Actuarial Applications of Hierarchical Modeling

CAS RPM Seminar
Chicago
March, 2010

Jim Guszcza
Deloitte Consulting LLP

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.

Topics

Hierarchical Model Theory

Hierarchical Model Theory

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
  - Income by profession
  - Claim severity by injury type
  - Churn rate by agency
  - Multiple years of loss experience by policyholder.
  - …

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

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

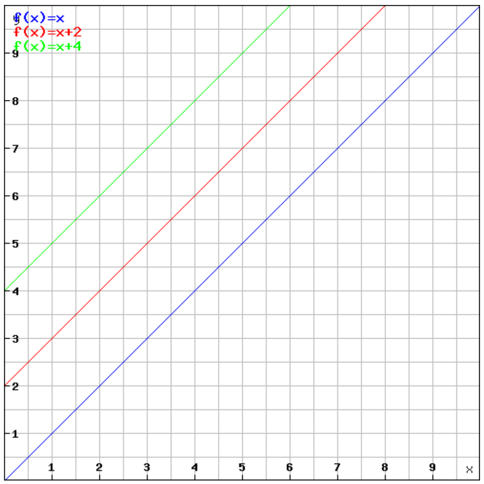
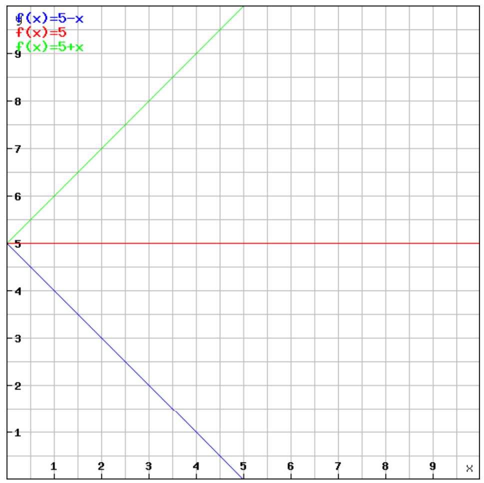
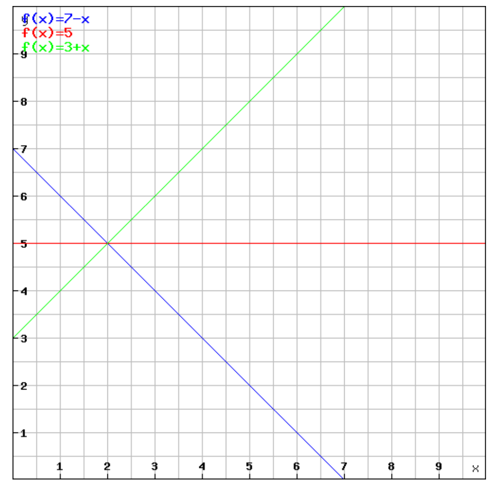
What's in a Name?

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

- I prefer the "hierarchical/multilevel model" terminology because it evokes the way models-within-models are used to reflect levels-within-levels of ones data.

- An important special case of hierarchical models involves multiple observations through time of each unit.
  - Here group membership is the repeated observations belonging to each individual.
  - Time is the covariate.

Varying Slopes and Intercepts

Random Intercept Model	Random Slope Model	Random Intercept / Random Slope Model
		

Random Intercept Model:
- Intercept varies with group
- Slope stays constant

Random Slope Model:
- Intercept stays constant
- Slope varies by group

Random Intercept / Random Slope Model:
- Intercept and slope vary by group

- Each line represents a different group

Common Hierarchical Models

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

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

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

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

Parameters and Hyperparameters

- We can rewrite the random intercept model this way:

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

- Suppose there are 100 levels: $j = 1, 2, \ldots, 100$ (e.g. SIC bytes 1-2)
- This model contains 101 parameters: $\{\alpha_1, \alpha_2, \ldots, \alpha_{100}, \beta\}$.
- And it contains 4 <u>hyper</u>parameters: $\{\mu_\alpha, \beta, \sigma, \sigma_\alpha\}$.

- Here is how the hyperparameters relate to the parameters:

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

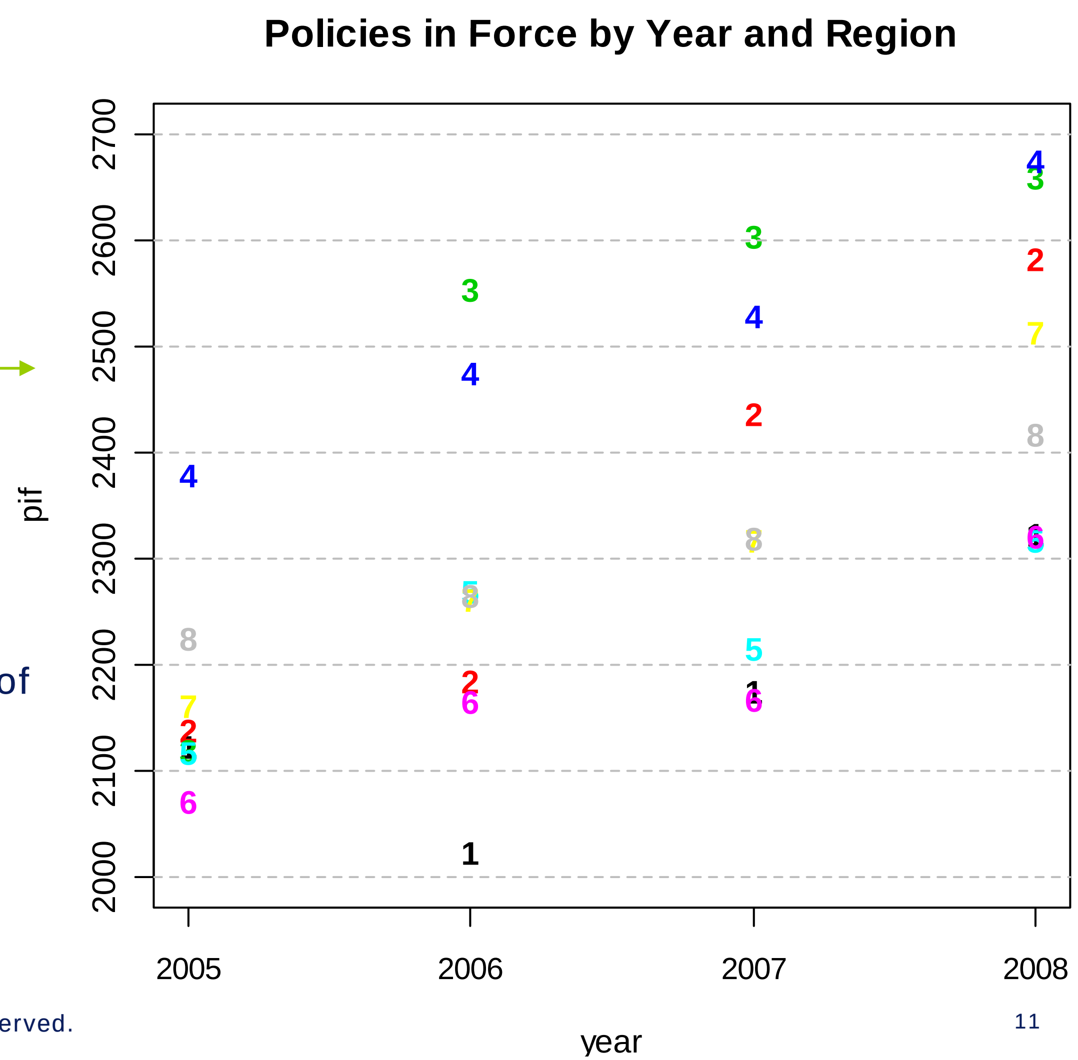
- Does this formula look familiar?

