Machine Learning in Reserving Institute and Faculty of Actuary Working Group

September 15, 2020
Sarah MacDonnell, FIA
Jacqueline Friedland, FCIA, FCAS, FSA
Gráinne McGuire, FIAA
Nigel Carpenter, FIA
Kevin Kuo

1

Introduction to Machine Learning (ML) in Reserving Working Party

2

The ML in Reserving working party

- No easy answers
- Why bother?
- How can we help?
  - Research, facilitate, co-operate
- Aim to be a global hub
  - Broad church

3

Agenda

- Survey of reserving actuaries on ML in reserving
  - UK – Sarah MacDonnell
  - Canada – Jacque Friedland
- Foundations Workstream
  - Gráinne McGuire
- Advanced paper walkthrough
  - Nigel Carpenter
- Super brief annotated bibliography of neural net + reserving papers
  - Kevin Kuo

4

Survey of UK Actuaries on Machine Learning (ML) in Reserving

5

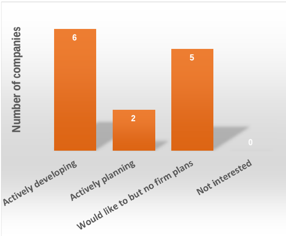
UK survey

- "Our starting premise is that whilst machine learning techniques are widespread in pricing, they are not being adopted 'on the ground' in reserving."
- UK personal lines (motor) companies only
  - 13 respondents – representative of the sector
  - Typically half hour interviews over the phone

6

ML in reserving – UK motor insurers

- No one is at the stage of embedding it as part of the reserving process
- But everyone is keen to do something



7

Why so much interest?

- Deeper understanding
  - benefit is outside of reserving and estimate itself, eg early warning system
  - not use to set IBNR, but for development patterns/case strengthening conversations
  - do more to understand; deep dives, still reliant on person

"the necessity to move away from chain ladder techniques
Is ever more clear and present"

8

Why so much interest?

- Efficiency
  - automatically highlight issues
  - faster identifying of costs
  - accuracy and speed – motivation
- Use additional data that is becoming available

9

Why are we slow to develop techniques?

- Resource/time limitations
  - BAU, projects (eg regulation), firefighting

"not good at freeing people up to explore"

- Accessibility of knowledge
  - lack of experience
  - papers too extreme
  - anchoring, willingness to move away from chain ladder

"complicated and a lot of work"

10

Why are we slow to develop techniques?

- Explainability/black box
  - explainability/might create volatility - human element smooths
  - if change case estimate philosophy, how deal with change, how feed in info?
  - comfort around changes in methodology (internal and external stakeholders)
  - need deep understanding to explain, validation

  - regulation: lack of transparency
  - regulation: requirement to document method

11

Insights

– Individual methods are a black box, and take more time, but they give more understanding than triangles.
They need massive computer power, it is complicated and is a lot of work.

– Triangular methods are simpler and lend themselves more to automation, there are some quick wins.

– The structure of the model matters and will differ between the ML method employed. Ie what you fit is more important than the method.

12

Internal relationships

- ML techniques likely to be used in pricing (and claims) in the UK
- Very little collaboration on ML
  - 4 companies said they had data science teams
    - but have limited collaboration with reserving
  - Initiative for ML research coming from reserving teams
    - reserving teams have quite a lot of autonomy
    - only one company said management was engaged with ML
  - Many global companies but virtually no apparent collaboration with overseas colleagues

13

Help from actuarial organisations?

- Almost half were not looking to the IFoA for help
- Specific requests
  - sharing methods
  - to see what other people are doing
  - ideas
  - list of papers
  - practical examples
  - possible use cases - eg improving segmentation
  - guides, practical, specific to motor
  - R skills

14

Other

- Generally happy with data available
- Companies are becoming more used to using open source software (R and Python)

15

Summary

- A great deal of interest/enthusiasm
- At an early stage - lots still to do
- Significant barriers: time/resource and difficulty to learn
- Virtually no external push factors, eg from management
- Limited internal collaboration

16

Next steps?

- Raise awareness outside of actuarial reserving
- Make learning accessible
- Make resources available
- Individual responsibility to upskill?
  Companies to support more?

