Institute and Faculty of Actuaries

Bayesian Hierarchical Models for Loss Development

Michael Cooney, Jake Morris and Cynon Sonkkila

31 October 2017

Introduction to Bayesian methods

*"Modern Bayesian methods provide **richer information**, with **greater flexibility** and broader applicability than 20th century methods.*

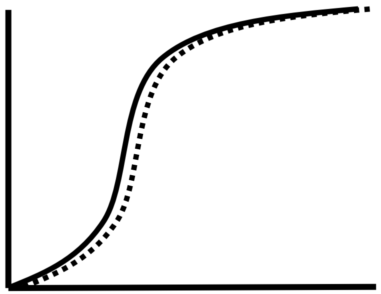
*Bayesian methods are intellectually **coherent and intuitive** ...[and] readily computed..."*

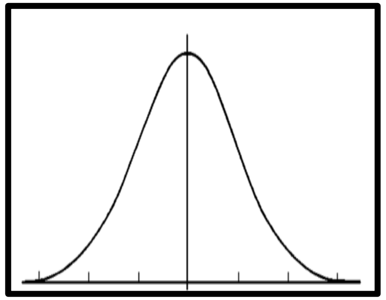
John K. Kruschke

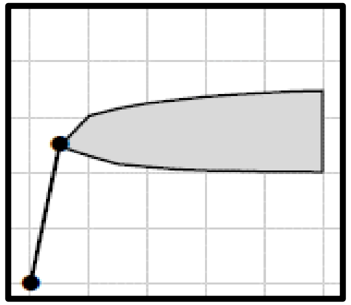
Introduction to Bayesian methods

- Actuarial work is fundamentally assumptions-based:
  - Data + $\underbrace{\text{Models + Judgements}}_{\textbf{Assumptions}}$ = Predictions $\longleftrightarrow$ **Assumptions**

- **Key challenge:** updating assumptions as new information arises
  - Are existing assumptions still relevant?
  - To what extent should we react? Are we consistent?







Introduction to Bayesian methods

- Standard actuarial problems → credibility theory:

$$\boxed{Estimate = Z\bar{X}_i + (1 - Z)\mu}$$

- OK, but we might also like…
  - Model flexibility, e.g. nonlinearities, time-series, …
  - Full distribution of estimates (reflecting uncertainty in $\bar{X}_i$ & μ):

> *“**Given** [our] estimate of future payments and …*
> ***current state of knowledge**, what is the probability that final payments will be no larger than the given value?”*

Introduction to Bayesian methods

- Bayes' theorem **(probability)**:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

Events

- Bayes' theorem **(inference)**:

$$p(\theta|X) \propto p(\theta)L(X;\theta)$$

Posterior ∝ Prior x Likelihood

RVs
(Data are fixed)

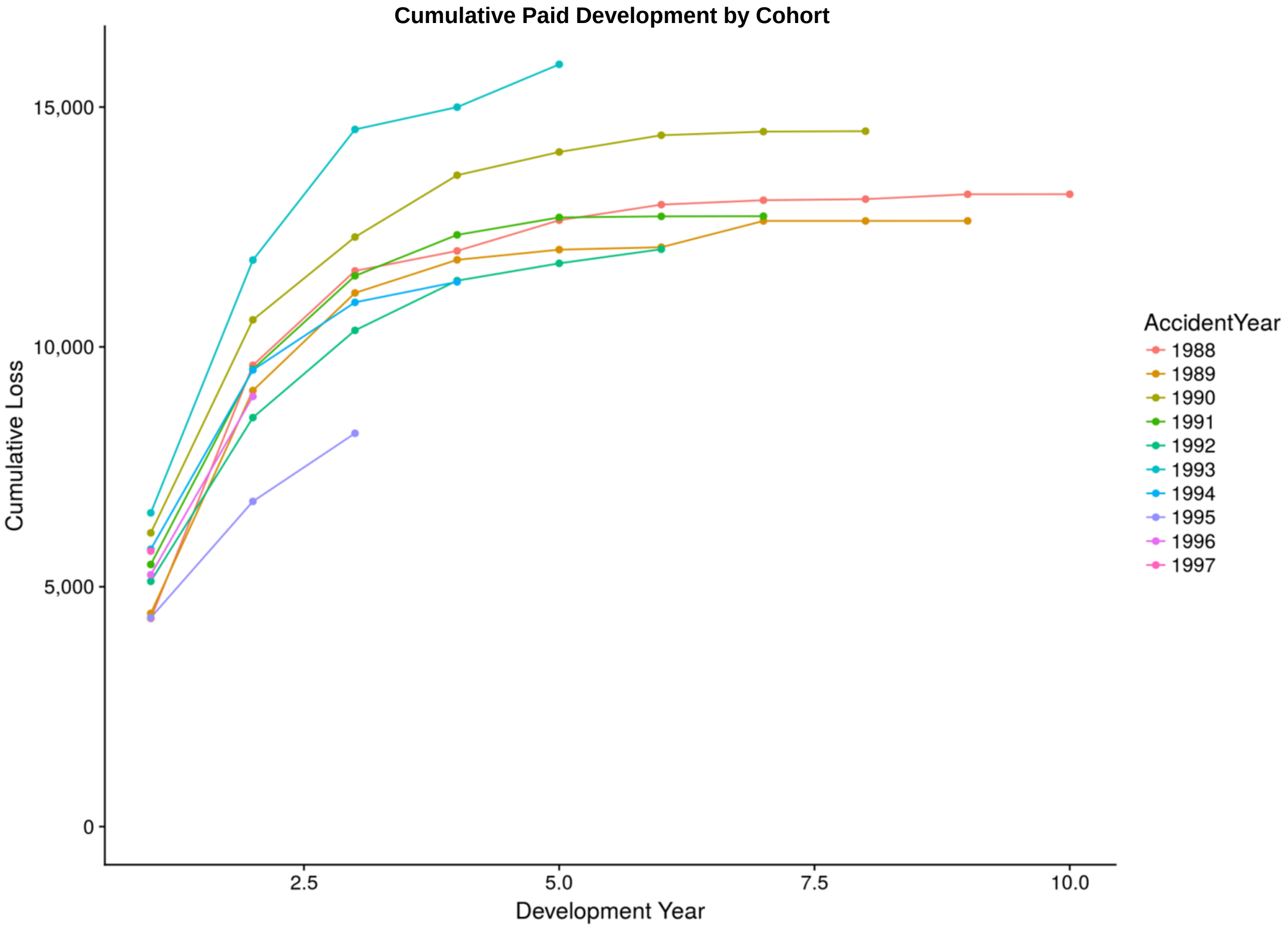
- For actuaries:

$$p(ULR|Inc) \propto p(ULR)L(Inc;ULR)$$

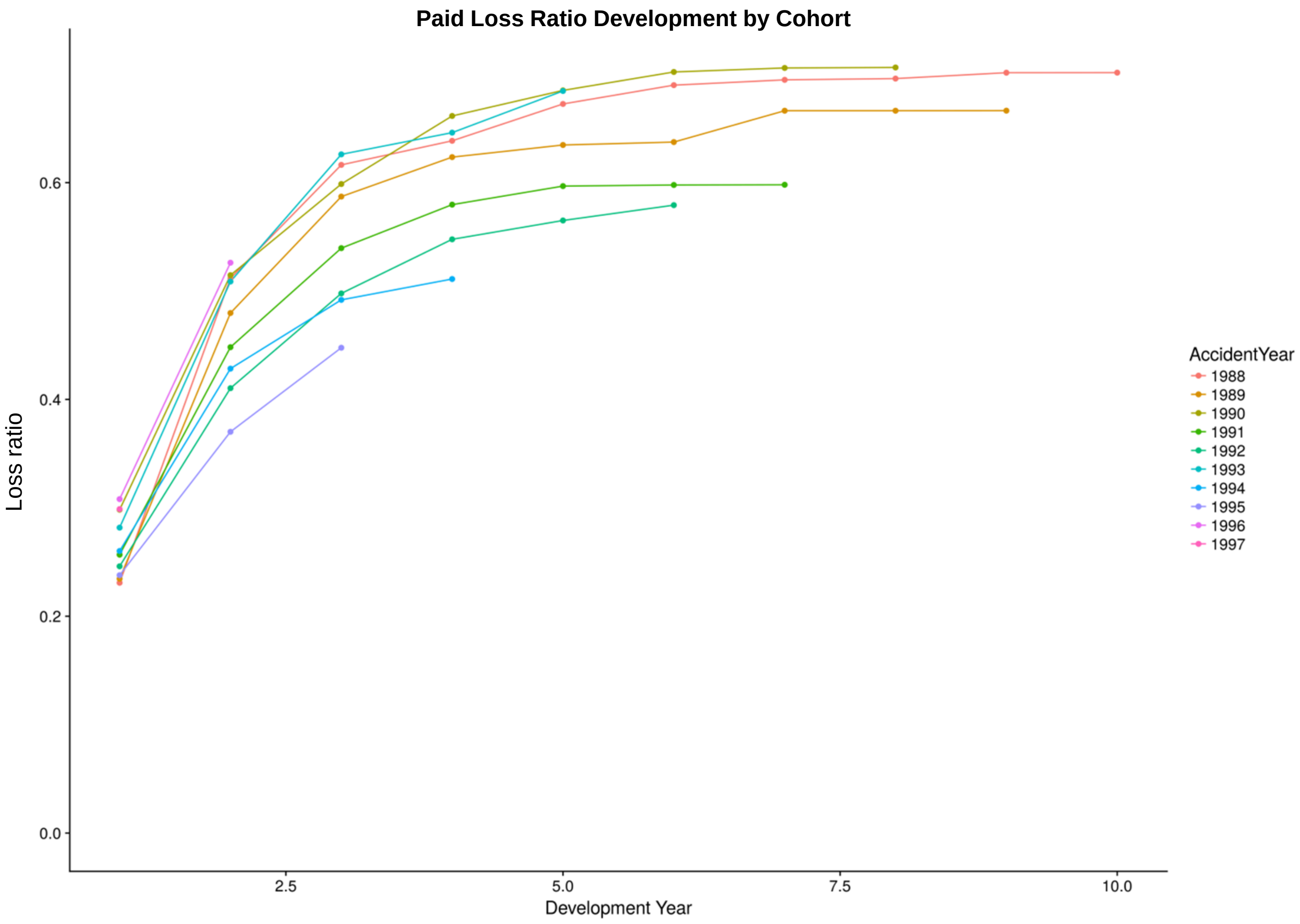
ULR unknown & probability modelled

Application: Loss Development Models

Cumulative losses by Cohort



Loss Ratios by Cohort

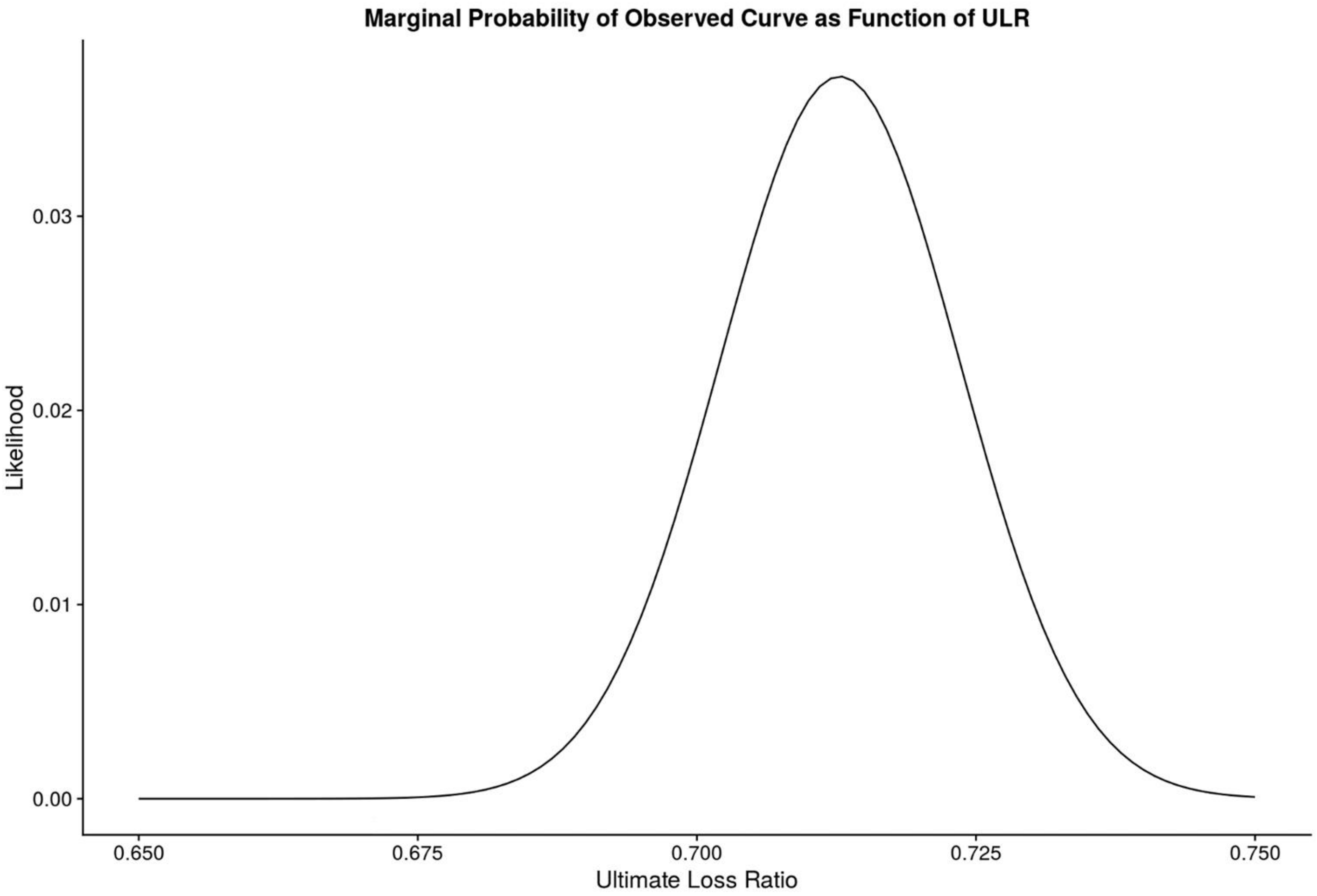


A Potential Model

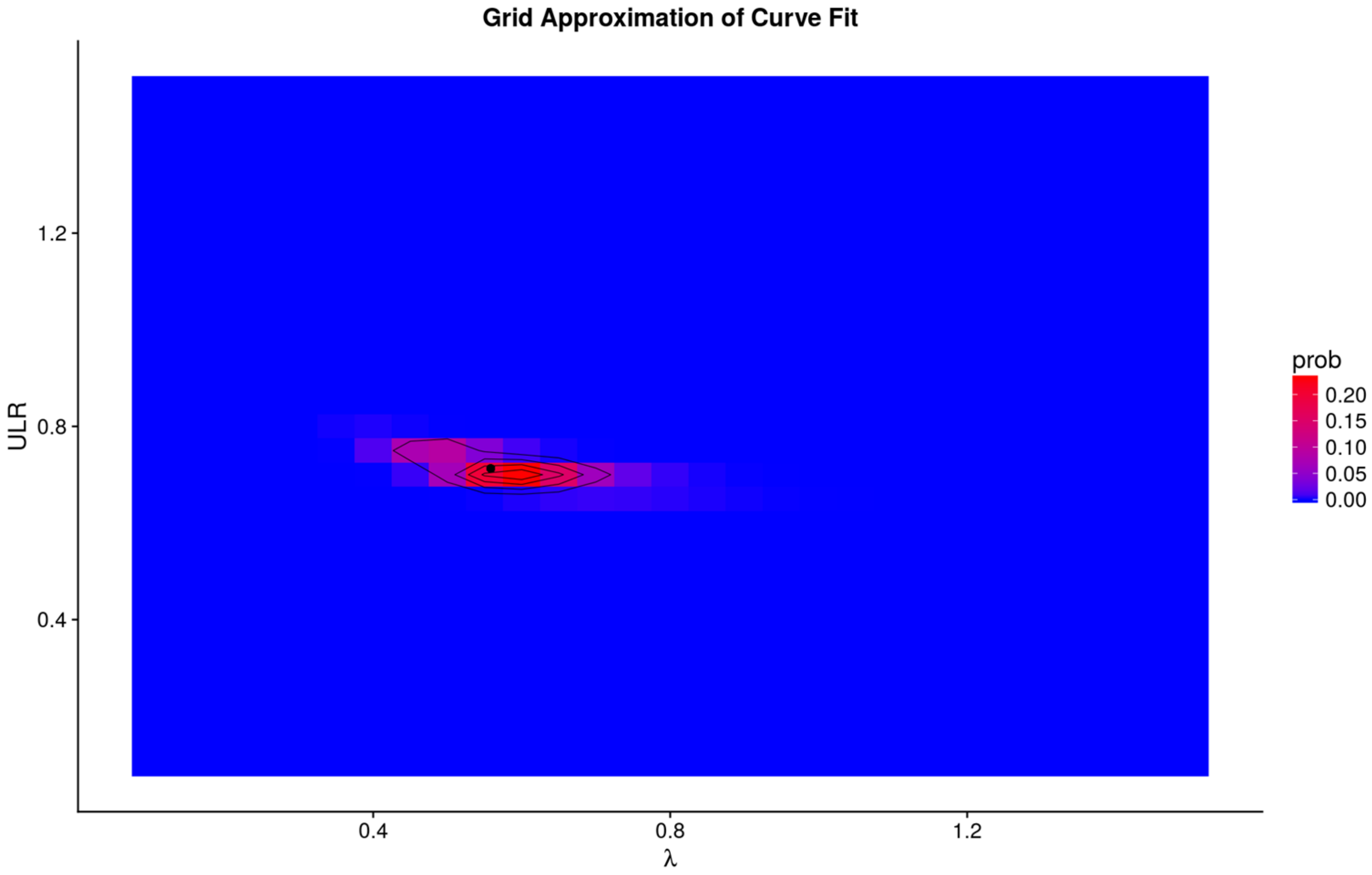
$$\text{Loss}(t) = \text{Premium} \times \text{Ultimate Loss Ratio} \times \text{GF}(t)$$

