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

Universidade de São Paulo, Brazil
9 January 2009

Outline

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) $\mathbf{x}_{it}$
  - 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:

Distribution of Claims, by Claim Type Observed

Value of M	1	2	3	4	5	6	7
Claim by Combination	(y_1)	(y_2)	(y_3)	(y_1, y_2)	(y_1, y_3)	(y_2, y_3)	(y_1, y_2, y_3)
Percentage	0.4	73.2	12.3	0.3	0.1	13.5	0.2

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:

$$\begin{aligned} f(N, \mathbf{M}, \mathbf{y}) &= f(N) \times f(\mathbf{M}|N) \times f(\mathbf{y}|N, \mathbf{M}) \\ \text{joint} &= \text{frequency} \times \text{conditional claim-type} \\ &\quad \times \text{conditional severity.} \end{aligned}$$

- This natural decomposition allows us to investigate/model each component separately.
- Frees and Valdez (2009), Hierarchical Insurance Claims Modeling, *Journal of the American Statistical Association*, to appear.
- Frees, Shi and Valdez (2009), Actuarial Applications of a Hierarchical Insurance Claims Model, *ASTIN Bulletin*, submitted.

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 use 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 $\mathbf{x}_{it}$
- The data available therefore consist of

$$\{e_{it}, \mathbf{x}_{it}, N_{it}, M_{itj}, y_{itjk}\}.$$

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 -
  - ages $\geq$21, 22-25, 26-35, 36-45, 46-55, 56-65, $\leq$66.
- The NCD applicable for the calendar year - 0%, 10%, 20%, 30%, 40%, and 50%.

Random effects negative binomial count model

- Let $\lambda_{it} = e_{it} \exp\left(\mathbf{x}'_{\lambda,it}\beta_\lambda\right)$ be the conditional mean parameter for the $\{it\}$ observational unit, where
  - $\mathbf{x}_{\lambda,it}$ is a subset of $\mathbf{x}_{it}$ 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

$$(1-p_{it})/p_{it} = \lambda_{it}\sigma = e_{it} \exp(\mathbf{x}'_{\lambda,it}\beta_\lambda)\sigma.$$

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 $V_m = V_{it,m} = \mathbf{x}'_{M,it}\beta_{M,m}$.

- For our purposes, the covariates in $\mathbf{x}_{M,it}$ 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).

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

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

  where $z = (\ln y - \mu)/\sigma$.
- μ is a location parameter, σ is a scale parameter and α_1 and α_2 are shape parameters.
- 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

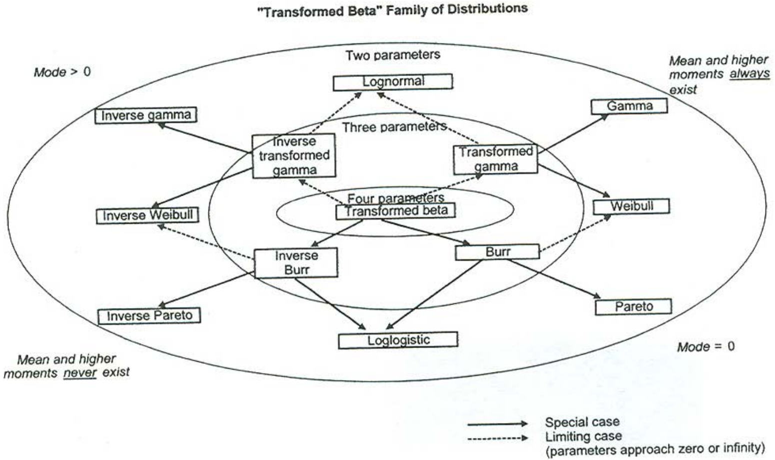


Fig. 4.7 Distributional relationships and characteristics.

