Spring 2015

Application of loglinear models to claims triangle runoff data

Netanya Lee Martin

APPLICATION OF LOGLINEAR MODELS TO CLAIMS TRIANGLE RUNOFF DATA

by

NETANYA LEE MARTIN

A THESIS

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN APPLIED MATHEMATICS WITH AN EMPHASIS IN STATISTICS

2015

Approved by

Dr. Akim Adekpedjou, Advisor
Dr. Gregory Gelles
Dr. V.A. Samaranayake

ABSTRACT

In this thesis, we presented in detail different aspects of Verrall's chain ladder method and their advantages and disadvantages. Insurance companies must ensure there are enough reserves to cover future claims. To that end, it is useful to estimate mean expected losses. The chain ladder technique under a general linear model is the most widely used method for such estimation in property and casualty insurance. Verrall's chain ladder technique develops estimators for loss development ratios, mean expected ultimate claims, Bayesian premiums, and Bühlmann credibility premiums. The chain ladder technique can be used to estimate loss development in cases where data has been collected from a population but the statistician has no information on which to base a parametric prior distribution (empirical Bayesian estimation).

ACKNOWLEDGMENTS

Many thanks to all the individuals who helped me along the way: my academic advisor, Dr. Akim Adekpedjou, boyfriend Clifford Long, parents Gary and Devona Martin, and instructors Sister Sharon Hamsa and Janet Wyatt who encouraged a formative interest in mathematics.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 MOTIVATION

An insurance company must ensure as first order of business that there are suitable reserves available to meet the demand by outstanding claims. Insurance companies make promises to policyholders to pay out monetary amounts if certain events (claims) occur. Events may be, for example, due to a car accident for which the policyholder was not at fault (auto insurance), a tree limb damaging a policyholder's house (home insurance), or a monetary benefit after an insured party dies (life insurance). There are two major types of insurance: property & casualty (non-life) and life insurance. For property and casualty (non-life) insurance, one buys a policy to cover for an unexpected loss or partial loss of the covered property due to accident, storm damage, theft, etc. Life insurance covers claims due to death. Insurance companies can suddenly face the possibility of paying claims, which can be large. Large claims can be due to events such as a car accident in which the policyholder's car was totaled or a tornado destroying the policyholder's home. The company needs to have sufficient reserves to cover each claim. There is a need for forecasting future values of those claims based on past experience.

The chain ladder technique is the most common general linear model technique to forecast the claims. Some other modeling possibilities mentioned by Verrall [1] include Zehnwirth's (1985) Gamma curve with $\gamma_{ij} = \nu_i + \delta_i \log j + \gamma_i (j-1) + e_{ij}$ and Ajne's (1989) exponential tail. Sahasrabudhe [2] introduced the short-tail Claims Simulation Model as an alternative technique when the assumptions underlying the chain ladder technique's link ratios are inappropriate for the data. He applied his technique specifically to professional indemnity insurance, where the majority of claims events fall under the incurred but not reported (IBNR) category. Recent developments within the chain ladder framework were summarized by Schmidt [3] in a survey of chain ladder techniques (cf. Schmidt [4]). Schmidt [5] and Zocher [6] extended the existing chain ladder methods, including the Bornhuetter-Ferguson, Cape Cod, and Panning's methods. Another major development to the field in recent years was the Munich chain ladder method, which was developed by Mack [7] and solved the IBNR problem.

The chain ladder technique first arranges the data in a run-off triangle, as shown in Figure 1.1. Run-off triangles are primarily used as an estimation method for non-life insurance claims. The data in the run-off triangle can be truncated or not. It can also be given in incremental or cumulative claims. The observed data (incremental claims) are denoted as Z_{ij}, where i is the underwriting year and j is the delay year. The underwriting year, or year of business, is the year in which a policy

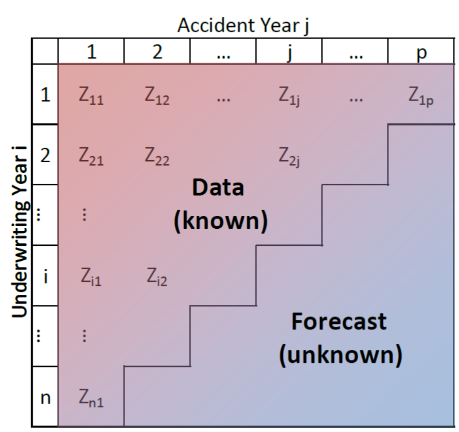

Figure 1.1: Claims Run-Off Triangle

was written, whereas the delay or development year is the number of years until the accident. There is often a reporting delay between the occurrence of an event and the time the claims is reported to the insurer, and a settlement delay until the date at which the insurer settles the loss with the policyholder and pays out the monetary benefit. In this case, the delay year is the time until the claim is reported.

The problem is how to forecast outstanding claims in the lower right half. Sums of data from the run-off triangle, cumulative claims $C_{ij} := \sum_{k=1}^{j} Z_{ik}$ are looked at, and the chain ladder technique is applied to the C_{ij}'s, where Z_{ij} is treated as a random variable. Underlying parameters of claims distributions can then be estimated using the technique. The forecast will give an estimate of the reserve needed to cover the future reported claims amount (cf. Verrall [1]). The forecast could also estimate the amount of reserve necessary to pay off claims that have been reported by the policyholder but not yet settled. The method for forecasting IBNR claims is similar to Verrall's method. IBNR claims could be incurred, for example, due to an accident which costs the policyholder less than the deductible.

A paper by Richard Verrall[1] laid out a framework for analyzing data using the chain ladder technique, extending the range of analysis to loglinear models. Loglinear analysis is a good model for data involving multiple variables in multi-way contingency tables, where no distinction is made between independent and dependent variables - just the association is shown. Also the variables are easily broken into discrete categories by year or possibly months depending on the settlement delays in the line of business. Verrall [1] first tested the loglinear model for goodness of fit, which allows easy comparison of different models, and used the model to forecast outstanding claims. Using the technique in his previous paper [8], Verrall [1] improved the chain ladder technique by extending it using Bayesian methods. The chain ladder technique needed improvement for several reasons, including an improved connection between accident years which prevented over-parameterization and unstable forecasts. The loglinear model needed an allowance for change in parameters over time. He also showed how the chain ladder linear model gives upper prediction bounds on total outstanding claims.

1.2 THE MODEL

The run-off triangle method for claims estimation relies on past claims data. This includes complete monetary amounts due to specific claims and dates for the event, report, and settlement of each claim. Estimating future claims using the run-off triangle only makes sense if all accident years follow the same loss development model (cf. Schmidt [3]). Thus, it implicitly assumes that patterns of claims occurring in the past will continue into the future – that is, a homogenous development pattern. There is no point in forecasting future claims if this assumption is not met. To ensure a homogenous development pattern, lines of business should be segmented so that the observed claims for any run-off triangle stem from a homogenous population. It can be further segmented among populations if there are different claims handling processes resulting in a different pattern of settlement or reporting delays (cf. Wiendorfer [9]). Insurance company data is usually proprietary and therefore difficult to obtain. Some papers and actuarial manuals use fictional portfolios of claims with which to work. The data set for Verall's work, Table (1.1), originated in a paper by Taylor and Ashe [10] given in incremental claims each delay year.[1] The first column is delay year 1, not delay year zero.

Table 1.1: The Taylor-Ashe Data (1983)

Z_{ij}	Delay yr 1	Delay yr 2	Delay yr 3	Delay yr 4	Delay yr 5	Delay yr 6	Delay yr 7	Delay yr 8	Delay yr 9	Delay yr 10
Business yr 1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
Business yr 2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	
Business yr 3	290507	1001799	926219	1016654	750816	146923	495992	280405		
Business yr 4	310608	1108250	776189	1562400	272482	352053	206286			
Business yr 5	443160	693190	991983	769488	504851	470639				
Business yr 6	396132	937085	847498	805037	705960					
Business yr 7	440832	847631	1131398	1063269						
Business yr 8	359480	1061648	1443370							
Business yr 9	376686	986608								
Business yr 10	344014									

[1] Table (1.1) as shown here was published in Verrall's 1994 paper[1] and is available in Excel format online.

Exposure factors are a subjective, potential percentage of a loss to a specific asset if a specific threat is realized. If PL is the ceding company policy limit, AP is the treaty attachment point, and Lim is the treaty limit, the *exposure factor* is defined as[2]

$$\text{Exposure Factor } = \left(E[X; \min\{PL, AP + \text{Lim}\}] - \frac{E[X; \min\{PL, AP\}]}{E[X; PL]} \right). \quad (1.1)$$

To connect successive development years (delay years until the accident), one looks at the ratios of claims in successive delay years. The resulting development factor is called the link ratio. Multiple quantities can be used to estimate the development factor: the proportion of the ultimate cumulative claims losses settled in a particular development year (development pattern for incremental claims losses settled), the proportion of ultimate cumulative claims losses settled by a particular development year (development pattern for cumulative claims losses), and the ratio of cumulative claims losses settled by a particular development year to the cumulative claims losses settled by the previous development year (cumulative claims loss factor). Verrall used the last ratio, and estimated the development factor for delay year j as

$$\text{Link Ratio } = \widehat{\lambda}_j = \frac{\sum_{i=1}^{t-j+1} C_{ij}}{\sum_{j=1}^{t-j+1} C_{i,j-1}}, \quad (1.2)$$

$$\text{where } E[C_{ij} | C_{i1}, C_{i2}, ..., C_{i,j-1}] = \lambda_j C_{i,j-1} \qquad \text{for } j = 2, ..., t.$$

Note that the numerator in Equation (1.2) looks at delay year j but the denominator at delay year $j - 1$. Both have the same business (underwriting) year. The

[2]For the purpose of example, the ceding company is giving up some portion of their risk to a reinsurer. If the ceding company, for example, has a policy limit of \$5 Million, keeps \$3 Million of the risk and gives up \$2 Million of their risk to the reinsurer; then $PL = 5$ Million, $AP = 2$ Million, $Lim = 3$ Million.

assumption that patterns of claims occurring in the past continue into the future induces an assumption that development of settled claims losses follows the same pattern for every claims occurrence year – i.e., that the cumulative claims loss settlement factor for a specific development year is the same for all claims occurrence years.The cumulative claims pattern is shown in Table (1.2). By way of example, the cumulative claims amount C_{13} for business year 1 and accident year 3 is calculated as $\sum_{k=1}^{3} Z_{1k} = 1735330$. Using Equation (1.2), the link ratio estimate

Table 1.2: Cumulative Claims Pattern

C_{ij}	Delay yr 1	Delay yr 2	Delay yr 3	Delay yr 4	Delay yr 5	Delay yr 6	Delay yr 7	Delay yr 8	Delay yr 9	Delay yr 10
Business yr 1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463
Business yr 2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	
Business yr 3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315		
Business yr 4	310608	1418858	2195047	3757447	4029929	4381982	4588268			
Business yr 5	443160	1136350	2128333	2897821	3402672	3873311				
Business yr 6	396132	1333217	2180715	2985752	3691712					
Business yr 7	440832	1288463	2419861	3483130						
Business yr 8	359480	1421128	2864498							
Business yr 9	376686	1363294								
Business yr 10	344014									

$\hat{\lambda}_9 = (3833515 + 5339085)/(3606286 + 4914039) = 1.076552$. The rest of the link ratios are calculated as shown in Table (1.3).

Why would one want this link ratio estimate anyway? The link ratio estimate allows for a fairly straightforward forecast of $E[C_{ij}|$ past values of $C_{ij}]$. This is an estimate of the cumulative claims for underwriting year i and accident year j. The most recent loss $C_{i,t-i+1}$ is multiplied by the appropriate development factor estimate $\hat{\lambda}_j$. For example, $E[C_{67}|C_{61}, \ldots, C_{66}] = (1.\widehat{0862694})(4074998.6)$. The estimated loss forecasts are shown in lower right triangle of Table (1.3). Ease of calculation makes this a method allowed under Solvency II[3] to estimate reserves needed for non-life

[3]Solvency II is a law concerning the reserves European Union insurance companies must hold to prevent insolvency. See Wiendorfer [9] for a practical application of the chain ladder method under Solvency II.