- Model growth function as Weibull or Log-logistic
- Clark (2003)

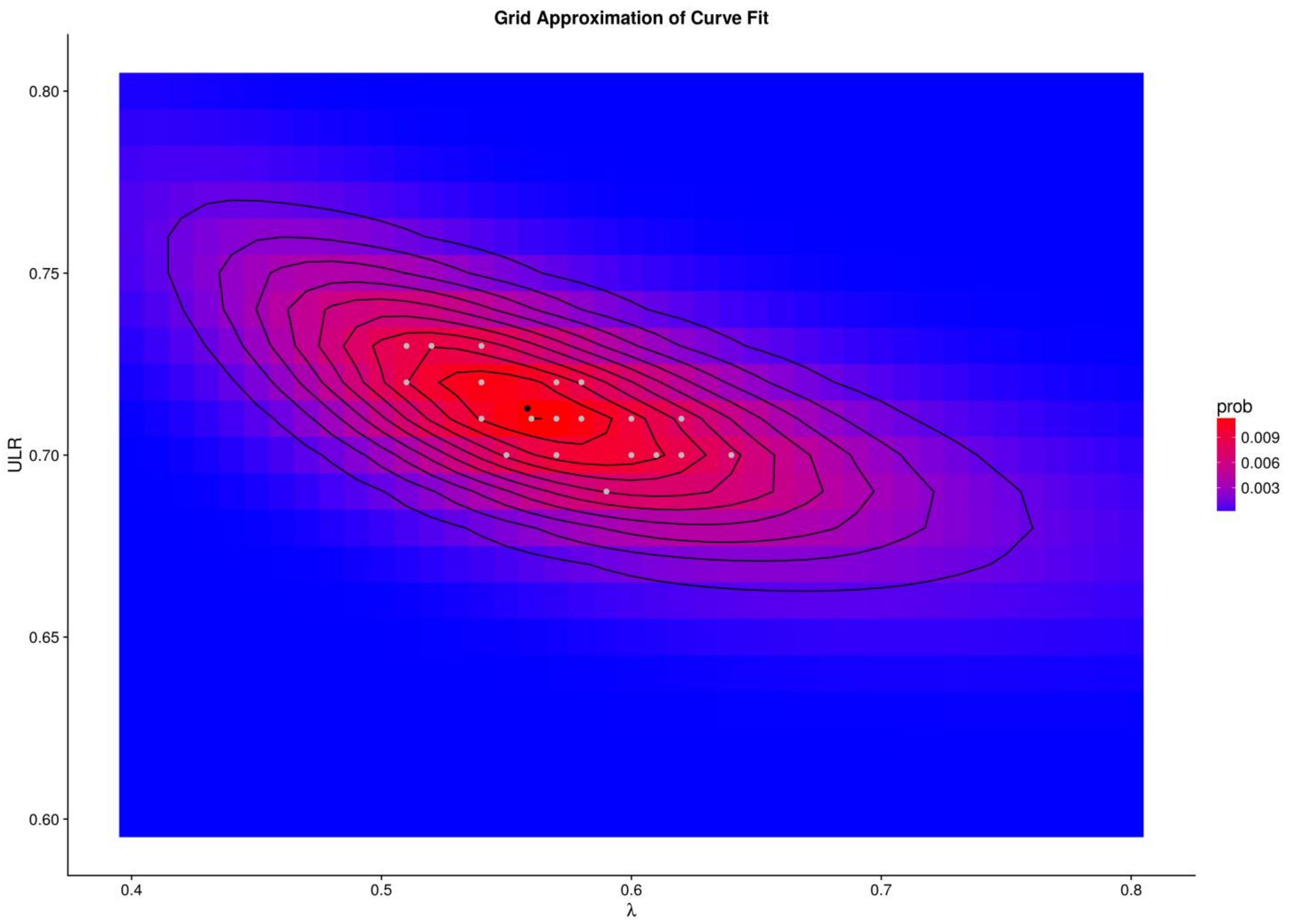
Grid Approximation



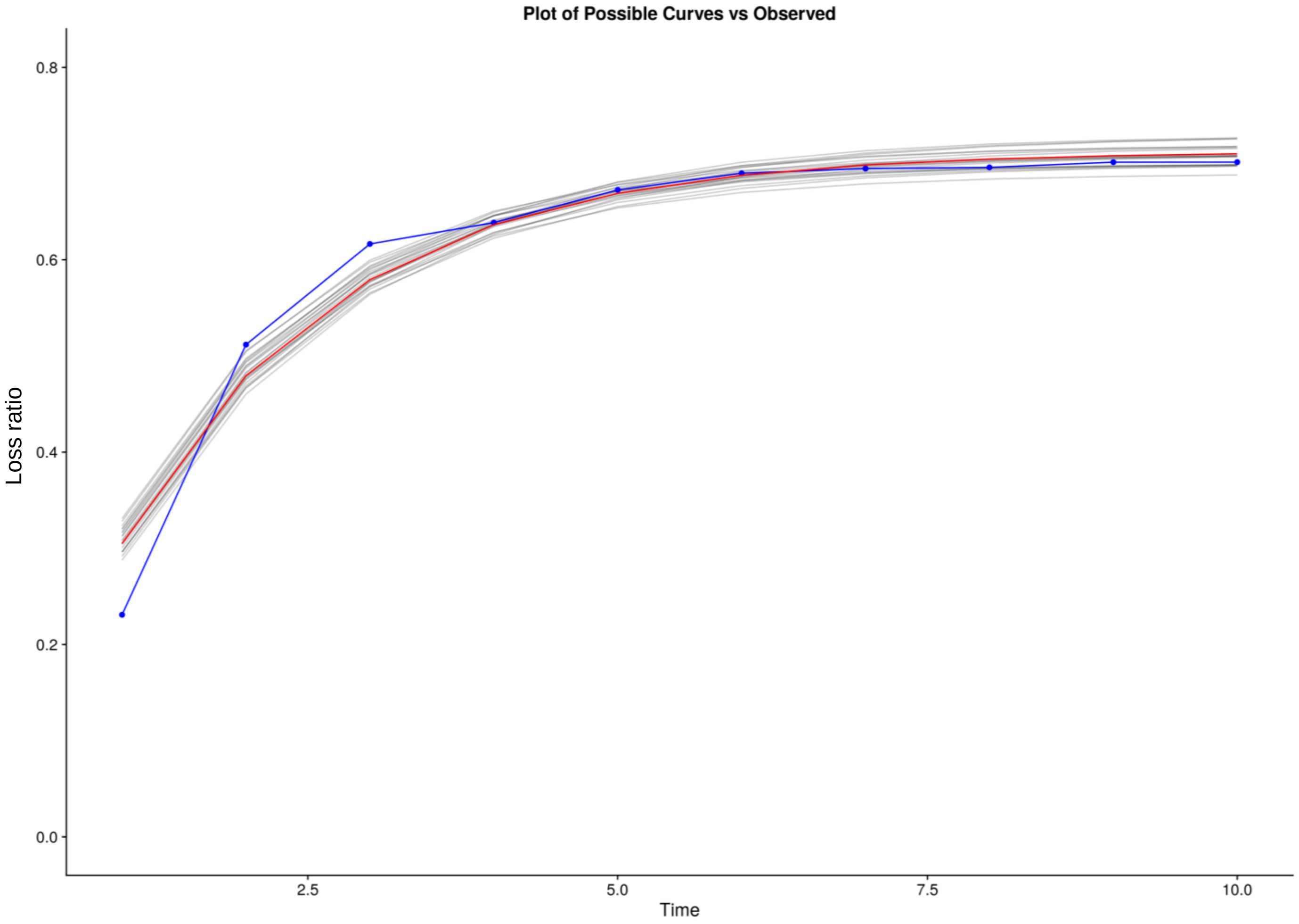
2D Grid Approximation



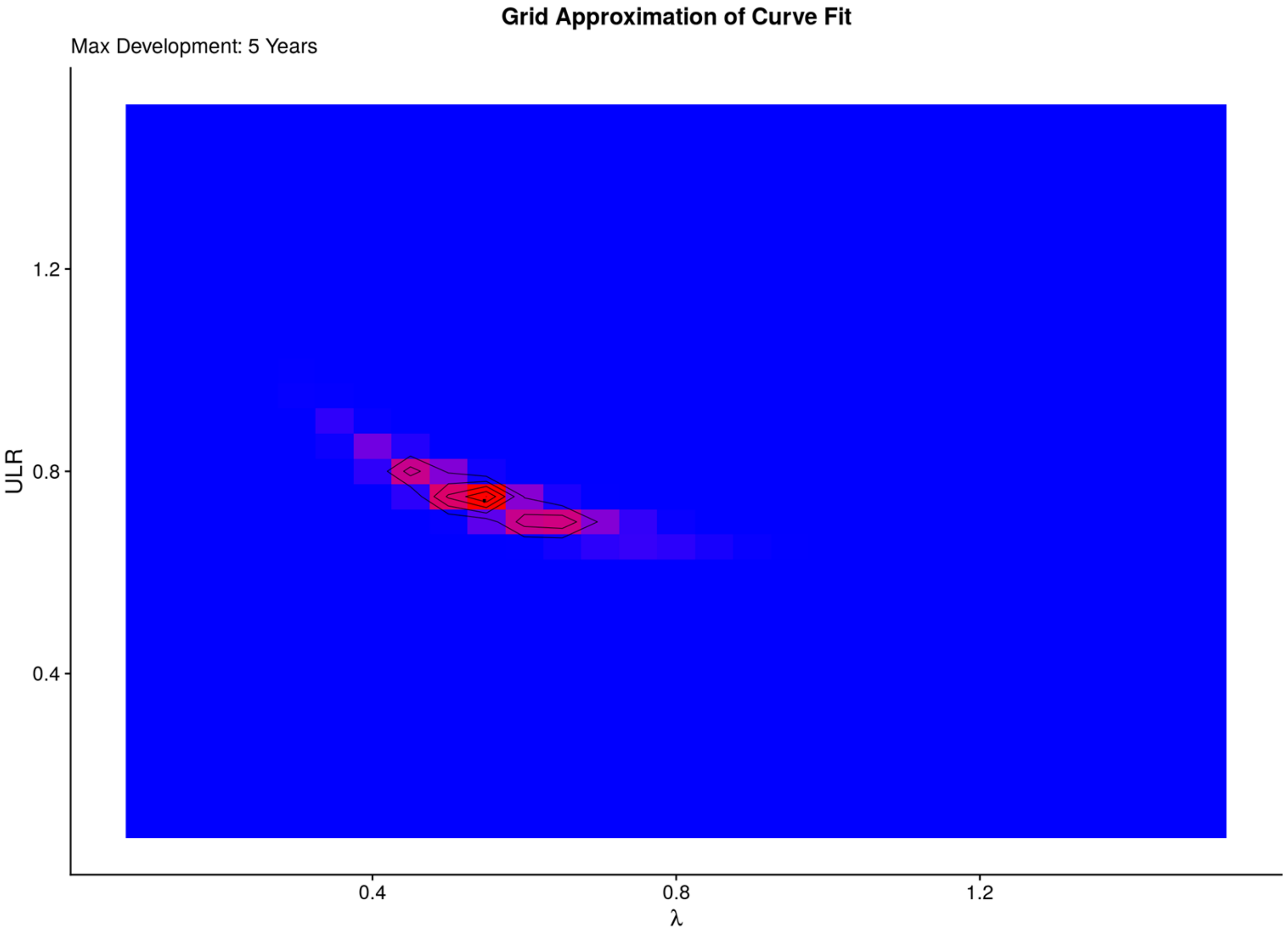
2D Grid Approximation



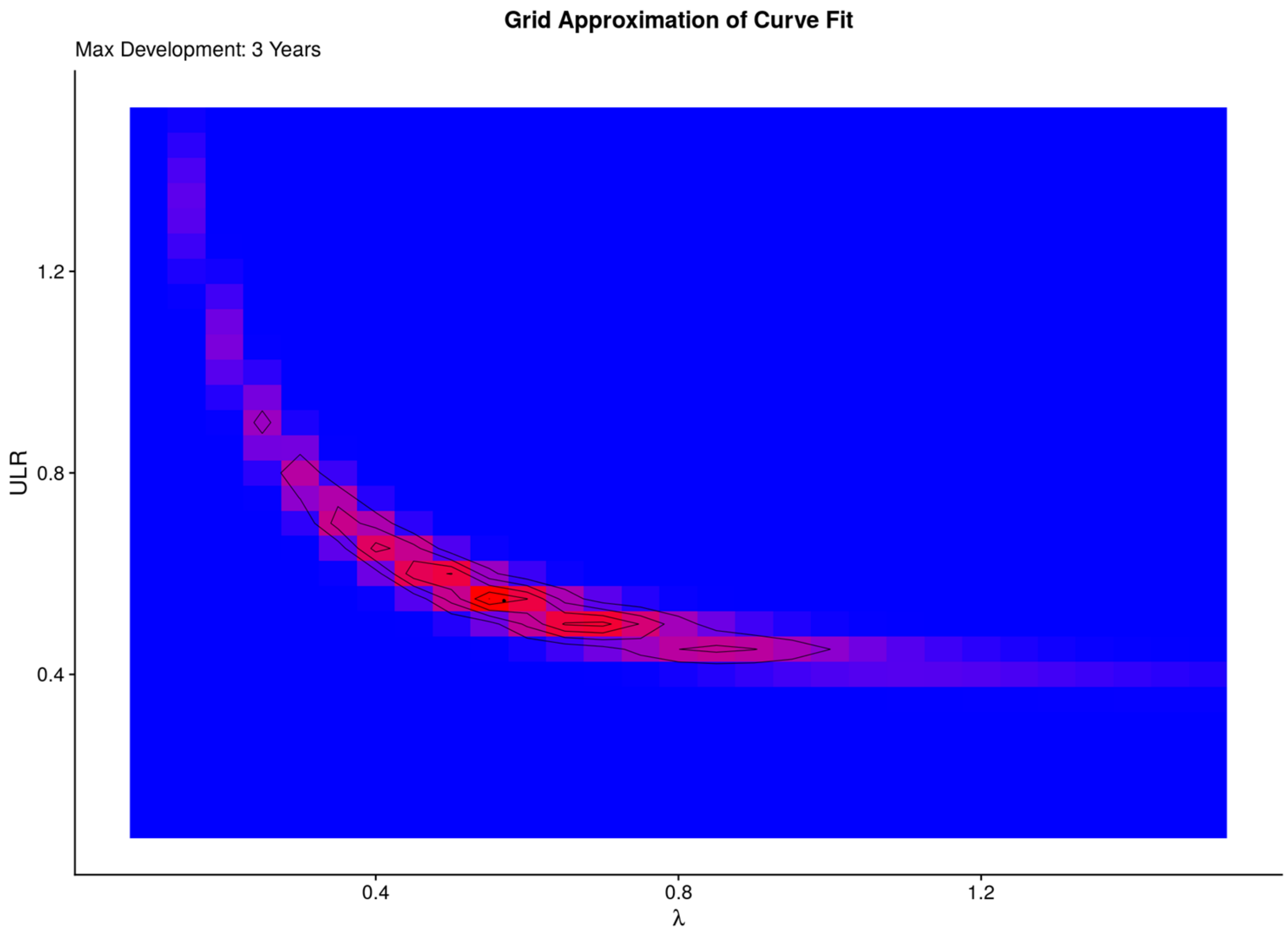
Observed vs Fitted Loss Ratio Development



