"Borrowing Strength"
Hierarchical Models in Actuarial Work

ASA – Chicago
January 24, 2012

Jim Guszcza, PhD, FCAS, MAAA
Deloitte Consulting
University of Wisconsin-Madison

Agenda

Context: Actuarial Science, Data Science, and Business Analytics

Actuarial Background: Minimum Bias and Credibility Theory

Multilevel/Hierarchical Modeling Concepts

Ratemaking Case Study

Loss Reserving 101

NLME Loss Reserving Model

Nonlinear Hierarchical Bayes Loss Reserving Model (Zhang-Dukic-Guszcza)

Context
Business Analytics, Data Science, Actuarial Science

How old is Business Analytics?

"Business Analytics" is Now Part of the Culture

From last week's Larry Summers op-ed in the NYT about future trends in education:

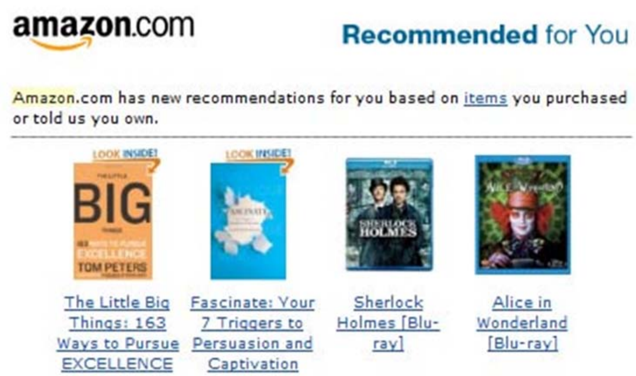
6. Courses of study will place much more emphasis on the analysis of data. Gen. George Marshall famously told a Princeton commencement audience that it was impossible to think seriously about the future of postwar Europe without giving close attention to Thucydides on the Peloponnesian War. Of course, we'll always learn from history. But the capacity for analysis beyond simple reflection has greatly increased (consider Gen. David Petraeus's reliance on social science in preparing the army's counterinsurgency manual).

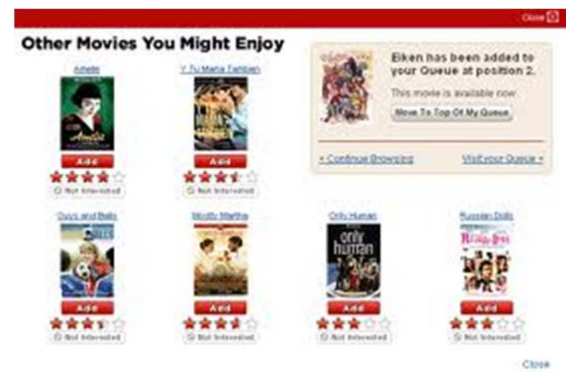
As the "Moneyball" story aptly displays in the world of baseball, the marshalling of data to test presumptions and locate paths to success is transforming almost every aspect of human life. It is not possible to make judgments about one's own medical care without some understanding of probability, and certainly the financial crisis speaks to the consequences of the failure to appreciate "black swan events" and their significance. In an earlier era, when many people were involved in surveying land, it made sense to require that almost every student entering a top college know something of trigonometry. Today, a basic grounding in probability statistics and decision analysis makes far more sense.

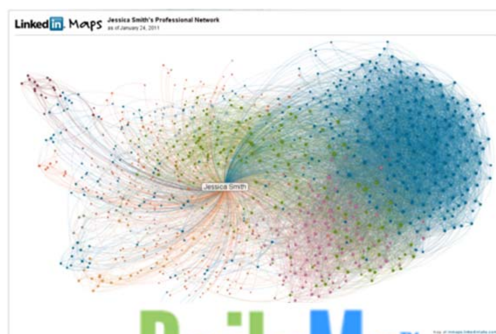
Business Analytics is Ubiquitous

- **Today, The analysis of data affects:**

- What we buy
- What we read
- What we watch
- How we network
- How we socialize
- The opinions we form
- Whom we date and marry





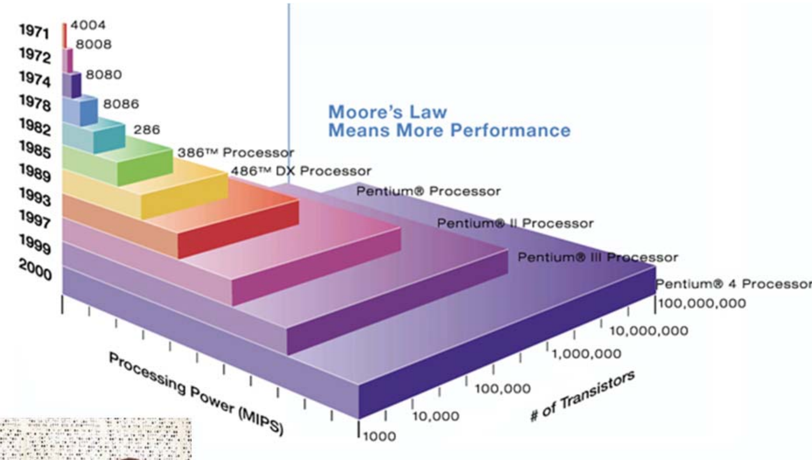


DailyMe™
Better News. Better You.™

eHarmony®
Love Begins Here™

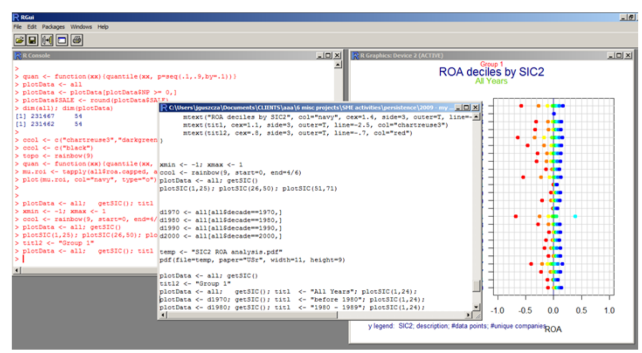
Why Now (Moore, Moore, Moore)

- Technology
  - Cost of storage and computing power has decreased exponentially



- Data
  - Big data: data exhaust captured from purchase behavior, RFID tags, GPS, internet surfing, unstructured text… is growing exponentially.
  - Companies are learning to do more with their internal data

- Software and algorithms
  - Ideas keep coming from statistics, economics, machine learning, marketing, …
  - Open-source analytics tools (eg R)



The Rise of Data Science

Or: "The Collision between Statistics and Computation"

- The skill set underlying business analytics is increasingly called **data science**.
- This definition conveys why data science goes beyond
  - traditional statistics
  - business intelligence [BI]
  - Information technology

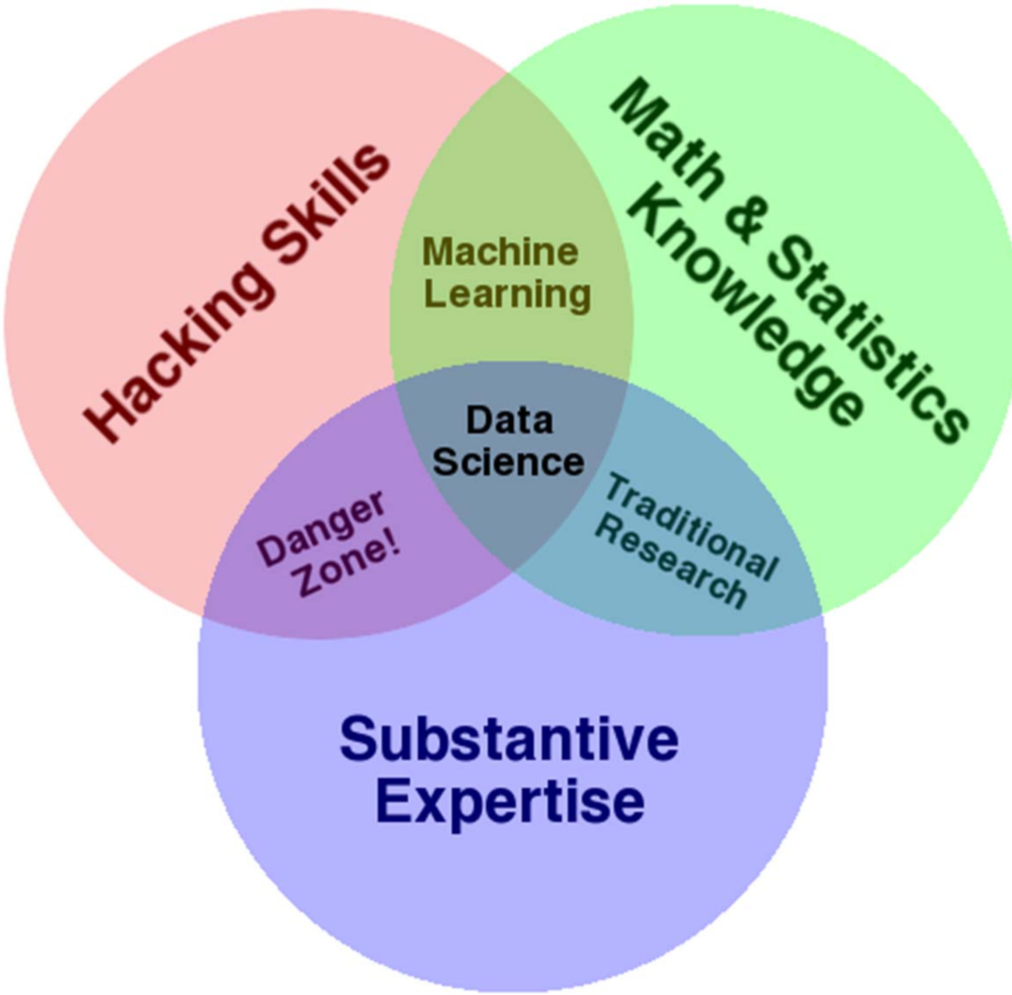


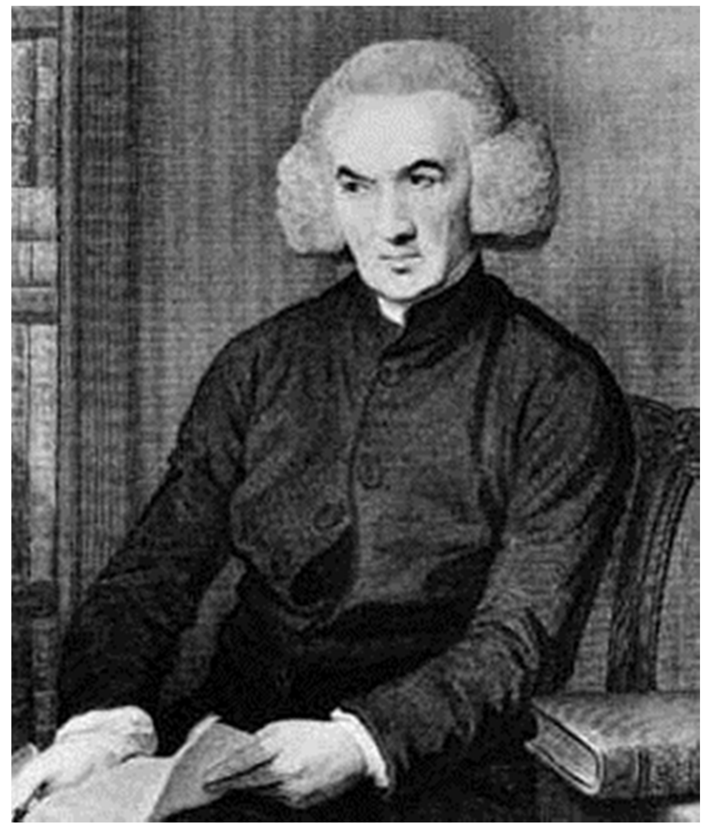
Image borrowed from Drew Conway's blog
http://www.dataists.com/2010/09/the-data-science-venn-diagram

How Old is Business Analytics?

- Richard Price was an 18th century Unitarian minister, moral philosopher, political pamphleteer, …
  - Associated with such figures as Thomas Jefferson, Benjamin Franklin, Thomas Paine, and David Hume

- Best known for presenting Thomas Bayes' original paper to the British Royal society.

- He was in effect also the world's first actuary:
  - Wrote treatises on pricing annuities and life insurance
  - Consulted for the Equitable in London
  - A 250 year-old example of "business analytics"



Actuarial Background

Minimum Bias and Generalized Linear Models
Credibility Theory and Bayesian Statistics
Multilevel/Hierarchical Models

Interplay Between Actuarial Science and Statistics

- Actuarial Science sometimes anticipates developments in mainstream statistics.

- Two examples come from General Insurance.
  - (aka non-life, aka property-casualty)

- Exhibit A: Bailey's Minimum Bias Method (circa 1960) anticipated the aspects of Generalized Linear Models (circa 1970).

- Exhibit B: Credibility theory injected core ideas of Bayesian statistics into actuarial practice many decades before the modern "Bayesian renaissance".

- The modern theory of Multilevel/Hierarchical models enables a practical unification of these separate cornerstones of General Insurance actuarial practice.

