Society of Actuaries in Ireland

A practical use of machine learning: Loss reserving models using the LASSO

3 April 2019

Disclaimer

The views expressed in this presentation are those of the presenter(s) and not necessarily those of the Society of Actuaries in Ireland or their employers.

Acknowledgements

Joint work by

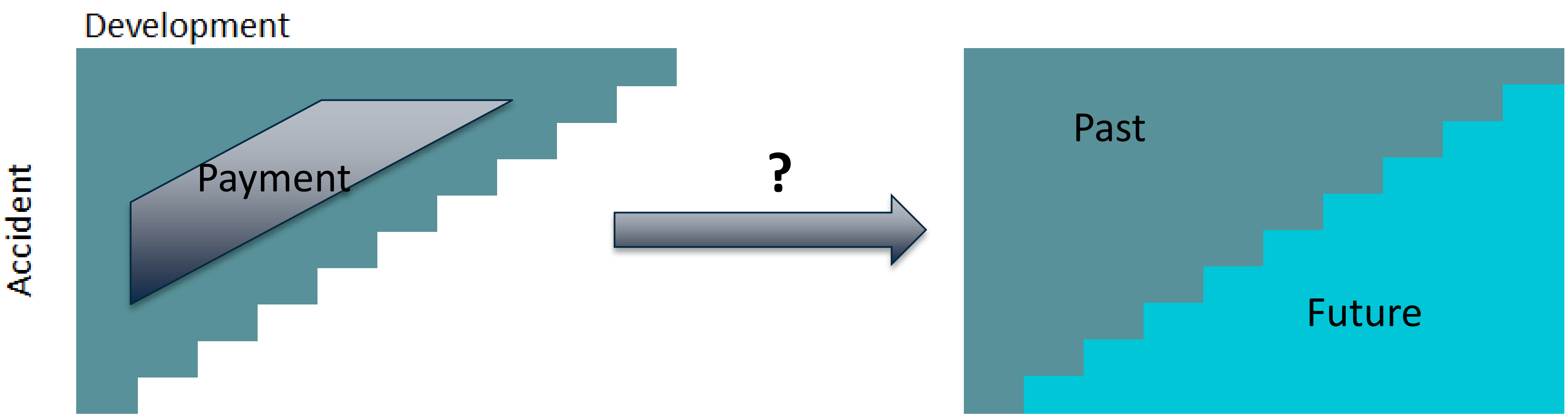
- Gráinne McGuire, Hugh Miller (Taylor Fry)
- Greg Taylor, Josephine Ngan (University of New South Wales)

Outline of presentation

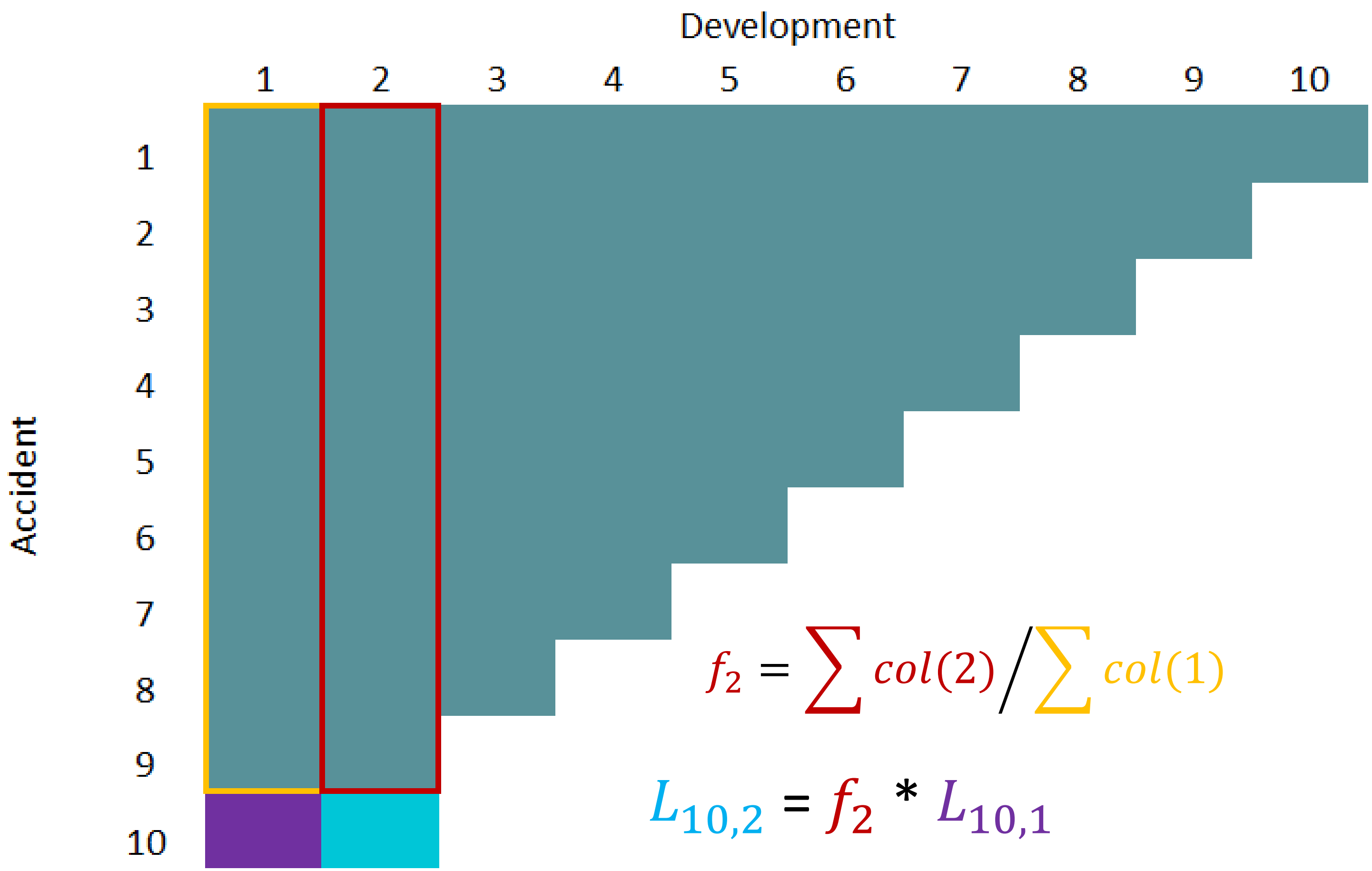
- Loss reserving basics
- Motivation
- Regularised regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

Loss reserving basics

Claims triangle



Simple method: Chain ladder



$$f_2 = \sum col(2) \Big/ \sum col(1)$$

$$L_{10,2} = f_2 * L_{10,1}$$

Outline of presentation

- Loss reserving basics
- Motivation
- Regularised regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

