Institute and Faculty of Actuaries

Granular Reserving Dialogistic in Machine Learning

Asif John

21 June 2018

Contents

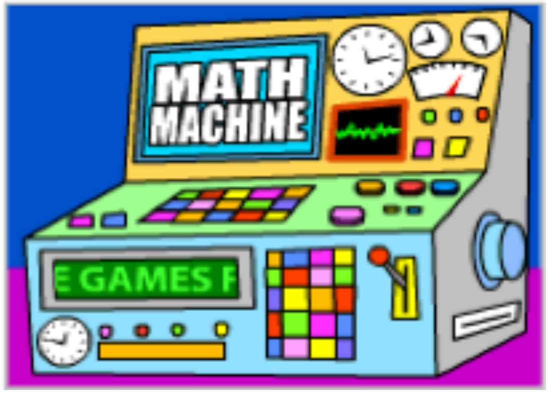
- Machine Learning Overview
- How can this benefit us?
- A proposed method
- How can this be implemented in our current process
  - Tools and Interface
- Demo – PoC (Proof of Concept)
- Insights:
  - Summary of Statistics
  - Graphical visualisation
  - Granularity of peril types
- Appendix – Case Studies
- Questions

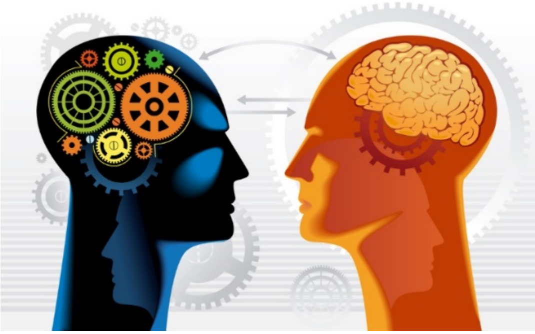
Institute
and Faculty
of Actuaries

Machine Learning

21 June 2018

Machine Learning overview





- Popularity of machine learning driving innovation
- Can Machine Learning be used for reserving?
- Reduce information loss and improve insight
  - Inability to understand the drivers of reserving results
  - Inability to adjust assumptions to claim characteristics
- Uptake limited by trade off of simplicity vs accuracy
- Companies now investigating different predictive techniques to mitigate the Mean Absolute Error (MAE)
- Machine learning 'blackbox' like but different machine learning methods which we can use:
  1. GBM (Gradient Boosting Machine)
  2. Decision Tree (the random forest)
  3. LASSO (least absolute shrinkage and selection operator)

How can this benefit us?

Estimate

- **One emerging view is that the errors in the reserving estimates can be explained much better by using machine learning on granular claims data.**

Classical

- The classical reserving methods use a one-size-fits-all approach, so it is difficult to learn from the actual vs expected. Machine learning could give insight here
- Machine learning models use the claims and exposure features which affect the development, frequency and severity.

Parameter

- Simply put, machine learning would use algorithms to estimate a different development factor for brain injury vs muscle injury
- Parameter estimation involves learning from historical granular data, minimising the errors and back-testing the parameters

Depth Analysis

- It therefore allows for a more in-depth analysis of the actual vs expected, e.g. brain injuries may have deteriorated worse than expected

Monthly

- Although machine learning models are computationally intensive and complex, they can be implemented very easily once built.
- Importantly, they can be rerun frequently within small intervals (say monthly) to monitor the actual vs expected.

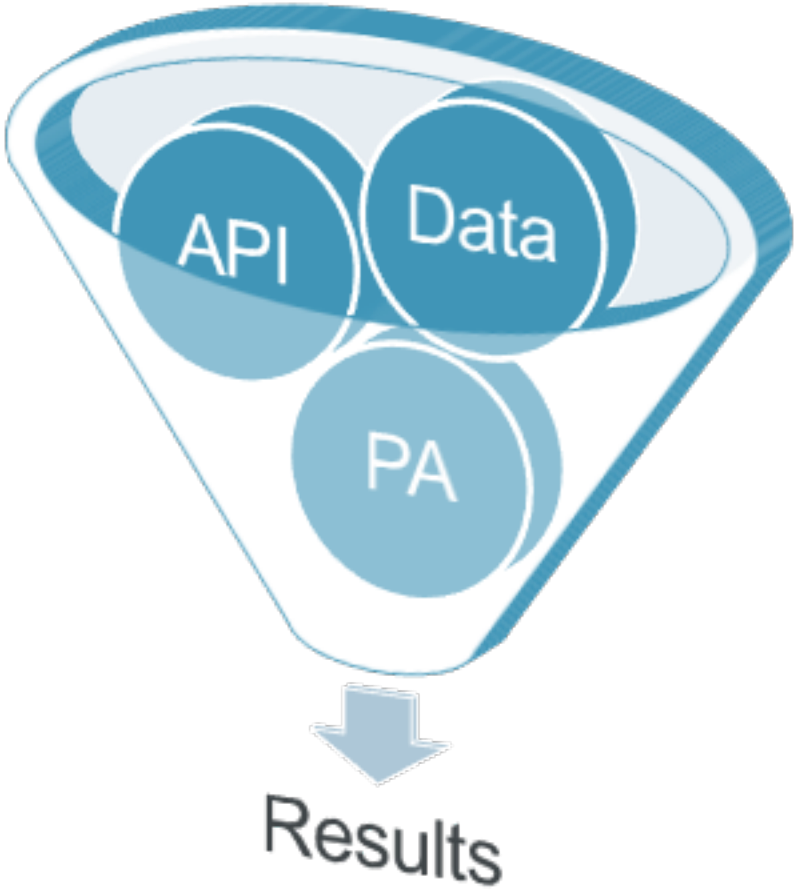
Validation

One suggestion from the working party is not for machine learning to replace the traditional reserving techniques, but rather to **validate and enhance them.**

Strategy - Value

Importantly, in this case machine learning models should be used to understand and explain the actual vs expected, and over time, help to develop more granular assumptions for traditional models such as loss ratios, development factors, frequency and severity.

How could we implement THIS?



