Taylor Fry

Risk pricing for Australian Auto

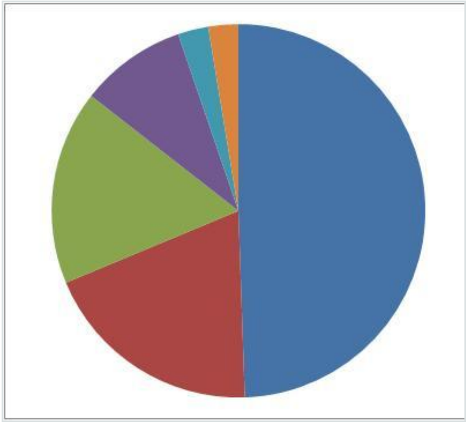
October 2011

Contents

Contents

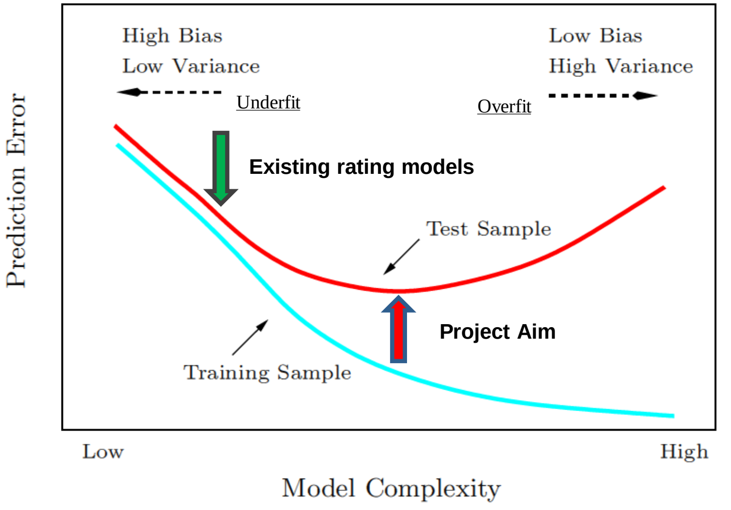
Scope

- Portfolio
  - Several brands, all Australian states
  - 9 million policy years worth of exposure
- Brief
  - Build a "state of the art" risk pricing model
- Coverage
  - All brands, all states
- This presentation concentrates on at-fault collision frequency



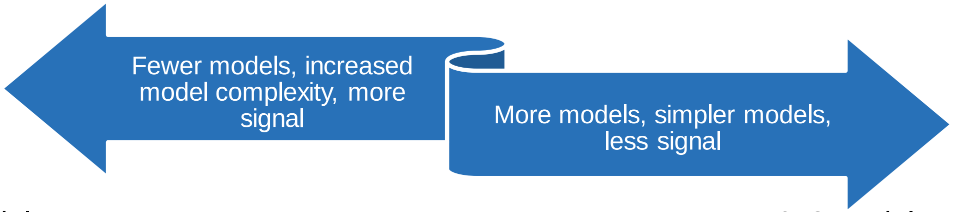
Modelling philosophy

- Our aim when determining how complex our model should be, is to minimise Prediction Error in Test/Holdout Datasets



From Hastie, Tibshirani and Friedman

How many models?



- 2 models
  - Frequency
  - Size
- Size model potentially very complex
  - Many different perils
- Frequency model
  - Has to incorporate all peril specific predictors

- 672 models!
  - Close to existing practice
  - Very hard to maintain and update
- Each model is "simple", but
- Low signal
  - And unnecessary structure /parameter error

How many models?

Fewer models, increased model complexity, more signal

More models, simpler models, less signal

- 2 models
  - Frequency
  - Size
- Size model potentially very complex
  - Many different perils
- Frequency model
  - Has to incorporate all peril specific predictors

- 16 models
  - Frequency
  - Size
  - 8 perils
- Models complex
  - Need to incorporate brand/state interactions
- More signal
  - But does it justify the increased complexity of each model?

- 672 models!
  - Close to existing practice
  - Very hard to maintain and update
- Each model is "simple", but
- Low signal
  - And unnecessary structure /parameter error

Contents

Large universe of potential predictors

Policy Attributes

- Standard policy rating factors
  - Age, Gender, value,…
- Additional Driver Information
  - Number of drivers, age, family group, accident flag…

Policy Related History

- Policy and customer tenure
- Past motor claims history for each policy
- Other product holdings and claims history from same customer/address

Vehicle Related Data

- Vehicle attributes from Glass's Guide
  - Make
  - Model
  - Engine size
  - PWR
  - Shape
  - Dimensions
  - Tyre width and profile
  - Value …