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 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

Outline of presentation

- Loss reserving basics
- Motivation
- Regularised regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

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

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

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 and size of non-zero parameters)

  – Special case: $p = 1$, **Least Absolute Shrinkage and Selection Operator (LASSO)**

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

Favourite ML technique of many - transparent, interpretable model

LASSO: shrinkage and selection

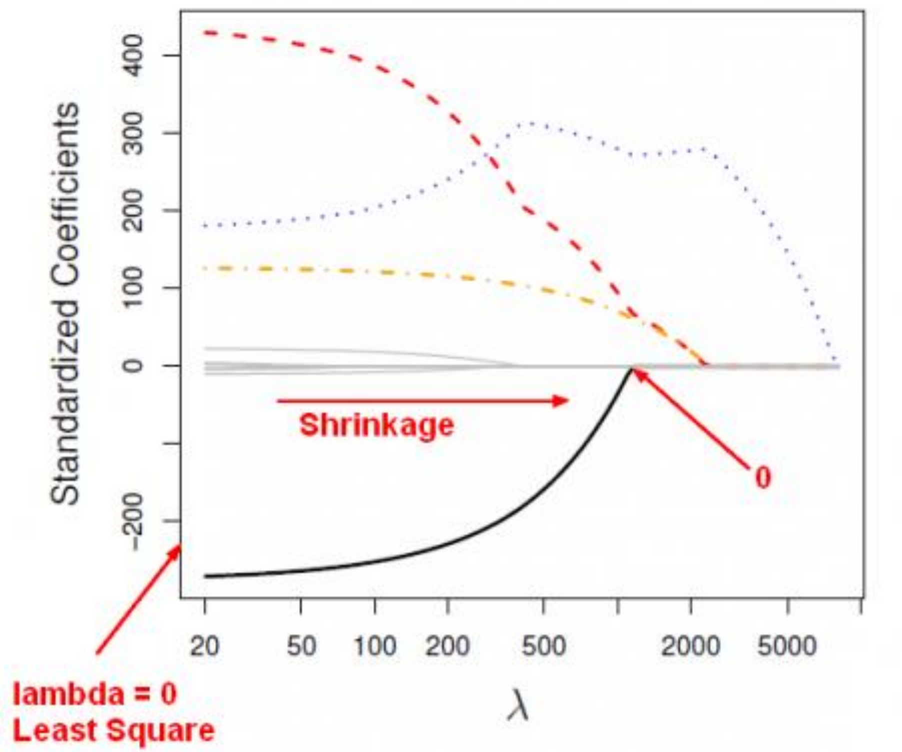


Image sourced from: https://gerardnico.com/data_mining/lasso

Outline of presentation

- Loss reserving basics
- Motivation
- Regularised regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

Synthetic data sets: construction

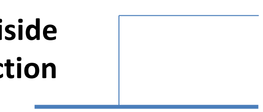
- Purpose of synthetic data sets is to introduce known trends and features, and then check the accuracy with which the LASSO is able to detect them
- 4 data sets with different underlying model structures considered
  - In increasing order of stress to the model
- Notation
  - k = accident quarter [AQ]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

Model formulation

- This is where nearly all the effort is – what predictors/regressors do we use?
- Easy part
  - 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
  - AQ splines are
    - max(0, k-1), max(0, k-2), …., max(0, k-39)
    - Similarly for DQ and PQ
  - AQ x DQ interactions are
    - I(k>1)*I(j>1), I(k>1)*I(j>2), ….., I(k>39)*I(j>39),
    - similarly for AQxPQ and DQxPQ

Spline

Heaviside function



Collinearity in terms – we will come back to that later

Model 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 all AQ basis functions are scaled by the same amount.

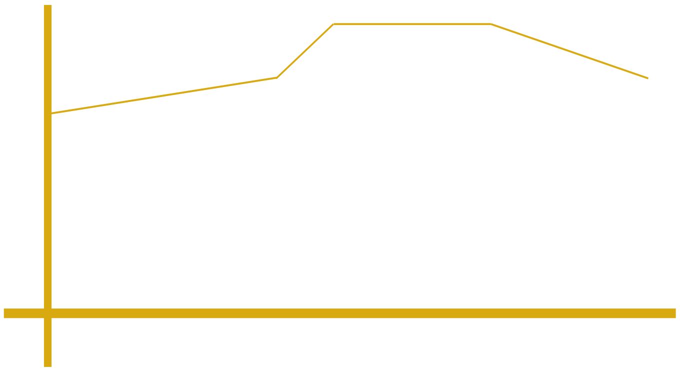
Model selection and performance measurement

- Model selection
  - For each λ, calculate 8-fold cross-validation error
  - Select model with minimum CV error
  - **Forecast with extrapolation of any PQ trend**
    - 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**
  - **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
  - Done in R
  - glmnet package for LASSO
  - ggplot2 for graphs