What can you do?

https://www.actuaries.org.uk/practice-areas/general-insurance/research-working-parties/general-insurance-machine-learning-reserving

17

Survey of Canadian Actuaries on ML in Reserving

Jacqueline Friedland, FCIA, FCAS, FSA
Jacque.friedland@gmail.com
416-820-4741

18

Composition of Survey Respondents

- Nine insurers including
  - Canadian and global companies
  - One reinsurer
- Six consulting firms including
  - Canadian and global firms
  - Three of the big 4 accounting firms
- Three telephone interviews and twelve email responses

19

Do you currently use any ML techniques for reserving?

- Where used, only on individual claims (no ML for triangles)
- Insurers/reinsurer
  - Four replied yes, two indicated that are actively investigating/planning for later in 2020, and three said no
  - Used for:
    - Booking of reserves by one insurer
    - Insight but not booking by three insurers
    - Allocation of IBNER at policy level by one insurer
- Consultants
  - Two replied yes, and four said no
  - One indicated in use for R&D purposes not client engagements
- Types of methods used include: GLM, boost, Taylor McGuire, and operational time models

20

Use of Stochastic Methods

- Not discussed in terms of ML
- Reported use by several insurers for provision for adverse deviation (PfAD) and IFRS 17 risk adjustment
- Noted in use by three consulting firms
- From personal experience in previous industry initiatives and client assignments, know that stochastic methods are also used in Canada (by insurers/reinsurers and consulting firms) for financial condition testing (FCT), formerly dynamic capital adequacy testing (DCAT)

21

Do you have contact with other areas of the business that might be using ML techniques?

- All nine insurers noted other teams within their companies that are using ML (particularly pricing, claims, and analytics)
- Two insurers spoke of use of ML to clean data
  - One to identify data errors at transactional level
  - One to prepare data for use in ML algorithm
- Five insurers noted collaboration of reserving team with other teams
- Two insurers spoke of environment in which ML work is done by a development team outside of reserving, which is supported by reserving subject matter experts
- Did not see similar responses from consulting firms

22

Do you have plans to introduce, or develop further, ML techniques for reserving?

- All nine insurers replied yes but with different time frames
  - Yes but not immediately – four
  - Yes currently investigating with plans for later in 2020 – two
  - Yes with no further comments – two
  - Yes with much activity extending to other coverages and provinces – one
- In responding,
  - Two insurers with most advanced use are focused on Ontario personal auto
  - Two insurers noted challenges with application to commercial lines
- Only one consulting firm replied yes
- One consulting firm noted current priority focus on IFRS 17

23

What barriers have you faced in the use of ML for reserving? Insurers' Responses (1 of 2)

- Insufficient IT platforms – response: upgraded systems capabilities
- Massive change as implications to so many stakeholders internal and external to company – response: formal change management program including steering committee and buy-in of senior management
- Need for speed in work associated with reserving and financial reporting deadlines – response: adjusted design of ML model
- Challenges in communications as difficult to explain methods and differences in results between traditional and ML techniques leads to
  - Lack of acceptance of results and default to traditional methods
  - Use of ML for insight instead of booking

24

What barriers have you faced in the use of ML for reserving? Insurers' Responses (2 of 2)

- Difficult to articulate cost/benefit unless linked to resource reduction, hard to demonstrate value in reserving area
- Resource constraints
  - Always other demands that take priority (e.g., IFRS 17, COVID-19)
  - Lack of resources with familiarity in ML methods
  - Even when there are expert ML resources, there are higher priorities than reserving
- View that there are more gains to be seen in activities related to automation than in ML
- Surprising absence of comments by insurers related to data

25

What barriers have you faced in the use of ML for reserving? Responses of Consultants

- Detailed data requirements – availability of granular and consistent data
- Ongoing need for/use of emergence patterns (reporting and paid)
- Requirements to produce exhibits that support analysis
- Challenge of finding most appropriate use of ML and how it fits best within reserving governance framework, such as
  - Segmentation
  - Making selections
  - Scenario testing

26

Is there any work in this area you are aware of that might be relevant to our research?

- Baudry and Robert, 2017 and 2019
- Wüthrich, Mario V, 2018
- Duval, Pigeon, 2019
- De Virgilis, Cerqueti, 2020
- Kuo, 2019
- Poon, 2019
- UQAM, Mathieu Pigeon
- ASTIN ML and Traditional Methods Synergy in Non-Life Reserving
- All recommendations provided by insurers, none offered by consultants
- Two insurers mentioned partnerships with universities for ML work

27

What would you like to see to help develop your knowledge or use of ML?

