Using micro-level automobile insurance data for macro-effects inference

Emiliano A. Valdez, Ph.D., F.S.A.
University of Connecticut

joint work with E.W. Frees*, P. Shi**, K. Antonio***

* University of Wisconsin – Madison
** Northern Illinois University
*** K.U. Leuven

Universidad de Barcelona, España
16-18 July 2012
Outline

Introduction

Model estimation
  Data
  Models of each component

Macro-effects inference
  Individual risk rating
  A case study
  Predictive distributions for portfolios
  Predictive distributions for reinsurance

Conclusion

Appendix A - Parameter Estimates

Appendix B - Singapore
  Insurance market

Intercompany experience data analysis
  References
A collection of work


"Policyholder" $i$ is followed over time $t = 1, \ldots, 9$ years

Unit of analysis "it" – a registered vehicle insured $i$ over time $t$ (year)

Have available: exposure $e_{it}$ and covariates (explanatory variables) $x_{it}$

- covariates often include age, gender, vehicle type, driving history and so forth

Goal: understand how time $t$ and covariates impact claims $y_{it}$.

Statistical methods viewpoint

- basic regression set-up - almost every analyst is familiar with:
  - part of the basic actuarial education curriculum

- incorporating cross-sectional and time patterns is the subject of longitudinal data analysis - a widely available statistical methodology
More complex data set-up

- Some variations that might be encountered when examining insurance company records
- For each “it”, could have multiple claims, \( j = 0, 1, \ldots, 5 \)
- For each claim \( y_{itj} \), possible to have one or a combination of three (3) types of losses:
  1. losses for injury to a party other than the insured \( y_{itj,1} \) - “injury”;
  2. losses for damages to the insured, including injury, property damage, fire and theft \( y_{itj,2} \) - “own damage”; and
  3. losses for property damage to a party other than the insured \( y_{itj,3} \) - “third party property”.
- Distribution for each claim is typically medium to long-tail
- The full multivariate claim may not be observed. For example:

<table>
<thead>
<tr>
<th>Distribution of Claims, by Claim Type Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of M</td>
</tr>
<tr>
<td>Claim by Combination</td>
</tr>
<tr>
<td>( y_1 )</td>
</tr>
<tr>
<td>Percentage</td>
</tr>
</tbody>
</table>
The hierarchical insurance claims model

- Traditional to predict/estimate insurance claims distributions:
  
  Cost of Claims = Frequency \times Severity

- Joint density of the aggregate loss can be decomposed as:

  \[ f(N, M, y) = f(N) \times f(M|N) \times f(y|N, M) \]

  \text{joint} = \text{frequency} \times \text{conditional claim-type} \times \text{conditional severity}.

- This natural decomposition allows us to investigate/model each component separately.
Model features

- Allows for risk rating factors to be used as explanatory variables that predict both the frequency and the multivariate severity components.
- Helps capture the long-tail nature of the claims distribution through the GB2 distribution model.
- Provides for a “two-part” distribution of losses - when a claim occurs, not necessary that all possible types of losses are realized.
- Allows to capture possible dependencies of claims among the various types through a $t$-copula specification.
Literature on claims frequency/severity

- There is large literature on modeling claims frequency and severity
  - Klugman, Panjer and Willmot (2004) - basics without covariates
  - Kahane and Levy (*JRI*, 1975) - first to model joint frequency/severity with covariates.
  - Coutts (1984) postulates that the frequency component is more important to get right.
    - Many recent papers on frequency, e.g., Boucher and Denuit (2006)

- Applications to motor insurance:
  - Brockman and Wright (1992) - good early overview.
  - Renshaw (1994) - uses GLM for both frequency and severity with policyholder data.
  - Pinquet (1997, 1998) - uses the longitudinal nature of the data, examining policyholders over time.
    - considered 2 lines of business: claims at fault and not at fault; allowed correlation using a bivariate Poisson for frequency; severity models used were lognormal and gamma.
  - Most other papers use grouped data, unlike our work.
Data

- Model is calibrated with detailed, micro-level automobile insurance records over eight years [1993 to 2000] of a randomly selected Singapore insurer.
  - Year 2001 data used for out-of-sample prediction
- Information was extracted from the policy and claims files.
- Unit of analysis - a registered vehicle insured $i$ over time $t$ (year).
- The observable data consist of
  - number of claims within a year: $N_{it}$, for $t = 1, \ldots, T_i$, $i = 1, \ldots, n$
  - type of claim: $M_{itj}$ for claim $j = 1, \ldots, N_{it}$
  - the loss amount: $y_{itjk}$ for type $k = 1, 2, 3$.
  - exposure: $e_{it}$
  - vehicle characteristics: described by the vector $x_{it}$
- The data available therefore consist of

$$\{e_{it}, x_{it}, N_{it}, M_{itj}, y_{itjk}\}.$$
Risk factor rating system

- Insurers adopt “risk factor rating system” in establishing premiums for motor insurance.

- Some risk factors considered:
  - vehicle characteristics: make/brand/model, engine capacity, year of make (or age of vehicle), price/value
  - driver characteristics: age, sex, occupation, driving experience, claim history
  - other characteristics: what to be used for (private, corporate, commercial, hire), type of coverage

- The “no claims discount” (NCD) system:
  - rewards for safe driving
  - discount upon renewal of policy ranging from 0 to 50%, depending on the number of years of zero claims.

- These risk factors/characteristics help explain the heterogeneity among the individual policyholders.
Covariates

- **Year**: the calendar year - 1993-2000; treated as continuous variable.
- **Vehicle Type**: automotive (A) or others (O).
- **Vehicle Age**: in years, grouped into 6 categories -
  - 0, 1-2, 3-5, 6-10, 11-15, ≥16.
- **Vehicle Capacity**: in cubic capacity.
- **Gender**: male (M) or female (F).
- **Age**: in years, grouped into 7 categories -
- **The NCD applicable for the calendar year**: 0%, 10%, 20%, 30%, 40%, and 50%.
Random effects negative binomial count model

- Let $\lambda_{it} = e_{it} \exp\left( x_{\lambda,it}' \beta_\lambda \right)$ be the conditional mean parameter for the \{it\} observational unit, where
  - $x_{\lambda,it}$ is a subset of $x_{it}$ representing the variables needed for frequency modeling.

- Negative binomial distribution model with parameters $p$ and $r$:
  - $\text{Pr}(N = k | r, p) = \binom{k + r - 1}{r - 1} p^r (1 - p)^k$.
  - Here, $\sigma = \frac{1}{r}$ is the dispersion parameter and $p = p_{it}$ is related to the mean through
    \[
    \frac{1 - p_{it}}{p_{it}} = \lambda_{it} \sigma = e_{it} \exp(x_{\lambda,it}' \beta_\lambda) \sigma.
    \]
Multinomial claim type

- Certain characteristics help describe the claims type.

- To explain this feature, we use the multinomial logit of the form

\[
\Pr(M = m) = \frac{\exp(V_m)}{\sum_{s=1}^{7} \exp(V_s)},
\]

where \( V_m = V_{it,m} = x'_{M,it} \beta_{M,m}. \)

- For our purposes, the covariates in \( x_{M,it} \) do not depend on the accident number \( j \) nor on the claim type \( m \), but we do allow the parameters to depend on type \( m \).

- Such has been proposed in Terza and Wilson (1990).

- An alternative model to claim type, multivariate probit, was considered in:
  - Young, Valdez and Kohn (2009)
Severity

- We are particularly interested in accommodating the long-tail nature of claims.
- We use the generalized beta of the second kind (GB2) for each claim type with density

\[ f(y) = \frac{\exp(\alpha_1 z)}{y|\sigma|B(\alpha_1, \alpha_2)[1 + \exp(z)]^{\alpha_1 + \alpha_2}}, \]

where \( z = (\ln y - \mu)/\sigma \).

