A Multilevel Analysis of Intercompany Claim Counts

joint work with K. Antonio (Amsterdam) and E.W. Frees (Wisconsin)

Emiliano A. Valdez, University of Connecticut

Wharton Business School, University of Pennsylvania - *Philadelphia*
September 11, 2008

Multilevel models

- Models that are extensions to regression whereby:
  - the data are generally structured in groups, and
  - the regression coefficients may vary according to the group.
- Multilevel refers to the nested structure of the data.
- Classical examples are usually derived from educational or behavioral studies:
  - e.g. students $\in$ classes $\in$ schools $\in$ communities
- The basic unit of observation is the 'level 1' unit; then next level up is 'level 2' unit, and so on.
- Some references for multilevel models: Gelman & Hill (2007), Goldstein (2003), Raudenbusch & Byrk (2002), Kreft & De Leeuw (1995).

Analysis of intercompany frequency data

- Our paper examines an intercompany database using multilevel models. We focus analysis on claim counts.
- The empirical data consists of:
  - financial records of automobile insurers over 9 years (1993-2001), and
  - policy exposure and claims experience of randomly selected 10 insurers.
- Source of data: General Insurance Association (GIA) of Singapore
- The multilevel model accommodates clustering at four levels: vehicles (v) observed over time (t) that are nested within fleets (f), with policies issued by insurance companies (c).

Benefits of intercompany data analysis

- The Society of Actuaries often collects intercompany data through experience studies.
  - However, often the analysis is limited to analysis of descriptive statistics and in tabular forms.
  - Formal statistical models are often lacking.
- We believe that a multilevel statistical model of intercompany data would be of interest to:
  - the direct insurers - comparing their own experience with their competitors;
  - reinsurers - ability to predict losses at various company level; and
  - regulators - examine loss experience of several companies e.g. to detect fraud by inspecting unusual levels of losses.

The motivation to use multilevel models

- Multilevel models allows us to account for variation in claims at the individual level as well as for clustering at the company level.
  - intercompany data models are of interest to insurers, reinsurers, and regulators.
- It also allows us to examine the variation in claims across 'fleet' policies:
  - policies whose insurance covers more than a single vehicle e.g. taxicab company.
  - possible dependence of claims of automobiles within a fleet.
- In general, it allows us to assess the importance of cross-level effects.

