ON CLAIMS RESERVING WITH MACHINE LEARNING TECHNIQUES

Vilma Guevara Härkönen

Actuary, Folksam

Folksam

AGENDA

Introduction

Methods

Results

Conclusions

INTRODUCTION

01

Insurance companies set aside money to be able to pay future compensation for damages or claims that may arise later. A reserve is like a financial buffer.

02

To estimate the size of the reserve, well-established traditional statistical models that have been industry standards for decades are used.

03

Machine learning methods are still relatively new and have not yet been widely used in the industry for reserving.

04

Implement three different machine learning models to estimate the reserve and cash flows and compare these with traditional models.

05

Machine learning provides more precise estimates of the reserve size and reduces systematic estimation errors compared to traditional models.

RESERVING MODELS

What do we know today?

- Number of claims to date
- Claim cost to date
- Date the damage occurred
- Date the damage is reported to the insurance company
- Date of payments so far

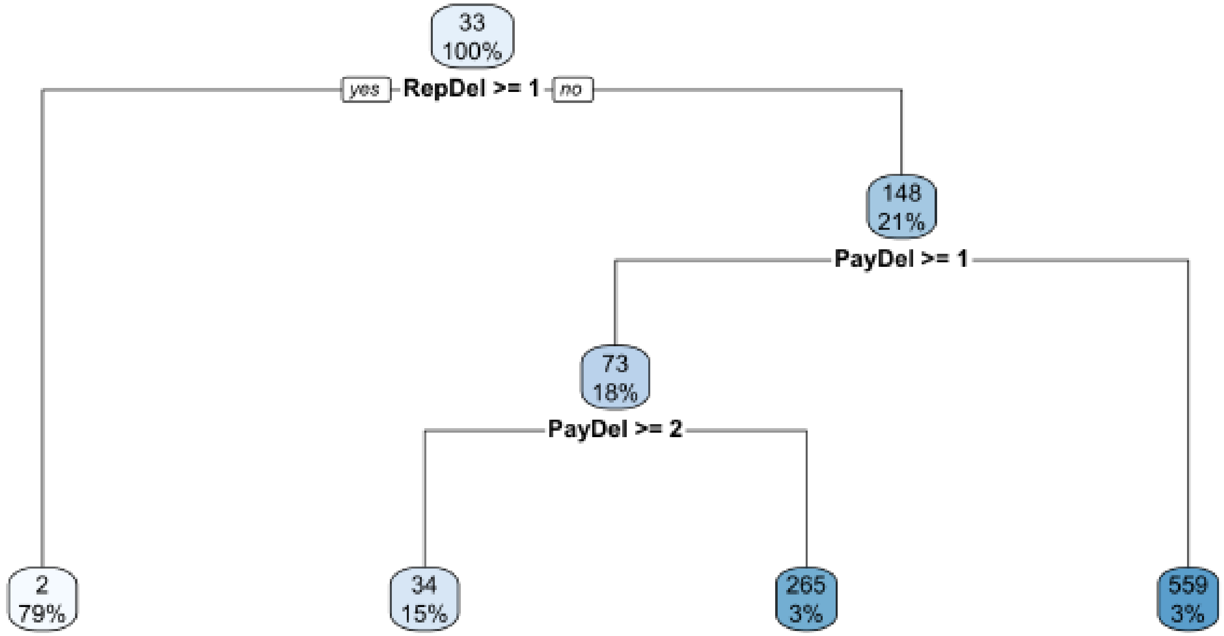
What do we want to achieve?

- Upcoming payments for claims that have already been reported
- The number of claims that have occurred but not yet been reported to the insurance company and their claim cost

How?

- Historical data
- Statistical models that estimate the expected number of claims and claim cost
- Traditionally Chain-Ladder
- Overdispersed Poisson model as an extention of Chain Ladder

GRADIENT BOOSTING MACHINES (GBM)



- Tree-based models where observations are divided with Yes/No questions
- Combine multiple weak estimators for a stronger estimator
- Fit multiple trees with low depth Four hyperparameters:
  - Depth
  - Bagging
  - Shrinkage
  - Minimum number of observations per leaf
- The model is fitted, and the hyperparameters that minimize the prediction error are chosen

