Multi-something Regression Models

Brian A. Fannin

March 23, 2014

The loss triangle

AY	12	24	36	48	60	72	84
2002	2,318	7,932	13,822	22,095	31,945	40,629	44,437
2003	1,743	6,240	12,683	22,892	34,505	39,320	
2004	2,221	9,898	25,950	43,439	52,811		
2005	3,043	12,219	27,073	40,026			
2006	3,531	11,778	22,819				
2007	3,529	11,865					
2008	3,409						

Dimensions: two

Variables: just the one

Construction of model factors

- Diligent copying of formulae
- Use of the OFFSET function
- VBA
- Change the data

The most important 3 words in the history of the actuarial profession ever

"Abandon your triangles!"

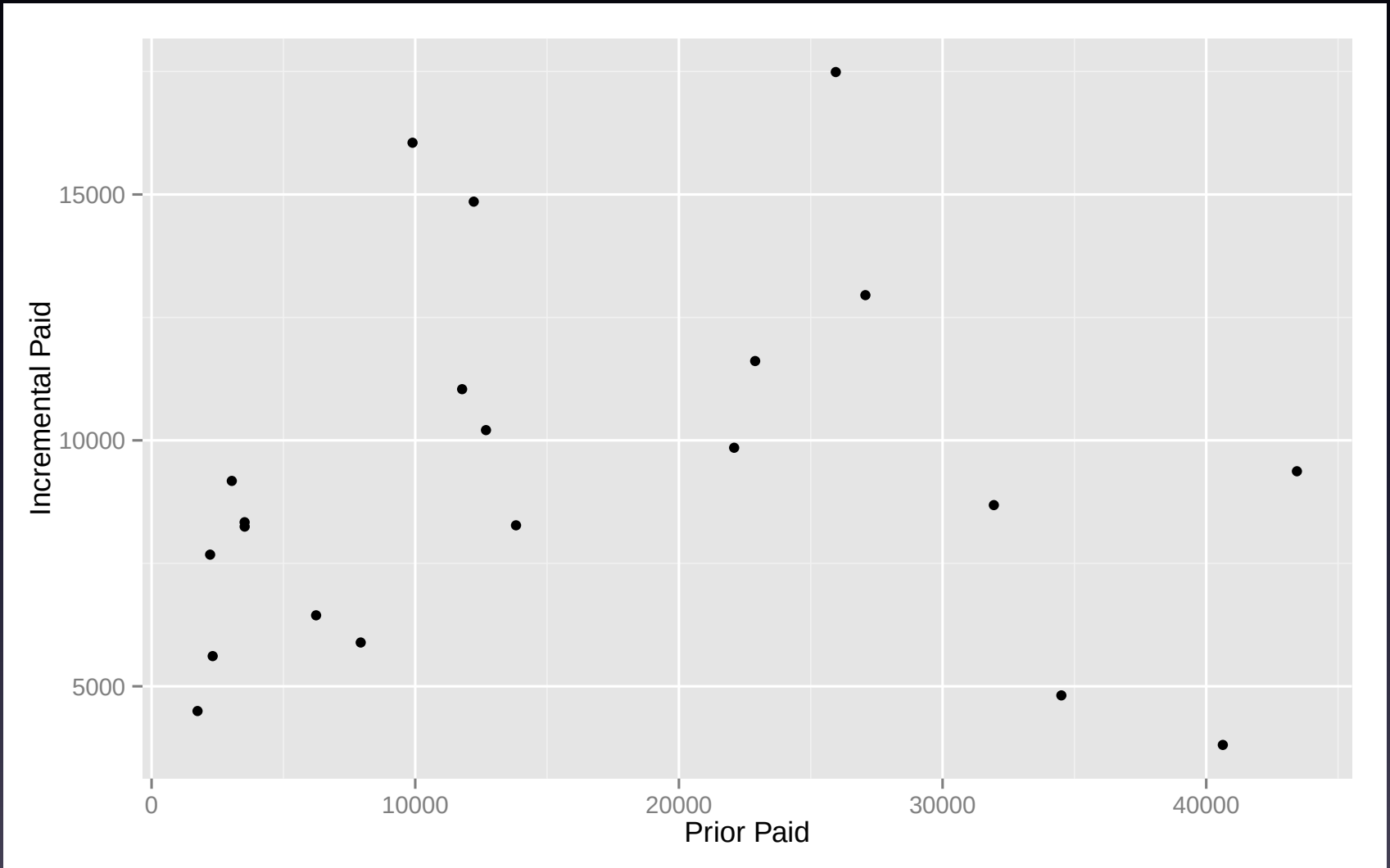
- Dave Clark

The long format

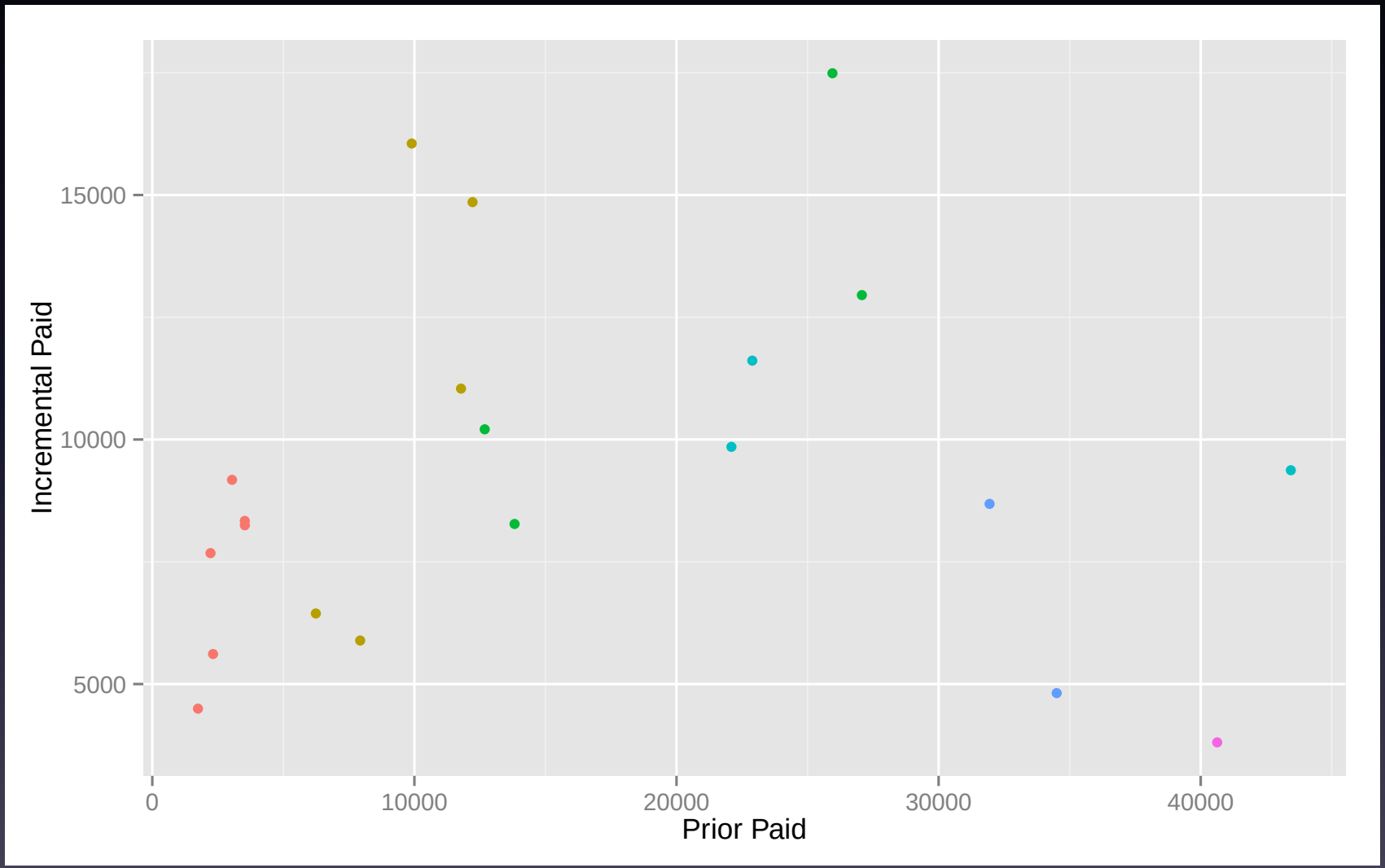
	A	B	C
1	AY	Month	Paid
2	2002	12	2,318
3	2003	12	1,743
4	2004	12	2,221
5	2005	12	3,043
6	2006	12	3,531
7	2007	12	3,529
8	2008	12	3,409
9	2002	24	7,932
10	2003	24	6,240
11	2004	24	9,898
12	2005	24	12,219
13	2006	24	11,778
14	2007	24	11,865
15	2002	36	13,822
16	2003	36	12,683
17	2004	36	25,950
18	2005	36	27,073
19	2006	36	22,819
20	2002	48	22,095
21	2003	48	22,892
22	2004	48	43,439
23	2005	48	40,026
24	2002	60	31,945
25	2003	60	34,505
26	2004	60	52,811
27	2002	72	40,629
28	2003	72	39,320
29	2002	84	44,437

Dimensions: two?
Variables: ??