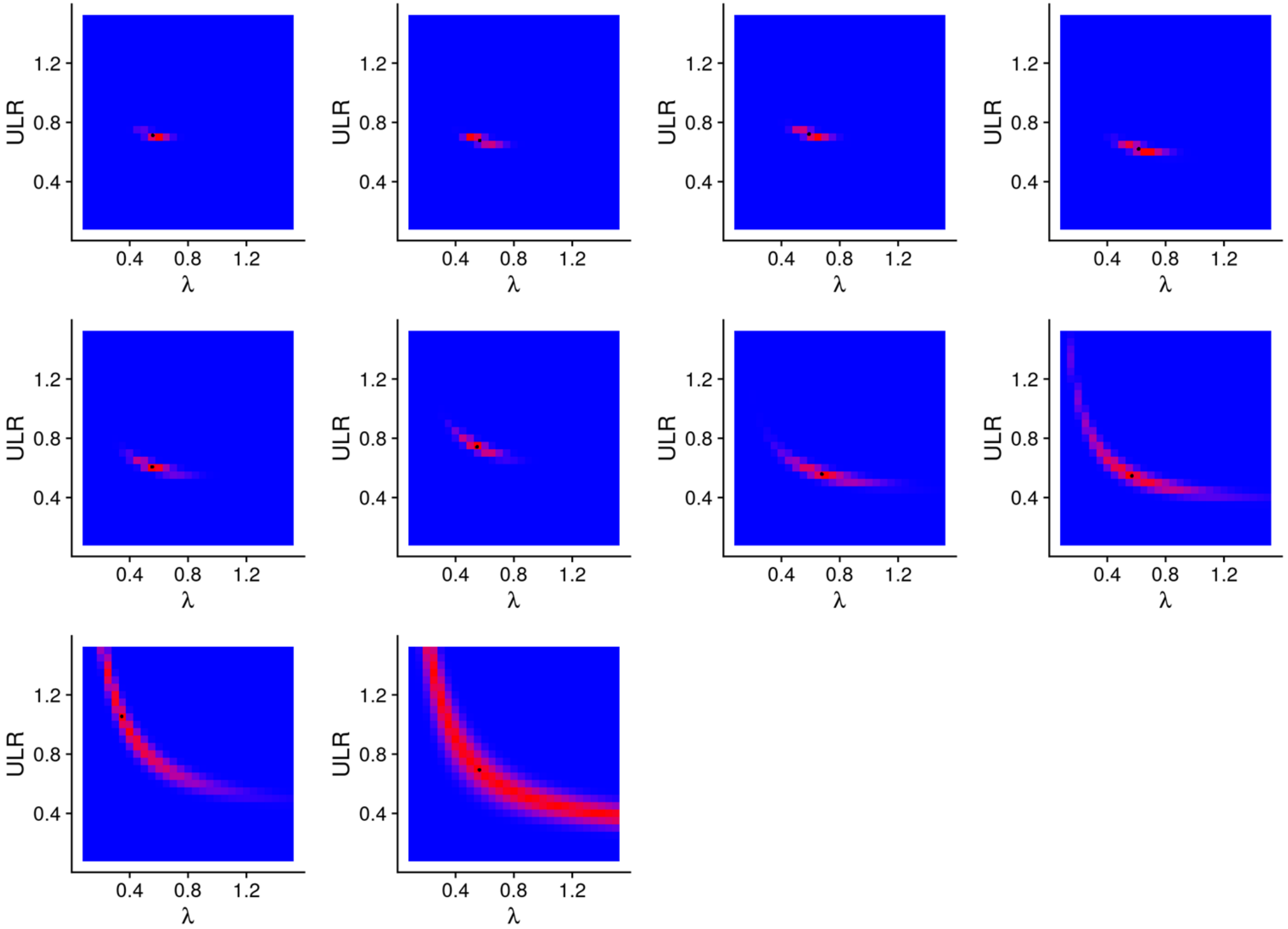
5 Years of Development Only



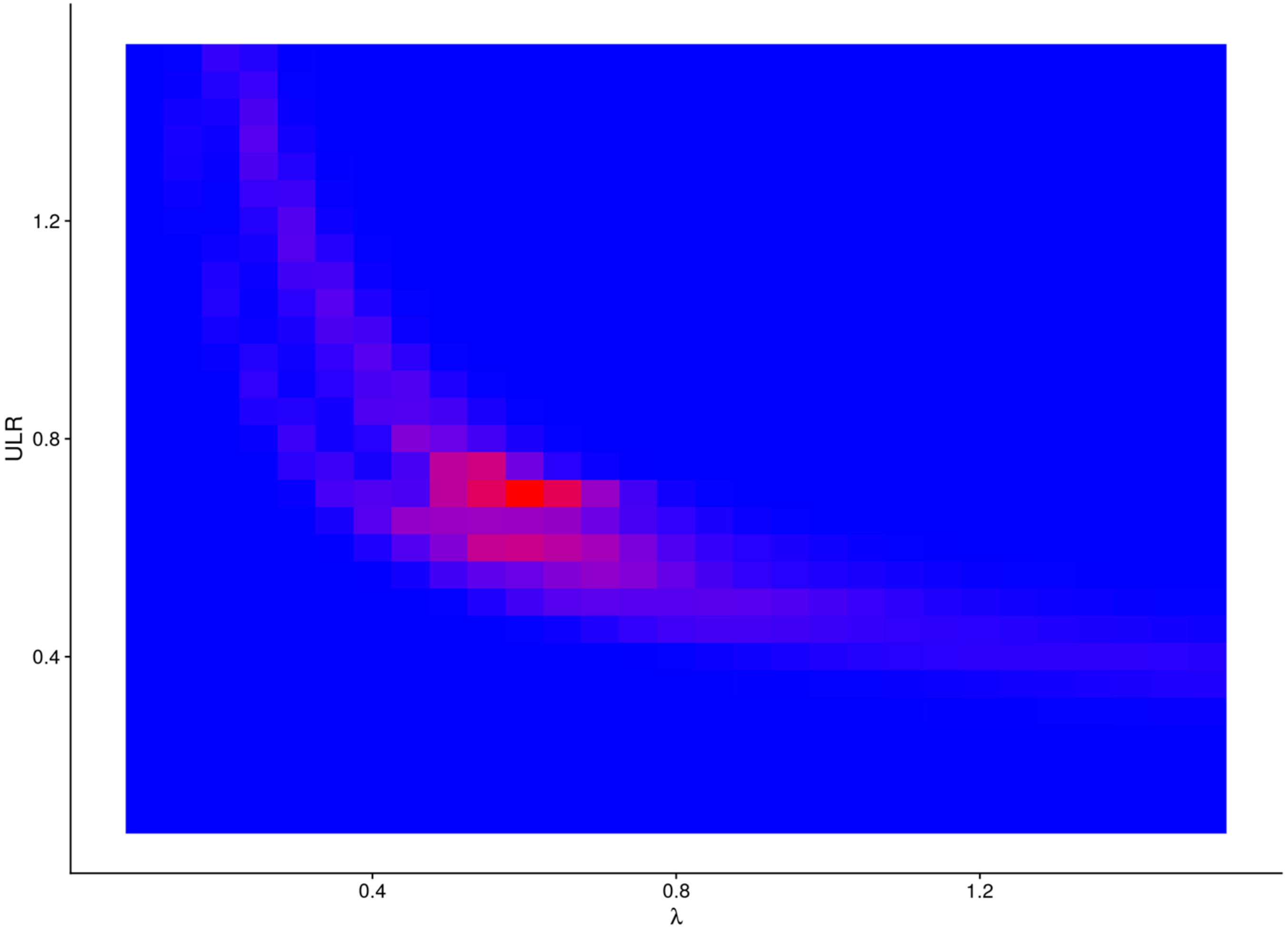
3 Years of Development



Different Accident Years



Sum of all Accident Years



Hierarchical Model Specification

$$\text{Loss}(Y,t) \sim \text{Normal}(\mu(Y,t),\ \sigma_Y)$$

$$\mu(Y,t) = \text{Premium}(Y) \times \text{LR}(Y) \times \text{GF}(t)$$

$$\sigma_Y = \text{Premium}(Y) \times \sigma$$

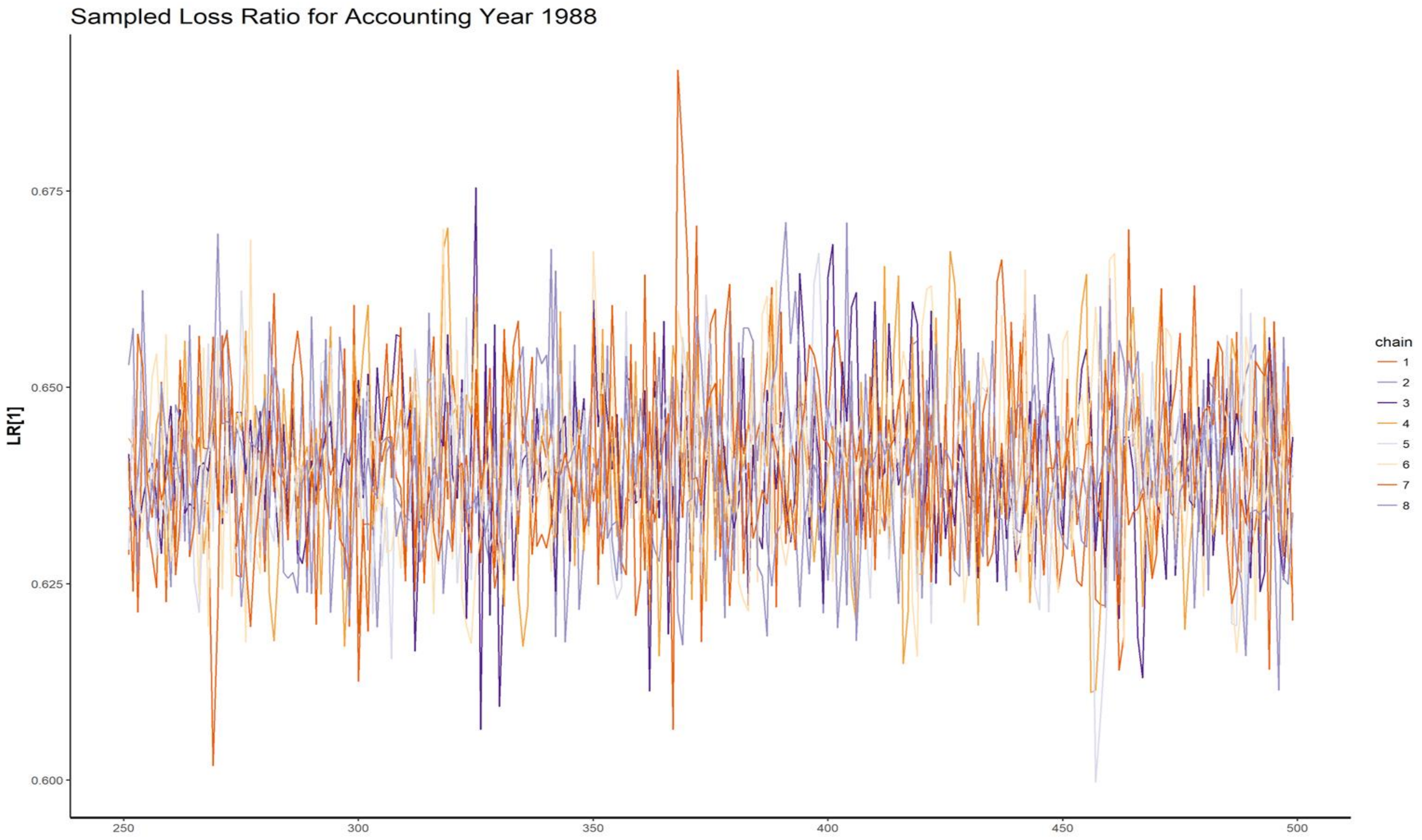
$$\text{LR}_Y \sim \text{Lognormal}(\mu_{LR},\sigma_{LR})$$

$$\mu_{LR} \sim \text{Normal}(0,0.5)$$

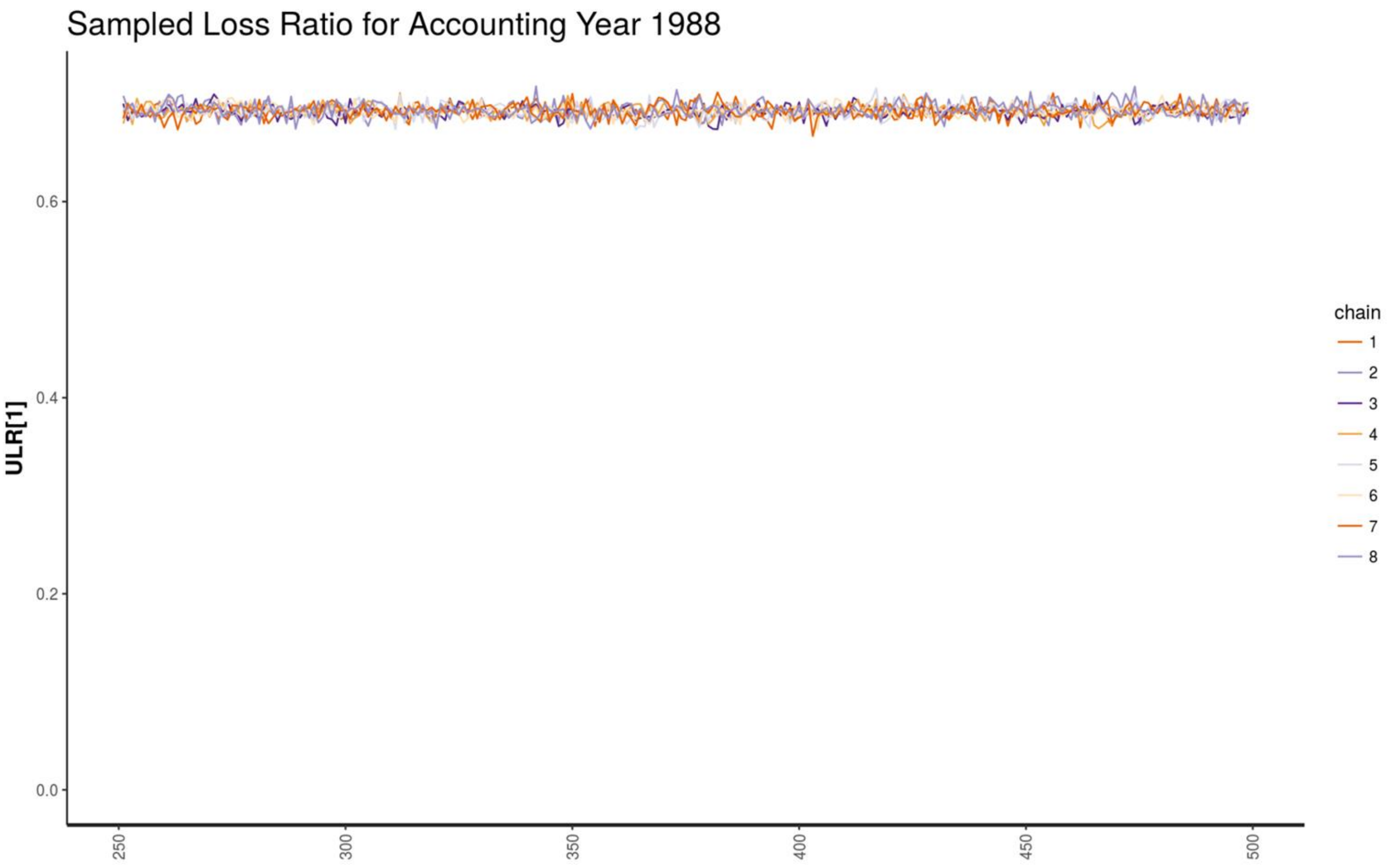
Posterior Sampling

- High-dimensional integrals
- Computationally infeasible
- Sample instead
- Stan (mc-stan.org)
- Hamiltonian Monte Carlo