Table 1.3: Cumulative Claims Forecast

C_{ij}	Delay yr 1	Delay yr 2	Delay yr 3	Delay yr 4	Delay yr 5	Delay yr 6	Delay yr 7	Delay yr 8	Delay yr 9	Delay yr 10	Delay yr 11
Business yr 1	357848	1124788	1735330	2218270	2745596	3319994	3466336	3606286	3833515	3901463	3901463
Business yr 2	352118	1236139	2170033	3353322	3799067	4120063	4647867	4914039	5339085	5339085	5339085
Business yr 3	290507	1292306	2218525	3235179	3985995	4132918	4628910	4909315	5285148.5	5378826.3	5378826.3
Business yr 4	310608	1418858	2195047	3757447	4029929	4381982	4588268	4835458	5205637.3	5297905.8	5297905.8
Business yr 5	443160	1136350	2128333	2897821	3402672	3873311	4207459.1	4434133.2	4773589.1	4858199.6	4858199.6
Business yr 6	396132	1333217	2180715	2985752	3691712	4074998.6	4426546.1	4665023.4	5022155.1	5111171.5	5111171.5
Business yr 7	440832	1288463	2419861	3483130	4088678.1	4513179.1	4902528.2	5166648.7	5562182.5	5660770.6	5660770.6
Business yr 8	359480	1421128	2864498	4174756.2	4900544.6	5409336.5	5875996.5	6192562.1	6666634.7	6784799	6784799
Business yr 9	376686	1363294	2382128.1	3471744.1	4075312.7	4498426.1	4886502.4	5149759.6	5544000.4	5642266.3	5642266.3
Business yr 10	344014	1200817.5	2098227.7	3057983.9	3589619.6	3962306.6	4304132.3	4536014.7	4883270.1	4969824.7	4969824.7
Link Ratio $\hat{\lambda}_j$		3.4906	1.7473	1.4574	1.1739	1.1038	1.0863	1.0539	1.0766	1.0177	

insurance; however, it is also a rough estimation method (cf. Verrall [1]). This gives rise to Verrall's estimate of the desired forecast:

$$\text{Expected Ultimate Loss } = \widehat{E[C_{it}]} = \left(\prod_{j=t-i+2}^{t} \widehat{\lambda}_j \right) C_{i,t-i+1}. \tag{1.3}$$

For example, $E[C_{6,6}] = \prod_{j=2}^{6} \lambda_j C_{6,1}$. Verrall's estimates are shown in Table (1.4). It results in a row effect, due to $\lambda_i; i = 2, ..., t$ and the row being considered,

Table 1.4: Forecast with Verrall's Expected Loss Method

C_{ij}	Delay yr 1	Delay yr 2	Delay yr 3	Delay yr 4	Delay yr 5	Delay yr 6	Delay yr 7	Delay yr 8	Delay yr 9	Delay yr 10	Sum	U_i
UW yr 1	357,848	1,124,788	1,735,330	2,218,270	2,745,596	3,319,994	3,466,336	3,606,286	3,833,515	3,901,463	26,309,426	32,724,489
UW yr 2	352,118	1,236,139	2,170,033	3,353,322	3,799,067	4,120,063	4,647,867	4,914,039	5,339,085	5,433,719	35,365,452	43,506,648
UW yr 3	290,507	1,292,306	2,218,525	3,235,179	3,985,995	4,132,918	4,628,910	4,909,315	5,251,748	5,378,826	35,324,229	43,673,409
UW yr 4	310,608	1,418,858	2,195,047	3,757,447	4,029,929	4,381,982	4,588,268	5,092,411	5,400,491	5,297,906	36,472,947	45,388,370
UW yr 5	443,160	1,136,350	2,128,333	2,897,821	3,402,672	3,873,311	4,365,888	4,298,439	4,628,949	4,858,200	32,033,123	39,378,637
UW yr 6	396,132	1,333,217	2,180,715	2,985,752	3,691,712	4,562,595	4,778,696	4,714,338	4,767,924	4,111,172	34,522,253	42,679,436
UW yr 7	440,832	1,288,463	2,419,861	3,483,130	-	-	-	-	-	-	-	-
UW yr 8	359,480	1,421,128	2,864,498	-	-	-	-	-	-	-	-	-
UW yr 9	376,686	1,363,294	-	-	-	-	-	-	-	-	-	-
UW yr 10	344,014	-	-	-	-	-	-	-	-	-	-	-
$\hat{\lambda}_j$		3.4906	1.7473	1.4574	1.1739	1.1038	1.0863	1.0539	1.0766	1.0177		

and also a column effect due to $\lambda_j; j = 2, ..., t$. The random variable $C_{i,t-i+1}$ =

latest cumulative claim = row effect. For this reason, Verrall also considered other models with row and column effects, such as ANOVA and the multiplicative model.[4]

One-way ANOVA compares the null hypothesis H_0: $\mu_1 = \mu_2 = \cdots = \mu_n$ to the alternative hypothesis H_A: There exists (i, j) where $i \neq j$ such that $\mu_i \neq \mu_j$. Examples of model factors are as shown in Figure (1.2).

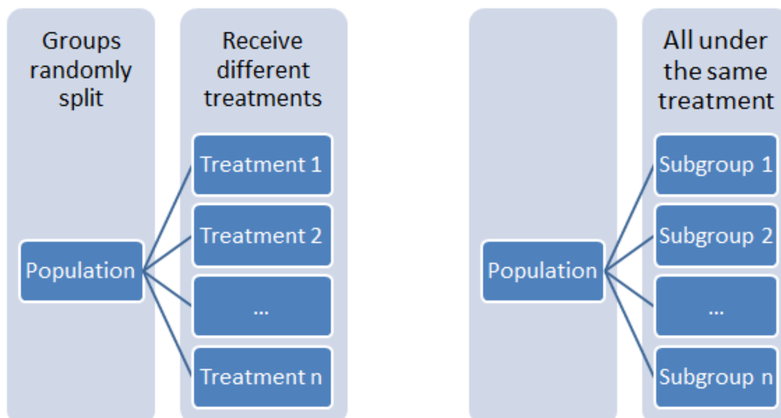

Figure 1.2: One-way ANOVA model possibilities

The chain ladder technique and two-way ANOVA are quite similar. Both can be represented by a linear regression model (cf. Christensen [11]) and both models have both a row and a column effect (cf. Verrall [1]). For two-way ANOVA, a setup looking at two factors (A and B) where Factor A has three different levels and Factor B has 3 levels could have the 9 treatment groups shown in Figure (1.3). It is also possible to have multiple observations per cell, though the incremental run-off triangle will have a single observation, $\log Z_{ij}$ per cell. The design set-up is randomized block, with underwriting year and accident year factors. There is a one-way ANOVA null hypothesis for both the row and column factors. For the row factor,

[4]ANOVA models can be represented by regular regression models of the form $Y = XB + \varepsilon$ where the matrix X is entirely 0's and 1's.[11][12]

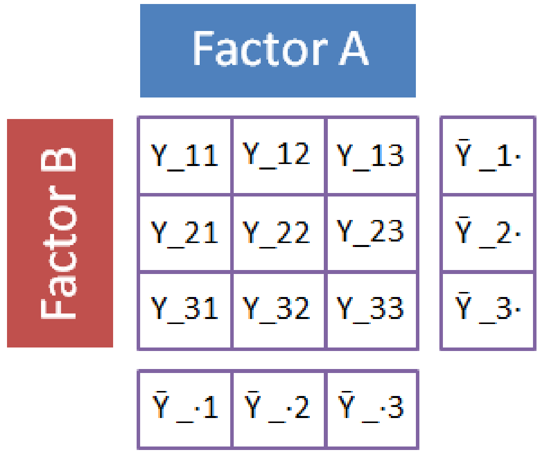

Figure 1.3: Two-way ANOVA model with single observations per cell

the null hypothesis is H_0: $\mu_{1\cdot} = \mu_{2\cdot} = \cdots = \mu_{k\cdot}$ versus H_A: At least one $\mu_{i\cdot} \neq \mu_{j\cdot}$. For the one-way ANOVA column factor, H_0: $\mu_{\cdot 1} = \mu_{\cdot 2} = \cdots = \mu_{\cdot l}$ where H_A: At least one $\mu_{\cdot i} \neq \mu_{\cdot j}$. These are equivalent to testing H_0: No main effect of row factor versus H_A: There is a main effect of row factor and H_0: No main effect of column factor versus H_A: There is a main effect of column factor. The third test for independence looks at possible interaction between the two factors. The null hypothesis is H_0: $(\mu_{ij} - \mu_{\cdot j}) = (\mu_{i\cdot} - \mu_{\cdot\cdot})$.[5] The assumptions for the tests include data being lognormally distributed, the model being applied to logged incremental claims, and unbiasedness of estimates.

The multiplicative model $Z = b_0(X_1^{b_1})(X_2^{b_2})\epsilon$, where b_0, b_1, $b_2 \geq 0$ and ϵ is the error term, is best used when the dependent variable Y is proportional to percentage changes in the independent variables X_i. It makes sense to use the multiplicative model for the claims triangle under the assumption that ultimate cumulative claims losses are proportional to claims in a particular development year (cf. Schmidt [5]). Using the additive general linear regression model, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ where

[5] This is achieved by testing the sum of squares error, SSE, for significance. SSE is large when there is column and row interaction. Kleinbaum describes *Tukey's test for additivity* to test for such interaction.[12]

$\beta = (\beta_0, \beta_1, \beta_2)$ is the regression parameter vector and ε is the error vector, as a framework, the multiplicative model can be turned into a linear model, $\log Z = \beta_0 + \beta_1 \log X_1 + \beta_2 \log X_2 + \varepsilon$, through the logarithmic transformation. Now if $Z_{ij} \sim$ lognormal, then $Y_{ij} = \log Z_{ij} \sim N(\mu, \sigma^2)$, where $E[Z_{ij}] = e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}$.

Verrall's multiplicative model sets up the data with parameters U_i for row i and S_j for column j. Then U_j is the expected total claim for business year i and S_j is the expected proportion of ultimate claims for development year j. The multiplicative model is given by $E[Z_{ij}] = U_i S_j$, where $\sum_{j=1}^{t} S_j = 1$. Kremer [13] derived a form for the expected ultimate loss, which can be used to estimate the entire reserve needed to pay out claims, given by:

$$\text{Expected Ultimate Loss } = U = E[C_{it}] \qquad \text{where } U_i = e^{\alpha_i} e^{\mu} \sum_{j=1}^{t} e^{\beta_j}. \tag{1.4}$$

Furthermore, the estimate of U is $\widehat{U_i} = e^{\hat{\alpha_i}} e^{\hat{\mu}} \sum_{j=1}^{t} e^{\hat{\beta_j}}$. The result in (1.4) is very similar to that of the chain ladder technique; however, Kremer's estimate $\widehat{U_i}$ is not unbiased. The relationships between the parameters for the loglinear model are:

$$S_j = \frac{\lambda_j - 1}{\prod_{i=j}^{t} \lambda_i}, \qquad S_1 = \frac{1}{\prod_{i=2}^{t} \lambda_i}, \text{ and} \qquad U_i = E[C_{it}].$$

Applying Kremer's estimate to the previous Taylor-Ashe data, the estimates for S_j (rounded to four decimal places) are calculated in Table (1.5) along with the forecasts based on proportionality factor S_j. The unrounded values for S_j sum to 1.

