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

Gráinne McGuire
Taylor Fry

Greg Taylor
University of New South Wales

06 September 2018

- Joint work with

Hugh Miller, Taylor Fry

Josephine Ngan, University of New South Wales

Institute and Faculty of Actuaries

Outline of presentation

- Motivation
- Regularized regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

Institute and Faculty of Actuaries

Outline of presentation

- **Motivation**
- Regularized regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

Institute and Faculty of Actuaries

Motivation

- We consider the modelling of claim data sets containing complex features
  - Where chain ladder and the like are inadequate (examples later)
- When such features are present, they may be modelled by means of a Generalized 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

Outline of presentation

- Motivation
- **Regularized regression and the LASSO**
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

Regularized regression and the LASSO

- Consider general GLM structure

Link function

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

Stochastic error (EDF)

- Regularized regression loss function becomes

p-th power of L_p-norm

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

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 Square Shrinkage Operator (LASSO)**

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

Institute and Faculty of Actuaries

Outline of presentation

- Motivation
- Regularized regression and the LASSO
- **Case studies**
  - **Synthetic data**
  - Real data
- Discussion
- Conclusions

Institute and Faculty of Actuaries

Synthetic data sets: construction

- Purpose of synthetic data sets is to introduce known trends and features, and then check the accuracy with which the lasso is able to detect them
- 4 data sets with different underlying model structures considered
  - In increasing order of stress to the model
- Notation
  - k = accident quarter ($= 1,2,..,40$)
  - j = development quarter ($= 1,2,..,40$)
  - $t = k + j - 1$ = payment quarter
  - Y_{kj} = incremental paid losses in (k, j) cell
  - $\mu_{kj} = E[Y_{kj}], \sigma_{kj}^2 = Var[Y_{kj}]$
  - Assumed that $ln\, \mu_{kj} = \alpha_k + \beta_j + \gamma_t$ (generalized chain ladder)

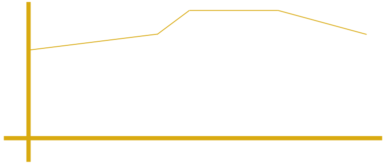
Institute and Faculty of Actuaries

Synthetic data sets: features

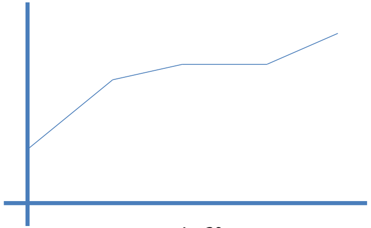
$$ln\,\mu_{kj} = \alpha_k + \beta_j + \gamma_t$$

- **Data set 1:** β_j follows Hoerl curve as function of j, γ_t=0 (no payment year effect), α_k as in diagram
- **Data set 2:** α_k, β_j as for data set 1, γ_t as in diagram
- **Data set 3:** α_k, β_j as for data sets 1&2, γ_t as for data set 2, AQ-DQ interaction (35% increase) as in diagram
- **Data set 4:** $ln\,\mu_{kj} = \alpha_k + \beta_j + \boldsymbol{\theta_j}\gamma_t$, α_k, β_j as for data sets 1-3, γ_t as for data sets 2&3, θ_j as in diagram