Deloitte.

Sample Hierarchical Model

Example

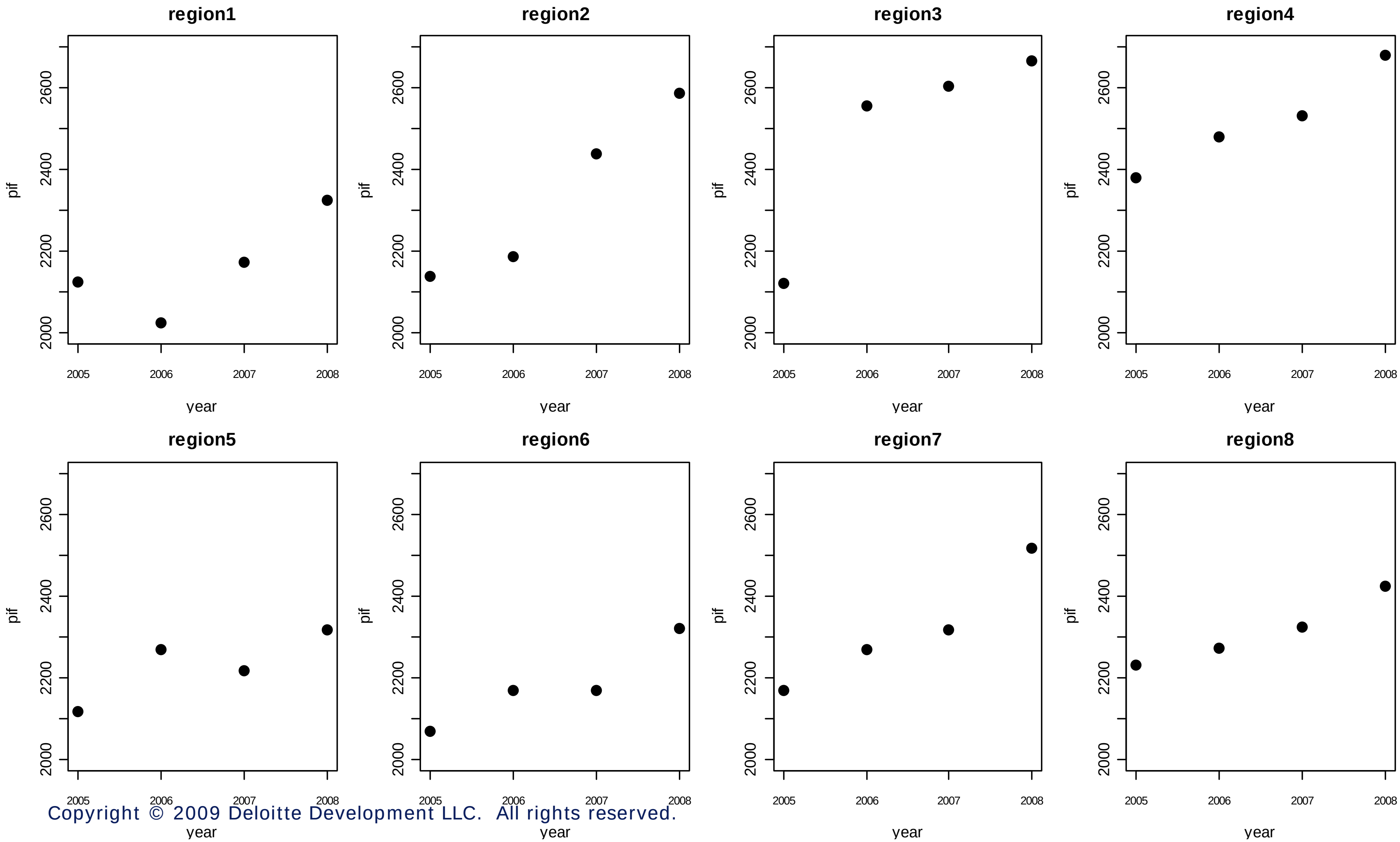
- Suppose we wish to model a company's policies in force, by region, for the years 2005-08.
- 8 * 4 = 32 data points.



- One way to visualize the data:
  - Plot all of the data points on the same graph, use different colors/symbols to represent region.
- Alternate way:
  - Use a trellis-style display, with one plot per region
  - More immediate representation of the data's hierarchical structure.
  - (see next slide)

Trellis-Style Data Display

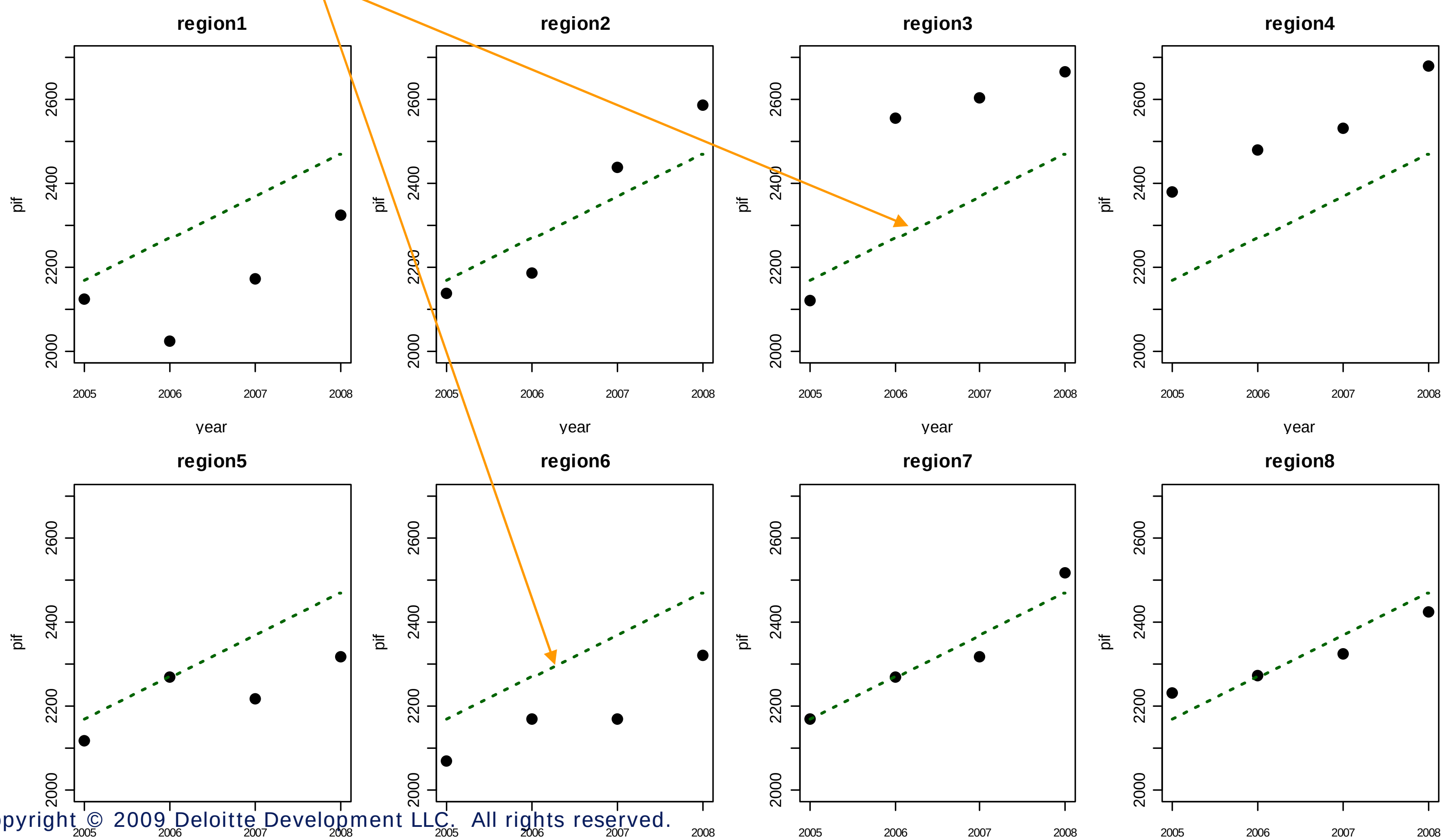
- We wish to build a model that captures the change in PIF over time.
- We must reflect the fact that PIF varies by region.



