Strategies for Working with Loss Development Factors

Uri Korn

CLRS, September 11, 2017

Loss Development Factors

- LDFs seem simple, but dealing with them often isn't:
  - Selection with volatile data
  - Different groupings – trade off between finer segmentation vs. greater volatility
  - How to handle changing LDFs
  - Handling mixes of retentions and limits

- But are necessary to analyze insurance data

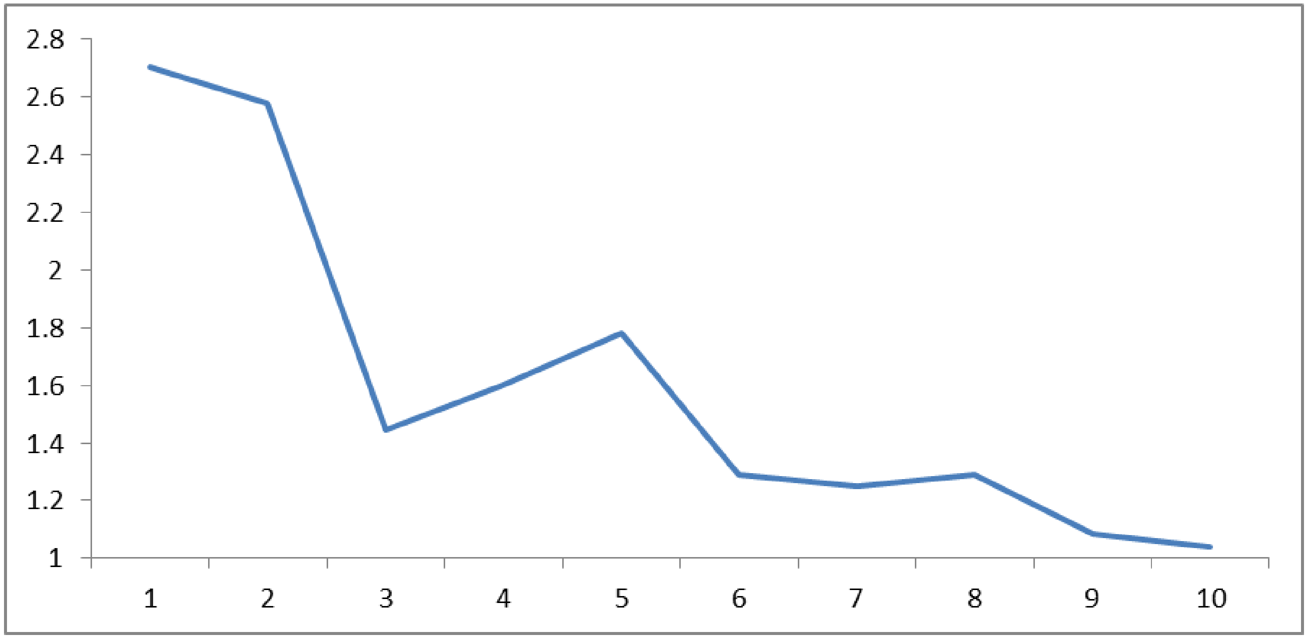
Outline

1. Dealing with volatility – Curve Fitting
2. Handling segmentations and credibility
3. Determining the best look-back period
4. Dealing with assorted limits and retentions

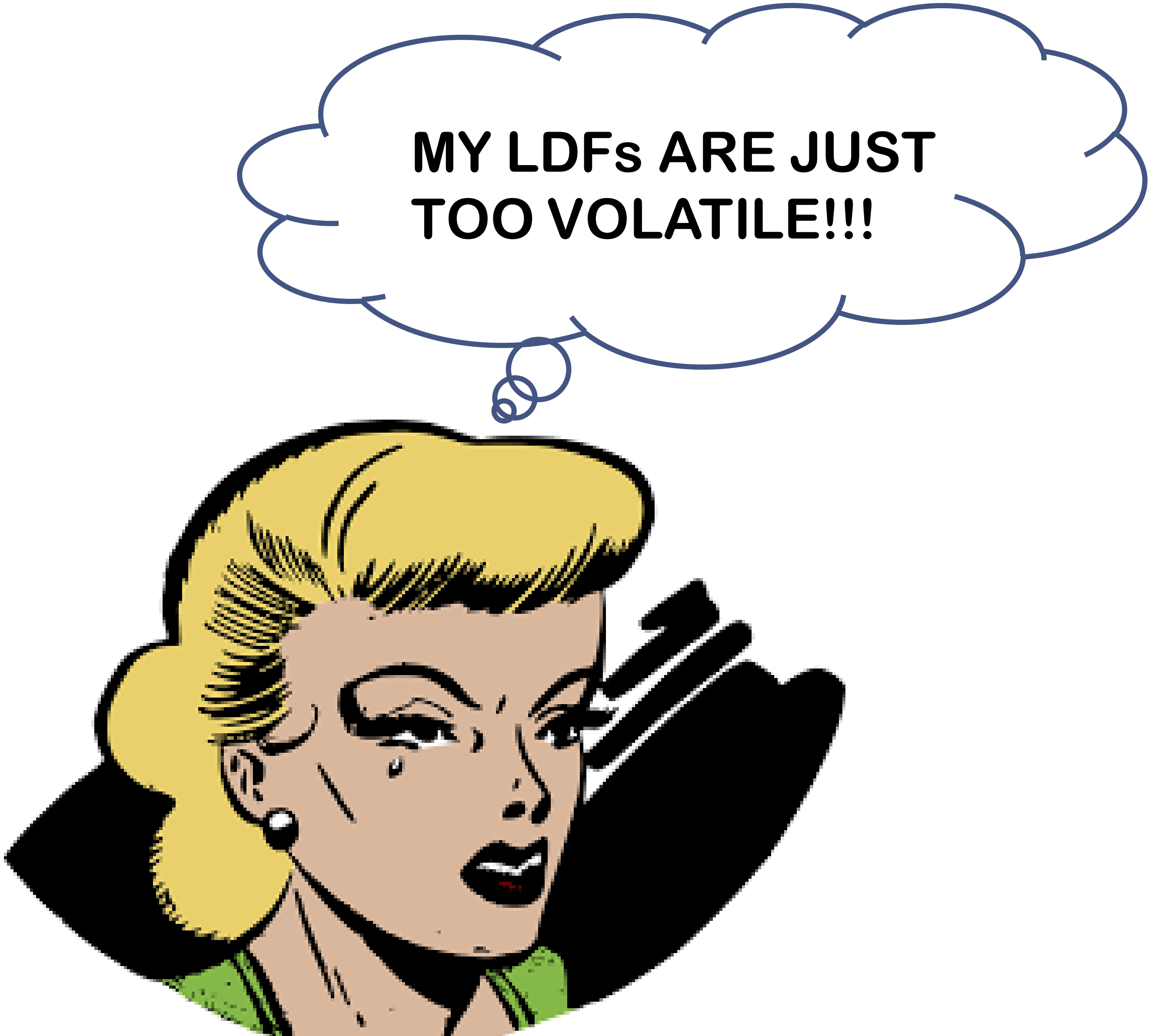
The LDF Ninja

Part 1) LDF Selection & Curve Fitting

LDF Selection



- Difficult to select LDFs when the data is volatile
- Subjective – repeating the task can result in different factors
- Can be time consuming
- Theoretical problem – over-parameterized (for those who care about that stuff…)


MY LDFs ARE JUST TOO VOLATILE!!!

Curves

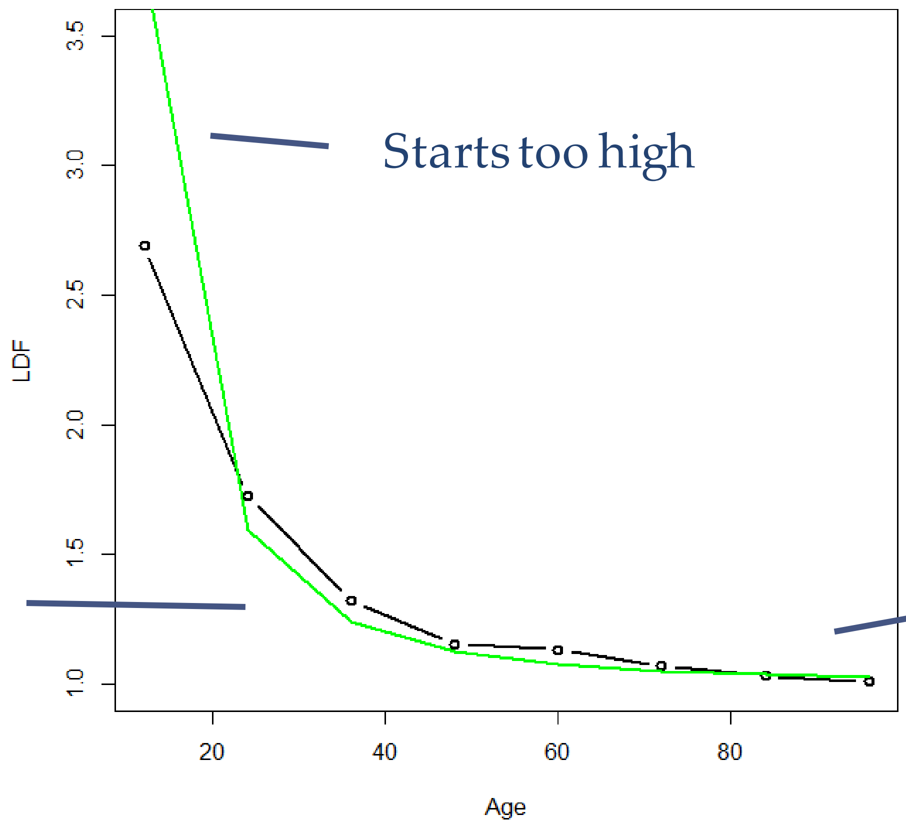
- Inverse Power Curve (IPOC) (Sherman 1984)

$$\log(LDF - 1) = A + B \times \log(age)^{*}$$

  - Nice simple curve that is easy to implement
  - But can often be a poor fit to the data

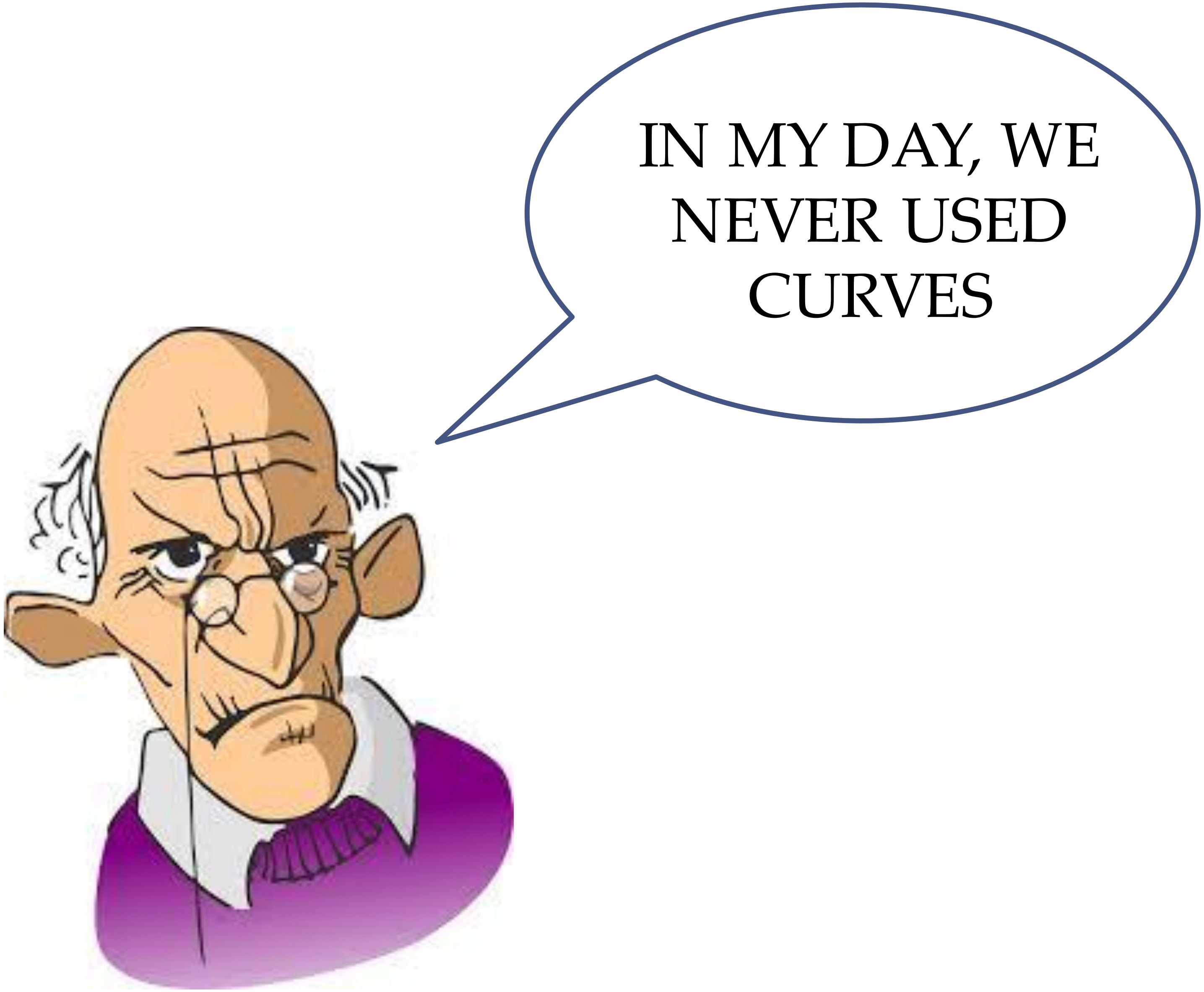
* Using *age* instead of *1 / age*, since the regression equations are equivalent. Also, *ignoring the c parameter and setting it to 0*