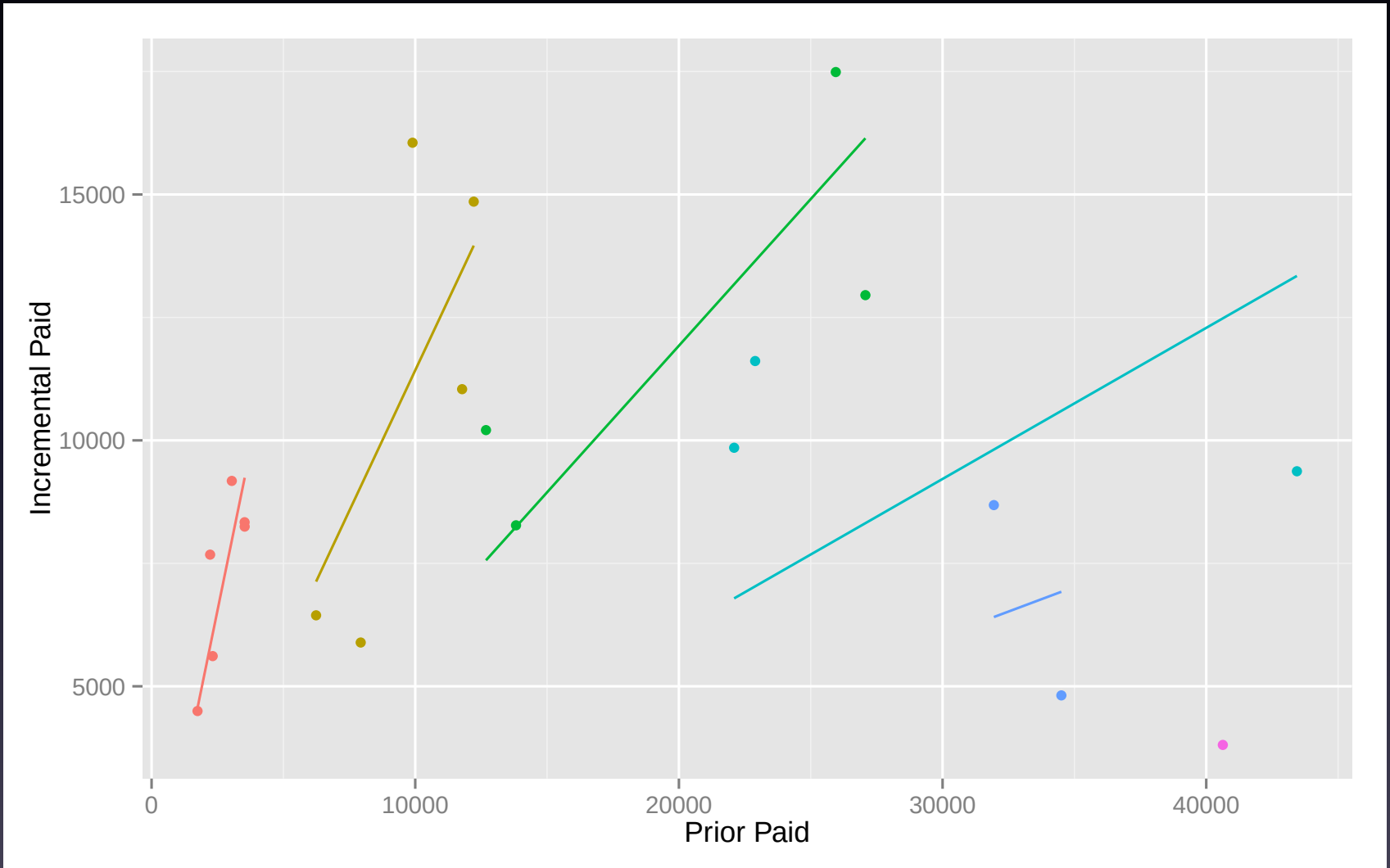
Plotting a long triangle



Grouping the plot



A plot looks like a model

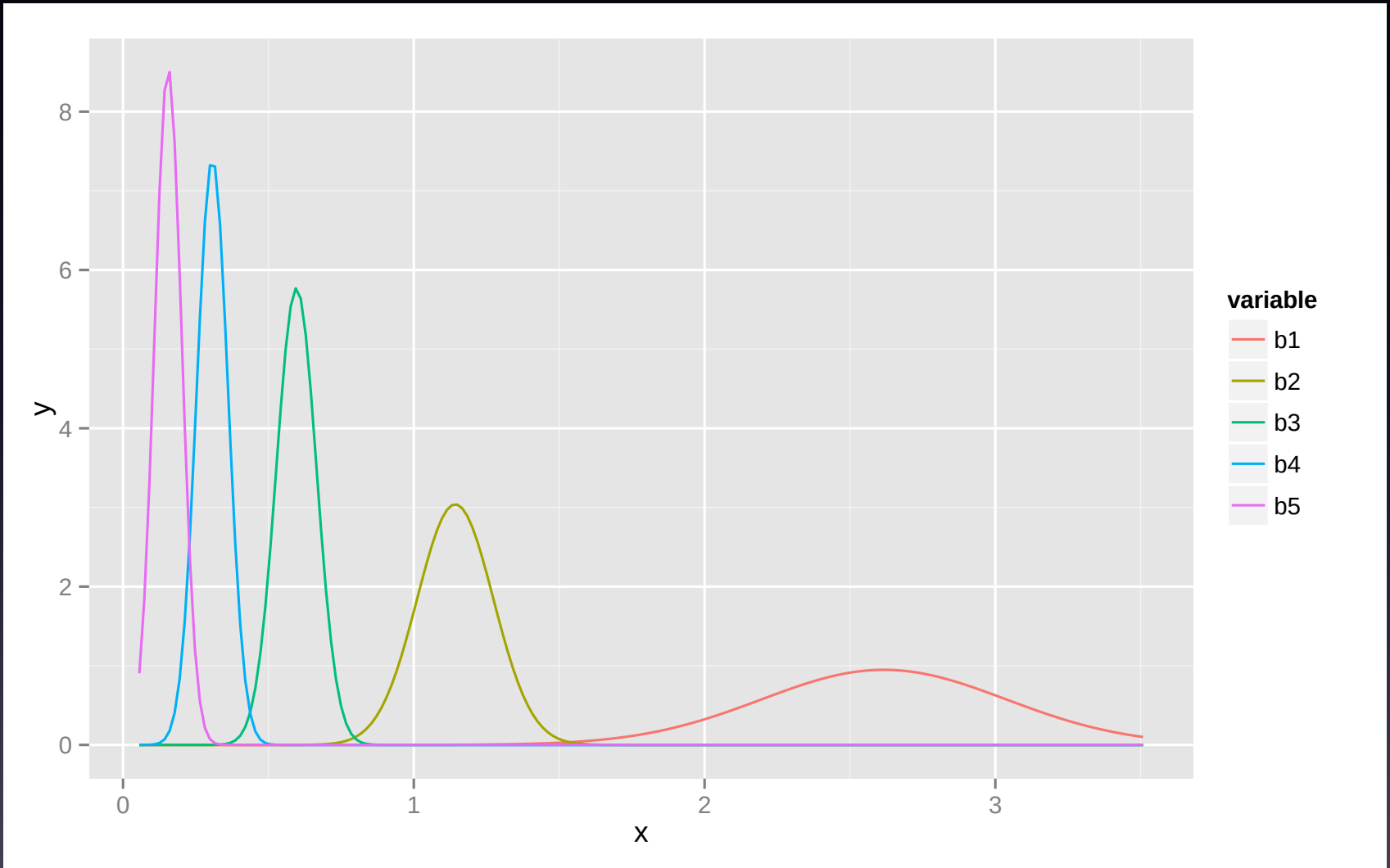


Most reserving methods are linear models

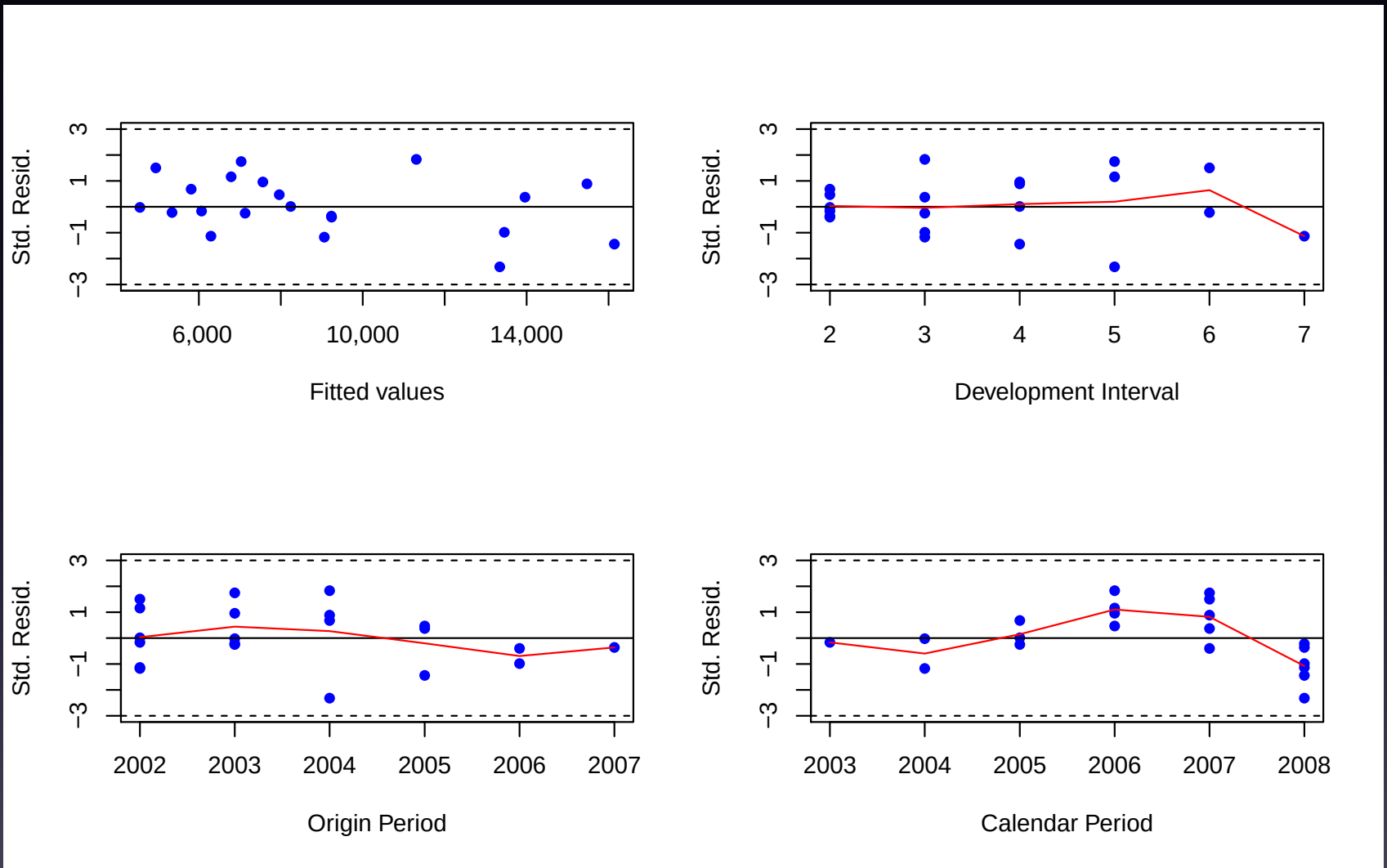
$$Y = \alpha + \beta_1 X_1 + \varepsilon$$

1. Linear model with specified parameters
   - Significance of individual model factors
   - Significance of model
2. Functional form of errors
3. Independence of errors
   - (Serial) correlation of errors
   - Homoskedasticity

Link ratio estimates are variables



Residual plots



Loads more we could do with linear regression:

- Heteroskedasticity
- Serial correlation
- Form of error term (skewed normal?)

