The Casualty Actuarial Society *Forum*
Winter 2006 Edition

To CAS Members:

This is the Winter 2006 Edition of the Casualty Actuarial Society *Forum*. It contains three Data Management Call Papers, two Research Working Party reports, and six additional papers.

The *Forum* is a nonrefereed journal printed by the Casualty Actuarial Society. The CAS is not responsible for statements or opinions expressed in the papers in this publication.

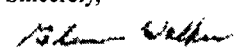
The CAS *Forum* is edited by the CAS Committee for the Casualty Actuarial Society *Forum*. Members of the committee invite all interested persons to submit papers on topics of interest to the actuarial community. Articles need not be written by a member of the CAS, but the paper's content must be relevant to the interests of the CAS membership. Members of the Committee for the Casualty Actuarial Society *Forum* request that the following procedures be followed when submitting an article for publication in the *Forum*:

1. Authors must submit papers in conformance with the research paper template established by the CAS. See the CAS Web Site at http://www.casact.org/research/template.htm for formatting instructions. Papers not conforming to the template will not be accepted.
2. Authors should submit a camera-ready original paper and two copies.
3. Authors should not number their pages.
4. All exhibits, tables, charts, and graphs should be in original format and camera-ready.
5. Authors should format graphs, tables, exhibits, and text in solid black and white. Avoid using color or gray-shaded graphs, tables, or exhibits as the printing process will produce muddy and moire patterns.
6. Authors should submit an electronic file of their paper using a popular word processing software (e.g., Microsoft Word and WordPerfect) for inclusion on the CAS Web Site.

The CAS *Forum* is printed periodically based on the number of call paper programs and articles submitted. The committee publishes two to four editions during each calendar year.

All comments or questions may be directed to the Committee for the Casualty Actuarial Society *Forum*.

Sincerely,



Glenn M. Walker, CAS *Forum* Chairperson

The 2006 CAS Data Management Call Papers
Presented at the
2006 Ratemaking Seminar
March 13-14, 2006
Marriott Salt Lake City Downtown
Salt Lake City, Utah
and Research Working Party Reports

The Winter 2006 Edition of the CAS *Forum* is a cooperative effort of the for the CAS *Forum* Committee, the Committee on Management Data and Information, and two CAS Research Working Parties: Correlations and Dependencies Among All Risk Sources, and Risk Transfer Testing.

The Data Management Call Papers will be presented at the 2006 CAS Ratemaking Seminar, March 13-14, 2006, in Salt Lake City, Utah.

*Non-CAS member of the committee

Data Management Call Papers

CAS Research Working Party Reports

Additional Papers

Distinguishing the Forest from the TREES: A Comparison of Tree Based Data Mining Methods

Richard Derrig, Ph.D. and Louise Francis, FCAS, MAAA

Richard Derrig, PhD,
OPAL Consulting LLC
41 Fosdyke Street
Providence
Rhode Island, 02906, U.S.A.
Phone: 001-401-861-2855
email: richard@derrig.com

Louise Francis, FCAS, MAAA
Francis Analytics & Actuarial Data Mining
706 Lombard Street
Philadelphia
Pennsylvania, 19147, U.S.A.
Phone: 001-215-923-1567
email: louise_francis@msn.com

Abstract

In recent years a number of "data mining" approaches for modeling data containing nonlinear and other complex dependencies have appeared in the literature. One of the key data mining techniques is decision trees, also referred to as classification and regression trees or CART (Breiman et al, 1993). That method results in relatively easy to apply decision rules that partition data and model many of the complexities in insurance data. In recent years considerable effort has been expended to improve the quality of the fit of regression trees. These new methods are based on ensembles or networks of trees and carry names like TREENET and Random Forest. Viaene et al (2002) compared several data mining procedures, including tree methods and logistic regression, for prediction accuracy on a small fixed data set of fraud indicators or "red flags". They found simple logistic regression did as well at predicting expert opinion as the more sophisticated procedures. In this paper we will introduce some available regression tree approaches and explain how they are used to model nonlinear dependencies in insurance claim data. We investigate the relative performance of several software products in predicting the key claim variables for the decision to investigate for excessive and/or fraudulent practices, and the expectation of favorable results from the investigation, in a large claim database. Among the software programs we will investigate are CART, S-PLUS, TREENET, Random Forest and Insightful Miner Tree procedures. The data used for this analysis are the approximately 500,000 auto injury claims reported to the Detailed Claim Database (DCD) of the Automobile Insurers Bureau of Massachusetts from accident years 1995 through 1997. The decision to order an independent medical examination or a special investigation for fraud, and the favorable outcomes of such decisions, are the modeling targets. We find that the methods all provide some predictive value or lift from the available DCD variables with significant differences among the methods and the four targets. All modeling outcomes are compared to logistic regression as in Viaene et al. with some model/software combinations doing significantly better than the logistic model.

Keywords: Fraud, Data Mining, ROC Curve, Variable Importance, Decision Trees

INTRODUCTION

In recent years a number of approaches for modeling data containing nonlinear and other complex dependencies have appeared in the literature. Many of the methods were developed by researchers from the computer science, artificial intelligence and statistics disciplines[1]. The methods have been widely characterized as *data mining* techniques. These procedures include several that should be of interest to actuaries dealing with large and complex data sets. The procedures of interest for the purposes of this paper are various varieties of classification and regression trees or CART. Viaene et al (2002) applied a wider set of procedures, including neural networks, support vector machines, and a classical general linear model, logistic regression, on a small single data set of insurance claim fraud indicators or "red flags" as predictors of suspicion of fraud. They found simple logistic regression did as well at predicting expert opinion on the presence of fraud as the more sophisticated procedures. Stated differently, the logistic model performed well enough in modeling the expert opinion of fraud that there was little need for the more sophisticated procedures[2].

A wide variety of statistical software is now available for implementing fraud and other predictive models through clustering and data mining. In this paper we will introduce a variety of Regression Tree data mining approaches[3] and explain how they are used to model nonlinear dependencies in insurance claim data. We also investigate the relative performance of several software products that implement these models. As an example of relative performance, we test for the key claim variables in the decision to investigate for excessive and/or fraudulent practices in a large claim database. The software programs we will investigate are CART, S-PLUS, TREENET, Random Forests, and Insightful Tree and Ensemble from the Insightful Miner package. Naïve Bayes and Logistic models are used as benchmarks. The data used for this analysis are the auto bodily injury liability claims reported to the Detailed Claim Database (DCD) of the Automobile Insurers Bureau of Massachusetts from accident years 1995 through 1997[4]. Three types of variables are employed. Several variables thought to be related to the decision to investigate are included here as reported to the DCD, such as outpatient provider medical bill amounts. A few variables are included that are derived from publicly available demographic data sources, such as income per household for each claimant's zip code. Additional variables are derived by accumulating proportional statistics from the DCD; e.g., the distance from the claimant's zip code to the zip code of the first medical provider or claimant's zip code rank for the number of plaintiff attorneys per zip code. The decision to order an independent medical examination or a special investigation for fraud, and a favorable outcome for each, are the modeling target.

Eight modeling software results will be compared for effectiveness based on a standard procedure, the area under the receiver operating characteristic curve (AUROC). We find that the methods all provide some predictive value or lift from the DCD variables we make available, with significant differences among the eight methods and four targets. Modeling outcomes can be compared to logistic regression as in Viaene et al. but the results here are different. They show some software/methods can improve significantly on the predictive

ability of the logistic model. That result may be due to the relative richness of this data set and/or the types of independent variables at hand compared to the Viaene data. We show how "important" each variable is within each software/model tested[5] and note the type of data that are important for this analysis. This entire exercise should provide practicing actuaries with guidance on regression tree software and market methods to analyze complex nonlinear relationship commonly found in all types of insurance data.

The paper is organized as follows. Section 1 introduces the general notion of non-linear dependencies in insurance data. Section 2 describes the data set of Massachusetts auto bodily injury liability claims and variables used for illustrating the models and software implementations. Descriptions and illustrations of the data mining methods applied in the paper appear in Section 3 while the specific software procedures are covered in Section 4. Comparative outcomes for the variables ("importance") and software ("AUROC") are reported in Sections 5 and 6. We provide some interpretation of the results in terms of the decision to investigate within the Massachusetts data as an illustration of the usefulness of the modeling effort in Section 7. Implications for the use of the software models are discussed in section 8. Conclusions are shown in Section 9.

SECTION 1. NONLINEARITY IN INSURANCE DATA

Actuaries are nearly inseparable from data and data manipulation techniques. Data come in all forms as a matter of course. Numeric (loss ratios), categorical (injury types), and text (accident description) data all flood insurers on a daily basis. Reserving and pricing are two major functions of casualty actuaries. Reserving involves compiling and understanding through mathematical techniques historical patterns of a portfolio of insurance claims in order to predict an ultimate value. Pricing involves taking the best estimates of historical cost data on claims and expenses, combining that data with financial asset pricing models that include projecting future values in order to arrive at best estimates of all costs of accepting underwriting risk. Of course, actuaries continually look back at both analytic exercises to determine the accuracy of those estimates as the real accounting data develops over time.

Traditionally, actuarial models were confined to linear, multiplicative or mixed algebraic equations in the absence of the powerful computing environment we enjoy today. Those mostly manual methods provided crude approximations that sufficed when alternative methods were unavailable or non-existent. Simple deviations from linear relationships, such as escalating inflation, could be handled by simple transformations of the data (log transform) that allowed linear techniques to be applied to the data. Gradually, over time these transformation techniques became more sophisticated and could be applied to many problems with a variety of non-linear data[6].

Trend lines of time series data, such as claim severity or frequency, are generally amenable to linear techniques. However, data where interactions and cross correlations are essential to the modeling of the dynamics of the process underlying the data, require more

comprehensive techniques that yield more precision on more types of data complexities. Figure 1-1 shows a particular non-linear relationship between two insurance variables that would be difficult, if not impossible, to model with simple techniques. One purpose of this paper is to demonstrate a range of so-called artificial intelligence or statistical learning techniques that have been developed to handle complicated relationships within data sets.

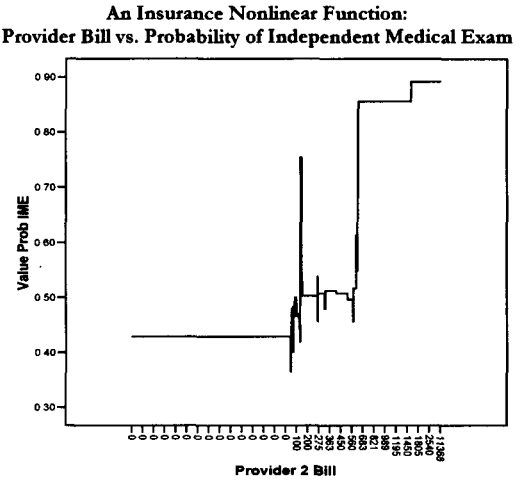


Figure 1 -1

Nearly all regression and econometric academic courses address the topic of nonlinearity, at least briefly. Students are instructed in methods to detect nonlinearity and how to model it. Detection generally involves using scatter plots of independent versus dependent variables or evaluating plots of residuals. Two methods of modeling nonlinearity that are generally taught: are 1) transformation of variables and 2) polynomial regression (Miller and Wichern[7], 1977, and Neter et al, 1985). For instance, if an examination of residual plots indicates that the magnitude of the residuals increases with the size of an independent variable, the log transformation is recommended. Polynomial regressions are considered useful approximations when a curvilinear relationship exists but its exact form is unknown.

A generalization of linear models known as Generalized Linear Models or GLM (McCullagh and Nelder, 1989) enabled the modeling of multivariate relationships in the presence of certain kinds of non-normality (i.e. where the random component is from the exponential family of distribution). The link function of GLMs formalizes the incorporation of certain nonlinear relationships into the modeling procedure: The transformations incorporated into the common GLMs are:

The identity link: $h(Y) = Y$

$$\text{The log link: } h(Y) = \ln(Y)$$
$$\text{The inverse link: } h(Y) = \frac{1}{Y} \qquad (1)$$
$$\text{The logit link: } h(Y) = \ln\left(\frac{Y}{1-Y}\right)$$
$$\text{The probit link: } h(Y) = \Phi(Y), \Phi \text{ denotes the normal CDF}$$

Of these transformations, the log and logit transformation appear frequently in the insurance literature. Because many insurance variables are right skewed, the log transformation is applied to attained approximate normality and homogeneity of variance. In addition, apriori or domain considerations (e.g., the relationship between the independent variables and the dependent variable is believed to be multiplicative) sometimes suggest the log transformation. The logit transform is commonly used when the dependent variable is binary.

Unfortunately, while the techniques cited above add significantly to the analyst's ability to model nonlinearity, they are not sufficient for many situations encountered in practice. In actual insurance data, complex nonlinear relationships are the rule rather than the exception. Some of the reasons the traditional approaches often do not provide a satisfactory approximation to nonlinear functions are:

- The form of the nonlinearity may be other than one of those permitted by the known transformations which produce linearity. Figure 1-1 displays one such non-linear function based on the insurance database used in this analysis.
- While a polynomial of adequate degree can approximate many complex functions, extrapolation beyond the data, or interpolation within the data, may be problematic, particularly for higher order polynomials.
- Determining the appropriate transformation (or polynomial) can be difficult if not impossible when there are many independent variables, and the appropriate relation between the target and each independent variable must be found.
- The relationship between a dependent variable and an independent variable may be confounded by a third variable due to interaction or correlations that are not simple to approximate.

To remedy these problems requires methods where:

- Any nonlinear relationship can be approximated.
- The analyst does not need to know the form of the nonlinearity.
- The effect of interactions can be easily determined and incorporated into the model.
- The method generalizes well on out-of-sample data for interpolation or extrapolation purposes.

The regression tree methods included in our analysis meet these conditions. Section 3 of this paper describes how each of our methods models nonlinearity. We now turn to a description of the data set we will use in this analysis.

SECTION 2. DESCRIPTION OF THE MASSACHUSETTS AUTO BODILY INJURY DATA

The database we will use for our analysis is a subset of the Automobile Insurers Bureau of Massachusetts Detail Claim Database (DCD); namely, those claims from accident years 1995-1997 that had closed by June 30, 2003 (AIB, 2004). All auto claims[8] arising from injury coverages: Personal Injury Protection (PIP)/ Medical payments excess of PIP[9], Bodily Injury Liability (BIL), Uninsured and Underinsured Motorist. While there are more than 500,000 claims in this subset of DCD data, we will restrict our analysis to the 162,761 third party BIL coverage claims. This will allow us to divide the sample into training, test, and holdout sub samples, each containing in excess of 50,000 claims[10]. The dataset contains fifty-four variables relating to the insured, claimant, accident, injury, medical treatment, outpatient medical providers (2 maximum), attorney presence, and three claims handling techniques for mitigating claims cost for their presence, outcome, and formulaic savings amounts.

The claims handling techniques tracked are: Independent Medical Examination (IME), Medical Audit (MA) and Special Investigation (SIU). IMEs are performed by licensed physicians of the same type as the treating physician[11]. They cost approximately $350 per exam with a charge of $75 for no shows. They are designed to verify claimed injuries and to evaluate treatment modalities. One sign of a weak or bogus claim is the failure to submit to an IME and, thus, an IME can serve as a screening device for detecting fraud and build-up claims. MAs are peer reviews of the injury, treatment and billing. They are typically done by physicians without a claimant examination, by nurses on insurers' staff or by third party organizations, but also from expert systems that review the billing and treatment patterns[12]. Favorable outcomes are reported by insurers when the damages are mitigated, the billing and treatment are curtailed, and when the claimant refuses to undergo the IME or does not show. In the latter two situations the insurer is on solid ground to reduce or deny payments under the failure-to-cooperate clause in the policy.[13]

Special Investigation (SIU) is reported when claims are handled through non-routine investigative techniques (accident reconstruction, examinations under oath and surveillance are examples), possibly including an IME or Medical Audit, on suspicion of fraud. For the most part, these claims are handled by Special Investigative Units (SIU) within the claim department or by some third party investigative service. Occasionally, companies will be organized so that additional adjusters, not specifically a part of the company SIU, may also conduct special investigations on suspicion of fraud. Both types are reported to DCD and we refer to both by the shorthand SIU in subsequent tables and figures. Favorable outcomes are reported for SIU if the claim is denied or compromised based on the SIU investigation.

For purposes of this analysis and demonstration of non-linear models and software, we employ twenty-one potentially predicting variables and four target variables. Thirteen predicting variables are numeric, two from DCD fields (F), eight derived from internal demographic type data (DV), and three variables derived from external demographic data (DM) as shown in Table 2-1.

Auto Injury Liability Claim Numeric Variables						
Variable	**N**	**Type**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
Provider 1_BILL	162,761	F	0	1,861,399	2,671.92	6,640.98
Provider 2_BILL	162,761	F	0	360,000	544.78	1,805.93
Age	155,438	DV	0	104	34.15	15.55
Report Lag	162,709	DV	0	2,793	47.94	144.44
Treatlag	147,296	DV	1	9	3.29	1.89
HouseholdsPerZipcode	118,976	DM	0	69,449	10,868.87	5,975.44
AverageHouseValue Per Zip	118,976	DM	0	1,000,001	166,816.75	77,314.11
IncomePerHousehold Per Zip	118,976	DM	0	185,466	43,160.69	17,364.45
Distance (MP1 Zip to CLT. Zip)	72,786	DV	0	769	38.85	76.44
Rankatt1 (rank att/zip)	129,174	DV	1	3,314	150.34	343.07
Rankdoc2 (rank prov/zip)	109,387	DV	1	2,598	110.85	253.58
Rankcity (rank claimant city)	118,976	DV	1	1,874	77.37	172.76
Rnkpcity (rank provider city)	162,761	DV	0	1,305	30.84	91.65
Valid N (listwise)	70,397					
N = Number of non missing records; F=DCD Field, DV = Internal derived variable, DM = External derived variable						

Source: Automobile Insurers Bureau of Massachusetts, Detail Claim Database, AY 1995-1997 and Authors' Calculations.

Table 2-1

Eight predicting variables, and four target variables (IME and SIU, Decision and Favorable Outcome for each), are categorical variables, all taken as reported from DCD and as shown in Table 2-2.

Auto Injury Liability Claim Categorical Variables			
Variable	**N Type**	**Type**	**Description**
Policy Type	162,761	F	Personal 92%, Commercial 8%
Emergency Treatment	162,761	F	None 9%, Only 22%, w Outpatient 68%
Health Insurance	162,756	F	Yes, 15%, No 26%, Unknown 60%
Provider 1 – Type	162,761	F	Chiro 41%, Physical Th. 19%, Medical 30%, None 10%
Provider 2 – Type	162,761	F	Chiro 6%, Physical Th. 6%, Medical 36%, None 52%
2001 Territory	162,298	F	Rating Territories 1 (2.2%) Through 26 (1.3%); Territory 1-16 by increasing risk, 17-26 is Boston
Attorney	162,761	F	Attorney present (89%), no attorney (11%)
Susp1 (SIU Done)	162,761	F	Special Investigation Done (7%), No SIU (93%)
Susp2 (IME Done)	162,761	F	Independent Medical Examination Done (8%), No IME (92%)
Susp3 (SIU Favorable)	162,761	F	Special Investigation Favorable (3.4%), Not Favorable/Not Done (95.6%)
Susp4 (IME Favorable)	162,761	F	Independent Medical Exam Favorable (4.4%), Not Favorable/Not Done (96.6%)
Injury Type	162,298	F	Injury Types (24) including minor visible (4%), strain or sprain, back and/or neck (81%), fatality (0.4%), disk herniation (1%) and others
N = Number of non missing records F= DCD Field			
Note: Descriptive percentages may not add to 100% due to rounding			

Source: Automobile Insurers Bureau of Massachusetts, Detail Claim Database, AY 1995-1997 and Authors' Calculations.

Table 2-2

Similar claim investigation variables are now being collected by the Insurance Research Council in their periodic sampling of countrywide injury claims (IRC, 2004a, pp 89-104)[14]. Nationally, about 4% and 2% of BI claims involved IMEs and SIU respectively, only one-half to one-quarter of the Massachusetts rate. Most likely, this is because (1) a majority of other states have a full tort system and so BIL contains all injury claims and (2) Massachusetts is a fairly urban state with high claim frequencies and more dubious claims[15]. In fact, the most recent IRC study shows (IRC, 2004b, p25) Massachusetts has the highest percentage of BI claims in no-fault states that are suspected of fraud (23%) and/or buildup (41%). It is therefore, entirely consistent for the Massachusetts claims to exhibit more non-routine claim handling techniques. Favorable outcomes average about 67% when an IME is done or a claim is referred to SIU. We now turn to descriptions of the types of models, and the software that implements them, in the next two sections before we describe how they are applied to model the IME and SIU target variables.

SECTION 3. MODELS FOR NON-LINEAR DEPENDENCIES

How models handle nonlinearity

Traditional actuarial and statistical techniques often assume that the functional relationship between the independent variables and the dependent variable is linear or that some transformation of the data exists that can be treated as linear. Insurance data often contain

variables where the relationship among variables is nonlinear. Typically when nonlinear relationships exist, the exact nature of the nonlinearity (i.e., where some transformation can be used to establish linearity) is not known. In the field of data mining, a number of nonparametric techniques have been developed which can model nonlinear relations without any assumption being made about the nature of the nonlinearity. We cover how each of our methods reviewed in this paper models nonlinearities in the following two examples. The variables in this example were selected because of a known nonlinear relationship between independent and dependent variables.

Ex. 1 The dependent variable, a numeric variable, is total paid losses and the independent variable is provider 2 bill. Table 3-1 displays average paid losses at various bands of provider 2 bill[16].

Ex. 2 The dependent variable, a binary categorical variable, is whether or not an independent medical exam is requested and the independent variable again is provider 2 bill.

Nonlinear Example Data

Provider 2 Bill (Banded)	Avg Provider 2 Bill	Avg Total Paid	Percent IME
Zero	-	9,063	6%
1 – 250	154	8,761	8%
251 – 500	375	9,726	9%
501 - 1,000	731	11,469	10%
1,001 - 1,500	1,243	14,998	13%
1,501 - 2,500	1,915	17,289	14%
2,501 - 5,000	3,300	23,994	15%
5,001 - 10,000	6,720	47,728	15%
10,001+	21,350	83,261	15%
All Claims	545	11,224	8%

Table 3-1

Trees

Trees, also known as classification and regression trees (CART) fit a model by recursively partitioning the data into two groups, one group with a higher value on the dependent variable and the other group with a lower value on the dependent variable. Each partition of the tree is referred to as a node. When a parent node is split, the two children nodes, or "leaves" of the tree, are each more homogenous (i.e., less variable) with respect the dependent variable[17]. A goodness of fit statistic is used to select the split which maximizes the difference between the two nodes. When the independent variable is numeric, such as provider 2 bill, the split takes the form of a cutpoint, or threshold: $x \geq c$ and $x < c$ as in Figure 3-1.

CART Example of Parent and Children Nodes
Total Paid as a Function of Provider 2 Bill

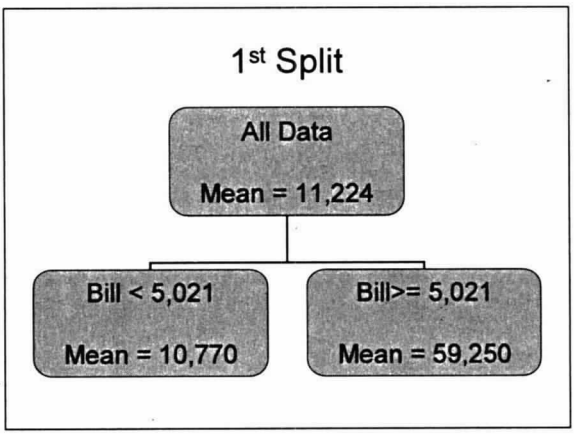


Figure 3-1

The cutpoint c is found by evaluating all possible values for splitting the numeric variable into higher and lower groups, and selecting the value that optimizes the split in some manner. When the dependent variable is numeric, the split is typically based on the value which results in the greatest reduction in residual sum of squares. For this example, all values of provider 2 bill are searched and a split is made at the value $5,021. All claims with provider 2 bill less than $5,021 go to the left node and "predict" a total paid of $10,770 and all claims with provider 2 bill greater than $5,021 go to the right node, and "predict" a total paid of $59,250. This is depicted in Figure 3-1. The tree graph shows that the total paid mean is significantly lower for the claims with provider 2 bills less than $5,021.

One statistic often used as a goodness of fit measure to optimize tree splits is sum squared error or the total squared deviation of actual values around the predicted values. The selected cutpoint is the one which produces the largest reduction in total sum squared errors (SSE). That is, for the entire database the total squared deviation of paid losses around the predicted value (i.e., the mean) of paid losses is 4.95×10^{13}. The SSE declines to 4.66×10^{13} after the data are partitioned using $5,021 as the cutpoint. Any other partition of the provider bill produces a larger SSE than 4.66×10^{13}. For instance, if a cutpoint of $10,000 is selected, the SSE is 4.76×10^{13}.

The two nodes in Figure 3-1 can each be split into to children nodes and these can then be further split. The sequential splitting continues until no improvement in the goodness of fit statistic occurs. The nodes containing the result of all the splits resulting from applying a sequence of decision rules are the final nodes often referred to as terminal nodes. The terminal nodes provide the predicted values of the dependent variables. When the dependent

variable is numeric, the mean of the dependent variable at the terminal nodes is the prediction.

The curve of the predicted value resulting from a tree fit to total paid losses is a step function. As shown in Figure 3-2A, with only two terminal nodes, the fitted function is flat until $5,021, steps up to a higher value and then remains flat. Figure 3-2B displays the predicted values of a tree with 7 terminal nodes. The steps or increases are more gradual for this function.

CART Example with Two and Seven Nodes
Total Paid as a Function of Provider 2 Bill

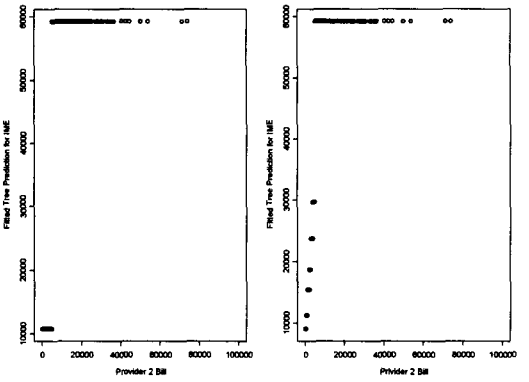


Figure 3-2A **Figure 3-2B**

The procedure for modeling data where the dependent variable is categorical (binary in our example) is similar to that of a numeric variable. For instance, one of the fraud surrogates is independent medical exam (IME) requested. The target class is claimants for whom an IME was requested and the non-target group of (presumably) legitimate claims is that where an IME was not requested. At each step, the tree procedure selects the split that best improves or lowers node impurity. That is, it attempts to partition the data into two groups so that one partition has a significantly higher proportion of the target category, IME requested, than the other node. A number of statistical goodness of fit statistics measures is used in different products to select the optimal split. These include entropy/deviance and Gini index (which is described later in this paper). Kantardzic (2003), Breiman et al (1993) and Venibles and Ripley (1999) describe the computation and application of the Gini index and entropy/deviance measures[18]. A score or probability can be computed for each node after a split is performed. This is generally estimated based on the number of observations in the target groups versus the total number of observations at the node. The score or probability

is frequently used to assign records to one of the two classes. Typically, if the model score exceeds a threshold such as 0.5, the record is assigned to the target class; otherwise it is assigned to the non-target class.

Figure 3-3A displays the result of using a tree procedure to predict a categorical variable from the AIB data. The graph shows that each time the data is split on provider 2 bill; one child node has a lower proportion and the other a higher proportion of claimants receiving IMEs. The fitted tree function is used to model a nonlinear relationship between provider bill and the probability that a claim receives an IME as shown in Figure 3-3B.

CART Example with Seven Nodes
IME Proportion as a Function of Provider 2 Bill

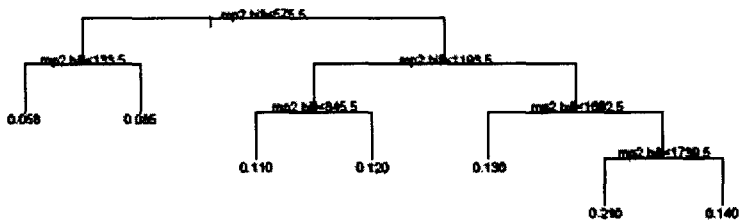


Figure 3-3A

CART Example with Seven Step Functions
IME Proportion as a Function of Provider 2 Bill

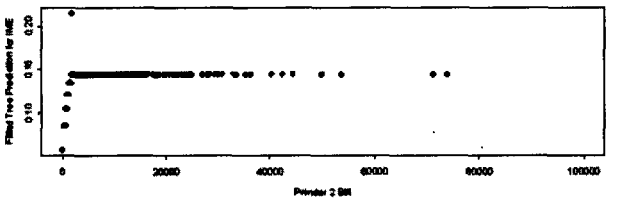


Figure 3-3B

Tree models use categorical as well as numeric independent variables in modeling complex data. However, because the levels on categorical data may not be ordered, all possible two-way splits of categorical variables must be considered before the data are partitioned.

Ensemble Models-Boosting

Ensemble models are composite tree models. A series of trees is fit and each tree improves the overall fit of the model. In the data mining literature the technique is often referred to as

"boosting" (Hastie et al 2001, Freidman, 2001). The method initially fits a small tree of say 5 to 10 terminal nodes on a training dataset. Typically, the user specifies the number of terminal nodes, and every tree fit has the same number of terminal nodes. The error, or difference between the actual and fitted values, is computed and used in another round of fitting as a dependent variable. The error is also used in the computation of the weight in subsequent rounds of fitting, with records containing larger errors receiving higher weighting in the next round of estimation.

One algorithm for computing the weight is described by Hastie et al[19]. Consider an ensemble of trees 1, 2, ...,M. The error for the m^{th} tree measures the departure of the actual from the fitted value on the test data after the m^{th} model has been fit. When the dependent variable is categorical, as it is in the fraud application in this paper, a common error measure used in boosting is:

$$err_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq F_m(\mathbf{x}_i))}{\sum_{i=1}^{N} w_i} \qquad (2)$$

where N is the total number of records, w_i is a weight (which is initialized to 1/N in the first round of fitting), I is an indicator function equal to zero if the category is correctly predicted and one if the class assigned is incorrect, y_i is the dependent variable, $\mathbf{x}$ is a matrix of predictors and $F_m(\mathbf{x})$ is the prediction for the i^{th} record of the m^{th} tree.

Then, the coefficient alpha is a function of the weight:

$$\alpha_m = \log(\frac{1-err_m}{err_n})$$

and the new weight is: (3)

$$w_{i,m+1} = w_m \exp(\alpha_m I(y_i \neq F_m(\mathbf{x})))$$

The process is performed many times until no further statistical improvement in the fit is obtained.

The specific boosting procedures implemented differ among different software products. For instance, TREENET (Freidman, 2001) uses stochastic gradient boosting. Stochastic gradient boosting incorporates a number of procedures which attempt to build a more robust model by controlling the tendency of large complex models to overfit the data. A key technique used is resampling. A new sample is randomly drawn from the training data each time a new tree is fit to the residuals from the prior round of model estimation. The goodness of fit of the model is assessed on data not included in the sample, the test data. Another procedure used by TREENET to control overfitting is *shrinkage* or *regularization*. A simple way to implement shrinkage is to apply a weight which is greater than zero and less than one to the contribution of each tree as it is added to the weighted average estimate.

Alternatively, the Insightful Miner Ensemble model employs a simpler implementation of boosting which applies non-stochastic boosting and uses all the training data in each round of fitting.

The final estimate resulting from an ensemble approach will be a weighted average of all the trees fit. Using a large collection of trees allows:

- Many different variables to be used. Some of these would not be used in smaller models[20].
- Many different models are used. The predictive modeling literature (Hastie et al., 2001, Francis, 2003a, 2003c) indicates that composites of multiple models perform better than the prediction of a single model[21].
- Different training and test records are used (with stochastic gradient boosting). This makes the procedure more robust to the influence of a few extreme observations.

The method of fitting many (often 100 or more) small trees results in fitted curves which are almost smooth. Figures 3-4A and 3-4B display two nonlinear functions fit to total paid and IME variables by the TREENET ensemble model.

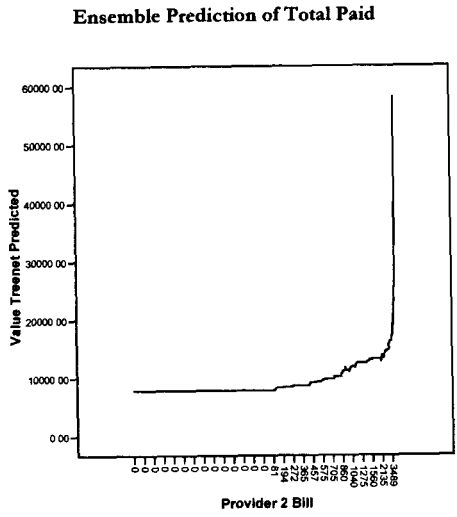


Figure 3-4A

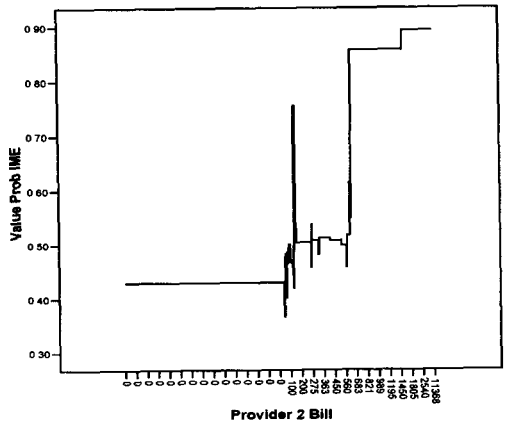


Figure 3-4B

Ensemble Models-Bagging

Bagging is an ensemble approach based on resampling or bootstrapping. Bagging is an acronym for "bootstrap aggregation" (Hastie et al., 2000). Bagging does not use the error from the prior round of fitting as a dependent variable or weight in subsequent rounds of fitting. Bagging uses recursive sampling of records in the data to fit many trees. For instance an analyst may decide to take a 50% of the data as a training set each time a model

is fit. Under bagging, 100 or more models may be fit, each one to a different sample. The trees fit are not necessarily small trees with 5 to 10 terminal nodes as with boosting and each tree may have a different number of terminal nodes. By averaging the predictions of a number of bootstrap samples, bagging reduces the prediction variance. The implementation of bagging used in this paper is known as Random Forest. In addition to using only a sample of the data each time a tree model is fit, Random Forest also samples from the variables. For the analysis in this paper, one third of the variables were sampled for each tree fit.

Figures 3-5A displays an ensemble Random Forest tree fit to total paid losses and Figure 3-5B displays a tree fit to IME.

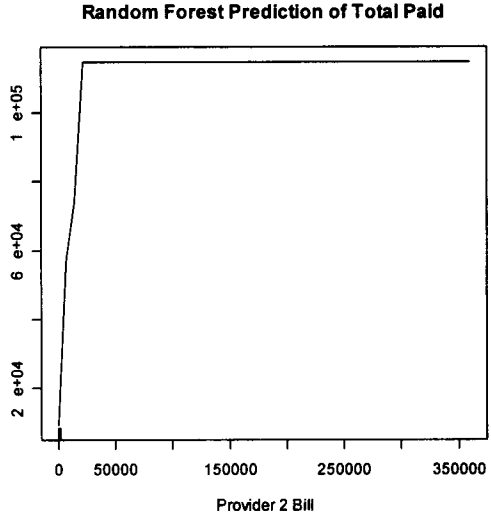


Figure 3-5 A

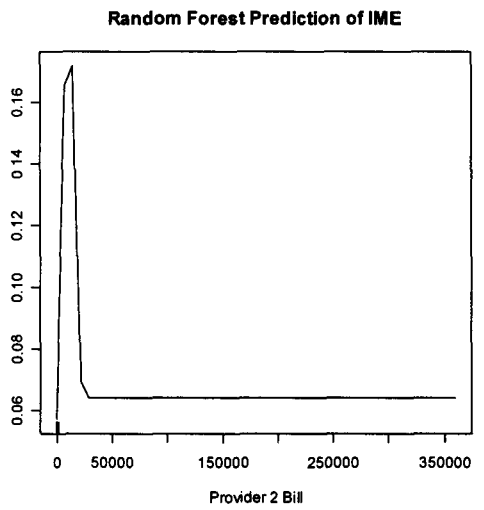


Figure 3-5 B

Naive Bayes

The Naïve Bayes method is a relatively simple and easy to implement method. In our comparison, we view it as a benchmark data mining method. That is, we are interested in how more complex methods improve performance (or not) against an approach where simplifying assumptions are made in order to make the computations more tractable. We also use logistic regression models as a second benchmark.

The Naïve Bayes method was developed for categorical data. Specifically, both dependent and independent variables are categorical. Therefore, its application to fitting nonlinear functions will be illustrated only for the categorical target variable IME. In order to utilize numeric predictor variables it was necessary to derive new categorical variables based on discretizing, or "binning", the distribution of data for the numeric variables[22].

The key simplifying assumption of the Naïve Bayes method is the assumption of independence. All predictor variables are assumed to act independently in influencing the target variable. Interactions and correlations among the predictor variables are not considered.

Bayes rule is used to estimate the probability that a record with given independent variable vector $X = \{x_i\}$ is in category $C = \{c_j\}$ of the dependent variable.

$$P(c_j \mid x_i) = P(x_i \mid c_j)P(c_j) / P(x_i) \qquad (4a)$$

Because of the Naive Bayes assumption of conditional independence, the probability that an observation will have a specific set of values for the independent variables is the product of the conditional probabilities of observing each of the values given category c_j

$$P(X \mid c_j) = \prod_I P(x_i \mid c_j) \qquad (4b)$$

The method is described in more detail in Kantardzic (2003). To illustrate the use of Naive Bayes in predicting discrete variables, the provider 2 bill data was binned into groups based on the quintiles of the distribution. Because about 50 percent of the claims have a value of zero for provider 2 bill, only four categories are created by the binning procedure. The new variable was used to estimate the IME targets. Figure 3-6 displays a bar plot of the predicted probability of an IME for each of the groups. Figure 3-7 displays the fitted function. This function is a step function which changes value at each boundary of a provider 2 bill bin.

Bayes Predicted Probability IME Requested vs. Quintile of Provider 2 Bill

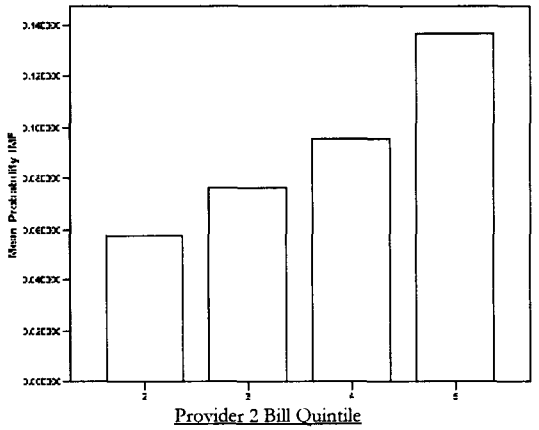


Figure 3-6

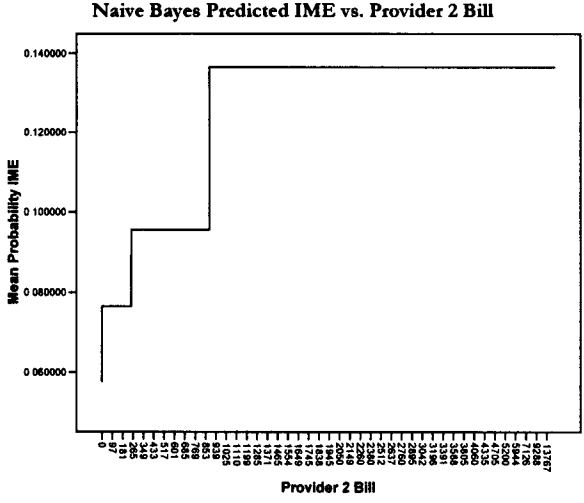


Figure 3-7

SECTION 4. SOFTWARE FOR MODELING NON-LINEAR DEPENDENCIES

Nonadditivity: interactions

Conventional statistical models such as regression and logistic regression assume not only linearity, but also additivity of the predictor variables. Under additivity, the effect of each variable can be can be added to the model one at a time. When the exact form of the relationship between a dependent and independent variable depends on the value of one or more other variables, the effects are not additive and one or more interactions exist. For instance, the relationship between provider 2 bill and IME may vary by type of injury (i.e. traumatic injuries versus sprains and strains). Interactions are common in insurance data (Weisberg and Derrig, 1998, Francis, 2003c).

With conventional linear statistical models, interactions are incorporated with multiplicative terms:

$$Y = a + b_1X_1 + b_2X_2 + b_3{*}X_1{*}X_2 \qquad (5)$$

In the case of a two-way interaction, the interaction terms appear as products of two variables. If one of the two variables is categorical, the interaction terms allow the slope of the fitted line to vary with the levels of the categorical variable. If both variables are continuous the interaction is a bilinear interaction (Jicard and Turrisi, 2003) and the slope of one variable changes as a linear function of the other variable. If both variables are categorical the model is equivalent to a two factor ANOVA with interactions.

The conventional approach to handling interactions has some limitations.

- Only a limited number of types of interactions can be modeled easily.
- If many predictor variables are included in the model, as is often the case in many predictive modeling applications, it can be tedious, if not impossible, to find all the significant interactions. Including all possible interactions in the model without regard to their significance likely results in a model which is over-parameterized.

The tree-based data mining techniques used in this paper each have efficient methods for handling interactions.

- Interactions are inherent in the method used by trees to partition data. Once data have been partitioned, different partitions can and typically do split on different variables and capture different interactions among the predictor variables. When the decision rules used by a tree to reach a terminal node involve more than one variable, in general, an interaction is being modeled.
- Ensemble methods incorporate interactions because they are based on the tree approach.
- Naïve Bayes, because it assumes conditional independence of the predictors, ignores interactions.
- Logistic regression incorporates interactions in the same way ordinary least squares regression does, with product interaction terms. In this fraud comparison study, no attempt was made to incorporate interaction terms as this procedure lacks an efficient way to search for the significant interactions.

Multiple predictors

Thus far, the discussion of the tree-based models concerned only simple one or two variable models. Extending the tree methods to incorporate many potential predictors is straightforward. For each tree fit, the method proceeds as follows:

- For each variable determine the best two-way partition of the data.
- Select the variable which produces the best improvement in the goodness of fit statistic to split the data at a particular node.
- Repeat the process until no further improvement in fit can be obtained.

Software for modeling nonlinear dependencies and testing the models

Four software products were included in our fraud comparison: They are CART, TREENET,
S-PLUS (R) and Insightful Miner[23].

CART and TREENET are Salford Systems stand-alone software products that each performs one technique. CART (Classification and Regression Trees) does tree analysis and TREENET applies stochastic gradient boosting using the method described by Freidman (2001). All the software tested produce SAS code[24] that can be used to implement the model

in a production stage. All the products contain a procedure for handling missing values using surrogate variables. At any given split point, CART and TREENET find the variable that is next in importance in influencing the target variable and they use this variable to replace the missing data. The specific statistic used to rank the variables and find the surrogates is described in Brieman et. al. (1993). Different versions of CART and TREENET handle different size databases. The number of levels of categorical variables affects how much memory is needed, as more levels necessitate more memory. The 128k version of each product was used for this analysis. With approximately 100,000 records in the training data, occasional memory problems were experienced and it became necessary to sample fewer records. One of the very useful features of the Salford Systems software is that all the products rank variables in importance[25].

S-PLUS and R are comprehensive statistical languages used to perform a range of statistical analyses including exploratory data analysis, regression, ANOVA, generalized linear models, trees and neural networks. Both S-PLUS and R are derived from S, a statistical programming language originally developed at Bell Labs. The S progeny, S-PLUS and R, are popular among academic statisticians. S-PLUS is a commercial product sold by Insightful which has a true GUI interface that facilitates easier handling of some functions. Insightful also supplies technical support. The S-PLUS programming language is widely used by analysts who do serious number crunching. They find it more effective, especially for processes that are frequently repeated. R is free open source statistical software that is supported largely by academic statisticians and computer science faculty. It has only limited GUI functionality and the data mining functions must be accessed through the language. Most code written for S-PLUS will also work for R. One notable difference is that data must be converted to text mode to be read by R (a bit of an inconvenience, but usually not an insurmountable one). Fox (2002) points out some of the differences between the two languages, where they exist. The S-PLUS procedures used here in the fraud comparison are found in both S-PLUS and R. However one ensemble tree method used, Random Forest, appears only to be available in R. The S-PLUS (R) procedures used were: the tree function for decision trees and the glm (generalized linear models) for logistic regression. S-PLUS (R) incorporates relatively crude methods for handling missing values. These include eliminating all records with a missing value on any variable, an approach which is generally not recommended (Francis 2005, Allsion 2002). S-PLUS also creates a new category for missing values (on categorical variables) and allows aborting the analysis if a missing value is found. In general, it is necessary to preprocess the data (at least the numeric variables where there is no missing value method[26]) to make a provision for the missing values. In the fraud comparison, a constant not in the range of the data was substituted into the variable and an indicator dummy variable for missing was created for each numeric variable with missing values. S-PLUS and R are generally not considered optimal choices for analyzing large databases. After experiencing some difficulty reading training data of about 100,000 records into S-PLUS, the database was reduced to contain only the variables used in the analysis. Once the data was read into S-PLUS, few problems were experienced. Another eccentricity is that the S-PLUS tree function can only handle 32 levels on any given categorical variable, so in the preprocessing the number of levels may need to be reduced[27]. The R Random Forest function incorporates a procedure that can be used to rank variables in importance. The

procedure produces an *impurity* statistic which can be used to rank the variables. The impurity is based on the Gini index for classification applications and mean squared error for numeric dependent variables. The S-PLUS tree function contains no built-in capability for ranking variables in importance. Therefore using the S-PLUS language, an algorithm was coded into S-PLUS to rank the variables. The method is described in Francis (2001) and Potts (2000). The procedure quantifies how much the error increases when a variable is removed from the model; the larger the increase in errors, the more important the variable.

The Insightful Miner is a data mining suite that contains the most common data mining tools: regression, logistic regression, trees, ensemble trees, neural networks and Naïve Bayes[28]. As mentioned earlier, Insightful also markets S-PLUS. However, the Insightful Miner has been optimized for large databases and contains methods (Naïve Bayes) which are not part of S-PLUS (R). The Naïve Bayes, Tree and Ensemble Tree procedures from Insightful Miner are used here in the fraud comparison. The insightful Miner has several procedures for automatically handling missing values. These are 1) drop records with missing values, 2) randomly generate a value, 3) replace with the mean, 4) replace with a constant and 5) carry forward the last observation. Each missing value was replaced with a constant. In theory, the data mining methods used, such as trees, should be able to partition records coded for missing from the other observations with legitimate categorical or numeric values and separately estimate their impact on the target variable (possible after allowing for interactions with other variables). Server versions of the Insightful Miner generate C code that can be used in deploying the model, but the version used in this analysis did not have that capability. As mentioned above some preprocessing was necessary for the Naïve Bayes procedure. Since Insightful Miner contains no procedure for ranking variables in importance, no rankings were provided for the Iminer methods.

Validating and Testing

It is common in data mining circles to partition the data into three groups (Hastie et al., 2001). One group is used for "training", or fitting the model. Another group, referred to as the validation set, is used for "testing" the fit of the model and re-estimating parameters in order to obtain a better model. It is common for a number of iterations of testing and fitting to occur before a final model is selected. The third group of data, the "holdout" sample, is used to obtain an unbiased test of the model's accuracy. An alternative approach to a validation sample that is especially appropriate when the sample size used in the analysis is relatively modest, is cross-validation. Cross-validation is a method involving holding out a portion of the training sample, say one fifth of the data, fitting a model to the remainder of the data and testing it on the held out data. In the case of 5-fold cross validation, the process is repeated five times and the average goodness of fit of the five validations is computed. The various software products and procedures have different methods for validating the models. Some (Insightful Miner Tree) only allow cross-validation. Others (TREENET) use a validation sample. S-PLUS (R) allows either approach[29] to be used (so a test sample of about 20% of the training data was used as we had a relatively large database). Neither validation sample nor cross-validation was used with Naïve Bayes, Logistic Regression or the Ensemble Tree.

In this analysis, approximately a third of the data, about 50,000 records, was used as the holdout sample for the final testing and comparison of the models. Two key statistics often used to compare models accuracy are sensitivity and specificity. *Sensitivity* is the percentage of events (i.e., claims with an IME or referred to a special investigation unit) that were predicted to be events. The *specificity* is the percentage of nonevents (in our applications claims believed to be legitimate) that were predicted to be nonevents. Both of these statistics should be high for a good model. Table 4-1, often referred to as a confusion matrix (Hastie et. al., 2001), presents an example of the calculation.

Sample Confusion Matrix: Sensitivity and Specificity

	True Class		
Prediction	**No**	**Yes**	**Row Total**
No	800	200	1,000
Yes	200	400	600
Column Total	1,000	600	

	Correct	**Total**	**Percent Correct**
Sensitivity	800	1,000	80%
Specificity	400	600	67%

Table 4-1

In the example confusion matrix, 800 of 1,000 non-events are predicted to be non-events so the sensitivity is 80%. The specificity is 67% since 400 of 600 true positives are accurately predicted.

SECTION 5. SOFTWARE RANKINGS OF "IMPORTANT" VARIABLES IN THE DECISION TO INVESTIGATE: IME AND SIU

The remainder of this paper is devoted to illustrating the usefulness and effectiveness of eight model/software combinations applied to our Example 2, the decision to investigate via IMEs or referral to SIU. We model the presence and proportion of favorable outcomes, of each investigative technique for the DCD subset of automobile bodily injury liability (third party) claims from 1995-1997 accident years.[30] We employ twenty-one potentially predicting variables of three types: (1) eleven typical claim variable fields informative of injury claims as reported, both categorical and numeric, (2) three external demographic variables that may play a role in capturing variations in investigative claim types by geographic region of Massachusetts, and (3) seven internal "demographic" variables derived from informative pattern variables in the database. Variables of type 3 are commonly used in predictive modeling for marketing purposes. The variables used for these illustrations are by no means optimal choices for all three types of variables. Optimization can be approached by other procedures (beyond the scope of this paper) that maximize information and minimize cross correlations and by variable construction and selection by domain experts.

The eight model/software combinations we will use here are of two categories: six tree models, and two benchmark models (Naïve Bayes and Logistic). They are:

1) TREENET
2) Iminer Tree
3) SPLUS Tree
4) CART
5) Iminer Ensemble
6) Random Forest
7) Naïve Bayes
8) Logistic

As described in Section 4, CART and TREENET are Salford Systems stand-alone software products that each performs one technique. CART (Classification and Regression Trees) does tree analysis, and TREENET applies stochastic gradient boosting to an ensemble of trees using the method described by Freidman (2001). The S-PLUS procedures used here in the fraud comparison are found in both S-PLUS and in a freeware version in R. These were: the tree function for decision trees, and the GLM (generalized linear models) for logistic regression.

Insightful Miner is a data mining suite. The Naïve Bayes, Tree and Ensemble Tree procedures, from Insightful Miner are used here in the fraud comparison.

Model performance is covered in the next section, section 6, as we first cover the ranking of variables by "importance" in relation to the target variables: the decision to perform an IME or a Special Investigation (SIU) and the favorable outcomes of each investigative technique. The training data of approximately 75,000 records was used in the ranking evaluations.

Data mining models are typically complex models where it is difficult to determine the relevance of predictors to the model result. One of the handy tasks that some of the data

mining software products perform is to rank the predictor variables by their importance to the model in predicting the dependent variable. Where the software did not supply a ranking, we omitted an importance ranking leaving five model/software determinations of importance for the twenty-one variables. Different procedures are used for different methods and different products.

Two software products, CART and TREENET supply importance rankings. The procedures used are:

CART: CART uses a goodness of fit measure, also referred to in the literature as an impurity measure, and computed over the entire tree, to determine a variable's importance. In this study the goodness of fit measure was the Gini Index defined below (Hastie, et al., p.271-272):

$$i(t) = 1 - \sum_i p_i \quad \text{i=the categories of the dependent variable and } p_i \text{ is the probability of class} \quad (6)$$

Each split of the tree lowers the overall value for the statistic. CART keeps track of the impurity improvement at each node for both the variable used in the split and for surrogate variables used as a replacement in the case of missing values. A consequence of this is that a variable not used for splitting may rank higher in importance than a variable that is.

TREENET: Because it is composed of many small CART trees, TREENET uses the same method as CART to compute importance rankings.

S-PLUS (R) does not supply an importance ranking, but the programming language can be used to program a procedure to compute rankings. A sensitivity value was computed for each variable in the model. The sensitivity is a measure of how much the predicted value's error increases when the variables are excluded from the model one at a time. However, instead of actually excluding variables and refitting the model, their values are fixed at a constant value. (See Francis, 2001 for a detailed recipe for applying the approach). The sensitivity statistic was used to rank the variables from the tree function. For the logistic regression, information about the variables contribution to sum of squared variation explained by the model was used to rank it. Like CART and TREENET, Random Forest uses an impurity measure (i.e., Gini Index) to produce an importance ranking.

Insightful Miner does not supply importance rankings. Unlike S-PLUS (R), the analytical methods are not accessed through the language but through a series of icons placed on a palate. Thus, we were not able to custom program a ranking procedure for application with the Iminer's modeling methods. The resulting importance rankings were used in Tables 5-1A & 5-2A for the decisions to investigate and 5-1B and 5-2B for the favorable outcomes.

Each of five model/software combination outputs allowed for the evaluation of the predicting variables in rank order of importance, when significant, together with a measure of the relative value of importance on a scale of zero (insignificant) to 100 (most significant

variable). Table 5-1A displays the importance results for predicting an IME using the five tree models while Table 5-1B displays those results for the remaining five model/software combinations, including the benchmark Naïve Bayes and Logistic. The predicting variables are listed in the order of importance in the TREENET model, where all variables are significant. The number of significant variables found ranges from a low of twelve variables (S-PLUS Tree) to all twenty one (TREENET).

Software Ranking of Variables for IME Decision By Importance Rank and Value					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Variable	**TREENET**	**S Plus Tree**	**CART**	**Random Forest**	**Logistic**
Provider 2 Bill	1 (100)	2 (91)	1 (100)	1 (100)	10 (1)
Attorneys Per Zip	2 (80)	5 (26)	13 (9)	6 (34)	11 (1)
Territory	3 (71)	4 (32)	11 (11)	3 (59)	*
Health Insurance	4 (61)	1 (100)	3 (68)	2 (84)	1 (100)
Injury Type	5 (50)	6 (24)	5 (47)	10 (18)	2 (51)
Provider 1 Bill	6 (47)	3 (51)	4 (58)	4 (59)	*
Provider 1 Type	7 (31)	9 (7)	*	12 (15)	6 (8)
Report Lag	8 (31)	7 (16)	8 (18)	8 (27)	13 (1)
Attorney	9 (25)	12 (3)	*	19 (5)	5 (18)
Age	10 (23)	*	17 (2)	17 (8)	*
Provider 2 Type	11 (19)	8 (9)	*	5 (42)	3 (47)
Income Household/Zip	12 (18)	*	10 (13)	11 (16)	9 (2)
Avg. Household Price/Zip	13 (17)	*	15 (5)	*	*
Providers per City	14 (17)	*	9 (15)	16 (9)	*
Claimants per City	15 (16)	*	*	7 (32)	12 (1)
Providers/Zip	16 (16)	*	*	15 (13)	8 (2)
Households/Zip	17 (16)	11 (3)	*	13 (15)	7 (2)
Treatment Lag	18 (14)	10 (4)	18 (2)	9 (24)	4 (24)
Distance MP1 Zip to Clt Zip	19 (13)	*	20 (0.1)	14 (14)	*
Emergency Treatment	20 (4)	*	7 (20)	18 (6)	*
Policy Type	21 (3)	*	19 (2)	20 (0)	*

Note: * represents insignificance of variable in the model.

Table 5-1A

The same set of model/software combinations was used with the same set of twenty-one predicting variables to predict the favorable outcome of the IME. Table 5-1B shows the importance of each of the 21 predictors for modeling favorable outcomes of IMEs.

Software Ranking of Variables for IME Favorable By Importance Rank and Value					
	(1)	(2)	(3)	(4)	(5)
Variable	**TREENET**	**S Plus Tree**	**CART**	**Random Forest**	**Logistic**
Provider 2 Bill	5 (64)	3 (22)	4 (37)	5 (49)	2 (13)
Attorneys Per Zip	11 (28)	*	11 (6)	13 (28)	11 (1)
Territory	2 (98)	2 (43)	12 (5)	1 (100)	4 (9)
Health Insurance	1 (100)	1 (100)	1 (100)	2 (71)	1 (100)
Injury Type	4 (76)	5 (10)	9 (15)	4 (67)	3 (13)
Provider 1 Bill	7 (45)	4 (15)	2 (51)	3 (70)	*
Provider 1 Type	8 (38)	9 (16)	5 (36)	10 (32)	5 (9)
Report Lag	6 (53)	8 (7)	18 (0)	6 (45)	8 (6)
Attorney	12 (25)	*	*	18 (3)	7 (8)
Age	13 (24)	*	19 (0)	9 (33)	*
Provider 2 Type	10 (29)	*	6 (30)	12 (31)	*
Income Household/Zip	20 (7)	11 (4)	17 (0)	8 (33)	10 (2)
Avg. Household Price/Zip	15 (16)	*	15 (0)	*	*
Providers per City	19 (8)	*	8 (17)	15 (23)	*
Claimants per City	9 (36)	12 (3)	13 (2)	16 (22)	13 (1)
Providers/Zip	17 (12)	13 (2)	7 (20)	11 (31)	*
Households/Zip	16 (15)	7 (7)	16 (0)	7 (37)	9 (2)
Treatment Lag	14 (22)	14 (1)	10 (6)	14 (28)	6 (8)
Distance MP1 Zip to Clt Zip	3 (78)	6 (8)	14 (1)	*	*
Emergency Treatment	18 (9)	10 (6)	3 (44)	17 (5)	12 (1)
Policy Type	21 (5)	*	*	*	*

Note: * represents insignificance of variable in the model.

Table 5-1B

The same set of five model/software combinations was used with the same set of twenty-one predicting variables to predict the use of special investigation or SIU. Tables 5-2A and 5-2B show the corresponding ranking of variables by importance for each of the five model combinations and two target variables, decision and favorable.

Software Ranking of Variables for SIU Decision By Importance Rank and Value					
	(1)	(2)	(3)	(4)	(5)
Variable	**TREENET**	**S Plus Tree**	**CART**	**Random Forest**	**Logistic**
Providers/Zip	1 (100)	1 (100)	8 (37)	3 (74)	*
Provider 2 Type	2 (98)	10 (3)	15 (34)	10 (30)	6 (39)
Territory	3 (92)	5 (18)	3 (84)	1 (100)	*
Health Insurance	4 (64)	3 (33)	7 (52)	6 (50)	7(28)
Provider 1 Bill	5 (59)	2 (51)	2 (85)	2 (89)	14 (2)
Injury Type	6 (52)	7 (6)	5 (59)	16 (5)	2 (71)
Attorney	7 (47)	8 (4.5)	17 (13)	18 (4)	3 (63)
Provider 1 Type	8 (38)	4 (29)	4 (69)	5 (51)	1 (100)
Age	9 (31)	*	*	17 (5)	*
Provider 2 Bill	10 (30)	*	1 (100)	4 (74)	13 (5)
Report lag	11 (28)	*	6 (54)	8 (10)	11 (17)
Average House Price	12 (28)	*	15 (18)	*	*
Attorneys/zip	13 (22)	6 (8)	14 (20)	9 (30)	12 (7)
Distance to Provider	14 (20)	*	19 (4)	15 (18)	4 (58)
Emergency Treatment	15 (19)	*	13 (27)	19 (4)	5 (49)
Income/Cap Household	16 (18)	11 (3)	9 (4.5)	13 (21)	9 (27)
Claimants per City	17 (17)	*	12 (30)	11 (26)	*
Treatment Lag	18 (16)	9 (34)	18 (12)	14 (20)	15 (2)
Households/Zip	19 (16)	*	16 (16)	12 (21)	8 (28)
Policy Type	20 (8)	*	*	20 (1)	*
Providers per City	21 (6)	12 (1)	11 (30)	7 (44)	10 (22)

Note: * represents insignificance of variable in the model.

Table 5-2A

Software Ranking of Variables for SIU Favorable By Importance Rank and Value					
	(6)	(7)	(8)	(9)	(10)
Variable	TREENET	S Plus Tree	CART	Random Forest	Logistic
Providers/Zip	10 (20)	10 (6)	12 (25)	7 (24)	13 (2)
Provider 2 Type	4 (41)	*	7 (35)	9 (18)	5 (21)
Territory	1 (100)	2 (94)	1 (100)	1 (100)	1 (100)
Health Insurance	13 (18)	6 (16)	*	15 (10)	7 (19)
Provider 1 Bill	6 (30)	13 (4)	15 (9)	5 (29)	14 (1)
Injury Type	3 (58)	5 (16)	6 (39)	16 (8)	3 (41)
Attorney	14 (16)	12 (4)	9 (27)	18 (6)	6 (20)
Provider 1 Type	5 (40)	1 (100)	3 (50)	3 (33)	2 (45)
Age	8 (22)	*	17 (7)	13 (13)	11 (2)
Provider 2 Bill	2 (66)	4 (18)	8 (32)	6 (26)	9 (3)
Report lag	7 (25)	7 (14)	19 (2)	2 (36)	12 (2)
Average House Price	15 (16)	*	13 (24)	*	*
Attorneys/zip	11 (19)	8 (14)	4 (45)	10 (17)	*
Distance to Provider	16 (15)	9 (14)	5 (39)	*	*
Emergency Treatment	21 (9)	3 (72)	14 (17)	14 (11)	4 (25)
Income/Cap Household	17 (14)	11 (5)	2 (61)	11 (16)	*
Claimants per City	12 (19)	*	11 (25)	12 (13)	15 (1)
Treatment Lag	19 (13)	*	18 (4)	17 (6)	*
Households/Zip	18 (13)	*	16 (9)	8 (19)	8 (5)
Policy Type	20 (10)	*	*	*	*
Providers per City	9 (21)	14 (3)	10 (26)	4 (31)	10 (2)

Note: * represents insignificance of variable in the model.

Table 5-2B

Clearly, in both instances of target variables the specific model and software implementation determines how to unwind the cross correlations to extract the most information for prediction purposes. For example, the distance between the claimant's zip code and the first outpatient provider (Distance) ranks low in importance (19/21) in the TREENET application for the IME decision target but it is quite important in the TREENET model for favorable IME outcome (3/21). Note, however, provider 2 bill is deemed highly important in all IME non-benchmark applications. One way to isolate the importance of each predicting variable is to tally a summary importance score across models. We will use a score of (21-rank)*(importance), with all insignificant variables assigned zero importance, summed over all relevant model combinations. For example, the variable provider 2 type would have a summary score relating to the IME target across the five tree models for a total importance score of 2,268. This scoring formula is typical of the ad hoc methods common to data mining analytics. The multiplicative form gives emphasis to both the categorical rank and the importance score in a dual monotone way. The numeric value of the score is less important than the final rankings of the variables. Tables 5-3A&B and 5-4A&B show the range of variable importance summary scores for all variables relative to the two targets, IME and SIU, respectively. The ranks of the variables according to the two summary scores are highly (Pearson) correlated as, for example, the decision summary ranks and favorable summary ranks have correlation coefficients of 0.65 for IME and 0.57 for SIU. The tables

also indicate the variable category of original DCD field (F), an internally derived variable (DV) and an external demographic variable (DM). The external demographic variables do not seem to be very informative in the presence of the field and derived variables chosen.

Important Variable Summarizations for IME Tree Models Applied to Decision and Favorable Targets					
			Total Score	Decision Score	Favorable Score
Variable	Variable type	Total Score	Rank	Rank	Rank
Health Insurance	F	17,206	1	2	1
Provider 2 Bill	F	10,820	2	1	4
Territory	F	7,871	3	5	2
Provider 1 Bill	F	6,726	4	4	3
Injury Type	F	6,084	5	6	5
Attorneys Per Zip	DV	3,102	6	3	15
Provider 2 Type	F	2,873	7	8	9
Report Lag	DV	2,859	8	16	7
Provider 1 Type	F	2,531	9	10	6
Distance MP1 Zip to Clt Zip	DV	1,655	10	11	8
Treatment Lag	DV	1,331	11	17	16
Emergency Treatment	F	1,216	12	7	10
Claimants per City	DV	1,146	13	14	13
Income Household/Zip	DM	987	14	13	17
Attorney	F	971	15	9	19
Households/Zip	DM	957	16	19	11
Age	F	881	17	12	14
Providers/Zip	DV	838	18	18	12
Providers per City	DV	719	19	20	18
Avg. Household Price/Zip	DM	262	20	15	20
Policy Type	F	4	21	21	21

Table 5-3

Important Variable Summarizations for SIU Tree Models Applied to Decision and Favorable Targets					
			Total Score	Decision Score	Favorable Score
Variable	**Variable Type**	**Total Score**	**Rank**	**Rank**	**Rank**
Territory	F	15,242	1	2	1
Provider 1 Type	F	9,965	2	4	2
Providers/Zip	DV	6,676	3	1	13
Provider 1 Bill	F	6,240	4	3	10
Provider 2 Bill	F	6,030	5	5	4
Injury Type	F	5,845	6	7	3
Provider 2 Type	F	4,753	7	8	6
Health Insurance	F	4,262	8	6	15
Emergency Treatment	F	3,039	9	13	5
Attorney	F	2,705	10	9	14
Report lag	DV	2,642	11	10	9
Providers per City	DV	2,275	12	12	10
Attorneys/zip	DV	2,183	13	14	8
Distance to Provider	DV	2,109	14	11	14
Income/Cap Household	DM	2,091	15	15	7
Claimants per City	DV	1,142	16	18	16
Households/Zip	DM	1,061	17	16	18
Age	F	830	18	19	17
Treatment Lag	DV	706	19	17	20
Average House Price	DM	648	20	20	9
Policy Type	F	19	21	21	21

Table 5-4

Additional Analyses

Most software allow for additional diagnostic tools that focus on the importance of individual variable levels in the predictive model. We focus on two such features: partial dependency plots and pruning of trees. Both features are designed to illustrate the contribution of each *level* of categorical variable and each *interval* of continuous variables created by the cut points. We illustrate the additional analyses using the Random Forest and S-PLUS's tree software.

Partial Dependence

The partial dependence plot is a useful way to visualize the effect of the values of a specific variable on a dependent variable when a complex modeling method such as Random Forest is used. The partial dependence plot is a graph of the marginal effect of a variable on the class probability. For a classification application (in Random Forest), the partial plot uses the logit or log of the odds ratio (the odds of being in the target category versus its compliment) rather than the actual probability.

$$f(x) = \log p_k(x) - \sum_{j=1}^{K} \log(p_j) \qquad (7)$$

Figures 5-1 and 5-2 show the partial dependence plot for the two IME targets for the most important variable in Table 5-4, territory.

Random Forest: IME Requested

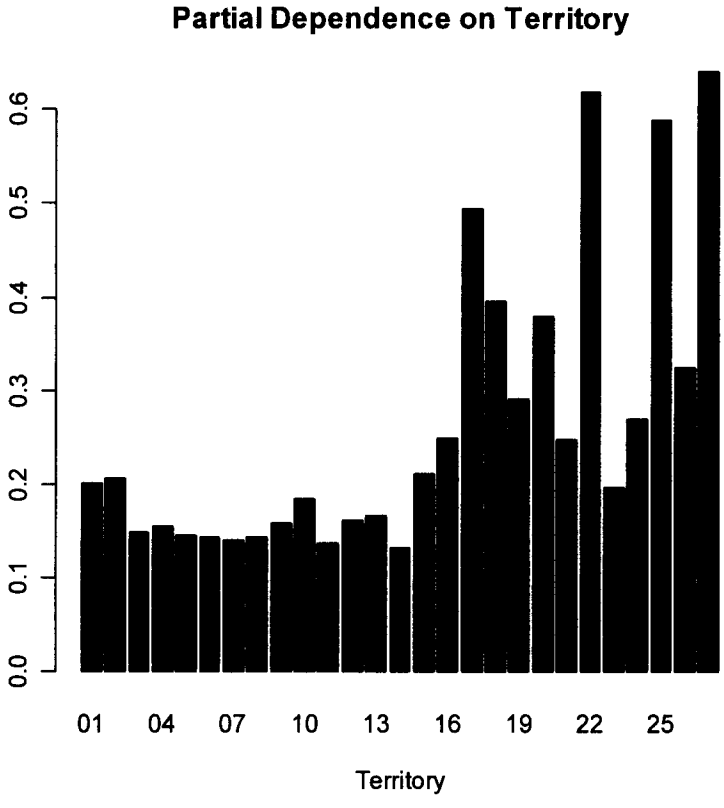


Figure 5-1

Random Forest: IME Favorable

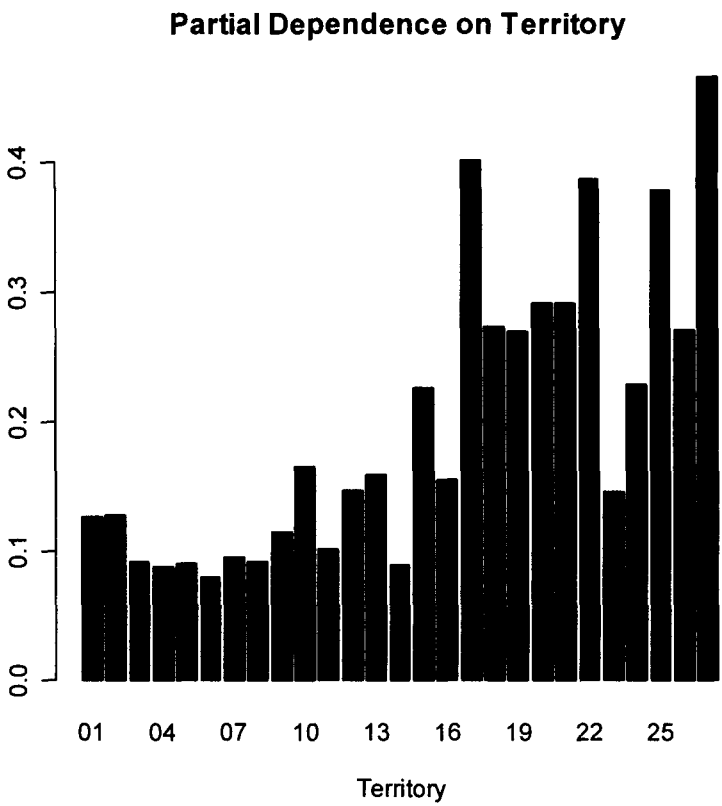


Figure 5-2

Both bar graphs have a distinctive right shift in the size of the partial dependency on the territory variable. This result is not surprising given that Massachusetts automobile territories are set every two years based upon the calculation of a single 5-coverage pure premium index for each of 350 towns. Towns are then grouped into 16 nearly homogenous territories with the index generally rising from territory 1 (lowest) to territory 16 (highest). Territories 17-26 are 10 individual parts of Boston that vary widely in this calculated pure premium index (Conger, 1987). Figure 5-3 shows a bar graph of the pure premium indices for the 26 territories used in this analysis for comparison purposes.

Massachusetts Rating Territories

Five Coverage Pure Premiums

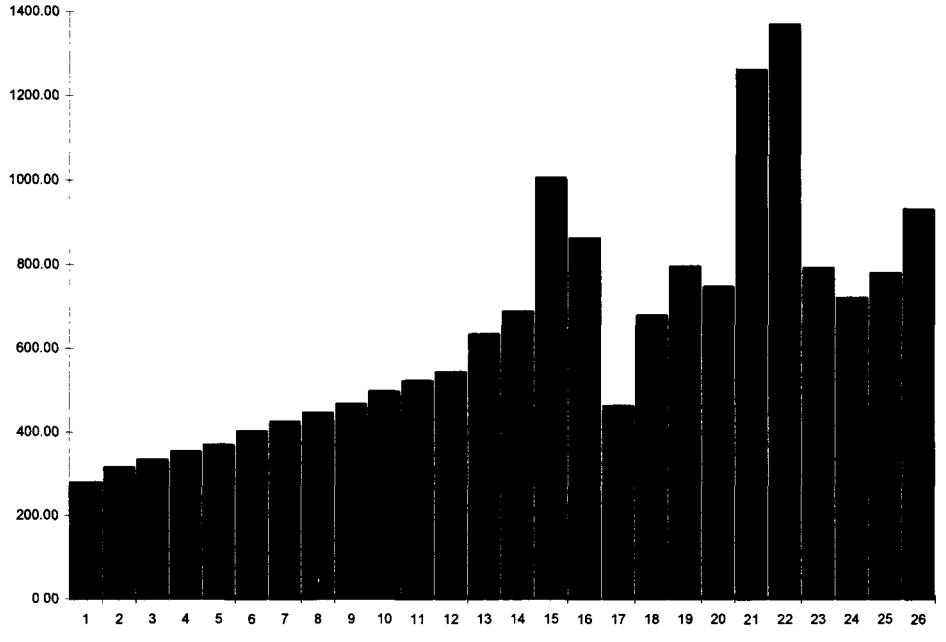


Figure 5-3

Figure 5-4 displays the proportion of claims with an IME requested (not marginal effects) by territory, superimposed on the pure premium territory levels. In contrast to the similarity of the marginal importance of the IME territory variable to the territory pure premiums, the proportions of claims with IME requested shown in Figure 5-4 show more uniformity across territories, indicating a real dependence on other important variables.

Massachusetts Rating Territories

Five Coverage Pure Premium vs IME Request Ratios

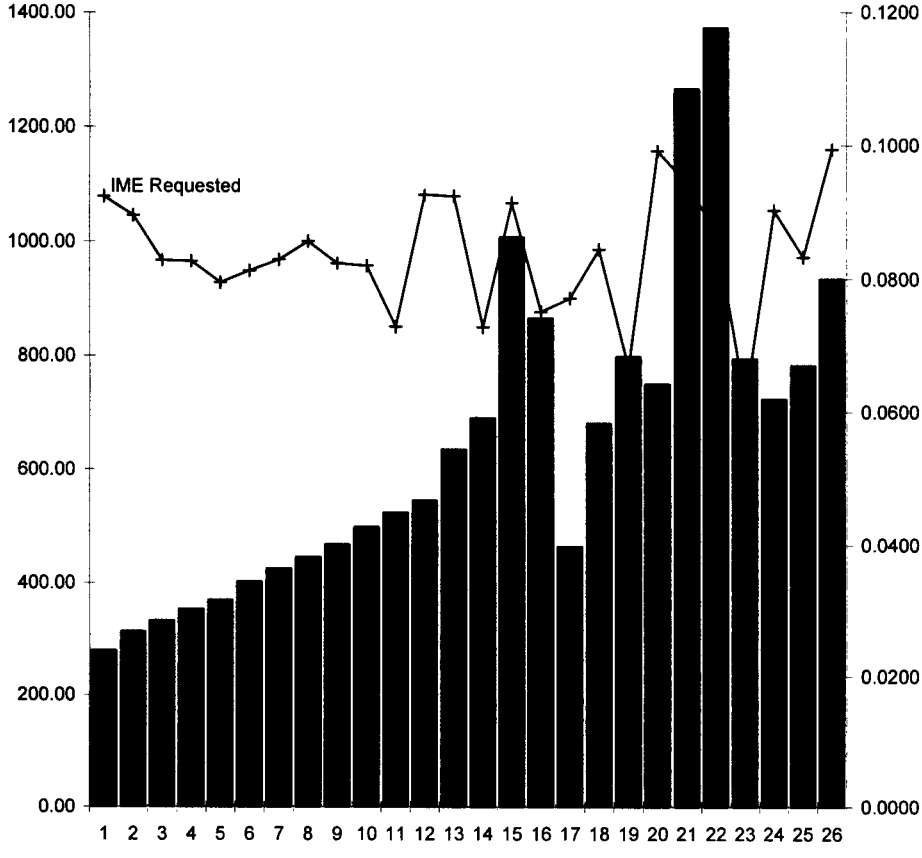


Figure 5-4

Pruning the Trees

Simple trees[31] that extend to a large number of terminal nodes are difficult to assess the full importance of individual variable levels because (1) later node splits may or may not be statistically significant depending on the software algorithms employed and (2) terminal nodes on the order of fifty plus may obscure the precise contribution of each variable level despite the importance value described above for the overall variable.

The full tree produced by the software can be *pruned* back to the "best" tree with a pre-determined number of nodes. For example, Figure 5-5 shows a best 10 node pruned tree from S-PLUS. It begins with the health insurance variable as the "root" node (Y/N to the left and U to the right)[32] and proceeds to make general node splits based only on the provider 2 bill amount. The universe of records is then classified by terminal node IME requested ratios ranging from 0.019 to 0.170. A similar pruned tree can be produced for the other three targets.

S-PLUS TREE: IME Requested
Best Ten Node Pruned Tree

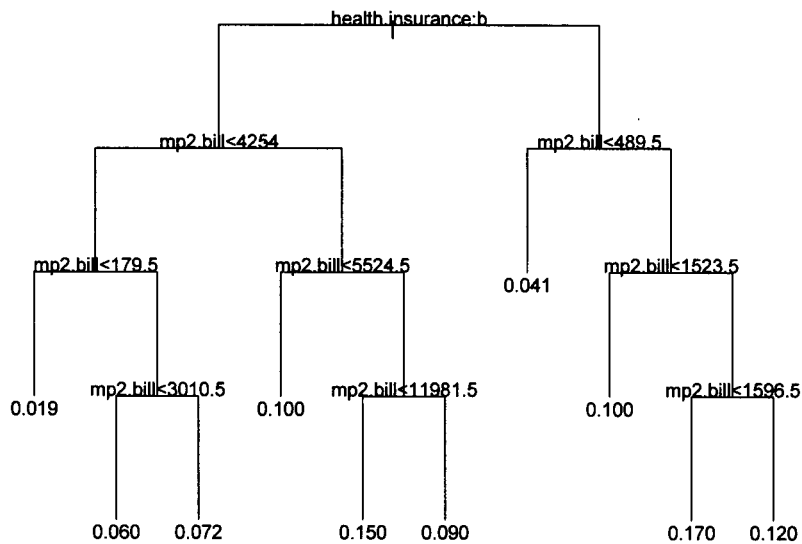


Figure 5-5

We next turn to consideration of model performance as a whole in section 6 with an interpretation of the models and variables relative to the problem at hand (example 2) in Section 7.

SECTION 6. ROC CURVES AND LIFT FOR SOFTWARE: TREES, NAIVE BAYES AND LOGISTIC MODELS

The sensitivity and specificity measures discussed in Section 4 are dependent on the choice of a cutoff value for the prediction. Many models score each record with a value between zero and one, though some other scoring scale can be used. This score is sometimes treated like a probability, although the concept is much closer in spirit to a fuzzy set measurement function[33]. A common cutoff point is .5 and records with scores greater than .5 are classified as events and records below that value are classified as non-events[34]. However, other cutoff

values can be used. Thus, if a cutoff lower than 50% were selected, more events would be accurately predicted and fewer non-events would be accurately predicted.

Because the accuracy of a prediction depends on the selected cutoff point, techniques for assessing the accuracy of models over a range of cutoff points have been developed. A common procedure for visualizing the accuracy of models used for classification is the receiver operating characteristic (ROC) curve[35]. This is a curve of sensitivity versus specificity (or more accurately 1.0 minus the specificity) over a range of cutoff points. It illustrates graphically the sensitivity or true positive rate compared to 1- specificity or false alarm rate. When the cutoff point is very high (i.e. 1.0) all claims are classified as legitimate. The specificity is 100% (1.0 minus the specificity is 0), but the sensitivity is 0%. As the cutoff point is lowered, the sensitivity increases, but so does 1.0 minus the specificity. Ultimately a point is reached where all claims are predicted to be events, and the specificity declines to zero (1.0 - specificity = 1.0). The baseline ROC curve (where no model is used) can be thought of as a straight line from the origin with a 45-degree angle. If the model's sensitivity increases faster than the specificity decreases, the curve "lifts" or rises above a 45-degree line quickly. The higher the "lift" or "gain"; the more accurate the model[36]. ROC curves have been used in prior studies of insurance claims and fraud detection regression models (Derrig and Weisberg, 1998 and Viaene et al., 2002). The use of ROC curves in building models as well as comparing performance of competing models is a well established procedure (Flach et al (2003)).

A statistic that provides a one-dimensional summary of the predictive accuracy of a model as measured by an ROC curve is the area under the ROC curve (AUROC). In general, AUROC values can distinguish good models from bad models but may not be able to distinguish among good models (Marzban, 2004). A curve that rises quickly has more area under the ROC curve. A model with an area of .50 demonstrates no predictive ability, while a model with an area of 1.0 is a perfect predictor (on the sample the test is performed on). For this analysis, SPSS was used to produce the ROC curves and area under the ROC curves. SPSS generates cutoff values midway between each unique score in the data and uses the trapezoidal rule to compute the AUROC. A non-parametric method was used to compute the standard error of the AUROC. The formula for the standard error[37] is:

$$SE(A) = \sqrt{\frac{A(1-A) + (n_+ - 1)(Q_1 - A^2) + (n_- - 1)(Q_2 - A^2)}{n_+ N}} \qquad (8)$$

Where n_+ is the number of events, n_- is the number of non-events, N is the sample size
A is the AUROC and scores are denoted as x

$$Q_1 = \frac{1}{n_- n_+^2} \sum_x n_- = j \times [n_{+>/}^2 + n_{+>/} + n_{+=/} + \frac{n_{+=/}^2}{3}]$$

$$Q_2 = \frac{1}{n_-^2 n_+} \sum_x n_+ = j \times [n_{->/}^2 + n_{->/} + n_{-=/} + \frac{n_{-=/}^2}{3}]$$

Tables 6-1A&B show the values of AUROC for each of eight model/software combinations in predicting a decision to investigate with an IME (6-1A) and an SIU (6-1B). for the Massachusetts auto bodily injury liability claims that comprise the holdout sample, about 50,000 claims. Upper and lower bounds for the "true" AUROC value are shown as the AUROC value ± two standard deviation determined by equation (7). TREENET, Random Forest both do well with AUROC values about 0.7, significantly better than the logistic model. The Iminer models (Tree, Ensemble and Naïve Bayes) generally have AUROC values significantly below the top two performers, with two (Tree and Ensemble) significantly below the Logistic and the Iminer Naïve Bayes benchmarks. CART also scores at or below the benchmarks and significantly below TREENET and Random Forest. On the other hand, S-PLUS (R) tree scores at or somewhat above the benchmarks.

Area Under the ROC Curve – IME Decision				
	CART Tree	**S-PLUS Tree**	**Iminer Tree**	**TREENET**
AUROC	0.669	0.688	0.629	0.701
Lower Bound	0.661	0.680	0.620	0.693
Upper Bound	0.678	0.696	0.637	0.708
	Iminer Ensemble	**Random Forest**	**Iminer Naïve Bayes**	**Logistic**
AUROC	0.649	703	0.676	0.677
Lower Bound	0.641	695	0.669	0.669
Upper Bound	0.657	711	0.684	0.685

Table 6-1A

Area Under the ROC Curve – IME Favorable				
	CART Tree	**S-PLUS Tree**	**Iminer Tree**	**TREENET**
AUROC	0.651	0.664	0.591	0.683
Lower Bound	0.641	0.653	0.578	0.673
Upper Bound	0.662	0.675	0.603	0.693
	Iminer Ensemble	**Random Forest**	**Iminer Naïve Bayes**	**Logistic**
AUROC	0.654	0.692	0.670	0.677
Lower Bound	0.643	0.681	0.660	0.667
Upper Bound	0.665	0.702	0.681	0.687

Table 6-1B

Tables 6-2A&B show the values of AUROC for the model/software combinations tested for the SIU dependent variable. We first note that, in general, the model predictions as measured by AUROC are significantly lower than for IME across all eight model/software combinations. This reduction in AUROC values may be a reflection of the explanatory variables used in the analysis; i.e., they may be more informative about claim build-up, for which IME is the principal investigative tool, than about claim fraud, for which SIU is the principal investigative tool.

Area Under the ROC Curve – SIU Decision				
	CART Tree	**S-PLUS Tree**	**Iminer Tree**	**TREENET**
AUROC	0.607	0.616	0.565	0.643
Lower Bound	0.598	0.607	0.555	0.634
Upper Bound	0.617	0.626	0.575	0.652
	Iminer Ensemble	**Random Forest**	**Iminer Naïve Bayes**	**Logistic**
AUROC	0.539	0.677	0.615	0.612
Lower Bound	0.530	0.668	0.605	0.603
Upper Bound	0.548	0.686	0.625	0.621

Table 6-2A

Area Under the ROC Curve – SIU Favorable				
	CART Tree	**S-PLUS Tree**	**Iminer Tree**	**TREENET**
AUROC	0.598	0.616	0.547	0.678
Lower Bound	0.584	0.607	0.555	0.667
Upper Bound	0.612	0.626	0.575	0.689
	Iminer Ensemble	**Random Forest**	**Iminer Naïve Bayes**	**Logistic**
AUROC	0.575	0.645	0.607	0.610
Lower Bound	0.530	0.631	0.593	0.596
Upper Bound	0.548	0.658	0.625	0.623

Table 6-2B

TREENET and Random Forest perform significantly better than all other model/software combinations on the favorable target variables. Both perform significantly better than the Logistic. Iminer Tree and Ensemble again do poorly on the IME and SIU Favorable holdout samples.

Figures 6-1 to 6-4 show the ROC curves for TREENET compared to the Logistic for both IME and SIU[38]. As we can see, a simple display of the ROC curves may not be sufficient to distinguish performance of the models as well as the AUROC values.

TREENET ROC Curve – IME
AUROC = 0.701

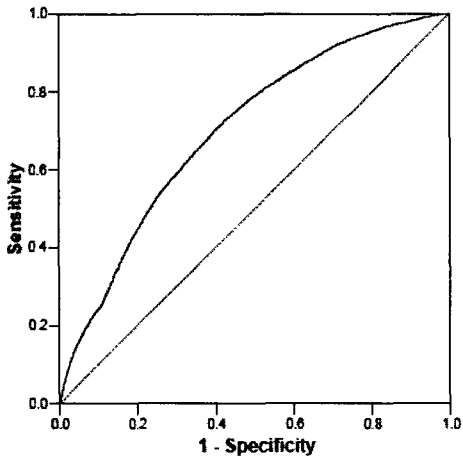


Figure 6-1

TREENET ROC Curve – SIU
AUROC = 0.677

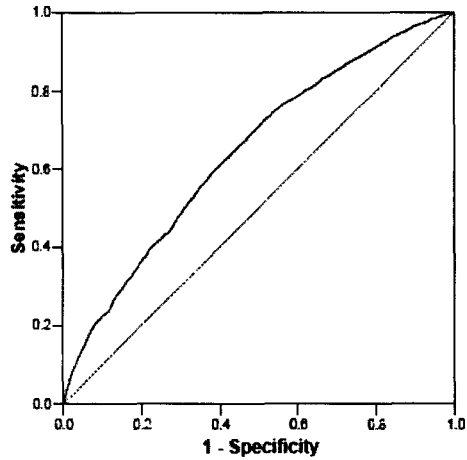


Figure 6-2

Logistic ROC Curve – IME
AUROC = 0.643

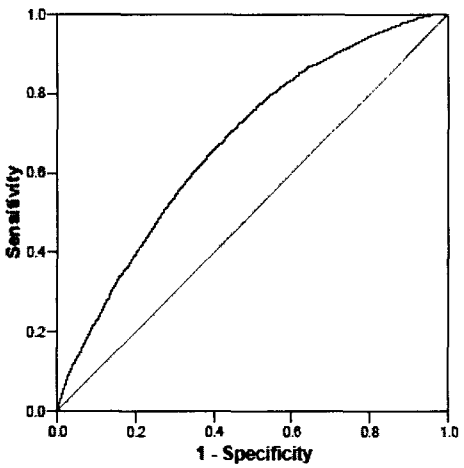


Figure 6-3

Logistic ROC Curve – SIU
AUROC = 0.612

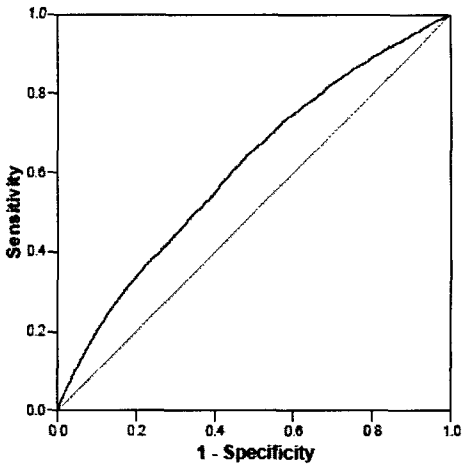


Figure 6-4

Finally, Table 6-3 displays the relative performance of the model/software combinations according to AUROC values and their ranks. With Naïve Bayes and Logistic as the benchmarks, TREENET, Random Forest and SPLUS Tree do better than the benchmarks while CART Tree, Iminer Tree, and Iminer Ensemble do worse.

Ranking of Methods By AUROC - Decision				
Method	**SIU AUROC**	**SIU Rank**	**IME Rank**	**IME AUROC**
Random Forest	0.645	1	1	0.703
TREENET	0.643	2	2	0.701
S-PLUS Tree	0.616	3	3	0.688
Iminer Naïve Bayes	0.615	4	5	0.676
Logistic	0.612	5	4	0.677
CART Tree	0.607	6	6	0.669
Iminer Tree	0.565	7	8	0.629
Iminer Ensemble	0.539	8	7	0.649

Table 6-3A

Ranking of Methods By AUROC - Favorable				
Method	**SIU AUROC**	**SIU Rank**	**IME Rank**	**IME AUROC**
TREENET	0.678	1	2	0.683
Random Forest	0.645	2	1	0.692
S-PLUS Tree	0.616	3	5	0.664
Logistic	0.610	4	3	0.677
Iminer Naïve Bayes	0.607	5	4	0.670
CART Tree	0.598	6	7	0.651
Iminer Ensemble	0.575	7	6	0.654
Iminer Tree	0.547	8	8	0.591

Table 6-3B

Finally, Figures 6-5A&B show the relative performance in a graphic. Procedures would work equally on both IME and SIU if they lie on the 45 degree line. To the extent that performance is better on the IME targets, procedures would be above the diagonal. Better performance is shown by positions farther to the right and closer to the top of the square. This graphic clearly shows that TREENET and Random Forest procedures do better than the other tree procedures and the benchmarks.

Plot of AUROC for SIU vs. IME Decision

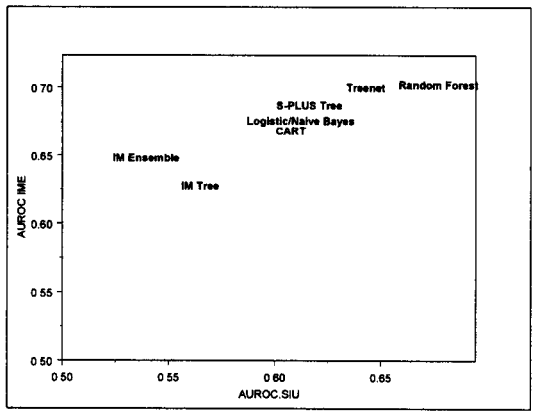


Figure 6-5A

Plot of AUROC for SIU vs. IME Favorable

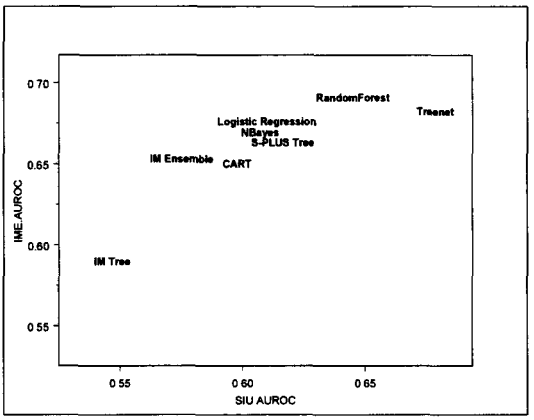


Figure 6-5B

SECTION 7. CONCLUSION

Insurance data often involves both large volumes of information and nonlinearity of variable relationships. A range of data manipulation techniques have been developed by computer scientists and statisticians that are now categorized as data mining, techniques with principal advantages being precisely the efficient handling of large data sets and the fitting of non-linear functions to that data. In this paper we illustrate the use of software implementations of CART and other tree-based methods, together with benchmark procedures of Naïve Bayes and Logistic regression. Those eight model/software combinations are applied to data arising in the Detail Claim Database (DCD) of auto injury liability claims in Massachusetts. Twenty-one variables were selected to use in prediction models using the DCD and external demographic variables. Four target categorical variables were selected to model: The decision to request an independent medical examination (IME) or a special investigation (SIU) and the favorable outcome of each investigation. The two decision targets are the prime claim handling techniques that insurers can use to reduce the asymmetry of information between the claimant and the insurer in order to distinguish valid claims from those involving buildup, exaggerated injuries or treatment, or outright fraud.

Eight modeling software results were compared for effectiveness of modeling the targets based on a standard procedure, the area under the receiver operating characteristic curve (AUROC). We find that the methods all provide some predictive value or lift from the predicting variables we make available, with significant differences among the eight methods and four targets. Seven modeling outcomes are compared to logistic regression as in Viaene et al. (2002) but the results here are different. They show some software/methods can improve on the predictive ability of the logistic model. TREENET, Random Forest and SPLUS Tree do better than the benchmark Naïve Bayes and Logistic methods, while CART

tree, Iminer tree, and Iminer Ensemble do worse. That some model/software combinations do better than the logistic model may be due to the relative size and richness of this data set and/or the types of independent variables at hand compared to the Viaene et al. data.

We show how "important" each variable is within each software/model tested and note the type of data that are important for this analysis. In general, variables taken directly from DCD fields and variables derived as demographic type variables based on DCD fields do better than variables derived from external demographic data. Variables relating to the injury and medical treatment dominate the highly important variables while the presence of an attorney, age of the claimant, and policy type, personal or commercial, are less important in making the decision to invoke these two investigative techniques.

No general conclusions about auto injury claims can be drawn from the exercise presented here except that these modeling techniques should have a place in the actuary's repertory of data manipulation techniques. Technological advancements in database assembly and management, especially the availability of text mining for the production of variables, together with the easy access to computer power, will make the use of these techniques mandatory for analyzing the nonlinearity of insurance data. As for our part in advancing the use of data mining in actuarial work, we will continue to test various software products that implement these and other data mining techniques (e.g. support vector machines).

REFERENCES

Allison, P., Missing Data, Sage Publications, 2002.

Automobile Insurers Bureau of Massachusetts, Detail Claim Database Claim Distribution Characteristics, Accident Years 1995-1997, Boston MA, 2004.

Brieman, L., J. Freidman, R. Olshen, and C. Stone, Classification and Regression Trees, Chapman Hall, 1993.

Conger, R.F., The Construction of Automobile Rating Territories in Massachusetts, *Casualty Actuarial Society Forum*, pp. 230-335, Fall 1987.

Derrig, R.A., and L. Francis, Comparison of Methods and Software for Modeling Nonlinear Dependencies: A Fraud Application, Working Paper, 2005.

Flach, P., H. Blockeel, C. Ferri, J. Hernandez-Orallo, and J. Struyf, Decision Support for Data Mining: An Introduction to ROC Analysis and Its Applications, Chapter 7 in Data

Mining and Decision Support, D. Mlandenic, Nada Lavrac, Marko Bohanec and Steve Moyle Eds., Kluwer Academic, N.Y., 2003.

Fox, J, An R and S-PLUS Companion to Applied Regression, SAGE Publications, 2002.

Francis, L.A., An introduction to Neural Networks in Insurance, Intelligent and Other Computational Techniques in Insurance, Chapter 2, A.F. Shapiro and L.C. Jain, Eds, World Scientific, pp. 51-1, 2003a.

Francis, L.A., Martian Chronicles: Is MARS better than Neural Networks? *Casualty Actuarial Society Forum,* Winter, pp. 253-320, 2003b.

Francis, L.A., Practical Applications of Neural Networks in Property and Casualty Insurance, Intelligent and Other Computational Techniques in Insurance, Chapter 3, A.F. Shapiro and L.C. Jain, Editors, World Scientific, pp. 104-136, 2003c.

Francis, L.A., Neural Networks Demystified, *Casualty Actuarial Society Forum,* Winter, pp. 254-319, 2001.

Francis, L.A., A Comparison of TREENET and Neural Networks in Insurance Fraud Prediction, presentation at CART Data Mining Conference, 2005.

Friedman, J., Greedy Function Approximation: The Gradient Boosting Machine, *Annals of Statistics,* 2001.

Hassoun, M.H., Fundamentals of Artificial Neural Networks, MIT Press, Cambridge MA, 1995.

Hastie, T., R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Springer, New York., 2001.

Hosmer, David W. and Stanley Lemshow, Applied Logistic Regression, John Wiley and Sons, 1989.

Insurance Research Council, Auto Injury Insurance Claims: Countrywide Patterns in Treatment Cost, and Compensation, Malvern, PA, 2004a.

Insurance Research Council, Fraud and Buildup in Auto Injury Insurance Claims, Malvern, PA, 2004b.

Jacard, J and Turrisi, Interaction Effects in Multiple Regression, SAGE Publications, 2003.

Kantardzic, M., Data Mining, John Wiley and Sons, 2003.

Lawrence, Jeannette, Introduction to Neural Networks: Design, Theory and Applications, *California Scientific Software,* 1994.

Marzban, C., A Comment on the ROC curve and the Area Under it as Performance Measures, 2004. *Weather and Forecasting,* Vol. 19, No. 6, 1106-1114.

McCullagh, P. and J. Nelder, General and Linear Models, Chapman and Hall, London, 1989.

Miller, R.B. and D.B. Wichern, Intermediate Business Statistics, Holden-Day, San Francisco, 1997.

Neter, John, Mihael H. Kutner, William Waserman, and Christopher J. Nachtsheim, Applied Linear Regression Models, 4th Edition, 1985.

Salford Systems, Data mining with Decision Trees: Advanced CART Techniques, Notes from Course, 1999.

Smith, Murry, (1996), Neural Networks for Statistical Modeling, International Thompson Computer Press, 1996.

Viaene, S., R.A. Derrig, and G. Dedene, A Case Study of Applying Boosting Naïve Bayes for Claim Fraud Diagnosis, IEEE Transactions on Knowledge and Data Engineering, 2004, Volume 16, (5), pp. 612-620, May.

Viaene, S., R.A. Derrig, and G. Dedene, Illustrating the Explicative Capabilities of Bayesian Learning Neural Networks for Auto Claim Fraud Detection, Intelligent and Other Computational Techniques in the Insurance, Chapter 10, A.F. Shapiro and L.C. Jain, Editors, World Scientific, pp. 365-399, 2003.

Viaene, S., B. Baesens, G. Dedene, and R. A. Derrig, A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Fraud Detection, *Journal of Risk and Insurance*, 2002, Volume 69, (3), pp. 373-421.

Weisberg, H.I. and R.A. Derrig, Methodes Quantitatives Pour La Detection Des Demandes D'Indemnisation Frauduleuses, *RISQUES*, No. 35, Juillet-Septembre, 1998, Paris.

Zhou, X-H, D.K. McClish, and N.A. Obuchowski, Statistical Methods in Diagnostic Medicine, John Wiley and Sons, New York, 2002.

[1] A good up-to-date and comprehensive source for a variety of data manipulation procedures is Hastie, Tibshirani, and Friedman (2001), Elements of Statistical Learning, Springer.

[2] They also found that augmenting the categorized red flag variables with some other claim data (e.g. age, report lag) improved the lift as measured by AUROC across all methods but the logistic model still did as well as the other methods (Viaene et al., 2002, Table 6, p.400-401).
[3] A wider set of data mining techniques is considered in Derrig, R.A. and L A. Francis, Comparison of Methods and Software Modeling Nonlinear Dependencies: A Fraud Application, Congress of Actuaries, Paris, June 2006
[4] See section 2 for an overview of the database and descriptions of the variables used for this paper.
[5] The relative importance of the independent variables in modeling the dependent variable within these methods are analogous to statistical significance or p-values in ordinary regression models.
[6] See, for example, 2004 Discussion Paper Program, Applying and Evaluating Generalized Linear Models, May 16-19, 2004, Casualty Actuarial Society.
[7] This was the text used by the Casualty Actuarial Society for the exam on applied statistics during the 1980s
[8] Claims that involve only third party subrogation of personal injury protection (no fault) claims but no separate indemnity payment or no separate claims handling on claims without payment are not reported to DCD.
[9] Combined payments under PIP and Medical Payments are reported to DCD.
[10] With a large holdout sample, we are able to estimate tight confidence intervals for testing model results in section 6 using the area under the ROC curve measure.
[11] This fact is a matter of Massachusetts law which does not permit IMEs by one type of physician, say an orthopedist, when another physician type is treating, say a chiropractor. This situation may differ in other jurisdictions.
[12] Because expert bill review systems became pervasive by 2003, reaching 100% in some cases, DCD redefined the reported MA to encompass only peer reviews by physicians or nurses for claims reported after July 1, 2003..
[13] The standard Massachusetts auto policy has a cooperation clause for IME both in the first party PIP coverage and in the third party BI liability coverage.
[14] The IRC also includes an index bureau check as one of the claims handling activities
[15] Prior studies of Massachusetts Auto Injury claim data for fraud content included Weisberg and Derrig (1998, Suspicion Regression Models) and Derrig and Weisberg (1998, Claim Screening with Scoring Models).
[16] See Section 5 for the importance of the provider 2 bill variable in the decision to investigate claims for fraud (SIU) and/or buildup (IME).
[17] There are Tree Software models that may split nodes into three or more branches. SPSS classification trees is an example of such software.
[18] For binary categorical data assumed to be generated from a binomial distribution, entropy and deviance are essentially the same measure. Deviance is a generalized linear model concept and is closely related to the log of the likelihood function.
[19] Hastie et al., p. 301 Note that Hastie et al. describe other error and weight functions. [endnote]
[20] Note that the ensemble tree methods employ all 21 variables in the models. See tables 5-1 and 5-2.
[21] The ROC curve results in Section 6 show that TREENET generally provides the best prediction models for the Massachusetts data.
[22] The numeric variables were grouped into five bins or into quintiles in this instance.
[23] The software product MARS also was used to rank variables in importance. MARS implements multivariate adaptive regression splines and is described in Francis (2003).
[24] The SAS code is generally relatively easy to edit if some other language is used to implement the model
[25] See Section 5 for the importance of variables in our study.
[26] S-PLUS would convert the numeric variable into a categorical variable with a level for every numeric value that is in the training data, including missing data, but the result would have far too many categories to be feasible.
[27] Generally by collapsing sparsely populated categories into an "all other" category

[28] It also contains some dimension reduction methods such as clustering and Principal Components which are also contained in S-PLUS.
[29] In general, some programming is required to apply either approach in S-PLUS (R)
[30] The data set is described in more detail in Section 2 above.
[31] Pruning is not feasible or necessary for the example tree methods such as TREENET or Random TREENET.
[32] The S-PLUS tree graph does not print out the values of categorical variables, although it displays the values of the numeric variables. For categorical variables letters are assigned and displayed instead of the category values.
[33] See Ostaszewski (1993) or Derrig and Ostaszewski (1999).
[34] One way of dealing with values equal to the cutoff point is to consider such observations as one-half in the event group and one-half in the non-event group
[35] A ROC curve is one example of a so-called "gains" chart.
[36] ROC curves were developed extensively for use in medical diagnosis testing in the 1970s and 1980s (Zhou et al. 2004 and more recently in weather forecasting (Marzban, 2004) and (Stephenson, 2000).
[37] The details of the formula were supplied by SPSS.
[38] All twenty ROC curves are available from the authors.

Acknowledgement: The authors gratefully acknowledge the production assistance of Eilish Browne and helpful comments from Rudy Palenik on a prior version of the paper.

Taming Text: An Introduction to Text Mining

Louise A. Francis, FCAS, MAAA

Abstract

Motivation. One of the newest areas of data mining is text mining. Text mining is used to extract information from free form text data such as that in claim description fields. This paper introduces the methods used to do text mining and applies the method to a simple example.

Method. The paper will describe the methods used to parse data into vectors of terms for analysis. It will then show how information extracted from the vectorized data can be used to create new features for use in analysis. Focus will be placed on the method of clustering for finding patterns in unstructured text information.

Results. The paper shows how feature variables can be created from unstructured text information and used for prediction

Conclusions. Text mining has significant potential to expand the amount of information that is available to insurance analysts for exploring and modeling data

Availability. Free software that can be used to perform some of the analyses describes in this paper is described in the appendix.

Keywords. Predictive modeling, data mining, text mining, statistical analysis

1. INTRODUCTION

Traditional statistical analysis is performed on data arrayed in spreadsheet format. That is, the data is arrayed in two dimensional matrices where each row represents a record and each column represents a feature or variable. Table 1-1 provides a sample of such a database. In Table 1-1, each row represents a claimant. The features are the variables claim number, accident date, claim status, attorney involvement, paid loss, outstanding loss, incurred loss, incurred allocated loss adjustment expenses (ALAE) and claimant state. As seen in Table 1-1, the data contain two key types of variables, quantitative or numeric variables such as incurred losses and incurred expenses and nominal or categorical variables such as claim status and state. Each numeric value denotes a specific quantity or value for that variable. Each value or category, whether numeric or alphanumeric, of a categorical variable embeds a coding that maps the value to one and only one category.[1] This data is structured data. Structured databases result from intentional design where the variables have proscribed definitions and the values of the variables have proscribed meaning.

[1] Note that categorical variables can contain numeric codes as well as string values as in the example. Each code for the categorical variables maps to a value. That is injury '01' may denote a back strain and injury '02' may denote a broken wrist for an injury type variable.

Sample Structured Data

Claim No	Accident Date	Status	Attorney	Paid	Outstanding	Incurred ALAE	Incurred Loss	State
199816	01/08/1999	C	Yes	37,284	0	11,021	37,284	NY
199843	01/16/1999	C	No	0	0	0	0	NY
200229	12/30/2002	O	No	195	0	3	195	CA
199868	09/19/1998	C	Yes	99,852	0	31,807	99,852	NJ
200327	05/19/2003	C	No	286	0	72	286	PA

Table 1-1

Another kind of data that is also present in corporate databases is unstructured data. This data typically has the appearance of free form text data. Examples of text data are claim description fields in claim files, the content of e-mails, underwriters written evaluation of prospective policyholders contained in underwriting files and responses to open ended survey questions on customer satisfaction survey. It has been estimated that 85% of corporate data is of the unstructured type (Robb, 2004). As Mani Shabrang of Dow Chemical says, "We are drowning in information but starved for knowledge" (Robb, 2004).

When data is unstructured there is no obvious procedure for converting the data which is composed of sequences of characters that vary in length and content in apparently random ways, to information that can be used for analysis and prediction. Manual intervention on the part of human beings may be able to convert some unstructured data to structured features which can be used to perform statistical analysis. Derrig *et al.* (1994) provide an example where claims experts reviewed claims files and scored the claims on a number of indicators of suspicion of fraud. Because of the effort required and difficulty of interpreting the unstructured text data, it is typically ignored for doing analysis. If information could be automatically extracted from unstructured data, a significant new source of data could become available to corporations.

In the field of data mining, *text mining* has been attracting increased interest in the insurance industry. Text mining refers to a collection of methods used to find patterns and create intelligence from unstructured text data.

In this paper, the key methods used in text mining will be presented. A simple application to a free form claim description field will be used to illustrate the text mining procedures.

1.1 Research Context

While text mining is relatively new, software for analyzing text data has been available

since the late 1990s from the major statistical software vendors such as SAS and SPSS. One of the most common uses of text mining procedures is in search engine technology. A user types in a word or phrase, which may include misspellings, and the search engine searches through a vast repository of documents to find the most relevant documents. Other applications include:

- Analysis of survey data
    - Text mining is used as an automated approach to coding information from open ended survey questions.
- Spam identification
    - The title line and contents of e-mails are analyzed to identify which are spam and which are legitimate (Hastie *et al.*, 2001).
- Surveillance
    - It is believed that a project referred to as ENCODA monitors telephone, internet and other communications for evidence of terrorism (Wikipedia, 2005).
- Call center routing
    - Calls to help desks and technical support lines are routed based on verbal answers to questions.
- Public health early warning
    - Global Public Health Intelligence Network (GPHIN) monitors global newspaper articles and other media to provide an early warning of potential public health threats including disease epidemics such as SARS, and chemical or radioactive threats. (Blench, 2005).
- Alias identification
    - The aliases of health care and other providers are analyzed to detect over billing and fraud. For instance, a bill may have been submitted by John Smith, J. Smith and Smith, John. The same approaches may be used to identify abuse by claimants, where given claimants submit numerous insurance claims under different aliases.

Text mining has evolved sufficiently that web sites are devoted to it and courses focusing solely on text mining are appearing in graduate school curricula. Text mining also occurs

frequently as a topic at data mining conferences. While the interest in text mining is relatively recent, Weiss *et al.* (2005) point out that text analysis dates back to at least the late 1950s where "automatic abstracting" of text information was studied. In the 1970s and 1980s, artificial intelligence researchers were interested in natural language processing. Many of these early efforts did not yield commercially useful results, so interest in text analysis declined. However, in the 1990s new developments in text mining tools led to a reawakened interest in the field.

In property and casualty insurance, literature on text mining is sparse. In 2002 Ellingsworth described the application of text mining to fraud identification (Ellingsworth, 2002). Kolyshkina (2005) described the use of text mining to create features for identifying serious claims.

This paper attempts to fill a gap in the actuarial literature on text mining. It will show that text mining combines string manipulation functions that are available in many modern programming languages, with commonly available statistical analysis methods.

Many of the statistical procedures described in this paper are described in the statistical and actuarial literature (Hastie *et al.* 2001, Kaufman, 1990) but have not heretofore been applied to unstructured text data. Derrig *et al.* (1994) and Francis (2001, 2003), Hayword (2002) have described analytical methods that can be applied to large insurance databases and are used in data mining. Berry and Linoff, (1997), Kaufman and Rousseeuw (1990) and Hastie *et al.* (2001) described some of the dimension reduction techniques that are utilized in text mining. The description of the methods used to derive information from terms in text data will make use of these dimension reduction techniques.

Much of the text mining literature focuses on search engine and other information retrieval method. This paper focuses, instead, on using text mining for prediction of important business outcomes. It will therefore not cover some of the key approaches that are applied primarily in information retrieval.

1.2 Objective

The objective of this paper is to introduce actuaries and other insurance professionals to the methods and applications of text mining. The paper shows that many of the procedures are straightforward to understand and utilize. Many of the procedures have been known in the statistics discipline for decades. The two methods described are k-means and hierarchical clustering.

1.3 Outline

The remainder of the paper proceeds as follows. Section 2 will discuss the data used in the exposition of the text mining methods. Section 2.1 presents the first phase of text mining: parsing and other string manipulations used to create terms for further analysis. Section 2.2 presents the methods used to create features or variables from the terms extracted in the first phase of the process. These features can then be used to perform additional analysis. The concept of dimension reduction is discussed in section 2.2.1. The two key methods of dimension reduction used in this paper, k-means clustering and hierarchical clustering are discussed in sections 2.2.2 and 2.2.3 respectively. Further considerations such as the number of clusters to retain and cluster naming to provide understanding of the features created by clustering are described in sections 2.2.4 and 2.2.5. Section 2.2.6 presents two simple examples of using variables derived from text mining for prediction. Results of the analysis are summarized and discussed in Section 3. Conclusions are presented in Section 4.

While many details of how text mining is performed will be presented, some analysts will want to acquire software specific for text mining. A discussion of text mining software is presented in the Appendix.

2. BACKGROUND AND METHODS

Text mining can be viewed as having two distinct phases: term extraction and feature creation. Term extraction makes heavy use of string manipulation functions but also applies techniques from computational linguistics. Actual content is a result of the feature creation process. Feature creation applies unsupervised learning methods that reduce many potential features into a much smaller number of final variables. These features are then potentially useable as dependent or predictor variables in an analysis.

The example employed to illustrate text mining uses simulated data from a general liability claims file[2]. The data contains one free form text field: injury description. In this simple example, there is no injury or body part code in the data and the only information about the nature of the injury is the free form text description. The injury description is a very brief description, generally containing only a few words. The data is representative of that which might be available from a small self insured exposure. While many claims

[2] The claim description field is very similar to actual claim descriptions in actual data from small self insurance programs. Other data, such as ultimate losses have been simulated, but are based on relationships actually observed in data.

databases contain larger accident and claim description fields, this data serves as a simple example of how text mining works. An example of the sample text data is shown in Table 2-1.

Sample Claim File Text Data

INJURY DESCRIPTION
BROKEN ANKLE AND SPRAINED WRIST
FOOT CONTUSION
UNKNOWN
MOUTH AND KNEE
HEAD, ARM LACERATIONS
FOOT PUNCTURE
LOWER BACK AND LEGS
BACK STRAIN
KNEE

Table 2-1

The sample data also contains other insurance related information: incurred losses, incurred loss adjustment expenses, accident year, status (open/closed) and whether or not an attorney is involved in the claim. There are approximately 2,000 records in the data. The values in the data are simulated, but are based on relationships observed in actual data for this line of business.

2.1 Term Extraction

During term extraction, character text is first parsed into words. The term extraction process also strips away words that convey no meaning such as "a" or "the". An additional part of the process involves finding words that belong together such as "John Smith".

When data is parsed, string functions are used to extract the words from the character string composing the text data. To do this, spaces, commas and other delimiters must be used to separate words. A simple example of parsing one record using Microsoft Excel string functions with blank spaces as delimiters is illustrated in Table 2.1-1. The total length of the character string is first determined using the "length" function. Then, the "find" function of Excel is used to find the first occurrence of a blank. This is shown in column (3). Next, the substring function is used to extract the first word from the text, using the position of the first blank (column (4)). The remainder of the term, after removing the first word is then extracted, again using the substring function (columns (5) and (6)). The process continues until every word has been extracted. The "iserr" function can be used to determine when no more blanks can be found in the field. The words extracted are shown in the highlighted area of the table.

Example of Parsing Claim Description

Full Description	Total Length	Location of Next Blank	First Word	Remainder Length 1
(1)	(2)	(3)	(4)	(5)
BROKEN ANKLE AND SPRAINED WRIST	**31**	7	BROKEN	24
Remainder 1		**2nd Blank**	**2nd Word**	**Remainder Length 2**
(6)		(7)	(8)	(9)
ANKLE AND SPRAINED WRIST		6	ANKLE	18
Remainder 2		**3rd Blank**	**3rd Word**	**Remainder Length 3**
(10)		(11)	(12)	(13)
AND SPRAINED WRIST		4	AND	14
Remainder 3		**4th Blank**	**4th Word**	**Remainder Length 4**
(14)		(15)	(16)	(17)
SPRAINED WRIST		9	SPRAINED	5
Remainder 4		**5th Blank**	**5th Word**	
(18)		(19)	(20)	
WRIST		0	WRIST	

Table 2.1-1

The result of parsing is data organized in spreadsheet format, i.e., a rectangular matrix containing indicator variables for the words extracted from the text field. For each word found in any record in the data a variable is created. The variable carries a value of 1 if a given record contains the word and a 0 otherwise.

Example of Terms Created

INJURY DESCRIPTION	BROKEN	ANKLE	AND	SPRAINED	WRIST	FOOT	CONTU-SION	UNKNOWN	NECK	BACK	STRAIN
BROKEN ANKLE AND SPRAINED WRIST	1	1	1	1	1	0	0	0	0	0	0
FOOT CONTUSION	0	0	0	0	0	1	1	0	0	0	0
UNKNOWN	0	0	0	0	0	0	0	1	0	0	0
NECK AND BACK STRAIN	0	0	1	0	0	0	0	0	1	1	1

Table 2.1-2

The example above displays data that could be created from an injury description text field. Each claim description is treated as a "bag of words" (Weiss *et al.*, 2005). The matrices resulting from parsing text data are typically sparse. That is, for most of the terms, most of the records contain a zero for that term and only a few records have a one.

The example shown is a relatively simple one. The claim description field is relatively short and contains no delimiters other than a blank space. However, other delimiters such as the comma and period occur frequently and need to be identified also. Some delimiters, such as a single apostrophe (as in I'll) and period (as in etc.) may be part of the words; so the complex rules for finding and using such delimiters must be coded into the program that parses the data.

Certain words occur very frequently in text data. Examples include "the" and "a". These words are referred to as "stopwords". The stopwords are words removed from the term collection because they have no meaningful content. By creating a list of such stopwords and eliminating

them, the number of indicator variables created is reduced. Table 2.1-3 displays a sample of stopwords used in this analysis. Many of these stopwords do not appear in the claim description data, but appear frequently in text data.

