Institute
and Faculty
of Actuaries

Self-assembling insurance claim models using regularised regression and machine learning

Gráinne McGuire
Taylor Fry

15 April 2019

Acknowledgements

Joint work by

- Gráinne McGuire, Hugh Miller (Taylor Fry)

- Greg Taylor, Josephine Ngan (University of New South Wales)

Outline of presentation

- Motivation
- Regularised regression and the LASSO
- Model specification
- Real data example
- Discussion
- Conclusions

Motivation

- We consider the modelling of claim data sets containing complex features
  - Where chain ladder and the like are inadequate
- When such features are present, they may be modelled by means of a Generalised Linear Model (GLM)
- But construction of this type of model requires many hours (perhaps a week) of a highly skilled analyst
  - Time-consuming
  - Expensive
- Objective is to consider more automated modelling that produces a similar GLM but at much less time and expense
- Note that we are not discussing stochastic case estimate type of models here – those that use individual claim characteristics to produce an estimate of the ultimate loss.
  - Our models mainly use accident, development and payment quarter effects

Institute and Faculty of Actuaries

Outline of presentation

- Motivation
- **Regularised regression and the LASSO**
- Model specification
- Real data example
- Discussion
- Conclusions

Institute and Faculty of Actuaries

Regularised regression and the LASSO

- Consider general GLM structure

$$y = h^{-1}(X\beta) + \varepsilon$$

Link function

Stochastic error (EDF)

- Regularised regression loss function becomes

$$L = -2\ell(y; X, \hat{\beta}) + \lambda \|\hat{\beta}\|_p^p$$

p-th power of L_p-norm

Log-likelihood

  - Penalty included for more coefficients and larger coefficients, so tends to force parameters toward zero
    - $\lambda \to 0$: model approaches conventional GLM
    - $\lambda \to \infty$: all parameter estimates approach zero
    - Intermediate values of λ control the complexity of the model (number of non-zero parameters)
  - Special case: $p = 1$, **Least Absolute Shrinkage and Selection Operator (LASSO)**

$$L = -2\ell(y; X, \hat{\beta}) + \lambda \sum_j |\beta_j|$$

Favourite ML technique of many - transparent, interpretable model

Institute and Faculty of Actuaries

LASSO: shrinkage and selection

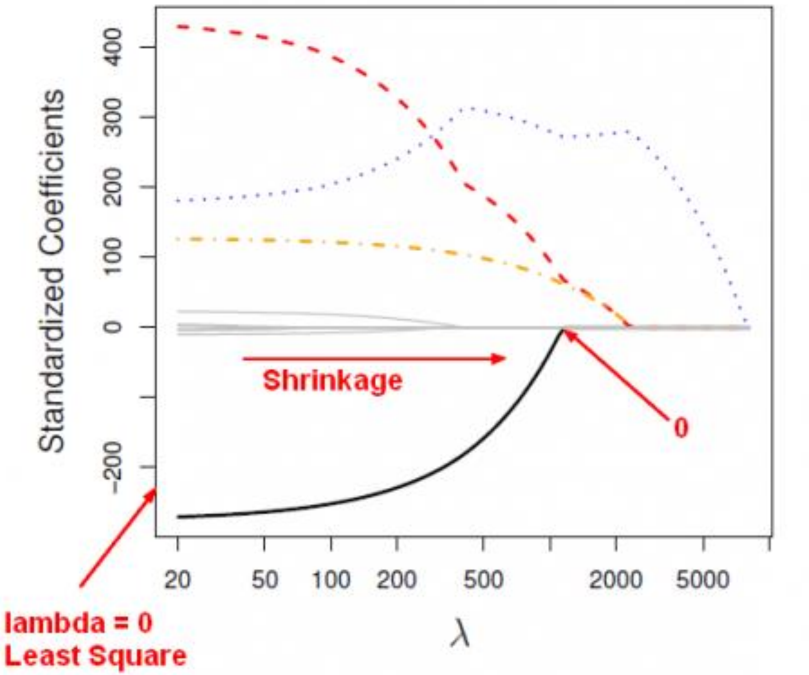


Image sourced from: https://gerardnico.com/data_mining/lasso

Outline of presentation

- Motivation
- Regularised regression and the LASSO
- **Model specification**
- Real data example
- Discussion
- Conclusions