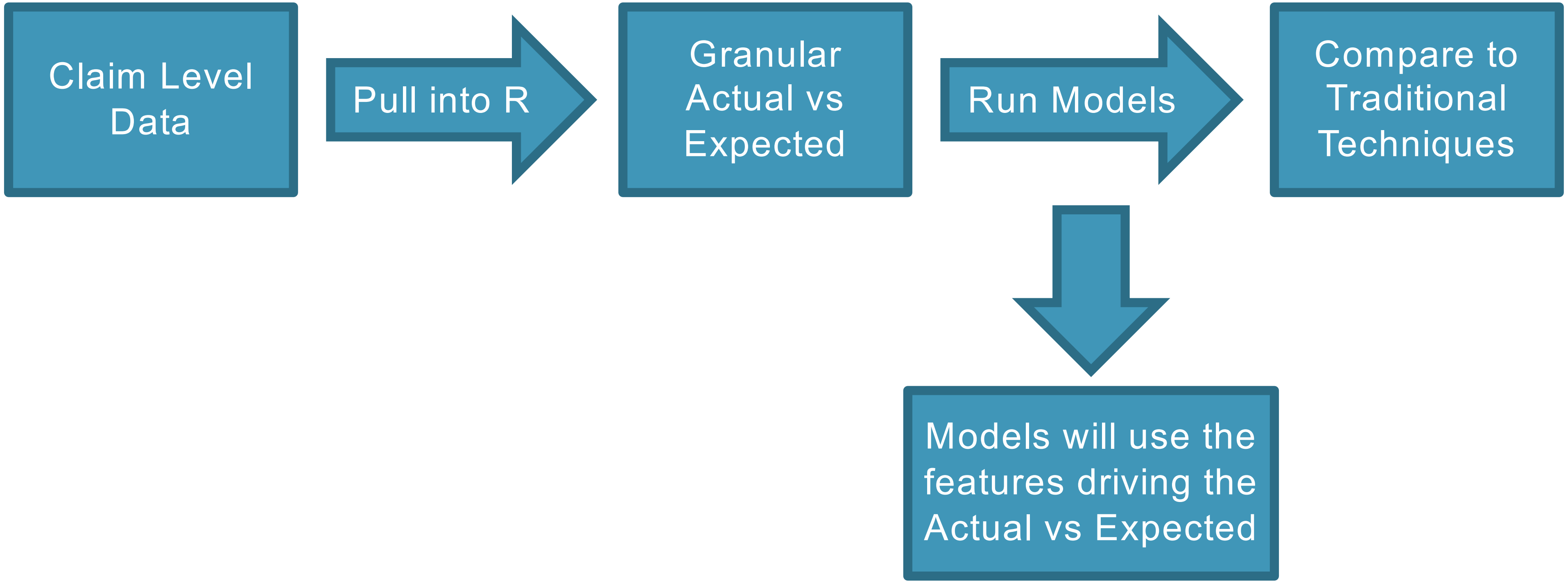
Tools and Interface:

Common Interface used for Reserving	Common ML tools
Excel	R and R-Shiny
Access	Python
Bespoke Interface	Spark
Other	RapidMiner

DEMO – PoC (Proof of Concept)

Flow chart

We show in here flow chart on how we could implement this in practical terms (with assumptions caveated). For example; every company has different interface and we will base it with excel tool and using R-Shiny



ML - Overview illustrative results

Summary Statistics

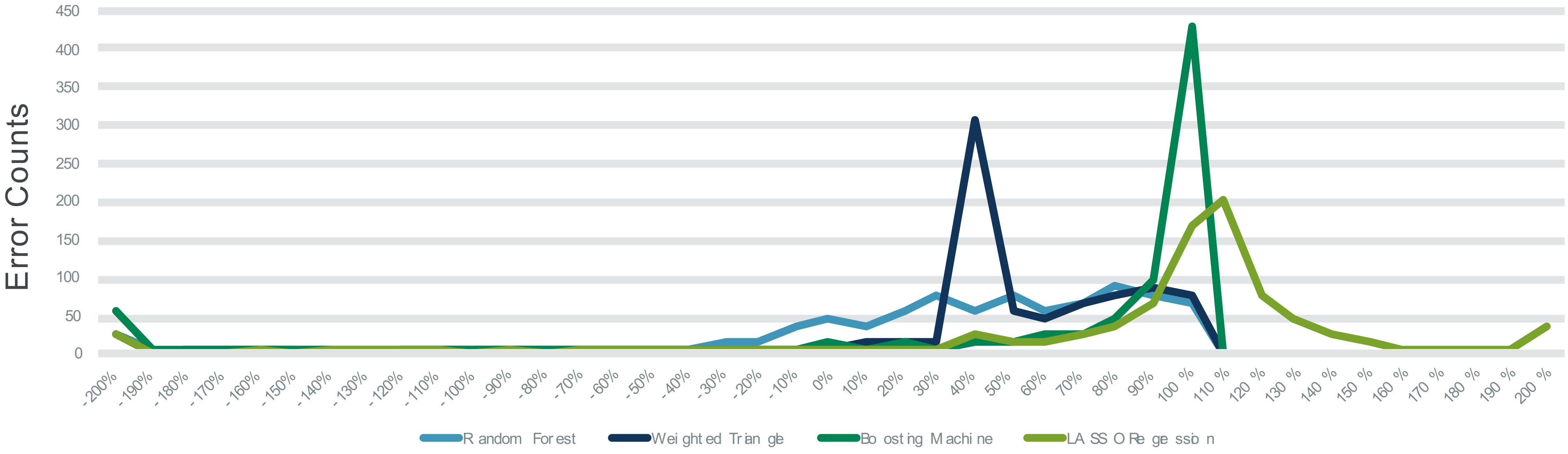
Method	Total Predicted	Actual	Actual vs Predicted	Mean Error %	Median Error %	Total Absolute Error	Absolute Error %
Triangle	16,764,770	15,685,367	1,079,403	7%	37%	12,474,066	80%
Forest	15,884,229	15,685,367	198,862	1%	43%	12,714,048	81%
GBM	15,639,526	15,685,367	(45,841)	0%	90%	20,462,309	130%
Lasso	25,064,981	15,685,367	9,379,614	60%	100%	32,916,272	210%

Comments

- Triangle = has lowest Absolute error but suffers higher mean error
- Forest = has slightly higher absolute error but very low mean error
- GBM = has lowest mean error but very high absolute errors, see predictions which are very sticky around mean mark
- Lasso regression = performs worst due to linear effect of the model, cannot capture the non-linear trends in the data