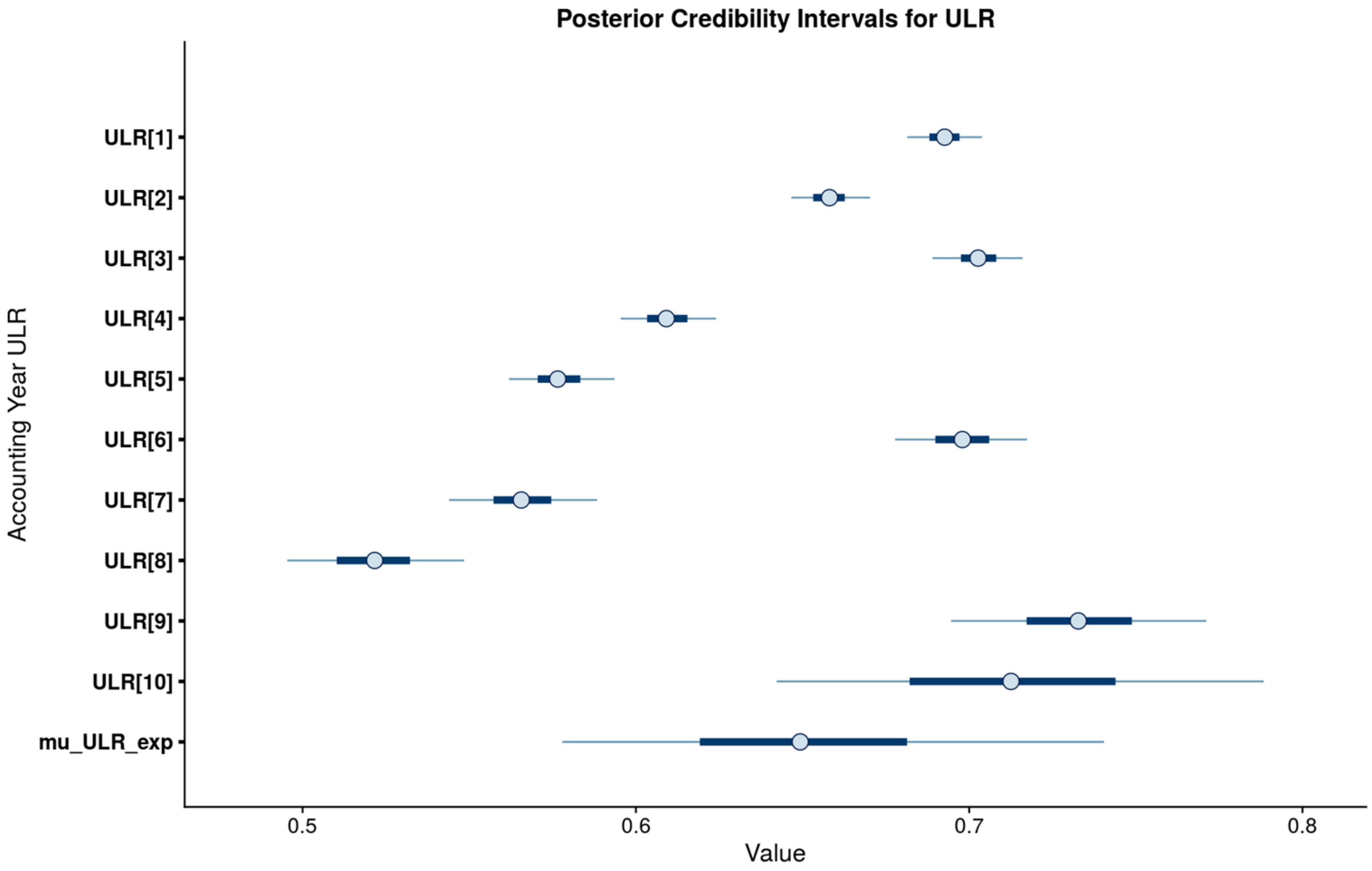
Outputs of MCMC



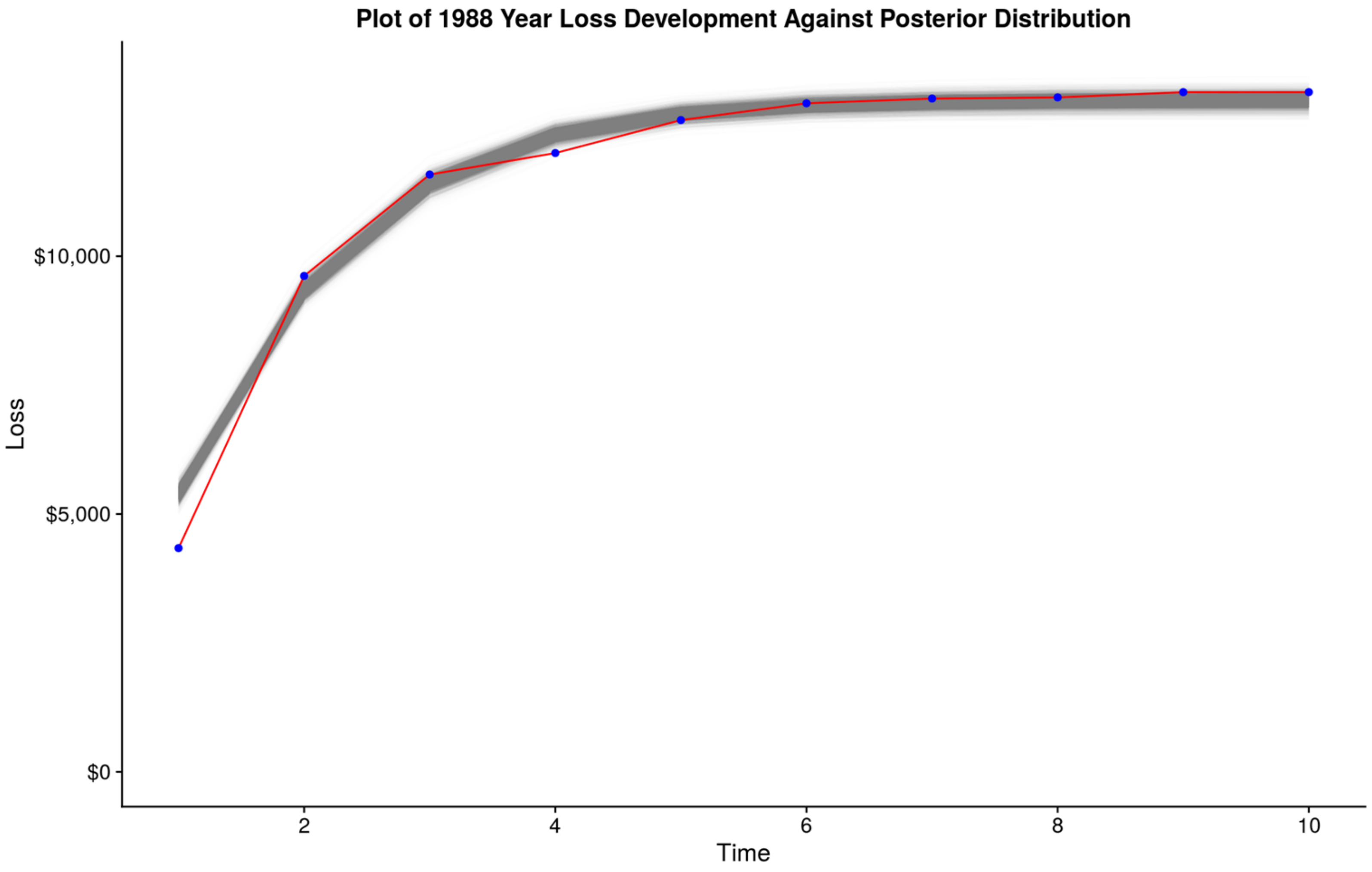
Outputs of MCMC



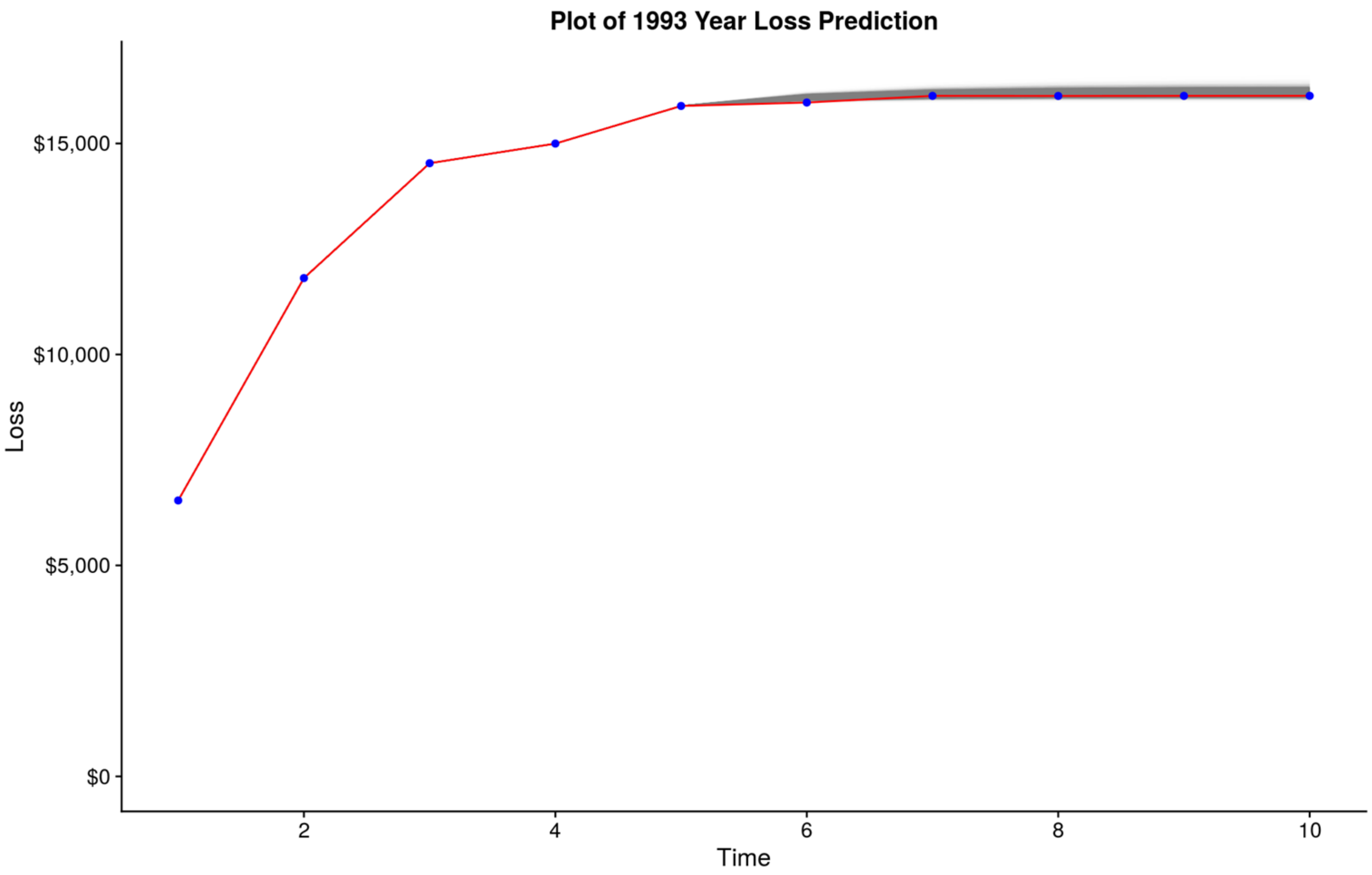
Parameter Inference



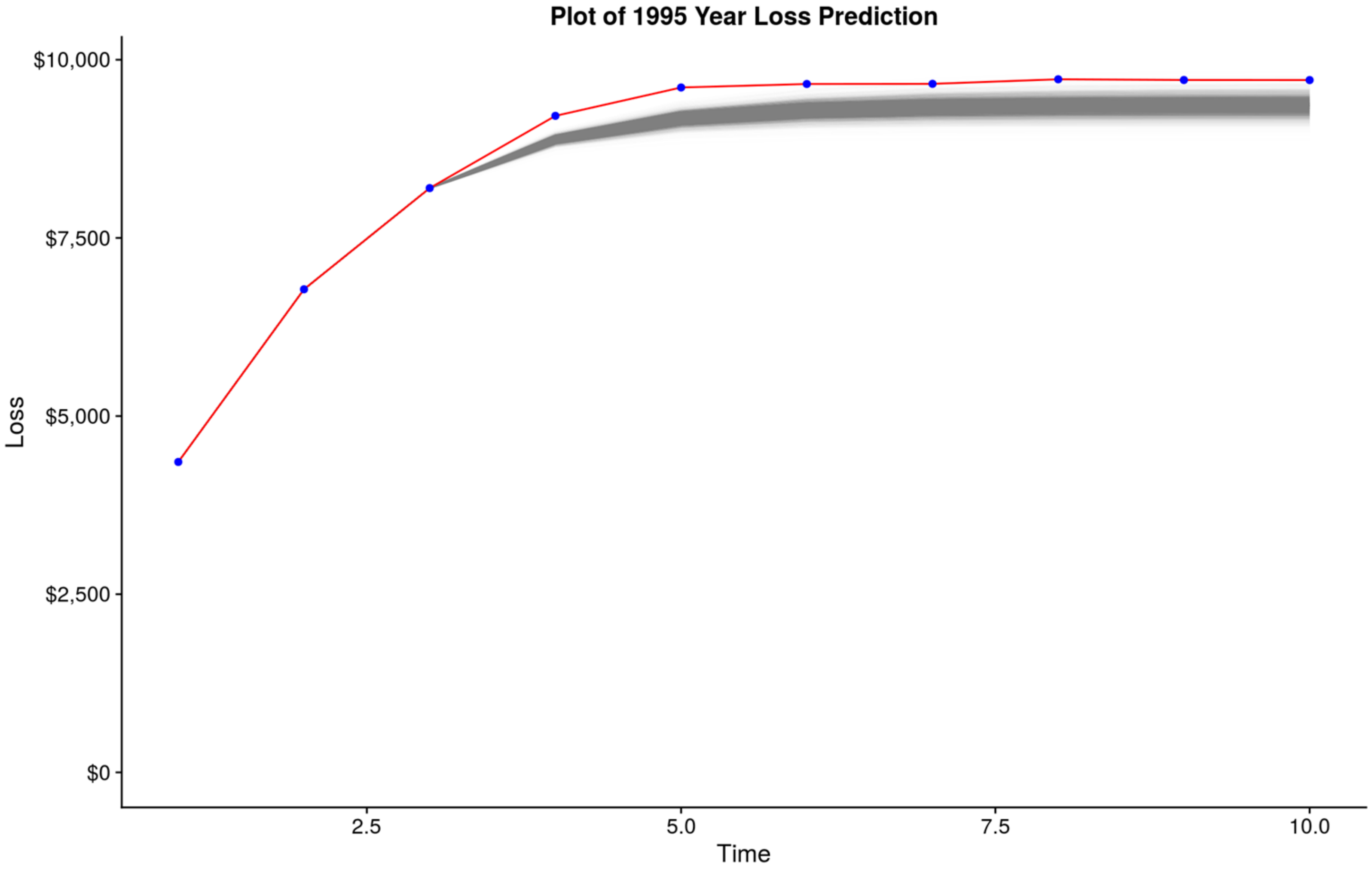
Sanity Check for 1988



Predictions for 1993

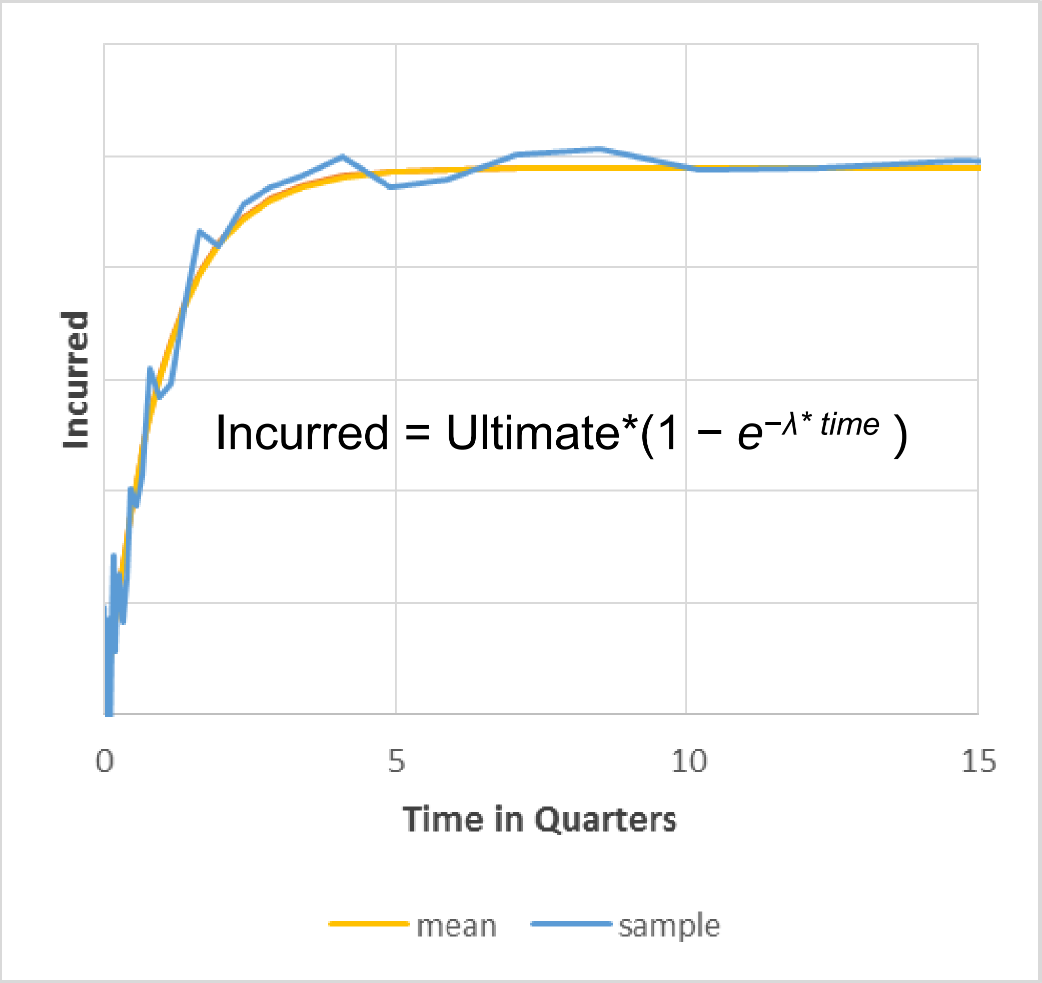


Predictions for 1995



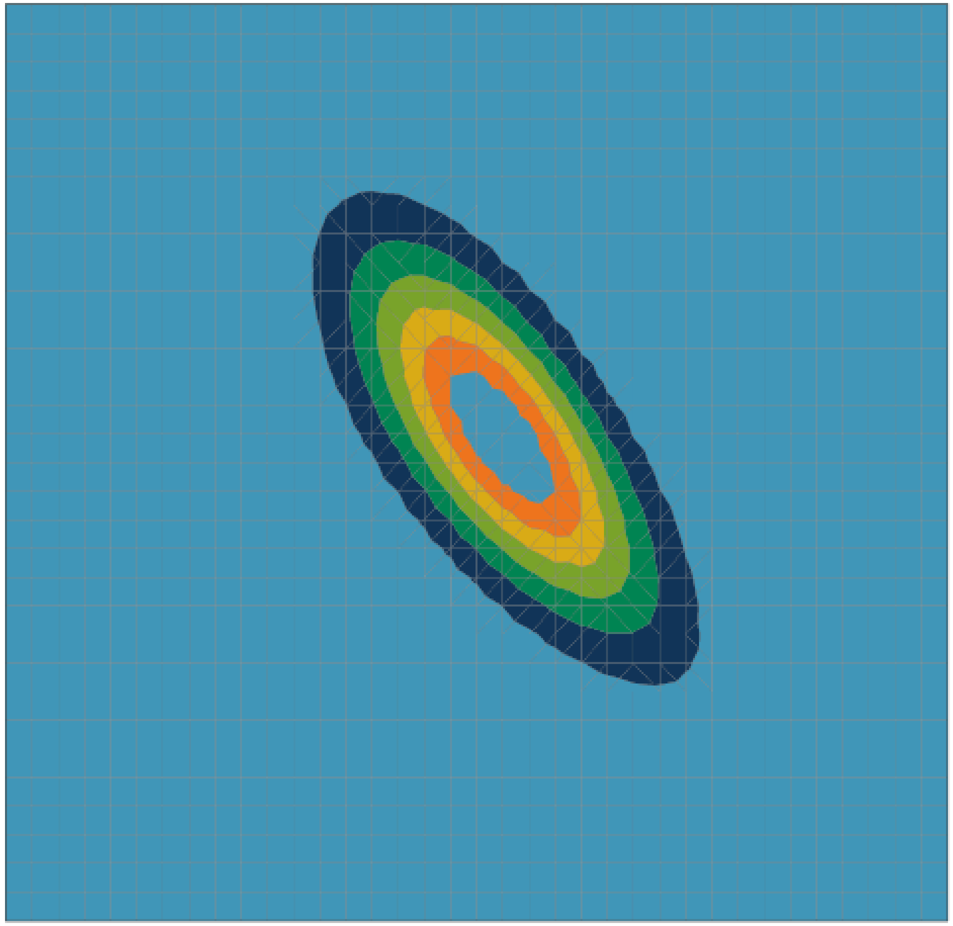
Recap

- We have seen how a time varying development pattern can be approximated using a CDF like curve.
- Relatively low number of parameters are needed.
- The example shown is for:
  - an exponential CDF fit
  - 2 parameters: Ultimate and Lambda.



