DISTRIBUTIONAL FORECASTING OF OUTSTANDING LIABILITIES WITH NEURAL NETWORKS

Muhammed Al-Mudafer (University of New South Wales, Australia)

In collaboration with:

Benjamin Avanzi (University of Melbourne, Australia)

Greg Taylor (University of New South Wales, Australia)

Bernard Wong (University of New South Wales, Australia)

CONTENTS

Paper: Stochastic loss reserving with mixed density neural networks

Code: Available on GitHub: **agi-lab/reserving-MDN-ResMDN**

INTRODUCTION

MACHINE LEARNING IN THE ACTUARIAL FIELD

- Machine learning has seen a rapid increase in application in the actuarial field, in areas such as mortality modelling, pricing and loss reserving
- Neural networks (NNs) have shown great potential thus far
- **We focus our work on NEURAL NETWORKS (NNs) and apply them to RESERVING with LOSS TRIANGLES**

NEURAL NETWORKS IN RESERVING: **POTENTIAL**

- Highly flexible, can capture complex trends in the data. Have outperformed the Chain Ladder in early applications
- Versatile modelling: Can learn from visual data, categorical data, time data, etc.
- Granular data: Neural networks thrive with large, granular data. We also demonstrate that it performs well with aggregate data

NEURAL NETWORKS IN RESERVING: **RISKS AND CURRENT CHALLENGES**

- Most NN reserving applications focus on giving **central estimates**; less focus on the **distribution**, which is critical for risk management
- **Black box!** NNs have low interpretability. Hard to justify results to stakeholders
- Can be **unpredictable** when extrapolating, due to their flexibility
- Require a **lot of data** – risky on a loss triangle
- **Addressing these risks can help NNs realise their potential, encourage implementation in practice**

NEURAL NETWORKS IN RESERVING: **ADDRESSING SHORTCOMINGS**

WE IMPLEMENT A NEURAL NETWORK THAT:

1. Predicts the **distribution** of outstanding claims
2. Is **easy** to fit
3. Can be adapted to an Actuarial – Neural Network hybrid approach for increased **interpretability**
4. Incorporates actuarial judgement, ensuring projections are **reasonable**
5. And most importantly, provides **accurate** reserve estimates (outperforms the chain ladder benchmark)

MODEL DESIGN AND DEVEOLPMENT

DATA: LOSS TRIANGLE

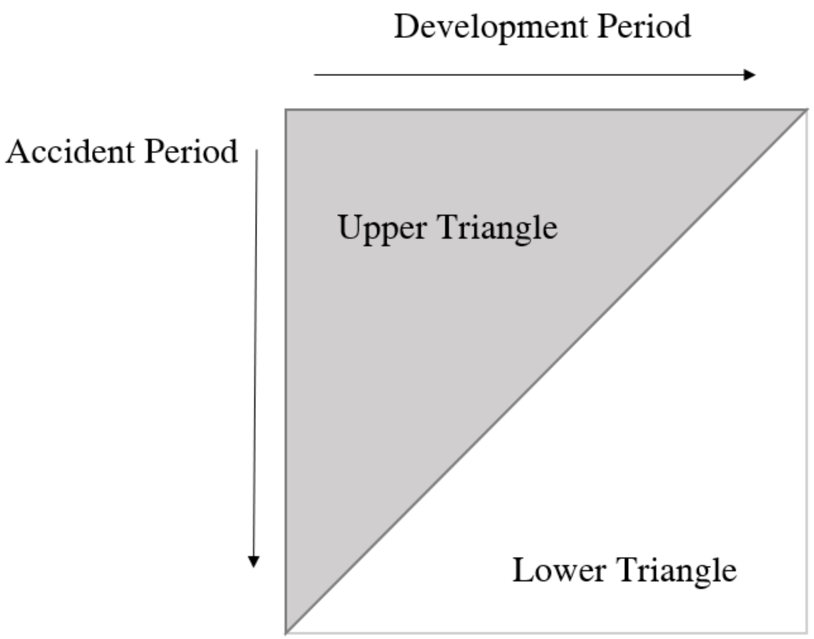


Figure 1: Aggregate loss triangle

- 40x40 loss triangles (Figure 1)
- Accident quarters (AQ): $\boldsymbol{i}$
- Development quarters (DQ): $\boldsymbol{j}$
- **Incremental Claims, $X_{i,j}$**
- **Benchmark: Stochastic Chain Ladder (CL)**

NEURAL NETWORKS

- A neural network consists of an input layer, hidden layers (with hidden nodes) and an output layer
- Nodes are connected by weights (arrows), which carry the input to the output
- Input $\boldsymbol{x}$ is passed through the network weights and hidden nodes, which transforms $\boldsymbol{x}$ into $\widehat{\boldsymbol{y}}$
- Weight parameters (arrows) are optimised to improve the fit $\widehat{\boldsymbol{y}}$

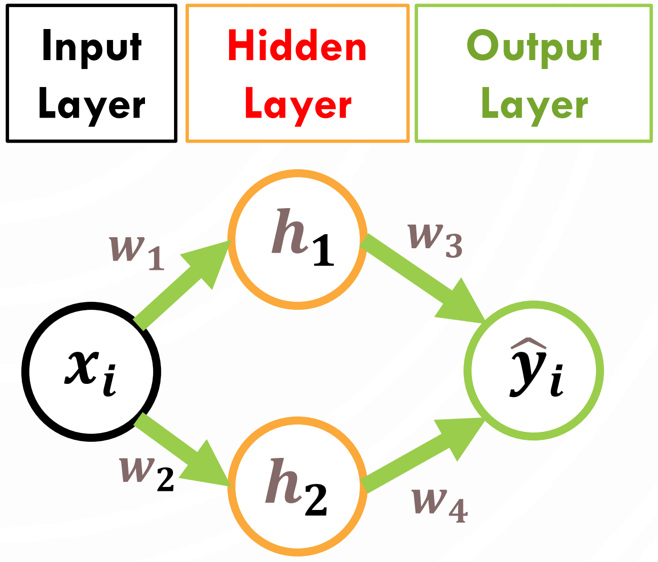


Figure 2: Example neural network

DISTRIBUTIONAL FORECASTING: MIXTURE DENSITY NETWORKS (MDN)

- Assume losses $X_{i,j}$ follow a Mixed Gaussian distribution, as such:
- $f_{\hat{X}_{i,j}}(x) = \sum_{k=1}^{K} \alpha_k \phi(x|\mu_k, \sigma_k)$
- Mixture Density Network (Figure 3) produces the (α, μ, σ) parameters as output

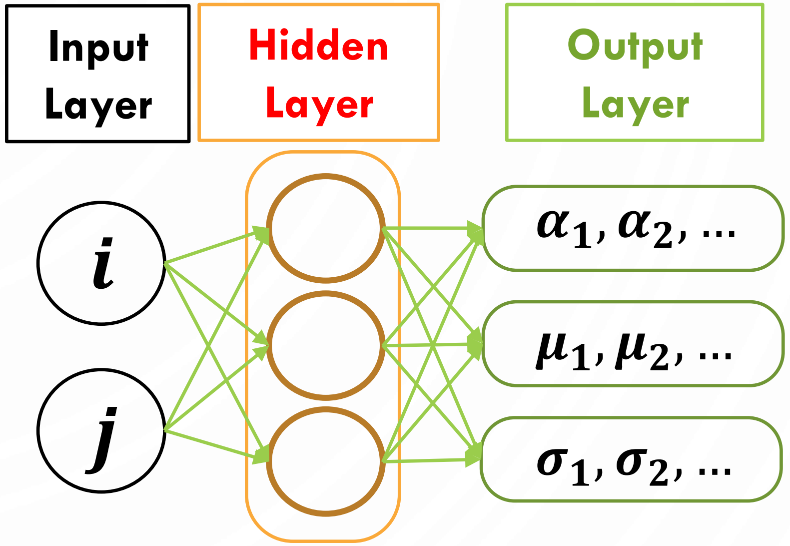


Figure 3: Mixture Density Network

MIXTURE DENSITY NETWORK (MDN): INTUITION

- Mixed Gaussian can approximate any distribution (given enough components)
- Relatively simplistic network
- Accurate central estimates

