Estimation of Individual Claim Liabilities

A comparison of Traditional and Machine Learning Methodologies

CASUALTY ACTUARIAL SOCIETY SPRING MEETING

May 11-13, 2020

Linda Brobeck & Marco De Virgilis

Presenters

Linda Brobeck, Director and Consulting Actuary, Pinnacle Actuarial Resources

Marco De Virgilis, Senior Actuarial Analyst, Allianz Global Corporate Specialty

Agenda

- Machine Learning Overview
- Traditional Claim Reserving
- Machine Learning in Claim Reserving
- Model Comparison

Machine Learning Overview

Machine Learning Overview

- Machine Learning
  - Supervised
    - Predictive
    - Target Variable
    - Task Driven
    - Regression, Classification
  - Unsupervised
    - Descriptive
    - No Target Variable
    - Data Driven
    - Clustering, Pattern Discovery, Dimension Reduction

Reinforcement – Algorithm Learns to React

Machine Learning Overview

Regression
- Ordinary Least Squares
- Linear, Logistic, Stepwise
- Multivariate Adaptive
- Locally Estimated Scatterplot Smoothed

Instance based
- k-Nearest Neighbor
- Learning Vector Quantization
- Self-Organizing Map
- Locally Weighted Learning

Regularization
- Ridge Regression
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Elastic Net
- Least-Angle Regression

Decision Trees
- Iterative Dichotomizer 3 (ID3)
- Chi-squared Automatic Interaction Detection (CHAID)
- Decision Stump
- Conditional Decision Trees

Bayesian
- Naïve, Gaussian Naïve, Multinomial Naïve
- Bayesian Belief Network
- Bayesian Network

Dimensionality Reduction
- Principal Component Analysis
- PC Regression
- Partial Least Squares
- Sammon Mapping
- Multidimensional Scaling
- Projection Pursuit
- Linear, Mixture, Quadratic, Flexible Discriminant Analysis

Association Rule Learning
- A priori algorithm
- Eclat algorithm

Clustering
- k-Means, k-Medians
- Expectation Maximization
- Hierarchical Clustering

Artificial Neural Network
- Perceptron
- Back-Propagation
- Hopfield Network
- Radial Basis Function Network

Deep Learning
- Deep Boltzmann Machine
- Deep Belief Networks
- Convolutional Neural Network
- Stacked Auto-Encoders

Machine Learning Pros & Cons

PROS:

- Find relationships in data
- Learning and predicting
- Sophistication
- Open-source software

CONS:

- Lack of transparency
- Over-fit
- Computational Cost

Case Study Techniques

POLLING QUESTION: Which of the following techniques that will be discussed today have you explored and/or used? (pick all that apply)

- Generalized Additive Models (GAM)
- Multivariate Regression Splines (MARS)
- K Nearest Neighbor (KNN)
- Gradient Boosting (GB)
- Artificial Neural Network (NN)

Generalized Additive Models (GAM)