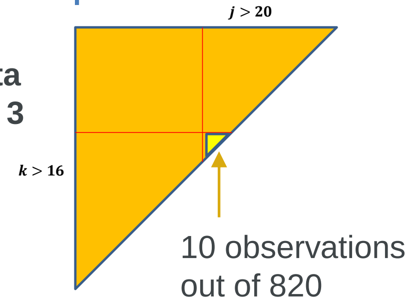
Data set 1



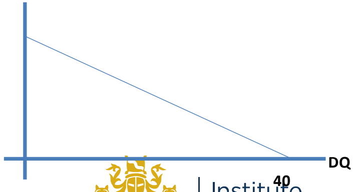
Data set 2



Data set 3



Data set 4



Model formulation, selection and performance measurement

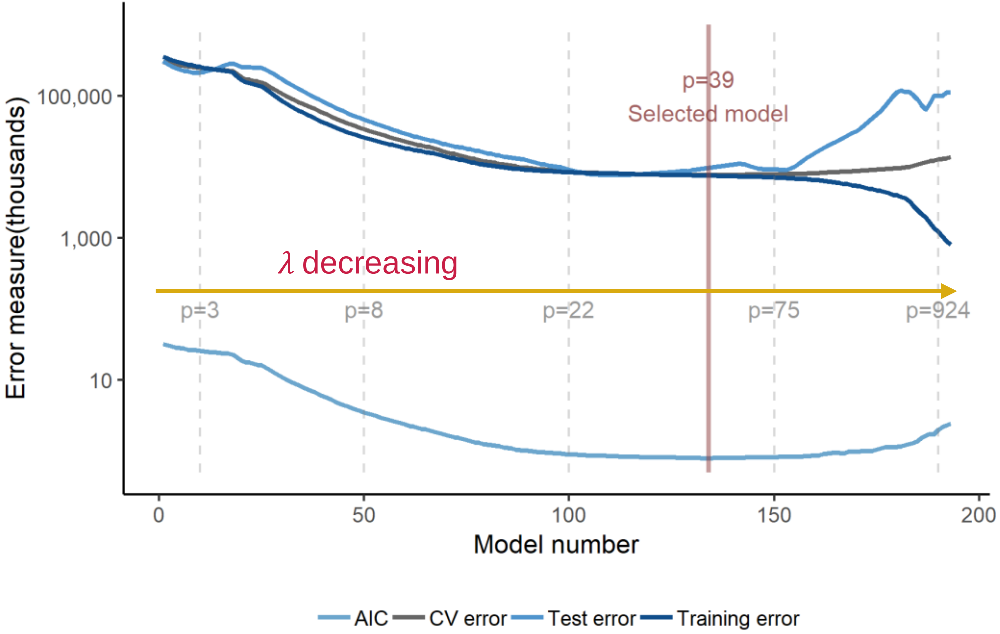
- Model formulation
  - Regressors consist of set of basis functions that form a vector space:
    - All single-knot linear spline functions of k, j, t
    - All 2-way interactions of Heaviside functions of k, j, t
- Model selection
  - For each λ, calculate 8-fold cross-validation error
  - Select model with minimum CV
  - **Forecast with extrapolation of any PQ trend** (to be discussed later)
- Model performance
  - **AIC**
  - **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)