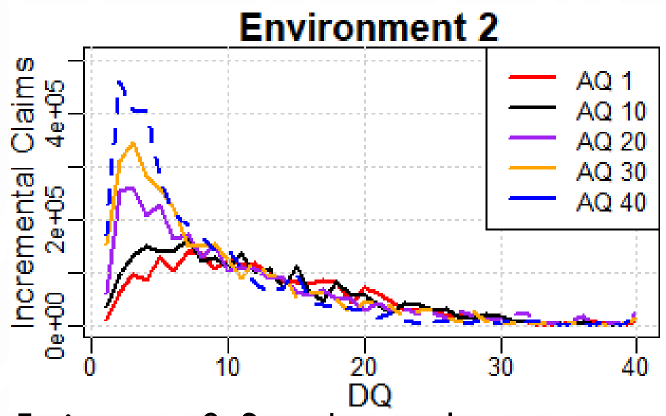
DATA SPECIFICATIONS

Simulated data, using the SynthETIC Simulator, **4 different environments:**

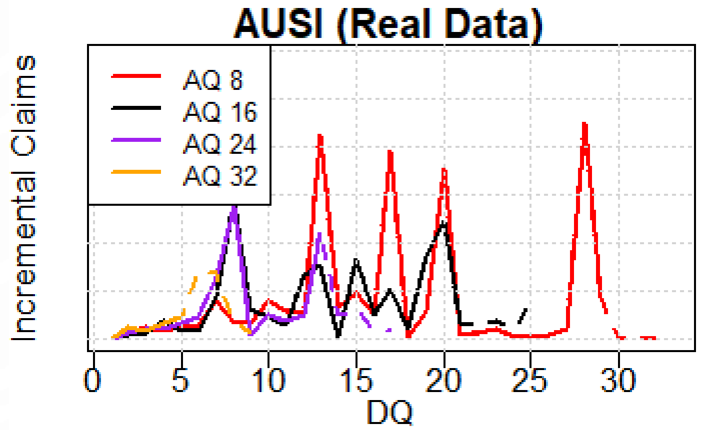
1. Simple, short tail claims
2. Gradually speeding up claim settlement
3. Inflation shock
4. Complex, volatile long tail claims

50 triangles for each simulated environment

- **Real data (AUSI)**, Auto Bodily Injury, long tail claims, 10 triangles
- **MDN run on 210 triangles**



Environment 2: Speed up settlement

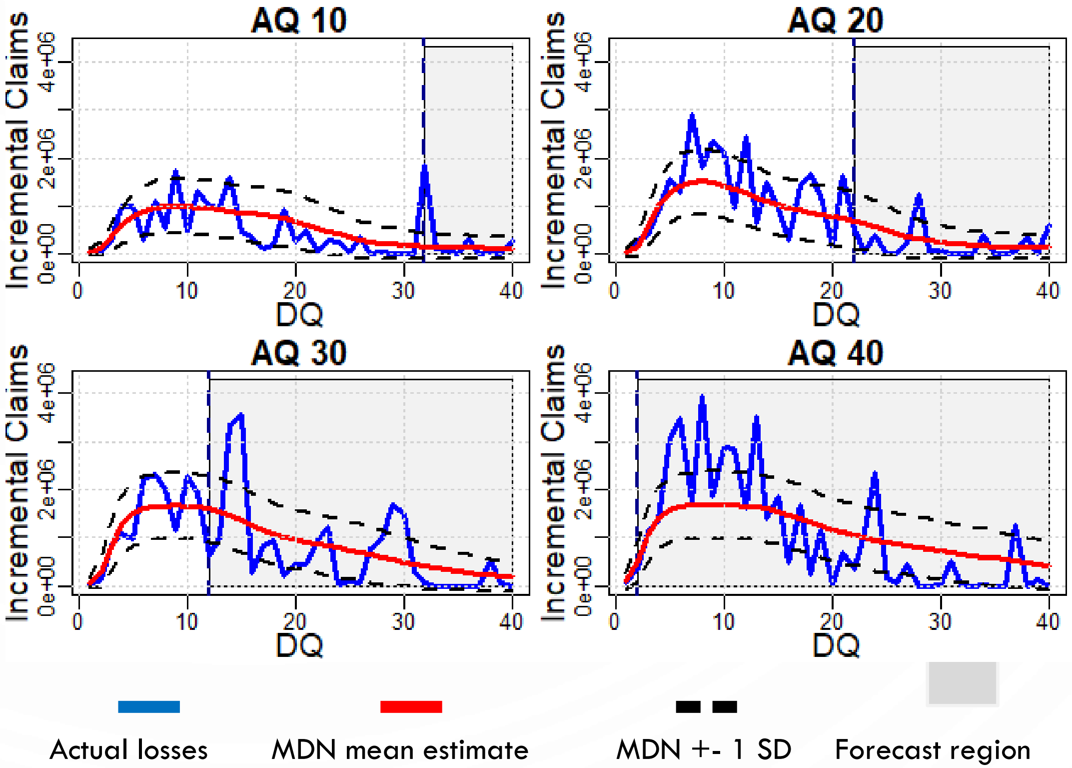


AUSI: Auto Bodily Injury, long tail claims

RESULTS

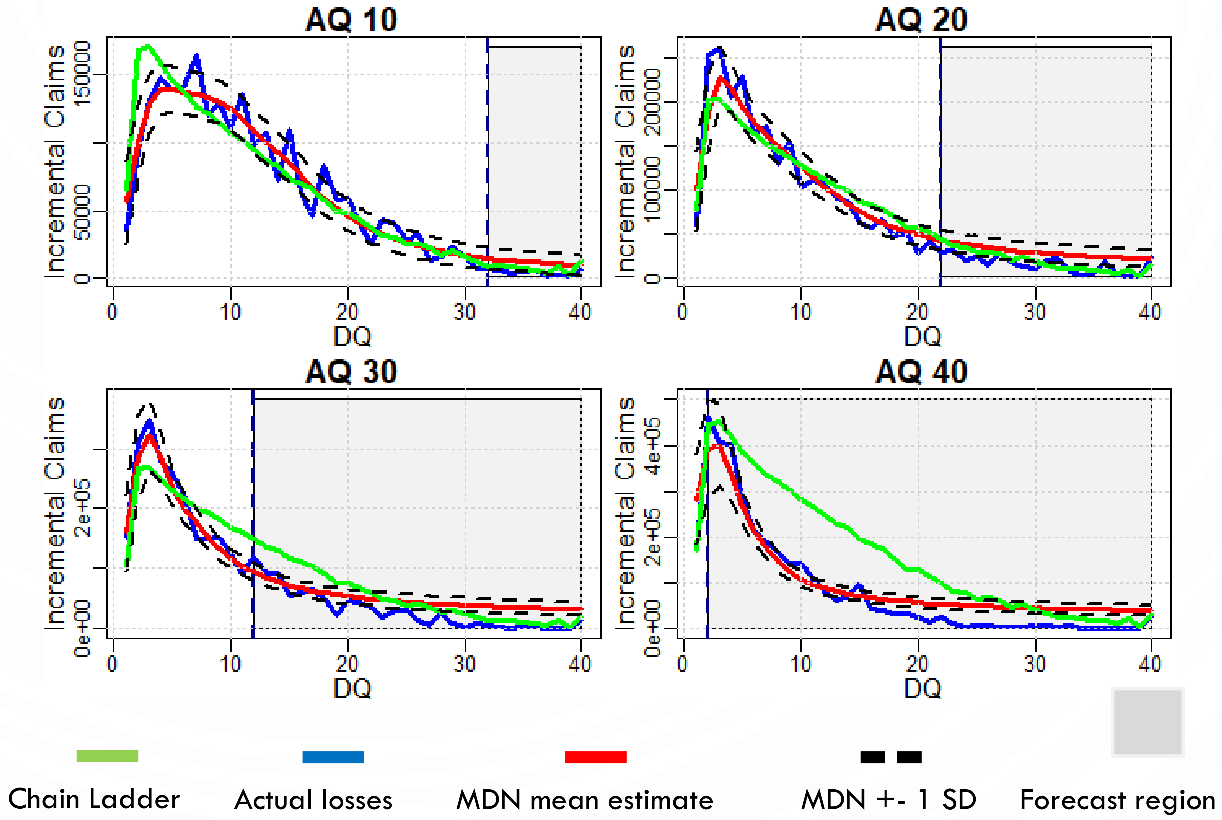
RESULTS – SMOOTH, ROBUST FORECASTS

- **Environment 4**
- Smooth, robust and accurate forecasts, even with volatile long tail claims