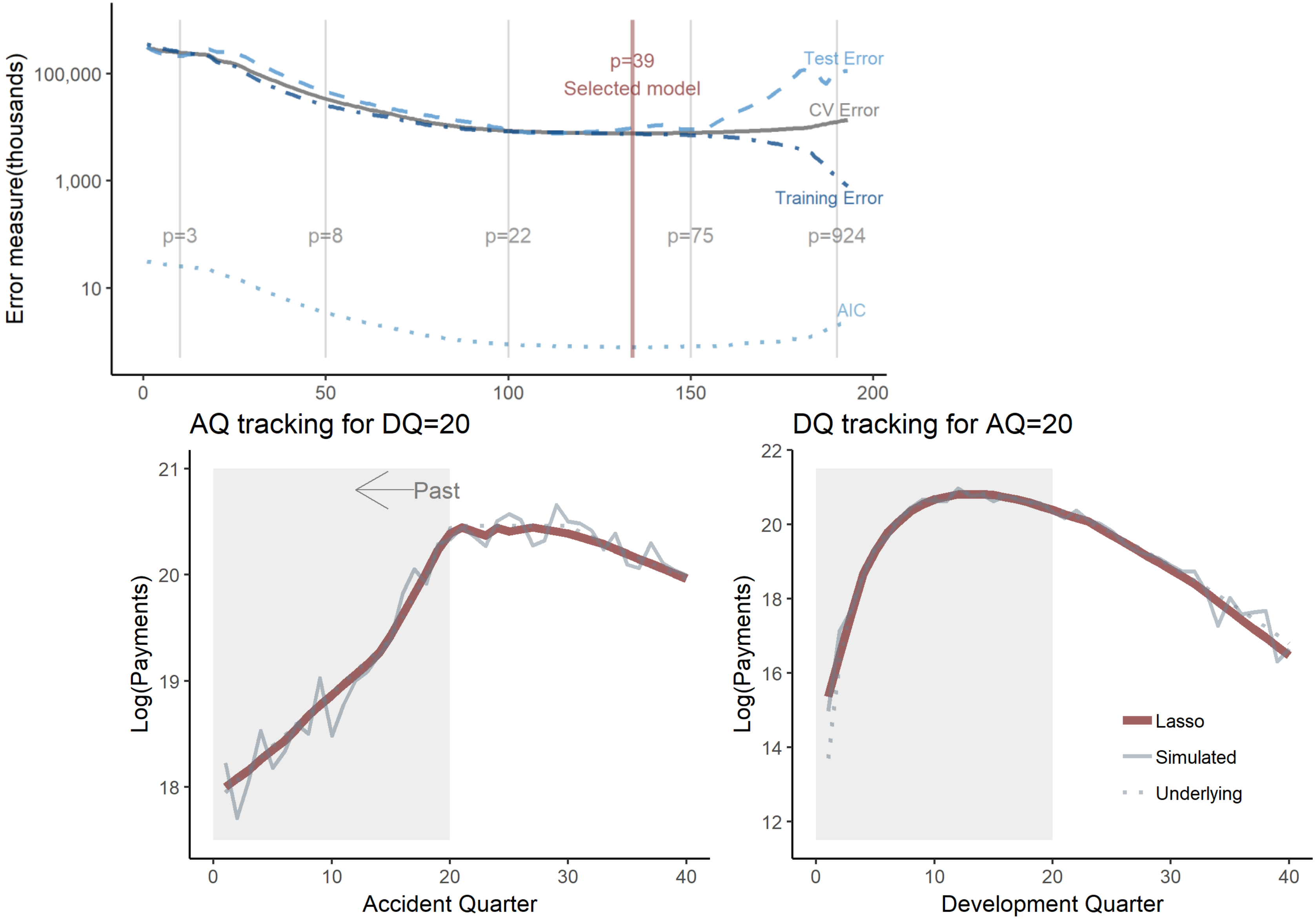
Synthetic data set 1

- $ln\ \mu_{kj} = \alpha_k + \beta_j + \gamma_t$
  - β_j follows Hoerl curve as function of j,
  - γ_t=0 (no payment year effect),
  - α_k as in diagram

Accident quarter effect



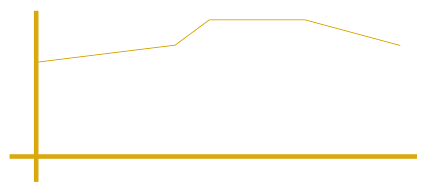
Synthetic data set 1 - results



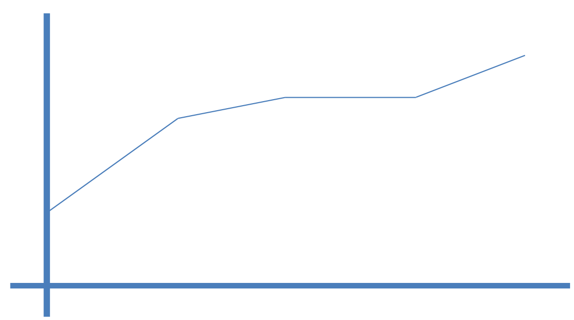
Synthetic data set 2

- $ln\ \mu_{kj} = \alpha_k + \beta_j + \gamma_t$
  - α_k, β_j as for data set 1,
  - γ_t as in diagram

