

Actuarial Applications of a Hierarchical Insurance Claims Model

Edward W. Frees
University of Wisconsin

Peng Shi
University of Wisconsin

Emiliano A. Valdez
University of Connecticut *†

February 17, 2008

Abstract

This paper demonstrates actuarial applications that can be performed when modern statistical methods are applied to detailed, micro-level automobile insurance records. We consider 1993-2001 data consisting of policy, claims and payment files from a major Singapore insurance company. A hierarchical statistical model, developed in prior work Frees and Valdez (2008), is fit using the micro-level data. This model allows us to study the accident frequency, loss type and severity jointly and to incorporate individual characteristics such as age, gender and driving history that explain heterogeneity among policyholders.

Based on this hierarchical model, one can analyze the risk profile of either a single policy (micro-level) or a portfolio of business (macro-level). This paper investigates three types of actuarial applications. First, we demonstrate the calculation of the predictive mean of losses for individual risk rating. This allows the actuary to differentiate prices based on policyholder characteristics. The nonlinear effects of coverage modifications such as deductibles, policy limits and coinsurance are quantified. Moreover, our flexible structure allows us to “unbundle” contracts and price more primitive elements of the contract, such as coverage type. The second application concerns the predictive distribution of a portfolio of business. We demonstrate the calculation of various risk measures, including value at risk and conditional tail expectation, that are useful in determining economic capital for insurance companies. Third, we examine the effects of several reinsurance treaties. Specifically, we show the predictive loss distributions for both the insurer and reinsurer under quota share and excess-of-loss reinsurance agreements. In addition, we present an example of portfolio reinsurance, in which the combined effect of reinsurance agreements on the risk characteristics of ceding and reinsuring company are described.

*Keywords: Long-tail regression, copulas, risk measures, reinsurance.

†The first author thanks the National Science Foundation (Grant Number SES-0436274) and the Assurant Health Insurance Professorship for funding to support this research. Please do not quote without the authors' permission

1 Introduction

Actuaries and other financial analysts that work with short term coverages such as automobile insurance and healthcare expenditures typically have massive amounts of in-company data. With modern computing equipment, analysts can readily access data at the individual policyholder level that we term “micro-level.” Actuaries use statistical models to summarize micro-level data that subsequently need to be interpreted properly for financial decision-making. For example, automobile insurers typically differentiate premium rates based on policyholder characteristics such as age, gender and driving history. Gourieroux and Jasiak (2007) have dubbed this emerging field the “microeconometrics of individual risk.”

Developing a statistical model to substantiate rate differentials may not be straightforward because policyholder characteristics can affect the frequency (number of accidents) and the severity (amount or intensity of the accident) in different ways. Moreover, policies differ based on the type of coverage offered as well as payment limitations (such as deductibles, upper limits and coinsurance). Actuaries are also interested in other financial measures in addition to those used for pricing. For example, actuaries use measures that summarize portfolio risks for capital allocation and solvency purposes. As another example, actuaries involved in managing risk through reinsurance use statistical models to calibrate reinsurance treaties.

This paper demonstrates analyses that can be used for pricing, economic capital allocation, solvency and reinsurance based on a statistical model of a specific database. The structure of the database encompasses policyholder, claim and payment files and is widely used. To illustrate, all property and casualty (general) insurers domiciled in Singapore use this structure to report their data to a quasi-governmental organization, the General Insurance Association (G.I.A.) of Singapore.

Our data are from a Singaporean insurance company. From the policyholder file, we have available several policyholder characteristics that can be used to anticipate automobile insurance claims. Additional data characteristics are in Section 2.2. For each policy i , we are interested in predicting:

- N_i - the number of losses and
- y_{ijk} - the loss amount, available for each loss, $j = 1, \dots, N_i$, and the type of loss $k = 1, 2, 3$.

When a claim is made, it is possible to have one or a combination of three types of losses. We consider: (1) losses for injury to a party other than the insured y_{ij1} , (2) losses for damages to the insured, including injury, property damage, fire and theft y_{ij2} , and (3) losses for property damage to a party other than the insured y_{ij3} . Occasionally, we shall simply refer to them as “injury,” “own damage” and “third party property.” It is not uncommon to have more than one type of loss incurred with each accident.

The hierarchical model was developed in our prior work (Frees and Valdez, 2008). This model allows for:

- risk rating factors to be used as explanatory variables that predict both the frequency and multivariate severity,
- the long-tail nature of the distribution of insurance claims through the GB2 (generalized beta of the second kind) distribution and
- the “two-part” distribution of losses. When a claim occurs, we do not necessarily realize all three types of losses. Each type of loss may equal zero or may be continuous and hence be comprised of “two parts.” Further, we allow for
- losses to be related through a t -copula specification.

To keep this paper self-contained, this statistical model is described in the Section 2.3. To give readers a feel for the historical precedents of this model, Section 2.1 provides a review of the literature.

We fit this model to 1993-2000 company data. The statistical model allows us to provide predictions for a 2001 portfolio of $n=13,739$ policies. Several risk rating factors known at the beginning of the year are used to develop predictions. The focus of this paper is to show how predictions from this statistical model can be used for rating and portfolio risk management.

Section 3 discusses individual risk rating by the predictive mean, a measure that actuaries base prices upon. To begin, the model allows us to quantify differences by risk rating factors. Moreover, although our data is based on a comprehensive coverage that offers protection for “injury,” “own damage” and “third party property,” we can compute predictive means for each type of coverage. This allows the analyst to assess the effects of “unbundling” the comprehensive coverage.

We also quantify the effect of other policy modifications, such as if the policy paid only for one claim during the year or the effect of policy limits such as deductibles, upper coverage limit (by type of coverage) and coinsurance. This type of flexibility allows the actuary to design a coherent pricing structure that is appealing to consumers.

Section 4 examines predictive distributions for a portfolio or group of policies. Here, we seek to gain insights into the effects of “micro-level” changes, such as the introduction of a policy deductible or upper limit, on the “macro-level” group basis. In principle, the group could be selected by geographic region, sales agent type or some other method of clustering. For illustrative purposes, we simply take random subsets of our portfolio of $n=13,739$ policies. Because of our concern of macro-level effects, Section 4 goes beyond simply the predictive mean and examines other summary measures of the distribution. We do this by Monte Carlo methods, where we simulate the loss frequency, type and severity for each policy in the group.

The portfolio summary measures include well-known tail summary measures such as the value at risk (VaR) and the conditional tail expectation (CTE). Thus, the tools that we propose can be used to determine economic capital for insurers as well as for solvency testing. For example, we are able to quantify the effect of imposing an upper limit on injury protection on the VaR for the portfolio. Although we allow for long-tail severity for individual policies, we can give practical guidance as to when approximate normality (via a central limit theorem) holds for a portfolio of

policies.

Section 5 retains the focus on portfolio of policies but now with an eye towards quantifying the impact of reinsurance treaties. Essentially, we are interested in “macro-level” changes on a group (macro-level) basis. We discuss proportional reinsurance as well as non-proportional reinsurance, and examine the effect of reinsurance agreements on the losses distribution. An example of portfolio reinsurance is given, and the combined effects of reinsurance agreements are investigated by showing the tail summary measures of losses for both insurer and reinsurer.

Section 6 provides a summary and closing remarks.

2 Hierarchical Insurance Claims

2.1 Literature Review

There is a rich literature on modeling the joint frequency and severity distribution of automobile insurance claims. To distinguish this modeling from classical risk theory applications (see, for example, Klugman, Panjer and Willmot, 2004), we focus on cases where explanatory variables, such as policyholder characteristics, are available. There has been substantial interest in statistical modeling of claims frequency, see Boucher and Denuit (2006) for a recent example. However, the literature on modeling claims severity, especially in conjunction with claims frequency, is less extensive. One possible explanation, noted by Coutts (1984), is that most of the variation in overall claims experience may be attributed to claim frequency (at least when inflation was small). Coutts (1984) also remarks that the first paper to analyze claim frequency and severity separately seems to be Kahane and Levy (1975).

Brockman and Wright (1992) provide an earlier overview of how statistical modeling of claims and severity can be helpful for pricing automobile coverage. For computational convenience, they focused on categorical pricing variables to form cells that could be used with traditional insurance underwriting forms. Renshaw (1994) shows how generalized linear models can be used to analyze both the frequency and severity portions based on individual policyholder level data. Hsiao et al. (1990) note the “excess” number of zeros in policyholder claims data (due to no claims) and compare and contrast Tobit, two-part and simultaneous equation models, building on the work of Weisberg and Tomberlin (1982) and Weisberg et al. (1984). However, all of these papers use grouped data, not individual policyholder level data as in this paper.

Pinquet (1997, 1998) provides a more modern statistical approach, fitting not only cross-sectional data but also following policyholders over time. Pinquet was interested in two lines of business, claims at fault and not at fault with respect to a third party. For each line, Pinquet hypothesized a frequency and severity component that were allowed to be correlated to one another. In particular, the claims frequency distribution was assumed to be bivariate Poisson. Severities were modeled using lognormal and gamma distributions. Also at the individual policyholder level, Frangos and Vrontos (2001) examined a claim frequency and severity model, using negative binomial and Pareto distributions, respectively. They used their statistical model to develop experience

rated (bonus-malus) premiums.