- \( \mu \) is a location, \( \sigma \) is a scale and \( \alpha_1 \) and \( \alpha_2 \) are shape parameters.
- With four parameters, distribution has great flexibility for fitting heavy tailed data.
- Introduced by McDonald (1984), used in insurance loss modeling by Cummins et al. (1990).
- Many distributions useful for fitting long-tailed distributions can be written as special or limiting cases of the GB2 distribution; see, for example, McDonald and Xu (1995).
GB2 Distribution

"Transformed Beta" Family of Distributions

Fig. 4.7 Distributional relationships and characteristics.

Source: Klugman, Panjer and Willmot (2004), p. 72

E.A. Valdez (U. of Connecticut) Barcelona Summer School, Day 1 16-18 July 2012
Heavy-tailed regression models

- Loss Modeling - Actuaries have a wealth of knowledge on fitting claims distributions. (Klugman, Panjer, Willmot, 2004) (Wiley)
  - Data are often “heavy-tailed” (long-tailed, fat-tailed)
  - Extreme values are likely to occur
  - Extreme values are the most interesting - do not wish to downplay their importance via transformation

- Studies of financial asset returns is another good example Rachev et al. (2005) “Fat-Tailed and Skewed Asset Return Distributions” (Wiley)

- Healthcare expenditures - Typically skewed and fat-tailed due to a few yet high-cost patients (Manning et al., 2005, J. of Health Economics)
GB2 regression

- We allow scale and shape parameters to vary by type and thus consider $\alpha_{1k}, \alpha_{2k}$ and $\sigma_k$ for $k = 1, 2, 3$.

- Despite its prominence, there are relatively few applications that use the GB2 in a regression context:
  - McDonald and Butler (1990) used the GB2 with regression covariates to examine the duration of welfare spells.
  - Beirlant et al. (1998) demonstrated the usefulness of the Burr XII distribution, a special case of the GB2 with $\alpha_1 = 1$, in regression applications.
  - Sun et al. (2008) used the GB2 in a longitudinal data context to forecast nursing home utilization.

- We parameterize the location parameter as $\mu_{ik} = x_{ik}' \beta_k$:
  - Thus, $\beta_{k,j} = \partial \ln \mathbb{E}(Y | x) / \partial x_j$
  - Interpret the regression coefficients as proportional changes.
Dependencies among claim types

- We use a parametric copula (in particular, the $t$ copula).

- Suppressing the $\{i\}$ subscript, we can express the joint distribution of claims $(y_1, y_2, y_3)$ as

  \[ F(y_1, y_2, y_3) = H(F_1(y_1), F_2(y_2), F_3(y_3)) . \]

- Here, the marginal distribution of $y_k$ is given by $F_k(\cdot)$ and $H(\cdot)$ is the copula.

- Modeling the joint distribution of the simultaneous occurrence of the claim types, when an accident occurs, provides the unique feature of our work.

Macro-effects inference

- Analyze the risk profile of either a single individual policy, or a portfolio of these policies.

- Three different types of actuarial applications:
  - Predictive mean of losses for individual risk rating
    - allows the actuary to differentiate premium rates based on policyholder characteristics.
    - quantifies the non-linear effects of coverage modifications like deductibles, policy limits, and coinsurance.
    - possible “unbundling” of contracts.
  - Predictive distribution of portfolio of policies
    - assists insurers in determining appropriate economic capital.
    - measures used are standard: value-at-risk (VaR) and conditional tail expectation (CTE).
  - Examine effects on several reinsurance treaties
    - quota share versus excess-of-loss arrangements.
    - analysis of retention limits at both the policy and portfolio level.
Individual risk rating

- The estimated model allowed us to calculate **predictive means** for several alternative policy designs.
  - based on the 2001 portfolio of the insurer of \( n = 13,739 \) policies.
- For alternative designs, we considered four random variables:
  - individuals losses, \( y_{ijk} \)
  - the sum of losses from a type, \( S_{i,k} = y_{i,1,k} + \ldots + y_{i,N_i,k} \)
  - the sum of losses from a specific event, 
    \[ S_{EVENT,i,j} = y_{i,j,1} + y_{i,j,2} + y_{i,j,3}, \text{ and} \]
  - an overall loss per policy,
    \[ S_i = S_{i,1} + S_{i,2} + S_{i,3} = S_{EVENT,i,1} + \ldots + S_{EVENT,i,N_i}. \]
- These are ways of “unbundling” the comprehensive coverage, similar to decomposing a financial contract into primitive components for risk analysis.
Modifications of standard coverage

- We also analyze modifications of standard coverage
  - deductibles $d$
  - coverage limits $u$
  - coinsurance percentages $\alpha$

- These modifications alter the claims function

$$g(y; \alpha, d, u) = \begin{cases} 
0 & y < d \\
\alpha(y - d) & d \leq y < u \\
\alpha(u - d) & y \geq u 
\end{cases}.$$
Calculating the predictive means

- Define $\mu_{ik} = E(y_{ijk}|N_i, K_i = k)$ from the conditional severity model with an analytic expression
  \[
  \mu_{ik} = \exp(x'_{ik}\beta_k) \frac{B(\alpha_{1k} + \sigma_k, \alpha_{2k} - \sigma_k)}{B(\alpha_{1k}, \alpha_{1k})}.
  \]

- Basic probability calculations show that:
  \[
  E(y_{ijk}) = \Pr(N_i = 1)\Pr(K_i = k)\mu_{ik},
  \]
  \[
  E(S_{i,k}) = \mu_{ik}\Pr(K_i = k) \sum_{n=1}^{\infty} n\Pr(N_i = n),
  \]
  \[
  E(S_{EVENT,i,j}) = \Pr(N_i = 1) \sum_{k=1}^{3} \mu_{ik}\Pr(K_i = k), \text{ and}
  \]
  \[
  E(S_i) = E(S_{i,1}) + E(S_{i,2}) + E(S_{i,3}).
  \]

- In the presence of policy modifications, we approximate this using simulation (Appendix A.2).
A case study

To illustrate the calculations, we chose at a randomly selected policyholder from our database with characteristic:

- 50-year old female driver who owns a Toyota Corolla manufactured in year 2000 with a 1332 cubic inch capacity.
- for losses based on a coverage type, we chose “own damage” because the risk factors NCD and age turned out to be statistically significant for this coverage type.

The point of this exercise is to evaluate and compare the financial significance.
Predictive means by level of NCD and by insured’s age

### Table 3. Predictive Mean by Level of NCD

<table>
<thead>
<tr>
<th>Type of Random Variable</th>
<th>Level of NCD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Individual Loss (Own Damage)</td>
<td>330.67</td>
</tr>
<tr>
<td>Sum of Losses from a Type (Own Damage)</td>
<td>436.09</td>
</tr>
<tr>
<td>Sum of Losses from a Specific Event</td>
<td>495.63</td>
</tr>
<tr>
<td>Overall Loss per Policy</td>
<td>653.63</td>
</tr>
</tbody>
</table>

### Table 4. Predictive Mean by Insured’s Age

<table>
<thead>
<tr>
<th>Type of Random Variable</th>
<th>Insured’s Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤ 21</td>
</tr>
<tr>
<td>Individual Loss (Own Damage)</td>
<td>258.41</td>
</tr>
<tr>
<td>Sum of Losses from a Type (Own Damage)</td>
<td>346.08</td>
</tr>
<tr>
<td>Sum of Losses from a Specific Event</td>
<td>479.46</td>
</tr>
<tr>
<td>Overall Loss per Policy</td>
<td>642.14</td>
</tr>
</tbody>
</table>
Macro-effects inference A case study