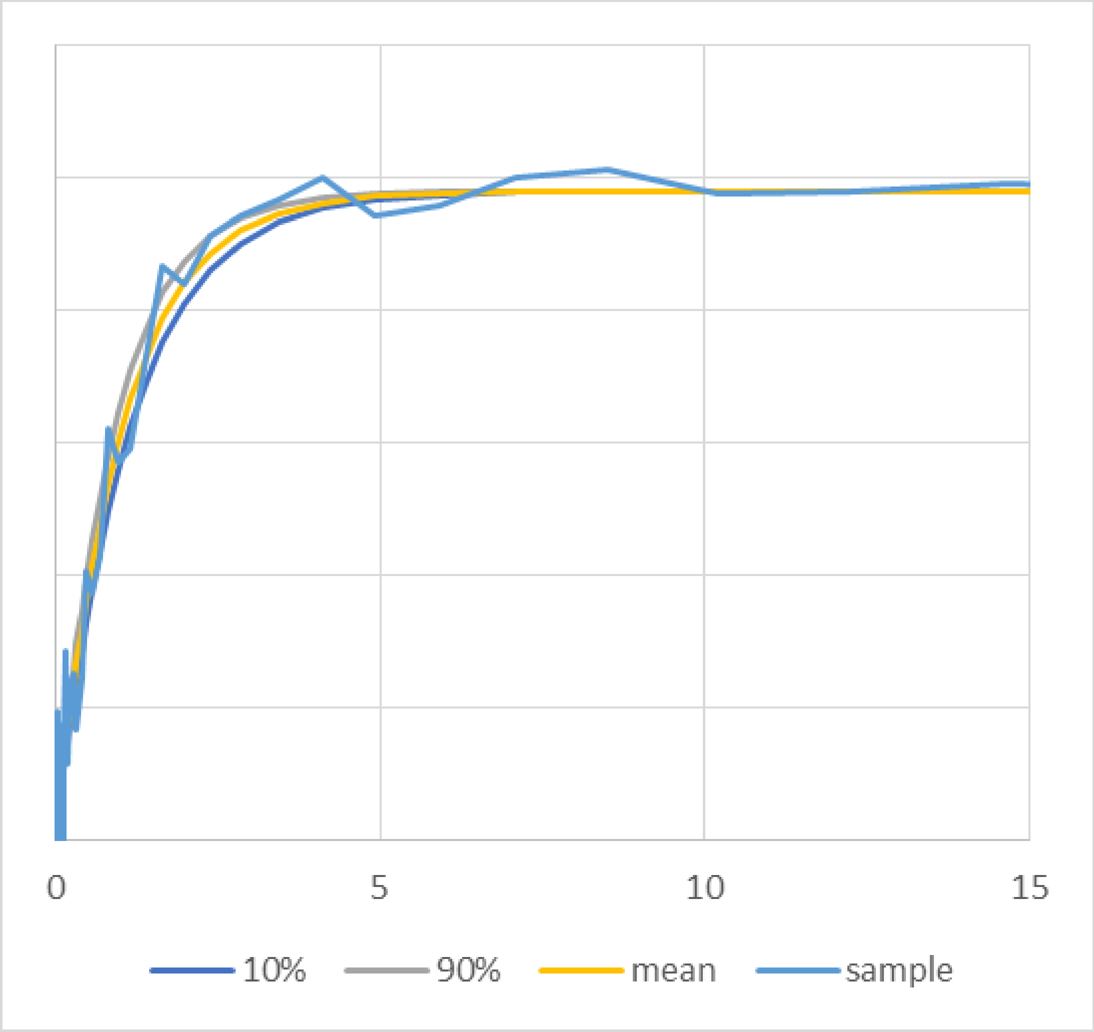
Recap

- Parameter Uncertainty can be represented using the likelihood function (of MLE fame).
- Presented here on a grid.
- MLE would be in the centre.



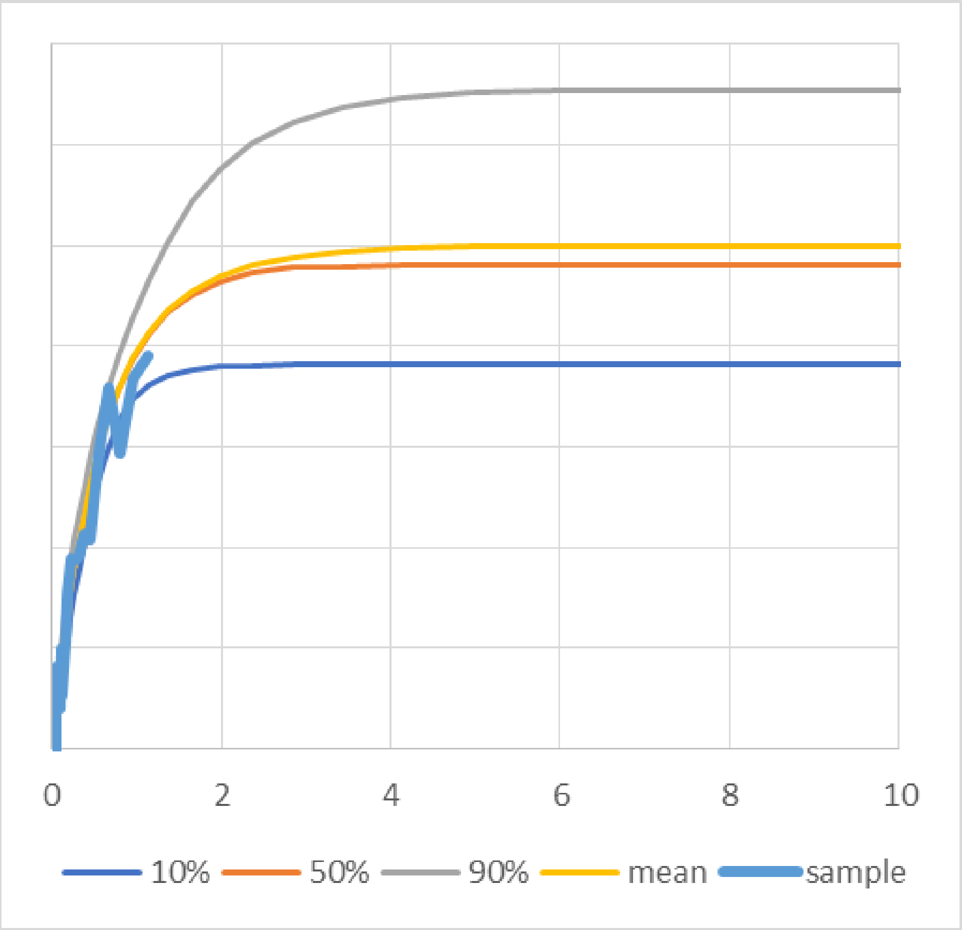
Recap

- Sampling is conducted on the Likelihood distribution
- Sample development curves give an envelope of reasonable development patterns that fit the data.



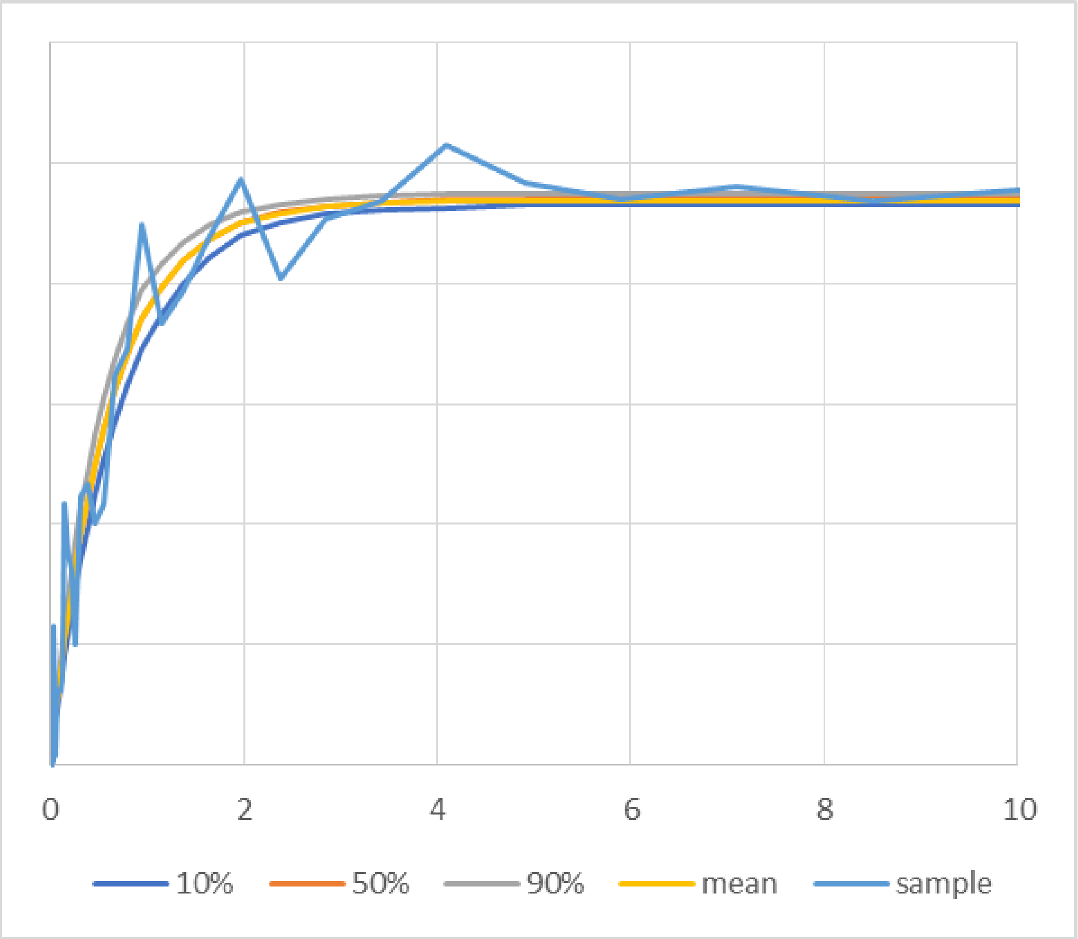
Recap

- Less data creates more uncertainty.



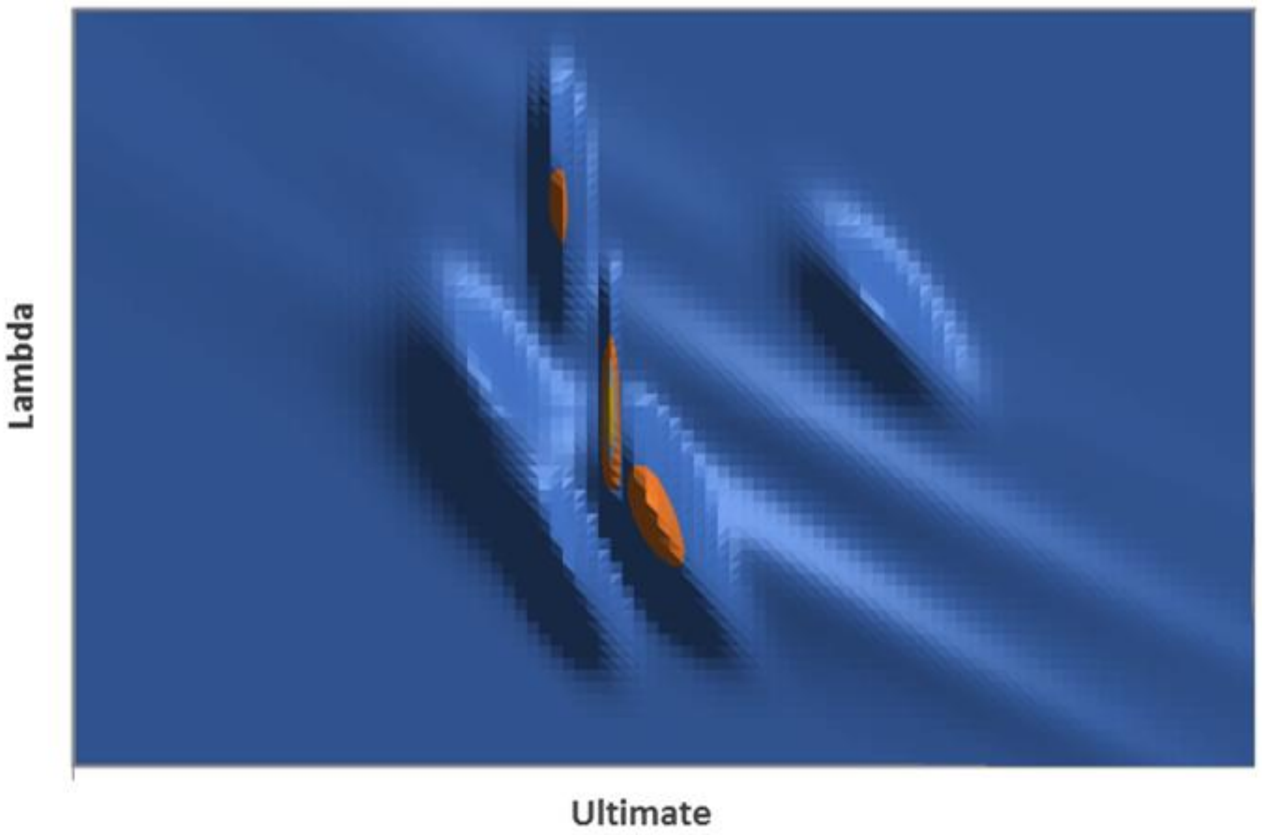
Recap

- More data creates less uncertainty.



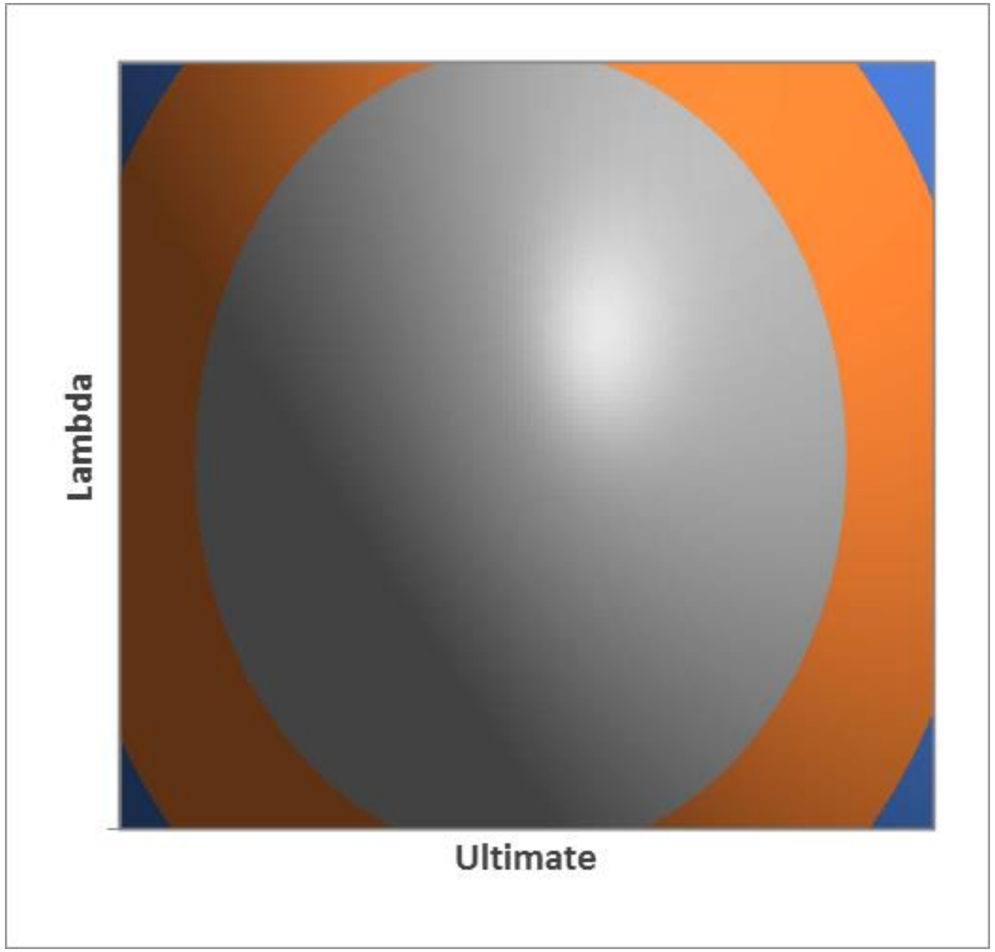
Recap

- Sharing credibility from year to year is incorporated using a prior.
- One method is to think of the prior years as samples “what might happen” for a new year.
- So the “sum” of previous years is a suggestion for a new year i.e. a prior.
- 14 years of data shown here as an example.
- Perhaps there is a better way?