IPOC Fit



Trouble making the "turn"

Tail too high

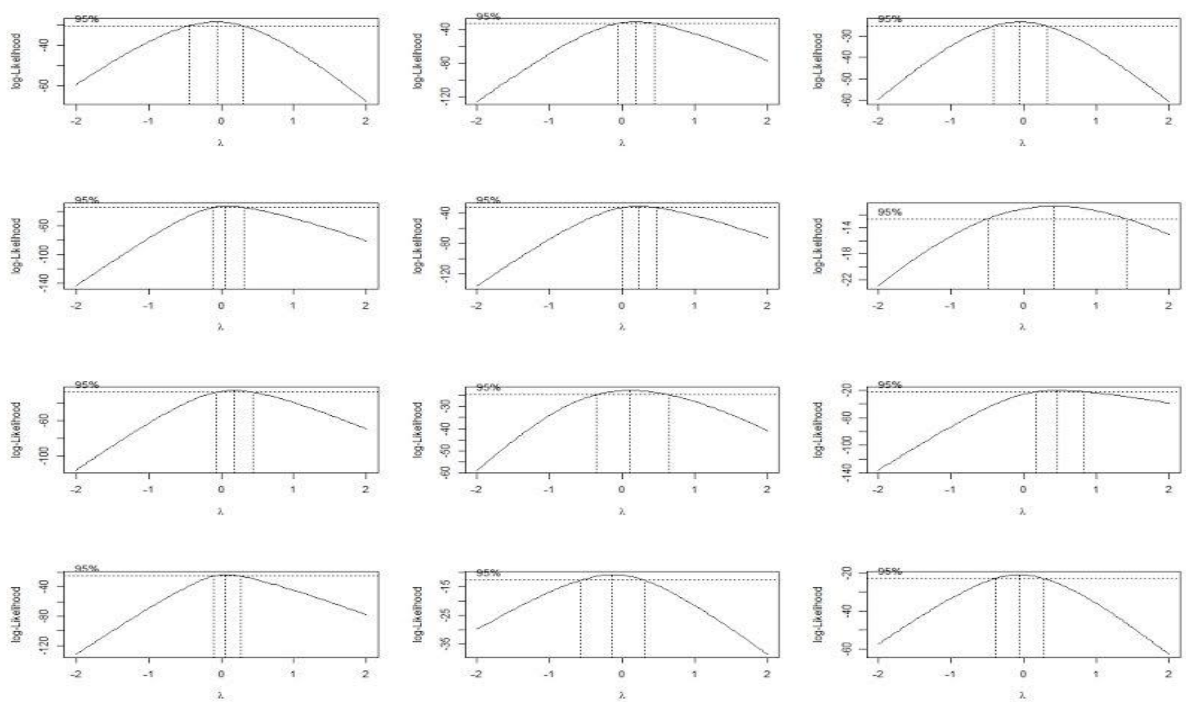

IN MY DAY, WE NEVER USED CURVES

DIPOC

- Modify the IPOC to be more appropriate for LDFs
- The DIPOC: Double [GLM] Inverse Power Curve:

  1) Use a Gamma GLM with a log-link instead of regular regression

    - Box Cox tests on actual LDFs:



DIPOC

2) The variance is not constant

- The variance should vary with the loss volume
  - Keeping with the Gamma family assumption, make the Coefficient of Variation (CoV) inversely proportional to the loss volume (used in the LDF denominator)

- The variance should vary with the age – the later LDFs are much more volatile than the earlier ones

→ Have a separate equation for the CoV

- The LDFs will be more volatile at the very early ages due to lower loss volume, and much more volatile at the later ages due to the variance equation

DIPOC

$$\text{Fitted LDF}_i = \exp(\, A + B \times \log(\text{age}_i)\,) + 1$$

$$\text{CoV Factor}_i = \exp(\, I + J \times \text{age}_i \,)$$

$$\text{CoV}_i = \text{CoV Factor}_i \,/\, \sqrt{\text{Loss}_i}$$

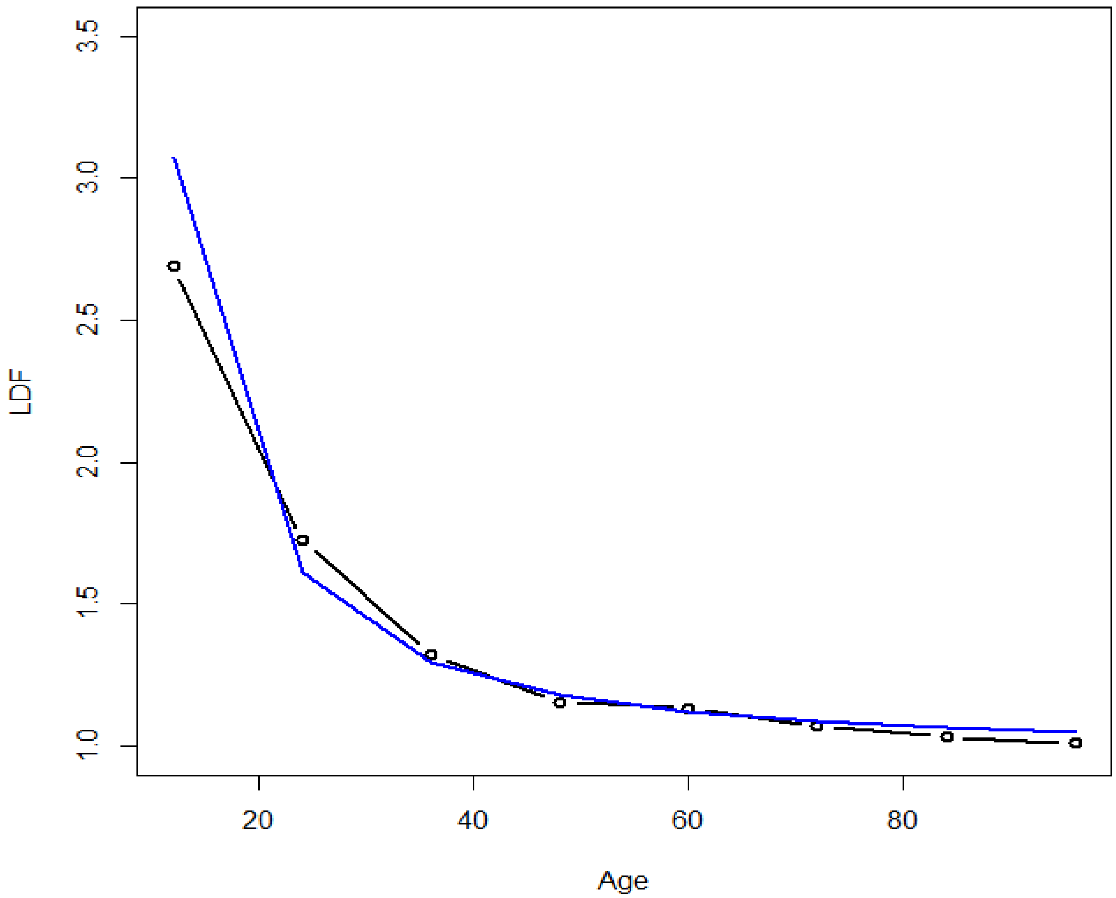
$$\alpha = 1 \,/\, \text{CoV}_i^2$$

$$\beta = \alpha \,/\, (\,\text{Fitted LDF}_i - 1\,)$$

$$\text{loglik} = \sum \log(\,\text{Gamma PDF}(\,\text{Actual LDF}_i - 1 \,,\, \alpha \,,\, \beta\,)\,)$$

- 4 Parameters to maximize: A, B, I, J
- Can be done in Excel via Solver

DIPOC Fit



Much better, although still not perfect…

DIPOC

- Handling negative development, i.e. LDFs less than 1
  - Is less of a problem if fitting on the average LDFs
  - Take out negatives, fit curve, and adjust for bias (by fitting another curve, for example)
  - Instead of subtracting 1 from each LDF, subtract a lower number, e.g. 0.9 (the base), and adjust the CoV accordingly. (Although results are mixed with this method)

$$Adjusted\ CoV\ Factor = \frac{CoV\ Factor \times (Fitted\ LDF - 1)}{Fitted\ LDF - Base}$$

Add Smoothing

- Extra smoothing can be added to the DIPOC via a Generalized Additive Model (GAM)/Smoothing Splines
  - (This concept is borrowed from England & Verall 2001)

- Call this curve the SMIPOC (Smoothed [Double GLM] Inverse Power Curve)

Additive Models

- Linear Models:

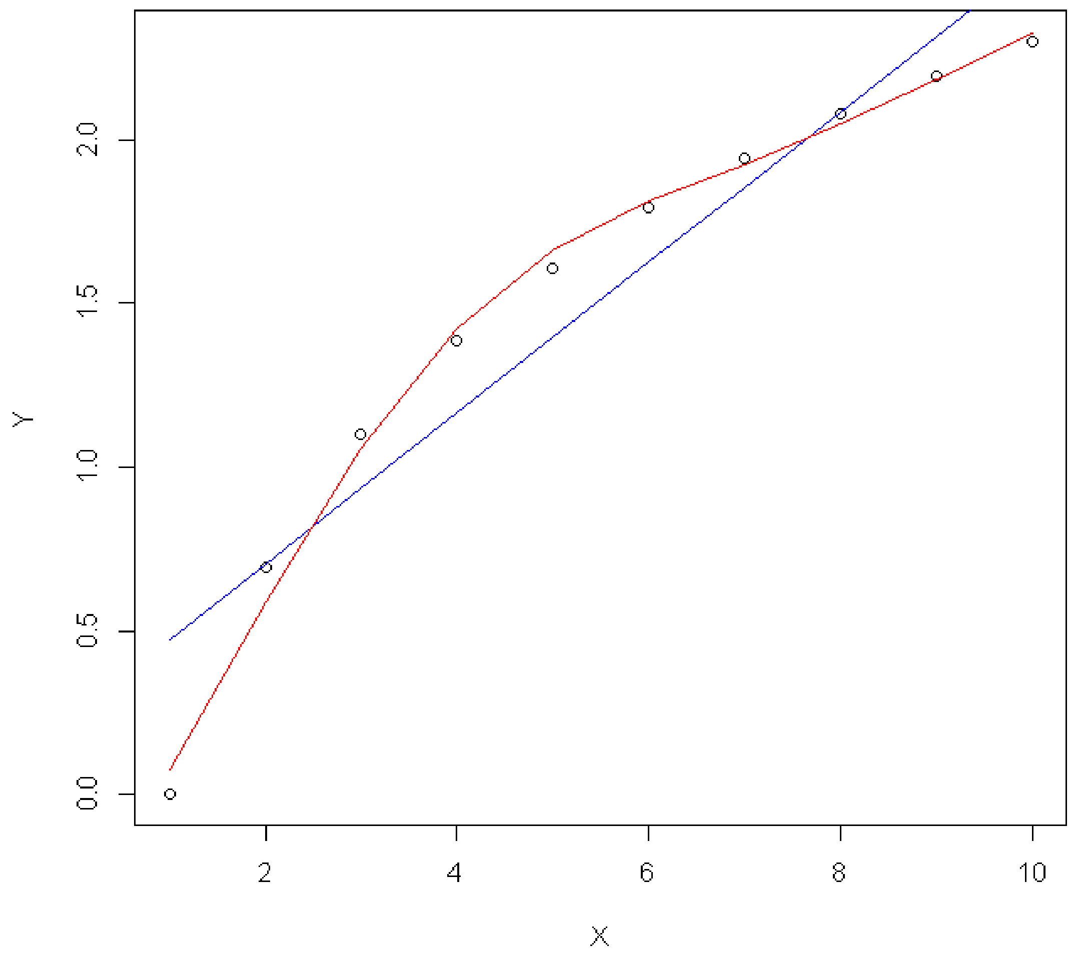
$$Y_i = \sum_j B_j X_{ij}$$

- Additive Models:

$$Y_i = \sum_j f(X_{ij})$$

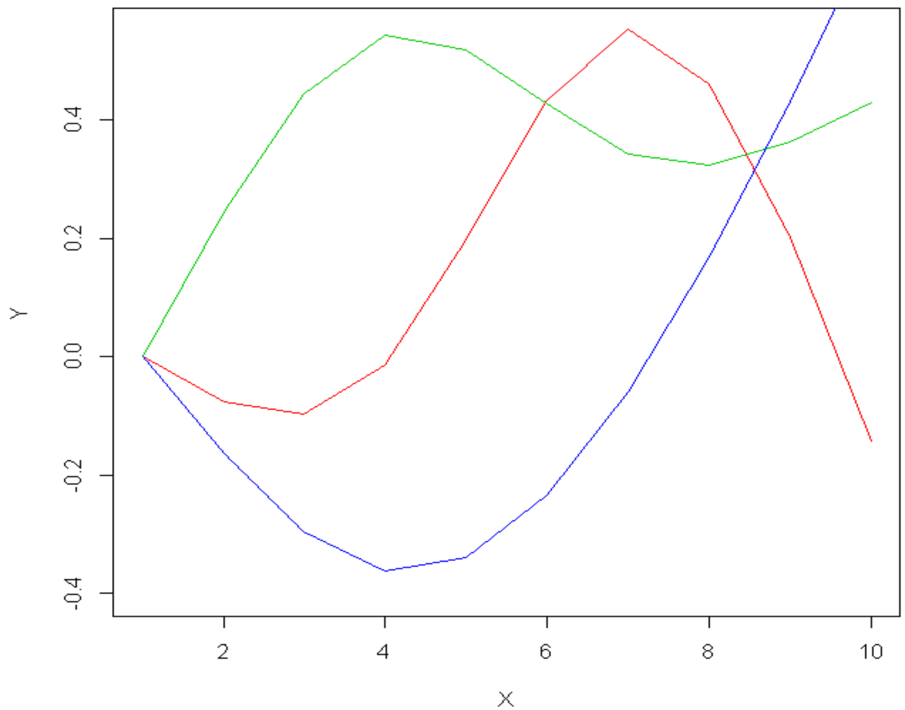
  - A common function choice is to use cubic splines to smooth the data

Additive Model Example



How do Additive Models work?