Exhibit A

Multivariate Ratemaking
From “Methods” to “Models”

Bailey's Minimum Bias Method

- In the early 1960s Robert Bailey and LeRoy Simon proposed a method for simultaneously deriving relativities for the many dimensions of an insurance rating plan.
  - E.g. older drivers should be charged less
  - Claim-free drivers should also be charged less
  - But deriving rate "relativities" via 1-way analyses leads to under/over-charging

- Given: r_{ij} = observed relative loss costs for the risks in cell (i,j)

- Goal: determine a set of multiplicative rating factors that approximate $\{r_{ij}\}$ as well as possible
  - Derive $\{a_1, \ldots, a_m, b_1, \ldots, b_n, \ldots\}$ where: $\hat{r}_{ij} = a_i b_j$

- Solution: iteratively compute these quantities until they converge

$$a_i = \frac{\sum_j w_{ij} r_{ij}}{\sum_j w_{ij} b_j} \quad , \quad b_j = \frac{\sum_i w_{ij} r_{ij}}{\sum_i w_{ij} a_i}$$

  - w_{ij} = Volume of exposures in cell (i,j)
  - Motivated by a "balance principle"

Bailey’s Minimum Bias Method

- We can recover the Bailey-Simon multiplicative model by maximizing Poisson likelihood.

$$L(a_i, b_j \mid r_{ij}, w_{ij}) = \prod_{i,j} \frac{e^{-a_i b_j \lambda_0 w_{ij}} \left(a_i b_j \lambda_0 w_{ij}\right)^{r_{ij}\lambda_0 w_{ij}}}{\left(r_{ij}\lambda_0 w_{ij}\right)!}$$

  - $\lambda_0 \equiv$ loss rate of the base cell
  - ➔ $r_{ij}\lambda_0 w_{ij}$ loss cost of the cell (i,j)

- Bailey’s multiplicative model is a purely algebraic approach to Poisson regression.

- Appeared a dozen years before Nelder and Wedderburn’s original paper on Generalized Linear Models (1972).

- Today GLM is an part of mainstream actuaries’ toolkits.

Exhibit B

Credibility Theory and (Bayesian) Hierarchical Models

Credibility Theory: Early

- Late 18th Century: Thomas Bayes and Pierre-Simon Laplace formulate the principles of "inverse probability"
  - Probabilistic inference from data to model parameters
  - Bayes' intellectual executor, Richard Price, became perhaps the world's first consulting actuary (Equitable Life Assurance company, London)
  - Price's – and perhaps Bayes' – thinking was influenced by the publication of David Hume's *Treatise on Human Nature* (1740)

- 1918: A. W. Whitney "The Theory of Experience Rating".
  - Advocated combining the claims experience of a single risk with that of a cohort (class, portfolio, …) of similar risks.

$$\overline{\mu} = Z \cdot \hat{\mu}_{risk} + (1 - Z) \cdot \hat{\mu}_{class}$$

  - Estimated pure premium should be a weighted average of the individual risk's claim experience with that of the cohort.
  - *Z*: a "credibility factor"
  - Z is of the form:

$$Z = \frac{w}{w + k}$$

  - Where *w* is a measure of volume, such as units of exposure or dollars of premium
  - Whitney suggested that *k* must be judgmentally determined

Credibility Theory: Early-Modern

- 1950: Arthur Bailey publishes “Credibility Procedures: Laplace’s Generalization of Bayes’ Rule and the Combination of Collateral Knowledge with Observed Data”.
  - “At present, practically all methods of statistical estimation appearing in textbooks… are based on an equivalent to the assumption that any and all collateral information or *a priori* knowledge is worthless. There have been rare instances of rebellion against this philosophy by practical statisticians who have insisted that they actually had a considerable store of knowledge apart from the specific observations being analyzed… However 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.”

- Bailey foreshadowed the Hans Bühlmann’s subsequent work.

- Bailey also quotes philosophical comments of Richard Price and Bertrand Russell on making inferences from available data.

- Further illustration that credibility theory was intimately connected with Bayesian methodology from its earliest days.

Credibility Theory: Mid-Century Modern

- 1967: Bühlmann's "greatest accuracy" Bayes credibility model.

- Assume:
  - Let X_{ij} denote dollars of loss associated with risk *i* at time *j*.
  - Assume $X_1, \ldots, X_m$ are iid, conditional on a parameter (vector) θ
  - Let $m(\theta_i)$ denote "risk premium": $m(\theta_i) \equiv E[X_{ij} | \theta_i]$

- Bühlmann minimizes mean squared errors: $E\left[m(\theta_i) - a - b\overline{X}_i\right]^2$

- … to arrive at an estimator for $m(\theta_i)$: $z_i \cdot \overline{X}_i + (1 - z_i) \cdot \mu$

- … where:

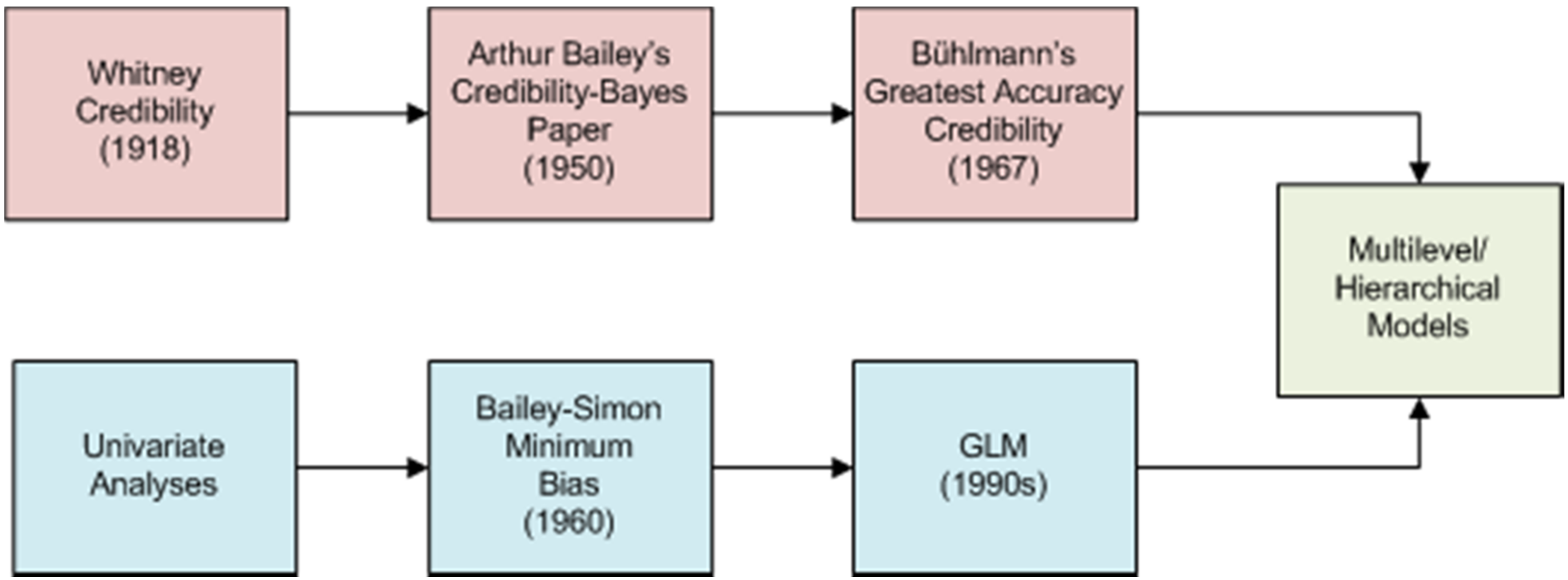
$$z_i = \frac{n_i}{n_i + k} \quad , \quad k = \frac{E\left[Var\left(X_{ij} | \theta_i\right)\right]}{Var\left(m\left(\theta_i\right)\right)}$$

- The within/between variances in *k* are estimated from the data

Summary View of the Historical Context

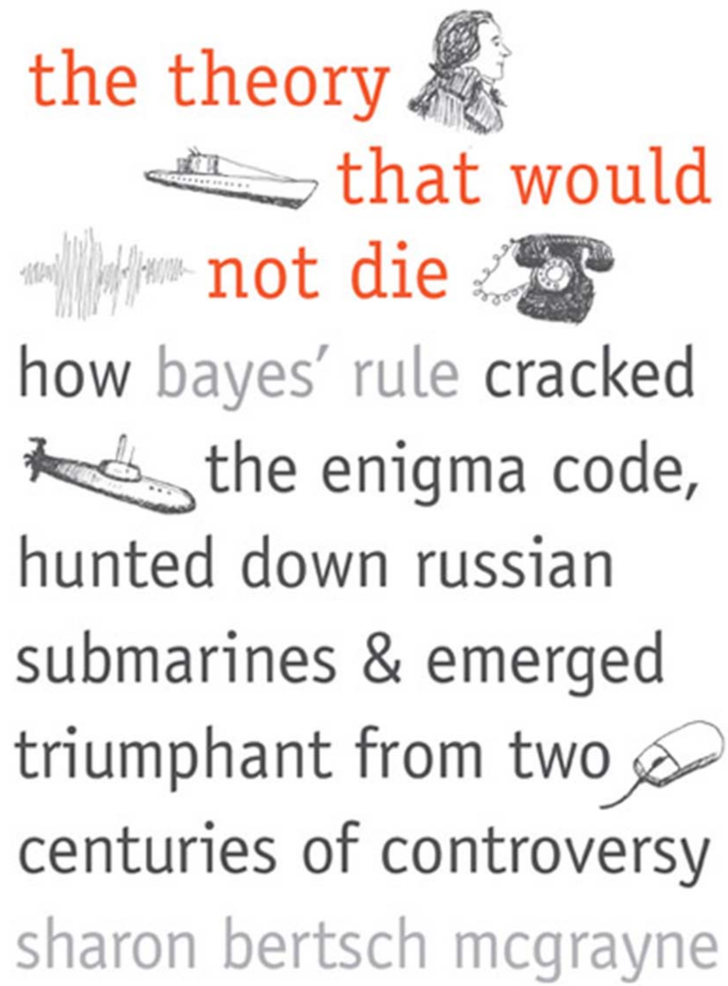
- Multilevel/Hierarchical models can be viewed as unifying two strands of the development of actuarial science.

- Note: Multilevel / Hierarchical models can be either empirical Bayes or fully Bayesian.
    - Today's focus is on the empirical Bayes case.



Advertisement

- McGrayne tells many of the “hidden histories” in the development of Bayesian statistics.
- Credibility theory was motivated by a crisis sparked by the first workers comp legislation… no data implied the need to “borrow strength”.
- Arthur Bailey’s Bayesian conversion predated Jimmy Savage’s!



Hierarchical Modeling Background

Hierarchical Data Structures
Hierarchical Models
Motivating Example

What is Hierarchical Modeling?

- Hierarchical modeling is used when one's data is *grouped* in some important way.
    - Claim experience by state or territory
    - Workers Comp claim experience by class code
    - Claim severity by injury type
    - Churn rate by agency
    - Income by profession
    - Test results by school/district/state
    - Repeated observations of the growth of a soybean plant
    - Multiple observations of cohorts of claims over time

- Often grouped data is modeled either by:
    - Building separate models by group
    - Pooling the data and introducing dummy variables to reflect the groups

- Hierarchical modeling offers a "middle way".
    - Parameters reflecting group membership enter one's model through appropriately specified *probability sub-models*.

What's in a Name?

- Hierarchical models go by many different names
  - Mixed effects models
  - Random effects models
  - Multilevel models
  - Longitudinal models
  - Panel data models

- The "hierarchical model" language is advantageous in that it evokes the way models-within-models are used to reflect levels-within-levels of ones data.

- Important special case of hierarchical models: multiple observations through time of each unit.
  - Here group membership is the individual that the repeated observations belong to.
  - Time is the covariate.

Common Hierarchical Models

- Notation:
  - Data points $(\boldsymbol{X}_i, Y_i)_{i=1\ldots N}$
  - $j[i]$: data point i belongs to group j.

- Classical Linear Model $\quad \boldsymbol{Y_i = \alpha + \beta X_i + \varepsilon_i}$
  - Equivalently: $Y_i \sim N(\alpha + \beta X_i, \sigma^2)$
  - Same α and β for every data point

- Random Intercept Model $\quad \boldsymbol{Y_i = \alpha_{j[i]} + \beta X_i + \varepsilon_i}$
  - Where $\alpha_j \sim N(\mu_\alpha, \sigma^2{}_\alpha)$ & $\varepsilon_i \sim N(0, \sigma^2)$
  - Same β for every data point; but α varies by group

- Random Intercept and Slope Model $\quad \boldsymbol{Y_i = \alpha_{j[i]} + \beta_{j[i]} X_i + \varepsilon_i}$
  - Where $(\alpha_j, \beta_j) \sim N(\mathrm{M}, \Sigma)$ & $\varepsilon_i \sim N(0, \sigma^2)$
  - Both α and β vary by group

