Institute and Faculty of Actuaries

London Market Pricing Actuaries, Data Science Affects You Too…

Buu Truong
Mark Lee

Insight Risk Consulting

08 June 2016

Contents

- Introduction
- Scarce data
  - An actuarial dilemma
  - Communicating uncertainty
  - Clustering for commonality
  - Classification can help
- Unlocking data
  - Data augmentation
  - Cleaning data.

Institute and Faculty of Actuaries

Introduction: We live for data

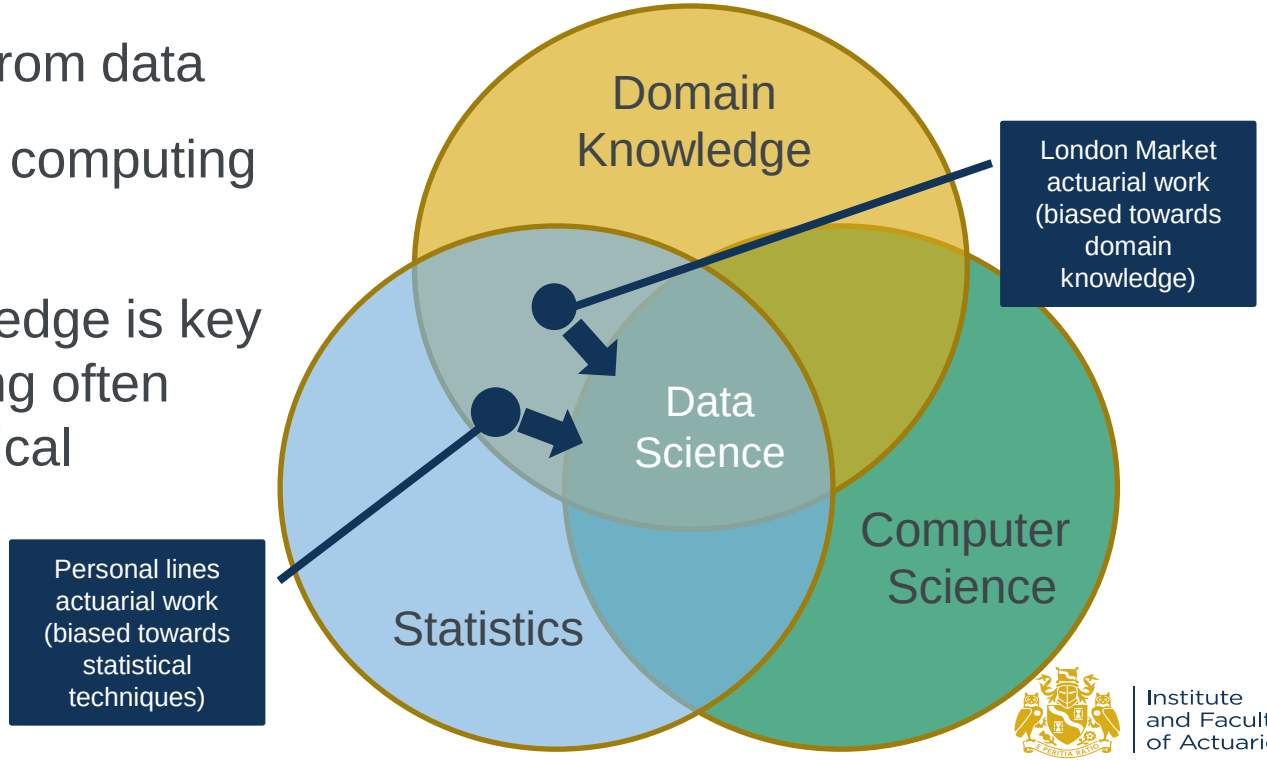
- Actuaries use and gain insights from data; it is where we add value!
- Two types of data problem:
  - Scarce data
  - Unlocking data
- Data science and statistical techniques can help.
- We cheated (a little) and are broad with our definition of data science.
- A theme for this presentation:

"Understand what your data can and can not help you with…"

Institute and Faculty of Actuaries

Introduction: Are we not data scientists already?

- Gaining insight from data
- Utilising modern computing resources
- Statistical knowledge is key (machine learning often known as statistical learning).



