A Multilevel Analysis of Intercompany Claim Counts

Katrien Antonio*† Edward W. Frees ‡ Emiliano A. Valdez §

May 15, 2008

Abstract

For the construction of a fair tariff structure in automobile insurance, insurers classify the risks that they underwrite. The idea behind this risk classification is to subdivide the portfolio into classes of risks with similar profiles. While some insurers may have sufficient historical data, several others may not have significant volume of experience data in order to produce reliable claims predictions to help enhance their risk classification systems. A database containing a pooled experience of several insurers thereby helps to produce a more fair, reliable, and equitable premium structure for all risks concerned. Research and analysis of such "intercompany" insurance experience data is lacking in both the actuarial and statistical literature. Its benefits goes beyond the insurer; reinsurers (i.e. insurers of insurers) together with regulators also benefit from statistical models of this type of data because they typically deal with analyzing the experience of a collection of insurers.

In this paper, we use multilevel models to analyze the data on claim counts provided by the General Insurance Association of Singapore, an organization consisting of most of the general insurers in Singapore. Our data comes from the financial records of automobile insurance policies followed over a period of nine years. The multilevel nature of the data is due to the following: a certain vehicle is observed

*Corresponding author: k.antonio@uva.nl (phone: +31 20 525 5020)

†University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.

‡University of Wisconsin in Madison, USA (jfrees@bus.wisc.edu).

§University of Connecticut, USA (valdez@math.uconn.edu).

over a period of years and is insured by a particular insurance company under a certain 'fleet' policy. Fleet policies are umbrella–type policies issued to customers whose insurance covers more than a single vehicle with a taxicab company being a typical example. We show how intercompany data lead to *a priori* premiums and *a posteriori* corrections to these initial premiums. Specific focus is made in understanding the intercompany effects using various count distribution models (Poisson, negative binomial, zero–inflated and hurdle Poisson). The performance of these various models is compared; we also investigated how to use the historical claims of a company, fleet and/or vehicle in order to correct for the premium initially set.

Keywords: actuarial science; hierarchical model; multilevel model; experience rating; bonus–malus factors; generalized count distributions.

1 Introduction

In many countries and for many lines of business, the insurance market is mature, or at least expanding less rapidly, and highly competitive. This strong competition induces insurers to classify risks that they underwrite in order to mitigate problems of adverse selection. To illustrate the nature of adverse selection in automobile insurance, a policyholder's prior driving history is commonly used as a risk rating factor. A person with a poor driving history may seek for a company that does not use this rating factor for pricing; use of the rating factor penalizes him or her for past mistakes in the form of higher premiums. Conversely, a person with a good driving history may seek companies that use this rating factor; these companies reward previous good experience with lower premiums. Companies that use a less refined classification system than their competitors tend to attract less desirable risks, which can have a spiraling effect on future claims. Risk classification systems allow insurers to price their products in a fair and equitable manner, and on a sound statistical basis.

Strong competition encourages insurers to utilize detailed classification systems, so refined that they may not have sufficient exposure to produce reliable claims predictions for all risks in the portfolio. To understand their claims distributions, it is common for several insurers to pool their experience, forming a database known as 'intercompany' data. With a database large enough to produce a refined classification system, fair and equitable premiums can be determined more reliably across all risks.

Although insurance companies compete for the same business, economic forces dictate that the loss experience of insurers can differ. During the sales process, insurers use different underwriting standards and pricing structures to attract different mixes of business. During claim settlements, insurers differ in their procedures (including legal) and the calculations of claims adjustments, thereby realizing different loss experience across companies. Moreover, there are issues of moral hazard, where an insured that has a policy with a company may have a different claims experience than if the insured were contracted with another company.

1.1 Multilevel Modeling

This paper examines an intercompany database using multilevel modeling. Specifically, we consider policy exposure and claims experience data derived from automobile insurance portfolios of a randomly selected sample of ten general (property and casualty) insurance companies in Singapore. Our data comes from the financial records of automobile insurance policies over a period of nine years, 1993-2001.

Multilevel modeling allows us to readily handle individual claims experience and account for clustering at the company level. It also allows us to examine commercial insurance policies by restricting considerations to 'fleet' policies. These are policies issued to customers whose insurance covers more than a single vehicle. A typical situation of 'fleet' policies is automobile insurance coverage provided to a taxicab company, where several taxicabs are insured under the same policy. A peculiar characteristic of these policies is the nature of the possible dependence of claims of automobiles within a fleet. Multilevel models, to an extent, capture this peculiarity. The unit of observation in our analysis is therefore a registered vehicle insured under a fleet policy. Our 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).

Ideas of multilevel modeling and inference are now well–developed in the statistics literature (Kreft and deLeeuw (1998), Snijders and Bosker (1999), Raudenbush and Bryk (2002), Goldstein (2003) and Gelman and Hill (2007)). Linear multilevel modeling also has a long history in the actuarial literature, as summarized in Norberg (1986). Norberg credits the idea to Jewell (1975), with early contributions by Taylor (1979) and Sundt (1980). As an example of classic multilevel insurance applications, Sundt briefly mentions (i) insurance claims from a person, with (ii) several people living in a household, (iii) where several houses are in a town, (iv) and many towns in a county and (v) several counties with a country. Norberg (1986) and Frees et al. (1999) discuss the connections between the statistical linear modeling and traditional actuarial literatures.

1.2 Count Data

This paper examines nonlinear models using insurance claim counts. The frequency component has been well analyzed in the actuarial literature, at least when cross-sectional and panel data structures are considered. For instance, the modern approach of fitting a claims count distribution to longitudinal data can be attributed to the work of Dionne and Vanasse (1989) who applied a random effects Poisson count model to automobile insurance claims. Pinquet (1997) and Pinquet (1998) extended this work, considering severity as well as frequency distributions. Pinquet et al. (2001) and Bolancé et al. (2003) introduced a dynamic element into the observed latent variable, again using Poisson regression.

Poisson regression is probably the most popular technique for regression with count data. However, recent research in actuarial science (see e.g. Yip and Yau (2005) and Boucher et al. (2007)) has highlighted the use of parametric distributions other than Poisson to accommodate features of actuarial panel data that are inconsistent with the Poisson distribution. These authors investigated the use of the negative binomial, zero-inflated and hurdle distributions for the analysis of cross-sectional and longitudinal claim counts. Cameron and Trivedi (1998), Winkelmann (2003), Yau et al. (2003) and Lee et al. (2006) discuss similar research in econometrics, respectively biostatistics. Further on, we extend these distributions towards the analysis of multilevel data with more than two levels.

A data analytic discussion of ratemaking for fleet data has received limited attention in the actuarial literature, Desjardins et al. (2001) and Angers et al. (2006) being the exceptions. They discuss the calculation of bonus–malus factors ('BMF') for a three level data set of claim counts on insured trucks in Québec. The econometric (or statistical) models used in their papers are Poisson regression models with random effects for vehicles and fleets.

1.3 Benefits of Intercompany Data

A multilevel model of intercompany data is of interest to insurance companies, regulators and reinsurers. Insurance companies can use the results of this paper to predict the number of claims not only for each vehicle but also for each fleet. Predictions at the fleet level are particularly important because contracts are written and hence premiums are exchanged for coverage at this level. Further, an insurance company can use a model of several companies to understand and possibly compare their experience with their competitors. To illustrate, given a specific risk class (such as female, aged 20-24 with poor driving experience), is the loss experience for the company high or low compared to the competition? This type of information is extremely useful in a competitive pricing environment.