Institute and Faculty of Actuaries

Model specification - formulation

- This is where nearly all the effort is – what predictors/regressors do we use?
- Consider a *n* x *n* triangle labelled by AQ, DQ, PQ
  - Regressors consist of set of basis functions that form a vector space:
    - All single-knot linear spline functions of AQ(*k*), DQ(*j*), PQ(*t*)
    - All 2-way interactions of Heaviside functions of k, j, t
  - AQ splines are
    - max(0, *k*-1), max(0, *k*-2), …., max(0, *k*-(n-1))
    - Similarly for DQ and PQ
  - AQ x DQ interactions are
    - I(*k*>1)*I(*j*>1), I(*k*>1)*I(*j*>2), ….., I(*k*>n-1)*I(*j*>n-1),
    - similarly for AQxPQ and DQxPQ
- Apply similar ideas if other variables available, e.g. operational time

Spline

Heaviside function

Potential collinearity in terms – we will come back to that later but remove/reduce if possible

Institute and Faculty of Actuaries

Model specification - formulation

- Hard part
  - Scaling!
  - $L = -2\ell(y; X, \hat{\beta}) + \lambda \sum_j |\beta_j|$

Regressors on different scales → Parameters on different scales → Influences selection

- Make standard deviations comparable?
  - Questionable here – we only have 3 fundamental regressors. Everything else is derived from these.
- Our approach:
  - **Base scaling on the original variables.**
  - **So (e.g.) all AQ basis functions are scaled by the same amount.**

Institute and Faculty of Actuaries

Model selection and performance measurement

- Model selection
  - For each λ, calculate 8-fold cross-validation error
  - Select model with minimum CV error
  - **Possibly forecast with extrapolation of any PQ trend especially if your basis functions are collinear**
    - Due to misallocation of effects between AQ and PQ (can happen due to including all of AQ/DQ/PQ in the model).
- Model performance
  - **Visual, standard actuard vs fitted diagnostics etc**
  - **Training error [sum of (actual-fitted)2/fitted values for training data set]**
  - **Test error [sum of (actual-fitted)2/fitted values for test data set]** (N.B. unobservable for real data)
- Model fitting (in R)
  - glmnet package for LASSO (glmnet() and cv.glmnet())
  - ggplot2 for graphs

Institute and Faculty of Actuaries

Outline of presentation

- Motivation
- Regularised regression and the LASSO
- Model specification
- **Real data example**
- Discussion
- Conclusions

Institute and Faculty of Actuaries

Description of the data set

- Motor Bodily injury (moderately long tail)
- (Almost) all claims from one Australian state
  - AQ 1994M9 to 2014M12
  - About 139,000 claims
  - Cost of individual claim finalisations, adjusted to 31 December 2014 $
    - Each claim tagged with:
      - Injury severity score ("maislegal") 1 to 6 and 9
      - Legal representation: maislegal set to 0 for unrepresented severity 1 claims
      - Its operational time **(OT)**, proportion of AQ's ultimate number of claims finalised up to and including it