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 α_{1k}, α_{2k} and σ_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 $\mu_{ik} = \mathbf{x}_{ik}' \beta_k$:
  - Thus, $\beta_{k,j} = \partial \ln \mathrm{E}\left(Y \vert\, \mathbf{x}\right) / \partial x_j$
  - 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

$$\mathrm{F}(y_1, y_2, y_3) = \mathrm{H}\left(\mathrm{F}_1(y_1), \mathrm{F}_2(y_2), \mathrm{F}_3(y_3)\right).$$

- Here, the marginal distribution of y_k is given by $\mathrm{F}_k(\cdot)$ and $\mathrm{H}(\cdot)$ 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.
- Some references are: Frees and Valdez (1998), Nelsen (1999).

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, $S_i = S_{i,1} + S_{i,2} + S_{i,3} = S_{EVENT,i,1} + ... + S_{EVENT,i,N_i}$.
- 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 α
- 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 $\mu_{ik} = \mathrm{E}(y_{ijk}|N_i, K_i = k)$ from the conditional severity model with an analytic expression

$$\mu_{ik} = \exp(\mathbf{x}_{ik}^{'}\beta_k)\frac{\mathrm{B}(\alpha_{1k}+\sigma_k, \alpha_{2k}-\sigma_k)}{\mathrm{B}(\alpha_{1k}, \alpha_{1k})}.$$

- Basic probability calculations show that:

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

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

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

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

- 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

Type of Random Variable	Level of NCD					
	0	10	20	30	40	50
Individual Loss (Own Damage)	330.67	305.07	267.86	263.44	247.15	221.76
Sum of Losses from a Type (Own Damage)	436.09	391.53	339.33	332.11	306.18	267.63
Sum of Losses from a Specific Event	495.63	457.25	413.68	406.85	381.70	342.48
Overall Loss per Policy	653.63	586.85	524.05	512.90	472.86	413.31

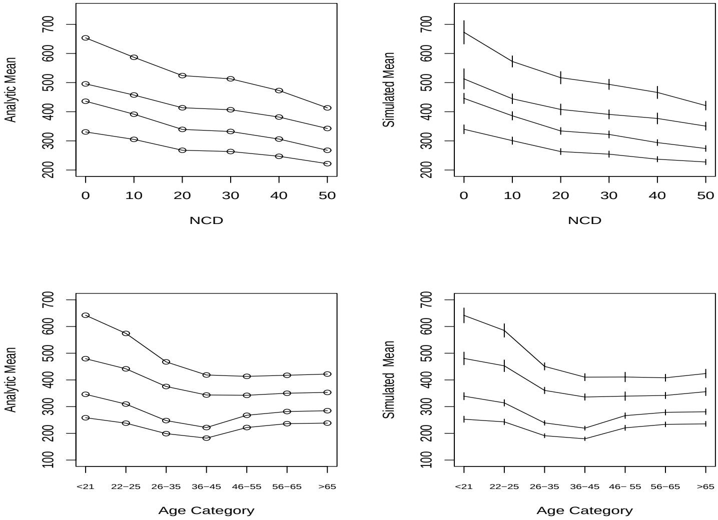
Table 4. Predictive Mean by Insured's Age

Type of Random Variable	Insured's Age						
	≤ 21	22-25	26-35	36-45	46-55	56-65	≥ 66
Individual Loss (Own Damage)	258.41	238.03	198.87	182.04	221.76	236.23	238.33
Sum of Losses from a Type (Own Damage)	346.08	309.48	247.67	221.72	267.63	281.59	284.62
Sum of Losses from a Specific Event	479.46	441.66	375.35	343.59	342.48	350.20	353.31
Overall Loss per Policy	642.14	574.24	467.45	418.47	413.31	417.44	421.93