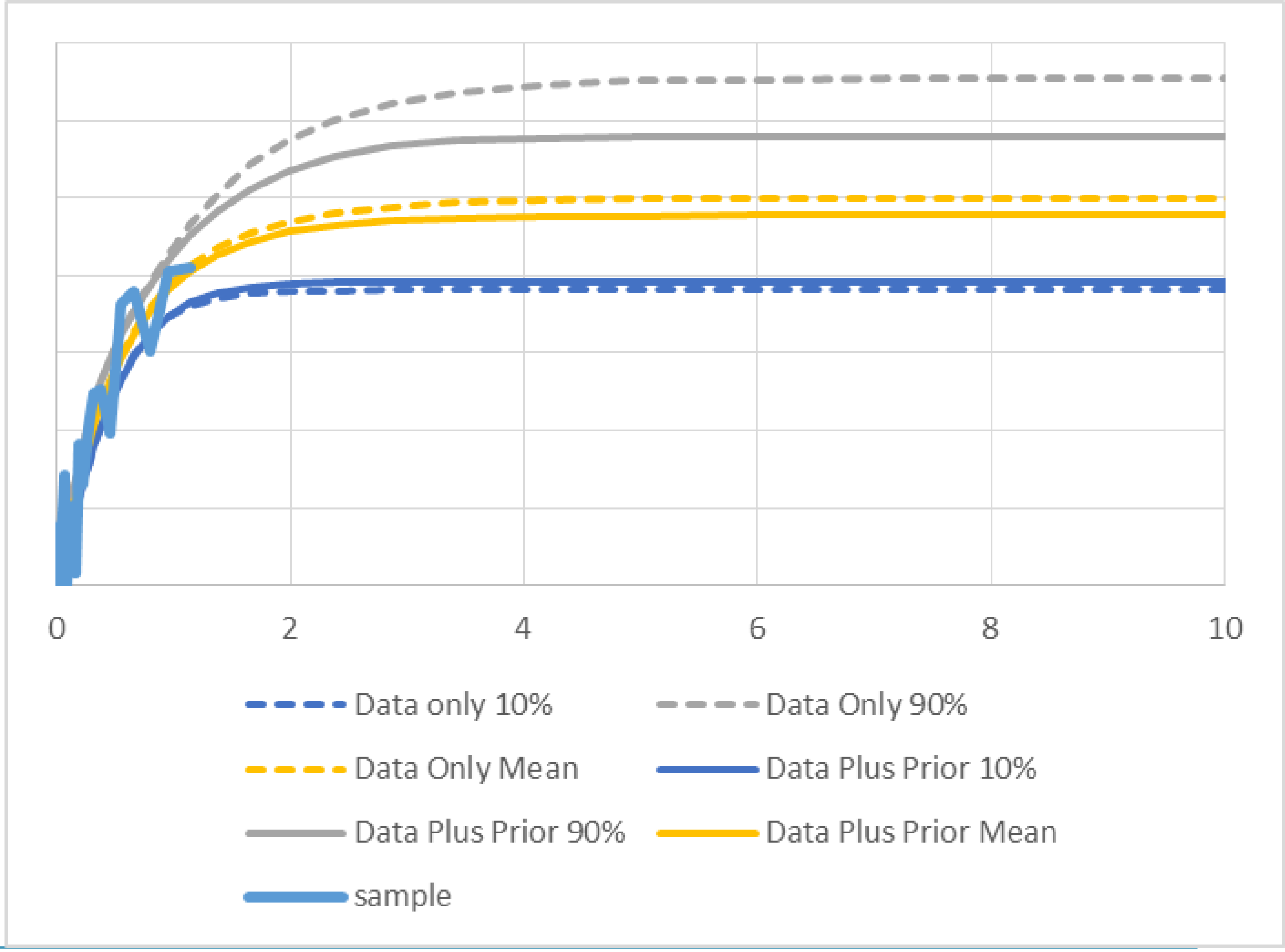
Recap

- The prior is better approximated as a smooth distribution.
- A lognormal prior for Lambda and Ultimate are shown here.
- The prior can be fitted as part of a hierarchical model.



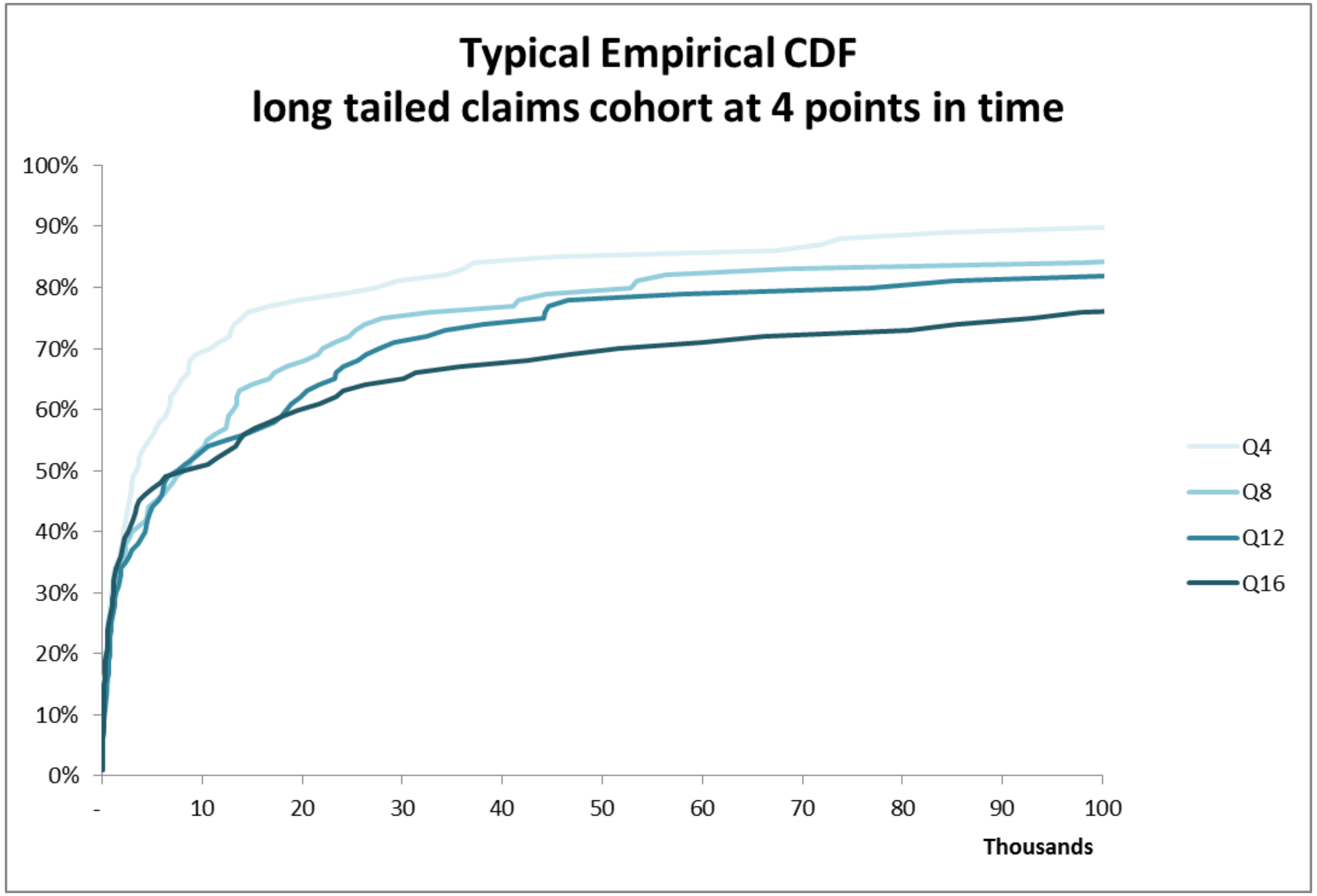
Recap

- Applying a prior based on other years lends credibility to a year with limited samples.
- Reduced mean and spread can result when a combination of the prior and data are used to estimate the range of reasonable ultimates.



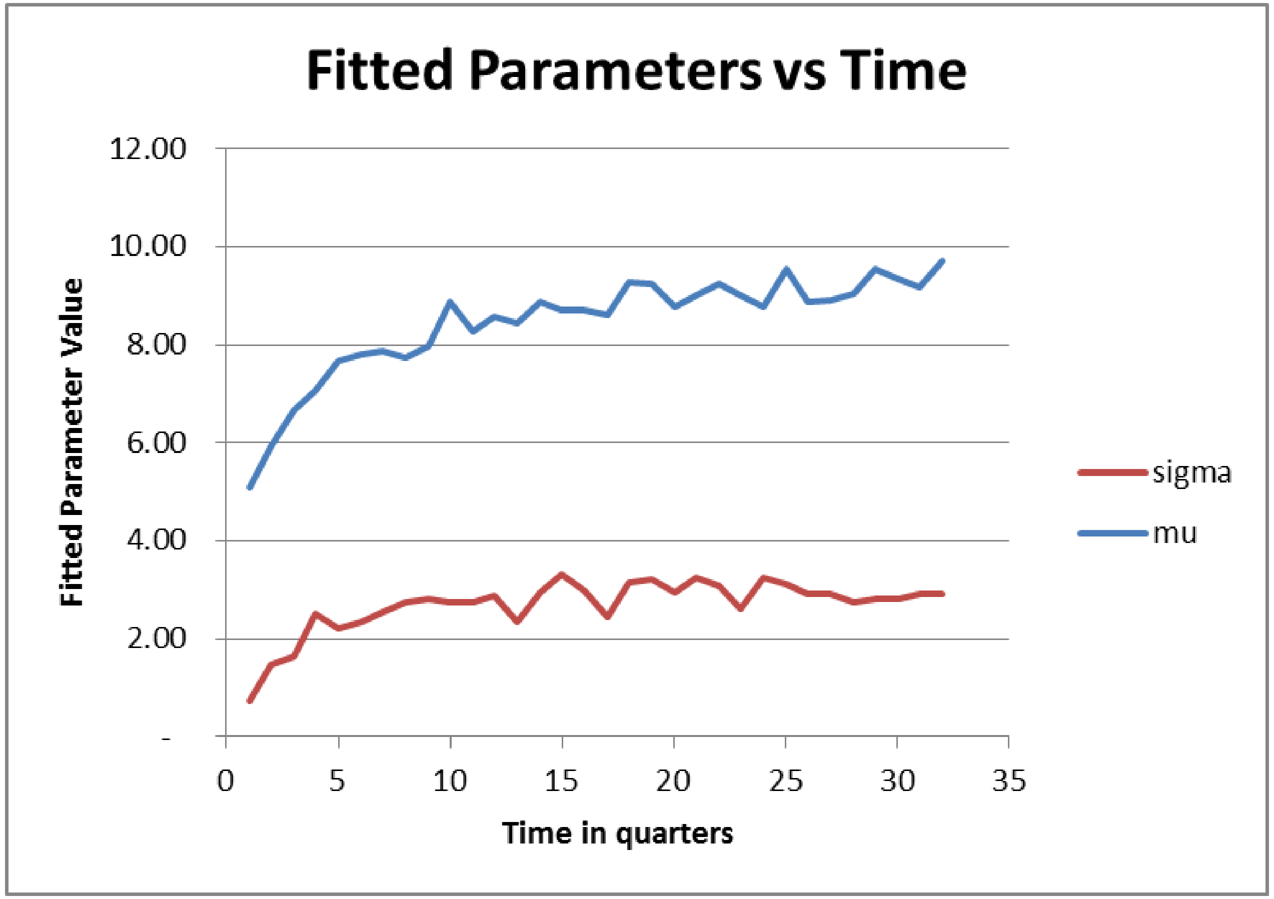
Extensions

- Instead of tracing the path of "incurred" development we can trace the path of some other parameters.
- Opposite is the path of a typical AY cohort of 100 claims as they develop.
- Mean and standard deviation tend to increase with time as heavier claims are reported later.



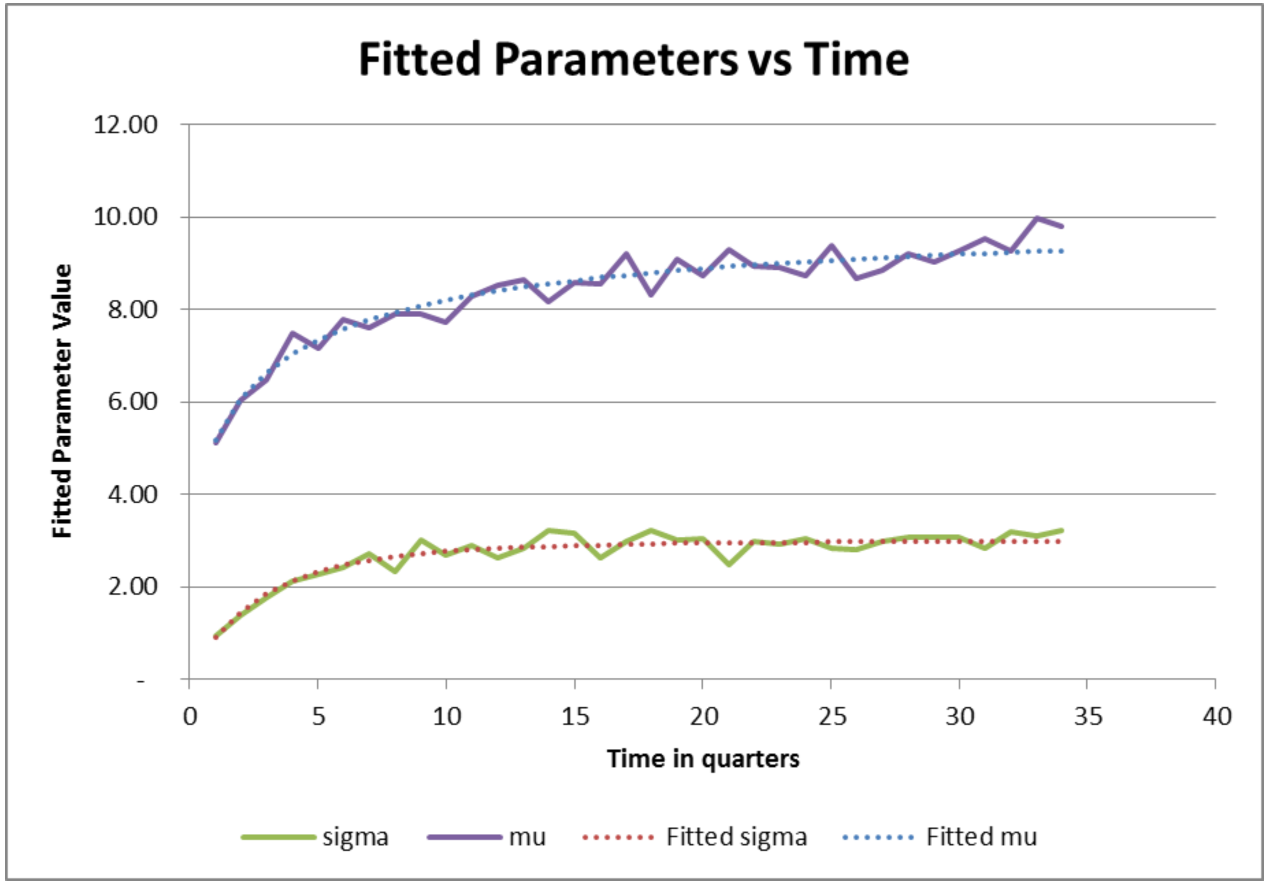
Time series of fitted parameters

- This behaviour can be represented as a trend in the fitted parameters.
- Here we fit a lognormal at each point in time and plot the parameters.
- Mu and Sigma trend much like a development curve.