Stopwords
A
And
Able
About
Above
Across
Aforementioned
After
Again

Table 2.1-3

Table 2.1-4 below presents a collection of words obtained from parsing the injury description data into single words and removing stop words.

Parsed Words	
HEAD	INJURY
LACERATION	NONE
KNEE	BRUISED
UNKNOWN	TWISTED
L	LOWER
LEG	BROKEN
ARM	FRACTURE
R	FINGER
FOOT	INJURIES
HAND	LIP
ANKLE	RIGHT
HIP	KNEES
SHOULDER	FACE
LEFT	FX
CUT	SIDE
WRIST	PAIN
NECK	INJURED

Table 2.1-4

Other issues affecting the usefulness of the data must be dealt with. One issue is multiple versions and spellings of words. Table 2.1-4 illustrates this. Both L and LEFT are used to denote left, R and RIGHT are used to denote right and the database has both the singular and plural versions of KNEE. In addition, it can be seen from the table that certain "words" stand for the same injury. For example, as a result of abbreviations used, FX and FRACTURE as well as BROKEN all denote the same injury The process referred to as

stemming is used to substitute one word, referred to as a stem (because in the example of knee and knees, both words have the same stem) for all versions of the term.

Once the words have been parsed, stopwords removed and stemming performed, the sparse matrix of term indicators is ready for the next step: feature creation. During the feature creation step, words and sequences of words are classified into groups that contain similar information.

In some text mining applications, especially those that attempt to understand the content contained in large documents, other analysis such as grammatical analysis is performed before progressing to the feature creation step. Such analysis will not be described here as it is not relevant to the example in this paper.

2.2 Feature Creation

Term extraction is the first step in deriving meaning or content from free form text. The next step is feature creation. Thus far, each line of text has been parsed into a "bag of words". The data can be represented as a rectangular array that has indicator variables for each term in the injury description. When we analyze the terms more carefully, we may find that some words such as "back strain" and "neck strain" denote similar injuries and are unlike "head trauma". Thus, occurrence or non-occurrence of specific words may tell us something useful about the nature and severity of the injury.

One of the most common techniques used to group records with similar values on the terms together is known as cluster analysis. Cluster analysis is an example of dimension reduction. Before describing cluster analysis, the concepts of dimension and of dimension reduction are introduced.

2.2.1 Dimension Reduction

Jacoby (1991) describes dimensions as "the number of separate and interesting sources of variation among objects"[3] There are two views as to sources of variation when dealing with a database organized in rectangular spreadsheet format: columns (or variables) and rows (or records). Table 2.2.1-1 displays the two views of dimensionality for a sample claims database. The arrow pointing to the right indicates that each column of data can be viewed as a separate dimension. The downward pointing arrow indicates that each row

[3] Jacoby, p. 27

or claimant also can be viewed as a dimension.

Two Ways of Viewing Dimension in a Database

	CLAIM NUMBER	DATE OF LOSS	STATUS	INCURRED LOSS
				VARIABLES
	1998001	09/15/97	C	407.81
	1998002	09/25/97	C	0.00
	1998003	09/26/97	C	0.00
	1998004	09/29/97	C	8,247.16
	1998005	09/29/97	C	0.00
	1998006	10/02/97	C	0.00
	1998007	10/10/97	C	0.00
	1998008	10/24/97	C	0.00
	1998009	10/29/97	C	21,211.66
	1998010	10/29/97	C	0.00
R	1998011	11/03/97	C	0.00
E	1998012	11/03/97	C	0.00
C	1998013	11/04/97	C	451.66
O	1998014	11/04/97	C	0.00
R	1998015	11/04/97	C	0.00
D	1998016	11/06/97	C	15,903.66
S	1998017	11/11/97	C	465.10

Table 2.2.1-1

Each column contains information about the claimants and is a potential variable in an actuarial or modeling analysis. Each column is a separate dimension. Often in a large database containing hundreds or even thousands of variables, many variables are highly correlated with each other and contain redundant information. The large number of variables can be reduced to a smaller number of components or factors using a technique such as factor analysis. For instance, Figure 2.2-1 displays three of the dimensions related to financial information in the sample claims data; ultimate incurred loss, ultimate allocated loss adjustment expense (ALAE) and ultimate incurred loss plus ALAE. It can be seen from the graph that the three dimensions are correlated, which one would expect, particularly when one of the variables is the sum of the other two. It is common in actuarial analysis (particularly with small databases) to work with only one of these variables; ultimate loss and ALAE. Thus the number of "dimensions" used in the analysis is reduced to one.

Scatterplot of Correlated Dimensions (Variables)

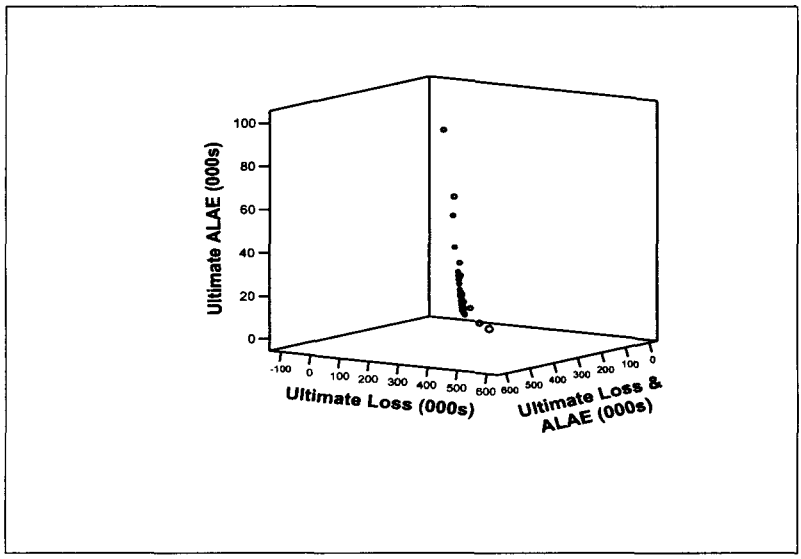


Figure 2.2-1

Each record in the data is also a dimension. When data is aggregated by accident year and development age in order to construct a loss development triangle, row-wise dimension reduction is taking place. The number of dimensions is reduced from the total number of records in the database to the number of cells in the loss development triangle.

2.2.2 K-means Clustering

In statistics, a formal procedure known as clustering is often used to perform dimension reduction along the rows. The objective of the technique is to group like records together. A common application of this method in property and casualty insurance is territory development. Policyholders are grouped into territories according to where they live and the territories are used in ratemaking. Both geographic information such as longitude and latitude and demographic information such as population density can be used for the territorial clustering. Cluster analysis is an unsupervised learning method; there is no dependent variable. Rather, records with similar values on the variables used for clustering are grouped together. In the territory example, policyholders living in high population density zip codes in the southeastern part of a state might be grouped together into one territory. In text mining, clustering is used to group together records with similar words or words with similar meanings.

Many different techniques for clustering exist. One of the most common methods is k-means clustering. When using k-means clustering the analyst specifies the number of clusters he/she wants (a discussion of how to make this choice is deferred until later). A statistical measure of dissimilarity between records is used to separate records that are the most dissimilar and group together records that are the most similar. Different measures of dissimilarity are used for numeric data as opposed to categorical data. Text data is generally viewed as categorical. However, when terms are coded as binary indicator variables, it is possible to apply the techniques that are used on numeric data. Moreover, text mining is commonly applied to documents containing large collections of words, such as academic papers and e-mail messages. Some words appear multiple times in such text data and the number of times a word appears may be recorded and used for text analysis instead of a binary indicator variable. Dissimilarity measures for both numeric and categorical data will be presented.

One of the most common measures of dissimilarity for numeric variables is Euclidian distance. The formula for Euclidian distance is shown below in equation 2.1. The Euclidian distance between two records is based on the variable-wise squared deviation between the values of the variables of the two records.

$$d_{i,j} = \left(\sum_{k=1}^{m} (x_{i,k} - x_{j,k})^2\right)^{1/2} \quad \text{i, j = records, m = number of variables} \tag{2.1}$$

The second dissimilarity measure Manhattan distance is shown in equation 2.2. This measure uses absolute deviations rather than squared deviations.

$$d_{i,j} = \sum_{k=1}^{m} | x_{i,k} - x_{j,k} | \quad \text{i, j = records, m = number of variables} \tag{2.2}$$

Table 2.2.2-1 displays a calculation of both measures using two sample records. The first record has injury "broken ankle and sprained wrist" and the second record has injury "contusion to back of leg". Binary variables indicating the presence or absence of words after the parsing of text are the variables used in the measure.

Euclidian and Manhattan Distance Between Two Records

Variable	Record 1 BROKEN ANKLE AND SPRAINED WRIST	Record 2 CONTUSION TO BACK OF LEG	Squared Difference	Absolute Difference
Back	0.000000	1.000000	1	1
Contusion	0.000000	1.000000	1	1
Head	0.000000	0.000000	0	0
Knee	0.000000	0.000000	0	0
Strain	0.000000	0.000000	0	0
Unknown	0.000000	0.000000	0	0
Laceration	0.000000	0.000000	0	0
Leg	0.000000	1.000000	1	1
Arm	0.000000	0.000000	0	0
Foot	0.000000	0.000000	0	0
Hand	0.000000	0.000000	0	0
Ankle	1.000000	0.000000	1	1
Shoulder	0.000000	0.000000	0	0
Hip	0.000000	0.000000	0	0
Left	0.000000	0.000000	0	0
Neck	0.000000	0.000000	0	0
Wrist	1.000000	0.000000	1	1
Cut	0.000000	0.000000	0	0
Fracture	1.000000	0.000000	1	1
Surgery	0.000000	0.000000	0	0
Finger	0.000000	0.000000	0	0
None	0.000000	0.000000	0	0
Broken	1.000000	0.000000	1	1
Trauma	0.000000	0.000000	0	0
Lower	0.000000	0.000000	0	0
Right	0.000000	0.000000	0	0
Total			7	7
Distance Measure			2.65	7

Table 2.2.2-1

Dissimilarity measures specific for categorical variables also exist: Table 2.2.2-2 displays the notation for comparing two records on all their binary categorical variables. For instance, the sum of all variables for which both records have a one is shown as "a" on the table. The counts of variables on which the two records agree are denoted "a" and "d". The counts of variables on which the two records disagree are denoted "b" and "c".

Crosstabulation of Counts for Two Records Binary Variables

		Record 1	
		1	0
Record 2	1	a	b
	0	c	d

Table 2.2.2-2

Simple matching is a dissimilarity measure that compares the total number of non matches to the total number of variables as shown in equation 2.3.

$$d_{i,j} = \frac{b+c}{a+b+c+d} \quad \text{i,j = records} \tag{2.3}$$

Another dissimilarity measure, shown in equation 2.4, is Rogers and Tanimoto. This measure gives more weight to disagreements than to agreements. In the example above (Table 2.2.2-1) where there are 7 disagreements and 19 agreements, the Rogers and Tanimoto dissimilarity measure is 0.43.

$$d_{i,j} = \frac{2(b+c)}{a+d+2(b+c)} \quad \text{i,j = records} \tag{2.4}$$

Instead of using a dissimilarity measure, some clustering procedures use a measure of similarity. A common measure of similarity is the cosine measure. The cosine statistic is a measure of covariance, but it is applied to records rather than to variables.

$$\text{cosine}_{i,j} = \frac{\sum_{k=1}^{m} (x_{i,k} * x_{j,k})}{\sqrt{\sum_{k=1}^{m} (x_{i,k}{}^2)} \sqrt{\sum_{k=1}^{m} (x_{j,k}{}^2)}} \quad \text{i,j = records, m = number of variables} \tag{2.5}$$

Rather than use binary indicator variables in the cosine calculation, this statistic typically uses a value referred to as the tf-idf statistic as x_{ij} in equation 2.5. The tf-idf (term frequency – inverse document frequency) statistic is based on the frequency of a given term in the

record. The static is normalized by being divided by the total number of times term appears in all records[4].

$$tf = \frac{n_i}{\sum_k n_k} \quad n_i = \text{number of times term i occurs,}$$
$$\text{tf-idf} = \frac{\text{tf}}{Df} \quad Df \text{ is the document frequency} \tag{2.6}$$

There are several ways to count the document frequency (denoted Df in equation 2.6) or the frequency of a term in a database (Wikipedia, 2005). A common method counts the number of records[5] in which the term appears divided by the total number of records. Sometimes the log of the inverse of the document frequency is used in the calculation. This statistic is more appropriate for applications involving larger collections of words, i.e., where each record is an entire document. The tf-idf method was not used in the analysis in this paper.

K-means clustering using Euclidian distance was applied to the matrix of extracted terms from the injury descriptions. Each cluster that is created from a k-means clustering procedure has a center referred to as the centroid. The centroid is the vector of average values for the cluster for each variable entering the clustering procedure. In the case of binary variables coded as either zero or one, the centroid is the cluster's frequency for each term or the proportion of all records in the cluster which contain the term. For example, the clustering procedure was used to create two classes or clusters. The clusters' frequencies for each term are displayed in the Table 2.2.2-3. From the table, it can be seen that none of the claims in Cluster 1 contain the word "back" and all of the claims in Cluster 2 contain the word. In addition, Cluster 1 contains a much higher percentage of claims with the words "contusion" "unknown" and "laceration" while Cluster 2 contains a much higher proportion of records with the word "strain". Thus, when k-means clustering is used to create two clusters, a cluster with a high representation of claims with back injuries is partitioned from claims with other injuries.

[4] In much of the text mining literature, the term "document" is a synonym for "record", because the unit of observation is often an entire document, such as a newswire article

[5] Frequently when this statistic is used, each record is a document. See footnote 4 above.

Frequencies for Two Clusters

Cluster Number	back	contusion	head	knee	strain	unknown	laceration
1	0.00	0.15	0.12	0.13	0.05	0.13	0.17
2	1.00	0.04	0.11	0.05	0.40	0.00	0.00

Table 2.2.2-3

Frequency statistics for three clusters are displayed in Table 2.2.2-4. Again, one group, Cluster 2, is a cluster with 100% back injuries. Cluster 3 contains a high proportion of claims with knee injuries, while contusions and unknown are the most common injuries in Cluster 1. As will be discussed in more detail in Section 2-5, examining cluster statistics such as those in Tables 2.2.2-3 and 2.2.2-4 assist the analyst in assigning labels to cluster.

Frequencies for Three Clusters

Cluster Number	back	contusion	head	knee	strain	unknown	laceration
1	0.00	0.17	0.14	0.04	0.05	0.16	0.19
2	1.00	0.04	0.11	0.05	0.40	0.00	0.00
3	0.00	0.07	0.04	0.48	0.09	0.00	0.05

Table 2.2.2-4

Viewing the statistics for three clusters versus two clusters, it is clear that there is refinement in the definition of the injury clusters when progressing from two to three clusters. Determining how many clusters to use is something of an art. If too many clusters are estimated, the model is over parameterized and is fitting noise as well as pattern. If too few clusters are created, the data are not adequately modeled. This topic is discussed in more detail in Section 2.2.4.

2.2.3 Hierarchical Clustering

Though less common than k-means clustering, hierarchical clustering is another common method applied in text mining to cluster terms in order to discover content (in this case, to create features that can be used for further analysis). Hierarchical clustering is a stepwise procedure that begins with many clusters and sequentially combines clusters in close proximity to each other until no further clusters can be created. Typically, hierarchical clustering begins with every observation as a single cluster and terminates with one cluster

containing all the records.[6] Hierarchical clustering procedures produce dendograms or tree-like visualizations of the stages of clustering which assist the analyst in determining the final number of clusters to select. Hierarchical clustering can be applied to either records or variables. Because the visualization of the results is easier to display, the results of clustering by variable are displayed in Figure 2.2-2. The figure displays the tree like figure or dendogram that results from clustering ten of the injury terms. For this dendogram, Euclidian distance was used.

The dendogram displays the cluster groupings created at each step of the clustering process. The left-hand side of the dendogram under the label CASE lists the variables and their numbers (based on order in the database). The left-most portion of the dendogram is a line representing a terminal branch of the tree. There is one branch for each variable, as each variable is its own cluster at the beginning of the clustering process. Moving to the right, the branch for arm and the branch for foot are connected, indicating that a new cluster is created in step one by combing arm and foot. Forming a cluster composed of these two variables indicates that the distances between arm and foot are smaller than the distances between any other possible combination of two variables. Next on the dendogram, the branch for leg is connected to the branch containing arm and foot, indicating that at the second step, a new cluster is created by combining leg with the arm-foot cluster. The stepwise process of creating new clusters by combining together smaller clusters at each step continues until there is only one cluster containing all the variables. This is shown at the right side of the dendogram where a branch containing back and strain are connected to a branch containing all other variables (i.e., at the next to last step the two group cluster partitions the terms "back" and/or "strain" from all other injuries).

Table 2.2.3-1 presents a matrix of proximity (i.e. distance) measures which underlie the dendogram clusters. These are the distances used to cluster the variables. For example, the distances between arm and foot, which are clustered together in the first step is 6.708. This compares to the distance of 8.888 between arm and back, which only cluster together in the last step, where all variables are combined into one cluster.

[6] Hierarchical clustering can also proceed in the opposite direction, from one cluster with all the data to many clusters

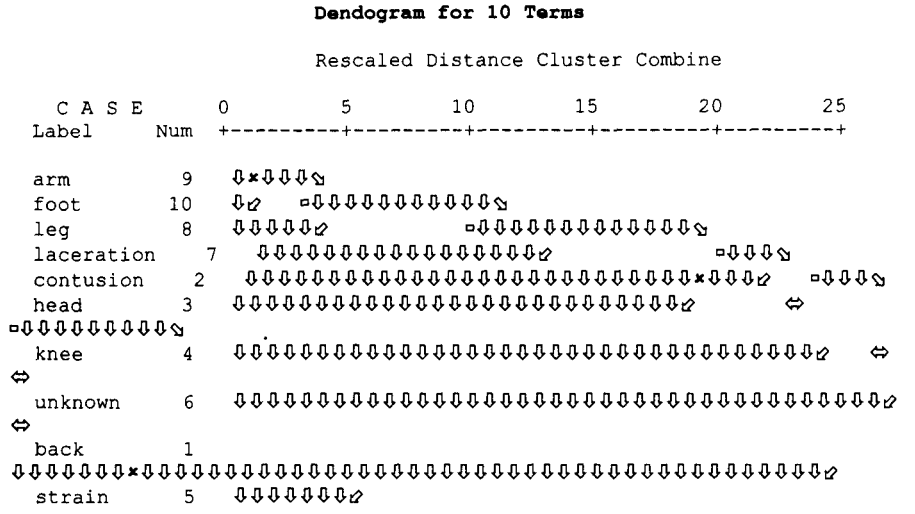


Figure 2.2-2

Proximity Matrix

	Matrix File Input									
Case	Back	Contusion	Head	Knee	Strain	Unknown	Laceration	Leg	Arm	Foot
back	0.000	10.000	9.327	9.539	7.071	9.747	9.747	9.055	8.888	8.602
contusion	10.000	0.000	7.937	8.062	9.274	9.220	9.110	8.246	8.185	7.211
head	9.327	7.937	0.000	9.055	9.000	8.944	8.124	8.185	8.000	7.810
knee	9.539	8.062	9.055	.000	8.307	8.832	8.602	8.307	8.000	7.681
strain	7.071	9.274	9.000	8.307	0.000	8.775	8.775	7.746	7.810	7.616
unknown	9.747	9.220	8.944	8.832	8.775	0.000	8.718	8.185	8.000	7.550
laceration	9.747	9.110	8.124	8.602	8.775	8.718	0.000	7.550	8.000	7.141
leg	9.055	8.246	8.185	8.307	7.746	8.185	7.550	0.000	7.000	6.928
arm	8.888	8.185	8.000	8.000	7.810	8.000	8.000	7.000	0.000	6.708
foot	8.602	7.211	7.810	7.681	7.616	7.550	7.141	6.928	6.708	0.000

Table 2.2.3-1

The hierarchical clustering of the terms in the data provides insight into word combinations in the data that tend to occur together and or tend to be associated with similar injuries. However, for the purpose of classifying records into injury categories, it is more typical to cluster case-wise rather than variable-wise. Thus, hierarchical clustering was also used to cluster the injury description records.

2.2.4 Number of Clusters

Determination of the number of clusters to retain is often something of art. One approach involves viewing the cluster centers to determine if the clusters from a given grouping appear meaningful. Another procedure for determining the number of clusters involves comparing the performance of different clustering schemes on an auxiliary target variable of interest. Here, ultimate incurred losses and ALAE is one variable of interest that may help in the decision. Figure 2.2-3 displays how the mean ultimate loss and ALAE varies by cluster for four and eight cluster groupings.

A forward stepwise regression was run to determine the best cluster size. Stepwise regression is an automated procedure for selecting variables in a regression model. Forward stepwise regression begins with a null model or model that has no predictors. The procedure then tests all possible independent variables that can be used on a one-variable regression model. The variable which improves the goodness of fit measure the most is the variable entered in step one. In step two, all 2-variable regressions are fit using the variable selected in step one and the variables not selected in step one. The variable which produces the largest improvement in goodness of fit is then selected and entered into the model. The process continues until no further significant improvement in fit can be obtained.

Average Ultimate Loss and ALAE by Cluster for 4 and 8 Clusters

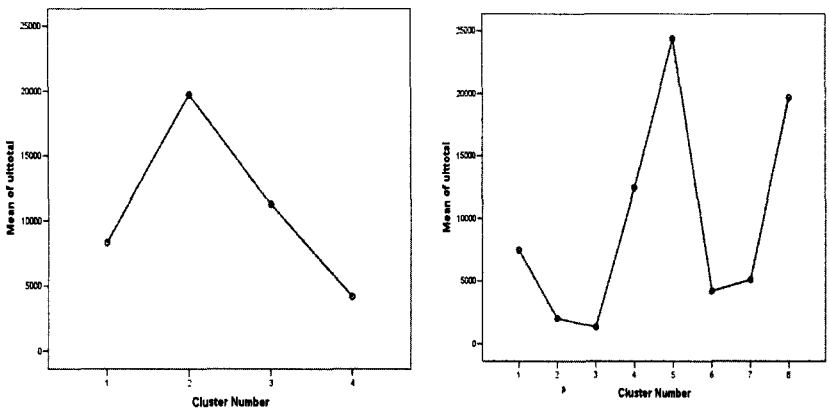


Figure 2.2-3

A common goodness of fit measure used in stepwise regression is the F-statistic:

$$F = \frac{\text{MS Regression}}{\text{MS Residual}} = \frac{SS_{reg} / p}{SS_{res} / (N - p - 1)} \qquad (2.7)$$

where p=number of variables,
N=number of observations,
SS=sum of squared deviation

The F statistic is the ratio of mean squared error of the regression (the amount of variance explained by the regression) divided by the mean square error of the residual (the amount of unexplained variation). When used in stepwise regression, after the first variable is entered, the change in F statistic is used. The user typically selects a significance level such as 5% that is used as a threshold for entering variables into the regression.

When using stepwise regression to select the number of clusters to use, the possible predictor variables in the regression are the clusters created by 2 category cluster, 3 category cluster, etc. Since the objective is to find the optimal number of clusters, the regression is run on each of the category cluster variables and the category cluster with the best fit is selected. For the purposes of this paper, only the first step of the stepwise regression was

performed, i.e., only the one variable supplying the best fit of all the one variable regressions was retained.[7] Stepwise regression provides a quick and efficient method for determining the number of clusters. The stepwise procedure determined that a regression with seven groups produced the best fit. Regression also ascertained that k-means clustering produced clusters that were better predictors of ultimate losses and ALAE than hierarchical clustering.

A more formal approach is to use a statistical test to determine the optimum number of clusters. One such test is the BIC (Swartz Bayesian Information Criterion) (Chen and Gopalakrishnan, 2004). The statistic is used to compare two models at a time. The statistic chooses between a simpler model and a more complex model by comparing their adjusted or penalized likelihood function. A penalty related to the number of variables in the model is applied in order to control for overfitting. When applying the statistic, it is common to treat the data as if from a multivariate normal distribution:

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad (2.8)$$

where $\mathbf{X}$ is a vector of random variables $\boldsymbol{\mu}$ is the centroid (mean) of the data and $\boldsymbol{\Sigma}$ is the variance-covariance matrix

[7] Because of the small size of the data only one injury code variable was created.

The formula for the BIC statistic is:

$$BIC = \log L(X, M) - \lambda \frac{1}{2} \text{p*log(N)} \qquad (2.9)$$

where log(L(X,M)) is the logliklihood function for a model, p is the number of parameters, N the number of records, λ is a penalty parameter, often equal to 1

For a cluster analysis, each cluster has a likelihood function based on the cluster's centroid and variance-covariance matrix. For instance, in comparing a one-group cluster to a two-group cluster, a likelihood based on the overall centroid of all the data and the overall variance-covariance matrix is compared to a two group likelihood based on the centroids and variance-covariance matrices of the two clusters. The second model has twice as many parameters as the first. If the BIC increases significantly using two clusters compared to one cluster, a two group clustering is preferred.

Most of the software used in this analysis did not implement the BIC statistic to determine cluster size. However, one procedure, the SPSS two-step clustering procedure intended for categorical and mixed categorical-numeric data did implement the procedure. A two-step clustering procedure breaks the clustering process into two steps 1) create a dissimilarity matrix which may be done differently for categorical as opposed to numeric data and 2) use the dissimilarity matrix to cluster the data. When this procedure was applied, it produced a clustering with three groups. The two-step clusters had a significant correlation with ultimate losses and ALAE, though this correlation was not as high as that for the best k-means cluster.

The end result of clustering of the claim description field in the data is to introduce one new feature or variable. This variable is a categorical variable indicating to which of the cluster groupings or classes a record is assigned. This new variable can be viewed as an injury type coding. In the application in section 2.2.6, the seven cluster grouping will be used, but other choices could have been made. Note that while only one cluster grouping of the injury descriptions was selected, there may be situations where the analyst prefers to use multiple new features derived from the clustering procedure, each with a different number of groups.

2.2.5 Naming the Clusters

For each cluster, it can be informative to determine which word or words are important in defining the cluster. Examining the frequencies of each word for each of the clusters can be used to gain insight into the clusters. Figure 2.2-4 displays the frequencies of the words "back" and "strain" for the seven-group cluster. The graph is a population pyramid. The

graph displays visually with bars a crosstabulation of back versus strain by cluster group. That is, a bar displays the count for a zero or one on back, versus zero or one on strain for each of the seven injury cluster groupings. The bars appearing under one for back, one for strain or one for both back and strain denote injury groups that contain the words back, strain or both. From the graph it can be seen that Cluster 4 has a relatively high count of both the words back and strain and Cluster 6 has a high representation of the word back, but not strain.

Table 2.2.5-1 presents frequencies of key words for the seven-group cluster. The table displays the proportion of records for each cluster which contain the words. Words that have high representation within a cluster have been highlighted. From the table it can be seen that Cluster 1 has a high representation of the word unknown. Cluster 2 has a high representation of the word contusion. Cluster 4 has a high representation of the words back and strain. Cluster 7 also has a high representation of the word strain, but the word back has a low representation. A conclusion is that Cluster 4 appears to be back strains while Cluster 7 is largely other strains, and includes a high representation of the word leg.

Frequencies of the Words Back and Strain by Cluster

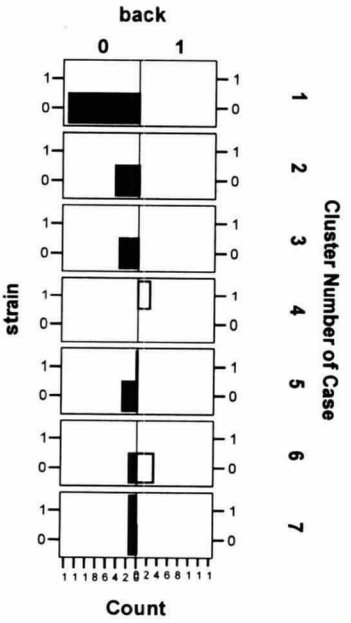


Figure 2.2-4

Frequency of Terms by Cluster

Cluster	Back	Contusion	head	knee	strain	unknown	laceration	Leg
1	0.000	0.000	0.000	0.095	0.000	**0.277**	0.000	0.000
2	0.022	**1.000**	0.261	0.239	0.000	0.000	0.022	0.087
3	0.000	0.000	0.162	0.054	0.000	0.000	**1.000**	0.135
4	**1.000**	0.000	0.000	0.043	**1.000**	0.000	0.000	0.000
5	0.000	0.000	0.065	**0.258**	0.065	0.000	0.000	0.032
6	0.681	0.021	**0.447**	0.043	0.000	0.000	0.000	0.000
7	0.034	0.000	0.034	0.103	**0.483**	0.000	0.000	**0.655**
Weighted Average	0.163	0.134	0.120	0.114	0.114	0.108	0.109	0.083

Table 2.2.5-1

A procedure involving tabulation of the frequency of the words within the cluster can be automated. The most commonly occurring words can be identified and used to label the cluster.

2.2.6 Using the Features Derived from Text Mining

A major objective of text mining is to create new information that has predictive value. The simple illustration in this paper mined an injury description field and assigned each claim to one of seven cluster groups based on the words in the injury description. The cluster group is a new independent variable that can be used to predict a dependent variable of interest to the analyst. Potential variables of interest in a claims database include financial variables such as losses and loss adjustment expenses, whether or not there has been subrogation or recovery on the claim and whether or not the claim is likely a fraud or abuse claim. The database used in this exercise is representative of what might be available in cases where a third party claims adjuster supplied data to a self insured entity. It is therefore smaller and less rich than what one would find in a large insurance company database. This simple example focuses on the financial variables in the data.

The example uses the new injury feature added by the text mining procedure to predict the likelihood that a claim will be a serious claim. One application of data mining in the literature (Derrig, 2004) uses models to score claims early in the life of the claim. The objective is to identify claims that are likely to be the most costly to the company and apply more resources to those claims. For this analysis, a serious claim is defined as a claim whose total losses plus allocated loss adjustment expenses exceeds $10,000. Approximately 15% of the claims in the data exceed this threshold. A histogram of claim severities is shown in Figure 2.2-5. The histogram indicates that the severity distribution is right skewed and heavy tailed. Approximately 98% of loss dollars are due to claims defined as serious. (See the pie chart in Figure 2.2-6).

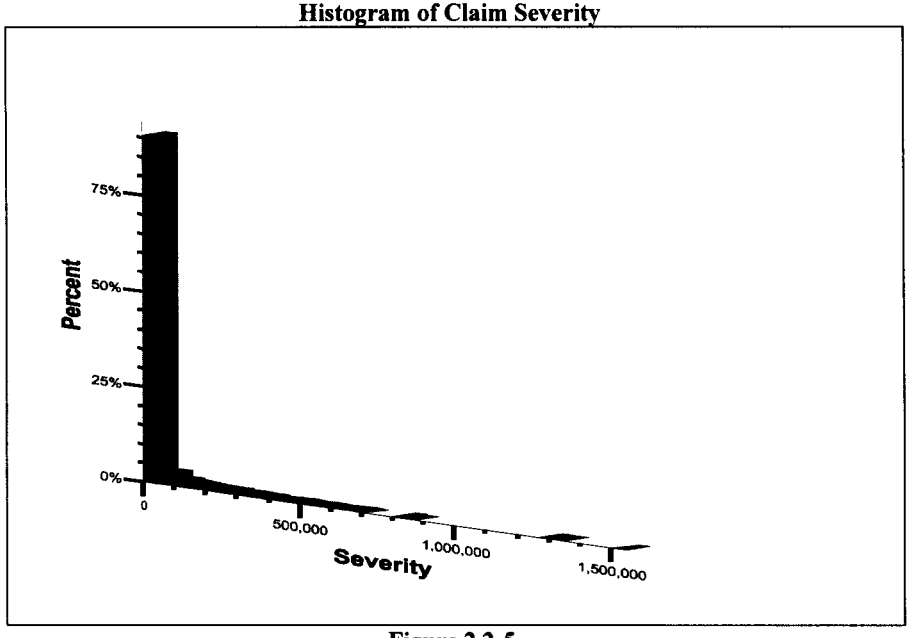

Histogram of Claim Severity
Percent
75%
50%
25%
0%
0
500,000
1,000,000
1,500,000
Severity

Figure 2.2-5

Percent of Loss Dollars: Serious vs. Non Serious Claims

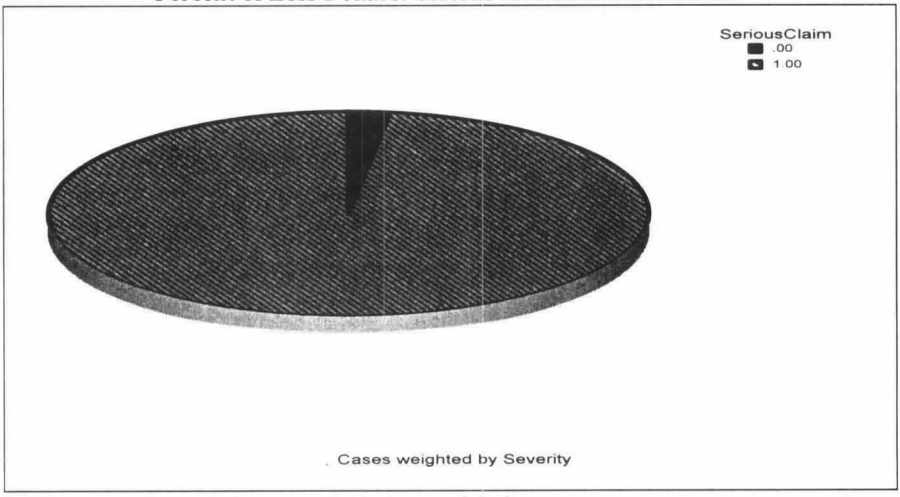


Figure2.2-6

Logistic regression was used to predict the likelihood of a claim being a serious claim. Much of the current actuarial literature dealing with modeling large complex databases focuses on generalized linear models (See CAS Discussion Paper Program, 2004). A modeling procedure that is one of the options found within the family of generalized linear models is logistic regression. The logistic regression procedure functions much like ordinary linear regression, but under logistic regression the dependent variable is categorical or ordered categorical, not numeric. Logistic regression is a popular choice from the family of generalized linear models for performing classification. With categorical variables, a value of one can be assigned to observations with a category value of interest to the researcher (i.e., serious claims) and zero to all other claims. Typically the objective is to score each observation with a probability the claim will fall into the target category, category one. The probability the claim will have a value of 1 lies in the range 0 to 1. This probability is denoted p(y). The model relating p(y) to a vector of independent variables **x** is:

$$\ln\left(\frac{p(y)}{1-p(y)}\right) = b_0 + b_1X_1 + b_2X_2 \ldots + b_nX_n \qquad (2.10)$$

The ratio $\frac{p(y)}{1-p(y)}$ is referred to as the odds ratio and the quantity $\ln\left(\frac{p(y)}{1-p(y)}\right)$ is known as the logit function or logit transformation.

The reader is referred to the extensive literature on logistic regression for further details (Hosmer 1989, Venibles and Ripley 1999). Once a linear model has been fit, the predicted value will be on the logit transformed scale. To use the predictions as probabilities, they must be transformed back to the original scale. If $\hat{f}(\mathbf{x})$ is the logistic predicted value, the transformation $e^{\hat{f}(\mathbf{x})}/(1+e^{\hat{f}(\mathbf{x})})$ must be applied.

Other analytical methods such as CART (Brieman *et al.*, 1990) could also be applied although the data in this example likely does not lend itself to complex approaches meant for larger, complex databases. Two variables were used to predict the probability of a serious claim: attorney involvement and the injury variable derived from text mining. Because the sample database is relatively small only a main effects[8] model was fit (a model with interaction terms was tested and found not to be significant). This means the model fit was of the following form:

$$Y = B_0 + B_1 Attorney + B_2 Injury\_Group \tag{2.11}$$

The injury group used in the model is the injury grouping of seven clusters created by the text mining analysis. The attorney variable denotes whether an attorney is involved in the claim.

The logistic regression found both variables to be significant. The table below shows the average model probability of a serious claim for both the serious and non-serious claims. It can be seen that the model predicts a much higher probability, on average, for the serious groups of claims than the non-serious group of claims.

[8] In a main effects model there are no interactions incorporated into the model.

Mean Probability of Serious Claim vs. Actual Value

	Actual Value	
	1	0
Avg Prob	0.31	0.01

Table 2.2.6-1

One other application of the text variable is illustrated. A simple analysis of variance (ANOVA) was used to predict ultimate losses and ALAE. An ANOVA is a linear model where the dependent variable is numeric and the independent variables are categorical. ANOVA is like a linear regression with categorical predictor variables. The form of the model is:

$$Y = B_0 + B_1 Attorney + B_2 Injury\_Group + B_3 Attorney * Injury\_Group \quad (2.12)$$

where Y is ultimate losses and ALAE trended to a common date

Note this model includes the interaction term attorney * injury group. The results are displayed in the Table 2.2.6-2. In this regression, both attorney involvement and injury group as well as the interaction between attorney and injury are significant. As an alternative to the classification procedure illustrated in the previous example, such a regression could be used to identify serious claims (i.e., the claims with high predicted values for the ultimate losses and ALAE). Another application of models that predict ultimate losses and ALAE is estimating reserves for insurance exposures. Heckman (1999) and Taylor (2004) introduced methods of reserving that utilized individual claims data. There are two components to using claim data to estimate ultimate losses for a reserving application:

- Estimate ultimate losses for claims already reported using the individual information for each claim's independent variables. Historic information on more mature claims is used to develop a model for less mature claims
- Estimate ultimate losses for claims that have occurred, but because of a long reporting lag, have not yet been reported. In order to estimate ultimate values for unreported claims the actuary needs:
    - An estimate of unreported claims (perhaps derived from a claim development triangle)

- An estimate of the proportion of claims in each category of the key predictor variables (i.e., an estimate of the proportion within each attorney/injury type combination). Recent historical patterns could be used to derive such estimates

Results for Regression of Ultimate Losses and ALAE on Attorney and Injury

Parameter Estimates

Dependent Variable: Ultimate Loss & ALAE

Parameter	*B*	*Std. Error*	*t*	*Sig.*	*95% Confidence Interval*	
					Lower Bound	Upper Bound
Intercept	2975.08	74.56	39.90	0 00	2790.20	3159.97
[attorney=.000000]	-2924.58	453.27	-6.45	0.00	-4048.49	-1800.67
[attorney=1.000000]	0.00	.	.	.	.	.
[QCL6= 1]	18426.20	80.08	230.10	0.00	18227.64	18624.77
[QCL6= 2]	10504.67	153.40	68.48	0.00	10124.31	10885.03
[QCL6= 3]	6506.90	214.04	30.40	0.00	5976.17	7037.63
[QCL6= 4]	1175.95	112.17	10.48	0.00	897.81	1454.08
[QCL6= 5]	37081.94	89.64	413.67	0.00	36859.67	37304.22
[QCL6= 6]	74620.90	79.82	934.92	0.00	74422.99	74818.81
[QCL6= 7]	0.00	.	.	.	.	.
[attorney=.000000] * [QCL6= 1]	-16537.20	530.17	-31.19	0.00	-17851.81	-15222.59
[attorney=.000000] * [QCL6= 2]	-10123.91	556.53	-18.19	0.00	-11503.88	-8743.95
[attorney=.000000] * [QCL6= 3]	-3934.19	607.54	-6.48	0.00	-5440.64	-2427.74
[attorney=.000000] * [QCL6= 4]	-675.90	719.76	-0.94	0.35	-2460.60	1108.80
[attorney=.000000] * [QCL6= 5]	-36860.96	673.40	-54.74	0.00	-38530.71	-35191.21
[attorney=.000000] * [QCL6= 6]	-63147.92	567.22	-111.33	0.00	-64554.39	-61741.44
[attorney=.000000] * [QCL6= 7]	0	.	.	.	.	.
[attorney=1.000000] * [QCL6= 1]	0	.	.	.	.	.
[attorney=1.000000] * [QCL6= 2]	0	.	.	.	.	.
[attorney=1.000000] * [QCL6= 3]	0	.	.	.	.	.
[attorney=1.000000] * [QCL6= 4]	0	.	.	.	.	.
[attorney=1.000000] * [QCL6= 5]	0	.	.	.	.	.
[attorney=1.000000] * [QCL6= 6]	0	.	.	.	.	.
[attorney=1.000000] * [QCL6= 7]	0	.	.	.	.	.
A	This parameter is set to zero because it is redundant.					

Table 2.2.6-2

3. RESULTS AND DISCUSSION

In this paper, a very simple example of text mining was used as an illustration of the underlying concepts and methods. The illustration has shown that the basic procedures underlying text mining are straightforward to understand and implement. The two key technologies that are used are 1) string manipulation and processing functions that are part of nearly all programming languages and 2) classical statistical procedures for dimension reduction, such as clustering, that are included within nearly all statistical software packages.

In the illustration, text mining was used to add an injury type code to a database that contained only a free form text field describing the injury.

The injury description was then used as an independent variable in two simple predictive models. In more realistic situations text mining has the potential to add significantly to the information available to the analyst in large modeling projects. For instance, many large insurance company databases contain one or more free form claim description fields or narratives describing the accident and the circumstances associated with the accident. These narratives often contain information not contained in injury, cause of loss or other coding. This may be particularly true when new types of claims or new patterns of claiming behavior are beginning to emerge Ellingsworth and Sullivan (2003) describe applications used to provide an understanding of rising homeowner claims and suspicious and possibly fraudulent auto claims at a large insurance company. When analytical approaches using only structured information coded into the company's database were unsuccessful in explaining the patterns, they turned to text mining. Ellingsworth and Sullivan provided the following hypothetical example of text from a claim description field:

> "The claimant is anxious to settle; mentioned his attorney is willing to negotiate. Also willing to work with us on loss adjustment expenses (LAE) and calculating actual cash value. Unusually familiar with insurance industry terms. Claimant provided unusual level of details about accident, road conditions, weather, etc. Need more detail to calculate the LAE."

Certain terms in the text such as "anxious", "settle" and "familiar" may provide clues to suspicious claims that cannot be found in the structured data in the claims database. Mining the text data for such terms significantly improved the ability of Ellingsworth and Sullivan to model the patterns in the data.

Text mining has become sufficiently prominent that the major vendors of statistical and data mining software tools (such as SAS, SPSS and Insightful) offer text mining products. Some of these tools are very powerful and are capable of processing data from large document collections. While a discussion of software tools for text mining is postponed to the Appendix of this paper, acquisition of powerful text mining software may be unnecessary for smaller applications such as in this paper. That is, when the "documents" being analyzed are relatively modest in size, as many claim description data are, methods developed for applications on larger documents such as academic papers and news service articles may be more than is needed. The analyses in this paper were performed using free text mining software along with statistical procedures available in SPSS13.0 and S-PLUS 6.2. The author believes that there are many situations where text mining can be used to augment the amount of information available for analysis and that for smaller applications, it is

unnecessary to acquire expensive specialized tools.

4. CONCLUSIONS

The purpose of this paper is to educate actuaries on the potential for using text mining for insurance applications. That is, the intent is to provide a basic introduction to the new area of text mining. It was shown that relatively uncomplicated methods underlie the main procedures used to perform text mining. It is widely believed that a large percentage of data is contained in unstructured form. Text mining has the potential to add significantly to the amount of data available for analysis. Some of this data includes adjuster claim description notes, loss prevention specialist notes and underwriter notes.

The field of text mining is one that is undergoing rapid development. New methods are being developed to improve on simple clustering as a means of classifying text data. These include methods based on discriminant analysis (Howland and Park, 2004), methods that use principal components analysis and single value decomposition (Snellert and Blondel, 2004), and linkage based methods that dynamically update (Aggarwal, 2005). Note that the methods used in this paper perform row-wise dimension reduction and cluster similar records. Methods based on factor analysis, principal components analysis and single value decomposition can perform column-wise or term-wise dimension reduction. While these methods were not described or illustrated in this paper, they show promise for improving the classification of text information. Another area under development that may expand the applicability of text mining is handwriting recognition and optical character recognition (Wikipedia, 2005). Many PDAs read handwritten entries. Microsoft Windows XP and Office XP also have handwriting recognition capability. Improvements in handwriting and optical character recognition could permit scanning and mining of handwritten and typed notes currently stored in paper files and not currently accessible from computerized databases.

Acknowledgment

The author acknowledges the assistance of Michael Francis, Virginia Lambert, Rudy Palenik and Jane Taylor in editing this paper.

Appendix A – Software for Text Mining

This appendix describes the author's experiences with several text mining tools. The tools covered are 1) commercial text mining products, 2) a free text mining tool and 3) programming languages. The author found only categories 2) and 3) to be useful in this analysis, although commercial text mining products may prove invaluable in tasks involving

larger, more complex data sets.

The task of locating appropriate software for use in text mining proved to be something of a challenge. A number of software options were investigated in preparation for undertaking the analysis in this paper. Davi *et al.* (2005) gave a favorable review to two text mining packages; WordStat and SAS Text Miner. As the SAS Text Miner package is sold bundled with the Enterprise Miner, a large, relatively expensive data mining suite intended for large application, no attempt was made to acquire or test it. Therefore WordStat, a modestly price product was investigated. The WordStat web site allows prospective customers to download a trial version of the software. The user can use the demo software for 30 days or 10 uses. The latter limitation of 10 uses proved to be the more severe limiting factor. During this study about 5 of the 10 uses were consumed in figuring our how to read data into the text mining module of the software. Once the data were read, simple term extraction was performed and some simple descriptive statistics were created. However, the author was unable to apply clustering procedures to create an injury description feature or to output terms for analysis in other software. Thus, other options were investigated as WordStat was unable to provide the functionality needed for this study.

Other vendors of text mining software (SPSS and Insightful) felt their text mining software was inappropriate for the purposes of this study[9]. After relatively little success with other options, the free package TMSK was used to perform many of the tasks for the text mining analysis in this paper.

TMSK is a free product available to purchasers of the book Text Mining (Weiss *et al.* 2005). It can be downloaded from the author's web site using passwords supplied with the book. This software is very handy for performing term extraction. It comes with lists containing stop words and stem words that are automatically applied during running of the program and can be used to do feature creation using k-means clustering. Certain other analytical tasks not covered in this paper are also included. However, a certain amount of persistence is required to obtain useful results from the software. Some of these features of this program the user needs to be aware of are:

- The user must have the programming language Java on his/her computer. Java can be downloaded for free from the Sun Microsystems web site: http://java.sun.com/.

[9] The software is intended primarily for much larger scale complex applications, and is intended for use on a server making it difficult to install and use initially. Many of these are not major issues when being applied to large scale applications for which these packages are intended.

- The program will only run in DOS mode, i.e., in the command window. On many windows systems the command prompt is accessed by looking under accessories in the program listing.
- The program will only read xml files. For this analysis, the injury description field of the example data was saved to an xml file format within Microsoft Excel. More recent versions of Adobe Acrobat can also save text in xml format.
- The results of term extraction are output to what is referred to as a "sparse vector". Table A-1 displays a snapshot of what a sparse vector looks like. The sparse vector is a condensed representation of the terms extracted, containing an entry only when the term is present for the record. The notation on the first row of Table A-1 indicates that for record 1 of the example data, Term 15 occurred once, Term 20 occurred once and Term 21 occurred once. The analytical procedures included with TMSK read and process the sparse vector data. However, in order to use a statistical procedure other than the ones that come with TMSK, it is necessary to read and parse this output in some other programming language and associate the correct term and correct record with the position indicator and row from the table.
- The manual indicates that the user can add additional stem words to the list maintained by TMSK. However, during this analysis, this feature did not appear to function, so some additional stemming was performed in other software.

Sparse Vector Representation of Terms Extracted

15@1	20@1	21@1
1@1	2@1	8@1
6@1		
1@1	23@1	
1@1		

Table A-1

Most of analysis after term extraction was performed in SPSS and S-PLUS. (TMSK could have been used for clustering, but more output and analysis than this package provides was needed for this paper). Most general purpose statistical packages provide clustering procedures that can be used in feature creation.

Text miners may also want to program the steps required for term extraction themselves.

Most programming languages, including those popular for statistical applications, such as S-PLUS, R (an open source analysis package), and SAS contain string manipulation function that can be used to parse words from text data. Initially, some investment in programming effort would be required to eliminate stopwords and perform stemming. The book *Text Mining* (Weiss *et al.*, 2005) contains pseudo code that can be referenced for programming many of the text mining procedures.

Two programming languages, Perl and Python have become popular for processing text data. Both languages are free and can be downloaded from the appropriate web site (www.perl.com and www.python.org). Because these languages are used so frequently for text processing, functions have already been developed and made available to users that handle many of the term extraction tasks.

In summary, text mining is a relatively new application, and software for performing text mining is relatively undeveloped compared to other data mining applications. When using one the data mining suites, the text miner may want to use text mining capabilities sold with the suite. These have not been tested as part of this study. The text miner may also wish to use free software or one of the programming languages that specialize in text processing.

5. REFERENCES

[1] Aggarawal, C, "On Learning Startegies for Topic Specific Web Crawling" in *Next Generation of Data Mining Applications*, Wiley, pp. 447 – 468, 2005

[2] Blench, M and Proulx, L, "Global Public Health Intelligence Network (GPHIN): An Early Warning Intelligence Information Management Model", Government Texhnologies Forum and Health Technologies Forum, August 30, 2005

[3] Casualty Actuarial Society Discussion Paper Program on Applying and Evaluating Generalized Linear Models, 2004

[4] Chen, S and Gopalakrishnan, "Speaker, Environment and Channel Change Detection and Clustering via the Bayesian Information Criterion", www.nist.gov/speech/publications/darpa98/html/bn20/bn20.htm, 1998

[5] Davi, A, Haughton, D, Nasr, N, Shah, G, Skaletsky , M and Spack, R., "A Review of Two Text Mining Packages: SAS Text Mining and WordStat", *American Statistician*, Feb 2005

[6] Derrig, R, Weisberg H, and Chen, X, 1994, "Behavioral Factors and Lotteries Under No-Fault with a Monetary Threshold: A Study of Massachusetts Automobile Claims", *Journal of Risk and Insurance*, June, 1994, 61:2: 245-275

[7] Derrig, R and Ostaszewski, "Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification", *Journal of Risk and Insurance*, pp. 447-482, 1995

[8] Derrig, R, "Fraud Fighting Actuaries: Mathematical Models for Insurance Fraud Detection", 2004 CAS Predictive Modeling Seminar

[9] De Vel, "Mining E-mail Authorship", KDD Workshop on Text Mining, 2000

[10] Howland, P, and Park, H, "Cluster Preserving Dimension Reduction Methods for Efficient Classification of Text Data", in *Survey of Text Mining*, Springer, 2004

[11] Ellingsworth, M, "An Overview of Fraud Modeling", Insurance Fraud Conference, November 2002

[12] Ellingsworth, M and Sullivan, D, "Text Mining Improves Business Intelligence and Predictive Modeling

in Insurance", DMReview, July, 2003
[13] Francis, L, "Neural Networks Demystified", CAS Discussion Paper Program, Winter 2001
[14] Hayward, G, "Mining Insurance Data to Promote Traffic Safety and Better Match Rates to Risk", *Casualty Actuarial Society Forum*, Winter 2002, *pp. 31 – 56.*
[15] Hastie, T, Tibshirani, R and Friedman, *J., The Elements of Statistical Learning*, Springer, 2001
[16] Heckman, P., "Aggregate Reserve Distribution from Detailed Process Model for Individual Claims", Casualty Loss Reserve Seminar, 2000
[17] Hoffman, P, *Perl for Dummies*, Wiley, 2003
[18] Holler, K, Somner, D, and Trahair, G, "Something Old, Something New in Classification Ratemaking With a New Use of GLMs for Credit Insurance", *Casualty Actuarial Society Forum, Winter 1999, pp. 31-84.*
[19] Hosmer, D, and Lemshow, S, *Applied Logistic Regression*, John Wiley and Sons, 1989
[20] Jacoby, William, *Data Theory and Dimension* Analysis, Sage Publications 1991.
[21] Ellingsworth, M and Sullivan, D, "Text Mining Improves Business Intelligence and Predictive Modeling in Insurance", DMReview, July, 2003
[22] Kauffman, L, Rousseeuw, *Finding Groups in Data, Wiley*, 1990
[23] Kolyshkina, I, "Text Mining Challenges and Treenet Impact of Textural Information on Claims Cost Prediction", presentation at Second International Salford Systems Data Mining Conference, March 2005
[24] Manning, C, Schutze, H, *Foundations of Statistical Natural Language Processing*, MIT Pres, 1999
[25] Metz, D, *Text Processing in Python*, Addison Wesley, 2003
[26] Miller, T, *Data and Text Mining*, Pearson Prentice Hall, 2003
[27] Mitchell, R, "Anticipation Game", Computer World, 2005
[28] Robb, D, "Text Mining Tools Take on Unstructured Data, *Computer World*, June 2004
[29] Senellart, P and Blondel, V, "Automatic Discovery of Similar Words" in *Survey of Text Mining*, Springer, 2004
[30] Sullivan, D, *Document Warehousing and Text Mining*, Wiley, 2001
[31] Taylor, G and McGuire, G, "Loss Reserving with GLMs: A Case Study", *2004 CAS Discussion Paper Program*
[32] Venebles, W.N. and Ripley, B.D., Modern *Applied Statistics with S-PLUS*, third edition, Springer, 1999
[33] Weisberg, H and Derrig, R, "Pricing No-Fault and Bodily Injury Coverages Using Micro-Data and Statistical Models", *Casualty Actuarial Society Forum, 1993*
[34] Weiss, Shalom, Indurkhya, Nitin, Zhang, Tong and Damerau, Fred, *Text Mining*, Springer, 2005
[35] Wikipedia, www.wikipedia.org, 2005

Biography of the Author

Louise Francis is a Consulting Principal at Francis Analytics and Actuarial Data Mining, Inc. She is involved data mining projects as well as conventional actuarial analyses. She has a BA degree from William Smith College and an MS in Health Sciences from SUNY at Stony Brook. She is a Fellow of the CAS and a Member of the American Academy of Actuaries. She is currently chair of the CAS Committee on the Theory of Risk, and is a frequent presenter at actuarial and industry symposia. She published three previous papers in the Data Management, Quality and Technology Call Paper Program: "Neural Networks Demystified" (2001) and "Martian Chronicles: Is MARS Better than Neural Networks" (2003) and "Dancing with Dirty Data: Methods for Exploring and Cleaning Data (2005).

Variable Reduction for Predictive Modeling with Clustering

Robert Sanche, and Kevin Lonergan, FCAS

Abstract

Motivation. Thousands of variables are contained in insurance data warehouses. In addition, external sources of information could be attached to the data contained in data warehouses. When actuaries build a predictive model, they are confronted with redundant variables which reduce the model efficiency (time to develop the model, interpretation of the results, and inflate variance of the estimates). For these reasons, there is a need for a method to reduce the number of variables to input in the predictive model.

Method. We have used *proc varclus* (SAS/STAT®) to find clusters of variables defined at a geographical level and attached to a database of automobile policies. The procedure finds cluster of variables which are correlated between themselves and not correlated with variables in other clusters. Using business knowledge and $1\text{-}R^2_{ratio}$, cluster representatives can be selected, thus reducing the number of variables. Then, the cluster representatives are input in the predictive model.

Conclusions. The procedure used in the paper for variable clustering quickly reduces a set of numeric variables to a manageable reduced set of variable clusters.

Availability. *proc varclus* from SAS/ STAT® has been used for this study. We found an implementation of variable clustering in R, function varclus, while we did not experiment with it.

Keywords. variable reduction, clustering, statistical method, data mining, predictive modeling.

1. INTRODUCTION

Over the last decade, insurance companies have gathered a vast amount of data in their data warehouses. Some of this information is well-known by the actuaries because it is used for other purposes, e.g. pricing of the policy. Also, there are many sources of external data (demographics, financial, meteorological...) available from vendors. The external sources are typically not as familiar to the actuary as the data from the data warehouses. This vast amount of information is available to create a predictive model. The objective of the predictive model could be to improve the pricing or reserving process, but also to analyze profitability, fraud, catastrophe, and any insurance operation. This amount of information from multiple sources provides numerous variables for the modeling project contemplated.

When a modeling project involves numerous variables, the actuary is confronted with the need to reduce the number of variables in order to create the model. The variables have sometimes an unknown behavior with the objective of the modeling project. In addition, when there is a multitude of variables, it becomes difficult to find out the relationship between variables.

Too many variables reduce the model efficiency. With many variables there is a potential

of overfitting the data. The parameter estimates of the model are destabilized when variables are highly correlated between each other. Also, it is much more difficult to have an explainable model when there are many variables. Finally, creating models with all possible combinations of variables is exhaustive, but this approach would take indefinite time when there are thousands of variables. An intermediate approach to the exhaustive search would also take a lot of time and some combinations of variables could be overseen.

Suppose you want to reduce the number of variables to a smaller set of variable clusters for efficiency, you can use variable clustering. Variable clustering provides groups of variables where variables in a group are similar to other variables in the same group and as dissimilar as possible to variables in another group.

1.1 Research Context

This paper addresses the initial stage of every predictive modeling project performed by an actuary, i.e. variable selection. Then, the variables selected would become inputs to predictive modeling techniques, such as, linear regression, generalized linear model, a neural network, to name a few.

A technical description of the variable clustering algorithm, *proc varclus*, is included in the SAS/STAT® User's Guide.[1] The method is not found in many textbooks on multivariate techniques, it mostly started as an implementation in statistical software.[2]

This paper is focused on variable clustering, but the example could be used, for example, in the context of complement to territorial relativities for automobile insurance. This complement would be obtained from a predictive model based on variables defined at some geographical level. The variables were selected using variable clustering on multiple sources of information, usually not used in pricing, attached to an automobile policy database. If the objective of the predictive model is to predict cost by territory, it makes sense to use fact (demographics, consumer expenditure, weather ...) variables selected from the variable clustering on the multiple sources, defined at some geographies (e.g. county), to complement territorial relativities.

The example provided in the paper is a simplification of a variable reduction problem. Many more variables would be clustered in a real life study.

Note that the variables used in the example have some intuitive relation to automobile

[1] SAS/STAT® 9.1 User's Guide p. 4797
[2] **Pasta paper 205**

insurance cost, although generally the variables presented to the variable clustering procedure are not previously filtered based on some educated guess. All the demographics, consumer expenditure, and weather variables are used in the clustering analysis. Filtering of variables is typically done after the variable clusters have been created. When there is a multitude of variables, it is more difficult to recognize irrelevant variables than to recognize redundant variables. A variable is considered irrelevant if it is not predictive for the specific predictive model. When the actuary deals with unknown data, a large number of the variables turn out to be irrelevant. A variable is redundant when it is highly correlated with another potential variable.

1.2 Objective

More and more actuaries use advanced statistical methods to create insurance models. This paper provides a tool; variable clustering, that can be added to the arsenal of the actuarial miners. Traditionally, PCA have been used for variable reduction by creating a set of components (weighted linear combinations of the original variables) which are difficult to interpret.

Typically, in the clustering literature, there is a rule for selecting the cluster representative, the $1\text{-}R^2_{ratio}$. Business knowledge from subject matter expert should also complement this rule to guide the selection of variables. For this reason, someone could decide to use more than one variable per cluster. Even though the clustering procedure provides diagnostic measures, there are reasons for using more than one variable per cluster. One of them is that the maximum number of clusters is a parameter provided by the user of the procedure. Also, for communication to users of the predictive model, an alternate variable may provide a better intuitive interpretation of the model than the cluster representative.

We should point out that the variable clustering works only with numeric variables. However, there are ways to convert categorical variables into numeric variables. For example, the hamming distance converts categorical variables into a numeric variable. Conversion of categorical variables is not covered in this paper.

We suggest options (*centroid* without *cov*) to the procedure of variable clustering which turn out to produce a scale-invariant method. Otherwise it would probably be necessary to rescale the ranges of the variables (with *proc standard*).

1.3 Outline

The remainder of the paper proceeds as follows. Section 2 will provide an overview of

clustering and more precisely the variable clustering. We will describe shortly the variable clustering algorithm used in this paper. Section 3 will provide an example of variable reduction in the context of automobile insurance. We will use variable clustering and will explain how variables can be selected to reduce their number. In section 4, we conclude the study. In Appendix A, we include an example of the SAS code and in Appendix B we include the procedure's output.

2. CLUSTERING

2.1 Clustering

"**Cluster Analysis** is a set of methods for constructing a sensible and informative classification of an initially unclassified set of data, using the variable values observed on each individual"[3]

In general, the goal of a cluster analysis is to divide a data set into groups of similar characteristics, such that observations in a group are as similar as possible and as dissimilar to observations in another group. Variable clustering, however, does not divide a set of data; instead it splits a set of variables with similar characteristics using a set of subject data.

Clustering is an unsupervised learning technique as it describes how the data is organized without using an outcome[4]. As a comparison, regression is a supervised learning technique as there is an outcome used to derive the model. Most data mining techniques are supervised learning techniques. Unsupervised techniques are only useful when there is redundancy in the data (variables).

At the basis of clustering is the notion of similarity. Without supervision, there is no response to say that occurrence *a* is similar to occurrence *b*. If there was a response associated with each occurrence; it could be used to compare *a* and *b* responses to induce similarity between both. Similarity: Two occurrences are similar if they have common properties.

For example, one occurrence is a car, another occurrence is a motorcycle and the last occurrence is a bicycle. First, lets say we have only number of wheels as a property. Then we would cluster the motorcycle and the bicycle since they have the same number of wheels. However, if we add the number of cylinders and fuel consumption, then the motorcycle is

[3] Everitt 1998
[4] Hastie p.2

more similar to the car. Similarity can be measured by distance measure (Euclidian distance, Manhattan or city block distance ...) or correlation type metrics.

There are two classes of clustering methods:

- Hierarchical: This class of clustering produces clusters that are hierarchically nested within clusters from previous iterations. This is the most commonly used clustering technique.
- Partitive: This class of clustering divides data in clusters by minimizing an error function of the distance between the observation vectors and the reference vectors (centroid - initial guess). This clustering technique requires elaborate selection of parameters and evaluation of the error function for all possible partition is impractical.

There are two approaches to hierarchical clustering:

- Agglomerative
    1. Start with each observation as its own cluster
    2. Compute the similarity between clusters
    3. Merge the clusters that are most similar
    4. Repeat from step 2 until one cluster is left
- Divisive
    1. Start will all observations assigned to one cluster
    2. Compute the similarity between clusters
    3. Split the cluster that are least similar
    4. Repeat from step 2 until each observation is a cluster

2.2 Variable Clustering

The procedure used in this paper for variable clustering is both a divisive algorithm and iterative algorithm. The procedure starts with a single cluster and recursively divides existing clusters into two sub-clusters until it reaches the stopping criteria, producing a hierarchy of disjoint clusters.

As mentioned previously, the procedure starts with all variables in one cluster. Based on the smallest percentage of variation explained by its cluster component a cluster is

chosen for splitting. The chosen cluster is split in two clusters by finding the first two principal components and assigning each variable to the component with which it has the higher correlation. The assignment follows a hierarchical structure with the approach presented in this paper. The clustering stops when the maximum number of clusters is attained or reached a certain percentage of variation explained.

3. VARIABLE CLUSTERING EXAMPLE

After the multiple sources of data (demographics, consumer expenditures, meteorological ...) are attached to the auto policy database, variable clustering can be performed to reduce the number of variables. The SAS code is included in Appendix A. The rule dictates to select the variable with the minimum $1\text{-}R^2_{\text{ratio}}$ as the cluster representative. The $1\text{-}R^2_{\text{ratio}}$ is defined below.

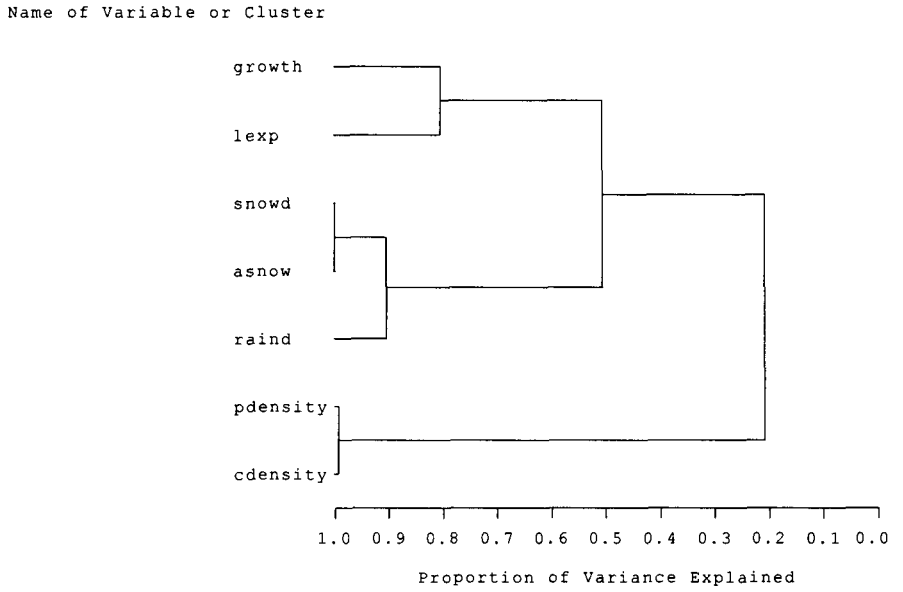
$$1\text{-}R^2_{\text{ratio}} = (1\text{-}R^2_{own})/(1\text{-}R^2_{nearest}) \tag{3.1}$$

Intuitively, we want the cluster representative to be as closely correlated to its own cluster ($R^2_{own} \to 1$) and as uncorrelated to the nearest cluster ($R^2_{nearest} \to 0$). Therefore, the optimal representative of a cluster is a variable where $1\text{-}R^2_{\text{ratio}}$ tends to zero.

Below, we include an extract of the output from *proc varclus* (see Appendix B for additional output from the procedure) with three clusters. Based on the $1\text{-}R^2_{\text{ratio}}$ we should select variables snowd, cdensity, and lexp as cluster representatives.

3 Clusters		R-squared with			
Cluster	**Variable**	**Own Cluster**	**Next Closest**	**1-R**2 Ratio**	
Cluster 1	**Raind**	0.5995	0.0426	0.4183	
	Snowd	0.8976	0.0317	0.1058	Choose
	Asnow	0.8940	0.0314	0.1095	
Cluster 2	**Pdensity**	0.9804	0.0228	0.0201	
	Cdensity	0.9804	0.0113	0.0199	Choose
Cluster 3	**Growth**	0.6459	0.0911	0.3896	
	Lexp	0.6459	0.0013	0.3546	Choose

After *proc varclus*, we have created a tree using *proc tree* which shows how the variable clusters are created. The variables are displayed vertically. The proportion of variance explained at each clustering level is displayed on the horizontal axis.



In that example, variables with similar factual attributes were clustered together; weather variables are in the same cluster and density variables are in the same cluster. Even with more variables, similar grouping patterns are observed.

If we consider three clusters; snowd, asnow and raind would all be in one cluster as they are on the same branch of the tree. The variable snowd would be the cluster representative since it has the lowest $1\text{-}R^2_{\text{ratio}}$.

The number of variables has been reduced and, now, we can efficiently create a predictive model to solve the problem at hand using linear regression, GLM[5], or neural network[6].

4. CONCLUSIONS

Given hundreds of variables, in order to create a predictive model the variable clustering

[5] Holler
[6] Francis

procedure runs quickly and produces satisfying results. We were able to reduce the number of variables using this procedure in order to efficiently create a predictive model. An efficient model was defined as followed:

- Interpretable
- Stable
- Timely

With this procedure, the modeling process is sped up significantly. The hierarchies produced by this procedure are easily interpretable with the tree output. Subject-matter experts usually do not have expertise to analyze statistical output in table form, but given the cluster hierarchy in tree output, can easily uncover alternate cluster representatives or eliminate irrelevant input. Other variable reduction techniques (e.g. PCA) do not create interpretable and disjoint clusters.

Appendix A: Code

```
* Example of variable clustering ;

%let varlist= pdensity cdensity growth /* demographics   */
              lexp                     /* expenditures   */
              raind snowd asnow        /* weather        */

proc varclus data='C:\example.sas7bdat' outtree=tree centroid maxc=6;
var &varlist ;
weight exp;
run;

axis1 label=(angle=0 rotate=0) minor=none;
axis2 minor=none order=(0 to 1 by 0.10);

proc tree data=tree horizontal vaxis=axis1 haxis=axis2;
height _propor_;
run;
```

Appendix B: Ouput

Cluster summary:

Cluster summary gives the number of variables in each cluster. The variation explained by the cluster is displayed. The proportion of variance explained is the variance explained divided by the total variance of the variables in the cluster.

Also displayed, is the summary are the R^2 of each variable with its own cluster, its closest cluster, and the $1\text{-}R^2_{\text{ratio}}$

Cluster Summary for 3 Clusters				
Cluster	**Members**	**Cluster Variation**	**Variation Explained**	**Proportion Explained**
1	3	3	2.371253	0.7904
2	2	2	1.960732	0.9804
3	2	2	1.291809	0.6459

Total variation explained = 5.623794 Proportion = 0.8034

3 Clusters		R-squared with			
Cluster	**Variable**	**Own Cluster**	**Next Closest**	**1-R**2 Ratio**	**Variable Label**
Cluster 1	**Raind**	0.5995	0.0426	0.4183	Rain 2
	Snowd	0.8976	0.0317	0.1058	Snow 2
	Asnow	0.8940	0.0314	0.1095	Snow 1
Cluster 2	**Pdensity**	0.9804	0.0228	0.0201	Pop density
	Cdensity	0.9804	0.0113	0.0199	Car density
Cluster 3	**Growth**	0.6459	0.0911	0.3896	Pop growth
	Lexp	0.6459	0.0013	0.3546	Leg expenditures

Standardized scoring coefficients:

The standardized scoring coefficients predict clusters from the variables. If a variable is not in a cluster, then the coefficient is zero. SAS does not provide unstandardized scoring coefficients.

Standardized Scoring Coefficients				
Cluster		1	2	3
Pdensity	Pop density	0.000000	0.504982	0.000000
Cdensity	Car density	0.000000	0.504982	0.000000
Growth	Pop growth	0.000000	0.000000	0.622137
Lexp	Leg expenditures	0.000000	0.000000	0.622137
Raind	Rain 2	0.374930	0.000000	0.000000
Snowd	Snow 2	0.374930	0.000000	0.000000
Asnow	Snow 1	0.374930	0.000000	0.000000

Cluster Structure:

The cluster structure gives the correlation between the variables and the clusters.

Cluster Structure				
Cluster		1	2	3
Pdensity	Pop density	-.069069	0.990134	-.151107
Cdensity	Car density	-.082041	0.990134	-.106496
Growth	Pop growth	-.301845	-.204659	0.803682
Lexp	Leg expenditures	-.036435	-.004435	0.803682
Raind	Rain 2	0.774267	-.102212	-.206297
Snowd	Snow 2	0.947393	-.044943	-.177956
Asnow	Snow 1	0.945502	-.056370	-.177070

Inter-Cluster Correlation:

This table provides the correlations between the clusters.

Inter-Cluster Correlations			
Cluster	1	2	3
1	1.00000	-0.07631	-0.21046
2	-0.07631	1.00000	-0.13008
3	-0.21046	-0.13008	1.00000

Cluster 3 will be split because it has the smallest proportion of variation explained, 0.645904, which is less than the PROPORTION=1 value.

Final summary:

Cluster summary and the other tables are listed for each number of clusters up to the maximum of clusters (option *maxc*). This table is listed at the end of the output and summarizes for each number of clusters the total variation and proportion explained by the clusters, the minimum proportion explained by a cluster, the minimum R^2 for a variable and the maximum $1\text{-}R^2_{\text{ratio}}$ for a ratio.

Number of Clusters	Total Variation Explained by Clusters	Proportion of Variation Explained by Clusters	Minimum Proportion Explained by a Cluster	Minimum R-squared for a Variable	Maximum 1-R**2 Ratio for a Variable
1	1.454308	0.2078	0.2078	0.0007	.
2	3.539308	0.5056	0.3157	0.0300	1.0124
3	5.623794	0.8034	0.6459	0.5995	0.4183
4	6.331985	0.9046	0.7904	0.5995	0.4349
5	6.952514	0.9932	0.9804	0.9804	0.0205
6	6.991782	0.9988	0.9959	0.9959	0.0058

5. REFERENCES

[1] B.S. Everitt , *The Cambridge Dictionary of Statistics*, 1998
[2] David J. Pasta, Diana Suhr, "Creating Scales from Questionnaires: PROC VARCLUS vs. Factor Analysis," *SUGI 29 Proceedings*, 2004, Paper 205-29
[3] Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning*, Springer, 2001
[4] SAS Institute Inc. 2004, SAS/STAT® 9.1 User's Guide, NC: SAS Institute
[5] Holler, "Something Old, Something New in Classification Ratemaking with a Novel Use of GLMs for Credit Insurance," *Casualty Actuarial Society Forum, Winter 1999.*
[6] Francis, "Neural Network Demystified," *Casualty Actuarial Society Forum, Winter 2001.*

Abbreviations and notations

PCA, principal component analysis
GLM, generalized linear model
proc, procedure in SAS

Biographies of the Authors

Robert Sanche is a Consultant with Tillinghast a business of Towers Perrin. He is responsible for predictive modeling projects. Prior to joining Tillinghast, he developed class plans for personal lines automobile using multivariate techniques with Travelers and The Hartford. He has also worked for GMAC Insurance and AXA in personal lines doing data mining and ratemaking respectively. He has degrees in Mathematics (Actuarial Science) and Computer Science (Operations Research) from Université de Montréal.

Kevin Lonergan graduated from Southern Connecticut State University in 1969 with BS, 1972 with MS. He taught mathematics in high school from 1969 to 1980. He has developed a new automobile product at turn of century. ACAS 1982. FCAS 1983.

The Report of the Research Working Party on Correlations and Dependencies Among All Risk Sources

Introduction

Glenn G. Meyers, Co-Chair, Research Working Party on Correlations and Dependencies Among All Risk Sources

Enterprise risk management requires the quantification of the total risk of an enterprise. As we undertake this task of quantification we first attempt to quantify the risk of individual parts of the enterprise. Examples of "individual parts" of an insurance enterprise could include the losses arising from its new business, its loss reserves or its asset portfolio. To properly combine these risks one needs to consider the "correlation" between the risks. We put the term "correlation" in quotes to draw attention to the fact that we are not restricting ourselves to the linear correlation that we all study in introductory statistics. This report considers a variety of ways for different risks to "move together."

There are three aspects of this problem that deserve some discussion.

Formulating models of correlated risks. An example of this includes inflation affecting the losses of different lines of insurance causing them to be correlated. A second example is where a correlation between inflation and interest rates that drives the correlation between the losses and assets of an insurance company.

Combining the models of correlated risks. This aspect refers to the mathematical techniques that are needed to combine to obtain the combined distribution of all the individual parts.

Parameterizing the models of correlated risks. If correlation matters, we should be able to find data somewhere that reflects this correlation and use it to parameterize a model that describes this data.

As the working party began its discussions, we quickly found out that these aspects of the correlation problem could not be treated in isolation. In the end, individual authors took the lead and produced four separate papers (or "chapters") that make up this report. Here is a summary of each paper.

1. "Correlation and Aggregate Loss Distributions with an Emphasis on the Iman-Conover Method" by Stephen J. Mildenhall. This paper gives a grand tour of a variety of multivariate models exhibiting correlation that should be of interest to the casualty actuary. The focus of the paper is the Iman-Conover method which can take arbitrary marginal (or individual risk) distributions and derive a multivariate distribution that has an arbitrary rank correlation matrix.

2. "Aggregating Bivariate Claim Severities with Numerical Fourier Inversion" by David L. Homer. This paper uses the very powerful technique of Fourier transforms to calculate the aggregate loss distributions with correlated claim severity distributions. Once we have settled on a standard set of models to describe the stochastic nature of the insurance business, the techniques described in this paper can significantly reduce the time needed to compute the distribution of an insurer's total losses.

3. "The Common Shock Model for Correlated Insurance Losses" by Glenn G. Meyers. This paper addresses the problem of estimating the correlations between lines of insurance. It takes the data from several insurers and produces stable estimates of parameters underlying the collective risk model for correlated insurance losses. And along the way it provides on how the parameters of the collective risk model change as the size of the risk changes.

4. "Serial Correlation of Interest and Inflation Rates" by Hans E. Waszink. This paper discusses an approach to modeling the present value of reserves under the impact of uncertain interest and inflation rates. The dependence between interest rates and inflation rates is modeled using copulas. The paper also shows how to test the goodness of fit of data to members of the class of Archimedean copulas.

In addition to the chapter authors listed above, there were several others who contributed to the work of the Correlation Working Party by either presenting ideas or by reviewing papers. These individuals are listed below.

Shawna Ackerman
Kevin Dickson
Lijia Guo
Leigh Halliwell
Roger Hayne
Philip Heckman
Daniel Heyer
Youngju Lee
Christopher Monsour
James Rech
Kevin Shang
Chuck Thayer
Emil Valdez

The Report of the Research Working Party on Correlations and Dependencies Among All Risk Sources

Part 1

Correlation and Aggregate Loss Distributions With An Emphasis On The Iman-Conover Method

Stephen J. Mildenhall, Member, CAS Research Working Party on Correlations and Dependencies Among All Risk Sources

November 27, 2005

Abstract

Motivation. The CAS Research Working Party on Correlation and Dependencies Among All Risk Sources has been charged to "lay the theoretical and experimental foundation for quantifying variability when data is limited, estimating the nature and magnitude of dependence relationships, and generating aggregate distributions that integrate these disparate risk sources."

Method. The Iman-Conover method represents a straight forward yet powerful approach to working with dependent random variables. We explain the theory behind the method and give a detailed step-by-step algorithm to implement it. We discuss various extensions to the method, and give detailed examples showing how it can be used to solve real world actuarial problems. We also summarize pertinent facts from the theory of univariate and multivariate aggregate loss distributions, with a focus on the use of moment generating functions. Finally we explain how Vitale's Theorem provides a sound theoretical foundation to the Iman-Conover method.

Availability. The software used to generate the paper's examples is available at http://www.mynl.com/wp.

Keywords. Dependency, correlation, aggregate loss distributions, fast Fourier transform.

Chapter 1

INTRODUCTION

The Iman Conover (IC) Method is a practical, down-to-earth approach to dealing with dependent random variables. It should be part of every actuary's toolkit.

When two variables X and Y are positively correlated there is a tendency for large values of X to be associated with large values of Y. Knowing how the large values are associated would make it possible to work in reverse: by ordering samples from X and Y so that large-large matches and small-small matches are more likely would result in a bivariate sample with positive correlation. The Iman-Conover (IC) method gives a way of determining reasonable associations, and hence inducing correlation between samples of variables. It is ideally suited to simulation work where marginal distributions are sampled independently but must be combined to achieve some desired level of correlation. The IC method is used by the popular @Risk software package to induce correlation.

Before describing the IC method, we begin with a review of measures of correlation and association in Chapter 2. Then, in Chapter 3 we describe several useful techniques for working with univariate and multivariate aggregate loss distributions. These include formulae to compute moments of aggregates using moment generating functions, a discussion of mixed Poisson counting distributions, approximating univariate aggregates using the shifted gamma and shifted lognormal distributions, Fast Fourier transform methods, and computing correlation coefficients related to multivariate aggregates in a variety of situations.

Next we turn to a description of the IC method itself, which can simplistically be described as follows. Given two samples of n values from known marginal distributions X and Y and a desired correlation between them, first determine a sample from some reference bivariate distribution that has exactly the desired linear correlation. Then re-order the samples from X and Y to have the same rank order as the reference distribution. The output will be a sample from a bivariate distribution with the correct marginal distributions and with rank correlation coefficient equal to that of a bivariate distribution which, in turn, has exactly the desired correlation coefficient. Since linear correlation and rank correlation are typically close, the output has approximately the desired correlation structure. What makes the IC method work so effectively is the existence of easy algorithms to determine samples from reference distributions with prescribed correlation structures. Obviously the method can then be extended to work with samples from multivariate distributions in any dimension.

In their original paper, Iman and Conover [21] point out that their method has several desirable properties.

1. It is very simple to understand and implement.

2. It is distribution free; it may be used with equal facility on all types of input distributions.

3. It can be applied to any sampling scheme for which correlated input variables could logically be considered. That is, the output sample contains the same values as the input, only with a different pairing. Hence in Latin hyper cube sampling, the integrity of the intervals is maintained.

4. The marginal distributions remain intact.

The actual IC method involves some enhancements over such a naive implementation, and we give full details in Chapter 4. We give a step-by-step example to explain how the method works in practice in Section 4.3. In Section 4.4 we show how the basic IC method can be extended, and illustrate the impact these extensions have on the types of multivariate distributions the method produces. Section 4.5 compares the IC method with the normal copula method described in Wang [37].

Chapter 5 gives a detailed practical example which computes the bivariate distribution of ceded and retained losses in a book with an excess of loss reinsurance structure. Such a bivariate distribution is necessary to compute the net underwriting result if the reinsurance contains any variable feature like a sliding scale commission, profit commission or annual aggregate deductible.

Chapter 6 discusses the theoretical underpinnings of the IC method in a more technical manner. It can be ignored with impunity by readers more interested in practice than theory.

Appendix A discusses practical computational issues and describes some freely available software which can be used to implement the IC method in Excel.

Some sections are marked with an asterisk. These I regard as interesting, but not "core". The remaining, un-starred sections, contain core facts which I think every actuary working with correlation and aggregate losses should know.

When modeling correlation the reader should remember that the model must follow an understanding of reality. Model building, especially modeling correlation, must start with an understanding of the underlying processes. Graphical representations are often useful to aid understanding and help communicate results. It may be necessary to build pre-models to understand the underlying processes and use these to parameterize quicker, more computationally efficient, implementations. The IC method is ideal here: having understood the drivers of correlation and estimated an appropriate correlation matrix and suitable multivariate structure, the IC method can be used to produce correlated samples with blistering speed. However, the reader should not succumb to the temptation to estimate a

200×200 correlation matrix using data and expect to get realistic result from it. It will be more noise than signal.

In conclusion, we recommend the IC method as being fast, simple to understand, and efficient. We also recommend the use of a shifted gamma or shifted lognormal distribution to model univariate aggregate losses.

Throughout the paper bold face roman variables represent vectors (lower case) and matrices (upper case). Sections in a smaller typeface are optional, more technical discussions. Sections marked with an asterisk* contain non-core material.

Acknowledgement. I would like to thank Christopher Monsour, Chuck Thayer, Leigh Halliwell, Roger Hayne, Phil Heckman, and Kevin Shang for their helpful comments and suggestions.

Chapter 2

CORRELATION AND ASSOCIATION

Before discussing specific measures of correlation and association, recall that two random variables X and Y are independent if

$$\Pr(X \in A, Y \in B) = \Pr(X \in A)\Pr(Y \in B) \tag{2.1}$$

for all suitable sets A and B. It is possible to prove that X and Y are independent if and only if

$$\mathrm{E}[f(X)g(Y)] = \mathrm{E}[f(X)]\mathrm{E}[g(Y)] \tag{2.2}$$

for all functions f and g.

See Wang [37] and Press et al. [31] for more information on the definitions and terms described here.

2.1 Correlation, Rank Correlation and Kendall's Tau

There are three common measures of association (more commonly called simply correlation) between two random variables X and Y: linear or Pearson correlation, rank correlation and Kendall's Tau.

The linear or Pearson correlation coefficient, usually just called correlation, is defined as

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)} = \frac{\mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))]}{\sigma(X)\sigma(Y)} \tag{2.3}$$

where $\sigma(X)$ is the standard deviation of X. By the Cauchy-Schwarz inequality the correlation coefficient always lies in the range $[-1,1]$. The correlation coefficient is sometimes called the Pearson correlation coefficient or linear correlation coefficient. Perfect correlation, when $\rho = \pm 1$, occurs if and only if $Y = aX + b$ for constants $a > 0$ (resp. $a < 0$) and b. The correlation coefficient is a natural measure of association when X and Y come from a bivariate normal distribution because it is enough to completely specify the dependence between the marginals. Needless to say, such pleasant results do not hold in general! For a multidimensional distribution the correlation matrix has i,jth element equal to the correlation coefficient of the i and jth marginals.

A related measure is the covariance coefficient defined as

$$\omega(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{E}(X)\mathrm{E}(Y)}. \tag{2.4}$$

By (2.2) independent variables are uncorrelated. However, the converse is not true. The classic counter-examples of uncorrelated but dependent variables include

- X a standard normal and $Y = X^2$,
- (X,Y) uniformly distributed over a circle of radius one centered at the origin, and
- (X,Y) distributed with a bivariate t-distribution with zero correlation.
- Let $X, X_1, X_2, \ldots$ be identically distributed random variables with mean zero and let N be a counting distribution. Then $A = X_1 + \cdots X_N$ and N are uncorrelated but not independent. If X and X_i have a non-zero mean then $\mathrm{Cov}(A,N) = \mathrm{E}(X)\mathrm{Var}(N)$.

The correlation coefficient of a bivariate sample (X_i, Y_i), $i = 1, \ldots, n$, is defined as

$$\rho = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2}} \tag{2.5}$$

where $\bar{X} = n^{-1} \sum_i X_i$ and similarly for $\bar{Y}$.

Let X be an $n \times r$ matrix representing an n-fold sample of $1 \times r$ vectors. Suppose that the means of the columns of X are all zero (subtract the means if necessary). Then the variance-covariance matrix of X is simply $n^{-1}X'X$, where X' denotes matrix transpose.

The second measure of association we consider is rank correlation. Given a sample $X_1, \ldots, X_n$ of observations from a random variable X the rank order statistics $X_{(1)}, \ldots, X_{(n)}$ are a permutation of the original observations ordered so that $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Call j the rank of $X_{(j)}$. The rank correlation, or Spearman rank correlation, of a sample is defined as the correlation of the ranks of the sample. Rank correlation lies in the range $[-1, 1]$ because it is a correlation. It is invariant under strictly monotonic transforms of X and Y, so for example the rank correlation of a sample (X, Y) is the same as the transformed samples $(\log(X), \log(Y))$ or $(\exp(X), \exp(Y))$. Rank correlation is a nonparametric measure of association because it is invariant under transformation. For continuous random variables rank correlation can also be computed as

$$12\mathrm{E}[(F_X(X) - 0.5)(F_Y(Y) - 0.5)] \tag{2.6}$$

where F_X (resp. F_Y) is the distribution function of X (resp. Y).

The third common measure of association is called Kendall's tau. Kendall's tau looks at concordances and discordances between pairs of data points (x_i, y_i) and (x_j, y_j). A pair of observation-pairs is concordant if $(x_i - x_j, y_i - y_j)$ lies in the upper right hand or lower left hand quadrants of the plane, and discordant otherwise. Take all $n(n-1)/2$ distinct pairs of data from the sample and count the number of concordances c and discordances d, except that if the ranks of the x's are the same the pair is called an extra y pair and if the ranks of the y's are the same the pair is an extra x pair. If the ranks of both x and y are the same the pair

does not count at all. Let e_x and e_y be the number of extra x and y pairs. Kendall's tau is defined in Press et al. [31] as

$$\tau = \frac{c-d}{\sqrt{c+d+e_y}\sqrt{c+d+e_x}}. \tag{2.7}$$

Kendall's tau can also be computed as

$$\tau(X,Y) = 4\int_0^1 \int_0^1 F_{X,Y}(x,y)d^2F_{X,Y}(x,y) - 1 \tag{2.8}$$

provided singularities are handled appropriately, see Wang [37]. The Kendall's tau of a sample (X_i, Y_i), $i = 1, \ldots, n$ can be computed as

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} \text{sign}((X_i - X_j)(Y_i - Y_j)) \tag{2.9}$$

where $\text{sign}(z)$ is 1, 0 or -1 when z is positive, zero or negative.

The statistics of Kendall's tau are covered in more detail by Conover, [6]. Conover points out that if the null hypothesis that (X, Y) are independent is true, the distribution of tau approaches the normal quite rapidly. Hence the normal approximation for tau is better than the one for Spearman's rho under the null hypothesis. He also points out that tau has a natural interpretation in terms of the probability that an observation is concordant versus discordant.

Equation (2.9) is precisely consistent with the definition in Equation (2.7) only when there are no ties. In the no-ties case, (2.9) is the form that Kendall proposed in his 1938 paper. When there are ties, (2.9) ignores ties in either X or Y, but it counts every pair of observations in the total used in the denominator.

Equation (2.7) accounts explicitly for ties without distorting the answer unduly, and it always provides an answer regardless of the number of ties in the sample. Conover's method fails when every pair results in a tie in the rank of the Xs. On the other hand, if the ranks of all the Xs are tied, X should not really be termed a "variable", much less a "random variable"!

Conover's alternative to (2.9) is to use a different method to account for ties. If the Y ranks are tied, he adds $1/2$ to both c and d. If the X ranks are tied, the

comparison is dropped entirely, adding nothing to c or to d. Otherwise, c gets incremented for positive signs and d counts the negatives in the sum. Conover's final statistic is

$$\tau = \frac{c-d}{c+d}. \tag{2.10}$$

Conover's statistic adopts a form of Kendall's tau that was introduced by Goodman and Kruskal [14]. Equation (2.10), which is also called the gamma coefficient, can attain the values $+1$ and -1 even in the presence of ties in the sample data.

There are several relationships between these measures of correlation, particularly if the sample comes from a multivariate normal distribution. For example if (X, Y) are bivariate normal with correlation ρ then

$$\tau(\Phi(X), \Phi(Y)) = \frac{2}{\pi} \arcsin(\rho) \tag{2.11}$$

and the rank correlation

$$\text{rankCorr}(\Phi(X), \Phi(Y)) = \frac{6}{\pi} \arcsin(\rho/2). \tag{2.12}$$

Similar results hold for samples from any elliptically contoured distribution, see Fang and Zhang [11], Embrechts, Lindskog and McNeil [9] and Embrechts, McNeil and Straumann [10].

2.2 Comonotonicity*

Two random variables X and Y are comonotonic if there exists a third variable Z and non-decreasing functions f and g such that $X = f(Z)$ and $Y = g(Z)$. For example, if X and Y are two different excess layers on the same risk then they are comonotonic. A stock and an option on it have comonotonic payouts. Comontonicity represents a high level of association between two values, but it need not result in a high level of linear correlation.

Some authors propose that risk measures r should be sub-additive, $r(X+Y) \leq r(X) + r(Y)$, with the tag-line "mergers cannot cause risk". Coherent measures require sub-additivity, see Artzner et al. [2]. Others authors propose additivity for comonotonic risks $r(X + Y) = r(X) + r(Y)$ if X and Y are comonotonic, see Wang [36].

2.3 Measures for Non-Normal Variables*

Linear correlation is the perfect measure of association for normally distributed random variables. It does not deal so effectively with non-normal variables. However, any continuous random variable X with distribution function F can be transformed into a normal variable Y via

$$Y = \Phi^{-1}(F(X)). \tag{2.13}$$

It therefore makes sense to transform non-normal variables using (2.13) and then to compute correlations between the transformed variables. If X is already a normal variable then (2.13) simply normalizes X to mean 0 and standard deviation 1.

Normalizing transformations are related to the IC method and the normal copula method as we will explain with Theorem 2 below. The normalizing transformation has been used in the literature by Wang [38] and Sornette et al. [32] amongst others.

Chapter 3

GENERAL PROPERTIES OF AGGREGATE LOSS DISTRIBUTIONS

Here we gather together some useful techniques for working with aggregate distributions, modeling correlation, parameter uncertainty, and so forth. Many of the techniques we introduce here will be used in the extended example, given in the Chapter 5. We introduce the negative multinomial distribution and we provide an introduction to Fast Fourier Transform (FFT) methods in both one and two dimensions. We begin with a discussion of moment generating functions and mixed Poisson frequency distributions.

We will use the following notation. The variance of a random variable X is $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$. The standard deviation is $\sigma(X) = \sqrt{\mathrm{Var}(X)}$. The coefficient of variation (CV) of X is $\mathrm{CV}(X) = \sigma(X)/\mathrm{E}(X)$. The skewness of X is $\mathrm{E}[(X - \mathrm{E}(X))^3]/\sigma(X)^3$.

3.1 Moment Generating Functions

The moment generating function of a random variable X is defined as

$$M_X(\zeta) = \mathrm{E}(\exp(\zeta X)). \tag{3.1}$$

The moment generating function is related to the characteristic function of X which is defined as $\phi_X(\zeta) = \mathrm{E}(\exp(i\zeta X)) = M_X(i\zeta)$. ϕ is guaranteed to converge for all real ζ and so is preferred in certain situations.

Moment generating functions get their name from the fundamental property that

$$\left.\frac{\partial^n M_X}{\partial \zeta^n}\right|_{\zeta=0} = \mathrm{E}(X^n) \tag{3.2}$$

for all positive integers n provided the differential exists.

Let F be the distribution function of X. Feller [12, Section XVII.2a] shows that if F has expectation μ then ϕ, the characteristic function of F, has a derivative ϕ' and $\phi'(0) = i\mu$. However the converse is false. Exactly what does hold is spelt out in the next theorem.

Theorem 1 *(Pitman) The following are equivalent.*

1. $\phi'(0) = i\mu$.
2. *As* $t \to \infty$, $t(1 - F(t) + F(-t)) \to 0$ *and*

$$\int_t^{-t} x dF(x) \to \mu. \tag{3.3}$$

   $F(-t) := \lim F(s)$ *as* $s \uparrow t$.

3. *The average* $(X_1 + \cdots + X_n)/n$ *tends in probability to* μ, *that is* $\Pr(|(X_1 + \cdots + X_n)/n - \mu| > \epsilon) \to 0$ *as* $n \to \infty$.

Note that the condition for the limit in (3.3) to exist is weaker than the requirement that $\mathrm{E}(X)$ exists if X is supported on the whole real line. For the expectation to exist requires $\int_{-\infty}^{\infty} x dF(x)$ exists which means $\lim_{t\to-\infty} \lim_{s\to\infty} \int_t^s x dF(x)$.

The moment generating function of a bivariate distribution (X_1, X_2) is defined as

$$M_{X_1,X_2}(\zeta_1, \zeta_2) = \mathrm{E}(\exp(\zeta_1 X_1 + \zeta_2 X_2)). \tag{3.4}$$

It has the property that

$$\left.\frac{\partial^{m+n} M_{X_1,X_2}}{\partial \zeta_1^m \partial \zeta_2^n}\right|_{(0,0)} = \mathrm{E}(X_1^m X_2^n) \tag{3.5}$$

for all positive integers n, m.

The MGF of a normal variable with mean μ and standard deviation σ is $M(\zeta) = \exp(\mu\zeta + \sigma^2\zeta^2/2)$. The MGF of a Poisson variable with mean n is

$$M(\zeta) = \exp(n(e^\zeta - 1)), \tag{3.6}$$

a fact we will use repeatedly below.

See Feller [12] and Billingsley [3] for more information on moment generating functions, characteristic functions and modes of convergence.

3.2 Mixed Poisson Frequency Distributions

Here we consider some basic facts about mixed Poisson frequency distributions. Let G be a non-negative mixing distribution with $\mathrm{E}(G) = 1$ and $\mathrm{Var}(G) = c$. The variable c is called the contagion. Let N be a claim count random variable where the conditional distribution of $N|G = g$ is Poisson with mean gn for some non-negative real n. We will call N a G-mixed Poisson random variable.

By (3.6), the MGF of a G-mixed Poisson is

$$M_N(\zeta) = \mathrm{E}(e^{\zeta N}) = \mathrm{E}(\mathrm{E}(e^{\zeta N}|G)) = \mathrm{E}(e^{nG(e^{\zeta}-1)}) = M_G(n(e^{\zeta}-1)) \tag{3.7}$$

since $M_G(\zeta) := \mathrm{E}(e^{\zeta G})$. Thus

$$\mathrm{E}(N) = M_N'(0) = nM_G'(0) = n, \tag{3.8}$$

because $\mathrm{E}(G) = M_G'(0) = 1$, and

$$\mathrm{E}(N^2) = M_N''(0) = n^2 M_G''(0) + nM_G'(0) = n^2(1+c) + n. \tag{3.9}$$

Hence

$$\mathrm{Var}(N) = n(1+cn). \tag{3.10}$$

Finally

$$\mathrm{E}(N^3) = M_N^{(3)}(0) = n^3 M_G^{(3)}(0) + 3n^2 M_G''(0) + nM_G'(0) \tag{3.11}$$

from which it is easy to compute the skewness.

We can also assume G has mean n and work directly with G rather than nG, $\mathrm{E}(G) = 1$. We will call both forms mixing distributions.

Table 3.1: Parameterizations of the Gamma Distribution

Model	Density	MGF	Mean	Var
(a) α, β	$\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}$	$(1-\beta t)^{-\alpha}$	$\alpha\beta$	$\alpha\beta^2$
(b) α, β	$\frac{x^{\alpha-1}\beta^{\alpha}e^{-x\beta}}{\Gamma(\alpha)}$	$(1-t/\beta)^{-\alpha}$	α/β	α/β^2
(c) α, θ	$\frac{x^{\alpha-1}e^{-x/\theta}}{\theta^{\alpha}\Gamma(\alpha)}$	$(1-t\theta)^{-\alpha}$	$\alpha\theta$	$\alpha\theta^2$

3.3 Gamma and Negative Binomial Variables

Recall that a negative binomial is a gamma-mixed Poisson: if $N|G$ is distributed as a Poisson with mean G, and G has a gamma distribution, then the unconditional distribution of N is a negative binomial. Both the gamma and negative binomial occur in the literature with many different parameterizations. The main ones are shown in the Tables 3.1 and 3.2.

In Table 3.1 model (a) is used by Microsoft Excel, Wang [37] and Johnson et al. [22, Chapter 17]. Model (b) is used by Bowers et al. [4]. Model (c) is used by Klugman, Panjer and Willmot in the Loss Models text [25]. Obviously model (c) is just model (a) with a change of notation.

In Table 3.2 model (a) is used by Wang and Loss Models, (b) by Johnson et al. [24, Chapter 5] and (c) by Bowers et al. [4] and Excel. In model (c) the parameter r need not be an integer because the binomial coefficient can be computed as

$$\binom{r+x-1}{x} = \frac{\Gamma(r+x)}{\Gamma(r)x!},$$

an expression which is valid for all r. The cumulative distribution function of the negative binomial can be computed using the cumulative distribution of the beta

distribution. Using the model (c) parameterization, if N is negative binomial p, r then

$$\Pr(N \leq k) = \text{BETADIST}(p; r, k+1) := \frac{1}{B(r, k+1)} \int_0^p u^{r-1}(1-u)^k du$$

where B is the complete beta function. See Johnson, Kotz and Kemp [24, Eqn. 5.31] for a derivation. BETADIST is the Excel beta cumulative distribution function.

The name negative binomial comes from an analogy with the binomial. A binomial variable has parameters n and p, mean np and variance npq, where $p + q = 1$. It is a sum of n independent Bernoulli variables B where $\Pr(B = 1) = p$ and $\Pr(B = 0) = q = 1 - p$. The MGF for a binomial is $(q + pe^{\zeta})^n$ and the probabilities are derived from the binomial expansion of the MGF. By analogy the negative binomial can be defined in terms of the negative binomial expansion of $(Q - Pe^{\zeta})^{-k}$ where $Q = 1 + P$, $P > 0$ and $k > 0$.

For the actuary there are two distinct ways of looking at the negative binomial which give very different results and it is important to understand these two views. First there is the contagion view, where the mixing distribution G has mean n and variance c producing a negative binomial with mean n and variance $n(1 + cn)$. (In fact G is a gamma with model (a) parameters $\alpha = r$ and $\beta = 1/r$.) The word contagion was used by Heckman and Meyers [17] and is supposed to indicate a "contagion" of claim propensity driven by common shock uncertainty, such as claim inflation, economic activity, or weather. Here the variance grows with the square of n and the coefficient of variation tends to $\sqrt{c} > 0$ as $n \to \infty$. Secondly, one can consider an over-dispersed family of Poisson variables with mean n and variance vn for some $v > 1$. We call v the variance multiplier. Now the coefficient of variation tends to 0 as $n \to \infty$. The notion of over-dispersion and its application in modeling is discussed in Clark and Thayer [5] and Verrall [34].

Table 3.2: Parameterizations of the Negative Binomial Distribution

Model	Density	MGF	Mean	Var
(a) α, β	$\binom{\alpha+x-1}{x}\left(\frac{\beta}{1+\beta}\right)^x\left(\frac{1}{1+\beta}\right)^\alpha$	$(1-\beta(e^t-1))^{-\alpha}$	$\alpha\beta$	$\alpha\beta^2$
(b) P, k	$\binom{k+x-1}{x}\left(\frac{P}{Q}\right)^x\left(\frac{Q-P}{Q}\right)^k$	$(Q-Pe^t)^{-k}$	kP	kPQ
(c) $p, r>0$	$\binom{r+x-1}{x}p^r q^x$	$\frac{p^r}{(1-qe^s)^r}$	rq/p	rq/p^2

$Q = P+1$, $q = 1-p$, $0 < p < 1$ and $r > 0$, and $P = 1/(\beta+1)$.

Table 3.3: Fitting the Negative Binomial Distribution

Model	Parameters	Variance Multiplier		Contagion	
		Scale	Shape	Scale	Shape
(a)	r, β	$r = m/(v-1)$	$\beta = v-1$	$r = 1/c$	$\beta = cn$
(b)	k, P	$k = m/(v-1)$	$P = v-1$	$k = 1/c$	$P = cn$
(c)	r, p	$r = m/(v-1)$	$p = 1/v$	$r = 1/c$	$p = 1/(1+cn)$

3.4 Aggregate Distributions

Let $A = X_1 + \cdots + X_N$ be an aggregate distribution, where N is a G-mixed Poisson and X_i are iid severity random variables. Then, proceeding by analogy with 3.7, we have

$$\begin{aligned} M_A(\zeta) &= \mathrm{E}(\exp(\zeta(X_1 + \cdots X_N))) \\ &= \mathrm{EE}(\exp(\zeta(X_1 + \cdots X_N))|N) \\ &= \mathrm{E}(\mathrm{E}(\exp(\zeta X_1)^N)) \\ &= \mathrm{E}(\mathrm{E}(M_X(\zeta)^N|G)) \\ &= \mathrm{E}(\exp(nG(M_X(\zeta) - 1))) \\ &= M_G(n(M_X(\zeta) - 1)) \end{aligned} \tag{3.12}$$

Thus

$$\mathrm{E}(A) = M_A'(0) = nM_G'(0)M_X'(0) = n\mathrm{E}(X) \tag{3.13}$$

and

$$\begin{aligned} \mathrm{E}(A^2) &= M_A''(0) \\ &= n^2 M_G''(0)M_X'(0)^2 + nM_G'(0)M_X''(0) \\ &= n^2\mathrm{E}(G^2)\mathrm{E}(X)^2 + n\mathrm{E}(X^2). \end{aligned} \tag{3.14}$$

Hence, using the fact that $\mathrm{E}(G^2) = 1 + c$, we get

$$\begin{aligned} \mathrm{Var}(A) &= n^2\mathrm{E}(G^2)\mathrm{E}(X)^2 + n\mathrm{E}(X^2) - n^2\mathrm{E}(X)^2 \\ &= n^2 c\mathrm{E}(X)^2 + n\mathrm{E}(X^2) \\ &= (\mathrm{Var}(N) - \mathrm{E}(N))\mathrm{E}(X)^2 + \mathrm{E}(N)\mathrm{E}(X^2) \\ &= \mathrm{Var}(N)\mathrm{E}(X)^2 + \mathrm{E}(N)\mathrm{Var}(X). \end{aligned} \tag{3.15}$$

Continuing along the same vein we get

$$\begin{aligned} \mathrm{E}(A^3) =& \mathrm{E}(N)\mathrm{E}(X^3) + \mathrm{E}(N^3)\mathrm{E}(X)^3 + 3\mathrm{E}(N^2)\mathrm{E}(X)\mathrm{E}(X^2) \\ &- 3\mathrm{E}(N)\mathrm{E}(X)\mathrm{E}(X^2) - 3\mathrm{E}(N^2)\mathrm{E}(X)^3 + 2\mathrm{E}(N)\mathrm{E}(X)^3. \end{aligned} \tag{3.16}$$

and so we can compute the skewness of A—remember that $\mathrm{E}[(A - \mathrm{E}(A))^3] = \mathrm{E}(A^3) - 3\mathrm{E}(A^2)\mathrm{E}(A) + 2\mathrm{E}(A)^3$. Further moments can be computed using derivatives of the moment generating function.

Having computed the mean, CV and skewness of the aggregate using these equations we can use the method of moments to fit a shifted lognormal or shifted gamma distribution. We turn next to a description of these handy distributions.

3.5 Shifted Gamma and Lognormal Distributions

The shifted gamma and shifted lognormal distributions are versatile three parameter distributions whose method of moments parameters can be conveniently computed by closed formula. The examples below show that they also provide a very good approximation to aggregate loss distributions. The shifted gamma approximation to an aggregate is discussed in Bowers et al. [4]. Properties of the shifted gamma and lognormal distributions, including the method of moments fit parameters, are also shown in Daykin et al. [7, Chapter 3].

Let L have a lognormal distribution. Then $S = s \pm L$ is a shifted lognormal, where s is a real number. Since s can be positive or negative and since L can equal $s + L$ or $s - L$, the shifted lognormal can model distributions which are positively or negatively skewed, as well as distributions supported on the negative reals. The key facts about the shifted lognormal are shown in Table 3.4. The variable η is a solution to the cubic equation

$$\eta^3 + 3\eta - \gamma = 0$$

where γ is the skewness.

Let G have a gamma distribution. Then $T = s \pm G$ is a shifted gamma distribution, where s is a real number. Table 3.1 shows some common parametric forms for the gamma distribution. The key facts about the shifted gamma distribution are also shown in Table 3.4.

The exponential is a special case of the gamma where $\alpha = 1$. The χ^2 is a special case where $\alpha = k/2$ and $\beta = 2$ in the Excel parameterization. The Pareto is a mixture of exponentials where the mixing distribution is gamma.

Table 3.4: Shifted Gamma and Lognormal Distributions

Item	Shifted Gamma	Shifted Lognormal
Parameters	s, α, θ	s, μ, σ
Mean m	$s + \alpha\theta$	$s + \exp(\mu + \sigma^2/2)$
Variance	$\alpha\theta^2$	$m^2 \exp(\sigma^2 - 1)$
CV, ν	$\sqrt{\alpha}\beta/\gamma m$	$\exp((\sigma^2 - 1)/2)$
Skewness, γ	$2/\sqrt{\alpha}$	$\gamma = \nu(\nu^2 + 3)$
Method of Moments Parameters		
η	n/a	$\eta = u - 1/u$ where $u^3 = \sqrt{\gamma^2 + 4}/2 + \gamma/2$
Shift variable, s	$m - \alpha\beta$	$m(1 - \nu\eta)$
α or σ	$4/\gamma^2$	$\sqrt{\ln(1 + \eta^2)}$
β or μ	$m\nu\gamma/2$	$\ln(m - s) - \sigma^2/2$

3.6 Excess Frequency Distributions

Given a ground-up claim count distribution N, what is the distribution of the number of claims exceeding a certain threshold? We assume that severities are independent and identically distributed and that the probability of exceeding the threshold is q. Define an indicator variable I which takes value 0 if the claim is below the threshold and the value 1 if it exceeds the threshold. Thus $\Pr(I = 0) = p = 1 - q$ and $\Pr(I = 1) = q$. Let M_N be the moment generating function of N and N' is the number of claims in excess of the threshold. By definition we can express N' as an aggregate

$$N' = I_1 + \cdots + I_N. \tag{3.17}$$

Thus the moment generating function of N' is

$$\begin{aligned} M_{N'}(\zeta) &= M_N(\log(M_I(\zeta))) \\ &= M_N(\log(p + qe^{\zeta})) \end{aligned} \tag{3.18}$$

Using indicator variables I is called p-thinning by Grandell [15].

Here are some examples.

Let N be Poisson with mean n. Then

$$M_{N'}(\zeta) = \exp(n(p + qe^{\zeta} - 1)) = \exp(qn(e^{\zeta} - 1)) \tag{3.19}$$

so N' is also Poisson with mean qn—the simplest possible result.

Next let N be a G-mixed Poisson. Thus

$$\begin{aligned} M_{N'}(\zeta) &= M_N(\log(p + qe^{\zeta})) \\ &= M_G(n(p + qe^{\zeta} - 1)) \\ &= M_G(nq(e^{\zeta} - 1)). \end{aligned} \tag{3.20}$$

Hence N' is also a G-mixed Poisson with lower underlying claim count nq in place of n.

In particular, if N has a negative binomial with parameters P and c (mean cP, $Q = 1 + P$, moment generating function $M_N(\zeta) = (Q - Pe^{\zeta})^{-1/c}$), then N' has parameters qP and c. If N has a Poisson-inverse Gaussian distribution with parameters μ and β, so

$$M_N(\zeta) = \exp\left(-\mu(\sqrt{1 + 2\beta(e^{\zeta} - 1)} - 1)\right),$$

then N is also Poisson inverse Gaussian with parameters μq and βq.

In all cases the variance of N' is lower than the variance of N and N' is closer to Poisson than N in the sense that the variance to mean ratio has decreased. For the general G-mixed Poisson the ratio of variance to mean decreases from $1 + cn$ to $1 + cqn$. As $q \to 0$ the variance to mean ratio approaches 1 and N' approaches a Poisson distribution. The fact that N' becomes Poisson is called the law of small numbers.

3.7 Negative Multinomial Distribution and Related Frequency Distributions*

When we consider mixed Poisson distributions we often regard G as carrying inter-risk correlation, or more evocatively "contagion", information about weather, the state of the economy and inflation, gas prices etc. Hence if we have two related frequency variables N_1 and N_2 we should expect to use the same G and produce a bivariate mixed Poisson where, conditional on $G = g$, N_i has a Poisson distribution with mean $n_i g$ and N_1 and N_2 are conditionally independent. The MGF of such a distribution will be

$$\begin{aligned} M(\zeta_1, \zeta_2) &= \mathrm{E}(e^{\zeta_1 N_1 + \zeta_2 N_2}) \\ &= \mathrm{E}(\mathrm{E}(e^{\zeta_1 N_1 + \zeta_2 N_2} | G)) \\ &= \mathrm{E}_G(\mathrm{E}(e^{\zeta_1 N_1} | G)\mathrm{E}(e^{\zeta_2 N_2} | G)) \\ &= \mathrm{E}_G(\exp(G(n_1(e^{\zeta_1} - 1) + n_2(e^{\zeta_2} - 1)))) \\ &= M_G(n_1(e^{\zeta_1} - 1) + n_2(e^{\zeta_2} - 1)). \end{aligned} \tag{3.21}$$

For example, if G is a gamma random variable with MGF

$$M_G(\zeta) = (1 - \beta\zeta)^{-k} \tag{3.22}$$

(mean $k\beta$, variance $k\beta^2$) we get a bivariate frequency distribution with MGF

$$\begin{aligned} M(\zeta_1, \zeta_2) &= [1 - \beta(n_1(e^{\zeta_1} - 1) + n_2(e^{\zeta_2} - 1))]^{-k} \\ &= [1 + \beta \sum_i n_i - \beta \sum_i n_i e^{\zeta_i}]^{-k} \\ &= (Q - \sum_i P_i e^{\zeta_i})^{-k} \end{aligned} \tag{3.23}$$

where $P_i = \beta n_i$, $P = \sum_i P_i$ and $Q = 1 + P$. Equation (3.23) is the moment generating function for a negative multinomial distribution, as defined in Johnson, Kotz and Kemp [23]. The negative multinomial distribution has positively correlated marginals as expected given its construction with a common contagion G.

The form of the moment generating function for negative multinomial distribution can be generalized allowing us to construct multivariate frequency distributions $(N_1, \ldots, N_t)$ where

1. Each N_i is a negative binomial.

2. The sum $N_1 + \cdots + N_t$ under the multivariate distribution is also negative binomial. (In general, the sum of independent negative binomials will not be negative binomial.)

3. The N_i are correlated.

We will call such multivariate frequencies, with common mixing distributions, G-mixed multivariate Poisson distributions.

3.7.1 Evolution of Claims Over Time*

Here is an application of the NMN distribution. If A is an aggregate distribution representing ultimate losses we may want to determine a decomposition $A = \sum_t D_t$ into a sum of losses paid at time t for $t = 1, \ldots, T$.

If $A = X_1 + \cdots + X_N$ has a compound Poisson distribution then such a decomposition is easy to arrange. Let π_t be the expected proportion of ultimate losses paid at time t, so $\sum_{t=1}^{t=T} \pi_t = 1$. By definition we mean

$$\mathrm{E}(D_t) = \pi_t \mathrm{E}(A). \tag{3.24}$$

(Equation (3.24) is a different assumption to

$$\mathrm{E}(D_t) = \pi_t \mathrm{E}(A|\text{information available at } t-1) = \pi_t A_{t-1},$$

which is closer to the problem actually faced by the reserving actuary. Our π_t's are prior estimates assumed known at time 0. These types of differences have interesting implications for actuarial methods and they are explored further in Mack [28].) Now we seek a decomposition

$$A = D_1 + D_2 + \cdots + D_T \tag{3.25}$$

but we know only (3.24). The simplest approach is to assume that severity X is independent of time and that $\pi_t n$ of the total n claims are paid at time t. If we

further assume that the number of claims paid at time t is also Poisson, then the moment generating function of $D_1 + \cdots + D_T$ is given by

$$\begin{aligned} M_{D_1+\cdots+D_T}(\zeta) &= \prod_t \exp(\pi_t n(M_X(\zeta) - 1)) \\ &= \exp(n(\sum_t \pi_t M_X(\zeta) - 1)) \\ &= \exp(n(M_X(\zeta) - 1)) \\ &= M_A(\zeta). \end{aligned} \tag{3.26}$$

Thus we have a very simple decomposition for (3.25): the individual D_t are independent compound Poisson variables with expected claim count $\pi_t n$ and severity distribution X.

Moving one step further, it is often observed in practice that average severity increases with t so the assumption that X is fixed for all t is unrealistic. It may be better to assume that losses which close at time t are samples of a random variable X_t. As above, we assume that the expected number of such losses is $\pi'_t n$ where n is the expected ultimate number of claims, and π'_t adjusts the original π_t for the difference in average severity $\mathrm{E}(X)$ vs. $\mathrm{E}(X_t)$. Now

$$\begin{aligned} M_{D_1+\cdots+D_T}(\zeta) &= \prod_t \exp(\pi'_t n(M_{X_t}(\zeta) - 1)) \\ &= \exp(n(\sum_t \pi'_t M_{X_t}(\zeta) - 1)) \\ &= \exp(n(M_{X'}(\zeta) - 1)) \\ &= M_A(\zeta) \end{aligned} \tag{3.27}$$

where X' is a mixture of the X_t with weights π'_t. Equation (3.27) is a standard result in actuarial science, see Bowers et al. [4].

If we try to replicate the compound Poisson argument using a negative binomial distribution for N we will clearly fail. However if X is defined as a mixture of X_t with weights π_t, as before, then we can write

$$M_{D_1,\ldots,D_T}(\zeta_1, \ldots, \zeta_T) = (Q - \sum_t P\pi_t M_{X_t}(\zeta_t))^{-k} \tag{3.28}$$

and so

$$M_A(\zeta) = M_{D_1,\ldots,D_T}(\zeta,\ldots,\zeta) = (Q - \sum_t P_t M_{X_t}(\zeta))^{-k} = (Q - PM_X(\zeta))^{-k} \tag{3.29}$$

where $P_t := \pi_t P$. Equation (3.28) is the MGF for a negative multinomial distribution, as defined in the previous section and Johnson, Kotz and Kemp [23]. As we have seen the negative multinomial distribution has positively correlated marginals, in line with our prior notions of liability dynamics. It therefore provides a good model for the decomposition of ultimate losses into losses paid each period.

3.7.2 Related Multivariate Frequency Distributions*

We can use the same trick with other mixing distributions than the gamma. The Poisson inverse Gaussian (PIG) distribution is an inverse Gaussian mixture of Poissons, just as the negative binomial distribution is a gamma mixture. The MGF is

$$M(\zeta) = \exp(-\tau(\sqrt{1+\beta(e^{\zeta}-1)}-1)). \tag{3.30}$$

The mean is $\tau\beta$ and the variance is $\tau\beta(1+\beta)$. We can define a multivariate PIG (MPIG) by

$$M(\zeta_1,\ldots,\zeta_T) = \exp(-\tau(\sqrt{1+\sum\beta_i(e^{\zeta_i}-1)}-1)). \tag{3.31}$$

Sichel's distribution is an generalized-inverse Gaussian mixture of Poissons. The MGF is

$$M(\zeta) = \frac{K_\gamma(\omega\sqrt{1-2\beta(e^{\zeta}-1)})}{K_\gamma(\omega)(1-2\beta(e^{\zeta}-1))^{\gamma/2}}. \tag{3.32}$$

The mean and variance are given in Johnson, Kotz and Kemp [24, page 456]. Clearly we can apply the same techniques to get another multivariate frequency distribution.

