Casualty Actuarial Society
E-Forum, Winter 2019

The CAS *E-Forum*, Winter 2019

The Winter 2019 edition of the CAS *E-Forum* is a cooperative effort between the CAS *E-Forum* Committee and various CAS committees, task forces, working parties and special interest sections. This *E-Forum* contains Report 13 of the CAS Risk-Based Capital Dependencies and Calibration Working Party and the 2018 CAS Reserves Call Papers. (For Reports 1-12 of the CAS Risk-Based Capital Dependencies and Calibration Working Party, visit https://www.casact.org/pubs/forum/.) The Reserves Call Papers were presented at the Casualty Loss Reserve Seminar (CLRS), which was held September 5-7, 2018, in Anaheim, California.

Risk-Based Capital Dependencies and Calibration Research Working Party

Allan M. Kaufman, *Chairperson*

Karen H. Adams
Emmanuel Theodore Bardis
Jess B. Broussard
Robert P. Butsic
Pablo Castets
Damon Cham, *Actuarial Student*
Joseph F. Cofield
Jose R. Couret
Orla Donnelly
Chris Dougherty
Nicole Elliott
Brian A. Fannin
Sholom Feldblum
Kendra Felisky
Dennis A. Franciskovich
Timothy Gault
Dean Guo
Jed Nathaniel Isaman
Shira L. Jacobson
Shiwen Jiang
James Kahn
Alex Krutov
Terry T. Kuruvilla
Apundeep Singh Lamba
Giuseppe F. LePera
Zhe Robin Li
Lily (Manjuan) Liang
Thomas Toong-Chiang Loy
Eduardo P. Marchena
Mark McCluskey
James P. McNichols
Glenn G. Meyers
Daniel M. Murphy
Douglas Robert Nation
G. Chris Nyce
Jeffrey J. Pfluger
Yi Pu
Ashley Arlene Reller
David A. Rosenzweig
David L. Ruhm
Andrew Jon Staudt
Timothy Delmar Sweetser
Anna Marie Wetterhus
Jennifer X. Wu Jianwei
Xie Ji Yao
Linda Zhang
Christina Tieyan Zhou
Karen Sonnet, *Staff Liaison*

Committee on Ratemaking

Sandra J. Callanan, *Chairperson*
Ronald S. Lettofsky, *Vice Chairperson*

Morgan Haire Bugbee
Caryn C. Carmean
William M. Carpenter
James Chang
Sang Suk Cho
Donald L. Closter
Christopher L. Cooksey
Greg Frankowiak
Ray Yau Kui Ho
Anthony Hovest
Aditya Khanna
Duk Inn Kim
Dennis L. Lange
Shan Lin
Daniel W. Lupton
Harsha S. Maddipati
Joshua Jacob Newkirk
Allan I. Schwartz
Emily Ruth Stoll
Jane C. Taylor
Gary Joseph Wierzbicki
Todd F. Witte
Karen Sonnet, *Staff Liaison*

CAS *E-Forum*, Winter 2019

Table of Contents

For information on submitting a paper to the *E-Forum*, visit http://www.casact.org/pubs/forum/.

Risk-Based Capital Line of Business Diversification: Current RBC Approach vs. Correlation Matrix Approach

Report 13 of the CAS Risk-Based Capital (RBC) Research Working Parties
Issued by the RBC Dependencies and Calibration Working Party (DCWP)

Abstract: The NAIC RBC Formula treatment of line of business (LOB) diversification (referred to in this paper as the CoMaxLine% Approach) is very different from the Solvency II Standard Formula treatment. In this paper we show that, notwithstanding the differences, the NAIC RBC Formula, the correlation matrix approach used in Solvency II[1] and the Herfindahl-Hirschman Index (HHI), widely used in economics, all produce similar risk-based capital underwriting risk values, for most companies.

To the extent that there are differences between the CoMaxLine% and correlation matrix approaches, the differences are due, in part, to the fact that CoMaxLine% calculates diversification based on premium or reserve volume while the correlation matrix approach calculates diversification based on premium risk or reserve risk. To examine this feature of the RBC Formula, we also apply the CoMaxLine% idea to risk by LOB rather than volume by LOB. We refer to that as CoMaxLine%-Risk. The differences between CoMaxLine%-Risk and the correlation matrix approach are smaller than the differences to the RBC CoMaxLine% Approach.

This is one of several papers being issued by the Risk-Based Capital (RBC) Dependencies and Calibration Working Party.

Keywords: Risk-Based Capital, Capital Requirements, Analyzing/Quantifying Risks, Assessing/Prioritizing Risks, Integrating Risks, dependency, correlation.

1. INTRODUCTON

The Property & Casualty NAIC RBC Formula ("RBC Formula") has six main risk categories, R_0 – R_5. Underwriting (UW) risk is represented in two of these categories, R_4[2] and R_5, reserve risk and premium risk, respectively. In this work, we focus on the UW risk elements, R_4 and R_5. Following the RBC Formula, we calculate the UW portion of the

[1] Using a limited number of correlation matrix values, e.g., only 25% and 50% in the Solvency II Standard Formula and 25%, 50%, 75% and 100% in our RBC equivalent matrix.

[2] When applied, the pure reserve risk component is combined with a portion of the reinsurance credit risk component. This paper deals with the pure reserve risk component of R_4.

Company Action Level RBC Value[3,4] as the square root of R_4 squared plus R_5 squared[5] and refer to the resulting quantity as the RBC UW Risk Value.[6]

R_4 and R_5 are first calculated by line of business (LOB). The all-lines R_4, the reserve risk charge, is the sum of the R_4 risk charges by LOB, multiplied by a Loss Concentration Factor (LCF). The all-lines R_5, the premium risk charge, is the sum of the R_5 risk charges by LOB, multiplied by a Premium Concentration Factor (PCF).[7]

For each company, the LCF calculation uses the ratio of (a) the largest of the 19 LOB[8] reserves, to (b) the total all-lines reserves.[9] Similarly, for each company, the PCF calculation uses the ratio of (a) the largest of the 19 LOB written premiums, to (b) the total all-lines written premium.[10] The LCF and PCF are values between 0.0 and 1.0 that represent the degree of concentration across LOBs, within R_4 and R_5, respectively. A company with greater diversification across its LOBs will have smaller LCF and PCF values than a less diversified company.

We refer to this method of measuring concentration as the Company Maximum Line Percentage of Business or the "CoMaxLine% Approach." We refer to the ratios computed as the CoMaxLine%$_{PREMIUM}$ and the CoMaxLine%$_{RESERVES}$, or CoMaxLine% generically for either.

The CoMaxLine% Approach in the NAIC RBC Formula is very different in concept from the Solvency II Standard Formula correlation matrix approach. In this paper we show that,

---

[3] That is the Company Action Level RBC as if the R_0-R_3 and R_3-Reinsurance Credit Risk RBC values were zero.

[4] In all cases in the paper, when we refer to "RBC UW Risk Value" we refer to the Company Action Level RBC. The RBC value in the Annual Statement is the Authorized Control Level, equal to 50% of the Company Action Level.

[5] Note that we compare diversification formulas using the UW portion of RBC rather than the total RBC value. Had we compared using the total RBC value, the percentage differences between companies would have appeared smaller than the differences displayed in Tables 3-1, 3-2, and 3-3 below.

[6] The RBC Formula treats premium risk and reserve risk as independent risks. We are not testing alternatives to the way that the RBC Formula combines premium risk and reserve risk.

[7] The LCF and PCF are applied to the sum of the LOB RBC amounts, where those RBC amounts reflect the investment income offset, the own-company experience adjustment, and the loss sensitive contract adjustment.

[8] There are 22 LOBs in the Annual Statement Schedule P. In the RBC forms, those are consolidated into 19 LOBs. Other Liability Occurrence and Other Liability Claims-Made LOBs are combined and treated as one LOB. Products Occurrence and Products Claims-Made are combined and treated as one LOB. Reinsurance: nonproportional assumed property and reinsurance: nonproportional assumed financial LOBs are combined and treated as one LOB. NAIC, 2010, "Property and Casualty Risk-Based Capital Forecasting & Instructions." page 19.

[9] The reserves used to compute the ratio are the reserves for unpaid claims and claim expenses, net of reinsurance, as of the most recent year-end including both adjusting and other expenses and defense and cost containment expenses.

[10] The premiums used in this calculation are the most recent year's written premiums net of reinsurance.

notwithstanding the conceptual differences, the NAIC RBC Formula, the correlation matrix approach used in Solvency II and the Herfindahl-Hirschman Index (HHI), widely used in economics to measure concentration, produce similar RBC UW Risk Values, for most companies.

This paper is focused solely on a comparison of the RBC UW Risk Values produced by several methods of reflecting diversification among lines of business. In this paper we do not evaluate the CoMaxLine% parameters or the parameters for other methods of measuring concentration.[11]

In Section 2. we describe the alternative diversification approaches. In Section 3, we compare the UW Risk RBC Values, by company, that result from the different approaches.

1.1 Terminology, Assumed Reader Background and Disclaimer

This paper assumes the reader is generally familiar with the property/casualty RBC Formula.[12]

In this paper we use the term "diversification" rather than its complement[13] "concentration" unless the context makes the alternative clearer.

Although the term "multi-line insurance company" is commonly used to refer to an insurer that is well-diversified across LOBs, in this paper we will use the term more broadly to refer to any company for which the diversification credit is greater than zero.

References to "we" and "our" mean the principal authors of this paper.

The "working party" and "DCWP" refer to the CAS RBC Dependencies and Calibration Working Party.

The analysis and opinions expressed in this report are solely those of the principal authors, and are not those of the authors' employers, the Casualty Actuarial Society, or the American Academy of Actuaries.

Nether the authors nor DCWP make recommendations to the NAIC or any other body. This material is for the information of CAS members, policy makers, actuaries and others who might make recommendations regarding the future of the P&C RBC Formula. In particular,

[11] In DCWP Report 14 we evaluate the CoMaxLine% parameters.

[12] For a detailed description of the formula and its basis, see Feldblum, Sholom, NAIC Property/Casualty Insurance Company Risk-Based Capital Requirements, Proceedings of the Casualty Actuarial Society, 1996 and NAIC, Risk-Based Capital Forecasting & Instructions, Property Casualty, 2010.

[13] A company with a concentration ratio of 80% can equivalently be described as a having a diversification ratio of 20%, 100%-80%.

we expect that the material will be used by the American Academy of Actuaries.

This paper is one of a series of articles prepared under the direction of the DCWP.

2. Alternative Diversification Formulas

RBC Diversification Approach

The RBC Formula uses the CoMaxLine% Approach and a maximum diversification credit (MDC) of 30% to calculate PCFs and LCFs as follows:

$$PCF_{COMPANY} = 0.7 + 0.3 * CoMaxLine\%_{PREMIUM,\ COMPANY}$$

$$LCF_{COMPANY} = 0.7 + 0.3 * CoMaxLine\%_{RESERVES,\ COMPANY}$$

These can also be written as:

$$PCF_{COMPANY} = 1.0 - 0.3 * (1.0 - CoMaxLine\%_{PREMIUM,\ COMPANY})$$

$$LCF_{COMPANY} = 1.0 - 0.3 * (1.0 - CoMaxLine\%_{RESERVES,\ COMPANY})$$

Thus, the company diversification credit is 0.3 * (1 - CoMaxLine%).

For mono-line companies, CoMaxLine% and the PCF/LCF are 1.00. The maximum credit of 30% would be achievable only if there were an infinite number of LOBs. Since there are 19 statutory lines of business used in the RBC Formula the smallest value of CoMaxLine% is 1/19 = 5.3%, the smallest value of PCF or LCF is 71.6% (0.7 + 0.3 * 5.3%), and the maximum achievable diversification credit is 28.4%, (100% - 71.6%).

Alternatives to the CoMaxLine% Approach

Looking at the treatment of diversification in regulatory capital formulas developed in other regulatory regimes, the UK Individual Capital Adequacy Standard (UK ICAS) can be thought of as the simplest. In UK ICAS there is no premium or reserve risk diversification adjustment. Instead, LOB risk factors were selected to represent the LOB risk when combined with a typical LOB distribution.[14]

The CoMaxLine% Approach can be viewed as one step more complex than the UK ICAS in that it recognizes different levels of diversification.

From the risk theory perspective, the natural approach to diversification is to combine risk

[14] Solvency – Models, Assessment and Regulation, Arne Sandström, 2006, Taylor & Francis Group, LLC, p 161-164, http://docslide.us/documents/solvency-models-assessment-and-regulation.html;
Also at NAIC, SMI, Country Comparisons, UK,
http://www.naic.org/documents/committees_smi_int_solvency_uk.pdf

charges by LOB using correlation[15] factors between each pair of LOBs. Individual company economic capital models (called 'internal models' in Solvency II) often use this pairwise correlation matrix approach. The Solvency II Standard Formula uses the pairwise correlation matrix approach. The correlation matrix approach, if applied in the RBC Formula, would require 171 parameters since 19 LOBs are used. In contrast to the correlation matrix approach, the RBC Formula CoMaxLine% Approach might be described as simple, perhaps too simple, and ad hoc.

One difference between the CoMaxLine% Approach and the correlation matrix approach, as normally applied, is that the degree of diversification in the correlation matrix approach is based on risk by LOB while the degree of diversification in the CoMaxLine% Approach is based on volume (premium amount or reserve amount) by LOB. Therefore, as another alternative to CoMaxLine% and correlation matrix approaches, we also consider a CoMaxLine%-Risk Approach, in which we apply the CoMaxLine% Approach to LOB risk rather than LOB volume, when calculating the LCF and PCF for a company.[16]

Finally, the Herfindahl-Hirschman Index (HHI) is widely used by economists to measure concentration. HHI considers the relative proportions of all LOBs, the largest, second largest, third largest, and so on.[17] HHI is more complex than the CoMaxLine% Approach in that it recognizes the extent of diversification for the 2nd, 3rd, 4th, etc. largest LOBs.[18] HHI is simpler than the correlation matrix approach in that HHI does not recognize differences in the extent of the diversification between different pairs of LOBs.[19]

---

[15] We use the term correlation matrix approach to describe a factor method or copula method for computing total risk by combining several individual risks. In using the term, we do not intend to imply that the assumptions related to linear correlation are appropriate.

[16] For CoMaxLine%-Risk, as for CoMaxLine%, the risk charge after diversification equals the sum of the risk charges over all LOBs times the PCF and LCF determined using the risk version of CoMaxLine% for premium risk and reserve risk, respectively.

[17] HHI equals the sum of the squares of the LOB shares of total. For example, if there is only one LOB, HHI is 1.0, as is the case for the CoMaxLine%. With two lines split 25% and 75% HHI is 0.25^2 plus 0.75^2 or 0.625 compared to the CoMaxLine% of 0.750, i.e., HHI shows more diversification. With three lines split 50%, 25% and 25% HHI is 0.50^2 plus 0.25^2 plus 0.25^2 or 0.375, more diversification than the CoMaxLine% of 0.5. With two lines split 50% and 50% HHI and the CoMaxLine% are both 0.5.

[18] The HHI is sometimes applied to only the n-th largest segments, e.g., the degree of diversification among the top ten LOBs. The HHI index applied to the single largest segment would be very similar to the CoMaxLine%. HHI can be written as p_1^2+p_2^2+p_3^2…+p_n^2. The truncated HHI limited to one element would be p_1^2. CoMaxLine% is p_1. HHI is always less than or equal to CoMaxLine%.

[19] For HHI, as for CoMaxLine%, the risk charge after diversification equals the sum of the risk charges over all LOBs times the PCF and LCF determined using the HHI formula, separately for premium risk and reserve risk.

3. Effect of Alternative Diversification Formulas

We now look at the extent to which the different methods of measuring diversification for R_4 and R_5 produce different RBC UW Risk Values. For each company that filed a 2010 Annual Statement, we calculate the all-lines value for R_4 and for R_5 before diversification using the 2010 RBC Formula.[20] We then use each of the following approaches to calculate the effect of diversification across LOBs, arriving at R_4 and R_5, after diversification, for each company:

a. CoMaxLine% based on volume (as applied in the NAIC RBC Formula)
b. CoMaxLine%-Risk
c. Correlation matrix
d. HHI

Using the values of R_4 and R_5, after diversification, for each company, for each of the four approaches, we calculate the RBC UW Risk Value.[21] Appendix 1 provides more details regarding the data used and the simplifying steps taken in applying the RBC Formula with each of the four diversification approaches.

3.1 Correlation vs. CoMaxLine%

In this section, we compare the results of using the CoMaxLine% Approach (based on volume) to the results of using the correlation matrix approach.

To apply the correlation matrix approach, we construct a set of pairwise correlation factors, called a correlation matrix. Following the Solvency II approach, we use values of 25% or 50% for most of the 171 LOB-pairs.[22] For several LOB-pairs that we consider very highly correlated we select correlation factors of 75% or 100%.[23]

Appendix 1/Exhibit 1 shows our correlation matrix. Appendix 1/Exhibit 2 shows the Solvency II Standard Formula LOB correlation matrix, for comparison.

For each company with a 2010 Annual Statement, we apply both the CoMaxLine% Approach and the correlation matrix approach to produce the two alternative RBC UW Risk Values. The company-by-company differences between the two diversification approaches

[20] We calculate the Company Action Level of RBC.
[21] We are not testing alternatives to the way that the RBC Formula combines premium risk and reserve risk.
[22] "Advice for Band 2 Implementing Measures on Solvency II: SCR Standard Formula Article 111(d) Correlations," (former Consultation Paper 74), January 2010, pp 39-44. See Appendix 1 for further discussion of the origin of the Solvency II correlation matrix.
[23] We select pairwise correlations of 100% for claims-made and occurrence medical malpractice and for general liability, special liability and products liability. We select pairwise correlations of 75% between special property and homeowners, between private passenger automobile liability and automobile physical damage and between commercial automobile liability and automobile physical damage.

have two parts:

- the overall industry-wide difference, and
- the remaining difference for each individual company after normalizing to remove the industry-wide difference.

We measure the first part by computing the total US industry-wide RBC UW Risk Value that each approach produces, using the 30% MDC in the CoMaxLine% Approach and using the parameters specified in Appendix 1 / Exhibit 1 in the correlation matrix approach. We find that the industry-total RBC UW Risk Value is $106.2 billion with the CoMaxLine% Approach and $100.6 billion with the correlation matrix approach. We find that increasing the 30% MDC to 39.1% in the CoMaxLine% Approach decreases the RBC UW Risk Value to $100.6 billion, equal to the correlation matrix-based RBC UW Risk Value.[24]

In this analysis, we are more interested in the second part, the differences in diversification credit by company that remain after controlling for the overall effect on the total industry-wide RBC UW Risk Value. Therefore, we look at the company-by-company differences between the CoMaxLine% Approach with a MDC of 39.1%, and the correlation matrix approach using the parameters specified in Appendix 1 / Exhibit 1.

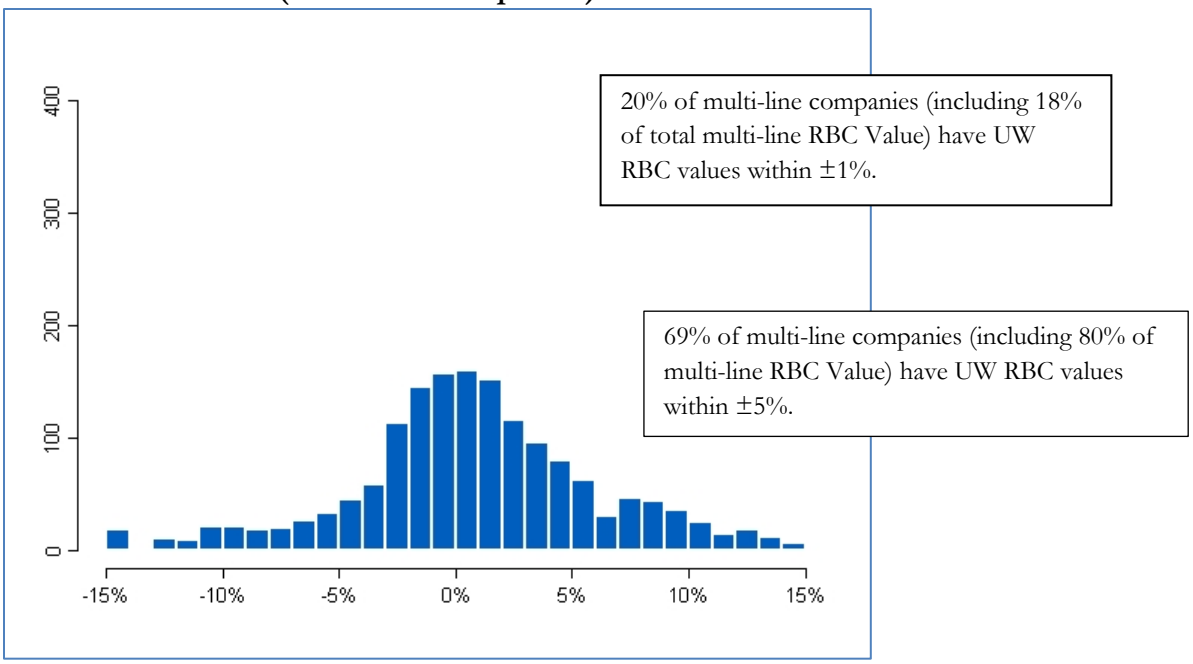
Looking at the differences, we observe a sizable number of cases where the UW risk values are the same regardless of the diversification structure. These zero differences arise for companies that have zero UW risk (i.e. due to having zero premium and reserves in all lines) and for mono-line companies.[25,26] We focus on multi-line companies, where the choice of diversification formula can affect the RBC UW Risk Value. The histogram in Table 3-1 below includes multi-line companies only and shows the distribution of percentage differences in RBC UW Risk Values by company.

[24] The CoMaxLine% Approach with a 30% MDC produces approximately the same total RBC as a correlation matrix with all pairwise correlations of 50%. Our selected correlation matrix has correlations at, generally, 50% or 25%. Thus, the average correlation in the matrix is lower than 50%. The resulting diversification is higher than the CoMaxLine% Approach with 30%. Therefore, an equivalent CoMaxLine% formula would need a MDC greater than 30%, as is the case.

[25] Including some companies that are so close to mono-line that the effect rounds to zero within $1k.

[26] We also remove some companies with significant negative premiums/reserves that would distort the comparisons among diversification methods.

Table 3-1
2010 RBC UW Risk Value Differences by Company[27]
Distribution of Number of Companies
Correlation matrix approach versus CoMaxLine% Approach (39.1% MDC)
(Multi-line Companies)



X-axis = Percentage difference between RBC UW Risk Values based on CoMaxLine% Approach and RBC UW Risk Values based on correlation matrix approach.
Y-axis = Number of companies, in buckets of 1% difference in RBC UW Risk Value.

We find that:

- For 33% of companies, with 3% of total industry-wide RBC UW Risk Value, the difference between diversification approaches is zero because they have zero UW risk (14.8%) or because they are mono-line (18.6%). These companies are excluded from the histogram.
- For 20% of the multi-line companies, with 18% of the industry-wide multi-line RBC UW Risk Value, the differences are less than ±1%.
- For 69% of the multi-line companies, with 80% of the industry-wide multi-line RBC UW Risk Value, the differences are less than ±5%.
- The differences are greater than 10% for only 10% of the multi-line companies constituting about 9% of the industry-wide multi-line RBC UW Risk Value.

[27] Positive differences represent companies for which the correlation matrix approach produces a higher RBC UW Risk Value than the CoMaxLine% Approach.

- Considering all companies, even those companies which are mono-line, or which have zero premium and reserves, we find that for 46% of all companies, with 20% of the total RBC UW Risk Value, the differences are less than ±1%. For 79% of all companies, with 79% of the total RBC UW Risk Value, the differences are less than ±5%.

Differences of 5% might be considered small as a practical matter. In addition, we consider the differences to be small for several statistical reasons. First, the differences are not large compared to the inherent accuracy of the risk factors which are used to calculate R_4 and R_5 for each individual LOB. Moreover, the systematic variation in LOB risk factors due to LOB-size, LOB-age, and other factors discussed in DCWP Reports 6-9 is larger than the variation shown here from using a different diversification approach. Finally, correlation matrix values have inherent uncertainty, particularly in that the values are largely calibrated by expert judgment with only limited data.

3.2 Correlation Matrix versus CoMaxLine%-Risk

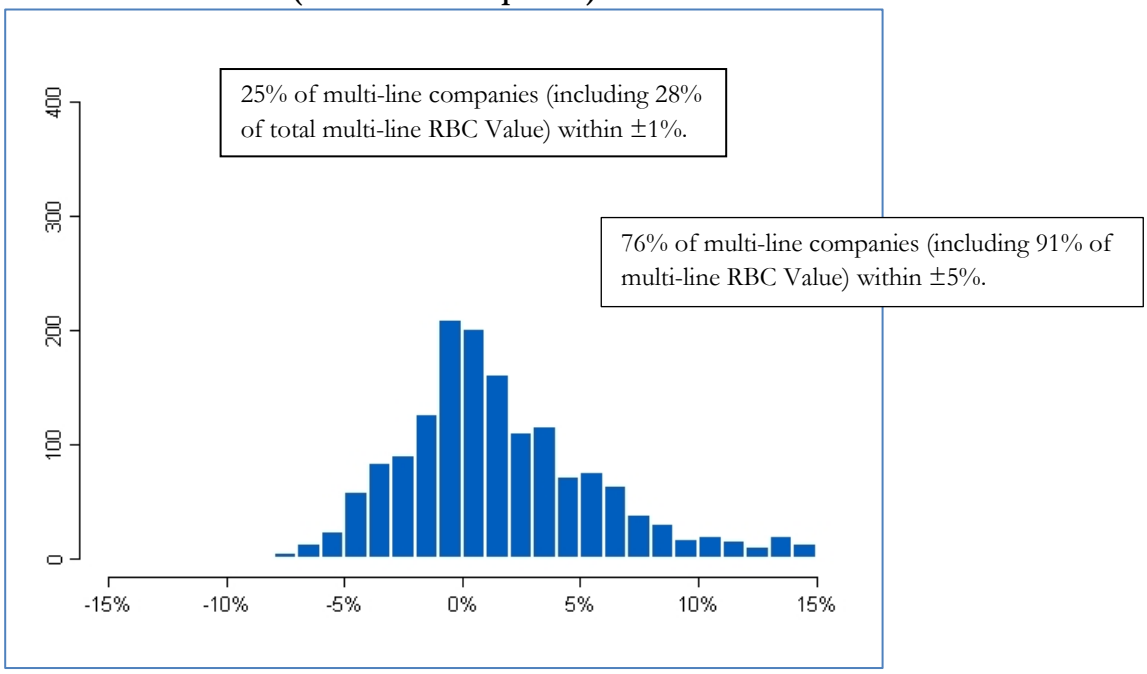
The difference between the correlation matrix approach and the CoMaxLine% Approach is due, in part, to the fact that the degree of diversification in the correlation matrix approach is based on risk by LOB while the degree of diversification in the CoMaxLine% Approach is based on volume (premium amount or reserve amount) by LOB.

In this section we evaluate the effect of that difference by comparing CoMaxLine%-Risk to the correlation matrix approach, company-by-company.

First, to calibrate the CoMaxLine%-Risk approach, we determine that with a MDC of 44.4% the industry-wide RBC UW Risk Value produced by CoMaxLine%-Risk is the same as the total industry-wide RBC UW Risk value from the correlation matrix approach ($100.6 billion). Then, as we did with the NAIC CoMaxLine% Approach, we examine the company-by-company differences between CoMaxLine%-Risk and the correlation matrix approach that remain when both produce the same total industry-wide RBC UW Risk Value.

The histogram in Table 3-2, below, shows the distribution of differences, company-by-company, in the same format as Table 3-1. As was the case in Table 3-1, Table 3-2 excludes mono-line companies and companies with zero RBC UW Risk Values.

Table 3-2
2010 RBC UW Risk Value Differences by Company[28]
Distribution of Number of Companies
Correlation matrix approach versus CoMaxLine%-Risk Approach (44.4% MDC)
(Multi-line Companies)



X-axis = Percentage difference between RBC UW Risk Values based on CoMaxLine%-Risk Approach and RBC UW Risk Values based on correlation matrix approach.
Y-axis = Number of companies, in buckets of 1% difference in RBC UW Risk Value.

Comparing Table 3-1 and Table 3-2 we see that the percentage of multi-line companies with CoMaxLine%-Risk within 5% of the correlation matrix approach is 76%, 7 percentage points more than with the CoMaxLine% Approach. Also, the percentage of RBC UW Risk Value of multi-line companies with CoMaxLine%-Risk within 10% of the correlation matrix approach is 93%, 3 percentage points more than with the CoMaxline% approach.

3.3 HHI vs. CoMaxLine%

In this section, we compare the results of using the CoMaxLine% Approach to the results of using the HHI approach. In Appendix 1, we describe how we calculate the RBC UW Risk Values using the HHI approach.

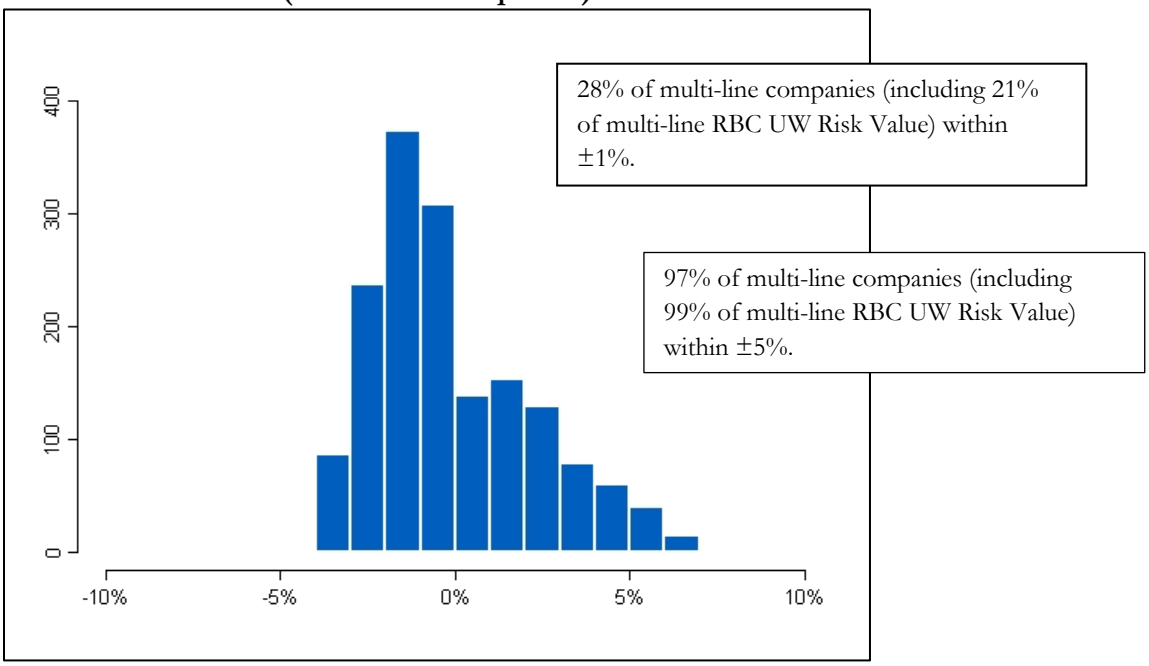
[28] Positive differences represent companies for which the correlation matrix approach produces a higher RBC UW Risk Value than the CoMaxLine%-Risk Approach.

For each company with a 2010 Annual Statement, we apply both the CoMaxLine% Approach and the HHI approach to produce the RBC UW Risk Values by company. Similar to the discussion in Section 3.1, the differences company-by-company between the two diversification approaches have two parts, and we are interested in the differences that remain after controlling for the overall difference in the industry-wide RBC UW Risk Values. We again focus on the companies with non-zero differences in RBC UW Risk Values.

The industry-wide RBC UW Risk Value produced by the HHI approach, with a MDC of 30%, is $101.5 billion. The industry-wide RBC UW Risk Value produced by the CoMaxLine% Approach would be $101.5 billion if the MDC were increased from 30% to 37.7%.

The histogram in Table 3-3, below, shows the distribution of differences, company-by-company, in the same format as Tables 3-1 and 3-2. As was the case in those tables, Table 3-3 excludes mono-line companies and companies with zero RBC UW Risk Values.

Table 3-3
2010 RBC UW Risk Value Differences by Company
Distribution of Number of Companies
HHI approach versus CoMaxLine% Approach (37.7% MDC)
(Multi-Line companies)



X-axis = Percentage difference between RBC UW Risk Values based on CoMaxLine% Approach and RBC UW Risk Values based on HHI approach.
Y-axis = Number of companies, in buckets of 1% difference in RBC UW Risk Value.

We find that:

- 33% of all companies are excluded from the histogram because they are not multi-line.
- For 28% of the multi-line companies, with 21% of the industry-wide multi-line RBC UW Risk Value, the differences are less than ±1%.
- For 97% of the multi-line companies, with 99% of the industry-wide RBC UW Risk Value, the differences are less than ±5%.
- There are no companies where the differences are greater than 10%.
- Considering all companies, even those companies which are mono-line, or which have zero premium and reserves, we find that for 52% of all companies, with 23% of the total RBC UW Risk Value, the differences are less than ±1%. For 97% of all companies, with 99% of the total RBC UW Risk Value, the differences are less than ±5%.

3.4 Further Observations

An analysis of why the three methods discussed in this report produce similar results is beyond the scope of this paper. However, in this section we discuss some of the factors that contribute to that result.

First, the diversification credits are zero for mono-line companies, regardless of method.

Second, the correlation matrix values for LOB-pairs are not highly varied. It is possible that the differences would be wider if the correlation matrix values were more varied, but we have not explored that possibility.

Third, the diversification element is only one part of the RBC UW Risk Value. The dollar weighted average diversification credit for all multi-line companies is 20%.[29] Differences in diversification credit are thus "diluted" in the total calculation. For multi-line companies with little diversification credit, even large percentage differences in diversification credit have a small effect on total RBC UW Risk Value.

Finally, the diversification formula has the greatest effect on the most diversified companies, and we find that the differences between the CoMaxLine% Approach and the correlation matrix approach decrease as company diversification increases.[30]

Appendix 2, Exhibit 3, Box A, shows the RBC UW Risk Value, the dollars of diversification

[29] Appendix 2/Exhibit 3/Box A/Column "All".

[30] Appendix 2/Exhibit 4/Box D/trend in columns from least diversified to most diversified/in rows -5 to +5, -10 to +10 and -25 to +25.

credit and the average diversification credit for all companies combined and for companies within each company diversification band. Box B shows the same information by RBC UW Risk Value. Boxes C and D show the corresponding information based on the CoMaxLine%-Risk measure of diversification.

In Appendix 2, Exhibit 4 we show the proportions of companies where UW Risk RBC Values varies by 5% or less, 10% or less and 25% or less, for the CoMaxLine% Approach versus the correlation matrix approach, by company size band (measured by RBC UW Risk Value) and by company diversification band. In Appendix 2, Exhibit 4 we also show the proportion of companies where the dollar diversification amount varies by 5% or less, 10% or less and 25% or less, for the CoMaxLine% Approach versus correlation matrix approach, by company size band (measured by RBC UW Risk Value) and by diversification band.

We say the CoMaxLine% Approach is closer to the correlation matrix approach for size/diversification cells where the proportion of companies within the 5% variation, 10% variation and 25% variation bands is higher. We see that RBC UW Risk Value from the CoMaxLine% Approach is closer to the correlation matrix approach for the larger companies (Box C) and for the more diversified companies (Box D).

In Appendix 2, Exhibit 5 we show the data for CoMaxLine%-Risk versus the correlation matrix approach as we did in Exhibit 4 for CoMaxLine% versus the correlation matrix approach. We see that CoMaxLine%-Risk is generally closer to the correlation matrix approach than was the case for the CoMaxLine% Approach.