2.2 Data

Our statistical model of insurance claims is based on detailed, micro-level automobile insurance records. Specifically, we analyzed information from three databases: the policy, claims and payment files. The policy file consists of all policyholders with vehicle insurance coverage purchased from a general insurer during the observation period. Each vehicle is identified with a unique code. This file provides characteristics of the policyholder and the vehicle insured, such as age and gender and type and age of vehicle insured. The claims file provides a record of each accident claim that has been filed with the insurer during the observation period and is linked to the policyholder file. The payment file consists of information on each payment that has been made during the observation period and is linked to the claims file. Although it is common to see that a claim will have multiple payments made, we do not use that information for this paper and aggregate all payments that arise from each accident event. See Antonio et al. (2006) for a recent description of the claims “run-off” problem.

The policyholder file provides several characteristics to help explain and predict automobile accident frequency, type and severity. These characteristics include information about the vehicle, such as type and age, as well as person level characteristics, such as age, gender and prior driving experience. Table 1 summarizes these characteristics. These characteristics are denoted by the vector \mathbf{x}_{it} and will serve as explanatory variables in our analysis. Person level characteristics are largely unavailable for commercial use vehicles, so the explanatory variables of personal characteristics are only used for observations having non-commercial purposes. We also have the exposure e_{it} , measured in (a fraction of) years, which provides the length of time throughout the calendar year for which the vehicle had insurance coverage.

| Table 1. Description of Risk Rating Factors | |
|--|---|
| Covariates | Description |
| Year | The calendar year, from 1993-2001 inclusive |
| Vehicle Type | The type of vehicle insured, either automobile (A) or others (O). |
| Vehicle Age | The age of vehicle, in years, grouped in six categories. |
| Vehicle Capacity | The cubic capacity of the vehicle |
| Gender | The gender of the policyholder, either male or female |
| Age | The age of the policyholder, in years, grouped in to seven categories. |
| NCD | No Claim Discount. This is based on the previous accident record of the policyholder. |
| | The higher the discount, the better is the prior accident record. |

With the information in the claims file, we potentially observe a trivariate claim amount, one claim for each type. For each accident, it is possible to have more than a single type of claim incurred; for example, an automobile accident can result in damages to a driver’s own property as well as damages to a third party who might be involved in the accident.

To provide focus, we restrict our considerations to “non-fleet” policies; these comprise about 90% of the policies for this company. These are policies issued to customers whose insurance covers a single vehicle. In contrast, fleet policies are issued to companies that insured several vehicles, for example, coverage provided to a taxicab company, where several taxicabs are insured. See Angers et al. (2006) and Desjardins et al. (2001) for discussions of fleet policies. The unit of observation in our analysis is therefore a registered vehicle insured, broken down according to their exposure in each calendar year 1993 to 2001. In order to investigate the full multivariate nature of claims, we further restrict our consideration to policies that offer comprehensive coverage, not merely for only third party injury or property damage.

In summary, the hierarchical insurance claims model is based on observable data consisting of

$$\{e_{it}, \mathbf{x}_{it}, N_{it}, y_{itk}, k = 1, 2, 3, t = 1, \dots, T_i, i = 1, \dots, n\}.$$

2.3 Statistical Model

The statistical model, from Frees and Valdez (2008), consists of three components - each component uses explanatory variables to account for heterogeneity among policyholders. The first component uses a negative binomial regression model to predict accident probability. The second component uses a multinomial logit model to predict type of losses, either third party injury, own damage, third party property or some combination. The third component is for severity; here, a GB2 distribution is used to fit the marginal distributions and a t -copula is used to model the dependence of the three types of claims.

It is customary in the actuarial literature to condition on the frequency component when analyzing the joint frequency and severity distributions. See, for example, Klugman, Panjer and Willmot (2004). Frees and Valdez (2008) incorporate an additional claims type layer to handle the many zeros in each claims distribution claims (known as “two-part” data) as well as accounting for the possibility of multivariate claims. Specifically, conditional on having observed at least one type of claim, the random variable M describes which of the seven combinations is observed. Table 2 provides potential values of M .

| Value of M | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------------------------------|---------|---------|---------|--------------|--------------|--------------|-------------------|
| Claim by Combination Observed | (y_1) | (y_2) | (y_3) | (y_1, y_2) | (y_1, y_3) | (y_2, y_3) | (y_1, y_2, y_3) |

We are now in a position to describe the full predictive model. Suppressing the $\{i\}$ subscripts, the joint distribution of the dependent variables is:

$$f(N, \mathbf{M}, \mathbf{y}) = f(N) \times f(\mathbf{M}|N) \times f(\mathbf{y}|N, \mathbf{M})$$

joint = frequency \times conditional claim type \times conditional severity,

where $f(N, \mathbf{M}, \mathbf{y})$ denotes the joint distribution of $(N, \mathbf{M}, \mathbf{y})$.

We now discuss each of the three components. The parameter estimates corresponding to each component are provided in Appendix A.1.

2.3.1 Frequency Component

For our purposes, we use standard count models. For these models, one uses $\lambda_{it} = e_{it} \exp(\mathbf{x}'_{\lambda, it} \boldsymbol{\beta}_{\lambda})$ to be the conditional mean parameter for the it^{th} observational unit. Here, the vector $\mathbf{x}_{\lambda, it}$ is a subset of \mathbf{x}_{it} , representing the variables needed for frequency modeling. The amount of exposure, e_{it} , is used as an offset variable because a driver may have insurance coverage for only part of the year. We use the negative binomial distribution with parameters p and r , so that $\Pr(N = k) = \binom{k+r-1}{r-1} p^r (1-p)^k$. Here, $\sigma = r^{-1}$ is the dispersion parameter and $p = p_{it}$ is related to the mean through $(1 - p_{it})/p_{it} = \lambda_{it} \sigma = e_{it} \exp(\mathbf{x}'_{\lambda, it} \boldsymbol{\beta}_{\lambda}) \sigma$.

2.3.2 Claims Type Component

The multinomial logit regression model allows us to incorporate explanatory variables into our explanations of the claim type. This model is of the form $\Pr(M = m) = \exp(V_m) / \left\{ \sum_{s=1}^7 \exp(V_s) \right\}$, where $V_m = V_{it, m} = \mathbf{x}'_{M, it} \boldsymbol{\beta}_{M, m}$. Note that for our application, the covariates in $\mathbf{x}_{M, it}$ do not depend on the accident number j nor on the claim type m although we allow parameters ($\boldsymbol{\beta}_{M, m}$) to depend on m . This portion of the model was proposed by Terza and Wilson (1990).

2.3.3 Severity Component

To accommodate the long-tail nature of claims, we use the GB2 distribution for each claim type. This has density function

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

where $z = (\ln y - \mu)/\sigma$ and $\text{B}(\alpha_1, \alpha_2) = \Gamma(\alpha_1)\Gamma(\alpha_2)/\Gamma(\alpha_1 + \alpha_2)$, the usual beta function. Here, μ is a location parameter, σ is a scale parameter and α_1 and α_2 are shape parameters. This distribution is well known in actuarial modeling of univariate loss distributions (see for example, Klugman, Panjer and Willmot, 2004). With four parameters, the distribution has great flexibility for fitting heavy tailed data. 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).

We use this distribution but 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 the prominence of the GB2 in fitting distributions to univariate data, there are relatively few applications that use the GB2 in a regression context. Recently, Sun et al. (2008) used the GB2 in a longitudinal data context to forecast nursing home utilization.

To accommodate dependencies among claim types, we use a parametric copula. See Frees and Valdez (1998) for an introduction to copulas. Suppressing the $\{i\}$ subscripts, we may write 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 linking the marginals to the joint distribution. We use a trivariate t -copula with an unstructured correlation matrix.

3 Predictive Means for Individual Risk Rating

Given a set of risk rating factors such as in Table 1, one basic task is to arrive at a fair price for a contract. In setting prices, often the actuary is called upon to quantify the effects of certain policy modifications. As a basis of fair pricing is the predictive mean, this section calculates predictive means for several alternative policy designs that may be of interest.

For alternative designs, we consider four random variables:

- individuals losses, y_{ijk}
- the sum of losses from a type, $S_{i,k} = y_{i,1,k} + \dots + y_{i,N_i,k}$
- 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, $S_i = S_{i,1} + S_{i,2} + S_{i,3} = S_{EVENT,i,1} + \dots + S_{EVENT,i,N_i}$.

Our database is from comprehensive policies, associated with the fourth random variable. The other random variables represent different ways of “unbundling” this coverage, similar to decomposing a financial contract into primitive components for risk analysis. The first random variable can be thought of as claims arising from a policy that covers losses from a single accident of a certain type. The second represents claims from a policy that covers all losses within a year of a certain type. The third variable corresponds to a policy that covers all types of losses from a single accident.

