Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Reserving for the next 100 years: Thinking outside the triangle

Nov 11, 2014

Don Mango

Jessica Leong

Dave Clark

Jim Guszcza

A tale of
two actuaries

With what we have to work with, I think we have done a fair job in setting reserves

With what we have to work with, I think we have done a fair job in setting reserves

[But] we are still relying on the same methods that I started my career with thirty years later!

With what we have to work with, I think we have done a fair job in setting reserves

[But] we are still relying on the same methods that I started my career with thirty years later!

… we must move to stochastic methods.

With what we have to work with, I think we have done a fair job in setting reserves

[But] we are still relying on the same methods that I started my career with thirty years later!

… we must move to stochastic methods.

You can program a computer to play musical notes, but only a musician can make music. I feel the same way about loss reserving methods.

Give an actuary …the tools to dig at what's behind that triangle and the result will be better than Mack's attempt…

Give an actuary …the tools to dig at what's behind that triangle and the result will be better than Mack's attempt…

[But there’s] …too much reliance on "actuarial judgement"

Give an actuary …the tools to dig at what's behind that triangle and the result will be better than Mack's attempt…

[But there’s] …too much reliance on "actuarial judgement"

It’s as if reserving has been stuck in producing new versions of the Windows operating platform, during which time pricing created the smartphone.

[True]. Predictive Analytics has worked nicely for auto, where our reserving record is pretty good, too,

[True]. Predictive Analytics has worked nicely for auto, where our reserving record is pretty good, too,

[But] the jury is still out on predictive modeling …for long tail lines. Finding the right model parameters for long-tail lines is tricky.

[True]. Predictive Analytics has worked nicely for auto, where our reserving record is pretty good, too,

[But] the jury is still out on predictive modeling …for long tail lines. Finding the right model parameters for long-tail lines is tricky.

[What] "newer" methods …provide [a] better understanding…[?]

Boot Strapping

Boot Strapping

incremental cost per closed

Boot Strapping

incremental cost per closed

Hierarchical linear models

Boot Strapping

incremental cost per closed

Hierarchical linear models

Neural networks

Boot Strapping

incremental cost per closed

Hierarchical linear models

Neural networks

support vector machines

Boot Strapping

incremental cost per closed

Hierarchical linear models

Neural networks

support vector machines

restricted boltzmann machines

Boot Strapping

incremental cost per closed

Hierarchical linear models

Neural networks

support vector machines

restricted boltzmann machines

erm…[?]

The End

Cue: Orchestra

Cue: Polite applause

Reserving for the next 100 years: Thinking outside the triangle

Don Mango

Jessica Leong

Dave Clark

Jim Guszcza

Agenda

1. Do we have a problem?
2. What is it?
3. What are some solutions?

Agenda

1. Do we have a problem?
2. What is it?
3. What are some solutions?

Do we have a problem?

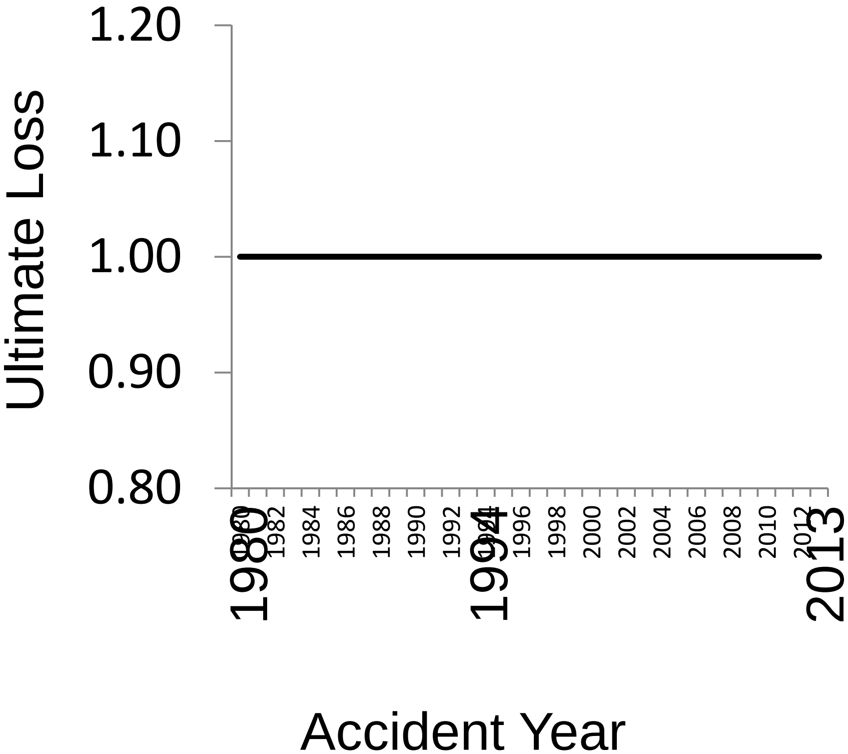
Do we have a problem?

“There isn't much of a reserving "problem." The methods we've been utilizing over these last decades make decent estimates for what we believe the cost will be.”

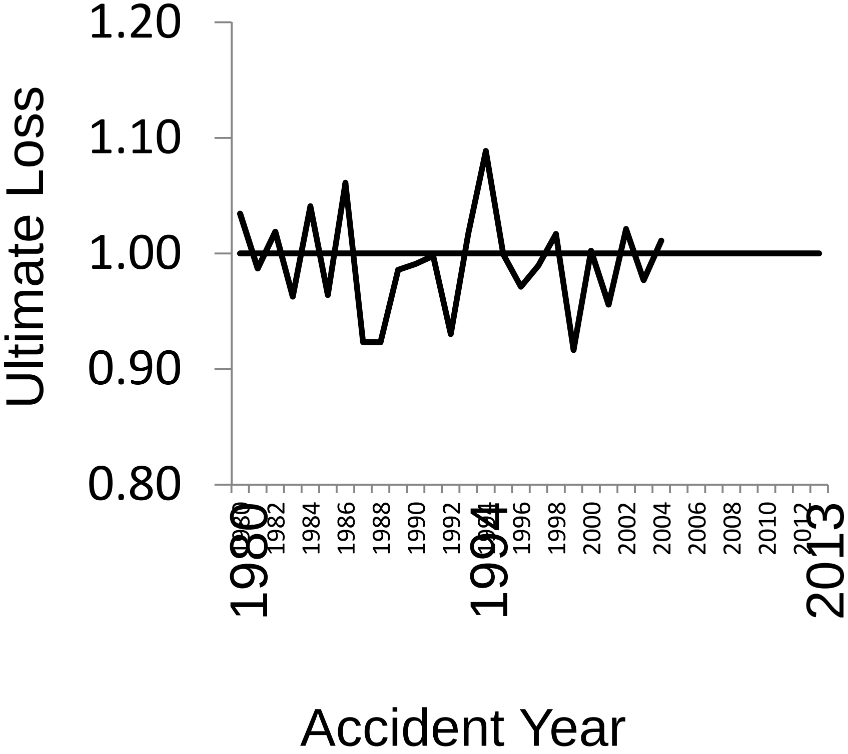
Reserving theory	Reserving practice

← Reserving theory | Reserving practice →

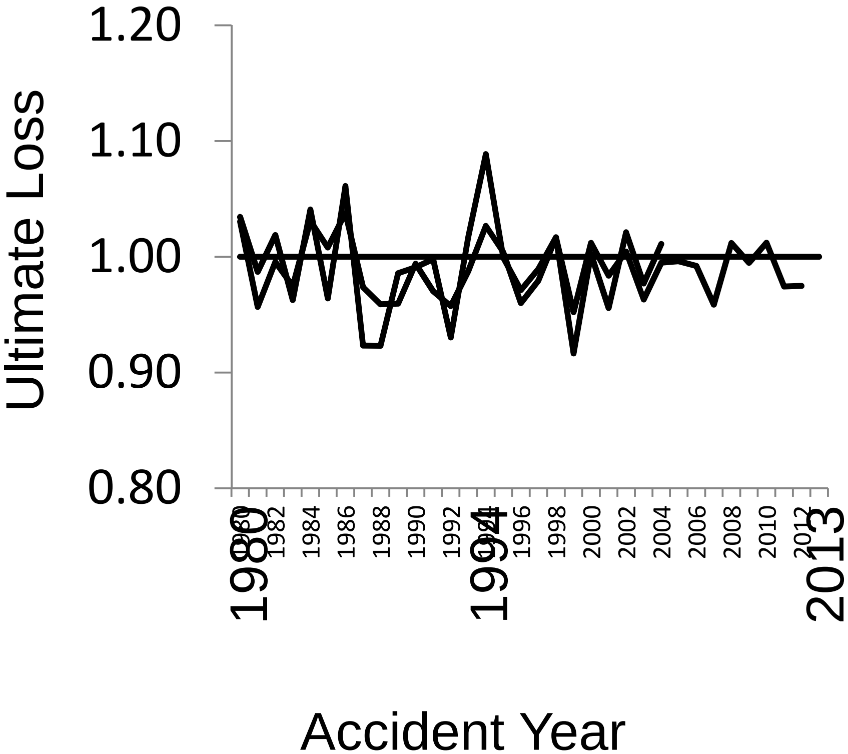
Reserving theory | Reserving practice



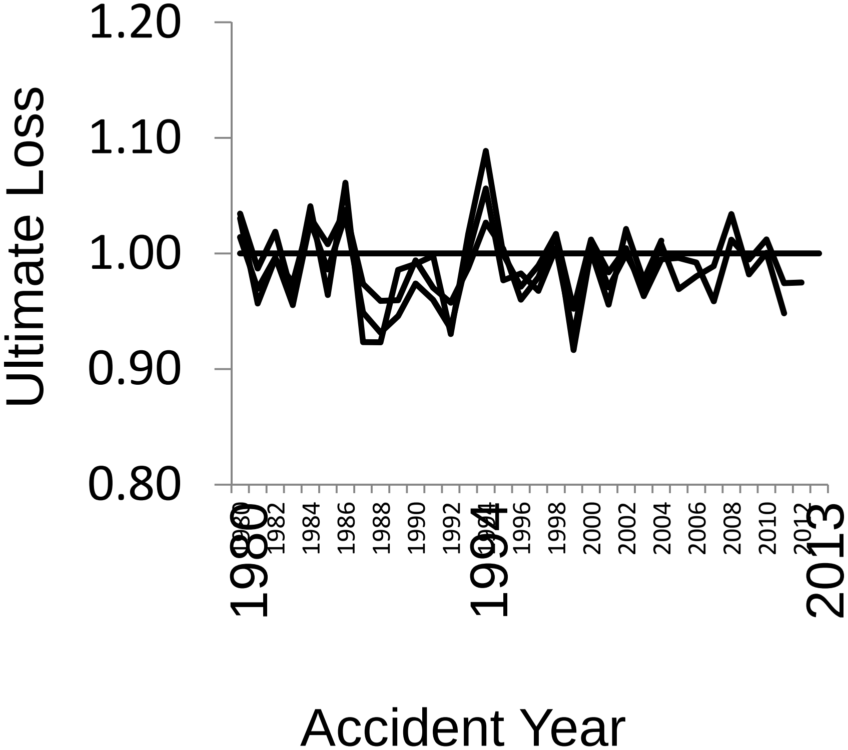
← Reserving theory | Reserving practice →