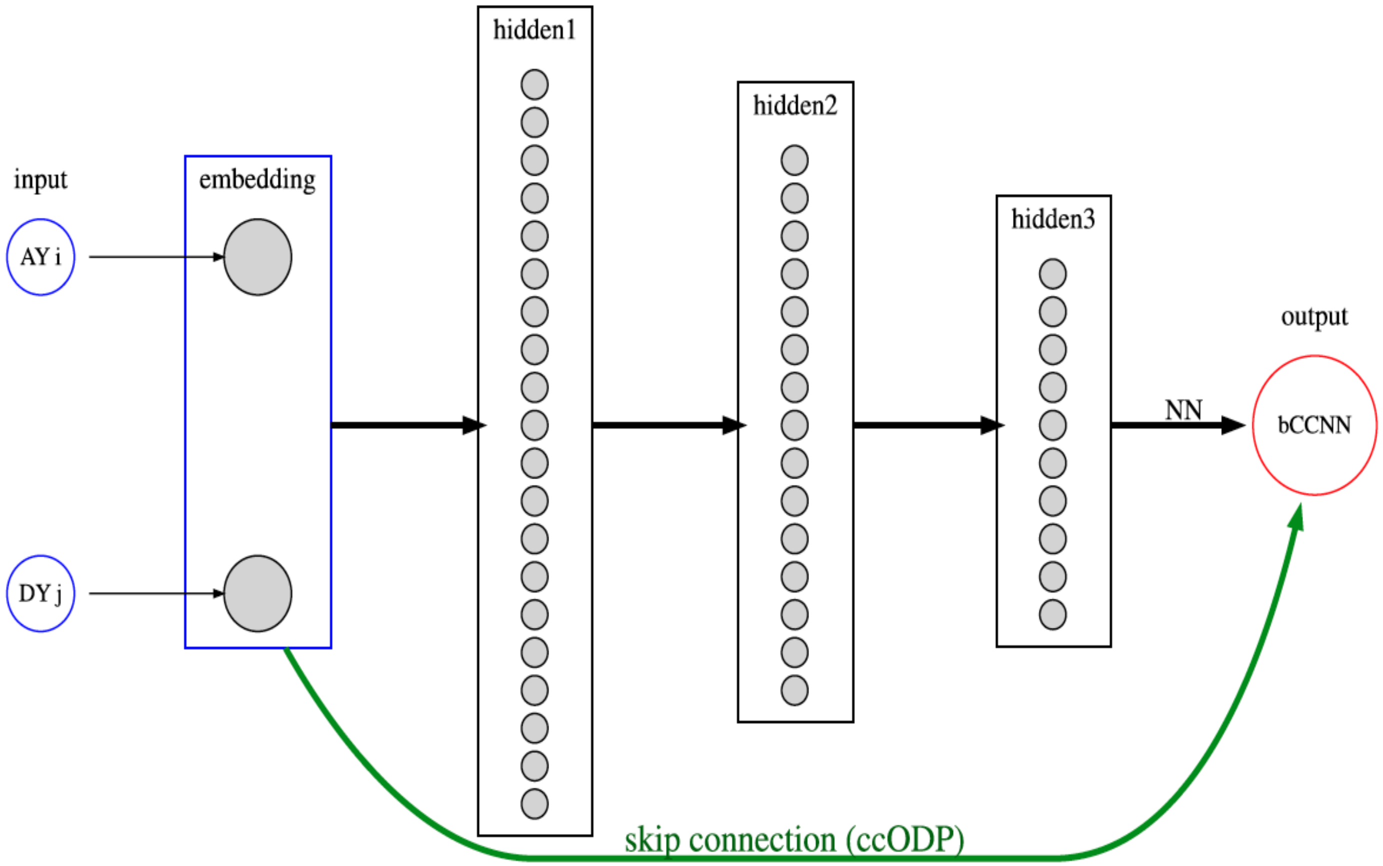
NEURAL NETWORKS MATHEMATICS

- Algorithm inspired by neuroscience

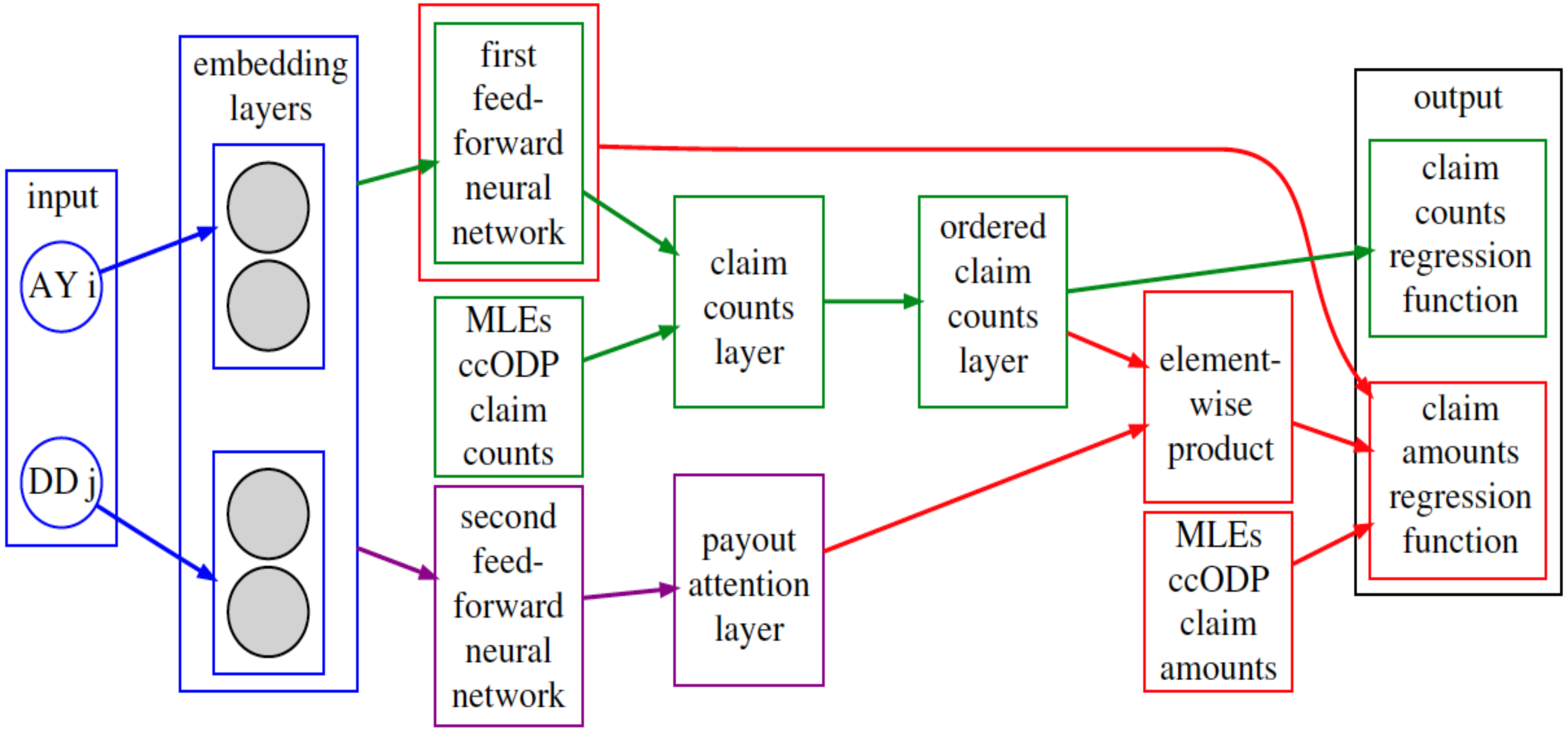
$$z^{(h)} = f^{(h)}\left(b_h + \left\langle w_h, z^{(h-1)}(\mathbf{x})\right\rangle\right),$$

where $z^{(h-1)} \in \mathbb{R}^{q_{h-1}}$, $b_h \in \mathbb{R}^{q_h}$ is the bias vector, $w_h \in \mathbb{R}^{q_{h-1} \times q_h}$ is the weight matrix and $f^{(h)}$ is some function, also known as the activation function, which is applied element-wise.

NEURAL NETWORKS (NN)



ARCHITECTURE OF A DOUBLE NEURAL NETWORK



DATA

Simulated

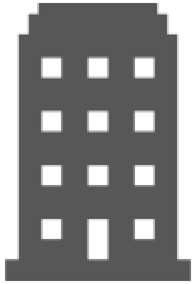
Six different products

12 claim years

Claims occurred between 1994 to 2005

Easier to predict

Från: Gabrielli, A. & Wüthrich, M. V. (2018). An Individual Claims History Simulation Machine. Risks, 6(2):29.