We also examine the effect of standard coverage modifications that consist of (1) deductibles d , (2) coverage limits u and (3) coinsurance percentages α . As in Klugman et al. (2004), we define the function

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

From the conditional severity model, we define $\mu_{ik} = E(y_{ijk} | N_i, K_i = k)$. The random variable K_i indicates the type, for $K_i = 1, 2, 3$. Then, basic probability calculations show that:

$$E(y_{ijk}) = \Pr(N_i = 1)\Pr(K_i = k)\mu_{ik}, \tag{2}$$

$$E(S_{i,k}) = \mu_{ik} \Pr(K_i = k) \sum_{n=1}^{\infty} n \Pr(N_i = n), \quad (3)$$

$$E(S_{EVENT,i,j}) = \Pr(N_i = 1) \sum_{k=1}^3 \mu_{ik} \Pr(K_i = k), \quad \text{and} \quad (4)$$

$$E(S_i) = E(S_{i,1}) + E(S_{i,2}) + E(S_{i,3}). \quad (5)$$

Thus, to compute means, we only need to calculate the probability of number and type of loss, as well as the expected loss given the type of loss in case of accident. Appendix A.1 provides the necessary coefficients and probabilities estimates.

To provide baseline comparisons, we emphasize the simplest situation, a policy without deductible, coverage limits and coinsurance modifications. In this case, from the severity distribution, we have an analytic expression for the conditional mean of the form

$$\mu_{ik} = \exp(\mathbf{x}_{ik}' \beta_k) \frac{B(\alpha_{1k} + \sigma_k, \alpha_{2k} - \sigma_k)}{B(\alpha_{1k}, \alpha_{1k})}, \quad (6)$$

where $\beta_k, \alpha_{jk}, \sigma_k$ are parameters of the GB2 severity distribution (see Section 2.3.3). With policy modifications, we approximate μ_{ik} via simulation (see Section A.2).

The predictive means in equations (2)-(5), by level of no claims discount (NCD) and insured's age, are shown in Tables 2 and 3, respectively. The calculation is based on a randomly selected observation from the 2001 portfolio. The policyholder is a 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. (Appendix A.1 shows that both NCD and age are statistically significant variables in at least one component of the predictive model, either frequency, type or severity.) The point of this section is to understand their economic significance.

Using equation (6), Table 3 shows that the insured who enjoys a higher no claim discount has a lower expected loss; this result holds for all four random variables. This is consistent with our intuition and, as shown in the Appendix, with the statistical model fitting results. To what extent are these overall results driven by the frequency versus the severity components? Appendix A.3 shows how to analyze this question (for this particular test case).

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 |

| 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 |

Table 4 presents a more complex nonlinear pattern for insured's age. For each random variable, predictive means are at their highest at the youngest age group, decrease as age increases and are stable for age group 46-55 and on. Figure 1 presents a graphical display in the lower left-hand corner. Appendix A.3 presents further analysis.

When different coverage modifications are incorporated, we need to simulate the amount of losses to calculate predictive means. Tables 5 and 6 show the effects of different policy designs under various combinations of NCD and insured's age. The first row under each of the four random variables corresponds to the predictive mean for the policy without any coverage modification; here, readers will notice a slight difference between these entries and the corresponding entries in Tables 3 and 4. This is due to simulation error. We used 5,000 simulated values.

To understand the simulation error, Figure 1 compares the analytic and simulated predictive means. Here, we consider the case of no deductible, policy limits and coinsurance, so that the analytic result in equation (5) is available. The upper two panels shows the relationship between predictive mean and NCD, whereas the lower two panels are for insured's age. The two panels on the left are for the analytic result, whereas the two panels on the right are for the simulation results. For the simulated results, the lines provide the 95% confidence intervals. The width of these lines show that the simulation error is negligible for our purposes - for other purposes, one can always reduce the simulation error by increasing the simulation size.

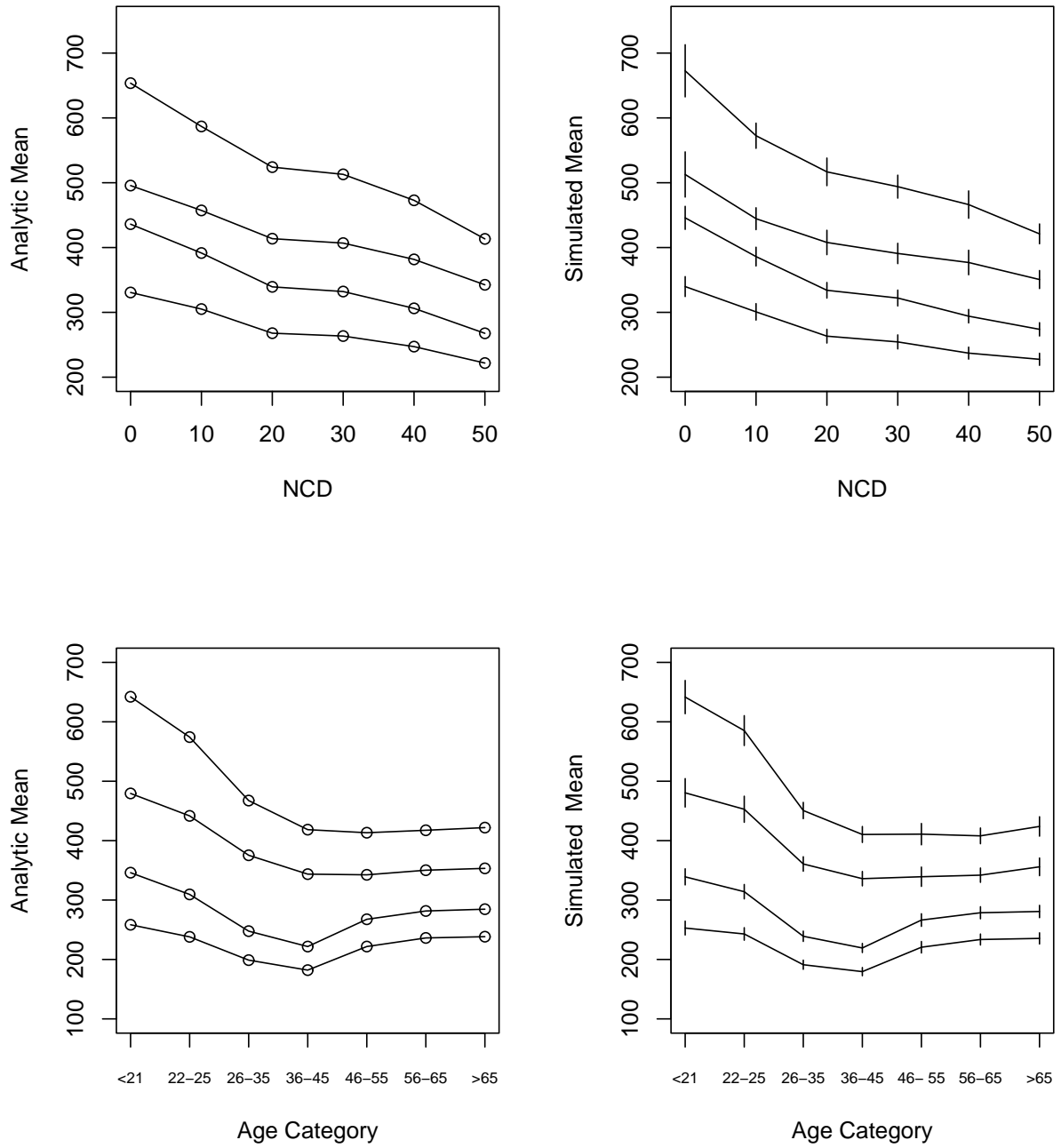


Figure 1: Simulated Predictive Mean and Confidence Interval. The left-hand panels provide analytical means, the right-hand panels are based on simulated values.

Table 5 shows the simulated predictive mean at different levels of NCD under various coverage modifications. As expected, any of a greater deductible, lower policy limit or smaller coinsurance results in a lower predictive mean. Deductibles and policy limits change the predictive mean nonlinearly, whereas coinsurance changes the predictive mean linearly. For example, the predictive mean decreases less when the deductible increases from 250 to 500, compared to the decrease when deductible increases from 0 to 250. This pattern applies to all the four loss variables at all NCD levels. The effect of policy limit depends on the expected loss. For random variables with a small expected loss (e.g. individual loss and sum of losses from a type), there is little difference in predictive means between policies with a 50,000 limit and no limit. In contrast, for random variables with large expected losses (e.g. sum of losses from a specific event and overall losses), the difference in predictive means can be greater when limit increases from 25,000 to 50,000 than an increase from 50,000 to no limit.

Table 6 shows the effect of coverage modifications on the predictive mean for the insured at different age categories. As with Table 5, any of a higher deductible, lower coverage limit or lower coinsurance percentage results in a lower predictive mean. The combined effect of three kinds of coverage modifications can be derived from the three marginal effects. For example, when the insured's age is in the 26-35 category, the predictive mean of individual loss with deductible 250, coverage limit 25,000 and coinsurance 0.75 can be calculated from predictive mean of individual loss with deductible 250 and predictive mean of individual loss with coverage limit 25,000, that is $(170.54 + 189.64 - 191.13) * 0.75 = 126.79$. Similar results can be derived for all the four random variables under different NCD or insured's age values.