The long format again

DirectEP	CededEP	NetEP	CumulativeIncurred	CumulativePaid	IncrementalIncurred	IncrementalPaid	PriorIncurred	PriorPaid	Company	Line
1,070	87	983	179	72	179	72	NA	NA	Agway Ins Co	Auto
5,140	885	4,255	2,791	505	2,612	433	179	72	Agway Ins Co	Auto
190	154	36	51	20	(2,740)	(485)	2,791	505	Agway Ins Co	Auto
2,355	905	1,450	1,178	230	1,127	210	51	20	Agway Ins Co	Auto
603	223	380	216	40	(962)	(190)	1,178	230	Agway Ins Co	Auto
722	135	587	528	147	312	107	216	40	Agway Ins Co	Auto
139	11	128	74	43	(454)	(104)	528	147	Agway Ins Co	Auto
1,560	418	1,142	555	178	481	135	74	43	Agway Ins Co	Auto
95	19	76	49	20	(506)	(158)	555	178	Agway Ins Co	Auto
808	5	803	350	123	301	103	49	20	Agway Ins Co	Auto
1,590	314	1,275	825	592	475	469	350	123	Agway Ins Co	Auto
1,358	69	1,289	747	360	(78)	(232)	825	592	Agway Ins Co	Auto
99,852	35,329	######	46,246	9,352	45,499	8,992	747	360	Agway Ins Co	Auto
7,820	2,008	5,812	3,087	952	(43,159)	(8,400)	46,246	9,352	Agway Ins Co	Auto
217	46	171	75	17	(3,012)	(935)	3,087	952	Agway Ins Co	Auto
11,846	2,033	9,813	8,838	1,680	8,763	1,663	75	17	Agway Ins Co	Auto
166	29	137	124	34	(8,714)	(1,646)	8,838	1,680	Agway Ins Co	Auto
9,326	864	8,462	6,579	1,925	6,455	1,891	124	34	Agway Ins Co	Auto
117	-	117	74	15	(6,505)	(1,910)	6,579	1,925	Agway Ins Co	Auto
1,899	388	1,511	1,420	577	1,346	562	74	15	Agway Ins Co	Auto
899	19	880	351	126	(1,069)	(451)	1,420	577	Agway Ins Co	Auto
1,179	175	1,004	487	62	136	(64)	351	126	Agway Ins Co	Auto
2,598	65	2,533	2,234	980	1,747	918	487	62	Agway Ins Co	Auto
494	36	458	71	-	(2,163)	(980)	2,234	980	Agway Ins Co	Auto
629	-	629	349	52	278	52	71	-	Agway Ins Co	Auto
30,939	715	######	20,225	4,381	19,876	4,329	349	52	Agway Ins Co	Auto

Multi-dimensional vs. multivariate

- Dimensions = specification of where something happens
- Variable = what happens there

Multivariate regression

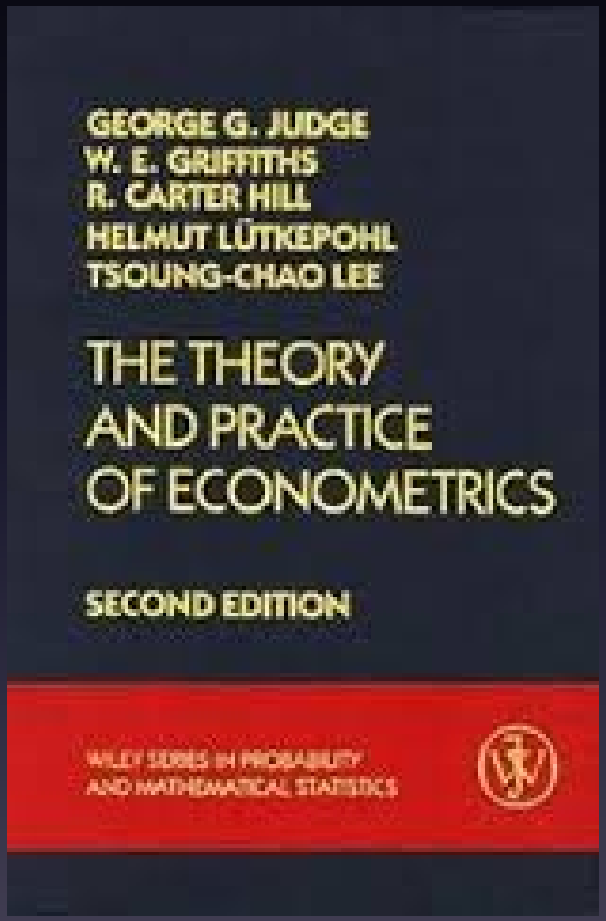
Multivariate regression is EASY.
Multi-dimensional, not so much.

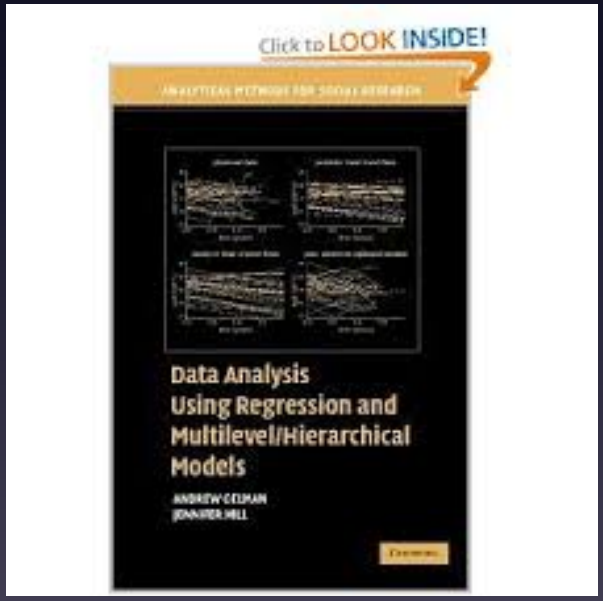
Multivariate regression in reserving

This is difficult:

- Very few degrees of freedom
- Stochastic regressors
- Claim report latency

Regression

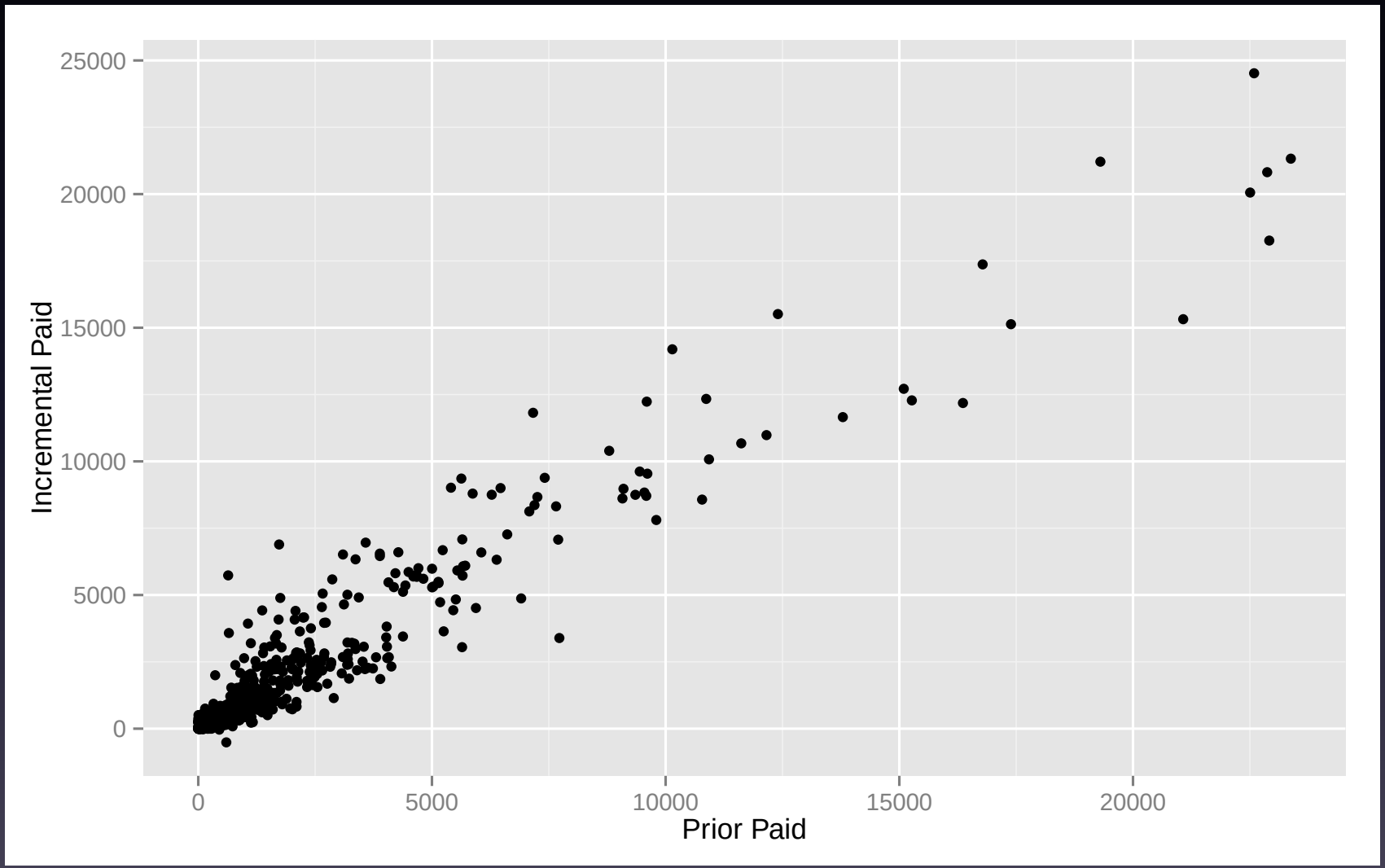

GEORGE G. JUDGE
W. E. GRIFFITHS
R. CARTER HILL
HELMUT LÜTKEPOHL
TSOUNG-CHAO LEE
THE THEORY AND PRACTICE OF ECONOMETRICS
SECOND EDITION


Click to LOOK INSIDE!
Data Analysis Using Regression and Multilevel/Hierarchical Models