The Poisson-Pascal distribution is a Poisson-stopped sum of negative binomials. It has moment generating function

$$M(\zeta) = \exp(\theta((1 - P(e^{\zeta} - 1))^{-k} - 1)) \tag{3.33}$$

and so will also yield another multivariate family. The mean and variance are given by

$$\mu = \theta k P \tag{3.34}$$

$$\mu_2 = \theta k P(Q + kP). \tag{3.35}$$

3.7.3 Excess count interpretation of G-mixed multivariate Poisson distributions*

The reader has probably realized that a G-mixed multivariate Poisson seems closely related to a single G-mixed Poisson and a series of indicator variables, combining results from the previous sub-sections with Section 3.6. Let N be G-mixed Poisson with parameter n and $\mathrm{Var}(G) = c$. Let (N_1, N_2) be G-mixed bivariate Poisson with parameters n_1 and n_2 and the same G, so the MGF of (N_1, N_2) is

$$M_1(\zeta_1, \zeta_2) = M_G(n_1(e^{\zeta_1} - 1) + n_2(e^{\zeta_2} - 1)). \tag{3.36}$$

Finally let (I, J) be a bivariate distribution supported on $\{0, 1\} \times \{0, 1\}$ with

$$\Pr(I = 0, J = 0) = p_{00}$$
$$\Pr(I = 1, J = 0) = p_{10}$$
$$\Pr(I = 0, J = 1) = p_{01}$$
$$\Pr(I = 1, J = 1) = p_{11}$$

and $\sum p_{ij} = 1$.

We can define a new bivariate distribution from (I, J) and N as

$$(M_1, M_2) = (I_1, J_1) + \cdots + (I_N, J_N). \tag{3.37}$$

The MGF of (M_1, M_2) is

$$M_2(\zeta_1, \zeta_2) = M_G(n(p_{11}e^{\zeta_1+\zeta_2} + p_{10}e^{\zeta_1} + p_{01}e^{\zeta_2} + p_{00}). \tag{3.38}$$

Thus, if $p_{11} = 0$ we see the single-frequency sum of the bivariate (M_1, M_2) is actually a G-mixed bivariate Poisson. If $p_{00} = 0$ then $n = n_1 + n_2$, otherwise $(1 - p_{00})n = n_1 + n_2$ and there are some extraneous "zero" claims. However, if $p_{11} \neq 0$ then the single frequency sum is not a G-mixed bivariate Poisson.

Here is an interesting interpretation and application of (I, J). We can regard I as an indicator of whether a claim has been reported at time t and J and indicator of whether the claim is closed. Then

$\Pr(I = 0, J = 0) =$ meaningless

$\Pr(I = 1, J = 0) =$ reported claim which closes without payment

$\Pr(I = 0, J = 1) =$ claim not yet reported which closes with payment

$\Pr(I = 1, J = 1) =$ claim reported and closed with payment.

Combining with a distribution N of ultimate claims we can use (3.37) to produce $(M_1, M_2) = (I_1 + \cdots + I_N, J_1 + \cdots + J_N)$—a bivariate distribution of (claims reported at time t, ultimate number of claims)! Note the value $(0, 0)$ is a meaningless annoyance (it scales n) and we assume $p_{00} = 0$. The three other parameters can easily be estimated using standard actuarial methods.

Given such a bivariate and a known number of claims reported we can produce a posterior distribution of ultimate claims. Furthermore, in all these techniques we can extend the simple count indicators (I, J) to be the distribution of case incurred losses and ultimate losses. Then we would get a bivariate distribution of case incurred to date and ultimate losses. I believe there is a lot of useful information that could be wrought from these methods and that they deserve further study. They naturally give confidence intervals on reserve ranges, for example.

We end with a numerical example illustrating the theory we have developed and showing another possible application. Rather than interpreting p_{ij} as reported

and ultimate claims we could interpret them as claims from line A and line B, where there is some expectation these claim would be correlated. For example A could be auto liability and B workers compensation for a trucking insured. Let $c = 0.02$ be the common contagion and $n = 250$. Then let

$$\begin{aligned}
\Pr(I = 0, J = 0) &= 0 \\
\Pr(I = 1, J = 0) &= 0.45 \\
\Pr(I = 0, J = 1) &= 0.05 \\
\Pr(I = 1, J = 1) &= 0.50.
\end{aligned}$$

We interpret I as indicating a workers compensation claim and J as indicating an auto liability claim. The distribution says that when there is an auto liability claim ($J = 1$) there is almost always an injury to the driver, resulting in a workers compensation claim ($I = 1$). However, there are many situations where the driver is injured but there is no liability claim—such as back injuries. Overall we expect $250(0.45 + 0.50) = 237.5$ workers compensation claims and $250(0.05 + 0.5) = 137.5$ auto liability claims and 250 occurrences.

We will consider the single-frequency bivariate distribution and the negative multinomial. We have seen that the negative multinomial distribution will be slightly different because $p_{11} \neq 0$. The appropriate parameters are $n_1 = 250(p_{10} + p_{11}) = 237.5$ and $n_1 = 250(p_{01} + p_{11}) = 137.5$. Figure 3.1 shows the negative multinomial bivariate (top plot) and the single-frequency bivariate aggregate of (I, J) (bottom plot). Because of the correlation between I and J, $p_{11} = 0.5$, the lower plot shows more correlation in aggregates and the conditional distributions have less dispersion. Figure 3.2 shows the two marginal distributions, which are negative binomial $c = 0.02$ and mean 237.5 and 137.5 respectively, the sum of these two variables assuming they are independent (labelled "independent sum"), the sum assuming the negative multinomial joint distribution ("NMN Sum") which is identical to a negative binomial with $c = 0.02$ and mean $350 = 237.5 + 137.5$, the total number of claims from both lines, and finally, the sum with dependent (I, J) ("bivariate sum"). The last sum is not the same as the negative binomial sum; it has a different MGF.

Figure 3.2 also shows the difference between the sum of two independent negative binomials with means n_1 and n_2 and contagion c and a negative binomial with mean $n_1 + n_2$ and contagion c. The difference is clearly very material in the tails and is an object lesson to modelers who subdivide their book into homogeneous parts but then add up those parts assuming independence. Such an approach is *wrong* and must be avoided.

As the contagion c increases the effects of G-mixing dominate and the difference between the two bivariate distributions decreases, and conversely as c decreases to zero the effect is magnified. The value $c = 0.02$ was selected to balance these two effects.

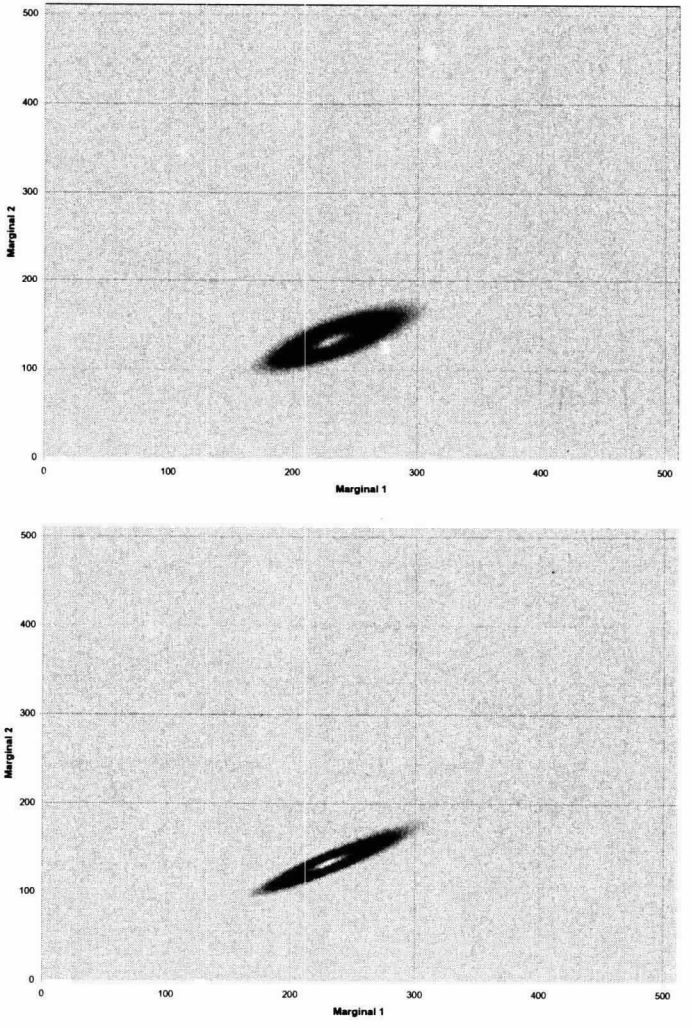


Figure 3.1: Comparison of negative multinomial (top) and single frequency bivariate claim count (bottom) bivariate distributions.

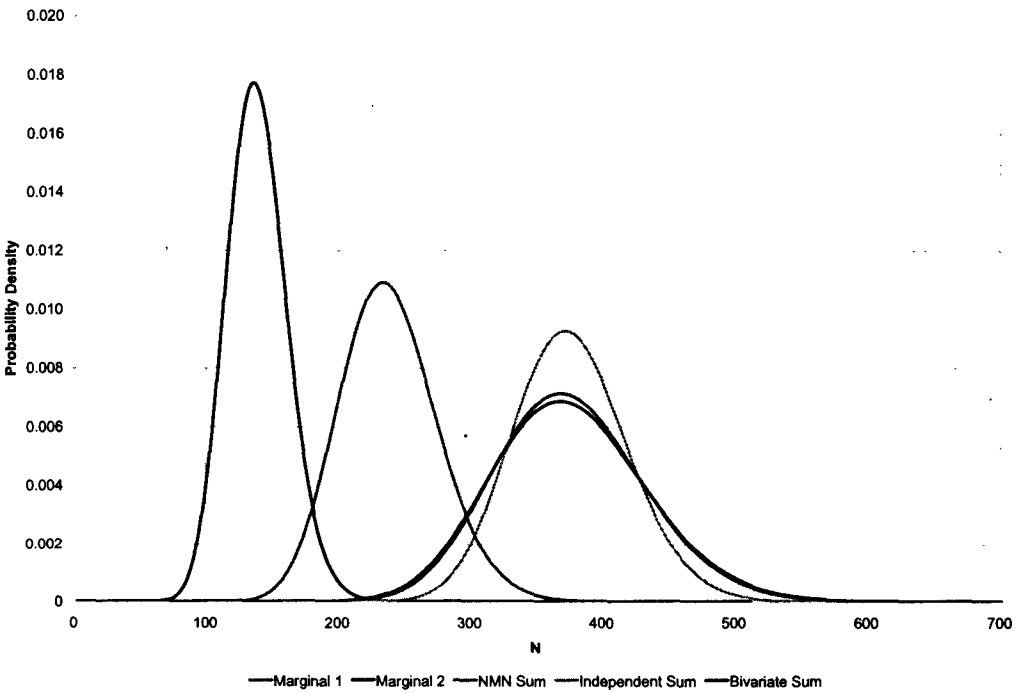


Figure 3.2: Comparison of negative multinomial and single frequency bivariate claim count marginal and total distributions.

3.8 Fast Fourier Transforms

The FFT method is a miraculous technique for computing aggregate distributions. It is especially effective when the expected claim count is relatively small and the underlying severity distribution is bounded. These assumptions are true for many excess of loss reinsurance treaties, for example. Thus the FFT is very useful when quoting excess layers with annual aggregate deductibles or other variable features. The FFT provides a discrete approximation to the moment generating function.

To use the FFT method, first "bucket" (or quantize) the severity distribution into a density vector $\mathbf{x} = (x_1, \ldots, x_m)$ whose length m is a power of two $m = 2^n$. Here

$$x_i = \Pr((i - 1/2)b < X < (i + 1/2)b) \tag{3.39}$$

$$x_1 = \Pr(X < b/2), \quad x_m = \Pr(X > (m - 1/2)b) \tag{3.40}$$

for some fixed b. We call b the bucket size. Note $\sum_i x_i = 1$ by construction. The FFT of the $m \times 1$ vector $\mathbf{x}$ is another $m \times 1$ vector $\hat{\mathbf{x}}$ whose jth component is

$$\sum_{k=0}^{2^n-1} x_k \exp(2\pi i j k / 2^n). \tag{3.41}$$

The coefficients of $\hat{\mathbf{x}}$ are complex numbers. It is also possible to express $\hat{\mathbf{x}} = \mathbf{F}\mathbf{x}$ where $\mathbf{F}$ is an appropriate matrix of complex roots of unity, so there is nothing inherently mysterious about a FFT. The trick is that there exists a very efficient algorithm for computing (3.41). Rather than taking time proportional to m^2, as one would expect, it can be computed in time proportional to $m \log(m)$. The difference between $m \log(m)$ and m^2 time is the difference between practically possible and practically impossible.

You can use the inverse FFT to recover $\mathbf{x}$ from its transform $\hat{\mathbf{x}}$. The inverse FFT is computed using the same equation (3.41) as the FFT except there is a minus sign in the exponent and the result is divided by 2^n. Because the equation is essentially the same, the inversion process can also be computed in $m \log(m)$ time.

The next step is magic in actuarial science. Remember that if N is a G-mixed Poisson and $A = X_1 + \cdots + X_N$ is an aggregate distribution then

$$M_A(\zeta) = M_G(n(M_X(\zeta) - 1)). \tag{3.42}$$

Using FFTs you can replace the *function* M_X with the discrete approximation *vector* $\hat{\mathbf{x}}$ and compute

$$\hat{\mathbf{a}} = M_G(n(\hat{\mathbf{x}} - 1)) \tag{3.43}$$

component-by-component to get an approximation vector to the function M_A. You can then use the inverse FFT to recover an discrete approximation $\mathbf{a}$ of A from $\hat{\mathbf{a}}$! See Wang [37] for more details.

Similar tricks are possible in two dimensions—see Press et al. [31] and Homer and Clark [20] for a discussion.

The FFT allows us to use the following very simple method to qualitatively approximate the density of an aggregate of dependent marginals $X_1, \ldots, X_n$ given a correlation matrix Σ. First use the FFT method to compute the sum S' of the X_i as though they were independent. Let $\mathrm{Var}(S') = \sigma'^2$ and let σ^2 be the variance of the sum of the X_i implied by Σ. Next use the FFT to add a further "noise" random variable N to S' with mean zero and variance $\sigma^2 - \sigma'^2$. Two obvious choices for the distribution of N are normal or shifted lognormal. Then $S' + N$ has the same mean and variance as the sum of the dependent variables X_i. The range of possible choices for N highlights once again that knowing the marginals and correlation structure is not enough to determine the whole multivariate distribution. It is an interesting question whether all possible choices of N correspond to actual multivariate structures for the X_i and conversely whether all multivariate structures correspond to an N. (It is easy to use MGFs to deconvolve N from the true sum using Fourier methods; the question is whether the resulting "distribution" is non-negative.)

Heckman and Meyers [17] used Fourier transforms to compute aggregate distributions by numerically integrating the characteristic function. Direct inversion of the Fourier transform is also possible using FFTs. The application of FFTs is not completely straight forward because of certain aspects of the approximations involved. The details are very clearly explained in Menn and Rachev [29]. Their method allows the use of FFTs to determine densities for distributions which have analytic MGFs but not densities—notably the class of stable distributions.

3.9 Correlated Aggregate Distributions

Here we extend some of the ideas in Section 3.7.3 from plain frequency distributions to aggregate distributions. Begin with bivariate aggregate distributions. There are two different situations which commonly arise. First we could model a bivariate severity distribution and a univariate count distribution:

$$(A, B) = (X_1, Y_1) + \cdots + (X_N, Y_N). \tag{3.44}$$

Equation (3.44) arises naturally as the distribution of losses and allocated expense, ceded and retained losses, reported and ultimate claims, and in many other situations. Secondly we could model

$$(A, B) = (X_1 + \cdots + X_M, Y_1 + \cdots + Y_N) \tag{3.45}$$

where X_i and Y_j are independent severities and (M, N) is a bivariate frequency distribution. (3.45) could be used to model losses in a clash policy.

We will use the following notation. $A = X_1 + \cdots + X_M$ and $B = Y_1 + \cdots + Y_N$ are two aggregate distributions, with X_i iid and Y_j iid, but neither X and Y nor M and N necessarily independent. Let $\mathrm{E}(X) = x$ and $\mathrm{E}(Y) = y$, $\mathrm{Var}(X) = v_x$ and $\mathrm{Var}(Y) = v_y$. Let $\mathrm{E}(M) = m$, $\mathrm{E}(N) = n$, c be the contagion of M and d that of N. Hence $\mathrm{Var}(M) = m(1 + cm)$ and $\mathrm{Var}(N) = n(1 + dn)$.

Will now calculate the correlation coefficient between A and B in four situations.

3.9.1 Correlated Severities, Single Frequency

Assume that the bivariate severity distribution $(X.Y)$ has moment generating function $M_{(X,Y)}(\zeta, \tau)$. Also assume that the claim count distribution N is a G-mixed Poisson. Then, just as for univariate aggregate distributions, the MGF of the bivariate aggregate (A, B) is

$$M_{(A,B)}(\zeta, \tau) = M_G(n(M_{(X,Y)}(\zeta, \tau) - 1)). \tag{3.46}$$

Therefore, since $\mathrm{E}(G) = 1$ and $\mathrm{E}(G^2) = 1 + c$,

$$\begin{aligned} E(AB) &= \frac{\partial^2 M_{(A,B)}}{\partial\zeta\partial\tau}\Big|_{(0,0)} \\ &= M_G''(0)n^2\frac{\partial M_{(X,Y)}}{\partial\zeta}\frac{\partial M_{(X,Y)}}{\partial\zeta} + M_G'(0)n\frac{\partial^2 M_{(X,Y)}}{\partial\zeta\partial\tau} \\ &= (1+c)n^2xy + n\mathrm{E}(XY) \\ &= (1+c)n^2xy + n\mathrm{Cov}(X,Y) + nxy. \end{aligned} \tag{3.47}$$

The value of $\mathrm{Cov}(X,Y)$ will depend on the particular bivarate severity distribution.

For example, suppose that Z represents ground up losses, X represents a retention to a and Y losses excess of a (per ground up claim), so $Z = X + Y$. Then (X,Y) is a bivariate severity distribution. Since Y is zero when $Z \le a$ we have $\mathrm{Cov}(X,Y) = (a-x)y$.

3.9.2 Bivariate Frequency

The second method for generating correlated aggregate distributions is to use a bivariate frequency distribution. So, suppose (M,N) has a G-mixed bivariate Poisson distribution. The variance of A is given by Equation (3.15). To compute the covariance of A and B write the bivariate MGF of (A,B) as

$$M_{(A,B)}(\zeta,\eta) = M(\zeta,\eta) = M_G(m(M_X(\zeta)-1)+n(M_Y(\eta)-1)) = M_G(\psi(\zeta,\eta)) \tag{3.48}$$

where the last equality defines ψ. Then, evaluating at the partial derivatives at zero, we get

$$\begin{aligned} \mathrm{E}(AB) &= \frac{\partial^2 M}{\partial\zeta\partial\eta} \\ &= \frac{\partial^2 M_G}{\partial t^2}\frac{\partial\psi}{\partial\zeta}\frac{\partial\psi}{\partial\eta} + \frac{\partial M_G}{\partial t}\frac{\partial^2\psi}{\partial\zeta\partial\eta} \\ &= (1+c)mxny. \end{aligned} \tag{3.49}$$

Hence

$$\mathrm{Cov}(A,B) = \mathrm{E}(AB) - \mathrm{E}(A)\mathrm{E}(B) = cmnxy. \tag{3.50}$$

3.9.3 Parameter Uncertainty

It is common for actuaries to work with point estimates as though they are certain. In reality there is a range around any point estimate. We now work through one possible implication of such parameter uncertainty. We will model $\mathrm{E}(A) = R$ and $\mathrm{E}(B) = S$ with R and S correlated random variables, and A and B conditionally independent given R and S. We will assume for simplicity that the severities X and Y are fixed and that the uncertainty all comes from claim counts. The reader can extend the model to varying severities as an exercise. R and S pick up uncertainty in items like the trend factor, tail factors and other economic variables, as well as the natural correlation induced through actuarial methods such as the Bornheutter-Ferguson.

Suppose $\mathrm{E}(R) = r$, $\mathrm{E}(S) = s$, $\mathrm{Var}(R) = v_r$, $\mathrm{Var}(S) = v_s$ and let ρ be the correlation coefficient between R and S.

By (3.15) the conditional distribution of $A|R$ is a mixed compound Poisson distribution with expected claim count R/x and contagion c. Therefore the conditional variance is

$$\begin{aligned} \mathrm{Var}(A|R) &= \mathrm{E}(M|R)\mathrm{Var}(X) + \mathrm{Var}(M|R)\mathrm{E}(X)^2 \\ &= R/x v_x + R/x(1 + cR/x)x^2 \\ &= xR(1 + v_x/x^2) + cR^2, \end{aligned} \tag{3.51}$$

and the unconditional variance of A is

$$\begin{aligned} \mathrm{Var}(A) &= \mathrm{E}(\mathrm{Var}(A|R)) + \mathrm{Var}(\mathrm{E}(A|R)) \\ &= \mathrm{E}(xR(v_x/x^2 + 1) + cR^2) + \mathrm{Var}(R) \\ &= xr(v_x/x^2 + 1) + c(v_r + r^2) + v_r. \end{aligned} \tag{3.52}$$

Next, because A and B are conditionally independent given R and S,

$$\begin{aligned} \mathrm{Cov}(A, B) &= \mathrm{E}(\mathrm{Cov}(A, B|R, S)) + \mathrm{Cov}(\mathrm{E}(A|R), \mathrm{E}(B|S)) \\ &= \mathrm{Cov}(R, S). \end{aligned} \tag{3.53}$$

Note Equation (3.53) is only true if we assume $A \neq B$.

3.9.4 Parameter Uncertainty and Bivariate Frequency

Finally, suppose $\mathrm{E}(A) = R$, $\mathrm{E}(B) = S$ with R and S correlated parameters *and* conditional on (R, S) suppose that (M, N) has a G-mixed bivariate Poisson distribution. By (3.50) $\mathrm{Cov}(A, B|R, S) = cRS$. The unconditional variances are as given in (3.52). The covariance term is

$$\begin{aligned}\mathrm{Cov}(A, B) &= \mathrm{E}(\mathrm{Cov}(A, B|R, S)) + \mathrm{Cov}(\mathrm{E}(A|R), \mathrm{E}(B|S)) \\ &= c\mathrm{E}(RS) + \mathrm{Cov}(R, S) \\ &= (1 + c)\mathrm{Cov}(R, S) + crs \\ &= \rho\sqrt{v_r v_s}(1 + c) + crs. \end{aligned} \tag{3.54}$$

3.10 Severity is Irrelevant

In some cases the actual form of the severity distribution is essentially irrelevant to the shape of the aggregate distribution. Consider an aggregate with a G-mixed Poisson frequency distribution. If the expected claim count n is large and if the severity is tame (roughly tame means "has a variance"; any severity from a policy with a limit is tame; unlimited workers compensation may not be tame) then particulars of the severity distribution diversify away in the aggregate. Moreover the variability from the Poisson claim count component also diversifies away and the shape of the aggregate distribution converges to the shape of the frequency mixing distribution G. Another way of saying the same thing is that the normalized distribution of aggregate losses (aggregate losses divided by expected aggregate losses) converges in distribution to G.

We can prove these assertions using moment generating functions. Let X_n be a sequence of random variables with distribution functions F_n and let X another random variable with distribution F. If $F_n(x) \to F(x)$ as $n \to \infty$ for every point of continuity of F then we say F_n converges weakly to F and that X_n converges in distribution to F.

Convergence in distribution is a relatively weak form of convergence. A stronger form is convergence in probability, which means for all $\epsilon > 0$ $\Pr(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$. If X_n

converges to X in probability then X_n also converges to X in distribution. The converse is false. For example, let $X_n = Y$ and X be binomial 0/1 random variables with $\Pr(Y = 1) = \Pr(X = 1) = 1/2$. Then X_n converges to X in distribution. However, since $\Pr(|X - Y| = 1) = 1/2$, X_n does not converge to X in probability.

It is a fact that X_n converges to X if the MGFs M_n of X_n converge to the MFG of M of X for all t: $M_n(t) \to M(t)$ as $n \to \infty$. See Feller [12] for more details. We can now prove the following theorem.

Proposition 1 *Let N be a G-mixed Poisson distribution with mean n, G with mean 1 and variance c, and let X be an independent severity with mean x and variance $x(1+\gamma^2)$. Let $A = X_1 + \cdots + X_N$ and $a = nx$. Then A/a converges in distribution to G, so*

$$\Pr(A/a < \alpha) \to \Pr(G < \alpha) \tag{3.55}$$

as $n \to \infty$. Hence

$$\sigma(A/a) = \sqrt{c + \frac{x(1+\gamma^2)}{a}} \to \sqrt{c}. \tag{3.56}$$

Proof: By (3.12)

$$M_A(\zeta) = M_G(n(M_X(\zeta) - 1)) \tag{3.57}$$

and so using Taylor's expansion we can write

$$\begin{aligned}
\lim_{n\to\infty} M_{A/a}(\zeta) &= \lim_{n\to\infty} M_A(\zeta/a) \\
&= \lim_{n\to\infty} M_G(n(M_X(\zeta/nx) - 1)) \\
&= \lim_{n\to\infty} M_G(n(M'_X(0)\zeta/nx + R(\zeta/nx))) \\
&= \lim_{n\to\infty} M_G(\zeta + nR(\zeta/nx))) \\
&= M_G(\zeta)
\end{aligned}$$

for some remainder function $R(t) = O(t^2)$. Note that the assumptions on the mean and variance of X guarantee $M'_X(0) = x = \mathrm{E}(X)$ and that the remainder term in Taylor's expansion actually is $O(t^2)$. The second part is trivial.

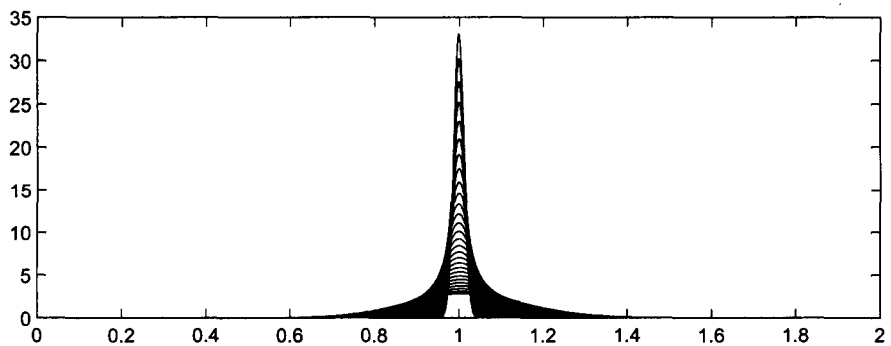


Figure 3.3: Theoretical distribution of scaled aggregate losses with no parameter or structure uncertainty and Poisson frequency.

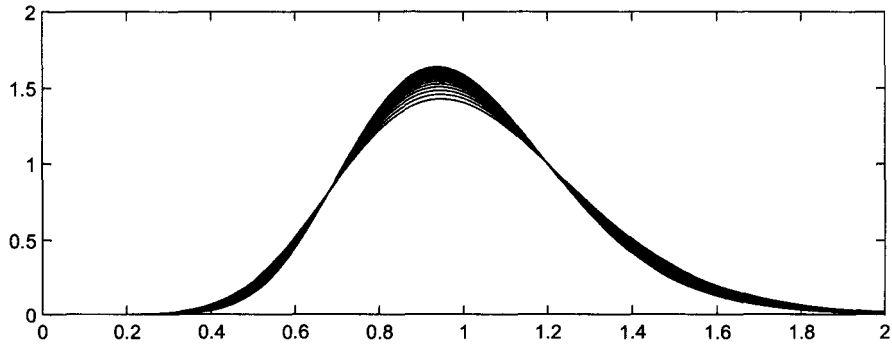


Figure 3.4: Theoretical distribution envelope of scaled aggregate losses with a gamma mixed Poisson frequency with mixing variance $c = 0.0625$.

The proposition implies that if the frequency distribution is actually a Poisson, so the mixing distribution G is $G = 1$ with probability 1, then the loss ratio distribution of a very large book will tend to the distribution concentrated at the expected, hence the expression that "with no parameter risk the process risk completely diversifies away."

Figures 3.3 and 3.4 illustrate the proposition, showing how aggregates change shape as expected counts increase.

In Figure 3.3 $G = 1$ and the claim count is Poisson. Here the scaled distributions get more and more concentrated about the expected value (scaled to 1.0).

In Figure 3.4 G has a gamma distribution with variance 0.0625 (asymptotic CV of 25%). Now the scaled aggregate distributions converge to G.

It is also interesting to compute the correlation between A and G. We have

$$\begin{aligned} \mathrm{Cov}(A,G) &= \mathrm{E}(AG) - \mathrm{E}(A)\mathrm{E}(G) \\ &= \mathrm{E}\mathrm{E}(AG|G) - nx \\ &= \mathrm{E}(nxG^2) - nx \\ &= nxc, \end{aligned} \tag{3.58}$$

and therefore

$$\mathrm{Corr}(A,G) = nxc/\sqrt{nx\gamma + n(1+cn)}\sqrt{c} \to 1 \tag{3.59}$$

as $n \to \infty$.

The proposition shows that in some situations severity is irrelevant to large books of business. However, it is easy to think of examples where severity is very important, even for large books of business. For example, severity becomes important in excess of loss reinsurance when it is not clear whether a loss distribution effectively exposes an excess layer. There, the difference in severity curves can amount to the difference between substantial loss exposure and none. The proposition does *not* say that any uncertainty surrounding the severity distribution diversifies away; it is only true when the severity distribution is known with certainty. As is often the case with risk management metrics, great care needs to be taken when applying general statements to particular situations!

Chapter 4

THE IMAN-CONOVER METHOD

Here is the basic idea of the Iman-Conover method. Given samples of n values from two known marginal distributions X and Y and a desired correlation ρ between them, re-order the samples to have the same rank order as a reference distribution, of size $n \times 2$, with linear correlation ρ. Since linear correlation and rank correlation are typically close, the re-ordered output will have approximately the desired correlation structure. What makes the IC method work so effectively is the existence of easy algorithms to determine samples from reference distributions with prescribed linear correlation structures.

Section 4.1 explains the Choleski trick for generating multivariate reference distributions with given correlation structure. Section 4.2 gives a formal algorithmic description of the IC method.

4.1 Theoretical Derivation

Suppose that $\mathbf{M}$ is an n element sample from an r dimensional multivariate distribution, so $\mathbf{M}$ is an $n \times r$ matrix. Assume that the columns of $\mathbf{M}$ are uncorrelated, have mean zero, and standard deviation one. Let $\mathbf{M}'$ denote the transpose of $\mathbf{M}$. These assumptions imply that the correlation matrix of the sample $\mathbf{M}$ can be computed as $n^{-1}\mathbf{M}'\mathbf{M}$, and because the columns are independent, $n^{-1}\mathbf{M}'\mathbf{M} = \mathbf{I}$. (There is no need to scale the covariance matrix by the row and column standard

deviations because they are all one. In general $n^{-1}\mathbf{M}'\mathbf{M}$ is the covariance matrix of $\mathbf{M}$.)

Let $\mathbf{S}$ be a correlation matrix, i.e. $\mathbf{S}$ is a positive semi-definite symmetric matrix with 1's on the diagonal and all elements ≤ 1 in absolute value. In order to rule out linearly dependent variables assume $\mathbf{S}$ is positive definite. These assumptions ensure $\mathbf{S}$ has a Choleski decomposition

$$\mathbf{S} = \mathbf{C}'\mathbf{C} \tag{4.1}$$

for some upper triangular matrix $\mathbf{C}$, see Golub [13] or Press et al. [31]. Set $\mathbf{T} = \mathbf{MC}$. The columns of $\mathbf{T}$ still have mean zero, because they are linear combinations of the columns of $\mathbf{M}$ which have zero mean by assumption. It is less obvious, but still true, that the columns of $\mathbf{T}$ still have standard deviation one. To see why, remember that the covariance matrix of $\mathbf{T}$ is

$$n^{-1}\mathbf{T}'\mathbf{T} = n^{-1}\mathbf{C}'\mathbf{M}'\mathbf{M}\mathbf{C} = \mathbf{C}'\mathbf{C} = \mathbf{S}, \tag{4.2}$$

since $n^{-1}\mathbf{M}'\mathbf{M} = \mathbf{I}$ is the identity by assumption. Now $\mathbf{S}$ is actually the correlation matrix too because the diagonal is scaled to one, so the covariance and correlation matrices coincide. The process of converting $\mathbf{M}$, which is easy to simulate, into $\mathbf{T}$, which has the desired correlation structure $\mathbf{S}$, is the theoretical basis of the IC method.

It is important to note that estimates of correlation matrices, depending on how they are constructed, need not have the mathematical properties of a correlation matrix. Therefore, when trying to use an estimate of a correlation matrix in an algorithm, such as the Iman-Conover, which actually requires a proper correlation matrix as input, it may be necessary to check the input matrix does have the correct mathematical properties.

Next we discuss how to make $n \times r$ matrices $\mathbf{M}$, with independent, mean zero columns. The basic idea is to take n numbers $a_1, \ldots, a_n$ with $\sum_i a_i = 0$ and $n^{-1}\sum_i a_i^2 = 1$, use them to form one $n \times 1$ column of $\mathbf{M}$, and then to copy it r times. Finally randomly permute the entries in each column to make them independent as columns of random variables. Iman and Conover call the a_i "scores".

They discuss several possible definitions for the scores, including scaled versions of $a_i = i$ (ranks) and a_i uniformly distributed. They note that the shape of the output multivariate distribution depends on the scores. All of the examples in their paper use normal scores. We will discuss normal scores here, and consider alternatives in Section 4.4.1.

Given that the scores will be based on normal random variables, we can either simulate n random standard normal variables and then shift and re-scale to ensure mean zero and standard deviation one, or we can use a stratified sample from the standard normal, $a_i = \Phi^{-1}(i/(n+1))$. By construction, the stratified sample has mean zero which is an advantage. Also, by symmetry, using the stratified sample halves the number of calls to Φ^{-1}. For these two reasons we prefer it in the algorithm below.

The correlation matrix of $\mathbf{M}$, constructed by randomly permuting the scores in each column, will only be approximately equal to $\mathbf{I}$ because of random simulation error. In order to correct for the slight error which could be introduced Iman and Conover use another adjustment in their algorithm. Let $\mathbf{E} = n^{-1}\mathbf{M}'\mathbf{M}$ be the actual correlation matrix of $\mathbf{M}$ and let $\mathbf{E} = \mathbf{F}'\mathbf{F}$ be the Choleski decomposition of $\mathbf{E}$, and define $\mathbf{T} = \mathbf{M}\mathbf{F}^{-1}\mathbf{C}$. The columns of $\mathbf{T}$ have mean zero, and the covariance matrix of $\mathbf{T}$ is

$$\begin{aligned} n^{-1}\mathbf{T}'\mathbf{T} &= n^{-1}\mathbf{C}'\mathbf{F}'^{-1}\mathbf{M}'\mathbf{M}\mathbf{F}^{-1}\mathbf{C} \\ &= \mathbf{C}'\mathbf{F}'^{-1}\mathbf{E}\mathbf{F}^{-1}\mathbf{C} \\ &= \mathbf{C}'\mathbf{F}'^{-1}\mathbf{F}'\mathbf{F}\mathbf{F}^{-1}\mathbf{C} \\ &= \mathbf{C}'\mathbf{C} \\ &= \mathbf{S}, \end{aligned} \tag{4.3}$$

and hence $\mathbf{T}$ has correlation matrix exactly equal to $\mathbf{S}$, as desired. If $\mathbf{E}$ is singular then the column shuffle needs to be repeated.

Now the reference distribution $\mathbf{T}$ with exact correlation structure $\mathbf{S}$ is in hand, all that remains to complete the IC method is to re-order the each column of the input distribution $\mathbf{X}$ to have the same rank order as the corresponding column of $\mathbf{T}$.

4.2 Algorithm

Here is a more algorithmic description of the IC method. The description uses normal scores and the Choleski method to determine the reference distribution. As we discussed above, it is possible to make other choices in place of these and they are discussed in Section 4.4. We will actually present two versions of the core algorithm. The first, called "Simple Algorithm" deals with the various matrix operations at a high level. The second "Detailed Algorithm" takes a more sophisticated approach to the matrix operations, including referencing appropriate Lapack routines [1]. Lapack is a standard set of linear algebra functions. Software vendors provide very high performance implementations of Lapack, many of which are used in CPU benchmarks. Several free Windows implementations are available on the web. The software described in the Appendix uses the Intel Performance http://www.intel.com/software/products/perflib/. The reader should study the simple algorithm first to understand what is going in the IC method. In order to code a high performance implementation you should follow the steps outlined in the detailed algorithm. Both algorithms have the same inputs and outputs.

Inputs: An $n \times r$ matrix $\mathbf{X}$ consisting of n samples from each of r marginal distributions, and a desired correlation matrix $\mathbf{S}$.

The IC method does not address how the columns of $\mathbf{X}$ are determined. It is presumed that the reader has sampled from the appropriate distributions in some intelligent manner. The matrix $\mathbf{S}$ must be a correlation matrix for linearly independent random variables, so it must be symmetric and positive definite. If $\mathbf{S}$ is not symmetric positive semi-definite the algorithm will fail at the Choleski decomposition step. The output is a matrix $\mathbf{T}$ each of whose columns is a permutation of the corresponding column of $\mathbf{X}$ and whose approximate correlation matrix is $\mathbf{S}$.

Simple Algorithm:

1. Make one column of scores $a_i = \Phi^{-1}(i/(n+1))$ for $i = 1, \ldots, n$ and rescale to have standard deviation one.

2. Copy the scores r times to make the score matrix $\mathbf{M}$.

3. Randomly permute the entries in each column of $\mathbf{M}$.

4. Compute the correlation matrix $\mathbf{E} = n^{-1}\mathbf{M}'\mathbf{M}$ of $\mathbf{M}$.

5. Compute the Choleski decomposition $\mathbf{E} = \mathbf{F}'\mathbf{F}$ of $\mathbf{E}$.

6. Compute the Choleski decomposition $\mathbf{S} = \mathbf{C}'\mathbf{C}$ of the desired correlation matrix $\mathbf{S}$.

7. Compute $\mathbf{T} = \mathbf{M}\mathbf{F}^{-1}\mathbf{C}$. The matrix $\mathbf{T}$ has exactly the desired correlation structure by Equation (4.3).

8. Let $\mathbf{Y}$ be the input matrix $\mathbf{X}$ with each column reordered to have exactly the same rank ordering as the corresponding column of $\mathbf{T}$.

Detailed Algorithm:

1. Compute the Choleski decomposition of $\mathbf{S}$, $\mathbf{S} = \mathbf{C}'\mathbf{C}$, with $\mathbf{C}$ upper triangular. If the Choleski algorithm fails then $\mathbf{S}$ is not a valid correlation matrix. Flag an error and exit. Checking $\mathbf{S}$ is a correlation matrix in Step 1 avoids performing wasted calculations and allows the routine to exit as quickly as possible. Also check that all the diagonal entries of $\mathbf{S}$ are 1 so $\mathbf{S}$ has full rank. Again flag an error and exit if not. The Lapack routine DPOTRF can use be used to compute the Choleski decomposition. In the absence of Lapack, $\mathbf{C} = (c_{ij})$ can be computed recursively using

$$c_{ij} = \frac{s_{ij} - \sum_{k=1}^{j-1} c_{ik}c_{jk}}{\sqrt{1 - \sum_{k=1}^{j-1} c_{jk}^2}} \tag{4.4}$$

for $1 \le i \le j \le n$—since all the diagonal elements of S equal one. The empty sum $\sum_0^0 = 0$ and for $j > i$ the denominator of (4.4) equals c_{ii} and the elements of $\mathbf{C}$ should be calculated from left to right, top to bottom. See Wang [37, p. 889] or Herzog [19].

2. Let $m = \lfloor n/2 \rfloor$ be the largest integer less than or equal to $n/2$ and $v_i = \Phi^{-1}(i/(2m+1))$ for $i = 1, \ldots, m$.

3. If n is odd set

$$\mathbf{v} = (v_m, v_{m-1}, \ldots, v_1, 0, -v_1, \ldots, -v_m)$$

and if n is even set

$$\mathbf{v} = (v_m, v_{m-1}, \ldots, v_1, -v_1, \ldots, -v_m).$$

Here we have chosen to use normal scores. Other distributions could be used in place of the normal, as discussed in Section 4.4.1. Also note that by taking advantage of the symmetry of the normal distribution halves the number of calls to Φ^{-1} which is relatively computationally expensive. If multiple calls will be made to the IC algorithm then store $\mathbf{v}$ for use in future calls.

4. Form the $n \times r$ score matrix $\mathbf{M}$ from r copies of the scores vector $\mathbf{v}$.

5. Compute $m_{xx} = n^{-1} \sum_i v_i^2$, the variance of $\mathbf{v}$. Note that $\sum_i v_i = 0$ by construction.

6. Randomly shuffle columns $2, \ldots, r$ of the score matrix $\mathbf{M}$. Knuth [26, pp.139–41] gives the following algorithm for a random shuffle, which we have implemented it in Visual Basic.

```
'' vtemp[0 to n-1] is the array being shuffled.
'' vtemp[j] is the end, you work backwards up the
''    array shuffling each element.
'' Rnd() returns a uniform random variable
'' between zero and one.

dim j as long, vtemp[0 to n-1] as double
dim temp as double, u as double
''
''  populate vtemp
''
```

```
j=n-1
do while j > 0
  u = Rnd()
  k = CLng(j * u)
  temp = vtemp[j]
  vtemp[j] = vtemp[k]
  vtemp[k] = temp
  j=j-1
loop
```

7. Compute the correlation matrix $\mathbf{E}$ of the shuffled score matrix $\mathbf{M}$. Each column of $\mathbf{M}$ has mean zero, by construction, and variance m_{xx}. The correlation matrix is obtained by dividing each element of $\mathbf{M}'\mathbf{M}$ by m_{xx}. The matrix product can be computed using the Lapack routine DGEMM. If $\mathbf{E}$ is singular repeat step 6.

8. Determine Choleski decomposition $\mathbf{E} = \mathbf{F}'\mathbf{F}$ of $\mathbf{E}$ using the Lapack routine DPOTRF. Because $\mathbf{E}$ is a correlation matrix it must be symmetric and positive definite and so is guaranteed to have a Choleski root.

9. Compute $\mathbf{F}^{-1}\mathbf{C}$ using the Lapack routine DTRTRS to solve the linear equation $\mathbf{FA} = \mathbf{C}$ for $\mathbf{A}$. Solving the linear equation avoids a time consuming matrix inversion and multiplication. The routine DTRTRS is optimized for upper triangular input matrices.

10. Compute the correlated scores $\mathbf{T} = \mathbf{M}\mathbf{F}^{-1}\mathbf{C} = \mathbf{MA}$ using DGEMM. The matrix $\mathbf{T}$ has exactly the desired correlation structure.

11. Compute the ranks of the elements of $\mathbf{T}$. Ranks are computed by indexing the columns of $\mathbf{T}$ as described in Section 8.4 of [31]. Let $r(k)$ denote the index of the kth ranked element of $\mathbf{T}$. See Appendix B for VBA code to perform indexing.

12. Let $\mathbf{Y}$ be the $n \times r$ matrix with ith column equal to the ith column of the input matrix $\mathbf{X}$ given the same rank order as $\mathbf{T}$. The re-ordering is performed using the ranks computed in the previous step. First sort the input columns into ascending order if they are not already sorted and then set $\mathbf{Y}_{i,k} = \mathbf{X}_{i,r(k)}$.

Outputs: The output of the algorithm is a matrix $\mathbf{Y}$ each of whose columns is a permutation of the corresponding column of the input matrix $\mathbf{X}$. The rank correlation matrix of $\mathbf{Y}$ is identical to that of a multivariate distribution with correlation matrix $\mathbf{S}$.

4.3 Simple Example of Iman-Conover

Having explained the IC method, we now give a simple example to explicitly show all the details. The example will work with $n = 20$ samples and $r = 4$ different marginals. The marginals are samples from four lognormal distributions, with parameters $\mu = 12, 11, 10, 10$ and $\sigma = 0.15, 0.25, 0.35, 0.25$. The input matrix is

$$\mathbf{X} = \begin{pmatrix} 123,567 & 44,770 & 15,934 & 13,273 \\ 126,109 & 45,191 & 16,839 & 15,406 \\ 138,713 & 47,453 & 17,233 & 16,706 \\ 139,016 & 47,941 & 17,265 & 16,891 \\ 152,213 & 49,345 & 17,620 & 18,821 \\ 153,224 & 49,420 & 17,859 & 19,569 \\ 153,407 & 50,686 & 20,804 & 20,166 \\ 155,716 & 52,931 & 21,110 & 20,796 \\ 155,780 & 54,010 & 22,728 & 20,968 \\ 161,678 & 57,346 & 24,072 & 21,178 \\ 161,805 & 57,685 & 25,198 & 23,236 \\ 167,447 & 57,698 & 25,393 & 23,375 \\ 170,737 & 58,380 & 30,357 & 24,019 \\ 171,592 & 60,948 & 30,779 & 24,785 \\ 178,881 & 66,972 & 32,634 & 25,000 \\ 181,678 & 68,053 & 33,117 & 26,754 \\ 184,381 & 70,592 & 35,248 & 27,079 \\ 206,940 & 72,243 & 36,656 & 30,136 \\ 217,092 & 86,685 & 38,483 & 30,757 \\ 240,935 & 87,138 & 39,483 & 35,108 \end{pmatrix}. \tag{4.5}$$

Note that the marginals are all sorted in ascending order. The algorithm does not actually require pre-sorting the marginals but it simplifies the last step.

The desired target correlation matrix is

$$\mathbf{S} = \begin{pmatrix} 1.000 & 0.800 & 0.400 & 0.000 \\ 0.800 & 1.000 & 0.300 & -0.200 \\ 0.400 & 0.300 & 1.000 & 0.100 \\ 0.000 & -0.200 & 0.100 & 1.000 \end{pmatrix}. \tag{4.6}$$

The Choleski decomposition of **S** is

$$\mathbf{C} = \begin{pmatrix} 1.000 & 0.800 & 0.400 & 0.000 \\ 0.000 & 0.600 & -0.033 & -0.333 \\ 0.000 & 0.000 & 0.916 & 0.097 \\ 0.000 & 0.000 & 0.000 & 0.938 \end{pmatrix}. \tag{4.7}$$

Now we make the score matrix. The basic scores are $\Phi^{-1}(i/21)$, for $i = 1, \ldots, 20$. We scale these by 0.868674836252965 to get a vector **v** with standard deviation one. Then we combine four **v**'s and shuffle randomly to get

$$\mathbf{M} = \begin{pmatrix} -1.92062 & 1.22896 & -1.00860 & -0.49584 \\ -1.50709 & -1.50709 & -1.50709 & 0.82015 \\ -1.22896 & 1.92062 & 0.82015 & -0.65151 \\ -1.00860 & -0.20723 & 1.00860 & -1.00860 \\ -0.82015 & 0.82015 & 0.34878 & 1.92062 \\ -0.65151 & -1.22896 & -0.65151 & 0.20723 \\ -0.49584 & -0.65151 & 1.22896 & -0.34878 \\ -0.34878 & -0.49584 & -0.49584 & -0.06874 \\ -0.20723 & -1.00860 & 0.20723 & 0.65151 \\ -0.06874 & 0.49584 & 0.06874 & -1.22896 \\ 0.06874 & -0.34878 & -1.22896 & 0.49584 \\ 0.20723 & 0.34878 & 0.65151 & 0.34878 \\ 0.34878 & -0.06874 & -0.20723 & 1.22896 \\ 0.49584 & -1.92062 & -0.82015 & -0.20723 \\ 0.65151 & 0.20723 & 1.92062 & -1.92062 \\ 0.82015 & 1.00860 & 1.50709 & 1.50709 \\ 1.00860 & -0.82015 & -1.92062 & 1.00860 \\ 1.22896 & 1.50709 & 0.49584 & -1.50709 \\ 1.50709 & 0.06874 & -0.06874 & 0.06874 \\ 1.92062 & 0.65151 & -0.34878 & -0.82015 \end{pmatrix}. \tag{4.8}$$

As described in Section 4.1, **M** is approximately independent. In fact **M** has covariance matrix

$$\mathbf{E} = \begin{pmatrix} 1.0000 & 0.0486 & 0.0898 & -0.0960 \\ 0.0486 & 1.0000 & 0.4504 & -0.2408 \\ 0.0898 & 0.4504 & 1.0000 & -0.3192 \\ -0.0960 & -0.2408 & -0.3192 & 1.0000 \end{pmatrix} \tag{4.9}$$

and $\mathbf{E}$ has Choleski decomposition

$$\mathbf{F} = \begin{pmatrix} 1.0000 & 0.0486 & 0.0898 & -0.0960 \\ 0.0000 & 0.9988 & 0.4466 & -0.2364 \\ 0.0000 & 0.0000 & 0.8902 & -0.2303 \\ 0.0000 & 0.0000 & 0.0000 & 0.9391 \end{pmatrix} \tag{4.10}$$

Thus $\mathbf{T} = \mathbf{M}\mathbf{F}^{-1}\mathbf{C}$ is given by

$$\mathbf{T} = \begin{pmatrix} -1.92062 & -0.74213 & -2.28105 & -1.33232 \\ -1.50709 & -2.06697 & -1.30678 & 0.54577 \\ -1.22896 & 0.20646 & -0.51141 & -0.94465 \\ -1.00860 & -0.90190 & 0.80546 & -0.65873 \\ -0.82015 & -0.13949 & -0.31782 & 1.76960 \\ -0.65151 & -1.24043 & -0.27999 & 0.23988 \\ -0.49584 & -0.77356 & 1.42145 & 0.23611 \\ -0.34878 & -0.56670 & -0.38117 & -0.14744 \\ -0.20723 & -0.76560 & 0.64214 & 0.97494 \\ -0.06874 & 0.24487 & -0.19673 & -1.33695 \\ 0.06874 & -0.15653 & -1.06954 & 0.14015 \\ 0.20723 & 0.36925 & 0.56694 & 0.51206 \\ 0.34878 & 0.22754 & -0.06362 & 1.19551 \\ 0.49584 & -0.77154 & 0.26828 & 0.03168 \\ 0.65151 & 0.62666 & 2.08987 & -1.21744 \\ 0.82015 & 1.23804 & 1.32493 & 1.85680 \\ 1.00860 & 0.28474 & -1.23688 & 0.59246 \\ 1.22896 & 1.85260 & 0.17411 & -1.62428 \\ 1.50709 & 1.20294 & 0.39517 & 0.13931 \\ 1.92062 & 1.87175 & -0.04335 & -0.97245 \end{pmatrix} \tag{4.11}$$

An easy calculation will verify that $\mathbf{T}$ has correlation matrix $\mathbf{S}$, as required.

To complete the IC method we must re-order each column of $\mathbf{X}$ to have the same rank order as $\mathbf{T}$. The first column does not change because it is already in ascending order. In the second column, the first element of $\mathbf{Y}$ must be the 14th element of $\mathbf{X}$, the second the 20th, third 10th and so on. The ranks of the other elements are

$$\begin{pmatrix} 14 & 20 & 10 & 18 & 11 & 19 & 17 & 13 & 15 & 8 & 12 & 6 & 9 & 16 & 5 & 3 & 7 & 2 & 4 & 1 \\ 20 & 19 & 16 & 4 & 14 & 13 & 2 & 15 & 5 & 12 & 17 & 6 & 11 & 8 & 1 & 3 & 18 & 9 & 7 & 10 \\ 18 & 6 & 15 & 14 & 2 & 8 & 9 & 13 & 4 & 19 & 10 & 7 & 3 & 12 & 17 & 1 & 5 & 20 & 11 & 16 \end{pmatrix}'$$

and the resulting re-ordering of $\mathbf{X}$ is

$$\mathbf{T} = \begin{pmatrix} 123,567 & 50,686 & 15,934 & 16,706 \\ 126,109 & 44,770 & 16,839 & 25,000 \\ 138,713 & 57,685 & 17,620 & 19,569 \\ 139,016 & 47,453 & 35,248 & 20,166 \\ 152,213 & 57,346 & 20,804 & 30,757 \\ 153,224 & 45,191 & 21,110 & 24,019 \\ 153,407 & 47,941 & 38,483 & 23,375 \\ 155,716 & 52,931 & 17,859 & 20,796 \\ 155,780 & 49,420 & 33,117 & 27,079 \\ 161,678 & 58,380 & 22,728 & 15,406 \\ 161,805 & 54,010 & 17,265 & 23,236 \\ 167,447 & 66,972 & 32,634 & 24,785 \\ 170,737 & 57,698 & 24,072 & 30,136 \\ 171,592 & 49,345 & 30,357 & 20,968 \\ 178,881 & 68,053 & 39,483 & 16,891 \\ 181,678 & 72,243 & 36,656 & 35,108 \\ 184,381 & 60,948 & 17,233 & 26,754 \\ 206,940 & 86,685 & 25,393 & 13,273 \\ 217,092 & 70,592 & 30,779 & 21,178 \\ 240,935 & 87,138 & 25,198 & 18,821 \end{pmatrix}. \tag{4.12}$$

The rank correlation matrix of $\mathbf{Y}$ is exactly $\mathbf{S}$. The actual linear correlation is only approximately equal to $\mathbf{S}$. The achieved value is

$$\begin{pmatrix} 1.00 & 0.85 & 0.26 & -0.11 \\ 0.85 & 1.00 & 0.19 & -0.20 \\ 0.26 & 0.19 & 1.00 & 0.10 \\ -0.11 & -0.20 & 0.10 & 1.00 \end{pmatrix}, \tag{4.13}$$

a fairly creditable performance given the input correlation matrix and the very small number of samples $n = 20$. When used with larger sized samples the IC method typically produces a very close approximation to the required correlation matrix, especially when the marginal distributions are reasonably symmetric.

4.4 Extensions of Iman-Conover

Following through the explanation of the IC method shows that it relies on a choice of multivariate reference distribution. A straightforward method to compute a reference is to use the Choleski decomposition method Equation (4.2) applied to certain independent scores. The example in Section 4.3 used normal

scores. However nothing prevents us from using other distributions for the scores provided they are suitably normalized to have mean zero and standard deviation one. We explore the impact of different choices of score distribution on the resulting multivariate distribution in Section 4.4.1.

Another approach to IC is to use a completely different multivariate distribution as reference. There are several other families of multivariate distributions, including the elliptically contoured distribution family (which includes the normal and t as a special cases) and multivariate Laplace distribution, which are easy to simulate from. We explore the impact of changing the reference distribution in Section 4.4.2. Note that changing scores is actually an example of changing the reference distribution; however, for the examples we consider the exact form of the new reference is unknown.

4.4.1 Alternative Scores

The choice of score distribution has a profound effect on the multivariate distribution output by the IC method. Recall that the algorithm described in Section 4.2 used normally distributed scores. We now show the impact of using exponentially and uniformly distributed scores.

Figure 4.1 shows three bivariate distributions with identical marginal distributions (shown in the lower right hand plot), the same correlation coefficient of 0.643 ± 0.003 but using normal scores (top left), exponential scores (top rigtht) and uniform scores (lower left). The input correlation to the IC method was 0.65 in all three cases and there are 1000 pairs in each plot. Here the IC method produced bivariate distributions with actual correlation coefficient extremely close to the requested value.

The normal scores produce the most natural looking bivariate distribution, with approximately elliptical contours. The bivariate distributions with uniform or exponential scores look unnatural, but it is important to remember that if all you know about the bivariate distribution are the marginals and correlation coefficient all three outcomes are possible.

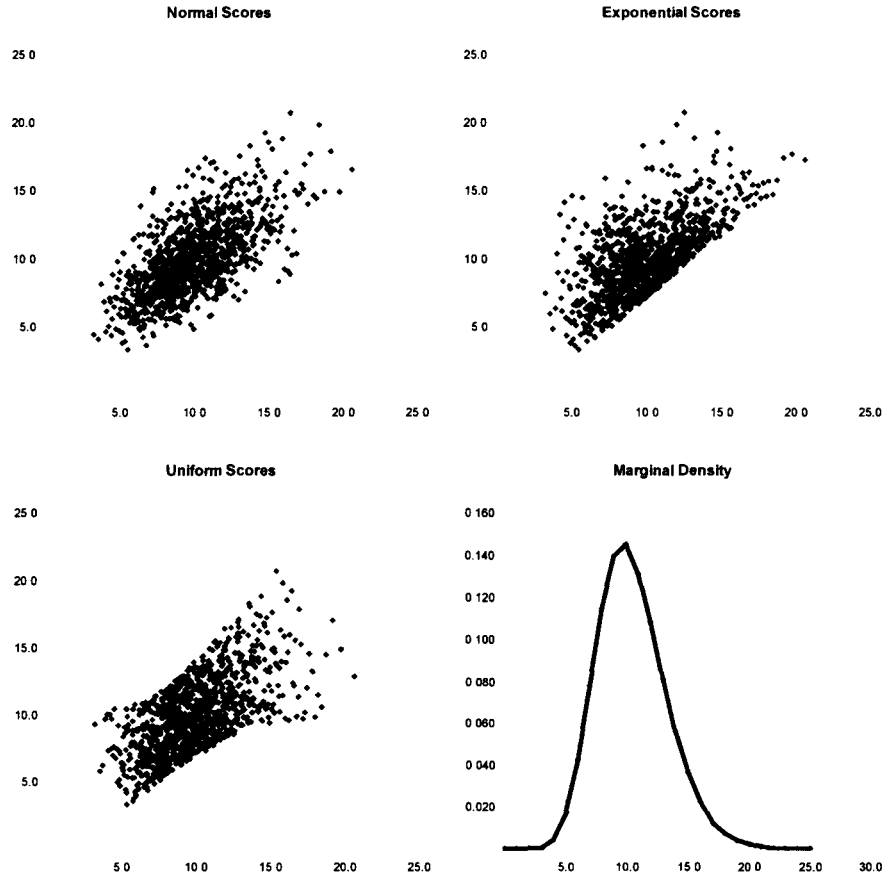


Figure 4.1: Bivariate distributions with normal, uniform and exponential scores.

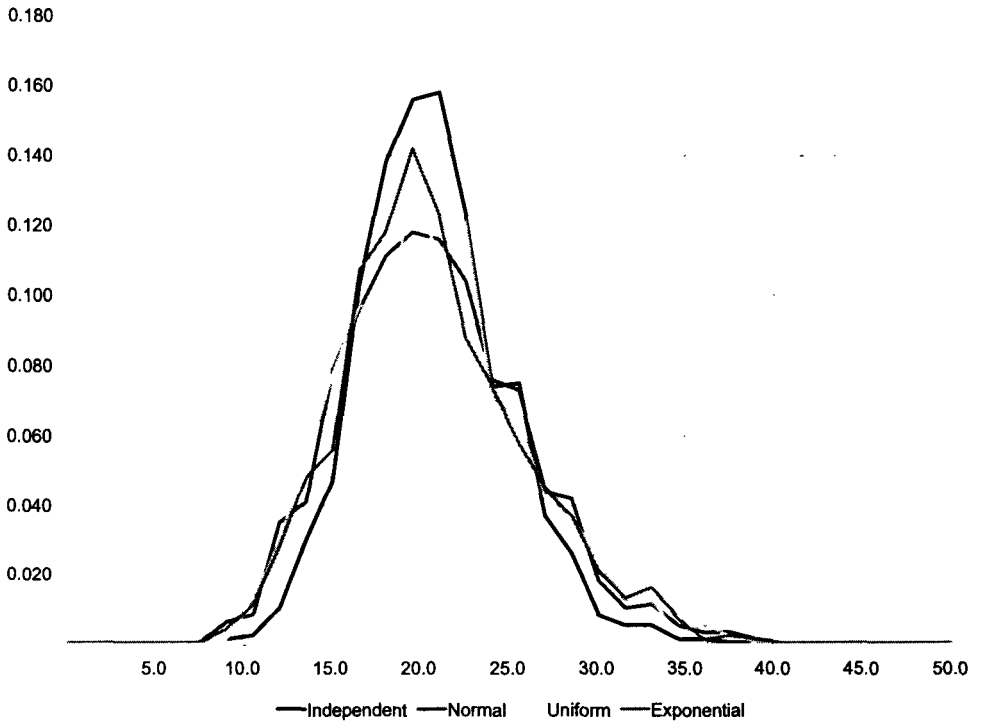


Figure 4.2: Sum of marginals from bivariate distributions made with different score distributions.

Figure 4.2 shows the distribution of the sum of the two marginals for each of the three bivariate distributions in Figure 4.1 and for independent marginals. The sum with exponential scores has a higher kurtosis (is more peaked) than with normal scores. As expected all three dependent sums have visibly thicker tails than the independent sum.

Iman and Conover considered various different score distributions in their paper. They preferred normal scores as giving more natural looking, elliptical contours. Certainly, the contours produced using exponential or uniform scores appear unnatural. If nothing else they provide a sobering reminder that knowing the marginal distributions and correlation coefficient of a bivariate distribution does not come close to fully specifying it!

4.4.2 Multivariate Reference Distributions

The IC method needs some reference multivariate distribution to determine an appropriate rank ordering for the input marginals. So far we have discussed using the Choleski decomposition trick in order to determine a multivariate normal reference distribution. However, any distribution can be used as reference provided it has the desired correlation structure. Multivariate distributions that are closely related by formula to the multivariate normal, such as elliptically contoured distributions and asymmetric Laplace distributions, can be simulated using the Choleski trick.

Elliptically contoured distributions are a family which extends the normal. For a more detailed discussion see Fang and Zhang [11]. The multivariate t-distribution and symmetric Laplace distributions are in the elliptically contoured family. Elliptically contoured distributions must have characteristic equations of the form

$$\Phi(\mathbf{t}) = \exp(i\mathbf{t}'\mathbf{m})\phi(\mathbf{t}'\mathbf{S}\mathbf{t}) \tag{4.14}$$

for some $\phi : \mathbf{R} \to \mathbf{R}$, where $\mathbf{m}$ is an $r \times 1$ vector of means and $\mathbf{S}$ is a $r \times r$ covariance matrix (nonnegative definite and symmetric). In one dimension the

elliptically contoured distributions coincide with the symmetric distributions. The covariance is $\mathbf{S}$, if it is defined.

If $\mathbf{S}$ has rank r then an elliptically contoured distribution $\mathbf{x}$ has a stochastic representation

$$\mathbf{x} = \mathbf{m} + R\mathbf{T}'\mathbf{u}^{(r)} \tag{4.15}$$

where $\mathbf{T}$ is the Choleski decomposition of $\mathbf{S}$, so $\mathbf{S} = \mathbf{T}'\mathbf{T}$, $\mathbf{u}^{(r)}$ is a uniform distribution on the sphere in $\mathbf{R}^r$, and R is a scale factor independent of $\mathbf{u}^{(r)}$. The idea here should be clear: pick a direction on the sphere, adjust by $\mathbf{T}$, scale by a distance R and finally translate by the means $\mathbf{m}$. A uniform distribution on a sphere can be created as $\mathbf{x}/\|\mathbf{x}\|$ where $\mathbf{x}$ has a multivariate normal distribution with identity covariance matrix. (By definition, $\|\mathbf{x}\|^2 = \sum_i x_i^2$ has a χ_r^2 distribution.) Uniform vectors $\mathbf{u}^{(r)}$ can also be created by applying a random orthogonal matrix to a fixed vector $(1, 0, \ldots, 0)$ on the sphere. Diaconis [8] describes a method for producing random orthogonal matrices.

The t-copula with ν degrees of freedom has a stochastic representation

$$\mathbf{x} = \mathbf{m} + \frac{\sqrt{\nu}}{\sqrt{S}}\mathbf{z} \tag{4.16}$$

where $S \sim \chi_\nu^2$ and $\mathbf{z}$ is multivariate normal with means zero and covariance matrix $\mathbf{S}$. Thus one can easily simulate from the multivariate t by first simulating multivariate normals and then simulating an independent S and multiplying.

The multivariate Laplace distribution is discussed in Kotz, Kozubowski and Podgorski [27]. It comes in two flavors: symmetric and asymmetric. The symmetric distribution is also an elliptically contoured distribution. It has characteristic function of the form

$$\Phi(\mathbf{t}) = \frac{1}{1 + \mathbf{t}'\mathbf{S}\mathbf{t}/2} \tag{4.17}$$

where $\mathbf{S}$ is the covariance matrix. To simulate from (4.17) use the fact that $\sqrt{W}\mathbf{X}$ has a symmetric Laplace distribution if W is exponential and $\mathbf{X}$ a multivariate normal with covariance matrix $\mathbf{S}$.

The multivariate asymmetric Laplace distribution has characteristic function

$$\Psi(\mathbf{t}) = \frac{1}{1 + \mathbf{t}'\mathbf{S}\mathbf{t}/2 - i\mathbf{m}'\mathbf{t}}. \tag{4.18}$$

To simulate from (4.18) use the fact that

$$\mathbf{m}W + \sqrt{W}\mathbf{X} \tag{4.19}$$

has a symmetric Laplace distribution if W is exponential and $\mathbf{X}$ a multivariate normal with covariance matrix $\mathbf{S}$ and means zero. The asymmetric Laplace is not an elliptically contoured distribution.

Figure 4.3 compares IC samples produced using a normal copula to those produced with a t-copula. In both cases the marginals are normally distributed with mean zero and unit standard deviation. The t-copula has $\nu = 2$ degrees of freedom. In both figures the marginals are uncorrelated, but in the right the marginals are not independent. The t-copula has pinched tails, similar to Venter's Heavy Right Tailed copula [33]

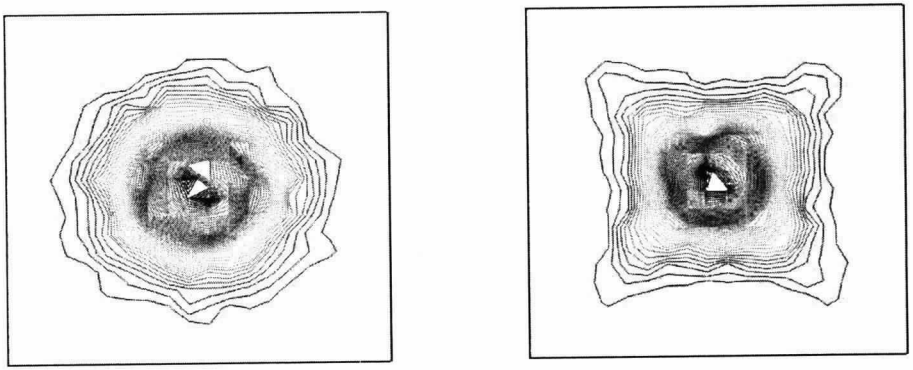


Figure 4.3: IC samples produced from the same marginal and correlation matrix using the normal and t copula reference distributions.

4.4.3 Algorithms for Extended Methods

In Section 4.4.2 we described how the IC method can be extended by using different reference multivariate distributions. It is easy to change the IC algorithm to incorporate different reference distributions for t-copulas and asymmetric Laplace distributions. Follow the detailed algorithm to step 10. Then use the stochastic representation (4.16) (resp. 4.19 for the Laplace): simulate from the scaling distribution for each row and multiply each component by the resulting number, resulting in an adjusted **T** matrix. Then complete steps 11 and 12 of the detailed algorithm.

4.5 Comparison With the Normal Copula Method

By the normal copula method we mean the following algorithm, described in Wang [37] and Herzog [19].

Inputs: A set of correlated risks $(X_1, \ldots, X_r)$ with marginal cumulative distribution functions F_i and Kendall's tau $\tau_{ij} = \tau(X_i, X_j)$ or rank correlation coefficients $r(X_i, X_j)$.

Algorithm:

1. Convert Kendall's tau or rank correlation coefficient to correlation using

$$\rho_{ij} = \sin(\pi\tau_{ij}/2) = 2\sin(\pi r_{ij}/6) \tag{4.20}$$

   and construct the Choleski decomposition $\mathbf{S} = \mathbf{C}'\mathbf{C}$ of $\mathbf{S} = (\rho_{ij})$.

2. Generate r standard normal variables $\mathbf{Y} = (Y_1, \ldots, Y_r)$.

3. Set $\mathbf{Z} = \mathbf{YC}$.

4. Set $u_i = \Phi(Z_i)$ for $i = 1, \ldots, r$.

5. Set $X_i = F_i^{-1}(u_i)$.

Outputs: The vectors $(X_1, \ldots, X_r)$ form a sample from a multivariate distribution with prescribed correlation structure and marginals F_i.

The Normal Copula method works because of the following theorem from Wang [37, Theorem 2].

Theorem 2 *Assume that $(Z_1, \ldots, Z_k)$ have a multivariate normal joint probability density function given by*

$$f(z_1, \ldots, z_k) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp(-\mathbf{z}'\Sigma^{-1}\mathbf{z}/2), \tag{4.21}$$

$\mathbf{z} = (z_1, \ldots, z_k)$, with correlation coefficients $\Sigma_{ij} = \rho_{ij} = \rho(Z_i, Z_j)$. Let $H(z_1, \ldots, z_k)$ be their joint cumulative distribution function. Then

$$C(u_1, \ldots, u_k) = H(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_k)) \tag{4.22}$$

defines a multivariate uniform cumulative distribution function called the normal copula.

For any set of given marginal cumulative distribution functions $F_1 \ldots, F_k$, the set of variables

$$X_1 = F_1^{-1}(\Phi(Z_1)), \ldots, X_k = F_1^{-1}(\Phi(Z_k)) \tag{4.23}$$

have a joint cumulative function

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = H(\Phi^{-1}(F_x(u_1)), \ldots, \Phi^{-1}(F_k(u_k)) \tag{4.24}$$

with marginal cumulative distribution functions $F_1, \ldots, F_k$. The multivariate variables $(X_1, \ldots, X_k)$ have Kendall's tau

$$\tau(X_i, X_j) = \tau(Z_i, Z_j) = \frac{2}{\pi} \arcsin(\rho_{ij}) \tag{4.25}$$

and Spearman's rank correlation coefficients

$$rkCorr(X_i, X_j) = rkCorr(Z_i, Z_j) = \frac{6}{\pi} \arcsin(\rho_{ij}/2) \tag{4.26}$$

In the normal copula method we simulate from H and then invert using (4.23). In the IC method with normal scores we produce a sample from H such that $\Phi(z_i)$ are equally spaced between zero and one and then, rather than invert the distribution functions, we make the jth order statistic from the input sample correspond to $\Phi(z) = j/(n+1)$ where the input has n observations. Because the jth order statistic of a sample of n observations from a distribution F approximates $F^{-1}(j/(n+1))$ we see the normal copula and IC methods are doing essentially the same thing.

While the normal copula method and the IC method are confusingly similar there are some important differences to bear in mind. Comparing and contrasting the two methods should help clarify how the two algorithms are different.

1. Theorem 2 shows the normal copula method corresponds to the IC method when the latter is computed using normal scores and the Choleski trick.

2. The IC method works on a given sample of marginal distributions. The normal copula method generates the sample by inverting the distribution function of each marginal as part of the simulation process.

3. Though the use of scores the IC method relies on a stratified sample of normal variables. The normal copula method could use a similar method, or it could sample randomly from the base normals. Conversely a sample could be used in the IC method.

4. Only the IC method has an adjustment to ensure that the reference multivariate distribution has exactly the required correlation structure.

5. IC method samples have rank correlation exactly equal to a sample from a reference distribution with the correct linear correlation. Normal copula samples have approximately correct linear and rank correlations.

6. An IC method sample must be taken in its entirety to be used correctly. The number of output points is fixed by the number of input points, and the sample is computed in its entirety in one step. Some IC tools (@Risk, SCARE) produce output which is in a particular order. Thus, if you sample the nth observation from multiple simulations, or take the first n samples, you will not get a random sample from the desired distribution. However, if you select random rows from multiple simulations (or, equivalently, if you randomly permute the rows output prior to selecting the nth) then you will obtain the desired random sample. It is important to be aware of these issues before using canned software routines.

7. The normal copula method produces simulations one at a time, and at each iteration the resulting sample is a sample from the required multivariate distribution. That is, output from the algorithm can be partitioned and used in pieces.

In summary remember these differences can have material practical consequences and it is important not to misuse IC method samples.

Chapter 5

EXAMPLES

We now give an extended example which applies the IC method and some of the other methods introduced in Chapter 3. The example will compute the bivariate distribution of retained and ceded losses, where the reinsurance is an excess of loss cover. Such a bivariate distribution would be useful for a ceding company trying to determine its distribution of net underwriting results if the reinsurance included a variable feature such as a swing rate, sliding scale commission, annual aggregate deductible or profit commission.

The example will apply the following methods and techniques:

M1. Computing aggregate distributions using FFTs.

M2. Compare aggregate distributions computed using FFTs (essentially exact) with method of moments shifted lognormal and shifted gamma approximations.

M3. Computing the bivariate distribution of ceded and retained losses using two dimensional FFTs.

M4. Computing the bivariate distribution of ceded and retained losses using the IC method.

M5. Compare the FFT and IC method.

M6. Illustrate the effect on the IC method of using a t reference distribution.

The first two examples compute univariate marginal distributions, the fundamental inputs to the IC method. The next five examples compute multivariate distributions in various ways.

The reader should take away two lessons from these examples. First, the FFT method works incredibly well for small claim counts and thin tailed severity distributions. In particular, any severity distribution with an occurrence limit is thin tailed. Second, the shifted gamma and shifted lognormal approximations to an aggregate distribution are exceedingly good in all but the most extreme cases. Extreme cases include a very small claim count (say less than five) or a thick tailed severity distribution.

5.1 Example Parameters

The input parameters for the example are as follows. Severity is modeled using a lognormal variable with $\mu = 9.0$ and $\sigma = 2.0$. Underlying policies have a \$1M policy limit. The excess of loss reinsurance attaches at \$200,000 and has a limit of \$800,000; thus it pays all losses between \$200,000 and the policy limit, ground-up. The ground-up expected loss is \$25M.

The nth moments of the layer y excess a of for losses with density f, viz.

$$\mathrm{E}[\min(y, \max(X-a, 0))^n] = \int_a^{a+y} (x-a)^n f(x)dx + y^n \Pr(X > a+y), \quad (5.1)$$

can be computed using

$$\int_a^{a+y} (x-a)^n f(x)dx = \sum_{k=0}^{n} \binom{n}{k} (-a)^k \int_a^{a+y} x^{n-k} f(x)dx \quad (5.2)$$

reducing the problem to computing simpler integrals. For the lognormal, the integral $\lambda(n) := \int_a^{a+y} x^n f(x)dx$ equals

$$\lambda(n) = \exp(n\mu + n^2\sigma^2/2)\Phi((\log(a+y) - \mu - n\sigma^2)/\sigma) \quad (5.3)$$

if $a = 0$ and

$$\lambda(n) = \exp(n\mu+n^2\sigma^2/2)(\Phi((\log(a+y)-\mu-n\sigma^2)/\sigma)-\Phi((\log(a)-\mu-n\sigma^2)/\sigma)) \tag{5.4}$$

for $a > 0$. Then by the above formula for the lognormal we have

$$\mathrm{E}[\min(y, \max(X-a, 0))^n] = y^n(1-\Phi((\log(a+y)-\mu)/\sigma))+\sum_{k=0}^{k=n}\binom{n}{k}(-a)^k\lambda(n-k). \tag{5.5}$$

Using these formulae we find that the expected ground up loss is \$47,439, the expected retained loss is \$31,591 and the expected ceded loss is \$15,848. The probability of attaching the reinsurance is 0.054463 and so the excess severity, conditional on an excess claim, is \$290,985.

The expected claim count is $n = 526.00 = 25000000/47439$. We will model claims using a negative binomial with contagion parameter $c = 0.0625$ which by the discussion in Section 3.3 corresponds to an asymptotic CV of 25% for the aggregate distribution. The parameters of the negative binomial are $r = 1/c = 16$ and $p = 1/(1 + cn) = 0.0295$, using the Bowers et al. [4] parameterization. The moments of the negative binomial are

$$\begin{aligned} E(N) &= (1-p)r/p \\ E(N^2) &= (p-1)r((p-1)r-1)/p^2 \\ E(N^3) &= (1-p)r((p-1)r((p-1)r-3)-p+2)/p^3 \end{aligned}$$

(computed using symbolic differentiation of the moment generating function using a computer algebra program).

Now we can compute the moments of the gross aggregate distribution using (3.14) and (3.16). Writing $A = X_1 + \cdots + X_N$ the results are

$$\begin{aligned} \mathrm{E}(A) &= \mathrm{E}(N)\mathrm{E}(X) \\ \mathrm{E}(A^2) &= \mathrm{E}(N)\mathrm{E}(X^2) + \mathrm{E}(X)^2\mathrm{E}(N^2) - \mathrm{E}(N)\mathrm{E}(X)^2 \\ \mathrm{E}(A^3) &= \mathrm{E}(N)\mathrm{E}(X^3) + \mathrm{E}(X)^3\mathrm{E}(N^3) + 3\mathrm{E}(X)\mathrm{E}(N^2)\mathrm{E}(X^2) - \\ &\quad 3\mathrm{E}(N)\mathrm{E}(X)\mathrm{E}(X^2) - 3\mathrm{E}(X)^3\mathrm{E}(N^2) + 2\mathrm{E}(N)\mathrm{E}(X)^3. \end{aligned}$$

From these expressions we can compute the variance, standard deviation, CV and skewness of frequency, severity and aggregate losses using the definitions at the beginning of Chapter 3. The results are shown in Table 5.1.

Table 5.1: Frequency, Severity and Aggregate Distribution Statistics

Severity	**Ground-Up**	**Retained**	**Ceded**
E(X)	47,439.0	31,591.0	290,985.3
CV(X)	2.7217	1.6745	0.9513
skew(X)	5.2380	2.2351	0.8375
Frequency			
E(X)	527.0	527.0	28.7
CV(X)	0.2538	0.2538	0.3120
skew(X)	0.5001	0.5001	0.5123
Aggregate			
E(X)	25,000,000.0	16,648,209.8	8,351,790.2
CV(X)	0.2801	0.2640	0.3590
skew(X)	0.5128	0.5018	0.5543

5.2 Univariate Methods—Computing Marginal Distributions

5.2.1 Fast Fourier Transform Methods

To compute the aggregate distribution using Fast Fourier Transforms (FFT) we first have to "bucket" the severity distributions. We will use 4,096 buckets (the number must be a power of two for the FFT to work at peak efficiency) and a bucket size $b = 12,500$. The largest loss that we can produce is therefore \$51.1875M which will be adequate for our example. The easiest way to bucket the severity is to compute the cumulative distribution function F at $b/2, 3b/2, \ldots$ and then take differences. The coefficients of bucketed distribution must sum to one. The distribution for ceded losses is actually the conditional distribution given an excess loss, $F(x)/(1 - F(a))$ where a is the attachment and F is the ground up severity distribution. The first few terms in the bucketed severities are shown in Table 5.2

There are slight errors introduced when you bucket the severity distribution, particularly for the retained losses. The mean of the retained severity is 1.9% lower than the actual; the CV is 2.8% higher and the skewness is 1.5% lower. The

Table 5.2: Bucketed Severity Distributions

Bucket Start	Bucket Mid-Point	Ground-Up	Retained	Ceded
0	6,250	0.448350	0.448350	0.030800
12,500	18,750	0.214215	0.214215	0.056797
25,000	31,250	0.087561	0.087561	0.051170
37,500	43,750	0.050294	0.050294	0.046328
50,000	56,250	0.033252	0.033252	0.042130
62,500	68,750	0.023819	0.023819	0.038467
75,000	81,250	0.017980	0.017980	0.035252
87,500	93,750	0.014089	0.014089	0.032415
100,000	106,250	0.011353	0.011353	0.029899
112,500	118,750	0.009351	0.009351	0.027658
125,000	131,250	0.007839	0.007839	0.025653
137,500	143,750	0.006667	0.006667	0.023853
150,000	156,250	0.005740	0.005740	0.022230
162,500	168,750	0.004993	0.004993	0.020762
175,000	181,250	0.004382	0.004382	0.019431
187,500	193,750	0.003876	0.003876	0.018219
200,000	206,250	0.003452	0.056238	0.017114
212,500	218,750	0.003093	0.000000	0.016103
225,000	231,250	0.002787	0.000000	0.015176
237,500	243,750	0.002523	0.000000	0.014323
250,000	256,250	0.002295	0.000000	0.013538
262,500	268,750	0.002095	0.000000	0.012814
275,000	281,250	0.001920	0.000000	0.012143
287,500	293,750	0.001765	0.000000	0.011522
300,000	306,250	0.001628	0.000000	0.010946
etc.				

Table 5.3: Shifted Gamma Approximations to the Aggregate Distributions

Parameter	Ground-Up	Retained	Ceded
s	-2.341E+06	-8.907E+05	-2.468E+06
α	15.19659	15.88163	13.02156
β	1.7780E+06	1.0842E+06	8.3084E+05

excess severity is virtually exact—because the bucket size is small relative to the features of the distribution. The ground-up severity is in between. The smaller the bucket size the lower these discretization errors will be, but on the other hand the less "space" available for the aggregate distribution. Selecting a bucket size which is an exact divisor of the limit will greatly help improve the accuracy of the discretized severity distribution. To determine if your bucket size is appropriate look at the moments of the FFT aggregate relative to the exact moments and plot a graph of the output density. It is usually obvious when the method has failed.

Next we have to take the Fast Fourier Transform of the three 4096×1 severity vectors. We will assume the reader has a computer routine available which will compute FFTs—see Appendix A for one freely available implementation. Then you apply the moment generating function of the frequency distribution (see Table 3.2) row-by-row to the transformed severity. Note that you will be working with complex numbers. Finally you apply the inverse FFT to get a vector of real numbers. Because of the form of the input you are guaranteed that the output will be real and will sum to 1.

5.2.2 Method of Moments and the Shifted Gamma and Lognormal Distributions

In Section 3.5 we introduced the shifted gamma and lognormal distributions and gave explicit expressions for their method-of-moments parameters in terms of mean, CV and skewness. In our example the gross, retained and ceded fits are shown in Table 5.3 for the shifted gamma, 5.4 for the shifted lognormal, and 5.5 for the lognormal.

Table 5.4: Shifted Lognormal Approximations to the Aggregate Distributions

Parameter	Ground-Up	Retained	Ceded
s	-1.636E+07	-9.872E+06	-8.057E+06
μ	17.52370	17.07988	16.59690
σ	0.16811	0.16463	0.18122

Table 5.5: Lognormal Fits to the Aggregate Distributions

Parameter	Ground-Up	Retained	Ceded
μ	16.98348	16.57454	15.87726
σ	0.27554	0.26015	0.34819

Figure 5.1 shows a comparison of the shifted gamma fits (denoted with an asterisk in the legend) with the FFTs. For each of the total, ground-up loss, retained loss and ceded or excess loss the fits appear essentially perfect. On a log-scale, Figure 5.2, we see that the fits are again essentially perfect except for disagreement for small losses. However, the disagreement actually shows an error in the FFT; probabilities for losses greater than the largest bucket size (approximately $50M) wrap around in the FFT and re-appear as small losses, thus the FFT picture is actually inaccurate. The wrapping phenomenon is an example of aliasing; it is the same effect that causes wagon wheels to appear to rotate backwards in Western movies. See Hamming [16] for more details. The shifted gamma approximation is recommended in Bowers et al. [4].

Figure 5.3 shows the shifted lognormal fit. Although not quite as good as the shifted gamma, the fit is still very close. A log scale (not shown) would show that the shifted lognormal is somewhat thicker in the extreme tail. The fact that the shifted gamma does a better job in the tail should not be a surprise since the negative binomial uses a gamma mixing distribution.

Finally, Figure 5.4 shows a comparison of the FFTs with a regular two parameter lognormal. The lognormal is too skewed (peaks too soon) and does not match the true shape of the aggregate well. Using a shifted gamma or shifted lognormal distribution gives a much more satisfactory fit to the true aggregate for very little extra work.

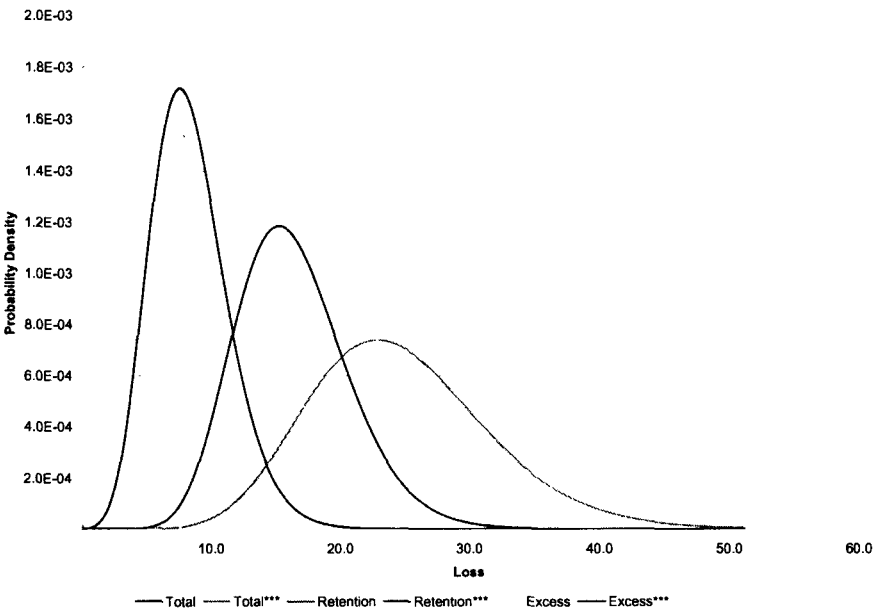


Figure 5.1: FFT vs. shifted gamma approximations for total, retained and ceded losses, illustrating that the gamma is an almost perfect fit.

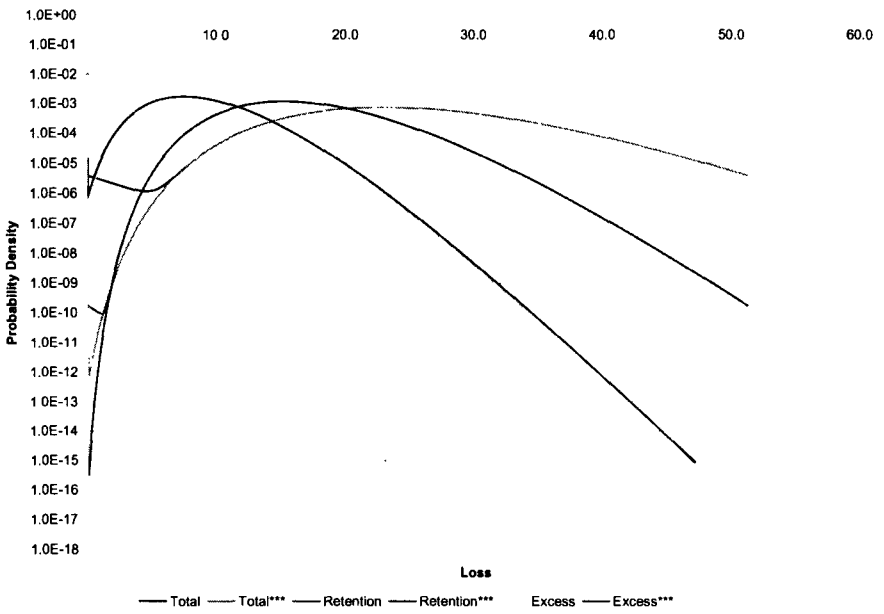


Figure 5.2: FFT vs. shifted gamma approximations for total, retained and ceded losses on a log density scale.

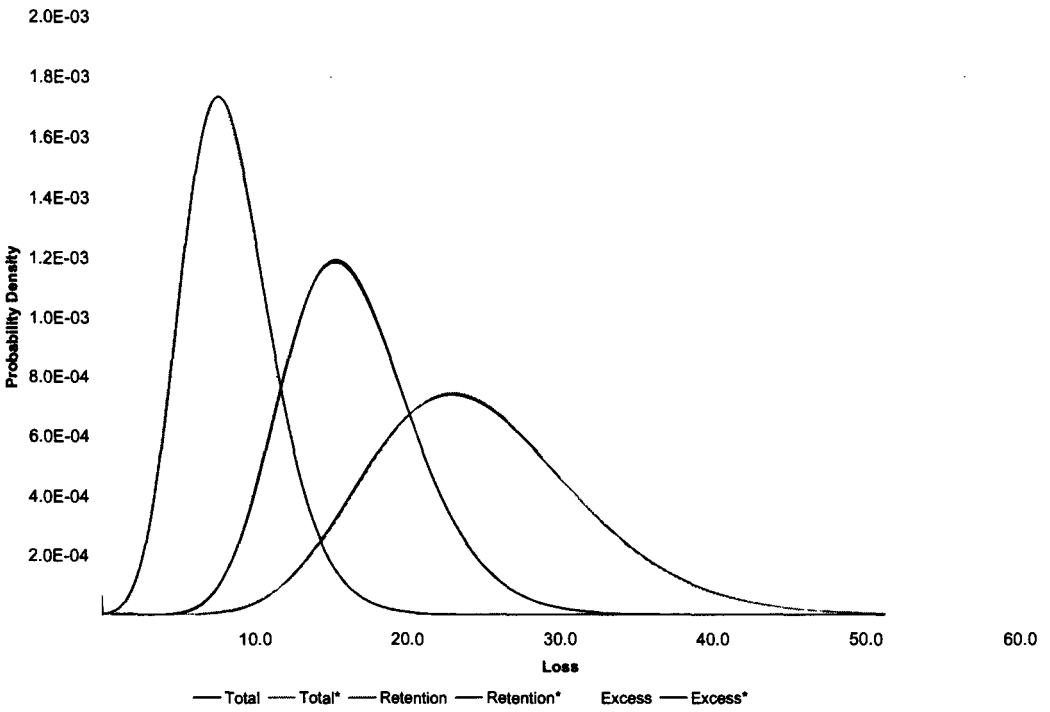


Figure 5.3: FFT vs. shifted lognormal approximations for total, retained and ceded losses.

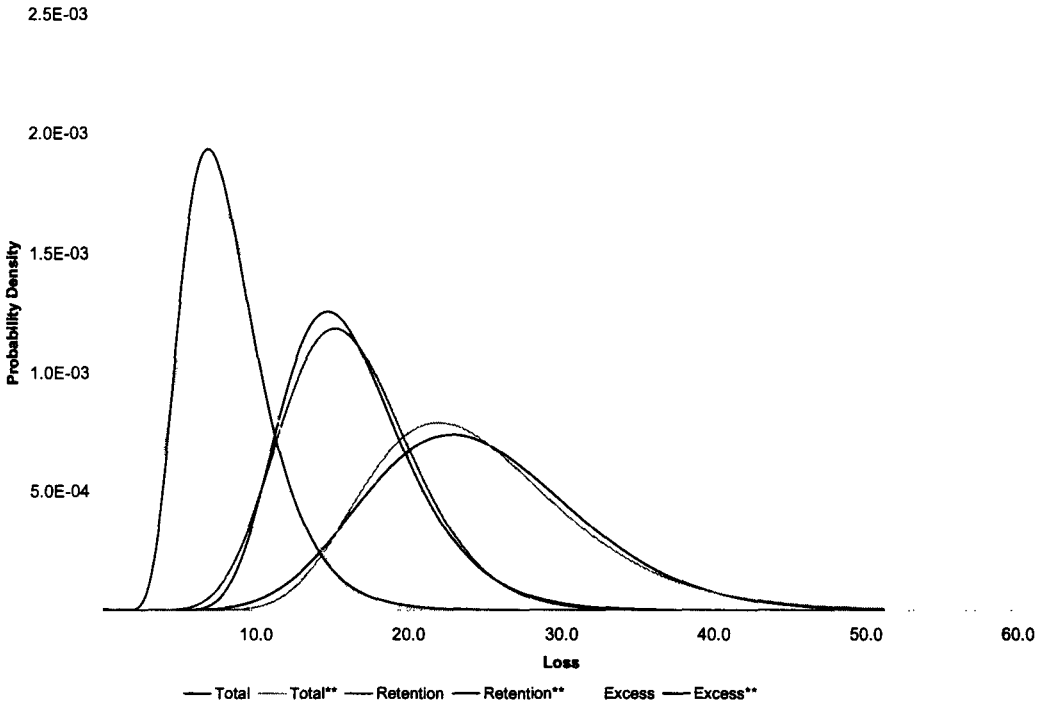


Figure 5.4: FFT vs. lognormal approximations for total, retained and ceded losses, illustrating that the lognormal is a poorer fit than the shifted lognormal.

5.3 Multivariate Methods and the IC Method

Now we have the marginal distributions we need we can apply the IC method to determine the bivariate distribution of retained and ceded losses.

5.3.1 Fast Fourier Transform Methods

In order to have a benchmark for the IC method we begin by computing the exact bivariate distribution of ceded and retained losses using two dimensional FFTs. The two dimensional FFT method is far more limited than the one dimensional version because it is impractical to use discretized distributions larger than 4096×4096—the size we will use here. One is caught by the need for a small bucket size to capture the shape of the ground-up severity and the need for enough buckets to capture the whole aggregate distribution.

The method for two dimensional FFTs is essentially the same as for one dimension: compute a discretized version of the input severity distribution, which will be a matrix rather than a column vector, apply the FFT, apply the MGF of the frequency distribution term-by-term, and then apply the inverse FFT. The resulting distribution is shown in Figure 5.5.

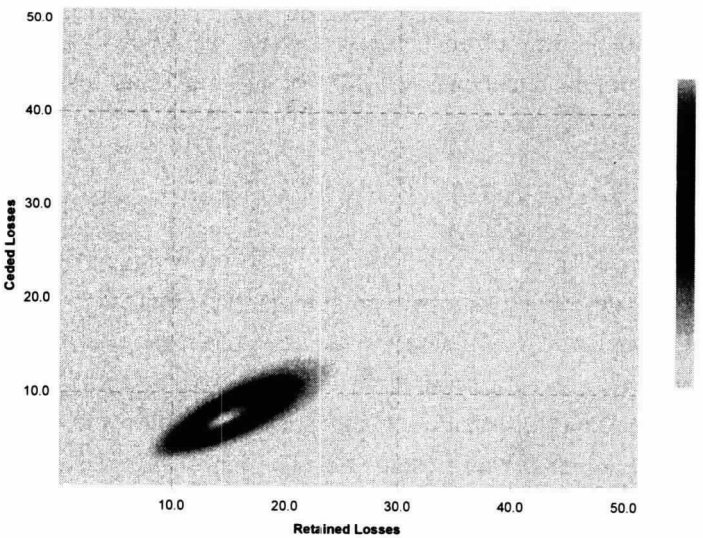


Figure 5.5: Two dimensional FFT estimate of bivariate distribution of ceded and retained losses.

5.3.2 IC Method

Next we apply the IC method to the marginal retained and ceded distribution computed in the previous section. Individual ceded and retained losses are a good example of comonotonic variables, since they are both increasing functions of gross losses. Aggregate ceded and retained losses will not generally be comonotonic. To apply IC we need the correlation coefficient between ceded and retained losses which can be computed using (3.47). The only missing value from that equation is the covariance between retained severity R and ceded severity C. However, because of the simple form of the bivariate severity, viz. ceded losses are zero until gross losses hit the retention $a = \$200,000$ and then ceded losses increase, the covariance is easy to compute:

$$\mathrm{Cov}(R, C) = \mathrm{E}(RC) - \mathrm{E}(R)\mathrm{E}(C) = \mathrm{E}(C)(a - \mathrm{E}(R)). \tag{5.6}$$

Substituting into (3.47) gives a correlation of 0.786 between aggregate retained losses and aggregate ceded losses. We can now apply the Iman Conover method. Here we used samples of 10,000 observations from the univariate distributions of ceded and retained losses. The result of the IC method will be a 10000×2 matrix sample from the bivariate distribution. In order to visualize the result we produced a bivariate histogram, as shown in Figure 5.6. The approximation is very similar to the previous "exact" FFT contour plot, as you can see if you overlay the two plots.

The IC method underlying Figure 5.6 used a normal copula reference distribution. As we have already discussed there are many other possible reference distributions we could chose to use. Figure 5.7 shows the resulting two dimensional histogram if we use a t-copula with two degrees of freedom, which is a very extreme choice. Just as we saw in Figure 4.3 the result of using a t-copula is to introduce more extreme value dependence and the contours have a pinched look—both in the slope 1 and slope -1 directions.

Clearly the normal copula IC method produces bivariate distribution closer to the FFT actual than the t-copula, which should not be a surprise. There is no

generator of extreme tail correlation in our example. However, in other modeling situations, such as modeling the bivariate movement of stock prices or foreign exchange movements, there may be empirical evidence of strong tail correlation and a t-copula (or other non-normal) copula approach would be more appropriate.

Finally, Figure 5.8 shows the distribution of the sum of ceded and retained losses using the normal-copula, t-copula, and actual dependence relationships. As expected the normal copula model is closest to the actual. The t-copula sum is too peaked and is more thick tailed than the actual distribution.

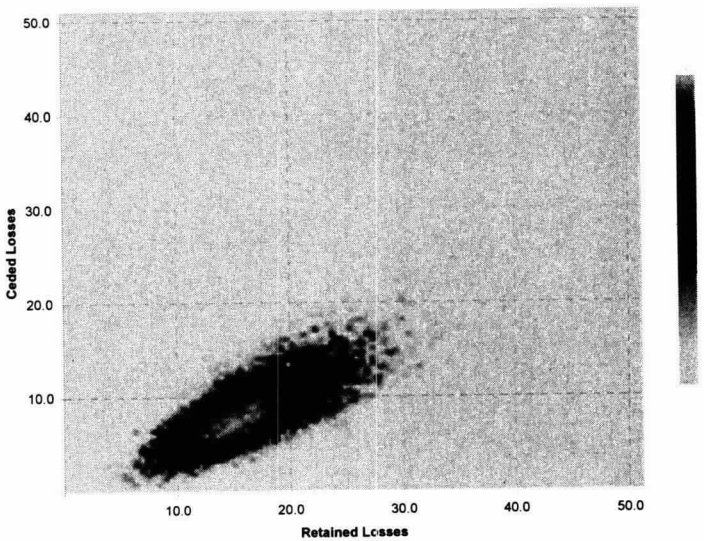


Figure 5.6: Iman-Conover approximation to bivariate distribution of ceded and retained losses.

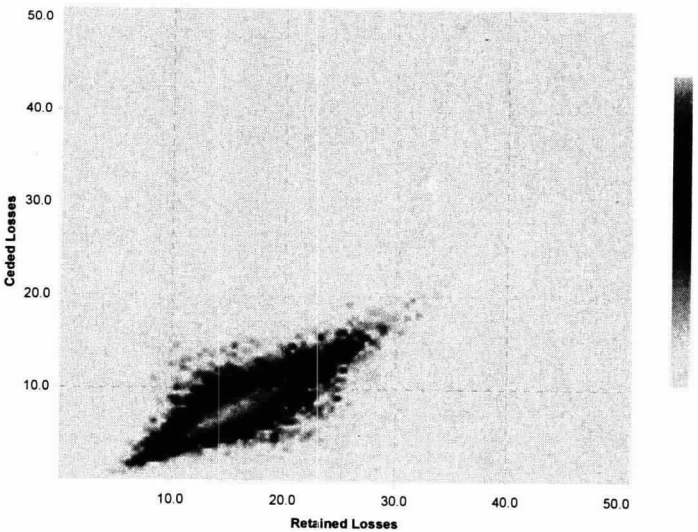


Figure 5.7: Iman-Conover approximation to bivariate distribution of ceded and retained losses using the t-copula as a reference distribution.

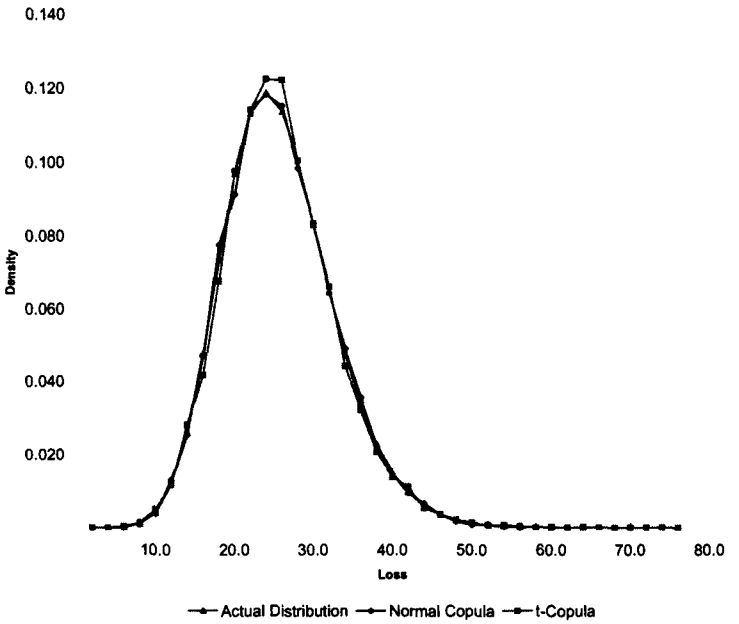


Figure 5.8: Distribution of total losses (ceded + retained) under normal copula, t-copula and actual.

Chapter 6

THEORETICAL UNDERPINNINGS OF THE IMAN-CONOVER METHOD

The theoretical foundations of the Iman-Conover method are elegantly justified by Vitale's Theorem [35]. We will state Vitale's theorem, explain its relationship to the IC method, and sketch the proof. The result should give a level of comfort to practitioners using a simulation approach to modeling multivariate distributions. It is not necessary to follow the details laid out here in order to understand and use the IC method, so the uninterested reader can skip the rest of the section. The presentation we give follows Vitale's original paper [35] closely.

Functional dependence and independence between two random variables are clearly opposite ends of the dependence spectrum. It is therefore surprising that Vitale's Theorem says that any bivariate distribution (U, V) can be approximated arbitrarily closely by a functionally dependent pair (U, TU) for a suitable transformation T.

In order to explain the set up of Vitale's theorem we need to introduce some notation. Let n be a power of 2. An interval of the form $((j-1)/n, j/n)$ for some $n \geq 1$ and $1 \leq j \leq n$ is called a dyadic interval of rank n. An invertible (Borel)

measure-preserving map which maps by translation on each dyadic interval of rank n is called a permutation of rank n. Such a T just permutes the dyadic intervals, so there is a natural correspondence between permutations of n elements and transformations T. If the permutation of dyadic intervals has a single cycle (has order n in the symmetric group) then T is called a cyclic permutation.

Theorem 3 *(Vitale) Let U and V be uniformly distributed variables. There is a sequence of cyclic permutations $T_1, T_2, \ldots$ such that $(U, T_n U)$ converges in distribution to (U, V) as $n \to \infty$.*

Recall convergence in distribution means that the distribution function of $(U, T_n U)$ tends to that of (U, V) at all points of continuity as $n \to \infty$.

The proof of Vitale's theorem is quite instructive and so we give a detailed sketch.

The proof is in two parts. The first constructs a sequence of arbitrary permutations T_n with the desired property. The second part shows it can be approximated with cyclic permutations. We skip the second refinement.

Divide the square $[0,1] \times [0,1]$ into sub-squares. We will find a permutation T such that the distributions of (U, V) and (U, TU) coincide on sub-squares. Reducing the size of the sub-squares will prove the result.

Fix n, a power of two. Let $I_j = ((j-1)/n, j/n)$, $j = 1, \ldots, n$. We will find an invertible permutation T such that

$$\Pr(U \in I_j, TU \in I_k) = \Pr(U \in I_j, V \in I_k) := p_{jk} \tag{6.1}$$

for $j, k = 1, \ldots, n$. Define

$$I_{j1} = ((j-1)/n, (j-1)/n + p_{j1}) \tag{6.2}$$

$$I_{j2} = ((j-1)/n + p_{j1}, (j-1)/n + p_{j1} + p_{j2}) \tag{6.3}$$

$$\cdots \tag{6.4}$$

$$I_{jn} = ((j-1)/n + p_{j1} + \cdots + p_{j,n-1}, j/n) \tag{6.5}$$

and

$$\tilde{I}_{j1} = ((j-1)/n, (j-1)/n + p_{1j}) \tag{6.6}$$

$$\tilde{I}_{j2} = ((j-1)/n + p_{1j}, (j-1)/n + p_{1j} + p_{2j}) \tag{6.7}$$

$$\cdots \tag{6.8}$$

$$\tilde{I}_{jn} = ((j-1)/n + p_{1j} + \cdots + p_{n-1,j}, j/n). \tag{6.9}$$

By construction the measure of I_{jk} equals the measure of $\tilde{I}_{kj}$. The invertible map T which sends each I_{jk} to $\tilde{I}_{kj}$ by translation is the map we need because

$$\Pr(U \in I_j, T(U) \in I_k) = \Pr(U \in I_j, U \in T^{-1}(I_k)) \tag{6.10}$$

$$= \Pr(U \in I_j \cap T^{-1}(\bigcup_l \tilde{I}_{kl})) \tag{6.11}$$

$$= \Pr(U \in \bigcup_l I_j \cap I_{lk}) \tag{6.12}$$

$$= \Pr(U \in I_{jk}) \tag{6.13}$$

$$= p_{jk}, \tag{6.14}$$

since the only I_{lk} which intersects I_j is I_{jk} by construction, and U is uniform. The transformation T is illustrated schematically in Table 6.1 for $n = 3$. The fact 3 is not a power of 2 does not invalidate the schematic!

If each p_{jk} is a dyadic rational then T is a permutation of the interval. If not then we approximate and use some more heavy duty results (a 1946 theorem of Birkhoff on representation by convex combinations of permutation matrices) to complete the proof.

Vitale's theorem can be extended to non-uniform distributions.

Corollary 1 *(Vitale) Let U and V be arbitrary random variables. There is a sequence of functions $S_1, S_2, \ldots$ such that $(U, S_n U)$ converges in distribution to (U, V) as $n \to \infty$.*

Let F be the distribution function of U and G for V. Then $F(U)$ and $G(V)$ are uniformly distributed. Apply Vitale's theorem to get a sequence of functions T_n. Then $S_n = G^{-1}T_nF$ is the required transformation.

Table 6.1: Schematic of the Vitale transformation for $n = 3$

I_3	$\tilde{I}_{33}$								p_{33}
	$\tilde{I}_{32}$					p_{23}			
	$\tilde{I}_{31}$			p_{13}					
I_2	$\tilde{I}_{23}$							p_{32}	
	$\tilde{I}_{22}$				p_{22}				
	$\tilde{I}_{21}$		p_{12}						
I_1	$\tilde{I}_{13}$						p_{31}		
	$\tilde{I}_{12}$				p_{21}				
	$\tilde{I}_{11}$	p_{11}							
		I_{11}	I_{12}	I_{13}	I_{21}	I_{22}	I_{13}	I_{31}	I_{32}	I_{33}
		I_1			I_2			I_3		

Appendix A

SOFTWARE IMPLEMENTATIONS

Having laid out the IC method and given some explicit examples, we now discuss implementation issues. We will follow the Detailed Algorithm laid out in Section 4.2.

A.1 General Implementation and Design Considerations

A good general rule in writing software is to ensure that the steps which execute most frequently are coded as efficiently as possible. Cutting 50% from the execution time of a step which runs once and takes 1 second will have a barely perceptible impact. Cutting 50% from a step which takes 10 msecs, but executes 10,000 times will have a material and perceptible impact. See Hennessy and Patterson [18] for more discussion.

Matrix and linear algebra operations can be hard to code efficiently because of the design of modern computer chips and the strains matrix operations put on memory management. Modern CPUs have on-chip cache memory, which operates very quickly. Processors are "smart" enough to partially anticipate future memory calls and ensure the relevant locations have been pre-loaded into cache.

For example, arrays are usually stored in contiguous blocks of memory, and if you ask for $x[i]$, it is likely you will ask for $x[i+1]$. Processors will pull in a block of memory each side of $x[i]$ to speed operation. If the required memory is not in cache the CPU has a "cache-miss". These are very costly and result in a lot of lost processing time. Certain operations used extensively in the IC algorithm tend to generate lots of cache-misses: matrix multiplication being the worst (you pull in a row and a column; only one of these will be contiguous in memory). There are ways around these problems, but they are not ways you would want to navigate yourself! Fortunately, professional software designers spend a lot of effort to code matrix operations as efficiently as possible. Many processor benchmarks use matrix operations, so chip manufacturers have a vested interest here.

The Lapack [1] package is an example of a very efficiently coded set of matrix algebra operations. It is build on BLAS, Basic Linear Algebra Subprograms, which implements fundamental operations like matrix multiplication. Lapack implementations are available for most platforms, including Windows. See http://www.netlib.org/lapack for a non-commercial implementation. See http://-www.intel.com/software/products/mkl for a version optimized for Intel processors. It will automatically multi-thread operations if there are two or more CPUs available.

The implementation in Section 4.2 describes the appropriate Lapack functions for all the matrix operations, such as Choleski decomposition and solving a system of linear equations. I cannot claim that the implementation is optimal, but it is very fast.

A.2 SCARE

SCARE, a Simulating, Correlated Aggregation and Risk Engine, is a COM object (DLL) program which can be used from Excel/VBA. It implements the IC method, some other useful copula functions and bivariate normal and t distributions. It can be downloaded from www.mynl.com/wp. It was originally designed and implemented for Scor Re US, who have kindly given permission for it to

be made available to CAS members as part of the Working Party on Correlation effort.

Before programming with SCARE you need to install it and then reference it from your VBA project. Within the VBA editor, click Tools, References to bring up the References dialog. Make sure the SCARE library is selected, as shown in Figure A.1. In your application the location should show as C:/Program Files/Scare/Bin. Libraries are listed alphabetically, except those in use by an open project, which appear at the top.

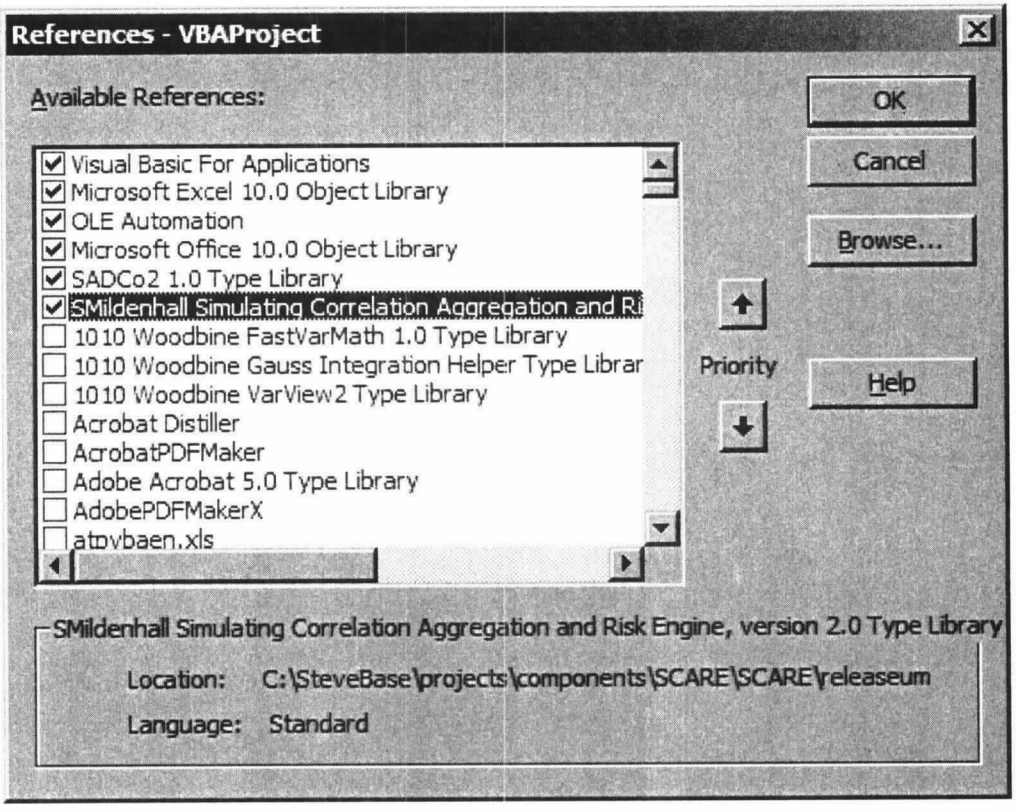


Figure A.1: Adding a reference to the SCARE component in Excel VBA.

A.3 SCARE Functions

All of the functions exposed by the SCARE DLL are described in detail in the User's Guide [30]. Here we give a quick overview of the key functions aimed at the Excel user. The SCARE DLL functions can call be referenced from VBA but they cannot be called directly from an Excel spreadsheet.

A.4 SCARE Wrapper Functions

A wrapper function "wraps" a COM function call so that it can be used directly from a worksheet[1]. For example, here is how we wrap the sort-by-aggregate function. The scSortByAgg function takes four arguments: the input multivariate array, an indicator to sort in ascending or descending order, a holder for the sorted output with the aggregate appended in the last column and a holder for the sorted output without the aggregate appended. Here is the VBA code.

```
Function scSortByAgg(v, ad As Long)
    Dim xx as new Shuffler
    Dim w, agg
    ' ad=1 ascending order, ad=-1 descending order
    xx.SortByAgg v, ad, w, agg
    scSortByAgg = w
End Function
```

The argument v is the multivariate sample and the argument ad is +/-1 for ascending or descending order respectively. Within the function, new variables w and agg are defined to hold the answers, and xx is defined as a new Shuffler object to access the member function. The SortByAgg method of Shuffler is then called. Finally scSortByAgg=w sets the answer to be returned to Excel.

In a spreadsheet, the function would be called as =scSortByAgg(A1:D100,1) input as an array function in a range large enough to hold the answer. Array functions are entered using control+shift+enter, rather than just enter. They appear in the spreadsheet as {=scSortByAgg(A1:D100,1)}.

[1]Excel 2002 will automatically create wrappers for all functions using Tools, Add-Ins and selecting the Automation button.

The implementation in the Wrappers module of SCARE.xla uses a single variable xx which is shared between all the functions in the work book. It is defined as a private variable at the beginning of the workbook.

Several variables below are defined as "variants". Variants are a useful hold-all variable type in VB/VBA. Almost all method output variables are Variants. They can hold a single number, a string, an array of numbers, or a mixed array, or even a reference to an entire open Excel application. Code such as v = Range("a1:b10").value will set v equal to a 2 x 10 variant array. Depending on the contents of the range it could be an array of doubles or an array of variants. Code like set v = Range("a1:b10") sets v equal to a reference to the range object. It is then possible to write v.ClearContents to clear Range("a1:b10") or v.Value = 10 to set all the cells A1:B10 equal to the value 10. Variants need to be used with care. In some situations they are convenient, fast and efficient—but in others they are convenient, slow and inefficient. Their use in SCARE is confined to the former.

The main various functions in SCARE.xla are as follows.

```
Function scSortByAgg(v, ad As Long)
```

Sums the input $n \times r$ multivariate density over columns to get an $n \times 1$ aggregate. Sorts the whole input array by the aggregate. Use ad=1 to sort in ascending order, and ad=-1 for descending order.

```
Public Function scCholeski(x As Variant) As Variant
```

Returns the Choleski decomposition of the input $r \times r$ matrix x. Note that the C++ object only populates the upper half of the matrix. The VBA method "tidies-up" that return by zero filling the lower portion.

```
Function scCorr(v As Variant) As Variant
```

Computes the mean by column, covariance matrix and correlation matrix of the input $n \times r$ multivariate density v. Only the correlation matrix is returned to Excel, but it would be easy for the user to alter to return the means vector or the covariance matrix.

```
Public Function scIsCovMat(x As Variant) As Boolean
```

Tests input matrix X to determine if it is positive definite. Returns true or false.

```
Function scNormDist(x As Double) As Double
```

Computes $\Phi(x)$ the standard normal distribution function evaluated at x. It is more accurate, and from VBA quicker, than the worksheet function NormDist.

```
Function scNormInv(x As Double) As Double
```

Computes the inverse standard normal probability distribution function at probability level x. Also quicker and more accurate than the built in functions.

```
Function scTDist(nu As Long, x As Double) As Double
```

Computes the t-distribution function with ν degrees of freedom at x.

```
Function scTInv(nu As Long, x As Double) As Double
```

Computes the inverse to the t distribution function with ν degrees of freedom at probability level x.

```
Function scBVN(h As Double, k As Double, rho As Double) As Double
```

Computes the probability $\Pr(X < h, Y < k)$ where (X, Y) have a bivariate normal distribution with standard normal marginals and correlation coefficient ρ.

```
Function scBVT(nu As Long, h As Double, k As Double, rho As Double)
As Double
```

Computes the bivariate t distribution function $\Pr(X < h, Y < k)$ where (X, Y) have a bivariate t distribution with $n\nu$ degrees of freedom and correlation ρ.

```
Function scQuickShuffle(rgIn, corrIn) As Variant
```

Returns the input $n \times r$ range rgIn re-ordered to have correlation approximately equal to the $r \times r$ correlation matrix corrIn.

```
Function scQuickShuffleParam(spec, n As Long, covMat) As Variant
```

As scQuickShuffle, except the input values are parameters for shifted lognormal variables. The spec input is a $n \times r$ array where there are n input lines of business and nine columns. The meaning of the nine columns is as follows:

1. Not used
2. ± 1, where $+1$ is used to represent losses and -1 liabilities
3. μ
4. σ
5. s, the shift parameters
6. 0 or 1 indicator where 1 means there is layer and attachment information for the current row.
7. Layer value
8. Attachment value; the sample is from a shifted lognormal with parameters s, μ and σ, conditional on losses being greater than the attachment. The attachment is subtracted and losses are limited by layer value input.
9. Not used.