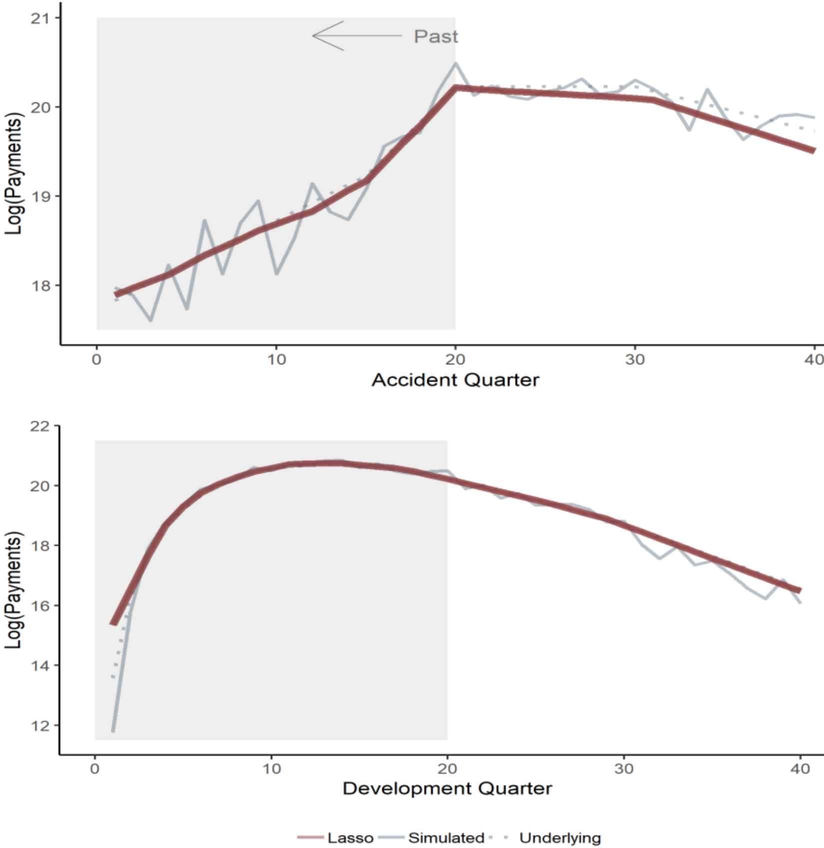
Spline

Heaviside function

Institute and Faculty of Actuaries

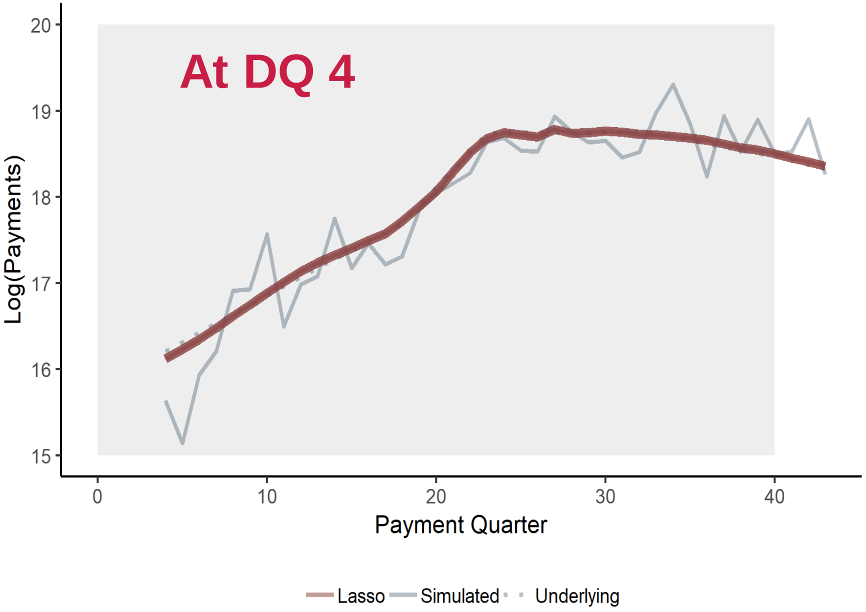
Synthetic data set 1: results

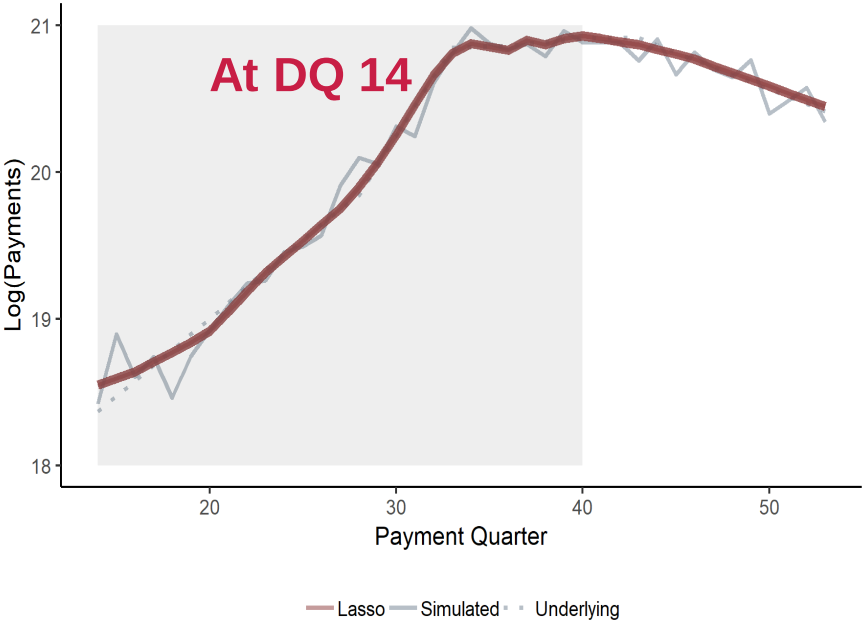




Institute and Faculty of Actuaries

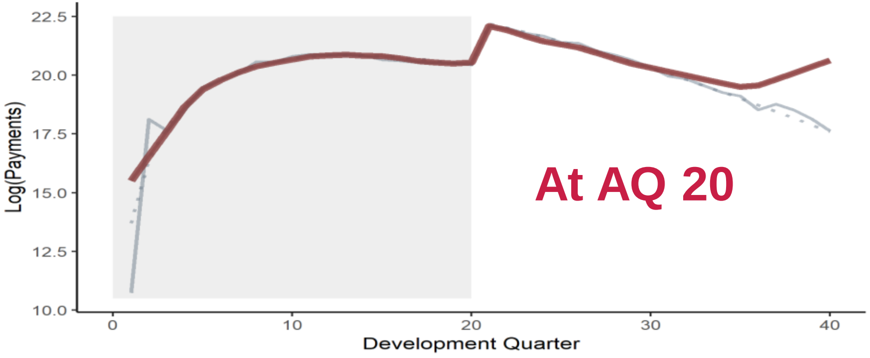
Synthetic data set 2: results

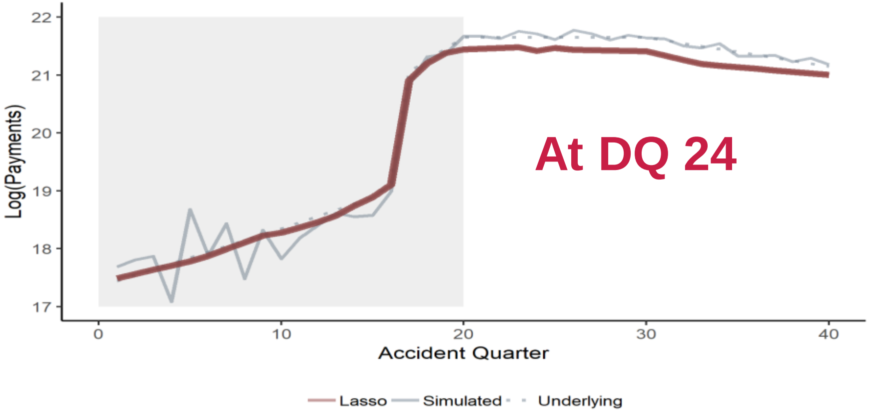