Location Based Data

- 2006 Census Data
- 3rd party socio-demographic segmentation
- Weather history

Partitioning and sampling

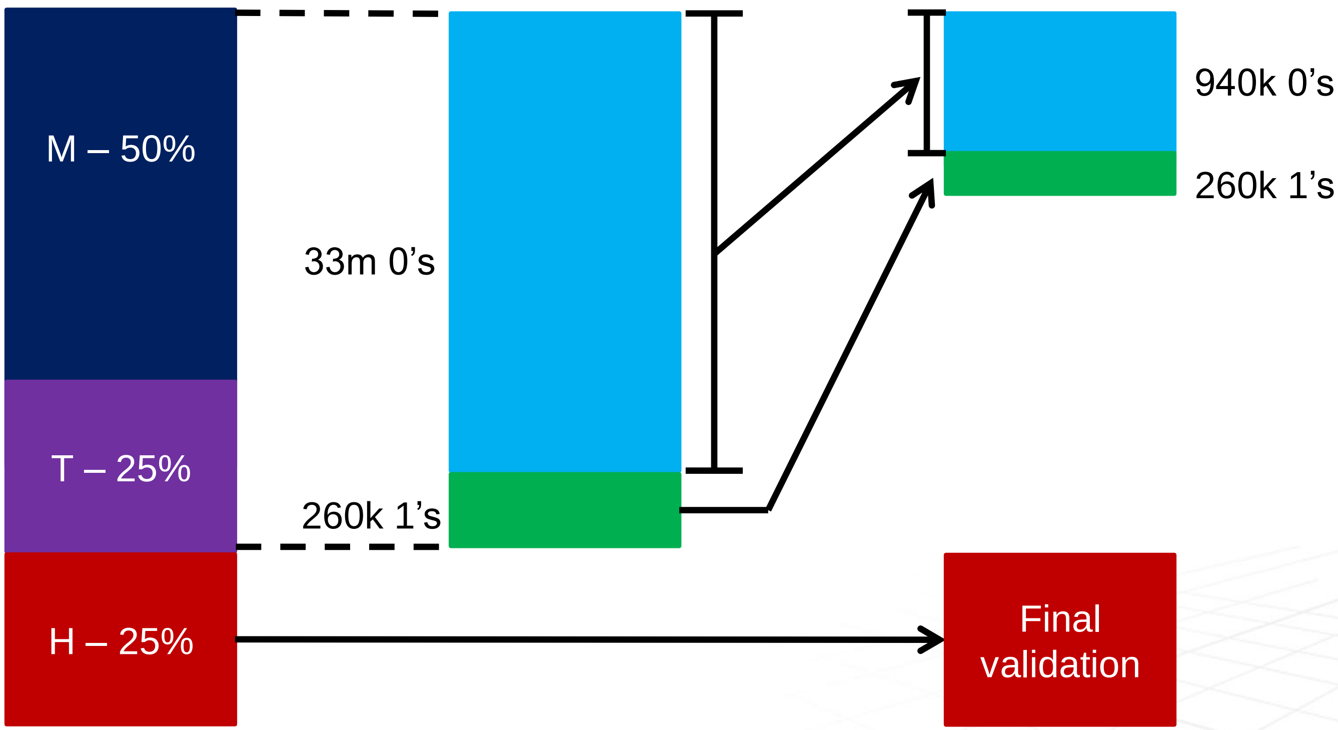
- 44M records, 3000 potential predictors
  - SAS dataset is 1400GB uncompressed!

- Need to reduce this to a size that we can deal with
  - Partition and sample "horizontally"
    - Learn, test and holdout sets
    - Keep all the "1"s but only some of the zeros

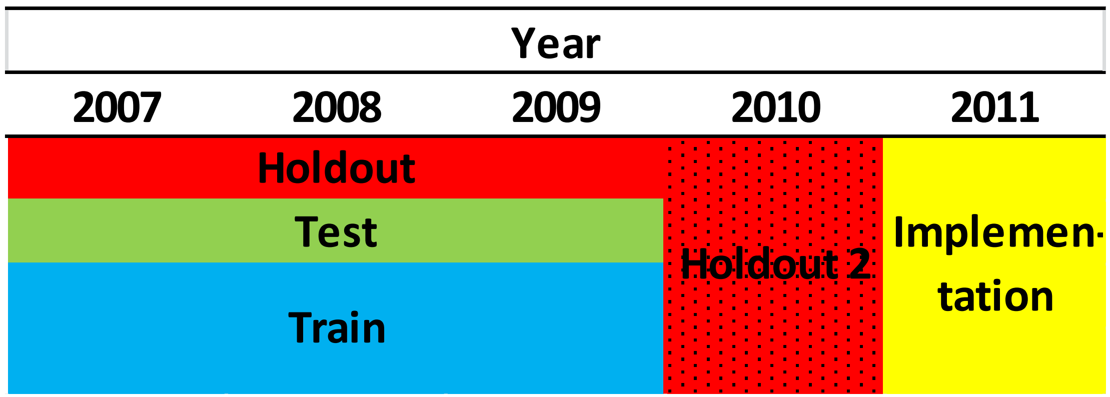
  - Partition and sample "vertically" for variable selection

- Have a strategy for time based testing

Partitioning and sampling - horizontal



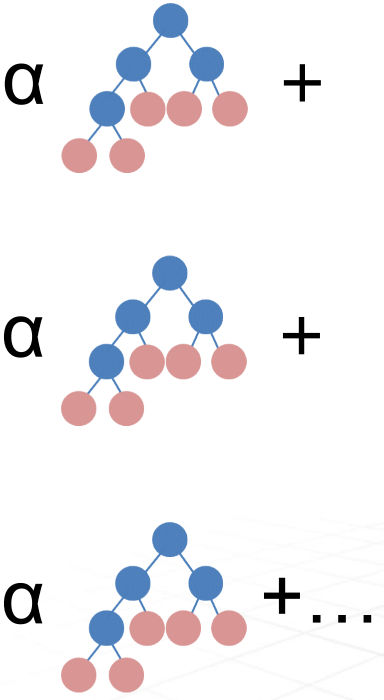
Time based testing



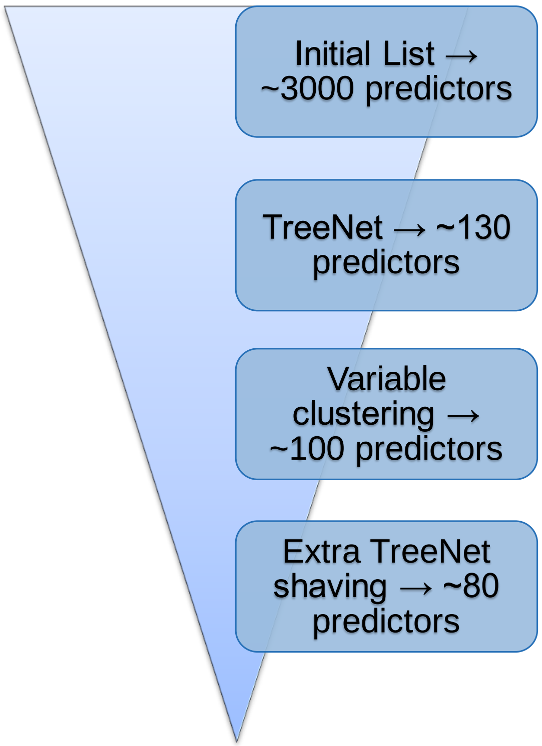
Contents

Filtering tool of choice - Treenet

- What is Treenet? Or MART or GBM
  - A gradient based boosting algorithm using small decision trees as the base learners
  - Performs both classification and regression, with various choices of loss functions
  - Falls into the general class of ensemble based predictive models
- Gives very good “out of the box” models which often take some time and effort to surpass, <u>but</u>
  - Maximum capacity (of Salford implementation) is about 300 variables and 1M observations
  - Overfits generally, and
  - Gives preferential treatment to high level categorical variables