Our contribution

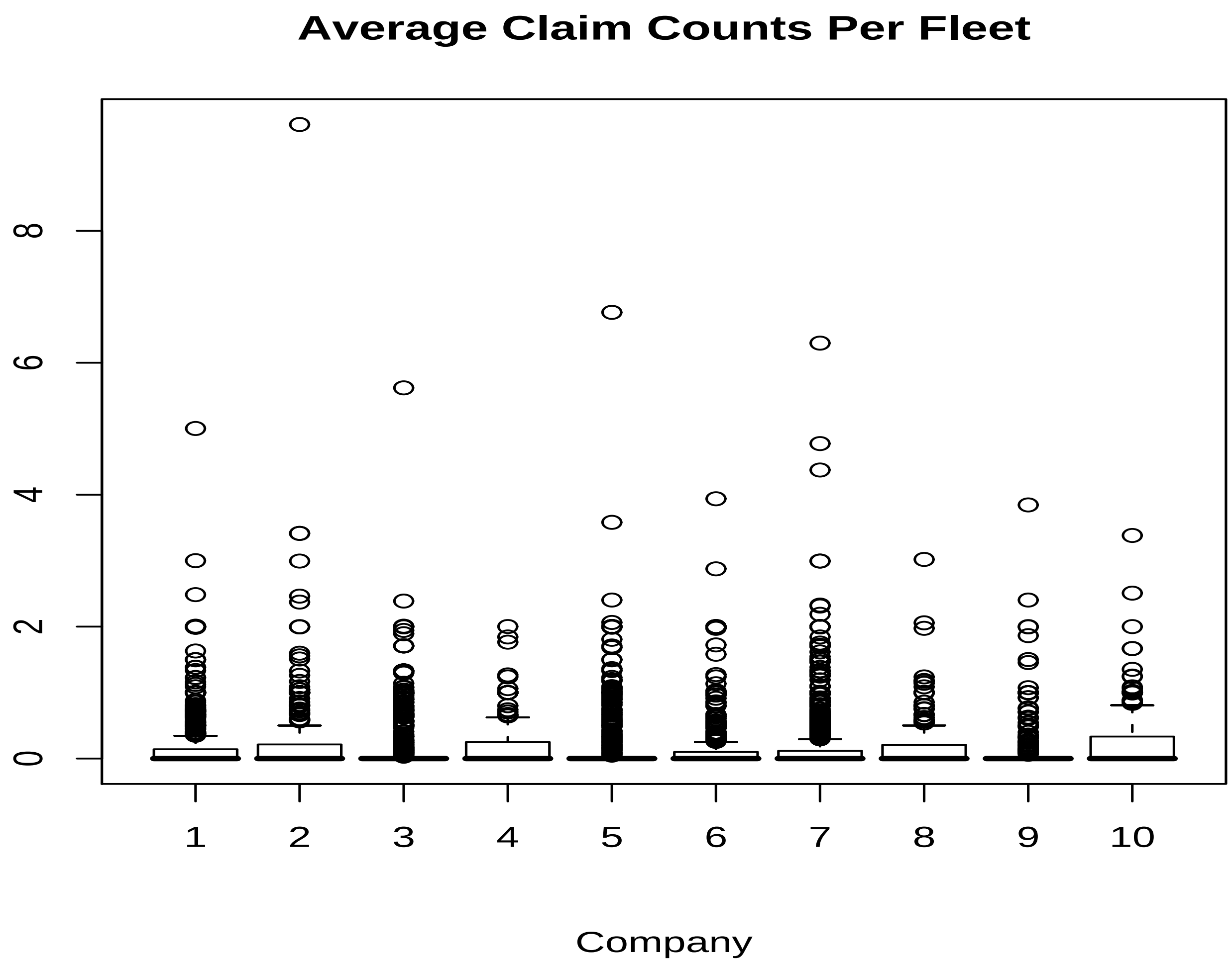
- We develop the connection between hierarchical credibility and multilevel statistics, a discipline generally unknown in actuarial science.
  - We go beyond the 2-level structure often found in panel data.
- We extended applications (to more than two levels) of generalized count distribution models in actuarial science:
  - Poisson, Negative Binomial, Zero-inflated Poisson, Hurdle Poisson
- We provide modeling and detailed analysis of intercompany data on fleets which has been scarce in the actuarial literature.

Data characteristics

Table 1: *Claims by company*

		Percentage of Claims by Company									
Count	All	1	2	3	4	5	6	7	8	9	10
0	87.82	88.27	81.68	94.68	87.71	89.43	88.83	87.44	86.86	88.78	87.28
1	10.49	10.23	15.11	4.96	10.55	9.3	9.74	11.09	11.13	9.57	10.85
2	1.41	1.3	2.73	0.3	1.43	0.96	1.1	1.26	1.62	1.37	1.71
3	0.22	0.18	0.36	0.06	0.29	0.19	0.2	0.19	0.34	0.24	0.17
4	0.04	0.03	0.12	0	0	0.06	0.1	0.02	0.05	0.04	0
5	0.01	0	0	0	0.02	0.06	0.04	0	0	0	0
# Claims	5,557	528	1,096	191	603	398	669	891	318	328	535
# Obs.	39,120	3,920	4,951	3,327	4,191	3,225	5,105	6,251	2,040	2,487	3,623
# Exp.	30,560	3,106	4,440	2,480	3,240	2,497	3,978	5,023	1,635	1,505	2,656
Mean	0.14	0.17	0.25	0.08	0.19	0.16	0.17	0.18	0.19	0.22	0.20
# Fleet	6,763	841	270	1,229	270	1,279	646	1,286	335	268	339

Figure 1



Vehicle level covariates

Table 3: *Vehicle level explanatory variables*

Categorical Covariate	Description	Percentage		
Vehicle Type	Car	54%		
	Motor	41%		
	Truck	5%		
Private Use	Vehicle is used for private purposes	31%		
	Vehicle is used for other than private purposes	69%		
NCD	'No Claims Discount' at entry in fleet: based on previous accident record of policyholder. The higher the discount, the better the prior accident record.			
	NCD $= 0$	83%		
	NCD > 0	17%		
SwitchPol	1 if vehicle changes fleet	55%		
	0 if vehicle enters fleet for first time or stays in the same fleet	45%		
Continuous Covariate		**Minimum**	**Mean**	**Maximum**
Vehicle Age	The age of the vehicle in years, at entry in fleet	0	4.22	33
Cubic Capacity	Vehicle capacity for cars and motors	124	1,615	6,750
Tonnage	Vehicle capacity for trucks	1	7.6	61
TLengthEntry	Time (in years) vehicle was in the sample, before entering the fleet	0	0.35	6.75
TLength	(Exposure) Fraction of calendar year for which insurance coverage is purchased	0.006	0.78	1

