LDF Curve-Fitting: Incorporating Prior Information

David R. Clark, FCAS

Casualty Loss Reserve Seminar – September 2019

Agenda

- Loss Development and Growth Curves
  - Description of Model
  - LDF and Cape Cod Forms
  - Variance Definition and Fitting Criteria

- Blending Client Data with Benchmark Curves
  - Bayesian Credibility
  - Cross Validation
  - Example

Loss Development and Growth Curves

Business Context:

Goal is to improve estimation of loss development patterns for individual clients.

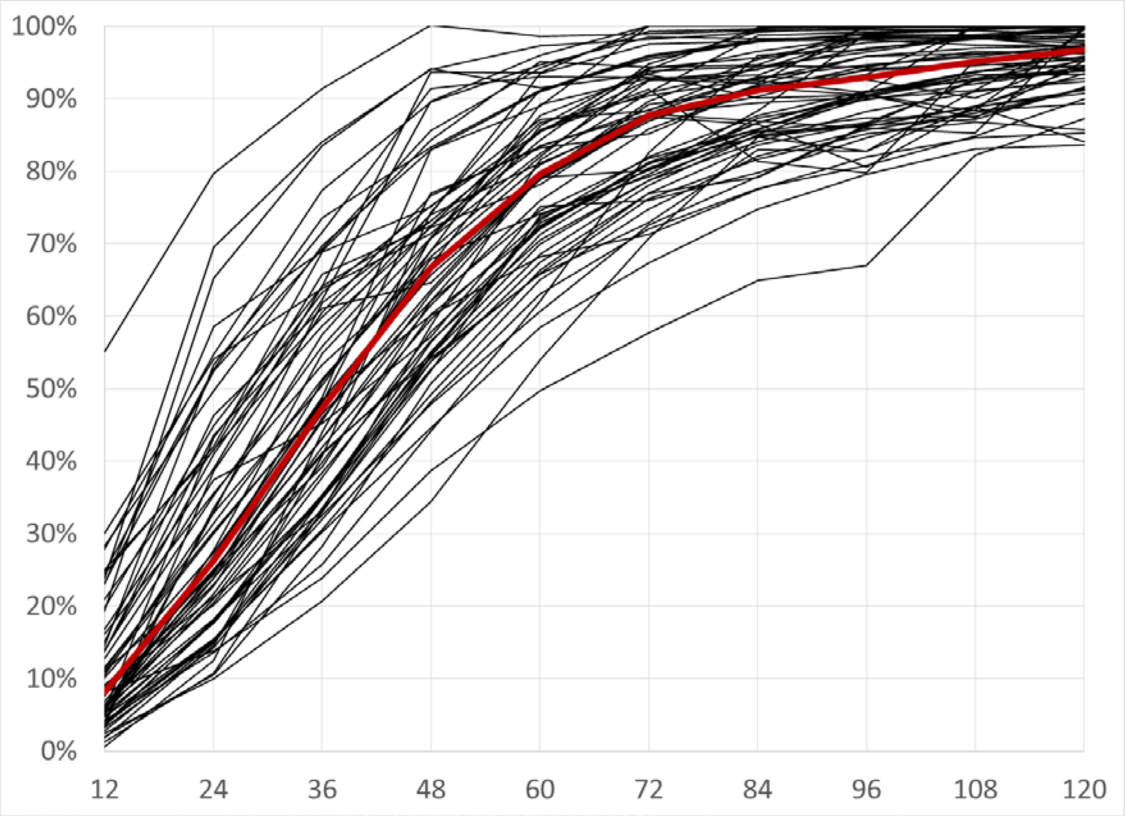
A mathematical model is introduced and expanded to include:

- Exposure information
- Benchmark patterns

Principle: Including more information improves our estimate

Loss Development and Growth Curves

Description of Model



Loss development patterns trace the aggregate amount of paid or case incurred loss over time.

General pattern is increasing, ultimately reaching 100%.

Note: This example shows a collection of third party liability (TPL) paid claims for US insurance companies. Each line is a different company.

Median of the sample shown in red.

Source: SNL Schedule P

Loss Development and Growth Curves
Description of Model

The pattern of development can be modelled as a “growth curve” smoothly moves from 0% to 100% over time.

Two forms are useful:

- Weibull (aka “Craighead”)

$$G(t|\theta,\omega) = 1 - exp\left(-\left(\frac{t}{\theta}\right)^{\omega}\right)$$

- Loglogistic (aka “inverse power”)

$$G(t|\theta,\omega) = \left[1 + \left(\frac{\theta}{t}\right)^{\omega}\right]^{-1} = 1 - \left(\frac{\theta^{\omega}}{t^{\omega} + \theta^{\omega}}\right)$$

Each of these has a “scale” parameter (theta), and a “shape” parameter (omega).

Loss Development and Growth Curves
Description of Model

Advantages of using fitted curves:

- Reduce over-parameterization
- Smooth and extrapolate the pattern – including help with tail factor selection
- Handle irregular evaluation dates
  (e.g., latest diagonal less than 12 months)
- Estimate variance or ranges

Loss Development and Growth Curves

Description of Model

Limitations and Model Assumptions:

- Pattern is the same for all accident years
- Growth curve form selection is correct
- Requires curve-fitting engine: fast, but generally not “real time”

Loss Development and Growth Curves

LDF and Cape Cod Forms

We begin with incremental losses in each “cell” of a development triangle.

“LDF Form” lets each accident year stand on its own.

The expected loss is estimated based on the growth curve and an estimated ultimate loss for each accident year.

Number of parameters = 2 + number of accident years

$$\hat{\mu}_{i,t} = Ult_i \cdot [G(t|\theta,\omega) - G(t-1|\theta,\omega)]$$

Loss Development and Growth Curves

LDF and Cape Cod Forms

An alternative form is based on the “Cape Cod” or “Additive” method, if we have an exposure base for each year.

Onlevel (indexed) premium may be used.

This form reduces the number of parameters to be estimated to three:

ELR, Scale (theta), and Shape (omega)

$$\hat{\mu}_{i,t} = Premium_i \cdot ELR \cdot [G(t|\theta,\omega) - G(t-1|\theta,\omega)]$$

Loss Development and Growth Curves

LDF and Cape Cod Forms

The names for these two methods are based on the fact that the Ultimate Loss and ELR parameters are based on the MLE estimates.