Estimates $\widehat{S_i}$ and $\widehat{U_i}$, can be obtained by applying a linear model to the logged incremental claims data: $E[Z_{ij}] = U_i S_j$, so that $\log(Z_{ij}) = Y_{ij} \Rightarrow E(Y_{ij}) = \mu + \alpha_i + \beta_j$. For an alternate derivation of estimates from the multiplicative model, see Schmidt [5]. There are some assumptions with the loglinear model; namely, errors are identically distributed around zero with standard deviation σ (cf. Christensen [11] and Verrall

Table 1.5: Forecast with Kremer's Proportionality Factor

C_{ij}	Delay yr 1	Delay yr 2	Delay yr 3	Delay yr 4	Delay yr 5	Delay yr 6	Delay yr 7	Delay yr 8	Delay yr 9	Delay yr 10	Sum	U_i
UW yr 1	357,848	1,124,788	1,735,330	2,218,270	2,745,596	3,319,994	3,466,336	3,606,286	3,833,515	3,901,463	26,309,426	32,724,489
UW yr 2	352,118	1,236,139	2,170,033	3,353,322	3,799,067	4,120,063	4,647,867	4,914,039	5,339,085	5,339,085	35,270,818	43,410,336
UW yr 3	290,507	1,292,306	2,218,525	3,235,179	3,985,995	4,132,918	4,628,910	4,909,315	4,996,331	4,996,331	34,686,318	43,009,164
UW yr 4	310,608	1,418,858	2,195,047	3,757,447	4,029,929	4,381,982	4,588,268	5,027,075	5,116,179	5,116,179	35,941,572	44,828,489
UW yr 5	443,160	1,136,350	2,128,333	2,897,821	3,402,672	3,873,311	4,472,371	4,900,094	4,986,947	4,986,947	33,228,007	40,644,809
UW yr 6	396,132	1,333,217	2,180,715	2,985,752	3,691,712	4,630,424	5,346,582	5,857,912	5,961,742	5,961,742	38,345,932	46,727,227
UW yr 7	440,832	1,288,463	2,419,861	3,483,130	4,822,390	6,048,606	6,984,105	7,652,043	7,787,673	7,787,673	48,714,776	58,100,020
UW yr 8	359,480	1,421,128	2,864,498	4,655,372	6,455,358	8,084,255	9,334,595	10,227,326	10,408,603	10,408,203	64,209,217	75,956,842
UW yr 9	376,686	1,363,294	3,229,074	5,247,880	7,265,684	9,113,170	10,522,646	11,528,998	11,733,346	11,733,346	72,114,124	84.790,938
UW yr 10	344,014	1,423,771	3,372,319	5,480,681	7,587,997	9,517,440	10,989,441	12,040,436	12,253,849	12,253,849	75,263,797	88,552,349
λ_j		3.4906	1.7473	1.4574	1.1739	1.1038	1.0863	1.0539	1.0766	1.0177		
$\prod \lambda_j$		14.4466	4.1387	2.3686	1.6252	1.3845	1.2543	1.1547	1.0956	1.0177		
Proportion S_j	0.0692	0.1724	0.1806	0.1931	0.1070	0.0750	0.0688	0.0467	0.0699	0.0174		
$Sum(S_j)$	.0692	.2416	.4222	.6153	.7223	.7973	.8661	.9127	.9826	1		

[1]). Setting parameters $\alpha_1 = \beta_1 = 0$ ensures a non-linear design matrix. The loglinear model can be written in the familiar form $Y = X\beta + \varepsilon$, where Y is the vector of logged incremental claims, X is the design matrix, β is the parameter vector, and ε is the error vector. The logged data is shown in Table (1.6). Verrall [14] fit a loglinear

Table 1.6: Logged Taylor-Ashe Data

$\log(Z_{ij})$	Delay yr 1	Delay yr 2	Delay yr 3	Delay yr 4	Delay yr 5	Delay yr 6	Delay yr 7	Delay yr 8	Delay yr 9	Delay yr 10
UW yr 1	12.7879	13.5502	13.3221	13.0877	13.1756	13.2612	11.8937	11.8490	12.3337	11.1265
UW yr 2	12.7717	13.6922	13.7471	13.9838	13.0075	12.6792	13.1765	12.4919	12.9600	
UW yr 3	12.5794	13.8173	13.7389	13.8320	13.5289	11.8977	13.1143	12.5440		
UW yr 4	12.6463	13.9183	13.5622	14.2617	13.5153	12.7715	12.2370			
UW yr 5	13.0017	13.4491	13.8075	13.5535	13.1320	13.0619				
UW yr 6	12.8895	13.7505	13.6500	13.5986	13.4673					
UW yr 7	12.9964	13.6502	13.9390	13.8769						
UW yr 8	12.7924	13.8753	14.1825							
UW yr 9	12.8392	13.8020								
UW yr 10	12.7484									

model $\log Z_{ij} = X_{ij}\beta + \varepsilon_{ij}$ to the logged incremental claims, where i is the underwriting year, j is the delay year, the matrix X_{ij} contains explanatory variables, β is the parameter vector, $Var[\varepsilon] = \sigma^2 I$ is the error vector, and errors ε_{ij}, ε_{kl} are independent. The run-off triangle data is assumed to be loglinear, with independent identically distributed (iid) errors. The homogeneity assumption (identically distributed errors) can occasionally be violated if the sample size is large enough. In matrix notation, $Y = X\beta + \varepsilon$ for the first three years of data (cf. Verrall [8]) would look like

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} 12.7879 \\ 13.5502 \\ 13.3221 \\ 12.7717 \\ 13.6922 \\ 12.5794 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \end{bmatrix}. \tag{1.5}$$

The standard error for the three years of data shown above is $\hat{\sigma} = 0.07911$. For the full model, X is a 55×19 design matrix. The standard error, which can be computed in a program such as RStudio is $\hat{\sigma} = 0.3409$. The least-squares estimate is $\hat{\beta} = (X'X)^{-1}X'y$ (Christensen [11]). For the three years of data,

$$\hat{\beta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 12.7483105 \\ 0.0629650 \\ -0.1689275 \\ 0.8414070 \\ 0.5737915 \end{bmatrix}.$$

For the full model,

$$\hat{\beta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9 \\ \alpha_{10} \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \\ \beta_{10} \end{bmatrix} = \begin{bmatrix} 12.519840036 \\ 0.361002000 \\ 0.282241500 \\ 0.171194452 \\ 0.282223613 \\ 0.311746273 \\ 0.392047931 \\ 0.480267443 \\ 0.345165464 \\ 0.228599964 \\ 0.911189000 \\ 0.938718562 \\ 0.964980568 \\ 0.383200324 \\ -0.004910349 \\ -0.118069524 \\ -0.439277869 \\ -0.053511036 \\ -1.393340036 \end{bmatrix}.$$

Verrall [1] also examined enhancing the stability of predictions using Bayesian methodology. Often mentioned in research during the last decade, the use of prior estimators of ultimate claims losses and prior estimators of the cumulative claims loss settlement factors (development factors) can improve reliability of the estimation. These prior estimators can be based on run-off triangle data, another data source like market statistics or data obtained through a reinsurer, or the insurer's personal experience. Verrall uses data contained within the run-off triangle itself.

2 VERRALL'S ESTIMATORS

2.1 ESTIMATION OF RESERVES

Using the "quick and dirty" method mentioned in the introduction, the reserve can be estimated using the cumulative claims forecasts from Table (1.3). First use the cumulative claims triangle to forecast the incremental claims triangle, shown in Table (2.1). Then use the incremental claims forecast to estimate the claim loss settlement

Table 2.1: Forecasted Incremental Claims

Z_{ij}	Delay yr 1	Delay yr 2	Delay yr 3	Delay yr 4	Delay yr 5	Delay yr 6	Delay yr 7	Delay yr 8	Delay yr 9	Delay yr 10
UW yr 1	357848	766940	610542	482940	527326	574398	146342	139950	227229	67948
UW yr 2	352118	884021	933894	1183289	445745	320996	527804	266172	425046	4914039
UW yr 3	290507	1001799	926219	1016654	750816	146923	495992	280405	5004743	374083
UW yr 4	310608	1108250	776189	1562400	272482	352053	206286	4629172	576465	4721440
UW yr 5	443160	693190	991983	769488	504851	470639	3736820	697313	4076276	781924
UW yr 6	396132	937085	847498	805037	705960	3369039	1057508	3607516	1414639	3696532
UW yr 7	440832	847631	1131398	1063269	3025409	1487770	3414758	1751891	3810292	1850479
UW yr 8	359480	1061648	1443370	2731386	2169158	3240178	2635819	3556744	3109891	3674908
UW yr 9	376686	986608	1395520	2076224	1999089	2499337	2387165	2762595	2781406	2860860
UW yr 10	344014	856804	1241424	1816560	1773060	2189247	2114886	2421129	2462141	2507684

amount for each future year. The incremental forecasts are summed diagonally across the triangle for each future year reserve estimate. For example, for year 11, the estimated claim loss settlement is $856804 + 1395520 + 2731386 + \ldots + 4914039 = 29662932$. The reserve can be calculated from the estimated claim loss settlement amounts for upcoming years, as in Table (2.2). However, since the cumulative claim forecasts were biased estimates, this method also results in a biased estimate for the reserve. See Wiendorfer [9] for a more detailed explanation of this calculation method.

Verrall's [1] problem was reversing the log transformation to get unbiased estimates on the original scale. Since $C_i \sim$ lognormal, then $Y_i = \log C_i \sim N(\mu, \sigma^2)$ and

Table 2.2: Estimated Claim Loss Settlements

yr 11	29,662,932
yr 12	9,679,945
yr 13	22,875,817
yr 14	10,856,669
yr 15	15,639,980
yr 16	9,837,850
yr 17	8,877,443
yr 18	5,323,001
yr 19	2,507,684
Total	115,261,321

$E[C] = \hat{\theta} = e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}$, where $\hat{\mu}$ and $\hat{\sigma}^2$ are the MLE estimates. The estimate $\hat{\mu}$ is biased, but $\hat{\theta}$ is asymptotically unbiased. In claims reserving, n is not usually large, which may lead to biased estimators. To overcome this problem, Verrall used a result attributed to Finney [15] for an unbiased estimate $\hat{\theta}$ using $g_m(t)$: $\hat{\theta} = \exp(\hat{\mu}) g_m\left(\frac{1}{2}(1-\frac{1}{n})s^2\right)$, where s^2 is the standard error estimate of σ^2, and

$$g_m(t) = \sum_{k=0}^{\infty} \frac{m^k(m+2k)}{m(m+2)\cdots(m+2k)} \frac{t^k}{k!}, \tag{2.1}$$

$m = n - 1$ being the degrees of freedom associated with the distribution of $\hat{\sigma}^2$ (cf. Bradu and Mundlak [16]).

To estimate the reserve, we first need an unbiased estimate for mean incremental claims. To proceed, we assume that the data is lognormally distributed, $Z_i \overset{iid}{\sim}$ lognormal, with mean $E[Z_i] = \theta$. This implies that $Y_i = \log Z_i \sim N(\mu, \sigma^2)$ and $\theta = e^{\mu+\frac{1}{2}\sigma^2}$. We must estimate θ and find the mean squared error, or find σ^2 if $\hat{\theta}$ is unbiased. The maximum likelihood estimates are $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2$. This means θ can be estimated as a plug-in estimator $\hat{\theta} = e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}$. Since $\hat{\theta}$ is biased, the results of Finney (2.1), come in handy: Now $\tilde{\theta} = e^{\hat{\mu}} g_m\left(\frac{1}{2}(1-\frac{1}{n})s^2\right)$ is unbiased for θ where $s^2 = \frac{n}{n-1}\hat{\sigma}^2$. Verrall's [1] plug-in estimator for the variance is

$$Var[\tilde{\theta}] = \tilde{\tau}^2 = e^{2\hat{\mu}} \left\{ \left(g_m \left[\frac{1}{2} \left(1 - \frac{1}{n} \right) s^2 \right] \right)^2 - g_m \left(\left[1 - \frac{2}{n} \right] s^2 \right) \right\}. \tag{2.2}$$

Next, for the claims runoff triangles, the data is based on both the business year i and delay year j. If Z_{ij} is distributed lognormally with mean $\theta_{ij} = E[Z_{ij}] = e^{(X_{ij}\beta + \frac{1}{2}\sigma^2)}$, then $Y_{ij} = \log Z_{ij} \overset{indep}{\sim} N(X_{ij}\beta, \sigma^2)$. Then the expected value of the logged data is $E[Y_{ij}] = X_{ij}\beta$ and the variance is $Var[Y_{ij}] = \sigma^2$, where X_{ij} is a row vector of the variables and β is a column vector of parameters. When we look at Y as the vector of observations of our data, $E[Y] = X\beta$, where X is an $n \times p$ matrix. The errors in this model are assumed to be *iid* normally distributed, with variance $\sigma^2 I$. The problem of a biased estimator $\hat{\theta}$ for the mean θ of the data Z_{ij} mentioned above arises again here, and it is dealt with similarly by using Finney's $g_m(t)$ along with an unbiased estimator derived by Bradu and Mundlak [16]. Bradu and Mundlak's unbiased estimator of $e^{Z\beta + a\sigma^2}$, where Z is a row vector, is $e^{Z\hat{\beta}} g_m \left([a - \frac{1}{2} Z(X'X)^{-1}Z'] s^2 \right)$. From this Verrall [1] derived the unbiased estimate of $E[Z_{ij}] = \theta_{ij}$:

$$\tilde{\theta}_{ij} = e^{\left(X_{ij}\hat{\beta}\right)} g_m \left(\frac{1}{2} \left[1 - X_{ij}(X'X)^{-1}X'_{ij} \right] s^2 \right), \tag{2.3}$$

along with a matching estimate for variance:

$$\begin{aligned} Var[\tilde{\theta}_{ij}] = \tilde{\tau}^2_{ij} = e^{\left(2X_{ij}\hat{\beta}\right)} \Bigg\{ & \left[g_m \left(\frac{1}{2} (1 - X_{ij}(X'X)^{-1}X'_{ij}) s^2 \right) \right]^2 \\ & - g_m \left(\left[1 - 2X_{ij}(X'X)^{-1}X'_{ij} \right] s^2 \right) \Bigg\}. \end{aligned} \tag{2.4}$$

