Business School

Never Stand Still | Business School | Risk and Actuarial Studies

Self-assembling insurance claim models

Greg Taylor
School of Risk and Actuarial Studies, UNSW Australia
Sydney, Australia

Gráinne McGuire
Hugh Miller
Taylor Fry Analytics & Actuarial Consulting, Sydney, Australia

Casualty Loss Reserve Seminar
Chicago, USA, September 18-20, 2016.

Overview

- Motivation
- Regularized regression
  - In general
  - The lasso
- Loss reserving framework and notation
- Application of lasso to loss reserving
- Test on simulated data
  - Data set
  - Results
- Further testing and development
- Conclusion

Motivation (1)

- Consider loss reserving on the basis of a conventional **triangular data set**
  - e.g. paid losses, incurred losses, etc.
- A model often used is the **chain ladder**
  - This involves a very simple model structure
  - It will be inadequate for the capture of certain claim data characteristics from the real world (illustrated later)

Motivation (2)

- When such features are present, they may be modelled by means of a **Generalized Linear Model (GLM)** (McGuire, 2007; Taylor & McGuire, 2004, 2016)
- 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

Overview

- Motivation
- **Regularized regression**
  - In general
  - The lasso
- Loss reserving framework and notation
- Application of lasso to loss reserving
- Test on simulated data
  - Data set
  - Results
- Further testing and development
- Conclusion

Regularized regression: in general

- Consider a standard (multivariate) linear regression problem, expressed in vector and matrix form:

$$y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

- Estimation of parameter vector β by $\hat{\beta}$
  - OLS loss function is

$$\left(y - X\hat{\beta}\right)^T \left(y - X\hat{\beta}\right) = \left\|y - X\hat{\beta}\right\|_2^2 \qquad \text{[least squares]}$$

  where $\|\bullet\|_p$ denotes the ***p*-norm**: $\|z\|_p = \left[\left(\sum_j |z_j|^p\right)\right]^{1/p}$

  - regularized regression loss function is

$$\left\|y - X\hat{\beta}\right\|_2^2 + \lambda\left\|\hat{\beta}\right\|_p^p$$

Penalty for poor fit | **Tuning constant ($\lambda \geq 0$)** | **Penalty for additional parameters**

Regularized regression: the lasso

- Regularized regression loss function (previous slide)

$$\|y - X\hat{\beta}\|_2^2 + \lambda\|\hat{\beta}\|_p^p$$

- Special cases
  - $p = 0$: OLS regression (no penalty)
  - $p = 2$: Ridge regression
  - **$p = 1$: Lasso (Least Absolute Shrinkage and Selection Operator)**
- Adaptation to Generalized Linear Models (**GLM**)
  - GLM takes form

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

  - Regularized regression loss function becomes

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

Link function

Stochastic error (EDF)

Log-likelihood

Formal derivations of the GLM lasso

- Constrained parameters
  - Fit GLM by MLE subject to parameter constraint $\|\hat{\beta}\|_1 = \sum_j |\beta_j| \leq const.$
- Random effects GLM
  - MAP (maximum a posteriori) estimation of β when parameters subject to random effects with independent Laplace distributed priors:

$$pdf\ of\ \beta_j = exp\left(½\lambda|\beta_j|\right)$$

Application of GLM lasso

- Regularized GLM regression loss function (earlier slide)

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

- Lasso version ($p = 1$)

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

- The second member of the loss function tends to force parameters to zero
  - $\lambda \rightarrow 0$: model approaches conventional GLM
  - $\lambda \rightarrow \infty$: all parameter estimates approach zero
  - Intermediate values of λ control the complexity of the model (number of non-zero parameters)

Overview

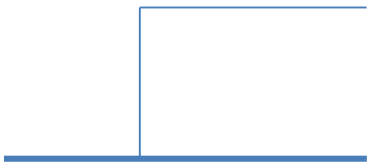
- Motivation
- Regularized regression
  - In general
  - The lasso
- **Loss reserving framework and notation**
- Application of lasso to loss reserving
- Test on simulated data
  - Data set
  - Results
- Further testing and development
- Conclusion

Loss reserving framework and notation (1)

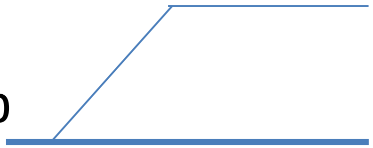
- Experimental simulated data sets
  - Incremental quarterly paid claim triangles (40x40)
- Notation
  - $k =$ accident quarter
  - $j =$ development quarter
  - $t = k + j - 1 =$ payment quarter
  - $Y_{kj} =$ incremental paid losses in (k, j) cell
  - $\mu_{kj} = E[Y_{kj}], \sigma_{kj}^2 = Var[Y_{kj}]$
    - Assumed that $ln\, \mu_{kj} = \alpha_k + \beta_j + \gamma_t$ (generalized chain ladder)
    - The σ_{kj}^2 are selected throughout to be consistent with the Mack formulation of the chain ladder model

Loss reserving framework and notation (2)

- It is necessary to decide the set of regressors to be included in the model, i.e. the rows of the design matrix X
- Each regressor is some function of k, j
- The present application includes all of the following regressors (**"basis functions"**)
  - All unit step functions of k, j or t
    - Various locations of the step



  - All unit gradient ramp functions of k, j or t
    - Various locations of the start and end of the ramp
    - Combinations of these **(linear splines)** can approximate smooth curves



  - (later) interactions between the step basis functions

Overview

- Motivation
- Regularized regression
  - In general
  - The lasso
- Loss reserving framework and notation
- **Application of lasso to loss reserving**
- Test on simulated data
  - Data set
  - Results
- Further testing and development
- Conclusion

Application of lasso to loss reserving

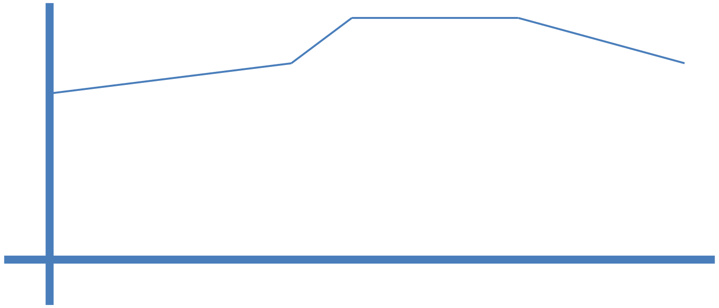
- 4 data sets with different underlying model structures (of μ_{kj}) considered
  - In increasing order of stress to the model
  - Lasso applied to each dataset
    - Once the tuning constant λ selected, models are self-assembling
  - Interest in examination of the extent to which the model self-assembles the structures concealed in the data
  - Also compare model forecasts with those from the chain ladder

Overview

- Motivation
- Regularized regression
  - In general
  - The lasso
- Loss reserving framework and notation
- Application of lasso to loss reserving
- **Test on simulated data**
  - Data set
  - Results
- Further testing and development
- Conclusion

