVERSITY OF HAMBURG

- Contact me if this is of interest

THANK YOU

- **Mario Wüthrich**
- **Andreas Tsanakas**
- **Michael Merz**
- **Mathias Lindholm**
- **Kevin Kuo**
- **Nicolai von Rummell**
- **Louis Rossouw**

REFERENCES

- Agarwal, R., Frosst, N., Zhang, X., Caruana, R., & Hinton, G. E. (2020). Neural additive models: Interpretable machine learning with neural nets. ArXiv.
- Cheng, J., Dong, L., & Lapata, M. (2016). Long short-term memory-networks for machine reading. EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings, 551–561. https://doi.org/10.18653/v1/d16-1053
- Denuit, M., Charpentier, A., & Trufin, J. (2021). Autocalibration and Tweedie-dominance for Insurance Pricing with Machine Learning. Retrieved from http://arxiv.org/abs/2103.03635
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., … Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Retrieved from http://arxiv.org/abs/2010.11929
- Gabrielli, A., Richman, R., & Wüthrich, M. V. (2019). Neural network embedding of the over-dispersed Poisson reserving model. Scandinavian Actuarial Journal. https://doi.org/10.1080/03461238.2019.1633394
- Gao, G., Wang, H., & Wüthrich, M. V. (2021). Boosting Poisson regression models with telematics car driving data. Machine Learning, 1–30. https://doi.org/10.1007/s10994-021-05957-0
- Guillen, M., Nielsen, J. P., & Pérez-Marín, A. M. (2021). Near-miss telematics in motor insurance. Journal of Risk and Insurance. https://doi.org/10.1111/jori.12340
- Kadra, A., Lindauer, M., Hutter, F., & Grabocka, J. (2021). Regularization is all you Need: Simple Neural Nets can Excel on Tabular Data. Retrieved from http://arxiv.org/abs/2106.11189
- Kuo, K., & Richman, R. (2021). Embeddings and Attention in Predictive Modeling. ArXiv. Retrieved from http://arxiv.org/abs/2104.03545
- Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems, 2017-Decem, 6403–6414.
- LeCun, Y., & Misra, I. (2021). Self-supervised learning: The dark matter of intelligence. Retrieved June 29, 2021, from https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/
- Lindholm, M., Richman, R., Tsanakas, A., & Wuthrich, M. V. (2020). Discrimination-Free Insurance Pricing. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3520676
- Marino, M., & Levantesi, S. (2020). Measuring Longevity Risk Through a Neural Network Lee-Carter Model. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3599821
- Merz, M., Richman, R., Tsanakas, A., & Wüthrich, M. V. (2021). Interpreting Deep Learning Models with Marginal Attribution by Conditioning on Quantiles. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3809674
- Richman, R, von Rummell, N., & Wüthrich, M. V. (2019). Believing the Bot - Model Risk in the Era of Deep Learning. Actuarial Society of South Africa Convention 2019. Retrieved from https://www.actuarialsociety.org.za/wp-content/uploads/2019/10/2019-RichmanVRummellWuthrich-FIN.pdf
- Richman, Ronald. (2021). Mind the Gap-Safely Incorporating Deep Learning Models into the Actuarial Toolkit. SSRN Electronic Journal. Retrieved from https://ssrn.com/abstract=3857693
- Richman, Ronald, & Wüthrich, M. V. (2020). Nagging Predictors. Risks, 8(3), 83. https://doi.org/10.3390/risks8030083
- Rizzi, S., Halekoh, U., Thinggaard, M., Engholm, G., Christensen, N., Johannesen, T. B., & Lindahl-Jacobsen, R. (2019). How to estimate mortality trends from grouped vital statistics. International Journal of Epidemiology, 48(2), 571–582. https://doi.org/10.1093/ije/dyy183
- Rossouw, L., & Richman, R. (2019). Using Machine Learning to Model Claims Experience and Reporting Delays for Pricing and Reserving. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3465424
- Schnürch, S., & Korn, R. (2021). Point and Interval Forecasts of Death Rates Using Neural Networks. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3796051
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 2017-Decem, 5999–6009.
- Vaughan, J., Sudjianto, A., Brahimi, E., Chen, J., & Nair, V. N. (2018). Explainable neural networks based on additive index models. ArXiv. Retrieved from http://arxiv.org/abs/1806.01933
- Wilson, A. G., & Izmailov, P. (2020). Bayesian Deep Learning and a Probabilistic Perspective of Generalization. ArXiv. Retrieved from https://github.com/izmailovpavel/
- Wüthrich, M. V. (2019). Bias regularization in neural network models for general insurance pricing. European Actuarial Journal, 1–24. https://doi.org/10.1007/s13385-019-00215-z
- Wüthrich, M. V. (2017). Covariate selection from telematics car driving data. European Actuarial Journal, 7(1), 89–108.
- Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., … Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. 32nd International Conference on Machine Learning, ICML 2015, 3, 2048–2057.

Ron is an experienced actuary and risk manager, currently Managing Head of Insurance Actuarial at SA Taxi. Before this he was an Associate Director at QED Actuaries and Consultants, Africa's largest independent actuarial consulting firm, where he was responsible for client work on life and general insurance clients and performing research into applications of machine learning and AI to actuarial and insurance topics. Prior to this, he led the Enterprise Risk Management and Actuarial Functions for the AIG group within Africa.

Ron is a Fellow of the Institute and Faculty of Actuaries (IFoA) and the Actuarial Society of South Africa (ASSA), holds practicing certificates in Short Term Insurance and Life Insurance from ASSA, and a Masters of Philosophy in Actuarial Science, with distinction, from the University of Cape Town.

Ron chairs the Actuarial Society of South Africa's ERM committee and is a member of the ASTIN Board.

ABOUT ME

Ron Richman
SA Taxi

EAA e-Conference on
Data Science & Data Ethics

29 June 2021

Contact

Ron Richman

SA Taxi

+27 79 133 7248

ron@ronaldrichman.co.za