Appendix B

VBA CODE FOR INDEXING

```
Private Sub indexx(n As Long, arr, colNo As Long, indx() As Long)

'  Indexes an array arr[1..n], i.e., outputs the array indx[1..n] such
'  that arr[indx[j]] is in ascending order for j = 1, 2, . . . ,N. The
'  input quantities n and arr are not changed.  Translated from [31].

    Const m As Long = 7
    Const NSTACK As Long = 50

    Dim i As Long, indxt As Long, ir As Long, itemp As Long, j As Long
    Dim k As Long, l As Long
    Dim jstack As Long, istack(1 To NSTACK) As Long
    Dim a As Double

    ir = n
    l = 1

    For j = 1 To n
        indx(j) = j
    Next j

    Do While 1
        If (ir - l < m) Then
            For j = l + 1 To ir
                indxt = indx(j)
                a = arr(indxt, colNo)
                For i = j - 1 To l Step -1
                    If (arr(indx(i), colNo) <= a) Then Exit For
                    indx(i + 1) = indx(i)
```

```
            Next i
            indx(i + 1) = indxt
        Next j
        If (jstack = 0) Then Exit Do
        ir = istack(jstack)
        jstack = jstack - 1
        l = istack(jstack)
        jstack = jstack - 1
    Else
        k = (l + ir) / 2
        itemp = indx(k)
        indx(k) = indx(l + 1)
        indx(l + 1) = itemp
        If (arr(indx(l), colNo) > arr(indx(ir), colNo)) Then
            itemp = indx(l)
            indx(l) = indx(ir)
            indx(ir) = itemp
        End If
        If (arr(indx(l + 1), colNo) > arr(indx(ir), colNo)) Then
            itemp = indx(l + 1)
            indx(l + 1) = indx(ir)
            indx(ir) = itemp
        End If
        If (arr(indx(l), colNo) > arr(indx(l + 1), colNo)) Then
            itemp = indx(l)
            indx(l) = indx(l + 1)
            indx(l + 1) = itemp
        End If
        i = l + 1
        j = ir
        indxt = indx(l + 1)
        a = arr(indxt, colNo)
        Do While 1
            Do
                i = i + 1
            Loop While (arr(indx(i), colNo) < a)
            Do
                j = j - 1
            Loop While (arr(indx(j), colNo) > a)
            If (j < i) Then Exit Do
            itemp = indx(i)
            indx(i) = indx(j)
            indx(j) = itemp
        Loop
        indx(l + 1) = indx(j)
```

```
            indx(j) = indxt
            jstack = jstack + 2
            If (jstack > NSTACK) Then MsgBox (''NSTACK too small in indexx.")
            If (ir - i + 1 >= j - l) Then
                istack(jstack) = ir
                istack(jstack - 1) = i
                ir = j - 1
            Else
                istack(jstack) = j - 1
                istack(jstack - 1) = l
                l = i
            End If
        End If
    Loop

End Sub
```

Bibliography

[1] E. Anderson. *LAPACK Users' Guide (www.netlib.org/lapack)*. SIAM, 1999.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, pages 203–28, 1999.

[3] P. Billingsley. *Probability and Measure*. J. Wiley and Sons, second edition, 1986.

[4] N. Bowers, H. Gerber, J. Hickman, D. Jones, and C. Nesbitt. *Actuarial Mathematics*. Society of Actuaries, 1986.

[5] D. R. Clark and C. A. Thayer. A primer on the exponential family of distributions. *Casualty Actuarial Society Spring Forum*, pages 117–148, 2004.

[6] W. J. Conover. *Practical nonparametric statistics*. John Wiley and Sons, third edition, 1999.

[7] C. D. Daykin, T. Pentikainen, and M. Pesonen. *Practical Risk Theory for Actuaries*. Chapman and Hall, 1994.

[8] P. Diaconis. Patterns in eigenvalues: The 70th Josiah Willard Gibbs Lecture. *Bulletin of the American Mathematical Society*, 40(2):155–178, 2003.

[9] P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. *ETHZ Zurich*, 2001.

[10] P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependency in risk managment: Properties and pitfalls. *ETHZ Zurich*, 1998.

[11] K.-T. Fang and Y.-T. Zhang. *Generalized Multivariate Analysis*. Springer-Verlag/Science Press Beijing, 1990.

[12] W. Feller. *An Introduction to Probability Theory and Its Applications, two volumes*. J. Wiley and Sons, second edition, 1971.

[13] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, third edition, 1996.

[14] L. Goodman and W. Kruskal. Measures of association for cross-classifications. iii: Approximate sample theory. *Journal of ASA*, 58:310–364, 1963.

[15] J. Grandell. *Mixed Poisson Processes*. Chapman and Hall, 1977.

[16] R. W. Hamming. *Digital Filters*. Dover, third edition, 1989.

[17] P. Heckman and G. G. Meyers. The calculation of aggregate loss distributions from claim severity and claim count distributions. *Proceedings of the Casualty Actuarial Society*, LXXX:22–61, 1983.

[18] J. L. Hennessy and D. A. Patterson. *Computer Architecture: A Quantitative Approach*. Morgan Kaufmann, second edition, 1996.

[19] T. Herzog. Part 3 study note. *Education and Examination Committee of the Society of Actuaries*, 1984.

[20] D. L. Homer and D. R. Clark. Insurance applications of bivariate distributions. *Proceedings of the Casualty Actuarial Society*, XC:274–307, 2003.

[21] R. Iman and W. Conover. A distribution-free approach to inducing rank correlation amoung input variables. *Commun. Statist. Simula. Computa.*, 11(3):311–334, 1982.

[22] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 1. John Wiley and Sons, second edition, 1994.

[23] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Discrete Multivariate Distributions*. J. Wiley and Sons, 1996.

[24] N. L. Johnson, S. Kotz, and A. W. Kemp. *Univariate Discrete Distributions*. J. Wiley and Sons, second edition, 1992.

[25] S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss Models, From Data to Decisions*. John Wiley and Sons, 1998.

[26] D. Knuth. *The Art of Computer Programming, vol. 2 Seminumerical Algorithms*. Addison-Wesley, 1999.

[27] S. Kotz, T. Kozubowski, and K. Podgorski. *The Laplace Distribution and Generalizations*. Birkhauser, 2001.

[28] T. Mack. The stochastic model underlying the chain ladder method. *ASTIN Bulletin*, Vol. 21(1):22–61, 1991.

[29] C. Menn and S. T. Rachev. Calibrated FFT-based density approximations for α-stable distributions. *Elsevier Science*, 2004.

[30] S. J. Mildenhall. *SCARE: Simulating Correlated Aggregation and Risk Engine User's Guide*. www.mynl.com/wp, 2002.

[31] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipies in C, The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.

[32] D. Sornette, P. Simonetti, and J. V. Andersen. ϕ^q-field theory for portfolio optimization: "fat-tails" and non-linear correlations (working title: Non-linear covariance matrix and portfolio theory for non-gaussian multivariate distributions). *Physics Report*, 335(2):19–92, 2000.

[33] G. G. Venter. Tails of copulas. *Proceedings of the Casualty Actuarial Society*, LXXXIX:68–113, 2002.

[34] R. J. Verrall. A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims resering. *NAAJ*, 8 No. 3:67–89, 2004.

[35] R. Vitale. *On Stochastic Dependence and a Class of Degenerate Distributions*, pages 459–469. In Topics in Statistical Dependence, Edited by H. Block, A. Sampson and T. Savits. Institute of Mathematical Statistics, Hayward CA, 1990.

[36] S. Wang. Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26:71–92, 1995.

[37] S. Wang. Aggregation of correlated risk portfolios: Models and algorithms. *Proceedings of the Casualty Actuarial Society*, LXXXV:848–939, 1998.

[38] S. Wang. A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, 67(1):15–36, 2000.

Biography

Stephen Mildenhall is Executive Vice President at Aon Re Services in Chicago IL. He has an undergraduate degreee in Mathematics from the University of Warwick, UK, and a masters and PhD in Mathematics from the University of Chicago, IL. He is a Fellow of the Casualty Actuarial Society, an Associate of the Society of Actuaries and Member of the American Academy of Actuaries. His research interests include capital allocation and applications of statistics to reserving and ratemakeing problems.

The Report of the Research Working Party on Correlations and Dependencies Among All Risk Sources

Part 2

Aggregating Bivariate Claim Severities With Numerical Fourier Inversion

David L. Homer Member, CAS Research Working Party on Correlations and Dependencies Among All Risk Sources

AGGREGATING BIVARIATE CLAIM SEVERITIES WITH NUMERICAL FOURIER INVERSION

DAVID L. HOMER

Abstract

This chapter will apply continuous Fourier transforms to compute the bivariate aggregate claims distribution arising from a bivariate severity distribution and a univariate claim count distribution.

1. INTRODUCTION

This chapter will apply continuous Fourier transforms to compute the bivariate aggregate claims distribution arising from a bivariate severity distribution and a univariate claim count distribution.

Section 1 provides a general description of univariate aggregate claims methods followed by a general description of bivariate aggregate claims methods.

Section 2 provides a brief summary of the univariate Fourier transform method applied by Heckman and Meyers [3] since this will provide the foundation for the bivariate method presented in section 3. The abbreviation "HM" will be used for "Heckman and Meyers". Section 4 presents examples.

1.1. Univariate Methods

There are several methods described in the actuarial literature for computing the *univariate* aggregate loss distribution arising from a univariate severity distribution and a univariate claim count distribution. These methods include HM's numerical Fourier

inversion [3], discrete Fourier transforms as discussed by Wang [10] and Robertson [8], and Panjer's recursive techniques [7].

Heckman and Meyers' numerical Fourier inversion method uses a severity distribution with claim size intervals of constant density and a possible point mass at the maximum claim size. The claim count model is Binomial, Poisson, or Negative Binomial. This method works best when the expected claim counts are large because the numeric integral computed by this method coverges more quickly when the claim counts are large.

The basic discrete Fourier transform method requires a discrete claim size distribution with claim sizes at equally spaced intervals. It works best when the expected claim counts are small because of computer memory constraints. The interval size must be small enough to accurately represent the claim size distribution while the largest claim size represented must be large enough to capture the aggregate distribution. This generally means a large number of intervals are required and limited computer memory can make computations for large claim counts impractical.

Robertson's method is a clever adaptation of the basic discrete Fourier transform for application with claim size distributions with equally spaced intervals of constant density. This is nearly the same claim size model used by HM, but with a few additional limitations. There is no point mass allowed at the maximal claim size and the intervals of constant density must have uniform width. The claim count model is a finite list of probabilities. This method works best when the expected claim counts are small because of computer memory constraints.

Additional calculations are required to correct the basic discrete Fourier transform for the non-discrete severity density. In practice, the cost of the additional calculations may outweigh the benefit, if any, of using severities with intervals of constant density. However, since Robertson's method is exact it is extremely useful for checking methods like the HM method which has an error term. The testing must be done with examples with a moderate number of expected claims since the HM method works best with a large number and Robertson's method works best with a small number. In this paper we will use a two-dimensional application of Robertson's

method to compute the error of the two-dimensional extension of HM.

The recursive technique uses a discrete severity distribution with uniformly spaced claim sizes. The claim count model includes the Binomial, Poisson and Negative Binomial distributions.[1]. This method works well when the expected claim counts are small for reasons similar to those given for discrete Fourier transform methods.

In the methods described above, a pair of risk collections—each with its own severity and claim count distribution—would be aggregated assuming the collections were independent. Heckman and Meyers also allow a *mixing parameter* that reflects parameter risk in the scale of the aggregate distribution and induces a correlation between collections. Wang [10] and Meyers [6] discuss the univariate aggregation of correlated collections.

1.2. Bivariate Methods

The actuarial literature also describes the computation of *bivariate* aggregate distributions. Homer and Clark [4] describe bivariate examples using two-dimensional discrete Fourier transforms. Sundt [9] extends Panjer recursions to multiple dimensions. Walhin [11] describes an application of two-dimensional Panjer recursions. Like their univariate counterparts, these methods work best when the expected claim counts are small due to computer memory constraints.

This chapter extends the HM method to bivariate aggregate distributions. As with the univariate method, this extension works best when the expected claim counts are large because the numeric integrals computed converge more quickly with large claim counts.

The following sections will provide a brief review of the HM univariate method, develop the bivariate method, and present some examples.

[1]The claim count model for recursion technique includes a larger group of distributions which are the members of the $(a, b, 0)$ or $(a, b, 1)$ classes as described by Klugman et al [3]. The HM method can be modified to use (a, b, x) members.

2. UNIVARIATE NUMERICAL FOURIER INVERSION

2.1. Univariate Collective Risk Model

The collective risk model describes aggregate claims for a collection of risks with a claim count or frequency distribution and a claim size or severity distribution. The individual claims sizes X_k are independent and identically distributed (iid). The individual claim sizes are also independent of the claim count N. The aggregate losses are

$$Z = X_1 + ... + X_N. \tag{2.1}$$

This model may be used to describe the aggregate losses for a single line or book of business.

2.2. Univariate Aggregate Characteristic Function

The aggregate loss distribution is conveniently described through its characteristic function in terms of the characteristic function of the claim size distribution and the probability generating function of the claim count distribution.

Recall that the characteristic function (cf) for a distribution is defined as

$$\phi_X(t) = E(e^{itX}), \tag{2.2}$$

and that the probability generating function (pgf) for a discrete distribution is defined as

$$PGF_N(t) = E(t^N). \tag{2.3}$$

The aggregate loss characteristic function $\phi_Z(t)$ is equal to the composition of the claim count probability generating function $PGF_N(t)$ with the claim size characteristic function $\phi_X(t)$,

$$\begin{aligned} \phi_Z(t) &= E(e^{itZ}) \\ &= E(e^{X_1+...+X_N}) \\ &= E_N(\phi_X(t)^N|N) \\ &= PGF_N(\phi_X(t)). \end{aligned} \tag{2.4}$$

The cdf $F(z)$ of Z can be obtained from $\phi_Z(t)$ when it is continuous

$$F(z) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty |\phi_Z(t)| \frac{\sin(zt - \arg(\phi_Z(t)))}{t} dt. \tag{2.5}$$

Although $\phi(t)$ is complex, Equation 2.5 is real valued; $|\phi|$ is the *modulus* of ϕ and $\arg(\phi)$ is its *argument*. The right hand side of 2.5 yields $F(z) - Pr(Z = z)/2$ at steps when $F(z)$ is not continuous. Given $\phi(t)$, $F(z)$ is obtained via numeric integration.

Equation 2.5 is equivalent to HM equation 6.5. By applying a scale change of variable $t \to t/\sigma$ and substituting $f(t) = |\phi(t/\sigma)|$ and $g(t) = \arg(\phi(t/\sigma)))$ into equation 2.5 we get HM equation 6.5,

$$F(z) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{f(t)}{t} \sin(tz/\sigma - g(t)) dt. \tag{2.6}$$

2.3. Univariate Severity Model

The severity density is approximated to make the calculation of $\phi_x(t)$ easy. It is approximated with n intervals (a_k, a_{k+1}) of constant density d_k $(k = 1, ...n)$ and an optional point mass p at the maximal claim size a_{n+1} such that

$$\sum_{k=1}^{n} d_k(a_{k+1} - a_k) + p = 1. \tag{2.7}$$

Figure 2.1 shows a sample severity density with two intervals (a_1, a_2) and (a_2, a_3) and a point mass at a_3. With this severity model we easily obtain

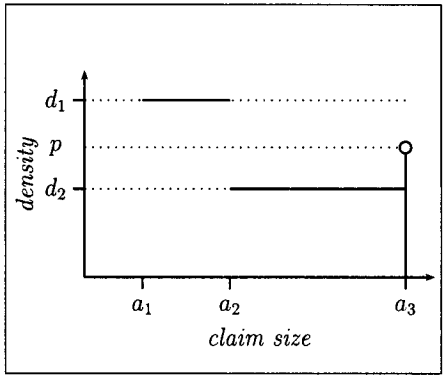
$$\begin{aligned} \phi_X(t) &= E_X(e^{itX}) \\ &= \sum_{k=1}^{n} d_k \frac{e^{ita_{k+1}} - e^{ita_k}}{it} + pe^{ita_{n+1}}. \end{aligned} \tag{2.8}$$

2.4. Univariate Numerical Inversion

Heckman and Meyers integrate 2.5 using five point Gaussian quadrature with special treatment of the portion of the integral near zero. We will extend this to two dimensions using five point quadrature first along one dimension and again along the second dimension.

FIGURE 2.1

UNIVARIATE SEVERITY DENSITY — INTERVALS OF CONSTANT DENSITY



3. BIVARIATE NUMERICAL FOURIER INVERSION

3.1. Bivariate Collective Risk Model

The collective risk model can be extended to model two collections of risks and their dependencies. There are two forms for this extension.

The first form is the *bivariate severity form.* It is useful for modeling aggregate losses together with the corresponding aggregate adjustment expenses. This form uses a single claim count distribution and a bivariate claim size distribution. While the bivariate pair (X_k, Y_k) may have any dependency structure, the pairs arising from different claims are assumed to be iid. The claim size pairs are also independent from the claim count N. The aggregate loss pair is

$$(Z_x, Z_y) = (X_1 + ... + X_N, Y_1 + ... + Y_N). \tag{3.1}$$

The second form is the *bivariate count form.* It is useful for modeling two risk collections with different but related claim counts. The claim size severities X_k and Y_j are separately iid and also independent from each other. The claim counts for each

risk collection arise from a bivariate claim count distribution. The claim count pair (M, N) is independent from each of the claim sizes. The aggregate pair is

$$(Z_x, Z_y) = (X_1 + ... + X_M, Y_1 + ... + Y_N). \tag{3.2}$$

This chapter will focus on the bivariate severity form, but the methods presented here can also be applied to the bivariate count form.

3.2. The Bivariate Aggregate Characteristic Function

The aggregate characteristic function for the bivariate severity form of the collective risk model is a composition of the claim count pgf with the bivariate severity characteristic function.

$$\begin{aligned}
\phi_{Z_x,Z_y}(s,t) &= E(e^{isZ_x+itZ_y}) \\
&= E(e^{is(X_1+...+X_N)+it(Y_1+...+Y_N)}) \\
&= E(e^{isX_1+itY_1}...e^{isX_N+itY_N}) \\
&= E_N(\phi_{X,Y}(s,t)^N|N) \\
&= PGF_N(\phi_{X,Y}(s,t))
\end{aligned} \tag{3.3}$$

For the bivariate count form, Wang [10] gives the aggregate characteristic function.

$$\phi_{Z_x,Z_y}(s,t) = PGF_{M,N}(\phi_X(s), \phi_Y(t)). \tag{3.4}$$

Where $PGF_{M,N}(s,t)$ is the bivariate claim count pgf.

Appendices A and B develop an expression for $F(z_x, z_y)$ in terms of $\phi_{Z_x,Z_y}(s,t)$ when F is continuous,

$$F(x,y) = \frac{1}{2}\big(F(x) + F(y)\big) - \frac{1}{4} + \frac{1}{4\pi^2}I, \tag{3.5}$$

where

$$\begin{aligned}
I &= 2\int_0^\infty \int_0^\infty \big(|\phi(s,t)|\cos(sx + ty - \arg(\phi(s,t))) - \\
&\quad |\phi(s,-t)|\cos(sx - ty - \arg(\phi(s,-t)))\big)\frac{dsdt}{(is)(it)}.
\end{aligned} \tag{3.6}$$

When F is not continuous, the right hand side of 3.5 yields $F(z_x, z_y) + m/4$, where m is a correction for probability mass that lies along the lines $Z_x = z_x$ and $Z_y = z_y$, and

$$\begin{aligned} m &= Pr(Z_x > z_x \cap Z_y = z_y) && (3.7)\\ &\quad + Pr(Z_y > z_y \cap Z_x = z_x) && (3.8)\\ &\quad - Pr(Z_x \leq z_x \cap Z_y = z_y) && (3.9)\\ &\quad - Pr(Z_y < z_y \cap Z_x = z_x). && (3.10) \end{aligned}$$

3.3. Bivariate Severity Model

In an extension of the univariate severity model, the bivariate severity density will be approximated with rectangles of constant density. That is, the severity domain will be divided into mn rectangles $(a_j, a_{j+1}) \times (b_k, b_{k+1})$ of constant density $d_{j,k}$ $(j = 1...m)(k = 1...n)$. Like the one dimensional case, this simplifies the calculation of $\phi_{X,Y}(s,t)$,

$$\begin{aligned} \phi_{X,Y}(s,t) &= E(e^{\imath sX + \imath tY}) \\ &= \sum_{j=1}^{m}\sum_{k=1}^{n} \int_{b_k}^{b_{k+1}} \int_{a_j}^{a_{j+1}} d_{j,k} e^{\imath sx + \imath ty} dx dy \\ &= \sum_{j=1}^{m}\sum_{k=1}^{n} d_{j,k} \frac{e^{\imath s a_{j+1}} - e^{\imath s a_j}}{\imath s} \frac{e^{\imath t b_{k+1}} - e^{\imath s b_k}}{\imath t}. \end{aligned} \quad (3.11)$$

Figure 3.1 shows a sample bivariate density.

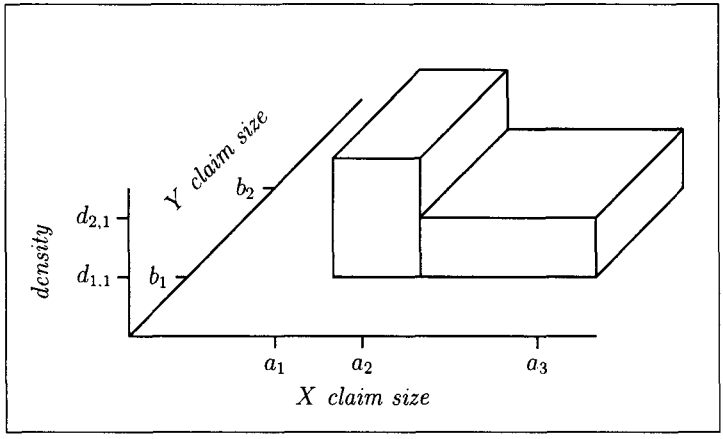
Here we have not included mass points or mass lines, but it is possible to do so.

3.4. Bivariate Numerical Fourier Inversion

We will make use of two-dimensional five point Gaussian quadrature. Appendix C provides additional descriptions of two-dimensional quadrature. Sample code will also be provided in a spreadsheet that can be downloaded from the CAS Web site. It will follow key elements of the HM code fairly closely.

FIGURE 3.1

BIVARIATE SEVERITY DENSITY—RECTANGLES OF CONSTANT DENSITY



In particular, HM split the line into 256 intervals of width $h = 2\pi\sigma/x_{max}$. We will split the grid into rectangles of widths $h_x = \pi/x_{max}$ and $h_y = \pi/y_{max}$ respectively. We are using half of the HM interval and trying to economize on the total number of rectangles. We leave out the additional factor of σ which is the standard deviation of the aggregate distribution and is not required. Heckman and Meyers additionally split the first interval into 5 smaller intervals $(0, h/16), (h/16, h/8), (h/8, h/4), (h/4, h/2), (h/2, h)$. This is helpful because the integrand changes rapidly near zero.

As suggested by HM it is speculated that the key source of error in this method is truncation error, since the integrals are from zero to infinity, but our algorithm must stop at a finite values. Errors in our sample calculations will be computed with comparisons to known values.

4. BIVARIATE EXAMPLES

This section presents two examples. The first example applies the 2d inversion technique to a bivariate severity and a claim count distribution allowing only a single claim. Thus, the aggregate distribution is the same as the bivariate severity and the error is readily computed.

The second example applies the 2d inversion to the same bivariate severity with a moderate number of expected counts. This result is compared to an exact calculation produced by a two dimensional version of Robertson's method [8].

4.1. Example 1—Exactly One Claim

Table 4.1 shows a sample bivariate severity distribution. If we also assume the claim count distribution has a 100% probability of 1 claim the resulting aggregate distribution computed by our method is shown in Table 4.2. This method should reproduce Table 4.1. The error is shown in Table 4.3.

TABLE 4.1

SAMPLE BIVARIATE SEVERITY CUMULATIVE DISTRIBUTION FUNCTION

$F(x,y)$		y				
		0	200	600	800	1,200
	0	0.0000	0.0000	0.0000	0.0000	0.0000
x	200	0.0000	0.4705	0.7557	0.7845	0.8120
	400	0.0000	0.4858	0.8243	0.8621	0.8990
	600	0.0000	0.4917	0.8540	0.8964	0.9380
	1,000	0.0000	0.4953	0.8735	0.9190	0.9640
	2,000	0.0000	0.4991	0.8949	0.9440	0.9930
	3,000	0.0000	0.4996	0.8978	0.9474	0.9970
	5,000	0.0000	0.5000	0.9000	0.9500	1.0000

TABLE 4.2

AGGREGATION OF SAMPLE BIVARIATE SEVERITY CDF WITH 100% PROBABILITY OF 1 CLAIM

$F(z_x, z_y)$		z_y 0	200	600	800	1,200
	0	0.0000	0.0047	0.0076	0.0079	0.0082
z_x	200	0.0011	0.4652	0.7485	0.7774	0.8046
	400	0.0012	0.4850	0.8236	0.8617	0.8985
	600	0.0012	0.4909	0.8535	0.8961	0.9377
	1,000	0.0012	0.4946	0.8731	0.9189	0.9639
	2,000	0.0012	0.4984	0.8945	0.9439	0.9929
	3,000	0.0012	0.4989	0.8975	0.9474	0.9969
	5,000	0.0012	0.4993	0.8996	0.9499	0.9999

TABLE 4.3

ERROR FOR EXAMPLE 1 AGGREGATE CDF

Error		z_y 0	200	600	800	1,200
	0	0.0000	0.0047	0.0076	0.0079	0.0082
z_x	200	0.0011	(0.0054)	(0.0072)	(0.0072)	(0.0074)
	400	0.0012	(0.0008)	(0.0007)	(0.0005)	(0.0005)
	600	0.0012	(0.0008)	(0.0005)	(0.0002)	(0.0003)
	1,000	0.0012	(0.0007)	(0.0004)	(0.0001)	(0.0001)
	2,000	0.0012	(0.0007)	(0.0004)	(0.0001)	(0.0001)
	3,000	0.0012	(0.0007)	(0.0004)	(0.0001)	(0.0001)
	5,000	0.0012	(0.0007)	(0.0004)	(0.0001)	(0.0001)

4.2. Example 2—Variable Claim Counts

In this example we use the claim size distribution from Example 1 and a claim count distribution with a maximum claim size. This allows us to compute the exact answer using an alternative method based on Robertson's one-dimensional method. Appendix D provides a brief discussion of a 2d Robertson method. In addition, sample R code showing the 2d Robertson calculation will be made available for downloading. Table 4.4 shows the count distribution. Table 4.5 shows the exact calculation based on the Robertson method. Table 4.6 shows the result from numerical Fourier inversion. The error is shown in Table 4.7. The errors are substantially smaller that those from Example 1 and this is attributed to the larger claim counts forcing the integrand to converge to zero more quickly.

5. CONCLUSION

Numerical Fourier inversion is a viable technique for exploring claim dependencies. When the claim counts are large, it may be more efficient than other techniques such as discrete Fourier transforms, recursion, or simulation.

Additional development is possible for alternate severity structures such as a bivariate distribution for primary and excess claim portions. Given the aggregate characteristic function, conditional expected values can also be computed. These calculations could have potential applications in reserving and surplus allocation.

6. ACKNOWLEDGEMENTS

The author would like to thank the members of the Correlation Working Party for encouragment and helpful suggestions. Particular thanks go to Glenn Meyers, Stephen Mildenhall, Shawna Ackerman, and Chuck Thayer.

TABLE 4.4

EXAMPLE 2—CLAIM COUNT DISTRIBUTION

Count	Probability	Cumulative
0	0.000	0.000
1	0.000	0.000
2	0.000	0.000
3	0.000	0.000
4	0.000	0.000
5	0.000	0.000
6	0.000	0.000
7	0.000	0.000
8	0.100	0.100
9	0.100	0.200
10	0.100	0.300
11	0.100	0.400
12	0.100	0.500
13	0.100	0.600
14	0.100	0.700
15	0.100	0.800
16	0.100	0.900
17	0.100	1.000
Mean	12.500	
Std	2.872	
Var	8.250	

TABLE 4.5

EXAMPLE 2—EXACT SOLUTION FROM 2D ROBERTSON METHOD

$F(z_x, z_y)$		z_y 1,000	2,000	4.000	6,000	10,000
	1,000	0.0009	0.0195	0.0550	0.0566	0.0567
z_x	2,000	0.0017	0.0577	0.3249	0.3924	0.3951
	3,000	0.0019	0.0688	0.4850	0.6649	0.6782
	4,000	0.0019	0.0724	0.5613	0.8193	0.8431
	5,000	0.0019	0.0737	0.5925	0.8928	0.9239
	6,000	0.0019	0.0744	0.6073	0.9287	0.9639
	8,000	0.0019	0.0747	0.6170	0.9547	0.9935
	15,000	0.0019	0.0747	0.6185	0.9601	1.0000

TABLE 4.6

EXAMPLE 2—AGGREGATE CDF FROM NUMERICAL FOURIER INVERSION

$F(z_x, z_y)$	0	z_y 1,000	2,000	4.000	6,000	10,000
	1,000	0.0009	0.0195	0.0550	0.0566	0.0567
z_x	2,000	0.0017	0.0577	0.3249	0.3924	0.3951
	3,000	0.0019	0.0688	0.4850	0.6649	0.6782
	4,000	0.0019	0.0724	0.5613	0.8193	0.8431
	5,000	0.0019	0.0737	0.5925	0.8928	0.9239
	6,000	0.0019	0.0744	0.6073	0.9287	0.9639
	8,000	0.0019	0.0747	0.6170	0.9547	0.9935
	15,000	0.0019	0.0747	0.6185	0.9601	1.0000

TABLE 4.7

EXAMPLE 2—ERROR

Error		z_y 1,000	2,000	4,000	6,000	10,000
	1,000	0.00000	0.00000	0.00000	0.00000	0.00000
z_x	2,000	0.00000	0.00000	0.00000	0.00000	0.00000
	3,000	0.00000	(0.00000)	(0.00000)	(0.00000)	(0.00000)
	4,000	0.00000	0.00000	0.00000	0.00000	0.00000
	5,000	0.00000	0.00000	0.00000	0.00000	0.00000
	6,000	0.00000	0.00000	0.00000	0.00000	0.00000
	8,000	0.00000	0.00000	0.00000	0.00000	0.00000
	15,000	0.00000	0.00000	0.00000	0.00000	0.00000

7. BIOGRAPHY

David Homer is a Principal and Vice President of the Reinsurance business of Towers Perrin. He holds a bachelor's degree in mathematics and physics from Swarthmore College and a master's degree in mathematics from the University of Pennsylvania.

David's current work involves pricing and modeling reinsurance solutions. He is a past winner of the Dorweiller Prize.

REFERENCES

[1] Abramowitz, Milton, and Irene A. Stegun, "Handbook of Mathematical Functions," New York, Dover Publications, Inc., 1970.

[2] Burden, Richard L., and J. Douglas Faires, "Numerical Analysis," 3rd ed., Boston, Prindle, Weber and Schmidt, 1985.

[3] Heckman, Philip E. and Glenn G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *PCAS* LXX, 1983, pp. 22–61.

[4] Homer, David L. and David R. Clark, "Insurance Applications of Bivariate Distributions," *PCAS* XC, 2003, pp. 274–307.

[5] Klugman, Stuart A., Harry H. Panjer, and E. Willmot Gordon, *Loss Models: From Data to Decisions*, New York, John Wiley & Sons, Inc., 1998.

[6] Meyers, Glenn G., "Discussion of Shaun Wang's Aggregation of Correlated Risk Portfolios: Models and Algorithms," *PCAS* LXXXV, 1999, pp. 781–805.

[7] Panjer, Harry H., "Recursive Evaluation of a Family of Compound Distributions," *ASTIN Bulletin*, Vol. 12, No. 1, 1981, pp. 22–26.

[8] Robertson, John P., "The Computation of Aggregate Loss Distributions," *PCAS* LXXIX, 1992, pp. 57–133.

[9] Sundt, Bjørn, "On Multivariate Panjer Recursions," *ASTIN Bulletin*, Vol. 29, No. 1, 1999, pp. 29–45.

[10] Wang, Shaun S., "Aggregation of Correlated Risk Portfolios: Models and Algorithms," *PCAS* LXXXV, 1998, pp. 848–939.

[11] Walhin, Jean-François, "On the Optimality of Multiline Excess of Loss Covers," *CAS Forum*, Spring 2003, pp. 231-243.

APPENDIX A

TWO-DIMENSIONAL INTEGRATION FORMULA

Consider the integral

$$I = \int_0^\infty \int_0^\infty e^{\imath sx}(e^{ity}\phi(-s,-t) - e^{-ity}\phi(-s,t)) - e^{-isx}(e^{ity}\phi(s,-t) - e^{-ity}\phi(s,t))\frac{dsdt}{isit}. \quad (A.1)$$

Substitute the integral form for ϕ and apply Fubini's theorem to change the order of integration. Then,

$$I = \int_{-\infty}^\infty \int_{-\infty}^\infty \int_0^\infty \int_0^\infty e^{is(x-u)}(e^{it(y-v)} - e^{\imath t(v-y)}) - e^{\imath s(u-x)}(e^{it(y-v)} - e^{it(v-y)})\frac{dsdtdF(u,v)}{isit}. \quad (A.2)$$

Since

$$\int_0^\infty \frac{e^{isx} - e^{-isx}}{is} ds = \pi \mathrm{sgn}(x), \quad (A.3)$$

where

$$\mathrm{sgn}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}, \quad (A.4)$$

$$
\begin{aligned}
I &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \pi \mathrm{sgn}(x-u)(e^{it(y-v)} - e^{it(v-y)})\frac{dt}{it} dF(u,v) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \pi^2 \mathrm{sgn}(x-u)\mathrm{sgn}(y-v) dF(u,v) \\
&= \int_{-\infty}^{x}\int_{-\infty}^{y} \pi^2 dF(u,v) - Pr(Z_x \le z_x \cap Z_y = z_y) \\
&\quad -Pr(Z_y < z_y \cap Z_x = z_x) \\
&\quad -\int_{-\infty}^{x}\int_{y}^{\infty} \pi^2 dF(u,v) + Pr(Z_y > z_y \cap Z_x = z_x) \\
&\quad -\int_{x}^{\infty}\int_{-\infty}^{y} \pi^2 dF(u,v) + Pr(Z_x > z_x \cap Z_y = z_y) \\
&\quad +\int_{x}^{\infty}\int_{y}^{\infty} \pi^2 dF(u,v) \\
&= \pi^2(F(x,y) - (F(x,\infty) - F(x,y)) - (F(\infty,y) - F(x,y)) \\
&\quad +(1 - F(\infty,y) - F(x,\infty) + F(x,y)) + m). \qquad \text{(A.5)}
\end{aligned}
$$

Where m is a correction for probability mass that lies along the lines $Z_x = z_x$ and $Z_y = z_y$ when $F(z_x, z_y)$ is not continuous, since $sgn(0) = 0$.

$$
\begin{aligned}
m &= Pr(Z_x > z_x \cap Z_y = z_y) && \text{(A.6)} \\
&\quad +Pr(Z_y > z_y \cap Z_x = z_x) && \text{(A.7)} \\
&\quad -Pr(Z_x \le z_x \cap Z_y = z_y) && \text{(A.8)} \\
&\quad -Pr(Z_y < z_y \cap Z_x = z_x). && \text{(A.9)}
\end{aligned}
$$

So,

$$I = \pi^2(4F(x,y) - 2(F(x) + F(y)) + 1 + m). \qquad \text{(A.10)}$$

Finally,

$$F(x,y) = \frac{1}{2}\big(F(x) + F(y)\big) - \frac{1}{4} + \frac{1}{4\pi^2}I - \frac{m}{4}. \qquad \text{(A.11)}$$

APPENDIX B

EXPANSION OF I FOR NUMERICAL INTEGRATION

Appendix A provides an expression for the bivariate cdf $F(x,y)$.

$$F(x,y) = \frac{1}{2}\big(F(x)+F(y)\big) - \frac{1}{4} + \frac{1}{4\pi^2} I, \tag{B.1}$$

where,

$$\begin{aligned} I \;=\; & \int_0^\infty \int_0^\infty \Big(e^{isx}\left(e^{ity}\phi(-s,-t) - e^{-ity}\phi(-s,t)\right) \\ & -e^{-isx}\left(e^{ity}\phi(s,-t) - e^{-ity}\phi(s,t)\right)\Big)\frac{dsdt}{(is)(it)}. \end{aligned} \tag{B.2}$$

It will be helpful to write ϕ in polar form and make use of a few symmetries. Let

$$R(s,t) = |\phi(s,t)| \tag{B.3}$$

$$\theta(s,t) = \arg(\phi(s,t)), \tag{B.4}$$

then,

$$\phi(s,t) = R(s,t)e^{i\theta(s,t)} = E(e^{isx+ity}). \tag{B.5}$$

We have the complex conjugate of ϕ

$$\overline{\phi(s,t)} = R(s,t)e^{-i\theta(s,t)} \tag{B.6}$$

$$= \overline{E(e^{isx+ity})} \tag{B.7}$$

$$= E(e^{-isx-ity}) \tag{B.8}$$

$$= R(-s,-t)e^{i\theta(-s,-t)}. \tag{B.9}$$

Thus,

$$\theta(s,t) = -\theta(-s,-t) \tag{B.10}$$

$$\theta(s,-t) = -\theta(-s,t) \tag{B.11}$$

$$R(s,t) = R(-s,-t) \tag{B.12}$$

$$R(s,-t) = R(-s,t). \tag{B.13}$$

Now writing I with ϕ in polar form

$$I = \int_0^\infty \int_0^\infty \left(e^{\imath sx}\left(e^{\imath ty}R(-s,-t)e^{\imath\theta(-s,-t)} - e^{-\imath ty}R(-s,t)e^{\imath\theta(-s,t)}\right) - e^{-\imath sx}\left(e^{\imath ty}R(s,-t)e^{\imath\theta(s,-t)} - e^{-\imath ty}R(s,t)e^{\imath\theta(s,t)}\right)\right)\frac{dsdt}{(\imath s)(\imath t)} \tag{B.14}$$

and simplifying using equations B.10-B.13,

$$I = \int_0^\infty \int_0^\infty \left(R(s,t)\left(e^{\imath sx+\imath ty-\imath\theta(s,t)} + e^{-\imath sx-\imath ty+\imath\theta(s,t)}\right) - R(s,-t)\left(e^{\imath sx-\imath ty-\imath\theta(s,-t)} + e^{-\imath sx+\imath ty+\imath\theta(s,-t)}\right)\right)\frac{dsdt}{(\imath s)(\imath t)} \tag{B.15}$$

we can now write

$$I = \int_0^\infty \int_0^\infty \left(R(s,t)2cos(sx+ty-\theta(s,t)) - R(s,-t)2cos(sx-ty-\theta(s,-t))\right)\frac{dsdt}{(\imath s)(\imath t)}. \tag{B.16}$$

In terms of ϕ we have

$$I = \int_0^\infty \int_0^\infty \left(|\phi(s,t)|2\cos(sx+ty-\arg(\phi(s,t))) - |\phi(s,-t)|2\cos(sx-ty-\arg\phi(s,-t))\right)\right)\frac{dsdt}{(\imath s)(\imath t)}. \tag{B.17}$$

APPENDIX C

GAUSSIAN QUADRATURE FOR TWO DIMENSIONS

This section will develop the formulae for two-dimensional Gaussian quadrature. The basic form to be approximated is

$$I = \int_a^b \int_c^d f(x,y)dxdy. \tag{C.1}$$

Using a change of variables, change the integral domain from the rectangle $[a,b]\times[c,d]$ to $[-1,1]\times[-1,1]$.

$$u = \frac{1}{b-a}(2x-a-b). \tag{C.2}$$

$$v = \frac{1}{d-c}(2y-c-d). \tag{C.3}$$

Equation C.1 becomes

$$I = \int_{-1}^1 \int_{-1}^1 f\left(\frac{u(b-a)+a+b}{2}, \frac{v(d-c)+c+d}{2}\right)\frac{(b-a)(d-c)}{4}dudv. \tag{C.4}$$

The integral C.4 is now computed as a double sum,

$$I = \sum_{i=1}^5 \sum_{j=1}^5 w_i w_j f\left(\frac{x_i(b-a)+a+b}{2}, \frac{x_j(d-c)+c+d}{2}\right)\frac{(b-a)(d-c)}{4} + \epsilon. \tag{C.5}$$

The error term ϵ depends on how well $f(x,y)$ can be approximated by polynomials of finite degree (nine or less for five point Gaussian quadrature). By choosing sufficiently small intervals ϵ can be made small. See [2] for additional details.

The quadrature values x_i and w_i are taken from Abramowitz and Stegun [1].

TABLE C.1

ABSCISSAS AND WEIGHTS FOR FIVE POINT GAUSSIAN QUADRATURE

k	x_k	w_k
1	-0.90617 98459 38664	0.23692 68850 56189
2	-0.53846 93101 05683	0.47862 86704 99366
3	0.00000 00000 00000	0.56888 88888 88889
4	+0.53846 93101 05683	0.47862 86704 99366
5	+0.90617 98459 38664	0.23692 68850 56189

APPENDIX D

TWO-DIMENSIONAL ROBERTSON METHOD

This appendix provides a brief discussion of extending Robertson's method [8] to two dimensions. It begins with a summary of the one dimensional method.

Robertson's method computes the aggregate distribution for a finite claim count distribution and a claim size distribution with equal width and constant density intervals. The method is exact and it uses discrete Fourier transforms.

A more basic application of the discrete Fourier transform requires a discrete claim size distribution with claim sizes at integral intervals.

Robertsons's method uses the usual discrete Fourier technique to compute convolutions, but adds a correction to reflect the constant density claim size intervals. The method is quite clever and it is not hard to develop an intuition to see why it works.

Consider a discrete random variable X with integral size intervals of width I. Now add a random variable U that is uniform on the interval I. The result $X + U$ is a random variable with claim size intervals of constant density.

This observation can be applied to develop the aggregate distribution with claim size distribution F_{X+U} and claim count distribution P. Note that the sum of n independent copies of $X + U$ has the same distribution as the sum of n independent copies of X plus n independent copies of U. The aggregate cumulative distribution

function is then

$$F(z) = \sum_{n=0}^{n=n_{max}} P(n) F_{X+U}^{(n)}(z) \tag{D.1}$$

$$= \sum_{n=0}^{n=n_{max}} P(n)(F_X^{(n)}(z) * F_U^{(n)}(z)). \tag{D.2}$$

The quantity $F_X^{(n)}(z)$ can be computed with the discrete Fourier transform and Robertson explains how $F_U^{(n)}(z)$ can be obtained. For integral values of z the convolution of the two is

$$F_{X+U}^{(n)}(z) = \sum_{j=0}^{j=z} (f_X^{(n)}(j) F_U^{(n)}(z-j)). \tag{D.3}$$

Now consider $F_{X+U}^{(n)}$ for integral values of z,

$$F_{X+U}^{(n)}(z) - F_{X+U}^{(n)}(z-1) = \sum_{j=0}^{j=z} (f_X^{(n)}(j)(F_U^{(n)}(z-j) - F_U^{(n)}(z-j-1)) \tag{D.4}$$

Robertson explains that the differences $(F_U^{(n)}(z-j) - F_U^{(n)}(z-j-1))$ are the factors a_{z-j}^n where,

$$a_0^n = 1/n! \qquad n \geq 1, \tag{D.5}$$

$$a_j^1 = 0 \qquad j \geq 1, \tag{D.6}$$

$$a_j^n = (1/n)((n-j)a_{j-1}^{n-1} + (j+1)a_j^{n-1} \quad n \geq 2,\ j \geq 1. \tag{D.7}$$

The right hand side of equation D.4 is the convolution of $f^{(n)}$ with a_j^n and can be computed using discrete Fourier transforms.

The two-dimensional extension works in exactly the same way by considering the discrete random pair (X, Y) with integral size intervals of widths I and J. By adding an independent pair (U, V) where U is uniform on I and V is uniform on J, we get the random pair $(X+U, Y+V)$, which has claim size rectangles of constant density. The two-dimensional correction factors for the nth convolution are outer products of the one-dimensional correction factors, since U and V are independent.

$$a_{(i,j)}^n = a_i^n a_j^n \tag{D.8}$$

Sample R-code will be submitted with this chapter for downloading.

The Report of the Research Working Party on Correlations and Dependencies Among All Risk Sources

Part 3

The Common Shock Model for Correlated Insurance Losses

Glenn Meyers, FCAS, MAAA, Ph.D.

Abstract

This chapter discusses an approach to the correlation problem where losses in different lines of insurance are linked by a common variation (or shock) in the parameters of each line's loss model. The chapter begins with a simple common shock model and graphically illustrates the effect of the magnitude of the shocks on correlation. Next it describes some more general common shock models that involve common shocks to both the claim count and claim severity distributions. It derives formulas for the correlation between lines of insurance in terms of the magnitude of the common shocks and the parameters of the underlying claim count and claim severity distributions. Finally, it shows how to estimate the magnitude of the common shocks. A feature of this estimation is that it uses the data from several insurers.

1. Introduction

In the study of insurer enterprise risk management, "correlation" has been an important, but elusive phenomenon. Those who have tried to model insurer risk assuming independence have almost always understated the variability that is observed in publicly available data. Most actuaries would agree that "correlation" is the major missing link to the realistic modeling of insurance losses.

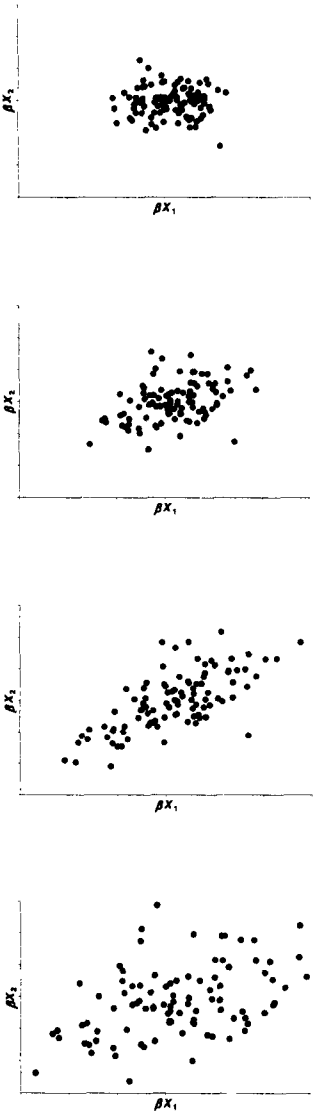
This chapter discusses an approach to the correlation problem where losses in different lines of insurance are linked by a common variation (or shock) in the parameters of each line's loss model. Here is an outline of what is to follow.

- I will begin with a simple common shock model and graphically illustrate the effect of the magnitude of the shocks on correlation.
- Next I will describe some more general common shock models that involve common shocks to both the claim count and claim severity distributions. I will derive formulas for the correlation between lines of insurance in terms of the magnitude of the common shocks and the parameters of the underlying claim count and claim severity distributions.
- Finally, I will show how to estimate the magnitude of the common shocks. A feature of this estimation is that it uses the data from several insurers.

2. A Simple Common Shock Model

Let X_1 and X_2 be independent positive random variables. Also let β be a positive random variable with mean 1 and variance b. If $b > 0$, the random variables βX_1 and βX_2 tend to be larger when β is large, and tend to be smaller when β is small. Hence the random variables βX_1 and βX_2 are correlated. Figures 1-4 below illustrate this graphically.

I will refer to the β as the "common shock" and refer to the b as the magnitude of the common shocks. Figures 1-4 illustrate graphically that coefficient of correlation depend upon b and the volatility of the random variables X_1 and X_2.



I will now derive formulas for the coefficient of correlation between the random variables βX_1 and βX_2. This derivation will be detailed and I believe that it is worth the reader's time to master these details in order to appreciate much of what is to follow.

Let's begin with the derivation of two general equations from which I will derive much of what follows. These equations calculate the global covariance (or variance) in terms of the covariances (or variances) that are given conditionally on a parameter θ.

$$\begin{aligned} Cov[X,Y] &= E[X \cdot Y] - E[X] \cdot E[Y] \\ &= E_\theta\left[E[X \cdot Y \mid \theta]\right] - E_\theta\left[E[X \mid \theta]\right] \cdot E_\theta\left[E[Y \mid \theta]\right] \\ &= E_\theta\left[E[X \cdot Y \mid \theta] - E[X \mid \theta] \cdot E[Y \mid \theta]\right] \\ &\quad + E_\theta\left[E[X \mid \theta] \cdot E[Y \mid \theta]\right] - E_\theta\left[E[X \mid \theta]\right] \cdot E_\theta\left[E[Y \mid \theta]\right] \\ &= E_\theta\left[Cov[X,Y \mid \theta]\right] + Cov_\theta\left[E[X \mid \theta], E[Y \mid \theta]\right] \end{aligned} \tag{1}$$

An important special case of this equation occurs when $X = Y$.

$$Var[X] = E_\theta\left[Var[X \mid \theta]\right] + Var_\theta\left[E[X \mid \theta]\right] \tag{2}$$

Now let's apply Equations 1 and 2 to the common shock model given at the beginning of this section.

$$\begin{aligned} Cov[\beta X_1, \beta X_2] &= E_\beta\left[Cov[\beta X_1, \beta X_2 \mid \beta]\right] + Cov_\beta\left[E[\beta X_1 \mid \beta], E[\beta X_2 \mid \beta]\right] \\ &= E_\beta\left[\beta^2 Cov[X_1, X_2]\right] + Cov_\beta\left[\beta E[X_1], \beta E[X_2]\right] \\ &= E_\beta\left[\beta^2 \cdot 0\right] + E[X_1] \cdot E[X_2] \cdot Cov_\beta[\beta, \beta] \\ &= E[X_1] \cdot E[X_2] \cdot b \end{aligned} \tag{3}$$

$$\begin{aligned} Var[\beta X_1] &= E_\beta\left[Var[\beta X_1 \mid \beta]\right] + Var_\beta\left[E[\beta X_1 \mid \beta]\right] \\ &= E_\beta\left[\beta^2 \cdot Var[X_1]\right] + Var_\beta\left[\beta \cdot E[X_1]\right] \\ &= Var[X_1] \cdot E_\beta\left[\beta^2\right] + E[X_1]^2 \cdot Var_\beta[\beta] \\ &= Var[X_1] \cdot (1+b) + E[X_1]^2 \cdot b \end{aligned} \tag{4}$$

Similarly:
$$Var[\beta X_2] = Var[X_2] \cdot (1+b) + E[X_2]^2 \cdot b. \tag{5}$$

Equations 3-5 can now be plugged into the following equation for the coefficient of correlation.

$$\rho[\beta X_1, \beta X_2] = \frac{Cov[\beta X_1, \beta X_2]}{\sqrt{Var[\beta X_1] \cdot Var[\beta X_2]}} \tag{6}$$

Plugging Equations 3-5 into Equation 6 yields a simple expression if we give X_1 and X_2 identical distributions with a common coefficient of variation, CV.

$$\rho[\beta X_1, \beta X_2] = \frac{b}{(CV)^2 \cdot (1+b) + b} \tag{7}$$

The coefficients of correlation given in Figures 1-4 were calculated using Equation 7.

At this point we can observe that the common shock model, as formulated above, implies that the coefficient of correlation depends not only the magnitude of the shocks, but also the volatility of the distributions that receive the effect of the random shocks.

3. The Collective Risk Model

The collective risk model describes the distribution of total losses arising from a two-step process where: (1) the number of claims is random; and (2) for each claim, the claim severity is random. In this section I will specify a particular version of the collective risk model. In the next section I will subject both the claim count and claim severity distributions to common shocks across different lines of insurance and calculate the correlations implied by this model.

Let's begin by considering a Poisson distribution with mean λ and variance λ for the claim count random variable, N. Let χ be a random variable with mean 1 and variance c. The claim count distribution[1] for our version of the collective risk model will be defined by the two-step process where; (1) χ is selected at random; and (2) the claim count is selected at random from a Poisson distribution with mean $\chi\lambda$. The mean of this distribution is λ. I will refer to the parameter c as the contagion parameter.

[1] If χ has a gamma distribution, it is well known that this claim count distribution is the negative binomial distribution. None of the results derived in this paper will make use of this fact.

Using Equation 2, one calculates the variance of N as:

$$\begin{aligned} Var[N] &= E_\chi\left[Var[N \mid \chi]\right] + Var_\chi\left[E[N \mid \chi]\right] \\ &= E_\chi[\chi\lambda] + Var_\chi[\chi\lambda] \\ &= \lambda + c \cdot \lambda^2 \end{aligned} \tag{8}$$

Let Z_i be a random variable for claim severity for the i^{th} claim. We will assume that each Z_i is identically distributed with mean μ and variance σ^2. For random claim count, N, let:

$$X = Z_1 + \ldots + Z_N.$$

The mean of X is $\lambda\mu$. Using Equation 2 we calculate the variance of X as:

$$\begin{aligned} Var[X] &= E_N\left[Var[X \mid N]\right] + Var_N\left[E[X \mid N]\right] \\ &= E_N\left[N \cdot \sigma^2\right] + Var_N[N \cdot \mu] \\ &= \lambda \cdot \sigma^2 + \mu^2 \cdot \left(\lambda + c \cdot \lambda^2\right) \\ &= \lambda \cdot \left(\sigma^2 + \mu^2\right) + c \cdot \lambda^2 \cdot \mu^2 \end{aligned} \tag{9}$$

At this point, I would like to introduce a notion of risk size and specify my assumptions on how the parameters of this model change with risk size.

1. The size of the risk is proportional to the expected claim count, λ.

2. The parameters of the claim severity distribution, μ and σ, are the same for all risk sizes.

3. The contagion parameter, c, is the same for all risk sizes.

I do not claim that these assumptions are applicable to all situations. For example, increasing the size of an insured building will expose an insurer to a potentially larger property insurance claim.

I do believe these assumptions are applicable in the context of this chapter, enterprise risk management. As an insurer increases the number of risks that it insureds, its total expected claim count, λ, increases. If each risk that it adds on is similar to its existing risks, it is

reasonable to expect μ and σ to be the same. One way to think of the contagion parameter, c, is as a measure of the uncertainty in the claim frequency. I believe it is reasonable to think this uncertainty applies to all risks simultaneously.

While a set of assumptions may sound reasonable, ultimately one should empirically test the predictions of such a model. I will do so below after I complete the description of my proposed model.

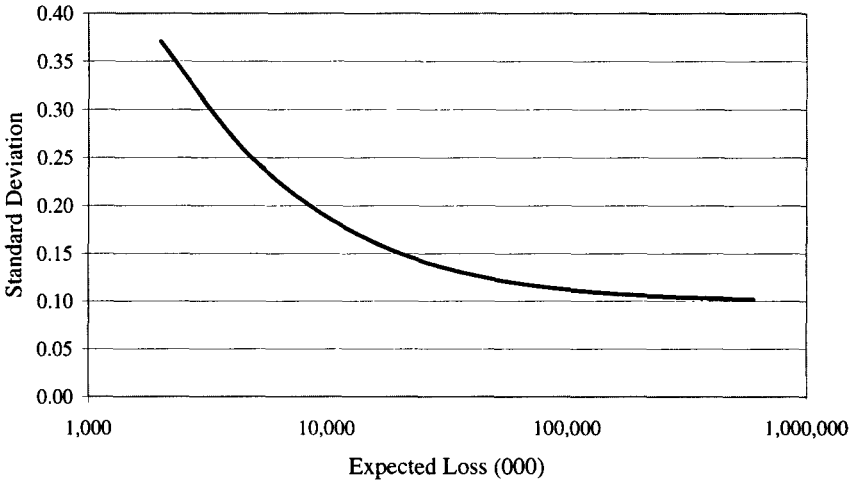
If the risk size is proportional to the expected claim count, λ, under the above assumptions it is also proportional to the expected loss $\lambda \cdot \mu$. In this chapter let's define the loss ratio as the ratio of the random loss X to its expected loss $E[X] = \lambda \cdot \mu$.

Equation 10 shows that the standard deviation of the loss ratio, $R = X/E[X]$ decreases asymptotically to $\sqrt{c}$ as we increase the size of the risk. Figure 5 below illustrates this graphically.

$$\text{Standard Deviation}[R] = \frac{\sqrt{\lambda \cdot (\sigma^2 + \mu^2) + c \cdot \lambda^2 \cdot \mu^2}}{\lambda \cdot \mu} \xrightarrow[\lambda \to \infty]{} \sqrt{c} \qquad (10)$$

Figure 5

Loss Ratios for the Collective Risk Model
$\mu = 15{,}000$ $\sigma = 60{,}000$ $c = 0.01$



4. Common Shocks in the Collective Risk Model

I will now apply the ideas underlying the common shock model described in Section 2, to the collective risk model described in Section 3. I will start with the claim count distributions.

Let N_1 and N_2 be two claim count random variables with $E[N_i] = \lambda_i$ and $Var[N_i] = \lambda_i + c_i \cdot \lambda_i^2$ for $i = 1$ and 2.

Let α be a random variable with $E[\alpha] = 1$ and $Var[\alpha] = g$.

I now introduce common shocks into the joint distribution of N_1 and N_2 by selecting N_1 and N_2 from claim count distributions with means $\alpha \cdot \lambda_1$ and $\alpha \cdot \lambda_2$ respectively and variances $\alpha \cdot \lambda_1 + c_1 \cdot (\alpha \cdot \lambda_1)^2$ and $\alpha \cdot \lambda_2 + c_2 \cdot (\alpha \cdot \lambda_2)^2$. Let's calculate the covariance matrix for N_1 and N_2.

Using Equation 2 to calculate the diagonal elements yields:

$$\begin{aligned} Var[N_i] &= E_\alpha\left[Var[N_i \mid \alpha]\right] + Var_\alpha\left[E[N_i \mid \alpha]\right] \\ &= E_\alpha\left[\alpha \cdot \lambda_i + c_i \cdot \alpha^2 \cdot \lambda_i^2\right] + Var_\alpha\left[\alpha \cdot \lambda_i\right] \\ &= \lambda_i + c_i \cdot \lambda_i^2 \cdot (1+g) + \lambda_i^2 \cdot g \\ &= \lambda_i + \lambda_i^2 \cdot (c_i + g + c_i \cdot g) \end{aligned} \tag{11}$$

Using Equation 1 to calculate the off-diagonal elements yields:

$$\begin{aligned} Cov[N_1, N_2] &= E_\alpha\left[Cov[N_1, N_2 \mid \alpha]\right] + Cov_\alpha\left[E[N_1 \mid \alpha], E[N_2 \mid \alpha]\right] \\ &= E_\alpha[0] + Cov_\alpha[\alpha \cdot \lambda_1, \alpha \cdot \lambda_2] \\ &= g \cdot \lambda_1 \cdot \lambda_2 \end{aligned} \tag{12}$$

Now let's add independent random claim severities, Z_1 and Z_2 to our common shock model. Here are the calculations for the elements of the covariance matrix for the total loss random variables X_1 and X_2.

$$\begin{aligned} Var[X_i] &= E_{N_i}\left[Var[X_i \mid N_i]\right] + Var_{N_i}\left[E[X_i \mid N_i]\right] \\ &= E_{N_i}\left[N_i \cdot \sigma_i^2\right] + Var_{N_i}\left[N_i \cdot \mu_i\right] \\ &= \lambda_i \cdot \sigma_i^2 + \mu_i^2 \cdot \left(\lambda_i + \lambda_i^2 \cdot (c_i + g + c_i \cdot g)\right) \\ &= \lambda_i \cdot \left(\sigma_i^2 + \mu_i^2\right) + \lambda_i^2 \cdot \mu_i^2 \cdot (c_i + g + c_i \cdot g) \end{aligned} \tag{13}$$

$$\begin{aligned} Cov[X_1, X_2] &= E_\alpha\left[Cov[X_1, X_2 \mid \alpha]\right] + Cov_\alpha\left[E[X_1 \mid \alpha], E[X_2 \mid \alpha]\right] \\ &= E_\alpha[0] + Cov_\alpha[\alpha \cdot \lambda_1 \cdot \mu_1, \alpha \cdot \lambda_2 \cdot \mu_2] \\ &= g \cdot \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \end{aligned} \tag{14}$$

Finally, let's multiply the claim severity random variables, Z_1 and Z_2, by a random variable β with $E[\beta] = 1$ and $\mathrm{Var}[\beta] = b$. Here are the calculations for the elements of the covariance matrix for the total loss random variables X_1 and X_2.

$$\begin{aligned} Var[X_i] &= E_\beta\left[Var[X_i \mid \beta]\right] + Var_\beta\left[E[X_i \mid \beta]\right] \\ &= E_\beta\left[\lambda_i \cdot \beta^2 \cdot \left(\sigma_i^2 + \mu_i^2\right) + \lambda_i^2 \cdot \beta^2 \cdot \mu_i^2 \cdot \left(c_i + g + c_i \cdot g\right)\right] + Var_\beta\left[\lambda_i \cdot \beta \cdot \mu_i\right] \\ &= \left(\lambda_i \cdot \left(\sigma_i^2 + \mu_i^2\right) + \lambda_i^2 \cdot \mu_i^2 \cdot \left(c_i + g + c_i \cdot g\right)\right) \cdot E\left[\beta^2\right] + \lambda_i^2 \cdot \mu_i^2 \cdot Var[\beta] \\ &= \lambda_i \cdot \left(\mu_i^2 + \sigma_i^2\right) \cdot (1+b) + \lambda_i^2 \cdot \mu_i^2 \cdot \left(c_i + g + b + c_i \cdot g + c_i \cdot b + g \cdot b + c_i \cdot g \cdot b\right) \end{aligned} \tag{15}$$

$$\begin{aligned} Cov[X_1, X_2] &= E_\beta\left[Cov[X_1, X_2 \mid \beta]\right] + Cov_\beta\left[E[X_1 \mid \beta], E[X_2 \mid \beta]\right] \\ &= E_\beta\left[g \cdot \lambda_1 \cdot \beta \cdot \mu_1 \cdot \lambda_2 \cdot \beta \cdot \mu_2\right] + Cov_\beta\left[\lambda_1 \cdot \beta \cdot \mu_1, \lambda_2 \cdot \beta \cdot \mu_2\right] \\ &= g \cdot \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot E\left[\beta^2\right] + \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot Var[\beta] \\ &= \lambda_1 \cdot \mu_1 \cdot \lambda_2 \cdot \mu_2 \cdot (b + g + b \cdot g) \end{aligned} \tag{16}$$

I now complete my description of this version of the collective risk model with the following two assumptions.

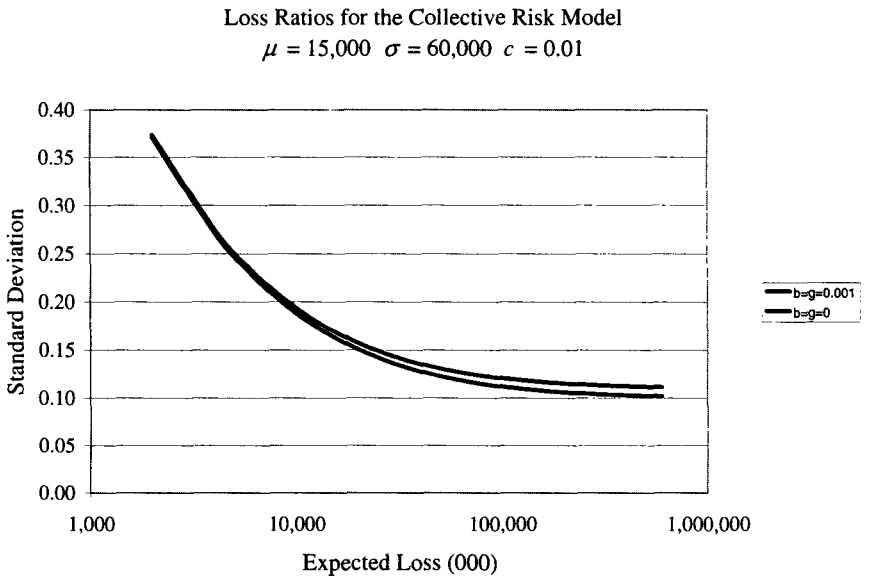
1. b and g are the same for all risk sizes.
2. b and g are the same for all lines of insurance.

The parameters b and g represent parameter uncertainty that applies across lines of insurance and it seems reasonable to assume that this uncertainty is independent of the size of risk. I made the second assumption to keep the math simple without sacrificing the main themes of this chapter. In practice I have allowed g to vary by line of insurance. I will leave it as an exercise to the reader to show that you can replace g in Equations 14 and 16 with $\sqrt{g_1 \cdot g_2}$ when the coefficient of correlation between α_1 and α_2 is equal to one.

Now I will illustrate the implications of this model for loss ratios as we vary the size of risk. My example will assume that μ = 16,000, σ = 60,000 and c = 0.010 for each line of insurance. The additional parameters will be $b = g = 0.001$. In the final sections, I will show that these are reasonable choices of the parameters.

First let's note that since b and g are small compared to c, introducing b and g into the model has little effect on the standard deviation of the loss ratio, although what effect there is, increases with the size of the risk. This is illustrated by Figure 6.

Figure 6

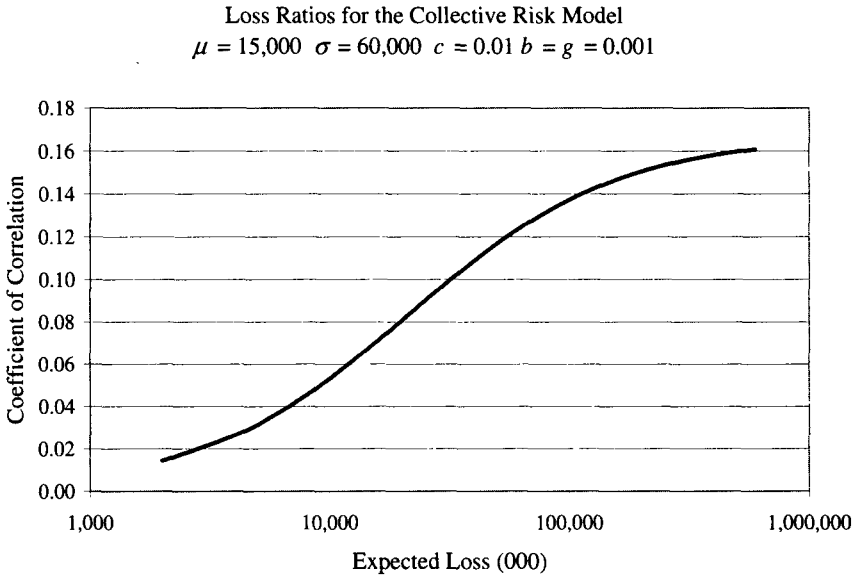


However, the coefficient of correlation, as defined by:

$$\rho[R_1, R_2] = \frac{Cov[R_1, R_2]}{\sqrt{Var[R_1] \cdot Var[R_2]}};$$

increases significantly as you increase the size of the risk. In Figure 7 below, it is almost negligible for small risks.

Figure 7



When I show similar exhibits to other actuaries, I often find that their expectations of the coefficient of correlations are much higher. My best rationale for these expectations is that most expect a positive number between 0 and 1, and 0.5 seems like a good choice.

Even so, these (perhaps) seemingly small correlations can have a significant effect for a multiline insurer seeking to manage its risk as I shall now illustrate.

Let's consider the covariance matrix for an insurer writing n lines of business.

$$\begin{pmatrix} Var[X_1] & Cov[X_1,X_2] & \ldots & Cov[X_1,X_n] \\ Cov[X_2,X_1] & Var[X_2] & \ldots & Cov[X_2,X_n] \\ \ldots & \ldots & \ldots & \ldots \\ Cov[X_n,X_2] & Cov[X_n,X_2] & \ldots & Var[X_n] \end{pmatrix}$$

The standard deviation of the insurer's total losses, $X_1 + \ldots + X_n$, is the square root of the sum of the elements of the covariance matrix. If $b = g = 0$, this sum consists of the n variances along the diagonal. If b and/or g $\neq 0$, then there are $n^2 - n$ off-diagonal covariances included in the sum. As n increases, so does the effect of even a "small" correlation. This is illustrated in Figures 8 and 9.

Figure 8

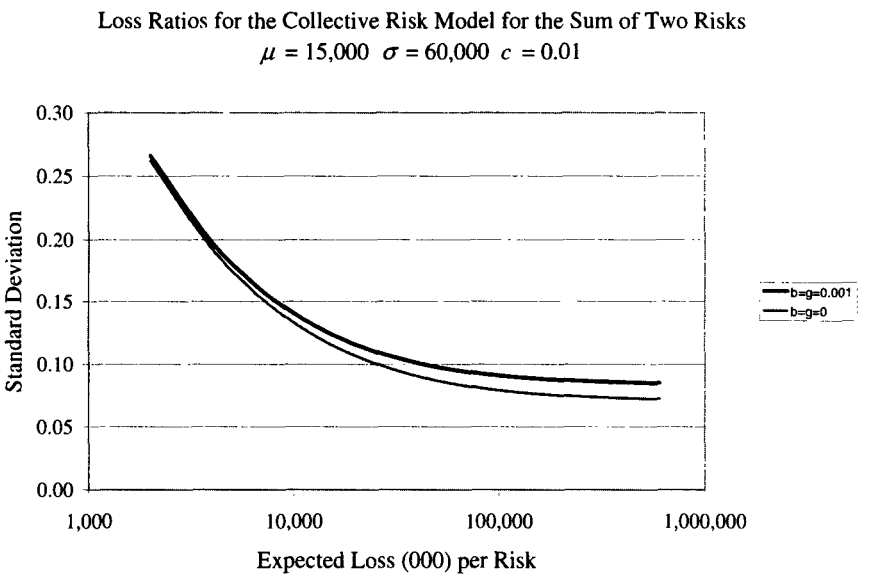
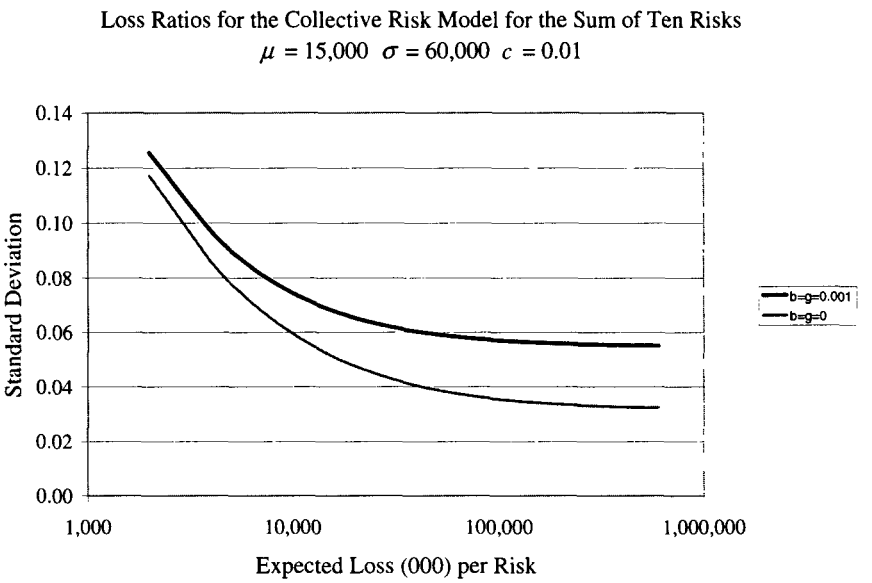


Figure 9



5. An Empirical Test of the Model

The collective risk model, as defined above, makes predictions about how the volatility and correlation statistics of loss ratios vary with insurer characteristics. These predictions should, at least in principle, be observable when one looks at a sizeable collection of insurance companies. In this section I will demonstrate that data that is publicly available on Schedule P is consistent with the major predictions of this model.

Data in Schedule P includes net losses, reported to date, and net premium by major line of insurance over a 10-year period of time. With Schedule P data for several insurers I calculated various statistics such as standard deviations and coefficients of correlation between lines of insurance for several insurers. Testing the model consisted of comparing these statistics with available information about each insurer.

But first I will discuss some of the difficulties with Schedule P data and discuss how, in work done jointly with Fred Klinker (see Meyers, Klinker and Lalonde [3] for details), we dealt with these difficulties.

Schedule P premiums and reserves vary in largely predictable ways due to conditions that are present in the insurance market. These conditions are often referred to as the underwriting cycle. The underwriting cycle contributes an artificial volatility to underwriting results that lies outside the statistical realm of insurance risk. The measures insurance managers take to deal with the statistical realm of insurance risk, i.e. reinsurance and diversification, are different than those measures they take to deal with the underwriting cycle.

We dealt with these difficulties by first using paid, rather than incurred, losses and estimating the ultimate incurred losses with industrywide paid loss development factors. Next we attempted to smooth out differences in loss ratios that we deemed "predictable." Appendix A in the Meyers *et. al.* paper referenced above describes this process in greater detail.

After making the above adjustments, two other difficulties should be discussed. First, the use of industrywide loss development factors removes the volatility that takes place after the report date of the loss. As such, we should expect the volatilities we measure to understate the ultimate volatility.

Second, Schedule P losses are reported net of reinsurance. In addition, policy limits are not reported. Rather than incorporate this information directly into our estimation, we did sensitivity tests of our model varying limits and reinsurance provisions over realistic scenarios.

Here I present results for commercial automobile liability insurance. I feel this is a good choice because: (1) it is a shorter tailed line than general liability and the underestimation of volatility will not be as great; (2) the use of reinsurance is not as great as it is in the general liability lines of insurance; and (3) commercial auto is not as prone to catastrophes as the property lines of insurance.

5.1 Standard Deviation of Loss Ratios vs. Size of Insurer

As illustrated in Figure 5, the collective risk model predicts that the standard deviation of insurer loss ratios should decrease as the size of the insurer increases. In Figure 10 we can see that this prediction is consistent with the observed standard deviations calculated from the Schedule P data described above. In this figure we plotted the empirical standard deviation of 55 commercial auto insurers against the average (over the 10 years of reported data) expected loss for the insurer[2].

Figure 10 also includes the standard deviations predicted by the collective risk model. The series denoted by "LowLim" used claim severity distribution parameters taken from a countrywide ISO claim severity distributions evaluated at the $300,000 occurrence limit. In this series I set $c = 0.007, g = 0.0005$ and $b = 0$. See Section 6 below for my commentary on selecting b and g.

Now we (at ISO) know from data reported to us that, depending on the subline (e.g. light and medium trucks or long-haul trucks), typically 65% to 90% of all commercial auto insurance policies are written at the $1 million policy limit. But since I also believed that the Schedule P data understates the true volatility of the loss ratios, I selected the $300,000 policy limit for the test.

For the sake of comparison, the series "HiLim" represents a judgmental adjustment that one might use to account for problems with the Schedule P data. I used claim severity distribution parameters taken from a countrywide ISO claim severity distributions evaluated at the $1,000,000 occurrence limit. In this series I set $c = 0.010, g = 0.0010$ and $b = 0$.

Figure 11 provides a comparable plot of loss ratios simulated from a collective risk model using the same parameters I used for the "LowLim" series.

The two plots both suggest that the Schedule P data is well represented by the collective risk mode — for an individual line of insurance.

[2] Since the expected loss varies by each observation of annual losses, the annual loss ratios are not identically distributed according to the collective risk model. I don't think this is a serious problem here since the volume of business is fairly consistent from year to year.

Figure 10

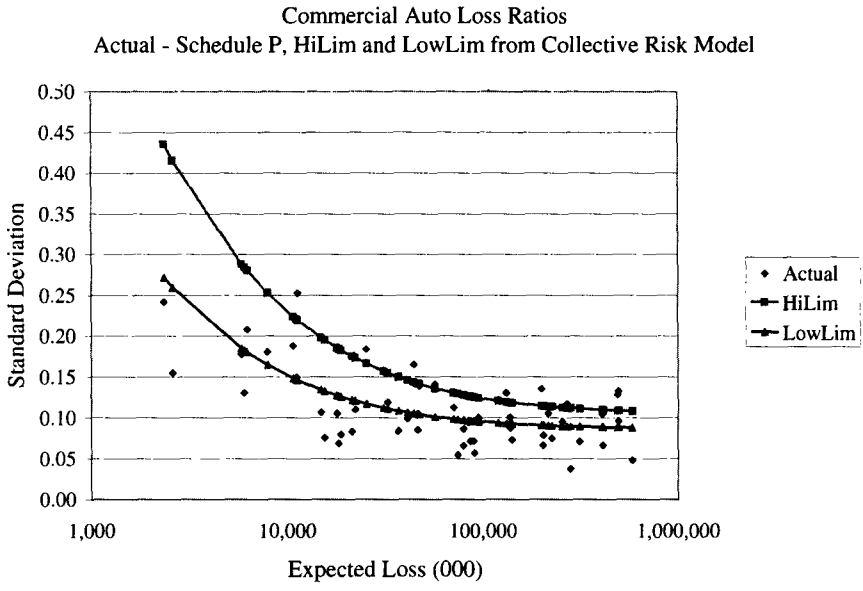
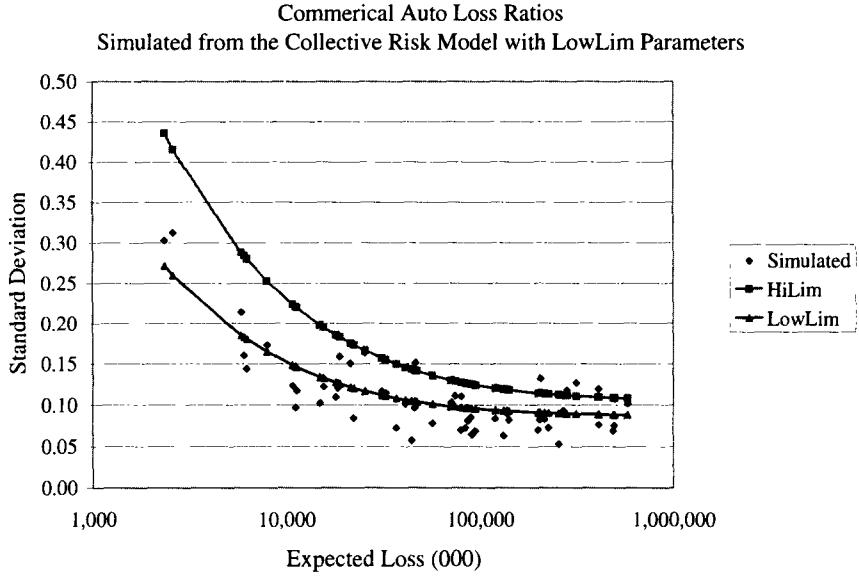


Figure11



5.2 Coefficients of Correlation vs. the Size of the Insurer

As Figure 7 illustrates, a second prediction of the collective risk model is that the coefficients of correlation will increase with the size of the insurer. In Figure 12 below, we plotted the empirical coefficient of correlation between commercial auto and personal auto for 38 insurers of both lines, against the average (over 10 years of experience reported for the two lines of insurance) expected loss. A comparable plot based on simulated data from the model underlying the URM is in Figure 13[3].

We observe that the coefficient of correlation is a very volatile statistic for both the empirical data and the simulated data which has a built-in assumption consistent with our hypothesis. This serves to illustrate the difficulty in measuring the effect of correlation in insurance data.

To provide a deeper analysis of the correlation problem I will make the assumption that the common shock random variables α and β operate on all insurers simultaneously. For random loss ratios R_1 and R_2:

$$E\left[(R_1-1)\cdot(R_2-1)\right]=\frac{Cov\left[X_1,X_2\right]}{\lambda_1\cdot\mu_1\cdot\lambda_2\cdot\mu_2}=b+g+b\cdot g; \tag{17}$$

which I derived from Equation 16.

Now we have already established that the standard deviation of loss ratios decreases with the size of the insurer. Thus the denominator of:

$$\rho\left[R_1,R_2\right]=\frac{E\left[(R_1-1)\cdot(R_2-1)\right]}{Std\left[R_1\right]\cdot Std\left[R_2\right]}$$

should decrease. If we can demonstrate with the Schedule P data, that the numerator does not also decrease, we can conclude that the prediction that coefficients of correlation will increase is consistent with the Schedule P data. It is to this we now turn.

[3] It may seem odd that the predicted correlation curve is not smooth. It is not smooth because the horizontal axis is the average of the commercial auto and the personal auto expected loss, while the actual split between the two expected losses varies significantly between insurers.

Figure 12

Commercial Auto and Personal Auto Loss Ratios
Actual – Schedule P, Predicted – Collective Risk Model

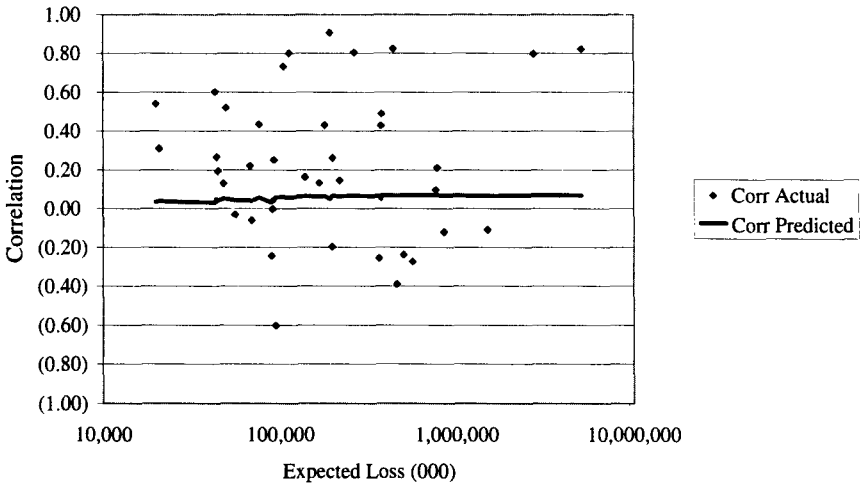
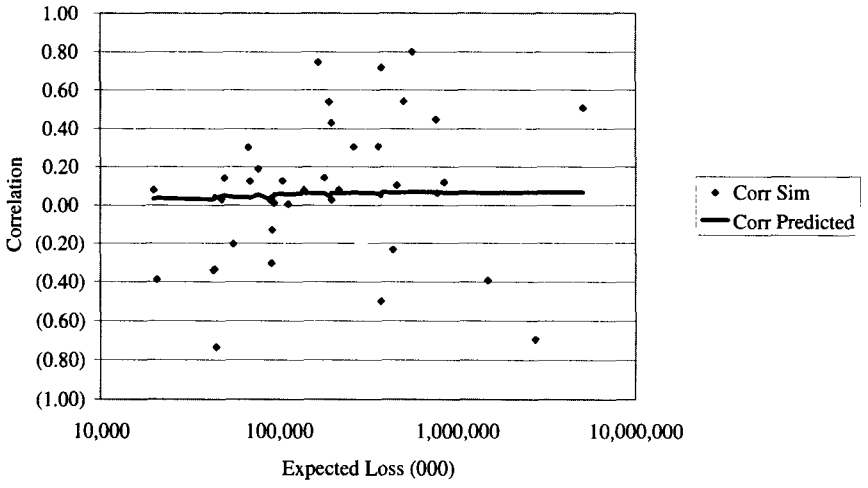


Figure 13

Commercial Auto and Personal Auto Loss Ratios
Simulated and Predicted from the Collective Risk Model



The data used in the test that $(R_1-1)\cdot(R_2-1)$ was independent of insurer size consisted of all possible pairs (15,790 in all) of r_1 and r_2, and the associated expected losses, taken from the same year and different insurers. I fit a line[4] to the ordered pairs

$$(\textit{Average Size of the Insurer}, (r_1-1)\cdot(r_2-1))$$

and obtained a slope of $+1.95\times10^{-10}$. This slightly positive slope means that an increasing coefficient of correlation is consistent with the Schedule P data.

Equation 17 also provides us with a way to estimate the quantity $b + g + b\,g$. One simply has to calculate the weighted average of the 15,970 products of $(r_1-1)\,(r_2-1)$, 0.00054. Since the 15,790 observations are not independent, the usual tests of statistical significance do not apply. To test the statistical significance of this result, I simulated 200 weighted averages using the "LowLim" parameters (except that $b = g = 0$) with the result that the highest weighted average was 0.000318. Thus we can reject the hypothesis that $b + g + b\,g = 0$.

I did one final simulation with the "LowLim" parameters (except that b = 0 and $g = 0.00054$) and calculated 200 slopes, with the result that the slope of 1.95×10^{-10} was just below the 49th highest. Thus this slope would not be unusual if the collective risk model is the correct model.

6. The Role of Judgment in Selecting Final Parameters

Historically, most actuaries have resorted to judgment in the quantification of correlation. This chapter was written in the hope of supplying some objectivity to this quantification. My employer, Insurance Services Office (ISO), has worked on quantifying this correlation. We have conducted analyses similar to the one described above for several lines of business using both Schedule P data and individual insurer data reported to ISO. In the end, no data set is perfect for the job and we end up making some judgments. Here are some of the considerations we made in selecting our final models. Comments are always welcome.

[4] I used a weighted least squares fit, using the inverse of the product of the predicted standard deviations of the loss ratio as the weights. This gives the higher volume, and hence more stable, observations more weight.

- We have reason to believe that the data we observe understates the ultimate variability since there are some claims that have yet to be settled. As a result we judgmentally increased the c, b and g parameters in the final model.

- Since the estimation procedure described provides an estimate of $b + g + b\,g$, it is impossible to distinguish between the claim frequency common shocks, as quantified by g, and the claim severity common shocks as quantified by b. A lot of work has been done with claim severity and claim frequency trend and one can look to uncertainties is these trends when selecting the final parameters.

- While one might argue that the distinction between claim frequency common shocks and claim severity common shocks is unimportant, the way we apply them does make a difference. For claim frequency we group the various lines of insurance judgmentally, with some support from the data. For example, the same common shock for claim frequency applies to personal and commercial auto, but different common shocks apply to the commercial liability lines. We apply claim severity shocks across all lines. Meyers, Klinker and Lalonde [2003] describe this model more fully.

Accounting data such as Schedule P may not be the best source for such analyses, but if we cannot see the effect of correlation in the accounting data, I would ask, do we need to worry about correlation? I believe that the analysis in this chapter demonstrates that we do need to consider correlation between lines of insurance.

7. Acknowledgements

This chapter is largely an exposition of work that appeared in a series of prior papers that I will now describe. A significant advance in the correlation literature was made by Shaun Wang [4] with the publication of his work on a project that was sponsored by the CAS. It is in this paper that I first heard the term "common shock model." I rather quickly followed up with two related papers. In Meyers[1], I originally developed the model that is described in Section 4 of this paper, and in Meyers[2] I developed methodology to parameterize the model with data that was "theoretically" available. A few years later we —

Meyers, Klinker and Lalonde [3] — followed up with another methodology to parameterize the model with data that was actually available. The original version of this methodology is described in Appendix A and Fredrick Klinker deserves the lion's share of the credit for developing it. I would described Section 5 as a minor improvement to this methodology.

References

[1] Meyers, Glenn G., 1999, "A Discussion of Aggregation of Correlated Risk Portfolios: Models & Algorithms, by Shaun S. Wang," 1999, *PCAS,* Vol. LXXXVI, 781-805.
http://www.casact.org/pubs/proceed/proceed99/99705.pdf

[2] Meyers, Glenn G., "Estimating Between Line Correlations Generated by Parameter Uncertainty," *CAS Forum*, Summer 1999, 16-82.
http://www.casact.org/pubs/forum/99sforum/99sf197.pdf

[3] Meyers, Glenn G., Fredrick L. Klinker and David A. Lalonde, "The Aggregation and Correlation of Insurance Risk," *CAS Forum*, Summer 2003, 16-82.
http://www.casact.org/pubs/forum/03sforum/03sf015.pdf

[4] Wang, Shaun S., "Aggregation of Correlated Risk Portfolios: Models & Algorithms," *PCAS*, 1998, Vol. LXXXV, 848-939.
http://www.casact.org/pubs/proceed/proceed98/980848.pdf

Biography

Glenn Meyers is the Chief Actuary for ISO Innovative Analytics. He holds a bachelor's degree in mathematics and physics from Alma College in Alma, Mich., a master's degree in mathematics from Oakland University, and a Ph.D. in mathematics from the State University of New York at Albany. Glenn is a Fellow of the Casualty Actuarial Society and a member of the American Academy of Actuaries. Before joining ISO in 1988, Glenn worked at CNA Insurance Companies and the University of Iowa.

Glenn's current responsibilities at ISO include the development of scoring products. Prior responsibilities have included working on ISO Capital Management products, increased limits and catastrophe ratemaking, ISO's, and Property Size-of-Loss Database (PSOLD), ISO's model for commercial property size of loss distributions.

Glenn's work has been published in *Proceedings of the Casualty Actuarial Society (CAS).* He is a three-time winner of the Woodward-Fondiller Prize, a two-time winner of the Dorweiller Prize and a winner of the Dynamic Financial Analysis Prize. He is a frequent speaker at CAS meetings and seminars.

His service to the CAS includes membership on various education and research committees. He currently serves on the International Actuarial Association Solvency Committee and the CAS Board of Directors.

The Report of the Research Working Party on Correlations and Dependencies Among All Risk Sources

Part 4
Serial Correlations of Interest and Inflation Rates

Hans E. Waszink AAG, MSc.

Abstract

This chapter discusses an approach to model the value of an outstanding, discounted liability under the impact of uncertain interest and inflation rates. Interest and inflation rates are modeled separately as time series to take into account autocorrelation. Subsequently, the dependence between interest and inflation is modeled using copulas. The goodness of fit of some copulas can be evaluated on the basis of historic data using a quantile plot. This is done for the Gumbel, Clayton and Independent copulas. The Gumbel copula, which gives the best fit, is then compared with the Normal copula to show that the two copulas are very similar with the parameters chosen. The distribution of the required reserve is shown under four different copula assumptions: comonotonicity, which represent the best case, countermonotonicity which represents the worst case, and the Gumbel and Normal copulas which represent more realistic scenarios. The choice of copula has considerable impact on the higher percentiles of the required reserve, and the adopted approach is effective in selecting a suitable copula.

1. INTRODUCTION

In this chapter the following are investigated:

1. Correlations between the same variable, i.e. interest or inflation, at different points in time (autocorrelation).
2. Correlations between interest and inflation over an extended period of time.
3. Impact of these correlations on the present value of a discounted and inflated liability.

The effect of both types of correlations is demonstrated in a case study investigating the effect of interest and inflation rates fluctuations on outstanding claims liabilities. Interest and inflation rates are modeled as time series. Time series models are commonly used for variables of which observations are available sequentially in time, and consecutive observations are dependent. Both these properties typically apply to interest as well as inflation rates.

A simple example of a time series is an autoregressive process of order 1 (AR(1)) which is given below:

$$X(t) = a + bX(t\text{-}1) + \varepsilon(t), \quad t = 1,\ldots,T$$

with

$X(.)$: array of stochastic variables, $t = 0,1,\ldots,T$, $X(0)$ a given constant.
$\varepsilon(t)$: random error within period $(t\text{-}1,t)$, with $N(0,\sigma)$ distribution.
a,b: model parameters.

It can be shown that this structure defines a correlation structure between all $X(t)$, with correlations depending on b and σ and the elapsed time between observations. More complex time series models are often required to adequately capture specific characteristics such as cyclicality or heteroskedasticity.

2. OUTSTANDING LIABILITY UNDER UNCERTAIN INTEREST AND INFLATION RATES

We consider the value of an outstanding claims reserve as the present value of inflated and discounted future claim payments. Interest and inflation rates are modeled as random variables. As a starting point, we use uninflated projections of future claim payments in each future payment period. These can be derived from triangular reserving methods which include an explicit inflationary effect.

Define:

C(t): Uninflated, fixed and given cashflow projection at time *t*.
Inf(t): Inflation rate in period *(t,t+1)*, *t = 0,1,2,...*
Int(t): Interest rate in period *(t, t+1)*, *t = 0,1,2,...*

Ac(t): Actual cashflow at time *t*.

Ac(t) is equal to:

$$Ac(t) = C(t) \times \prod_{\tau=0}^{t-1} [1 + Inf(\tau)], \quad t = 1,2,3,\ldots$$

For simplicity it is assumed that *Ac(t)* is the product of the cashflow projection *C(t)*, which is fixed and given, and future inflation rates only. Therefore the only uncertain factor in actual future cashflows is future inflation which can represent general inflation, superimposed inflation or a line-specific inflation. In this study we have used medical inflation, a line-specific inflation impacting on health insurance related liabilities.

The inflation rates represent a component of systematic risk in the cash flow projection, i.e. they affect all individual claims simultaneously and to the same extent. To relax the assumption that inflation is the only uncertain factor affecting future cashflows, additional components of unsystematic risk can be added without any difficulty, however these are excluded here.

Df(t): Discount factor in period *(t,t+1)*, *t=0,1,2,...*:

$$Df(t) = \frac{1}{1 + Int(t)}$$

RR(t): Required reserve at time *t* , *t=0,1,2,...*:

$$RR(t) = \sum_{s>t} [Ac(s) \times \prod_{\tau=0}^{s-t-1} Df(t+\tau)]$$

The required reserve is the total of all actual future cashflows discounted at actual future interest rates. Obviously *RR(t)* is not known in advance as it is a function of *C(t)*, *Inf(t)* and *Int(t)* with future interest and inflation rates unknown.

The distribution of the *RR(t)* is a function of the marginal distributions of the interest and inflation rates after time *t* and the dependencies between interest rates in different periods, the dependence between inflation rates in different periods, and the dependence between inflation and interest rates in the same period and in different periods.

3. MODELLING THE DISTRIBUTION OF INTEREST AND INFLATION RATES AND THEIR MUTUAL DEPENDENCE

3.1 Interest rates

A discrete version of the CIR[1]-model for a single interest rate is used. A single interest rate is used for simplicity, although the CIR-model allows for the generation of the entire yield curve with full dependence between different maturities. Different yield curve structures can be generated using various other interest rate models of a similar time series structure.

The discrete CIR-model is a time-series model of the following form:

$$Int(t) = max\{0, Int(t-1) + a[b - Int(t-1)] + \sqrt{Int(t-1)}\varepsilon_{int}(t)\}$$

with

$Int(t)$:	the interest rate in the period $(t,t+1)$
a:	the average speed of reversion to the long term mean interest rate;
b:	the long term mean interest rate.
$\varepsilon_{int}(t)$:	random deviation in period $(t,t+1)$. The $\varepsilon_{int}(t)$ are mutually independent with marginal distributions $N(0,\sigma^2)$.

The model has several desirable properties such as:

- Interest rates are mean reverting;
- Interest rates are non-negative.
- Interest rates are heteroskedastic, i.e. variance increases with mean.
- Interest rates at adjacent points in time are correlated.
- Confidence intervals widen for interest rates projections further into the future.

For the parameterization of the time series, we have used 3 year interest rates on US government securities which are shown in appendix I. The estimated parameters are shown in appendix II, simulated autocorrelations of interest rates are shown in appendix III.

[1] Cox Ingersoll Ross, see Kaufmann (2001)

3.2 Inflation rates

For inflation rates a second order autoregressive process (AR(2)) is used:

$$Inf(t) = c_0 + c_1\, Inf(t\text{-}1) + c_2\, Inf(t\text{-}2) + \varepsilon_{inf}(t)$$

with

c_0, c_1, c_2: model parameters.
$\varepsilon_{inf}(t)$: random deviations in period $(t,t+1)$.

The $\varepsilon_{inf}(t)$ are mutually independent with identical marginal distributions $N(0,\sigma^2)$.

Some properties of the AR(2) model are:

- If $c_2 < 0$, inflation rates may exhibit cyclicality.
- Observations at adjacent points in time are correlated.
- Confidence intervals widen for projections further into the future.

For the parameterization of the time series, we have used US medical care index figures provided by the Bureau of Labor Statistics, which are shown in appendix I. The estimated parameters are shown in appendix II, simulated autocorrelations of inflation rates are shown in appendix III.

The (analytically determined) autocorrelations between the inflation rate in time period 1 and all other periods, derived from the time series parameterization, are shown below:

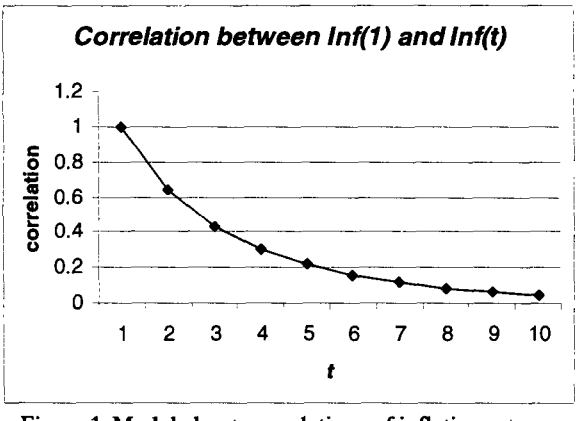


Figure 1: Modeled autocorrelations of inflation rates

As the parameter c_2 is very close to, and not significantly different from 0, there is no cyclical pattern in the correlation structure and the process is virtually identical to an AR(1) process.

3.3 Dependence between inflation and interest rates

The dependence relation between interest and inflation rates in the same period is determined by, on the one hand, the structure of the time series model for both interest and inflation rates, and on the other hand by the dependence between the random errors $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ in the same period.

Both time series as well as the dependence between them are parameterized on the basis of actual historic data shown in appendix I. It can be expected that there is a dependency between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ as changes in both inflation and interest rates are driven by the same or related macro-economic factors. Various types of dependence relationships, i.e. copulas, can be used to model the dependency between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$. We assume that the dependence relation is the same for all values of t, hence does not change over time.

Given that both error terms are assumed to follow a Normal distribution, the simplest form of dependency is the linear correlation which corresponds to the Normal copula. However the Normal copula does not always capture dependencies in the tail of the distributions appropriately[2] hence the Gumbel and Clayton copulas are also investigated

As $\varepsilon_{int}(t)$ and $\varepsilon_{int}(t')$ are independent if $t \neq t'$, so are $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t')$. Thus the choice of the time series models for interest and inflation rates together with the copula representing the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ fully define the joint distribution of interest and inflation rates. As $RR(t)$ is fully determined by the deterministic uninflated cashflows $C(t)$ in combination with interest and inflation rates during the projection period, the distribution of all $RR(t)$ is fully defined by the joint distribution of inflation and interest rates and $C(t)$. The distribution of $RR(t)$ is derived by means of simulation.

For the uninflated cashflow projection $C(t)$ we set $C(t) = 1$ for $t = 1,2,\ldots,10$ and 0 otherwise. For the choice of the copula defining the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$, several alternative scenarios are investigated:

1. $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ are comonotonic, i.e. the dependence between the two is maximum. As both $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ are Normal random variables, the linear correlation between them is 100%. This is the best case scenario for the insurer with respect to the dependence between the two error terms. The underlying assumption is that random deviations of interest rates are fully correlated with random deviations of inflation rates, hence unexpected increases in inflation are always accompanied by unexpected increases in interest rates. As increases in inflation rates lead to increases in $RR(.)$ whereas increases in interest rates lead to a decreases of $RR(.)$, the comonotonic assumption implies that there always is a compensating effect of the two random errors on the liability for the insurer.

[2] See Embrechts (2001)

Therefore this scenario represents a best case for the insurer with respect to the occurrence of extremely high values of *RR(t)*.

2. In the second scenario, the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ is assumed to be 'countermonotonic'[3], meaning unexpected increases in inflation rates are always accompanied by unexpected decreases in interest rates. Contrary to the first alternative, this scenario represents the worst case with respect to the occurrence of extremely high values of *RR(.)*, as the effects of unexpected inflation in any particular period are aggravated by lower interest earnings in the same period.

3. In the third scenario, the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ is parameterized on the basis of historic observations. Historic observations of the error terms are obtained by substituting observed historic interest/inflation rates in the time series equations for *Int(t)* and *Inf(t)*. Sufficient credible historic data needs to be available to justify a choice and parameterization of a copula in this way.

   The copula chosen here is the Gumbel copula, with parameter $\alpha = 1.4$. Appendix IV shows the fit of the Gumbel and Clayton copulas, on the basis of which the Gumbel copula is the preferred choice. Appendix V shows correlations between inflation and interest rates under the Gumbel copula.

4. In the fourth alternative, the dependence between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ is modeled as a multivariate Normal distribution, with the dependence between the two random variables fully characterized by their linear correlation coefficient.

The simulated results of each of the four methods are shown in figure 2 below[4], with BC (Best Case) , WC (Worst Case), Gumbel and Normal depicting *RR(0)* in alternatives 1-4 respectively. As the graphs of alternative 3 and 4 seem to overlap completely, the right tail is shown in more detail in figure 3.

[3] Characterization of comonotonicity and countermonotonicity can be found in Denuit (2003)
[4] Results were generated using IGLOO software.

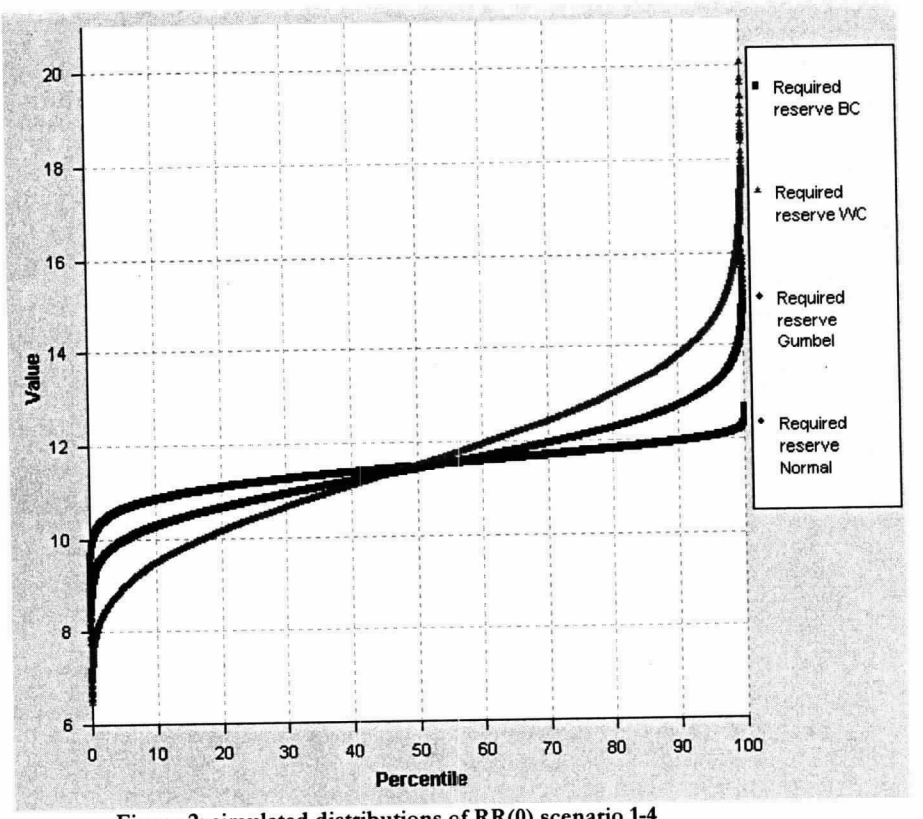


Figure 2: simulated distributions of RR(0) scenario 1-4

In alternative 4, a linear correlation between $\varepsilon_{int}(t)$ and $\varepsilon_{inf}(t)$ of 0.44 is applied. This is the historically observed correlation between the residuals. In alternative 3, the Gumbel copula is parameterized using the algorithm described in Valdez (1998). The Gumbel copula in this case gives rise to the same linear correlation of 0.44 as the Normal copula.

The right tail of the distributions resulting from the Normal and Gumbel copulas are shown below.

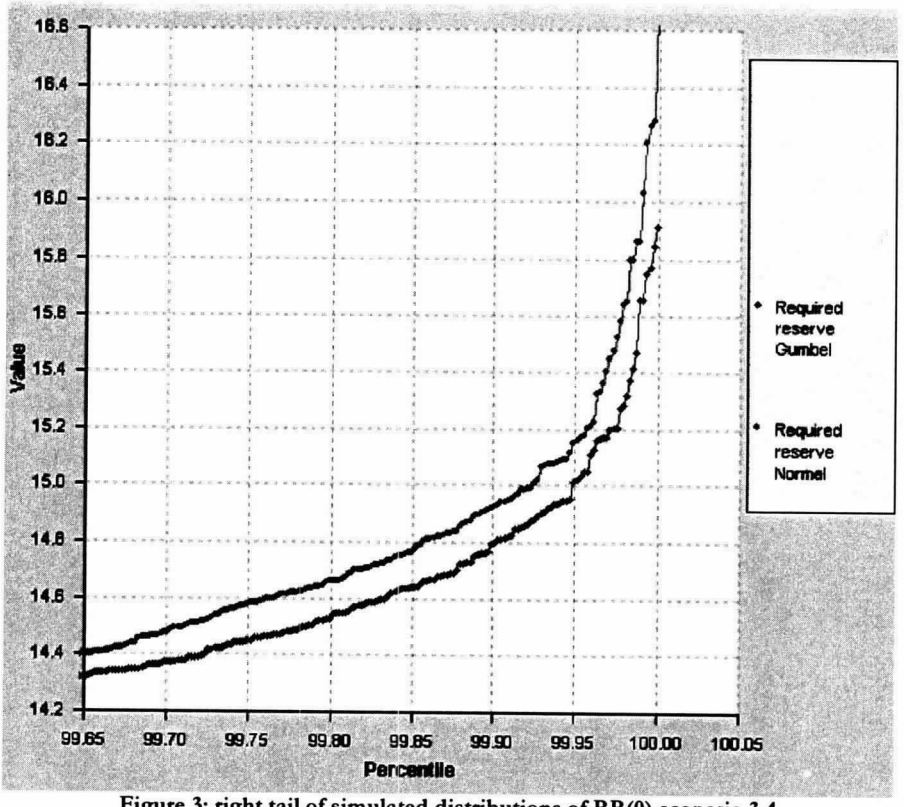


Figure 3: right tail of simulated distributions of RR(0) scenario 3,4

A marginal difference between the two reserves can be observed. The fact that the difference between the distributions under the two copulas is so small suggests that the two copulas generate very similar dependence structures. This is confirmed by the simulated rank scatter plots of the two copulas shown in figure 4 below.

A rank scatter plot shows simulated pairs of uniform random variables under a given dependence structure between the two variables. When realizations are spread evenly across the square, this indicates a low degree of dependence. A high degree of dependence is indicated by concentrations of points in certain parts of the square. For example tail correlation leads to a higher concentration of realizations in the corners of the square.

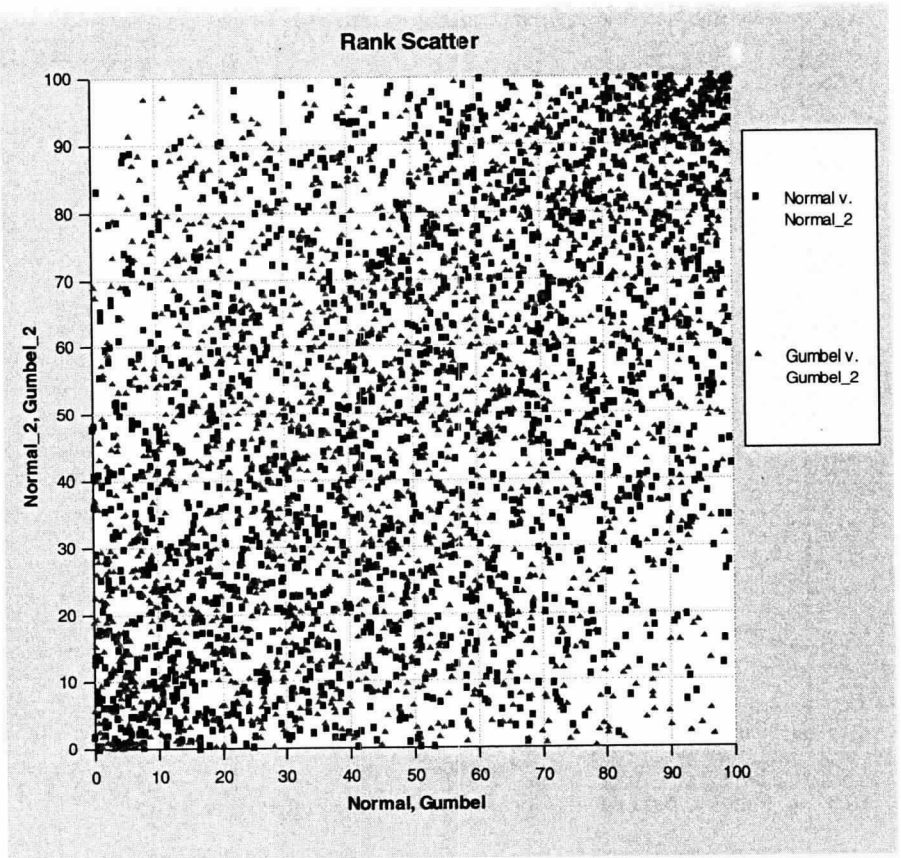


figure 4: rank scatter plot of simulated Normal and Gumbel copulas.

The two scatter plots shown in figure 4 show very similar patterns, both with a slightly lower density of points towards the upper left hand and lower right hand corner, and higher towards the other two corners. This indicates the dependence structures simulated by the two copulas are very similar.

4. RESULTS AND DISCUSSION

Dependence between interest and inflation rates has a considerable impact on the distribution of the required reserve. The parameterization of the copulas in alternatives 3 and 4 require a sufficiently large history of reliable data, and one needs to assume that the dependence structure does not change over time. The approach in alternative 2 however provides an upper bound with regard to the dependence between the random errors of the two time series. Hence alternative 2 may be preferable if a prudent approach is sought and historic data are not considered sufficiently reliable.

The difference between the Normal and the Gumbel copula and the impact on the distribution of the required reserve is minimal. The Gumbel copula gives a better fit to the data than the Clayton copula. A fit of the Normal copula can not be shown in the same way as it does not belong to the family of so-called 'Archimedean' copulas, see Valdez (1998).

Parameterization of an interest rate model based on historically observed rates may lead to results which are inconsistent with current market rates. Also, the use of a one-factor model can be regarded as too simplistic. However additional prudence can be built in by reducing the long term mean parameter b for example on the basis of projections by an economic forecasting bureau.

The long term average interest rate parameter b of 6.7% appears high in the current environment, and leads to a continuous upward trend in the projected future interest rate. Reducing b to 3% leads to an increase of the liability by about 6% across the distribution. Alternatively the CIR model can be parameterized on the basis of the current yield curve but this would not allow for the measurement of the correlation with inflation rates. Such measurement requires the availability of simultaneous observations of interest and inflation over an extended historic period.

Supplementary Material

Two spreadsheets are attached. One contains the parameterization of the time series, the other the parameterization of the copulas and the quantile plot. They can be viewed in the Publications section of the CAS Web Site (**www.casact.org**) under 2006 Winter Forum.

Appendix I

1. Medical Inflation rates:

Area: *U.S. city*
Item: *Medical care*
Source:http://data.bls.gov/servlet/SurveyOutputServlet?data_tool=latest_numbers&series_id=CUUR0000SAM&output_view=pct_1mth

2. Interest rates:

```
Rate of interest in money and capital markets
Federal Reserve System
Long-term or capital market
Government securities
Federal
Constant maturity
Three-year
Not seasonally adjusted
```
Twelve months ending December

Source: http://www.federalreserve.gov/releases/h15/data.htm#fn12

Year	Medical Inflation rate (%)	Interest rate (%)
1962	2.02	3.47
1963	2.42	3.67
1964	2.02	4.03
1965	2.83	4.22
1966	6.69	5.23
1967	6.38	5.03
1968	6.27	5.68
1969	6.05	7.02
1970	7.44	7.29
1971	4.80	5.66
1972	3.45	5.72
1973	5.52	6.96
1974	12.56	7.84
1975	9.70	7.5
1976	10.14	6.77
1977	8.73	6.68
1978	8.73	8.29
1979	10.25	9.70
1980	10.03	11.51
1981	12.45	14.46
1982	11.02	12.93
1983	6.38	10.45
1984	6.48	11.92
1985	6.48	9.64
1986	7.87	7.06
1987	5.75	7.68
1988	6.91	8.26
1989	8.73	8.55
1990	9.70	8.26
1991	7.97	6.82
1992	6.48	5.30
1993	5.22	4.44
1994	4.80	6.27
1995	3.97	6.25
1996	3.04	5.99
1997	3.04	6.10
1998	3.45	5.14
1999	3.76	5.49
2000	4.18	6.22
2001	4.70	4.09
2002	4.90	3.10
2003	3.66	2.10
2004	4.28	2.78

Appendix II Parameterization of the time series

Interest rates

Parameters of the CIR model are:

a	0.085823
b	6.684528
σ	0.450159

Parameters are estimated by minimizing the sum of squared residuals on the basis of the data in appendix I.

Inflation rates

parameter	Estimate	Standard error
c_0	1.3470	
c_1	0.8441	0.13
c_2	- 0.0806	0.13
σ'	1.7151	

N.B. as c_2 is very small and not significantly different from 0, an AR(1) process (with $c_2 = 0$) will produce very similar results.

Parameter estimates are derived as[5]:

$$c_1 = \frac{r_1(1-r_2)}{1-r_1^2}$$

$$c_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

with r_1 and r_2 estimates of the first and second order autocorrelation:

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}, \; k = 1,2 \text{ and } n \text{ the number of observations } x.$$

c_0 is estimated such that the mean inflation rate is stationary and equal to the historical average:

$$c_0 \quad c_0 = \frac{1}{n}\sum_{t=1}^{n} Inf(t)(1-c_1-c_2).$$

[5] See Box (1994)

Appendix III Simulated autocorrelations of interest and inflation rates

	Interest[1]	Interest[2]	Interest[3]	Interest[4]	Interest[5]	Interest[6]	Interest[7]	Interest[8]	Interest[9]	Interest[10]
Interest [1]	1.000	0.660	0.509	0.416	0.352	0.306	0.267	0.230	0.206	0.186
Interest [2]	0.660	1.000	0.766	0.625	0.529	0.457	0.399	0.347	0.310	0.277
Interest [3]	0.509	0.766	1.000	0.815	0.688	0.593	0.519	0.453	0.400	0.357
Interest [4]	0.416	0.625	0.815	1.000	0.842	0.724	0.633	0.556	0.490	0.435
Interest [5]	0.352	0.529	0.688	0.842	1.000	0.857	0.749	0.657	0.579	0.514
Interest [6]	0.306	0.457	0.593	0.724	0.857	1.000	0.868	0.760	0.670	0.594
Interest [7]	0.267	0.399	0.519	0.633	0.749	0.868	1.000	0.876	0.773	0.685
Interest [8]	0.230	0.347	0.453	0.556	0.657	0.760	0.876	1.000	0.881	0.780
Interest [9]	0.206	0.310	0.400	0.490	0.579	0.670	0.773	0.881	1.000	0.884
Interest [10]	0.186	0.277	0.357	0.435	0.514	0.594	0.685	0.780	0.884	1.000

	Inflation[1]	Inflation[2]	Inflation[3]	Inflation[4]	Inflation[5]	Inflation[6]	Inflation[7]	Inflation[8]	Inflation[9]	Inflation[10]
Inflation [1]	1.000	0.629	0.420	0.289	0.206	0.151	0.108	0.071	0.055	0.043
Inflation [2]	0.629	1.000	0.709	0.493	0.351	0.254	0.185	0.128	0.098	0.072
Inflation [3]	0.420	0.709	1.000	0.739	0.528	0.383	0.280	0.197	0.144	0.105
Inflation [4]	0.289	0.493	0.739	1.000	0.754	0.547	0.398	0.290	0.209	0.151
Inflation [5]	0.206	0.351	0.528	0.754	1.000	0.760	0.555	0.402	0.291	0.211
Inflation [6]	0.151	0.254	0.383	0.547	0.760	1.000	0.760	0.550	0.401	0.291
Inflation [7]	0.108	0.185	0.280	0.398	0.555	0.760	1.000	0.765	0.561	0.410
Inflation [8]	0.071	0.128	0.197	0.290	0.402	0.550	0.765	1.000	0.767	0.560
Inflation [9]	0.055	0.098	0.144	0.209	0.291	0.401	0.561	0.767	1.000	0.767
Inflation [10]	0.043	0.072	0.105	0.151	0.211	0.291	0.410	0.560	0.767	1.000

Appendix IV Quantile plot for Independent, Gumbel and Clayton copulas

A quantile plot (also know as Q-Q plot) can be used to inspect the goodness of fit of Archimedean copulas, and is derived as follows.

