Business School

Never Stand Still | Business School | Risk and Actuarial Studies

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

TAYLOR FRY

UNSW Australia | AGSM
Business School

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

TAYLOR FRY
UNSW AUSTRALIA

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
  - McGuire G, Taylor G, & Miller H (2021). Self-assembling insurance claim models using regularized regression and machine learning. (with G McGuire & H Miller). **Variance**, 14(1).

Overview

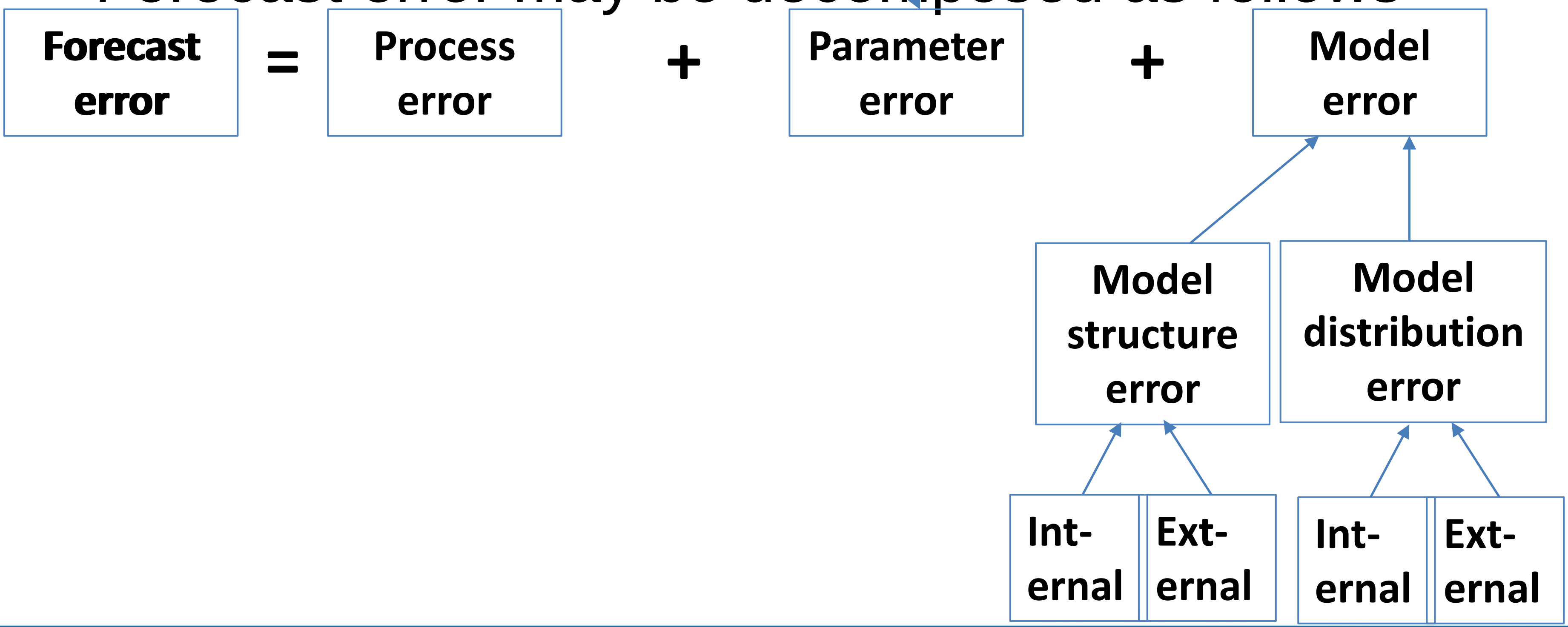
- Motivation
- **Components of forecast error**
- Internal model error
- Model set generation
- LASSO
- Bootstrapping the LASSO
- Numerical illustrations
- Conclusions