Filtering – general process

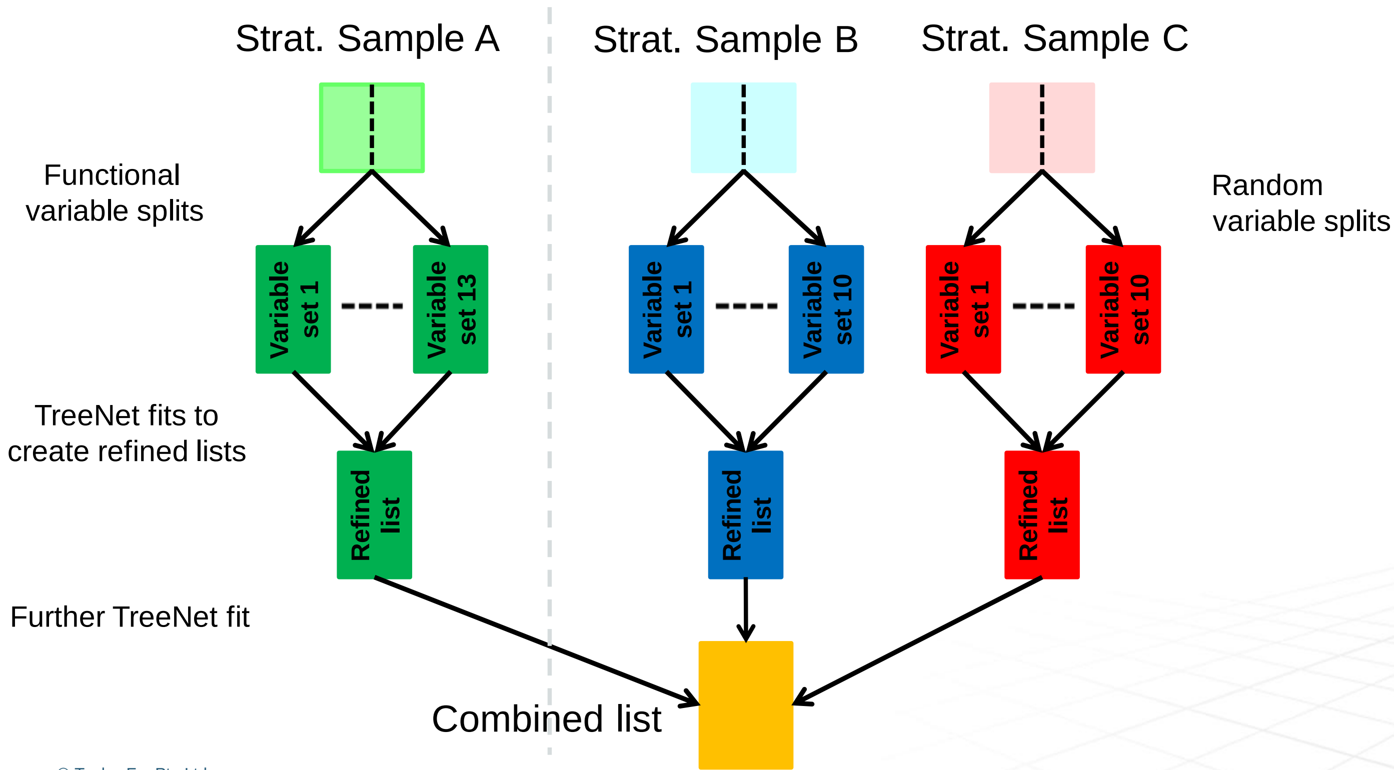


- Partition dataset (horizontal) and variables (vertical)
- For each variable partition
  - Main rating variables included in each partition
  - High level categorical variables grouped or excluded
- Fit Treenet models to each partition
- The best variables from each partition enter the “super group”
- Using Treenet, the super-group gets shaved from the bottom until performance on the test set peaks

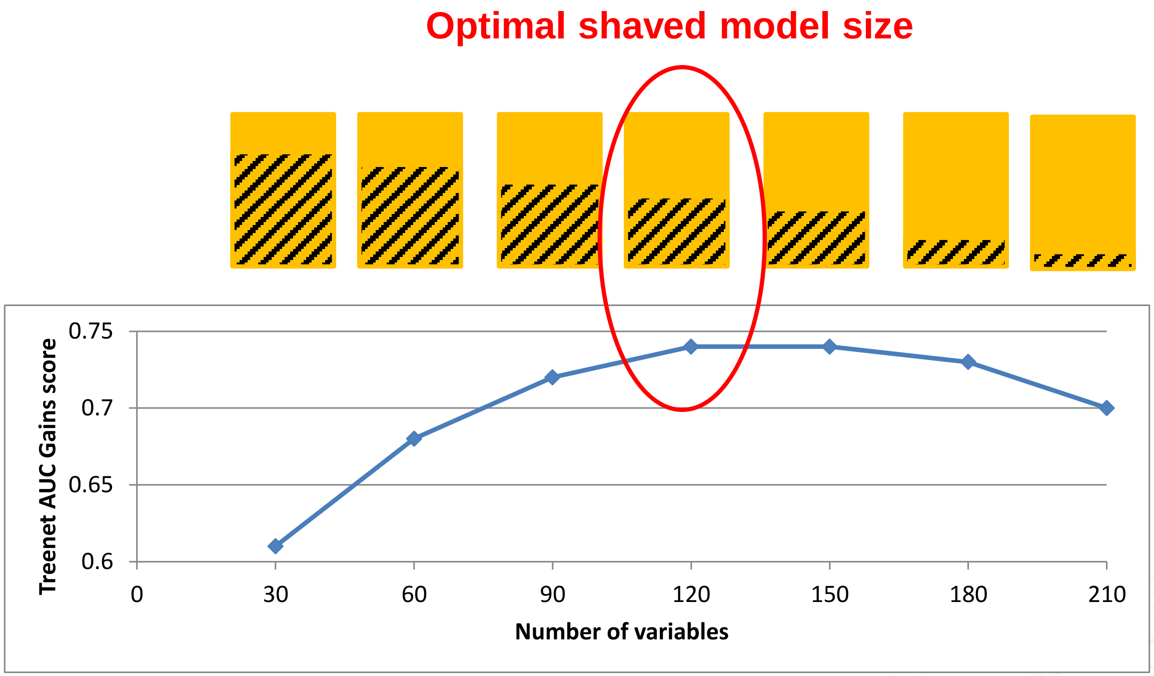
- Proc Varclus (SAS) used to discard correlated variables