Now if $\hat{\beta} = (X'X)^{-1}X'y$ as before (cf. Christensen [11]) and we take s^2 to be the standard error estimate, then for the smaller 3-year model of logged data, the estimate for $E[Z_{2j}]$ is

$$\begin{aligned}\tilde{\theta}_{2j} &= e^{\left(X_{2j}\hat{\beta}\right)} g_m\left(\frac{1}{2}\left[1 - X_{2j}(X'X)^{-1}X'_{2j}\right] s^2\right) \\ &= (797883.2) g_m(0.0007822418) \\ &= (797883.2)\sum_{k=0}^{\infty} \frac{4^k(4+2k)}{4(6)\ldots(4+2k)} \frac{(.0007822418)^k}{k!}.\end{aligned}$$

To estimate the reserve, we now need an estimator for mean total outstanding claims, which was derived by Verrall [14]. Begin by totaling the claims by business year. Let R_i be the total claims for business year i and R be the total outstanding claims for the entire run-off triangle. Then $\tilde{R}_i = \sum_{j=t-i+2}^{t} \tilde{\theta}_{ij}$ is an unbiased estimate of total outstanding claims in business year i, where $\tilde{\theta}_{ij} = \widetilde{E[Z_{ij}]}$ as in (2.3) and Z_{ij} are incremental claims. An unbiased estimate for the variance derived by Verrall using the plug-in technique is

$$\begin{aligned}\widetilde{Var[R_i]}_{\text{unbiased}} = Var[\tilde{R}_i] &= Var\left[\sum_{j=t-i+2}^{t} \tilde{\theta}_{ij}\right] \\ &= \sum_{j=t-i+2}^{t}\left[Var[\tilde{\theta}_{ij}] + 2\sum_{k=j+1}^{t} Cov(\tilde{\theta}_{ij}, \tilde{\theta}_{ik})\right] \\ &= \sum_{j=t-i=2}^{t}\left[\tilde{\tau}_{ij}^2 + 2\sum_{k=j+1}^{t} \tilde{\tau}_{ijk}\right], \qquad (2.5)\end{aligned}$$

where $\tilde{\tau}_{ijk} = Cov(\tilde{\theta}_{ij}, \tilde{\theta}_{ik})$ (2.6) is an unbiased estimate derived the same way as Verrall [1] derived the mean claims estimate $\tilde{\tau}_{ij}^2 = Var[\tilde{\theta}_{ij}]$ in (2.4). Verrall's variance for total outstanding claims in business year i turns out to be exactly the same as the variance given by Bradu and Mundlak:

$$\tau_{ij}^2 = Var[\tilde{\theta}_{ij}] = E[\tilde{\theta}_{ij}^2] - \left(E[\tilde{\theta}_{ij}]\right)^2 = \tilde{\theta}_{ij}^2 - \theta_{ij}^2 = \tilde{\theta}_{ij}^2 - e^{2X_{ij}\beta+\sigma^2}.$$

Therefore,

$$\tilde{\theta}_{ij}^2 = e^{2X_{ij}\beta}\left\{\left[g_m\left(\frac{1}{2}[1-X_{ij}(X'X)^{-1}X_{ij}]'\right)s^2\right]^2 - g_m\left([1-2X_{ij}(X'X)^{-1}X_{ij}']s^2\right)\right\}.$$

Now,

$$\tau_{ijk} = Cov(\tilde{\theta}_{ij}, \tilde{\theta}_{ik}) = E[\tilde{\theta}_{ij}\tilde{\theta}_{ik}] - E[\tilde{\theta}_{ij}]E[\tilde{\theta}_{ik}] = \tilde{\theta}_{ij}\tilde{\theta}_{ik} - \theta_{ij}\theta_{ik}$$

implies

$$\begin{aligned}\tilde{\tau}_{ijk} = e^{(X_{ij}+X_{ik})\hat{\beta}}\left\{g_m\left(\frac{1}{2}[1-X_{ij}(X'X)^{-1}X_{ij}']s^2\right)g_m\left(\frac{1}{2}[1-X_{ik}(X'X)^{-1}X_{ik}']s^2\right)\right.\\ \left. -g_m\left(\left[1-\frac{1}{2}(X_{ij}+X_{ik})(X'X)^{-1}X_{ij}+X_{ik}\right]s^2\right)\right\}. \quad (2.6)\end{aligned}$$

To derive an estimate of total outstanding claims for the entire triangle,

$$E[\tilde{R}] = \sum_{i=2}^{t}\tilde{\theta}_{ij} = \sum_{i=2}^{t}\left\{e^{\left(X_{ij}\hat{\beta}\right)}g_m\left(\frac{1}{2}\left[1-X_{ij}(X'X)^{-1}X_{ij}'\right]s^2\right)\right\}, \qquad (2.7)$$

where $\tilde{\theta}_{ij} = \widetilde{E[Z_{ij}]}$ as in (2.3). Then with the entire lower triangle forecasted as shown above using Equation (2.3) and perhaps a computer program to quickly calculate values for $g_m(t)$, we could calculate an unbiased estimate for the total reserve more accurate than the current "quick and dirty" method shown at the beginning of this chapter. Verrall [1] also derived the variance

$$\begin{aligned}Var[\tilde{R}] = \sum_{i=2}^{t}\sum_{j=t-i+2}^{t}\widetilde{Var(Z_{ij})} = \sum_{i=2}^{t}\sum_{j=t-i+2}^{t}e^{\left(2X_{ij}\hat{\beta}\right)}\left\{g_m\left(2\left[1-X_{ij}(X'X)^{-1}X_{ij}'\right]s^2\right)\right.\\ \left. - g_m\left(\left[1-2X_{ij}(X'X)^{-1}X_{ij}'\right]s^2\right)\right\}. \quad (2.8)\end{aligned}$$

The term inside the sum of (2.8), $\widetilde{Var[Z_{ij}]}$, is derived similarly to (2.4).

Verrall [1] also found prediction intervals for total outstanding claims. A 95% upper confidence bound on R implies Prob$(R \leq \tilde{R} + k) = .95$ where $E[\tilde{R}] = R$.

Straightforward algebra further implies that Prob$(R - \tilde{R} \leq k) = .95$, where $E[R - \tilde{R}] = E[R] - E[\tilde{R}] = 0$. Similarly, $Var[R - \tilde{R}] = Var[R] + Var[\tilde{R}]$ due to the assumption of independent claims. Since R and $\tilde{R}$ are combinations of a large number of lognormal random variables, a normal approximation for the 95% confidence interval gives $\tilde{R} + 1.645\sqrt{MSE} = \tilde{R} + 1.645\sqrt{Var[R] + Var[\tilde{R}]}$.

2.2 ESTIMATION OF DEVELOPMENT FACTORS

Recall that development factors, also referred to as link ratios, are ratios of claims in successive delay years. The link ratio estimate, given in 1.2, allows for a forecast of $E[C_{ij}|j$ past values of $C_{ij}]$. When estimating outstanding claims, unbiased estimators are necessary. When comparing several sets of runoff patterns, unbiasedness is not critical. The maximum likelihood estimator (MLE) can provide a good, though biased estimate. Consider development factors λ_j and proportions of ultimate claims S_j. As mentioned above from Kremer's [13] paper, $S_1 = \frac{1}{\sum_{i=2}^{t} \lambda_i}$ and $S_j = \frac{\lambda_j - 1}{\sum_{i=j}^{t} \lambda_i} = \frac{\lambda_j - 1}{\prod_{i=2}^{t} \lambda_i}$. Additionally, Kremer showed that $S_j = \frac{e^{\beta_j}}{\sum_{i=1}^{t} e^{\beta_t}}$, where the β_i are the column parameters and $\beta_1 = 0$. Verrall [17] had previously shown that $\lambda_j = 1 + \frac{e^{\beta_j}}{\sum_{i=1}^{j-1} e^{\beta_i}}$. For both S_j and λ_j, he obtained maximum likelihood estimates $\hat{S}_j$ and $\hat{\lambda}_j$ by plugging in the MLE estimates β_j and $\beta_t = 0$. If the variance-covariance matrix of β is $Var[\beta]$, then the variance-covariance matrix of link ratios, λ_j, and proportions of ultimate claims, S_j, are as follows:

$$Var[\lambda] = \left(\frac{\partial \lambda}{\partial \beta}\right) Var[\beta] \left(\frac{\partial \lambda}{\partial \beta}\right),$$
$$Var[S] = \left(\frac{\partial S}{\partial \beta}\right) Var[\beta] \left(\frac{\partial S}{\partial \beta}\right),$$

where λ is a vector of link ratios $\lambda = [\lambda_2, \ldots, \lambda_t]'$, S is a vector of ultimate claims proportions $S = [S_1, \ldots, S_t]'$, and β is the parameter vector described in the general linear model, Equation 1.5. In addition, Verrall showed the following result:

$$\left(\frac{\partial \lambda}{\partial \beta}\right)_{jk} = \frac{\partial \lambda_j}{\partial \beta_k} = \begin{cases} 0 & = 0 & k > j \\ \frac{e^{\beta_j}}{\sum_{i=1}^{j-1} e^{\beta_t}} & = \lambda_j - 1 & k = j \\ \frac{e^{\beta_j} e^{\beta_k}}{\sum_{i=1}^{j-1} e^{\beta_i}} & = -(\lambda_j - 1)(\lambda_k - 1) & k < j \end{cases}$$

$$\left(\frac{\partial S}{\partial \beta}\right)_{jk} = \frac{\partial S_j}{\partial \beta_k} = \begin{cases} \frac{-e^{\beta_j} e^{\beta_k}}{\sum_{i=1}^{t} e^{\beta_i}} & = -S_j S_k & k \neq j \\ \frac{e^{\beta_j}\left(\sum_{i=1}^{t} e^{\beta_i} - e^{\beta_i}\right)}{\sum_{i=1}^{t} e^{\beta_i}} & = S_j(1 - S_j) & k = j. \end{cases}$$

3 BAYESIAN METHODOLOGY

3.1 BAYESIAN ANALYSIS

Let $X = (X_1, \ldots, X_n)$ be a random sample, $x = (x_1, \ldots, x_n)$ be observed data drawn from a population, and θ be a parameter describing risk characteristics within the population. Risk parameter θ can be accident rating class, age, or residence in a geographic location prone to tornadoes or hurricanes. In Bayesian analysis, instead of making inference about the parameter θ using the observed data, parameter θ is assumed to be random with a probability distribution called the *prior distribution*, $\pi(\theta)$, which is based on an analyst's belief about the population prior to taking the data sample $x = (x_1, \ldots, x_n)$. Then after observing the data, the prior distribution $\pi(\theta)$ is updated based on the sample using Bayesian technique and called the *posterior distribution*. The posterior distribution, the conditional distribution of θ given the sample x, is:

$$\pi(\theta|x) = \frac{f(x,\theta)}{f(x)} = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta} \tag{3.1}$$

where $f(x|\theta)$ is the *sampling distribution.* The distribution of claims or losses X given risk parameter θ is $f(x|\theta)$. Because $\int f(x|\theta)\pi(\theta)d\theta$ is independent of θ, then we have $\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$. See Casella [18] for further development of classical Bayesian analysis.

Assume we wish to set a rate to cover any new claims X_{n+1} using the *predictive distribution* $f(X_{n+1}|x)$; that is, the conditional probability of a new observation X_{n+1} given the data x. The risk parameter, θ, is unknown. Experience and policyholders corresponding to different exposure periods are assumed independent. The observation random variables $X_1, X_2, \ldots, X_n, X_{n+1}$ conditional on θ are independent but not necessarily identically distributed, where $X_j = f(x_j|\theta)$ for $j = 1, 2, \ldots, n, n+1$.