Folksam

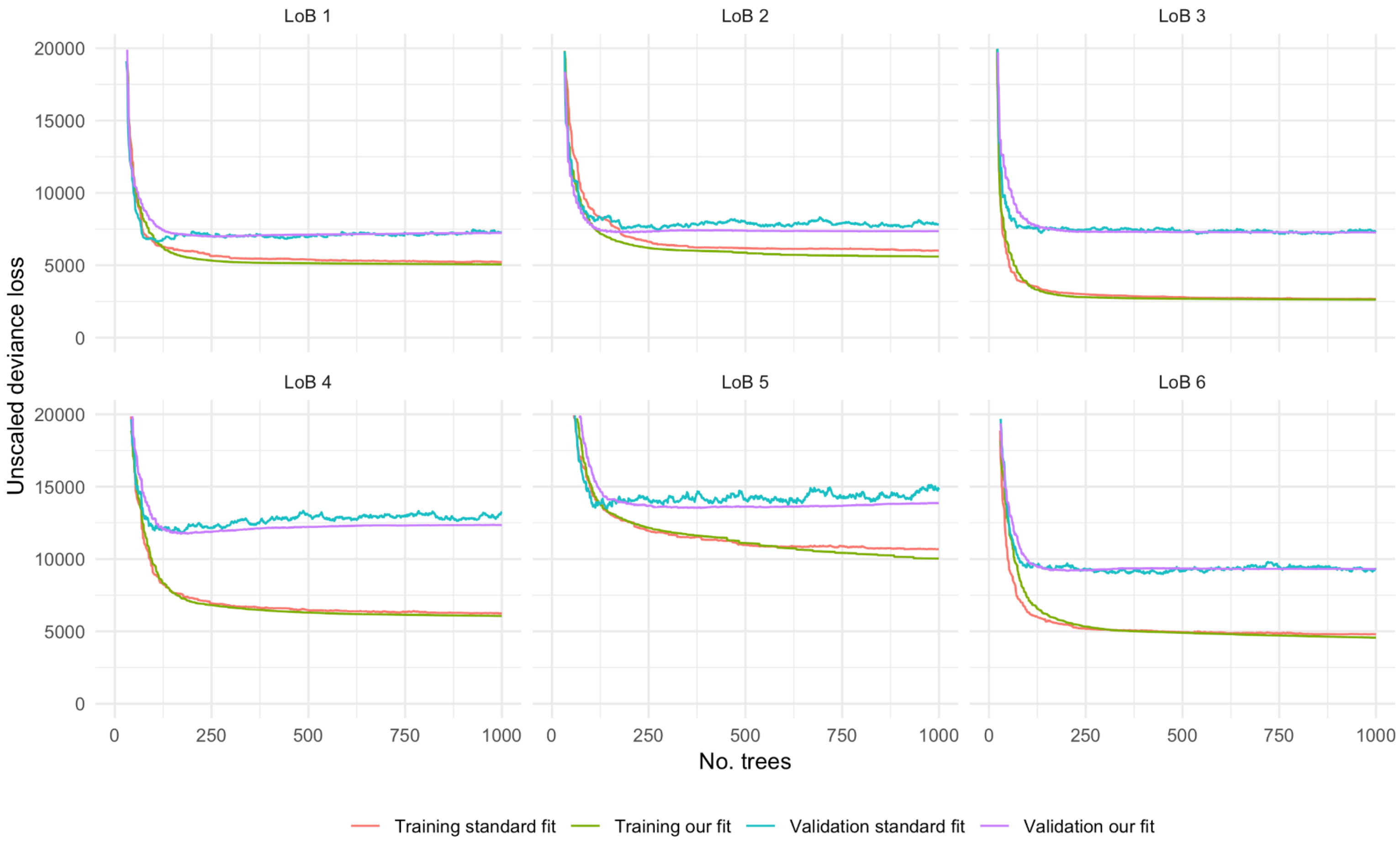
Three different products

10 to 16 claim years

Claims occurred between 1990-2007

More challenging to predict future payments

TRAINING OF THE MODELS



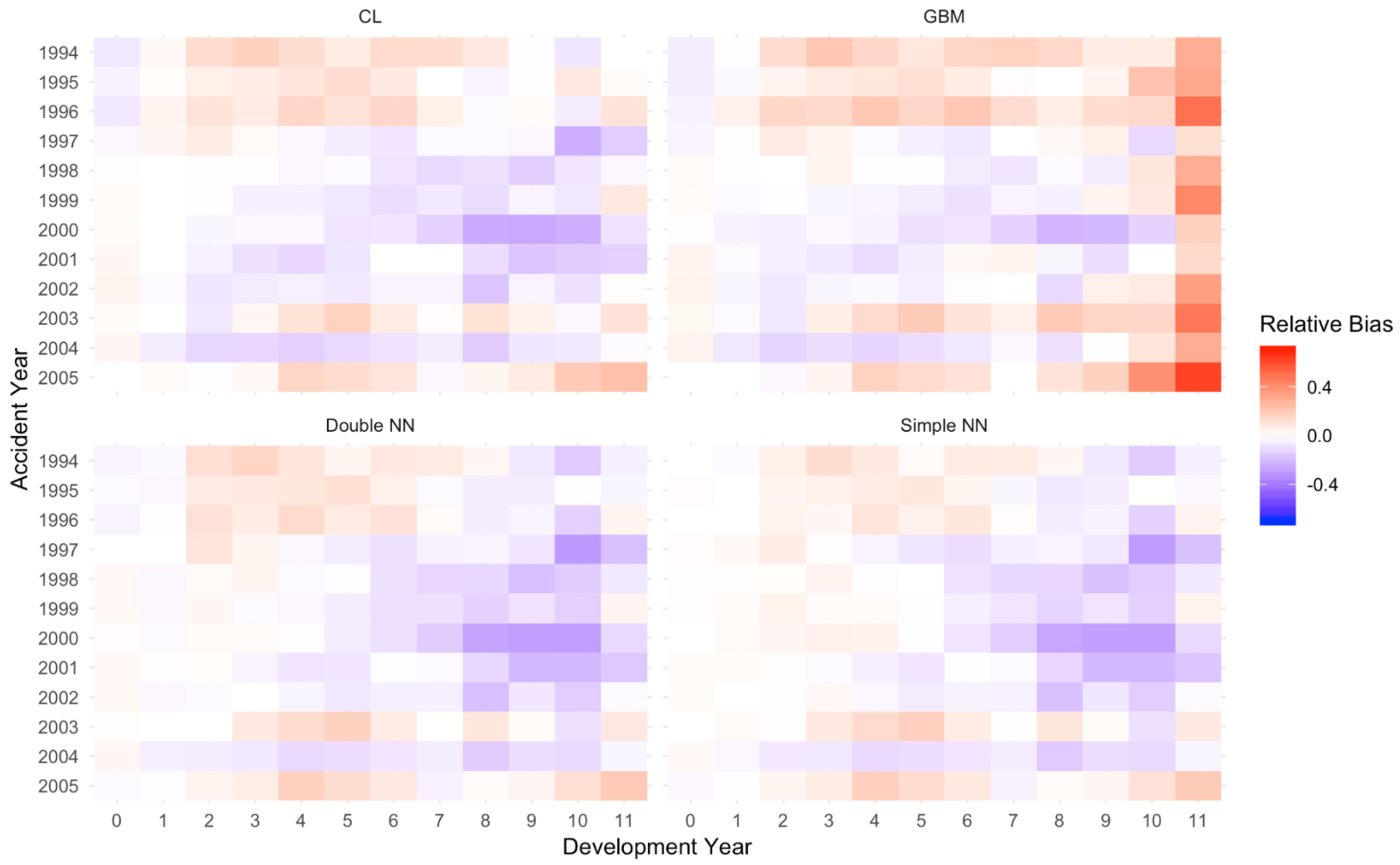
PREDICTED RESERVES SIMULATED

Model	Type	LoB 1	2	3	4	5	6
True	Reserve	39 689	37 037	16 878	71 630	72 548	31 117
CL	Reserve	38 569	35 460	15 692	67 574	70 166	29 409
CL	Bias %	-2.82	-4.26	-7.02	-5.66	-3.28	-5.49
ODP	Reserve	38 308	35 151	15 452	67 055	69 470	29 115
ODP	Bias %	-3.48	-5.10	-8.45	-6.39	-4.24	-6.44
GBM	Reserve	39 697	37 229	16 367	72 667	71 433	32 114
GBM	Bias %	**0.02**	**0.52**	**-3.03**	1.44	-1.54	**3.20**
Simple NN	Reserve	41 268	34 779	15 356	71 682	70 649	29 336
Simple NN	Bias %	3.98	-6.10	-9.02	**0.07**	-2.62	-5.73
Double NN	Reserve	40 029	35 959	15 686	69 509	72 512	30 047
Double NN	Bias %	0.85	-2.91	-7.06	-2.96	**-0.05**	-3.44