Regulators and reinsurers typically deal with several companies and so would also benefit from a single model representing the experience of many companies. Regulators are concerned with establishing fair pricing of insurance policies and ensuring that insurers have sufficient assets to meet contractual obligations. A single model can help regulators examine loss experiences of several companies, using covariate information to comparably account for the risks underwritten by these companies. Moreover, regulators can use these comparisons for detecting fraud and further inspecting unusually high or low losses (that may be suspect as indicated by the risk rating factors as covariates).

Reinsurers are the 'insurers of insurance companies' and they take on layers of risks so that insurers are able to diversify their loss exposure. Naturally, reinsurers are interested in the loss distributions that they are accepting. Predictions at the company level are important to prices charged by reinsurers.

In the USA the Society of Actuaries ('SOA') collects intercompany data through experience studies. As noted in Iverson et al. (2007), "one of the key elements that led to the creation of the Actuarial Society of America in 1889 was the need for an independent body to collect and report upon experience." The SOA publishes descriptive statistics based on the data collected from participating insurers and these "intercompany reports of experience are considered a proxy for the state of the industry with companies using these results to benchmark their own experience" (Iverson et al. (2007)). Despite the various parties (insurers, reinsurers, regulators and actuarial organizations like the SOA) interested in the analysis of intercompany data on claim statistics, sound statistical research in this area is still lacking.

The primary contributions of the research in this paper are threefold. Firstly, we develop the connection between hierarchical credibility and multilevel statistics, a discipline that is generally unknown in actuarial science. We go beyond the two level structures often found in panel data. Credibility is a classical actuarial approach for experience rating (and Hickman and Heacox (1999) claimed it to be one of the cornerstones of actuarial mathematics). Secondly, with the growing popularity of the generalized count distributions in actuarial science, we extended their applications towards more than two level data sets. The statistical estimation poses some challenges, but Bayesian estimation techniques allowed us to meet these. Thirdly, we provide modeling and a detailed analysis of intercompany data on fleets, which, as alluded earlier, has been rather scarce in the actuarial literature.

The paper has been structured as follows. Section 2 gives background on the data used in the analysis. Model specification, data analysis and prediction for claim counts is covered in Section 3 and 4. Section 5 concludes.

2 Intercompany Insurance Claims Data

2.1 Background

We investigate a data set with policy exposure information, covariates and claim counts registered for vehicle insurance portfolios of general insurance companies in Singapore. With 'exposure' is meant the fraction of a year during which the policy holder pays for insurance. The source of this intercompany data set is the General Insurance Association ('GIA') of Singapore (see the organization's webpage http://www.gia.org.sg), an organization consisting of general insurers in Singapore. In Singapore motor insurance is compulsory and it is not surprising to find it to be one of the most important general insurance lines of business.

Two files were examined: the policy and the claims files. The policy file consists of records of policyholders with vehicle insurance coverage purchased from a general insurer during the period 1993-2001. Each vehicle is identified with a unique code. In general, the file provides characteristics of the policyholder and the vehicle insured. However, for fleet policies, no information on the driver of the vehicle is available, since a vehicle may be used by several drivers. Thus, the unit of observation in our analysis is a registered vehicle insured, broken down according to their exposure in each calendar year 1993 to 2001. The claims file provides a record of each accident claim that has been filed with the insurer during the observation period and is linked to the policyholder file.

All policies in the sample have a comprehensive coverage that includes coverage for third party injury and property damage as well as damage to one's own vehicle. Each vehicle is followed over a period of (maximum) nine years: January 1993 until December 2001. However, not every vehicle is in the sample during the full period (see infra for concrete statistics). Vehicles may switch fleets and companies and enter and exit the panel. Vehicles with doubtful information on some variables such as the vehicle capacity or the year the car was manufactured were removed from the sample and therefore ignored in the analysis.

The hierarchical structure of the data lends itself naturally to multilevel modeling with four different levels for this data set. At the highest level, we analyze ten insurance companies (using c to denote a company). For confidentiality reasons, these ten companies, labeled 1 to 10, were randomly drawn from 27 companies available in the GIA's entire database. At the next level, from these 10 companies, we consider 6,763 fleets (f). Level two consists of 16,437 vehicles (v) that we observe over time (t), for a total of 39,120 level one observations.

2.2 Data Characteristics

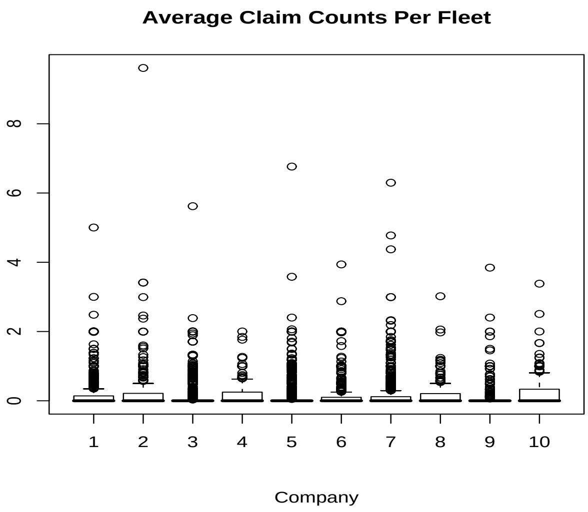
The empirical distribution of the observed claim counts in Table 1 (2^{nd} column) illustrates that about 88% percent of the observations are zeros. At most 5 claims during one time period have been observed. Table 1 describes the distribution of claim counts per company. '# Obs.' gives the number of observations per company and '# Exp.' the total exposure time per company (in policy years). These statistics show that the companies are roughly the same size, with no one company dominating the market. The 'Mean' is the sum of claim counts divided by the total exposure. This suggests substantial differences among companies; the average claim count for company 3 is quite small compared to companies 2, 9 and 10.

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 shows the distribution of claim counts at the fleet level. Specifically, for each fleet, the average claim count (per unit of exposure) was computed and the distribution of these averages appears in Figure 1, by company. One can observe company effects in Table 1 in the sense that the average number of claims reported by company 3 is very low, whereas the averages from companies 2, 9 and 10 are rather high. Company 9 is special in the sense that it has the lowest exposure and yet one of the largest claims per unit of exposure. However, at fleet level (see Figure 1), 81% of the fleets in this company reported zero claims in total (compare this with e.g. company 10 where only 51% of the fleets stayed claim–free during the observation period).

Figure 1: *Boxplots of a fleet's average number of claims, by company*



Fleets buy and sell vehicles periodically and switch insurance companies. To explore this behavior, we can explore the length of time that a vehicle, our unit of analysis, stays in the sample with the same fleet and company. For example, 55% of all vehicles in the sample never switched fleet nor company during their period of insurance that is registered in the data set, 21% switched once and 24% switched two or more times. Statistics for the total length of stay of a vehicle within the same fleet ('Cum(ulative) L(ength) Fleet') and the total observation period in the sample ('Cum(ulative) L(ength) Sample') (in years) are in Table 2. 'Prem(ium)' gives the premium paid per unit of exposure. Striking in the latter table is that vehicles – on average – only stay for a short period within the same fleet. Averages of the total length of their exposure period in the sample are just slightly higher.