← Reserving theory | Reserving practice →

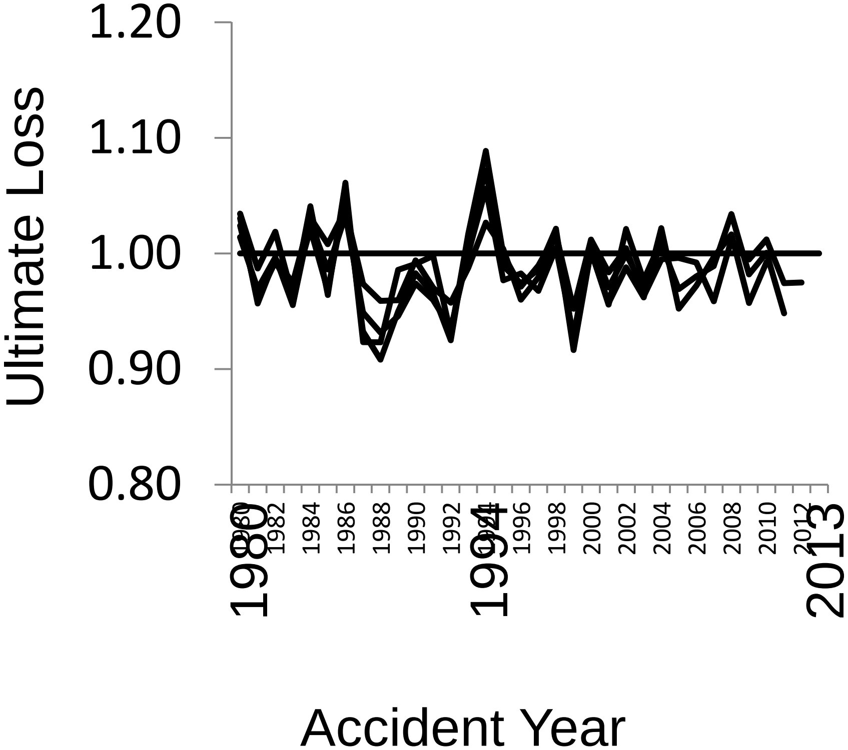


← Reserving theory | Reserving practice →

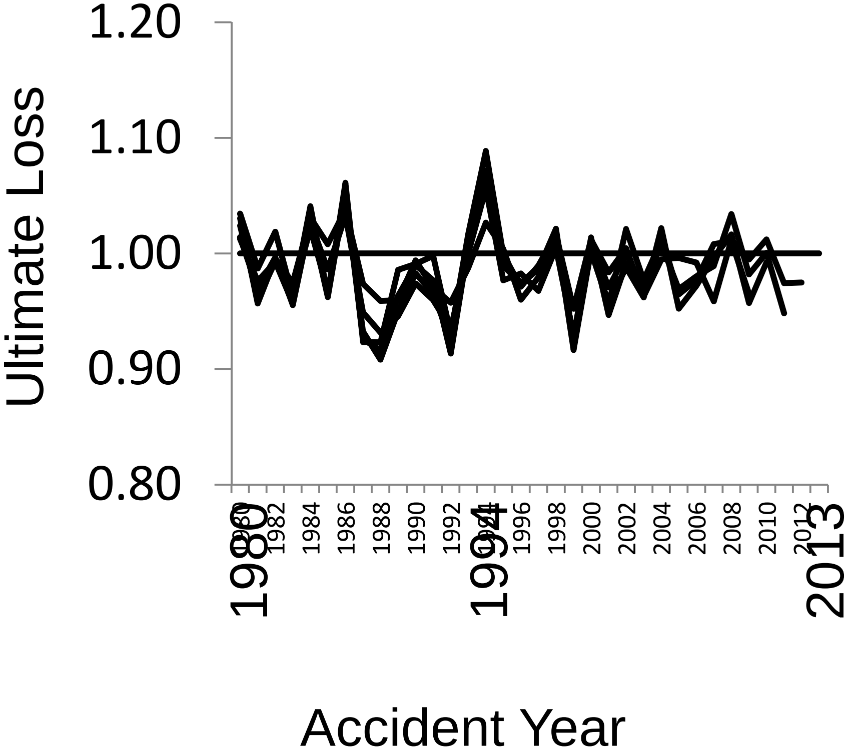


Reserving theory

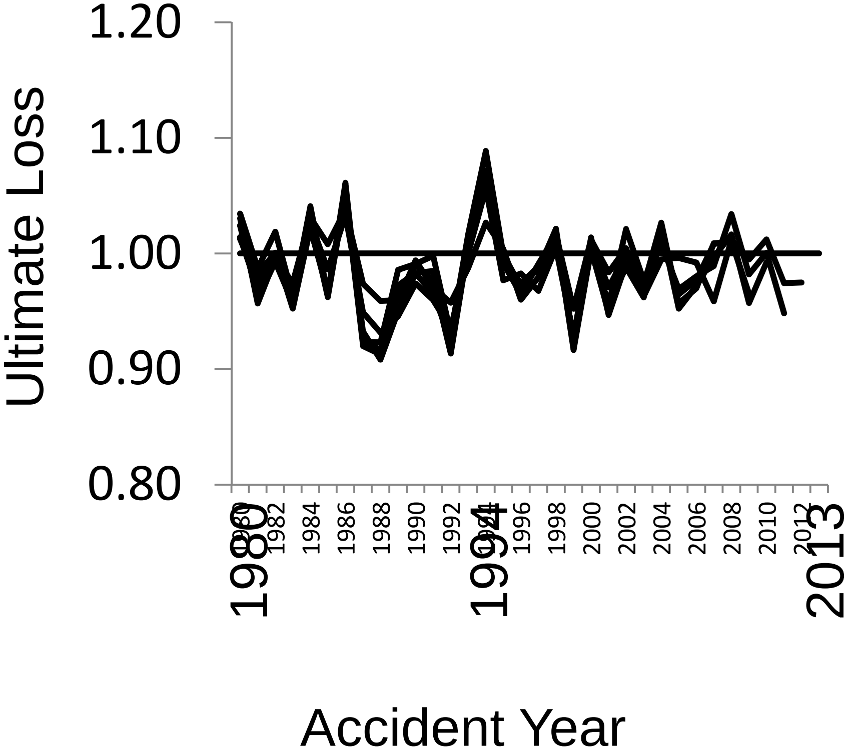
Reserving practice



← Reserving theory | Reserving practice →

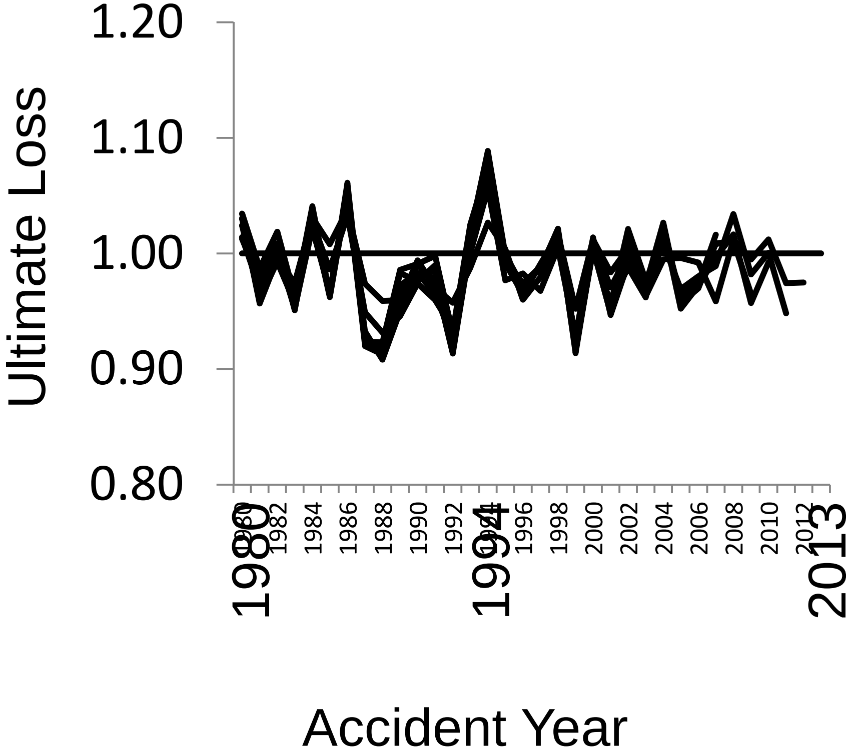


⟵ Reserving theory | Reserving practice ⟶

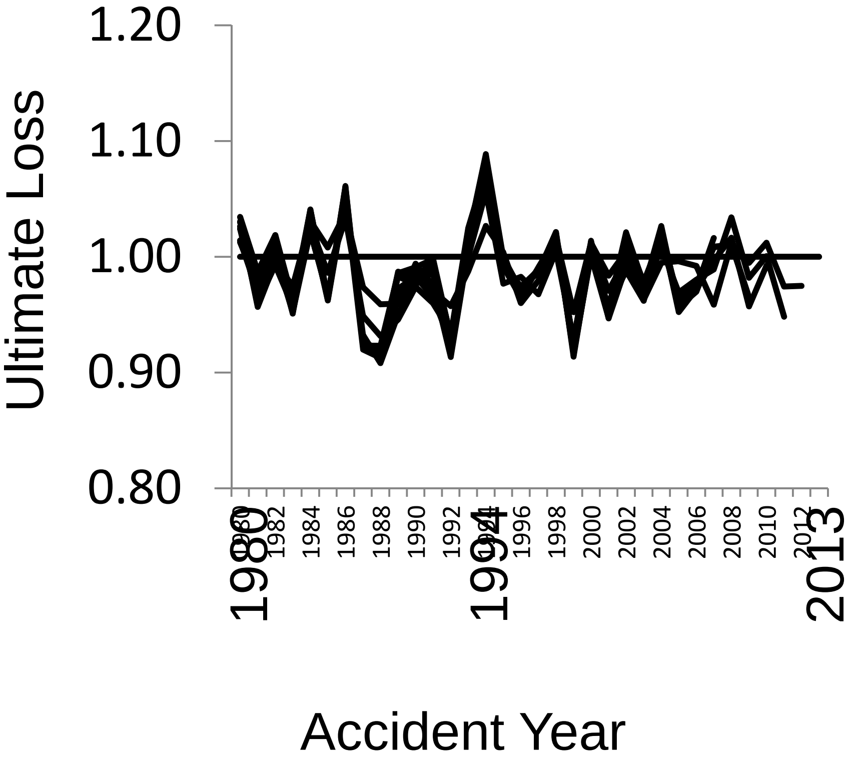


← Reserving theory

Reserving practice →

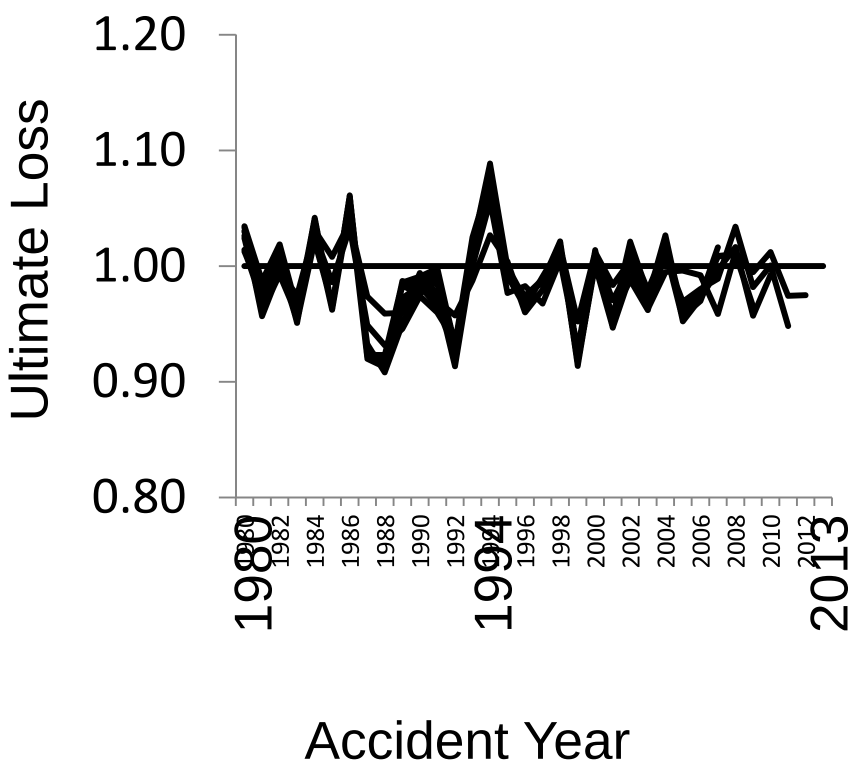


← Reserving theory | Reserving practice →



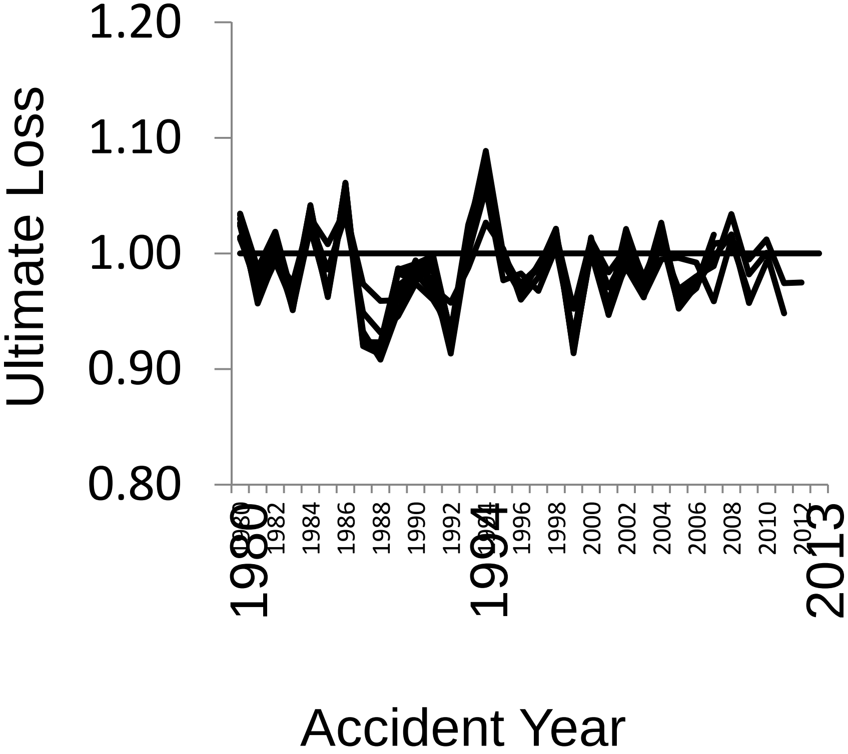
⟵ Reserving theory

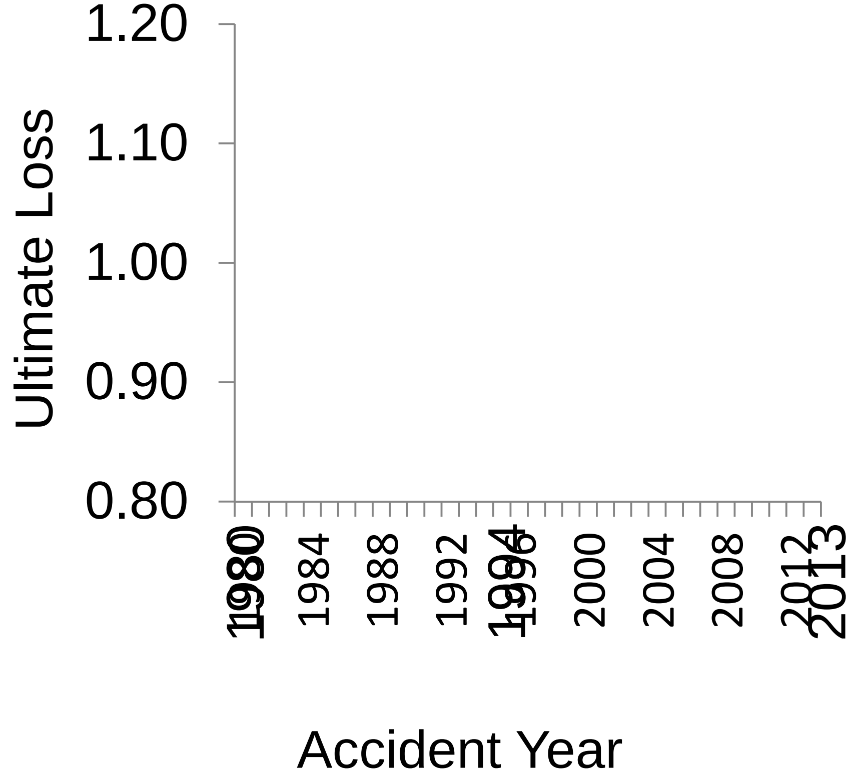
Reserving practice ⟶



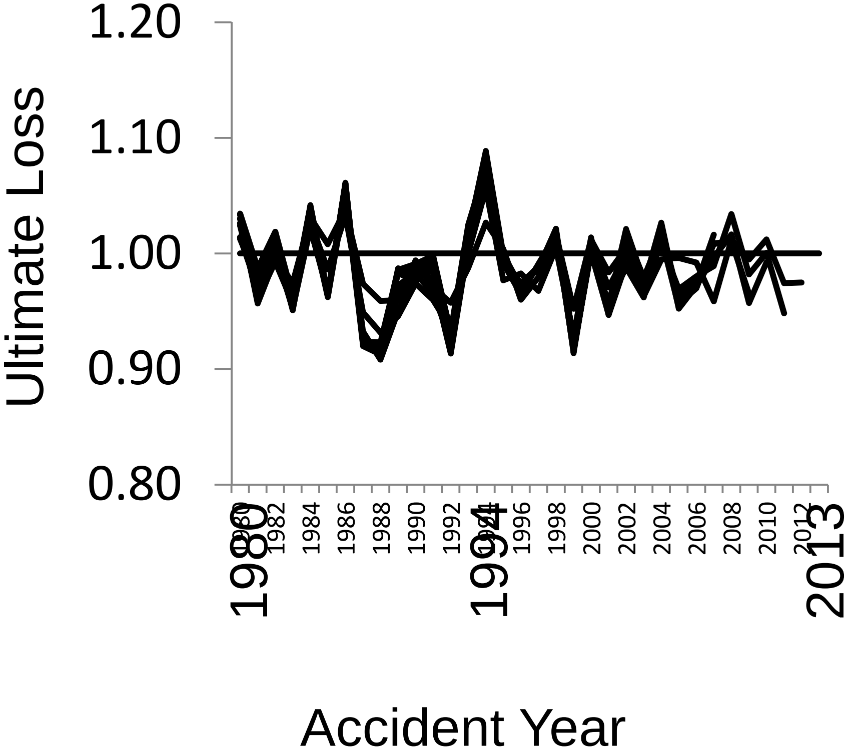
Reserving theory

Reserving practice

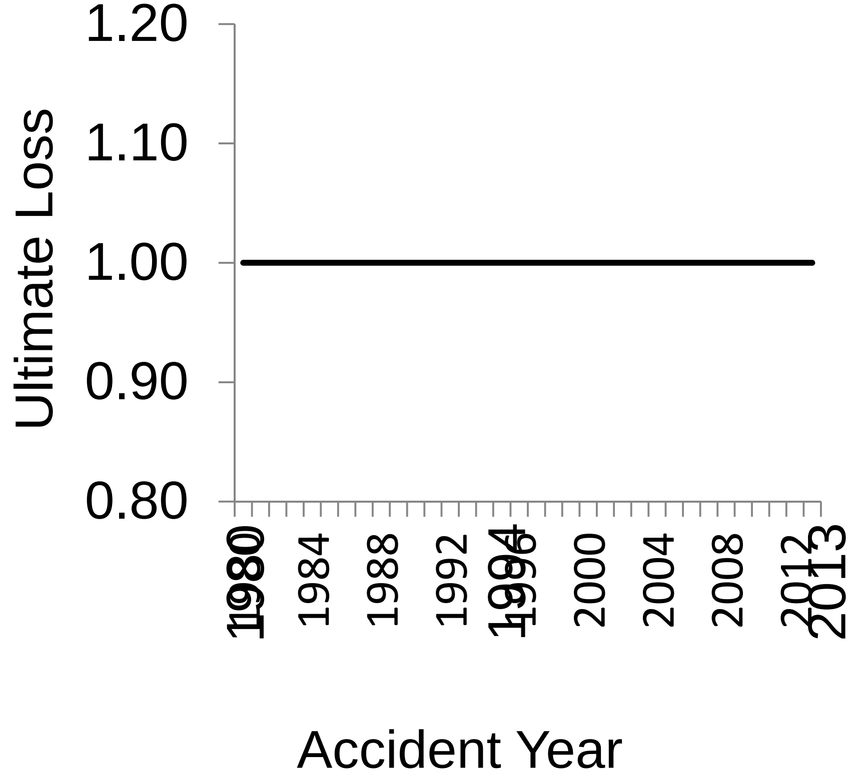




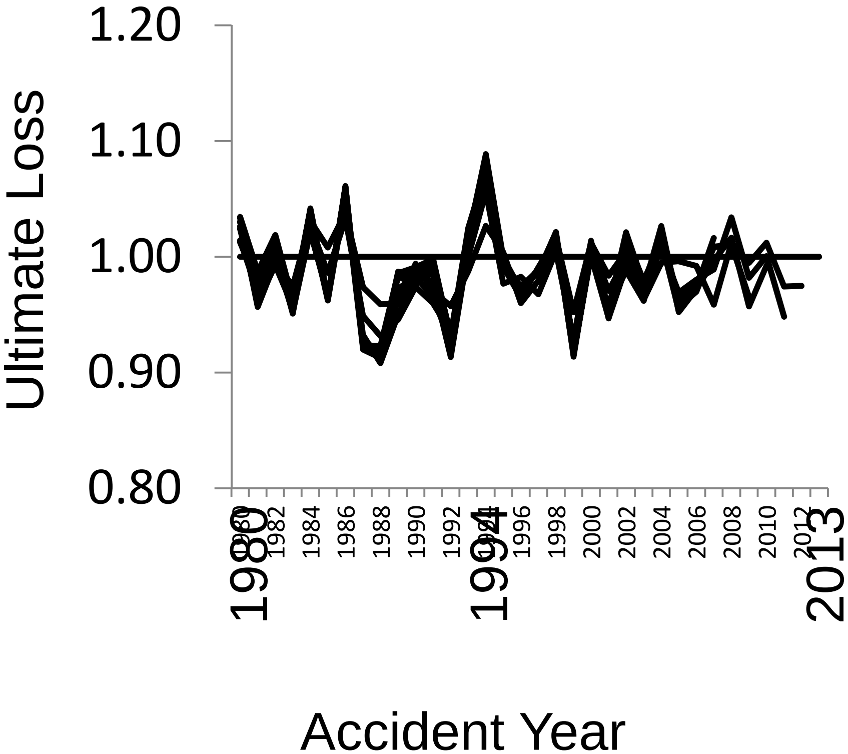
Reserving theory



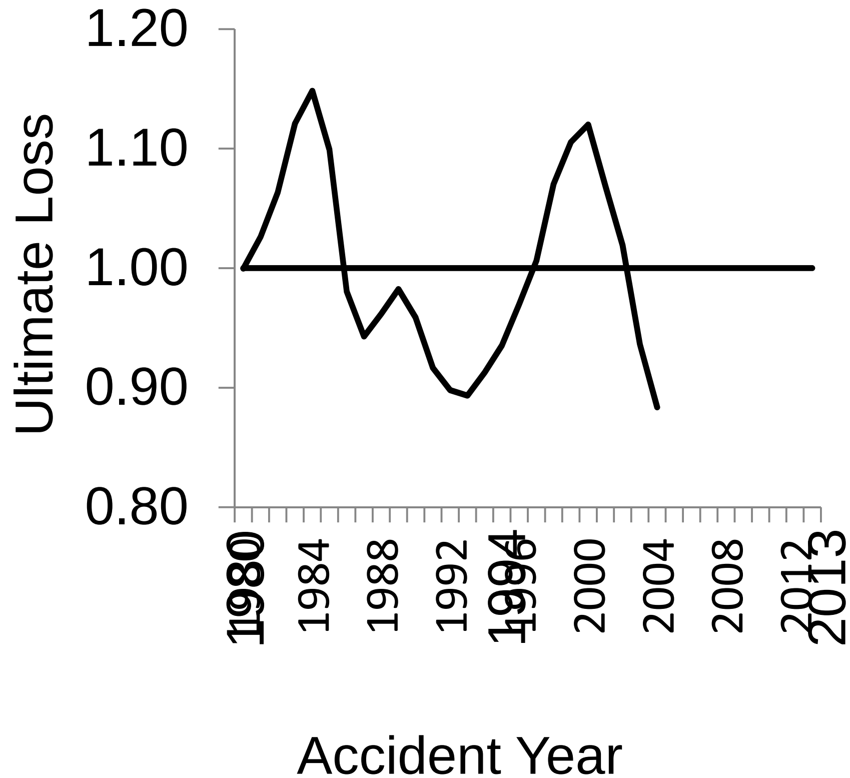
Reserving practice



← Reserving theory

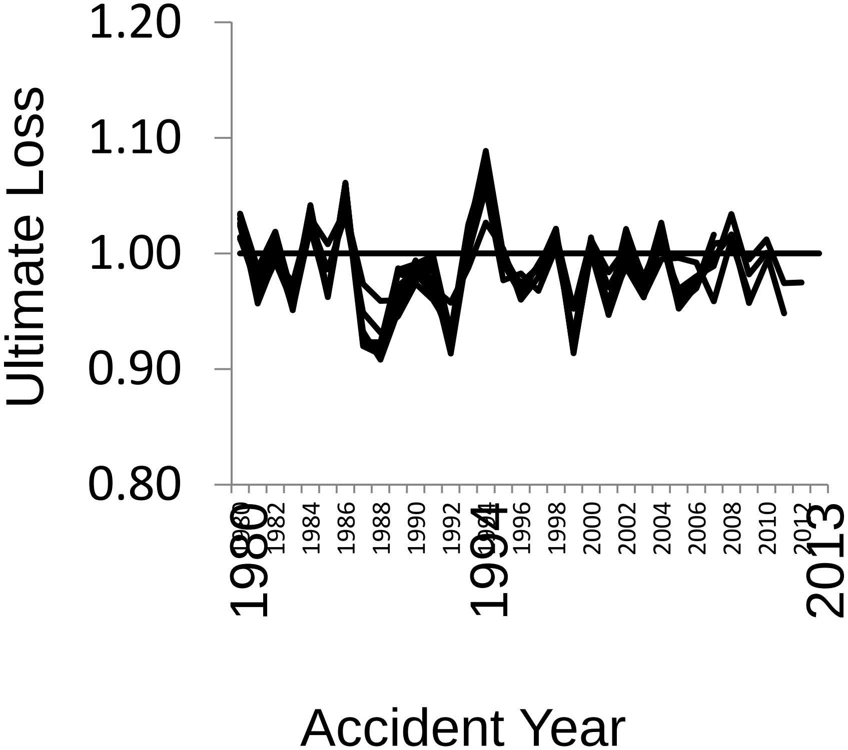


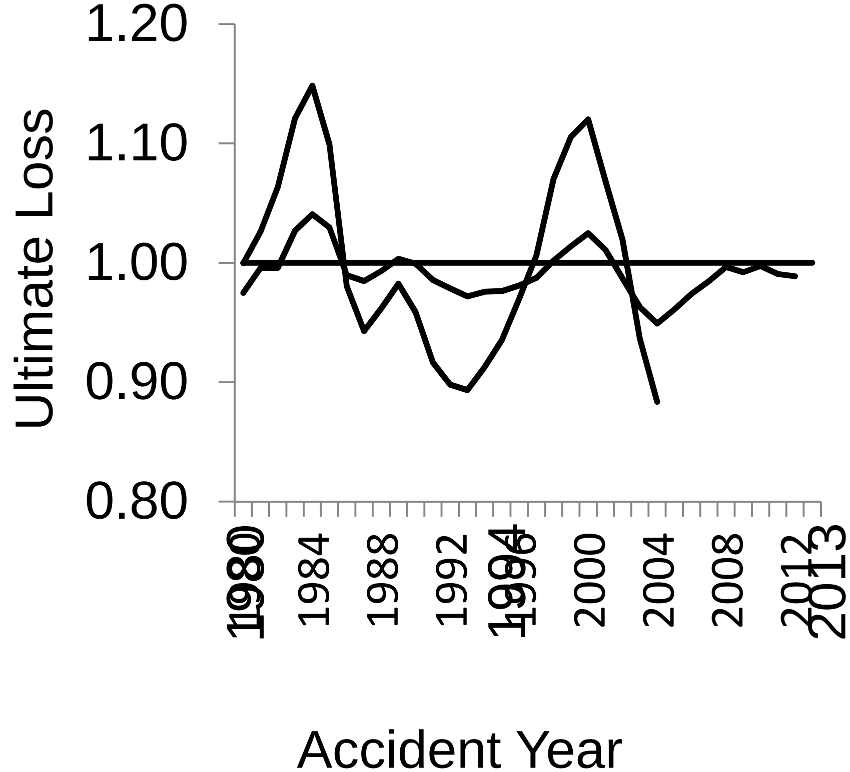
Reserving practice →



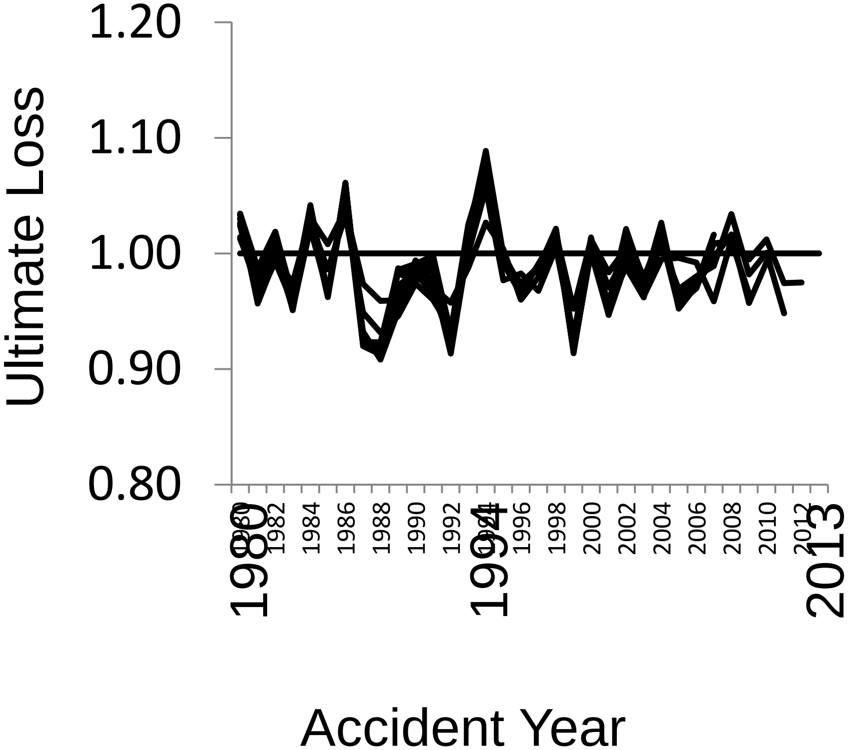
Reserving theory

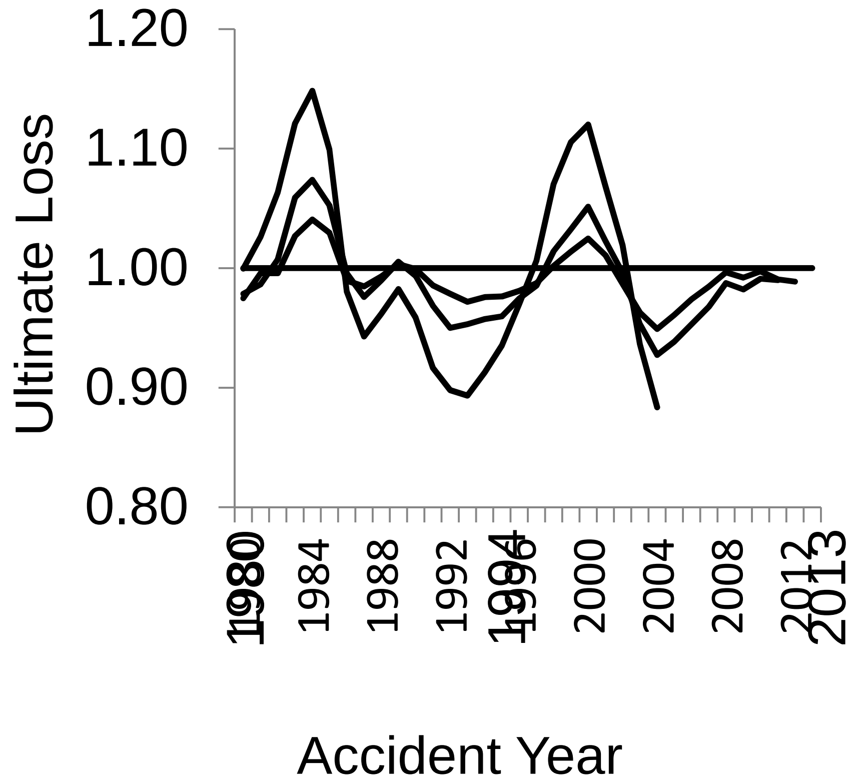
Reserving practice





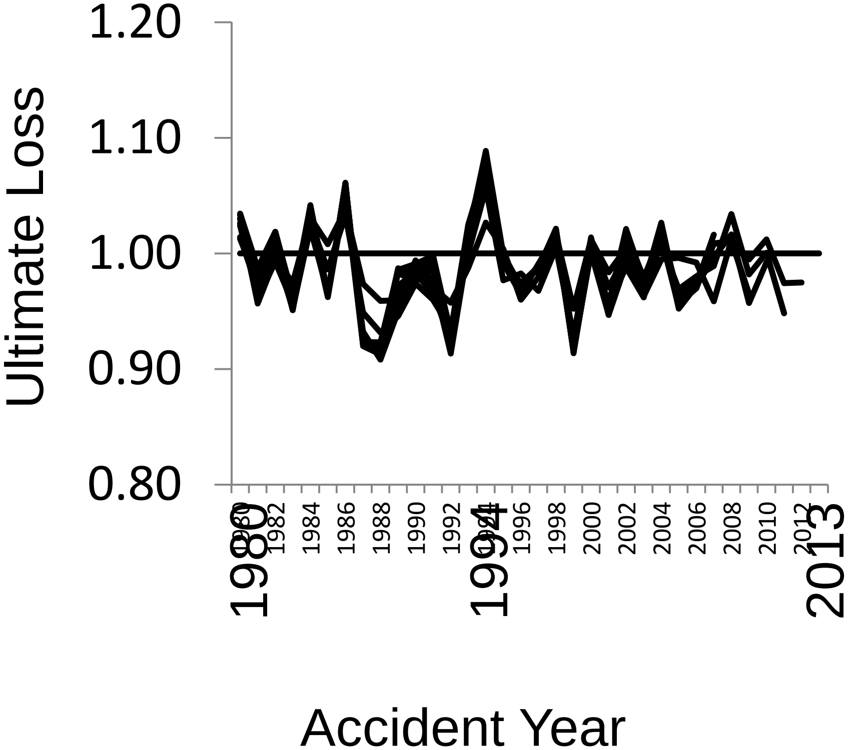
Reserving theory | Reserving practice

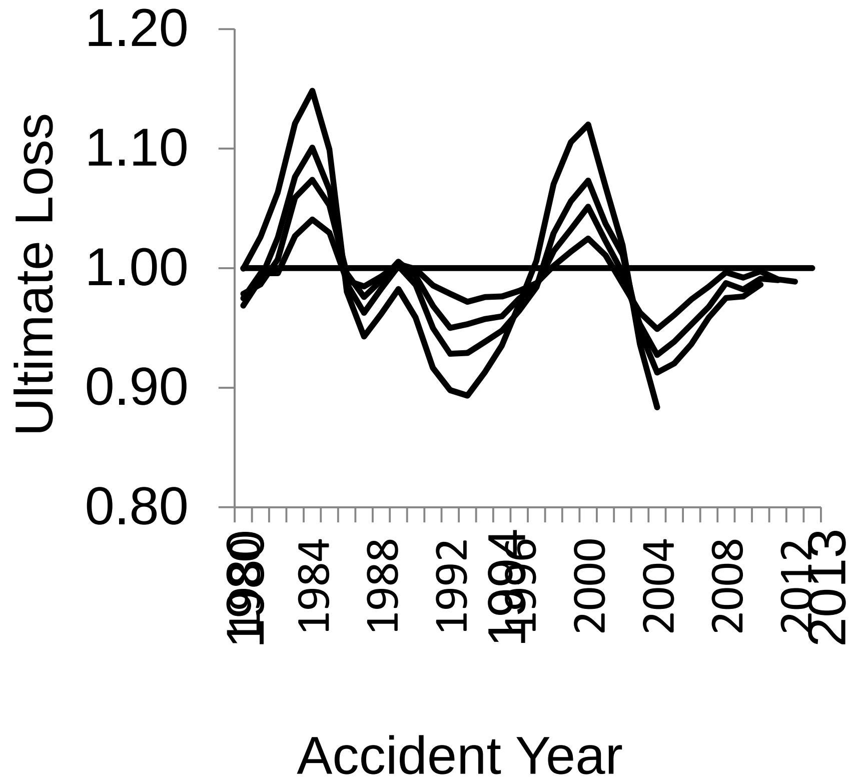




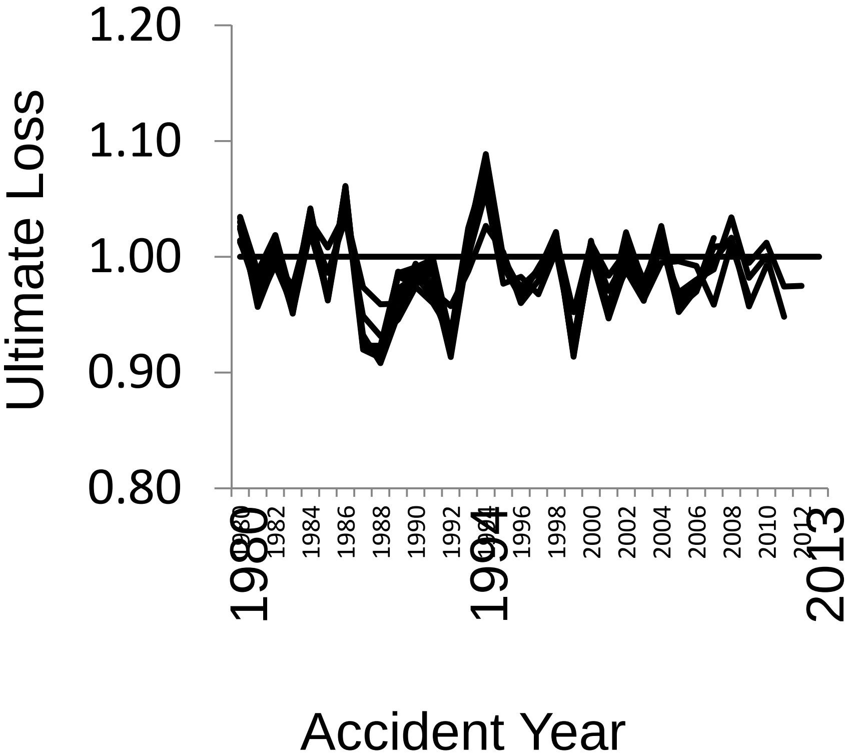
← Reserving theory

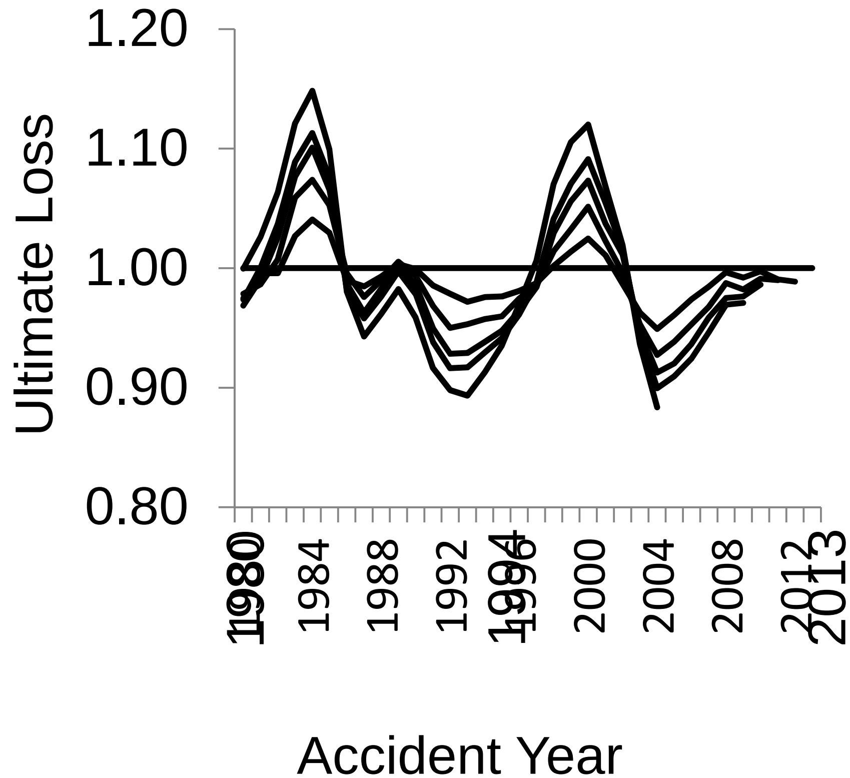
Reserving practice →