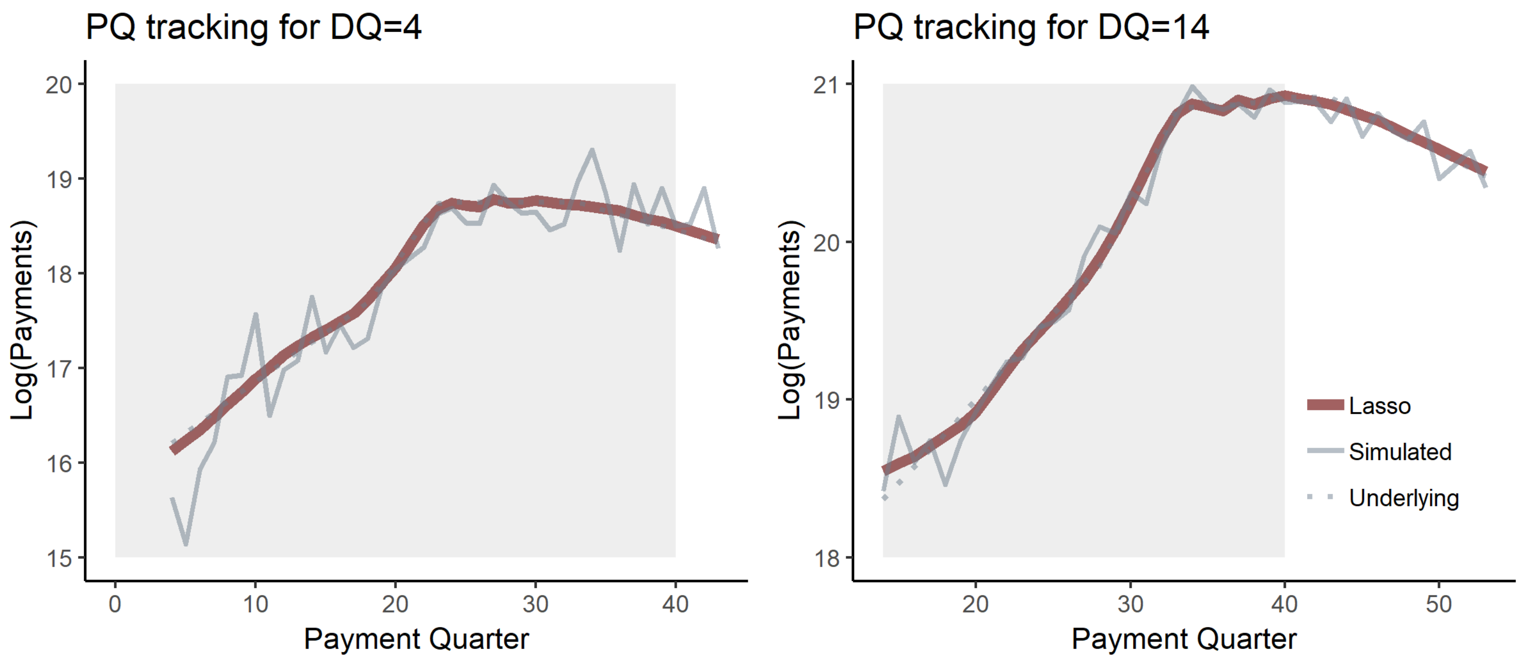
Accident quarter effect



Payment quarter effect



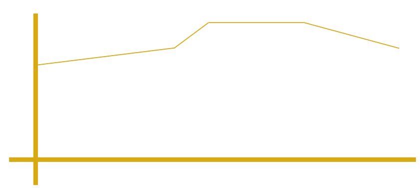
Synthetic data set 2



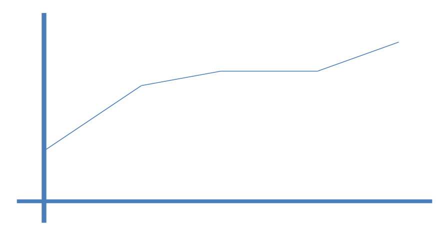
Synthetic data set 3

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

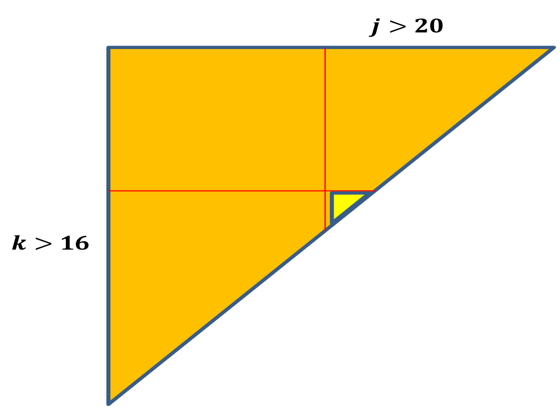
Accident quarter effect



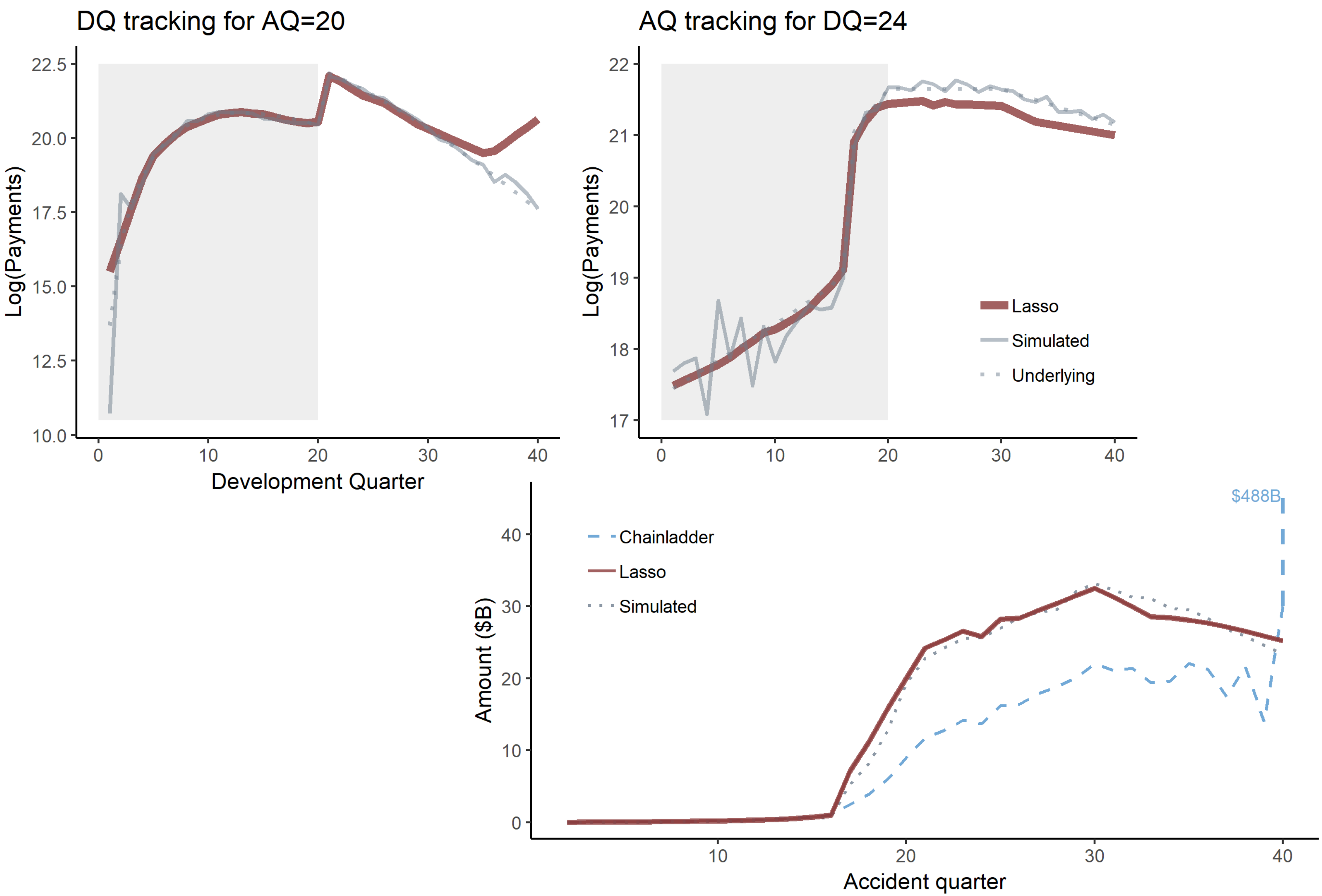
Payment quarter effect



AQ – DQ interaction



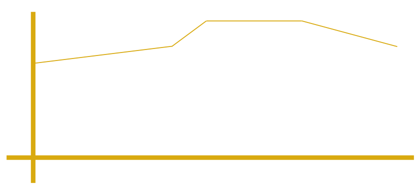
Synthetic data set 3



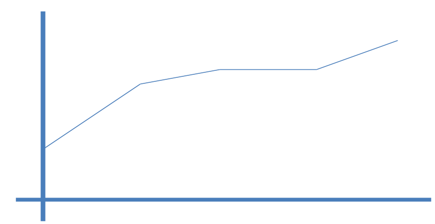
Synthetic data set 4

- $ln\ \mu_{kj} = \alpha_k + \beta_j + \boldsymbol{\theta_j}\gamma_t,$
  - α_k, β_j as for data sets 1-3, γ_t as for data sets 2&3,
  - θ_j as in diagram