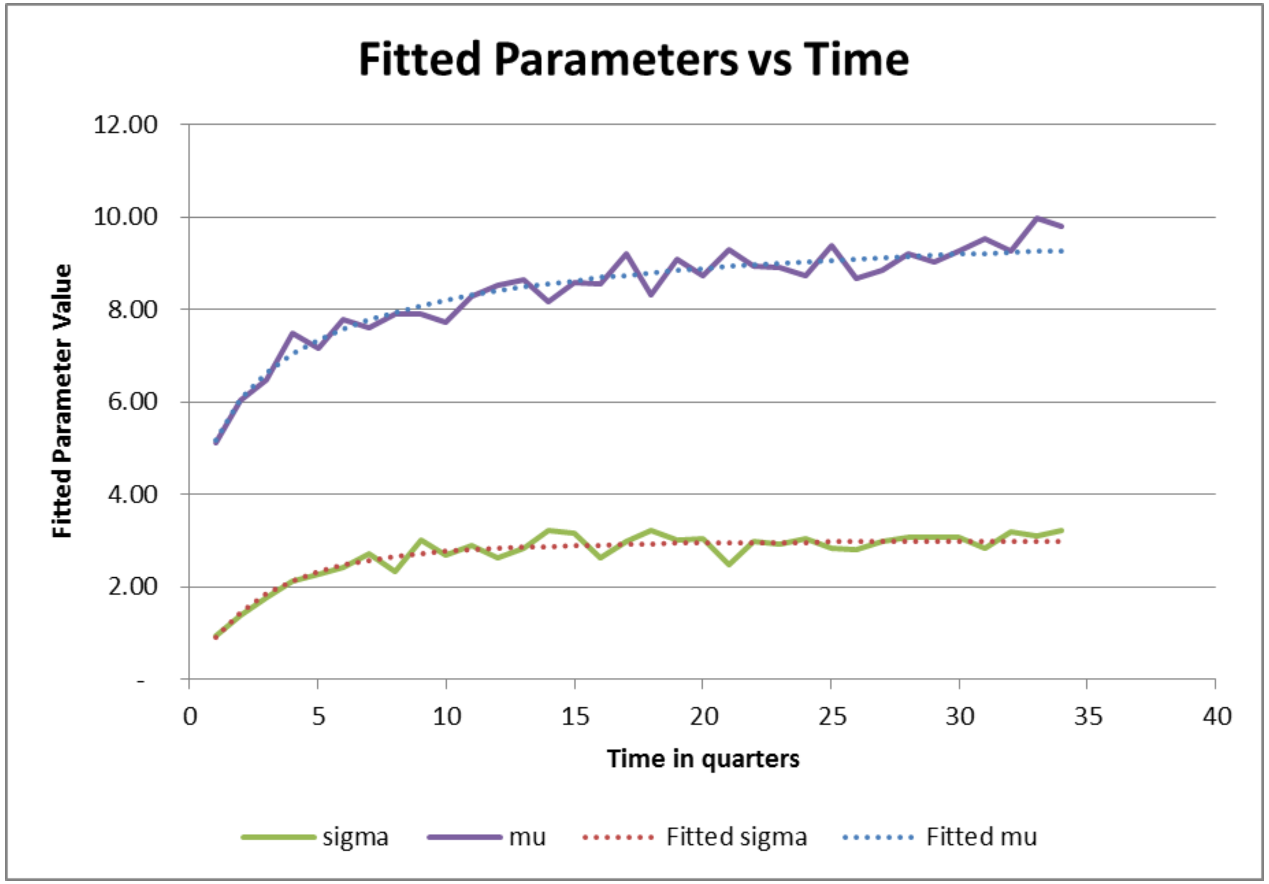
Time series regression

- This behaviour can be modelled with a growth curve.
- Here mu and sigma are fitted using a lognormal CDF with a start and end parameter.
- For example:
  - Mu = start + (end-start)* LNCDF($mu_{development}$,$sigma_{development}$)
- In this case the “ultimate” distribution of claims are given by the ultimate mu and sigma (the “end” parameter).



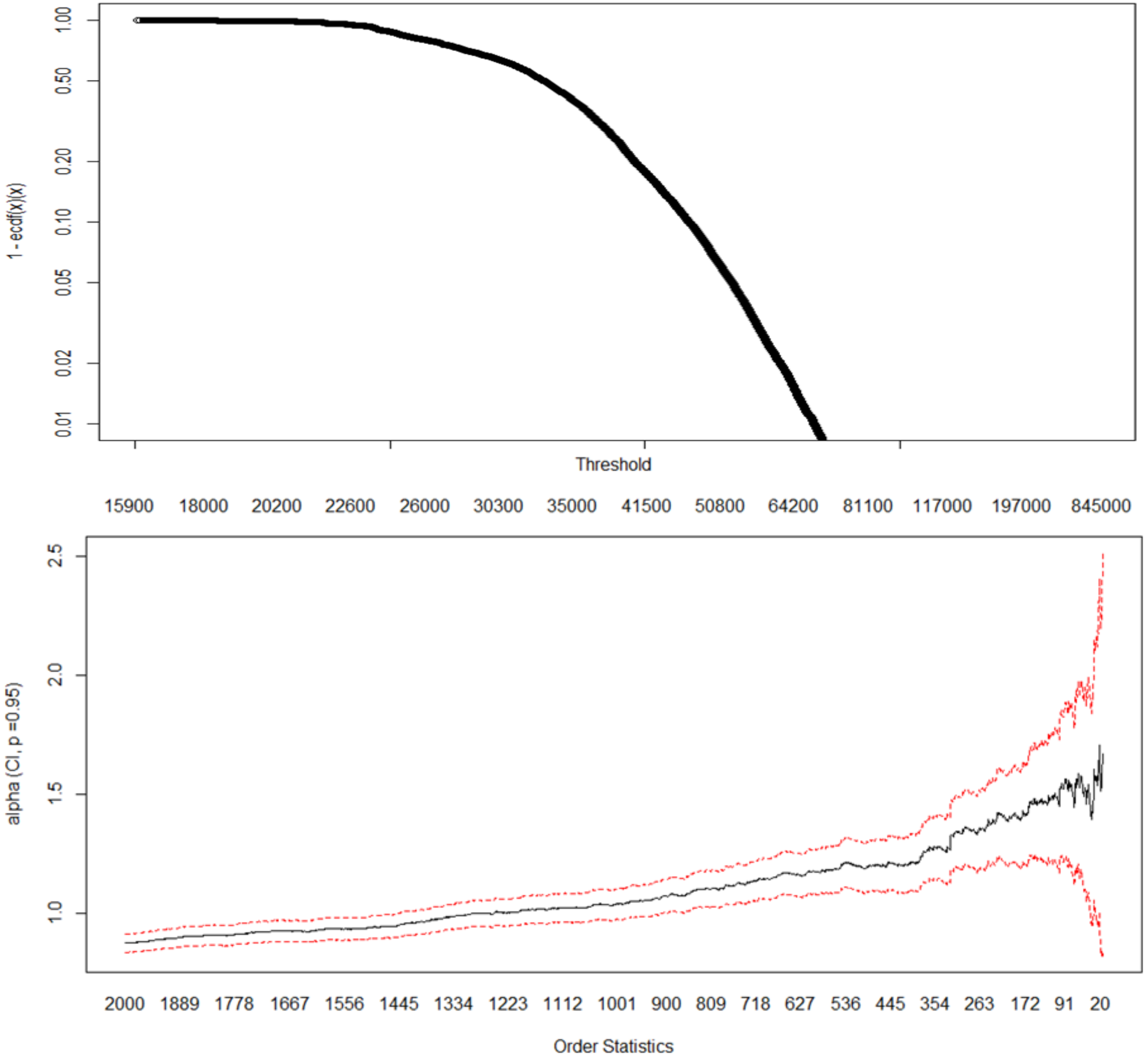
Time series regression

- Such a model does not require development factors.
- Bayesian techniques are best used to estimate mu and sigma.
- The ultimate expected mu and sigma are then given with parameter uncertainty.
- Recent years where there is limited data would utilise a credible prior based on previous years as before.



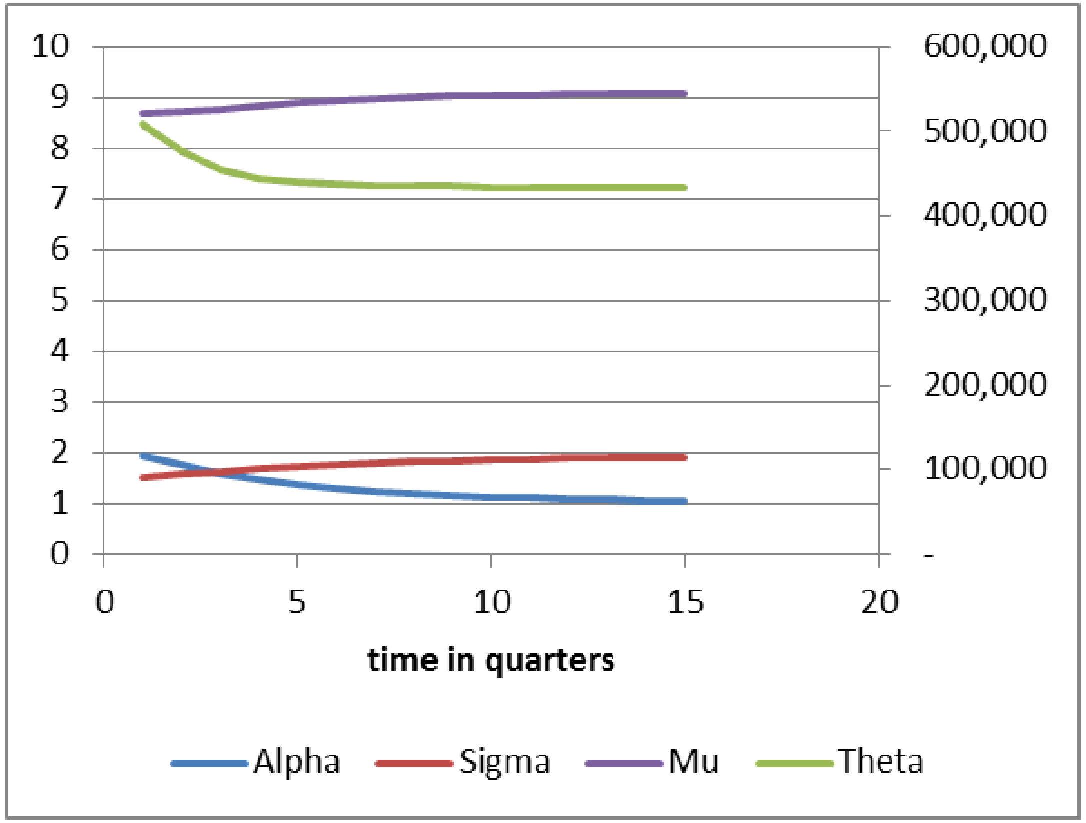
Further extensions to treat extreme events

- Some loss data may be described using Extreme Value Theory type distributions.
- A typical plot is a log-log survival plot or Hill plot shown opposite.
- Linear behaviour on the upper plot would better be modelled by a Pareto distribution.
- A Hill plot (lower right) would show stable fitted alpha above some level.



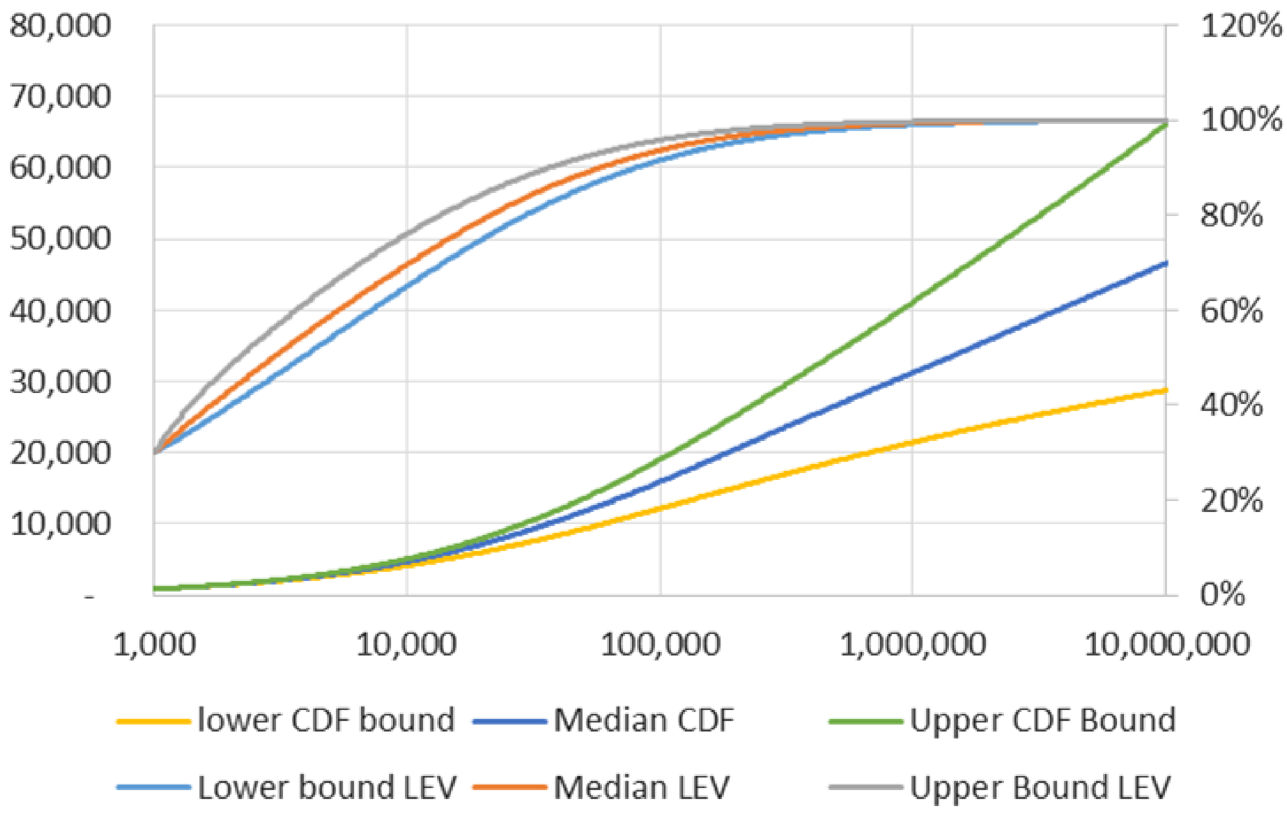
Fitting a Lognormal-Pareto distribution

- The method for fitting a lognormal distribution with a Pareto tail is outlined by Teodorescu, S. (2009).
- The model has three parameters:
  - Alpha, the Pareto distribution parameter
  - Theta, the level at which the Pareto distribution will be fitted
  - Sigma, one of the lognormal parameters
- Other parameters are fixed due to the requirement for the distributions to be continuous and smooth at theta.
- These three parameters are modelled through time just as mu and sigma previously.
- An example fit is shown opposite.



Estimating Ultimate CDF/LEV for a Typical Dataset

- After MCMC we calculate fitted CDFs and LEVs with error bounds.
- Of note is the level of error in the LEV (and therefore any ILF) for the upper layers.



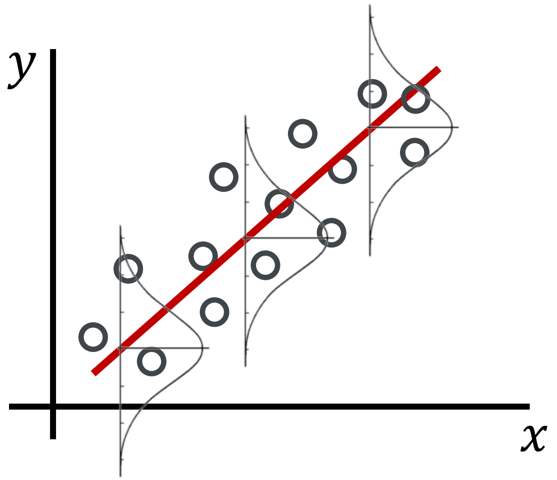
Summary

- Long tailed claims can be modelled as a distribution that changes through time to some ultimate position.
- Each parameter of the distribution can be modelled through time using a growth curve.
- A Lognormal distribution with a Pareto distribution acting in the tail may be useful for including treatment for extreme events seamlessly in your severity model.
- Hierarchical models are useful for projecting undeveloped claims without development factors.
- MCMC methods can provide reasonable measures of uncertainty for parameters such as portfolio ILFs.
- Uncertainty in ILFs may then be useful for credibility based pricing for excess layers using frequency/severity models.