Data set 1: set-up

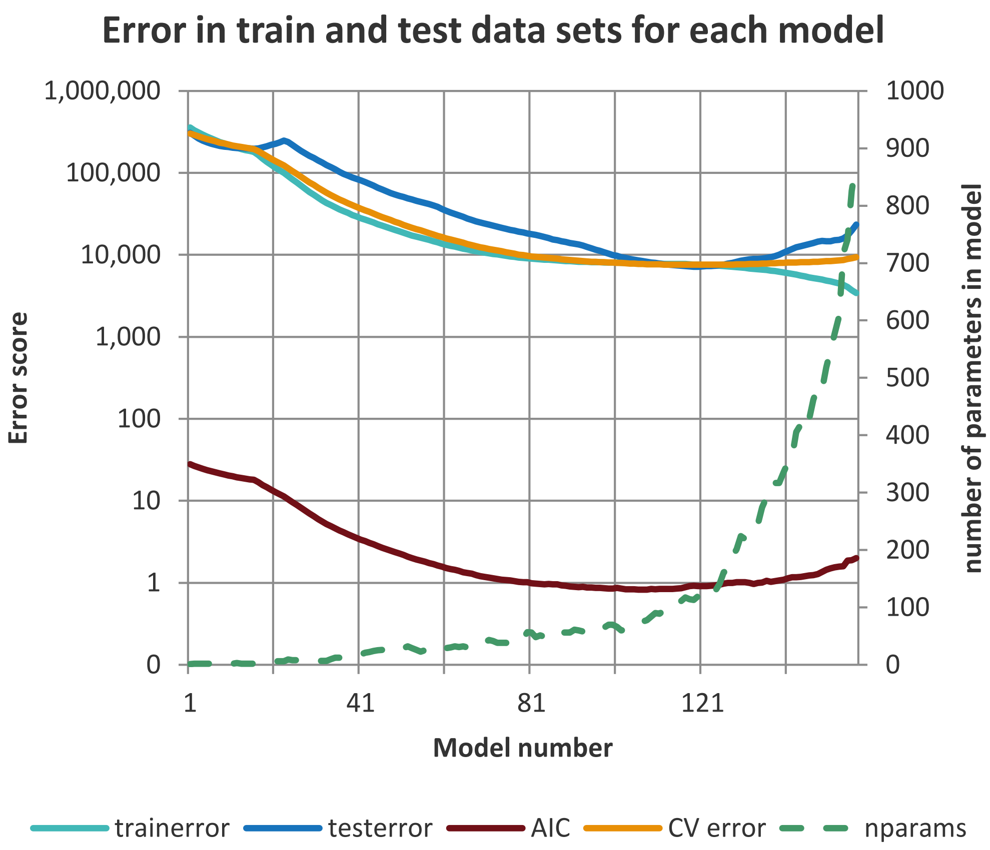
- Recall $ln\, \mu_{kj} = \alpha_k + \beta_j + \gamma_t$
- Observations (both past and future, whole square, not just triangle) simulated according to $Y_{kj} \sim ODP(\mu_{kj}, \phi)$
  - Assumed that the Y_{kj} are in constant dollar values (inflation corrected)
    - Any calendar quarter trend represents **superimposed inflation ("SI")**
  - Upper triangle forms training data set
  - Lower triangle forms test data set
- Assume
  - β_j follows Hoerl curve as function of j
  - γ_t=0 (no payment year effect)
  - α_k appears as in diagram



Data set 1: model selection

- The number of basis functions (regressors) was 2,380
- Model fitted to the training data set for a large number of values of tuning constant λ
  - As λ increases, number of non-zero parameters decreases
- Model performance (for any given λ) measured by:
  - 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 in practice)
  - 8-fold cross-validation error based on training data set

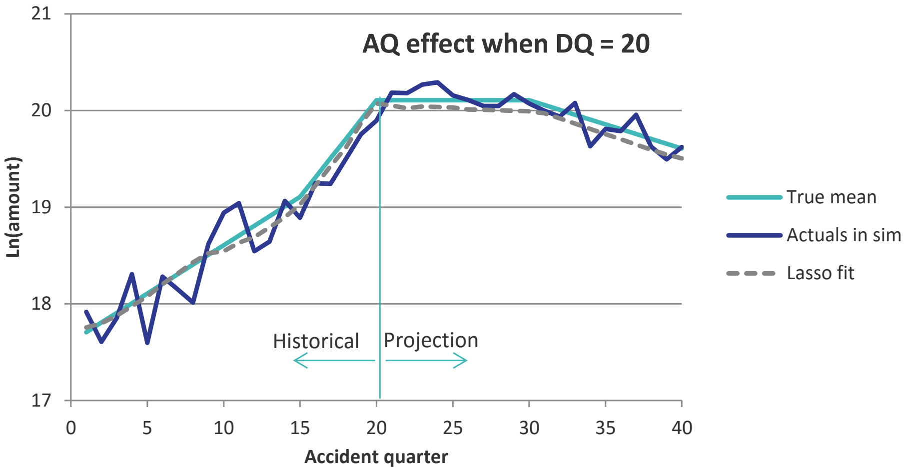
Data set 1: model selection (cont'd)



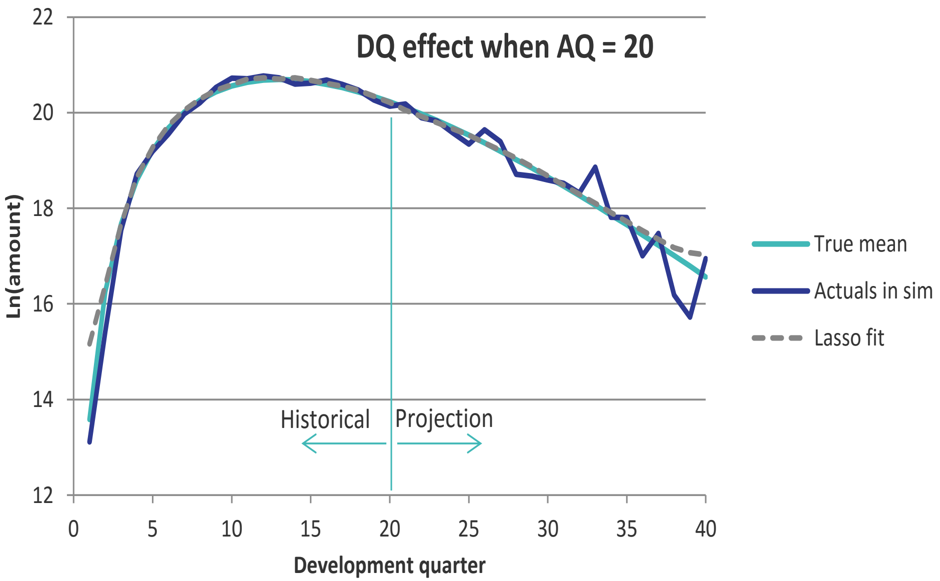
- Model selected to minimize CV error
  - Other approaches are possible to reduce model complexity
- Selected model contains 152 non-zero parameters