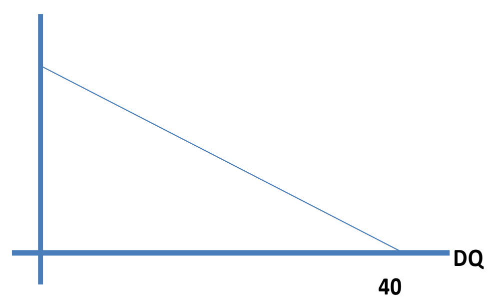
Accident quarter effect



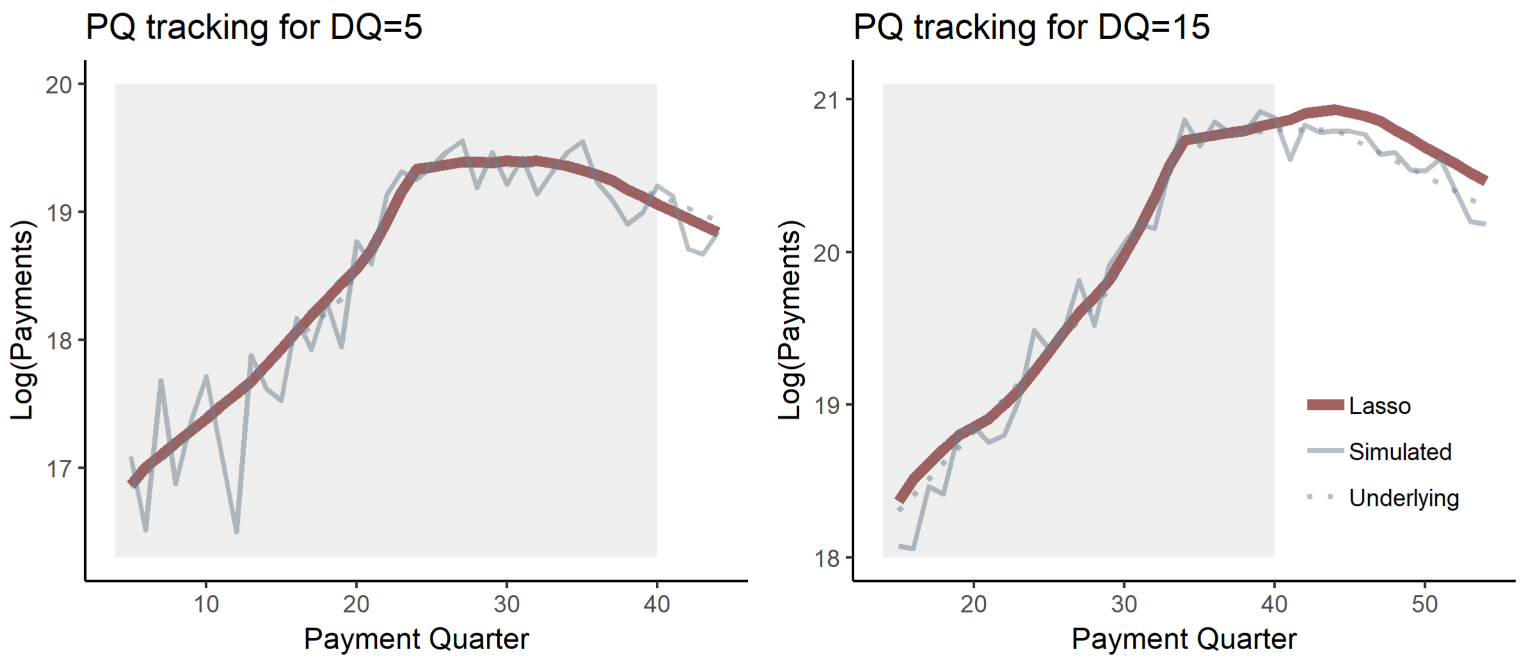
Payment quarter effect



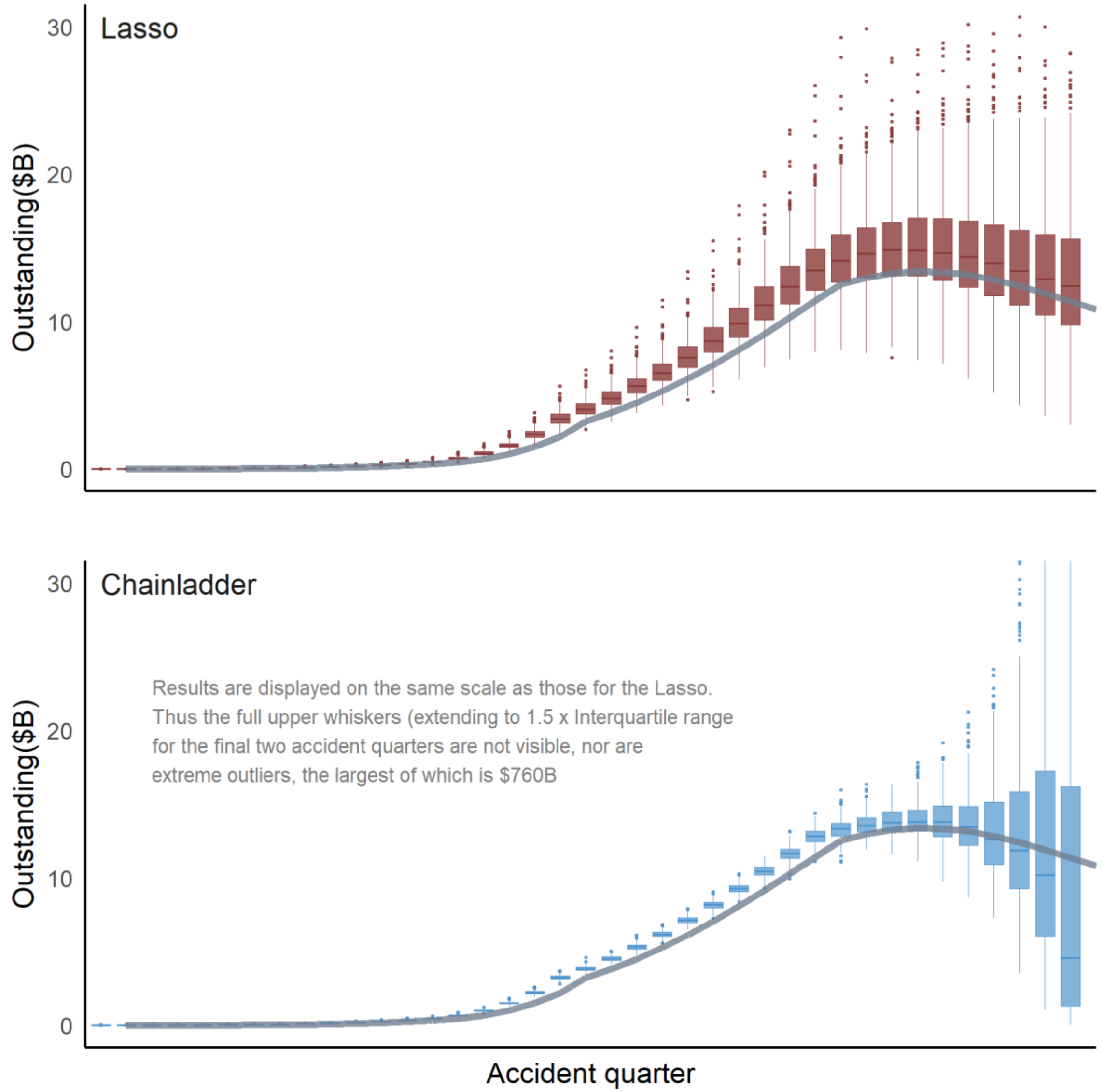
Modification to superimposed inflation



Synthetic data set 4



Synthetic data set 4



Outline of presentation

- Loss reserving basics
- Motivation
- Regularised regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