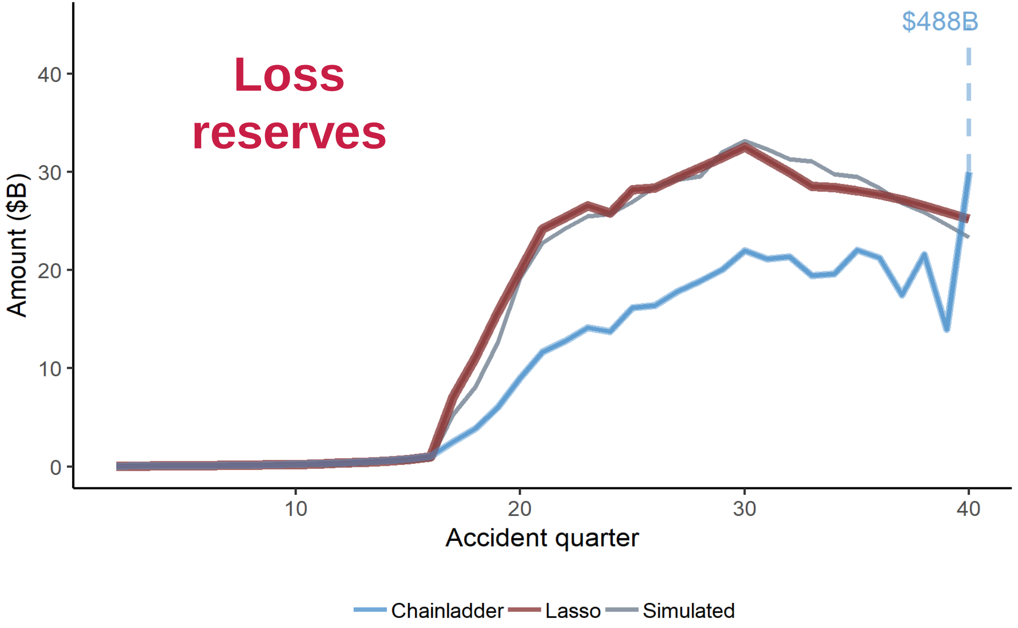


Institute and Faculty of Actuaries

Synthetic data set 3: results

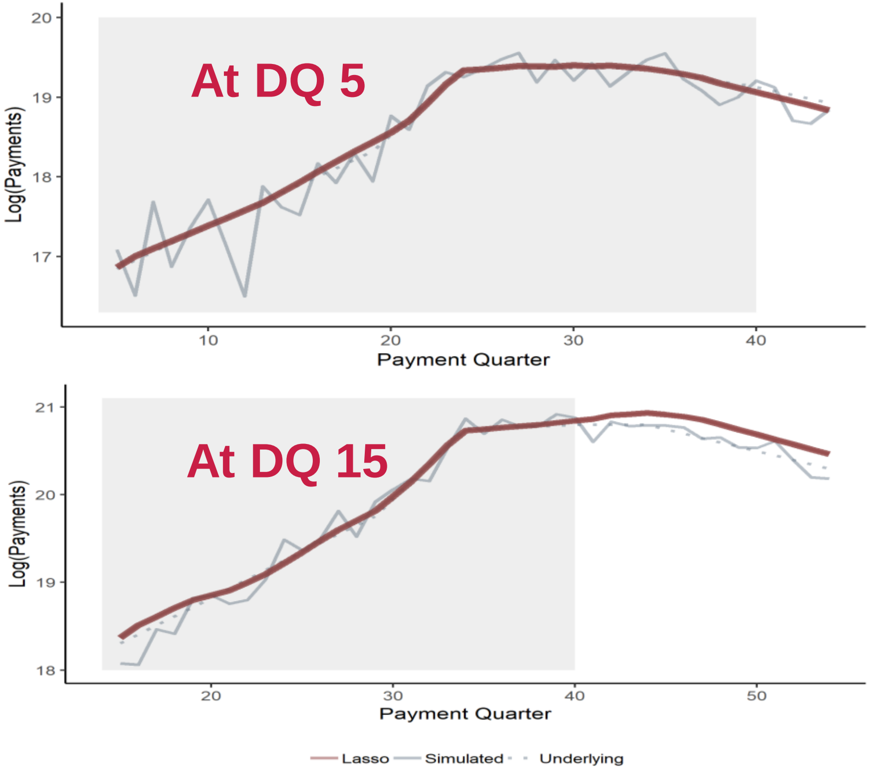






Institute and Faculty of Actuaries

Synthetic data set 4: results



Outline of presentation

- Motivation
- Regularized regression and the LASSO
- Case studies
  - Synthetic data
  - **Real data**
- Discussion
- Conclusions

Institute and Faculty of Actuaries

Real data: nature of 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 finalizations, 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 finalized up to and including it

Institute and Faculty of Actuaries

Real data: known data features

- Collectively, presenters have worked continually with data set for about 17 years
- 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