Forecast error and its components

- Definition:

forecast error = observation – forecast

- Forecast error may be decomposed as follows



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

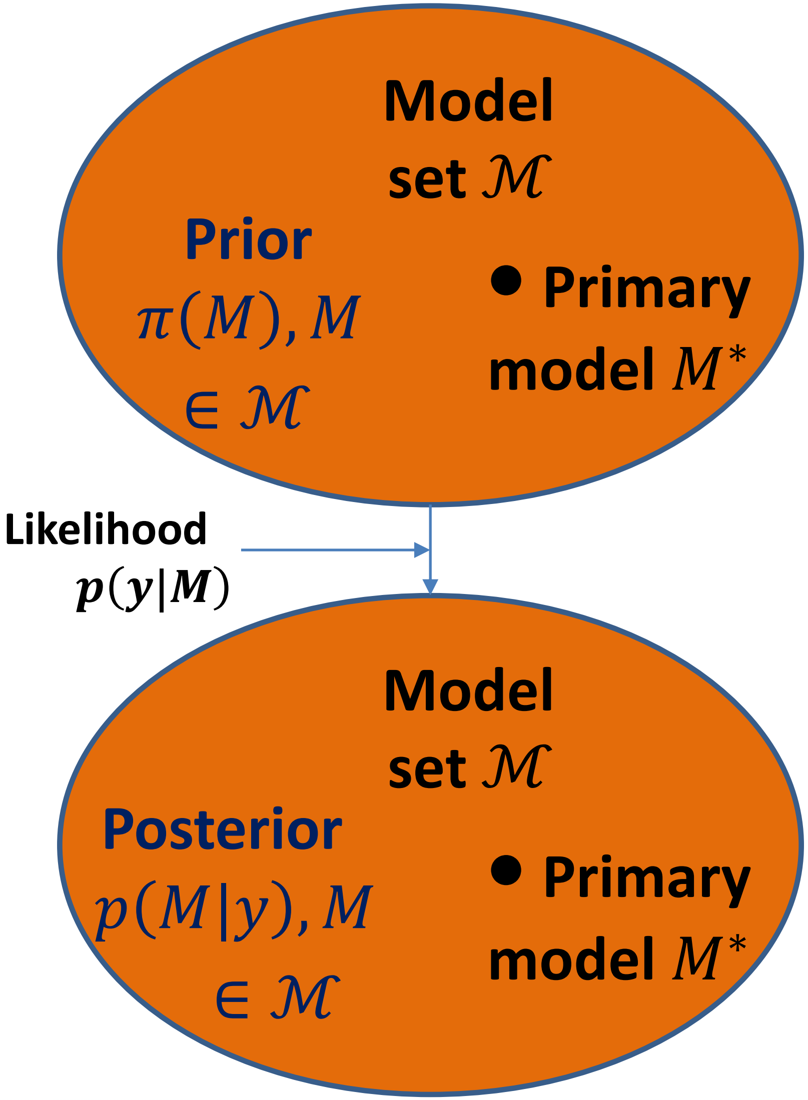
- **Process error**: irreducible error due to stochastic nature of process
- **Parameter error**: sampling error in model parameter estimates
- **Model error**: error induced by incorrect model specification
  - **Model structure error**: error induced by incorrect specification of algebraic form of 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

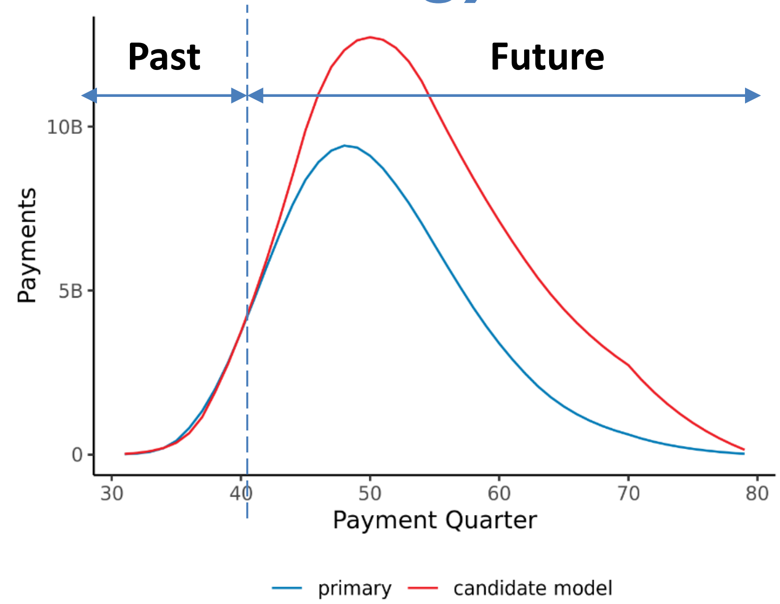
Internal model error (Bayesian setting): essential ingredients (1)