Archimedean copulas are of the form:

$C_\varphi(u,v) = \varphi^{-1}(\varphi(u)+ \varphi(v))$ with $0 < u,v \leq 1$ and φ a convex decreasing function with domain (0,1].

For two random variables X and Y with dependence defined by the Archimedean copula C_φ, it can be shown that the random Variable $Z = C_\varphi(F_X(X), F_Y(Y))$ has the following distribution function:

$$F_Z(z) = z - \varphi(z)/\varphi'(z).$$

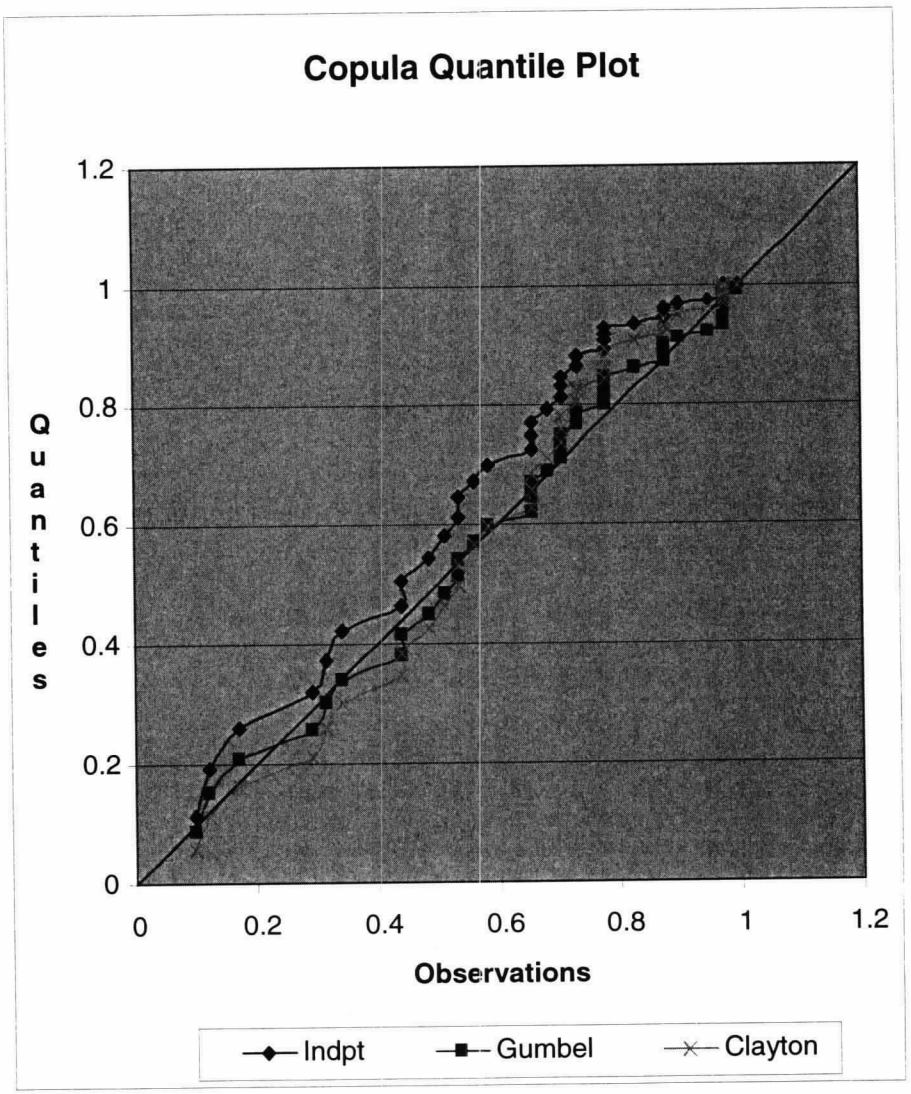
This implies that, assuming the dependence between X and Y is described by a given Archimedean copula C_φ, the variable Z should follow the distribution function given above. Hence comparing n ordered (pseudo)-observations of Z with the percentiles of the distribution function of Z in a Q-Q plot allows for inspection of the goodness of fit of the assumed distribution of Z hence of the copula function C_φ. The observations of Z are derived from the observations of X and Y and the relation $Z = C_\varphi(F_X(X), F_Y(Y))$. The process of constructing the quantile plot and the underlying theory can be found in Valdez (1998).

The interpretation of the Q-Q plot is no different than the Q-Q plot for any other single random variable. The closer observations are to the corresponding percentiles of the theoretical distribution, the better the fit of the distribution. Hence a Q-Q plot showing a pattern close to the straight line through the origin and (1,1) indicates a good fit of the distribution.

The copulas used are:

Gumbel:	$C(u,v) = \exp\{-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{1/\alpha}\}$,	$\varphi(u) = (-\ln u)^\alpha$
Clayton:	$C(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$,	$\varphi(u) = u^{-\alpha} - 1$
Independent:	$C(u,v) = uv$,	$\varphi(u) = -\ln u$.

[2]



Parameterization:

Gumbel (α)	1.41716
Clayton (α)	0.62773

Appendix V Linear correlations of interest and inflation rates

Simulated linear correlations between interest and inflation rates in alternative 3 under the Gumbel copula are as follows:

	Interest[1]	Interest[2]	Interest[3]	Interest[4]	Interest[5]	Interest[6]	Interest[7]	Interest[8]	Interest[9]	Interest[10]
Inflation [1]	0.436	0.289	0.214	0.176	0.148	0.124	0.105	0.094	0.083	0.070
Inflation [2]	0.277	0.430	0.322	0.259	0.219	0.187	0.159	0.140	0.126	0.111
Inflation [3]	0.188	0.314	0.431	0.346	0.290	0.248	0.214	0.187	0.169	0.151
Inflation [4]	0.131	0.220	0.324	0.423	0.356	0.305	0.263	0.229	0.204	0.183
Inflation [5]	0.094	0.158	0.236	0.324	0.420	0.360	0.312	0.274	0.242	0.220
Inflation [6]	0.067	0.114	0.169	0.234	0.320	0.417	0.362	0.320	0.284	0.254
Inflation [7]	0.052	0.086	0.127	0.173	0.238	0.323	0.414	0.366	0.323	0.288
Inflation [8]	0.035	0.059	0.089	0.121	0.169	0.236	0.319	0.409	0.360	0.319
Inflation [9]	0.027	0.045	0.068	0.090	0.127	0.175	0.237	0.316	0.403	0.358
Inflation [10]	0.022	0.037	0.052	0.069	0.096	0.131	0.177	0.235	0.313	0.398

4. REFERENCES

[1] Valdez, Emiliano A, and Edward W. Frees "Understanding Relationships Using Copulas" *North American Actuarial Journal*, 1998, 1-25.

[2] Box, George E, Gwilym M. Jenkins and Gregory C. Reinsel "Time Series Analysis" Prentice-Hall 1994.

[3] Nelsen, Roger B "An Introduction to Copulas" Springer 1999.

[4] Denuit, Michel, and Jan Dhaene "Simple Characterizations of Comonotonicity and Countermonotonicity by Extremal Correlations" Catholic University Leuven 2003. http://www.econ.kuleuven.ac.be/tew/academic/actuawet/pdfs/Corel_Denuit_Dhaene_BAB.pdf

[5] Wang, Shaun S "Aggregation of Correlated Risk Portfolios: Models and Algorithms" *PCAS* 1998, Vol. LXXXV, 848-939.

[6] Embrechts, Paul, Filip Lindskog, and Alexander McNeil "Modelling Dependence with Copulas and Applications to Risk Management" RiskLab September 2001.

[7] Kaufmann, Roger, Andreas Gadmer and Ralf Klett "Introduction to Dynamic Financial Analysis". Astin Bulletin, Vol. 31, No. 1 (May 2001) pages 213-249. http://www.casact.org/library/astin/vol31no1/213.pdf

[8] Cherubini, Umberto, Elisa Luciano and Walter Vecchiato "Copula Methods in Finance". Wiley Finance 2004.

Biography of Hans Waszink

Hans Waszink works for Lloyds TSB General Insurance in the United Kingdom, where he is responsible for capital adequacy, investment strategy and reinsurance. He holds a Master's degree in Mathematics from the University of Groningen in the Netherlands, and a Master's degree in Actuarial Science from the University of Amsterdam, the Netherlands. He is a full member of the Actuarial Society of The Netherlands. Prior to joining Lloyds TSB in 2003, Hans worked for ING Group headquarters as a general insurance actuary, and as an actuarial consultant for William M. Mercer.

Hans currently serves on the IAA (International Actuarial Association) Solvency Subcommittee and the ABI (Association of British Insurers) Non-life Capital Working Party. Hans was co-author of the IAA publication 'A Global Framework for Insurer Solvency Assessment' which was published in 2004. He is currently pursuing a part-time MBA at London Business School.

Risk Transfer Testing of Reinsurance Contracts: Analysis and Recommendations

CAS Research Working Party on Risk Transfer Testing

Abstract

This paper was prepared in response to a call from the American Academy of Actuaries Committee on Property and Liability Financial Reporting (COPLFR). The call requested ideas about how to define and test for risk transfer in short duration reinsurance contracts as required by FAS 113 and SSAP 62. These accounting standards require that a reinsurance contract must satisfy one of two conditions in order to qualify for reinsurance accounting treatment: 1) the contract must transfer "substantially all" of the underlying insurance risk, or failing that, 2) it must at least transfer "significant" insurance risk. The paper presents methods to test for both conditions, but the main focus is on testing for "significant" risk transfer. The shortcomings of the commonly used "10-10" test are discussed and two alternative testing frameworks are presented as significant improvements over "10-10". The first of these, which is presented in detail, is based on the expected reinsurer deficit (*ERD*). Conceptually, that approach is a refinement and generalization of "10-10" that addresses its major shortcomings. The second framework, based on the right tail deviation (*RTD*), is presented more briefly. It has certain desirable properties but at the cost of greater complexity.

Keywords: risk transfer testing, FAS 113, "10-10" test, downside risk, expected reinsurer deficit (*ERD*), right tail deviation (*RTD*), tail value at risk (*TVaR*), parameter uncertainty

1. INTRODUCTION

The purpose of this paper is to propose an improved framework for testing short-duration reinsurance contracts for risk transfer compliance with FAS 113. Under that accounting statement, reinsurance accounting is allowed only for those indemnity contracts that transfer insurance risk. The aim of the paper is to present a theoretically sound but practical approach to determining whether a contract meets the risk transfer requirements of FAS 113.

1.1 Context

The working party that prepared this paper was formed by the CAS to respond to a call by the American Academy of Actuaries Committee on Property and Liability Financial Reporting (COPLFR) for the submission of actuarially sound ideas about how to define and test for risk transfer in reinsurance transactions. The American Academy call arose out of the need for a constructive response from the actuarial profession following some widely publicized cases of alleged abuse of finite reinsurance and related accounting principles. Those cases have led to renewed scrutiny of reinsurance contracts to ascertain whether they comply with the existing accounting requirements and to a broader inquiry as to whether

FAS 113 goes far enough in specifying the manner in which contracts will be accounted for either as reinsurance or otherwise.

In a letter dated June 13, 2005, and addressed to members of the CAS, the chair of COPLFR framed the request as follows:

"Property/casualty actuaries interested in contributing suggestions...are asked to submit responses to one or more of the following questions:

1. What is an effective test for risk transfer? (Respondents are asked to focus on actuarial methodology and provide examples as appropriate.)
2. What criteria should be used to determine whether a reinsurance contract transfers significant risk to the reinsurer? (Respondents are asked to focus on decision criteria used to evaluate the results of the test described in question #1.)
3. What safe harbors, if any, should be established so that a full risk transfer analysis does not have to be completed for each and every reinsurance contract (i.e., in what instances is risk transfer "reasonably self-evident" and therefore cash flow testing is not necessary to demonstrate risk transfer)?
4. What are the advantages and disadvantages of the suggested approach versus other approaches commonly used?"

There is very little published actuarial literature on the subject. The only significant paper appears to be the one prepared in 2002 by the CAS Valuation, Finance, and Investments Committee entitled, "Accounting Rule Guidance Statement of Financial Accounting Standards No. 113—Considerations in Risk Transfer Testing"[1]. That paper provided an excellent summary of FAS 113 and the risk transfer testing methods that emerged in response (including the "10-10" test) as well as a discussion of a number of alternative methods. However, the paper was fairly muted in its criticism of "10-10", and it did not strongly advocate replacing it with an alternative.

In this paper we seek to respond to all four of the questions posed by COPLFR. The members of the working party believe the time has come to be explicit about the shortcomings of the "10-10" test that has come into common use and to advocate its replacement with a better framework. Accordingly, in this paper we include an extensive critique of the "10-10" test and describe two frameworks, one in detail and the other in

summary, that would be significant improvements over "10-10". We also identify methods for determining whether individual reinsurance contracts should be subject to detailed testing.

The frameworks described in the paper primarily address the issue of developing a more consistent and rigorous quantitative approach for the evaluation of risk transfer. As a result, the approaches described might reduce the potential for accounting mistakes simply by virtue of the higher level of clarity and consistency that result from their application. But the working party wants to make it very clear that no quantitative methodology will ever be fully successful in detecting intentional attempts at fraud or accounting abuse. Regulators and auditors face a difficult but necessary task in ferreting out the motives and intent of the producers of accounting statements. Actuaries are important partners and advisors in the area, especially in areas such as risk transfer. But it would be a mistake to think that actuaries or any other quantitative expert can provide a formula that reduces the analysis of intent, good or bad, to a simple (or even complex) calculation. This is important, because many of the alleged acts that have topped recent headlines are in fact much more about bad intent than risk transfer. No matter how good this working party's work, the methodologies developed here would not likely have prevented many of the alleged abuses, at least not without other efforts to discern the intent of the transactions.

At the same time, it is important to remember that in most reinsurance transactions the parties are acting in good faith and their intentions are good. Just as a mathematical test cannot identify bad intent, it cannot by itself discern the likely good intent of the parties. Therefore, the failure of a contract to meet a quantitative risk transfer test should not result in denial of reinsurance accounting treatment to a transaction without a thorough review of the all aspects of the deal, including the question of intent.

1.2 Disclaimers

While this paper is the product of a CAS working party, its findings do not necessarily represent the official view of the Casualty Actuarial Society. Moreover, while we believe the approaches we describe are very good examples of how to address the issue of risk transfer, we do not claim they are the only acceptable ones.

In the course of the paper, in order to make our ideas as clear as possible, we present a number of numerical examples that require assumptions about the distribution of losses and

appropriate threshold values for the risk transfer tests we describe. We recognize that any loss model we choose is an approximation to reality at best and might even be a poor one, and that with respect to the decision about appropriate risk transfer threshold values, other constituencies, including regulators, accountants and outside auditors have a key role to play. In making such assumptions for purposes of illustration, we are not necessarily endorsing any particular loss model or threshold value.

In many of our examples we display the results of calculations to two decimal places, which suggests an unreasonably high level of precision. We do so only in order to highlight the differences in what are frequently very small numbers. We are not suggesting that use of two decimal places is appropriate in the practical application of the methods we describe.

Throughout the paper we use the FAS 113 definition of the reinsurer's loss, which ignores brokerage and the reinsurer's internal expenses. Our use of that definition should not be construed to mean that we endorse that definition for any purpose other than testing reinsurance contracts for compliance with FAS 113.

1.3 Organization of Paper

The paper is structured in nine sections.

Section 1 describes the impetus for and context of the paper as well as a summary of the risk transfer requirements of FAS 113, which we treat as a reasonable framework for evaluating risk transfer, subject to a fair interpretation of the critical elements of "reasonably possible" and "substantially all". To meet the FAS 113 risk transfer requirements, a contract must satisfy one of two conditions: 1) the reinsurer must assume "substantially all" of the underlying insurance risk, or 2) the reinsurer must assume "significant" insurance risk and it must be "reasonably possible" that the reinsurer may realize a "significant" loss.

In Section 2 we present a systematic approach for determining whether "substantially all" of the underwriting risk has been transferred under a reinsurance contract. If "substantially all" the risk has been transferred, then the contract meets the risk transfer requirement of FAS 113 without it being necessary to show that the risk transfer is "significant". This section partially addresses the third question.

In Section 3 we present a detailed critique of the "10-10" test itself and how it has been applied in practice. We first describe the emergence of the "10-10" approach as a method of testing contracts for "significant" risk. Then we illustrate the application of the "10-10"

benchmark to three reinsurance contracts that clearly contain risk, including a property catastrophe contract and two quota shares of primary portfolios. All the tested contracts "fail" the "10-10" test, implying that the test is flawed. In the context of one of the examples we also emphasize the importance of taking parameter uncertainty into account in the risk assessment. Finally, we point out some unintended consequences of "10-10", namely that it implicitly imposes price controls on reinsurance contracts. We conclude that "10-10" is inadequate as a measure of risk and therefore unsuitable as a universal test for determining the "significance" of risk transfer. At best, one may argue that "10-10" is a sufficient test for risk transfer. It is not, however, a necessary condition.

Section 4 discusses two specific shortcomings of "10-10" and describes a different approach that addresses those shortcomings, thus addressing the first, second and fourth questions to varying degrees. The improved test we present here is based on the *expected reinsurer deficit* (*ERD*), which incorporates present value underwriting loss frequency and severity into a single measure. The loss severity embedded in the *ERD* is the tail value at risk (*TVaR*) measured at the economic breakeven loss ratio. We show that the *ERD* test is effectively a variable *TVaR* standard. We point out that a "significance" threshold of $ERD \geq 1\%$ has the merit of a certain amount of continuity with the "10-10" but without that test's major shortcomings. In order to address concerns that "10-10" might not be a strict enough standard, we also suggest the possibility of a supplemental minimum downside requirement. However, we do not advocate retesting of contracts already on the books that have already been found to pass "10-10".

Section 5 shows the application of the *ERD* test to the same contracts tested in Section 3 as well as to additional quota share contracts with loss ratio corridors or loss ratio caps, as well as to excess swing-rated contracts and individual risks. Using an illustrative standard of $ERD \geq 1\%$, we show that contracts that most people would consider risky receive a "passing" score, with one exception. This further addresses the first two questions.

Section 6 discusses the identification of contracts subject to the "significant" risk requirement, but which do not require individual testing, and thus addresses the third question. The NAIC is considering a requirement that the CEO and CFO attest that a risk transfer analysis has been completed for all reinsurance contracts, except those for which it is "reasonably self-evident" that significant risk has been transferred. We seek to put some definition to "reasonably self-evident". In this section we illustrate the application of the

ERD ≥ 1% test to several classes of reinsurance contracts with certain structural features. We show, using conservative assumptions, that 1) standard catastrophe excess of loss treaties, 2) contracts covering individual risks and 3) certain other excess of loss reinsurance structures, could all be "pre-qualified" as meeting the "significant" risk requirement (unless there is reason to believe they include other features that might affect the amount of risk transferred). We also describe an additional approach that could potentially be used to further expand the set of such contracts.

Section 7 discusses the possible evolution of risk measurement beyond the application to risk transfer testing that is the focus of this paper. This section offers an alternative way to address the first two questions. It briefly presents a framework proposed based on *right tail deviation* (*RTD*) that tightly links risk transfer testing and risk loading. We present two examples. While the *RTD*-based approach has theoretical appeal, it has the drawback of being more complex and thus less understandable to a non-actuarial audience than the *ERD* approach.

Section 8 is a summary of the key points of the paper.

Section 9 provides suggested priorities for areas of further research.

Appendix A gives the mathematics underlying the *ERD* test. Appendix B explains the comparison between S&P 500 equity risk and quota share reinsurance risk (which is used in examples in Sections 3 and 5). References are listed in Section 10, which follows the appendices.

1.4 Background

FAS 113 ("Accounting and Reporting for Reinsurance of Short-Duration and Long-Duration Contracts") was implemented in 1993[1] to prevent, among other things, abuses in GAAP accounting for contracts that have the formal appearance of reinsurance but do not transfer significant insurance risk and thus should not be eligible for reinsurance accounting. FAS 113 amplified the earlier requirement of FAS 60 that reinsurance accounting only applies to contracts that transfer insurance risk. SSAP 62, which largely incorporates the same language as FAS 113, was implemented shortly thereafter to address the same issues

[1] It was issued in December 1992 for implementation with respect to financial statements for fiscal years commencing after December 15, 1992. Since insurance companies generally have fiscal years that coincide with calendar years, in effect it was implemented for the 1993 fiscal year.

with respect to statutory accounting. Our references to FAS 113 should be understood to refer collectively to FAS 113 and SSAP 62.

In order for a contract to qualify for reinsurance accounting treatment in accordance with FAS 113, it must transfer insurance risk from an insurer to a reinsurer. To meet the risk transfer requirement, a reinsurance contract must satisfy one of two conditions:

1. It must be evident that "the reinsurer has assumed substantially all of the insurance risk relating to the reinsured portion of the underlying insurance contracts" (paragraph 11), or
2. The reinsurer must "assume significant insurance risk under the reinsured portions of the underlying insurance contracts"(paragraph 9a) and it must be "reasonably possible that the reinsurer may realize a significant loss from the transaction" (paragraph 9b).

We are aware that our presentation of the two FAS 113 conditions in this order (i.e., first the paragraph 11 condition and then the paragraph 9 condition) is unusual. In practice, the "significant" risk requirement has often been considered first, and only if the contract "fails" is paragraph 11 considered. However, because part of our aim is to determine how to avoid testing every contract, we find it useful to start with the consideration of whether the contract meets the risk transfer requirement by virtue of "substantially all" the underlying risk having been transferred. If it does, then the "significant" risk question does not need to be considered at all. Accordingly, throughout the paper we will present and work with the FAS 113 risk transfer conditions in that conceptual order.

This paper is not intended to be a critique of FAS 113. We treat FAS 113 as it is currently constructed as a reasonable framework for evaluating risk transfer, subject to a fair interpretation of the critical elements of "reasonably possible" and "substantially all", despite some reservations about its focus on the financial effects (excluding brokerage and internal expenses) of a transaction on the reinsurer alone.

While all reinsurance contracts must satisfy the requirements of FAS 113, it is up to each company to determine which contracts should be subjected to detailed testing and which contracts clearly satisfy the requirements of FAS 113 based upon inspection. In this paper we describe an approach that can help guide both ceding companies and reinsurers through that decision process.

2. DETERMINING WHETHER THE CONTRACT TRANSFERS "SUBSTANTIALLY ALL" UNDERLYING INSURANCE RISK

We suggest it makes sense to begin by determining whether the contract meets the FAS 113 condition of transferring "substantially all" the insurance risk. If it does, then the contract meets the risk transfer requirement. If it does not, then the contract is subject to the other condition that the risk transfer must be "significant".

What is the "insurance risk relating to the...underlying insurance contracts?" We see it as the *downside risk* associated with the cedent's portfolio of insurance, i.e., the exposure faced by the underwriter to incurring a loss. If the downside risk assumed by the reinsurer is essentially the same as that faced by the cedent with respect to the original unreinsured portfolio, then the contract transfers "substantially all" the insurance risk.

The trivial case is a quota share or other proportional contract with a flat ceding commission equal to the ceding company's expense ratio, where there are no features such a sliding scale commission, profit commission, loss ratio corridor or aggregate loss ratio limit. In such a case, the comparison between the ceding company's position and that of the reinsurer is obvious. The contract clearly transfers not only "substantially all" the risk to the reinsurer but literally all of it. Facultative reinsurance is often written on this basis, but more often than not, quota share treaties include one or more of the features identified above.

Sliding scale and/or profit commission features are often used by reinsurers as incentives to reinforce the ceding company's motivation to underwrite its business in a disciplined way. Their use can promote a win-win situation for the ceding company and the reinsurer. These and other features such as loss ratio corridors or caps appear frequently in traditional reinsurance contracts as a means of making otherwise unattractive treaties acceptable to the reinsurance market. Usually the context for incorporation of caps or corridors is poor historical underwriting experience in the portfolio for which reinsurance is being sought. The ceding company believes it has taken the necessary corrective actions to turn the portfolio around, but the reinsurance market is skeptical. The inclusion of caps and corridors in a reinsurance contract can often make it possible for a ceding company that has confidence in its own business plan to obtain the reinsurance capacity it requires to execute that plan. Sometimes, but not always, such features have the effect of taking "too much" risk out of a reinsurance deal to allow the "substantially all" requirement to be met. We need

to be able to compare the downside risk in the ceding company's unreinsured policies with the downside risk of the reinsurer.

We describe two ways of making this comparison – there may be other good methods as well – and illustrate them with an example. The first method is easier to understand but is not always conclusive, while the second method is somewhat more complicated but can always be applied.

Method 1 – Comparison of All Underwriting Downside Scenarios

Compare the cedent's underwriting margin over a range of loss ratios on the original unreinsured portfolio to the reinsurer's underwriting margin over the same range of loss ratios. The cedent's underwriting margin is defined as 100% less its unreinsured loss ratio less its actual expense ratio on the unreinsured portfolio[2]. The reinsurer's underwriting margin is defined as 100% less its assumed loss ratio less the ceding commission[3]. If the cedent's margin equals or exceeds the reinsurer's margin for the loss ratios that imply an underwriting loss, then clearly the reinsurer has assumed "substantially all" of the insurer's downside risk. Even if the cedent's margin is less than the reinsurer's margin, if that difference is small (as it is in Example 2.1), then the "substantially all" test may be met. Note that unless there are significant cash flow differences between the ceding company and the reinsurer, it is not necessary to conduct a full analysis of cash flows, since they will affect both parties in the same way.

Method 2 – Comparison of Cedent and Reinsurer Expected Underwriting Deficits

Compare the expected underwriting deficits (*EUD*) of the cedent and the reinsurer. The *EUD* can be calculated either directly as the pure premium of an aggregate excess of loss

[2] Expenses before reinsurance divided by premiums before reinsurance. Whether expenses should be marginal or average is a matter of debate.

[3] This definition of the reinsurer's underwriting margin does not reflect other expenses of the reinsurer, including brokerage and internal expenses. While this approach to measuring the reinsurer's profitability is consistent with the FAS 113 definition, it does not reflect economic reality.

cover attaching at the breakeven loss ratio or as the product of the frequency and severity of underwriting loss, (*Freq*(*UL*) and *Sev*(*UL*), respectively)[4].

If the *EUD* faced by the reinsurer is greater than or equal to the *EUD* of the cedent, then the "substantially all" test is clearly met. Because "substantially all" is less than "all", if the *EUD* faced by the reinsurer is within a small tolerance of the expected underwriting deficit faced by the cedent, say, within 0.1%, then we would also say the "substantially all" test is met.

Let's consider an example to illustrate these two methods.

Example 2.1: Non-Standard Auto Share with Sliding Scale Commission

Suppose a quota share of a non-standard auto portfolio is under consideration. The ceding commission is on a sliding scale. A minimum commission of 19.5% is payable if the loss ratio is 73% or higher. The commission slides up at a rate of one point for every one point of reduction in the loss ratio ("1:1 slide") below 73%, up to 30% at a loss ratio of 62.5%. The commission increases above 30% at a rate of 0.75% for every one point of loss ratio reduction ("0.75:1 slide") below 62.5%, up to a maximum commission of 39%, which is achieved at a loss ratio of 50.5% or lower. The ceding company's direct expense ratio on the subject business is 20%, so at the minimum ceding commission of 19.5%, it recoups virtually all of its direct costs. Its underwriting breakeven loss ratio is 80%. The reinsurer's FAS 113 underwriting breakeven loss ratio (i.e., ignoring brokerage and reinsurer internal expenses) is 80.5%.

The results of Method 1 are given in Table 1 and the accompanying Chart 1. The table compares the ceding company's expense ratio and underwriting margin on the unreinsured portfolio over a wide range of loss ratios to the reinsurer's ceding commission expense and underwriting margin at the same loss ratios. The accompanying chart compares the ceding company's margin and the reinsurer's margin graphically. From Table 1 and Chart 1 we see that above an 80% loss ratio (the ceding company's breakeven on the unreinsured portfolio), the ceding company's margin and reinsurer's margin are virtually undistinguishable, which

[4] If x represents the loss ratio and B is the underwriting breakeven loss ratio, then

$EUD = \int_B^\infty (x - B)\, f(x)\, dx = Freq(UL) \cdot Sev(UL)$, where $Freq(UL) = \int_B^\infty f(x)\, dx$ and $Sev(UL)$ is the "tail value at risk" (*TVaR*) at the underwriting breakeven: $Sev(UL) = \int_B^\infty (x - B)\, f(x)\, dx \,/\, \int_B^\infty f(x)\, dx$

indicates the reinsurer has assumed "substantially all of the insurance risk" of the reinsured policies.

TABLE 1

"Substantially All" Risk Transfer Analysis - Method 1
Comparison of Reinsurer vs. Cedent Margins
Example 2.1

Subject Loss Ratio	Cedent Expense Ratio	Cedent Margin	Reinsurance Ceding Commission	Reinsurer Margin
30.0%	20.0%	50.0%	39.0%	31.0%
50.5%	20.0%	29.5%	39.0%	10.5%
62.5%	20.0%	17.5%	30.0%	7.5%
73.0%	20.0%	7.0%	19.5%	7.5%
80.0%	20.0%	0.0%	19.5%	0.5%
80.5%	20.0%	-0.5%	19.5%	0.0%
100.0%	20.0%	-20.0%	19.5%	-19.5%

CHART 1
Cedent and Reinsurer Margins
Example 2.1

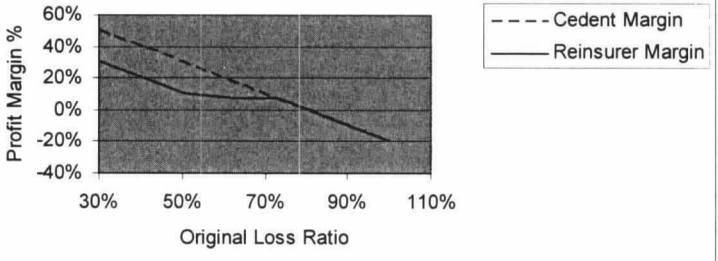


Table 2 summarizes the Method 2 comparison of expected underwriting deficits. It shows the insurer's and reinsurer's comparative underwriting downside risk by examining their respective *Freq(UL)*, *Sev(UL)* and *EUD*. In this example, the ceding company's frequency of underwriting loss is 11.28% vs. 10.45% for the reinsurer. The ceding company's underwriting loss severity is 8.33% vs. the reinsurer's 8.48%. The ceding company's *EUD* is 0.94% vs. the reinsurer's *EUD* of 0.89%[5]. While these measures vary slightly between the ceding company and the reinsurer, they are clearly very close. Thus, we would say that Method 2 also indicates that the reinsurer has assumed "substantially all" of

TABLE 2

"Substantially All" Risk Transfer Analysis - Method 2
Reinsurer vs. Cedent Margins in Downside Scenarios
Example 2.1

	Breakeven Loss Ratio	*Freq(UL)*	*Sev(UL)*	*EUD*
Cedent	80.0%	11.3%	8.3%	0.940%
Reinsurer	80.5%	10.5%	8.5%	0.886%
Difference	-0.5%	0.8%	-0.2%	0.054%

the ceding company's downside risk and the contract therefore meets the risk transfer requirements of FAS 113.

We conclude that in this example either Method 1 or Method 2 indicates the contract transfers "substantially all" the underlying insurance risk to the reinsurer.

While this approach works most naturally for quota share contracts, it can potentially be applied to excess of loss treaties as well. In that case, the reinsurer's *EUD*, calculated in the same way as above in the quota share case as a ratio to the ceded premium, should be compared to the cedent's *EUD* on the portion of the original subject portfolio which is exposed to the same risks as the excess of loss reinsurance contract. If the reinsurer's *EUD*

[5] Losses have been modeled using a lognormal distribution modified for parameter uncertainty, the details of which are not important for this example.

is close to or greater than the cedent's, then the reinsurer can be judged to have assumed "substantially all" the cedent's insurance risk in this context. For example, suppose the portion of original insurance risk assumed by a catastrophe reinsurance contract covering a portfolio of business has a 1% probability of a claim of a certain size. In that case the reinsurance of that portion of the risk also requires no more than a 1% probability of loss of the same size, because the *EUD*s of the ceding company and the reinsurer are the same with respect to the original catastrophe exposure.

If our argument about the applicability of the comparative *EUD* approach to excess of loss contracts and contracts with loss ratio caps is not found to be compelling, note that in section 6 we will also demonstrate that catastrophe reinsurance and some other contracts with aggregate loss limitations can meet the "significant" risk requirement under many circumstances.

Finally, there is a case to be made that, to the extent that a ceding insurance company is limited in its ability to meet net losses by its surplus, it is reasonable to allow a similar limitation of the reinsurer's aggregate liability. If this is accepted, then it is possible to calculate the minimum loss ratio cap that can be imposed by the reinsurer without violating the condition that "substantially all" of the underlying risk has been transferred. This potentially represents a third way of determining whether the "substantially all" risk transfer condition has been met.

For example, suppose a ceding company enters into a whole account quota share reinsurance arrangement that results in a net premium to surplus ratio of 200%. If the quota share has a ceding commission of 25% (approximating the ceding company expenses), then a loss ratio cap as low as 125% would be consistent with the transfer of "substantially all" of the risk, because at a combined ratio of 150% the ceding company has lost all of its surplus. Naturally such an interpretation would have to be made after due consideration of all other relevant features of the reinsurance contract in question.

If a contract does not meet the "substantially all" test, then it is subject to the second FAS 113 condition that "significant risk" must be transferred in order for the contract to qualify for reinsurance accounting. We now turn our attention to the question of what constitutes "significant" risk.

3. "SIGNIFICANT" RISK TRANSFER AND THE "10-10" TEST

3.1 "10-10" and its Shortcomings

A contract that does not meet the FAS 113 requirement for risk transfer by transferring "substantially all" the underlying insurance risk is subject to the second condition that "significant" risk be transferred. The so-called "10-10" test emerged in the years following the implementation of FAS 113 as a common benchmark for determining whether a reinsurance contract satisfies the requirement of a reasonable chance of "significant" loss to the reinsurer, which the test defines as "at least a 10% chance of a 10% loss". "10-10" is usually referred to as a "risk transfer" test, which implies an understanding of "risk" as a measure of exposure to loss rather than as exposure to volatility of results. "10% chance of a 10% loss" is usually interpreted to mean that the underwriting loss at the 90th percentile (of the probability distribution of underwriting results[6]) must be at least 10% of the ceded reinsurance premiums, where both underwriting loss and premiums are understood to be present values. Another term for "the underwriting loss at the 90th percentile" is "the value at risk" at the 90th percentile" or "$VaR_{90\%}$" with respect to the underwriting result. Accordingly, the "10-10" test can also be succinctly described as requiring $VaR_{90\%} \geq 10\%$.

The "10-10" benchmark arose as an informal method for testing whether purported reinsurance contracts contained sufficient risk transfer to meet the requirements of FAS 113 under the reasonable chance of significant loss criterion. It was not intended to be a universally applicable risk transfer test. Indeed, it has long been recognized that many reinsurance contracts having the characteristics of low underwriting loss frequency but high severity (such as property catastrophe excess of loss reinsurance) fail "10-10" on the basis that the probability of a 10% loss is less than 10%. In addition, if they do not meet FAS 113 risk transfer requirements by virtue of transferring "substantially all" risk, ordinary quota share reinsurance of many primary insurance portfolios (e.g., low limits private passenger auto), which have the characteristics of high frequency of underwriting loss but relatively low severity, may also fail. Until recently that was not seen as a problem because experienced practitioners understood the target of FAS 113 to be highly structured contracts that limited the transfer of insurance risk. As a consequence, traditional reinsurance contracts were typically not even tested.

[6] Low percentiles represent better results; high percentiles represent poorer results. Underwriting losses are represented as positive numbers. References to "underwriting results" and "underwriting losses" should be understood to refer to present values.

In the wake of the recent revelations of new accounting abuses related to "reinsurance contracts" apparently involving little or no risk transfer, the situation has changed. There is greater sentiment now that (a) more contracts should be routinely tested for significant risk transfer and (b) "10-10" is not a stringent enough standard. The view that "10-10" may not be stringent enough arises in part from the fact that some highly structured contracts have been carefully engineered to allow for exactly a 10% probability of a 10% loss and little or no possibility of a loss greater than 10%.

It is clear from the failure of the "10-10" benchmark to correctly identify both catastrophe excess of loss and some quota share reinsurance as risky and its failure to flag certain highly structured contracts as not significantly risky that "10-10" is insufficiently discriminating to serve as a universal measure of risk transfer in reinsurance contracts. We need a better test for measuring significant risk transfer in contracts that are subject to that requirement.

The interpretation of FAS 113's paragraph 9b is a critical issue. Paragraph 64 states that "an outcome is reasonably possible if its probability is more than remote." Despite this definition, the expectation appears to have developed that "reasonably possible" means a probability substantially greater than "remote". While the accounting literature gives no specific guidance on these probabilities, a 10% chance has come to be widely accepted as the smallest probability that should be categorized as "reasonably possible." It is our position that a different interpretation of "reasonably possible" is more appropriate, one that depends on the context of the risk and recognizes that some weight should be given to loss scenarios that, while rare, are not remote.

In particular, we propose that, in establishing the threshold probability for "reasonably possible", consideration must be given to the probability of loss (and indeed the size of that loss) arising from the reinsured portions of the underlying insurance contracts. For example, in the context of catastrophe reinsurance, "reasonably possible" should be associated with a probability that reflects the inherently low probability of the covered event. For other reinsured portfolios, where the inherent probability of loss is greater, "reasonably possible" is appropriately associated with a higher probability value.

This interpretation goes a long way toward eliminating the apparent inconsistency of according reinsurance accounting to some contracts that do not satisfy an invariant probability threshold of 10%. That property catastrophe contracts are typically accorded

reinsurance accounting treatment even though they often do not meet a "reasonable possibility" requirement, defined as 10%, implicitly reflects this kind of interpretation.

In section 4 we will present a framework for capturing the interaction between the "reasonably possible" and "significant loss" components of paragraph 9b in a way that automatically makes the appropriate contextual adjustment without having to resort to situation-based arguments.

First, let us continue our critique of "10-10".

3.2 Illustration of the Shortcomings of "10-10"

Through a series of examples we will show why "10-10" is an unsatisfactory test for establishing whether or not a reinsurance contract transfers significant risk. Example 3.1 illustrates the application of the test to a property catastrophe contract and shows that it "fails" to transfer significant risk. Example 3.2 illustrates the application (and misapplication) of "10-10" to a low volatility primary quota share, given a set of historical loss ratio experience. We also use that example to warn of the pitfalls of simply fitting a loss distribution to on-level loss ratio experience and using that for risk transfer analysis. Example 3.3 shows that a quota share of an insurance portfolio having the volatility characteristics of the S&P 500 would frequently fail the "10-10" test.

We begin with the property catastrophe example.

Example 3.1: Property Catastrophe Excess of Loss Reinsurance

A property catastrophe reinsurance contract paying a premium equal to 10% of the limit[7] is typically priced to a loss ratio of around 50%. That implies an expected loss of 5% of the limit. Catastrophe reinsurance contracts, especially for higher layers, run loss free or have small losses in most years but occasionally have a total limit loss. This pattern is illustrated by the simplified catastrophe loss distribution shown in Table 3 below.

[7] This is frequently referred to as a "10% rate on line".

TABLE 3
Catastrophe Loss Distribution for Example 3.1

Loss as % of Limit	Loss as % of Premiums	Probability of Given Loss
0%	0%	67%
5%	50%	20%
10%	100%	10%
100%	1000%	3%
5%	50%	100%

The loss at the 90th percentile of the catastrophe loss distribution is 100% of premiums. Assuming standard reinstatement premium provisions, the 90th percentile of the underwriting result distribution is an underwriting profit of 10% of premiums (100% original premiums plus 10% reinstatement premiums minus 100% loss). This contract fails the "10-10" test.

There is universal agreement among accountants, regulators, insurers, reinsurers and rating agencies that contracts like this one are risky. Clearly, the failure of "10-10" to identify the contract in this example as risky is an indication of a problem with "10-10" and not the contract.

Example 3.2: Primary Quota Share Reinsurance

Assume a cedent and reinsurer have negotiated a quota share treaty on a primary insurance portfolio. The treaty has a ceding commission of 25%. Does the treaty contain "significant" risk transfer[8]?

[8] Let's assume the treaty does not meet the condition of transferring "substantially all" of the underlying risk, perhaps because the cedent's expenses are substantially greater than the ceding commission. As a result the treaty is subject to the "significant" risk transfer requirement.

To measure the risk transferred we need to model the prospective underwriting result. Because the underwriting result is the breakeven loss ratio minus the actual loss ratio, the key to modeling the underwriting result is the probability distribution of the prospective loss ratio x. There are a number of reasonable actuarial methods for modeling prospective loss ratios[9]. In actuarial pricing applications the principal focus is on the mean of the prospective loss ratio distribution. Not much attention is paid to the full distribution. In contrast, risk transfer analysis requires the full distribution. This means there are pitfalls associated with using the output from the pricing analysis for the risk transfer analysis without full consideration of the issues affecting the full loss ratio distribution.

Let's review the underwriting experience analysis of the insurance portfolio that is the subject matter of the quota share. Five years of loss ratio experience is available together with information of varying quality about historical loss development and claim trends as well as the rate level history and the cedent's expectation of rate actions during the treaty period. This is summarized in Table 4, which shows the reported, estimated ultimate and estimated ultimate "on-level" loss ratios[10] together with the loss development, premium on-level and loss on-level factors used in the analysis. The means, variances and standard deviations of the on-level loss ratios x_i and their natural logs $\ln x_i$ are tabulated using the assumption that exposure has been constant over the experience period.

The historical experience has been poor. Given the ceding commission of 25% and ignoring brokerage and internal expenses (as per FAS 113), the reinsurer's present value breakeven loss ratio is 75%[11]. Three of the five years have estimated ultimate loss ratios significantly greater than 75% and in two of the years the loss ratio is over 75% even on a reported basis. The good news is that the ceding company has taken action to increase rates significantly, which results in estimated on-level loss ratios that are much lower than the actual historical loss ratios. The on-level mean of 70.67% compares very favorably with the

[9] The models we use for the purposes of illustrating the issues related to risk transfer testing are not intended to be prescriptive and are independent of the risk measurements we describe.

[10] This means the loss ratios have been adjusted to reflect the projected premium rate and claim cost levels expected to apply during the treaty term.

[11] Note that given typical brokerage of 1.5% and internal expenses of 3% to 5%, reinsurers would regard their real breakeven loss ratio as 68.5% to 70.5%, depending on expenses. As we shall see, this treaty is a breakeven or slightly worse than breakeven proposition and would not be attractive to most reinsurers.

historical mean of about 80%. Moreover, the on-level loss ratios are not very variable as indicated by the standard deviations of 7.45% with respect to x and 10.88% with respect to $\ln x$.

TABLE 4
On-Level Loss Ratio Experience
For Quota Share in Example 3.2

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Accident Year	Reported L/R	Age to Ult Factors	Est Ult L/R	Prem On-Level Factors	Loss On-Level Factors	On-Level L/R x_i	$\ln x_i$
1	92.8%	1.039	96.4%	1.963	1.364	67.0%	-0.401
2	75.6%	1.048	79.3%	1.737	1.307	59.7%	-0.516
3	77.0%	1.095	84.3%	1.376	1.246	76.4%	-0.269
4	61.2%	1.141	69.9%	1.139	1.181	72.5%	-0.321
5	52.5%	1.415	74.3%	1.061	1.111	77.8%	-0.251
				Mean	$\bar{x}$	70.7%	-0.352
				*Var**	s^2	0.554%	1.18%
*Unbiased				*St. Dev.**	s	7.45%	10.88%

We are first going to illustrate *how not to apply* the "10-10" benchmark in this scenario. We do this in order to point out the problems associated with this approach, which we believe may be in relatively common use.

Let's assume the underlying random process governing the prospective loss ratio is lognormal. Then the "best fit" distribution, given the on-level loss ratio experience, is defined by parameters $\mu = \bar{x}$ and $\sigma = s$. From this it is easy to determine whether the present value underwriting loss corresponding to $VaR_{90\%}$ exceeds 10%. If B is the present

value FAS 113 breakeven loss ratio and FV and PV represent "future value" and "present value" operators, respectively, then from the characteristics of the lognormal distribution we know that

$$N^{-1}(90\%) = \frac{ln[FV(B + VaR_{90\%})] - \mu}{\sigma} \tag{3.1}$$

which implies

$$VaR_{90\%} = PV(e^{\mu + N^{-1}(90\%)\cdot\sigma}) - B \tag{3.2}$$

If ceded loss payments lag ceded premium payments by one year on average, the risk free interest rate is 5%, $\mu = \bar{x} = -0.3518$ and $\sigma = s = 10.88\%$, then formula (3.2) implies

$$\begin{aligned} VaR_{90\%} &= e^{(-0.3518)+(1.2815)(0.1088)} \cdot (1.05^{-1}) - .75 \\ &= 2.02\% \end{aligned}$$

Since "10-10" requires $VaR_{90\%} \geq 10\%$, according to this analysis the quota share treaty in this example does not transfer "significant" risk. In fact, the $VaR_{90\%}$ of 2.02% suggests that the treaty contains hardly any risk at all. Yet when we look back at the historical experience, we see that the reinsurer would have lost more than 10% in one year and would have lost money over the entire period. The conclusion that the reinsurer does not face a "reasonable possibility of significant loss" seems strange.

Why did we get this result? There are two reasons. The first, as we hinted at the beginning, has to do with inadequacies in the loss model we selected. The second has to do with shortcomings in the "10-10" test itself.

Let's discuss the problem with the approach we described for identifying a loss ratio model. Fundamentally, the problem is that we fitted a single distribution to the on-level loss ratios and then used that distribution as though we knew with certainty that it is the correct one. In that case the only source of risk being modeled is process risk, because we have assumed we have the correct model. In fact, there are multiple sources of parameter uncertainty, some of which we enumerate below:

- The ultimate loss estimates might be wrong;
- The rate level history might be inaccurate;
- The prospective rate changes assumptions might be wrong;
- The historical claim trend estimates might be inaccurate;
- The prospective claim trend assumptions might be wrong;
- The experience period might be too short to include rare but very large losses;
- The prospective loss ratios might not be lognormally distributed;
- The lognormal assumption is right, but the "best fit" distribution is not the actual;
- Cash flow timing assumptions, particularly regarding claims, might be wrong;
- The prospective exposure mix might be different from expected;
- For multi-year reinsurance contracts, the level of parameter uncertainty from all sources increases as the length of the coverage period increases.

In any actuarial application where the knowledge of the loss distribution itself and not just its mean is important, it is very important that the modeling be based on loss models that incorporate parameter uncertainty, which is an important and frequently underestimated source of risk[12]. Risk transfer testing, given its dependence on the right tail of the loss ratio distribution is one of those applications.

Accordingly, actuaries should be cautious about placing too much confidence in a single distribution fitted to estimated loss ratios. Where the estimates are the result of applying large development and/or on-level factors, the likelihood of parameter error is especially large, and appropriately large adjustments must be made to the distribution to account for it.

While it is beyond the scope of this paper to discuss specific methods for estimating the impact of parameter uncertainty, for the sake of illustration, suppose the effect of reflecting parameter uncertainty in the current example is to increase σ in the lognormal model to 15%. If we constrain μ such that $E(x)$ remains unchanged, then $\mu = -0.3571$ and formula (3.1) yields $VaR_{90\%} = 5.76\%$, which still fails to meet the "10-10" threshold for

[12] Kreps[2] and Van Kampen [3] provide examples of large effects in loss reserve estimates and aggregate excess pure premiums, respectively, due to the recognition of parameter uncertainty.

"significant" risk transfer. In this case, an adjustment to try to take account of parameter uncertainty is not sufficient to show "significant" risk transfer in the contract, at least if we use "10-10" to measure it.

The next example brings into question the appropriateness of the "10-10" criterion of $VaR_{90\%} \geq 10\%$ by examining its implications for how we think about stock market risk.

Example 3.3: Primary Quota Share Reinsurance (Volatility of S&P 500)

Assume we are considering a quota share treaty on a second primary insurance portfolio. As in Example 3.2 the treaty ceding commission is 25%, which implies a FAS 113 breakeven present value loss ratio of 75%. Suppose this portfolio has the distributional and volatility characteristics commonly attributed to the S&P 500 equity index and an on-level loss ratio of 70%. This implies an assumption that the prospective loss ratio is lognormally distributed[13] with a mean of 70%. Let's also assume the claim payments lag premiums by one year. In order to pass the "10-10" test, which requires a present value loss ratio of at least 85% at the 90th percentile, if the risk free interest rate is 5%, the minimum value of the lognormal σ parameter is about 21%[14].

Actual annualized volatility in the price of the S&P 500 index exchange traded fund (symbol SPY) between early May 2004 and early May 2005 was 10.64%.[15] On May 4, 2005, the broadly based CBOE Volatility Index (VIX), a measure of the expected annualized volatility in the S&P 500 stock index implied by the market pricing of index options, closed at 13.85%. The market was using a higher estimate of future volatility for pricing purposes than that observed in the recent past, which might reflect an adjustment for parameter uncertainty or simply the opinion that volatility would increase. Both estimates of σ fall

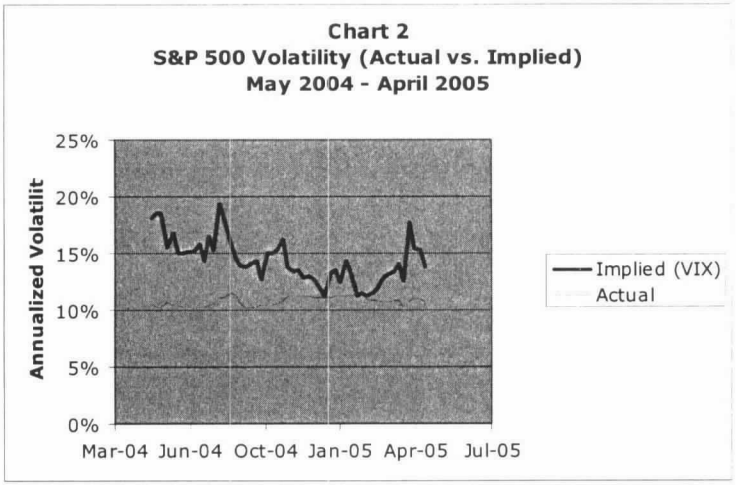
[13] For a discussion of the basis for this assumption, see Appendix B.

[14] $\sigma = \frac{ln[(.85)(1.05)] - \mu}{N^{-1}(.9)}$ and $\mu = ln(.70) - .5\sigma^2$ imply $\sigma = 20.6\%$ or 236%, the former being the only reasonable solution in this context. This threshold assumes a ceding commission of 25%, a risk free interest rate of 5% and lognormal stock prices. The threshold will vary depending on the parameters.

[15] Calculated as the annualized standard deviation of weekly log returns $\ln(P_w / P_{w-1})$ between May 2004 and May 2005.

below the threshold of 21% required to pass "10-10", implying that a "quota share" of the S&P 500 index[16] would fail to meet the FAS 113 requirement for significant risk transfer!

This is not merely a temporary aberration. During the period from early May 2004 through early May 2005 the actual volatility observed on a one-year look-back basis averaged 10.77%. Over the same time period, VIX averaged 14.39%. Chart 2 shows this graphically. The persistent pattern of VIX greater than actual historical volatility suggests that VIX reflects an adjustment for parameter uncertainty rather than a forecast that volatility will increase.

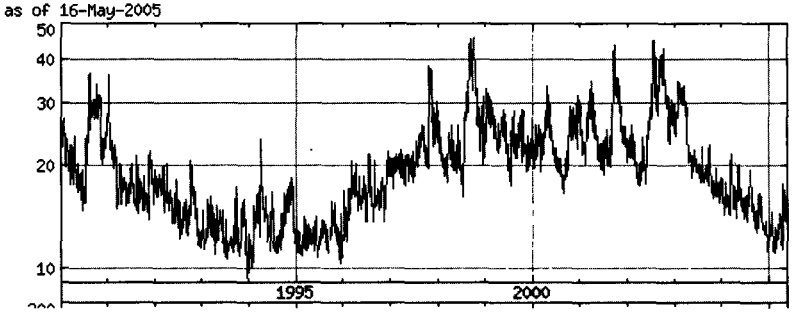


Source: Yahoo! Finance

Over a longer period of time the market opinion of the prospective volatility of the S&P 500 has varied considerably, ranging from a high of about 50% in 2002 to a low of about 9% in 1993[17]. Chart 3 shows this graphically.

[16] We put "quota share" in quotation marks because the S&P 500 index transaction comparable to a quota share of an insurance portfolio involves a short sale. Since a short sale is usually considered to be even riskier than a long position, the failure to "pass" a risk transfer test is all the more surprising. See Appendix B for details.

CHART 3



Source: Yahoo! Finance

Chart 4 shows the probability of a present value loss of 10% or more on the quota share of this example, given σ = VIX values as of the last trading day of each year from 1990 through 2004 plus May 4, 2005. It shows that the probability exceeds 10%, given the VIX values at the end of 1990 and those for every December from 1996 through 2002. However, the probability is less than 10%, given the VIX values from every December 1991 through 1995 and those for December 2002 and 2003 as well as that for May 2005[18]. Almost no one would argue that an investment in equities, even in a diversified portfolio such as the S&P 500, is not risky. Yet the implication of the "10-10" benchmark is that a quota share reinsurance that has the same volatility characteristics ascribed to the S&P 500 by the options market over the period since 1990 would have been considered risky only about half the time! Unless the intention is to set the bar for "significant" risk at a level higher than the typical volatility of the S&P 500, we must conclude that the "10-10" criterion is an inadequate measure of significant risk.

[17] For more information about VIX and its calculation, see the white paper published by the CBOE, which is available at its website: http://www.cboe.com/micro/vix/vixwhite.pdf. The paper included the history between 1990 and August 2003.

[18] The data underlying Chart 4 can be found in Appendix B.

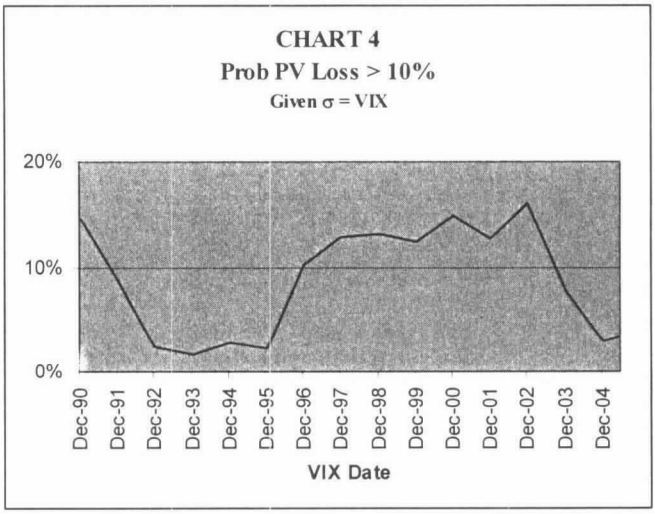


Table 5 illustrates the "10-10" analysis for a quota share of a portfolio whose loss ratio has the volatility characteristics of the S&P 500, for two volatility scenarios: 9% (representing the low end of the VIX range since 1990) and 13.85% (representing the VIX value on May 4, 2005). The ceding commission is 25%. The table shows (a) the loss at the 90th percentile of the present value underwriting result distribution, and (b) the probability of a present value loss of 10% or more, for σ = 9% and 13.85%. Both of these volatility scenarios fail to meet the "10-10" threshold for significant risk transfer.

If σ = 9%, which represents the low end of the range of S&P 500 implied volatility since 1990, the quota share actually has a negative loss (i.e., small profit) at the 90th percentile ("10% chance of a (0.49%) or greater loss") and a miniscule 0.30% probability of a 10% loss or more. This scenario fails the "10-10" test badly!

For σ = 13.85% Table 5 shows a 10% chance of a 3.85% or greater loss and a 3.41% chance of a 10% loss or more. This contract scenario also fails "10-10" by a long way[19].

[19] Note that even at an expected loss ratio of 75%, which is the treaty breakeven point, there is a 10% chance of only a 9.49% or greater loss. See Appendix B (Table B-2) for details about the sensitivity of the analysis to changes in the expected loss ratio assumption.

TABLE 5
"10-10" Risk Transfer Analysis
for Quota Share in Example 3.3
Given Portfolio with Volatility of S&P 500

VIX	σ	(a) 90th Percentile P.V. Underwriting Loss	(b) Probability of ≥ 10% P.V. Underwriting Loss
Low	9.00%	(0.49%)	0.30%
May 2005	13.85%	3.85%	3.41%

For further discussion of the comparability of quota share reinsurance with the S&P 500, see Appendix B.

3.3 Unintended Consequences: The Impact of "10-10" on Reinsurance Pricing

There is a further troubling implication of "10-10". It implicitly imposes price controls on reinsurance contracts at such a low level that, if that benchmark were to be enforced as a rule, reinsurance capacity for certain types of business is likely to be reduced, if not eliminated entirely.

To illustrate this we will assume the prospective loss ratio is lognormally distributed[20]. The mean of a lognormal distribution is given by

$$E(x) = e^{\mu+0.5\sigma^2} \tag{3.3}$$

If we solve for μ in formula (3.1) and substitute the result for the μ in formula (3.3) we obtain the formula for $E(x)$ constrained by $VaR_{90\%} = 10\%$:

[20] We choose the lognormal merely for purposes of illustration. A different distribution might be more appropriate.

$$E(x) = Exp\{\ln[FV(B + VaR_{90\%})] + N^{-1}(90\%) \cdot \sigma + 0.5\sigma^2\} \qquad (3.4)$$

For example, in the treaty scenario with no ceding commission, $B + VaR_{90\%} = 110\%$, and the minimum permissible loss ratio is:

$$E(x) = Exp\{\ln[FV(110\%)] + 1.2815 \cdot \sigma + 0.5\sigma^2\} \qquad (3.5)$$

Table 6 is a tabulation of the minimum permissible loss ratios allowed by "10-10" for a range of values of σ and average net claim payment lags of zero, one year, two years and three years. Chart 5 is a graphical representation of the data in Table 6. We see that for small values of σ and claim lags of a year or more, the minimum permissible loss ratios are greater than 100%, implying the reinsurer is required to price its business at an underwriting loss even before taking into account brokerage and its own internal expenses. Even at somewhat higher values of σ that might correspond to certain excess of loss business, the reinsurers's net underwriting margins (after typical brokerage of 10% and comparable internal expenses) are quite low.

For example, given $\sigma = 9\%$ and assuming no claim payment lag (and hence no investment income), the reinsurer's minimum permissible loss ratio is 98.4%. That implies a maximum allowable margin before brokerage and internal expenses of 1.6%. The maximum permissible loss ratio rises as the claim payment lag increases. The effect of the $VaR_{90\%} = 10\%$ constraint is that all the investment income earned as a result of the claim payment lag is credited to the cedent, and the present value of the reinsurer's margin remains at 1.6%. For example, given a three-year payment lag and a 5% interest rate, the breakeven loss ratio is 115.8% and the minimum permissible loss ratio is 113.9%, which leaves a future value margin for the reinsurer of 1.9%. The present value of that 1.9% is 1.6%. Clearly, given brokerage costs and internal expenses, no reinsurer could afford to write business at such a meager margin.

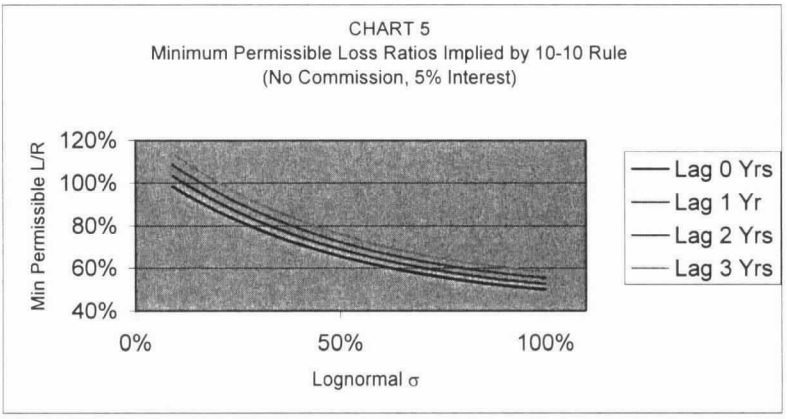
TABLE 6

Minimum Permissible Loss Ratio
Implied by "10-10"

Contracts with No Ceding Commission
Interest at 5% per annum

By σ and Claim Lag

σ	No Lag	1 Yr Lag	2 Yr Lag	3 Yr Lag
9.0%	98.4%	103.3%	108.5%	113.9%
10.0%	97.3%	102.1%	107.2%	112.6%
11.0%	96.1%	100.9%	106.0%	111.3%
12.0%	95.0%	99.8%	104.7%	110.0%
13.0%	93.9%	98.6%	103.5%	108.7%
14.0%	92.8%	97.5%	102.4%	107.5%
15.0%	91.8%	96.4%	101.2%	106.3%
20.0%	86.8%	91.2%	95.8%	100.5%
25.0%	82.4%	86.5%	90.8%	95.4%
30.0%	78.3%	82.3%	86.4%	90.7%
40.0%	71.4%	74.9%	78.7%	82.6%
50.0%	65.7%	69.0%	72.4%	76.0%
60.0%	61.0%	64.1%	67.3%	70.7%
75.0%	55.7%	58.5%	61.4%	64.5%
100.0%	50.3%	52.9%	55.5%	58.3%



In light of our earlier discussion of parameter uncertainty, it may well be that σ values as low as 9% will never be used in practice. However, the problem remains to some extent at higher values of σ. For example, for $\sigma = 30\%$ the maximum gross reinsurer's margin is 21.7% (100% less the minimum loss ratio with no claims lag). If the reinsurance is on an excess of loss basis, brokerage is likely to be 10% and internal expenses are likely to be a similar amount. That leaves only 1.7% as a net present value margin for the reinsurer, which is not likely to be attractive.

3.4 Section Summary

The discussion in this section should make it clear that the "10-10" benchmark is a flawed measure of "significant" risk transfer. The test used to measure risk transfer should accurately distinguish between contracts that clearly contain significant risk from those that don't. That "10-10" fails to identify both catastrophe reinsurance treaties and contracts with the characteristics of equity investments as risky tells us that it is a poor test. "10-10" also implies very restrictive caps on reinsurance pricing that can never have been intended. At the same time it has received criticism from the other direction that it does not do an adequate job of screening out contracts that meet its minimum requirements but in such a contrived way that the intent of FAS 113 is thwarted. For all of these reasons it makes sense to identify a better test than "10-10", which we seek to do in the next section.

4. TOWARD A BETTER TEST

There are at least two major shortcomings of the "10-10" test. First, the focus on the present value loss only at the 90th percentile ($VaR_{90\%}$) ignores the information in the remainder of the tail represented by the percentiles beyond the 90th. A better test would take account of the loss potential in the right tail of the distribution, which sometimes can be extreme (as in the case of catastrophe reinsurance). Second, both the 10% probability and 10% loss thresholds are arbitrary. The risk transfer test should be generalized to allow for both low frequency-high severity (e.g., 5%-20%) and high frequency-low severity (e.g., 20%-5%) combinations.

The first shortcoming could be remedied by replacing $VaR_{90\%}$ with the mean severity of present value underwriting losses at and beyond the 90th percentile, a measure known as the "tail value at risk" or $TVaR_{90\%}$[21]. This measure of severity incorporates the information about the loss potential in the right tail that the "10-10" test misses. Indeed, the 2002 VFIC paper suggested replacing $VaR_{90\%}$ in the "10-10" test with $TVaR_{90\%}$. However, simply replacing $VaR_{90\%}$ with $TVaR_{90\%}$ is not by itself a full solution to the problems associated with "10-10", because it leaves unaddressed that test's second shortcoming that the 10% thresholds wrongly screen out low frequency-high severity and high frequency-low severity contracts.

That second shortcoming can be corrected by relaxing the requirement that the probability of loss and the severity of loss must both exceed 10%. We can do this by making use of the fact that the *expected reinsurer deficit* (*ERD*)[22] is equal to the probability (or *frequency*) of the present value underwriting loss times its average *severity*, where the latter is *TVaR* measured at the economic breakeven point. Since *ERD* incorporates information about both the frequency and severity of the reinsurer's downside risk into a single measure, it makes sense to use that measure to define a threshold for measurement of significant risk transfer rather than to define it in terms of frequency and severity separately:

---

[21] Also known as the "tail conditional expectation" or "*TCE*", *TVaR* has been praised by VFIC[1], Meyers [4], and others as a coherent measure of risk as well as for its incorporation of the information contained in the right tail of the distribution.

[22] The *ERD* is the expected cost of all present value underwriting loss scenarios. It is also the expected value of Mango's [5] contingent capital calls. Conceptually, it is related to the *EUD* defined in Section 2, but the *EUD* is defined in nominal terms and the *ERD* is defined in present value terms.

$$ERD = Freq \times Sev \geq A \tag{4.1}$$

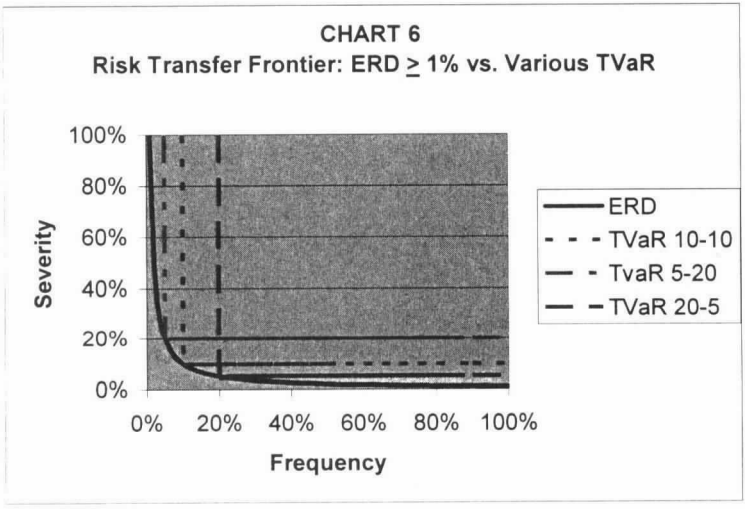
where A represents the threshold above which a contract is considered to have provisionally "passed" the "significant" risk transfer test and below which is it is considered to have "failed". *Freq* and *Sev* refer to the frequency of present value loss and the average severity of such loss, respectively. See Appendix A for the mathematical definitions of all the elements of formula (4.1).

This approach, which we will refer to as the "*ERD* Test", addresses both shortcomings of the "10-10" test by (a) reflecting the full right tail risk in the definition of severity and (b) replacing separate frequency and severity requirements with a single integrated measure that treats low frequency-high severity, high frequency-low severity and moderate frequency-moderate severity contracts in the same way.

We will illustrate the application of the *ERD* test with a threshold A of 1%, because it has the merit of a certain amount of continuity with the "10-10" test[23]. The way to think about that is that first we have changed the $VaR_{90\%} \geq 10\%$ embodied in the "10-10" test to $TVaR_{90\%} \geq 10\%$. Then we have generalized the *TVaR* standard to allow contracts having a wide variety of frequency-severity combinations, including 5%-20%, 10%-10% and 20%-5%, to meet the requirement for "significant" risk transfer. $ERD \geq 1\%$ is effectively a variable *TVaR* standard that defines "significant" as $TVaR_{1-Freq} \geq \frac{1\%}{Freq}$. One implication of this is that any contract that passes "10-10" will also pass a standard of $ERD \geq 1\%$.

Chart 6 shows the "significant" risk transfer frontiers for $ERD \geq 1\%$ and three *TVaR* standards ("10-10" as well as "5-20" and "20-5") plotted in terms of frequency and severity. Frequency-severity combinations above and to the right of the frontiers represent "significant" risk. We see that a fixed *TVaR* "10-10" standard would exclude contracts with loss frequencies less than 10% and severities less than 10% that the *ERD* standard would accept as "significant". As a generalized *TVaR* standard, a $ERD \geq 1\%$ standard would accept $TVaR_{95\%} \geq 20\%$ or $TVaR_{90\%} \geq 10\%$ or $TVaR_{80\%} \geq 5\%$, etc.

[23] Whether that is the proper threshold warrants further research.



To address the issue of contracts that have been engineered to remove most or all of the potential for a loss greater than 10% in the right tail, which some criticize as too small, we suggest consideration of a supplemental requirement that there be the potential for a reinsurer loss of some minimum threshold, say, 15% or 20% of premiums. That would eliminate very low loss ratio caps.

We are not advocating that every reinsurance contract be tested for significant risk transfer. It should be possible to conclude that some contracts have adequate risk transfer without formally testing them. In section 6 we will suggest some ways to do that. However, we *are* suggesting that the *ERD* test (possibly together with the supplemental test) could be applied to all contracts that are subject to the "significant" risk transfer requirement with the confidence that it would produce consistently reasonable results.

We believe the *ERD* test (with or without the supplemental component), if adopted, should only be applied prospectively and not to contracts already on the books.

5. ILLUSTRATION OF THE *ERD* TEST

In this section we apply the proposed test to the contracts used in the examples of Section 3 as well as several additional examples.

Example 5.1: Property Catastrophe Excess of Loss Reinsurance

If we apply the *ERD* test to the catastrophe reinsurance contract described in Example 3.1, that contract now easily passes muster for risk transfer. Again assuming normal reinstatement premium provisions, which call for an additional premium equal to the original premium times the proportion of the limit that has been exhausted, *Freq*=3%, *Sev*= $TVaR_{97\%}$=800% and *ERD* =24%. Because of the large contribution from *Sev* to *ERD*, this contract now easily surpasses the standard of $ERD \geq 1\%$.

TABLE 7

ERD / Max Downside

For Standard Cat XL Contracts

By Rate on Line

Rate on Line	Poisson λ	*ERD* *	Reinsurer Max Downside*
1.0%	0.5%	49.0%	19545%
2.0%	1.0%	48.0%	9678%
3.0%	1.5%	47.0%	6364%
4.0%	2.1%	46.0%	4651%
5.0%	2.6%	45.1%	3726%
7.5%	3.9%	42.6%	2373%
10.0%	5.3%	40.2%	1711%
12.5%	6.7%	37.9%	1315%
15.0%	8.1%	35.6%	1051%
20.0%	11.1%	31.0%	723%
25.0%	14.2%	26.6%	530%
30.0%	17.5%	22.3%	402%
40.0%	24.6%	14.2%	246%
50.0%	32.4%	6.6%	157%

* Ratio to expected premium

Assumptions.
- One reinstatement of limit for 100% A.P.
- Investment income effects ignored
- Poisson model with parameter λ
- Expected loss ratio 50%

In fact, using conservative assumptions, contracts having the same structure as the standard property catastrophe treaty[24] can be shown to exceed the $ERD \geq 1\%$ threshold (as well as a supplemental minimum potential downside threshold) if the upfront rate on line $ROL \leq 50\%$. Table 7 summarizes the *ERD* and potential downside values (ignoring investment income) for contracts having rates on line ranging from 1% to 50%, based on the simplifying assumptions that the expected loss ratio is 50%, all claims are total limit losses and that claims are Poisson distributed. On the basis that every rate on line in Table 7 easily passes the *ERD* test even without the supplemental downside requirement, we suggest that any reinsurance contract having this structure be deemed to meet the requirements for "significant" risk transfer. Clearly, such contracts are subject to the "significant" risk transfer requirement, but because we have, in effect, pre-qualified them as a class, the requirement to demonstrate significant risk transfer can be waived.

Example 5.2: Primary Quota Share Reinsurance

We applied the *ERD* test to the primary quota share contract described in Example 3.2. Again assuming a one-year net claim payment lag[25], a 5% interest rate and a lognormal σ of 15%, we calculated the frequency and severity, respectively, of present value underwriting loss to be 21.53% and 6.91%, which corresponds to an *ERD* of 1.49%[26]. This *ERD* value surpasses the $ERD \geq 1\%$ standard. Moreover, because there is no limit on the reinsurer downside potential, it would meet the suggested supplemental requirement. Therefore, this contract meets the "significant" risk transfer requirement.

Example 5.3: Primary Quota Share Reinsurance (Volatility of S&P 500)

In this example we test the same quota share that was the subject of Example 3.3. That quota share covered an insurance portfolio with the same loss ratio volatility as an S&P 500 index investment. The ceding commission is 25%. The frequency, severity and *ERD*

[24] The standard property catastrophe treaty provides two loss limits, the second one paid for with a contingent "reinstatement" premium at the same rate on line as the first one.

[25] Using this simplifying assumption, we can focus on the present value of the losses only, measured at the time the premium is received, because the present value factor applicable to premiums and losses for the period up to the premium receipt date is the same. The ratio of discounted *ERD* to discounted premium using the full claim and premium payment lags is equal to the ratio of discounted *ERD*, using the net claim lag, to undiscounted premium.

[26] If the prospective loss ratio is lognormally distributed, $ERD = PV[E(x) \cdot N(d1) - FV(B) \cdot N(d2)]$, where N is the normal cdf, $d1 = [ln(E(x)/FV(B)) + 0.5\,\sigma^2]/\sigma$ and $d2 = d1 - \sigma$.

characteristics of such a portfolio are summarized in Table 8 for the two volatility scenarios modeled in Example 3.3. For volatility of 13.85% the $ERD \geq 1\%$ standard is met. However, at the historically low volatility of 9%, a portfolio with S&P 500 volatility characteristics has an *ERD* of only 0.28% and thus fails the $ERD \geq 1\%$ standard by a wide margin. That creates a conundrum – is it ever reasonable to consider the S&P 500 to be without risk? If not, a 1% threshold for *ERD* is too high.

TABLE 8

ERD Risk Transfer Analysis for Quota Share in Example 5.3
Given Portfolio with Volatility of S&P 500

σ	*Freq*	*Sev*	*ERD*
9.00%	8.8%	3.2%	0.28%
13.85%	17.9%	6.0%	1.07%

Next, we will use the *ERD* test to assess quota share contracts with features such as loss ratio caps and corridors that reduce the loss exposure of the reinsurer. These features appear frequently in traditional reinsurance contracts as a means of making otherwise unattractive treaties acceptable to the reinsurance market.

Example 5.4: Reinsurance with 25% Ceding Commission and 5-Point Loss Ratio Corridor

Table 9 shows the downside risk measures *Freq*, *Sev* and *ERD* for a quota share or excess contract that provides a 25% ceding commission and requires the ceding company to retain any losses that fall within a five point loss ratio corridor from 75% to 80%. We assume the prospective loss ratio is lognormally distributed, with a mean of 70% and a range of values for σ. Claim payments are assumed to lag premium payments by one year.

Table 9 shows that for lower volatility business, represented here by lognormal σ values of 10% and 15%, a treaty with the 5 point loss ratio corridor removes enough risk from the deal that the *ERD* falls below 1%, indicating that the risk transfer is not significant. For the σ values of 25% and higher, the *ERD* significantly exceeds the 1% threshold. Clearly, the

effect of a loss ratio corridor depends on the characteristics of the reinsured business, and in some circumstances such treaty feature is entirely appropriate.

TABLE 9

ERD Risk Transfer Analysis for Contract With 25% Ceding Commission and Loss Ratio Corridor from 75% to 80%

σ	*Freq*	*Sev*	*ERD*
10%	3.1%	3.2%	0.10%
15%	9.1%	6.0%	0.59%
20%	15.6%	9.2%	1.43%
25%	19.7%	12.6%	2.47%
30%	22.4%	16.2%	3.63%
40%	25.6%	23.9%	6.13%
50%	26.9%	32.4%	8.74%

Example 5.5: Reinsurance with 25% Ceding Commission and 95% Loss Ratio Cap

We now consider the effect of an aggregate loss ratio cap of 95% (instead of a loss ratio corridor) on the same subject matter business discussed in Example 5.4. Table 10 shows frequency, severity and *ERD* for σ values ranging from 10% to 50%. Except for the case of $\sigma = 10\%$ (where *ERD* =0.41%) the aggregate loss ratio cap is at a high enough level that the 1% threshold is exceeded, and for the higher values of σ by a wide margin.

Note that in the case of $\sigma = 10\%$, the *ERD* associated with a contract with no loss ratio cap is also 0.41%, indicating that the cap at 95% has no significant effect on the risk transferred to the reinsurer. On that basis, the contract with a 95% cap transfers

"substantially all" the risk in the underlying portfolio, and even though it does not transfer "significant" risk, it meets the risk transfer requirements of FAS 113.

TABLE 10

ERD Risk Transfer Analysis for Contract With 25% Ceding Commission and Loss Ratio Cap of 95%

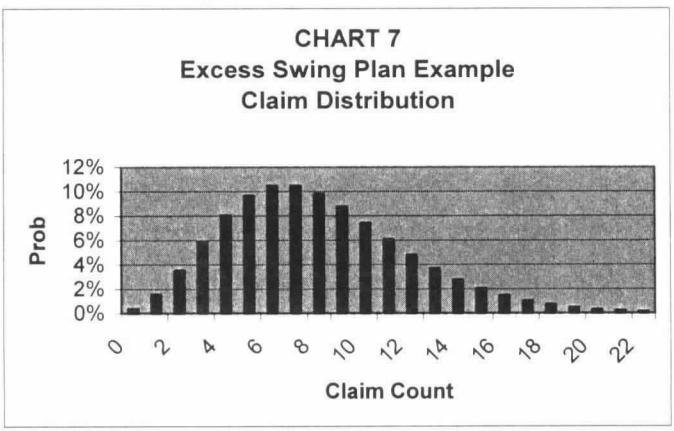
σ	*Freq*	*Sev*	*ERD*
10%	11.0%	3.8%	0.41%
15%	19.5%	6.5%	1.27%
20%	24.5%	8.9%	2.18%
25%	27.6%	10.7%	2.94%
30%	29.4%	12.0%	3.53%
40%	31.1%	13.8%	4.29%
50%	31.4%	14.9%	4.69%

Example 5.6: Excess Swing-Rated Reinsurance

It is common for "working layer" excess of loss reinsurance to be structured on a "swing-rated" basis, which means the premium is based in part on the losses ceded to the treaty. Typically, the premium formula calls for ceded claims to be multiplied by a loading factor to reflect a margin for the reinsurer, subject to a minimum and maximum. In primary insurance this structure is known as a "retrospective experience rating plan". The purpose of such plans is to allow the ceding company to fund its own excess claims up to the point beyond which it would become too painful and to cede the excess claims beyond that point to the reinsurer. To the extent that the excess claims experience is good, the ceding company benefits from a lower rate. Reinsurers often like these plans because they provide strong

incentives, both positive and negative, to the ceding company to minimize excess claims. Ceding companies often find these plans attractive because they believe their realized rate will be significantly less than under a flat-rated plan.

While minimizing risk transfer is not usually the driving force behind the structuring of a swing plan, such a structure typically does transfer less risk than a flat-rated excess of loss treaty covering the same business. To illustrate this, suppose the expected excess losses are $4 million. If the total premiums on the subject portfolio are $50 million, this can be expressed as a loss cost of 8%. For the sake of discussion let's assume the excess claim count can be modeled using a negative binomial distribution with an mean of 8 claims[27] and that only total limit claims are possible. The claim distribution is shown graphically in Chart 7.



Suppose the swing plan calls for an excess reinsurance premium equal to excess claims times 100/80, subject to a minimum of 4% of subject premiums and a maximum of 16%. That results in the excess rate distribution shown in Chart 8. The expected value of the premium rate under this plan is 9.71%. The alternative is a contract with a flat rate of 11.43%.

[27] Specifically, using the Microsoft Excel function for the negative binomial probability, Prob(COUNT)= NEGBINOMDIST(COUNT, 8, 0.5)

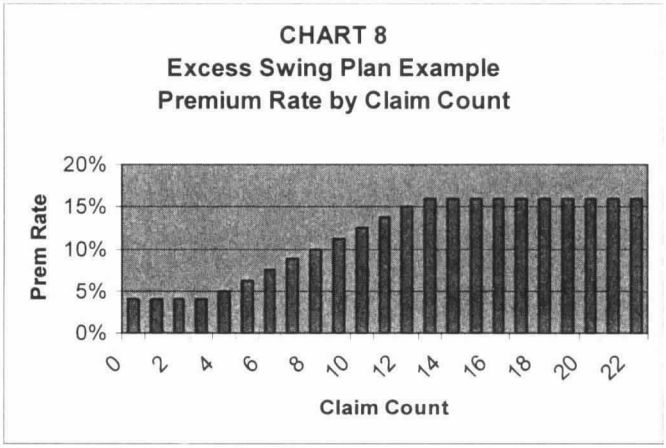


Table 11 summarizes the *ERD* analysis for both the flat-rated and swing-rated plans, assuming that there is a negligible claim payment lag. It shows that the swing plan has an *ERD* of 0.97%, just under the 1% threshold for significant risk. With some minor restructuring this contract would be able to pass the *ERD* test. In contrast, the flat-rated plan has an *ERD* of 4.70%, which is well above the threshold. Note that the mean severity of loss faced by the reinsurer is greater in the case of the swing plan than in the flat-rated plan, but because the probability of loss is much lower, the swing plan *ERD* falls below the threshold for "significant" risk. This is a good illustration of why severity (*TVaR*) by itself is an unreliable indicator of risk.

TABLE 11

ERD Risk Transfer Analysis

Swing-Rated vs. Flat-Rated Excess

Plan	Rate	*Freq*	*Sev*	*ERD*
Swing	9.71%	3.2%	30.4%	0.97%
Flat	11.43%	18.0%	26.2%	4.70%

Example 5.7: Individual Risks

One of the well known drawbacks of the "10-10" test is that if it were applied to individual insurance contracts or facultative reinsurance contracts, it would in almost all cases indicate that they do not contain "significant" risk, which strikes virtually everyone as unreasonable. In this example, using simplifying but not unreasonable assumptions we will show that the *ERD* test correctly identifies individual risk contracts as containing significant risk.

We assume that a portion of the premium for every individual risk contract is attributable to the potential for a limit loss. Since it is very large losses rather than partial losses that are most likely to put the insurer or reinsurer into deficit, we will ignore the potential for small losses and focus on limit losses. Let's assume that the pure premium for total limit losses is 10% of the total premium. Since a limit loss can occur only once in a policy period, let's assume the probability of such a loss is Bernoulli distributed with a probability equal to this 10% times the total premium rate on line (i.e., the total premium divided by the limit). From that we can calculate the *ERD* and the maximum downside potential.

The results are shown in Table 12 for rates on line ranging from 0.5% up to 83.33%. We see that any individual risk paying a rate on line of less than 83.33% would exceed a $ERD \geq 1\%$ standard for "significant" risk. We display such a wide range of rates on line, because we want to show that virtually all individual risks, ranging from personal lines policies to large commercial policies with a high level of premium funding, can be shown to meet the "significant" risk requirement using the *ERD* test.

Above a rate on line of 83.33%, the maximum downside falls below 20% of premium, which is a potential threshold for our proposed minimum downside requirement. Thus, individual risks with rates on line above 83.33% would fail to show "significant" risk. While this is a highly idealized example and further research would be appropriate to refine the methodology, we believe it is sufficiently realistic to "pre-qualify" virtually all individual risk contracts as containing significant risk and thus make it unnecessary to test them individually.

TABLE 12

ERD / Max Downside
For Individual Risk Contracts

By Rate on Line

Rate on Line	Limit Loss Prob	*ERD*	Reinsurer Max Downside
0.5%	0.05%	9.95%	19900%
1.0%	0.10%	9.90%	9900%
2.5%	0.25%	9.75%	3900%
5.0%	0.50%	9.50%	1900%
10.0%	1.00%	9.00%	900%
25.0%	2.50%	7.50%	300%
50.0%	5.00%	5.00%	100%
75.0%	7.50%	2.50%	33%
83.3%	8.33%	1.67%	20%

Assumptions.
- Investment income effects ignored
- Bernoulli probability of limit loss
- Total limit loss ratio 10%

5.1 Section Summary

In this section we have shown that the *ERD* test produces mostly reasonable results when applied to a variety of reinsurance structures covering insurance portfolios having a wide range of risk characteristics. Using the $ERD \geq 1\%$ standard together with reasonable contract assumptions we have demonstrated that catastrophe excess of loss reinsurance and individual risk contracts generally contain significant risk, which is a common sense result that eludes the "10-10" test. We also showed that loss ratio corridors and loss ratio caps are acceptable under some circumstances but not under others, and similarly that swing-rated excess reinsurance must be structured with care to ensure that it transfers significant risk while still meeting the reinsurer's and ceding company's other goals. The only unreasonable result we produced was that a quota share contract with a ceding commission of 25% and the prospective volatility characteristics of the S&P 500 (as measured by VIX) does not always meet the "significant" risk requirement. VIX has ranged as low as 9% in the period

since 1990. Volatility parameters below about 13% produce *ERD* results (in the quota share we tested) that suggest insignificant levels of risk. This is an anomalous result because it suggests that under some circumstances an investment related to the S&P 500 index should not be considered risky, a conclusion that does not seem reasonable.

In summary, given these results and the findings in Section 4, we conclude that:

1. The *ERD* methodology described here, with a 1% threshold for significant risk transfer, is numerically comparable to the "10-10" benchmark;
2. The *ERD* methodology is qualitatively superior to that benchmark; and
3. If the 1% *ERD* method were adopted as a de facto standard replacing the "10-10", we would consider that a significant improvement.

6. IDENTIFICATION OF CONTRACTS SUBJECT TO "SIGNIFICANT" RISK REQUIREMENT THAT DO NOT REQUIRE INDIVIDUAL TESTING

Apart from those contracts for which it can be demonstrated that they transfer "substantially all" the risk inherent in the underlying insurance policies, all purported reinsurance contracts are subject to the requirement that they transfer "significant" risk. Unless a contract is tested, it is impossible to know whether or not it meets the requirement. However, the implication that it is necessary to test every single reinsurance contract is daunting. For many ceding companies buying excess of loss reinsurance, it might even be impossible. Ceding companies often buy excess coverage not only to transfer risk but also to obtain pricing for excess exposure they themselves do not fully understand, which they can factor into their own insurance rates. Under such circumstances, to ask ceding companies to model such exposure to demonstrate compliance with FAS 113 seems unreasonable.

Ideally, we would like to find a way to partition the set of all reinsurance contracts subject to the "significant" risk requirement into the subset containing those that we can reasonably expect will pass if they were tested and the subset comprising all other contracts. The former subset would be exempt from individual testing, while the latter subset would have to be tested individually. The purpose of this section is to begin to identify elements of the first subset of contracts that do not require individual testing.

Example 6.1: Individual Risk and Catastrophe Excess of Loss Contracts

In Section 5 we showed that 1) standard catastrophe excess of loss contracts and 2) individual risk contracts, generally possess *ERD* characteristics that indicate these two classes of contracts meet the "significant" risk requirement, and that it is therefore unnecessary to test contracts within those classes individually.

Example 6.2: Other Excess of Loss Contracts

By virtue of analysis similar to that for individual risk and catastrophe excess of loss contracts, it is possible to add a further large subset of excess of loss contracts (treaty and facultative) to the category of contracts that do not require individual testing. Table 13 summarizes the *ERD* analysis for excess of loss contracts with no ceding commission and rates on line ranging from 1% to 500% and aggregate limits no less than one full limit or 200% of premiums, whichever is greater. The term "rate on line" is most frequently used in connection with catastrophe excess of loss treaties and other excess contracts where the ratio of premium to limit[28] is far less than 100%, so a rate on line of 500% might be surprising. However, it is common for "working layer" excess of loss contracts to be priced with the expectation that there will be between several and many claims during the coverage period. Under typical pricing assumptions, a 500% rate on line implies the expectation that excess claims will be equivalent to about three total limits losses.

Our analysis assumes a Poisson distribution for claim frequency and that all claims are limit losses. Theoretically, we should use a negative binomial, but because that makes the tail fatter and thus easier to pass the *ERD* test, the Poisson assumption is conservative. We assume an expected loss ratio of 70%, another conservative assumption. In a competitive market the expected loss ratio can be expected to be higher, especially for the higher rate on line business. We assume an interest rate of 5% and a 5-year claim payment lag (which makes this analysis suitable for reasonably long tail as well as short tail business).

On the basis that every rate on line in Table 13 from 1% to 500% passes the *ERD* test even without the supplemental downside requirement coming into play, we suggest that any excess of loss contract having this structure (and no loss sensitive or other features that might call the contract's status into question) be deemed to meet the requirements for

[28] Note that the limit used in the denominator is the risk or occurrence limit, depending on the coverage, not the aggregate limit except in the case of aggregate excess coverage.

TABLE 13

Expected Reinsurer Deficit / Max Downside

For Long/Short Tail XL Contracts with
Aggregate Limit ≥ One Limit or 200% Loss Ratio

By Rate on Line

Rate on Line	Poisson λ	Expected Reinsurer Deficit*	Reinsurer Max P.V. Downside*
1.0%	0.7%	54.0%	7735%
2.5%	1.8%	52.6%	3034%
5.0%	3.5%	50.5%	1467%
10.0%	7.0%	46.2%	684%
15.0%	10.5%	42.1%	422%
25.0%	17.5%	34.3%	213%
50.0%	35.0%	16.7%	57%
75.0%	52.5%	6.9%	57%
100.0%	70.0%	8.8%	57%
200.0%	140.0%	5.0%	57%
300.0%	210.0%	2.9%	57%
400.0%	280.0%	1.8%	57%
500.0%	350.0%	1.3%	57%

* Ratio to premium

Assumptions.
- Loss cap of greater of one limit or 200% L/R
- No ceding commission
- Poisson model with parameter λ
- Claim payment lag 5 years
- Interest rate 5% per annum
- Expected loss ratio 70%

"significant" risk transfer. Excess of loss contracts with no aggregate limit clearly fall into this category as well. All such contracts are subject to the "significant" risk transfer

requirement. However, because we have, in effect, pre-qualified them as a class, the requirement to demonstrate significant risk transfer can be waived.

Example 6.3: Contracts with Expected Loss Ratios Above a Minimum Permissible Loss Ratio Threshold

There is a further general approach to expanding the set of contracts subject to "significant" risk testing that do not need to be tested individually. In Section 3 we noted that one unreasonable implication of the "10-10" test is a cap on reinsurance pricing at such a low level that, if it were enforced, would likely lead to a reduction of reinsurance capacity. The $ERD \geq 1\%$ standard we have proposed also implies a cap on reinsurer margins. Fortunately, the *ERD* standard we have illustrated implies a significantly higher maximum permissible present value margin for the reinsurer than the "10-10" test does.

Table 14 shows maximum permissible present value margins and corresponding minimum permissible loss ratios implied by $ERD \geq 1\%$ for claim lags of zero, one year, two years and three years with respect to contracts for which the prospective loss ratio can be modeled using a lognormal distribution[29]. The results are shown for σ values ranging from 9% to 100%. Note that for each value of σ, the permissible loss ratios increase in nominal terms with the claim lag, but the present values are all the same. The allowable margins for the σ values at the low end of the range might make reinsurance of such low risk portfolios impossible unless the reinsurance is structured to meet the "substantially all" risk transfer test. For example, the maximum permissible present value margin for $\sigma = 9\%$ of only 7.1%, while much higher than the 1.6% permitted under "10-10"[30], does not allow a reinsurer much, if any, upside potential, after deducting brokerage and internal expenses. That is one reason to consider the possibility that an *ERD* threshold of 1% might be too high. On the other hand, in light of our discussion in Section 3 about parameter uncertainty, it might turn out to be the case that realistic prospective estimates of σ will, in practice, generally exceed the low end of the range, making this concern irrelevant.

[29] Where the lognormal assumption is not appropriate, similar tables could be constructed for other loss ratio models.

[30] See Table 6. It is worth noting that the $ERD \geq 3\%$ mentioned in the 2002 VFIC paper as a possible threshold would result in an even lower maximum permissible present value margin of 1.2%! A threshold of 3% is clearly too high.

TABLE 14

Maximum Margins / Minimum Permissible Loss Ratios
Implied by $ERD \geq 1\%$

Contracts with No Ceding Commission
Interest at 5% per annum

Tabulated by σ and Claim Lag

	Max	Minimum Permissible Loss Ratio			
	P.V.	Lag 0	Lag 1	Lag 2	Lag 3
σ	Margin	Yrs	Yr	Yrs	Yrs
9.0%	7.1%	92.9%	97.5%	102.4%	107.5%
10.0%	8.4%	91.6%	96.2%	101.0%	106.0%
11.0%	9.7%	90.3%	94.8%	99.6%	104.6%
12.0%	11.0%	89.0%	93.5%	98.2%	103.1%
13.0%	12.3%	87.7%	92.1%	96.7%	101.6%
14.0%	13.6%	86.4%	90.8%	95.3%	100.1%
15.0%	14.9%	85.1%	89.4%	93.9%	98.6%
20.0%	21.3%	78.7%	82.7%	86.8%	91.1%
25.0%	27.4%	72.6%	76.2%	80.0%	84.0%
30.0%	33.2%	66.8%	70.1%	73.6%	77.3%
40.0%	43.7%	56.3%	59.1%	62.1%	65.2%
50.0%	52.6%	47.4%	49.8%	52.2%	54.9%
60.0%	60.1%	39.9%	41.9%	44.0%	46.2%
75.0%	69.1%	30.9%	32.5%	34.1%	35.8%
100.0%	79.5%	20.5%	21.6%	22.6%	23.8%

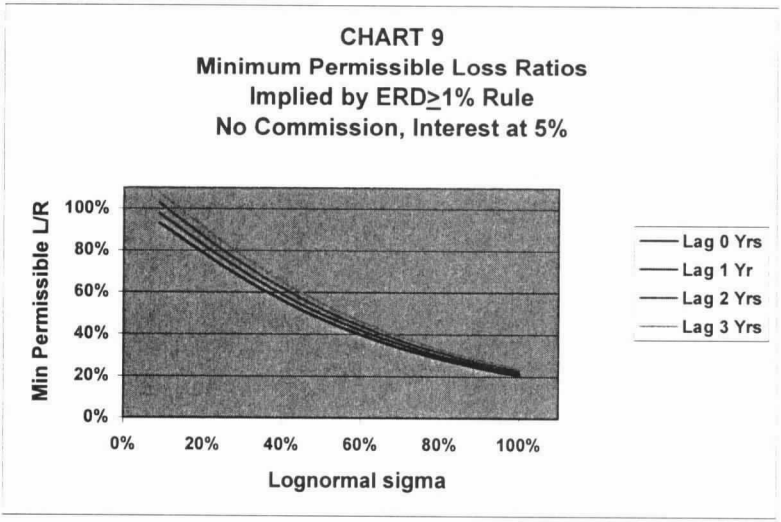
The maximum margins implied by $ERD \geq 1\%$ for larger values of σ seem more reasonable. For example, for $\sigma = 30\%$, the allowable present value margin is 33.2%, which is a more reasonable ceiling[31].

The implication of this for our present discussion is that if a contract with no ceding commission is priced to an expected loss ratio that is greater than the minimum permissible loss ratio shown for the relevant σ and claim lag (and the other assumptions are reasonable), then the contract will meet the $ERD \geq 1\%$ standard that indicates significant risk transfer. We present this as an illustration of how the subset of contracts that do not

[31] In contrast, a threshold of $ERD \geq 3\%$ implies a maximum permissible present value margin of 22.0%, which is about the same as that implied by "10-10".

require detailed testing for significant risk transfer could be expanded beyond the catastrophe excess of loss, individual risk and other excess of loss contracts we identified earlier. Any contract that is priced to an expected loss ratio that exceeds the minimum permissible loss ratio would be exempt from individual testing. Additional research is necessary to fully realize this approach.

Chart 9 shows the minimum permissible loss ratios in Table 14 graphically.



Example 6.4: Contracts with Immaterial Premiums

Contracts or programs that involve the cession of small amounts of premium should be exempt from individual testing, unless there is reason to suspect that they might materially distort either the ceding company's or reinsurer's financial statements. A reasonable definition of small might be the smaller of $1 million and 1% of total gross premiums. The rationale for this exception is that small premium cessions by definition have a very limited impact on either party's financial statements. Any distortion resulting from minimal risk transfer below the significance threshold would be immaterial.

7. POSSIBLE EVOLUTION OF RISK TRANSFER MEASUREMENT

The context of the paper is risk transfer testing. However, the notion of risk transfer is also integral to the pricing of insurance and reinsurance products. Risk transfer is what gives rise to risk premiums and the potential for profit. Many methods already exist for explicitly or implicitly adding a profit load to a reinsurance contract. It seems reasonable that a risk loading method used to determine needed profits could be turned into a risk transfer test as well. Although this paper does not address the issue directly, the *ERD* risk transfer test described in earlier sections of this paper measures tail value at risk (*TVaR*), which is a valid method for producing risk and profit loads. In fact, given the coherent nature of *TVaR*, it is considered a superior method for risk loading by many practitioners.

At least one major insurance company has used the *ERD* framework in pricing and enterprise risk management for several years, in the form of the *risk coverage ratio* (*RCR*) described by Ruhm [6]. In practice, that risk measure has produced results for the company that are reasonable and consistent across a broad variety of actual risks, due in large part to its good technical properties and its relative transparency.

As noted before, this working party is not endorsing any single specific method for risk transfer testing. Thus, rather than doing more work on our *ERD* example to show its full implications for risk loading, we will show another (much briefer) example here where risk loading and risk transfer testing are tightly linked.

The approach we examine here is based on the *right tail deviation* (*RTD*), a framework proposed by Wang and developed from concepts he has written about extensively [7] [8].

For a given aggregate distribution function $F(x)$ (derived from some convolution of frequency and severity distributions), we transform the distribution using the following formula:

$$F^*(x) = 1 - \sqrt{1 - F(x)} \tag{7.1}$$

Because $0 < F(x) < 1$ for all x, it is fairly easy to see that $F^*(x) < F(x)$ for all x, which implies the following expected value relationship:

$$E^*(x) \geq E(x) \tag{7.2}$$

The interpretation is that the transform has "loaded" the original distribution for risk. The difference between E^* and E is the risk load, for any layer of the distribution. Thus, we can use E^* instead of E to represent a fully risk loaded pure premium. The reason this approach is appealing is that the transformed distribution is itself another loss distribution, meaning that all the ordinary mathematics of loss distributions carry over. Relating this to financial mathematics, it is generally assumed that assets like equities are themselves transformed distributions, although this is not usually explicitly stated. The transform in the financial economic model is the so-called state price, which enforces no-arbitrage pricing [9].

If one wants to think about the risk load independently, it is easily captured as:

$$RTD(x) = E^*(x) - E(x) \tag{7.3}$$

Under this approach, the risk load *RTD* might be adjusted (i.e. multiplied) by some constant factor α to produce the final profit load. Note that Wang has generalized this model to consider other exponents of transformation (i.e. instead of just the power of 0.5, any power between 0 and 1 exclusive).

There are a couple of ways in which the *RTD* could be used to devise a risk transfer test. One way would be to treat αRTD as the maximum permissible reinsurer's margin consistent with "significant" risk transfer. That is essentially the same approach that was described in Example 6.3. The difference is that in that example, we derived the risk load consistent with a "significant" risk transfer threshold of $ERD \geq 1\%$, whereas here we would determine the risk load component αRTD first and then effectively determine the risk transfer threshold that is consistent with it.

A second way would be to devise a risk transfer test that compares the full premium (not just the margin) with a multiple of αRTD using the following procedure, which is similar to one outlined by Wang:

1. Compute expected loss of the contract under the untransformed distribution $F(x)$;
2. Note the premium for the deal (however computed—allows for market pricing);
3. Compute *RTD* for the deal using the transformed distribution and formula (7.3);

4. Define the *maximum qualified premium* as some multiple of *RTD* (Wang suggests 3-5x[32]);
5. The "significant" risk transfer threshold is defined as "*maximum qualified premium* ≥ *premium*"[33].

We will look at two examples of this approach. The first is the catastrophe excess of loss contract described in Examples 3.1 and 5.1. The second example addresses a questionable scheme for creating a reinsurance structure that apparently meets the "significant" risk transfer requirement by combining two unrelated coverages to produce just enough risk transfer to pass. This is an important example, because this method separates the reinsurance premium into higher risk and lower risk components and thus has potential to identify highly structured reinsurance contracts that satisfy other quantitative tests but do not meet the spirit of FAS 113[34].

Example 7.1: Property Catastrophe Excess of Loss Reinsurance

If we apply the *RTD* qualified premium approach to the property catastrophe excess of loss example discussed in Examples 3.1 and 5.1, we see that the contract easily meets this *RTD*-based risk transfer requirement. Table 15 shows the catastrophe loss distribution originally shown in Table 3 with an additional column for the "transformed" probability based on the $F^*(x)$ determined from formula 7.1. $E^*(x)$, expressed both in terms of premiums and limit, is shown at the bottom of the table as 203% and 20%, respectively.

[32] The issue of the appropriate multiplier of *RTD* warrants further research. A multiple of 4 appears to imply that traditional quota shares like those discussed in Examples 3.2 and 3.3 do not contain significant risk transfer, which suggests the effective threshold may be set too low.

[33] Wang has a suggested giving partial credit in cases where the maximum qualified premium is less than the actual reinsurance premium. However, we prefer to focus on the risk characteristics of the contract as a whole.

[34] This comes at the cost of some complexity. The subdivision into risky and less risky components depends on the values chosen for α, the multiplier for αRTD, and the exponent in formula (7.1), choices that are made more difficult by the fact that it is difficult to ascribe an intuitive meaning to these parameters.

TABLE 15 Catastrophe Loss Distribution Example 7.1			
Loss as % of Limit	Loss as % of Premiums	Actual Probability of Given Loss	Transformed Probability* of Given Loss
0%	0%	67%	43%
5%	50%	20%	21%
10%	100%	10%	19%
100%	1000%	3%	17%
5%	50%	100%	100%
20%*	203%*		

In terms of premium, *RTD*=203%-50%=153%. Using a multiplier of 4x, the "qualified" premium proportion is 612%, which is well in excess of the threshold of 100% required for significant risk transfer.

Example 7.2: "Highly Structured" Mix of Low Risk and High Risk Portfolios

We now move on to the example of potential manipulation. In this case, the deal structure consists of a base portfolio with very little risk mixed with a highly risky catastrophe layer. The overall structure is designed to barely pass risk transfer using the "10-10" criterion.

The low risk portfolio has expected losses of $8 million with lognormal σ value of only 1%. To maximize the low risk nature of this portfolio, its premium is $8 million—no load for expense or profit at all.

The catastrophic portfolio we add to this deal is a $1.6 million layer with a 12.5% chance of loss. For simplicity, if a loss occurs, it is a total loss. Thus, the expected loss for this piece is $200,000. Let's assume the premium is $500,000, for a 40% expected loss ratio.

First, let us consider the two pieces separately. The low risk portfolio has an untransformed expected loss of $8 million and a transformed expected loss of $8.1 million. The maximum qualified premium is only $0.4 million, leaving $7.6 million unqualified. This piece falls far short of the "significant" risk standard.

The catastrophic portfolio has an untransformed expected loss of $200,000 and a transformed expected loss of $666,000. The maximum qualified premium is well in excess of the actual premium of $500,000, thus easily meeting the *RTD*-based "significant" risk standard.

Now consider the combined distribution. The combined contract has a premium of $8.5 million. A 10% loss over this would be an attachment of $9.35 million, and the probability of this occurring is 12.5% (very close to the cat loss alone, of course). Thus, this contract passes the "10-10" test. But Wang's method gets closer to the truth. The transformed expected losses are only $8.65 million vs. $8.2 million untransformed, producing maximum qualified premiums of only $1.8 million, leaving $6.7 million unqualified, well short of the 100% required for "significant" risk transfer.

Note that this method penalizes the combination even more than the sum of the components (the *RTD* of the combined deal is $450,000, whereas the sum of the *RTD*s of the two deals is about $570,000)[35]. It is not clear whether this phenomenon, i.e., the *RTD*-based approach of the highly contrived structure being less than sum of the *RTD* of the separate components, represents the general case. However, it does suggest the intriguing possibility that this approach could perhaps be developed into a quantitative test to detect reinsurance structures that appear to pass certain quantitative threshold, but which do not meet the spirit of FAS 113.

This is as far as we will pursue the *RTD* ideas here. The *RTD* approaches have some appeal and added properties that the *ERD* method does not, at the cost of increased complexity. As noted previously, the working party is not specifically advocating any particular method. This example shows that other methods could be used instead of the *ERD* example that we have examined in some detail. Ultimately, a combination of market and regulatory factors will determine what methods are actually deployed.

[35] This is due to the diversification of the combined deal, which is of course the correct treatment.

8. SUMMARY

The purpose of this paper has been to contribute constructively to the thinking about what should be understood by the term "risk transfer" in the context of FAS 113 by framing a comprehensive response to the four questions posed by COPLFR.

In particular, we have responded to the first two questions by describing two approaches for assessing the significance of risk transfer that are superior to the "10-10" test that is in common use. The first approach, which we have described and illustrated in detail, is based on the expected reinsurer deficit (*ERD*). The second approach, which we outline more briefly, is based on the concept of right tail deviation (*RTD*). We have responded to the third "safe harbor" question in two parts. First, we have described a framework for determining whether a purported reinsurance contract meets the FAS 113 risk transfer requirement by virtue of the cession of "substantially all" of the underlying insurance risk to the reinsurer. Second, we have begun to identify groups of contracts that are subject to the "significant" risk requirement of FAS 113, but which can be exempted from detailed individual testing, because we have established that contracts falling within the group can reasonably be expected to pass the "significance" test, if they were actually tested.

In particular, the following classes of contracts fall into the category of transferring "substantially all" of the original insurance risk, unless they include features that reduce the reinsurer's *expected underwriting deficit* (*EUD*) below that which the cedent would face on its unreinsured portfolio:

- Proportional facultative reinsurance with effective ceding commissions no less than cedent expenses;
- Proportional treaties with effective minimum ceding commissions no less than cedent expenses;
- Proportional facultative or treaty reinsurance for which it can be shown that the reinsurer's *EUD* is essentially the same as the cedent's *EUD* on the unreinsured subject portfolio, irrespective of whether the contract includes a loss ratio corridor, loss ratio cap or other risk mitigating feature;
- Excess of loss facultative or treaty reinsurance for which it can be shown that the reinsurer's *EUD* is essentially the same as the cedent's *EUD* on the portion of the

original subject portfolio that is exposed to the same risks as the excess of loss contract;

- Whole account quota share contracts with loss ratio caps no lower than the point at which the ceding company would exhaust its surplus.

To address the question of how to measure "significant" risk transfer, we have proposed an *ERD* test as an improvement over the "10-10" test, which arose in the 1990s as a way to test "finite risk" reinsurance contracts for compliance with FAS 113. The "10-10" test was not originally intended to be applied to traditional reinsurance contracts, and usually it was not. In the wake of recent real and alleged reinsurance accounting abuses, there is an increasing sentiment that a wider class of reinsurance contracts beyond those classified as "finite" need to be tested for significant risk transfer. Because it has come into widespread use, the "10-10" test has become the de facto standard for reinsurance risk transfer testing, despite the fact that it has never been endorsed by any professional body nor subjected to serious critical scrutiny.

We have also addressed COPLFR's fourth question. Throughout the paper we have discussed the advantages of our described approaches over the "10-10" test that is commonly used today. We have demonstrated that "10-10" is inadequate for use as a universal risk transfer test, because it cannot correctly identify contracts that are clearly risky. We have proposed an improved alternative test based on the concept of the *expected reinsurer deficit*, or *ERD*, which incorporates both frequency and severity of underwriting loss into a single measure. The embedded severity measure is the *TVaR* at the economic breakeven point. *TVaR* has the advantages over *VaR* of reflecting all the information in the right tail of the underwriting result distribution as well as being a coherent measure of risk.

We have shown that the proposed $ERD \geq 1\%$ threshold correctly classifies as "risky[36] a quota share treaty that has the loss ratio volatility characteristics of the S&P 500 stock index. This is important because the standard for assessing reinsurance risk should be consistent with those in other financial markets.

We have also shown that low frequency-high severity reinsurance contracts (such as catastrophe excess of loss treaties) and high frequency-low severity contracts (such as traditional primary quota share treaties) pass the *ERD* test, provided loss mitigating features

[36] Provided the risk characteristics of the treaty are not too distorted by a large ceding commission.

such as loss ratio caps and/or corridors do not remove too much risk from the contracts (in which case a "failure" is entirely appropriate).

In summary, while we are not explicitly endorsing any single model or framework, because the *ERD* methodology described here (with a 1% risk transfer threshold) is numerically comparable to the current "10-10" benchmark and is superior in almost every way to that benchmark, if the 1% *ERD* method were adopted as a de facto standard replacing the "10-10", we would consider that a good outcome.

To address the concern in some quarters that the *ERD* test is not always stringent enough with respect to the potential for a large loss by the reinsurer, we have suggested consideration of a supplemental requirement that the reinsurer face a minimum downside potential of 15% or 20% of premiums.

Among contracts that are subject to the "significant" risk transfer requirement, under the "significance" standard embodied in $ERD \geq 1\%$ the classes of contracts listed below would not be subject to individual testing, because they have already been found to meet the requirement under very general conditions. It is therefore possible to say about contracts falling into the categories on the list below that the significance of their risk transfer is "reasonably self-evident". This is a preliminary list. We believe it may be possible to expand it considerably.

- Individual risk contracts;
- Short tail excess of loss treaties in the standard catastrophe excess structure, i.e., one reinstatement of the limit for 100% additional premium, with rates on line of up to 50%;
- Other excess of loss contracts with aggregate limits of no less than the greater of one occurrence (or risk) limit and 200% of premiums, no ceding commissions, and rates on line of up to 500%;
- Proportional and excess contracts having an expected loss ratio above the minimum permissible loss ratio implied by the $ERD \geq 1\%$ standard (or other standard as may be agreed);
- Contracts involving immaterial premiums.

Other contracts should be considered for significance testing, even if they appear to fall into one of the safe harbor categories, for the simple reason that they have greater potential to attract attention, and it is better to be prepared. This group includes, for example, 1) contracts involving large premium cessions, 2) those which, when accounted for as reinsurance, would substantially alter surplus or the ratio of premiums to surplus, and 3) contracts involving unusual structures, especially those that look contrived (e.g., a primary quota share combined with catastrophe protection on a different portfolio). Contracts in category 3 may be structured to narrowly meet the quantitative requirements for "significant" risk transfer, but they might still reasonably be disqualified on other grounds. Thus, a quantitative risk transfer test such as the *ERD* will not be adequate in all cases. However, we believe the *ERD* would do a good job of discriminating between contracts with significant risk and those without significant risk in all but cases involving contrived structures.

We have also pointed out that other risk transfer tests besides *ERD* can and should be considered, particularly in the context of reconciling risk transfer testing to the process of determining risk and profit loads. One such example, based on the *right tail deviation*, has certain desirable properties but comes at the cost of greater complexity. Other approaches could surely be used and should be the subject of future research.

It is important to remember that any risk transfer test requires a model of the prospective underwriting results and the related cash flows. In cases where there is relevant and credible loss experience, identifying a model is often straightforward, though it is always important to appropriately adjust the historical loss experience to prospective levels and to be conscious of the uncertainty in the model parameters. Where there is little or no relevant historical experience, the model must be chosen on the basis of the similarity of the subject portfolio to other ones with the same general characteristics. In such cases there will be greater uncertainty about the parameters, which should be reflected in the structure of the model.

9. SUGGESTED PRIORITIES FOR FURTHER RESEARCH

The *ERD* test proposed in this paper should be seen as an example of a reasonable framework for assessing the significance of risk transfer in reinsurance contracts. We have demonstrated that it is a clear improvement over "10-10", but we do not claim that it is the only reasonable approach. Indeed, we briefly described another promising, albeit more

complicated, method, namely, Wang's *RTD* framework. There may be others. We urge the CAS to encourage further research on this subject, perhaps through a call for papers.

We recommend the following research priorities in order to quickly arrive at a more effective assessment of risk transfer according to FAS 113 as well as to provide for continuing research in relation to future improvements.

9.1 Immediate "Level 1" Research – Consensus on Thresholds

1. Determination of an appropriate pass threshold for the comparison methodologies presented in Section 2 to determine whether or not "substantially all" of the insurance risk has been transferred. This may include determining a single applicable testing methodology (i.e., limiting the test to just one of the two methods presented);

2. Determination of an appropriate "pass" threshold framework for the *ERD* test presented in Section 4. In particular, is the 1% threshold illustrated in this paper appropriate, or would some other threshold be more appropriate? In addition, should there be a supplemental requirement that the reinsurer's potential loss be greater than or equal to some minimum amount? (We considered a minimum underwriting loss of 20% in some of our examples.);

3. Determination of the contract categories and financial characteristics of contracts that will not be required to be individually tested for "significant" risk transfer (because they have previously been analyzed and found generally to pass the significance test). This depends on item 2. Given a standard of $ERD \geq 1\%$, we demonstrated that individual risks, short tail excess of loss contracts in the standard catastrophe excess of loss structure within a certain rate on line range, other excess treaties within a certain rate on line range that have aggregate limits that are not too large, and other contracts with expected loss ratios above a minimum permissible loss ratio threshold, should not be required to be individually tested because we have determined they will pass if they were tested. It may be possible to expand that set of contracts "pre-qualified" for "significant" risk in that same way. If an *ERD* threshold different from 1% is adopted, the set of contracts that can be pre-qualified for "significant" risk may change.

9.2 On-Going "Level 2" Research – Other Methods

1. Continued research on methodologies and thresholds for determining whether or not "substantially all" of the insurance risk has been transferred;

2. Continued research for methodologies that assess risk transfer within the "reasonably possible" chance of a "significant" loss. As stated earlier, the Wang transformation could be one example of such a method;

3. Continued research into appropriate methods for incorporation of parameter uncertainty into models used for risk transfer testing.

Appendix A

Definition of Downside Risk Measures

Suppose B represents the amount of (present value) claims corresponding to the reinsurer's economic "breakeven" point, before taking into account brokerage and internal expenses (the FAS 113 definition):

$$B = P - C \tag{A.1}$$

where P represents the ceded premiums and C represents the ceding commissions payable on ceded premiums, if any. If $C = 0$, then the breakeven loss amount is equal to the premiums.

Let x denote the random variable for the prospective losses. (It may be more convenient in practice to work with loss ratios, but here we are using loss dollars.) Then the expected cost of FAS-113-defined present value loss scenarios PV($Loss > 0$) (which ignore all reinsurer expenses other than ceding commissions), also known as the present value expected reinsurer deficit or *ERD*, expressed as a dollar amount, is:

$$ERD = E[(PV(Loss) > 0)] = PV \int_{FV(B)}^{\infty} (x - FV(B)) \cdot f_x(x) dx \tag{A.2}$$

As the pure premium cost of underwriting loss scenarios, *ERD* is a measure of the reinsurer's underwriting downside risk[37].

[37] Note that the *ERD* is the expected present value of the contingent capital calls described by Mango [5].

The probability or frequency of the insurer incurring a present value loss $PV(Loss) > 0$ is:

$$Freq = \Pr ob[PV(Loss) > 0] = \int_{FV(B)}^{\infty} f_x(x)dx \tag{A.3}$$

The expected severity of underwriting loss, given $PV(Loss) > 0)$, is

$$Sev = E[(PV(Loss) \mid PV(Loss > 0)]$$

$$= \frac{\int_{FV(B)}^{\infty} (x - FV(B)) f_x(x)dx}{\int_{FV(B)}^{\infty} f_x(x)dx}$$

$$= \frac{ERD}{\Pr ob[PV(Loss > 0)]} \tag{A.4}$$

Note that *Sev* is the Tail Value at Risk (for present value underwriting loss) described by Meyers [4] as a coherent measure of risk and by the CAS Valuation, Finance, and Investments Committee [1] for potential use in risk transfer testing of finite reinsurance contracts. Meyers (p. 239) gives the following formula for $TVaR_\alpha$:

$$TVaR_\alpha = VaR_\alpha + \frac{EPD(VaR_\alpha)}{1-\alpha} \tag{A.5}$$

At the present value breakeven loss point B, $\alpha = F_x(B) = \int_0^{FV(B)} f_x(x)dx$. The present value loss at the breakeven loss is zero, implying $VaR_\alpha = 0$. That leaves only the second term. Because $EPD(VaR_{F_x(B)}) = ERD$ and $1-\alpha = 1-F_x(B) = \Pr ob[PV(loss > 0)]$, when the variable of interest is present value underwriting loss, (A.5) equates to formula (A.4).

For a quota share with no loss ratio caps or corridors, the reinsurer's loss ratio is identical to the ceding company's loss ratio on the subject portfolio and their distributions are identical[38]:

$$f_x(x) = f_y(y)$$

If there are no loss ratio caps or corridors, it is often still convenient to express the random variable x for the reinsurer's loss ratio in terms of the subject portfolio's loss ratio random variable y. For example, given a 5-point loss ratio corridor between 75% and 80% with respect to the subject portfolio, the reinsurer's loss ratio $x(y)$ is:

$$x(y) = \begin{cases} y & \text{if } y \le 75\% \\ 75\% & \text{if } 75\% < y < 80\% \\ y - 5\% & \text{if } y \ge 80\% \end{cases}$$

In this case, given $B = 75\%$, formula (A.2) for ERD would be expressed in terms of y as follows:

$$ERD = PV \int_{FV(B_y)}^{\infty} (y - FV(B_y)) \cdot f_y(y)dy$$

$$ERD = PV \int_{FV(80\%)}^{\infty} (y - FV(80\%)) \cdot f_y(y)dy$$

where $B_y = B + 5\%$. Similarly, Formulas (A.3) for frequency and (A.4) and severity can be expressed in terms of y.