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

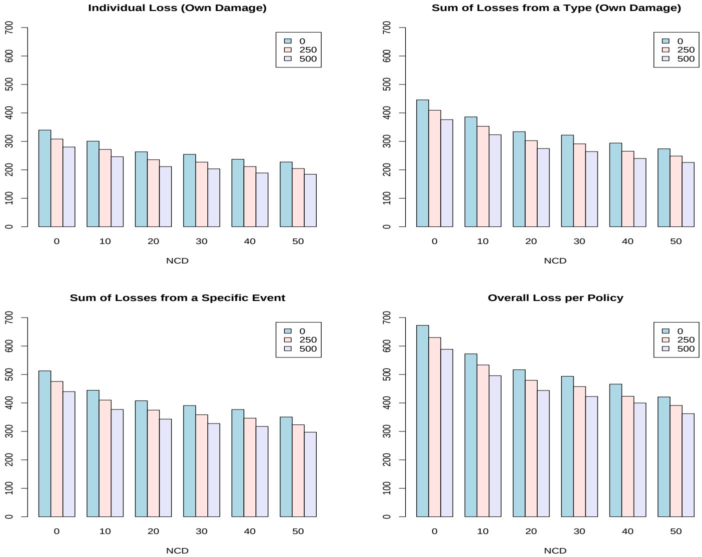


Coverage modifications by level of NCD

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

Coverage Modification			Level of NCD					
Deductible	Limits	Coinsurance	0	10	20	30	40	50
Individual Loss (Own Damage)								
0	none	1	339.78	300.78	263.28	254.40	237.10	227.57
250	none	1	308.24	271.72	235.53	227.11	211.45	204.54
500	none	1	280.19	246.14	211.32	203.43	188.94	184.39
0	25,000	1	331.55	295.08	260.77	250.53	235.42	225.03
0	50,000	1	337.00	300.00	263.28	254.36	237.10	227.27
0	none	0.75	254.84	225.59	197.46	190.80	177.82	170.68
0	none	0.5	169.89	150.39	131.64	127.20	118.55	113.78
250	25,000	0.75	225.00	199.51	174.76	167.43	157.33	151.50
500	50,000	0.75	208.05	184.02	158.49	152.54	141.70	138.07
Sum of Losses from a Type (Own Damage)								
0	none	1	445.81	386.04	334.05	322.09	294.09	273.82
250	none	1	409.38	352.94	302.65	291.29	265.41	248.43
500	none	1	376.47	323.36	274.82	264.12	239.90	225.93
0	25,000	1	434.86	378.55	330.50	316.57	291.78	270.39
0	50,000	1	442.35	385.05	333.98	321.87	294.07	273.40
0	none	0.75	334.36	289.53	250.54	241.56	220.56	205.37
0	none	0.5	222.91	193.02	167.03	161.04	147.04	136.91
250	25,000	0.75	298.82	259.09	224.32	214.33	197.33	183.75
500	50,000	0.75	279.75	241.77	206.06	197.94	179.91	169.13
Sum of Losses from a Specific Event								
0	none	1	512.74	444.50	407.84	390.87	376.92	350.65
250	none	1	475.56	410.12	374.90	358.54	346.58	323.41
500	none	1	439.84	377.11	343.33	327.64	317.47	297.37
0	25,000	1	483.88	433.28	394.80	380.54	359.31	340.67
0	50,000	1	494.20	442.06	401.99	388.21	367.02	348.79
0	none	0.75	384.55	333.38	305.88	293.15	282.69	262.98
0	none	0.5	256.37	222.25	203.92	195.44	188.46	175.32
250	25,000	0.75	335.02	299.17	271.39	261.15	246.73	235.08
500	50,000	0.75	315.98	281.00	253.11	243.74	230.68	221.64
Overall Loss per Policy								
0	none	1	672.68	572.51	516.77	493.93	466.26	421.10
250	none	1	629.88	533.50	479.64	457.56	432.43	391.14
500	none	1	588.55	495.85	443.87	422.63	399.85	362.37
0	25,000	1	634.81	555.90	499.72	479.90	445.04	408.81
0	50,000	1	649.67	568.30	509.52	490.46	454.84	418.92
0	none	0.75	504.51	429.39	387.58	370.45	349.69	315.82
0	none	0.5	336.34	286.26	258.39	246.96	233.13	210.55
250	25,000	0.75	444.01	387.67	346.94	332.65	308.41	284.14
500	50,000	0.75	424.16	368.72	327.46	314.37	291.32	270.15