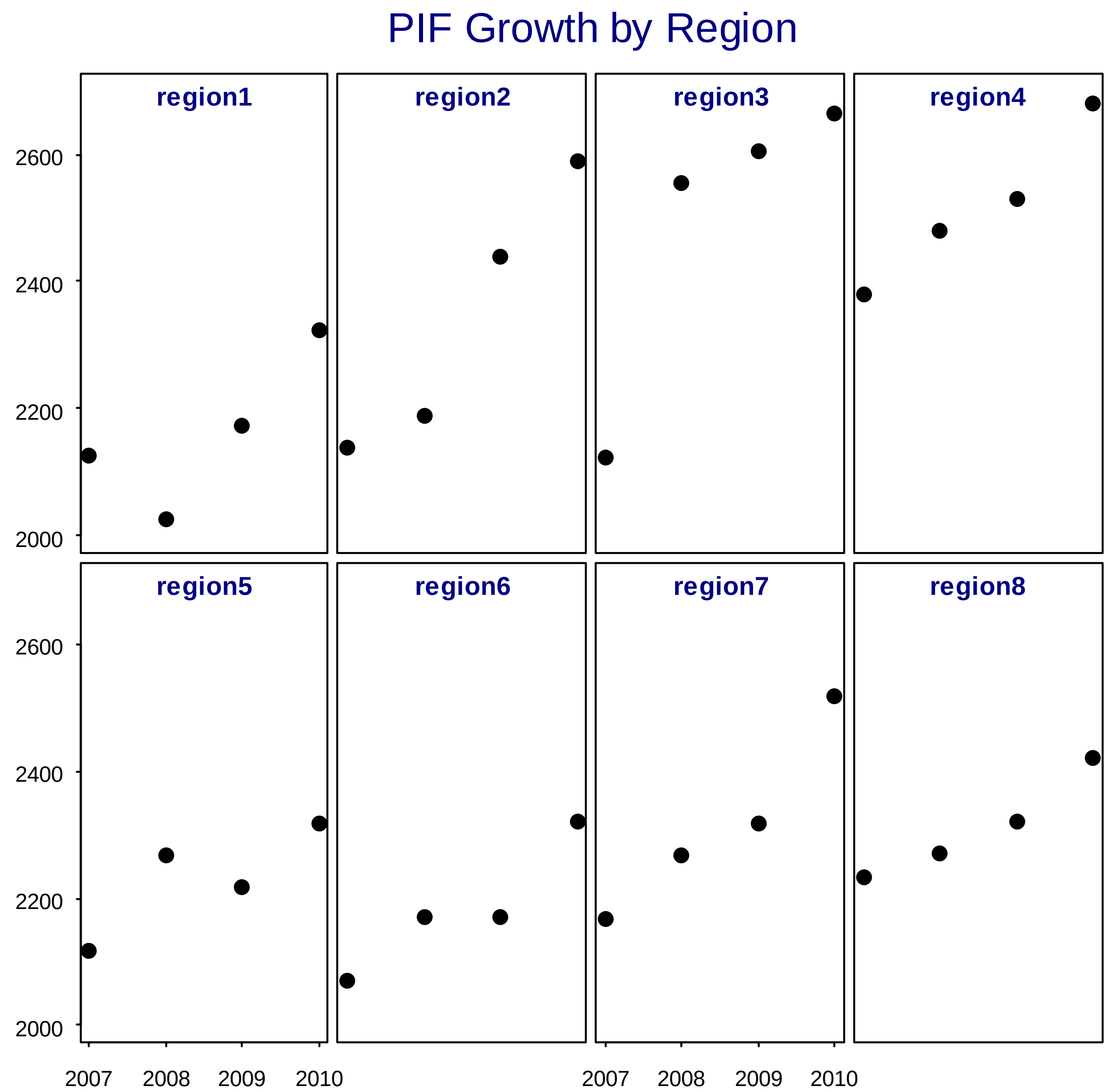
$$Y_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]} \cdot X_i, \sigma^2\right) \quad where \quad \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \Sigma\right) \quad , \quad \Sigma = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix}$$

Example: Policies In-Force Growth by Region

- Simple example: Growth in in-force policies [PIF] by region from 2007-10
- 32 data points
  - 4 years
  - 8 regions

region	2005	2006	2007	2008
1	2124	2024	2174	2324
2	2138	2188	2438	2588
3	2121	2554	2604	2666
4	2380	2480	2530	2680
5	2118	2268	2218	2318
6	2070	2170	2170	2320
7	2167	2267	2317	2517
8	2232	2272	2322	2422

- But we could as easily have 80 or 800 regions
  - Our model would not change



Classical Linear Model

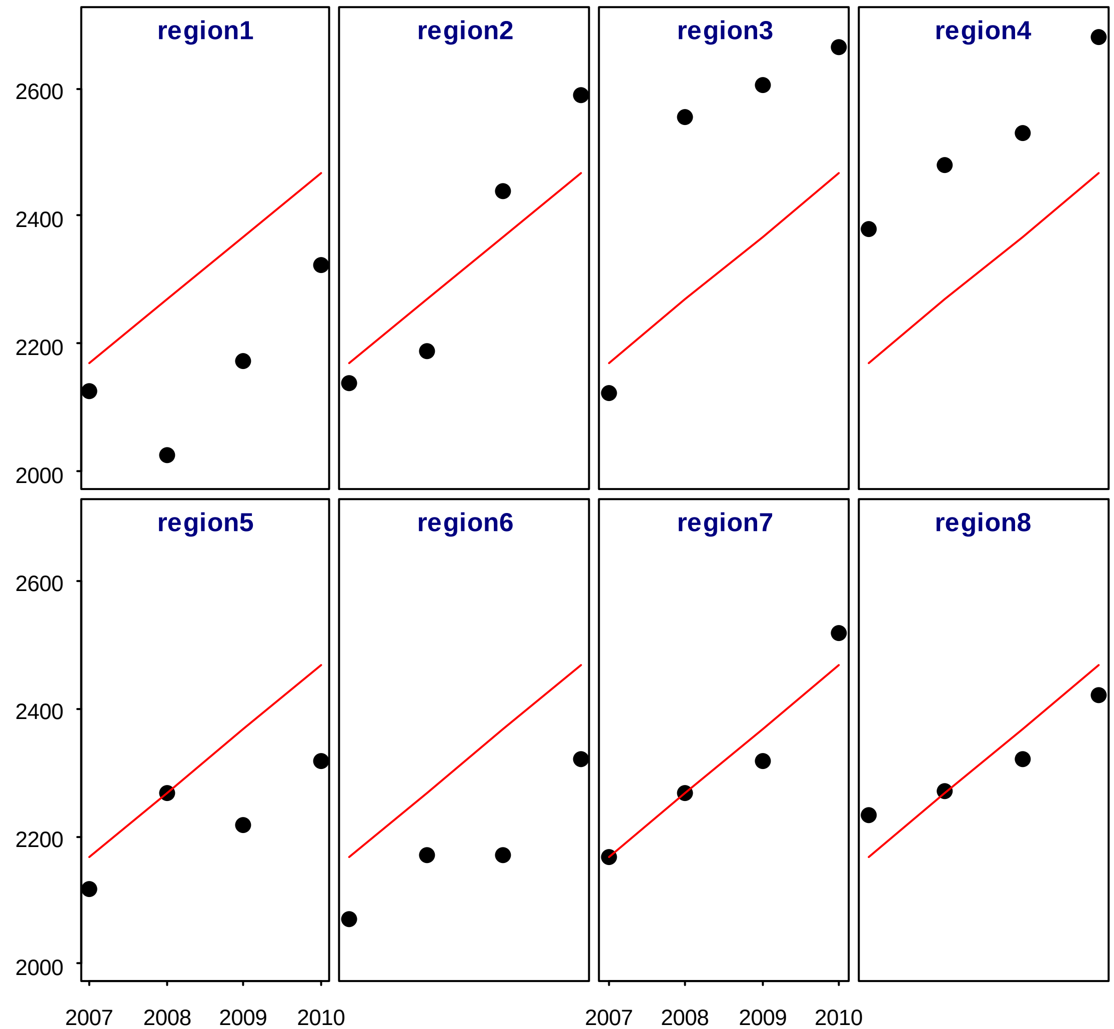
- Option 1: the classical linear model

- Complete Pooling
  - Don't reflect region in the model design
  - Just throw all of the data into one pot and regress

- $Y_i = \alpha + \beta X_i + \varepsilon_i$
  - i.e.: $Y_i \sim N(\alpha + \beta X_i, \sigma^2)$
  - Same α and β for every data point

- This obviously doesn't cut it
  - But nor do we want to fit 8 separate regression models

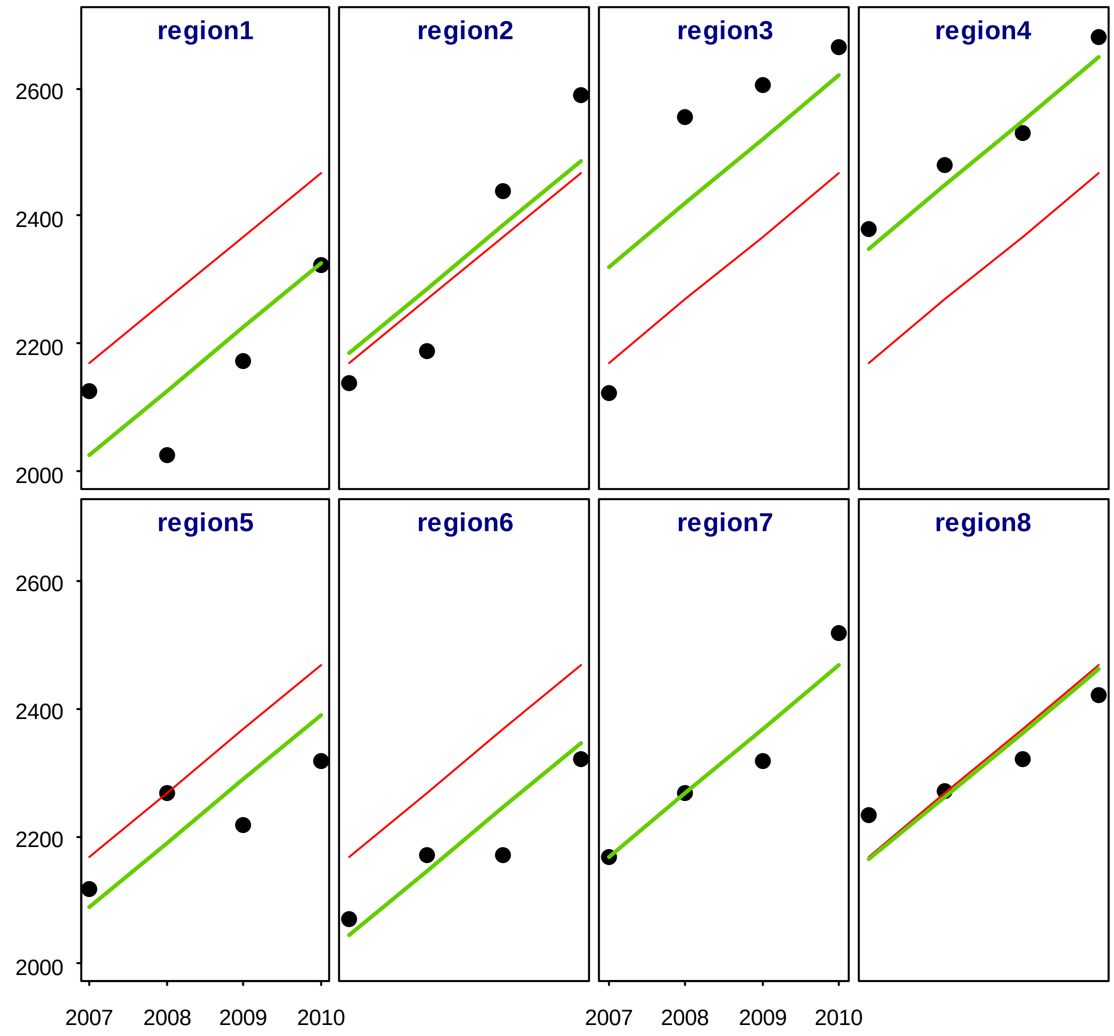
$$Y_i \sim N(\alpha + \beta X_i, \sigma^2)$$



Randomly Varying Intercepts

- Option 2: random intercept model
- $Y_i = \alpha_{j[i]} + \beta X_i + \varepsilon_i$
- This model has 9 parameters:
  $\{\alpha_1, \alpha_2, \ldots, \alpha_8, \beta\}$
- And it contains 4 hyperparameters:
  $\{\mu_\alpha, \beta_2, \sigma, \sigma_\alpha\}$.
- **A big improvement**

$$Y_i \sim N(\alpha_{j[i]} + \beta X_i, \sigma^2) \qquad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$$



Randomly Varying Intercepts and Slopes

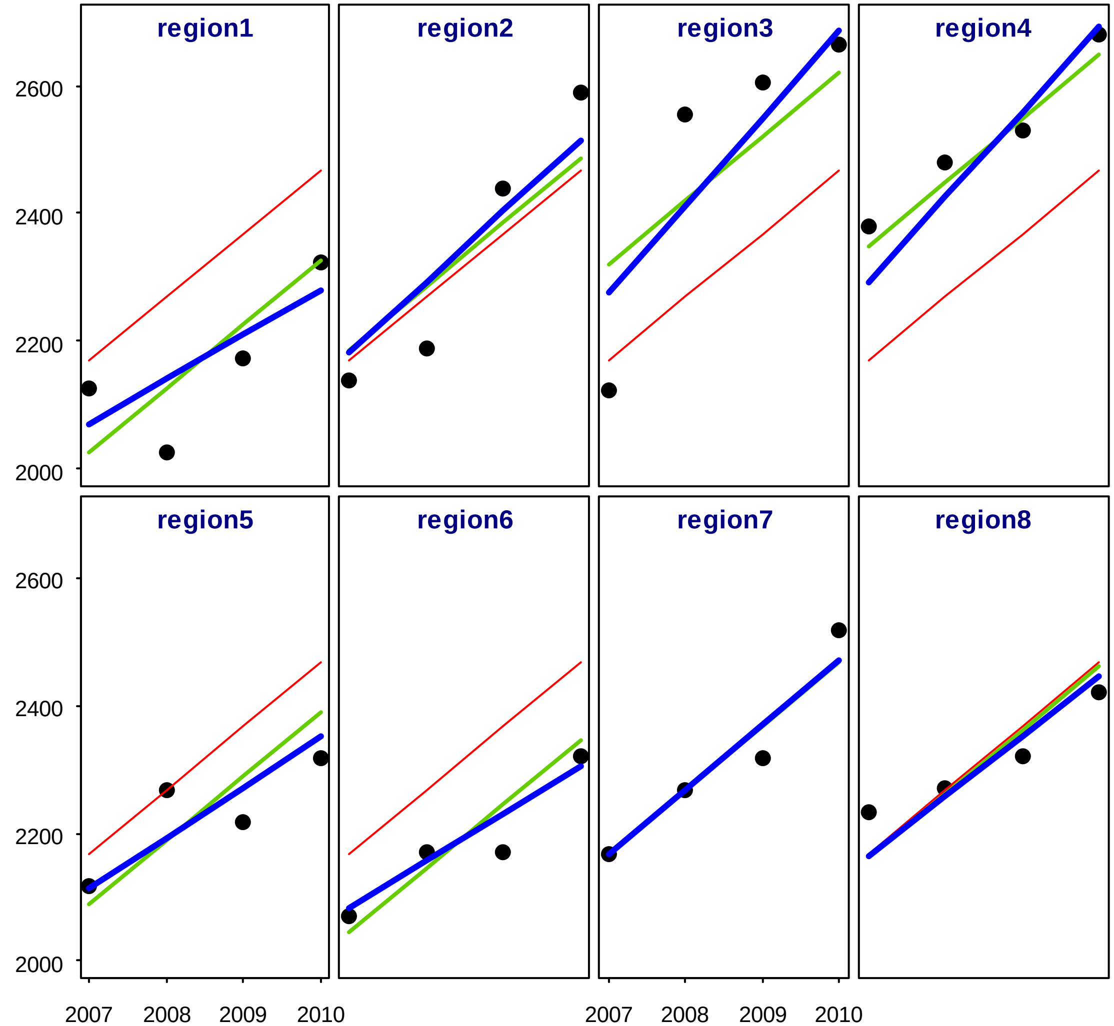
$$Y_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]} \cdot X_i, \sigma^2\right) \quad where \quad \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \Sigma\right) \quad , \quad \Sigma = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix}$$