SAMPLE RESULTS – CAPTURED TRENDS

- **Environment 2:** Claim settlement speed increases
- MDN captured that change
- Chain Ladder (CL), assumes homogeneity, failed to capture change



QUANTITATIVE ANALYSIS: MEAN ESTIMATES

- Central estimate accuracy for $\boldsymbol{X_{i,j}}$
- RMSE metric (Root Mean Squared Error)
- Lower RMSE = more accurate fit
- Figure 4 shows the MDN had a lower RMSE than the Chain Ladder in the overwhelming majority of triangles in all environments

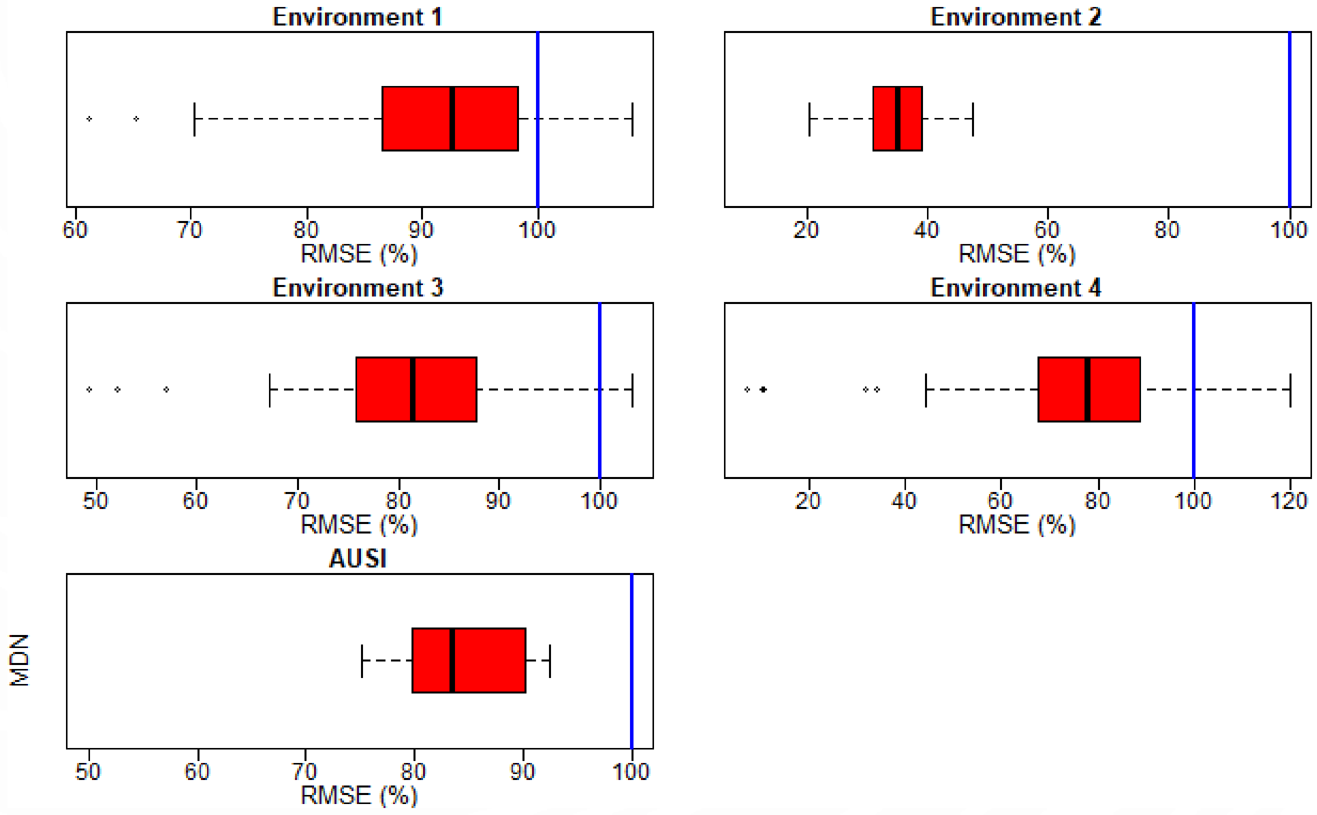


Figure 4: Boxplots of (RMSE (MDN)/ RMSE (Chain Ladder)) over 50 (10 for AUSI) triangles

QUANTITATIVE PERFORMANCE OVERALL

- MDN consistently produced **more accurate reserve estimates** than the Chain Ladder
- Log-likelihood measures distributional accuracy.
- MDN had a higher log-likelihood than Chain Ladder in majority of triangles
- Hence, MDN's **distributional forecast consistently outperformed** the stochastic Chain Ladder

Environment	Mean RMSE (% of Chain Ladder)	RMSE – Triangles MDN Outperformed CL (%)	Log-Likelihood – Triangles MDN Outperformed CL (%)
1	90	80	80
2	35	100	100
3	79	100	96
4	52	92	96
AUSI (Real data)	84	100	100

PRACTICAL CONSIDERATIONS

RESMDN: GLM – MDN HYBRID MODEL

- Follows the work of Wuethrich & Merz (2019), Gabrielli et al (2020)
- Hybrid model: GLM forms backbone, while the Neural Network boosts the GLM's residuals (Figure 5)
- NN picks up trends which the GLM missed
- Combines GLM interpretability with NN modelling power
- **Output** = **GLM** + **NN boosting**

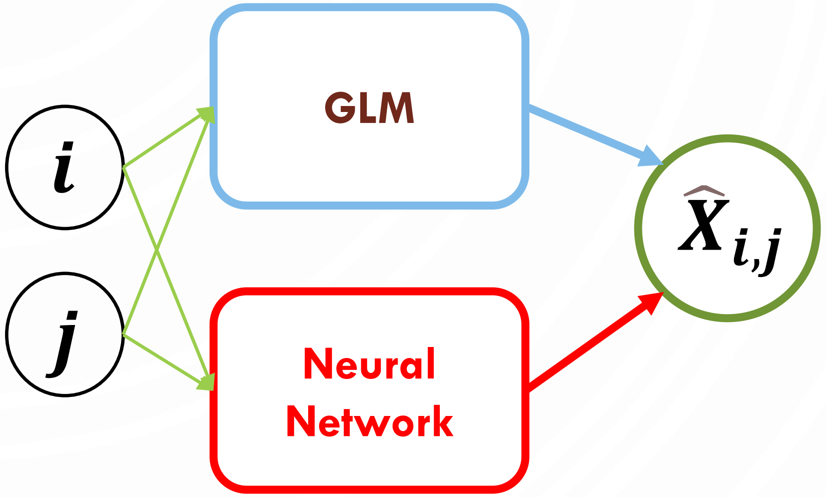
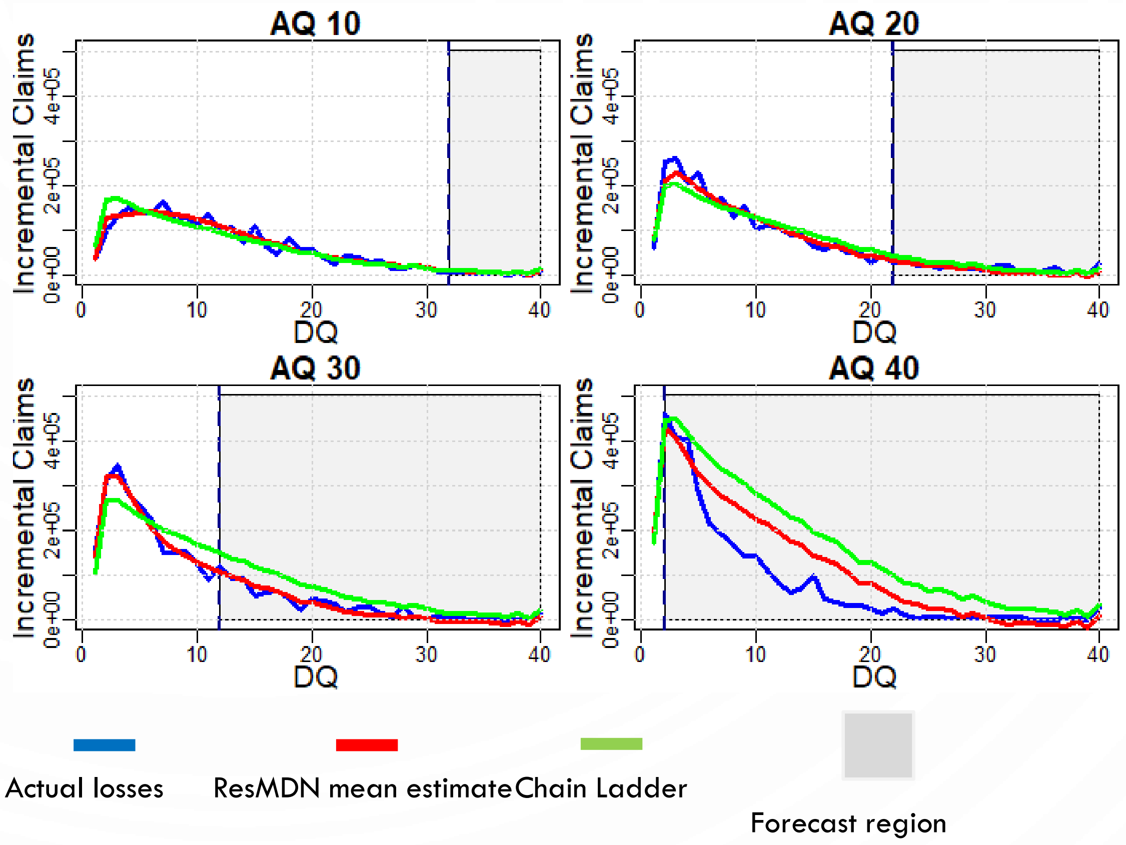