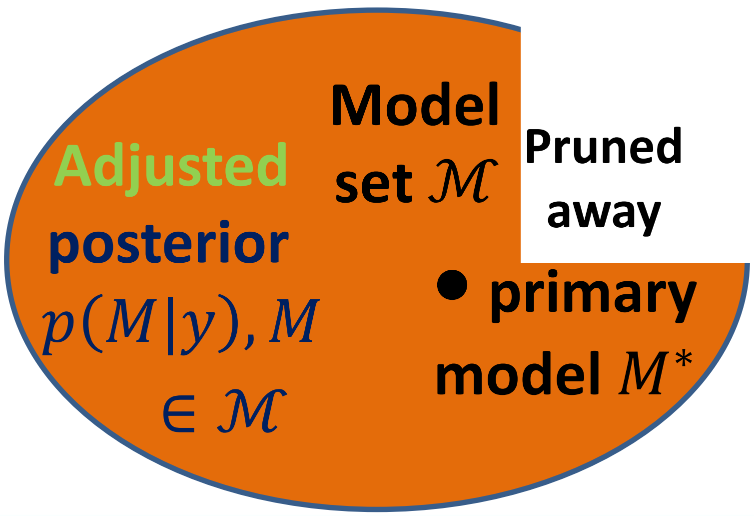
- **Model set:** population of candidate models
  - containing **primary model**
    - Model whose error is to be measured
- **Prior distribution:** on model set
- **Posterior distribution** follows from:
  - Prior distribution
  - Model likelihood (of data *y*)
- Form of **Bayesian Model Average**



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





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)

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

where

- y = data vector
- h = link function
- β = parameter vector
- X = design matrix
- ε = stochastic disturbance with $\mathrm{E}[\varepsilon] = 0$

– Estimate β by minimization of **regularized** negative log-likelihood

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = \mathbf{arg}\,\min_{\boldsymbol{\beta}}[\ell(\boldsymbol{y}|\boldsymbol{\beta}) + \boldsymbol{\lambda}^{T}|\boldsymbol{\beta}|]$$

where

- ℓ = negative log-likelihood **(NLL)**
- $|.|$ operates elementwise on β
- λ = **penalty parameter vector** with non-negative components

LASSO: Bayesian interpretation

- Assume parameter vector β subject to **Laplace prior** distribution

$$\boldsymbol{\pi(\beta) \propto exp\left(-\lambda^T|\beta|\right)}$$

where

  – $|.|$ operates elementwise on β

- Prior distribution:
  – Consists of a two-sided exponential distribution for each β_j
  – $Var\left[\beta_j\right] = 2/\lambda_j^2$ (small variance ⇒ large penalty)
- In this framework, **maximum a posteriori (MAP) estimator of $\boldsymbol{\beta}$ is same as LASSO estimator**

LASSO: selection of penalty parameter (1)

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = \mathbf{arg}\,\min_{\boldsymbol{\beta}}[\ell(\boldsymbol{y}|\boldsymbol{\beta}) + \boldsymbol{\lambda}^{T}|\boldsymbol{\beta}|]$$

- In practice λ vector is unknown
- We choose $\boldsymbol{\lambda}^{T} = \boldsymbol{\lambda}(\mathbf{0}, \mathbf{1}, \dots, \mathbf{1})$
  - λ now a scalar
  - No penalty on intercept parameter; equal penalties on all other parameters
- Selection of value for scalar λ still required
- $\lambda = 0 \iff$ maximum likelihood
  - As $\lambda \uparrow$, model simplifies, NLL increases