PREDICTED RESERVES FOLKSAM

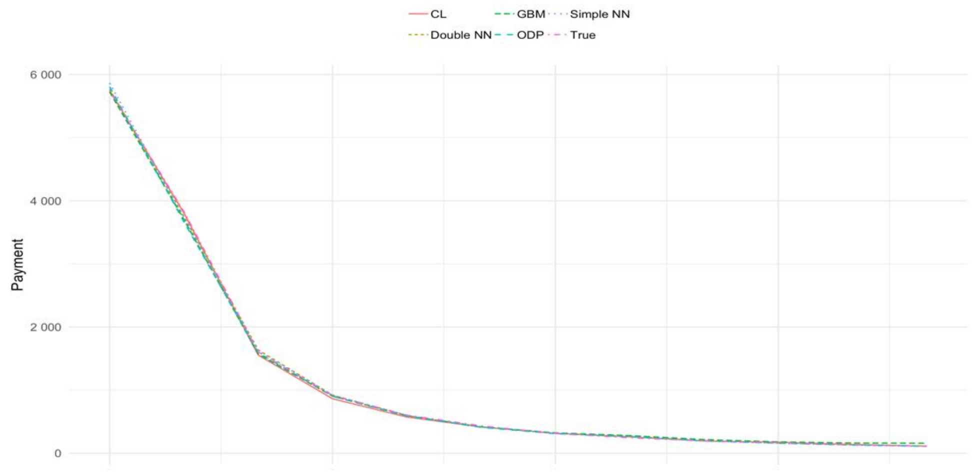
		LoB		
Model	Type	1	2	3
True	Reserve	734 200	135 241	486 714
CL	Reserve	401 572	131 799	375 972
CL	Bias %	-45.30	-2.55	-22.75
ODP	Reserve	459 873	141 439	374 627
ODP	Bias %	-37.36	4.58	-23.03
GBM	Reserve	783 878	145 219	394 719
GBM	Bias %	**6.77**	7.38	**-18.90**
Simple NN	Reserve	593 404	135 031	388 580
Simple NN	Bias %	-19.18	**-0.15**	-20.16
Double NN	Reserve	474 484	138 474	375 805
Double NN	Bias %	-35.37	2.39	-22.78