[38] Because it is easier to compare the cedent and reinsurer positions if we use loss ratios rather than loss dollars, this part of the discussion is in terms ratios to premiums.

Appendix B

Discussion of Analogy to Stock Market Risk

In this appendix we compare S&P 500 equity risk[39] to the risk in a quota share reinsurance treaty. We begin by discussing the basis of the lognormal assumption. Then, in Example B.1, we show how the cash flows and economics of the quota share described in Example 3.3 can be replicated by an S&P 500 index transaction. That transaction takes the form of a short sale. In that scenario, the short seller loses money if the S&P 500 index closes higher than its level at the time of the short sale, just as the reinsurer loses money if the actual loss ratio exceeds the breakeven loss ratio. The appendix also includes Table B-1, which shows the data underlying Chart 4 and Table B-2, which shows the sensitivity of "10-10" test results for the quota share in Example 3.3 to the expected loss ratio.

Basis of Lognormal Assumption

It is possible, perhaps even likely, that stock prices are not lognormally distributed. However, stock price movements are commonly assumed by financial economists to follow Brownian motion through continuous time, which implies that stock returns over infinitesimal time intervals are normally distributed and stock prices are lognormally distributed after any finite time interval. For example, see Hull [10] Chapter 11 (p. 228) and Baxter-Rennie [11] Chapter 3 (p. 51). The latter says, "It is not the only model for stocks...but it is simple and not that bad." The Black-Scholes call option pricing formula was originally derived using a Brownian motion assumption. It has subsequently been shown that it can also be derived from the assumption that "asset prices are lognormally distributed under the martingale measure Q."[Ibid, p. 181].

At the same time there is some disagreement with the Brownian motion/lognormal assumption. See for example Peters [12], Chapter 3 (p. 27), who presented evidence that the distribution of actual stock market *returns* has a higher peak and fatter tails than predicted by a normal distribution and found, "The stock market's probability of a three-sigma event is roughly twice that of the Gaussian random numbers." [Ibid, p. 29]. He argues that because "capital market theory is, in general, dependent on normally distributed

[39] In order to simplify the discussion we ignore dividends, which could easily be incorporated in the example, but at the cost of complicating the comparison.

returns"[Ibid, p. 25], the Efficient Market Hypothesis, Capital Asset Pricing Model and Modern Portfolio Theory all rest on a shaky foundation. We don't take a position in that debate. However, we do wish to point out that our use of a lognormal distribution is consistent with the mainstream view.

The fact is that doubling the probability at the three-sigma level does not have a significant practical effect. We can adjust for Peter's finding of a fatter tail in the stock return distribution. A Student's *t* distribution with 30 degrees of freedom has twice the probability of a three-sigma event as the corresponding normal. It has a higher peak and fatter tails.

If we replace the lognormal stock price model with a "log t" model, "10-10" test values for the Example 3.3 quota share with $\sigma = 9\%$ and $\sigma = 13.85\%$ still fall far short of the significance threshold. For $\sigma = 9\%$, the 90th percentile result is still a small profit of 0.29% and the probability of a 10% loss rises to just 0.51%. For $\sigma = 13.85\%$, we find a 90th percentile loss of 4.17% and a probability of a 10% loss of 3.91%. These values are only slightly higher than those arising from the lognormal model. There is no practical effect of the non-normality observed by Peters.

Example B.1: Replicating a Quota Share with 25% Ceding Commission

Suppose the quota share in Example 3.3 involves ceded premiums of $10 million. Given a ceding commission of 25%, the net proceeds to the reinsurer total $7.5 million. Similarly, if S&P 500 "spiders" (symbol SPY) are trading at $117 a share (as they were in early May 2005), a short sale of 64,103 shares also yields net proceeds to the seller of $7.5 million. The expected loss ratio on the quota share is 70%, implying expected losses of $7 million. Claim payments are expected to lag premiums by one year. This is equivalent to the short seller estimating the expected value of SPY in one year's time as $109.20, or $7 million in total for the short position. (A short seller would generally not short the stock if he did not expect it to decline.) In order for the reinsurer to suffer a $1 million present value loss (10% of the ceded premiums), given a risk free interest rate of 5%, the loss ratio would need to reach 85% times 1.05, or 89.25%. In order for the short seller to incur a $1 million present value

loss, the stock price would have to reach $139.23[40]. These are the threshold levels for "passing" the "10-10" test.

As discussed in Example 3.3, in order for either the loss ratio to exceed 89.25% or the stock price to exceed $139.23 with a probability of 10% (these being fundamentally identical scenarios), the lognormal σ parameter must be at least 20.6%.

If we remove the 25% ceding commission from the quota share terms and instead provide for a premium cession net of a 25% expense allowance, then the "10-10" threshold for a 10% / $750,000 present value loss to the reinsurer is 82.5% times 1.05, or 86.63%. The comparable "10-10" threshold for the short seller is a stock price of $135.14. Exceeding these thresholds requires a σ value of at least 17.9%.

Data Underlying Chart 4

Table B-1 shows the data underlying Chart 4, which plots the probability of a 10% present value loss on the quota share defined in Example 3.2, given a 70% expected loss ratio, 25% ceding commission and σ values equal to VIX as of the last trading day of each year from 1990 through 2004 plus May 4, 2005.

[40] $1 million loss amounts to $15.60 per share, implying a present value share price of $132.60 and a future value share price of $139.23.

TABLE B-1
"10-10" Risk Transfer Analysis
for Quota Share in Example 2.3
Given Portfolio with Volatility of S&P 500 VIX
Data Underlying Chart 4

VIX Date	VIX	(a) 90th Percentile P.V. Underwriting Loss	(b) Probability of ≥10% P.V. Underwriting Loss
Dec 1990	26.4%	15.3%	14.6%
Dec 1991	19.3%	8.8%	8.8%
Dec 1992	12.6%	2.7%	2.3%
Dec 1993	11.7%	1.9%	1.6%
Dec 1994	13.2%	3.3%	2.8%
Dec 1995	12.5%	2.7%	2.3%
Dec 1996	20.9%	10.3%	10.3%
Dec 1997	24.0%	13.1%	12.9%
Dec 1998	24.4%	13.5%	13.2%
Dec 1999	23.4%	12.6%	12.4%
Dec 2000	26.9%	15.7%	14.9%
Dec 2001	23.8%	12.9%	12.7%
Dec 2002	28.6%	17.3%	16.1%
Dec 2003	18.3%	7.9%	7.8%
Dec 2004	13.3%	3.4%	2.9%
May 2005	13.9%	3.9%	3.4%

Sensitivity of "10-10" Test Values to Expected Loss Ratio Assumption

Table B-2 shows the sensitivity of the values shown in Table 5 to changes in the expected loss ratio. It shows that our conclusions with respect to the "10-10" test apply even with high assumed levels for the expected loss ratio. For example, even in the case of no expected profit and the higher May 2005 implied volatility levels, the "10-10" rule is not met.

TABLE 5
"10-10" Risk Transfer Analysis
for Quota Share in Example 2.3
Given Portfolio with Volatility of S&P 500
Sensitivity to Expected Loss Ratio

VIX	σ	Expected Loss Ratio	(a) 90th Percentile P.V. Underwriting Loss/(Profit)	(b) Prob of ≥ 10% P.V. Underwriting Loss/(Profit)
Low	9.00%	65%	(5.81%)	0.02%
Low	9.00%	67.5%	(3.15%)	0.08%
Low	9.00%	70%	(0.49%)	0.30%
Low	9.00%	62.5%	2.18%	0.93%
Low	9.00%	75%	4.84%	2.40%
May 2005	13.85%	65%	(1.78%)	0.92%
May 2005	13.85%	67.5%	1.04%	1.85%
May 2005	13.85%	70%	3.85%	3.41%
May 2005	13.85%	62.5%	6.67%	5.82%
May 2005	13.85%	75%	9.49%	9.25%

10. References

[1] CAS Valuation, Finance, and Investments Committee "Accounting Rule Guidance Statement of Financial Accounting Standards No. 113--Considerations in Risk Transfer Testing", *Casualty Actuarial Society Forum*, Volume: Fall, 2002, pp. 305-338.

[2] Kreps, Rodney E., "Parameter Uncertainty in (Log)Normal Distributions", PCAS LXXXIV, 1997, p. 553-580, http://www.casact.org/pubs/proceed/proceed97/97553.pdf

[3] Van Kampen, Charles E., "Estimating the Parameter Risk of a Loss Ratio Distribution", Casualty Actuarial Society Forum, Volume: Spring, 2003, p. 177-213, http://www.casact.org/pubs/forum/03spforum/03spf177.pdf

[4] Meyers, Glenn G., "The Cost of Financing Insurance", *Casualty Actuarial Society Forum*, Volume: Spring, 2001, pp. 221-264, http://www.casact.org/pubs/forum/01spforum/01spf221.pdf.
[5] Mango, Donald F., "Capital Consumption: An Alternative Methodology for Pricing Reinsurance", *Casualty Actuarial Society Forum*, Volume: Winter, 2003, pp. 351-379, http://www.casact.org/pubs/forum/03wforum/03wf351.pdf
[6] Ruhm, David L., "Risk Coverage Ratio: A Leverage-Independent Method of Pricing based on Distribution of Return," ASTIN Colloquium, July 2001, http://www.actuaries.org/ASTIN/Colloquia/Washington/Ruhm.pdf
[7] Wang, Shaun S., "Premium Calculation by Transforming the Layer Premium Density", *ASTIN Bulletin*, Volume: 26:1 (1996), pp. 71-92, http://www.casact.org/library/astin/vol26no1/71.pdf
[8] Wang, Shaun S., "A Universal Framework for Pricing Financial and Insurance Risks", *ASTIN Bulletin*, Volume: 32:2 (2002), pp. 213-234, http://www.casact.org/library/astin/vol32no2/213.pdf
[9] Panjer, Harry H., et al, Financial Economics, Schaumburg, IL, The Actuarial Foundation, 1998.
[10] Hull, John, *Options, Futures and Other Derivatives*, Third (International) Edition, London, Prentice Hall International, Inc., 1997.
[11] Baxter, Martin and Andrew Rennie, *Financial Calculus*,, London, Cambridge University Press, 1996.
[12] Peters, Edgar E., *Chaos and Order in the Capital Markets*, New York, John Wiley & Sons, Inc., 1991

Abbreviations and notations

10-10, 10% chance of 10% loss benchmark
CAS, Casualty Actuarial Society
COPLFR, Committee on Property and Liability Financial Reporting
E(x), expected value of x
E(x)*, expected value of transformed x
ERD, expected reinsurer deficit
EUD, expected underwriting deficit
F(x), aggregate distribution function
F(x)*, transformed aggregate distribution function
FAS 113, Financial Accounting Standard No. 113
Freq, probability of present value loss
Freq(UL), probability of underwriting loss
FV, future value operator
$N(z)$, standard normal distribution function
N^{-1}(prob), standard normal inverse dist function
PV, present value operator
RTD, right tail deviation
S&P 500, Standard & Poor's 500 stock index
Sev, severity of present value loss
Sev(UL), severity of underwriting loss
SSAP, Statement of Statutory Accounting Principles
TVaR, tail value at risk
$TVaR_\alpha$, tail value at risk at α probability level
VaR, value at risk
VaR_α,, value at risk at α probability level

Biographies of Working Party Members

Michael Wacek (Working Party Chair) is President of Odyssey America Reinsurance Corporation based in Stamford, CT. Over the course of 20 years in the industry, including nine years in the London Market, Mike has seen the business from the vantage point of a primary insurer, reinsurance broker and reinsurer. He has a BA from Macalester College (Math, Economics), is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries. He has authored several papers.

John Aquino is Executive Vice President of Benfield, Inc. in Chicago. He has over twenty years of experience in the reinsurance industry and actuarial consulting. John has also been a frequent speaker at industry meetings. He has served as President of the Midwest Actuarial Forum and also on various committees of the CAS. He holds a bachelor's degree in Math and a MBA in Finance from the University of Chicago and is a Fellow member of the CAS and a Member of the American Academy of Actuaries.

Todd Bault is Senior Research Analyst covering non-life insurance equities for Sanford C. Bernstein & Co., LLC, a New York-based research firm. In 2004, Todd was ranked the #1 non-life insurance analyst by Institutional Investor's annual poll. Risk measurement and quantification is a favorite research topic. Todd is a Fellow of the CAS and a Member of the American Academy of Actuaries.

Paul Brehm is Senior Vice President and Instrat Manager for Guy Carpenter. Paul spent 22 years at St Paul Travelers, most recently as the Chief Actuary. He holds a BS degree in Economics from the University of Minnesota. He is a Fellow of the CAS and a Member of the American Academy of Actuaries. Paul is a former chair of the CAS Valuation, Finance, and Investment Committee and has authored several papers.

Elizabeth Hansen is a Managing Director at Guy Carpenter, based in Minneapolis, MN. She is responsible for all quantitative resources in the Mid-America region as a regional manager of Instrat. Elizabeth holds a bachelor's degree in Mathematics from Luther College in Decorah, Iowa. She is a Fellow of the Casualty Actuarial Society, a member of the American Academy of Actuaries and a frequent presenter at industry conferences.

Pierre Laurin is Senior Vice President and Director of Reinsurance with Zurich in North America. He is responsible for the corporate reinsurance treaties and has underwriting oversight over all business unit related treaties. He has a degree in actuarial science from the University of Laval in addition to a Master of Science from Western University. He is a Fellow of the CAS, of the CIA and a Member of the American Academy of Actuaries. He has participated on the CAS examination committee, and is a frequent presenter at industry symposia on reinsurance topics.

Mark Littmann is a principal with PricewaterhouseCoopers LLP, based in the firm's Hartford, CT, office. His practice areas have included reserving, financial modeling, valuations, benchmarking claims and actuarial practices, financial reporting, and actuarial software systems. He has been the actuarial group's thought-leader for evaluating the internal control implications for actuaries arising from Section 404 of the Sarbanes-Oxley Act. He co-authored the firm's paper on the practical implications of implementing fair value accounting for property/casualty loss reserves. He earned a BA degree in Mathematics and Economics from Valparaiso University. He is a Fellow of the CAS and a Member of the American Academy of Actuaries.

Karen Pachyn is Senior Vice President and Chief Pricing Actuary for the North American Broker segment of GE Insurance Solutions. In this role, she leads a team of actuaries pricing treaty reinsurance in the U.S. and Canada. She is an FCAS, MAAA and CPCU, and is currently involved in the CAS Committee on Special Interest Seminars. She has previously participated on a number of other CAS Committees and is a past President of the Midwestern Actuarial Forum. She has a BA from Illinois Wesleyan University.

Deborah Rosenberg is the Deputy Chief Casualty Actuary for the New York State Insurance Department. She is a Fellow of the CAS and a Member of the American Academy of Actuaries. Deborah is the current Vice President of Administration for the CAS and also a member of the Task Force on Publications and the Task Force on Reserving Principles.

David Ruhm is Assistant Vice President at Hartford Investment Management Company in Hartford, CT. His areas of responsibility include portfolio risk management, financial modeling and enterprise risk management. He has a bachelor's degree in Mathematics from the University of California, San Diego and is a Fellow of the CAS. David has published several papers on risk theory and capital management. He participates on the CAS Theory of Risk Committee, and is a frequent presenter at industry conferences.

Mark van Zanden is a structured risk underwriter with Catlin Insurance Company Ltd. in Bermuda. He is responsible for originating, structuring, analyzing and underwriting highly tailored (re)insurance transactions. He has a degree in Mathematics and Statistics from the University of Western Ontario in London, Canada. He is a Fellow of the CAS and of the Canadian Institute of Actuaries. He is a CFA charterholder. He has over ten years of experience designing and analyzing alternative risk transfer (re)insurance transactions.

Estimating Tail Development Factors: What to do When the Triangle Runs Out

Joseph Boor, FCAS

Abstract:: There are several methods in use today for estimating tail factors. However, most of them are discussed as adjuncts to papers that primarily deal with other subjects. This paper will present a wide variety of method in an understandable format, and includes copious examples

Keywords: Loss reserving, tail factors

1. INTRODUCTION

In many loss reserve analyses, especially those involving long-tail casualty lines, the loss development triangle may end before all the claims are settled and before the final costs of any year are known. For example, it is quite common to analyze U. S. workers compensation loss reserve needs using the ten years of data available in Schedule P of the US NAIC-mandated Annual Statement, while knowing that some of the underlying claims may take as long as fifty years to close. In response to this, actuaries supplement the 'link ratios' they obtain from the available triangle data with a 'tail factor' that estimates the development beyond the last stage of development (last number of months of maturity, usually) for which a link ratio could be calculated.

The tail factor is used just like a link ratio in that it estimates (1.0 + ratio of (final costs after all claims are closed) to (the costs as of the last development stage used)). It is of course included in the product of all the remaining link ratios beyond any given stage of development in calculating a loss development factor to ultimate for that stage of development.

This paper will discuss the methods of computing (really estimating to be precise) tail factors in common usage today. It will also suggest both improvements in existing methods and a new method. It will begin with the simplest class of methods and move forward in increasing complexity.

There are four groups of methods that will be presented:

1. The Bondy (repeat-the-last link-type) methods
2. The Algebraic methods (methods based on algebraic relationships between the paid and incurred triangles)
3. Use of Benchmark Data
4. Curve Fitting Methods

As part of the discussion, commentary on the advantages and disadvantages of each individual method, as well as each class of methods will be included. When the opportunity to discuss an improvement or enhancement that applies to multiple methods presents itself, a brief digression on the enhancement will be included.

2. GROUP 1 - THE BONDY-TYPE METHODS

The Bondy methods all arose from an approach published by Martin Bondy prior to the 1980s. In what was thought to be a period where development decayed rapidly from link ratio to link ratio, he promulgated a practice of simply repeating the last link ratio for use as the tail factor. Since then, several variations of his method that all base the tail factor on the last available link ratio have arisen.

2.1 The Bondy Method

As explained above, the original Bondy method involves simply using the last link ratio that could be estimated from the triangle (the link ratio of the last development stage present in the triangle, or the last stage where the triangle data could be deemed reliable for estimation) as the tail factor. This 'repeat the last link ratio' approach probably seems crude and unreasonable for long-tailed lines, where link ratios decay slowly. However, for fast decaying lines (such as an accident year[1] analysis of automobile extended warranty) this method may work when used as early as thirty-six or forty-eight months of maturity. It must be recognized, though, that in long-tailed lines the criticism is usually justified.

To truly understand this method it also may be best viewed in historical context. The author of the method, Martin Bondy, developed this method well prior to the 1980's. It is commonly believed that during the 1960s and certainly part of the 1970s the courts proceeded at a faster pace and, ignoring the long-tail asbestos, environmental, and mass tort issues that would eventually emerge, general liability was believed to have a much shorter tail than we see today.

It is also of interest to note that there is a theoretical foundation that supports this in certain circumstances. If one assumes that the 'development portion' of the link ratios (the link ratios minus one) are decreasing by one-half at each stage of development, and the last link ratio is fairly low, then the theoretically correct tail factor to follow a link ratio of 1+d is:

$(1+.5d)\times(1+.25d)\times(1+.125d)\times(1+.0625d)\times\ldots\ldots$

Or

$1+(.5+.25+.125+.0625+\ldots)\times d$ + terms involving d^2, d^3, etc.

[1] It should be noted that policy year automobile extended warranty represents an entirely different situation.

Which, per the interest theorem $v+v^2+v^3+....=v/(1-v)$ is equivalent to:

$1+1\times d$ + terms involving d^2, d^3, etc.

Since d is 'small', the other terms will be smaller by an order of magnitude, making the implied tail factor under these assumptions very close to the Bondy tail factor, a repetition of the last link ratio, 1+d. So the Bondy tail factor is 'nearly' equivalent to the tail implied by what will later be called the 'exponential decay' method, with a 50% decay constant.

Of note, this involves two basic assumptions. First that the link ratios decay in proportion to the remaining 'development portion' of the link. Of note, in the absence of any information whatsoever about the decay, that would be as reasonable an assumption as one could reasonably make. Second, that the decay constant is 50%. Again, in the absence of any data whatsoever, one-half would be as reasonable an assumption as one could possibly make. Of course, we do have data in the link ratios before the tail, but it is important to understand this theoretical basis for the Bondy tail factor.

2.2 The Modified Bondy Method

In this method, the last link ratio available from the triangle, call it 1+d, is modified by multiplying the development portion by 2. The result is a development factor like 1+2d. Alternately, the last entire link ratio may be squared, which yields nearly the same value. This has many of the same issues and applications as the basic Bondy method, but it does yield a larger tail than the Bondy method itself. However, for long-tail lines it is still not what would be considered a truly conservative approach, as we will see later. The assumption here is ' The Bondy method seems to underestimate, it should be increased, the easiest thing to do is to multiply the development portion by two.'.

A little algebra and the $v+v^2+v^3+....=v/(1-v)$ theorem show that this is functionally equivalent to 'exponential decay' with a decay coefficient of $^2/_3$.

2.3 Advantages and Disadvantages of the Bondy Methods

The primary advantages of the Bondy methods are that they are extremely simple to execute and easy to understand. Further, they involve relatively straightforward assumptions. However, a major disadvantage is that they tend to greatly underestimate tails of long-tailed, slow-decaying lines.

3. GROUP 2 - THE ALGEBRAIC METHODS

These methods involve initially computing some algebraic quantity that in turn describes a relationship between some aspect of the paid and incurred loss triangles. Then that quantity can be used to generate a tail factor estimate. As with the Bondy method, and almost all tail factor estimation methods, they are based on assumptions. However, in this case each is based on some relatively simple and fairly logical assumption that some numerical relationship known to be true in one circumstance will be true in another.

3.1 Equalizing Paid and Incurred Development Ultimate Losses

This method is the first method discussed with a full theoretical background. It is most useful when incurred loss development essentially stops after a certain stage (i.e., the link ratios are near to unity or unity). Then, due to the absence of continuing development, the current case incurred (sometimes called reported) losses are a good predictor of the ultimate losses for the older or oldest years without a need for additional tail factor development. A tail factor suitable for paid loss development can then be computed as the ratio of the case incurred losses to-date for the oldest (accident[2]) year in the triangle divided by the paid losses to-date for the same (accident) year. That way, the paid and incurred development tests will produce exactly the same ultimate losses for that oldest year.

This method relies on one axiomatic (meaning plainly true rather than an assumption as such) assumption and two true assumptions. The axiomatic assumption is that the paid loss and incurred loss development estimates of incurred loss are estimating the same quantity, therefore the ultimate loss estimates they produce should be equal. The second assumption (the first true assumption) is that the incurred loss estimate of the ultimate losses for the oldest year is accurate. The last assumption is that the other years will show the same development in the tail as the oldest year.

This method may also be generalized to the case where case incurred losses are still showing development near the tail. In that case, the implied paid loss tail factor is

(incurred loss development ultimate loss estimate for the oldest year) / (paid losses to-date for the oldest year).

Of course, in that instance the incurred loss development estimate for the oldest (accident) year is usually the case incurred losses for the oldest year multiplied by an incurred loss tail factor developed using other methods.

This method has a substantial advantage in that it is based solely on the information in the triangle itself and needs no special assumptions. Its weakness is that you must already have a reliable estimate of the ultimate loss for the oldest year before it can be used. An ancillary weakness flows from the assumptions underlying this method. Specifically, if the initial incurred loss development test is driven by a tail factor assumption, this becomes a test that is also based on not only that assumption, but also the assumption that the ratio of the case incurred loss to the paid loss will be the same for the less mature years once they reach the older level of maturity where you are equalizing the paid and incurred loss estimates.

[2] Accident year is used here for illustration. Under similar circumstances, this method would also work in policy year, reinsurance contract year, etc. development.

3.1.1 An example:

Assume that it is just after year-end of 2000. You have pulled the incurred loss triangle from a carrier by subtracting part 4 of Schedule P from part 2 of Schedule P. You have also pulled a paid loss triangle from part 3 of Schedule P. The triangles cover 1991-2000, so 1991 is the oldest year. Say for the sake of argument that the incurred loss link ratios you develop are 2.0 for 12-24 months, 1.5 for 24-36, 1.25 for 36-48, 1.125 for 48-60, 1.063 for 60-72, 1.031 for 72-84, 1.016 for 84-96, 1.008 for 96-108, and 1.004 for 108-120. This conveniently happens to match the exponential decay discussed for the Bondy method, so it makes sense to use 1.004 for the tail factor for development beyond 120 months. Now assume that the latest available (i.e., at 12/31/2000, or 120 months maturity) the case incurred loss[3] for 1991 is $50,000,000 and the corresponding paid loss is $40,000,000. The incurred test ultimate using the 1.004 tail factor is $50,200,000. The paid loss tail factor to equalize the ultimate would be $50,200,000 divided by $40,000,000 or 1.255.

3.1.2 Improvement 1 - using multiple years to develop the tail factor

As stated earlier, the previous method assumes that the current ratio of case incurred loss to paid loss that exists in the oldest year (1991 evaluated at 12/31/2000 in the example above) will apply to the other years when they reach that same level of maturity. For a large high dollar volume triangle with relatively low underlying policy limits that may be a reasonable assumption, but for many reserving applications the 120 month ratio of case incurred to paid loss may depend on whether a few large, complex claims remain open or not. Therefore, it may be wise to supplement the tail factor derived from the oldest available year with that implied by the following year or even the second following year. This method is particularly useful when the later development portion of the triangle has some credibility, but the individual link ratio estimates from the development triangle are not fully credible.

The process of doing so is fairly straightforward. You merely compute the tail factor for each succeeding year by the method above, and divide each by the remaining link ratios in the triangle.

An example using the data above may help clarify matters. Given the data above, assume that 1992 has $50,000,000 of paid loss and $60,000,000 of case incurred loss. Also, assume that your best estimate of the 108-120 paid loss link ratio is 1.01. The incurred loss estimate of the ultimate loss, using the 108-120 link ratio (1.004) and the incurred loss tail factor (also 1.004) is $60,000,000×1.004×1.004, or $60,480,960. The estimated (per incurred loss development) ultimate loss to paid loss ratio at 108 months would then be $60,480,960/$50,000,000, or approximately 1.210. So, 1.210 would then be the tail factor estimate for 108 months. Dividing out the 108-120 paid link ratio (assumed above to be 1.01) gives a tail factor for 120 months of 1.21/1.01 = 1.198. By comparison, the previous analysis using 1991 instead of 1992 gave a 120-month tail factor estimate of 1.255. So it is possible that either 1991 has a high number of claims remaining open, or that 1992 has a low number. Both indicate tail factors in the 120-125 approximate range, though. So averaging

[3] To be technically correct, this would be loss and defense and cost containment under 2003 accounting rules.

the estimates might be prudent. Further, the use of averaging greatly limits the impact of any unusually low or high case reserves that may be present in the oldest year in the triangle.

Note also, that the improvement above involved computing an alternate tail factor using the year with one year less maturity. A similar analysis could also be performed on the next oldest year, 1993, except that two incurred development link ratios plus the tail factor are needed to compute the incurred loss estimate of ultimate. Correspondingly, two paid loss link ratios need to be divided out of the (incurred loss ultimate estimate)/(paid loss to-date ratio for 1993) to estimate the 120 month paid loss tail factor

3.1.2.1 An important note

Further, in this case the improvement involved reviewing the tail factors at various ages from the equalization of paid and incurred loss estimates of the ultimate loss. The core process involves computing tail factors at different maturies, then dividing by the remaining link ratios to place them all at the same maturity. As such, it can also be used in the context of other methods for computing tail factors that will be discussed later in this paper.

3.1.3 A brief digression – the primary activity within each development stage

When using multiple years to estimate a tail factor, it is relatively important that the years reflect the same general type of claims department activity as that which takes place in the tail. For example, in the early 12 to 24 month stage of workers compensation, the primary development activity is the initial reporting of claims and the settlement and closure of small claims. The primary factors influencing development are how quickly the claims are reported and entered into the system, and the average reserves (assuming the claims department initially just sets a 'formula reserve', or a fixed reserve amount for each claim of a given type such as medical or lost time) used when claims are first reported. In the 24 to 36-48 month period, claims department activity is focused on ascertaining the true value of long-term claims and settling medium-sized claims. After 48-60 months most of the activity centers on long-term claims. So, the 12-24 link ratio has relatively little relevance for the tail, as the driver behind the link ratio is reporting and the size of initial formula reserves rather than the handling of long-term cases. Similarly, if the last credible link ratio in the triangle is the 24 to 36 or 36 to 48 link ratio, that triangle may be a poor predictor of the required tail factor.

3.2 The Sherman-Boor[4] Method - Adjusting the Ending Case Using Ratios of Paid Loss to Case Reserve Disposed of

This method, developed by Sherman in Section X of [3] and independently by the author, is the one method that relies solely on the triangle itself and does not require a pre-existing ultimate loss estimate, involve curve-fitting assumptions, or require external data. For data triangles with high statistical reliability as predictors, this can represent the optimum estimation process.

This method involves simply determining the ratio of case reserves to paid loss for the oldest year in the triangle, then adjusting the case reserves by an estimate of the ratio of the unpaid loss to carried case reserves. In essence, the case reserves of the oldest accident year are 'grossed up' to estimate the true unpaid loss using a factor. The estimate of the (true unpaid loss)/(case reserves) factor is based on how many dollars of payments are required to 'eliminate' one dollar of case reserves.

The mathematical formula requires computing a triangle containing incremental rather than cumulative paid losses. In effect, for each point in the paid loss triangle, one need only subtract the previous value in the same row (the first column is of course unchanged). The next step begins with a triangle of case reserves. The incremental case reserve disposed of is calculated as the case reserve in the same row before the data point, less the current case reserve. That represents (as the beginning case reserve – the ending case reserve) the case reserve disposed of. Then the ratios of incremental paid to reserve disposed of at the same points in the triangles are computed. Reviewing these, the adjustment ratio for the ending case reserves is estimated.

3.2.1 An example

Reviewing an example may help the reader follow the calculations discussed earlier. This method requires two triangles, one of paid loss and one of case reserves. Consider the following set of triangles:

Cumulative Paid Loss Triangle						
	12	24	36	48	60	72
1991	1,000	2,000	2,500	2,800	2,950	3,100
1992	1,100	2,400	3,000	3,500	3,900	
1993	1,300	2,500	3,000	3,400		
1994	1,200	2,300	3,100			
1995	1,400	2,800				
1996	1,490					

[4] Of note, this method was first published by Richard Sherman, FCAS in 1984 and developed independently by the author in 1987. Of note, the author used some business materials that contained precursors to this method in 1984-1986 that were developed by a firm of which Mr. Sherman was a principal.

Triangle of Case Reserves Outstanding (Cumulative Case Incurred-Cumulative Paid)						
	12	24	36	48	60	72
1991	1,500	1,300	900	750	600	500
1992	2,000	1,700	1,300	900	600	
1993	1,900	1,700	1,300	1,000		
1994	2,100	2,100	1,500			
1995	2,300	2,000				
1996	2,500					

First, we compute the incremental paid loss triangle. We begin with a given cell in the cumulative paid loss triangle, and then we subtract the previous cell in the same row of the cumulative paid loss triangle. That produces the following triangle.

Incremental Paid Loss Triangle						
	12	24	36	48	60	72
1991	1,000	1,000	500	300	150	150
1992	1,100	1,300	600	500	400	
1993	1,300	1,200	500	400		
1994	1,200	1,100	800			
1995	1,400	1,400				
1996	1,490					

Then we subtract the current cell from the previous cell in the case reserve triangle to obtain the triangle of case reserves disposed of.

Triangle of Incremental Case Reserves Disposed Of						
	12	24	36	48	60	72
1991		200	400	150	150	100
1992		300	400	400	300	
1993		200	400	300		
1994		100	600			
1995		300				
1996						

Then we divide the actual final costs paid (the incremental paid loss), by the assumption-based case reserves eliminated.

Ratio of Paid Loss to Reserves Eliminated						
	12	24	36	48	60	72
1991		500%	125%	200%	100%	150%
1992		433%	150%	125%	133%	
1993		600%	125%	133%		
1994		1100%	133%			
1995		467%				
1996						

Because the early development involves not just elimination of case reserves through payments, but also substantial emergence of IBNR claims, the 12 and 36 columns are presumably distorted. In many lines the 48 month column would still be heavily affected by newly reported large claims, but presumably this is medium-tail business. Looking at the various ratios it would appear that they average around 140%, so we will use that as our adjustment factor for the case reserves.

Pulling the $500 of case left on the 1991 year at 72 months, and the cumulative paid on the 1991 year of $3,100, the development portion of the paid loss tail factor would be ($500/$3,100)×140% = .161×140% = .226. So, the paid loss tail factor would be 1.226.

For the incurred loss tail factor, first note that only the 'development portion' of the 140%, or 40%, need be applied (the remaining case is already contained in the incurred). Second, a ratio of the case reserves to incurred loss is technically needed (replacing 1.61 with $500/($500+$3,100) = .139). Multiplying the two numbers creates an estimate of the development portion of the tail at .4×.139=.056. So, the incurred loss tail factor estimate would be 1.056.

3.2.2 An Important Note

As is the case with most of the other methods, this method has strengths and weaknesses. Significant strengths of this method are that it requires only the data already in the triangle and that it does not require additional assumptions. The weakness is that it can be distorted if the adequacy of the ending case has changed significantly from the previous year. The reader is advised to also follow Improvement 1 and also evaluate the tail at the next-to-oldest year.

4. GROUP THREE - METHODS THAT USE BENCHMARK DATA

A common solution to the ratemaking problems generated by data with partial statistical reliability (credibility) is to supplement the claims data with a 'complement of credibility'. Of course, tail factor estimation problems stem more from a lack of any data at all after the oldest development stage in the triangle rather than from partially reliable data. But, we can adopt a similar strategy and add outside data in the form of benchmark development factors.

4.1 Directly Using Tail Factors From Benchmark Data

As noted above, many actuaries review benchmark data in selecting tail factors[5]. Benchmark data may come from one of several sources. Perhaps the most common is the use of the data triangles that can be developed from Best's Aggregates and Averages for each of the Schedule P lines. The two larger rating bureaus, the National Council on Compensation Insurance and Insurance Services Office; as well as the Reinsurance Association of America, all publish benchmark loss development data. At its simplest, this method involves copying the derived remaining development factor at the maturity desired for the tail factor.

It is important to note, though, that the quality of the benchmark tail factor as an estimate of the tail depends on how closely the tail development of the benchmark mirrors the tail development of the book of business being analyzed. Considerations such as differences in the way claims are adjusted or reserved, differences in the potential for long-developing high value claims, differences in the initial reporting pattern of claims (claims-made vs. occurrence, whether or there is an innately long discovery period or not, etc.), and differences in the adjudication process of litigated claims can all cause differences in development patterns. It is important to consider those factors along with the statistical reliability of the benchmark triangle when selecting the most appropriate benchmark tail factor.

4.2 Using Benchmark Tail Factors Adjusted to Company Development Levels

One way to address differences between the benchmark development pattern and the development pattern of a given book of business is to try to adjust the benchmark data to more closely mirror the subject book of business. A common practice is to review the relativities of link ratios from the triangle being analyzed to benchmark link ratios. Of course, there is not a tail factor for the triangle being analyzed (we are trying to estimate one). So, instead we can review the quotients (relativities) of subject triangle link ratios to those of the benchmark data at the development stages prior to the tail development stage. The relativities from those stages are used to estimate a adjustment multiplier for the benchmark tail factor. Of note, generally just the development portions ('d' of 1+d) are compared in all the relativities we compute.

[5] It is also common for actuaries to review benchmark data to supplement the portion of the reserve triangle following 72, 60, 48, or even 36 months when the overall triangle has medium credibility and hence has less than medium credibility in the portion that is dominated by activity on a smallish number of claims.

4.2.1 An example

An example will help to illustrate how the process works. Consider the following two patterns:

	Link Ratio Estimated By Triangle	Benchmark Link Ratio
Months of Maturity		
12	2.000	2.000
24	1.450	1.350
36	1.200	1.150
48	1.150	1.100
60	1.100	1.050
72	1.080	1.030
84	1.050	1.025
96	1.035	1.020
108	1.010	1.010
Tail		1.050

We then simply compute the relativity quotient of the 'development portion' of our triangle-based link ratios to the development portion of the matching benchmark link ratios. Noting that 1+1 = 100%, .45+.35 = 129%, .2+.15 = 133%, etc.

Months of Maturity	Link Ratio Estimated by Triangle	Benchmark Link Ratio	Relativity of Triangle Development to Benchmark
12	2.000	2.000	100%
24	1.450	1.350	129%
36	1.200	1.150	133%
48	1.150	1.100	150%
60	1.100	1.050	200%
72	1.080	1.030	267%
84	1.050	1.025	200%
96	1.035	1.020	175%
108	1.010	1.010	100%
Tail		1.050	
Chosen Ratio			175%
Implied Tail	1.088		

In the case above, we judgmentally select that the triangle development is roughly 175% of benchmark based on the 60 through 108 month relativities. So the .05 development portion of the benchmark tail becomes $.05 \times 1.75 = .0875 \cong .088$. Consequently the entire tail factor, including unity, is 1.088.

4.2.2 Another important note

It is important to consider that adjusting the benchmark tail for actual triangle link ratios is only helpful as long as the link ratios, or at least the broad pattern of link ratios has statistical reliability (predictive accuracy). If not, the uncertainty surrounding the true long-term link ratios of the block of business will cause the adjusted tail factor to lack predictive accuracy.

4.3 *Advantages and Disadvantages of Using Benchmark Data*

When a good benchmark tail factor is available, this is both one of the easiest and also among the most useful methods. However, it is often difficult to find a perfect match in terms of all the factors (claims handling, case reserving, potential for large claims, etc.) that affect loss development. Adjusting the benchmark improves the fit markedly. One could even think of the process of adjusting the benchmark as that of fitting a curve to the link ratios, where the family of curves you are fitting from consists of various relativity-adjusted versions of the benchmark. If the benchmark is remotely related to the book of business being analyzed, that family of curves should be a superior choice to the highly assumption-driven curve families discussed later under curve fitting.

On the other hand, it is often very difficult to obtain the more-mature data needed to create a reliable benchmark tail factor. So, for tail factors beginning at 108 or 120 months, it may be very difficult to find a suitable benchmark.

5. GROUP 4 -THE CURVE FITTING METHODS

As good students of numerical analysis, actuaries long ago realized that they could attempt to extrapolate the tail development by fitting curves to the development before the tail, then using the fitted curve to extrapolate the additional tail development. Some methods have been developed that fit a curve to the paid or incurred loss. Other methods fit to the link ratios. What they all have in common is that they begin with some assumption about the development decay that gives rise to a family of curves, and then select the coefficient(s) that specify the particular member of the family of curves that best fits the data. As with most extrapolations, they are as good as the assumptions that underlie them.

5.1 McClenahan's Method-Exponential Decay of Paid Loss Itself

McClenahan's method (as discussed in [1]) fits a curve to a set of data per an assumption that the incremental paid loss of a single accident year will decay exponentially over increasing maturities of the accident year. In effect, that there was some decay rate 'p' and that the next month's payout on the accidents in a given month would always be 'p' times the current month's payments on that given accident month. He combined that with an assumption that no payments occurr in the first few months of a claim. Putting those pieces together mathematically, he inferred that the payments in a given incremental month of maturity (call it 'm') were

$Ap^{(m-a)}q$.

In this case A is a constant of proportionality and 'p', ($0<p<1$, $q=(1-p)$) represents the decay rate[6] and 'a'[7] represents the average lag time until claims begin to be paid. A theorem from the study of compound interest states that

$$\sum_{m=a}^{\infty} Ap^{(m-a)}q = A\sum_{i=0}^{\infty} p^i q = Aq/(1-p) = Aq/q = A.$$

So A is actually the ultimate loss for the entire year.

Then, under this assumption, the additional payments or incurrals beyond x months are theoretically determined by the basic formula, at least once p and a are estimated. And there are several ways to estimate p and a. For convenience, p is monthly, but p^{12}, the annual decay rate, may be defined as 'r'[8]. Then r may be estimated by reviewing the ratios of incremental paid between m+12 and m+24 months to the incremental paid between m and m+12 months. McClenahan advised that 'a' could be estimated by simply reviewing the average report lag[9] (average date of report-average date of occurrence) for the line of business.. Then, a curve of the form

Ar^y,

where y is the maturity of the accident year in years before each amount of incremental paid can be fit to the incremental dollar amounts of paid loss (or incurred loss, as long as no downward development in incurred loss is present in the development pattern).

Then, McClenahan shows that the percentage remaining unpaid for an entire twelve month accident year at m months of (returning to $p = r^{1/12}$) is

$$(1-p)\times(p^{m+1-a} + p^{m+1-a-1} + p^{m+1-a-2} + \ldots + p^{m+1-a-11})/(12\times(1-p)) = p^{m-a-10}(1-p^{12})/12q$$

The tail factor at m months is of course unity divided by the percentage paid at m months, or

1/(100% - percentage unpaid at m months).

---

[6] McClenahan's model actually incorporates additional variables for trend, etc that may be collapsed into 'p' for purposes of this analysis.

[7] In Mclenahan's original paper, 'd' is used instead of 'a'. But, since I have used 'd' to denote the development portion of the link ratio or development factor, I am using 'a' to denote the average payment lag.

[8] Please note that the usage of 'r' in this context is different than the usage in McClenahan's original paper. It is used merely because it represents an annual rate.

[9] Note that 'a' applies on a month-by-month basis. So it is technically incorrect to say that the average lag between the beginning of all loss reporting for an accident year is six months (the average lag between inception of the accident year and loss occurrence, at least for a full twelve month accident year) plus 'a' months. To simplify the calculations, the first twelve months can be excluded from the fit

Substituting our formula for the unpaid at 12 months, McClenahan's method produces a tail factor of

$1/\{1 - [p^{m-a-10}(1-p^{12})/12q]\}$

Some algebra reduces that to

$12q/\{12q - p^{m-a-10}(1-p^{12})\}$,

which provides a nice closed form[10] expression for the tail.

An Example:

Assume that you begin with an 8-year triangle, and generate the following link ratios:

12-24	5.772
24-36	1.529
36-48	1.187
48-60	1.085
60-72	1.042
72-84	1.022
84-96	1.012

The first step is to covert them to a form of dollars paid (remember that there are different paid amounts for different accident years, so we just begin with one hundred dollars for the curve fitting and multiply by the successive link ratios.

			Equivalent
Development	Link	Beginning	Cumulative
Stage	Ratio	Maturity	Paid
12-24	5.772	12	$100.00
24-36	1.529	24	$577.23
36-48	1.187	36	$882.45
48-60	1.085	48	$1,047.38
60-72	1.042	60	$1,136.50
72-84	1.022	72	$1,184.66
84-96	1.012	84	$1,210.68
		96	$1,224.75

[10] It should be noted that while a closed form expression makes the calculations easy, for some audiences, it may be preferable to show the projected link ratios, at least until they are overwhelmingly close to unity.

Then subtract successive cumulative paid amounts to obtain 'normalized to $100 of first year paid' incremental dollars at each stage of development that mirror the actual link ratios.

Development Stage	Link Ratio	Beginning Maturity	Equivalent Cumulative Paid	Incremental Paid (Difference)
12-24	5.772	12	$100.00	$100.00
24-36	1.529	24	$577.23	$477.23
36-48	1.187	36	$882.45	$305.22
48-60	1.085	48	$1,047.38	$164.93
60-72	1.042	60	$1,136.50	$89.12
72-84	1.022	72	$1,184.66	$48.16
84-96	1.012	84	$1,210.68	$26.02
		96	$1,224.75	$14.06

Then ratios of the successive 'normalized' incremental paid amounts can be taken.

Development Stage	Link Ratio	Beginning Maturity	Equivalent Cumulative Paid	Incremental Paid (Difference)	Year to Year Ratio
12-24	5.772	12	$100.00	$100.00	
24-36	1.529	24	$577.23	$477.23	4.7723
36-48	1.187	36	$882.45	$305.22	0.6396
48-60	1.085	48	$1,047.38	$164.93	0.5404
60-72	1.042	60	$1,136.50	$89.12	0.5404
72-84	1.022	72	$1,184.66	$48.16	0.5404
84-96	1.012	84	$1,210.68	$26.02	0.5404
		96	$1,224.75	$14.06	0.5404

As one can see, in this contrived example, the development stage-to-stage ratio is a constant $r = .5404$. It's twelve root p is $p = r^{1/12} = .95$.

That of course only provides p, the average delay must be found as well. Because the answer is contrived to have a=7 months, a= 7 months will work perfectly[11] for this example, but note that McClenahan suggests merely using the report delay for the book of business to determine 'a'.

Using a= 7 months and p = .95, the computed tail factor is

$$12q/\{12q - .95^{m-a-1(1}(1-.95^{12})\},= .6/\{.6 - .017385(1-.5404)\}= 1.0135.$$

If one reviews the link ratios prior to this, it certainly appears to be reasonable. In fact, extending the payout to additional stages of development will confirm its accuracy.

[11] An interested reader can confirm that a=7 months and p=.95 yields the exact link ratios above.

5.1.1 Advantages and disadvantages of McClenahan's method

At its core, McClenahan's method involves three basic assumptions: First, it assumes that the pattern of paid loss will be a constantly decreasing pattern, at least after all the initial report lags are finished. Second, he assumes that the reduction will always occur in proportion to the size of the most current payout (exponential decay). Third, he assumes that the exponent of decay is constant throughout the entire payout pattern. Logically speaking, if one knew nothing about the individual pattern of the data, but was forced to make some assumptions, those assumptions would seem to be about as minimal and reasonable as possible (excepting perhaps the third). But it is important to remember that they are assumptions and as such will color the predictions the method generates. They do suggest exponential decay of the paid amounts, and exponential decay is a relatively fast decay relative to other forms of asymptotic (far out in the tail) decay. Moreover, it does seem that in practice the decay in paid loss often seems to 'stall out' and show less decay near the tail.

5.1.2 Improvement 2 - exact fitting to the oldest year

A common problem with fitted curves is that the combination of the curve assumptions and the data in the middle of the triangle may create a curve that varies significantly from the development factors at the older stages. McClenahan's method is relatively unique in that the curve is fit to the incremental paid, rather than the link ratios (as will be done in most of the later methods). Nevertheless, we can often improve the quality of the tail prediction by comparing the fitted value to the actual incremental paid loss at the latest stage.

This approach is especially helpful when the curve does not match the shape of the data itself. For example, assume that the assumption of a constant decay rate does not hold. Say the initial year-to-year decay was high at between 36[12] and 48 months, 48 and 60 months, etc., but the decay rate at 84 to 96 months and 96 to 108 months, etc. is much less (i.e., a higher decay factor). Then, the last incremental payments (say between 108 and 120) may be much higher percentagewise than what is implied by the fitted curve.

In that case[13], one need merely multiply the 'development portion' of the tail factor (the tail factor minus one) times the ratio of the actual 108 to 120 increment to the fitted increment. Of course, unity (one) must be added to the final result to produce a proper tail factor.

[12] Note that because of the delay a before payments, etc. begin, the apparent decay between 12 and 24 months and 24 to 36 months is a distortion of the true annual decay.

[13] Assuming that the data has enough volume for the 108 to 120 link ratio to have full credibility.

For example, in the above data, the last incremental data shown is from 96 to 108 months. In that case the fitted value equals the actual 'normalized' value equals $14.06 per a 96 to 108 link ratio of 1.012 and decay rate of .5404. But what if we had the same decay rate overall, but the link ratio from 96 to 108 was 1.018. In that case, the incremental paid would be $21.09, or 150% of the fitted value of $14.06. Then the adjusted tail factor would be:

1+150%(fitted tail factor-1) = 1+150%(1.0135-1)=1+150%×.0135=1.0203.

Note that in the case of McClenahan's method, the ratio used for 'exact fitting' is the ratio of actual to fitted paid loss. In the later methods, where a curve is fit to the 'development portions', a ratio of development portions should be used to produce the exact fit to the last link ratio.

5.1.3 Improvement 1 (using multiple years to estimate the tail) can enhance improvement 2

For McClenahan's method, and all the curve-fitting methods, improvement 1 can only be done in connection with improvement 2. In essence, the concept is to create an exact fit to the next-to-oldest link ratio or 'normalized' paid loss, and perhaps the third-to-last link ratio as well. Then, the implied tail factors can be averaged or otherwise combined into a single tail factor indication. This method is particularly useful when the 'tail' of the triangle has some credibility, but the individual link ratio estimates from the development triangle are not fully credible.

			Equivalent	Incremental	Year	Revised	Equivalent	Incremental
Dev	Link	Ending	Cumulative	Paid	to Year	Link	Cumulative	Paid
Stage	Ratio	Maturity	Paid	(Difference)	Ratio	Ratio	Paid	(Difference)
12-24	5.772	12	$100.00	$100.00		5.772	$100.00	$100.00
24-36	1.529	24	$577.23	$477.23	4.7723	1.529	$577.23	$477.23
36-48	1.187	36	$882.45	$305.22	0.6396	1.187	$882.45	$305.22
48-60	1.085	48	$1,047.38	$164.93	0.5404	1.085	$1,047.38	$164.93
60-72	1.042	60	$1,136.50	$89.12	0.5404	1.042	$1,136.50	$89.12
72-84	1.022	72	$1,184.66	$48.16	0.5404	1.044	$1,184.66	$48.16
84-96	1.012	84	$1,210.68	$26.02	0.5404	1.018	$1,236.79	$52.13
		96	$1,224.75	$14.06	0.5404		$1,259.05	$22.26

For example, the table above contains the data cited in the original example of McClenahan's Method (5.1) as the first set of link ratios, equivalent cumulative paid, etc. But, beginning with the 'Revised Link Ratio' column it contains alternate link ratios, etc. for 72 months and later. Using that data, one would still conclude that the fitted annual decline is.5404. But, now the last link is 1.018 (as in 5.1.2 – Improvement 2) instead of 1.012, and that the next-to-last (penultimate) 72-84 link is 1.044 instead of 1.022. In this case, the implied normalized incremental paid between 72 and 84 monthsis now $52.13 instead of the original $26.02. $52.13 is approximately twice $26.02, so the 72-84 activity would imply a tail factor of

1+200%(fitted tail factor –1) = 1+200%(1.0135-1) = 1+200%×.0135 = 1.0270.

The implied tail factor per the 84-96 link ratio is very close to the 1.0203 of the previous example. Note that the normalized paid loss in the 84-96 stage is $22.26 now or roughly 158% of paid loss. That implies a tail factor of

1+158%(1.0135-1) =1+158%×.0135 = 1.0213.

So, averaging the two, a tail factor in the range of 1.024 might be optimal.

5.2 Skurnick's[14] Simplification of McClenahan's Method

Skurnick's approach in [3] is essentially the same as McClenahan's. The difference is that Skurnick does not include the delay constant. Further, Skurnick does not calculate a single decay rate for the entire triangle using selected link ratios. Rather Skurnick fits a curve to each accident year and uses each curve as the sole mechanism of projecting each year's ultimate losses. Mathematically, his tail factor reduces to

$$(1-r)\Big/(1-r-r^{y})$$

where r and y are as before. In this case y denotes the number of years of development at which the tail factor will apply.

An Example

Consider the following incremental loss payouts:

Development	Accident Year	
Stage	1992	1991
12	4000	1000
24	2000	2000
36	1000	1000
48	500	500
60	250	250
72	125	125
84	62.5	62.5
96		31.25

[14] This method is also referred to as the 'Geometric Curve' method.

For illustration of the curve fitting process, the 1992 data produces the following table, when a curve is fit to the natural logarithms of the paid loss in each year (using the identity $\ln(A \times r^y) = \ln(A) + y \times \ln(r)$).

				Fitted Line					
Development	Stage	Amount	Log of	Ln(A) =	8.987	EXP = A =	8000	Fitted	Fit
Stage	in Years	Paid	Amount	Ln(r) =	-.693	EXP = r =	0.5	Curve	Error
12	1	4,000	8.29405					4,000	0
24	2	2,000	7.600902					2,000	0
36	3	1,000	6.907755					1,000	0
48	4	500	6.214608					500	0
60	5	250	5.521461					250	0
72	6	125	4.828314					125	0
84	7	63	4.135167					63	0

The tail factor is then $(1-.5)/(1-.5-.5^7)=.5/(1-.5-.007813) = 1.0159$.

The above is of course a contrived example. But consider the more typical case of the 1991 accident year. In this case, the payments begin low, then decrease after reaching a 'hump' in the 24 month stage. The eventual rate of decrease is still .5, but the curve fit produces:

				Fitted Line					
Development	Stage	Amount	Log of	Ln(A) =	8.294	EXP = A =	4000	Fitted	Fit
Stage	in Years	Paid	Amount	Ln(r) =	-0.578	EXP = r =	0.56123	Curve	Error
12	1	1,000	6.907755					2,245	-1,245
24	2	2,000	7.600902					1,122	878
36	3	1,000	6.907755					561	439
48	4	500	6.214608					281	219
60	5	250	5.521461					140	110
72	6	125	4.828314					70	55
84	7	63	4.135167					35	27
96	8	31	3.442019					18	14

Because of the hump shape 'r' is computed at a higher (i.e., less decay) value, .5613. Hence the tail factor is much larger at

$(1-.5613)/(1-.5613-.5613^7) = .4387/(.4387-.017554) = 1.0417$.

5.2.1 Advantages and disadvantages of Skurnick's method

The primary advantage of Skurnick's method, at least relative to McClenahan's method, is that the calculations are much simpler. But correspondingly, this method involves not only all of the assumptions underlying McClenahan's method; a constantly decreasing pattern, exponential decay, and a lack of trend in the decay rate; it adds the assumption of no lag between the accident date and when payments begin. The last assumption is clearly untrue in the vast majority of cases.

As shown above, an additional major disadvantage is that it does not accommodate 'hump shaped' patterns well. The problems with hump-shaped curves serve as an introduction to the next improvement.

5.2.2 Improvement 3 – limit curve fitting to the more mature years

Skurnick's method is a prime candidate for this approach, because it is so common to have a 'hump-shaped' payout curve, whereas by the very nature of the exponential curve, exponential curves are monotonically decreasing. So, it is logical to refocus the tail estimation process, putting primary emphasis on the type of claims activity occurring near the tail.

Going back to the 'Brief Digression' on types of claims activity, the type of claims activity most closely associated with the tail does not begin until after 48 or 60 months. So, it would be logical to just fit the development curve to the paid after 60 months. The result of performing that limited fit on the 1991 data used to illustrate Skurnick's method is shown below.

				Fitted Line					
Development	Stage	Amount	Log of	Ln(A) =	8.987	EXP = A =	8000	Fitted	Fit
Stage	in Years	Paid	Amount	Ln(r) =	-0.69	EXP = r =	0.5	Curve	Error
72	6	125	4.82831					125	0
84	7	63	4.13517					63	0
96	8	31	3.44202					31	0

As expected, this produces the correct decay rate value of 'r' = .5, and the corresponding tail factor of 1.0159.

5.2.1 A note of caution

The above improvement is logical and generally works well with large volume high-credibility data. When the triangle is of 'medium'[15] size and has a fairly high cap on loss size, the triangle will not have full credibility. Therefore, a fit to paid data directly out of the triangle will likely lead to poor tail factor estimates. Of note, Skurnick's method is not the only method where this will yield poor tail estimates. It will happen with all the curve-fitting methods.

5.2.3. Improvements 1 and 2 applied to Skurnick's method

These improvements and their processes have likely been discussed enough earlier in this paper to eliminate a need for examples. Logically, both improvements may be applied while using Skurnick's method.

Method 1, using multiple ending years can be applied by simply fitting the curve to all the payments but the last year, computing the corresponding tail factor for the next-to-last stage of development, and dividing by the last link ratio.

Method 2 can be performed just as it was in McClenahan's method. For example, in the poor curve fit obtained when fitting to all of the 1991 data, the 'development portion' of the fitted tail, 1.0417-1=.0417 could be multiplied by the ratio of the actual incremental paid loss in the 96-108 stage (31, holding the place of the exact value 31.25) to the fitted value (rounded to 18). Note though, that the 'corrected' tail factor is even further off at $1+31\times.417/18 = 1.0718$. This illustration of when improvement 2 does not improve the tail factor prediction is intended to further show what happens when the type of curve fitted is a poor match for the pattern of the data.

5.3 Exponential Decay of the Development Portion of the Link Ratios[16]

This method is the first of several methods that extrapolate the tail factor off the loss development link ratios rather than the paid loss. This method was referred to briefly in the discussion of the Bondy method as a possible source of theoretical underpinnings for the two Bondy methods. The process is very simple. Given a set of link ratios $1+d_1$, $1+d_2$, $1+d_3$,.... $1+d_y$, a curve of the form

$$D \times r^m$$

where D is the fitted development portion of the first link ratio and r is the decay constant, is fit to the d_m's. The easiest way to do so is by using a regression to the natural logarithms of the d_m's. Then, for an ending d_y of small size, the additional development can be estimated by using the previous approach of

[15] It is very difficult to quatify 'medium' in a manner that will work across the different lines of insurance and still be meaningful years in the future. At the time this was written, an example of a 'medium' volume triangle might be a very large workers compensation self-insurance fund.

[16] This method was outlined in Sherman's paper, but likely was already heavily used by actuaries before Sherman's paper was published..

$$\prod_{m=y+1}^{\infty}(1+ D\times r^m) = \prod_{m=1}^{\infty}(1+ d_y r^m) \cong 1 + d_y \sum_{m=y+1}^{\infty} r^m = 1+ d_y r/(1-r).$$

This also automatically introduces Improvement 2 by fitting exactly to the last point. Similar algebra would show that the tail factor is approximated by

$1+ D\times r^{y+1}/(1-r)$.

For an ending d_y of larger size, it may be necessary to simply project the link ratios for the next fifteen or so years (until the additional tail is immaterial), then multiply them all together to create a tail factor.

5.3.1 An example

Consider the following sample link ratio data.

Development Stage	Stage in Years	Link Ratio
12	1	1.5
24	2	1.25
36	3	1.125
48	4	1.0625
60	5	1.03125
72	6	1.015625
84	7	1.007813

The astute reader will notice that is a pattern similar to that underlying the Bondy method. In any event, to fit our exponential curve to the development portion, we first subtract unity to obtain the development portion of each link ratio.

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'
12	1	1.5	0.5
24	2	1.25	0.25
36	3	1.125	0.125
48	4	1.0625	0.0625
60	5	1.03125	0.03125
72	6	1.015625	0.015625
84	7	1.007813	0.0078125

Then, as a precursor to curve fitting, we take the natural logarithms of the development portions, or "d's".

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d'
12	1	1.5	0.5	-0.69315
24	2	1.25	0.25	-1.38629
36	3	1.125	0.125	-2.07944
48	4	1.0625	0.0625	-2.77259
60	5	1.03125	0.03125	-3.46574
72	6	1.015625	0.015625	-4.15888
84	7	1.007813	0.0078125	-4.85203

Then, we fit a line to those logarithms. Standard commercial spreadsheet software produces:

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d'	Fitted Curve Values	
					Slope	-0.6931
					Intercept	0.0000
12	1	1.5	0.5	-0.69315		
24	2	1.25	0.25	-1.38629		
36	3	1.125	0.125	-2.07944		
48	4	1.0625	0.0625	-2.77259		
60	5	1.03125	0.03125	-3.46574		
72	6	1.015625	0.015625	-4.15888		
84	7	1.007813	0.0078125	-4.85203		

Then, our 'D', or development portion at time zero, is the exponent of the intercept, and the rate of reduction, 'r' is the exponent of the slope. Calculating the exponents and the fitted curve, we get:

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d'	Fitted Curve Values Slope	-0.6931		Fitted Curve
					Intercept	0.0000		
12	1	1.5	0.5	-0.69315				1.50000
24	2	1.25	0.25	-1.38629	r = exp(slope)		0.5	1.25000
36	3	1.125	0.125	-2.07944	D = exp(intercept)		1	1.12500
48	4	1.0625	0.0625	-2.77259				1.06250
60	5	1.03125	0.03125	-3.46574				1.03125
72	6	1.015625	0.015625	-4.15888				1.01563
84	7	1.007813	0.0078125	-4.85203				1.00781
	8							1.00391
	9							1.00195
	10							1.00098
	11							1.00049
	12							1.00024
	13							1.00012
	14							1.00006
	15							1.00003
	16							1.00002
	17							1.00001
	18							1.00000
	19							1.00000
	20							1.00000
	21							1.00000
	22							1.00000

Then, for reference we compute the tail factor using both the 'quick' formula usable for small remaining 'development portions', and by multiplying the fifteen fitted link ratios that make up the tail.

Quick Formula Tail	
1+1×(.5^8)/(1-.5) =	1.00781
Product of 8-22 Links	1.00783

As one can see, the difference is negligible.

5.3.2 A more realistic example

The previous example was contrived to make the mathematics clear. Consider the following set of more realistic data.

Development Stage	Stage in Years	Link Ratio
12	1	2.000
24	2	1.250
36	3	1.090
48	4	1.050
60	5	1.040
72	6	1.030
84	7	1.028
96	8	1.020

A curve can be fit to the data using the methodology employed in the previous example.

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d'	Fitted Curve Values			Fitted Curve	Fit Error
					Slope	-0.4415			
					Intercept	-0.5723			
12	1	2	1	0.0000				1.3628	-0.6372
24	2	1.25	0.25	-1.3863	r = exp(slope)		0.643042	1.2333	-0.0167
36	3	1.09	0.09	-2.4079	D = exp(intercept)		0.56422	1.1500	0.0600
48	4	1.05	0.05	-2.9957				1.0965	0.0465
60	5	1.04	0.04	-3.2189				1.0620	0.0220
72	6	1.03	0.03	-3.5066				1.0399	0.0099
84	7	1.028	0.028	-3.5756				1.0257	-0.0023
96	8	1.02	0.02	-3.9120				1.0165	-0.0035
108	9	1.018	0.018	-4.0174				1.0106	-0.0074

Note that the fit errors exhibit some cyclic behavior, negative as a group at first, then positive from 3-6 years, then negative again at 7-9 year maturities. This suggests that the

curve may be a poor fit. That is borne out by the relationship of the tail factor estimates with and without exact fit to the last link ratio:

Quick Formula Tail	
1+D×(r^10)/(1-r) =	1.019108
Product of 8-22 Est. Links	1.019226
After exact fit to last link	
1+.0191×.018/.01061	1.032403

Once again the 'quick approximation' to the tail is almost identical to the precise tail indicated by exponential decay. However, note that because of the poor fit of the curve near the tail, the use of Improvement 2 (exact fitting to the last link ratio) produces a markedly different tail factor. The question of which tail factor is best must now be answered.

To do so, Improvement 3 (fitting the curve solely to the mature years) is in order. In this case, the curve will simply be fit to years 4 (48 months) and beyond. That produces the following fit;

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d'	Fitted Curve Values			Fitted Curve	Fit Error
					Slope	-0.2073			
					Intercept	-2.1900			
48	4	1.05	0.05	-2.9957				1.0488	-0.0012
60	5	1.04	0.04	-3.2189	r = exp(slope)		0.812748	1.0397	-0.0003
72	6	1.03	0.03	-3.5066	D = exp(intercept)		0.111915	1.0323	0.0023
84	7	1.028	0.028	-3.5756				1.0262	-0.0018
96	8	1.02	0.02	-3.9120				1.0213	0.0013
108	9	1.018	0.018	-4.0174				1.0173	-0.0007

Which produces the following tail estimates:

Quick Formula Tail	
1+D×(r^10)/(1-r) =	1.075166
Product of 10-24 Est. Links	1.075813
After exact fit to last link	
1+.075×.018/.0173	1.078035

Due to the low fit errors, as long as the 48-120 development triangle data that generated the link ratios is credible, this would strongly suggest that a tail factor of around 1.075 is needed. Note also that the 'quick approximation also works well in this instance. In summary, this example illustrates the importance of restricting use of the fitted curve to the portion of the development data that it can reasonably fit.

5.3.3 Advantages and disadvantages of this method

A primary advantage of this method is it's simplicity. The assumption of exponential decay is relatively easy to understand. The calculations have moderate complexity, but an illustration of the fitted values can readily give laypeople comfort that the method is being executed correctly. Of note, this method is 'asymptotically equal' to both McClenahan's and Skurnick's methods, yet is much simpler to execute. That also leads to it's major disadvantage. Because it assumes such a quick decay of the link ratios (exponential decay is faster decay than $1/x$, $1/x^2$, $1/x^3$, etc.), it can easily underestimate the tail.

5.4 Sherman's Method - Fitting an Inverse Power Curve to the Link Ratios

This method, the last[17] of the curve fitting approaches to be discussed, was first articulated by Sherman [2]. Sherman noted[18], while fitting a curve from the McClenahan-Skurnick-Exponential Decay family, that the 'decay ratios' (ratios of successive development portions of link ratios) were not constant as suggested by expoential decay. Rather, as one went further out in the development pattern, the decay ratios rose towards unity (i.e. there was less and less decay as one went further out in the curve). Looking at the data, it appeared that asymptotically, the decay ratios approached unity. Based on this, he posited an 'inverse power' curve of the form $1+a\times t^b$ (t representing the maturity in years) for the link ratios. Sherman then investigated the quality of curve fit to actual industry data for several families of curves, including the inverse power curve. The family that he found generally fit best were the so-called 'inverse power' curves.

The process of fitting an inverse power curve is very similar to that used to fit the exponential curve, excepting that the 'independent variable' used in the curve fit is ln(t). More specifically, the identity

$$\ln(1+d-1) = \ln(d) \equiv \ln(1+a\times t^b-1) = \ln(a\times t^b) = \ln(a) + b\times\ln(t)$$

can be used to create an opportunity to base the fitted curve on a simple regression.

[17] Sherman also discussed the fitting of a lognormal curve to the cumulative paid (or implied cumulative paid) and the fit of a logarithmic curve to the link ratios. However, the lognormal fit does not lend itself to easy spreadsheet mathematics, and the logarithmic fit to the link ratios does not produce a unique tail factor. Further, a Sherman discussed, the inverse power curve is a preferable approach.

[18] Mr. Sherman discusses this in Section III of [3].

Unfortunately, this author is not aware of any simple closed form approximation to the tail this curve generates, so the tail factor must be estimated by multiplying together the successive link ratios after the tail begins until the impact of additional link ratios is negligible.

5.4.1 An example

This may best be illustrated by using the initial dataset used for the exponential decay approach:

Development Stage	Stage in Years	Link Ratio
12	1	1.5
24	2	1.25
36	3	1.125
48	4	1.0625
60	5	1.03125
72	6	1.015625
84	7	1.007813

The first step is to calculate the development portion of each link ratio and take natural logarithms of the result.

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d'
12	1	1.5	0.5	-0.6931
24	2	1.25	0.25	-1.3863
36	3	1.125	0.125	-2.0794
48	4	1.0625	0.0625	-2.7726
60	5	1.03125	0.03125	-3.4657
72	6	1.015625	0.015625	-4.1589
84	7	1.007813	0.0078125	-4.8520

Those will represent the 'dependent variable' in our regression. Then for the independent variable, we take natural logarithms of the development stage/beginning maturity for the link ratio in years.

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d' 'X'	Log of Stage in Yrs 'Y'
12	1	1.5	0.5	-0.6931	0.0000
24	2	1.25	0.25	-1.3863	0.6931
36	3	1.125	0.125	-2.0794	1.0986
48	4	1.0625	0.0625	-2.7726	1.3863
60	5	1.03125	0.03125	-3.4657	1.6094
72	6	1.015625	0.015625	-4.1589	1.7918
84	7	1.007813	0.0078125	-4.8520	1.9459

Then, we compute the regression parameters.

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d' 'X'	Log of Stage in Yrs 'Y'	Fitted Curve Parameters		
12	1	1.5	0.5	-0.6931	0.0000	Slope =	-2.10512	=b
24	2	1.25	0.25	-1.3863	0.6931	Intercept =	-0.20881	
36	3	1.125	0.125	-2.0794	1.0986	a = exp(intercpt)	0.811553	
48	4	1.0625	0.0625	-2.7726	1.3863			
60	5	1.03125	0.03125	-3.4657	1.6094			
72	6	1.015625	0.015625	-4.1589	1.7918			
84	7	1.007813	0.0078125	-4.8520	1.9459			

Following that, we compute the fitted curve values and the fit error.

Development Stage	Stage in Years	Link Ratio	Fitted Curve Parameters			Fitted Curve	Fit Error
			Slope =	-2.10512	=b		
12	1	1.5	Intercept =	-0.20881		1.8116	0.3116
24	2	1.25	a = exp(intercept)	0.811553		1.1886	-0.0614
36	3	1.125				1.0803	-0.0447
48	4	1.0625				1.0438	-0.0187
60	5	1.03125				1.0274	-0.0038
72	6	1.015625				1.0187	0.0030
84	7	1.007813				1.0135	0.0057
	8					1.0102	
	9					1.0080	
	10					1.0064	
	11					1.0052	
	12					1.0043	
	13					1.0037	
	14					1.0031	
	15					1.0027	
	16					1.0024	
	17					1.0021	
	18					1.0018	
	19					1.0016	
	20					1.0015	
	21					1.0013	
	22					1.0012	

And, the tail factor estimates are:

Fitted Tail =	1.056977	
Exact Fit to last link		
1+0.056977×0.007813/0.0135		
=	1.032975	.

Even with the utility this adds in the fit, the initial fit produces a tail factor of over 1.05, when the previous exponential decay analysis suggested only 1.00781. The exact fit correction, though, does produce a number that is much closer to the theoretical tail.

Again, one approach is to fit solely to the mature years. That approach produces the following regression calculations:

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d'	Log of Stage in Yrs	Fitted Curve Parameters		
				'X'	'Y'	Slope =	-3.69867	=b
48	4	1.0625	0.0625	-2.7726	1.3863	Intercept =	2.413854	
60	5	1.03125	0.03125	-3.4657	1.6094	a = exp(intercpt)	11.17696	
72	6	1.015625	0.015625	-4.1589	1.7918			
84	7	1.007813	0.0078125	-4.8520	1.9459			

And then it produces the following fitted curve:

Development Stage	Stage in Years	Link Ratio	Fitted Curve Parameters			Fitted Curve	Fit Error
			Slope =	-3.69867	=b		
48	4	1.0625	Intercept =	2.413854		1.0663	0.0038
60	5	1.03125	a = exp(intercpt)	11.17696		1.0290	-0.0022
72	6	1.015625				1.0148	-0.0008
84	7	1.007813				1.0084	0.0006
	8					1.0051	
	9					1.0033	
	10					1.0022	
	11					1.0016	
	12					1.0011	
	13					1.0008	
	14					1.0006	
	15					1.0005	
	16					1.0004	
	17					1.0003	
	18					1.0003	
	19					1.0002	
	20					1.0002	
	21					1.0001	
	22					1.0001	

And, the tail it produces, although it remains higher than the theoretical tail (at a certain level, the slower decay of the inverse power curve as compared to an exponential curve makes it inevitable that it will produce a higher tail) is much closer to the theoretical tail.

Fitted Tail =	1.017077	
Exact Fit to last link		
1+0.017077×0.007813/0.0084		
=	1.015884	

5.4.2 The more realistic example

Going back to the exponential decay, a tail was fit to the more realistic link ratios shown below:

Development Stage	Stage in Years	Link Ratio
12	1	2
24	2	1.25
36	3	1.09
48	4	1.05
60	5	1.04
72	6	1.03
84	7	1.028
96	8	1.02
108	9	1.018

As in the previous example, we fit an inverse power curve:

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d' 'X'	Log of Stage in Yrs 'Y'	Fitted Curve Parameters		
						Slope =	-1.82492	=b
12	1	2	1	0.0000	0.0000	Intercept =	-0.18424	
24	2	1.25	0.25	-1.3863	0.6931	a = exp(intercpt)	0.83174	
36	3	1.09	0.09	-2.4079	1.0986			
48	4	1.05	0.05	-2.9957	1.3863			
60	5	1.04	0.04	-3.2189	1.6094			
72	6	1.03	0.03	-3.5066	1.7918			
84	7	1.028	0.028	-3.5756	1.9459			
96	8	1.02	0.02	-3.9120	2.0794			
108	9	1.018	0.018	-4.0174	2.1972			

And then we compute the fitted curve values for the link ratios that comprise the tail. Since the link ratios decay so slowly, we project thirty years of additional development instead of fifteen.

Development Stage	Stage in Years	Link Ratio	Fitted Curve Parameters			Fitted Curve	Fit Error
			Slope =	-1.82492	=b		
12	1	2.000	Intercept =	-0.18424		1.8317	-0.1683
24	2	1.250	a = exp(intercpt)	0.83174		1.2348	-0.0152
36	3	1.090				1.1120	0.0220
48	4	1.050				1.0663	0.0163
60	5	1.040				1.0441	0.0041
72	6	1.030				1.0316	0.0016
84	7	1.028				1.0149	1.0239
96	8	1.020				1.0111	1.0187
108	9	1.018				1.0151	-0.0029
	10					1.0124	
	11					1.0105	
	12					1.0089	
	13					1.0077	
	14					1.0067	
	15					1.0059	
	16					1.0053	
	17					1.0047	
	18					1.0043	
	19					1.0039	
	20					1.0035	
	21					1.0032	
	22					1.0030	
	23					1.0027	
	24					1.0025	
	25					1.0023	
	26					1.0022	
	27					1.0020	
	28					1.0019	
	29					1.0018	
	30					1.0017	
	31					1.0016	
	32					1.0015	
	33					1.0014	
	34					1.0013	
	35					1.0013	
	36					1.0012	
	37					1.0011	
	38					1.0011	
	39					1.0010	

That produces the following tail data.

Fitted Tail =	1.114487
Exact Fit to last link	
1+0.11451×0.018/0.0151	
=	1.136502

For comparison, the final 'best estimates' using the exponential decay were in the 1.03-1.05 range. But, those best estimates were based off a fit to just the mature years. So, let us fit the curve solely to the 48+ month data.

Development	Stage	Link	Development	Log of	Log of	Fitted Curve Parameters		
Stage	in Years	Ratio	Portion 'd'	d'	Stage in Yrs			
				'X'	'Y'	Slope =	-1.28108	=b
48	4	1.05	0.05	-2.9957	1.3863	Intercept =	-1.18688	
60	5	1.04	0.04	-3.2189	1.6094	a = exp(intercpt)	0.305171	
72	6	1.03	0.03	-3.5066	1.7918			
84	7	1.028	0.028	-3.5756	1.9459			
96	8	1.02	0.02	-3.9120	2.0794			
108	9	1.018	0.018	-4.0174	2.1972			

However, in this case, the tail is even higher, per the fit

Development	Stage	Link	Fitted Curve Parameters			Fitted	Fit
Stage	in Years	Ratio				Curve	Error
			Slope =	-1.28108	=b		
48	4	1.05	Intercept =	-1.18688		1.0517	0.0017
60	5	1.04	a = exp(intercpt)	0.305171		1.0388	-0.0012
72	6	1.03				1.0307	0.0007
84	7	1.028				1.0252	-0.0028
96	8	1.02				1.0213	0.0013
108	9	1.018				1.0183	0.0003
	10					1.0160	
	11					1.0141	
	12					1.0126	
	13					1.0114	
	Etc.					Etc.	

Multiplying the link ratios that comprise the tail factor together, the estimated tail is:

Fitted Tail =	1.208566
Exact Fit to last link	
1+0.2086×0.018/0.0183	
=	1.20518

So, this illustrates how this method is generally more conservative than the exponential decay method.

5.4.3 Advantages and disadvantages of Sherman's method

Relative to the other curve-fitting methods, this method's primary strengths and weaknesses stem from it's source, although that is mitigated by the fact that in choosing the form of the mathematical curve family that was used (the inverse power curve), Sherman relied heavily on actual data. Specifically, he noted that exponential decay factors flattened heavily (i.e., rose toward unity) at later ages. So, he chose the inverse power curve as his model to reduce the decay at later ages. In a sense, Sherman designed the inverse power curve with an eye toward mathematically correcting an observed deficiency in the exponential decay method. The approach he used to correct exponential decay[19] was merely to find a curve that roughly matched the data he observed. So, since the inverse power approach is based on actual properties of the observed development link ratio curves, and appears to have superior fit to the data, it should arguably be a better predictor of the tail. But on the other hand it also gives no single simple assumption (such as decay proportional to development portion size) that we can test the data against. In other areas, the fit looks a little more mathematically complex to the outsider, but is no more computationally difficult for the practitioner than exponential decay of the link ratios.

5.5 Sherman's Revised Method – Adding Lag to the Inverse Power Curve

In his study of the inverse power curve, Sherman [3] noted that the fit could sometimes be improved by adding a lag parameter to the curve. He used the formula

$1+d \cong 1+a\times(t-c)^b$.

In this case, the mechanics of fitting the curve are somewhat more complex. An example will illustrate the process.

[19] Sherman effectively replaced $1+D\,r^t$ from exponential decay with $1+a\times t^b$. Note that a in the inverse power curve plays the same role as D in exponential decay, so really he just replaced r^t, with a constant decay ratio of r by t^b, with a decay rate of $\left((t+1)\div t\right)^b$, which is asymptotically one.

5.5.1 Example of fitting an inverse power curve with lag

We first set the lag equal to one (unity) to begin the process, then fit the an inverse power curve reflecting that lag

Development Stage	Stage in Years	Link Ratio	Development Portion 'd'	Log of d'	Stage Minus Lag	Log of Rev. Stage in Yrs	Fitted Curve Parameters		
							Lag =	1	
							Slope =	-1.0273	=b
48	4	1.05	0.05	-2.9957	3.0000	1.0986	Intercept =	-1.8324	
60	5	1.04	0.04	-3.2189	4.0000	1.3863	a = exp(intercpt)	0.1600	
72	6	1.03	0.03	-3.5066	5.0000	1.6094			
84	7	1.028	0.028	-3.5756	6.0000	1.7918			
96	8	1.02	0.02	-3.9120	7.0000	1.9459			
108	9	1.018	0.018	-4.0174	8.0000	2.0794			

Then we compute the link ratios on the fitted curve, and the total squared fit error as well

Development Stage	Stage in Years	Link Ratio	Fitted Curve Parameters			Fitted Curve	Fit Error	Squared Error
			Lag =	1				
			Slope =	-1.027387872	=b			
48	4	1.05	Intercept =	-1.832444677		1.0385	-0.0115	1.32E-04
60	5	1.04	a = exp(intercpt)	0.160021887		1.0306	-0.0094	8.79E-05
72	6	1.03				1.0254	-0.0046	2.12E-05
84	7	1.028				1.0217	-0.0063	4.00E-05
96	8	1.02				1.0189	-0.0011	1.22E-06
108	9	1.018				1.0167	-0.0013	1.58E-06
								2.84E-04

We note that the total fit error associated with a lag of one is .000284.

Next, in order to estimate the optimum lag, we use a bisection process, following the process above for different potential lags; finding the lowest value of the squared error across a group of values; and progressively narrowing the range. The computations were as follows, and only 7 steps were needed. For reference, at each step of the process the lowest value of the fit error as well as the two adjacent values (the three values generated by the lag points that will be carried to the next step of the process) are in bold.

Stage 1		Stage 2		Stage 3	
	Squared		Squared		Squared
Lag	Error	Lag	Error	Lag	Error
		-1	7.06E-04	-0.5	1.58E-04
-1	**7.06E-04**	-0.5	**1.58E-04**	-0.25	**4.86177E-05**
0	**1.32E-05**	0	**1.32E-05**	0	**1.32454E-05**
1	**2.84E-04**	0.5	**9.22E-05**	0.25	**3.28903E-05**
2	7.50E-04	1	2.84E-04	0.5	9.22146E-05
3	1.08E-03				

Stage 4		Stage 5		Stage 6	
	Squared		Squared		Squared
Lag	Error	Lag	Error	Lag	Error
-0.25	4.86177E-05	-0.125	2.2949E-05	-0.0625	1.62499E-05
-0.125	**2.2949E-05**	-0.0625	**1.625E-05**	-0.03125	1.43034E-05
0	**1.32454E-05**	0	**1.3245E-05**	**0**	**1.32454E-05**
0.125	**1.72333E-05**	0.0625	**1.366E-05**	**0.03125**	**1.30419E-05**
0.25	3.28903E-05	0.125	1.7233E-05	**0.0625**	**1.36599E-05**

Stage 7					
	Squared				
Lag	Error	**Final Selection**		**0.02**	
0	**1.32454E-05**				
0.015625	**1.30389E-05**				
0.03125	**1.30419E-05**				
0.046875	1.32502E-05				
0.0625	1.36599E-05				

Note that as the fit error changes little near the minimum point, a rounded value is acceptable.

Then, that lag value may be used in the final curve fit.

Development Stage	Stage in Years	Link Ratio	Fitted Curve Parameters			Fitted Curve
			Slope =	-1.253797784	=b	
48	4	1.05	Intercept =	-1.23628766		1.0511
60	5	1.04	a = exp(intercpt)	0.290460507		1.0386
72	6	1.03				1.0307
84	7	1.028				1.0253
96	8	1.02	Fitted Tail =	1.230663894		1.0214
108	9	1.018				1.0185
	10		Exact Fit to last link			1.0162
	11		1+0.2307×0.018/0.0185			1.0144
	12		=	1.224464865		1.0129
	13					1.0117
	14					1.0106
	15					1.0097
	16					1.0090
	17					1.0083
	18					1.0077
	19					1.0072
	20					1.0068
	21					1.0064
	22					1.0060
	23					1.0057
	24					1.0054
	25					1.0051
	26					1.0049
	27					1.0047
	28					1.0045
	29					1.0043
	30					1.0041
	31					1.0039
	32					1.0038
	33					1.0036
	34					1.0035
	35					1.0034
	36					1.0032
	37					1.0031
	38					1.0030
	39					1.0029

Which provides a slightly smaller tail.

Fitted Tail =	1.2306
Exact Fit to last link	
1+0.2307×0.018/0.0185	
=	1.2244

5.5.2 Advantages and disadvantages of introducing lag in the inverse power curve

Summarizing, we can note that while the lag factor may sometimes mitigate the size of the tail, the inverse power in general tends to produce a higher tail than the exponential fit. Although it has not been illustrated herein with actual data, the inverse power curve also generally indicates higher tail factors than McClenahan's and Skurnick's methods, as those methods tend to produce results that are very similar to that of the exponential decay[20]. As before, the inverse power curve's main attraction is that it simply seems to fit the data better. However, in introducing lag it is clear that much computational complexity is added. The practitioner should evaluate whether the additional complexity produces large gains in the accuracy of the estimated tail factor.

6. SUMMARY

Several different methods for assessing tail development were presented, as well as some refinements. Hopefully, this will help the reader in his or her actuarial practice.

[20] That is because they are simply based on exponential decay of the payments rather than the link ratios. A little analysis will show that their decay patterns are about equal for 'large' maturities. If in doubt, simply consider their asymptotic properties.

Appendix 1-Tail Factor Methods Based on Counts

A.1 Introduction

Although they are less commonly used, there are several methods for estimating tail factors that are based on counts. Among these are the Sherman-Diss method, the projected unpaid severity method, and what is really a older year ultimate loss selection method instead of a tail factor method for unexpectedly low open counts that is based on maximum possible costs per claim. All of these methods do have demonstrated limitations, though. So, it is just as important to understand the limitations of each method as it is to understand the methods themselves.

A.2 The Sherman-Diss Method

The Sherman-Diss method described in [4] is a specific example of what could become a class of methods that project the open claim counts at future times, and the cost per claim at each future period. For the first step, this method involves projecting the likelihood that each 'mature' (near the tail maturity) workers compensation claim will still be open next year, the following year, the year following that, etc. using life (mortality) tables and the claimant's current age. Then, the indemnity (wage replacement) benefits each would receive in each future period (if they are still alive to collect benefits as estimated using the life table) is estimated using each worker's current annual benefit, plus an estimate of any inflation in the benefit (should any be allowed under the law of the injured worker's state). The total indemnity tail would then be calculated by extending the probability of each claimant's survival at each future period (the expected open claim counts) times the annual indemnity benefit. For the medical benefits allowed claimants under the workers compensation laws, the probability of survival to each future period is extended by the current medical inflated by an appropriate medical inflation factor. The extension of probability of survival times medical benefits produce the dollars of medical tail.

A.2.1 Pros and cons

Due to the complexity of the calculations and the status of this discussion as an appendix rather than the main paper, an example will not be provided. However, some discussion of this and the other methods in this appendix is certainly in order.

When Sherman and Diss compared their method to other tail factor methods (primarily the curve-fitting methods) on some specific workers compensation data, they found that it produced much higher tail factors than the other methods. However, when they tested their method retrospectively against actual dollar emergence on some Western state fund data, they found that as claimants achieved advanced ages (roughly at thirty to forty-plus years of development) the medical became much higher than that predicted by their method. Per their studies, it appears that as claimants achieve advanced ages, unexpected (at least per life tables and medical) additional development occurs because the main injury may cause related illnesses that are exacerbated by age and because family or spousal care for severely injured claimants must be replaced by nursing home care as the caregivers age and become infirm. So, at least for direct and unlimited workers compensation benefits, it appears that many common methods produce an inadequate tail, but that this method does not fully solve the problem.

Also note that this 'open claim count' method is suitable only for lines where benefits are paid as long as claims remain open. To this author's knowledge, the only lines of insurance that have that feature are workers compensation and disability.

Further, this method was designed for direct and unlimited claim costs, when most insurers purchase some form of specific excess reinsurance that caps the insurer's costs at some 'net retention'. Note however, that method could be revised by accumulating the total projected costs paid to each claimant and eliminating the claim once the net retention is reached[21]. In so doing, each claim would be effectively capped at the retention.

Lastly, this method only directly produces a tail factor for the mature years. If there is a low volume of claims remaining open in the older years (as is often the case), the results of this method will not be a reliable statistic for projecting the tail on the later years (i.e., they will lack credibility).

Qualifications aside, this method does create a powerful tool in the right circumstances. Futher, as time goes by it is possible that other 'remaining open count'-based methods will be developed.

[21] Of note, it is also appropriate to build in any projected costs that exceed the limit of per claim reinsurance purchased.

A.3 The Unclosed Count Method

This method also requires qualifications, but is worth discussion. Just as in workers compensation, the open status of a claim is related to payments. In most other lines the majority of payments occur at the time of claim closing. So, it is reasonable to suppose that there would be a method based on the number of claims yet to close and the average cost of each of those claims. Of course, while it may be relatively easy to estimate the number of claims that will close in the future as long as the actuary is certain that no further IBNR claims will materialize; it is usually very difficult if not impossible to estimate the average costs of closing each claim. However, in some limited circumstances, the average paid loss per closed claim of the oldest accident year may have reliably and permanently plateaued. In those specific circumstances (and only those circumstances), it would be appropriate to multiply the number of unclosed claims by the average paid loss per closed claim from the latest twelve months for the given accident year.

A.3.1 Pros and cons

This method cannot be discussed without discussing the tremendous detraction posed by blithely assuming that the current average paid loss per closed claim will equal the average cost of disposing of the open claim inventory. The author has personally seen general liability data of about 48 months maturity and fairly low volume where the average paid per claim had leveled off at around $5,000 per claim, where only four claims were open, but they were all $20,000+ claims. One major problem was that the maturity was only 48 months. So, the actuary is strongly cautioned to use this only for data of at least 96 months maturity, preferably 120 months, and to carefully review whether the remaining open claims are of the same type, class, average demand, etc. as the claims closed between, say, 96 and 120 months.

The actuary is also cautioned that if the data volume is not overwhelming large, the percentage of claims left open for the older years now may not match the percentage of claims left open at 120 months or so on the more recent years once they reach the 120 month stage. For example, if only four or so claims are left open on the older years, they will lack statistical validity (a form of credibility) in predicting what will be open when the more recent years reach the same development stage. Therefore, they will lack validity in predicting the tail factors for the more current years.

All that being said, under the right circumstances this can be a useful method. One must simply make sure that the set of underlying assumptions hold in whatever circumstance the actuary is using this method.

A.4 The Maximum Possible Loss Method

This method is a variant of the unclosed count method. It, however, does not so much create a tail factor as it does establish a maximum tail for the older years. The core idea of this method is that, given that the maximum net liability of an insurer is some net retention 'R', the liability for all the open claims should not be more than the sum of R-paid to date across all the open claims. So, to use it, given that an accident year is sufficiently mature for no IBNR claims to be reasonably possible, the remaining amounts to reach the retention (R-paid to date) are summed across all remaining open claims in the accident year. The result is not so much an estimate of the tail factor as an upper bound on tail development for that specific year. So, if application of the tail factor to a given year suggests more development than is 'possible' per the remaining amounts to reach the retention in the accident year, the ultimate unpaid loss for that accident year might be capped at the amounts remaining to reach the retention.

In the (fairly unusual) event that there are enough claims left open for this to be a statistically valid predictor of the development of the more recent years, it could be used in estimating the tail factor for all the accident years. But, one would have to be certain that this finding was statistically consistent with the initial tail factor analysis. For example, if the initial tail factor came from a curve fitting, it might be reasonable statistically that the curve fitting was simply using the wrong curve. However, if the initial tail factor came from a 'paid over disposed' method that also used the actual data in the triangle itself, the tail findings would suggest the data is internally inconsistent. In that case, greater care must be taken to understand which method is most accurate for the tail factor to be applied to the more recent years.

A.4.1 Pros and cons

This method improves on the average unpaid loss method by virtue of the fact that the amount to reach the retention need not be estimated. Rather, it is fact. However, it only produces an upper bound, not an actual best estimate.

Like the average unpaid loss method, there are often statistical reliability issues when making inferences about the tail factors of the more recent years. But, one cannot readily dispute the results as an upper bound for the older years on which the method is applied, at least as long as one is certain the prospect of additional IBNR claims is immaterial. So, like the average unpaid loss method, one must be very careful to make sure the proper assumptions hold when using it. But, unlike the average unpaid loss method, it has far more certainty surrounding the loss sizes.

Appendix 2-Developing Case Reserves on the Older Years

This method is also not so much a method for estimating the tail factors to use in incurred or paid loss development as it is a method for estimating ultimate losses in the very mature years. The scenario this addresses is that of a medium-to-low credibility (medium-to-low volume of losses in relation to the net retention) loss triangle. In that scenario it is not unusual for the remaining unpaid loss in the mature to vary significantly depending on whether a large claim, or a few large claims, or no large claims happen to have occurred and still be open in the late development stage. In such circumstances, the standard application of a tail factor may not work simply because there are not enough open claims in the mature years, or even open claims expected in the tail factor, for the law of large numbers to apply. In that case, some recognition of the specific cases remaining open (assuming no further reopenings or IBNR claims) will make the resulting ultimate loss predictions for the older years more accurate.

The process is fairly simple. Given a ratio of what it actually costs to close cases vs. the case reserves held from the 'paid loss to reserve disposed of' method, one simply multiplies that ratio times the case reserve to obtain an estimate of the unpaid loss on each of the very mature years. The ultimate loss estimate for each of those years would simply be the derived unpaid loss estimate plus the paid-to-date for each year.

A word of caution is in order, however. Remember that this method was used to estimate the ultimate loss because the law of large numbers did not work. Therefore, the unpaid losses derived using this method lack credibility in estimating the tail factor for the less mature years. So, if this method is used because the remaining unpaid losses are driven by 'luck of the draw'[22], it is illogical to use the unpaid losses from this method to estimate a tail factor for the less mature years.

Pros and cons

This method's inherent advantage is it's usefulness in low credibility situations. It's disadvantage is that it does not truly produce a tail factor, just some estimates of ultimate loss for the older years. Further, it assumes no reopenings or true IBNR claims. So, it must be used with great caution and respect for it's limitations.

[22] The astute reader will note that the adjusted case reserves are exactly what is used to develop a tail factor in the 'paid loss to reserve disposed of' method. But note that in that instance the tail is presumably based on case reserves that are large enough to have reasonable credibility.

Bibliography

[1] McClenahan, Charles L., 'A Mathematical Model for Loss Reserve Analysis', *Proceedings of the Casualty Actuarial Society*, **1975** Vol: LXII Page(s): 134-153,

[2] Skurnick, David ., Review of 'A Mathematical Model for Loss Reserve Analysis' by Charles McClenahan, *Proceedings of the Casualty Actuarial Society*, **1976** Vol: LXIII Page(s): 125-127

[3] Sherman, Richard E., 'Extrapolating, Smoothing and Interpolating Development Factors', *Proceedings of the Casualty Actuarial Society*, **1984** Vol: LXXI Page(s): 122-155

[4] Sherman, Richard E. and Diss, Gordon F., 'Estimating the Workers Compensation Tail', *Casualty Actuarial Society Forum*, **Winter 2004** Page(s) 207-282

A DATABASE IN 3-D

Alfred D. Commodore, ACAS

Abstract

Three-dimensional geometry and calculus are useful conceptual and analytical tools for working with valuations of insurance statistics. Geometry can be used to provide a pictorial representation of a database and illustrate differences between calendar, exposure/accident and policy year concepts. Calculus can be used to estimate on-level, trended and developed statistics used in ratemaking and reserving. Also, three dimensions lead to several practical two-dimensional methods when the generality of three dimensions is not required.

1. INTRODUCTION

An insurer keeps track of various information over the course of doing business. Databases are maintained recording the result of exposure base, premium and claims transactions for financial reporting, statistical reporting, ratemaking and reserving purposes, for example. The goal of this paper is to develop a pictorial representation of a database as three dimensions (3-D) with values attached to points. The result is a simple, conceptual and analytical tool useful for working with valuations of insurance statistics. Our approach is summarized as follows:

- Section 2 develops 3-D as a pictorial representation of a database.
- Section 3 includes conceptual applications of Section 2 using an example from ratemaking.
- Section 4 includes closing comments.
- The Appendix provides further discussion of the 3-D approach and analytical applications of Section 2 using methods from basic calculus.

2. A DATABASE IN 3-D

A variety of exposure base, premium and claims transactions make up a database and are collected to produce reports. In this section we: 1) represent a database as points in 3-D with values attached; and 2) represent a report as a collection of points in 3-D with a collective value attached.

2.1 A Database in 3-D

For the transaction: "On a policy issued 1-1-04 covering a claim incurred 8-1-04, pay $5,000 on 2-1-05."; we identify three dates and a value: 1) "1-1-04", the policy date or date the policy was written or issued; 2) "8-1-04", the exposure date or date the policy was in force and exposed to loss; 3) "2-1-05", the valuation date or date the transaction was made; and 4) "$5,000 paid losses", the amount and type of statistic. By defining xyz-space coordinates x, y and z as **policy date**, **exposure lag** and **valuation lag**, respectively – so that x+y and x+y+z are defined as exposure date and valuation date, respectively – we can represent the transaction as the "valued" point in 3-D, ((1-1-04,7,6),5000), length 1 on an axis equal to 1 month.

TABLE 1
TRANSACTION, ((X,Y,Z),S), 3-D COORDINATES

Combinations	Interpretations*
• x, y, z	• policy date, exposure lag, valuation lag
• x+y, x+y+z	• exposure date, valuation date
• y+z, z	• policy age, exposure age

*Policy, exposure and valuation date are synonymous with issue/written, in-force and transaction date, respectively. Exposure date and lag become claim (or accident) date and lag, respectively, when dealing with claim statistics. Note the two interpretations for the z-coordinate as a type of lag and age.

2.1.a. *In 3-D represent the transaction: "To a policy written, exposed and valuated at times x_o, x_o+y_o and $x_o+y_o+z_o$, respectively, assign s_o statistic units."; as valued point $((x_o,y_o,z_o),s_o)$.*

Table 1 lists the various definitions and interpretations of xyz-coordinate combinations used throughout the paper. We plot several transactions on a policy issued 1-1-04 next, length 1 on an axis equal to 1 month [1].

2.1.a.1. The initial transaction to record $4,320 in written premium is made at issue so that the valuation date equals the policy date; ((1-1-04,0,0),4320).

2.1.a.2. The transaction to record $12 in earned premium on 3/1/04 is made at the time the policy is in-force on 3/1/04; ((1-1-04,2,0),12) [2].

2.1.a.3. On 6/1/04 corrective transactions are made since the policy was actually written for $8,640 and recorded in error. Transaction ((1-1-04,0,5),4320) corrects for written premium transaction 2.1.a.1 and transaction ((1-1-04,2,3),12) corrects for earned premium transaction 2.1.a.2.

2.1.a.4. An endorsement for additional coverage for $400 written premium is processed mid-term on 9/1/04; ((1-1-04,8,0),400).

2.1.a.5. The policy is cancelled without penalties on 12/1/04 with a full refund of a month's premiums unearned on the base policy for $720(=$8,640/12) and on the endorsement for $100(=$400/4); ((1-1-04,0,11),-720) & ((1-1-04,8,3),-100).

2.1.a.6. The transaction to record a claim incurred 8/1/04 is made on 8/15/04; ((1-1-04,7,0.5),1).

A transaction's policy and valuation dates are recorded in practice. Exposure date – unless associated with a date of loss – is not recorded and is an abstract concept we introduce for purposes of the presentation. Imagine that an annual policy is made up of 365 separate daily policies, 8,760 separate hourly policies, etc., until we view the policy as a **post**[3] continuum. We introduced exposure date so that transactions tracked by policy in practice could be tracked by post in theory.

[1] We invite the reader to express each transaction in the form provided by 2.1.a by identifying the three dates and value, and plot all transactions in the same 3-D diagram.
[2] This transaction is "implicit" as no actual record is made to record earnings.
[3] Mnemonic for "a Policy expOSed to loss at a point in Time".

We extend 2.1.a to apply to a database of transactions. For simplicity we assume all the transactions in the database are for the same statistic.

2.1.b. *In 3-D represent a database of transactions for a single statistic by $D=(P,f)$.*

In 2.1.b, P is the collection of points, (x,y,z), resulting from plotting transactions using 2.1.a and f is the rule assigning value to points of P, $f(x,y,z)$, derived as the sum of the statistic over all transactions assigned to point (x,y,z). Thus, $D=(P,f)$ is a collection of points of the form, $((x,y,z),f(x,y,z))$. Figure 1(a) is a generic database D with the f-values assumed color-coded to simplify the illustration. Note:

2.1.b.1 D ranges from the collectible (e.g., written premium transactions in company records) to the hypothetical (e.g., written premium projections).

2.1.b.2 D (or more precisely P in $D=(P,f)$) appears differently by type of statistic. Under some basic conventions D is: 1) confined to the x-axis for written statistics; 2) confined to the xy-plane for earned statistics; and 3) unrestricted in space for claim statistics. [4]

2.1.b.3 D appears differently by coverage for the same statistic. Private Passenger Auto Physical Damage paid losses close rather quickly when compared to Workers Compensation paid losses. Transaction activity that occurs long after the accident date is reflected in larger z-coordinate values for points in D. Thus, we would expect D^* for auto paid losses to be generally closer to the xy-plane than D^{**} for Workers Compensation paid losses.

2.2 A Report in 3-D

A report involves collecting statistical information from specific database transactions. In 3-D this amounts to identifying a subset of space along with the total statistic value associated with that subset. The report: "Accident Year 2001 as of 12/31/2002 totaled $31.3 Million Paid Losses."; is a very basic type of report we call a **valuation** characterized by: 1) a **data organization** (i.e., Accident Year 2001); 2) a **status** (i.e., "as of 12/31/2002"); and 3) a statistic **level** (i.e., $31.3 Million Paid Loss).

2.2.a. *In 3-D represent a valuation by O.V.s.*

In 2.2.a, O and V (O containing V) are subsets of space determined by the valuation's data organization and status, respectively, and level s assigned to V is the total statistic for points $P\cap V$. [5] We often write V for O.V.s, O and s understood.

---

[4] The conventions arise under the assumptions written, earned and claim statistic transactions ((x,y,z),s) are made *only* at policy inception (i.e., x+y+z equals x), moment in force (i.e., x+y+z equals x+y) and after the date of loss (i.e., x+y+z exceeds x+y), respectively, and so reduce to forms ((x,0,0),s), ((x,y,0),s) and ((x,y,z>0),s), respectively - assuming non-negative coordinates x, y and z. 2.1.a.3 includes counterexamples for the written and earned statistic conventions. As for a counterexample to the claim statistic convention, confine hypothetical transactions to record ultimate claim counts to the xy-plane.

[5] If $P\cap V=\varnothing$, then s is undefined. We consider two valuations distinct if they differ at either of O or V but *equivalent* if their V's have the same intersection with P. We can talk about the level or equivalence (to another subset) of an arbitrary subset A by considering the valuation where A=O=V.

TABLE 2
VALUATION, O.V.S, ILLUSTRATED

Valuation Data Organizations, O, as Subsets of Space	Valuation Statuses, V, as Subsets of O
• Policy issued at time t with term k: All points with policy date x=t and exposure lag y at most k, (t, y≤k,z).	• As of date *t*: V is all points of O with valuation date x+y+z at most *t*.
• Post/Accident at exposure lag e on a policy issued at time t: All points with policy date x=t and exposure/accident lag y=e, (t,e,z).	• At policy age *a*: V is all points of O with policy age y+z at most *a*.
• Exposure/Accidents at time t: All points with exposure/accident date x+y=t, (x,t-x,z).	• At exposure/accident age *e*: V is all points of O with exposure/accident age z at most *e*.
• Time t: All points with valuation date x+y+z=t, (x,y,t-x-y).	• Over calendar period t1 to t2: V is all points of O with valuation date x+y+z from t1 to t2, inclusive.
• Policy, Exposure/Accident or Calendar Period t1 to t2: All points with policy, exposure/accident or valuation date from t1 to t2, inclusive.	• At ultimate: V equals O. This is equivalent to *t*, *a* and *e* becoming infinite in the as-of-date-t, at-policy-age-*a* and at-exposure-age-*e* statuses, respectively.

Note: Level s is the collective value of P∩V. See Table 1 for various date, lag and age definitions.

Thus, using 2.2.a our valuation example is represented by drawing all points, (x,y,z), with accident date x+y in the year 2001 and assigning $31.3 Million Paid Losses to the subset with valuation date x+y+z at most 12/31/2002.[6] We close this section with comments on the primary valuation data organizations and statuses, which we define with drawing instructions in Table 2 and illustrate in Figure 1.

2.2.a.1. A **policy** provides coverage during its term, coverage at any point mid-term referred to as a **post**[7]. Figure 1(b) shows the annual policy written 12/31/04, which by Table 2 is drawn as all points with policy date x equal to 12/31/04 and exposure lag y at most 1 year. The post (and the **claim** it covers) on 6/30/05 is drawn as all points in the policy with exposure/claim date x+y equal to 6/30/05.

2.2.a.2. An **exposure**[8] is coverage from all policies in force at a point in time and so concurrent claims are covered by an exposure. Figure 1(b) shows the exposure (and concurrent claims) on 1/1/03, drawn as all points with exposure date x+y equal to 1/1/03.

2.2.a.3. A point in **time**, itself, is drawn as all points with a given valuation date. Figure 1(b) shows time 12/31/02 as all points with valuation date x+y+z equal to 12/31/02.

2.2.a.4. Finally, policies, exposures and times combine to form **periods** of the same. Figure 1(b) shows Calendar Year (CY) 2002, Exposure/Accident Year (E/AY) 2003 and Policy Year (PY) 2004.

2.2.a.5. Valuation **statuses** for data organizations in 2.2.a.1-2.2.a.4 indicate which points to collect for a total statistic. **As-of-date, at-policy-age** and **at-exposure-age** statuses collect transactions for a data organization thru a certain valuation date, age of underlying policies and age of underlying

[6] The valuation is drawn in Figure 2.
[7] Introduced in Section 2.1 and equal to the intersection of a policy and an exposure as point sets.
[8] We use the term "exposure" as a coverage concept. Other uses (e.g., type of insured, insurance coverage limit) are found in the literature.

exposures, respectively. For example, in Figure 1(c) the status: 1) as-of-date-w=1/1/04 collects transactions with valuation date $x+y+z \leq 1/1/04$; 2) at-policy-age-w=18 months collects transactions with policy age $y+z \leq 18$; and 3) at-exposure-age-w=18 months collects transactions with exposure age $z \leq 18$. **Over-calendar-period** status identifies transactions made on or between two dates, and **at-ultimate** status takes into account all transactions, past, present and future.

3. APPLICATIONS

In this section we apply $D=(P,f)$ conceptually using the geometry of P to illustrate ideas from ratemaking. Analytical applications, where we consider how f behaves on P using calculus in 1, 2 or 3 dimensions as required, are reserved for the Appendix.

XYZ Company proposes a rate change for PY 2004 to be effective 1-1-04. They estimated ultimate losses and ultimate earned premiums (written using the current manual) for PY 2004 to be $45.0 million and $67.3 million, respectively. With a permissible loss ratio target of 65%, a rate level change of 2.9%(=45.0/67.3/0.65-1) was indicated. We sketch their approach as follows, using Figure 2 as a guide.

3.a. LOSS PROJECTION (Figure 2(a)): PY 2004 ultimate losses are estimated from AY(=EY) 2001 losses. As of the latest valuation, 12-31-02, AY 2001 paid losses are $31.3 million. Development to ultimate would add another $9.6 million. Finally, $4.1 million trends or conforms the AY 2001 experience to the PY 2004 basis. The result is $45.0 (=31.3+9.6+4.1) million in estimated ultimate losses for PY 2004.

3.b. PREMIUM PROJECTION (Figure 2(b)): PY 2004 ultimate earned premium is estimated from CY 2001 earned premium. CY 2001 earned premium was $60.3 million. Of that amount, $45.9 million was earned on policies written under the current manual that became effective 1-1-01. The remaining $14.4 million would be increased $0.6 million if underlying policies had been written using the current manual. Finallly, $6.3 million trends the CY 2001 experience to the PY 2004 basis. The result is $67.3 (=60.3+0.6+6.3) million in estimated ultimate earned premium for PY 2004.

In practice we might use several AYs and CYs in pricing PY 2004, applying some weighted average of the results to derive our final estimate. Moreover, instead of using AY 2001 losses and CY 2001 premium, we could use losses and premiums from the same data organization.

4. CONCLUSIONS

We represented a database pictorially by assuming three dates and a statistic value could be associated with each transaction. Note:

4.a. In constructing a database in 3-D we first draw all policies then populate those policies with transactions. A policy, exposure and post/claim are infinite in height. We generally assume D consists of points with non-negative xyz-coordinates. By allowing z to be less than zero we can represent transactions made prior to policy issuance. For example, transaction ((1-1-04,0,-1),$1,000) is a $1,000 premium renewal received 1 month in advance of the 1-1-04 renewal date.

4.b. In 3-D the origin should, in principle, correspond to a date before or on the effective date of the very first policy written. Policy term may be unlimited (e.g., a title insurance policy) and the picture for a policy is independent of the claim "trigger" (e.g., occurrence or claims-made triggers).

4.c. The list of valuation data organizations and statuses in Table 2 is not exhaustive, but representative of the valuations that often arise in practice. A variety of valuations can be found in Schedule P of the NAIC Annual Statement. CY, AY, EY and PY valuations can be found in Schedule P, Parts 1, 2, 6 and 7, respectively. Moreover, we consider PY, EY and CY the fundamental data organizations, with AY a special case of EY when we are dealing with claim statistics.

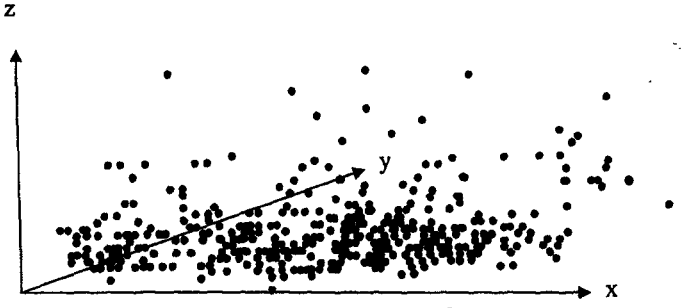
4.d. Development triangles in Schedule P for AY, EY and PY are on an as-of-date-t basis. We could also set up development triangles on an at-policy-age or at-exposure/accident-age basis. All three approaches partition a data organization using planes at ever increasing height in the z or "development" direction.

REFERENCES

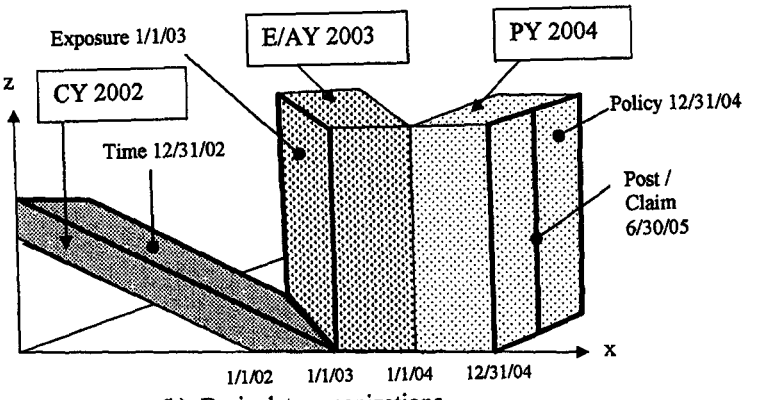
[1] Miller, D. L. and Davis, G. E., " A Refined Model for Premium Adjustment," PCAS LXIII, 1976, pp. 117-124.

[2] Cook, C. F., "Trend and Loss Development Factors," PCAS LVII, 1970, pp. 1-14.

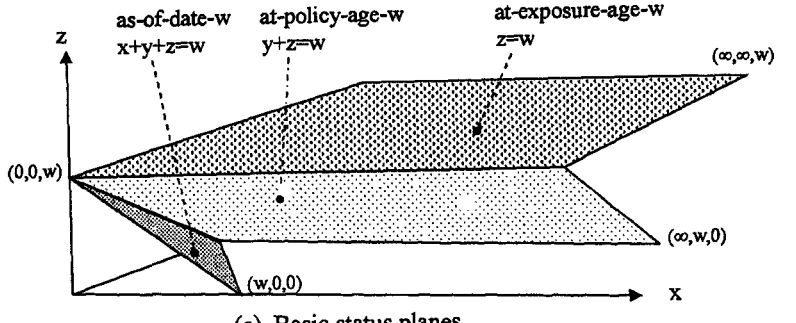
FIGURE 1
A DATABASE IN 3-D



(a) A database in 3-D, D=(P,*f*)

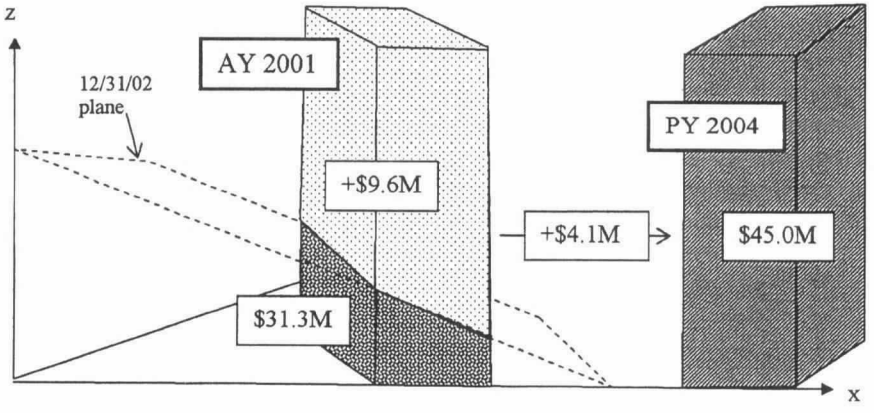


(b) Basic data organizations

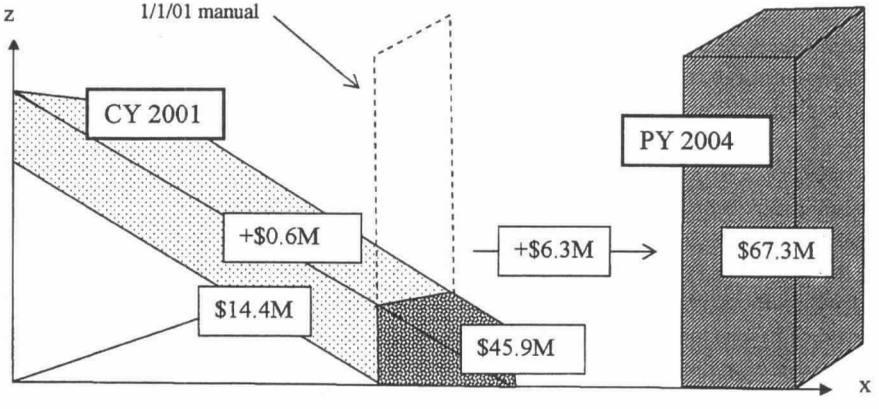


(c) Basic status planes

FIGURE 2
RATEMAKING ILLUSTRATED



(a) Loss Projection: Accident Year 2001 to Policy Year 2004



(b) Premium Projection: Calendar Year 2001 to Policy Year 2004

APPENDIX

ANALYTICAL APPLICATIONS

Let D=(P,f) be a database in 3-D [9] and O.V.s. a valuation. We assume set V and density f sufficiently defined so that we may calculate s as the integral of f over V. We apply densities for various statistics for D in 1-, 2- and 3-dimensions, dimensions limiting valuation variety.[10] Five basic analytical applications are described below:

A.1. *Developed, On-Level & Trended Statistics*: Developing involves determining the change in level for two valuations of the same data organization. On-leveling involves determining the change in level for some V under different rules f_1 & f_2. Finally, trending involves determining the change in level for V_1 & V_2 under the same rule f.

A.2. *Average Value and Average Point*: Define the *average value* of V as s/|V| and the *average point* or *center* of V as that with average coordinates, v*=(x*,y*,z*) [11]. The center provides the average policy, exposure/claim and valuation date, for example. The center for $f(x,y)$ constant or uniform over certain V in the xy-plane is readily determined as follows:

a. For V a rectangle or parallelogram, x* is the midpoint of the x range and y* the midpoint of the y-range for V.
b. For V an isosceles right triangle, x* is 1/3 into the x-range and y* 1/3 into the y-range of V, starting from the vertex at the right angle.

For the examples that follow, in xyz-space length 1 on an axis equals 1 year and x=0 and x=1 correspond to dates 1/1/00 and 1/1/01, respectively.

EXAMPLE 1. As an example of an on-level calculation, we estimate the change in earned premium for CY 2001 at actual and current rate levels, given: i) policies are annual term; ii) the manual effective 1/1/01 represents the current rate level; and iii) D_{EP} for earned premium is confined to the xy-plane with density $f_{EP}(x,y)=(4000x)(\$100)$ for policies written prior to 1/1/01 and $f_{EP}(x,y)=(4000x)(\$100)(1.2)$, thereafter.[12] In Figure 3(a) we show CY 2001 split by rate level. The desired factor is On-level EP ÷ Actual EP, calculated as follows:

---

[9] With **R** the set of real numbers, $f: P \subseteq \mathbf{R}^3 \rightarrow \mathbf{R}^1$ implies D is the graph of f, a subset of $\mathbf{R}^3 \times \mathbf{R}^1$.

[10] Two subsets A and B of space are said to be equivalent if they have the same intersection with P, written "A≡B". For example, D confined to the x-axis, xy-plane and xz-plane implies CY≡E/AY≡PY, CY≡E/AY and E/AY≡PY, respectively.

[11] |V|>0 is the content (i.e., length, area or volume) of V and v* is defined only when f>0 on V with x*, y* and z* calculated as the integral of xf/s, yf/s and zf/s over V, respectively.

[12] Here: i) 4000x is the density for written exposure base units (e.g., car years, payroll, stadium seats, etc.) earned uniformly with respect to lag y; ii) $100 is the average written premium; and iii) factor 1.2 represents a 20% increase in rate level on 1/1/01.

$$OEP\ /\ AEP\ = \int_0^1\int_{1-y}^{2-y} 480{,}000\,xdxdy\ /[\int_0^1\int_{1-x}^{1} 400{,}000\,xdydx\ +\ \int_1^2\int_0^{2-x} 480{,}000\,xdydx\ ]$$

$$1.059 = \$480{,}000\ /\ [\$400{,}000(1/3)+\$480{,}000(2/3)].$$

If exposure writings were uniform instead of increasing (e.g., replace 4,000x with constant 4,000 in the two densities) then the on-level factor becomes:

$$1.091 = \$480{,}000\ /\ [\$400{,}000(1/2)+\$480{,}000(1/2)].$$

Example 1 is taken from [1]. In that paper the authors perform the same on-level factor calculations, however using a geometric orientation based on the traditional parallelogram method. *For D confined to the xy-plane, the transformation sending point (x,y) to (x+y,y) results in illustrations under the traditional parallelogram method.*[13]

EXAMPLE 2. As an example of a trend calculation, we estimate the change in ultimate loss ratios between PY 2001 @12/31/01 and PY 2004, given: i) ½-year policy terms; and ii) 3% accident year loss cost trend; and iii) 1% on-level policy year premium trend. We assume:

- ultimate loss and ultimate earned premium densities take the forms $f_{UL}(x,y)=wl\times(1.03)^{x+y}$ and $f_{EP}(x,y)=wp\times(1.01)^{x}$, respectively, w, l and p exposure base unit, loss cost and premium constants, respectively[14]; and
- loss and premium levels for V are estimated by $s\equiv f(v^*)|V|$ where v^* is the uniform center of V, so that $l(103)^{x^*+y^*}/p(1.01)^{x^*}$ is the loss ratio estimate for V.