Table 2: *Length of stay (in years) of a vehicle within a fleet and within the sample. Premiums paid are per unit of exposure.*

		Company									
		1	2	3	4	5	6	7	8	9	10
Cum.L.	Min	0.008	0.005	0.011	0.005	0.011	0.005	0.008	0.005	0.008	0.011
Fleet	Mean	0.994	0.949	1.065	1.135	0.963	0.969	0.934	0.94	1.142	0.979
	Max	2.504	2	2.219	6	2.225	2.285	2.244	2.459	2.999	2
Cum.L.	Min	0.008	0.027	0.011	0.005	0.011	0.005	0.011	0.033	0.008	0.016
Sample	Mean	1.821	2.096	1.77	1.675	1.355	2.106	2.309	1.831	1.414	1.778
	Max	7.753	5.922	7.467	7.331	5.505	8	7.174	7.567	5.008	7.99
Prem.	Min	0.01	0.0312	0.01	0.025	0.032	0.014	0.004	0.014	$1.8*10^{-4}$	0.045
	Mean	1.522	0.952	1.232	1.124	1.553	1.2	1.107	1.366	0.854	1.422
	Max	11.648	59.562	56.129	7.124	156.411	96.963	42.8	188.935	5.635	12.379

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

Other measurable characteristics at the level of the vehicle are summarized in Table 3. No specific information at the level of the fleet or company is available (such as the branch where the fleet is operating, details on the financial structure of the company, et cetera). For instance, Angers et al. (2006) uses the sector of activity of the carrier as an explanatory variable in their regression analysis. To compensate this lack, averages at the level of the fleet and company are created. These are listed in Table 4. The averages in the upper part of the table are computed at fleet level, e.g. 'AvPrem' is the total premium paid by all vehicles in the fleet, divided by the total period (in years) for which insurance is guaranteed by the fleet.

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

3 Models and Results

We use multilevel modeling for this 4–level data set (vehicles followed over time, grouped in a fleet policy issued by a company). Multilevel models allow us to incorporate explicit knowledge about the hierarchical structure of the data by specifying random effects for the vehicle and/or fleet and/or company. The random effects are included to deal with the (apparent) contagion resulting from unobservable effects at the various levels in the data. For instance, the missions assigned to a vehicle or the behavior of its drivers (in case this group is small) may influence the riskiness of a vehicle. At the level of the fleet, guidelines on driving hours, mechanical check-ups and loading instructions may influence the number of accidents reported by vehicles in the fleet.

Due to the huge percentage of zeros reported in Table 1, we will not only investigate Poisson regression, but also negative binomial, zero–inflated Poisson and hurdle Poisson models. These distributions are specified below:

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

- (**Zero–inflated Poisson**)

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

- (**Hurdle Poisson**)

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

In the sequel of this paper we will use 'Poi' to denote a Poisson distribution, 'NB' for a negative binomial, 'ZIP' for a zero–inflated Poisson and 'Hur' for a hurdle Poisson distribution with parametrization as given above. In the absence of covariate information, these count distribution models are each fitted to the 'rough' data set. A comparison of their performance is illustrated in Table 5. Table 5 suggests that the negative binomial is the best candidate model followed closely by the hurdle Poisson model.

Table 5: *Observed and expected claim counts for the various count distributions (likelihood–based estimation)*

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

Categorized versions of the covariates are used in our multilevel specifications (see Table 6 and 7). Using categorizations is in line with tarification practice in insurance companies and with the literature on non-life insurance (see e.g. Desjardins et al. (2001), Angers et al. (2006) and Denuit et al. (2007)).

An overview of the likelihood specifications involved in panel data models for generalized count distributions may be found in Antonio (2007) (pages 93–120). Both econometric (with conjugate distributions for the random effects) as well as general statistical specifications are discussed there. A collection of useful formulas with respect to prediction for panel data is also available.

Section 3.1 is a summary of the multilevel models that we investigated for our data. The corresponding results are presented and discussed in Section 3.2.

Table 6: *Vehicle level categorizations*

Covariate	**Categorization**	
Vehicle Age	0<=VAgeEntry<=4	Reference
	VAgeEntry>4	
Cubic Capacity	0<VehCapCubic<=1500	Reference
	VehCapCubic>1500	
Tonnage	0<Tonnage<=2	Reference
	2<Tonnage<=8	
	8<Tonnage	
TLengthEntry	TLengthEntry=0	Reference
	TLengthEntry>0	
Year	Year<=1997	Reference
	1997<Year<=1999	
	1999<Year	
	coverage is purchased.	

3.1 Model Specifications

3.1.1 Hierarchical Poisson Models

Starting point are hierarchical Poisson models with random intercepts for the vehicle, fleet and company.

Jewell's hierarchical model The first model is a Bayesian implementation of Jewell's hierarchical credibility model for counts. Jewell's credibility scheme (see Dannenburg

Table 7: *Fleet level categorizations*

Covariate	**Categorizations**	
AvNCD	AvNCD=0	
	AvNCD>0	Reference
AvTLengthEntry	AvTLengthEntry=0	Reference
	0<AvTLengthEntry	
AvTLength	AvTLength<=0.7	Reference
	AvTLength> 0.7	
AvVAge	0<= AvVAgeEntry <=4	
	4<AvAgeEntry	
FleetCap	FleetCap=1	Reference
	1<FleetCap<= 15	
	15<FleetCap	

et al. (1996) and Antonio and Beirlant (2007)) is the traditional actuarial approach for experience rating with hierarchical data. It is distribution–free in its original specification, but can be interpreted as a random effects model under normality assumptions (see Frees et al. (1999)). Our specification is given in (1), where $e_{c,f,v,t}$ is an exposure variable that gives the length of time during calendar year t for which the vehicle has insurance coverage. $Y_{c,f,v,t}$ denotes the claims observed in year t for vehicle v, which is insured under fleet f in company c.

- **Jewell's model**

$$Y_{c,f,v,t} \sim \text{Poi}(\lambda_{c,f,v,t}) \quad \text{with} \quad \lambda_{c,f,v,t} = e_{c,f,v,t} \exp\left(\eta_{c,f,v,t}\right)$$
$$\text{and} \quad \eta_{c,f,v,t} = \gamma + \epsilon_c + \epsilon_{c,f} + \epsilon_{c,f,v}, \tag{1}$$

where 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.

- **Random effects distributions used in** (1)

$$\epsilon_c \sim N(0,\sigma_c^2),\ \epsilon_{c,f} \sim N(0,\sigma_{c,f}^2) \text{ and } \epsilon_{c,f,v} \sim N(0,\sigma_{c,f,v}^2). \tag{2}$$

Hierarchical Poisson models We now incorporate risk factors corresponding with the four levels in the data set and random effects for the vehicle, fleet and company. In general:

- **Poisson model**

$$Y_{c,f,v,t} \sim \text{Poi}(\lambda_{c,f,v,t}) \quad \text{with} \quad \lambda_{c,f,v,t} = e_{c,f,v,t} \exp\left(\eta_{c,f,v,t} + \epsilon_c + \epsilon_{c,f} + \epsilon_{c,f,v}\right)$$
$$\text{and} \quad \eta_{c,f,v,t} := \gamma + \boldsymbol{X}_c\boldsymbol{\beta}_4 + \boldsymbol{X}_{c,f}\boldsymbol{\beta}_3 + \boldsymbol{X}_{c,f,v}\boldsymbol{\beta}_2 + \boldsymbol{X}_{c,f,v,t}\boldsymbol{\beta}_1. \tag{3}$$

Hereby ϵ_c, $\epsilon_{c,f}$ and $\epsilon_{c,f,v}$ are as previously defined. The explanatory variables used in $\eta_{c,f,v,t}$ are:

- $\boldsymbol{X}_{c,f,v,t}$: Year;
- $\boldsymbol{X}_{c,f,v}$: VehicleType, Capacity Cubic, Tonnage, VAgeEntry, TLengthEntry, Private, SwitchPol;
- $\boldsymbol{X}_{c,f}$: AvPrem, AvTLength, AvTLengthEntry, AvNCD; and
- $\boldsymbol{X}_c$: / (intercept γ is included in (3)).