Data set 1: results

- AQ tracking



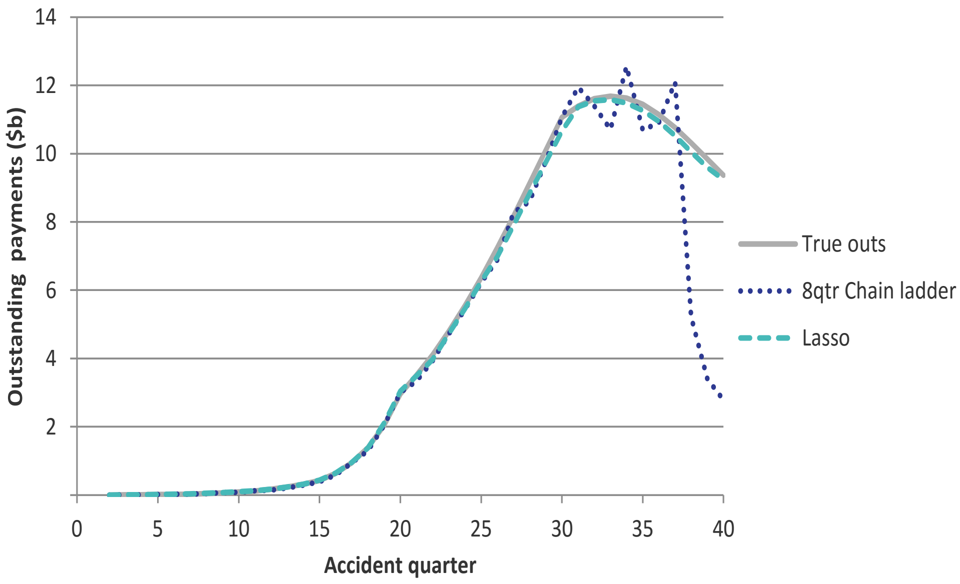
- DQ tracking



- Tracking appears reasonable

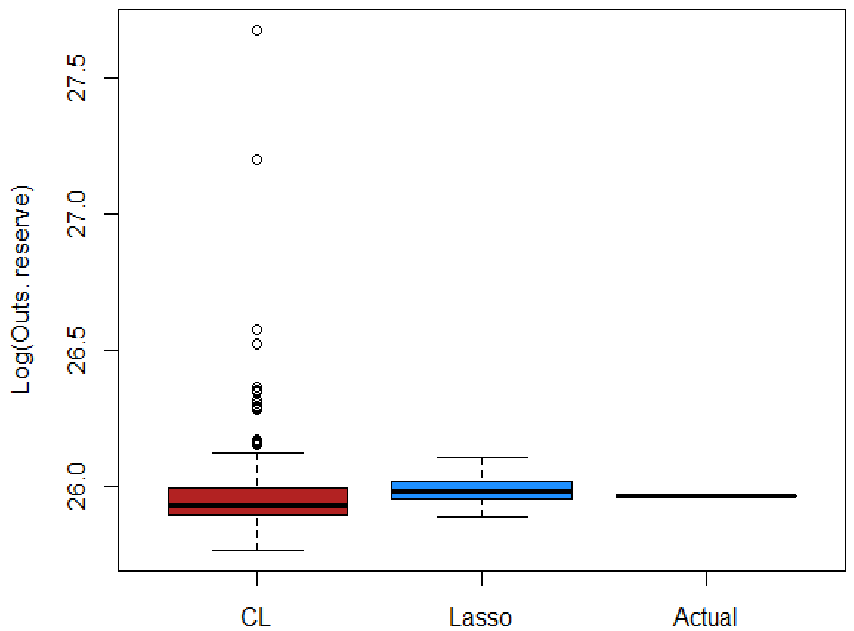
Data set 1: results (cont'd)

- Loss reserve by AQ



- Lasso forecast appears satisfactory

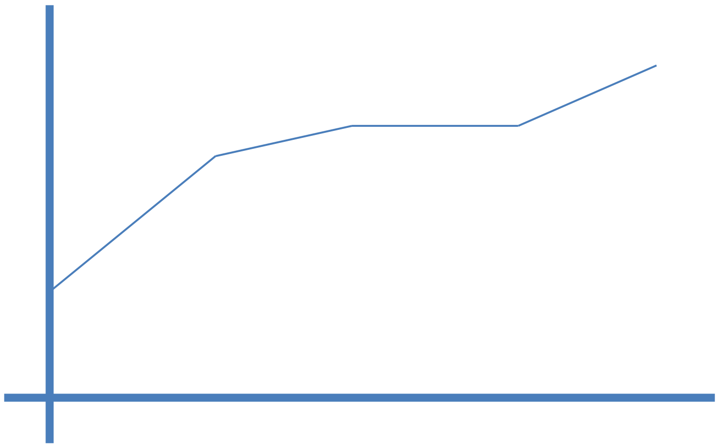
- Distribution of total loss reserve



- Lasso forecast tighter than chain ladder

Data set 2: set-up and model selection

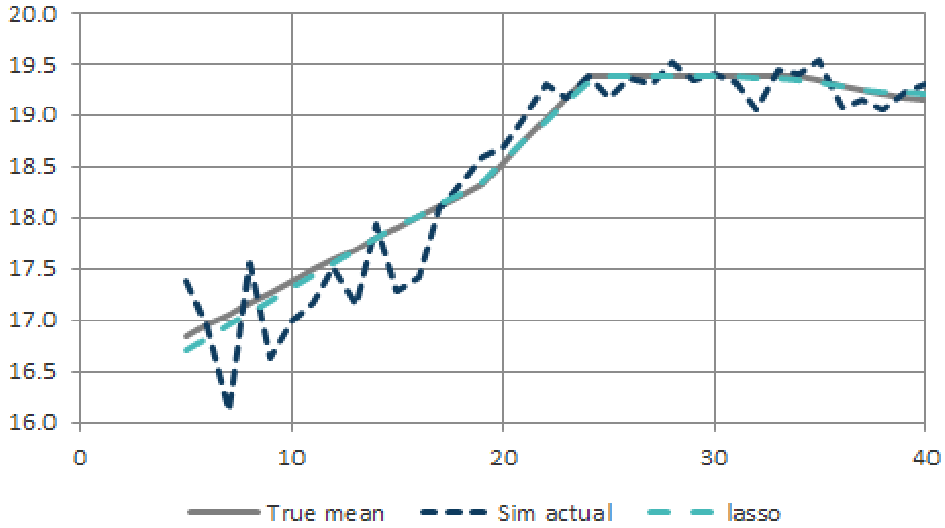
- Recall $ln\, \mu_{kj} = \alpha_k + \beta_j + \gamma_t$
- Assume
  - α_k as for data set 1
  - β_j as for data set 1
  - γ_t appears as in diagram