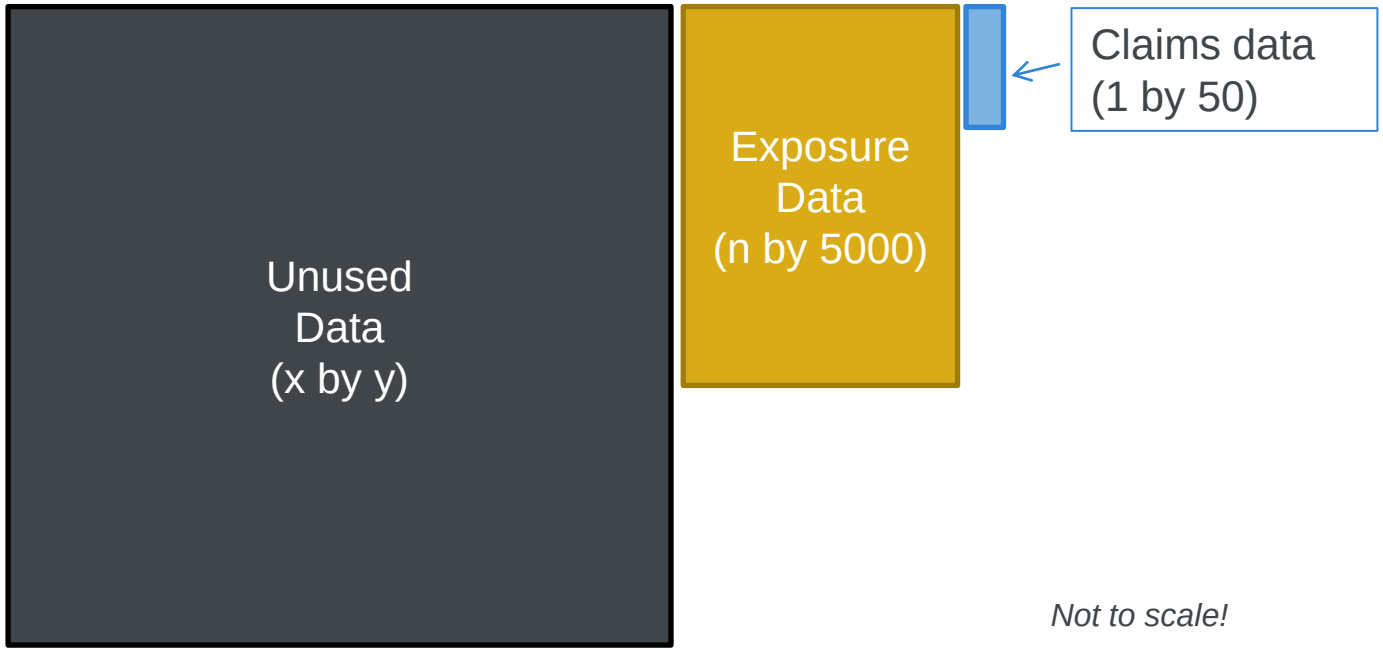
Institute and Faculty of Actuaries

Introduction: Data science

- Data science ≠ Big data!
- Data science can be used to generate predictions, understand your data and present your data.
- If you have enough data for a GLM, then *supervised regression machine learning* will be of interest – exciting opportunities for personal lines pricing in particular.
- Today, we assume that you have too little data for a GLM, and focus on other *supervised classification machine learning*, *semi-supervised/ unsupervised machine learning* and *statistical approaches*.

Institute and Faculty of Actuaries

Introduction: London Market data



Not to scale!

Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma

- Pricing actuaries need to make a number of necessary assumptions when parameterising any model from historical data.
- These are often not explicitly discussed and we have included an example for consideration.
- The parameter error conversation is improving but it is still relatively immature compared with in other disciplines.
- This is ‘just statistics’, but respecting data is a core competency of both actuarial and data science.

Institute and Faculty of Actuaries

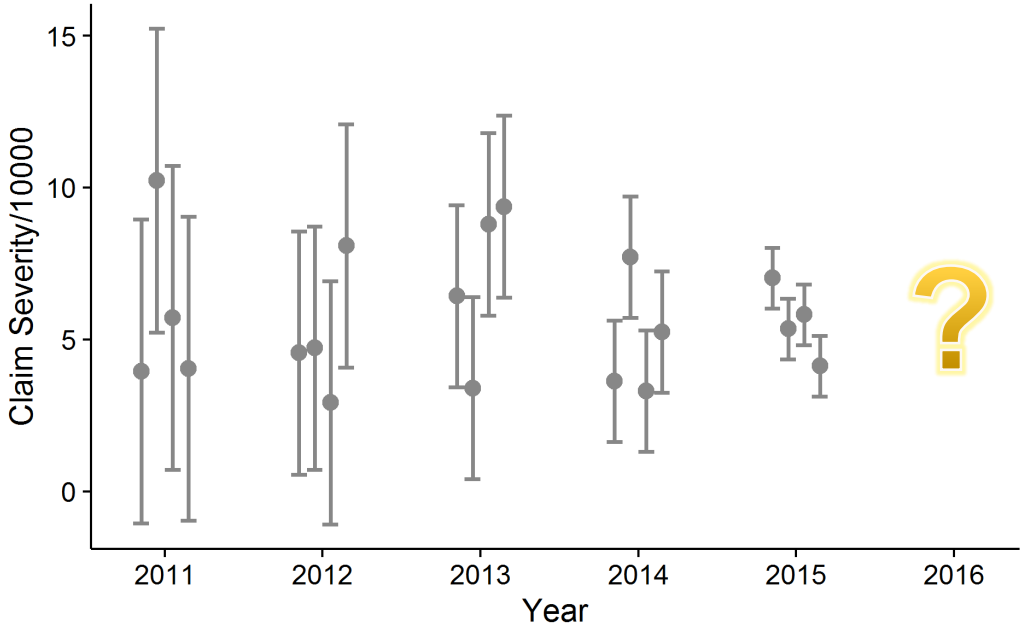
Scarce data: An actuarial dilemma

- A typical severity curve parameterisation exercise (the untold story):
  - We on-level historical data assuming we know about underlying trends e.g. inflation. *We do not recognise in our prediction the uncertainty around the on-levelling process.*
  - As there is limited data, we don't carve out some data for later testing. *As there is no hold-out data set we are likely to over-fit to the data.*
  - We choose a curve to enable stochastic sampling later on. *This parametric method reduces parameter error but increases model error. This approach improves transparency.*
  - We discard some significant outliers which distort our prediction. *This may be reasonable but it is difficult to be certain and it's often material.*

Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma

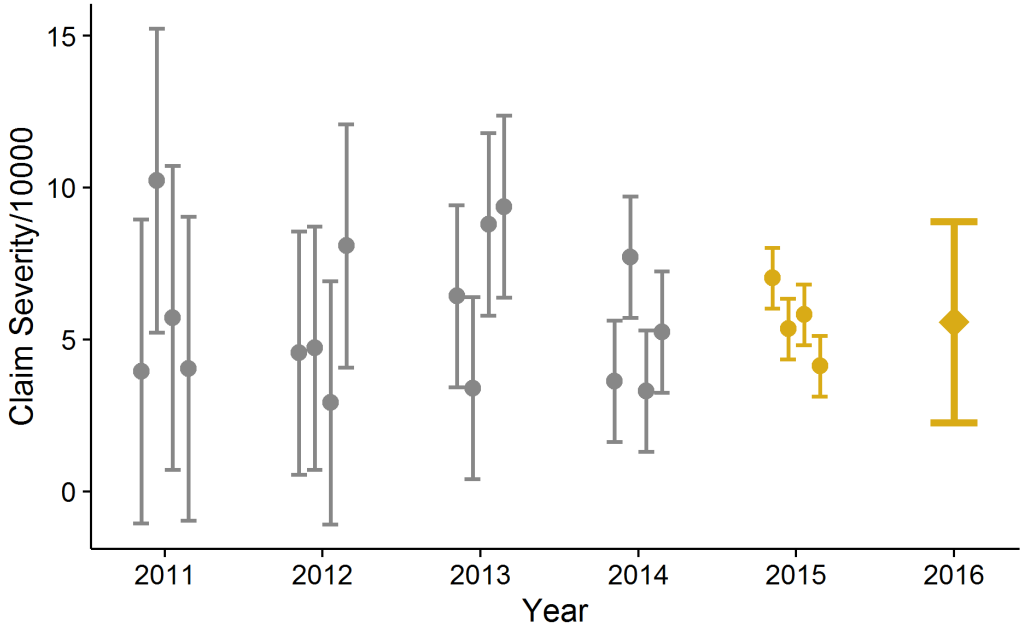
- Chart of data and parameter uncertainty around the on-levelled data
- Shows the impact on the parameter error of excluding historical data - itself subject to uncertainty.



Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma

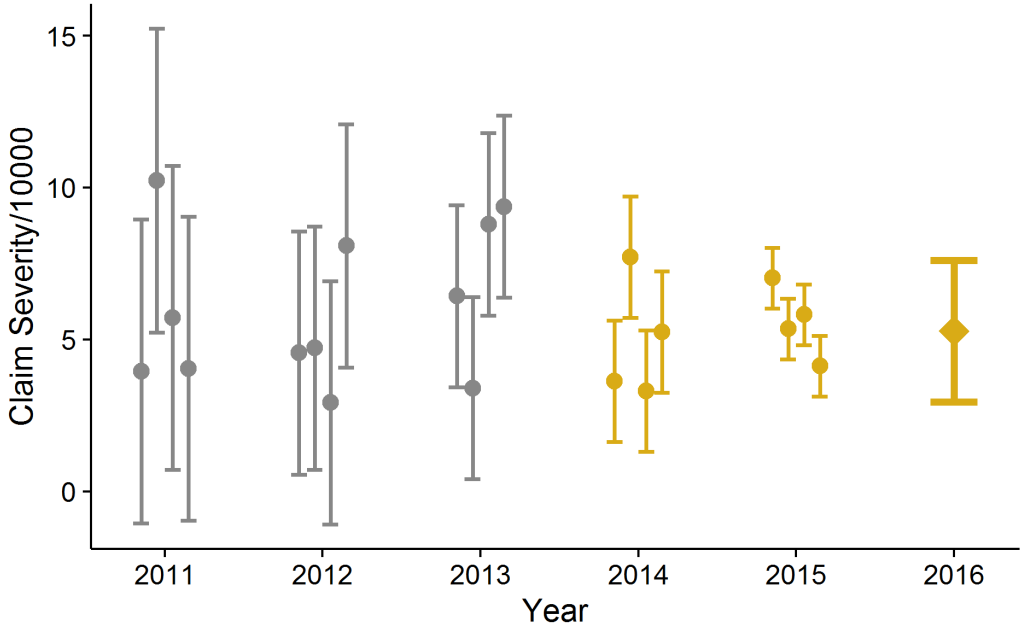
- Chart of data and parameter uncertainty around the on-levelled data
- Shows the impact on the parameter error of excluding historical data - itself subject to uncertainty.



Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma

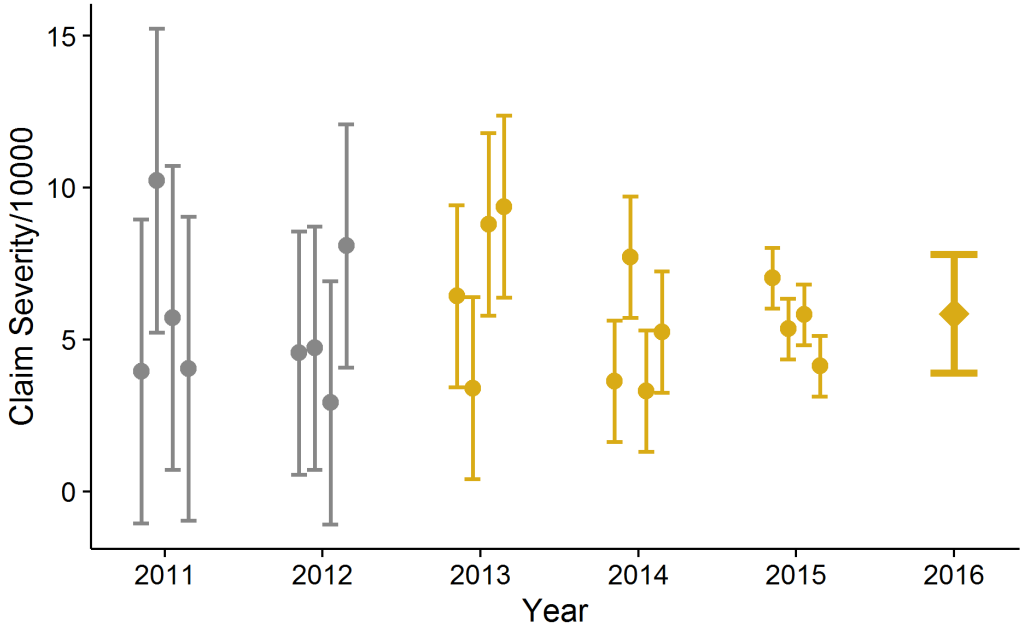
- Chart of data and parameter uncertainty around the on-levelled data
- Shows the impact on the parameter error of excluding historical data - itself subject to uncertainty.



Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma

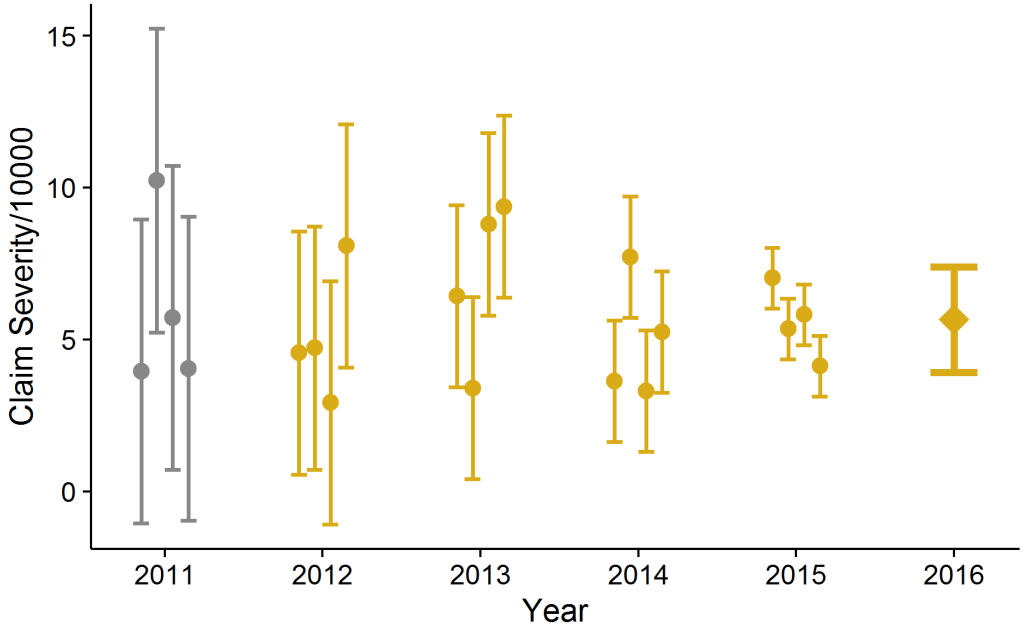
- Chart of data and parameter uncertainty around the on-levelled data
- Shows the impact on the parameter error of excluding historical data - itself subject to uncertainty.



Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma

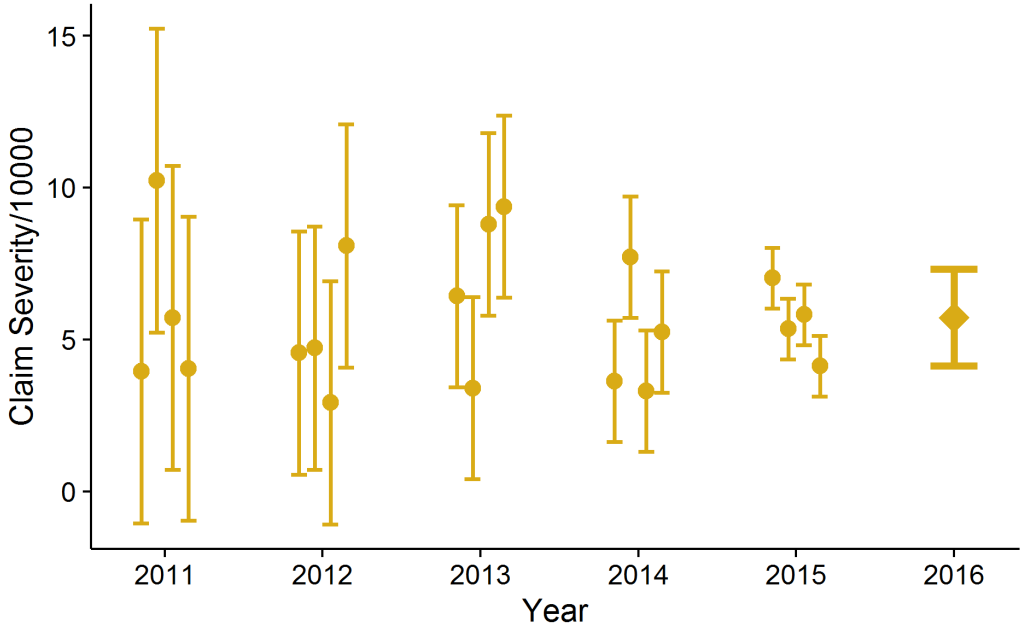
- Chart of data and parameter uncertainty around the on-levelled data
- Shows the impact on the parameter error of excluding historical data - itself subject to uncertainty.



Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma

- Chart of data and parameter uncertainty around the on-levelled data
- Shows the impact on the parameter error of excluding historical data - itself subject to uncertainty.



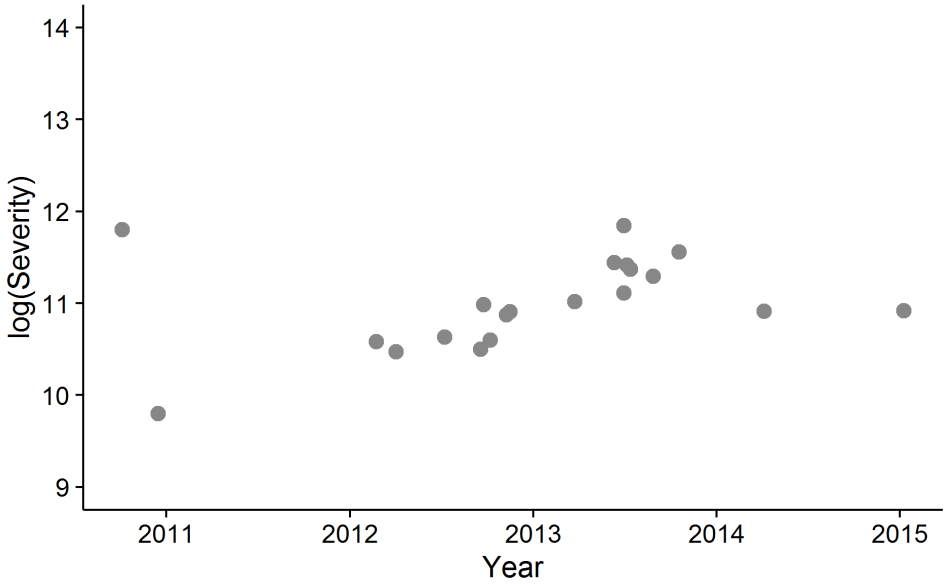
Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma

- Major pitfall with scarce data is over-fitting; this is fitting to the natural noise in the data.
  - Parametric models less likely to over-fit than flexible machine-learning algorithms (e.g., trees, forests and neural networks) and they utilise domain knowledge.
  - Regularisation of parametric models helps reduce over-fitting.
- Small data sets are more susceptible to outliers:
  - Standard statistical techniques – number of standard deviations from the mean, and other more advanced methods (median deviation of the medians)
  - Clustering algorithms or projection methods – principal components analysis
  - Use robust regression methods (down-weight the influence of outliers compared to ordinary least-squares regression).

Institute and Faculty of Actuaries

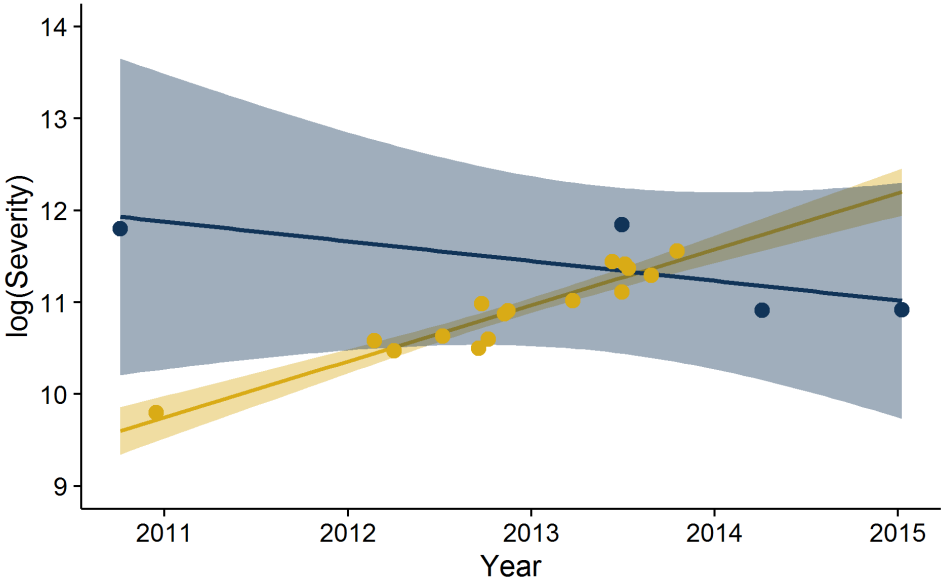
Scarce data: An actuarial dilemma



- Do your outliers represent a different risk?
- This data could have been extracted from an insured that had two distinct types of business, of which the first is no longer ongoing. This approach has many potential uses.

Institute and Faculty of Actuaries

Scarce data: An actuarial dilemma



- Do your outliers represent a different risk?
- This data could have been extracted from an insured that had two distinct types of business, of which the first is no longer ongoing. This approach has many potential uses.

Institute and Faculty of Actuaries

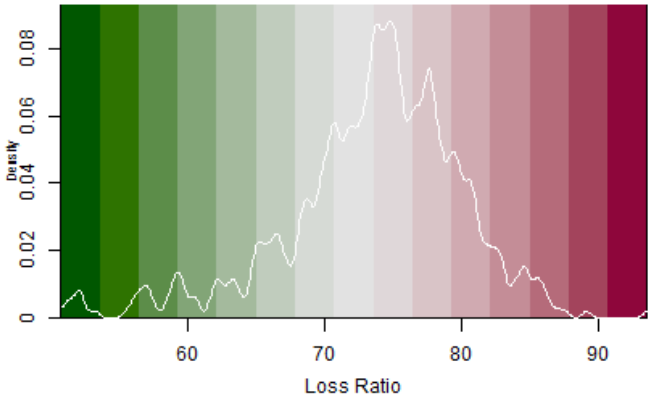
Scarce data: Communicating uncertainty

- As actuaries, we should effectively communicate the uncertainties in our work. Ignoring underlying variability and over-fitting lead to over-confidence.
- There are approaches to better understand uncertainty without discarding data:
  - Cross-validation – unbiased but tends to overestimate variance
  - Bootstrap methods – better variance characteristics
  - Bayesian methods – use your prior expectations.
- We should price for this uncertainty through a risk load…