ML – Overview Error Distribution

Comparison of methods

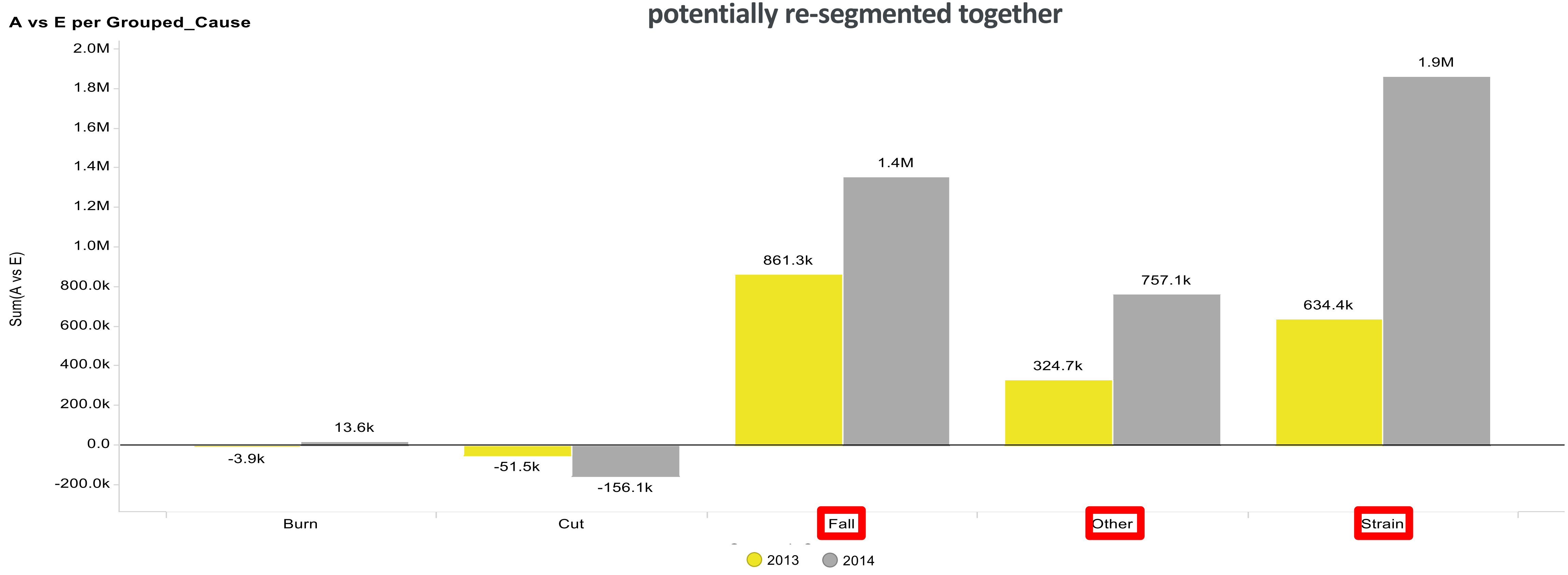


Commentary

- Employer's Liability Bodily Injury
- Large losses are not capped, large loss is >100K
- Prediction Error is (Actual - Expected)/Expected
- Total Claims 4815, split into 3972 Training 843 Tested (for prediction error check performance)
- Variables used - Incurred, Paid, Case, Type of Injury, Part of Body, State

Overview Granular A vs E – Bodily Injury – Total (losses)

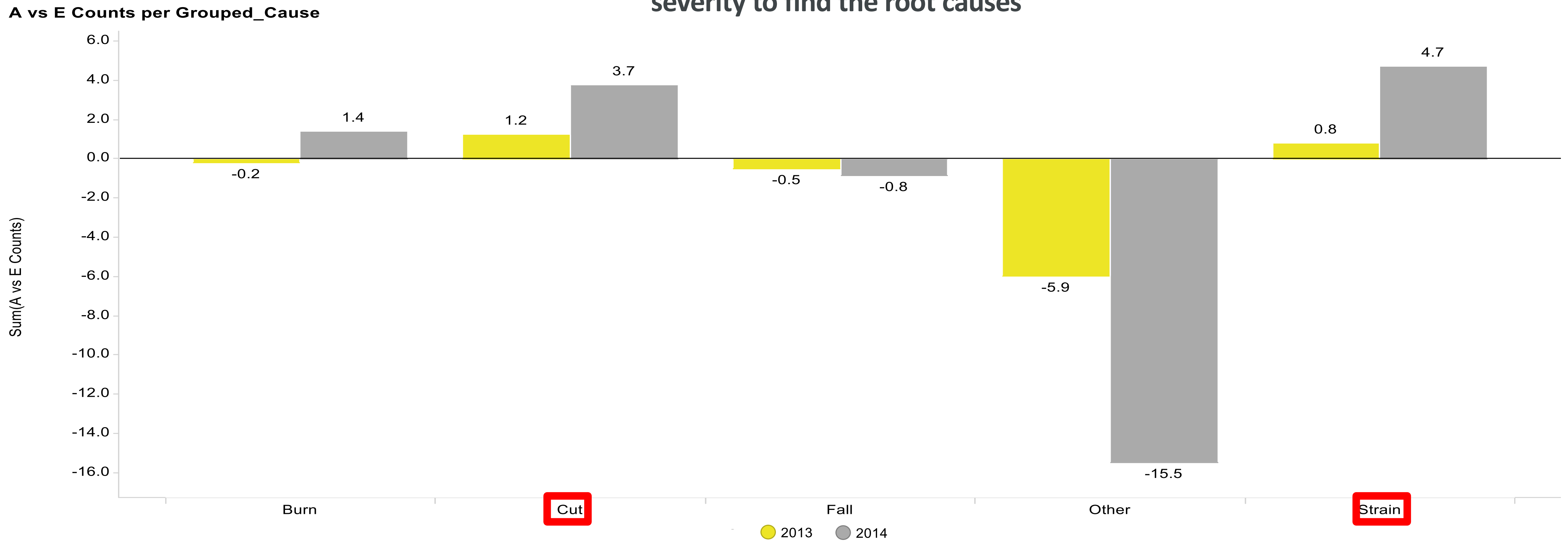
Claim types/injuries that consistently show adverse development can be potentially re-segmented together



Advantages – Easy insights into drivers of adverse development, also feeds back valuable information from reserving to business planning and analytics