However, for fleets with only one vehicle (there are 6,245 of such fleets in the sample) the general specification of the linear predictor is slightly modified. With only one vehicle per fleet, the vehicle and fleet level coincide. Therefore, no fleet random effect is included for such fleets, neither are averages at the level of the fleet used. Apart from the model in (3), we also consider a Poisson hierarchical model with a random effect for the company and fleet but without a random effect for the vehicle.

Our distributional assumptions for the random effects in the Poisson hierarchical models (as in (3)) slightly differ from those traditionally used in multilevel modelling:

- **Random effects distributions for** (3)

$$\epsilon_c \sim N(-\frac{\sigma_c^2}{2}, \sigma_c^2), \ \epsilon_{c,f} \sim N(-\frac{\sigma_{c,f}^2}{2}, \sigma_{c,f}^2), \ \text{and } \epsilon_{c,f,v} \sim N(-\frac{\sigma_{c,f,v}^2}{2}, \sigma_{c,f,v}^2). \tag{4}$$

Using the specifications in (3) and (4), the *a priori* mean, $\text{E}[Y_{c,f,v,t}]$, is given by

$$\begin{aligned}\text{E}[Y_{c,f,v,t}] &:= \lambda^{\text{prior}}_{c,f,v,t} \\ &= e_{c,f,v,t}\exp\left(\eta_{c,f,v,t}\right).\end{aligned} \quad (5)$$

The *a posteriori* premium, $\text{E}[Y_{c,f,v,t}|\epsilon_c, \epsilon_{c,f}, \epsilon_{c,f,v}]$ then becomes

$$\text{E}[Y_{c,f,v,t}|\epsilon_c, \epsilon_{c,f}, \epsilon_{c,f,v}] = \lambda^{\text{prior}}_{c,f,v,t}\exp\left(\epsilon_c\right)\exp\left(\epsilon_{c,f}\right)(\epsilon_{c,f,v}), \quad (6)$$

In our Bayesian analysis, the posterior distributions of (5) and (6) are used for ratemaking. Examples follow in Section 4. Specification (6) explicitly shows how *a posteriori* corrections are made to the *a priori* premium.

Priors Prior distributions used in the Bayesian analysis are selected as follows (similar specifications are used for the other models discussed in this Section):

(i) for the regression coefficients in $\boldsymbol{\beta}_4, \ldots, \boldsymbol{\beta}_1$: a normal prior with a variance of 10^6;

(ii) for the inverse of the variance components: gamma priors $\Gamma(0.001, 0.001)$.

3.1.2 Hierarchical Negative Binomial Model

Because the negative binomial provides a good fit to the 'raw' count data in Table 5, a negative binomial multilevel regression model is considered as well. This regression model uses the same covariate information as in Section 3.1.1. Thus,

- **Negative binomial model**

$$\begin{aligned}Y_{c,f,v,t} &\sim \text{NB}(\mu_{c,f,v,t}, \tau) \\ \text{where } \mu_{c,f,v,t} &= e_{c,f,v,t}\exp\left(\eta_{c,f,v,t} + \epsilon_c + \epsilon_{c,f}\right)\end{aligned} \quad (7)$$

where random effects for the company and fleet are used. Apart from this specification, we also examined the possible inclusion of an extra variability component at the level of the vehicle. However, convergence of the MCMC updates for this variance component could not be obtained within a reasonable number of iterations. The simpler model in (7) is therefore preferred. The random effects distributional assumptions in (4) result in $\text{E}[Y_{c,f,v,t}] = \mu^{\text{prior}}_{c,f,v,t} = e_{c,f,v,t}\exp\left(\eta_{c,f,v,t}\right)$. Prior distribution for τ was $\tau = \exp\left(\tau^\star\right)$ with $\tau^\star \sim N(0, 10^6)$.

3.1.3 Hierarchical Zero-Inflated Poisson Models

Two types of zero-inflated Poisson models were investigated. For the first specification, we have:

- **Zero-inflated Poisson model (1)**

$$\begin{aligned} Y_{c,f,v,t} &\sim \text{ZIP}(p, \lambda_{c,f,v,t}) \\ \text{where } \lambda_{c,f,v,t} &= e_{c,f,v,t} \exp\left(\eta_{c,f,v,t} + \epsilon_c + \epsilon_{c,f}\right). \end{aligned} \quad (8)$$

Prior specifications are similar as before, completed with $p \sim \text{Beta}(1,1)$ as prior for the additional proportion of zeroes. Using the random effects distributions in (4), the *a priori* premium for this ZIP model is $\text{E}[Y_{c,f,v,t}] = (1-p)\lambda^{\text{prior}}_{c,f,v,t}$ with $\lambda^{\text{prior}}_{c,f,v,t}$ as in (5). *A posteriori* $(1-p)\lambda^{\text{prior}}_{c,f,v,t} \exp\left(\epsilon_c\right) \exp\left(\epsilon_{c,f}\right)$ is used.

In a second hierarchical ZIP model we let the extra proportion of zeros be fleet-specific and use $p_{c,f} \sim \text{Beta}(1,b)$. Prior for b is $\log\left(b\right) \sim N(0, 10^6)$. Thus,

- **Zero-inflated Poisson model (2)**

$$\begin{aligned} Y_{c,f,v,t} &\sim \text{ZIP}(p_{c,f}, \lambda_{c,f,v,t}) \\ \text{where } \lambda_{c,f,v,t} &= e_{c,f,v,t} \exp\left(\eta_{c,f,v,t} + \epsilon_c + \epsilon_{c,f}\right), \end{aligned} \quad (9)$$

which results in an *a priori* premium $\text{E}[Y_{c,f,v,t}] = (1 - \frac{1}{1+b})\lambda^{\text{prior}}_{c,f,v,t}$ and *a posteriori* $(1 - p_{c,f})\lambda^{\text{prior}}_{c,f,v,t} \exp\left(\epsilon_c\right) \exp\left(\epsilon_{c,f}\right)$.

3.1.4 Hierarchical Hurdle Poisson Model

The hurdle Poisson model for panel data is extended towards multilevel data through the inclusion of level specific explanatory variables and random effects. The following specification is used:

- **Hurdle Poisson model**

$$\begin{aligned} Y_{c,f,v,t} &\sim \text{Hur}(p_{c,f,v,t}, \lambda_{c,f,v,t}) \\ \text{where } p_{c,f,v,t} &= \text{logit}\left(\eta_{c,f,v,t,\text{Bin}} + \epsilon_{c,\text{Bin}} \epsilon_{c,f,\text{Bin}}\right), \\ \text{and } \lambda_{c,f,v,t} &= e_{c,f,v,t} \exp\left(\eta_{c,f,v,t} + \epsilon_c + \epsilon_{c,f}\right). \end{aligned} \quad (10)$$

The risk factors in $\eta_{c,f,v,t,\text{Bin}}$ are

- intercept, Year;
- VehicleType, Private, VehicleAge, Capacity Cubic, Tonnage;
- AvTLength, AvTLengthEntry, AvPrem, AvNCD, SwitchPol;

and for $\eta_{c,f,v,t}$ the same set of explanatory variables is included (though this is not necessary). Note that this second part of the model (over the hurdle) is fitted to only 12% of the original data set. Distributional specifications for the random effects are those in (4). The random effects in the zero and non-zero part are independent of each other. These specifications however do not lead to an explicit expression for the *a priori* mean.