4. GLOSSARY

Annual Statement	US NAIC Annual Statement
CoMaxLine%	The NAIC measure of concentration, the percentage of a company's total premium or reserves from its single largest LOB.
CoMaxLine% Approach	The NAIC method of determining diversification credit across LOBs. It is (1.0 – CoMaxLine%) times 30%.
CoMaxLine%-Risk Approach	CoMaxLine% Approach based on risk charge size by LOB rather than premium or reserve volume by LOB.
Correlation	We use that term to characterize methods of combining LOB risk charges to produce an all-lines risk charge or combining premium risk and reserve risk to produce total risk using 'correlation factors.' The use of the term does not imply that the assumptions underlying individual and joint distributions of the parameters are satisfied.
Correlation Factor	A factor used to express the relationship between individual risks to produce the risk parameter of interest for the combined risk. The use of the term does not imply that the assumptions underlying individual and joint distributions of the parameters are satisfied.
Correlation Matrix	A matrix of correlation factors, typically one factor for each pair of LOBs.
DCWP	Risk-Based Capital Dependency and Calibration Working Party of the Casualty Actuarial Society
LCF	Loss Concentration Factor, as calculated in the 2010 RBC Formula, applicable to reserve risk. Based on the CoMaxLine% Approach.
LOB	Schedule P Lines of Business used in the RBC Formula. Note that three pairs of Schedule P LOBs are combined; occurrence and claims Other Liability (Line H), occurrence and claims-made Products Liability (Line R), and Reinsurance: nonproportional property and Reinsurance: nonproportional financial (Lines P and N, respectively).
Loss sensitive business adjustment	An element of the RBC Formula that reduces the risk charge if unfavorable experience can be offset by increases in income on loss sensitive business.
MDC	Maximum Diversification Credit, 30% in the 2010 RBC Formula
NAIC	National Association of Insurance Commissioners
Own company adjustment, or 50/50 rule	For each company and LOB, premium risk and reserve risk are based 50% on factors calibrated on industry data and 50% on industry data adjusted by the ratio of company experience to industry experience for the most recent 10 years (if 10 years of company data is available, otherwise, there is no adjustment).
PCF	Premium Concentration Factor as calculated in the 2010 RBC Formula. Based on the CoMaxLine% Approach.
R_0	Asset Risk – Insurance affiliate investment and (non-derivative) off-balance sheet risk.
R_1	Asset Risk – Fixed Income Investments
R_2	Asset Risk – Equity

R_3	Credit risk (non-reinsurance plus one half of Reinsurance Credit Risk)
R_3-Reinsurance Credit Risk	See Reinsurance Credit Risk
R_4	UW – Reserve risk plus one half of reinsurance credit risk,[31] including growth risk. This paper uses R_4 without the reinsurance credit risk adjustment and without growth risk.
R_5	UW – Premium risk, including growth risk. This paper uses R_5 without growth risk.
RBC	Risk-Based Capital
RBC Formula or Formula	The 2010 NAIC Property-Casualty RBC Formula
RBC Value	The Company Action Level amount calculated from the RBC Formula.
RBC UW Risk Value	The Company Action Level amount calculated for the UW risk components of the RBC Formula.
Reinsurance Credit Risk	An element of R_3, representing both credit risks related to reinsurance financial capacity and the difference in premium and reserve risk between companies with varying levels of ceded reinsurance.
Solvency II	EU regulation and related implementing measures.
Standard Formula	A formula determining capital requirements under Solvency II, RBC or other regulatory capital systems.
UW	Underwriting
UW risk	Underwriting risk – the combination of premium risk and reserve risk.

[31] The 'transfer' from credit risk to reserve risk applies only if the pure reserve risk component is larger than the reinsurance credit risk, as is the case for most companies.

5. Authors

Principal Authors: Ronald Wilkins, Allan M. Kaufman

Assistance provided by Natalie Atkinson, Damon Chom, Kean Mun Loh and Ji Yao.

Work was supported by the DCWP party with 2015-2017 membership as follows:

Allan M Kaufman, Chair	
Natalie S. Atkinson	Giuseppe F. LePera
Joseph F. Cofield	Kean Mun Loh
Jordan Comacchio	Ronald Wilkins
Sholom Feldblum	Jennifer X. Wu
CAS Staff – Karen Sonnet	

6. References

DCWP Reports

[1.] DCWP Report 1 Overview of Dependencies and Calibration in the RBC Formula, CAS *E-Forum*, Winter 2012, Volume 1, http://www.casact.org/pubs/forum/12wforum/DCWP_Report.pdf.

[2.] DCWP Report 2, 2011 Research – Short Term Project, CAS *E-Forum*, Winter 2012 Volume 1, www.casact.org/pubs/forum/12wforum/RBC_URWP_Report.pdf.

[3.] DCWP Report 3, Solvency II Standard Formula and NAIC RBC, CAS *E-Forum*, Fall/2012, http://www.casact.org/pubs/forum/12fforumpt2/RBC-DCWPRpt3.pdf.

[4.] DCWP Report 4, A Review of Historical Insurance Company Impairments, CAS *E-Forum*, Fall 2012, http://www.casact.org/pubs/forum/12fforumpt2/RBC-DCWPRpt4.pdf.

[5.] DCWP Report 5, An Economic Basis for P/C Insurance RBC Measures, CAS *E-Forum*, Summer/2013, http://www.casact.org/pubs/forum/13sumforum/01RBC-econ-report.pdf.

[6.] DCWP Report 5, An Economic Basis for P/C Insurance RBC Measures, http://www.casact.org/pubs/forum/13sumforum/01RBC-econ-report.pdf.

[7.] DCWP Report 6, Premium Risk Charges – Improvements to Current Calibration Method, CAS *E-Forum*, Fall 2013, http://www.casact.org/pubs/forum/13fforum/01-Report-6-RBC.pdf.

[8.] DCWP Report 7, Reserve Risk Charges – Improvements to Current Calibration Method, CAS *E-Forum*, Winter 2014, http://www.casact.org/pubs/forum/14wforum/Report-7-RBC.pdf.

[9.] DCWP Report 8, Differences in Premium Risk Factors by Type of Company, CAS *E-Forum*, Spring 2014, http://www.casact.org/pubs/forum/14spforum/01-RBC-Dependencies-Calibration-Working-Party.pdf.

[10.] DCWP Report 9, Differences in Premium and Reserve Risk Charges by Ceded Reinsurance Usage, CAS *E-Forum*, Fall 2014, http://www.casact.org/pubs/forum/14fforumv2/DCWP__Report.pdf.

[11.] DCWP Report 10, Reserve Risk Charges – Standard Formula vs. Individual Company Assessments, CAS *E-Forum*, Winter 2015, http://www.casact.org/pubs/forum/15wforum/DCWP-Report.pdf.

[12.] DCWP Report 11, RBC UW Risk Safety Levels – Actual vs. Expected, http://www.casact.org/pubs/forum/16wforum/DCWP-Report.pdf Add all DCWP Reports.

[13.] DCWP Report 12, Insurance Risk-Based Capital with a Multi-Period Time Horizon. CAS *E-Forum*, Spring 2016, http://www.casact.org/pubs/forum/16spforum/Working-Party-Report.pdf.

National Association of Insurance Commissioners (NAIC)

[14.] NAIC, "Solvency Modernization Initiative: Country Comparison Analysis: United Kingdom," NAIC November 2009, 1-8. http://www.naic.org/documents/committees_smi_int_solvency_uk.pdf.

[15.] NAIC, "Risk Based Capital General Overview," July 15, 2009, http://www.naic.org/documents/committees_e_capad_RBCoverview.pdf.

[16.] NAIC, "Property and Casualty Risk-Based Capital Forecasting & Instructions," 2010.

[17.] NAIC, "U.S.-EU Dialogue Project: A Comparison of the Two Regulatory Regimes and the Way Forward," NAIC Center for Insurance Policy and Research Newsletter April 2013, 7-11. http://www.naic.org/cipr_newsletter_archive/vol7_us_eu_dialogue.pdf.

[18.] NAIC, "IAIS Insurance Capital Standard Public Consultation Document: Final NAIC Comments," NAIC February 2015, 1-18. http://www.naic.org/documents/committees_g_related_naic_comments_iais_ics_draft.pdf.

Other

[19.] Chief Risk Officer Forum, June 2005, "A framework for incorporating diversification in the solvency assessment of insurers."

[20.] Ferri, Antoni, Lluis Bermudez and Montserrat Guillen. 2011. "A Correlation Sensitivity Analysis for non-life underwriting risk module SCR," ASTIN presentation June 2011. http://www.actuaries.org/ASTIN/Colloquia/Madrid/Papers/Bermudeza_Ferri_Guillen.pdf.

[21.] Financial Services Authority (United Kingdom). 2003. "Enhanced Capital Requirements and Individual Capital Assessments for Life Insurers," FSA, Consultation Paper 195, 1-329. http://www.fsa.gov.uk/pubs/cp/cp195.pdf.

[22.] Groupe Consultatif Actuariel Européen, 2005. "Diversification," Groupe Consultatif Actuariel Européen Technical Paper, October 2005, 1-13. http://actuary.eu/documents/diversification_oct05.pdf.

[23.] International Actuarial Association, Insurer Solvency Assessment Working Party. 2004. "A Global Framework for Insurer Solvency Assessment," International Actuarial Association Research Report, 2004, 1-179. http://www.actuaries.org/LIBRARY/papers/global_framework_insurer_solvency_assessment-public.pdf.

[24.] Kaufman, Allan M. and Elise C. Liebers, "NAIC Risk Based Capital Efforts in 1990-91", CAS Forum, 1992.

[25.] Lloyd's. "Solvency II: 2015 Year-End Standard Formula Exercise Guidance Notes," Lloyd's, 2016, 1-28. https://www.lloyds.com/~/media/files/the%20market/operating%20at%20lloyds/solvency%20ii/2016%20guidance/2015%20yearend%20standard%20formula%20submission%20guidance%20february%202016published.pdf.

[26.] Lloyd's. 2016. "Standard Formula SCR and MCR Calculation Template for use in the 2015 Year-End Exercise (Excel Template)," Lloyd's April 2016. http://www.lloyds.com/~/media/files/the%20market/operating%20at%20lloyds/solvency%20ii/2016%20guidance/2015_yesf_synd_v62.xlsx.

[27.] Panjer, Harry, "Capital Requirements for Insurers - Incorporating Correlations", CAS-SOA ERM-Capital Symposium, Presentation, July 2003.

[28.] Sharara, Ishmael, Mary Hardy, and David Saunders, "Regulatory Capital Standards for Property and Casualty Insurers under the U.S., Canadian and Proposed Solvency II (Standard) Formulas," Sponsored by CAS, CIA, and SOA Joint Risk Management Section, University of Waterloo, 2010.

[29.] Sutherland-Wong, Christian and Michael Sherris, "Risk-Based Regulatory Capital for Insurers: A Case Study," International Actuarial Association, 2004

[30.] Sandström, Arne, Solvency – Models, Assessment and Regulation, 2006, Taylor & Francis Group, LLC, http://docslide.us/documents/solvency-models-assessment-and-regulation.html\.

[31.] Sandström, Arne, "Solvency—A Historical Review and Some Pragmatic Solutions," *Swiss Association of Actuaries Bulletin* 1, 2007, http://www.actuaries.ch/de/mitgliedschaft/bulletin.htm.

CEIOPS (EIOPA)

[32.] EIOPA general website with links to EIOPA and CEIOPS (predecessor to EIOPA) documents. http://ec.europa.eu/finance/insurance/solvency/solvency2/index_en.htm.

[33.] CEIOPS, "QIS5 Technical Specifications Annex to Call for Advice from CEIOPS on QIS5," July 2010, https://eiopa.europa.eu/fileadmin/tx_dam/files/consultations/QIS/QIS5/QIS5-technical_specifications_20100706.pdf.

[34.] CEIOPS, "Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula—Article 111 l, Simplified calculations in the Standard Formula," January, 2010, https://eiopa.europa.eu/CEIOPS-Archive/Documents/Advices/CEIOPS-L2-Advice-Simplifications-for-SCR.pdf.

[35.] CEIOPS, "Annexes to the QIS5 Technical Specifications," July 2010, https://eiopa.europa.eu/fileadmin/tx_dam/files/consultations/QIS/QIS5/Annexes-to-QIS5-technical_specifications_20100706.pdf.

[36.] CEIOPS/EIOPA Web page with links to QIS 5 forms and spreadsheets, 2010, https://eiopa.europa.eu/consultations/qis/quantitative-impact-study-5/spreadsheets-and-it-tools/index.html.

[37.] CEIOPS, "Solvency II Final L2 Advice, Index," https://eiopa.europa.eu/publications/sii-final-l2-advice/index.html.

[38.] CEIOPS, "Solvency II Calibration Paper," (CEIOPS Main background document for Level 2 advice as to calibration), April 2010, https://eiopa.europa.eu/fileadmin/tx_dam/files/publications/submissionstotheec/CEIOPS-Calibration-paper-Solvency-II.pdf.

[39.] CEIOPS, "Solvency II Final L2 Advice, Index," https://eiopa.europa.eu/publications/sii-final-l2-advice/index.html.

[40.] CEIOPS, "Solvency II Calibration Paper," (CEIOPS Main background document for Level 2 advice as to calibration), April 2010, https://eiopa.europa.eu/fileadmin/tx_dam/files/publications/submissionstotheec/CEIOPS-Calibration-paper-Solvency-II.pdf.

[41.] CEIOPS, "Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula Calibration of Non-life Underwriting Risk," April 2010, https://eiopa.europa.eu/fileadmin/tx_dam/files/consultations/consultationpapers/CP71/CEIOPS-DOC-67-10_L2_Advice_Non_Life_Underwriting_Risk.pdf.

[42.] CEIOPS, "Advice for Level 2 Implementing Measures on Solvency II: SCR Standard Formula Article 111(d) Correlations," (former Consultation Paper 74), January 2010, https://eiopa.europa.eu/fileadmin/tx_dam/files/consultations/consultationpapers/CP74/CEIOPS-L2-Advice-Correlation-Parameters.pdf.

[43.] CEIOPS/EIOPA Web page with links to QIS 5 forms and spreadsheets, 2010, https://eiopa.europa.eu/consultations/qis/quantitative-impact-study-5/spreadsheets-and-it-tools/index.html.

[44.] EIOPA, "Annexes to the EIOPA Report on QIS5 (Fifth Quantitative Impact Study for Solvency II)," EIOPA, March 2011, 1-29. https://eiopa.europa.eu/Publications/Reports/QIS5_Annexes_Final.pdf.

[45.] EIOPA, "Calibration of the Premium and Reserve Risk Factors in the Standard Formula of Solvency II, Report of the Joint Working Group on Non-Life and Health Non-Similar to Life Techniques (NSLT) Calibration," EIOPA, December 2011, 1-77. https://eiopa.europa.eu/Publications/Reports/EIOPA-11-163-A-Report_JWG_on_NL_and_Health_non-SLT_Calibration.pdf.

[46.] EIOPA, "Calibration of the Premium and Reserve Risk Factors in the Standard Formula of Solvency II, Report of the Joint Working Group on Non-Life and Health Non-Similar to Life Techniques (NSLT) Calibration: Annex 6_2: Averaging and Combined Approach," EIOPA, December 2011, 1-14. https://eiopa.europa.eu/Publications/Reports/EIOPA-11-163-C-Annex_6_2_Report_JWG_on_NL_and_Health_non-SLT_Calibration.pdf.

[47.] EIOPA, "EIOPA Report on the Fifth Quantitative Impact Study (QIS5) for Solvency II," EIOPA, March 2011, 1-153. https://eiopa.europa.eu/Publications/Reports/QIS5_Report_Final.pdf.

[48.] EIOPA and NAIC, "EU-U.S. Dialogue Project, Technical Committee Reports Comparing Certain Aspects of the Insurance Supervisory and Regulatory Regimes in the European Union and the United States," EIOPA and NAIC, December 2012, 1-130. http://www.naic.org/documents/eu_us_dialogue_report_121220.pdf.

Appendix 1 - Calculation of 2010 RBC UW Risk Values by Company

In Section 3, we compare the RBC UW Risk Values from the RBC Formula with the RBC UW Risk Values from alternative formulas in which we replace the CoMaxLine% calculation with correlation matrix, CoMaxLine%-Risk and HHI calculations. We use 2010 Annual Statement data by company[32] to determine the company-by-company RBC UW Risk Values as described below.

For each LOB individually:

- We obtain 2010 net written premium and net loss and loss adjustment expense reserves by LOB from the Annual Statement.
- We use Schedule P Part 2 reserve runoff to calculate the own-company adjustment factors for reserve risk.
- We use Schedule P Part 1 LRs to calculate the own-company adjustment factors for premium risk.
- We use Schedule P Parts 7A and 7B to calculate the loss-sensitive contract adjustment for premium risk.
- For each LOB, we apply the premium risk factor, the reserve risk factor, the premium and reserve investment income offsets, the own company adjustments, and loss sensitive contract adjustment, in accordance with the 2010 RBC Formula.

[32] For this purpose, we considered individual company legal entities. We do not use the NAIC groups or DCWP-pooled companies.

- The premium calculation includes extra steps in that premium risk factors by LOB are converted to the premium risk charge by LOB using the all-lines company expense ratio.

All LOBs combined

- We determine the all-lines combined risk values for premium and reserves using the PCFs and LCFs by company, respectively.

  As explained in Section 2, for each company, the PCFs and LCFs will be values between 71.6% and 100.0% using the CoMaxLine% Approach.

Simplifications

- We do not apply the growth risk charge
- We do not apply the own-company adjustment for 2-Year LOBs, as the necessary data is not in Schedule P.
- The reserve risk component does not include the R_3-Reinsurance Credit Risk amount that is transferred to R_4.

Correlation Matrix Approach

To estimate the RBC UW Risk Values for the correlation matrix approach we first calculate the results by LOB as described above, using all-lines company expenses for each LOB.[33]

We combine the LOB risk charges applying correlation matrix, Appendix 6A/Exhibit 6-1[34] to the risk charges by LOB.

CoMaxLine%-Risk Approach

To estimate the RBC UW Risk Values for the CoMaxLine%-Risk Approach we first calculate the premium risk and reserve risk values by LOB in accordance with RBC Formula as described above for the correlation matrix approach.

We calculate CoMaxLine%-Risk using the dollar amounts of premium risk and reserve risk, by LOB, rather than using the dollar amounts of premium and reserves.

We calculate the PCFs/LCFs from the CoMaxLine%s-Risk.

HHI Alternative

To estimate the RBC UW Risk Values for the HHI approach we first calculate the results by LOB as described above.

[33] When the RBC Formula was constructed it was decided to use company total expenses rather than LOB expenses in the premium UW risk calculation because the LOB expenses are not available in the Annual Statement. The expenses by LOB are produced one month later in the Insurance Expense Exhibit.

[34] In mathematical terms, we take the LOB risk charges as a 19x1 vector; multiply it by the 19x19 correlation matrix and multiple that by the LOB risk charges, in dollars, as a 1x19 vector. LCF and PCF factors are not used in the correlation matrix approach.

We calculate the PCFs/LCFs using the HHI values rather than CoMaxLine%. The HHI concentration value equals the sum of the squares of the LOB shares of total. For example, if there is only one LOB, HHI is 1.0, as is the case for CoMaxLine%. With two lines split 25% and 75% HHI is 0.25 ^2 plus 0.75^2 or 0.625 compared the CoMaxLine% of 0.750, i.e., it shows less concentration/more diversification. With three lines split 50%, 25% and 25% HHI is 0.50^2 plus 0.25^2 plus 0.25^2 or 0.375, less concentration/more diversification than the CoMaxLine% of 0.5.

To combine the LOBs, we replace the CoMaxLine%s with the HHI values.

- For each LOB, we apply the premium risk factor, the reserve risk factor, the premium and reserve investment income offsets, the own company adjustments, and loss sensitive contract adjustment, in accordance with the 2010 RBC Formula.

Company Selection

There are 2,434 companies with 2010 Annual Statements in our data set. Of those, 50 companies have significantly negative premium or reserves for some LOBs.[35] The RBC Formula substitutes zero for negative values. For our work, we eliminate those 50 companies, leaving 2,384 companies in our analysis. Of those, 360 have zero UW Risk RBC and 402 have zero diversification credit in the CoMaxLine%, CoMaxLine%-Risk and HHI calculations. The remaining 1,622 companies provide information on how the diversification formulas affect RBC UW Risk Values.

[35] Negative in total for all lines combined or with large enough negative values to potentially distort one or more of the diversification formulas we are testing.

Appendix 1/Exhibit 1
Selected DCWP Correlation Matrix – Applied By the DCWP to US NAIC LOBs for this Study

LOB/LOB	HO	PPA	CA	WC	CMP	M-Occ	M-CM	SL	OL	SP	Phy	Fid	Other	Int'l	Re Prop	Re- Liab	Prod	FG	Warranty
HO	100%	25%	25%	25%	50%	25%	25%	25%	25%	**75%**	50%	25%	25%	25%	25%	25%	25%	25%	25%
PPA	25%	100%	50%	25%	25%	25%	25%	25%	25%	25%	**75%**	25%	25%	25%	25%	25%	25%	25%	25%
CA	25%	50%	100%	50%	50%	25%	25%	50%	50%	25%	**75%**	25%	25%	25%	25%	25%	50%	25%	25%
WC	25%	25%	50%	100%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%
CMP	50%	25%	50%	25%	100%	25%	25%	50%	50%	50%	25%	25%	25%	25%	25%	25%	50%	25%	25%
M-Occ	25%	25%	25%	25%	25%	100%	**100%**	50%	50%	25%	25%	25%	25%	25%	25%	25%	50%	25%	25%
M-CM	25%	25%	25%	25%	25%	**100%**	100%	50%	50%	25%	25%	25%	25%	25%	25%	25%	50%	25%	25%
SL	25%	25%	50%	25%	50%	50%	50%	100%	**75%**	25%	25%	25%	25%	25%	25%	50%	**100%**	25%	25%
OL	25%	25%	50%	25%	50%	50%	50%	**75%**	100%	25%	50%	50%	25%	50%	25%	50%	**100%**	25%	25%
SP	**75%**	25%	25%	25%	50%	25%	25%	25%	25%	100%	25%	25%	25%	25%	50%	25%	25%	25%	25%
Phy	50%	**75%**	**75%**	25%	25%	25%	25%	25%	50%	25%	100%	25%	25%	25%	25%	25%	25%	25%	25%
Fid	25%	25%	25%	25%	25%	25%	25%	25%	50%	25%	25%	100%	25%	25%	25%	50%	25%	25%	25%
Other	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	100%	25%	25%	25%	25%	25%	25%
Int'l	25%	25%	25%	25%	25%	25%	25%	25%	50%	25%	25%	25%	25%	100%	25%	25%	25%	25%	25%
Re Prop	25%	25%	25%	25%	25%	25%	25%	25%	25%	50%	25%	25%	25%	25%	100%	25%	25%	25%	25%
Re- Liab	25%	25%	25%	25%	25%	25%	25%	50%	50%	25%	25%	50%	25%	25%	25%	100%	50%	25%	25%
Prod	25%	25%	50%	25%	50%	50%	50%	**100%**	**100%**	25%	25%	25%	25%	25%	25%	50%	100%	25%	25%
FG	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	100%	25%
Warranty	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	25%	100%

Note: Off diagonal values other than 25%, 50% are in bold.

LOB Definitions

LOB	Abbreviation	LOB	Abbreviation	LOB	Abbreviation
Homeowners/Farmowners	HO	Special Liab	SL	International	Int'l
Priv. Passenger Auto	PPA	Other Liab-Occ and CM	OL	Reinsurance-Fin and Prop	Re Prop
Commercial Auto	CA	Spec Property	SP	Reinsurance-Liab	Re Liab
Workers Compensation	WC	Auto Physical Damage	Phy	Products Liability-Occ and CM	Prod
Commercial Multi-peril	CMP	Fidelity & Surety	Fid	Financial/Mortgage Guarantee	FG
Medical Prof Liab - Occ	M-Occ	Other	Other	Warranty	Warranty
Medical Prof Liab - CM	M-CM				

Solvency II Correlation Matrix

The Solvency II Standard Formula uses a correlation matrix to specify LOB diversification. Appendix 1/Exhibit 2A lists the Solvency II 12 non-life LOBs

Appendix 1/Exhibit 2A Solvency II LOBs[36]

1	Motor vehicle liability	7	Legal expenses
2	Other motor	8	Assistance
3	Marine, aviation and transport	9	Miscellaneous financial loss
4	Fire and other damage to property	10	NP casualty reinsurance
5	General liability	11	NP marine, aviation and transport reinsurance
6	Credit and suretyship	12	NP property reinsurance

Direct LOBs include proportional reinsurance of the same type.
NP = Non-proportional

Appendix 1/Exhibit 2B below shows the Solvency II Standard Formula LOB correlation matrix for those 12 LOBs.[37]

Appendix 1/Exhibit 2B
Solvency II Standard Formula Correlation Matrix for Premium and Reserves

LOB/LOB	1	2	3	4	5	6	7	8	9	10	11	12
1	100%	50%	50%	25%	50%	25%	50%	25%	50%	25%	25%	25%
2	50%	100%	25%	25%	25%	25%	50%	50%	50%	25%	25%	25%
3	50%	25%	100%	25%	25%	25%	25%	50%	50%	25%	50%	25%
4	25%	25%	25%	100%	25%	25%	25%	50%	50%	25%	50%	50%
5	50%	25%	25%	25%	100%	50%	50%	25%	50%	50%	25%	25%
6	25%	25%	25%	25%	50%	100%	50%	25%	50%	50%	25%	25%
7	50%	50%	25%	25%	50%	50%	100%	25%	50%	50%	25%	25%
8	25%	50%	50%	50%	25%	25%	25%	100%	50%	25%	25%	50%
9	50%	50%	50%	50%	50%	50%	50%	50%	100%	25%	50%	25%
10	25%	25%	25%	25%	50%	50%	50%	25%	25%	100%	25%	25%
11	25%	25%	50%	50%	25%	25%	25%	25%	50%	25%	100%	25%
12	25%	25%	25%	50%	25%	25%	25%	50%	25%	25%	25%	100%

The factors equal to 1.0, along the diagonal, represent the correlation between the LOB and itself. In the Solvency II 3rd Quantitative Impact Analysis (QIS3), the factors were calibrated with data from one country, supplemented by expert judgment. The factors appear to primarily represent an expert judgment on whether the LOB pairwise correlation is lower (0.25) or higher (0.50).

In the Solvency II 4th Quantitative Impact Analysis (QIS4) analysis, the factors were sensitivity

36 http://www.lloyds.com/~/media/files/the%20market/operating%20at%20lloyds/solvency%20ii/2016%20guidance/2015_yesf_synd_v62.xlsx. " Non-Life & NSLT Health P&R"

37 Ibid. Tab "Non-Life and Health UW Risk"

tested with additional analysis assuming a minus or plus 25 percentage points adjustment to each "non-diagonal" value. These changes resulted in capital requirements that were 25% lower and 21% higher (respectively) than the proposed QIS4 factors.[38] After this sensitivity analysis was completed, the selected factors were maintained at the QIS3 level *"translating the broad support there is around these parameters and the lack of more evidence for changing the correlations".*[39] Thus, the overall level appears to rely heavily on expert judgment much like the 30% MDC in the RBC Formula.

[38] CEIOPS-DOC-70/10, Annex B, pages 38-44
[39] CEIOPS-DOC-70/10 (Page 44, paragraph B.31)

Appendix 2 – Comparisons between CoMaxLine%, CoMaxLine%- Risk, and Correlation Matrix Approaches

Appendix 2/Exhibit 3

Appendix 2/Exhibit 3, below, shows the dollar amount of RBC UW Risk Value, the dollar amount of diversification credit, and the average diversification credit by company-size and by company-diversification band, separately for the CoMaxLine% Approach and the CoMaxLine%-Risk Approach. We define the size and diversification bands below.

RBC UW Risk Value Size Bands

We show the data, in seven company-size bands. The bands A through E divide the 1,622 multi-line companies into five groups with approximately 325 companies in each band. Band A has the smallest 20% of multi-line companies. Band E has the largest 20% of multi-line companies. In addition, we show two other informational bands. "Tiny" is for the 75 smallest multi-line companies. This column is for information only, as we include the 75 in band A. "Jumbo" is for the 75 largest multi-line companies. This column is for information, as we include the 75 in band E.

Columns: %Diversification Size Bands

We show the data, in seven company-diversification bands. The bands A through E divide the 1,622 multi-line companies into five groups with approximate 325 multi-line companies in each band. Band A has the least diversified multi-line companies, those with the lowest percentage diversification credits. Band E has the most diversified 20% of multi-line companies, those with the highest percentage diversification credits. In addition, we show two other bands. The column "75 Least Diversified" is for the 75 multi-line companies with the lowest, non-zero, diversification percentages. This column is for information as we include the 75 in band A. The column "75 Most Diversified" is for the 75 multi-line companies with the largest diversification credit %. This column is also for information, as we include the 75 in band E.

Distribution of RBC UW Risk Value and Diversification Amount

Appendix 2/Exhibit 3, has four "boxes," labeled A, B, C and D. Within each box we show the dollar amount of RBC UW Risk Value, the percentage of RBC UW Risk Value by size band or diversification band, the dollar amount of diversification credit and the average diversification credit.

Boxes A and C show the data in company-diversification bands, for CoMaxLine% and CoMaxLine%-Risk approaches, respectively. Boxes B and D show the data in RBC UW Risk Size bands, for CoMaxLine% and CoMaxLine%-Risk approaches, respectively.

Some key features of the summary are the following:

- The weighted average percentage diversification across all multi-line companies is 20%, for both the CoMaxLine% Approach and the CoMaxLine%-Risk Approach (the same value appears in boxes A, B, C, and D in the "All" column).

- For the 75 most diversified multi-line companies, the average diversification percentage is 30% for CoMaxline% (Box A), and 32% for CoMaxLine%-Risk (Box C).
- For CoMaxLine%, the total RBC UW Risk Value is $97,975 million, excluding mono-line companies. Of that amount, $64,659 million, or 66%, relates to the 75 largest multi-line companies. $87,567 million of that amount, or 89%, relates to the largest 20% of multi-line companies (Box B. RBC UW Risk Size Bands/Column E).
- For CoMaxLine%, the total RBC UW Risk Value is essentially the same as for CoMaxLine%-Risk because we calibrated the CoMaxLine% MDC to achieve that result. The distribution by RBC UW Value size bands for CoMaxLine%-Risk is similar to the distribution for CoMaxLine%.
- For CoMaxLine%, nearly all of the diversification credit, $22 million of $24 million, arises from size band E, the 20% largest companies by RBC UW Risk Value (Box B/Column E).

Appendix 2/Exhibit 4 – CoMaxline% and Correlation Matrix by Size and Diversification Bands

In Appendix 2/Exhibit 4, we compare RBC UW Risk Value and dollar diversification credit amounts for the CoMaxLine% Approach to the corresponding values for the correlation matrix approach. We show the information for all companies, and separately in size and diversification bands, defined above.

In each column, we show the percentage of multi-line companies with percentage difference in RBC UW Risk Value (Boxes A and B) and percentage difference in dollar diversification credit (Boxes C and D) in bands ±5%, ±10%, and ±25%, for CoMaxline% versus correlation matrix approaches. Boxes A and C show the information by RBC WW Risk Value Size Band. Boxes B and D show the information by % Diversification Band.

Appendix 2/Exhibit 4/Box A/Column "All" shows that the RBC UW Risk Values differ from the corresponding correlation matrix values by more than 5% for only 31% of all multi-line companies and for 26%, of the largest 20% of multi-line companies (Box A/column E). The values differ by more than 10% for 10% of multi-line companies overall and for 9% of the largest 20% of multi-line companies. (Box A, columns "All" and "E").

The percentage differences in diversification will be larger than the percentage difference in RBC UW Risk Value. Therefore, the differences in diversification amount will be higher than the differences in RBC UW Risk Values. In fact, the percentage difference in diversification amount is more than 5% for 86% of multi-line companies, more than 10% for 71% of multi-line companies and more than 25% for 48% of multi-line companies (Box C or D/column "All").

For the most diversified multi-line companies, band E, that are potentially the most affected by differences in the diversification formula, the percentage change in dollars of diversification is more

than 5% for 66% of multi-line companies, but more than 10% for only 28% of multi-line companies and more than 25% for only 6% of multi-line companies; much fewer than for all multi-line companies combined. For the least diversified multi-line companies, band A, the difference in dollars of diversification is greater than 25% for 83% of multi-line companies (Box D), but in that case, the average diversification percentage is only 3% (Exhibit 3/Box A).

Appendix 2/Exhibit 5- CoMaxline%-Risk and Correlation Matrix by Size and Diversification Bands

Appendix 2/Exhibit 5 compares CoMaxLine%-Risk to the correlation matrix approach, showing the same information as Exhibit 4.

In many respects, the patterns in Exhibit 5 are similar to the patterns in Exhibit 4, but the CoMaxLine%-Risk and correlation matrix approaches are closer than is the case for the CoMaxLine% Approach versus the correlation matrix approach.

Appendix 2/Exhibit 3
CoMaxLine% and CoMaxLine%-Risk
RBC UW Risk Values and Diversification Amounts

CoMaxLine%								
A. Percentage Diversification Bands								
Item	**All**	**75 Least Diversified (memo)**	**A**	**B**	**C**	**D**	**E**	**75 Most Diversified (memo)**
RBC UW Risk Value	97,975	956	5,249	15,939	19,364	30,805	26,617	4,274
% of RBC UW Risk Value	100%	1%	5%	16%	20%	31%	27%	4%
$ of Diversification	23,901	3	141	1,747	3,702	8,618	9,693	1,819
Avg % Diversification	20%	0%	3%	10%	16%	22%	27%	30%
B. RBC UW Risk Size Bands								
Item	**All**	**Tiny (memo)**	**A**	**B**	**C**	**D**	**E**	**Jumbo (memo)**
RBC UW Risk Value	97,975	8	218	928	2,523	6,739	87,567	64,659
% of RBC UW Risk Value	100%	0%	0%	1%	3%	7%	89%	66.0%
$ of Diversification	23,901	1	33	163	480	1,364	21,861	16,354
Avg % Diversification	20%	12%	13%	15%	16%	17%	20%	20%

CoMaxLine% - Risk								
C. Percentage Diversification Bands								
Item	**All**	**75 Least Diversified (memo)**	**A**	**B**	**C**	**D**	**E**	**75 Most Diversified (memo)**
RBC UW Risk Value	97,990	691	7,297	17,477	26,467	21,652	25,097	4,864
% of RBC UW Risk Value	100%	1%	7%	18%	27%	22%	26%	5%
$ of Diversification	23,886	2	243	1,907	4,798	6,405	10,533	2,296
Avg % Diversification	20%	0%	3%	10%	15%	23%	30%	32%
D. RBC UW Risk Size Bands								
Item	**All**	**Tiny (memo)**	**A**	**B**	**C**	**D**	**E**	**Jumbo (memo)**
RBC UW Risk Value	97,990	8	215	921	2,490	6,661	87,703	65,120
% of RBC UW Risk Value	100%	0%	0%	1%	3%	7%	90%	66%
$ of Diversification	23,886	1	37	168	522	1,455	21,703	15,794
Avg % Diversification	20%	13%	15%	15%	17%	18%	20%	20%

Appendix 2/Exhibit 4
% Difference from CoMaxLine% Approach to Correlation Matrix Approach

A. Change in RBC UW Risk Value by RBC UW Risk Value Size Band								
% Change in RBC UW Risk Value	RBC UW Risk Size Bands							
	All	Tiny (memo)	A	B	C	D	E	Jumbo (memo)
-5 to +5	69%	51%	64%	64%	67%	77%	74%	81%
-10 to +10	90%	89%	88%	88%	89%	95%	91%	91%
-25 to +25	100%	100%	100%	100%	100%	100%	100%	100%
Greater than ±5%	31%	49%	36%	36%	33%	23%	26%	19%
Greater than ±10%	10%	11%	12%	12%	11%	5%	9%	9%
Greater than ±25%	0%	0%	0%	0%	0%	0%	0%	0%

B. Change in RBC UW Risk Value by % Diversification Band								
% Change in RBC UW Risk Value	Percentage Diversification Bands							
	All	75 Least Diversified (memo)	A	B	C	D	E	75 Most Diversified (memo)
-5 to +5	69%	99%	96%	59%	53%	60%	78%	84%
-10 to +10	90%	99%	98%	94%	82%	79%	97%	93%
-25 to +25	100%	100%	100%	100%	100%	100%	100%	100%
Greater than ±5%	31%	1%	4%	41%	47%	40%	22%	16%
Greater than ±10%	10%	1%	2%	6%	18%	21%	3%	7%
Greater than ±25%	0%	0%	0%	0%	0%	0%	0%	0%

C. Change in $ Diversification by RBC UW Risk Value Size Band								
% Change in Div $	RBC UW Risk Size Bands							
	All	Tiny (memo)	A	B	C	D	E	Jumbo (memo)
-5 to +5	14%	4%	7%	12%	15%	19%	18%	20%
-10 to +10	29%	9%	16%	20%	26%	38%	45%	53%
-25 to +25	52%	25%	35%	47%	48%	63%	69%	80%
Greater than ±5%	86%	96%	93%	88%	85%	81%	82%	80%
Greater than ±10%	71%	91%	84%	80%	74%	62%	55%	47%
Greater than ±25%	48%	75%	65%	53%	52%	37%	31%	20%

D. Change in $ Diversification by % Diversification Band								
% Change in Div $	Percentage Diversification Bands							
	All	75 Least Diversified (memo)	A	B	C	D	E	75 Most Diversified (memo)
-5 to +5	14%	1%	3%	7%	10%	16%	34%	57%
-10 to +10	29%	5%	10%	13%	17%	34%	72%	83%
-25 to +25	52%	17%	19%	33%	48%	68%	94%	93%
Greater than ±5%	86%	99%	97%	93%	90%	84%	66%	43%
Greater than ±10%	71%	95%	90%	87%	83%	66%	28%	17%
Greater than ±25%	48%	83%	81%	67%	52%	32%	6%	7%

Appendix 2/Exhibit 5
% Difference from CoMaxLine% - Risk Approach to Correlation Matrix Approach

A. Change in RBC UW Risk Value by RBC UW Risk Value Size Band								
% Change in RBC UW Risk Value	RBC UW Risk Size Bands							
	All	Tiny (memo)	A	B	C	D	E	Jumbo (memo)
-5 to +5	76%	55%	68%	72%	73%	82%	85%	91%
-10 to +10	93%	91%	89%	89%	94%	96%	97%	97%
-25 to +25	100%	100%	100%	100%	100%	100%	100%	100%
Greater than ±5%	24%	45%	32%	28%	27%	18%	15%	9%
Greater than ±10%	7%	9%	11%	11%	6%	4%	3%	3%
Greater than ±25%	0%	0%	0%	0%	0%	0%	0%	0%

B. Change in RBC UW Risk Value by % Diversification Band								
% Change in RBC UW Risk Value	Percentage Diversification Bands							
	All	75 Least Diversified (memo)	A	B	C	D	E	75 Most Diversified (memo)
-5 to +5	76%	100%	98%	67%	61%	69%	84%	93%
-10 to +10	93%	100%	100%	96%	83%	87%	98%	100%
-25 to +25	100%	100%	100%	100%	100%	99%	100%	100%
Greater than ±5%	24%	0%	2%	33%	39%	31%	16%	7%
Greater than ±10%	7%	0%	0%	4%	17%	13%	2%	0%
Greater than ±25%	0%	0%	0%	0%	0%	1%	0%	0%

C. Change in $ Diversification by RBC UW Risk Value Size Band								
% Change in Div $	RBC UW Risk Size Bands							
	All	Tiny (memo)	A	B	C	D	E	Jumbo (memo)
-5 to +5	21%	11%	13%	15%	19%	26%	31%	32%
-10 to +10	35%	13%	21%	29%	31%	42%	50%	51%
-25 to +25	58%	28%	43%	52%	58%	64%	74%	76%
Greater than ±5%	79%	89%	87%	85%	81%	74%	69%	68%
Greater than ±10%	65%	87%	79%	71%	69%	58%	50%	49%
Greater than ±25%	42%	72%	57%	48%	42%	36%	26%	24%

D. Change in $ Diversification by % Diversification Band								
% Change in Div $	Percentage Diversification Bands							
	All	75 Least Diversified (memo)	A	B	C	D	E	75 Most Diversified (memo)
-5 to +5	21%	0%	4%	8%	10%	31%	51%	60%
-10 to +10	35%	5%	16%	15%	16%	47%	79%	91%
-25 to +25	58%	16%	26%	30%	56%	81%	98%	100%
Greater than ±5%	79%	100%	96%	92%	90%	69%	49%	40%
Greater than ±10%	65%	95%	84%	85%	84%	53%	21%	9%
Greater than ±25%	42%	84%	74%	70%	44%	19%	2%	0%

DeepTriangle: A Deep Learning Approach to Loss Reserving

Kevin Kuo[a]

[a]*RStudio, 250 Northern Ave, Boston, MA 02210, United States*

Abstract

We propose a novel approach for loss reserving based on deep neural networks. The approach allows for joint modeling of paid losses and claims outstanding, and incorporation of heterogeneous inputs. We validate the models on loss reserving data across lines of business, and show that they improve on the predictive accuracy of existing stochastic methods. The models require minimal feature engineering and expert input, and can be automated to produce forecasts more frequently than manual workflows.

