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

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

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

*University of Wisconsin – Madison
** University of Amsterdam

13th International Congress on Insurance: Mathematics and Economics
Istanbul, Turkey 27-29 May 2009
Outline

1 Introduction
2 Model estimation
   - Data
   - Models of each component
3 Macro-effects inference
   - Individual risk rating
   - A case study
   - Predictive distributions for portfolios
4 Conclusion
5 Appendix A - Parameter Estimates
6 Appendix B - Singapore
Basic data set-up

- “Policyholder” $i$ is followed over time $t = 1, \ldots, 9$ years
- Unit of analysis “$it$”
- 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 (including GLM) - 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>Percentage</td>
</tr>
</tbody>
</table>
The hierarchical insurance claims model

- Traditional to predict/estimate insurance claims distributions:
  \[
  \text{Cost of Claims} = \text{Frequency} \times \text{Severity}
  \]

- 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.
Papers


- Antonio, Frees and Valdez (2009), A Hierarchical Model for Micro-Level Stochastic Loss Reserving, also being presented separately at this conference.
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

- Large literature on modeling claims frequency and severity:
  - 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.

- 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 use for out-of-sample prediction
- Information was extracted from the policy, claims and payment 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: automobile (A) or others (O).
- Vehicle Age: in years, grouped into 6 categories -
  - 0, 1-2, 3-5, 6-10, 11-15, $\geq 16$.
- Vehicle Capacity: in cubic capacity.
- Gender: male (M) or female (F).
- Age: in years, grouped into 7 categories -
  - ages $\leq 21$, 22-25, 26-35, 36-45, 46-55, 56-65, $\geq 66$.
- 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 $r$:
  - $\Pr(N = k | r, p) = \binom{k + r - 1}{r - 1} p^r (1 - p)^k$.
  - Here, $\sigma = r^{-1}$ is the dispersion parameter and
  - $p = p_{it}$ is related to the mean through
    $$(1 - p_{it})/p_{it} = \lambda_{it} \sigma = e_{it} \exp(x'_{\lambda, it} \beta \lambda) \sigma.$$
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).

- Alternative to model claim type was considered in:
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 \( z = (\ln y - \mu)/\sigma \), with location \( \mu \), scale \( \sigma \), and shape parameters \( \alpha_1 \) and \( \alpha_2 \).
- With four parameters, the 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

Mean and higher moments always exist
Mean and higher moments never exist

Mode > 0

Two parameters
Lognormal
Gamma

Three parameters
Transformed gamma
Weibull

Four parameters
Transformed beta
Burr

Mode = 0

Inverse gamma
Inverse transformed gamma
Inverse Weibull
Inverse Pareto
Inverse Burr
Loglogistic

Special case
Limiting case (parameters approach zero or infinity)

Fig. 4.7 Distributional relationships and characteristics.

Source: Klugman, Panjer and Willmot (2004), p. 72
Figure: GB2 density for varying parameters

- σ = 10
- σ = 5
- σ = -5
- σ = -10

- μ = 0
- μ = log(2)
- μ = log(3)
- μ = log(4)

- α₁ = 0.5
- α₁ = 1
- α₁ = 5
- α₁ = 10

- α₂ = 2
- α₂ = 1.5
- α₂ = 1
- α₂ = 0.5
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, $\beta_{k,j} = \partial \ln \mathbb{E} \left( Y \mid x \right) / \partial x_j$
  - 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, \( S_{EVENT,i,j} = y_{i,j,1} + y_{i,j,2} + y_{i,j,3} \)
    - 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 $\mu_{ik} = \mathbb{E}(y_{ijk}|N_i, K_i = k)$ 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:

$$
\mathbb{E}(y_{ijk}) = \Pr(N_i = 1) \Pr(K_i = k)\mu_{ik},
$$

$$
\mathbb{E}(S_{i,k}) = \mu_{ik} \Pr(K_i = k) \sum_{n=1}^{\infty} n \Pr(N_i = n),
$$

$$
\mathbb{E}(S_{EVENT,i,j}) = \Pr(N_i = 1) \sum_{k=1}^{3} \mu_{ik} \Pr(K_i = k), \text{ and}
$$

$$
\mathbb{E}(S_i) = \mathbb{E}(S_{i,1}) + \mathbb{E}(S_{i,2}) + \mathbb{E}(S_{i,3}).
$$

- 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.
### 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>
Predictive means and confidence intervals
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

Frees, Shi, Antonio & Valdez (WI/CT/Ams) Using Micro-Level Automobile Data IME, 27-29 May 2009 26 / 44
The effect of deductible, by insured’s age

- **Individual Loss (Own Damage)**
- **Sum of Losses from a Type (Own Damage)**
- **Sum of Losses from a Specific Event**
- **Overall Loss per Policy**
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; $\text{VaR}(\alpha)$ gives the $100(1 - \alpha)$ percentile of the distribution.
  - The Conditional Tail Expectation (CTE) is the expected value conditional on exceeding the $\text{VaR}(\alpha)$.