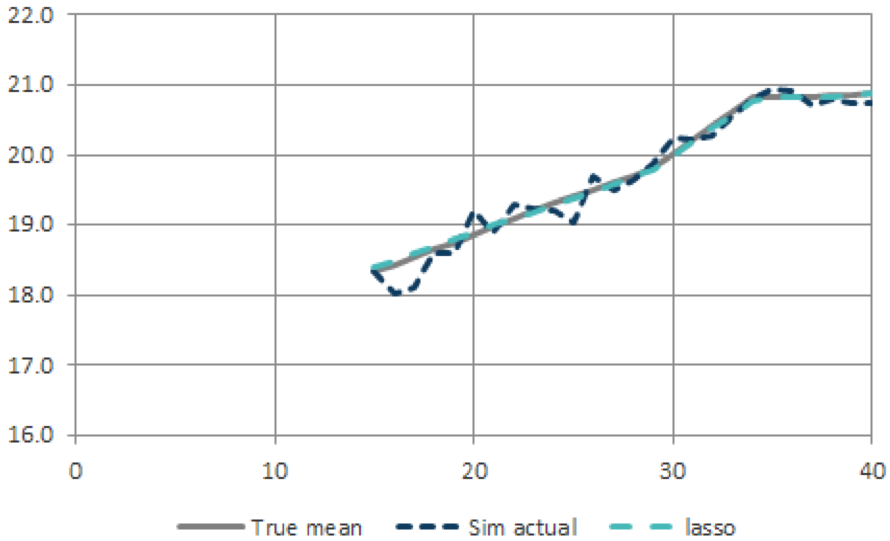
- Model includes:
  - about 3,200 basis functions
    - Experimentation suggests inclusion of an **unpenalized** constant SI term ($\gamma_t = t$) in regression
  - 84 non-zero parameters

Data set 2: results

- CQ tracking : at DQ 5



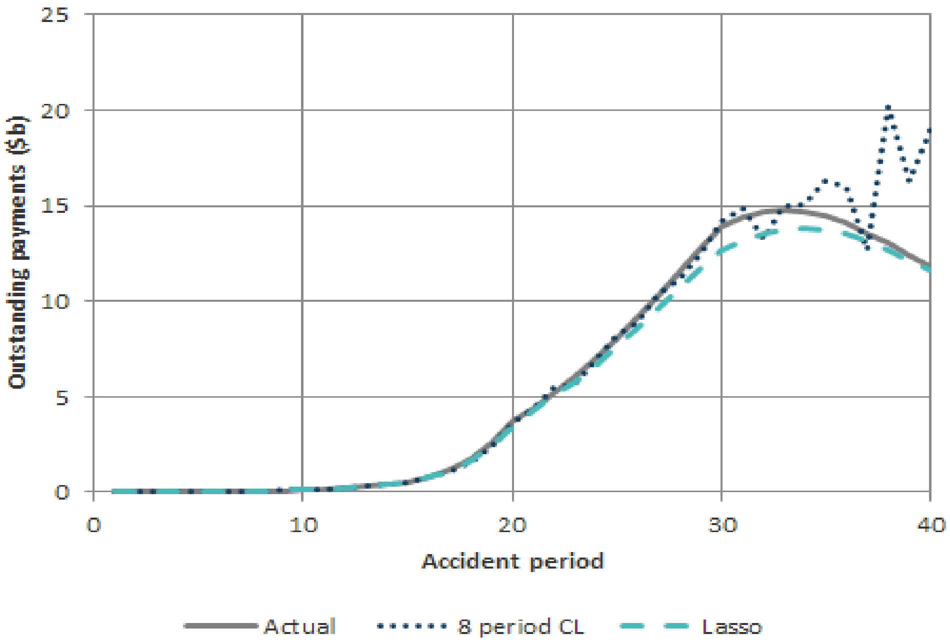
- CQ tracking : at DQ 15



- Tracking again appears reasonable

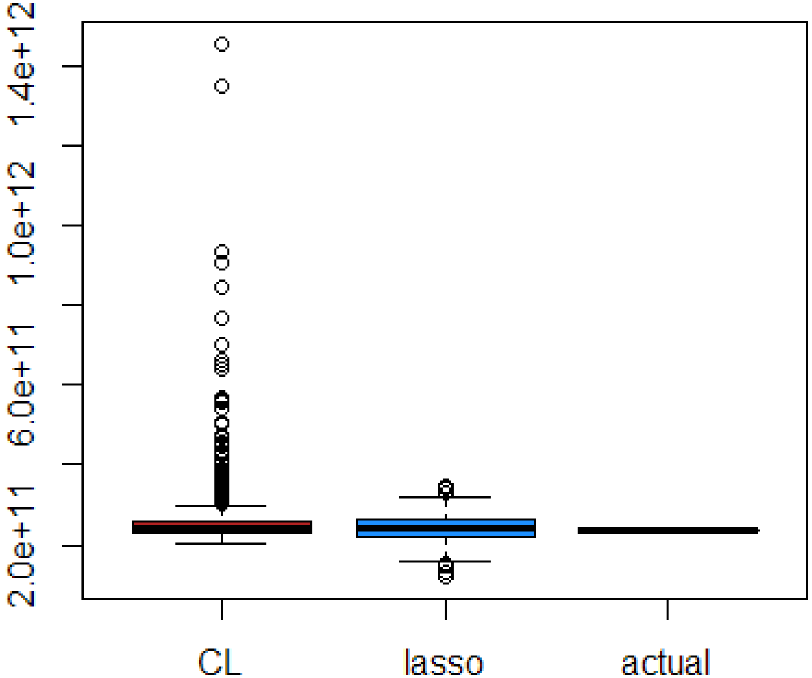
Data set 2: results (cont'd)

- Loss reserve by AQ



- Chain ladder now based on last 8 calendar quarters
- Lasso CQ trends stopped at last diagonal
  - Hence lasso biased downward relative to CL