The effect of deductible, by NCD



Coverage modifications by level of NCD and age

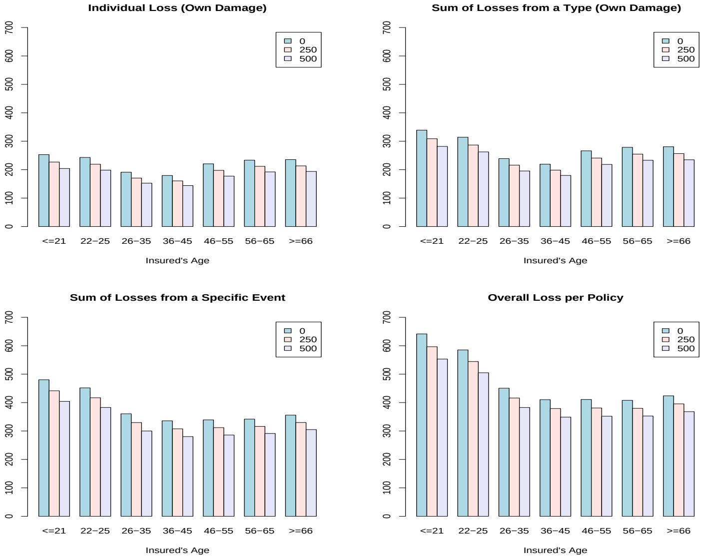
- Now we only use simulation.
- As expected, any of a greater deductible, lower policy limit or smaller coinsurance results in a lower predictive mean.
- Coinsurance changes the predictive means linearly.
- The analysis allows us to see the effects of deductibles and policy limits on long-tail distributions!!!

Coverage modifications by insured's age

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

Coverage Modification			Level of Insured's Age						
Deductible	Limits	Coinsurance	≤21	22-25	26-35	36-45	46-55	56-65	≥66
Individual Losses (Own Damage)									
0	none	1	252.87	242.94	191.13	179.52	220.59	233.58	235.44
250	none	1	226.93	219.16	170.54	160.61	197.57	211.76	213.42
500	none	1	204.13	198.39	152.52	144.00	177.44	192.24	193.78
0	25,000	1	246.94	238.24	189.64	178.33	217.14	230.52	232.35
0	50,000	1	250.64	242.62	191.13	179.46	219.32	233.38	235.44
0	none	0.75	189.65	182.21	143.35	134.64	165.44	175.19	176.58
0	none	0.5	126.43	121.47	95.57	89.76	110.29	116.79	117.72
250	25,000	0.75	165.75	160.84	126.79	119.57	145.60	156.52	157.75
500	50,000	0.75	151.42	148.56	114.39	107.95	132.12	144.03	145.34
Sum of Losses from a Type (Own Damage)									
0	none	1	339.05	314.08	239.04	219.34	266.34	278.61	280.74
250	none	1	308.86	286.80	215.95	198.39	240.96	254.71	256.59
500	none	1	281.82	262.57	195.44	179.74	218.47	233.12	234.84
0	25,000	1	331.01	307.77	236.54	217.53	262.13	274.59	276.51
0	50,000	1	336.33	313.60	238.89	219.16	264.92	278.29	280.67
0	none	0.75	254.29	235.56	179.28	164.50	199.75	208.96	210.55
0	none	0.5	169.53	157.04	119.52	109.67	133.17	139.31	140.37
250	25,000	0.75	225.61	210.37	160.08	147.43	177.56	188.02	189.27
500	50,000	0.75	209.33	196.57	146.47	134.67	162.79	174.60	176.08
Sum of Losses from a specific Event									
0	none	1	480.49	452.84	360.72	336.00	339.24	341.88	355.91
250	none	1	441.68	417.13	329.75	307.68	312.02	316.15	329.97
500	none	1	404.35	382.86	300.06	280.46	285.91	291.37	305.06
0	25,000	1	461.26	434.27	356.68	329.88	326.36	335.92	341.76
0	50,000	1	471.44	444.84	360.30	333.98	331.88	341.66	351.95
0	none	0.75	360.37	339.63	270.54	252.00	254.43	256.41	266.93
0	none	0.5	240.24	226.42	180.36	168.00	169.62	170.94	177.95
250	25,000	0.75	316.83	298.92	244.28	226.17	224.35	232.65	236.87
500	50,000	0.75	296.48	281.14	224.73	208.83	208.91	218.37	225.83
Overall Loss per Policy									
0	none	1	641.63	585.21	450.69	410.37	410.93	408.05	423.90
250	none	1	596.61	544.40	416.07	379.07	380.98	379.93	395.52
500	none	1	553.07	505.04	382.74	348.87	352.15	352.76	368.17
0	25,000	1	616.34	561.58	444.58	402.51	394.26	399.93	406.63
0	50,000	1	630.29	575.81	449.98	407.74	401.61	407.27	419.34
0	none	0.75	481.22	438.91	338.02	307.78	308.20	306.04	317.92
0	none	0.5	320.82	292.60	225.34	205.19	205.46	204.03	211.95
250	25,000	0.75	428.49	390.58	307.48	278.41	273.23	278.86	283.69
500	50,000	0.75	406.30	371.73	286.52	259.68	257.13	263.98	272.71