Option 1: Simple Regression

- The easiest thing to do is to pool the data across groups -- **i.e. simply ignore region**
- Fit a simple linear model
- Alas, this model is <u>not</u> appropriate for all regions

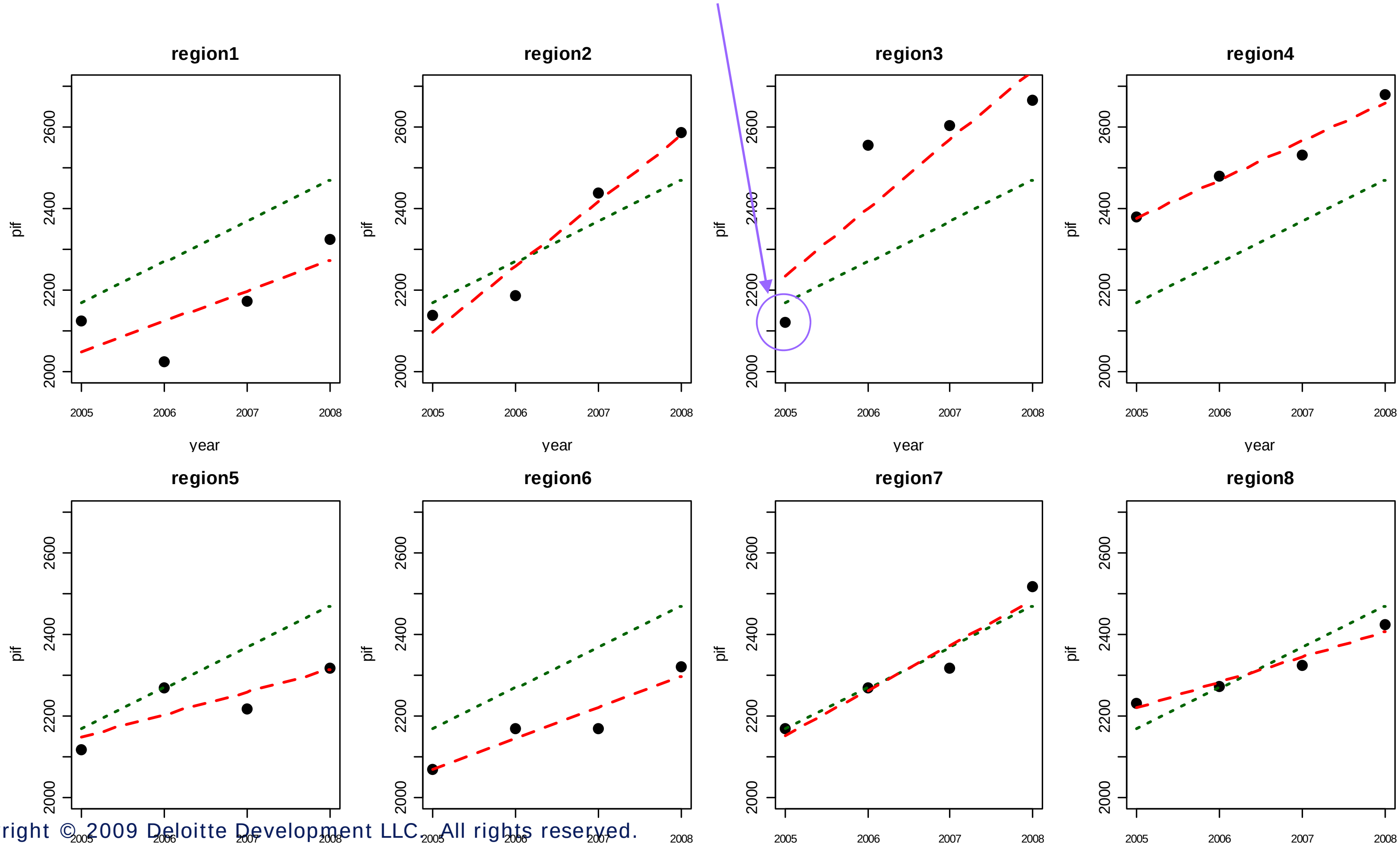
$$PIF = \alpha + \beta t + \varepsilon$$



Option 2: Separate Models by Region

- At the other extreme, we can fit a separate simple linear model for each region.
- Each model is fit with 4 data points.
- Introduces danger of over-fitting the data.