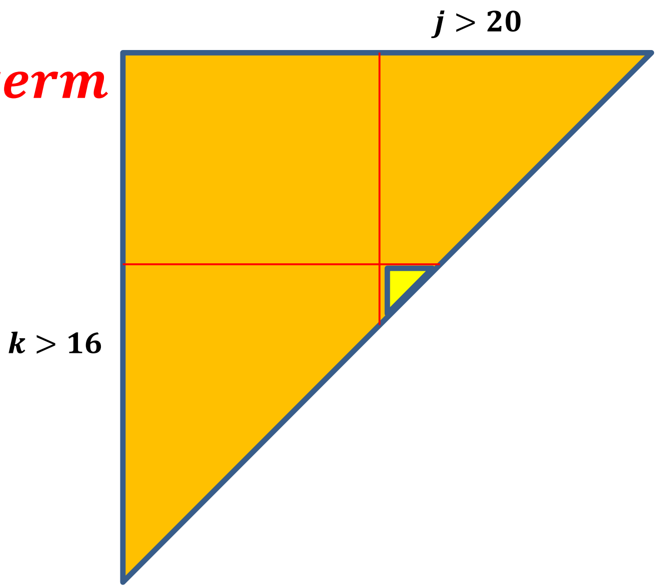
- Total loss reserve



- Chain ladder highly volatile

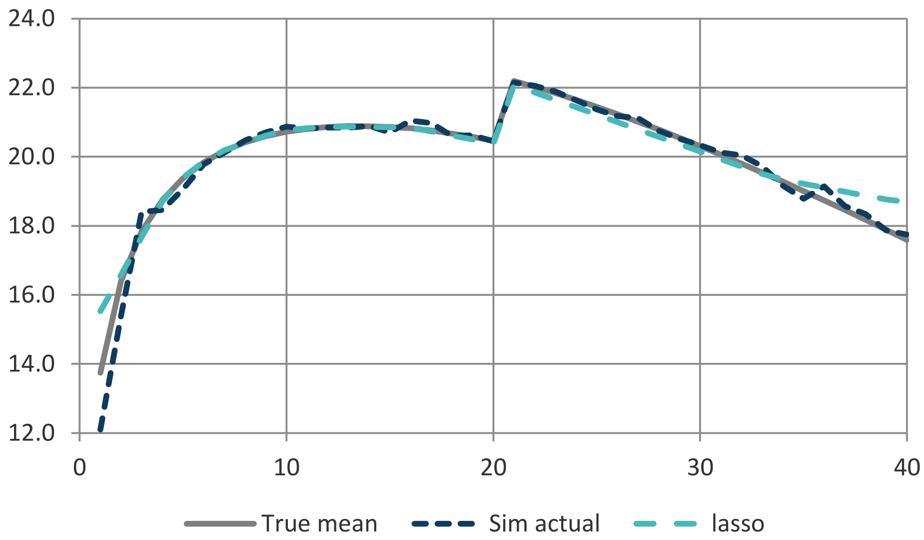
Data set 3: set-up and model selection

- This time $ln\ \mu_{kj} = \alpha_k + \beta_j + \gamma_t + \textit{interaction term}$
- Assume
  - α_k as for data sets 1 & 2
  - β_j as for data sets 1 & 2
  - γ_t as for data set 2
  - Interaction between AQ and DQ
    - For $k > 16$, β_j increases by 0.3 for $j > 20$
    - Difficult to detect: affects only 6 cells in the triangle of 820 cells
- Model includes:
  - about 3,200 basis functions
  - 103 non-zero parameters



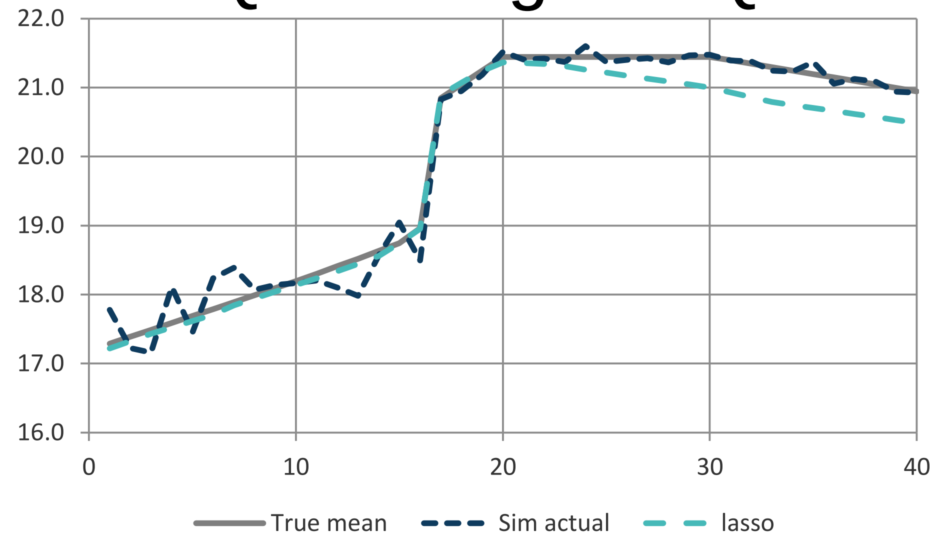
Data set 3: results

- DQ tracking at AQ 25



- DQ tracking surprisingly accurate

- AQ tracking at DQ 25



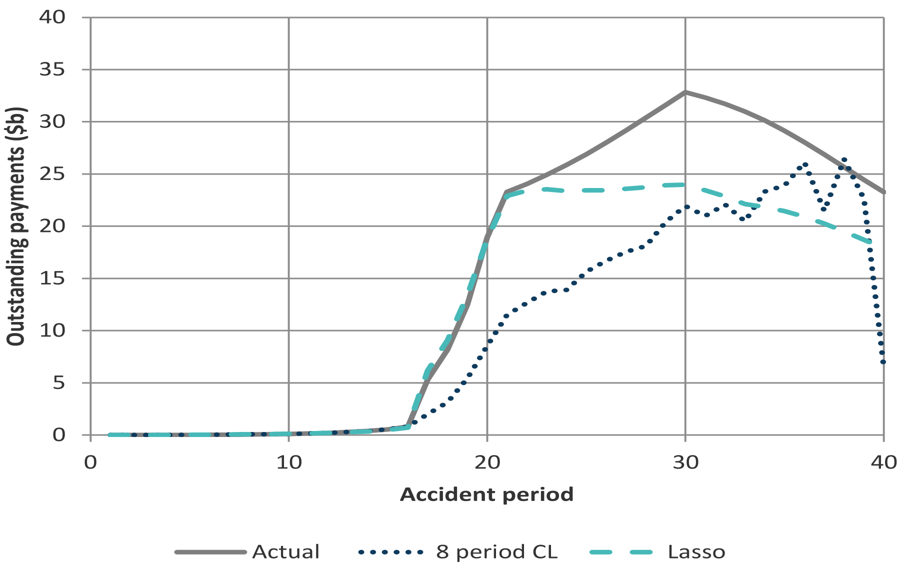
- Though under-estimation of tail at higher AQs

Taylor Fry
Analytics & Actuarial Consulting

UNSW AUSTRALIA

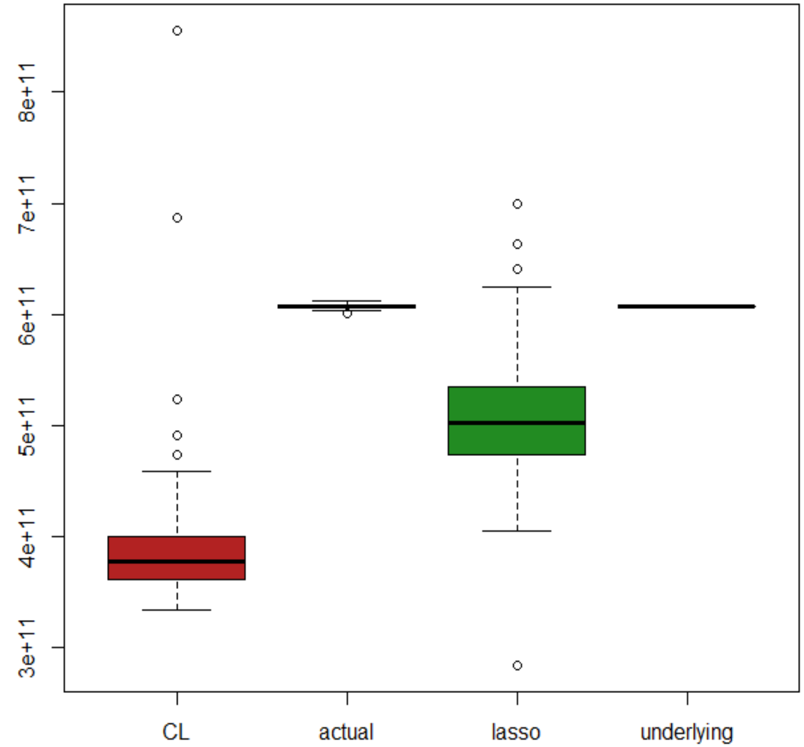
Data set 3: results (cont'd)