The effect of deductible, by insured's age



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

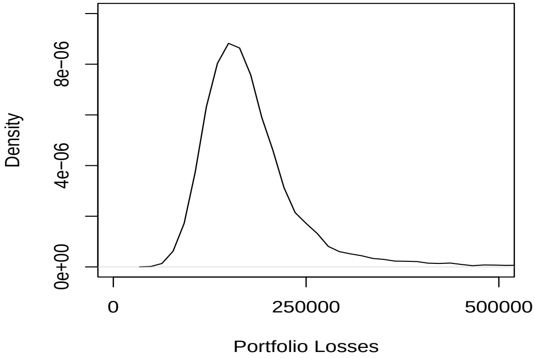


Figure: Simulated Predictive Distribution for a Randomly Selected Portfolio of 1,000 Policies.

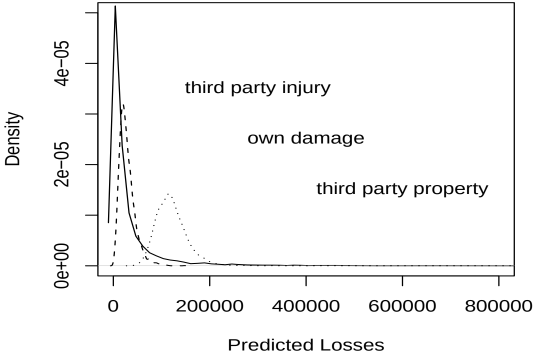


Figure: Simulated Density of Losses for Third Party Injury, Own Damage and Third Party Property of a Randomly Selected Portfolio.

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 - α) 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

Coverage Modification Deductible	Limit	VaR(90%)	Lower Bound	Upper Bound	VaR(95%)	Lower Bound	Upper Bound	VaR(99%)	Lower Bound	Upper Bound
0	none	258,644	253,016	264,359	324,611	311,796	341,434	763,042	625,029	944,508
250	none	245,105	239,679	250,991	312,305	298,000	329,689	749,814	612,818	929,997
500	none	233,265	227,363	238,797	301,547	284,813	317,886	737,883	601,448	916,310
1,000	none	210,989	206,251	217,216	281,032	263,939	296,124	716,955	581,867	894,080
0	25,000	206,990	205,134	209,000	222,989	220,372	225,454	253,775	250,045	256,666
0	50,000	224,715	222,862	227,128	245,715	243,107	249,331	286,848	282,736	289,953
0	100,000	244,158	241,753	247,653	272,317	267,652	277,673	336,844	326,873	345,324
250	25,000	193,313	191,364	195,381	208,590	206,092	211,389	239,486	235,754	241,836
500	50,000	199,109	196,603	201,513	219,328	216,395	222,725	259,436	255,931	263,516
1,000	100,000	197,534	194,501	201,685	224,145	220,410	229,925	287,555	278,601	297,575

Table 8. CTE by Percentile and Coverage Modification with a Corresponding Standard Deviation

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

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**

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

How important is the copula?