- Option 3: random slope and intercept model
- $Y_i = \alpha_{j[i]} + \beta_{j[i]} X_i + \varepsilon_i$
- This model has 16 parameters:

  $\{\alpha_1, \alpha_2, \ldots, \alpha_8,$
  $\beta_1, \beta_2, \ldots, \beta_8,\}$

  - Note that 8 separate models would also yield 16 parameters
- And it contains 6 hyperparameters:

  $\{\mu_\alpha, \mu_\beta, \sigma, \sigma_\alpha, \sigma_\beta, \sigma_{\alpha\beta}\}$



Compromise Between Complete Pooling & No Pooling

$$PIF = \alpha + \beta t + \varepsilon$$

Complete Pooling

- Ignore group structure altogether

$$\{PIF = \alpha^k + \beta^k t + \varepsilon^k\}_{k=1,2,..,8}$$

No Pooling

- Estimating one model for each group

Compromise

Hierarchical Model

- Estimates parameters using a compromise between complete pooling and no pooling.

$$Y_i \sim N\left(\alpha_{j[i]} + \beta_{j[i]} \cdot X_i, \sigma^2\right) \quad where \quad \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix}, \Sigma\right) \quad , \quad \Sigma = \begin{bmatrix} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{bmatrix}$$

Enhanced Credibility

- Let's focus on the random intercept model:

$$Y_i \sim N(\alpha_{j[i]} + \beta X_i, \sigma^2) \qquad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$$

- This model can contain a large number of parameters: $\{\alpha_1, \alpha_2, \ldots, \alpha_J, \beta\}$.
- And it contains 4 <u>hyper</u>parameters: $\{\mu_\alpha, \beta_2, \sigma, \sigma_\alpha\}$.

- The hyperparameters relate to the parameters in a familiar way:

$$\hat{\alpha}_j = Z_j \cdot (\bar{y}_j - \hat{\beta}\bar{x}_j) + (1 - Z_j) \cdot \hat{\mu}_\alpha \quad where \quad Z_j = \frac{n_j}{n_j + \hat{\sigma}^2 / \hat{\sigma}_\alpha^2}$$

- **The multilevel/hierarchical modeling framework unifies of two pillars of actuarial practice:**
  - **GLM modeling**
  - **Bühlmann-style credibility**

The Middle Way

- The random intercept model is a compromise between the pooled-data model (option 1) and the separate models for each region (option 2).

$$\hat{\alpha}_j = Z_j \cdot (\bar{y}_j - \hat{\beta}\bar{t}_j) + (1 - Z_j) \cdot \hat{\mu}_\alpha \quad where \quad Z_j = \frac{n_j}{n_j + \hat{\sigma}^2 / \hat{\sigma}_\alpha^2}$$

- As $\sigma_\alpha \rightarrow 0$, the random intercept model → complete pooling

- As $\sigma_\alpha \rightarrow \infty$, the random intercept model → separate models by group

- In principle it's nearly always appropriate to use hierarchical models
    - Rather than a judgment call, the data tells us the degree to which the groups should be fit using separate models or a single common model
    - It is no longer an all-or-nothing decision

- Multilevel/hierarchical modeling should – will? – be a standard part of the practicing actuary's toolkit.

Hierarchical Modeling for Ratemaking

Ratemaking and Generalized Linear Models

- Personal insurance (auto, home) policies are today commonly priced using Generalized Linear Models.

- Actuarially fair premium = E[loss] + expense/profit load

- Insurance losses have a frequency and severity component:

$$Y_i = X_{i1} + X_{i2} + \ldots + X_{iN} \quad , \quad N = 0,1,2,\ldots$$

- Ratemaking: regress Y_i on risk factors
  - Majority of policies have "exact zeros" corresponding to no loss
  - Losses conditional on claim have skewed distribution

- Common approaches:
  - Separate frequency/severity models (e.g. Poisson/Gamma)
  - Model pure premium directly using compound gamma-Poisson (Tweedie) GLM models.
  - Tweedie models: $\mathrm{Var}(Y_i) = \varphi \mathrm{E}[Y_i]^p$, $p \in (1,2)$

Example: Modeling Claim Frequency

- Personal auto dataset.
- 67K observations.
- Build Poisson claim frequency models.

```
> all[1:10,]
   exposure numclaims veh_value veh_age gender agecat area veh_body body_type
1  0.3039014         0      1.06       3      F      2    C    HBACK     HBACK
2  0.6488706         0      1.03       2      F      4    A    HBACK     HBACK
3  0.5694730         0      3.26       2      F      2    E      UTE       UTE
4  0.3175907         0      4.14       2      F      2    D    STNWG     STNWG
5  0.6488706         0      0.72       4      F      2    C    HBACK     HBACK
6  0.8542094         0      2.01       3      M      4    C    HDTOP     HDTOP
7  0.8542094         0      1.60       3      M      4    A    PANVN     PANVN
8  0.5557837         0      1.47       2      M      6    B    HBACK     HBACK
9  0.3613963         0      0.52       4      F      3    A    HBACK     HBACK
10 0.5201916         0      0.38       4      F      4    B    HBACK     HBACK
>
> dim(all)
[1] 67856     9
```

- AREA and BODY_TYPE are highly categorical values.
  - We can treat these as dummy variables or as random intercepts.
  - Note several levels of Body Type have few exposures.

```
> round(tapply(exposure, area, sum))
   A    B    C    D    E    F
7597 6298 9578 3820 2772 1736
> round(tapply(exposure, veh_body, sum))
  BUS CONVT COUPE HBACK HDTOP MCARA MIBUS PANVN RDSTR SEDAN STNWG TRUCK   UTE
   26    33   319  8810   783    59   317   409    12 10445  7638   844  2106
```

Model #1: Standard Poisson Regression

- We build a 4-factor model
  - Vehicle Value
  - Driver Age
  - Area (territory)
  - Vehicle body type

- Many levels of AREA, BODY_TYPE are not statistically significant.

- **Note:** levels of BODY_TYPE with few exposures have large GLM parameters.

- **Dilemma:** should we exclude these levels, judgmentally temper them, or keep them as-is?

```
Call:
glm(formula = numclaims ~ veh_value + factor(agecat) + area +
    body_type, family = poisson, data = all, offset = log(exposure))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9701  -0.4528  -0.3460  -0.2212   4.5247

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1.676697   0.059593 -28.136  < 2e-16 ***
veh_value          0.054132   0.012378   4.373 1.22e-05 ***
factor(agecat)2   -0.174371   0.054157  -3.220 0.001283 **
factor(agecat)3   -0.233137   0.052857  -4.411 1.03e-05 ***
factor(agecat)4   -0.260159   0.052727  -4.934 8.05e-07 ***
factor(agecat)5   -0.479397   0.059082  -8.114 4.89e-16 ***
factor(agecat)6   -0.460072   0.067566  -6.809 9.81e-12 ***
areaB              0.054467   0.042804   1.272 0.203213
areaC              0.006597   0.038995   0.169 0.865651
areaD             -0.110542   0.052933  -2.088 0.036768 *
areaE             -0.031239   0.057866  -0.540 0.589301
areaF              0.060685   0.066114   0.918 0.358675
body_typeBUS       0.877358   0.317783   2.761 0.005765 **
body_typeCONVT    -0.979685   0.588638  -1.664 0.096048 .
body_typeCOUPE     0.355757   0.118525   3.002 0.002686 **
body_typeHBACK    -0.030187   0.037553  -0.804 0.421495
body_typeHDTOP     0.052380   0.090219   0.581 0.561518
body_typeMCARA     0.467935   0.260606   1.796 0.072564 .
body_typeMIBUS    -0.126886   0.151430  -0.838 0.402079
body_typePANVN     0.037731   0.123999   0.304 0.760910
body_typeRDSTR     0.296033   0.579598   0.511 0.609522
body_typeSTNWG    -0.026440   0.041465  -0.638 0.523710
body_typeTRUCK    -0.065282   0.092729  -0.704 0.481426
body_typeUTE      -0.222763   0.066394  -3.355 0.000793 ***
```

Model #2: Random Intercepts for Area and Body Type

- Rather than use dummy variables for AREA and BODY_TYPE we can introduce "random effects".
- Methodology equally applicable even with many more levels.

```
> summary(m2)
Generalized linear mixed model fit by the Laplace approximation
Formula: numclaims ~ veh_value + factor(agecat) + (1 | area) + (1 | veh_body)
   Data: all
   AIC   BIC logLik deviance
 25409 25492 -12696    25391
Random effects:
 Groups   Name        Variance  Std.Dev.
 veh_body (Intercept) 0.0109110 0.104456
 area     (Intercept) 0.0016531 0.040658
Number of obs: 67856, groups: veh_body, 13; area, 6

Fixed effects:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.67722    0.06624 -25.319  < 2e-16 ***
veh_value        0.05003    0.01172   4.268 1.97e-05 ***
factor(agecat)2 -0.17358    0.05410  -3.209  0.00133 **
factor(agecat)3 -0.23397    0.05276  -4.435 9.23e-06 ***
factor(agecat)4 -0.26008    0.05266  -4.939 7.84e-07 ***
factor(agecat)5 -0.47950    0.05900  -8.128 4.38e-16 ***
factor(agecat)6 -0.46323    0.06742  -6.871 6.37e-12 ***
```

```
> ranef(m2)
$veh_body
       (Intercept)
BUS    0.061306648
CONVT -0.046777680
COUPE  0.155021044
HBACK -0.024148049
HDTOP  0.035785954
MCARA  0.055752923
MIBUS -0.040128201
PANVN  0.018846328
RDSTR  0.008698423
SEDAN  0.004750781
STNWG -0.015622911
TRUCK -0.037829055
UTE   -0.165545254

$area
   (Intercept)
A  0.002021295
B  0.035785439
C  0.006824017
D -0.051164202
E -0.012967832
F  0.021033130
```

Model #3: Add Vehicle Value Random Slope

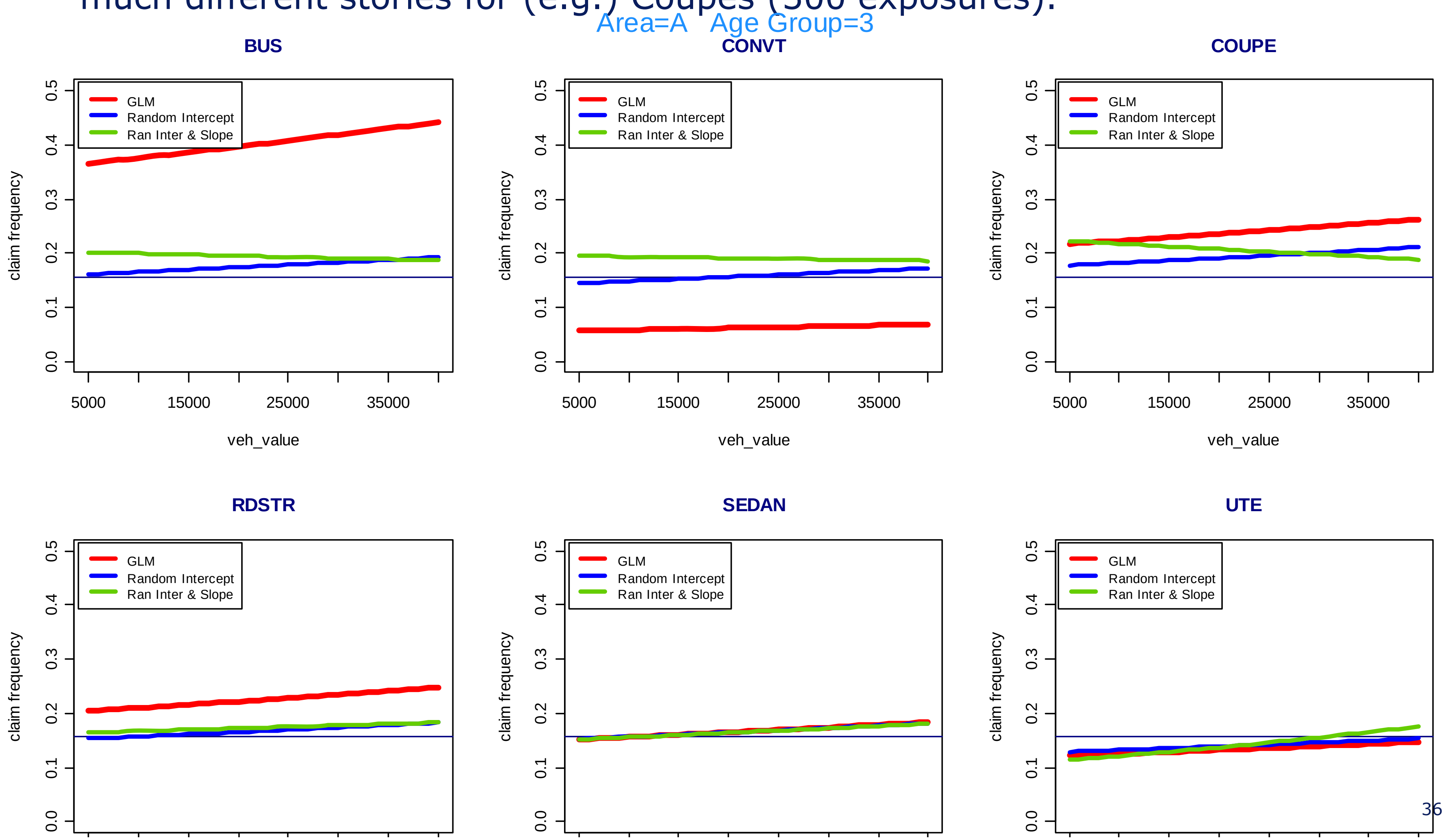
- **Intuition:** Relationship between vehicle value and claim frequency might vary by type of vehicle.
- **Response:** Introduce random slopes for VEH_VALUE.

```
> summary(m3)
Generalized linear mixed model fit by the Laplace approximation
Formula: numclaims ~ veh_value + factor(agecat) + (1 | area) +
   Data: all
   AIC   BIC logLik deviance
 25409 25510 -12694    25387
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 veh_body (Intercept) 0.0618265 0.248649
          veh_value   0.0031765 0.056360 -1.000
 area     (Intercept) 0.0015220 0.039012
Number of obs: 67856, groups: veh_body, 13; area, 6

Fixed effects:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.61442    0.09993 -16.156  < 2e-16 ***
veh_value        0.03544    0.02240   1.582  0.11359
factor(agecat)2 -0.17204    0.05407  -3.182  0.00146 **
factor(agecat)3 -0.23130    0.05271  -4.388 1.14e-05 ***
factor(agecat)4 -0.25756    0.05263  -4.894 9.89e-07 ***
factor(agecat)5 -0.47587    0.05895  -8.073 6.88e-16 ***
factor(agecat)6 -0.45767    0.06738  -6.792 1.10e-11 ***
```

```
> ranef(m3)
$veh_body
      (Intercept)    veh_value
BUS    0.25480949 -0.057756752
CONVT  0.21485769 -0.048701020
COUPE  0.36562617 -0.082875169
HBACK -0.09898229  0.022435959
HDTOP -0.02703973  0.006128999
MCARA  0.11200410 -0.025387566
MIBUS -0.13195792  0.029910427
PANVN -0.06120002  0.013871989
RDSTR  0.02546871 -0.005772900
SEDAN -0.06570074  0.014892151
STNWG -0.10148617  0.023003505
TRUCK -0.09823058  0.022265573
UTE   -0.39124661  0.088682462

$area
   (Intercept)
A  0.002499680
B  0.035031096
C  0.007269752
D -0.049034245
E -0.012770462
F  0.018773112
```

