Casualty Actuarial Society
E-Forum, Summer 2020

The CAS *E-Forum*, Summer 2020

The Summer 2020 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 three submissions in response to a call for papers on reserving issued by the CAS Reserves Committee (CASCOR). Also included are five independent research papers.

Committee on Reserving

Julie Ann Lederer, *Chairperson*

John P. Alltop
Denise M. Ambrogio
Diego Fernando Antonio
Rachel C. Dolsky
Adam Michael Gerdes
Karl Goring
Isabelle Guerard
Ziyi Jiao
William J. Lakins
Thomas James Marshall
Martin Menard
Kelly L. Moore
Chandrakant C. Patel
Marc B. Pearl
Stefan Joseph Peterson
Christopher James Platania
Camilo Andres Rodriguez
Ryan P. Royce
Ernest I. Wilson
Brian A. Fannin, *Staff Liaison*

CAS *E-Forum*, Summer 2020

Table of Contents

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

Introduction to Data Visualization

John Deacon, Annie Fan, Brian Fannin, Jennifer Levine, Keith Quigley, Patrick Yu

Abstract: This paper summarizes some of the literature on the topic of basic data visualization techniques. We emphasize the importance of knowing the audience, and focusing on what message is intended to be sent. We provide visual examples of graph types and describe when to use the different types for different situations. We identify several decluttering and accentuating techniques and we share some of the basic research on how the human eye and brain work to interpret visual information. We provide a before-and-after example of the basic data visualization techniques, to show how much improvement can be achieved in delivering the intended message.

keywords: data visualization, communication, gestalt principles

1. Definition of "Data Visualization"

Visualizations are everywhere. Data and information are becoming more accessible all the time. To absorb value from all this data, we may leverage the power of "data visualization". Simply: **data visualization = data represented by a visual image**. Some of the most common and basic data visualization applications include bar graphs, line graphs, and scatterplots. The general design concepts in this publication apply to any type of data visualization in any medium, e.g. business software such as Excel or PowePpoint, programming languages, data visualization software like Tableau, and websites.

1.1. Why Do We Need Data Visualization?

With the explosion of data and tools to analyze it, we need skills to **communicate the message most effectively**. By applying proven techniques, you can create data visualizations that maximize your audience's absorption of your intended message(s).

Actuaries excel at analyzing data and drawing conclusions from their work. However, visual communication skills (i.e., how to make effective graphs, charts, tables, etc.) are not represented on the CAS exam syllabus, nor are they regularly taught or reinforced. Nevertheless, actuaries must *regularly* communicate complex patterns, trends, conclusions, ideas, and concepts with each other or non-actuarial business partners. With some knowledge and practice, actuaries can create effective data visualization.

We can do the best work with the best data resulting in the best conclusion, but if we can't communicate it effectively, the message is **lost**.

Many authors have produced books, papers, websites, and blogs on data visualization. We will summarize their general design concepts in this paper. The approach we will take is as follows:

- First, we will focus on the **audience**, those individuals with whom we will be communicating.
- Next, we will review some of the basic **visualization types and how to choose** from among them based on your data and your needs.
- Finally, we will share thoughts about **how to design a visualization** for maximum impact.

2. Consider the audience and the message

When we create data visualizations to present quantitative information, instead of jumping straight to the designing stage, we should first understand "the context for the need to communicate" (Knaflic 2015). The context can be split into two parts: the target audience and the intended message.

2.1. Who is the audience?

If an image is displayed in an empty room, is it ever seen? An audience is a critical component in the creation of a data visualization. Without them, the image does not really exist at all. By the same token, a different audience may require a different sort of visual representation. Without a reasonable consideration of the consumers of our work, it may as well not exist. To properly consider our audience, we ask some basic questions.

2.1.1. What is their role? What is their background?

Actuaries can use different kinds of visualizations to communicate findings and recommendations to various groups such as other actuaries, regulators, underwriters, brokers, executives and many more. A common pitfall is that we tend to generalize the target audience group to "internal and external stakeholders" or "anyone who may be interested" (Knaflic 2015). These generic ("internal or external") groups include audiences with different roles and backgrounds.

We must tailor the visualizations to a specific audience group to maximize the effectiveness of the message. To narrow the target audience group, Knaflic recommends identifying the decision maker. Ask yourself: "What information and what considerations drive the decision-maker"? Asking the question: "Who is the audience?" may also determine whether the presentation should be live/interactive or whether it is appropriate to send the audience the visualizations to view on their own.

2.1.2. How technical are they? What do they already know?

After defining the target audience group, the presenter should know the audience's technical knowledge level. Without careful attention to data visualization, actuarial analyses can be too technical and difficult for non-actuarial groups to comprehend fully, especially with the large collection of actuarial acronyms, terms, and methods. When designing visualizations for a non-actuarial audience, consider putting yourself in their shoes by asking yourself "What is their perspective?" and "What do they know?". Even technical audiences, like actuaries, will understand and absorb the complex information more effectively if it is presented simply.

2.2. What is the message?

When crafting visuals, a content creator should continually ask themselves and fellow workers **"What is the message?"**. Asking this question and the other related questions below helps refine the content, complexity, and nature of visual presentations.

2.2.1. What does the audience need know to influence the decision?

After we understand the target audience's role and technical background, we should think about how to make the visualization relevant for the audience. In other words, what information needs to be included in the visualization to help the audience make decisions. Knaflic clarifies an important distinction between *exploratory* analysis and *explanatory* analysis. Exploratory analysis refers to the preliminary work we do "to understand the data and figure out what might be noteworthy or interesting to highlight to others" (Knaflic 2015), and explanatory analysis refers to the "specific thing

you want to explain" or the "specific story you want to tell" (Knaflic 2015). We should focus on the **explanatory** aspect of our work in the visualizations to influence the audience's decision. For example, an underwriter may prefer to see how a change in the selected tail factor affects the bottom line loss ratio, whereas another actuary may want to see all the indicated tail factors underlying the change in selected tail factor. Think about what impacts your audience and motivates their decisions.

2.2.2. How much detail do they really need? How much will suit them?

According to Knaflic, the communication mechanism determines how much control the presenter has on the way the target audience consumes the information, which in turn determines the level of detail a visualization needs (Knaflic 2015). For example, during a live presentation, the presenter has the most control over the flow of the storytelling process. If the audience has a question about a particular point, the presenter can answer right away and provide more details in person, so the **visualizations in the presentation deck do not need to be over-filled with details**. However, for documents viewed independently by the audience, the author has less control over when and how the audience will interpret the visualization. In this situation, more consideration and care is needed to create visualizations that communicate the intended message. Our tendency may be to provide more detail, but that may confuse, rather than clarify the message.

2.2.3. What is the audience supposed to DO with the information?

It's important to remember that after we inform the audience about key insights on a topic, we should guide them to take action(s) to resolve it. We can explicitly state the next steps, provide recommendations, or encourage discussions for situations where the next step is unclear (Knaflic 2015). It is reasonable to include an action statement *within* the title or visual.

3. The Visualization Framework

Having considered our audience, let's take a moment to talk about how data visualization actually comes into being. Most actuaries have generated a time series or bar chart and may find the process fairly straightforward. It is, but this simplicity belies the mathematical structure and set of decisions that underly the mapping of data to image.

3.1. The Tangibility of Data

Because we interact with it so often, it may seem as though "data" is tangible. This sense comes from the way we interact with it: a visual representation of data. This image is generated from the photons of light emitted by your laptop, the bits of ink on a sheet of paper, or the electric charge stored in a computer's memory. But these things are not actually the real-world elements, they are simply a representation of them. A "policyholder" is not a record in a database and would likely object to being referred to as such!

Consider Table 1. This shows the first ten (of five thousand) records of a list of policyholders. Note that this is only a subset of the potential information which we might display. There could be claims, an address, demographics, marketing touch points, or any number of other characteristics. The data is no less real even though we choose to display a subset of it.

Table 1: The first ten policyholders

territory	policyholder	effective_date	expiration_date	premium
Athens	Reyes	2001-08-15	2002-08-15	11,561
Athens	Villagomez	2001-07-21	2002-07-21	7,578
Sparta	al-Abdullah	2001-12-20	2002-12-20	10,294
Sparta	Jefferson	2001-12-17	2002-12-17	9,133
Sparta	el-Mina	2001-05-22	2002-05-22	9,812
Athens	Hamlin	2001-02-15	2002-02-15	10,702
Sparta	Hitchcock	2001-09-26	2002-09-26	8,917
Athens	Lease	2001-12-27	2002-12-27	10,743
Sparta	Rivera	2001-04-21	2002-04-21	8,620
Athens	al-Sami	2001-04-11	2002-04-11	8,554

We are accustomed to thinking of the set of recorded properties of real-world entities - like what is shown in Table 1 as being the true, accurate "data". Free yourself of that notion. Once you have taken that step, it will be easier to accept the authenticity of visualization as tangible "data".

3.2. The visual space

When we create a visualization, we have a fixed space in which to place our data. It will be most useful to think of it as a rectangle, though this is not strictly required. Indeed, spherical coordinates, or certain mapping projections will result in non-rectangular plotting spaces. However, most visualizations are expressed sthis way.

A rectangular space gives us two dimensions to work with. That is it, just two. We can introduce additional information via attributes like color and shape, but there are only two physical dimensions. Returning to our policy data, we can see one possible representation of the dimensions of the visual space in Figure 1. There is nothing within the space yet; we must begin by defining its size and scale.

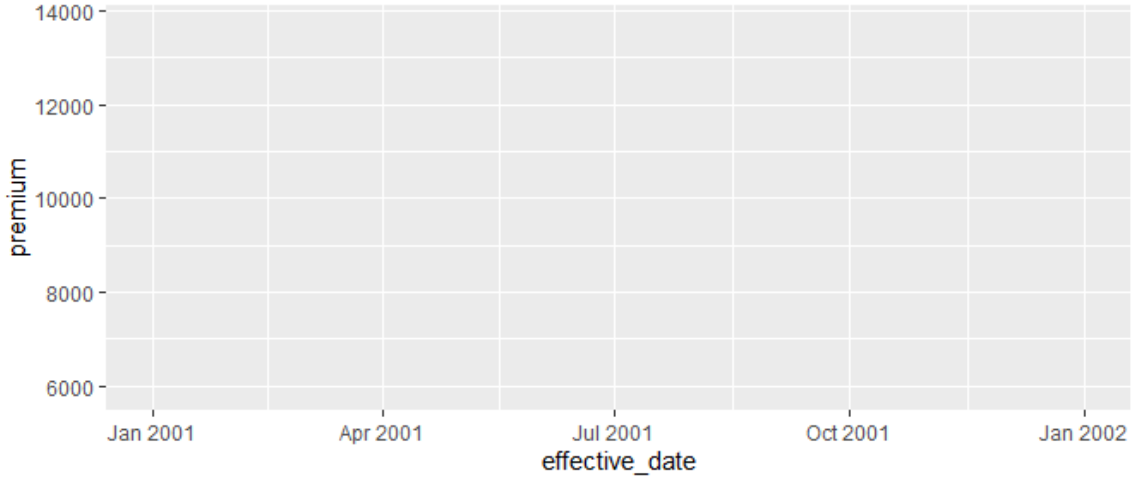


Figure 1: An empty visual space

We have naturally assumed that time should run along the horizontal, or x axis, and that premium should run along the vertical, or y axis. There is no requirement that we do so; it is simply a choice that we have made. Just because it is sensible should not lead us to think that we are adhering to any obvious rule. It works, but we would not be violating any natural law by transposing the axes, as shown in Figure 2.

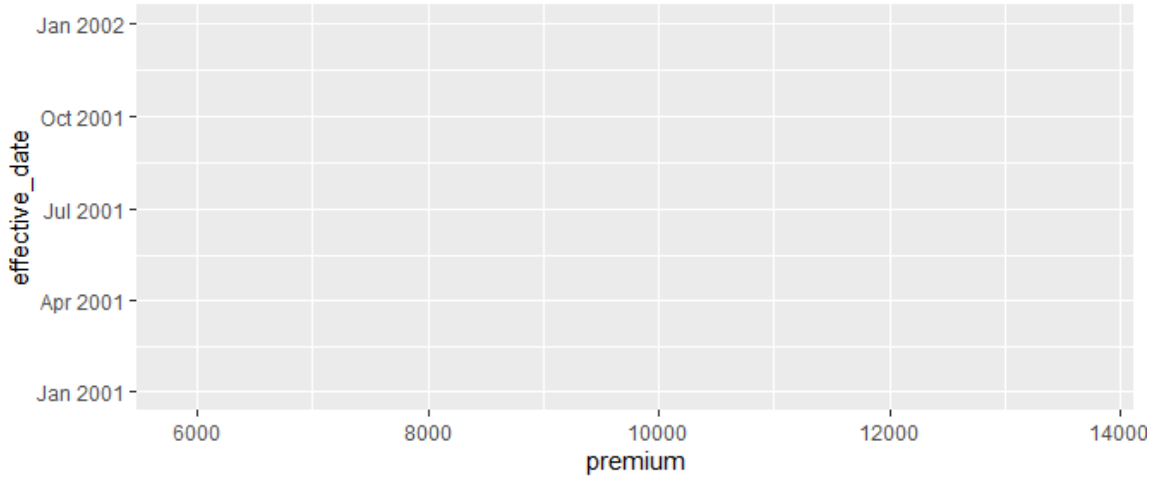


Figure 2: Transposing the axes

Take a moment and compare the linear distance available for the effective date in Figures 1 and 2. Resist the temptation to think of that distance in temporal terms, and think only in physical terms. That is, think of distance as inches, centimeters, or pixels. There is more space in Figure 1 than there is in Figure 2 - our rectangle is longer than it is wide. However, there is a precisely equal amount of *temporal space*. We are able to do this by relying on the computer's ability to scale the range of our data to the range of the axis.

3.3. Categorical axes

Physical distance is generally reckoned as coming from the set of real numbers. However, our plotting space is not bound by this rule. A dimension may be categorical - something which maps to the set of natural numbers - instead. The policy effective month is an obvious candidate for a categorial axis.

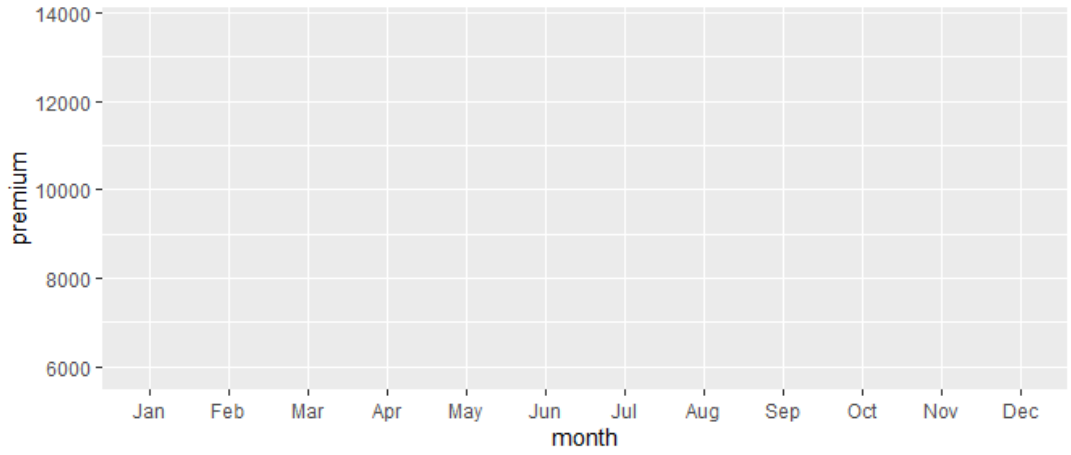


Figure 3: Categorical data along the x-axis

There is often a natural preference for orderings categories, but this, too, is something which the analyst must choose. In Figure 3, we arranged the months in date order, but we could have chosen something different. In Figure 4, the months have been randomly sorted.

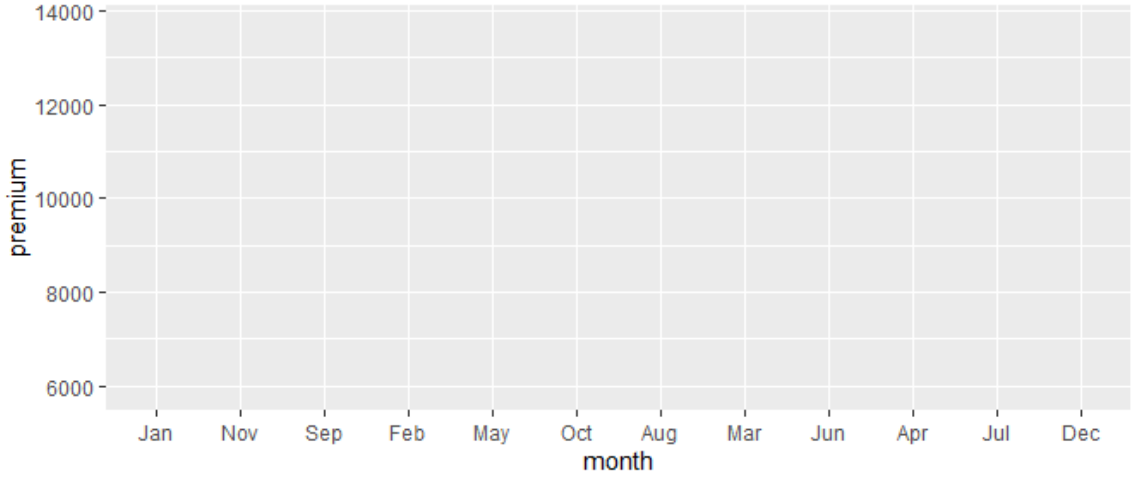


Figure 4: Months arranged randomly

We could just as well have chosen territory as one of the axes. Consider what information could be shown in Figure 5.

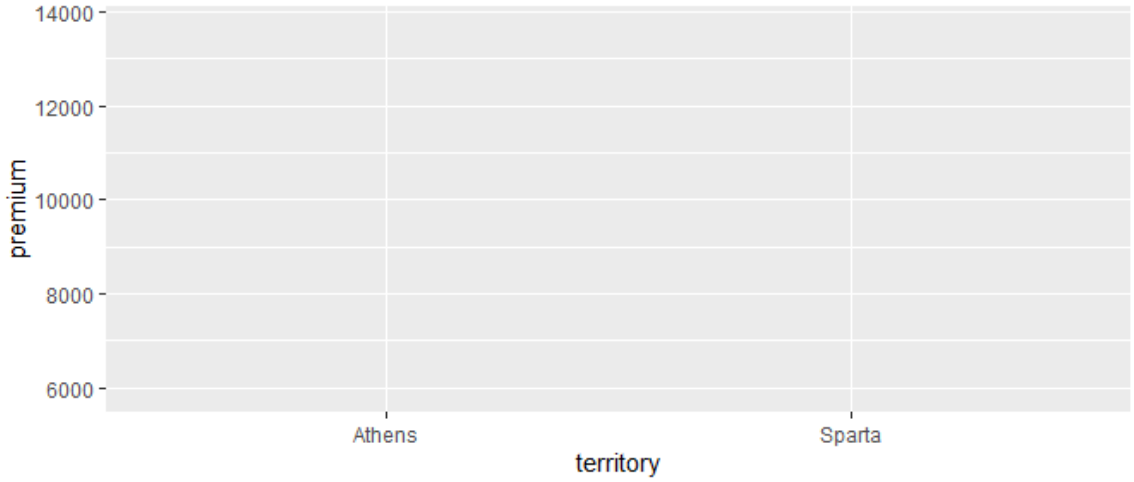


Figure 5: Another option for categorical data on the x-axis

3.4. Mapping

The figures so far lack an obvious element: nothing appears in the plotting space. The process of expressing data visually requires a mapping of data to a visual element. We have already mapped two elements: the policy effective date (or month) and premium. In this case, we need do nothing more than inform our visual rendering engine what geometric shape to place in the plotting space. In Figure 6, we see the data represented as points. Each point corresponds to a single policyholder.

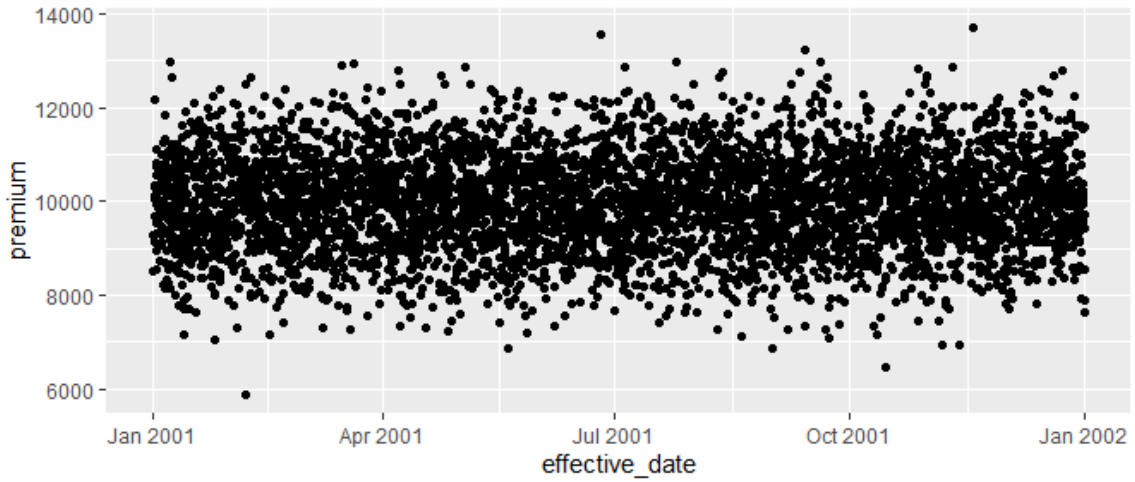


Figure 6: A geometric shape as tangible data

Figure 6 contains *more* information than Table 1, yet consumes less physical space. One might argue that Table 1 contains more detail and this is true, to a point. It is valid to say that Table 1 allows one to identify the specific premium and policy effective date for the first ten policyholders. If that is what the audience wants, you have succeeded in a way that Figure 6 has not. However, as we learned in Section 2, it is unlikely that is what your audience wants. What Figure 6 does is answer the following questions:

- Is the business growing?
- Is the business contracting?
- Is there any seasonality to our revenue?
- What is the average size of premium?
- Are there any policies which are particularly large or small?

We can use the same amount of visual space to introduce more information. As mentioned above, the two axis dimensions are already being used, but we could map a data element to color or shape. We see an example of this in Figure 7.

Figure 7: Data represented via color

In this section, we have shown some of the mechanics of bringing a data visualization into being. We have also shown the result of one particular set of choices. In the next section, we will explore more options for visual representation and consider how to choose among them.

4. How to Choose the Type of Data Visualization

Now that we have established the audience and the message, it is time to design the visualizations. This section will help you identify the best forms for the data and message.

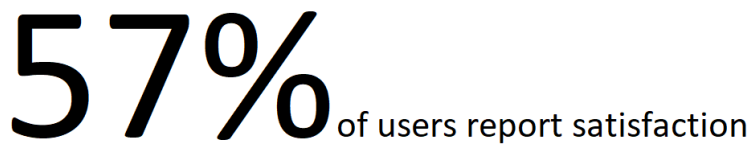
4.1. Determine the Best Type of Visualization to Use

Authors such as (Yau 2011) have described how to select the particular graph type that fits the data and message. Using the guide below, you can then experiment and decide which type of visualization to use. Remember that the ultimate judge of the appropriateness of a visual is: will the audience understand and act on the message? Repeatedly ask yourself: "What is the message?", and decide whether it is clearly delivered. Try creating multiple visualizations for that message and ask a trusted colleague for feedback about which visualization works best for the intended message.

4.2. Examples of the Main Types of Visuals and Key Uses

4.2.1. Text

There is no need to introduce geometric abstraction, when only a few key figures are involved. Simple text will suffice when you have one or two numbers to show.



4.2.2. Tables

Similarly, a simple table may be ideal when the data is simple and the user is interested in looking up precise values.

TABLE

	X	Y	Z
Category I	10%	22%	30%
Category II	12%	10%	15%
Category III	5%	12%	10%
Category IV	3%	10%	5%
Category V	15%	15%	15%

Visualization may be introduced to communicate larger messages. For example, one may add color to the table. The hue is mapped to particular attributes of the data in order to emphasize relative importance — low and high values — or some other message. Because of their similarity to weather visualizatoin, these exhibits are commonly referred to as "heatmaps".

HEATMAP

	X	Y	Z
Category I	10%	22%	30%
Category II	12%	10%	15%
Category III	5%	12%	10%
Category IV	3%	10%	5%
Category V	15%	15%	15%

4.2.3. Line

Line graphs imply continuous data or a connection between the points. Line graphs can be augmented with a shaded range or confidence interval around the observations, or a model fit to the data. A basic line graph, like the one shown in Figure 8 are best used with time series and trends. They are also well suited to display of a deviation relationship such as a difference from plan. In a smoothed form, line graphs are familiar to actuarial students as expressing probability distributions.

LINE

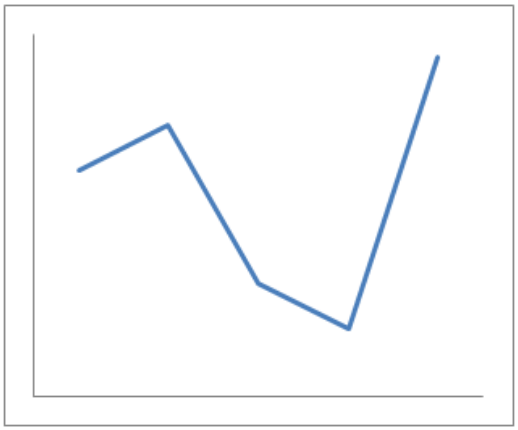


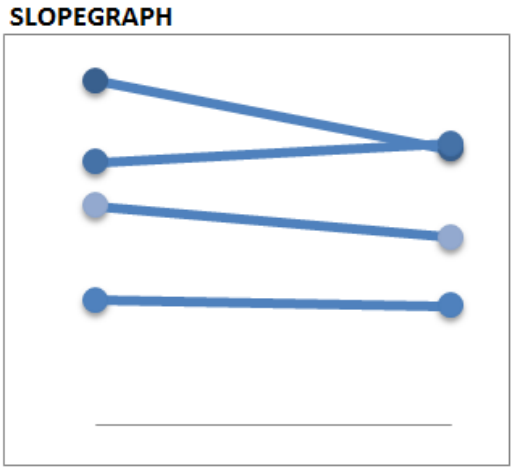
Figure 8: A standard line graph

Sparklines are line graphs which are presented in a minimalist fashion, often included as a column in a table of other information, as shown in Figure 9. This can be valuable for seeing trends at-a-glance or highlighting minimum and maximum values.

Reg A	Reg B	Reg C	Change
10	6	7	
6	9	10	
3	5	9	
9	7	4	
2	3	3	
5	5	6	

Figure 9: A data table with sparklines

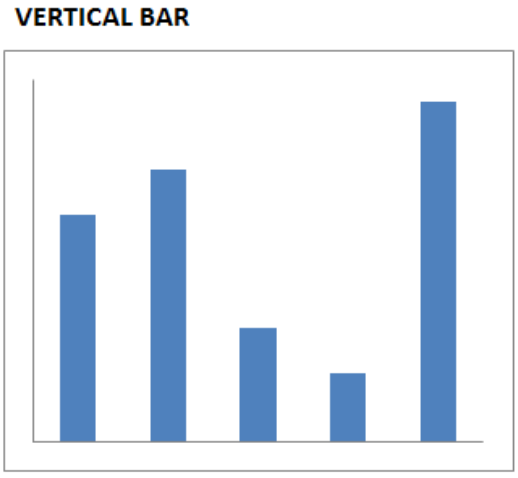
A slopegraph is composed by mapping categorical data to the x-axis. The categories may originate from something quantitative like a point in time, or they could relate to regions, or rating classes. There are typically only two categories and rarely more than three. Slopegraphs can be useful when showing data for just two time periods and when you want to easily show relative decreases or increases among several categories. These graphs can get cluttered, and should be avoided when there are too many overlapping lines.



4.2.4. Bar

Similar to a slopegraph, a bar chart also maps categorical data to the x-axis. However, the treatment is such that more than a few columns may be displayed.

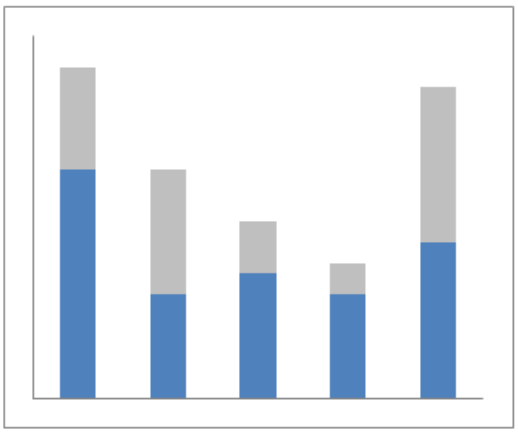
Vertical bar charts are useful for time series graphs where you want to focus on the comparison between values of individual points, rather than on the overall pattern of values over time. Vertical bar charts can also be useful in showing deviation relationships (e.g., difference from plan) at a point in time.



Additional categorical data may be introduced with a stacked vertical bar chart. Use caution in using stacked vertical bar charts, as it can be hard to compare sizes or values, especially if the baseline for a

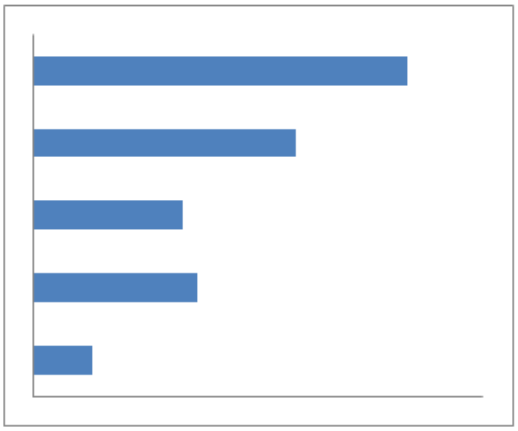
given 'stack' is not the same. Thus, stacked vertical bar charts are best if there is a key category of focus, which should be placed as the bottom set of bars. Consider using absolute numbers in stacked vertical bar charts or using stacked vertical bars that each sum to 100%. Avoid using more than two or three sub-categories within bars, as this may obscure the intended message.

STACKED VERTICAL BAR



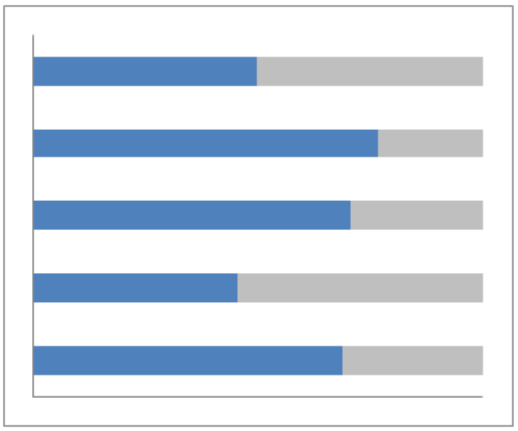
Horizontal bar charts are great for displaying categorical data, especially if using the categories to rank the values. Horizontal bars are also useful for long category names that won't fit as well for a vertical bar format.

HORIZONTAL BAR



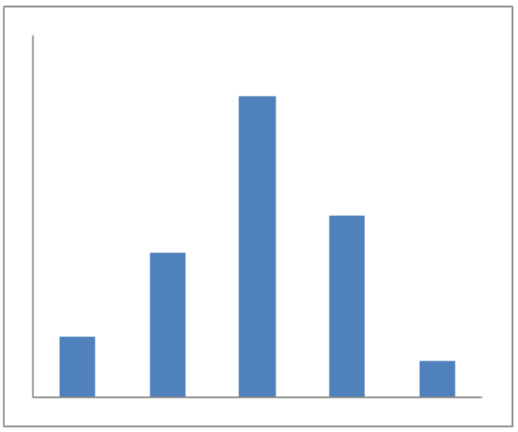
As with vertical bar charts, a horizontal bar chart may use stacking to encode categorical data along a different (data) dimension.

STACKED HORIZONTAL BAR



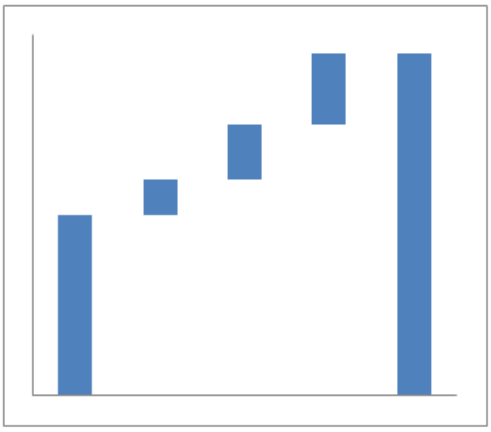
A histogram is a vertical bar chart used to display an empirical, or sample distribution.

HISTOGRAM



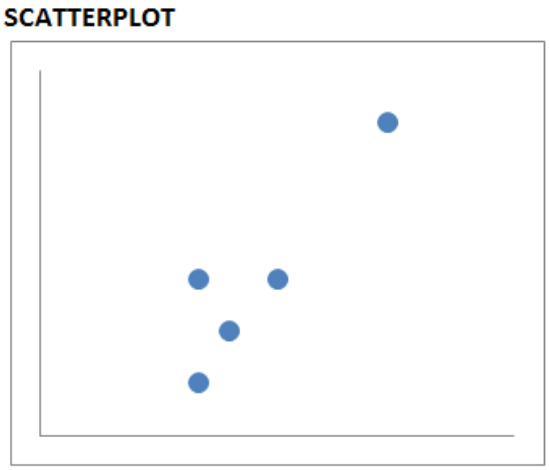
Waterfall charts are useful to show a starting point, incremental increases and/or decreases, and an ending point.

WATERFALL

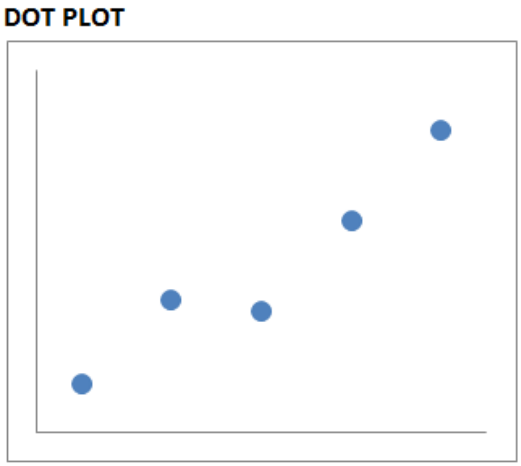


4.2.5. Scatterplot

1. Exploratory data analysis
2. Correlation between two variables
3. Time series



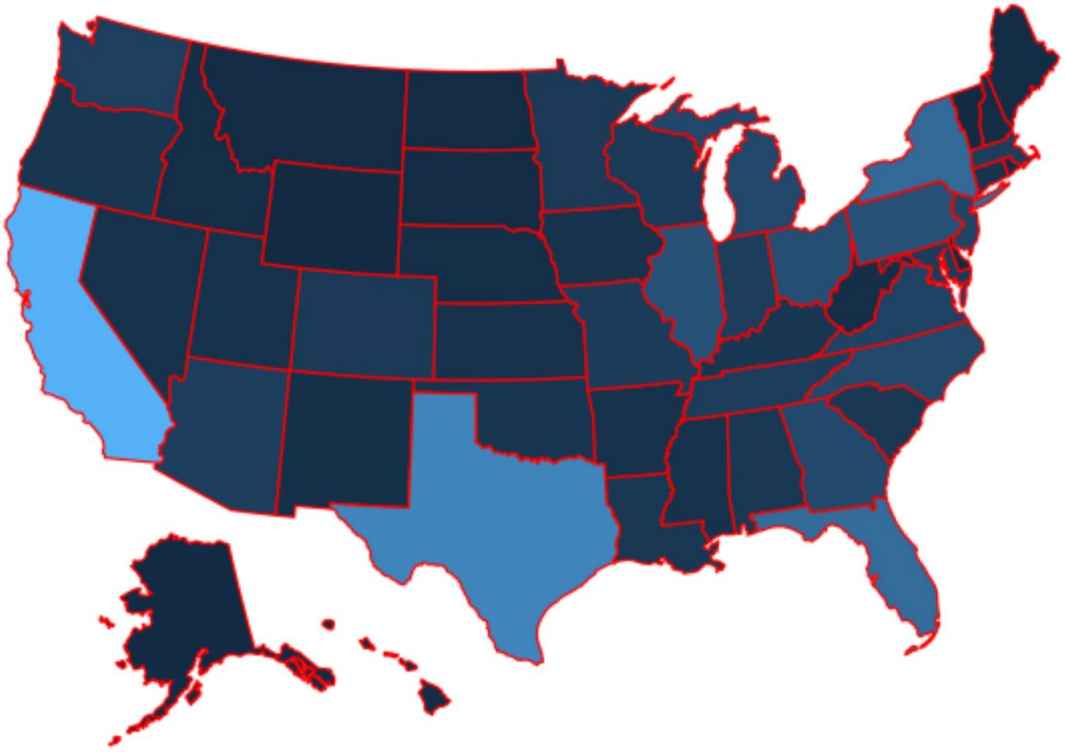
Scatterplots are useful to show the relationship between two variables or the correlations between them. Use care when using scatterplots as they are not well understood by all audiences.



Dot plots are useful for nominal comparative relationships when you want to highlight differences that would be hard to see in a bar graph that has zero as the baseline. Dot pots are also useful for showing time series data not representing consistent intervals of time.

4.2.6. Maps

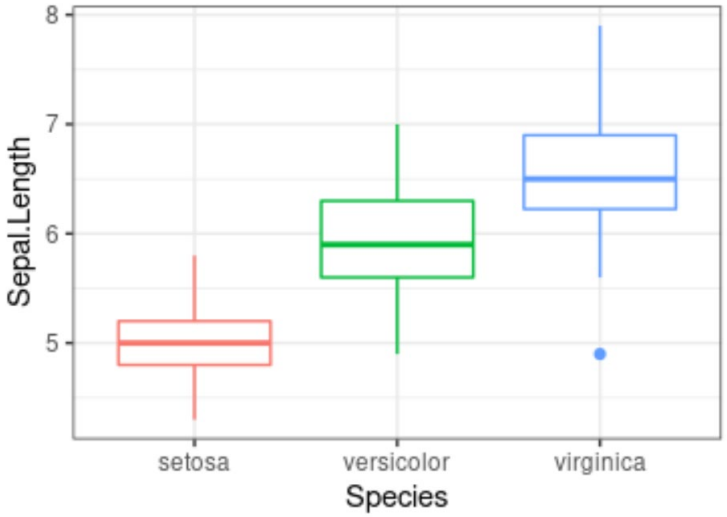
Maps are valuable for displaying geospatial information. One can use intensities of color by state or region to encode data. Maps are often preferred to tables which convey the same data. This leverages the fact that the viewer approaches a map with prior knowledge of the identity of the shapes being displayed. This permits the viewer to scan in any direction to "look up" the value associated with a particular territory. If the *order* of data is being emphasized, then one should consider a table or a bar chart.



4.2.7. Hybrid types

Some displays combine point and lines into hybrid shapes. Box plots and violin plots are two such examples. They are both used to display central tendencies of sample data, while also indicating outliers and other elements of the observed distribution.

Box plots display a large amount of distribution data in a single graphic: the highest and lowest values, the spread of values from highest to lowest, the median, and the 25th and 75th percentiles. Use box plots with care as percentiles are not readily understood by all audiences; consider simplifying to only show what is needed.



A violin plot shows more detail than a boxplot, similar to the difference between a density plot and a histogram.

4.2.8. Small multiples

Small multiples may also be called facet, or trellis plots. These exhibits repeat the same plot more than once, with subsets of data in each plot, as demonstrated in Figure 10. The plots are placed side by side to facilitate comparison. The subsetting is based on some categorical element like territory, vehicle class, predictor, etc. Small multiples are used to compare data across categories, to observe relationships or the correlation of each combination of two categories.

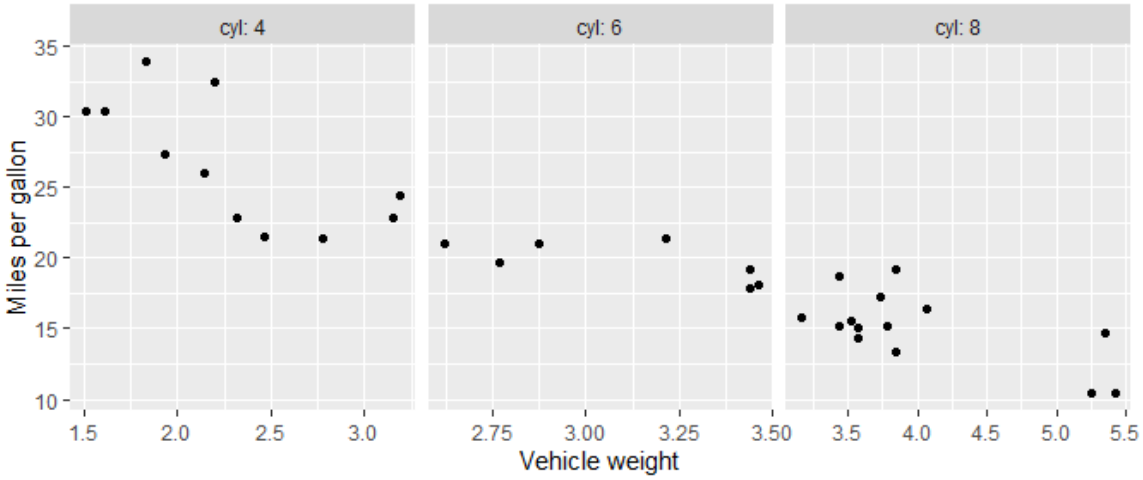


Figure 10: Small multiples

4.3. Visualizations to Avoid

Many common visualizations do not actually communicate the message effectively. These are less effective due to the science of how the human eye and brain work, and what is most effectively deciphered. Below are some common graph types, with a brief description of why they should be avoided.

- Pies and Doughnuts - the human eye can not accurately compare angles and area of pie or doughnut slices. Try stacked bar.
- 3D - due to the perspective, the human eye cannot accurately compare relative values three dimensionally. Try using color.
- Area - the human eye does not decipher area effectively. Try vertical bars.
- Double-Axis - too much for a reader to process quickly and effectively. Try separating the graphs vertically with the same x-axis and different y-axes.

4.3.1. Questions to help determine which type of visualization to use:

- What is the **simplest** way the information can be conveyed?
- Have I tried other ways of displaying the information?
- Have I shown a draft visualization to another person for feedback on how effectively it conveys the message?

5. How to design the visualization

In this section we provide some techniques for designing a data visualization to render a crisp graphic that highlights the message. Two techniques described in this section are:

- **Decluttering** to remove what's not needed, and
- **Accentuating** to highlight what is crucial

5.1. The Importance of Decluttering

Clutter is the enemy in a graphic. Lines, colors, fonts, etc. in default graphics tend to be overly busy. One must expend the effort to declutter data visualizations to maximize the clarity of the message. Below, we describe a few concepts related to decluttering: cognitive load and data-to-ink ratio.

5.1.1. Cognitive Load

Cognitive load is a concept discussed in (Knaflic 2015). Consuming information takes brain power. A viewer has limits on their short-term working memory. Viewers can only store three to four pieces of information at a time, which means that cluttered and poorly designed data visualizations increase a viewer's 'cognitive load' and reduce what they can retain. The data visualization author must design with the audience's cognitive load in mind. Without this consideration, the message gets lost altogether or takes the audience more time and energy to absorb (i.e. more than they may be willing or able to expend). Decluttering reduces cognitive load by minimizing what the brain needs to see and process, which enables the reader to see the relevant data and message more clearly. Think "less is more".

5.1.2. Data-to-Ink Ratio

Data-to-ink-ratio is a concept discussed in *The Visual Display of Quantitative Information* (Tufte 2001). Each dot of 'ink' on the visualization is valuable, and can be used to either display the data or to clutter a visualization. The author should create data visualizations with a focus on using 'ink' to display data in order to convey the message, compared to the amount of 'ink' used for text, lines, borders, or color/shading.

5.2. How to Declutter

The following aspects of a graph are relatively easy to declutter:

- Borders - delete them since they are generally not needed, and obscure key data ink
- Axis lines - delete them since the eye automatically sees a line created by the vertical or horizontal labels, or bars
- Gridlines - keep only if they serve a clear purpose, otherwise delete
- Axis tick marks - remove the horizontal or vertical line marks '-'
- Gray color - use for data or text that is not the *primary* focus of the visualization
- Axis names - use concise yet descriptive names to describe axes
- Titles - combine the title and the message into a single phrase

Legends are commonly used in graphs. However, the viewer must move their eyes back and forth from the graphed data to the legend, which takes time and increases cognitive load. If possible, place the label describing the data directly adjacent to the data. Apply the same color, weight, line-type, etc. for the text of the data label as used for its data (direct labeling). This approach may not work for all graph types or for all data. For example, if the lines in a graph are crossing each other, then direct labeling might not work effectively.

Y-axis labels should not use vertical text. It is common to see y-axis labels that are oriented at a 90-degree angle from the x-axis, reading upwards. No one naturally reads this way. Vertical orientation is more difficult to read and is a form of cognitive load; at a minimum it slows down the reader's ability to quickly identify the axis. Instead, try arranging the vertical-axis label horizontally at the top left of the axis the way we naturally read, or include the vertical axis label within the graph title or a graph footnote.

5.3. How to Accentuate to Make the Data 'Pop'

We have any tools to draw our viewers attention, and highlight elements that important for message to stand out. We'll discuss two (similar) schools of thought that describe ways that the human eye and brain perceive our world.

5.3.1. Gestalt principles

(Gestalt School of Psychology, 1912). These principles identified how the eyes and brain work to visually connect things together and make sense of our world. The following principles apply:

- Proximity - The brain naturally groups together items that are closer together. We can design our data visualizations to direct the patterns our viewer sees, by placing the relevant items in close proximity.
- Similarity - Objects with similar size, color, shape, font, or angular orientation are perceived by the brain to be part of the same group.
- Enclosure - Using some type of border or shading can render data or multiple objects to be associated as grouped objects.
- Continuity - The brain will attempt to enclose things (lines, objects) that are not fully enclosed by a solid line. The brain will 'fill in' a dashed line to perceive it as enclosing something if it can be interpreted that way. The brain may perceive a border to exist when objects are lined up; for

example, bars lined up along the x-axis could serve as a graph's x-axis border even if the border isn't shown.

- Connection - Elements that are visually connected are perceived to be related. This principle is commonly used in line graphs to literally 'connect the dots' for the viewer.

5.3.2. Pre-attentive Attributes (Few 2012)

The phrase 'quantitatively perceived' means that the human eye and brain are programmed to perceive a specific set of visual attributes very quickly and with a *high rate of accuracy* as to the quantity expressed by the visual attribute. Some attributes may be quantitatively perceived generally (e.g. area), but the brain and eye cannot be relied to be accurate compared to other attributes.

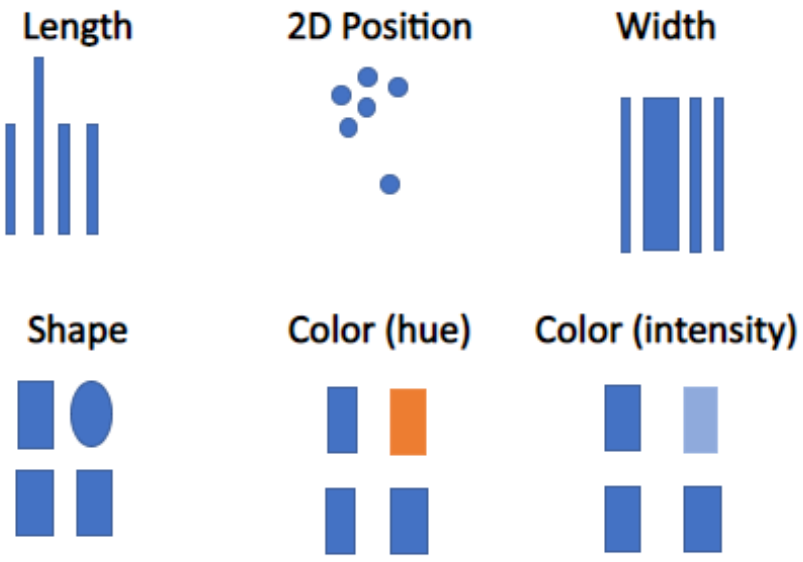
Attributes of Form

- Length - can be quantitatively perceived.
- Width - can be quantitatively perceived, but limited in accuracy.
- Area & Volume - *cannot* be quantitatively perceived.
- Spatial position - can be perceived and contrasted in vertical and horizontal position fairly well, but only in *two* dimensions not three dimensions
- Shape - *cannot* be quantitatively perceived.
- Location in 2-Dimensions- *cannot* be quantitatively perceived.

Attributes of Color

- Hue refers to color. Hue can be described by the location on a standard color wheel. Color combinations that work well together, and are distinct enough from one another, can be found on the website ColorBrewer.org.
- Intensity refers to the 'fullness' of a color (saturation), and lightness or darkness of a given color.

Below are some examples of how pre-attentive attributes could be used to distinguish and compare values or objects.

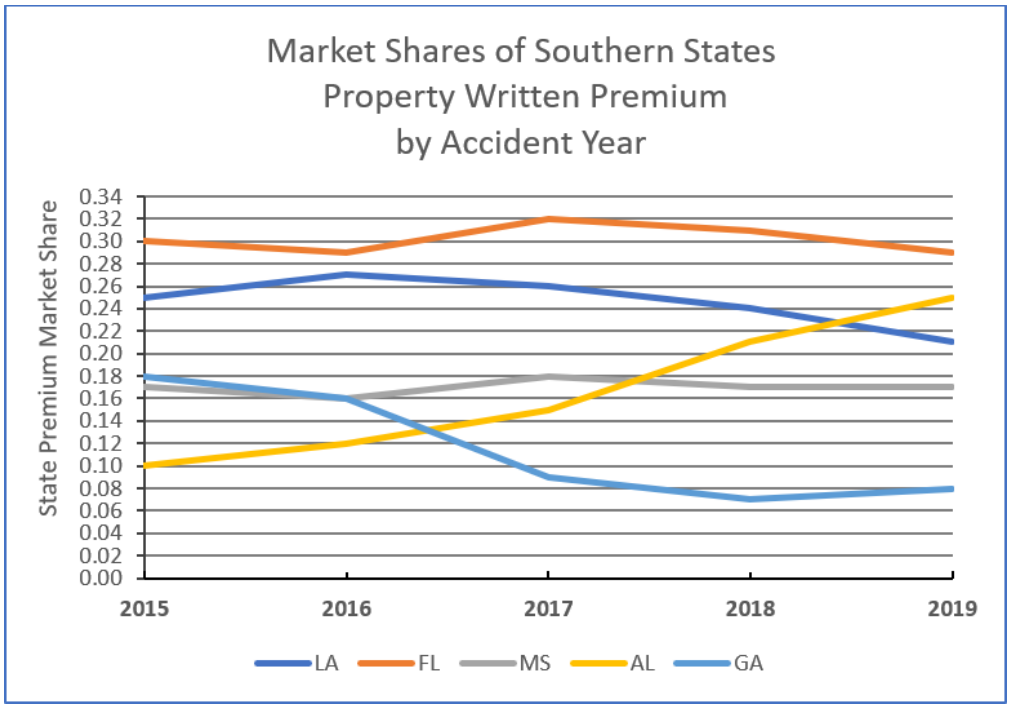


- Note that *area, volume, angle, and depth are all omitted from the list of pre-attentive attributes*, since the human eye cannot easily decipher differences in these. This is why we should generally avoid pie, 3D, area, and bubble charts. There are some exceptions. For example, some designers say it might be OK to use pie charts with only two variables. Others might say that it is OK to use treemap and bubble charts sparingly, to show generalities rather than precise differentiation between very similar quantities.

5.4. Before-and-After Example of Data Visualization

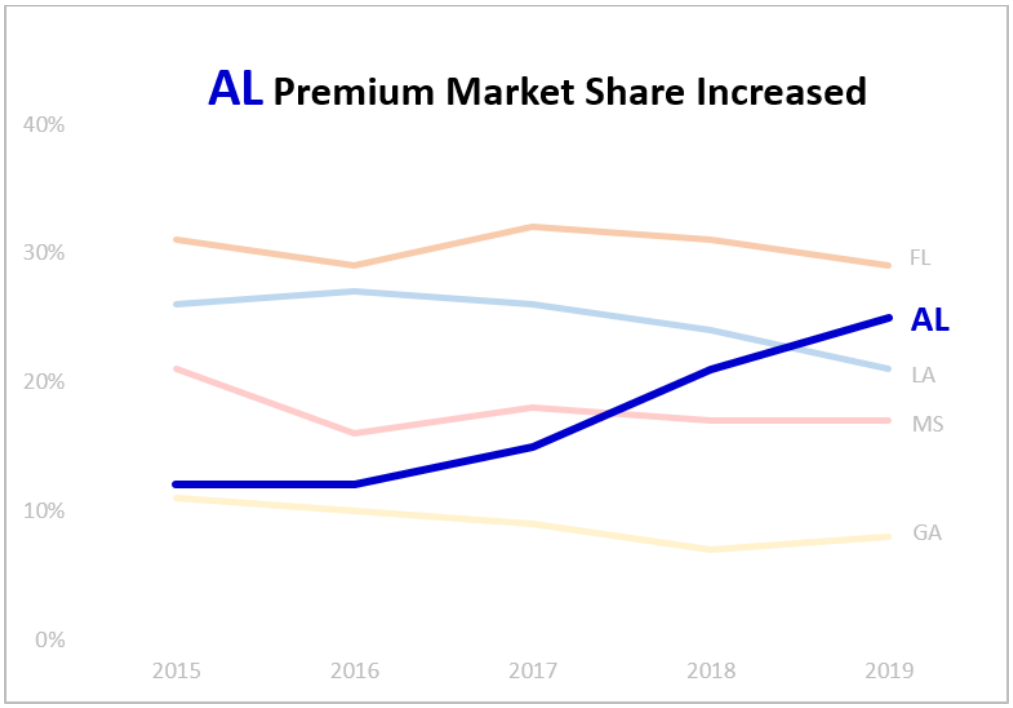
In the two graphs below, we show a *before* and *after* version of a graph.

Before we've applied data visualization basic techniques using default settings, we can hardly see the data, and we certainly cannot tell what the message is or even what to focus on.