Figure 5: Diagram of ResMDN

SAMPLE RESULTS: BOOSTING CHAIN LADDER

- **Environment 2:** Claim settlement speed increases
- ResMDN corrected the Chain Ladder's errors, to an extent
- Results are easier to understand and justify



DISCUSSION

DISCUSSION

THE MIXTURE DENSITY NETWORK:

- Provides accurate mean and distributional forecasts
- Performs well with a 40x40 triangle
- Captured trends, outperformed the Chain Ladder in a variety of test environments
- Can be hybridised with a GLM for more interpretability
- Tested on hundreds of triangles, consistent results

SOME CONSIDERATIONS

- MDN isn't perfect. Missed some trends, especially ones embedded in little data
- Results came after experimentation. Several techniques developed to tackle the MDN's shortcomings, outlined in attached paper

CONCLUSION - FUTURE POTENTIAL OF NEURAL NETWORKS IN RESERVING

With actuarial supervision and guidance, the MDN is a powerful modelling tool for reserving (among other actuarial fields), and has future potential. For example:

- Learning from multiple triangles simultaneously; creating a 'reserving brain' that can be imported to any individual triangle
- Fitting different mixture distributions, such as Gamma, Pareto

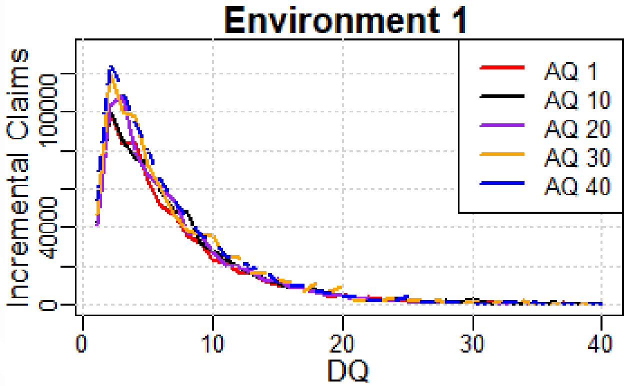
THANK YOU

REFERENCES

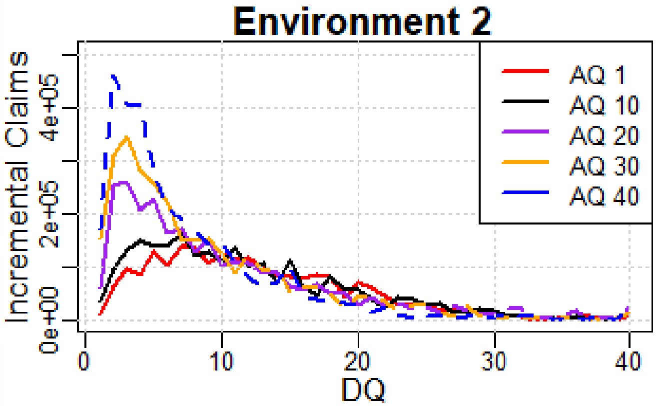
- Gabrielli, A., Richman, R., Wuethrich, M.V., 2020. Neural network embedding of the over-dispersed poisson reserving model.Scandinavian Actuarial Journal 2020, 1–29. URL:http://www.tandfonline.com/doi/abs/10.1080/03461238.2019.1633394.
- Wuethrich, M.V., Merz, M., 2019. Yes, we cann! ASTIN Bulletin: The Journal of the IAA 49, 1–3.