- More Treenet shaving

Initial variable filtering



Filtering – variable shaving



Correlated variables

Problem

- Highly correlated variables are undesirable in GLMs:
  - Misleading parameters and significance
  - Odd shapes
  - Longer fit times
- Removal one of a pair of closely correlated variable generally has negligible impact on model performance

Solution

- Apply hierarchical variable clustering (Proc Varclus) to identify sets of correlated variables
- Keep one or two variables from each cluster – based on rank in TreeNet importance lists
- Rerun TreeNet on reduced variable list to check performance not materially degraded
- Produce correlation matrix to check that no major correlations remain

Contents

Modelling - techniques

- Machine learning models
  - Non-parametric
  - Easy to build
  - Adequate fit with little effort
  - Push button to update

  - Performance often not as good as good GLM
  - Can be over-parameterised
  - Little insight (except decision trees)
  - Recalibration can lead to large changes at the observation level

  - TF preferred for variable selection and interaction searches

- Generalised Linear Models (GLMs)
  - Structure, form, INSIGHT
  - Equal or better model with fewer parameters
  - More stable over time
  - Push button to update but with some care

  - Harder to build good models
  - Can be very hard if the structure is complex (but this is very rare)
  - May not pick all structural nuances

  - TF preferred for main model

Modelling - GLM fitting strategy

Fit “Saturated Model”, including key interactions

Find correct error structure (distribution and mean/variance relationship)

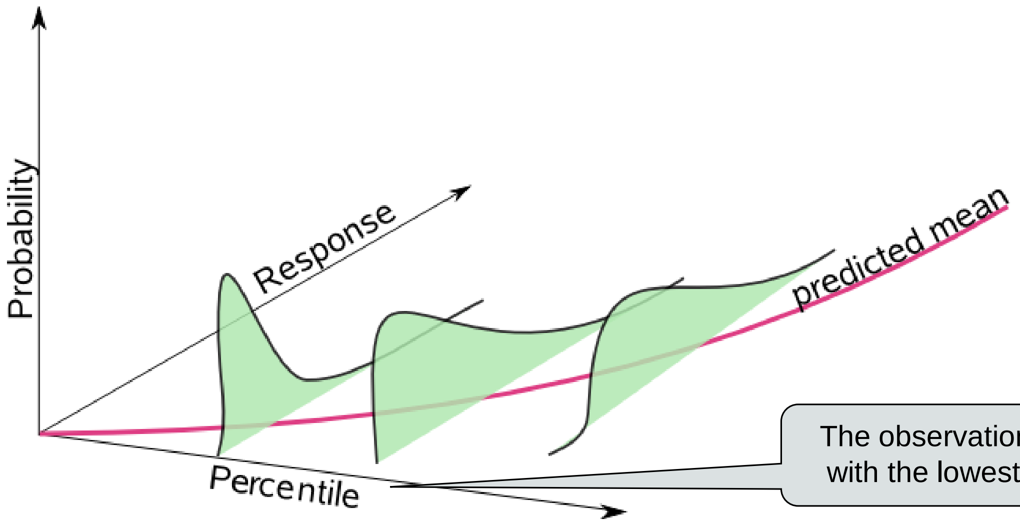
Simplify Continuous Predictors

Simplify Categorical Predictors

Search for more interactions

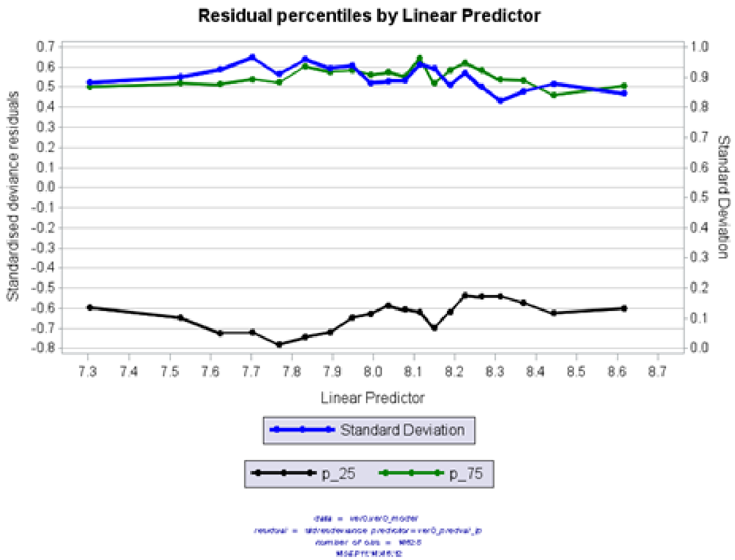
- All GLMs fitted in SAS using TF custom macros