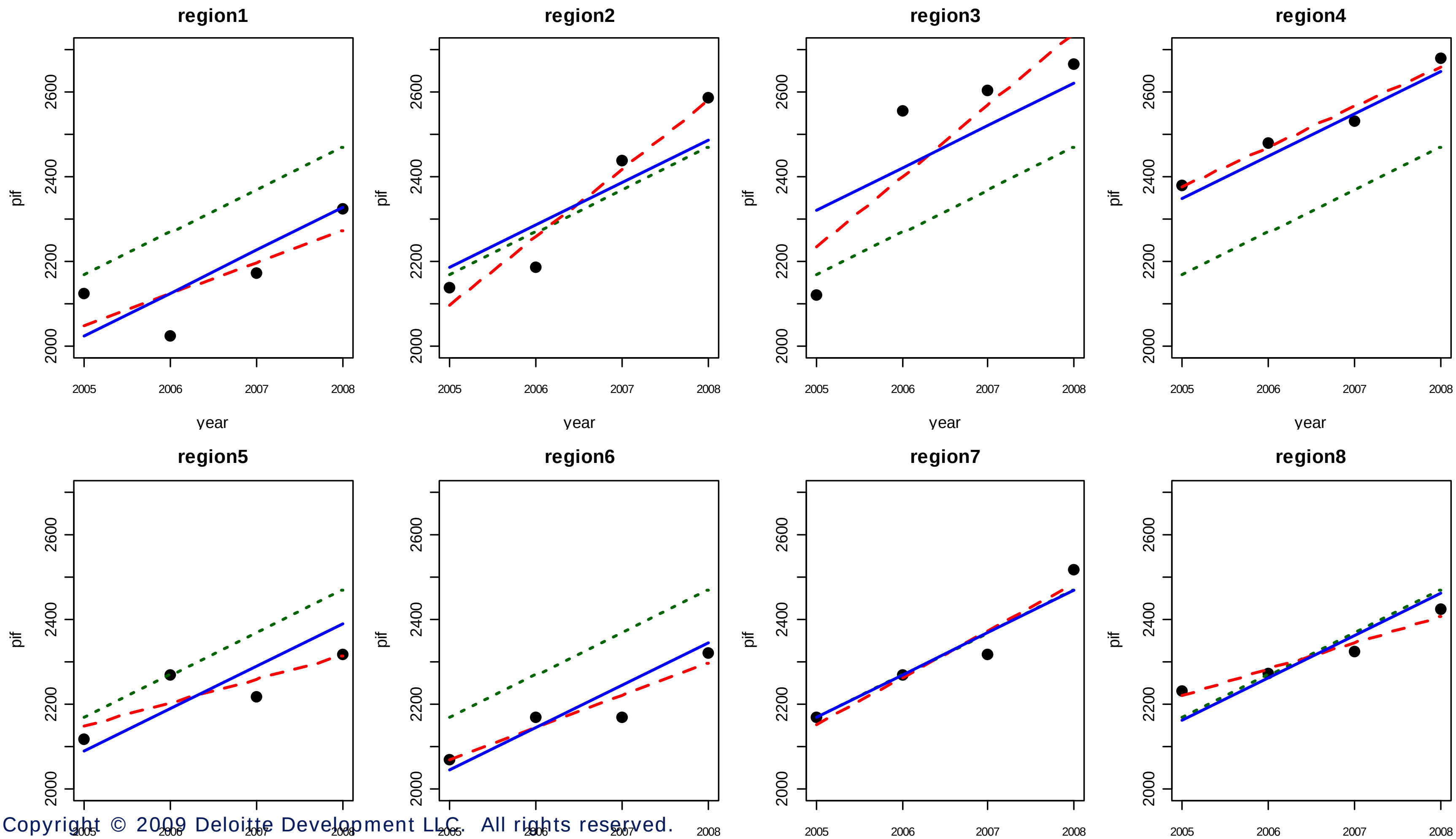
$$\left\{ PIF = \alpha^k + \beta^k t + \varepsilon^k \right\}_{k=1,2,..,8}$$



Option 3: Random Intercept Hierarchical Model

- Compromise: Reflect the region group structure using a hierarchical model.

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



Compromise Between Complete Pooling & No Pooling

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

Complete Pooling

- Ignore group structure altogether

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

No Pooling

- Estimating one model for each group

Compromise

Hierarchical Model

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

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

Option 1b: Adding Dummy Variables

- Question: of course it'd be crazy to fit a separate SLR for each region.
- But what about adding 8 region dummy variables into the SLR?

$$PIF = \gamma_1 R_1 + \gamma_2 R_2 + \ldots + \gamma_8 R_8 + \beta t + \varepsilon$$

- If we do this, we need to estimate 9 parameters instead of 2.

- In contrast, the random intercept model contains 4 <u>hyper</u>parameters: μ_α, β, σ, σ_α

- Now suppose our example contained 800 regions. If we use dummy variables, our SLR potentially requires that we estimate 801 parameters.

- But the random intercept model will contain the same 4 hyperparameters.

Varying Slopes

- The random intercept model is a compromise between a “pooled” SLR and a separate SLR by region.

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

- But there is nothing sacred about the intercept term: **we can also allow the slopes to vary by region**.

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

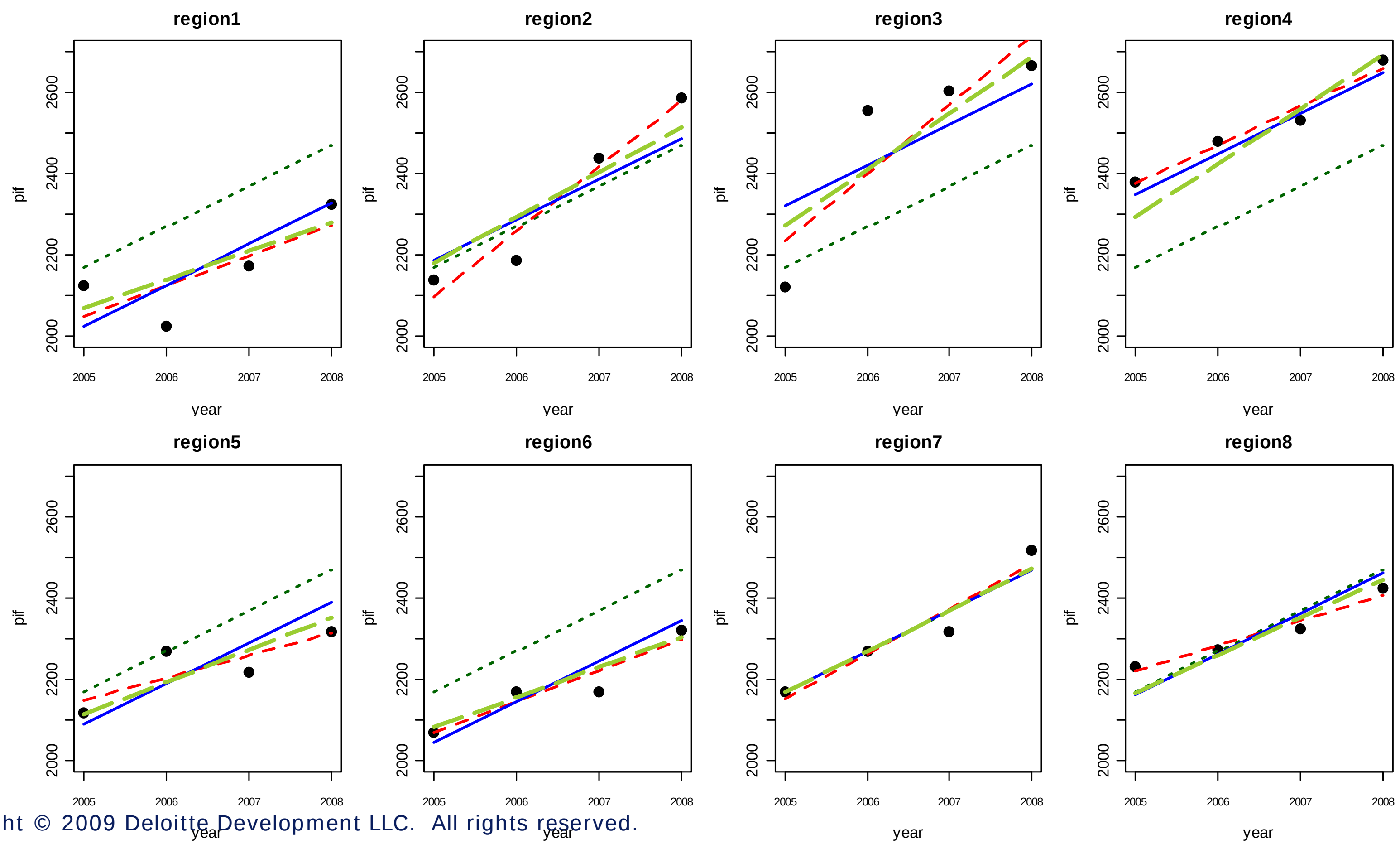
- In the dummy variable option (1b) this would require us to interact region with the time *t* variable… i.e. it would return us to option 2.
  - Great danger of overparameterization.