APPENDIX

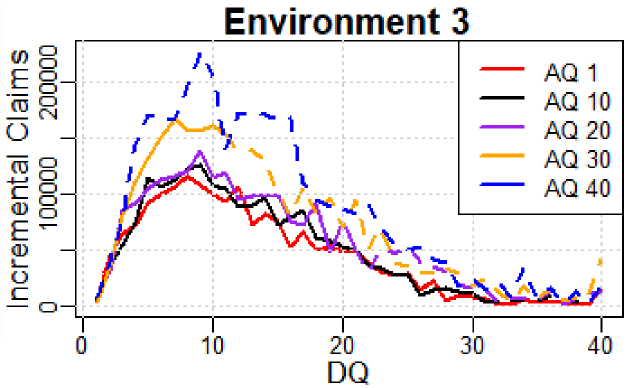
LOSS TRIANGLES



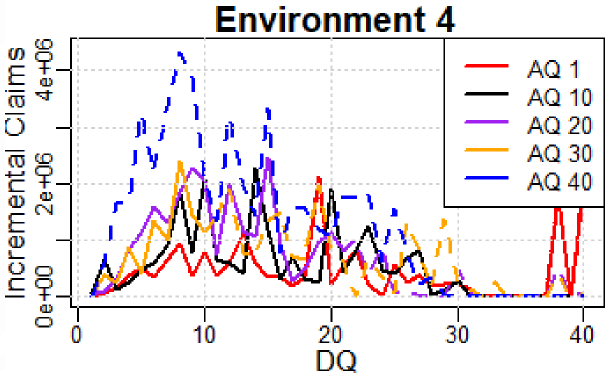
Environment 1: Simple, short tail



Environment 2: Speed up settlement



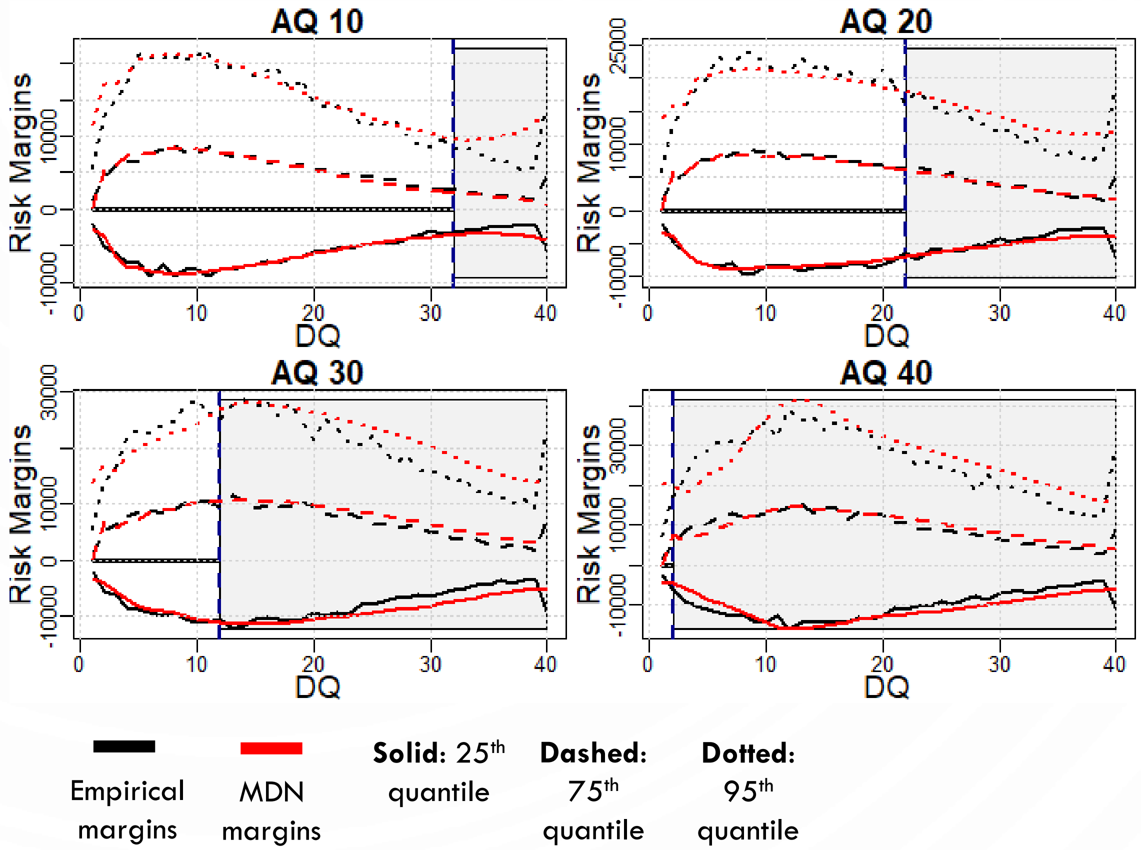
Environment 3: Increase in inflation



Environment 4: Complex, volatile long tail

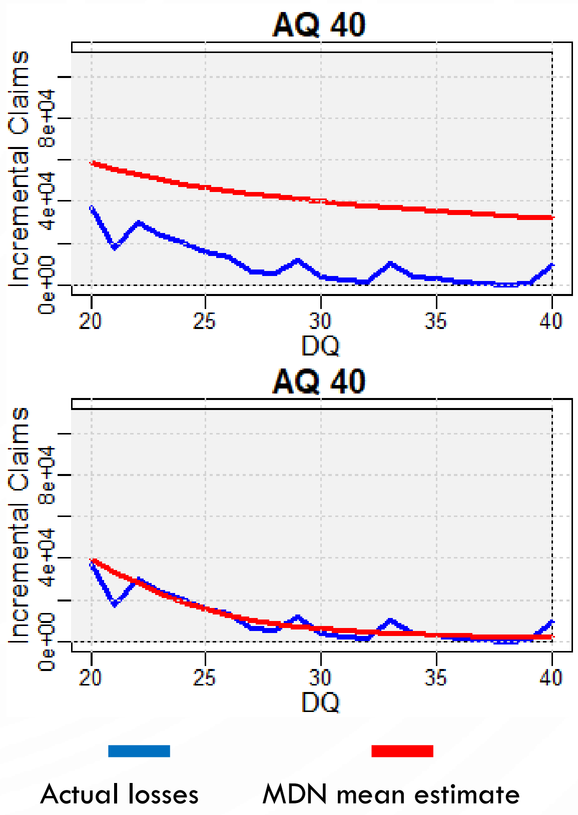
RESULTS – ACCURATE DISTRIBUTION ESTIMATES

- MDN accurately quantifies the risk margins (shape) of incremental claims, $\hat{X}_{i,j}$
- Hence, the distributional forecast is accurate



INCORPORATING ACTUARIAL JUDGEMENT

- Output can be constrained to boundaries set by the practitioner.
- Useful in case the MDN gives clearly unreasonable projections, or misses a visible trend
- Ensures forecasts are always reasonable
- Allows practitioner to control the model's macro-behaviour

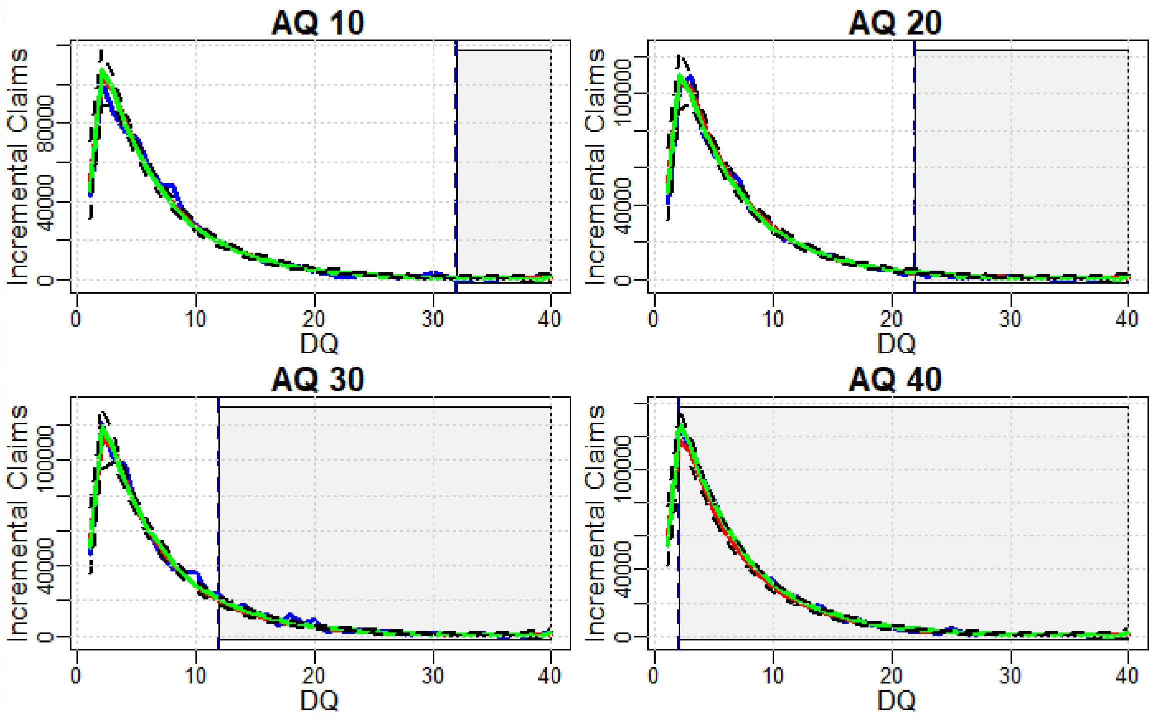


BEFORE

AFTER

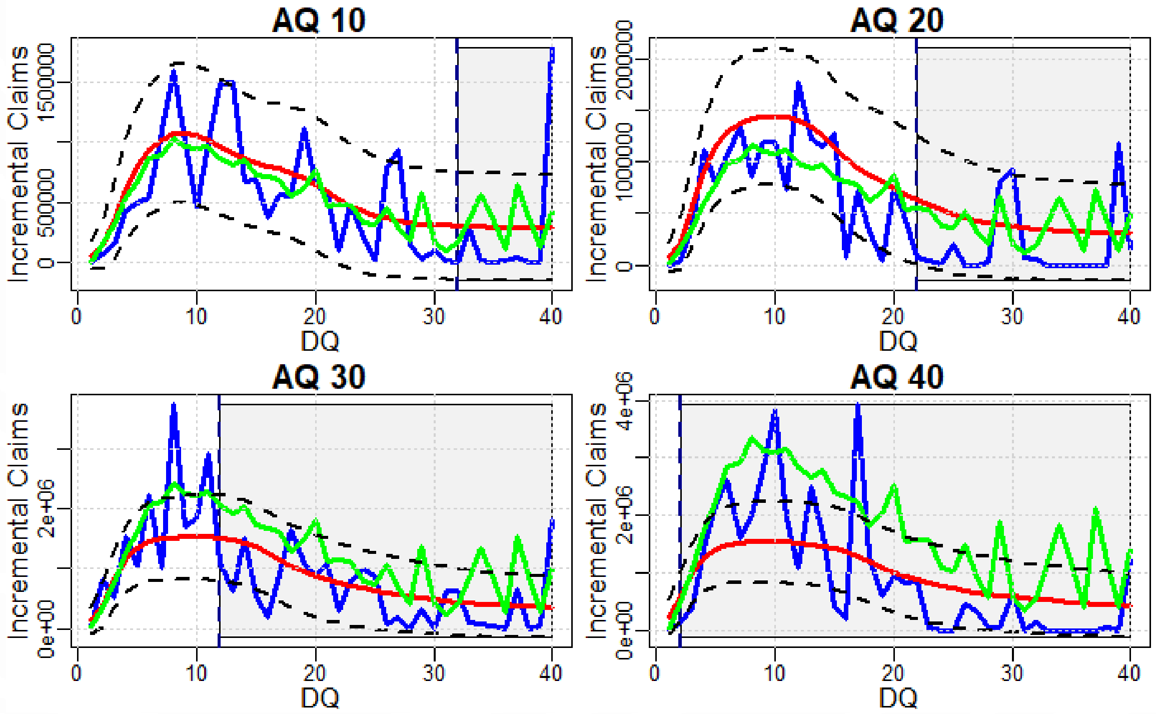
ENVIRONMENT 1

- Comparing results for Environment 1 (simple, short tail claims
- Both perform well, as data is homogeneous



ENVIRONMENT 4

- Comparing results for Environment 6 (complex, long tail claims)
- Chain Ladder is volatile, MDN gives smooth, accurate output



QUANTITATIVE ANALYSIS: DISTRIBUTIONAL ACCURACY

- Distributional accuracy of $f_{\hat{X}_{i,j}}$
- Log Score (Log-Likelihood) used
- Higher log score = better distributional fit
- Similarly, Figure 10 shows the MDN had a higher log score in the majority of triangles in all environments.