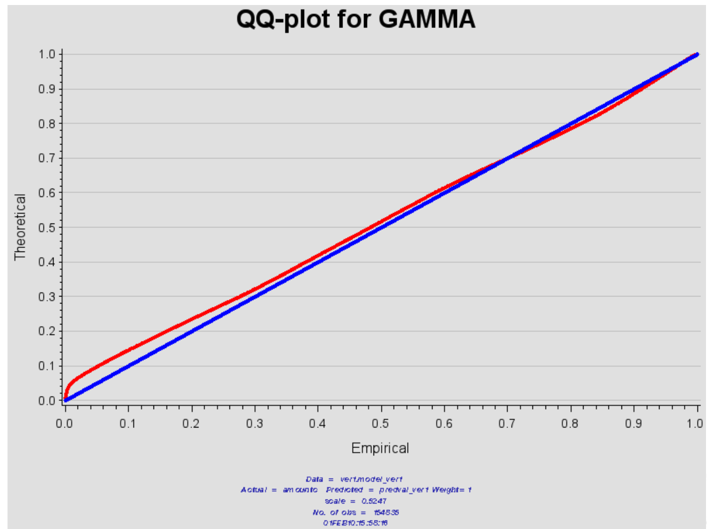
Mean/variance and distribution



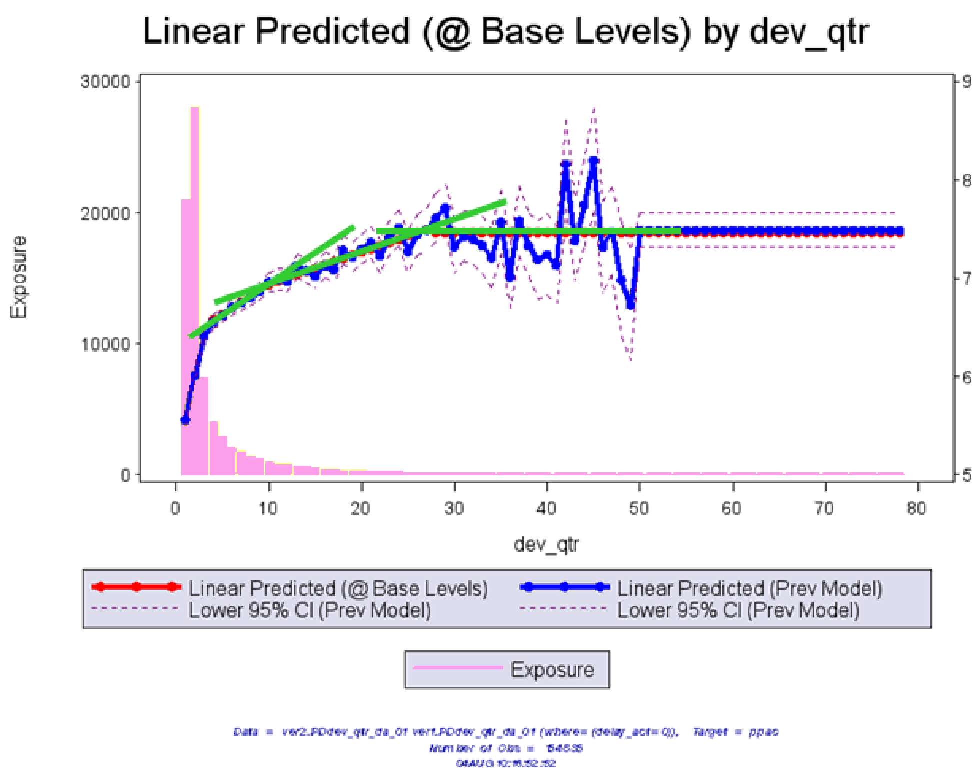
The observations are sorted with the lowest μ_i at the left

Variance diagnostics

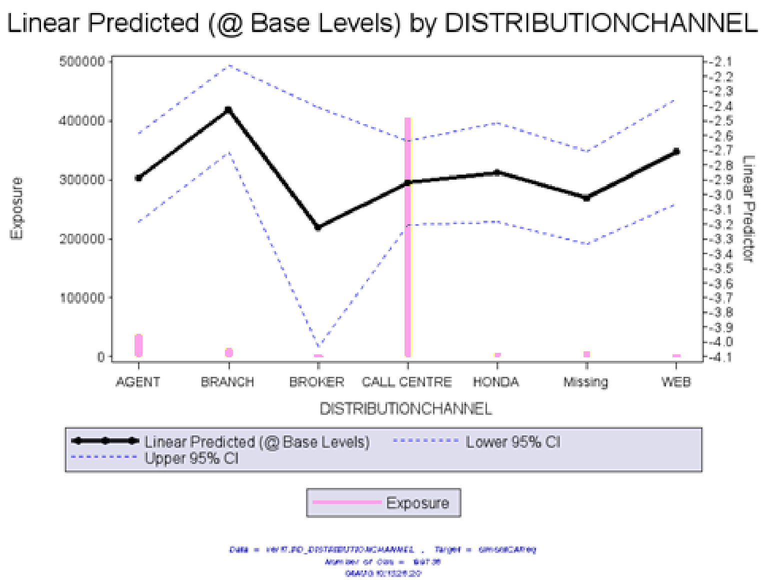




Simplifying continuous variables



Simplifying categorical variables



	pvalue						
	C_distributionchannel						
	AGENT	BRANCH	BROKER	CALLCENTR	HONDA	Missing	WEB
R_distributionchannel							
AGENT	.	.	.	.	.	.	.
BRANCH	0.00	.	.	.	.	.	.
BROKER	0.39	0.04	.	.	.	.	.
CALLCENTR	0.45	0.00	0.43	.	.	.	.
HONDA	0.70	0.00	0.35	0.43	.	.	.
Missing	0.08	0.00	0.61	0.13	0.11	.	.
WEB	0.13	0.01	0.20	0.06	0.31	0.02	.

Contents

Extensive search for interactions

Brute force search

- Main effects models are splines plus binaries
- Form list of all pairwise interactions (6,000!)
- Fit a GLM for each interaction, using main effects prediction as an offset
- Order using AIC and p-value (a bit unreliable)
- Fit candidates in main model for final evaluation

Testing suspected interactions manually

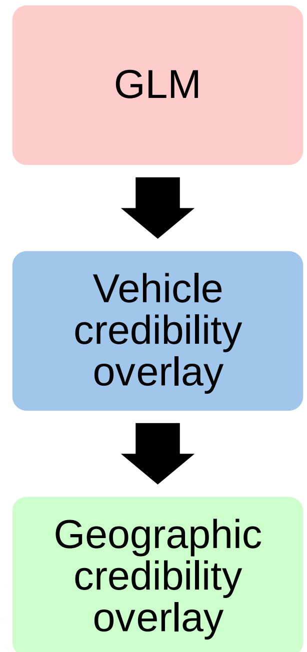
Machine learning insights

- A Treenet model with two node trees is a "main effects model". This can be used to evaluate the effect of interactions
- Fit Treenet models to residuals, view the model and use to test further interactions
- Also used rule fit

Contents

Vehicle and geographic overlays

- Final model
  - 54 predictors and 200 parameters, including
  - 100 state/brand interactions and 36 other interactions
- Vehicle characteristics and geography largely accounted for and integrated
- Still possible residual effects from high level categorical variables
  - Make, model and zone?