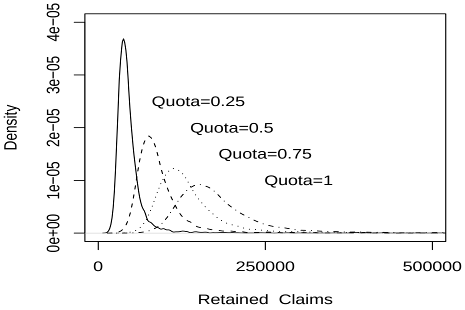
Very!!

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

	VaR			CTE		
Copula	90%	95%	99%	90%	95%	99%
Effects of Re-Estimating the Full Model						
Independence	359,937	490,541	1,377,053	778,744	1,146,709	2,838,762
Normal	282,040	396,463	988,528	639,140	948,404	2,474,151
t	258,644	324,611	763,042	468,850	652,821	1,537,692
Effects of Changing Only the Dependence Structure						
Independence	259,848	328,852	701,681	445,234	602,035	1,270,212
Normal	257,401	331,696	685,612	461,331	634,433	1,450,816
t	258,644	324,611	763,042	468,850	652,821	1,537,692

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.



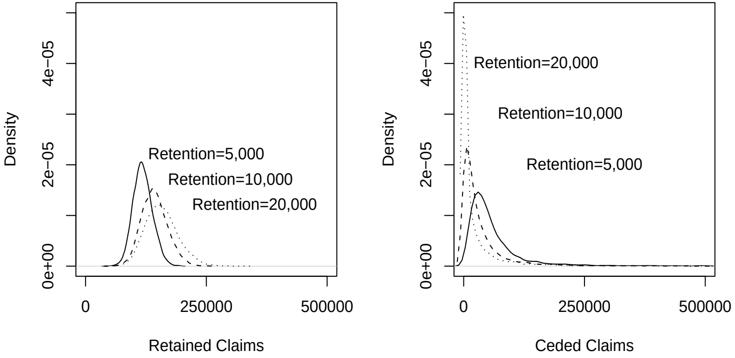


Figure: Distribution of Losses for the Insurer and Reinsurer under Excess of Loss Reinsurance. The losses are simulated under different primary company retention limits. The left-hand panel is for the insurer and right-hand panel is for the reinsurer.

Table 11. Percentiles of Losses for Insurer and Reinsurer under Reinsurance Agreement

Percentile for Insurer											
Quota	Policy Retention	Portfolio Retention	1%	5%	10%	25%	50%	75%	90%	95%	99%
0.25	none	100,000	22,518	26,598	29,093	34,196	40,943	50,657	64,819	83,500	100,000
0.5	none	100,000	45,036	53,197	58,187	68,393	81,885	100,000	100,000	100,000	100,000
0.75	none	100,000	67,553	79,795	87,280	100,000	100,000	100,000	100,000	100,000	100,000
1	10,000	100,000	86,083	99,747	100,000	100,000	100,000	100,000	100,000	100,000	100,000
1	10,000	200,000	86,083	99,747	108,345	122,927	140,910	159,449	177,013	188,813	200,000
1	20,000	200,000	89,605	105,578	114,512	132,145	154,858	177,985	200,000	200,000	200,000
0.25	10,000	100,000	21,521	24,937	27,086	30,732	35,228	39,862	44,253	47,203	53,352
0.5	20,000	100,000	44,803	52,789	57,256	66,072	77,429	88,993	100,000	100,000	100,000
0.75	10,000	200,000	64,562	74,810	81,259	92,195	105,683	119,586	132,760	141,610	160,056
1	20,000	200,000	89,605	105,578	114,512	132,145	154,858	177,985	200,000	200,000	200,000
Percentile for Reinsurer											
Quota	Policy Retention	Portfolio Retention	1%	5%	10%	25%	50%	75%	90%	95%	99%
0.25	none	100,000	67,553	79,795	87,280	102,589	122,828	151,972	194,458	250,499	486,743
0.5	none	100,000	45,036	53,197	58,187	68,393	81,885	102,630	159,277	233,998	486,743
0.75	none	100,000	22,518	26,598	29,093	36,785	63,771	102,630	159,277	233,998	486,743
1	10,000	100,000	0	8,066	16,747	36,888	63,781	102,630	159,277	233,998	486,743
1	10,000	200,000	0	0	992	5,878	18,060	43,434	97,587	171,377	426,367
1	20,000	200,000	0	0	0	0	2,482	24,199	78,839	151,321	412,817
0.25	10,000	100,000	68,075	80,695	88,555	104,557	127,652	161,743	215,407	292,216	541,818
0.5	20,000	100,000	45,132	53,298	58,383	68,909	84,474	111,269	167,106	245,101	491,501
0.75	10,000	200,000	23,536	28,055	31,434	39,746	54,268	81,443	135,853	209,406	462,321
1	20,000	200,000	0	0	0	0	2,482	24,199	78,839	151,321	412,817