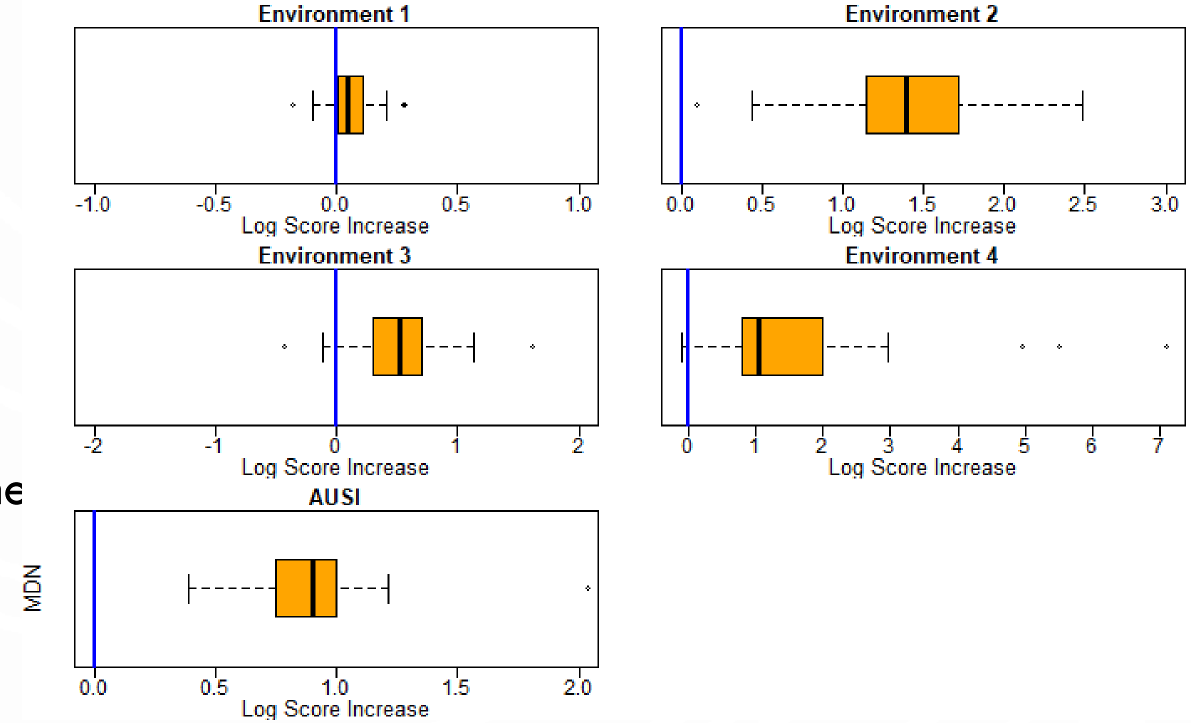
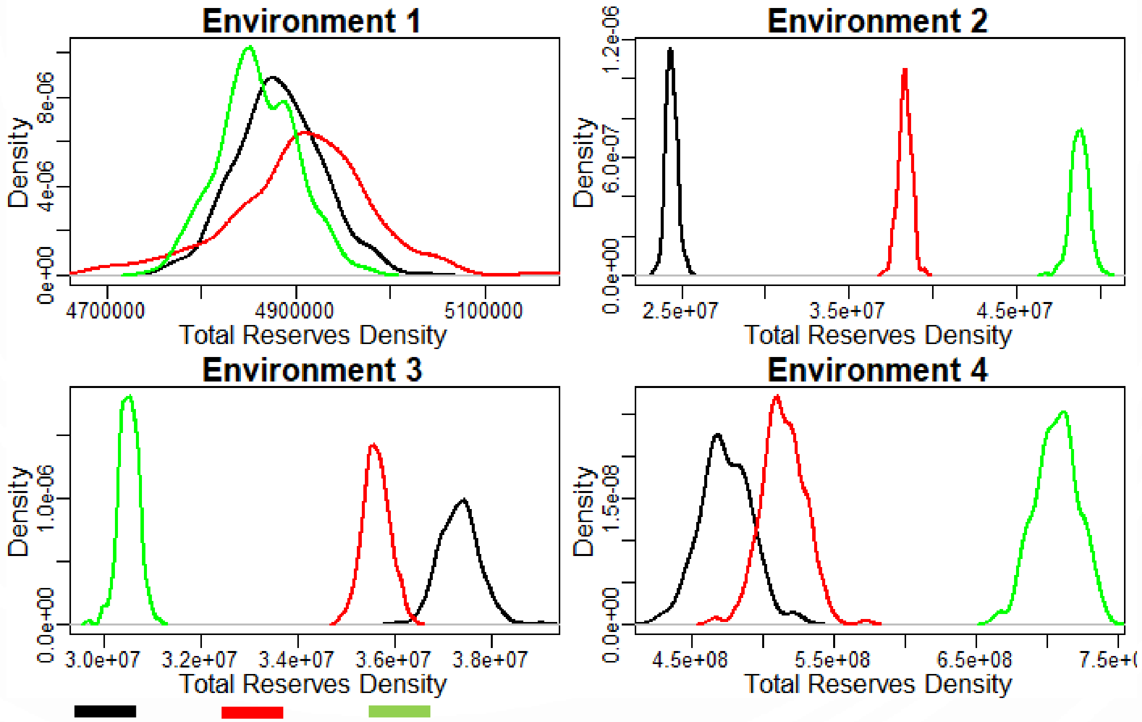


Figure 10: Boxplots of (LogScore(MDN) – LogScore(Chain Ladder))
Over 50 (10 for AUSI) triangles

RESULTS – TOTAL RESERVE ESTIMATES

- MDN had better mean and dispersion estimates of total reserves than the Chain Ladder
- Also more accurate **75% and 99.5% quantiles**