Comparing Tables 5 and 6, the effect of NCD has a greater effect on the predictive mean than that of insured's age, in the sense that the range of predictive means is greater under alternative NCD levels compared to age levels. For example, for a policy with a 500 deductible, 50,000 policy limit and 0.75 coinsurance, the predictive mean of individual losses is 138.07 from Table 5 and 132.12 from Table 6 (the difference is due to simulation error). For this policy, the predictive mean varies between 138.07 and 208.05 under various NCD levels, and varies between 107.95 and 151.42 under alternative insured's age categories. Under other coverage modifications we observe similar results.

**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 |

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 |

4 Predictive Distributions for Portfolios

4.1 Predictive Distribution

Actuaries are trained to look beyond the mean - to manage risk, one should understand a risk's entire distribution. This section examines the predictive distribution for a portfolio of risks.

In contrast, Section 3 dealt with a single illustrative contract. For a single contract, there is a large mass at zero (about 92% for many combinations of risk rating factors) and thus each of the random variables introduced in Section 3 had a large discrete component as well as continuous severity component. For a Bernoulli random variable, it is known that the mean determines the distribution. Because of the analogy between Bernoulli random variables and the Section 3 random variables, analysts have historically tended to focus on the mean as the first step in understanding the distribution.

One cannot make that case for a portfolio of risks. As noted in Section 1, the portfolio could be selected by geographic region, sales agent type or some other method of clustering. For illustrative purposes, we have randomly selected 1,000 policies from our 2001 sample. If the concern is with overall losses, we wish to predict the distribution of $S = S_1 + \dots + S_{1000}$. Clearly the predictive mean provides only one summary measure.

Studying the distribution of S is a well-known problem in probability that receives substantial attention in actuarial science, see for example, Klugman et al. (Chapter 6, 2004). The problem is to analyze the *convolution* of distribution functions. Unlike textbook treatments, for our application the distribution functions for S_1, \dots, S_{1000} are nonidentical, each having a discrete and highly non-normal continuous component. Thus, although analytic methods are feasible, we prefer to use simulation methods to compute the predictive distribution of S . For each policy, using known risk rating factors and estimated parameters, we simulated the event of an accident and type of loss, as well as the severity of losses. These components are then used to develop simulated values of each of the four types of random variables introduced in Section 3. As in Section 3, we report results based on 5,000 replications. Further details of the simulation procedure are in Appendix A.2.

Figure 2 summarizes the results for the overall loss, S . This figure shows that by summing over 1,000 policies, the discrete component is no longer evident. It is also interesting to see that the portfolio distribution is still long-tail. Elementary statistics texts, citing the central limit theorem, typically state that the normal is a good approximation for the distribution of the sum based on 30 to 50 i.i.d. draws. The distribution of the sum shown in Figure 2 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.

For comparison purposes, the density for actual losses in year 2001 is shown in Figure 6 in Appendix A.3. Like the Figure 2 simulated density, the Figure 6 density also has long tails and is not approximately normal.

As described in Section 3, there may be situations in which the analyst is interested in the claims from each type of coverage instead of the total claims in a comprehensive policy. In this

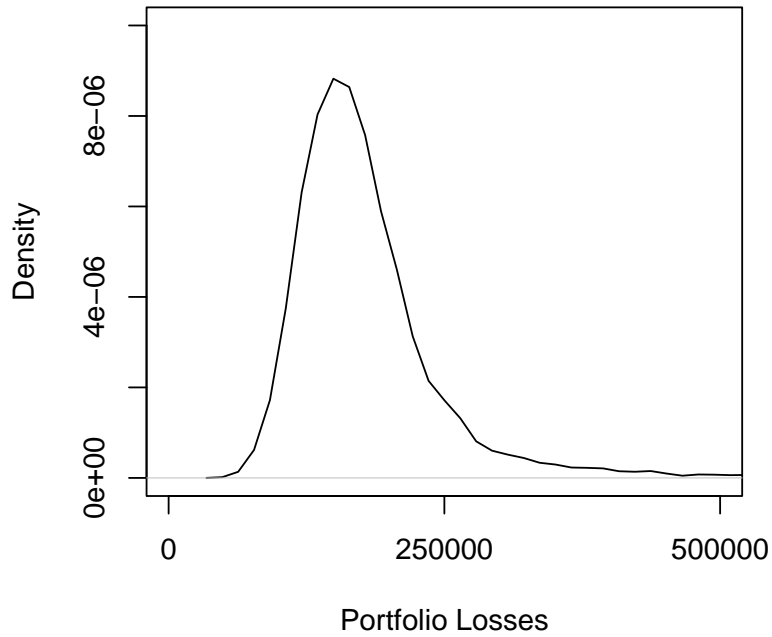


Figure 2: Simulated Predictive Distribution for a Randomly Selected Portfolio of 1,000 Policies.

case, one should consider the random variables $S_k = \sum_{i=1}^{1000} S_{i,k}$, where $k = 1, 2, 3$ corresponds to third party injury, own damage and third party property claims.

As shown in Figure 3, the distributions for the three types of losses are quite different in terms of their skewness and kurtosis as well as other properties. The density for third party injury and own damage have higher peaks and are more positively skewed than third party property. The density of third party injury has a heavier tail. The distributions for actual losses in year 2001 for the three types of losses are shown in Figure 7 in Appendix A.3. Figure 7 also show the significant differences among densities of three types of losses. This type of analysis can provide useful information about the risk profile of different coverages within a comprehensive policy, as well as the risk profile of different lines of business. For example, the analyst may consider which type of coverage should be incorporated and which type should be eliminated, so that the product can be tuned to meet the company's risk management requirement, regulatory policy or other strategic goals.

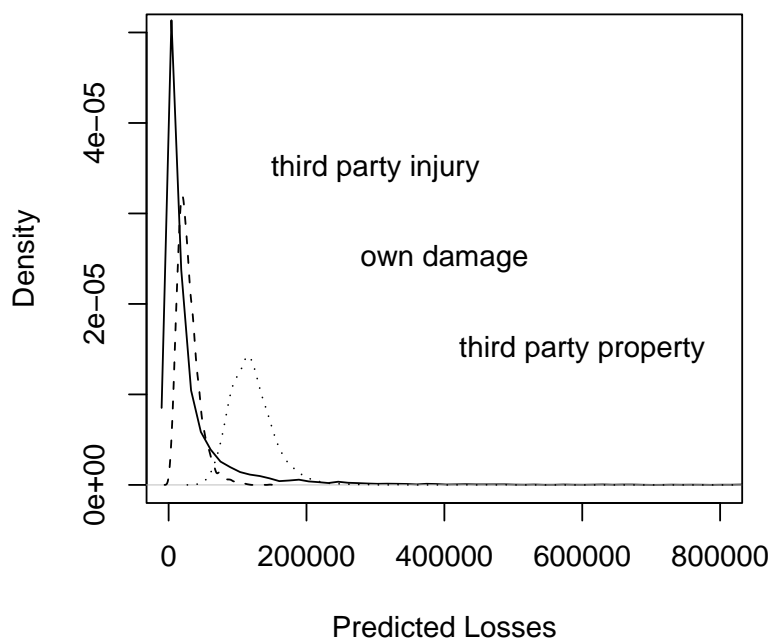


Figure 3: Simulated Density of Losses for Third Party Injury, Own Damage and Third Party Property of a Randomly Selected Portfolio.

4.2 Risk Measures

Graphical displays of distributions help analysts understand patterns, linear and nonlinear; numerical summary measures complement these displays by providing precise information of selected aspects of the distribution. This section examines the Value-at-Risk (VaR) and Conditional Tail Expectation (CTE), two numerical risk measures focusing on the tail of the distribution that have been widely used in both actuarial and financial work. The VaR is simply a quantile or percentile; $VaR(\alpha)$ gives the $100(1 - \alpha)$ percentile of the distribution. The $CTE(\alpha)$ is the expected value conditional on exceeding the $VaR(\alpha)$. See, for example, Hardy (2006) for further background on these and related risk measures.

In addition to the effects of coverage modifications on predictive mean investigated in Section 3, we are also interested in their effects on the distribution of losses. In this section we focus on VaR and CTE , shown in Tables 7 and 8, respectively. The calculations are based on the randomly selected portfolio of policies as investigated above. The results in first row of two tables are corresponding to a policy without deductibles and limits.

Table 7 shows the VaR s at different quantiles and coverage modifications, with a corresponding

95% confidence interval. The results are consistent with expectations. First, larger deductibles and smaller policy limits decrease the VaR in a nonlinear way. The marginal effect of the deductible on VaR decreases as the deductible increases; for example, the VaR difference between deductibles 0 and 250 is larger than the VaR difference between deductibles 250 and 500. Similarly, the marginal effect of policy limits also decreases as the policy limit increases.