Institute and Faculty of Actuaries

Known data features

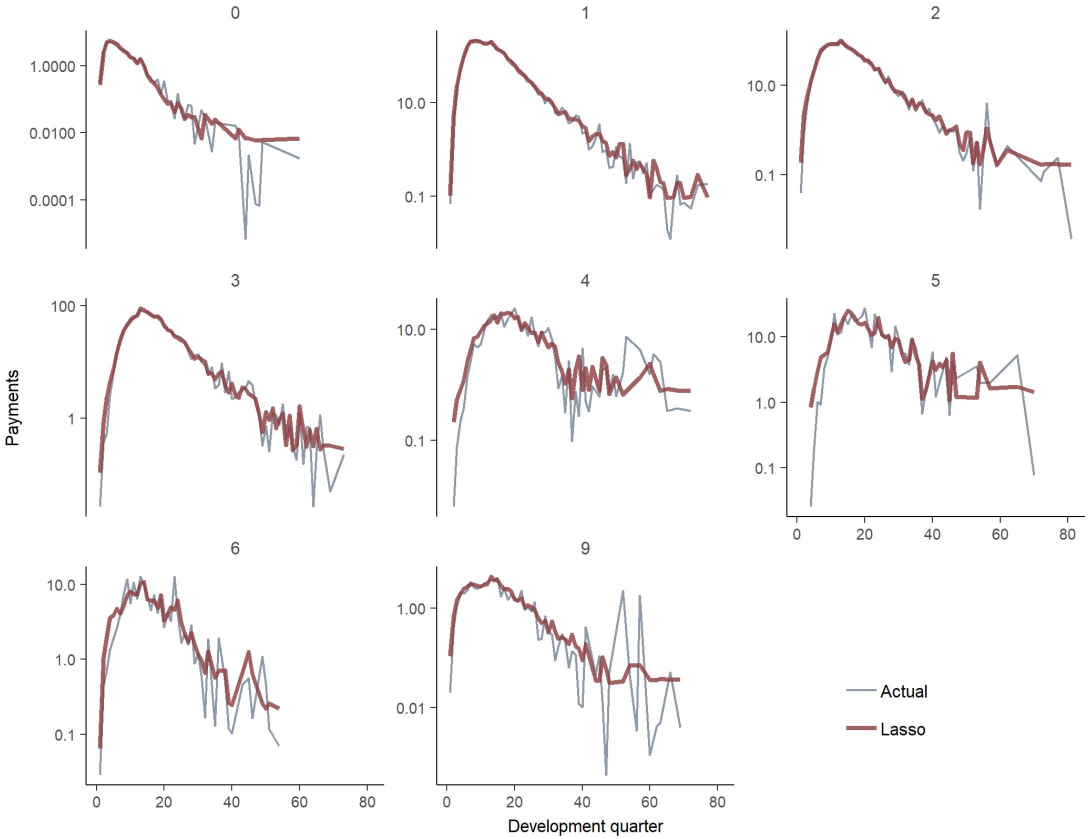
- The **Civil Liability Act** affected AYs ≥ 2003
  - Eliminated many small claims
  - Reduced the size of some other small to medium claims
- There have been periods of material change in the rate of claim settlement
- There is clear evidence of superimposed inflation **(SI)**
  - This has been irregular, sometimes heavy, sometime non-existent
  - SI has tended to be heavy for smallest claims, and non-existent for largest claims

Real data: lasso model

- Lasso applied to the data set summarised into quarterly cells
  - This summary is not theoretically essential but reduces computing time
- Basis functions:
  - Indicator function for severity score (maislegal)
  - All single knot linear splines for OT, PQ
  - All 2-way interactions of maislegal*(OT or PQ spline)
  - All 3-way interactions maislegal*(AQ*OT or PQ*OT Heaviside)
- Model contains 94 terms
  - Average of about 12 per injury severity
- By comparison, the custom-built consultant's GLM included 70 terms
- Forecasts do NOT extrapolate any PQ trend
  - Same basis as GLM forecast