- Too many border and axis lines, and the gridlines actually *hide* the lines representing the data
- Too many Y-axis labels, the color is too dark and/or lines are too thick, all draw our attention away from the data and increase cognitive load
- Color is used to differentiate the lines, and while they are color-coded with the legend, they each have bright colors so it isn't clear where to focus
- The legend requires the eye to look back and forth from the data lines to the legend
- The title is long and cumbersome, and we still don't know what the message is

After we have applied basic data visualization techniques we can clearly tell what the message is from the title and color choices. The axis amounts and information is visible but given less emphasis with gray or muted color.



- No gridlines, border, or axis lines.
- Fewer Y-axis labels; still allowing for reference for the lines
- Color is used to differentiate the lines, but muted tones for other states and darker for the focus state.
- The legend is omitted entirely, rather each line is direct-labeled
- The title tells the reader what the message is, and uses color-matching with the line and line label.

Which graph would you rather see as a viewer? With some effort and knowledge, you can see dramatic improvements in delivering the message, while reducing the cognitive load of the viewer.

6. Conclusion

With these basic principles and examples, you can start creating impactful, sleek, modern data visualizations that convey your message effectively.

Think about your audience, what they know, and what they *need* to know. Keep asking yourself and others **"What's the Message?"** throughout the data visualization design process. Your focus should be on identifying and displaying those things that best convey the message to your intended audience.

Use the visualization that matches the type of data and message. Line graphs are good for time series and (smoothed) distributions. Bar graphs have many uses, including comparing values by category, and parts-to-a-whole with stacked bars. Use horizontal bar graphs when the data labels are long, since we read horizontally (not vertically, or on a 45-degree angle).

Declutter visualizations, by removing extraneous axis lines, tickmarks, borders. Remove everything that does not help tell the message, or use gray font or line/bar color to de-emphasize non-essential elements. Let the data itself be the star of the show.

Accentuate the data so your message pops out for your audience. Draw attention to the most important data using length, width, shape, color, and other techniques to distinguish the data and highlight the message. Use Preattentive Attributes when designing your visualization (use length and relative position in 2D, rather than area and volume, for example). Use Gestalt principles, which reflect the science of how the human eye perceives lines, bars, dots, etc. in space.

Be creative and have fun creating data visualizations that enlighten your audience and convey your message!

7. Acknowledgements

This paper comes from the work and discussions of the CAS Data Visualization Working Party. The membership is comprised of the following individuals:

Damian Bailey	Jennifer Levine	Keith Quigley
Julian Coleman	Khanh Luu	Tanveer Quraishi
John Deacon	Madeline Main	Kyle Reed
Marc-Andre Desrosiers	Paul Majchrowski	Rajesh Sahasrabuddhe
Chunyang Fan	Paul Mayfield	Raymond Tobias
Brian A. Fannin	Sameer Nahal	Karthik Tumuluru
Navarun Jain	Bradford Nichols	Jeffrey White
Benyamin Kosofsky	Brian O'Reilly	Patrick Yu
Clifford Lau	David Payne	Arthur Zachary

8. References

[1] Black, Rob. 2018. "Data Visualisation as a Powerful Means of Communication."

[2] Bolch, Charlotte A, and Tim Jacobbe. 2019. "Investigating Levels of Graphical Comprehension Using the Locus Assessments." *Numeracy* 12: 8.

[3] Cairo, Alberto. 2012. *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders.

[4] ———. 2016. *The Truthful Art: Data, Charts, and Maps for Communication*. New Riders.

[5] Campbell, Mary Pat. 2016. "The Why of Data Visualization." https://www.soa.org/news-and-publications/newsletters/compact/2016/march/com-2016-iss-52/the-why-of-data-visualization/.

[6] Cleveland, William S. 1985. *The Elements of Graphing Data*. Second. Hobart Press.

[7] ———. 1993. *Visualizing Data*. Hobart Press.

[8] Evergreen, Stephanie D. H. 2016. *Effective Data Visualization*. Sage Publications.

[9] Few, Stephen. 2012. *Show Me the Numbers*. Second. Analytics Press.

[10] ———. 2020. "Visual Business Intelligence." https://www.perceptualedge.com/blog/.

[11] Fung, Kaiser. n.d. "Junk Charts." https://junkcharts.typepad.com/.

[12] Gelman, Andrew, and Anthony Unwin. 2013. "Infovis and Statistical Graphics: Different Goals, Different Looks." *Journal of Computational and Graphical Statistics* 22. https://doi.org/10.1080/10618600.2012.761137.

[13] Healy, Kieran. 2019. *Data Visualization: A Practical Introduction*. Princeton University Press.

[14] Knaflic, Cole Nusbaumer. 2015. *Storytelling with Data*. Wiley.

[15] Tufte, Edward R. 1997. *Visual and Statistical Thinking: Displays of Evidence for Making Decisions*. Graphics Press.

[16] ———. 2001. *The Visual Display of Quantitative Information*. Graphics Press LLC.

[17] ———. 2006. *Beautiful Evidence*. Graphics Press LLC.

[18] Wickham, Hadley. 2010. "A Layered Grammar of Graphics." *Journal of Computational and Graphical Statistics* 19 (1): 3–28. https://doi.org/10.1198/jcgs.2009.07098.

[19] Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science*. O'Reilly.

[20] Wilkinson, Leland. 2005. *The Grammar of Graphics*. Springer.

[21] Yau, Nathan. 2011. *Visualize This: The Flowingdata Guide to Design, Visualization, and Statistics*. Wiley Publishing, Inc.

Segmenting Closed Claim Payment Data to Estimate Loss and ALAE Reserves for Construction Defects

by James B. Kahn, FCAS, MAAA, Brad Tumbleston, FCAS, and Wilson Townsend

Abstract: Actuaries looking to assess an entity's ultimate liability for Construction Defect (CD) exposure often find themselves with many difficulties, not the least of which is the relative scarcity of accompanying literature within the Publications of the Casualty Actuarial Society. Industry meetings still continue to have concurrent sessions around the complexities of CD estimation, with a few presentations available for use.

This paper provides a few new considerations for estimating CD liability as well as providing suggestions and enhancements to techniques that are occasionally considered. Specifically, it better allows a user to consider historical emergence, understand and amend selections and assumptions based on changes to the portfolio's history, and consider future changes to assumptions.

Noteworthy to consider, the methodology described within this publication:

(i) Notes particular instance of the time from when a claim is reported and how long it remains open prior to closure. We have observed that this relationship is a significant driver of differences in characteristics of claims, and assumptions should be reflected accordingly.
(ii) Selectively limits loss claim severity to temper or avoid distortions that accompany unusual large claim emergence that may not be reflective of the overall characteristics of the book of business being reviewed.
(iii) Further segments the traditional Closed with Payment claim grouping into Closed with Payment (regardless of ALAE payment) and Closed with Paid ALAE only, allowing separate frequency and severity selections for each. This split considers the different historical emergence patterns we have observed for the different segments themselves.
(iv) Provides a platform to amend reporting assumptions should the historical relationships between Accident/Loss Year and Report Year be different from current or future time periods.
(v) Provides for the establishment of an excess "load" based on historical and current observations to handle severity limitations noted earlier. It could similarly consider a load for the liability associated with future reopened claims noted from historical percentages; and
(vi) Utilizes relatively simple data collection that can be used even with information from an Inception to Date loss run for Loss, ALAE, and Claims activity.

1. INTRODUCTION

Accurately estimating future CD liabilities is different from estimating other commercial lines due to the complexity of construction, the nature of the litigation, and the number of parties. The list of issues includes:

- Differing contract terms and conditions at the outset of the building process.
- Late reporting due to the long latency periods that complicates the identification of accident dates.
- The large number of developers, contractors, and subcontractors involved with multiple attorneys.
- The quality of workmanship and training of contractors and subcontractors.
- Aggressive homeowner's counsel.
- Litigious nature of CD claims.
- Multi-family homes versus single-family homes.

- Evolving and changing terms and conditions of the insurance coverage.
- "Risk transfer" contractual mechanisms between general contractors (GCs) and subcontractors including Additional Insured (AI) endorsements and contractual indemnity resting unequally among the defendants.
- A multitude of changes to insurance coverage, like the advent of Owner Controlled Insurance Program or "Wrap" coverage, changing the landscape of CD litigation.

Insurers responded to these rapid changes by changing the mix of business they underwrote, adding restrictive policy language including limiting the number of policies triggered by any one case, and increasing self-insured retentions (SIRs) and deductibles. Many insurers stopped underwriting residential exposures or construction risks altogether. Because of many of these rapid changes, those who review CD business often note vast differences of development and emergence patterns even within books of CD. Additionally, with no meaningful industry information available for comparison purposes from standard schedules within the Annual Statement, CD business is often one of the more difficult segments to project using traditional actuarial standard reserving techniques.

This paper lays out a methodology to address the various uncertainties, by segmenting the data and performing projections of ultimate reserve liability separately for each segment. It is also the intent of the authors to lead the readers to think of additional data segmentation that could be considered to deal with the impact of various changes and differences within CD books of business.

As a result of a few of the issues previously cited, actuaries have often used a frequency/severity type of Report Year methodology similar to what is used in other lines of business. Changes in the makeup of the book can lead to difficulties in considering future claim emergence. In addition, since CD business is not written on a stand-alone basis, considerations of frequency to an exposure base is not relevant for future projections. Once estimates of claims are determined, understanding of future claims that close with indemnity payment could have their own distortions if the makeup of the book has changed over time (whereby the future claim reports may not be reflective of what has been seen historically in the data). A few of the uncertainties seen for many CD methods (in particular, the number of future reported claims) are not specifically addressed within this paper. However, changes to the historical emergence can be more readily observed, with anticipated future changes considered more rapidly. Accordingly, we have found this method to be more responsive than many others.

Estimating severity has its own challenges as well. The value of homes that have incurred damage is often independent of the contractors' historical work. Any changes to the mix of business, new home values, size of development or project, etc. from historical amounts will have an impact. This is especially challenging when dealing with a report year methodology, since its result likely contains several policy years' worth of information.

CD portfolios have among the highest percentage of allocated loss adjustment expense (ALAE) or defense and cost containment (DCC) costs (collectively referred to as "ALAE" throughout the paper) of any reviewed segments, with overall costs equal to or even excess of the loss itself. ALAE costs are not only large when associated with underlying loss payments, but expenditures also need to be understood for claims that will be closed with no indemnity payment but with an ALAE payment. The latter may even be relevant when considering the emergence of claim reports beyond the relevant statutes of limitations or statutes of repose (similar to a statute of limitation, but where the deadline is based on the occurrence of a certain event that doesn't itself cause harm or is incapable of discovery with reasonable efforts) of the states where the claims are reported. The ALAE patterns between these two situations (indemnity paid, indemnity not paid) may be quite different in understanding exposure for future claim expense liabilities.

Transfer of risk adds another component of ALAE that often needs to be segmented due to its different characteristics. ALAE incurred in the defense of the named insured (NI) and ALAE incurred to defend additional insureds do not behave similarly. Understanding the book, the prevalence of AI endorsements, and the mix of contractors is critical to understanding the differences in the amount of ALAE.

The authors discuss ways of addressing the issues above in the main body of the paper. We also note that data segmentation and collection for this technique are relatively simple and can be used with a loss run that is easily compiled and maintained for both current and future evaluation periods. Coordination and communication between the claims professionals and actuarial teams are critical to the process and are the only ways for an actuary to truly understand some of the shifts and trends within a portfolio.

The authors' hope is that the techniques discussed within can either be utilized or at least amend current techniques for specific CD books to better estimate liabilities and scenario test reasonable alternatives accordingly.

2. BACKGROUND ON CURRENT METHODOLOGIES AND CAS HISTORICAL PUBLICATIONS

It has been close to 20 years since Green et al [1] published "Reserving for Construction Defects" as a reserving call paper (Green paper). That paper detailed background on CD at the time, with methodologies to project reserve liability and concluded by wondering whether the observed CD emergence from the late 1990s would continue in future periods. Now with the benefit of hindsight, we can conclusively say that CD continues to be as unique a segment as ever, with many aspects that would need special consideration in making projections.

There have not been many, if any, specific CAS publications on performing CD liability projections since the Green paper. CAS conferences still have occasional presentations pertaining to CD liabilities,

providing many specific examples of relevant changes to background and landscape, with fewer providing detail on methodologies to project ultimate reserve liability.

This paper will take the readers through the "Kahn-Tumbleston-Townsend" (KTT) methodology to address a number of CD projection nuances while describing situations that led to the design. A number of considerations, differences, and enhancements to existing methodologies, including those considered by Green, were noted earlier in the Abstract. The authors have found that the methodology detailed has consistently performed better than others we have used (such as estimation of IBNR using estimated exposure distributions) with a better ability to understand and incorporate changes within the particular segments being reviewed.

Any user of this publication should feel free to adapt as needed to their own books of business, including any further segmentation of data and classes that may be deemed relevant and/or credible.

3. CONSTRUCTION DEFECTS 101

Special care and consideration should be given when projecting liability for CD segments, given a few unique characteristics of the losses themselves as well as the way corresponding coverage would be applied. Many of the peculiarities have become so systemic in some states, that it is actually quite rare that any condo, townhome, or development would completely escape litigation in today's litigious environment.

3.1. What Defines a Defect?

A construction defect is the failure of a building or any building component to be erected in a reasonably workman-like manner, or to perform in the manner intended by the manufacturer or reasonably expected by the buyer which proximately causes damage to the structure. For insurance purposes, a defect is only for resulting property damage in third-party liability situations.

Defects themselves are defined to be either "patent" where the defect is readily observable or evident or "latent" where the defect is present but not readily detectible by reasonable inspection. A typical example of a latent defect in a CD case is water infiltration behind drywall, which would not be readily evident to the homeowner. In the cases of latent defects, it most likely would take a significantly longer time for the defect to become detectible, even with the implementation of routine inspections to a property.

3.2. Homebuilding

The process of building homes, whether it be single-family homes or multi-family homes (including condominiums, townhouses, apartments, and mixed-use developments) involves multiple contractors with specific expertise. Builders or developers usually initiate projects and hire general contractors or

project managers to oversee the construction. For developments with hundreds of units, this is a complex job requiring planning and coordinating resources under tight timelines. Mass graders prepare the land for development and lots for building. Concrete subcontractors pour foundations, framers construct the buildings, and roofers work to make the building weather tight. Once these structural components are put into place, window installers, stucco subcontractors, plumbers, electricians, multiple flooring specialists, finish carpenters, and other subcontractors (of which there could be many) finish the interior. Fine graders, landscapers, and irrigation companies complete the outside of the home. While a typical CD claim may involve ten subcontractors, larger developments can have twenty or more.

3.3. Number of Years Eligible to File Claims

Statutes of limitations and statutes of repose serve to limit the time period a lawsuit may be brought. Patent defects often require that legal action be filed three years from the date of first notice. This period can be from one to six years, with three years being a typical threshold. Latent defects, which are often not noticed for several years, are given special consideration under the law. In instances of latent defects, statutes of repose often provide homeowners with a ten-year window in which to file claims. In some states, this could be as few as six years. It is possible to have costs beyond the various statutes of repose. Examples include, but are not limited to, (1) rules that grant an additional two years after discovery to file a claim if it is within the window of allowed claims, and (2) additional ALAE costs required to deny liability, even if denials are based on statutes of repose. A few states have stretched the statute deadlines even further with expensive processes for a defendant to escape liability.

3.4. Notice and Opportunity to Repair

A mandatory notice and opportunity to repair process was instituted in many states after builders and contractors faced numerous lawsuits without ever receiving notice of an issue. Beginning in the early 2000s[1], these laws generally required homeowners to provide notice to builders and gave them the right to repair the alleged defects prior to filing CD claims. The acts usually apply to residential construction and improvements to real property where building components do not meet industry standards that are specifically enumerated in the statutes. These acts typically prescribe a process for the homeowner to notify the developer/builder of the alleged deficiencies and give the builder time to remedy the alleged defect. Rather than serving to resolve defect claims earlier, these processes often add to the time between claim reporting and ultimate closing.

[1] SB 800 California Civil Code 895 – 945.5

3.5. Insurance Coverage for CD

Insurance coverage is triggered[2] by an "occurrence" resulting in covered "property damage" during the policy period in most states[3]. The most prevalent trigger calls for the insured to establish actual physical injury during the policy period. In a claim where water infiltration is the main driver of the litigation, each policy is generally triggered from the time the home was completed, e.g., when the water infiltration first started, to the filing of the lawsuit, if the insured can establish physical injury during each policy period. Courts rely on the definition of occurrence, which states in part, "continuous or repeated exposure to substantially the same general harmful conditions." Other states still rely on "manifestation" which only triggers the single policy in effect at the time of the discovery.

Understanding these subtle differences often affect reporting patterns and the amount of time a claim remains open. Insurance coverage applies only for the resulting property damage caused by the contractor's alleged negligent work. Commercial General Liability (CGL) insurance does not cover the actual work or product installed by the contractor. The language in the Insurance Services Office (ISO), Inc. Commercial General Liability Coverage Form [4] excludes "damage to your product and damage to your work." In almost all cases, the liability or negligence for workmanship of the contractor that leads to property damage to another component is covered. However, the work directly performed on the product installed or the work to install it is excluded from coverage. CGL polices are not guarantees and do not act as bonds.

3.6. Exposure Period of Latent Claims and Differences by Jurisdiction

Following the second Montrose Decision in California in 1995[4], most states adopted a "continuous trigger" of coverage where coverage existed from the time the property damage first took place (often when a home was completed) until a lawsuit was filed. In such cases, each carrier would be responsible for their portion of the property damage over all of the insured years for the loss. Indemnity was often allocated to each carrier providing coverage for the years following the home completion until the suit was filed. They would then be potentially responsible for their equitable share of the total damages awarded or settled.

Some jurisdictions continue to rule that when the damage first manifests, the date of such discovery is the sole trigger of coverage. A minority of rulings over the past twenty years have expanded coverage. For the purpose of monitoring claims history and making projections for future liabilities based on reasonable loss patterns, it is important to be able to separate the various claims by jurisdiction to understand the loss history as well as the expected future claims and liability emergence.

[2] Trigger of Coverage Chart, (https://www.alfainternational.com/insurance-law-compendia)

[3] Insurance Services Office, Inc., Commercial General Liability Coverage Form, CG 001 ed. 07/98, Appendix C

[4] Montrose Chemical Corp. V. Admiral Ins. Co., 10 Cal.4th 645, 42 Cal.Rptr.2d 324 (1995)

3.7. Owner Controlled Insurance Programs

In response to the growing volume and the expense of CD claims, a new insurance product was developed: Owner Controlled Insurance Programs (OCIPs), commonly known as "Wrap" policies. Wraps were intended to insure a single project or development rather than focus on a single insured. While the owners or developers controlled the insurance program, subcontractors working on the project were required to enroll in the Wrap and pay their portion of the premium. In practice, Wraps were issued on many commercial and municipal projects and later on in the residential projects. Meant to limit cross litigation between defendants, the Wraps worked well on some projects where the defendants could align behind a focused defense. In reality, some of the old problems of cross litigation and indemnification suits are still problematic in driving up costs. Good claims management is needed for the policies to operate as intended.

3.8. Differences in Programs or Books of Business

Several factors drive inconsistencies between programs which make it difficult to rely on data from other programs. Who originated the business or whether origination was from a broker, agent, or Managing General Agent (MGA), can determine underwriting differences, rating differences, and mix of business. The mix of business between high severity claims of developers and general contractors, or the lower severity and higher volume of artisan subcontractor claims are distinctions that must be understood. Even within the subcontractor category, there are significant differences in exposure from structural and building envelope subcontractors or artisan contractors. Underwriting changes are frequently made over the history of a program. These changes can be driven by underwriting results from high loss and combined ratios; poor litigation results when defending policy language; and insurer risk appetite. Over the last twenty years, insurers and risk bearers have sought to limit coverage to a single policy year by adding prior work, prior acts, and pre-existing damage exclusions, some of which have succeeded in limiting coverage, while others have not stood up to judicial scrutiny. Some carriers tried to reduce their risk by eliminating condominiums, townhouses, apartments, and even residential homes from coverage. Other endorsements were utilized to limit coverage by excluding roofers, any subcontractor involved in moving earth or preparing the land for concrete pads/foundations, building envelope subcontractors (such as stucco professionals), and to assure insureds stayed within their business classifications. Common endorsements and exclusions are shown on *Appendix B*.

3.9. Costs Associated with CD Claims

The duty to defend is greater than the duty to indemnify.[5] As CD claims have proportionately high ALAE payments due to litigation practices of the developers/general contractors who often name subcontractors as cross or third-party defendants without completing an investigation into the alleged property damage, questions as to date of occurrence and whether the statute of repose is still valid, whether or not insurance coverage pertains, which of several contractors may be ultimately liable, and the allocation of damages to triggered policies negotiated between insurers. In addition, many of these claims close with no loss payment which also lead to an overall higher ALAE to loss ratio. Some insurers and large self-insureds have entered into variable or fixed fee types of arrangements with defense counsel where a series of payments are made to correspond with the level of work. Subtle differences exist within these types of arrangements.

Historically, ALAE costs were associated with claims where construction experts were needed to testify on differences as to what building procedures should have been performed as well as whether building standards were followed. These types of costs are less frequent in today's environment as carriers look to limit the total cost of the claim. However, certain trades remain very expensive to defend — concrete/foundation subcontractors for example typically require soil engineers and geotechnical engineers regarding claims of compression and concrete strength.

3.10. Risk Transfer

The developers and GCs employ their leverage contractually so they can be indemnified by their subcontractors in the case of any property damage "arising out of the work" of the subcontractor. Under the terms of the contract, the subcontractors are required to name the developer and general contractors as an AI on a subcontractor's policy. In some instances, developers include a Prevailing Party clause in the contract giving them another chance to recover their litigation costs. As a result, a majority of litigation costs incurred by a developer and general contractor are passed to the subcontractors' insurers. Over the last twenty years, there have been many changes to the ISO AI Endorsement, in the law, and how insurers manage the resultant data.

Over time, many of the same coverage questions continued to emerge in new jurisdictions, requiring more experts. Anecdotally, we see that the additional number of experts reviewing similar issues has resulted in a modest decrease in such costs. Still, it continues to be common practice for a GC to bring in several experts on many related issues, especially with large construction projects. Understanding the differences between costs incurred by the NI (in this case the subcontractors) versus the AI, and segmenting that data appropriately in any type of analysis or projection, can be critical in some programs where AI Endorsements were issued freely.

[5] Gray vs. Zurich Insurance Co., 65 Cal.2d 263 (1966)

3.11. Potential Large Claim Emergence Throughout the History of the Portfolio

Common insurance lines of business such as Private Passenger Automobile tend to have their more severe loss emergence the longer a reported claim is from the beginning of an accident date or policy inception date. Easier-to-settle claims are usually closed quicker and tend to be of smaller value. In CD, large claims can emerge very close to the project completion date or several years thereafter. The nature of CD losses does not necessarily lend itself to large claim emergence several years after a particular property is completed. In fact, many CD claims reported close to the statutes of limitation or statutes of repose are denied almost immediately. As another example, a latent claim could emerge whether or not the specific property is a high-cost or low-cost home. Losses are rarely for full value of a property. As such, there is less of a likelihood that emerged losses would vary as much based on costs of the property. Additionally, if early signs of damage are recognized, expenditures could potentially be mitigated.

3.12. Changes to Nature of Originally Filed Claim

Patent losses are often filed initially to establish the beginning claims docket. Following this, further inspection of properties may unearth additional defects (both patent and latent are possible). Accordingly, in these situations, it is possible that the final claim or set of claims could conceivably have little resemblance to the claim initially reported.

3.13. High Percentage of Claims That Close Without Payment

There are higher ratios of claims that are closed with no payment (CwoP) for this segment than we see in most other portfolios given the duty of insurers to defend litigation due to coverage issues, statutes of repose for filing claims, and exclusionary language within insurance contracts. The approach and filing practices of the homeowners' attorneys, naming many subcontractors often without allegation of defects attributable to their work, also contributes. The large percentage of CwoP claims, as well as understanding historical shifts in these ratios over time, should be considered in reserve estimates and understood by relevant interested parties such as underwriters and claims personnel.

4. UNIQUE ACTUARIAL CONSIDERATIONS REGARDING CONSTRUCTION DEFECTS

4.1 Use of Report Year Methodologies

With the difficulty in determining accident date assignment for latent claims, CD liability projections are often performed on a report year type basis, estimating historical exposure bases (even though underlying policies would rarely, if ever, be written on a claims-made basis like most actuarial segments that use report year methodologies). The difficulty in defining accident years (and subsequently policy

years) for CD as well as the potential latent emergence of covered losses (many reporting greater than 5 years following the building of the properties in question) adds to the reasons to consider report year methodologies with the various projection methods.

4.2. Pitfalls of Utilizing Premium as an Exposure Base

Projections for CD coverage have rarely been standardized. Experience is difficult to quantify because CD coverage itself is not specifically written as a stand-alone policy. It is often a subset of General Liability coverage, where it becomes difficult to assign an appropriate exposure base (such as earned premium) to accompanying losses. Report Year CD methods often estimate a calendar/accident year premium and estimate the premium earning for the various years into future report years based on approximating the emergence of historical reported claims. Actuaries then perform projections using the premium as an exposure base for Bornhuetter-Ferguson methods or to determine a frequency of future claims for a frequency/severity type of method.

These premium based methodologies rely on a number of difficult assumptions:

4.2.1. Estimation of Premium Itself

Estimating premium by accident year is often subjective in determining the percentage of CD premium from within overall contractors premium.

4.2.2. Impact of Historical Rate Changes

Ideally, we should understand historical rate changes to the premium used so that there is an appropriate comparison of on-level premium for the different accident years. It is often very difficult to determine rate changes for the different types of contractors classes, and the application of schedule credits could conceivably mask the overall reduction in premium from one period to another. The methodology itself would assume that reduction in premium corresponds to an exact decrease in losses by the same percentage. Failure to properly account for rate changes would add another level of complication to these types of projections.

4.2.3. Complications from Shifts in Mix of Business

Contractors business notoriously is driven by macro conditions such as the economy, housing starts, etc., with peaks and valleys of historical writings. Given the changing population over time, it is certainly possible that the projection of claim reports would not behave identically for different contractors or building projects that enter and exit the population over time. Accordingly, the fundamental assumptions as to premium earning would be violated in such a situation.

4.2.4. Use of Multiple Earnings Patterns

With the various accident year premium exposures being allocated to future report years based on perceived earnings, each resulting report years' estimated premium consists of several historical policy years that may have very little in common; this phenomenon could continue indefinitely. Given the composition of this estimated premium, it becomes extremely difficult to understand underlying causes that may cause relevant report years' experience (such as observed loss ratios or frequency) to look different among the various periods. Changes within a book by policy year or accident year (such as fundamental changes in underlying exposure, terms and conditions, etc.), would be difficult to understand and quantify with these types of methods. When we also consider the uncertainty of estimating historical rate changes, it becomes difficult to rely upon methods that utilize premium estimates with any degree of certainty.

4.3. ALAE Costs with or Without Corresponding Loss Payments

As we noted, there are often extensive ALAE payments in CD as disputes arise as to coverage, ultimate liabilities, validity of endorsement exclusions, application of statutes of limitation, where to file claims, etc. We have found that sometimes ALAE costs correspond to matters that are much simpler to handle. For instance, events that could lead to early denial include obvious statutes or where policy periods pertained to periods significantly different than the filed claims. Often, there would not be corresponding loss payments associated with such ALAE payments. Actuarial techniques shown have rarely, if ever, considered these differences by segregating their various projections.

4.4. Claims That Close Without Payment

As noted earlier, CD claims often result in closings without any associated payments. It is important for actuaries to consider CwoP claims in projections of future claim counts, with many methods often considering a companion severity estimate for claims that close with given payments. Given the multiple policy years of coverage that make up a corresponding report year, changes to business mix can sometimes have a profound impact on CwoP percentages (as well as any other CD related characteristics) that may ultimately bear little resemblance to historical observations:

4.4.1. Changes to State Mix

Any changes to state mix over the course of time could create situations where the future CwoP ratios look different than historical averages. Recall that states have different allowable triggers, with some considering coverage for continuous triggers, whereas other state laws allow only a claim manifestation trigger. Additionally, some states may allow additional time for filing a valid claim (leading to a possibility of claims closing with a payment) whereas others would not (as such, a claim would be closed with no payment).

4.4.2. Changes to Law

Changes to law (either tort or legislative) can have an impact on CwoP liability as coverage viewpoints may change to make establishment of liability either more or less difficult than what was observed historically. The insurance industry for example has seen non-CD examples such as molestation claims where the statute of limitation has lengthened, allowing formerly barred claims to be filed. Similar changes have taken place to historical laws (such as allowing additional years to file claims) or rules concerning CD that may need to be considered for future evaluations. Any other changes such as revisions to triggers, excluded classes, choice of law, etc. can impact liability considerations.

4.4.3. Claim Filings for One Time Events with Low Cost Claims

On occasion, the authors have found situations leading to an increase in smaller claims for a specific event such as a class action. Any such event should receive consideration for further data segmentation. This one-time increase in claim counts (that do settle with small indemnity payments) needs to be understood for projections as this type of distortion would impact historical claim frequency (higher), CwoP percentage (lower), and historical severity (lower).

4.4.4. Consideration of Policy Limit Caps

It may be possible in some situations where numerous valid claims would be filed historically, but the corresponding insurance policy cap would have already been reached. As such, there would be no more loss coverage remaining, even if there remain appropriate types of claims open or being filed.

4.4.5. Use of Endorsements and Exclusions in Policy Language

Certain types of losses that were previously covered as valid claims, may not be allowable under revised language or exclusionary endorsements. In such cases, previously covered losses may not have valid coverage with the use of court tested exclusions.

4.4.6. Changes to CwoP Percentages as Claims Approach the Statutes of Repose

We have seen evidence in some books of business where additional claims were filed as the opportunity to file claims began to run out. If these claims are less likely to have valid coverage and are only being filed because of the ending of the filing opportunity, it would be more likely for such claims to close without corresponding loss payment. The possibility also exists that these claims do result in much smaller settlements (simply reducing the operational and ALAE costs), but where similar percentages of claims close with payment. It would be worthwhile to understand the impact of these filings prior to making projections for these specific CD books of business.

4.4.7. Relevant Operational Changes

Any relevant operational changes in procedure or definition should be fully understood to project future CwoP percentages. As an example, the authors have observed situations where several successive policy years existed for a given insured, and an insurance company's practice was to assign a filed claim to each historical policy year. Upon discovery of which policy would have the ultimate liability, the remaining assigned claims would be closed with no payment. Supposing in this example, the company had several old existing policies that were constantly renewed, but have recently begun to grow by writing new policies. In the updated situation, as policies were renewed, the practice of assigning claims to each policy period would as a result have fewer impacted policy periods. Upon the final assignment, the CwoP percentage could be much lower than what was seen historically, as there would be fewer policy periods that would close the claim with no loss payment.

Any other noteworthy changes (definitions of claim counts, inconsistencies upon definitions or claims handling within historical TPAs, coding issues, etc.) should be understood and adjusted as needed.

4.5. Differences by State

As mentioned, there are important differences by state in terms of deemed occurrence dates (continuous trigger, manifestation, etc.), risk transfer, and various reporting statutes. Accordingly, it is important for any projections to consider segmentation of data into similar portfolios. Such groupings could include separation of individual states or possible groupings of states with similar characteristics in treatment of CD claims. Any corresponding changes to historical state mix should be understood including the potential impact on assumptions.

Another issue is that some states allow for suits against the insurers in addition to the trades (direct action states). These often increase costs dramatically. The venue is critical in determining exposure as some states are widely known for being more plaintiff-friendly. Exposure in these jurisdictions can be dramatically higher.

4.6. Difficulty Benchmarking to Other Peer Companies

Information in the marketplace is difficult to access as publications such as an Annual Statement consider CD losses within the General Liability segments. Publications of specific studies pertaining to only CD losses are even more rarely produced or shared within the insurance industry.

5. PROPOSED ACTUARIAL SOLUTION AND CORRESPONDING METHODOLOGY

Because of the elements noted earlier, projections of CD liabilities are continually re-evaluated in practice. Report year methodologies essentially behave like calendar year methodologies and accordingly would be quite sensitive to both frequency components and severity estimations. Additionally, report year methods by their very nature are made up of several historical policies or accident year periods, which can be inconsistent in a segment like CD with all the potential historical changes (over several policy years) noted earlier.

We already mentioned changes that can take place including (1) changes in historical coverage and liability as a result of the application of limitations and endorsements, (2) internal operational changes including claims practices or definitions of terms such as claims, closed claims, ALAE, etc., (3) changes in mix of business being written both in terms of the underlying nature of business as well as location of liability or insured parties, and (4) changes in laws and torts through judicial rulings and case review. All these elements can have profound impact over time, and lead to instability with many actuarial methodologies used to project CD liabilities.

In particular, elements of establishing appropriate exposure bases for CD liabilities, including approximation of premium and their accompanying earning patterns, often have pitfalls as noted in earlier sections.

It is the intention for the KTT methodology to be better able to segment and identify changes within the various books of business as well as perform projections that may better capture elements that are more meaningful in understanding observed and future emergence. The KTT methodology exclusively uses Paid Loss and ALAE methodologies for closed claims, which the authors believe gives a more reflective viewpoint on observed historical emergence within the various CD segments. **The motivation for this change is that upon examining the history of several internal CD portfolios, we determined that the lag between report year and close year was the most predictive factor for both whether a claim will close with payment and the average severity of such claims that do close with payment.** The change in severity over time to closure was much more of an indicator of increases to larger claims than time since an accident or time since construction. This was another incentive for the use of payment methodologies in our projections.

Anyone using the KTT methodology should feel free to further split data into more homogenous segments to the extent their data would be credible enough to do so.

Given how impactful any changes from historical practice and makeup to a book of business can be for estimations within this segment, it is absolutely critical that open communication takes place between actuaries and other departments such as claims or underwriting both from a proactive sense (where actuaries can adjust assumptions based on input from these specialties) as well as an observed sense

(situations in which the actuaries can better explain observations of changes to data and projections to other interested parties).

5.1. Data Elements, Indicators and Requirements

The data elements needed for the projection are largely items that are captured in typical systems and may even be simpler than what is needed in many CD projection methods currently used by actuaries. Additional elements can be established early in the creation process and would likely be available in most companies' typical flat files:

5.1.1. Data Requirements

This KTT projection methodology does not require transactional data. Instead, all that is needed is an inception-to-date (ITD) loss run with the following fields:

- Claim ID
- Loss Year /Accident Year
- Report Year
- Close Year
- ITD Paid Loss
- ITD Paid ALAE
- Case Loss
- Case ALAE

See *Appendix A, Section 1: Data Requirements* for examples.

5.1.2. Loss Limitations

With Construction Defect liabilities, large loss emergence and manifestation can occur at any point in a traditional loss triangle, making it difficult to select loss development factors or severities. Issues related to poor workmanship are often independent of whether an incident is on a lower cost starter home or a more expensive home that may or may not have utilized expensive products. This contradicts many typical rules of thumb for actuarial thinking mentioned historically by Salzmann [8] and others that associate higher severity amounts as the age of a claim becomes greater.

The sporadic large loss emergence for CD portfolios, can be mitigated by selecting a loss limit - a split between what we'll consider "limited loss" versus what we consider "excess loss". The limited losses are used in the method's triangle based projection methods, and a load for excess loss will be incorporated following.

The loss limit should be selected such that the distortions from extremely large losses are dampened. For example, one might choose a loss limit such that ninety percent of claims closed with a paid loss do

not exceed the limit. If the book does not exhibit extremely large losses with high variation in timing, the loss limit may not be necessary.

This limitation would also provide an opportunity to specifically account for known large losses that have individual case reserves that claims handlers would be comfortable with. By removing these large losses and then including them as specific individual estimates later, distortions for these known outliers can be considered in a more reliable estimate (similar to how property catastrophe claims are often handled).

5.1.3. Derived Data Fields

Following the establishment of a loss run and loss limits, we can now look to derive additional fields as seen on *Appendix A, Section 1: Data Requirements*.

5.1.3.1. Open or Closed Claims

Claims can be classified as Open or Closed, based either on the Close Year or on the case reserves (where a case reserve of $0 would signify a closed claim). Closed claims can then be partitioned into three categories:

- Closed with Paid Loss (CwPL) – whether or not ALAE has also been paid
- Closed with Paid ALAE Only (CwPAO)
- Closed without Payment (CwoP)

5.1.3.2. Excess Loss Identifier

Once the CwPL claims have been identified, another indicator should be added for Closed with Paid Excess Loss (CwPL XS) claims. Note that these CwPL XS claims are a subset of the CwPL counts, rather than additional claims themselves.

5.1.3.3. Paid Loss Segmentation

Paid Loss can now be partitioned into three categories:

- Limited Paid Loss on CwPL claims
- Excess Paid Loss on CwPL claims
- Paid Loss on Open claims

5.1.3.4. Paid ALAE Segmentation

Paid ALAE is partitioned into three categories:

- Paid ALAE on CwPL claims
- Paid ALAE on CwPAO claims
- Paid ALAE on Open claims

5.2. Claim Count Projection Methodologies

This section details the many elements included in the projections of ultimate reserve liability. As noted earlier, this method largely tracks and considers methodologies involving claim counts and historical closed paid severity calculations. Consideration is also given for claim closures including those that may or may not close with a loss payment, with a similar breakdown within ALAE for those that close with or without a loss payment:

5.2.1. Reported Claim Count Projection

Following the establishment of a loss run and loss limits, we can now look to derive additional fields related to claim counts.

Using the loss run, a triangle of cumulative reported counts aggregated by loss year and by age of the report year from the loss year, can be constructed. This would be the only triangle in the method aggregated by loss year - all others rely on the report year.

The projected ultimate reported claim counts methodology uses traditional chain ladder development techniques. The next step in the process is to transform the triangle from cumulative counts to incremental counts, and then "square the triangle" - that is, fill the empty space below and to the right of the triangle with the projected future reported claim counts. Any standard methodology should suffice; see the Friedland [3] text for an entry-level walkthrough of many such methodologies.

Moving up and to the right for each diagonal would correspond to a particular report year. The sum of each diagonal determines the ultimate reported claim counts by report year. This process can be seen on *Appendix A, Section 2: Reported Claim Count Projection*.

5.2.2. Ratios of Closed Claim Counts

The next step in the process is to construct a triangle of cumulative closed claims further segmented by age of the close year from the report year. Using this triangle, a second triangle is constructed that divides each value from the first triangle by the ultimate reported claim count for the corresponding report year. The resulting second triangle depicts the cumulative close ratios for each report year.

The user would then need to select a cumulative close ratio for each age, and transform those selections from a cumulative basis to an incremental one. Similar to any other methodology, one should consider whether the incremental selections would make intuitive sense. For example, if the age 48-60 incremental closed ratio is three times larger than the 36-48 ratio, does that seem reasonable? Is there a known story that explains the observation, or is further investigation required? Policy language and exclusions, as well as changes over time as discussed earlier would likely have strong impacts on the closure ratios.

When this process is complete, apply the selected incremental closed ratios to open and yet-to-be-reported claim counts to project incremental closed claim counts by age of close year from report year.

Appendix A, Section 3: Closed Claim Count Projection shows a calculation for both future report years as well as the allocation procedure to better understand existing report years.

5.2.3. Note on Characteristics of the Triangles

From this point forward in the methodology, all triangles will have the following characteristics:

- Incremental, rather than cumulative
- Aggregated by report year and by age of the close year from the report year
- Contain only data from closed claims as a consequence of the last point

There are a number of resulting ramifications that should be considered. First, paid amounts and case reserves for open claims will not be found in any triangles from this point forward. Second, when comparing the same triangle at two different valuation dates (e.g., this year's analysis versus last year's), there may be significant differences within the interiors of the triangles. This is because the close year of a claim could change over time. As an example, suppose the claim was initially dismissed within a year of being opened, but was subsequently reopened and closed at a later period with Paid Loss. Changes to interiors of triangles result in a continual restatement of historical values to be most reflective of the latest situations. Finally, as all triangles are compiled on an incremental basis, there may be few claims beyond a certain age. It may be uncommon for claims to take more than five years to close after being reported to the insurer. In such a case, the actuary should consider combining all ages 60 months or greater.

5.2.4. Closed with Payment Ratios

The user will need to construct the following three triangles, understanding that they will also need to be complied on an incremental basis and then aggregated by report year and age of the close year from the report year.

- Closed Claim Counts
- Closed with Paid Loss (CwPL) Claim Counts
- Closed with Paid ALAE Only (CwPAO) Claim Counts

The authors have noticed considerable differences across all calculated functions between ALAE only claims, ALAE claims where a loss payment was made as well, and loss payments. Dividing the CwPL Claim Counts triangle by the Closed Claim Counts triangle derives a triangle of incremental CwPL ratios. The user would then select an incremental CwPL ratio for each age, similar to how one would select loss development ratios from a traditional loss development triangle.

Intuitively, one would think that there would be a positive correlation between the length of time it takes a claim to close and the likelihood that this claim closes with a Paid Loss. Nuisance claims would often be denied without a loss payment very quickly after being reported to the insurer. On the other extreme, claims that would remain open for five years or more would typically have a higher rate of being

closed with a Paid Loss. If the selected CwPL ratios do not reflect this correlation, the incremental CwPL ratio triangle can be an excellent tool for discussions with the Claims Department about why actual observations would deviate from intuition.

Finally, the user would need to derive a triangle of incremental CwPAO ratios in an identical manner, and select a ratio for each age. The same intuitions about the correlation between time-to-close and likelihood of closing with paid ALAE only may not hold. It may hold for the first few ages, but particularly as the CwPL ratio increases by age, the ceiling for CwPAO ratio selections must decrease because the sum of the CwPL and CwPAO ratios for any given age cannot exceed 100%. The derivation of CwPL and CwPAO frequencies can be seen on *Appendix A, Section 4: CwPL and CwPAO Ratio Selections*.

5.2.5. Claim Count Projections by Category

Earlier in this method and descriptions, reported claim counts were projected for future report years and future closed claim counts were projected by report year and by age of close year from report year. The next step is to apply the appropriate selected incremental CwPL and CwPAO ratios to the future closed claim counts. The resulting projections seen on *Appendix A, Section 5: Future Count Projections* are CwPL and CwPAO claim counts by report year and by age of close year from report year.

Projections in this format have the advantage of being ready-made for "actual versus expected" types of analyses, in that the "expected" closed (or CwPL, or CwPAO) claim counts for each future calendar year, which may ordinarily be difficult to derive, would be observed simply by taking the sums of the projected diagonals.

5.3. Loss Related Projections, Severities, and Considerations

Developing final estimates of ultimate loss liability involves utilizing the development of future closed claim counts and combining them with an estimate of severities for future years. Those pertaining to the loss only component are noted in this section:

5.3.1. Limited Loss Severities

The user now constructs a triangle of limited Paid Loss amounts on closed claims, and similarly aggregates these by report year and by age of close year. Dividing this triangle by the incremental CwPL claim count triangle determined from the previously detailed steps results in a triangle displaying CwPL limited loss severities. Finally, one should select a CwPL limited loss severity for each age, similarly to how one would select loss development ratios from traditional methodologies and severity triangles as shown on *Appendix A, Section 6: Limited Paid Loss Projection*.

Intuitively, there would be a positive correlation between the length of time it takes a claim to close with paid loss and the size of the Paid Loss. Small claims tend to be resolved relatively quickly after being reported to the insurer, while larger claims are more likely to enter litigation and therefore take significantly

longer to settle. If the selected CwPL limited loss severities do not reflect this correlation, one could consider whether the selected loss limit is set at too high a value (allowing larger losses to potentially distort earlier ages). If one concludes this to not be the situation, the CwPL limited loss severity triangle can be an excellent tool for discussions with the Claims Department about why practice deviates from intuition.

5.3.2. Excess Loss Frequency and Severity Determination

Similar to the limited loss calculations, one would need to construct a triangle of incremental closed with paid excess loss (CwPL XS) claim counts aggregated by report year and by age of close year. Next, we compare this triangle to the incremental CwPL claim count triangle from previous steps to derive scenarios for an overall CwPL XS frequency. For example, one could divide the sum of the entire first triangle by the sum of the entire second triangle to derive an all year weighted average frequency. Alternatively, one could divide the sum of the bottom X rows of the first triangle by the sum of the bottom X rows of the second triangle to derive an X-report-year weighted average. We would then select a single overall CwPL XS frequency.

To determine severities, we'd need to construct a triangle of excess Paid Loss amounts on closed claims aggregated by report year and by age of close year. Comparing this triangle to the incremental CwPL XS claim count triangle would lead the user to select a single overall CwPL XS severity (similar to the way the CwPL XS frequency was selected). This is detailed further in *Appendix A, Section 7: Excess Paid Loss Projection*.

5.3.3. Ultimate Loss Projections

Earlier, future CwPL claim counts were projected by report year and by age of close year from report year. Next, the user applies the appropriate selected CwPL limited loss severities to the future CwPL claim counts. The result is the projection of future CwPL limited Paid Loss by report year and by age of close year from report year, seen on *Appendix A, Section 6: Limited Paid Loss Projection*.

The CwPL limited Paid Loss projection format also has the advantage of already determining future values for an "actual versus expected" analysis, similar to the earlier projections for CwPL and CwPAO claim counts.

Finally, to determine the overall ultimate loss projection, one needs to multiply the sum of all future CwPL claim counts by the selected CwPL XS frequency and severity to arrive at the "load" for excess losses. This is shown on *Appendix A, Section 7: Excess Paid Loss Projection*. This excess load could be allocated to the limited losses in a format similar to the future limited loss or it could be left as a stand-alone reserve without being assigned to specific report years, close years, or calendar years. The advantages and disadvantages of either approach will vary according to specific circumstances.

5.4. ALAE Related Projections, Severities, and Considerations

Developing final estimates of ALAE liability involves utilizing the development of future closed claim counts and combining them with an estimate of severities for future years. Those pertaining to the ALAE only component are noted in this section. Of note, the KTT methodology does not consider partial payments for ALAE within the observed payment history that derives the historical severities. Rather, ALAE payments are only considered upon the year a claim is ultimately closed. Accordingly, the user may need to separately compile historical ALAE payments on open claims and adjust the ALAE IBNR. This would avoid estimating a future liability for amounts that have already been paid. We note that interim ALAE payments on claims that remain open are significantly more common than interim loss payments as claims are often litigated for several years until the final loss would be paid. To the extent that interim loss payments have been made on open claims, a similar procedure could be considered.

5.4.1. ALAE Severities

The user creates a triangle of paid ALAE on CwPL claims aggregated by report year and by age of close year. Dividing this triangle by the incremental CwPL claim count triangle from previous steps derives a triangle of CwPL ALAE severities. One then selects a CwPL ALAE severity for each age.

The same process uses the triangles of paid ALAE on CwPAO claims and the previously constructed incremental CwPAO claim count triangle in order to analyze and select CwPAO ALAE severities. Projections are found on *Appendix A, Section 8: Paid ALAE Projection for Future CwPL Claims* and *Appendix A, Section 9: Paid ALAE Projection for Future CwPAO Claims*.

No explicit adjustment is made for a limitation of ALAE payments. A user can make such an adjustment if they consider it relevant.

The development of separate ALAE severities for claims that do and do not close with Paid Loss may aid strategy in addition to considering unique characteristics of each subset of ALAE in making projections. As an example, suppose that claims that close without Paid Loss after 60 months have an average ALAE spend of $50,000, while claims that close after 60 months with Paid Loss have an average ALAE spend of $20,000 and an average loss severity of $90,000. These results may indicate an advantage of increasing the amount spent for defense.

5.4.2. Ultimate ALAE Projection

Earlier in this method description, future CwPL claim counts and future CwPAO claim counts were projected by report year and by age of close year from report year. The next step would be to apply the appropriate selected CwPL ALAE severities to the future CwPL claim counts and the appropriate selected CwPAO ALAE severities to the future CwPAO claim counts. The result is the projection of future paid ALAE by report year and by age of close year from report year seen on *Appendix A, Section 8: Paid ALAE Projection for Future CwPL Claims* and *Appendix A, Section 9: Paid ALAE Projection for Future CwPAO Claims*.

Similar to the earlier projections of CwPL and CwPAO claims counts and CwPL limited Paid Losses, the ALAE projection can be utilized easily for commonplace "actual versus expected" analyses.

5.5. Summary of Projections

The final step in the method is to aggregate the results of the projections by report year and in total as seen in *Appendix A, Section 10: Summary of Projections*.

The current claim counts, projected future counts, and ultimate counts should be shown for each of the count types used in the analysis: reported, closed, CwPL, CwPL XS, CwPAO, and CwoP. While the CwoP counts are not used explicitly in this method, they are easily derived by subtracting the CwPL and CwPAO counts from the total closed counts and may be useful in discussions about the reasonableness of results.

Similarly, for loss and ALAE, the current, projected future, and ultimate paid amounts should be shown for each of the categories analyzed: CwPL limited loss, CwPL XS loss, CwPL ALAE, and CwPAO ALAE. Additionally, the current amounts of paid loss and ALAE on open claims and the case reserves for loss and ALAE must be included, as these are required to determine the current total paid amounts and the IBNR estimate.

The examples shown in *Appendix A* do not allocate future CwPL XS counts or future CwPL XS paid loss to report year and are instead left as excess loads shown only in the totals. The user may prefer to perform a different allocation. One possible method could apply a portion of excess counts and losses to known large open claims and then allocate the remaining portion to future report years using CwPL counts as a distribution basis.

Finally, because this method requires a number of partitions of the data that could be unfamiliar to others, this exhibit should include a simplified version that shows paid loss and ALAE, case loss and ALAE, and IBNR loss and ALAE, as well as ultimate totals on a combined basis.

6. CONCLUDING REMARKS

The authors would recommend that users of this method consider further refinement of data to split into various different programs or books of business to the extent that any such segmentation (state, commercial/residential construction, AI/NI, general contractor/subcontractor, etc.) is relatively homogeneous and would be appropriate given credibility considerations.

A couple of future potential enhancements are noted below to consider in potentially making business decisions and/or appropriate estimations and conclusions:

6.1. Reopened Claim Counts

If claims reopen at a high enough frequency, it may be worthwhile to consider an explicit adjustment and projection accordingly. Determining this could be accomplished simply by adding "Reopen Year" as a named field to the loss run. It would then be possible to build reopened claim count triangles, which could be used to select ratios of reopened claims to prior closed claims by age. Reopened claims could then be projected, which would be considered in addition to open and future reported counts when determining future closed counts.

6.2. Adjusting for Negative IBNR Indications for Specific Report Years

Given that the case reserves are not considered in any part of this projection method, it is possible that specific large losses are already known and may not settle for a number of years. If the load for excess losses is allocated to report year without consideration for these known claims, the projected future loss for the noted report year could potentially be lower than the booked case reserves. The user could adjust by allocating the overall excess estimated loss amount as appropriate. One such approach could be to allocate the known case loss reserves exceeding the selected loss limit (with the remaining excess being assigned to the remaining years). Other allocations may consider more judgmental assumptions, including consideration of specific estimates for known large cases as discussed previously.

6.3. Final Thoughts

The methodologies and considerations here should give the reader an appropriate methodology to project CD liabilities, have meaningful internal company discussions, and ultimately make more appropriate business decisions. None of the assumptions supporting these projections should be considered within a vacuum both in terms of observed trends or knowledge of upcoming environmental changes.

A segment like CD is perpetually changing, and the understanding of recent and future changes can lead to a more reflective projection or even adjustments to a selected method or assumptions. The authors would welcome seeing the method detailed within this paper used by others as a "beginning" for estimating their own changing situations. We look forward to welcoming further thoughts.

7. REFERENCES

[1] Green, Michael D., Larrick, Michael, Wettstein, Carolyn D., and Bennington, Toby L., "*Reserving for Construction Defects*", Casualty Actuarial Society Forum, 2001, Fall, pages 105-152.
[2] Bornhuetter, Ronald L. and Ferguson, Ronald E., "*The Actuary and IBNR*", *PCAS* LIX, 1972, pages 181-195.
[3] Friedland, J.F., "*Estimating Unpaid Claims Using Basic Techniques*", Casualty Actuarial Society, Third Version, July 2010.
[4] Insurance Services Office, Inc., Commercial General Liability Coverage Form, 1997, CG 00 01 07 98
[5] Kozlowski, Ronald T., Towers Watson & Co., and Williams, Cedric, Deloitte Consulting LLP, "*Construction Defect Overview*", CLRS, Boston, MA, September 15-17, 2013, pages 15 – 33.
[6] Kozlowski, Ronald T., Consultant, Hornyak, Scott, Willis Tower Watson, Yu, Patrick, Willis Tower Watson, "*Are Construction Defect Claims Still Hammering the Industry*?", CLRS, Chicago, IL., September 19-20, 2016, pages 6-36.

[7] Kozlowski, Ronald T., RTK Actuarial Services and Allen, Emily, Liberty Mutual Insurance, "*Construction Defects: Learning from A Deep Dive*", CLRS, Anaheim, CA, September 5-7, 2018, pages 7-17.
[8] Salzmann, Ruth E., "*Estimating Liabilities for Losses and Loss Adjustment Expenses*", Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984.
[9] Flaherty, Brian chapter co-author with Duncan, Chris, CLCS, Farmer Woods Group "*Construction Risk Management Through Insurance*", Arizona Construction Law Practice Manual, 3rd Ed., 2016.

Abbreviations and notations

AI, Additional insured

ALAE, Allocated loss adjustment expenses

CAS, Casualty Actuarial Society

CD, Construction defect

CDF, Cumulative loss development factor

CGL, Commercial general liability

CwoP, Closed without payment

CwP, Closed with a payment

CwPAO, Closed with payment for ALAE only

CwPL, Closed with Paid Loss

CwPL XS, Closed with Paid Loss on claims with an excess component

DCC, Defense and cost containment

EIFS, Exterior insulation and finishing systems

GC, General contractor

IBNR, Incurred but not reported loss (i.e., all unreported development beyond case reserves)

ID, Identification code

ISO, Insurance Services Office

ITD, Incurred to date

KTT, Kahn-Tumbleston-Townsend Construction Defect liability projection method

LDF, Age-to-age loss development factor

MGA, Managing general agency

NI, Named insured

OCIP, Owners Controlled Insurance Program

SIR, Self-insured retention

TPA, Third party administrators

Acknowledgment

The authors would like to thank Carey McGowan for her assistance throughout this process in creating this final product as well as Ron Kozlowski, Matt Kunish, Marc Pearl, Chris Platania, and Ernie Wilson for their diligent review of this paper and several previous submissions.

In particular, special acknowledgment and thanks are also necessary for Mike Bryant, Kevin Follett, Joe Mak, and the rest of the great people from the Claims Department at RiverStone. You are truly Industry leaders in understanding Construction Defect claims, typical and unusual situations, and coverage issues with potential impact on current and future events.

Any remaining errors within this paper are exclusively those of the authors.