Second, under each combination of deductible and policy limit, the confidence interval becomes wider as the VaR percentile increases. This result is in part because of the heavy tailed nature of the losses. Third, policy limits exert a greater effect than deductibles on the tail of the distribution. This can be seen by comparing the $VaRs$ in the last three rows in Table 7. The three policy designs consist of an increasing deductible (which decreases VaR) and an increasing policy limit (which increases VaR); the overall results show an increasing effect on VaR . Fourth, the policy limit exerts a greater effect than a deductible on the confidence interval capturing the VaR .

Table 8 shows $CTEs$ by percentile and coverage modification with a corresponding standard deviation. The results are consistent with Table 7. Either a larger deductible or a smaller policy limit results in a lower CTE and the effect is nonlinear. The range of policy limits explored has a greater influence on CTE than the range of deductibles. The sparsity of data combined with the heavy tailed nature of the distribution result in the greater standard deviations at higher percentiles. Finally, the decrease in the policy limits reduces the standard deviation of the CTE substantially whereas changes in the deductible have little influence.

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 | 312,305 | 298,000 | 329,689 | 763,042 | 625,029 | 944,508 | 749,814 | 612,818 | 929,997 | 737,883 | 601,448 | 916,310 | 581,867 | 894,080 | |
| 250 | none | 245,105 | 239,679 | 250,991 | 301,547 | 284,813 | 317,886 | 281,032 | 263,939 | 296,124 | 253,775 | 250,045 | 256,666 | 286,848 | 282,736 | 289,953 | 336,844 | 326,873 | 345,324 | 239,486 | 235,754 | 241,836 |
| 500 | none | 233,265 | 227,363 | 238,797 | 222,989 | 220,372 | 225,454 | 245,715 | 243,107 | 249,331 | 272,317 | 267,652 | 277,673 | 208,590 | 206,092 | 211,389 | 219,328 | 216,395 | 222,725 | 224,145 | 220,410 | 229,925 |
| 1,000 | none | 210,989 | 206,251 | 217,216 | 206,990 | 205,134 | 209,000 | 244,158 | 241,753 | 247,653 | 193,313 | 191,364 | 195,381 | 199,109 | 196,603 | 201,513 | 197,534 | 194,501 | 201,685 | | | |
| 0 | 25,000 | 206,990 | 205,134 | 209,000 | 224,715 | 222,862 | 227,128 | 244,158 | 241,753 | 247,653 | 193,313 | 191,364 | 195,381 | 199,109 | 196,603 | 201,513 | 197,534 | 194,501 | 201,685 | | | |
| 0 | 50,000 | 224,715 | 222,862 | 227,128 | 244,158 | 241,753 | 247,653 | 193,313 | 191,364 | 195,381 | 199,109 | 196,603 | 201,513 | 197,534 | 194,501 | 201,685 | | | | | | |
| 0 | 100,000 | 244,158 | 241,753 | 247,653 | 206,990 | 205,134 | 209,000 | 224,715 | 222,862 | 227,128 | 244,158 | 241,753 | 247,653 | 193,313 | 191,364 | 195,381 | 199,109 | 196,603 | 201,513 | 197,534 | 194,501 | 201,685 |
| 250 | 25,000 | 193,313 | 191,364 | 195,381 | 206,990 | 205,134 | 209,000 | 224,715 | 222,862 | 227,128 | 244,158 | 241,753 | 247,653 | 193,313 | 191,364 | 195,381 | 199,109 | 196,603 | 201,513 | 197,534 | 194,501 | 201,685 |
| 500 | 50,000 | 199,109 | 196,603 | 201,513 | 206,990 | 205,134 | 209,000 | 224,715 | 222,862 | 227,128 | 244,158 | 241,753 | 247,653 | 193,313 | 191,364 | 195,381 | 199,109 | 196,603 | 201,513 | 197,534 | 194,501 | 201,685 |
| 1,000 | 100,000 | 197,534 | 194,501 | 201,685 | 206,990 | 205,134 | 209,000 | 224,715 | 222,862 | 227,128 | 244,158 | 241,753 | 247,653 | 193,313 | 191,364 | 195,381 | 199,109 | 196,603 | 201,513 | 197,534 | 194,501 | 201,685 |

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

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

4.3 Dependence

One way to examine the role of dependence is to decompose the comprehensive coverage into more “primitive” coverages for the three types of claims (third party injury, own damage and third party property). As in derivative securities, we call this “unbundling” of coverages. We are able to calculate risk measures for each unbundled coverage, as if separate financial institutions owned each coverage, and compare them to risk measures for the bundled coverage that the insurance company is responsible for. The results are shown in Table 9. For the bundled (comprehensive) coverage, the VaR and CTE are from the first row of Tables 7 and 8, respectively, for policies without coverage modifications.

As anticipated, the risk measures for bundled coverages are smaller than the sum of unbundled coverages, for both risk measures and all percentiles. One of the important purposes of risk measures is to determine economic capital which is the amount of capital that banks and insurance companies set aside as a buffer against potential losses from extreme risk event. The implication of this example is that by bundling different types of coverage into one comprehensive policy, the insurers can reduce the economic capital for risk management or regulatory purpose. Another perspective is that this example simply demonstrates the effectiveness of economies of scales; three small financially independent institutions (one for each coverage) require in total more capital than a single combined institution (one for the bundled coverage). The interesting thing is that this is true even though the dependencies among the three coverages are positive, as shown in Appendix A1.

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

| Unbundled Coverages | VaR | | | CTE | | |
|----------------------------------|---------|---------|-----------|---------|-----------|-----------|
| | 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 |

The statistical model described in Section 2.3 with parameter estimates in Appendix A1 show strong significance evidence of positive relations among the three coverage types, third party injury, own damage and third party property. However, the model is complex, using copulas to assess this nonlinear dependence. How important is the dependence for the financial risk measures? To quantify this issue, Table 10 shows the VaR and CTE for different copula models. The independence copula comes from treating the three lines as unrelated, the normal copula captures the correlation among different coverages within a comprehensive policy compared with independence copula, whereas the t -copula captures the heavy-tail features of the risk compared to the normal copula. As a sensitivity analysis, we incorporated the copulas in two ways. In the top portion of the table, we assumed that the specified copula was consistently used for the estimation and prediction portions.

For the bottom portion, we assumed that the (correct but more complex) t -copula was used for estimation with the specified copula used for prediction. The idea behind the bottom portion was that a statistical analysis unit of a company may perform a more rigorous analysis using a t -copula and another unit within a company may wish to use this output for quicker calculations about their financial impact.

| Table 10. <i>VaR</i> and <i>CTE</i> for Bundled Coverage by Copula | | | | | | |
|--|---------|---------|-----------|---------|-----------|-----------|
| Copula | VaR | | | CTE | | |
| | 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 |

Table 10 shows that the copula effect is large and increases with the percentile. Of course, the upper percentiles are the most important to the actuary for many financial implications.

Table 10 demonstrates a large difference between assuming independence among coverages and using a t -copula to quantify the dependence. We found, when re-estimating the full model under alternative copulas, that the marginal parameters changed to produce significant differences in the risk measures. Intuitively, one can think of estimation and prediction under the independence copula to be similar to “unbundled” coverages in Table 9, where we imagine separate financial institutions accepting responsibility for each coverage. In one sense, the results for the independence copula are somewhat counterintuitive. For most portfolios, with positive correlations among claims, one typically needs to go out further in the tail to achieve a desired percentile, suggesting that the *VaR* should be larger for the t -copula than the independence copula.

To reinforce these findings, the bottom half of the table reports results when the marginal distributions are unchanged, yet the copula differs. Table 10 shows that the *VaR* is not affected by the choice of copula; the differences in Table 10 are due to simulation error. In contrast, for the *CTEs*, the normal and t -copula give higher values than the independence copula. This result is due to the higher losses in the tail under the normal and t -copula models. Although not displayed here, we re-ran this portion of the analysis with 50,000 simulation replications (in lieu of 5,000) and verified these patterns.

When applied to economic capital, these results indicate that the independence copula leads to more conservative risk measures while the t -copula leads to more aggressive risk measures. To determine which model to be used to calculate the economic capital may depend on the purpose of the capital, either for interval, risk management or regulatory use. It may also depend on the trade-off among model simplicity, estimation accuracy and computational complexity.

5 Predictive Distributions for Reinsurance

Reinsurance, an important risk management tool for property and casualty insurers, is another area of application where predictive distributions can be utilized. We examine two types of reinsurance agreements: quota share reinsurance and excess of loss reinsurance. In addition, we investigate a simple portfolio reinsurance. We simulate the number of accident, type of losses and severity of losses, and then allocate the losses between insurer and reinsurer according to different reinsurance agreements.

Quota share reinsurance is a form of proportional reinsurance which specifies that a fixed percentage of each policy written will be transferred to the reinsurer. The effect of different quota shares on the retained claims for the ceding company is examined and presented in Figure 4. The distributions of retained claims are derived assuming the insurer retains 25%, 50%, 75% and 100% of its business. In Figure 4 a quota of 0.25 means the insurer retains 25% of losses and cedes 75% to the reinsurer. The curve corresponding to quota of 1 represent the losses of insurer without a reinsurance agreement, as in Figure 2. As we can see, the quota share reinsurance does not change the shape of the retained losses, only the location and scale. For example, if the insurer ceded 75% of losses, the retained losses will shift left and the variance of retained losses will be $1/16$ times the variance of original losses. We did not present the distribution of losses for reinsurer, because under quota share reinsurance, the insurer and reinsurer shares the losses proportionally.