Institute and Faculty of Actuaries

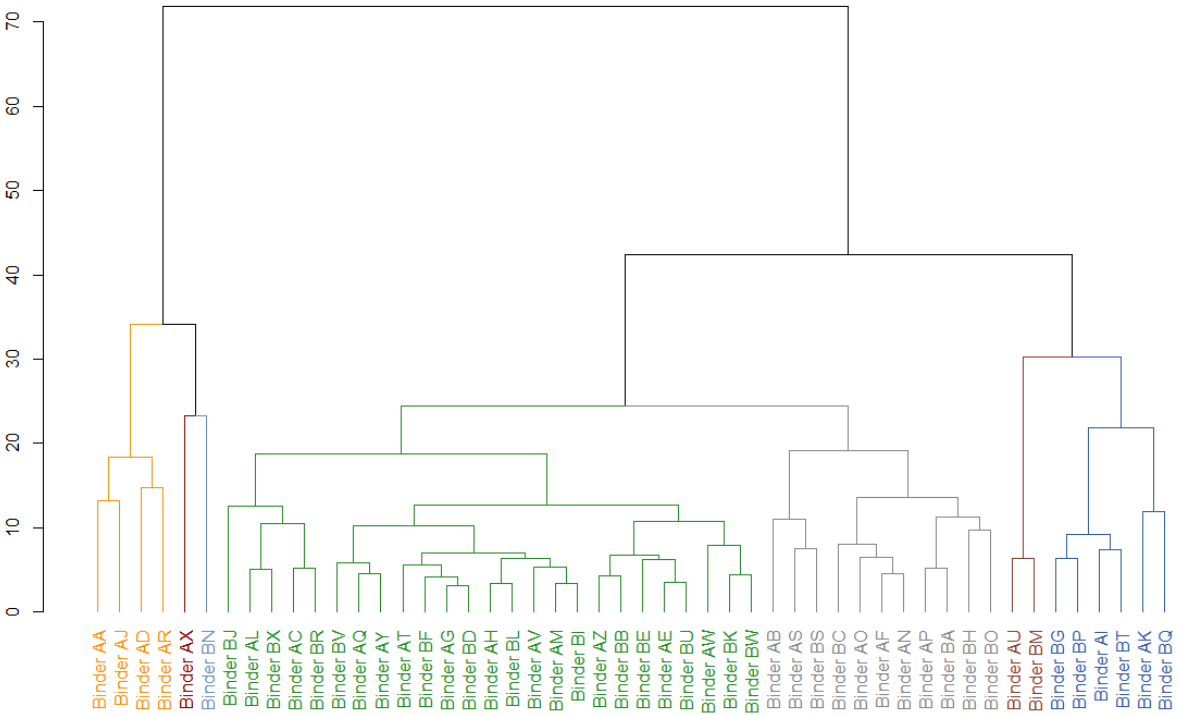
Scarce data: Clustering for commonality

- Clustering is an unsupervised technique that allows the machine to find patterns and relationships that may not be easily visible.
- Our example has over 50 binders each with 11 years of loss ratios.
- Use Hierarchical Cluster Analysis (HCA) to group similar performances over time.



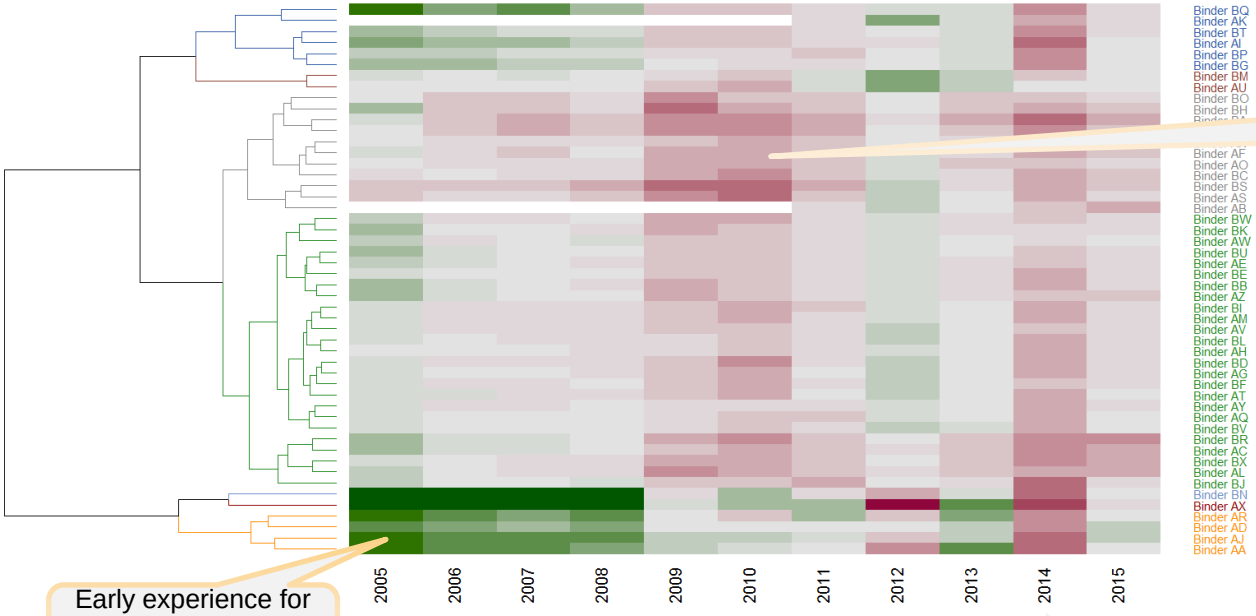
Institute and Faculty of Actuaries

Scarce data: Clustering for commonality



Institute and Faculty of Actuaries

Scarce data: Clustering for commonality



2009 and 2010 were poor years for this group – why?

There are perhaps 7 groups. What do they have in common such that we can name the groups?