Overview Granular A vs E – Bodily Injury – Counts

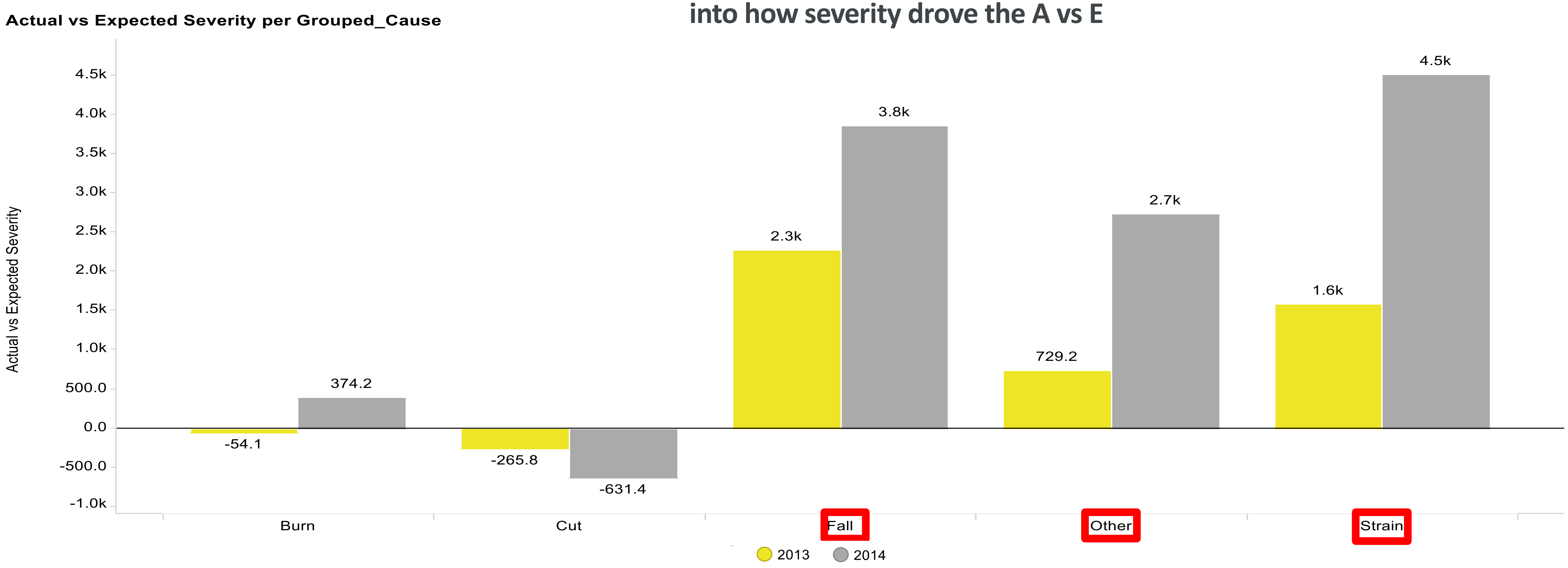
This adverse development can be further broken down into frequency and severity to find the root causes



For example, here we find counts A vs E is not significant, so it is actually severity that is driving the A vs E. So we can examine the severity data closely

Overview Granular A vs E – Bodily Injury – Severity

Looking into the Actual versus Expected severity gives us more insights into how severity drove the A vs E



This can feed back *valuable* information into the reserving process, business planning as well as pricing analytics

Institute and Faculty of Actuaries

APPENDIX – Case Studies

21 June 2018

Introduction

- Classical Reserving Techniques –
  current use of granular claim-level and exposure-level data in reserving is very limited?

  For e.g., claims are grouped into segments based on:
  - Line of Business
  - Claim origin year
  - Maturity of claims
  - Attritional vs Large
  - Peril – PD, BI, etc.

- This impedes the reserving process in two major ways:
  - Inability to understand the drivers of reserving results –
    there is a need to better understand the A vs E movements, which cannot be done until claims are segmented by characteristics that actually drive development – cause of claim, location of claim

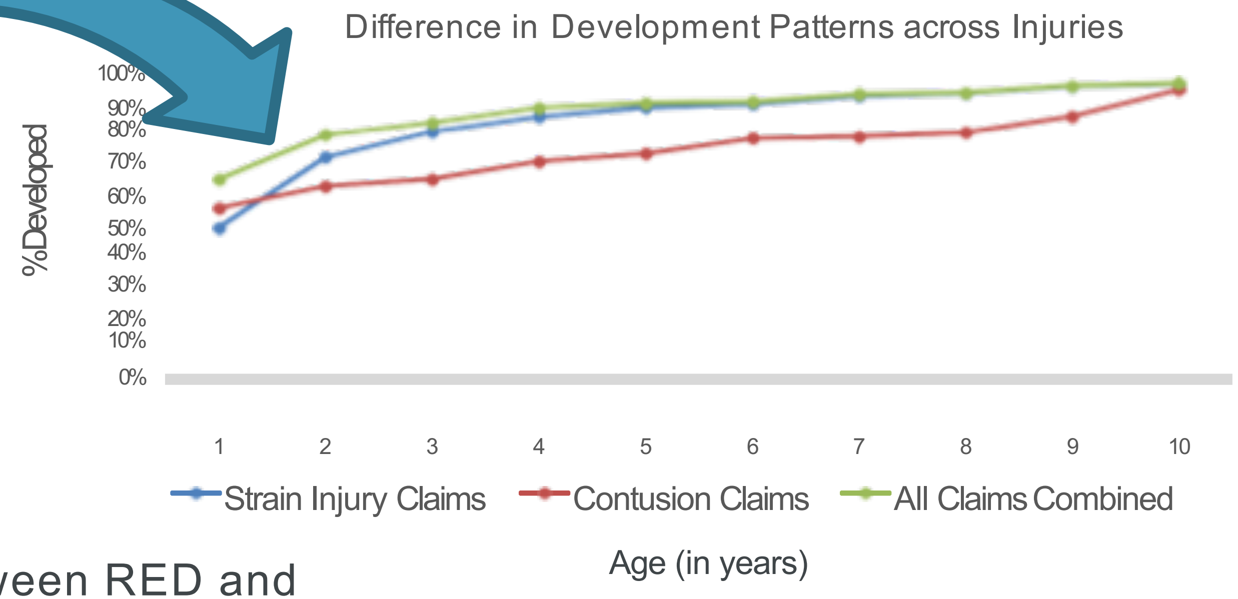
  - Inability to adjust assumptions to claim characteristics (claim type, exposure type)

Real-World Examples

- Following examples illustrate the need for granular reserving:
  - Motor Liability: Claims caused by rear end car collisions are very different than claims caused by head on collisions – both in terms of severity and development patterns
  - Employer`s Liability: Soft tissue injuries such as lower back strains develop very differently compared to minor injuries like contusions (bruises), lacerations (cuts), etc.
- Apart from looking at claim features, exposure features also matter:
  - Motor Liability: Loss Ratios in highly litigious regions will be higher than other regions
  - Employer`s Liability: Loss Ratios for high risk industries like Construction, etc. could be higher than Loss Ratios for low risk industries such as Restaurants, Clerical work, etc.