Model Comparison

- **Shrinkage:** The hierarchical model estimates (green, blue) are less extreme than the standard GLM estimates.
- **Different stories:** All models agree for (e.g.) Sedans (10K+ exposures) but tell much different stories for (e.g.) Coupes (300 exposures).



Hierarchical Modeling for Loss Reserving

Loss Reserving and its Discontents

Models vs Methods
Need for Variability Estimates

Loss Reserving 101

- The largest balance sheet liability for property-casualty insurers is the provision set aside to pay the claims for which it is liable.

- Claim amounts “develop” over time
  - claims that have occurred but not yet been reported [IBNR]
  - Lawsuits / judicial proceedings
  - Ongoing disability claims for workers compensation claims
  - …

- Major job for actuaries: estimate the ultimate dollars of loss that will be paid for claims incurred in a given year.
  - “accident year”

Loss Reserving 101

- CAS statement of the problem: "given our current state of knowledge, what is the probability that [an entity's] final payments will be no larger than [a] given value?"
  - A point estimate isn't really enough
  - We want a "range"
- Current practice perhaps lags what is possible.
- In practice, spreadsheet-based projection methods are often used.
- 2003 Standard & Poor's report: suggested "naivety or knavery" on the part of the actuarial profession for inadequately reserved companies.

The Chain-Ladder: A Baseline Reserving Methodology

- Here is a garden-variety loss triangle:

		Cumulative Losses in 1000's												
AY	*premium*	*12*	*24*	*36*	*48*	*60*	*72*	*84*	*96*	*108*	*120*	*CL Ult*	*CL LR*	*CL res*
1988	2,609	404	986	1,342	1,582	1,736	1,833	1,907	1,967	2,006	2,036	2,036	0.78	0
1989	2,694	387	964	1,336	1,580	1,726	1,823	1,903	1,949	1,987		2,017	0.75	29
1990	2,594	421	1,037	1,401	1,604	1,729	1,821	1,878	1,919			1,986	0.77	67
1991	2,609	338	753	1,029	1,195	1,326	1,395	1,446				1,535	0.59	89
1992	2,077	257	569	754	892	958	1,007					1,110	0.53	103
1993	1,703	193	423	589	661	713						828	0.49	115
1994	1,438	142	361	463	533							675	0.47	142
1995	1,093	160	312	408								601	0.55	193
1996	1,012	131	352									702	0.69	350
1997	976	122										576	0.59	454
chain link		2.365	1.354	1.164	1.090	1.054	1.038	1.026	1.020	1.015	1.000	12,067		1,543
chain ldf		4.720	1.996	1.473	1.266	1.162	1.102	1.062	1.035	1.015	1.000			
growth curve		21.2%	50.1%	67.9%	79.0%	86.1%	90.7%	94.2%	96.6%	98.5%	100.0%			

- Cumulative losses summarized to the accident year/development period level

- The chain-ladder is a commonly used reserving technique
  - A simple way of projecting historical loss development patterns into the future
  - Calculate $t_j \rightarrow t_{j+1}$ "link ratios"
  - Piecewise approximation of the overall historical loss development pattern
  - Apply to (less mature) losses from more recent years

Loss Reserving and its Discontents

- Much loss reserving practice is "pre-theoretical" in nature.
    - Techniques like chain ladder, BF, and Cape Cod aren't performed in a statistical modeling framework.

- Traditional **methods** aren't necessarily optimal from a statistical POV.
    - Potential of over-fitting small datasets.
    - Difficult to assess goodness-of-fit, compare nested models, etc.
    - Often no concept of out-of-sample validation or diagnostic plots.

- Related point: traditional methods produce point estimates only.

    → Reserve variability estimates in practice are often ad hoc.

- Stochastic reserving: build statistical **models** of loss development.
    - Attempt to place loss reserving practice on a sound scientific footing.
    - Field is developing rapidly.
    - Today: explore non-linear hierarchical models (aka "nonlinear mixed effects models") as natural, parsimonious models of the loss development process.
    - Initially motivated by Dave Clark's paper [2003] as well as nonlinear mixed effects model [NLME] theory.

The Chain-Ladder is a type of GLM

- The cumulative loss triangle in symbolic form:

i	t_1	t_2	$\ldots$	t_{I-1}	t_I	$\ldots$ t_∞	Premium
1	$y_1(t_1)$	$y_1(t_2)$	$\cdots$	$y_1(t_{I-1})$	$y_1(t_I)$		p_1
2	$y_2(t_1)$	$y_2(t_2)$	$\cdots$	$y_2(t_{I-1})$			p_2
$\vdots$	$\vdots$	$\vdots$	$\ldots$				$\vdots$
i	$y_i(t_1)$	$\cdots$	$y_i(t_{I+1-i})$				p_i
$\vdots$	$\vdots$	$\vdots$					$\vdots$
$I-1$	$y_{I-1}(t_1)$	$y_{I-1}(t_2)$					p_{I-1}
I	$y_I(t_1)$						p_I

- Let's recast this triangle in terms of <u>incremental</u> losses: $z_i(t_j) = y_i(t_j) - y_i(t_{j-1})$

- **We can replicate the chain-ladder solution with an over-dispersed Poisson [ODP] GLM model with *I* row and *I* column effects**
  - e.g. England-Verrall 2001

$$\log\left(E[z_i(t_j)]\right) = \mu + \alpha_i + \beta_j \qquad \forall i, j \in 1, \ldots, I$$

What Do You See?

- Let's look at the loss triangle with fresh eyes.
- We would like to do stochastic reserving the "right" way.
- What considerations come to mind?

Cumulative Losses in 1000's

AY	*premium*	*12*	*24*	*36*	*48*	*60*	*72*	*84*	*96*	*108*	*120*
1988	2,609	404	986	1,342	1,582	1,736	1,833	1,907	1,967	2,006	2,036
1989	2,694	387	964	1,336	1,580	1,726	1,823	1,903	1,949	1,987	
1990	2,594	421	1,037	1,401	1,604	1,729	1,821	1,878	1,919		
1991	2,609	338	753	1,029	1,195	1,326	1,395	1,446			
1992	2,077	257	569	754	892	958	1,007				
1993	1,703	193	423	589	661	713					
1994	1,438	142	361	463	533						
1995	1,093	160	312	408							
1996	1,012	131	352								
1997	976	122									

Some Essential Features of Loss Reserving

- **Repeated measures**
  - The dataset is inherently longitudinal in nature.

Cumulative Losses in 1000's

AY	*premium*	*12*	*24*	*36*	*48*	*60*	*72*	*84*	*96*	*108*	*120*
1988	2,609	404	986	1,342	1,582	1,736	1,833	1,907	1,967	2,006	2,036
1989	2,694	387	964	1,336	1,580	1,726	1,823	1,903	1,949	1,987	
1990	2,594	421	1,037	1,401	1,604	1,729	1,821	1,878	1,919		
1991	2,609	338	753	1,029	1,195	1,326	1,395	1,446			
1992	2,077	257	569	754	892	958	1,007				
1993	1,703	193	423	589	661	713					
1994	1,438	142	361	463	533						
1995	1,093	160	312	408							
1996	1,012	131	352								
1997	976	122									

- **A "Bundle" of time series**
  - A loss triangle is a collection of time series that are "related" to one another…
  - … but no guarantee that the same development pattern is appropriate to each one

- **Non-linear**
  - Each year's loss development pattern in inherently non-linear
  - Ultimate loss (ratio) is an asymptote

- **Incomplete information**
  - Few loss triangles contain all of the information needed to make forecasts
  - Most reserving exercises must incorporate judgment and/or background information

→ ***Loss reserving is inherently Bayesian***

Towards a More Realistic Stochastic Reserving Framework

- How many stochastic loss reserving techniques reflect <u>all</u> of these considerations?
  1. Repeated Measures (Isn't loss reserving a type of <u>longitudinal data analysis</u>?)
  2. Multiple Time Series
  3. Non-linear (Are GLMs really appropriate?)
  4. Incomplete information ("Bayes or Bust"!)

1-2 ➔ We need to build <u>**hierarchical models**</u>

3 ➔ Our hierarchical models should be <u>**non-linear**</u> (growth curves)

4 ➔ Our non-linear hierarchical models should be <u>**Bayesian**</u>

Cumulative Losses in 1000's

AY	premium	12	24	36	48	60	72	84	96	108	120
1988	2,609	404	986	1,342	1,582	1,736	1,833	1,907	1,967	2,006	2,036
1989	2,694	387	964	1,336	1,580	1,726	1,823	1,903	1,949	1,987	
1990	2,594	421	1,037	1,401	1,604	1,729	1,821	1,878	1,919		
1991	2,609	338	753	1,029	1,195	1,326	1,395	1,446			
1992	2,077	257	569	754	892	958	1,007				
1993	1,703	193	423	589	661	713					
1994	1,438	142	361	463	533						
1995	1,093	160	312	408							
1996	1,012	131	352								
1997	976	122									

Origin of the Approach: Dave's Idea + Random Effects

LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach

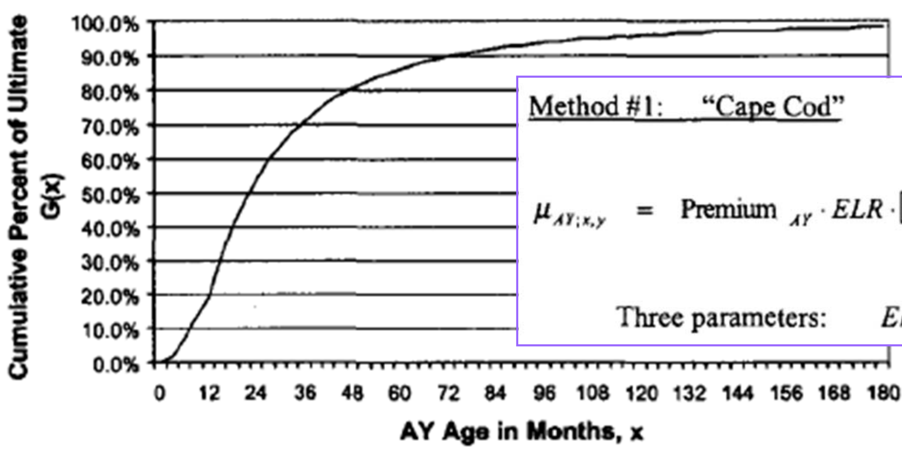
or

How to Increase Reserve Variability with Less Data

David R. Clark
American Re-Insurance

2003 Reserves Call Paper Program

$G(x) \;=\; 1/LDF_x$ = cumulative % reported (or paid) as of time x

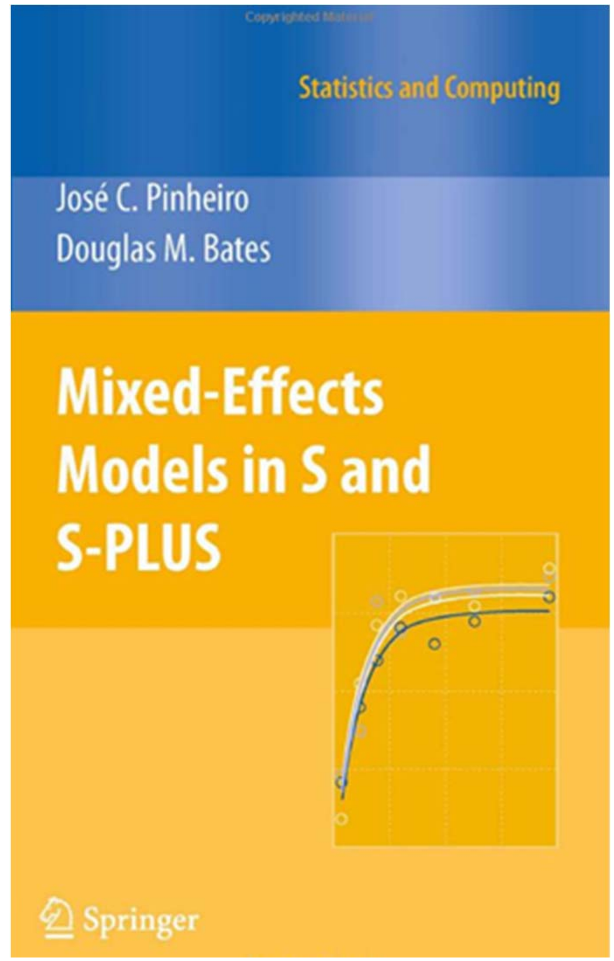


Method #1: "Cape Cod"

$$\mu_{AY;x,y} \;=\; \text{Premium}_{AY} \cdot ELR \cdot [G(y \mid \omega,\theta) - G(x \mid \omega,\theta)]$$