- Loss reserve by AQ



- Chain ladder now based on last 8 calendar quarters
- CL and lasso both under-estimate
  – But CL under-estimation greater

- Total loss reserve

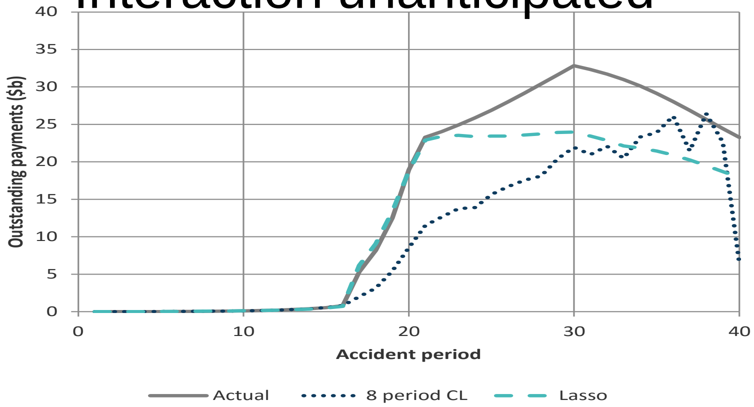


- Chain ladder highly volatile

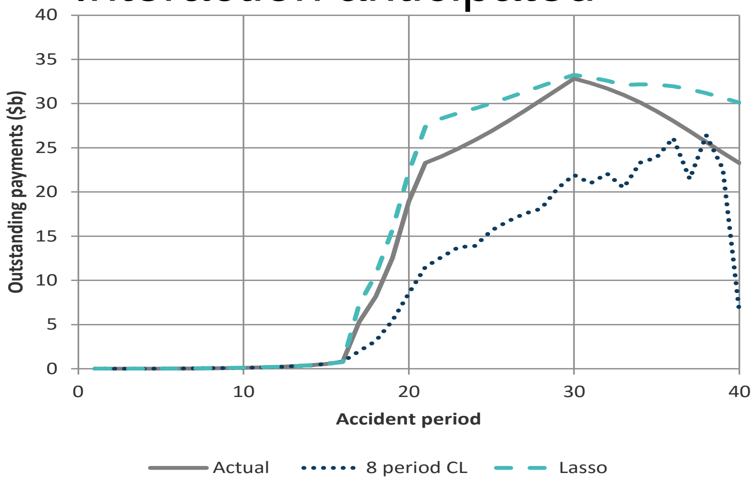
Data set 3: results (cont'd)

- The AQxDQ interaction has been penalized like all other regressors
- In practice, one might be able to anticipate the change
  - e.g. a legislated benefit change, taking effect in AQ 17
- In this case, one could apply **no penalty** to the interaction

- Loss reserve by AQ
  - Interaction unanticipated

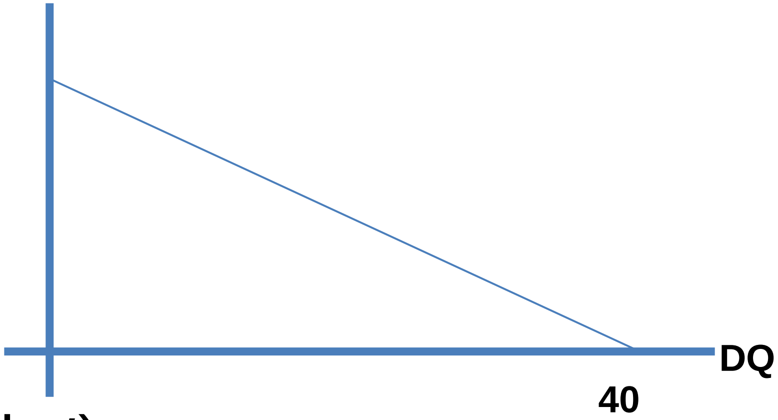


  - Interaction anticipated



Data set 4: set-up and model selection

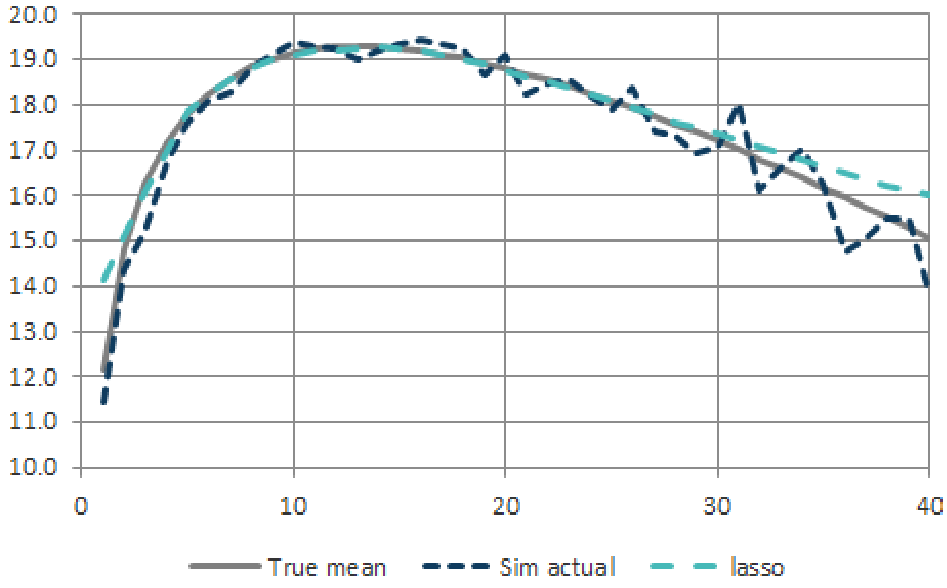
- This time $ln\, \mu_{kj} = \alpha_k + \beta_j + \boldsymbol{\lambda_j} \gamma_t$
- Assume
  - α_k as for data sets 1-3
  - β_j as for data sets 1-3
  - γ_t as for data sets 2 & 3
  - λ_j (multiplier that varies SI with j)



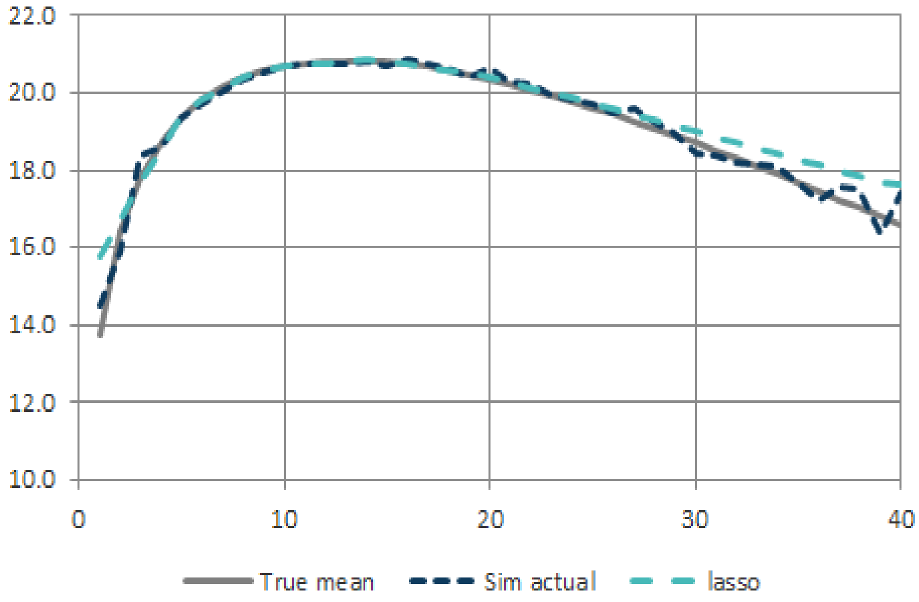
- Model includes:
  – about 3,200 basis functions
  – 87 non-zero parameters