Our interest lies in $X_{n+1}|\Theta = \theta \sim f(x_{n+1}|\theta)$. Since θ is unknown, conditioning on θ is not possible. The next best thing is to condition on the known value x, using the predictive distribution $f(x_{n+1}|x)$. This distribution is relevant for risk analysis, management, and decision making (cf. Klugman [19]). The predictive distribution can then be computed as:

$$f_{X_{n+1}|X=x}(x_{n+1}|x) = \int \underbrace{f_{X_{n+1}|\Theta}(x_{n+1}|\theta)}_{\substack{\text{PDF of new observation}\\ \text{given the parameter value}}} \underbrace{\pi_{\Theta|X=x}(\theta|x)}_{\substack{\text{posterior}\\ \text{distribution}}} d\theta. \tag{3.2}$$

The derivation of (3.2) is as follows (cf. Klugman [19] and Weishaus [20]):

$$\begin{aligned} f(x,\theta) = \quad f(x_1, x_2, \ldots, x_n|\theta)\pi(\theta) &= \left\{ \prod_{j=1}^{n} f(x_j|\theta) \right\} \pi(\theta) \\ f(x_1, x_2, \ldots, x_n) &= \int \prod_{j=1}^{n} f(x_j|\theta)\pi(\theta)d\theta \\ f(x_1, x_2, \ldots, x_{n+1}) &= \int \prod_{j=1}^{n+1} f(x_j|\theta)\pi(\theta)d\theta \\ f(x_{n+1}|x) &= \frac{\int \prod_{j=1}^{n+1} f(x_j|\theta)\pi(\theta)d\theta}{f(x)} = \frac{\int f(x_{n+1}|\theta) \prod_{j=1}^{n} f(x_j|\theta)\pi(\theta)d\theta}{f(x)}. \end{aligned}$$

Since the posterior distribution of θ given x is in the form of (3.1), then

$$f(x_{n+1}|x) = \frac{\int f(x_{n+1}|\theta)f(x,\theta)d\theta}{f(x)} = \int f(x_{n+1}|\theta)\pi(\theta|x)d\theta,$$

proving Equation (3.2).

One determines the *Bayesian premium*, the predictive expected value, using both

$$E[X_{n+1}|X = x] = \int x_{n+1} f(x_{n+1}|x)dx_{n+1}$$

and (3.3)

$$E[X_{n+1}|X = x] = \int \mu_{n+1}(\theta)\pi(\theta|x)d\theta$$

(cf. Weishaus [20]). The motivation for (3.3) is that $X = x$ is observed for a policyholder and we wish to predict its mean, X_{n+1}. If θ is known, the *hypothetical mean* (individual premium) is

$$\mu_{n+1}(\theta) = E[X_{n+1}|\Theta = \theta] = \int x_{n+1} f(x_{n+1}|\theta) dx_{n+1}. \tag{3.4}$$

The *pure* (or collective) *premium*,

$$\mu_{n+1} = E[X_{n+1}] = E[E[X_{n+1}|\Theta]] = E[\mu_{n+1}(\Theta)], \tag{3.5}$$

is the mean of individual premiums (cf. Klugman [19] and Weishaus [20]). This premium is used when nothing is known about the policyholder, as it does not depend on the individual's risk parameter, θ, and does not use x, the data collected from the individual. Because θ is unknown, a Bayesian premium should be used. The derivation for the Bayesian premium (3.3) is as follows (cf. Klugman [19]):

$$\begin{aligned}
E[X_{n+1}|X = x] &= \int x_{n+1} f(x_{n+1}|x) dx_{n+1} \\
&= \int x_{n+1} \left\{ \int f(x_{n+1}|\theta) \pi(\theta|x) d\theta \right\} dx_{n+1} \\
&= \int \underbrace{\int x_{n+1} f(x_{n+1}|\Theta) dx_{n+1}}_{E[X_{n+1}|\Theta]} \pi(\theta|x) d\theta \\
&= \int \mu_{n+1}(\theta) \pi(\theta|x) d\theta.
\end{aligned}$$

The pure premium is $\mu_{n+1} = E[X_{n+1}]$, but policyholders prefer it when insurance companies charge the individual premium $\mu_{n+1}(\theta)$, which is the hypothetical mean using the hypothetical parameter θ associated with the policyholder. Since θ is unknown, this is impossible. Instead the company must condition on the past data

x, which leads to the Bayesian premium $E[X_{n+1}|x]$. The problem is that it is difficult to evaluate the Bayesian premium because it often requires numerical integration.

There are several purposes for Bayesian analysis. The statistician may have information from previously taken data which may help specify a prior parametric distribution for the parameters in the model before collecting a data sample. Verrall [8] finds results for Bayesian estimators when both prior information is known and no prior information is assumed. The Bayesian estimators have a simpler form than the unbiased estimators of mean claims derived using Finney's $g_m(t)$ (2.1). Since lognormal models are completely described by the mean and variance, one only needs the mean and variance of the prior distribution to make some inference. Assuming the data is lognormally distributed $Z_{ki|\theta} \sim$ lognorm (θ, σ^2), and the posterior distribution normally distributed $\theta|X \sim N(m, r^2)$, where σ^2 and r^2 are known, and X refers to the data; then $E[Z_{ki}|X] = e^{m+\frac{1}{2}\sigma^2+\frac{1}{2}r^2}$ and $Var[Z_{ki}|X] = e^{2m+\sigma^2+r^2}\left(e^{\sigma+r^2}-1\right)$. Under squared error loss, the Bayes estimate is the mean of the posterior distribution (cf. Klugman [19]). The ANOVA models Verrall [1] used to analyze claims runoff triangles can be represented in linear form (cf. Christensen [11]). Verrall analyze them from a Bayesian point of view, and developed the Bayes estimate of outstanding claims for business year i and Bayes estimate of the variance:

$$\widehat{E[Z_{ki}|X]}_{\text{Bayes}} = \sum_{j>n-i+1} E[Z_{kij}|X],$$

$$\widehat{Var[Z_{ki}|X]}_{\text{Bayes}} = \sum_{j>n-i+1} \left\{ Var[Z_{kij}|X] + 2\sum_{k>j} Cov(Z_{ij}, Z_{ik}|X) \right\}.$$

3.2 BÜHLMANN CREDIBILITY

Another reason for Bayesian analysis is developing estimators with a credibility theory interpretation. This could be used to give partial weight of experience to the population from which the sample is drawn and partial weight to the individual policyholder. If the statistician lacks prior information, the variance of the prior distribution could be set larger and still analyzed from a Bayesian perspective. As the variance (2.2) becomes large, the estimators tend to the ordinary least squares estimator (cf. Verrall [1]). Here the statistician could use empirical Bayes prior estimators.[1]

Difficulty with numerical evaluation of the Bayesian premium led to an alternate suggestion in 1967 by Bühlmann [21]: a weighted least squares estimate of the Bayesian predictive mean (cf. Weishaus [20]). The motivation is to use $f_{X_{n+1}|\Theta}(x_{n+1}|\theta)$ or hypothetical mean $\mu_{n+1}(\theta)$ for estimation of the following year's claims. The company has observed x, so approximate $\mu_{n+1}(\theta)$ by a linear function of past data. Estimators are restricted to ones of the form of a weighted average in order to minimize squared error loss. The weighted average is $\alpha_0 + \sum_{j+1}^{n} \alpha_j X_j$, where values of $\alpha_0, \ldots, \alpha_n$ are chosen to minimize the squared error loss

$$Q = E_{X,\Theta}\left[\left(\mu_{n+1}(\Theta) - \alpha_0 - \sum_{j=1}^{n} \alpha_j X_j\right)^2\right]. \tag{3.6}$$

Set $\frac{\partial Q}{\partial \alpha_0} = 0$. This implies $E[\mu_{n+1}(\Theta)] = \tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j E[X_j]$ where $\tilde{\alpha}_0, \ldots, \tilde{\alpha}_n$ are values that minimize Q. Then using iterated expectation, derive

$$E[X_{n+1}] = E\left[E[X_{n+1}|\Theta]\right] = E[\mu_{n+1}(\Theta)] = \underbrace{\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j E[X_j],}_{\text{Must be unbiased for } E[X_{n+1}]} \tag{3.7}$$

the first of two normal equations. Even though $E[X_{n+1}]$ must be unbiased, the credibility estimate may be a biased estimator $\mu_{n+1}(\theta) = E[X_{n+1}|\theta]$. The bias averages out over all $\theta_i \in \Theta$, so it is generally accepted to reduce the overall mean-squared

[1] This section on Bühlmann Credibility extensively uses Bühlmann [21], Klugman [19], and Weishaus [20]. Notation is consistent with Klugman. Other souces not cited here but referenced by Weishaus include Herzog's *Introduction to Credibility Theory* and Mahler-Dean's study notes.

error. For $\frac{\partial Q}{\partial \alpha_i} = 0$,

$$E[\mu_{n+1}(\Theta)X_i] = \tilde{\alpha}_0 E[X_i] + \sum_{j=1}^{n} \tilde{\alpha}_j E[X_i X_j] = \ldots = E[X_i X_{n+1}]. \tag{3.8}$$

This implies that

$$Cov[X_i, X_{n+1}] = \sum_{j=1}^{n} \tilde{\alpha}_j Cov[X_i, X_j] \text{ for } i = 1, \ldots, n. \tag{3.9}$$

This is the second of the two normal equations. They can be solved for $\tilde{\alpha}_0, \tilde{\alpha}_1, \ldots, \tilde{\alpha}_n$ to yield the *credibility premium* $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j$. One can also show that the solutions $\alpha_0, \ldots, \alpha_n$ satisfy the normal equations and minimize Q, making the credibility premium the best linear estimator of $E[X_{n+1}|\Theta]$ (the hypothetical mean), $E[X_{n+1}|X]$ (the Bayesian premium), and X_{n+1}. See Bühlmann [21] and Klugman [19].

For each policyholder (conditional on Θ), past losses are independent and identically distributed $X_i \overset{iid}{\sim} f(x|\theta)$. The hypothetical mean is denoted as $\mu(\theta) = E[X_j|\Theta = \theta]$, the *process variance* (hypothetical variance) as $Var[\theta] = Var[X_j|\Theta = \theta]$, the expected value of hypothetical means as $\mu = E[\mu(\Theta)]$, the expected value of the process variance as $\nu = E[Var(\Theta)]$, and variance of the hypothetical means as $a = Var[\mu(\Theta)]$ (cf. Klugman [19] and Weishaus [20]). In the case where nothing is known about the policyholder–i.e., there is no information about θ, use the collective premium $\mu = E[X_j] = E\left[E[X_j|\Theta]\right] = E[\mu(\Theta)]$. Then $Var[X_j] = E\left[Var[X_j|\Theta]\right] + Var\left[E[X_j|\Theta]\right] = E[\nu(\theta)] + Var[\mu(\Theta)] = \nu + a$. For $i \neq j$, $Cov[X_i, X_j] = a$. The credibility premium is $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j = Z\overline{X} + (1 - Z)\mu$ where $Z = \frac{n}{n+k}$ and $k = \frac{\nu}{a} = \frac{E[Var[X_j|\Theta]]}{Var[E[X_j|\Theta]]}$. Now the credibility factor $Z = \frac{n}{n+k}$ is known as the Bühlmann Credibility Factor (cf. Klugman [19] and Weishaus [20]).

The usefulness of Bühlmann's solution is found in $\lim_{n\to\infty} Z = 1$, so the credibility factor gives more weight to $\overline{X}$ than to μ as more past data accumulates (cf. Klugman

[19]). This keeps the policyholder invested in keeping their own premium down, but also gives the insurance company the collective premium as a good starting point. If $\mu(\Theta) = E[X_j|\Theta]$ has small variability – that is, it does not vary to a great extent with values of Θ – then $\nu = E[v(\Theta)]$ is large relative to $a = Var[\mu(\Theta)]$. This causes $k = \frac{\nu}{a}$ to be large and $Z = \frac{n}{n+k}$ to be closer to zero. For a homogeneous population, μ is of more value in helping to predict X_{n+1}. For a heterogeneous population, the opposite happens: a is large relative to ν, which makes k small and Z gives more weight to $\overline{X}$. This makes logical sense and is of great practical use. See the discussion in Klugman [19].