Applying (A.2.a) & (A.2.b) to Figure 3(b), the centers under uniformity for PY 2001 @12/31/01 and PY 2004 are $(1.389,0.222)$[15] and $(4.5,0.25)$, respectively. The desired trend factor estimate is therefore 1.06378 ($= (1.03^{(4.75-1.611)} / 1.01^{(4.5-1.389)})$.

Example 2 supports a common calculation made in practice. Using actual in place of approximate integrations results in a trend factor of 1.06380. Integration has the advantages of following directly from the density assumptions and differentiating between V's with the same uniform center.

---

[13] For example, apply the transformation to Figure 3. Several 2-D plotting methods also arise from "collapsing" 3-D. In particular, we note mappings of (x,y,z) to 2-D planes (x,y), (x+y,y), (x,y+z), (x+y,z) and (x+y+z,y).

[14] $l(1.03)^t$ is the result of an exponential fit of a series of accident year average loss costs. With the on-leveling adjustment treated separately, $p(1.01)^t$ is the result of an exponential fit of a series of policy year average earned premiums at current rate level.

[15] The center for EY 2001 is the weighted average of centers for its components from PY 2000 and PY 2001. Thus, solving for (x_o,y_o) in: $(1.25,.25)= .25(1\text{-}1/6,.50\text{-}1/6)+.75(x_o,y_o)$; yields $(x_o,y_o)=(1.389,.222)$.

EXAMPLE 3. We provide the calculations behind the XYZ Company loss and premium values presented in 3.a and 3.b from Section 3. In Figure 3(b) we show the "footprint" of C/E/AY 2001 and PY 2004 in the xy-plane. Assume ½ year policy terms. Let D_{PL} for paid losses be 3-dimensional with density, $f_{PL}(x,y,z)=196{,}000(\$400)e^{.03x+.015y-z}$. Let D_{EP} for earned premium be 2-dimensional with density $f_{EPa}(x,y) = 196{,}000(\$575)e^{.04x}$ for x prior to 1/1/01 and $f_{EPb}(x,y) = 196{,}000(\$600)e^{.04x}$ thereafter. The three loss and three premium integrations required are as follows:

$$AY2001\_PL@12/31/02 = \int_0^{.5}\int_{1-y}^{2-y}\int_0^{3-x-y} f_{PL}(x,y,z)\,dz\,dx\,dy = \$31{,}329{,}071$$

$$AY2001\_PL@ultimate = \int_0^{.5}\int_{1-y}^{2-y}\int_0^{\infty} f_{PL}(x,y,z)\,dz\,dx\,dy = \$40{,}852{,}442$$

$$PY2004\_PL@ultimate = \int_0^{.5}\int_4^{5}\int_0^{\infty} f_{PL}(x,y,z)\,dz\,dx\,dy = \$45{,}036{,}196$$

$$CY2001\_EP@actual = \int_0^{.5}\int_{1-y}^{1} f_{EPa}(x,y)\,dx\,dy + \int_0^{.5}\int_1^{2-y} f_{EPb}(x,y)\,dx\,dy$$

$$= \$14{,}444{,}165 + \$45{,}977{,}558 = \$60{,}421{,}723$$

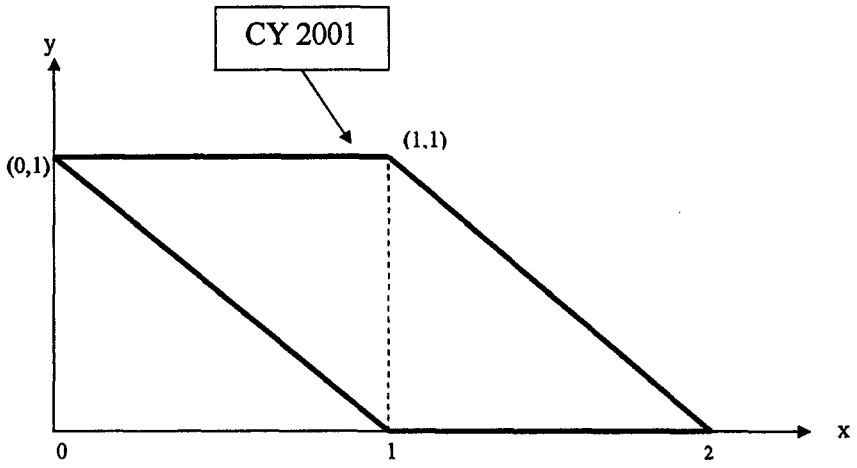
$$CY2001\_EP@on-level = \int_0^{.5}\int_{1-y}^{2-y} f_{EPb}(x,y)\,dx\,dy = \$61{,}049{,}730$$

$$PY2004\_EP@on-level = \int_4^{5}\int_0^{.5} f_{EPb}(x,y)\,dy\,dx = \$67{,}301{,}286.$$

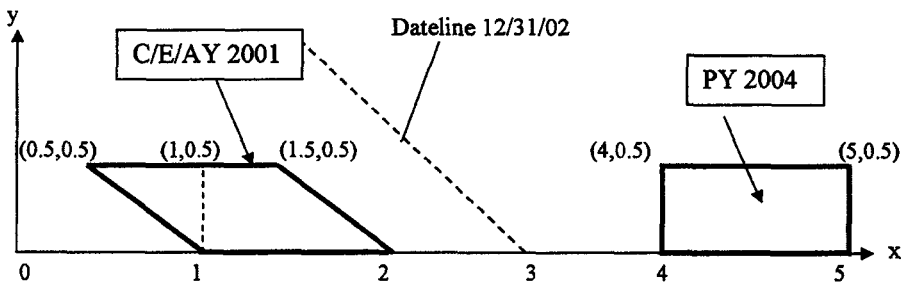
Example 3 densities were developed from assumptions on the rate at which units are written and earned, and the relationship between premiums and losses for a given risk. The densities took the forms $f_{EP}(x,y) = w(x)e(x,y)p(x,y)$ and $f_{PL}(x,y,z) = w(x)e(x,y)l(x,y)q(z)$, where:

- $w(x)=10^5e^{.01x}$ is the rate at which units are written at time x;
- $e(x,y)=1.96=2(.98)$ is the rate at which units written at time x are earned at lag y. The integral of e(x,y)dy over term $0\le y\le 1/2$ equals 98% due to a 2% cancellation rate on average;
- p(x,y) is the post premium: the product of a base rate (\$575 for x prior to 1/1/01 and \$600 thereafter) and premium relativity $e^{.03x}$.
- l(x,y) is the post loss cost: the product of base loss cost \$400, loss cost relativity $e^{.005x}$ and inflation factor thru the date of loss $e^{.015(x+y)}$.
- $q(z)= e^{-z}$ is the portion of loss l(x,y) paid at lag z. The integral of q(z) over $0\le z\le\infty$ equals 100%.

FIGURE 3
EXAMPLES 1, 2 & 3 DIAGRAMS



(a) Example 1 diagram



(b) Examples 2 & 3 diagram

Reinsurance Involving Partial Risk Transfer
Addressing the Accounting Difficulties

Spencer M. Gluck, FCAS, MAAA

Abstract: The paper proposes a measure for risk transfer, the portion or percentage of risk transferred ("**PRT**") that varies between 0% and 100%. Such measure would provide a superior basis for a binary decision between reinsurance accounting and deposit accounting (with a likely critical value of 50%). A preferred approach would be to use **PRT** as the basis for continuous accounting. The paper differentiates between "natural" reinsurance contract provisions that do not limit risk transfer and "structural" contract provisions that may limit risk transfer. The **PRT** measures the risk-limiting impact of the structural provisions by comparing risk distributions before and after the application of structural provisions. **PRT** is 100% for contracts without structural provisions. The risk to be measured is defined as potential adverse deviation from the amounts reflected in accounting values. Fixed reinsurance contract provisions that are accounted for without uncertainty provide no potential for adverse deviation and do not affect **PRT**. The paper includes a discussion and critique of the FAS 113 definition of risk transfer, and finds two fundamental flaws: (1) the definition is based on an absolute measure of the riskiness of the ceded cash flows, so that reinsurance of low risk subject portfolios often fails even though nearly all the risk is transferred, while reinsurance of high risk subject portfolios often passes even though the risk transfer is severely limited; and (2) the focus on reinsurer profitability includes fixed amounts that are unrelated to risk, and thereby includes an implicit standard for reinsurance pricing that is an inappropriate role for accounting. The paper includes examples of the application of PRT and several other risk transfer measures to a range of underlying cash flows and reinsurance contract structures.

Introduction and Summary

Reinsurance contracts frequently contain any number of risk limiting provisions, which may call into question the validity of reducing net losses and premiums by showing them as having been ceded to the reinsurance, i.e. "reinsurance accounting". Many or most such contracts cede some, but not all of the relevant risk, which the author describes as partial risk transfer.[1]

There are concerns that some partial risk transfer contracts have been used to manipulate financial statements. Yet there are many legitimate uses of partial risk transfer, and more that may develop in the future as sophisticated tools for risk management. Furthermore, there may be risks for which reasonably priced reinsurance is available only with risk-limiting provisions. The author's view is that opportunities for financial statement manipulation arise from inaccurate accounting. The author's proposal for more accurate accounting would substantially eliminate opportunities for manipulation while allowing the legitimate use and further development of structured risk transfer techniques.

[1] More common terms are "structured risk" and "finite risk". The author prefers partial risk transfer, which corresponds more directly with the basis of the approach. Partial risk transfer includes many traditional risk sharing

Currently, the accounting choice is whether or not the contract in question has enough risk transfer to qualify as reinsurance, and therefore be eligible for reinsurance accounting. FAS 113[1], for U.S. GAAP, and SSAP 62[2], for SAP,[2] provide guidance for making this choice.

The author's central thesis is that the degree of risk transfer in a reinsurance contract can be described by a relatively simple and intuitive measure called "the percentage of risk transferred" or "***PRT***", which should be the basis for the above decision. The central provisions for defining risk transfer in FAS 113 are found to be fundamentally flawed.

Section I:

- develops the underlying basis for the central thesis,
- contrasts the approach with FAS 113,
- defines the approach specifically, and
- applies the approach, along with several others, to a range of hypothetical cash flow models and hypothetical reinsurance contracts.

The second aspect of the central thesis is that the two available accounting choices are appropriate for 100% risk transfer and 0% risk transfer, but that neither is truly appropriate for partial risk transfer. Section II illustrates how the measure developed in Section I can be used to develop appropriate accounting for partial risk transfer contracts.

[2] The relevant language is generally identical in FAS 113 and SSAP 62. For brevity, references hereinafter will be to FAS 113.

Outline

Section I - Defining and Measuring Risk Transfer in Reinsurance Contracts

1.1 Risk Transfer and Accounting
- 1.1.1 Risk and Balance Sheets/Income Statements
- 1.1.2 Risk and Net Premiums and Losses
- 1.1.3 Reinsurance Accounting vs. Deposit Accounting
- 1.1.4 The Relevant Risk
- 1.1.5 Partial Risk Transfer

1.2 The FAS 113 Definition of Risk Transfer – Discussion and Critique
- 1.2.1 Measuring Risk Rather than Risk Transfer
- 1.2.2 Re-Pricing the Reinsurance

1.3 The Percentage of Risk Transfer ("*PRT*") Approach
- 1.3.1 Defining 100% Risk Transfer: Natural vs. Structural Contract Provisions
- 1.3.2 The Applicable Cash Flows
- 1.3.3 The Risk Model
- 1.3.4 Adverse Deviation from Accounting Values
- 1.3.5 Risk Measures and Co-Measures I
- 1.3.6 The Percentage of Risk Transferred ("*PRT*")
- 1.3.7 Some Advantages of the *PRT* Approach

1.4 Risk Measures and Co-Measures II
- 1.4.1 Definitions and Examples
- 1.4.2 Measures and Co-Measures Applied

1.5 Examples Comparing Risk Transfer Measures: *PRT* vs. "Absolute" Risk Measures
- 1.5.1 The Risk Transfer Measures
- 1.5.2 The Subject Business Models
- 1.5.3 The Reinsurance Contracts
- 1.5.4 Risk Transfer Measures Applied to Subject Business
- 1.5.5 Risk Transfer Measures Applied to Quota-Share Contracts
- 1.5.6 Risk Transfer Measures Applied to Structured Aggregate Excess Contracts
- 1.5.7 Conclusion

1.6 Examples Using *PRT* with Various Risk Measures and Co-Measures

Section II – Accounting for Partial Risk Transfer Reinsurance

Section I - Defining and Measuring Risk Transfer in Reinsurance Contracts

1.1 Risk Transfer and Accounting

The effects of risk transfer accounting are subdivided into two basic categories:

- Effects on overall reported equity and income; and
- Effects on reported net premiums and losses.

1.1.1 Risk and Balance Sheets/Income Statements

In the most straightforward case, consider the reinsurance premium net of ceding commission to be the sum of the mean discounted ceded losses and the reinsurer's margin. The initial impact of reinsurance on balance sheets and income statements consists of a cost - the reinsurer's margin, and a gain - the difference between the ceded losses and their mean discounted value. While reinsurance accounting and deposit accounting differ on the timing of the recognition of the cost, our primary focus is on the gain, or more specifically, on the cession of incurred losses and loss reserves. The impact on incurred losses will be controlled by the impact on loss reserves.

Loss reserves for most P/C liabilities are recorded at estimated nominal (undiscounted) value, i.e., an estimate of the sum of future outgoing cash flows. It is important to distinguish the reserve from the liability itself. The liability is more complex, the sum total of the insurer's obligations under the relevant policies. The reserve is simply a valuation of the liability, possibly a surrogate for a market value.

If the same future cash flows were not estimates, but simply future payment obligations that were fixed in amount and timing, then it is clear that the value of those obligations would be the discounted value of the future payments, and the liability would be accounted for as such. The accounting difference between an at-risk insurance liability and the corresponding no-risk liability is precisely the discount. The (unrecognized) discount then is the required risk load. It

exists precisely because the liabilities are subject to insurance risk and would not exist if they were not.

Under an "economic value" accounting concept (not currently applicable under U.S. GAAP or SAP), the implicit risk margin in the unrecognized discount may be replaced by an explicitly discounted reserve and an explicit risk margin. The issues to be discussed subsequently regarding ceding the reserve and its associated risk margin would be equally applicable if the risk margin were converted from implicit to explicit.

1.1.2 Risk and Net Premiums and Losses

For shorter tail business, where loss reserves and their implicit risk margin are small, the choice of accounting will have little impact on overall equity or income. However, the characterization of premiums and losses as having been ceded (or not) affects the reported net premiums, losses, and loss reserves. Various measures of capital adequacy used by rating agencies, regulators, and other publics use net premiums, net losses, net loss reserves, etc. as measures of the risk to which a company is exposed.[3] Accounting for premiums and losses as ceded when the corresponding risk has not been ceded, or has been partially ceded, distorts these measures.

1.1.3 Reinsurance Accounting vs. Deposit Accounting

When accounting for a ceded reinsurance contract (perhaps we should say a purported reinsurance contract), we currently have two options: reinsurance accounting or deposit accounting.

Under reinsurance accounting, reserves are ceded on the same basis that they are established: in most cases at undiscounted, and therefore implicitly risk-loaded, value. Since the net recorded liability for the ceded cash flows is reduced to zero, the underlying assumption is clear – that the liability itself has been ceded, both at the recorded estimate and at all other possible outcomes.

The risk load has been 100% eliminated, which is appropriate only if 100% of the risk has been ceded. Similarly, since premiums and losses have been 100% ceded, capital adequacy measures, regulatory ratios, etc. also assume a 100% cession of the related risk.

For contracts that do not qualify for reinsurance accounting, the AICPA Statement of Position 98-7[3] and SSAP No. 75[4] provide rules for deposit accounting under GAAP and SAP, respectively. The "interest method" is prescribed for all reinsurance contracts under SAP. Under GAAP, the same method is prescribed except for contracts that transfer underwriting but not timing risk, or that have indeterminate risk. The interest method assumes that no reinsurance transaction has occurred, in other words, that 0% of the risk has been ceded.

1.1.4 The Relevant Risk

For equity and income, the choice between reinsurance accounting and deposit accounting hinges on whether it is appropriate to eliminate (by cession) the risk load imbedded in the carried loss reserves. To discuss whether this risk has been ceded, we must define the relevant risk more precisely. What risk does this risk load provide for?

The author believes that it is fairly clear that the relevant risk is the risk of inaccuracy in the estimate that is on the balance sheet. If we consider only downside risk to be important, then it is the risk of inadequacy of the estimate. If we view the balance sheet value as a surrogate for market value, the risk load is the amount in addition to the discounted value required to fund the mean losses that an assumer of the liability would require to compensate for the risk of inadequacy in the mean estimate.

This description of risk is consistent with a concept of risk as related to economic or financial losses. The risk as defined above is the risk of the insurer realizing losses subsequent to the statement date related to the loss reserves to be ceded.

---

[3] This paper does not necessarily endorse the validity of any particular capital adequacy measure. For example, capital adequacy measures that use net premiums as a surrogate for underwriting risk have a number of

While the previous paragraphs refer to loss reserves, we will normally view risk prospectively, i.e. at the inception of the reinsurance contract, before statement values are established. How do we define the risk of future losses? If the expected losses create an underwriting loss, then actual losses worse than expected create a future loss. If the expected losses create an underwriting profit, then actual losses worse than breakeven create a future loss.

All further analysis herein will be based on a definition of risk as adverse deviation from actual or expected statement values. For prospective losses, adverse deviation is measured relative to expected losses or underwriting breakeven losses, whichever is higher.

Note that fixed amounts, which create no accounting uncertainty as to their value, are not relevant. In particular, ceded premiums, to the extent that they are not contingent on losses, will be accounted for in their normal straightforward manner with no risk of accounting inaccuracy. The size of those fixed premiums, and therefore of the reinsurer's profit margin, does not affect the question of whether the insurer has retained or ceded the risk for its losses, only the question of at what cost. Whatever the cost, that cost will be expensed under normal accounting procedures, and therefore creates no additional risk for the insurer.

1.1.5 Partial Risk Transfer

Many reinsurance contracts have risk-sharing provisions (e.g., retrospective rating, adjustable commissions, profit sharing, refundable experience accounts), and/or risk limiting provisions (e.g., aggregate limits, sub-limits, additional premiums). These provisions may reduce, but not necessarily eliminate, the transfer of risk. In such cases, neither of the assumptions underlying the available accounting options – 100% risk transfer or 0% risk transfer – is precisely accurate.

The question before us is stated narrowly: Given that we have only these two options, which shall we use? A likely answer is: The one that is more nearly accurate. In other words, does the contract more nearly transfer 100% of the risk or 0% of the risk?

---

imperfections and potential distortions that shall not be discussed further.

In order to answer this question, we need to estimate, for any reinsurance contract, the portion, or percentage, of the risk that has been transferred ("***PRT***"). In fact, a reasonable definition of ***PRT*** is fairly simple, and the modeling required to estimate the value is no more complex or difficult than the modeling required to perform risk transfer testing under FAS 113 as currently written. Both require the same risk model of the underlying cash flows.

Once the ***PRT*** has been estimated, the choice of accounting treatment can be decided by comparing the ***PRT*** to a critical value. A critical value of 50% would seem to best answer the question of which accounting treatment is more nearly accurate, though other critical values might be chosen.

The above test will provide a practical, intuitive answer to the narrow question which will, in the author's opinion, represent a significant improvement to current practice. It will minimize the degree of accounting inaccuracy to the extent possible under the constraint that we have only the two accounting treatments to choose from. Nonetheless, it must be recognized that neither of the available accounting treatments is in fact designed for partial risk transfer, and both will be inaccurate to some degree. The definition and estimation of the ***PRT*** can also provide the basis for practical accounting for partial risk transfer. While this is a larger change to current accounting practice, the difficulties that arise from inaccurate accounting for partial risk transfer cannot be eliminated until partial risk transfer reinsurance is formally recognized and appropriate accounting is promulgated.

A previous reference to measuring risk transferred by comparing "before" and "after" distributions is noted in the report of the CAS Valuations, Finance and Investment Committee ("VFIC") [5]. The reference is to an approach described for testing the basis risk in catastrophe derivatives [6].

1.2 The FAS 113 Definition of Risk Transfer – Discussion and Critique

The well known FAS 113 definition of adequate risk transfer is that it must be "*reasonably possible that the reinsurer may realize a significant loss from the transaction*" [1]. The determination must be

based on all cash flows to the reinsurance contract, whether characterized as losses, premiums, expenses, etc., but transactional expenses and the reinsurer's expenses are not included.[4] The terms "*reasonably possible*" and "*significant loss*" are not specifically defined, but some guidance is given and the well known "10/10" rule is frequently applied to test whether a contract meets the FAS 113 definition.

The 10/10 rule has frequently been discussed and criticized and a number of potentially superior risk measures have been suggested. The author's critique is more fundamental: The FAS 113 definition of risk transfer is fundamentally flawed, not just because of problems with the risk measures, but because the wrong risk is being measured.

The two fundamental defects:

1. The definition of risk transfer does not contain the concept of risk transfer. Rather, the FAS 113 definition sets an absolute standard of the required level of assumed risk. A test of risk transfer requires a comparison of "before" and "after" risk. No single absolute standard can produce results that are meaningful regardless of the riskiness of the underlying cash flows.

2. The definition is influenced by fixed profit margins paid to the reinsurer. As discussed in the previous section, in determining proper accounting from the cedant's perspective, the relevant risk is the risk that the amounts carried in the cedant's financial statements are inadequate. Fixed profit margins are irrelevant. Furthermore, it is inappropriate for the risk transfer analysis to be influenced by the analyst's implicit second-guessing of the reinsurance pricing, which is unavoidably the case when applying the FAS 113 definition.

Each of these defects is further explored below:

1.2.1 Measuring Risk Rather than Risk Transfer

[4] While the definition is stated from the reinsurer's perspective, the exclusion of transactional and reinsurer's expenses actually convert it to the cedant's perspective. A more accurate expression would be "reasonably possible

A problem that may arise from the FAS 113 definition that has been frequently discussed by others is that obvious risk transfers of low risk portfolios may not pass. FAS 113 provides that obvious 100% risk transfer contracts need not be tested. The specific language is that the previous test would not apply if "*the reinsurer has assumed substantially all of the insurance risk relating to the reinsured portion of the underlying insurance contracts*" [1]. Unstructured quota-share contracts are generally accepted to fall within this "safe harbor". While such contracts need not be tested, it would nevertheless be desirable if such contracts would pass the test.

A number of practitioners have explored risk measures that should be superior to the 10/10 rule. Whatever the risk measure, a critical value must be selected, and "obviously risky enough" contracts should pass. Even with a fairly low threshold, unstructured quota-shares of stable, profitable business may still fail – the solution will still be imperfect and the exception will still be required.

But the corresponding problems at the other end of the risk spectrum, which have rarely been explored, may be even more significant. Imagine that the underlying ceded cash flows are extremely risky long-tailed payments. Because of the long tail, the distinction between discounted and undiscounted reserves (the implicit risk margin) is large and the choice of accounting treatment is highly material. Let us further assume that the reinsurance contract is highly structured so that only 20% of the risk is transferred. If we have set the critical value of the risk measure low enough so that a modestly risky quota-share will pass (as we must), then 20% of the risk on these extremely risky cash flows will also pass. If so, the cedant will be eligible for reinsurance accounting and will record on its books a 100% cession of the relevant reserves including a 100% elimination of the risk margin, even though in fact 80% of the risk has been retained -- a material accounting inaccuracy. This example is hardly purely hypothetical.

---

that the *cedant* may realize a significant *gain* from the transaction."

The example demonstrates that there is no absolute standard of riskiness, no matter how good the risk measure, that can apply equally to all incoming cash flows which themselves contain various degrees of risk.

1.2.2 Re-Pricing the Reinsurance

The author has already presented a first principles case that the relevant risk is the risk in the cedant's financial statements, and that fixed premium amounts are irrelevant to the issue of whether the cedant's risk has been transferred. Risk relates only to uncertainty.

A significant problem with the FAS 113 definition is that the risk analysis in this approach inherently includes an opinion on the appropriateness of the reinsurance pricing. There should be no better measurement of value than the actual price agreed to by a willing buyer and a willing seller in a free market. Furthermore, there may be any number of valid reasons, in volatile and cyclical markets, for a buyer to agree to pay a more conservative price at any given time. Accounting should be concerned with properly recording the actual price paid, not passing judgment on it, and any inherent "re-pricing" of the reinsurance is undesirable.

For example, in the past year, we have seen several cases where risk transfer has been questioned by auditors for straightforward casualty excess-of-loss contracts without adjustable provisions. Assuming that the FAS 113 "safe harbor" does not clearly apply in this case, the auditors were simply diligently applying the provisions of FAS 113. In these cases, the FAS 113 test failed simply because the analyst's risk model implied that the reinsurance was overpriced. Apparently, the consensus of the assuming and ceding companies was otherwise.

1.3 The Percentage of Risk Transfer ("*PRT*") Approach

To define ***PRT's*** between 0% and 100%, we first require a definition of 100% risk transfer. The author presumes that the meaning of 0% risk transfer is self-evident, and no more discussion is necessary.

1.3.1 Defining 100% Risk Transfer: Natural vs. Structural Contract Provisions

Practitioners have a fairly good idea regarding the meaning of 100% risk transfer as well. The safe harbor provision of FAS 113 provides a starting point. Recalling that language, the reinsurer must have "*assumed substantially all of the insurance risk relating to the reinsured portion of the underlying insurance contracts.*" The definition may be adequate, but could be clarified. For example, it should be clear that a traditional per-claim excess-of-loss reinsurance contract is covered, even though the per-claim retentions and limits in the reinsurance contract do not necessarily correspond to provisions in the underlying insurance contract, and might not be considered as defining the "*reinsured portion.*" Yet per-claim retentions and limits are not generally believed to be risk-limiting structures.

To more specifically define 100% risk transfer, we introduce the concept of "natural provisions" of a reinsurance contract. These would be generally defined as provisions that do not limit the losses ceded to the contract in a way that the cedant's own liability, as it relates to premiums and losses that would be ceded to such contract, is not similarly limited. We introduce the term "structural provisions" to refer to provisions that involve risk-limiting or risk sharing. Any reinsurance contract containing only natural provisions would be deemed to contain 100% risk transfer.

The author's suggested list of natural provisions:

- Percentage multipliers (e.g. quota-share, surplus share);
- Deductibles, retentions, limits, on a per claim, per claimant, or per risk or per basis;
- Deductibles, retentions, limits, on a per occurrence basis in some cases;
- Exclusions applied on a policy or coverage basis;
- Deductibles or retentions in the aggregate for all or subsets of the subject losses.

We describe the losses that would be ceded to a contract applying only the natural provisions as being in their "natural form".

Structural provisions are those that limit the ceded losses in ways that the cedant's own liability for such losses is not similarly limited or that create additional cash flows contingent upon the natural form losses. Common provisions of this type include:

- Aggregate limits, applied to the total of natural form losses or sub-limits applying to a subset of the natural form losses;
- Corridors, whether applying to the total natural form losses or a subset;
- Limits on an occurrence basis in some cases;
- Exclusions on a type of claim basis;
- Additional premiums;
- Experience accounts and profit sharing provisions;
- Retrospective rating;
- Sliding scale commissions;
- Limited reinstatements;
- Reinstatement premiums.

Neither list is necessarily exhaustive, and new types of provisions may be developed. Ultimately, the determination of whether a provision is considered natural or structural will have to be made by applying the basic principles. Hopefully, it will usually be a fairly straightforward matter.

Note, for example, that per occurrence limits have been included in both lists. In the context of catastrophe reinsurance, occurrence limits are natural. There is no cession of premiums or losses that implies that a risk has been eliminated when in fact it has not. On the other hand, in the context of quota-share reinsurance, a catastrophe occurrence limit or exclusion is structural. Ceding premiums and losses under the quota-share implies that the risks associated with those premiums and losses are also ceded, and the provision limits the risk that is transferred.

Note that for the most part, aggregate provisions are considered structural. An exception has been suggested for aggregate deductibles or retentions as these are not viewed as risk-limiting.

The reader may notice that the list of structural provisions includes a number of risk-sharing and risk-limiting provisions that are common features of traditional reinsurance. In particular, limited reinstatements and reinstatement premiums are universal in catastrophe reinsurance and common in some other high risk reinsurance; nevertheless, they are technically structural as they limit ceded risk in a way that the cedant's own risk is not limited. However, as commonly practiced, the exhaustion of available reinstatements occurs only at very remote probabilities and reinstatement premiums are not typically a large percentage of ceded losses; therefore, the risk limiting effect of these provisions is not likely to be substantial.

Having now defined 100% risk transfer, we are ready to measure partial risk transfer, for contracts containing structural provisions.

1.3.2 The Applicable Cash Flows

Given that natural provisions are not risk-limiting, the analysis of risk transfer is an analysis of the impact of structural provisions. For ease of expression, we will use the familiar terms "gross", "ceded", and "net", relative to the structural provisions, with all values reflecting the natural provisions.

Let L be a random vector (i.e. a string of values) representing the cash flows for losses subject to a reinsurance contract.

Gross:

Let: $g(L)$ = the net present value of the losses that would be ceded to that contract applying only natural provisions, gross of structural provisions.

For convenience, we have combined the processes of applying the natural provisions and taking the net present value into a single function.

Ceded:

Let: $c(L)$ = the net present value of the cash flows ceded to the contract, applying all provisions, both natural and structural.

The ceded cash flows may include premium refunds or other favorable cash flows not accounted for as ceded losses, e.g. favorable commission adjustments (for compactness, we will refer to all such adjustments as refunds). For certain calculations we require these to be separately identified. Therefore, we define $c_r(L)$ as the net present value of refunds, and $c_o(L) = c(L) - c_r(L)$ as the net present value of other ceded cash flows, i.e. loss recoveries less unfavorable adjustments.

Net:

Let: $n(L) = g(L) - c(L)$ = the net present value of the net cash flows to the cedant arising from natural losses, i.e. the net cash flows due to structural provisions.[5]

Also, let $n_o(L) = g(L) - c_o(L)$, and $n_r(L) = -c_r(L)$. As for the ceded, we have separately identified the net cash flows arising from refunds.

FAS 113 requires that all cash flows, no matter how characterized, be included in the analysis. In the above, all such cash flows would be included in $c(L)$, and consequently in $n(L)$. That approach can be used here as well; however, fixed cash flows will have no impact. Only contingent cash flows, i.e. cash flows that can vary based on the value of L, are essential.

1.3.3 The Risk Model

[5] Sign convention: Ceded losses under $g(L)$ and $c(L)$ have positive values reflecting positive cash flows to the cedant. Positive values of $n(L)$ are unfavorable, reflecting decreased cash flows to the cedant due to the structural provisions. For example, if the structural provision is a loss limitation, then $c(L)$ will sometimes be smaller than $g(L)$. The resulting positive value of $n(L)$ indicates an unfavorable cash flow effect. If the structural provision is a premium refund, then $c(L)$ may sometimes exceed $g(L)$. The resulting negative value of $n(L)$ indicates a favorable cash flow effect.

As with FAS 113, we require a risk model giving the probability distribution of L and the resulting probability distributions of $g(L)$, $c(L)$, and $n(L)$.

Given the book of business that the insurer expects to write and intends to cede, and the reinsurer intends to reinsure, the goal of the risk model is to reflect all of the uncertainty in L, including the uncertainty in both the amount and timing of the payments.

Risk is often sub-divided into "process" and "parameter" risk.

Process Risk: Given that L is the result of a random process, the process risk refers to the risk arising from the randomness of that process. Typically, the random process will be described by a mathematical model which allows the analyst to calculate (often by simulation) the effects of the random process.

Parameter Risk: The remaining risk relates to the uncertainty about the model of the random process. The term "parameter risk" is often used to broadly describe this remaining risk. More generally, the risk relates to the uncertainty in both the parameters and the form of the risk model. For example, if the total of the payments in L is modeled as a lognormal distribution with a certain mean and variance, there will be uncertainty as to whether the parameters (i.e. mean and variance) are correct as well as whether the lognormal is the correct form for the distribution. The portion of the risk model relating to uncertainty in payment timing may be more complex and more uncertain in its parameters and form.

Underlying types of risk that contribute to parameter risk may include:

- Data Risks: The amount, stability, and applicability of available data.
- Market Risks: Uncertain market impact on pricing, underwriting, risk selection.
- Economic Risks: The impact of uncertain future inflation, employment, etc.

Actual risk model structures and estimation are beyond the scope of this paper.

An important exception is that it would be inappropriate to include the risk that the company will write a different from expected book of business, e.g., a different mix of classes, coverages, policy limits, etc. This is not a risk that reinsurance is necessarily expected to absorb. Reinsurers may include provisions, some of which may be structural in form, to protect them against the cedant altering its book of business. For example, a sub-limit on a hazardous class of business may be set at a level that is remote relative to the intended book, but would be significantly risk-limiting if that class were to grow dramatically. The impact of the provision is appropriately measured against the intended book only.

1.3.4 Adverse Deviation from Accounting Values

Adverse deviation is defined relative to the financial statements. Typically, financial statement values correspond to a single loss scenario. Accordingly, we define adverse deviation relative to a base cash flow scenario, corresponding to the expected losses or the underwriting breakeven losses, whichever is higher. Let a be the vector representing the base cash flow stream.

Base Values:

Gross: Define the base value for $g(L)$ as $b_g = g(a)$. Frequently $b_g = \mathbf{E}[g(L)]$, but not necessarily in all cases.

Net: Define the base value for $n(L)$ as $b_n = n_o(a)$ ***minus the carried asset for refunds*** under cash flow scenario a (assuming reinsurance accounting). Note that an asset has a negative sign relative to net losses. Here we are using the distinction between cash flows related to refunds (n_r) and other cash flows (n_o). The distinction is necessary since the carried asset for refunds is frequently less than $n_r(a)$ – see example 2 below. If $n_r(a)$ were included in the base, it would

result in an adverse deviation whenever the refund was less than its expected value, even if no asset were carried for the refund.

Note that b_n will often be neither $n(a)$ nor $\mathbf{E}[n(L)]$, and may frequently be zero. Two examples for illustration:

1. The structural feature is an aggregate limit larger than $\sum a$. $n(L)$ is zero for $\sum L \leq$ the limit, and positive for $\sum L >$ the limit. $\mathbf{E}[n(L)]$ is therefore positive, but at scenario a, there are no net losses. $n_0(a)$ is zero, and thus $b_n = 0$.
2. The structural feature is a premium refund based on an experience account that accrues interest. At scenario a, a refund would be due, given accrual of interest, meaning that $n_r(a)$ would be negative. Further assume that no refund would be due at scenario a if accrual of interest were ignored. Under these circumstances, normally no asset is carried for the premium refund, and therefore $b_n = n_0(a) \square n(a)$.

Adverse Deviation:

The adverse deviations for $g(L)$, $n(L)$, and $c(L)$ are defined as:

$d_g = g(L) - b_g$, if positive, and zero otherwise;

$d_n = n(L) - b_n$, if positive, and zero otherwise; and

$d_c = d_g - d_n$.

Negative values are eliminated for d_g to reflect the basic principal that risk is defined by adverse results only. A negative value for d_n indicates that the effect of structural provisions is more favorable than is reflected in the accounting values (typically a premium refund larger than the asset – if any – carried for it), which does not increase the cedant's downside risk. Negative values for d_n are eliminated so that favorable effects of structural provisions cannot decrease the risk transfer measure.

Note that:

$$g(L) = n(L) + c(L); \text{ and}$$
$$d_g = d_n + d_c.$$

Note also that if fixed cash flows have been included in $c(L)$ and therefore in $n(L)$, they will be identical in the base values and all other values and will not affect the adverse deviations. In keeping with previously stated principles, these adverse deviations represent the relevant risk we intend to measure.

1.3.5 Risk Measures and Co-Measures I

Given a random variable, X, a risk measure, $r(X)$ is a function applied to the distribution of X that returns a single value.

Next assume that X is itself the sum of a number of random variables, i.e.:

$$X = \sum X_i.$$

For a broad class of risk measures, there are corresponding "co-measures" that can be applied to the sub-variables X_i.[6] The most common example of a risk measure and co-measure is variance and covariance. Co-measures provide a mathematically sound basis for allocating risk among sub-variables that may be dependant.

For risk measure $r(X)$, denote the corresponding co-measure applied to the sub-variable X_i as $r_i(X_i)$. The essential property of co-measures is additivity, i.e.:

$$r(X) = \sum r_i(X_i),$$

regardless of the nature of any dependencies among the X_i's.

[6] See Kreps [7].

In our specific case, $r(d_g) = r_n(d_n) + r_c(d_c)$. Thus, co-measures provide a basis for allocating the risk in the losses gross of structural provisions to the net and ceded losses after the application of structural provisions.

Another useful property of co-measures is that, for any constant k,

$$\text{if } X_i = kX, \quad \text{then} \quad r_i(X_i)/r(X) = k.$$

Thus, a co-measure applied to an x% quota-share is x% of the risk measure applied to a 100% share.

A more complete definition of co-measures along with examples of actual risk measures and co-measures follows the next section.

1.3.6 The Percentage of Risk Transferred ("*PRT*")

Simply stated, the *PRT* is the portion of the risk associated with the natural losses, gross of the structural provisions, which is still ceded after the application of the structural provisions. Specifically:

Let r be a risk measure with corresponding co-measure.

The percentage of risk transferred is then defined as:

$$PRT = 1.0 - \frac{r_n(d_n)}{r(d_g)}$$

or equivalently,

$$PRT = \frac{r_c(d_c)}{r(d_g)}$$

With PRT defined, adequate risk transfer to qualify as reinsurance would be defined as a value of PRT in excess of a selected critical value. A natural choice for the critical value may be 50%, as previously discussed.

1.3.7 Some Advantages of the *PRT* Approach

1. Risk transfer is reduced to a simple single number with an intuitive meaning.

2. Safe harbors for obvious risk transfer contracts are an integral part of the risk transfer definition, rather than exceptions.

3. The approach is equally valid regardless of the relative riskiness of the subject losses.

4. The approach is unaffected by profit margins and expenses. The approach avoids the second-guessing of the reinsurance pricing that is implicit in the FAS 113 definition.

1.4 Risk Measures and Co-Measures II

1.4.1 Definitions and Examples:

Define a risk measure r applied to a random variable X as:

$$r(X) = \mathrm{E}[\, w(X) \cdot l(X) \mid \mathit{Condition}\ (X)\,],$$

where l is a linear function and w is a weighting function. Note that the weights, w, may be a function of X and are unrestricted as to form. The condition may also be functionally dependant on X.

For a sub-variable X_i, the corresponding co-measure is:

$$r_i(X_i) = \mathrm{E}[\, w(X) \cdot l(X_i) \mid \mathit{Condition}(X) \,].^{7}$$

Note that the weights and the condition depend only on X, not X_i, and are identical to the weights and condition in $r(X)$.

As an example, consider variance:

$$\mathit{Variance}(X) = \mathrm{E}[(X - \mathrm{E}(X))^2] = \mathrm{E}[(X - \mathrm{E}(X)) \cdot (X - \mathrm{E}(X))]$$

In this form, the first occurrence of $(X - \mathrm{E}(X))$ can be considered the weight and the second occurrence the linear function. There is no condition.

Next, consider covariance:

$$\mathit{Covariance}(X_i,X) = \mathit{Variance}_i(X_i) = \mathrm{E}[(X - \mathrm{E}(X)) \cdot (X_i - \mathrm{E}(X_i))]$$

Note that the weight is dependant only on X and is identical to the weight used in variance, and the linear function is applied to X_i. Thus, covariance satisfies the definition of a co-measure relative to variance.

By adding a condition, we define the semi-variance:

[7] This is one formulation consistent with the framework presented in [7]. The separate condition is convenient for our use, but could have been subsumed in the weights.

$$\textit{Semi-variance}(X) = \mathrm{E}[\,(X - \mathrm{E}(X))^2 \mid (X > \mathrm{E}(X)\,],$$

with the average restricted to the values greater than the mean. The corresponding co-measure is:

$$\textit{Semi-variance}_i(X_i) = \mathrm{E}[\,(X - \mathrm{E}(X)) \cdot (X_i - \mathrm{E}(X_i)) \mid (X > \mathrm{E}(X)\,]$$

Again, the condition is based on X, not X_i.

1.4.2 Measures and Co-Measures Applied

We next consider actual applications, applied to the problem at hand.

Mean Square Adverse Deviation ("MSAD")

Define:

$$\textit{MSAD}(d_g) = \mathrm{E}[\,d_g^2 \mid d_g > 0\,].$$

Recall that $d_g = g(L) - b_g$ for positive values. Often, $b_g = \mathrm{E}[g(L)]$, in which case,

$$\textit{MSAD}(d_g) = \textit{Semi-variance}(g(L)).$$

The corresponding co-measure applied to d_n is:

$$\textit{MSAD}_n(d_n) = \mathrm{E}[\,d_n \cdot d_g \mid d_g > 0\,]$$

The condition is again based on d_g rather than d_n. Therefore, the average may (and often will) include values of $d_n = 0$.

Expected Adverse Deviation ("EAD")

Eliminating the quadratic weight from ***MSAD*** leaves us with the simpler Expected Adverse Deviation:

$$EAD(d_g) = \mathrm{E}[\, d_g \mid d_g > 0\,],$$

with the corresponding co-measure:

$$EAD_n(d_n) = \mathrm{E}[\, d_n \mid d_g > 0\,].$$

Tail Value at Risk ("TVaR")

TVaR is a popular risk measure for capital adequacy. It is similar to ***EAD***, except the borderline condition is a percentile of the distribution. Normally, relatively high percentiles are used, reflecting a belief that the most significant risk is exposure to extreme events.

Define ***VaR-p*** (d_g), the "Value at Risk," as the p^{th} percentile of the distribution of d_g. Then,

$$TVaR\text{-}p(d_g) = \mathrm{E}[\, d_g \mid d_g > VaR\text{-}p(d_g)\,]$$

with the corresponding co-measure:

$$TVaR\text{-}p_n(d_n) = \mathrm{E}[\, d_n \mid d_g > VaR\text{-}p(d_g)\,]$$

Of the above three choices, the author's preference is for ***MSAD***.

TVaR and other tail-oriented measures are often used for measuring capital needs. In the context of measuring risk transfer, the measures have several drawbacks. One is that the selected percentile is arbitrary, which may not be desirable for a single measure to be widely applied. Another is that these measures, when used with relatively high percentiles, are responsive only to a small portion of the distribution, and many structural risk-limiting provisions may be ignored.

EAD is at the other end of the spectrum, considering the entire downside of the distribution without any greater weight to values in the tail. Most models for pricing risk assume that more extreme values have greater impact.

MSAD, like ***EAD***, includes the entire downside of the distribution, and will therefore be responsive to any risk limiting provisions. ***MSAD*** is quadratically weighted, so that values toward the tail of the distribution have more impact. It is a relatively conventional risk measure, closely related to semi-variance, with the difference that deviations are measured from an accounting value which may differ from the mean.

Some practitioners believe that the quadratic weighting of ***MSAD*** does not give sufficient weight to the tail. The structure of co-measures can accommodate more complex weighting schemes, including tail-heavier weights, as well as risk loading methods based on transformations of the probability distribution. The VFIC paper [5] discusses two such transforms, the Wang Transform [8], and an Exponential Transform [9]. While such transforms are normally applied to the entire distribution, they could applied as measures and co-measures to the distributions of d_g and d_n to develop corresponding ***PRT***'s.

1.5 Examples Comparing Risk Transfer Measures: *PRT* vs. "Absolute" Risk Measures

The following examples use four measures to evaluate risk transfer: ***PRT*** and three different "absolute" risk measures. The absolute measures in this case refer to risk measures applied to the distribution of reinsurer's profit, as defined by FAS 113. They are described as absolute measures since they apply to the riskiness of a single distribution, as contrasted with ***PRT*** which is based on a comparison of riskiness in "before" and "after" distributions. The measures are applied to four different illustrative models of underlying subject losses with different degrees of

volatility, and up to five different reinsurance contract structures. All measures are based on 10,000 simulations.

1.5.1 The Risk Transfer Measures

In all cases below, the reinsurer's result is calculated according to the FAS 113 rules, i.e., the net present value of all cash flows to the reinsurer, however characterized, but without deducting transaction costs and without allowance for the reinsurer's internal expenses. All present values are at 4%. We will characterize a net loss to the reinsurer as a negative result.

1. **VaR-90**: The reinsurer's result as a percentage of ceded premium at the 90th percentile (adverse) of the distribution (given the above sign convention, this is actually the 10th percentile). Applying a critical value of -10% yields the "10/10" rule.

2. **TVaR-90**: The expected value of the reinsurer's result as a percentage of ceded premium, given reinsurer's result less than VaR-90. There is no standard critical value. 10% of the ceded premium has been suggested as a "more correct" 10/10 rule; however this is invariably less strict than the 10/10 rule. The VFIC paper suggests -25%, though this seems unusually high.[8] A range of -10% to -15% appears more in line with other measures.

3. **Expected Reinsurer's Deficit ("ERD")**: The expected value of the reinsurer's result as a percentage of ceded premium, given a reinsurer's result less than zero, multiplied by the probability that the reinsurer's losses are greater than zero. Equivalently:

$$ERD = \int_{x<0} xf(x)dx \Big/ NPV(Cededpremium)$$

[8] The VFIC paper calculates a TVaR-90 of 42% for a quota-share with 10% volatility, similar to one of the examples used herein. However, that quota-share may be under priced. A graph appears to indicate that the reinsurer's median discounted profit is zero, meaning that the reinsurer's mean profit will be less than zero, even before consideration of transaction costs or the reinsurer's internal expenses. This illustrates the difficulties with using risk transfer measures sensitive to the reinsurance pricing.

Again, there is no standard critical value. In subsequent discussion we will use a range of -1.0% to -2.0%.

4. ***PRT***, using ***MSAD*** as the risk measure.

1.5.2 The Subject Business Models

M1: Low volatility, short payment pattern.

M2: Modest volatility, modest payment pattern.

M3: Higher volatility, longer payment pattern (e.g., primary casualty).

M4: High risk, long payment pattern (e.g., excess casualty).

Table 1 summarizes the assumptions for the various models:

Table 1
Summary of Subject Business Models

		Model			
		M1	M2	M3	M4
Premium		$100	$100	$100	$100
Expenses		$30	$30	$30	$30
Expected Losses		$68	$69	$73	$83
CV		5%	10%	20%	40%
Underwriting Profit		2.0%	1.0%	-3.0%	-13.0%
Profit Including Discount		3.6%	4.3%	6.0%	11.3%
Payout	1	90%	50%	20%	1%
	2	10%	30%	20%	3%
	3		15%	10%	5%
	4		5%	10%	7%
	5			10%	7%
	6			10%	7%
	7			8%	7%
	8			6%	7%
	9			4%	7%
	10			2%	6%
	11				6%
	12				6%
	13				6%
	14				5%
	15				5%
	16				5%
	17				4%
	18				3%
	19				2%
	20				1%

In all cases, the aggregate loss distribution is presumed to be lognormal. Payment patterns are at fixed percentages for all scenarios.

The assumptions are illustrative, not based on any specific source. In the author's opinion, none of the subject business is assumed to be unusually profitable.

1.5.3 The Reinsurance Contracts

Quota-Share Contracts:

C1: With aggregate limit 35% over expected losses.

C2: With aggregate limit 10% over expected losses.

C3: With "corridor" (losses not covered) from 5% to 15% over expected losses and aggregate limit 35% over expected losses.

Table 2
Quota-Share Contracts

Contract		Model M1	M2	M3	M4
C1	Ceded Premium	$100	$100	$100	$97
	Ceding Commission	30%	30%	30%	30%
	Loss Ratio at Limit	103%	104%	108%	118%
C2	Ceded Premium	$100	$100	$97	$92
	Ceding Commission	30%	30%	30%	30%
	Loss Ratio at Limit	78%	79%	83%	93%
C3	Ceded Premium	$100	$99	$97	$94
	Ceding Commission	30%	30%	30%	30%
	Loss Ratio at Limit	103%	104%	108%	118%
	Loss Ratio at Corridor Bottom	73%	74%	78%	88%
	Loss Ratio at Corridor Top	83%	84%	88%	98%

Note that the ceding commission rate has been set equal to the expense ratio on the subject business. Ceded premiums have been reduced from $100 proportional to the reduction in expected losses from limits and corridors.

Structured Aggregate Excess of Loss Contracts:

C4: Aggregate retention and limit;
Attaches within expected losses;
Upfront premium plus additional premiums as a percentage of ceded losses;
Fixed margin is deducted from upfront premium;
Refundable experience account accrues interest at 4%.

C5: Same as C4, plus another layer of additional premiums on subject losses extending beyond the policy limit.

Table 3
Structured Aggregate Excess of Loss Contracts

Contract		Model M3	Model M4
C4	Upfront Premium	$9.00	$5.50
	Margin	$3.00	$4.00
	Retention	63.0%	76.0%
	Loss Ratio at Limit	98.0%	136.0%
	A.P.Rate	59.0%	47.5%
	AP Attachment L/R	73.0%	83.0%
	AP Exhaustion L/R	98.0%	136.0%
C5	2nd A.P.Rate	12.5%	12.5%
	2nd AP Attachment L/R	93.0%	126.0%
	2nd AP Exhaustion L/R	113.0%	146.0%

These contracts have no ceding commission.

1.5.4 Risk Transfer Measures Applied to Subject Business

Before applying the risk transfer measures to the reinsurance contracts, it is interesting to first apply these measures to the subject business to be ceded (excluding ***PRT***, which is not defined in this case):

Table 4
Summary of Risk Transfer Measures
Applied to Subject Business

	Model			
	M1	**M2**	**M3**	**M4**
Loss Probability	**14.15%**	**24.91%**	**28.97%**	**25.50%**
VaR-90	**-0.73%**	**-4.35%**	**-10.85%**	**-19.13%**
TVaR-90	**-2.46%**	**-7.98%**	**-19.54%**	**-38.73%**
ERD	**-0.26%**	**-1.09%**	**-2.85%**	**-5.13%**

The difficulties with the absolute risk transfer measures can be anticipated. All measures produce values well below any likely threshold for M1. 10% volatility without unusual profitability seems like a level of risk that should "pass", but the 10/10 rule and **TVaR-90** fail for M2 as well, while the **ERD** passes only marginally at the low end of the range.

1.5.5 Risk Transfer Measures Applied to Quota-Share Contracts

We next apply the various measures to the three quota-share reinsurance contracts.

Table 5
Summary of Risk Transfer Measures
Quota-Share Contracts

		Model			
Contract		M1	M2	M3	M4
C1	Loss Probability	13.83%	24.78%	29.27%	29.19%
	PRT-MSAD	100.00%	100.00%	94.85%	63.86%
	VaR-90	-0.71%	-4.17%	-10.88%	-17.77%
	TVaR-90	-2.43%	-7.94%	-17.78%	-21.60%
	ERD	-0.26%	-1.06%	-2.65%	-3.82%
C2	Loss Probability	13.83%	24.78%	34.86%	34.27%
	PRT-MSAD	98.98%	78.65%	51.44%	31.72%
	VaR-90	-0.70%	-4.35%	-6.49%	-7.24%
	TVaR-90	-2.33%	-5.40%	-7.48%	-9.95%
	ERD	-0.25%	-0.83%	-1.77%	-1.91%
C3	Loss Probability	13.83%	27.69%	34.86%	25.12%
	PRT-MSAD	67.99%	52.21%	62.16%	48.82%
	VaR-90	-0.68%	-1.53%	-5.65%	-14.22%
	TVaR-90	-1.26%	-2.38%	-12.60%	-17.72%
	ERD	-0.14%	-0.43%	-1.66%	-2.88%

The contract C1 aggregate limit 35% over the mean has no discernable impact when applied to the lower volatility M1 and M2 models. As the volatility increases with M3 and M4, the risk limiting impact of the aggregate limit increases. This effect can be seen as the percentage of risk transferred decreases to 95% for M3 and down to 64% for the volatile M4 model.

The C1 contract applied to M1 fails the risk transfer test for all of the absolute risk measures, even though substantially all the risk is transferred. For M2, most still fail or marginally pass. As the underlying business gets riskier in the M3 and M4 models, results on these risk transfer tests improve significantly, even as the aggregate limit becomes less remote and has

more risk limiting impact. The tests based on absolute risk measures are more sensitive to the level of risk in the underlying business than to the degree of risk transfer.

The same pattern persists as we move to more significant risk limiting features. In each case, the risk limiting impact of the features becomes more significant when applied to the higher volatility cash flows, as is reflected in the declining ***PRT***. In each case, the absolute risk measures increase due to the increased underlying risk, even though a smaller percentage of that risk is being transferred.

1.5.6 Risk Transfer Measures Applied to Structured Aggregate Excess Contracts

Next, consider the application of the highly structured reinsurance contracts C4 and C5 to the riskier cash flows of models M3 and M4.

Summary of Risk Transfer Measures
Structured Aggregate Excess Contracts

		Model	
Contract		**M3**	**M4**
C4	**Loss Probability**	**24.80%**	**21.16%**
	PRT-MSAD	**22.89%**	**18.35%**
	VaR-90	**-10.51%**	**-10.91%**
	TVaR-90	**-15.76%**	**-21.00%**
	ERD	**-2.53%**	**-3.09%**
C5	**Loss Probability**	**24.80%**	**21.16%**
	PRT-MSAD	**19.36%**	**13.19%**
	VaR-90	**-10.74%**	**-10.73%**
	TVaR-90	**-11.56%**	**-11.91%**
	ERD	**-2.17%**	**-1.94%**

While risk transfer measures based on absolute risk levels may often "fail" a contract which transfers nearly all the risk when it is applied to relatively stable business, the effect is just the

opposite when applied to higher volatility business. In these cases, contracts with features that eliminate most of the risk can still pass.

In the case of C4, only 23% and 18% of the risk is transferred for M3 and M4, respectively. Yet the 10/10 test is a marginal pass and the other tests would also appear to pass at likely critical values.

Even though less than 25% of the risk is transferred, the C4 contracts are fairly risky for the reinsurer, especially relative to their small margins. The accounting distortion is that the losses accounted for as ceded are oversized relative to the risk absorbed by the reinsurer.

The C4 contract leaves the reinsurer with substantial tail risk, which is addressed in C5. Another layer of additional premium attaches just above the 90th percentile and extends beyond the policy limit, protecting the reinsurer from the acceleration risk caused by worsening loss ratios beyond the policy limit. The technique succeeds in further risk reduction, now bringing the ***PRT***'s to 19% and 13%. Yet the 10/10 rule is unaffected (as intended in the design of the feature). The more sophisticated **TVaR** and **ERD** tests respond to the additional risk reduction, with the more tail-oriented **TVaR** showing the greater effect. Despite the additional risk limitations, the **ERD** still produces a passing score and the **TVaR** may as well, depending on choice of critical value.

1.5.7 Conclusion

In conclusion, the ***PRT*** test appears to logically and consistently identify the impact of structural features that limit risk transfer. The measures based on absolute standards invariably underestimate risk transfer for more stable subject business and overestimate risk transfer for more volatile subject business.

1.6 Examples Using *PRT* with Various Risk Measures and Co-Measures

The following tables present the results of ***PRT***, applied to the same models and contracts as the previous section, with one exception. We have removed the aggregate limit from the 3rd contract (the corridor). We use the following risk measures (with their corresponding co-measures):

- ***MSAD***
- ***EAD***
- ***TVaR-90***
- ***TVaR-95***
- ***TVaR-98***

The results are presented without a great deal of additional comment. With each risk measure, the pattern of ***PRT***'s as the risk models and contracts change conform to a reasonable pattern of decreasing risk transfer as the risk-limiting provisions become more significant.

The results are not identical, however. The measures respond to the "heart" and the "tail" of the distribution to different degrees, consistent with their design.

Table 7
PRT's -- Comparison of Risk Measures
Quota-Share Contracts

Contract		Model M1	M2	M3	M4
C1	MSAD	100.00%	100.00%	94.85%	63.86%
	EAD	100.00%	100.00%	97.92%	78.14%
	TVaR-90	100.00%	100.00%	95.84%	61.56%
	TVaR95	100.00%	100.00%	93.13%	52.25%
	TVaR98	100.00%	100.00%	85.94%	43.72%
C2	MSAD	98.98%	78.65%	51.44%	31.72%
	EAD	99.63%	87.82%	64.46%	43.90%
	TVaR-90	99.47%	76.45%	44.00%	29.53%
	TVaR95	99.16%	64.80%	38.16%	25.37%
	TVaR98	98.31%	54.41%	32.65%	21.27%
C3	MSAD	67.51%	40.25%	61.89%	83.62%
	EAD	79.97%	46.61%	55.05%	71.67%
	TVaR-90	71.55%	34.37%	64.16%	84.14%
	TVaR95	58.01%	34.00%	70.58%	87.42%
	TVaR98	46.51%	41.60%	75.93%	90.01%

Table 8
PRT's -- Comparison of Risk Measures
Structured Aggregate Excess Contracts

Contract		Model M3	Model M4
C4	MSAD	22.89%	18.35%
	EAD	23.84%	17.93%
	TVaR-90	23.23%	19.27%
	TVaR95	21.76%	18.85%
	TVaR98	19.53%	17.50%
C5	MSAD	19.36%	13.19%
	EAD	21.71%	14.52%
	TVaR-90	18.97%	13.31%
	TVaR95	15.85%	11.35%
	TVaR98	12.82%	10.41%

Some observations:

- In most cases ***MSAD*** produces results similar to ***TVaR-90***.
- Aggregate limits affect only the tail of the distribution, and are most penalized by the more tail-oriented ***TVaR*** measures, for example the low aggregate limit of the C2 contract applied to the moderately high risk M3 model.
- The combination of low corridor and no limit (C3), when applied to high risk models M3 and M4, decreases risk more in the heart of the distribution than the tail. In this case, the least tail-oriented measure, ***EAD***, indicates the greatest reduction in risk transfer.
- The first highly structured contract, C4, dramatically reduces risk in the heart and the tail of the distribution and all measures are similar.
- The second highly structured contract, C5, has an additional feature that mitigates the tail risk. Especially for risk model M4, risk transfer is significantly lowered. The effect of the tail-protecting feature is the smallest for the EAD and the largest for the more tail-oriented measures.

In conclusion, ***PRT*** is demonstrated to work acceptably well with a variety of risk measures. Assuming that it is desirable to have a single measure to be used universally, the author's preference continues to be for ***MSAD***, which works consistently and appears to strike the best compromise between responsiveness to the whole downside of the distribution and emphasis on the significance of the tail.

Section II – Accounting for Partial Risk Transfer Reinsurance

2.1 The Case for Continuous Accounting

The problem addressed so far is to find the best possible solution given the significant accounting constraint that there are two types of accounting available – one that is appropriate for 100% risk transfer and another for 0% risk transfer – and that our only option is to choose one or the other. The difference between these approaches can sometimes be very large – and for large enough contracts it can be material to the company's financial statements.

If the difference between the two accounting treatments is material, then it is likely that half that difference is material as well. Regardless of which accounting treatment is used, the accounting for a contract with 50% risk transfer will be materially inaccurate, one way or another. The author's suggestion of a critical value of 50% to define adequate risk transfer is simply to cut the worst case inaccuracy to the lowest possible number.

Using the 50% critical value, there could continue to be motivation to design 51% risk transfer contracts to take advantage of the 100% risk transfer accounting. 49% risk transfer contracts are no less problematic. The cedant may get no credit in its financial statements or solvency tests for a significant reduction in risk. And a reinsurer that assumes a 49% risk transfer contract that is ineligible for reinsurance accounting will be assuming significant risk while its financial statements reflect that it has assumed none.[9]

Another significant problem is the point of discontinuity itself. If the difference in accounting treatment has a large impact, and the estimated $\boldsymbol{PRT}$ is close to the critical value, then a large material difference will turn on a decision requiring a precision of estimation that simply doesn't exist.

[9] This last point illustrates that there is no such thing as a "safe", "conservative" choice for a critical value. Whenever deposit accounting is conservative for the cedant, it is aggressive for the reinsurer.

Thus, the binary choice between reinsurance accounting and deposit accounting may not be an adequate solution. A continuous solution would provide more appropriate accounting for partial risk transfer contracts. The availability of ***PRT*** can provide a basis for such a continuous accounting solution.

2.2 Goals of Partial Risk Transfer Accounting

The author has considered the following two goals of appropriate accounting for partial risk transfer:

- Income statements and balance sheets that are undistorted in total, i.e., accurate total income and equity; and
- Proper characterization of ceded premiums and ceded losses.

2.3 Bifurcation to Achieve Continuous Accounting

2.3.1 Proportional Bifurcation

The simplest approach, which would require no new development of basic accounting rules, is to apply a weighted average of the two accounting procedures already available, i.e. proportional bifurcation. The approach would be to simply divide all 100% values proportional to ***PRT*** and ***1-PRT***, with the amounts proportional to ***PRT*** accounted for as reinsurance and the amounts proportional to ***1-PRT*** accounted for using deposit accounting. For the deposit accounting, the "interest method," which corresponds best to zero risk transfer, would be most appropriate.

The First Objective -- Income and Equity:

As discussed much earlier, the income and equity effects of reinsurance are gains related to the cession of losses and costs related to the reinsurer's margin. Since there are no gains from ceding losses under deposit accounting, under proportional bifurcation using ***PRT***, the initial gain from ceding losses would be reduced proportional to the reduction in risk transfer, exactly as intended. As for the cost of the reinsurer's margin, this is expensed up front under reinsurance accounting but implicitly expensed over the life of the cash flows under deposit accounting. Thus proportional bifurcation will cause a deferral of a portion of this cost. This is not necessarily our intent, and could be remedied with a slightly more complex solution. However, to the extent that this is considered an imperfection, it is not a serious one, and may not warrant the additional complexity.

The Second Objective -- Losses and Premiums:
Net losses under proportional bifurcation will be in proportion to the percentage of the risk retained, exactly as intended.

Net premiums resulting from the proportional subdivision of premiums will not be perfectly reflective of net underwriting risk retained, so the second objective will not be perfectly satisfied for net and ceded premium.

Two imperfections related to the proportional subdivision of premium: The first imperfection is that the reinsurer's margin would be expected to be reduced if the risk is reduced. It would probably be preferable to allocate the margin entirely to reinsurance accounting, rather than subdivided. The second imperfection, related to over-funding, is in the opposite direction. As will be discussed in a subsequent section, income, equity, and ceded losses are not distorted by over-funding. However, if the reinsurance is over-funded with a refund provision, then the premium allocated to reinsurance accounting will be overstated to some degree.

In the author's view, none of the imperfections noted is likely to be significant, and simple proportional bifurcation will provide a major improvement in accounting accuracy compared with current practice. A modestly more complex solution can be devised for the income issue and the first premium issue discussed above, although the second premium issue is more

difficult. In any case none of the imperfections are likely to be large enough to warrant the additional complexity.

2.3.2 What Contracts Should Be Bifurcated?

Bifurcation would increase accounting workloads and complexity and it makes sense to limit its application. Many reinsurance contracts have structural features that have modest risk-limiting effects. At the other extreme, there may be some contracts determined to have minimal risk transfer. In order to avoid unnecessary bifurcation, the author suggests that contracts with $PRT > 80\%$ or $PRT < 20\%$ be accounted for with reinsurance accounting or deposit accounting, respectively, with bifurcation limited to $20\% < PRT < 80\%$.

Such a threshold would also reduce the need for unnecessary testing. It will be fairly obvious in some cases that structural provisions will not reduce risk transfer by more than the threshold value, and minimal testing may be required.

2.3.3 Should Risk Transfer Be Reevaluated?

If PRT were to become an explicit factor in reinsurance accounting, the PRT would presumably be evaluated at the inception of the reinsurance contract and that value would become fixed for accounting purposes at the inception of the contract. The issue of possible reevaluation of the PRT would not be retrospective from inception, but only prospectively relating to remaining loss reserves. To the extent the PRT changed, that change would affect only the remaining loss reserves, not any previously accounted for amounts, such as premiums or loss payments.

In the author's view, this idea is cumbersome and impractical and would appear to be an idea to be avoided. However, the discussion is included for the theoretical completion of the concept.

The amount of remaining risk transferred for ceded loss reserves will change as the contract progresses over time. The change in the remaining risk transferred can be illustrated with a simple aggregate limit example. Suppose that an aggregate limit set above the expected loss ratio is originally estimated to have a 40% risk-limiting effect (i.e. 60% ***PRT***). Two years later, the ultimate losses are known with much more accuracy and have developed below the original expected losses. The aggregate limit now appears quite remote and 95% of the remaining risk is transferred. Or conversely, losses have developed much worse than the original expected losses and ultimate losses are now estimated to be at the aggregate limit, leaving no more coverage available. To the extent that there are still ceded reserves, almost none of the risk related to the remaining reserves is transferred. While these situations may be realistic, it would be hard to imagine that the increase in accounting accuracy would warrant reevaluating risk transfer on all contracts.

But perhaps it should be considered in a few special cases. An obvious candidate is a multi-line contract combining long and short tail business. For example, assume that such a contract, mixing property and casualty but not readily bifurcated in the more traditional sense, is estimated at its inception to transfer 50% of the risk and is accounted for with a 50% proportional bifurcation. Let us further assume that almost all the risk comes from exposure to property catastrophes, and that at the end of the year there has been no such catastrophe. There may be a significant cession of casualty reserves at a discount, but little or no risk transfer remaining. Conversely, if property catastrophe losses have occurred, a much larger degree of risk may be ceded on the remaining casualty reserves.

2.4 Comments on Related Topics

2.4.1 Over-Funding

A common technique for reducing risk to the reinsurer is over-funding, i.e., charging a conservative premium with refund provisions. The refund may be based on an "experience account" which includes interest credited on ceded funds. This technique may allow a reinsurer

and cedant to come to terms without resolving differences of opinion on likely losses, or may simply be used to lower the risk premium charged.

Over-funding may be accomplished by charging a large upfront premium, through a contingent additional premium feature, or a combination of the two. To the extent that contingent additional premiums are charged, the outgoing cash flows will be included in the calculation of ***PRT*** and the value of ***PRT*** will be reduced.

To the extent that over-funding is accomplished through an increase in upfront premiums, it will probably have no effect on ***PRT***, as only downside risks are measured, and premium refunds usually have no impact. This may appear counterintuitive, as over-funding clearly reduces the risk to the reinsurer.

Nevertheless, contingent refunds cannot cause a future loss for the cedant. To the extent that the risk related to ceded losses is covered by the reinsurance, it is appropriate to cede the losses and their associated risk margin, i.e. to apply reinsurance accounting. Whether the risk related to the ceded losses is covered from funds provided by the cedant or risk taken by the reinsurer is immaterial. As long as the cedant has expensed the premiums ceded, there is no increased risk of inadequacy in the financial statement values.

Under current accounting, the cedant records an asset for future refunds only to the extent that the current ceded loss estimate indicates that a refund will be due *without including future investment income credited to an experience account.* This asset, when applicable, prevents over-funding from causing a deferral of income. The exclusion of future investment income is also necessary – including it in the calculation of the asset would have a similar effect to discounting the loss reserve while retaining the risk.

In conclusion, premium refunds are not important when determining ***PRT*** since they do not affect downside risk. When reinsurance accounting is applied to reinsurance that includes over-

funding, the net effects on balance sheets, income statements, and ceded losses are not significantly distorted.[10]

2.4.2 Underwriting Risk and Timing Risk

Both GAAP and SAP require the separate consideration of whether underwriting and timing risk have been transferred, as well as whether the overall degree of risk transfer is adequate. The approach herein is focused only on the overall risk. In the author's view, the distinction between underwriting and timing risk is often artificial. If a continuous approach to risk transfer accounting were adopted so that the degree of risk transfer were specifically reflected, perhaps the distinctions between underwriting and timing risk would be unnecessary.

2.4.3 Accounting for Retroactive Reinsurance

There are substantial restrictions in GAAP and Statutory accounting when the liabilities ceded are related to losses incurred in the past, e.g., loss portfolio transfers ("LPT's"). In fact, GAAP essentially applies deposit accounting to all retroactive reinsurance, as if no risk transfer is possible. This punitive accounting undoubtedly has its historical roots in past abuses, but otherwise appears to have no sound basis.

LPT's are often legitimate risk transfer motivated reinsurance contracts. There are any number of valid motivations, such as moving risky liabilities to better diversified and capitalized companies. LPT's are still done despite punitive accounting. But it would be hard to imagine that the accounting is not suppressing the market for legitimate retroactive reinsurance.

As we have demonstrated in the examples, FAS 113 is not effective in preventing financial engineering for prospective reinsurance, nor would it be effective for retroactive reinsurance if the present restrictions were eliminated. The improved accounting recommended herein would

[10] Overall equity and income will be undistorted, as will ceded losses and loss reserves. Ceded premiums may be overstated to some degree. As with other imperfections on the premium side, this problem may not be significant enough to warrant a more complex procedure.

effectively prevent the type of abuses that were committed long ago, and the punitive accounting, which is itself highly inaccurate, could be eliminated.

2.4.4 Policing

Punitive accounting for retroactive reinsurance under GAAP might be considered an example of policing by accounting – the idea is not to account accurately, but to prevent abuse. Regulators have more direct police powers. Insurance executives may have to increasingly describe the intent of reinsurance transactions. While improved disclosure by financial executives is beneficial, the author is not entirely comfortable with police powers to regulate intent.

With more accurate accounting, regulation of intent would be less necessary. Bad behavior will still be possible; policing will still be needed. But with better accounting rules, policing can be about following the rules.

References

[1] FAS 113

[2] SSAP No. 62

[3] AICPA Statement of Position 98-7

[4] SSAP No. 75

[5] CAS Valuation, Finance, and Investments Committee, "Accounting Rule Guidance Statement of FAS No. 113 – Considerations in Risk Transfer Testing", Casualty Actuarial Society Forum, Volume: Fall, 2002, pp. 305-338

[6] AAA Index Securitization Task Force, "Evaluating the Effectiveness of Index-Based Insurance Derivatives in Hedging Property/Casualty Insurance Transactions", October 4, 1999

[7] Kreps, Rodney, "Riskiness Leverage Models", to be published in the Proceedings of the Casualty Actuarial Society.

[8] Wang, Shaun, "A Risk Measure that Goes Beyond Coherence," Institute of Insurance and Pension Research, Research Report No. 18, University of Waterloo.

[9] Kreps, Rodney and Oakley E. Van Slyke, "When Is Risk "Significant" in a Reinsurance Contract," an unpublished paper.

A Multivariate Bayesian Claim Count Development Model With Closed Form Posterior and Predictive Distributions

Stephen J. Mildenhall

Abstract

We present a rich, yet tractable, multivariate Bayesian model of claim count development. The model combines two conjugate families: the gamma-Poisson distribution for ultimate claim counts and the Dirichlet-multinomial distribution for emergence. We compute closed form expressions for all distributions of actuarial interest, including the posterior distribution of parameters and the predictive multivariate distribution of future counts given observed counts to date and for each of these distributions give a closed form expression for the moments. A new feature of the model is its explicit sensitivity to ultimate claim count variability and the uncertainty surrounding claim count emergence. Depending on the value of these parameters, the posterior mean can equal the Bornhuetter-Ferguson or chain-ladder reserve. Thus the model provides a continuum of models interpolating between these common methods. We give an example to illustrate use of the model.

JEL Classification: G - Financial Economics; G220 - Insurance; Insurance Companies

Keywords: Loss Development, Chain-Ladder Method, Bornhuetter-Ferguson Method, Dirichlet-multinomial, Poisson-gamma

1 INTRODUCTION

We present a Bayesian model of claim count development. The model is rich enough to provide a realistic model for the practitioner but at the same time it is mathematically tractable and we give explicit equations for the posterior and predictive distributions. The predictive distribution is an example of a generalized power series distribution and a generalized hypergeometric distribution. The

method in the paper will be of interest to practicing actuaries because it is easy to implement and it provides explicit posterior distributions for unreported claims, and hence Bayesian means and confidence intervals, and a rationale for choosing between existing reserving methods. The model is theoretically interesting because the posterior mean generalizes three common reserving methods (the peg, the Bornhuetter-Ferguson and the chain-ladder) in an intuitive and insightful manner.

Actuaries today are asked to provide a distribution of potential outcomes or a confidence interval around the point estimates they have traditionally supplied. The push towards greater quantification of uncertainty is particularly marked in the property and casualty loss reserving practice. Understanding reserve uncertainty and linking the pricing actuary's prior estimate of ultimate losses to the reserving actuary's posterior estimates is therefore becoming more and more important.

These recent demands on the profession have played up some shortcomings of the traditional chain-ladder method of determining loss reserves. The chain-ladder method is simple to apply and easy to explain, and is the *de facto* standard reserve method. Mack's 1993 paper [15] showing how to compute the standard error of chain ladder reserves was an important enhancement to the method. However, the chain-ladder is still not well suited to providing explicit posterior distributions, nor does it provide diagnostic information to assess model fit. The latter point is a severe weakness in practice. There is no *one* chain-ladder method; the technique can be applied to a variety of different loss development triangles in slightly different ways. (Academic discussions usually assume link ratios are stable over time—something rarely seen in practice—and use the weighted average of all years link ratios.) When the various chain-ladder related estimates do not agree there is no statistical guidance on which method to prefer. The shortcomings of the chain-ladder have been discussed in the literature. Mack [14] identifies the stochastic assumptions which underlie the chain-ladder method. Venter [25] discusses the assumptions required for the chain-ladder estimates to produce least-squares optimal reserve estimates, and discusses some alternative methods when

the conditions are not met. Renshaw and Verrall [22] describe a statistical procedure which is exactly equivalent to the chain-ladder in almost all circumstances. We will discuss their model more in Section 6.

In order to address these shortcomings, and respond to the demand for more precise quantification of uncertainty, both practicing actuaries and academics have explored alternative models. Zehnwirth [30] and Zehnwirth and Barnett [31] construct general linear models of reserve development based on log-incremental data. Kunkler [12] uses a mixed model to include zero claims in a log-incremental model. England and Verrall [7] and Wright [29] discuss generalized linear models, the latter taking an operation time point of view. Norberg [19] models the claims process as a non-homogeneous marked Poisson process. There has been considerable interest in Bayesian models of development. Reserving involves the periodic update of estimates based on gradually emerging information—a naturally Bayesian situation. Bayesian methods have been explored by Robbin [23], de Alba [5], Dellaportas and Ntzoufras [20], Renshaw and Verrall [22], and Verrall [26, 27], amongst others. Stephens et al. [24] use a survival time approach to modeling claim closure in a Bayesian framework. As Robbin points out, the mathematics of Bayesian models often becomes intractable. One advantage of this paper's model is the closed mathematical form of all the distributions of interest. For the less tractable models the WinBUGs MCMC system has been applied. See Verrall [27] for a very detailed explanation of how to do this.

Despite all of these advances, no model has come close to challenging the chain ladder method. In part this reflects the difficulties a new method faces before it becomes accepted practice. It also reflects the technical complexity of some of the alternative models. Practicing actuaries can be uncomfortable with the assumptions[1] and the number of parameters. The chain-ladder method has one parameter for each development period: the link-ratios and the tail factor. Regression models may produce a model with fewer parameters, but the model itself is often selected from a very large number of potential models. This can lead to generalization error where a particular model can over-fit artifacts in a small data

[1] A difficulty with *explicit* assumptions is the disquiet they can cause!

set. The advantages of "simple" models are discussed in Balasubramanian [2] and Domingos [6].

There is, therefore, a need for a simple statistical model of loss development to augment and enhance the chain-ladder method. A new model should have a similar number of parameters to the chain-ladder, should be fit to the data using a statistical technique such as maximum likelihood, should be able to incorporate prior information from the pricing department, and should be easy to update with observed loss information as it becomes available. We will present such a model for claim count development in this paper. The model is introduced in Section 2. The explicit form of the marginal distributions of claims reported in each period is proved in Section 3. Sections 4 and 5 prove results about the conditional and predictive distributions. Section 6 discusses where our model fits within the continuum of reserving models, from the "book plan" peg method through the chain-ladder method. Section 7 discusses parameter estimation. Section 8 applies the model to a specific triangle. Finally, Section 9 will discuss extending the model to loss development, rather than just claim count development.

This paper focuses on the theoretical development of a new claim count model. However, I want to stress that this model is easy to use in practice and that it provides useful and powerful reserving diagnostics.

Notation

The following notational convention will be use extensively in the paper. For any n-tuple $x_1, \ldots, x_n$ define

$$x(t) = \sum_{i=1}^{t} x_i, \quad x'(t) = \sum_{i=t+1}^{n} x_i$$

and let $x := x(n)$. Thus $x = x(n) = x(t) + x'(t)$ for all $t = 1, \ldots, n$. This notation will apply to B, b, π, and v. It will be re-iterated before it is used.

The letters p and $q := 1 - p$ will be used as parameters of a gamma distribution and will never have subscripts.

2 THE GPDM BAYESIAN CLAIM COUNT MODEL

This section introduces the new gamma-Poisson Dirichlet-multinomial (GPDM) claim count model. The GPDM is a combination of a gamma-Poisson random variable to model total ultimate claims and a Dirichlet-multinomial multivariate distribution to model incremental claims by report period. For a particular accident year, let B_i be the incremental number of claims reported in period $i = 1, \ldots, n$. We assume that nth report is ultimate and will not model further claim emergence. Let

$$B(n) = B_1 + \cdots + B_n \tag{1}$$

denote the ultimate number of claims.

The GPDM is defined as a combination of two conjugate models. The ultimate number of claims $B(n)$ conditional on $\Lambda = \lambda$ has a Poisson distribution with mean λ. Λ, the prior ultimate claim count, has a gamma distribution. Conditional on $B(n)$ and parameters $\Pi_1 = \pi_1, \ldots, \Pi_{n-1} = \pi_{n-1}$, $\pi_n = 1 - \sum_{i-1}^{n-1} \pi_i$, the claim emergence vector $(B_1, \ldots, B_n)$ has a multivariate multinomial distribution with parameters $B(n), \pi_1, \ldots, \pi_n$. Here π_t is the expected proportion of claims reported in period t and $B(n)$ is the number of ultimate claims. $\Pi_1, \ldots, \Pi_{n-1}$ have a Dirichlet prior distribution. The full vector of parameters is $\Theta = (\Lambda, \Pi_1, \ldots, \Pi_{n-1})$. Conditional on

$$\Theta = \theta := (\lambda, \pi_1, \ldots, \pi_{n-1}) \qquad \pi_n = 1 - \sum_{i=1}^{n-1} \pi_i. \tag{2}$$

the GPDM probability density is

$$\Pr(B_1, \ldots, B_n \mid \Theta = \theta) = \frac{e^{-\lambda} \lambda^{b(n)}}{b_1! \ldots b_n!} \pi_1^{b_1} \ldots \pi_n^{b_n} \tag{3}$$

where $b(n) = \sum_{i=1}^{n} b_i$ and the two $b(n)!$ terms have cancelled.

The prior densities for the parameter vector $\Theta = (\Lambda, \Pi_1, \ldots, \Pi_{n-1})$ are

$$\Lambda \sim \text{Gamma}(r, p/q), \quad q = 1 - p, \tag{4}$$

$$\Pr(\Lambda = \lambda) = \frac{p^r}{q^r \Gamma(r)} \lambda^{r-1} e^{-\lambda p/q} \tag{5}$$

and

$$(\Pi_1, \ldots, \Pi_{n-1}) \sim \text{Dirichlet}(v_1, \ldots, v_n), \tag{6}$$

$$\Pr(\Pi_1 = \pi_1, \ldots, \Pi_{n-1} = \pi_{n-1}) = \frac{\Gamma(v_1 + \cdots + v_n)}{\Gamma(v_1) \cdots \Gamma(v_n)} \pi_1^{v_1 - 1} \cdots \pi_n^{v_n - 1}. \tag{7}$$

Λ and the Π_i are *a priori* independent.

The form of the gamma distribution in Eqn. (4) is chosen so that the negative binomial predictive distribution for $B(n)$ has density

$$\Pr(B(n) = b(n)) = \binom{r + b(n) - 1}{b(n)} p^r q^{b(n)}. \tag{8}$$

Thus $\text{E}(B(n)) = rq/p$ and $\text{Var}(B(n)) = rq/p^2$. If $\text{E}(B(n)) = m$ then $p = r/(r+m)$, $q = m/(r+m) = 1/(1+m/r)$ and $\text{Var}(B(n)) = m(1+m/r)$. The coefficient of variation of the gamma distribution is $1/\sqrt{r}$. The expression $1/r$ is sometimes called the contagion, see Mildenhall [17, Section 2.2].

Compared to traditional methods of reserving the GPDM includes two new parameters: r which controls the variability of ultimate claim counts and the extra Dirichlet paramter which controls the variability of claim emergence. The Bornhuetter-Ferguson method of reserving, by contrast, assumes a prior estimate of the ultimate number of claims but no measure of its variability. The chain-ladder does not assume a prior estimate of ultimate claims but gives full credibility to observed claim emergence, corresponding to a high degree of confidence in estimates of Π_i. These two extra parameters determine the behavior of the GPDM model.

Pricing actuaries often have prior estimates of expected frequency because the frequency-severity approach is a common pricing method. Thus reserving actuaries can usually obtain a prior mean for the number of ultimate claims expected

from a block of business. We want to be able to incorporate this information into our claim count model. Eqn. (3) assumes that the ultimate claim count, $B(n)$, has a negative binomial distribution. The r parameter for $B(n)$ is a measure of the inhomogeneity of insureds or of non-diversifiable parameter risk; it could be estimated based on line of business studies. The negative binomial has been suggested as a more flexible alternative to the Poisson distribution for modeling claim counts by many authors, including Klugman, Panjer and Willmot [10]. Also see the references in Johnson et al. [9].

The second part of Eqn. (3) is the multinomial with Dirichlet conjugate prior. Basic properties of Dirichlet-multinomial (DM) are given in Bernardo and Smith [3] and Johnson et al. [8, Section 35.13.1]. For more details on the Dirichlet see Kotz et al. [11]. The Dirichlet distribution has n free parameters (compared to only $n-1$ free π_i because of the condition $\sum_i \pi_i = 1$), and the extra parameter controls uncertainty in the proportions. When $n = 2$ the Dirichlet becomes a beta distribution.

The DM distribution with parameters $(b(n); v_1, \ldots, v_n)$ has predictive probability density function

$$\Pr(B_1 = b_1, \ldots, B_n = b_n) = \frac{b(n)!}{b_1! \cdots b_n!} \frac{\Gamma(\sum v_i)}{\Gamma(b(n) + \sum v_i)} \prod_{i=1}^{n} \frac{\Gamma(b_i + v_i)}{\Gamma(v_i)} \tag{9}$$

where $b(n) = \sum_i b_i$. We can write Eqn. (9) more succinctly using the Pochhammer symbol $(r)_k$. For a real r and non-negative integer k define

$$(r)_k := r(r+1)\cdots(r+k-1) = \frac{\Gamma(r+k)}{\Gamma(r)}. \tag{10}$$

Then Eqn. (9) becomes

$$\Pr(B_1 = b_1, \ldots, B_n = b_n) = \frac{b(n)!}{b_1! \cdots b_n!} \frac{1}{(\sum v_i)_{b(n)}} \prod_{i=1}^{n} (v_i)_{b_i}. \tag{11}$$

We will use the Pochhammer symbol extensively.

The marginal distributions of the DM are beta-binomial mixtures. Let $v = \sum_i v_i$. Then

$$\mathrm{E}(B_i) = mv_i/v, \tag{12}$$

$$\mathrm{Var}(B_i) = \frac{b(n)+v}{1+v}\left(\frac{b(n)v_i(v-v_i)}{v^2}\right) \tag{13}$$

$$\mathrm{Cov}(B_i, B_j) = -\frac{b(n)+v}{1+v}\frac{b(n)v_iv_j}{v^2} \tag{14}$$

$$\mathrm{Corr}(B_i, B_j) = -\sqrt{\frac{v_iv_j}{(v-v_i)(v-v_j)}}. \tag{15}$$

The marginal and conditional distributions of a DM are also DMs. See Johnson et al. [8, Section 35.13.1] for these facts.

The next lemma, and its obvious generalizations, follows from the properties of the Dirichlet and multinomial distributions. We will use it several times in various guises.

Lemma 1 *Let $B_1, \ldots, B_n \mid \Theta$ be a GPDM. Then $B_1 + B_2, B_3, \ldots, B_n \mid \Theta' = (\lambda, \pi_1 + \pi_2, \pi_3, \ldots, \pi_n)$ is also a GPDM.*

Proof: This follows from [8, Chapter 35 Section 13.1]. ■

We end this section by computing the predictive distribution of $B_1, \ldots, B_n$ given no observations. Let $v = \sum v_i$. Then

Proposition 1 *Let $B_1, \ldots, B_n \mid \Theta$ be a GPDM. Then*

$$\Pr(B_1 = b_1 \ldots, B_n = b_n) = \frac{b(n)!\Gamma(v)}{\Gamma(b(n)+v)}\prod_{i=1}^{n}\frac{\Gamma(b_i+v_i)}{b_i!\Gamma(v_i)}\binom{r+b(n)-1}{b(n)}p^r q^{b(n)}. \tag{16}$$

Proof: We have

$$
\begin{aligned}
&\Pr(B_1 = b_1, \ldots, B_n = b_n) \\
&= \int \cdots \int \Pr(B_1, \ldots, B_n \mid \Theta) f(\lambda) f(\pi_1, \ldots, \pi_n) d\lambda d\pi_1 \ldots d\pi_{n-1} \\
&= \int \cdots \int \left(\frac{e^{-\lambda} \lambda^{b(n)}}{b(n)!} \right) \left(\frac{b(n)!}{b_1! \ldots b_n!} \pi_1^{b_1} \ldots \pi_n^{b_n} \right) \frac{p^r}{\Gamma(r) q^r} \lambda^{r-1} e^{-p\lambda/q} \\
&\quad \times \frac{\Gamma(v)}{\prod \Gamma(v_i)} \prod_{i=1}^{n} \pi_i^{v_i - 1} d\lambda d\pi_1 \ldots d\pi_{n-1} \\
&= \frac{b(n)!}{b_1! \ldots b_n!} \frac{\Gamma(v)}{\prod \Gamma(v_i)} \frac{p^r}{\Gamma(r) q^r b(n)!} \\
&\quad \times \int \cdots \int \left(\int e^{-\lambda(1+p/q)} \lambda^{b(n)+r-1} d\lambda \right) \prod_{i=1}^{n} \pi_i^{b_i + v_i - 1} d\pi_1 \ldots d\pi_{n-1} \\
&= \frac{b(n)! \Gamma(v)}{\Gamma(b(n) + v)} \prod_{i=1}^{n} \frac{\Gamma(b_i + v_i)}{b_i! \Gamma(v_i)} \binom{r + b(n) - 1}{b(n)} p^r q^{b(n)}
\end{aligned}
$$

since the inner integral with respect to λ equals $\Gamma(b(n) + r) q^{b(n)+r}$ and $1 + p/q = 1/q$. ■

Eqn. (16) can also be written more compactly as

$$
\Pr(B_1 = b_1, \ldots, B_n = b_n) = p^r q^b \frac{(r)_{b(n)}}{(v)_{b(n)}} \prod_{i=1}^{n} \frac{(v_i)_{b_i}}{b_i!}. \tag{17}
$$

3 MARGINAL DISTRIBUTIONS

The GPDM is a tractable distribution because it is possible to write down closed-form and easy-to-compute expressions for its conditional marginal distributions and its predictive distribution of future claims given observed claims to date. The marginals are necessary to compute likelihoods from whole or partial claim count development triangles. The predictive distributions provide a conditional distribution for ultimate claims given counts to date. We now prove these two important results, starting with marginal distributions in this section.

The marginal and conditional distributions of an GPDM are hypergeometric distribution and use the Gaussian hypergeometric functions ${}_2F_1(a, b; c; z)$. Mathematicians and actuaries today may not be as familiar with hypergeometric functions as their counter-parts would have been 50 or 100 years ago. Given this lack of familiarity expressions involving ${}_2F_1$ can be a little forbidding. It is important to remember that ${}_2F_1$ is no more mysterious than the other functions built-in to most calculators and spreadsheets. Indeed, it is very easy to program ${}_2F_1$ into a spreadsheet and use it like a built-in function. The properties of ${}_2F_1$ we use, together with pseudo-code to compute it, are given in Appendix A.

The next proposition computes the marginal distribution of $B_1, \ldots, B_t$ for $t < n$. Obviously an analogous result would hold for any subset of the B_i. Remember that $v = \sum_{i=1}^{n} v_i$, $v'(t) = \sum_{i=t+1}^{n} v_i$, $b(t) = \sum_{i=1}^{t} b_i$ and $\pi(t) = \sum_{i=1}^{t} \pi_i$.

Proposition 2 *Let $B_1, \ldots, B_n \mid \Theta$ have a GPDM distribution. If $t \leq n-1$ then the marginal distribution of $(B_1, \ldots, B_t \mid \Theta)$ is also GPDM with*

$$\Pr(B_1 = b_1, \ldots, B_t = b_t \mid \Theta = (\lambda, \pi_1, \ldots, \pi_{n-1})) \tag{18}$$

$$= \Pr(B_1 = b_1, \ldots, B_t = b_t \mid (\pi(t)\lambda, \frac{\pi_1}{\pi(t)}, \ldots, \frac{\pi_t}{\pi(t)})). \tag{19}$$

The predictive marginal of $(B_1, \ldots, B_t)$ is

$$\Pr(B_1 = b_1, \ldots, B_t = b_t) = p^r q^{b(t)} \frac{\Gamma(v)}{\Gamma(b(t)+v)} \frac{\Gamma(b(t)+r)}{\Gamma(r)} \prod_{i=1}^{i=t} \frac{\Gamma(b_i+v_i)}{\Gamma(v_i) b_i!}$$

$$\times {}_2F_1(v'(t), b(t)+r; b(t)+v; q) \tag{20}$$

$$= p^r q^{b(t)} \frac{(r)_{b(t)}}{(v)_{b(t)}} \prod_{i=1}^{i=t} \frac{(v_i)_{b_i}}{b_i!} {}_2F_1(v'(t), b(t)+r; b(t)+v; q).$$

Proof: Using Lemma 1 we can sum the unobserved variables $(B_{t+1}, \ldots, B_n)$ and, without loss of generality, assume that $t = n-1$.

Then $\Pr(B_1 = b_1, \ldots, B_t = b_t \mid \Theta)$

$$= \sum_{b_n \geq 0} \Pr(B_1 = b_1, \ldots, B_n = b_n \mid \Theta)$$

$$= \sum_{b_n \geq 0} \left(b(n)! \prod_{i=1}^{n} \frac{\pi_i^{b_i}}{b_i!} \right) \frac{e^{-\lambda} \lambda^{b(n)}}{b(n)!} = \left(\prod_{i-1}^{n-1} \frac{\pi_i^{b_i}}{b_i!} \right) \left(\sum_{b_n \geq 0} \frac{\lambda^{b_n} \pi_n^{b_n}}{b_n!} \right) e^{-\lambda} \lambda^{b(n-1)}$$

$$= \frac{\pi_1^{b_1} \ldots \pi_t^{b_t}}{b_1! \ldots b_t!} \lambda^{b(t)} e^{-\pi(t)\lambda}$$

$$= b(t)! \frac{(\pi_1/\pi(t))^{b_1} \ldots (\pi_t/\pi(t))^{b_t})}{b_1! \ldots b_t!} \frac{(\pi(t)\lambda)^{b(t)} e^{-\pi(t)\lambda}}{b(t)!}$$

$$= \Pr\big(B_1 = b_1, \ldots, B_t = b_t \mid (\pi(t)\lambda, \frac{\pi_1}{\pi(t)}, \ldots, \frac{\pi_t}{\pi(t)})\big).$$

Next, using Proposition 1 and remembering $t = n - 1$, we have $\Pr(B_1 = b_1, \ldots, B_t = b_t)$

$$= \sum_{b_n \geq 0} \Pr(B_1 = b_1, \ldots, B_t = b_t, B_n = b_n)$$

$$= \sum_{b_n \geq 0} \frac{b(n)! \Gamma(v)}{\Gamma(b(n) + v)} \prod_{i=1}^{n} \frac{\Gamma(b_i + v_i)}{b_i! \Gamma(v_i)} \binom{r + b(n) - 1}{b(n)} p^r q^{b(n)}$$

$$= \Gamma(v) p^r q^{b(t)} \prod_{i=1}^{t} \frac{\Gamma(b_i + v_i)}{b_i! \Gamma(v_i)} \sum_{b_n \geq 0} \frac{b(n)!}{b_n!} \frac{\Gamma(b_n + v_n)}{\Gamma(v_n) \Gamma(b(n) + v)} \frac{\Gamma(r + b(n))}{\Gamma(r) b(n)!} q^{b_n}$$

$$= p^r q^{b(t)} \frac{\Gamma(v) \Gamma(b(t) + r)}{\Gamma(b(t) + v) \Gamma(r)} \prod_{i=1}^{t} \frac{\Gamma(b_i + v_i)}{b_i! \Gamma(v_i)} \sum_{b_n \geq 0} \frac{(v_n)_{b_n} (b(t) + r)_{b_n}}{(b(t) + v)_{b_n} b_n!} q^{b_n}$$

$$= p^r q^{b(t)} \frac{\Gamma(v) \Gamma(b(t) + r)}{\Gamma(b(t) + v) \Gamma(r)} \prod_{i=1}^{t} \frac{\Gamma(b_i + v_i)}{b_i! \Gamma(v_i)} \, {}_2F_1(v_n, b(t) + r; b(t) + v; q)$$

since $b(n) = b(n-1) + b_n = b(t) + b_n$. ■

To evaluate Eqn. (20) use the log-gamma function and convert the product of gamma functions into a sum and difference of log's and then exponentiate. This avoids potential over- or under-flow problems.

It follows that the marginal distribution of B_1 is

$$\Pr(B_1 = b_1) = p^r q^{b_1} \frac{\Gamma(v)\Gamma(b_1 + r)}{\Gamma(b_1 + v)\Gamma(r)} \frac{\Gamma(b_1 + v_1)}{b_1!\Gamma(v_1)} \, {}_2F_1(v_2, b_1 + r; b_1 + v; q). \quad (21)$$

Since the two components of a GPDM are *a priori* independent Eqn. (12) implies the mean of B_1 is

$$\begin{aligned} \mathrm{E}(B_1) &= \mathrm{E}(\mathrm{E}(B_1 \mid B(n))) \\ &= \mathrm{E}(B(n)\Pi_1) = \mathrm{E}(\Lambda\Pi_1) = \mathrm{E}(\Lambda)\mathrm{E}(\Pi_1) \\ &= v_1 m / v. \end{aligned} \quad (22)$$

The variance of B_1 can be computed using Eqn. (13):

$$\begin{aligned} \mathrm{Var}(B_1) &= \mathrm{E}(\mathrm{Var}(B_1 \mid B(n))) + \mathrm{Var}(\mathrm{E}(B_1 \mid B(n))) \\ &= \frac{v_1(v - v_1)}{v^2(1 + v)} \mathrm{E}(B(n)^2) + \frac{v_1(v - v_1)}{v(1 + v)} \mathrm{E}(B(n)) + \frac{v_1^2 m(1 + m/r)}{v^2} \\ &= \frac{m v_1}{v} + \frac{m^2 v_1 (v(1 + r^{-1}) - v_1(1 - r^{-1}v))}{v^2(1 + v)} \end{aligned} \quad (23)$$

since $\mathrm{E}(B(n)) = m$, $\mathrm{Var}(B(n)) = m(1 + mr^{-1})$ and $\mathrm{E}(B(n)^2) = m + m^2(1 + r^{-1})$. Similarly, the covariance of B_1 and B_2 can be computed using Eqn. (14):

$$\begin{aligned} \mathrm{Cov}(B_1, B_2) &= \mathrm{E}(\mathrm{Cov}(B_1, B_2 \mid B(n))) + \mathrm{Cov}(\mathrm{E}(B_1 \mid B(n)), \mathrm{E}(B_2 \mid B(n))) \\ &= -\mathrm{E}\left[\frac{B(n) + v}{1 + v} \frac{B(n) v_i v_j}{v^2}\right] + \mathrm{Cov}(v_1 B(n)/v, v_2 B(n)/v) \\ &= m^2 \frac{v_1 v_2 (r^{-1}v - 1)}{v^2(1 + v)}. \end{aligned} \quad (24)$$

Eqn. (14) shows that the covariance between two marginals of a Dirichlet-multinomial is always negative. Eqn. (24) shows that the covariance between two marginals of a GPDM is negative if $v < r$, and positive otherwise. It becomes positive because the effect of the common mixing through the gamma prior for λ overwhelms the negative correlation given $B(n)$.

We will show in Section 6 that when $r = v$ the GPDM produces the Bornhuetter-Ferguson reserve; when $r > v$, and there is less uncertainty in the prior ultimate

than emergence, it favors the peg method; and when $r < v$ it favors the chain-ladder method. Which of these methods is indicated depends on the data being analyzed. Common practice favors the chain-ladder and Bornhuetter-Ferguson methods; the peg method is rarely used. Thus we expect to find that $r \leq v$ in data.

When $r \to \infty$ the variance of the gamma prior tends to zero and the ultimate claim count distribution tends to a Poisson with mean λ. The the marginal distribution of $B_1, \ldots, B_t$ becomes

$$\Pr(B_1 = b_1, \ldots, B_t = b_t) = e^{-\lambda}\lambda^{b(t)}\frac{\Gamma(v)}{\Gamma(b+v)}\prod_{i=1}^{i=t}\frac{\Gamma(b_i+v_i)}{\Gamma(v_i)b_i!} \times {}_1F_1(v'(t); b(t)+v; q) \tag{25}$$

where ${}_1F_1$ is a confluent hypergeometric function.

4 POSTERIOR DISTRIBUTIONS

In this section we consider the posterior distribution of Θ given observed development data:

$$\Pr(\Theta \mid \text{data}) = \frac{\Pr(\text{data} \mid \Theta)\Pr(\Theta)}{\Pr(\text{data})} \propto \Pr(\text{data} \mid \Theta)\Pr(\Theta). \tag{26}$$

When we are trying to identify the posterior distribution we can ignore any variable which is not a function of the parameters Θ.

Our data consists of multivariate observations of development data $B_1, \ldots, B_n$. However for all but the oldest accident year we only have a partial observation $B_1, \ldots, B_t$ for some $t < n$ with which to update the distribution of Θ. Recall that the prior distribution of $\Theta = (\Lambda, \Pi_1, \ldots, \Pi_n)$ is

$$\Pr(\Theta) = \Gamma(r, p/q) \times \text{Di}(v_1, \ldots, v_n) \tag{27}$$

where Λ has a gamma distribution, the proportions Π_i have a Dirichlet distribution and the two distributions are *a priori* independent. The next proposition shows how to update the prior distribution of Θ given a partial observation of claim counts. Let $\pi(t) = \sum_{i=1}^{t}\pi_i$, $\pi'(t) = \sum_{i=t+1}^{n}\pi_i$ and $b(t) = \sum_{i=1}^{t} b_i$.

Proposition 3 *Let $B_1, \ldots, B_n \mid \Theta$ have a GPDM distribution and let $t \leq n$. Then the posterior distribution of Θ given a partial observation $B_1, \ldots, B_t$ has density*

$$\Pr(\Theta = (\lambda, \pi_1, \ldots \pi_{n-1}) \mid B_1 = b_1, \ldots, B_t = b_t) = \kappa\, \lambda^{b(t)+r-1} e^{-\lambda(p/q+\pi(t))} \pi_1^{b_1+v_1-1} \cdots \pi_t^{b_t+v_t-1} \pi_{t+1}^{v_{t+1}-1} \cdots \pi_n^{v_n-1} \quad (28)$$

where

$$\kappa = \Gamma(v + b(t)) \left(q^{r+b(t)} \Gamma(r + b(t)) \prod_{i=1}^{t} \Gamma(b_i + v_i) \times \prod_{i=t+1}^{n} \Gamma(v_i) \times {}_2F_1(v'(t), b(t) + r; b(t) + v; q) \right)^{-1} \quad (29)$$

In Eqn. (28) the distribution of Λ is dependent on the distribution of observed claims through the term $\pi(t)$, so the two have become entangled. This is the reserving conundrum: counts through t periods are higher than expected; is this because we have observed a greater proportion of ultimate claims than expected or because ultimate claims will be higher than expected? Our model will show how to answer this question. When $t = n$, and we have a full observation, the posterior is no longer entangled because $\pi(n) = 1$; the posterior distribution is again a product of independent gamma and Dirichlet distributions.

Proof: Using Eqn. (3), the prior distribution for Θ, and the multinomial expansion in the penultimate step, we have $\Pr(\Theta \mid B_1, \ldots, B_t)$

$$\begin{aligned}
&\propto \Pr(B_1, \ldots, B_t \mid \Theta) \Pr(\Theta) \\
&= \sum_{b_{t+1}, \ldots, b_n} \frac{1}{b_1! \cdots b_n!} \pi_1^{b_1} \cdots \pi_n^{b_n} \lambda^{b(n)} e^{-\lambda} \lambda^{r-1} e^{-p\lambda/q} \pi_1^{v_1-1} \cdots \pi_n^{v_n-1} \\
&\propto \sum_{b \geq 0} \frac{\lambda^b}{b!} \left(\sum_{b_{t+1}+\cdots+b_n=b} \frac{b! \pi_{t+1}^{b_{t+1}} \cdots \pi_n^{b_n}}{b_{t+1}! \cdots b_n!} \right) \lambda^{b(t)+r-1} e^{-\lambda(1+p/q)} \\
&\qquad \times \pi_1^{b_1+v_1-1} \cdots \pi_t^{b_t+v_t-1} \pi_{t+1}^{v_{t+1}-1} \cdots \pi_n^{v_n-1} \\
&= \left(\sum_{b \geq 0} \frac{\lambda^b \pi'(t)^b}{b!} \right) \lambda^{b(t)+r-1} e^{-\lambda(1+p/q)} \pi_1^{b_1+v_1-1} \cdots \pi_t^{b_t+v_t-1} \pi_{t+1}^{v_{t+1}-1} \cdots \pi_n^{v_n-1} \\
&= \lambda^{b(t)+r-1} e^{-\lambda(\pi(t)+p/q)} \pi_1^{b_1+v_1-1} \cdots \pi_t^{b_t+v_t-1} \pi_{t+1}^{v_{t+1}-1} \cdots \pi_n^{v_n-1}.
\end{aligned}$$

To evaluate the constant κ use the exact form of the conditional and unconditional marginal distributions given in Proposition 2. ■

Figure 1 is a contour plot of the prior and posterior distribution of (Π_1, Λ). The left hand column shows the prior distributions with prior mean 250 and $\mathrm{E}(\Pi_1) = 0.5$. The middle column shows the posterior given an observation 40 below expected and the right hand column the posterior given an observation 40 above expected. The four rows show different degrees of precision in the priors.

Row 1. $r = 10$ and $v = 15$, so both priors have a moderately high uncertainty. Since $r < v$ the model gives weight to the chain-ladder method, so the posterior distributions lie north-east to south-west. Both are still relatively diffuse, reflecting the lack of information in the priors. The correlation between Π_1 and Λ in the posterior densities is very clear.

Row 2. $r = 10$ and $v = 50$, so the emergence is known with more prior certainty than the ultimate. The prior is now stretched along the y-axis, ultimate claims. Since emergence is known more precisely, this method is closer to the chain-ladder method (100% confidence in observed losses). In the picture we see the two posterior distributions lie north-east to south-west, corresponding to the chain-ladder method

Row 3. $r = 50$ and $v = 15$, so the prior ultimate is known with more certainty than the emergence. Now the prior is stretched along the x-axis, emergence. This method is closer to the peg method. The two posterior distribution lie east-west, corresponding to the less weight given to the observed claim information.

Row 4. $r = v = 50$, so both ultimate and emergence are known with more confidence. Compared to row 1 the prior is far more concentrated. Since $r = v$ this method reproduces the Bornhuetter-Ferguson—see below.

In the left hand column Λ and Π_1 are uncorrelated in all four examples.

The next corollary computes the exact Bayesian reserve: the expected number of unreported claims given claims to date. It is an important result and we will discuss it further in Section 4. The corollary assumes $n = 2$ and $t = 1$; using Lemma 1 we can reduce any particular reserving problem to this case.

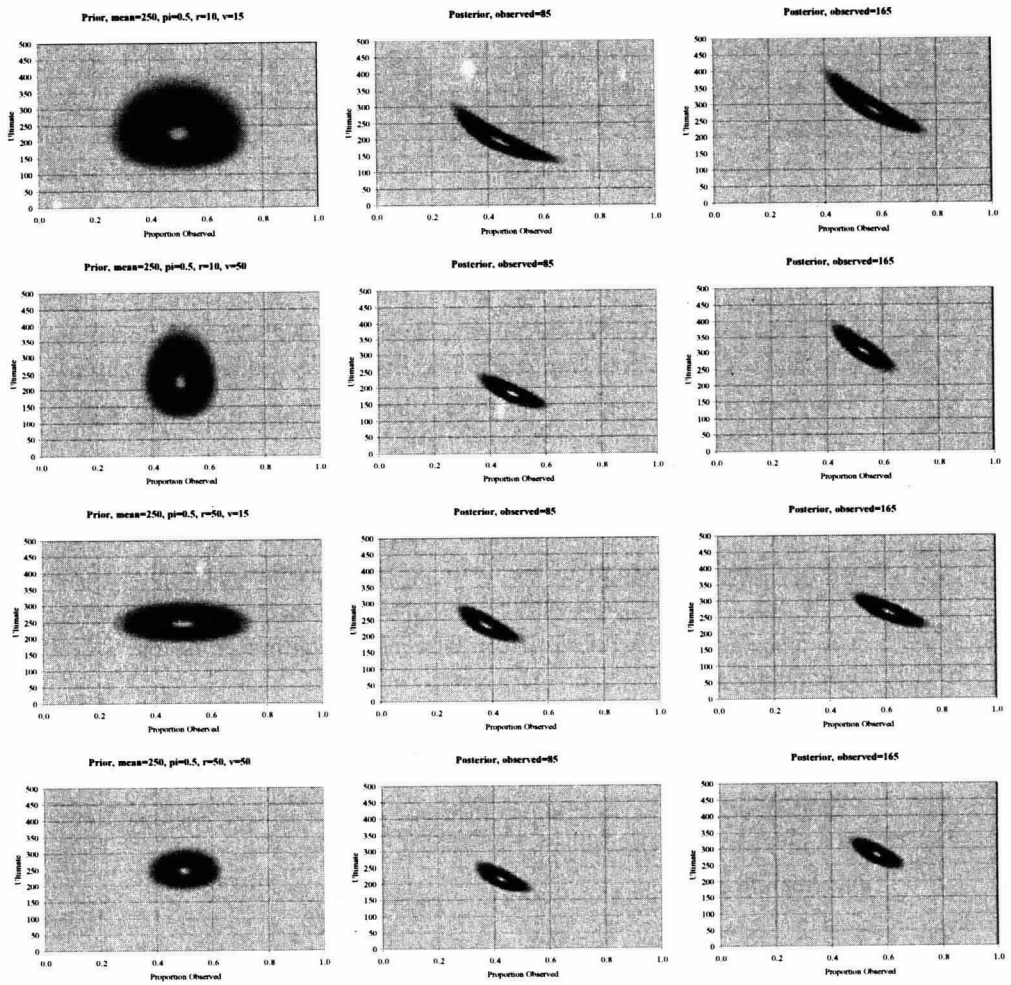


Figure 1: Prior and posterior density of Π_1 vs $\Lambda = \mathrm{E}(B(n))$ for various values of r and v and observed counts. Prior mean equals 250 and $\mathrm{E}(\Pi_1) = 0.5$. Left hand column shows prior density. Middle column shows posterior given observed counts 40 below expected; right hand column posterior given counts 40 above expected.

Corollary 1 *Let $n = 2$ and $t = 1$. Then*

$$\mathrm{E}(B_2 \mid B_1 = b_1) = q\frac{v_2(b_1 + r)}{b_1 + v}\frac{{}_2F_1(v_2 + 1, b_1 + r + 1; b_1 + v + 1; q)}{{}_2F_1(v_2, b_1 + r; b_1 + v; q)}. \quad (30)$$

Proof: By definition

$$\mathrm{E}(B_2 \mid B_1 = b_1) = \mathrm{E}((1 - \Pi_1)\Lambda \mid (\Theta \mid B_1)). \quad (31)$$

Now use the explicit form of the posterior distribution of $(\Theta \mid B_1)$ given in the proposition and integrate with respect to λ to get

$$\mathrm{E}(B_2 \mid B_1 = b_1) = \kappa \int_0^1 \int_0^\infty \lambda^{b_1+r} e^{-\lambda(\pi_1+p/q)} \pi_1^{b_1+v_1-1}(1 - \pi_1)^{v_2} d\lambda d\pi_1 \quad (32)$$

$$= \kappa \int_0^1 \frac{\Gamma(b_1 + r + 1)}{(\pi_1 + p/q)^{b_1+r+1}} \pi_1^{b_1+v_1-1}(1 - \pi_1)^{v_2} d\pi_1. \quad (33)$$

Substitute $w = 1 - \pi_1$ and re-arrange to get

$$\kappa q^{b_1+r+1}\Gamma(b_1 + r + 1) \int_0^1 w^{v_2}(1 - w)^{b_1+v_1-1}(1 - qw)^{-(b_1+r+1)} dw. \quad (34)$$

The result follows from Euler's integral representation of hypergeometric functions Eqn. (65). ∎

We can write Eqn. (30) as

$$(b_1 + r)\left(\frac{{}_2F_1(v_2, b_1 + r + 1; b_1 + v; q)}{{}_2F_1(v_2, b_1 + r; b_1 + v; q)} - 1\right) \quad (35)$$

using Whittaker and Watson [28, Chapter 14, Ex. 1]. Since b_1 is claims observed to date, the Bayesian expected ultimate is

$$\mathrm{E}(\Lambda \mid B_1 = b_1) = b_1 f + r(f - 1) \quad (36)$$

where f is the ratio of hypergeometric functions. Thus f is acting like a loss development factor, but one which is a function of b_1. It is interesting that the Bayesian estimate does not go through the origin because of the constant r term.

Using the same approach we can compute all moments of the posterior distribution.

Corollary 2 *Let $n = 2, t = 1$ and let a, b be non-negative integers. Then*

$$\mathrm{E}(\Lambda^a \Pi_1^b) = q^a \frac{(b_1 + r)_a (b_1 + v_1)_b}{(b_1 + v)_b} \frac{{}_2F_1(v_2, b_1 + r + a; b_1 + v + b; q)}{{}_2F_1(v_2, b_1 + r; b_1 + v; q)}. \quad (37)$$

5 PREDICTIVE DISTRIBUTIONS

The next proposition gives an expression for the predictive distribution

$$(B_{t+1}, \ldots, B_n \mid B_1, \ldots B_t).$$

Remember that $b(t) = \sum_{i=1}^{t} b_i$, $b'(t) = \sum_{i=t+1}^{n} b_i$ and $v = \sum_{i=1}^{n} v_i$.

Proposition 4 *Let $B_1, \ldots, B_n \mid \Theta$ have a GPDM distribution and let $1 \leq t \leq n-1$. Then the conditional distribution of $(B_{t+1}, \ldots, B_n)$ given $B_1, \ldots, B_t$ is*

$$q^{b'(t)} \frac{(b(t)+r)_{b'(t)}}{(b(t)+v)_{b'(t)}} \prod_{i=t+1}^{n} \frac{(v_i)_{b_i}}{b_i!} \, {}_2F_1(v'(t), b(t)+r; b(t)+v; q)^{-1}. \qquad (38)$$

Proof: Recall that

$$\begin{aligned}
&\Pr(B_{t+1}, \ldots, B_n \mid B_1, \ldots B_t) \\
&\qquad = \int \Pr(B_{t+1}, \ldots, B_n \mid B_1, \ldots B_t, \Theta) f(\Theta \mid B_1, \ldots B_t) d\Theta \\
&\qquad = \int \frac{\Pr(B_1, \ldots, B_n \mid \Theta)}{\Pr(B_1, \ldots, B_t \mid \Theta)} \frac{\Pr(B_1, \ldots, B_t \mid \Theta) f(\Theta)}{\Pr(B_1, \ldots, B_t)} d\Theta \\
&\qquad = \frac{\Pr(B_1, \ldots . B_n)}{\Pr(B_1, \ldots, B_t)}
\end{aligned}$$

Combine this with Proposition 2 and the definition of the GPDM and then cancel to complete the proof. ■

Proposition 4 shows the predictive distribution does not depend on the individual observed values $b_1, \ldots, b_t$ but only on their sum $b(t) = b_1 + \cdots b_t$. Thus the GPDM model has a kind of Markov property that the future development depends only on the total number of claims observed to date, and not on how those claims were reported over time.

Considering the probability distribution of the sum $B_{t+1} + \cdots + B_n$ given $B_1 + \cdots + B_t$ gives us the following corollary which we shall need later. This corollary can also be proved using induction and properties of the binomial coefficients.

Corollary 3

$$\sum_{b_1+\cdots+b_n=b} \left(\prod_{i=1}^{n} \frac{(v_i)_{b_i}}{b_i!} \right) = \frac{(v_1+\cdots+v_n)_b}{b!} \tag{39}$$

Using Lemma 1 we can add $B_{t+1}, \ldots, B_n$ and reduce to the case $n = 2, t = 1$. Then Eqn. (38) gives the conditional distribution of unreported claims B_2 given claims reported to date b. This provides a closed form expression for the posterior distribution which is exactly the distribution required for claim count reserving.

Corollary 4

$$\Pr(B_2 = b_2 \mid B_1 = b) = q^{b_2} \frac{(b+r)_{b_2}}{(b+v)_{b_2}} \frac{(v_2)_{b_2}}{b_2!} \, {}_2F_1(v_2, b+r; b+v; q)^{-1}. \tag{40}$$

The probabilities $\Pr(B_2 = j \mid B_1 = b)$ can be computed recursively using

$$\Pr(B_2 = j+1 \mid B_1 = b) = \Pr(B_2 = j \mid B_1 = b) \frac{q}{j+1} \frac{(b+r+j)(v_2+j)}{(b+v+j)} \tag{41}$$

for $j \geq 0$ and

$$\Pr(B_2 = 0 \mid B_1 = b) = {}_2F_1(v_2, b+r; b+v; q)^{-1}. \tag{42}$$

Figure 2 shows six examples of the density $B_2 \mid B_1$ for various values of v and r. They are the two key shape parameters. For comparison, each plot also has a Poisson with the same mean 30.305 as the $r = 100, v = 1$ frequency.

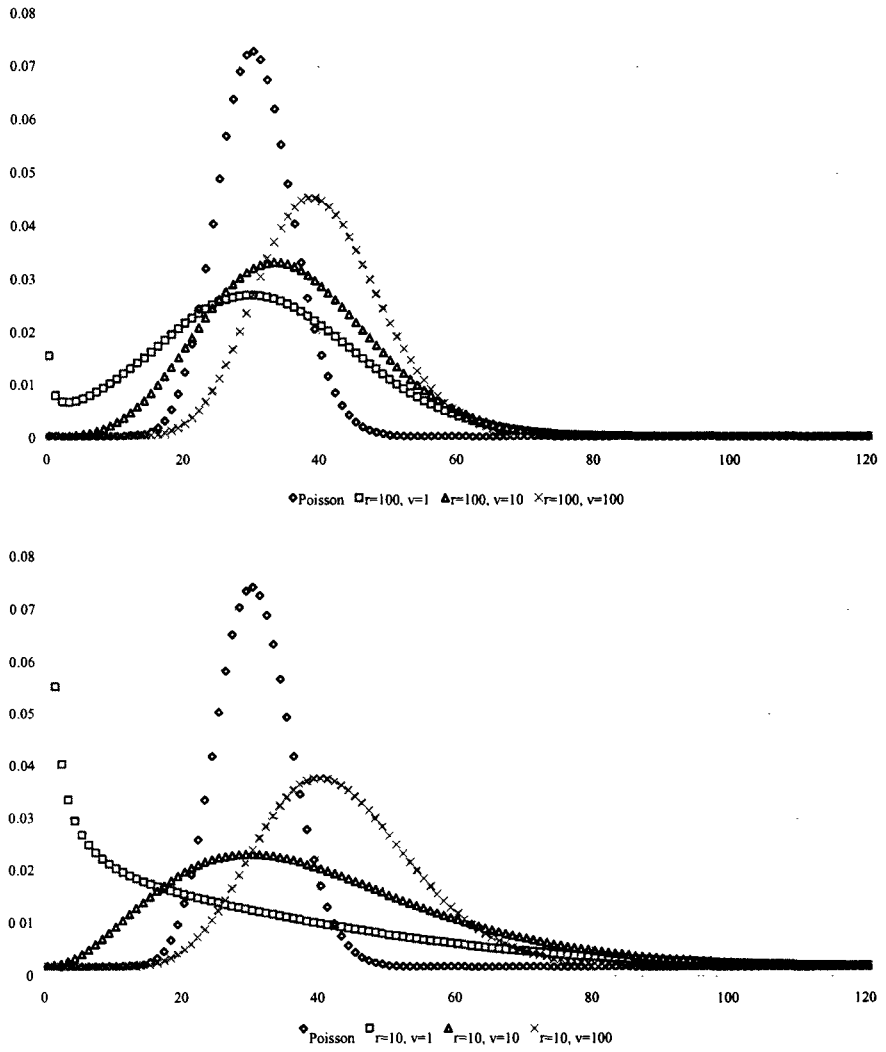


Figure 2: $(B_2 \mid B_1)$ for various values of r and v. $n = 100$, $b = 65$, and $v_1/v = 0.6$.