- 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>
<thead>
<tr>
<th>Coverage Modification Limit</th>
<th>VaR(90%) Lower Bound</th>
<th>VaR(90%) Upper Bound</th>
<th>VaR(95%) Lower Bound</th>
<th>VaR(95%) Upper Bound</th>
<th>VaR(99%) Lower Bound</th>
<th>VaR(99%) Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 none</td>
<td>258,644</td>
<td>253,016</td>
<td>264,359</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250 none</td>
<td>245,105</td>
<td>239,679</td>
<td>250,991</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 none</td>
<td>233,265</td>
<td>227,363</td>
<td>238,797</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,000 none</td>
<td>210,989</td>
<td>206,251</td>
<td>217,216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 25,000</td>
<td>206,990</td>
<td>205,134</td>
<td>209,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 50,000</td>
<td>224,715</td>
<td>222,862</td>
<td>227,128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 100,000</td>
<td>244,158</td>
<td>241,753</td>
<td>247,653</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250 25,000</td>
<td>193,313</td>
<td>191,364</td>
<td>195,381</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 50,000</td>
<td>199,109</td>
<td>196,603</td>
<td>201,513</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,000 100,000</td>
<td>197,534</td>
<td>194,501</td>
<td>201,685</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 8. CTE by Percentile and Coverage Modification with a Corresponding Standard Deviation

<table>
<thead>
<tr>
<th>Coverage Modification Deductible Limit</th>
<th>CTE(90%) Standard Deviation</th>
<th>CTE(95%) Standard Deviation</th>
<th>CTE(99%) Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 none</td>
<td>468,850 22,166</td>
<td>652,821 41,182</td>
<td>1,537,692 149,371</td>
</tr>
<tr>
<td>250 none</td>
<td>455,700 22,170</td>
<td>639,762 41,188</td>
<td>1,524,650 149,398</td>
</tr>
<tr>
<td>500 none</td>
<td>443,634 22,173</td>
<td>627,782 41,191</td>
<td>1,512,635 149,417</td>
</tr>
<tr>
<td>1,000 none</td>
<td>422,587 22,180</td>
<td>606,902 41,200</td>
<td>1,491,767 149,457</td>
</tr>
<tr>
<td>0 25,000</td>
<td>228,169 808</td>
<td>242,130 983</td>
<td>266,428 1,787</td>
</tr>
<tr>
<td>0 50,000</td>
<td>252,564 1,082</td>
<td>270,589 1,388</td>
<td>304,941 2,762</td>
</tr>
<tr>
<td>0 100,000</td>
<td>283,270 1,597</td>
<td>309,661 2,091</td>
<td>364,183 3,332</td>
</tr>
<tr>
<td>250 25,000</td>
<td>213,974 797</td>
<td>227,742 973</td>
<td>251,820 1,796</td>
</tr>
<tr>
<td>500 50,000</td>
<td>225,937 1,066</td>
<td>243,608 1,378</td>
<td>277,883 2,701</td>
</tr>
<tr>
<td>1,000 100,000</td>
<td>235,678 1,562</td>
<td>261,431 2,055</td>
<td>315,229 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>
<thead>
<tr>
<th>Unbundled Coverages</th>
<th>VaR 90%</th>
<th>VaR 95%</th>
<th>VaR 99%</th>
<th>CTE 90%</th>
<th>CTE 95%</th>
<th>CTE 99%</th>
</tr>
</thead>
<tbody>
<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>Copula</th>
<th>90% VaR</th>
<th>95% VaR</th>
<th>99% VaR</th>
<th>90% CTE</th>
<th>95% CTE</th>
<th>99% CTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>359,937</td>
<td>490,541</td>
<td>1,377,053</td>
<td>778,744</td>
<td>1,146,709</td>
<td>2,838,762</td>
</tr>
<tr>
<td>Normal</td>
<td>282,040</td>
<td>396,463</td>
<td>988,528</td>
<td>639,140</td>
<td>948,404</td>
<td>2,474,151</td>
</tr>
<tr>
<td>t</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>

**Table 10. VaR and CTE for Bundled Coverage by Copula**

**Effects of Re-Estimating the Full Model**

**Effects of Changing Only the Dependence Structure**
Intercompany experience data

- Singapore database is an intercompany database - allows us to study claims pattern that vary by insurer.
- We use multilevel regression modeling framework:
  - a four level model
  - levels vary by company, insurance contract for a fleet of vehicles, registered vehicle, over time
- This work focuses on claim counts, examining various generalized count distributions including Poisson, negative binomial, zero-inflated and hurdle Poisson models.
- Not surprisingly, we find strong company effects, suggesting that summaries based on intercompany tables must be treated with care.
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