For LDF method:

$$\widehat{Ult}_i = \left(\frac{1}{G(t_i|\theta,\omega)}\right) \cdot \frac{\sum_t c_{i,t}}{Premium_i}$$

For Cape Cod method:

$$\widehat{ELR} = \frac{\sum_i \sum_t c_{i,t}}{\sum_i Premium_i \cdot G(t_i|\theta,\omega)}$$

Loss Development and Growth Curves

Variance Definition and Fitting Criteria

To fit the model, it is convenient to borrow the "over-dispersed Poisson" (ODP) from the world of Generalized Linear Models (GLM).

The ODP GLM model can reproduce the traditional chain-ladder method when the model has a separate parameter for each development period.

The key assumption is that variance of loss is proportional to the mean, and that this proportion is a constant for all years and development ages.

Loss Development and Growth Curves
Variance Definition and Fitting Criteria

The ODP GLM fit is performed using “quasi-likelihood”.

We borrow the form of the Poisson likelihood to estimate the parameters of the growth curve. When needed, the variance-to-mean ratio is separately approximated using a chi-square statistic.

This allows for a quick model fit, and an approximation for standard error on the parameters.

$$\sum_{i,t} c_{i,t} \cdot ln(\hat{\mu}_{i,t}) - \hat{\mu}_{i,t}$$

$$\widehat{\sigma^2} \approx \frac{1}{n-p} \cdot \sum_{i,t} \frac{\left(c_{i,t} - \hat{\mu}_{i,t}\right)^2}{\hat{\mu}_{i,t}}$$

Variance Definition and Fitting Criteria

Should the uncertainty in the parameters change our selected development pattern?

In original paper, we recommend using the growth curves with the "best" fitted parameters:

$$G\left(t|\hat{\theta}_{MLE}, \hat{\omega}_{MLE}\right)$$

An alternative is to use the expected growth curve, averaging over all the possible parameters:

$$E_{\theta,\omega}[G(t|\theta,\omega)] = \int G(t|\theta,\omega) \cdot f(\theta,\omega)\, d\theta\, d\omega$$

Blending Client Data with Benchmark Curves

Bayesian Credibility

The model described so far works well for a development triangle on a single client. The growth curve smooths out the pattern and is robust against [some] negative development or outlier points.

For small clients or volatile lines of business, the fit does not always help. The model parameters can “blow up” and produce unrealistic results.

We want to stabilize the results by incorporating more information. Thomas Bayes to the rescue!

Blending Client Data with Benchmark Curves
Bayesian Credibility

Bayesian credibility follows Bayes' Theorem (technically: Laplace's generalization of Bayes' original work).

$$Pr(\theta_k|data) = \frac{Pr(data|\theta_k) \cdot Pr(\theta_k)}{\sum_k Pr(data|\theta_k) \cdot Pr(\theta_k)}$$

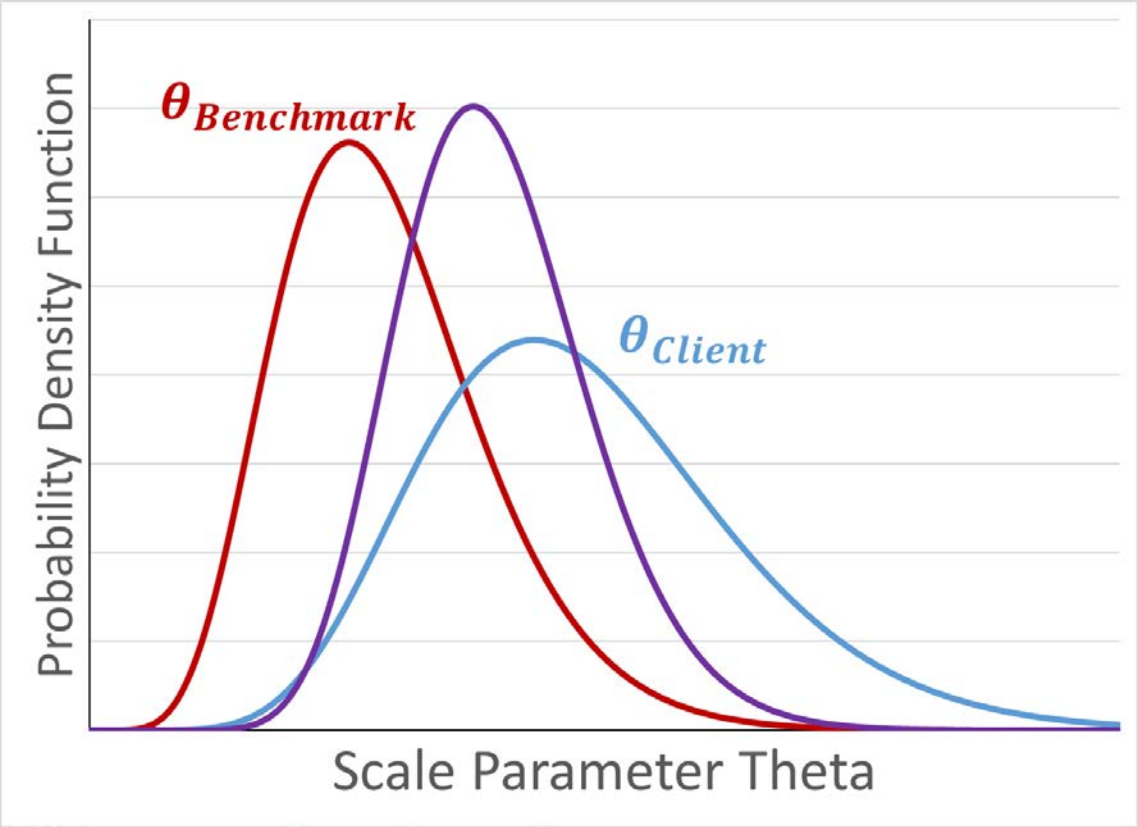
The idea is that we have some prior knowledge of the range of possible parameter values. When new data comes in, we gain more knowledge about the parameters.