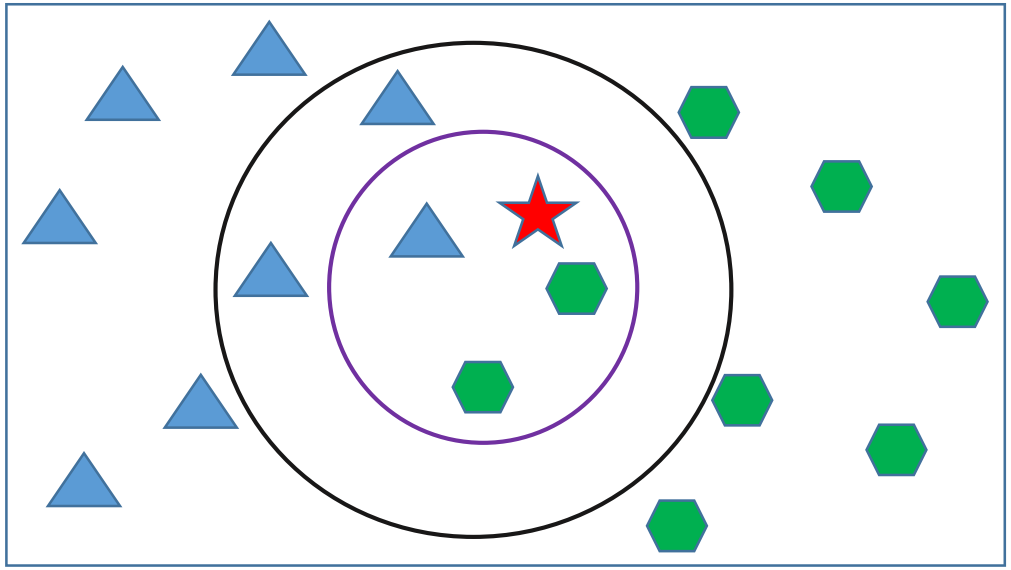
- **Replaces estimation of linear form parameters with smooth linear or non-linear functions**
  - $\eta = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p)$
  - $\mu = g^{-1}(\eta)$
- **Goodness-of-fit**
  - Conceptually equivalent to sum of squares in ordinary linear regression
- **Functions f_i can be:**
  - Parametric with a specified form (i.e., a polynomial)
  - Non-Parametric
  - Each f_i can be a different function

Multivariate Regression Splines (MARS)

- Models built in two phases – forward and backward pass
- FORWARD PASS
  - Start with just the intercept
  - Repeatedly add basis function in pairs
  - Find pair that maximizes reduction in sum-of-squares residual error
  - Add terms until change in residual error is very small
- BACKWARD PASS
  - Remove terms one by one, deleting the least effective term

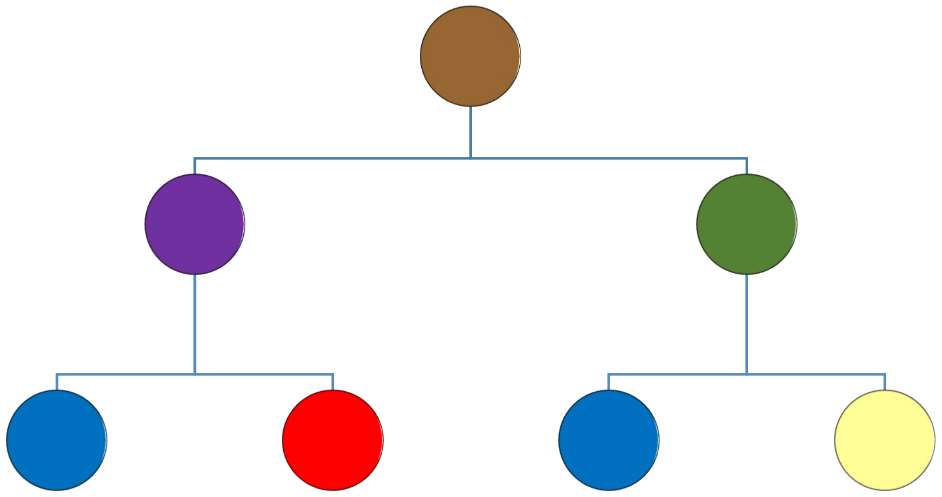
K Nearest Neighbor (KNN)

- KNN captures the idea of similarity
- Classifies a data point based on how its neighbors are classified
- How to chose K?