3.2 Results

3.2.1 Hierarchical Poisson Models

Estimated claim frequencies are in Table 8. In the table we compare the results obtained with Jewell's hierarchical model (full version), Jewell's hierarchical model with just fleet and company specific intercepts, a Poisson regression without random effects and the Poisson regression model with random intercepts for vehicles, fleets and companies. The table reports estimated claim counts obtained from hierarchical Poisson analysis. For every count model, two parallel chains were run; 15,000 iterations each with burn-in of 500 observations. 'RE' stands for random effects. We conclude that the Poisson multilevel model in (3) outperforms the other specifications.

3.2.2 Hierarchical Negative Binomial, Zero-Inflated and Hurdle Poisson Models

Table 9 compares estimated claims amounts for the preferred hierarchical Poisson model (3) and alternatives described in Sections 3.1.1- 3.1.4. On the basis of this table, the negative binomial, zero-inflated Poisson and hurdle Poisson outperformed hierarchical Poisson with the negative binomial (7) performing the best. We also remark that, among the negative binomial, zero-inflated Poisson and hurdle Poisson, the hurdle Poisson regression models have the advantage of allowing for the fastest MCMC sampling.

Table 8: *Estimated claim counts obtained from Bayesian Poisson hierarchical analyses. (95% credibility intervals are given in parens.)*

Num.	Obs. Freq.	Poisson No RE	Poisson Jewell (No Veh. RE)	Poisson Jewell (1)	Poisson Multilevel (3)
0	34,357	34,020	34,180	34,300	34,310
		(33,900;34,140)	(34,060;34,290)	(34,180;34,410)	(34,200;34,430)
1	4,104	4,665	4,395	4,204	4,176
		(4,564;4,767)	(4,299;4,491)	(4,104;4,306)	(4,081;4,273)
2	551	404	490	529	536
		(385;423)	(467;514)	(505;554)	(511;560)
3	86	26	52	76	79
		(24;28)	(47;57)	(68;84)	(71;87)
4	17	1.42	6	13	14
		(1.24;1.7)	(5;7)	(10;15)	(11;16)
5	5	0.088	0.69	2	3
		(0.05;0.22)	(0.49;1.03)	(1.8;3.37)	(2;4)

With the exception of the Poisson model, the estimation of a variability component at the level of the vehicle was problematic; the corresponding chains could not converge in a reasonable number of iterations. Hereby it is important to recall the descriptive statistics in Table 2; for the general insurers present in the data the cumulative period of exposure of a vehicle in a certain fleet was – on average – around 1. Thus, in general, the number of repeated measurements for a single vehicle in the same fleet is very small. Apart from this, the NB, ZIP and hurdle Poisson models – compared to the Poisson – incorporate already an extra heterogeneity or overdispersion component in their model specifications.

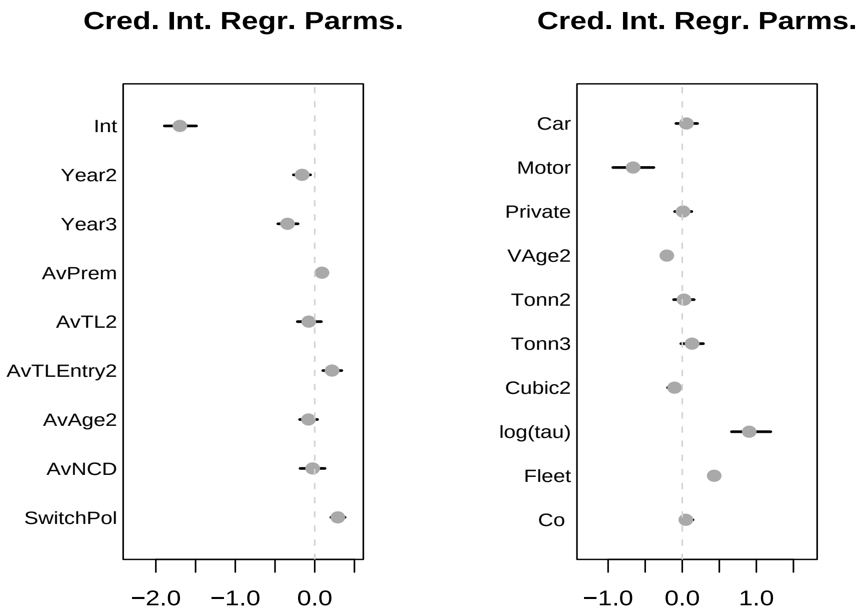
95% credibility intervals for the parameters used in the negative binomial model (7) are displayed in Figure 2. Similar credibility intervals for the parameters used in the zero–inflated regression (8) are in Figure 3. For the hurdle model (10), Figure 4 is included. These results are based on 2 parallel chains with 15,000 iterations each, after a burn-in of 1,000 iterations. To illustrate the mixing and convergence of the chains, Figure 6 is attached at the end of this paper. Similar plots were obtained for the other multilevel models.

In agreement with the findings in Frees and Valdez (2008) (who investigate a non fleet data set from the GIA), few risk factors have a statistically significant effect on the

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 2: *95% credibility intervals for negative binomial hierarchical model* (7).



average number of claims. Some observations are (to name a few): switching fleets (see ‘SwitchPol’) has a positive and significant effect on the number of claims, motors report significantly less claims than the reference class (i.e. trucks) and heavy trucks (‘Tonnage$>$8’) report significantly more claims. This is not surprising because switching

Figure 3: *95% credibility intervals for zero-inflated Poisson hierarchical model* (8).

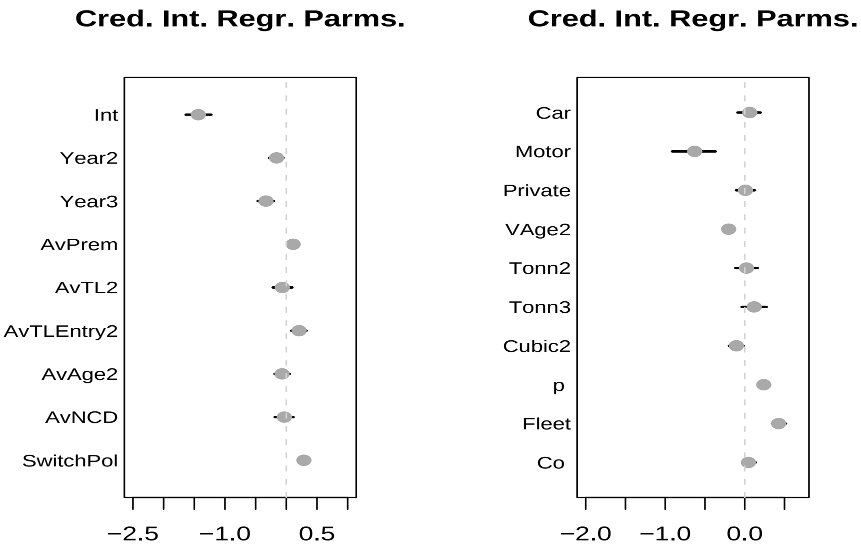
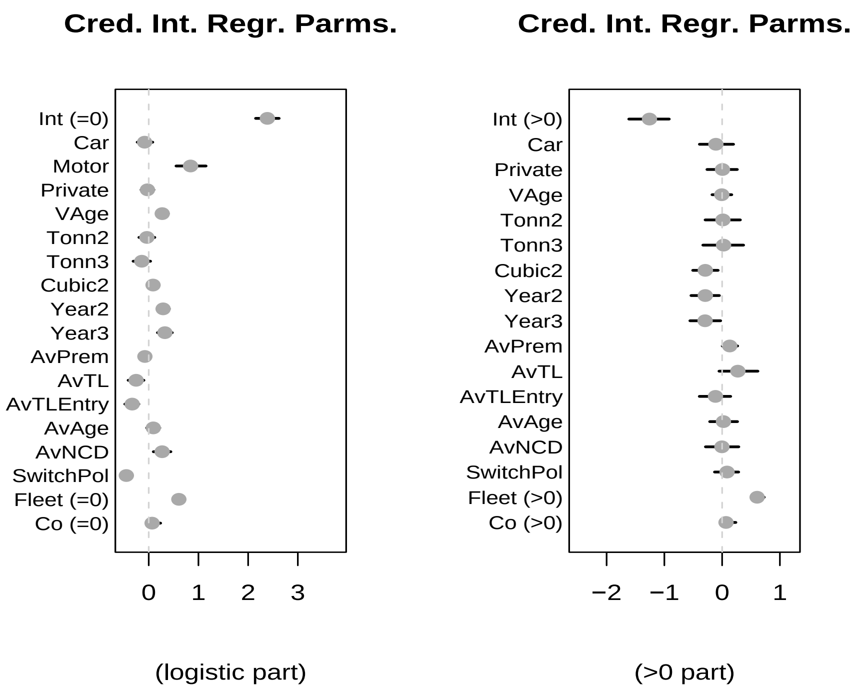