Predictive means by level of NCD and by insured’s age

- **NCD**
  - Predictive means decrease as NCD increases
  - Predictive means increase as the random variable covers more potential losses
  - Confidence intervals indicate that 5,000 simulations is sufficient for exploratory work

- **Age**
  - Effect of age is non-linear.
Predictive means and confidence intervals
## Coverage modifications by level of NCD

### Table 5. Simulated Predictive Mean by Level of NCD and Coverage Modifications

<table>
<thead>
<tr>
<th>Coverage Modification</th>
<th>Deductible Limits</th>
<th>Coinsurance</th>
<th>Level of NCD</th>
<th>Individual Loss (Own Damage)</th>
<th>Sum of Losses from a Type (Own Damage)</th>
<th>Sum of Losses from a Specific Event</th>
<th>Overall Loss per Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td><strong>Individual Loss</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 none</td>
<td>none</td>
<td>1</td>
<td>1</td>
<td>339.78</td>
<td>300.78</td>
<td>263.28</td>
<td>254.40</td>
</tr>
<tr>
<td>250 none</td>
<td>none</td>
<td>1</td>
<td>0.75</td>
<td>331.55</td>
<td>295.08</td>
<td>260.77</td>
<td>250.53</td>
</tr>
<tr>
<td>500 none</td>
<td>none</td>
<td>1</td>
<td>0.5</td>
<td>337.00</td>
<td>300.00</td>
<td>263.28</td>
<td>254.36</td>
</tr>
<tr>
<td>0 25,000</td>
<td>none</td>
<td>1</td>
<td>0.75</td>
<td>254.84</td>
<td>225.59</td>
<td>197.46</td>
<td>190.80</td>
</tr>
<tr>
<td>0 50,000</td>
<td>none</td>
<td>1</td>
<td></td>
<td>169.89</td>
<td>150.39</td>
<td>131.64</td>
<td>127.20</td>
</tr>
<tr>
<td>0 25,000</td>
<td>none</td>
<td>0.5</td>
<td>0.75</td>
<td>225.00</td>
<td>199.51</td>
<td>174.76</td>
<td>167.43</td>
</tr>
<tr>
<td>0 50,000</td>
<td>none</td>
<td>0.5</td>
<td></td>
<td>208.05</td>
<td>184.02</td>
<td>158.49</td>
<td>152.54</td>
</tr>
<tr>
<td><strong>Sum of Losses</strong></td>
<td></td>
<td></td>
<td>0</td>
<td>245.80</td>
<td>216.04</td>
<td>186.00</td>
<td>179.91</td>
</tr>
<tr>
<td>250 none</td>
<td>none</td>
<td>1</td>
<td>0.75</td>
<td>445.81</td>
<td>386.04</td>
<td>334.05</td>
<td>322.09</td>
</tr>
<tr>
<td>500 none</td>
<td>none</td>
<td>1</td>
<td></td>
<td>475.56</td>
<td>410.12</td>
<td>374.90</td>
<td>358.54</td>
</tr>
<tr>
<td>0 25,000</td>
<td>none</td>
<td>1</td>
<td>0.75</td>
<td>434.35</td>
<td>385.05</td>
<td>333.98</td>
<td>321.87</td>
</tr>
<tr>
<td>0 50,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>334.36</td>
<td>289.53</td>
<td>250.54</td>
<td>241.56</td>
</tr>
<tr>
<td>0 none</td>
<td>none</td>
<td>0.5</td>
<td></td>
<td>222.91</td>
<td>193.02</td>
<td>167.03</td>
<td>161.04</td>
</tr>
<tr>
<td>250 25,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>298.82</td>
<td>259.09</td>
<td>224.32</td>
<td>214.33</td>
</tr>
<tr>
<td>500 50,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>279.75</td>
<td>241.77</td>
<td>206.06</td>
<td>197.94</td>
</tr>
<tr>
<td><strong>Sum of Losses</strong></td>
<td></td>
<td></td>
<td>0</td>
<td>425.74</td>
<td>364.50</td>
<td>323.33</td>
<td>305.88</td>
</tr>
<tr>
<td>250 none</td>
<td>none</td>
<td>1</td>
<td>0.75</td>
<td>512.74</td>
<td>444.50</td>
<td>390.87</td>
<td>376.92</td>
</tr>
<tr>
<td>500 none</td>
<td>none</td>
<td>1</td>
<td></td>
<td>475.56</td>
<td>410.12</td>
<td>374.90</td>
<td>358.54</td>
</tr>
<tr>
<td>0 25,000</td>
<td>none</td>
<td>1</td>
<td>0.75</td>
<td>439.84</td>
<td>377.11</td>
<td>343.33</td>
<td>327.64</td>
</tr>
<tr>
<td>0 50,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>384.55</td>
<td>333.38</td>
<td>305.88</td>
<td>293.15</td>
</tr>
<tr>
<td>0 none</td>
<td>none</td>
<td>0.5</td>
<td></td>
<td>256.37</td>
<td>222.25</td>
<td>190.92</td>
<td>188.46</td>
</tr>
<tr>
<td>250 25,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>335.02</td>
<td>299.17</td>
<td>271.39</td>
<td>261.15</td>
</tr>
<tr>
<td>500 50,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>315.98</td>
<td>261.00</td>
<td>233.11</td>
<td>224.67</td>
</tr>
<tr>
<td><strong>Sum of Losses</strong></td>
<td></td>
<td></td>
<td>0</td>
<td>508.72</td>
<td>440.60</td>
<td>391.33</td>
<td>378.92</td>
</tr>
<tr>
<td>250 none</td>
<td>none</td>
<td>1</td>
<td>0.75</td>
<td>672.68</td>
<td>572.51</td>
<td>516.77</td>
<td>493.93</td>
</tr>
<tr>
<td>500 none</td>
<td>none</td>
<td>1</td>
<td></td>
<td>629.88</td>
<td>533.50</td>
<td>479.64</td>
<td>457.56</td>
</tr>
<tr>
<td>0 25,000</td>
<td>none</td>
<td>1</td>
<td>0.75</td>
<td>588.55</td>
<td>495.85</td>
<td>443.87</td>
<td>422.63</td>
</tr>
<tr>
<td>0 50,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>649.67</td>
<td>568.30</td>
<td>509.52</td>
<td>490.46</td>
</tr>
<tr>
<td>0 none</td>
<td>none</td>
<td>0.5</td>
<td></td>
<td>504.51</td>
<td>429.39</td>
<td>387.58</td>
<td>370.45</td>
</tr>
<tr>
<td>250 25,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>444.01</td>
<td>387.67</td>
<td>346.94</td>
<td>332.65</td>
</tr>
<tr>
<td>500 50,000</td>
<td>none</td>
<td>0.75</td>
<td></td>
<td>424.16</td>
<td>368.72</td>
<td>327.46</td>
<td>314.37</td>
</tr>
</tbody>
</table>
The effect of deductible, by NCD

Individual Loss (Own Damage)

Sum of Losses from a Type (Own Damage)

Sum of Losses from a Specific Event

Overall Loss per Policy
Coverage modifications by level of NCD and age

Now we only use simulation.

As expected, any of a greater deductible, lower policy limit or smaller coinsurance results in a lower predictive mean.

Coinsurance changes the predictive means linearly.