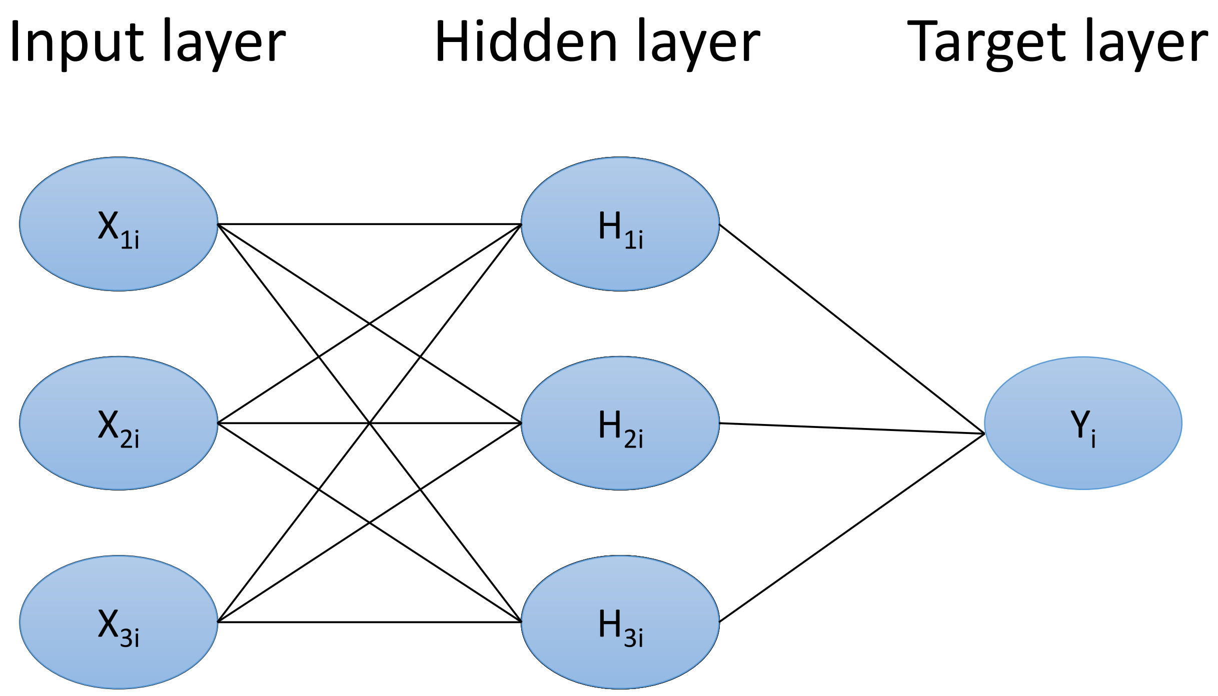


Gradient Boosting (GB)

- Decision trees built sequentially
- New tree is built on the residual of prior tree(s)



Artificial Neural Network (NN)



Traditional Claim Reserving

Run-Off Triangles: An Overview

	0	1	2	3	4	5	6	7	8	9
1981	5,012	8,269	10,907	11,805	13,539	16,181	18,009	18,608	18,662	18,834
1982	106	4,285	5,396	10,666	13,782	15,599	15,496	16,169	16,704	
1983	3,410	8,992	13,873	16,141	18,735	22,214	22,863	23,466		
1984	5,655	11,555	15,766	21,266	23,425	26,083	27,067			
1985	1,092	9,565	15,836	22,169	25,955	26,180				
1986	1,513	6,445	11,702	12,935	15,852					
1987	557	4,020	10,946	12,314						
1988	1,351	6,947	13,112							
1989	3,133	5,395								
1990	2,063									
LDF	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009	1.000

Run-Off Triangles: An Overview

AY	Latest	Ultimate	IBNR	Mack.S.E	CV(IBNR)
1981	18,834	18,834	0	0	
1982	16,704	16,858	154	143	0.928
1983	23,466	24,083	617	592	0.959
1984	27,067	28,703	1,636	713	0.436
1985	26,180	28,927	2,747	1,452	0.529
1986	15,852	19,501	3,649	1,995	0.547
1987	12,314	17,749	5,435	2,204	0.405
1988	13,112	24,019	10,907	5,354	0.491
1989	5,395	16,045	10,650	6,332	0.595
1990	2,063	18,402	16,339	24,566	1.503

Run-Off Triangles: Advantages and Disadvantages

- Advantages:
  - Easy Implementation
  - Stabilizes Experience
  - Intuitive and Interpretable

- Disadvantages:
  - Information compression, e.g. 10 years worth of data to 55 data points
  - Unreliable recent results
  - Loss of information

Machine Learning in Claim Reserving

Individual Claim Reserving

Looking at claim data at the individual level can overcome the main drawback of standard triangle techniques, bringing these advantages:

- Timely Estimates:
  - There is no need to wait for each AY year to sufficiently develop
- Extensive Use of Data:
  - Data is usually recorded anyway, it would be clever to actually use it.