Bayesian linear estimation has similarity with credibility estimators of risk premiums. The credibility premium mentioned is linear, a weighted average. There is also an assumption that some parameters are exchangeable. This is very similar to the independent and identically distributed variable assumption, and arises naturally under the chain ladder model, where future samples will behave similarly to past samples (see the assumptions in section 1.1). This affects estimates due to shrinkage toward a central value. This gives stability, as shrinkage is the greatest when the number of observations is small (cf. Verrall [1]). Regard the runoff rows as a set of risks and use the Bühlmann risk credibility estimator. Verrall starts from the runoff triangles and proceeds to credibility formula via the linear models. The advantage is that the linear model approach produces estimates of standard errors of the estimates. Constraints are $\alpha_0 = \beta_0 = 0$ from the first stage distribution are retained. These restraints ensure a non-singular design matrix and introduce asymmetry into the prior distribution. One can use a constraint such as $\sum \alpha_i = \sum \beta_j = 0$ to avoid this (cf. Kremer [13] and Christensen [11] for introducing constraints to general linear models). One can also apply the constraint at the second stage and use the prior distribution $\alpha_i \sim N(0, \sigma_\alpha^2)$, for $i = 1, \ldots, t$.

Verrall's Bayesian estimates for the ANOVA model use a 2-stage Bayesian model, assuming some prior information or knowledge. Recall that the chain ladder model can be described in linear form as in Equation (1.5), where $Y|\beta \sim N(X\beta, \Sigma)$:

$$\overbrace{\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{bmatrix}}^{Y} = \overbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ \beta_3 \end{bmatrix}}^{X\beta} + \overbrace{\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \end{bmatrix}}^{\varepsilon}.$$

The log-linear model implies that $Y_i = \log Z_i$ where $Y_i \sim N(X_{ij}\beta, \sigma^2)$ and $Z_i \sim$ Lognormal$(X_{ij}\beta, \sigma^2)$, $E[Z_i] = e^{\left(X_{ij}\beta + \frac{\sigma^2}{2}\right)}$ and $Var[Z_i] = e^{\left(2X_{ij}\beta+\sigma^2\right)}(e^{\sigma^2} - 1)$. The prior information is quantified in the prior distribution on β: $\beta|\theta_i \sim N(I_i\theta_i, \Sigma_i)$. Similar sets of data may all give information on the individual parameters. Here, I_1 is an identity matrix, θ_1 is a vector containing the prior estimates, and Σ_1 is the diagonal variance-covariance matrix. When there are nonzero covariances, the prior distribution becomes $\beta|\theta_1 \sim N(\theta_1, \Sigma_1)$. If the errors are independent, then the variance-covariance matrix is the diagonal matrix $\Sigma = \sigma^2 I_n$, where I_n is the $n \times n$ identity matrix.

Verrall [1] showed that the Bayes estimate of the parameter vector is the solution $\tilde{\beta}$ of

$$\underbrace{(\sigma^{-2}X'X + \Sigma_1^{-1})}_{Var[\tilde{\beta}]^{-1}} \tilde{\beta} = \sigma^{-2}X'X\tilde{\beta} + \Sigma_1^{-1}\theta_1. \tag{3.10}$$

where $[\sigma^{-2}X'X + \Sigma_1^{-1}]^{-1}$ is the variance-covariance matrix of $\tilde{\beta}$. Then the Bühlmann credibility premium is $\tilde{\beta} = Z\tilde{\beta} + (1 - Z)\theta_1$ where Z is a $p \times p$ matrix. The credibility factor Z is $[\sigma^{-2}X'X + \Sigma_1^{-1}]^{-1}\sigma^{-2}X'X = \frac{\sigma^{-2}X'X}{\sigma^{-2}X'X+\Sigma_1^{-1}} = Z$. Recall that, the credibility factor $Z = \frac{m}{m+\frac{v}{a}} = \frac{1}{1+\frac{v}{am}}$, where $E[Var(\overline{X}|\theta)] = \frac{v}{m}$ and the variance of the hypothetical means is a. It is not possible to write a credibility formula separately for each factor in the form $\tilde{\alpha}_j = Z\tilde{\alpha}_j + (1-Z)\theta_j$ (cf. Klugman [19]). To estimate the variance σ^2, use $s^2 = \frac{(\nu - X\tilde{\beta})'(\nu - X\tilde{\beta})}{n+2}$ and substitute it into $[\tilde{\sigma}^{-2}X'X + \Sigma_1^{-1}]\tilde{\beta} = [s^{-2}X'X + \Sigma_1^{-1}]\tilde{\beta} = s^{-2}X'X\tilde{\beta} + \Sigma_1^{-1}\theta_1$ and solve for $\tilde{\beta}$ to get the Bayes estimates. To do this numerically, start with $s^2 = 0$ and iterate between $[s^{-2}X'X + \Sigma_1^{-1}]\tilde{\beta}$ and $s^{-2}X'X\tilde{\beta} + \Sigma_1^{-1}\theta_1$.[1]

There is also a generalization of the Bühlmann model, the Bühlmann-Straub model. For the Bühlmann model, past claims experience comprise independent and identically distributed components with respect to each past year, so there is no variation in exposure or size. The model does not reflect if the claims experience reflects only a portion of a year, if benefits change mid-way through the year, or if the size of a group in group insurance changes over time. The Bühlmann-Straub model is more appropriate when each X_j is an average of m_j independent (conditional on Θ) random variables, each with mean $\mu(\theta)$ and variance $v(\theta)$. For example the m_j independent random variables could be months the policy was in force, the number of individuals in a group, or the amount of premium income for a policy in year j. Both Klugman [19] and Weishuas [20] contain a discussion of Bühlmann-Straub.

Assume $X_1, \ldots, X_n$ are independent (cf. Bühlmann [21]), conditional on Θ with common mean $\mu(\theta) = E[X_j|\Theta = \theta]$ and conditional variances $Var[X_j|\Theta = \theta] = \frac{\nu(\theta)}{m_j}$, where m_j is a known constant measuring exposure, proportional to the size of the risk. As before, let $\mu = E[\mu(\Theta)]$, $\nu = E[Var(\Theta)]$, and $a = Var[\mu(\Theta)]$. For unconditional moments, $E[X_j] = \mu$, $Cov[X_i, X_j] = a$, and $Var[X_j] = E\left[Var[X_j|\Theta]\right] + Var\left[E[X_j|\Theta]\right]$ implies $Var[X_j] = E\left[\frac{\nu(\Theta)}{m_j}\right] + Var\left[\mu(\Theta)\right] = \frac{\nu}{m_j} + a$ (cf. Klugman [19]).

To obtain the credibility premium $\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j$, solve normal equations (3.7) and (3.9) to obtain $\tilde{\alpha}_0, \tilde{\alpha}_1, \ldots, \tilde{\alpha}_n$. Klugman [19] gives a very detailed explanation of the solution. Define $m = m_1 + m_2 + \cdots + m_n$ to be the total exposure. Recall that the unbiasedness Equation (3.7) is $E[X_{n+1}] = \tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j E[X_j]$ where $E[X_j] = E\left[E[X_j|\Theta]\right] = E[\mu(\Theta)] = \mu$. This implies that $\mu = \tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j \mu$, which implies $\sum_{j=1}^{n} \tilde{\alpha}_j = 1 - \frac{\tilde{\alpha}_0}{\mu}$. Also recall that $Cov[X_i, X_j] = \sum_{j=1}^{n} \tilde{\alpha}_j Cov[X_i, X_j]$ for $i = 1, \ldots, n$ (3.9). Set

$$a = Cov[X_i, X_{n+1}] = \sum_{\substack{j=1 \\ j \neq i}}^{n} \tilde{\alpha}_j a + \tilde{\alpha}_i (a + \frac{\nu}{m_i}) = \sum_{j=1}^{n} \tilde{\alpha}_j a + \frac{\nu \tilde{\alpha}_i}{m_i}.$$

This implies that

$$\tilde{\alpha}_i = \frac{a}{\nu} m_i \left(a - \sum_{j=1}^{n} \tilde{\alpha}_j \right) = \frac{a}{\nu} m_i \left(\frac{\tilde{\alpha}_0}{\mu} \right)$$

and (3.11)

$$a - \frac{\tilde{\alpha}_0}{\mu} = \sum_{j=1}^{n} \tilde{\alpha}_j = \sum_{i=1}^{n} \tilde{\alpha}_i = \frac{a}{\nu} \frac{\tilde{\alpha}_0}{\mu} \sum_{i=1}^{n} m_i = \frac{a \tilde{\alpha}_0 m}{\mu \nu}.$$

Equations (3.11) result in closed form solutions (3.12) to the unbiased normal equation:

$$\text{Normal Equation Solutions:} \quad \begin{cases} \tilde{\alpha}_j &= \frac{a\tilde{\alpha}_0}{\mu\nu} m_j &= \frac{m_j}{m + \frac{v}{a}} \\ \tilde{\alpha}_0 &= \frac{\mu}{1 + \frac{am}{\nu}} &= \frac{\frac{\nu}{a}}{m + \frac{\nu}{a}} \mu. \end{cases} \tag{3.12}$$

Let $k = \frac{\nu}{a}$, $Z = \frac{m}{m+k}$, and $\overline{X} = \sum_{j=1}^{n} \frac{m_j}{m} X_j$ as before. Then the Bühlmann-Straub credibility premium can be derived from the unbiased normal equation, by plugging solutions (Equation (3.12)) into the Bühlmann-Straub credibility premium, as shown in Klugman [19] and Bühlmann [21]:

$$\begin{aligned}
\tilde{\alpha}_0 + \sum_{j=1}^{n} \tilde{\alpha}_j X_j &= \frac{\frac{\nu}{a}}{m + \frac{\nu}{a}}\mu + \sum_{j=1}^{n} \frac{m_j}{m + \frac{v}{a}} X_j \\
&= \frac{k}{m+k}\mu + \sum_{j=1}^{n} \frac{m_j}{m+k} X_j \\
&= Z\overline{X} + (1 - \overline{Z})\mu. \qquad (3.13)
\end{aligned}$$

Note that (3.13) is still in the form of credibility premium $P_c = Z\overline{X} + (1-Z)M$. The Bühlmann-Straub credibility factor Z depends on m, the exposure associated with the policyholder. $\overline{X}$ is the weighted average of X_j, the average loss of m_j group members in year j, with weights proportional to m_j so $m_j X_j$ is the total loss of the group in year j. $\overline{X}$ is then the overall average loss per group member over the n years. The credibility premium to charge the group in year $n+1$ is then $P_c = m_{n+1}[Z\overline{X}+(1-Z)\mu]$, where m_{n+1} is the number of group members in year $n+1$.

For a single observation $\overline{x}$, the process variance is $Var[\overline{x}|\theta] = \sum_{j=1}^{n} \frac{m_j^2}{m^2}\frac{\nu(\theta)}{m}$, which implies $E[V[\overline{x}|\theta]] = \frac{\nu}{m}$. The variance of hypothetical means is a, which implies $k = \frac{\nu}{am}$. Finally, the Bühlmann-Straub Credibility Factor $Z = \frac{a}{a+\frac{\nu}{am}} = \frac{m}{m+\frac{\nu}{a}}$, with weights inversely proportional to the conditional variance of each X_j. The assumptions in Bühlmann-Straub are better than previous models, but are too restrictive to accurately represent reality. Large risks generally do not behave as independent aggregations of small risks; they are far more variable. The model can be generalized by letting the variance $\mu(\Theta)$ depend on the exposure. This may be reasonable if size of a risk affects its tendency to produce claims different from the mean.

Exact credibility occasionally arises in the Bühlmann-Straub situation. When the Credibility Premium (using best linear approximation) and Bayesian Premium (using squared error loss) are equal, the approximation is referred to as exact – or exact credibility. Verrall's paper did not delve into this topic any further.

3.3 CREDIBILITY THEORY

Limited[2] fluctuation credibility theory is a method for assigning full or partial credibility to a policyholder's experience. It may be that the remaining weight comes from some other information, such as occupation.[3] Greatest accuracy credibility theory was formalized by Bühlmann [21] in 1967. He used least squares estimation, which relies on a geometric, not statistical, argument (cf. Christensen [11]). The unknown parameters can be estimated by data using nonparametric estimation or semi-parametric estimation upon assuming a particular distribution.

One can either assume X_j are the number of claims or losses experienced in period $j \in \{1, 2, \ldots, n\}$ of a policyholder or that X_j is the experience from the j^{th} policy group or j^{th} member of a group or class. Assumptions are that X_j is homogenous (cf. Bühlmann [21]) or as Klugman [19] puts it, stable over time across members of the group or class. Mathematically, $E[X_j] = \xi$; and that X_j are the same for all j, $Var[X_j] = \sigma^2$. Past experience is collected, and the goal is to decide the value of ξ. Then $\bar{X} = \frac{x_1+x_2+\ldots+x_n}{n}$ and $E[\bar{X}] = \xi Var[\bar{X}] = \frac{\sigma^2}{n}$ (cf. Bühlmann).