- Adding random slopes adds considerable flexibility at the cost of only two additional hyperparameters.
  - Random slope only: μ_α, β, σ, σ_α
  - Random slope & intercept: μ_α, μ_β, σ, σ_α, σ_β, $\sigma_{\alpha\beta}$

Option 4: Random Slope & Intercept Hierarchical Model

- We can similarly include a sub-model for the slope β.

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



Does Adding Random Slopes Improve the Model?

- How do we determine whether adding the random slope term improves the model?

1. Graphical analysis and judgment:
    - the random slopes arguably yield an improved fit for Region 5.
    - but it looks like the random slope model might be overfitting Region 3.
    - Other regions a wash

2. Out of sample lift analysis.

3. Akaike information Criterion [AIC]: -2*LL + 2*d.f.
    - Random intercept AIC: 380.40
    - Random intercept & slope AIC: 380.64
    - Slight deterioration ➔ better to select the random intercept model.

- Random slopes don't help in this example, but it is a very powerful form of variable interaction to consider in one's modeling projects.

Parameter Comparison

- It is important to distinguish between each model's *parameters* and *hyperparameters*.

	α, β		$\mu_\alpha, \beta, \sigma, \sigma_\alpha$		$\mu_\alpha, \mu_\beta, \sigma, \sigma_\alpha, \sigma_\alpha, \sigma_{\alpha\beta}$	
	SLR		random intercept		random intercept & slope	
region	intercept	slope	intercept	slope	intercept	slope
1	2068.0	100.1	1911.3	100.1	1999.3	70.3
2	2068.0	100.1	2087.8	100.1	2070.2	111.2
3	2068.0	100.1	2236.1	100.1	2137.0	137.4
4	2068.0	100.1	2267.3	100.1	2159.6	133.2
5	2068.0	100.1	1980.3	100.1	2033.1	79.3
6	2068.0	100.1	1932.3	100.1	2008.9	73.8
7	2068.0	100.1	2066.8	100.1	2066.3	101.2
8	2068.0	100.1	2061.8	100.1	2069.5	94.1

- SLR: 2 parameters and 2 hyperparameters
- Random intercept: 9 parameters and 4 hyperparameters
- Random intercept & slope: 16 parameters and 6 hyperparameters

- **How do the hyperparameters relate to the parameters?**

Connection with Credibility Theory

Hierarchical Models and Credibility Theory

- Let's revisit the random intercept model.

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

- This is how we calculate the random intercepts $\{\alpha_1, \alpha_2, \ldots, \alpha_8\}$:

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

- Therefore: each random intercept is a **credibility-weighted average** between:
  - The intercept for the pooled model (option 1)
  - The intercept for the region-specific model (option 2)
  - As $\sigma_\alpha \to 0$, the random intercept model $\to$ option 1 (complete pooling)
  - As $\sigma_\alpha \to \infty$, the random intercept model $\to$ option 2 (separate models)

Bühlmann's Credibility and Random Intercepts

- If we remove the time covariate (t) from the random intercepts model, we are left with a very familiar formula:

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

- **Therefore: Bühlmann's credibility model is a specific instance of hierarchical models.**

- Similarly for Bühlmann-Straub and Hachemeister.

- **Hierarchical models give one a practical way to integrate credibility theory into one's GLM modeling activities.**

Example: The Hachemeister Data

- Number of claims & average severity for 5 states, over 12 quarters.
- Rich structure allows us to fit the three classic credibility models.
  - Bühlmann — random intercept, no weight
  - Bühlmann-Straub — random intercept, weighted
  - Hachemeister — random slope & intercept model, weighted
- These are three special cases of hierarchical models.

Hachemeister Claim Severity Data

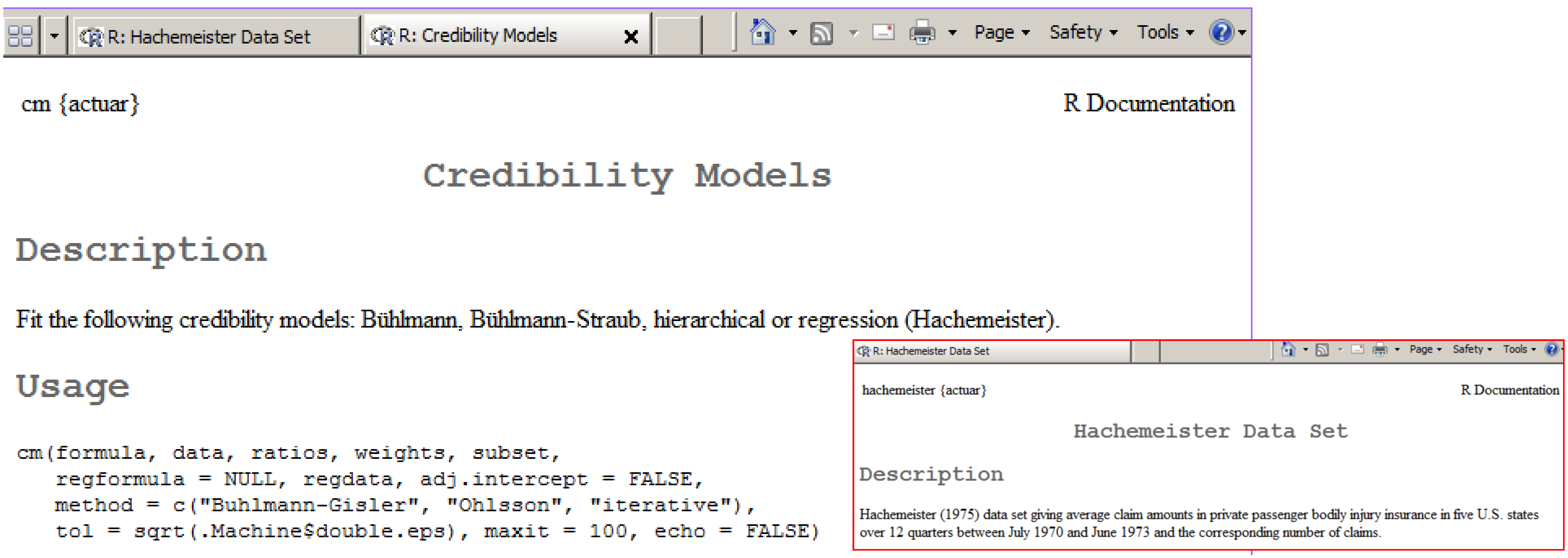
	state 1		state 2		state 3		state 4		state 5		combined	
quarter	#claims	severity	#claims	severity	#claims	severity	#claims	severity	#claims	severity	#claims	severity
1	7,861	1,738	1,622	1,364	1,147	1,759	407	1,223	2,902	1,456	13,939	1,622
2	9,251	1,642	1,742	1,408	1,357	1,685	396	1,146	3,172	1,499	15,918	1,579
3	8,706	1,794	1,523	1,597	1,329	1,479	348	1,010	3,046	1,609	14,952	1,690
4	8,575	2,051	1,515	1,444	1,204	1,763	341	1,257	3,068	1,741	14,703	1,882
5	7,917	2,079	1,622	1,342	998	1,674	315	1,426	2,693	1,482	13,545	1,827
6	8,263	2,234	1,602	1,675	1,077	2,103	328	1,532	2,910	1,572	14,180	2,009
7	9,456	2,032	1,964	1,470	1,277	1,502	352	1,953	3,275	1,606	16,324	1,836
8	8,003	2,035	1,515	1,448	1,218	1,622	331	1,123	2,697	1,735	13,764	1,853
9	7,365	2,115	1,527	1,464	896	1,828	287	1,343	2,663	1,607	12,738	1,893
10	7,832	2,262	1,748	1,831	1,003	2,155	384	1,243	3,017	1,573	13,984	2,024
11	7,849	2,267	1,654	1,612	1,108	2,233	321	1,762	3,242	1,613	14,174	2,027
12	9,077	2,517	1,861	1,471	1,121	2,059	342	1,306	3,425	1,690	15,826	2,156