Fleet and company level covariates

Table 4: *Fleet and company level explanatory variables*

Covariate	Description	Minimum	Mean	Maximum
Fleet Level				
AvNCD	Average of No Claims Discount at entry in the fleet	0	6.3	50
AvTLengthEntry	Average of TLengthEntry	0	0.59	6.75
AvTLength	Average of cumulative time period spent in fleet	0	1	3.64
AvVAge	Average of vehicle age at entry in the fleet	0	4.75	27.33
AvPrem	Average of premium paid per unit of exposure	0.01	1.3	59.56
FleetCap	Number of vehicles in the fleet	1	4.56	1,092
Company Level				
NumFleets	Number of fleets in the company	268	942	1,286
NumVeh	Number of vehicles in the company	1,319	3,084	5,394
NumCars,	Number of cars, trucks and motorcycles	391	1,652	4,453
NumTrucks,	in the company	224	1,259	3,019
NumMotors		0	170	888

Count distribution models

- **(Poisson)** $\mathsf{Pr}_{\mathsf{Poi}}(Y = y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}$;
- **(Negative binomial)**
  $\mathsf{Pr}_{\mathsf{NB}}(Y = y|\mu,\tau) = \frac{\Gamma(y+\tau)}{y!\Gamma(\tau)} \left(\frac{\tau}{\mu+\tau}\right)^{\tau} \left(\frac{\mu}{\mu+\tau}\right)^{y}$;
- **(Zero–inflated Poisson)**

$$\mathsf{Pr}_{\mathsf{ZIP}}(Y = y|p,\lambda) = \begin{cases} p + (1-p)\mathsf{Pr}_{\mathsf{Poi}}(Y = 0|\lambda) & y = 0, \\ (1-p)\mathsf{Pr}_{\mathsf{Poi}}(Y = y|\lambda) & y > 0; \end{cases}$$

- **(Hurdle Poisson)**

$$\begin{aligned} \mathsf{Pr}_{\mathsf{Hur}}(Y = 0|p,\lambda) &= p & y = 0, \\ \mathsf{Pr}_{\mathsf{Hur}}(Y = y|p,\lambda) &= \frac{1-p}{1-\mathsf{Pr}_{\mathsf{Poi}}(0|\lambda)}\mathsf{Pr}_{\mathsf{Poi}}(Y = y|\lambda) & y > 0. \end{aligned}$$

Fitted models without covariates

Table 5: *Observed and expected claim counts*

Num. Claims	Obs. Freq.	Poisson	NB	ZIP	Hurdle Poi
0	34,357	33,940	34,362	34,357	34,357
1	4,104	4,821	4,079	4,048	4,048
2	551	342	577	641	641
3	86	16	86	68	68
4	17	1	13	5	5
≥ 5	0	2	0	0	
Mean	0.142	0.142	0.142	0.142	0.142
Variance	0.171	0.142	0.17	0.17	0.17
-2 Log Lik	/	34,032	33,536	45,815	33,582
AIC	/	34,034	33,540	45,819	33,586

Models considered

- Hierarchical Poisson models which include
  - Jewell's hierarchical model
- Hierarchical Negative Binomial model
- Hierarchical Zero-Inflated Poisson model
- Hierarchical Hurdle Poisson model

Model spec. of the hierarchical ZIP model

- While the specifications of all models considered are in the paper, here we focus on the hierarchical zero-inflated model.

$$
\begin{aligned}
Y_{c,f,v,t} &\sim \mathsf{ZIP}(p, \lambda_{c,f,v,t}) \\
\text{where } \lambda_{c,f,v,t} &= e_{c,f,v,t} \exp\left(\eta_{c,f,v,t} + \epsilon_c + \epsilon_{c,f}\right) \\
\text{and } \eta_{c,f,v,t} &:= \gamma + \boldsymbol{X}_c \boldsymbol{\beta}_4 + \boldsymbol{X}_{c,f} \boldsymbol{\beta}_3 + \boldsymbol{X}_{c,f,v} \boldsymbol{\beta}_2 + \boldsymbol{X}_{c,f,v,t} \boldsymbol{\beta}_1 .
\end{aligned}
$$

- Here γ is the intercept, ϵ_c is a random company effect, $\epsilon_{c,f}$ is a random effect for the fleet within the company and $\epsilon_{c,f,v}$ is a random effect for the vehicle within the fleet.