Let M be the manual premium. This means that if one ignored the past data (assumed no credibility), one would charge M based on past experience from similar (not individual) policyholders. Partial credibility involves choosing a combination of M and $\bar{X}$, similar to choosing between the individual and collective premium in section 3.1. To choose between $\bar{X}$ and M, lean toward M if $\bar{X}$ is variable (has a large σ^2) but lean toward $\bar{X}$ if it is less variable and more stable. If there is reason to believe the policyholder will behave differently than the group which produced M, more weight should be given to $\bar{X}$. For full credibility, one relies entirely on past data, but there are certain conditions which must apply (cf. Klugman [19]).

[2]This section also makes wide use of Bühlmann [21] and Klugman [19], following the last section in using Klugman's notation.

[3]Mowbray (1914) first suggested it in connection with worker's compensation insurance.[19]

Stability means the difference between $\bar{X}$ and ξ is small relative to ξ with high probability. Klugman [19] derived a mathematical expression for full credibility to describe this situation. Let $\sqrt{\lambda_0} > 0$ and $0 < p < 1$; then,

$$\text{Prob}(-\sqrt{\lambda_0}\xi \leq \bar{X} - \xi \leq \sqrt{\lambda_0} * \xi) \geq p \quad \Rightarrow \quad \text{Prob}\left(\left|\frac{\bar{X}-\xi}{\sigma\sqrt{n}} \leq \frac{\sqrt{\lambda_0}\xi\sqrt{n}}{\sigma}\right|\right) \geq p.$$

If we let

$$y_p = \inf_y \left\{ P\left(\left|\frac{\bar{X}-\xi}{\sigma\sqrt{n}}\right| \leq y\right) \geq p \right\}$$

where y_p is the smallest value of y which satisfies this equation, then Klugman's condition for full credibility is

$$\frac{\sqrt{\lambda_0}\xi\sqrt{n}}{\sigma} \geq y_p \quad \Rightarrow \quad \frac{\sqrt{\lambda_0 n}}{y_p} \geq \frac{\sigma}{\xi}$$

where $\frac{\sigma}{\xi}$ is the coefficient of variation. Since $Var[\bar{X}] = \frac{\sigma^2}{n} \leq \frac{\lambda_0 \xi^2}{y_p^2}$, then the minimum sample size is

$$n \geq \left(\frac{\sigma}{\xi}\right)^2 \frac{y_p^2}{\lambda_0}$$

where λ is the exposure factor (see Equation (1.1)) required for full credibility. In other words, n must be at least this big to have full credibility. Klugman [19] discusses how to approximate $\bar{X}$ by the normal distribution $\frac{(\bar{X}-\xi)}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$. Then y_p is the $\left(\frac{p+1}{2}\right)^{\text{th}}$ percentile of the standard normal distribution. If the number of claims are being considered, $E[X_j] = \lambda$. However, if as in Verrall's [1] paper, the total claim amount is being considered, $E[X_j] = \lambda \times (\text{ claim size distribution}) = \lambda f(x)$. This should not be seen as unusual given the similarity to Verrall's estimate for expected ultimate loss (1.3). Also see Schmidt [3] for estimation of mean claims using development factors.

Partial credibility means the sample size n may not be large enough for full credibility, so the credibility premium is a mixture of $\bar{X}$ and M. Then the credibility premium is $P_c = Z\bar{X} + (1 - Z)M$ where $Z \in [0, 1]$ is the credibility factor. The Bühlmann [21] credibility factor is $Z = \frac{n}{n+k}$. Klugman [19] gives another formula for the Bühlmann credibility factor by controlling the variance of the credibility premium P_c.

$$\frac{\xi^2}{\lambda_0} = Var[P_c] = Z^2 \frac{\sigma^2}{n} \quad \Rightarrow \quad Z = \min\left\{\frac{\xi}{\sigma}\sqrt{\frac{n}{\lambda_0}}, 1\right\}$$

Greatest accuracy credibility has similar assumptions (cf. Klugman [19]). There are n exposure units of past claims $x_1, x_2, \ldots, x_n$ and a manual rate $M = \mu$ for the policyholder. When past experience indicates μ may not be appropriate as the manual rate (i.e., $\bar{X}$ and $E[X]$ could be very different from μ), then it may create a problem. The question is, should next year's net premium per exposure unit be based on μ, $\bar{X}$, or a combination of the two? One must decide if the policyholder different from assumptions made while calculating μ, or is random chance the difference between μ and $\bar{X}$.

The assumptions for greatest accuracy credibility are as follows (cf. Klugman [19]): The policyholder is in a homogenous risk class, based on underwriting. The rating class is characterized by a risk parameter, θ, which varies by policyholder. There is a probability mass (or density) function, $\pi(\theta)$, which varies across the rating class. If θ is a scalar parameter, $\Pi(\theta) = \text{Prob}(\Theta \leq \theta)$, which is the proportion of policyholders with a risk parameter less than or equal to θ. The last assumption that $\pi(\theta)$ is known, can be relaxed but the relevant characteristics of the risk structure $\pi(\theta)$ can be found within the population.

3.4 EMPIRICAL BAYES

Verrall's [8] empirical Bayes estimates for the chain ladder linear model uses a 3-stage Bayesian model, with an improper or vague prior distribution.[1] This method allows for the possibility that there may be no prior information on which to base a prior parametric distribution. The Bühlmann parameters are instead estimated entirely from empirical data. This is possible because Bühlmann credibility is based on least squares estimation, a geometric and not statistical method (cf. Weishaus [20]). The method uses an improper prior distribution (a distribution, the CDF of which may not sum to 1) at the third stage for row parameters and an improper prior distribution at the second stage for overall mean and column parameters. The same assumptions for overall mean and column parameters apply as for maximum likelihood estimation.

Those assumptions for maximum likelihood estimation are that the row parameters are iid (same as in credibility theory when assigning risk parameters to each risk) (cf. Verrall [8]). Estimates produced combine information from not just rows but the triangle as a whole. The prior distribution (second stage) is estimated from the data, which means the estimates are empirical Bayes estimates. The linear model for the chain ladder method is $y|\beta \sim N(X\beta, \sigma^2 I)$ with constraint $\alpha_1 = \beta_1 = 0$ as before. Errors are assumed to be iid. As in credibility theory, a structure is put onto row parameters $\alpha_2, \alpha_3, \ldots, \alpha_t$ - assume these are iid. For overall mean μ and column parameters $\beta_2, \beta_3, \ldots, \beta_t$, the same assumptions apply as for MLE, but the estimators are different because of the row parameter treatment (cf. Verrall [1]).

Verrall [8] proved the following result for 3-stage models (using improper prior) in section 2.2 of his earlier 1990 paper. Define the prior distribution for a vector of hyper-parameters: $\theta_1|\theta_2 \sim N(I_2\theta_2, \Sigma_2)$, where θ_2 is a p_2-dimensional vector, I_2 is a $p_1 \times p_2$ matrix, Σ_2 is a $p_1 \times p_1$ matrix. Also let D refer to the data. Verrall assumes all the parameters follow a parametric distribution, with the following assumptions:

$$y|\beta \sim N(X\beta, \Sigma), \qquad \beta|\theta_1 \sim N(I_1\theta_1, \Sigma_1), \text{ and } \qquad \theta_1|\theta_2 \sim N(I_2\theta_2, \Sigma_2).$$

Those further imply the posterior distribution of β is $\beta|y,\theta_2 \sim N(Dd, D)$, $D^{-1} = X'\Sigma^{-1}X + [\Sigma_1 + I_1\Sigma_2 I_1']^{-1}$ and $d = X'\sigma^{-1}y + [\Sigma_1 + I_1\Sigma_2 I_1']^{-1} I_1 I_2 \theta_2$. The Bayesian estimate (posterior mean) is a weighted average of the maximum likelihood estimator and prior mean (cf. Verrall, and Weishaus [20]):

$$\hat{\beta} = \left[X'\Sigma^{-1}X + (\Sigma_1 + I_1\Sigma_2 I_1')^{-1}\right]^{-1} \left[X'\Sigma^{-1}X\hat{\beta} + (\Sigma_1 + I_1\Sigma_2 I_1')^{-1} I_1 I_2 \theta_2\right] \quad (3.14)$$

which can be viewed as a credibility formula with credibility factor

$$Z = \frac{X'\Sigma^{-1}X}{X'\Sigma^{-1}X + (\Sigma_1 + I_1\Sigma_2 I_1')^{-1}}. \quad (3.15)$$

The weight given to the MLE depends on the inverse of the dispersion matrix of $\hat{\beta}$, which is $X'\Sigma^{-1}X$. Verrall [8] uses a vague third-stage prior with the 3-stage Bayesian model. With the same three assumptions as before and an additional assumption of $\Sigma_2^{-1} = 0$, the posterior distribution of $\beta|y$ is $\beta|y \sim N(D_0 d_0, D_0)$ where $D_0^{-1} = \underbrace{X'\Sigma^{-1}X}_{\sigma^{-2}X'X} + \Sigma_1^{-1} - \Sigma_1^{-1} I_1 (I_1'\Sigma_1^{-1} I)^{-1} I_1^{-1} \Sigma_1^{-1}$ and $d_0 = \underbrace{X'\Sigma^{-1}y}_{\sigma^{-2}X'y}$. These are the parameters used for empirical Bayes estimation of the parameters. They have a credibility theory interpretation similar to the estimators used in the premium setting by Bühlmann.

For the prior distribution, $\beta|(\omega,\theta,\xi) \sim N(I_1\theta_1, \Sigma_1)$, where

$$I_1\theta_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & 1 & & \vdots \\ \vdots & \vdots & \vdots & \\ 0 & 1 & 0 & \\ \vdots & 0 & 1 & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega \\ \theta \\ \xi_2 \\ \vdots \\ \xi_t \end{bmatrix} \text{ and } \Sigma_1 = \begin{bmatrix} \sigma^2\mu & & & & & & \\ & \sigma^2\alpha & & & & & \\ & & \ddots & & & & \\ & & & \sigma^2\alpha & & & \\ & & & & \sigma^2\beta & & \\ & & & & & \ddots & \\ & & & & & & \sigma^2\beta \end{bmatrix}.$$

Here, Σ_1 is taken so $\sigma^{-2}\mu \to 0$ and $\sigma^{-2}\beta \to 0$. Let ψ be the mean of the common distribution of the row parameters $\alpha_2, \ldots, \alpha_t$. An improper prior distribution is used for ψ. As $\sigma^{-2}\mu \to 0$ and $\sigma^{-2}\beta \to 0$, the third stage distribution is not needed for ω and $\xi_2, \ldots, \xi_t$. So the model is a combination of 2nd (prior distribution) and 3rd (improper prior) models.

The Bayes estimate, $\hat{\beta}$, of β, was proved by Verrall in the appendix of his 1990 paper[8] to be

$$\hat{\beta} = \sigma^{-2}X'X + \begin{bmatrix} 0 & & & & & & \\ & \sigma^{-2}\alpha & & & & & \\ & & \ddots & & & & \\ & & & \sigma^{-2}\alpha & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}^{-1} \times$$

$$\left\{ \sigma^{-2}X'X\beta + \begin{bmatrix} 0 & & & & & & \\ & \sigma^{-2}\alpha & & & & & \\ & & \ddots & & & & \\ & & & \sigma^{-2}\alpha & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \tilde{\alpha}_0 \\ \vdots \\ \tilde{\alpha}_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\} \qquad (3.16)$$

where $\tilde{\alpha}_0 = \frac{1}{t-1}\sum_{i=2}^{t} \tilde{\alpha}_i$.

The credibility interpretation of Verrall's estimate is that the empirical Bayes estimates of the row parameter are in the general form of credibility estimates – a weighted average of the MLE estimates and average of the estimates from all the rows.