- Performs a transformation on one or more of the independent variables via a cubic spline function
- In the previous example, the following 3 numeric sequences were created and used instead of the numbers from 1 to 10
- Once these new sequences are generated, a linear model can be used



Additive Models

- Good at adapting to the data, especially when a parametric form is hard to find – as in the case of LDFs
- A downside is that they can sometimes over-fit the data
  - Significance tests and eyeing the data can help avoid this

Additive Models in Excel

- One way to implement this in Excel is to generate the spline variables in R and paste them in to Excel

- To generate a spline sequence in R with 2 degrees of freedom (i.e. 2 variables) starting at the 2nd age and ending at the 20th, but having a tail that extends to the 40th:

```
library(splines)
ns( log(2:40),
  Boundary.knots=c(log(2),log(20)), df=2 )
```

SMIPOC

- DIPOC with additional smoothing

$$log(LDF - 1) = A + B \times s(\, log(t)\,)$$

Where *s* is a cubic spline function
Note that the spline is called on the logarithm of age

Can also be written as, where t(i) is each new generated spline sequence on the log of age:

$$log(LDF - 1) = A + Bt^{(1)} + Ct^{(2)}$$

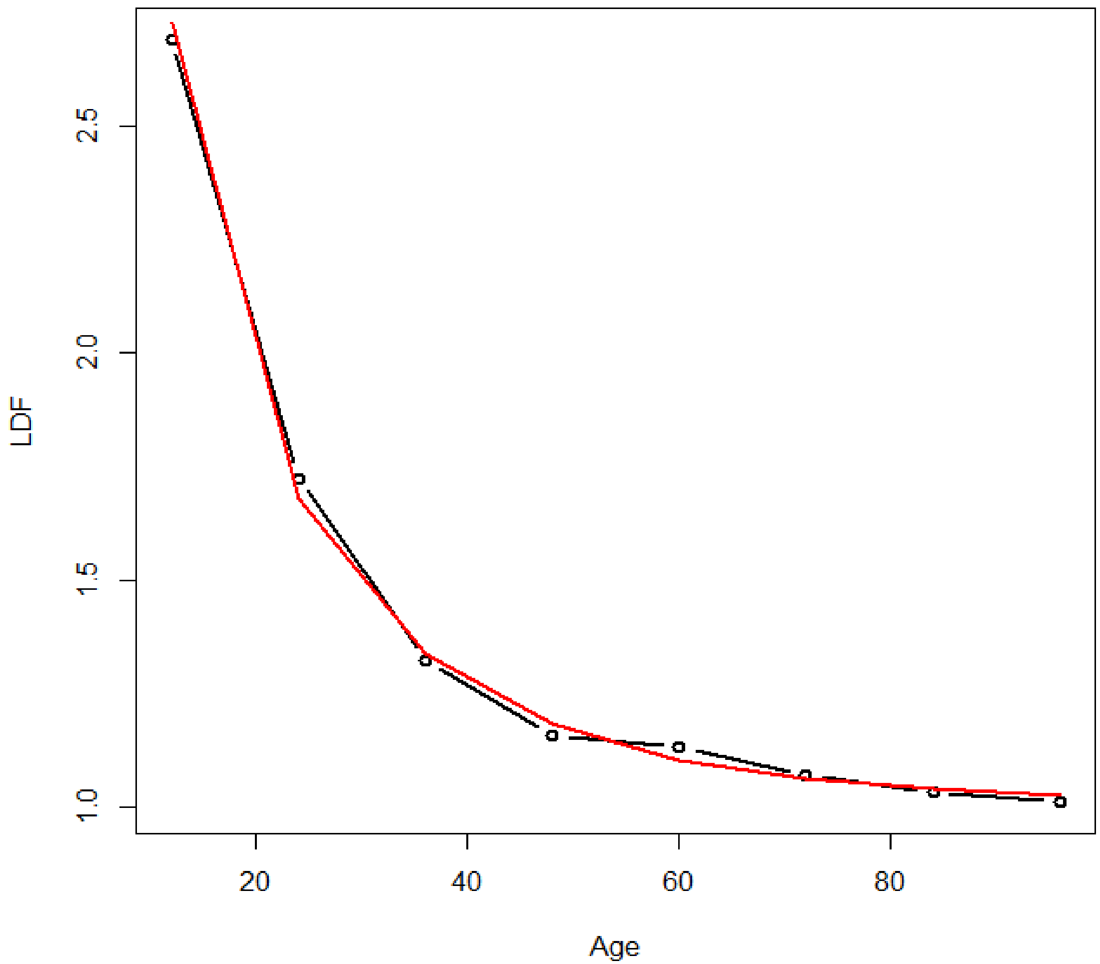
This curve adds an additional parameter to the DIPOC. (More can be added if needed.)

SMIPOC

- A likelihood ratio test can be performed to see if the additional parameter is significant and preferred over the DIPOC

1 – Chisq CDF(2 x Difference in Log-likelihoods, df=1)

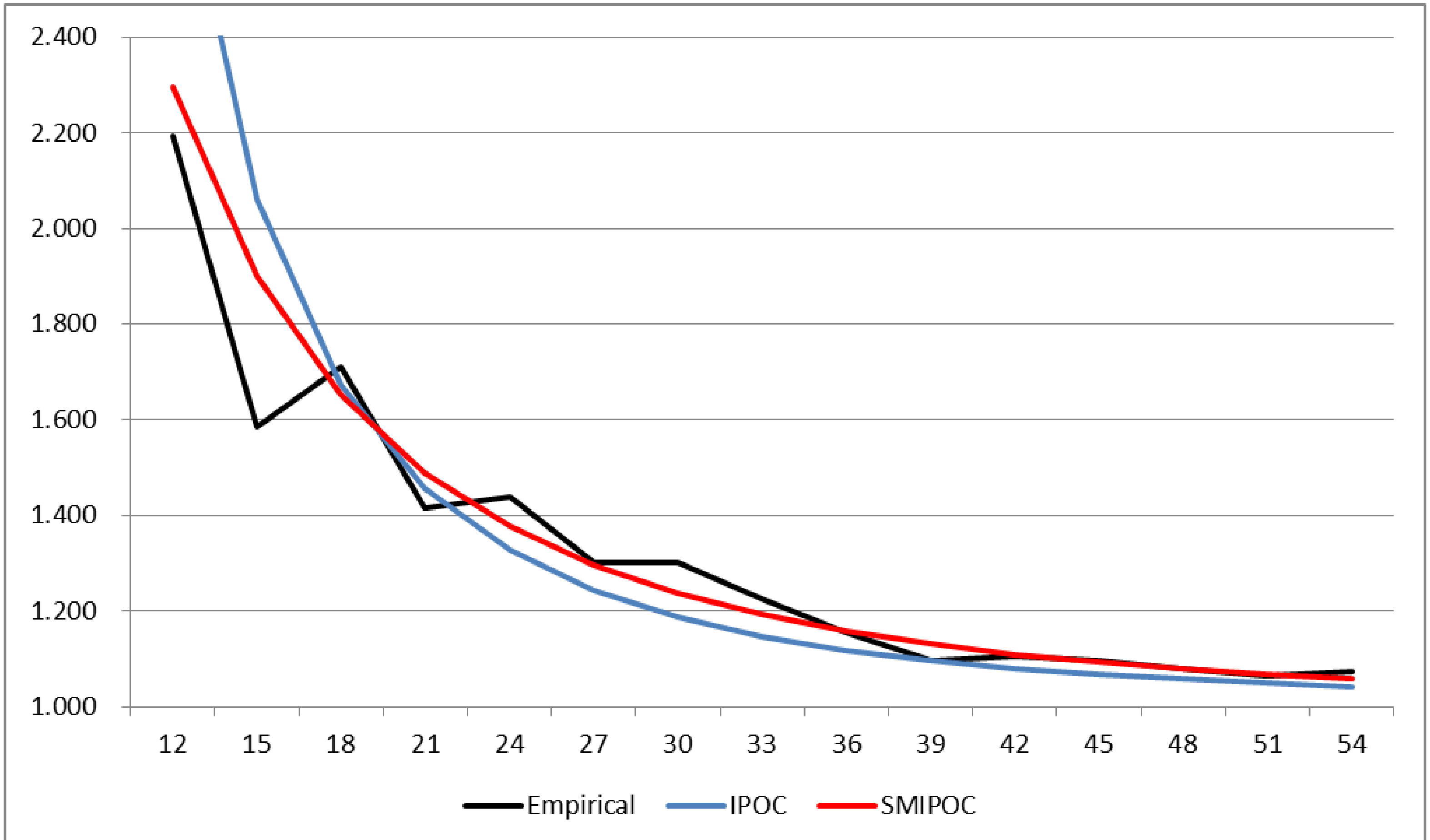
SMIPOC Fit



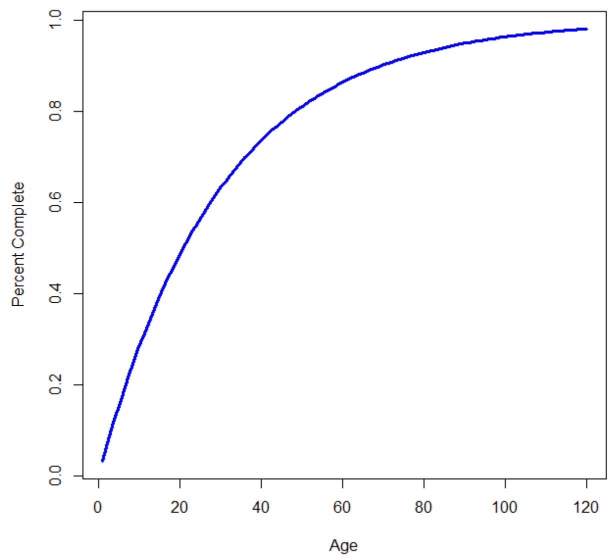
Much better!

Real Data Example

(That has been modified)



A Distribution Approach



- Clark (2003) shows how distributions can be fit to the percentage completion data as an alternative to using curves
- To convert: LDF(age) = 1 / CDF(age)
- The percentages are calculated by dividing by the expected ultimates using either the Chain Ladder or Bornheutter-Ferguson estimates

A Distribution Approach

- Here, we'll be using Clark's distribution approach to model loss development, but with a different method for fitting the distribution

- (An advantage of this approach is that it allows breaking up triangle segments into finer groups than can be done with the IPOCs, since the latter model on the actual LDFs)

Right Truncation

- This approach is borrowed from a method used to fit the arrival times of reported claims (Korn 2015)
    - Recall left truncation: Used to model policy retentions, since the number and sizes of the claims below the retention are not known
    - Reported claims on the other hand are right truncated
        - The number and timing of future claims are also not known
    - Instead of fitting individual claims, this same approach can be used to fit the arrival times of each individual reported or paid dollar in the triangle

Right Truncation (Cont.)