- The $\boldsymbol{X}$'s are the explanatory variables defined accordingly on page 15 of paper.

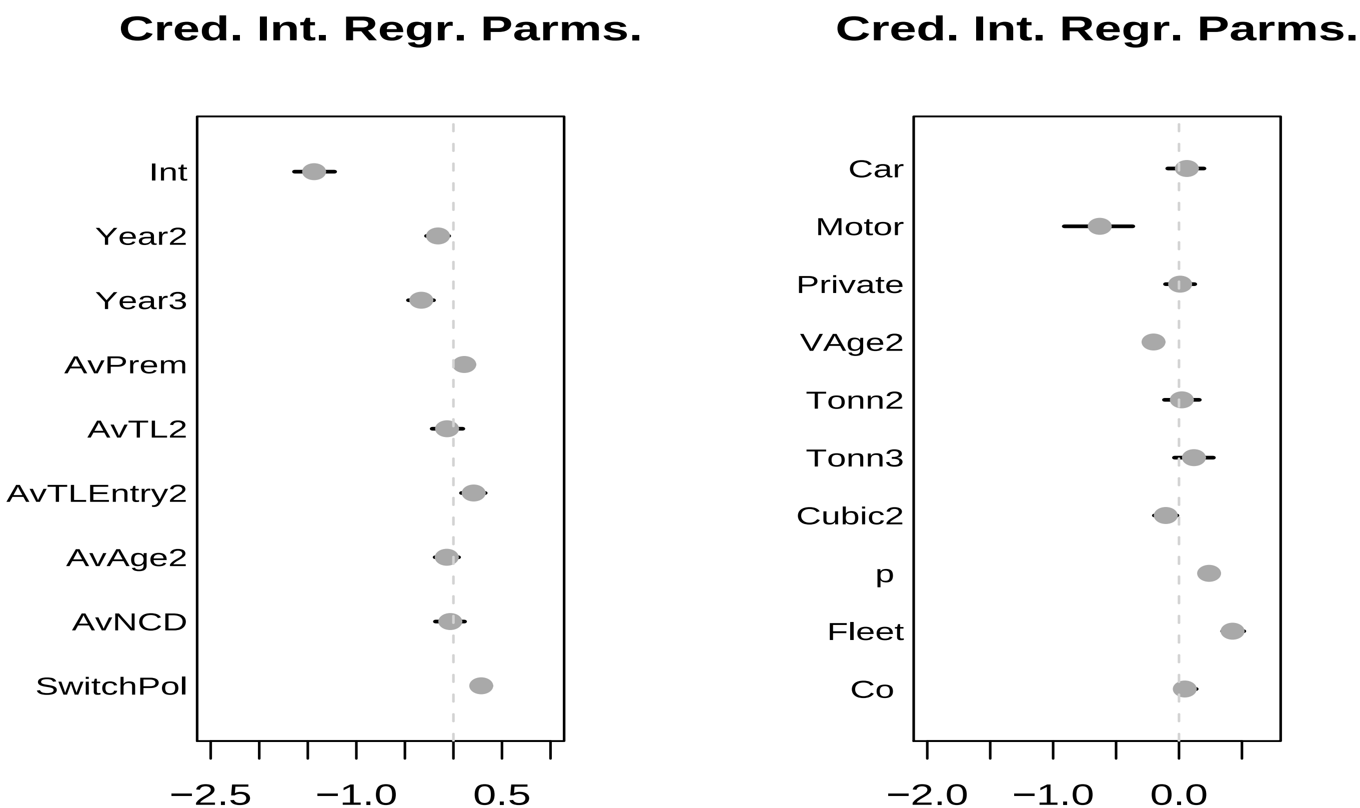
Comparing the fitted models

Table 9: *Estimated claim counts obtained from Bayesian hierarchical analyses*

Num. Claims	Obs. Freq.	Poisson (3)	NB (7)	ZIP (8)	Hurdle Poi (10)
0	34,357	34,310	34,365	34,350	34,360
		(34,200;34,430)	(34,240;34,490)	(34,230;34,470)	(34,230;34,480)
1	4,104	4,176	4,086	4,092	4,139
		(4,081;4,273)	(3,978;4,196)	(3,979;4,207)	(4,025;4,253)
2	551	536	560	584	540
		(511;560)	(532,588)	(551;618)	(505;576)
3	86	79	88	79	73
		(71,87)	(78,99)	(70;89)	(64;82)
4	17	14	16	11	10
		(11,16)	(13,20)	(9;13)	(8;13)
≥ 5	5	3	4	2	2
		(2,4)	(2,4.5)	(1;2)	(1;2.4)

Figure 3

95% credibility intervals for the hierarchical ZIP model



A Priori Premiums and A Posteriori Corrections

Table 11: *Results for the ZIP model*

Co.	Fleet	Vehicle	*A Priori* (Exp.)	*A Posteriori*	BMF	Acc. Cl. Fleet (Exp.)	Claim free Years
4	1,590	6,213	0.2156 (1)	0.3653	1.69	7 (15.25)	10.4
		6,261	0.2156 (1)	0.3653			
1	4,370	10,104	0.1404 (1)	0.218	1.56	7 (21.5)	16.5
		5,841	0.1404 (1)	0.218			
		7,152	0.1715 (1)	0.2663			
5	4,673	9,350	0.07942 (0.5)	0.106	1.33	6 (18.5)	17
		12,131	0.07942 (0.5)	0.106			
		12,210	0.07942 (0.5)	0.106			
4	6,592	1,656	0.1066 (1)	0.1898	1.78	12 (40)	32.3
		15,329	0.1099 (1)	0.1956			
		2,577	0.1302 (1)	0.2319			
2	1,485	11,122	0.01672 (0.08)	0.03961	2.4	17 (40)	31.7
		10,782	0.01223 (0.08)	0.02867			
		11,063	0.01494 (0.08)	0.03539			
3	4,672	12,007	0.06814 (0.334)	0.0705	1.03	5 (20.4)	16.1
		8,367	0.06814 (0.334)	0.0705			
		11,958	0.06814 (0.334)	0.0705			
5	1,842	1,826	0.1486 (1)	0.1244	0.84	2 (16)	14
		1,569	0.1486 (1)	0.1244			