Figure 4: *95% credibility intervals for hurdle Poisson regression model* (10).



of policies is often motivated by increases or decreases in premiums resulting from the variability of claims. Regarding the averages at the fleet level: fleets with higher (average) premiums report more claims on average and staying longer within the same fleet (see 'AvTLength') has a negative effect on the number of claims. Again, this is consistent with fair actuarial premium calculations whereby premiums are directly linked to claims. Parameter estimates reported for the Poisson (3) and negative binomial model (7) closely correspond since they use the same specification for $\mathrm{E}[Y_{c,f,v,t}]$.

4 *A Priori* Premiums and *A Posteriori* Corrections

The very aim of the data analysis is a sound statistical approach to premium rating for intercompany data. How should a reinsurer or regulator translate the company effects that became apparent in the descriptive Table 1 into an accurate prediction for the expected number of claims? The posterior distribution of the random company effects is used for this purpose. Fleet effects can be taken into account in a similar manner. In Section 4.2 it is motivated that the different distributions used in Section 3.1 represent different styles of penalizing for past claims. For instance, the zero-inflated model with fleet-specific $p_{c,f}$ (see (9)) and the hurdle Poisson model in (10) not only use the number of past claims, but also the claim-free period of a fleet. The various distribution models considered in this paper give the user the choice of which style is suitably adoptable to its philosophy.

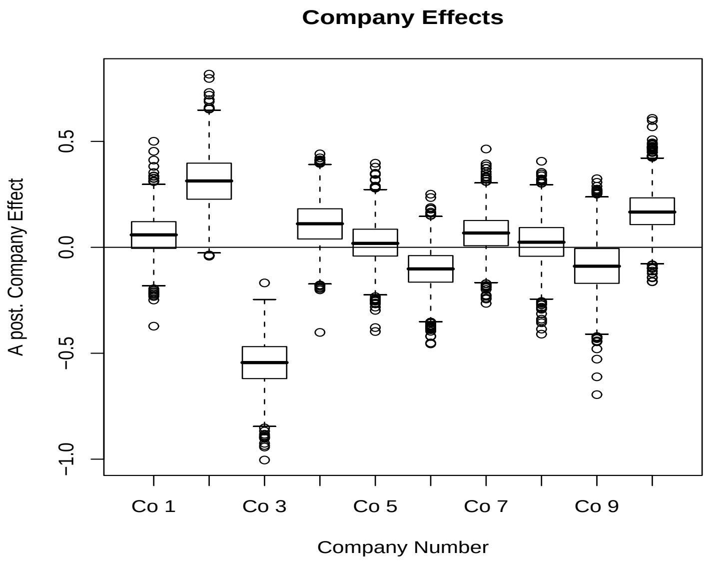
4.1 Posterior distributions for the random effects

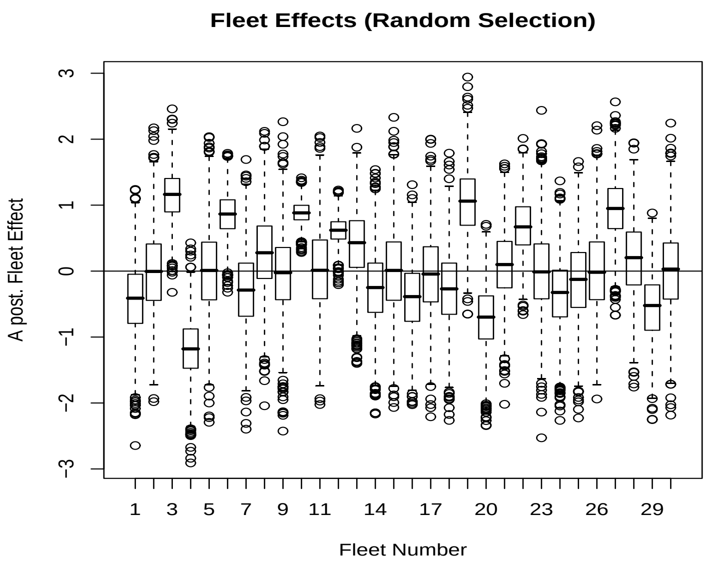
Figure 5 illustrates the *a posteriori* distributions for the company effects as well as for a random selection of fleet effects. The underlying model specification is Jewell's model (see (1)). We briefly comment:

- The company effects in the upper plot of Figure 5 are in line with the statistics from Table 1.
- Some descriptive statistics for the fleet effects in the lower plot of Figure 5: for fleet 1 zero claims are reported on a period of 11 years of exposure, for fleet 3 9 claims are reported on a period of 9 years of exposure, for fleet 4 zero claims

are reported on a period of 53 years of exposure and for fleet 10 41 claims are reported on a period of 100 years of exposure.

Figure 5: *Illustration of a posteriori distributions of company effects and a random selection of fleet effects.*





4.2 *A posteriori* corrections to *a priori* premiums

We select some vehicle and fleet profiles and examine how the reported claims history *a posteriori* corrects the *a priori* premium. Both premiums are calculated for the various model specifications introduced in Section 3. Recall from Section 3 that $\mathrm{E}[Y_{c,f,v,T+1}]$ is used for the *a priori* premium and $\mathrm{E}[Y_{c,f,v,T+1}|\epsilon_c, \epsilon_{c,f}, \epsilon_{c,f,v}]$ is used for the *a posteriori* premium. In the scenarios shown below 'SwitchPol' is always equal to 0.

4.2.1 Results for Poisson Hierarchical Model with Company, Fleet and Vehicle Effects

Let us start from the model in (3). To correct the *a priori* premium, this model uses the history of the vehicle, the history of the fleet to which it belongs and the history of the company. The results are printed in Table 10. Hereby the bonus–malus factor ('BMF') is the ratio (*a posteriori* premium/*a priori* premium). BMFs are standards used in the insurance industry for penalizing or rewarding customers according to their historical claims experience. Apparently, a BMF > 1 indicates some penalty required, while a BMF < 1 indicates the opposite. See Lemaire (1995) for more details.