- Recall the likelihood for left truncation:
  - Divide by s(x)
- Similarly, for right truncation:
  - Divide by CDF(x)

Right Truncation

- The log-likelihood for the losses in yellow would be:

	Age			
AY	**12**	**24**	**36**	**48**
2013	100	140	150	155
2014	110	154	165	
2015	121	169		
2016	133			

(154 - 110) x

log[(CDF(24) – CDF(12)) / CDF(36)]

Right Truncation

- If only the latest 2 diagonals are used:

	Age			
AY	**12**	**24**	**36**	**48**
2013	100	140	150	155
2014	110	154	165	
2015	121	169		
2016	133			

(154 - 110) x
log[(CDF(24) – CDF(12)) / (CDF(36) – CDF(12)]

(Now the data is both right and left truncated)

RIPOD

- Notice how only the CDFs/survival functions are needed to calculate the likelihood – this gives flexibility in defining the curve

- RIPOD: Right Truncated Inverse Power Distribution

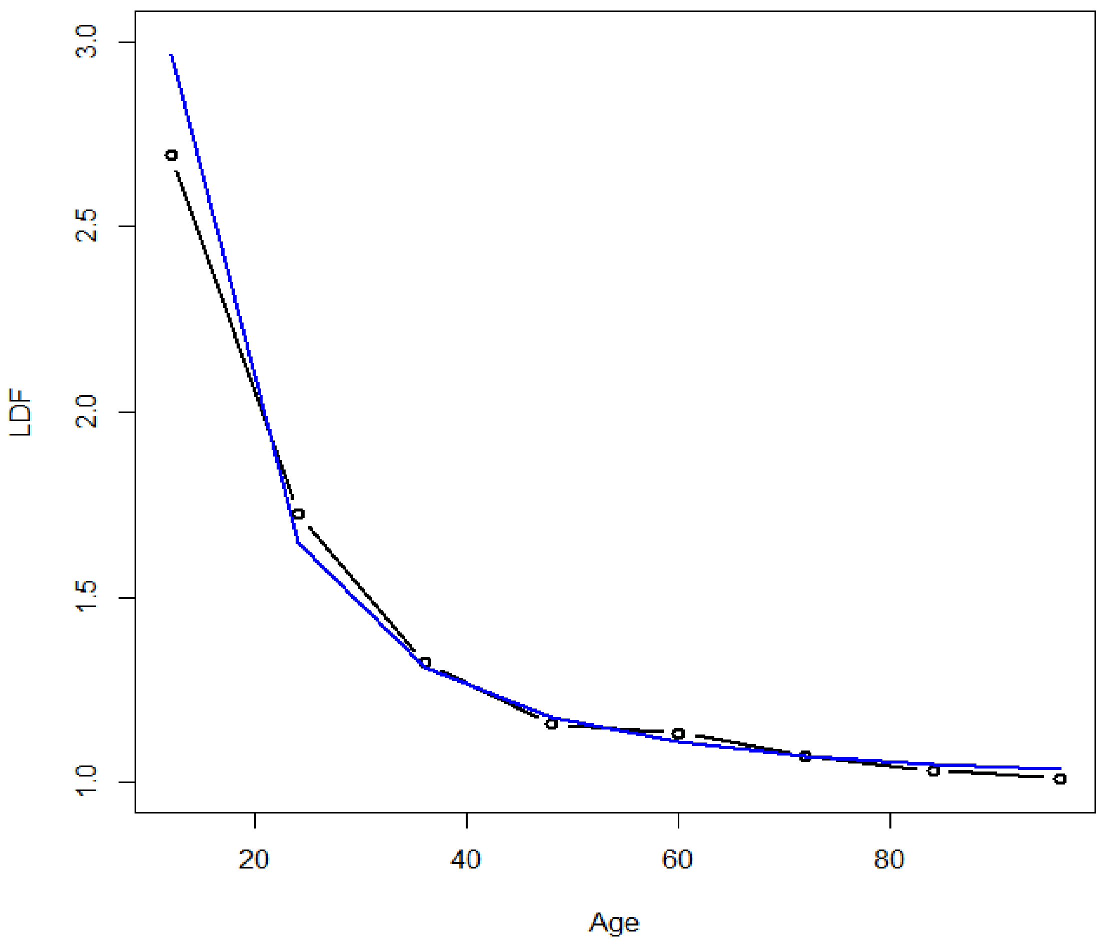
$$s(age) = ilogit(A + B \times log(age))$$

- Using this result, calculate the total log-likelihood (using the formula shown before) weighted by the incremental losses

RIPOD (Cont.)

- This is equivalent to the log-logistic distribution that Clark mentions
- Will work even if some of the incremental losses are negative (but not if the expectation is)

RIPOD Fit



Not bad, except that the fitted tail is too high
(This behavior of the log-logistic curve is noted by Clark as well)

SMIPOD

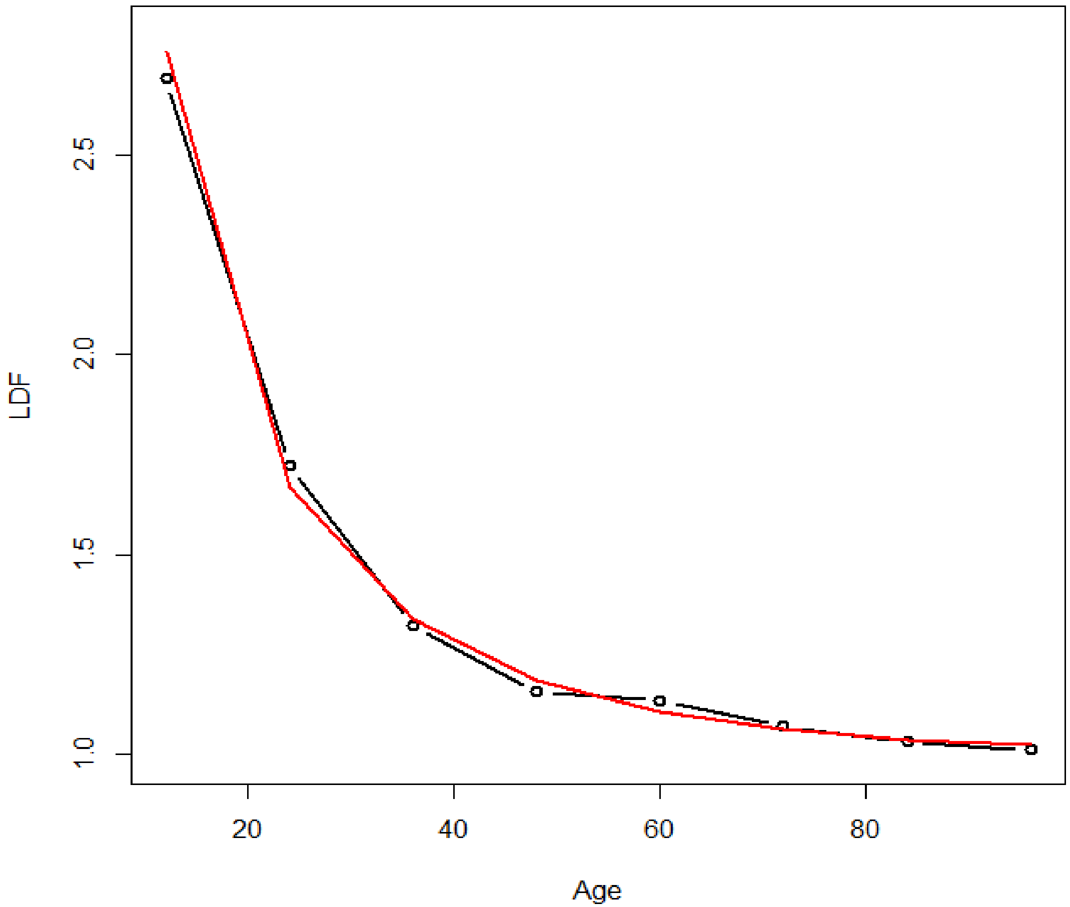
- SMIPOD: Smoothed [Right Truncated] Inverse Power Distribution
- RIPOD with smoothing, similar to the SMIPOC
- Usually produces a lower tail estimate than the RIPOD

SMIPOD

- Fitted maximum likelihood parameter values are harder to find than the RIPOD

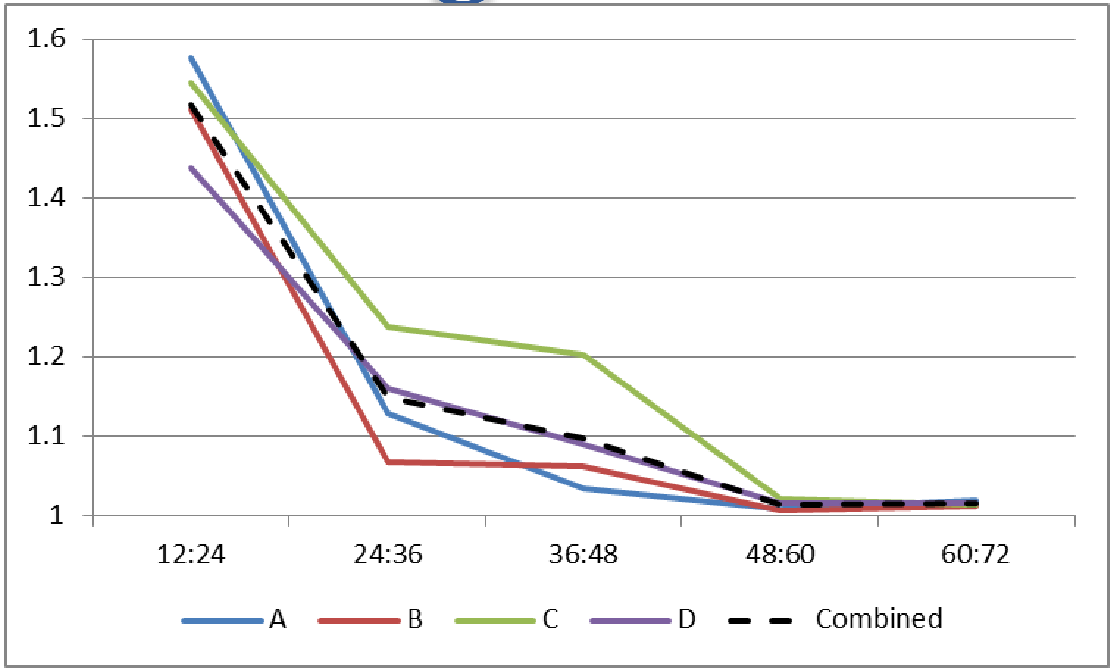
- Suggested fitting procedure:
  - First fit parameters that minimize the absolute/squared difference between this and the RIPOD LDFs
  - Use these parameters as the starting values for the SMIPOD maximum likelihood routine

SMIPOD Fit

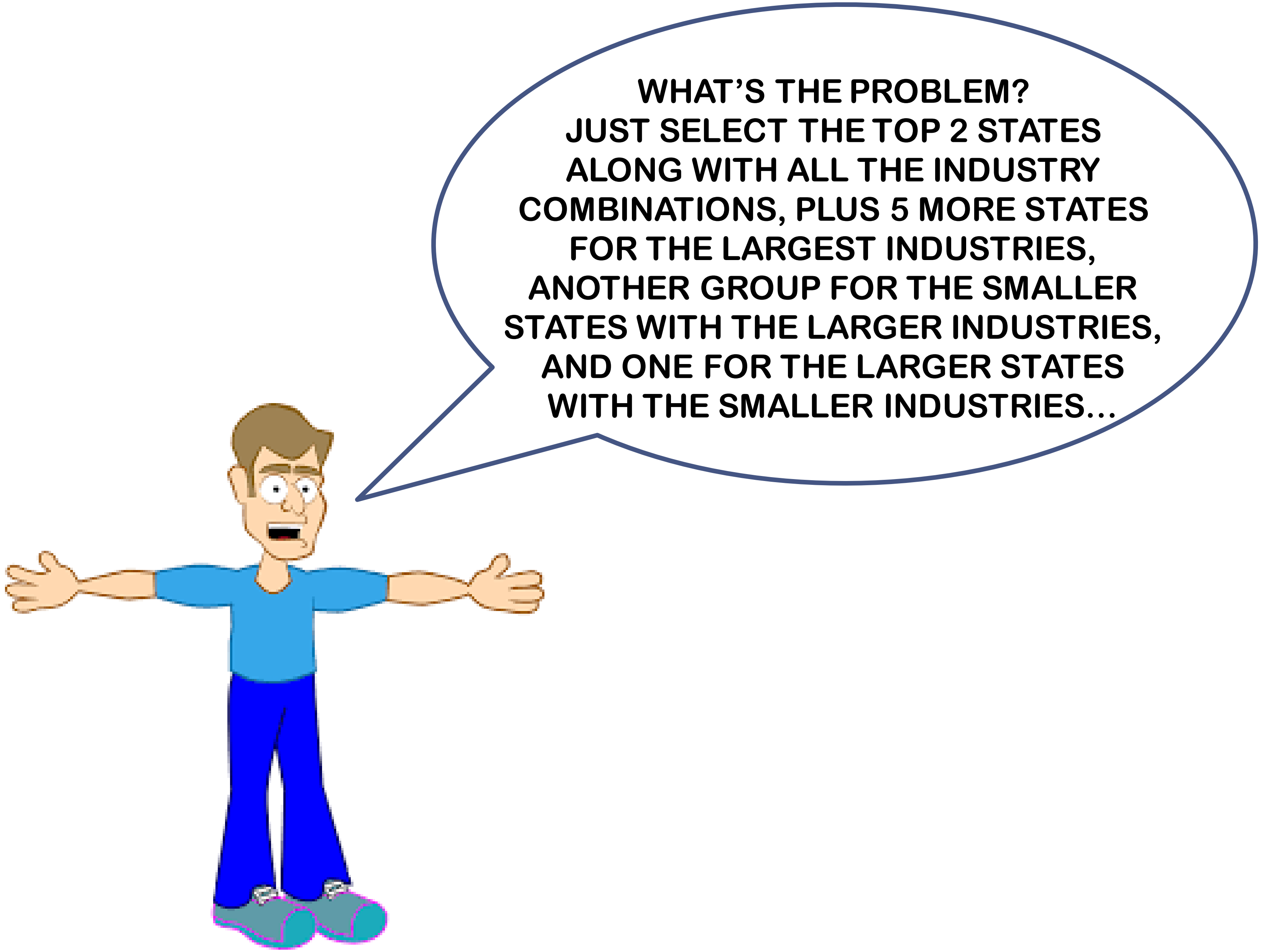


Part 2) Segmentations & Credibility

Data Segmentation



- Common scenarios:
  - Heterogeneous segments too small/volatile to analyze on their own
  - Similar "credible" segments that may be able to benefit from sharing information
  - Different dimensions in the data where it would be almost impossible to separate out each combination, e.g. industry, state, account size
- The most robust and straightforward solution is to use credibility