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

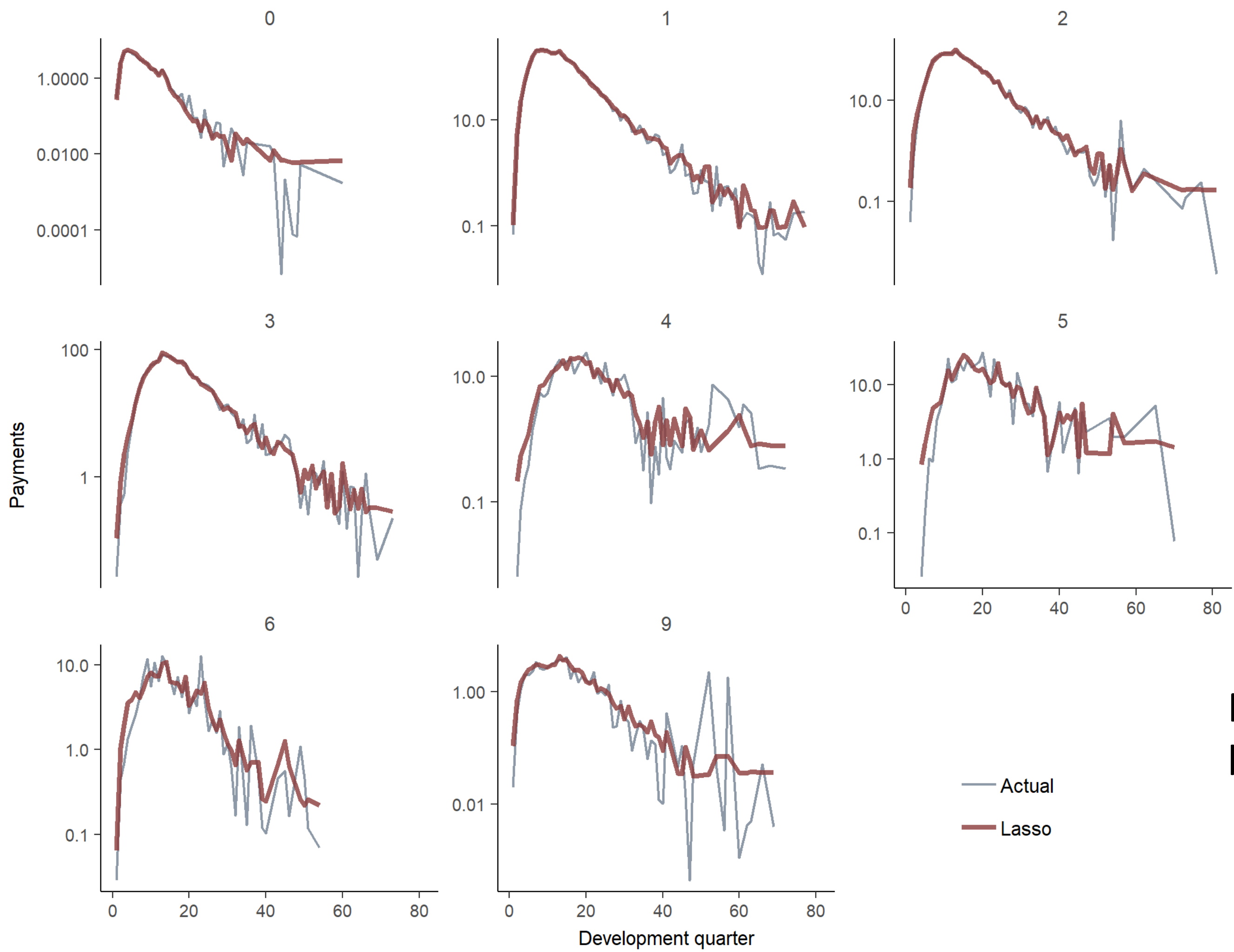
Known data features

- Collectively, presenters have worked continually with data set for about 17 years
- The **Civil Liability Act** affected AYs $\geq\ 2003$
  - 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

LASSO Model

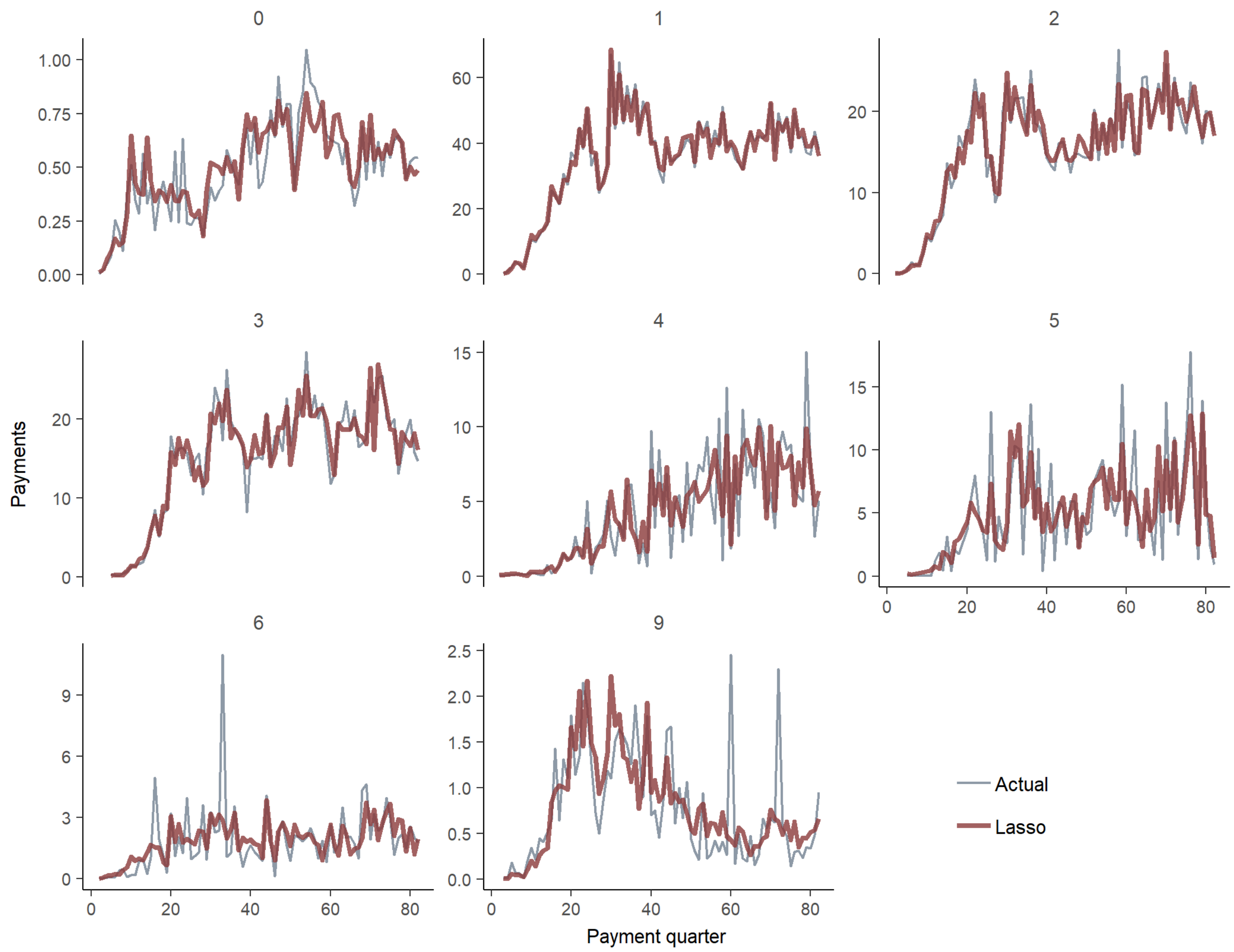
- LASSO applied to the data set summarised into quarterly cells
  - This summary is not theoretically essential but reduces computing time
- Basis functions:
  - Indicator function for severity score (maislegal)
  - All single knot linear splines for OT, PQ
  - All 2-way interactions of maislegal*(OT 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**
  - Less collinearity in basis functions used than the synthetic data examples
  - Less potential for misallocation of effects