The analysis allows us to see the effects of deductibles and policy limits on long-tail distributions!!!
## Coverage modifications by insured’s age

**Table 6. Simulated Predictive Mean by Insured’s Age and Coverage Modifications**

<table>
<thead>
<tr>
<th>Coverage Modification</th>
<th>Individual Losses (Own Damage)</th>
<th>Sum of Losses from a Type (Own Damage)</th>
<th>Sum of Losses from a specific Event</th>
<th>Overall Loss per Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level of Insured’s Age</td>
<td>≤21</td>
<td>22-25</td>
<td>26-35</td>
</tr>
<tr>
<td>Deductible</td>
<td>Limits</td>
<td>Coinsurance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 none</td>
<td>1</td>
<td>252.87</td>
<td>242.94</td>
<td>191.13</td>
</tr>
<tr>
<td>250 none</td>
<td>1</td>
<td>226.93</td>
<td>219.16</td>
<td>170.54</td>
</tr>
<tr>
<td>500 none</td>
<td>1</td>
<td>204.13</td>
<td>198.39</td>
<td>152.52</td>
</tr>
<tr>
<td>0 25,000</td>
<td>1</td>
<td>246.94</td>
<td>238.24</td>
<td>189.64</td>
</tr>
<tr>
<td>0 50,000</td>
<td>1</td>
<td>250.64</td>
<td>242.62</td>
<td>191.13</td>
</tr>
<tr>
<td>0 none</td>
<td>0.75</td>
<td>189.65</td>
<td>182.21</td>
<td>143.35</td>
</tr>
<tr>
<td>0 none</td>
<td>0.5</td>
<td>126.43</td>
<td>121.47</td>
<td>95.57</td>
</tr>
<tr>
<td>250 25,000</td>
<td>0.75</td>
<td>165.75</td>
<td>160.84</td>
<td>126.79</td>
</tr>
<tr>
<td>500 50,000</td>
<td>0.75</td>
<td>151.42</td>
<td>148.56</td>
<td>114.39</td>
</tr>
<tr>
<td>0 none</td>
<td>1</td>
<td>339.05</td>
<td>314.08</td>
<td>239.04</td>
</tr>
<tr>
<td>250 none</td>
<td>1</td>
<td>308.86</td>
<td>286.80</td>
<td>215.95</td>
</tr>
<tr>
<td>500 none</td>
<td>1</td>
<td>281.82</td>
<td>262.57</td>
<td>195.44</td>
</tr>
<tr>
<td>0 25,000</td>
<td>1</td>
<td>331.01</td>
<td>307.77</td>
<td>236.54</td>
</tr>
<tr>
<td>0 50,000</td>
<td>1</td>
<td>336.33</td>
<td>313.60</td>
<td>238.89</td>
</tr>
<tr>
<td>0 none</td>
<td>0.75</td>
<td>254.29</td>
<td>235.56</td>
<td>179.28</td>
</tr>
<tr>
<td>0 none</td>
<td>0.5</td>
<td>169.53</td>
<td>157.04</td>
<td>119.52</td>
</tr>
<tr>
<td>250 25,000</td>
<td>0.75</td>
<td>225.61</td>
<td>210.37</td>
<td>160.08</td>
</tr>
<tr>
<td>500 50,000</td>
<td>0.75</td>
<td>209.33</td>
<td>196.57</td>
<td>146.47</td>
</tr>
<tr>
<td>0 none</td>
<td>1</td>
<td>480.49</td>
<td>452.84</td>
<td>360.72</td>
</tr>
<tr>
<td>250 none</td>
<td>1</td>
<td>441.68</td>
<td>417.13</td>
<td>329.75</td>
</tr>
<tr>
<td>500 none</td>
<td>1</td>
<td>404.35</td>
<td>382.86</td>
<td>300.06</td>
</tr>
<tr>
<td>0 25,000</td>
<td>1</td>
<td>461.26</td>
<td>434.27</td>
<td>356.68</td>
</tr>
<tr>
<td>0 50,000</td>
<td>1</td>
<td>471.44</td>
<td>444.84</td>
<td>360.30</td>
</tr>
<tr>
<td>0 none</td>
<td>0.75</td>
<td>360.37</td>
<td>339.63</td>
<td>270.54</td>
</tr>
<tr>
<td>0 none</td>
<td>0.5</td>
<td>240.24</td>
<td>226.42</td>
<td>180.36</td>
</tr>
<tr>
<td>250 25,000</td>
<td>0.75</td>
<td>316.83</td>
<td>298.92</td>
<td>244.28</td>
</tr>
<tr>
<td>500 50,000</td>
<td>0.75</td>
<td>296.48</td>
<td>281.14</td>
<td>224.73</td>
</tr>
</tbody>
</table>

E.A. Valdez (U. of Connecticut) Barcelona Summer School, Day 1

16-18 July 2012 30 / 68
The effect of deductible, by insured’s age

Macro-effects inference  A case study
Predictive distribution

- For a single contract, the prob of zero claims is about 7%.
  - This means that the distribution has a large point mass at zero.
  - As with Bernoulli distributions, there has been a tendency to focus on the mean to summarize the distribution
- We consider a portfolio of randomly selected 1,000 policies from our 2001 (held-out) sample
- Wish to predict the distribution of $S = S_1 + \ldots + S_{1000}$
  - The central limit theorem suggests that the mean and variance are good starting points.
  - The distribution of the sum is not approximately normal; this is because (1) the policies are not identical, (2) have discrete and continuous components and (3) have long-tailed continuous components.
  - This is even more evident when we “unbundle” the policy and consider the predictive distribution by type
Figure: Simulated Predictive Distribution for a Randomly Selected Portfolio of 1,000 Policies.
Figure: Simulated Density of Losses for Third Party Injury, Own Damage and Third Party Property of a Randomly Selected Portfolio.
Risk measures

- We consider two measures focusing on the tail of the distribution that have been widely used in both actuarial and financial work.
  - The Value-at-Risk ($VaR$) is simply a quantile or percentile; $Var(\alpha)$ gives the $100(1 - \alpha)$ percentile of the distribution.
  - The Conditional Tail Expectation ($CTE$) is the expected value conditional on exceeding the $Var(\alpha)$.
- Larger deductibles and smaller policy limits decrease the $VaR$ in a nonlinear way.
- Under each combination of deductible and policy limit, the confidence interval becomes wider as the $VaR$ percentile increases.
- Policy limits exert a greater effect than deductibles on the tail of the distribution.
- The policy limit exerts a greater effect than a deductible on the confidence interval capturing the $VaR$. 
### Table 7. $VaR$ by Percentile and Coverage Modification with a Corresponding Confidence Interval

<table>
<thead>
<tr>
<th>Coverage Modification</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deductible Limit</td>
<td>VaR(90%)</td>
<td>VaR(95%)</td>
<td>VaR(99%)</td>
<td>VaR(99%)</td>
<td>VaR(99%)</td>
<td>VaR(99%)</td>
<td>VaR(99%)</td>
<td>VaR(99%)</td>
</tr>
<tr>
<td>0 none</td>
<td>258,644</td>
<td>253,016</td>
<td>264,359</td>
<td>324,611</td>
<td>311,796</td>
<td>341,434</td>
<td>763,042</td>
<td>625,029</td>
</tr>
<tr>
<td>250 none</td>
<td>245,105</td>
<td>239,679</td>
<td>250,991</td>
<td>312,305</td>
<td>298,000</td>
<td>329,689</td>
<td>749,814</td>
<td>612,818</td>
</tr>
<tr>
<td>500 none</td>
<td>233,265</td>
<td>227,363</td>
<td>238,797</td>
<td>301,547</td>
<td>284,813</td>
<td>317,886</td>
<td>737,883</td>
<td>601,448</td>
</tr>
<tr>
<td>1,000 none</td>
<td>210,989</td>
<td>206,251</td>
<td>217,216</td>
<td>281,032</td>
<td>263,939</td>
<td>296,124</td>
<td>716,955</td>
<td>581,867</td>
</tr>
<tr>
<td>0 25,000</td>
<td>206,990</td>
<td>205,134</td>
<td>209,000</td>
<td>222,989</td>
<td>220,372</td>
<td>225,454</td>
<td>253,775</td>
<td>250,045</td>
</tr>
<tr>
<td>0 50,000</td>
<td>224,715</td>
<td>222,862</td>
<td>227,128</td>
<td>245,715</td>
<td>243,107</td>
<td>249,331</td>
<td>286,848</td>
<td>282,736</td>
</tr>
<tr>
<td>0 100,000</td>
<td>244,158</td>
<td>241,753</td>
<td>247,653</td>
<td>272,317</td>
<td>267,652</td>
<td>277,673</td>
<td>336,844</td>
<td>326,873</td>
</tr>
<tr>
<td>250 25,000</td>
<td>193,313</td>
<td>191,364</td>
<td>195,381</td>
<td>208,590</td>
<td>206,092</td>
<td>211,389</td>
<td>239,486</td>
<td>235,754</td>
</tr>
<tr>
<td>500 50,000</td>
<td>199,109</td>
<td>196,603</td>
<td>201,513</td>
<td>219,328</td>
<td>216,395</td>
<td>222,725</td>
<td>259,436</td>
<td>255,931</td>
</tr>
<tr>
<td>1,000 100,000</td>
<td>197,534</td>
<td>194,501</td>
<td>201,685</td>
<td>224,145</td>
<td>220,410</td>
<td>229,925</td>
<td>287,555</td>
<td>278,601</td>
</tr>
</tbody>
</table>
Table 8. *CTE* by Percentile and Coverage Modification with a Corresponding Standard Deviation