4.2.2 Results for Poisson Hierarchical Model with Company and Fleet Effects

This model does not use the history of the vehicle separately, but relies on the history of the whole fleet to which the vehicle belongs and on the history of the company. The results are not displayed here due to space restrictions (see Antonio (2007), page 141 for full details). We briefly comment that, for instance, in Table 10, for fleet 6,592, the BMFs for all vehicles are > 1, but the BMF for the vehicle that reports 2 claims is much higher (=3.27) than the BMF for the claim-free vehicles (=1.45). Checking the corresponding results for the hierarchical Poisson model with just company and fleet random effects, the BMF for all vehicles is > 1 and in between those reported in Table 10.

4.2.3 Negative Binomial, Zero-Inflated Poisson and Hurdle Poisson Hierarchical Models with Company and Fleet Effects

For the scenarios used in Table 10, the NB, ZIP and hurdle poisson models (see (7), (8), (9) and (10)) and their corresponding *a priori* and *a posteriori* premiums and bonus–malus factors are calculated. We only display the results for (9) (see Table 11). The

Table 10: *Results for Poisson model in* (3).

Co.	Fleet	Vehicle	*A Priori* (Exp.)	*A Posteriori*	BMF	Acc. Cl. Fleet (Exp.)	Acc. Cl. Veh. (Exp.)
4	1,590	6,213	0.2141 (1)	0.6325	2.95	7 (15.25)	3 (1)
		6,261	0.2141 (1)	0.3641	1.7		0 (1.22)
1	4,370	10,104	0.1428 (1)	0.2895	2.03	7 (21.5)	2 (4.24)
		5,841	0.1428 (1)	0.1946	1.36		1 (7)
		7,152	0.1773 (1)	0.3675	2.07		3 (7)
5	4,673	9,350	0.07956 (0.5)	0.145	1.82	6 (18.5)	1 (1)
		12,131	0.07956 (0.5)	0.11	1.38		0 (1)
		12,210	0.07956 (0.5)	0.1102	1.39		0 (1)
4	6,592	1,656	0.1059 (1)	0.1529	1.44	12 (40)	0 (1.8)
		15,329	0.11 (1)	0.1597	1.45		0 (1.7)
		2,577	0.1315 (1)	0.4301	3.27		2 (2)
2	1,485	11,122	0.01683 (0.08)	0.09327	5.54	17 (40)	4 (1)
		10,782	0.01223 (0.08)	0.02652	2.17		0 (1)
		11,063	0.01519 (0.08)	0.03234	2.13		0 (1)
3	4,672	12,007	0.07028 (0.334)	0.06179	0.88	5 (20.4)	0 (2)
		8,367	0.07028 (0.334)	0.08484	1.21		1 (1.334)
		11,958	0.07028 (0.334)	0.06294	0.9		0 (2)
5	1,842	1,826	0.1473 (1)	0.1024	0.7	2 (16)	0 (2)
		1,569	0.1473 (1)	0.09015	0.61		0 (4)
6	5,992	1,906	0.183 (1)	0.2373	1.3	7 (21)	1 (5)
		1,889	0.183 (1)	0.2886	1.58		2 (6)
9	5,823	1,020	0.109 (1)	0.07598	0.7	2 (16)	0 (3)
		1,056	0.109 (1)	0.07591	0.7		0 (3)
		1,025	0.109 (1)	0.09828	0.9		1 (3)
10	3,564	15,564	0.1929 (1)	0.1797	0.93	2 (17)	2 (5)
		14,831	0.1553 (1)	0.1249	0.8		1 (4)
		15,194	0.1553 (1)	0.08599	0.55		0 (7)
10	3,568	1,119	0.1512 (1)	0.1391	0.92	3 (19.25)	1 (3.7)
		1,206	0.1512 (1)	0.1934	1.28		3 (5.75)
		1,540	0.1512 (1)	0.09953	0.66		0 (5)

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.

other results are not printed, due to space restrictions, but are available in Antonio (2007).

Tables 10 and 11 illustrate the calculations of the bonus–malus factors that can be used to update premiums based on claims experience. As a result of the statistical modeling, Table 11 shows BMFs calculated at the fleet level, a natural point in the hierarchy because it is at this level where an insurance contract between a fleet and

Table 11: *Results for the zero–inflated Poisson model in* (9).

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.

insurance company is written. Hence, fleet level BMFs can be used for premium renewals. Table 10 shows BMFs calculated at the vehicle level. This information could also be used for contracts written at the fleet level; as the fleet composition changes through the retirement or sale of vehicles, the total fleet premium should reflect the changing composition of vehicles. Vehicle level BMFs will allow prices to depend on the vehicle composition of fleets. We anticipate that pricing actuaries will find both set of findings useful.

Some brief comments:

- *a priori* premiums obtained with the different model specifications closely correspond;

- the zero-inflated model with fleet-specific $p_{c,f}$ (see (9)) and the hurdle poisson model in (10) take the claim-free period of a fleet into account. For panel data this feature was made explicit in the derivations in Antonio (2007) (pages 105–107) and the references mentioned there. Compare the results for fleet 4,673 between the various specifications: in the Poisson, NB and ZIP with p fixed, the BMF for this fleet is about 1.5. In ZIP model (9) this drops to 1.33 and in the hurdle model even to 0.9. That is because these last two model specifications not only use the number of registered claims, but also the claim-free periods (which is here 17 out of a total of 18.5 years).

In Table 12 some artificial scenarios are investigated. Fleet 4,672 (see also Tables 10 and 11) originally belongs to company 3, but the observations corresponding with this fleet were switched to different companies (namely company 2, 6, 7 and 10). Results are for the hierarchical negative binomial model in (7). The *a priori* premiums closely correspond, but *a posteriori* premiums reflect the company differences that became apparent in Figure 5.

Table 12: *Fleet 4,672 under different companies: results for negative binomial model in* (7).

	Company				
	2	3	6	7	10
A Priori	0.0695	0.071	0.07126	0.0712	0.07102
A Posteriori	0.0911	0.06967	0.08112	0.08383	0.08778

5 Conclusion

This paper presents a multilevel analysis of a four–level intercompany data set on claim counts for fleet policies. The data come from the General Insurance Association in Singapore and the observations are from 10 Singaporean general insurers. We build multilevel models using generalized count distributions (Poisson, negative binomial, hurdle Poisson and zero–inflated Poisson) and use Bayesian estimation techniques. The effect of explanatory variables at the different levels in the data set is investigated. We find that in all models considered, there is the importance of accounting for the effects of the various levels. The results also indicate possible different styles for penalizing or rewarding past claims. 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. A comparison of these calculated premiums results in bonus–malus factors which are important in establishing experience–rated premiums. Insurers, reinsurers and regulators can use the methodology recommended in this paper to study the differences in riskiness among fleets and companies. *A posteriori* predictions for a specific fleet, vehicle or company can be readily calculated from the estimated multilevel models.

References

J.-F. Angers, D. Derjardins, G. Dionne, and F. Guertin. Vehicle and fleet random effects in a model of insurance rating for fleets of vehicles. *ASTIN Bulletin*, 36(1):25–77, 2006.

K. Antonio. *Statistical Tools for Non-Life Insurance: Essays on Claims Reserving and Ratemaking for Panels and Fleets*. PhD Thesis, Katholieke Universiteit Leuven, Belgium, 2007.

K. Antonio and J. Beirlant. Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, 40(1):58–76, 2007.

C. Bolancé, M. Guillén, and J. Pinquet. Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on random effects. *Insurance: Mathematics and Economics*, 33:273–282, 2003.