Institute and Faculty of Actuaries

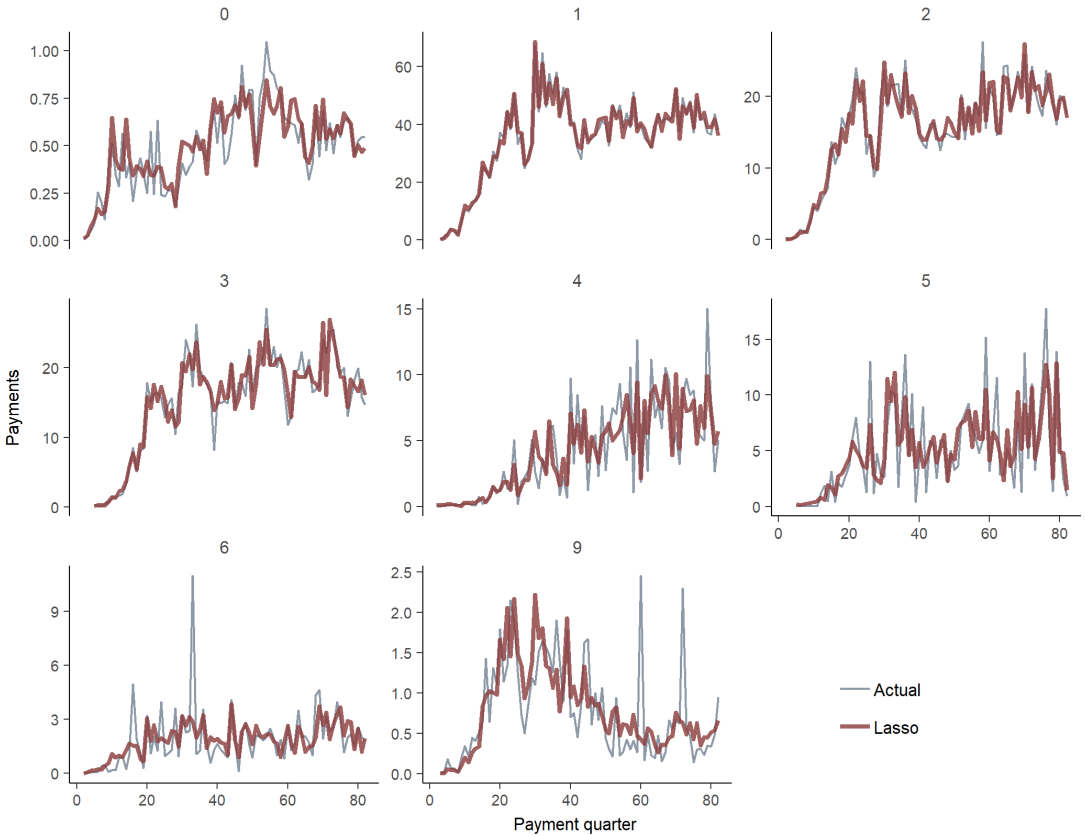
Actual vs fitted - DQ



Payments have been scaled.