<table>
<thead>
<tr>
<th>Coverage Modification Limit</th>
<th>CTE(90%)</th>
<th>Standard Deviation</th>
<th>CTE(95%)</th>
<th>Standard Deviation</th>
<th>CTE(99%)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 none</td>
<td>468,850</td>
<td>22,166</td>
<td>652,821</td>
<td>41,182</td>
<td>1,537,692</td>
<td>149,371</td>
</tr>
<tr>
<td>250 none</td>
<td>455,700</td>
<td>22,170</td>
<td>639,762</td>
<td>41,188</td>
<td>1,524,650</td>
<td>149,398</td>
</tr>
<tr>
<td>500 none</td>
<td>443,634</td>
<td>22,173</td>
<td>627,782</td>
<td>41,191</td>
<td>1,512,635</td>
<td>149,417</td>
</tr>
<tr>
<td>1,000 none</td>
<td>422,587</td>
<td>22,180</td>
<td>606,902</td>
<td>41,200</td>
<td>1,491,767</td>
<td>149,457</td>
</tr>
<tr>
<td>0 25,000</td>
<td>228,169</td>
<td>808</td>
<td>242,130</td>
<td>983</td>
<td>266,428</td>
<td>1,787</td>
</tr>
<tr>
<td>0 50,000</td>
<td>252,564</td>
<td>1,082</td>
<td>270,589</td>
<td>1,388</td>
<td>304,941</td>
<td>2,762</td>
</tr>
<tr>
<td>0 100,000</td>
<td>283,270</td>
<td>1,597</td>
<td>309,661</td>
<td>2,091</td>
<td>364,183</td>
<td>3,332</td>
</tr>
<tr>
<td>250 25,000</td>
<td>213,974</td>
<td>797</td>
<td>227,742</td>
<td>973</td>
<td>251,820</td>
<td>1,796</td>
</tr>
<tr>
<td>500 50,000</td>
<td>225,937</td>
<td>1,066</td>
<td>243,608</td>
<td>1,378</td>
<td>277,883</td>
<td>2,701</td>
</tr>
<tr>
<td>1,000 100,000</td>
<td>235,678</td>
<td>1,562</td>
<td>261,431</td>
<td>2,055</td>
<td>315,229</td>
<td>3,239</td>
</tr>
</tbody>
</table>
Unbundling of coverages

- Decompose the comprehensive coverage into more “primitive” coverages: third party injury, own damage and third party property.
- Calculate a risk measure for each unbundled coverage, as if separate financial institutions owned each coverage.
- Compare to the bundled coverage that the insurance company is responsible for.
- Despite positive dependence, there are still economies of scale.

Table 9. \( VaR \) and \( CTE \) by Percentile for Unbundled and Bundled Coverages

<table>
<thead>
<tr>
<th></th>
<th>( VaR )</th>
<th></th>
<th>( VaR )</th>
<th></th>
<th>( CTE )</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
</tr>
<tr>
<td>Third party injury</td>
<td>161,476</td>
<td>309,881</td>
<td>1,163,855</td>
<td>592,343</td>
<td>964,394</td>
<td>2,657,911</td>
</tr>
<tr>
<td>Own damage</td>
<td>49,648</td>
<td>59,898</td>
<td>86,421</td>
<td>65,560</td>
<td>76,951</td>
<td>104,576</td>
</tr>
<tr>
<td>Third party property</td>
<td>188,797</td>
<td>209,509</td>
<td>264,898</td>
<td>223,524</td>
<td>248,793</td>
<td>324,262</td>
</tr>
<tr>
<td>Sum of Unbundled Coverages</td>
<td>399,921</td>
<td>579,288</td>
<td>1,515,174</td>
<td>881,427</td>
<td>1,290,137</td>
<td>3,086,749</td>
</tr>
<tr>
<td>Bundled (Comprehensive) Coverage</td>
<td>258,644</td>
<td>324,611</td>
<td>763,042</td>
<td>468,850</td>
<td>652,821</td>
<td>1,537,692</td>
</tr>
</tbody>
</table>
How important is the copula?

Very!!

<table>
<thead>
<tr>
<th>Table 10. $VaR$ and $CTE$ for Bundled Coverage by Copula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copula</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Independence</td>
</tr>
<tr>
<td>Normal</td>
</tr>
<tr>
<td>$t$</td>
</tr>
</tbody>
</table>

Effects of Re-Estimating the Full Model

Effects of Changing Only the Dependence Structure
Quota share reinsurance

- A fixed percentage of each policy written will be transferred to the reinsurer
- Does not change the shape of the retained losses, only the location and scale
- Distribution of Retained Claims for the Insurer under Quota Share Reinsurance. The insurer retains 25%, 50%, 75% and 100% of losses, respectively.
Figure: Distribution of Losses for the Insurer and Reinsurer under Excess of Loss Reinsurance. The losses are simulated under different primary company retention limits. The left-hand panel is for the insurer and right-hand panel is for the reinsurer.
### Table 11. Percentiles of Losses for Insurer and Reinsurer under Reinsurance Agreement

#### Percentile for Insurer

<table>
<thead>
<tr>
<th>Quota</th>
<th>Policy Retention</th>
<th>Portfolio Retention</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>none</td>
<td>100,000</td>
<td>22,518</td>
<td>26,598</td>
<td>29,093</td>
<td>34,196</td>
<td>40,943</td>
<td>50,657</td>
<td>64,819</td>
<td>83,500</td>
<td>100,000</td>
</tr>
<tr>
<td>0.5</td>
<td>none</td>
<td>100,000</td>
<td>45,036</td>
<td>53,197</td>
<td>58,187</td>
<td>68,393</td>
<td>81,885</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
</tr>
<tr>
<td>0.75</td>
<td>none</td>
<td>100,000</td>
<td>67,553</td>
<td>79,795</td>
<td>87,280</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>100,000</td>
<td>86,083</td>
<td>99,747</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>200,000</td>
<td>86,083</td>
<td>99,747</td>
<td>108,345</td>
<td>122,927</td>
<td>140,910</td>
<td>159,449</td>
<td>177,013</td>
<td>188,813</td>
<td>200,000</td>
</tr>
<tr>
<td>1</td>
<td>20,000</td>
<td>200,000</td>
<td>89,605</td>
<td>105,578</td>
<td>114,512</td>
<td>132,145</td>
<td>154,858</td>
<td>177,985</td>
<td>200,000</td>
<td>200,000</td>
<td>200,000</td>
</tr>
<tr>
<td>0.25</td>
<td>10,000</td>
<td>100,000</td>
<td>21,521</td>
<td>24,937</td>
<td>27,086</td>
<td>30,732</td>
<td>35,228</td>
<td>39,862</td>
<td>44,253</td>
<td>47,203</td>
<td>53,352</td>
</tr>
<tr>
<td>0.5</td>
<td>20,000</td>
<td>100,000</td>
<td>44,803</td>
<td>52,789</td>
<td>57,256</td>
<td>66,072</td>
<td>77,429</td>
<td>88,993</td>
<td>100,000</td>
<td>100,000</td>
<td>100,000</td>
</tr>
<tr>
<td>0.75</td>
<td>10,000</td>
<td>200,000</td>
<td>64,562</td>
<td>74,810</td>
<td>81,259</td>
<td>92,195</td>
<td>105,683</td>
<td>119,586</td>
<td>132,760</td>
<td>141,610</td>
<td>160,056</td>
</tr>
<tr>
<td>1</td>
<td>20,000</td>
<td>200,000</td>
<td>89,605</td>
<td>105,578</td>
<td>114,512</td>
<td>132,145</td>
<td>154,858</td>
<td>177,985</td>
<td>200,000</td>
<td>200,000</td>
<td>200,000</td>
</tr>
</tbody>
</table>