6	5,992	1,906	0.1816 (1)	0.2333	1.28	7 (21)	16
		1,889	0.1816 (1)	0.2333			
9	5,823	1,020	0.1091 (1)	0.09044	0.83	2 (16)	14.25
		1,056	0.1091 (1)	0.09044			
		1,025	0.1091 (1)	0.09044			
10	3,564	15,564	0.1919 (1)	0.1475	0.77	2 (17)	15
		14,831	0.157 (1)	0.1207			
		15,194	0.157 (1)	0.1207			
10	3,568	1,119	0.1508 (1)	0.135	0.90	3 (19.25)	16.25
		1,206	0.1508 (1)	0.135			
		1,540	0.1508 (1)	0.135			

Note: 'Acc. Cl. Fleet' and 'Acc. Cl. Veh.' are accumulated number of claims at fleet and vehicle levels, respectively. 'Exp.' is exposure at year level, in parenthesis.

Concluding remarks

- This paper presents a multilevel analysis of a four–level intercompany data set on claim counts for fleet policies.
- We build multilevel models using generalized count distributions (Poisson, negative binomial, hurdle Poisson and zero–inflated Poisson) and use Bayesian estimation techniques.
- We find that in all models considered, there is the importance of accounting for the effects of the various levels.
- To demonstrate the usefulness of the models, we illustrate how *a priori* rating (using only *a priori* available information) and *a posteriori* corrections (taking the claims history into account) for intercompany data can be calculated on a sound statistical basis.

Some useful references

Gelman, A. and Hill, J., 2007, *Applied Regression and Multilevel (Hierarchical) Models*, Cambridge University Press.

Goldstein, H., 2003, *Multivariate Statistical Models*, Oxford University Press.

Kreft, I.G. and deLeeuw, J., 1998, *Introducing Multilevel Modeling*, London: Sage Publications.

Raudenbush, S.W. and Byrk, A.S., 2002, *Hierarchical Linear Models: Applications and Data Analysis Methods*, Thousand Oaks: Sage Publications.