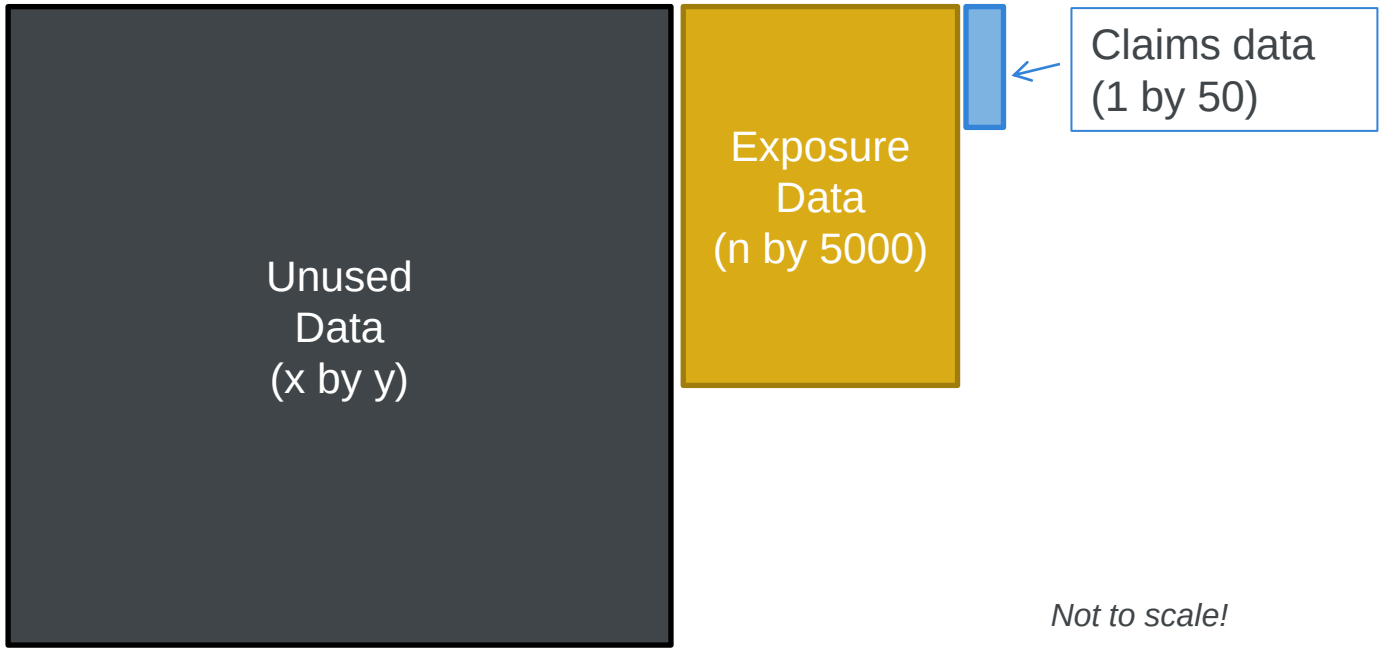
Semi-supervised methods can extend this unsupervised method, and help put labels of "good", "okay" and "bad" to your clusters.

Early experience for this group is positive.

2014 is a poor year across all groups

Institute and Faculty of Actuaries

Introduction: London Market data



Not to scale!

Institute and Faculty of Actuaries

Scarce data: Classification can help

- Given data limitations, London Market techniques often classify data and then adjust rates as a secondary process. Focus is therefore on ‘classification methods’ and not ‘regression methods’.
- Many rating factors in London Market pricing are categorical factors with ‘one-way’ relationships with price.
- Whether to create consistency or create benchmarking, classification methods can help rate risks based on similar characteristics.

Institute and Faculty of Actuaries

Scarce data: Classification can help

Following example might be employer's liability. The rating method is:
Price = Exposure*Base*Modifier (or 400=1,000,000*0.0005*0.8).
where 0.5 per mille is the rate per exposure and 0.8 is the factor for the low modifier.

The data features are not directly used having been replaced by the underwriter modifier.

Exposure	Feature A	Feature B	Feature C					Underwriter's Modifier
1,000,000	45	A	3					Low
900,000	43	S	3					Low
1,200,000	42	C	2					Low
2,000,000	36	C	3					High
1,500,000	19	D	4					Medium
700,000	38	B	3					High

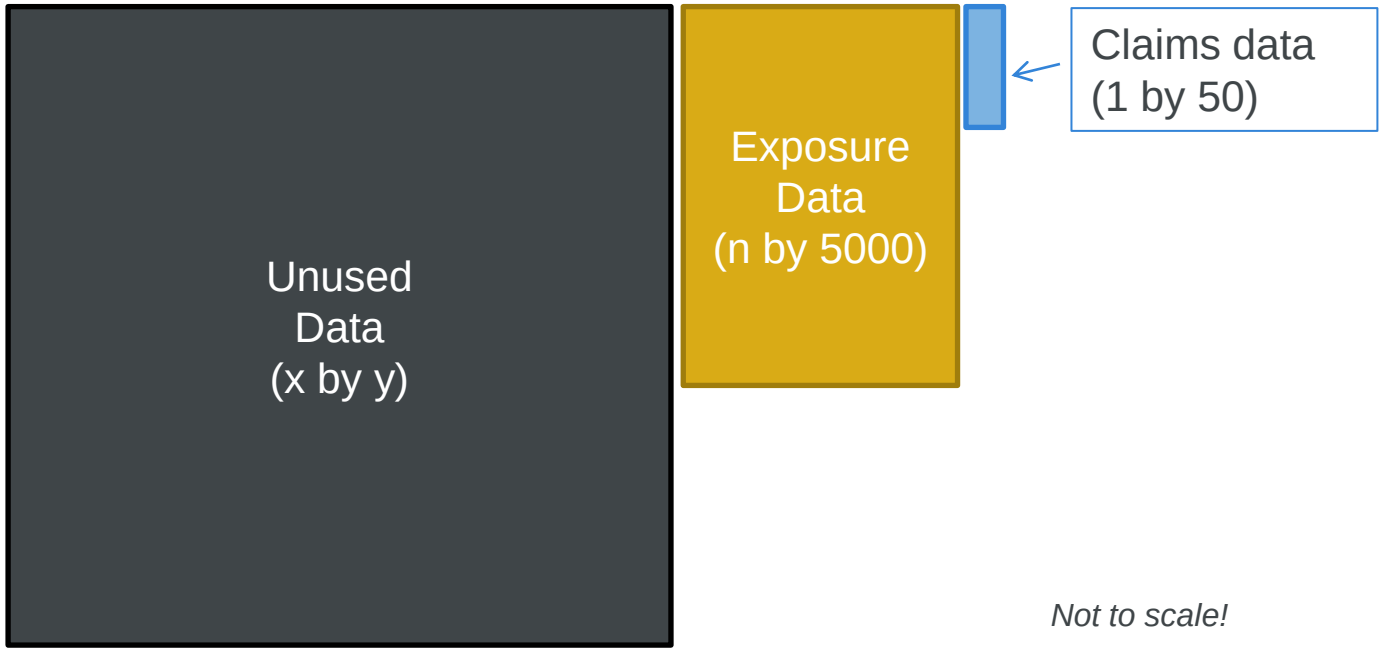
Scarce data: Classification can help

- How does the underwriter assign category of the modifier? Can a machine learn the categories from the other data?
  - Automate the process for initial categorisation of new business
  - Challenge the underwriter for consistency
  - Can the process be extended to parameterise the modifier?
- A ‘confusion matrix’ outlines a summary of method performance:

Prediction \ Actual	Low	Medium	High
Low	598	13	24
Medium	16	303	42
High	14	17	445

Institute and Faculty of Actuaries

Introduction: London Market data

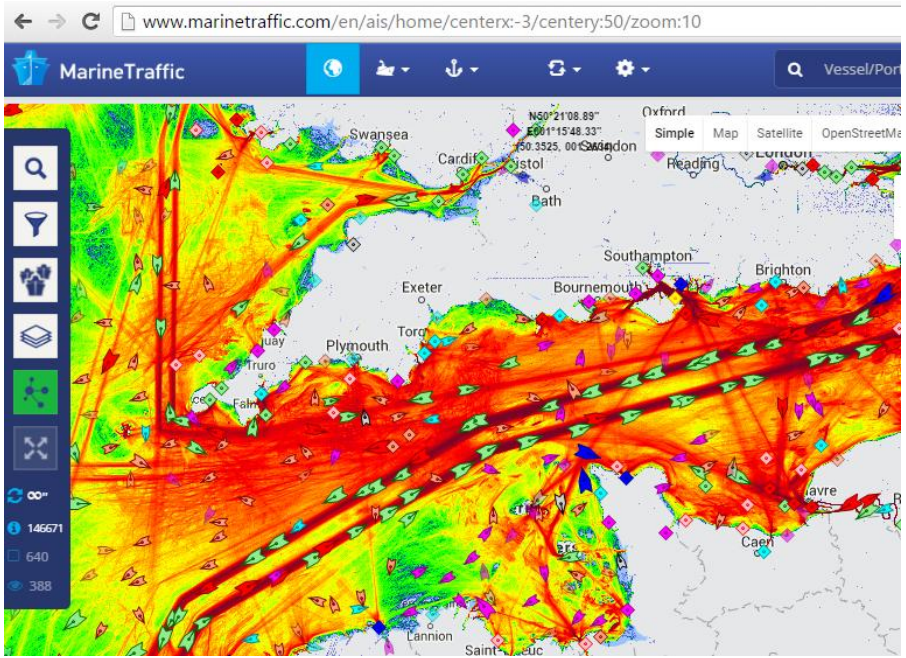


Not to scale!

Institute and Faculty of Actuaries

Unlocking data: Data augmentation

- There exists real time and historical vessel location data.
- New information has the potential to radically improve pricing. This really is big data.
- This can apply across any industry. Think of the data that must exist in pharmaceutical companies that might help product liability.



Institute and Faculty of Actuaries

Unlocking data: Data augmentation

- Slightly further afield is soft underwriting information in broker presentations.
- Unstructured data is a growing field where ‘machines’ are getting better and making this type of data usable.
- Company financial information could be a useful predictor of claims experience. This might be the equivalent to the way credit scoring impacted the motor market.

Institute and Faculty of Actuaries

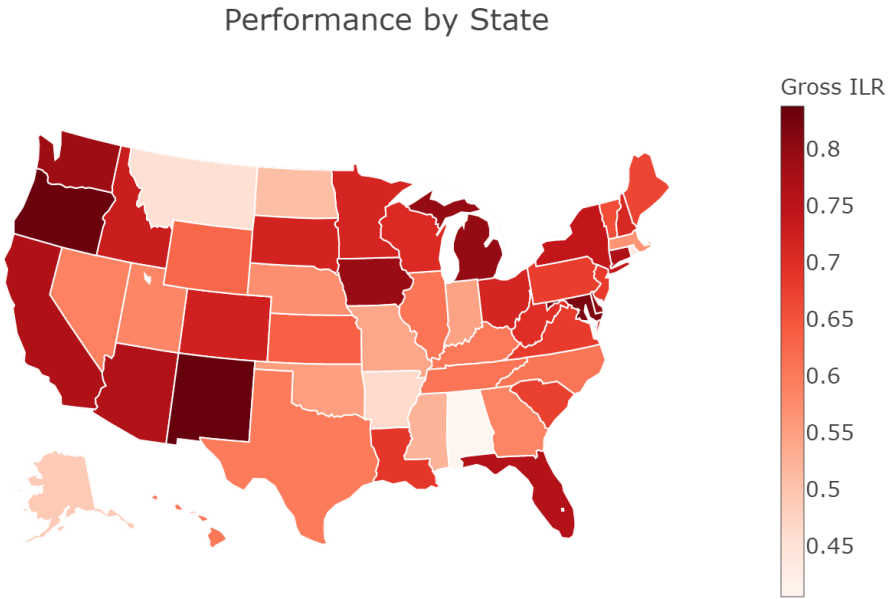
Unlocking data: Cleaning and presenting data

- Using some of the methods discussed already, data science techniques can help fill gaps in data or create benchmarks.
  - Policy level data could be used to estimate benchmark rate changes.
  - Catastrophe models need assumptions to fill in missing data; data science can add more rigour to this process.

Institute and Faculty of Actuaries

Unlocking data: Cleaning and presenting data

- Presenting data has never been easier. Powerful and accessible visualisation tools are now available.



Institute and Faculty of Actuaries

How to learn more

- We have some options for you:
  - I’m curious but will ask a question over coffee without 100 people listening.
  - I’ve a pressing question, and I’ll ask it now.
  - I’m interested – let’s have a coffee to discuss in depth.
  - I’m really interested – I want to do your data science course and be a data science guru.
- What type of data science problem is this?

www.insightriskconsulting.co.uk

Institute and Faculty of Actuaries