- Similar comments from insurers, reinsurer and consultants
- Use cases and examples of successful real world applications, including
  - Code
  - Advantages and disadvantages
  - Approaches used and challenges faced
  - Proper attention to shortcomings and difficulties to overcome ("avoid the sales pitch")
- Discussions of interpretability of results
- Highlighting important variables that significantly influence results, especially for individual claims reserving methods
- ML methods for late reported claims (pure IBNR)
- Roadmap for how to move from (a) not using ML to (b) using some input from ML to (c) full implementation of ML
- Focus on the practical (much theory available)
- Tutorials, simple examples that outline the steps of a ML algorithm

28

How happy are you with the data you have available that would allow you to apply ML techniques?

- Answers differ by insurers, reinsurer, and consultants
- Surprising number of insurers generally satisfied with data available
- Insurers split on satisfaction with data
  - High quality, rich data available, particularly from newer claims platforms (e.g., Guidewire)
  - Legacy systems and systems from acquisitions can present issues
  - Greater challenges cited with data for commercial lines vs. personal auto
- Reinsurer further removed from source data but nevertheless progressing on ML
- Consultants responded more often about limited volume and quality of data
  - Three issues related to data:
    - Quality of rich data, consistency of data over time, ability to access data quickly and cost efficiently
    - Few insurers have all three
  - Commented on cost-sensitivity of client engagements

29

Focus on Data Quality – Insurers' Comments

- Data availability is just one step in data journey
- Many insurers spoke of quality with a focus on data entry of Claims function
- Quality is key from point of entry
- One respondent stated that insurers need to "implement a data driven culture where accuracy of entered data primes above productivity metrics"

30

What is your organization's attitude to using open source software such as R and Python?

- Most insurers and consultants open to use of open source software
- Several insurers noted issues of security and governance (including need for vetting by IT) that must be addressed for use of open source software
- Some insurers have established dedicated environment
  - Operations with open source code separated from other company operations
  - For one insurer, data must be anonymized for use in this separated space
- Only one consultant noted preference for commercial software that offers dedicated training and support

31

Foundations Workstream

Gráinne McGuire, FIAA
grainne.mcguire@taylorfry.com.au

32

Why?

33

How?

34

What? – Articles on ML basics

- Introductory topics
  - Getting started
  - Data manipulation
  - Graphing tools
  - Simulating data
- Methods
  - GLMs
  - Regularised regression / LASSO
  - Tree based methods (decision trees, random forests, gradient boosting)
- Links to learning resources

35

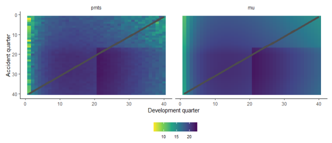
What? – worked example

- Data
  - Simulated 40x40 traditional triangles
  - Examples run quickly – but will not display full potential of ML techniques
- Modelling
  - **Focus on demonstrating use of techniques rather than search for best model**
- Performance
  - Some visual tools, measurement
  - Simulated data -> future outcomes available

36

Data

- Data have
  - AQ effects
  - DQ effects
  - PQ effects
  - AQ*DQ interaction
  - Described at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3241906



37

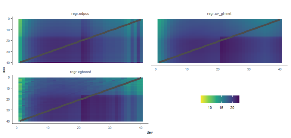
Models used

- 5 models (3 shown here)
  - **Volume all chain ladder**
    - **regr.odpcc**
  - **LASSO**
    - **regr.cv_glmnet**
  - Decision tree
  - Random forest
  - **XGBoost**
    - **regr.xgboost**
- Model fitting
  - Train, validate and test data sets
  - Hyper-parameters
  - Cross-validation
  - Code provided

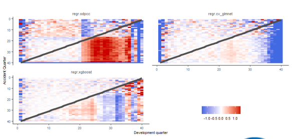
38

Model fit

Fitted values (logged)



Actual/expected



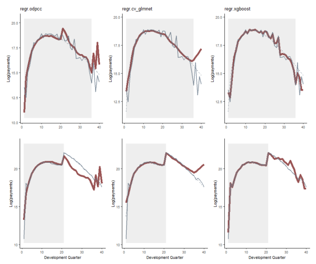
Values bounded by (25%, 400%)

39

Model fit – tracking by development quarter

AQ = 5

Prior to interaction

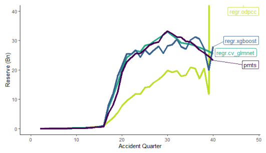


AQ = 20

After interaction

40

Reserves



Model	Reserve	Ratio to true	RMSE
True	608		
Regr.odpcc (CL)	563	93%	1.7E9
Regr.cv_glmnet	627	103%	2.5E8
Regr.xgboost	607	100%	3.5E8

41

Where and when?