Credibility overlays

- Credibility overlays used to model residual vehicle and geographic variation

Vehicle Credibility Model

Make → Model → Variant

Geographic Credibility Model

Statistical district (SD) → Stat sub-district (SSD) → SLA

- Each level in the hierarchy has a relativity factor applied, representing a portion of the actual observed relativity.

1.1 Observed Relativity for Holden → (Portion depends on exposure and tuning parameter) → **1.04** Adopted Relativity for Holden

Vehicle credibility overlay - details

- Adaptation of approach suggested by Ohlsson, Scandinavian Actuarial Journal, 2008 (for log link)

$y^* = f_{make} f_{model} f_{variant} y$ $\qquad f_{make=k} = (1\text{-}z_k)+z_k r_k$ $\qquad z_k = w_k/(w_k + T)$

$w_k = \sum_k y$, for Poisson model

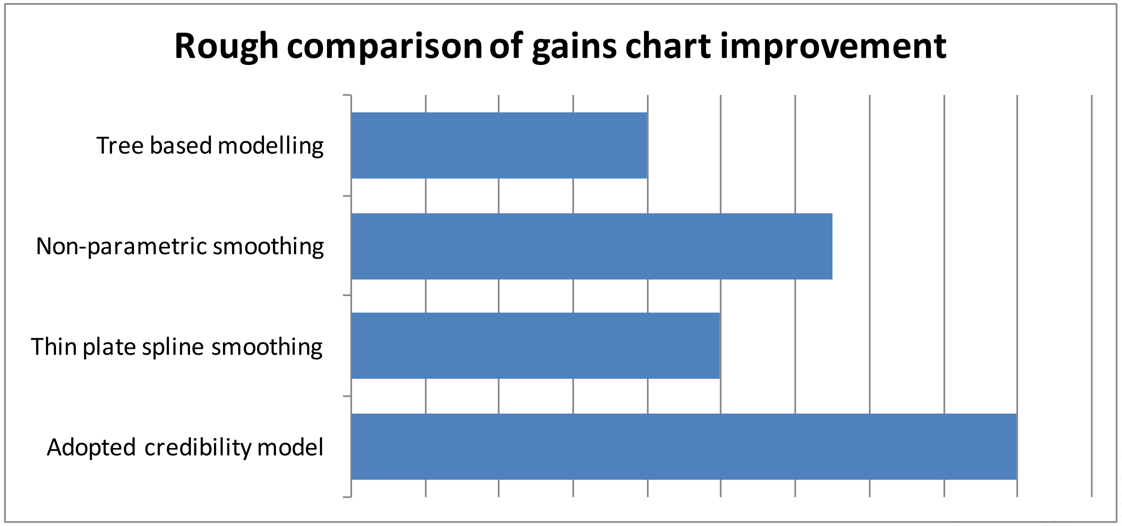
- Relativities are done sequentially, using test set to set optimal T (see other TF presentation at this seminar)

- Combined assigned relativities
  - Within +/- 15%
  - 95% within +/- 10%

Alternative geographic overlays

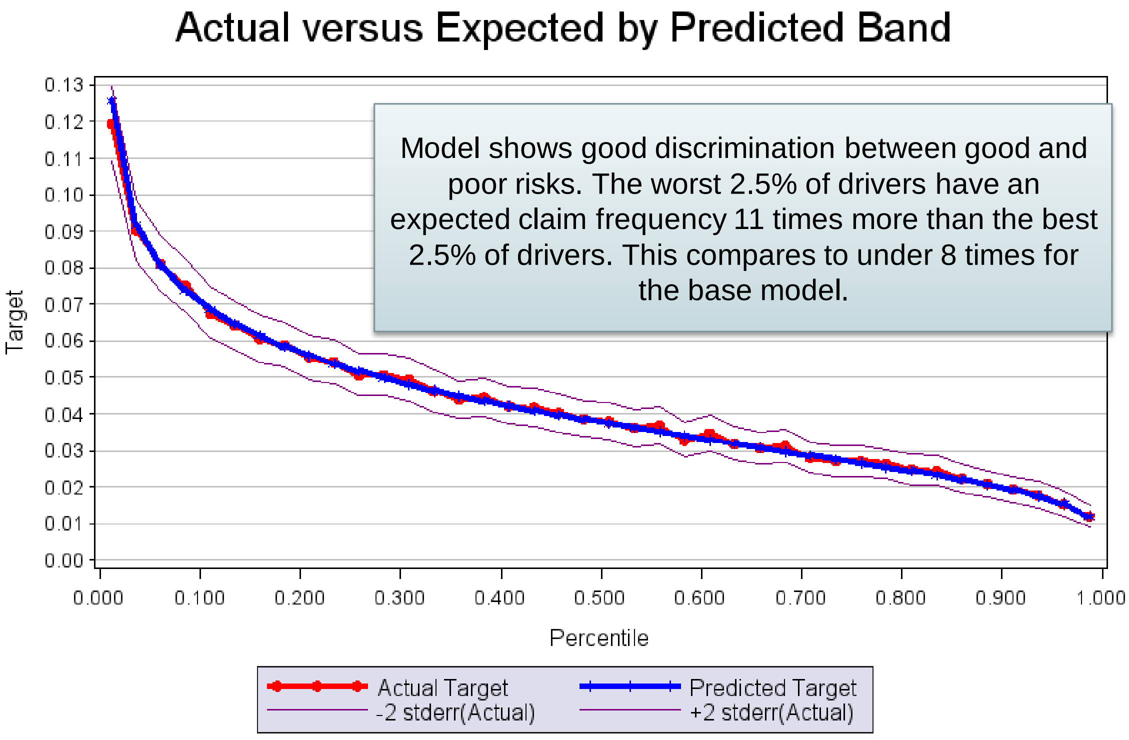
Tree based latitude/longitude modelling	• Taking the residual as the target, and latitude and longitude plus rotations as predictors, build a tree-based model. • Attempted using both random forests and treenet. The latter performed more strongly
Non-parametric kernel smoothing	• For a given location, form a prediction by taking the weighted average residual of the 500 nearest CCDs • Weights defined by a kernel function. More weight given to closest CCDs. • Number of CCDs chosen empirically • Postcodes (rather than CCDs) also trialled, giving inferior performance
Thin plate spline (Whittaker) smoothing	• Using latitude and longitude as predictors, fit a thin spline model to the average residuals by postcode. • Too computationally intensive to apply at a more granular level
Alternative credibility setups	• Other types of hierarchies and variables explored • Adopted the most successful