Real-World Examples

- E.g. of differing development patterns between claim types:



The % difference between RED and Green at DY2 is circa 33%

Industry Needs

- Need for machine learning in reserving:
  - Need to segment different claim types appropriately
  - Claim segmentation will improve Chain Ladder and Average Cost Per Claim (ACPC)
  - Exposure type will also affect claim patterns and loss ratios (e.g. State in personal lines, industry in commercial lines)
  - Granular reserving improves insight into risk profile, thus improves pricing, capital modelling and risk management
  - Need to explain actual vs expected movements by attributing deviations to underlying causes/claim types

IFRS 17

IFRS 17 starts with a few **key principles**:

- Current market-consistent valuation of expected future cash flows
- Reflects risk associated with those cash flows

=> discounted, risk-adjusted cash flows

- No recognition of profit until services are provided, losses recognised immediately
- Transparency through disclosures
- Revenue metrics consistent with other non-insurance industry revenue metrics

Overview:

- General measurement model often referred to as the Building Block Approach (BBA)
- Simplifications exist for eligible contracts:
  - Premium Allocation Approach (PAA) for unexpired risk component
  - PAA with undiscounted expired risk
- Recognition of contracts - earliest of start of coverage and premium receipt (plus onerous contract test)
- Applies to outwards reinsurance too
- More granularity required (level of aggregation)
- Detailed disclosure requirements

Machine Learning – A Solution

- How can machine learning help solve this problem?

  - It is not possible to apply different assumptions to every claim type

  - need to account for homogeneity & credibility

  - Machine Learning techniques help identify claim features that are important, and help determine optimal segments

  - Our research will demonstrate a fool-proof, industry-tested way of segmenting claims to improve reserve accuracy, and will demonstrate the improvement in Actual vs Expected

Case Study: Clustering and Chain-Ladder

- Use of Clustering to improve chain ladder reserving:
  - We performed our research on Worker's Compensation LOB (the US equivalent of Employer's Liability) since this class has been known to experience adverse development.
  - Our dataset consists of transactional level data for 18,922 Worker's Compensation claims, with key claims characteristics such as Cause of Injury, Nature of Injury, Body Part Injured, Location where the claim occurred, Occupation of the worker, Industry of the employer, etc.

Case Study: Clustering - Considerations

- Key considerations for the clustering algorithm:

  - Which variables should we consider for clustering claims? – This is a key decision that every organization has to make, based on the Line of Business and the availability of data

  - Should we group claims with similar injuries? Similar industries? Similar occupations?

  - The deciding factor: Which variables drive development? – exploratory analysis and supervised learning techniques help identify variables that affect claim severity & development

Case Study: Clustering - Key Variables

Our exploratory data analysis gave us the 4 key variables:

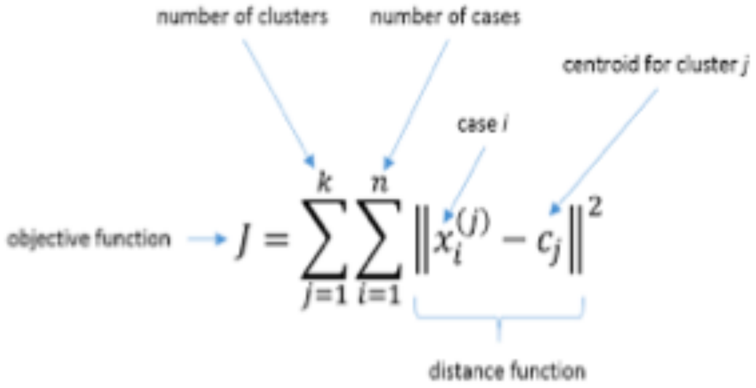
- Cause of Injury – Certain events (e.g. Fall, Slip and Trip) lead to injuries that emerge late – and hence have very different patterns and severity than other simple injuries (like Burns)
- Nature of Injury – This is the actual injury type suffered by the worker – for e.g. Falls can cause fractures, strains, lacerations, etc. – each of which have very different patterns and severity
- Body Part – Soft tissue injuries emerge late, WC regulations
- Location/State – difference in litigation culture, WC regulations

Case Study: Clustering - Dimensions

- Once we finalized the variables, it was important to determine the dimensions to be used for the clustering algorithm:

  - K-means Clustering Algorithm was used: This algorithm groups N observations into K clusters, in which each observation belongs to the cluster with the nearest mean.

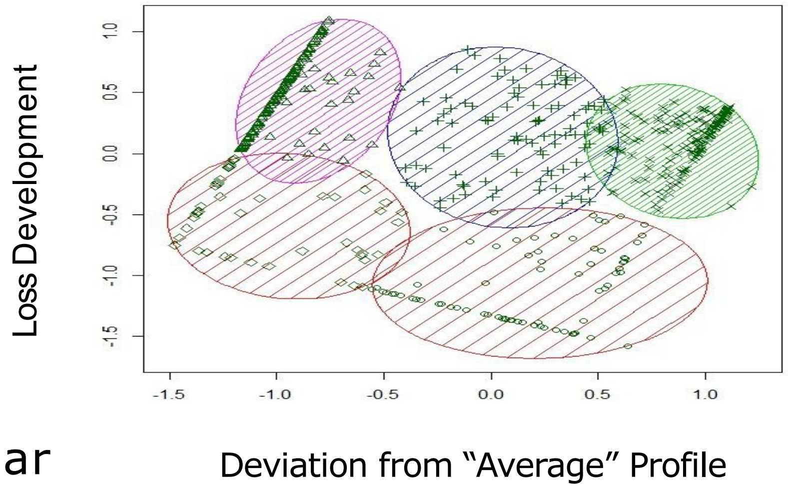
    - Clusters the data into k groups where k is predefined.
    - Select k points at random as cluster centers.
    - Assign objects to their closest cluster center according to the *Euclidean distance* function.
    - Calculate the centroid or mean of all objects in each cluster.
    - Repeat steps until the same points are assigned to each cluster in consecutive rounds