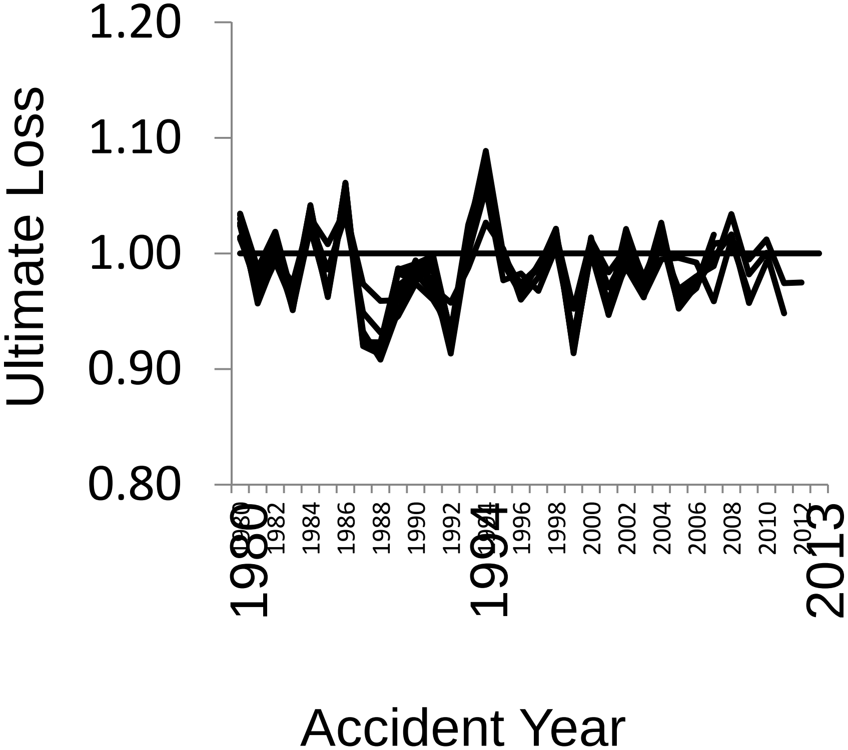


Reserving theory | Reserving practice

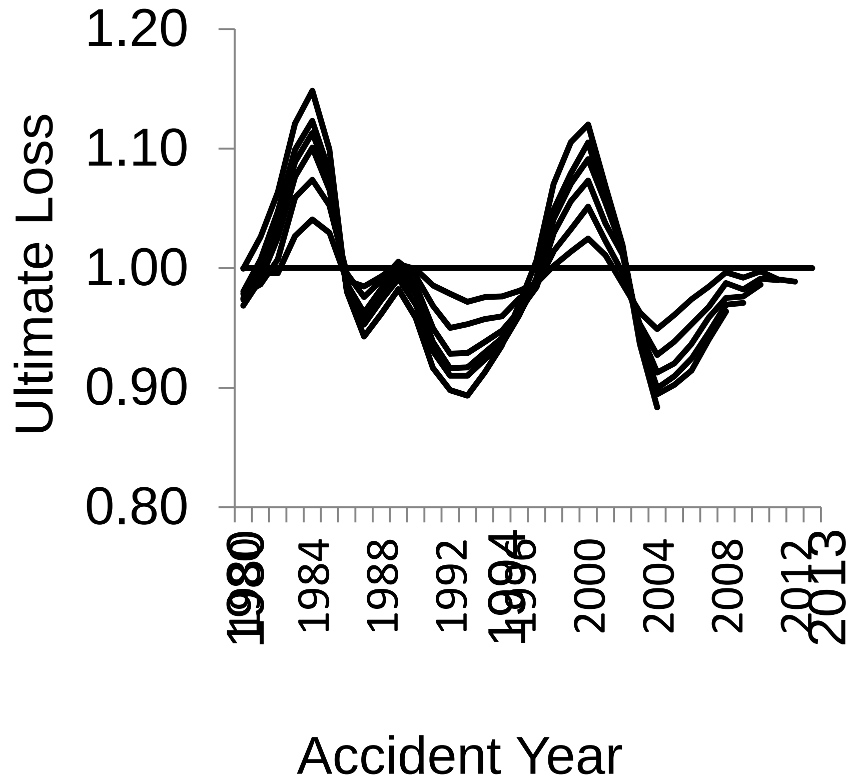




Reserving theory

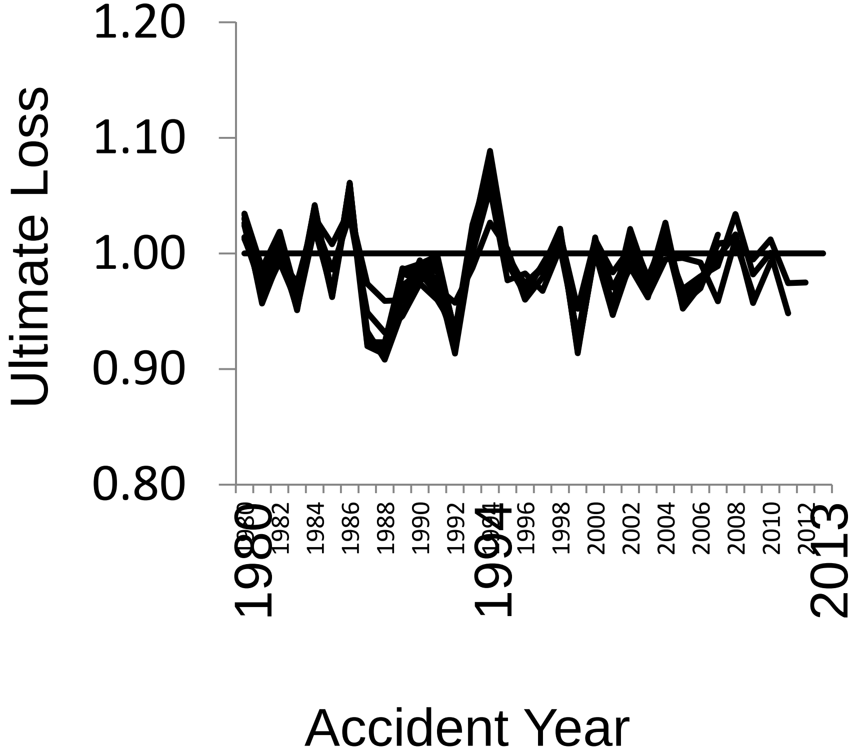


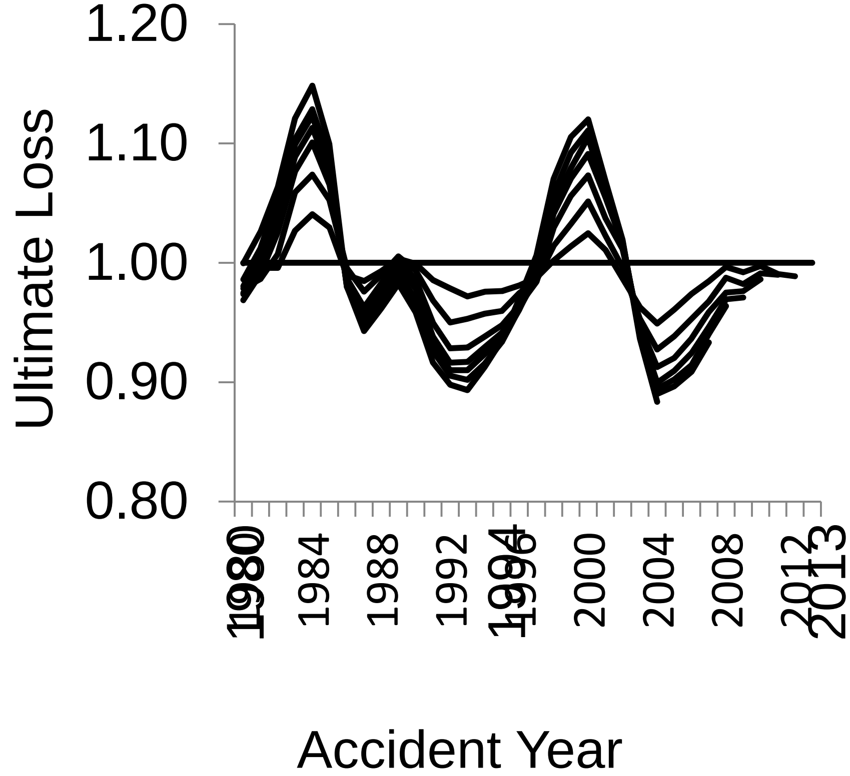
Reserving practice



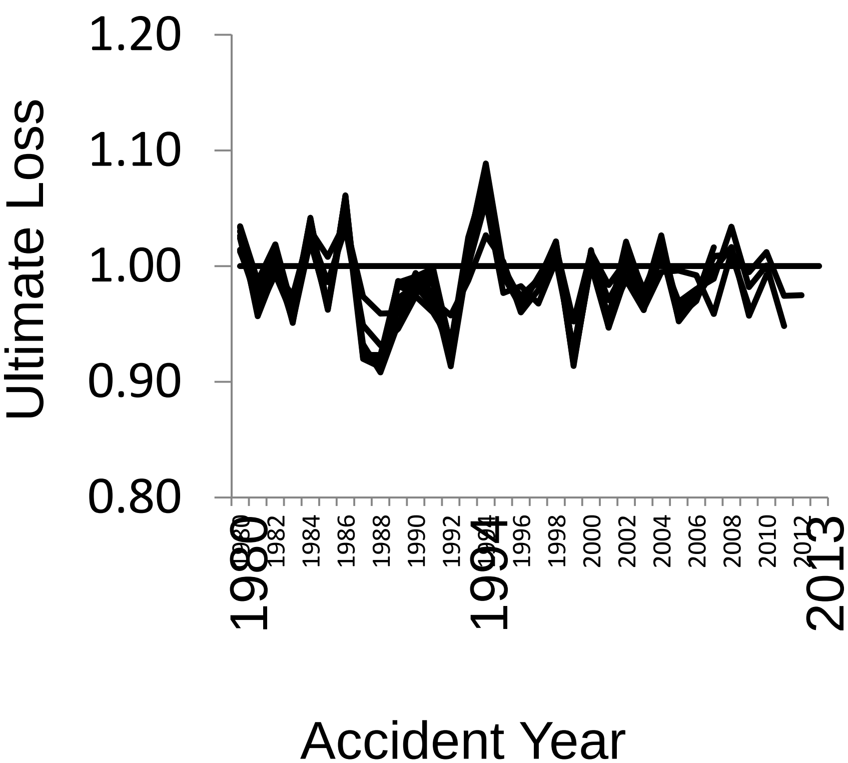
Reserving theory

Reserving practice

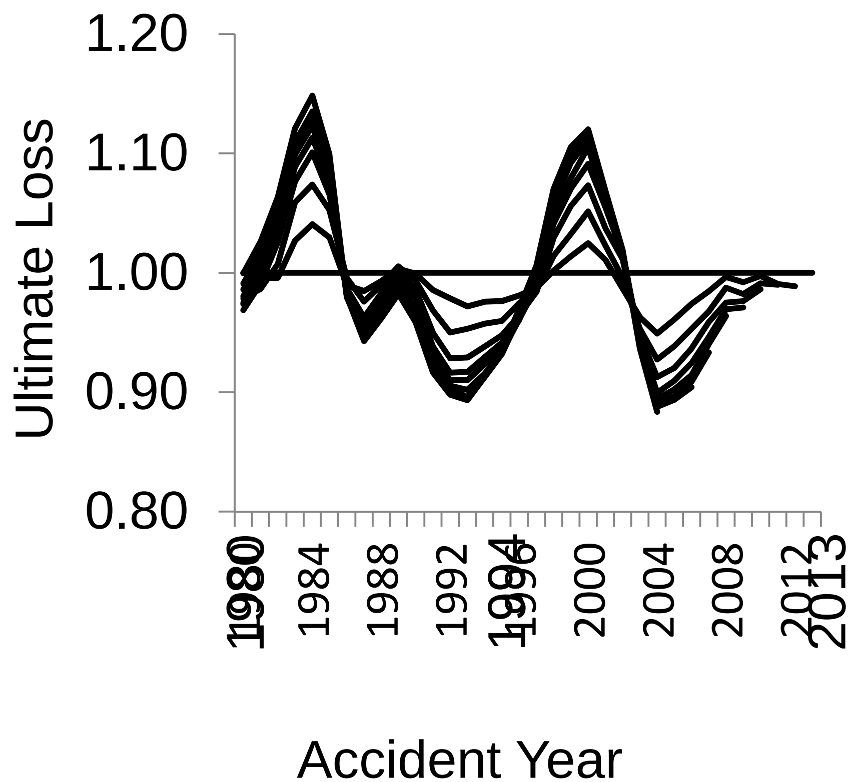


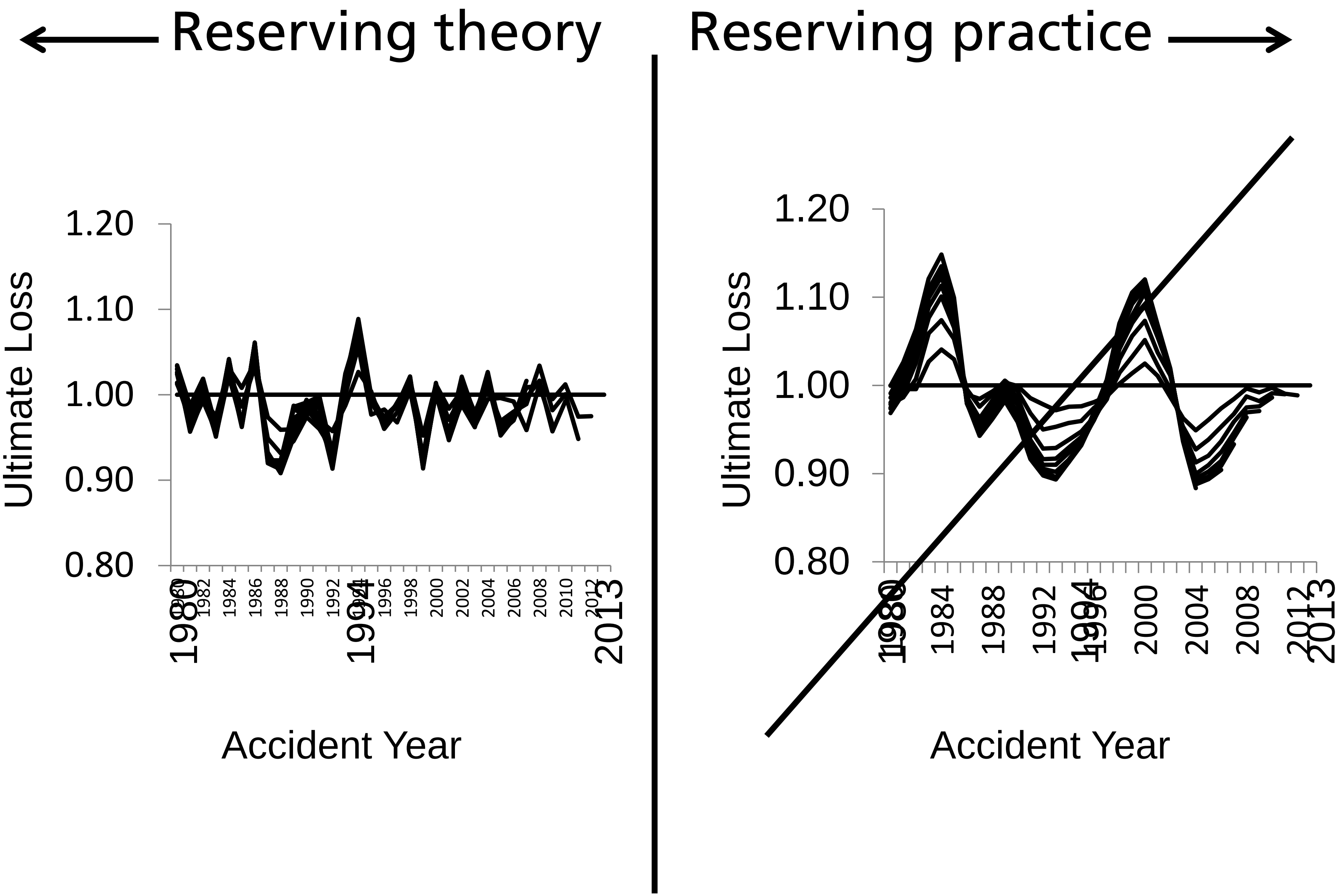


← Reserving theory



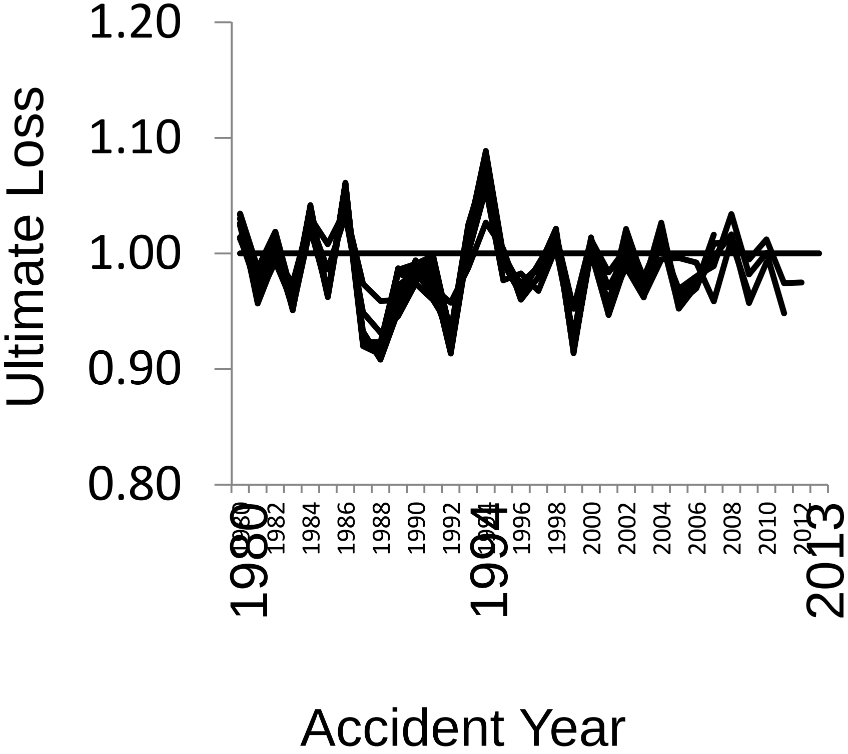
Reserving practice →

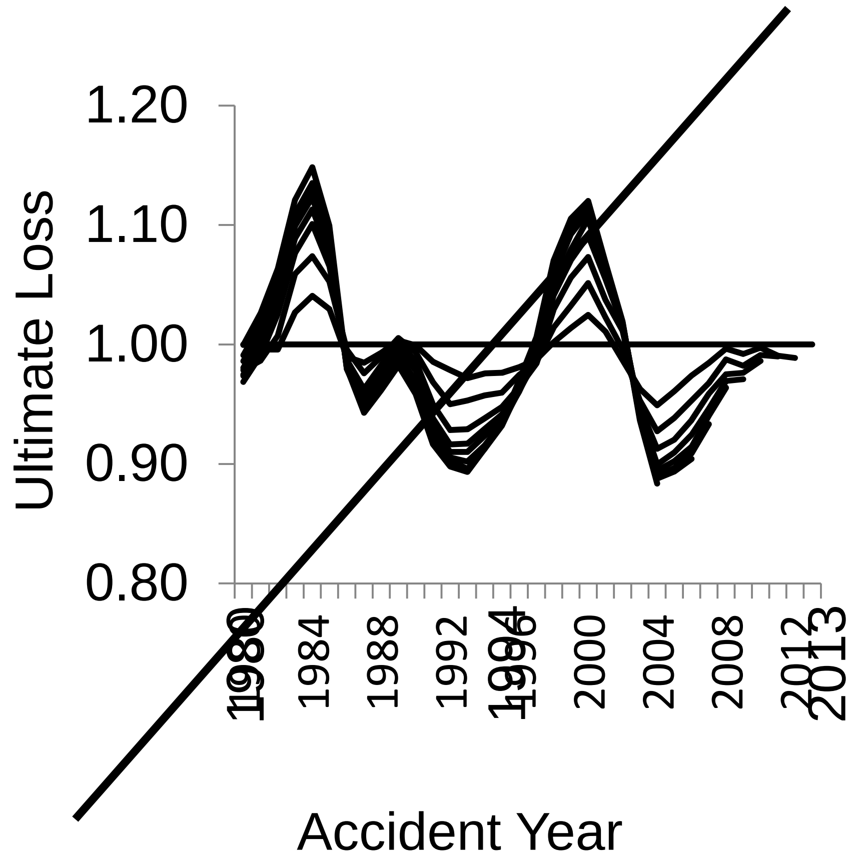




Reserving theory
Reserving practice
Ultimate Loss
1.20
1.10
1.00
0.90
0.80
1980
1994
2013
Accident Year
Ultimate Loss
1.20
1.10
1.00
0.90
0.80
1980
1984
1988
1992
1994
1996
2000
2004
2008
2012
2013
Accident Year

Reserving theory Reserving practice





Agenda

1. Do we have a problem?
2. What is it?
3. What are some solutions?

What’s holding us back?

1. We don’t forecast into the future

“Accurate prediction is very hard, especially when it involves the future.”

2. Pricing and reserving perpetuate the cycle

“Garbage in garbage out. More care needs to be taken in compiling pricing information to be used by reserving.”

3. We need a better way to incorporate judgment

What’s holding us back?

1. We don’t forecast into the future

“Accurate prediction is very hard, especially when it involves the future.”

2. Pricing and reserving perpetuate the cycle