Alternative geographic overlays

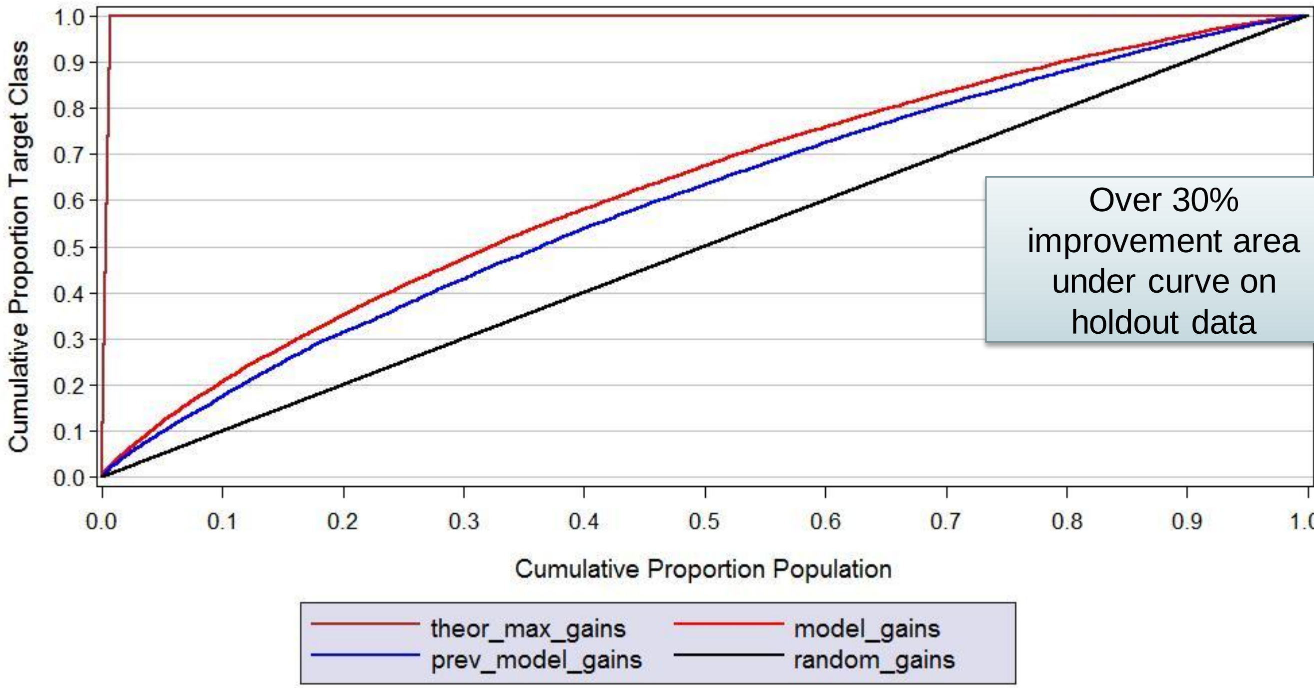


Contents

Overall fit diagnostic

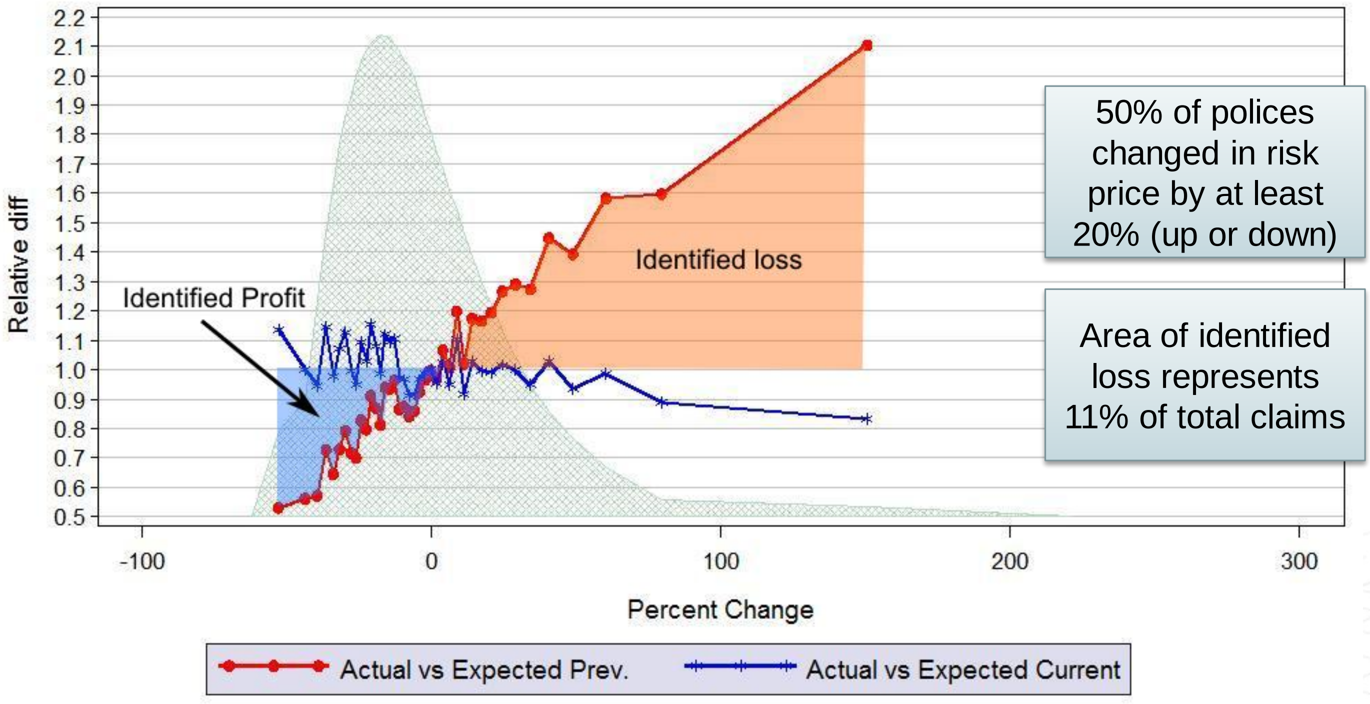


Comparison with recalibrated base



Over 30% improvement area under curve on holdout data

Comparison with recalibrated base

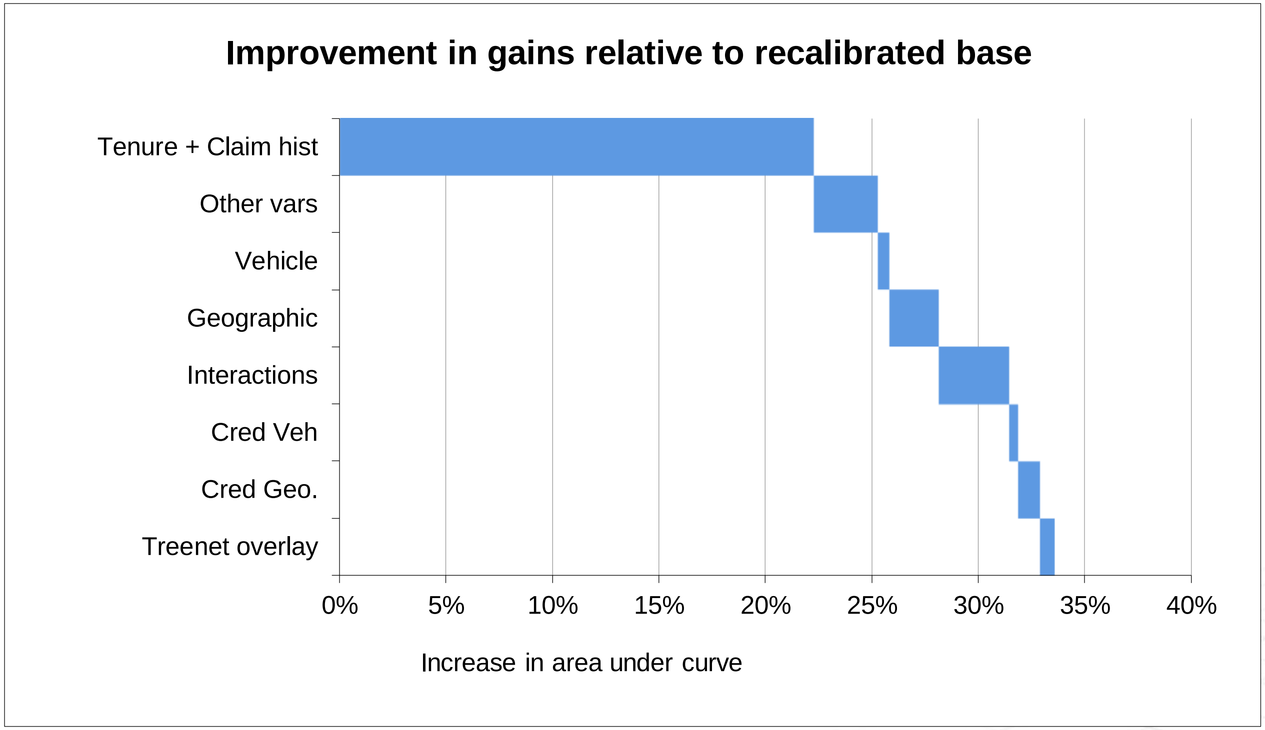


50% of polices changed in risk price by at least 20% (up or down)

Area of identified loss represents 11% of total claims

Contents

Sources of value



Contents

Final tests

Has the multi-brand, multistate structure cost predictive power?

- AvsE diagnostics by brand and state
- Comparison with a model fitted only on one brand one state