Keywords: loss reserving, machine learning, neural networks
JEL: G22

1. Introduction

In the loss reserving exercise for property and casualty insurers, actuaries are concerned with forecasting future payments due to claims. Accurately estimating these payments is important from the perspectives of various stakeholders in the insurance industry. For the management of the insurer, the estimates of unpaid claims inform decisions in underwriting, pricing, and strategy. For the investors, loss reserves, and transactions related to them, are essential components in the balance sheet and income statement of the insurer. And, for the regulators, accurate loss reserves are needed to appropriately understand the financial soundness of the insurer.

There can be time lags both for reporting of claims, where the insurer is not notified of a loss until long after it has occurred, and for final development of claims, where payments continue long after the loss has been reported. Also, the amounts of claims are uncertain before they have fully developed. These factors contribute to the difficulty of the loss reserving problem, for which extensive literature exists and active research is being done. We refer the reader to England and Verrall (2002) for a survey of the problem and existing techniques.

Deep learning has garnered increasing interest in recent years due to successful applications in many fields (LeCun, Bengio, and Hinton 2015) and has recently made its way into the loss reserving literature. Wüthrich (2018b) augments the traditional chain ladder method with neural networks to incorporate claims features, and Gabrielli, Richman, and Wuthrich (2018) embeds the over-dispersed Poisson (ODP) model into a neural network.

Email address: kevin.kuo@rstudio.com (Kevin Kuo)
Preprint submitted to CAS E-Forum *March 5, 2019*

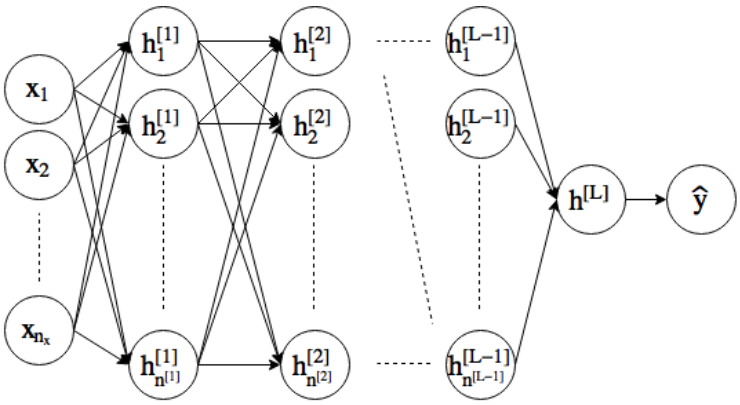


Figure 1: Feedforward neural network.

In developing our framework, which we call DeepTriangle, we also draw inspiration from the existing stochastic reserving literature: Quarg and Mack (2004) utilize incurred loss along with paid loss data, Miranda, Nielsen, and Verrall (2012) incorporate claim count information in addition to paid losses, and Avanzi et al. (2016) consider the dependence between lines of business within an insurer's portfolio.

The approach that we develop differs from existing works in many ways, and has the following advantages. First, it enables joint modeling of paid losses and claims outstanding for multiple companies simultaneously in a single model. In fact, the architecture can also accommodate arbitrary additional inputs, such as claim count data and economic indicators, should they be available to the modeler. Second, it requires no manual input during model updates or forecasting, which means that predictions can be generated more frequently than traditional processes, and, in turn, allows management to react to changes in the portfolio sooner.

2. Neural network preliminaries

For comprehensive treatments of neural network mechanics and implementation, we refer the reader to Goodfellow, Bengio, and Courville (2016) and Chollet and Allaire (2018). In order to establish common terminology used in this paper, we present a brief overview in this section.

We motivate the discussion by considering an example feedforward network with fully connected layers represented in Figure 1, where the goal is to predict an output y from input $x = (x_1, x_2, \ldots, x_{n_x})$, where n_x is the number of elements of x. The intermediate values, $h_j^{[l]}$, known as hidden units, are organized into layers, which try to transform the input data into representations that successively become more useful at predicting the output. The nodes in the figure are computed, for each layer $l = 1, \ldots, L$, as

$$h_j^{[l]} = g^{[l]}(z_j^{[l]}), \tag{1}$$

where

$$z_j^{[l]} = \sum_k w_{jk}^{[l]} h_k^{[l-1]} + b_j^{[l]}. \tag{2}$$

Here, the index j spans $\{1, \ldots, n^{[l]}\}$, where $n^{[l]}$ denotes the number of units in layer l. The functions $g^{[l]}$ are called activation functions, whose values $h_j^{[l]}$ are known as activations. The $w^{[l]}$ values come from matrices $W^{[l]}$, with dimensions $n^{[l]} \times n^{[l-1]}$. Together with the biases $b_j^{[l]}$, they represent the weights, which are learned during training, for layer l.

For $l = 1$, we define the previous layer activations as the input, so that $n^{[0]} = n_x$. Hence, the calculation for the first hidden layer becomes

$$h_j^{[1]} = g^{[1]} \quad \sum_k w_{jk}^{[1]} x_k + b_j^{[1]} \Bigg). \tag{3}$$

Also, for the output layer $l = L$, we compute the prediction

$$\hat{y} = h_j^{[L]} = g^{[L]} \quad \sum_k w_{jk}^{[L]} h_k^{[L-1]} + b_j^{[L]} \Bigg). \tag{4}$$

We can then think of a neural network as a sequence of function compositions $f = f_L \circ f_{L-1} \circ \cdots \circ f_1$ parameterized as $f(x; W^{[1]}, b^{[1]}, \ldots, W^{[L]}, b^{[L]})$.

Each neural network model is specified with a specific loss function, which is used to measure how close the model predictions are to the actual values. During model training, the parameters discussed above are iteratively updated in order to minimize the loss function. Each update of the parameters typically involves only a subset, or mini-batch, of the training data, and one complete pass through the training data, which includes many updates, is known as an epoch. Training a neural network often requires many passes through the data.

3. Neural architecture for loss reserving

As shown in Figure 2, DeepTriangle is a multi-task network with two prediction goals: claims outstanding and paid loss. We construct one model for each line of business and each model is trained on data from multiple companies.

3.1. Training Data

Let indices $1 \leq i \leq I$ denote accident years and $1 \leq j \leq J$ denote development years under consideration. Also, let $\{P_{ij}\}$ and $\{OS_{ij}\}$ denote the *incremental* paid losses and the *total* claims outstanding, respectively.

Then, at the end of calendar year I, we have access to the observed data

$$\{P_{ij} : i = 1, \ldots, I; j = 1, \ldots, I - i + 1\} \tag{5}$$

and

$$\{OS_{ij} : i = 1, \ldots, I; j = 1, \ldots, I - i + 1\}. \tag{6}$$

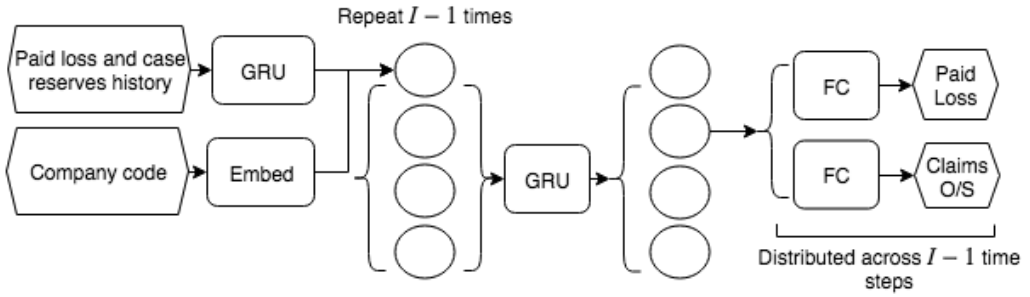


Figure 2: DeepTriangle architecture. *Embed* denotes embedding layer, *GRU* denotes gated recurrent unit, *FC* denotes fully connected layer.

Assume that we are interested in development through the Ith development year; in other words, we only forecast through the eldest maturity in the available data. The goal then is to obtain predictions for future values $\{\widehat{P}_{ij} : i = 2, \ldots, I; j = i+1, \ldots, I\}$ and $\{\widehat{OS}_{ij} : i = 2, \ldots, I; j = i+1, \ldots, I\}$. We can then determine ultimate losses for each accident year $i \in 1, \ldots, I$ by calculating

$$\widehat{UL}_i = \left(\sum_{j=1}^{I-i+1} P_{ij} \right) + \left(\sum_{j=I-i+2}^{I} \widehat{P}_{ij} \right). \tag{7}$$

3.2. Response and predictor variables

In DeepTriangle, each training sample is associated with an accident year-development year pair, which we refer to thereinafter as a *cell*. The response for the sample associated with accident year i and development year j is the sequence

$$(Y_{i,j}, Y_{i,j+1}, \ldots, Y_{i,I-i+1}), \tag{8}$$

where each $Y_{ij} = (P_{ij}, OS_{ij})/NPE_i$, where NPE_i denotes the net earned premium for accident year i. Working with loss ratios makes training more tractable by normalizing values into a similar scale.

The predictor for the sample contains two components. The first component is the observed history as of the end of the calendar year associated with the cell:

$$(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,j-1}). \tag{9}$$

In other words, for each accident year and at each evaluation date for which we have data, we attempt to predict future development of the accident year's paid losses and claims outstanding based on the observed history as of that date. While we are ultimately interested in P_{ij}, the paid losses, we include claims outstanding as an auxiliary output of the model. Since the two quantities are related, we expect to obtain better performance by jointly training than predicting each quantity independently (Collobert and Weston 2008).

The second component of the predictor is the company identifier associated with the experience. Because we include experience from multiple companies in each training

iteration, we need a way to differentiate the data from different companies. We discuss handling of the company identifier in more detail in the next section.

3.3. Model Architecture

DeepTriangle utilizes a sequence-to-sequence architecture inspired by Sutskever, Vinyals, and Le (2014) and Srivastava, Mansimov, and Salakhutdinov (2015).

We utilize gated recurrent units (GRU) (Chung et al. 2014), which is a type of recurrent neural network (RNN) building block that is appropriate for sequential data. A graphical representation of a GRU is shown in Figure 3, and the associated equations are as follows:

$$\tilde{h}^{<t>} = \tanh(W_h[\Gamma_r h^{<t-1>}, x^{<t>}] + b_h) \tag{10}$$

$$\Gamma_r^{<t>} = \sigma(W_r[h^{<t-1>}, x^{<t>}] + b_r) \tag{11}$$

$$\Gamma_u^{<t>} = \sigma(W_u[h^{<t-1>}, x^{<t>}] + b_u) \tag{12}$$

$$h^{<t>} = \Gamma_u^{<t>}\tilde{h}^{<t>} + (1 - \Gamma_u^{<t>})h^{<t-1>}. \tag{13}$$

Here, $h^{<t>}$ and $x^{<t>}$ represent the activation and input values, respectively, at time t, and σ denotes the logistic sigmoid function defined as

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \tag{14}$$

W_h, W_r, W_u, b_h, b_r, and b_u are the appropriately sized weight matrices and biases to be learned.

We first encode the sequential predictor with a GRU to obtain a summary of the historical values. We then repeat the output $I - 1$ times before passing them to a decoder GRU. The factor $I - 1$ is chosen here because for the Ith accident year, we need to forecast $I - 1$ timesteps into the future. Each timestep of the decoded sequence is then concatenated with the company embedding before being passed to two subnetworks, corresponding to the two prediction outputs, of fully connected layers, each of which shares weights across the timesteps.

The company code input is first passed to an embedding layer. In this process, each company is mapped to a fixed length vector in $\mathbb{R}^k$, where k is a hyperparameter. The mapping is learned during the training of the entire network instead of a separate data preprocessing step. Companies that are similar in the context of our claims forecasting problem are mapped to vectors that are close to each other in terms of Euclidean distance. Intuitively, one can think of this representation as a proxy for characteristics of the companies, such as size of book and case reserving philosophy. Categorical embedding is a common technique in deep learning that has been successfully applied to recommendation systems (Cheng et al. 2016) and retail sales prediction (Guo and Berkhahn 2016). In the actuarial science literature, Richman and Wuthrich (2018) utilize embedding layers to capture characteristics of regions in mortality forecasting, while Gabrielli, Richman, and Wuthrich (2018) apply them to lines of business factors in loss reserving.

Rectified linear unit (ReLU) (Nair and Hinton 2010), defined as

$$x \mapsto \max(0, x), \tag{15}$$

is used as the activation function for the fully connected layers, including both of the output layers.

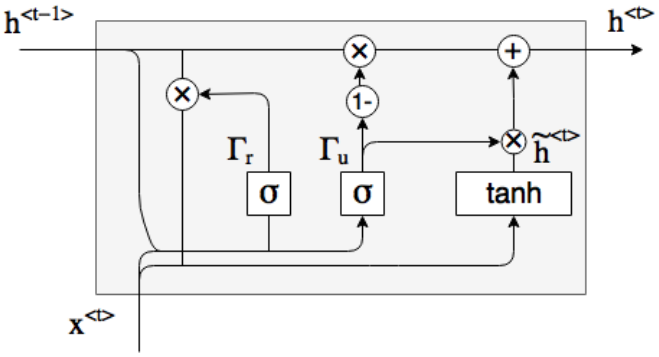


Figure 3: Gated recurrent unit.

3.4. Deployment considerations

While one may not have access to the latest experience data of competitors, the company code predictor can be utilized to incorporate data from companies within a group insurer. During training, the relationships among the companies are inferred based on historical development behavior. This approach provides an automated and objective alternative to manually aggregating, or clustering, the data based on knowledge of the degree of homogeneity among the companies.

If new companies join the portfolio, or if the companies and associated claims are reorganized, one would modify the embedding input size to accommodate the new codes, leaving the rest of the architecture unchanged, then refit the model. The network would then assign embedding vectors to the new companies.

Since the model outputs predictions for each triangle cell, one can calculate the traditional age-to-age, or loss development, factors (LDF) using the model forecasts. Having a familiar output may enable easier integration of DeepTriangle into existing actuarial workflows.

Insurers often have access to richer information than is available in regulatory filings, which underlies the experiments in this paper. For example, in addition to paid and incurred losses, one may include claim count triangles so that the model can also learn from, and predict, frequency information.

4. Experiments

4.1. Data

We validate the modeling approach on data from National Association of Insurance Commissioners (NAIC) Schedule P triangles (Meyers and Shi 2011). The dataset corresponds to claims from accident years 1988-1997, with development experience of 10 years for each accident year.

Following Meyers (2015), we restrict ourselves to a subset of the data which covers four lines of business (commercial auto, private personal auto, workers' compensation, and other liability) and 50 companies in each line of business. This is done to facilitate comparison to existing results.

We use the following variables from the dataset in our study: line of business, company code, accident year, development lag, incurred loss, cumulative paid loss, and net earned premium. Claims outstanding, for the purpose of this study, is derived as incurred loss less cumulative paid loss.

We use data as of year end 1997 for training, and evaluate predictive performance on the development year 10 ultimates.

4.2. Evaluation metrics

We aim to produce scalar metrics to evaluate the performance of the model on each line of business. To this end, for each company and each line of business, we calculate the actual and predicted ultimate losses as of development year 10, for all accident years combined, then compute the root mean squared percentage error (RMSPE) and mean absolute percentage error (MAPE) over companies in each line of business. Percentage errors are used in order to have unit-free measures for comparing across companies with vastly different sizes of portfolios. Formally, if $\mathcal{C}_l$ is the set of companies in line of business l,

$$MAPE_l = \frac{1}{|\mathcal{C}_l|} \sum_{C \in \mathcal{C}_l} \left| \frac{\widehat{UL}_C - UL_C}{UL_C} \right|, \tag{16}$$

and

$$RMSPE_l = \sqrt{\frac{1}{|\mathcal{C}_l|} \sum_{C \in \mathcal{C}_l} \left(\frac{\widehat{UL}_C - UL_C)}{UL_C} \right)^2} \tag{17}$$

where $\widehat{UL}_C$ and UL_C are the predicted and actual cumulative ultimate losses, respectively, for company C.

An alternative approach for evaluation could involve weighting the company results by the associated earned premium or using dollar amounts. However, due to the distribution of company sizes in the dataset, the weights would concentrate on a handful of companies. Hence, to obtain a more balanced evaluation, we choose to report the unweighted percentage-based measures outlined above.

4.3. Implementation and training

The loss function for the each output is computed as the average over the forecasted time steps of the mean squared error of the predictions. The losses for the outputs are then averaged to obtain the network loss. Formally, for the sample associated with cell (i, j), we can write the per-sample loss as

$$\frac{1}{I - i + 1 - (j-1)} \sum_{k=j}^{I-i+1} \frac{(\widehat{P_{ik}} - P_{ik})^2 + (\widehat{OS_{ik}} - OS_{ik})^2}{2}. \tag{18}$$

For optimization, we use the AMSGrad (Reddi, Kale, and Kumar 2018) variant of Adam with a learning rate of 0.0005. We train each neural network for a maximum of 1000 epochs with the following early stopping scheme: if the loss on the validation set does not improve over a 200-epoch window, we terminate training and revert back to the

Table 1: Performance comparison of various models. DeepTriangle and AutoML are abbreviated do DT and ML, respectively.

Line of Business	Mack	ODP	CIT	LIT	ML	**DT**
MAPE						
Commercial Auto	0.060	0.217	0.052	0.052	0.068	**0.043**
Other Liability	0.134	0.223	0.165	0.152	0.142	**0.109**
Private Passenger Auto	0.038	0.039	0.038	0.040	0.036	**0.025**
Workers' Compensation	0.053	0.105	0.054	0.054	0.067	**0.046**
RMSPE						
Commercial Auto	0.080	0.822	0.076	0.074	0.096	**0.057**
Other Liability	0.202	0.477	0.220	0.209	0.181	**0.150**
Private Passenger Auto	0.061	0.063	0.057	0.060	0.059	**0.039**
Workers' Compensation	0.079	0.368	0.080	0.080	0.099	**0.067**

weights on the epoch with the lowest validation loss. The validation set used in the early stopping criterion is defined to be the subset of the training data that becomes available after calendar year 1995. For each line of business, we create an ensemble of 100 models, each trained with the same architecture but different random weight initialization. This is done to reduce the variance inherent in the randomness associated with neural networks.

We implement DeepTriangle using the keras R package (Chollet, Allaire, and others 2017) with the TensorFlow (Abadi et al. 2015) backend. Code for producing the experiment results is available online.[1]

4.4. Results and discussion

In Table 1 we tabulate the out-of-time performance of DeepTriangle against other models: the Mack chain-ladder model (Mack 1993), the bootstrap ODP model (England and Verrall 2002), an AutoML model, and a selection of Bayesian Markov chain Monte Carlo (MCMC) models from Meyers (2015) including the correlated incremental trend (CIT) and leveled incremental trend (LIT) models. For the stochastic models, we use the means of the predictive distributions as the point estimates to which we compare the actual outcomes. For DeepTriangle, we report the averaged predictions from the ensembles.

The AutoML model is developed by automatically searching over a set of common machine learning techniques. In the implementation we use, it trains and cross-validates a random forest, an extremely-randomized forest, a random grid of gradient boosting machines, a random grid of deep feedforward neural networks, and stacked ensembles thereof (The H2O.ai team 2018). Details of these algorithms can be found in Friedman, Hastie, and Tibshirani (2001). Because the machine learning techniques produce scalar outputs, we use an iterative forecasting scheme where the prediction for a timestep is used in the predictor for the next timestep.

We see that DeepTriangle improves on the performance of the popular chain ladder and ODP models, common machine learning models, and Bayesian stochastic models.

[1] https://github.com/kevinykuo/deeptriangle

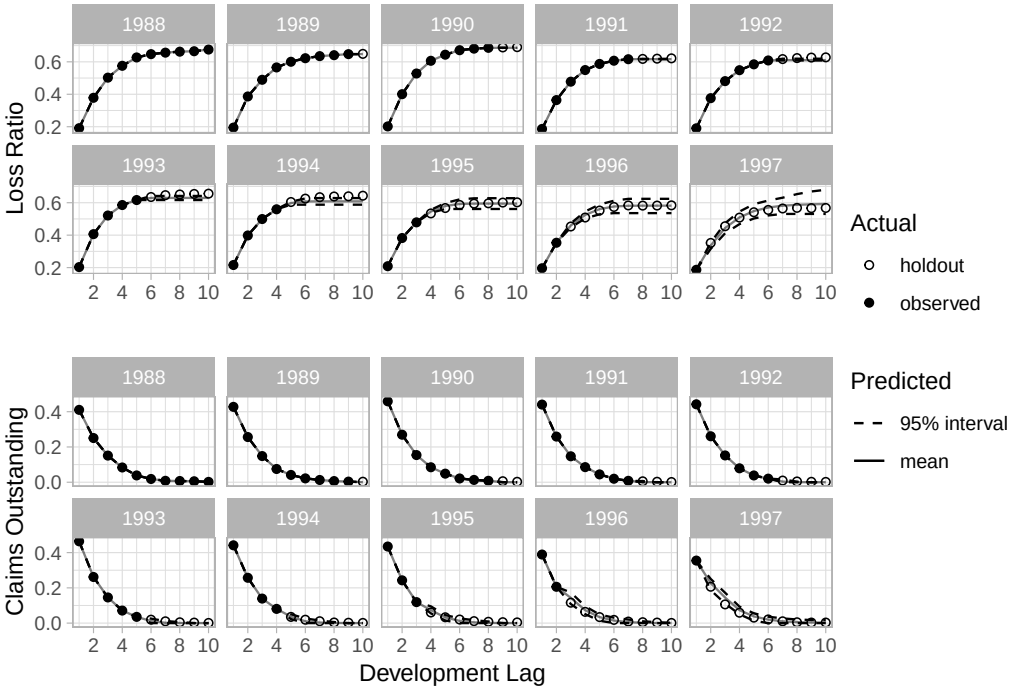


Figure 4: Development by accident year for Company 1767, commercial auto.

In addition to aggregated results for all companies, we also investigate qualitatively the ability of DeepTriangle to learn development patterns of individual companies. Figures 4 and 5 show the paid loss development and claims outstanding development for the commercial auto line of Company 1767 and the workers' compensation line of Company 337, respectively. We see that the model captures the development patterns for Company 1767 reasonably well. However, it is unsuccessful in forecasting the deteriorating loss ratios for Company 337's workers' compensation book.

We do not study uncertainty estimates in this paper nor interpret the forecasts as posterior predictive distributions; rather, they are included to reflect the stochastic nature of optimizing neural networks. We note that others have exploited randomness in weight initialization in producing predictive distributions (Lakshminarayanan, Pritzel, and Blundell 2017), and further research could study the applicability of these techniques to reserve variability.

5. Conclusion

We introduce DeepTriangle, a deep learning framework for forecasting paid losses. Our models are able to attain performance comparable, by our metrics, to modern stochastic reserving techniques without expert input. By utilizing neural networks, we can incorporate multiple heterogeneous inputs and train on multiple objectives simultaneously, and also allow customization of models based on available data.

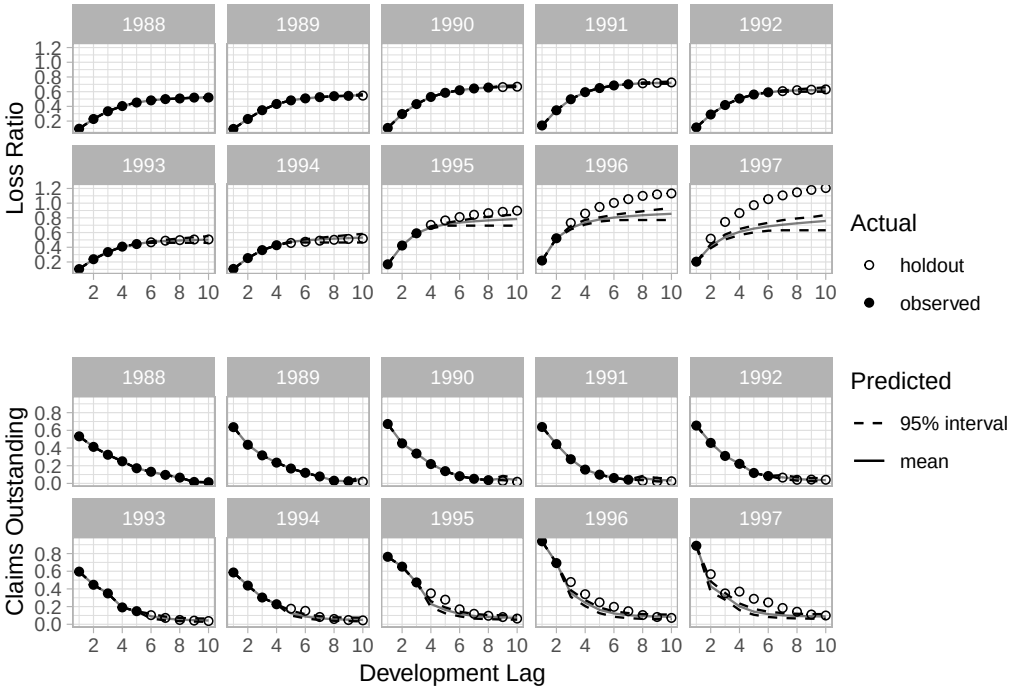


Figure 5: Development by accident year for Company 337, workers' compensation.

We analyze an aggregated dataset with limited features in this paper because it is publicly available and well studied, but one can extend DeepTriangle to incorporate additional data, such as claim counts.

Deep neural networks can be designed to extend recent efforts, such as Wüthrich (2018a), on applying machine learning to claims level reserving. They can also be designed to incorporate additional features that are not handled well by traditional machine learning algorithms, such as claims adjusters' notes from free text fields and images.

While this study focuses on prediction of point estimates, future extensions may include outputting distributions in order to address reserve variability.

References

Abadi, Martı'n, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, et al. 2015. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems." https://www.tensorflow.org/.

Avanzi, Benjamin, Greg Taylor, Phuong Anh Vu, and Bernard Wong. 2016. "Stochastic Loss Reserving with Dependence: A Flexible Multivariate Tweedie Approach." *Insurance: Mathematics and Economics* 71. Elsevier: 63–78.

Cheng, Heng-Tze, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, et al. 2016. "Wide & Deep Learning for Recommender Systems." *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems - DLRS 2016*. ACM Press. https://doi.org/10.1145/2988450.2988454.

Chollet, Francois, and JJ Allaire. 2018. *Deep Learning with R*. Manning Publications.

Chollet, François, JJ Allaire, and others. 2017. "R Interface to Keras." https://github.com/rstudio/keras; GitHub.

Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling."

Collobert, Ronan, and Jason Weston. 2008. "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning." In *Proceedings of the 25th International Conference on Machine Learning*, 160–67. ACM.

England, Peter D, and Richard J Verrall. 2002. "Stochastic Claims Reserving in General Insurance." *British Actuarial Journal* 8 (3). Cambridge University Press: 443–518.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. 10. Springer series in statistics New York, NY, USA:

Gabrielli, Andrea, Ronald Richman, and Mario V Wuthrich. 2018. "Neural Network Embedding of the over-Dispersed Poisson Reserving Model." *Available at SSRN*.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press Cambridge.

Guo, Cheng, and Felix Berkhahn. 2016. "Entity Embeddings of Categorical Variables." *CoRR* abs/1604.06737. http://arxiv.org/abs/1604.06737.

Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. 2017. "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles." In *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, 6402–13. Curran Associates, Inc. http://papers.nips.cc/paper/7219-simple-and-scalable-predictive-uncertainty-estimation-using-deep-ensembles.pdf.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553). Nature Publishing Group: 436.

Mack, Thomas. 1993. "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." *Astin Bulletin* 23 (2): 213–25.

Meyers, Glenn. 2015. *Stochastic Loss Reserving Using Bayesian MCMC Models*. Casualty Actuarial Society.

Meyers, Glenn, and Peng Shi. 2011. "Loss Reserving Data Pulled from NAIC Schedule P." http://www.casact.org/research/index.cfm?fa=loss_reserves_data.

Miranda, Marı'a Dolores Martı'nez, Jens Perch Nielsen, and Richard Verrall. 2012. "Double Chain Ladder." *ASTIN Bulletin: The Journal of the IAA* 42 (1). Cambridge

University Press: 59–76.

Nair, Vinod, and Geoffrey E Hinton. 2010. "Rectified Linear Units Improve Restricted Boltzmann Machines." In *Proceedings of the 27th International Conference on Machine Learning (Icml-10)*, 807–14.

Quarg, Gerhard, and Thomas Mack. 2004. "Munich Chain Ladder." *Blätter Der DGVFM* 26 (4). Springer: 597–630.

Reddi, Sashank J., Satyen Kale, and Sanjiv Kumar. 2018. "On the Convergence of Adam and Beyond." In *International Conference on Learning Representations*. https://openreview.net/forum?id=ryQu7f-RZ.

Richman, Ronald, and Mario V Wuthrich. 2018. "A Neural Network Extension of the Lee-Carter Model to Multiple Populations." *Available at SSRN 3270877*.

Srivastava, Nitish, Elman Mansimov, and Ruslan Salakhutdinov. 2015. "Unsupervised Learning of Video Representations Using Lstms." *CoRR* abs/1502.04681. http://arxiv.org/abs/1502.04681.

Sutskever, Ilya, Oriol Vinyals, and Quoc V Le. 2014. "Sequence to Sequence Learning with Neural Networks." In *Advances in Neural Information Processing Systems*, 3104–12.

The H2O.ai team. 2018. *H2o: R Interface for H2o*. http://www.h2o.ai.

Wüthrich, Mario V. 2018a. "Machine Learning in Individual Claims Reserving." *Scandinavian Actuarial Journal*. Taylor & Francis, 1–16.

———. 2018b. "Neural Networks Applied to Chain–Ladder Reserving." *European Actuarial Journal* 8 (2). Springer: 407–36.

Back-Testing the ODP Bootstrap & Mack Bootstrap Models

Mark R. Shapland, FCAS, FSA, FIAI, MAAA

Abstract

Motivation. Distributions of unpaid claims are gaining importance within the actuarial community as management, regulators, and others look to the actuarial profession for a quantitative approach to evaluating risk. Actuaries have historically applied their judgment to determine if a best estimate is reasonable, but how do we know if the models used to produce distributions are reasonable? Determining if a distribution is reasonable is a much more complex task than for a point estimate. Is the model producing a reasonable estimate at the 95th percentile? Is it producing reasonable distribution shapes? In effect, actuarial judgment shifts focus from a single point estimate to the entire distribution and we must rely, at least in part, on the proposition that "if the theory is acceptable then the distribution is acceptable." Therefore, the purpose of this paper is to determine if the theory really holds up in practice.

There are five objectives of this research. First, by greatly expanding the database used to back-test models the testing can provide more evidence to validate (or not) prior research and address any weaknesses in the prior research. Second, all of the prior research focused only on the estimate of a single outcome (i.e., the ultimate for the current accident year), so this research expands the testing for every possible estimate, e.g., each accident period, each calendar period, each incremental cell, etc. Third, more models were tested and some of the model assumptions were tested in order to expand our understanding of the predictive value of different models. Fourth, recent proposals to address model weaknesses were examined to assess their viability. Fifth, a new proposal for using this research to benchmark unpaid claim estimates will be put forth.

Method. The estimated distribution of possible outcomes for various models based on the ODP Bootstrap model and the Mack Bootstrap model are saved and compared to the actual outcome up to 9 years later – i.e., a single back-test. While the result from a single data set is not indicative of the quality of the original estimate, comparing results for a large number of data sets does provide an indication of the quality of the model.

Results. Based on the back-testing, all tested models appear to underestimate the width of the "true" distribution but some of the models tested appeared to get closer to the "true" distribution than others and the tested adjustments to the model assumptions seem to improve the results, which is a desirable quality. Another key result is to show how the insurance underwriting cycle also impacts the results of the back testing.

Conclusions. The major results from prior similar research is confirmed, but the volume of this research has led to a new approach to benchmarking both deterministic and stochastic unpaid claim estimates in practice.

Keywords. Back-test, benchmark, bootstrap, chain ladder, Mack model, over-dispersed Poisson, reserve variability, systemic risk, underwriting cycle.

1. INTRODUCTION

Enterprise Risk Management has been at the leading edge of effectively managing insurance and other risk bearing operations for many years. Its use is expected to grow, and perhaps accelerate, into the foreseeable future as regulators and rating agencies focus on risk based approaches. One of the key metrics in any risk model used for ERM is the variability of unpaid claims as these are normally the largest liability in the balance sheet. While many

stochastic models provide the diagnostic tools for calibrating the assumptions of the model, there are no tools for gauging the quality of unpaid claim variability estimates. After vetting the theory underlying a model, the only way to gain valuable insights into the quality of the model is to back-test the results to see how well the models predicted the actual outcomes.

To calculate a distribution of possible outcomes and select unpaid claim estimates at a confidence level of say 75%, or to demonstrate that reserves are at such a level, is not a straight-forward process. Indeed, it is not a process that can be performed exactly or by using purely statistical approaches. Reasons for this include, but are not limited to, the following:

- Uncertainty in reserving can be attributable to three types of risk: process risk, parameter risk and model risk. Of these, process risk, and to a lesser extent parameter risk, can be assessed statistically and then only to the extent permitted by the volume and quality of available data. Model risk does not, in general, follow clear statistical patterns.
- Where process or parameter risk can be assessed statistically, the available historical data will not show the full breadth of the possible outcomes (i.e., variability). The resulting uncertainty in any outcomes will increase the further one moves away from the mean.
- New lines of business will have little or no data on which to assess variability due to process or parameter risk. The statistical credibility of the data for small volumes of business will also be limited.
- The assessment of process or parameter risk can be distorted by historic data including the effect of systemic risks, e.g., changes in case law that affect claim settlement amounts.

Therefore, any assessment of reserves at a particular confidence level will require the reserving actuary to exercise judgment to a significant degree. This is similar to how actuaries currently assess deterministic unpaid claim estimates, where actuaries use tools (such as the Chain Ladder ("CL") and the Bornhuetter-Ferguson ("BF") methods) to calculate a central estimate of the claims liabilities. Based in part on their knowledge of the strengths and weaknesses of the methods, they exercise considerable judgment in selecting factors and parameters, in adjusting for trends and for known or expected distortions, and in selecting the amounts to be booked.