Three parameters: $ELR,\ \omega,\ \theta$

+



Components of Our Approach

- **Growth curves** to model the loss development process (Clark 2003)
  - Parsimony; obviates need for tail factors
- Loss reserving treated as **longitudinal data analysis** (Guszcza 2008)
  - Parsimony; similar approach to non-linear mixed effects models used in biological and social sciences
- Further using the hierarchical modeling framework to simultaneously model **multiple loss triangles** (Zhang-Dukic-Guszcza 2011)
  - "Borrow strength" from other loss reserving triangles
  - Similar in spirit to credibility theory
- Building a **fully Bayesian** model by assigning prior probability distributions to all hyperparameters (Zhang-Dukic-Guszcza 2011)
  - Provides formal mechanism for incorporating background knowledge and expert opinion with data-driven indications.
  - Results in full predictive distribution of all quantities of interest
  - Conceptual advantages: Bayesian paradigm treats data as fixed and parameters are randomly varying

Hierarchical Growth Curve Loss Reserving Model (Empirical Bayes)

Hierarchical Modeling for Loss Reserving

- Here is our Schedule P loss triangle:

Cumulative Losses in 1000's

AY	premium	12	24	36	48	60	72	84	96	108	120	CL Ult	CL LR	CL res
1988	2,609	404	986	1,342	1,582	1,736	1,833	1,907	1,967	2,006	2,036	2,036	0.78	0
1989	2,694	387	964	1,336	1,580	1,726	1,823	1,903	1,949	1,987		2,017	0.75	29
1990	2,594	421	1,037	1,401	1,604	1,729	1,821	1,878	1,919			1,986	0.77	67
1991	2,609	338	753	1,029	1,195	1,326	1,395	1,446				1,535	0.59	89
1992	2,077	257	569	754	892	958	1,007					1,110	0.53	103
1993	1,703	193	423	589	661	713						828	0.49	115
1994	1,438	142	361	463	533							675	0.47	142
1995	1,093	160	312	408								601	0.55	193
1996	1,012	131	352									702	0.69	350
1997	976	122										576	0.59	454
chain link		2.365	1.354	1.164	1.090	1.054	1.038	1.026	1.020	1.015	1.000	12,067		1,543
chain ldf		4.720	1.996	1.473	1.266	1.162	1.102	1.062	1.035	1.015	1.000			
growth curve		21.2%	50.1%	67.9%	79.0%	86.1%	90.7%	94.2%	96.6%	98.5%	100.0%			

- Let's model this as a longitudinal dataset.
- Grouping dimension: Accident Year (AY)

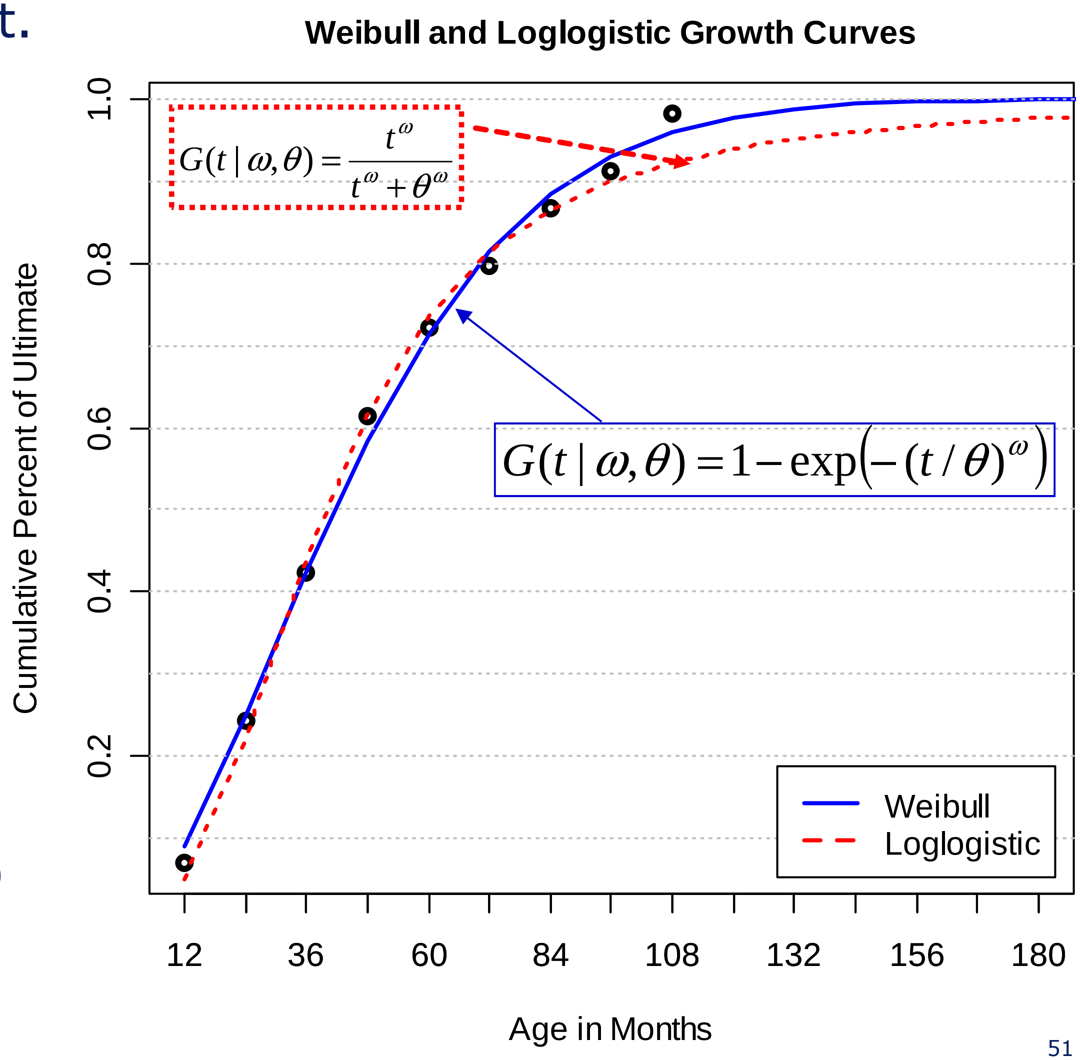
- **We can build a parsimonious non-linear model that uses random effects to allow the model parameters to vary by accident year.**

Growth Curves

- We want to build that reflects the **non-linear** nature of loss development.
  - GLM shows up a lot in the stochastic loss reserving literature.
  - But… are GLMs natural models for loss triangles?

- Growth curves
  - 2-parameter curves
  - θ = scale
  - ω = shape
  - See Clark [2003]

- Heuristic idea:
  - We fit these curves to the LDFs
  - Add random effects
  - Allows ultimate loss (ratio) and/or θ and/or ω to vary randomly by year.



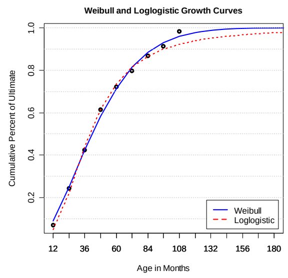
Baseline Model: Heuristics

- Basic intuition is familiar: $(\text{CumLoss}_{\text{AY},t}) * (\text{LDF}) = \text{Ultimate loss}$

Baseline Model: Heuristics

- Basic intuition is familiar: $(\text{CumLoss}_{AY,t}) * (\text{LDF}) = \text{Ultimate loss}$

➔ $y_i(t_j) = (\text{Ult loss}_{AY}) * (1 / \text{LDF}_t)$



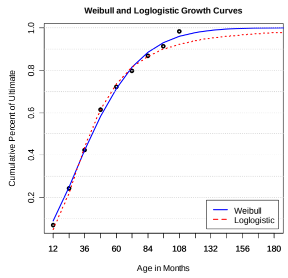
➔ $y_i(t_j) = (\text{Ult loss}_{AY}) * G_{\omega,\theta}(t)$

Baseline Model: Heuristics

- Basic intuition is familiar: $(\text{CumLoss}_{AY,t}) * (\text{LDF}) = \text{Ultimate loss}$

➔ $y_i(t_j) = (\text{Ult loss}_{AY}) * (1 / \text{LDF}_t)$

➔ $y_i(t_j) = (\text{Ult loss}_{AY}) * G_{\omega,\theta}(t)$



$$y_i(t_j) = \gamma_i * p_i\left[1 - \exp\left(-(t_j/\theta)^{\omega}\right)\right] + \varepsilon_i(t_j)$$
$$\gamma_i \sim N\left(\gamma, \sigma_\gamma^2\right)$$
$$\varepsilon_i(t_j) = \rho\varepsilon_i(t_{j-1}) + \delta_i(t_j)$$

- $p_i \equiv$ premium for AY i (given)
- $\gamma_i \equiv$ ult loss ratio for AY I

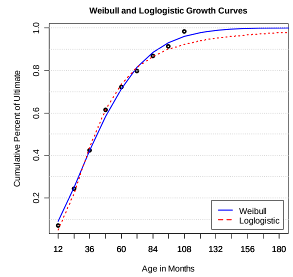
➔ $p_i \cdot \gamma_i$ = E[ult $loss]

Baseline Model: Heuristics

- Basic intuition is familiar: ($CumLoss_{AY,t}$) * (LDF) = Ultimate loss

➔ $y_i(t_j)$ = (Ult $loss_{AY}$) * (1 / LDF_t)

➔ $y_i(t_j)$ = (Ult $loss_{AY}$) * $G_{\omega,\theta}(t)$



$$y_i(t_j) = \gamma_i * p_i\left[1 - \exp\left(-(t_j/\theta)^{\omega}\right)\right] + \varepsilon_i(t_j)$$
$$\gamma_i \sim N\left(\gamma, \sigma_\gamma^2\right)$$
$$\varepsilon_i(t_j) = \rho\varepsilon_i(t_{j-1}) + \delta_i(t_j)$$

- p_i ≡ premium for AY i (given)
- γ_i ≡ ult loss ratio for AY *I*

➔ $p_i \cdot \gamma_i$ = E[ult $loss]

- The "growth curve" part comes in by using $G(t)$ instead of LDFs.
  - Think of LDF's as a rough piecewise linear approximation to a $G(t)$

- The "hierarchical" part comes in because we can let γ, θ, and/or ω vary by AY (using sub-models).

Other “Random Effects”

- Our model so far:

$$y_i(t_j) = \gamma_i * p_i\left[1-\exp\left(-(t_j/\theta)^{\omega}\right)\right] + \varepsilon_i(t_j)$$
$$\gamma_i \sim N\left(\gamma, \sigma_\gamma^2\right)$$
$$\varepsilon_i(t_j) = \rho\varepsilon_i(t_{j-1}) + \delta_i(t_j)$$

- What if we want to include other random effects in the model?
- It’s easily done:

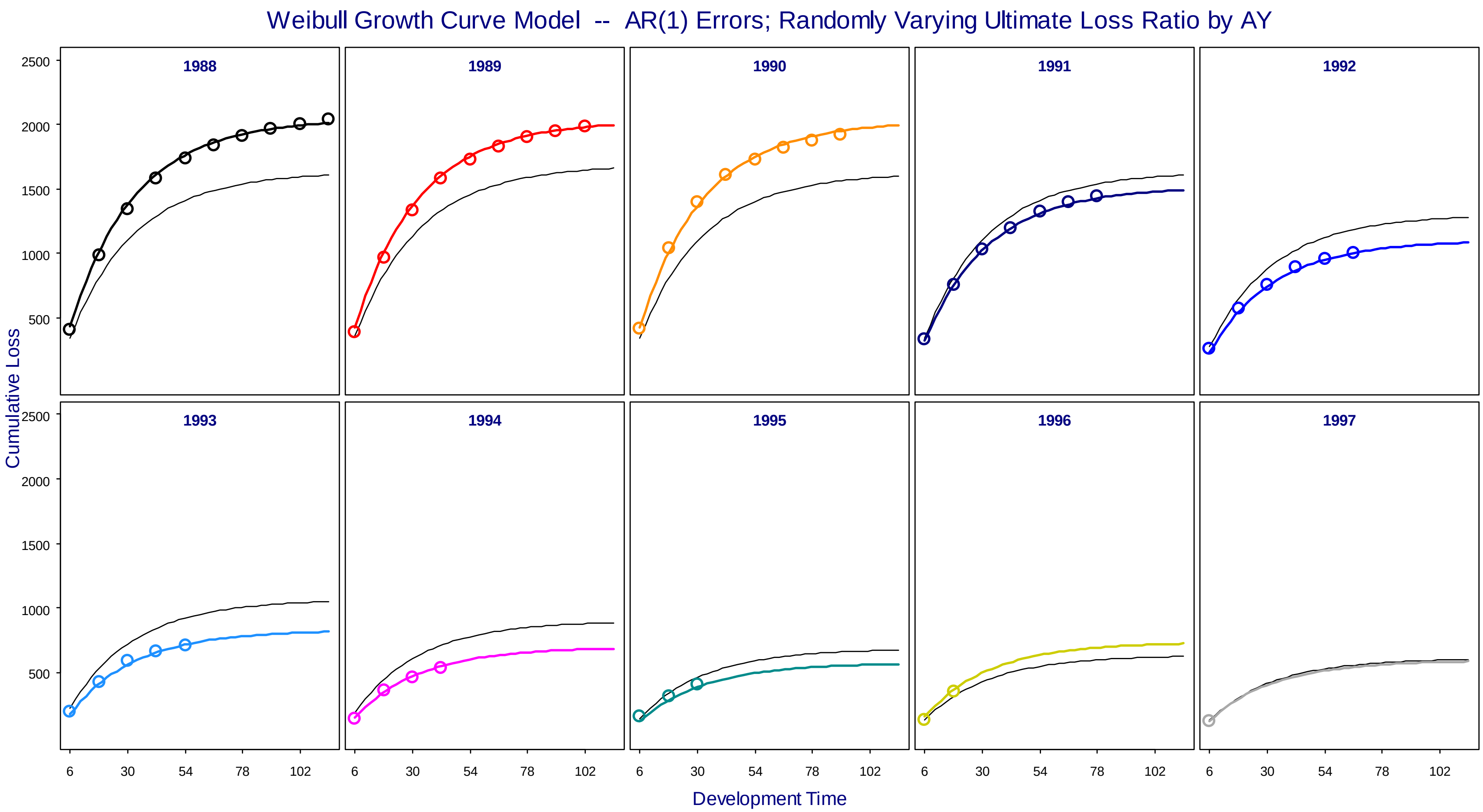
$$y_i(t_j) = \gamma_i \cdot p_i\left[1-\exp\left(-(t_j/\theta_i)^{\omega}\right)\right] + \varepsilon_i(t_j)$$
$$\begin{pmatrix}\gamma_i\\ \omega_i\end{pmatrix} \sim N\left(\begin{pmatrix}\gamma\\ \omega\end{pmatrix}, \Sigma\right) \quad , \quad \Sigma = \begin{pmatrix}\sigma_\gamma^2 & \sigma_{\gamma,\omega}\\ \sigma_{\gamma,\omega} & \sigma_\omega^2\end{pmatrix}$$
$$\varepsilon_i(t_j) = \rho\varepsilon_i(t_{j-1}) + \delta_i(t_j)$$

- Here we add a “random shape” effect to let ω vary by AY.
  - This is analogous to letting slope vary in a linear model.
  - When we try this we find that ω does not vary significantly by AY.
  - Analogous process found that “shape” (θ) does not significantly vary by AY either.

Baseline Model Performance

$$y_i(t_j) = \gamma_i * p_i\left[1 - \exp\left(-(t_j/\theta)^{\omega}\right)\right] + \varepsilon_i(t_j)$$
$$\gamma_i \sim N\left(\gamma, \sigma_\gamma^2\right)$$
$$\varepsilon_i(t_j) = \rho\varepsilon_i(t_{j-1}) + \delta_i(t_j)$$

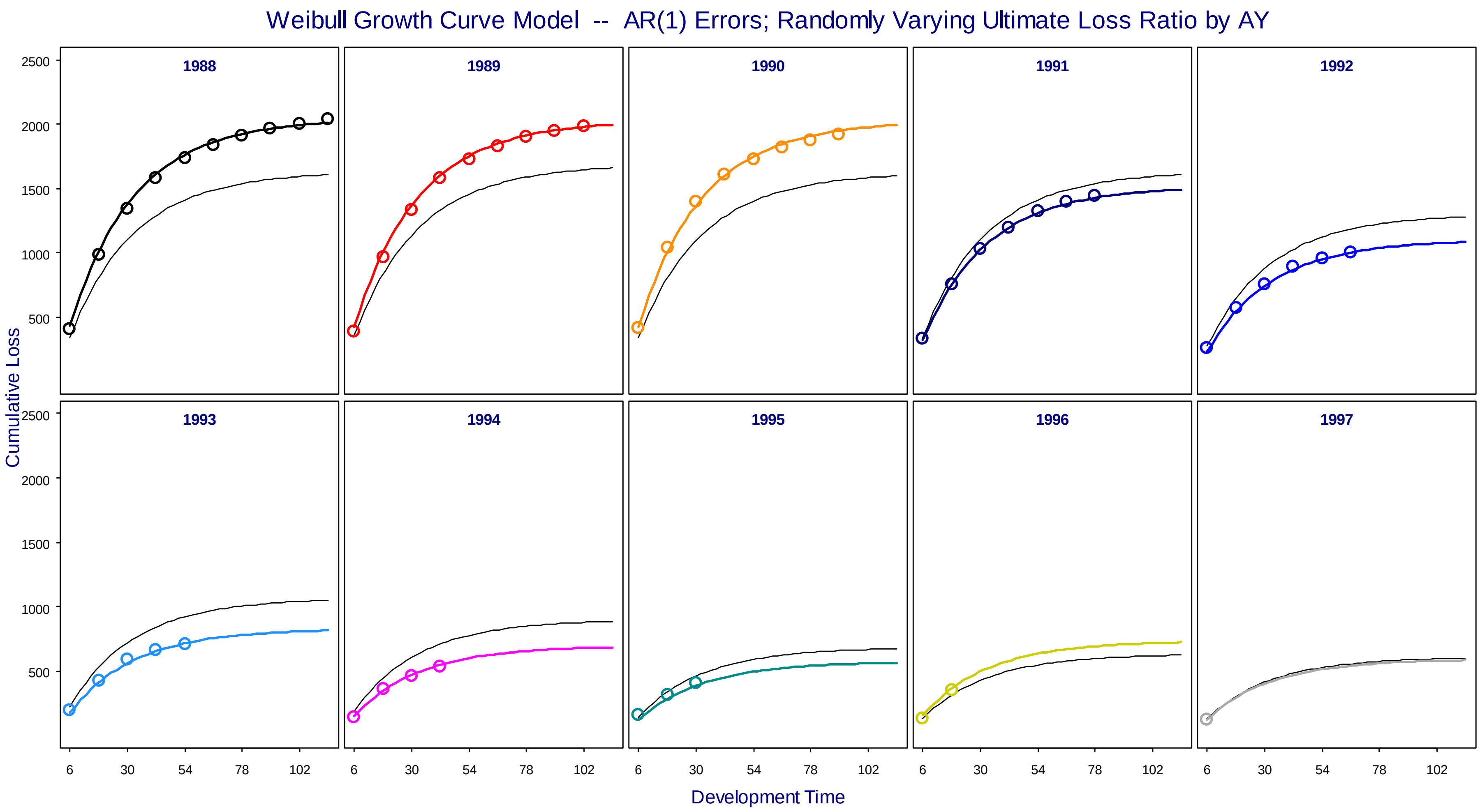
- Random LR effects allow a "custom fit" growth curve for each AY
- Yet the model is vary parsimonious
- The model contains only 6 hyperparameters, but fits the loss triangle very well
- Parsimony is achieved because the model is well suited to the data (not ad hoc)



Baseline Model Performance

$$y_i(t_j) = \gamma_i * p_i\left[1 - \exp\left(-(t_j/\theta)^{\omega}\right)\right] + \varepsilon_i(t_j)$$
$$\gamma_i \sim N\left(\gamma, \sigma_\gamma^2\right)$$
$$\varepsilon_i(t_j) = \rho\varepsilon_i(t_{j-1}) + \delta_i(t_j)$$

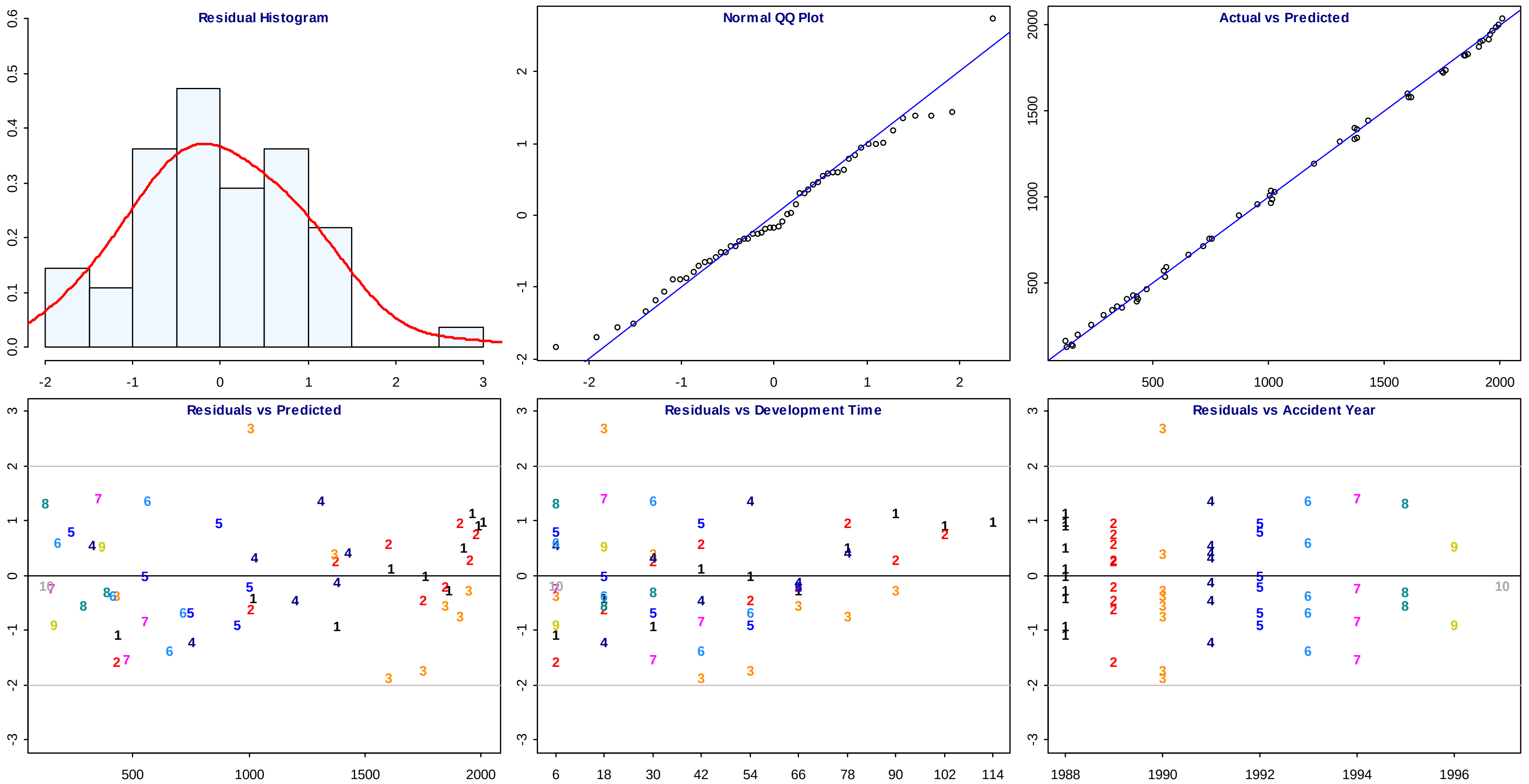
- We have estimated the parameters: {γ; ω ; θ; σ_γ ; ρ; σ}
- Random effects are added to ultimate LR parameter γ — Analogous to random intercepts
- Random shape (ω) effects don't seem necessary — Analogous to random slopes



Residual Diagnostics

- Residual diagnostics suggest a reasonable fit.
- Note: without the AR(1) error structure, the “development time” plot (bottom middle) would show wavy patterns of residuals.
- … but model risk could still be lurking …

Weibull Growth Curve Model -- AR(1) Errors; Randomly Varying Ultimate Loss Ratio by AY



Model Results

Chain Ladder Analysis

AY	premium	6	18	30	42	54	66	78	90	102	114	CL Ult	CL res
1988	2,609	404	986	1,342	1,582	1,736	1,833	1,907	1,967	2,006	2,036	2,036	0
1989	2,694	387	964	1,336	1,580	1,726	1,823	1,903	1,949	1,987		2,017	29
1990	2,594	421	1,037	1,401	1,604	1,729	1,821	1,878	1,919			1,986	67
1991	2,609	338	753	1,029	1,195	1,326	1,395	1,446				1,535	89
1992	2,077	257	569	754	892	958	1,007					1,110	103
1993	1,703	193	423	589	661	713						828	115
1994	1,438	142	361	463	533							675	142
1995	1,093	160	312	408								601	193
1996	1,012	131	352									702	350
1997	976	122										576	454
chain link		2.365	1.354	1.164	1.090	1.054	1.038	1.026	1.020	1.015	1.000	12,067	**1,543**
chain ldf		4.720	1.996	1.473	1.266	1.162	1.102	1.062	1.035	1.015	1.000		
growth curve		21.2%	50.1%	67.9%	79.0%	86.1%	90.7%	94.2%	96.6%	98.5%	100.0%		

Parameters and Estimated Reserves - Weibull Model

AY	prem	dev	LR	omega	theta	growth	reported	eval120	eval240	ULT	reserves
1988	2609	114	0.78	0.96	26.55	98.3%	2,036	2,017	2,045	2,045	9
1989	2694	102	0.75	0.96	26.55	97.4%	1,987	2,004	2,031	2,032	44
1990	2594	90	0.78	0.96	26.55	96.1%	1,919	2,001	2,028	2,029	110
1991	2609	78	0.58	0.96	26.55	94.1%	1,446	1,497	1,518	1,518	72
1992	2077	66	0.53	0.96	26.55	91.0%	1,007	1,088	1,103	1,103	95
1993	1703	54	0.49	0.96	26.55	86.2%	713	818	829	829	117
1994	1438	42	0.48	0.96	26.55	78.9%	533	688	697	697	164
1995	1093	30	0.53	0.96	26.55	67.5%	408	567	575	575	167
1996	1012	18	0.73	0.96	26.55	49.7%	352	725	735	735	384
1997	975.9	6	0.61	0.96	26.55	21.2%	122	587	595	596	474
total								11,991		12,159	**1,636**