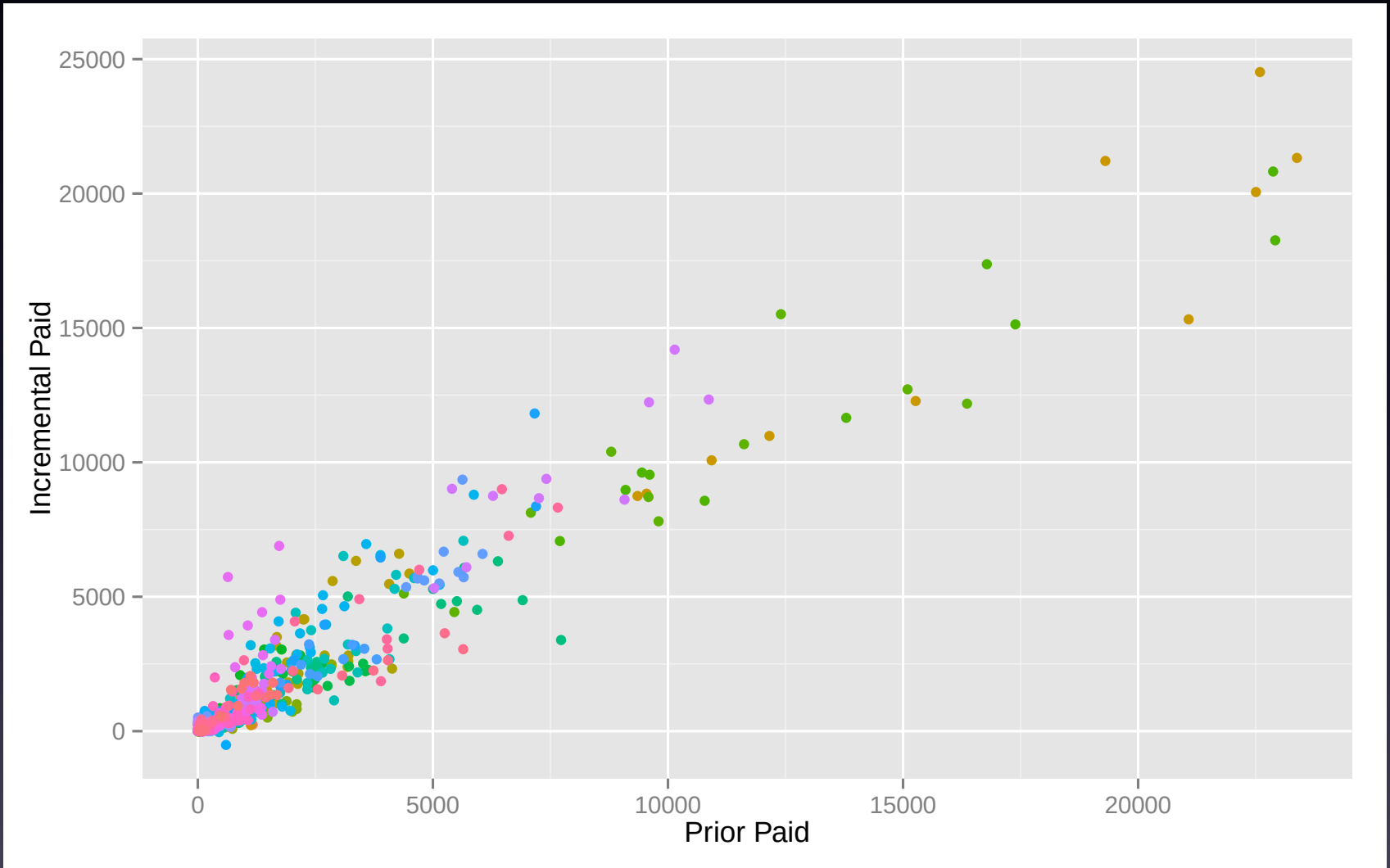
Multi-dimensional regression

$$y_i \sim N(\alpha_{j[i]} + \beta_{j[i]} x_i, \sigma_y^2)$$
$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$$
$$\beta_j \sim N(\mu_\beta, \sigma_\alpha^2)$$

12-24 Months for 92 Auto companies



Grouped


25000
20000
15000
10000
5000
0
Incremental Paid
0
5000
10000
15000
20000
Prior Paid

3 ways to fit multi-dimensional data

1. Lump all the data together - Pooled
2. Model each group separately - Individual
3. Weight both estimates together - Mixed

Pooled

$$IncrementalPaid = \alpha + \beta PriorPaid + \epsilon$$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	100.6437	30.4961	3.30	0.0010
PriorPaid	0.9597	0.0095	100.97	0.0000

Individual

This is effectively 90 different regressions. Here are results for the first 5.

	Estimate	Std..Error	t.value
Agway Ins Co	0.61	0.91	0.66
Alaska Nat Ins Co	0.76	0.24	3.20
Alliance Mut Ins Co	1.75	10.94	0.16
American Contractors Ins Grp	1.22	1.28	0.95
American Resources Ins Co Inc	0.78	0.91	0.86
Antilles Ins Co	0.85	0.97	0.87

Multi-dimensional regression

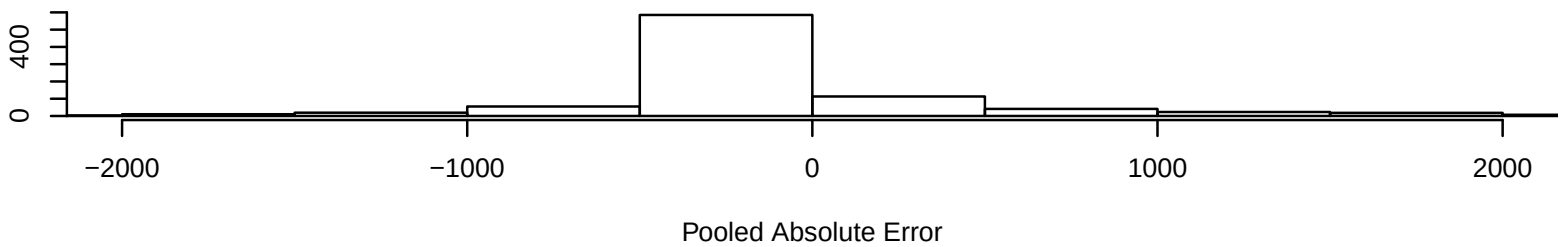
$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$$

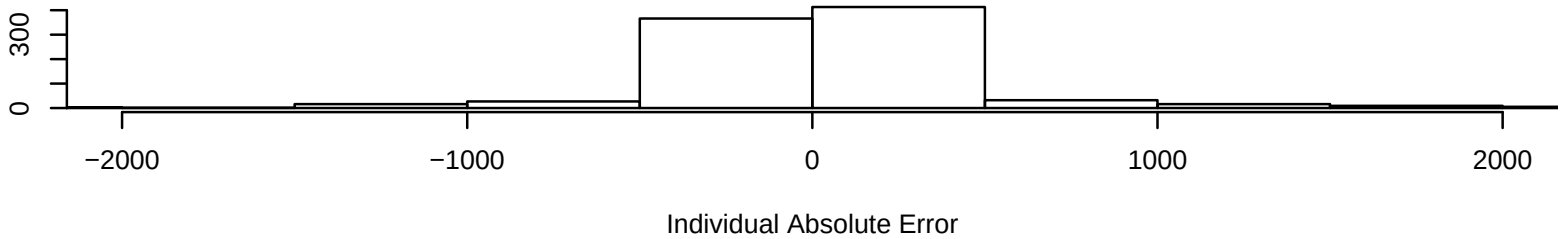
$$\hat{\alpha}_j \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