LASSO: selection of penalty parameter (2)

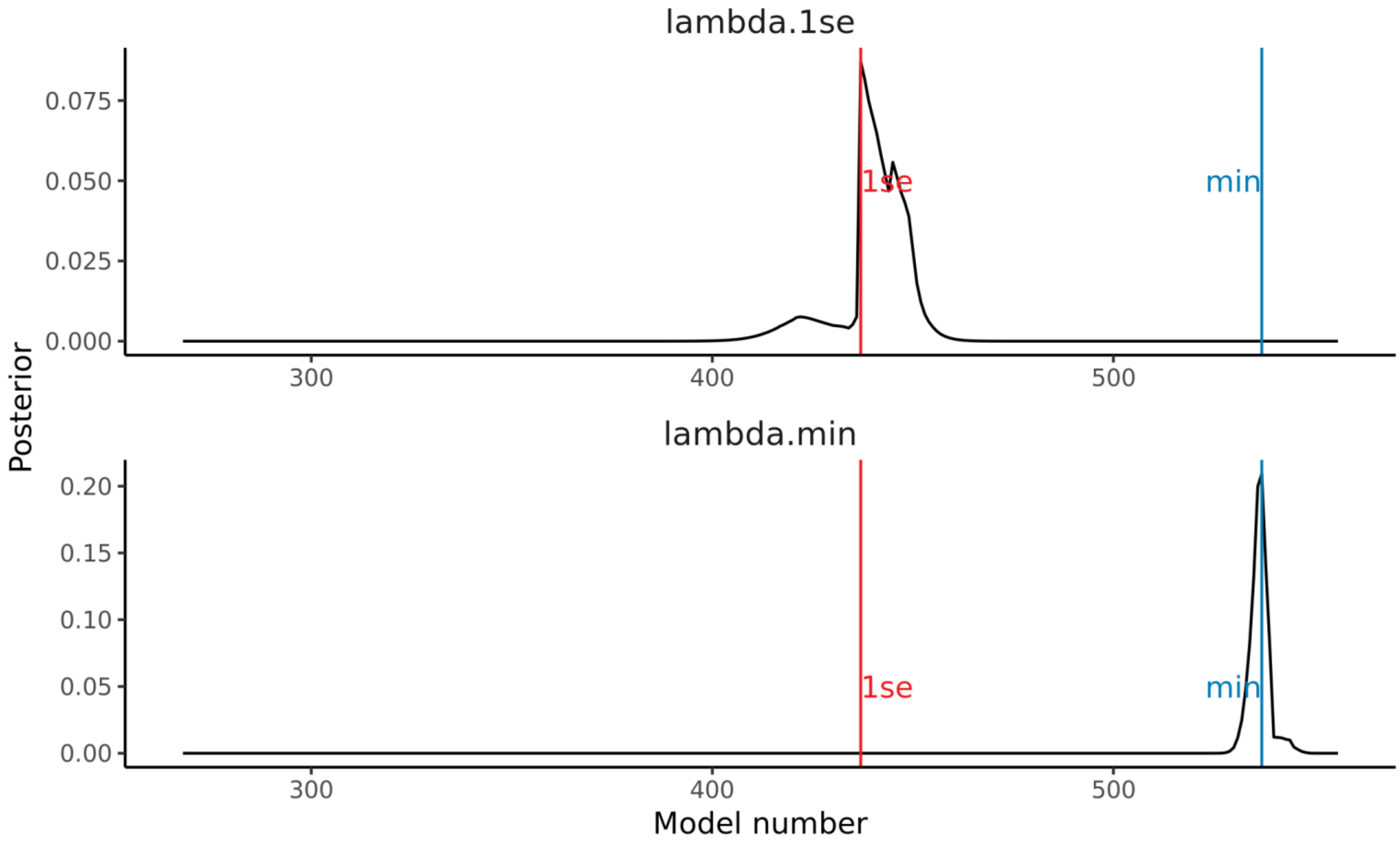
- For given λ, let
  - $C(\lambda)$ denote the average k-fold **cross-validation error** (we use $k = 8$) for parameterization $\hat{\beta}(\lambda)$
  - $S(\lambda)$ denote standard deviation of the **cross-validation error** across the k folds
- 4 alternative values of scalar λ are selected:

1) **"lambda.min':** $\boldsymbol{\lambda_{min} = \arg\min_{\lambda} C(\lambda)}$
2) **"1se":** $\boldsymbol{\lambda_{1se} = \arg\max_{\lambda} \{C(\lambda) \le C(\lambda_{min}) + S(\lambda_{min})\}}$
3) **"simple" :** $\boldsymbol{\lambda_{simp} = \max\{\lambda: p(M_{1se}|y;\lambda) > \delta\}}$
4) **"complex" :** $\boldsymbol{\lambda_{comp} = \min\{\lambda: p(M_{1se}|y;\lambda) > \delta\}}$

We use $\delta = 0.0005$

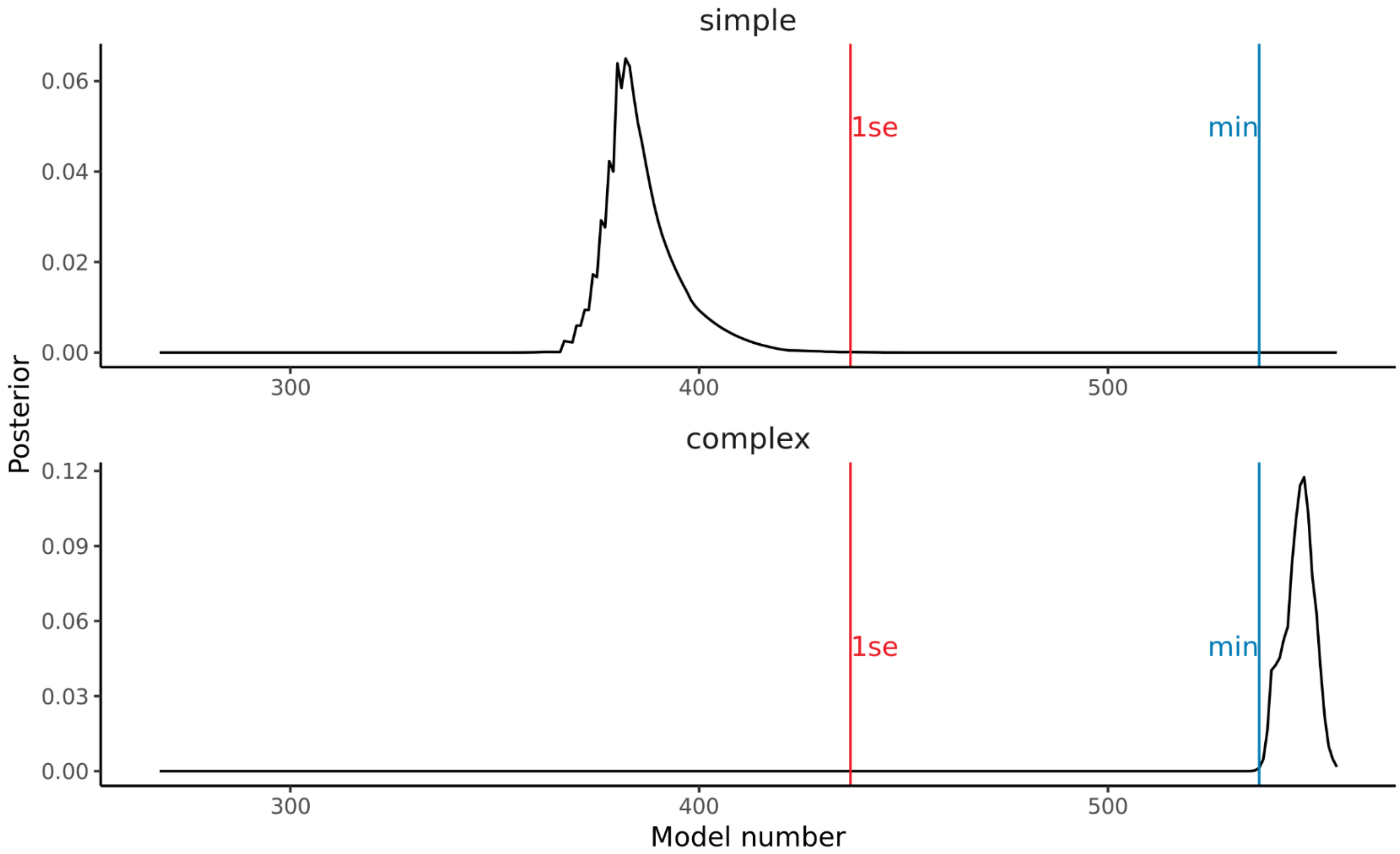
LASSO: selection of penalty parameter (3)