For our application, the prior knowledge is our benchmark curve including some measure of uncertainty when it is used for a given client.

Prior is updated when we see the client's development data.

Blending Client Data with Benchmark Curves

Bayesian Credibility



The blended curve is the *posterior* distribution of the shape parameter theta.

It represents the uncertainty in the shape parameter after both the benchmark and the client information have been included.

Blending Client Data with Benchmark Curves
Bayesian Credibility

Key idea: We have a mixture model.

There is a standard statistical model (in this case the ODP model with growth curve). There is also a mixing distribution on the parameters of the model.

The total variance therefore has two components:

- Expected process variance = random outcomes
- Parameter variance = “prior” uncertainty around the parameters themselves

The relative magnitude of these two variances determines how much credibility we assign to the client data relative to the benchmark.

Blending Client Data with Benchmark Curves

Bayesian Credibility

We will assume that each company has a growth curve that follows a Weibull pattern. All companies share the same "shape" (omega) but have different "scale" (theta) parameters.

The spread of the "scale" parameters is centered and controlled by a prior mixing distribution. The mixing distribution is an inverse transformed gamma, where the alpha parameter controls the spread.

The mixed version is a Burr curve.

$$G(t|\theta,\omega) = 1 - exp\left(-\left(\frac{t}{\theta}\right)^{\omega}\right)$$

$$f(\theta) = \frac{\omega \cdot \beta^{\alpha \cdot \omega}}{\theta^{\alpha \cdot \omega + 1} \cdot \Gamma(\alpha)} \cdot exp\left(-\left(\frac{\beta}{\theta}\right)^{\omega}\right)$$

Blending Client Data with Benchmark Curves

Bayesian Credibility

A growth curve following a “Burr” shape adds one additional parameter (alpha). The other curves are special cases.

The Burr can also be derived as a “mixed” version of the Weibull curve.

Burr

$$G(t|\theta,\omega) = 1 - \left(\frac{\alpha\cdot\theta^{\omega}}{t^{\omega}+\alpha\cdot\theta^{\omega}}\right)^{\alpha}$$

$\alpha = 1$

$\alpha \to \infty$

Loglogistic

$$G(t|\theta,\omega) = 1 - \left(\frac{\theta^{\omega}}{t^{\omega}+\theta^{\omega}}\right)$$

Weibull

$$G(t|\theta,\omega) = 1 - exp\left(-\left(\frac{t}{\theta}\right)^{\omega}\right)$$

Blending Client Data with Benchmark Curves

Bayesian Credibility

A “benchmark” growth curve takes the “Burr” form:

$$G(t) \;=\; 1 - \left(\frac{\alpha \cdot \theta_{Benckmark}^{\omega}}{t^{\omega} + \alpha \cdot \theta_{Benckmark}^{\omega}}\right)^{\alpha}$$

$$\alpha \approx \frac{1}{CV_{\theta_{Benchmark}^{\omega}}^{2}}$$

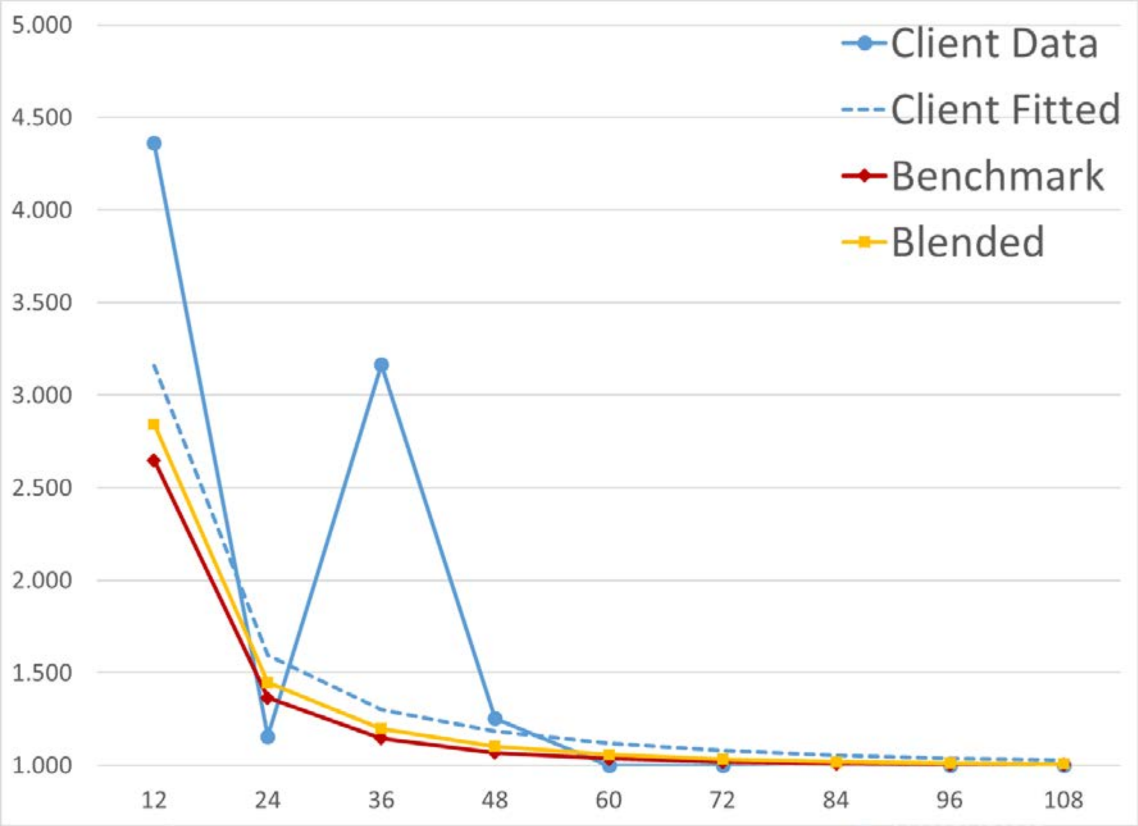
The blended growth curve includes additional terms from the fit to client data:

$$G(t) \;=\; 1 - \left(\frac{\alpha \cdot \theta_{Benckmark}^{\omega} + c \cdot \theta_{Client}^{\omega}}{t^{\omega} + \alpha \cdot \theta_{Benckmark}^{\omega} + c \cdot \theta_{Client}^{\omega}}\right)^{\alpha + c}$$

$$c \approx \frac{1}{CV_{\theta_{Client}^{\omega}}^{2}}$$

Blending Client Data with Benchmark Curves

Bayesian Credibility



The age-to-age factors (link ratios) for an individual client are shown in blue in this graph.