“Garbage in garbage out. More care needs to be taken in compiling pricing information to be used by reserving.”

3. We need a better way to incorporate judgment

What’s holding us back?

1. We don’t forecast into the future

“Accurate prediction is very hard, especially when it involves the future.”

2. We need to bring in external information

“There aRE TOO MANY VARIABLES WHICH CAN RANDOMLY AFFECT OUTCOMES”

3.

What’s holding us back?

1. We don’t forecast into the future

“Accurate prediction is very hard, especially when it involves the future.”

2. We need to bring in external information

“There aRE TOO MANY VARIABLES WHICH CAN RANDOMLY AFFECT OUTCOMES”

3. We need a better way to incorporate judgment

“ [T]here is still …resistance to new methods... We convince ourselves that since we use a lot of judgement with the old methods, what's the point of using a new method when I'm going to overwrite the outcomes with my judgement anyway.”

Agenda

1. Do we have a problem?
2. What is it?
3. What are some solutions?

What’s the problem?

1. We don’t forecast into the future

GLMs

2. We need to bring in external information

GLMs with external variables

3. We need a better way to incorporate judgment

Bayesian

What’s the ~~problem~~ solution?

1. We don’t forecast into the future

GLMs

2. We need to bring in external information

GLMs with “latent variables”

3. We need a better way to incorporate judgment

Bayesian principles

What’s the ~~problem~~ solution?