- 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 λ
  - The model set may be taken as the set of all models generated by the values of λ considered
- Now consider a single selected λ
  - This defines a prior $\boldsymbol{\pi(M), M \in \mathcal{M}}$ on the model set $\mathcal{M}$
  - Together with data y, this defines a posterior $\boldsymbol{p(M|y), M \in \mathcal{M}}$ 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^*}

TAYLOR FRY
UNSW AUSTRALIA

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 λ
  - 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:

Accident periods	Gate	Payment periods	Gate
Last 2	[0.75,1.33]	Next 2	[0.91,1.10]
Last 5	[0.80,1.25]	Next 5	[0.87,1.15]
Last 10	[0.83,1.20]	Next 10	[0.83,1.20]

- 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

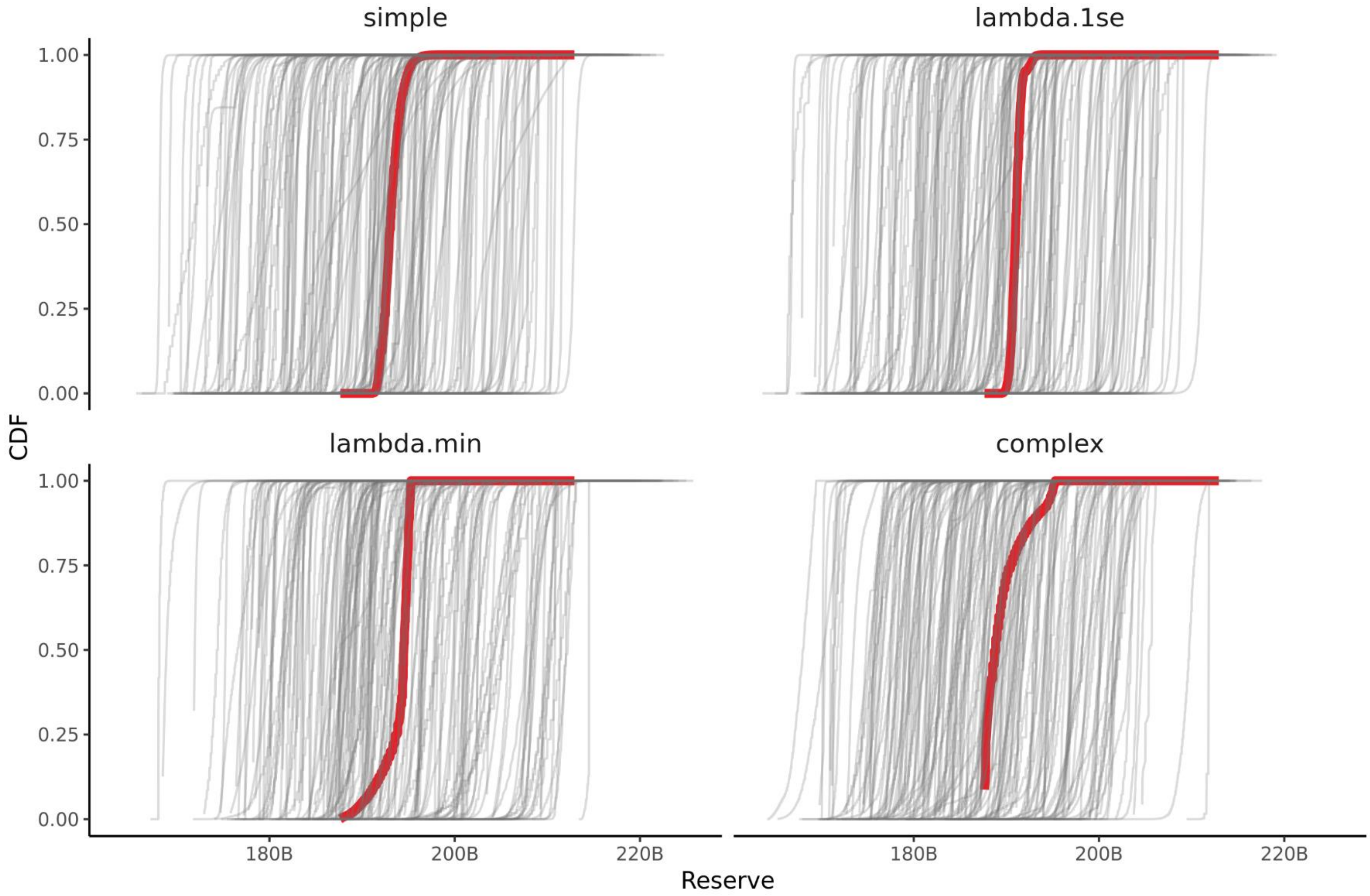
Numerical results: data set 1

Model	Prior	True reserve ($B)	Fore-cast ($B)	Estimated internal model error
Primary	Simple	190	193	0.51%
	1se	190	191	0.30%
	Lambda.min	190	194	0.82%
	Complex	190	190	1.14%
Boot-Strap	Simple	190	191	0.37%
	1se	190	189	0.32%
	Lambda.min	190	192	0.41%
	Complex	190	188	0.53%

Not too much differ-ence over priors

- **Very small model error for data compatible with chain ladder**

Numerical results: data set 1 : posterior cdf’s



TAYLOR FRY
UNSW AUSTRALIA

Numerical results for all data sets 1-4