Biographies of Authors

James B. Kahn is Vice President, Leader Actuarial at RiverStone Resources LLC, a company specializing in runoff insurance, including providing solutions and third party administrative (TPA) claims management, expertise, and independent adjusting services to interested parties. He has a B.A. degree in Mathematics (with an application in Probability and Statistics) from the Johns Hopkins University in Baltimore, MD. He is a Fellow of the CAS and a Member of the American Academy of Actuaries. Mr. Kahn has 30 years of actuarial experience with more than 12 years directly related to runoff insurance functions. In addition, he has held various positions at consulting firms and both primary and reinsurance companies. He has authored previous papers for the CAS, served on several committees including the CAS Committee on Reserves, and been a speaker at a number of CAS conferences.

Mr. Kahn may be contacted at james_kahn@trg.com; RiverStone Resources, LLC.

Brad Tumbleston is an Actuarial Manager at RiverStone Resources LLC. He has a B.S. degree in Mathematics and a B.A. degree in Philosophy, both earned at the University of Florida in Gainesville, FL. He is a Fellow of the CAS with 12 years of actuarial experience, including 8 years in commercial run-off reserving, with a specialty in Construction Defect Liability projections.

Mr. Tumbleston may be contacted at brad_tumbleston@trg.com; RiverStone Resources, LLC.

Wilson Townsend is Vice President, Director of Operations for RiverStone Resources LLC Services. RiverStone is a provider of RiskSmart Run-Off SM solutions, third party administrative claim management, and Independent Adjusting services. Over a 32-year career, including 21 in management, Mr. Townsend has focused on Commercial Liability claims, including Construction Defect, Commercial Auto, Errors & Omissions, and Directors and Officers claims, as well as Primary and Excess Workers' Compensation claims. Over the last 19 years at RiverStone, he has served in several management roles including as a Technical Claim Manager, Business Lead of a large business transformation project, and leading a new initiative to develop services for the third-party market. Mr. Townsend earned a B.A. degree in Political Science from the University of Massachusetts at Amherst.

Mr. Townsend may be contacted at wilson_townsend@trg.com; RiverStone Resources, LLC.

Appendix A: Projection of Ultimate Liabilities

APPENDIX A, SECTION 1: Data Requirements
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019 ($000's)

Exhibit A1-1: Abridged Fictional Construction Defect Loss Run and Derived Fields

Abridged Loss Run

Claim_No	[A] Loss Year	[B] Report Year	[C] Close Year	[D] Paid Loss	[E] Paid ALAE	[F] Case Loss	[G] Case ALAE
CN000000001	2010	2010	2011	0	3	0	0
CN000000043	2010	2012	2014	584	22	0	0
CN000000059	2012	2012	2014	23	20	0	0
CN000000102	2010	2013	2013	0	1	0	0
CN000000149	2012	2014	2014	0	1	0	0
CN000000203	2010	2015	2016	0	0	0	0
CN000000222	2014	2015	2015	0	1	0	0
CN000000286	2014	2016	2016	0	1	0	0
CN000000299	2015	2016	2016	0	1	0	0
CN000000324	2015	2016	2018	0	0	0	0
CN000000335	2015	2016	2016	0	1	0	0
CN000000374	2013	2017	2018	20	17	0	0
CN000000393	2016	2017		0	0	59	48
CN000000414	2010	2017	2018	36	17	0	0
CN000000466	2015	2017		0	0	228	43

Derived Fields

Claim_No	[H] Loss-to-Report Lag	[I] Report-to-Close Lag	[J] Reported Count	[K] Open Count	[L] CwPL Count	[M] CwPL XS Count	[N] CwPAO Count	[O] CwPL Ltd Paid Loss	[P] CwPL XS Paid Loss	[Q] CwPL Paid ALAE	[R] CwPAO Paid ALAE
CN000000001	12	24	1	0	0	0	1	0	0	0	3
CN000000043	36	36	1	0	1	1	0	200	384	22	0
CN000000059	12	36	1	0	1	0	0	23	0	20	0
CN000000102	48	12	1	0	0	0	1	0	0	0	1
CN000000149	36	12	1	0	0	0	1	0	0	0	1
CN000000203	72	24	1	0	0	0	0	0	0	0	0
CN000000222	24	12	1	0	0	0	1	0	0	0	1
CN000000286	36	12	1	0	0	0	1	0	0	0	1
CN000000299	24	12	1	0	0	0	1	0	0	0	1
CN000000324	24	36	1	0	0	0	0	0	0	0	0
CN000000335	24	12	1	0	0	0	1	0	0	0	1
CN000000374	60	24	1	0	1	0	0	20	0	17	0
CN000000393	24		1	1	0	0	0	0	0	0	0
CN000000414	96	24	1	0	1	0	0	36	0	17	0
CN000000466	36		1	1	0	0	0	0	0	0	0

Derivations

[H] = ([B] - [A] + 1) * 12
[I] = ([C] - [B] + 1) * 12
[J] = 1

[K] = 1 if [C] is empty; otherwise 0
[L] = 1 if [K] = 0 and [D] > 0; otherwise 0
[M] = 1 if [L] = 1 and [D] > 200; otherwise 0
[N] = 1 if [K] = 0 and [D] = 0 and [E] > 0; otherwise 0

[O] = the lesser of [L] * [D] and 200 (the selected loss limit)
[P] = [L] * [D] - [O]
[Q] = [L] * [E]
[R] = [N] * [E]

Exhibit A1-2: Summary of Data

			Current Counts				Current Loss and ALAE							
Rept Yr	Reported	Closed	CwPL	CwPL XS	CwPAO	CwoP	CwPL Ltd Loss	CwPL XS Loss	CwPL ALAE	CwPAO ALAE	Open Paid Loss	Open Paid ALAE	Case Loss	Case ALAE
2010	12	12	3	0	6	3	175	0	81	21	0	0	0	0
2011	19	19	8	0	7	4	490	0	324	68	0	0	0	0
2012	32	32	10	2	18	4	713	509	349	114	0	0	0	0
2013	40	40	15	0	11	14	881	0	724	20	0	0	0	0
2014	54	53	24	0	21	8	1,590	0	1,068	82	0	21	30	56
2015	93	91	30	3	40	21	2,041	818	932	125	0	0	130	63
2016	108	97	27	1	43	27	1,478	166	811	124	10	19	900	623
2017	134	103	30	0	50	23	1,206	0	747	153	26	107	1,534	1,079
2018	112	71	4	0	42	25	135	0	62	71	33	85	1,379	1,219
2019	105	45	2	0	30	13	26	0	11	30	26	65	1,421	951
Total	709	563	153	6	268	142	8,734	1,493	5,108	809	96	297	5,394	3,991

Abbreviations

CwPL Closed with Paid Loss
CwPL XS Closed with Excess Paid Loss (a subset of CwPL)
CwPAO Closed with Paid ALAE Only
CwoP Closed without Pay

APPENDIX A, SECTION 2: Reported Claim Count Projection
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019

Exhibit A2-1: Cumulative Reported Counts by Age of Loss Year

Loss Yr	12	24	36	48	60	72	84	96	108	120
2010	12	22	35	43	52	69	80	86	90	91
2011	9	19	31	45	64	82	94	95	100	
2012	9	22	32	48	64	81	85	88		
2013	7	13	23	40	66	83	91			
2014	15	30	48	74	100	133				
2015	16	33	55	84	102					
2016	11	22	34	47						
2017	14	25	35							
2018	8	15								
2019	7									

Exhibit A2-2: Age-to-Age Development Factors

Loss Yr	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120
2010	1.833	1.591	1.229	1.209	1.327	1.159	1.075	1.047	1.011
2011	2.111	1.632	1.452	1.422	1.281	1.146	1.011	1.053	
2012	2.444	1.455	1.500	1.333	1.266	1.049	1.035		
2013	1.857	1.769	1.739	1.650	1.258	1.096			
2014	2.000	1.600	1.542	1.351	1.330				
2015	2.063	1.667	1.527	1.214					
2016	2.000	1.545	1.382						
2017	1.786	1.400							
2018	1.875								
2019									
Selected	1.997	1.582	1.482	1.363	1.292	1.113	1.040	1.050	1.011

Exhibit A2-3: Cumulative Reported Counts by Age of Loss Year, Developed to Ultimate
Derived from Exhibits A2-1 and A2-2

Loss Yr	12	24	36	48	60	72	84	96	108	120
2010	12	22	35	43	52	69	80	86	90	91
2011	9	19	31	45	64	82	94	95	100	101
2012	9	22	32	48	64	81	85	88	92	93
2013	7	13	23	40	66	83	91	95	100	101
2014	15	30	48	74	100	133	148	154	162	164
2015	16	33	55	84	102	132	147	153	161	163
2016	11	22	34	47	64	83	92	96	101	102
2017	14	25	35	52	71	92	102	106	111	112
2018	8	15	24	36	49	63	70	73	77	78
2019	7	14	22	33	45	58	65	68	71	72

Exhibit A2-4: Incremental Reported Counts by Age of Loss Year, Developed to Ultimate
Derived from Exhibits A2-3

Loss Yr	12	24	36	48	60	72	84	96	108	120
2010	12	10	13	8	9	17	11	6	4	1
2011	9	10	12	14	19	18	12	1	5	1
2012	9	13	10	16	16	17	4	3	4	1
2013	7	6	10	17	26	17	8	4	5	1
2014	15	15	18	26	26	33	15	6	8	2
2015	16	17	22	29	18	30	15	6	8	2
2016	11	11	12	13	17	19	9	4	5	1
2017	14	11	10	17	19	21	10	4	5	1
2018	8	7	9	12	13	14	7	3	4	1
2019	7	7	8	11	12	13	7	3	3	1

Exhibit A2-5: Ultimate Reported Counts by Report Year
Sums of diagonals of Exhibit A2-4

Rept Yr	Ult Rptd Counts	Rept Yr	Ult Rptd Counts	Rept Yr	Ult Rptd Counts	Rept Yr	Ult Rptd Counts
2010	12	2015	93	2020	104	2025	16
2011	19	2016	108	2021	85	2026	8
2012	32	2017	134	2022	69	2027	4
2013	40	2018	112	2023	50	2028	1
2014	54	2019	105	2024	31		

APPENDIX A, SECTION 3: Closed Claim Count Projection
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019

Exhibit A3-1: Cumulative Closed Counts by Age of Report Year
Ultimate Reported Counts taken from Exhibit A2-5

Rept Yr	12	24	36	48	60	72	84	96	108	120	Ult Rptd Counts
2010	4	10	10	12	12	12	12	12	12	12	12
2011	6	10	12	15	18	19	19	19	19		19
2012	12	21	28	31	32	32	32	32			32
2013	15	29	33	37	40	40	40				40
2014	18	28	36	44	53	53					54
2015	42	62	76	88	91						93
2016	37	70	86	97							108
2017	39	76	103								134
2018	50	71									112
2019	45										105

Exhibit A3-2: Cumulative Closed Ratios (Closed / Ultimate Reported)

Rept Yr	12	24	36	48	60	72	84	96	108	120
2010	33%	83%	83%	100%	100%	100%	100%	100%	100%	100%
2011	32%	53%	63%	79%	95%	100%	100%	100%	100%	
2012	38%	66%	88%	97%	100%	100%	100%	100%		
2013	38%	73%	83%	93%	100%	100%	100%			
2014	33%	52%	67%	81%	98%	98%				
2015	45%	67%	82%	95%	98%					
2016	34%	65%	80%	90%						
2017	29%	57%	77%							
2018	45%	63%								
2019	43%									
Selected	37%	64%	78%	91%	98%	100%	100%	100%	100%	100%
Incrmntl	37%	27%	14%	13%	8%	1%	0%	0%	0%	0%

Exhibit A3-3: Incremental Closed Counts by Report-to-Close Lag, Developed to Ultimate
Above the dotted line: actual incremental closed counts by age
RYs 2016 and prior, age 60+: ultimate reported minus cumulative closed at age 48
All other: ultimate reported minus current closed for the report year, allocated to future age proportional to incremental close ratios, adding actual current closed counts in for age 60+ for RYs 2015 and prior

Rept Yr	12	24	36	48	60+	Ult Rptd Counts
2010	4	6	0	2	0	12
2011	6	4	2	3	4	19
2012	12	9	7	3	1	32
2013	15	14	4	4	3	40
2014	18	10	8	8	10	54
2015	42	20	14	12	5	93
2016	37	33	16	11	11	108
2017	39	37	27	18	13	134
2018	50	21	15	15	11	112
2019	45	26	13	12	9	105
2020	38	28	14	13	10	104
2021	31	23	11	11	8	85
2022	25	19	9	9	6	69
2023	18	14	7	6	5	50
2024	11	8	4	4	3	31
2025	6	4	2	2	2	16
2026	3	2	1	1	1	8
2027	1	1	1	1	0	4
2028	0	0	0	0	0	1

Examples
RY 2015, age 60+: 93 - 88 = 5
RY 2016, age 60+: 108 - 97 = 11
RY 2017, age 48:
(134 - 103) * 13% / (13% + 8% + 1%) = 18

RY 2020, age 36:
(104 - 0) * 14% / (37% + 27% + … + 1%)
= 14

APPENDIX A, SECTION 4: Closed with Paid Loss (CwPL) and Closed with Paid ALAE Only (CwPAO) Ratio Selections
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019

Exhibit A4-1: Incremental Closed Counts by Report-to-Close Lag
Derived from Exhibit A3-1

Rept Yr	12	24	36	48	60+
2010	4	6	0	2	0
2011	6	4	2	3	4
2012	12	9	7	3	1
2013	15	14	4	4	3
2014	18	10	8	8	9
2015	42	20	14	12	3
2016	37	33	16	11	
2017	39	37	27		
2018	50	21			
2019	45				

Exhibit A4-2: Incremental CwPL Counts by Report-to-Close Lag

Rept Yr	12	24	36	48	60+
2010	0	1	0	2	0
2011	0	1	1	2	4
2012	0	3	4	2	1
2013	0	4	4	4	3
2014	1	2	6	6	9
2015	1	8	8	10	3
2016	1	6	11	9	
2017	0	15	15		
2018	1	3			
2019	2				

Exhibit A4-3: Incremental CwPL Ratios (CwPL / Closed)

Rept Yr	12	24	36	48	60+
2010	0%	17%	0%	100%	0%
2011	0%	25%	50%	67%	100%
2012	0%	33%	57%	67%	100%
2013	0%	29%	100%	100%	100%
2014	6%	20%	75%	75%	100%
2015	2%	40%	57%	83%	100%
2016	3%	18%	69%	82%	
2017	0%	41%	56%		
2018	2%	14%			
2019	4%				
Selected	2%	28%	63%	81%	100%

Exhibit A4-4: Incremental CwPAO Counts by Report-to-Close Lag

Rept Yr	12	24	36	48	60+
2010	2	4	0	0	0
2011	2	3	1	1	0
2012	8	6	3	1	0
2013	8	3	0	0	0
2014	12	5	2	2	0
2015	24	10	5	1	0
2016	22	15	4	2	
2017	28	15	7		
2018	31	11			
2019	30				

Exhibit A4-5: Incremental CwPAO Ratios (CwPAO / Closed)

Rept Yr	12	24	36	48	60+
2010	50%	67%	0%	0%	0%
2011	33%	75%	50%	33%	0%
2012	67%	67%	43%	33%	0%
2013	53%	21%	0%	0%	0%
2014	67%	50%	25%	25%	0%
2015	57%	50%	36%	8%	0%
2016	59%	45%	25%	18%	
2017	72%	41%	26%		
2018	62%	52%			
2019	67%				
Selected	62%	47%	28%	16%	0%

APPENDIX A, SECTION 5: Future Count Projections
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019

Exhibit A5-1: Incremental Future Closed Counts by Report-to-Close Lag
Exhibit A3-3 minus Exhibit A4-1

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	1
2015	0	0	0	0	2
2016	0	0	0	0	11
2017	0	0	0	18	13
2018	0	0	15	15	11
2019	0	26	13	12	9
2020	38	28	14	13	10
2021	31	23	11	11	8
2022	25	19	9	9	6
2023	18	14	7	6	5
2024	11	8	4	4	3
2025	6	4	2	2	2
2026	3	2	1	1	1
2027	1	1	1	1	0
2028	0	0	0	0	0

Exhibit A5-2: Incremental Future CwPL Counts by Report-to-Close Lag
Product of Exhibit A5-1 and Selected CwPL Ratios in Exhibit A4-3

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	1
2015	0	0	0	0	2
2016	0	0	0	0	11
2017	0	0	0	15	13
2018	0	0	10	12	11
2019	0	7	8	10	9
2020	1	8	9	11	10
2021	1	6	7	9	8
2022	1	5	6	7	6
2023	0	4	4	5	5
2024	0	2	3	3	3
2025	0	1	1	2	2
2026	0	1	1	1	1
2027	0	0	0	0	0
2028	0	0	0	0	0

Exhibit A5-3: Incremental Future CwPAO Counts by Report-to-Close Lag
Product of Exhibit A5-1 and Selected CwPAO Ratios in Exhibit A4-5

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	0
2015	0	0	0	0	0
2016	0	0	0	0	0
2017	0	0	0	3	0
2018	0	0	4	2	0
2019	0	12	4	2	0
2020	24	13	4	2	0
2021	20	11	3	2	0
2022	16	9	3	1	0
2023	12	6	2	1	0
2024	7	4	1	1	0
2025	4	2	1	0	0
2026	2	1	0	0	0
2027	1	1	0	0	0
2028	0	0	0	0	0

APPENDIX A, SECTION 6: Limited Paid Loss Projection
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019 ($000's)

Exhibit A6-1: Incremental CwPL Counts by Report-to-Close Lag
Identical to Exhibit A4-2

Rept Yr	12	24	36	48	60+
2010	0	1	0	2	0
2011	0	1	1	2	4
2012	0	3	4	2	1
2013	0	4	4	4	3
2014	1	2	6	6	9
2015	1	8	8	10	3
2016	1	6	11	9	
2017	0	15	15		
2018	1	3			
2019	2				

Exhibit A6-2: Ltd Paid Loss on CwPL Claims by Report-to-Close Lag

Rept Yr	12	24	36	48	60+
2010	0	38	0	137	0
2011	0	29	51	170	239
2012	0	108	297	239	69
2013	0	188	123	204	366
2014	11	50	220	466	843
2015	10	251	388	939	453
2016	8	193	742	534	
2017	0	652	554		
2018	27	108			
2019	26				

Exhibit A6-3: CwPL Limited Loss Severities

Rept Yr	12	24	36	48	60+
2010	0	38	0	68	0
2011	0	29	51	85	60
2012	0	36	74	120	69
2013	0	47	31	51	122
2014	11	25	37	78	94
2015	10	31	49	94	151
2016	8	32	67	59	
2017	0	43	37		
2018	27	36			
2019	13				
Selected	14	38	48	77	99

Exhibit A6-4: Incremental Future CwPL Counts by Report-to-Close Lag
Identical to Exhibit A5-2

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	1
2015	0	0	0	0	2
2016	0	0	0	0	11
2017	0	0	0	15	13
2018	0	0	10	12	11
2019	0	7	8	10	9
2020	1	8	9	11	10
2021	1	6	7	9	8
2022	1	5	6	7	6
2023	0	4	4	5	5
2024	0	2	3	3	3
2025	0	1	1	2	2
2026	0	1	1	1	1
2027	0	0	0	0	0
2028	0	0	0	0	0

Exhibit A6-5: Incremental Future Ltd Paid Loss by Report-to-Close Lag
Product of Exhibit A6-4 and Selected CwPL Limited Loss Severities in Exhibit A6-3

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	99
2015	0	0	0	0	197
2016	0	0	0	0	1,084
2017	0	0	0	1,123	1,285
2018	0	0	471	926	1,059
2019	0	272	391	769	881
2020	12	297	428	841	963
2021	10	243	350	688	787
2022	8	197	284	558	639
2023	6	143	206	404	463
2024	3	89	128	251	287
2025	2	46	66	129	148
2026	1	23	33	65	74
2027	0	11	16	32	37
2028	0	3	4	8	9

APPENDIX A, SECTION 7: Excess Paid Loss Projection
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019 ($000's)

Exhibit A7-1: Incremental CwPL Counts by Report-to-Close Lag
Identical to Exhibit A4-2

Rept Yr	12	24	36	48	60+
2010	0	1	0	2	0
2011	0	1	1	2	4
2012	0	3	4	2	1
2013	0	4	4	4	3
2014	1	2	6	6	9
2015	1	8	8	10	3
2016	1	6	11	9	
2017	0	15	15		
2018	1	3			
2019	2				

Exhibit A7-2: Incremental CwPL XS Counts by Report-to-Close Lag

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	1	1	0
2013	0	0	0	0	0
2014	0	0	0	0	0
2015	0	0	0	2	1
2016	0	0	1	0	
2017	0	0	0		
2018	0	0			
2019	0				

Exhibit A7-3: Excess Paid Loss on CwPL Claims by Report-to-Close Lag

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	384	125	0
2013	0	0	0	0	0
2014	0	0	0	0	0
2015	0	0	0	444	374
2016	0	0	166	0	
2017	0	0	0		
2018	0	0			
2019	0				

Exhibit A7-4: CwPL XS Frequency and Severity Selections

	Frequency	Severity
All-Year Weighted Average	3.9%	249
Wtd Avg of RYs 2015-2019	4.3%	246
Selection	3.9%	249

Exhibit A7-5: CwPL XS Load Development

Future CwPL Counts	244	*Sum of Exhibit A6-4*
Future CwPL XS Counts	10	*Product of Future CwPL Counts and Selected CwPL XS Frequency in Exhibit A7-4*
Future CwPL Excess Paid Loss	2,381	*Product of Future CwPL XS Counts and Selected CwPL XS Severity in Exhibit A7-4*

APPENDIX A, SECTION 8: Paid ALAE Projection for Future Closed with Paid Loss (CwPL) Claims
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019 ($000's)

Exhibit A8-1: Incremental CwPL Counts by Report-to-Close Lag
Identical to Exhibit A6-1

Rept Yr	12	24	36	48	60+
2010	0	1	0	2	0
2011	0	1	1	2	4
2012	0	3	4	2	1
2013	0	4	4	4	3
2014	1	2	6	6	9
2015	1	8	8	10	3
2016	1	6	11	9	
2017	0	15	15		
2018	1	3			
2019	2				

Exhibit A8-2: Paid ALAE on CwPL Claims by Report-to-Close Lag

Rept Yr	12	24	36	48	60+
2010	0	9	0	72	0
2011	0	16	17	57	234
2012	0	73	78	94	103
2013	0	89	110	287	239
2014	3	89	125	197	654
2015	4	133	221	416	158
2016	4	121	264	423	
2017	0	329	418		
2018	3	60			
2019	11				

Exhibit A8-3: CwPL ALAE Severities

Rept Yr	12	24	36	48	60+
2010	0	9	0	36	0
2011	0	16	17	29	58
2012	0	24	20	47	103
2013	0	22	27	72	80
2014	3	45	21	33	73
2015	4	17	28	42	53
2016	4	20	24	47	
2017	0	22	28		
2018	3	20			
2019	5				
Selected	4	21	25	44	69

Exhibit A8-4: Incremental Future CwPL Counts by Report-to-Close Lag
Identical to Exhibit A6-4

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	1
2015	0	0	0	0	2
2016	0	0	0	0	11
2017	0	0	0	15	13
2018	0	0	10	12	11
2019	0	7	8	10	9
2020	1	8	9	11	10
2021	1	6	7	9	8
2022	1	5	6	7	6
2023	0	4	4	5	5
2024	0	2	3	3	3
2025	0	1	1	2	2
2026	0	1	1	1	1
2027	0	0	0	0	0
2028	0	0	0	0	0

Exhibit A8-5: Incremental Future CwPL Paid ALAE by Report-to-Close Lag
Product of Exhibit A8-4 and Selected CwPL ALAE Severities in Exhibit A8-3

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	69
2015	0	0	0	0	139
2016	0	0	0	0	763
2017	0	0	0	645	905
2018	0	0	244	532	746
2019	0	155	203	442	620
2020	3	169	222	483	678
2021	3	138	181	395	554
2022	2	112	147	321	450
2023	2	81	107	232	326
2024	1	50	66	144	202
2025	1	26	34	74	104
2026	0	13	17	37	52
2027	0	7	9	19	26
2028	0	2	2	5	7

Exhibit A9-1: Incremental CwPAO Counts by Report-to-Close Lag

Identical to Exhibit A4-4

Rept Yr	12	24	36	48	60+
2010	2	4	0	0	0
2011	2	3	1	1	0
2012	8	6	3	1	0
2013	8	3	0	0	0
2014	12	5	2	2	0
2015	24	10	5	1	0
2016	22	15	4	2	
2017	28	15	7		
2018	31	11			
2019	30				

Exhibit A9-2: Paid ALAE on CwPAO Claims by Report-to-Close Lag

Rept Yr	12	24	36	48	60+
2010	2	19	0	0	0
2011	2	13	14	40	0
2012	8	40	34	32	0
2013	8	12	0	0	0
2014	13	12	36	22	0
2015	23	33	51	17	0
2016	25	44	24	30	
2017	27	69	58		
2018	35	36			
2019	30				

Exhibit A9-3: CwPAO ALAE Severities

Rept Yr	12	24	36	48	60+
2010	1	5	0	0	0
2011	1	4	14	40	0
2012	1	7	11	32	0
2013	1	4	0	0	0
2014	1	2	18	11	0
2015	1	3	10	17	0
2016	1	3	6	15	
2017	1	5	8		
2018	1	3			
2019	1				
Selected	1	4	10	20	30

Exhibit A9-4: Incremental Future CwPAO Counts by Report-to-Close Lag

Identical to Exhibit A5-3

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	0
2015	0	0	0	0	0
2016	0	0	0	0	0
2017	0	0	0	3	0
2018	0	0	4	2	0
2019	0	12	4	2	0
2020	24	13	4	2	0
2021	20	11	3	2	0
2022	16	9	3	1	0
2023	12	6	2	1	0
2024	7	4	1	1	0
2025	4	2	1	0	0
2026	2	1	0	0	0
2027	1	1	0	0	0
2028	0	0	0	0	0

Exhibit A9-5: Incremental Future CwPAO Paid ALAE by Report-to-Close Lag

Product of Exhibit A9-4 and Selected CwPAO ALAE Severities in Exhibit A9-3

Rept Yr	12	24	36	48	60+
2010	0	0	0	0	0
2011	0	0	0	0	0
2012	0	0	0	0	0
2013	0	0	0	0	0
2014	0	0	0	0	0
2015	0	0	0	0	0
2016	0	0	0	0	0
2017	0	0	0	59	0
2018	0	0	43	48	0
2019	0	47	36	40	0
2020	25	51	39	44	0
2021	20	42	32	36	0
2022	16	34	26	29	0
2023	12	25	19	21	0
2024	7	15	12	13	0
2025	4	8	6	7	0
2026	2	4	3	3	0
2027	1	2	2	2	0
2028	0	0	0	0	0

APPENDIX A, SECTION 10: Summary of Projections
Company XYZ Construction Defect Portfolio, Evaluated as of December 31, 2019 ($000's)

Exhibit A10-1: Summary of Projections

	Current Counts						Future Counts						Ultimate Counts					
Rept Yr	Reported	Closed	CwPL	CwPL XS	CwPAO	CwoP	Reported	Closed	CwPL	CwPL XS	CwPAO	CwoP	Reported	Closed	CwPL	CwPL XS	CwPAO	CwoP
2010	12	12	3	0	6	3	0	0	0		0	0	12	12	3		6	3
2011	19	19	8	0	7	4	0	0	0		0	0	19	19	8		7	4
2012	32	32	10	2	18	4	0	0	0		0	0	32	32	10		18	4
2013	40	40	15	0	11	14	0	0	0		0	0	40	40	15		11	14
2014	54	53	24	0	21	8	0	1	1		0	0	54	54	25		21	8
2015	93	91	30	3	40	21	0	2	2		0	0	93	93	32		40	21
2016	108	97	27	1	43	27	0	11	11		0	0	108	108	38		43	27
2017	134	103	30	0	50	23	0	31	28		3	0	134	134	58		53	23
2018	112	71	4	0	42	25	0	41	33		7	2	112	112	37		49	27
2019	105	45	2	0	30	13	0	60	34		18	8	105	105	36		48	21
2020							104	104	38		43	22	104	104	38		43	22
2021							85	85	31		35	18	85	85	31		35	18
2022							69	69	25		29	15	69	69	25		29	15
2023							50	50	18		21	11	50	50	18		21	11
2024							31	31	11		13	7	31	31	11		13	7
2025							16	16	6		7	3	16	16	6		7	3
2026							8	8	3		3	2	8	8	3		3	2
2027							4	4	1		2	1	4	4	1		2	1
2028							1	1	0		0	0	1	1	0		0	0
Total	709	563	153	6	268	142	368	514	244	10	181	89	1,077	1,077	397	16	449	231

	Current Loss and ALAE								Future Loss and ALAE				Ultimate Loss and ALAE			
Rept Yr	CwPL Ltd Loss	CwPL XS Loss	CwPL ALAE	CwPAO ALAE	Open Pd Loss	Open Pd ALAE	Case Loss	Case ALAE	CwPL Ltd Loss	CwPL XS Loss	CwPL ALAE	CwPAO ALAE	CwPL Ltd Loss	CwPL XS Loss	CwPL ALAE	CwPAO ALAE
2010	175	0	81	21	0	0	0	0	0		0	0	175		81	21
2011	490	0	324	68	0	0	0	0	0		0	0	490		324	68
2012	713	509	349	114	0	0	0	0	0		0	0	713		349	114
2013	881	0	724	20	0	0	0	0	0		0	0	881		724	20
2014	1,590	0	1,068	82	0	21	30	56	99		69	0	1,689		1,138	82
2015	2,041	818	932	125	0	0	130	63	197		139	0	2,238		1,071	125
2016	1,478	166	811	124	10	19	900	623	1,084		763	0	2,562		1,574	124
2017	1,206	0	747	153	26	107	1,534	1,079	2,408		1,550	59	3,614		2,297	212
2018	135	0	62	71	33	85	1,379	1,219	2,456		1,522	92	2,590		1,584	163
2019	26	0	11	30	26	65	1,421	951	2,313		1,420	123	2,339		1,430	153
2020									2,541		1,556	159	2,541		1,556	159
2021									2,077		1,271	130	2,077		1,271	130
2022									1,686		1,032	106	1,686		1,032	106
2023									1,222		748	76	1,222		748	76
2024									757		464	47	757		464	47
2025									391		239	24	391		239	24
2026									195		120	12	195		120	12
2027									98		60	6	98		60	6
2028									24		15	2	24		15	2
Total	8,734	1,493	5,108	809	96	297	5,394	3,991	17,547	2,381	10,967	836	26,282	3,874	16,075	1,645

	Loss	ALAE	Total
Paid	10,323	6,214	16,537
Case	5,394	3,991	9,385
IBNR	14,438	7,516	21,954
Ultimate	30,156	17,720	47,876

Derivations

Current Counts and Current Loss and ALAE are identical to Exhibit A1-2

Future Reported Counts derived in Exhibit A2-5
Future Closed Counts derived in Exhibit A5-1
Future CwPL Counts derived in Exhibit A5-2
Future CwPL XS Counts derived in Exhibit A7-5
Future CwPAO Counts derived in Exhibit A5-3
Future CwoP Counts = Future Closed - Future CwPL Counts - Future CwPAO Counts

Future CwPL Ltd Loss derived in Exhibit A6-5
Future CwPL XS Loss derived in Exhibit A7-5
Future CwPL ALAE derived in Exhibit A8-5
Future CwPAO ALAE derived in Exhibit A9-5

Ultimate Counts and Ultimate Loss and ALAE derived by adding Current and Future

Paid Loss = Current CwPL Ltd Loss + Current CwPL XS Loss + Open Paid Loss
Paid ALAE = Current CwPL ALAE + Current CwPAO ALAE + Open Paid ALAE
Ultimate Loss = Ultimate CwPL Ltd Loss + Ultimate CwPL XS Loss
Ultimate ALAE = Ultimate CwPL ALAE + Ultimate CwPAO ALAE
IBNR Loss = Ultimate Loss - Paid Loss - Case Loss
IBNR ALAE = Ultimate ALAE - Paid ALAE - Case ALAE

Appendix B: Common CD Endorsements and Exclusions

The following list of Endorsements and Exclusions are commonly used to change the scope of coverage, either by limiting or transferring liability associated with Contractors. Differences may exist within the language of insurers' endorsements. However, all may change evaluation of liabilities in Construction Defect cases, specifically regarding changes from previously seen emergence:

Condominium/Town House Exclusion – In order to avoid large projects with multiple residential units, insurers started including the exclusions to limit coverage for multi-family projects. Apartment complexes were added to the endorsement by many insurers to further limit exposure to multi-family dwellings.

Continuous Injury Exclusion – This exclusion pertains to latent types of claims where the ongoing damage takes place throughout a specified period whereby an incident or occurrence (or accordingly multiple incidents or occurrences) could be considered to be taking place at any given point of time during a policy period.

Contractors Limitation Endorsement – Insurers sometimes look to combine several common industry exclusions within one specific endorsement labelled as a "contractor's limitation endorsement". These now typically include common CD types of exclusions (EIFS, mold, residential construction, etc.) that are often included on a stand-alone basis.

Exterior Insulation and Finishing Systems (EIFS) Exclusion – Insurers provide language for exclusion to damage associated with Exterior Insulation and Finishing Systems. Such claims associated with exterior walls (including stucco) have been quite rampant in terms of CD litigation, especially in regions which are prone to high humidity (where water condensation often leads to drywall damage or mold to the underlying property).

Known Loss Exclusion – Known injury or damage endorsements specifically exclude coverage for losses or potential losses, which the insured was aware of prior to the policy period.

Mold Exclusion – Insurers have recently been providing language to exclude mold damage. There are sometimes inconsistencies both in terms of types of contractors that would have the exclusion as well as broadness of coverage. Many use the standard form provided by ISO that pertains to fungi, mold, or bacteria that causes injury or damage. This form generally also excludes associated cleanup costs associated with bacteria or fungi.

Prior Work Exclusion – The exclusion eliminates coverage for injury or damage resulting from the insured's work that was completed before a stated date.

Residential Construction Exclusion – Insurers exclude coverage for specific types of coverage (apartments, condos, construction projects of X units or more, etc.) including being an additional insured on the policy of a subcontractor. These endorsements were often included on commercial subcontractors' policies.

Machine Learning, Regression Models, and Prediction of Claims Reserves

Mathias Lindholm, Richard Verrall, Felix Wahl and Henning Zakrisson

Abstract

The current paper introduces regression based reserving models that allow for separate RBNS and IBNR reserves based on aggregated discrete time data containing information about accident years, reporting years, and payment delay, since reporting. All introduced models will be closely related to the cross-classified over-dispersed Poisson (ODP) chain-ladder model. More specifically, two types of models are introduced (i) models consisting of an explicit claim count part, where payments, in a second step, are modelled conditionally on claim counts, and (ii) models defined directly in terms of claim payments without using claim count information. Further, these general ODP models will be estimated using regression functions defined by (i) tree-based gradient boosting machines (GBM), and (ii) feed-forward neural networks (NN). This will provide us with machine learning based reserving models that have interpretable output, and that are easy to bootstrap from. In the current paper we will give a brief introduction to GBMs and NNs, including calibration and model selection. All of this is illustrated in a longer numerical simulation study, which shows the benefits that can be gained by using machine learning based reserving models.

Keywords. Claims reserving, Reported But Not Settled Claims, Incurred But Not Reported Claims, Gradient Boosting Machines, Neural Networks

1. INTRODUCTION

Claims reserving is a major component of the assessment of e.g. solvency and capital adequacy. The uncertainty of the reserves has become an important part of this process, and many statistical models and approaches have been proposed in the literature. Underpinning all of these are the assumptions about the claims processes, and they are often used in conjunction with management information and intervention. This may require judgments to be made about how much to trust the data, what to change from what the data tells you before simulating possible future outcomes, and so on. In this paper, we seek to explore some of these issues using different regression based reserving models combined with machine learning techniques. In order to assess how well these approaches work, we have used the neural network based simulation machine calibrated to Swiss insurance data introduced in [12].

To be more specific, the methodological starting point of the present paper is regression based reserving models that allow for separate reported but not settled (RBNS) reserves and incurred but not settled (IBNR) reserves that make use of count data. Examples of such models are the ones introduced in [26, 22, 27, 17]. Something all these models have in common is that they rely on a description of detailed claim dynamics, but the resulting reserve predictors may be defined only in terms of incremental yearly accident year and development year data for payments and number of reported claims. In particular, the assumptions about individual claim dynamics result in aggregated incremental payments that can be expressed as *conditional* general(ized) linear models, conditional on observed claim counts. Intuitively this means that the observed claim counts will function as exposures. It is also important to stress that due to the detailed constructive motivations used in [26, 22, 27, 17] all model parameters have clear interpretations that are easy to communicate to non-actuaries. Further, due to the conditional linear model structure it is possible to estimate all parameters using e.g. (quasi-) likelihood theory or generalized least squares techniques. Without going into details, the models from [26, 22, 27, 17] allow us to produce

- RBNS reserves, by predicting future remaining payments stemming from the already observed claim counts,
- IBNR reserves, by first predicting the expected number of IBNR claim counts, and then assuming that claim payments behave like the previously described RBNS payments.

In the paper [27], payments are assumed to occur according to a discrete time Poisson process, which allows us to obtain an over-dispersed Poisson (ODP) model for the incremental payments, together with an ODP model for the claim counts. If one wishes, this latter model may be parametrized so that it coincides with a standard cross-classified Poisson chain-ladder model, see e.g. [24]. Thus,

- the model from [27] may be expressed in terms of two standard ODP regression models,
- separate RBNS and IBNR reserves are straightforward to produce,
- all parameters have meaningful interpretations that can be communicated to non-experts.

The main contribution of the present paper is that we will use the ODP regression formulation of the model from [27] together with the underlying modelling ideas to formulate regression based machine learning reserving models that

(a) produce separate RBNS and IBNR reserves,

(b) are interpretable and, to some extent, possible to combine with expert judgement,

(c) are easy to bootstrap.

The methodological focus will be on tree-based gradient boosting machines (GBM) and neural networks (NN), which will be given a brief introduction in this paper, mainly following the exposition in [7]. The overall ambition is to provide a smooth transition from using more standard Poisson regression type models to more complex algorithmic versions, and to describe how these more complex methods may be calibrated, evaluated, interpreted and compared with more standard models. With this said, our focus will primarily be on how to define models and combine these with complex algorithmic regression functions, and how these can be estimated, discussing underlying principles. All numerical illustrations will be based on simulated data produced using a neural network model calibrated to Swiss insurance data, see [12], which will make all our results reproducible.

The disposition of the paper is as follows. Section 1.1 gives a literature overview, Section 2 introduces the regression based reserving models that will be used, and which will be combined with machine learning methods in Section 3. In Section 4 a longer simulation study is used to illustrate how the regression based machine learning models can be calibrated and how to evaluate their performance. Concluding remarks are given in Section 5, and supplementary analyses and technical remarks are given in the appendix.

1.1 Literature overview

The modelling approach taken in the present paper can be thought of as being on an intermediate level compared with pure "macro" models, such as chain-ladder technique models (e.g. [19, 24]), and pure "micro" (individual claim) models, such as point process models (e.g. [2, 23, 16, 1]). For a deeper discussion on pros and cons with using individual level information in claims reserving, see e.g. [1, 25]. The current paper is primarily focused on modelling based on information on the level used in [4], where not only information about accident years and reporting years is used, but also information concerning payment delay since reporting – still, working in discrete time. We believe that this level of granularity is the first natural extension when moving away from pure (discrete-time) macro models. Further, from a practical point of view, this level of information is easy to deal with, still corresponding to standard grouping and aggregation of data, which is much less detailed and complex

to handle than continuous-time micro level data. Due to this, the modelling approach taken here will remain close to those used for standard macro reserving models, and in particular be close to the cross-classified ODP (ccODP) chain-ladder model, see [24].

When turning to machine learning techniques, focus will be on tree-based gradient boosting machines (GBM), see e.g. [9], and feed-forward neural networks (NN), see e.g. [13], and general references that cover both techniques (and much more) are e.g. [14, 7]. Both of these techniques belong to the class of "supervised learning" methods that can be thought of as methods to perform flexible curve (or surface) fitting, subject to a chosen loss function. This makes these methods natural to combine with the ODP regression models we have in mind, since given an ODP model structure, flexible machine learning based regression functions can be used to approximate the true, unobservable mean function, by choosing an ODP likelihood as loss function. How this is can be done in general for GBMs and NNs is discussed in e.g. [7, Ch. 17.3, Ch. 18]. When turning to reserving, [11] uses this idea and introduces a feed-forward neural network extension of the ODP chain-ladder model from [24], and describe how to calibrate and "train" the NN based on partially observed claims data. See also [30] for a condensed discussion on how to combine GLMs and NNs. As mentioned previously, there is a close connection between the ccOPD chain-ladder model and the so-called collective reserving model (CRM) from [27], which will make it natural to use similar techniques in the current setting. Other approaches working with deep neural networks on aggregated data can be found in e.g. [29] which discusses an extension of Mack's distribution-free chain-ladder model, [15] which uses aggregated data and gated recurrent units, and [10] which is closer to the ccODP model but also include a neural network model for count data. An example of an individual loss reserving model using deep neural networks is given in [28]. For examples of other machine learning techniques applied to individual loss reserving, see e.g. [18, 5] which uses trees, and [6] which uses tree-based gradient boosting machines. For a more general discussion of individual loss reserving and machine learning techniques, see e.g. [25] and the references therein.

2. REGRESSION BASED RESERVING MODELS

The basic building block in the present paper is regression models, and in particular over-dispersed Poisson (ODP) regression models. That is, if we let $\boldsymbol{Y}$ denote the response of interest, and let $\boldsymbol{c} = (c_1, \ldots, c_p)'$ be an arbitrary covariate vector such that $\boldsymbol{c} \in \boldsymbol{C}$, we then say that $\boldsymbol{Y} \mid \boldsymbol{c} \sim \text{ODP}(\boldsymbol{\mu}(\boldsymbol{c}; \boldsymbol{\beta}), \boldsymbol{\rho})$ with *link-function* $\boldsymbol{u}(\cdot)$, if

$$\boldsymbol{u}(\mathbb{E}[\boldsymbol{Y} \mid \boldsymbol{c}]) = \boldsymbol{\mu}(\boldsymbol{c}; \boldsymbol{\beta}),$$

where $\mu(\boldsymbol{c};\boldsymbol{\beta}) > 0$ is some arbitrary mean function, and

$$\mathbb{E}[Y \mid \boldsymbol{c}] = \text{Var}(Y \mid \boldsymbol{c})/\rho.$$

Note that we will sometimes stress the dependence on the parameters by using the notation $\mathbb{E}[Y \mid \boldsymbol{c};\boldsymbol{\beta}]$ and $\text{Var}(Y \mid \boldsymbol{c};\boldsymbol{\beta},\rho)$.

Remark 1

(a) The "Poisson" part of ODP comes from that if $Y \mid \boldsymbol{c} \sim \text{Po}(\mu(\boldsymbol{c};\boldsymbol{\beta}))$*, it holds that* $\mathbb{E}[Y \mid \boldsymbol{c}] = \text{Var}(Y \mid \boldsymbol{c})$*. From this perspective, the parameter* ρ *introduces an additional degree of freedom which allows for "over-dispersion". For more on ODP models, see e.g. [20] or [8, 3], in an actuarial context.*

(b) ODP models may be estimated using quasi-likelihood theory so that $\mu(\boldsymbol{c};\boldsymbol{\beta})$ *may be estimated with a* standard *Poisson likelihood. The over-dispersion parameter* ρ *may thereafter be estimated using Pearson (or deviance) residuals.*

Concerning reserving models more specifically, as mentioned in the introduction the so-called "collective reserving model" (CRM) from [27] will serve as the starting point for our analyses. Let $X_{i,j}$ denote the total amount of payments from claims that occurred during accident year i, $i = 1,\dots,m$, that where made during development year j, $j = 0,\dots,m-1+d$, where d corresponds to the maximal delay of a payment since the time since reporting. Further, let $N_{i,j}$ denote the total number of claims from accident year i, $i = 1,\dots,m$, that are reported j, $j = 0,\dots,m-1$, years later. Moreover, let $\mathcal{N}_0 := \sigma\{N_{i,j} : i + j \leq m, i = 1,\dots,m, j = 0,\dots,m-1\}$ and $\mathcal{N} := \sigma\{N_{i,j} : i = 1,\dots,m, j = 0,\dots,m-1\}$. This allows us to define our baseline model: This allows us to define our baseline model:

Model 1 *The CRM based on* $X_{i,j}$ *and* $N_{i,j}$ *data from [27] can be written on the form*

$$X_{i,j} \mid \mathcal{N} \sim \text{ODP}\left(\sum_{k=0}^{j\wedge d} \psi_{i,j-k,k} N_{i,j-k},\ \varphi\right),$$

and

$$N_{i,j} \sim \text{ODP}\left(\nu_{i,j},\ \phi\right),$$

that is,

$$\mathbb{E}[X_{i,j} \mid \mathcal{N}] = \sum_{k=0}^{j\wedge d} \psi_{i,j-k,k} N_{i,j-k} = \text{Var}(X_{i,j} \mid \mathcal{N})/\varphi,$$

and

$$\mathbb{E}[N_{i,j}] = \nu_{i,j} = \text{Var}(N_{i,j})/\phi.$$

All $\boldsymbol{X_{i,j}}$ are assumed to be conditionally independent, given $\boldsymbol{\mathcal{N}}$, and all $\boldsymbol{N_{i,j}}$ are assumed to be independent.

Remark 2

(a) Note that $\mathbb{E}[X_{i,j} \mid \mathcal{N}]$ is given by a sum, which makes it impossible to estimate the $\psi_{i,j,k}$s using anything but the identity link-function, whereas the parameters of the $N_{i,j}$ model can be estimated using any link-function.

(b) Going from the micro-level assumptions to this formulation of the CRM relies on a number of assumptions, which can be summarised as that all individual claims occur independently, all claims with the same accident year and reporting year have the same payment dynamics, and all payments occur independently between claims. For more on these assumptions, see [27]. For the purposes of this paper, the important thing is to think about the model, what the parameters mean in practice and how they can be estimated and used to produce reserve estimates etc.

(c) Based on the individual claim dynamics from [27], the interpretation of $\psi_{i,j,k}$ is that it corresponds to the average amount paid for a claim that occurred during accident year i that was reported j years after the accident occurred, and paid k periods after reporting. Thus, $\psi_{i,j,k}$ may be represented as $\psi_{i,j,k} = \mu_{i,j,k}\lambda_{i,j,k}$, where $\mu_{i,j,k}$ denotes the expected size of a single payment for a claim that occurred in year i, was reported j years later, and paid k years after reporting, and $\lambda_{i,j,k}$ denotes the expected number of claim payments in analogy with $\mu_{i,j,k}$. Further, from the definition of Model 1 there is a single *over-dispersion parameter φ. From [27] it follows that φ equals*

$$\varphi := \frac{\sigma^2_{i,j,k} + \mu^2_{i,j,k}}{\mu_{i,j,k}},$$

implying that $\sigma^2_{i,j,k} := \mu_{i,j,k}(\varphi - \mu_{i,j,k}) \geq 0$ must hold, i.e. it is not possible to choose both $\mu_{i,j,k}$ and $\sigma^2_{i,j,k}$ freely – either, will in this sense, determine the other when assuming a constant φ.

(d) Note that by having a single φ which is independent of the indices i, j, and k, means that from a quasi-likelihood perspective, all the $\psi_{i,j,k}$s can (in theory) be estimated using a standard Poisson *likelihood, and in a second step φ may be estimated using e.g. Pearson or deviance residuals.*

(e) *The model for $\boldsymbol{X}_{i,j}$ conditional on $\mathcal{N}$ given by Model 1 corresponds to an ODP model with identity link function expressed in terms of $\psi_{i,j,k}$s, and may be estimated using standard statistical software such as* R*. However, the fully general parametrization in terms of $\psi_{i,j,k}$ does not allow for estimation given $\boldsymbol{X}_{i,j}$. Instead, one is either forced to consider more restricted models, such as*

$$\psi_{i,j,k} = \mu\lambda_{j,k},$$

or

$$\psi_{i,j,k} = \alpha_i\beta_j\gamma_k \tag{1}$$

or to consider data on a more granular level, which we will get back to further on. Moreover, given that the $\psi_{i,j,k}$s have been estimated using maximum quasi-likelihood, to separate these into their constituent components, e.g. $\psi_{i,j,k} = \mu\lambda_{j,k}$, there is need of additional information, such as e.g. an estimate of μ. Similarly, for the count model, the perhaps most common parametrisation is obtained by assuming $\nu_{i,j} = \alpha_i\beta_j$ with $\sum_j \beta_j = 1$ we may estimate the parameters using a log-link function, i.e. $\log(\mathbb{E}[N_{i,j}]) = \widetilde{\alpha}_i + \widetilde{\beta}_j$, which exactly corresponds to the cross-classified over-dispersed Poisson chain-ladder model of [24].

Concerning reserve estimators, let $R_i^{\mathcal{R}}$ denote the outstanding RBNS claim payments for accident year i and let $R_i^{\mathcal{I}}$ denote the corresponding outstanding IBNR claim payments, i.e.

$$R_i := R_i^{\mathcal{R}} + R_i^{\mathcal{I}}.$$

From Eq. (6) – (9) in [27] it follows that

$$h_i^{\mathcal{R}}(\boldsymbol{\psi}; \mathcal{N}_0) := \mathbb{E}\left[R_i^{\mathcal{R}} \mid \mathcal{N}_0\right] = \sum_{j=m-i+1}^{m-1+d} \sum_{k=j-m+i}^{j\wedge d} \psi_{i,j-k,k} N_{i,j-k}, \tag{2}$$

$$h_i^{\mathcal{I}}(\boldsymbol{\psi}, \boldsymbol{\nu}) := \mathbb{E}\left[R_i^{\mathcal{I}} \mid \mathcal{N}_0\right] = \sum_{j=m-i+1}^{m-1+d} \sum_{k=0\vee(j-m+1)}^{(j-m+i-1)\wedge d} \psi_{i,j-k,k}\nu_{i,j-k}, \tag{3}$$

which, by replacing all unknown model parameters with estimators, gives us the following *computable* RBNS- and IBNR-reserve estimators

$$\widehat{R}_i^{\mathcal{R}} := h_i^{\mathcal{R}}(\widehat{\boldsymbol{\psi}}; \mathcal{N}_0) \tag{4}$$

and

$$\widehat{R}_i^{\mathcal{I}} := h_i^{\mathcal{I}}(\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\nu}}). \tag{5}$$

Although Eq. (2) and (3) are slightly more general than the formulation in [27], the proofs are identical, and it is also possible to calculate RBNS and IBNR reserve process variances explicitly following the arguments in [27], see Appendix C.2 for details.

Remark 3

(a) Eq. (2) *and* (3) *are expressed in terms of* $\psi_{i,j,k}$*s whose index combinations have not yet been observed. Consequently, apart from that the general formulation of Model 1 is not possible to estimate (see Remark 2(c) and (e)), the model is neither possible to use for prediction unless further structural assumptions are made such as e.g.* $\psi_{i,j,k} = \mu\lambda_{j,k}$ *or* $\psi_{i,j,k} = \alpha_i\beta_j\gamma_k$.

(b) The aggregate level formulation of the CRM from Model 1 is not sufficiently detailed in order to be able to derive Eq. (2) *and* (3)*. In particular, unless the micro-level assumptions discussed in Remark 2(b) are fulfilled,* it is not possible *to claim that the reserves given by* (4) *and* (5) *actually correspond to proper RBNS and IBNR reserves.*

As commented on in Remark 2(c) and (e) it is not possible to allow freely varying individual claim payment distributions, and one aspect of this (relating to Remark 2(d) and the assumption of a single over-dispersion parameter) is that such models may be impossible to estimate. This, however, is partly a consequence of having too many parameters in relation to modelling based on aggregated $X_{i,j}$ level data, and partly due to that a too flexible parametrization does not allow us to observe all data needed w.r.t. specific (i, j, k)-combinations. Thus, by using slightly less aggregated data, introducing $X_{i,j,k}$, which denotes the total amount of payments from accident year i, that come from claims that are reported with j years delay, and that are paid k years after reporting, it is natural to define the following model:

Model 2 *The CRM based on* $X_{i,j,k}$ *and* $N_{i,j}$ *data from [27] can be written on the form*

$$X_{i,j,k} \mid \mathcal{N} \sim \text{ODP}\left(\psi_{i,j,k}N_{i,j},\ \varphi\right),$$

and

$$N_{i,j} \sim \text{ODP}\left(\nu_{i,j},\ \phi\right),$$

that is,

$$\mathbb{E}[X_{i,j,k} \mid \mathcal{N}] = \psi_{i,j,k}N_{i,j} = \text{Var}(X_{i,j,k} \mid \mathcal{N})/\varphi,$$

and

$$\mathbb{E}[N_{i,j}] = \nu_{i,j} = \text{Var}(N_{i,j})/\phi.$$

All $X_{i,j,k}$ *are assumed to be conditionally independent, given* $\mathcal{N}$*, and all* $N_{i,j}$ *are assumed to be independent.*

But, a closer inspection of the derivations in [27] (as well as in [26, 22, 17]) $X_{i,j,k}$ level data is actually what is used in intermediate steps. This level of granularity has also been used in e.g. [4] and is still considerably less demanding to use than the continuous time micro-data needed in e.g. [23, 1].

Remark 4

(a) Model 2 uses less aggregated data than Model 1, see Remark 2(b), and these two models will not *produce the same parameter estimates. Concerning parametrisations, note that Model 2 can be expressed as*

$$\log(\mathbb{E}[X_{i,j,k} \mid N_{i,j}]) = \log(N_{i,j}) + \widetilde{\psi}_{i,j,k},$$

or

$$\mathbb{E}[X_{i,j,k} \mid N_{i,j}] = N_{i,j}\, \mathrm{e}^{\widetilde{\psi}_{i,j,k}},$$

making it natural to estimate the $\psi_{i,j,k}$s an ODP model with log-link function and the $\log(N_{i,j})$s as offsets, conditionally on the $N_{i,j}$s.

(b) Model 2 will produce the same theoretical RBNS and IBNR reserves as Model 1. The same holds true for the process variances, which are identical with those for Model 1. For more on this, see Appendix C.2.