The high volatility of the client data means there is little credibility assigned to their data.

The blended pattern (yellow) is still a change from the benchmark (red).

Blending Client Data with Benchmark Curves
Bayesian Credibility

For the credibility constant assigned to the client data:

This is derived from the standard error of the fitted growth curve parameters.

However, this standard error may be understated and need some adjustment.

- The model assumes incremental losses are independent and identically distributed (i.i.d.)
- The use of the Hessian matrix to approximate parameter variance is a lower bound that becomes accurate in large samples, but our data is small sample
- The variance/mean parameter is assumed to be fixed and known

Blending Client Data with Benchmark Curves

Bayesian Credibility

How do we set the spread around the benchmark parameter?

Subjective Bayes:

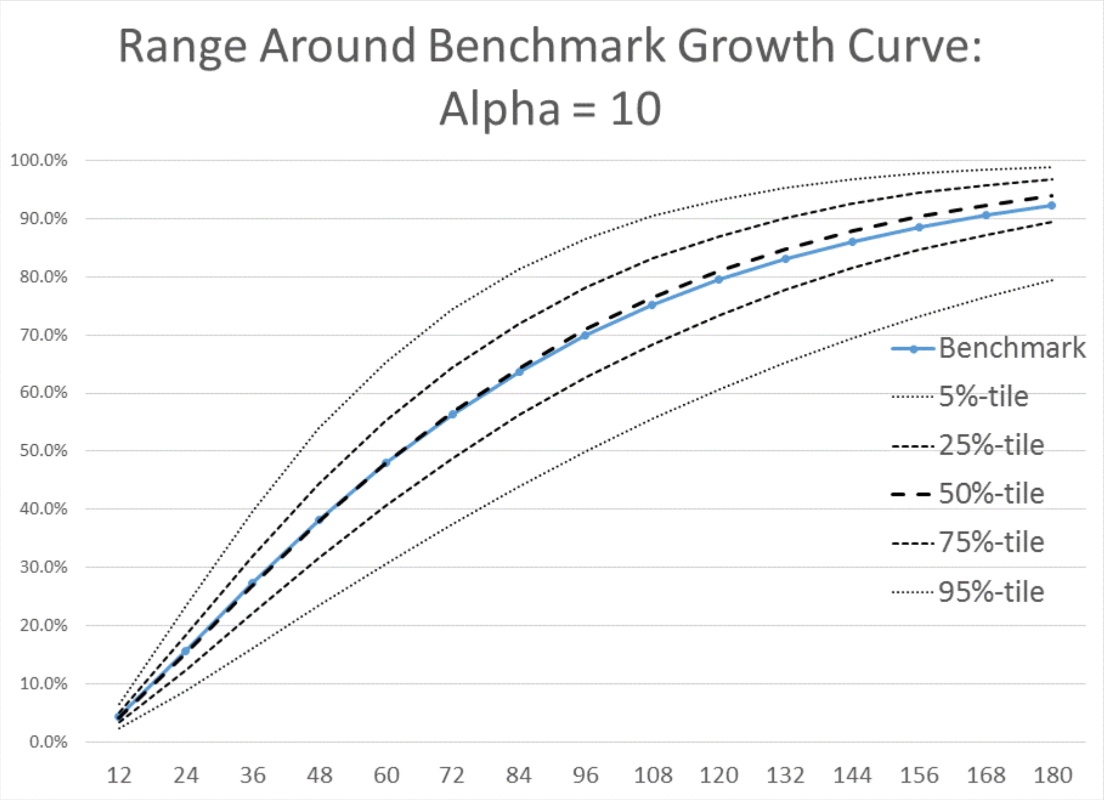
- Business expertise selects the range of possible values
  For example: how much faster or slower than average can a company settle its claims?

Empirical Bayes:

- How much actual spread is there among the companies?
- This is equivalent to what data scientists call estimation via “cross validation”

Blending Client Data with Benchmark Curves

Bayesian Credibility

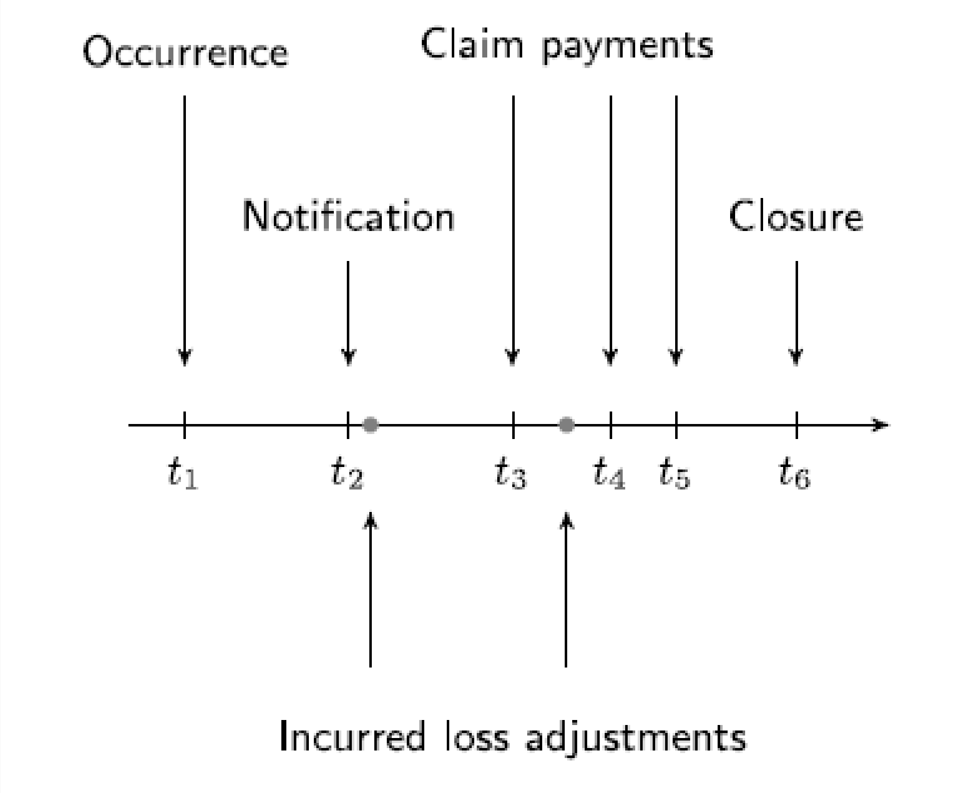


Big alpha means a narrow range of possible values around the benchmark.