  - This algorithm will group similar injury types into 1 segment. But how is similarity defined?
  - Two injuries might be similar in terms of frequency but very different in terms of severity

Case Study: Clustering - Dimensions

- Here, we introduce the concept and importance of using "dimensions" in clustering
- Dimensions: numerical quantities that define features of data. E.g. Frequency, Loss Ratio
- "Similarity" has to be defined in terms of the key dimensions that matter in reserving:
  - Loss Development Patterns (CDF/LDF)
  - Deviation from "average" claim profile
  - Frequency
  - Loss Ratio
  - Severity
- E.g.: Two industries may be grouped in the same cluster…
- If Loss Ratio and Frequency in both industries are similar
- Variables can be clustered on more than 2 dimensions



Case Study: Clustering - Results

- Once the dimensions are determined, the clustering algorithm should not be run like a "black box".
  - It is important to validate the results of the algorithm against expectations
  - and find out the reasons for differences, if any.

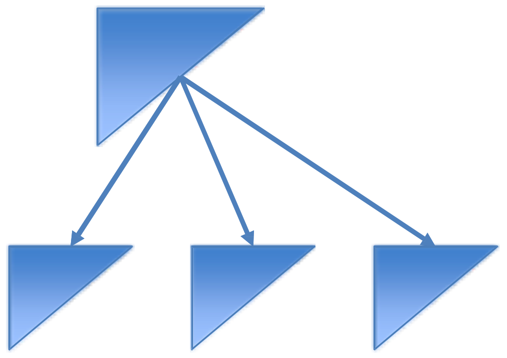
- The table shows results of clustering "Cause of Injury"
- Development = Cumulative Development Factor (CDF)
- Deviation = % Difference from the average severity:
  - Strains are highest risk claims…
  - And hence a separate cluster
  - "Fall" & "Motor" grouped into one.
  - (Make sense based on the dimensions of Development, Deviation.
  - As expected, "Other" in jury have lower development factor.

Cause of Injury	Development	Deviation	Cluster
Cause – Other	1.410	-24.4%	3
Cause – Strain	1.941	17.8%	1
Cause – Fall	1.707	5.0%	2
Cause – Motor	1.816	2.5%	2

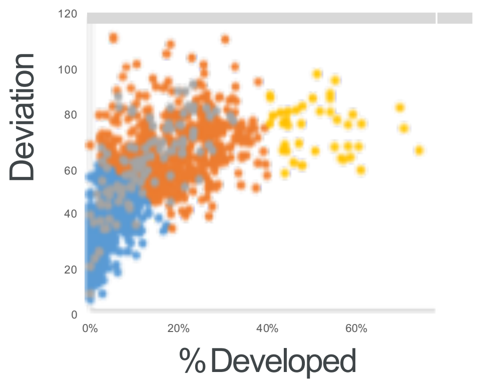
Evaluating the results of Clustering

Case Study: Segmenting the Triangles

- The next step was to segregate the aggregate triangle into 3 separate triangles based on results of the clustering algorithm:



Segmentation of Aggregate Triangle

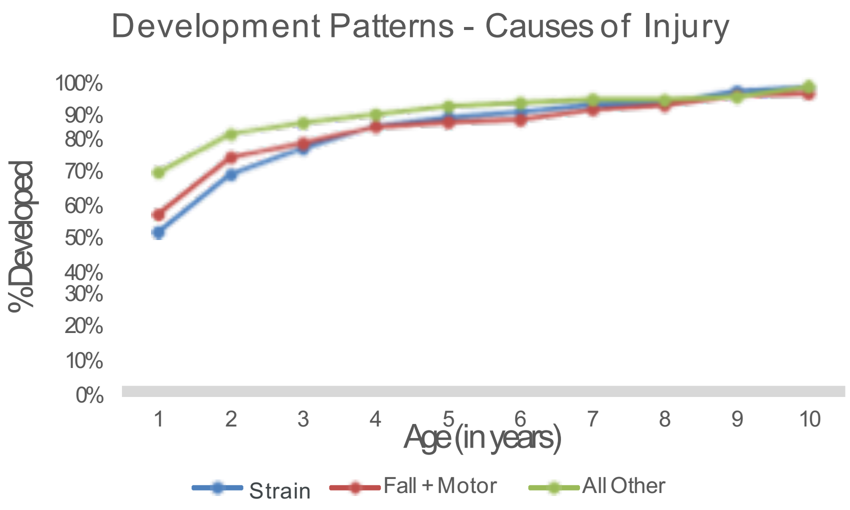


A different view of clustering results

Case Study: Evaluating the Triangles

- An evaluation framework – are the new patterns as expected?
  - Strains – slow development
  - Fall + Motor – faster
  - All Other – fastest

- An evaluation framework – a new A vs E:
  - Compare results from aggregate
  - Against results from granulars
  - Granular may not always be better