However, note that by working with $X_{i,j,k}$ data it is no longer necessary to include $N_{i,j}$ count data information in order to be able to produce separate RBNS and IBNR reserves, also recall Remark 3(b), since

$$R_i^{\mathcal{R}} := \sum_{j=0}^{m-i} \sum_{k>m-(i+j)} X_{i,j,k},$$

and

$$R_i^{\mathcal{I}} := \sum_{j=m-i+1}^{m-1} \sum_{k} X_{i,j,k}.$$

This last observation suggests to introduce the following model:

Model 3 *Let $X_{i,j,k}$ be defined as an over-dispersed model according to*

$$X_{i,j,k} \sim \mathrm{ODP}\left(\psi_{i,j,k},\ \varphi\right),$$

with

$$\mathbb{E}[X_{i,j,k}] = \psi_{i,j,k} = \mathrm{Var}(X_{i,j,k})/\varphi,$$

where all $X_{i,j,k}$ are assumed to be independent.

Note that Model 3 does not only have very simple RBNS and IBNR reserve predictors, being sums of suitably indexed $\psi_{i,j,k}$s, but also the corresponding process variances are easily obtained due to that all $X_{i,j,k}$s are assumed to be independent. This independence assumption may, of course, be questioned, since in reality the underlying claims dynamics should correspond to something similar to the micro-level assumptions that underpin the CRM model (Model 1). Nevertheless, Model 3 defines a flexible statistical model in its own right, and whose performance will be evaluated in Section 4. Still, as for Model 1 and Model 2, Model 3 will need additional parameter restrictions in order to be able to be used in practice.

Before proceeding further, note that Model 1 - Model 3

(i) have enough flexibility to be able to capture e.g. "inflation" or "calendar year" effects using e.g. $\psi_{i,j,k} := \mu_{i+j}\lambda_{j,k}$ (although these parameters may be hard to estimate in practice due to the number of parameters, and hard to extrapolate for prediction purposes),

(ii) are straightforward to bootstrap, see Appendix C.2 for details.

Concerning the problems with estimating $\psi_{i,j,k}$ s and extrapolating these for reserve predictions, these problems can at least partly be addressed by replacing the $\psi_{i,j,k}$ s with general functional forms that can be estimated using regression based machine learning techniques.

3. MACHINE LEARNING BASED RESERVING MODELS

In Section 2 three different over-dispersed Poisson models have been introduced and discussed, Model 1 - Model 3. All of the models are flexible w.r.t. parametrisation and estimation, and may be used to produce separate RBNS and IBNR reserves using simple formulas. Still, the models' most general formulations will have too many parameters for reliable estimation, but the general $\psi_{i,j,k}$ parametrisation will not lend itself to extrapolation, see Remark 2(c) and (e), and Remark 3(a). Moreover, even though it would be possible to estimate all $\psi_{i,j,k}$ s, these estimators will likely be unstable. We may instead use functional forms, but then the question is, *which* functional forms will serve as appropriate approximations.

In the current section these questions will be addressed by using tree-based gradient boosting machines (GBM) and neural network (NN) regression functions combined with the ODP models from Section 2.

We will start by discussing these questions in a reserving context without explic-

itly stating any particular machine learning technique. The specific GBM and NN estimated models will be discussed in Section 3.3 and 3.4 below.

3.1 General considerations

The general focus will be on reserving models based on $X_{i,j,k}$-level data with regression functions $\psi(i, j, k; \boldsymbol{\theta})$ and $\nu(i, j; \boldsymbol{\theta})$ corresponding to general functional forms defined in terms of a parameter vector $\boldsymbol{\theta}$, which we will try to approximate using complex algorithmic techniques. Further, all models that we will consider will be ODP models parametrised using log-link functions, unless stated otherwise. In practice, this simply means that the $\psi_{i,j,k}$s and $\nu_{i,j}$s from Section 2 are defined according to $\psi_{i,j,k} := \psi(i, j, k; \boldsymbol{\theta})$ and $\nu_{i,j} := \nu(i, j; \boldsymbol{\theta})$, and that a specific link-function is being used. That is, Model 2 may be parametrized according to

$$\begin{cases} X_{i,j,k} \mid N_{i,j} & \sim \text{ODP}(\log(N_{i,j}) + \psi(i, j, k; \boldsymbol{\theta}),\ \varphi), \\ N_{i,j} & \sim \text{ODP}(\nu(i, j; \boldsymbol{\theta}),\ \phi), \end{cases}$$

where

$$\mathbb{E}[X_{i,j,k} \mid N_{i,j}] = N_{i,j} \exp\{\psi(i, j, k; \boldsymbol{\theta})\},$$

(recall Remark 4(a)) and

$$\mathbb{E}[N_{i,j}] = \exp\{\nu(i, j; \boldsymbol{\theta})\}.$$

Similarly, Model 3 can be parametrized as

$$X_{i,j,k} \sim \text{ODP}(\psi(i, j, k; \boldsymbol{\theta}),\ \varphi),$$

with

$$\mathbb{E}[X_{i,j,k}] = \exp\{\psi(i, j, k; \boldsymbol{\theta})\}.$$

These parametrizations of Model 2 and Model 3 are very close to the regression function parametrizations that will be introduced and used when fitting the GBMs and NNs discussed in Section 3.2. Note that using this specific parametrization will not change the interpretability of e.g. ψ, see Remark 2(c). Furthermore, this fact will also hold when these regression functions are estimated by using the functional forms described by the machine learning methods. Consequently, due to the possibility to interpret the underlying model structure, it will be possible to include expert opinion into this type of machine learning based regression models, e.g. by scaling estimated ψs based on expert opinions.

At this point we have not yet discussed any particular machine learning techniques in detail, but the overarching idea is that these techniques will allow us to

model complex functional relationships such as higher order non-linear relationships between covariates without explicitly stating *how* these should look. In general these methods involve a lot of parameters defining the functional forms that are being used to approximate the true regression functions (ψ and ν). Due to this, it is of great importance to discuss how to avoid overfitting, which is done in the next section.

3.2 Estimation and calibration of machine learning models

Before going into details on specific machine learning methods and reserving, let us take a step back to the general regression setting that was starting point of Section 2. That is, let $Y \mid \boldsymbol{c} \sim \text{ODP}(\mu(\boldsymbol{c};\boldsymbol{\beta}), \rho)$ where $\mu(\boldsymbol{c};\boldsymbol{\beta})$ is some unknown functional form that we want to approximate, which in particular means for a specfic link-function $u(\cdot)$ it holds that

$$\mu(\boldsymbol{c};\boldsymbol{\beta}) = u(\mathbb{E}[Y \mid \boldsymbol{c};\boldsymbol{\beta}]).$$

Moreover, recall from Remark 1 that when estimating the mean function in an ODP model it is possible to do this separately from the over-dispersion parameter ρ by using a standard Poisson-likelihood. Consequently, what we will do next is to use machine learning techniques to

- approximate the unknown $\mu(\boldsymbol{c};\boldsymbol{\beta})$ function by using a flexible function class $f(\boldsymbol{c};\boldsymbol{\gamma})$,
- do the optimization of $f(\boldsymbol{c};\boldsymbol{\gamma})$ using a loss function corresponding to a Poisson likelihood.

The machine learning techniques to be discussed below will introduce flexible functional forms $f(\boldsymbol{c};\boldsymbol{\gamma})$, where the parameter vector $\boldsymbol{\gamma}$ will be (potentially very) high-dimensional and it is, therefore, important to avoid overfitting. The general approach to tackle overfitting of complex algorithmic methods is to use out-of-sample validation, see e.g. [7, Ch. 12]. This is done by splitting the available data into two sub-sets: one used for training ("parameter estimation") and one used for validating the predictive performance of the trained model. Further, the numerical procedures that will be introduced below to estimate $\boldsymbol{\gamma}$ will be based on iterative numerical updating procedures. A problem with these procedures when it comes to complex algorithmic methods is that, unless you stop the iterative numerical procedure used for approximating $\mu(\boldsymbol{c};\boldsymbol{\beta})$ (or in our reserving models the functions $\psi(\cdot)$ and $\nu(\cdot)$) early, you will likely end up with a fully saturated model - a model with a unique parameter for each observed data point. In the extreme situation $\boldsymbol{\gamma}$ may have more parameters than observations (which will be the case in our reserving situation discussed in

Section 4). This premature stopping is what is often referred to as "early stopping", and can be thought of as an informal regularization technique, see e.g. [7, Ch. 18.2]. Consequently, what often is done in practice, is that the number of optimization steps used are determined based on the trained model's out-of-sample performance based on the validation data. Unless done properly, from a statistical perspective this may result in a kind of cheating – you are *not* allowed to *estimate* a parameter, in this context the number of iterations, based on validation data. Still, given a sufficient amount of data you can always split your data set into three parts: one part used for training, one part used for determining the number of steps that the optimization procedure should be run, corresponding to a pseudo-validation data set, and a final third sub-set used for proper out-of-sample validation once the model has been properly trained (without any prior peeking!). This procedure allows us to evaluate different parametrizations and configurations of our machine learning models and serves as a generic way of doing model selection. Further, since the models of interest belong to the family of ODP models, it is natural to evaluate the performance w.r.t. deviance residuals. This corresponds to that the parameter vector $\boldsymbol{\gamma}$ is optimized w.r.t. the standard ODP likelihood loss function. That is, this choice corresponds to standard maximum likelihood optimization, but w.r.t. complex parametric function $f(\boldsymbol{c}; \boldsymbol{\gamma})$.

When trying to implement these ideas in a reserving context the first problem is that we ideally would like to have fully developed payment data to be used to evaluate the prediction of outstanding claim payments. This is rarely the case. In order to at least partly circumvent this issue we adopt the procedure from [11]:

(i) Split the *un-aggregated* individual claim data into two parts,

(ii) Depending on the model that you want to train, aggregate the separate individual data sets into $\boldsymbol{X}_{i,j}$, $\boldsymbol{X}_{i,j,k}$, and $\boldsymbol{N}_{i,j}$ form. Let $\boldsymbol{X}_{i,j}^{(t)}$ denote "training" data and let $\boldsymbol{X}_{i,j}^{(v)}$ denote "validation" data, and analogously for $\boldsymbol{X}_{i,j,k}$ and $\boldsymbol{N}_{i,j}$.

(iii) Train your models on $\boldsymbol{X}_{i,j}^{(t)}$, $\boldsymbol{X}_{i,j,k}^{(t)}$, and $\boldsymbol{N}_{i,j}^{(t)}$ data, validate the model performance based on using $\boldsymbol{X}_{i,j}^{(v)}$, $\boldsymbol{X}_{i,j,k}^{(v)}$, and $\boldsymbol{N}_{i,j}^{(v)}$ data.

Remark 5

(a) The above algorithm is only described in terms of splitting the historic data in two, but you may, of course, make another implicit split of the data set used to construct e.g. $\boldsymbol{X}_{i,j}^{(t)}$ *to be able to carry out the above described three step procedure.*

(b) *By using the above algorithm we will train our models based on observed in-sample data, and we will use another set of in-sample data not used for training to construct "out-of-sample" validation data. This is most likely sub-optimal, but the best we can hope for.*

(c) *It is important that the mechanism used for splitting the original data sets into sub-sets does not create severe imbalances in terms of overall exposures per accident year. The procedure used in [11] corresponds to a random 50 / 50 split of the original individual claim data.*

We will return to Remark 5(b) when discussing GBM and NN models in more detail.

Next, we will describe two specific choices of functions $f(\boldsymbol{c};\boldsymbol{\gamma})$ using GBMs and NNs.

3.3 Tree-based gradient boosting machines

The current section will give a brief introduction to regression trees and tree based gradient boosting. A comprehensive introduction to these subjects can be found in e.g. [14, 7], which form the basis for the current exposition. More technical parts of the exposition can be found in Appendix C.1.

To start off, we consider the situation with data is in the form of pairs $(y_i, \boldsymbol{c}_i)$, $i = 1, \ldots, m$, where y_i corresponds to the observed response i with corresponding covariate vector $\boldsymbol{c}_i := (c_{i,1}, \ldots, c_{i,p})' \in \boldsymbol{C}$. As a notational convention all references to $\boldsymbol{c} = (c_1, \ldots, c_p)'$ are with respect to a general covariate vector. In the current paper we will focus on binary regression trees. A binary regression tree is, like any other regression model, a model that takes a covariate vector $\boldsymbol{c}$ and parameter vector $\boldsymbol{\gamma}$ as input and produces an expected value of $Y \mid \boldsymbol{c}$ as output. More specifically, we will focus on ODP models, which means that given a link-function $u(\cdot)$, we let

$$r(\boldsymbol{c};\boldsymbol{\gamma}) := u(\mathbb{E}[Y \mid \boldsymbol{c};\boldsymbol{\gamma}]),$$

and where the distribution of $Y \mid \boldsymbol{c}$ will imply a specific loss function to be used when estimating $\boldsymbol{\gamma}$ – in our case a Poisson likelihood. More specifically, in its simplest form, a binary regression tree of "depth" k will partition $\boldsymbol{C}$ into 2^k different regions $\mathcal{A}_j$, $j = 1, \ldots, 2^k$, where each region is assigned a single value δ_j, which means that $r(\boldsymbol{c};\boldsymbol{\gamma})$ may be represented according to

$$r(\boldsymbol{c};\boldsymbol{\gamma}) := \sum_{j=1}^{2^k} \delta_j \mathbb{1}_{\{\boldsymbol{c}\in\mathcal{A}_j\}}, \tag{6}$$

where $\boldsymbol{\gamma}$ is the vector containing all δ_js and all additional parameters needed to define the $\mathcal{A}_j$s. Consequently, given that the $\mathcal{A}_j$s are known, in our ODP-regression situation, we would get that the δ_js are obtained by optimizing

$$\delta_j := \arg\min_{\delta} \sum_{i:\boldsymbol{c}_i \in \mathcal{A}_j} L(y_i, r(\boldsymbol{c}_i; \boldsymbol{\gamma})),$$

where

$$r(\boldsymbol{c}_i; \boldsymbol{\gamma}) = \delta_j, \qquad i : \boldsymbol{c}_i \in \mathcal{A}_j,$$

and where $\boldsymbol{L}$ corresponds to the Poisson log-likelihood function[1]. Further, to see how to construct the $\mathcal{A}_j$s, we will use the "tree" interpretation of this regression procedure. That is, the reason why this is referred to as a regression "tree" is because the procedure to arrive at the 2^k partitioning splits of $\boldsymbol{C}$ can be represented as a binary decision tree consisting of k binary decisions, where each decision is based on a single component of the covariate vector $\boldsymbol{c}$. That is, the $\mathcal{A}_j$ regions can be thought of as being constructed using a sequence of k binary decisions expressed in terms of $\mathcal{A}_{l,j}, l = 1, \ldots, 2^j, j = 1, \ldots, k$, and the $\mathcal{A}_j$ regions are often referred to as "leaves". An example of a binary decision tree of depth 2 is shown in Figure 1, where $\boldsymbol{c} = (c_1, c_2)$ and where

$$\mathcal{A}_{l,j} := \mathcal{A}(\pi_{l,j}, \kappa_{l,j}) = \{c_{\pi_{l,j}} \leq \kappa_{l,j}\}.$$

Further, Figure 1 implies that it will be computationally complex to estimate all δ_js and all $\mathcal{A}_{l,j}$s simultaneously. As a consequence of this, a so-called "greedy" algorithm is used. This is a sequential algorithm where the tree is allowed to grow from depth 1 to k, where the $\mathcal{A}_{l,j}$s, together with the associated δ_js, are optimized for the *next* depth level of the tree. That is, given that you have reached depth s of the tree, you *only* optimize the $\mathcal{A}_{s,j}$s and the associated δ_js as if these would be the terminal depth of the tree, leaving the previously estimated $\mathcal{A}_{l,j}, l < s$, fixed. This depth-by-depth level optimization is continued until you reach depth k, which gives you your final $\mathcal{A}_j$ and δ_js. Note that in practice $\mathcal{A}_{l,j} = \mathcal{A}(\pi_{l,j}, \kappa_{l,j}) = \{c_{\pi_{l,j}} \leq \kappa_{l,j}\}$, meaning that you optimize w.r.t. *which* covariate component to base your split on ($\pi_{l,j}$), and *which* covariate threshold to base your split on ($\kappa_{l,j}$). This procedure allows the same covariate to appear on different levels in the same tree, which is seen in Figure 1. Also note that each $\mathcal{A}_{l,j}$ is associated with *two* δ-values, depending on which binary decision that was made based on $\mathcal{A}_{l,j}$. For more on this, see Appendix C.1 and the previously mentioned references.

The above described procedure for how to construct a tree will always produce a tree with 2^k leaves. It is, however, common to use so-called "pruning", where the

[1]Note that the objective function used for optimizing δ_j corresponds to the sample average estimator of $\mathbb{E}[L(Y, r(\boldsymbol{c}; \boldsymbol{\gamma})) \mid \boldsymbol{c}]$.

δ-values associated with leaves estimated based on e.g. few observations are removed. Further, trees, as the one depicted in Figure 1, are easy to interpret in terms of the $\mathcal{A}_j$ regions, and trees allow for non-trivial interactions between the covariates. Moreover, trees scale nicely when having many covariates, i.e. $\boldsymbol{c} = (c_1, \ldots, c_p)$ with large p, but trees that are estimated with large depth may be unstable w.r.t. overfitting, introducing large estimation error variance. Due to this, techniques such as random forests and bagging are used, which are based on averaging trees of lower depth, where not all covariates are allowed to be used in each optimization step when building the trees.

Another alternative to building a single complex tree-based regression model is to use a sequential learning procedure known as gradient boosting. Unlike forests which average a large number of smaller (approximately uncorrelated) regression trees, tree-based gradient boosting instead uses trees of low depth, e.g. $k \leq 10$, to approximate the pointwise gradient of the loss function given the current fit. That is, let $\widehat{G}_0(\boldsymbol{c}) = 0$, and for each $b, b = 1, \ldots, \beta$, calculate

$$g_i = -\frac{\partial}{\partial z} L(y_i, z)\Big|_{z=\widehat{G}_{b-1}(\boldsymbol{c}_i)}, i = 1, \ldots, m,$$

and fit a tree with low depth to these gradients according to

$$\widehat{\gamma}_b = \arg\min_{\gamma} \sum_{i=1}^{m} (g_i - r(\boldsymbol{c}_i; \gamma))^2.$$

The fit is thereafter updated according to

$$\widehat{G}_b(\boldsymbol{c}) = \widehat{G}_{b-1}(\boldsymbol{c}) + \epsilon r(\boldsymbol{c}; \widehat{\gamma}_b),$$

where $\epsilon > 0$ corresponds to the so-called shrinkage factor or learning rate, which after β iterations results in

$$\widehat{G}_\beta(\boldsymbol{c}) := G_\beta(\boldsymbol{c}; \epsilon, \widehat{\gamma}_b, b = 1, \ldots, \beta) = \epsilon \sum_{b=1}^{\beta} r(\boldsymbol{c}; \widehat{\gamma}_b).$$

The above described procedure corresponds to Algorithm 17.4 in [7], and this means that our gradient boosting machine predictor $f^{\text{GBM}}(\boldsymbol{c}; \widehat{\gamma})$ is given by

$$f^{\text{GBM}}(\boldsymbol{c}; \widehat{\gamma}) := \epsilon \sum_{b=1}^{\beta} r(\boldsymbol{c}; \widehat{\gamma}_b).$$

In Section 4 the `R`-package `gbm` will be used, and we refer to the documentation of this software for the precise implementation of the exact GBM configuration that is

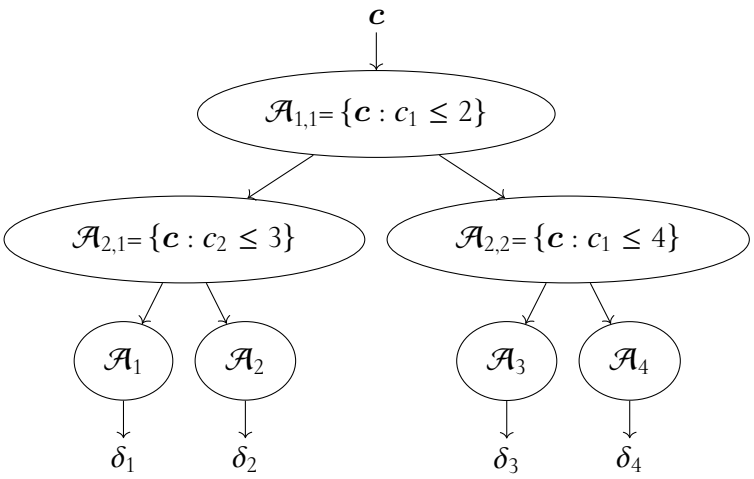


Figure 1: Example of tree of depth 2, where $\mathcal{A}_1 = \{\boldsymbol{c} : \boldsymbol{c} \in \mathcal{A}_{1,1} \cap \mathcal{A}_{2,1}\} = \{\boldsymbol{c} : \boldsymbol{c} \in \{c_1 \leq 2, c_2 \leq 3\}\}$, and analogously for $\mathcal{A}_2$, $\mathcal{A}_3$, and $\mathcal{A}_4$.

being used. Further, as discussed in Section 3.2, how to choose the learning rate ϵ, the depth of the trees used for gradient approximations, and the number of iterations β, can be done by splitting the data into training and validation sets. We will return to this in more detail in Section 4.

3.3.1 Reserving models

From the general definition of Model 2, using log-link functions according to the discussion on parametrizations in Section 3.1 and Remark 4(a), it follows that its GBM estimated version will produce the following estimates

$$\begin{cases} \widehat{\mathbb{E}}[X_{i,j,k}|N_{i,j}] &= N_{i,j}\exp\{f_{\psi}^{\text{GBM}}(i,j,k;\widehat{\boldsymbol{\theta}})\} \\ &= N_{i,j}\exp\{\epsilon\sum_{b=1}^{\beta} r_{\psi}(i,j,k;\widehat{\gamma}_b)\} \\ \widehat{\mathbb{E}}[N_{i,j}] &= \exp\{f_{\nu}^{\text{GBM}}(i,j;\widehat{\boldsymbol{\theta}})\} \\ &= \exp\{\epsilon\sum_{b=1}^{\beta} r_{\nu}(i,j;\widehat{\gamma}_b)\} \end{cases} \qquad \text{(M2 - GBM)}$$

and where $\widehat{\varphi}$ and $\widehat{\phi}$ corresponds to estimates based on e.g. Pearson or deviance residuals. Note that this is all the information that is needed in order to produce separate RBNS and IBNR reserves based on Eq. (4) and (5). Further, when it comes to bootstrapping the GBM-estimated version of Model 2, we will use Model 2, but using the "plug-in" estimates from (M2 - GBM). For more details on this, see Appendix C.2.

Analogously, the GBM estimated version of Model 3 will produce

$$\begin{aligned}\widehat{\mathbb{E}}[X_{i,j,k}] &= \exp\{f_{\psi}^{\text{GBM}}(i,j,k;\widehat{\boldsymbol{\theta}})\} \\ &= \exp\{\epsilon \sum_{b=1}^{B} r_{\psi}(i,j,k;\widehat{\gamma}_b)\}.\end{aligned} \qquad \text{(M3 - GBM)}$$

3.4 Neural networks

As in Section 3.3 we start with the general setting with a response $\boldsymbol{Y}$ that we want to regress on $\boldsymbol{c} \in \boldsymbol{C}$, and the neural network (NN) model will define a general regression function parametrised using a specific link-function $u(\cdot)$ according to

$$f^{\text{NN}}(\boldsymbol{c};\boldsymbol{w}) = u(\mathbb{E}[Y \mid \boldsymbol{c};\boldsymbol{w}]),$$

defining the mean of, in our reserving context, an ODP model. The reason for using "$\boldsymbol{w}$" instead of the previous "$\boldsymbol{\gamma}$", is because $\boldsymbol{w}$ will correspond to weights. We will here focus on so-called *feed-forward* neural networks and base our exposition on [14, 13, 7]. A feed-forward neural network is most easily understood from an example as the one given in Figure 2. Based on this figure it is natural to give $f^{\text{NN}}(\boldsymbol{c};\boldsymbol{w})$ a recursive representation, which leads us to the following definition of $f^{\text{NN}}(\boldsymbol{c};\boldsymbol{w})$ for a general feed-forward neural network with λ layers:

$$\begin{aligned} f^{\text{NN}}(\boldsymbol{c};\boldsymbol{w}) &= a^{(\lambda)}, \\ a_j^{(l)} &= g^{(l)}(w_{j,0}^{(l-1)} + \sum_{i=1}^{p_{l-1}} w_{j,i}^{(l-1)} a_i^{(l-1)}), \quad 2 \le l \le \lambda - 1,\ j = 1,\dots,p_l, \\ a_j^{(1)} &= c_j, \quad j = 1,\dots,p, \end{aligned} \qquad (7)$$

where the ws are weights to be estimated, p_l corresponds to the number of "neurons" (or nodes) in layer l, and the $g^{(l)}$-functions are the (parameter-free) so-called "activation functions", that often have a sigmoid shape, such as the hyperbolic tangent function (tanh). Further, the first layer is called the "input" layer, the last layer is called the "output" layer, and all intermediate layers are called "hidden" layers. A feed-forward neural network with more than 2 hidden layers is what usually is referred to as being a "deep" (feed-forward) neural network. Further, from the above it is clear that this type of model, consisting of a large number of parameters and iterative composition of functions, will be very flexible making them so-called "universal approximators" (Cybenko's universality theorem). Moreover, due to that $f^{\text{NN}}(\boldsymbol{c};\boldsymbol{w})$ is expressed in terms of compositions of $g^{(l)}$-functions, it is possible to obtain explicit gradients using so-called "backpropagation", see e.g. [7, Alg. 18.1]. By using these gradients the parameter estimates can be updated iteratively. Usually not all input data

is used in each gradient update step, but rather the gradient update is based on a sample of input data ("batch") chosen at random. This is primarily used if input data is very large. Due to this the concept of "epochs" has been introduced, where an epoch corresponds to that all input data has been used for a parameter estimate update. That is, if all data is used in each step of the gradient updating procedure, an epoch is the same as a standard iteration. If batches are used, several gradient updates will be done within each epoch.

The flexibility of NN models is partly due to the large number of parameters, often thousands, which will make the models heavily overparametrized, with the risk of overfitting if you run too many epochs (iterations). The problem with overparametrization can be addressed by using regularization techniques, by adding e.g. Ridge or Lasso penalization to the likelihood, see e.g. [14, 7]. A technique related to regularization that will be used in the current paper is to use what is known as "dropout", which corresponds to that a fraction of all weights for each layer in the network are randomly set to zero in each epoch.

Further, recall that the starting point of the current paper is GLM models. One way to improve the performance of a feed-forward NN in this context is to use the approach taken in [11, 30] where one first fit a GLM model, and use these parameters as offsets (or non-trainable parameters) in an NN model. That is, the NN model will essentially be fitted to the residuals from the GLM model, which may be thought of as so-called regression or residual boosting the GLM model using a NN model, see e.g. [7, Ch. 16.7, Ch. 17.2]. Given that the GLM model provides a reasonable fit, the NN model will likely need less training. Still, this procedure will *not* provide initial weights $\boldsymbol{w}$.

Concerning other aspects of tuning of NN models, you can choose the number of layers, the number of neurons (or nodes) in a layer, activation functions (that often are different for different layers), and the number of epochs used for training. All of these decisions can be evaluated using the techniques discussed in Section 3.2 by splitting the data into training and validation sets. We will comment more on this in Section 4.

3.4.1 Reserving models

In analogy with the GBM estimated models from Section 3.3.1, using log-link functions, it follows that the NN estimated version of model 2 will produce

$$\begin{cases} \widehat{\mathbb{E}}[X_{i,j,k}|N_{i,j}] &= N_{i,j} \exp\{f_{\psi}^{\mathrm{NN}}(i,j,k;\widehat{\boldsymbol{w}})\} \\ \widehat{\mathbb{E}}[N_{i,j}] &= \exp\{f_{\nu}^{\mathrm{NN}}(i,j;\widehat{\boldsymbol{w}})\} \end{cases} \qquad \text{(M2 - NN)}$$

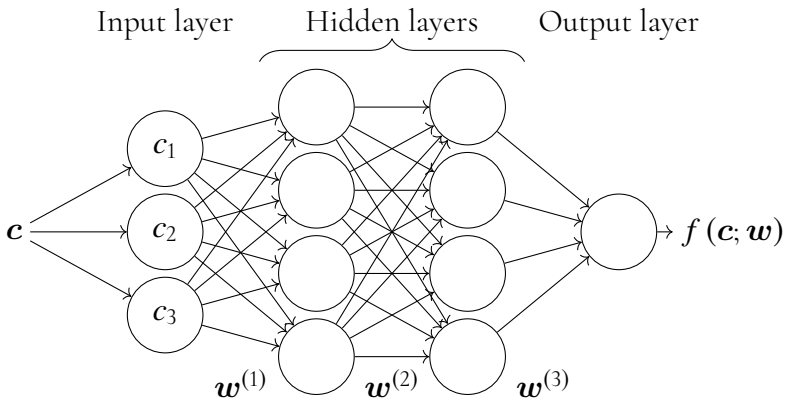


Figure 2: Example of feed-forward neural network taking $\boldsymbol{c} = (c_1, c_2, c_3)'$ as input covariates with $\lambda = 4$ layers, where 2 are hidden layers with $p_2 = p_3 = 4$ neurons, where $\boldsymbol{w}^{(l)}$, $l = 1, 2, 3$, corresponds to weights.

with $f_\bullet^{\text{NN}}(\cdot\,; \widehat{\boldsymbol{w}})$ from (7), and where $\widehat{\varphi}$ and $\widehat{\phi}$ correspond to estimates based on e.g. Pearson or deviance residuals. Again, as discussed in Section 3.3.1 this is sufficient information in order to be able to produce separate RBNS and IBNR reserves, as well as to produce bootstrapped reserves.

As before, the NN estimated version of Model 3 will produce

$$\widehat{\mathbb{E}}[X_{i,j,k}] = \exp\{f_\psi^{\text{NN}}(i, j, k; \widehat{\boldsymbol{w}})\}. \qquad \text{(M3 - NN)}$$

4. NUMERICAL ILLUSTRATION

In this section, we give an illustration of how we, in practice, can use the machine learning techniques described in this paper. For this, we consider six different portfolios of simulated data generated by the individual claims history simulation machine of [12]. These six portfolios will be referred to as Lines of Business (LoBs) 1-6 and are identical to those in [11]. For a detailed description of these LoBs and their characteristics, see Section 2.3 and the appendix in [11]. For `R`-code showing how to generate the individual-level data, see Listing 1 in the same paper. We will not describe the data and simulation machine in detail here. In short, the simulation machine uses neural networks calibrated on Swiss insurance data to simulate synthetic claims histories. The 6 LoBs are generated independently of each other, with the differences between them being that the first three use one stochastic generator and the last three another one. The difference between the LoBs within the same group, i.e. LoBs 1-3 and 4-6, are that their underlying features, which are the size, growth, and covariate structure of the portfolios, are different. The synthetic data

that we will look at here can be simulated following Listing 1 of [11] using seed 75, implemented using the `R`-package `keras`.

As mentioned in Section 3.2, to alleviate overfitting, we perform early stopping in training our GBMs and NNs, which we illustrate further on in this section. To do this, we need to split the data into a training and a validation set, and we do this following [11] by, for each LoB and AY, dividing the individual claims into two datasets of equal size according to Listing 2 in [11]. As is described in [11], the two datasets should be of equal size since parameters may be volume-dependent, see also the discussion in Section 3.2 above. In Listing 1 of the present paper, we have included code that aggregates the individual-level data from the simulation machine into data corresponding to the $\boldsymbol{X_{i,j,k}}$s and $\boldsymbol{N_{i,j}}$s.

Table 1 shows the number of individual claims and the total payments generated for each LoB, together with the proportion of these that are RBNS. From this table, we can note that extended reporting and payment delays are not a prominent feature of the data generated by this simulation machine since most claims are RBNS at the current time. Still, the simulated data contain claims that have longer reporting and payment delays, see e.g. Figure 3.

Table 1: Number of generated individual claims per LoB.

	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
Number of individual claims	250,040	250,197	99,969	249,683	249,298	99,701
Percent RBNS	99.4%	99.4%	99.2%	99.2%	99.2%	99.1%
Percent RBNS in last AY	93.6%	93.7%	93.3%	91.5%	91.5%	91.6%
Total payments	285,989	278,621	108,345	429,344	437,728	171,482
Total outstanding payments	39,689	37,037	16,878	71,630	72,548	31,117

Given that we simulate data, we know all payments, both current and future. Therefore we will be able to compare the reserves of the different models and methods to what the real outstanding amounts are. One metric we will use to assess the performances is the relative bias (error). The relative bias of a prediction $\widehat{\boldsymbol{R}}$ of the true reserve $\boldsymbol{R}$ is defined as $\frac{R-\widehat{R}}{R}$. In Table 2, we show the reserves of four benchmark models, the chain-ladder (CL) technique, the standard CRM using $\boldsymbol{X_{i,j,k}}$-data (Model 2) with a log-link and the parametrization $\log \psi_{i,j,k} = \alpha_k$, the CRM using the parametrization $\log \psi_{i,j,k} = \alpha_i + \beta_j + \gamma_k$ — from now on referred to as the generalized CRM (GCRM) — and the neural network of [11] (GRWNN). The standard CRM severely underestimates the true reserve, which is not necessarily surprising since the CRM assumes that payments are only dependent on the payment delay. Therefore, this underestimation could be due to there being some accident year or

reporting delay effect in the data that the standard CRM cannot capture. We see that the simple extension to the GCRM already alleviates this problem to a large extent and performs similarly to the CL. Although the neural network (GRWNN) yields the smallest biases across the board; however, it does not allow for the computation of individual RBNS and IBNR reserves. It is also not as interpretable as the conditional ODP models discussed in this paper, which can be connected to detailed individual-level assumptions. In Table 7 in the appendix, a version of Table 2 for IBNR and RBNS reserves are shown, and it is clear that the GCRM performs better than the standard one. It seems that we, in general, underestimate RBNS reserves while overestimating IBNR reserves.

Table 2: True reserve compared to CL, the CRM and [11] (GRWNN) reserves. Relative bias (error) of the predictions in the parentheses.

	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
True reserves	39,689	37,037	16,878	71,630	72,548	31,117
CL reserves	38,569	35,460	15,692	67,574	70,166	29,409
	(-2.82)	(-4.26)	(-7.02)	(-5.66)	(-3.28)	(-5.49)
CRM reserves	32,485	29,901	13,040	55,782	59,390	24,403
	(-18.15)	(-19.27)	(-22.74)	(-22.12)	(-18.14)	(-21.58)
GCRM reserves	38,293	35,117	15,448	66,961	69,397	29,104
	(-3.52)	(-5.18)	(-8.47)	(-6.52)	(-4.34)	(-6.47)
GRWNN	39,233	35,899	15,815	70,219	70,936	30,671
	(-1.15)	(-3.07)	(-6.30)	(-1.97)	(-2.22)	(-1.43)

It is difficult to say why we see these underestimated RBNS and overestimated IBNR reserves, except for the CRM. For the CRM, it is clear that only taking payment delay into account leaves a lot of room for error. If there is some accident year effect, for instance, then we would have no way of capturing it. The GCRM, on the other hand, can catch this potential accident year effect. Nonetheless, it still underestimates the reserve for all LoBs. This underestimation could be due to many different reasons. For instance, as [11] notes, there may be some interaction that the cross-classified model structure cannot capture. That is, there may be some interaction effect between, for instance, payment and reporting delay that affects how large payments are. This suspicion is strengthened by Figure 3, which shows the average development of the payments from an individual claim as a function of the payment delay and the first three periods of reporting delay for each LoB. If there is no interaction between reporting and payment delay, the curves in this figure should overlap, which they do not. The cross-classified model structure does not allow us to capture this difference in patterns. In addition to interaction effects, there could also be some

calendar year or inflation effect, which these cross-classified model structures likely will not capture.

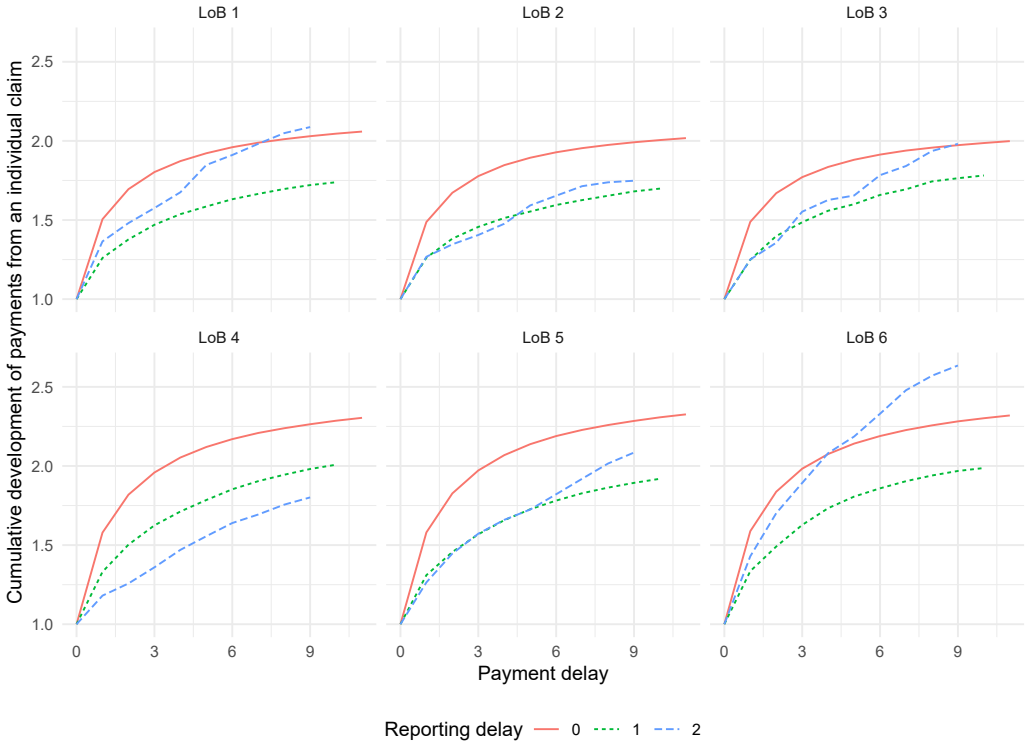


Figure 3: Average cumulative development of the payments from an individual claim as a function of payment delay, stratified on LoB and the first three periods of reporting delay, for all claims in all accident years.

In Figure 4, we show heatmaps of the relative biases of the predicted payments from the two versions of the CRM in each combination of accident year and development year for LoB 1. The CRM overestimates payments in the upper left triangle, while the opposite is true in the bottom right triangle. The GCRM has a much weaker, but similar pattern. A potential explanation is that the GCRM has picked up on some accident year effect, while not being able to capture some calendar year effect. To illustrate this further, Figure 5 shows how the average total size of the payments for individual claims evolves over the accident years. We see that there is a clear trend upwards for all of the LoBs. This trend upwards could be due to some type of inflation effect that the (generalized) CRM cannot capture.

We now move on to the machine learning methods, i.e. the GBM and NN. We

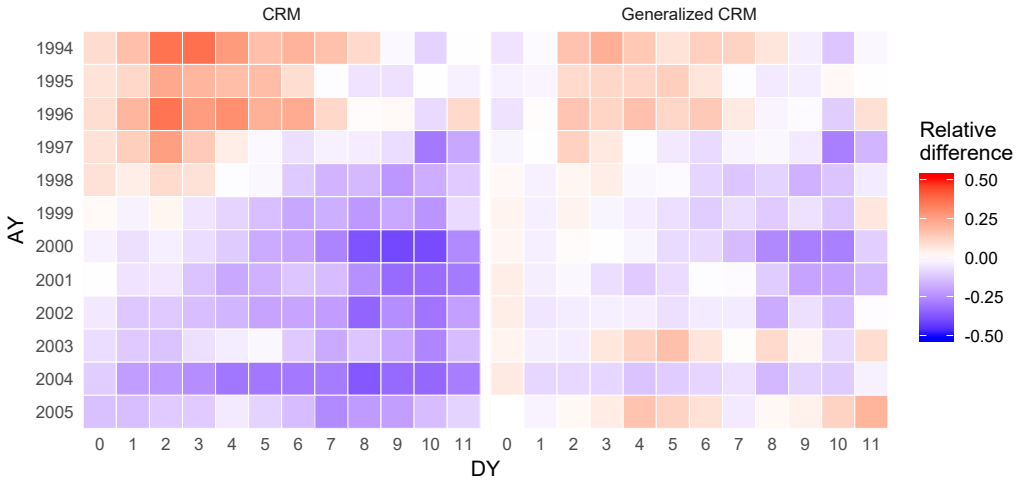


Figure 4: Heatmap for the relative biases (errors) of the prediction within a specific accident year and development year combination for LoB 1 using the CRM and GCRM.

will look at tunings of these algorithms that are, in some sense, standard. For the neural networks, we use the architecture of [11]. Their code, see Listing 4 in [11], can be used without modification for the model of the number of reported claims. However, for the payment part of the model, we have to adapt it to our situation with more granular data. This adaptation is straightforward to implement given the already available code from [11], which uses the `R`-package `keras`; see Listing 3 in the appendix.

For the number of epochs in training the neural networks, we set an upper limit at 10,000. If we consider LoB 1 and the number of reported claims part of Model M2 - NN, running 10,000 epochs takes around 47 seconds while it takes 174 for the payment part of the model[2] Therefore, in the worst-case scenario, we run 20,000 epochs taking around 220 seconds, which is still a feasible amount of computation time. However, if we want to compute an MSEP using bootstrapping, then we would need to repeat this fitting procedure several times. Assume we are satisfied with 100 simulations in the bootstrap, a rather small number for estimating the MSEP, which may be skewed and heavy-tailed, the bootstrap would take ca 6 hours to run. If we increase the number of bootstrap simulations to a thousand, the run-time would be about two and a half days for a single line of business. Now, given a faster computer,

[2]These and all other computations in this section are performed on a stationary PC with an Intel Core i7-7700 processor @ 3.60Ghz and 8GB RAM.

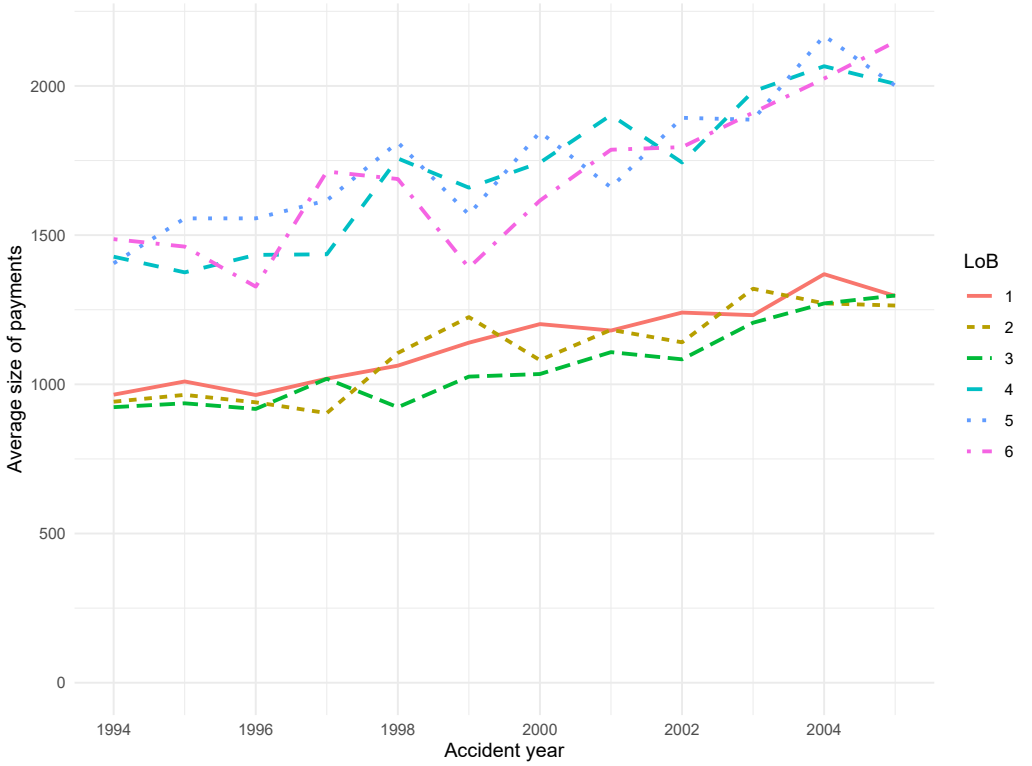


Figure 5: Average size of payments within the different combinations of reporting and payment delay for each accident year.

this number would naturally be smaller. For this illustration, we will stick to a maximum of 10,000 epochs, and simply state that the performance of the neural networks might be a lower bound on how well they could perform given more computation time.

For the GBMs, we start with the standard tuning in the `gbm`-function of the `gbm`-package in `R` and perform a rudimentary tuning of these by slightly changing one tuning parameter at a time and investigating how this affects the validation loss and actual out-of-sample performance (lower right triangle). This latter out-of-sample check cannot be done in practice. Nonetheless, since we can simulate data, we can investigate how tuning these parameters seem to perform over a number of simulated data sets generated by the same underlying dynamics.

The standard set of parameters in the `gbm`-function is a tree depth of 1 with a shrinkage factor of 0.1, a bagging fraction of 0.5, and a minimum of 10 observations per node. We will call this the *standard tuning* and now discuss how our tuning re-

lates to it and how we arrive at it. First of all, we set the maximum number of trees to 10,000 and then decide how many are needed based on the validation loss when fitting the models to the training data. For the GBMs, for both the payment part of the model and the number of reported claims part, our tunings seem to suggest that we get similar results with a smaller shrinkage of 0.01 as with 0.1, although then we need about ten times more trees. This can be seen in the upper left graph of Figure 6, where we illustrate the training and validation loss as a function of the number of trees in the GBM for the payment part of Model M2 - GBM applied to LoB 5 with our tuning using these two shrinkage factors. We see that the results are almost identical, with the larger shrinkage factor allowing us to reach the minimum validation loss faster. Whichever shrinkage factor we choose, the algorithm is reasonably quick. Fitting 10,000 trees takes ca half a second, which is feasible even within a bootstrap — increasing the number of trees to 100,000 increases the computation time tenfold, which should not be a problem in practice. Although, seeing as the results are almost identical, to keep computation times within reason, we settle for a shrinkage factor of 0.1. In practice, if the validation loss looks volatile, one usually decreases the shrinkage factor, making the loss more stable (smooth). It should be noted that the computation time here is orders of magnitude smaller than for the neural networks.

In the upper right graph of Figure 6, we illustrate the effect of varying the bagging fraction. In our case, bagging does not seem to help us out, only resulting in noisy deviance paths without, seemingly, decreasing the validation loss or yielding better out-of-sample performance. We, therefore, set the bagging fraction to 1, i.e. no bagging at all. Further, our data is quite small, and we seem to be helped by decreasing the minimum observation per node to 1. This can be a problem since it is possible that we base predictions on parameters estimated using only one observation, and therefore possibly introducing a large amount of variance in the predictions. However, the aggregated data that we use is small, especially taking into account that reporting and payment delays are so fast, and we thus have many zeroes – in particular for the claim counts. It can, therefore, be the case that we miss effects in the data by forcing ten observations per node, i.e. we introduce bias. How to decide on this trade-off between bias and variance is up to the practitioner, possibly by minimizing an MSEP. Based on the bottom left graph of Figure 6, there does not seem to be much of a difference anyway, so we decide to allow for one observation per node.

Finally, the depths of the trees are slightly more complicated. Looking at the validation data in the bottom right graph of Figure 6, we would be led to believe that depth 2 is better than 1. This seems to be the case for the number of reported claims. However, for the payment part of the model, this then leads to awful out-of-sample (lower triangle) performance, which would seem to indicate that the GBM overfits interactions in the observed data that are not representative of future payment pat-

terns. This is not obvious from the training and validation loss, although the distinct decrease in the training loss, which is not seen in the validation loss, may be seen as an indication of overfitting. Further, the reason for illustrating the tuning for LoB 5 is that not all of the other LoBs has the pattern where a tree depth of 2 yields a lower validation loss than a tree depth of 1. For some LoBs, the validation loss becomes worse when using a tree depth of 2.

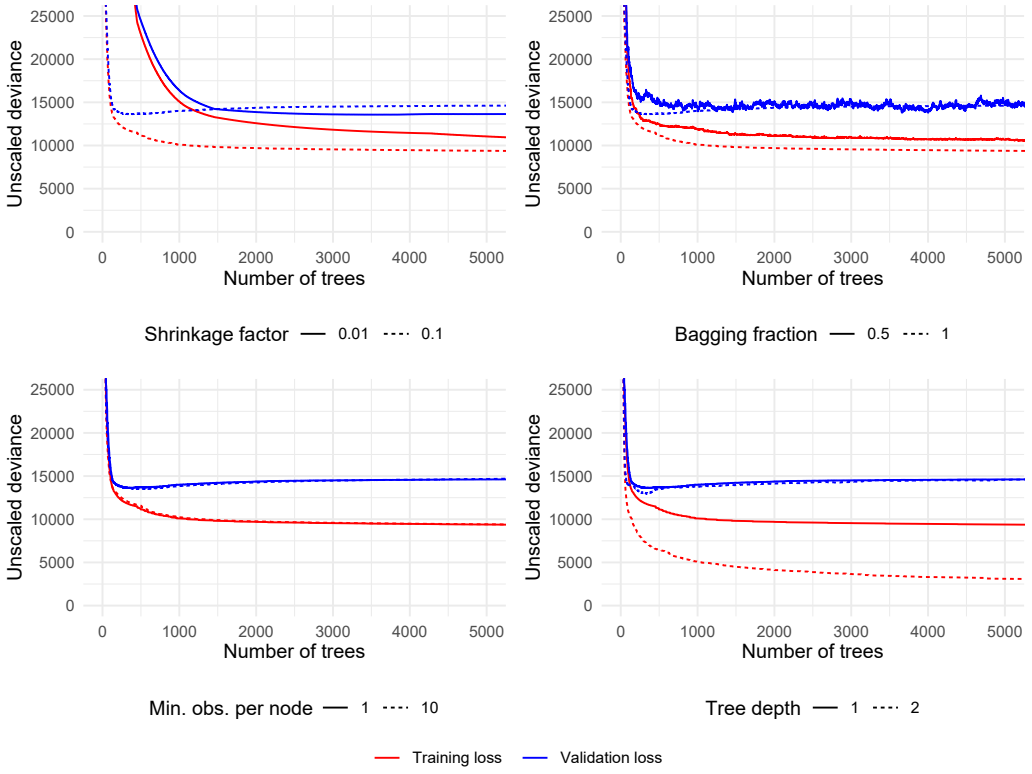


Figure 6: Training and validation loss for our tuning of the GBM for the payment part of Model 2 when varying the shrinkage factor (upper left),the bagging fraction (upper right), the minimum number of observations per node (lower left), and the tree depth (lower right).

Finally, it is straightforward to include transformations of the data when fitting the GBMs. As we have seen, there is a potential inflation (or calendar year) effect in the data, see Figure 4. Therefore, in addition to including the accident year, the reporting delay, and the payment delay as covariates, we also include the calendar time of reporting, i.e. the sum of the accident year and the reporting delay. Technically the GBM should be able to capture this transformed feature by itself; however, including it does seem to help performance quite a bit. Nonetheless, the GBM without this

inflation feature still performs well.

Another part of the calibration of machine learning methods is early stopping, discussed in Section 3.2, i.e. the choice of the number of trees in the GBMs and the number of epochs in training the NNs. [11] describe how to choose the number of epochs when training a neural network, which we follow closely. For this reason, we refrain from repeating what they have already described, and simply add that in addition to what they do, we choose the number of epochs by finding the minimum validation loss using a central simple moving average with window size 100. We use this moving average since the validation loss becomes volatile when using the dropout in the NNs. Without showing a graph, we note that the convergences are very slow for the incurred claims part of the NN models. The number of epochs at which the minimum (moving average) validation losses are attained can be seen in Table 3, in which we report (for Model 2) the number of epochs used for the NNs and the number of trees for the GBMs. Almost all LoBs need close to the maximum of 10,000 epochs and allowing for more epochs may yield better results than those we will acquire. For the payment part of the model, the minimum losses are reached much earlier for most LoBs.

Now let us look at the analogous choice of the number of trees for the GBMs. In Figure 7, we illustrate these choices for Model 2 by showing the training and validation loss (unscaled deviance) as a function of the number of trees for our and the standard tuning. The dotted black lines indicate the minimum of the validation loss as well as its position for our tuning. It is clear that our tuning and the standard tuning yield similar results, with our tuning generally being more stable and achieving slightly smaller losses. It is also clear that the validation losses for each LoB reach a point where they do not decrease any more and, most often, start to increase. This point is where we begin to overfit by adding more trees. Thus, this will be the number of trees that will be used in the GBM model when making the final prediction of the outstanding payments based on all data (training + validation). As was mentioned above, Table 3 shows these numbers of trees.

Table 3: Number of epochs used in training the NNs and the number of trees used in the GBMs for Model 2.

Number of epochs/trees	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
Payment part NN	8,825	405	225	9,895	331	227
Incurred claims part NN	9,950	9,704	9,357	9,946	9,751	1371
Payment part GBM	279	200	251	171	357	206
Incurred claims part GBM	635	383	912	310	360	467

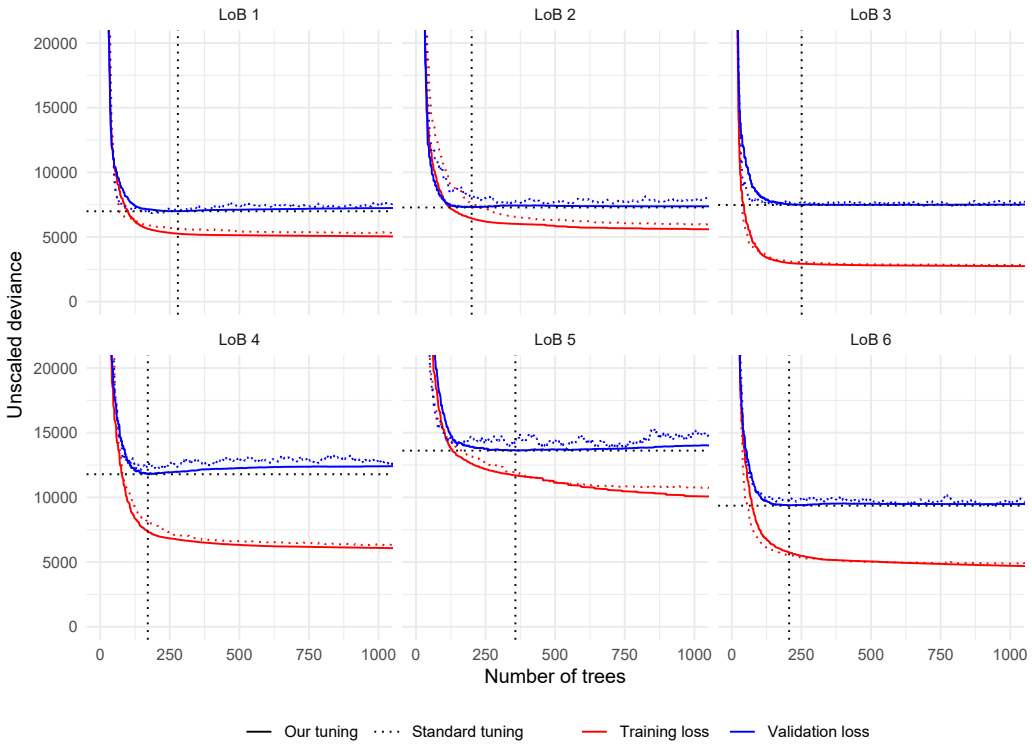


Figure 7: Out-of-sample validation analysis for over-fitting (determining the number of trees) in the GBM for Model 2. The dotted vertical and horizontal lines show the minimum of the validation loss and the number of trees at which it is obtained for our tuning.