Small alpha means a wider range of patterns across companies.

Alpha=1 is a very wide range and reproduces the log-logistic curve.

Blending Client Data with Benchmark Curves
Bayesian Credibility



Source: Pigeon, et al (2014)

Shifting the scale parameter (theta) can be interpreted as a company speeding up or slowing down the steps of the processing of an “average” claim.

Other considerations:

- Mix of from-ground-up and excess policies
- Use of Third Party Administrators
- Technology / automated claims

Blending Client Data with Benchmark Curves
Cross Validation

We note that credibility theory is similar to “shrinkage” methods (Ridge Regression, LASSO, etc.) where parameters are constrained or “regularized” to stay within a given range. This is accomplished by introducing a tuning parameter that shrinks parameters towards an average value.

Cross validation is a method for estimating this tuning parameter.

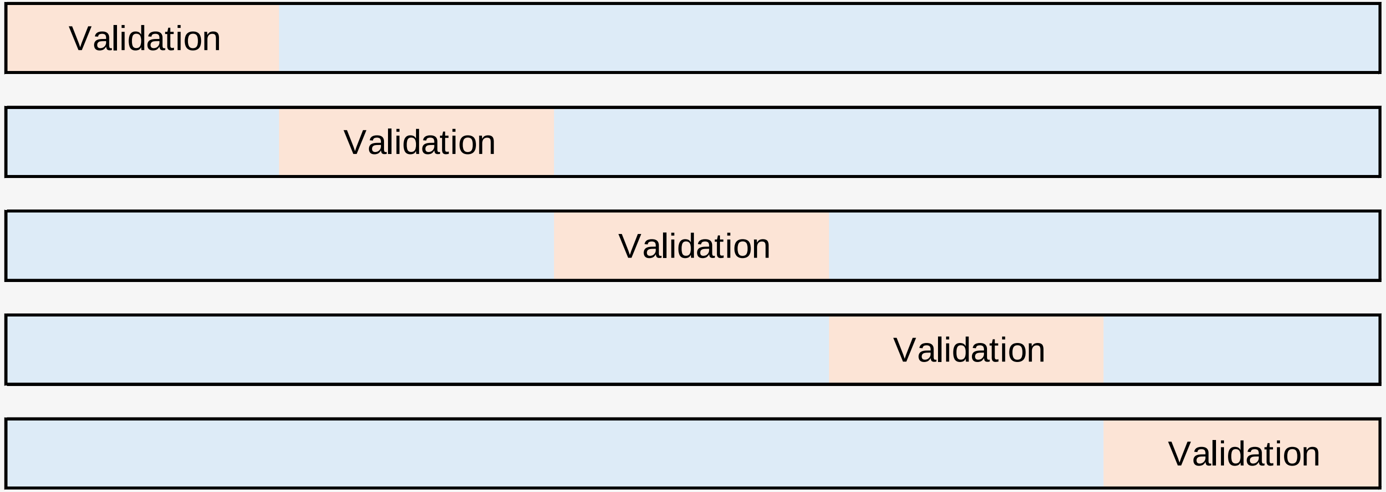
- Split the data into training and validation (hold-out) sets
- For a grid of tuning parameters, estimate the error in predicting the hold-out data
- Repeat for multiple validation sets
- Select the tuning parameter that produces the smallest prediction error

See for example James, et al “An Introduction to Statistical Learning” pages 214-227 of the sixth edition.

Blending Client Data with Benchmark Curves
Cross Validation

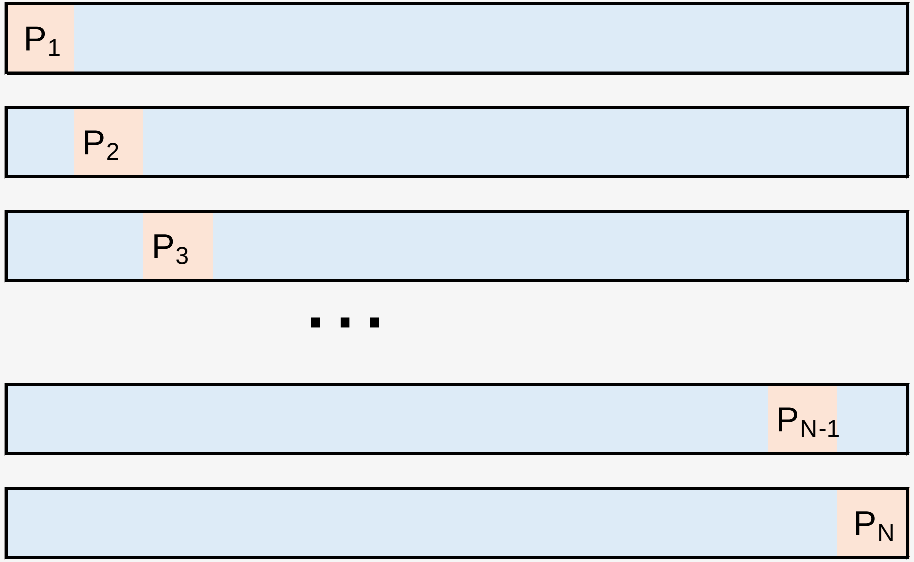
Five-fold cross validation runs the fitting exercise five time, each time leaving one fifth of the data out of the fit.