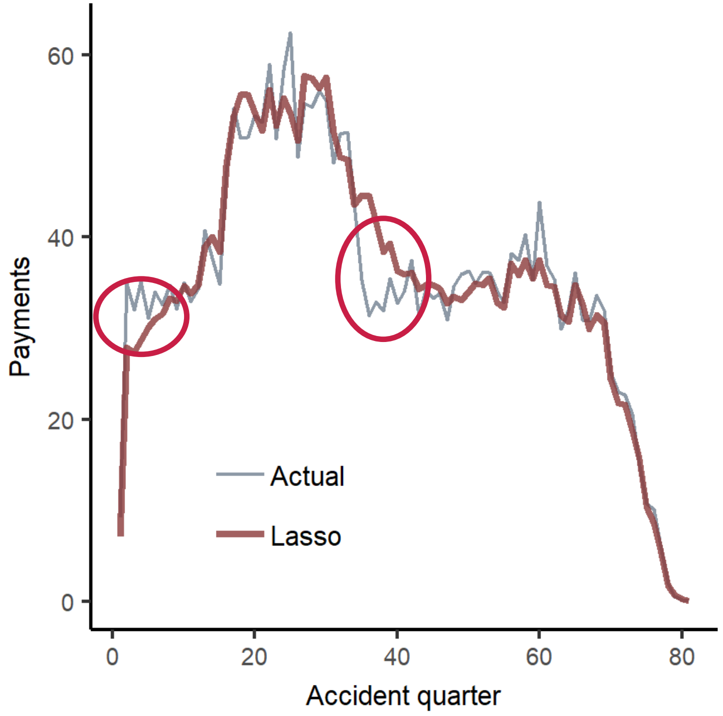
Institute and Faculty of Actuaries

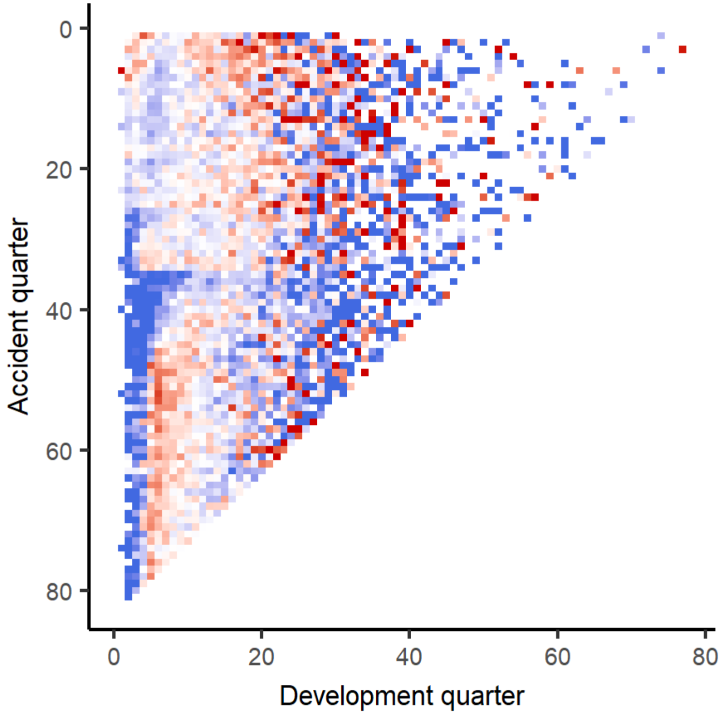
Actual vs fitted - PQ



Institute and Faculty of Actuaries

Actual vs fitted – AQ – injury severity 1

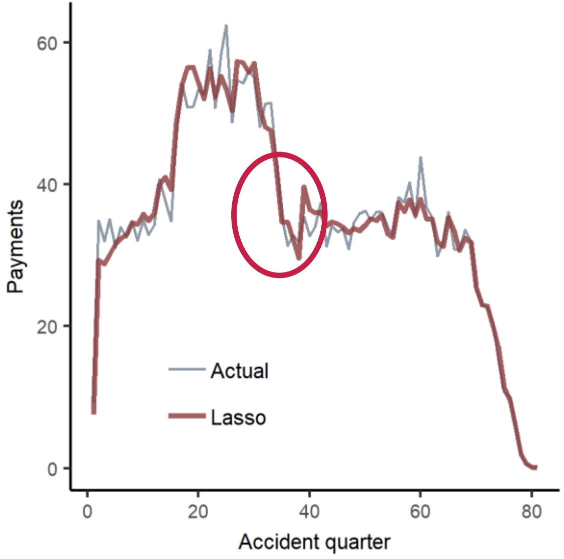


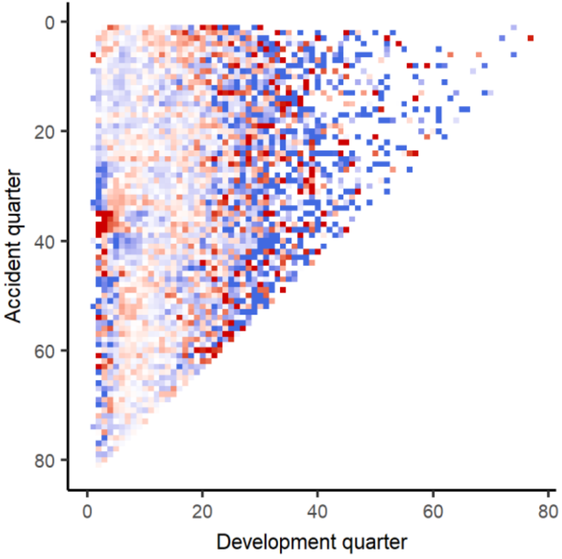


Institute and Faculty of Actuaries

Model misfit: known data features

- Failure of fit results from data features that were known in advance
  - Legislative change affecting AQ ≥ 35
- Perverse to ignore it in model formulation
- Introduce a few simple interactions between injury severity, AQ, OT **without penalty**
  - Brief side investigation required to formulate these
- Model fit considerably improved