It follows from Eqn. (40) that the probability generating function[2] of $B_2 \mid B_1 = b$ is

$$G(z) = \frac{{}_2F_1(v_2, b+r; b+v; zq)}{{}_2F_1(v_2, b+r; b+v; q)}. \tag{45}$$

Therefore $B_2 \mid B_1 = b$ is a generalized power series distribution and a generalized hypergeometric probability distribution according to the classification in Johnson et al. [8]. It does not, however, appear in Table 2.4 of [8].

Differentiating G, using Equations 63 and 64 for the derivatives of the hypergeometric function, gives the factorial moments of $B_2 \mid B_1 = b$:

$$\mathrm{E}(B_2 \mid B_1 = b) = \frac{qv_2(b+r)}{b+v} \; \frac{{}_2F_1(v_2+1, b+r+1; b+v+1; q)}{{}_2F_1(v_2, b+r; b+v; q)}, \tag{46}$$

which reproduces Corollary 1, and more generally

$$\mu_{(k)}(B_2 \mid B_1 = b) = \frac{q^k(v_2)_k(b+r)_k}{(b+v)_k} \; \frac{{}_2F_1(v_2+k, b+r+k; b+v+k; q)}{{}_2F_1(v_2, b+r; b+v; q)}. \tag{47}$$

6 THE CONTINUUM OF RESERVING METHODS

Corollary 1 is very important. It provides a Bayesian estimate of unreported claims given claims to date which is exactly the quantity the reserving actu-

[2]The probability generating function of a nonnegative discrete random variable X is defined as

$$G(z) = \mathrm{E}(z^X).$$

The (descending) kth factorial moment of a random variable X is defined as

$$\mu_{(k)}(X) = E(X(X-1)\cdots(X-k+1)).$$

Factorial moments can be computed from the probability generating function by differentiating:

$$\mu_{(k)} = \frac{d^k G(z)}{dz^k}\Big|_{z=1}. \tag{43}$$

It is easy to compute the central moments and moments about zero from the factorial moments. For example

$$\mathrm{Var}(X) = \mu_{(2)} + \mu - \mu^2. \tag{44}$$

See Johnson et al. [9] for more general relationships.

ary must estimate. In this section we show that special or limiting cases of the GPDM include the peg, the Bornhuetter-Ferguson, Benktander, and the chain-ladder methods. Then we compare the GPDM to traditional methods over practical ranges of the parameters r and v. The model confirms the suggestion in Renshaw and Verrall [22] that the chain-ladder is just one of many appropriate methods. A schematic showing how the GPDM interpolates between other reserving methods is shown in Figure 3.

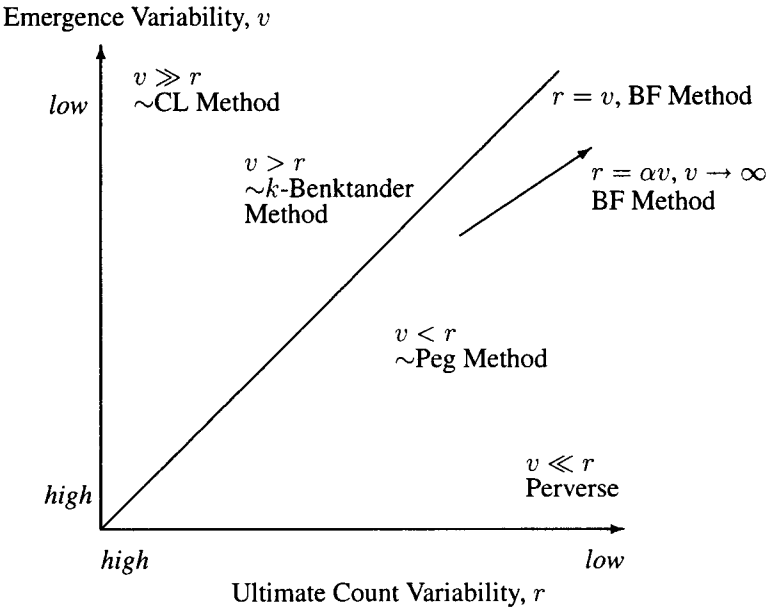


Figure 3: Schematic showing the behavior of the GPDM reserve as (r, v) vary. Low r (resp. v) corresponds to high uncertainty in ultimate counts (resp. claim emergence). The $r = v$ diagonal is exactly the Bornhuetter-Ferguson method.

Using Lemma 1 we can reduce each accident year to the case $n = 2, t = 1$. B_1 denotes observed claims and B_2 unreported claims. $(B_1, B_2 \mid \Theta)$ has a GPDM distribution; $\Theta = (\Lambda, \Pi_1)$. Λ has a gamma distribution with mean m, the prior

expected number of ultimate claims, and variance $m(1 + m/r)$. Π_1 has a beta distribution with parameters v_1 and v_2. Let $v = v_1 + v_2$ and $\pi := \mathrm{E}(\Pi_1) = v_1/v$. Per Section 2, the parameters of the gamma distribution are r and p/q where $p = r/(r+m)$ and $q = 1-p = m/(r+m)$. Higher values of r and v correspond to lower variances of Λ and Π_1 respectively. As $r \to \infty$ the claim count distribution tends to a Poisson.

We are going to compare the six estimates of unreported claims: GPDM, peg, chain-ladder, Bornhuetter-Ferguson, k-Benktander, and linear least squares. Conditional on observed claims to date of b estimated unreported claims for each method are denoted $b'_*(b)$ where $* = g, p, c, b, k, l$ indicates the method. The estimate of ultimate claims corresponding to each method is therefore simply $b + b'_*(b)$.

1. The GPDM method

$$b'_g(b) = \mathrm{E}(B_2 \mid B_1 = b) = (b+r)\left(\frac{{}_2F_1(v_2, b+r+1; b+v; q)}{{}_2F_1(v_2, b+r; b+v; q)} - 1\right). \tag{48}$$

2. The peg method

$$b'_p(b) := (m-b)^+. \tag{49}$$

   The peg ultimate is insensitive to observed data—until observed claims exceed the peg! The peg is an extreme reserving method. It ignores actual emergence completely.

3. The chain-ladder method

$$b'_c(b) := \frac{(1-\pi)b}{\pi} \tag{50}$$

   see Mack [16] or Renshaw and Verrall [22]. π is usually estimated from the data as a product of link ratios. Each link ratio is the weighted average development from one period to the next over all available accident periods. The chain-ladder method is at the opposite extreme to the peg method. It completely ignores prior estimates of ultimate counts.

4. The Bornhuetter-Ferguson method estimate

$$b'_b(b) := m(1-\pi) \tag{51}$$

see Mack [16]. This estimate of unreported claims is completely insensitive to the observation b. The Bornhuetter-Ferguson method is sometimes regarded as an extreme, but it is actually a middle-ground method between the chain-ladder and peg methods.

5. The k-Benktander method, $k = 0, 1, 2, \ldots$

$$b'_k(b) := (1-(1-\pi)^k)b'_c(b) + (1-\pi)^k b'_b(b) \tag{52}$$

see Mack [16]. When $k = 0$ this reduces to the Bornhuetter-Ferguson. As $k \to \infty$, $b'_k(b) \to b'_c(b)$. The Benktander methods are all linear in b. They are a credibility weighting of the Bornhuetter-Ferguson and chain-ladder methods.

6. The linear least squares, or greatest accuracy credibility, estimate

$$b'_l(b) := \alpha + \beta b \tag{53}$$

where α and β are chosen to minimize the expected squared error. This approach is described in Klugman et al. [10, Section 5.4] from a credibility perspective and in Murphy [18] from a linear least squares loss development perspective. Solving by differentiating $\mathrm{E}((B_2 - \alpha - \beta B_1)^2)$ with respect to α and β and setting to zero gives

$$\alpha = \mathrm{E}(B_2) - \beta \mathrm{E}(B_1) \qquad \beta = \frac{\mathrm{Cov}(B_1, B_2)}{\mathrm{Var}(B_1)}. \tag{54}$$

In order to actually compute α and β we need a bivariate distribution for B_1 and B_2; we use the GPDM. The variance and covariance are computed in Eqn. (23) and Eqn. (24). By construction b'_l will be the least squares line through b'_g. When $r = v$, and B_1 and B_2 are uncorrelated, b'_l reduces to the Bornhuetter-Ferguson.

Neither the chain-ladder nor the Bornhuetter-Ferguson method is sensitive to the relative variance of ultimate losses $B_1 + B_2$ and the proportion of claims observed Π_1. This is a weakness that can be illustrated by considering two hypothetical situations. In the first, the ultimate is estimated with low confidence but the claim reporting pattern is very predictable, so $r < v$. We would favor the chain-ladder estimate over the prior m. This corresponds to the second row in Figure 1. In the second situation, the ultimate claim count distribution is known with a high confidence, but the reporting pattern is estimated with less confidence, so $r > v$. Here we would weigh the prior estimate m more than the chain-ladder which relies on the proportion reported. This corresponds to the third row in Figure 1. Corollary 1 provides a probabilistic model of these intuitions that continuously interpolates from one extreme to the other. The GPDM captures and models the process behind the actuarial judgment of selecting appropriate reserves. By providing a quantification of what is currently a judgmental process the model should be of great value to the practicing actuary.

Here are six examples of how the GPDM behaves for different values of r and v. They are illustrated in Figure 3.

1. For fixed r, $b'_g(b) \to m(1-\pi)(b+r)/(m\pi+r)$ as $v \to \infty$. *Proof:* As $v \to \infty$

$$\begin{aligned} {}_2F_1(v_2, b+r; b+v; q) &= {}_2F_1((1-\pi)v, b+r; b+v; q) \\ &\to {}_2F_1(1, b+r; 1; (1-\pi)q) \\ &= (1-(1-\pi)q)^{-(b+r)} \end{aligned}$$

by Eqn. (66). Therefore

$$b'_g(b) \to \frac{q(1-\pi)(b+r)}{1-(1-\pi)q} = \frac{m(1-\pi)(b+r)}{m\pi+r}. \tag{55}$$

We can write this limit as a credibility weighting of the chain-ladder and Bornhuetter-Ferguson with credibility $z = m\pi/(m\pi+r)$ given to the chain ladder:

$$\frac{m(1-\pi)(b+r)}{m\pi+r} = \left(\frac{m\pi}{m\pi+r}\right)\frac{(1-\pi)b}{\pi} + \left(\frac{r}{m\pi+r}\right)((1-\pi)m). \tag{56}$$

This equation corresponds to the k-Benktander method with

$$k = \frac{\log(r/(m\pi + r))}{\log(1 - \pi)}. \tag{57}$$

2. As $r \to 0$ the GPDM reserve tends to the chain ladder reserve, $b'_g(b) \to b'_c(b)$. *Proof:* Use the limit of the hypergeometric function as $r \to 0$ and $q \to 1$.

3. If $r = v$ then the GPDM reserve equals the Bornhuetter-Ferguson reserve, $b'_g(b) = b'_b(b)$. *Proof:* Applying Eqn. (66) to Eqn. (46) gives

$$b'_g(b) = \frac{qv_2}{1 - q} = \frac{qv_2}{p}. \tag{58}$$

Since $r = v$, $q/p = m/v$ and so

$$b'_g(b) = \frac{mv_2}{v} = m(1 - \pi) \tag{59}$$

as required. The case $r = v$ represents an exact balance between the uncertainty in ultimate losses and claim count emergence which reproduces the Bornhuetter-Ferguson. By Eqn. (24) it also represents the case when B_1 and B_2 are uncorrelated.

4. If $r = \alpha v$ for a constant α then as $v \to \infty$ the GPDM reserve converges to the Bornhuetter-Ferguson: $b'_g(b) \to b'_b(b)$.

5. For fixed small $v > 0$ the GPDM reserve is close to the peg reserve as $r \to \infty$. See the figures below.

6. As $v \to 0$ and $r \to \infty$ the GPDM reserve tends to zero if $b > 0$ and m if $b = 0$. This is a perverse kind of reserve! It is possible to prove this analytically, but heuristically the reason is that as $v \to 0$ the Dirichlet distribution becomes concentrated at the corners. Thus the claims all tend to be reported at once. So if any claims have been reported then no more are expected. On the other hand, if none have been reported they should all still be held in reserve.

The fourth point needs elaborating because it appears to contradict the main result of Renshaw and Verrall [22]. Their model assumes incremental claims B_i have a Poisson distribution and hence emergence is modeled with a multinomial distribution. They prove their model reproduces the chain-ladder reserve when the parameters are determined using maximum likelihood. As $r \to \infty$ the GPDM ultimate has a Poisson distribution. As $v \to \infty$ the Dirichlet prior becomes a degenerate distribution, so the DM becomes a multinomial distribution conditional on ultimate counts $B(n)$. In this situation B_i will also have a Poisson distribution and so in the limit the GPDM model appears to be the same as Renshaw and Verrall's, and yet it gives the Bornhuetter-Ferguson reserve and not the chain-ladder. The reconciliation of this apparent contradiction is that Renshaw and Verrall fit the emergence pattern (means of the multinomial) and the prior accident year means from the data. If we interpret these parameters as prior estimates then their model produces exactly expected emergence—see [22, Eqn. (2.4)]. In the GPDM model the emergence pattern and accident year means are given *a priori*. As $v, r \to \infty$ both parameters become certain. If losses emerge exactly as expected then the chain-ladder and Bornhuetter-Ferguson methods agree and so the GPDM would also give the chain-ladder reserve. However, actual emergence from the GPDM need not be exactly equal to expected because the means and emergence are specified *a priori*. Note that in the Poisson-multinomial model ($r, v \to \infty$) B_1 and B_2 are independent so the linear least squares method also reproduces the Bornhuetter-Ferguson.

These mathematical limits of the GPDM method are mainly of academic interest. However, the way the GPDM interpolates between the common reserving methods for realistic values of r and v is of practical interest because it provides analytical guidance to supplement actuarial judgment. We now explore that interpolation.

Figure 4 shows a plot of $b'_g(b)$ against b for $r = 25$, $\pi = 0.45$, $m = 110$ and $v = 0.1, 1, 10, 25. 100, 1000$. Each plot also shows the peg, chain-ladder, Bornhuetter-Ferguson, k-Benktander and linear least squares reserves. The value of k is determined by Eqn. (57). Figure 4 ties back to the six points we made about b'_g.

- The four standard methods do not change with v. The linear least squares method is sensitive to v and is a line through b'_g as expected.
- Point 1 is illustrated by $v = 100$ and $v = 1000$. The GPDM method tends to the predicted k-Benktander method line for larger k. If we had plotted large v and small r the GPDM line would eventually convert up to the chain-ladder line, per Point 2.
- Point 3 is illustrated by $v = 25 = r$; the GPDM line lies underneath the Bornhuetter-Ferguson line.
- The fact that the GPDM favors the peg method when $v < r$ and the chain-ladder method when $v > r$ is shown in the increasing slope of the GPDM line from the first plot to the last.
- Point 5 is illustrated by $v = 1$: b'_g is close to the peg method.
- Point 6 is illustrated by $v = 0.1$ which shows $b'_g \to 0$ for larger b.

Figures 5 and 6 are two views of the bivariate density of (B_1, B_2) computed with $m = 110$ claims, $r = 25$ and $\pi = 0.45$, so $\mathrm{E}(B_1) = 49.5$ and $\mathrm{E}(B_2) = 109.5$. The nine contour plots correspond to $v = 0.1, 1, 2.5, 5, 10, 25, 100, 1000, 10000$. As expected, when $v < 25$ B_1 and B_2 are negatively correlated. When $v = 25$ they are uncorrelated and when $v > 25$ they are positively correlated. The posterior distribution of $B_2 \mid B_1 = b_1$ is simply a re-scaled vertical slice through these distributions, so the reader should be able to connect these plots with the plots of $b'_g(b)$. The cases $v = 0.1$ and $v = 1$ help explain how the GPDM reacts given extreme uncertainty in the payout pattern. The 3-d plot explains why the contour plot seems to disappear: the probability becomes concentrated along the axes.

This completes our theoretical investigation of the properties of the GPDM distribution. We have produced easy-to-compute expressions for the marginal and conditional distributions and written down the mean of the posterior distribution of unreported claims given claims observed to date. Next we show how the GPDM can be used in practice by applying it to a particular claim count development triangle.

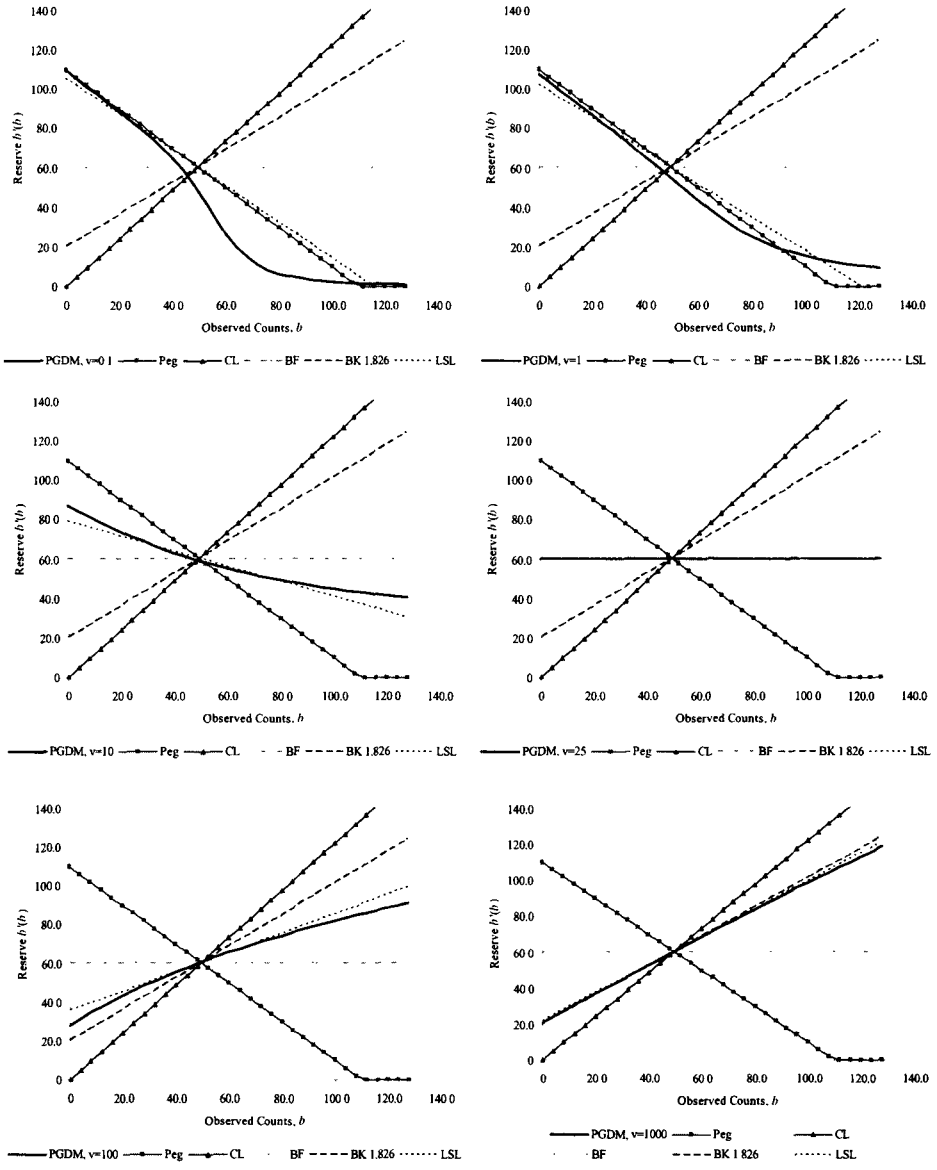


Figure 4: $b'_p(b)$ for $r = 25$, $\pi = 0.45$, $\mathrm{E}(\Lambda) = 110$ and $v = 0.1, 1, 10, 25, 100, 1000$, compared with the peg, chain-ladder, Bornhuetter-Ferguson, Benktander k and linear least squares methods. $k = 1.826$ determined by Eqn. (57).

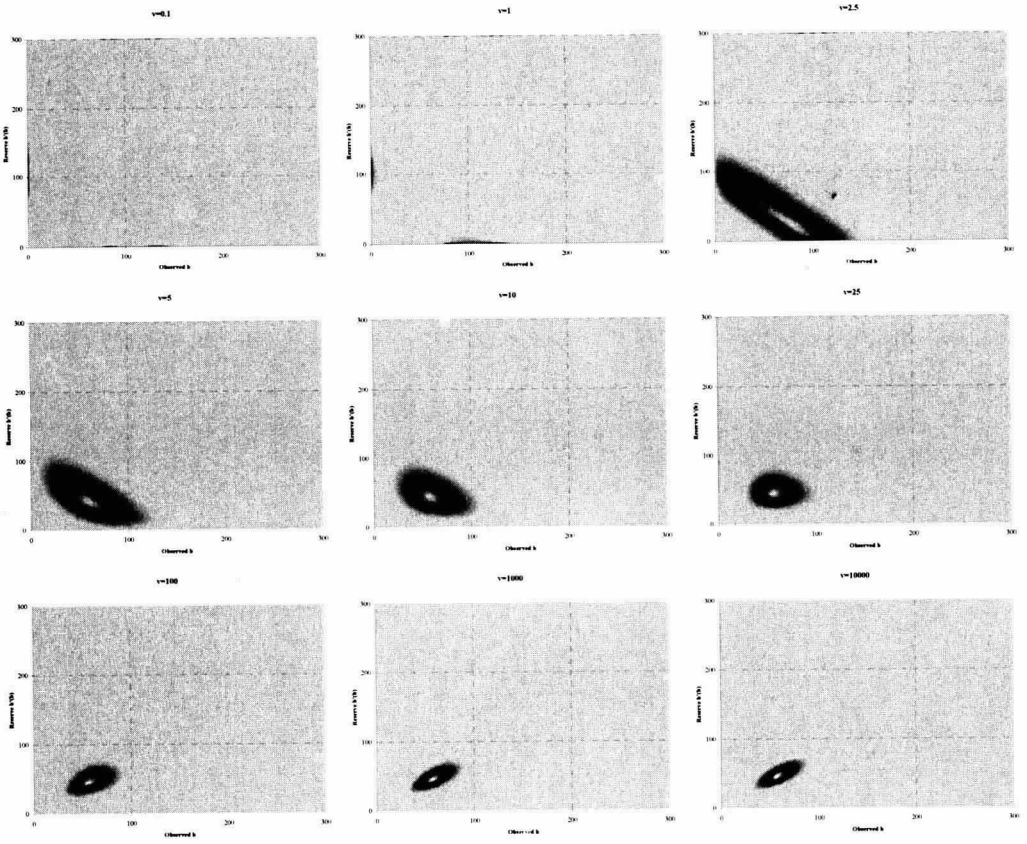


Figure 5: Contour plots of the bivariate density (B_1, B_2) with $r = 25$, $m = 110$, $\pi = 0.45$ shown for $v = 0.1, 1, 2.5, 5, 10, 25, 100, 1000, 10000$.

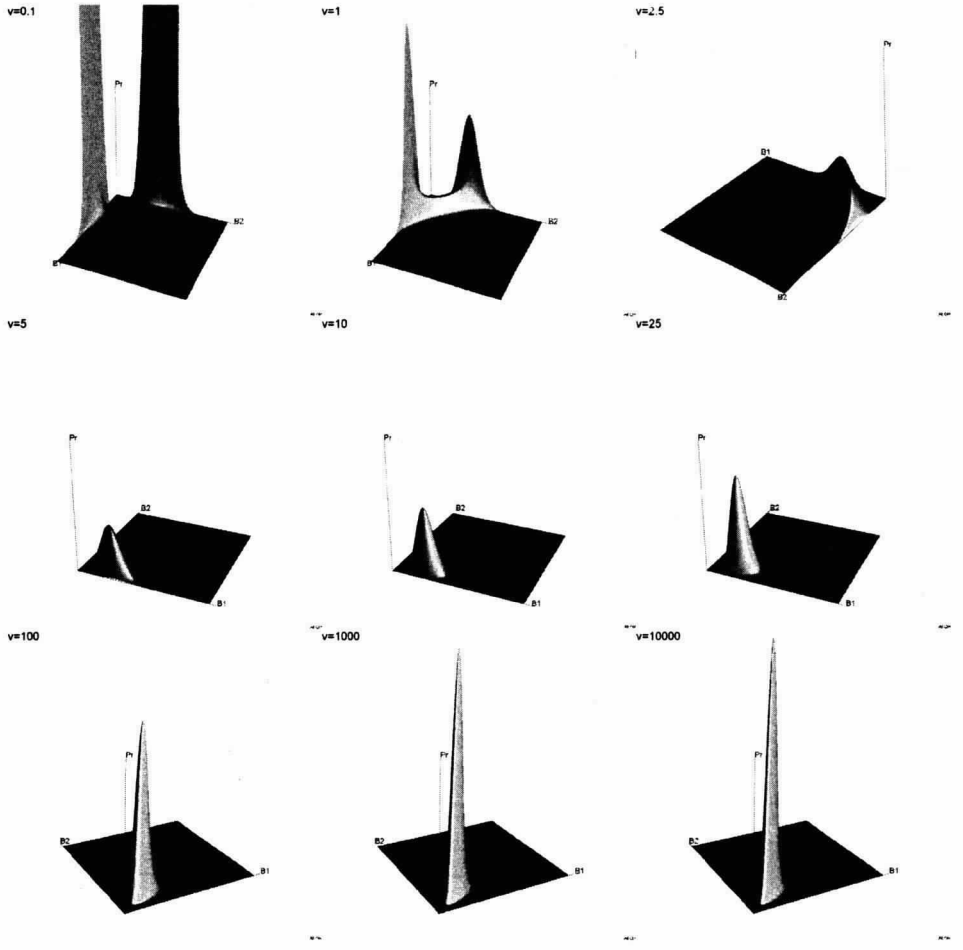


Figure 6: Three dimensional density plots of the bivariate density (B_1, B_2) with $r = 25$, $m = 110$, $\pi = 0.45$ shown for $v = 0.1, 1, 2.5, 5, 10, 25, 100, 1000, 10000$. The z-scales are all the same. The orientations vary by plot.

7 PARAMETER ESTIMATION

The GPDM model for n periods of development uses $n + 2$ parameters; of these the n development-related parameters $v_1, \ldots, v_n$ would usually be shared across multiple accident years. The prior mean m of the ultimate distribution would vary by accident year and r would generally be considered common. Thus to model a development triangle with n accident years and development periods there are $2n + 1$ parameters. If there is a good exposure measure then prior mean could be modeled as a common frequency times exposure and that would reduce the number of parameters to $n + 2$.

Reasonable initial estimates for m should be available from the pricing department. A view of r could be driven by a macro line-of-business level study. Alternatively we could take r to be very small corresponding to a non-informative prior for the ultimate.

Kotz et al. [11] discuss using sample moments to estimate the parameters v_t of a Dirichlet-multinomial. Let M'_{1t} be the sample mean of the proportion of claims observed in the tth period (computed with respect to the chain-ladder, for example), and let M'_{21} be the mean of the square of the proportion of claims observed in the first period. Then reasonable starting parameters are

$$\widehat{v}_t = \frac{(M'_{11} - M'_{21})M'_{1t}}{M'_{21} - (M'_{11})^2}, \quad i = 1, \ldots, n-1 \tag{60}$$

$$\widehat{v}_n = \frac{(M'_{11} - M'_{21})(1 - \sum_{t=1}^{n-1} M'_{1t})}{M'_{21} - (M'_{11})^2}. \tag{61}$$

Alternatively taking $v_1 = \cdots = v_n = 1$ gives a prior emergence distribution equal over all periods, which could be regarded as a non-informative prior.

Table 1: Incremental Claim Count Data

Year	1	2	3	4	5	6	7	8	9	10	b
1990	40	124	157	93	141	22	14	10	3	2	606
1991	37	186	130	239	61	26	23	6	6		714
1992	35	158	243	153	48	26	14	5			682
1993	41	155	218	100	67	17	6				604
1994	30	187	166	120	55	13					571
1995	33	121	204	87	37						482
1996	32	115	146	103							396
1997	43	111	83								237
1998	17	92									109
1999	22										22

Table 2: Loss Development Factors

AY	1 : 2	2 : 3	3 : 4	4 : 5	5 : 6	6 : 7	7 : 8	8 : 9	9 : 10
1990	4.100	1.957	1.290	1.341	1.040	1.024	1.017	1.005	1.003
1991	6.027	1.583	1.677	1.103	1.040	1.034	1.009	1.008	
1992	5.514	2.259	1.351	1.081	1.041	1.021	1.007		
1993	4.780	2.112	1.242	1.130	1.029	1.010			
1994	7.233	1.765	1.313	1.109	1.023				
1995	4.667	2.325	1.243	1.083					
1996	4.594	1.993	1.352						
1997	3.581	1.539							
1998	6.412								
Wtd. Avg.	5.055	1.930	1.350	1.134	1.035	1.023	1.011	1.007	1.003

8 EXAMPLE

We now give an example to illustrate the use of the GPDM to estimate the distribution of unreported claims.

The incremental claim count data is shown in Table 1 and the claim count development factors are shown in Table 2. The right hand column shows total counts observed to date b. This data was analyzed by de Alba [5]. Using a Bayesian model he found a mean of 919 outstanding claims with a standard deviation of 79.51.

We use Eqn. (20) to compute the likelihood of each row of the development triangle and then determine the maximum likelihood estimate parameters. Initial parameter values were $r = 25$, the chain ladder estimates for the prior means m_i by accident year, and the estimates of v_t given in the previous section. The starting values and maximum likelihood estimates for m_i are shown in Table 3. Table 4 shows the same thing for v_t along with the incremental reporting patterns for both estimates. The maximum likelihood estimator for r is 1625458.8 which is much closer to Poisson than the starting value and $v = 129.018$ so the model has $r > v$. Clearly the development pattern for this triangle is quite erratic, and so a low v is expected. One reason that r is so large is the use of a different variable m_i for each accident year. These parameters absorb some of the claim count variability and increase r.[3]

Table 5 shows the GPDM, chain ladder and Bornhuetter-Ferguson reserves, and the standard deviation and coefficient of variation of the GPDM reserve. The overall reserve is slightly lower than the chain ladder. It is interesting that the reserves are actually higher for the older years and lower for the more recent years.

Figure 7 shows the distribution of the GPDM reserve. This distribution is the sum of the reserve distributions for each accident year, assuming they are independent. Figure 8 shows the evolution of the predictive distribution of ultimate claims for the oldest accident year, as more and more claim information becomes available.

[3]Exposure information was not available for this triangle, but estimating an exposure base produced a modeled $\hat{r} = 196.1$, lowered the reserve to 889 from 895 and increased the standard deviation of the reserve to 55.4 from 40.5. The estimate of v declined slightly.

Table 3: Starting and maximum likelihood estimates of m for each accident year.

Year	CL Mean	Prior Mean m
1990	606.0	606.0
1991	716.4	718.2
1992	689.0	692.5
1993	616.7	621.6
1994	596.2	601.8
1995	520.8	527.1
1996	485.1	487.9
1997	391.9	390.0
1998	347.9	339.8
1999	355.0	333.0

Table 4: Starting and Maximum Likelihood Estimates for v_i with implied incremental and cumulative proportion of claims reported

t	1	2	3	4	5	
Initial v_t	13.195	52.531	60.500	43.094	22.834	
Incremental	0.064	0.253	0.292	0.208	0.110	
Cumulative	0.064	0.317	0.608	0.816	0.926	
MLE v_t	8.477	32.702	36.891	26.322	13.367	
Incremental	0.066	0.253	0.286	0.204	0.104	
Cumulative	0.066	0.319	0.605	0.809	0.913	
t	6	7	8	9	10	$v = \sum v_i$
Initial v_t	6.636	4.428	2.225	1.384	0.686	207.514
Incremental	0.032	0.021	0.011	0.007	0.003	
Cumulative	0.958	0.979	0.990	0.997	1.000	
MLE v_t	4.488	3.010	1.729	1.246	0.786	129.018
Incremental	0.035	0.023	0.013	0.010	0.006	
Cumulative	0.948	0.971	0.984	0.994	1.000	

Table 5: Comparison of Reserve Estimates ($r = 1625458.8$, $LL = -221.42$)

Year	b	m	$b'_g(b)$	Std. Dev.	CV	$b'_c(b)$	$b'_b(b)$
1991	714	718	4	5.018	1.191	2	2
1992	682	693	11	7.671	0.728	7	7
1993	604	622	18	9.300	0.529	13	13
1994	571	602	31	11.688	0.380	25	25
1995	482	527	45	12.910	0.287	39	39
1996	396	488	92	15.965	0.174	89	90
1997	237	390	153	16.774	0.110	155	154
1998	109	340	231	17.344	0.075	239	233
1999	22	333	311	18.072	0.058	333	312
Total	4423		895	40.465	0.045	902	876

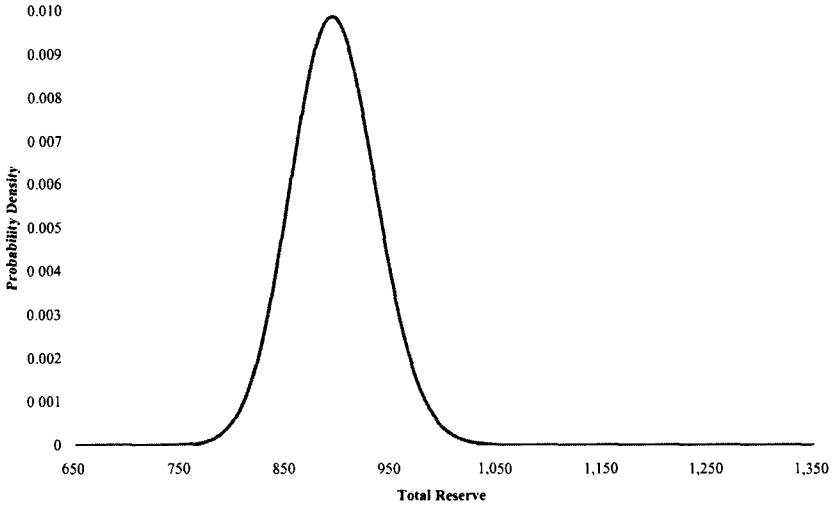


Figure 7: Distribution of total reserve.

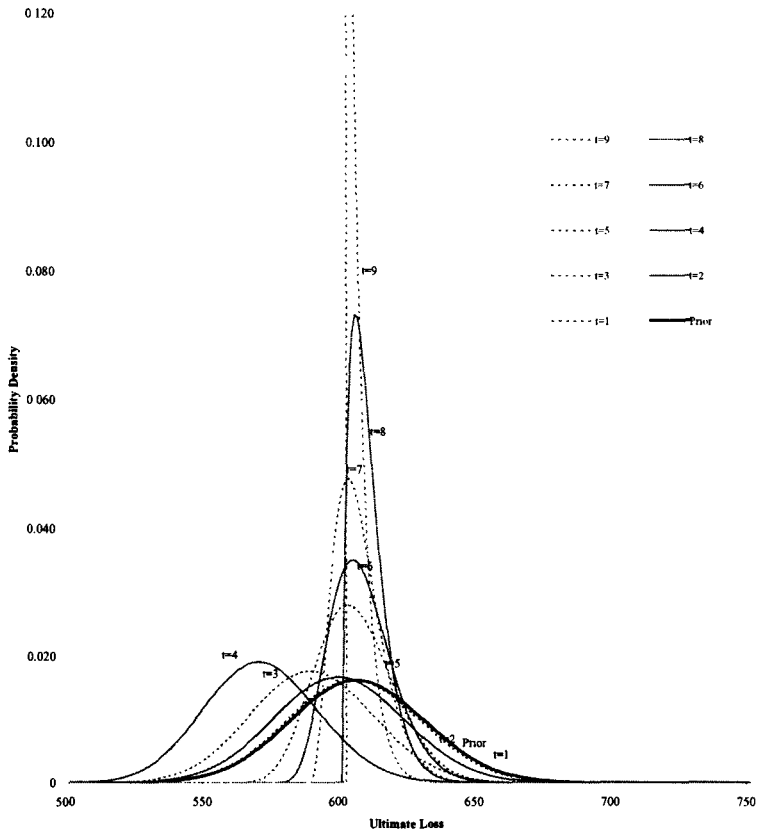


Figure 8: Predictive distribution of ultimate losses for oldest accident year starting with prior and adding observed losses for each development period.

9 EXTENSION TO LOSS DEVELOPMENT

The GPDM model applies to claim counts. Understanding claim counts can be a hard problem and the power of the model for working with counts should not be discounted. Nonetheless an extension to loss development is desirable. There will not be a similarly tractable model for loss development—just as there is no analog of the Poisson-gamma model for aggregate loss distributions. However, the general philosophy of the GPDM model, that the appropriate reserve depends on the relative variance of ultimate losses and loss emergence, carries over intact to losses and the problem is to determine a suitable bivariate distribution for observed and unobserved claims. Once that bivariate distribution is in hand numerical methods can be used to produce predictive reserve distributions. There are at least two approaches we could take.

Firstly, like Renshaw and Verrall [22], we can just use the GPDM directly to model losses. This is actually a more rational assumption than it seems. For a large book of business with a "tame" severity distribution (for example, where all policies have a low limit) the severity quickly diversifies and the normalized distribution of ultimate losses converges in distribution to the distribution of Λ as the book gets larger, see Daykin et al. [4, Appendix C] or Mildenhall [17, Section 2.10]. This method would be particularly appropriate when the maximum severity is of the same order of magnitude as the average severity because the diversification would occur more quickly. Working layer excess of loss reinsurance is an example.

The second approach is to try and determine a bivariate distribution for B_1 and B_2 and work with it numerically. Here we need a distribution of severity at tth report and ultimate. This could be estimated directly from a transactional loss database. The severity component would be combined with a mixed count emergence model like the GPDM. The aggregate distributions could be computed numerically using Fourier or fast Fourier transforms, or simulation. Alternatively, given the model specification and conditional severity distribution, we could use WinBUGs and MCMC techniques—see Verrall [27]. Understanding individual claim severity development is a great opportunity for further actuarial research in loss development. Since claim databases for most lines (except workers compensation) are much smaller than exposure databases this is also a practical thing to do.

10 CONCLUSIONS

We have introduced the GPDM model of claim count development and computed many of its important actuarial properties. The GPDM model incorporates estimates of the variability of ultimate claims and the claim emergence pattern into its estimates of reserves. Selecting between different reserve estimates is something usually done via actuarial judgment. The GPDM model can help bolster actuarial judgment by supplying a well-defined analytic selection framework.

The model includes the chain-ladder and Bornhuetter-Ferguson methods as special cases, and also closely approximates the peg method and k-Benktander methods. Thus it provides a rich modeling framework for the practitioner.

The GPDM is a statistical model of claim development which can be fit using maximum likelihood. Given an exposure base, it can also be used to fit ultimates in the presence of covariates, again also using maximum likelihood. The model is easy to use and provides full posterior distributions rather than just a point estimate and standard deviation.

A Appendix: Hypergeometric Functions

The hypergeometric function ${}_2F_1$ is defined as

$$ {}_2F_1(a,b;c;q) = \sum_{k\geq 0} \frac{(a)_k (b)_k}{(c)_k k!} q^k. \tag{62}$$

The notation $(a,b;c;q)$ indicates there are two variables in the numerator, one in the denominator and one argument (there are generalizations the reader can readily imagine). The series is absolutely convergent for $|q| < 1$ and conditionally convergent for $|q| = 1$. In our applications q is real and $0 < q < 1$, so convergence is not an issue. Hypergeometric functions have been described as a staple of nineteenth century math; a glance at any table of mathematical equations will explain why. The facts we use are gathered from Abramowitz and Stegun [1, Chapter 15] and Lebedev [13].

The hypergeometric function is very easy to compute for $|q| < 1$. The following algorithm, taken from Press et al. [21], will compute ${}_2F_1(a,b;c;q)$ for $a > 0$, $b > 0$, $c > 0$ and $0 < q < 1$ to machine accuracy.

```
Initialize: f = 1,  g = 1,  i = 1
do
      g = g *  q * a * b / c / i
      f = f + g
      a = a + 1
      b = b + 1
      c = c + 1
      i = i + 1
while g > 0
return f
```

Because the series defining ${}_2F_1$ is absolutely convergent it can be differentiated term by term, giving

$$\frac{dF}{dq} = \frac{ab}{c}\,{}_2F_1(a+1, b+1; c+1; q) \tag{63}$$

and more generally

$$\frac{d^nF}{dq^n} = \frac{(a)_n(b)_n}{(c)_n}\,{}_2F_1(a+n, b+n; c+n; q). \tag{64}$$

Euler's integral representation of ${}_2F_1$ is

$${}_2F_1(a, b; c; q) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tq)^{-a}dt \tag{65}$$

[1, Chapter 15.3]. We will use the result

$${}_2F_1(a, b; b; q) = {}_2F_1(b, a; b; q) = (1-q)^{-a} \tag{66}$$

from [1, Chapter 15.1]. This can be seen by considering the sum of the probabilities of a negative binomial distribution.

References

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions*. Dover, ninth edition, 1970.

[2] Vijay Balasubramanian. A geometric formulation of Occam's razor for inference of parametric distributions. *Princeton Preprints*, arxiv:adap-org/9601001 v1, 1996.

[3] José M. Bernardo and Adrian F. M. Smith. *Bayesian Statistical Decision Theory*. J. Wiley and Sons, 1994.

[4] C. D. Daykin, T. Pentikainen, and M. Pesonen. *Practical Risk Theory for Actuaries*. Chapman and Hall, 1994.

[5] Enrique De Alba. Bayesian estimation of outstanding claims reserves. *NAAJ*, 6 No. 4:1–20, 2002.

[6] Pedro Domingos. The role of Occam's razor in knowledge discovery. *Data Mining and Knowledge Discovery*, 3:409–425, 1999.

[7] Peter D. England and Richard J. Verrall. A flexible framework for stochastic claims reserving. *Proceedings of the Casualty Actuarial Society*, LXXXVIII:1–38, 2001.

[8] Norman L. Johnson, Samuel Kotz, and N. Balakrishnan. *Discrete Multivariate Distributions*. J. Wiley and Sons, 1996.

[9] Norman L. Johnson, Samuel Kotz, and Adrienne W. Kemp. *Univariate Discrete Distributions*. J. Wiley and Sons, second edition, 1992.

[10] Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot. *Loss Models, From Data to Decisions*. John Wiley and Sons, 1998.

[11] Samual Kotz, Norman L. Johnson, and N. Balakrishnan. *Continuous Multivaritate Distributions*, volume 1. John Wiley and Sons, second edition, 2000.

[12] Michael Kunkler. Modelling zeros in stochastic reserve methods. *Insurance Mathematics & Economics*, 34:23–35, 2004.

[13] N. N. Lebedev. *Special Functions and Their Applications*. Dover, New York, NY, 1972.

[14] Thomas Mack. The stochastic model underlying the chain ladder method. *ASTIN Bulletin*, Vol. 21(1):22–61, 1991.

[15] Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, Vol. 23 No. 2:213–25, 1993.

[16] Thomas Mack. Credible claims reserves: The Benktander method. *ASTIN Bulletin*, Vol. 20 No. 2:333–347, 2000.

[17] Stephen J. Mildenhall. Correlation and aggregate loss distributions with an emphasis on the Iman-Conover method. *Casualty Actuarial Society Winter Forum*, 2006.

[18] Daniel M. Murphy. Unbiased loss development factors. *Proceedings of the Casualty Actuarial Society*, LXXXI:154–222, 1994.

[19] Ragnar Norberg. Prediction of outstanding liabilities in non-life insurance. *ASTIN Bulletin*, Vol. 23, No. 1:95–115, 1993.

[20] Ioannis Ntzoufras and Petros Dellaportas. Bayesian modelling of outstanding liabilities incorporating claim count uncertainty. *NAAJ*, 6 No. 1:113–128, 2002.

[21] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipies in C, The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.

[22] A. E. Renshaw and R. J. Verrall. A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4, IV:903–923, 1998.

[23] Ira Robbin. A Bayesian credibility formula for IBNR claims. *Proceedings of the Casualty Actuarial Society*, LXXIII:129–164, 1986.

[24] David A. Stephens, Martin J. Crowder, and Petros Dellaportas. Quantification of automobile insurance liability: a Bayesian failure time approach. *Insurance Mathematics & Economics*, 34:1–21, 2004.

[25] Gary G. Venter. Testing the assumptions of age-to-age factors. *Proceedings of the Casualty Actuarial Society*, LXXXIV:807–847, 1998.

[26] R. J. Verrall. Bayes and empirical Bayes estimation for the chain ladder method. *ASTIN Bulletin*, 20 No. 2:218–243, 1990.

[27] R. J. Verrall. A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims resering. *NAAJ*, 8 No. 3:67–89, 2004.

[28] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*. Cambridge University Press, Cambridge, fourth edition, 1927.

[29] Thomas S. Wright. Stochastic claims reserving when past claim numbers are known. *Proceedings of the Casualty Actuarial Society*, LXXIX:255–361, 1992.

[30] Ben Zehnwirth. Probabilistic development factor models with application to loss reserve variability, prediction intervals and risk based capital. *Casualty Actuarial Society Forum*, Spring:447–605, 1994.

[31] Ben Zehnwirth and Glen Barnett. Best estimates for reserves. *Proceedings of the Casualty Actuarial Society*, LXXXVII:245–321, 2000.

Biography of the Author

Stephen Mildenhall is an Executive Vice President at Aon Re Services in Chicago Illinois. Aon Re Services is the technical arm of Aon Re, the world's largest reinsurance intermediary. At Aon Re, Stephen heads a team of actuaries providing a variety of actuarial services for clients, including reinsurance pricing and structure evaluation, dynamic financial analysis, capital allocation and predictive modeling. Prior to joining Aon Re, Stephen worked at Kemper Insurance and CNA Insurance. He has an undergraduate degree in Mathematics from the University of Warwick, UK, and a Masters and PhD in Mathematics from the University of Chicago. Stephen is a Fellow of the Casualty Actuarial Society, an Associate of the Society of Actuaries and Member of the American Academy of Actuaries. His research interests include risk theory, capital determination, allocation and optimization, and applications of statistics to reserving and ratemaking problems.

December 8, 2005

Incorporating Systematic Risk Into The RMK Framework

Trent Vaughn, FCAS, MAAA

Abstract
The RMK pricing algorithm provides a method for pricing insurance contracts or reinsurance deals. This paper discusses the incorporation of systematic, or non-diversifiable, risk into the RMK framework.

1. A SIMPLE EXAMPLE OF THE RMK METHOD

Ruhm/Mango (2003) present a simple illustration of the RMK pricing algorithm. Specifically, this simple example assumes that the insurance (or reinsurance) company writes two risks, each with the following state-dependent loss vector:

State	Risk 1 Loss	Risk 2 Loss	Portfolio Loss	Probability
1	$100	$100	$200	35%
2	$100	$200	$300	15%
3	$200	$100	$300	25%
4	$200	$200	$400	25%

The RMK algorithm incorporates an adjustment for risk by means of a set of *outcome-specific weights*. For this example, Ruhm/Mango utilize the following set of risk-averse outcome weights:

Portfolio Outcome	Risk-Averse Outcome Weight
$200	0.500
$300	1.000
$400	1.250

These risk-averse outcome weights are similar to Mango's (2003) concept of a *cost function*. Mango points out that such a function can be interpreted as a corporate utility function; that is, in some sense, management has determined that a $300 aggregate loss is "twice as bad" as a $200 aggregate loss.[1]

These risk-averse weights are then normalized (scaled so that their expected value is one) to produce the following vector of normalized weights:

[1] Fama and Miller (1972) point out the many theoretical difficulties involved in interpreting and determining a "corporate" utility-of-wealth function. However, for purposes of this paper, we will assume that such a function has been determined by some means.

Portfolio Outcome	Normalized Weight
$200	0.563
$300	1.127
$400	1.408

The RMK method then determines the "risk load" for each of the two risks according to the following formula:

Risk Load for Risk i = Cov(Ri,Z),

where Ri is the loss amount for each of the two risks (i = 1, 2),
and Z is the vector of normalized weights.

Thus, the resulting risk load is $13.38 for Risk 1 and $12.11 for Risk 2. The final premium is then determined by discounting the expected loss for each risk (at the risk-free rate of interest), then adding the risk load. Assuming that losses are payable at the end of one year, and a risk-free interest rate of 2%, the final RMK premiums are as follows:

Premium for Risk 1 = $150/1.02 + $13.38 = $160.44
Premium for Risk 2 = $140/1.02 + $12.11 = $149.36

2. THE FINANCIAL PRICING METHOD

By comparison, let's utilize a financial pricing method to price each of the risks in the previous example. If we ignore default costs, then the financial premium formula reduces to the following equation:

Premium = Present Value of Expected Loss (at risk-adjusted rate) + Capital * Cost of Capital

According to the Capital Asset Pricing Model (CAPM), the risk-adjusted discount rate for the loss amount depends on the relationship between the loss random variable and the return on the market portfolio. Let's assume the following state-specific returns (Rm) for the market portfolio:

State	Return on Market Portfolio (Rm)
1	+25%
2	+10%
3	+4%
4	-5%

For each of the two risks, the present value of the expected loss at the risk-adjusted rate, or PV(Ri), is determined according to the certainty-equivalent version of the CAPM:[2]

PV(Ri) = E(Ri) / (1 + Rf) – [Lambda * Cov(Ri,Rm)] / (1+Rf),

Where E(Ri) is the expected loss amount for each of the two risks,
Rf is the risk-free rate of interest,
And Lamda is the "market price of risk" given by:
Lambda = [E(Rm) – Rf] / Var(Rm)

According to our assumptions regarding the return on the market portfolio, we calculate the following values:

Cov(R1,Rm) = -5.25
Cov(R2,Rm) = -3.75
Lambda = 5.56
PV(R1) = $175.65
PV(R2) = $157.68

In the financial formula, the "cost of capital" is primarily due to double taxation and agency costs. Let's arbitrarily assume that the cost of capital is 10% of the required capital. The required capital for each policy is generally determined by allocating the total capital down to the risk, or policy, level. In the financial method, this allocation method is generally based on some form of Option Pricing Theory (OPT). However, for simplicity, let's assume that a total capital amount is $200, and that it will be allocated in proportion to the expected loss amount for each risk. The premium for each risk is then given as follows:

Premium for Risk 1 = $175.65 + 10% of $103.45 = $186.00
Premium for Risk 2 = $157.68 + 10% of $96.55 = $167.33

3. EXPLAINING THE DIFFERENCES BETWEEN THE RMK AND FINANCIAL PREMIUMS

In the previous two sections, the Financial method resulted in a much higher required premium for each policy than the RMK method. There are two major reasons for this discrepancy.

First, the RMK method requires a "calibration" to ensure that the resulting combined ratio and return on equity are in accordance with the overall corporate objectives. Mango (2003) discusses the issue of calibration in detail, but the procedure is outside the scope of this paper. Presumably, the overall return implied by management's risk-averse outcome weights would determined; if this overall return falls short of corporate targets, there would need to be a feedback loop back to management to adjust the weights. The

[2] See the Appendix to Chapter 9 of Brealey and Myers (2000) for a derivation of this formula.

procedure would continue until a set of weights had been identified that resulted in an acceptable corporate return.

Second, the financial method incorporates additional data and assumptions regarding the state-return on the market portfolio. In other words, the financial method incorporates the "systematic risk" of the loss variables, whereas the RMK method did not. Since these loss variables possessed a negative covariance with the market return, the incorporation of systematic risk resulted in an increase in the required premium.

This begs the question: can we adjust the RMK method from Section 1 to reflect the market return data – and the "systematic risk" of the loss variables? This question will be explored in the following section.

4. A METHOD FOR REFLECTING SYSTEMATIC RISK IN THE RMK ALGORITHM

Mango (2004) presents a simplified flow-chart method for incorporating systematic risk into the RMK framework. Essentially, the method combines the results of the insurer's underwriting portfolio and the insurer's asset portfolio to produce a state-specific *net income* distribution. This net income distribution then serves as the *reference portfolio* for the RMK application.

In order to determine this net income distribution, we need to develop some assumptions regarding the insurer's investment (or asset) portfolio. For this example, let's assume that 80% of the insurer's assets are invested in risk-free bonds, earning the risk-free rate of 2%; the remaining 20% of the insurer's assets are invested in the market portfolio, earning the state-specific returns provided in Section 2.

Since we are now dealing with net income, management's risk preferences must be stated in terms of various net income amounts (as opposed to aggregate loss amounts). Let's assume that management has developed the risk-averse outcome weights as a function of various net income amounts. Again, there is an intuitive interpretation of this risk aversion function.[3] For instance, let's say that the outcome-specific weight is 1.25 for net income of $50 and 0.25 for net income of $150; in this sense, management views a net income result of only $50 as being "five times as bad" as a higher net income result of $150.

In this case, the RMK method requires an iterative approach, since the resulting premium amount impacts both the underwriting income and the investment income.[4] With the asset allocation assumptions above – together with some assumed values for the risk-averse outcome weights -- the resulting premium is $171.77 for Risk 1 and $160.47 for Risk 2. The following chart and formulas provide the details of the calculation:

[3] And, again, we will ignore the theoretical and practical difficulties involved in determining this function.
[4] Investment income is impacted since total assets are equal to total premium plus total surplus.

State	Probability	Aggregate Loss	Market Return	Net Income	Mgt. Risk Weight	Normalized Weight (Z)
1	0.35	$200	25%	$167.37	0.25	0.230
2	0.15	$300	10%	$51.41	1.25	1.149
3	0.25	$300	4%	$45.02	1.25	1.149
4	0.25	$400	-5%	-$64.56	2.00	1.839

Risk Load for Risk 1 = Cov(R1,Z) = $24.71
Risk Load for Risk 2 = Cov(R2,Z) = $23.22
Premium for Risk 1 = $150/1.02 + $24.71 = $171.77
Premium for Risk 2 = $140/1.02 + $23.22 = $160.47

Also, it may be helpful to illustrate the calculation of the net income amount for state 1. In this state, the income variables are as follows:

Underwriting Income = Total Premium – Aggregate Loss = $332.24 - $200 = $132.24
Total Assets = Total Premium + Surplus = $332.24 + $200 = $532.24
Assets Invested in Market Portfolio: 20% of $532.24 = $106.45
Assets Invested in Bond Portfolio: = 80% of $532.24 = $425.79
Investment Income from Market Portfolio = 25% return on $106.45 = $26.62
Investment Income from Bond Portfolio = 2% return on $425.79 = 8.52
Total Income[5] = $132.24 + $26.62 + $8.52 = $167.38

Net income for the other states is determined in a similar manner.

5. POTENTIAL PROBLEMS WITH THE MANGO ADJUSTMENT FOR SYSTEMATIC RISK

In some sense, the method in Section 4 does provide an adjustment for systematic risk, since the insurer's net income depends (to a certain extent) on the return on the market portfolio. However, the sensitivity of the insurer's net income to the market return will depend on the insurer's asset allocation. For example, if the insurer is invested entirely in risk-free bonds, then net income will be unaffected by market return.

Moreover, in a practical situation, the insurance company invests in many more asset types than simply a "market portfolio" and risk-free bonds. Insurers may invest in corporate bonds, some sampling of common and preferred stocks, real estate, etc. In addition, the insurer's common stock portfolio may not be fully diversified, but invested in only a handful of individual stock holdings. In this case, the net income approach will reflect the risk characteristics of the insurer's asset portfolio, but it would be incorrect to say that it has "incorporated systematic risk" into the analysis.

[5] This is actually total income prior to federal income taxes. We are ignoring federal income taxes in this example.

As an alternative, we may wish to calculate the insurer's net income distribution by fully utilizing the return on the representative market portfolio -- that is, make the assumption that the insurer is 100% invested in the market portfolio. This approach is still subject to the following drawbacks:

1. Instead of using the assumption of 100% in the market portfolio, we could have used some other hypothetical mixture, such 75% in the market portfolio and 25% in risk-free bonds. It isn't clear which representative mixture best incorporates "systematic risk" into the net income distribution. And, in general, the resulting risk loads (and premiums) will vary on the basis of the assumed allocation.
2. It becomes much harder to provide any intuitive meaning to the *risk-averse outcome weight*. The subject of these outcome weights is now a complicated intermingling of the market return volatility and the insurance portfolio volatility – and may bear little resemblance to the actual net income result for the insurance company in any particular state.
3. There are already a variety of financial approaches for reflecting the systematic risk of a cash flow (e.g. CAPM, APT, Fama-French Three Factor Models). These models are not based on judgmental assessments of management's risk preferences, but financial theories regarding equilibrium in capital markets. By combining the adjustments for systematic risk and insurance risk into one step, we are not able to utilize these financial theories regarding systematic risk.

It is possible, in theory, to determine a set of risk-averse outcome weights for the RMK procedure that will duplicate the premiums from the financial model.[6] This, however, provides little guidance to the actuary who is pricing a reinsurance deal "from scratch". That is, assuming that the answer is not known in advance, the pricing actuary must determine a set of risk-averse outcome weights from a reference portfolio that has little (if any) intuitive or practical meaning.

6. AN ALTERNATIVE METHOD FOR INCORPORATING SYSTEMATIC RISK INTO RMK

As an alternative to the method in Section 4, we can accommodate systematic risk within an RMK framework simply by discounting the expected losses at a risk-adjusted discount rate. In other words, simply utilize the RMK risk loads from Section 1, but adjust the discount rate for the losses in accordance with financial theory.

For instance, according to the certainty-equivalent version of the CAPM, the present value of expected losses for each of the two risks was given as follows (per Section 2):

PV(R1) = $175.65
PV(R2) = $157.68

[6] Assuming that the surplus allocation in the financial model is additive, which it is in this case.

According to the RMK method of Section 2, the risk loads (excluding systematic risk) were $13.38 for Risk 1 and $12.11 for Risk 2. By adding these risk loads to the present value (at the risk-adjusted rate) of losses, we get the following premiums:

Premium for Risk 1 = $175.65 + $13.38 = $189.03
Premium for Risk 2 = $157.68 + $12.11 = $169.79

By using this method, we can also "bridge the gap" between the financial method and the RMK method. Specifically, in the financial method in Section 2, we allocated capital in proportion to expected losses. As an alternative, let's allocate capital in accordance with the risk-averse outcome weights assigned to the various aggregate loss amounts in Section 1. In other words, allocate capital to each risk in proportion to that risk's relative contribution to the covariance between the aggregate loss outcome and the Z-vector. For instance, the percentage of capital allocated to Risk 1 is Cov(R1,Z) / Cov (Aggregate Portfolio Outcome,Z) = $13.38 / $25.49 = 52.5%.[7] According to the Z-vector from Section 1, the $200 capital would then be allocated at $104.99 for Risk 1 and $95.01 for Risk 2. The financial premiums then become:

Risk 1 Premium = $175.65 + 10% of $104.99 = $186.15
Risk 2 Premium = $157.68 + 10% of $95.01 = $167.18

Lastly, the final reconciliation issue is simply a problem of "calibration" (per the terminology in Mango). That is, there is no reason to expect that the total capital ($200) and cost of capital (10%) in the financial model will produce the same ROE as the RMK method. But, on an individual policy level, the ratio between the premiums for each risk is the same. Thus, we can complete the reconciliation by changing either the total capital or the cost of capital in the financial model. Let's change the cost of capital to 12.75%, which will complete the reconciliation:

Risk 1 Premium = $175.65 + 12.75% of $104.99 = $189.03
Risk 2 Premium = $157.68 + 12.75% of $95.01 = $169.79

7. SUMMARY

This paper has presented two proposed methods for incorporating systematic risk into the RMK pricing algorithm. The Mango (2004) method is plagued by an assortment of theoretical and practical problems. In short, the best method for incorporating systematic risk into the RMK framework is simply to discount the expected losses at a risk-adjusted hurdle rate. This risk-adjusted rate can be determined by any one of the common financial pricing models, including the CAPM, the APT, or the Fama-French Three Factor Model.

[7] Also, note that this is just the ratio of each of the individual risk loads to the total risk load.

REFERENCES

Brealey, Richard A., and Myers, Stewart C., *Principles of Corporate Finance*, 6th Edition, McGraw-Hill, 2000.

Fama, Eugene F., and Miller, Merton H., *The Theory of Finance*, Holt, Rinehart & Winston, 1972.

Mango, Donald F., "Capital Consumption: An Alternative Method for Pricing Reinsurance," *Winter Forum*, Casualty Actuarial Society, 2003.

Mango, Donald F., Unpublished Power Point Presentation, 2004.

Ruhm, David, and Mango, Donald F., "A Risk Charge Based on Conditional Probability," Bowles Symposium, 2003.

A Portfolio Theory of Market Risk Load

Yingjie Zhang, PhD, FCAS

Abstract

In insurance pricing, it is convenient to split the total risk load for a policy into the market risk load and the insurer specific risk load, and calculate each separately. The market risk load represents an equilibrium price on a competitive insurance market. A portfolio theory is developed along the line of the classic CAPM, where a policy's market risk load is a function of its systematic risk and the risk load of the entire insurance market. The model is mathematically proved. As a corollary a formula for the risk adjusted discount rate is obtained. Issues about the real world application and testing are also discussed.

1 Introduction

Risk load calculation is important in insurance princing. As long as risk is transfered in an insurance transaction, a risk load should be included in the premium. The purpose of the risk load is to reward the insurers for taking the insurance risk. An insured pays a certain amount of premium to eliminate the uncertainty in future loss costs, and an insurer collects the premium and assumes the responsibility of paying any claims. Since both the insured and the insurer are risk averse, the insured is willing to pay a premium greater than the expected loss, and the insurer needs that additional premium to justify taking the risk. The size of the risk load depends on the riskiness of the insured loss and the competition on the insurance market.

In the actuarial literature the calculation of risk load has experienced considerable change. In the classic premium principles, a risk load is determined by the volatility of the insured loss itself, and the volatility is measured by the variance or the standard deviation [15]. Although these methods are still used, they have been considered inadequate. As pointed out in Feldblum [8], they measure the

insured's risk but not the insurer's risk, inconsistant with the purpose of risk load. More reasonable risk load formulas were proposed in [8] and [12], which took account of not only the volatility of the policy loss but also the company insurance portfolio and the market competition. These articles were inspired by the modern financial theory, especially the Capital Asset Pricing Model (CAPM). Under the assumptions that the insurance market is competitive and the market players are "rational" decision makers, supply and demand determine an equilibrium risk load. These methods better reflect the insurer perspective of risk loads. They are among the first attempts to extend the modern financial theory to insurance.

A recent COTOR review article [6] lays out a framework for the study of all risk components in premium. Underwriting risks come from various sources. Risks resulting from the uncertainty in an insured loss and the economic conditions of the insurance market do not rely on the particular insurer with which the policy is insured. The frictional cost of capital, on the other hand, is one of the risk items related to the capital structure of a particular insurer. So it is natural to split the insurer total risk load into two classes, the market risk loads and the insurer specific risk loads, which may be calculated separately. The following split is given in [6]

$$\begin{aligned}\text{premium} \;=\; &\text{expected loss} + \text{market risk premium} \\ &+ \text{risk management cost} + \text{expected default} + \text{expenses}.\end{aligned}$$

The total risk load consists of the second and the third term on the right hand side. (The expected default is a reduction to premium, so is not considered part of the risk load.) The market risk premium is just another name for the market risk load. The risk management cost includes all risks steming from an insurer's holding capital.

The market risk load is the subject of this paper. (Here the word "market" means the insurance market, not the total financial market.) The market equilibrium approach, whose power has been demonstrated repeatedly in modern financial theory, will be employed to derive a risk load model. (The same approach, however, seems less effective in studying the insurer specific risk load, since companies have different line-of-business composition and different capital adequacy.) The paper is structured as follows. In Section 2, we examine the concept of market

risk load in detail. The market risk load is determined by a market equilibrium where no arbitrage opportunities exist. It is not related to the line-of-business composition and the capital amount in any particular insurance company. Section 3 reviews various risk load models in the literature. We focus on the market equilibrium approach and the CAPM-related models. The CAPM idea seems widely applicable. But for a model to work it is necessary to reexamine the assumptions and perferrably provide a mathematical proof. Each risk load model by and large reflects one of the two pricing views: the actuarial view and the financial view. The former addresses the risk/return of the insurance companies, and the latter that of the shareholders.

In Section 4 we develope a portfolio theory for the market risk load. The derivation is parallel to the CAPM. The market risk load for a policy is a function of its systematic risk, defined in line with the β parameter in CAPM. The risk load is also in proportion to the overall market risk load, so is influenced by the level of competition on the insurance market, and in particular, by the underwriting cycles. A corresponding equation for the risk adjusted discount rate is derived in Section 5. Just like the CAPM, our model may not be a perfect fit in the real insurance market. In Section 6 we discuss what may happen when some of the theoretical assumptions fail. Modifications seem necessary to obtain more realistic models. Empirical testing of this or any other insurance models is difficult, due partly to the settlement lag and the data limitation. Finally, a mathematical proof is given in the appendix.

2 Market Risk Load

Market risk loads represent equilibrium prices in a competitive market. To develope a theory for the market risk load, we assume there exists an ideal insurance market. Insureds and insurers are risk averse. Insureds pay a premium to transfer their future uncertain loss to the insurance market. They are willing to pay a risk load in addition to the expected loss. The size of the risk load is commensurate with the risk transfered. On the other hand, insurers enter the insurance market to make a profit. They accept a premium, invest the proceeds in the financial market, and pay any claims. Because of the uncertainty of the future loss, an insurer demands a risk load over and above the expected loss. In a competitive market,

the insureds shop around to pay the lowest possible risk load, while the insurers collect the highest possible risk load from each policyholder and select the policies to minimize the total insurance risk. Further assume the market is efficient, so that insureds and insurers have perfect information regarding the expected loss and the risk of any policy, and they can easily access the entire market. Under these conditions there exists an equilibrium risk load for each policy. This is the market risk load.

The real insurance market has inadequate competition and efficiency. The insureds do not have sufficient information about price, so they may not find the lowest one. Insurers are limited by underwriting expertise and regulation, so they only write a few business lines and charge noncompetitive rates. Besides, without a frictionless trading mechanism, it is not possible to reach the equilibrium prices. Nevertheless, the market risk load is still a useful concept. It represents a fair premium to both insureds and insurers. It may not be reached, but can be unboundedly approached with improvement in market competition and efficiency. In a market segment where risk securitization is in place, the market risk load may be practically realized. CAT call spreads and bonds are examples of successful securitization.

The market risk load avoids the consideration of line-of-business composition and capital structure of a particular insurer. (In other words, we imagine "abstract" insurers that have unlimited and costless access to capital. They are able to minimize the total insurance risk by diversification, and they charge risk loads only to cover the uncertainty risk in the claims.) This allows a portfolio theory to be developed. On the other hand, the insurer specific risk varies with a different set of risk factors. The frictional cost of capital is one important component of the insurer specific risk, examples of which include taxation and agency costs. [23] gives a detailed analysis of the frictional cost. Premium charge for the frictional cost is a function of the capital amount allocated to the individual policies. Recent development in capital allocation includes [17], [20] and [27]. In practice, many companies also charge policyholders additional premium to compensate for their more risky line-of-business composition. Large and multiline insurers have a higher degree of diversification, so demands relatively lower risk loads, while small and monoline insurers require higher risk loads. It seems unreasonable to

charge the policyholders for an insurer's own inefficiency. Yet since the actual competition on the market is inadequate, companies are able to obtain this extra premium from unknowing policyholders. In actuarial literature, quantification of insurer specific risk loads is less studied. The market equilibrium approach seems powerless here.

In the rest of the paper we omit the insurer specific risk load, and focus only on the market risk load. The term "risk load" and "market risk load" may be used interchangeablly. We also ignore all expenses. Therefore, the premium has the following expression

$$\text{premium} = \text{expected loss} + \text{market risk load}.$$

Venter [22] discusses constraints imposed on premium in a competitive market, where any arbitrage activity must be short-lived. In equilibrium state the market is arbitrage-free. A necessary condition for an arbitrage-free market is that the premiums are additive, meaning that the total premium for a group of policies, whether independent or not, equals the sum of the individual premiums. This implies that the market risk loads are additive. Notice that when the insurer specific risk loads are included, the total risk loads do not have the additive property. Because of the diversification effect, the insurer specific risk load of a portfolio is likely to be lower than the sum of that of the individual policies. (It makes sense, however, for the total risk load to be additive within an insurance company.) [6] also has an interesting discussion on additivity.

Diversification is an important concept in modern financial theory. There are many forms of diversification in the insurance world. The market risk loads provide a simple one. When policies are combined into a portfolio, the portfolio risk load is the sum of the individual risk loads. However, as long as the policy losses are not perfectly correlated, the risk of the portfolio, represented by the standard deviation or other reasonable measures, is less than the sum of the individual risks. So it is to an insurer's advantage to write a large volume of multiline insurance portfolio. Greater diversification effect may be achieved by insuring many negatively correlated risks.

3 Review of Risk Load Models

On the surface the insurance market is analogous to the securities market. The insurance policies are like the securities, and the insurers the investors. An insurer's charging a risk load is similar to an investor's demanding a risk premium for a risky asset. Therefore, there is a great temptation in applying the securities pricing techniques to the insurance pricing.

Much research has done to extend the classic CAPM to risk load calculation. Among the P&C actuaries, the Feldblum article [8] was influential and inspired a great deal of discussion. It provides a CAPM-like model to calculate the risk loads by line. It argues that the CAPM has many advantages over other methods like the standard deviation, the probability of ruin, or the utility functions. However, as commented later on ([13] [21]), [8] contains some conceptual difficulties and the risk load formula is not convincing. One significant conceptual flaw in [8] is that it "simply borrows the CAPM notation while ignoring the underlying message of the CAPM paradigm" [21]. This subtle and important point warrants further explanation.

A basic CAPM assumption is that the investors are risk averse. They select the securities to maximize the portfolio return and minimize its risk. The selection process by many small investors produces a market equilibrium where the security returns are given by the classic CAPM. The CAPM is intuitively appealing and can be mathematically proved. It is also extensively tested with empirical data. Many modifications are proposed in response to the unfavorable test results. The current status of the issue is summarized well in [6] and [1]. The argument in [8], however, ignores the shareholders of the firm and the returns required by the financial market. In that setting the classic CAPM is not applicable. [8] replaces the investor/security pair by the firm/line-of business-pair, and restates the CAPM in terms of the latter. Without carefully examining the CAPM assumptions or providing a mathematical proof, this approach becomes simply "borrowing notation", which often leads to erroneous results. In a different context, Mildenhall [16] spells out the error of borrowing notations from the option pricing paradigm to the insurance pricing.

The classic CAPM is a cornerstone of the modern financial theory. Its eco-

nomic implication extends far beyond the formula itself. Even in situations the model is not directly applicable, its insights may still prove useful. Meyers [12] provides a risk load formula using the frequency and severity. The formula is derived along the line of the CAPM, from the equilibrium in a competitive insurance market. It is used by ISO in the calculation of the increased limit factors. The risk load problem is closely related to that of the risk adjusted discount rate. Butsic [3] derives a formula for the risk adjusted discount rate that looks similar to the CAPM. While in the classic CAPM the risk adjustment usually increases the rate of return (a positive β), the risk adjustment in a discount rate formula is negative, decreasing the discount rate for uncertain losses. (More discussion on this in Section 5.) Kulik [11] reviews many other CAPM related insurance applications.

The CAPM is based on the mean-variance optimization. The market risk load is also studied using other utility functions. The Bühlmann *economic premium* principle is one example [2]. The economic premium is equivalent to the market risk load. If P is an economic premium for an insured loss X, then $P - E(X)$ is the market risk load in our definition. [2] uses an exponential utility function. [25] contains some new development.

Venter [22] proposes two risk load principles satisfying the additive condition: the covariance principle and the adjusted distribution principle. Our portfolio theory is an example of the former. The adjusted distribution principles have been studied extensively. Two of the well-known adjustments are the PH-transform and the Wang transform [24] [26]. An adjusted distribution readily produces risk loads for multiple coverage layers, which are consistent in the following sense: a higher layer always has a higher risk load relative to the expected loss in that layer. Butsic [4] calculates the risk loads for excess layers using a generalized PH-transform. Usually the transforms contain one or more parameters to be determined according to the market conditions. It may be able to use a market risk load principle, such as developed in [2] or in this paper, to parameterize a transform. [25] is insightful in this regards.

The COTOR [6] distinguishes two views of the pricing paradigm. The actuarial view assumes the insurers are risk averse. They make underwriting selections and actively manage the risk/return of their insurance portfolio. The financial

view looks at the broader financial market. The shareholders of the insurance companies are risk averse. They choose to invest in the stocks of the insurance companies as well as other industries according to a preset utility function. In other words, the actuarial price is determined by the insurers with the insureds' fairness in mind, while the financial economic price is set on the market of all financial assets. The classic CAPM, being investor focused, has been used to build financial pricing models [5] [7]. In contrast, the economic premium principle [2] is purely actuarial. The goal of Feldblum [8] is also to construct an actuarial pricing model.

It is pointed out in [6] that the two insurance pricing views are converging. But so far they are still separate for the most part. The financial models ignore the mutual selection between the firms and the policyholders. The actuarial methods address the mutual selection but pay little attention to the shareholder welfare. The two theories complement each other in pricing practice. In the following section, we develope a portfolio theory within the actuarial pricing paradigm. Unlike [8], we price for the market risk only. It is necessary to limit our scope to derive a precise result. The model is similar to the classic CAPM. But it is about the insurer/insured relationship instead of the investor/security relationship in the CAPM.

4 A Portfolio Theory

We derive a risk load formula parallel to the classic CAPM. Our presentation follows a standard text book [18] (Chapter 8). The setting and the result are confined to the basic form.

Consider a one-period model where policies are written and premiums are collected at time 0 and losses are paid at time 1. At time 0, a loss payment at time 1 is viewed as a random variable. Assume at time 0 the market has complete knowledge of the random losses. In the context of mean-variance analysis, this means all market players know the mean, the variance and the covariance of all policy losses.

Assume an insurance market contains N policies with random losses $X_1, \ldots, X_N$,

which will be paid at time 1. The total market loss is thus a random variable $X^M = X_1 + \cdots + X_N$. Assume the market premium for policy i is P_i, which is charged at time 0. Then the total market premium is $P^M = P_1 + \cdots + P_N$. Further assume there is a risk free asset with rate of return r_f. So an insurer collects premium P_i, invests it in the risk free asset, receives $P_i(1 + r_f)$ at time 1 and pays any claim. The rate of return on premium is

$$R_i = \frac{P_i(1 + r_f) - X_i}{P_i} = r_f + \frac{P_i - X_i}{P_i},$$

where the first term is the investment rate of return and the second the underwriting rate of return. The mean and the covariance of the random returns are

$$\mu_i = E(R_i) = (1 + r_f) - \frac{E(X_i)}{P_i}, \tag{4.1}$$

$$\sigma_{ij} = Cov(R_i, R_j) = Cov\left(\frac{X_i}{P_i}, \frac{X_j}{P_j}\right) = \frac{1}{P_i P_j} Cov(X_i, X_j). \tag{4.2}$$

Now assume an insurer is allowed to insure any fraction of a policy, as in quota share treaties, and an insurer can borrow and lend any amount at the risk free rate. An insurance portfolio thus consists of a_i portion of loss X_i and a borrowed amount w, where $0 \leqslant a_i \leqslant 1, i = 1, \ldots, N$ (more on this condition in the appendix), and w may be positive or negative. $a_i P_i$ is the premium charge for insuring loss $a_i X_i$. A negative w means an amount of $|w|$ is lended.

At time 0, the portfolio has a total asset equal to $w + \sum_{i=0}^{N} a_i P_i$. When w is negative, assume $|w|$ is small so that $w + \sum_{i=0}^{N} a_i P_i > 0$. (An insurer can lend no more than its collected premium.) The asset is invested risk free and receives a rate of return r_f. At time 1, a loss $\sum_{i=1}^{N} a_i X_i$ is paid and the borrowed amount returned together with an earned interest. So the rate of return of this portfolio is

$$\begin{aligned} R_{\text{portfolio}} &= \frac{(w + \sum_{i=1}^{N} a_i P_i)(1 + r_f) - \sum_{i=1}^{N} a_i X_i - w(1 + r_f)}{w + \sum_{i=1}^{N} a_i P_i} \\ &= \frac{\sum_{i=1}^{N} a_i (P_i(1 + r_f) - X_i)}{w + \sum_{i=1}^{N} a_i P_i} \\ &= \frac{\sum_{i=1}^{N} a_i P_i R_i}{w + \sum_{i=1}^{N} a_i P_i}. \end{aligned} \tag{4.3}$$

Let us examine this setup. The return $R_{\text{portfolio}}$ is essentially a return on premium,

except for the amount w in the denominator. The return on premium is a reasonable measure of the insurer profit. In a competitive insurance market, not only the insurers select insureds, but the insureds choose among the insurers as well. The mutual selection mechanism forces the market to attain an equilibrium such that no insurer is allowed an excessive return on premium, no matter what initial wealth (capital) the insurer has. Each insured is charged an amount of premium commensurate to its market risk. Ignoring the capital structure of the insurer is both necessary and reasonable in studying the market risk loads. The inclusion of an amount w in the portfolio asset is needed for a closed form solution. An insurer should be allowed to use borrowing and lending to adjust its risk and return relationship. Lending at the risk free rate is practically achievable, but borrowing at the rate is less realistic. A similar issue also appears with the classic CAPM.

The mean and the variance of the portfolio return are

$$\begin{aligned}\mu_{\text{portfolio}} &= E(R_{\text{portfolio}}) \\ &= \frac{1}{w + \sum_{i=1}^{N} a_i P_i} \sum_{i=1}^{N} a_i P_i \mu_i \,, \qquad (4.4) \\ \sigma^2_{\text{portfolio}} &= Var(R_{\text{portfolio}}) \\ &= \frac{1}{(w + \sum_{i=1}^{N} a_i P_i)^2} \sum_{i,j=1}^{N} a_i a_j P_i P_j \sigma_{ij} \,. \qquad (4.5)\end{aligned}$$

We seek insurance portfolios that have a maximum $\mu_{\text{portfolio}}$ for a given $\sigma_{\text{portfolio}}$, or a minimum $\sigma_{\text{portfolio}}$ for a given $\mu_{\text{portfolio}}$. These are called the *efficient portfolios*. More formally, a portfolio is efficient with respect to a given $\tau \geqslant 0$ if the following quantity is maxmized

$$2\tau\, \mu_{\text{portfolio}} - \sigma^2_{\text{portfolio}} \,. \qquad (4.6)$$

The number τ represents the risk preference of an insurer. Notice that if a portfolio is efficient then a multiple of the portfolio is also efficient. This is easily seen since multiplying $a_1, \ldots, a_N$ and w by the same positive number does not change either $\mu_{\text{portfolio}}$ or $\sigma^2_{\text{portfolio}}$.

The mean-variance criterion (4.6) was used in the classic CAPM. It is also applicable in our setting. The variance captures the volatility risk of a firm. (The volatility is a significant risk. Reference [6], p.190, argues that volatility in earnings is harmful because of increased tax liability, reduced opportunity of benefiting

from deductions, and more costly funds from investors.) In addition, if a portfolio consists of a large number of policies, its return is approximately symmetrically distributed, alought the individual loss distributions are not symmetrical. (The total loss for a line of business is often modeled with a lognormal distribution. If the line is large the lognormal usually has a small CV, and the distribution is close to be symmetrical.) So the variance (or the standard deviation) is an appropriate risk measure.

Assume the $N \times N$ variance-covariance matrix $\Sigma = (\sigma_{ij})$ is positive definite. (A variance-covariance matrix is always nonnegative definite. A necessary and sufficient condition for Σ to be positive definite is that none of the linear combinations of the losses $X_1, \ldots, X_N$ is risk free. In particular, if a ground-up loss X is split into a primary loss X^p and an excess loss X^e, then either X or the pair X^p and X^e may be included in the model, but not all three.) If all insurers make rational decisions so that each chooses an efficient insurance portfolio, according to its own risk preference, then the following equation holds

$$P_i - \frac{E(X_i)}{1+r_f} = \frac{Cov(X_i, X^M)}{Var(X^M)} \cdot \left(P^M - \frac{E(X^M)}{1+r_f} \right). \tag{4.7}$$

This is our model for the market risk load. $P_i - E(X_i)/(1+r_f)$ is the risk load (at time 0) for the ith policy and $P^M - E(X^M)/(1+r_f)$ the overall market risk load. The appearance of the factor $1+r_f$ in the formula is because X_i is valued at time 1 while P_i is at time 0. $(X_i/(1+r_f)$ is called in [9] the (random) present value of X_i.) The equation will be proved in an appendix. The fact that all insurers choose efficient portfolios implies that the entire insurance market portfolio is efficient. Equation (4.7) actually follows from the efficiency of the market portfolio.

Equation (4.7) looks similar to the CAPM, and its proof is parallel to that of the CAPM. But the difference is noticeable. The investor/security pair in the classic CAPM is replace here by the insurer/insured pair. The basic assumption in the CAPM is that the investors are risk averse, and they select securities to minimize the risk for a given return. Here in the market risk load theory the shareholder is ignored. The insurers are assumed risk averse. They manage underwriting results and take risk control measures to minimize the total risk contained in the insurance portfolio.

As discussed in Section 3, there are two distinct views of the insurance pricing. The classic CAPM is a basis of the financial pricing approach, while the above model takes an actuarial point of view. It has been noticed that the two pricing views are not entirely consistent [6] [21]. Since the shareholders can easily select the securities and diversify their investment portfolio, they do not require the company to mitigate its risk. And the risk control is undesirable because it is always costly. But in practice, risk control and underwriting supremacy are among the very goals of the company management. With the help of recent development in Dynamic Financial Analysis, it becomes more probable to optimize the insurance portfolio, to improve the reinsurance structure, or to make more efficient use of the company capital. This apparent contradiction is explained in [6]. Because of imperfection in the financial market, it costs the shareholders if a company experiences financial distress or excessive profit volatility. Company value "will increase as long as the costs associated with the practice of risk management do not exceed the benefits of the risk management program" [6]. Neither the finanical view nor the actuarial view alone gives a complete picture of the insurance price. Integration of the two sides appears to be a challenging task.

The overall market risk load in (4.7) is usually positive due to risk aversion. For most policies, the random loss is positively correlated with the overall market, so the risk load is positive. The model provides an *economic* risk load in the sense of Bühlmann [2]. The risk load reflects not only the risk of the loss itself but also the market conditions. General economic environment and the level of competition on the insurance market are reflected in the overall risk load $P^M - E(X^M)/(1+r_f)$. An underwriting cycle is just a cyclic change in the overall risk load. P^M is high when the market is "hard", and is low when it is "soft". Model (4.7) states that a change in $P^M - E(X^M)/(1+r_f)$ causes a proportional change in the risk load of an individual policy. The overall market has a higher influence on an individual risk load if the correlation is high.

As in the investment theory, it is the covariance $Cov(X_i, X^M)$, rather than the variance or the standard deviation of X_i, that determines the risk load of X_i. Each X_i can be split as follows

$$X_i = X_i^{\text{sys}} + X_i^{\text{uns}},$$

where $X_i^{\text{sys}} = Cov(X_i, X^M)/Var(X^M) \cdot X^M$ is the *systematic* component and $X_i^{\text{uns}} = X_i - X_i^{\text{sys}}$ the *unsystematic* component. It is easy to verify that $Cov(X_i^{\text{uns}}, X^M) = 0$. Equation (4.7) implies that X_i^{uns} has no impact on the risk load of X_i. An unsystematic component may increase the total risk of a policy (calculated with the standard deviation or other risk measures), but does not warrant an additional risk charge, since it can be diversified away. In reality, however, diversification is not achieved in a single insurance company. It can only be done in the entire market. Increasing the premium volume and including more classes and territories, a company may attain a higher level of diversification. A small or monoline insurance company has a competitive disadvange because its insurance portfolio contains significant amount of unsystematic risk. However, even a very large insurance portfolio has a much lower degree of diversification than an average financial market player. Main reasons include that writing a policy is much more expensive than buying a share of stock, and that the insurance risks are more numerous and more heterogeneous.

In the investment world, an unsystematic risk means that it is uncorrelated with the total financial market. In [5] and [7], the same concept is used in insurance: the part of risk contained in the underwriting profit is called unsystematic if it is uncorrelated with the total financial market. This paper focuses on the insurance market instead. We implicitly assume the aggregate impact of the broad financial market on the policy losses is incorporated in the overall market risk load $P^M - E(X^M)/(1+r_f)$. (The overall market risk load serves as a "catch-all" term.) This definition of unsystematic risk is closer to the insurance practice. Underwriters usually consider a policy's correlation with other policies rather than with investment assets. However, it is possible to generalize our model to include all financial assets. Instead of assuming the premiums grow at the risk free rate, we may allow them to be invested in any financial instruments. The theory should develope similarly.

Schnapp [19] derives a pricing model similar to (4.7) using a heuristic approach. He noticed another conceptual difference between (4.7) and the classic CAPM. Both models provide a "reward" to the risk takers commensurate with the size of the risk. The CAPM defines "risk" in terms of the uncertainty in the future stock price. But the uncertainty in price is a result of the uncertainty in

company business. So the CAPM is about a "derived" risk. On the other hand, in our model (4.7) "risk" is related to the randomness of the loss variable X_i, the "original" risk. It is a more fundamental form of risk.

Equation (4.7) is simplified if a policy is not correlated with the rest of the market. If $Cov(X_i, X^M - X_i) = 0$ then $Cov(X_i, X^M) = Var(X_i)$ and (4.7) reduces to

$$P_i = \frac{E(X_i)}{1+r_f} + Var(X_i) \cdot \frac{1}{Var(X^M)} \left(P^M - \frac{E(X^M)}{1+r_f} \right). \tag{4.8}$$

This is a classic variance principle. Thus the variance principle is economically sound if a policy is uncorrelated with the market. But it oversimplifies in general. Equation (4.8) also provides a multiplier in the variance principle, which is a function of the overall market conditions.

Model (4.7) also explains other real world observations. If X_i is a random loss of a catastrophe coverage, then the risk load is expected to be large. The classic risk load priciples would support this by reasoning that $Var(X_i)$ is large. Equation (4.7) may explain more. Since a catastrophic event may simultaneously triger many policies and multiple coverages like property, business interruption, workers compensation, life and medical, it has high correlation with the overall market. So $Cov(X_i, X^M - X_i)$ is also large. Therefore, in (4.7), $Cov(X_i, X^M) = Var(X_i) + Cov(X_i, X^M - X_i)$ is a large number, which results in a high risk load.

Notice that on the right hand side of model (4.7), $E(X^M)$, P^M and $Var(X^M)$ are all very large numbers. We may restate (4.7) in the following more manageable format.

$$\frac{P_i}{E(X_i)} - \frac{1}{1+r_f} = \beta_i \cdot \left(\frac{P^M}{E(X^M)} - \frac{1}{1+r_f} \right), \tag{4.9}$$

where

$$\beta_i = Cov\left(\frac{X_i}{E(X_i)}, \frac{X^M}{E(X^M)} \right) / Var\left(\frac{X^M}{E(X^M)} \right). \tag{4.10}$$

β_i has been called a *loss beta* in the literature, which parallels the asset beta in the classic CAPM. β_i is different from the underwriting beta in [7], Section 4.

5 Risk Adjusted Discount Rate

A risk load model directly leads to a formula for the risk adjusted discount rate. If a policy loss is certain in both amount and timing, the risk load is zero and the economic premium equals the present value of the loss discounted at the risk free rate. If the loss is uncertain, however, the premium usually includes a positive risk load, and the premium is conventionally viewed as the present value of loss discounted at a rate *lower* than the risk free rate. This rate is called a risk adjusted discount rate.

Calculation of risk adjusted discount rates has been discussed in the actuarial literature. Butsic [3] derives an equation of the following form

$$\text{risk adjusted discount rate} = \text{risk free rate} - \text{risk adjustment}.$$

The size of the risk adjustment is in direct proportion to the riskiness of the claim payment cash flow. This formula is in the same spirit as the classic CAPM. Here the risk adjusted discount rate is used to discount uncertain claim payments, a cash outflow, while the CAPM calculates a rate to discount the future cash inflow. In the above equation the risk adjustment reduces the risk free rate. The CAPM, on the other hand, produces an upward rate adjustment for risk.

We use equation (4.9) to calculate the risk adjusted discount rate. By definition, in our one-period model, a discount rate for X_i is a rate r_i satisfying

$$P_i = \frac{E(X_i)}{1+r_i}.$$

A similar equation holds for the overall market discount rate r^M. Subtituting these into (4.9) we have

$$\frac{1}{1+r_i} - \frac{1}{1+r_f} = \beta_i \cdot \left(\frac{1}{1+r^M} - \frac{1}{1+r_f} \right). \tag{5.1}$$

It is convenient to introduce the risk adjusted discount factors $v_i = 1/(1+r_i)$, $v_f = 1/(1+r_f)$ and $v^M = 1/(1+r^M)$. Then equation (5.1) becomes

$$v_i = v_f + \beta_i (v^M - v_f). \tag{5.2}$$

In general, v^M is greater than v_f and β_i is positive. So equation (5.2) produces a positive risk adjustment for the discount factor. Summing up both sides in

equation (5.1), we have

$$\frac{r_f - r_i}{1 + r_i} = \beta_i \frac{r_f - r^M}{1 + r^M}.$$

Assuming for a loss X_i we have $r_i \approx r^M$, then the above equation approximately reduces to

$$r_i = r_f + \beta_i (r^M - r_f). \tag{5.3}$$

The risk adjustment is negative because r^M is less than r_f. Equation (5.3) is in the form of Butsic [3]. Our derivation shows that (5.3) is only an approximation, while equation (5.2), given in terms of the discount factors, is an exact relationship.

Note that the above discount rate correspond to the market risk load, not the total risk load. Discounting by this rate yields the market value of losse. As mentioned in Section 1, the complete premium also includes the insurer specific risk load. Therefore, a more precise term for the above rate would be the risk adjusted *market* discount rate. The discount rate for the complete premium is even smaller than the market discount rate, for an additional risk adjustment is included.

The practical use of the risk adjusted discounting is mostly for multiple-payment claims. For instance, one often estimates the annual payout pattern of a business line and then use a selected discount rate to calculate the present value of liability. The above derivation shows it is inappropriate to use one discount rate for all future years. There is a distinct discount rate for each year commensurate with the riskiness of that year's partial payment. Halliwell [9] argues against any use of the risk adjusted discounting. He proposes to start from the random present value and use the utility theory.

6 Validity of the Model

The risk load model (4.7) has many desirable features and is mathematically proved. But its validity does not directly follow, since the assumptions do not all hold in the real world. In this section, we reexamine the key assumptions and discuss issues related to empirical testing. (4.7) and the classic CAPM share many practical problems. But there are also significant differences.

In Section 4 we assume the insurance market is competitive and is efficient regarding the pricing information. In reality, most policyholders have little knowl-

edge about price. They unknowingly overpay premiums. In the mean time, insurance companies are inadequately diversified because of expense and capital concerns. They have to charge extra amount of risk loads for the remaining unsystematic risk. Therefore, the actual market risk loads are probably higher than needed. (On the other hand, recent industry data show that the P&C insurance as a whole has been less profitable compared with other industries, which seems to indicate the risk loads are charged too low. But this is an issue in the classic CAPM paradigm, not related to our model.) This difficulty does not appear in the context of the classic CAPM, because the financial market is much more efficient.

Another key assumption in (4.7) is that firms attempt to optimize the mean-variance criterion (4.6). The mean-variance is also used in the classic CAPM. As discussed in Section 4, it captures the volatility risk and is especially preferable if the insurance portfolio consists of many small policies. However, this criterion is less effective if the catastrophic or other large losses have a significant impact on the firm. If the risk is highly skewed, the potential damage from tail events is not captured by the variance alone. To remedy this problem higher moment CAPMs have been developed, first in the investment world, and then extended to insurance [10]. The same idea may be used here to add higher moments into equation (4.7).

In practice, model (4.7) should not be applied to individual policies, unless a policy is very large and is stable over time. It may be used to calculate the market risk load for a line of business, or any stable portfolio of policies. Since it is linear with respect to X_i and P_i, equation (4.7) can be stated with respect to any insurance portfolio. All policies need not be written at the same point in time. But the policies in the portfolio and those in the entire market should be comparable, meaning their policy terms and effective dates are similarly distributed within a common time period. It is also convenient to discount the loss of each policy to the policy inception date using the risk free rate. The portfolio version of equation (4.9) is

$$\frac{P_{\text{portfolio}}}{E(X_{\text{portfolio}})} - 1 = \beta_{\text{portfolio}} \left(\frac{P^M}{E(X^M)} - 1 \right), \tag{6.1}$$

where

$$\beta_{\text{portfolio}} = Cov\left(\frac{X_{\text{portfolio}}}{E(X_{\text{portfolio}})}, \frac{X^M}{E(X^M)}\right) / Var\left(\frac{X^M}{E(X^M)}\right). \qquad (6.2)$$

Note the change of notation here: $X_{\text{portfolio}}$ and X^M are not evaluated at the end of the time period, but at the same time as the premiums are evaluated. The ratio $P^M/E(X^M)$ is larger in a hard market and smaller in a soft one. If $\beta_{\text{portfolio}}$ is positive, the market cycle produces similar cyclic change in the portfolio price. Data from a rating agency may be used as a proxy for the overall market.

Since first derived forty years ago, the classic CAPM has been tested extensively. The implications of the test results are widely debated. Not all empirical evidences support the model. The unfavorable ones have led to many modifications of the original model, e.g., redefining β or adding other risk factors. But no single model in any modified version has been statistically confirmed. Nonetheless, the CAPM is still widely used in the financial world. [1] (Chapter 13) and [6] discuss historical development of testing the CAPM. Empirical testing of our model (4.7) or (6.1) has parallel issues. It also poses additional problems because of the nature of insurance business and the (generally inferior) data source.

The first problem with any tests is that insurance claims take many years to settle. Exact values of $X_{\text{portfolio}}$ and X^M are often not known within a reasonable length of time. (In particular, since X^M contains all liability claims, it takes even longer to fully develope.) Using the latest estimates to substitute for the exact values brings about additional randomness. So the quality of the test is inevitably compromised.

Another difficulty is that the market risk load cannot be singled out from the premium. In pricing, usually a total profit and contingency loading is explicitly built into the premium. (In formula, premium = expected loss + expense + total loading.) The total loading is the sum of the market risk load, the insurer specific risk load, and any profit provision over and above the risk loads. But equation (6.1) should only include the market risk load, the value of which cannot be recovered from the historical data.

Yet another challenge comes from the calculation of the expected losses $E(X_{\text{portfolio}})$ and $E(X^M)$ (or the expected loss ratios $E(X_{\text{portfolio}}/P_{\text{portfolio}})$ and $E(X^M/P^M)$). These expected values find further use in estimating the variance and the covariance in (6.2). An expected loss is a forecast made at one point in time, using all available information up to that point, on the average future claim payment. It is not observable from the experience. In the testing of the classic CAPM, the expected returns are statistically estimated from the actual returns. On the financial market stocks are actively traded everyday. Monthly average returns are satisfactory estimates for the monthly expected returns. Average returns of many months are available for the regression analysis. So the CAPM can be tested with reasonable precision. (Chapter 13 of [1] describes a regression using 60 months of data.) In insurance, however, observations are usually made once a year. Using actual losses or loss ratios to estimate the expected values requires many years of data. But such a time span normally would include several pricing cycles. So there is not enough stable samples for the statistical estimation. A discussion of the issue is also seen in [14]. Future expected losses are required inputs in many DFA models. The current projection methods are little more than educated guess.

7 Conclusions

It is convenient to split the total risk load into the market risk load and the insurer specific risk load. Market risk load can be studied using the market equilibrium approach. Our equation (4.7) is mathematically proved parallel to the classic CAPM. Its compact form, intuitive meaning and consistency with real world observations make it an attractive model. Although modifications seem necessary for more accurate calculations, I believe the model itself can provide a guidance and insights to the insurance pricing, similar to the role the CAPM has played in the financial world.

The expected value and the variance of loss, and the covariance between losses, are basic inputs for our model and all other DFA models. Estimation of these values requires both statistical and nontraditional tools. Better techniques need to be developed for the models to become truely useful in company decision making processes.

I wish to thank the CORP reviewers for the thorough review of my earlier draft. Their comments have helped me to understand the literature and to improve my presentation. Suggestions from Shaun Wang and Gary Venter are appreciated. Discussion with Frank Schnapp is also helpful.

References

[1] Bodie, Zvi, Alex Kane and Alan Marcus, *Investments*, McGraw-Hill/Irwin, New York, NY, 2002.

[2] Bühlmann, Hans, "An Economic Premium Principle," *ASTIN Bulletin* 11, 1, 1980, pp. 52-60.

[3] Butsic, Robert, "Determining the Proper Interest Rate for Loss Reserve Discounting: An Economic Approach," *CAS Discussion Paper Program*, May 1988, pp. 147-188.

[4] Butsic, Robert, "Capital Allocation for Property-Liability Insurers: A Catastrophe Reinsurance Application," *CAS Forum*, Spring 1999, 1-70.

[5] Cummins, J.David, "Asset Pricing Models for Insurance Ratemaking," *ASTIN Bulletin*, 20, 2, 1990, pp. 125-166.

[6] Cummins, J.David, Richard Phillips, Robert Butsic, and Richard Derrig, "The Risk Premium Project (RPP) Phase I and II Report," *CAS Forum*, Fall 2000, pp. 165-230.

[7] D'Arcy, Stephen and Michael Dyer, "Ratemaking: A Financial Economics Approach," *PCAS*, LXXXIV, 1997, pp. 301-390.

[8] Feldblum, Sholom, "Risk Loads for Insurers," *PCAS*, LXXVII, 1990, pp. 160-195.

[9] Halliwell, Leigh, "The Valuation of Stochastic Cash Flows," *CAS Forum*, Spring 2003, pp. 1-68.

[10] Kozik, Thomas and Aaron Larson, "The N-Moment Insurance CAPM," *PCAS*, LXXXVIII, 2001, pp. 39-63.

[11] Kulik, John, "A Practical Application of Modern Portfolio Theory to Capital Allocation," *CAS Forum*, Spring 1999, pp. 137-184.

[12] Meyers, Glenn, "The Competitive Market Equilibrium Risk Load for Increased Limits Ratemaking," *PCAS*, LXXVIII, 1991, pp. 163-200.

[13] Meyers, Glenn, Discussion of Feldblum "Risk Loads for Insurers," *CAS Forum*, Winter 1996, 85-95.

[14] Meyers, Glenn, Fredrick Klinker, and David Lalonde, "The Aggregation and Correlation of Insurance Exposure," *CAS Forum*, Summer 2003, pp. 15-82.

[15] Miccolis, Robert, "On the Theory of Increased Limits and Excess of Loss Pricing," *PCAS*, LXIV, 1977, pp. 27-59.

[16] Mildenhall, Stephen, Discussion of Wacek "Application of the Option Market Paradigm to the Solution of Insurance Problems," *PCAS*, LXXXVII, 2000, pp. 162-187.

[17] Myers, Steward and James Read Jr., "Capital Allocation for Insurance Companies," *Journal of Risk and Insurance*, 68, 4, 2001, pp. 545-580.

[18] Panjer, Harry, (editor), *Financial Economics*, The Actuarial Foundation, Schaumburg, IL, 1998.

[19] Schnapp, Frank, "Pricing for Systematic Risk," *CAS Forum*, Fall 2004, pp. 459-475.

[20] Sherris, Michael, "Solvency, Capital Allocation and Fair Rate of Return in Insurance," preprint, 2004. Available from the CAS website http://www.casact.org/research/summaries/allocation.htm

[21] Vaughn, Trent, Discussion of Feldblum "Risk Loads for Insurers," *CAS Forum*, Summer 2003, pp. 263-277.

[22] Venter, Gary, "Premium Calculation Implications of Reinsurance Without Arbitrage," *ASTIN Bulletin*, 21, 2, 1991, pp. 223-230.

[23] Venter, Gary, Discussion of Myers and Read "Capital Allocation for Insurance Companies," *CAS Forum*, Fall 2003, pp. 459-478.

[24] Wang, Shuan, "Implementation of Proportional Hazards Transforms in Ratemaking," *PCAS*, LXXXV, 1998, pp. 940-976.

[25] Wang, Shaun, "Equilibrium Pricing Transforms: New Results Using Bühlmann's 1980 Economic Model," *ASTIN Bulletin*, 33, 1, 2003, pp. 55-73.

[26] Wang, Shaun, "A Universal Framework for Pricing Financial and Insurance Risks," *ASTIN Bulletin*, 32, 2, 2002, pp. 213-234.

[27] Zhang, Yingjie, "Allocating Capital Using the Expected Value of Default," *14th Annual International AFIR Colloquium*, Society of Actuaries, 2004, pp. 541-558. Also available from the AFIR website http://afir2004.soa.org/afir_papers.htm

Appendix: Proof of Equation (4.7)

We prove equation (4.7) under the assumptions stated in Section 4. Our presentation follows the proof of the classic CAPM in [18], Chapter 8.

To increase readability vectors and matrices are used whenever needed. Let us first introduce the following (column) vectors and a matrix

$$\boldsymbol{R} = (R_1, \ldots, R_N)^T \text{ is the vector of returns,}$$
$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^T \text{ is the vector of mean returns,}$$
$$\Sigma = (\sigma_{ij}) \text{ is the } N \times N \text{ variance-covariance matrix.}$$

An insurance portfolio is represented by a pair $(\boldsymbol{a}, w)$, where $\boldsymbol{a} = (a_1, \ldots, a_N)^T$, $0 \leqslant a_i \leqslant 1$ for $i = 1, \ldots, N$, and $w + \sum_{j=1}^{N} a_j P_j > 0$. a_i is the portion of loss X_i included in the portfolio, and w is the amount borrowed. Call a pair $(\boldsymbol{a}, w)$ a *pseudo*-portfolio if the above condition $0 \leqslant a_i \leqslant 1$ is replaced by $-1 \leqslant a_i \leqslant 1$, and all other conditions stay the same. A pseudo-portfolio is not an insurance portfolio if some $a_i < 0$. We can think of an extended insurance market where an insurer can bet with other insurers on the loss of a policy, so that it makes sense to hold a_i portion of a policy even if $a_i < 0$.

For a given pseudo-portfolio $(\boldsymbol{a}, w)$, define a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$ by

$$\alpha_i = \frac{a_i P_i}{w + \sum_{j=1}^{N} a_j P_j}. \tag{A.1}$$

Under the assumption $w + \sum_{j=1}^{N} a_j P_j > 0$, if $a_i > 0$, $= 0$, or < 0, then $\alpha_i > 0$, $= 0$, or < 0, respectively. Conversely, for a given $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$, any pair $(\boldsymbol{a}, w)$ satisfying (A.1) has the form

$$a_i = \frac{\alpha_i}{P_i} \cdot A, \qquad i = 1, \ldots, N,$$
$$w = (1 - \sum_{j=1}^{N} \alpha_j) \cdot A, \tag{A.2}$$

where A is a positive number. It is easy to see $w + \sum_{i=1}^{N} a_i P_i = A$ and α_i and a_i have the same sign. If A is small, then all $|a_i|$s are less than 1.

In Section 4, the utility function (4.6) is a function of the pair $(\boldsymbol{a}, w)$, through the equations (4.4) and (4.5). Denote this function by $F_\tau(\boldsymbol{a}, w)$, that is,

$$F_\tau(\boldsymbol{a}, w) = 2\tau\,\mu_{\text{portfolio}} - \sigma^2_{\text{portfolio}}\,.$$

Equations (4.4) and (4.5) can be restated in terms of $\boldsymbol{\alpha}$, defined in (A.1),

$$\begin{aligned}\mu_{\text{portfolio}} &= \boldsymbol{\alpha}^T\boldsymbol{\mu},\\ \sigma^2_{\text{portfolio}} &= \boldsymbol{\alpha}^T\Sigma\boldsymbol{\alpha}.\end{aligned}$$

So the utility function (4.6) has the following expression

$$\begin{aligned}G_\tau(\boldsymbol{\alpha}) &= 2\tau\,\mu_{\text{portfolio}} - \sigma^2_{\text{portfolio}}\\ &= 2\tau\,\boldsymbol{\alpha}^T\boldsymbol{\mu} - \boldsymbol{\alpha}^T\Sigma\boldsymbol{\alpha}.\end{aligned}$$

Consider the following two optimization problems. An efficient insurance portfolio is determined by

$$\max\{F_\tau(\boldsymbol{a}, w)\,|\,0 \leqslant a_i \leqslant 1, \text{for } i = 1, \ldots, N, w + \sum_{i=1}^{N} a_i P_i > 0\}. \tag{A.3}$$

Or, stated in terms of $\boldsymbol{\alpha}$

$$\max\{G_\tau(\boldsymbol{\alpha})\,|\,\alpha_i \geqslant 0, \text{for } i = 1, \ldots, N\}. \tag{A.4}$$

The following lemma shows (A.3) and (A.4) are equivalent.

Lemma 1 If a pair $(\boldsymbol{a}, w)$ is a solution of the optimization problem (A.3), then $\boldsymbol{\alpha}$, given by (A.1), is a solution of the optimization problem (A.4). Conversely, if $\boldsymbol{\alpha}$ is a solution of (A.4), then there exists a number $A > 0$, so that the pair $(\boldsymbol{a}, w)$, given by (A.2), is a solution of (A.3).

The proof is straightforward. We also need parallel statements for pseudo-portfolios. An "efficient" pseudo-portfolio is a pair $(\boldsymbol{a}, w)$ defined by

$$\max\{F_\tau(\boldsymbol{a}, w)\,|\,-1 \leqslant a_i \leqslant 1, \text{for } i = 1, \ldots, N, w + \sum_{i=1}^{N} a_i P_i > 0\}. \tag{A.3a}$$

Stated in terms of $\boldsymbol{\alpha}$ yields an unconditional optimization problem

$$\max G_\tau(\boldsymbol{\alpha}). \tag{A.4a}$$

(A.3a) and (A.4a) are equivalent in the following sense.

Lemma 1a If a pair $(\boldsymbol{a}, w)$ is a solution of the optimization problem (A.3a), then the corresponding $\boldsymbol{\alpha}$ is a solution of the optimization problem (A.4a). Conversely, if $\boldsymbol{\alpha}$ is a solution of (A.4a), then there exists a number $A > 0$, so that the corresponding pair $(\boldsymbol{a}, w)$ is a solution of (A.3a).

Since $G_\tau(\boldsymbol{\alpha})$ is a quadratic function, the optimization problems (A.4) and (A.4a) are much easier to solve than (A.3) and (A.3a). Equation (4.7) will be proved in two steps. First, assuming there exists an efficient insurance portfolio, we show (4.7) holds. Then we prove an efficient insurance portfolio indeed exists; in fact, the overall insurance market portfolio is efficient.

Step 1. We work with the insurance portfolios and the optimization problems (A.3) and (A.4). The following assumption is needed.

Assumption A. There exists a solution $(\boldsymbol{a}^*, w^*)$ to the optimization problem (A.3), for some $\tau = \tau^*$, such that $\boldsymbol{a}^* = (a_1^*, \ldots, a_N^*)^T$ is a positive vector, that is, $a_i^* > 0$ for all $i = 1, \ldots, N$.

$\boldsymbol{a}^*$ being a positive vector means that this portfolio contains a nonzero fraction of every loss X_i. The reason to make the assumption is as follows. If $\boldsymbol{a}$ is a positive vector, then the corresponding $\boldsymbol{\alpha}$ is also positive. So $\boldsymbol{\alpha}$ lies in the interior of the region $\{\boldsymbol{\alpha} \,|\, \alpha_i \geqslant 0, \text{for } i = 1, \ldots, N, \}$. If the maximum in problem (A.4) is reached at $\boldsymbol{\alpha}$, then $\boldsymbol{\alpha}$ satisfies

$$\frac{\partial}{\partial \alpha_i} G_\tau(\boldsymbol{\alpha}) = 0, \qquad i = 1, \ldots, N.$$

Taking partial derivatives of the quadratic function, yields

$$\tau \boldsymbol{\mu} - \Sigma \boldsymbol{\alpha} = 0. \tag{A.5}$$

Since $G_\tau(\boldsymbol{\alpha})$ is a negative-definite quadratic function, (A.5) gives the one and only $\boldsymbol{\alpha}$ maximizing $G_\tau(\boldsymbol{\alpha})$.

Under Assumption A, the corresponding $\boldsymbol{\alpha}^*$ satisfies (A.5), i.e.,

$$\tau^* \boldsymbol{\mu} - \Sigma \boldsymbol{\alpha}^* = 0.$$

Solving for $\boldsymbol{\alpha}^*$, we have

$$\boldsymbol{\alpha}^* = \tau^* \Sigma^{-1} \boldsymbol{\mu}.$$

Since $\boldsymbol{\alpha}^*$ is a positive vector, the vector $\Sigma^{-1}\boldsymbol{\mu}$ must also be positive. So for any $\tau > 0$, the vector

$$\boldsymbol{\alpha} = \tau\Sigma^{-1}\boldsymbol{\mu}.$$

is positive and satisfies (A.5). It is thus the only solution to the optimization problem (A.4), with respect to τ. From the lemma we conclude that a pair $(\boldsymbol{a}, w)$ is a solution to the optimization problem (A.3) if and only if the corresponding $\boldsymbol{\alpha}$ satisfies equation (A.5).

Now we invoke a market clearing mechanism to prove equation (4.7). Assume there are K insurers, each selecting an efficient insurance portfolio according to its own risk preference. Let the kth insurer hold a portfolio $(\boldsymbol{a}^{(k)}, w^{(k)})$, with respect to $\tau^{(k)} > 0$, where $\boldsymbol{a}^{(k)} = (a_1^{(k)}, \ldots, a_N^{(k)})^T$. Then $(\boldsymbol{a}^{(k)}, w^{(k)})$ is a solution of (A.3) with $\tau = \tau^{(k)}$. The corresponding $\boldsymbol{\alpha}^{(k)} = (\alpha_1^{(k)}, \ldots, \alpha_N^{(k)})^T$ must satisfy (A.5),

$$\tau^{(k)}\boldsymbol{\mu} - \Sigma\boldsymbol{\alpha}^{(k)} = 0. \tag{A.6}$$

If the market clears, then the K portfolios add up to the overall market portfolio. Thus

$$\sum_{k=1}^{K} \boldsymbol{a}^{(k)} = (1, \ldots, 1)^T. \tag{A.7}$$

Let $w^M = w^{(1)} + \cdots + w^{(K)}$. Call the pair $\boldsymbol{a}^M = (1, \ldots, 1)^T$ and w^M the market portfolio. Then the corresponding $\boldsymbol{\alpha}^M$ is given by

$$\begin{aligned} \alpha_i^M &= \frac{P_i}{w^M + \sum_{j=1}^{N} P_j} \\ &= \frac{P_i}{w^M + P^M}, \qquad i = 1, \ldots, N. \end{aligned} \tag{A.8}$$

We introduce the following notation for any k

$$c^{(k)} = \frac{w^{(k)} + \sum_{j=1}^{N} a_j^{(k)} P_j}{w^M + P^M}. \tag{A.9}$$

Then $c^{(k)} > 0$ for $k = 1, \ldots, K$, and from (A.7), $\sum_{k=1}^{K} c^{(k)} = 1$. For any $i = 1, \ldots, N$ we have

$$\begin{aligned} \sum_{k=1}^{K} c^{(k)} \alpha_i^{(k)} &= \sum_{k=1}^{K} c^{(k)} \frac{a_i^{(k)} P_i}{w^{(k)} + \sum_{j=1}^{N} a_j^{(k)} P_j} \\ &= \sum_{k=1}^{K} \frac{a_i^{(k)} P_i}{w^M + P^M} = \frac{P_i}{w^M + P^M} = \alpha_i^M. \end{aligned}$$

Or in vector form

$$\sum_{k=1}^{K} c^{(k)} \boldsymbol{\alpha}^{(k)} = \boldsymbol{\alpha}^{M}. \tag{A.10}$$

Let $\tau^M = \sum_{k=1}^{K} c^{(k)} \tau^{(k)}$. Then from (A.6) and (A.10),

$$\tau^M \boldsymbol{\mu} = \Sigma \boldsymbol{\alpha}^M. \tag{A.11}$$

(A.11) implies the overall market portfolio is an efficient portfolio. Substituting (4.1) and (4.2) into (A.11), yields

$$\begin{aligned} \tau^M \left(1 + r_f - \frac{E(X_i)}{P_i}\right) &= \sum_{j=1}^{N} \frac{1}{P_i P_j} Cov(X_i, X_j) \frac{P_j}{w^M + P^M} \\ &= \frac{1}{P_i} \cdot \frac{1}{w^M + P^M} Cov(X_i, X^M). \end{aligned}$$

Or,

$$\tau^M((1 + r_f)P_i - E(X_i)) = \frac{1}{w^M + P^M} Cov(X_i, X^M). \tag{A.12}$$

Summing up (A.12) over i, yields

$$\tau^M((1 + r_f)P^M - E(X^M)) = \frac{1}{w^M + P^M} Cov(X^M, X^M). \tag{A.13}$$

Dividing (A.12) by (A.13) on both sides and rearranging terms we obtain (4.7).

Step 2. (4.7) has been proved in Step 1 under Assumption A. Now we show the assumption is indeed true; in fact the overall market portfolio is such an $\boldsymbol{\alpha}^*$. We start with the pseudo-portfolios and the optimization problems (A.3a) and (A.4a).

Let each of the K insurers hold an efficient pseudo-portfolio $(\boldsymbol{a}^{(k)}, w^{(k)})$, with respect to $\tau^{(k)} > 0$. Lemma 1a says the corresponding $\boldsymbol{\alpha}^{(k)}$ is a solution of the unconditional optimization problem (A.4a). Thus $\boldsymbol{\alpha}^{(k)}$ satisfies (A.6). If the (extended) market clears, (A.7) holds. Again define $\boldsymbol{\alpha}^M$ and $c^{(k)}$ by (A.8) and (A.9). The condition $w^{(k)} + \sum_{i=1}^{N} a_i^{(k)} P_i > 0$ implies $w^M + P^M > 0$, $\alpha_i^M > 0$ for $i = 1, \ldots, N$, and $c^{(k)} > 0$ for $k = 1, \ldots, K$. Using the same argument as in Step 1, we again derive equation (A.11), with $\tau^M = \sum_{k=1}^{K} c^{(k)} \tau^{(k)} > 0$.

(A.11) means $\boldsymbol{\alpha}^M$ is a solution to the optimization problem (A.4a), with respect to τ^M. But $\boldsymbol{\alpha}^M$ corresponds to the overall market portfolio $(\boldsymbol{a}^M, w^M)$. So

$(\boldsymbol{a}^M, w^M)$ is a solution of the optimization problem (A.3a). Since each $a_i^M = 1$, $(\boldsymbol{a}^M, w^M)$ is also a solution of the original problem (A.3). This proves Assumption A holds with $(\boldsymbol{a}^*, w^*) = (\boldsymbol{a}^M, w^M)$. Furthermore, from (A.11) we have $\boldsymbol{\alpha}^M = \tau^M \Sigma^{-1}\boldsymbol{\mu}$. So $\Sigma^{-1}\boldsymbol{\mu}$ is a positive vector. (A.6) gives $\boldsymbol{\alpha}^{(k)} = \tau^{(k)}\Sigma^{-1}\boldsymbol{\mu}$, which is also a positive vector. Thus the corresponding $\boldsymbol{a}^{(k)}$ is positive. This proves the original efficient pseudo-portfolios $(\boldsymbol{a}^{(k)}, w^{(k)})$ are actually efficient insurance portfolios. Therefore, the argument in Step 1 is entirely valid here. Proof of (4.7) is complete.

(The above proof reveals a very important property of the efficient portfolios: the overall market portfolio is essentially the only efficient portfolio. Any other efficient portfolio must be a fraction of the market portfolio; that is, it contains the same fraction of all policies. A different borrowing amount w produces an efficient portfolio corresponding to a different τ; and an efficient portfolio with respect to any τ is constructed this way with a suitable w.)