While GBMs are still non-standard in the insurance industry, they are surprisingly accessible given, for instance, familiarity with GLMs and `R`. Listing 2 in the appendix shows how to use the GBMs in practice when fitting Model 2. This short snippet of code is a lot more accessible than the corresponding code for the neural networks using the `R`-package `keras`. Note that the `keras` code in Listing 3 is only for the aggregated payments.

Table 4 shows the reserves produced by these machine learning methods, as well as the benchmark models and the true reserves. Both machine learning methods seem to be performing well, especially the GBM for Model 2. In Figure 8, we show a heatmap of the relative biases of the predictions generated by the GBM for Model 2. There is not any visible pattern remaining in the data, which is in line with its excellent performance. The only cell with a notable error is the bottom-right cell; however, due to the quick reporting of claims, not many payments are made in that cell, and

the importance of predicting it with a small margin of error is not crucial.

Table 4: True reserve compared to benchmark models and the GBMs and NNs. Relative biases of the reserve predictions in the parentheses.

	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
True reserves	39,689	37,037	16,878	71,630	72,548	31,117
CL reserves	38,569	35,460	15,692	67,574	70,166	29,409
	(-2.82)	(-4.26)	(-7.02)	(-5.66)	(-3.28)	(-5.49)
CRM reserves	32,485	29,901	13,040	55,782	59,390	24,403
	(-18.15)	(-19.27)	(-22.74)	(-22.12)	(-18.14)	(-21.58)
GCRM reserves	38,293	35,117	15,448	66,961	69,397	29,104
	(-3.52)	(-5.18)	(-8.47)	(-6.52)	(-4.34)	(-6.47)
GRWNN	39,233	35,899	15,815	70,219	70,936	30,671
	(-1.15)	(-3.07)	(-6.30)	(-1.97)	(-2.22)	(-1.43)
GBM (Model 2)	39,697	37,253	16,508	72,679	71,828	31,941
	(0.02)	(0.58)	(-2.19)	(1.46)	(-0.99)	(2.65)
GBM (Model 2) without inflation	38,324	37,053	16,327	73,386	70,486	32,100
	(-3.44)	(0.04)	(-3.26)	(2.45)	(-2.84)	(3.16)
GBM (Model 3)	40,114	35,729	15,761	69,448	72,418	30,061
	(1.07)	(-3.53)	(-6.62)	(-3.05)	(-0.18)	(-3.39)
NN (Model 2)	41,587	37,587	15,680	71,155	71,309	28,984
	(4.78)	(1.48)	(-7.10)	(-0.66)	(-1.71)	(-6.86)
NN (Model 3)	39,757	38,719	16,245	70,916	74,600	28,943
	(0.17)	(4.54)	(-3.75)	(-1.00)	(2.83)	(-6.99)

The conclusions based on these reserves are, of course, only based on one dataset. Therefore, we simulate new datasets for each seed 1-100, and for each of these, repeat the reserve computations. We do this for the CL, CRM, GCRM, and the best performing GBM (Model M2 - GBM) and NN (Model M2 - NN). Figure 9 shows boxplots of the relative biases of these reserves for the simulated data using seeds 1-100. All of these data sets have been generated using the same underlying model, and are hence i.i.d., and we see that LoB 3 and 6 are those with most variation, probably due to the lower portfolio sizes generating fewer claims. Further, recall Figure 3, which tells us that, e.g., the realised simulated reporting delays may fluctuate quite a bit between simulations. In the appendix, Figures 10 and 11 show the corresponding graphs for the RBNS and IBNR reserve predictions. The GBM model seems to be the one with the least amount of bias; however, the NN has a predictor with smaller variance. Which of these should be deemed most important is to be decided by the practitioner using the model. However, we usually pick models based on the mean squared error of prediction, as this naturally balances the importance of bias and variance. Table 5

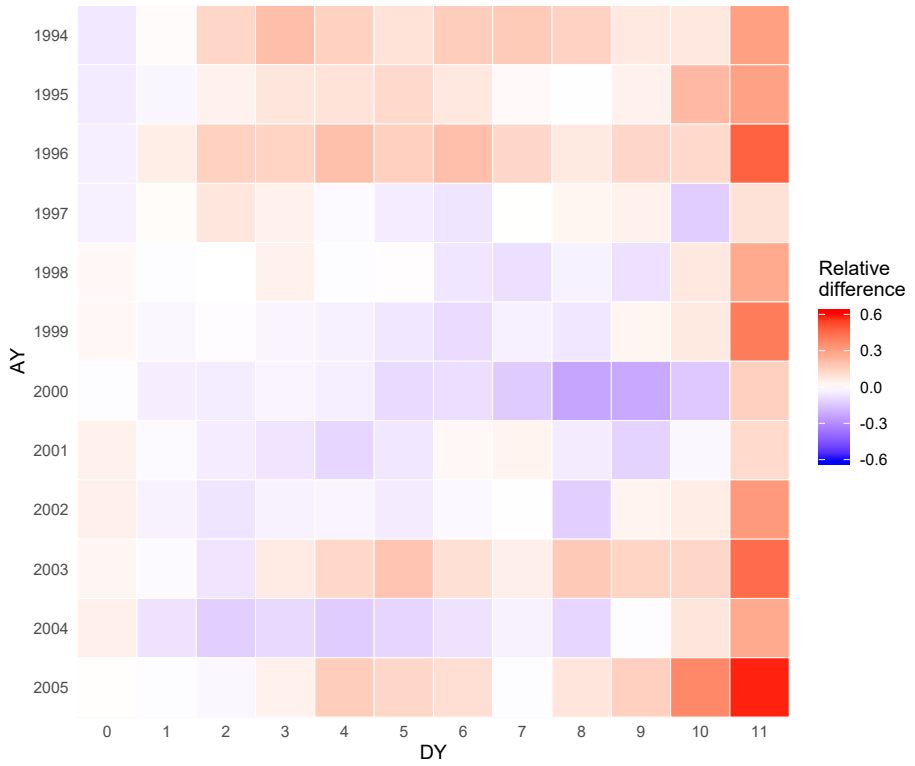


Figure 8: Heatmap for the relative biases (errors) of the prediction within a specific accident year and development year combination for LoB 1 using the GBM.

shows the average bias and the (true) root mean squared error of prediction across the simulations. The NN yields the smallest root mean squared error of prediction (RMSEP) except for LoB 3 and 6 where CL has slightly smaller RMSEP. Nonetheless, CL does not allow for separate RBNS and IBNR predictions, speaking in favor of the NN. It is, of course, also important to note that we have not allowed ourselves to run more than 10,000 epochs and that we have not performed any tuning of the neural networks, which could yield even better predictions.

4.1 Conditional mean squared error of prediction

We end this section by estimating the conditional MSEP for a subset of the models in this paper. First, for some predictor $\widehat{\boldsymbol{R}}$ of the outstanding payments $\boldsymbol{R}$, the

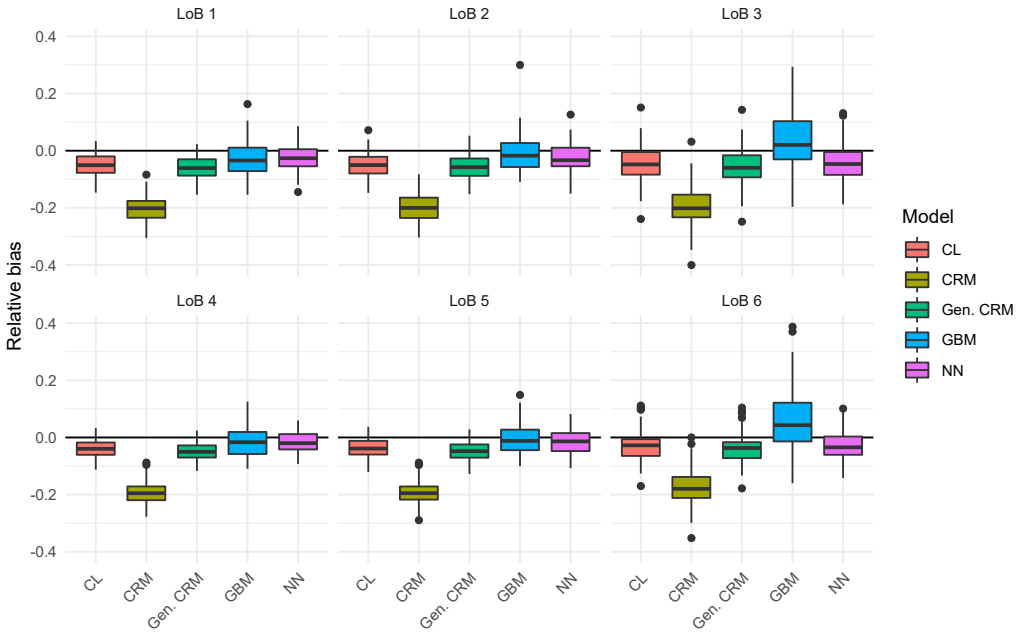


Figure 9: Boxplots of relative biases from the 100 simulations.

Table 5: Biases and RMSEP from the simulations using seeds 1-100 for the CL, CRM, GCRM, and the GBM and NN for Model 2.

bias	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
CL	-2,035	-1,945	-886	-2,927	-2,724	-1,051
CRM	-8,196	-8,017	-3,606	-14,310	-14.133	-5,889
GCRM	-2,452	-2,328	-1,073	-3.672	-3.507	-1,384
GBM	-1,162	-553	664	-937	-497	1,878
NN	-1,077	-1,109	-840	-1,212	-1,352	-1,064
RMSEP	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
CL	2,649	2,587	1,507	3,716	3,775	2,045
CRM	8,453	8,284	3,885	14,652	14,574	6,311
GCRM	2,979	2,878	1,630	4,342	4,375	2,231
GBM	2,767	2,574	1,854	3,897	4,019	3,832
NN	2,252	2,225	1,517	2,769	3,391	2,107

conditional MSEP is defined as

$$\begin{aligned}\mathrm{MSEP}(R,\widehat{R}|\mathcal{N}_0) &:= \mathbb{E}\left[\left(R-\widehat{R}\right)^2\middle|\mathcal{N}_0\right] \\ &= \mathrm{Var}\,(R|\mathcal{N}_0) + (R-\mathbb{E}[\widehat{R}|\mathcal{N}_0])^2,\end{aligned}$$

where the first term is usually referred to as the process variance and the second term as the estimation error. Given the ODP assumptions, it is straightforward to compute the process variance analytically and to utilize a bootstrap to estimate the estimation error. We detail how the process variances can be computed analytically and give a cursory description of the bootstrap in Appendix C.2. We refer the reader to [21] for further information on how to bootstrap CRM-type models.

The estimators of the conditional MSEPs for the CL, the GCRM, and the NN and GBM for Model 2, are given in Table 6. It should be noted that these MSEPs are estimated by analytical calculations and parametric bootstrapping that assume the fitted models to be the true ones – recall the discussion in Section 3.3.1 and Section 3.4.1, when introducing Model M2 - GBM and Model M2 - NN. Remember that Table 5 shows the true MSEPs based on simulating new datasets form the simulation machine, and are therefore taking model error into account. However, it is essential to note that these are unconditional MSEPs and are therefore not directly comparable to the estimated conditional ones in Table 6. Nonetheless, the conditional and unconditional MSEPs are likely not going to differ considerably, and comparing the two tables indicates that the machine learning methods give much more accurate representations of their true MSEPs than the CL, CRM, and GCRM do.

Table 6: Root mean squared error of prediction based on analytical calculation of the process variance and bootstrapping of the estimation error for the CL, the GCRM, and the GBM and NN for Model 2. n_{boot} is the number of bootstrap samples used in estimating the estimation error. Note that for the NNs we only use 100 due to the computation time being much longer than for the other models. The numbers in the parentheses are run-times (in seconds) of the bootstrap simulations.

	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6	n_{boot}
CL	1,092 (3.4)	1,313 (3.4)	460 (3.3)	2,145 (3.3)	1,900 (3.3)	995 (3.2)	1,000
GCRM	1,044 (16)	1,147 (15)	670 (15)	1,691 (15)	2,430 (15)	1,254 (15)	1,000
GBM	1,854 (33)	2,691 (24)	1,210 (37)	5,802 (20)	4,715 (33)	2,890 (25)	1,000
NN	1,828 (9,060)	2,676 (4,800)	733 (4,600)	3,053 (9,700)	3,895 (4,900)	1,425 (910)	100

5. CONCLUDING REMARKS

In the present paper we have introduced regression based reserving models that can produce separate RBNS and IBNR reserves based on aggregated data, and described how their regression functions can be modelled using complex algorithmic

machine learning techniques, including calibration and model selection. Our focus has been on GBMs and feed-forward NNs, trying to use "standard" tuning to as a wide extent as possible. Still, we have described how more detailed tuning can be conducted, which may be needed in in real world applications. In the numerical illustration of Section 4, focus has been on GBM models, since we are not aware of work in this direction based on aggregated claims data. Concerning the NN models, we have used the architecture (number of hidden layers, number of neurons per layer, etc.) from [11], but the techniques discussed for choosing different epochs applies to how to choose different architectures as well. For more on this, see e.g. [10]. One obstacle in the reserving context is that model estimation and calibration is carried out based on partially observed claims data, making it a risk to overfit to historical claims development patterns. This problem is not restricted to machine learning techniques, but is a general problem, although the risk may be higher when using very flexible complex models in these situations. In the current paper we have used the training and validation setup from [11], see also Section 3.2 above together with Remark 5. An alternative could be to use other splits of data, such as removing the last diagonal from the training data.

The overall conclusion from the paper based on Section 4 is that by using off-the-shelf software and standard tuning, machine learning techniques can be used to improve the predictive performance also in aggregated claims reserving modelling, still producing interpretable output that can be communicated to non-experts, and whose regression functions (e.g. the $\psi_{i,j,k}$s) may be altered based on expert opinions, see the discussion in Section 3.1.

A. NUMERICAL ILLUSTRATION APPENDIX

Table 7: RBNS and IBNR reserves. Relative biases of the reserve predictions in the parentheses.

IBNR reserves	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
True	1,596	1,537	603	3,593	2,739	1,048
CRM	1,867 (16.99)	1,821 (18.42)	921 (52.60)	3,823 (6.39)	4,001 (46.09)	1,817 (73.48)
GCRM	1,846 (15.68)	1,800 (17.09)	1,012 (67.64)	3,569 (-0.67)	3,479 (27.05)	1,643 (56.83)
GBM (Model 2)	1,643 (2.96)	1,652 (7.42)	957 (58.51)	3,197 (-11.02)	2,894 (5.67)	1,447 (38.11)
GBM (Model 3)	2,501 (56.75)	2,114 (37.48)	1,133 (87.76)	4,504 (25.35)	4,667 (70.41)	2,044 (95.08)
NN (Model 2)	1,780 (11.55)	1,773 (15.29)	955 (58.23)	3,293 (-8.34)	3,292 (20.20)	1,535 (46.55)
NN (Model 3)	1,889 (18.37)	1,624 (5.64)	930 (54.15)	3,407 (-5.17)	3,479 (27.05)	1,621 (54.74)
RBNS reserves	LoB 1	LoB 2	LoB 3	LoB 4	LoB 5	LoB 6
True	38,093	35,500	16,275	68,037	69,810	30,069
CRM	30,619 (-19.62)	28,080 (-20.90)	12,119 (-25.53)	51,959 (-23.63)	55,390 (-20.66)	22,586 (-24.89)
GCRM	36,447 (-4.32)	33,317 (-6.15)	14,436 (-11.29)	63,392 (-6.83)	65,918 (-5.57)	27,461 (-8.67)
GBM (Model 2)	38,054 (-0.10)	35,601 (0.28)	15,551 (-4.44)	69,482 (2.12)	68,934 (-1.25)	30,495 (1.41)
GBM (Model 3)	37,613 (-1.26)	33,615 (-5.31)	14,628 (-10.11)	64,944 (-4.55)	67,751 (-2.95)	28,017 (-6.83)
NN (Model 2)	39,807 (4.50)	35,814 (0.88)	14,725 (-9.52)	67,861 (-0.26)	68,017 (-2.57)	27,448 (-8.72)
NN (Model 3)	37,868 (-0.59)	37,095 (4.49)	15,315 (-5.90)	67,509 (-0.78)	71,121 (1.88)	27,322 (-9.14)

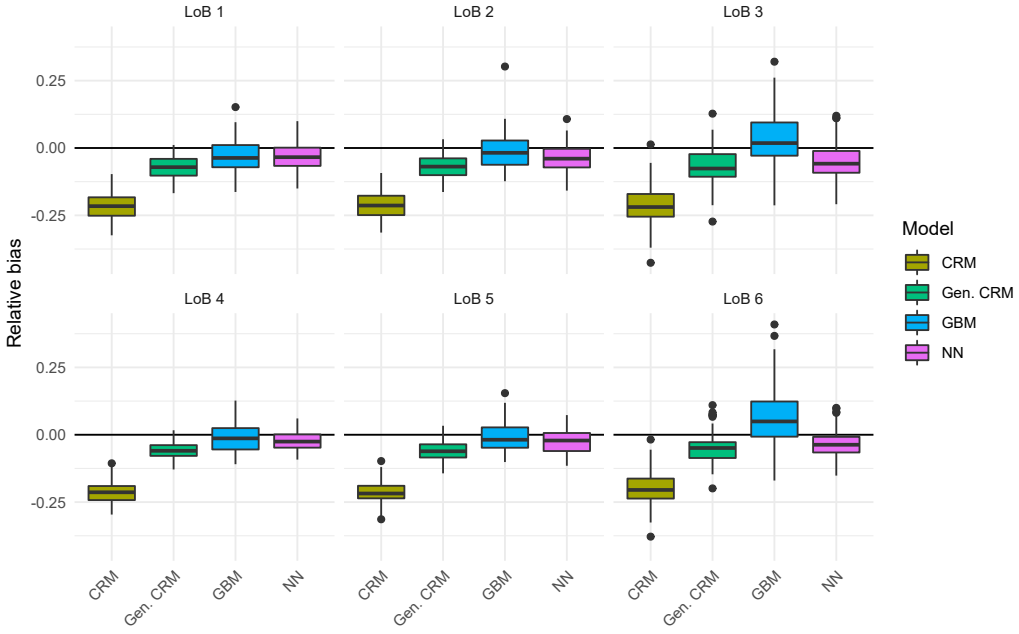


Figure 10: Boxplots of RBNS relative biases from the 100 simulations.

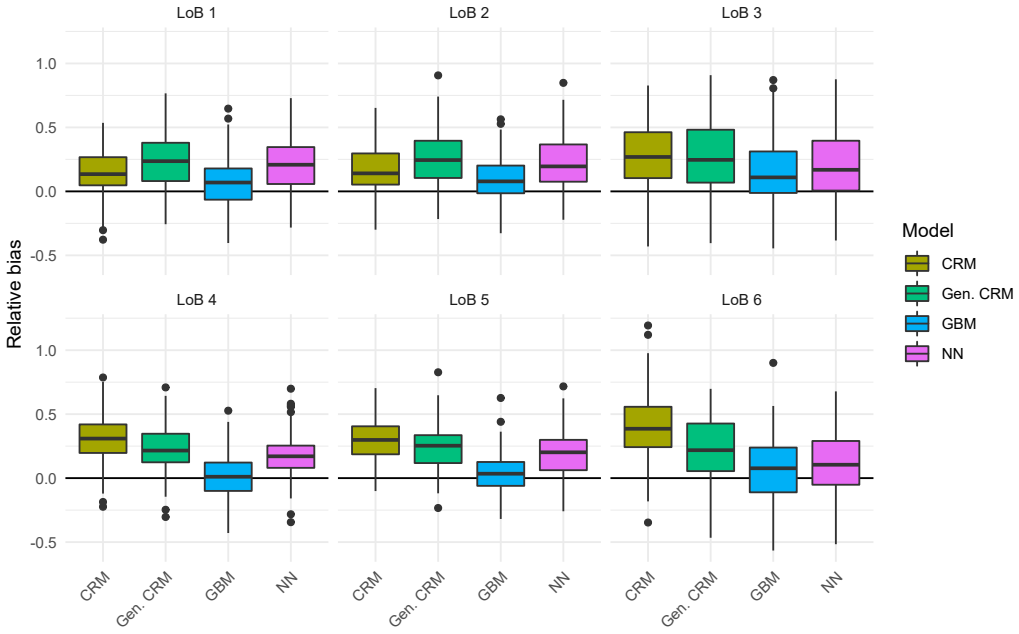


Figure 11: Boxplots of IBNR relative biases from the 100 simulations.

B. R-CODE

Listing 1: Aggregation of micro data. The data.frame `output` is generated by Listing 1 in GRW.

```
library("dplyr")
library("reshape2")
library("tidyr")

# Aggregate the micro data
df_aggregated <- output %>%
  dplyr::mutate(N = 1) %>%
  dplyr::select(LoB, AY, RepDel, N, dplyr::starts_with('Pay')) %>%
  dplyr::group_by(LoB, AY, RepDel) %>%
  dplyr::summarise_all(sum) %>%
  reshape2::melt(id = c("LoB", "AY", "RepDel", "N"),
                 variable.name = "PayDel",
                 value.name = "paid") %>%
  dplyr::mutate(PayDel = as.numeric(substr(PayDel, start = 4, stop = 5)),
                PayDel = PayDel - RepDel) %>%
  dplyr::filter(PayDel >= 0)

# Add cells without any reported claims
dat_X <- df_aggregated %>%
  tidyr::complete(LoB = 1:6,
                  AY = 1994:2005,
                  RepDel = 0:11,
                  PayDel = 0:11) %>%
  dplyr::group_by(LoB, AY, RepDel) %>%
  tidyr::fill(N) %>%
  dplyr::mutate(paid = ifelse(is.na(paid), 0, paid),
                N = ifelse(is.na(N), 0, N)) %>%
  dplyr::filter(PayDel + RepDel <= 11)

# Create dataset for only the number of reported claims
dat_N <- dat_X %>%
  subset(select = c(LoB, AY, RepDel, N)) %>%
  dplyr::distinct()
```

Listing 2: Fitting the GBMs for Model 2 to LoB 1.

```
library("gbm")

# Fit GBM to the number of reported claims for LoB 1
m_N_GBM <- gbm(
  formula = N ~ AY + RepDel,
  data = dat_N %>% dplyr::filter(LoB == 1, AY + RepDel + PayDel <= 2005),
  distribution = "poisson",
  interaction.depth = 2,
  shrinkage = 0.1,
  bag.fraction = 1,
  n.trees = 10000,
  n.minobsinnode = 1
)

# Fit GBM to the claims payments for LoB 1
m_X_GBM <- gbm(
  formula = paid ~ offset(log(N)) + AY + RepDel + PayDel + I(AY + RepDel),
  data = dat_X %>% dplyr::filter(LoB == 1, AY + RepDel + Paydel <= 2005),
  distribution = "poisson",
  interaction.depth = 1,
  shrinkage = 0.1,
  bag.fraction = 1,
  n.trees = 10000,
  n.minobsinnode = 1
)
```

Listing 3: Neural network architecture of GRW adapted to the payment part of Model 2.

```
# Declare features
AccYear <- layer_input(shape = c(1), dtype = "int32", name = "AccYear")
RepDel <- layer_input(shape = c(1), dtype = "int32", name = "RepDel")
PayDel <- layer_input(shape = c(1), dtype = "int32", name = "PayDel")
LogN <- layer_input(shape = c(1), dtype = "float32", name = "LogN")

# Define the embedding layers
AY_embed <- AccYear %>%
  layer_embedding(
    input_dim = 12, output_dim = 1, trainable = FALSE, input_length = 1,
    weights = list(array(alpha_ODP_train, dim = c(12, 1))), name = "AY_embed"
  ) %>%
  layer_flatten(name = "AY_flat")

RepDel_embed <- RepDel %>%
  layer_embedding(
    input_dim = 12, output_dim = 1, trainable = FALSE, input_length = 1,
    weights = list(array(beta_ODP_train, dim = c(12, 1))), name = "RepDel_embed"
  ) %>%
  layer_flatten(name = "RepDel_flat")

PayDel_embed <- PayDel %>%
  layer_embedding(
    input_dim = 12, output_dim = 1, trainable = FALSE, input_length = 1,
    weights = list(array(gamma_ODP_train, dim = c(12, 1))), name = "PayDel_embed"
  ) %>%
  layer_flatten(name = "PayDel_flat")

# Concatenate the embedding layers and add them to the CC model part
concate0 <- list(AY_embed, RepDel_embed, PayDel_embed) %>%
  layer_concatenate(name = "concate0")
CC0 <- list(AY_embed, RepDel_embed, PayDel_embed) %>% layer_add(name = "CC0")

# Define the 3 hidden layers of the NN model part
NN0 <- concate0 %>%
  layer_dense(units = 20, activation = "tanh", name = "hidden1") %>%
  layer_dropout(0.1, name = "dropout1") %>%
  layer_dense(units = 15, activation = "tanh", name = "hidden2") %>%
  layer_dropout(0.1, name = "dropout2") %>%
  layer_dense(units = 10, activation = "tanh", name = "hidden3") %>%
  layer_dropout(0.1, name = "dropout3")
```

```
# Define the bCCNN model using the skip connection for the CC part
Response <- list(CC0, NN0, LogN) %>% layer_concatenate(name = "concate1") %>%
  layer_dense(units = 1, activation = k_exp, name = "Response",
              weights = list(array(c(1, rep(0, 10), 1), dim = c(10 + 1 + 1, 1)),
                             array(intercept_ODP_train, dim = c(1))))

# Define and compile the model
model <- keras_model(inputs = c(AccYear, RepDel, PayDel, LogN),
                     outputs = c(Response))
model %>% keras::compile(optimizer = optimizer_rmsprop(), loss = "poisson")

fit <- model %>% keras::fit(x = dat_train_lst, y = y_train, epochs = 1000,
                            batch_size = batch_size,
                            validation_data = validation_data)
```

C. TECHNICAL DETAILS

C.1 Details on fitting a regression tree

Let us consider depth l of a tree that is being estimated, which implies that there are $j = 1, \ldots, 2^l$ decision sets $\mathcal{A}_{l,j} = \mathcal{A}(\pi_{l,j}, \kappa_{l,j})$ to be determined together with 2^{l+1} δ-values. That is, for each $\mathcal{A}_{l,j}$ there are two δ-values, $\delta_{l,j}^{(1)}$ which is assigned if the condition is fulfilled, and $\delta_{l,j}^{(0)}$ if the decision is not fulfilled. Let

$$\mathcal{I}_{l,j} = \{i : \text{ observation } i \text{ should be evaluated using } \mathcal{A}_{l,j}, i = 1, \ldots, m\},$$

where m corresponds to the maximal number of observations. This allows us to express the optimization problem for depth level l of the tree according to

$$\min_{\boldsymbol{\pi}_l, \boldsymbol{\kappa}_l} \min_{\boldsymbol{\delta}_l^{(1)}, \boldsymbol{\delta}_l^{(0)}} \sum_{j=1}^{2^l} \sum_{i \in \mathcal{I}_{l,j}} \left(\mathbb{1}_{\{\boldsymbol{c}_i \in \mathcal{A}_{l,j}\}} L(y_i, \delta_{l,j}^{(1)}) + \mathbb{1}_{\{\boldsymbol{c}_i \notin \mathcal{A}_{l,j}\}} L(y_i, \delta_{l,j}^{(0)}) \right).$$

This optimization step is repeated until the pre-defined tree depth k is reached.

The final $\mathcal{A}_j$ regions are obtained by taking the intersection of all $\mathcal{A}_{l,j}$ sets following the branches of the tree leading to this specific leaf. See also Figure 1.

C.2 Details on computing the conditional MSEP

In this section, we detail the estimation of the conditional MSEP by describing how one can, given the ODP assumptions, analytically compute the process variance and bootstrap to get an estimator of the estimation error.

Let us first detail how we can compute the process variances analytically, focusing on Model 2. By the ODP assumption, it holds that the process variances of the outstanding RBNS payments, i.e. those $X_{i,j,k}$s with indices $i + j \leq m$, are given by

$$\mathrm{Var}(X_{i,j,k}|\mathcal{N}_0) = \mathrm{Var}(X_{i,j,k}|N_{ij}) = \varphi\psi_{i,j,k}N_{i,j}. \tag{8}$$

Since all cells are, conditionally on the N_{ij}s, independent by assumption, it is straightforward to compute the RBNS variances by summing over appropriate indices. The RBNS variance for accident year i is

$$\begin{aligned}\mathrm{Var}(R_i^{\mathcal{R}}|\mathcal{N}_0) &= \sum_{j=0}^{m-i}\sum_{k>m-i-j}\mathrm{Var}(X_{i,j,k}|\mathcal{N}_0)\\ &= \varphi\sum_{j=0}^{m-i}\sum_{k>m-i-j}\psi_{i,j,k}N_{i,j}.\end{aligned} \tag{9}$$

The accident years are independent, so we can get the variance of the total outstanding RBNS payments by summing these variances over all accident years.

The IBNR variances are only slightly more complicated to calculate. By variance decomposition, the process variances of the outstanding IBNR payments, i.e. those $X_{i,j,k}$s with indexes $i + j > m$, are

$$\begin{aligned}\mathrm{Var}(X_{i,j,k}|\mathcal{N}_0) &= \mathrm{Var}(X_{i,j,k})\\ &= \mathrm{Var}(\mathbb{E}[X_{i,j,k}|N_{i,j}]) + \mathbb{E}[\mathrm{Var}(X_{i,j,k}|N_{i,j})]\\ &= \psi_{i,j,k}^2\,\mathrm{Var}(N_{i,j}) + \varphi\psi_{i,j,k}\mathbb{E}[N_{i,j}]\\ &= \psi_{i,j,k}^2\phi\nu_{i,j} + \varphi\psi_{i,j,k}\nu_{i,j}\\ &= (\psi_{i,j,k}\phi + \varphi)\psi_{i,j,k}\nu_{i,j}.\end{aligned} \tag{10}$$

Since we are not conditioning on the N_{ij}s for the IBNR claims, $X_{i,j,k}$ and $X_{i,j,k'}$ are dependent. Therefore we may not add these variances together as we did for the RBNS claims. Instead, the process variance for the outstanding IBNR payments in

accident year i is given by

$$\begin{aligned}
\mathrm{Var}(R_i^I \,|\mathcal{N}_0) &= \sum_{j=m-i+1}^{m-1} \mathrm{Var}\left(\sum_k X_{i,j,k}\right) \\
&= \sum_{j=m-i+1}^{m-1} \left(\mathbb{E}\left[\mathrm{Var}\left(\sum_k X_{i,j,k}\middle| N_{i,j}\right)\right] + \mathrm{Var}\left(\mathbb{E}\left[\sum_k X_{i,j,k}\middle| N_{i,j}\right]\right)\right) \\
&= \sum_{j=m-i+1}^{m-1} \left(\mathbb{E}\left[\sum_k \mathrm{Var}\left(X_{i,j,k}\middle| N_{i,j}\right)\right] + \mathrm{Var}\left(\sum_k \psi_{i,j,k} N_{i,j}\right)\right) \\
&= \sum_{j=m-i+1}^{m-1} \left(\mathbb{E}\left[\sum_k \varphi\psi_{i,j,k} N_{i,j}\right] + \left(\sum_k \psi_{i,j,k}\right)^2 \phi\nu_{i,j}\right) \\
&= \sum_{j=m-i+1}^{m-1} \left(\varphi + \phi \sum_k \psi_{i,j,k}\right) \nu_{i,j} \sum_k \psi_{i,j,k} \qquad (11)
\end{aligned}$$

As for the RBNS variances, these IBNR variances can be added together to get the variance of the total outstanding payments from IBNR claims. Finally, since the RBNS and IBNR claims are independent, to get the variance of the total outstanding payments, we add the RBNS and IBNR variances together. Now we have all the ingredients to compute the process variance, and we, therefore, move on to estimating the estimation error using a parametric bootstrap.

For the parametric bootstrap, we follow the algorithms in Section 6.2 in [21] to get bootstrap samples of the RBNS and IBNR reserve predictions in CRM-type models. These algorithms, adapted to our situation, is given at the end of this section. To follow these algorithms, we need to simulate new in-sample data. There are many possible ways of simulating from these ODP models. A standard approach is to use the following (which is used in e.g. [10] and [11]): If

$$\phi N_{i,j} \sim \mathrm{Po}(\nu_{i,j}/\phi),$$

then

$$N_{i,j} \sim \mathrm{ODP}(\nu_{i,j}, \phi).$$

That is, we may simulate from a Poisson distribution with mean $\nu_{i,j}/\phi$ and then multiply the observations with the dispersion parameter ϕ to get a random sample from $\mathrm{ODP}(\nu_{i,j}, \phi)$. For this, however, we need an estimator of ϕ. Here we will use the standard one based on the Pearson statistic (see e.g. p. 328 in [20]) which, for the

number of reported claims part of the model, is given by

$$\widehat{\phi} := \frac{1}{n - p_\nu} \sum_{i+j \le m} \frac{(N_{i,j} - \widehat{\nu}_{i,j})^2}{\widehat{\nu}_{i,j}},$$

where p_ν is the number of parameters ($p_\nu = 2m - 1$ for a cross-classified structure), and n is the number of observations in the upper left triangle. The estimator $\widehat{\varphi}$ is defined analogously as

$$\widehat{\varphi} := \frac{1}{n - p_\psi} \sum_{i+j+k \le m} \frac{(X_{i,j,k} - \widehat{\psi}_{i,j,k} N_{i,j})^2}{\widehat{\psi}_{i,j,k} N_{i,j}},$$

where n now is the sample size of the $X_{i,j,k}$s and p_ψ the number of parameters used in estimating the $\psi_{i,j,k}$s, which would be $3m - 2$ in a cross-classified model structure.

By computing the process variances according to (9) and (11), and bootstrapping the estimation error using the below algorithms, we have all the ingredients needed to estimate the conditional MSEP. To compute the estimation error for the total reserve, we use the bootstrap samples acquired in Step 6 in the two algorithms below by computing the sample average

$$\frac{1}{B} \sum_{b=1}^{B} (\widehat{R}^{\mathcal{R}} + \widehat{R}^{\mathcal{I}} - (\widehat{R}^{\mathcal{R},*}_{(b)} + \widehat{R}^{\mathcal{I},*}_{(b)}))^2$$

The following algorithm follows the underlying principles from [21]:

Algorithm RBNS

Step 1. *Estimation of the parameters.* Estimate the payment part parameters of the model using the original data to get the estimators $\widehat{\psi}_{i,j,k}$ and $\widehat{\varphi}$.

Step 2. *Bootstrapping the data.* Keep the same counts $N_{i,j}$, but generate new bootstrapped aggregated payments $\{X^*_{i,j,k} : i+j+k \le m\}$ by simulating from $\text{Po}(\widehat{\psi}_{ijk} N_{ij} / \widehat{\varphi})$ and multiplying by $\widehat{\varphi}$.

Step.3 *Bootstrapping the parameters.* Compute the estimators $\widehat{\psi}^*_{i,j,k}$ using $\{N_{i,j} : i + j \le m\}$ and the bootstrap data $\{X^*_{i,j,k} : i + j + k \le m\}$.

Step 4. *Bootstrapping the RBNS predictions.* Using the original incurred claims $\{N_{i,j} : i + j \le m\}$ and the bootstrap parameters $\widehat{\psi}^*_{i,j,k}$, compute the RBNS reserve prediction $\widehat{R}^{\mathcal{R},*}$ according to (4).

Step 5. *Monte Carlo approximation.* Repeat steps 2-4 $\boldsymbol{B}$ times to get an approximate bootstrap distribution of the RBNS reserve from the bootstrapped $\{\widehat{R}^{\mathcal{R},*}_{(b)}\}_{b=1}^{B}$.

Algorithm IBNR

Step 1. *Estimation of the parameters.* Estimate the parameters of the model using the original data to get the estimators $\widehat{\nu}_{i,j}$, $\widehat{\psi}_{i,j,k}$, $\widehat{\phi}$, and $\widehat{\varphi}$.

Step 2. *Bootstrapping the data.* Generate new bootstrapped data $\{N^*_{i,j} : i+j+k \leq m\}$ and $\{X^*_{i,j,k} : i+j+k \leq m\}$ by simulating the $X^*_{i,j,k}$s exactly as described in Step 2 of the RBNS algorithm above and the $N^*_{i,j}$s by simulating from $\text{Po}(\widehat{\nu}_{i,j}/\widehat{\phi})$ and multiplying by $\widehat{\phi}$.

Step.3 *Bootstrapping the parameters.* Compute the estimators $\widehat{\psi}^*_{i,j,k}$ using $\{N_{i,j} : i+j \leq m\}$ and the bootstrap data $\{X^*_{i,j,k} : i+j+k \leq m\}$, and compute the estimators $\widehat{\nu}^*_{i,j}$ using $\{N^*_{i,j} : i+j \leq m\}$.

Step 4. *Bootstrapping the IBNR predictions.* Using the bootstrap parameters $\widehat{\nu}_{i,j}$ and $\widehat{\psi}_{ijk}$, compute the IBNR reserve prediction $\widehat{R}^{\mathcal{I},*}$ according to (5).

Step 5. *Monte Carlo approximation.* Repeat steps 2-4 $\boldsymbol{B}$ times to get an approximate bootstrap distribution of the IBNR reserve from the bootstrapped $\{\widehat{R}^{\mathcal{I},*}_{(b)}\}_{b=1}^{B}$.

REFERENCES

[1] Katrien Antonio and Richard Plat. Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014(7):649–669, 2014.

[2] Elja Arjas. The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin: The Journal of the IAA*, 19(2):139–152, 1989.

[3] John T Bonsignore, Joseph O Marker, Julie Sims, Yisheng Bu, Gary V Nickerson Greg Taylor, Sandie Cagley, Bruce E Ollodart, Gary G Venter, David R Clark, Dianne M Phelps, et al. The analysis and estimation of loss & alae variability: A summary report. 2005.

[4] H Bühlmann, Rene Schnieper, and Erwin Straub. Claims reserves in casualty insurance based on a probabilistic model. *Bulletin of the Swiss Association of Actuaries*, 80:21–45, 1980.

[5] Massimo De Felice and Franco Moriconi. Claim watching and individual claims reserving using classification and regression trees. *Risks*, 7(4):102, 2019.

[6] Francis Duval and Mathieu Pigeon. Individual loss reserving using a gradient boosting-based approach. *Risks*, 7(3):79, 2019.

[7] Bradley Efron and Trevor Hastie. *Computer age statistical inference*, volume 5. Cambridge University Press, 2016.

[8] Peter D England and Richard J Verrall. Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3):443–518, 2002.

[9] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[10] Andrea Gabrielli. A neural network boosted double overdispersed poisson claims reserving model. *ASTIN Bulletin: The Journal of the IAA*, 50(1):25–60, 2020.

[11] Andrea Gabrielli, Ronald Richman, and Mario V Wüthrich. Neural network embedding of the over-dispersed poisson reserving model. *Scandinavian Actuarial Journal*, pages 1–29, 2019.

[12] Andrea Gabrielli and Mario V Wüthrich. An individual claims history simulation machine. *Risks*, 6(2):29, 2018.

[13] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. Available at http://www.deeplearningbook.org.

[14] Trevor Hastie, Jerome Friedman, and Robert Tibshirani. *The elements of statistical learning*. Springer series in statistics New York, NY, USA:, 2008.

[15] Kevin Kuo. Deeptriangle: A deep learning approach to loss reserving. *Risks*, 7(3):97, 2019.

[16] Christian Roholte Larsen. An individual claims reserving model. *ASTIN Bulletin: The Journal of the International Actuarial Association*, 37(01):113–132, 2007.

[17] Mathias Lindholm and Richard Verrall. On distribution-free reserving and partial information. *Stockholm university research reports in Mathematical statistics*, 2019:14, 2019.

[18] Olivier Lopez, Xavier Milhaud, Pierre-E Thérond, et al. Tree-based censored regression with applications in insurance. *Electronic journal of statistics*, 10(2):2685–2716, 2016.

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

[20] Peter McCullagh and John A Nelder. *Generalized linear models*. Chapman & Hall/CRC, 1989.

[21] María Dolores Martínez Miranda, Bent Nielsen, Jens Perch Nielsen, and Richard Verrall. Cash flow simulation for a model of outstanding liabilities based on claim amounts and claim numbers. *ASTIN Bulletin: The Journal of the IAA*, 41(1):107–129, 2011.

[22] María Dolores Martínez Miranda, Jens Perch Nielsen, and Richard Verrall. Double chain ladder. *ASTIN Bulletin: The Journal of the IAA*, 42(1):59–76, 2012.

[23] Ragnar Norberg. Prediction of outstanding liabilities in non-life insurance 1. *ASTIN Bulletin: The Journal of the IAA*, 23(1):95–115, 1993.

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

[25] Greg Taylor. Loss reserving models: Granular and machine learning forms. *Risks*, 7(3):82, 2019.

[26] Richard Verrall, Jens Perch Nielsen, and Anders Hedegaard Jessen. Prediction of rbns and ibnr claims using claim amounts and claim counts. *ASTIN Bulletin: The Journal of the IAA*, 40(2):871–887, 2010.

[27] Felix Wahl, Mathias Lindholm, and Richard Verrall. The collective reserving model. *Insurance: Mathematics and Economics*, 87:34–50, 2019.

[28] Mario V Wüthrich. Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, 2018(6):465–480, 2018.

[29] Mario V Wüthrich. Neural networks applied to chain–ladder reserving. *European Actuarial Journal*, 8(2):407–436, 2018.

[30] Mario V Wüthrich and Michael Merz. Yes, we cann! *ASTIN Bulletin: The Journal of the IAA*, 49(1):1–3, 2019.

Joint Model for Individual-Level Loss Reserving: An Empirical Analysis

A. Nii-Armah Okine, Edward W. Frees (FSA), Peng Shi (FSA)

In non-life insurance, actuarial analysts commonly encounter situations where the duration of settlement is positively associated with the size of payments for individual claims. The history of paid losses could help predict both the settlement time and outstanding payments, hence ignoring the payment-settlement association could lead to inaccurate reserve prediction. This paper introduces a joint model framework where the payment process and the settlement process of the claim are joined via shared latent variables to help improve prediction accuracy. We present a detailed empirical analysis using data from a property insurance provider. The joint model is fitted to a training dataset, and the fitted model is used to predict the future development of open claims. The prediction results from an out-of-sample data show the joint model framework outperforms existing reserving models that ignore the payment-settlement association. We also propose a novel form of cross-validation for longitudinal data named double cross-validation.

Keywords. Joint longitudinal-survival, micro-level loss reserving, RBNS reserve, dependence modeling, cross-validation.

1. INTRODUCTION

The loss reserve estimate is the most substantial liability on a non-life insurer's balance sheet (Grace and Leverty 2012). Therefore, the accuracy in the reserve prediction is critical for insurers to prevent insolvency issues and remain competitive as reserve estimates do affect not only pricing decisions but also the decision making of internal management, regulators, and investors (Friedland 2010).

In claims management, it is common that small claims are settled faster than large claims, because large and complicated claims naturally require experienced adjusters, demand special expertise, involve multiple interested parties, and are more likely to be litigated. As a result, the duration of settlement and size of payments for individual claims are often positively correlated. The payment-settlement association has important implications for the loss reserving practice. In loss reserving, actuaries predict the outstanding liabilities based on the claim history that is only observed up to a valuation date. When the settlement time and claim size are correlated and not accounted for, the historical claims that actuaries use for model building will not be representative of future payments, because large claims with longer settlement times will be more likely to be censored (not settled) by the

valuation date, a type of selection bias. Specifically, when larger claims take more time to settle, outstanding payments would be underestimated if the selection bias in the sampling procedure is not accounted for. Similarly, one would expect overestimation of future payments if the claim size and settlement time were negatively correlated.

Further, the payment-settlement association means that payment history may help predict settlement time, which in turn feeds back into the prediction of unpaid losses. Then the relation between the two processes allows for the dynamic prediction of outstanding liabilities. The prediction is dynamic because, at a future date, when more information becomes available, the settlement time and ultimate payment predictions can be updated.

This paper introduces the joint model (JM) for longitudinal and time-to-event data into the micro-level loss reserving literature to account for the payment-settlement association. We present a detailed empirical analysis of the joint model framework using data from a property insurance provider with the focus on Reported But Not Settled (RBNS) reserve prediction.

1.1 Research Context

In non-life insurance, loss reserve prediction is usually based on macro-level models that use aggregate loss data summarized in a run-off triangle and the chain-ladder (CL) method is the most commonly used macro-level model (Wüthrich and Merz 2008). The main strengths of the macro-level models are that they are easy to implement and interpret, while the limited ability to handle heterogeneity and environmental changes are the most significant drawbacks, which may lead to inaccurate predictions. In this context, "environmental changes" refers to changes in the insurer's business that can affect loss reserving, for example, underwriting practices, claims processing, mix of products, and so forth. When practicing actuaries believe that the losses are heterogeneous, then they are often segmented by specific discrete characteristics and compiled into multiple triangles. This approach to addressing heterogeneity becomes problematic when the source of heterogeneity is not clear or is a continuous variable. Further, the reduction in the number of claims in each portfolio can lead to credibility issues. Also, Friedland (2010) examined the effects of environmental changes on reserve prediction and found that the chain-ladder type methods are appropriate only in a steady state (stable environment). In the case of environmental changes, some of the commonly-used macro-models can generate a

reserve estimate without material errors. To handle environmental changes, macro-level methods consider either expected loss techniques that allow actuaries to incorporate a priori reserve estimate or trending techniques that treat environmental change as a trend to adjust the development projections. However, highly dependent on actuaries' judgments, both techniques could lead to problematic reserve estimates (Jin, 2014).

Recently, micro-level reserving techniques have gained traction as they allow an analyst to use the information on the policy, the individual claim, and the development process in the prediction of outstanding liabilities. Granular covariate information allows one to account for both claim and policy specific effects, and thus naturally captures claim heterogeneities and environmental changes. Under the individual-level reserving literature, the marked Poisson processes (MPP) framework introduced by (Arjas 1989; Jewell 1989; Norberg 1993, 1999) and first applied in Antonio and Plat (2014) constitute the dominant family of research. The MPP represents events, such as claims or claim payments, as a collection of time points on a timeline with some additional features (called marks) measured at each point. Generalized linear models (GLMs), in conjunction with survival analysis, have also been applied to the loss reserving problem (Taylor and Campbell 2002; Taylor and McGuire 2004; Taylor, McGuire, and Sullivan 2008). In addition, a growing stream of research for individual-level reserving focuses on using machine learning algorithms. The machine learning algorithms do not assume a structural form for the claims data and provide a data-driven approach. Wüthrich (2018a) illustrates the use of regression trees, and Wüthrich (2018b) focused on the application of neural networks for individual-level reserving.

1.2 Objective

In this paper, we fit the joint model to a training dataset, and the association between the payment history and settlement time is captured, which helps to accurately predict the settlement time and the ultimate amount of unsettled losses. The RBNS prediction performance of the JM is compared to existing reserving models using out–of–sample data. Because of the time dimension involved with the RBNS reserve prediction, the traditional cross-validation techniques cannot be used to evaluate the robustness of the prediction results. Thus, we introduce a novel form of cross-validation for longitudinal data, which we call *double cross-validation*.

Concerning the CL method, the payment-settlement association is not entirely ignored. Under a steady state where the development triangle is homogeneous, if the length of the triangle is appropriate, there should not be any bias introduced because payment sizes and settlement times are correlated. However, there may be a bias for newer companies that may not have an extensive history. Further, the JM framework improves the accuracy in unpaid losses prediction compared to macro-level models by leveraging claim level granular information to control for heterogeneity and environmental changes. The JM framework also offers an improvement over the existing individual-level reserving models by explicitly accounting for the payment-settlement association, hence addressing the issue of selection bias.

1.3 Outline

The remainder of the paper proceeds as follows. Section 2 presents the joint model framework and describes the property insurance claims dataset and its important characteristics that motivate the joint modeling framework. Section 3 provides estimation results from the joint model using a training dataset and prediction results using a hold-out sample. Section 4 concludes.

2. BACKGROUND AND METHODS

2.1 Joint Model Framework

The existing micro-level reserving methods do not explicitly capture the dependence between the payment history and settlement process. We further extend the literature by introducing the joint longitudinal-survival model (JM) framework to allow for such association.

The joint model has been proposed in the medical statistics literature for modeling longitudinal and survival outcomes when the two components are correlated (Elashoff, Li, and Li 2016). Two general frameworks have received extensive attention, the pattern mixture model and the selection model (Little 2008). These two frameworks differ in the way the joint distribution is factorized. In the former, the joint distribution is specified using the marginal distribution of time-to-event outcome and the conditional distribution of longitudinal outcomes given the time-to-event outcome. In contrast, the joint distribution in the latter is specified using models for the marginal distribution of longitudinal

outcomes and the conditional distribution of time-to-event outcome given longitudinal outcomes. Diggle and Kenward (1994) were first to apply selection models to non-random drop-out in longitudinal studies by allowing the drop-out probabilities to depend on the history of the measurement process up to the drop-out time. The two model families are primarily applied with discrete drop-out times and cannot be easily extended to continuous time.

The properties of the joint models have been well-developed in the biomedical literature in clinical studies (Ibrahim, Chu, and Chen 2010) and non-clinical studies (Liu 2009). Tsiatis and Davidian (2004), Yu et al. (2004), and Verbeke, Molenberghs, and Rizopoulos (2010) give excellent overviews of joint models. Besides, Rizopoulos (2010) and Rizopoulos (2016) develop R packages for joint models.

2.1.1 General Framework

In this section, we introduce the JM framework to the loss reserving problem, focusing on a subset of selection models called shared-parameter models. In shared-parameter models, a latent random effects vector $\mathbf{b}_i$ is used to capture the association between the longitudinal and the time-to-event outcomes (Rizopoulos 2012). For the loss reserving problem, the longitudinal sub-model represents the payment process for a given claim where the sequence of payments from a reported claim forms the longitudinal outcomes, and the survival sub-model drives the settlement process of the claim where the settlement time of the claim is the time-to-event outcome of interest.

For the ith claim ($i = 1, \dots, N$), we set the time origin for a claim as its reporting time. We denote T_i^* and c_i as the settlement time and valuation time, respectively. Assuming c_i is independent of T_i^*, define $T_i = \min(T_i^*, c_i)$ and $\Delta_i = I(T_i^* < c_i)$, where $I(A) = 1$ when A is true and $I(A) = 0$ otherwise. The pair (T_i, Δ_i) makes up the observable time-to-settlement outcomes for claim i, where Δ_i indicates whether the claim has been closed by the valuation time; if so, T_i indicates the settlement time. Let $\{\mathbf{Y}_i(t): 0 \le t \le T_i^*\}$ be the payment process, and $\mathbf{Y}_i^* = \{Y_{it}, t \in \tau_i^*\}$ be the vector of the complete cumulative payments for claim i with n_i^* payments at times $\tau_i^* = \{t_{ij}; j = 1, \dots, n_i^*\}$. Assume there are n_i payments by the time of valuation, we define $\tau_i = \{t_{ij}; j = 1, \dots, n_i\}$ as the observable payment times and denote $\mathbf{Y}_i = \{Y_{it}, t \in \tau_i\}$ the vector of cumulative payments at

observed time of payments. Further denote $\mathbf{Y}_i^+ = \{Y_{it}, t \in \tau_i^+\}$ the vector of cumulative payments at future times $\tau_i^+ = \{t_{ij}; j = n_i + 1, \ldots, n_i^*\}$ after the valuation time. Then the joint distribution $f_{\mathbf{Y}_i^*, T_i^*}(\mathbf{y}_i^*, t_i^*)$ is given by:

$$f_{\mathbf{Y}_i^*, T_i^*}(\mathbf{y}_i^*, t_i^*) = \int f(\mathbf{y}_i^* | \mathbf{b}_i) f(t_i^* | \mathbf{b}_i) dF(\mathbf{b}_i). \tag{2.1}$$

Figure 1 provides a graphical illustration of the cumulative payment process whose experience jumps at the time of each payment from the time of reporting to settlement. The left panel presents a closed claim where the entire development process of the claim is observed before the valuation time, i.e. $(\Delta_i = 1, n_i = n_i^*)$. The right panel provides an example of an open claim where only a part of the development process of the claim is observed at the valuation time, i.e. $(\Delta_i = 0, n_i \leq n_i^*)$.

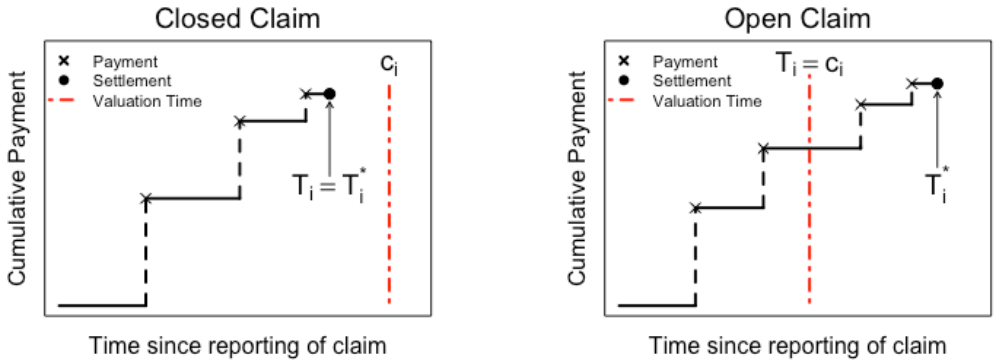


Figure 1: Graphical illustration of the cumulative payment process from the time of reporting to settlement.