The best tuning parameter reducing the error when tested against each hold-out set.



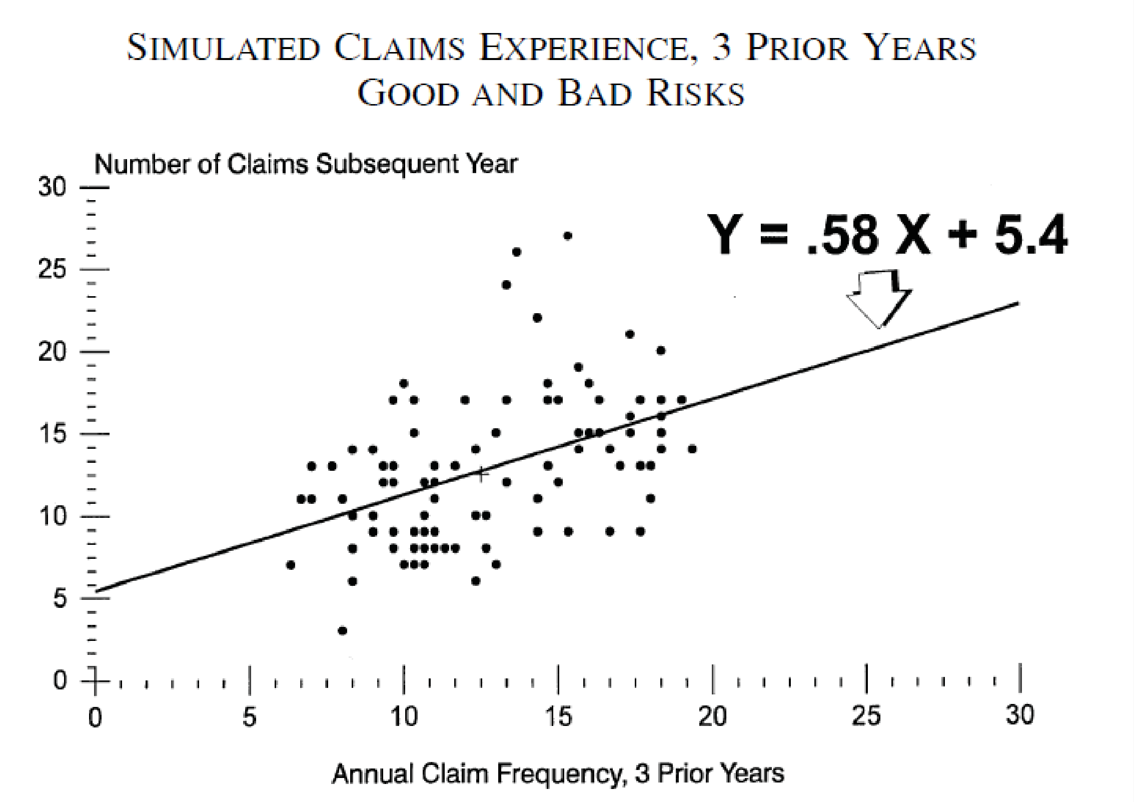
Blending Client Data with Benchmark Curves
Cross Validation

“Leave one out” (LOO) is the extreme version of cross validation, in which we calculate a validation error for each point in the data.



Blending Client Data with Benchmark Curves

Cross Validation



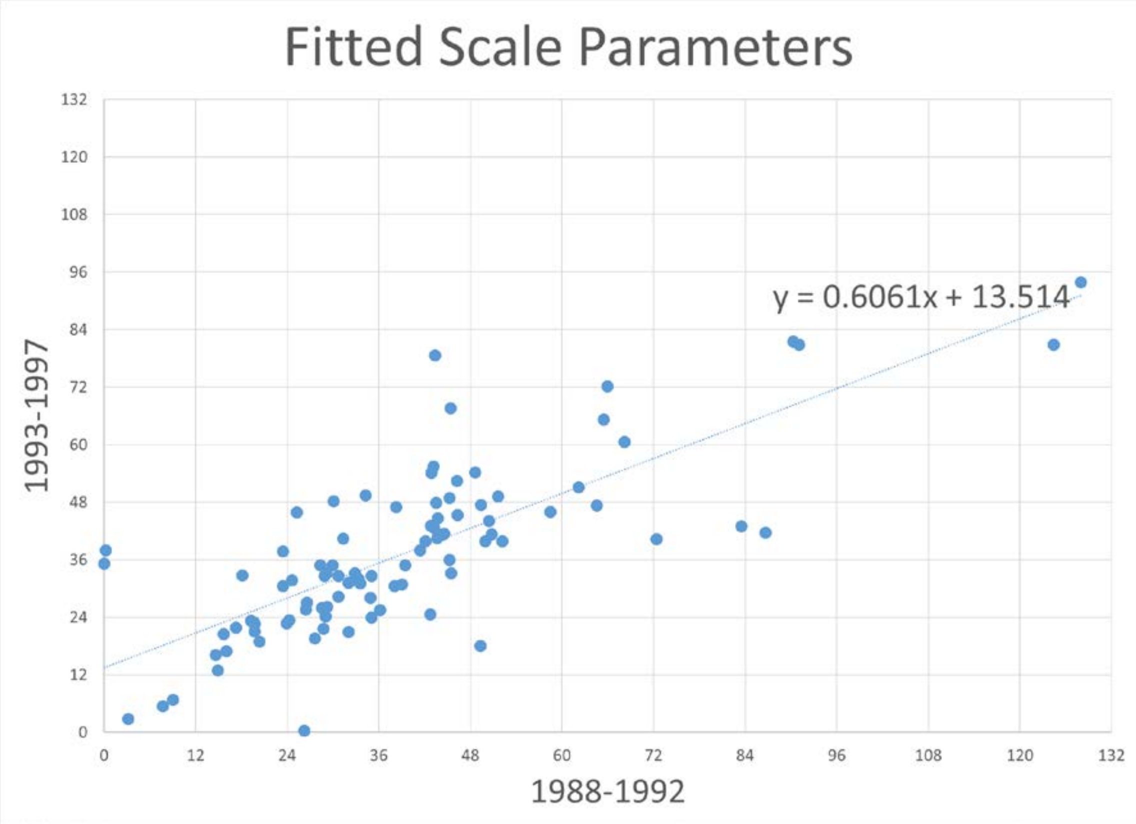
Actuaries often use similar technique, viewed as a time series with the prospective period being the “hold-out” data.