- Bootstrap results for 1se and lambda.min

Data set	Prior	Forecast					
		True ($B)	Mean ($B)	Internal model error (CoV)	Parameter error (CoV)	Process error (CoV)	Total error (CoV)
1	1se	190	189	0.32%	5.30%	3.29%	6.24%
	lambda.min	190	192	0.41%	5.15%	2.75%	5.85%
2	1se	238	252	1.45%	10.00%	3.93%	10.84%
	lambda.min	238	240	1.79%	8.83%	4.69%	10.16%
3	1se	608	703	2.27%	11.23%	5.71%	12.80%
	lambda.min	608	589	2.12%	11.19%	5.27%	12.54%
4	1se	216	243	1.37%	8.63%	4.01%	9.62%
	lambda.min	216	252	1.81%	12.54%	5.08%	13.65%

Numerical results for all data sets 1-4: comparison of LASSO and GLM forecasts

SANITY CHECK

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

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

Law of total variance:

$$Var[\hat{R}] = E_M[Var[\hat{R}]|M] + Var_M[E[\hat{R}]|M]$$

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

Accident periods	Gate	Payment periods	Gate
Last 2	[0.75,1.33]	Next 2	[0.91,1.10]
Last 5	[0.80,1.25]	Next 5	[0.87,1.15]
Last 10	[0.83,1.20]	Next 10	[0.83,1.20]

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

Data set	Prior	Total error (CoV)	
		Original gates	Widened gates
1	LASSO 1se	6.24%	7.50%
	lambda.min	5.85%	7.03%
2	LASSO 1se	10.84%	15.23%
	lambda.min	10.16%	17.57%
3	LASSO 1se	12.80%	17.86%
	lambda.min	12.54%	19.22%
4	LASSO 1se	9.62%	15.00%
	lambda.min	13.65%	19.43%

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