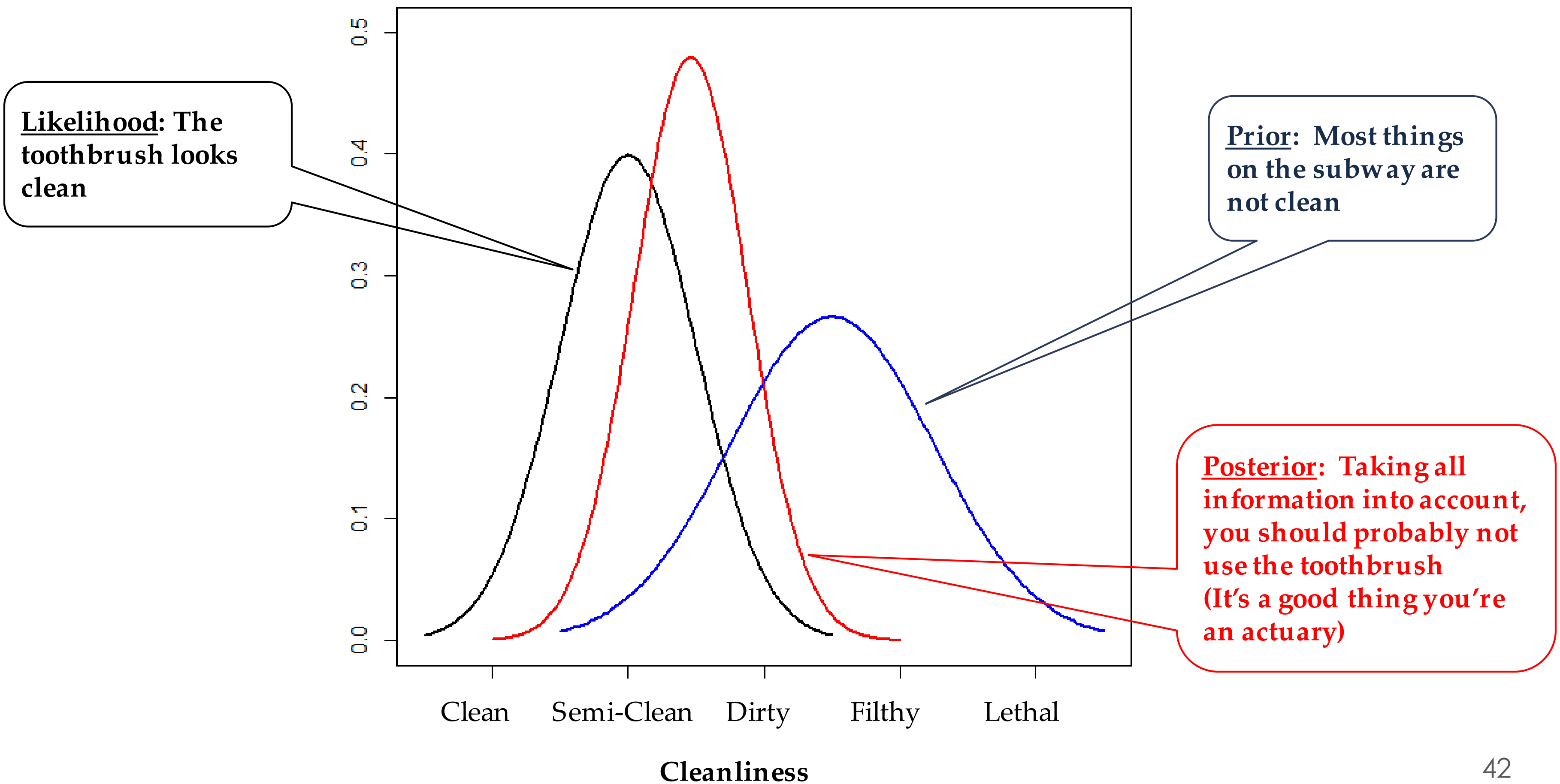

WHAT'S THE PROBLEM? JUST SELECT THE TOP 2 STATES ALONG WITH ALL THE INDUSTRY COMBINATIONS, PLUS 5 MORE STATES FOR THE LARGEST INDUSTRIES, ANOTHER GROUP FOR THE SMALLER STATES WITH THE LARGER INDUSTRIES, AND ONE FOR THE LARGER STATES WITH THE SMALLER INDUSTRIES…

Bayesian Credibility

f(Params | Data)	~	f(Data | Params)	x	f(Params)
Posterior	~	Likelihood	x	Prior

Bayesian Credibility Example

- You find a toothbrush on the subway!
- It looks semi-clean!
- Should you use it?



Bayesian Credibility for a Regression Curve/Distribution

- Performs credibility weighting on the parameters simultaneously while fitting the curve/distribution
- This is done by adding another component to the log-likelihood which pushes each parameter closer to the mean

Likelihood

$$\sum PDF(x, p1, p2) +$$

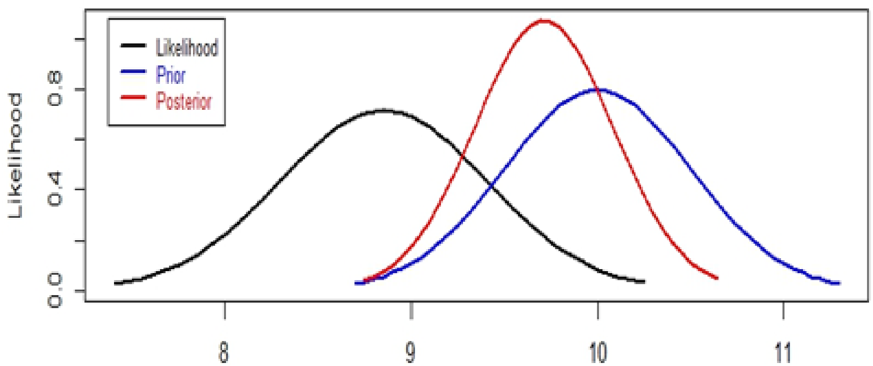
$$Norm(p1, Porfolio\ p1, Between\ Var1) + Norm(p2, Porfolio\ p2, Between\ Var2)$$

Prior

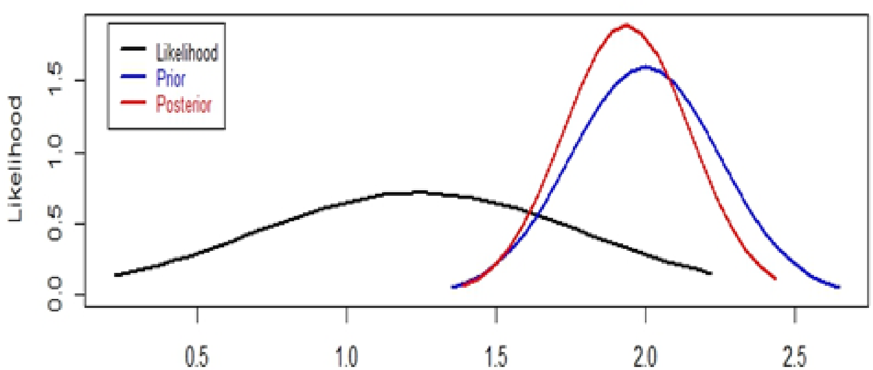
Where:
PDF is the <u>logarithm</u> of the probability density function
Norm is the <u>logarithm</u> of the normal probability density function

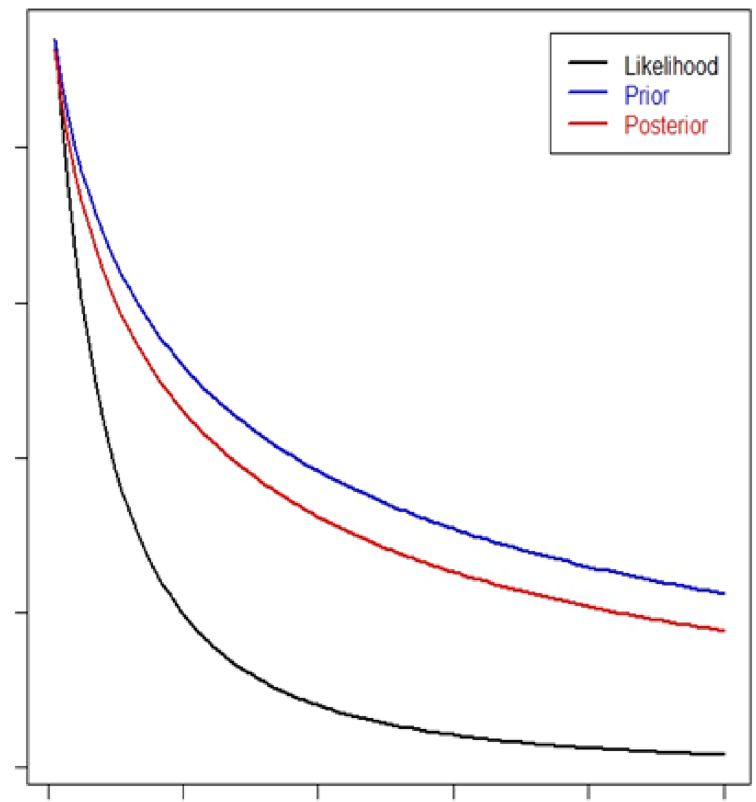
Bayesian Credibility on a Curve or Distribution

A



B

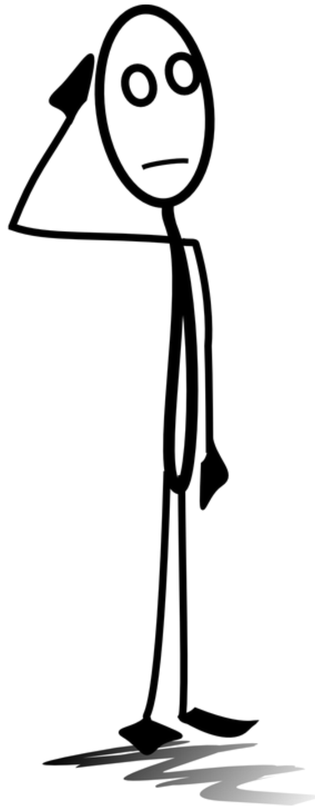




Bayesian Credibility

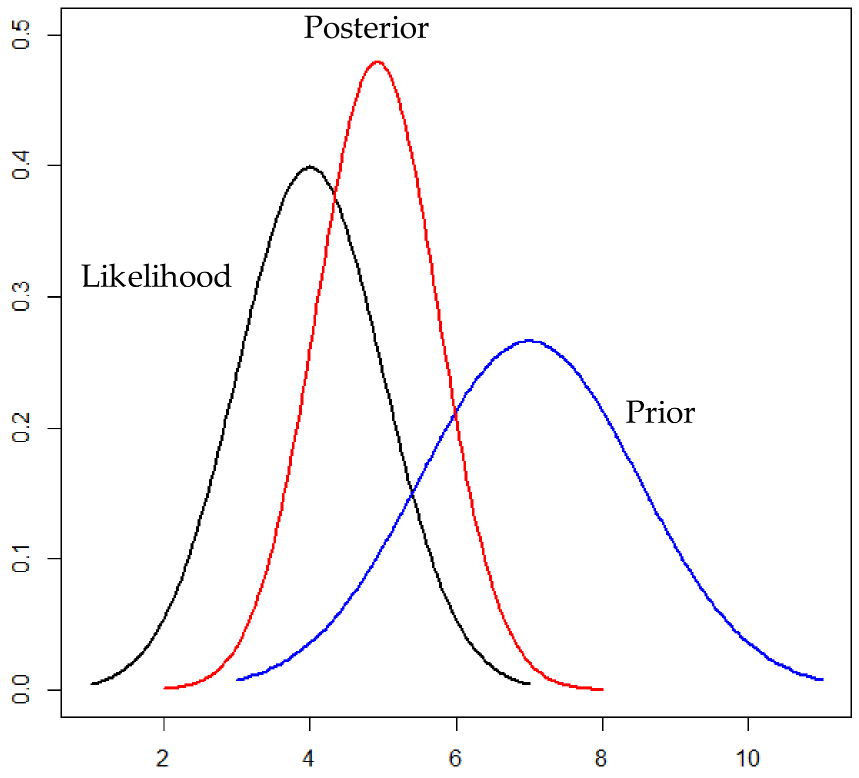
But this requires specialized software to run??