We can perform a similar exercise with the development data by company.

Source: Mahler, “A Graphical Illustration of Experience Rating Credibility”, PCAS 1998

Blending Client Data with Benchmark Curves

Cross Validation



For each company, we split the data into two blocks of years. The credibility for an "average" company is the slope of the fitted line.

$$\theta_{93-97} = Z \cdot \theta_{88-92} + (1 - Z) \cdot \theta_{Benchmark}$$

Blending Client Data with Benchmark Curves

Cross Validation

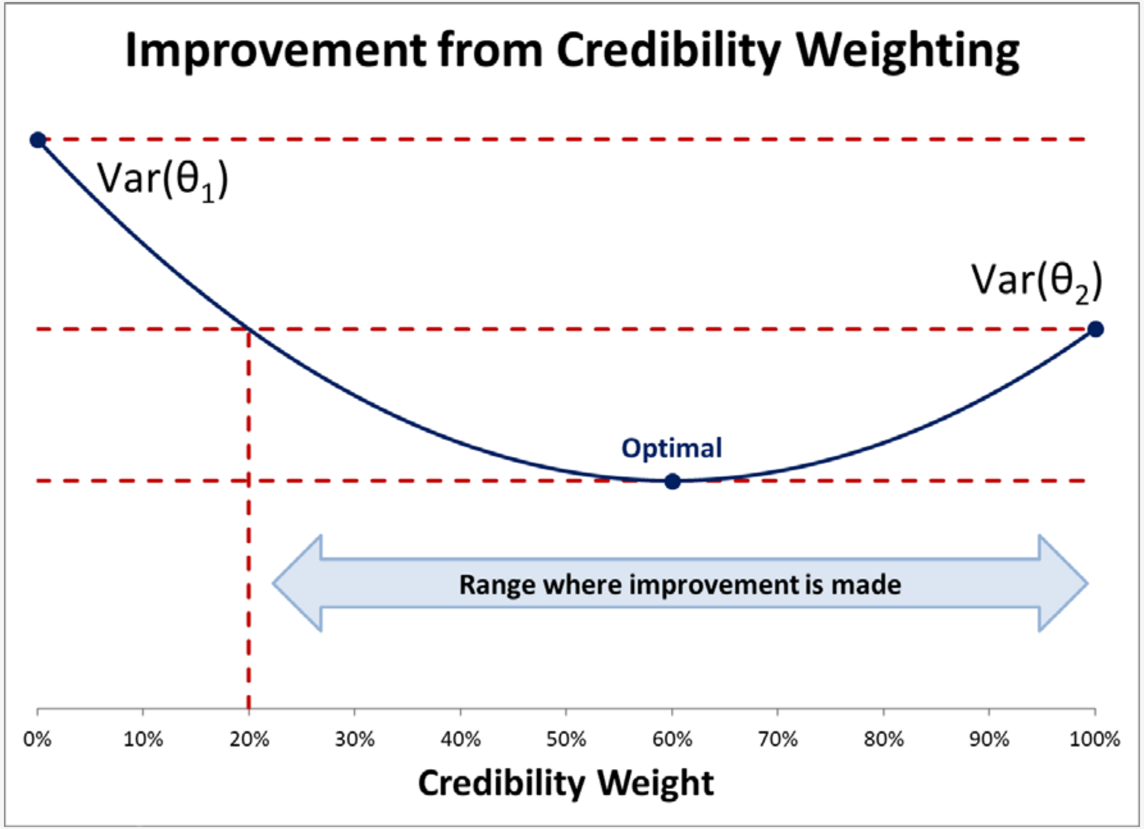
Fortunately, credibility values and shrinkage tuning parameters can be forgiving. A credibility formula with a “k” that is imprecise is generally still better than either extreme of no-pooling or complete pooling.

> “When estimating the Bayesian credibility parameter k, the estimate need not be extremely precise. For many practical applications, the estimate of k can be wrong by as much as a factor of two in either direction and still produce a fairly good estimate of the quantity, e.g., frequency, severity, pure premium, etc., that credibility is being used to estimate.”
>
> Mahler, “*An Actuarial Note on Credibility Parameters*”, PCAS 1986

A similar conclusion was found in the original 1970 paper by Hoerl and Kennard, “*Ridge Regression: Biased Estimation for Nonorthogonal Problems*”

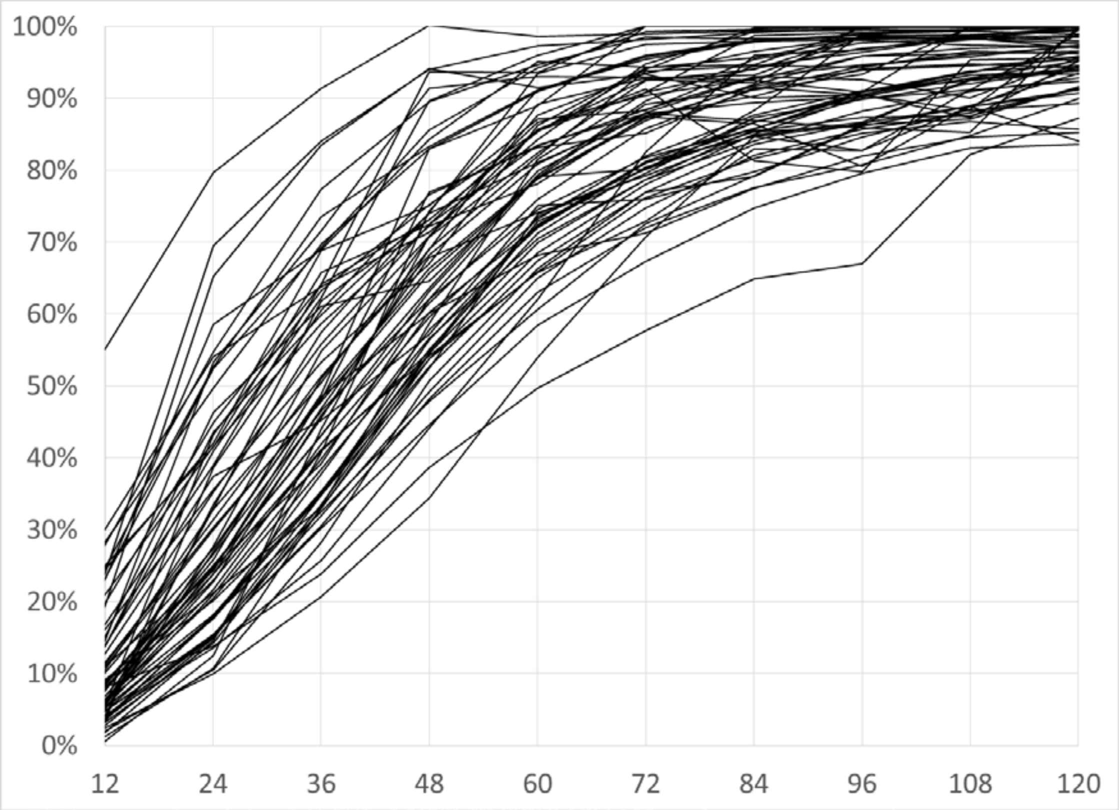
Blending Client Data with Benchmark Curves