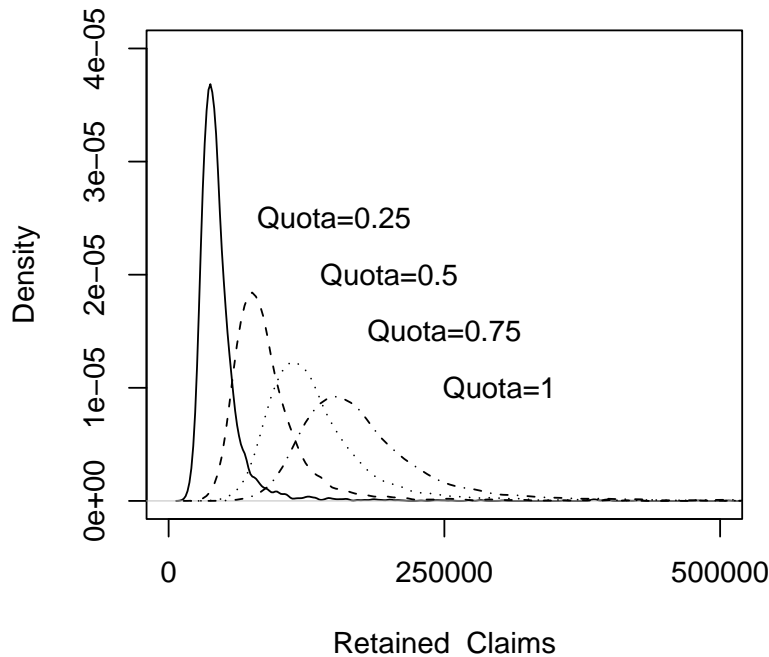


Figure 4: Distribution of Retained Claims for the Insurer under Quota Share Reinsurance. The insurer retains 25%, 50%, 75% and 100% of losses, respectively.

Excess of loss is a nonproportional reinsurance under which the reinsurer will pay the ceding company for all the losses above a specified dollar amount, the retention limit. The retention limit is similar to the deductible in a primary policy, the reinsurer will assume all the losses above it. Figure 5 shows the effect of different retention limits on the losses of insurer and reinsurer. Losses are simulated and limits of 5,000, 10,000 and 20,000 per policy are imposed. Unlike quota share arrangements, the retention limit changes the shape of the distribution for both the insurer and reinsurer. The lower the retention limit, the more business the insurer cedes so that losses for insurer become less skewed with thinner tails because the losses in the tail of distribution become the responsibility of reinsurer. Correspondingly, as the retention limit decreases, the distribution of losses for the reinsurer exhibits fatter tails. This is because the reinsurer retains a larger portion of the claim.

Figures 4 and 5 provide insights on how the various types of reinsurance agreement will affect their risk profile of the insurer and reinsurer. Through such analyses, the insurer can choose proper forms of reinsurance to manage its risk portfolio, and the reinsurer can decide upon the amount of risk to be underwritten. In practice, there are many other types of reinsurance contracts.

The above analysis focused on the reinsurance agreements where reimbursements are based on

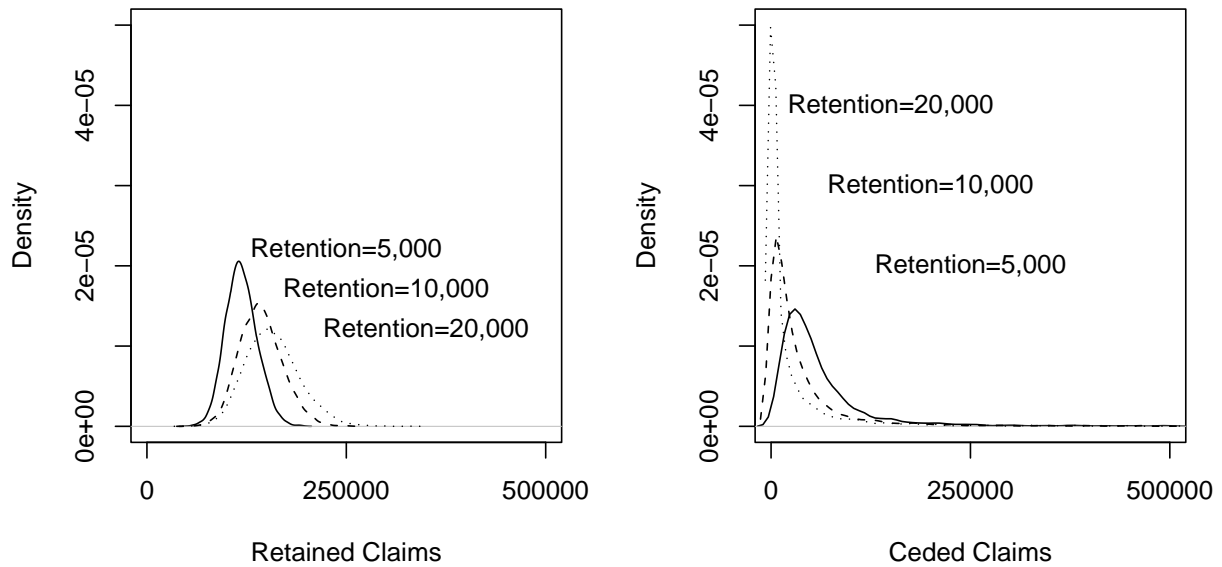


Figure 5: Distribution of Losses for the Insurer and Reinsurer under Excess of Loss Reinsurance. The losses are simulated under different primary company retention limits. The left-hand panel is for the insurer and right-hand panel is for the reinsurer.

losses for each policy. As another example, we consider a case where both policy reinsurance and portfolio reinsurance are involved. Portfolio reinsurance generally refers to the situation where the reinsurer is insuring all risks in a portfolio of policies, such as a particular line of business. For the purpose of illustration, we still consider the randomly selected portfolio which consists of 1000 policies from held-out-sample in year 2001. We now assume that there is a limit for the entire portfolio of business, the reinsurer will assume all the losses exceeding the portfolio limit. In addition, both quota share and policy limit have been applied to the individual policy in the portfolio. The combined effect of these coverage modifications on the risk profile for insurer and reinsurer are investigated and the results are presented in Table 11.

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 |

Table 11 presents percentiles of losses for the insurer and reinsurer under various reinsurance arrangements. As before, we see the long tail nature in the total losses of the portfolio and the interactive effect of coverage modifications on the claim distribution. In the first three rows where there is no policy retention limit, the insurer and reinsurer share the losses proportionally before the total losses reach the portfolio limit; that is why the 25th percentile for the insurer is equal to the 75th percentile for the reinsurer. When the total losses reach portfolio limit, the amount above it should be ceded to reinsurer. The second three rows for insurer and reinsurer shows the interactive effect of policy and portfolio retention limits. Both limits reduce the heavy tailed nature of losses for the insurer. Under a policy retention limit of 10,000, a greater portfolio retention limit has less effect on changing the tail behavior of losses for insurer. However the effect of portfolio retention limit (say 200,000) depends on the policy limit in the sense that the effect will be bigger under a greater policy limit (20,000). The last four rows of percentiles for both insurer and reinsurer shows the combined effect of coinsurance, policy limit and portfolio limit.

Table 11 provides useful information about the distribution of potential losses for insurer and reinsurer. This can help the ceding company understand the risk characteristics of retained business. The insurer can choose appropriate reinsurance agreements including setting proper policy retention limit, selecting the right quota and deciding the aggregate stop loss, to manage the risk more efficiently. The results also can be helpful in determine the reinsurance premium and to help reinsurer to assess the risk of the business they assumed from ceding company.

6 Summary and Conclusions

This paper follows our prior work (Frees and Valdez, 2008) where a statistical hierarchical model was introduced using detailed, micro-level automobile insurance records. In this paper, we demonstrate the financial implications of the statistical modeling.

We examined three types of applications that commonly concern actuaries. The first was individual risk rating. We examined the effect of coverage modifications including deductibles, coverage limits and coinsurance. We showed how to apply our techniques to rating “unbundled” coverages, very much akin to financial derivatives. We examined both analytic and simulated means.

Our second type of application dealt with estimating financial risk measures for portfolios of policies. In this paper we focused on the value at risk, VaR , and conditional tail expectation, CTE , although our approach could be easily extended to other risk measures. We assessed the effects of some of statistical assumptions, such as the copula, on these measures.

The third type of application was to provide predictive distributions for reinsurance. We examined the combined effect of coverage modifications for policies and portfolios on the claim distributions of the insurer and reinsurer, by examining the tail summary for the losses of portfolio generated using Monte Carlo simulation.

This paper demonstrates some of the interesting financial analyses of concern to actuaries

that can be accomplished with detailed micro-level data and advanced statistical models. With simulation, we can easily calculate predictive distributions for several financial risk measures. A limitation of this paper is that we did not explicitly incorporate estimation error into our predictive distributions (see, for example, Cairns (2000)). On one hand, one might argue that we had many observations available for estimation and that the resulting standard errors for the statistical model were inherently small. On the other hand, one could say that the statistical model incorporates many parameters whose joint uncertainty should be accounted for. We view this as an interesting area for future research.