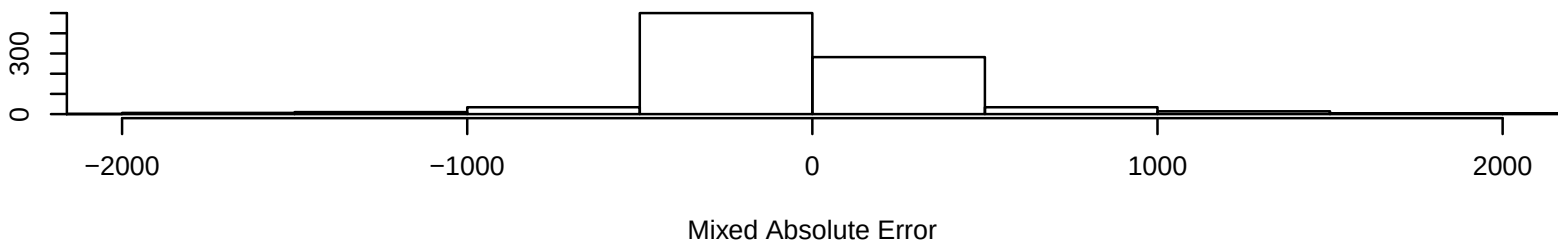
Mixed model fit

	(Intercept)	PriorPaid
Agway Ins Co	-20.13	0.88
Alaska Nat Ins Co	48.26	0.78
Alliance Mut Ins Co	33.62	0.90
American Contractors Ins Grp	75.99	0.90
American Resources Ins Co Inc	9.14	0.89
Antilles Ins Co	12.02	0.89

Model comparison

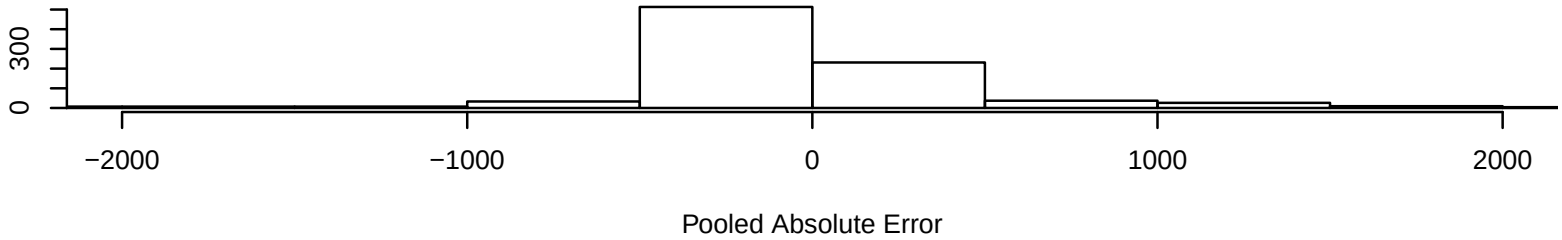


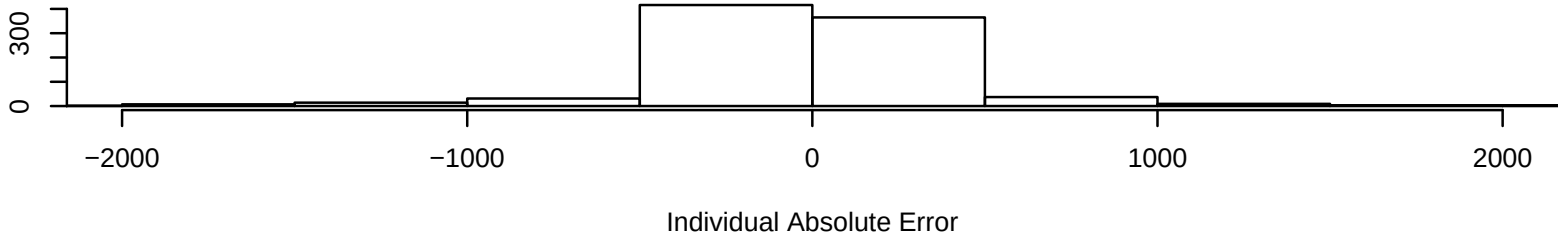


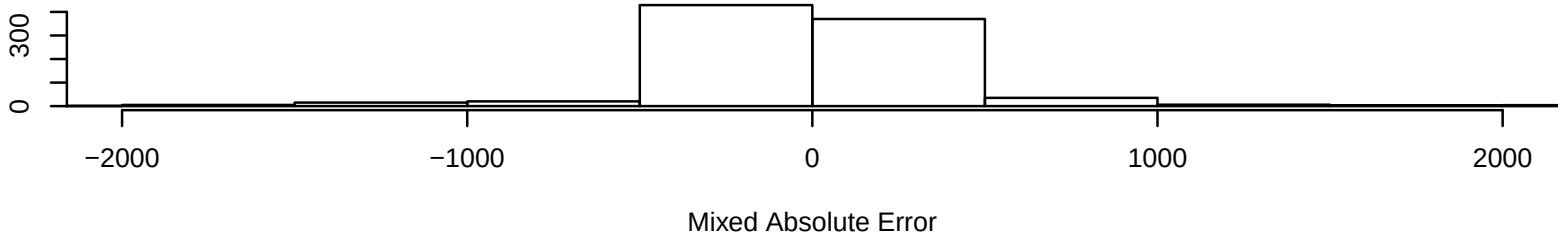


How about another predictor?

Additive method results







Model comparison

	Multiplicative	Additive
Pooled	439.90	362.75
Individual	262.20	251.07
Mixed	256.85	236.89

Thoughts

- How much data are you throwing away?
- How often do statistical diagnostics drive model decisions?
- How often does technology dictate what's in your statistical toolkit?

Thanks!