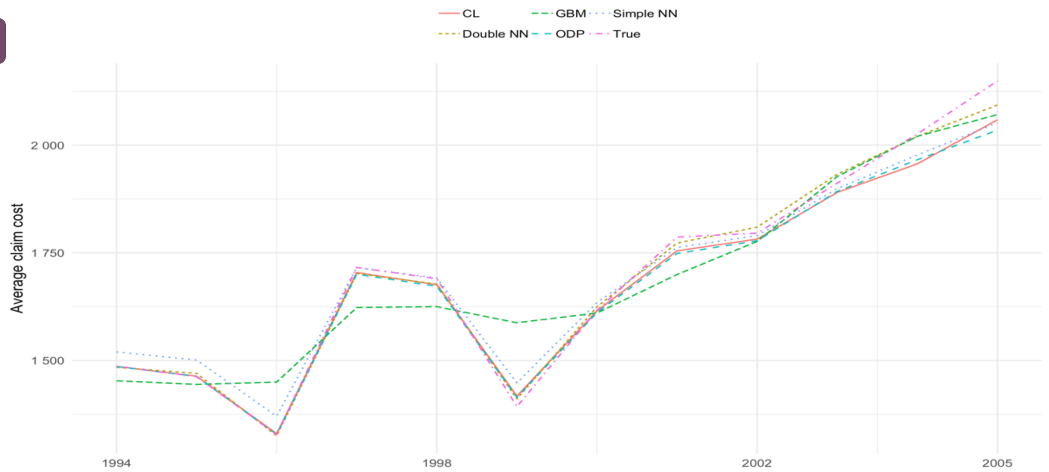
ESTIMATION ERROR WITH DIFFERENT MODELS



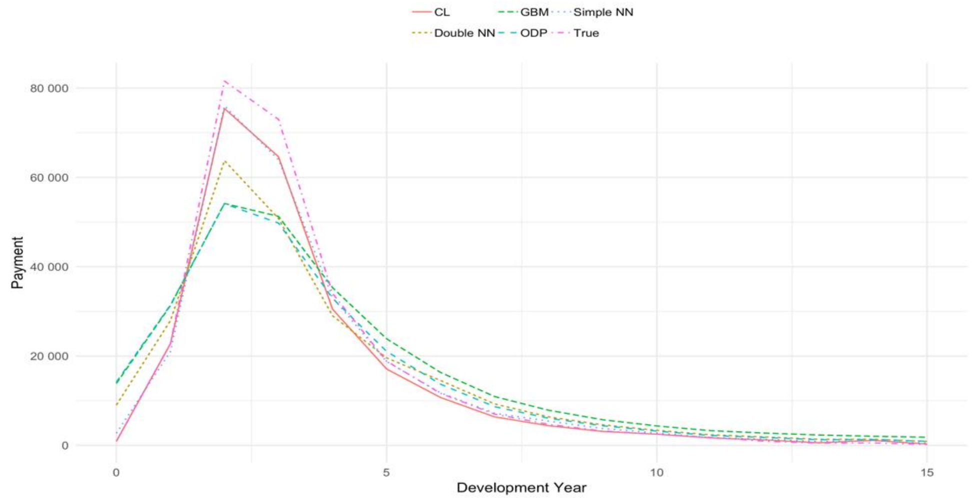
ESTIMATED CASHFLOW AND AVERAGE CLAIM COST

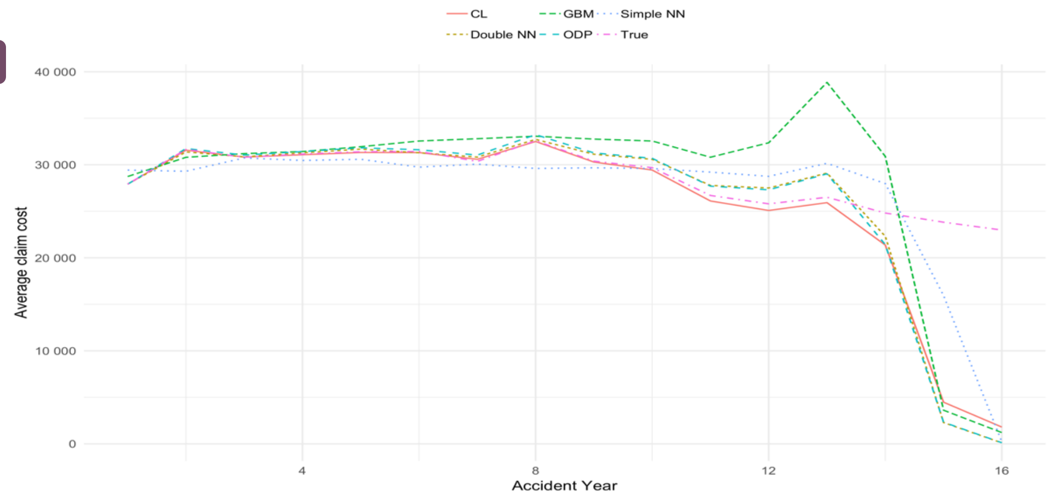
Simulated





Folksam





RMSEP

	LoB					
MSEP	1	2	3	4	5	6
CL	1 120	1 287	480	2 195	2 000	953
ODP	1 017	1 158	617	1 706	2 353	1 162
GBM	1 746	2 699	1 263	5 877	4 597	3 052
NN	1 881	1 794	750	3 216	3 709	1 305
Process Variance	1	2	3	4	5	6
CL	587	715	245	1 110	1 000	491
ODP	560	633	330	877	1 208	600
GBM	683	751	410	1 030	2 031	746
NN	442	598	327	677	1 130	598
Estimation Error	1	2	3	4	5	6
CL	954	1 069	412	1 894	1 732	817
ODP	849	969	521	1 463	2 019	995
GBM	1 607	2 593	1 195	5 786	4 124	2 959
NN	1 828	1 691	675	3 144	3 533	1 160

CONCLUSIONS

Machine learning has the potential of improving reserving

Which model is the best?

Point estimation? GBM

Least variation? CL/NN

Minimize estimation error? CL/ODP

Easy implementation? GBM

Improvements

Different divisions of training and validation data

Categorical explanatory variables

Fine-tuning of neural network

THANK YOU