Predicting Ultimate Cost

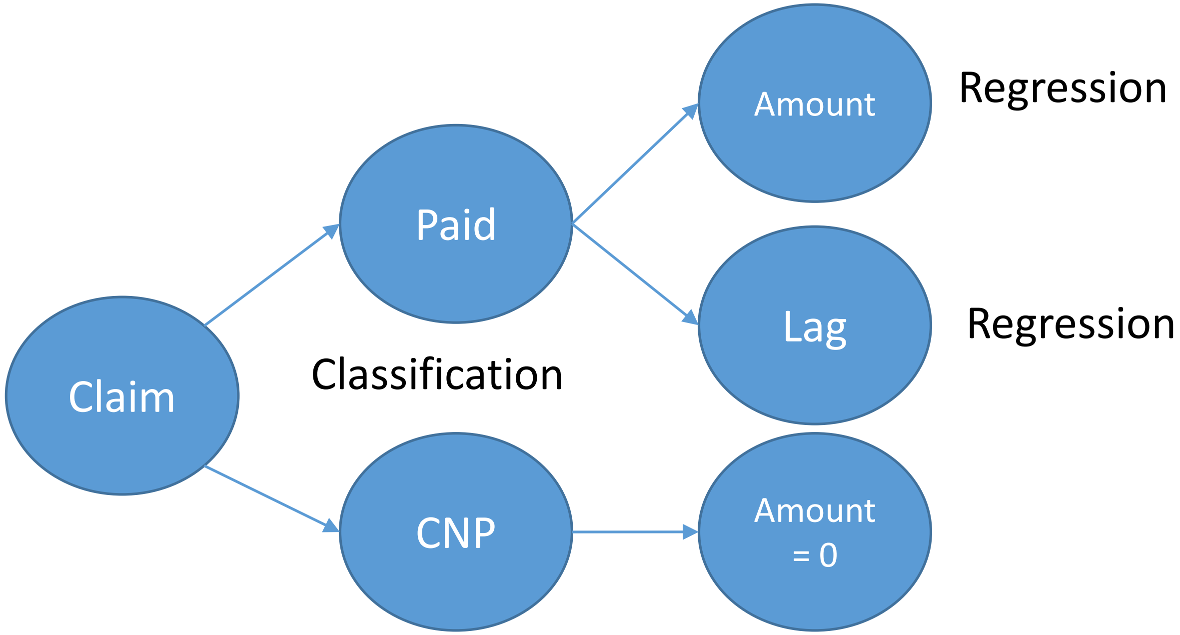
- Target: To predict the ultimate cost of the claims when they are initially reported.
- At this stage, there is no paid amount and an initial case reserve is established.
- In one case they will be settled and paid. This is the amount that needs to be estimated.
- On the other hand, claims could be closed with no payment (CNP), and therefore there will not be any payment.

Predicting Ultimate Costs

- We will have three stacked models:
  - First, each claim will be classified if it will be paid or closed with no payment.
  - If the claim will be paid, a second model will estimate the final claim amount, ie. ultimate cost.
  - A third model will also estimate the timing of such payment.

Modeling Framework

- The framework is built on one **classifier** and two **regressors.**



Data example

Acc_Date	Rep_Date	Car_Man	Lat	Long	Acc_Time	DUI	BAC	...	Case Res

Ultimate Cost	Time Lag

Model Comparison

Models Implemented

- The following models have been implemented:

  - General Additive Models, GAM
  - Multivariate Regression Splines, MARS
  - K Nearest Neighbor, KNN
  - Gradient Boosting, GB
  - Neural Networks, NN

  - Each different model has been used to classify claims and to predict both ultimate cost and lags.
  - Then, the performance have been compared and the best performing models have been chosen.