2.1.2 Longitudinal sub-model of claim payments

We specify a generalized linear mixed effect model (GLMMs) for the cumulative payments Y_{it}, where the claim-specific unobserved heterogeneity is accounted for through a vector of random effects $\mathbf{b}_i$. The GLMMs extend GLMs by including random or subject-specific effects in addition to the traditional fixed effects in the structure for the mean (Antonio and Beirlant 2007). See Frees (2004)

and Molenberghs and Verbeke (2006) for details on GLMMs. Given the random effects $\mathbf{b}_i$, the cumulative payment Y_{it} is assumed to be independent and from the exponential family. Using a link function $g(\cdot)$, the conditional mean is specified as a linear combination of covariates given by:

$$\eta_{it} = g(E[Y_{it}|\mathbf{b}_i]) = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\mathbf{b}_i. \tag{2.2}$$

Here, $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ are the vectors of covariates in the fixed and random effects, respectively, and $\boldsymbol{\beta}$ is the vector of the parameters for the fixed effects. In this model, we assume $\mathbf{b}_i$ are independent and follow a multivariate normal distribution.

2.1.3 Survival sub-model of claim settlement

The association between the claim payment process and the settlement process is introduced through the effects of η_{it} on the hazard of settlement. Then the time-to-settlement outcome T_i^* of a claim is modeled using a proportional hazard model specified as:

$$h_i(t) = h_0(t)\exp\{\boldsymbol{\gamma}'\mathbf{w}_{it} + \alpha\eta_{it}\}, \tag{2.3}$$

From (2.3), the survival function of T_i^* is:

$$S_i(t) = \exp\left(-\int_0^t h_0(s)\exp\{\boldsymbol{\gamma}'\mathbf{w}_{it} + \alpha\eta_{it}\}\,ds\right). \tag{2.4}$$

where $h_0(t)$ is the baseline hazard, $\mathbf{w}_{it}$ is a vector of covariates and $\boldsymbol{\gamma}$ is the vector of the corresponding regression coefficients. The strength of the association is measured by α, where a positive payment-settlement association given by a negative α implies larger payments take a longer time to settle and vice versa. For the baseline hazard in (2.3) we consider the Weibull model or a more flexible model where the baseline hazard is approximated using splines. The Weibull baseline is given by:

$$h_0(t) = \lambda k t^{k-1}, \tag{2.5}$$

where λ is the scale parameter, and k is the shape parameter. For the baseline hazard using splines, we have:

$$\log h_0(t) = \lambda_0 + \sum_{k=1}^{K} \lambda_k B_k(t, q). \tag{2.6}$$

Here, $\lambda = (\lambda_0, \lambda_1, \cdots, \lambda_K)$ are the spline coefficients, $B_k(\cdot)$ is a B-spline basis function, q denotes the degree of the B-spline basis function, and $K = q + m$; where m is the number of interior knots.

2.2 Joint Model Estimation

In the joint model, it is assumed that the vector of time-independent random effects $\mathbf{b}_i$ underlies both the longitudinal and survival processes. This means that conditioning on the shared random effects $\mathbf{b}_i$, the joint likelihood for unknown parameters can be formulated as separate models for the longitudinal payment process and the settlement process. The parameters of the joint model are estimated using a likelihood-based method, and the likelihood function for the observables $(t_i, \delta_i, \mathbf{y}_i)$ of claim i is shown as:

$$\begin{aligned} \mathrm{L}(\boldsymbol{\theta}; t_i, \delta_i, \boldsymbol{y}_i) &= \int f(\boldsymbol{y}_i|\mathbf{b}_i; \boldsymbol{\theta}) f(t_i, \delta_i|\mathbf{b}_i; \boldsymbol{\theta}) dF(\mathbf{b}_i; \boldsymbol{\theta}) \\ &= \int \left[\prod\nolimits_{t \in \tau_i} (\boldsymbol{y}_{it}|\mathbf{b}_i; \boldsymbol{\theta}) \right] f(t_i, \delta_i|\mathbf{b}_i; \boldsymbol{\theta}) f(\mathbf{b}_i; \boldsymbol{\theta}) d\mathbf{b}_i, \end{aligned} \tag{2.7}$$

where

$$f(t_i, \delta_i|\mathbf{b}_i; \boldsymbol{\theta}) = (h_i(t_i|\mathbf{b}_i; \boldsymbol{\theta}))^{\delta_i} S_i(t_i|\mathbf{b}_i; \boldsymbol{\theta}). \tag{2.8}$$

Here, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ summarizes the parameters of the longitudinal sub-model including both regression coefficients and variance components, and $\boldsymbol{\theta}_2$ summarizes the parameters of the survival sub-model that includes baseline hazard, regression coefficients, and association between claim payment and settlement.

2.3 Prediction Using Joint Model

At the valuation time, an open claim i is characterized by the time since reporting c_i and the longitudinal claim history $\mathcal{Y}_i(c_i) = \{y_{it}, 0 \le t \le c_i\}$. Since the claim is open, the settlement time $T_i^* > c_i$. With the fitted joint model, we can obtain the RBNS reserve prediction for an open claim at the valuation time, $\hat{R}_i^{RBNS}(c_i)$, using the following steps:

a) Predict the future time when the ith claim will be settled, $\hat{u}_i$, given $T_i^* > c_i$ and $\mathcal{Y}_i(c_i)$ using an estimate of the conditional survival probability shown as:

$$\hat{\pi}_i(u|c_i) = \frac{S_i\left(u|\hat{\eta}_{iu}, \mathbf{w}_{iu}; \widehat{\boldsymbol{\theta}}\right)}{S_i\left(c_i|\hat{\eta}_{ic_i}, \mathbf{w}_{ic_i}; \widehat{\boldsymbol{\theta}}\right)}, \tag{2.9}$$

where the $\hat{S}_i(\cdot)$ is an estimate of (2.4) using the MLE estimates $\widehat{\boldsymbol{\theta}}$, $u > c_i$, $\hat{\eta}_{iu} = \mathbf{x}'_{iu}\widehat{\boldsymbol{\beta}} + \mathbf{z}'_{iu}\hat{\mathbf{b}}_i$ and $\hat{\eta}_{ic_i} = \mathbf{x}'_{ic_i}\widehat{\boldsymbol{\beta}} + \mathbf{z}'_{ic_i}\hat{\mathbf{b}}_i$. Here, $\widehat{\boldsymbol{\beta}}$ are the fixed effects maximum likelihood estimates and $\hat{\mathbf{b}}_i$ are the empirical Bayes estimate for the random effects. The time-to-settlement for a RBNS claim, $\hat{u}_i = E(T^*|T_i^* > c_i, \mathcal{Y}_i(c_i); \widehat{\boldsymbol{\theta}})$ is given by:

$$\hat{u}_i = \int_{c_i}^{\infty} \hat{\pi}_i(u|c_i)\, du. \tag{2.10}$$

b) Predict the ultimate payment, given $\mathcal{Y}_i(c_i)$ and $T_i^* > c_i$ using:

$$\hat{Y}_i(u) = g^{-1}\left(\mathbf{x}'_{iu}\widehat{\boldsymbol{\beta}} + \mathbf{z}'_{iu}\hat{\mathbf{b}}_i\right). \tag{2.11}$$

Here, $g^{-1}(\cdot)$ is the inverse of the link function, and $\{\mathbf{x}_{iu}, \mathbf{z}_{iu}\}$ are covariates. With the time-to-settlement $\hat{u}_i$, $\hat{Y}_i^{ULT}(\hat{u}_i)$ is the predicted ultimate amount of the claim.

c) With the cumulative payment for the ith claim at valuation time, $Y_i(c_i)$, we have:

$$\hat{R}_i^{RBNS}(c_i) = \hat{Y}_i^{ULT}(\hat{u}_i) - Y_i(c_i). \tag{2.12}$$

The total RBNS reserve amount is given by:

$$\hat{R}^{RBNS}(c) = \sum_{i=1}^{m} \hat{R}_i^{RBNS}(c_i). \quad (2.13)$$

Here, m is the number of open claims at the valuation time, i.e. $m = \sum_{i=1}^{N} I(\delta_i = 0)$.

2.4 Data

The data analyzed in this paper is from the Wisconsin Local Government Property Insurance Fund (Wisconsin LGPIF). The Wisconsin LGPIF was established to make property insurance available for local government units, such as counties, cities, towns, villages, school districts, library boards, etc. The Wisconsin LGPIF offers three major types of coverage for local government properties: building and contents, inland marine (construction equipment), and motor vehicles. The Fund closed in 2017. When it was operational, on average, it wrote approximately $25 million in premiums and $75 billion in coverage each year; and it insured over a thousand entities.

Exposure information is available from January 1, 2006, to December 31, 2013, and we focus on claims from the building and contents coverage. The training data contain claims that have occurred and were reported between January 1, 2006, and December 31, 2009, which we call the training sample. The training sample contains 3,393 reported claims, including 129 claims reported but with no payment transaction by the valuation date, and 34 claims with partial payments but not settled by the valuation date. The validation sample contains actual payments from January 1, 2010, to December 31, 2013, on claims reported between January 1, 2006, and December 31, 2009. Thus, the validation sample contains 163 claims, with total actual unpaid losses of $4,511,490. The validation sample is used to evaluate the quality of the reserve predictions from the fitted models.

Table 1 describes variables in the Wisconsin LGPIF dataset, including the covariate information about the policy, policy-holder, claim, and transactions used in the model building. There were other covariates like coverage amount not shown here because they were not statistically significant in either the survival or longitudinal sub-models of the fitted joint model. Figure 2 plots the distribution of ultimate payments against the settlement time in quarters (days$/366 \times 4$) for claims in the training dataset. The settlement time of a claim is defined as the closed date minus the reported date plus one day. The solid line in the right panel is the fit of the loess scatterplot smoother. Both plots suggest a

strong positive relation between ultimate payment and settlement time, i.e., it takes longer to close larger claims. The insight provided from the payment-settlement association plot in Figure 2 shows that we can do a better job in reserve prediction by incorporating the payment-settlement association in the prediction process as the development of payment may yield early indications of an impending settlement.

Table 1: Description of variables in the Wisconsin LGPIF data.

Variable	Description
ClaimNum	Unique number of claim
TransNum	Unique transaction number
	Outcome Variables
PaymentBeforeDeductible	Incremental payments before subtracting deductibles
ClaimStatus	Status of claim (Open/Closed)
ClosedDate	Settlement date for claims that are already closed
	Claim/Transaction Covariates
LnInitialEst	Initial estimates for reported claims in logarithmic of dollars
LossDate	Date of claim occurrence
LossYear	Year of claim occurrence
LossQtr	Quarter in the year of claim occurrence
ReportedDate	Date claim was reported to the Fund
ReportDelay	Reporting delay (ReportedDate-LossDate)
AcctgDate	Date for transactions
CauseCode	A code to identify the peril type of each claim
TimeToPayment	Time from reporting to payment (AcctgDate-ReportedDate)
	Policy/Policyholder Covariates
LnPolicyDed	Deductible for the policy in logarithmic of dollars
EntityType	Categorical variable that is one of six types: Village, City, County, Misc, School, or Town
CountyCode	A code to identify which of the 72 counties the entity belongs to
Region	Categorical variable which identifies region the county belongs to: Northern,Northeastern, Southeastern, Southern,or Western

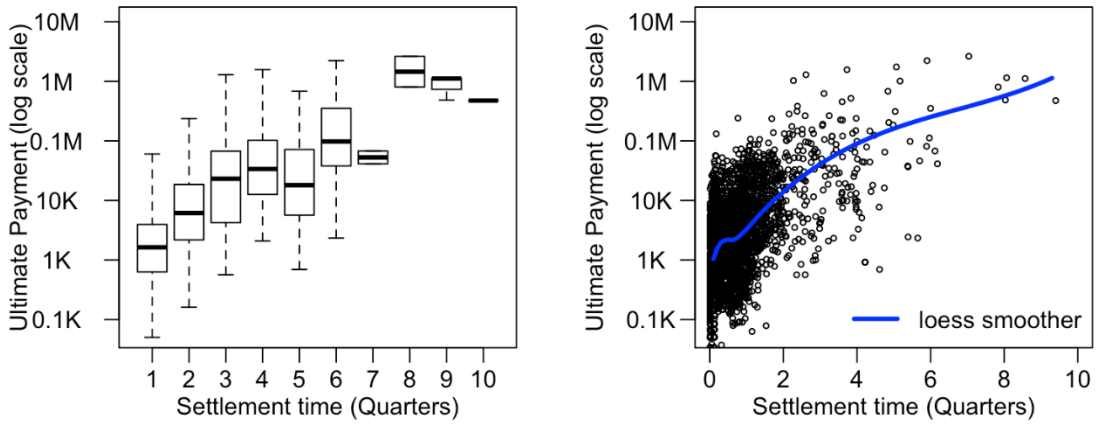


Figure 2: Distribution of ultimate payments by settlement time using data from a property insurance provider.

The left panel of Figure 3 shows the number of claims that occurred in each quarter from January 2006 to December 2009. Similar seasonal fluctuations are observed over each year, with the lowest occurrence in the winter season. The right panel of Figure 3 shows the distribution of the reporting delays in quarters. Approximately 75% of the claims are reported within the first quarter of the accident occurrence, but the distribution appears to be highly skewed to the right. Note that a reporting delay of zero corresponds with reporting on the day of occurrence. Also, with the valuation date assumed to be December 31, 2009, the low number of reported claims in the year 2009, as seen in the left panel of Figure 3, is due to Incurred But Not Reported (IBNR) claims.

Figure 3: Left Panel: Number of claims occurred in each quarter from January 2006 to December 2009. Right Panel: Reporting Delay.

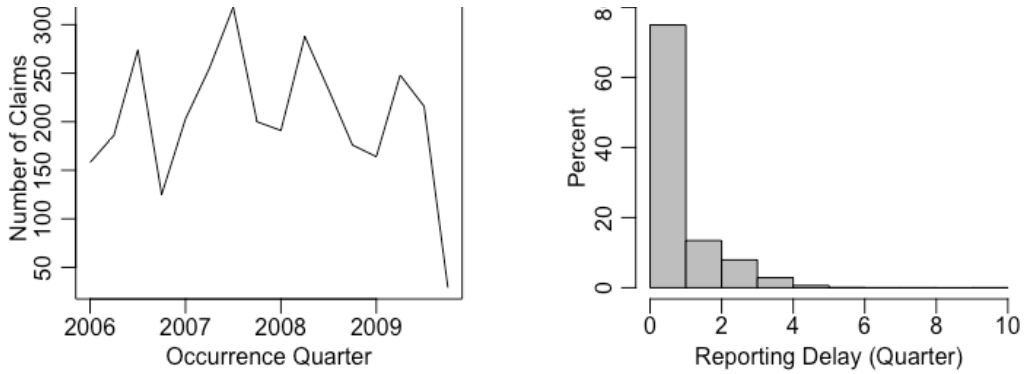


Table 2 summarizes the distribution of the continuous covariates and the two outcomes of interest, i.e., the ultimate losses and the settlement duration. The significant associations between the continuous covariates and the outcomes of interest, as shown by the Spearman correlations (ρ_S), suggest that they will be useful for predicting outstanding losses. Also, it is seen that the deductible, and initial estimate distributions are right-skewed. To handle the skewness, we will utilize logarithmic transformations of deductibles and initial estimates.

Table 2: Summary statistics for closed claims.

	Min.	Median	Mean	Max.	Ultimate Loss (ρ_S)	Settlement Time (ρ_S)
Ultimate Loss (Claim Severity)	25	2,203	14,133	2,633,822	-	0.49
Settlement Time (Days)	1	38	66	861	0.49	-
Deductible	500	1,000	12,297	100,000	-0.28	-0.21
Initial Estimate	30	2,500	9,545	1,000,000	0.93	0.51
Reporting Delay (Days)	0	28	66	864	-0.29	-0.55

3. RESULTS AND DISCUSSION

3.1 Estimation Results

The joint longitudinal-survival framework is applied to the micro-level reserving problem using the property data from the Wisconsin LGPIF. We begin by fitting a base model where for the longitudinal sub-model, we assume the observed cumulative payments follow a Log-Normal distribution, i.e. $y_{it} \sim$ Lognormal(η_{it}, σ^2), and fit a proportional hazard model with a Weibull baseline hazard for the survival sub-model. We also assume a random intercept longitudinal sub-model to account for unobserved claim-specific effects, where the random effects follow a normal distribution, $\mathcal{N}(0, \nu)$. See the Appendix for the estimation results for the base model.

3.1.1 Evaluation of survival sub-model fit

The correct specification of the survival sub-model is necessary to obtain accurate prediction results. In the base model, the survival baseline function assumes a Weibull model. In this section, we compare the overall survival sub-model fit using the Weibull baseline function to a more flexible survival sub-model with a spline baseline model. The spline baseline model was fitted with five equally-spaced internal knots in the quantiles of the observed event times.

To examine the overall fit of the survival model, we compare the Kaplan-Meier estimate of the Cox-Snell residuals from both survival sub-models to the function of the unit exponential distribution graphically (Rizopoulos 2012). Figure 4 plots the fit for the survival sub-model with a Weibull and spline sub-models assuming Log-Normal distribution for the longitudinal sub-model. The solid line is the Kaplan-Meier estimate of the survival function of the Cox-Snell residuals, and the dashed line is the survival function of the unit exponential distribution. It can be seen that both the Weibull and spline baseline fits the data very well. However, we choose the Weibull model because it is easier to interpret its components.

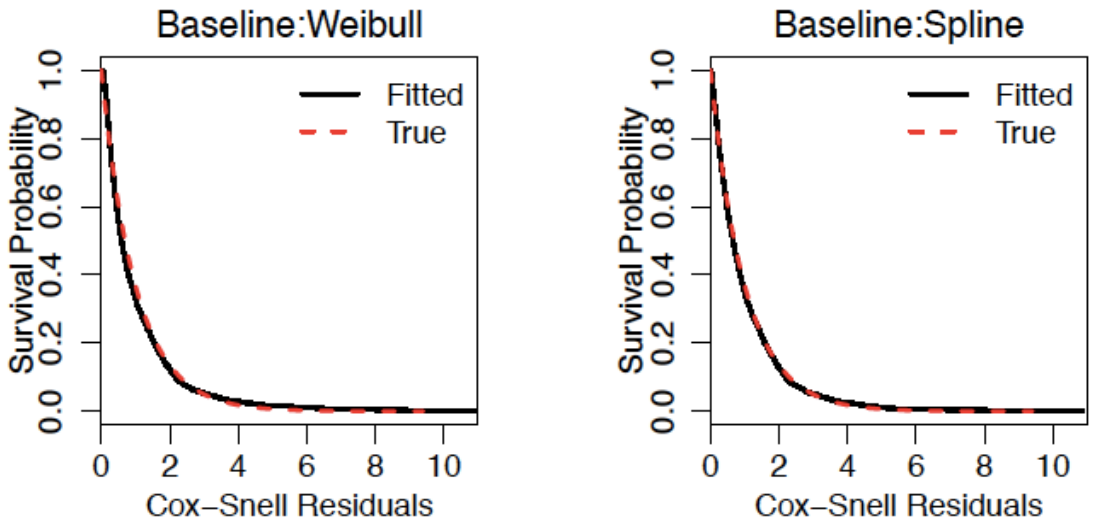


Figure 4: Evaluating the goodness of fit for survival sub-model

3.1.2 Evaluation of longitudinal sub-model fit

For the evaluation of the longitudinal sub-model fit in the base model, we first investigate whether the fit of the longitudinal sub-model can be improved by assuming a Gamma regression with dispersion parameter $1/\sigma$ and a log-link. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for the joint model with Log-Normal distribution and a linear trend are 74,117 and 74,488, respectively, and that of the Gamma model with a linear trend are 73,887 and 74,258, respectively. Therefore, a comparison using AIC and BIC suggests the Gamma model offers a better fit.

We also investigate whether the fit of the longitudinal sub-model can be improved by using a non-linear payment trend in the systematic component of the Gamma model. The left panel of Figure 5 plots the observed trend overlayed with the fitted linear trend, and the right panel plots the observed trend overlayed with the fitted non-linear trend using splines with an internal knot at payment time 5 (payment time in quarters). The AIC and BIC are 73,850 and 74,240 for the non-linear payment trend using splines. The AIC and BIC suggest a slightly better fit with the non-linear trend, and since the correct specification of the payment trend plays a critical role in the prediction of unpaid losses, we choose the model with the non-linear payment trend.

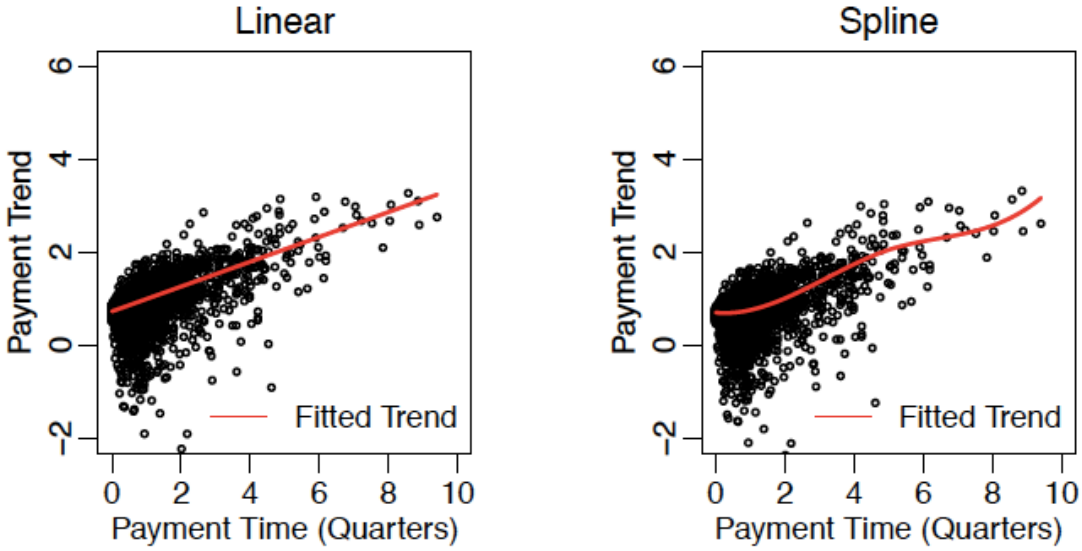


Figure 5: Evaluation of payment trend under the longitudinal sub-model.

The estimation results for the final fitted joint model, where y_{it} follows a Gamma distribution with a non-linear payment trend and log link in the longitudinal sub-model and a Weibull baseline in the survival sub-model, are given in Table 3. We present the parameter estimates and standard errors of the continuous covariates. For the categorical covariates, we present their likelihood ratio test statistic, the degrees of freedom, and p-value to test the importance of the categorical variable in each sub-model. For the survival sub-model, the association parameter $\alpha = -0.407$, and it measures the percentage reduction in hazard or risk of the settlement while expected payments increases by one percent. Note that α is highly significant at a 5% significance level and being negative in the hazard model means that the association between the settlement time and payment size is positive.

Table 3: Estimation results for final joint model: Assuming Gamma distribution with a log link and non-linear payment trend for longitudinal sub-model and a Weibull baseline survival sub-model.

Longitudinal sub-model			Survival sub-model		
Variable	Estimate	Std. Error	Variable	Estimate	Std. Error
(Intercept)	0.704	0.108	LnInitialEst	-0.069	0.060
B_1	-0.118	0.089	LnPolicyDed	0.010	0.013
B_2	1.876	0.168	ReportDelay	0.351	0.019
B_3	1.561	0.291			
B_4	2.465	0.343			
LnInitialEst	0.894	0.009			
LnPolicyDed	0.029	0.007			
ReportDelay	-0.012	0.014	α(association)	-0.407	0.067
Variance Components			Weibull Baseline Hazard		
shape (σ)	5.276		λ	50.159	
$\nu^{(1/2)}$	0.417		k	1.459	
Number of Payments	3,891		Number of Claims	3,264	
Categorical Variables					
Variable	LRT	df (p-value)	Variable	LRT	df (p-value)
CauseCode	93.550	9 (<0.0001)	CauseCode	93.430	9 (<0.0001)
Region	24.100	4 (0.0001)	Region	59.860	4 (<0.0001)
EntityType	9.720	5 (0.0837)	EntityType	64.840	5 (<0.0001)
LossQtr	4.120	3 (0.2486)	LossQtr	25.090	3 (0.0001)
LossYear	11.860	3 (0.0079)	LossYear	23.600	3 (<0.0001)

3.2 Out-of-Sample Validation

With our validation data that spans from January 1, 2010, to December 31, 2013, we can follow the actual future development trajectory of the RBNS claims after the valuation date and compare them to the predictions from the joint model and other reserving models. In this paper, we also consider two different estimation techniques known as the Independent model and the Two-Stage model. The Independent model sets $\alpha = 0$ in the survival sub-model and estimates the longitudinal and survival sub-model separately. The Two-Stage model's framework is similar to that of the joint model, but the Two-Stage model's parameters are estimated in two stages. The first stage estimates the longitudinal sub-model, and the second stage estimates the survival sub-model holding parameter estimates from the first stage fixed.

Further, we present results from the MPP model. The MPP framework allows for the modeling of the entire claim process, including occurrence, reporting, and development after reporting. Here, after reporting, the transaction occurrence times, the type of transaction, and the transaction's payment amount are considered to be the marks (features of interest). Different models are specified for each component of the MPP model. Detailed model specifications for the MPP are provided in the Appendix. To understand the impact of open claims on the prediction of unsettled losses, we also provide results from a model that employs a GLM for ultimate payments in the payment sub-model, and a survival sub-model that is modeled separately setting $\alpha = 0$. In addition, we provide results from the chain-ladder model.

3.2.1 Point prediction

To get the RBNS reserve estimate from the fitted joint model, we follow the prediction routine in section 2.3. Given that we are using the splines in the longitudinal sub-model, prediction for the ultimate losses is continued linearly for predicted settlement times greater than the largest observed payment times. The Gamma distribution is assumed for the longitudinal sub-model for the JM, Independent, the Two-Stage, and the GLM model. The prediction routine for the Independent, the Two-Stage model and the GLM model is similar to that of the joint model. For the MPP, we specify a discrete survival model with piece-wise constant hazard rates for the transaction occurrence, a logit model for the transaction type, and a Gamma regression for the incremental payments and follow the prediction routine for the RBNS reserve in Antonio and Plat (2014). The prediction routine simulates the next transaction's exact time, the transaction type (payment to a settlement, or intermediate payment), and the corresponding payment. For the chain-ladder, we employ a modified version of the Mack chain-ladder model (Mack, 1993), where claims in the run-off triangle are aggregated using reporting year and observation year instead of the occurrence year and development year. Then projections made from these development factors give us RBNS reserve estimates. Mack's model can be considered as a weighted linear regression. The analysis was performed in R following the ChainLadder package (Carrato et al., 2020), based on the run-off triangles provided in the Appendix.

Table 4 presents the reserve error, which is the expected RBNS reserve minus the actual unpaid losses and the error as a percentage of the actual unpaid losses for JM and other models. For all models

except the chain-ladder model, the estimated micro-level model is used to predict the RBNS estimate of each open claim and then aggregated to obtain the reserve estimate for the portfolio. In our out-of-sample data, we consider two claims as "unusual claims" because they each had payments totaling over a million dollars at the valuation date. These claims were caused by hail damage to buildings of a school in the year 2007 and a roof collapse of a building in the year 2008 with total payments at the valuation date of $5,398,051 and $1,802,742, respectively. Further, the ultimate amounts of these claims are $6,615,117 and $1,842,242, respectively. At the valuation date, the average total payment of open claims, including the unusual claims, is $60,668, and that of open claims without the unusual claims is $16,696. Naturally, the analyst will remove these unusual claims before any prediction exercise, but as a robustness check, we provide the prediction results with and without the unusual claims.

From the results, JM produced the least percentage reserve error at 0.41% without the unusual claims and a very competitive percentage error of -7.24% with the unusual claims. The results from the MPP are also competitive compared to JM. The performance of the Two-Stage and Independent model, in comparison to the JM, emphasizes that when the association between the payment process and settlement process and the endogenous nature of the payments process is ignored, it leads to inaccurate prediction of unpaid losses. Without any surprise, the GLM method, which only utilizes the ultimate payment for settled claims and ignores the payment-settlement association did not perform well. The results also show the chain-ladder method did not perform well in estimating the unpaid losses with the unusual claims but was very competitive without the unusual claims.

Table 4: RBNS reserve point prediction results for the validation sample.

	RBNS Estimate	Error %	RBNS Estimate	Error %
True Reserve	3,254,924		4,511,490	
JM Error	13,505	0.41	-326,721	-7.24
MPP Error	450,747	13.85	57,901	1.28
Two-Stage Error	74,910	2.30	451,714	10.01
Independent (JM with $\alpha = 0$) Error	725,206	22.28	-506,418	-11.23
GLM (Closed claims) Error	1,937,817	59.53	3,187,220	70.65
Chain-Ladder Error	262,232	8.06	1,261,332	27.96

The left panel of Figure 6 shows the comparison of the distribution of actual ultimate losses and predicted ultimate losses from JM over time. It can be seen that JM provides accurate predictions over time. Another advantage of the joint model is that we can use it to predict the time to settlement for open claims, which will be particularly useful in the run-off operation of an insurer. The right panel of Figure 6 provides a comparison of actual settlement times and predicted settlement times using the joint model. The joint model accurately predicts the settlement times with a Spearman correlation coefficient of 83%.

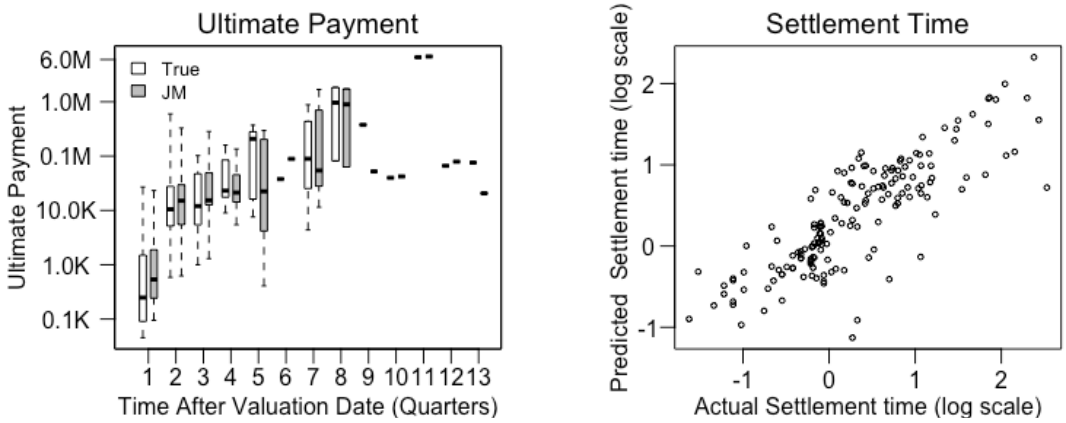


Figure 6: Left Panel: Distribution of the true and predicted ultimate payment over time (with unusual claims). Right Panel: Comparison of actual settlement times and predicted settlement times using JM (with unusual claims).

3.2.2 Predictive distribution

Here, we are not only interested in the expected value of prediction but also the variability in prediction. As a measure of reserve uncertainty, we provide the standard error, which is the standard deviation of the predictive distribution accounting for only parameter uncertainty and the root mean squared error of prediction (RMSEP), which is the standard deviation of the predictive distribution after accounting for both parameter and process uncertainty (England and Verrall 2002). All the prediction results in this section are based on 10,000 replications.

The predictive distribution of the expected outstanding payments is obtained by incorporating only the parameter uncertainty. For the joint model, we assume that the parameter estimates can be

approximated by a multivariate normal distribution with the maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$ as mean and covariance matrix $\widehat{Var}(\widehat{\boldsymbol{\theta}})$. The routine for the distribution of the expected outstanding payments is elaborated in Algorithm 1. The total RBNS liability for each replication is obtained by adding the RBNS prediction for all claims. The predictive distribution routine for the Independent, the Two-Stage, and the GLM models follow a similar procedure as the JM. For the MPP, we repeat the prediction routine to predict the RBNS reserve in (Antonio and Plat 2014), but we take the expected values for the payment amounts model. To obtain the predictive distribution of the mean for the CL, we employ the bootstrapping algorithm in England and Verrall (2002) and implemented in R following the ChainLadder package (Carrato et al., 2020).

For the predictive distribution of losses, in addition to the parameter uncertainty, we introduce process uncertainty to match the randomness of the development of losses. We generate the ultimate payments using the process distribution in the longitudinal sub-model at each replication. We repeat the steps in Algorithm 1 for introducing parameter uncertainty and introduce process uncertainty by simulating the ultimate payments from the process distribution of the longitudinal sub-model given each simulated set of parameters. The RBNS liability is then calculated for each simulated ultimate value, and the total RBNS liability for each replication is obtained by adding the RBNS prediction for all claims. Again, the predictive distribution for the Independent, Two-Stage, and GLM models follows a similar procedure as the JM. For the MPP, we introduce process uncertainty by simulating payments from a Gamma distribution. We account for the process uncertainty in the CL method by simulating payments in the future cells in the run-off triangle from the over-dispersed Poisson (England and Verrall 2002).

Table 5: RBNS reserve predictive distribution results for the validation sample.(without unusual claims).

	Estimate	SE	RMSEP
True Reserve	3,254,924		
JM	3,268,429	430,847	2,078,825
MPP	3,705,671	630,956	1,381,430
Two-Stage	3,329,834	336,761	2,084,256
Independent (JM with $\alpha = 0$)	3,980,130	343,737	2,261,560
GLM (Closed claims)	5,192,741	791,364	4,849,763
Chain-Ladder	3,517,156	987,054	1,334,597

Table 5 presents the reserve estimate, which was reproduced from Table 4, the standard error, and RMSEP for the out-of-sample data without the unusual claims. From the results, the joint model produced a significantly lower standard error than that of the chain-ladder. The higher predictive uncertainty of the chain-ladder is due to the loss of information from data aggregation. Also, building the reserving model with only information from closed claims, as seen with the GLM method, leads to a higher predictive uncertainty. The standard error from the MPP is higher compared to the joint model because the MPP model is composed of three sub-models compared to two sub-models from the joint model. By accounting for the payment-settlement association, i.e. an additional parameter, the joint model produced a slightly higher standard error compared to the Independent technique. Further, the Two-Stage technique produced a slightly lower standard error than the joint model, which could be the result of the Two-Stage technique ignoring the endogenous nature of the payments process hence understating the standard error. Figure 7 presents an illustration of the predictive distribution of the expected outstanding payments focusing on the out-of-sample data without unusual claims, and it can be seen that the JM provides both accurate mean prediction and low predictive uncertainty.

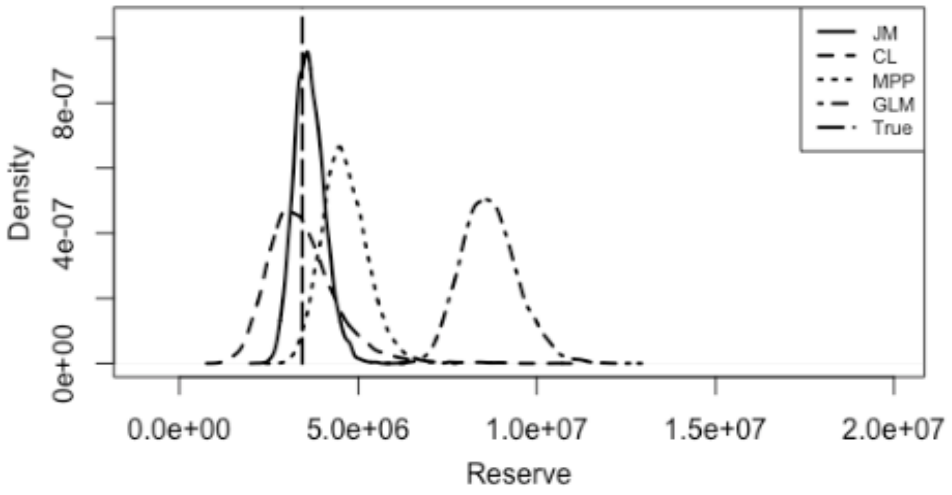


Figure 7: Predictive distribution of expected reserve estimates considering only parameter uncertainty (without unusual claims).

Figure 8 shows the predictive distribution after accounting for both parameter and the process uncertainty from the JM and other models focusing on the out-of-sample data without unusual claims. It can be seen that the joint model is associated with a higher process variance hence higher RMSEP compared to the MPP and the chain-ladder. The process variance from the joint model is higher because it is implemented using cumulative payments in the longitudinal sub-model. Also, as seen in Table 5, the process variance is higher for the micro-level models than the chain-ladder model.

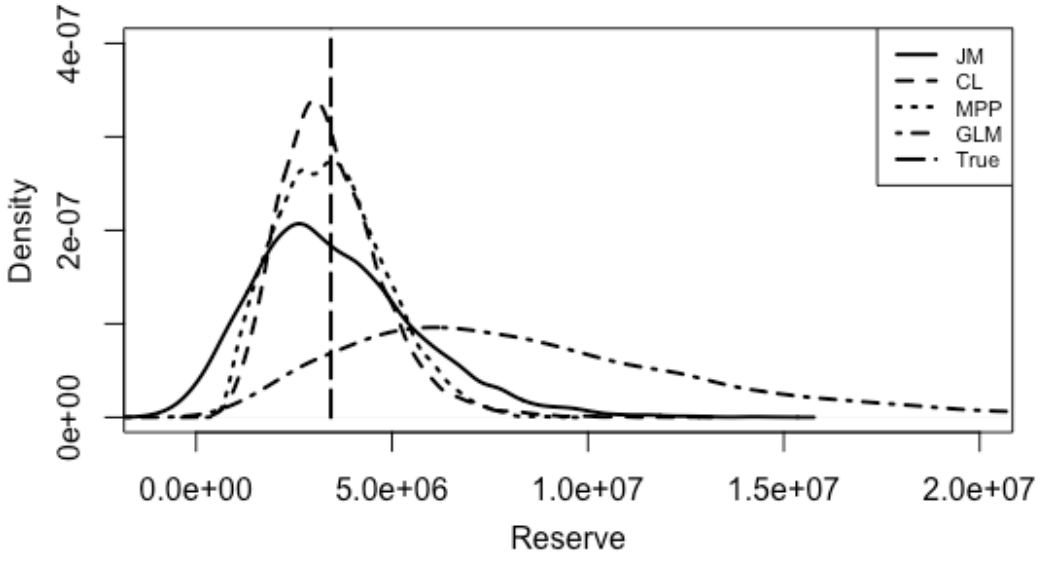


Figure 8: Predictive distributions (Parameter + Process Uncertainty) of the total RBNS reserve (without unusual claims).

3.2.3 Double cross-validation

In this subsection, we quantify the prediction error of different individual reserving methods using a novel out-of-sample validation method, which we call double-cross validation. The novelty of this approach comes from the longitudinal nature of the claims payment process, which makes it impossible to utilize traditional cross-validation techniques. Here, on the time dimension, we split the data by the valuation date. On the cross-section dimension, we split the data by the reporting date. Therefore, the training data contains payment from claims that have been reported by the valuation date. Then, the out-of-sample data comprises two parts. The first part contains payments made after the valuation date on claims reported before the valuation date; we call this the validation dataset. The second part contains payments from newly reported claims during the out-of-sample period, and we call this the test dataset. The routine for a K-fold double cross-validation technique is outlined in Algorithm 2. See Figure 9 for an example of 10-fold double cross-validation.

We obtain the prediction error percentages from both the validation and test datasets, and Table 6 provides the mean percentage error from the 10-fold double cross-validation for the JM and other micro-level models. Overall, the mean prediction percentage error for JM in the validation and test datasets are better than the results from other models, which highlights the robustness of the model.

Table 6: Mean percentage error from 10-fold double cross validation (without unusual claims).

	Validation data error %	Test data error %
JM	3.99	-17.74
MPP	47.72	-12.26
Two-Stage	26.59	-44.92
Independent (JM with $\alpha = 0$)	47.66	-25.25
GLM (Closed Claims)	83.92	87.32

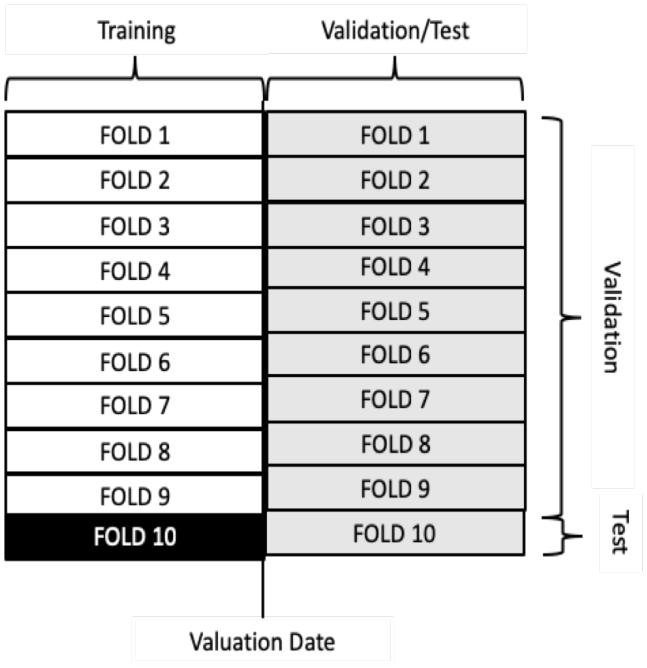


Figure 9: 10-fold double cross-validation technique.

3.3 Discussion on IBNR Reserving

This paper focuses on RBNS claims, so the practicing actuary would need to combine the joint model approach with a method for estimating IBNR claims to obtain the IBNR reserves. The general framework for estimating IBNR reserves can be broken down into two steps. The first step involves modeling the number of IBNR claims and their reporting delays with chain-ladder type strategies; for example, see Martínez-Miranda, Nielsen, and Verrall (2012) and Wüthrich (2018a). Further, Crevecoeur, Antonio, and Verbelen (2019) propose a granular approach to model the number of IBNR claims due to the heterogeneity of the reporting delay based on claim occurrence day and calendar day effects such as weekday and holiday effects. The second step involves modeling the development of the predicted IBNR claims with the proposed joint model fitted using the loss occurrence period and reporting delay as covariates.

Algorithm 1 Reserve predictive distribution for JM using Monte Carlo simulation.

Input: Valuation time c_i, observed data at valuation $(t_i, \delta_i, \boldsymbol{y}_i, \boldsymbol{w}_{ic_i}, \boldsymbol{x}_{ic_i}, \boldsymbol{z}_{ic_i})$, covariates at future time u $(\boldsymbol{w}_{iu}, \boldsymbol{x}_{iu}, \boldsymbol{z}_{iu})$, ML estimates $\hat{\boldsymbol{\theta}}$, $Var(\hat{\boldsymbol{\theta}})$, cumulative amount paid $Y_i(c_i)$, and empirical Bayes estimate $\hat{\boldsymbol{b}}_i$, number of draws K, and number of simulations L.

Output: $\left\{\hat{R}_i^{RBNS_l}(c_i), l = 1, ..., L\right\}$;

for $l = 1, ..., L$ **do**
- Generate $\boldsymbol{\theta}^l \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, Var(\hat{\boldsymbol{\theta}}))$;
- Generate $\boldsymbol{b}_i^l \sim f(\boldsymbol{b}_i | t_i, \delta_i, \boldsymbol{y}_i; \boldsymbol{\theta}^l)$;
- Calculate $S_i^l(c_i) = \exp\left(-\int_0^{c_i} h_0^l(s) \exp\{\boldsymbol{\gamma}'^l \boldsymbol{w}_{is} + \alpha^l \eta_{is}^l\} ds\right)$; where $\eta_{is}^l = \boldsymbol{x}_{is}' \boldsymbol{\beta}^l + \boldsymbol{z}_{is}' \boldsymbol{b}_i^l$ and $\{\alpha^l, \boldsymbol{\gamma}^l, \boldsymbol{\beta}^l\} \in \boldsymbol{\theta}^l$;
- **for** $k = 1, ..., K$ **do**
  - Generate $\pi_i(u|c_i) = U_k \sim \text{Uniform}(0, 1)$;
  - Calculate $u_{ik}^l = H_i^{-1}(-\log(U_k \times S_i^l(c_i)))$; where $H_i(u) = \int_0^u h_0^l(s) \exp\{\boldsymbol{\gamma}'^l \boldsymbol{w}_{is} + \alpha^l \eta_{is}^l\} ds$;
- **end for**
- **return** $\{u_{ik}^l; k = 1, ..., K\}$;
- Calculate $\hat{u}_i^l = K^{-1} \sum_{k=1}^K u_{ik}^l$;
- Generate $\hat{Y}_i^{ULT_l}(\hat{u}_i^l) = \exp(\hat{\eta}_{i\hat{u}_i^l})$; For parameter uncertainty.
- Generate $\hat{Y}_i^{ULT_l}(\hat{u}_i^l) \sim \text{Gamma}\left(\frac{\exp(\hat{\eta}_{i\hat{u}_i^l})}{\sigma^l}, \sigma^l\right)$; For parameter and process uncertainty. where $\hat{\eta}_{i\hat{u}_i^l} = \boldsymbol{x}_{i\hat{u}_i^l}' \boldsymbol{\beta}^l + \boldsymbol{z}_{i\hat{u}_i^l}' \boldsymbol{b}_i^l$ and $\{\boldsymbol{\beta}^l, \sigma^l\} \in \boldsymbol{\theta}^l$;
- Calculate $\hat{R}_i^{RBNS_l}(c_i) = \hat{Y}_i^{ULT_l}(\hat{u}_i) - Y_i(c_i)$;
- **return** $\left\{\hat{R}_i^{RBNS_l}(c_i), l = 1, ..., L\right\}$;

end for

Algorithm 2 Double cross-validation technique.

Input: Valuation time c and full dataset $\mathcal{D}_t = \{\mathcal{D}_t^T, \mathcal{D}_t^V\}$; where $\mathcal{D}_t^T$ is the training dataset, $\mathcal{D}_t^V$ is the validation dataset and t represent claim payment times;

Output: $\{\psi_k^V, k = 1, ..., K\}$ and $\{\psi_k^T, k = 1, ..., K\}$;

Split $\mathcal{D}_t$ into K groups, $\mathcal{D}_t = \{\mathcal{D}_t^k\}_{k=1}^K$;
for $k = 1, ..., K$ **do**
    Generate model building dataset $\mathcal{D}_t^S = \{\mathcal{D}_t^k\}_{k \neq k}$;
    Generate hold-out dataset $\mathcal{D}_t^H = \{\mathcal{D}_t^k\}_{k=k}$;
    Fit the model using training dataset $\mathcal{D}_t^{S_T} = \{\mathcal{D}_t^S\}_{t \leq c}$;
    Generate validation dataset $\mathcal{D}_t^{S_V} = \{\mathcal{D}_t^S\}_{t > c}$;
    Generate test dataset $\mathcal{D}_t^{H_T} = \{\mathcal{D}_t^H\}_{t > c}$;
    Calculate prediction error percentage ψ_k^V using $\mathcal{D}_t^{S_V}$;
    Calculate prediction error percentage ψ_k^T using $\mathcal{D}_t^{H_T}$;
    return $\{\psi_k^V, k = 1, ..., K\}$ and $\{\psi_k^T, k = 1, ..., K\}$;
end for

4. CONCLUSIONS

Actuarial analysts commonly encounter situations where the time of settlement is positively associated with the size of the claim. The payment-settlement association means settlement times will be impacted by paid losses, which affects the reserve prediction of open claims. Therefore, ignoring the payment-settlement association could lead to inaccurate predictions of outstanding payments.

In this paper, to incorporate the correlation between the payment and the settlement processes, the joint longitudinal-survival model (JM) framework was applied to the reserving problem using data from a property insurance provider. The prediction results from the joint model are compared to existing reserving models, and the results show that accounting for the payment-settlement association leads to better prediction accuracy and lower reserve uncertainty compared to models that ignore it.

We also introduced a novel cross-validation technique named double cross-validation as a result of the time dimension involved with claim development, which makes the use of the traditional cross-validation techniques impossible. The double cross-validation technique provides two datasets

(validation and test datasets) for the evaluation of the robustness of the models. The validation dataset contains outstanding payments for claims reported by the valuation date. The test data contains payments from newly reported claims during the out-of-sample period. Again, the joint model displayed superior prediction accuracy using both datasets compared to models that ignore the payment-settlement association, which highlights the robustness of the model.

The literature on joint models primarily focuses on the estimation aspect of inference. In this paper, we enrich the literature by applying joint models to the prediction of RBNS reserves. The application of this sophisticated model in a new setting will be of interest to statisticians and also to actuaries for accurate reserve prediction.

Acknowledgment

The authors are grateful to our reviewers Kelly Moore, Julie Lederer, and Denise Ambrogio, for their insightful comments and valuable suggestions to improve this paper.

Supplementary Material

Sample R code for estimation and prediction based on simulated data.

APPENDIX A: ESTIMATION RESULTS FOR BASE JOINT MODEL

Estimation results for the fitted joint model where y_{it} follows a Log-Normal distribution, with a Weibull baseline survival sub-model is given in Table 7.

Table 7: Estimation results for base joint model: Assuming Log-Normal distribution with a linear payment trend for the longitudinal sub-model and a Weibull baseline survival

Longitudinal sub-model			Survival sub-model		
Variable	Estimate	Std. Error	Variable	Estimate	Std. Error
(Intercept)	0.808	0.124	LnInitialEst	-0.349	0.041
TimeToPayment	0.288	0.012	LnPolicyDed	-0.003	0.012
LnInitialEst	0.853	0.010	ReportDelay	0.348	0.018
LnPolicyDed	0.028	0.008			
ReportDelay	0.029	0.014	α(association)	-0.086	0.044
Variance Components			Weibull Baseline Hazard		
σ	0.425		λ	35.449	
$v^{(1/2)}$	0.478		k	1.397	
Number of Payments	3,891		Number of Claims	3,264	
Categorical Variables					
Variable	LRT	df (p-value)	Variable	LRT	df (p-value)
CauseCode	90.500	9 (<0.0001)	CauseCode	95.050	9 (<0.0001)
Region	35.370	4 (<0.0001)	Region	51.500	4 (<0.0001)
EntityType	3.920	5 (0.5610)	EntityType	67.330	5 (<0.0001)
LossQtr	2.490	3 (0.4772)	LossQtr	27.420	3 (0.0001)
LossYear	17.05	3 (0.0007)	LossYear	19.95	3 (<0.0002)

APPENDIX B: DETAILS FOR MARKED POISSON PROCESS FOR RBNS

Under the Marked Poisson Process framework, the likelihood for the full development process of claim i is given by (Jin, 2014):

$$L = f_T \times f_{U|T} \times f_{X|T,U} = f_T \times f_{U|T} \times f_{V|T,U} \times f_{E|T,U,V} \times f_{p|T,U,V,E} \qquad \text{(A.1)}$$

where T and U represent the claim occurrence times and reporting delay respectively. However, with the focus on RBNS reserve prediction, we are interested in the claim development process X given by:

$$f_{X|T,U} = f_{V|T,U} \times f_{E|T,U,V} \times f_{p|T,U,V,E} \tag{A.2}$$

Where V denotes the transaction occurrence times, E denotes the type of transaction, and P denotes the payment amount of the transaction. The transaction occurrence times V are modeled by a discrete survival model with piecewise constant hazard rates. Following (Jin, 2014) and (Antonio and Plat, 2014), the first transactions are modeled with a hazard rate $g(t)$, and the later transactions are modeled with a different hazard rate $h(t)$. Let $[0, a_K]$ and $[0, b_L]$ be the interval for first and later transactions. Then we have:

$$g(t) = \sum_{k=1}^{K} g_k 1\{a_{k-1} < t \leq a_k\} \tag{A.3}$$

$$h(t) = \sum_{l=1}^{L} h_l 1\{b_{l-1} < t \leq b_l\} \tag{A.4}$$

With cumulative hazard rates given by:

$$G(t) = \int_0^t g(s)\, ds \tag{A.5}$$

$$H(t) = \int_0^t h(s)\, ds \tag{A.6}$$

Then the cumulative density functions of transaction occurrence times are given by:

$$\Pr(V_1 \leq t) = 1 - \exp\left(-\mathrm{G(t)}\right) \tag{A.7}$$

$$\Pr\left(V_J \leq t\right) = 1 - \exp\left(-\mathrm{H(t)}\right), \mathrm{j} > 1 \tag{A.8}$$

Let $a_K = N_1$ be regarded as the maximum waiting time to the first transaction, and $b_L = N_2$ is regarded as the maximum settlement delay. Then under these additional assumptions, the probability that the first transaction occurs at time $k, k = 1,2,\ldots, N_1$ is

$$\Pr(V_1 = k | V_1 \leq N_1) = \frac{\exp\left(-\mathrm{G}(k-1)\right) - \exp\left(-\mathrm{G}(k)\right)}{1 - \exp\left(-\mathrm{G}(N_1)\right)} \tag{A.9}$$

And given the occurrence time of the first transaction, $V_{j-1} = s$, the probability that transaction j occurs at time $k, k = s+1, s+2, \ldots, N_2$ is

$$\Pr\left(V_j = k | V_{j-1} = s, V_j < N_2\right) = \frac{\exp\left(-\mathrm{H}(k-1)\right) - \exp\left(-\mathrm{H}(k)\right)}{\exp\left(-\mathrm{H}(s)\right) - \exp\left(-\mathrm{H}(N_2)\right)} \qquad \text{(A.10)}$$

For the Wisconsin LGPIF training dataset, the maximum waiting time for the first transaction is 17 months, and the maximum settlement delay is 27 months. We assume that there is at most one transaction in each month, and the transactions can only occur at the end of a month. As noted in (Jin, 2014), this discrete setup is consistent with the fact that many insurers aggregate transactions on a monthly basis by the end of each month. Therefore, the piecewise-constant hazard rates is defined to have jumps every month, i.e. $a_1 = 0, a_2 = 1, \cdots, a_{17} = 17$ and $b_1 = 0, b_2 = 1, \cdots, b_{27} = 27$.

Furthermore, for the type of transactions E, we consider two types for claim i at time v; a payment transaction that leads to settlement ($e_{iv} = 1$) and an intermediate payment transaction ($e_{iv} = 0$). With an intermediate transaction, the claim development process continues. Given a transaction at time v, the transaction type is determined by a logit model that accommodates heterogeneity by incorporating random effects a_i.The probabilities also depend on the time of the transaction and covariates x_{iv} given by:

$$\Pr(e_{iv} = 1 | \boldsymbol{a}_i) = \pi(\mathbf{x'}_{iv}\boldsymbol{\beta} + \boldsymbol{a}_i) = \frac{1}{1 + \exp\left(-(\mathbf{x'}_{iv}\boldsymbol{\beta} + \boldsymbol{a}_i)\right)} \qquad \text{(A.11)}$$

To model the incremental payments P, we specify a Generalized Linear Mixed-Effects Model.