Note on Software

- Note that the Hachemeister data is included as part of Vincent Goulet's "actuar" R package.

- Actuar also contains a function cm() that computes the Bühlmann and Hachemeister credibility models.

- It is possible to replicate cm() credibility model results using the standard hierarchical modeling function lmer().



cm {actuar} R Documentation

Credibility Models

Description

Fit the following credibility models: Bühlmann, Bühlmann-Straub, hierarchical or regression (Hachemeister).

Usage

```
cm(formula, data, ratios, weights, subset,
   regformula = NULL, regdata, adj.intercept = FALSE,
   method = c("Buhlmann-Gisler", "Ohlsson", "iterative"),
   tol = sqrt(.Machine$double.eps), maxit = 100, echo = FALSE)
```

hachemeister {actuar} R Documentation

Hachemeister Data Set

Description

Hachemeister (1975) data set giving average claim amounts in private passenger bodily injury insurance in five U.S. states over 12 quarters between July 1970 and June 1973 and the corresponding number of claims.

Bühlmann Model

```
> predict(cm(~state, hachemeister, ratios=ratio.1:ratio.12)
[1] 2044.041 1518.588 1814.234 1375.987 1602.233
> h1 <- lmer(severity ~ (1|state), data=ddat)
> pred.h1 <- ranef(h1)$state[[1]] + fixef(h1) ; pred.h1
[1] 2044.096 1518.565 1814.255 1375.944 1602.223
```

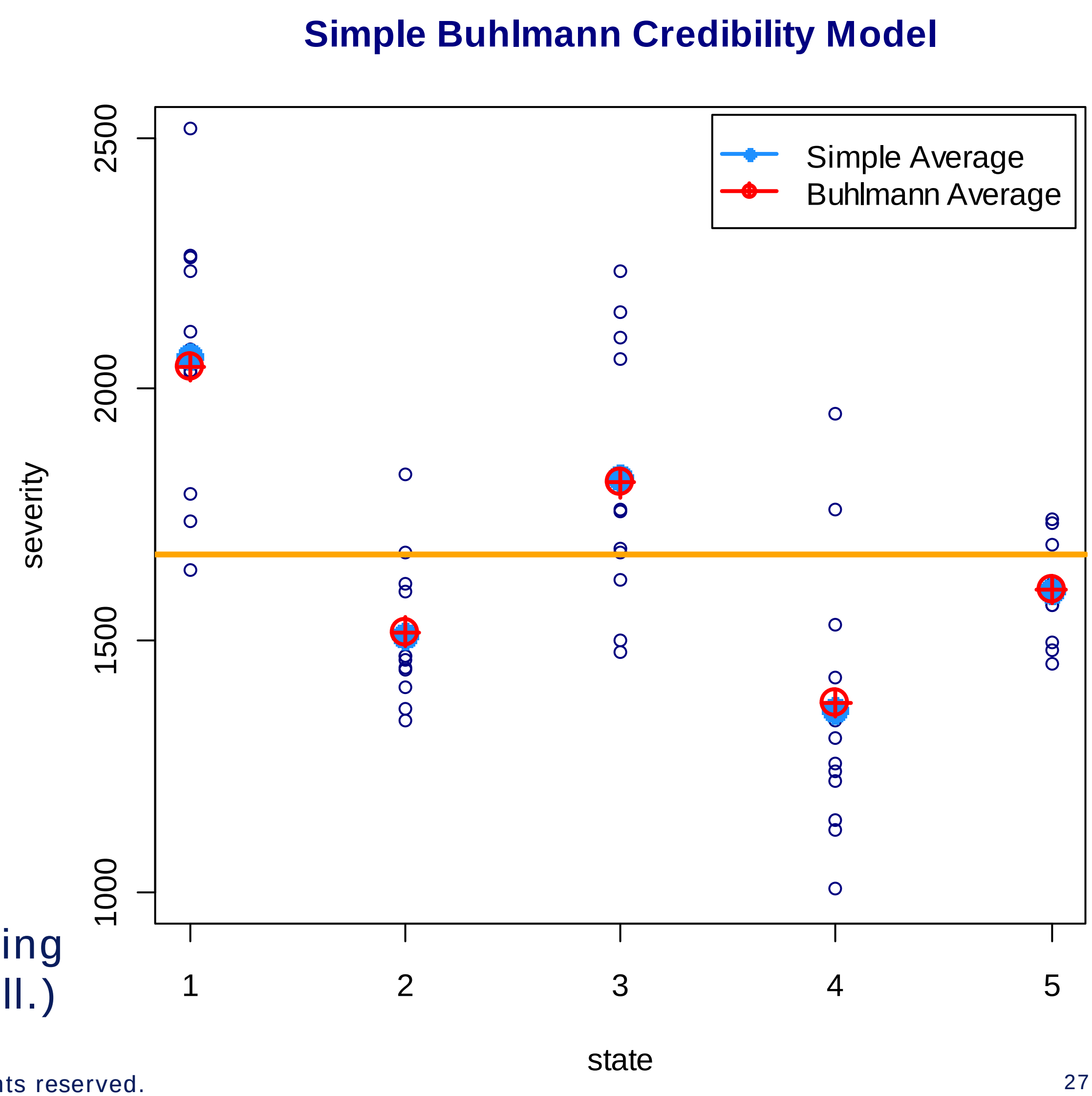
- To illustrate the Bühlmann model, we ignore the time structure.

- **No pooling:**
  - Regression model with 5 state dummy variables
  - i.e. simple average

- **Bühlmann credibility weighted estimates:**
  - Hierarchical model with random intercept.

$$\boxed{\begin{gathered} SEV \sim N(\alpha_{j[i]}, \sigma^2) \\ \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2) \end{gathered}}$$

- (These models are not appropriate because of varying number of claims in each cell.)