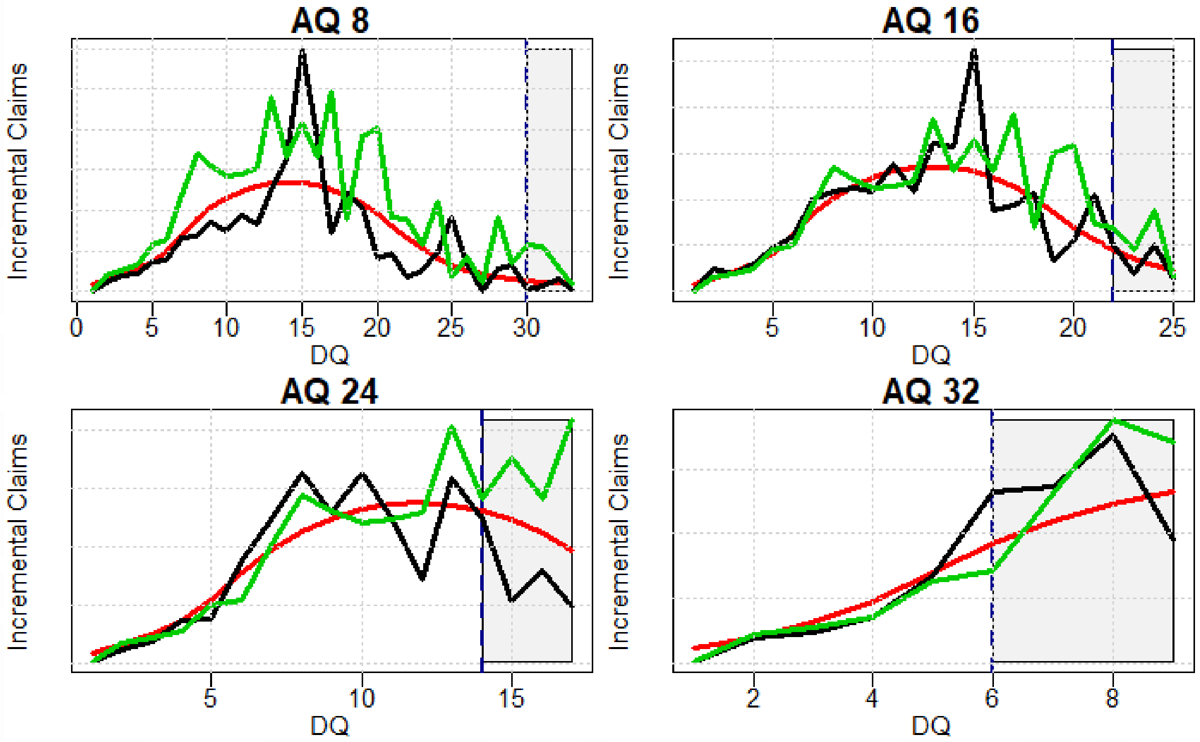
Empirical dist

MDN dist

Chain Ladder dist

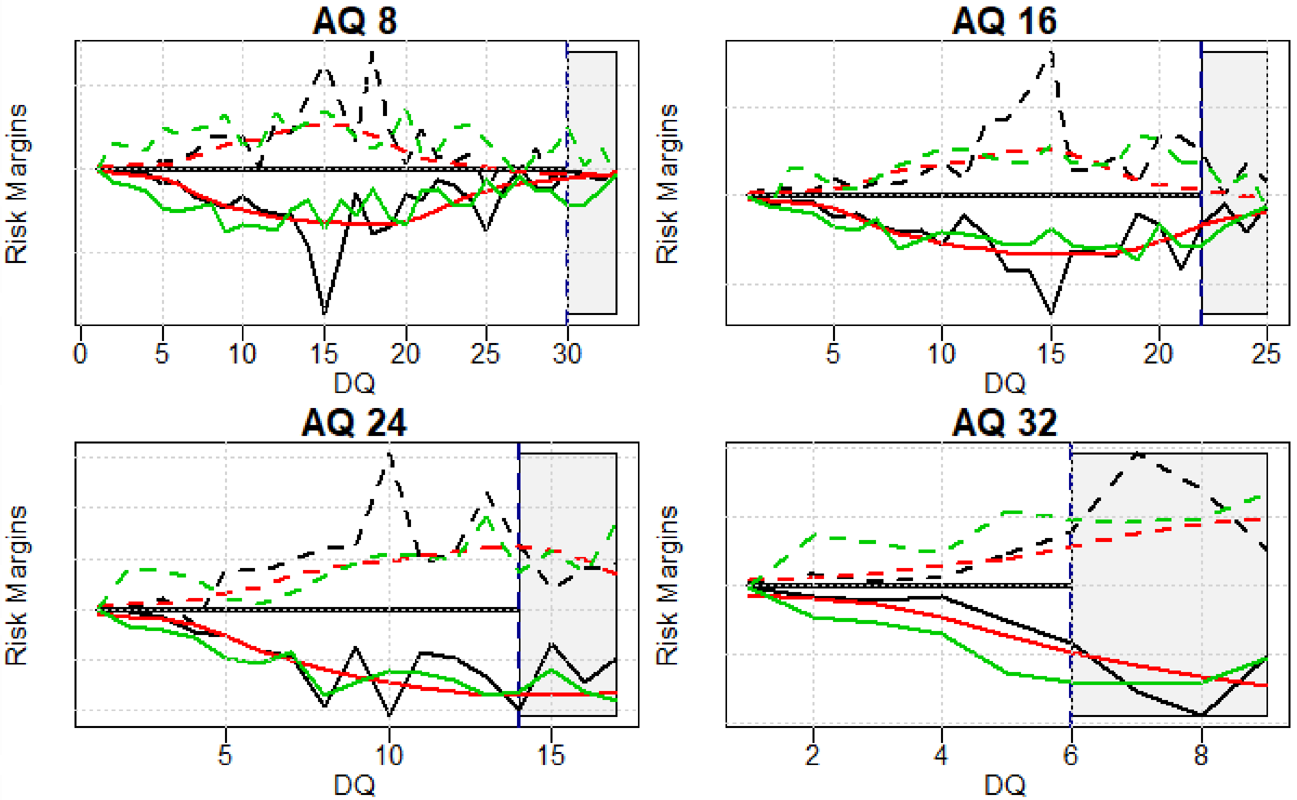
AUSI PLOTS

- Comparing the MDN and Chain Ladder on the AUSI set
- Black line is the empirical mean (based on 10 triangles)
- MDN gives smooth, more accurate results



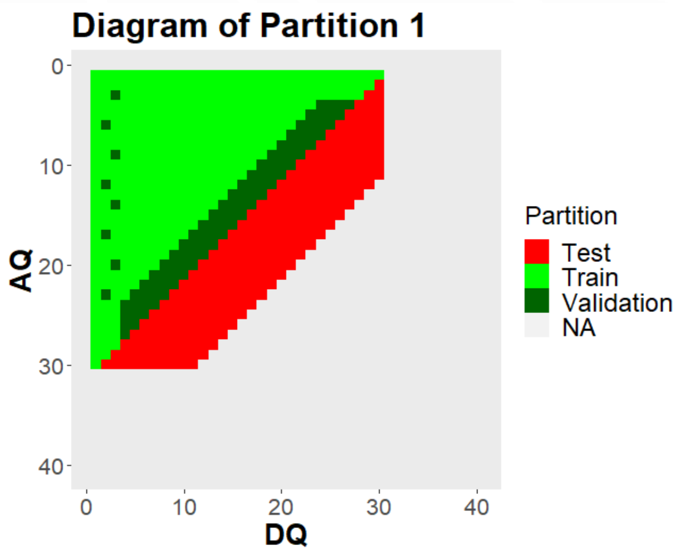
AUSI QUANTILES

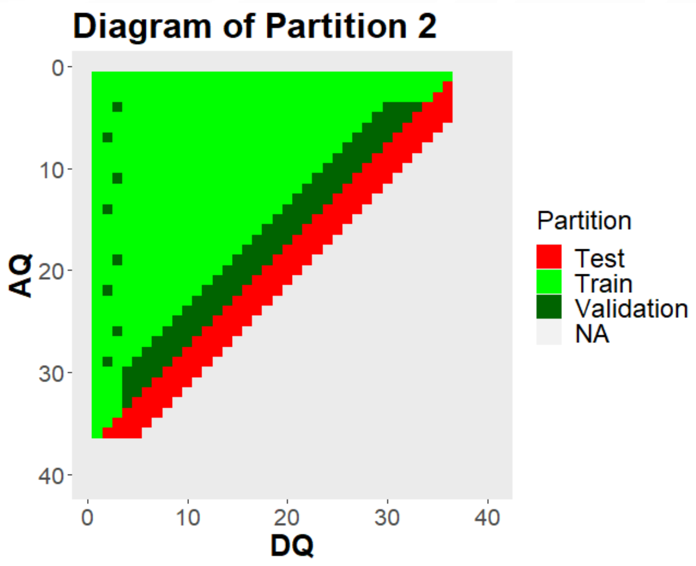
- Looking at the 25% and 75% risk margins for the AUSI set
- Again, MDN gives smooth, accurate results



