Model error and its estimation, with particular application to loss reserving

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

ASTIN Colloquium, 21-24 June 2022
Acknowledgements

• This is joint work with Gráinne McGuire, Taylor Fry, Sydney, Australia

• It has received financial support under Australian Research Council’s Linkage Projects funding scheme (project number LP130100723)

• I have benefited from discussions with colleagues:
  – Benjamin Avanzi
  – Bernard Wong
  – David Yu
Overview

• Motivation
• Components of forecast error
• Internal model error
• Model set generation
• LASSO
• Bootstrapping the LASSO
• Numerical illustrations
• Conclusions
Overview

- **Motivation**
- Components of forecast error
- Internal model error
- Model set generation
- LASSO
- Bootstrapping the LASSO
- Numerical illustrations
- Conclusions
Motivation (1)

- Loss reserve **risk margins**
  - Often set according to **Value at Risk (VaR)**
    - e.g. probability of sufficiency of loss reserve including risk margin = 75%
    - VaR a function of the distribution of liability
    - The variation about the mean of this distribution represents **forecast error**
    - **Model error** is a component of forecast error
    - This address discusses model error
Motivation (2)

• My ASTIN 2021 keynote address discussed the role of **model error** in formulation of a risk margin
  – Gave theoretical background
  – That was a work in progress
• The current address is a sequel and completes the earlier work
• The work draws heavily on
Overview

- Motivation
- **Components of forecast error**
- Internal model error
- Model set generation
- LASSO
- Bootstrapping the LASSO
- Numerical illustrations
- Conclusions
Forecast error and its components

- Definition:
  \[ \text{forecast error} = \text{observation} - \text{forecast} \]

- Forecast error may be decomposed as follows:
  \[ \text{Forecast error} = \text{Process error} + \text{Parameter error} + \text{Model error} \]

Model error includes:
- Model structure error
  - Internal
  - External

- Model distribution error
  - Internal
  - External
Forecast error components: definitions

• **Process error**: irreducible error due to stochastic nature of process

• **Parameter error**: sampling error in model parameter estimates

• **Model error**: error induced by incorrect model specification
  – **Model structure error**: error induced by incorrect specification of algebraic form of mean
  – **Model distribution error**: error induced by incorrect specification of model distribution
    • **Internal model error**: model error internal to past data
    • **External model error**: model error relating to future conditions (e.g. future superimposed inflation)
Forecast error components: definitions

• **Process error**: irreducible error due to stochastic nature of process

• **Parameter error**: sampling error in model parameter estimates

• **Model error**: error induced by incorrect model specification
  – **Model structure error**: error induced by incorrect specification of algebraic form of mean
  – **Model distribution error**: error induced by incorrect specification of model distribution
    • **Internal model error**: model error internal to past data
    • **External model error**: model error relating to future conditions (e.g. future superimposed inflation)
Overview

• Motivation
• Components of forecast error
• **Internal model error**
• Model set generation
• LASSO
• Bootstrapping the LASSO
• Numerical illustrations
• Conclusions
Internal model error (Bayesian setting): essential ingredients (1)

- **Model set**: population of candidate models
  - containing **primary model**
    - Model whose error is to be measured
- **Prior distribution**: on model set
- **Posterior distribution** follows from:
  - Prior distribution
  - Model likelihood (of data $y$)
- **Form of Bayesian Model Average**
Internal model error (Bayesian setting): essential ingredients (2)

- The posterior distribution on the model set ensures consistency with past data
  - Models that fit poorly are assigned low posterior probability
- However, some models may fit past data well, but extrapolate the future poorly
  - Hence a need to prune the model set to exclude models that do not extrapolate credibly

\[ \text{Adjusted posterior } p(M | y), M \in \mathcal{M} \]
Overview

• Motivation
• Components of forecast error
• Internal model error
• **Model set generation**
• LASSO
• Bootstrapping the LASSO
• Numerical illustrations
• Conclusions
Source of model set

• Can we find anything that generates a set of models from which the primary model can be regarded as a 1-sample, e.g.
  – Neural Network
    • Different models generated by different sets of hyperparameters
  – LASSO
    • We use LASSO (more to come)
Overview

• Motivation
• Components of forecast error
• Internal model error
• Model set generation
• **LASSO**
• Bootstrapping the LASSO
• Numerical illustrations
• Conclusions
**LASSO: formulation**