For stochastic models, with sufficient data the process and parameter risk would usually be evaluated using stochastic tools applied to the historic data. Different data (e.g., paid data

and incurred data) and different models would generate different results and different Coefficients of Variation ("CoVs"). Judgment is needed in deciding which CoVs would be appropriate to address model risk in addition to process and parameter risk.

Given that prior research has shown that the ODP Bootstrap and other models tend to underestimate the "true" variability, the actuary will need support for helping to inform their judgments about estimates of possible outcomes. Similar to benchmarks for deterministic assumptions, benchmark CoVs would be a very useful addition to the actuary's toolkit as a means of sense checking the estimated distributions. Thus, a primary use of this research is to provide benchmarks for distributions of possible outcomes for insurance data.

Even with benchmarks of CoVs by line of business, the actuary would need to combine these across all business lines. By definition, process risk should be independent of other risk factors (and across lines of business) but there may well be some degree of contagion (i.e., large losses that affect multiple lines of business) and/or correlation between the other factors. In order to combine the CoVs, correlation matrices will be required. Again, judgment is required, but another key benchmark from this research is estimated correlations based on industry data.

1.1 Research Context

Because it is such a critical part of effective actuarial practice, it seems likely that understanding the effectiveness of a method has been part of the research from the early days of actuarial science. For deterministic reserving methods, one of the earlier papers on the effectiveness of methods is Skurnick [18] and more recent examples include Forray [6] and Jing, Lebens, and Lowe [9]. For deterministic methods it is often enough to focus on the theory to understand the strengths and weaknesses of a method. For example, all actuaries learn early in their career that the chain ladder method will tend to underestimate the current period when the initial development period outcome is lower than average, and tend to overestimate the current period when the initial development period outcome is higher than average.

If we consider a triangle of data as illustrated in Graph 1.1, the goal of estimating unpaid claims is to estimate the unpaid amounts, $u(w,d)$, by projecting the cumulative amounts, $c(w,d)$.[1] The total reserve for an accident period, $R(w)$, can be estimated directly or indirectly

[1] For ease of exposition, the notation $c(w,d)$ and $u(w,d)$ does not specify cumulative or incremental values. The reader can infer cumulative or incremental values depending on their use.

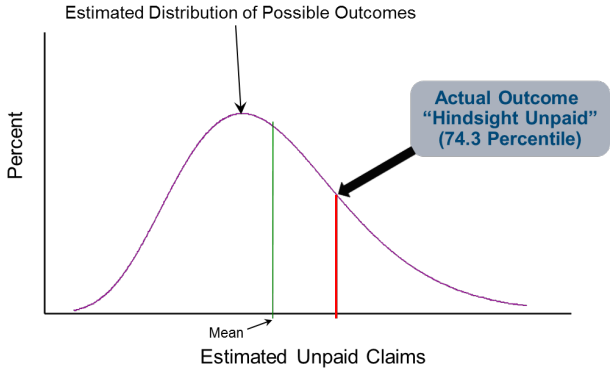
as a sum of the incremental unpaid amounts. For the chain ladder method, the estimation of *R(10)* is done using a factor times c(10,1), so it is easy to visualize how dependent this calculation is to the relative size of c(10,1).

Graph 1.1. Triangle of Data with Estimated Unpaid

	1	2	3	4	5	6	7	8	9	10	Total
1	c(1,1)	c(1,2)	c(1,3)	c(1,4)	c(1,5)	c(1,6)	c(1,7)	c(1,8)	c(1,9)	c(1,10)	
2	c(2,1)	c(2,2)	c(2,3)	c(2,4)	c(2,5)	c(2,6)	c(2,7)	c(2,8)	c(2,9)	*u(2,10)*	*R(2)*
3	c(3,1)	c(3,2)	c(3,3)	c(3,4)	c(3,5)	c(3,6)	c(3,7)	c(3,8)	*u(3,9)*	*u(3,10)*	*R(3)*
4	c(4,1)	c(4,2)	c(4,3)	c(4,4)	c(4,5)	c(4,6)	c(4,7)	*u(4,8)*	*u(4,9)*	*u(4,10)*	*R(4)*
5	c(5,1)	c(5,2)	c(5,3)	c(5,4)	c(5,5)	c(5,6)	*u(5,7)*	*u(5,8)*	*u(5,9)*	*u(5,10)*	*R(5)*
6	c(6,1)	c(6,2)	c(6,3)	c(6,4)	c(6,5)	*u(6,6)*	*u(6,7)*	*u(6,8)*	*u(6,9)*	*u(6,10)*	*R(6)*
7	c(7,1)	c(7,2)	c(7,3)	c(7,4)	*u(7,5)*	*u(7,6)*	*u(7,7)*	*u(7,8)*	*u(7,9)*	*u(7,10)*	*R(7)*
8	c(8,1)	c(8,2)	c(8,3)	*u(8,4)*	*u(8,5)*	*u(8,6)*	*u(8,7)*	*u(8,8)*	*u(8,9)*	*u(8,10)*	*R(8)*
9	c(9,1)	c(9,2)	*u(9,3)*	*u(9,4)*	*u(9,5)*	*u(9,6)*	*u(9,7)*	*u(9,8)*	*u(9,9)*	*u(9,10)*	*R(9)*
10	c(10,1)	*u(10,2)*	*u(10,3)*	*u(10,4)*	*u(10,5)*	*u(10,6)*	*u(10,7)*	*u(10,8)*	*u(10,9)*	*u(10,10)*	*R(10)*
Total											*R(T)*

This understanding of deterministic methods is largely possible because the focus of the method is a central estimate. For stochastic models, whose focus is the entire distribution, the same principles for the central estimate still apply, but understanding the entire distribution is impossible with a single observation. For example, it is common for a stochastic model to be used to simulate 10,000 possible outcomes for *R(10)*, but what if we later determine that the actual outcome was at the 74.3 percentile, as illustrated in Graph 1.2.

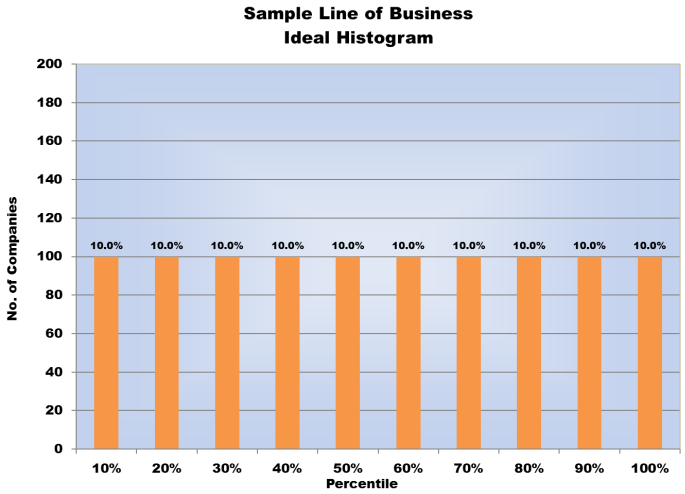
Graph 1.2. Back-Test of Estimated Distribution of Possible Outcomes



What does that tell us about the model? Did we get the mean wrong? What about the width of the distribution? We have no way to know with only one observation compared to our estimated distribution. Therefore, back-testing a large number of observations is essential to see if all parts of the distribution are represented in the outcomes. Another point to keep in mind is that when back-testing a model the mean of the estimated distribution is assumed to be the booked reserve even if the available data contains the actual booked reserve in order to test the efficacy of the model and not the judgment of the actuary selecting the reserve.

Testing all parts of a distribution can be illustrated graphically. For example, with 1,000 data sets, if the "true" distribution of the possible outcomes is fairly represented by the model then each decile group of the actual outcomes in a histogram should ideally contain 100 observations, as illustrated in Graph 1.3. For example, if the outcome from Graph 1.2 is included as one of the 1,000 datasets, then it would be one of the 100 in the bar labeled 80% (representing all outcomes greater than 70% and less than or equal to 80%) in Graph 1.3.

Graph 1.3. Ideal Histogram



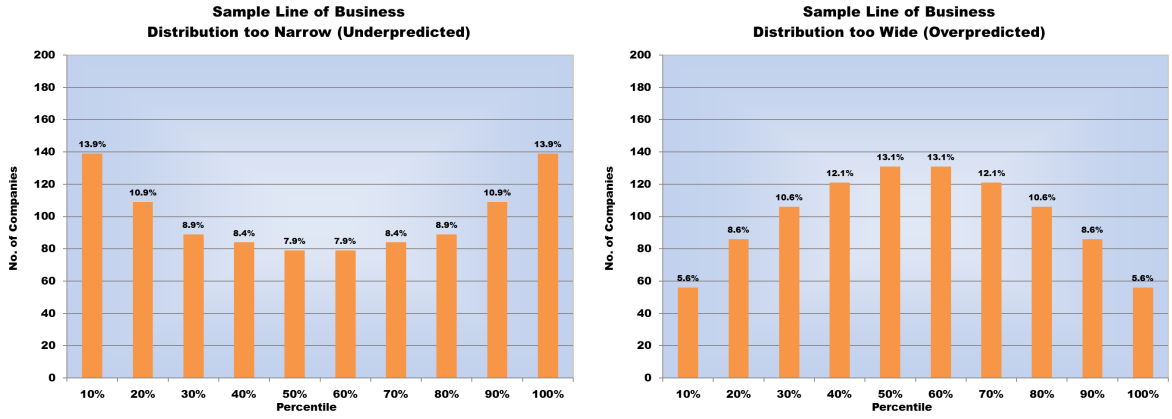
In Graph 1.3, like most of the similar graphs in the remainder of the paper, the percentiles along the X-axis are the decile groups for the percentile of the actual outcome compared to the estimated distribution. The Y-axis shows the number of companies or datasets in the bars, with the percent of the total number of companies or datasets as the bar labels.

If the model being back-tested is under predicting the "true" distribution then the histogram would show a higher than average number of observations at the extremes, say below the 20th percentile and above the 80th percentile, and it would show a lower than average number of observations in the middle percentiles. If the model being back-tested is over predicting the "true" distribution the histogram would show a lower than average number of observations at the extremes and higher than average observations in the middle percentiles. These two types of results are illustrated in Graph 1.4.

Of course, when back-testing real (or simulated) data the actual histograms will include random noise which could mask or partially mask the results, but typically the shape of the histogram will be indicative of the result even if random noise makes conclusions about a specific percentile problematic. The impact of random noise on the histogram can at least be

partially minimized by increasing the sample size to take advantage of the law of large numbers. This approach to understanding the effectiveness of stochastic models has been used by a number of researchers, but only a few key papers will be highlighted in Section 2.

Graph 1.4. Under & Over Prediction Histograms



1.2 Objectives

There are five objectives of this research. First, by expanding the database used to back-test models the testing can provide more evidence to validate (or not) prior research and address any weaknesses in the prior research. Second, all of the prior research focused only on the estimate of a single outcome, specifically the estimate of *R(10)* from Graph 1.1. For this research the outcomes for all possible estimates from Graph 1.1., e.g., each accident period, each calendar period, each incremental cell, etc., were included in the testing to see if any other insights can be gained by expanding the testing. Third, more models were tested and some of the model assumptions were tested in order to expand our understanding the predictive value of different models. Fourth, recent proposals to address model weaknesses were examined to assess their viability. Fifth, a new proposal for using this research to benchmark unpaid claim estimates will be put forth.

1.3 Outline

The remainder of the paper proceeds as follows. Section 2 will provide an overview of the prior research and proposed solutions. In Section 3, the data and the process used to validate it for the back-testing are described. Next, Section 4 will focus on the testing process. Then, in Section 5 the results of the back-testing are summarized, with additional details included as Appendix A. Finally, in Section 6 a process for using this research to benchmark unpaid claim estimates will be described.

2. OVERVIEW OF PRIOR TESTING

Other researchers have used back-testing to evaluate the quality of stochastic models, but providing an in depth review of prior work is beyond the scope of this paper. Since one of the objectives of this paper is to validate (or not) the prior research, some of the prior research is included in the References section for the interested reader and some highlights are included here. Note however, that the highlights discussed here are not intended to give a complete overview of these papers and other valuable insights could be gained by reading the original research papers.

Two early examples of back-testing stochastic models are the product of GIRO Working Parties [15, 16] in the U.K. in 2007 and 2008. The 2007 Working Party reviewed a number of models with a few real datasets, but also created simulated data (designed to meet all of the conditions/assumptions of the respective model) to more thoroughly test the ODP Bootstrap and Mack models. The 2008 Working Party expanded the simulation testing of the 2007 Working Party by creating a wider variety of simulated datasets (e.g., different triangle sizes). The back-testing was based on 10,000 samples of each simulated dataset for the ODP Bootstrap (paid chain ladder only) and closed form Mack models.

In theory at least, this testing was designed to see how well the model predicted outcomes for "perfect" data. The Working Parties also noted that simulated data was a good first step as it allows for controlled testing, but they also recognized that real data can include shocks and other anomalies which is likely to cause predicted results to be more inaccurate than simulated data. Interestingly, even with the "perfect" datasets the Working Parties concluded that:

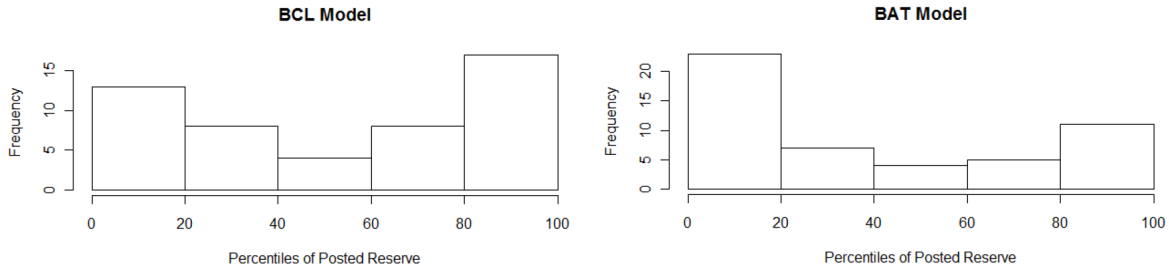
- The results for the Mack model exceeded the predicted 99th percentile 8.4% of the time for a 10 x 10 triangle, indicating the Mack model significantly under predicted the extreme outcomes. As the triangle size was increased to 100 x 100, the under prediction of the extreme outcomes reduced to 2.1% for the Mack model.
- The results for the ODP Bootstrap model exceeded the predicted 99th percentile 2.6% of the time, which also indicated an under prediction. As the triangle size increased for the ODP Bootstrap model the error rate stayed consistent.

In Meyers & Shi [12], the authors based their back-testing of the ODP Bootstrap model[2] on a database of 1997 Schedule P paid data from 50 companies. While the size of the

[2] The authors also proposed and tested a Bayesian Autoregressive Tweedie (BAT) model.

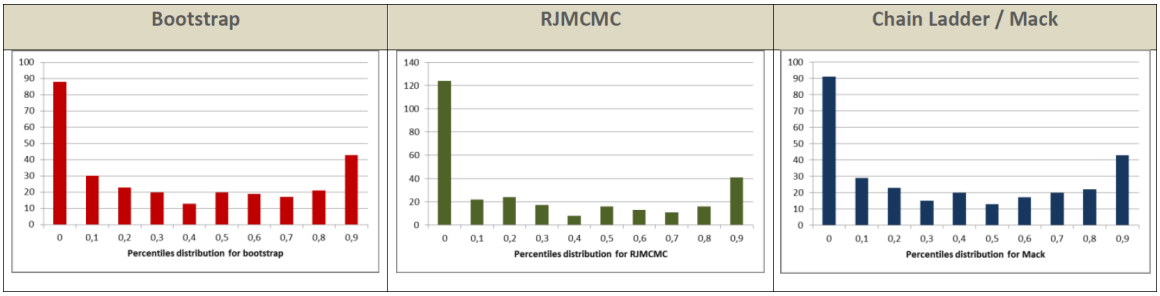
database was not sufficient to arrive at definitive conclusions, the authors recommended further testing and noted that their study "suggests that there might be environmental changes that no single model can identify" and "the actuarial profession cannot rely solely on stochastic loss reserve models to manage its reserve risk." To summarize the back-testing results, the authors included Graph 2.1, which show results for their tests of the Bootstrap Chain Ladder (BCL) model and the Bayesian Autoregressive Tweedie (BAT) model. Similar to the description above for Graph 1.3, the "frequency" label for the Y-axis represents the number of companies in each 20% group bar of the histograms. For the data as of 31 December 1997, only the current accident year was tested.

Graph 2.1. Percentile Results for Meyers & Shi



In Gremillet, Miehe & Zanón [7], the authors based their back-testing of the ODP Bootstrap[3] model on 296 triangles from four lines of business in the database created for the CAS by Meyers & Shi using 1997 Schedule P paid data. The authors concluded "it is core to have adjustments by actuaries prior to running the stochastic methods 'automatically'" and that "it seems that the 'actuary in the box' dream for stochastic reserves valuation is not yet happening…" To summarize the back-testing results, the authors included Graph 2.2, which show the results for the three models they tested. Similar to Meyers & Shi, only the current accident year was tested for the 1997 data.

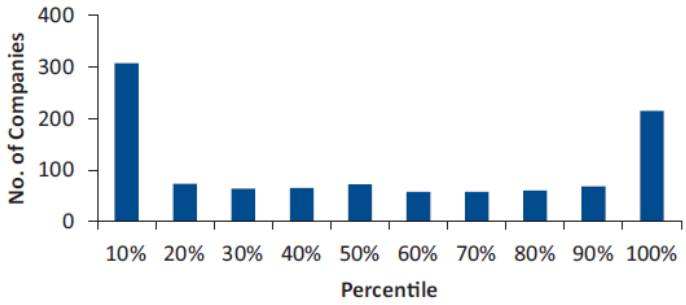
Graph 2.2. Percentile Results for Gremillet, Miehe & Zanón



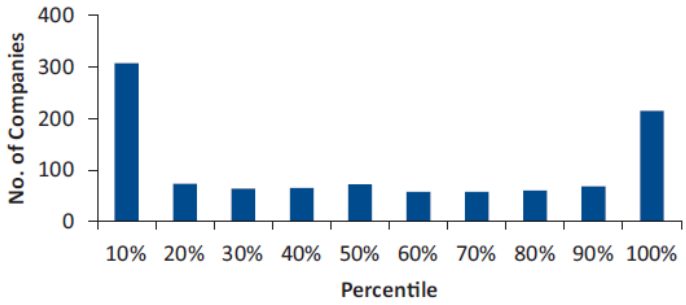
[3] The authors also tested the Reversible Jump Markov Chain Monte Carlo model and the Mack model.

In Leong, Wang & Chen [11], the authors based their back-testing of the ODP Bootstrap model on seven lines of business[4] for approximately 4,850 triangle datasets[5] from a database of Schedule P data from 1989 to 2002. The authors concluded "the popular ODP Bootstrap of the paid chain-ladder method is underestimating reserve risk" and that "it is because the bootstrap model does not consider systemic risk, or, to put it another way, the risk that future trends in the claims environment—such as inflation, trends in tort reform, legislative changes, etc.—may deviate from what we saw in the past". To summarize the back-testing results, the authors included Graph 2.3 showing the results for Homeowners and similar graphs for the other lines of business.

Graph 2.3. Results for Leong, Wang & Chen (HO – Paid CL – All Years)



Graph 2.4. Results for Leong, Wang & Chen (WC – Incurred CL – All Years)



In Leong, Wang & Chen [11], the authors then expanded their back-testing of the ODP Bootstrap to see if using the model for incurred data improved the models predictive power. The authors concluded "it appears that the incurred bootstrap model is also underestimating the risk of falling in these extreme percentiles" as illustrated in Graph 2.4 for Workers'

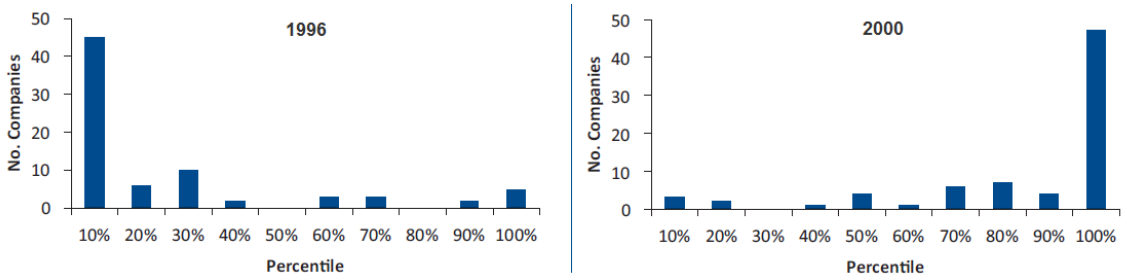
[4] For some years, data for Medical Professional Liability and Other Liability is split between Claims Made and Occurrence policies. In order to use this data consistently over all years, the parts are combined into one line of business. Thus, technically nine lines of business were included, but for four of the lines the splits were grouped to include both Claims Made and Occurrence.

[5] The authors do not state the actual number of datasets, but they do note that the line of business with the most data came from 78 companies and the line of business with the least data came from 21 companies. To estimate the total number of datasets, if we assume an average of 49.5 companies per line of business, 7 lines of business and 14 years, then 49.5 x 7 x 14 = 4,851.

Compensation.

An additional insight from the Leong, Wang & Chen [11] research was possible due to their use of multiple years to show that the reserving cycle has an impact on the results. The results in Graphs 2.3 and 2.4 are for all years combined, but results by year were also included by the authors such as Homeowners for 1996 and 2000 in Graph 2.5.

Graph 2.5. Results for Leong, Wang & Chen (HO – Paid CL – 1996 & 2000)



The authors also illustrated the reserving cycle for the industry in Graph 2.6 which shows that in 1996 the overall reserve level for the industry was too high and in 2000 it was too low. The left side histogram in Graph 2.5 corresponds to when the industry was over reserved and the back-testing resulted in a disproportionate number of outcomes less than 10%, which makes sense.[6] The right side histogram in Graph 2.5 also makes sense as the industry was under reserved in 2000, which leads to a disproportionate number of outcomes above 90%.[7]

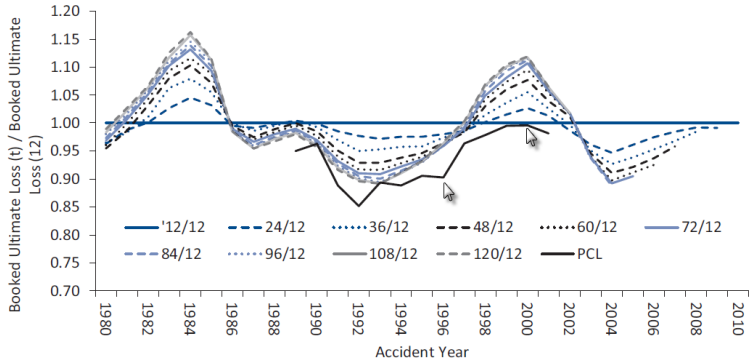
This insight led the authors to conclude that the ODP Bootstrap model only measures independent risk (arising from the randomness inherent in the insurance process) and not systemic risk (arising from the whole system). While there is a certain appeal to this conclusion, it seems the definition of systemic risk could be split into "internal systemic" risk (arising from within the modeling framework) and "external systemic" risk (arising from the outside the modeling framework). By using a broad definition of systemic risk the authors ignored weaknesses of the ODP Bootstrap model that contribute to this result. Their focus

[6] In Graph 2.6, the initial reserves at 12 months are the solid line for 1.00. For 1996, as the accident year matures the ultimate value gets lower and lower and at 120 months is a little over 90% of the ultimate at 12 months, i.e., the initial reserves were too high. In this case, if the initial mean of the simulated distributions was too high, say at 100, then when the final outcome is known if the mean should have been lower, say 90, then the odds that the actual random outcome is below 10% is increased, all else being equal.

[7] For 2000, as the accident year matures the ultimate value gets higher and higher and at 120 months is about 112% of the ultimate at 12 months, i.e., the initial reserves were too low. In this case, if the initial mean of the simulated distributions was too low, say at 100, then when the final outcome is known if the mean should have been higher, say 112, then the odds that the actual random outcome is above 90% is increased, all else being equal.

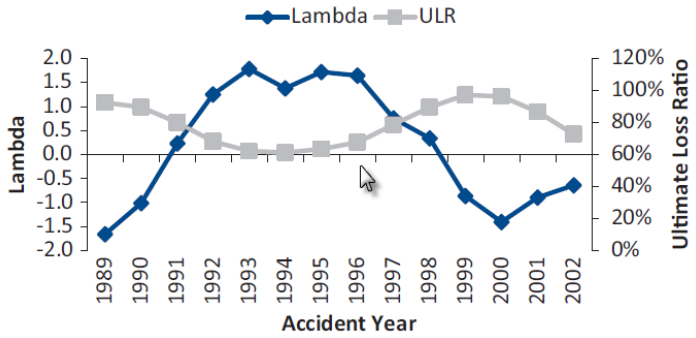
on systemic risk resulted in two methods to adjust for systemic risk in the ODP Bootstrap model, the systemic risk distribution method[8] and the Wang transform adjustment,[9] which allows the authors to show how the combination of both of these methods "fixed" the back-testing results.

Graph 2.6. Reserving Cycle in Leong, Wang & Chen



Digging deeper into the methods proposed by Leong, Wang & Chen [11] it seems that while their results do "correct" the back-testing results, the methods ignore a weakness of the ODP Bootstrap model, are backward looking only, and therefore should be used cautiously as a tool to adjust current ODP Bootstrap results. Starting with the variance adjustments, Graph 2.7 illustrates how when the ultimate loss ratio is less than the initial loss ratio (as in 1996) the variance is increased by Lambda, but this is not logical.

Graph 2.7. Comparison of Lambda and ULR in Leong, Wang & Chen



If the initial ultimate estimate is too high, a typical chain ladder method is likely to be

[8] The authors conclude that the ODP Bootstrap model only measures independent risk and not systemic risk. For the systemic risk distribution method a benchmark systemic risk distribution is estimated and combined with the independent risk distribution from the ODP Bootstrap model to obtain the total risk distribution.

[9] For the Wang transform adjustment, the authors note that the ODP Bootstrap estimate is biased (in their words "it is not assumed to be unbiased") so the adjustment tries to estimate the systemic bias over the course of the reserving cycle.

overestimating the central estimate due to a larger than average initial cumulative value in the current accident year. Extending this to the ODP Bootstrap, this overestimation from the chain ladder elements of the model also causes the variability of the future incremental values to be overestimated.[10] Logically, when the initial ultimate is overestimated the Lambda value should decrease the variance and vice versa. Thus, the Lambda proposed by the authors is above one when it should be below one and vice versa.[11]

Moving to the mean adjustments, the authors note that the Wang transform shifts the distribution to account for the estimation error. This implies that when the initial ultimate losses are overestimated (as in 1996) the shifting will reduce the mean of the distribution. Looking back at Graph 2.5 for 1996, the initial overestimation of the mean is a major contributor to why so many of the outcomes ended up in the lowest decile. Based on the combination of these two adjustments it makes sense that they "corrected" the historical biases in the model results. However, in order for this to have practical value looking forward the actuary would be required to guess at which part of the reserving cycle they are currently in and then select a Lambda which is opposite of what is indicated by the proposed formulas.

As noted above, this review of the Leong, Wang & Chen [11] paper indicates that their formulas should be used with caution when adjusting an estimated distribution from the ODP Bootstrap model, but this realization led to an alternative approach.[12] In summary, Leong, Wang & Chen [11] use a formula based approach for a single model to adjust an estimated distribution based solely on the data used for the estimated distribution. Alternatively, by using a very large database of outcomes from multiple models it becomes possible to create customizable benchmarks of unpaid claim distributions which can be used as a guide regardless of the model(s) being used by the actuary. Because of the cyclical bias in the mean noted above, another advantage of using benchmarks is that this approach assumes the actuary will address the bias in the mean and the benchmark can adjust for the remaining biases.

---

[10] To add process variance to the simulated outcomes, each future incremental value is assumed to be the mean and the variance is the mean times the Scale Parameter. Thus, if the future incremental values tend to be "too large" due to the chain ladder extrapolation of the first cell, then the variance of the sampled values will also tend to be "too large."

[11] The authors comment on the significant negative correlation between Lambda and the ultimate loss ratio at the end of Section 7.2.2.

[12] Of course the authors may refute this conclusion or perhaps use this review to revise their formulas to better address the issues driven by the reserve cycle.

3. DATA USED IN TESTING

The data used in this research includes net loss and ALAE data from nearly 31,000 real data sets (i.e., paid claim triangles, incurred claim triangles, earned premiums, etc.) for all 16 Schedule P lines of business, spanning 9 years from 1996 to 2004.[13] For each of these data sets, the actual results over the next nine years was also captured in the database used to back-test the efficacy of each model.

More specifically, data from 4,798 companies was downloaded from SNL for years spanning 1996 to 2013, but not all companies have data for all years as companies come and go over time. The data for all these companies was converted into 59,890 individual Company Files (i.e., CSV files by company by year), with each file containing Schedule P data triangles for all LOBs.[14] Processing all of this raw data to arrive at the data used for back-testing included several steps.

1. **Data Quality Tests** – In this step, each Company File was checked to determine which LOBs have complete data triangles for years spanning 1996 to 2004. For all key triangles, data quality tests include, but is not limited to, making sure there is non-zero data for each year and minimum data requirements of all models being tested are satisfied. Of the original 4,798 companies, only 2,716 had at least one LOB that passed this test in at least one of the years. For these 2,716 companies there were 79,573 "Data Quality" triangle sets, with the totals by LOB shown in Table 3.1.

2. **Data Validation Tests** – For each of the Data Quality triangle sets, additional tests were conducted to check the next 9 years to make sure none of the data in the original triangles changed over the next 9 years (i.e., to make sure pooling arrangements or other issues don't exist which would cause data to be invalid for testing purposes). The validation process reduced the total company count to 1,679 and for these remaining companies there were 30,707 "Valid Data" triangle sets, with the totals by LOB shown in Table 3.1.

3. **Create Complete Data** – For each of the Valid Data triangle sets, the data for the next 9 years was added to a new data file to speed up testing. Of course during simulation testing only the original triangles were used to parameterize the models, but having the

[13] The U.S. Annual Statement includes 22 lines of business in Schedule P, but there are only 16 lines of business containing 10 accident years of data. The remaining short tail lines are excluded from the research.

[14] If all 4,798 companies had data in all years there would be 86,364 (= 4,798 x 18) files, so there was no data at all about 30% of the time.

actual outcome speeds up the testing process.

4. **Save Diagnostics** – For each of the Valid Data triangle sets, the "optimal" hetero groups were found and diagnostics for all models were calculated and saved. These diagnostic tests were saved so that back-testing can include tests to determine the effectiveness of different diagnostics on assessing model parameters.[15]

Table 3.1. Summary of Datasets by LOB

Schedule P Line of Business	Quality	Valid	Ratio
Commercial Auto Liability	9,555	3,821	40.0%
Commercial Multi-Peril	9,955	4,130	41.5%
Homeowners & Farmowners	10,880	4,724	43.4%
International	317	123	38.8%
Medical Professional Liability - Claims Made	1,878	563	30.0%
Medical Professional Liability - Occurrence	1,465	481	32.8%
Other Liability - Claims Made	4,091	1,482	36.2%
Other Liability - Occurrence	10,923	4,160	38.1%
Products Liability - Claims Made	761	199	26.1%
Products Liability - Occurrence	3,996	1,220	30.5%
Private Passenger Auto Liability	10,075	3,962	39.3%
Reinsurance - Non-Proportional Assumed Financial	397	163	41.1%
Reinsurance - Non-Proportional Assumed Liability	1,758	611	34.8%
Reinsurance - Non-Proportional Assumed Property	2,123	989	46.6%
Special Lines	3,871	1,349	34.8%
Workers' Compensation	7,528	2,730	36.3%
Total All Lines	**79,573**	**30,707**	**38.6%**

For the 1,679 companies with at least one Valid Data triangle set, 1,182 of these companies had at least 2 LOBs with Valid Data for at least one year. For each company (and year) with 2 or more LOBs, the correlation between the residuals was also calculated and saved, both before and after the hetero group factor adjustments, for both paid and incurred data. This resulted in 195,228 pairs of LOBs with correlation values that were captured along with the P-Values and the Degrees of Freedom for all pairs for each company and year set of LOBs. A high level comparison of the data used in this research compared to prior research is shown in Table 3.2.

Table 3.2. Summary of Data by Author

Item	Meyers & Shi	Gremillet & Miehe	Leong, Wang & Chen	Shapland
Evaluation Periods	1	5	11	9
Models Tested	2	3	2	8
Lines of Business	1	4	9	16
Triangle Sets	50	296	~4,850	30,707

[15] Only limited back-testing related to the diagnostics has been completed to date. Future research will provide for more insights on the value of different diagnostic tests.

4. TESTING METHODOLOGY

Using each of the Valid Data triangle sets, the back-testing process starts by calculating the parameters for the six different ODP Bootstrap models[16] described in the Shapland [17] monograph and the Mack Bootstrap model as described in England & Verrall [5]. For all models, the residuals are based on the all year volume weighted average loss development factors, no tail factors were included, and no adjustments to the standard models were included. Because of the sheer volume of the test data, other than assumptions based on diagnostic tests it is nearly impossible to create assumptions tailored to the data in each data set. However, it is possible to use broad sets of assumptions that should be representative of what an analyst might select in practice in order to test how different broad sets of assumptions affects the results.[17]

For the Bornhuetter-Ferguson ODP Bootstrap models, the a priori loss ratios were based on the most recent ultimate loss ratios by year from Schedule P. While this does allow these models to benefit a bit from hindsight, one of the goals for these models was to remove as much of the cyclical bias as possible to see if this improved the accuracy of the models. As a counter to the foresight in the a priori loss ratios, the standard deviations were all set to zero for the preliminary tests.

For the Cape Cod ODP Bootstrap models, it is not possible to include rate level adjustment factors and trend factors based on the data are problematic without the ability to judgmentally review each factor or to set narrow ranges for the trend factors. Thus, all rate level factors were set to 1.0 and all trend factors were set to 2.5% per year.[18] For all tests a decay ratio of 90% was used and each accident year is given 100% weight so nothing is excluded. These assumptions for the Cape Cod models are not intended to be ideal in practice, but rather a reasonable baseline for which other broad sets of assumptions can be compared in future testing.

For the ODP Bootstrap family of models weighted results were also tested. For the weights by accident year, for the 7 oldest accident years the paid and incurred chain ladder

[16] As a technical note, the ODP Bootstrap modeling framework tested during all of the research described in Section 2 is from the original England & Verrall [3] paper that does not include various model enhancements introduced in subsequent papers. In addition, the incurred ODP Bootstrap tested in Leong, Wang & Chen [11] is essentially the paid ODP Bootstrap from England & Verrall [3] using incurred data and does not include the incurred to total unpaid steps described in Section 3.3.1 of Shapland [17].

[17] Only limited back-testing related to the broad sets of assumptions has been completed to date. Future research will provide for more insights on the value of different broad sets of assumptions.

[18] In other words, these assumptions assume there were no rate changes over the 10 years of history and all loss cost inflation is constant at 2.5%.

models were given equal weight. For the 3rd prior year, the paid and incurred chain ladder and Bornhuetter-Ferguson models were given equal weight. For the most recent 2 years, the paid and incurred Bornhuetter-Ferguson and Cape Cod models were given equal weight. While different weighting schemes by LOB would typically be used in practice, this weighting scheme was selected as being representative of a typical weighting scheme.

As a side note, it is also possible to test Aggregate results for each company with at least 2 LOBs of Valid Data, but the results from many different combinations of LOBs would not provide meaningful results without also segregating into groups with all the same LOBs. Instead of just testing the most recent accident year, i.e., only *R(10)* from Graph 1.1, the simulation output of these model tests was captured in great detail, i.e., by accident year, calendar year, calendar year runoff, loss ratios, and each incremental cell in Graph 1.1. Using all of the 10,000 iterations of simulated data, the final step is to compare the actual outcomes to the complete simulated distribution of possible outcomes to determine the percentile of actual outcome for each cell and combination of cells in Graph 1.1.

The companion files for the Shapland [17] monograph could be used to run all of the simulation tests, but those files are designed for educational purposes and not speed.[19] By way of comparison, the Excel model for just one ODP Bootstrap model takes about 15 minutes to run 10,000 iterations so even after completely automating the process it would take one computer over 7 years of continuous processing to finish all of the testing for all 8 models – i.e., the 6 ODP Bootstrap models, the weighted ODP Bootstrap and the Mack Bootstrap with paid data only.

In order to speed up this process commercial software was used, which reduced the total time for one computer from over 7 years to less than 43 days, much faster but still a long process. To reduce the elapsed time even further, the simulation tests were spread over 16 computers, which allowed the overall process to be effectively managed and cut the elapsed time to less than a week.[20]

The simulation back-tests with all of the standard assumptions noted above were considered the "Baseline" tests. Reviewing the baseline tests we found a significant number of simulations with extremely wide distributions. These extreme distributions are a

[19] The companion Excel files can be used to run each of the 6 ODP Bootstrap models and the weighted results but they do not include the Mack Bootstrap model. However, a similar Excel file could be created for the Mack Bootstrap model.