Where?

- https://www.actuaries.org.uk/practice-areas/general-insurance/research-working-parties/general-insurance-machine-learning-reserving
- https://institute-and-faculty-of-actuaries.github.io/mlr-blog/

When?

- Blog updated periodically with articles and worked examples

42

Advanced paper walkthrough

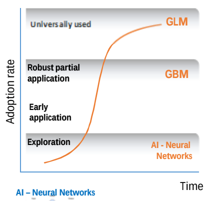
Nigel Carpenter, FIA

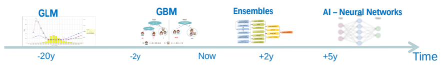
43

Context: Machine Learning in Pricing

Analytic Innovation: The imperative for accurate pricing drives the development and adoption of new analytic techniques.

- GLMs used for 20 years; now universal
- GBMs starting to be adopted
- Ensembles becoming possible
- Neural Networks some way off

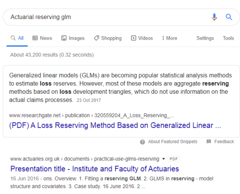




44

Reserving as a GLM

- Plenty of actuarial reserving research to show that the Chain-ladder can be formulated as a GLM.
- From GI Pricing we know that Machine Learning (GBMs and Neural Networks) outperforms GLMs.
- So where are all the Machine Learning in reserving papers?



45

Reserving & Machine Learning

Date	Title	Author	Rating	Comment
2016_09	Machine Learning Framework for Loss Reserving	KPMG	P	GBMs with aggregated data old approach to tuning and validation
2017_03	Machine Learning in Individual Claims Reserving	WUTHRICH	P	Individual claim transactions with decision trees but no IBNR
2017	Individual claim Development with Machine Learning	ASTIN	P	Old school Neural Networks on claim transactions
2017_12	Non parametric individual claim reserving in insurance	BAUDRY	PP	ML plus external data and IBNR, no code!
2018_05	Deep Triangle	KUO	PP	RNNs and code shared but complex!

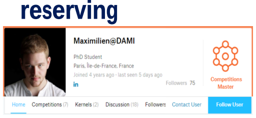
Growing number of good papers available up to 2018

Even more during 2019

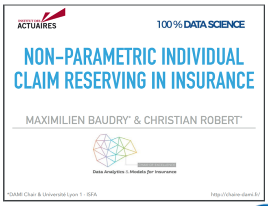
But awareness and accessibility can be difficult especially if you are new to Data Science.

46

BAUDRY: Non Parametric individual claim reserving



- Kaggle Master and PhD Student @ DAMI Paris.
- Expert knowledge in Machine Learning and Natural Language
- Supervisor Prof Christian Y Robert, provides Actuarial background.



47

BAUDRY: Non Parametric individual claim reserving

Baudry's approach uses extra info beyond traditional "triangle" style claims data.

$T_{0,p}$	Underwriting date
$t_i - T_{0,p}$	Exposure to reserve date
$F_{t_i,p}$	Policy Risk factors
$E_{T_{0,p}}$	External info at UW date
$E_{T_{1,p}}$	External info at Occurrence date
$E_{T_{2,p}}$	External info at Report date
$I_{t_i,p}$	Claim history up to valuation date

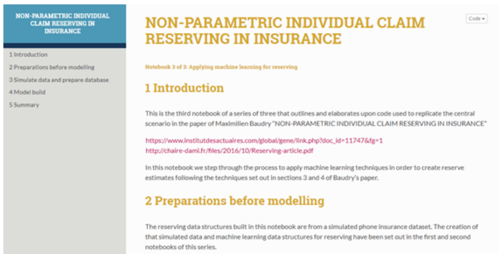
- Explicit use of this extra data, provides opportunities...
  - for the method to give improved results
  - to aid better understanding of influences on claim development

RBNS uses $(T_{0,p}, t_i - T_{0,p}, F_{t_i,p}, E_{T_{0,p}}, E_{T_{1,p}}, E_{T_{2,p}}, I_{t_i,p})$