Getting started

- Ordinary Least Squares (OLS)
- Independent variables ($\boldsymbol{X}$), parameters ($\boldsymbol{\beta}$)

$$y = \beta X + \epsilon$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$



- Restate as a probability model:

$$y \sim \mathcal{N}(\beta X, \sigma^2)$$

- Data are modelled as Normal with mean βX and variance σ^2
  - Equivalent, yet more intuitive

Getting started

- Auto claims data:

log_loss	lawyer	gender	seatbelt	age
3.6	yes	male	yes	50
2.4	no	female	yes	28
-1.1	no	male	yes	5
2.4	yes	male	no	32
-2.0	no	male	yes	30
…	…	…	…	…

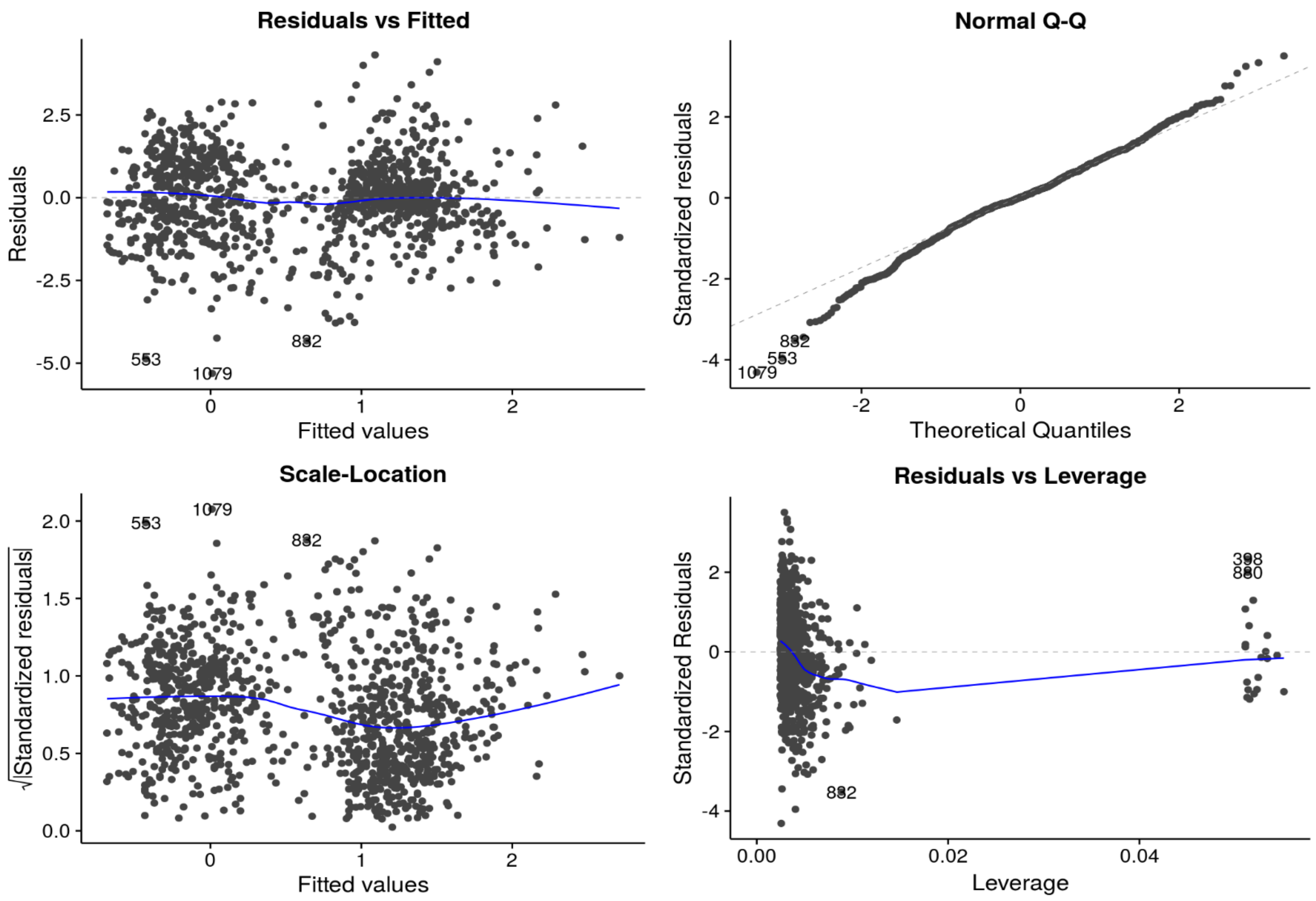
- Linear model for loss cost:

$$\text{log\_loss} = \beta_0 + \beta_1\text{lawyer} + \beta_2\text{gender} + \beta_3\text{seatbelt} + \beta_4\text{age}$$

- R implementation:

```
model_lm <- lm(log_loss ~ lawyer + seatbelt + gender + age,
               data = data)
```

Getting started



Getting started

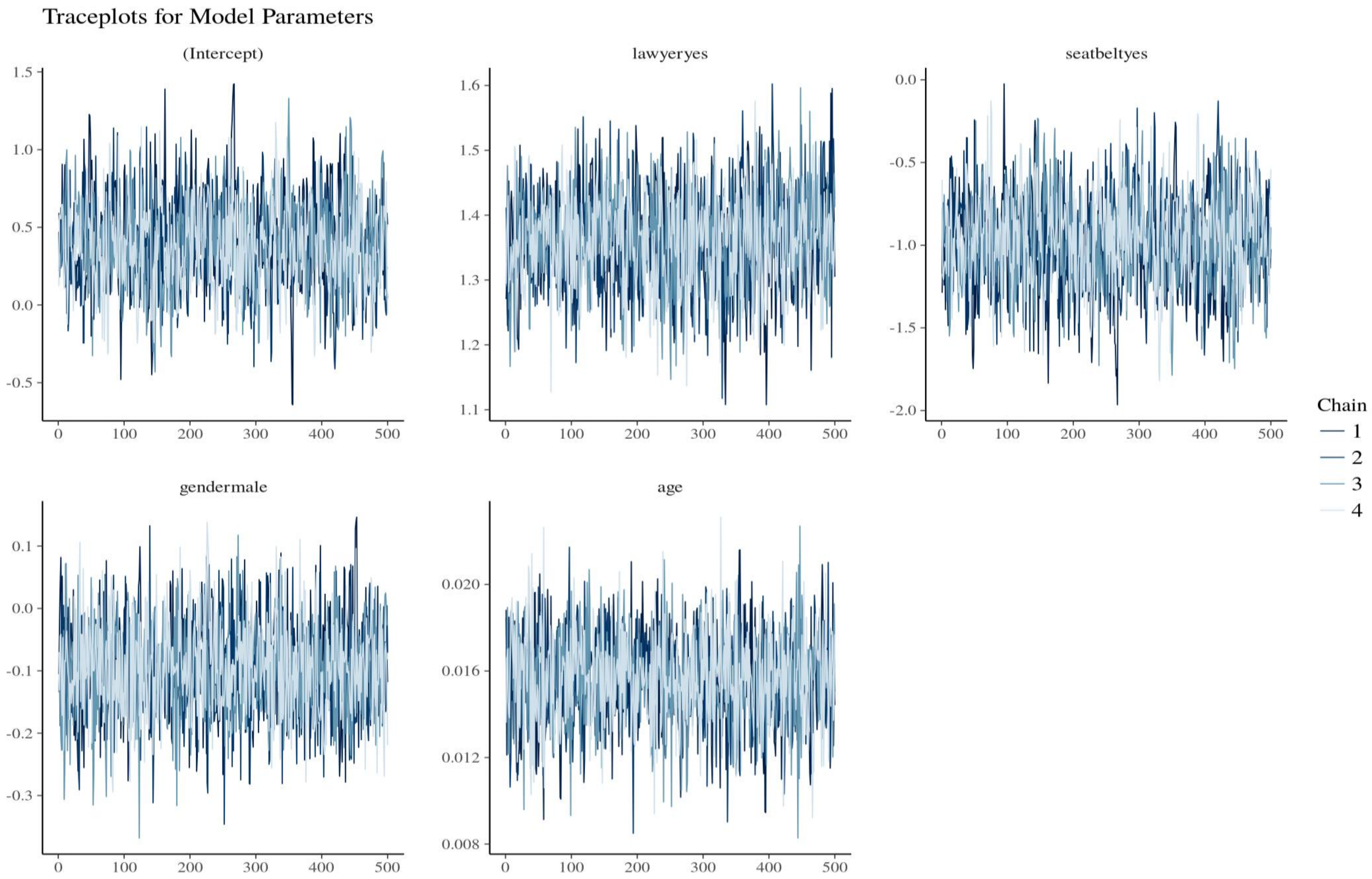
- Bayesian version?
  - Setting up from scratch in Stan → time/effort

- R package ‘rstanarm’ reduces coding requirements
  - Pre-built Stan models (e.g. linear models, GLMs, ANOVA…)
  - R syntax relatively simple:

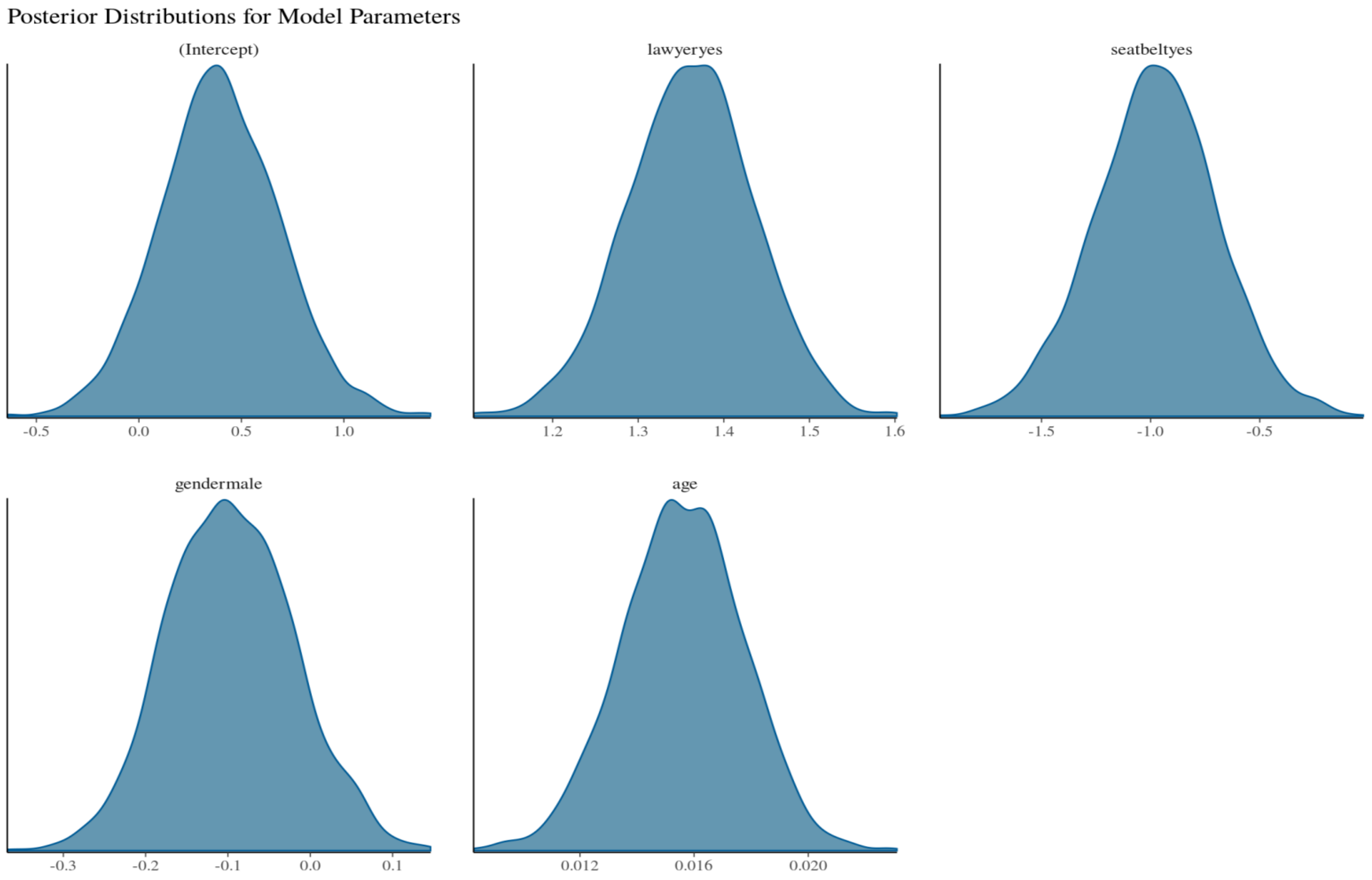
```
model_stanlm <- stan_lm(log_loss ~ lawyer + seatbelt + gender + age,
                        prior = R2(location = 0.8),
                        data = data)
```

- Offers various outputs…

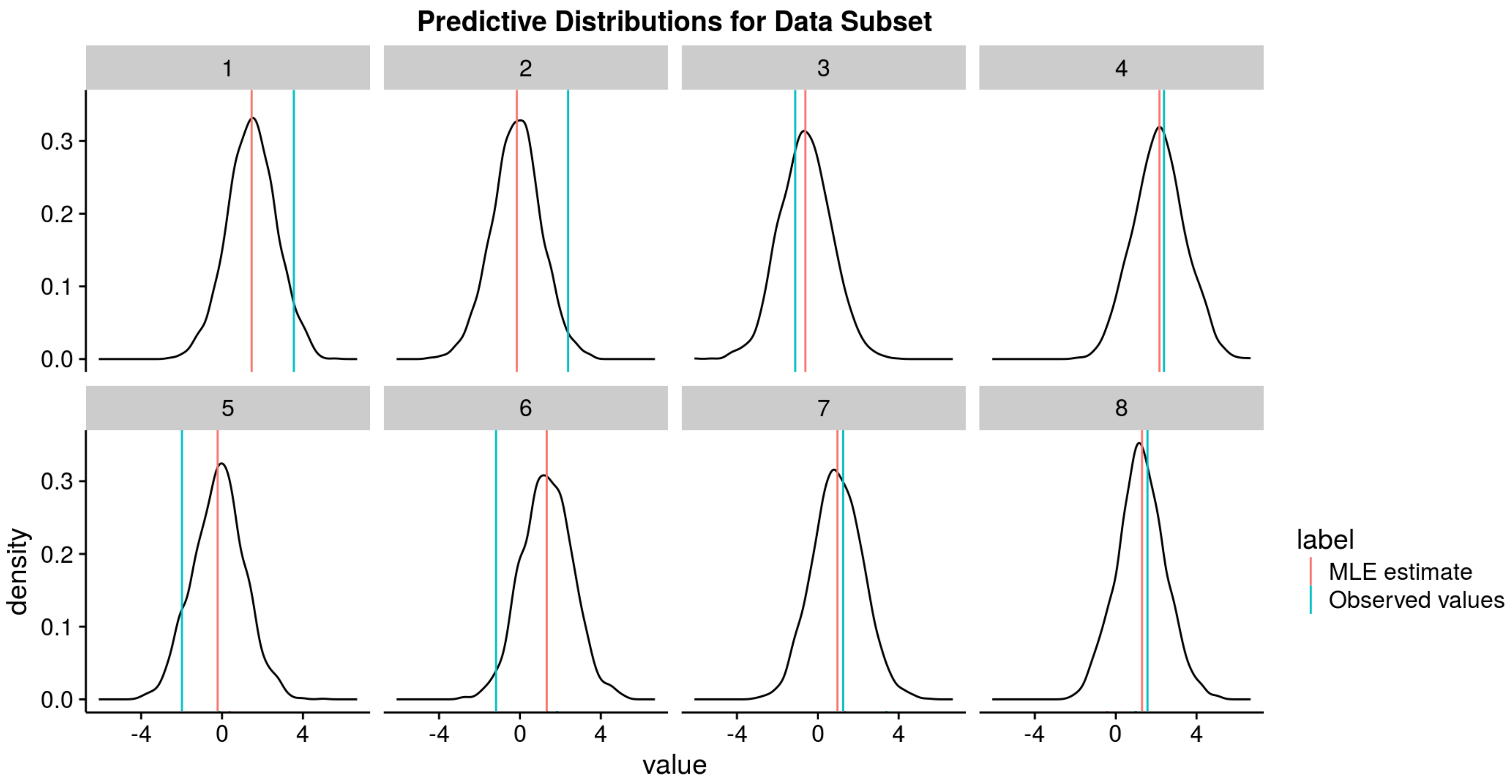
Getting started



Getting started



Getting started



Conclusions

- Bayesian methods offer a variety of benefits
  - Reflect uncertainty, model flexibility, external data/judgement, hierarchical models, …
- Numerous potential actuarial applications
  - Reserving, pricing, profitability studies, portfolio optimisation, ...
- Learning curve → ‘rstanarm’ a good place to start

> *“Scientific disciplines from astronomy to zoology are moving to Bayesian data analysis.* ***We should be leaders of the move, not followers****.”*
>
> \- *John K. Kruschke (2010)*

Further Reading

Hierarchical Growth Curve Models for Loss Reserving - Guszcza (*CAS Forum 2008)*

Hierarchical Compartmental Models for Loss Reserving - Morris (*CAS E-Forum, Summer 2016)*

On the Truncated Composite Lognormal-Pareto Model - Teodorescu (2009)

Doing Bayesian Data Analysis - John Kruschke

Statistical Rethinking - Richard McElreath

Data Analysis Using Regression and Multi-level/Hierarchical Models - Gelman and Hill

An Introduction to Statistical Learning - Tibshirani and Hastie

Stan Documentation - (tutorials, case studies, etc)

Modelling Loss Curves in Insurance with RStan (Stan Case Study) - Cooney

Open Actuarial - Various

Questions

Comments

The views expressed in this [publication/presentation] are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this [publication/presentation] and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this [publication/presentation].

The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this [publication/presentation] be reproduced without the written permission of the IFoA [*or authors, in the case of non-IFoA research*].