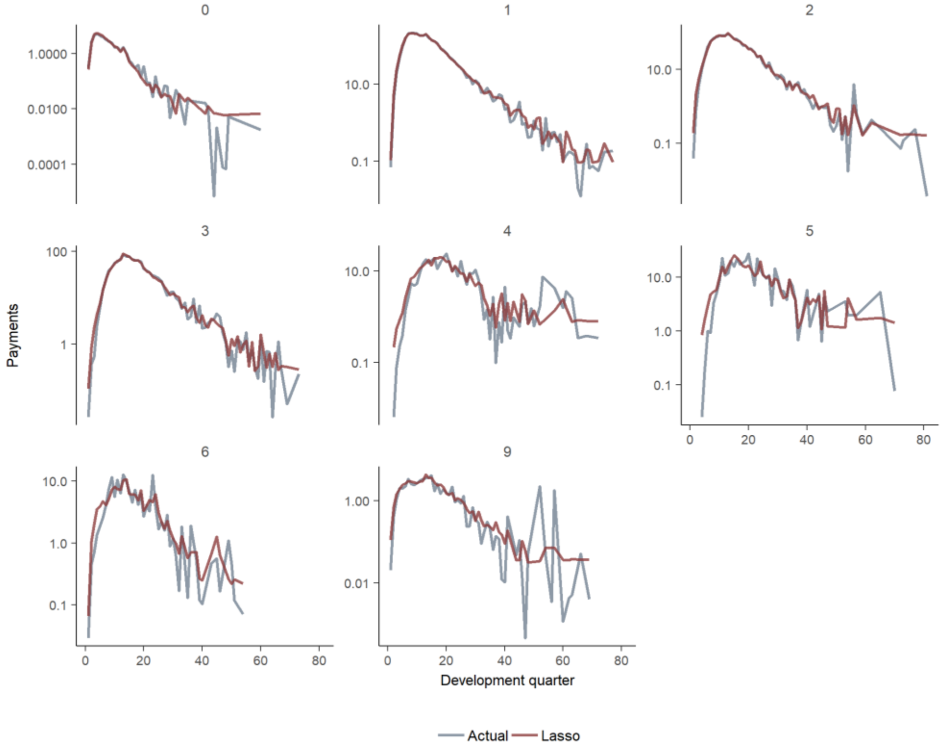
Institute and Faculty of Actuaries

Real data: lasso model

- Lasso applied to the data set summarized 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)
- **Forecasts do NOT extrapolate any PQ trend**
- Model contains 94 terms
  - Average of about 12 per injury severity
- By comparison, the custom-built consultant’s GLM included 70 terms