- Model form (same as GLM)
  
  \[
  y = h^{-1}(X\beta) + \varepsilon
  \]
  
  where
  
  - \( y \) = data vector
  - \( h \) = link function
  - \( \beta \) = parameter vector
  - \( X \) = design matrix
  - \( \varepsilon \) = stochastic disturbance with \( E[\varepsilon] = 0 \)

- Estimate \( \beta \) by minimization of **regularized** negative log-likelihood
  
  \[
  \hat{\beta}(\lambda) = \arg\min_{\beta} [\ell(y|\beta) + \lambda^T|\beta|]
  \]
  
  where
  
  - \( \ell \) = negative log-likelihood (NLL)
  - \( |. | \) operates elementwise on \( \beta \)
  - \( \lambda \) = **penalty parameter vector** with non-negative components
LASSO: Bayesian interpretation

- Assume parameter vector $\beta$ subject to Laplace prior distribution

$$
\pi(\beta) \propto \exp(-\lambda^T|\beta|)
$$

where
- $|.|$ operates elementwise on $\beta$

- Prior distribution:
  - Consists of a two-sided exponential distribution for each $\beta_j$
  - $\text{Var}[\beta_j] = 2/\lambda_j^2$ (small variance $\Rightarrow$ large penalty)

- In this framework, maximum a posteriori (MAP) estimator of $\beta$ is same as LASSO estimator
LASSO: selection of penalty parameter (1)

\[ \hat{\beta}(\lambda) = \arg \min_{\beta} \left[ \ell(y|\beta) + \lambda^T|\beta| \right] \]

- In practice \( \lambda \) vector is unknown
- We choose \( \lambda^T = \lambda(0, 1, ..., 1) \)
  - \( \lambda \) now a scalar
  - No penalty on intercept parameter; equal penalties on all other parameters
- Selection of value for scalar \( \lambda \) still required
- \( \lambda = 0 \iff \) maximum likelihood
  - As \( \lambda \uparrow \), model simplifies, NLL increases
LASSO: selection of penalty parameter (2)

- For given $\lambda$, let
  - $C(\lambda)$ denote the average k-fold cross-validation error (we use $k = 8$) for parameterization $\hat{\beta}(\lambda)$
  - $S(\lambda)$ denote standard deviation of the cross-validation error across the $k$ folds

- 4 alternative values of scalar $\lambda$ are selected:
  1) “lambda.min”: $\lambda_{\text{min}} = \arg \min_{\lambda} C(\lambda)$
  2) “1se”: $\lambda_{1\text{se}} = \arg \max_{\lambda} \{C(\lambda) \leq C(\lambda_{\text{min}}) + S(\lambda_{\text{min}})\}$
  3) “simple”: $\lambda_{\text{simp}} = \max \{\lambda: p(M_{1\text{se}}|y; \lambda) > \delta\}$
  4) “complex”: $\lambda_{\text{comp}} = \min \{\lambda: p(M_{1\text{se}}|y; \lambda) > \delta\}$

We use $\delta = 0.0005$
LASSO: selection of penalty parameter (3)

• Examples of “1se” and “lambda.min” models
LASSO: selection of penalty parameter (4)

• Examples of “simple” and “complex” models
LASSO: as generator of model set

• The LASSO generates one model for each value of \( \lambda \)
  – The model set may be taken as the set of all models
    generated by the values of \( \lambda \) considered
• Now consider a single selected \( \lambda \)
  – This defines a prior \( \pi(M), M \in \mathcal{M} \) on the model set \( \mathcal{M} \)
  – Together with data \( y \), this defines a posterior \( p(M|y), M \in \mathcal{M} \) on the model set \( \mathcal{M} \)
• For each model \( M \), there is an associated **expected loss reserve** \( E_M \)
  – The posterior \( p(M|y) \) induces a distribution on \( E_M, M \in \mathcal{M} \)
  – This is the distribution of internal model error around primary estimate \( E_M^* \)
Overview

• Motivation
• Components of forecast error
• Internal model error
• Model set generation
• LASSO
  • **Bootstrapping the LASSO**
• Numerical illustrations
• Conclusions
Bootstrap: motivation and process

• The LASSO generates an estimated distribution of internal model error, as described
  – However, this estimate is “thin” because it typically depends on a relatively small sample of models, perhaps 10-30

• Hence **bootstrap** to obtain multiple replications of the internal model error distributions
  – Re-sample **centred** standardized residuals from primary model
    • Checking carefully that they appear iid
Bootstrap: interpretation of results