Actual vs fitted - DQ

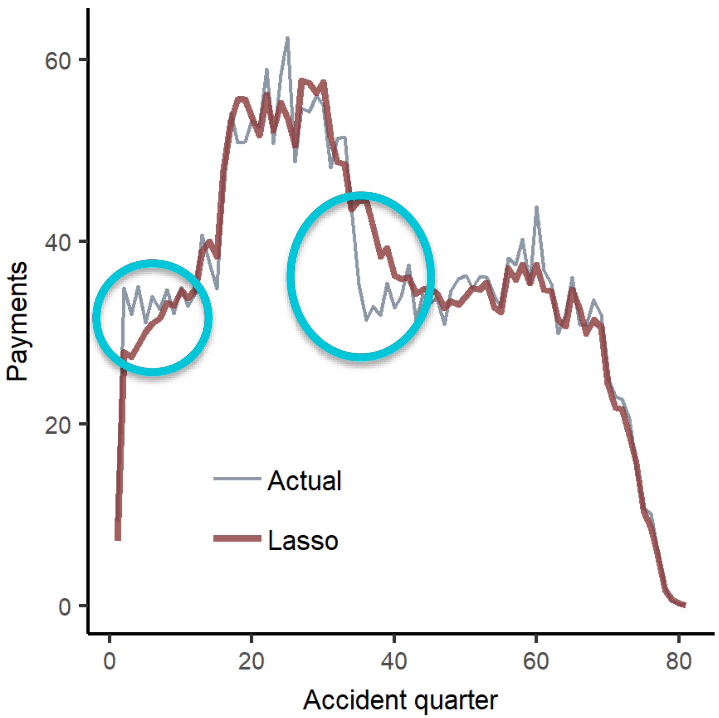


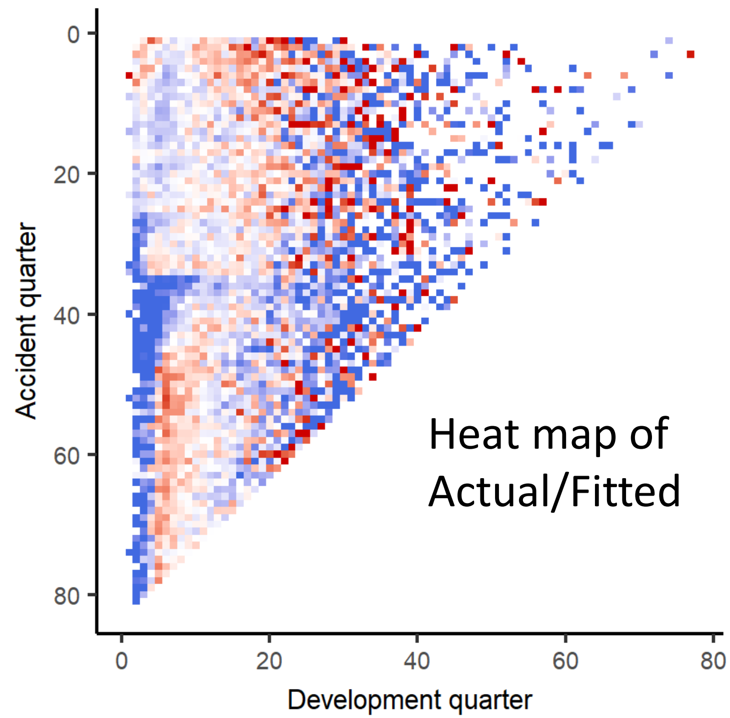
Payments have been scaled

Actual vs Fitted - PQ



Model fit by AQ (injury severity 1)