- Our papers show 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!!
### Table A.1. Fitted Negative Binomial 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 $\geq 16$</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 $\leq 21$</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 $\geq 66$</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 ≥ 6</th>
<th>non-automobile</th>
<th>automobile*age ≥ 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>0.128</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4.707</td>
<td>-0.024</td>
<td>-0.024</td>
<td>-0.153</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.281</td>
<td>-0.036</td>
<td>0.252</td>
<td>0.716</td>
<td>-0.201</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.052</td>
<td>-0.129</td>
<td>0.037</td>
<td>-0.349</td>
<td>0.338</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-1.628</td>
<td>0.132</td>
<td>0.132</td>
<td>-0.008</td>
<td>0.330</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.551</td>
<td>-0.089</td>
<td>0.032</td>
<td>-0.259</td>
<td>0.203</td>
<td></td>
</tr>
</tbody>
</table>
# The fitted conditional severity model

## Table A.4. Fitted Severity Model by Copulas

<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</td>
<td>Estimate</td>
</tr>
<tr>
<td></td>
<td>Standard Error</td>
<td>Copula</td>
<td>Standard Error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Estimate</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>$\alpha_{11}$</td>
<td>69.958</td>
<td>69.944</td>
<td>69.772</td>
</tr>
<tr>
<td>$\alpha_{21}$</td>
<td>392.362</td>
<td>129.664</td>
<td>392.496</td>
</tr>
<tr>
<td>intercept</td>
<td>34.269</td>
<td>34.094</td>
<td>31.915</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>$\alpha_{12}$</td>
<td>5.570</td>
<td>5.541</td>
<td>5.758</td>
</tr>
<tr>
<td>$\alpha_{22}$</td>
<td>12.383</td>
<td>12.555</td>
<td>13.933</td>
</tr>
<tr>
<td>intercept</td>
<td>1.987</td>
<td>2.005</td>
<td>2.183</td>
</tr>
<tr>
<td>year</td>
<td>-0.016</td>
<td>-0.015</td>
<td>-0.013</td>
</tr>
<tr>
<td>vehicle capacity</td>
<td>0.116</td>
<td>0.129</td>
<td>0.144</td>
</tr>
<tr>
<td>vehicle age $\leq 5$</td>
<td>0.107</td>
<td>0.106</td>
<td>0.107</td>
</tr>
<tr>
<td>automobile*NCD 0-10</td>
<td>0.102</td>
<td>0.099</td>
<td>0.087</td>
</tr>
<tr>
<td>automobile*age 26-55</td>
<td>-0.047</td>
<td>-0.042</td>
<td>-0.037</td>
</tr>
<tr>
<td>automobile*age $\geq 56$</td>
<td>0.101</td>
<td>0.080</td>
<td>0.084</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>$\alpha_{13}$</td>
<td>0.677</td>
<td>0.615</td>
<td>0.617</td>
</tr>
<tr>
<td>$\alpha_{23}$</td>
<td>1.383</td>
<td>1.528</td>
<td>1.324</td>
</tr>
<tr>
<td>intercept</td>
<td>1.071</td>
<td>1.035</td>
<td>0.841</td>
</tr>
<tr>
<td>vehicle age 1-10</td>
<td>-0.008</td>
<td>-0.054</td>
<td>-0.036</td>
</tr>
<tr>
<td>vehicle age $\geq 11$</td>
<td>-0.022</td>
<td>0.030</td>
<td>0.078</td>
</tr>
<tr>
<td>year</td>
<td>0.031</td>
<td>0.043</td>
<td>0.046</td>
</tr>
<tr>
<td>Copula</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{12}$</td>
<td>-</td>
<td>0.250</td>
<td>0.241</td>
</tr>
<tr>
<td>$\rho_{13}$</td>
<td>-</td>
<td>0.163</td>
<td>0.169</td>
</tr>
<tr>
<td>$\rho_{23}$</td>
<td>-</td>
<td>0.310</td>
<td>0.330</td>
</tr>
<tr>
<td>$\nu$</td>
<td>-</td>
<td>-</td>
<td>6.013</td>
</tr>
</tbody>
</table>
A bit about Singapore
Appendix B - 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: 4+ mn. Age structure: 0-14 yrs: 18%, 15-64 yrs: 75%, 65+ yrs 7%

- Birth rate: 12.79 births/1,000. Death rate: 4.21 deaths/1,000; Life expectancy: 80.1 yrs; male: 77.1 yrs; female: 83.2 yrs

- Ethnic groups: Chinese 77%, Malay 14%, Indian 7.6%; Languages: Chinese, Malay, Tamil, English
A bit about Singapore

- As of 2002: market consists of 40 general ins, 8 life ins, 6 both, 34 general reinsurers, 1 life reins, 8 both; also the largest captive domicile in Asia, with 49 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 2003:
  - total premiums: 15.4 bn; total assets: 77.4 bn [20% annual growth]
  - life insurance: annual premium = 499.8 mn; single premium = 4.6 bn
  - general insurance: gross premium = 5.0 bn (domestic = 2.3; offshore = 2.7)
- Further information: http://www.mas.gov.sg