#### Percentile for Reinsurer

<table>
<thead>
<tr>
<th>Quota</th>
<th>Policy Retention</th>
<th>Portfolio Retention</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>none</td>
<td>100,000</td>
<td>67,553</td>
<td>79,795</td>
<td>87,280</td>
<td>102,589</td>
<td>122,828</td>
<td>151,972</td>
<td>194,458</td>
<td>250,499</td>
<td>486,743</td>
</tr>
<tr>
<td>0.5</td>
<td>none</td>
<td>100,000</td>
<td>45,036</td>
<td>53,197</td>
<td>58,187</td>
<td>68,393</td>
<td>81,885</td>
<td>102,630</td>
<td>159,277</td>
<td>233,998</td>
<td>486,743</td>
</tr>
<tr>
<td>0.75</td>
<td>none</td>
<td>100,000</td>
<td>22,518</td>
<td>26,598</td>
<td>29,093</td>
<td>36,785</td>
<td>63,771</td>
<td>102,630</td>
<td>159,277</td>
<td>233,998</td>
<td>486,743</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>100,000</td>
<td>8,066</td>
<td>16,747</td>
<td>36,888</td>
<td>63,781</td>
<td>102,630</td>
<td>159,277</td>
<td>233,998</td>
<td>486,743</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>200,000</td>
<td>0</td>
<td>0</td>
<td>992</td>
<td>5,878</td>
<td>18,060</td>
<td>43,434</td>
<td>97,587</td>
<td>171,377</td>
<td>426,367</td>
</tr>
<tr>
<td>1</td>
<td>20,000</td>
<td>200,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2,482</td>
<td>24,199</td>
<td>78,839</td>
<td>151,321</td>
<td>412,817</td>
</tr>
<tr>
<td>0.25</td>
<td>10,000</td>
<td>100,000</td>
<td>68,075</td>
<td>80,695</td>
<td>88,555</td>
<td>104,557</td>
<td>127,652</td>
<td>161,743</td>
<td>215,407</td>
<td>292,216</td>
<td>541,818</td>
</tr>
<tr>
<td>0.5</td>
<td>20,000</td>
<td>100,000</td>
<td>45,132</td>
<td>53,298</td>
<td>58,383</td>
<td>68,909</td>
<td>84,474</td>
<td>111,269</td>
<td>167,106</td>
<td>245,101</td>
<td>491,501</td>
</tr>
<tr>
<td>0.75</td>
<td>10,000</td>
<td>200,000</td>
<td>23,536</td>
<td>28,055</td>
<td>31,434</td>
<td>39,746</td>
<td>54,268</td>
<td>81,443</td>
<td>135,853</td>
<td>209,406</td>
<td>462,321</td>
</tr>
<tr>
<td>1</td>
<td>20,000</td>
<td>200,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2,482</td>
<td>24,199</td>
<td>78,839</td>
<td>151,321</td>
<td>412,817</td>
</tr>
</tbody>
</table>
Concluding remarks

- Model features
  - Allows for covariates for the frequency, type and severity components
  - Captures the long-tail nature of severity through the GB2.
  - Provides for a “two-part” distribution of losses - when a claim occurs, not necessary that all possible types of losses are realized.
  - Allows for possible dependencies among claims through a copula
  - Allows for heterogeneity from the longitudinal nature of policyholders (not claims)

- Other applications
  - Could look at financial information from companies
  - Could examine health care expenditure
  - Compare companies’ performance using multilevel, intercompany experience
Micro-level data

- This paper shows how to use micro-level data to make sensible statements about “macro-effects.”
  - For example, the effect of a policy level deductible on the distribution of a block of business.

- Certainly not the first to support this viewpoint
  - Traditional actuarial approach is to development life insurance company policy reserves on a policy-by-policy basis.
  - See, for example, Richard Derrig and Herbert I Weisberg (1993) “Pricing auto no-fault and bodily injury coverages using micro-data and statistical models”

- However, the idea of using voluminous data that the insurance industry captures for making managerial decisions is becoming more prominent.
  - Gourieroux and Jasiak (2007) have dubbed this emerging field the “microeconometrics of individual risk.”
  - See recent ARIA news article by Ellingsworth from ISO

- Academics need greater access to micro-level data!!
The fitted frequency model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>-2.275</td>
<td>0.730</td>
</tr>
<tr>
<td>year</td>
<td>0.043</td>
<td>0.004</td>
</tr>
<tr>
<td>automobile</td>
<td>-1.635</td>
<td>0.082</td>
</tr>
<tr>
<td>vehicle age 0</td>
<td>0.273</td>
<td>0.739</td>
</tr>
<tr>
<td>vehicle age 1-2</td>
<td>0.670</td>
<td>0.732</td>
</tr>
<tr>
<td>vehicle age 3-5</td>
<td>0.482</td>
<td>0.732</td>
</tr>
<tr>
<td>vehicle age 6-10</td>
<td>0.223</td>
<td>0.732</td>
</tr>
<tr>
<td>vehicle age 11-15</td>
<td>0.084</td>
<td>0.772</td>
</tr>
<tr>
<td>automobile*vehicle age 0</td>
<td>0.613</td>
<td>0.167</td>
</tr>
<tr>
<td>automobile*vehicle age 1-2</td>
<td>0.258</td>
<td>0.139</td>
</tr>
<tr>
<td>automobile*vehicle age 3-5</td>
<td>0.386</td>
<td>0.138</td>
</tr>
<tr>
<td>automobile*vehicle age 6-10</td>
<td>0.608</td>
<td>0.138</td>
</tr>
<tr>
<td>automobile*vehicle age 11-15</td>
<td>0.569</td>
<td>0.265</td>
</tr>
<tr>
<td>automobile*vehicle age ≥16</td>
<td>0.930</td>
<td>0.677</td>
</tr>
<tr>
<td>vehicle capacity</td>
<td>0.116</td>
<td>0.018</td>
</tr>
<tr>
<td>automobile*NCD 0</td>
<td>0.748</td>
<td>0.027</td>
</tr>
<tr>
<td>automobile*NCD 10</td>
<td>0.640</td>
<td>0.032</td>
</tr>
<tr>
<td>automobile*NCD 20</td>
<td>0.585</td>
<td>0.029</td>
</tr>
<tr>
<td>automobile*NCD 30</td>
<td>0.563</td>
<td>0.030</td>
</tr>
<tr>
<td>automobile*NCD 40</td>
<td>0.482</td>
<td>0.032</td>
</tr>
<tr>
<td>automobile*NCD 50</td>
<td>0.347</td>
<td>0.021</td>
</tr>
<tr>
<td>automobile*age ≤21</td>
<td>0.955</td>
<td>0.431</td>
</tr>
<tr>
<td>automobile*age 22-25</td>
<td>0.843</td>
<td>0.105</td>
</tr>
<tr>
<td>automobile*age 26-35</td>
<td>0.657</td>
<td>0.070</td>
</tr>
<tr>
<td>automobile*age 36-45</td>
<td>0.546</td>
<td>0.070</td>
</tr>
<tr>
<td>automobile*age 46-55</td>
<td>0.497</td>
<td>0.071</td>
</tr>
<tr>
<td>automobile*age 56-65</td>
<td>0.427</td>
<td>0.073</td>
</tr>
<tr>
<td>automobile*age ≥66</td>
<td>0.438</td>
<td>0.087</td>
</tr>
<tr>
<td>automobile*male</td>
<td>-0.252</td>
<td>0.042</td>
</tr>
<tr>
<td>automobile*female</td>
<td>-0.383</td>
<td>0.043</td>
</tr>
<tr>
<td>$r$</td>
<td>2.167</td>
<td>0.195</td>
</tr>
</tbody>
</table>
The fitted conditional claim type model