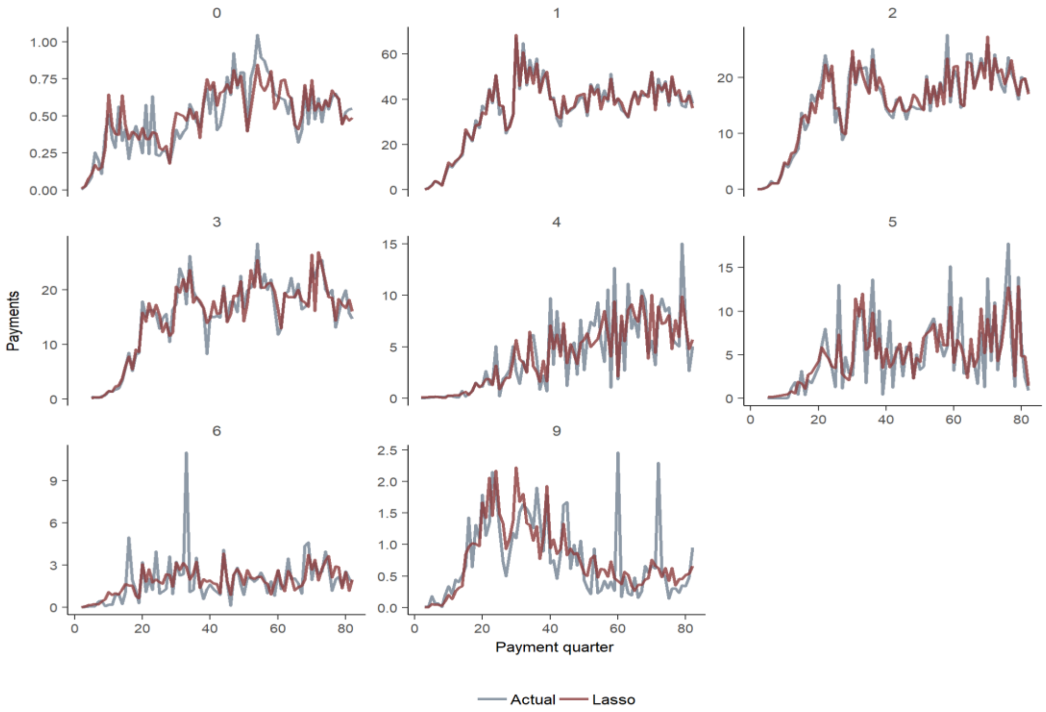
Institute and Faculty of Actuaries

Real data: model fit by DQ

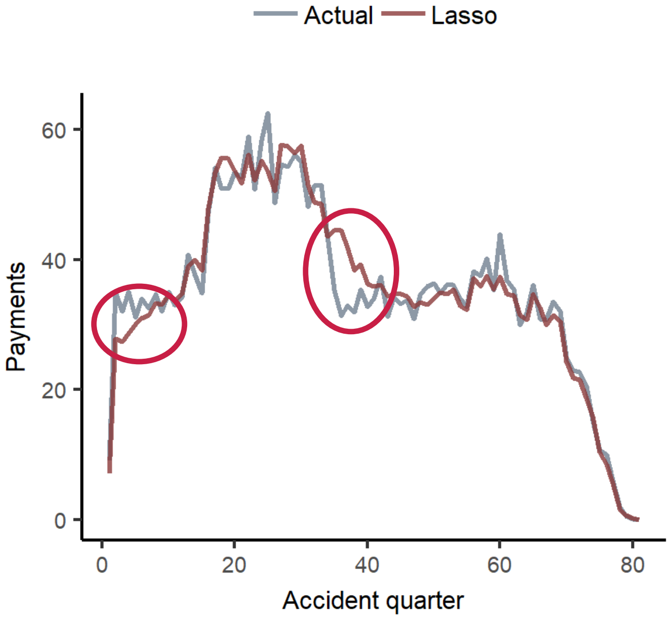


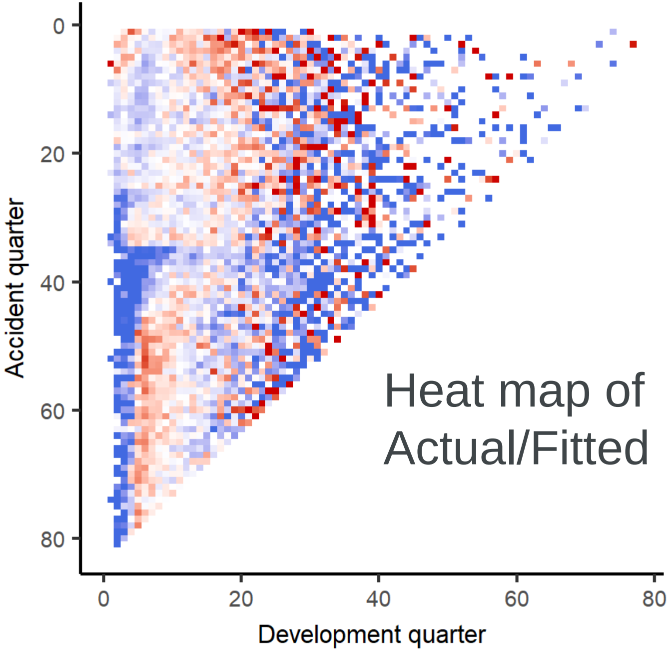
Payments have been scaled.

Real data: model fit by PQ



Real data: model fit by AQ (injury severity 1)