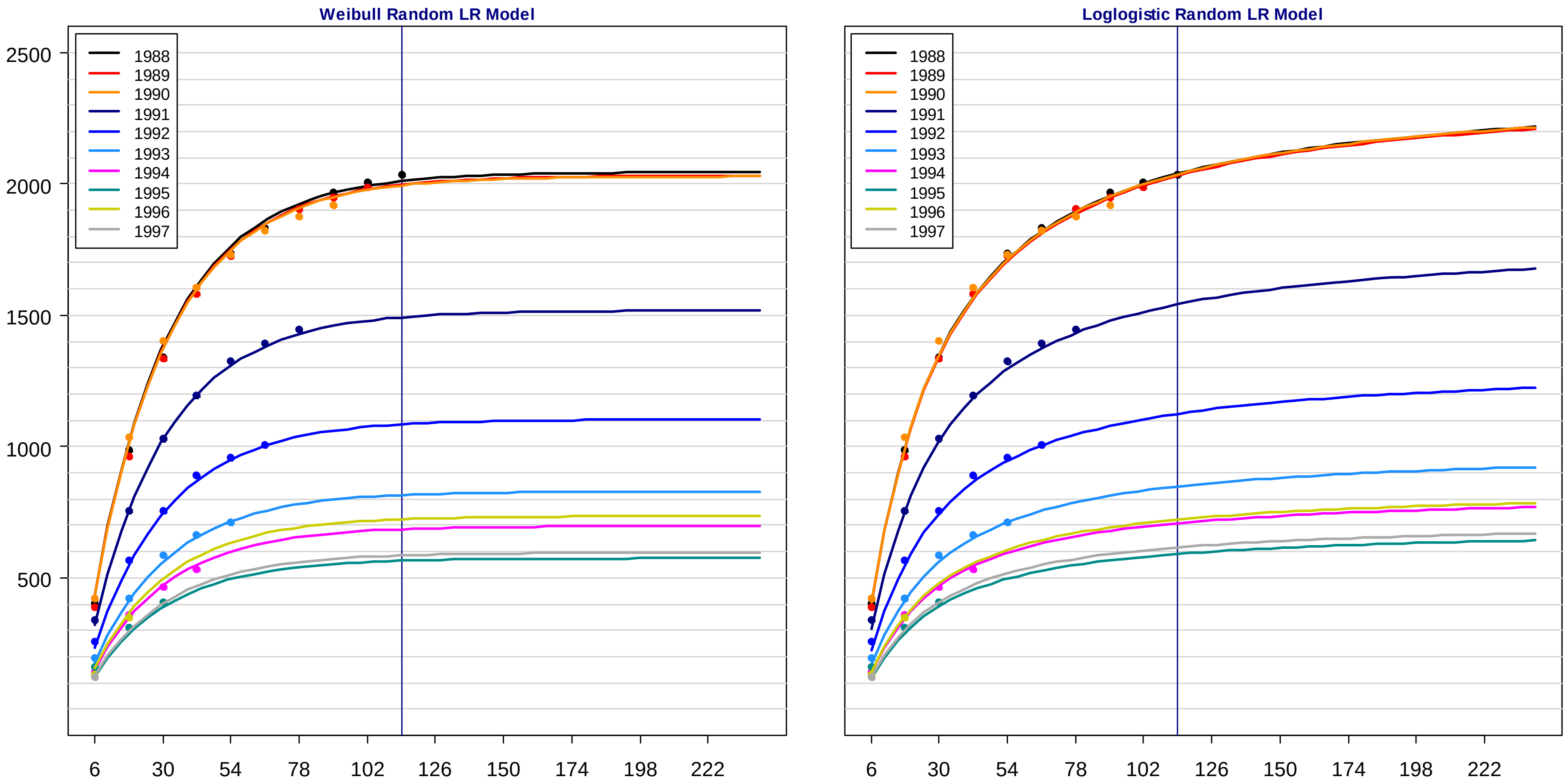
The overall Weibull reserve estimate is higher than that of the chain ladder because of "tail" development beyond 120 months

These are the 12 hierarchical model parameters.

Making Predictions is Difficult (Especially About the Future)

- Most loss reserve variability is due to "model risk"
- In this context, the most serious model risk is the choice of growth curves
- Both the Weibull and Loglogistic models fit the available data well
- But they extrapolate very differently

Comparison of Weibull vs Loglogistic Model Extrapolations



Bayes or Bust

"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's Working Party on Quantifying Variability in Reserve Estimates , 2004

- This can be read as a request for a Bayesian analysis
  - Bayesians (unlike frequentists) are willing to make **probability statements about unknown parameters**
  - **Ultimate losses are "single cases"** – difficult to conceive as random draws from a "sampling distribution in the sky".
    - Frequentist probability involved repeated trials of setups involving physical randomization.
    - In contrast it is meaningful to apply Bayesian probabilities to "single case events"
  - The Bayesian analysis yields an **entire posterior probability distribution** – not merely moment estimates

➔ Bayesian statistics is the ideal framework for loss reserving

- Remaining task: put prior probability distributions on model hyperparameters
- Hierarchical Bayes framework also allows us to bring in collateral information in the form of other loss triangles
  - Other companies, other regions, …
  - See Zhang-Dukic-Guszcza (2010)

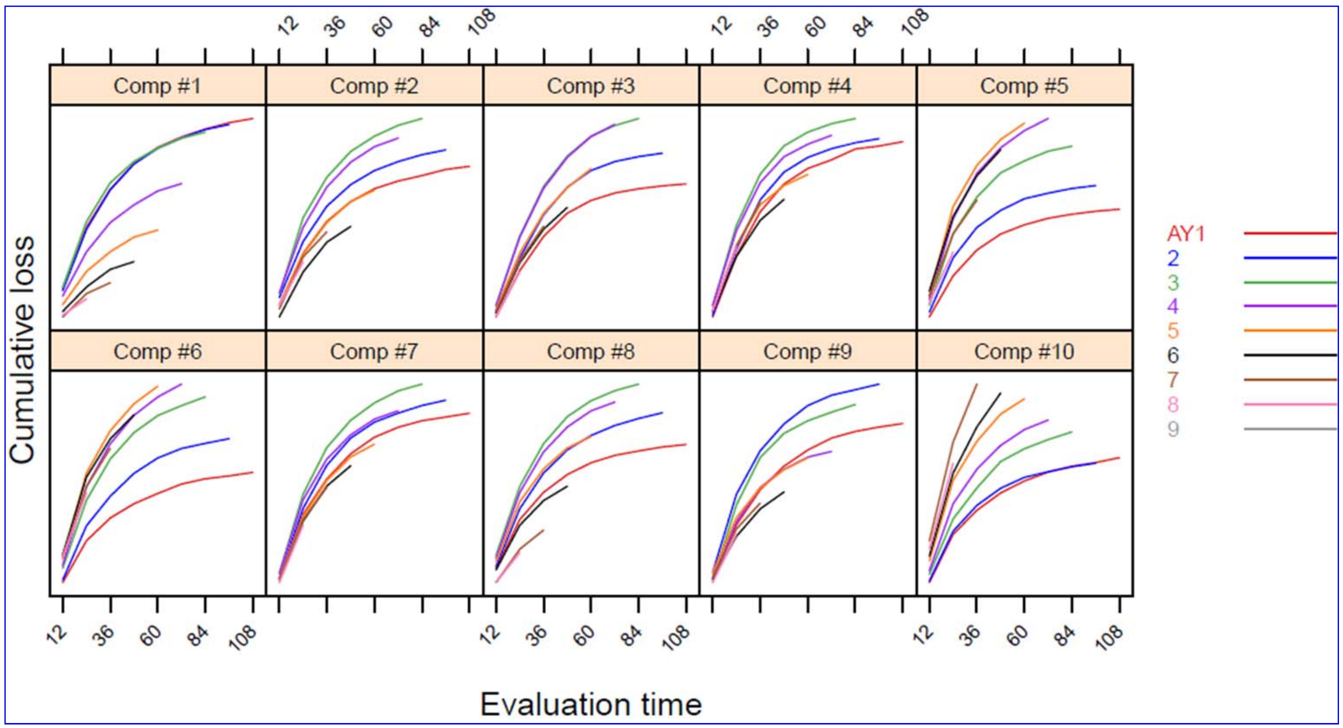
Hierarchical Bayes Growth Curve Loss Reserving Model

(Zhang-Dukic-Guszcza 2011)

Expanding the Analysis

- So let's do a fully Bayesian version of this model.
- While we're at it: let's add 9 more companies to the analysis. (Wayne's idea!)
  - Accident year (*i*) is already a "level" in our analysis
  - Company (*k*) is just another level
  - Further illustrates the power of the multilevel/hierarchical framework
- Use Schedule P Workers Comp data (1988-1997) for 10 companies
  - Development periods (12, 24, …, 120)
  - Set aside latest diagonal as holdout validation data

- This is a plot of the raw data
- 10 10x10 Δ's
- Visual inspection suggests that the Loglogistic is a better choice



The Nonlinear Hierarchical Bayes Model

- The basic idea remains the same
- Specify model on the log scale

$$\log y_{ik}(t_j) = \log \mu_{ik}(t_j) + \varepsilon_{ik}(t_j)$$

$$where:$$

$$\mu_{ik}(t_j) = p_{ik} \cdot \gamma_{ik} \cdot \frac{t_j^{\omega_k}}{t_j^{\omega_k} + \theta_k^{\omega_k}}$$

- "Growth curve" model form is the same
- We choose the log-logistic growth curve

- p_{ik} = premium for Accident Year *i*, company *k* (given)
- γ_{ik} = ultimate loss ratio for AY *i*, company *k*

➔ $p_{ik} \cdot \gamma_{ik}$ = ultimate $loss for AY *i*, company *k*

➔ $\mu_{ik}(t_j)$ = E[$loss] for AY *i*, company *k* at time t_j

The Nonlinear Hierarchical Bayes Model

- Add hierarchical structure, specify error structure, specify diffuse prior distributions.
- See Zhang-Dukic-Guszcza (2011) for more details.

$$\log y_{ik}(t_j) = \log \mu_{ik}(t_j) + \varepsilon_{ik}(t_j)$$
$$where:$$
$$\mu_{ik}(t_j) = p_{ik} \cdot \gamma_{ik} \cdot \frac{t_j^{\omega_k}}{t_j^{\omega_k} + \theta_k^{\omega_k}}$$

$$\varepsilon_{ik}(t_j) = \rho \cdot \varepsilon_{ik}(t_{j-1}) + \delta_{ik}(t_j)$$
$$\delta_{ik}(t_j) \sim N\left[0, \sigma_k^2 \cdot (1-\rho^2)\right]$$
$$\varepsilon_{ik}(t_0) \sim N(0, \sigma_k^2)$$

$$\rho \sim U(-1,1)$$
$$\sigma_k \sim U(0,100)$$

$$\log \gamma_{ik} \sim N\left(\gamma_k, \sigma_{\gamma.year}^2\right) \qquad \log(\gamma_k, \omega_k, \theta_k)' \sim N\left(\log(\gamma, \omega, \theta)', \Sigma\right)$$

$$\sigma_{\gamma.year} \sim U(0,100)$$
$$\log \gamma \sim N(0, 100^2)$$
$$\log \omega \sim N(0, 100^2)$$
$$\log \theta \sim N(0, 100^2)$$
$$\Sigma \sim Inv-Wishart_3(I)$$

MCMC from A to B

- Before 1990, Bayesian statistics was a lot more talk than action.
  - Unless you use conjugate priors, calculating posterior probability distributions is cumbersome at best, intractable at worst.

- Markov Chain Monte Carlo [MCMC]: a simulation technique used to solve high-dimensional integration problems.
  - Developed by physicists at Los Alamos
  - Introduced as a Bayesian computational technique by Gelfand and Smith in 1990
  - Gelfand and Smith sparked a renaissance in the practice of Bayesian statistics
  - As a result of the Gelfand and Smith MCMC approach, Bayesian statistics is now common in fields like:
    - Epidemiology, disease mapping
    - Political science, polling
    - Marketing science (hierarchical approach: customers are exchangeable units)
    - Genomics

- Rather than explicitly calculating high-dimensional, posterior densities, we estimate them by drawing repeated samples from them.
  - Construct a Markov chain whose equilibrium distribution is the posterior that we're trying to estimate.

References

Clark, David R. (2003) "LDF Curve Fitting and Stochastic Loss Reserving: A Maximum Likelihood Approach," CAS *Forum.*

Frees, Edward (2006). Longitudinal and Panel Data Analysis and Applications in the Social Sciences. New York: Cambridge University Press.

Gelfand, Alan and Adrian Smith (1990). "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*. http://home.gwu.edu/~stroud/classics/GelfandSmith90.pdf

Gelman, Andrew and Hill, Jennifer (2007). Data Analysis Using Regression and Multilevel / Hierarchical Models. New York: Cambridge University Press.

Guszcza, James. (2008). "Hierarchical Growth Curve Models for Loss Reserving," CAS *Forum*. http://www.casact.org/pubs/forum/08fforum/7Guszcza.pdf

Pinheiro, Jose and Douglas Bates (2000). Mixed-Effects Models in S and S-Plus. New York: Springer-Verlag.

Guszcza, James and Thomas Herzog (2010). "Enhanced Credibility: Actuarial Science and the Renaissance of Bayesian Thinking," *Contingencies*

Zhang, Yanwei, Vanja Dukic, James Guszcza (2011). "A Bayesian Nonlinear Model for Forecasting Insurance Payments," *JRSS A*.