[20] In theory the total elapsed time is less than a week, but in actuality stopping each computer periodically to save results in case of a crash, freeze or other operating system issue and retesting after a data quality review of the output extended the total time to about 2-3 weeks.

somewhat common occurrence in practice and typically result from "small" sample values in the first column that lead to extreme 12-24 month ATA factors (both positive and negative), which in turn lead to some extreme iterations (i.e., a more extreme version of the chain ladder weakness noted above). To address these extreme distributions, a second round of testing included adding constraints to limit the sample outcomes to zero (i.e., to remove negative incremental values) for selected triangle sets (referred to as the "Baseline with Limits" tests).[21] The process used to select triangle sets for adding this limit constraint were based on whether the width of the distributions exceeded a threshold to approximate when an actuary might use these constraints in practice, rather than simply adding this constraint to all triangle sets.

A third round of back-testing was done using all of the "Baseline with Limits" assumptions plus for all of the ODP Bootstrap models the optimal hetero group factors were applied to the modeling framework to test the impact of this common modeling option. This third set of tests are referred to as "Baseline Limits & Hetero".

5. TESTING RESULTS

Starting with the "Baseline" tests, the results for the ODP Bootstrap paid chain ladder for the current accident year (i.e., *R(10)* in Graph 1.1), for all lines of business, and all evaluation periods combined[22] are illustrated in Graph 5.1.

From Graph 5.1 it is clear that the results using significantly more data are still consistent with prior research. Two additional elements of this, and later, graphs are the red "bars" in the lowest and highest decile groups and the average percentile. The red "bars" represent the portion of their respective groups that exceeded the smallest or largest simulated possible outcome, respectively. For example, for the 10% bar the red portion represents the number of tests where the percentile for the actual outcome was less than 0% (i.e., less than the smallest simulated possible outcome). The average percentile is the average over all samples[23] and helps give a sense of how close the simulated means were to the "true" mean on average.

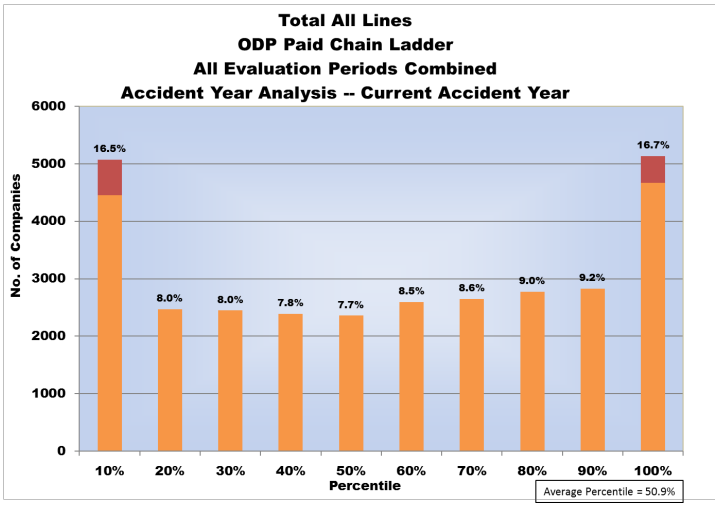
---

[21] This constraint on the simulation process is the third option described in section 4.1.1 of Shapland [17].

[22] For each evaluation period (e.g., 1996) the current accident year is always as of 12 months of development. Thus, while there are multiple evaluation dates the results for the current accident year for each evaluation date can be combined.

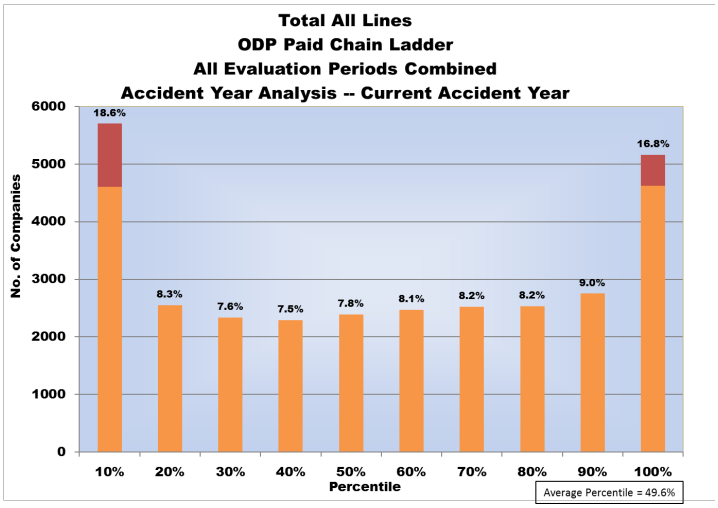
[23] For the samples below the minimum or above the maximum (i.e., represented by the red bars) the value used in the overall average percentile is 0% or 100%, respectively.

Graph 5.1. ODP Bootstrap Paid Chain Ladder – "Baseline"

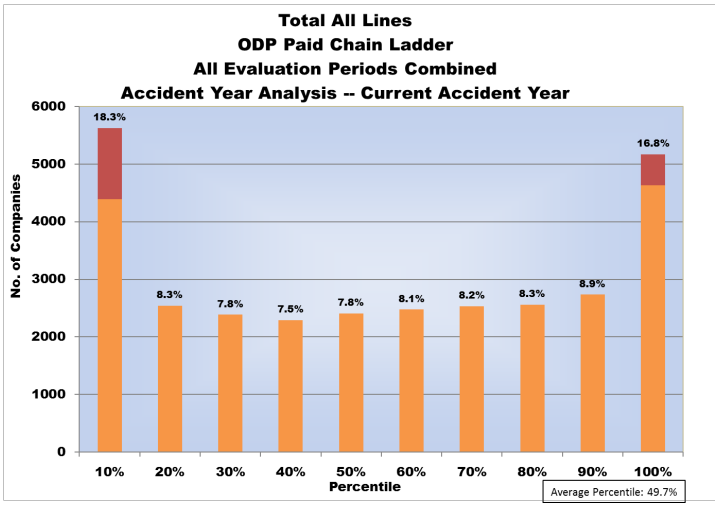


Moving to the "Baseline with Limits" tests the results for the ODP Bootstrap paid chain ladder for the current accident year, for all lines of business, and all evaluation periods combined are illustrated in Graph 5.2. Comparing Graph 5.2 with Graph 5.1 it makes sense that the "goal posts" at the extremes got higher, meaning the models further underestimated the "true" distributions, since the widest of the distributions in the "Baseline" tests were "narrowed" in the "Baseline with Limits" testing.

Graph 5.2. ODP Bootstrap Paid Chain Ladder – "Baseline with Limits"

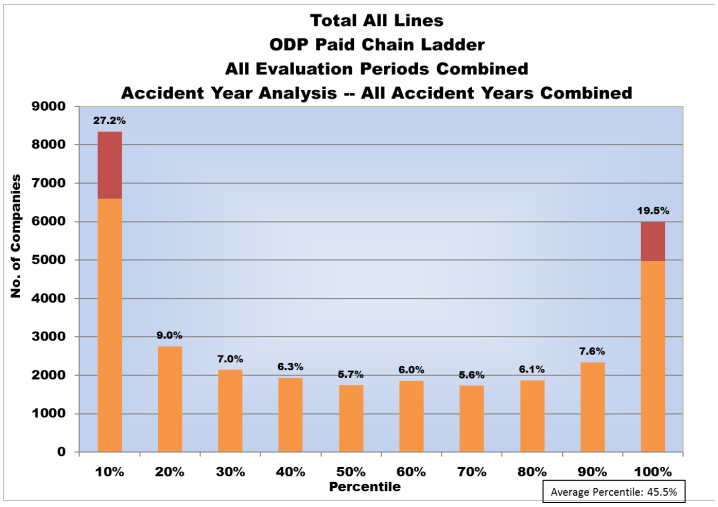


Graph 5.3. ODP Bootstrap Paid Chain Ladder – "Baseline Limits & Hetero"



At a high level the "Baseline Limits & Hetero" results for the ODP Bootstrap paid chain ladder for the current accident year, for all lines of business, and all evaluation periods combined are illustrated in Graph 5.3. The differences between Graph 5.3 and 5.2 are more subtle but a close inspection shows a slight improvement, which supports the use of heteroscedasticity adjustment factors in the ODP Bootstrap models. Admittedly, this support for using hetero factors is not strong but it is an improvement and rules out a negative conclusion (i.e., that hetero factors don't help). All of the results in the remainder of this paper are for the "Baseline Limits & Hetero" testing, but for simplicity this label is not included in any more graphs.

Graph 5.4. ODP Bootstrap Paid Chain Ladder – All Years Combined

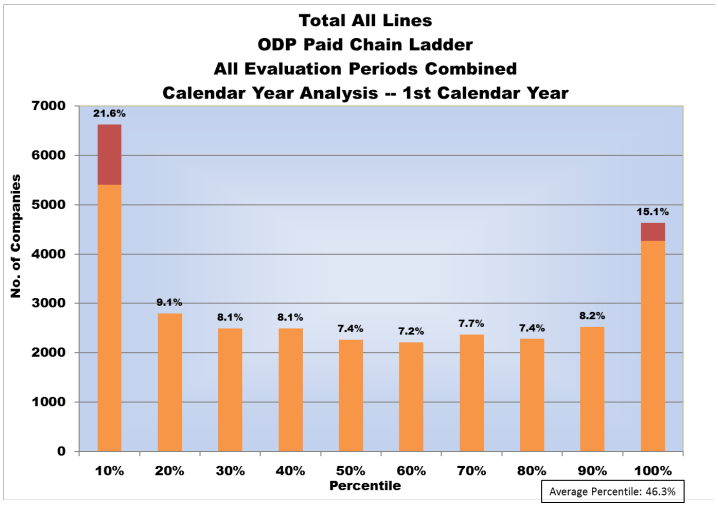


As we dig deeper into the back-testing results, a logical first dive would be to review results for prior accident years (i.e., *R(9)* to *R(2)* in Graph 1.1) to see if the estimation

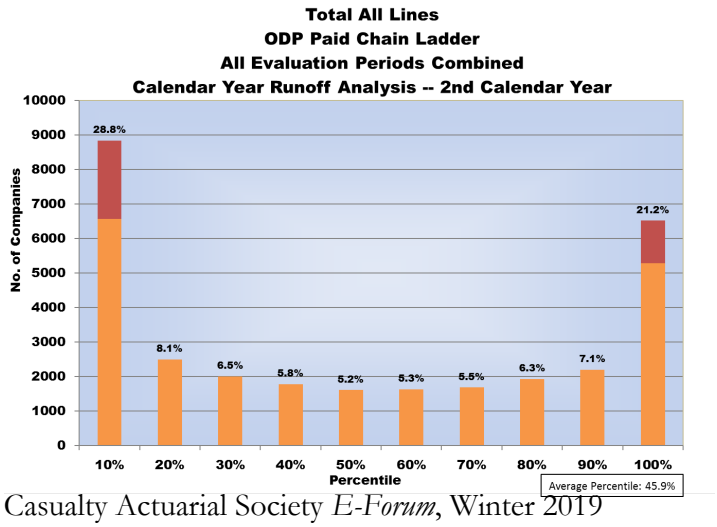
improves as the relative maturity of the accident year increases. The results by accident year are shown in Appendix A, but interestingly there is no improvement as the models predict fewer future periods. Similarly, combining all accident years (i.e., *R(T)* in Graph 1.1), as shown in Graph 5.4, does not improve the model predictions.

One of the insights from the Leong, Wang & Chen [11] paper was how the results were impacted by the reserving cycle. This impact was confirmed using this expanded database with the results by evaluation year shown in Appendix B. Consistent with the Leong, Wang & Chen results, the results by year show that the size of the "goal post" is predominantly in the lowest decile when the mean is being underestimated (e.g., in 1996) and shifts to being predominantly in the highest decile as the mean is overestimated. In addition, the average percentile shifts over the reserving cycle, which indicates how the estimates of the "true" mean change during the cycle.

Graph 5.5. ODP Bootstrap Paid Chain Ladder – First Calendar Year



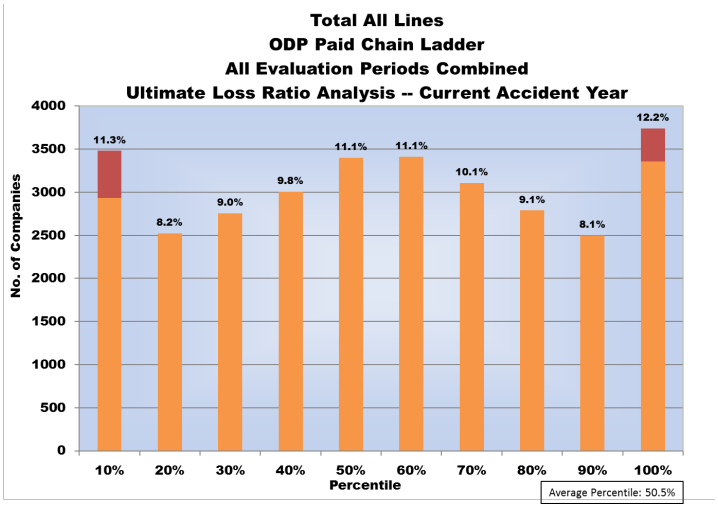
Graph 5.6. ODP Bootstrap Paid Chain Ladder – Calendar Year Runoff After 1 Year



In addition to looking at the predictions for the accident years, it was also possible to look at the calendar years (i.e., the sum of the diagonals in Graph 1.1), the calendar year runoff (i.e., the sum of all remaining diagonals as each diagonal is removed in Graph 1.1), and the time zero to ultimate loss ratios (i.e., the sum of an entire row in Graph 1.1). As might be expected after reviewing the accident year results, the calendar year results in Graph 5.5 and calendar year runoff results in Graph 5.6 are quite similar to the accident year results.[24]

For the time zero to ultimate loss ratio estimates by accident year shown in Graph 5.7 the predictions are much closer to the ideal histogram in Graph 1.3. This is an interesting result in the sense that the ODP Bootstrap predictions of the time zero to ultimate loss ratios appear to be more accurate than the predictions of the unpaid claims. To understand this we need to dig deeper into the results by incremental cell, which are shown in Appendix C. Interestingly, the results for the incremental cells reveals that the sampling of the incremental cells to create sample triangles for each iteration seems to produce more variability than observed in the data. On the other hand, since the model parameters are fit to the actual outcomes in the triangle perhaps seeing considerably more results in the middle decile groups is the expected result.

Graph 5.7. ODP Bootstrap Paid Chain Ladder – Ultimate Loss Ratio Current Year

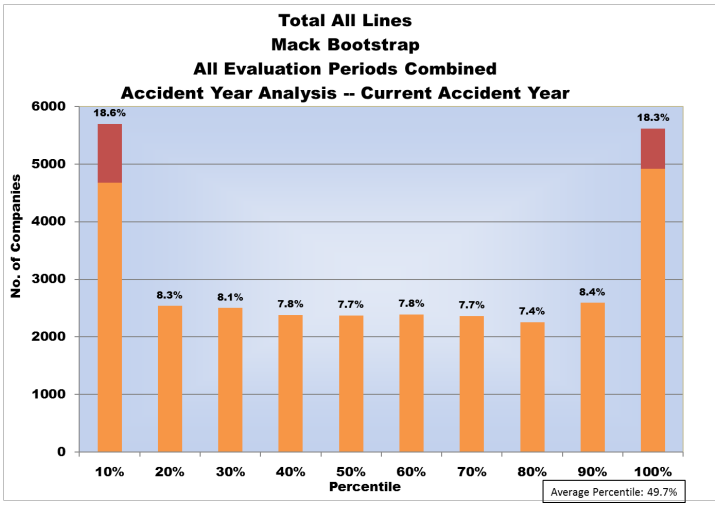


Now that we have dissected the ODP Bootstrap paid chain ladder model, we can compare this to the other models in the back-testing research. First, the results for the Mack Bootstrap paid chain ladder model are shown in Graph 5.8. Comparing Graph 5.8 with

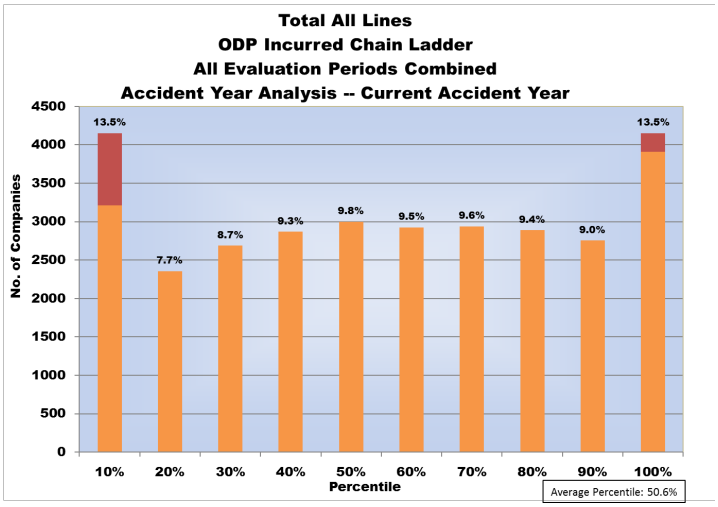
[24] Since Graph 5.5 is the first diagonal and Graph 5.6 is the sum of the remaining diagonals, the combination of these two graphs is the same as Graph 5.4.

Graph 5.3 shows that the Mack Bootstrap was a worse than the ODP Bootstrap, which is consistent with the findings of the GIRO Working Parties [15, 16].

Graph 5.8. Mack Bootstrap Paid Chain Ladder – Current Accident Year



Graph 5.9. ODP Bootstrap Incurred Chain Ladder – Current Accident Year



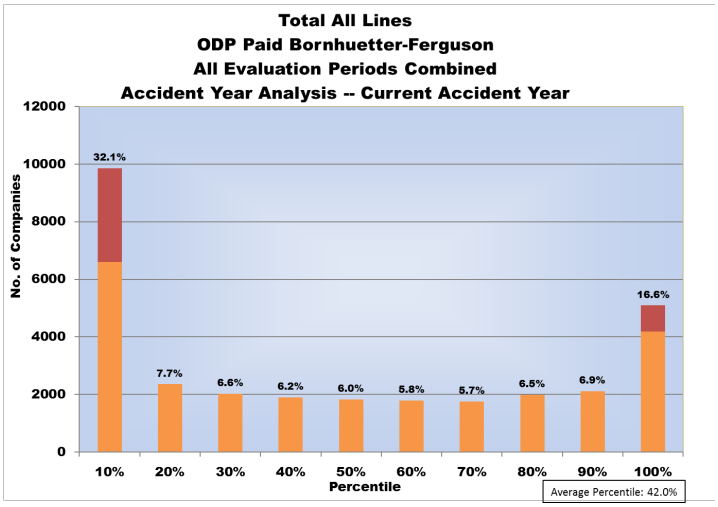
Next, the results for the ODP Bootstrap Incurred Chain Ladder model are shown in Graph 5.9. Comparing Graph 5.9 to Graph 5.3 there is a clear improvement in the predictive power of the incurred versus paid chain ladder versions of the ODP Bootstrap model.[25] Thinking about the mechanics of the ODP Bootstrap Incurred Chain Ladder model in Shapland [17] it seems fair to conclude that combining the variability of the paid and incurred data increases the relative variance of the unpaid estimates to come much closer to the ideal histogram in Graph 1.3. It is quite possible that the remaining "goal post" effect is

[25] This is inconsistent with the findings in Leong, Wang & Chen [11]. However, as mentioned in footnote 11, the algorithm being tested in this research is different.

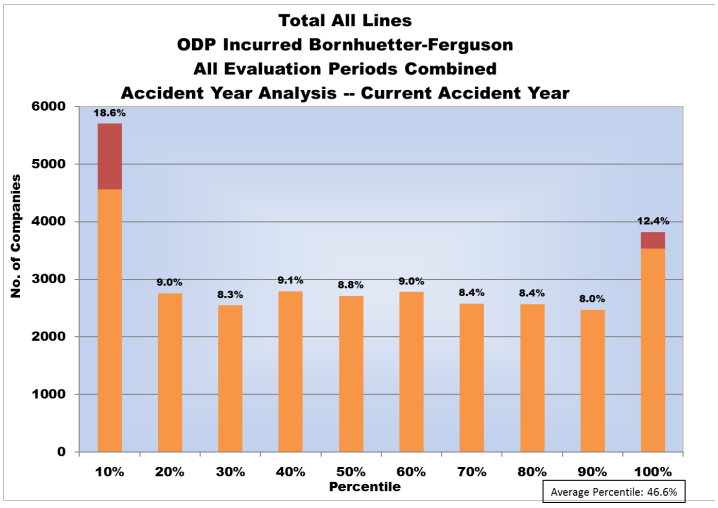
largely due to the mis-estimation of the mean during the reserving cycle.

Next, the results for the ODP Bootstrap Paid and Incurred Bornhuetter-Ferguson models are shown in Graph 5.10 and Graph 5.11, respectively. Comparing Graph 5.10 with Graph 5.3 and Graph 5.11 with Graph 5.9, respectively, it appears as though the Bornhuetter-Ferguson models are less predictive than their chain ladder counterparts are. This is inconclusive, however, since the variance assumption was set to zero during the current back-testing and it is easy to show that using a zero variance assumption will reduce the variability of the estimated unpaid claim distribution. Thus, conclusions about the predictive power of the ODP Bootstrap Bornhuetter-Ferguson models will need to wait until more testing can be completed.

Graph 5.10. ODP Bootstrap Paid Bornhuetter-Ferguson – Current Accident Year



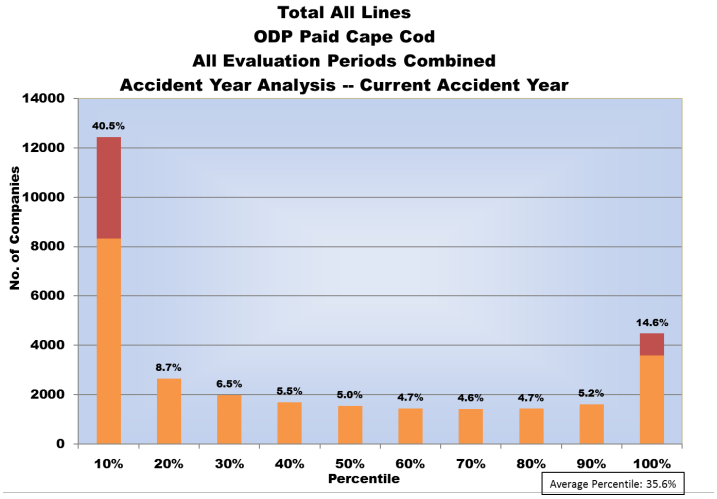
Graph 5.11. ODP Bootstrap Incurred Bornhuetter-Ferguson – Current Accident Year



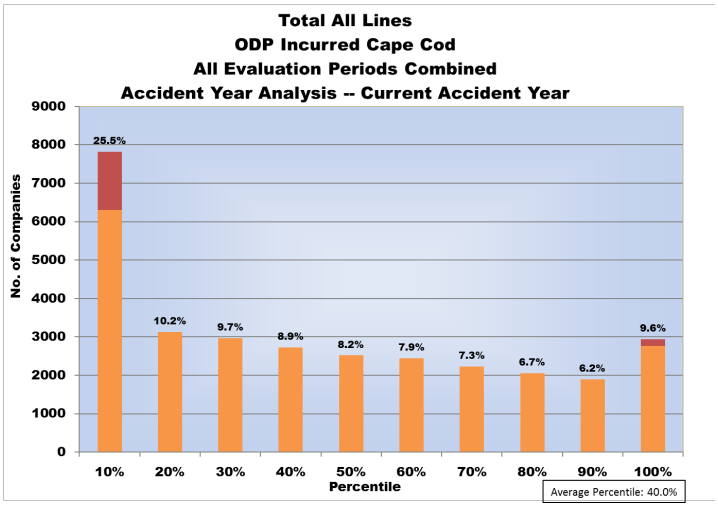
Next, the results for the ODP Bootstrap Paid and Incurred Cape Cod models are shown

in Graph 5.12 and Graph 5.13, respectively. Comparing Graph 5.12 with Graph 5.3 and Graph 5.13 with Graph 5.9, respectively, it appears as though the Cape Cod models are less predictive than their chain ladder counterparts are. This may also be inconclusive, however, since only one set of parameters has been tested so far. Thus, conclusions about the predictive power of the ODP Bootstrap Cape Cod models will need to wait until more testing can be completed.

Graph 5.12. ODP Bootstrap Paid Cape Cod – Current Accident Year



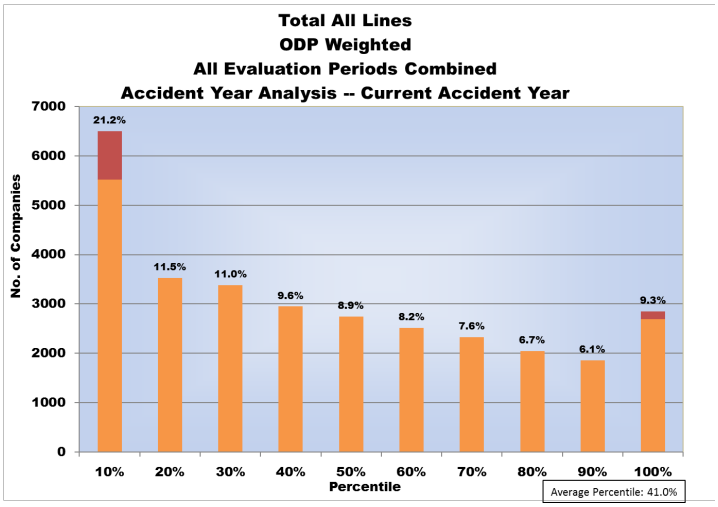
Graph 5.13. ODP Bootstrap Incurred Cape Cod – Current Accident Year



Finally, the results for the weighted combination of all six ODP Bootstrap models are shown in Graph 5.14. Comparing Graph 5.14 with Graphs 5.3, 5.9, 5.10, 5.11, 5.12, and 5.13, you can visualize how Graph 5.14 results from a combination of the other models. This seems promising as even with the deficiencies noted for each model individually the weighted results look like they are better that the sum of the parts. More importantly, this

demonstrates how weighting multiple models, to at least partially address model risk, can improve the results compared to a single model. As other assumptions for the Bornhuetter-Ferguson and Cape Cod models are tested, another avenue for future research will be considerations on how to apply Bayesian analysis to selecting the model weights.

Graph 5.14. ODP Bootstrap Weighted Models – Current Accident Year



All of the results presented in this Section and Appendices A, B, and C are for all lines of business combined. To show that the results are similar by line of business, the results by line of business for the ODP Bootstrap Paid Chain Ladder and Incurred Chain Ladder models are in Appendix D. It is possible to show many more details and combinations for all of these results, but this massive increase will be accompanied by an increase in random noise and will likely add little value beyond what we can already see at the higher level.

6. BENCHMARKS BASED ON TEST RESULTS

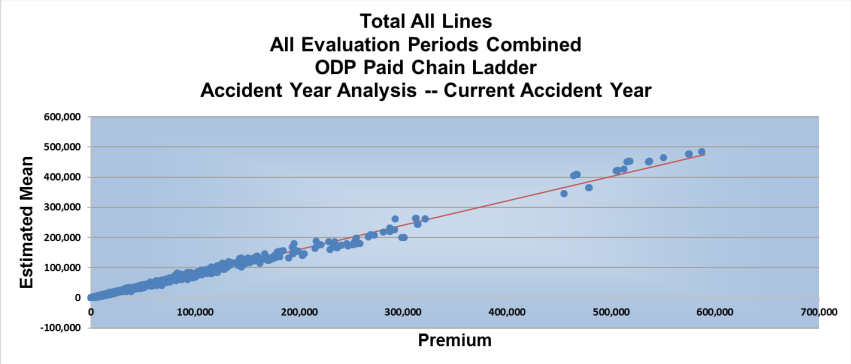
Even with the expansion of the research database, this research has confirmed the findings of prior authors. Thinking about the impact of the reserving cycle, it appears as though the results are strongly influenced by the internal systemic risks of the ODP Bootstrap modeling framework which, like the deterministic chain ladder, leads to the cycle of under and over estimation of the mean and in synch with this a lower and higher estimation of the variance. Even after potential corrections for the internal systemic risks, the ODP Bootstrap model is generally not accounting for the external systemic risks. On the other hand, it appears that some of the variations on the ODP Bootstrap framework may be significantly better at addressing the internal systemic risks.

In order to use this information in practice, one approach might be to consider how the formulas proposed by Leong, Wang & Chen [11] could be improved to separately address internal and external systemic risks. However, even with formula improvements, on a forward-looking basis the actuary is still faced with trying to understand which part of the reserving cycle they are currently in. Of course, knowing were one is in the reserving cycle is an issue no matter what the approach, but with a significantly larger database another way forward is possible.

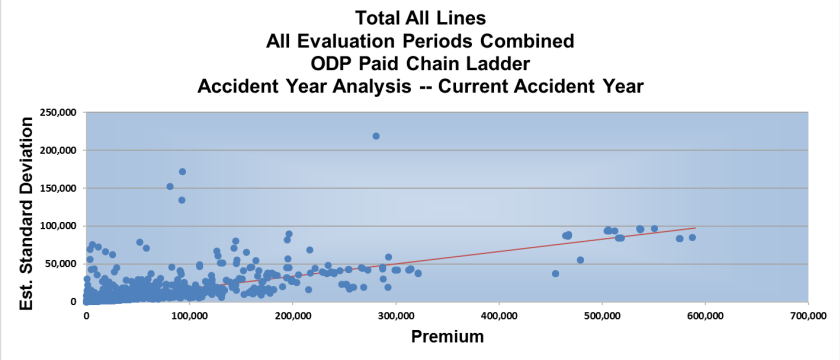
6.1. Unpaid Claim Benchmarks

Rather than try to create a precise formula for giving the "correct" distribution, we can take a page out of the deterministic reserving playbook and create benchmarks to help guide the judgment of the opining actuary. For example, consider Graphs 6.1 and 6.2, which illustrate the range of mean and standard deviation estimates from the ODP Bootstrap paid chain ladder model over the entire database for the most recent accident year. For Graph 6.1, it is not surprising that the mean unpaid is closely in line with the premium, with the deviations along the slope of the trend line representing differences in loss ratio by company.

Graph 6.1. ODP Bootstrap Means – Current Accident Year



Graph 6.2. ODP Bootstrap Standard Deviations – Current Accident Year

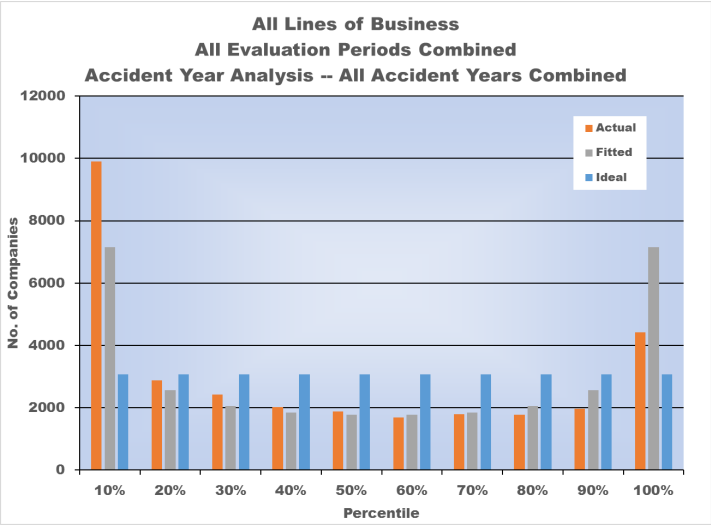


In Graph 6.2, it is not surprising that the standard deviations also increase in line with the premium, but the deviations around the trend line are more pronounced, which is likely due to the mixture of all lines of business, but at least to some degree a few of these could be considered outliers. A more important ingredient of Graph 6.2 is that the slope of the trend line is much lower, which confirms that the Coefficient of Variation is consistent with statistical principles, meaning for smaller companies the standard deviation is a larger percentage of the mean compared to larger companies.

The results shown in Graphs 6.1 and 6.2 are consistent for all other views of the data discussed in Section 5 (i.e., for each accident year, each calendar year, all years combined, etc.). In addition, similar graphs by line of business are also consistent with Graphs 6.1 and 6.2, except that they are more specific to the data for each line of business. This new insight lead to the idea of combining regression results (based on pure premiums instead of premiums) by line of business to create a benchmark algorithm for the means and related standard deviations by accident year, calendar year, etc., which at a minimum reflects the independent risks in the data.

As these regression results are based on the original simulation results, without any further adjustment the benchmarks would also reflect the biases shown in the back-testing results. In order to adjust for this bias an optimal variance correction factor was included similar to the factors proposed by Leong, Wang & Chen [11], except that the factor does not change each year during the reserving cycle. As an example, consider Graph 6.3 for all accident years and all lines of business combined.

Graph 6.3. ODP Bootstrap Bias Adjustment – All Accident Years Combined



For the fitted results in Graph 6.3 the optimal adjustment factor is 1.755, meaning the

benchmark standard deviations would be increased by 75.5%. There are variations in the optimal factor when looking at individual accident years, but they are reasonably consistent so only one factor based on the total of all years combined is used for the unpaid claims benchmarks. As noted for Graph 5.7, the results for the time zero to ultimate loss ratios are much closer to the ideal histogram so a lower adjustment factor is appropriate for the loss ratio benchmarks.

Because of the cyclical bias it is not possible to increase the factor to the point where the ideal histogram is achieved. However, this does seem to address the variance mis-estimation component of the internal systemic risk and external systemic risk to the extent that external systemic risks have influenced the outcomes in this research database. Assuming this is correct, the remaining "goal post" shape of the fitted histogram in Graph 6.3 is due to the mis-estimation of the mean during the reserving cycle that can be addressed by the actuary as part of the selection of the booked reserves.

It is possible that future research could help the actuary further understand the timing of the reserving cycles, but assuming the actuary can use caution to ensure their modeling assumptions are not being biased by the reserving cycle, the unpaid claim distribution benchmarks can be used as a guide to assess an estimated distribution from any stochastic model. For example, consider the results in Table 6.1 which compare standard results for the ODP Bootstrap paid and incurred chain ladder models with the corresponding benchmarks using commercial auto data from a randomly selected company in the research database.

Table 6.1. Comparison of ODP Bootstrap with Benchmark Unpaid

			ODP Bootstrap Paid Chain Ladder			ODP Bootstrap Incurred Chain Ladder			Unpaid Claim Benchmark		
Accident Year	Earned Premium	A priori Loss Ratio	Mean	Standard Error	CoV	Mean	Standard Error	CoV	Mean	Standard Error	CoV
2008	83,943	55.0%	125	194	154.9 %	135	216	160.7 %			
2009	94,343	55.0%	225	267	118.5 %	234	293	125.3 %	669	1,325	198.1 %
2010	115,098	55.0%	568	453	79.8 %	593	503	84.9 %	1,184	1,540	130.0 %
2011	126,714	55.0%	975	639	65.5 %	1,010	717	71.0 %	1,960	2,055	104.8 %
2012	138,148	55.0%	2,564	978	38.1 %	2,618	1,206	46.1 %	3,632	2,689	74.0 %
2013	156,046	55.0%	6,222	1,648	26.5 %	6,404	2,455	38.3 %	7,301	4,475	61.3 %
2014	173,621	55.0%	13,146	2,529	19.2 %	14,781	4,841	32.7 %	15,027	7,609	50.6 %
2015	181,416	55.0%	27,524	3,888	14.1 %	32,868	9,345	28.4 %	28,179	11,947	42.4 %
2016	184,422	55.0%	45,759	5,518	12.1 %	49,668	15,204	30.6 %	48,125	18,504	38.4 %
2017	186,444	55.0%	66,947	9,017	13.5 %	80,709	24,784	30.7 %	75,007	27,104	36.1 %
Totals	**1,440,195**		**164,055**	**12,928**	**7.9 %**	**189,019**	**31,310**	**16.6 %**	**181,085**	**38,898**	**21.5 %**

The benchmark algorithm is based on 10 years of data so the earned premium and a priori loss ratios are used to enter pure premiums by year into the algorithm. The ODP Bootstrap results in Table 6.1 do include using the optimal hetero groups and a few other model options to replicate what an actuary could easily produce as a first draft of the unpaid claim distribution. Not surprisingly, the benchmark results indicate that the CoV should be higher compared to either of the ODP Bootstrap models.

Reviewing Table 6.1, three more observations can be made. First, the benchmark does not include the 9th prior accident year (i.e., 2008). This is due to tail factors being excluded from the back-testing to date, but future back-testing can include tail factors which will allow the benchmark algorithm to be expanded to include the 9th prior accident year (and an additional calendar year). Second, even though the data used in the back-testing was from 1996-2004, the algorithm is independent of the year and based on fitting to distributions by size of exposure (e.g., pure premiums) so using the algorithm to create a benchmark for 2017 makes sense as long as the a priori loss ratios are reasonable given the reserving cycle.