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!!

Intercompany experience data

- "A Multilevel Analysis of Intercompany Claim Counts" - joint work with K. Antonio and E.W. Frees.
- 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.

The fitted frequency model

Table A.1. Fitted Negative Binomial Model

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

The fitted conditional claim type model

Table A.2. Fitted Multi Logit Model

	Parameter Estimates				
Category(M)	intercept	year	vehicle age ≫6	non-automobile	automobile*age ≫46
1	1.194	-0.142	0.084	0.262	0.128
2	4.707	-0.024	-0.024	-0.153	0.082
3	3.281	-0.036	0.252	0.716	-0.201
4	1.052	-0.129	0.037	-0.349	0.338
5	-1.628	0.132	0.132	-0.008	0.330
6	3.551	-0.089	0.032	-0.259	0.203

The fitted conditional severity model

Table A.4. Fitted Severity Model by Copulas

Parameter	Types of Copula					
	Independence		Normal Copula		t-Copula	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Third Party Injury						
σ_1	0.225	0.020	0.224	0.044	0.232	0.079
α_{11}	69.958	28.772	69.944	63.267	69.772	105.245
α_{21}	392.362	145.055	392.372	129.664	392.496	204.730
intercept	34.269	8.144	34.094	7.883	31.915	5.606
Own Damage						
σ_2	0.671	0.007	0.670	0.002	0.660	0.004
α_{12}	5.570	0.151	5.541	0.144	5.758	0.103
α_{22}	12.383	0.628	12.555	0.277	13.933	0.750
intercept	1.987	0.115	2.005	0.094	2.183	0.112
year	-0.016	0.006	-0.015	0.006	-0.013	0.006
vehicle capacity	0.116	0.031	0.129	0.022	0.144	0.012
vehicle age $\ll$5	0.107	0.034	0.106	0.031	0.107	0.003
automobile*NCD 0-10	0.102	0.029	0.099	0.039	0.087	0.031
automobile*age 26-55	-0.047	0.027	-0.042	0.044	-0.037	0.005
automobile*age $\gg$56	0.101	0.050	0.080	0.018	0.084	0.050
Third Party Property						
σ_3	1.320	0.068	1.309	0.066	1.349	0.068
α_{13}	0.677	0.088	0.615	0.080	0.617	0.079
α_{23}	1.383	0.253	1.528	0.271	1.324	0.217
intercept	1.071	0.134	1.035	0.132	0.841	0.120
vehicle age 1-10	-0.008	0.098	-0.054	0.094	-0.036	0.092
vehicle age $\gg$11	-0.022	0.198	0.030	0.194	0.078	0.193
year	0.031	0.007	0.043	0.007	0.046	0.007
Copula						
ρ_{12}	-	-	0.250	0.049	0.241	0.054
ρ_{13}	-	-	0.163	0.063	0.169	0.074
ρ_{23}	-	-	0.310	0.017	0.330	0.019
ν	-	-	-	-	6.013	0.688

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