Bayesian Credibility



Or does it??

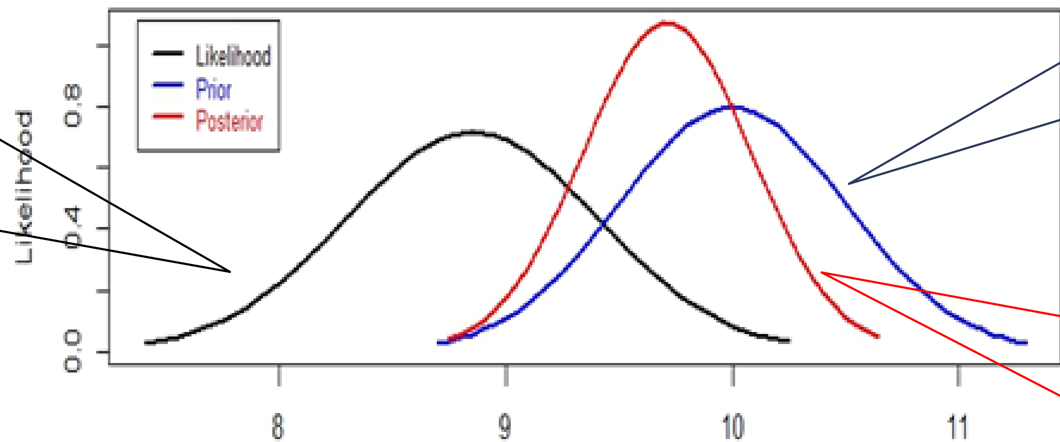
Normal Conjugate



- If both the likelihood and the prior are normally distributed, the posterior is normally distributed as well

Implementing Bayesian Credibility via MLE

MLE parameters are approximately normally distributed (asymptotically)

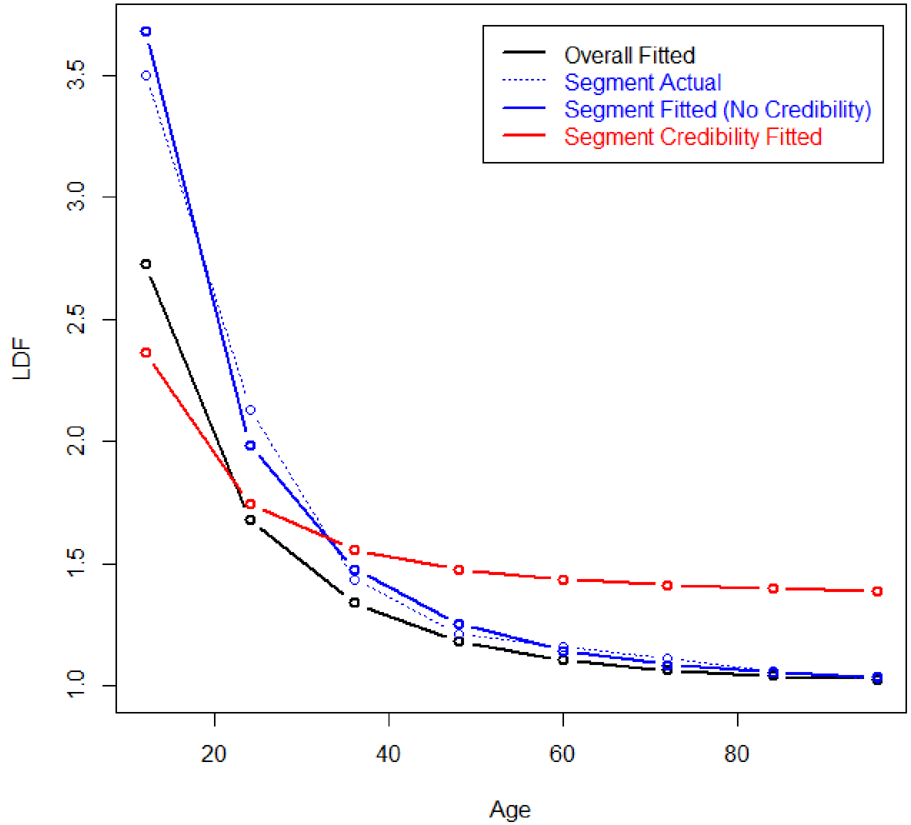


Normal prior on the parameters (the common assumption)

This is a conjugate prior and the posterior is normally distributed as well

- Since the result is normal, the mode equals the mean
- MLE, which returns the mode, also returns the mean in this case
- The result will match that returned from using specialized Bayesian software!

Credibility – Naïve Approach



- Applying Bayesian credibility and credibility weighting the parameters directly will often give funny results (especially for spline variables)
- Results often do not lie in between the segment and the overall mean, which is non-intuitive

Solution: Reparameterize the Curve

- For the IPOC, take any 2 different LDFs of the curve at 2 different ages:

$$B = \frac{log(\frac{LDF_1 - 1}{LDF_2 - 1})}{log(\frac{t_1}{t_2})} \qquad A = log(LDF_1 - 1) - B\ log(t_1)$$

- Given these 2 LDFs, A & B can be solved for
- Any LDF of the curve can now be determined
- Since this is the case, it is possible to reparameterize the curve and consider these 2 LDFs as the curve parameters

Inversion Formulas for 3 Parameter SMIPOC

$$log(LDF - 1) = A + Bt^{(1)} + Ct^{(2)}$$

$$X = t_1^{(1)} - t_2^{(1)}$$

$$Y = t_1^{(2)} - t_2^{(2)}$$

$$W = t_2^{(1)} - t_3^{(1)}$$

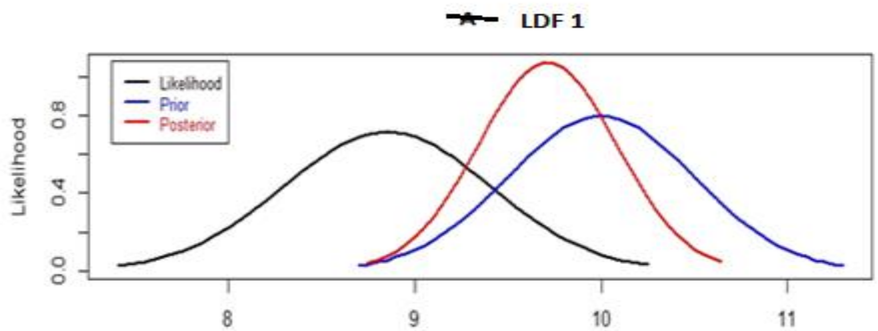
$$Z = t_2^{(2)} - t_3^{(2)}$$

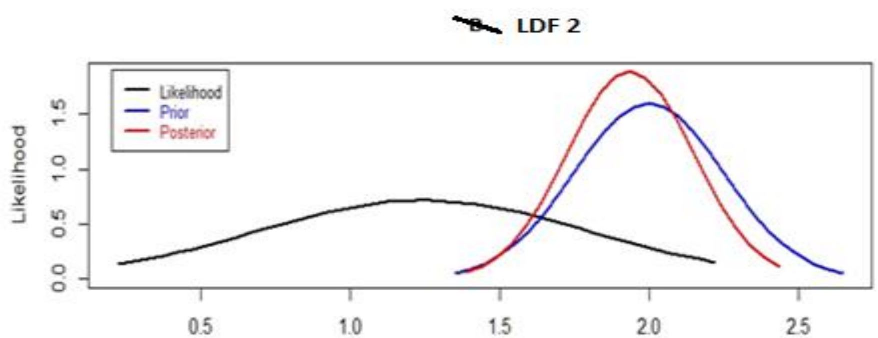
$$C = \frac{\frac{X}{W}\, log(\frac{LDF_2}{LDF_3}) - log(\frac{LDF_1}{LDF_2})}{\frac{XZ}{W} - Y}$$

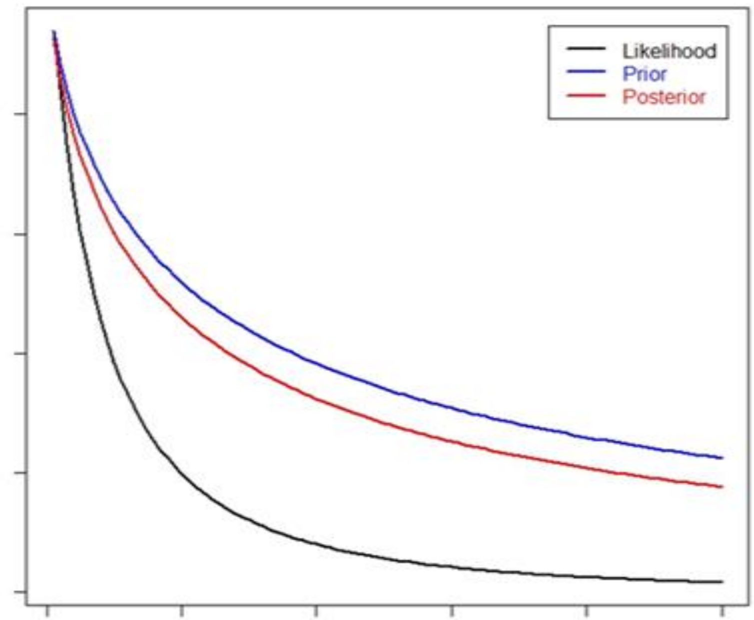
$$B = \frac{\frac{Y}{Z}\, log(\frac{LDF_2}{LDF_3}) - log(\frac{LDF_1}{LDF_2})}{\frac{YW}{Z} - X}$$

$$A = log(LDF_1) - Bt_1^{(1)} - Ct_1^{(2)}$$

Bayesian Credibility on a Curve or Distribution







- Instead of calculating the prior component (the penalty from deviating from the mean) on the original parameters, calculate the prior on the LDF parameters

- Since the 2 LDF parameters can be calculated once A & B are known, it is not actually necessary to invert the curve; the results will be equivalent

Solution: Reparameterize the Curve

- Suggested to initially choose ages evenly distributed along the curve and tweak as needed

- Or – calculate the prior on all of the LDFs taken to the power of 2/n or 3/n (not really statistically correct, but can be a good practical tool)

Credibility

- The total log-likelihood for the DIPOC/SMIPOC is equal to:

$$\sum \log(\text{Gamma PDF}(\text{Actual LDF}_i - 1, \alpha, \beta)) +$$

$$\sum_{c\,=\,Ages\ Used\ For\ Credibility\ Weighting} N(\,Fitted\ LDF_c, Complement\ LDF_c, Between\ Variance_c\,)$$

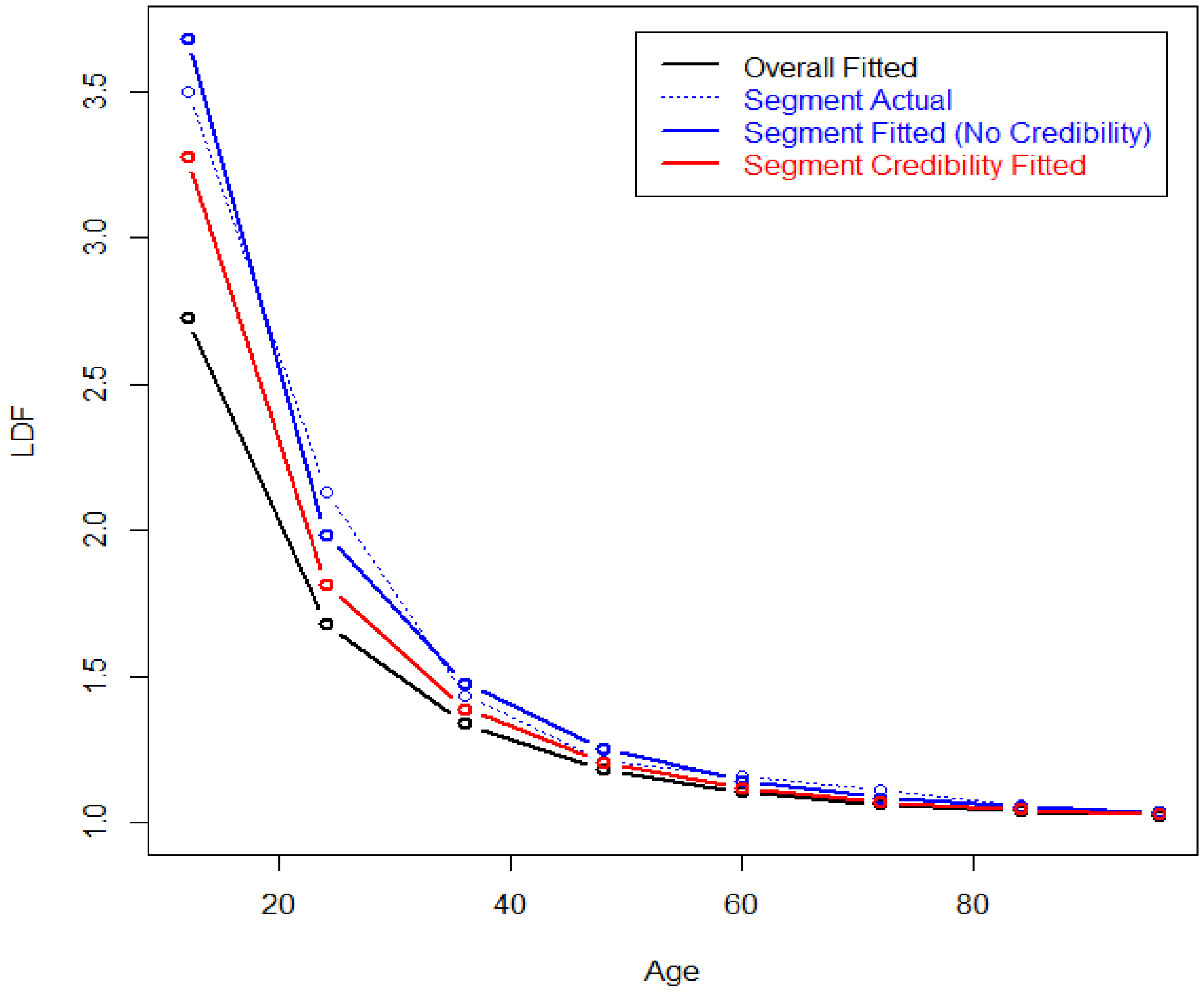
Where N is the logarithm of the Normal PDF