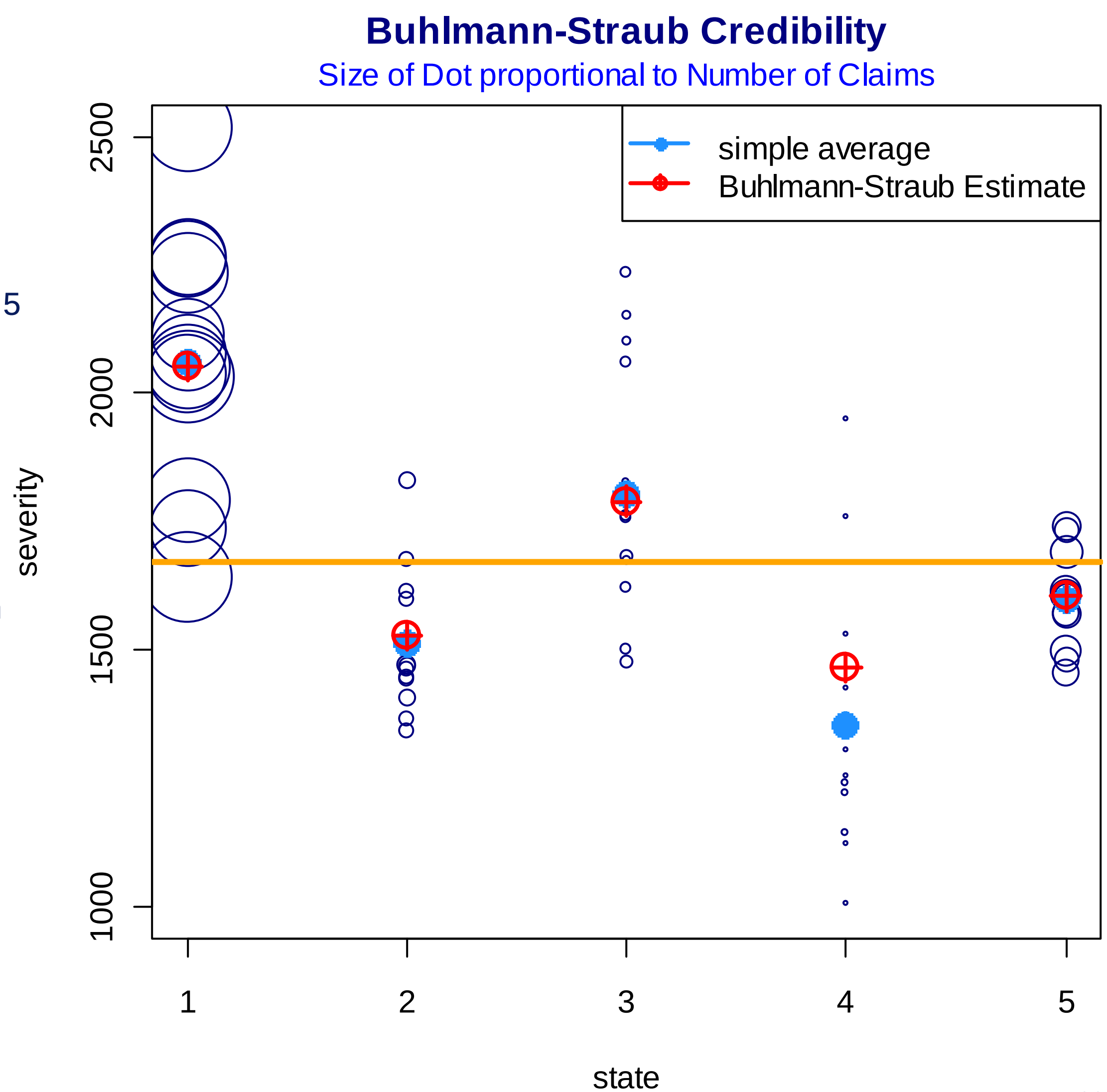
Bühlmann-Straub Model

- We continue to ignore the time structure but reflect the varying number of claims in each cell.

- **No pooling:**
  - Weighted Regression model with 5 state dummy variables
  - i.e. weighted averages

- **Bühlmann-Straub estimates:**
  - Weighted Hierarchical model with random intercept.

$$SEV \sim N\left(\alpha_{j[i]}, \sigma^2 / clmcnt_i\right)$$
$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$$



Hachemeister Model

- Finally we reflect the time structure of the data.

- **No pooling:**
  - Separate Regressions of severity on time

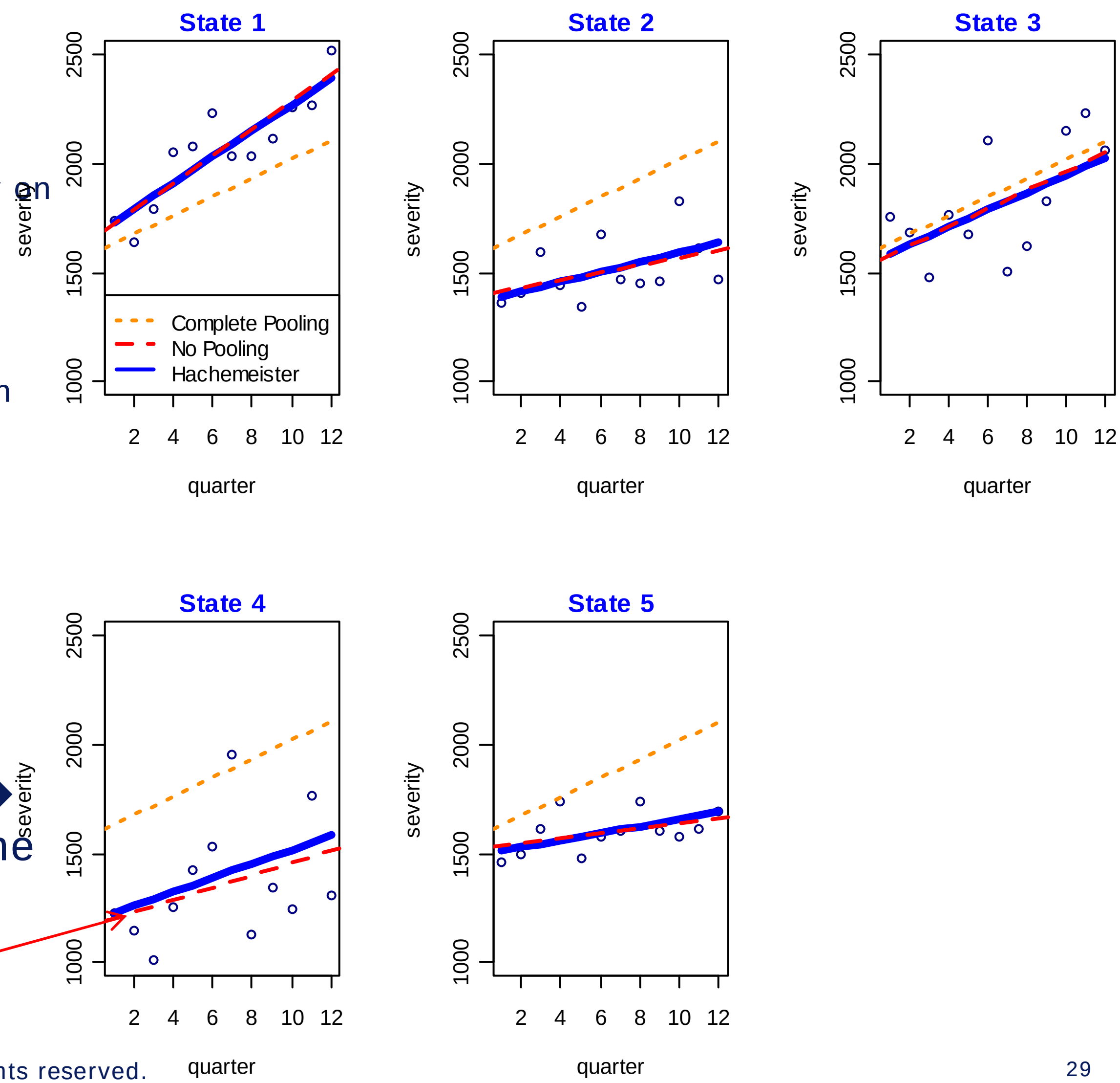
- **Hachemeister:**
  - Weighted Hierarchical model with random intercept and slope.