1. We don’t forecast into the future

GLMs

2. We need to bring in external information

GLMs with “latent variables”

3. We need a better way to incorporate judgment

Bayesian principles

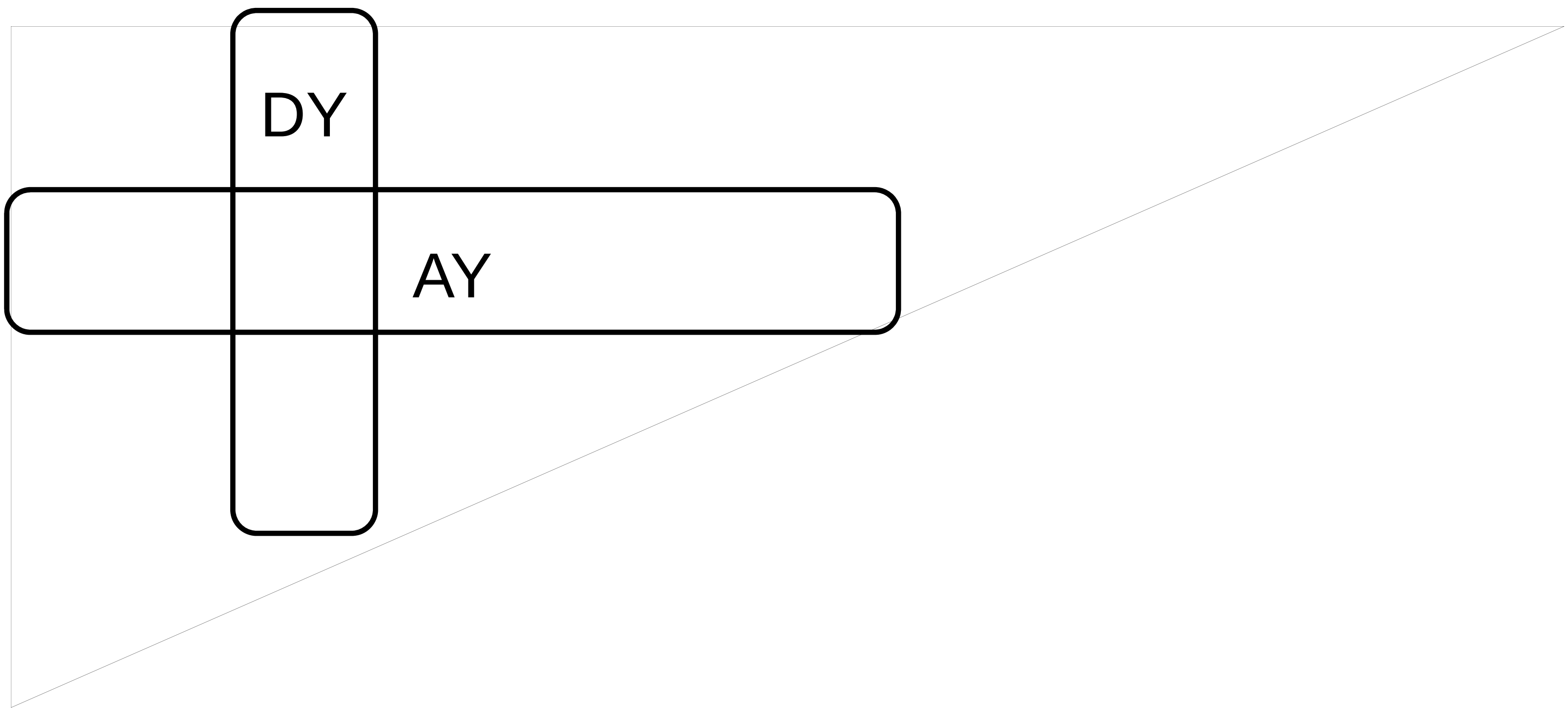
Generalized Linear Models

Jessica Leong, FIAA, FCAS

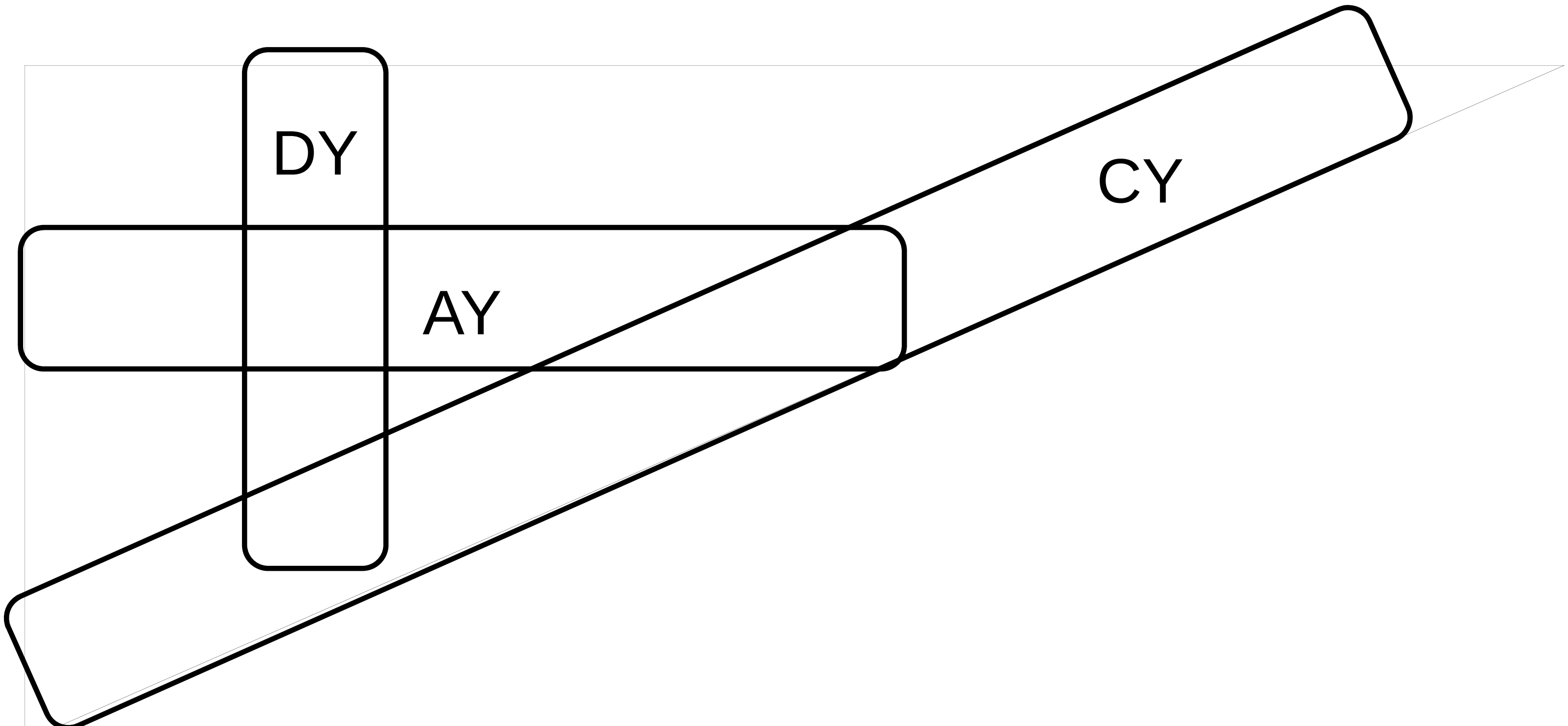
Predictive Analytics Execution Lead

Zurich Insurance NA

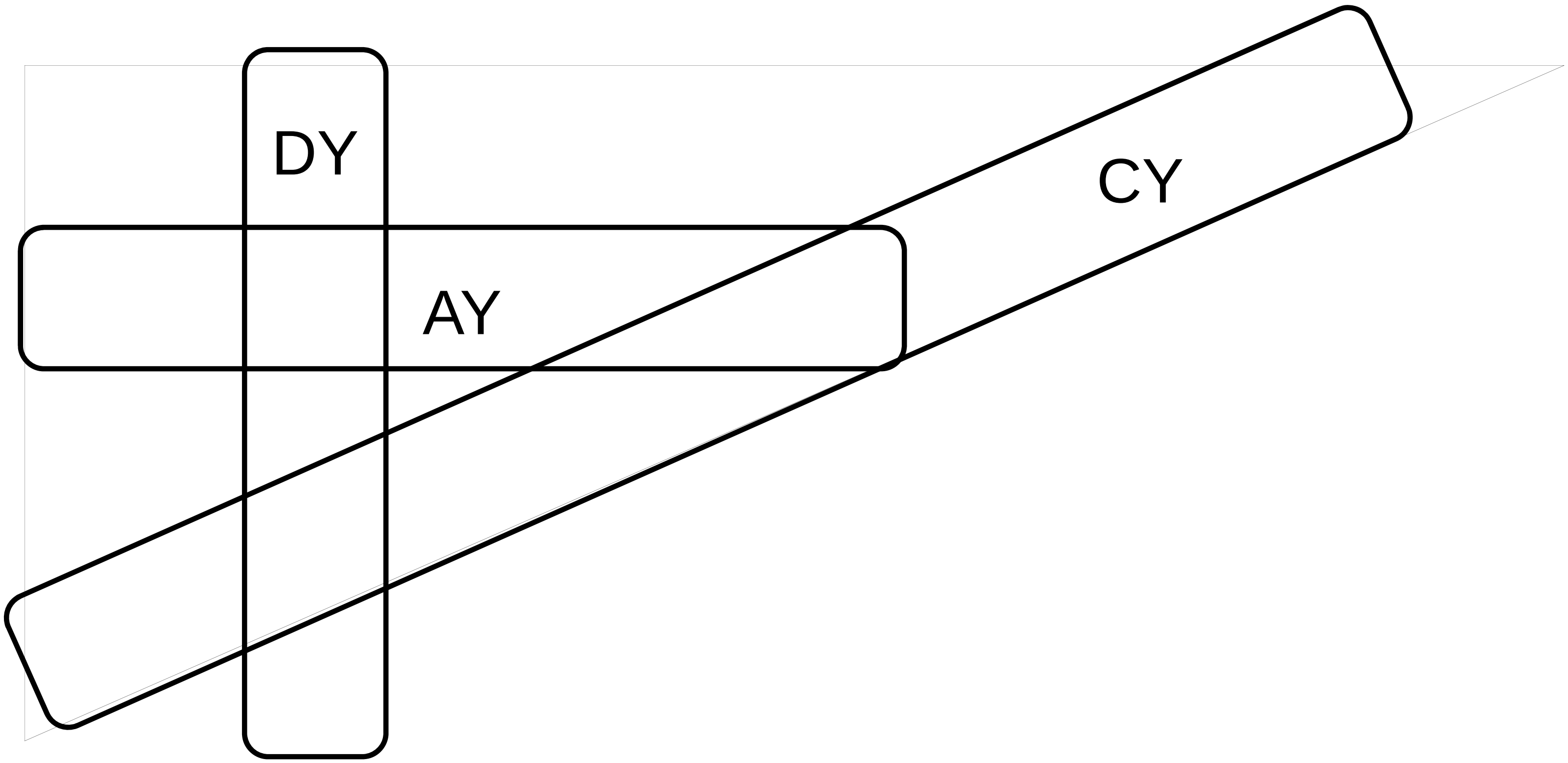
Generalized Linear Models



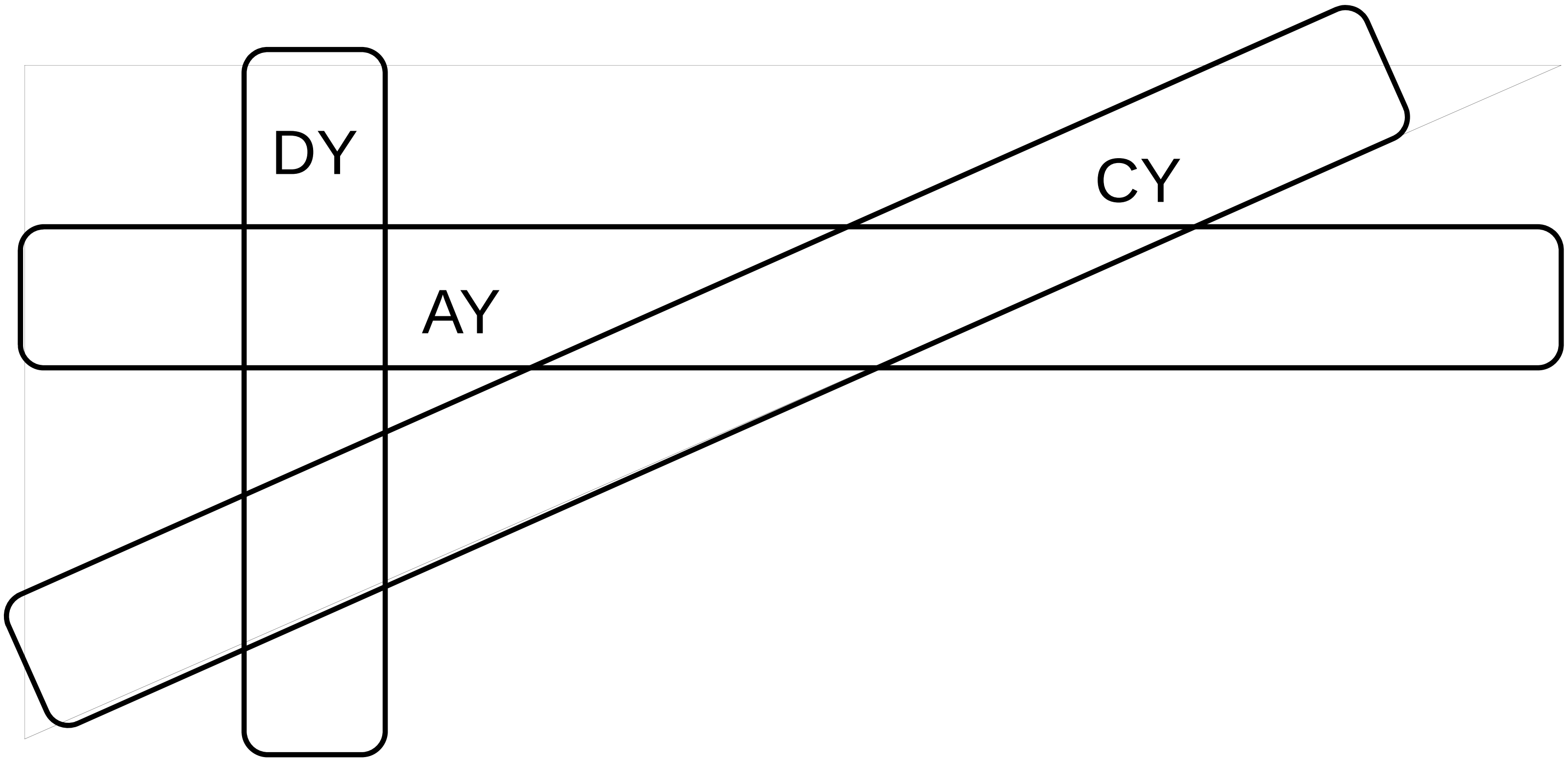
Generalized Linear Models



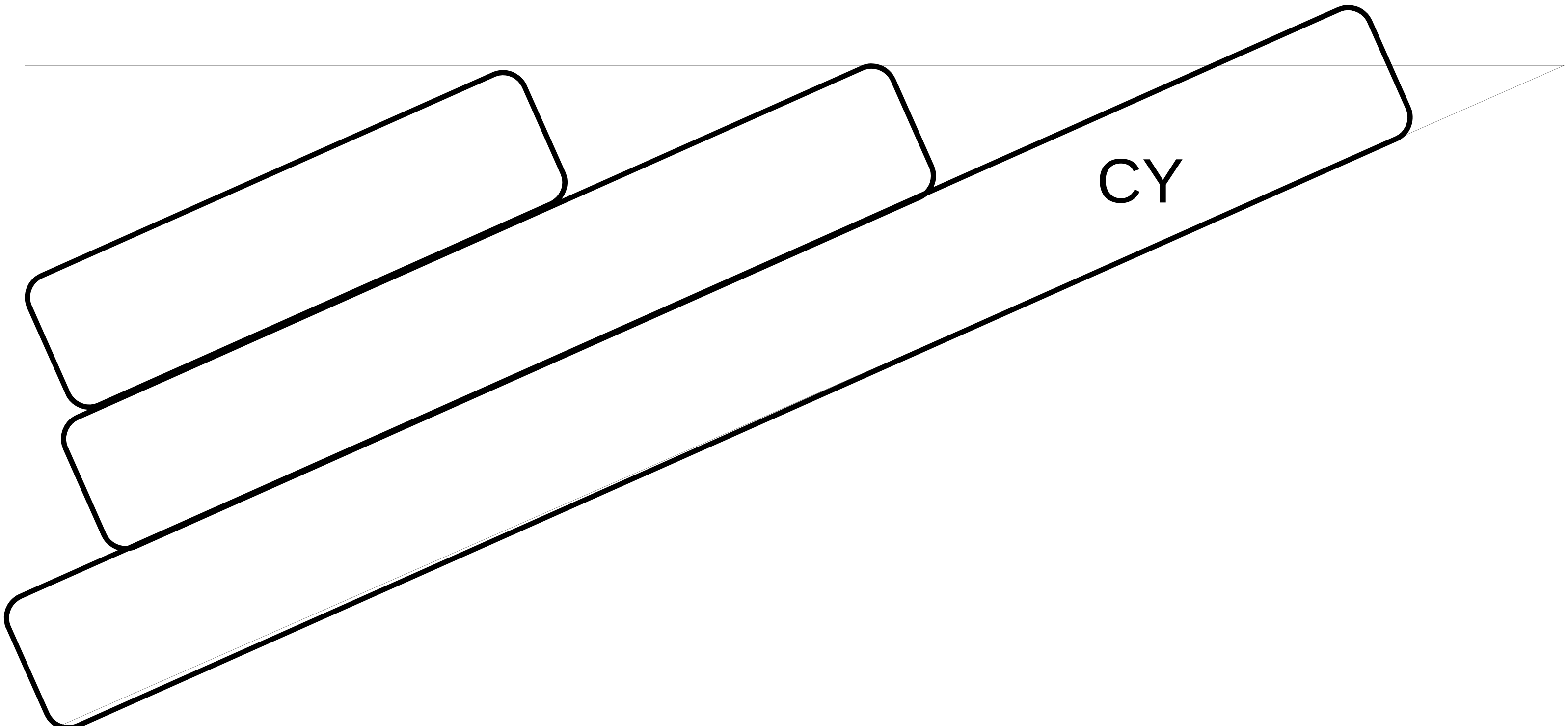
Generalized Linear Models



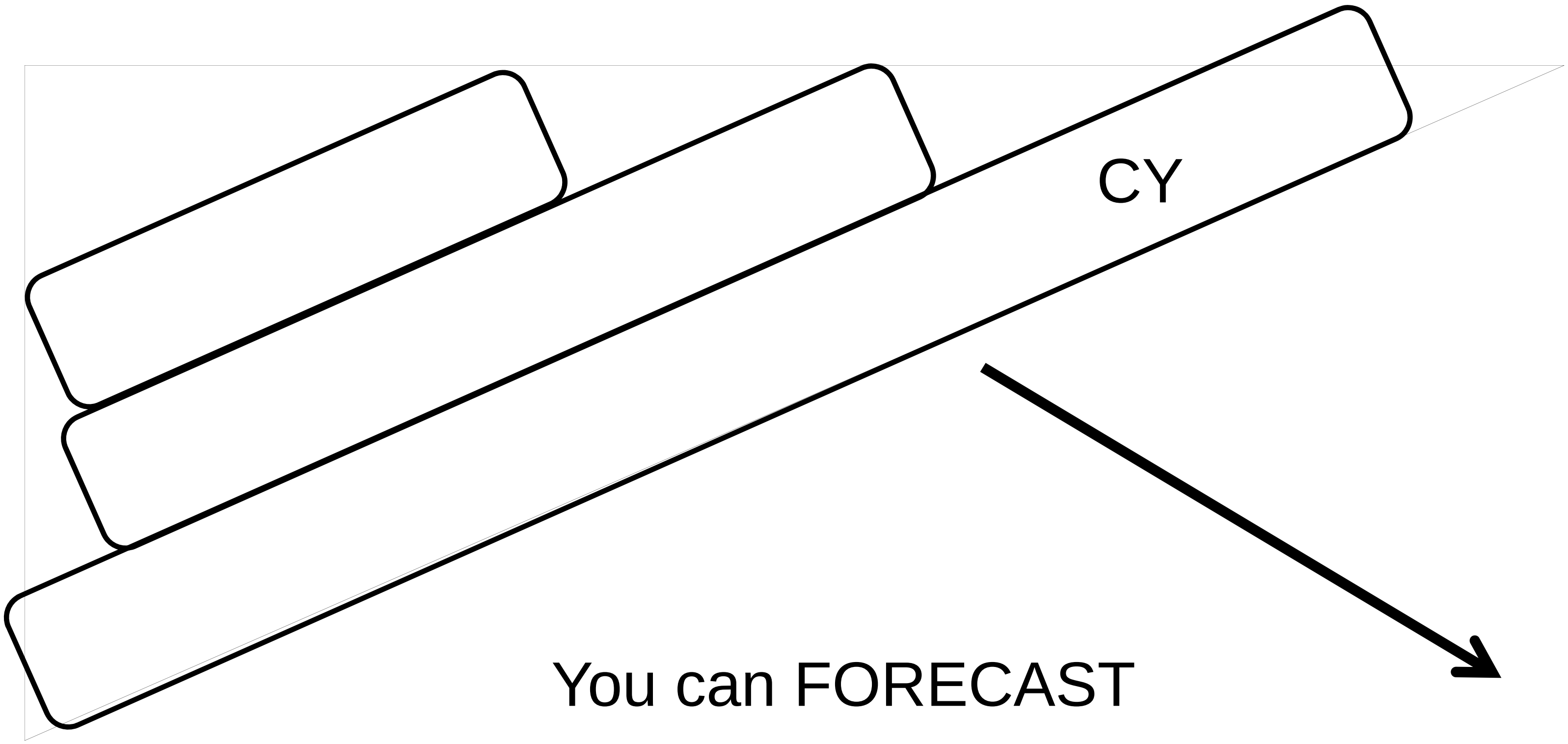
Generalized Linear Models



Generalized Linear Models



Generalized Linear Models



Scary!