References

- Angers, Jean-François, Denise Desjardins, Georges Dionne and Francois Guertin (2006). Vehicle and fleet random effects in a model of insurance rating for fleets of vehicles. *ASTIN Bulletin* 36(1), 25-77.
- Antonio, Katrien, Jan Beirlant, Tom Hoedemakers and Robert Verlaak (2006). Lognormal mixed models for reported claims reserves. *North American Actuarial Journal* 10(1), 30-48.
- Boucher, Jean-Philippe and Michel Denuit (2006). Fixed versus random effects in Poisson regression models for claim counts: A case study with motor insurance. *ASTIN Bulletin* 36(1), 285-301.
- Brockman, M.J. and T.S. Wright. (1992). Statistical motor rating: making effective use of your data. *Journal of the Institute of Actuaries* 119, 457-543.
- Cairns, Andrew J. G. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* 27 (3), 313-330.
- Coutts, S.M. (1984). Motor insurance rating, an actuarial approach. *Journal of the Institute of Actuaries* 111, 87-148.
- Desjardins, Denise, Georges Dionne and Jean Pinquet (2001). Experience rating schemes for fleets of vehicles. *ASTIN Bulletin* 31(1), 81-105.
- Frangos, Nicholas E. and Spyridon D. Vrontos (2001). Design of optimal bonus-malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bulletin* 31(1), 1-22.
- Frees, Edward W. and Emiliano A. Valdez (1998). Understanding relationships using copulas. *North American Actuarial Journal* 2(1), 1-25.
- Frees, Edward W. and Emiliano A. Valdez (2008). Hierarchical insurance claims modeling. To appear in *Journal of the American Statistical Association*. Available at <http://research3.bus.wisc.edu/course/view.php?id=129>.
- Gourieroux, Christian and Joann Jasiak (2007). *The Econometrics of Individual Risk: Credit, Insurance, and Marketing*. Princeton University Press, Princeton, NJ.
- Hsiao, Cheng, Changseob Kim and Grant Taylor (1990). A statistical perspective on insurance rate-making. *Journal of Econometrics* 44, 5-24.
- Kahane, Yehuda and Haim Levy (1975). Regulation in the insurance industry: determination of premiums in automobile insurance. *Journal of Risk and Insurance* 42, 117-132.

- Klugman, Stuart, Harry Panjer and Gordon Willmot (2004). *Loss Models: From Data to Decisions* (Second Edition), Wiley, New York.
- McDonald, James B. and Yexiao J. Xu (1995). A generalization of the beta distribution with applications. *Journal of Econometrics* 66, 133-152.
- Pinquet, Jean (1997). Allowance for cost of claims in bonus-malus systems. *ASTIN Bulletin* 27(1), 33-57.
- Pinquet, Jean (1998). Designing optimal bonus-malus systems from different types of claims. *ASTIN Bulletin* 28(2), 205-229.
- Renshaw, Arthur E. (1994). Modeling the claims process in the presence of covariates. *ASTIN Bulletin* 24(2), 265-285.
- Sun, Jiafeng, Edward W. Frees and Marjorie A. Rosenberg (2008). Heavy-tailed longitudinal data modeling using copulas. To appear in *Insurance: Mathematics and Economics*. Available at <http://research3.bus.wisc.edu/course/view.php?id=129>.
- Terza, Joseph V. and Paul W. Wilson (1990). Analyzing frequencies of several types of events: A mixed multinomial-Poisson approach. *The Review of Economics and Statistics* 108-115.
- Weisberg, Herbert I. and Thomas J. Tomberlin (1982). A statistical perspective on actuarial methods for estimating pure premiums from cross-classified data. *Journal of Risk and Insurance* 49, 539-563.
- Weisberg, Herbert I., Thomas J. Tomberlin and Sangit Chatterjee (1984). Predicting insurance losses under cross-classification: A comparison of alternative approaches. *Journal of Business & Economic Statistics* 2(2), 170-178.

AUTHOR INFORMATION:

Edward W. Frees

*School of Business
University of Wisconsin
Madison, Wisconsin 53706 USA
e-mail: jfrees@bus.wisc.edu*

Peng Shi

*School of Business
University of Wisconsin
Madison, Wisconsin 53706 USA
e-mail: pshi@bus.wisc.edu*

Emiliano A. Valdez

*Department of Mathematics
College of Liberal Arts and Sciences
University of Connecticut
Storrs, Connecticut 06268 USA
e-mail: valdez@math.uconn.edu*

A Appendix - The Predictive Model

The parameter estimates corresponding to each component of the hierarchical model are provided in Appendix A.1. These results are not directly comparable with Frees and Valdez (2008) because we selected a different Singaporean company. By examining data from a different company, we provide further validity of the model's robustness. Appendix A.2 describes the simulation procedure, on which the simulations throughout this work are based. Appendix A.3 provides several miscellaneous results to supplement the paper.

A.1 Parameter Estimates of Hierarchical Model

The parameter estimates for hierarchical predictive model are presented in Tables A.1, A.2 and A.4 for the frequency, type and severity components, respectively. The risk factors are based on those described in Table 1. For example, "vehicle age 3-5" in Table A.1 means that the insured's vehicle is 3, 4 or 5 years old, while the "vehicle age ≤ 5 " in Table A.4 means that the age is less than or equal to 5. A "*" signifies the interaction between the two risk factors. For example, "automobile*NCD 10" in Table A.1 indicates an insured who owns a private car with NCD equal to 10, and "automobile*NCD 0-10" in Table A.4 refers to a policyholder who has a private car with NCD equal to 0 or 10.

In Table A.2 we did not provide the standard errors to show the statistical significance of each parameter. Instead, we provide chi-square tests in Table A.3. In the multilogit model, one is concerned whether a covariate significantly differentiates the claim type probability across each category of M , not the significance of parameters within each category. In Table A.3, "vehicle age" is divided into two categories, less than 5 and equal or greater than 6. "automobile*age" represents three categories, non private car, private car with insured's age less than 45, private car with insured's age equal or greater than 46. Table A.3 shows that all variables are statistically significant.

| 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 |

| Category(M) | Parameter Estimates | | | | |
|-----------------|---------------------|--------|---------------------|----------------|-------------------------|
| | intercept | year | vehicle age $\gg 6$ | non-automobile | automobile*age $\gg 46$ |
| 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 |

| Source | DF | Chi-Square | Pr > ChiSq |
|------------------|-----|------------|------------|
| intercept | 6 | 2225.37 | <.0001 |
| year | 6 | 59.80 | <.0001 |
| vehicle age | 6 | 101.72 | <.0001 |
| automobile*age | 12 | 444.04 | <.0001 |
| Likelihood Ratio | 258 | 268.30 | 0.3168 |

| Parameter | Types of Copula | | | | | |
|-----------------------------|-----------------|----------------|---------------|----------------|------------------|----------------|
| | Independence | | Normal Copula | | <i>t</i> -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.2 Simulation Procedure

The simulation procedure used in this paper can be described in terms the three component hierarchical model.

We start the simulation for the frequency component of the hierarchical predictive model. The number of accidents N_i for policyholder i follows negative binomial distribution described in Section 2.3.1. We generate N_i using the probabilities $\Pr(N_i = k) = \binom{k+r-1}{r-1} p_i^r (1-p_i)^k$.

Then we generate the type of losses given an accident by simulating claim type variable M_i for each policyholder from the distribution $\Pr(M_i = m) = \frac{\exp(V_m)}{\sum_{s=1}^7 \exp(V_s)}$, which is described in Section 2.3.2.

Finally, we simulate the trivariate GB2 distribution, as follows:

- Generate (t_1, t_2, t_3) from a trivariate t -distribution using $\mathbf{t} = \mathbf{s}/\sqrt{W}$, where \mathbf{s} has a multivariate normal distribution with variance-covariance matrix Σ and W , independent of \mathbf{s} , follows a chi-square distribution with ν degrees of freedom.

- Generate (u_1, u_2, u_3) from the t -copula using $u_k = G_\nu(t_k)$, $k = 1, 2, 3$ where G_ν is the distribution function for a t -distribution with ν degrees of freedom.

- Calculate q_k , the u_k th percentile of a beta distribution with parameters α_{1k} and α_{2k} , $k = 1, 2, 3$. Here, α_{1k} and α_{2k} represent the shape parameters for a type k loss.

- Generate a realization of the trivariate GB2 distribution using

$$z_k = \exp(\mu_k)(q_k/(1-q_k))^{\sigma_k}, k = 1, 2, 3$$

where μ_k is the location parameter for a type k loss, defined by equation (6), and σ_k represents the scale parameter for a type k loss.