ROLLING ORIGIN MODEL VALIDATION

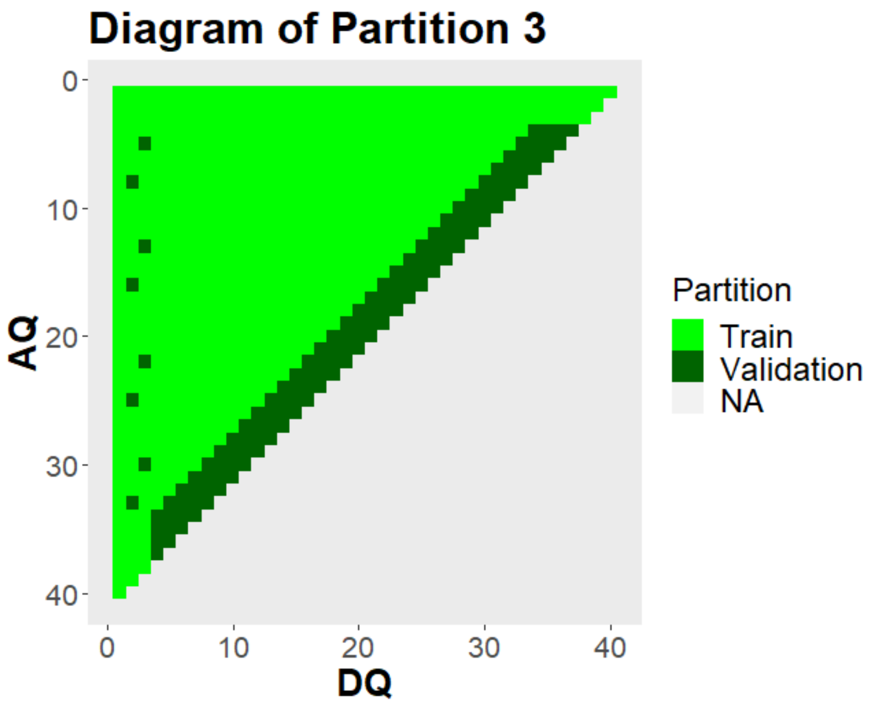
- Diagrams of the data partition into training, validation and testing sets





ROLLING ORIGIN 2

- Partition used in fitting the final model



MDN LOSS FUNCTION

$$NLLLoss(\mathbf{X}, \hat{\mathbf{X}}|\mathbf{w}) = -\frac{1}{|\mathbf{X}|} \sum_{i,j:X_{i,j} \in Train} ln(f_{\hat{X}_{i,j}}(X_{i,j}|\mathbf{w}))$$

- Negative Log-Likelihood loss function for the MDN
- We often add an MSE term, helps the MDN give better central estimate, capture trends

QUANTITATIVE METRICS

- RMSE: Calculated for incremental claims and for total reserves
- Log Score: Calculated for incremental claims
- Quantile Scores: Calculated for incremental claims and total reserves

$$RMSE(\mathbf{X},\hat{\mathbf{X}}) = \sqrt{\frac{\sum_{(i,j):X_{i,j}\in\mathbf{X}}(X_{i,j}-\hat{X_{i,j}})^2}{|\mathbf{X}|}}$$

$$RMSE(\mathbf{R},\hat{\mathbf{R}}) = \sqrt{\frac{\sum_{i=1}^{D}(R_i-\hat{R}_i)^2}{D}}$$

$$LogScore(\mathbf{X}, f_{\hat{\mathbf{X}}}) = \frac{\sum_{(i,j):X_{i,j}\in\mathbf{X}} ln(f_{\hat{X}_{i,j}}(X_{i,j}))}{|\mathbf{X}|}$$

$$QS(\hat{\boldsymbol{X}}_{\boldsymbol{q}},\mathbf{X}) = \frac{\sum_{(i,j):X_{i,j}\in\mathbf{X}}(\mathbf{1}(X_{i,j}<\hat{X}_{i,j,q})-q)(\hat{X}_{i,j,q}-X_{i,j})}{|\mathbf{X}|}$$

$$QS(\hat{\boldsymbol{R}}_{\boldsymbol{q}},\mathbf{R}) = \frac{\sum_{i=1}^{D}(\mathbf{1}(X_{i,j}<\hat{X}_{i,j,q})-q)(\hat{X}_{i,j,q}-X_{i,j})}{D}$$

CHOSEN MODELS

Environment	Model	λ_w	λ_σ	p	n	h	K
1	MDN	0	0.0001	0	60	4	2
1	ResMDN	0.001	0	0.1	100	1	5
2	MDN	0	0.0001	0.2	100	3	1
2	ResMDN	0	0	0	40	3	1
3	MDN	0	0*	0	80	3	4
3	ResMDN	0	0	0.1	60	2	1
4	MDN	0	0.0001	0.1	60	4	3
4	ResMDN	0	0	0.1	20	3	3
AUSI	MDN	0	0	0.2	100	2	4

Table A: The most accurate model design yielded by the algorithm listed in Section 3.2. λ_w represents the L2 weight regularisation coefficient, λ_σ represents the L2 activity regularisation coefficient on the σ output, p is the dropout rate, n is the number of neurons in each hidden layer, h is the number of hidden layers and K is the number of components in the mixture density. *see Appendix D

QUANTITATIVE RESULTS: INCREMENTAL CLAIMS

Environment	Model	Mean RMSE	RMSE (% of ccODP)	Mean LS	Mean QS (75%)	Mean QS (99.5%)
1	ccODP	1,695.0	100	-8.09	386.1	30.9
1	**MDN**	**1,527.5**	**90.1**	**-8.03**	**375.4**	**23.1**
1	ResMDN	1,766.6	104.2	-8.31	434.3	57.05
1	ResMDN-PC	1,697.2	100.1	-8.24	386.4	30.2
2	ccODP	52,168.8	100	-12.94	10,065.7	359.0
2	**MDN**	**18,150.7**	**34.8**	**-11.53**	**5,281.3**	**222.1**
2	ResMDN	36,862.6	70.7	-13.54	6,888.2	596.4
2	ResMDN-PC	39,734.1	76.2	-13.57	7,048.5	260.6
3	ccODP	16,778.8	100	-11.35	5,419.7	1,385.5
3	**MDN**	**13,223.8**	**78.8**	**-10.83**	**4,141.4**	**307.9**
3	ResMDN	15,937.0	95.0	-11.60	4,667.5	655.8
3	ResMDN-PC	16,030.3	95.5	-11.27	4,861.6	1,012.3
4	ccODP	1,272,623.9	100	-15.54	259,273.4	36,199.6
4	**MDN**	**657,283.5**	**51.6**	**-14.02**	**214,133.5**	**23,782.9**
4	ResMDN	1,274,700.9	100.2	-18.54	258,145.7	37,902.0
AUSI	ccODP	-	100	-14.04	124,976.3	19,324.8
AUSI	**MDN**	**-**	**84.2**	**-13.07**	**105,559.9**	**12,001.0**

Table 1: The average score, over 50 triangles, of each quantitative metric; the RMSE, log score (LS) and quantile scores (QS) for the 75% and 99.5% levels. The MDN outperformed the ccODP in all environments and metrics when the average is taken.

QUANTITATIVE RESULTS: TRIANGLES OUTPERFORMED

Environment	Model	RMSE	Log Score	Quantile Score (75%)	Quantile Score (99.5%)
1	MDN	80	80	62	92
1	ResMDN	34	6	60	64
2	MDN	100	100	100	100
2	ResMDN	100	36	100	42
3	MDN	100	96	100	100
3	ResMDN	80	62	90	88
4	MDN	92	96	82	76
4	ResMDN	8	6	54	40
AUSI	MDN	100	100	100	70

Table 2: The percentage of triangles in which the MDN outperformed the ccODP in that specific metric.

QUANTITATIVE RESULTS: TOTAL RESERVES

Environment	Model	RMSE ($\times$100,000)	QS(75%) ($\times$10,000)	QS(99.5%) ($\times$10,000)
1	**MDN**	**1.12**	**3.72**	**0.40**
1	ResMDN	4.21	11.26	8.73
1	ccODP	0.92	2.87	0.55
2	**MDN**	**92.6**	**228.5**	**49.2**
2	ResMDN	111.1	257.5	54.2
2	ccODP	238.8	601.6	125.0
3	**MDN**	**26.1**	**129.5**	**109.5**
3	ResMDN	39.1	169.6	154.2
3	ccODP	61.4	431.7	529.0
4	**MDN**	**807.0**	**2,098.7**	**602.7**
4	ResMDN	3,871.0	6,972.5	1,534.0
4	ccODP	3,823.7	6,802.6	1,506.4

Table 3: The RMSE and quantile scores (QS) at the 75% and 99.5% levels, calculated for total reserve estimates, $\hat{R}$. The ccODP outperforms for Environment 1, but the MDN outperforms otherwise.