Cross Validation



Even if we cannot estimate the optimal credibility perfectly, we can select a value that produces a blended estimate that is an improvement on either estimator alone.

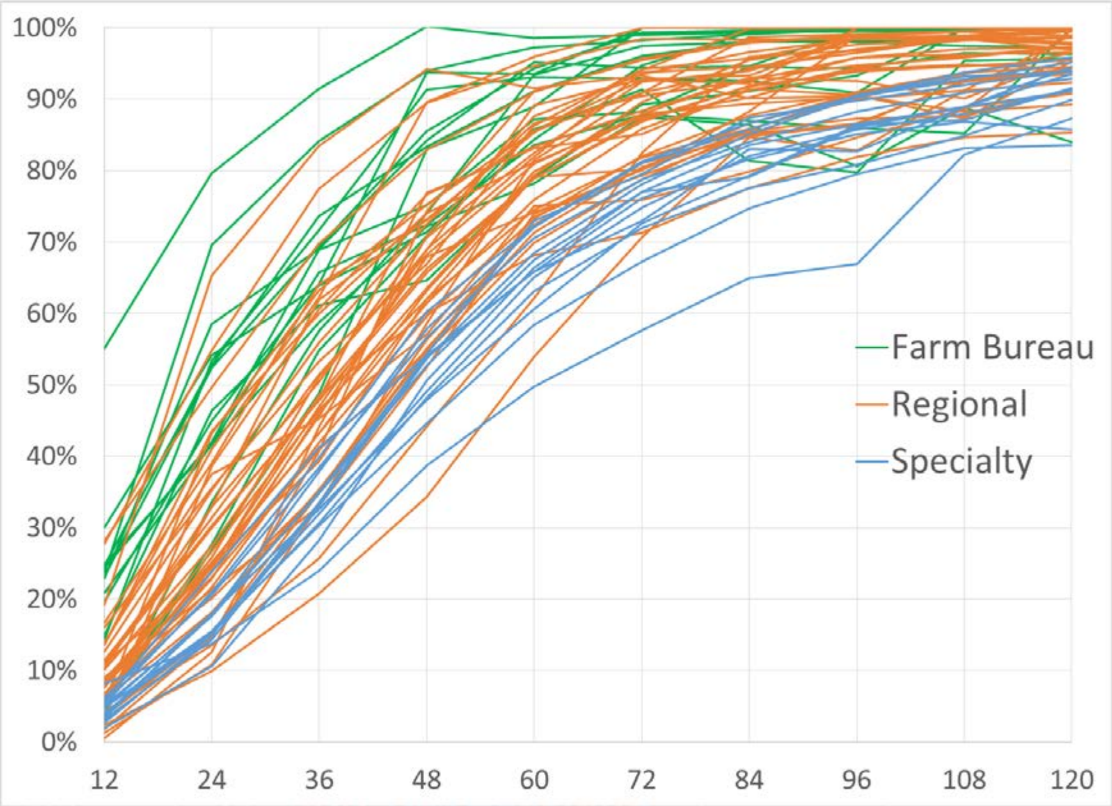
Blending Client Data with Benchmark Curves
Example



The spread of patterns by company is very wide.

This may imply that the credibility assigned to a single TPL benchmark pattern would be small.

Blending Client Data with Benchmark Curves
Example



Next goal is to better refine the benchmark into smaller cohorts of companies.

For this example, we can split the companies into three groups, and create a benchmark for each.

Research question: how best to create these cohorts?

Conclusion and Takeaways

- Fitting growth curves to individual client data is straightforward, but can be unreliable when the data is volatile or sparse
- Blending with benchmark patterns can stabilize the curve fit; and the mathematics is not difficult to derive
- More work is still needed on assembling the benchmark patterns

Thank you!

Selected References

Munich RE

Clark, D.R. "*LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach*" CAS Forum 2003.

https://www.casact.org/pubs/forum/03fforum/03ff041.pdf

[basis for first part of this presentation]

Craighead, D.H. "*Some aspects of the London reinsurance market in world-wide short-term business*" Journal of the Institute of Actuaries 1979.

[introduced the use of the Weibull curve for development data]

Pigeon, Mathieu, Katrien Antonio and Michel Denuit; "; *Individual Loss Reserving Using Paid-Incurred Data*" Insurance: Mathematics and Economics 58; 2014, 121-131.

Renshaw, A.E. and R. Verrall "*A Stochastic Model Underlying the Chain-Ladder Technique*" B.A.J. 1998

[good discussion showing relationship of chain-ladder method to GLM]

Sherman, R.E. "*Extrapolating, Smoothing and Interpolating Development Factors*" PCAS 1984.

https://www.casact.org/pubs/proceed/proceed84/84122.pdf

[introduced the use of inverse power (log-logistic) curve, though for link-ratios rather than for LDFs]