APPENDIX C: LOSS TRIANGLE FROM LGPIF DATA FOR RBNS RESERVE PREDICTION

Table 8 summarizes the cumulative amounts paid arising out of building and contents coverage from the LGPIF data that occurred and were reported between January 1, 2006, and December 31, 2009, organized by reporting quarters and observation quarters. Then projections made from the development factors give us RBNS reserve estimates. Table 9 provides the loss triangle without unusual claims.

Table 8: Observed historical cumulative losses from the LGPIF organized by reporting quarters and observation quarters (with unusual claims).

Reporting Quarter	Observation Quarter															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2006 Q1	1,264,447	1,682,188	1,682,188	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561
2006 Q2	2,009,311	3,370,260	4,761,115	5,174,541	5,873,088	6,559,705	6,818,143	6,818,143	7,003,825	7,003,825	7,003,825	7,003,825	7,003,825	7,003,825	7,003,825	
2006 Q3	1,402,769	2,406,855	2,482,578	2,523,994	2,615,403	2,670,567	2,670,567	2,670,567	2,670,567	2,670,567	2,670,567	2,670,567	2,670,567	2,670,567		
2006 Q4	806,480	1,207,453	1,239,388	1,572,304	4,227,033	4,227,033	4,227,033	4,481,423	4,481,423	4,481,423	4,481,423	4,481,423	4,481,423			
2007 Q1	1,135,006	1,788,184	1,917,445	2,123,798	2,590,466	2,654,564	2,654,564	2,720,276	2,720,276	2,720,276	2,720,276	2,720,276				
2007 Q2	705,146	1,350,025	1,841,137	2,010,819	6,004,549	6,232,118	6,619,945	6,696,295	8,035,846	8,035,846	8,035,846					
2007 Q3	1,100,841	1,854,040	2,193,211	2,234,338	2,612,670	2,637,070	2,637,070	2,637,070	2,671,639	2,744,121						
2007 Q4	1,893,020	2,951,579	3,380,631	3,725,657	3,756,556	3,758,990	3,765,436	3,953,195	3,976,862							
2008 Q1	1,488,889	2,944,819	3,234,739	3,675,551	3,696,906	3,812,083	4,787,062	5,041,970								
2008 Q2	1,516,620	3,537,807	4,642,772	5,173,249	5,209,948	5,249,361	5,249,361									
2008 Q3	1,375,480	2,860,584	3,612,216	3,762,930	3,792,355	3,796,699										
2008 Q4	1,046,145	1,537,226	1,805,164	1,956,216	1,956,216											
2009 Q1	1,277,018	1,779,555	1,971,026	2,058,902												
2009 Q2	816,927	1,625,055	1,810,738													
2009 Q3	1,396,415	1,816,822														
2009 Q4	450,633															

Table 9: Observed historical cumulative losses from the LGPIF organized by reporting quarters and observation quarters (without unusual claims).

Reporting Quarter	Observation Quarter															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2006 Q1	1,264,447	1,682,188	1,682,188	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561	1,722,561
2006 Q2	2,009,311	3,370,260	4,761,115	5,174,541	5,873,088	6,559,705	6,818,143	6,818,143	7,003,825	7,003,825	7,003,825	7,003,825	7,003,825	7,003,825	7,003,825	
2006 Q3	1,402,769	2,406,855	2,482,578	2,523,994	2,615,402	2,670,566	2,670,566	2,670,566	2,670,566	2,670,566	2,670,566	2,670,566	2,670,566	2,670,566		
2006 Q4	806,480	1,207,453	1,239,388	1,572,304	4,227,033	4,227,033	4,227,033	4,481,423	4,481,423	4,481,423	4,481,423	4,481,423	4,481,423			
2007 Q1	1,135,006	1,788,184	1,917,445	2,123,798	2,590,466	2,654,564	2,654,564	2,720,276	2,720,276	2,720,276	2,720,276	2,720,276				
2007 Q2	601,388	1,246,267	1,570,116	1,739,799	2,107,930	2,335,498	2,335,498	2,411,848	2,637,794	2,637,794	2,637,794					
2007 Q3	1,100,841	1,854,040	2,193,211	2,234,338	2,612,670	2,637,070	2,637,070	2,637,070	2,671,639	2,744,121						
2007 Q4	1,893,020	2,951,579	3,380,631	3,725,657	3,756,556	3,758,989	3,765,435	3,953,195	3,976,862							
2008 Q1	1,488,889	2,363,739	2,653,659	3,094,471	3,115,826	3,231,003	3,239,227	3,239,227								
2008 Q2	1,516,620	3,537,807	4,642,772	5,173,249	5,209,948	5,249,361	5,249,361									
2008 Q3	1,375,480	2,860,584	3,612,216	3,762,930	3,792,355	3,796,699										
2008 Q4	1,046,145	1,537,226	1,805,164	1,956,216	1,956,216											
2009 Q1	1,277,018	1,779,555	1,971,026	2,058,902												
2009 Q2	816,927	1,625,055	1,810,738													
2009 Q3	1,396,415	1,816,822														
2009 Q4	450,633															

5. REFERENCES

[1] Antonio, K., and J. Beirlant. 2007. "Actuarial Statistics with Generalized Linear Mixed Models." *Insurance : Mathematics and Economics* 40(1): 58–76.

[2] Antonio, K., and R Plat. 2014. "Micro-Level Stochastic Loss Reserving for General Insurance." *Scandinavian Actuarial Journal* 7: 649–69.

[3] Arjas, E. 1989. "The Claims Reserving Problem in Non-Life Insurance: Some Structural Ideas." *ASTIN Bulletin* 19(2): 139–52.

[4] Carrato, Alessandro, Fabio Concina, Markus Gesmann, Daniel Murphy, Mario Wüthrich, and Wayne Zhang. 2020. "Claims Reserving with R: ChainLadder-0.2.11 Package Vignette." *Https://Cran.r-Project.org/Web/Packages/ChainLadder/Vignettes/ChainLadder.pdf*.

[5] Crevecoeur, Jonas, Katrien Antonio, and Roel Verbelen. 2019. "Modeling the Number of Hidden Events Subject to Observation Delay." *European Journal of Operational Research* 277 (3): 930–44.

[6] Diggle, P., and M. Kenward. 1994. "Informative Drop-Out in Longitudinal Data Analysis." *Journal of the Royal Statistical Society, Series C, Applied Statistics* 43: 49–73.

[7] Elashoff, Robert M., Gang Li, and Ning Li. 2016. *Joint Modeling of Longitudinal and Time-to-Event Data*. Chapman; Hall/CRC.
[8] England, P. D., and R. J. Verrall. 2002. "Stochastic Claims Reserving in General Insurance." *British Actuarial Journal* 8(3): 443–518.
[9] Frees, Edward. 2004. *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. New York: Cambridge University Press.
[10] Friedland, Jacqueline. 2010. *Estimating Unpaid Claims Using Basic Techniques*. Casualty Actuarial Society.
[11] Grace, Martin F., and J. Tyler Leverty. 2012. "Property-Liability Insurer Reserve Error: Motive, Manipulation, or Mistake." *The Journal of Risk and Insurance* 79(2): 351–80.
[12] Ibrahim, Joseph G., Haitao Chu, and Liddy M. Chen. 2010. "Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data." *Journal of Clinical Oncology* 28(16): 2796–2801.
[13] Jewell, William S. 1989. "Predicting Ibnyr Events and Delays, Part I Continuous Time." *ASTIN Bulletin* 19(1): 25–55.
[14] Jin, X. 2014. "Micro-Level Stochastic Loss Reserving Models for Insurance." *The University of Wisconsin-Madison, ProQuest Dissertations Publishing*.
[15] Little, Roderick. 2008. "Selection and Pattern-Mixture Models." In *Longitudinal Data Analysis*, edited by Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs, 409–32. Boca Raton: CRC Press.
[16] Liu, Lei. 2009. "Joint Modeling Longitudinal Semi-Continuous Data and Survival, with Application to Longitudinal Medical Cost Data." *Statistics in Medicine* 28(6): 972–86.
[17] Mack, T. 1993. "Distribution Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." *ASTIN Bulletin* 23 (2): 213–25.
[18] Martínez-Miranda, M. D., J. P. Nielsen, and R. Verrall. 2012. "Double Chain Ladder." *ASTIN Bulletin* 42(1): 59–76.
[19] Molenberghs, G., and G. Verbeke. 2006. *Models for Discrete Longitudinal Data in: Springer Series in Statistics*. Springer, New York.
[20] Norberg, R. 1993. "Prediction of Outstanding Liabilities in Non-Life Insurance." *ASTIN Bulletin* 23(1): 95–115.
[21] ———. 1999. "Prediction of Outstanding Liabilities Ii. Model Variations and Extensions." *ASTIN Bulletin* 29(1): 5–25.
[22] Rizopoulos, D. 2010. "JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data." *Journal of Statistical Software (Online)* 35(9): 1–33.
[23] ———. 2012. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman; Hall/CRC.
[24] ———. 2016. "The R Package Jmbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using Mcmc." *Journal of Statistical Software* 72(1): 1–46.
[25] Taylor, G. C, and M. Campbell. 2002. "Statistical Case Estimation." *In Research Paper Number 104, the University of Melbourne, Australia*.
[26] Taylor, G. C, and G. McGuire. 2004. "Loss Reserving GLMs: A Case Study." *In Annual Meeting for the Casualty Actuarial Society, Spring 2004*.
[27] Taylor, G. C., G. McGuire, and J. Sullivan. 2008. "Individual Claim Loss Reserving Conditioned by Case Estimates." *Annals of Actuarial Science* 3(1-2): 215–56.
[28] Tsiatis, A., and M Davidian. 2004. "Joint Modeling of Longitudinal and Time-to-Event Data: An Overview." *Statistica Sinica* 14(3): 809–34.
[29] Verbeke, Geert, Geert Molenberghs, and Dimitris Rizopoulos. 2010. "Random Effects Models for Longitudinal Data." In *Longitudinal Research with Latent Variables*, edited by Kees van Montfort, Johan H. L. Oud, and Albert Satorra, 37–96. Berlin: Springer, Berlin, Heidelberg.
[30] Wüthrich, M. V. 2018a. "Machine Learning in Individual Claims Reserving." *Scandinavian Actuarial Journal* 2018(6): 465–80.
[31] ———. 2018b. "Neural Networks Applied to Chain-Ladder Reserving." *European Actuarial Journal* 8(2): 407–36.
[32] Wüthrich, M. V., and M. Merz. 2008. *Stochastic Claims Reserving Methods in Insurance*. John Wiley & Sons.
[33] Yu, M., N. Law, J. Taylor, and H. Sandler. 2004. "Joint Longitudinal-Survival-Curve Models and Their Application to Prostate Cancer." *Statistica Sinica* 14(3): 835–62.

Abbreviations and notations

AIC, Akaike information criterion	JM, joint model
BIC. Bayesian information criterion	LGPIF, local government property insurance fund
CL, chain ladder	MPP, marked Poisson process
GLMM, generalized linear mixed-effects model	RBNS, reported but not settled
GLM, generalized linear model	RMSEP, root mean squared error prediction
IBNR, Incurred but not reported	

AUTHORS

A. Nii-Armah Okine (Corresponding author)
Department of Mathematical Sciences
Appalachian State University
121 Bodenheimer Dr. Boone, NC 28608
Email: okinean@appstate.edu
A. Nii-Armah Okine is an assistant Professor at Appalachian State University. His research interests includes micro-level reserving, joint longitudinal-survival modeling, dependence modelling, micro insurance and machine learning.

Edward W. Frees
Department of Risk and Insurance
Wisconsin School of Business, University of Wisconsin-Madison
975 University Avenue, Madison WI 53706, USA
Email: jfrees@bus.wisc.edu
Edward W. (Jed) Frees is an emeritus professor, formerly the Hickman-Larson Chair of Actuarial Science at the University of Wisconsin-Madison. He is a Fellow of both the Society of Actuaries and the American Statistical Association.

Peng Shi
Department of Risk and Insurance
Wisconsin School of Business, University of Wisconsin-Madison
975 University Avenue, Madison WI 53706, USA
Email: pshi@bus.wisc.edu
Peng Shi is an associate professor in the Risk and Insurance Department at the Wisconsin School of Business. He is also the Charles & Laura Albright Professor in Business and Finance. Professor Shi is an Associate of the Casualty Actuarial Society (ACAS) and a Fellow of the Society of Actuaries (FSA).

Cash Flow and Unpaid Claim Runoff Estimates Using Mack and Merz-Wüthrich Models

Mark R. Shapland, FCAS, FSA, MAAA

Abstract

Motivation. For both Solvency II and IFRS 17 the actuary could use unpaid claim variability estimates for cash flows and the runoff of unpaid claims in addition to the more widely used accident year view of the unpaid claims. Under Solvency II, the concept of the one-year time horizon adds a new dimension to the estimates for unpaid claim distributions. Thus, the focus of this paper is to expand the accident year formulas developed by Mack, and modified by Merz & Wüthrich to address the 1-year time horizon, to include both runoff and cash flow formulas.

Method. This paper is based on a review of the foundational Mack and Merz-Wüthrich formulas and their decomposition into process variance and parameter uncertainty, per future diagonal. The decompositions are then used to show how modifications to the accident year formulas can be used to calculate the standard deviations for cash flow and unpaid claim runoff estimates. In addition, an alternative view of the covariance adjustment is developed to aide comparisons with other models.

Results. Merz & Wüthrich have previously addressed the runoff of the 1-year time horizon and England, Verrall & Wüthrich have proposed using this runoff of the Merz-Wüthrich formulas for risk margin estimates using the cost of capital method. In this paper, we will discuss how the original formulas can be modified to better fit the Solvency II time horizon concept.

Conclusions. While the Merz-Wüthrich formulas (and by extension the England, Verrall & Wüthrich formulas) are an elegant bridge between the 1-year time horizon and the ultimate time horizon developed by Mack, the alternative formulation for the runoff of the 1-year time horizon provides a better fit for the Solvency II environment.

Availability. In lieu of technical appendices, companion Excel workbooks are included that illustrate the calculations described in this paper. References to the Excel file "Mack & Merz-Wüthrich Runoff.xlsm" will be made in sections 3 to 5 of this paper in order to facilitate the comprehension of the formulas. The companion materials are summarized in the Supplementary Materials section and are available at https://www.casact.org/pubs/forum/20sumforum/.

Keywords. Reserve variability, chain ladder, prediction error, mean square error of prediction, cost of capital, risk margin, risk adjustment, Solvency II, IRFS 17, value at risk, tail value at risk, one-year time horizon.

1. INTRODUCTION

While there is now a large and growing volume of models that can be used for reserve variability estimates, one of the foundational models was introduced by Mack [5] in 1993. Because it is a closed form solution, which can be easily adapted in an Excel function or reserving software, it has gained widespread use.

The Solvency II regulatory regime in Europe introduced the concept of the 1-year time horizon and Merz & Wüthrich [7] took up the challenge of modifying the Mack formulas to directly estimate the reserve variability of the claim development result for a 1-year time horizon. Like the Mack

formulas, the Merz-Wüthrich models have gained widespread use for Solvency II.

For both the Mack and Merz-Wüthrich formulas, the papers only focus on an accident year view of the claim development, which is natural as this is the primary configuration for reserving data. Fortunately, as all the "parts" are included in the formulas it is a natural extension of these models to work out the calendar year formulas for calculating the variance of the cash flows and unpaid claim runoff. In addition to typical uses, examining both of these in more detail helps to decompose the 1-year time horizon, which includes both parameter and process variance for the next calendar year, to estimate possible outcomes, and only parameter variance for the remaining future calendar years, to estimate reserves contingent on the possible outcomes in the next calendar year (i.e., over a 1-year time horizon).

1.1 Research Context

The model developed by Mack is widely used by actuaries and the Mack papers are well supported with derivations and proofs. The models developed by Merz & Wüthrich [7, 8] are similarly well supported with derivations and proofs. Thus, this paper will focus on a high-level discussion of the modeling frameworks and will not reproduce the derivations and proofs as the reader can find these in the original papers.

This family of models is a distribution free method for calculating the variance of the chain ladder (CL) method by combining the process variance and parameter variance components of the mean squared error of prediction (MSEP):

$$\text{MSEP} \approx \sqrt{\textit{process variance} + \textit{parameter variance}} \tag{1.1}$$

The use of colors for the process variance and parameter variance components of the formulas is useful for clarifying the calculations and tracing the components through the various formulas.

1.2 Objective

The calendar year view of the standard formulas is an important addition to the actuarial literature to support cash flow and unpaid claim runoff calculations for different regulatory and financial reporting regimes, such as Solvency II and IRFS 17, in addition to enterprise risk management uses. A recent paper by England, Verrall & Wüthrich [3] examines how the ultimate and time horizon views of the Mack and Merz-Wüthrich models, respectively, are connected.

In England, Verrall & Wüthrich [3], the authors propose that the runoff of the time horizons using the Merz-Wüthrich formulas is ideal for uses such as the runoff of the capital for the cost of capital method of calculating a risk margin. We will examine this proposed use of the Merz-Wüthrich model

and suggest an alternative approach.

1.3 Outline

The remainder of the paper proceeds as follows. Section 2 will provide an overview of the notation used. In Section 3, the Mack model is described and then additional formulas for calculating the variance of the cash flows and runoff of the unpaid claims are specified. Next, Section 4 will focus on the Merz-Wüthrich models, which include the runoff of the time horizon beyond year one. Like the Mack model discussion, the cash flow formulas will be specified. Then, in Section 5 alternative formulas for the runoff of the time horizon beyond year one will be proposed. Finally, Section 6 will discuss conclusions based on results applying the formulas to a real dataset.

2. NOTATION

The notation in this paper is from the CAS Working Party on Quantifying Variability in Reserve Estimates Summary Report [1] since it is intended to serve as a basis for further research. Many models visualize loss data as a two-dimensional array, (w, d), with accident period or policy period w and development age d (think w = "when" and d = "delay").[1] For this discussion, it is assumed that the loss information available is an "upper triangular" subset for rows $w = 1,2, \dots, n$ and for development ages $d = 1,2, \dots, n$. The "diagonal" for which $w + d - 1$ equals the constant, k, represents the loss information for each accident period w as of accounting period k.[2]

For purposes of including tail factors, the development beyond the observed data for periods $d = n + 1, n + 2, \dots, u$, where u is the ultimate time period for which any claim activity occurs – i.e., u is the period in which all claims are final and paid in full – must also be considered.

The paper uses the following notation for certain important loss statistics:

$c(w, d)$: cumulative loss from accident year w as of age d.[3]

$q(w, d)$: incremental loss for accident year w from $d - 1$ to d.

$c(w, n) = U(w)$: total loss from accident year w when claims are at ultimate values at time n, or

---

[1] For a more complete explanation of this two-dimensional view of the loss information, see the *Foundations of Casualty Actuarial Science* [4], Chapter 5, particularly pages 210-226.

[2] Some authors define $d = 0,1, \dots, n - 1$ which intuitively allows $k = w$ along the diagonals, but in this case the triangle size is $n \times n - 1$ which is not intuitive. With $d = 1,2, \dots, n$ as defined in this paper, the triangle size $n \times n$ is intuitive, while $k = w + d - 1$ along the diagonals is less intuitive but still works. A way to think about this which helps tie everything together is to assume the w variables are the beginning of the accident periods and the d variables are at the end of the development periods. Thus, if years are used then cell $c(n, 1)$ represents accident year n evaluated at $12/31/n$, or essentially $1/1/n + 1$.

[3] The use of accident year is for ease of discussion. All of the discussion and formulas that follow could also apply to underwriting year, policy year, report year, etc. Similarly, year could also be half-year, quarter or month.

with tail factors the equivalent notation is $c(w,u) = U(w)$.[4]

$R(w)$: future development after age $n-w+1$ for accident year w, i.e., $= U(w) - c(w, n-w+1)$.

$F(d)$: factor applied to $c(w,d)$ to estimate $c(w,d+1)$.

$e(w,d)$: a random fluctuation, or error, which occurs at the w, d cell.

$E(x)$: the expectation of the random variable x.

$Var(x)$: the variance of the random variable x. Or, alternatively σ_x^2.

σ_x: the standard deviation of the random variable x.

$\hat{x}$: an estimate of the parameter x.

N: the total number of accident years.[5]

The notation does not distinguish paid vs. incurred, but if this is necessary, capitalized subscripts P and I could be used. The cumulative known data, D, used in the formulas in this paper can be illustrated as follows:

		d					
		1	2	3	…	n-1	n
w	1	c(1,1)	c(1,2)	c(1,3)	…	c(1,n-1)	c(1,n)
	2	c(2,1)	c(2,2)	c(2,3)	…	c(2,n-1)	
	3	c(3,1)	c(3,2)	c(3,3)			
	…	…	…				
	N-1	c(N-1,1)	c(N-1,2)				
	N	c(N,1)					

To better illustrate the perspectives related to time between the Mack and Merz-Wüthrich models, the following notation and terms are used:

t: "at time" is equivalent to the valuation date used for financial accounting, with $t = 0$ representing the current valuation date and $t = 1,2,3, \ldots$ representing future valuation dates.

T: "time horizon" is the period for which the full distribution, including both process and parameter variance, is estimated.

---

[4] This would imply that claims reach their ultimate value without any tail factor. This is generalized by changing n to $u = n + t$, where t is the number of periods in the tail.

[5] In a typical triangle the number of accident years, N, is the same as the number of development periods, n, but the number of development periods can be longer. Even when they are the same using N vs. n helps visualize the calculations in the formulas.

T': "time window" is the period between the valuation date and the time when the process variance and only a portion of the parameter variance is estimated.

3. MACK MODEL

Mack uses the common CL loss development model and demonstrates that, under specific assumptions, the best estimate of the age-to-age factors is the all-year volume weighted average:[6]

$$\hat{F}(d) = \frac{\sum_{j=1}^{N-d} c(j, d+1)}{\sum_{j=1}^{N-d} c(j, d)} \tag{3.1}$$

Further, given the best estimate of the age-to-age factors, the best estimate of the ultimate value, given the known data, is calculated from the product of the age-to-age factors:[7]

$$E[\hat{c}(w,n)|D] = c(w,d) \times \hat{F}(d) \times \hat{F}(d+1) \times \cdots \times \hat{F}(n-1) \tag{3.2}$$

3.1 Model Assumptions

For Mack's distribution free estimates of the variance, the formulas rest on three key assumptions. The first assumption is that the expected value of the next future cumulative value is the product of the previous cumulative value and the age-to-age factor:

$$E[\hat{c}(w,d+1)|D] = c(w,d) \times \hat{F}(d) \tag{3.3}$$

The second assumption is that the accident years are independent of one another:

$$\{c(i,1), c(i,2), \ldots, c(i,n)\} \;\&\; \{c(j,1), c(j,2), \ldots, c(j,n)\} \text{ are independent for all } i \neq j \tag{3.4}$$

The third assumption is that the variance of the next cumulative is proportional to the cumulative value:

$$Var[\hat{c}(w,d+1)|D] = c(w,d) \times \sigma_d^2 \tag{3.5}$$

Testing of these assumptions has been discussed by Mack and other authors so, like the proofs, the details of this testing are not included with this paper.[8]

3.2 Uncertainty by Accident Year

Building on these assumptions, the first step in calculating the total variance by accident year is to calculate the variance of the development periods, σ_d^2. Mack demonstrates that the unbiased estimator

[6] See step 2 in tabs "Mack", "M&W" and "Alternative" in the Excel file.
[7] See step 3 in tabs "Mack", "M&W" and "Alternative" in the Excel file.
[8] For example, see Venter [10].

of the variance of the development periods is calculated using formula (3.6).[9]

$$\hat{\sigma}_d^2 = \frac{1}{N-d-1} \times \sum_{j=1}^{N-d} c(j,d) \times \left\{ \frac{c(j,d+1)}{c(j,d)} - \hat{F}(d) \right\}^2 ; \quad 1 \le d \le n-2 \tag{3.6}$$

The interpretation of formula (3.6) is straightforward as this is the commonly used weighted standard deviation of the age-to-age factors, noting that $N-d$ is the number of individual age-to-age factors for development period d. For the last age-to-age factor, if $\hat{F}(n-1) = 1$ then we could assume that the development is finished and set $\hat{\sigma}_{n-1}^2 = 0$. However, if $\hat{F}(n-1) \neq 1$ then Mack suggested that the value for $\hat{\sigma}_{n-1}^2$ could be calculated by extrapolating using a loglinear regression of $\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_{n-2}$. Mack also suggested a simpler approach using formula (3.7), which is used in the examples that follow.

$$\hat{\sigma}_{n-1}^2 = \min\left[\frac{\hat{\sigma}_{n-2}^4}{\hat{\sigma}_{n-3}^2}, \min\{\hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2\} \right] \tag{3.7}$$

Using the estimated variances by development period, Mack then demonstrates that the MSEP for the reserves by accident year can be calculated using formula (3.8).[10]

$$Var\left[\hat{R}(w)\right] = \hat{c}(w,n)^2 \times \sum_{d=n+1-w}^{n-1} \frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \left\{ \frac{1}{\hat{c}(w,d)} + \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right\} \tag{3.8}$$

Reviewing the formula for the variance of the unpaid claims by accident year, (3.8), we can distinguish between the process variance component,[11] which is the variance of the column of observed development factors, and the parameter variance component,[12] which is the variance of the calculated weighted average development factors.

3.3 Total Uncertainty

To calculate the total variance for all accident years combined, we can rely on basic principles of statistics as the unpaid claim estimates are assumed to be the expected values, so the total estimated unpaid claims is the sum of the estimated unpaid claims by accident year, as shown in formula (3.9).

$$\hat{R}(tot) = \hat{R}(2) + \hat{R}(3) + \cdots + \hat{R}(N) \tag{3.9}$$

Similarly, the total variance for all accident years is the sum of the variances plus 2 times the covariance, as shown in formula (3.10).

---

[9] See step 5 in tabs "Mack", "M&W" and "Alternative" in the Excel file.
[10] See step 8 in tab "Mack" in the Excel file. The covariance adjustment in step 8 will be discussed in the next section.
[11] See step 8a in tab "Mack" in the Excel file.
[12] See step 8b in tab "Mack" in the Excel file.

$$Var\left[\hat{R}(tot)\right] = Var\left[\hat{R}(2)\right] + Var\left[\hat{R}(3)\right] + \cdots + Var\left[\hat{R}(N)\right] + 2 \times CoVariance \qquad (3.10)$$

Using these basic principles of statistics, Mack developed the formula for the total variance, as shown in formula (3.11), which completes the modeling framework.[13,14]

$$Var\left[\hat{R}(tot)\right] = \sum_{w=2}^{N} \left\{ Var\left[\hat{R}(w)\right] + 2\hat{c}(w,n)\left(\sum_{i=w+1}^{N} c(i,n)\right) \sum_{d=n+1-w}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right)\right\} \qquad (3.11)$$

It is convenient to segregate the "bottom" or covariance portion[15] of (3.11) when showing the results of the Mack calculations as this makes it easier for the user to quickly calculate the total uncertainty with and without the covariance adjustment (CVA) – i.e., assuming no correlation in the accident years. To illustrate the calculations in all formulas in this paper, we will use the triangle from Taylor & Ashe [9] as our sample data, shown in Table 3.1.

Table 3.1 – Sample Data Triangle

		d									
		1	2	3	4	5	6	7	8	9	10
w	1	357,848	1,124,788	1,735,330	2,218,270	2,745,596	3,319,994	3,466,336	3,606,286	3,833,515	3,901,463
	2	352,118	1,236,139	2,170,033	3,353,322	3,799,067	4,120,063	4,647,867	4,914,039	5,339,085	
	3	290,507	1,292,306	2,218,525	3,235,179	3,985,995	4,132,918	4,628,910	4,909,315		
	4	310,608	1,418,858	2,195,047	3,757,447	4,029,929	4,381,982	4,588,268			
	5	443,160	1,136,350	2,128,333	2,897,821	3,402,672	3,873,311				
	6	396,132	1,333,217	2,180,715	2,985,752	3,691,712					
	7	440,832	1,288,463	2,419,861	3,483,130						
	8	359,480	1,421,128	2,864,498							
	9	376,686	1,363,294								
	10	344,014									
	$\hat{F}(d)$	3.4906	1.7473	1.4574	1.1739	1.1038	1.0863	1.0539	1.0766	1.0177	
	$\hat{\sigma}_d$	400.35	194.26	204.85	123.22	117.18	90.48	21.13	33.87	21.13	

Using formulas (3.8) and (3.11), the results for the sample data triangle are shown in Table 3.2. While formula (3.11) can be used to directly calculate the total variance of 2,447,095, segregating the covariance adjustment allows us to also directly calculate the total variance assuming zero correlation of 2,038,397. The coefficient of variation (CoV) column is the standard deviation divided by the mean.

[13] In some sense this modeling framework is not yet complete, as it does not include the tail variability. For ease of exposition, the tail variability is ignored in the paper but for completeness the companion Excel files include tail variability. The companion Excel files also allow the user to include exposure adjustments and exclude outliers.

[14] See step 9 in tab "Mack" in the Excel file.

[15] See step 9a in tab "Mack" in the Excel file (or step 9b if exposure-adjusted data)

Table 3.2 – Mack Estimated Unpaid Claims and Standard Deviations

		$\hat{R}(w)$	$\sqrt{Var[\hat{R}(w)]}$	CoV	CVA	$\sqrt{Var[\hat{R}(w)']}$	CoV
w	1	-	-	0.0%	-	-	0.0%
	2	94,634	75,535	79.8%	-	75,535	79.8%
	3	469,511	121,699	25.9%	*81,086*	146,238	31.1%
	4	709,638	133,549	18.8%	*139,674*	193,246	27.2%
	5	984,889	261,406	26.5%	*176,876*	315,624	32.0%
	6	1,419,459	411,010	29.0%	*259,674*	486,168	34.3%
	7	2,177,641	558,317	25.6%	*388,850*	680,384	31.2%
	8	3,920,301	875,328	22.3%	*573,313*	1,046,368	26.7%
	9	4,278,972	971,258	22.7%	*721,693*	1,210,034	28.3%
	10	4,625,811	1,363,155	29.5%	*841,236*	1,601,833	34.6%
	CVA		*1,353,961*		*1,353,961*		
	Total	**18,680,856**	**2,447,095**	**13.1%**		**2,447,095**	**13.1%**
	Ex CVA		*2,038,397*	*10.9%*			

In addition to the commonly used display of the Mack estimates in the first three columns of Table 3.2, an interesting alternative is to include the covariance adjustment with the accident years.[16] Since the covariance adjustment in formula (3.11) includes portions related to each accident year, we can include the portion related to each accident year in an expansion of formula (3.8).

$$Var[\hat{R}(w)'] = \hat{c}(w,n)^2 \times \sum_{d=n+1-w}^{n-1} \frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \left\{ \frac{1}{\hat{c}(w,d)} + \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right\} + 2\hat{c}(w,n) \left(\sum_{i=w+1}^{N} c(i,n) \right) \sum_{d=n+1-w}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \quad (3.12)$$

When using formula (3.12) to include a portion of the covariance adjustment, the formula for the total variance shown in (3.10) is revised as shown in formula (3.13).

$$Var[\hat{R}(tot)] = Var[\hat{R}(2)'] + Var[\hat{R}(3)'] + \cdots + Var[\hat{R}(N)'] \quad (3.13)$$

This alternative view of the Mack estimates is also included in Table 3.2, starting with the column that shows the portion of the covariance adjustment "allocated" to each accident year.[17] Note that for the alternative view the CoVs exhibit a smoother transition from the oldest year to the most current year, which may make comparisons to other models more consistent.

[16] See last columns in step 8 in tab "Mack" in the Excel file.

[17] Technically, the CVA column is only the covariance portion of formula (3.12) and the alternative standard deviation column can be calculated from the square root of the sum of the squares of the original standard deviation column and the CVA column.

3.4 Unpaid Claim Runoff Uncertainty

Before looking at the formulas for the time horizon calculations introduced by Merz-Wüthrich, it is useful to start with the runoff of the unpaid claims.[18] The Mack runoff formulas can be used to calculate the risk margin using the cost of capital method and it will be a useful comparison to the runoff using the Merz-Wüthrich formulas.

In order to extend the Mack formulas for the runoff of the unpaid claims, we must first review the notation related to time. For this purpose we will designate the "at time" using a subscript $t = 0,1,\dots,u$ and we will designate the "time-horizon" within the formulas using a superscript $T = 1,2,\dots,U$. Including this new notation, we could restate the results from formulas (3.8) and (3.11) as $Var[\hat{R}_0^U(w)]$ and $Var[\hat{R}_0^U(tot)]$, respectively.[19,20] In this case, since we are starting from the end of the known data, D, the subscript is zero and because both the process and parameter variances are being calculated over the entire time horizon this is commonly referred to as the "ultimate" time horizon, which is designated with the superscript U.

If we start by running off the estimated unpaid claims, the notation in section 2 can be restated for $t = 1$, as shown in formula (3.14) for $w = 3,4,\dots,N$, and noting that the "new" latest diagonal is estimated by multiplying the last diagonal times the age-to-age factors from (3.1).

$$\hat{R}_1(w) = \hat{U}(w) - \hat{c}(w, N - w + 2) \tag{3.14}$$

We can generalize this further for any t as shown in formula (3.15) for $w = t + 2, t + 3, \dots, N$.

$$\hat{R}_t(w) = \hat{U}(w) - \hat{c}(w, N - w + t + 1) \tag{3.15}$$

Applying formulas (3.14) and (3.15) to the sample data we can show the runoff of the estimated unpaid claims in Table 3.3.

[18] As noted in Section 1.1, proofs for the original Mack formulas are not included with this paper and as the extensions in Sections 3.4 and 3.5 follow the same assumptions and formulations, although reorganized for the cash flows, they are included without proofs.

[19] Both the subscripts and superscripts are shown here as a bridge to the time horizon discussion that starts in section 4, but for any formula where the subscript is absent it can be assumed to be zero and when the superscript is absent it can be assumed to be U.

[20] In the Excel file (tab "Mack"), the steps / calculations are repeated for each valuation at time t = 1, 2, 3, etc.

Table 3.3 – Runoff of Estimated Unpaid Claims

		$\hat{R}_t(w)$								
	$t=$	0	1	2	3	4	5	6	7	8
w	1	-	-	-	-	-	-	-	-	-
	2	94,634	-	-	-	-	-	-	-	-
	3	469,511	93,678	-	-	-	-	-	-	-
	4	709,638	462,448	92,268	-	-	-	-	-	-
	5	984,889	650,741	424,066	84,611	-	-	-	-	-
	6	1,419,459	1,036,173	684,625	446,148	89,016	-	-	-	-
	7	2,177,641	1,572,093	1,147,592	758,242	494,122	98,588	-	-	-
	8	3,920,301	2,610,043	1,884,254	1,375,463	908,802	592,237	118,164	-	-
	9	4,278,972	3,260,138	2,170,522	1,566,954	1,143,840	755,764	492,507	98,266	-
	10	4,625,811	3,769,007	2,871,597	1,911,841	1,380,205	1,007,518	665,692	433,810	86,555
	Total	**18,680,856**	**13,454,320**	**9,274,925**	**6,143,258**	**4,015,986**	**2,454,107**	**1,276,363**	**532,076**	**86,555**

Running off the variance of the unpaid claims by accident year for $t = 1$, formula (3.8) can be restated as formula (3.16) for $w = 3,4,\dots,N$.

$$Var\left[\hat{R}_1(w)\right] = \hat{c}(w,n)^2 \times \sum_{d=n+2-w}^{n-1} \frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \left\{ \frac{1}{\hat{c}(w,d)} + \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right\} \tag{3.16}$$

Generalizing this further for any t is shown in formula (3.17) for $w = t+2, t+3, \dots, N$.[21]

$$Var\left[\hat{R}_t(w)\right] = \hat{c}(w,n)^2 \times \sum_{d=n+t+1-w}^{n-1} \frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \left\{ \frac{1}{\hat{c}(w,d)} + \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right\} \tag{3.17}$$

Similarly, running off the total variance of the unpaid claims for all accident years for $t = 1$, formula (3.11) can be restated as formula (3.18) for $w = 3,4,\dots,N$.

$$\begin{aligned} Var\left[\hat{R}_1(tot)\right] = \sum_{w=3}^{N} \Bigg\{ & Var\left[\hat{R}_1(w)\right] \\ & + 2\hat{c}(w,n) \times \left(\sum_{i=w+1}^{N} c(i,n) \right) \sum_{d=n+2-w}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \Bigg\} \end{aligned} \tag{3.18}$$

Generalizing this further for any t is shown in formula (3.19) for $w = t+2, t+3, \dots, N$.[22]

$$\begin{aligned} Var\left[\hat{R}_t(tot)\right] = \sum_{w=t+2}^{N} \Bigg\{ & Var\left[\hat{R}_t(w)\right] \\ & + 2\hat{c}(w,n) \times \left(\sum_{i=w+1}^{N} c(i,n) \right) \sum_{d=n+t+1-w}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \Bigg\} \end{aligned} \tag{3.19}$$

[21] See steps 8, 8a and 8b in tab "Mack" in the Excel file.
[22] See step 9 in tab "Mack" in the Excel file.

Applying formulas (3.16) to (3.19) to the sample data, we can show the runoff of the estimated standard deviations of the unpaid claims in Table 3.4.

Table 3.4 – Runoff of Estimated Standard Deviations of the Unpaid Claims

$\sqrt{Var[\hat{R}_t(w)]}$

	$t=$	0	1	2	3	4	5	6	7	8
w	1	-	-	-	-	-	-	-	-	-
	2	75,535	-	-	-	-	-	-	-	-
	3	121,699	74,931	-	-	-	-	-	-	-
	4	133,549	120,373	74,041	-	-	-	-	-	-
	5	261,406	125,695	113,131	69,186	-	-	-	-	-
	6	411,010	269,797	130,224	117,306	71,982	-	-	-	-
	7	558,317	437,273	287,714	139,969	126,301	78,029	-	-	-
	8	875,328	623,100	489,142	323,291	159,581	144,441	90,307	-	-
	9	971,258	785,070	557,224	436,400	287,117	139,643	125,999	77,826	-
	10	1,363,155	903,373	729,436	516,796	404,139	265,121	127,697	114,976	70,421
	CVA	*1,353,961*	*1,039,055*	*773,477*	*556,945*	*384,712*	*263,965*	*170,358*	*79,424*	-
	Total	**2,447,095**	**1,788,912**	**1,340,940**	**954,131**	**663,602**	**431,762**	**263,362**	**159,952**	**70,421**

As expected, the standard deviations decrease in a similar fashion to the estimated unpaid claims and when $t = 8$ there is no longer a covariance adjustment term since there is only one "cell" remaining. As another test of the entire runoff process, we can look at the coefficients of variation shown in Table 3.5.

Table 3.5 – Runoff of Coefficients of Variation of the Unpaid Claims

CoV

	$t=$	0	1	2	3	4	5	6	7	8
w	1	-	-	-	-	-	-	-	-	-
	2	79.8%	-	-	-	-	-	-	-	-
	3	25.9%	80.0%	-	-	-	-	-	-	-
	4	18.8%	26.0%	80.2%	-	-	-	-	-	-
	5	26.5%	19.3%	26.7%	81.8%	-	-	-	-	-
	6	29.0%	26.0%	19.0%	26.3%	80.9%	-	-	-	-
	7	25.6%	27.8%	25.1%	18.5%	25.6%	79.1%	-	-	-
	8	22.3%	23.9%	26.0%	23.5%	17.6%	24.4%	76.4%	-	-
	9	22.7%	24.1%	25.7%	27.9%	25.1%	18.5%	25.6%	79.2%	-
	10	29.5%	24.0%	25.4%	27.0%	29.3%	26.3%	19.2%	26.5%	81.4%
	Total	**13.1%**	**13.3%**	**14.5%**	**15.5%**	**16.5%**	**17.6%**	**20.6%**	**30.1%**	**81.4%**

From Table 3.5 we can see that the total coefficient of variation increases as we progress from $t = 0,1, \ldots ,8$. This makes sense statistically as estimates further in the future should be relatively more uncertain.

Adjusting the generalized formula (3.17) to include the covariance adjustment related to each

accident year, we can use formula (3.20).[23]

$$Var\left[\hat{R}_t(w)'\right] = \hat{c}(w,n)^2 \times \sum_{d=n+t+1-w}^{n-1} \frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \left\{ \frac{1}{\hat{c}(w,d)} + \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right\} + 2\hat{c}(w,n) \times \left(\sum_{i=w+1}^{N} c(i,n) \right) \sum_{d=n+t+1-w}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \tag{3.20}$$

Using formula (3.20), the runoff of the standard deviations in Table 3.4 can be restated as shown in Table 3.6.

Table 3.6 – Runoff of Estimated Standard Deviations of the Unpaid Claims

$\sqrt{Var[\hat{R}_t(w)']}$

	t =	0	1	2	3	4	5	6	7	8
w	1	-	-	-	-	-	-	-	-	-
	2	75,535	-	-	-	-	-	-	-	-
	3	146,238	74,931	-	-	-	-	-	-	-
	4	193,246	144,569	74,041	-	-	-	-	-	-
	5	315,624	182,890	136,340	69,186	-	-	-	-	-
	6	486,168	322,928	185,489	139,093	71,982	-	-	-	-
	7	680,384	516,048	342,289	197,511	149,869	78,029	-	-	-
	8	1,046,368	761,474	577,804	384,423	225,461	171,765	90,307	-	-
	9	1,210,034	960,541	700,295	528,807	351,362	210,222	156,485	77,826	-
	10	1,601,833	1,125,689	893,426	647,922	488,300	326,547	191,615	139,742	70,421
	Total	**2,447,095**	**1,788,912**	**1,340,940**	**954,131**	**663,602**	**431,762**	**263,362**	**159,952**	**70,421**

The coefficients of variation comparing the standard deviations in Table 3.6 to the expected values in Table 3.3 are shown in Table 3.7. As noted above for Table 3.2, there is a smoother transition of all CoVs from the oldest year to the most current year.

[23] See last columns in step 8 in tab "Mack" in the Excel file.

Table 3.7 – Runoff of Coefficients of Variation of the Unpaid Claims

		CoV								
	t =	0	1	2	3	4	5	6	7	8
w	1	-	-	-	-	-	-	-	-	-
	2	79.8%	-	-	-	-	-	-	-	-
	3	31.1%	80.0%	-	-	-	-	-	-	-
	4	27.2%	31.3%	80.2%	-	-	-	-	-	-
	5	32.0%	28.1%	32.2%	81.8%	-	-	-	-	-
	6	34.3%	31.2%	27.1%	31.2%	80.9%	-	-	-	-
	7	31.2%	32.8%	29.8%	26.0%	30.3%	79.1%	-	-	-
	8	26.7%	29.2%	30.7%	27.9%	24.8%	29.0%	76.4%	-	-
	9	28.3%	29.5%	32.3%	33.7%	30.7%	27.8%	31.8%	79.2%	-
	10	34.6%	29.9%	31.1%	33.9%	35.4%	32.4%	28.8%	32.2%	81.4%
	Total	**13.1%**	**13.3%**	**14.5%**	**15.5%**	**16.5%**	**17.6%**	**20.6%**	**30.1%**	**81.4%**

3.5 Cash Flow Uncertainty

In order to extend the Mack formulas for the uncertainty of the cash flows we need to focus on the calendar year diagonals where $k = w - d + 1$. Starting with the calendar year estimated unpaid claims, we can introduce new notation for cash flow, $CF(k)$, and use the formula shown in formula (3.21) for $k = N+1, N+2, \ldots, N+n$.[24]

$$\widehat{CF}(k) = \sum_{j=k-N}^{N} \begin{Bmatrix} \hat{c}(j,k+1-j) - c(j,k-j); & k = N+1 \\ \hat{c}(j,k+1-j) - \hat{c}(j,k-j); & k > N+1 \end{Bmatrix} \tag{3.21}$$

Of course, summing the estimated unpaid for all calendar years as shown in formula (3.22) should result in the same total estimated unpaid as in formula (3.9).

$$\widehat{CF}(tot) = \widehat{CF}(N+1) + \widehat{CF}(N+2) + \cdots + \widehat{CF}(N+n) \tag{3.22}$$

Reorganizing Mack's formula (3.8) for the variance of each accident year into a diagonal sum results in formula (3.23). Note, however, that while the variance for an accident year is based on the ultimate estimated amount for that accident year, for the calendar year each of the accident year component variances are based on the estimated cumulative amount for next year end, i.e., formula (3.23) uses $\hat{c}(j,k+1-j)^2$ instead of $\hat{c}(j,n)^2$.[25]

$$Var\left[\widehat{CF}(k)\right] = \sum_{j=k-N}^{N} \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \tag{3.23}$$

Like formula (3.10), the total variance for all calendar years is the sum of the variances plus 2 times the covariance, as shown in formula (3.24).

[24] See step 10 in tabs "Mack", "M&W" and "Alternative" in the Excel file.
[25] See step 11 in tab "Mack" in the Excel file. The covariance adjustment in step 11 will be discussed below.

$$Var\left[\widehat{CF}(tot)\right] = Var\left[\widehat{CF}(N+1)\right] + Var\left[\widehat{CF}(N+2)\right] + \cdots + Var\left[\widehat{CF}(N+n)\right] + 2 \times CoVariance \tag{3.24}$$

Similarly, formula (3.11) for the variance of the total of all accident years can be reorganized as the sum of the calendar years,[26] plus the differences between the accident year variances from formula (3.8) and calendar year variances from formula (3.23),[27] as shown in formula (3.25).

$$\begin{aligned} Var\left[\widehat{CF}(tot)\right] = \sum_{k=N+1}^{N+n} \Bigg\{ & Var\left[\widehat{CF}(k)\right] + 2\hat{c}(k-N \\ & + 1, n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \sum_{d=N+n-k}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \Bigg\} \\ & + \sum_{j=k-N}^{N} [\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \end{aligned} \tag{3.25}$$

Table 3.8 – Mack Estimated Cash Flows and Standard Deviations

		$\widehat{CF}(k)$	$\sqrt{Var[\widehat{CF}(k)]}$	CoV	CVA	$\sqrt{Var[\widehat{CF}(k)']}$	CoV
k	11	5,226,536	665,562	12.7%	*1,531,370*	1,669,750	31.9%
	12	4,179,394	609,716	14.6%	*1,015,053*	1,184,097	28.3%
	13	3,131,668	558,467	17.8%	*758,861*	942,208	30.1%
	14	2,127,272	445,167	20.9%	*521,368*	685,565	32.2%
	15	1,561,879	353,389	22.6%	*359,256*	503,933	32.3%
	16	1,177,744	248,729	21.1%	*234,931*	342,139	29.1%
	17	744,287	142,151	19.1%	*153,519*	209,224	28.1%
	18	445,521	118,457	26.6%	*81,200*	143,616	32.2%
	19	86,555	70,421	81.4%	-	70,421	81.4%
	CVA		*2,106,547*		*2,106,547*		
	Total	**18,680,856**	**2,447,095**	**13.1%**		**2,447,095**	**13.1%**

As with other modeling frameworks, the sums of the means and variances by diagonal should be consistent with the sums by row, as seen in Table 3.8. In other words, the totals in the first three columns of Table 3.8 are the same as the totals in the first three columns of Table 3.2. Ideally, the CoVs should increase steadily as the future diagonals should represent more uncertainty, i.e., as k increases from 11 to 19, but for the data in the example the CoVs are relatively consistent from $k = 11, \ldots, 18$ and then jump significantly for $k = 19$.

[26] See step 11a in tab "Mack" in the Excel file (or step 11c if exposure-adjusted data).
[27] See step 11b in tab "Mack" in the Excel file (or step 11d if exposure-adjusted data).

Like the adjustment of the accident year variance in formula (3.12), the expansion to formula (3.23) to include a portion of the covariance adjustment by calendar year is shown as formula (3.26).[28] The alternative view the CoVs in Table 3.8 exhibit a similar consistency from $k = 11, \dots, 18$ and then jump significantly for $k = 19$. Note that the covariance adjustment excludes the last diagonal, i.e., $i = w + N - 1$ in formula (3.11), so none of the CVA is allocated to $k = 19$ in Table 3.8.

$$\begin{aligned} Var\left[\widehat{CF}(k)'\right] = & \sum_{j=k-N}^{N} \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\ & + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \sum_{d=N+n-k}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \\ & + [\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \end{aligned} \tag{3.26}$$

After revising formula (3.23) to include a portion of the covariance adjustment, formula (3.24) for the total variance is revised as shown in formula (3.27).

$$Var\left[\widehat{CF}(tot)\right] = Var\left[\widehat{CF}(N+1)'\right] + Var\left[\widehat{CF}(N+2)'\right] + \cdots + Var\left[\widehat{CF}(N+n)'\right] \tag{3.27}$$

4. MERZ & WÜTHRICH MODEL

The premise of the 1-year time horizon is that if claims develop unfavorably over the subsequent 12 months and capital becomes impaired then management could intervene. Based on this premise as implemented for the Solvency II regime, the Merz & Wüthrich model calculates the uncertainty in the reserves after one year given the total uncertainty (i.e., the possible outcomes) during the first year. In other words, over a 1-year time horizon (i.e., the first diagonal), all possible outcomes should be considered and then the new reserves, conditional on each possible outcome, are calculated.

4.1 Uncertainty by Accident Year: One-Year Time Horizon

The formulas developed by Merz & Wüthrich [7] to calculate the unpaid claim uncertainty over a 1-year time horizon build on Mack's formulas and assumptions shown in (3.1) to (3.7). Starting with Mack's accident year uncertainty from (3.8), Merz-Wüthrich split the formula into components based on the first diagonal and the remaining diagonals as shown in (4.1).[29]

[28] See last columns in step 11 in tab "Mack" in the Excel file.
[29] See step 8 in tab "M&W" in the Excel file. The covariance adjustment in step 8 will be discussed in the next section.

$$Var\left[\hat{R}^1(w)\right] = \hat{c}(w,n)^2 \times \frac{\hat{\sigma}^2_{N+1-w}}{\hat{F}(N+1-w)^2} \times \left\{\frac{1}{c(w,N+1-w)} + \frac{1}{\sum_{j=1}^{w-1} c(j,N+1-w)}\right\}$$
$$+\hat{c}(w,n)^2 \times \sum_{d=n+2-w}^{n-1} \frac{\hat{\sigma}^2_d}{\hat{F}(d)^2} \times \left\{\alpha^1_d \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right\} \tag{4.1}$$

For the first diagonal, both the process and parameter uncertainty[30] are included such that the results will exactly match the Mack results for the first diagonal as in formula (3.23). For the remaining diagonals, only the parameter uncertainty[31] is included and it is also reduced a bit using a weight function, α^1_d, which is calculated using formula (4.2).[32]

$$\alpha^1_d = \frac{c(N+1-d,d)}{\sum_{j=1}^{N+1-d} c(j,d)}; for\ d = 1,2,\dots,N \tag{4.2}$$

The use of color for the weight components of the formulas is useful for clarifying the calculations and tracing the components through the various formulas.[33] The weight function can be thought of as an adjustment to the development factor, $F(d)$, and the parameter uncertainty for the years after the time horizon.

4.2 Total Uncertainty: One-Year Time Horizon

Adjusting the Mack formula (3.11) for the total uncertainty for the 1-year time horizon, Merz-Wüthrich developed formula (4.3), which also separates the covariance[34] into the first diagonal and remaining diagonal components.

$$Var\left[\hat{R}^1(tot)\right] = \sum_{w=2}^{N} \Bigg\{ Var\left[\hat{R}^1(w)\right]$$
$$+ 2\hat{c}(w,n) \times \left(\sum_{i=w+1}^{N} c(i,n)\right)$$
$$\times \left[\frac{\hat{\sigma}^2_{N+1-w}}{\hat{F}(N+1-w)^2} \times \frac{1}{\sum_{j=1}^{w-1} c(j,N+1-w)}\right.$$
$$\left. + \sum_{d=n+2-w}^{n-1} \left(\frac{\hat{\sigma}^2_d}{\hat{F}(d)^2} \times \alpha^1_d \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right)\right]\Bigg\} \tag{4.3}$$

Using formulas (4.1) and (4.3), the results for the sample data triangle are shown in Table 4.1.

[30] See steps 8a and 8c respectively in tab "M&W" in the Excel file.
[31] See step 8b in tab "M&W" in the Excel file.
[32] See step 6 in tab "M&W" in the Excel file.
[33] Alternatively, the weight functions could be colored as part of the parameter uncertainty but using a different color will help in later parts of the paper.
[34] See step 9a in tab "M&W" in the Excel file (or step 9b if exposure-adjusted data).

Comparing the results in the first three columns of Table 4.1 with the same columns in Table 3.2, note that for $w = 2$ the uncertainty is entirely for the first diagonal and, as such, the standard deviations are exactly the same.[35] For $w > 2$ the uncertainties in Table 4.1 are a combination of the first diagonal and the remaining diagonals and, as such, the standard deviations are less than those in Table 3.2. Finally, the covariance adjustment for the total uncertainty in Table 4.1 is also less than in Table 3.2, resulting in a total standard deviation of 1,778,968 compared to 2,447,095.

Table 4.1 – Merz-Wüthrich Estimated Unpaid Claims and Standard Deviations

		$\hat{R}^1(w)$	$\sqrt{Var[\hat{R}^1(w)]}$	CoV	CVA	$\sqrt{Var[\hat{R}^1(w)']}$	CoV
w	1	-	-	0.0%	-	-	0.0%
	2	94,634	75,535	79.8%	-	75,535	79.8%
	3	469,511	105,309	22.4%	*81,086*	132,910	28.3%
	4	709,638	79,846	11.3%	*129,729*	152,332	21.5%
	5	984,889	235,115	23.9%	*150,379*	279,093	28.3%
	6	1,419,459	318,427	22.4%	*226,186*	390,584	27.5%
	7	2,177,641	361,089	16.6%	*323,435*	484,763	22.3%
	8	3,920,301	629,681	16.1%	*441,515*	769,047	19.6%
	9	4,278,972	588,662	13.8%	*541,749*	800,010	18.7%
	10	4,625,811	1,029,925	22.3%	*600,426*	1,192,165	25.8%
	CVA		*1,025,050*		*1,025,050*		
	Total	**18,680,856**	**1,778,968**	**9.5%**		**1,778,968**	**9.5%**
	Ex CVA		*1,453,959*	*7.8%*			

Like the alternative view of the covariance adjustment by accident year for the Mack model, a portion of the covariance adjustment in formula (4.3) can be included with formula (4.1) as shown in formula (4.4).[36]

[35] As the Merz-Wüthrich formulas only address changes to the Mack standard deviations, the expected values are the same – i.e., $\hat{R}(w) = \hat{R}^1(w)$. The identical standard deviations for both Mack and Merz-Wüthrich for $w = 2$ is expected since the first diagonal includes both process and parameter variance for both formulas.