Why GLM?

- The GLM = chain ladder with calendar year trends -> explicit forecasting

Why GLM?

- The GLM = chain ladder with calendar year trends -> explicit forecasting
- Understand the real driver of reserve risk – changing economic trends

Why GLM?

- The GLM = chain ladder with calendar year trends -> explicit forecasting
- Understand the real driver of reserve risk – changing economic trends
- (Some) actuaries can do it, and (some) management have heard of it

Why don’t we use GLMs for reserving?

What’s the ~~problem~~ solution?

1. We don’t forecast into the future

GLMs

2. We need to bring in external information

GLMs with “latent variables”

3. We need a better way to incorporate judgment

Bayesian principles

Generalized Linear Models with Latent Variables

Dave Clark, FCAS

Senior Actuary

Munich Re

Latent Variables

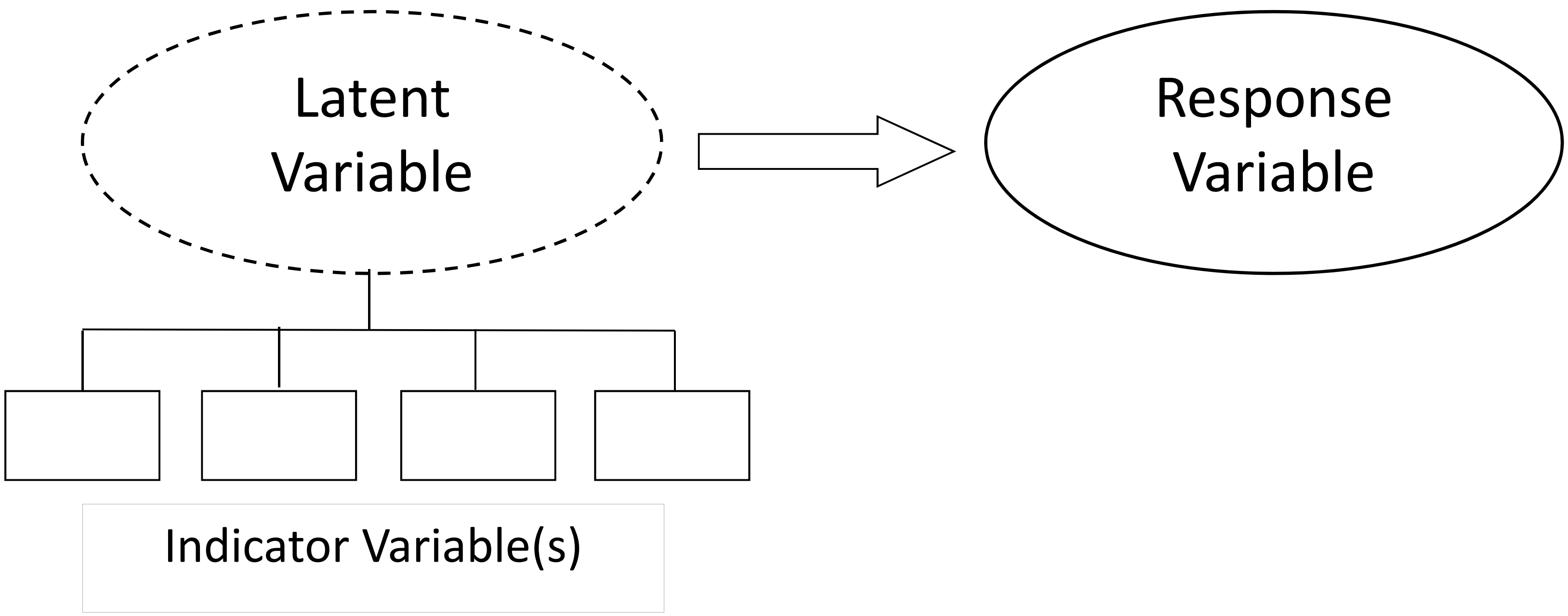
- A "Latent Variable" is one that is not directly observable, but can be approximated by a combination of other variables that can be measured, called "indicators."
- Non-Insurance examples:
  - Intelligence (does an IQ test measure this?)
  - Scholastic Aptitude
  - Job Satisfaction
  - Happiness
  - Credit-worthiness
  - Consumer Confidence

Latent Variables

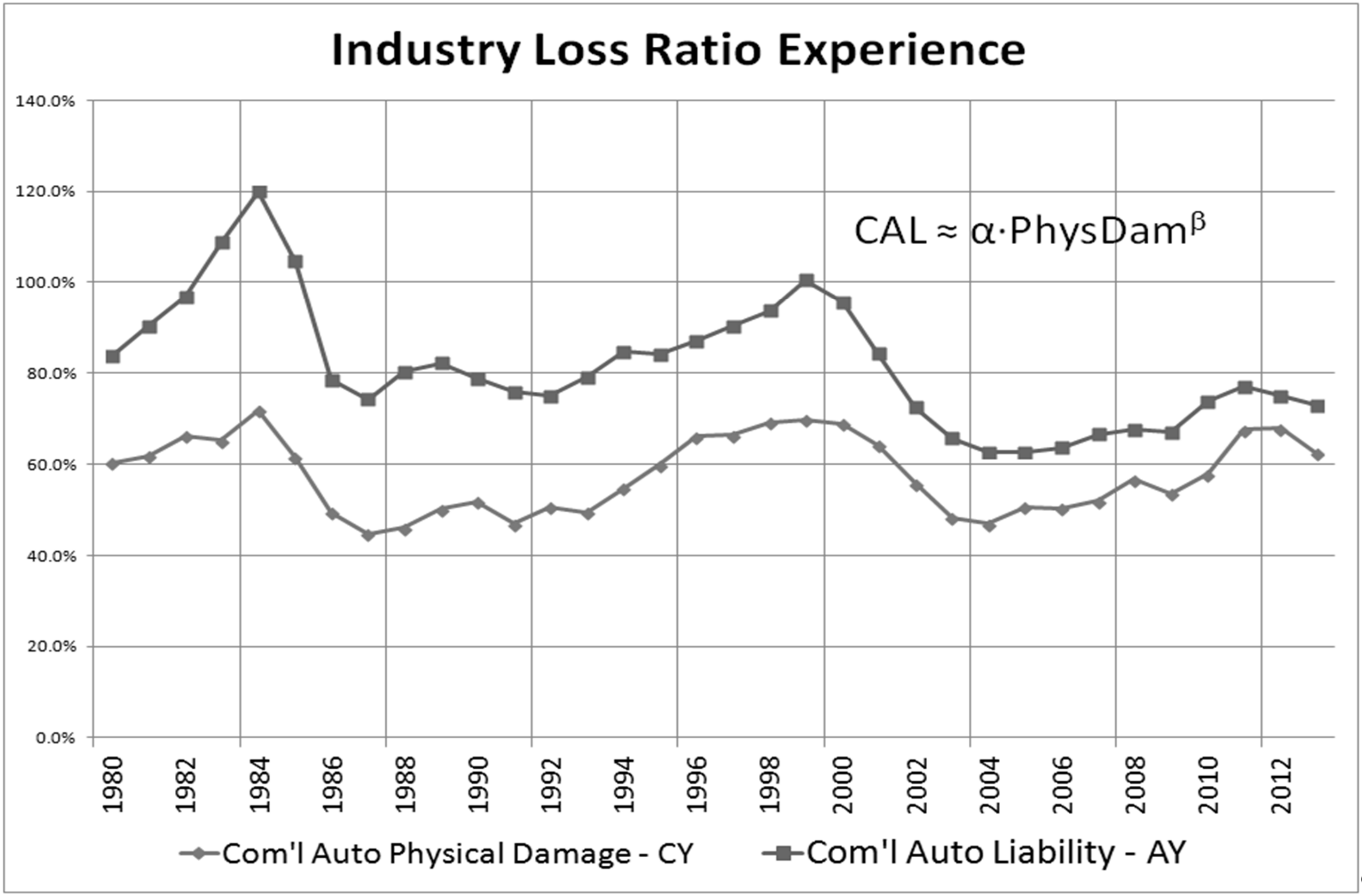
- For Insurance and Reserving:
  - "Exposure Base" - We really mean something like "propensity for loss" or some value that is proportional to expected loss.
    - Payroll, sales, miles driven, property value, etc, are indicators
  - "Market Cycle" – Is a "hard" or "soft" market directly measurable?
    - Market surveys, rate monitors, are indicators
- Key idea: We may not have true exposures for a reserve segment; we may not even have onlevel factors – but we do have indicators of these things.

Latent Variables

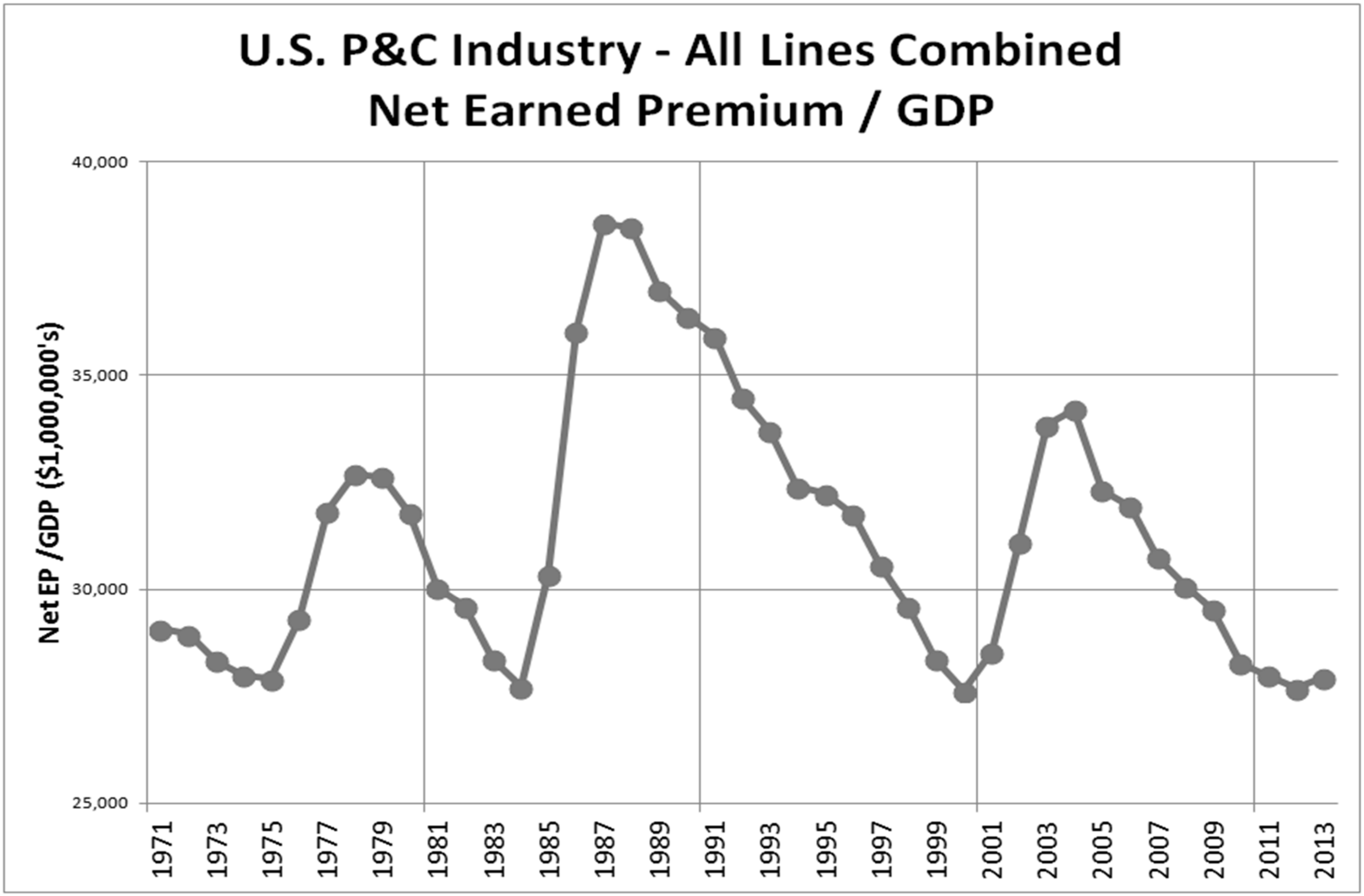
- Here is the magic:
- We do not need to have an exact historical exposure base or onlevel factors.
- We only need some indicators that are <u>correlated</u> with the onlevel factors.



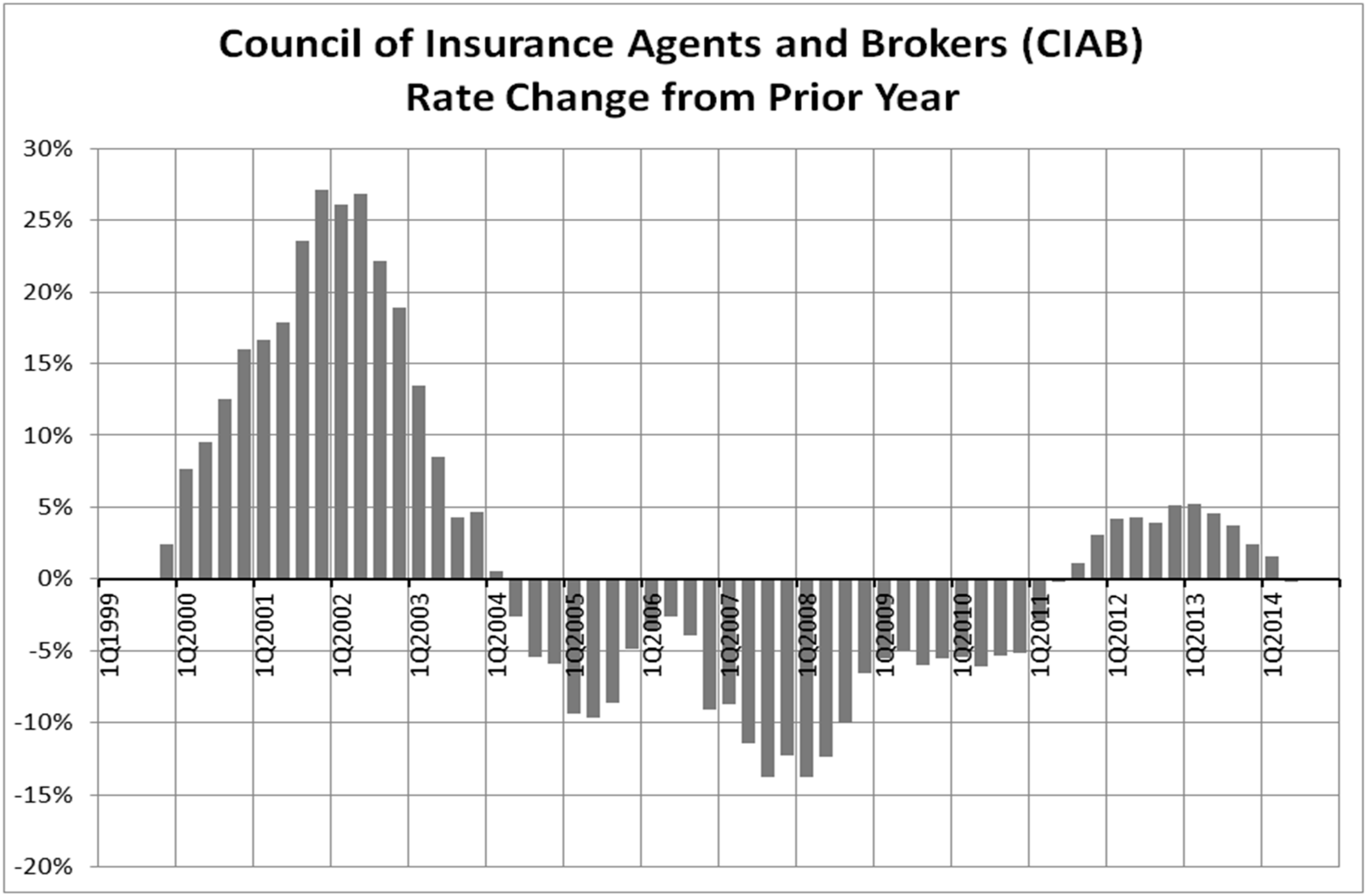
Latent Variables



Latent Variables



Latent Variables



Latent Variables

- We can consider the loss ratio by year for a given triangle to be a function of various external indices.
- The coefficients β are fit based on the data from reserve segment. This fits easily into a GLM.

$$E\left(LR_{year}\right) = \alpha \cdot \frac{CPI_{year}^{\beta 1}}{RateIndex_{year}^{\beta 2}}$$

$$= exp\left(ln(\alpha) + \beta_1 \cdot ln\left(CPI_{year}\right) + \beta_2 \cdot ln\left(RateIndex_{year}\right)\right)$$

Can use “latent variables” in reserving, like we do in pricing?

What’s the ~~problem~~ solution?

1. We don’t forecast into the future

GLMs

2. We need to bring in external information

GLMs with latent variables

3. We need a better way to incorporate judgment

Bayesian principles

Bayesian Principles

James Guszcza, FCAS
Chief Data Scientist
Deloitte

Bayesian Principles

Probably what we want

"Given any value (estimate of future payments) and our current state of knowledge, what is the probability that the final payments will be no larger than the given value?"