Case Study Techniques

POLLING QUESTION: Which technique performs the best in the **classification of a claim** resulting in a payment versus closed with no payment? (pick one)

- Generalized Additive Models (GAM)
- Multivariate Regression Splines (MARS)
- K Nearest Neighbor (KNN)
- Gradient Boosting (GB)
- Artificial Neural Network (NN)

RBNS Results (Reported + IBNER)

- Classification Performance:

AUC	GAM	MARS	KNN	GB	NN
Paid/CNP	0.8661	0.8748	0.8457	**0.9311**	0.9273

Case Study Techniques

POLLING QUESTION: Which technique performs the best in the estimating the **ultimate claim cost**? (pick one)

- Generalized Additive Models (GAM)
- Multivariate Regression Splines (MARS)
- K Nearest Neighbor (KNN)
- Gradient Boosting (GB)
- Artificial Neural Network (NN)

RBNS Results (Reported + IBNER)

- Regression Performance:

NRMSE	GAM	MARS	KNN	GB	NN
Claim Cost	0.1383	0.1380	0.2228	0.1337	**0.1291**

Case Study Techniques

POLLING QUESTION: Which technique performs the best in the estimating the **closing lag**?   (pick one)

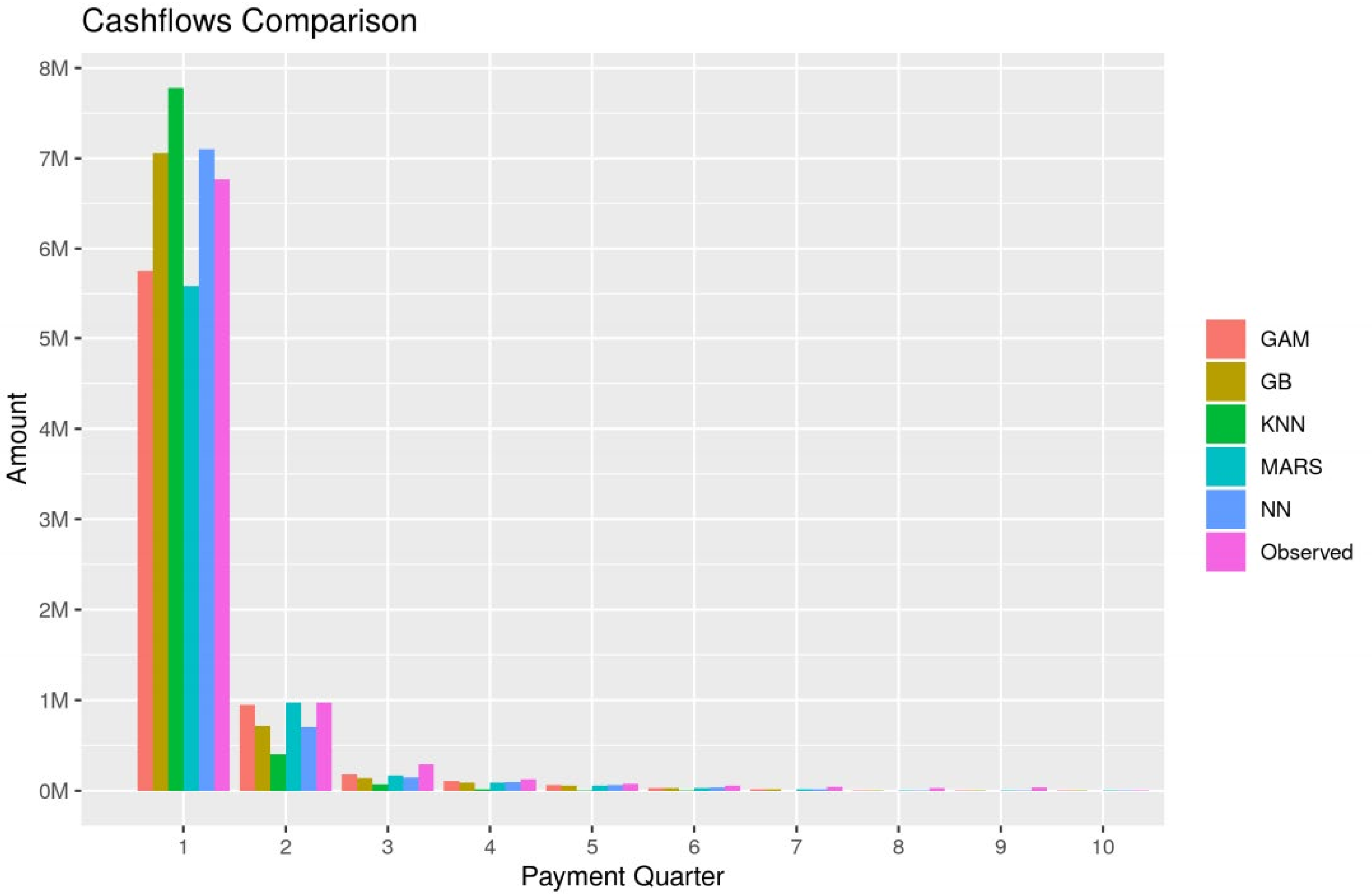
- Generalized Additive Models (GAM)
- Multivariate Regression Splines (MARS)
- K Nearest Neighbor (KNN)
- Gradient Boosting (GB)
- Artificial Neural Network (NN)