IBNR uses $(T_{0,p}, t_i - T_{0,p}, F_{t_i,p}, E_{T_{0,p}})$

48

BAUDRY: Notebook walkthrough


NON-PARAMETRIC INDIVIDUAL CLAIM RESERVING IN INSURANCE

1 Introduction
2 Preparations before modelling
3 Simulate data and prepare database
4 Model build
5 Summary

Code

NON-PARAMETRIC INDIVIDUAL CLAIM RESERVING IN INSURANCE

Notebook 3 of 3: Applying machine learning for reserving

1 Introduction

This is the third notebook of a series of three that outlines and elaborates upon code used to replicate the central scenario in the paper of Maximilien Baudry "NON-PARAMETRIC INDIVIDUAL CLAIM RESERVING IN INSURANCE"

https://www.institutdesactuaires.com/global/gene/link.php?doc_id=11747&fg=1
http://chaire-dami.fr/files/2016/10/Reserving-article.pdf

In this notebook we step through the process to apply machine learning techniques in order to create reserve estimates following the techniques set out in sections 3 and 4 of Baudry's paper.

2 Preparations before modelling

The reserving data structures built in this notebook are from a simulated phone insurance dataset. The creation of that simulated data and machine learning data structures for reserving have been set out in the first and second notebooks of this series.

49

Super brief annotated bibliography of neural net + reserving papers

Kevin Kuo

50

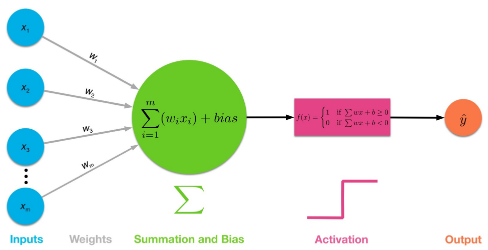
A few points before we dive in

- One does not simply jump from picking LDFs in Excel to fully automated individuals claims reserving systems – consider investing in robust/automated data/reporting pipelines first
- IMO none of the published methodologies will be deployed into production as-is, though concepts introduced will be incorporated
- But don't underestimate how quickly technology moves, even in the insurance industry
- ML on claims has applications beyond reserving
- There are approaches other than deep learning

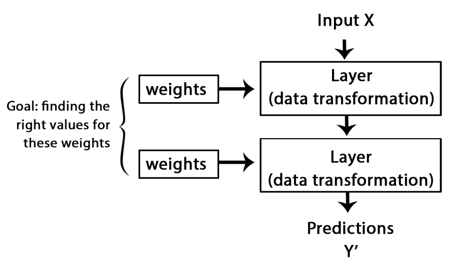
51

Crash course on neural nets

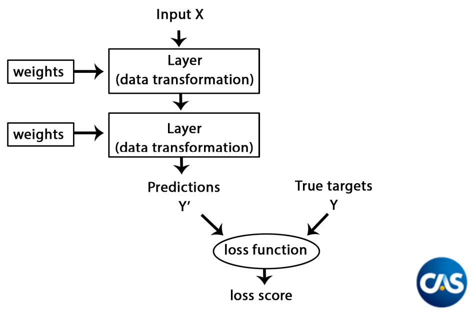
52



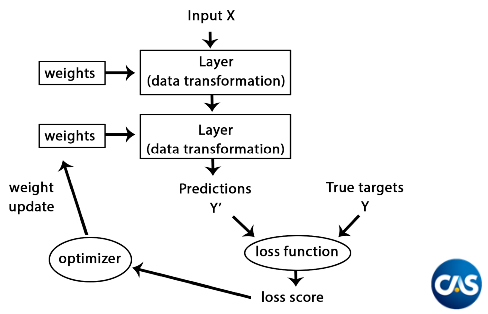
53



54



55



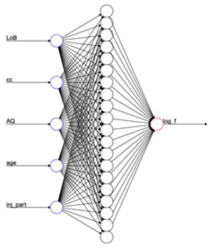
56

Papers

- 2017
  - Wüthrich, M. V. Neural Networks Applied to Chain-Ladder Reserving
- 2018
  - Kuo, K. DeepTriangle: A deep learning approach to loss reserving.
  - Gabrielli, A., R. Richman, and M. V. Wüthrich. Neural Network Embedding of the Over-Dispersed Poisson Reserving Model.
- 2019
  - Gabrielli, A. A neural network boosted double overdispersed Poisson claims reserving model.
- 2020
  - Kuo, K. Individual Claims Forecasting with Bayesian Mixture Density Networks.
  - Delong, L., M. Lindholm, and M. V. Wüthrich. Collective Reserving using Individual Claims Data.
  - Gabrielli, A. An individual Claims Reserving Model for Reported Claims.