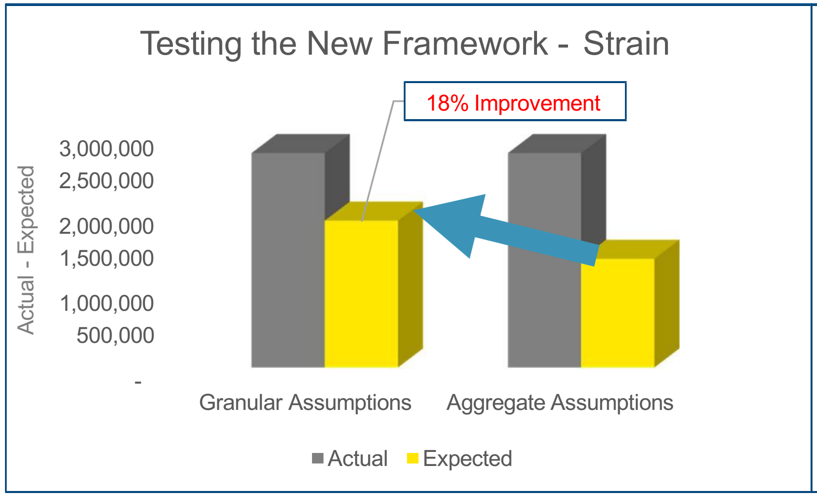


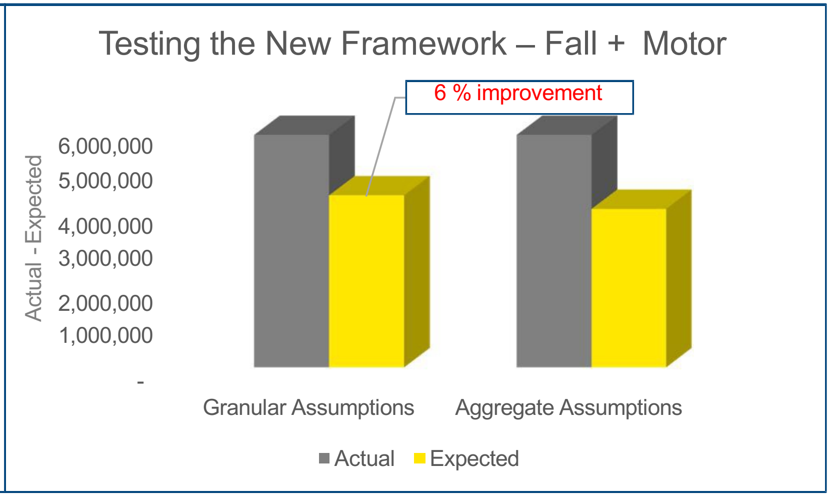
Case Study: Actual vs Expected

- An evaluation framework – a new actual vs expected:
  - A vs E using LDFs from aggregate vs granular triangles
  - It is not necessary that the new granular triangles will always explain the development better than aggregate
  - The success mantra – use granular assumptions where they explain A vs E better. Use aggregate assumptions otherwise

Case Study: Actual vs Expected

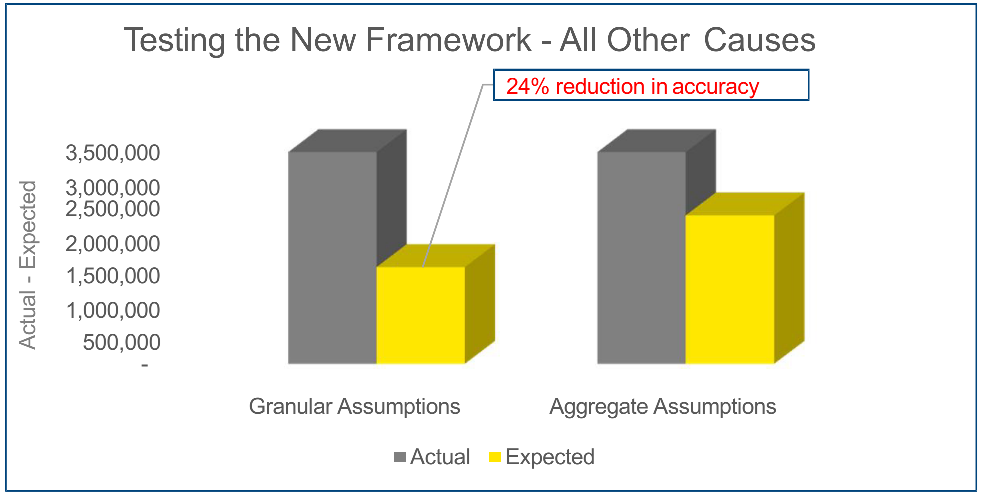
- "Strain", "Fall + Motor" claims - granular assumptions are better





Case Study: Actual vs Expected

- All Other causes of injury – aggregate assumptions are better

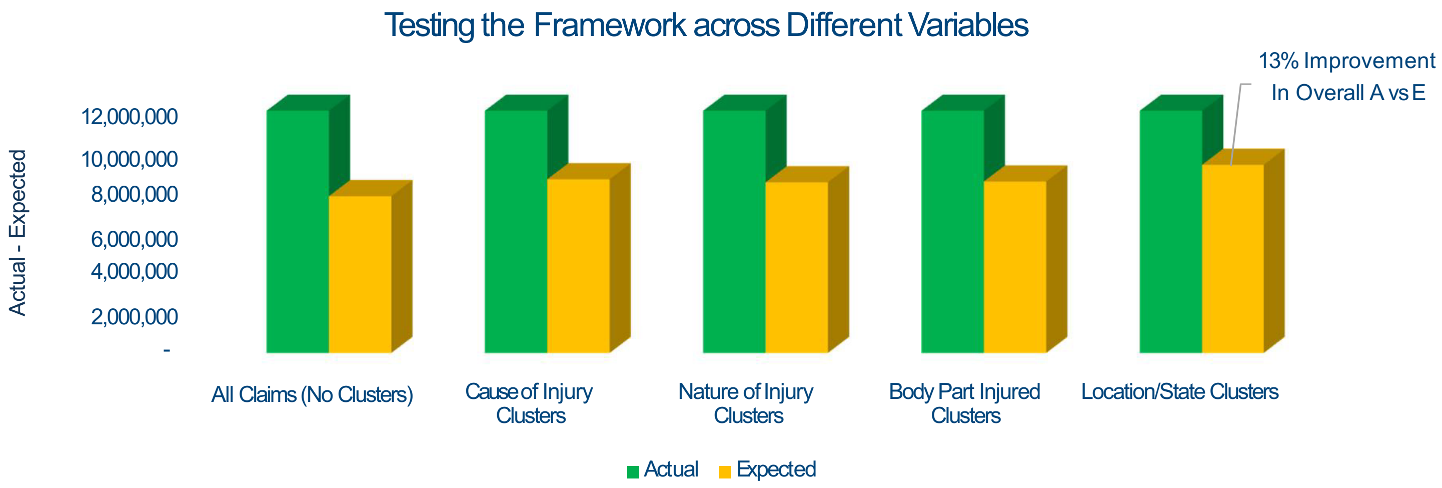


Case Study: Clustering and Chain Ladder

- Conclusions – establishing a comprehensive testing approach:
  - The success of using this machine learning framework lies not only in using granular assumptions to improve results
  - But also in recognizing areas where granular assumptions fail, and reverting to aggregate assumptions as appropriate
  - Combined appropriately, the results are more accurate
  - Conclusions should not be drawn on the basis of one year alone – back-testing should be done over multiple years