RBNS Results (Reported + IBNER)

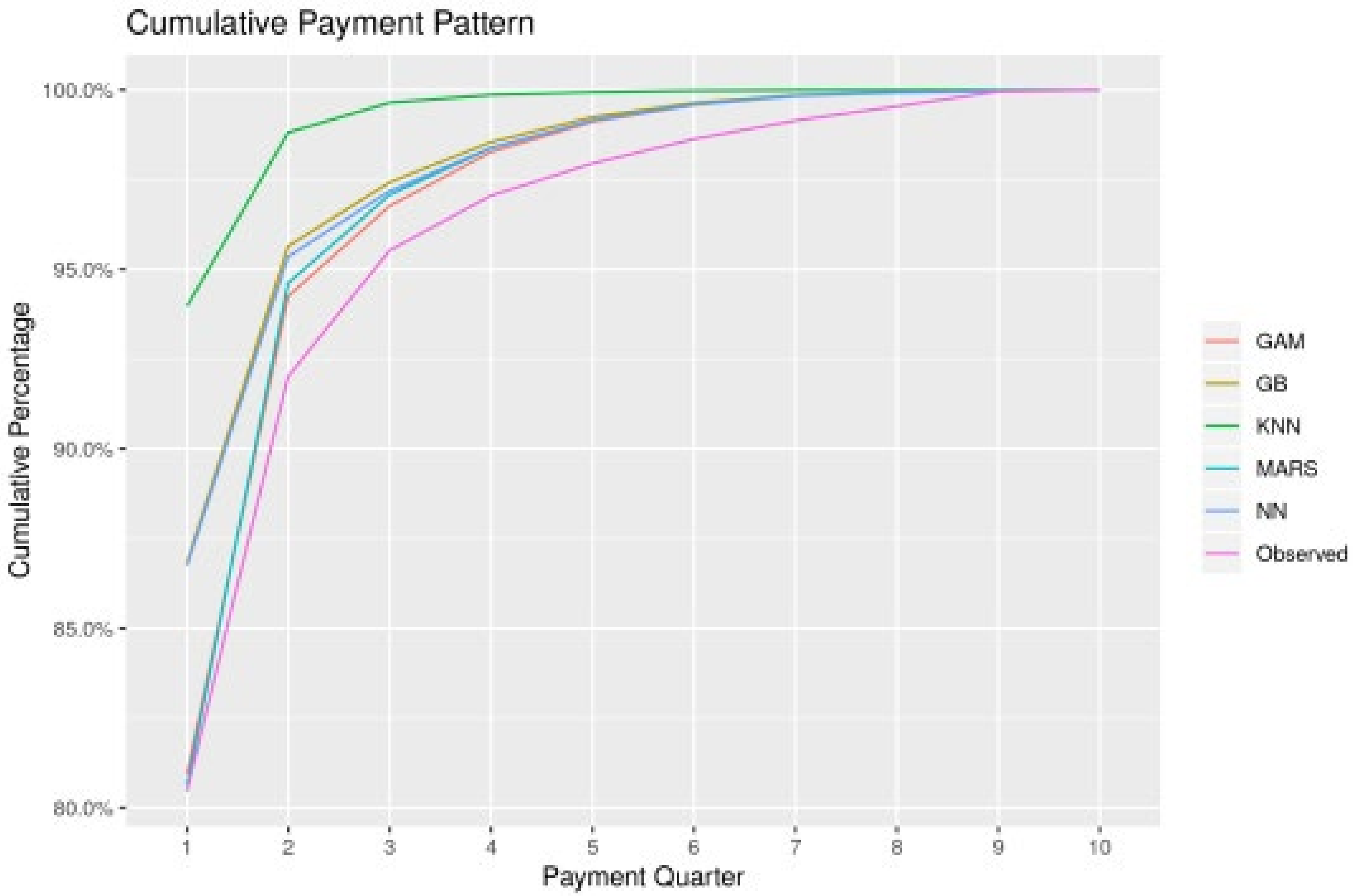
- Regression Performance:

NRMSE	GAM	MARS	KNN	GB	NN
Closing Lag	0.0988	0.1018	0.1014	**0.0793**	0.0990

RBNS Results (Reported + IBNER)



RBNS Results (Reported + IBNER)



IBNYR Estimates

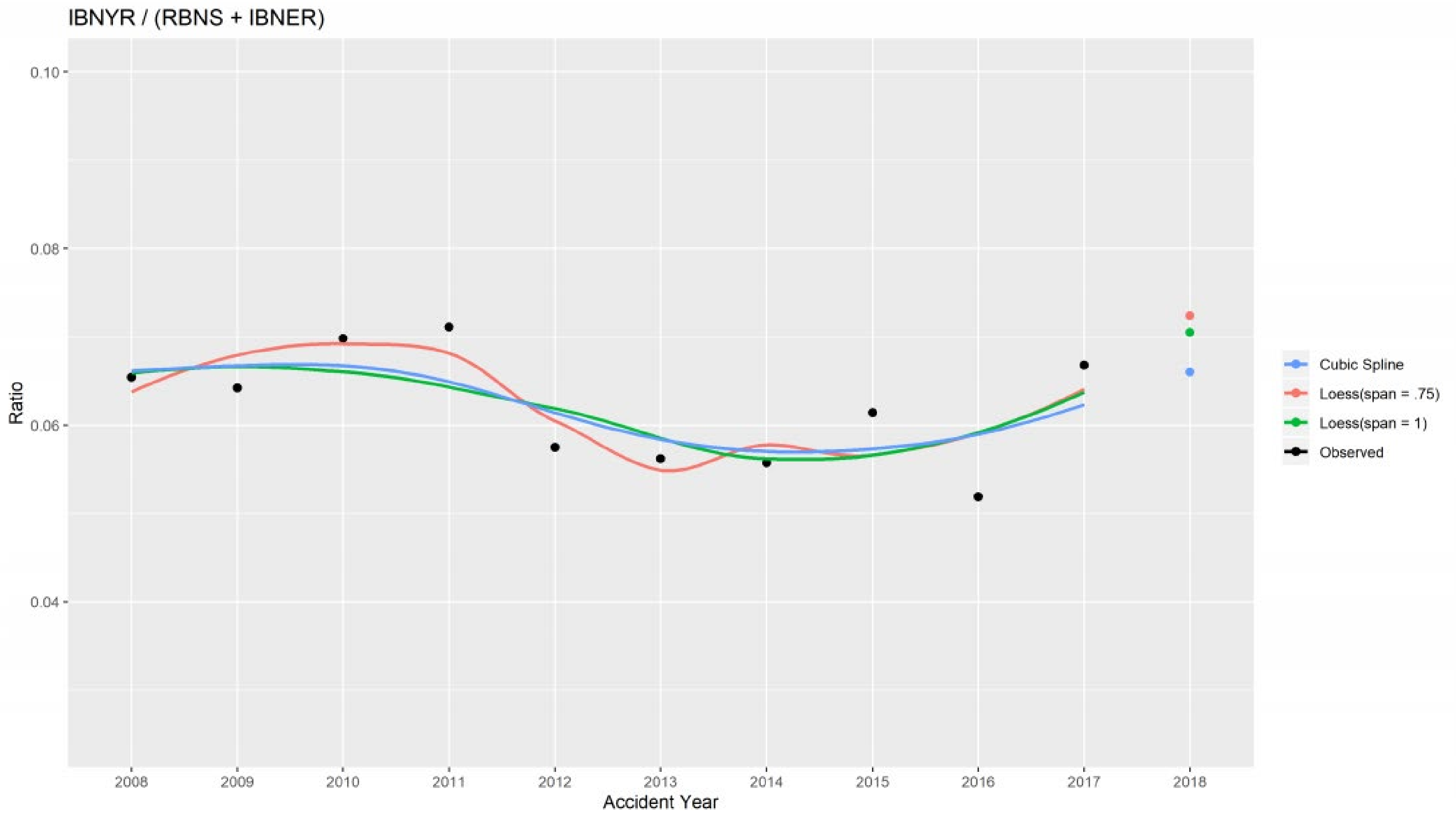
- At this point we have obtained estimates for claims that have been reported to the company.
- This includes RBNS (Reported but not settled) and IBNER (Incurred but not enough reported).
- We still have to produce estimates for IBNYR, (Incurred but not yet reported).
- Since the company does not have any records of these claims we have to follow a different approach.

IBNYR Estimates

Let's consider that the evaluation date is December 31, N.

1. Take the observed ultimate value of all the claims occurred in all the previous years and reported **by** year end (RBNS + IBNER).
2. Take the observed ultimate value of all the claims occurred in all the previous years and reported **after** year end (IBNYR).
3. Computing the ratios of these quantities, IBNYR/(RBNS + IBNER), we can have a time series for all previous AY's.
4. After estimating a value for year N, multiply this estimate by the level of ultimate amounts already predicted.
5. This will lead to an estimate of IBNYR.

IBNYR Results



2018 IBNYR	LOESS, span = .75	LOESS, span = 1	Cubic Spline
% Error	2.677%	2.646%	**2.475%**

Conclusions

- Results have a high level of accuracy.
- No reliance on individual point estimates.
- Early evaluation
  - Allows early decisions from management
- Future studies could explore the possibilities of predicting individual claim development.

Final Remarks

- We showed the potential of ML in estimation of policyholders' liabilities.
- It's not a one-fits-all recipe but it gives a framework of actions.
- Recent advances in computer power have allowed more extensive use of data in a wide variety of areas.
- We believe that it is very beneficial to explore these capabilities in the context of actuarial science.

Thank you for your attention

Further information available at

De Virgilis, M., Pierluigi C., Estimation of Individual Claim Liabilities. Casualty Actuarial Society, 2020.

https://www.casact.org/research/wp/papers/working-paper-Virgilis-Cerqueti-2020-01.pdf

Contacts:

Linda Brobeck: Lbrobeck@pinnacleactuaries.com

Marco De Virgilis: devirgilis.marco@gmail.com

Casualty Actuarial Society
4350 North Fairfax Drive, Suite 250
Arlington, Virginia 22203

www.casact.org