Can use the LDFs from a curve fitted to all of the data as the Complement LDF

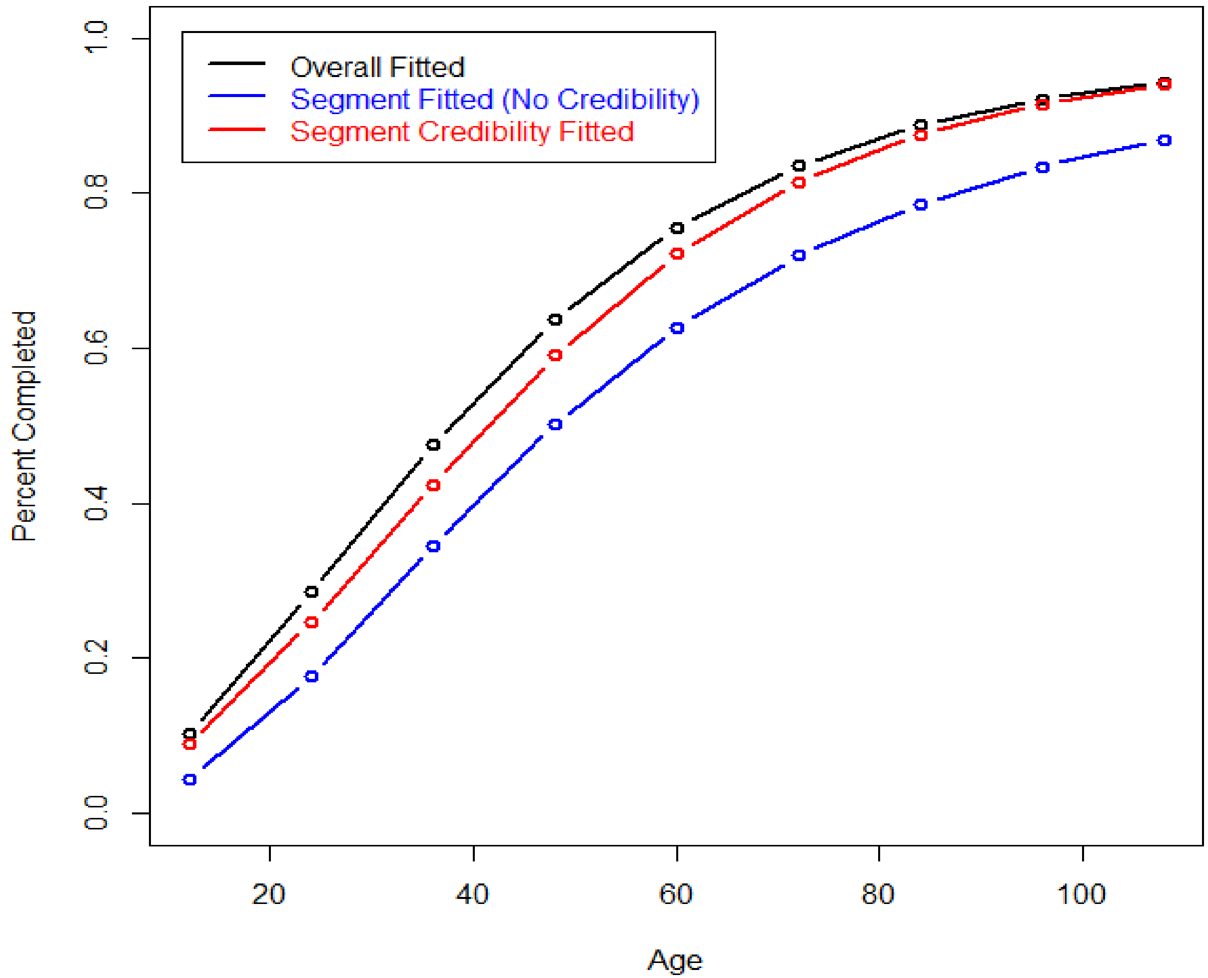
Credibility

- When fitting in Excel, suggested to first fit a curve to the overall data
  - Then run each segment separately using this as the complement of credibility
- When running in Stan/JAGS, run all segments together and let the model determine the complement of credibility
- When running in R, both options can be considered

SMIPOC Using Inversion Method



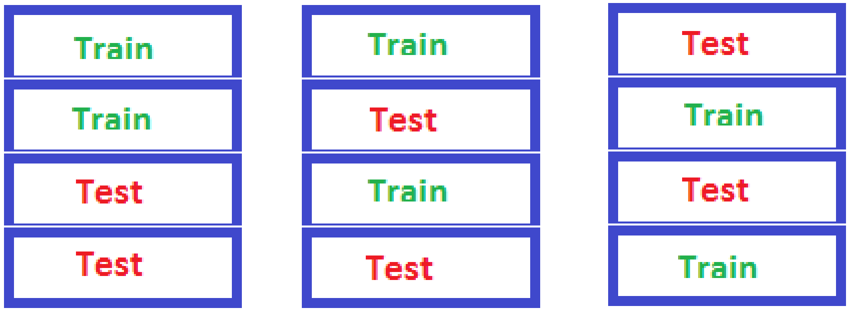
SMIPOD CDF Using Inversion Method



Calculating the Between Variance

- The between variance is needed for each LDF parameter
- Options:
  - Use the Buhlmann-Straub credibility formulas
    - Either on the selected LDF parameters, or on all of the LDFs and fit a curve
    - For the IPODs, divide so that the total volume equals the total claim count
    - Not completely lined up with method, but close
  - Build a Bayesian model
    - Requires specialized expertise
    - Can overestimate the between variance when the number of groups is small
  - Use cross validation
    - Easier
    - Does not suffer when number of groups is small
    - May work better with volatile data

Cross Validation



- Test multiple between variance values
- Fit on a fraction of the LDFs and calculate the likelihood on the remainder of the LDFs (without the prior component)
  - Use the same fractions for each set of variances being tested
- This can be done by either fitting on the individual LDFs, or by recalculating the average LDFs and fitting on those
- Options:
  - Do on each parameter using a grid search
  - Do on each parameter using random simulation (Bergstra & Bengio 2012)

Multidimensional Models

For each LDF parameter:

$$\log(LDF_{s,i} - 1)$$

$$= \log(Overall\ LDF\ Mean - 1) + State\ Coefficient_s + Industry\ Coefficient_i$$

Prior/Penalty:

$$\sum_{s=States} N(State\ Coefficient_s, 0, State\ Between\ Variance)$$

$$+ \sum_{i=Industries} N(Industry\ Coefficient_i, 0, Industry\ Between\ Variance)$$

- Total log-likelihood = sum of individual triangles' log-likelihoods
- Can handle cases where groups intersect
- Probably too many parameters for Excel. Use R or Bayesian software, etc.
- For between variance parameters, may need to use simulation approach, or use a ridge regression-type approach and use the same penalties for all groups (for each LDF parameter)

Continuous Variables

For each LDF parameter:

$$log(LDF\ Parameter\ -\ 1)\ =\ log(Overall\ LDF\ Mean\ -\ 1)\ +\ Coefficent\ x\ log(Account\ Size)$$

- If using IPOCs, use account size group. With IPODs, possible to use actual size of each account

Individual Account Credibility

- Can use these methods to smooth and credibility weight an individual client's LDFs with the portfolio's
- The between variances would be calculated by using a sample of actual accounts

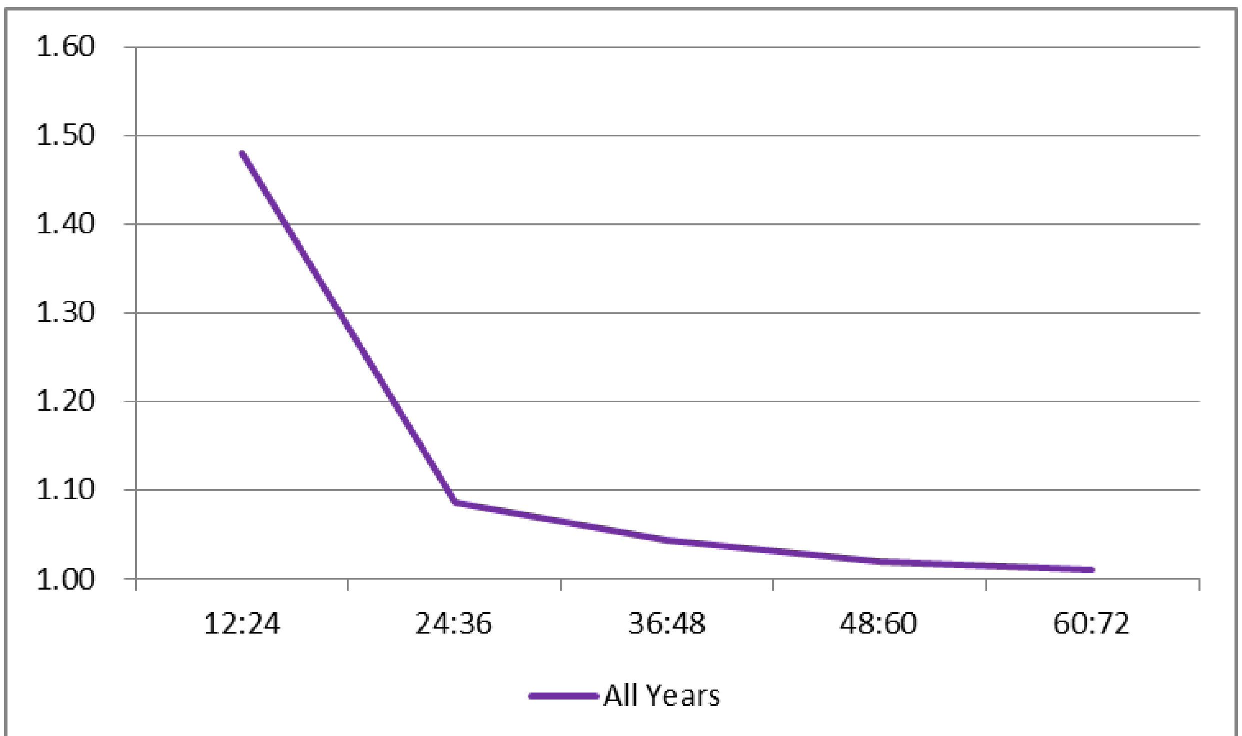
Part 3) Look-Back Period

Look-Back Period

- Determining the optimal look-back period to use can be difficult, especially with volatile data

Look-Back Period

	Age				
AY	**12:24**	**24:36**	**36:48**	**48:60**	**60:72**
2012	1.40	1.07	1.03	1.02	1.01
2013	1.40	1.07	1.05	1.02	
2014	1.40	1.10	1.05		
2015	1.60	1.10			
2016	1.60				