- All satisfactory

Has the separate peril structure cost predictive power?

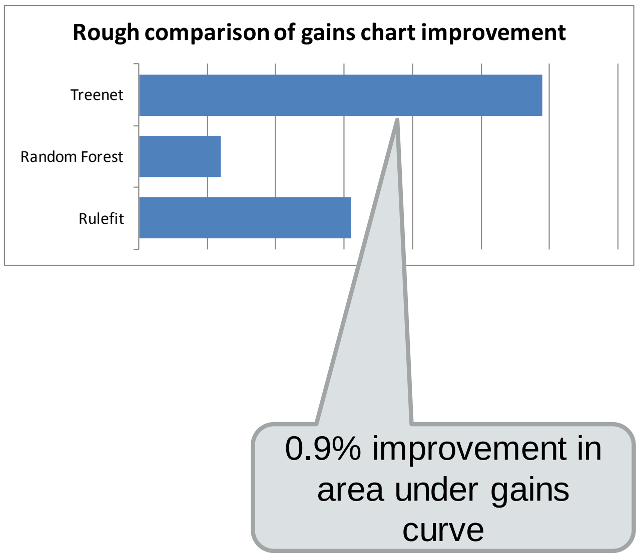
- Combine all models to get a single claim frequency and claim size
- Fit to residuals (take care with definition of residual) with Treenet and GLM
- Nothing compelling

Is there more to be gained?

- Compare final models to initial optimal Treenet models. Final models are better (but not by much)
- Fit to claim frequency and size residuals by peril to get "machine learning" overlay
- A little bit of extra performance

Machine overlays

Different machine learning overlays



Effect of Treenet overlay