J.-P. Boucher, M. Denuit, and M. Guillén. Risk classification for claim counts: a comparative analysis of various zero–inflated mixed poisson and hurdle models. *North American Actuarial Journal*, 11(4):110–131, 2007.

A.C. Cameron and P.K. Trivedi. *Regression analysis of count data*. Cambridge University Press, 1998.

D.R. Dannenburg, R. Kaas, and M. Goovaerts. *Practical actuarial credibility models*. Institute of actuarial science and econometrics, University of Amsterdam, 1996.

M. Denuit, X. Maréchal, S. Pitrebois, and J.-F. Walhin. *Actuarial Modelling Of Claim Counts: Risk Classification, Credibility and Bonus-Malus Scales*. Wiley, 2007.

D. Desjardins, G. Dionne, and J. Pinquet. Experience rating scheme for fleets of vehicles. *ASTIN Bulletin*, 31(1):81–105, 2001.

G. Dionne and C. Vanasse. A generalization of actuarial automobile insurance rating models: the Negative Binomial distribution with a regression component. *ASTIN Bulletin*, 19:199–212, 1989.

E.W. Frees and E.A. Valdez. Hierarchical insurance claims modeling. *To appear in Journal of the American Statistical Association*, 2008.

E.W. Frees, V.R. Young, and Y. Luo. A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24(3):229–247, 1999.

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

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

J.C. Hickman and L. Heacox. Credibility theory: the cornerstone of actuarial science. *North American Actuarial Journal*, 3(2):1–8, 1999.

B. Iverson, J. Luff, S. Siegel, and R. Stryker. The SOA: A place for research – answer to all your questions about the research done by the SOA. *The Actuary Magazine*, October/November, 2007.

W.S. Jewell. The use of collateral data in credibility theory: a hirerachical model. *Giornale dell'Instituto Italiano degli Attuari*, 38:1–6, 1975.

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

A. Lee, K. Wang, J. Scott, K. Yau, and G. McLachlan. Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods In Medical Research*, 15(1):47–61, 2006.

J. Lemaire. *Bonus–malus systems in automobile insurance*. Springer-Verlag, New York, 1995.

R. Norberg. Hierarchical credibility: analysis of a random effect linear model with nested classification. *Scandinavian Actuarial Journal*, pages 204–222, 1986.

J. Pinquet. Allowance for cost of claims in bonus-malus systems. *ASTIN Bulletin*, 27(1): 33–57, 1997.

J. Pinquet. Designing optimal bonus-malus systems from different types of claims. *ASTIN Bulletin*, 28(2):205–229, 1998.

J. Pinquet, M. Guillén, and C. Bolancé. Allowance for age of claims in bonus-malus systems. *ASTIN Bulletin*, 31(2):337–348, 2001.

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

T.A.B. Snijders and R.J. Bosker. *Multilevel Analysis: an introduction to basic and advanced multilevel modeling*. Sage Publications, London, 1999.

B. Sundt. A multi-level hierarchical credibility regression model. *Scandinavian Actuarial Journal*, pages 25–32, 1980.

G. C. Taylor. Credibility analysis of a general hierarchical model. *Scandinavian Actuarial Journal*, pages 1–12, 1979.

R. Winkelmann. *Econometric Analysis of Count Data*. Springer-Verlag, Berlin, 2003.

K. Yau, K. Wang, and A. Lee. Zero-inflated negative binomial mixed regression model. *Biometrical Journal*, 45(4):437–452, 2003.

K. Yip and K. Yau. On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, 36:153–163, 2005.

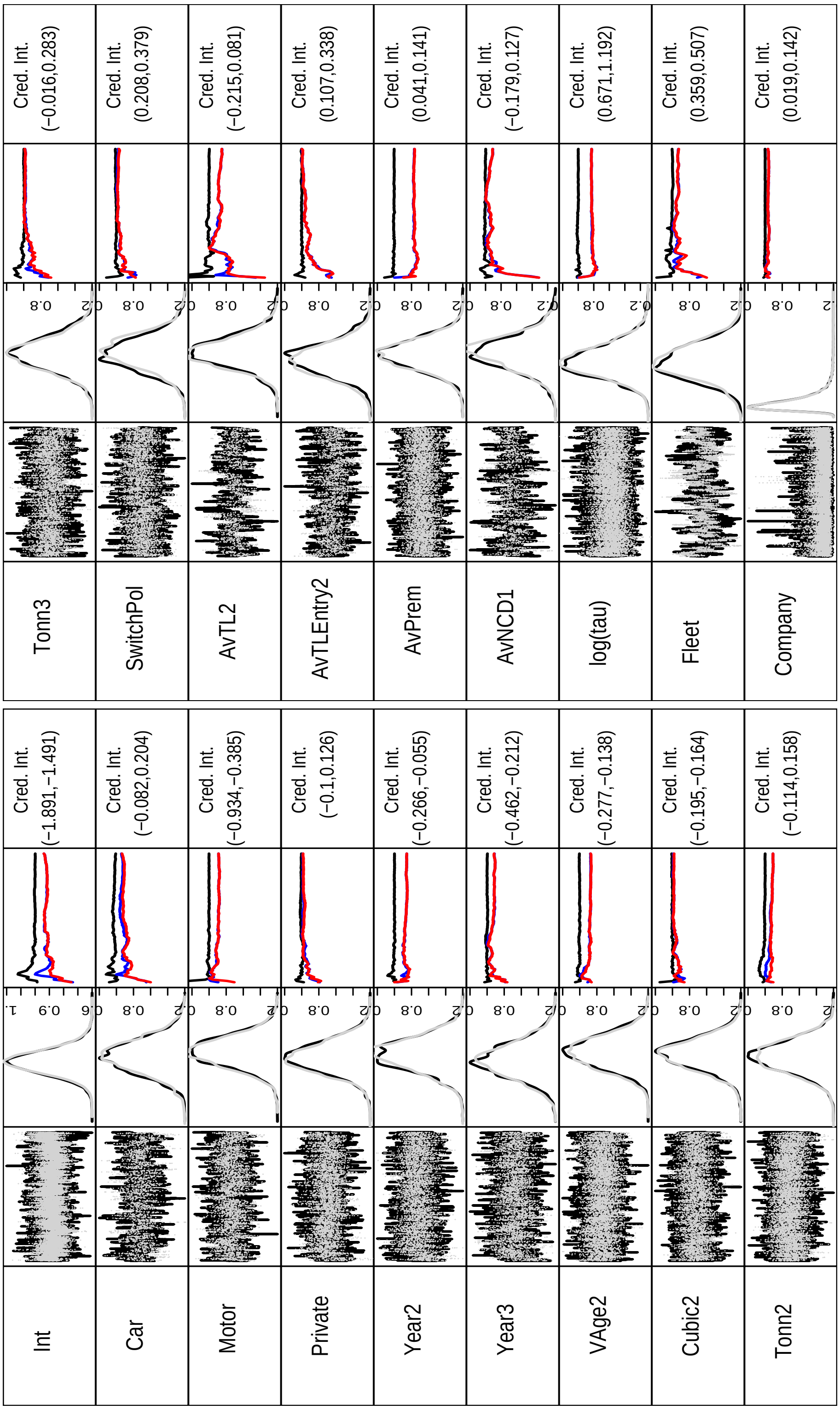


Figure 6: *Structure (7): convergence diagnostics for 2 chains with 15,000 iterations each. Burn-in of 500 iterations. Trace-plot, density, BGR convergence diagnostic and 95% credibility interval.*