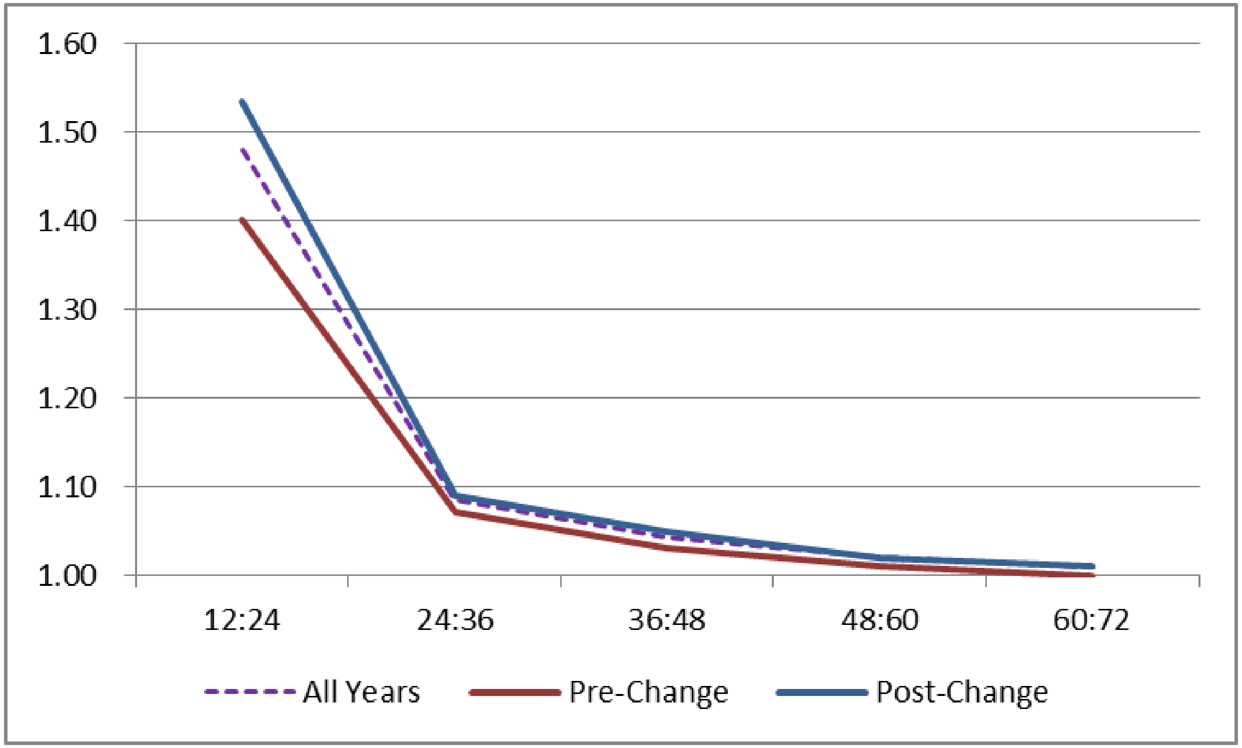


Look-Back Period

LDF Parameters A

LDF Parameters B

	Age				
AY	**12:24**	**24:36**	**36:48**	**48:60**	**60:72**
2012	1.40	1.07	1.03	1.02	1.01
2013	1.40	1.07	1.05	1.02	
2014	1.40	1.10	1.05		
2015	1.60	1.10			
2016	1.60				

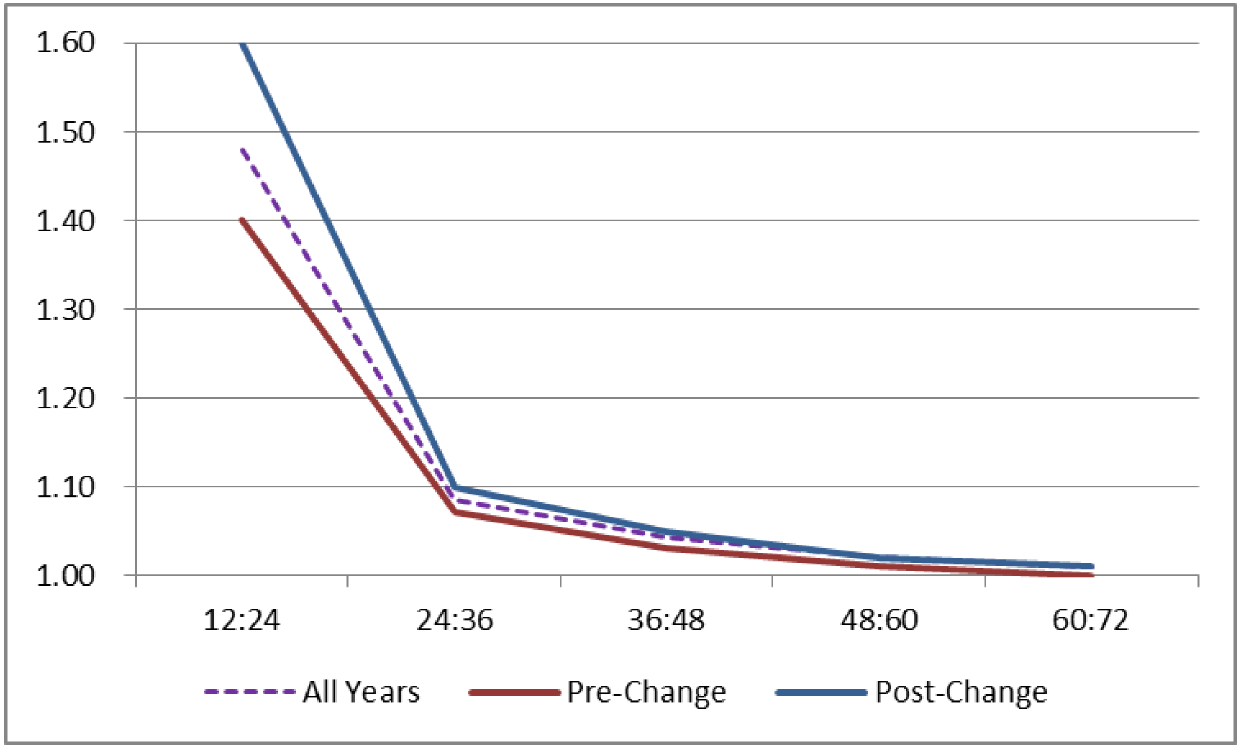


Look-Back Period

LDF Parameters A

LDF Parameters B

	Age				
AY	12:24	24:36	36:48	48:60	60:72
2012	1.40	1.07	1.03	1.02	1.01
2013	1.40	1.07	1.05	1.02	
2014	1.40	1.10	1.05		
2015	1.60	1.10			
2016	1.60				



Look-Back Period

- Choose the point with the highest likelihood
- Perform a significance test to see if change is statistically meaningful
    - Likelihood Ratio Test: 1 – Chisq CDF(2 * Likelihood Difference, df=Chg Parameters)

- Is only possible because we have fit a good curve to the LDFs (and possibly credibility weighting as well)
- Suggested to do this on the inverted LDF parameters
- (Is also possible to model gradual changes via a random walk…)

Part 4) Retentions, Limits, and Loss Caps

Retentions, Limits, and Loss Caps

- LDFs are often fit on data containing a mix of different retentions and limits

Retentions, Limits, and Loss Caps

- Solution: Use formulas to convert across different retentions and limits
  - (or possible to model as continuous variables too)

- This approach separates out frequency and severity
  - Assumes that claim count development factors (CCDFs) are known, as well as the severity distribution

Step 1

$$LDF_t = CCDF_t \times SDF_t$$

(Loss development consists of the arrival of new claims as well as increased severity of both the existing and new claims)

$$SDF_t = \frac{LDF_t}{CCDF_t}$$

NOTE: All LDFs in this section refer to age-ultimate factors

Step 2

$$SDF_t(c1) = \frac{LEV_T(c1)}{LEV_t(c1)}$$

Step 3

$$SDF_t(c1) = \frac{LEV(\theta;\, c1)}{LEV(\theta/a_t;\, c1)}$$

Solve for a(t)

$$SDF_t(c2) = \frac{LEV(\theta;\, c2)}{LEV(\theta/a_t;\, c2)}$$

$$LDF_t(c2) = CCDF_t \times SDF_t(c2)$$

Different Retentions As Well

- When calculating the expected average severities, take the retention into account as well
  - (This is the formula for the conditional severity)

$$\frac{LEV(AP + Cap) - LEV(AP)}{s(AP)}$$

- $CCDF_t(AP2) = CCDF_t(AP1) \times \frac{s_T(AP2) / s_T(AP1)}{s_t(AP2) / s_t(AP1)}$

Converting LDFs

- Use cases:
  - Using LDFs at a more stable layer, produce the LDFs for less stable layers, e.g. higher retentions
  - Use formulas for converting LDFs simultaneously while fitting a curve (or even credibility weighting) and model everything together

Simultaneous Fitting

- For IPOCs:
  - Choose a base level to represent the fitted parameters
  - Use these LDFs to back into the a(t) parameters for each age (from the age-to-ultimate factors)
    - To automate, create a table: 1 / a is always between 0 and 1
  - Use all of this to calculate the LDFs at other layers
  - For the CoV curve, add extra parameter(s) for the different layers
  - Calculate the log-likelihoods of each group and sum together

Simultaneous Fitting

- For IPODs:
  - Choose a base layer…
  - Calculate the LDFs = 1 / CDF
    - (Remember, the LDFs are the age-to-ult factors here)
  - Use formulas to convert LDFs to other layers
  - For each layer, CDF = 1 / LDF
  - Use CDFs to calculate the log-likelihoods for each layer and sum up
  - If necessary, can assign less weight to more volatile layers

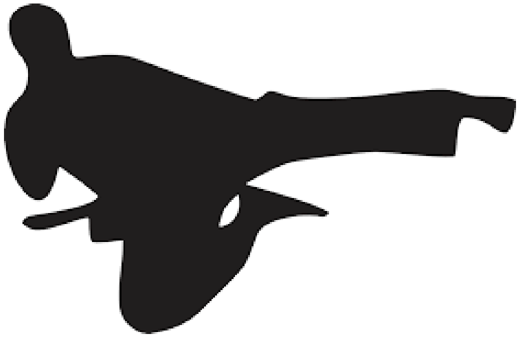
Simultaneous Fitting

- For the CCDFs:
  - Choose a base layer
  - Need separate parameters for fitting the CCDFs
    - Suggest using the right truncated approach with Gamma (or Weibull) distribution
  - Use the same *a* parameters along with formulas to convert CCDFs to different layers
  - Calculate the combined log-likelihood and add to total

- OR:
  - Calculate separately, before fitting the LDFs using the right truncated approach but treating the layer as continuous variable(s)

- OR:
  - Also calculate the CCDFs separately
  - Choose 2 base layers for which to setup parameters for
  - Use the relationships between them to calculate the a parameters
  - From these, all of the CCDFs for the other layers can be calculated

Recap



- Curve Fitting
  - DIPOC
  - SMIPOC
  - RIPOD
  - SMIPOD
- Triangle Segmentations
  - Credibility
    - Use Bayesian credibility with parameter inversion
  - Multidimensional Models
  - Continuous Variables

Recap

- Choosing the Optimal Look-Back Period
  - Fit using your curve of choice
  - Use the inverted formulas
  - Test the addition of more parameters at different points by CY, AY, etc.
  - Perform a significance test
  - If good, repeat

- Retentions, Limits, and Loss Caps
  - Frequency/Severity approach – CCDFs and severity distribution
  - Use the formulas to convert
  - Either fit the most stable layer and then convert, or model everything simultaneously

- A lot of pictures of ninjas, for some reason

Congratulations

Sources

- Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13, 281–305. http://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf
- Clark, David R. "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach," CAS Forum, Fall 2003. http://www.casact.org/pubs/forum/03fforum/03ff041.pdf
- Dean, C., "Topics in Credibility Theory," Education and Examination Committee of the Society of Actuaries, 2005, https://www.soa.org/files/pdf/c-24-05.pdf
- England P., and Verrall, R., "A Flexible Framework for Stochastic Claims Reserving," Proceedings of the Casualty Actuarial Society, 2001: LXXXVIII, pp. 1-38. http://www.casact.org/pubs/proceed/proceed01/01001.pdf
- England, P. D. and Verrall, R. J. 2002. Stochastic Claims Reserving in General Insurance. British Actuarial Journal 8:443-544.
- http://www.cassknowledge.com/sites/default/files/article-attachments/371~~richardverrall_-_stochastic_claims_reserving.pdf
- Gelman, A. Scaling regression inputs by dividing by two standard deviations. Stat. Med. 27: 2865–2873. 2008
- Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin. Bayesian Data Analysis (3rd Edition). CRC Press. 2014. pp. 128-132
- Klugman, S. A., H. H. Panjer, and G. E. Willmot, Loss Models: From Data to Decisions (1st Edition). New York: Wiley. 1998. pp. 62-66.
- Korn, Uri, "A Frequency-Severity Stoachastic Approach to Loss Development," CAS E-Forum, Sprint 2015. http://www.casact.org/pubs/forum/15spforum/Korn.pdf
- Lowe, S. P., and Mohrman, D. F., "Extrapolating, Smoothing and Interpolating Development Factors [Discussion]," Proceedings of the Casualty Actuarial Society Casualty Actuarial Society, 1985: LXXII, pp. 182-189. http://www.casact.org/pubs/proceed/proceed85/85182.pdf
- Sahasrabuddhe, R., "Claims Development by Layer: The Relationship between Claims Development Patterns, Trend and Claim Size Models," CAS E-Forum, Fall 2010, pp. 457-480
- https://www.casact.org/pubs/forum/10fforum/Sahasrabuddhe.pdf
- Sherman, R. E., "Extrapolating, Smoothing and Interpolating Development Factors," Proceedings of the Casualty Actuarial Society Casualty Actuarial Society, 1984: LXXI, pp. 122-155. https://www.beanactuary.com/pubs/proceed/proceed84/84122.pdf
- Siewert, J. J., "A Model for Reserving Workers Compensation High Deductibles," CAS Forum, Summer 1996, pp. 217-244.
- https://www.casact.org/pubs/forum/96sforum/96sf217.pdf
- Venables, W. N. znd Ripley, B. D. Modern Applied Statistics with S. Fourth Edition. Springer, New York. 2002. ISBN 0-387-95457-0.