After generating the three components of the predictive model, we can calculate the simulated losses of the three type from j th accident for policy holder i :

$$y_{ij1} = z_{ij1}(\mathbf{1}(M_i = 1) + \mathbf{1}(M_i = 4) + \mathbf{1}(M_i = 5) + \mathbf{1}(M_i = 7))$$

$$y_{ij2} = z_{ij2}(\mathbf{1}(M_i = 2) + \mathbf{1}(M_i = 4) + \mathbf{1}(M_i = 6) + \mathbf{1}(M_i = 7))$$

$$y_{ij3} = z_{ij3}(\mathbf{1}(M_i = 3) + \mathbf{1}(M_i = 5) + \mathbf{1}(M_i = 6) + \mathbf{1}(M_i = 7))$$

Incorporating the number of accidents, we have the claims of each type from a single policy:

$$S_{i,k} = y_{ijk} \sum_{n=0}^{\infty} \mathbf{1}(N_i > n), k = 1, 2, 3$$

and the total losses from a policy:

$$S_i = S_{i,1} + S_{i,2} + S_{i,3}$$

Also we can calculate the losses from a portfolio of policies:

$$S = \sum_{i=1}^m S_i$$

where m represent the size of this portfolio.

A.3 Supplemental Results

A.3.1 Predictive Means

To understand how the covariates affect the predictive mean and lead to the pattern shown in Tables 3 and 4, we need to ask the question: To what extent are these overall results driven by the frequency versus the severity models?

Table A.1 shows that an insured with an private car and higher NCD would be less likely to have an accident given other covariates held constant. The probability of having no accident is presented in Table A.5. In addition, NCD affects the severity of own damage losses. Tables A.4 and A.5 show that an insured with a NCD less than 20 is expected to have losses of own damage which is $\exp(0.087)$ times the losses of own damage for those with NCD equal or higher than 20, given an accident occurred.

| Table A.5. Effect of NCD on Analytic Predictive Mean | | | | | | |
|--|--------|--------|--------|--------|--------|--------|
| NCD | 0 | 10 | 20 | 30 | 40 | 50 |
| Probability of no accident under various NCD | | | | | | |
| No accident | 0.916 | 0.924 | 0.928 | 0.929 | 0.935 | 0.942 |
| Expected losses under various NCD | | | | | | |
| Third party injury | 10.669 | 10.669 | 10.669 | 10.669 | 10.669 | 10.669 |
| Own damage | 2.532 | 2.532 | 2.320 | 2.320 | 2.320 | 2.320 |
| Third party property | 2.765 | 2.765 | 2.765 | 2.765 | 2.765 | 2.765 |

NCD affects the probability of loss type and severity of loss, while insured's age affects probability of accident in addition to loss type probability and loss severity.

The pattern for Insured's Age is more complicated. Table A.6 shows the probability of no accident for insured in different age categories. The results indicate that younger driver tends to have more accident than older one assuming other variables are the same, except for the person in highest age category. In addition, the younger($\text{age} \leq 25$) and older($\text{age} \geq 56$) driver tends to incur higher expected losses of own damage, specifically $\exp(0.037)$ and $\exp(0.121)$ times that of middle-aged driver respectively, which can be derived from Table A.4. Different from NCD, age category also affects the probability of type, which is shown in Table A.6. The insured greater than 45 have is more likely to incur losses of third party injury and own damage, while less likely to incur losses of third party property. The combination of three components explains the common pattern for the four random variables: for the insured less than 46, the younger has greater predictive mean, while for those greater 45, the older has greater predictive mean. The difference in the pattern for the four random variables is that for individual losses of own damage and sum of losses from own damage, the predictive mean of insured between 36 and 45 is less than that of insured between 46 and 55, while the relationship of predictive mean for insured in these two age categories is opposite for sum of losses from a specific even and overall limit per policy. This is explained by noticing that the expected losses from own damage is smallest, so when calculating the total losses from three types, the influence of own damage is diluted.

| Table A.6. Effect of Age Category on Analytic Predictive Mean | | | | | | | |
|---|-----------|--------|--------|--------|--------|--------|-----------|
| Age | ≤ 21 | 22-25 | 26-35 | 36-45 | 46-55 | 56-65 | ≥ 66 |
| Probability of no accident under various age category | | | | | | | |
| No accident | 0.912 | 0.920 | 0.933 | 0.940 | 0.942 | 0.946 | 0.945 |
| Probability of losses type under various age category | | | | | | | |
| Third party injury | 0.027 | 0.027 | 0.027 | 0.027 | 0.031 | 0.031 | 0.031 |
| Own damage | 0.686 | 0.686 | 0.686 | 0.686 | 0.870 | 0.870 | 0.870 |
| Third party property | 0.408 | 0.408 | 0.408 | 0.408 | 0.277 | 0.277 | 0.277 |
| Expected losses under various age category | | | | | | | |
| Third party injury | 10.669 | 10.669 | 10.669 | 10.669 | 10.669 | 10.669 | 10.669 |
| Own damage | 2.407 | 2.407 | 2.320 | 2.320 | 2.320 | 2.618 | 2.618 |
| Third party property | 2.765 | 2.765 | 2.765 | 2.765 | 2.765 | 2.765 | 2.765 |

A.3.2 Actual Losses Distribution

Figure 6 shows the density for actual losses in year 2001. The losses show skewed and heavy tailed property which is consistent with Figure 2. This in some sense also justified the choice of GB2 for each claim type. As for the randomly selected portfolio in Section 4, the actual losses for the portfolio is 299755.33, which is corresponding to the 94th percentile of the distribution in Figure 2.

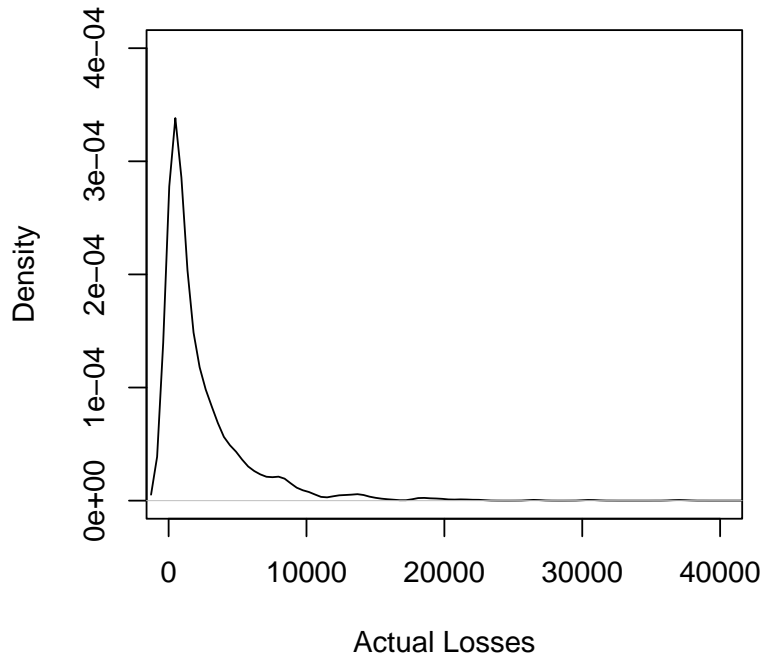


Figure 6: Distribution of Total Losses as of Year 2001 for the Insurance Company.

The actual losses distributions by claim type as of year 2001 for the insurance company are presented in Figure 7. All three types of losses present the heavy tailed nature in some extent. The losses of own damage have heaviest tail and the losses of third party property have largest peak.

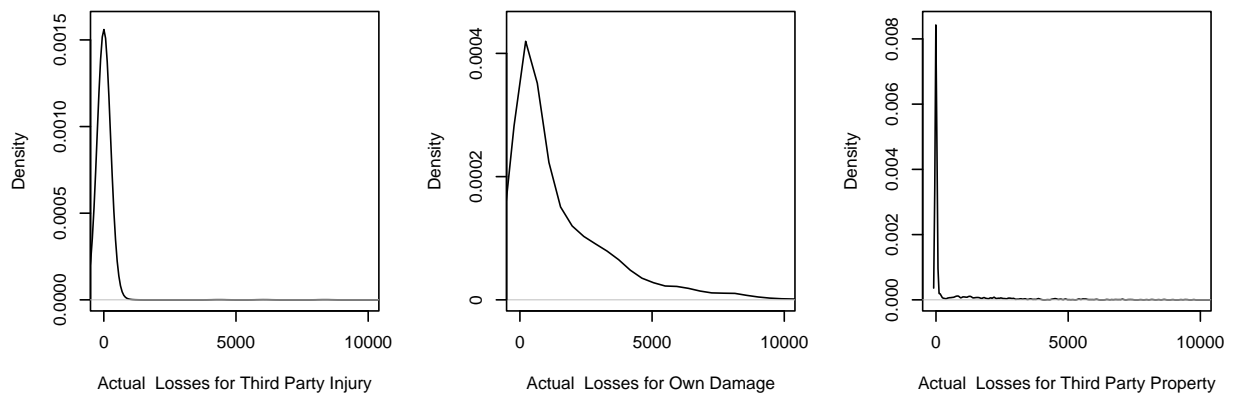


Figure 7: Distribution of Losses by Claim Type as of Year 2001 for the Insurance Company. The three panels are corresponding to the losses for third party injury, own damage and third party property respectively.