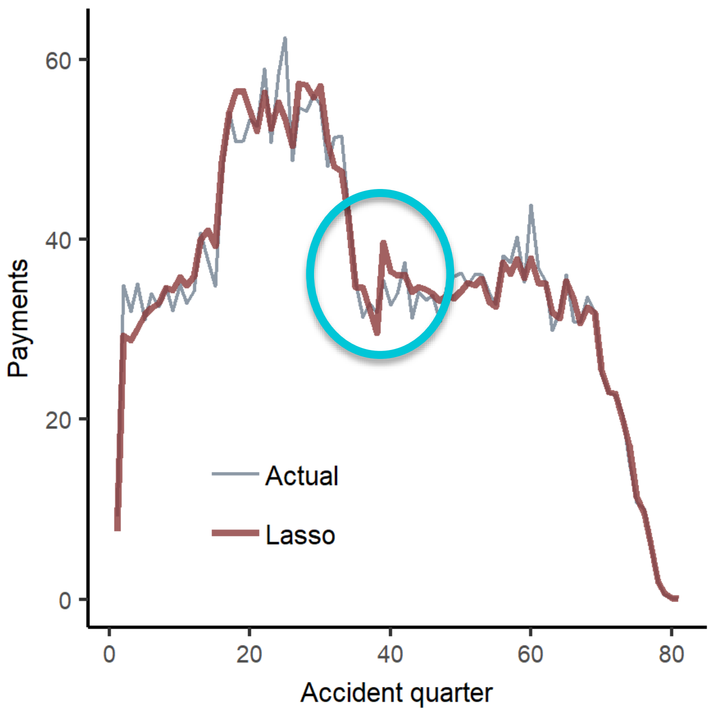


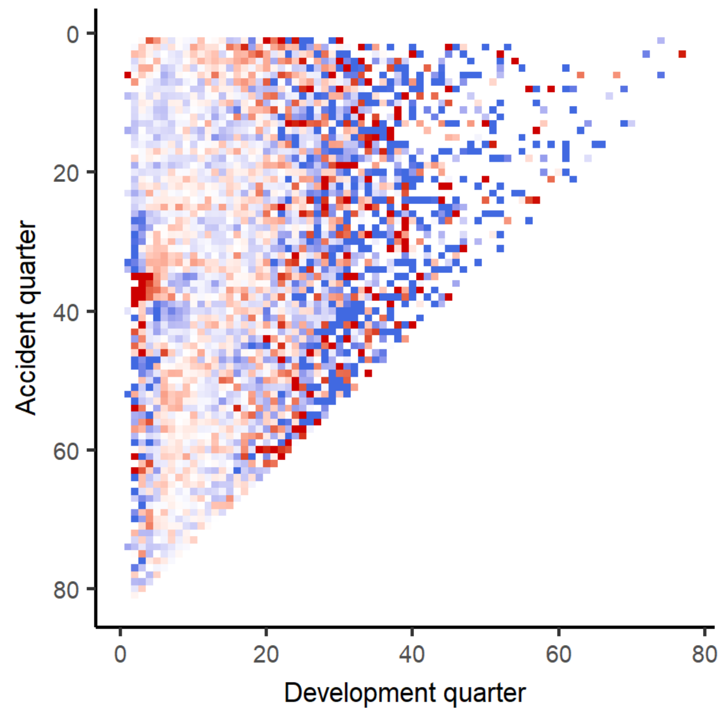


Heat map of Actual/Fitted

Model misfit: known data features

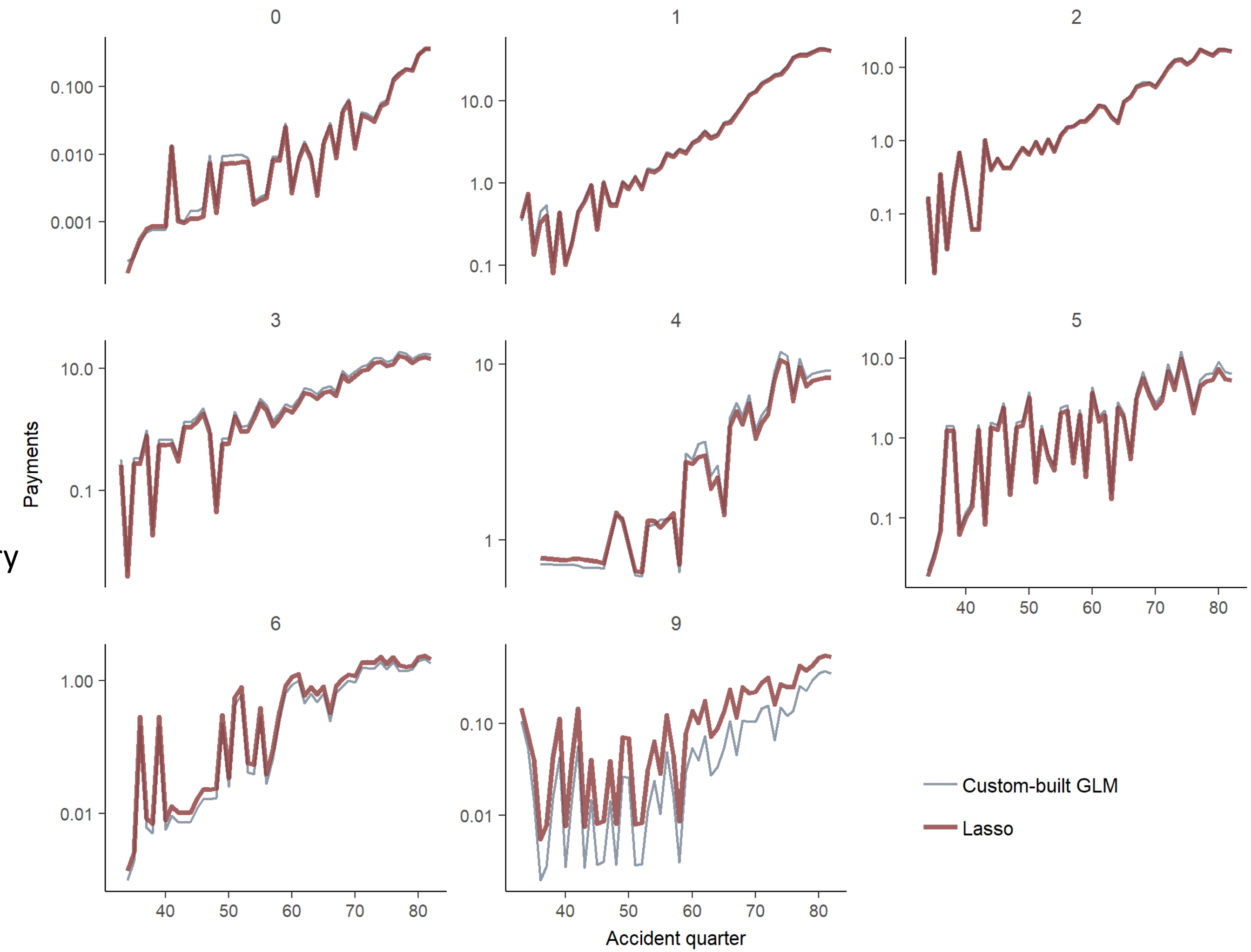
- Failure of fit results from data features that were known in advance
  - Legislative change affecting AQ $\geq$ 35
- Perverse to ignore it in model formulation
- Introduce a few simple interactions between injury severity, AQ, OT **without penalty**
  - Brief side investigation required to formulate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





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
  - Judgemental change in GLM forecast
- **BUT** consultant's analysis
  - More targeted
  - Less abstract
  - Conveys greater understanding of claim process



Outline of presentation

- Loss reserving basics
- Motivation
- Regularised regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

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.

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 visualisation
  - However, LASSO much more interpretable than a deep learning model

Miscellaneous

- 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
      - Sequential run. Could be speeded up with 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

Outline of presentation

- Loss reserving basics
- Motivation
- Regularised regression and the LASSO
- Case studies
  - Synthetic data
  - Real data
- Discussion
- Conclusions

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 recognise
  - 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
- As with any form of machine learning, model validation is important

Questions afterwards?

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

- Paper available at

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