-- Casualty Actuarial Society (2004)
Working Party on Quantifying Variability in Reserve Estimates

I read the above passage as a request for a Bayesian predictive distribution.

Why Bayes

"Modern Bayesian methods provide richer information, with greater flexibility and broader applicability than 20th century methods. Bayesian methods are intellectually coherent and intuitive. Bayesian analyses are readily computed with modern software and hardware."
-- John Kruschke, Indiana University Psychology

- Bayesian Data Analysis [BDA] frees us from relying on "procedural" approaches to data analysis.
- Today it is practical to estimate models that are as simple or complex as the situation demands.
- Output: full probability distribution estimates of <u>all</u> quantities of interest
  - Ultimate loss ratios by accident year
  - Outstanding loss amounts
  - Missing values of any cell in a loss triangle

The Fundamental Bayesian Principle

"For Bayesians as much as for any other statistician, parameters are (typically) fixed but unknown. It is the knowledge about these unknowns that Bayesians model as random…

… typically it is the Bayesian who makes the claim for inference in a particular instance and the frequentist who restricts claims to infinite populations of replications."

-- Andrew Gelman and Christian Robert

The Fundamental Bayesian Principle

"For Bayesians as much as for any other statistician, parameters are (typically) fixed but unknown. It is the knowledge about these unknowns that Bayesians model as random…

… typically it is the Bayesian who makes the claim for inference in a particular instance and the frequentist who restricts claims to infinite populations of replications."

-- Andrew Gelman and Christian Robert

Translation:

- **Frequentist:** Probability models the infinite replications of the data $\boldsymbol{X}$
- **Bayesian:** Probability models our partial knowledge about θ

Learning From Data

Suppose Persi tosses a coin 12 times and gets 3 heads.
What is the probability of heads on the 13th toss?

Learning From Data

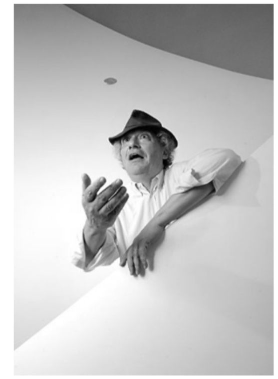
Suppose Persi tosses a coin 12 times and gets 3 heads.
What is the probability of heads on the 13th toss?

Frequentist analysis

$X_i \sim_{iid} Bern(\theta)$ ➔ $L(\theta | H = 3, n = 12) = \prod \theta^3 (1-\theta)^9$ ➔ $\hat{\theta}_{MLE} = \frac{1}{4}$

Learning From Data

Suppose Persi tosses a coin 12 times and gets 3 heads. What is the probability of heads on the 13th toss?



Frequentist analysis

$X_i \sim_{iid} Bern(\theta)$ ➔ $L(\theta|H=3, n=12) = \prod \theta^3 (1-\theta)^9$ ➔ $\hat{\theta}_{MLE} = \frac{1}{4}$

Thoughts

- "Parameter risk": 12 flips is not a lot of data ("credibility concerns")
- We've flipped other coins before… isn't that knowledge relevant?
- It would be nice to somehow "temper" the estimate of ¼ or "credibility weight" it with some other source of information
- It would be nice not to just give a point estimate and a confidence interval, but say things like: $Pr(L < \theta < U) = p$

Learning From Data

Suppose Persi tosses a coin 12 times and gets 3 heads.
What is the probability of heads on the 13th toss?

Bayesian analysis

$\theta \sim Beta(\alpha, \beta)$ ➔ $\theta \sim Beta(\alpha + 3, \beta + 9)$

Thoughts

- "Parameter risk": quantified by the posterior distribution
- Prior knowledge: encoded in the choice of $\{\alpha, \beta\}$
- Other data: maybe Persi has flipped other coins on other days… we could throw all of this (together with our current data) into a **hierarchical model**
- Mean what we say and say what we mean: $Pr(L < \theta < U) = p$ is a "**credibility interval**"… it's what most people think confidence intervals say… (but don't!)

Prior distributions: a feature, not a bug

"Your 'subjective' probability is not something fetched out of the sky on a whim; it is what your actual judgment should be, in view of your information to date and other people's information."

-- Richard Jeffrey, Princeton University

Prior distributions: a feature, not a bug

"Your 'subjective' probability is not something fetched out of the sky on a whim; it is what your actual judgment should be, in view of your information to date and other people's information."
-- Richard Jeffrey, Princeton University

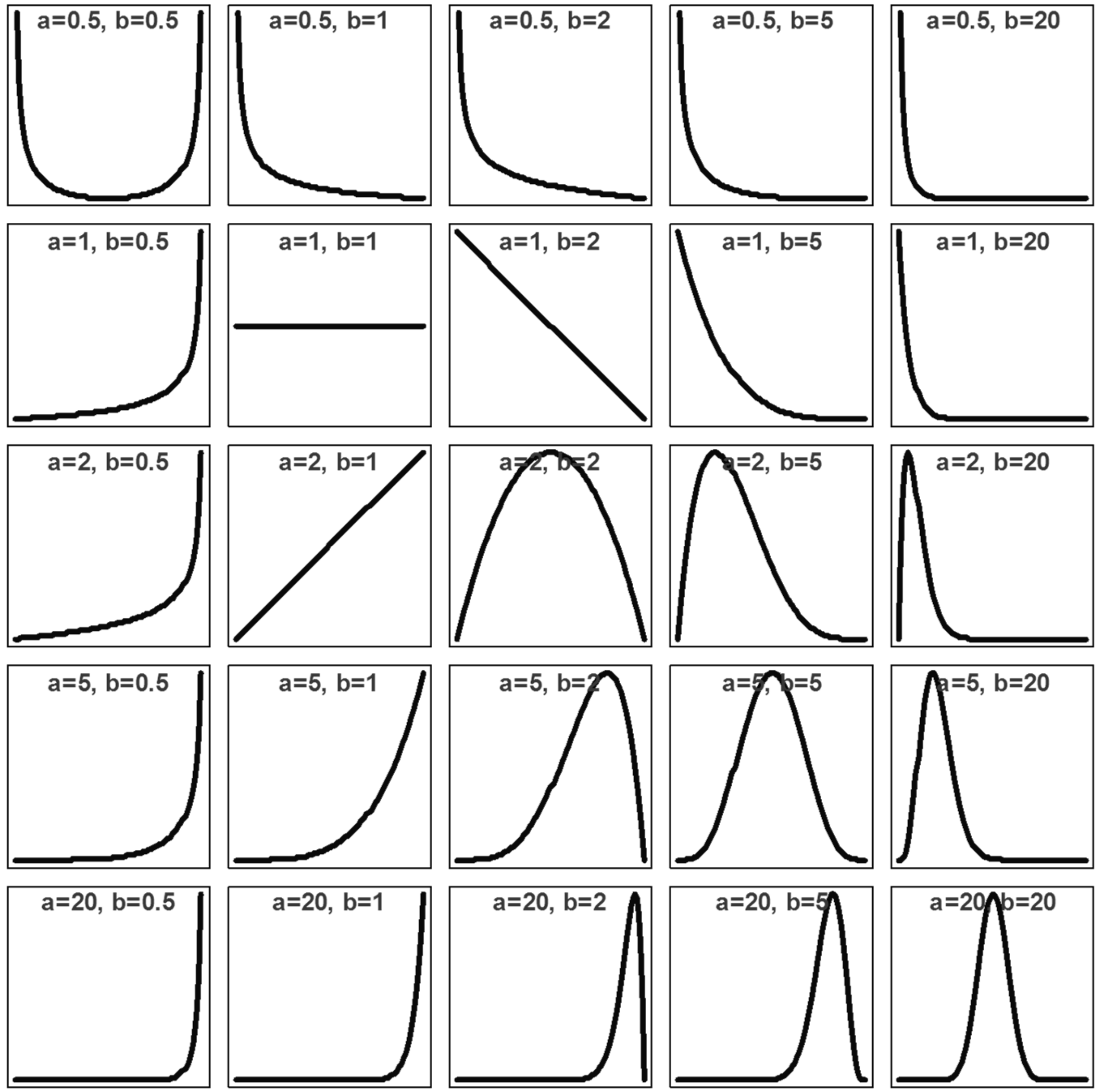
- "Subjective" probability is really "judgmental" probability

- The choice of likelihood function is also "subjective" In this sense
  - ODP (or other) distributional form
  - Inclusion of covariates
  - Trends
  - Tail factor extrapolations
  - ….

MahaBeta *(representing prior knowledge of the true probability of heads)*

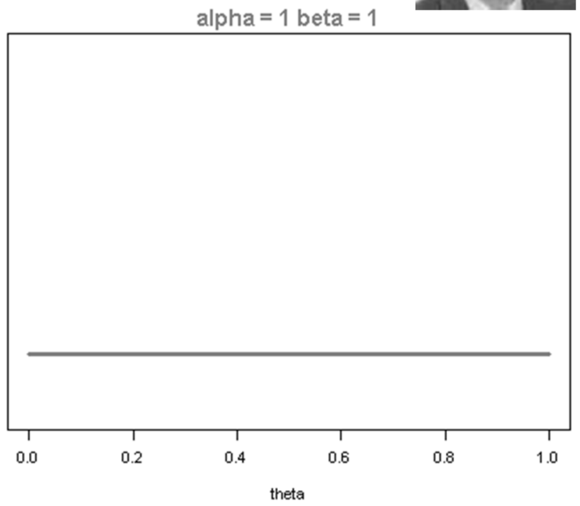
- Here is a gallery of Beta(α,β) distributions.
  - Defined on [0,1]
  - Very flexible
  - $E(\theta) = \frac{\alpha}{\alpha+\beta}$
  - $Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Think of choosing $\{\alpha, \beta\}$ as having flipped a coin $\alpha + \beta$ times and observing α heads



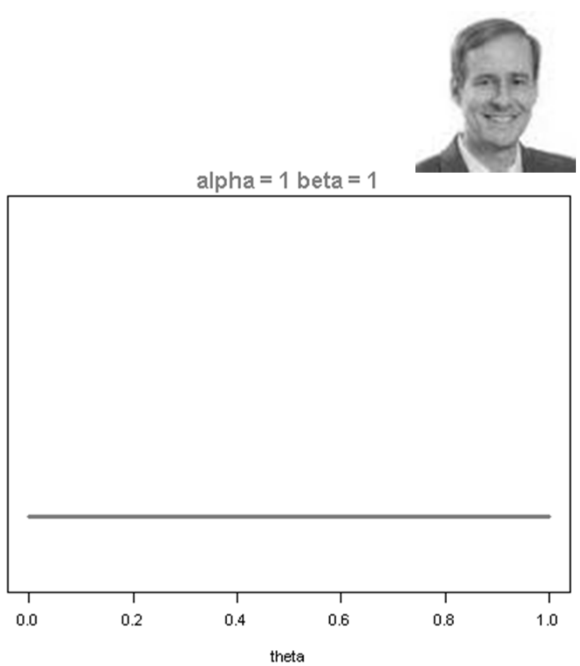
Three Scenarios

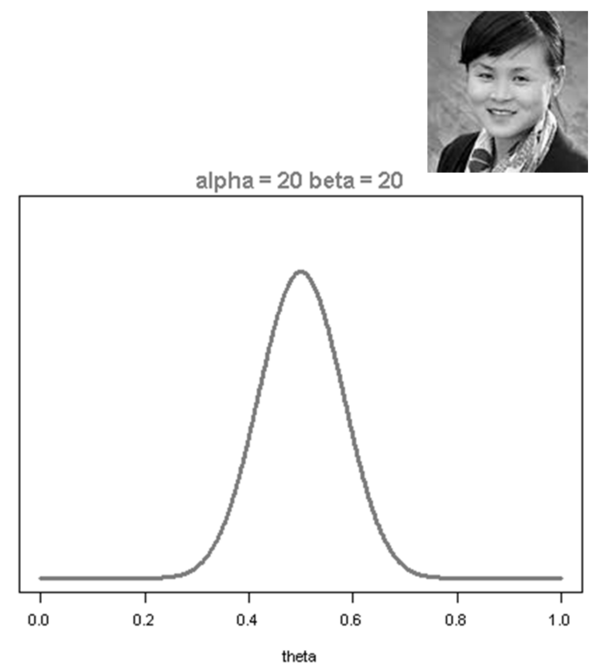
- Before Persi tossed the coin:
  - Dave “no commitments” Clark assumes that Persi’s coin is unique, so no prior knowledge about coin tosses is relevant to this case.



Three Scenarios

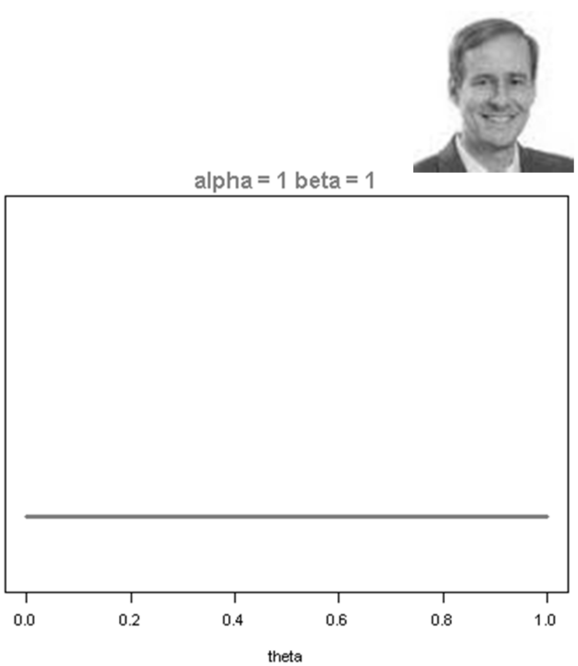
- Before Persi tossed the coin:
  - Dave “no commitments” Clark assumes that Persi’s coin is unique, so no prior knowledge about coin tosses is relevant to this case.
  - Jessica “realist” Leong thinks it’s fairly likely the true probability of heads $\approx {}^{1}/_{2}$

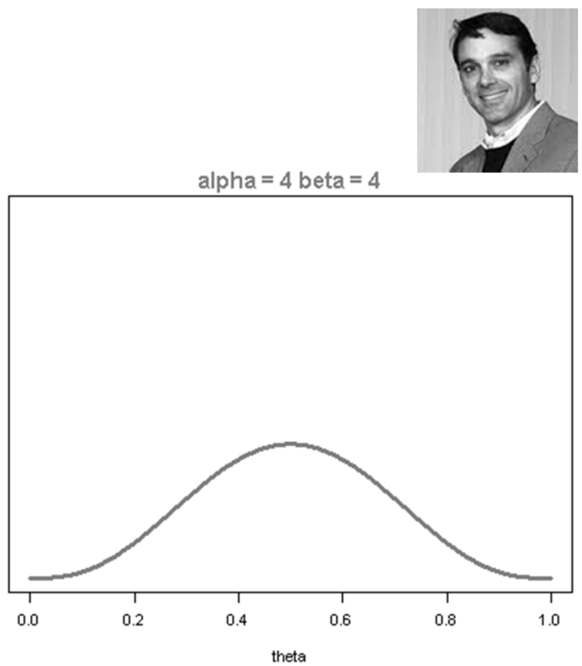


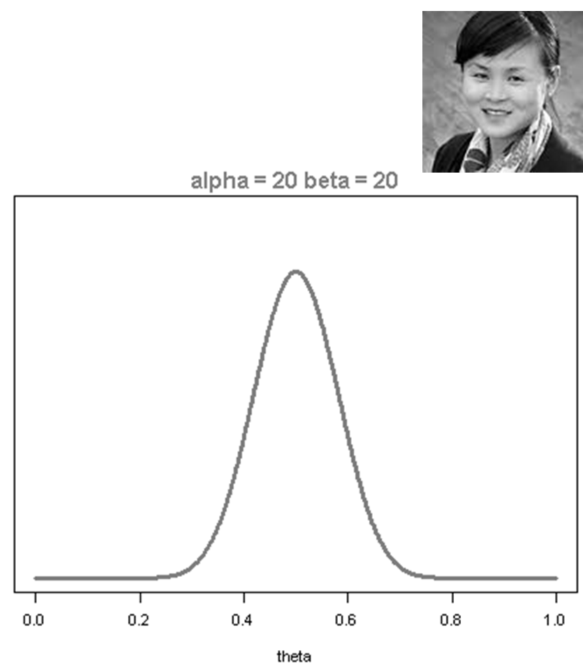


Three Scenarios

- Before Persi tossed the coin:
  - Dave “no commitments” Clark assumes that Persi’s coin is unique, so no prior knowledge about coin tosses is relevant to this case.
  - Jessica “realist” Leong thinks it’s fairly likely the true probability of heads $\approx\ ^1/_2$
  - Don “middle way” Mango takes an intermediate position.

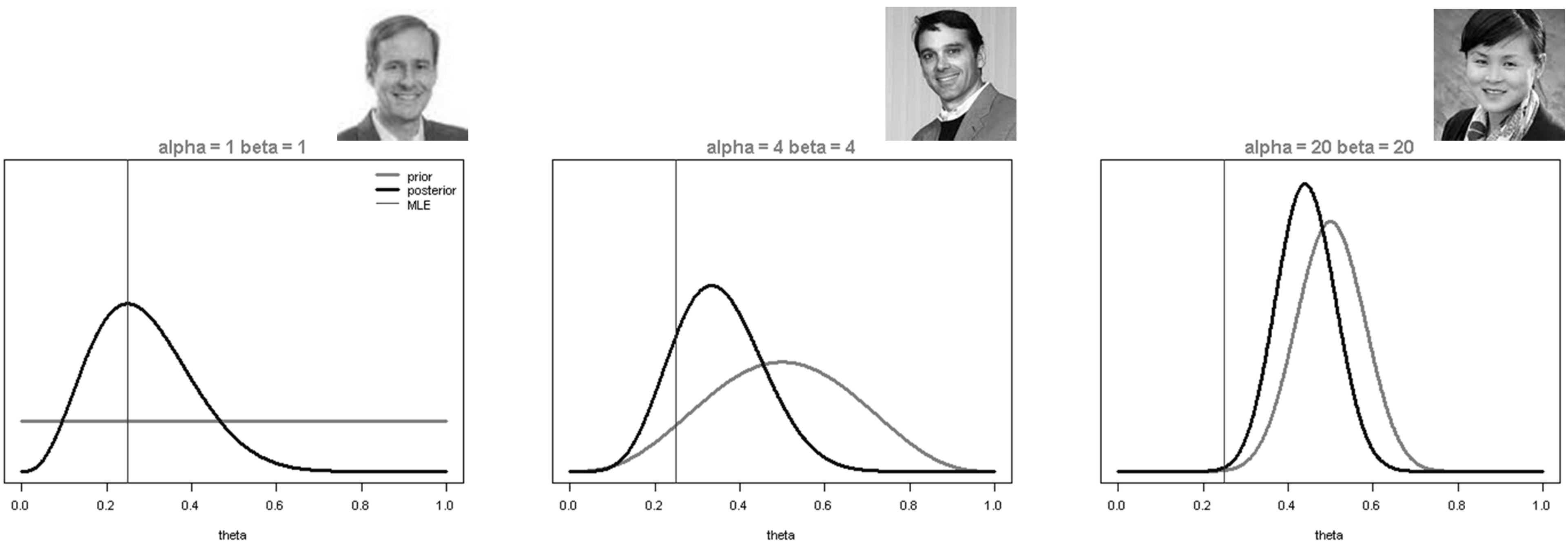






Three Scenarios

- Dave’s predictions about future tosses is determined entirely by the data.
- Jessica’s predictions are much less influenced by the data



In more familiar terms…

- Before Persi tossed the coin: $E(\theta) = \frac{\alpha}{\alpha+\beta}$

- After 3 heads in 12 tosses: $E(\theta|h = 3, t = 9) = \frac{\alpha+3}{\alpha+\beta+12}$

$$= Z\left(\frac{3}{12}\right) + (1 - Z)\left(\frac{\alpha}{\alpha+\beta}\right) \quad where \quad Z = \frac{12}{\alpha+\beta+12}$$

Choosing the $\{\alpha, \beta\}$ parameters is analogous to incorporating prior knowledge of ultimate loss ratio (or other parameters) in a loss reserving model.