Note that $X'X$ is not a diagonal or block-diagonal matrix. Estimation of $\mu, \beta_2, \ldots, \beta_t$ is tied up in estimating $\alpha_2, \ldots, \alpha_t$ and the reverse is true as well. Changing the estimates of row parameters forces change in other estimates, so this makes sense. The weights depend on the precision of the estimates, as in credibility theory. As before, σ^2 and $\sigma^2\alpha$ are replaced by model estimators s^2 and $s^2\alpha$ where

$$s^2 = \frac{\nu\lambda + (\nu - X\hat{\beta})'(\nu - X\hat{\beta})}{n + \nu + 2}$$

and

$$s^2\alpha = \frac{\nu_\alpha\lambda_\alpha + \sum_{i=2}^{t}(\hat{\alpha}_i - \alpha_.)^2}{t + \nu_\alpha + 1},$$

and $\nu, \lambda, \nu_\alpha, \lambda_\alpha$ are set by the prior distribution of the variances. As before with $s^2 = 0$, iterate between the two equations above and the Bayes estimate $\hat{\beta}$. The empirical Bayes assumptions can also be applied to column parameters, though Verrall mentions it has little practical use.

4 CONCLUSIONS

4.1 SUMMARY

Insurance companies must ensure as first order of business that there are enough reserves to cover future claims. For this reason, it is useful to estimate the mean expected losses. The run-off triangle method estimates losses based on past claims data, and is the most widely used method for such estimation in property and casualty insurance. Relying on development factors which are proportions of claims from year to year, the logged data from the multiplicative model in the chain ladder technique is well suited to being expressed in general linear model form. However, the chain ladder model also has limitations. Where the loss development factors in successive accident years are inconsistent, the chain ladder model may not be well suited to the data (see also Sahasrabuddhe [2] for alternate models and Schmidt [3] for assumptions underlying various chain ladder techniques).

Based on the work of Finney [15] and Bradu and Mundlak [16], Verrall's ([8], [14], [17], and [1]) chain ladder technique developed unbiased estimators for the claims in each business year, expected ultimate claims, and the variance for each estimator. Verrall also found maximum likelihood estimates for the development factors by taking partial derivatives of the ultimate claims proportions and link ratios (cf. Kremer [13]). Much of the current chain ladder technique involves a biased estimation not too dissimilar from the "quick and dirty" method discussed in the first two chapters. Verrall's method, which developed unbiased estimates first for the mean expected claims and then for the total reserve, is a significant improvement over currently accepted methods. With a computer program to quickly calculate or estimate values for Finney's infinite $g_m(t)$ sum (2.1), Verrall's technique could have a wide range of industry applications in loss reserving.

Bayesian analysis allows the statistician to make inference about a population based on some previously known knowledge. It also allows the statistician to predict a new observation given past data. Verrall combined the technique with Bühlmann [21] credibility to develop an estimator of the following year's claims based on the previous year's claims. Bühlmann credibility can be used to combine a previously known mean or premium based on the population from which the sample is drawn with a mean or premium based on the available data. The chain ladder technique can be used to estimate loss development in cases where data has been collected from a population but the statistician has no information on which to base a parametric prior distribution (empirical Bayesian estimation).

4.2 FURTHER RESEARCH

As mentioned above, Verrall's unbiased estimator for claims and expected ultimate claims suffers from a burden of tedious calculation, making it difficult to apply directly. For real data applications, a run-off triangle could easily reach into hundreds or thousands of rows and columns, depending on the time unit chosen and length of time the statistician wishes to apply historical data. Industry also often uses inexpensive and widely available applications. Instead of a direct calculation of Finney's $g_m(t)$, it may be useful to instead be able to accurately estimate it using one of the functions native to, say, MS Excel.

Catastrophic losses are extremely large losses arising from a single event catastrophe – for example, all losses resulting from the tornado which leveled a large portion of the city of Joplin, Missouri USA. It would be of interest to compare the loss development pattern for both catastrophic and non-catastrophic losses. The loss development pattern for catastrophic losses may spike higher but does not necessarily develop over time. For non-catastrophic losses, the development pattern does not spike but gradually increases over time. It is a useful application of the chain ladder technique to investigate the development patterns for both catastrophic and non-catastrophic losses. Investigation into speed of claims adjustment could also have useful industry applications. As the chain ladder technique is a linear model, regression analysis and comparison of residuals could inform a recommendation about which chain ladder technique is appropriately suited to each loss type.

BIBLIOGRAPHY

[1] Richard J Verrall. Statistical methods for the chain ladder technique. In *Casualty Actuarial Society Forum*, volume 1, pages 393–446, 1994.

[2] Rajesh V Sahasrabuddhe. Using a claim simulation model for reserving & loss forecasting for medical professional liability. In *Casualty Actuarial Society E-Forum*, pages 307–327, Winter 2007.

[3] Klaus D Schmidt. Methods and models of loss reserving based on run-off triangles: a unifying survey. *Forum*, pages 269–317, Fall 2006. Casualty Actuarial Society.

[4] Klaus D Schmidt. A bibliography on loss reserving, November 2013.

[5] Klaus D Schmidt. Bornhuetter-Ferguson as a general principle of loss reserving. Lehrstuhl für Versicherungsmathematik, Technische Universität Dresden, March 2008.

[6] Klaus D Schmidt and Mathias Zocher. The Bornhuetter-Ferguson principle. *Variance*, 2(1):85–110, 2008. Casualty Acturial Society.

[7] Gerhard Quarg and Thomas Mack. Munich chain ladder: A reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses. *Variance*, 2(2):266–299, 2008.

[8] Richard J Verrall. Bayes and empirical bayes estimation for the chain ladder model. *ASTIN Bulletin*, 20(2):217–243, 1990.

[9] Björn Wiendorfer. A practical guide to the use of the chain-ladder method for determining technical provisions for outstanding reported claims in non-life insurance. Working Paper Series 77, Fachhochschule des BFI, Wien, October 2012.

[10] Greg C Taylor and Frank R Ashe. Second moments of estimates of outstanding claims. *Journal of Econometrics*, 23(1):37–61, 1983.

[11] Ronald Christensen. *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer Texts in Statistics. Springer, fourth edition, 2011.

[12] David G Kleinbaum, Lawrence L Kupper, Azhar Nizam, and Keith E Muller. *Applied Regression Analysis and Other Multivariable Methods*. Brooks/Cole, Cengage Learning, fourth edition, 2008.

[13] Erhard Kremer. IBNR-claims and the two-way model of anova. *Scandinavian Actuarial Journal*, 1982(1):47–55, 1982.

[14] Richard J Verrall. On the unbiased estimation of reserves from loglinear models. *Insurance: Mathematics and Economics*, 10(1):75–80, 1991.

[15] DJ Finney. On the distribution of a variate whose logarithm is normally distributed. *Supplement to the Journal of the Royal Statistical Society*, pages 155–161, 1941.

[16] Dan Bradu and Yair Mundlak. Estimation in lognormal linear models. *Journal of the American Statistical Association*, 65(329):198–211, 1970.

[17] Richard J Verrall. Chain ladder and maximum likelihood. *Journal of the Institute of Actuaries*, 118:489–499, 1991.

[18] George Casella and Roger L Berger. *Statistical Inference*. Duxbury Advanced Series. Wadsworth Group, second edition, 2002.

[19] Stuart A Klugman, Harry H Panjer, and Gordon E Willmot. *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc, fourth edition, 2012.

[20] Abraham Weishaus. *Construction and Evaluation of Actuarial Models*. Exam C/Exam 4. Actuarial Study Materials (A.S.M.), fourteenth edition, 2012.

[21] Hans Bühlmann. Experience rating and credibility. *ASTIN Bulletin*, 4(03):199–207, July 1967. `http://journals.cambridge.org/article_S0515036100008989`.

[22] Manolis Bardis, Ali Majidi, and Daniel Murphy. A family of chain-ladder factor models for selected link ratios. *Variance*, 6(2):143–160, 2012.

[23] Kenneth Benoit. Linear regression models with logarithmic transformations. Technical report, Methodology Institute, London School of Economics, March 2011.

[24] Susanna Björkwall, Ola Hössjer, Esbjörn Ohlsson, and Richard Verrall. A generalized linear model with smoothing effects for claims reserving. *Insurance: Mathematics and Economics*, 49(1):27–37, 2011.

[25] Susanna Björkwall. *Stochastic claims reserving in non-life insurance: Bootstrap and smoothing models*. PhD thesis, Stockholm University, 2011.

[26] Anthony Boulter and Dawson Grubbs. Late claim reserves in reinsurance. Technical report, Swiss Reinsurance Company, Zurich, 2000.

[27] Christian Braun. The prediction error of the chain ladder method applied to correlated run-off triangles. *ASTIN Bulletin*, 34:399–424, 2004.

[28] ST Boris Choy, Jennifer SK Chan, and Udi E Makov. Model selection for loss reserves: The growing triangle technique. *Cutting Edge*, pages 35–40, January 2009. `http://www.life-pensions.com`.

[29] David R Clark. Basics of reinsurance pricing. *CAS Study Note*, pages 41–43, 1996.

[30] Jacqueline Friedland. *Estimating Unpaid Claims Using Basic Techniques*. Casualty Actuary Society, 3rd edition, July 2010.

[31] Alois Gisler and Mario V Wüthrich. Credibility for the chain ladder reserving method. *ASTIN Bulletin-Actuarial Studies in Non Life Insurance*, 38(2):565, 2008.

[32] Angela Jeansonne. Loglinear models. `http://userwww.sfsu.edu/efc/classes/biol710/loglinear/Log%20Linear%20Models.htm`, September 2002.

[33] Stuart A Klugman. Credibility for classification ratemaking via the hierarchical normal linear model. In *Proceedings of the Casualty Actuarial Society*, volume 74, pages 272–321, 1987.

[34] Dennis V Lindley and Adrian FM Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–41, 1972.

[35] Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, 23(2):213–225, 1993.

[36] Thomas Mack. Which stochastic model is underlying the chain ladder method? *Insurance: mathematics and economics*, 15(2):133–138, 1994.

[37] Maria Dolores Martínez-Miranda, Jens Perch Nielsen, and Richard Verrall. Double chain ladder and Bornhutter-Ferguson. *North American Actuarial Journal*, 17(2):101–113, 2013.

[38] María Dolores Martínez Miranda, Jens Perch Nielsen, Richard Verrall, and Mario V Wüthrich. Double chain ladder, claims development inflation and zero-claims. Cass Business School, May 2013.

[39] Daniel M Murphy. Chain ladder reserve risk estimators. In *CAS E-Forum Summer*, 2007.

[40] Paulo JR Pinheiro, João Manuel Andrade e Silva, and Maria de Lourdes Centeno. Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, 70(4):701–714, 2003.

[41] Arthur E Renshaw. Chain ladder and interactive modelling.(claims reserving and GLIM). *Journal of the Institute of Actuaries*, 116(03):559–587, 1989.

[42] Arthur E Renshaw and Richard J Verrall. A stochastic model underlying the chain-ladder technique. *British Actuarial Journal*, 4(04):903–923, 1998.

[43] Rajesh Sahasrabuddhe. Principles of the chain-ladder method selecting and updating claims development factors. In *Casualty Actuarial Society E-Forum*, pages 569–579, Fall 2008.

[44] Magda Schiegl. A model study about the applicability of the chain ladder method, 2013.

[45] Adrian FM Smith. A general Bayesian linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 67–75, 1973.

[46] Gregory Clive Taylor and Great Britain. *Testing Goodness-of-fit of an Estimated Run-off Triangle*. School of Economic and Financial Studies, Macquarie University, 1975.

[47] Richard J Verrall and Z Li. Negative incremental claims: chain ladder and linear models. *Journal of the Institute of Actuaries*, 120(01):171–183, 1993.

[48] Mario V Wüthrich, Michael Merz, and Natalia Lysenko. Uncertainty of the claims development result in the chain ladder method. *Scandinavian Actuarial Journal*, 2009(1):63–84, 2009.

[49] Jing Yang. Interpreting coefficients in regression with log-transformed variables. StatNews Nr.83 Cornell Statistical Consulting Unit, Cornell University, `http://www.cscu.cornell.edu`, June 2012.

INDEX

VITA

Netanya Martin graduated from Truman State University with a B.S. in Mathematics and a B.A. in German in 2005. She served in the Army Reserve from 2004 until 2015, achieving the rank of Captain, and worked in the Department of Military Science - Army ROTC as the Recruiting Operations Officer from 2010 to 2013. She served as the vice president of the Actuarial Club, while earning a Graduate Certificate in Actuarial Science from Missouri S&T in May 2014. She interned at Reinsurance Group of America during the summer of 2014 and began work as an actuarial analyst in the Corporate Actuarial & Risk department of SCOR Global Life Americas Reinsurance Company in January 2015. She expects to graduate from Missouri S&T with a M.S. in Applied Mathematics with an Emphasis in Statistics in May 2015.