The third observation from Table 6.1 is that the benchmarks are essentially based on the average loss development pattern from the industry data. Thus, it would be a reasonable critique to note that the loss development pattern for the company under review does not clearly match with a Schedule P line of business. Because this is such a common issue, the algorithm also includes an option to adjust benchmarks based on the loss development pattern assumption used by the actuary. In Table 6.2 the benchmark has been updated for the loss development pattern for the data being used in the example. Comparing Table 6.2 with Table 6.1, note that while the mean and standard deviations both decreased a bit the CoV essentially stayed the same or even increased a bit, which makes sense given the increased uncertainty. Also keep in mind that all benchmarks are only intended to serve as a guideline for the actuary and a perfect match is not a goal.

Table 6.2. Comparison of ODP Bootstrap with Benchmark Unpaid & Custom LDF Pattern

			ODP Bootstrap Paid Chain Ladder			ODP Bootstrap Incurred Chain Ladder			Unpaid Claim Benchmark		
Accident Year	Earned Premium	A priori Loss Ratio	Mean	Standard Error	CoV	Mean	Standard Error	CoV	Mean	Standard Error	CoV
2008	83,943	55.0%	125	194	154.9 %	135	216	160.7 %			
2009	94,343	55.0%	225	267	118.5 %	234	293	125.3 %	93	211	225.4 %
2010	115,098	55.0%	568	453	79.8 %	593	503	84.9 %	260	384	148.0 %
2011	126,714	55.0%	975	639	65.5 %	1,010	717	71.0 %	641	745	116.1 %
2012	138,148	55.0%	2,564	978	38.1 %	2,618	1,206	46.1 %	1,778	1,404	79.0 %
2013	156,046	55.0%	6,222	1,648	26.5 %	6,404	2,455	38.3 %	4,643	2,942	63.4 %
2014	173,621	55.0%	13,146	2,529	19.2 %	14,781	4,841	32.7 %	11,306	5,809	51.4 %
2015	181,416	55.0%	27,524	3,888	14.1 %	32,868	9,345	28.4 %	24,133	10,287	42.6 %
2016	184,422	55.0%	45,759	5,518	12.1 %	49,668	15,204	30.6 %	44,007	16,974	38.6 %
2017	186,444	55.0%	66,947	9,017	13.5 %	80,709	24,784	30.7 %	72,694	26,296	36.2 %
Totals	**1,440,195**		**164,055**	**12,928**	**7.9 %**	**189,019**	**31,310**	**16.6 %**	**159,555**	**34,460**	**21.6 %**

In order to illustrate how the benchmark algorithm responds to different input assumptions, Table 6.3 includes a comparison of the benchmarks from Table 6.2 with benchmarks based on only 10% of the original premiums (i.e., all other assumptions are the same). This shows how the benchmarks for a smaller company would compare to those for a larger company. Following statistical principles, and the regressions illustrated in Graph 6.1 and 6.2, with only 10% of the premium the mean is reduced by 90% but CoV increases to reflect the additional uncertainty.

Table 6.3. Comparison of Benchmarks by Size of Company

			Unpaid Claim Benchmarks						Unpaid Claim Benchmarks		
Accident Year	Earned Premium	A priori Loss Ratio	Mean	Standard Error	CoV	Accident Year	Earned Premium	A priori Loss Ratio	Mean	Standard Error	CoV
2008	83,943	55.0%				2008	8,394	55.0%			
2009	94,343	55.0%	95	214	224.8 %	2009	9,434	55.0%	10	47	495.6 %
2010	115,098	55.0%	262	387	147.8 %	2010	11,510	55.0%	26	91	346.5 %
2011	126,714	55.0%	616	719	116.8 %	2011	12,671	55.0%	62	166	269.1 %
2012	138,148	55.0%	1,735	1,374	79.2 %	2012	13,815	55.0%	173	289	166.8 %
2013	156,046	55.0%	4,525	2,874	63.5 %	2013	15,605	55.0%	452	523	115.6 %
2014	173,621	55.0%	11,154	5,736	51.4 %	2014	17,362	55.0%	1,115	877	78.6 %
2015	181,416	55.0%	23,905	10,194	42.6 %	2015	18,142	55.0%	2,390	1,369	57.3 %
2016	184,422	55.0%	43,759	16,882	38.6 %	2016	18,442	55.0%	4,376	2,253	51.5 %
2017	186,444	55.0%	72,465	26,216	36.2 %	2017	18,644	55.0%	7,246	3,436	47.4 %
Totals	**1,440,195**		**158,515**	**34,245**	**21.6 %**		**144,020**		**15,851**	**4,835**	**30.5 %**

As noted earlier, the benchmark algorithm includes more than the accident year unpaid claims, so Table 6.4 illustrates the cash flow and unpaid claim runoff benchmarks which would be comparable to the unpaid claim benchmarks in Table 6.2. The benchmark algorithm also includes time zero to ultimate loss ratios, but these are not illustrated in any of the Tables.

Table 6.4. Comparison of Unpaid, Cash Flow and Runoff Benchmarks

	Unpaid Claim Benchmarks				Cash Flow Benchmarks				Unpaid Claim Runoff Benchmarks		
Accident Year	Mean	Standard Error	CoV	Calendar Year	Mean	Standard Error	CoV	Calendar Year	Mean	Standard Error	CoV
2008								2017	158,515	34,245	21.6 %
2009	95	214	224.8 %	2018	61,896	16,101	26.0 %	2018	96,619	23,882	24.7 %
2010	262	387	147.8 %	2019	40,956	12,296	30.0 %	2019	55,663	16,877	30.3 %
2011	616	719	116.8 %	2020	24,529	9,031	36.8 %	2020	31,133	12,234	39.3 %
2012	1,735	1,374	79.2 %	2021	13,581	6,286	46.3 %	2021	17,552	8,841	50.4 %
2013	4,525	2,874	63.5 %	2022	7,252	4,308	59.4 %	2022	10,301	6,585	63.9 %
2014	11,154	5,736	51.4 %	2023	4,067	3,349	82.4 %	2023	6,234	5,381	86.3 %
2015	23,905	10,194	42.6 %	2024	2,437	2,827	116.0 %	2024	3,797	4,346	114.5 %
2016	43,759	16,882	38.6 %	2025	1,698	2,268	133.6 %	2025	2,099	4,164	198.3 %
2017	72,465	26,216	36.2 %	2026	2,099	4,164	198.3 %				
Totals	**158,515**	**34,245**	**21.6 %**		**158,515**	**34,245**	**21.6 %**				

6.2. Correlation Benchmarks

As noted at the end of Section 3, the data from 1,182 of the companies had at least 2 LOBs with Valid Data for at least one year. For each company (and year) with 2 or more LOBs, the correlation between the residuals was also calculated and saved, including the P-Values and the Degrees of Freedom, both before and after the hetero group factor adjustments, for both paid and incurred data. This database of 195,228 pairs of LOBs with correlation values were used to create separate benchmarks of correlation between Schedule P lines of business.

The correlation benchmarks include each year separately and all years combined, but only a sample from 1996 is illustrated in Table 6.5. In addition to calculating the sample average and standard deviations by pair, the number of pairs are also shown.

Table 6.5. Sample Correlation Benchmarks for 1996 – Paid After Hetero Adjustment – Raw Data

Mean Values	MPL-O	HO	WC	CA	PPA
MPL-O	100.0%	-0.5%	-10.9%	2.8%	-1.9%
HO	-0.5%	100.0%	4.0%	5.9%	11.8%
WC	-10.9%	4.0%	100.0%	11.9%	13.9%
CA	2.8%	5.9%	11.9%	100.0%	13.3%
PPA	-1.9%	11.8%	13.9%	13.3%	100.0%

Standard Deviations	MPL-O	HO	WC	CA	PPA
MPL-O	0.0%	11.1%	14.0%	16.1%	16.7%
HO	11.1%	0.0%	20.0%	18.9%	20.8%
WC	14.0%	20.0%	0.0%	23.5%	23.7%
CA	16.1%	18.9%	23.5%	0.0%	24.3%
PPA	16.7%	20.8%	23.7%	24.3%	0.0%

Count of Pairs	MPL-O	HO	WC	CA	PPA
MPL-O	-	57	62	59	48
HO	57	-	618	757	851
WC	62	618	-	688	570
CA	59	757	688	-	784
PPA	48	851	570	784	-

The P-Values are a measure of how significantly different from zero the correlation value is for each calculated pair. The lower the P-Value the more significantly different from zero the correlation. Thus, a second set of correlation benchmarks, using one minus the P-Value as the weights, were calculated for weighted means and weighted standard deviations. For comparison, the weighted benchmarks for the same sample are included in Table 6.6.

Table 6.6. Sample Correlation Benchmarks for 1996 – Paid After Hetero Adjustment – Weighted

Mean Values	MPL-O	HO	WC	CA	PPA
MPL-O	100.0%	0.0%	-16.2%	5.9%	-1.7%
HO	0.0%	100.0%	5.4%	9.5%	16.7%
WC	-16.2%	5.4%	100.0%	17.1%	18.9%
CA	5.9%	9.5%	17.1%	100.0%	19.3%
PPA	-1.7%	16.7%	18.9%	19.3%	100.0%

Standard Deviations	MPL-O	HO	WC	CA	PPA
MPL-O	0.0%	14.0%	14.6%	18.8%	18.6%
HO	14.0%	0.0%	23.6%	22.9%	22.9%
WC	14.6%	23.6%	0.0%	26.6%	26.0%
CA	18.8%	22.9%	26.6%	0.0%	27.1%
PPA	18.6%	22.9%	26.0%	27.1%	0.0%

Count of Pairs	MPL-O	HO	WC	CA	PPA
MPL-O	-	57	62	59	48
HO	57	-	618	757	851
WC	62	618	-	688	570
CA	59	757	688	-	784
PPA	48	851	570	784	-

While it was noted in Section 4 that aggregate simulations were not captured, and thus not available for additional benchmarks, it is quite straightforward to use the correlation benchmarks in conjunction with the unpaid benchmarks to create a customized aggregate unpaid benchmark. Finally, as noted above the Degrees of Freedom was also captured and could have been included as part of Tables 6.5 and 6.6. In practice, this would be a valuable benchmark for copulas used for aggregation as they are intended to strengthen the tail of the aggregate distribution given a selected correlation.

6.3. LDF Pattern Benchmarks

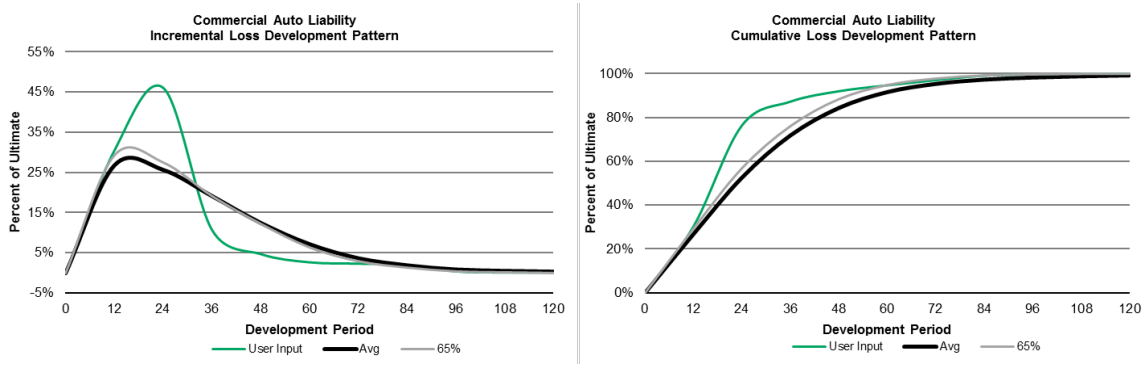
In addition to all of the simulation results, for each dataset the all year volume weighted average loss development pattern from the original paid triangle (actual), along with the implied pattern from the average of all the simulated sample paid triangles (simulated), were captured. Using all of the paid patterns by line of business, the mean and percentiles of these patterns can be used as LDF pattern benchmarks. For example, the development patterns for Commercial Auto sample used in the Tables in Section 6 are included in Table 6.7.

Table 6.7. Sample LDF Pattern Benchmarks – Commercial Auto

Development Periods:	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120	120+
Actual LDF Pattern:	30.4%	76.4%	87.2%	91.9%	94.6%	97.0%	99.0%	99.5%	99.7%	99.9%
Average LDF Pattern:	26.9%	52.6%	71.8%	84.3%	91.5%	95.2%	97.2%	98.1%	98.7%	99.1%
65% LDF Pattern:	29.3%	56.9%	76.1%	88.3%	94.8%	97.7%	99.1%	99.6%	99.8%	99.9%

The actual LDF pattern was calculated using the all year volume weighted average LDF factors from the sample dataset. The average and 65% LDF patterns are the average and 65th percentile from all of the simulated patterns in the database, respectively. By systematic testing and a little trial and error, the 65th percentile was found to be the best fit to the actual pattern. The patterns from Table 6.7 are illustrated in Graph 6.4, since one of the uses of LDF pattern benchmarks could be to help smooth the selection of age-to-age factors.

Graph 6.4. Comparison of Actual with Benchmark LDF Patterns

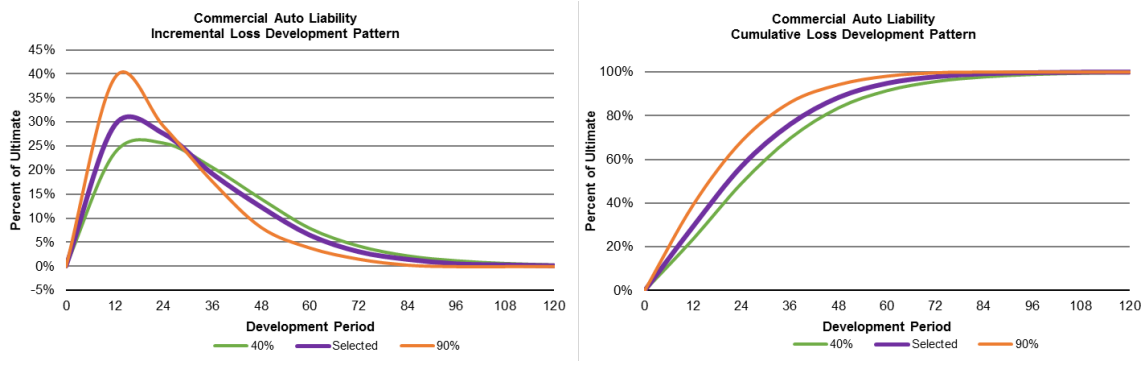


Once the actual LDF pattern has been smoothed, or a suitable percentile pattern has been selected, another use of the LDF pattern benchmarks is to help create a range of deterministic central estimates. For example, assuming the 65th percentile pattern is selected, the actuary could then base a deterministic range on the patterns which are 25 points above and below the 65th percentile as illustrated in Table 6.8 and Graph 6.5.

Table 6.8. Sample LDF Pattern Range – Commercial Auto

Development Periods:	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120	120+
40% LDF Pattern:	23.6%	49.1%	69.6%	83.6%	91.4%	95.5%	97.7%	98.8%	99.3%	99.5%
65% LDF Pattern:	26.9%	52.6%	71.8%	84.3%	91.5%	95.2%	97.2%	98.1%	98.7%	99.1%
90% LDF Pattern:	39.3%	68.5%	86.1%	94.3%	98.2%	99.7%	100.0%	100.0%	100.0%	100.0%

Graph 6.5. Range of Benchmark LDF Patterns



7. CONCLUSIONS

Using an extensive database pulled from historical Schedule P data, the results from back-testing various ODP Bootstrap models and the Mack Bootstrap model has confirmed similar prior research on how effective these models predict the distribution of possible outcomes. For the versions of the ODP Bootstrap model not previously tested, the back-testing results are both encouraging and inconclusive. In particular, for the ODP Bootstrap incurred chain ladder model, as described in Shapland [17], using both the paid and incurred data significantly improves the results. For the ODP Bootstrap Bornhuetter-Ferguson and Cape Cod models the results were inconclusive due to the need to test more model parameters. However, even with inconclusive results for four of the six ODP Bootstrap models, testing of weighted results demonstrated that weighing multiple models, to at least partially address model risk, is a significant improvement over using a single model.

Due to the size of the database used in the back-testing, the data allows us to use benchmarking algorithms as a guide when evaluating the estimated distribution of possible outcomes from any stochastic model. These benchmarking algorithms are quite sophisticated in the sense that they address the statistical properties of real data sets (e.g., more relative variance for smaller exposures) and can be customized to more closely approximate the data being analyzed (e.g., using selected ATA factors). Additional uses from the data include correlation benchmarks and LDF pattern benchmarks.

Acknowledgment

The author gratefully acknowledges the many authors listed in the References (and others not listed) that contributed to the foundation of the ODP bootstrap and Mack models, without which this research would not have been possible. He would like to thank all the peer reviewers, and is grateful to the CAS referees, in particular Lynne Bloom and Jiao Ziyi, for their comments that greatly improved the quality of the paper.

Supplementary Material

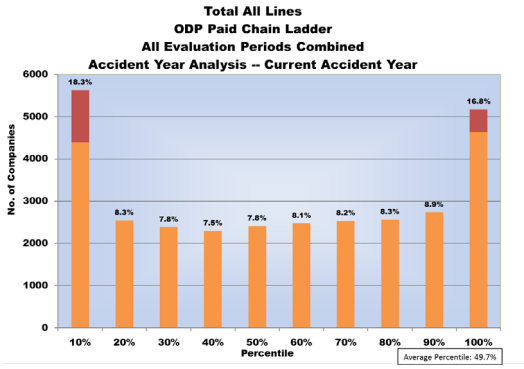
Given the vast size of the database used in this research and the proprietary nature of the results, no supplementary materials can be provided. However, the interested reader can contact the author to learn more about the proprietary benchmarks.

Appendix A – Back-Testing Results by Accident Year

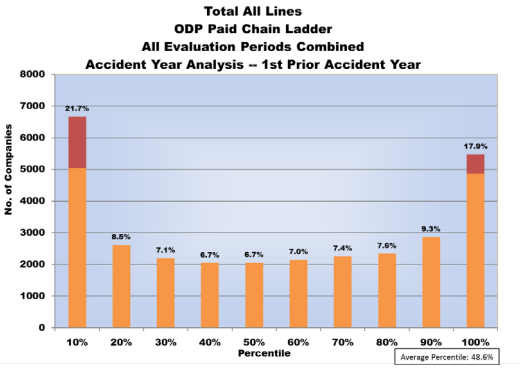
The back-testing results for the current accident year are shown in Graphs 5.3 and 5.9 for the ODP Bootstrap paid chain ladder and incurred chain ladder, respectively, and for completeness are repeated here in Graphs A.1 and A.10. All of the Graphs in Appendix A show results for the ODP Bootstrap paid chain ladder and incurred chain ladder models using the "Baseline Limits & Hetero" assumptions for all lines of business and all evaluation periods combined.

ODP Paid Chain Ladder:

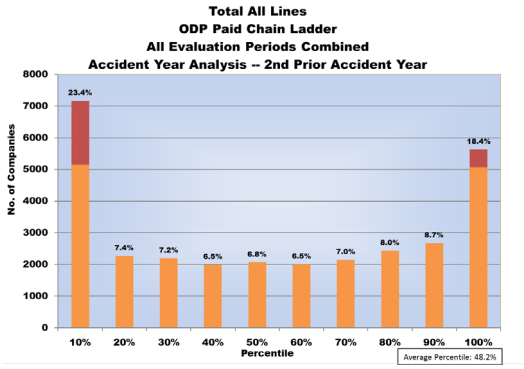
Graph A.1. Current Accident Year



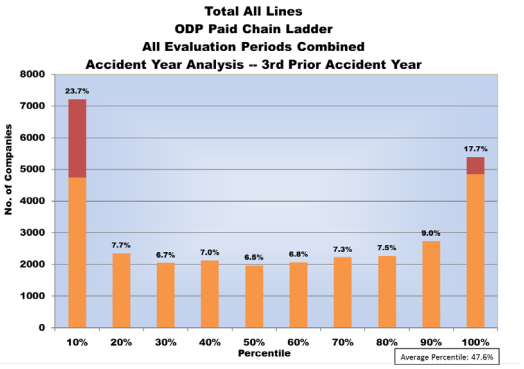
Graph A.2. 1st Prior Accident Year



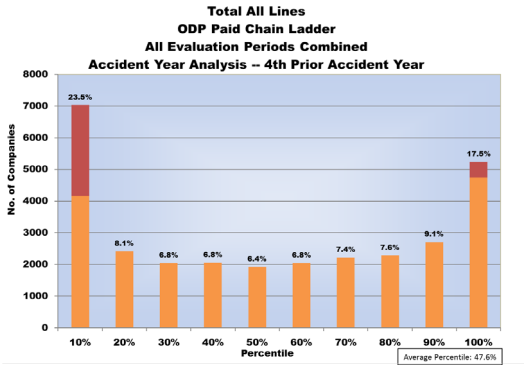
Graph A.3. 2nd Prior Accident Year



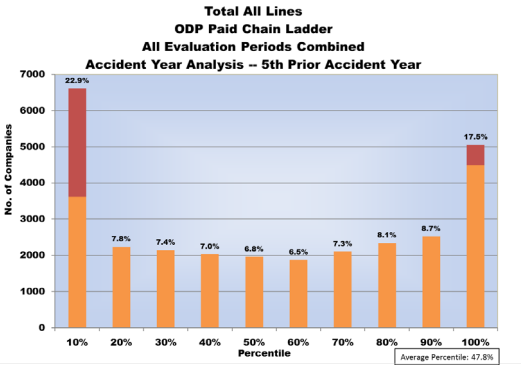
Graph A.4. 3rd Prior Accident Year



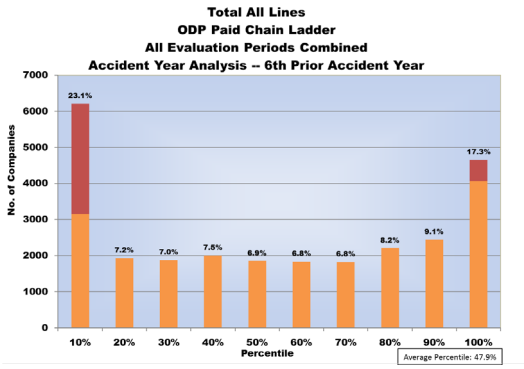
Graph A.5. 4th Prior Accident Year



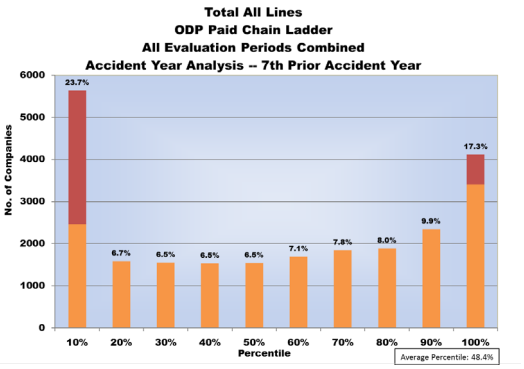
Graph A.6. 5th Prior Accident Year



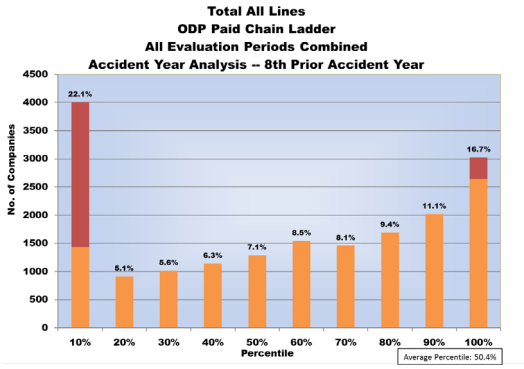
Graph A.7. 6th Prior Accident Year



Graph A.8. 7th Prior Accident Year

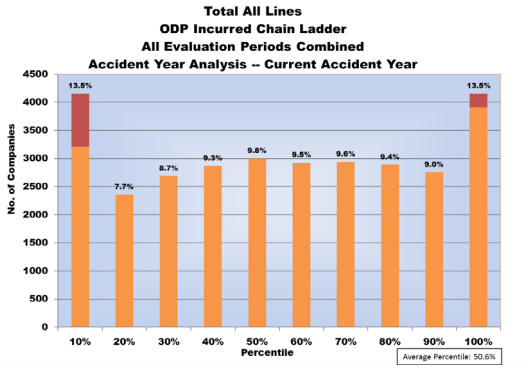


Graph A.9. 8th Prior Accident Year

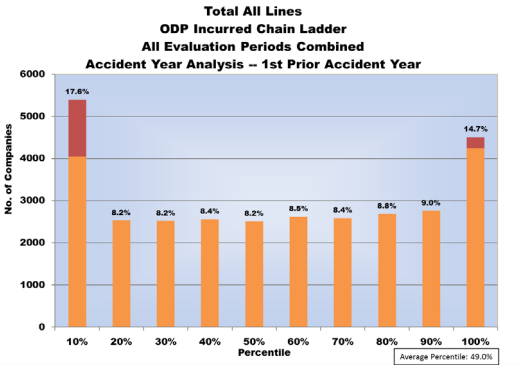


ODP Incurred Chain Ladder:

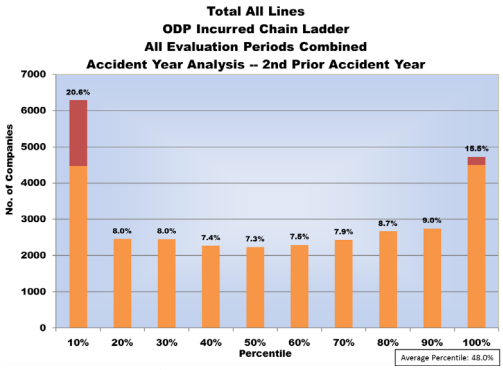
Graph A.10. Current Accident Year



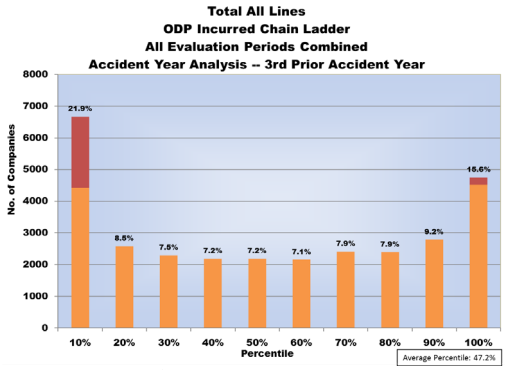
Graph A.11. 1st Prior Accident Year



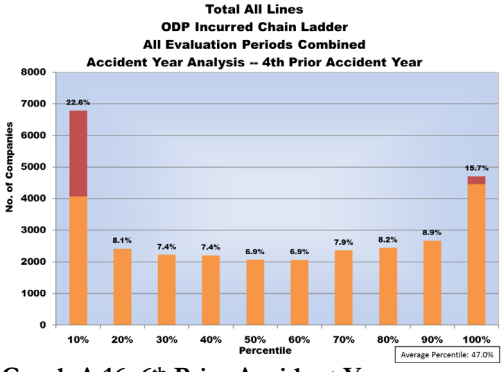
Graph A.12. 2nd Prior Accident Year



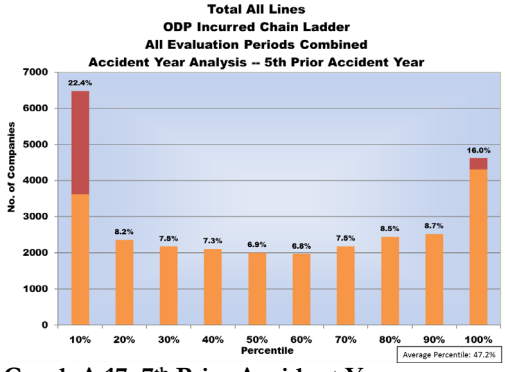
Graph A.13. 3rd Prior Accident Year



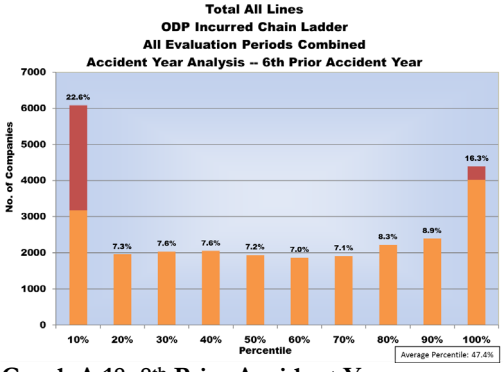
Graph A.14. 4th Prior Accident Year



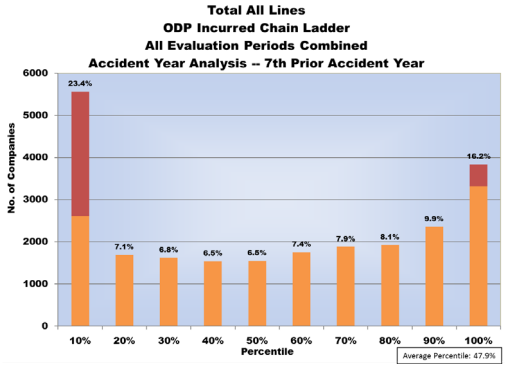
Graph A.15. 5th Prior Accident Year



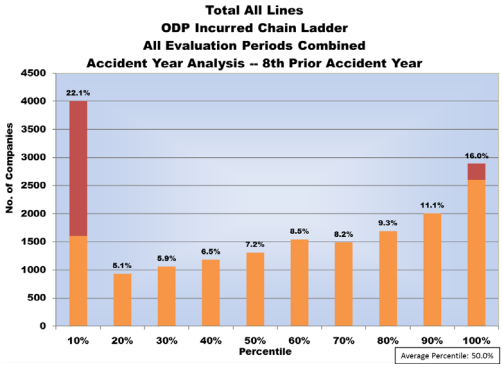
Graph A.16. 6th Prior Accident Year



Graph A.17. 7th Prior Accident Year



Graph A.18. 8th Prior Accident Year

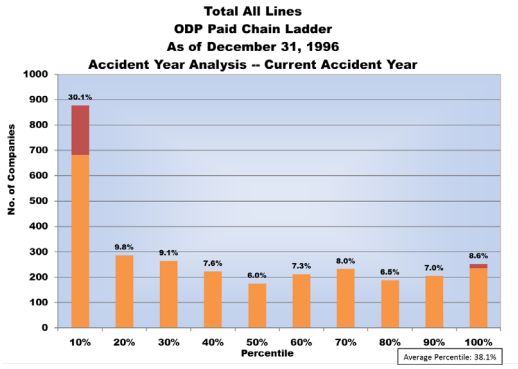


Appendix B – Back-Testing Results by Evaluation Year

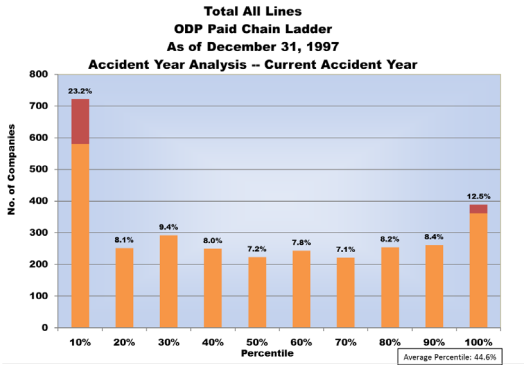
The back-testing results for the current accident year in Graphs 5.3 and 5.9, for the ODP Bootstrap paid chain ladder and incurred chain ladder, respectively, is for all evaluation years combined. All of the Graphs in Appendix B show results for the current accident year for the ODP Bootstrap paid chain ladder and incurred chain ladder models using the "Baseline Limits & Hetero" assumptions for all lines of business by evaluation periods.

ODP Paid Chain Ladder:

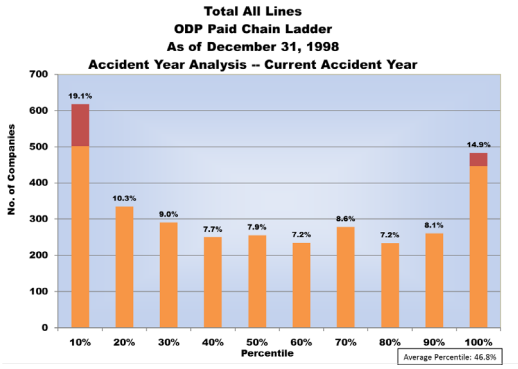
Graph B.1. Evaluation Year 1996



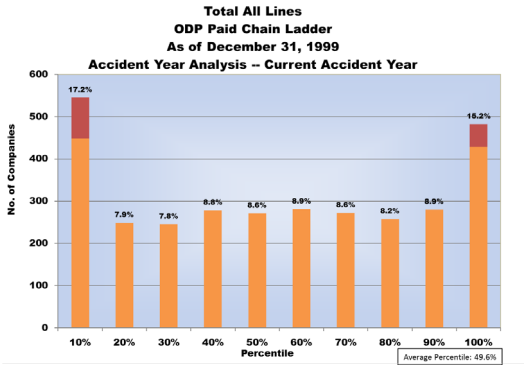
Graph B.2. Evaluation Year 1997



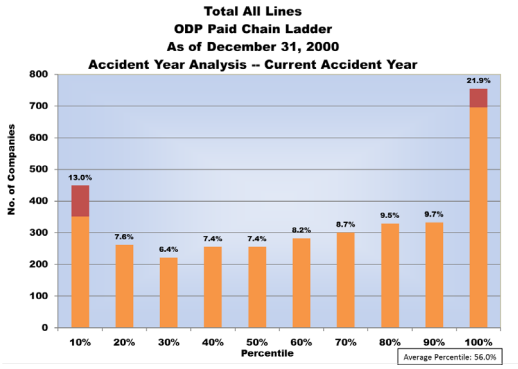
Graph B.3. Evaluation Year 1998



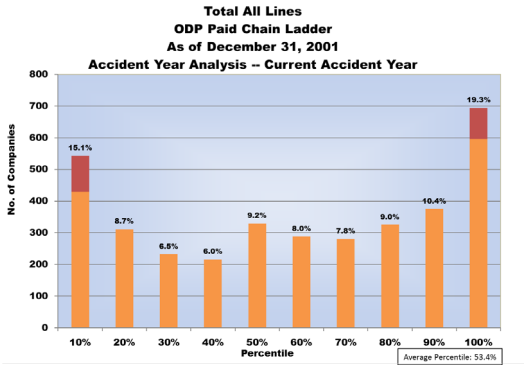
Graph B.4. Evaluation Year 1999



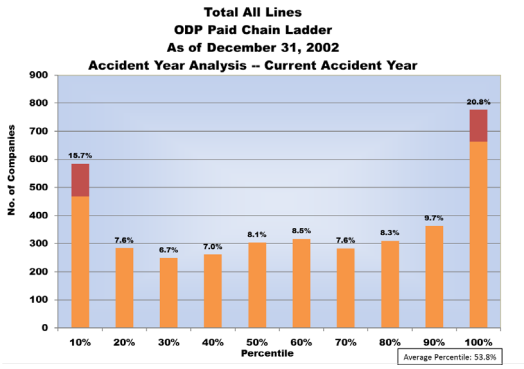
Graph B.5. Evaluation Year 2000



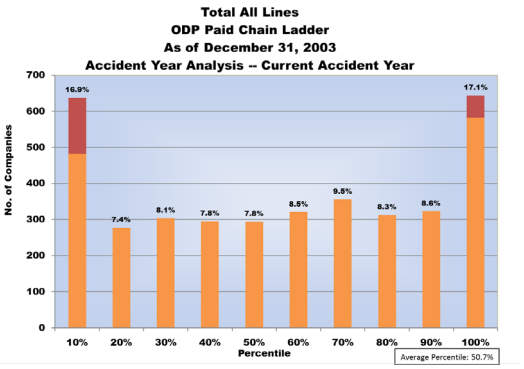
Graph B.6. Evaluation Year 2001



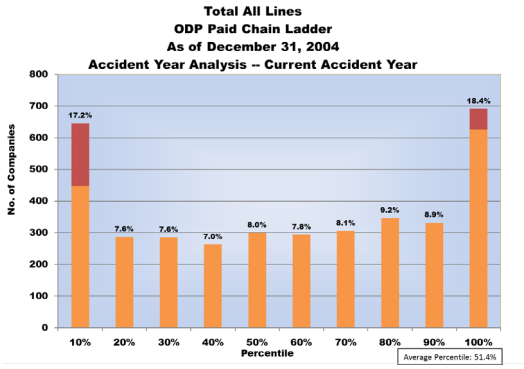
Graph B.7. Evaluation Year 2002



Graph B.8. Evaluation Year 2003

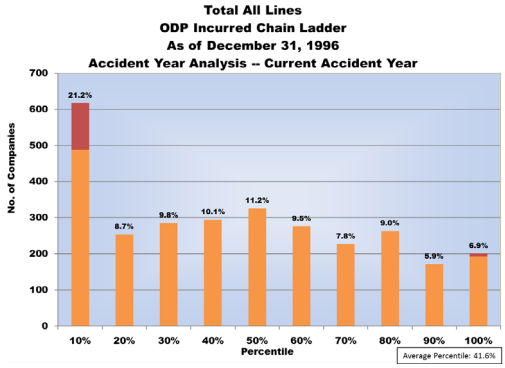


Graph B.9. Evaluation Year 2004

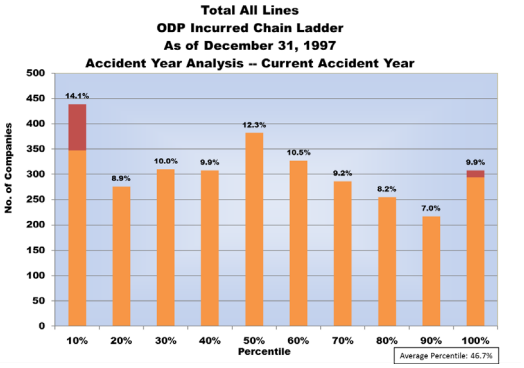


ODP Incurred Chain Ladder:

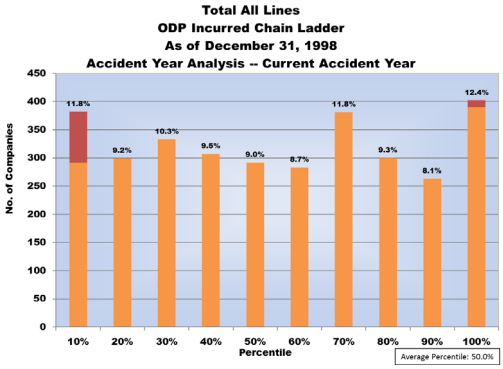
Graph B.10. Evaluation Year 1996



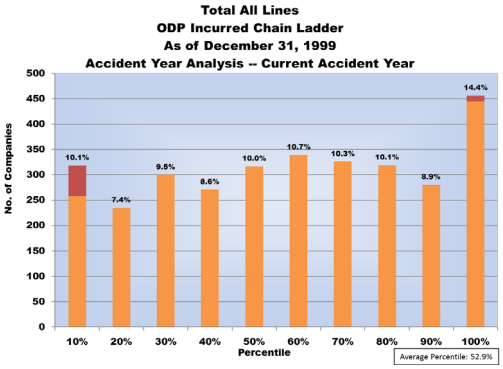
Graph B.11. Evaluation Year 1997



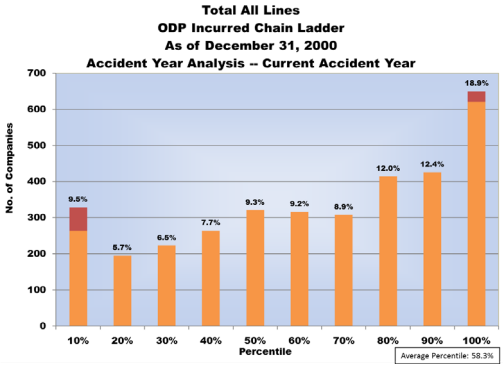
Graph B.12. Evaluation Year 1998



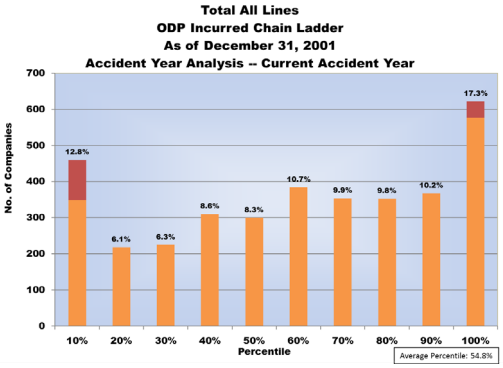
Graph B.13. Evaluation Year 1999



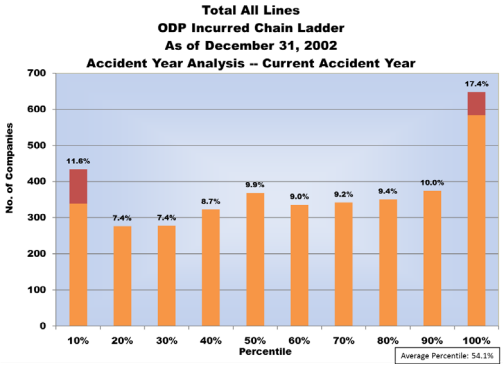
Graph B.14. Evaluation Year 2000



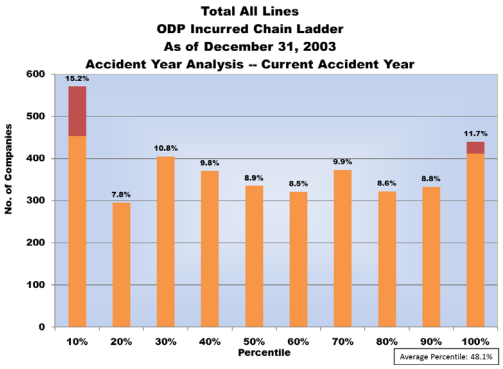
Graph B.15. Evaluation Year 2001



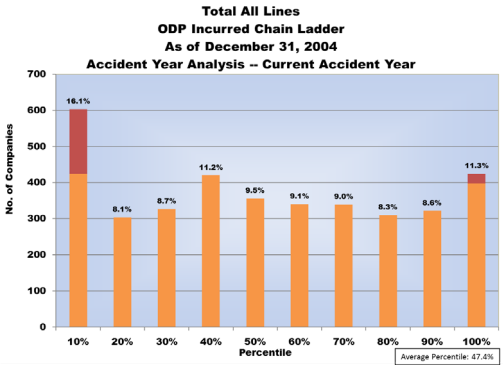
Graph B.16. Evaluation Year 2002



Graph B.17. Evaluation Year 2003



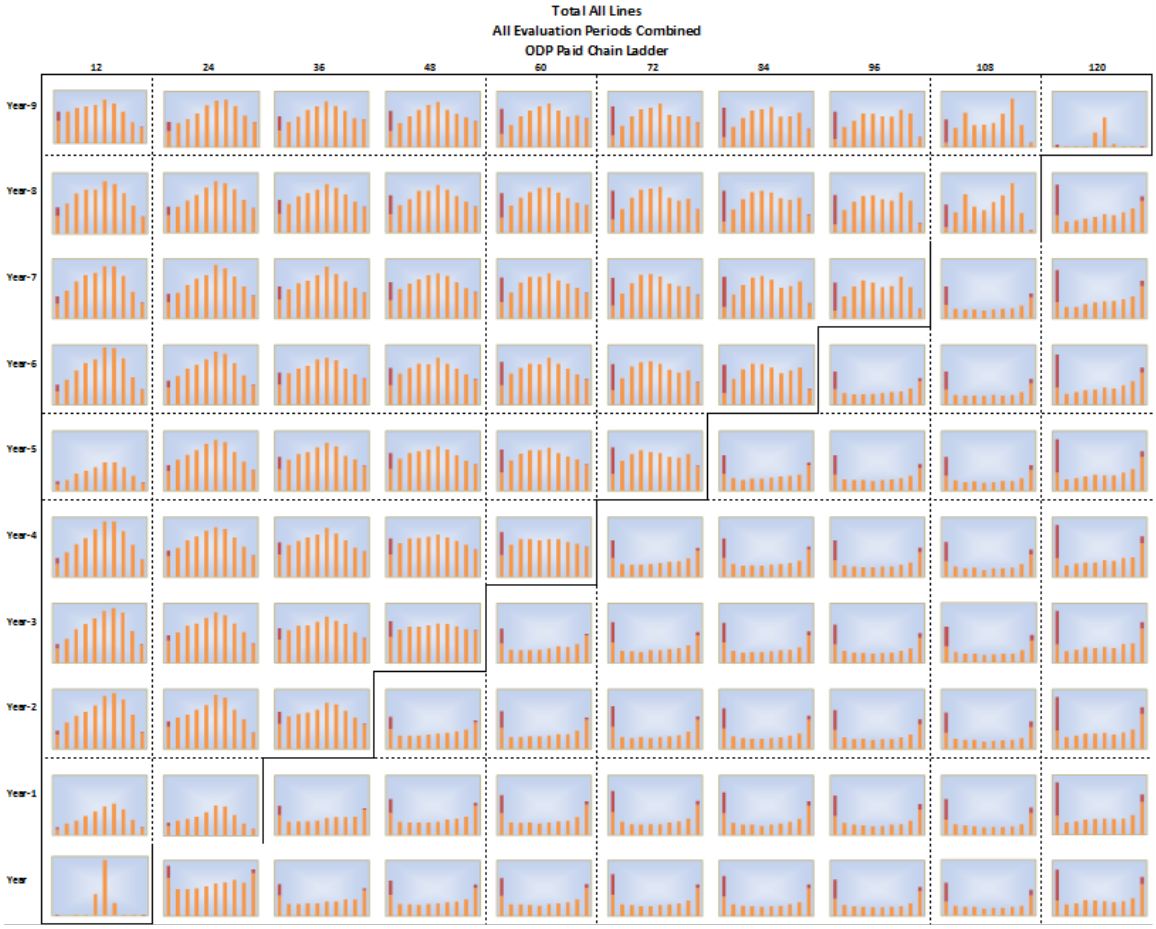
Graph B.18. Evaluation Year 2004



Appendix C – Back-Testing Results by Incremental Cell

The back-testing results in Appendix C show results for the ODP Bootstrap paid chain ladder and incurred chain ladder models using the "Baseline Limits & Hetero" assumptions for all lines of business and all evaluation periods combined.

Graph C.1. ODP Bootstrap Paid Chain Ladder by Incremental Cell



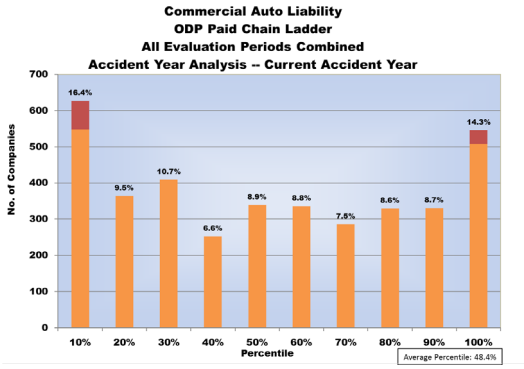
Graph C.2. ODP Bootstrap Incurred Chain Ladder by Incremental Cell

Appendix D – Back-Testing Results by Line of Business

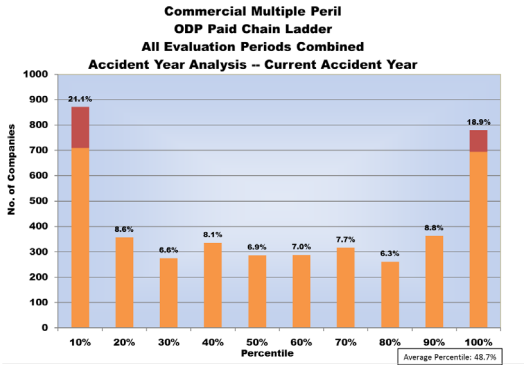
The back-testing results for the current accident year for all lines of business is included as Graphs 5.3 and 5.9 for the ODP Bootstrap paid chain ladder and incurred chain ladder, respectively. All of the Graphs in Appendix D show results for the ODP Bootstrap paid chain ladder and incurred chain ladder models using the "Baseline Limits & Hetero" assumptions for all evaluation periods combined, separately for each Schedule P line of business.

ODP Paid Chain Ladder:

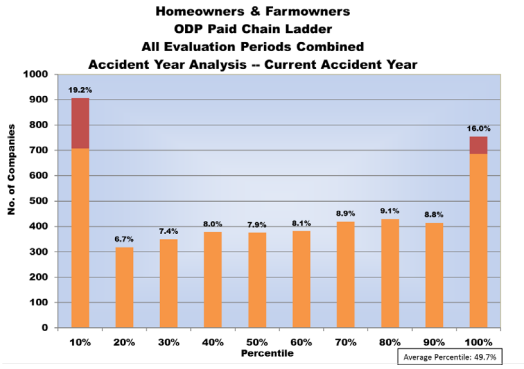
Graph D.1. Commercial Auto Liability



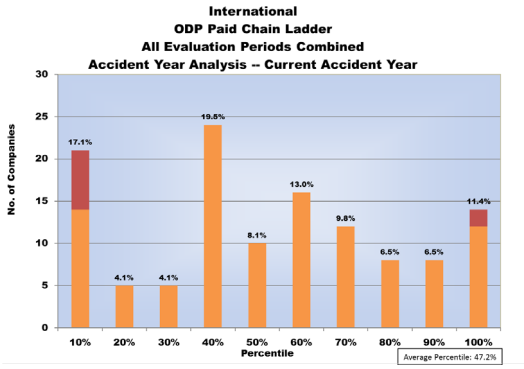
Graph D.2. Commercial Multiple Peril



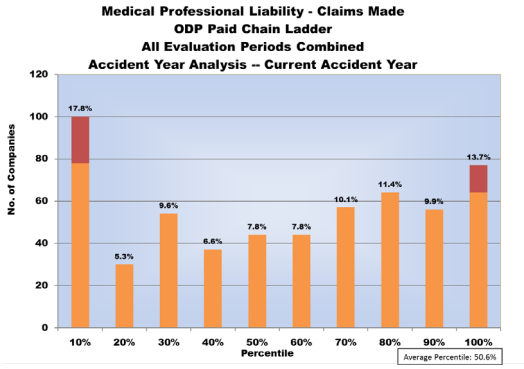
Graph D.3. Homeowners & Farmowners



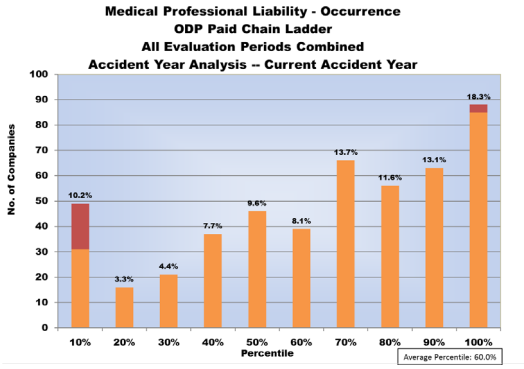
Graph D.4. International



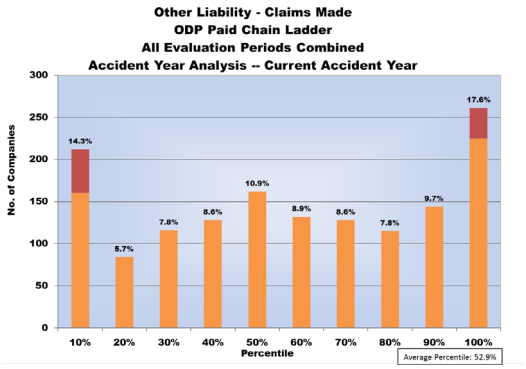
Graph D.5. Med. Prof. Liab. - Claims Made



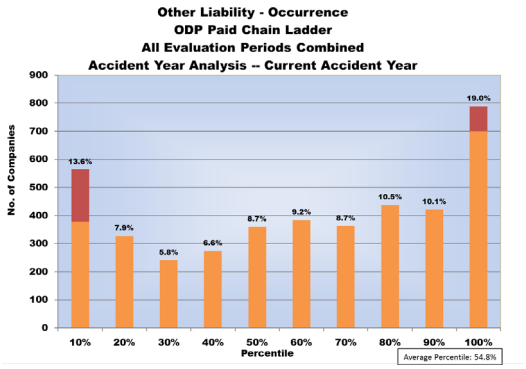
Graph D.6. Med. Prof. Liab. - Occurrence



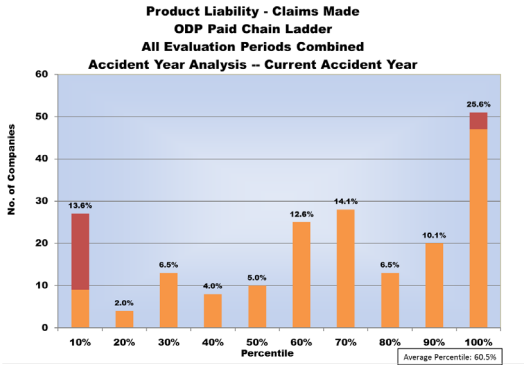
Graph D.7. Other Liability - Claims Made



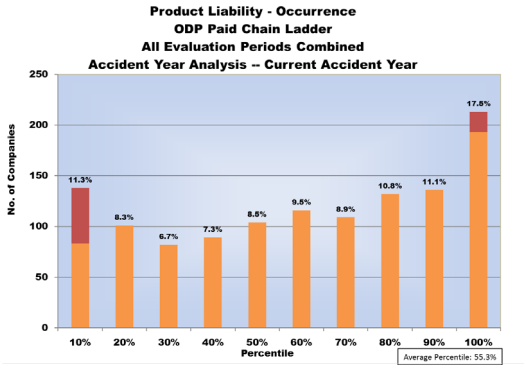
Graph D.8. Other Liability - Occurrence



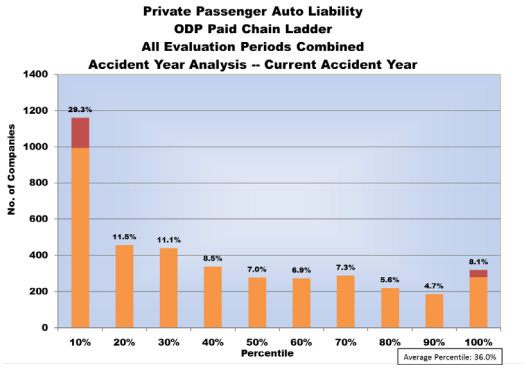
Graph D.9. Product Liability - Claims Made



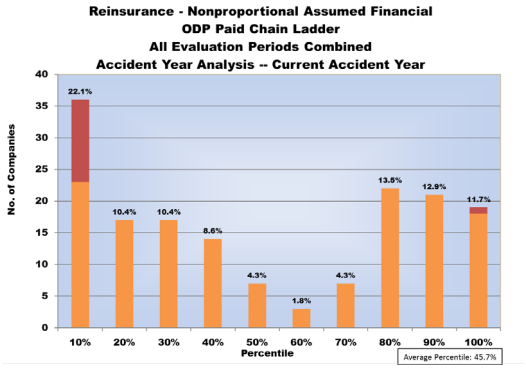
Graph D.10. Product Liability - Occurrence



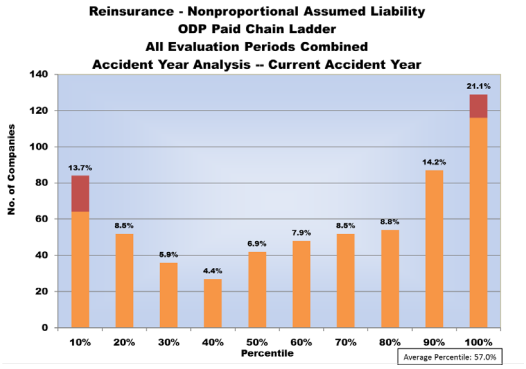
Graph D.11. Private Passenger Auto Liability



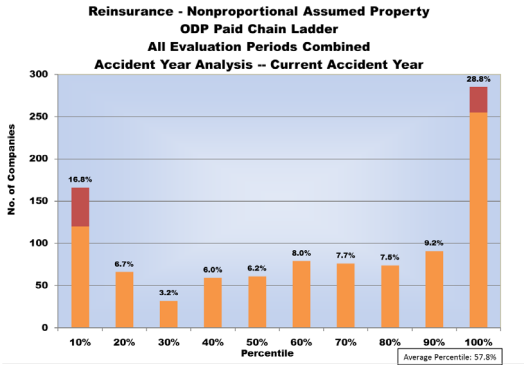
Graph D.12. Reins. – NP Assumed Financial



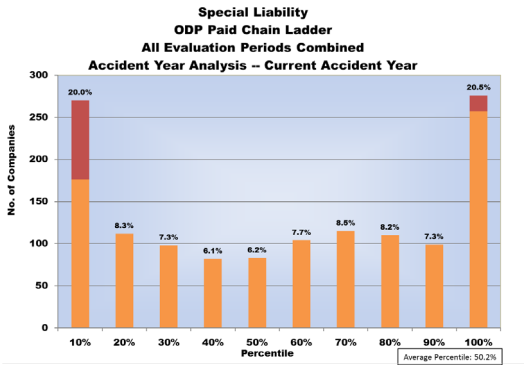
Graph D.13. Reins. – NP Assumed Liability



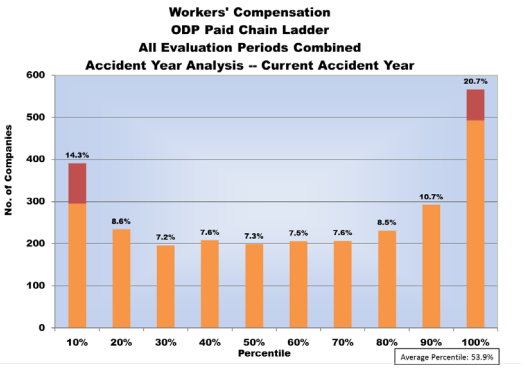
Graph D.14. Reins. – NP Assumed Property



Graph D.15. Special Liability

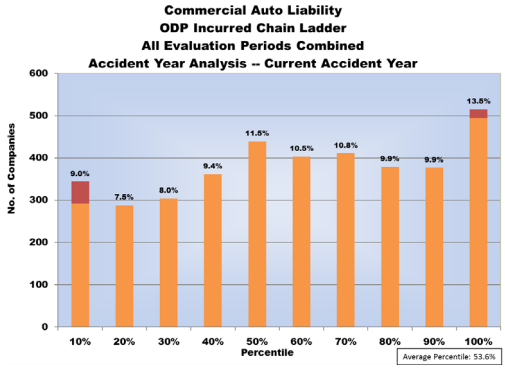


Graph D.16. Workers' Compensation

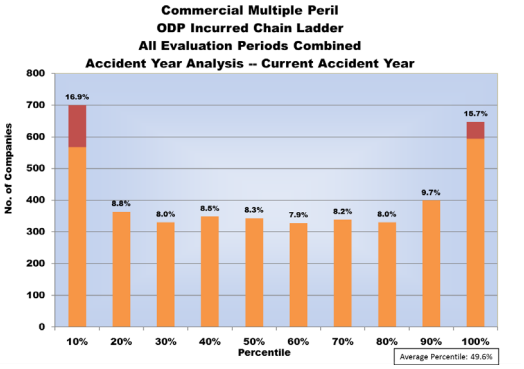


ODP Incurred Chain Ladder:

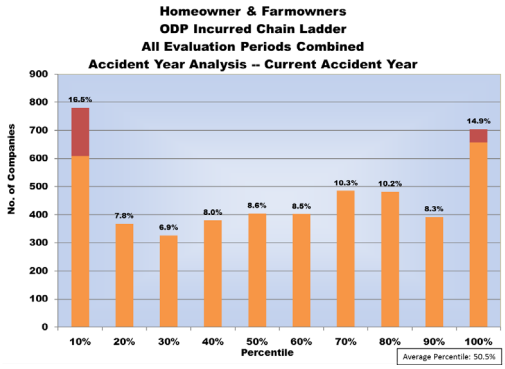
Graph D.17. Commercial Auto Liability



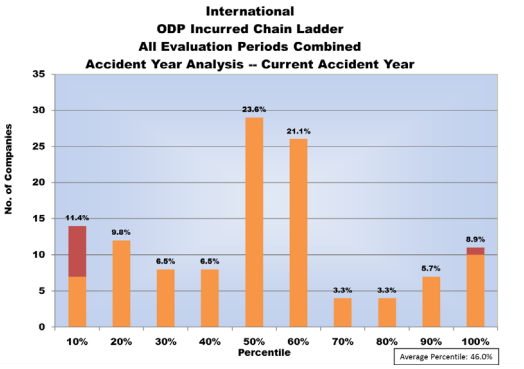
Graph D.18. Commercial Multiple Peril



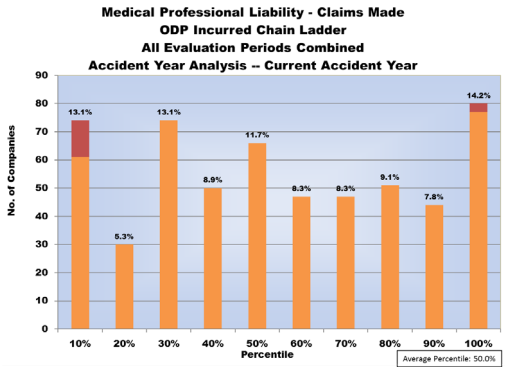
Graph D.19. Homeowners & Farmowners



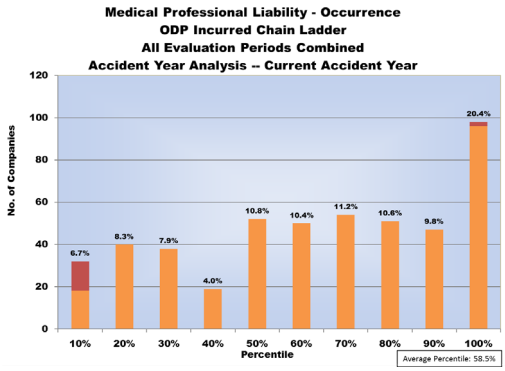
Graph D.20. International



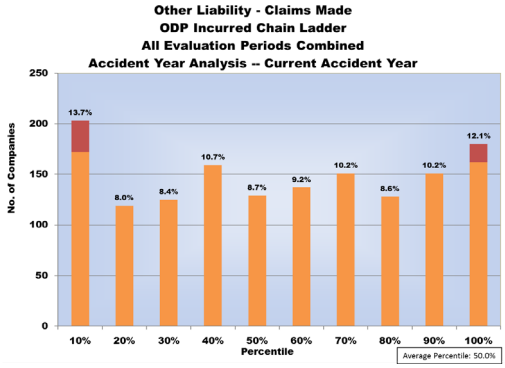
Graph D.21. Med. Prof. Liab. - Claims Made



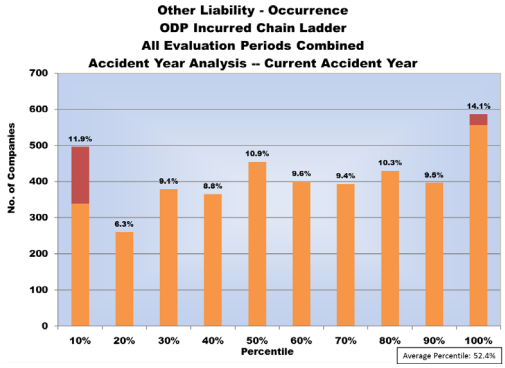
Graph D.22. Med. Prof. Liab. - Occurrence



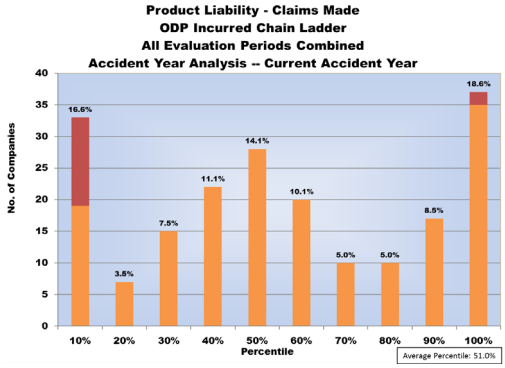
Graph D.23. Other Liability - Claims Made



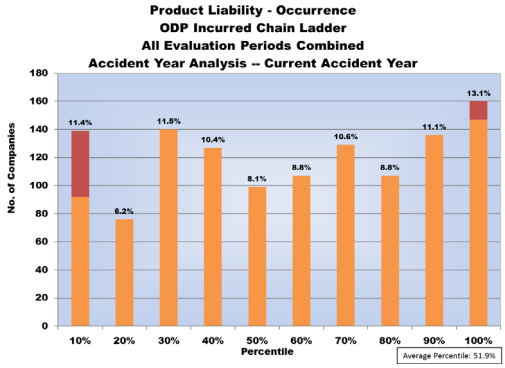
Graph D.24. Other Liability - Occurrence



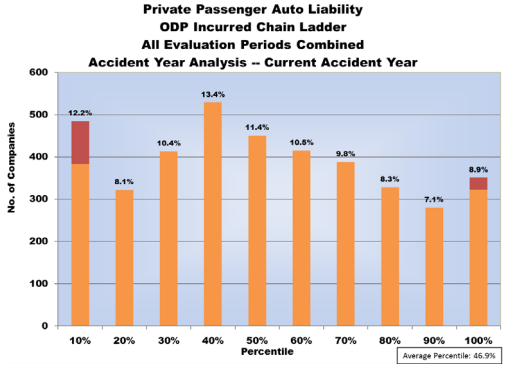
Graph D.25. Product Liability - Claims Made



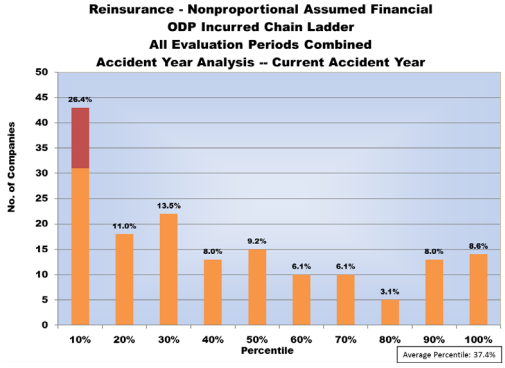
Graph D.26. Product Liability - Occurrence



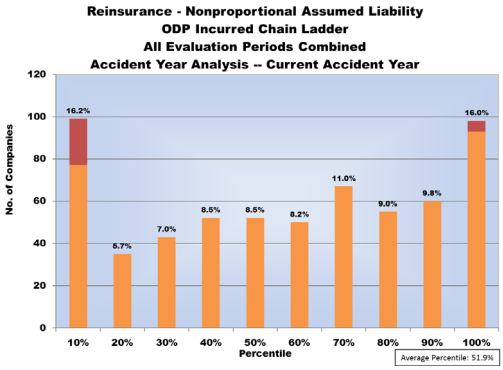
Graph D.27. Private Passenger Auto Liability



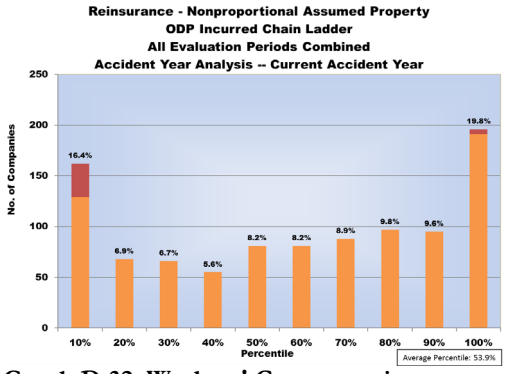
Graph D.28. Reins. – NP Assumed Financial



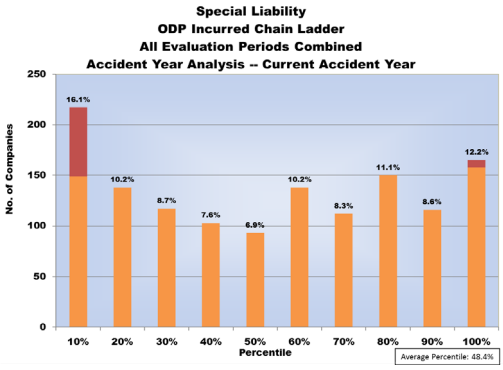
Graph D.29. Reins. – NP Assumed Liability



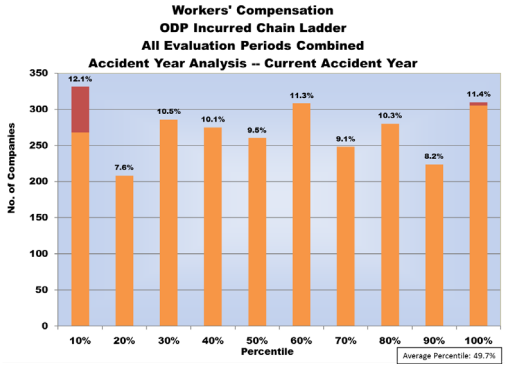
Graph D.30. Reins. – NP Assumed Property



Graph D.31. Special Liability



Graph D.32. Workers' Compensation



REFERENCES

[1] Bornhuetter, Ronald, and Ronald Ferguson. 1972. "The Actuary and IBNR." *Proceedings of the Casualty Actuarial Society* LIX: 181-195.

[2] CAS Working Party on Quantifying Variability in Reserve Estimates. 2005. "The Analysis and Estimation of Loss & ALAE Variability: A Summary Report." *CAS Forum* (Fall): 29-146.

[3] England, Peter D., and Richard J. Verrall. 1999. "Analytic and Bootstrap Estimates of Prediction Errors in Claims Reserving." Insurance: Mathematics and Economics 25: 281-293.

[4] England, Peter D., and Richard J. Verrall. 2002. "Stochastic Claims Reserving in General Insurance." British Actuarial Journal 8-3: 443-544.

[5] England, Peter D., and Richard J. Verrall. 2006. "Predictive Distributions of Outstanding Liabilities in General Insurance." Annals of Actuarial Science 1-2: 221-270.

[6] Forray, Susan J. 2012. "Looking Back to See Ahead: A Hindsight Analysis of Actuarial Reserving Methods." *CAS Forum* (Summer): 1-33.

[7] Gremillet, Marion, Pierre Miehe, and José Luis Vilar Zanón. 2014. "Back-Testing the Reversible Jump Markov Chain Monte Carlo & further extensions." Presented at 2014 ICA in Washington DC.

[8] Iman, R., and W. Conover. 1982. "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables." *Communications in Statistics--Simulation and Computation* 11(3): 311–334.

[9] Jing, Yi, Joseph R. Lebens, and Stephen P. Lowe. 2009. "Claim Reserving: Performance Testing and the Control Cycle." *Variance* 3-2: 161-193.

[10] Kirschner, Gerald S., Colin Kerley, and Belinda Isaacs. 2008. "Two Approaches to Calculating Correlated Reserve Indications Across Multiple Lines of Business." *Variance* 1: 15-38.

[11] Leong, (Weng Kah) Jessica, Shaun S. Wang and Han Chen. 2014. "Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data." *Variance* 8-2: 182-202.

[12] Meyers, G., and P. Shi. 2011. "The Retrospective Testing of Stochastic Loss Reserve Methods." CAS Forum (Summer):

[13] Mildenhall, Stephen J. 2006. "Correlation and Aggregate Loss Distributions with an Emphasis on the Iman-Conover Method." *Casualty Actuarial Society E-Forum* (Winter): 103-204.

[14] Pinheiro, Paulo J. R., João Manuel Andrade e Silva, and Maria de Lourdes Centeno. 2003. "Bootstrap Methodology in Claim Reserving." Journal of Risk and Insurance 70: 701-714.

[15] ROC/GIRO Working Party. 2007. "Best Estimates and Reserving Uncertainty." Institute of Actuaries.

[16] ROC/GIRO Working Party. 2008. "Reserving Uncertainty." Institute of Actuaries.

[17] Shapland, Mark R. 2016. "Using the ODP Bootstrap Model: A Practitioner's Guide." *CAS Monograph 4*.

[18] Skurnick, David. 1973. "A Survey of Loss Reserving Methods." *Proceedings of the Casualty Actuarial Society* LX: 16-58.

[19] Struzzieri, Paul J., and Paul R. Hussian. 1998. "Using Best Practices to Determine a Best Reserve Estimate." *Casualty Actuarial Society Forum* (Fall): 353-413.

[20] Venter, Gary G. 1998. "Testing the Assumptions of Age-to-Age Factors." *Proceedings of the Casualty Actuarial Society* LXXXV: 807-47.

Abbreviations and notations

APD, automobile physical damage
ATA, age-to-age
CL, chain ladder
DFA, dynamic financial analysis
GLM, generalized linear models
OLS, ordinary least squares
ERM, enterprise risk management

Biography of the Author

Mark R. Shapland is a Principal & Consulting Actuary in Milliman's Dubai office where he is responsible for various reserving and pricing projects for a variety of clients and was previously the lead actuary for the Property & Casualty Insurance Software (PCIS) development team at Milliman. He has a B.S. degree in Integrated Studies (Actuarial Science) from the University of Nebraska-Lincoln. He is a Fellow of the Casualty Actuarial Society, a Fellow of the Society of Actuaries, a Fellow of the Institute of Actuaries of India, and a Member of the American Academy of Actuaries. He was the leader of Section 3 of the Reserve Variability Working Party, the Chair of the CAS Committee on Reserves, co-chair of the Tail Factor Working Party, and co-chair of the Loss Simulation Model Working Party. He is also a co-developer and co-presenter of the CAS Reserve Variability Limited Attendance Seminar and has spoken frequently on this subject both within the CAS and internationally. He can be contacted at mark.shapland@milliman.com.

Enhancements to the Shane-Morelli Method to Provide Technical Guidance in Implementation and Proposed Solutions for Challenges Encountered in the Application to a Workers Compensation Tail

Shon Yim, PhD, ACAS

Dolph Zielinski, FCAS

Dawn (Morelli) Fowle, FCAS

Abstract

Motivation. Application of the Shane-Morelli method in practice for multiple reserve reviews revealed potential areas of refinement.