• Generate a bootstrap matrix
  – One row per bootstrap replication
  – Columns contain models for different values of \( \lambda \)
  – Columns can be reduced to summary statistics over the posterior distribution of models (dependent on selected prior), e.g. for each row
    • \( E[E_M; p] \) = posterior mean of the loss reserves \( E_M \)
    • \( Var[E_M; p] \) = posterior variance of the loss reserves \( E_M \)
    • Any other stuff of interest

• Interpretation
  – \( Var[E_M; p] \) measures internal model error (for the relevant replication)
  – \( E_p Var[E_M; p] \) = internal model error
  – \( Var_p E[E_M; p] \) = parameter error

\[ E_p, Var_p \] taken over rows
Bootstrap: pruning of results

• The most delinquent forms of extrapolation are likely to relate to:
  – The most recent accident periods
    • Where there is little accumulated data
  – Future payment periods

• We therefore set inclusion gates according to the following ratios of bootstrap replication future cash flows to those of the primary model:

<table>
<thead>
<tr>
<th>Accident periods</th>
<th>Gate</th>
<th>Payment periods</th>
<th>Gate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last 2</td>
<td>[0.75,1.33]</td>
<td>Next 2</td>
<td>[0.91,1.10]</td>
</tr>
<tr>
<td>Last 5</td>
<td>[0.80,1.25]</td>
<td>Next 5</td>
<td>[0.87,1.15]</td>
</tr>
<tr>
<td>Last 10</td>
<td>[0.83,1.20]</td>
<td>Next 10</td>
<td>[0.83,1.20]</td>
</tr>
</tbody>
</table>

• Note the subjective nature of the gates
Overview

• Motivation
• Components of forecast error
• Internal model error
• Model set generation
• LASSO
• Bootstrapping the LASSO
• **Numerical illustrations**
• Conclusions
Numerical illustrations: data

• 4 synthetic data sets from McGuire, Taylor & Miller (2021):
  – Full details there
  – **Data set 1**: chain ladder compatible
  – **Data set 2**: payment quarter effect included
  – **Data set 3**: AQ-DQ interaction added but only in 10 cells out of 820
  – **Data set 4**: superimposed inflation included, but with rate of SI depending on DQ
Numerical results: data set 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Prior</th>
<th>True reserve ($B)</th>
<th>Forecast ($B)</th>
<th>Estimated internal model error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Simple</td>
<td>190</td>
<td>193</td>
<td>0.51%</td>
</tr>
<tr>
<td></td>
<td>1se</td>
<td>190</td>
<td>191</td>
<td>0.30%</td>
</tr>
<tr>
<td></td>
<td>Lambda.min</td>
<td>190</td>
<td>194</td>
<td>0.82%</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>190</td>
<td>190</td>
<td>1.14%</td>
</tr>
<tr>
<td>Boot-</td>
<td>Simple</td>
<td>190</td>
<td>191</td>
<td>0.37%</td>
</tr>
<tr>
<td>Strap</td>
<td>1se</td>
<td>190</td>
<td>189</td>
<td>0.32%</td>
</tr>
<tr>
<td></td>
<td>Lambda.min</td>
<td>190</td>
<td>192</td>
<td>0.41%</td>
</tr>
<tr>
<td></td>
<td>Complex</td>
<td>190</td>
<td>188</td>
<td>0.53%</td>
</tr>
</tbody>
</table>

- Very small model error for data compatible with chain ladder
Numerical results: data set 1: posterior cdf’s
**Numerical results for all data sets 1-4**

- **Bootstrap results for 1se and lambda.min**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Prior</th>
<th>True ($B$)</th>
<th>Mean ($B$)</th>
<th>Internal model error (CoV)</th>
<th>Parameter error (CoV)</th>
<th>Process error (CoV)</th>
<th>Total error (CoV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1se</td>
<td>190</td>
<td>189</td>
<td>0.32%</td>
<td>5.30%</td>
<td>3.29%</td>
<td>6.24%</td>
</tr>
<tr>
<td></td>
<td>lambda.min</td>
<td>190</td>
<td>192</td>
<td>0.41%</td>
<td>5.15%</td>
<td>2.75%</td>
<td>5.85%</td>
</tr>
<tr>
<td></td>
<td>2se</td>
<td>238</td>
<td>252</td>
<td>1.45%</td>
<td>10.00%</td>
<td>3.93%</td>
<td>10.84%</td>
</tr>
<tr>
<td></td>
<td>lambda.min</td>
<td>238</td>
<td>240</td>
<td>1.79%</td>
<td>8.83%</td>
<td>4.69%</td>
<td>10.16%</td>
</tr>
<tr>
<td></td>
<td>3se</td>
<td>608</td>
<td>703</td>
<td>2.27%</td>
<td>11.23%</td>
<td>5.71%</td>
<td>12.80%</td>
</tr>
<tr>
<td></td>
<td>lambda.min</td>
<td>608</td>
<td>589</td>
<td>2.12%</td>
<td>11.19%</td>
<td>5.27%</td>
<td>12.54%</td>
</tr>
<tr>
<td></td>
<td>4se</td>
<td>216</td>
<td>243</td>
<td>1.37%</td>
<td>8.63%</td>
<td>4.01%</td>
<td>9.62%</td>
</tr>
<tr>
<td></td>
<td>lambda.min</td>
<td>216</td>
<td>252</td>
<td>1.81%</td>
<td>12.54%</td>
<td>5.08%</td>
<td>13.65%</td>
</tr>
</tbody>
</table>
Numerical results for all data sets 1-4: comparison of LASSO and GLM forecasts