Case Study: Clustering – Final results

- This framework was tested across all the 4 key variables –



Case Study: Clustering – Learnings

- Key learnings from this comprehensive case study and testing:

  - In some cases, it was better to remove some claims from the "All Claims" category (e.g. shoulder and knee injuries) – to make the "All Other" category more homogeneous, and apply the aggregate patterns to shoulder and knee injuries

  - Success mantra – <u>the key to success is to learn from failure</u> – apply granular assumptions only where it improves results over multiple tests, apply aggregate assumptions otherwise

Learnings - Optimizing the use of Big Data

- As demonstrated, the use of big data does not necessarily require a complete overhaul of classical reserving techniques

- Insurance companies should not use granular assumptions to overwrite their current reserving processes – but rather to validate current processes, and explain the A vs E movements

- Framework should be implemented alongside current process

- As improvements emerge over quarters, insurers can switch to consistently using this framework for appropriate classes

Conclusions: Implementation of framework

- The gradual incorporation of big data into classical reserving techniques will be facilitated by machine learning algorithms –

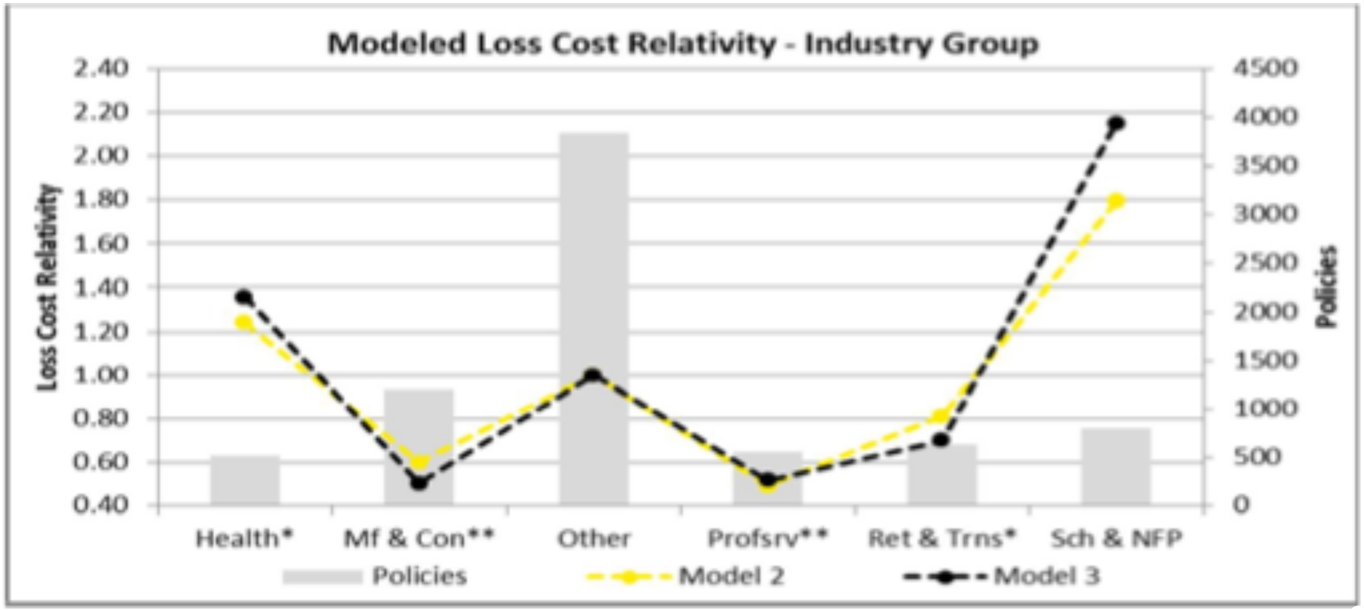
  - First to identify key variables for clustering/segmentation – GLMs, Random forests, etc. to measure variable importance

  - For example: key variables in motor liability would include – type of accident, type of injury, litigation, location/state, etc.

  - And then, use of clustering algorithms to segment aggregate data appropriately into homogeneous segments of data

Use Case 1: Clustering for B-F methods

- Use of the framework for selecting a-priori LR assumptions:
  - Group exposures with similar risk characteristics
  - Improve a-priori loss ratios
  - Critical for immature years
  - Clustering dimensions:
    - Loss Ratio
    - Frequency



Other Use Cases: For Actuarial Methods

- This Machine Learning framework can similarly be used alongside other classical reserving techniques such as:

  - Frequency-severity: Cluster exposures to improve a-priori frequency and severity selections (consider large loss propensity as a dimension for the clustering algorithm)

  - Stochastic Reserving: Cluster claims and exposures that exhibit similar characteristics of variability to get a more appropriate view of reserve risk and 1-in-200 scenarios

Other Use Cases: For Actuarial Functions

- This framework for granular assumptions should be used alongside other techniques for actuarial functions such as:

  - Capital Modelling: As discussed above, the machine learning framework can improve the calculation of reserve risk – this will also help improve calculation of diversification benefits (because all claim / exposure types do not behave similarly)

  - Pricing: Classical actuarial techniques used in pricing (e.g.: Burning Cost method) can be improved using this framework

Conclusions: Application of ML in Reserving

- Key principles to keep in mind for implementation –
  - IFRS 17 requires reserving at a more granular level
  - Selection of the right variables for segmentation is important
  - Need to select the right dimensions for clustering algorithms
  - Should not use machine learning algorithms as a “Black Box”
  - Support this framework using visualization of data & results
  - Need to use a blend of aggregate and granular assumptions
  - Need to use right software in the production environment

Institute and Faculty of Actuaries

Acknowledgment

Satraajeet Mukherjee &

AMJ GLOBAL SOLUTIONS

Contact:
Mobile: +447843412449
Email: asif@amjglobalsolutions.com

21 June 2018

Questions

Comments

The views expressed in this presentation are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation.

The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the IFoA.