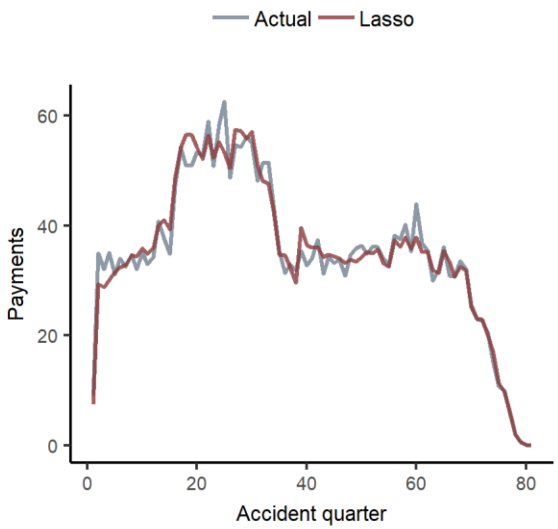


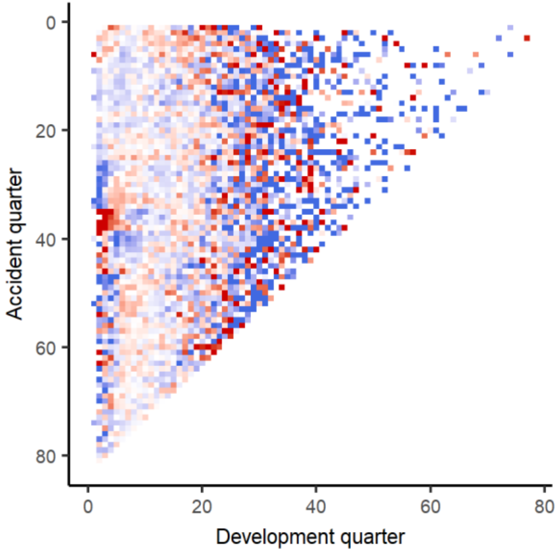


Institute and Faculty of Actuaries

Real data: 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

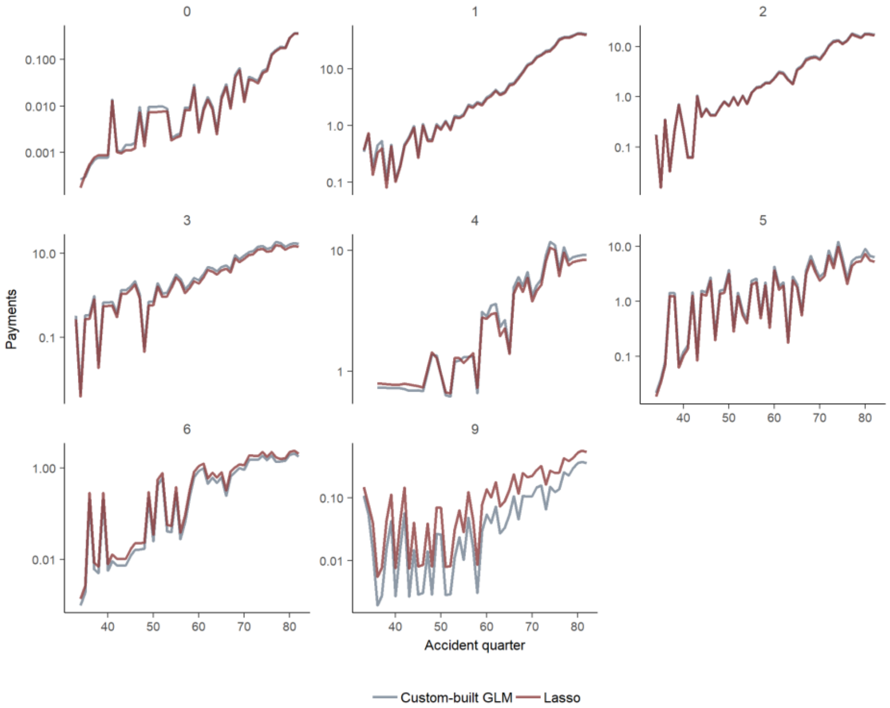




Institute and Faculty of Actuaries

Real data: Human vs Machine

- Same data set modelled with GLMs for many years as part of consulting assignment
  - Separate GLM 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
- **BUT** consultant's analysis
  - More targeted
  - Less abstract
  - Conveys greater understanding of claim process



Institute and Faculty of Actuaries

Outline of presentation

- Motivation
- Regularized regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- **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
  - Physical interpretation usually possible, but requires some analysis for visualization

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
      - 20 minutes without CV
      - 4½ hours with CV
  - 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
- Regularized regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- 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

Conclusions (2)

- Some changes of unusual types may be difficult for an unsupervised model to recognize
  - If these are foreseeable, a small amount of 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 supervision is recommended

Questions

Comments

TAYLOR FRY

UNSW AUSTRALIA Business School AGSM

Institute and Faculty of Actuaries