- GLM fitted to data containing same covariates as LASSO 1se model
- Forecast and estimated parameter error extracted

<table>
<thead>
<tr>
<th>Data set</th>
<th>Prior</th>
<th>Forecast</th>
<th></th>
<th></th>
<th>Parameter error (CoV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>True ($B$)</td>
<td>Mean ($B$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>LASSO 1se GLM</td>
<td>190</td>
<td>189</td>
<td>5.30%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>190</td>
<td>212</td>
<td>4.94%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>LASSO 1se GLM</td>
<td>238</td>
<td>252</td>
<td>10.00%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>238</td>
<td>284</td>
<td>7.49%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>LASSO 1se GLM</td>
<td>608</td>
<td>703</td>
<td>11.23%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>608</td>
<td>1007</td>
<td>3.56%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>LASSO 1se GLM</td>
<td>216</td>
<td>243</td>
<td>8.63%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GLM</td>
<td>216</td>
<td>274</td>
<td>8.40%</td>
<td></td>
</tr>
</tbody>
</table>

Law of total variance:

\[
\text{Var}[\hat{R}] = E_M[\text{Var}[\hat{R}|M]] + \text{Var}_M[E[\hat{R}|M]]
\]

where

\(\hat{R}\) = forecast

\(M\) = bootstrap pseudo-model

Apply to parameter error to see that bootstrap estimate will usually be greater
Sensitivity to selected inclusion gates

• Recall the selected inclusion gates

<table>
<thead>
<tr>
<th>Accident periods</th>
<th>Gate</th>
<th>Payment periods</th>
<th>Gate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last 2</td>
<td>[0.75, 1.33]</td>
<td>Next 2</td>
<td>[0.91, 1.10]</td>
</tr>
<tr>
<td>Last 5</td>
<td>[0.80, 1.25]</td>
<td>Next 5</td>
<td>[0.87, 1.15]</td>
</tr>
<tr>
<td>Last 10</td>
<td>[0.83, 1.20]</td>
<td>Next 10</td>
<td>[0.83, 1.20]</td>
</tr>
</tbody>
</table>

• Suppose we multiply (divide) each upper (lower) bound by 1.1

<table>
<thead>
<tr>
<th>Data set</th>
<th>Prior Gate</th>
<th>Prior Total error (CoV)</th>
<th>Widened Total error (CoV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LASSO 1se lambda.min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6.24%</td>
<td>7.50%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.85%</td>
<td>7.03%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10.84%</td>
<td>15.23%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.16%</td>
<td>17.57%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12.80%</td>
<td>17.86%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.54%</td>
<td>19.22%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9.62%</td>
<td>15.00%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.65%</td>
<td>19.43%</td>
<td></td>
</tr>
</tbody>
</table>
Overview

• Motivation
• Components of forecast error
• Internal model error
• Model set generation
• LASSO
• Bootstrapping the LASSO
• Numerical illustrations
• Conclusions
Conclusions (1)

• An estimate of internal model error of a loss reserve has been constructed that is
  – Rigorous
  – Based on Bayesian Model Averaging
• It is objective in all respects except one
  – The inclusion gates
    • One must be willing to select limits on credible extrapolations of claim experience
    • These must be selected carefully, as the estimated internal model error is sensitive to them
Conclusions (2)

• Part of model error leaks into parameter error
  – Both are estimated
  – They appear broadly reasonable relative to GLM estimation

• “Total” forecast error (model + parameter + process) has been estimated in numerical examples, but
  – WARNING: external model error not considered
    • This requires study by different means
Thank you