Data set 4: results

- DQ tracking at AQ 10

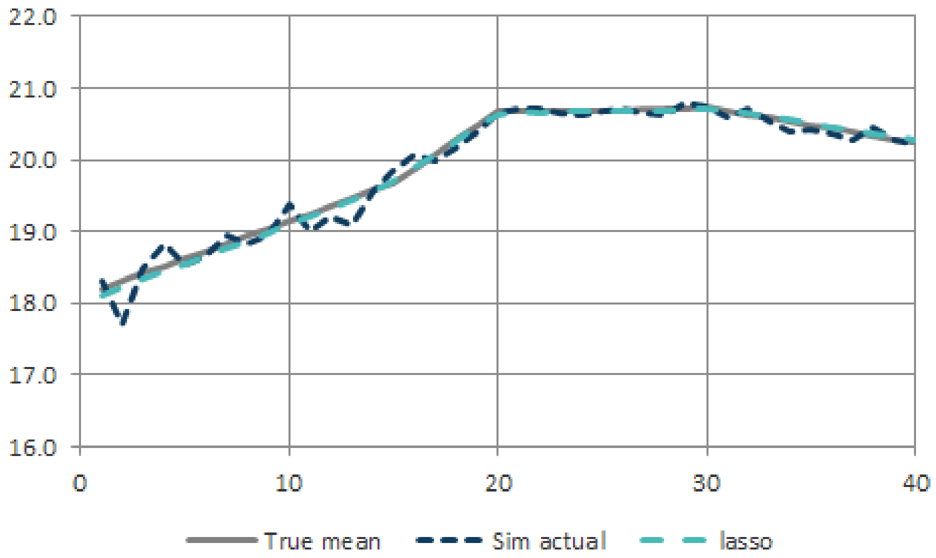


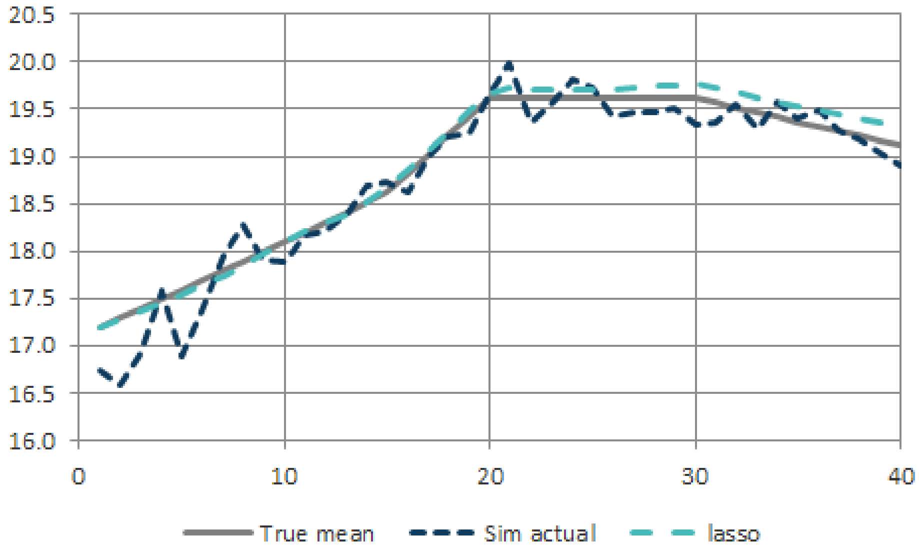
- DQ tracking at AQ 25



Data set 4: results (cont'd)

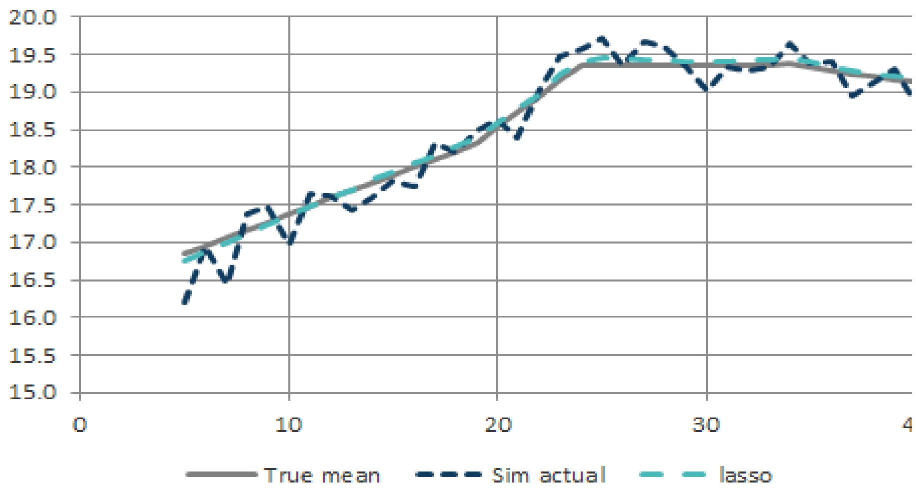
- AQ tracking at DQ 10
- AQ tracking at DQ 25



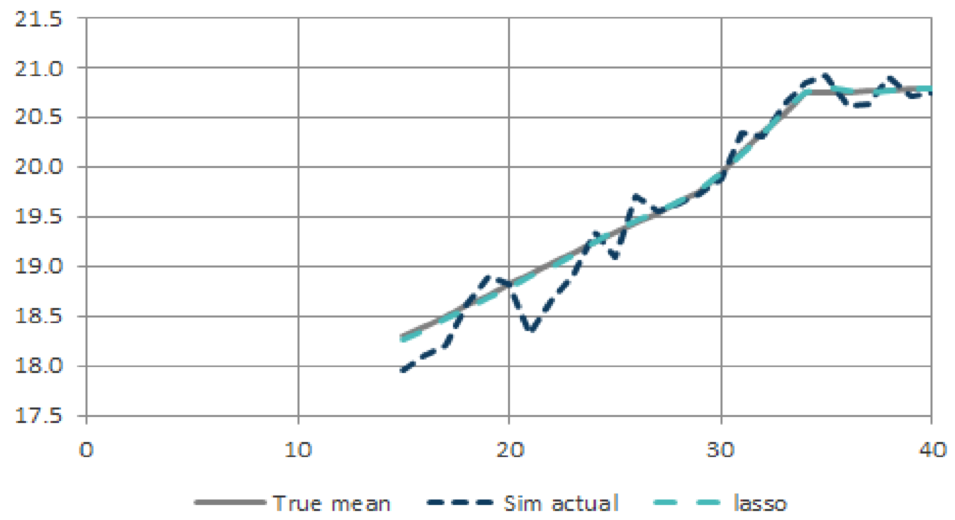


Data set 4: results (cont'd)