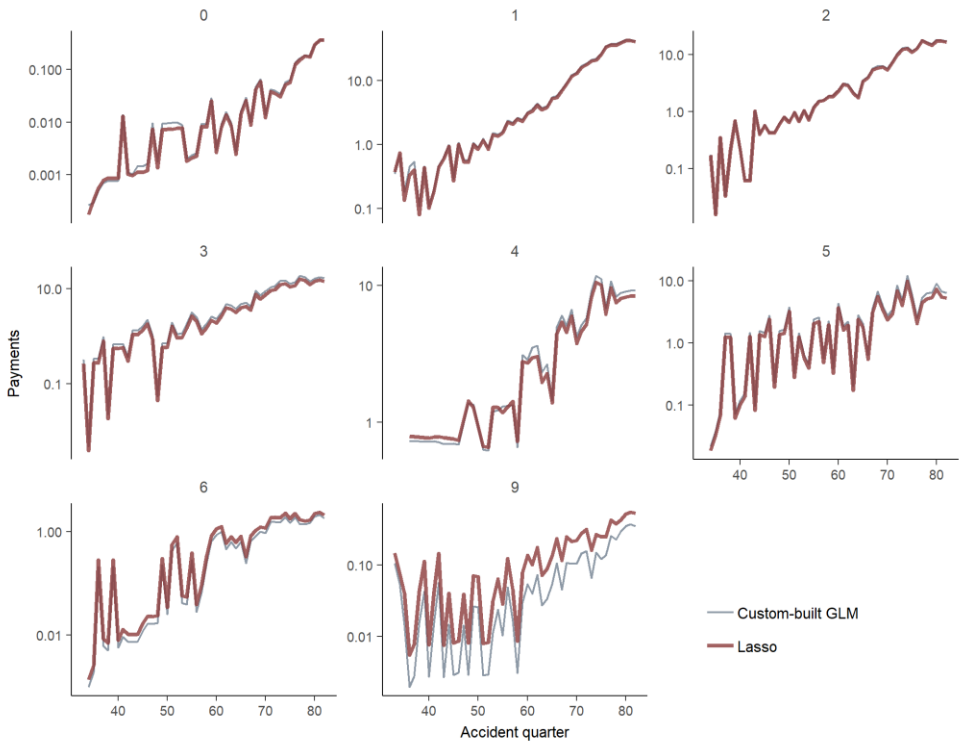




Institute and Faculty of Actuaries

Real data: Human vs Machine

- Same data set modelled with GLMs for many years as part of consulting assignment
  - Complex GLM with interactions for each injury severity
  - Many hours of skilled consultant's time
- Loss reserves from two sources very similar
  - Note that severity 9 is a small and cheap category
  - Some judgemental adjustments in GLM forecasts
- **BUT** consultant's analysis
  - More targeted
  - Less abstract
  - Conveys greater understanding of claim process



Outline of presentation

- Motivation
- Regularised regression and the LASSO
- Model specification
- Real data example
- **Discussion**
- Conclusions

Institute and Faculty of Actuaries

Discussion: feature selection

- How many covariates out of AQ, DQ, PQ should be included?
  - Usually at least 2
  - But 3 will generate collinearity
    - Enlarges model dimension
    - May cause mis-allocation of model features between among dimensions
    - So caution before introducing 3
- Make use of **feature selection** where features are known/strongly suspected

- **Implications for forecasting**
- Forecasts depend on future PQ effects
  - Should these be extrapolated?
  - How will forecasts be affected by mis-allocation?
- **Proposition.** Consider data set containing DQ and PQ effects but no AQ effect. Let $M1$ denote model containing explicit DQ, PQ effects but no AQ effect. Let $M2$ denote identical model except that also contains explicit AQ effects. Then, in broad terms, $M1$ and $M2$ will generate similar forecasts of future claim experience if each extrapolates future PQ effects at a rate representative of that estimated for the past by the relevant model.

Institute and Faculty of Actuaries

Discussion: interpretability

- Most machine learning models subject to the **interpretability problem**
  - Model is an abstract representation of the data
  - May not carry an obvious interpretation of model's physical features
- LASSO models
  - Physical interpretation usually possible, but requires some analysis for visualisation
  - Usually more interpretable than a deep learning model

Discussion: miscellaneous matters

- Prediction error
  - Bootstrap can be bolted onto lasso
  - Preference for non-parametric bootstrap
  - Computer-intensive if min CV chosen separately for each replication
    - Lasso for real data on a laptop
      - 20 minutes without CV
      - 4½ hours with 8-fold CV
      - Note that run times would significantly improve using parallelisation
  - Bootstrap will include at least part of internal model error, but not external model error
- Model thinning
  - Most appropriate distribution provided by lasso software *glmnet* is Poisson
  - Low significance hurdle
  - Reduce number of parameters by applying GLM with gamma error and same covariates as lasso
  - Model performance sometimes degraded, sometimes not
- Bayesian lasso
  - Lasso can be given a Bayesian interpretation
    - Laplacian prior with λ as dispersion parameter
  - Software (Stan) then selects λ according to defined performance criterion

Institute and Faculty of Actuaries

Outline of presentation

- Motivation
- Regularised regression and the LASSO
- Model specification
- Real data example
- Discussion
- **Conclusions**

Institute and Faculty of Actuaries

Conclusions (1)

- Objective was to develop an automated scheme of claim experience modelling
- Routine procedure developed
  - Specify basis functions and performance criteria
  - Then model self-assembles without supervision
- Tested against both synthetic and real data, with reasonable success
  - Lasso succeeds in modelling simultaneous row, column and diagonal features that are awkward for traditional claim modelling approaches
- Procedure is applicable to data of any level of granularity

Institute and Faculty of Actuaries

Conclusions (2)

- Some changes of unusual types may be difficult for a self-assembling model to recognise
  - If these are foreseeable, a small amount of human supervision might be added with minimal loss of automation
- Standard bootstrapping can be bolted on for the measurement of prediction error
  - Uniquely, this can be formulated so as to incorporate part of model error (internal systemic error) within estimated prediction error
- As with any form of unsupervised learning, strong back-end review and model validation is recommended

Institute and Faculty of Actuaries

Questions afterwards

- Contact:
  - grainne.mcguire@taylorfry.com.au

- Paper available at

  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3241906

Institute and Faculty of Actuaries

Questions

Comments

The views expressed in this [publication/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 [publication/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 [publication/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 [publication/presentation] be reproduced without the written permission of the IFoA [*or authors, in the case of non-IFoA research*].

Institute and Faculty of Actuaries