Bayesian Computation

And here's the thing: MCMC makes it practical

Why Isn't Everyone a Bayesian?

Before 1990: this sort of thing was often viewed as a parlor trick because of the need to analytically solve high-dimensional integrals:

$$f(Y \mid X) = \int f(Y \mid \theta) f(\theta \mid X) d\theta = \int f(Y \mid \theta) \left(\frac{f(X \mid \theta)\pi(\theta)}{\int f(X \mid \theta)\pi(\theta) d\theta} \right) d\theta$$

And here's the thing: MCMC makes it practical

Before 1990: this sort of thing was often viewed as a parlor trick because of the need to analytically solve high-dimensional integrals:

$$f(Y \mid X) = \int f(Y \mid \theta) f(\theta \mid X) d\theta = \int f(Y \mid \theta) \left(\frac{f(X \mid \theta)\pi(\theta)}{\int f(X \mid \theta)\pi(\theta) d\theta} \right) d\theta$$

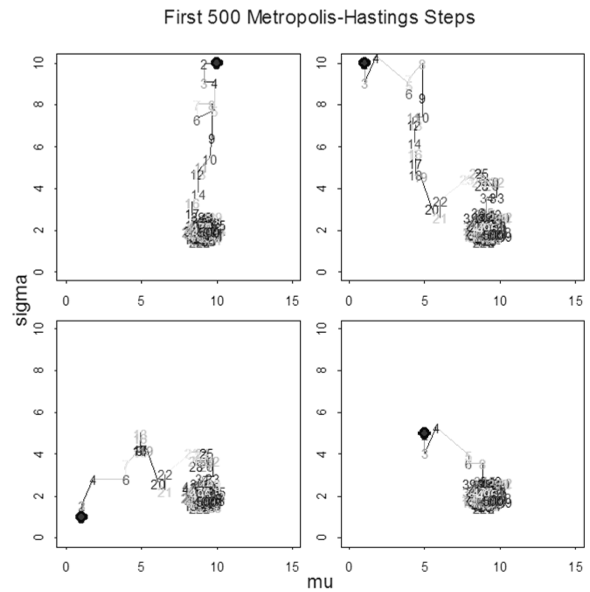
Why Isn't Everyone a Bayesian?

B. EFRON*

Originally a talk delivered at a conference on Bayesian statistics, this article attempts to answer the following question: why is most scientific data analysis carried out in a non-Bayesian framework? The argument consists mainly of some practical examples of data analysis, in which the Bayesian approach is difficult but Fisherian/frequentist solutions are relatively easy. There is a brief discussion of objectivity in statistical analyses and of the difficulties of achieving objectivity within a Bayesian framework. The article ends with a list of practical advantages of Fisherian/frequentist methods, which so far seem to have outweighed the philosophical superiority of Bayesianism.

And here's the thing: MCMC makes it practical

After 1990: MCMC introduces a "new world order":
Now we can <u>simulate</u> Bayesian posteriors.



Sampling-Based Approaches to Calculating Marginal Densities

ALAN E. GELFAND AND ADRIAN F. M. SMITH*

Journal of the American Statistical Association
June 1990, Vol. 85, No. 410, Theory and Methods

Bayesian Loss Reserving

Methodology: Sophisticated Simplicity

*"It is fruitful to start simply and complicate if necessary. That is, it is recommended that an initial, **sophisticatedly simple** model be formulated and tested in terms of explaining past data and in forecasting or predicting new data. If the model is successful… it can be put into use. If not, [it] can be modified or elaborated to improve performance…"*

-- Arnold Zellner, The University of Chicago

Methodology: Sophisticated Simplicity

"It is fruitful to start simply and complicate if necessary. That is, it is recommended that an initial, ***sophisticatedly simple*** *model be formulated and tested in terms of explaining past data and in forecasting or predicting new data. If the model is successful… it can be put into use. If not, [it] can be modified or elaborated to improve performance…"*

-- Arnold Zellner, The University of Chicago

This is precisely what Bayesian Data Analysis enables us to do.

Methodology: Sophisticated Simplicity

*"It is fruitful to start simply and complicate if necessary. That is, it is recommended that an initial, **sophisticatedly simple** model be formulated and tested in terms of explaining past data and in forecasting or predicting new data. If the model is successful… it can be put into use. If not, [it] can be modified or elaborated to improve performance…"*

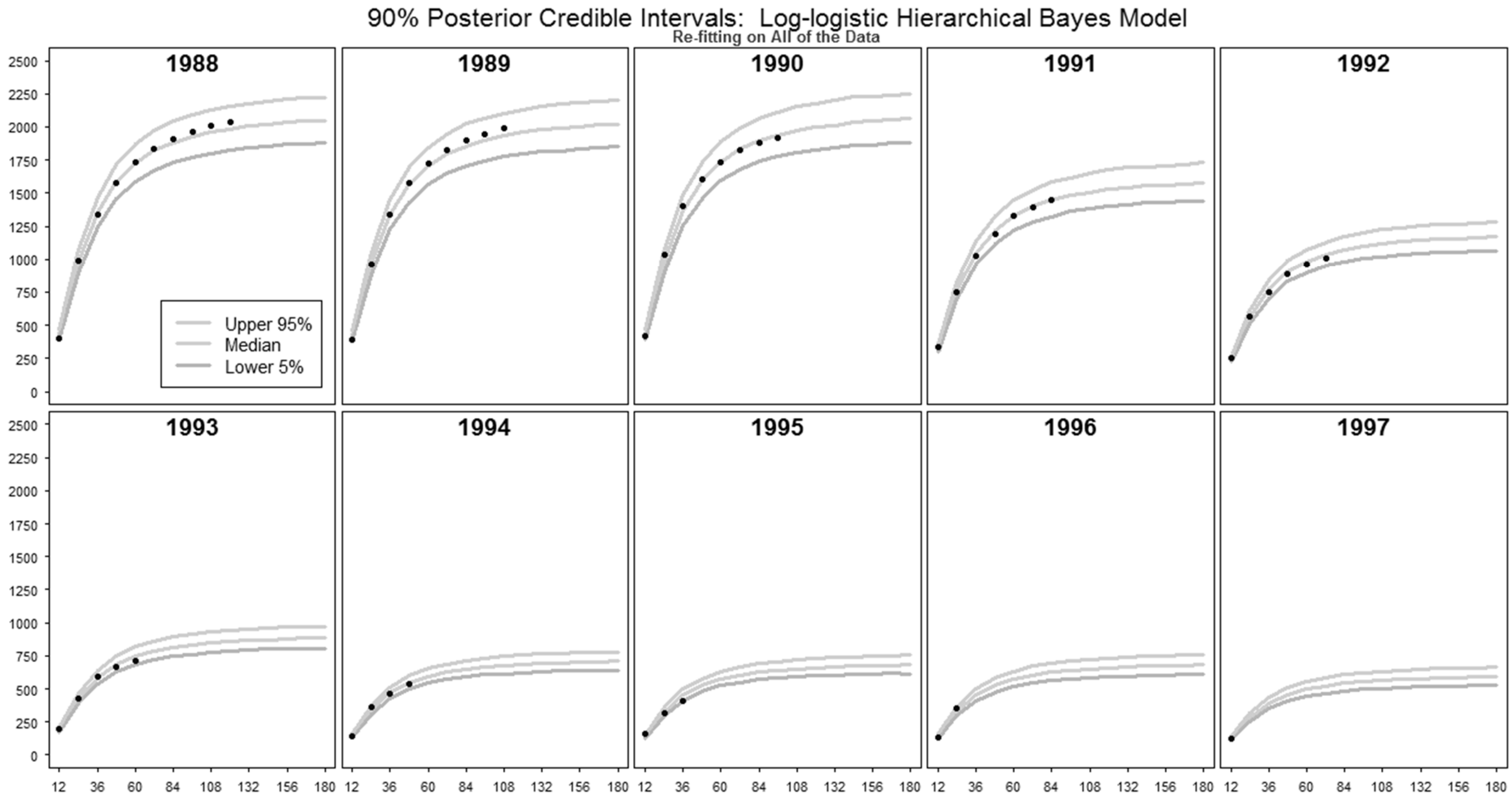
-- Arnold Zellner, The University of Chicago

Start with ODP (if that's what you like) and then add structure to account for:

- Other distributional forms (what's so sacred about GLM or exponential family?)
- Negative incremental incurred losses
- Nonlinear structure (e.g. growth curves)
- Hierarchical structure (e.g. fitting multiple lines, companies, regions)
- Prior knowledge
- Other loss triangles ("complement of credibility")
- CY and AY trends
- Autocorrelation
- …

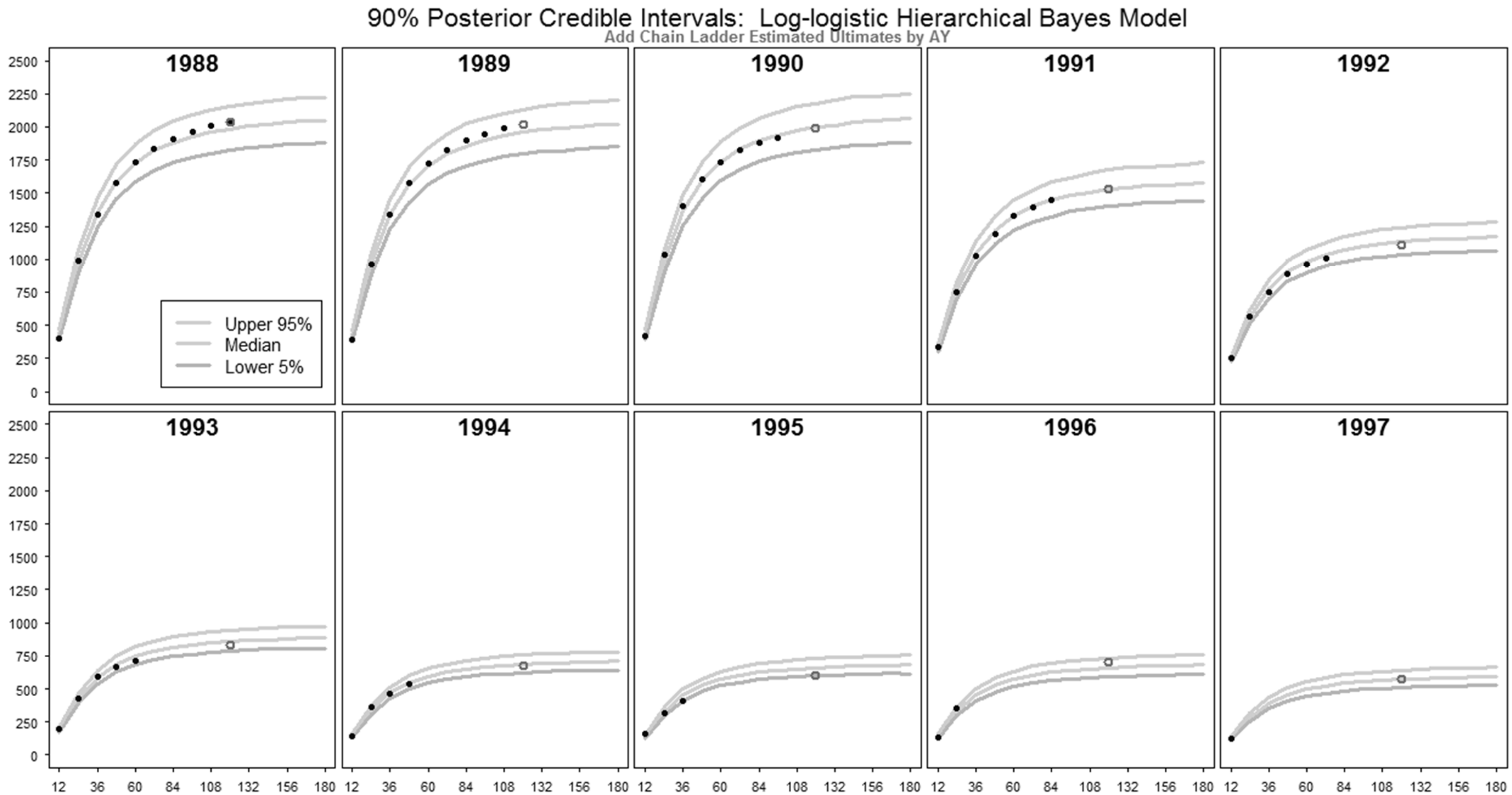
Example (with Wayne Zhang and Vanja Dukic)

- Posterior credible intervals of incremental losses – by accident year
  - Based on non-linear hierarchical growth curve model



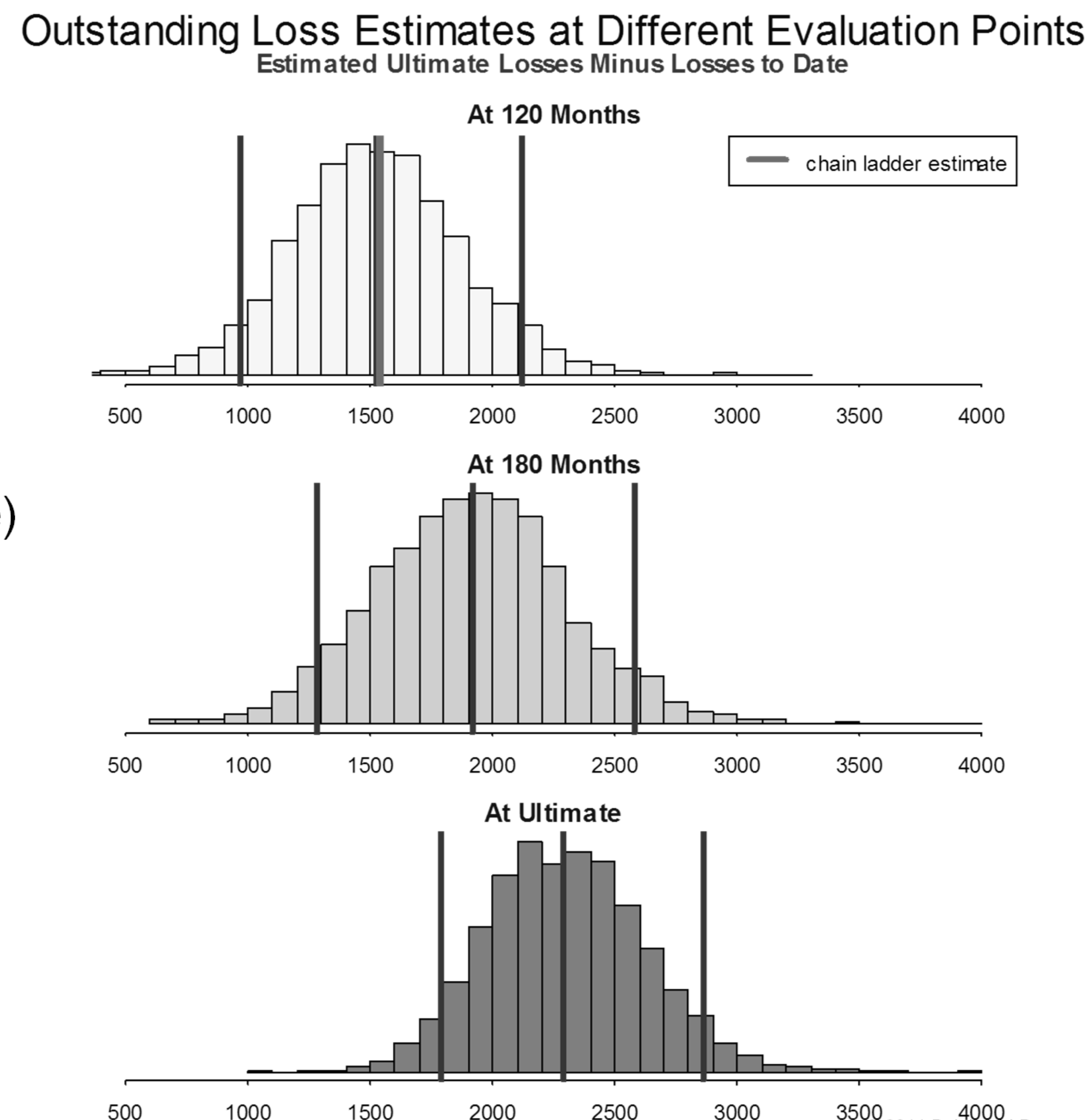
Example (with Wayne Zhang and Vanja Dukic)

- Posterior credible intervals of incremental losses – by accident year
  - Based on non-linear hierarchical growth curve model



Posterior distribution of aggregate outstanding losses

- Non-informative priors were used
- A full posterior distribution falls out of the analysis
  - No need for ad hoc simulations, settling for a point estimate with a confidence interval
- Use of non-linear (growth curve) model enables us to project beyond the range of the data
  - Choice of growth curves affects the estimates more than the choice of priors!
  - This choice "does the work of" a choice of tail factors



A Centennial Thought

Centennial thought:
Whither our profession's Bayesian heritage?

"Practically all methods of statistical estimation… are based on… the assumption that any and all collateral information or a priori knowledge is worthless. It appears to be only in the actuarial field that there has been an organized revolt against discarding all prior knowledge when an estimate is to be made using newly acquired data."

-- Arthur Bailey (1950)

Centennial thought:
Whither our profession's Bayesian heritage?

"Practically all methods of statistical estimation... are based on... the assumption that any and all collateral information or a priori knowledge is worthless. It appears to be only in the actuarial field that there has been an organized revolt against discarding all prior knowledge when an estimate is to be made using newly acquired data."

-- Arthur Bailey (1950)

... And today, in the age of MCMC, cheap computing, and open-source software...

Centennial thought: Whither our profession's Bayesian heritage?

"Practically all methods of statistical estimation... are based on... the assumption that any and all collateral information or a priori knowledge is worthless. It appears to be only in the actuarial field that there has been an organized revolt against discarding all prior knowledge when an estimate is to be made using newly acquired data."

-- Arthur Bailey (1950)

... And today, in the age of MCMC, cheap computing, and open-source software...

"Scientific disciplines from astronomy to zoology are moving to Bayesian data analysis. We should be leaders of the move, not followers."

-- John Kruschke, Indiana University Psychology (2010)

Can we use Bayesian techniques for reserving?

What’s the ~~problem~~ solution?

1. We don’t forecast into the future
GLMs

2. We need to bring in external information
GLMs with latent variables

3. We need a better way to incorporate judgment
Bayesian principles

Agenda

1. Do we have a problem?
2. What is it?
3. What are some solutions?