- CQ tracking at DQ 5

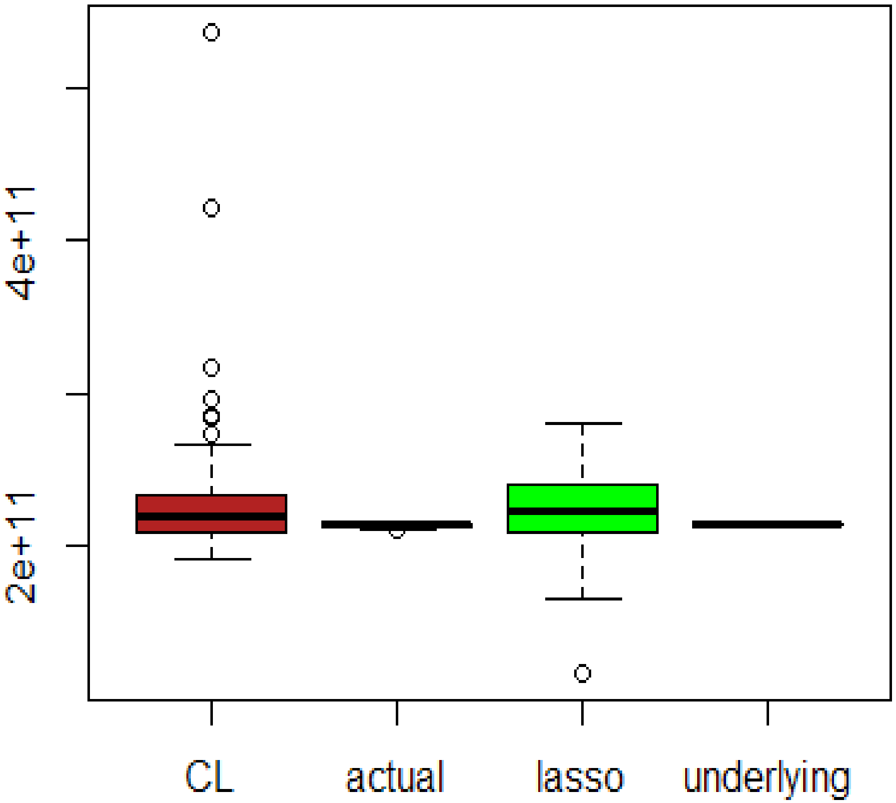


- CQ tracking at DQ 15



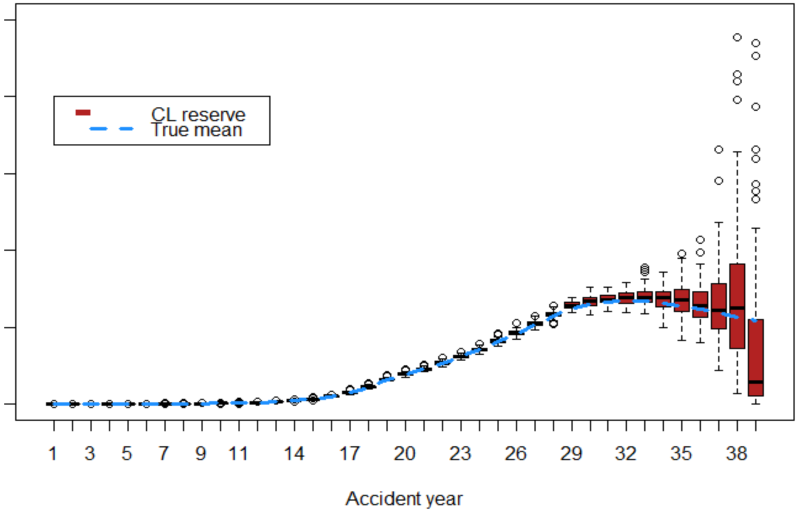
Data set 4: results (cont'd)

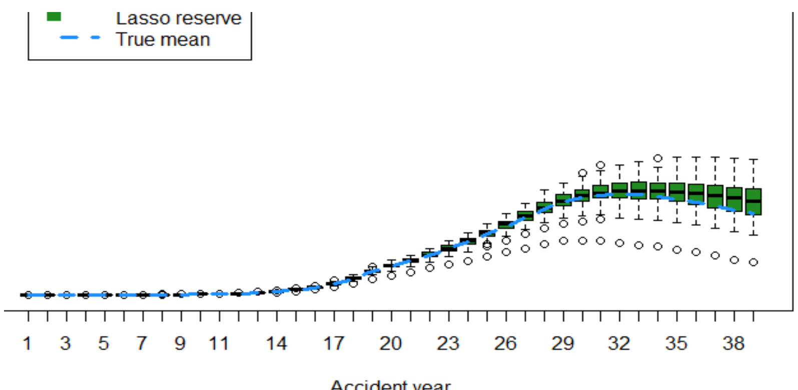
- Total loss reserve



- Chain ladder comparable with lasso but with some outlying forecasts

- Loss reserve by AQ





- Lasso more efficient predictor of individual accident quarters

Overview

- Motivation
- Regularized regression
  - In general
  - The lasso
- Loss reserving framework and notation
- Application of lasso to loss reserving
- Test on simulated data
  - Data set
  - Results
- **Further testing and development**
- Conclusion

Further testing and development

- Examination of additional scenarios
  - Particularly those likely to stress the chain ladder
- Different basis functions
  - e.g. Hoerl curve basis functions for DQ effects
- Consideration of future SI
  - How well adapted to extrapolation is the lasso?
- Robustification
- Multi-line reserving (with dependencies)
- Adaptive reserving
  - How might the lasso be adapted as a dynamic model?

Overview

- Motivation
- Regularized regression
  - In general
  - The lasso
- Loss reserving framework and notation
- Application of lasso to loss reserving
- Test on simulated data
  - Data set
  - Results
- Further testing and development
- **Conclusion**

Conclusion

- The lasso appears promising as a platform for self-assembling models
  - The model calibration procedure follows a routine and is relatively quick
    - Perhaps 30 minutes for routine calibration and examination of diagnostics
    - Perhaps an hour if one or two ad hoc changes require formulation and implementation
      - e.g. superimposed inflation, legislative change
- The lasso appears to track eccentric features of the data reasonably well
  - Including in scenarios where the chain ladder has little hope of an accurate forecast
- Some further experimentation required before full confidence can be invested in it as an automated procedure

References

- McGuire, G. 2007. “Individual Claim Modelling of CTP Data.” Institute of Actuaries of Australia XIth Accident Compensation Seminar, Melbourne, Australia. http://actuaries.asn.au/Library/6.a_ACS07_paper_McGuire_Individual%20claim%20modellingof%20CTP%20data.pdf
- Taylor, G., and G. McGuire. 2004. “Loss Reserving with GLMs: A Case Study.” Casualty Actuarial Society 2004 Discussion Paper Program, pp 327-392.
- Taylor, G., and G. McGuire. 2016. “Stochastic Loss Reserving Using Generalized Linear Models”. CAS Monograph Series, Number 3. Casualty Actuarial Society, Arlington VA.

Questions?