$$SEV \sim N\left(\alpha_{j[i]} + \beta_{j[i]} * t, \left(\sigma^2 / clmcnt_i\right)\right)$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \Sigma\right)$$

- State 4 has the least data ➔ the slope and intercept of the state 4 credibility-weighted regression line are most effected.

Corrected Hachemeister Model



Sample Applications

- Territorial ratemaking or including territory in a GLM analysis.
  - The large number of territories typically presents a problem.
- Vehicle symbol analysis
- WC or Bop business class analysis
- Repeated observations by policyholder
- Experience rating
- Loss reserving
  - Short introduction to follow

Summing Up

- Hierarchical models are applicable when one's data comes grouped in one or more important ways.

- A group with a large number of levels might be regarded as a "massively categorical value"...
  - Building separate models by level or including one dummy variable per level is often impractical or unwise from a credibility point of view.

- Hierarchical models offer a compromise between complete pooling and separate models per level.

- This compromise captures the essential idea of credibility theory.

- **Therefore hierarchical model enable a practical unification of two pillars of actuarial modeling:**
  - **Generalized Linear Models**
  - **Credibility theory**

Other thoughts

- The "credibility weighting" reflected in the calculation of the random effects represents a "shrinkage" of group-level parameters (α_j, β_j) to their means (μ_α, μ_β).
- The lower the "between variance" (σ_α^2) the greater amount of "shrinkage" or "pooling" there is.
- There is more shrinkage for groups with fewer observations (*n*).
- Panel data analysis is a type of hierarchical modeling ➔ this is a natural framework for analyzing longitudinal datasets.
  - Multiple observations of the same policyholder
  - Loss reserving: loss development is multiple observations of the same AY claims

Case Study
Hierarchical Poisson Regression

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             0.060605   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

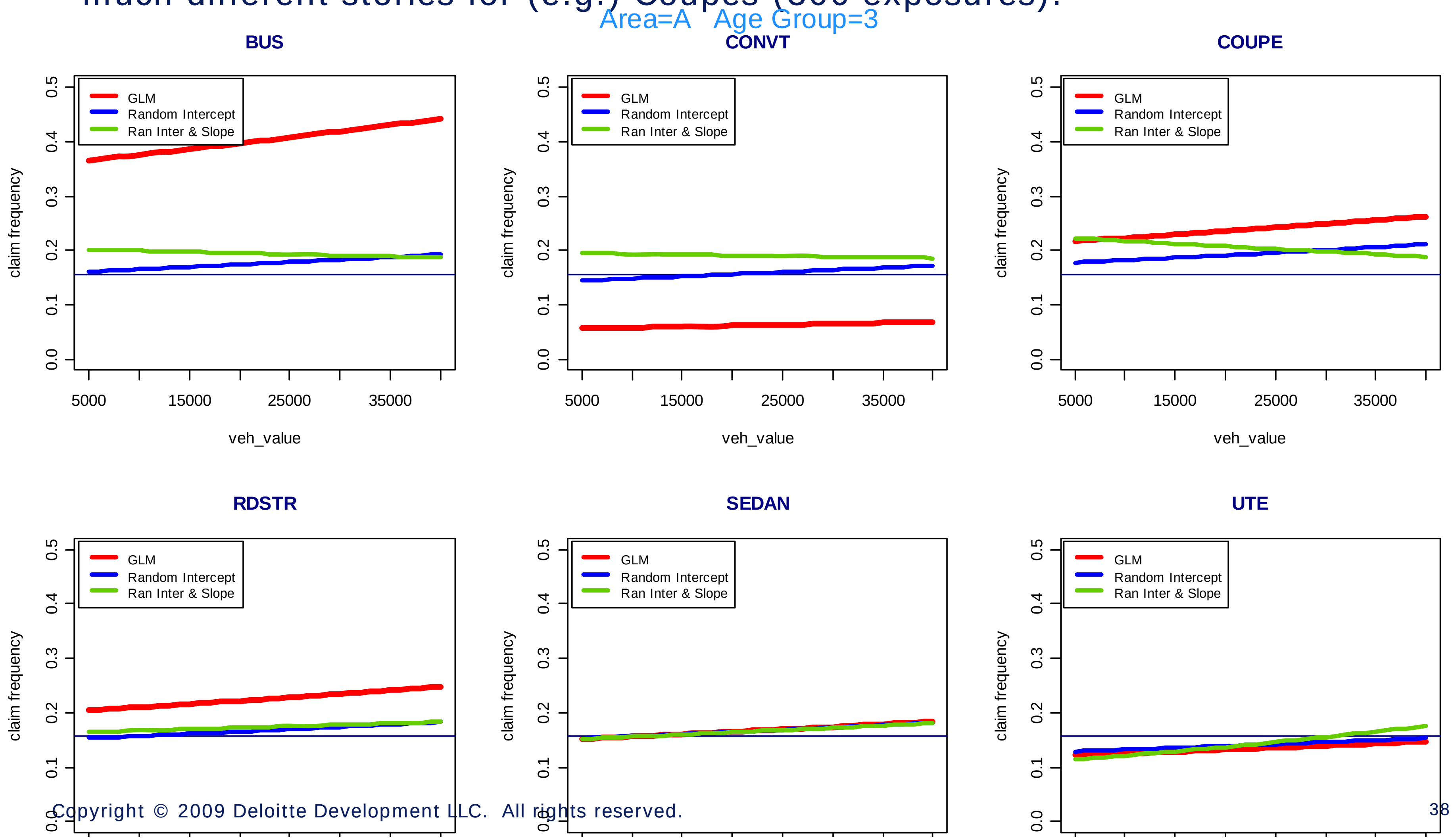
- **Intuition:** Relationship between vehicle value and claim frequency might vary by type of vehicle.
- **Response:** Introduce random slopes for VEH_VALUE.

```
> 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).



References

de Jong, Piet and Heller, Gillian (2008). Generalized Linear Models for Insurance Data. New York: Cambridge University Press.

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

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*.

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