### Table A.2. Fitted Multi Logit Model

<table>
<thead>
<tr>
<th>Category($M$)</th>
<th>intercept</th>
<th>year</th>
<th>vehicle</th>
<th>age $\geq 6$</th>
<th>non-automobile</th>
<th>automobile*age $\geq 46$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.194</td>
<td>-0.142</td>
<td>0.084</td>
<td>0.262</td>
<td></td>
<td>0.128</td>
</tr>
<tr>
<td>2</td>
<td>4.707</td>
<td>-0.024</td>
<td>-0.024</td>
<td>-0.153</td>
<td></td>
<td>0.082</td>
</tr>
<tr>
<td>3</td>
<td>3.281</td>
<td>-0.036</td>
<td>0.252</td>
<td>0.716</td>
<td></td>
<td>-0.201</td>
</tr>
<tr>
<td>4</td>
<td>1.052</td>
<td>-0.129</td>
<td>0.037</td>
<td>-0.349</td>
<td></td>
<td>0.338</td>
</tr>
<tr>
<td>5</td>
<td>-1.628</td>
<td>0.132</td>
<td>0.132</td>
<td>-0.008</td>
<td></td>
<td>0.330</td>
</tr>
<tr>
<td>6</td>
<td>3.551</td>
<td>-0.089</td>
<td>0.032</td>
<td>-0.259</td>
<td></td>
<td>0.203</td>
</tr>
</tbody>
</table>
The fitted conditional severity model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Independence</th>
<th>Types of Copula</th>
<th>t-Copula</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Normal Copula</td>
<td>Estimate</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>Standard Error</td>
<td>Standard Error</td>
</tr>
<tr>
<td>Third Party Injury</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.225</td>
<td>0.224</td>
<td>0.232</td>
</tr>
<tr>
<td></td>
<td>0.020</td>
<td>0.044</td>
<td>0.079</td>
</tr>
<tr>
<td>$\alpha_{11}$</td>
<td>69.958</td>
<td>69.944</td>
<td>69.772</td>
</tr>
<tr>
<td></td>
<td>28.772</td>
<td>63.267</td>
<td>105.245</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>392.362</td>
<td>392.372</td>
<td>392.496</td>
</tr>
<tr>
<td></td>
<td>145.055</td>
<td>129.664</td>
<td>204.730</td>
</tr>
<tr>
<td>intercept</td>
<td>34.269</td>
<td>34.094</td>
<td>31.915</td>
</tr>
<tr>
<td></td>
<td>8.144</td>
<td>7.883</td>
<td>5.606</td>
</tr>
<tr>
<td>Own Damage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.671</td>
<td>0.670</td>
<td>0.660</td>
</tr>
<tr>
<td></td>
<td>0.007</td>
<td>0.002</td>
<td>0.004</td>
</tr>
<tr>
<td>$\alpha_{12}$</td>
<td>5.570</td>
<td>5.541</td>
<td>5.758</td>
</tr>
<tr>
<td></td>
<td>0.151</td>
<td>0.144</td>
<td>0.103</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>12.383</td>
<td>12.555</td>
<td>13.933</td>
</tr>
<tr>
<td></td>
<td>0.628</td>
<td>0.277</td>
<td>0.750</td>
</tr>
<tr>
<td>intercept</td>
<td>1.987</td>
<td>2.005</td>
<td>2.183</td>
</tr>
<tr>
<td></td>
<td>0.115</td>
<td>0.094</td>
<td>0.112</td>
</tr>
<tr>
<td>year</td>
<td>-0.016</td>
<td>-0.015</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>vehicle capacity</td>
<td>0.116</td>
<td>0.129</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>0.031</td>
<td>0.022</td>
<td>0.012</td>
</tr>
<tr>
<td>vehicle age $\leq 5$</td>
<td>0.107</td>
<td>0.106</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>0.034</td>
<td>0.031</td>
<td>0.003</td>
</tr>
<tr>
<td>automobile*NCD 0-10</td>
<td>0.102</td>
<td>0.099</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>0.029</td>
<td>0.039</td>
<td>0.031</td>
</tr>
<tr>
<td>automobile*age 26-55</td>
<td>-0.047</td>
<td>-0.042</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td>0.027</td>
<td>0.044</td>
<td>0.005</td>
</tr>
<tr>
<td>automobile*age $\geq 56$</td>
<td>0.101</td>
<td>0.080</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>0.050</td>
<td>0.018</td>
<td>0.050</td>
</tr>
<tr>
<td>Third Party Property</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>1.320</td>
<td>1.309</td>
<td>1.349</td>
</tr>
<tr>
<td></td>
<td>0.068</td>
<td>0.066</td>
<td>0.068</td>
</tr>
<tr>
<td>$\alpha_{13}$</td>
<td>0.677</td>
<td>0.615</td>
<td>0.617</td>
</tr>
<tr>
<td></td>
<td>0.088</td>
<td>0.080</td>
<td>0.079</td>
</tr>
<tr>
<td>$\alpha_{23}$</td>
<td>1.383</td>
<td>1.528</td>
<td>1.324</td>
</tr>
<tr>
<td></td>
<td>0.253</td>
<td>0.271</td>
<td>0.217</td>
</tr>
<tr>
<td>intercept</td>
<td>1.071</td>
<td>1.035</td>
<td>0.841</td>
</tr>
<tr>
<td></td>
<td>0.134</td>
<td>0.132</td>
<td>0.120</td>
</tr>
<tr>
<td>vehicle age 1-10</td>
<td>-0.008</td>
<td>-0.054</td>
<td>-0.036</td>
</tr>
<tr>
<td></td>
<td>0.098</td>
<td>0.094</td>
<td>0.092</td>
</tr>
<tr>
<td>vehicle age $\geq 11$</td>
<td>-0.022</td>
<td>0.030</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>0.198</td>
<td>0.194</td>
<td>0.193</td>
</tr>
<tr>
<td>year</td>
<td>0.031</td>
<td>0.043</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Copula</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>-</td>
<td>-</td>
<td>0.250</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.241</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.054</td>
</tr>
<tr>
<td>$\rho_{13}$</td>
<td>-</td>
<td>-</td>
<td>0.163</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.169</td>
</tr>
<tr>
<td>$\rho_{23}$</td>
<td>-</td>
<td>-</td>
<td>0.310</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.330</td>
</tr>
<tr>
<td>$\nu$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.688</td>
</tr>
</tbody>
</table>
A bit about Singapore
A bit about Singapore