Method. A theoretical examination of curves best used to develop workers' compensation tail factors resulted in a proposed enhancement to this part of the original methodology. Further, a melding of theoretical and practical considerations gave rise to an approach for gradual introduction of mortality to the curve for determination of a more realistic tail. Finally, the determination of a process for updating the proposed model was developed for use by a practicing actuary.

Results. The paper presents a new curve for tail fits, the gradual introduction of mortality into the tail calculation and a process for regularly updating the model.

Conclusions. The updated approach provides an enhanced way of incorporating mortality into traditional aggregate reserving methods in a manner that can be readily explained to business partners.

Keywords. Workers' Compensation; Reserving Methods; Tail; Mortality; Gamma Curve; Shane-Morelli

1. INTRODUCTION

1.1 Background

Modeling of the tail development in worker compensation is well known for being challenging, but it is an important task and thus remains an active area of interest [1]. One approach is to fit a parametric curve to the loss development factors (i.e., LDFs or link ratios) and extrapolate the curve beyond the triangle in order to effectively achieve an extension of the chain ladder method. The limitation with this method is that many parametric curves, such as the popular inverse power, do not converge, so the extrapolation must be truncated at some selected point in the future. Unfortunately, the choice of the truncation point often drives significant swings in the estimate of reserves and can be highly subjective when selected solely based on judgment.

A recent paper by Shane and Morelli provides a practical solution to the truncation problem, hereafter referred to as the Shane-Fowle method [2]. Their method is based on the curve fit approach using the inverse power curve to fit the LDFs. Their key contribution is the use of the life expectancy

of the actual claimants to determine the duration of development in each accident year so that the extrapolation of the curve can be truncated using relevant information rather than simple judgment. The Shane-Fowle method is intuitive and can be applied readily. As with any model, however, application to new problems reveals challenges and potential for improvement. The purpose of this paper is to offer enhancements to the application of the Shane-Fowle method in the form of additional discussion around the selection of an appropriate curve, the gradual application of mortality to replace the truncation, and additional guidance around the updating of the underlying assumptions over time.

1.2 Scope

The analysis in this paper is limited to the development of workers' compensation claims in the form of the familiar aggregate runoff triangles. Both paid and incurred data are in scope for medical and indemnity, as are related defense and cost containment expenses. Our model is concerned only with the development of "lifetime claims," by which we mean claims expected to be paid regularly for the remainder of a claimant's life. Based on our experience, ten years is a reasonable threshold for the development age prior to which most shorter duration claims will have closed. This means that our analysis and model are not intended to be used on the most recent ten accident years.

1.3 Outline

The remainder of the paper proceeds as follows. In Section 2, we propose enhancements to the Shane-Fowle method. In particular, we first focus here on the selection of an appropriate curve, with discussion of the commonly known inverse power and exponential curves and the introduction of the gamma curve as a potential improvement. We then move on to a discussion of a method to gradually apply the effect of mortality, which we believe is an improvement to the truncation described in the Shane-Fowle method. In Section 3 we discuss business considerations, including the frequency of model and parameter updates, where we propose a "locked in" approach to the parameters for a pre-selected period of time to avoid overreactions to noise in the underlying data and a framework for understanding and explaining the changes in reserve estimates produced by the model. Section 4 is the conclusion, followed by appendices to support the main body of the paper.

2. ENHANCEMENTS TO THE SHANE-FOWLE METHOD

In this section, we present two enhancements to the original Shane-Fowle method. The first enhancement involves providing guidance to the selection of the curve fit; the second enhancement

offers a modification to the application of mortality.

2.1 Notation

The mathematical notation in this section closely follows the 2013 paper by the CAS Tail Factor Working Party [1].

development age in years.

$c(d)$ cumulative loss (paid or incurred) as of development age d.

$q(d)$ incremental loss from age $d-1$ to d.

$f(d)$ age-to-age LDF (or link ratio) such that $c(d+1) = c(d)f(d)$.

$v(d)$ "development portion" of the link ratio, such that $f(d) = 1 + v(d)$. For brevity, this is referred to as the development ratio in this paper.

annual decay rate of incremental losses such that $\beta = [q(d) - q(d+1)]/q(d)$.

2.2 Enhancement to the Curve Fit Selection

Citing the analysis by Sherman [3], Shane-Fowle chose the inverse power curve for fitting the age-to-age LDFs,

$$\hat{v}(d) = \quad , \tag{2.1}$$

where $\hat{v}(d)$ denotes the fitted development ratio at age d, and a and b are the fit parameters. The inverse power curve is a widely accepted choice for fitting LDFs [1], but justification for using it appears to be based on the observation that it often provides a reasonably good fit rather than providing sufficient information to allow the user to make an informed selection of the appropriate curve for the data being fit. Even if mortality is arguably the most important factor for a tail model, mortality doesn't start to dominate until much later in the tail, at claimant age 70 or so based on a review of the current life tables. Until then, the projection of loss is determined largely by the curve fit. Thus we believe it makes sense to invest more analysis into the selection of the curve fit.

2.2.1 Analysis of constant decay loss model

To gain some insights that will help guide our selection of the curve fit, we first conduct a simple analysis of a theoretical loss development model. This model will focus on only one accident year. We assume that the incremental loss $q(d)$ decays with age d at a constant rate of decay β such that the sequence of incremental losses starting at $d = 1$ are $q(1)$, $q(1)(1-\beta)$, $q(1)(1-\beta)^2$, and so

on. The decay rate β is restricted to the range [0, 1], which means that negative development and increasing incremental losses are not considered. The latter omission might be concerning if we were working with severity data subject to inflation, for example, but the Shane-Fowle method is based on aggregate runoff triangles where losses are expected to decay over time, making the scenario where $\beta < 0$ unlikely. This constant decay model was previously analyzed by McClenahan [4], but his effort was focused more on the formulation of the reserves. In contrast, our purpose is to understand how this particular development pattern relates to the curve fit of the development ratio.

Given a constant value of β, the development ratio can be expressed as

$$\nu(d) = \frac{\beta(1-\beta)}{1-(1-\beta)}. \tag{2.2}$$

The derivation of Equation 2.2 is provided in Appendix A. We will next evaluate this expression for the scenarios of no decay ($\beta = 0$) and significant decay ($\beta \to 1$).

No decay

At $\beta = 0$, Equation 2.2 yields the indeterminate expression $0/0$, so we use L'Hôpital's rule to evaluate the limit,

$$\lim_{\beta\to 0} \nu(d) = \lim_{\beta\to 0} \frac{-\quad (1-\beta)^{d-1} + (1-\beta)}{d(1-\beta)} \quad = \qquad . \tag{2.3}$$

This is the inverse power curve of Equation 2.1 with parameters $a = 1$ and $b = -1$.

Significant decay

At $\beta = 1$, Equation 2.2 collapses to the trivial solution of $\nu(d) = 0$, which is not meaningful since it equates to no development. So instead of evaluating at $\beta = 1$, we can see what happens as β approaches one. Looking at the denominator of Equation 2.2, the unity term will dominate over the $(1-\beta)^d$ term as β gets closer to one, and consequently Equation 2.2 will tend toward the expression $\nu(d) = \beta(1-\beta)^d$. This is the familiar form of the geometric distribution, which is the discrete analogue of the exponential curve.

To summarize, within the construct of the constant decay model, it is the decay rate that determines which curve fit is optimal. At one extreme, where losses do not decay, the inverse power curve provides the ideal fit. At the other extreme of significant decay, the exponential curve is ideal. In between, the appropriate fit is some blend of the two forms, with a gradual transition from the inverse

power to the exponential as β goes from 0 to 1. Appendix B illustrates this transition with a hypothetical example. As expected, the inverse power gives the perfect fit when there is no decay. As long as the decay is low ($\beta < 0.07$), the inverse power fits better than the exponential. But as β increases, the exponential curve becomes superior, and by the time $\beta = 0.2$ the exponential is already practically an ideal fit with an R^2 value of 1.000.

2.2.2 Curve fit selection

The insights from the constant decay loss model analysis can be used to guide the selection of an appropriate curve fit. Below, we discuss the inverse power and exponential forms, finishing with a recommendation for the gamma form.

Inverse power curve

The inverse power curve is ideal when incremental losses have zero or very low decay with development age. For aggregate triangles, this is most likely to happen when the claimant count and the loss run rate (i.e., dollar amount per claimant per year) both remain steady or decrease very slowly over time. Slowly decreasing claimant count is certainly possible for lifetime claims beyond ten years of development and prior to very mature ages, where mortality takes over. Steady run rates can be expected in indemnity paid data with fixed wage replacement payments and without cost of living adjustments. Medical paid data could also exhibit level run rates if most claims are in a steady state of routine treatments and if medical inflation is unremarkable. Finally, paid expense triangles often settle into regular increments in the lifetime development phase. The conclusion here is that the inverse power curve should always be considered for paid data.

Exponential curve

Even paid triangles can exhibit fast decaying incremental losses if the claimant attrition rate is high, which would likely be true for an aged population. Therefore, when the inverse power curve is clearly struggling to fit due to excessive decay, the exponential curve should be considered. With incurred triangles, significant decay may be expected at early ages in cases where a sizable initial case reserve is followed by smaller reserve adjustments in subsequent years. Decay could continue even into mature years to the extent that information about remaining liabilities generally improves over time and results in continually smaller reserve adjustments. In fact, it would be unusual for incremental incurred losses to have little or no decay so exponential curves may often be chosen over inverse power when it comes to incurred data.

Equation 2.2

At first, Equation 2.2 seems like an appealing choice since it effectively represents both the inverse power and the exponential curves as well as the spectrum in between. But there are two reasons against using it. The first is that the premise of this formula is the assumption of a constant decay rate, which is unlikely to be observed in practice. The second reason is that Equation 2.2 is a single parameter curve. As any practitioner can attest, an attempt to fit data as nonlinear as workers' compensation losses to any single parameter curve would likely be a futile exercise. Real data is messy with decay rates and other trends that often change unpredictably over time. For practical reasons, we would want a curve form with more than one parameter to be able to accommodate such data.

Gamma curve

A single curve that can fit both slow and fast decaying losses would be beneficial. We therefore recommend the following form[1]:

$$v(d) = \qquad . \tag{2.4}$$

This equation has the form of the gamma distribution. It consists of three parameters: the scale parameter A, the inverse power decay rate parameter b, and the exponential decay parameter r. This curve form has the following benefits and drawbacks:

- It contains both the inverse power and the exponential forms, so it can fit both low and high decaying loss patterns. The data itself will determine which curve form dominates, or the final form could be a blend of the two, but we do not have to manually choose one or the other.
- It is versatile in that it can be forced to be purely inverse power by constraining r to be zero, or it can be forced to be purely exponential by constraining b to be zero.
- It has three parameters, which allows for more flexibility than the inverse power or exponential alone, but also limits the number of parameters to avoid overfitting.
- Unlike the exponential and the inverse power, curve fitting for the gamma curve is not readily available through standard plotting applications such as Excel. While the choice of the curve fitting method is left to the readers' discretion, one option is to use an iterative algorithm for minimizing the sum of squared error using Excel's Solver function.
- The gamma curve is not commonly used for this purpose, and as a result, additional

[1] We credit Robert Ballmer, FCAS for suggesting this curve form.

communication with business partners regarding the justification of its application may need to take place.

2.3 Enhancement to the Application of Mortality

The second enhancement to the Shane-Fowle method has to do with the application of mortality in the loss projection. Shane and Fowle applied mortality by estimating the average remaining life expectancy of the underlying claimants and calculating a weighted average of life expectancy by accident year. They additionally introduced a useful method of evaluating life expectancy at desired percentiles, which allows the practitioner to judgmentally use percentiles in lieu of the expected value. The selected life expectancy for an accident year was then used to truncate the extrapolated curve fit to obtain the ultimate loss and the corresponding tail factor. Their approach has several merits, including being intuitively appealing and having high practical value. However, we see opportunities for improvement in view of the following considerations.

- The average of the individual life expectancies in a cohort will underestimate the life expectancy of the cohort. If we say that a cohort "dies" with its last remaining member, then the life expectancy of the cohort should exceed the average of the individual life expectancies. Consider the extreme example of a cohort of two people, one with a life expectancy of one year and the other with nineteen years. The average of their life expectancies is ten years. But the life expectancy of the cohort would be greater than ten years because the life expectancy of the youngest member is greater than ten years.
- A cohort in runoff dies off gradually, so it follows that the loss development should also be affected gradually by the force of mortality. Under Shane and Fowle's method of truncation, mortality is introduced at a single point in time – the point considered to be the ultimate. Prior to then, losses are assumed to develop without contribution from mortality other than the mortality which is embedded in the observed triangle. But mortality is nonlinear and observed mortality is a not a good predictor of future mortality.

2.3.1 Methodology for gradual application of mortality

To address the above considerations, we propose a gradual application of mortality[2]. This methodology requires the distribution of claimant ages as well as life tables for the claimant population. If gender is to be considered, and we believe it should, separate life tables for males and females are

[2] We again credit Robert Ballmer, FCAS for his significant role in the development of this methodology.

required along with the gender of each claimant. The number of years of projection is somewhat arbitrary as long as it is enough to cover the remaining life of all claimants. For illustrative purposes, one hundred years is likely a reasonable selection.

Prior to the projection, we assume that the link ratios have already been obtained using standard chain ladder methods and that a curve has already been fit to the development ratios. The steps below are then used to project losses for an individual accident year. For an illustration of the calculations for an example accident year, refer to Appendix C.

1. Group the claimants for the accident year and compute the group's current average age as well as its average age for every year of projection. The projected average group age is estimated using life contingencies whereby the group age at a future year is a weighted average of the member ages with the weights being equal to the probability of survival.

$$\text{age}_G(t) = \frac{\sum \quad \text{age}_k(t) S_k(t)}{\sum \quad (t)} \tag{2.5}$$

   In Equation 2.5, N is the number of members in the accident year and t is a time variable representing the number of years into the projection. $\text{age}_G(t)$ is the group average age at time t. $\text{age}_k(t)$ is the age of member k at time t, which is equal to member k's current age plus t. $S_k(t)$, the probability that member k will survive the next t years, can be calculated from the life tables. For example, if hypothetical member k is a 60-year-old male, $S_k(3) = p_{60}p_{61}p_{62}$, where p denotes the probability that a male at age x will survive to age $x+1$. Note that the group ages more slowly than an individual because the oldest members of the group are more likely to leave the group in the following year.

2. Estimate the group mortality rate for each year of projection by using the group average age calculated in the previous step to look up the mortality rate from the selected life table. If using gender-specific life tables, we can obtain a weighted average mortality rate using weights that reflect the gender split for the group.

3. Use the curve fit for the development ratios (Equation 2.4) to project annual incremental future losses. First, cumulative losses can be projected year after year using $c(d+1) = c(d)(1+v(d))$, starting with the current diagonal as the initial value. Then the incremental losses can be obtained by $q(d) = c(d) - c(d-1)$.

4. Use the incremental projected losses to compute an implied loss decay rate for each projected year.

5. The implied loss decay rate obtained in the previous step includes mortality observed in the triangle data. We want to remove this observed mortality in order to avoid double counting when we apply mortality later in a separate step. A precise evaluation of the observed mortality for each development age is impractically cumbersome, so we rely instead on the most recently observed mortality rate. This observed mortality rate is subtracted from the implied loss decay rate at every projected year. See Appendix C for additional detail.

6. Next we add the group mortality rate estimated in Step 2 to the restated implied loss decay rate from Step 5. This gives us the implied loss decay rate adjusted to include mortality. The subtle assumption made in this step is that loss decay and life mortality rates are additive. This assumption implies that the loss decay comes entirely from the loss of claimants and not from changes in the per claimant severity. This is a reasonable assumption for lifetime paid indemnity since many claimants receive a fixed wage replacement amount. Medical payments are also often steady due to routine treatments and medications although there is more variability here for a number of reasons, including the potential for medical technology and treatments to change over time, and the impact of claim handling practices around closed/reopened claims for routine infrequent treatments. The use of aggregate data should help smooth this volatility to some degree.

7. The final step is to apply the loss decay rate adjusted for mortality in Step 6 to estimate the future incremental losses. The future losses are then summed to produce the unpaid for the accident year.

2.3.2 Comparison of the gradual and truncated application of mortality

Using the projected incremental losses from the example in Appendix C, we can compare the gradual application of mortality described above against a truncated application. Figure 2.1 provides a visual representation to help understand the differences between the two methods. In the Appendix C example, the average life expectancy of the accident year cohort was computed to be 26 years, and that is where the truncated method stops projecting. On the other hand, the gradual application method projects incremental losses that taper down to zero over a period of 50 years. We can also see that losses from the gradual method decrease at a faster rate than losses from the truncation method due to the mortality adjustment applied every year. Comparing the two loss patterns, the gradual method also probably provides a more realistic cash flow projection. The total unpaid estimates for the gradual and truncated application are 1,047 and 964, respectively. It is not surprising that the gradual produced a higher unpaid estimate since it projects a longer life expectancy for the group (though also recall that the original Shane-Fowle method would have used an inverse power

which, in practice, would have yielded a higher unpaid estimate than shown here, so this is not a representation of the difference between the original method and our proposed enhancements). In practice, the actuary may choose to use either or both of the methods due to any number of considerations. For example, consider a scenario where the case reserves are presumed to be set adequately to cover all future payments at some point prior to the ultimate life expectancy of the cohort. In this case, the actuary may opt to use the truncated method for the incurred development, while the gradual method may be appropriate for paid development.

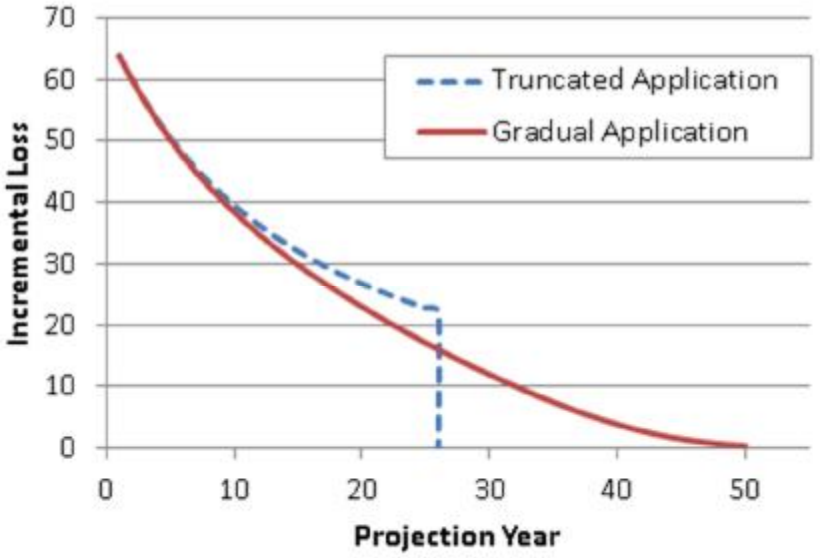


Figure 2.1: Projected incremental losses using the gradual and truncated application of mortality. Results are from the example presented in Appendix C. For clarity, curves are shown as continuous lines, but incremental losses are discrete for each projection year.

3. PRACTICAL CONSIDERATIONS RELATED TO APPLYING THE MODEL FOR SUBSEQUENT RESERVE REVIEW CYCLES

3.1 UPDATING THE MODEL

Since the "tail" analysis may often be one of the most leveraged assumptions in the projection to ultimate on long tail lines, it can also drive some of the most significant financial impacts.

Enhancements to the Shane-Morelli Method

Subsequent utilizations of the model described above will likely generate ultimate indications that differ from prior model results. This introduces the age-old debate around stability versus responsiveness. To address this concern, we will explore the concept of "locking in" parameters for an extended period of time with contemporaneous reasonability checks of those parameters.[3]

We propose that when the model is updated, two contemporaneous exercises take place:

- Exercise 1: Updating the Data

  Underlying data is updated, incorporating a new diagonal, or several new diagonals, of paid and incurred information as well as updated information on the distribution of claimant ages and genders on open claim. For this exercise, we lock down the parameters that generated the original loss development curve (specifically parameters *a, b,* and *r,* from equation 2.4 above). We then run through each step of section 2.3.1. with updated claims data. The resulting indications will then be driven entirely by underlying data changes and not changes to the model's parameters.

- Exercise 2: Re-Fitting the Parameters

  In a separate analysis, after updating the underlying data per Exercise 1, we go through the entire process as outlined in Section 2, above, from selecting a curve to the application of mortality. This should essentially be starting from scratch to determine the best model for the data without looking back at the prior results. In this way, the practitioner will have the impacts from updating the underlying data as well as the impacts from updating the model parameters.

For purposes of reducing bias, the practicing actuary should determine ahead of time how long the parameter assumptions will be locked. We believe a time period of three to five years is a reasonable starting point. This locking of the parameters insulates against responding to the year-to-year process noise introduced when the underlying data is updated and the curve fit parameters are refit.

Of course, locking in parameters without monitoring changes in the underlying data would expose one to missing changes in underlying trends or material shifts in development patterns. For this reason a determination should be made on a threshold to be used to measure whether or not parameters should be re-fit earlier than the predetermined timeframe. This threshold would be measured against

[3] We credit Michael Shane, FCAS for this suggestion.

the parameter refit impact, calculated as the total variance from Exercise 2, above, less the process variance from Exercise 1 above. Suggested benchmarks might include a percentage of capital or surplus, or perhaps a percentage of the underlying reserves being modeled. Additionally, if the model has been used for multiple years, this parameter refit impact should be monitored over time, since consistent directional impacts might indicate the need for a refit sooner than anticipated. In theory, one would expect these parameter refit impacts to oscillate around zero if no systemic change is taking place.

Alternatively, one could use a goodness of fit test with the new data and the locked parameters. One other option may be to use a Bayesian framework to assess the new parameters; given a prior assumption, test whether the new information suggests that the prior is no longer valid. The specific test to be used is left to the reader. Again, the determination should be made beforehand around the tolerance around the results of this test and the level at which a decision needs to be made on whether or not to refit the parameters. When the decision is made to refit the parameters, the model fit guidelines presented in Section 2 should again be utilized. The parameter locking process should also be essentially reset at the same time.

The authors would like to clarify that this approach assumes a single and consistent mortality table. In practice, there is likely more than one mortality table that may be used. We leave it to the reader to select the most appropriate mortality table for the purposes of this model. Discussion around which mortality table(s) to utilize are beyond scope of this paper. Further, note that if the underlying mortality table that the practitioner had decided to use for the prior iteration of this model was updated or changed, the practitioner should consider utilizing the most up-to-date mortality table in the next iteration of this model.

3.2 DISCUSSION OF RESULTS WITH BUSINESS PARTNERS

When the data is updated and new indications are generated, the discussion of results with business partners should be fairly intuitive. In theory, the drivers of changes to indications should be directly related to changes in open claim attributes, such as the distribution of ages and genders, or changes to the run rate of payments (e.g., escalating medical severity). This is one of the key benefits of the original Shane-Fowle model, since the resulting changes to indicated ultimates are often fairly obvious and reasonably easy to explain.

When the data and the parameters are both updated, the discussions could be more difficult. The completion of both Exercises 1 and 2, above, allow for the same insights as noted above that relate to changes in the open claims, leaving the actuary to explain the remaining changes as related to the updating of the parameters. If the parameters were updated due to shifts in the underlying data leading to parameter refit variances greater than the selected threshold, it is likely that specific internal or external drivers could be identified that explain the need to refit the parameters and the directional change to indications. These drivers might include:

- Internal – changes to the claim settlement practices, changes to case reserving standards, new cost or expense mitigation efforts, etc.
- External – changes to medical inflation, changes due to state specific reforms, changes to the legal environment due to court rulings, changes to life expectancies, etc.

When parameters are updated due to the predetermined passage of time, such specific drivers may be more difficult to find, but the impact of the parameter update on the overall indications is also likely to be less significant in this case.

4. CONCLUSIONS

The Shane-Fowle model described an intuitive method for incorporating mortality assumptions into otherwise standard actuarial methods to improve the resulting reserve estimates and offer insights for discussion with management and other business partners. This model extends and enhances the Shane-Fowle method in two ways. We introduce the use of the gamma curve, which results in a tail that is fit based on the underlying data with less subjectivity than the more traditional inverse power or exponential curves. Additionally, we propose a gradual application of mortality to the curve fit, which allows for the projection of incremental incurred or paid losses to converge to zero as expected, which is not achieved by either a straight curve fit or the truncation method originally proposed.

This paper has also included a proposal for the regular updating of the model that should facilitate discussion with business partners about the underlying cause of changes to estimates. This recommended approach cautions against modifying parameters overly often, proposing instead that parameters are updated either at predetermined intervals or when there are truly significant changes to the data that suggest the previous model might no longer be appropriate.

Acknowledgment

The authors gratefully acknowledge the generous instruction and support by Michael Shane, FCAS. We also thank Robert Ballmer, FCAS for his significant contributions to this work.

Supplementary Material

Included with this paper is a supplemental Excel tool titled WC Tail Model Template.

5. REFERENCES

[1] CAS Tail Factor Working Party, "The Estimation of Loss Development Tail Factors: A Summary Report," *Casualty Actuarial Society E-Forum*, **Fall 2013**, 1-111.

[2] Shane, Michael, and Dawn (Morelli) Fowle, "Using Life Expectancy to Inform the Estimate of Tail Factors for Workers' Compensation Liabilities," *Casualty Actuarial Society E-Forum,* **Fall 2013**, 1-12.

[3] Sherman, Richard E., "Extrapolating, Smoothing, and Interpolating Development Factors," *Proceedings of the Casualty Actuarial Society*, **1984**, Vol. LXXI, 122-155.

[4] McClenahan, Charles L, "A Mathematical Model for Loss Reserve Analysis," *Proceedings of the Casualty Actuarial Society*, **1975**, Vol. LXII, 134-153.

Biographies of the Authors

Shon Yim, ACAS is an actuarial consultant at CAN, where he has worked on reserving analytics and Long Term Care Insurance pricing. He received a doctoral degree in Mechanical Engineering from MIT. Prior to joining CNA, he worked for a commodity trading advisor, focusing on research and development of technical trading systems.

Dolph Zielinski, FCAS is an AVP and Actuary at CNA with over eight years of reserving experience across multiple lines, including Workers' Compensation. He received his master's degree in Applied Mathematics at Roosevelt University.

Dawn Fowle, FCAS is a senior actuarial consultant at Ernst & Young LLP (EY), where she provides reserving support for numerous clients. Prior to joining EY, Dawn worked for both ongoing and runoff insurers, with significant exposure to Workers' Compensation.

Disclaimers

Appendix A

In this appendix we derive a closed form expression for $v(d)$, the development portion of the link ratio. Link ratios for aggregate triangles are typically evaluated over multiple accident years, but this derivation will focus on the development in only a single accident year. Our analysis will impose only one condition: the incremental losses will decay with development age at a constant rate, β. We will derive the formula here on a discrete basis, but the continuous analogue is straightforward.

We denote the cumulative loss as of age d as $c(d)$. The incremental loss from age $d-1$ to d is $q(d)$. In equation form, $q(d) = c(d) - c(d-1)$. The link ratio, also known as the age-to-age loss development factor, is defined as $f(d) = c(d+1)/c(d)$. We are interested in the development ratio $v(d)$, which is equal to $f(d) - 1$. It can be shown easily that

$$v(d) = q(d+1)/c(d) \tag{A1}$$

Next we define the incremental loss decay rate β in equation form:

$$= \frac{q(d) - q(d+1)}{q(d)} = 1 - \frac{(d+1)}{q(d)} \tag{A2}$$

Recall that the decay rate is constant, which means that the incremental loss decreases by the same percentage of the previous incremental loss, regardless of the age. Equation A2 can be rearranged to yield the following expression:

$$(d+1) = (1-\beta)q(d) = (1-\beta)^d q(1), \tag{A3}$$

where the last equality comes from a recursive relation made possible by the assumption that β is constant. Next we recognize that $c(d)$ is simply the sum of q's:

$$c(d) = \quad q(k) = \quad (k+1) = \sum (1-\beta)^k q(1), \tag{A4}$$

where the last equality utilizes Equation A3. The right side of Equation A4 is the geometric series with the well-known solution

$$c(d) = q(1)\left[\frac{1-(1-\beta)}{} \right. . \tag{A5}$$

We next combine Equation A1 and A3 to get

$$v(d) = \frac{(1-\beta) \quad (1)}{c(d)} \ , \tag{A6}$$

and substituting Equation A5 for $c(d)$,

$$v(d) = \frac{(1-\beta) \quad (1)}{(1) \quad \frac{1-(1-\beta)^d}{}\Big]} \ . \tag{A7}$$

Finally, cancelling the $q(1)$ terms and rearranging yields the formula

$$v(d) = \frac{\beta(1-\beta)}{1-(1-\beta)} \ . \tag{A8}$$

Appendix B

In this appendix, we examine the inverse power and exponential curve fits for the development ratio of a loss pattern where the incremental losses are subject to a constant rate of decay. We are specifically interested in how the quality of fit responds as the decay rate changes. We conduct the analysis using a simple hypothetical loss development for a single accident year in which incremental losses start at 1,000 at age one and decrease annually by a decay rate of β. For example, a decay rate of 10% produces the following incremental and cumulative loss patterns. Note that we are only looking at development past the first ten years in keeping with the lifetime scope as identified in Section 1.2.

Development Age,	Incremental Loss, $q(d)$	Cumulative Loss, $c(d)$	Development Ratio, $v(d)$
10	387	6,513	0.054
11	349	6,862	0.046
12	314	7,176	0.039
13	282	7,458	0.034
14	254	7,712	0.030
15	229	7,941	0.026

Decay rate β = 0.1

Next, the development ratio as a function of the development age was fit to the inverse power curve,

$$\hat{v}(d) = ad^b, \tag{B1}$$

where a is the scale parameter. The more interesting parameter is b, which governs the rate of decay of the inverse power curve. The fitting was done by regression in log space using only data points for development years 10-29. The same data was also fit to the exponential curve,

$$\hat{v}(d) = \qquad . \tag{B2}$$

Parameter r controls the decay of the exponential curve. The fits to the inverse power and exponential curve were conducted for a range of decay rates. The quality of fit was measured with the R^2 metric in log space, but it was also confirmed that measuring R^2 in the original dimensions yields similar conclusions. The results are summarized below.

Decay Rate,	Inverse Power		R^2	Exponential		R^2
0.00	1.000	-1.000	1.000	0.155	-0.054	0.980
0.05	2.716	-1.537	0.997	0.157	-0.084	0.993
0.07	4.519	-1.801	0.995	0.161	-0.099	0.995
0.10	10.849	-2.248	0.991	0.170	-0.124	0.998
0.15	60.884	-3.116	0.986	0.193	-0.172	0.999
0.20	447.933	-4.119	0.983	0.226	-0.228	1.000

At zero decay rate, the inverse power curve provides a perfect fit with $b = -1$. The response to an increasing decay rate is that becomes more negative in an attempt to keep up with the faster decay, but this comes at the cost of the goodness of fit, as indicated by the decreasing R^2 value. On the other side of the table, it can be seen that the exponential is an inferior fit to the inverse power at a decay rate of zero, judging by the R^2 value. But the exponential's fit improves as the decay rate increases to the extent that above a decay rate of 7%, the exponential is better than the inverse power.

Appendix C

This appendix illustrates the projection of future losses using the gradual application of mortality as described in Section 2.3. This example uses hypothetical but realistic paid loss and claimant data. Suppose that we have the following information for an accident year:

- Cumulative loss at latest diagonal is 10,000.
- Development age at latest diagonal is 20 years.
- The development ratio has been fit to the curve in Equation 2.4, and the parameters have been estimated to be: $= 0.4,\quad = -1.37,\ r = -0.00165$.

The table below shows the calculations for this accident year. Note that some figures are rounded for display.

Proj Year [A]	Avg Age $\text{age}_G(t)$ [B]	Mort Rate [C]	Dev Age [D]	Fitted LDF $\hat{f}(d)$ [E]	Cumul Loss $c(d)$ [F]	Increm Loss [G]	Loss Decay Rate [H]	Adj Loss Decay [K]	Adj Increm Loss [L]
1	60.6	0.0056	20	1.0064	10,000	64			64
2	61.5	0.0062	21	1.0060	10,064	60	0.060	0.061	60
3	62.4	0.0068	22	1.0056	10,124	57	0.058	0.059	56
4	63.4	0.0075	23	1.0052	10,180	53	0.055	0.057	53
5	64.3	0.0082	24	1.0049	10,234	51	0.053	0.056	50
6	65.2	0.0088	25	1.0047	10,284	48	0.051	0.055	48
7	66.1	0.0094	26	1.0044	10,332	46	0.049	0.053	45
45	94.0	0.1695	64	1.0012	11,275	14	0.022	0.186	2
46	94.6	0.1872	65	1.0012	11,288	13	0.022	0.203	1
47	95.3	0.2039	66	1.0012	11,302	13	0.021	0.219	1
48	95.9	0.2039	67	1.0011	11,315	13	0.021	0.219	1
49	96.5	0.2204	68	1.0011	11,328	13	0.021	0.235	1
50	97.1	0.2394	69	1.0011	11,340	12	0.020	0.254	0

[A] Projection period in years.

[B] Projected average age of the group at the start of year t, calculated with Equation 2.5. This calculation requires a selected life table and the distribution of claimant ages.

[C] Mortality rate looked up from the life table using the age from column [B]. This represents the probability that a person will die within the next year.

[D] Development age at the start of year t in years, starting with age at the latest diagonal.

[E] = 1 + 0.4[D]$^{-1.37}$exp(-0.00165[D]) is the fitted link ratio using the given fit parameters with Equation 2.4.

[F] = [F previous]*[E previous] is the cumulative loss at the start of year t, obtained from the curve fit. First row is the latest diagonal.

[G] = [F next] – [F] is the incremental loss predicted for projection year t.

[H] = 1 – [G]/[G previous] is the implied loss decay rate between year $t-1$ and year t. Note that there will be discrepancies due to rounding in the table above.

[K] = [H] + ([C] – 0.0056) is the implied loss decay rate adjusted for mortality. The 0.0056 value is the latest observed mortality rate (first row of [C]), which is removed from the adjustment to avoid double counting mortality.

[L] = [L previous]*(1 – [K]) is the adjusted incremental loss.

The sum of column [L] is equal to 1,047. This is the total unpaid estimate for the accident year.

Appendix D

This appendix provides a comparison of reserving estimates using three different techniques: the enhanced Shane-Fowle method presented in this paper, the original Shane-Fowle method and a traditional method. The three methods were applied to example workers' compensation medical paid data from industry sources. All three methods utilized the same LDFs, which were five-year volume weighted averages excluding the first ten development years. The modeling distinctions among the three methods are as follows.

Enhanced Shane-Fowle: the gamma curve (Equation 2.4) was fit to the LDFs, and mortality was gradually applied as described in Section 2.3.

Original Shane-Fowle: the inverse power curve was used to fit the LDFs, and the curve was truncated at the life expectancy according to the average age of the accident year cohort.

Traditional: LDFs were used without curve fitting, and a tail factor was selected based on the incurred/paid ratio at the end of the triangle.

The table below shows the cumulative development factors (CDFs) and unpaid estimates using the three methods.

Acc Year	Cumul. Paid ($M)	Enhanced Shane-Fowle		Shane-Fowle		Traditional	
		CDF	Unpaid ($M)	CDF	Unpaid ($M)	CDF	Unpaid ($M)
1985	984	1.031	31	1.031	30	1.030	30
1986	1,133	1.035	40	1.034	38	1.033	37
1987	1,326	1.039	51	1.037	49	1.037	49
1988	1,530	1.041	63	1.041	62	1.040	61
1989	1,784	1.045	80	1.045	79	1.044	78
1990	2,029	1.048	98	1.047	96	1.048	97
1991	2,180	1.053	115	1.053	116	1.051	112
1992	1,741	1.058	101	1.057	99	1.055	97
1993	1,491	1.062	93	1.062	93	1.060	89
1994	1,449	1.068	99	1.068	98	1.065	95
1995	1,591	1.077	122	1.074	118	1.072	115
1996	1,681	1.081	136	1.077	130	1.081	136
1997	1,975	1.091	180	1.087	171	1.091	179
1998	2,585	1.100	258	1.094	243	1.101	261
1999	2,963	1.110	325	1.106	313	1.113	334
2000	3,486	1.121	422	1.114	396	1.127	442
2001	5,225	1.132	691	1.125	652	1.142	741
2002	5,342	1.149	798	1.141	756	1.158	842
2003	4,900	1.166	815	1.159	778	1.175	856
2004	3,915	1.184	720	1.177	695	1.194	760
2005	3,505	1.207	727	1.204	715	1.217	762
Total			5,967		5,727		6,171

In comparing the enhanced and the original Shane-Fowle estimates, it is interesting to recognize that the two enhancements made to the Shane-Fowle method had opposite effects on the unpaid estimate. The first enhancement of using the gamma curve in lieu of the inverse power lowered the estimate of future development because the exponential component of the gamma made the tail thinner. The second enhancement of the gradual application of mortality increased the estimate of development by projecting a longer life expectancy for the group.

Here we see an example where the three methods produce results that are not materially different from each other. This will not always be the case. For example if the gamma fit is very close to the inverse power, the first enhancement would not have much impact. If at the same time there is a wide distribution of claimant ages, the second enhancement would make a significant difference. The combined effect would be an overall material difference.