57

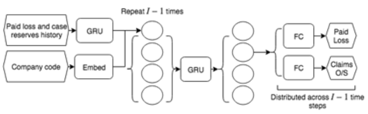
Neural Networks Applied to Chain-Ladder Reserving (Wüthrich 2017)



- Basically, predict age-to-age factors with a neural net

58

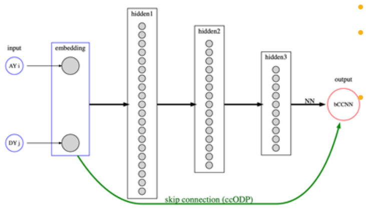
DeepTriangle (Kuo 2018)



- Applied to triangle data
- Trained on multiple companies simultaneously
- Sequence-to-sequence recurrent architecture

59

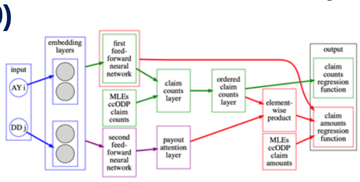
Neural net boosted ODP (Gabrielli et al. 2018)



- Applied to triangle data
- Learns residuals of cross-classified ODP applied to paid claims
- Single triangle or multiple LOBs simultaneously

60

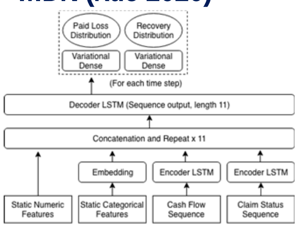
Neural net boosted double ODP (Gabrielli 2019)



- Also applied to triangle data
- Similar to previous work, but now learns paid amounts + claim counts simultaneously for better accuracy

61

Individual claims forecasting w/ Bayesian MDN (Kuo 2020)



- Applied to individual claims data
- Models distributions of paid loss + recovery for future time steps

62

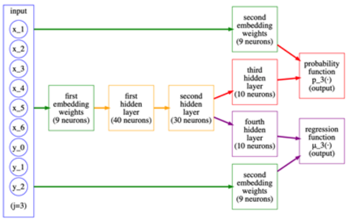
Collective reserving using individual claims data (Delong et al. 2020)

- No architecture figure in paper :(
- Basically a bunch of neural nets for modeling various processes in the life of a claim

modules serving a specific modeling purpose. The 1st network is introduced in Section 3, this network serves at predicting incurred but not reported (IBNR) claim counts. In Section 4 we introduce the 2nd network that serves at modeling the two dimensional process of payment dates and closings of reported but not settled (RBNS) claims. The 3rd and 4th networks in Section 5 model claim payments as well as recovery payments of RBNS claims: the 3rd network indicates whether we have a positive payment or a negative recovery payment, and the 4th network determines its size. In Section 6 we introduce the last two networks that serve at modeling amounts of IBNR claims: these claim amounts are modeled by a compound distribution with the 5th network indicating whether we have a zero IBNR claim or not, and the 6th network determining the amount of a non-zero IBNR claim. Our modeling is done on two different time

63

Individual claims reserving model for reported claims (Gabrielli 2020)



- Models expected probability of payment and payment amount for each time period

64

Papers

- 2017
  - Wüthrich, M. V. Neural Networks Applied to Chain-Ladder Reserving
- 2018
  - Kuo, K. DeepTriangle: A deep learning approach to loss reserving.
  - Gabrielli, A., R. Richman, and M. V. Wüthrich. Neural Network Embedding of the Over-Dispersed Poisson Reserving Model.
- 2019
  - Gabrielli, A. A neural network boosted double overdispersed Poisson claims reserving model.
- 2020
  - Kuo, K. Individual Claims Forecasting with Bayesian Mixture Density Networks.
  - Delong, L., M. Lindholm, and M. V. Wüthrich. Collective Reserving using Individual Claims Data.
  - Gabrielli, A. An individual Claims Reserving Model for Reported Claims.

65

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

www.casact.org

66