- Singa Pura: Lion city. Location: 136.8 km N of equator, between latitudes 103 deg 38’ E and 104 deg 06’ E. [islands between Malaysia and Indonesia]
- Size: very tiny [647.5 sq km, of which 10 sq km is water] Climate: very hot and humid [23-30 deg celsius]
- Population: nearly 5 mn. Age structure: 0-14 yrs: 16%, 15-64 yrs: 76%, 65+ yrs 8%
- Birth rate: 9.34 births/1,000. Death rate: 4.28 deaths/1,000; Life expectancy: 81 yrs; male: 79 yrs; female: 83 yrs
- Ethnic groups: Chinese 74%, Malay 13%, Indian 9%; Languages: Chinese, Malay, Tamil, English

---

1Updated: February 2010
Insurance market in Singapore

- As of 2009: market consists of 45 general ins, 8 life ins, 7 both, 17 general reinsurers, 2 life reins, 7 both; also the largest captive domicile in Asia, with 59 registered captives.
- Monetary Authority of Singapore (MAS) is the supervisory/regulatory body; also assists to promote Singapore as an international financial center.
- Insurance industry performance in 2009:
  - total premiums: 11.4 bn; total assets: 113.3 bn [20% annual growth]
  - life insurance: annual premium = 251.6 mn; single premium = 759.5 mn
  - general insurance: gross premium = 1.9 bn (domestic = 0.9; offshore = 1.0)
- Further information: http://www.mas.gov.sg

---

²Source: wikipedia
A multilevel analysis of intercompany claim counts
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 ∈ classes ∈ schools ∈ 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
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 ($c$).
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*

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

#### Table 3: Vehicle level explanatory variables

<table>
<thead>
<tr>
<th>Categorical Covariate</th>
<th>Description</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Type</td>
<td>Car</td>
<td>54%</td>
</tr>
<tr>
<td></td>
<td>Motor</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td>Truck</td>
<td>5%</td>
</tr>
<tr>
<td>Private Use</td>
<td>Vehicle is used for private purposes</td>
<td>31%</td>
</tr>
<tr>
<td></td>
<td>Vehicle is used for other than private purposes</td>
<td>69%</td>
</tr>
<tr>
<td>NCD</td>
<td>'No Claims Discount' at entry in fleet: based on previous accident record of policyholder. The higher the discount, the better the prior accident record.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NCD = 0</td>
<td>83%</td>
</tr>
<tr>
<td></td>
<td>NCD &gt; 0</td>
<td>17%</td>
</tr>
<tr>
<td>SwitchPol</td>
<td>1 if vehicle changes fleet</td>
<td>55%</td>
</tr>
<tr>
<td></td>
<td>0 if vehicle enters fleet for first time or stays in the same fleet</td>
<td>45%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuous Covariate</th>
<th>Minimum</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Age</td>
<td>0</td>
<td>4.22</td>
<td>33</td>
</tr>
<tr>
<td>Cubic Capacity</td>
<td>124</td>
<td>1,615</td>
<td>6,750</td>
</tr>
<tr>
<td>Tonnage</td>
<td>1</td>
<td>7.6</td>
<td>61</td>
</tr>
<tr>
<td>TLengthEntry</td>
<td>0</td>
<td>0.35</td>
<td>6.75</td>
</tr>
<tr>
<td>TLength</td>
<td>0.006</td>
<td>0.78</td>
<td>1</td>
</tr>
</tbody>
</table>

E.A. Valdez (U. of Connecticut) Barcelona Summer School, Day 1 16-18 July 2012 58 / 68
Fleet and company level covariates

Table 4: *Fleet and company level explanatory variables*

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

- **(Poisson)** $\Pr_{\text{Poi}}(Y = y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}$;

- **(Negative binomial)** $\Pr_{\text{NB}}(Y = y|\mu, \tau) = \frac{\Gamma(y+\tau)}{y!\Gamma(\tau)} \left( \frac{\tau}{\mu+\tau} \right)^\tau \left( \frac{\mu}{\mu+\tau} \right)^y$;

- **(Zero–inflated Poisson)**

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

- **(Hurdle Poisson)**

  $\Pr_{\text{Hur}}(Y = 0|p, \lambda) = p \quad y = 0,$
  
  $\Pr_{\text{Hur}}(Y = y|p, \lambda) = \frac{1 - p}{1 - \Pr_{\text{Poi}}(0|\lambda)} \Pr_{\text{Poi}}(Y = y|\lambda) \quad y > 0.$
## Fitted models without covariates

<table>
<thead>
<tr>
<th>Num. Claims</th>
<th>Obs. Freq.</th>
<th>Poisson</th>
<th>NB</th>
<th>ZIP</th>
<th>Hurdle Poi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>34,357</td>
<td>33,940</td>
<td>34,362</td>
<td>34,357</td>
<td>34,357</td>
</tr>
<tr>
<td>1</td>
<td>4,104</td>
<td>4,821</td>
<td>4,079</td>
<td>4,048</td>
<td>4,048</td>
</tr>
<tr>
<td>2</td>
<td>551</td>
<td>342</td>
<td>577</td>
<td>641</td>
<td>641</td>
</tr>
<tr>
<td>3</td>
<td>86</td>
<td>16</td>
<td>86</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>1</td>
<td>13</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>≥ 5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

|             | Mean       | 0.142   | 0.142  | 0.142   | 0.142     |
|             | Variance   | 0.171   | 0.142  | 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 specification 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.

\[ Y_{c,f,v,t} \sim ZIP(p, \lambda_{c,f,v,t}) \]

where \( \lambda_{c,f,v,t} = e_{c,f,v,t} \exp(\eta_{c,f,v,t} + \epsilon_c + \epsilon_{c,f}) \)

and \( \eta_{c,f,v,t} := \gamma + X_{c} \beta_4 + X_{c,f} \beta_3 + X_{c,f,v} \beta_2 + X_{c,f,v,t} \beta_1 \)

Here \( \gamma \) is the intercept, \( \epsilon_c \) 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 \( X \)'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*

<table>
<thead>
<tr>
<th>Num. Claims</th>
<th>Obs. Freq.</th>
<th>Poisson (3)</th>
<th>NB (7)</th>
<th>ZIP (8)</th>
<th>Hurdle Poi (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>34,357</td>
<td>34,310</td>
<td>34,365</td>
<td>34,350</td>
<td>34,360</td>
</tr>
<tr>
<td></td>
<td>(34,200;34,430)</td>
<td>(34,240;34,490)</td>
<td>(34,230;34,470)</td>
<td>(34,230;34,480)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4,104</td>
<td>4,176</td>
<td>4,086</td>
<td>4,092</td>
<td>4,139</td>
</tr>
<tr>
<td></td>
<td>(4,081;4,273)</td>
<td>(3,978;4,196)</td>
<td>(3,979;4,207)</td>
<td>(4,025;4,253)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>551</td>
<td>536</td>
<td>560</td>
<td>584</td>
<td>540</td>
</tr>
<tr>
<td></td>
<td>(511;560)</td>
<td>(532,588)</td>
<td>(551;618)</td>
<td>(505;576)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>86</td>
<td>79</td>
<td>88</td>
<td>79</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>(71,87)</td>
<td>(78,99)</td>
<td>(70,89)</td>
<td>(64,82)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>14</td>
<td>16</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>(11,16)</td>
<td>(13,20)</td>
<td>(9,13)</td>
<td>(8,13)</td>
<td></td>
</tr>
<tr>
<td>≥ 5</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(2,4)</td>
<td>(2,4.5)</td>
<td>(1,2)</td>
<td>(1,2.4)</td>
<td></td>
</tr>
</tbody>
</table>
Figure: 95% credibility intervals for the hierarchical ZIP model
### Table 11: Results for the ZIP model

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

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