[36] See last columns in step 8 in tab "M&W" in the Excel file.

$$Var\left[\hat{R}^1(w)'\right] = \hat{c}(w,n)^2 \times \frac{\hat{\sigma}^2_{N+1-w}}{\hat{F}(N+1-w)^2} \times \left\{\frac{1}{\hat{c}(w,N+1-w)} + \frac{1}{\sum_{j=1}^{w-1} c(j,N+1-w)}\right\}$$
$$+\hat{c}(w,n)^2 \times \sum_{d=n+2-w}^{n-1} \frac{\hat{\sigma}^2_d}{\hat{F}(d)^2} \times \left\{\alpha^1_d \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right\}$$
$$+2\hat{c}(w,n) \times \left(\sum_{i=w+1}^{N} c(i,n)\right) \times \left[\frac{\hat{\sigma}^2_{N+1-w}}{\hat{F}(N+1-w)^2} \times \frac{1}{\sum_{j=1}^{w-1} c(j,N+1-w)} + \sum_{d=n+2-w}^{n-1} \left(\frac{\hat{\sigma}^2_d}{\hat{F}(d)^2} \times \alpha^1_d \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right)\right] \quad (4.4)$$

This alternative view of the Merz-Wüthrich estimates is also included in Table 4.1, starting with the column that shows the portion of the covariance adjustment "allocated" to each accident year. Note that for the alternative view the CoVs exhibit a smoother transition from the oldest year to the most current year like the Mack alternative view.

4.3 Uncertainty by Accident Year: X-Year Time Horizon

The formulas developed by Merz & Wüthrich [7] above were subsequently extended in Merz & Wüthrich [8] to runoff the unpaid claim estimates for later time windows.[37] Starting with $T' = 2$, formula (4.1) is extended as shown in formula (4.5). In the Merz & Wüthrich [8] paper, the authors describe extensions of the "time horizon" for $T > 1$, but since the first diagonal in formula (4.5) does not include all of the process and parameter variances, in this paper we will refer to the extensions as "time windows" (and use the T' notation) to improve clarity between models.[38]

$$Var\left[\hat{R}^2(w)\right] =$$
$$\hat{c}(w,n)^2 \times \frac{\hat{\sigma}^2_{N+2-w}}{\hat{F}(N+2-w)^2} \times \left\{\frac{1}{\hat{c}(w,N+2-w)} + (1-\alpha^1_{N+2-w}) \times \frac{1}{\sum_{j=1}^{w-1} c(j,N+2-w)}\right\}$$
$$+\hat{c}(w,n)^2 \times \sum_{d=n+3-w}^{n-1} \frac{\hat{\sigma}^2_d}{\hat{F}(d)^2} \times \left\{\alpha^2_d \times (1-\alpha^1_d) \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right\} \quad (4.5)$$

Note that in the extension for $T' = 2$, one minus the weights for $T = 1$ are used and the formula for the weights for $T' = 2$ are as in formula (4.6). Note also that the calculation of the weights includes estimated cumulative values when $T > 1$.

---

[37] In the Excel file (tab "M&W"), the steps / calculations are repeated for each time horizon / window, T / T', all at time t=0.

[38] In some of the Tables that follow, the headers only refer to T' for simplicity but for $T = 1$ the conversion to T is implied.

$$\alpha_d^2 = \frac{\hat{c}(N+2-d,d)}{\sum_{j=1}^{N+2-d} c(j,d)}; for\ d = 2,3,\dots,N-1 \tag{4.6}$$

A further extension and generalization for $T' > 2$ is shown in formula (4.7)[39].

$$\begin{aligned} &Var\left[\hat{R}^{T'}(w)\right] = \\ &\hat{c}(w,n)^2 \times \frac{\hat{\sigma}_{N+T-w}^2}{\hat{F}(N+T-w)^2} \times \left\{ \frac{1}{\hat{c}(w,N+T-w)} + \prod_{m=1}^{T-1}(1-\alpha_{N+T-w}^m) \times \frac{1}{\sum_{j=1}^{w-1} c(j,N+T-w)} \right\} \\ &+\hat{c}(w,n)^2 \times \sum_{d=n+T+1-w}^{n-1} \frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \left\{ \alpha_d^T \times \prod_{m=1}^{T-1}(1-\alpha_d^m) \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right\} \end{aligned} \tag{4.7}$$

And the extension for the weight function is shown in formula (4.8).

$$\alpha_d^T = \frac{\hat{c}(N+T-d,d)}{\sum_{j=1}^{N+T-d} c(j,d)}; for\ d = T, T+1, \dots, N-T+1 \tag{4.8}$$

4.4 Total Uncertainty: X-Year Time Horizon

In Merz & Wüthrich [8] the extension of the total uncertainty for $T' = 2$ is shown in formula (4.9).

$$\begin{aligned} Var\left[\hat{R}^2(tot)\right] = \sum_{w=3}^{N} \Bigg\{ & Var\left[\hat{R}^2(w)\right] \\ & + 2\hat{c}(w,n) \times \left(\sum_{i=w+2}^{N} c(i,n) \right) \\ & \times \left[\frac{\hat{\sigma}_{N+2-w}^2}{\hat{F}(N+2-w)^2} \times (1-\alpha_{N+2-w}^1) \times \frac{1}{\sum_{j=1}^{w-1} c(j,N+2-w)} \right. \\ & \left. + \sum_{d=n+3-w}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \alpha_d^2 \times (1-\alpha_d^1) \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] \Bigg\} \end{aligned} \tag{4.9}$$

The further extension and generalization for $T' > 2$ is shown in formula (4.10).[40]

[39] See steps 8, 8a, 8b and 8c in tab "M&W" in the Excel file.
[40] See step 9a in tab "M&W" in the Excel file (or step 9b if exposure-adjusted data).

$$Var\left[\hat{R}^{T'}(tot)\right] = \sum_{w=T+1}^{N} \left\{ Var\left[\hat{R}^{T'}(w)\right] + 2\hat{c}(w,n)\left(\sum_{i=w+T}^{N} c(i,n)\right) \times \left[\frac{\hat{\sigma}^2_{N+T-w}}{\hat{F}(N+T-w)^2} \times \prod_{m=1}^{T-1}(1-\alpha^m_{N+T-w}) \times \frac{1}{\sum_{j=1}^{w-1} c(j,N+T-w)} + \sum_{d=n+T+1-w}^{n-1} \left(\frac{\hat{\sigma}^2_d}{\hat{F}(d)^2} \times \alpha^T_d \times \prod_{m=1}^{T-1}(1-\alpha^m_d) \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] \right\} \tag{4.10}$$

Applying formulas (4.1) to (4.10) to the sample data results in the standard deviations by year as shown in Table 4.2, with the results from Table 4.1 repeated in the first column of Table 4.2.

Table 4.2 – Runoff of Merz-Wüthrich Estimated Standard Deviations of the Unpaid Claims

$\sqrt{Var[\hat{R}^{T'}(w)]}$

	T'=	1	2	3	4	5	6	7	8	9	TOTAL
w	1	-	-	-	-	-	-	-	-	-	-
	2	75,535	-	-	-	-	-	-	-	-	75,535
	3	105,309	60,996	-	-	-	-	-	-	-	121,699
	4	79,846	91,093	56,232	-	-	-	-	-	-	133,549
	5	235,115	60,577	82,068	51,474	-	-	-	-	-	261,406
	6	318,427	233,859	57,825	82,433	51,999	-	-	-	-	411,010
	7	361,089	328,989	243,412	59,162	85,998	54,343	-	-	-	558,317
	8	629,681	391,249	359,352	266,320	64,443	94,166	59,533	-	-	875,328
	9	588,662	554,574	344,763	318,493	236,576	56,543	83,645	52,965	-	971,258
	10	1,029,925	538,726	511,118	317,142	293,978	218,914	51,661	77,317	49,055	1,363,155
	CVA	*1,025,050*	*676,444*	*449,236*	*288,887*	*164,691*	*92,828*	*57,595*	*24,085*	-	*1,353,961*
	Total	**1,778,968**	**1,177,727**	**885,178**	**607,736**	**428,681**	**267,503**	**128,557**	**96,764**	**49,055**	**2,447,095**

As expected, the standard deviations decrease in a similar fashion to the estimated unpaid claims and when $T' = 9$ there is no covariance adjustment term since there is only one "cell" remaining. An additional part of the results in Table 4.2 is the Total column, which is the square root of the sum of the squares of the other columns. The Total column is an important result as the complete runoff from Merz-Wüthrich are intended to reconcile with the results from Mack. Comparing the Total column in Table 4.2 with the results in Table 3.2 we see that the estimates are identical.

The expected runoff of the unpaid claims for Merz-Wüthrich is identical to the runoff for Mack, as previously shown in Table 3.3. Dividing the standard deviations in Table 4.2 by the means in Table 3.3 results in the runoff of the coefficients of variation shown in Table 4.3.

Table 4.3 – Runoff of Merz-Wüthrich Coefficients of Variation of the Unpaid Claims

		CoV									
	$T'=$	1	2	3	4	5	6	7	8	9	TOTAL
w	1	-	-	-	-	-	-	-	-	-	-
	2	79.8%	-	-	-	-	-	-	-	-	79.8%
	3	22.4%	65.1%	-	-	-	-	-	-	-	25.9%
	4	11.3%	19.7%	60.9%	-	-	-	-	-	-	18.8%
	5	23.9%	9.3%	19.4%	60.8%	-	-	-	-	-	26.5%
	6	22.4%	22.6%	8.4%	18.5%	58.4%	-	-	-	-	29.0%
	7	16.6%	20.9%	21.2%	7.8%	17.4%	55.1%	-	-	-	25.6%
	8	16.1%	15.0%	19.1%	19.4%	7.1%	15.9%	50.4%	-	-	22.3%
	9	13.8%	17.0%	15.9%	20.3%	20.7%	7.5%	17.0%	53.9%	-	22.7%
	10	22.3%	14.3%	17.8%	16.6%	21.3%	21.7%	7.8%	17.8%	56.7%	29.5%
	Total	**9.5%**	**8.8%**	**9.5%**	**9.9%**	**10.7%**	**10.9%**	**10.1%**	**18.2%**	**56.7%**	**13.1%**

Adjusting formulas (4.5) and (4.7) to include the covariance adjustment related to each accident year are left to the reader. Applying formula (4.4), and the extensions for formulas (4.5) and (4.7), the runoff of the standard deviations in Table 4.2 are restated in Table 4.4. The first column in Table 4.4 is from Table 4.1 but, more importantly, the Total column in Table 4.4 reconciles with the alternative view for Mack in Table 3.2.

Table 4.4 – Runoff of Merz-Wüthrich Estimated Standard Deviations of the Unpaid Claims

		$\sqrt{Var[\hat{R}^{T'}(w)']}$									
	$T'=$	1	2	3	4	5	6	7	8	9	TOTAL
w	1	-	-	-	-	-	-	-	-	-	-
	2	75,535	-	-	-	-	-	-	-	-	75,535
	3	132,910	60,996	-	-	-	-	-	-	-	146,238
	4	152,332	104,771	56,232	-	-	-	-	-	-	193,246
	5	279,093	103,950	90,942	51,474	-	-	-	-	-	315,624
	6	390,584	255,290	89,682	88,793	51,999	-	-	-	-	486,168
	7	484,763	377,458	258,077	86,475	91,743	54,343	-	-	-	680,384
	8	769,047	491,773	402,375	278,897	91,580	99,957	59,533	-	-	1,046,368
	9	800,010	658,702	429,906	356,254	247,299	81,487	89,102	52,965	-	1,210,034
	10	1,192,165	691,492	592,230	382,924	321,096	227,976	71,017	80,981	49,055	1,601,833
	Total	**1,778,968**	**1,177,727**	**885,178**	**607,736**	**428,681**	**267,503**	**128,557**	**96,764**	**49,055**	**2,447,095**

Table 4.5 – Runoff of Merz-Wüthrich Coefficients of Variation of the Unpaid Claims

		CoV									
	$T'=$	1	2	3	4	5	6	7	8	9	TOTAL
w	1	-	-	-	-	-	-	-	-	-	-
	2	79.8%	-	-	-	-	-	-	-	-	79.8%
	3	28.3%	65.1%	-	-	-	-	-	-	-	31.1%
	4	21.5%	22.7%	60.9%	-	-	-	-	-	-	27.2%
	5	28.3%	16.0%	21.4%	60.8%	-	-	-	-	-	32.0%
	6	27.5%	24.6%	13.1%	19.9%	58.4%	-	-	-	-	34.3%
	7	22.3%	24.0%	22.5%	11.4%	18.6%	55.1%	-	-	-	31.2%
	8	19.6%	18.8%	21.4%	20.3%	10.1%	16.9%	50.4%	-	-	26.7%
	9	18.7%	20.2%	19.8%	22.7%	21.6%	10.8%	18.1%	53.9%	-	28.3%
	10	25.8%	18.3%	20.6%	20.0%	23.3%	22.6%	10.7%	18.7%	56.7%	34.6%
	Total	**9.5%**	**8.8%**	**9.5%**	**9.9%**	**10.7%**	**10.9%**	**10.1%**	**18.2%**	**56.7%**	**13.1%**

The coefficients of variation comparing the standard deviations in Table 4.4 to the expected values in Table 3.3 are shown in Table 4.5. As noted above for Table 4.1, there is a smoother transition of

all CoVs from the oldest year to the most current year.

4.5 Cash Flow Uncertainty

The calculation of the cash flow uncertainty under the time horizon view is more complicated than for the ultimate view using Mack. The extension of the Mack formulas to calculate the cash flow uncertainty only requires one set of formulas as shown in (3.23) and (3.25). The extension of the Merz-Wüthrich formulas to calculate the cash flow uncertainty depends on the length of the time window, resulting in a different set of formulas for each of $T = 1,2,\dots,N-1$.[41]

Starting with the formulas for $T = 1$, it is more convenient to separate formula (3.23) into separate formulas for the first diagonal and remaining diagonals as shown in formula (4.11)[42].

$$Var\left[\widehat{CF}^1(k)\right] = \sum_{j=k-N}^{N} \begin{cases} \hat{c}(j,k+1-j)^2 \times \dfrac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \dfrac{1}{\hat{c}(j,k-j)} + \dfrac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k = N+1 \\ \hat{c}(j,k+1-j)^2 \times \dfrac{\hat{\sigma}_{N-j+1}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^1 \times \dfrac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k > N+1 \end{cases} \tag{4.11}$$

For the total uncertainty, it is also more convenient to separate formula (3.26) into separate formulas for the first diagonal and remaining diagonals as shown in formula (4.12).

$$Var\left[\widehat{CF}^1(tot)\right] = \sum_{k}^{N+n} \begin{cases} Var\left[\widehat{CF}^1(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \times \left[\dfrac{\hat{\sigma}_{k-N+1}^2}{\hat{F}(k-N+1)^2} \times \dfrac{1}{\sum_{j=1}^{w-1} c(j,k-N+1)} \right] & \\ \quad + [\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \dfrac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \dfrac{1}{\hat{c}(j,k-j)} + \dfrac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k = N+1 \\ Var\left[\widehat{CF}^1(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \times \left[\sum_{d=k-N+2}^{n-1} \left(\dfrac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \alpha_d^1 \times \dfrac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] & \\ \quad + [\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \dfrac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^1 + \dfrac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k > N+1 \end{cases} \tag{4.12}$$

Using formulas (4.11) and (4.12), the results for the sample data triangle are shown in Table 4.6.

[41] Similar to the Mack extensions, the extensions for Merz-Wüthrich follow the assumptions and formulations of the original papers so they are included without proofs.

[42] See step 11 in tab "M&W" in the Excel file. The covariance adjustment in step 11 will be discussed below.

Table 4.6 – Estimated Merz-Wüthrich Cash Flow and Standard Deviations for T=1

		$\widehat{CF}^1(k)$	$\sqrt{Var[\widehat{CF}^1(k)]}$	CoV	CVA	$\sqrt{Var[\widehat{CF}^1(k)']}$	CoV
k	11	5,226,536	665,562	12.7%	*1,531,370*	1,669,750	31.9%
	12	4,179,394	111,733	2.7%	*348,793*	366,252	8.8%
	13	3,131,668	108,154	3.5%	*284,901*	304,739	9.7%
	14	2,127,272	95,702	4.5%	*226,334*	245,735	11.6%
	15	1,561,879	83,976	5.4%	*177,520*	196,381	12.6%
	16	1,177,744	76,031	6.5%	*141,832*	160,926	13.7%
	17	744,287	67,017	9.0%	*109,047*	127,994	17.2%
	18	445,521	55,652	12.5%	*60,893*	82,493	18.5%
	19	86,555	40,213	46.5%	-	40,213	46.5%
	CVA		*1,632,904*		*1,632,904*		
	Total	**18,680,856**	**1,778,968**	**9.5%**		**1,778,968**	**9.5%**

Comparing the first three columns in Table 4.6 with Table 3.8, it makes sense that for $k = 11$, i.e., the first diagonal, the results are identical and, comparing Table 4.6 with Table 4.1, the total results are also identical as expected.

Like the alternative view of the covariance adjustment by calendar year for the Mack model, a portion of the covariance adjustment in formula (4.12) can be included with formula (4.11) as shown in formula (4.13).[43]

$$
\begin{aligned}
&Var\left[\widehat{CF}^1(k)'\right] \\
&= \sum_{j=k-N}^{N} \left\{ \begin{array}{r}
\hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}^2_{k-j}}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\
+2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \left[\frac{\hat{\sigma}^2_{k-N+1}}{\hat{F}(k-N+1)^2} \times \frac{1}{\sum_{j=1}^{w-1} c(j,k-N+1)} \right] \quad ;k = N+1 \\
+[\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}^2_{k-j}}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\
\hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}^2_{k-j}}{\hat{F}(k-j)^2} \times \left\{ \alpha^1_{k-j} \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\
+2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \left[+ \sum_{d=k-N+2}^{n-1} \left(\frac{\hat{\sigma}^2_d}{\hat{F}(d)^2} \times \alpha^1_d \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] \quad ;k > N+1 \\
+[\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}^2_{k-j}}{\hat{F}(k-j)^2} \times \left\{ \alpha^1_{k-j} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\}
\end{array} \right\}
\end{aligned}
\tag{4.13}
$$

This alternative view of the Merz-Wüthrich cash flow estimates is also included in Table 4.6, starting with the column that shows the portion of the covariance adjustment "allocated" to each

[43] See last columns in step 11 in tab "M&W" in the Excel file.

calendar year. Note that for the alternative view the CoVs exhibit a smoother transition from the first diagonal to the last diagonal like the Mack alternative view.

Continuing with the formulas for $T' = 2$, the formulas for the first diagonal and remaining diagonals are shown in formula (4.14).

$$Var\left[\widehat{CF}^2(k)\right] = \sum_{j=k-N}^{N} \left\{ \begin{array}{ll} \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + (1-\alpha_{k-j}^1) \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k = N+2 \\ \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^2 \times (1-\alpha_{k-j}^1) \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k > N+2 \end{array} \right\} \quad (4.14)$$

For the total uncertainty, the formulas for the first diagonal and remaining diagonals are shown in formula (4.15).

$$Var\left[\widehat{CF}^2(tot)\right] = \sum_{k}^{N+n} \left\{ \begin{array}{ll} Var\left[\widehat{CF}^2(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) & \\ \times \left[\frac{\hat{\sigma}_{k-N+1}^2}{\hat{F}(k-N+1)^2} \times (1-\alpha_{k-N+1}^1) \times \frac{1}{\sum_{j=1}^{w-1} c(j,k-N+1)} \right] & ;k = N+2 \\ +[\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + (1-\alpha_{k-j}^1) \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & \\ Var\left[\widehat{CF}^2(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) & \\ \times \left[\sum_{d=k-N+2}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \alpha_d^2 \times (1-\alpha_d^1) \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] & ;k > N+2 \\ +[\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^2 \times (1-\alpha_{k-j}^1) \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & \end{array} \right\} \quad (4.15)$$

Using formulas (4.14) and (4.15), the results for the sample data triangle are shown in Table 4.7.

Table 4.7 – Estimated Merz-Wüthrich Cash Flow and Standard Deviations for T'=2

		$\widehat{CF}^2(k)$	$\sqrt{Var[\widehat{CF}^2(k)]}$	CoV	CVA	$\sqrt{Var[\widehat{CF}^2(k)']}$	CoV
k	11						
	12	4,179,394	599,391	14.3%	*953,245*	1,126,031	26.9%
	13	3,131,668	86,156	2.8%	*213,751*	230,461	7.4%
	14	2,127,272	76,066	3.6%	*161,420*	178,445	8.4%
	15	1,561,879	62,836	4.0%	*114,746*	130,825	8.4%
	16	1,177,744	51,412	4.4%	*82,900*	97,548	8.3%
	17	744,287	38,525	5.2%	*60,077*	71,368	9.6%
	18	445,521	31,819	7.1%	*31,311*	44,641	10.0%
	19	86,555	20,602	23.8%	-	20,602	23.8%
	CVA		*1,002,522*		*1,002,522*		
	Total	**13,454,320**	**1,177,727**	**8.8%**		**1,177,727**	**8.8%**

Comparing Table 4.7 with Table 3.8, note that the standard deviation for $k = 12$, i.e., the first diagonal at time $t = 1$, in Table 4.7 is less than in Table 3.8, which makes sense since formula (4.17) for the first diagonal only includes a portion of the parameter uncertainty. Comparing Table 4.7 with the columns for $T' = 2$ in Tables 4.2 and 4.3, the totals are identical as expected. Table 4.7 also includes the alternative view of the covariance adjustment, but the derivation of the formula is left to the reader.

Continuing with the formulas for $T' > 2$, the formulas for the first diagonal and remaining diagonals are shown in formula (4.16).[44]

$$Var\left[\widehat{CF}^{T}(k)\right] = \sum_{j=k-N}^{N} \begin{Bmatrix} \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \prod_{m=1}^{T-1}(1-\alpha_{k-j}^{m}) \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ; k = N+T \\ \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^{T} \times \prod_{m=1}^{T-1}(1-\alpha_{k-j}^{m}) \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ; k > N+T \end{Bmatrix} \tag{4.16}$$

For the total uncertainty, the formulas for the first diagonal and remaining diagonals are shown in formula (4.17).

$$Var\left[\widehat{CF}^{T}(tot)\right] = \sum_{k}^{N+n} \begin{Bmatrix} Var\left[\widehat{CF}^{2}(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \\ \times \left[\frac{\hat{\sigma}_{k-N+1}^2}{\hat{F}(k-N+1)^2} \times \prod_{m=1}^{T-1}(1-\alpha_{k-j}^{m}) \times \frac{1}{\sum_{j=1}^{w-1} c(j,k-N+1)} \right] & ; k = N+T \\ + [\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \prod_{m=1}^{T-1}(1-\alpha_{k-j}^{m}) \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\ Var\left[\widehat{CF}^{2}(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \\ \times \left[\sum_{d=k-N+2}^{n-1} \left(\frac{\hat{\sigma}_{d}^2}{\hat{F}(d)^2} \times \alpha_{k-j}^{T} \times \prod_{m=1}^{T-1}(1-\alpha_{k-j}^{m}) \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] & ; k > N+T \\ + [\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^{T} \times \prod_{m=1}^{T-1}(1-\alpha_{k-j}^{m}) \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \end{Bmatrix} \tag{4.17}$$

Using formulas (4.16) and (4.17), the results for the sample data triangle will be similar to the results shown in Table 4.7, meaning the first diagonal will be less than the same diagonal in Table 3.6 and the totals will match the same time window in Tables 4.2 and 4.3.

4.6 A Comparison of Mack vs. Merz-Wüthrich

Now that we have reviewed the various formulas related to the Mack and Merz-Wüthrich models, it is instructive to compare the runoff for the two models using the totals from Tables 3.3, 3.4, 3.5,

[44] See last columns in step 11 in tab "M&W" in the Excel file.

4.2, and 4.3. As shown in Table 4.8, at time $t = 0$ (and $T = 1$) the standard deviation for the 1-year time horizon is 72.7% of the standard deviation for the ultimate time horizon. As previously discussed, this makes sense since the 1-year time horizon only includes the parameter variance beyond the first diagonal.

Table 4.8 – Comparison of Estimated Runoff for Mack and Merz-Wüthrich Models

		$\hat{R}_t(tot)$	$\sqrt{Var[\hat{R}_t(tot)]}$	CoV	$\sqrt{Var[\hat{R}^{T'}(tot)]}$	CoV	$Ratio$
$t=$	0	18,680,856	2,447,095	13.1%	1,778,968	9.5%	72.7%
	1	13,454,320	1,788,912	13.3%	1,177,727	8.8%	65.8%
	2	9,274,925	1,340,940	14.5%	885,178	9.5%	66.0%
	3	6,143,258	954,131	15.5%	607,736	9.9%	63.7%
	4	4,015,986	663,602	16.5%	428,681	10.7%	64.6%
	5	2,454,107	431,762	17.6%	267,503	10.9%	62.0%
	6	1,276,363	263,362	20.6%	128,557	10.1%	48.8%
	7	532,076	159,952	30.1%	96,764	18.2%	60.5%
	8	86,555	70,421	81.4%	49,055	56.7%	69.7%

In the England, Verrall and Wüthrich [3] paper, the authors discuss using the runoff of the time window standard deviations for the runoff of the capital requirement in the cost of capital method of calculating the risk margin under Solvency II.[45] While the runoff of the time window standard deviations clearly reconcile[46] with the Mack standard deviations, it does not appear as though the runoff of the time window standard deviations adhere to the time horizon concept used for Solvency II. Thus, Merz-Wüthrich would be more accurately described as a reasonable approximation for the runoff of the capital requirement.

To illustrate this issue, we start with $T = 1$ as shown in Table 4.6 and note that the first diagonal (i.e., for $k = 11$) is identical to the first diagonal in Table 3.8 since it includes both process and parameter uncertainty. The differences in the total uncertainty between Tables 4.6 and 3.8 is completely due to the remaining diagonals in Table 4.6 that only contain parameter uncertainty. This is the essence of the 1-year time horizon since the first diagonal should be an estimate of the total uncertainty and then we are concerned with estimating the change in reserves given the possible outcomes during the first year.

For the runoff of the time window, as we move to $T' = 2$ the same logic should continue to hold

[45] More specifically, the capital requirement is based on the 99.5th percentile of the 1-year time horizon unpaid claim distribution and the runoff of the capital requirement would be based on subsequent 99.5th percentiles as $T' = 1,2, \dots, N$.

[46] As shown in Table 4.2, the square root of the sum of the squares of the Merz-Wüthrich standard deviations by time window for each accident year and the total of all accident years are the same as the Mack standard deviations.

true, meaning after the first year is complete we would then want to estimate the total uncertainty for the next diagonal and the change in reserves given the possible outcomes during that second year. Following this logic, the first diagonal (i.e., for $k = 12$) in Table 4.7 should be identical to the second diagonal in Table 3.8. However, as seen in Table 4.7 the first diagonal is less than the second diagonal in Table 3.8 since it does not include all of the parameter uncertainty.

This issue can also be observed by comparing the oldest accident years for the runoff of the standard deviations in Tables 3.4 and 4.2. For example, in Table 3.4 the standard deviation for the oldest accident year when $t = 1$ is 74,931 and in Table 4.2 the standard deviation for the oldest accident year when $T' = 2$ is 60,996. Since both of these cells include only the first diagonal, the values should be the same. From the perspective of reconciling the runoff of Merz-Wüthrich with Mack this makes sense, but from the perspective of running off the required capital it does not make sense.

Another way to think about the runoff of the Merz-Wüthrich standard deviations is that they are always looking at the runoff from the perspective of the current time, or $t = 0$. From this perspective, in the second year (i.e., $T' = 2$) the first remaining diagonal (i.e., for $k = 12$ in Table 4.7) can be thought of as only containing enough uncertainty to reconcile with Mack at time $t = 0$. This perspective is also consistent with the total reserve notation in this section that does not contain a subscript implying that $t = 0$.

We can illustrate this issue with 2 other cases:

- If in the formula of α_d^T, $\hat{c}(N + T - d, d)$ is much greater than $\hat{c}(i, d)$ with i< $N + T - d$, then α_d^T tends to one and $(1 - \alpha_d^T)$ tends to 0. This would imply there is no more remaining parameter risk for $T' \geq 2$ linked to the diagonals $T' + 1$ which does not make sense in a run off of the required capital.
- We can also compare the case here with the re-reserving method or actuary in the box method (see Diers [2]). When simulating the T+1, …, T+N diagonals with the Bootstrap incrementals, a full Chain Ladder is applied, i.e., the full estimation risk is calculated even if it was already partially captured in the previous diagonal run-off.

5. TIME-HORIZON UNCERTAINY: AN ALTERNATIVE APPROACH

In order to calculate the runoff of the required capital under Solvency II, we need to revise formulas (4.5), (4.7), (4,9), and (4.10) to include all of the parameter uncertainty for the first diagonals as the reserves runoff for $t > 0$.

5.1 Uncertainty by Accident Year: X-Year Time Horizon

Starting with $T' = 2$, formula (4.5) must be revised as shown in formula (5.1),[47] except that to clearly note that we are concerned with a 1-year time horizon one year in the future the notation has also been revised to show that $t = 1$ and $T = 1$.

$$Var\left[\hat{R}_1^1(w)\right] = \hat{c}(w,n)^2 \times \frac{\hat{\sigma}_{N+2-w}^2}{\hat{F}(N+2\text{-}w)^2} \times \left\{\frac{1}{\hat{c}(w,N+2\text{-}w)} + \frac{1}{\sum_{j=1}^{w-1} c(j,N+2\text{-}w)}\right\}$$
$$+\hat{c}(w,n)^2 \times \sum_{d=n+3-w}^{n-1} \frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \left\{\alpha_d^2 \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right\} \quad (5.1)$$

Comparing formula (5.1) with formula (4.5), the one minus the weights for the first diagonal portions have been removed, but the weights for the remaining diagonals, as in formula (4.6), are still included. This formula for the second year is consistent with formula (4.1) for the first year. The generalization for $t > 1$ is shown in formula (5.2).[48]

$$Var\left[\hat{R}_t^1(w)\right] = \hat{c}(w,n)^2 \times \frac{\hat{\sigma}_{N+t+1-w}^2}{\hat{F}(N+t+1\text{-}w)^2} \times \left\{\frac{1}{\hat{c}(w,N+t+1\text{-}w)} + \frac{1}{\sum_{j=1}^{w-1} c(j,N+t+1\text{-}w)}\right\}$$
$$+\hat{c}(w,n)^2 \times \sum_{d=n+t+2-w}^{n-1} \frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \left\{\alpha_d^{t+1} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right\} \quad (5.2)$$

The generalization for $t > 1$ uses the weights as shown in formula (4.8), except that $T = t + 1$ using the new notation with both subscripts and superscripts.

5.2 Total Uncertainty: X-Year Time Horizon

The revised formula for the total uncertainty when $t = 1$ is shown in formula (5.3).[49]

$$Var\left[\hat{R}_1^1(tot)\right] = \sum_{w=3}^{N} \left\{ Var\left[\hat{R}_1^1(w)\right] \right.$$
$$\left. + 2\hat{c}(w,n)\left(\sum_{i=w+2}^{N} c(i,n)\right) \times \left[\frac{\hat{\sigma}_{N+2-w}^2}{\hat{F}(N+2-w)^2} \times \frac{1}{\sum_{j=1}^{w-1} c(j,N+2-w)} + \sum_{d=n+3-w}^{n-1}\left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \alpha_d^2 \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)}\right)\right]\right\} \quad (5.3)$$

[47] See last columns in step 8 in tab "Alternative" in the Excel file.

[48] In the Excel file (tab "Alternative"), the steps / calculations are repeated for a time horizon of $T = 1$ at each time $t = 1,2,3,\ldots$

[49] See last columns in step 8 and step 9a in tab "M&W" in the Excel file (or step 9b if exposure-adjusted data).

The generalization for $t > 1$ is shown in formula (5.4). Comparing formulas (5.3) and (5.4) with formulas (4.9) and (4.10), respectively, the one minus the weights terms have been removed similar to formulas (5.1) and (5.2).

$$Var\left[\hat{R}_t^1(tot)\right] = \sum_{w=t+2}^{N} \left\{ Var\left[\hat{R}_t^1(w)\right] + 2\hat{c}(w,n)\left(\sum_{i=w+t+1}^{N} c(i,n)\right) \times \left[\frac{\hat{\sigma}_{N+t+1-w}^2}{\hat{F}(N+t+1-w)^2} \times \frac{1}{\sum_{j=1}^{w-1} c(j, N+t+1-w)} + \sum_{d=n+t+2-w}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \alpha_d^{t+1} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] \right\} \tag{5.4}$$

Applying formulas (5.1) to (5.4) to the sample data results in the standard deviations by year as shown in Table 5.1, with the results from Table 4.1 repeated in the first column of Table 5.1.

Table 5.1 – Runoff of Alternative Estimated Standard Deviations of the Unpaid Claims

		$\sqrt{Var[\hat{R}_t^1(w)]}$									
	$t=$	0	1	2	3	4	5	6	7	8	TOTAL
w	1	-	-	-	-	-	-	-	-	-	-
	2	75,535	-	-	-	-	-	-	-	-	75,535
	3	105,309	74,931	-	-	-	-	-	-	-	129,247
	4	79,846	100,806	74,041	-	-	-	-	-	-	148,389
	5	235,115	68,535	93,353	69,186	-	-	-	-	-	271,067
	6	318,427	240,563	67,590	95,673	71,982	-	-	-	-	422,102
	7	361,089	336,607	255,033	70,558	102,361	78,029	-	-	-	574,697
	8	629,681	400,731	374,947	284,965	79,593	116,320	90,307	-	-	898,273
	9	588,662	562,933	356,774	334,233	253,564	69,171	101,939	77,826	-	993,953
	10	1,029,925	544,418	521,865	329,305	308,794	234,466	62,194	92,663	70,421	1,380,457
	CVA	*1,025,050*	*787,105*	*592,464*	*434,573*	*299,857*	*212,772*	*154,021*	*79,424*	-	*1,541,216*
	Total	**1,778,968**	**1,258,989**	**987,439**	**713,534**	**521,112**	**353,057**	**214,796**	**144,746**	**70,421**	**2,588,861**

As expected, the standard deviations runoff in a similar fashion to the estimated unpaid claims and when $t = 8$ there is no covariance adjustment term since there is only one "cell" remaining. An additional part of the results in Table 5.1 is the Total column, which is the square root of the sum of the squares of the other columns. The Total column shows that this time horizon runoff does not reconcile with the results from Mack, but that is not the intent.

The expected runoff of the unpaid claims is identical to the runoff for Mack, as previously shown in Table 3.3. Dividing the standard deviations in Table 5.1 by the means in Table 3.3 results in the runoff of the coefficients of variation shown in Table 5.2.

Table 5.2 – Runoff of Alternative Coefficients of Variation of the Unpaid Claims

		CoV									
	$t=$	0	1	2	3	4	5	6	7	8	TOTAL
w	1	-	-	-	-	-	-	-	-	-	-
	2	79.8%	-	-	-	-	-	-	-	-	79.8%
	3	22.4%	80.0%	-	-	-	-	-	-	-	27.5%
	4	11.3%	21.8%	80.2%	-	-	-	-	-	-	20.9%
	5	23.9%	10.5%	22.0%	81.8%	-	-	-	-	-	27.5%
	6	22.4%	23.2%	9.9%	21.4%	80.9%	-	-	-	-	29.7%
	7	16.6%	21.4%	22.2%	9.3%	20.7%	79.1%	-	-	-	26.4%
	8	16.1%	15.4%	19.9%	20.7%	8.8%	19.6%	76.4%	-	-	22.9%
	9	13.8%	17.3%	16.4%	21.3%	22.2%	9.2%	20.7%	79.2%	-	23.2%
	10	22.3%	14.4%	18.2%	17.2%	22.4%	23.3%	9.3%	21.4%	81.4%	29.8%
	Total	**9.5%**	**9.4%**	**10.6%**	**11.6%**	**13.0%**	**14.4%**	**16.8%**	**27.2%**	**81.4%**	**13.9%**

Adjusting formulas (5.1) and (5.2) to include the covariance adjustment related to each accident year are left to the reader. Applying formula (4.4), and the extensions for formulas (5.1) and (5.2), the runoff of the standard deviations in Table 5.1 are restated in Table 5.3.

Table 5.3 – Runoff of Alternative Estimated Standard Deviations of the Unpaid Claims

		$\sqrt{Var[\hat{R}_t^1(w)']}$									
	$t=$	0	1	2	3	4	5	6	7	8	TOTAL
w	1	-	-	-	-	-	-	-	-	-	-
	2	75,535	-	-	-	-	-	-	-	-	75,535
	3	132,910	74,931	-	-	-	-	-	-	-	152,577
	4	152,332	128,734	74,041	-	-	-	-	-	-	212,742
	5	279,093	136,650	120,436	69,186	-	-	-	-	-	340,379
	6	390,584	278,768	133,825	121,406	71,982	-	-	-	-	517,781
	7	484,763	406,147	290,998	138,420	130,333	78,029	-	-	-	725,855
	8	769,047	531,387	449,405	322,236	156,635	148,897	90,307	-	-	1,111,066
	9	800,010	695,112	485,883	409,150	291,078	149,109	137,853	77,826	-	1,287,906
	10	1,192,165	732,101	643,749	446,323	374,337	272,318	137,763	122,044	70,421	1,680,466
	Total	**1,778,968**	**1,258,989**	**987,439**	**713,534**	**521,112**	**353,057**	**214,796**	**144,746**	**70,421**	**2,588,861**

The coefficients of variation comparing the standard deviations in Table 5.3 to the expected values in Table 3.3 are shown in Table 5.4. As noted above for Table 4.1, there is a smoother transition of all CoVs from the oldest year to the most current year.

Table 5.4 – Runoff of Alternative Coefficients of Variation of the Unpaid Claims

		CoV									
	t =	0	1	2	3	4	5	6	7	8	TOTAL
w	1	-	-	-	-	-	-	-	-	-	-
	2	79.8%	-	-	-	-	-	-	-	-	79.8%
	3	28.3%	80.0%	-	-	-	-	-	-	-	32.5%
	4	21.5%	27.8%	80.2%	-	-	-	-	-	-	30.0%
	5	28.3%	21.0%	28.4%	81.8%	-	-	-	-	-	34.6%
	6	27.5%	26.9%	19.5%	27.2%	80.9%	-	-	-	-	36.5%
	7	22.3%	25.8%	25.4%	18.3%	26.4%	79.1%	-	-	-	33.3%
	8	19.6%	20.4%	23.9%	23.4%	17.2%	25.1%	76.4%	-	-	28.3%
	9	18.7%	21.3%	22.4%	26.1%	25.4%	19.7%	28.0%	79.2%	-	30.1%
	10	25.8%	19.4%	22.4%	23.3%	27.1%	27.0%	20.7%	28.1%	81.4%	36.3%
	Total	**9.5%**	**9.4%**	**10.6%**	**11.6%**	**13.0%**	**14.4%**	**16.8%**	**27.2%**	**81.4%**	**13.9%**

5.3 Cash Flow Uncertainty

Similar to the accident year formulas, cash flow formulas (4.11), (4.12), and (4.13) do not need to be revised. Moving to the formulas for $t = 1$, the formulas for the first diagonal and remaining diagonals are shown in formula (5.5).[50]

$$Var\left[\widehat{CF}_1^1(k)\right] = \sum_{j=k-N}^{N} \begin{cases} \hat{c}(j,k+1-j)^2 \times \dfrac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \dfrac{1}{\hat{c}(j,k-j)} + \dfrac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k = N+2 \\ \hat{c}(j,k+1-j)^2 \times \dfrac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^2 \times \dfrac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k > N+2 \end{cases} \tag{5.5}$$

For the total uncertainty, the formulas for the first diagonal and remaining diagonals are shown in formula (5.6).

$$Var\left[\widehat{CF}_1^1(tot)\right] = \sum_{k}^{N+n} \begin{cases} Var\left[\widehat{CF}^2(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \\ \quad \times \left[\dfrac{\hat{\sigma}_{k-N+1}^2}{\hat{F}(k-N+1)^2} \times \dfrac{1}{\sum_{j=1}^{w-1} c(j,k-N+1)} \right] \\ \quad + [\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \dfrac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \dfrac{1}{\hat{c}(j,k-j)} + \dfrac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k = N+2 \\ Var\left[\widehat{CF}^2(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \\ \quad \times \left[\sum_{d=k-N+2}^{n-1} \left(\dfrac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \alpha_d^2 \times \dfrac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] \\ \quad + [\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \dfrac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^2 \times \dfrac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ;k > N+2 \end{cases} \tag{5.6}$$

Using formulas (5.5) and (5.6), the results for the sample data triangle are shown in Table 5.5.

[50] See step 11 in tab "Alternative" in the Excel file.

Table 5.5 – Estimated Alternative Cash Flow and Standard Deviations for t=1

		$\widehat{CF}_1^1(k)$	$\sqrt{Var[\widehat{CF}_1^1(k)]}$	CoV	CVA	$\sqrt{Var[\widehat{CF}_1^1(k)']}$	CoV
k	11						
	12	4,179,394	609,716	14.6%	*1,015,053*	1,184,097	28.3%
	13	3,131,668	98,559	3.1%	*254,367*	272,794	8.7%
	14	2,127,272	87,848	4.1%	*197,935*	216,554	10.2%
	15	1,561,879	74,810	4.8%	*149,542*	167,210	10.7%
	16	1,177,744	64,972	5.5%	*115,815*	132,795	11.3%
	17	744,287	54,453	7.3%	*87,781*	103,298	13.9%
	18	445,521	45,194	10.1%	*48,293*	66,142	14.8%
	19	86,555	31,868	36.8%	-	31,868	36.8%
	CVA		*1,086,291*		*1,086,291*		
	Total	**13,454,320**	**1,258,989**	**9.4%**		**1,258,989**	**9.4%**

Comparing Table 5.5 with Table 3.8, note that the standard deviation for $k = 12$ in Table 5.5 is the same as in Table 3.6, which makes sense since formula (5.5) for the first diagonal includes all of the process and parameter uncertainty. Comparing Table 5.5 with the columns for $t = 1$ in Tables 5.1 and 5.2, the totals are identical as expected.

Like the alternative view of the covariance adjustment by calendar year for the Mack model, a portion of the covariance adjustment in formula (5.6) can be included with formula (5.5) as shown in formula (5.7).[51]

$$Var\left[\widehat{CF}_1^1(k)'\right] = \sum_{j=k-N}^{N} \begin{cases} \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\ +2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \times \left[\frac{\hat{\sigma}_{k-N+1}^2}{\hat{F}(k-N+1)^2} \times \frac{1}{\sum_{j=1}^{w-1} c(j,k-N+1)} \right] & ;k = N+2 \\ +[\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\ \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^2 \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\ +2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \times \left[\sum_{d=k-N+2}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \alpha_d^2 \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] & ;k > N+2 \\ +[\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^2 \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \end{cases} \quad (5.7)$$

This alternative view of the alternative cash flow estimates is also included in Table 5.5, starting with the column that shows the portion of the covariance adjustment "allocated" to each calendar

[51] See last columns in step 11 in tab "Alternative" in the Excel file.

year. Note that for the alternative view the CoVs exhibit a smoother transition from the first diagonal to the last diagonal similar to the Mack alternative view.

Continuing with the formulas for $t > 1$, the formulas for the first diagonal and remaining diagonals are shown in formula (5.8).

$$Var\left[\widehat{CF}_t^1(k)\right] = \sum_{j=k-N}^{N} \begin{cases} \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ; k = N+t+1 \\ \hat{c}(j,k+1-j)^2 \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^{t+1} \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} & ; k > N+t+1 \end{cases} \tag{5.8}$$

The formulas for the total uncertainty are shown in (5.9).

$$Var\left[\widehat{CF}_t^1(tot)\right] = \sum_{k}^{N+n} \begin{cases} Var\left[\widehat{CF}^2(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \\ \quad \times \left[\frac{\hat{\sigma}_{k-N+1}^2}{\hat{F}(k-N+1)^2} \times \frac{1}{\sum_{j=1}^{w-1} c(j,k-N+1)} \right] & ; k = N+t+1 \\ +[\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \frac{1}{\hat{c}(j,k-j)} + \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \\ Var\left[\widehat{CF}^2(k)\right] + 2\hat{c}(k-N+1,n) \times \left(\sum_{i=k-N+2}^{N} c(i,n) \right) \\ \quad \times \left[\sum_{d=k-N+2}^{n-1} \left(\frac{\hat{\sigma}_d^2}{\hat{F}(d)^2} \times \alpha_d^{t+1} \times \frac{1}{\sum_{j=1}^{N-d} c(j,d)} \right) \right] & ; k > N+t+1 \\ +[\hat{c}(j,n)^2 - \hat{c}(j,k+1-j)^2] \times \frac{\hat{\sigma}_{k-j}^2}{\hat{F}(k-j)^2} \times \left\{ \alpha_{k-j}^{t+1} \times \frac{1}{\sum_{i=1}^{N-j-1} c(i,k-j)} \right\} \end{cases} \tag{5.9}$$

Using formulas (5.8) and (5.9), the results for the sample data triangle will be similar to the results shown in Table 5.5, meaning the first diagonal will be equal to the same diagonal in Table 3.6 and the totals will match the same time horizon in Tables 5.1 and 5.2.

5.4 Comparison with Mack

Now that we have revised the formulas related to the Merz-Wüthrich models, it is instructive to compare the runoff for the two models using the totals from Tables 3.3, 3.4, 3.5, 5.1, and 5.2. As shown in Table 5.6, at time $t = 0$ (and $T = 1$) the standard deviation for the 1-year time horizon is 72.7% of the standard deviation for the ultimate time horizon. As previously discussed, this makes sense since the 1-year time horizon only includes the parameter variance beyond the first diagonal. In addition, the standard deviations for the last runoff period at time $t = 8$ are identical since there is no future diagonals at that point in time.

Table 5.6 – Comparison of Alternative Estimated Runoff with Mack Model

		$\hat{R}_t(tot)$	$\sqrt{Var[\hat{R}_t(tot)]}$	CoV	$\sqrt{Var[\hat{R}_t^1(tot)]}$	CoV	$Ratio$
$t=$	0	18,680,856	2,447,095	13.1%	1,778,968	9.5%	72.7%
	1	13,454,320	1,788,912	13.3%	1,258,989	9.4%	70.4%
	2	9,274,925	1,340,940	14.5%	987,439	10.6%	73.6%
	3	6,143,258	954,131	15.5%	713,534	11.6%	74.8%
	4	4,015,986	663,602	16.5%	521,112	13.0%	78.5%
	5	2,454,107	431,762	17.6%	353,057	14.4%	81.8%
	6	1,276,363	263,362	20.6%	214,796	16.8%	81.6%
	7	532,076	159,952	30.1%	144,746	27.2%	90.5%
	8	86,555	70,421	81.4%	70,421	81.4%	100.0%

In contrast to the runoff comparison with Merz-Wüthrich in Table 4.8, it does appear as though the runoff of the time horizon standard deviations adhere to the concepts used for Solvency II and, thus, is a better estimate for the runoff of the capital requirement.

5.5 Comparison of Risk Margins

As a final comparison, we can test how the different runoffs of the capital requirement affect the risk margins using the cost of capital method under Solvency II. Starting with the runoff from the Merz-Wüthrich method from Table 4.8, in Table 5.7 the lognormal distribution assumption is used to calculate the 99.5% Value at Risk (VaR). Using the VaR for each future year in the runoff, the costs of capital are calculated assuming an expected return of 6.0% and then the runoff of the cost of capital is discounted at 2.0%. Summing the discounted cost of capital over the runoff period results in a total discounted cost of capital of 891,587, which is 4.8% of the unpaid claims (i.e., 18,680,856) at $t = 0$.

Table 5.7 – Calculation of Risk Margin using Merz-Wüthrich Model

		$\hat{R}_t(tot)$	$\sqrt{Var[\hat{R}^{T\prime}(tot)]}$	99.5th Percentile	99.5% VaR	6.0% CoC	Discounted CoC
t =	0	18,680,856	1,778,968	23,753,426	5,072,570	304,354	301,328
	1	13,454,320	1,177,727	16,785,734	3,331,414	199,885	193,982
	2	9,274,925	885,178	11,799,479	2,524,553	151,473	144,092
	3	6,143,258	607,736	7,882,818	1,739,561	104,374	97,323
	4	4,015,986	428,681	5,252,966	1,236,980	74,219	67,836
	5	2,454,107	267,503	3,227,797	773,690	46,421	41,590
	6	1,276,363	128,557	1,645,023	368,659	22,120	19,425
	7	532,076	96,764	833,102	301,026	18,062	15,548
	8	86,555	49,055	293,233	206,679	12,401	10,464
	Total						**891,587**
	Percent of Unpaid Claims:						*4.8%*

In Table 5.8, the runoff using the alternative model from Table 5.6 is used to calculate the discounted cost of capital. Using all of the same assumptions noted above for Table 5.7, except for the standard deviation of the unpaid claims, the alternative model estimates the total discounted cost of capital at 1,007,157 or 5.4% of the unpaid claims at $t = 0$.

Table 5.8 – Calculation of Risk Margin using Alternative Model

		$\hat{R}_t(tot)$	$\sqrt{Var[\hat{R}_t^1(tot)]}$	99.5th Percentile	99.5% VaR	6.0% CoC	Discounted CoC
t =	0	18,680,856	1,778,968	23,753,426	5,072,570	304,354	301,328
	1	13,454,320	1,258,989	17,038,055	3,583,735	215,024	208,674
	2	9,274,925	987,439	12,123,409	2,848,484	170,909	162,580
	3	6,143,258	713,534	8,222,165	2,078,907	124,734	116,308
	4	4,015,986	521,112	5,555,442	1,539,456	92,367	84,424
	5	2,454,107	353,057	3,512,025	1,057,918	63,475	56,868
	6	1,276,363	214,796	1,935,777	659,413	39,565	34,745
	7	532,076	144,746	1,021,830	489,754	29,385	25,295
	8	86,555	70,421	421,013	334,458	20,067	16,933
	Total						**1,007,157**
	Percent of Unpaid Claims:						*5.4%*

Comparing Table 5.7 and 5.8, it makes sense that the risk margin is larger for the alternative method since the runoff is a bit slower.[52] While there could be situations where the alternative method results

[52] In this example, the risk margin is 13.0% larger but other examples could result in larger or smaller differences between the models.

in a faster runoff and a smaller risk margin, it seems like the most common result would be for the alternative method to result in a larger risk margin. In other words, in most situations the risk margin would be underestimated using the Merz-Wüthrich approximation for the runoff.

6. CONCLUSIONS

After reviewing the Mack and Merz-Wüthrich model formulas, the paper expands their usefulness by adding runoff and cash flow formulas. By comparing the runoff of the Merz-Wüthrich results to the Mack runoff it was demonstrated the Merz-Wüthrich does reconcile with the Mack in the sense that the variances of the time windows total to the Mack variances for the ultimate time horizon and is consistent in the context of a $t = 0$ view. However, to estimate the runoff of the required capital for the cost of capital method of calculating the risk margin under Solvency II, this formula would underestimate the risk when we consider the view at $t > 0$. In order to estimate the risk for $t > 0$, the first future year for each runoff period must include both the full process and parameter variance. Thus, an alternative set of formulas were derived and demonstrated to be consistent with concepts used for Solvency II. Finally, alternate views of the covariance adjustment were developed for all of the formulas that result in a smoother transition of the coefficients of variation and aide in comparisons to other models.

Acknowledgment

The author gratefully acknowledges the many authors listed in the References (and others not listed) that contributed to the foundation of the Mack and Merz-Wüthrich models, without which this research would not have been possible. He would like to thank all the peer reviewers, in particular Jeff Courchene, Roger Hayne, Pierre Miehe, Flavien Thery, and Louise Francis, and is grateful to the CAS referees, Isabelle Guerard, Ernest Wilson, and Chandrakant Patel, for their many comments that greatly improved the quality of the paper.

Supplementary Material

There are companion files designed to give the reader a deeper understanding of the formulas discussed in the paper and that were used to calculate all of the tables in this paper. The files are all in the "Mack & Merz-Wüthrich.zip" file. The files are:

Mack & Merz-Wüthrich Runoff.xlsm – this file contains the detailed calculations described in this paper for a single segment or line of business for a 10 x 10 triangle only. Data can be entered for a new triangle, exposures, a tail factor, and tail standard deviation.

Mack & Merz-Wüthrich Calc.xlsm – this file contains VBA functions that replicate all the calculations in the "Runoff" file for a segment or line of business for any size triangle. Data can be entered for a new triangle, exposures, a tail factor, and tail standard deviation.

Milliman Claim Variability Benchmarks (CVB) Functions – the algorithms in the paper have been developed into functions in the free version of the Milliman CVB Excel Add-In. If you are interested, you can request the free version by sending an email to actuarialsoftware@milliman.com.

Milliman Mind Application – the algorithms in the paper have been developed in a Milliman Mind app. If you are interested, you can request a free trial by sending an email to europeansoftware@milliman.com.

REFERENCES

[1] 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.

[2] Diers, Dorothea. 2019. "Stochastic re-reserving in multi-year internal models – An approach based on simulations." *ASTIN Colloquium*. 39.

[3] England, Peter D., Richard J. Verrall and Mario V. Wüthrich. 2018. "On the Lifetime and One-Year View of Reserve Risk, with Application to IRFS 17 and Solvency II Risk Margins."

[4] Foundations of Casualty Actuarial Science, 4th ed. 2001. Arlington, Va.: Casualty Actuarial Society.

[5] Mack, Thomas. 1993. "Distribution Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." ASTIN Bulletin 23-2: 213-225.

[6] Mack, Thomas. 1999. "The Standard Error of Chain Ladder Estimates: Recursive Calculation and Inclusion of a Tail Factor." ASTIN Bulletin, 29, 2, 361-366.

[7] Merz, Michael, and Mario V. Wüthrich. 2008. "Modeling the Claims Development Result for Solvency Purposes." Casualty Actuarial Society E-Forum (Fall): 542-568.

[8] Merz, Michael, and Mario V. Wüthrich. 2015. "Claims Run-off Uncertainty: The Full Picture." Swiss Finance Institute Research Paper No. 14-69. https://ssrn.com/abstract=2524352.

[9] Taylor, Greg and Frank Ashe. 1982. "Second Moments of Estimates of Outstanding Claims." *ASTIN Colloquium*. 23.

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

Abbreviations and notations

The abbreviations and notations used in the paper are listed here in alphabetical order.

CL, chain ladder
CoV, coefficient of variation
CVA, covariance adjustment (in the Excel file this is labeled as "CV Adj")
MSE, mean squared error of prediction
VaR, Value at Risk

Biography of the Author

Mark R. Shapland 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 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, the European Actuarial Academy's Stochastic Modelling Seminar, and has spoken frequently on this subject both within the CAS and internationally. He can be contacted at mrshapland@netzero.com.

Updating Increased Limits Factors for Trend Using Interpolation Along a Curve

Joseph A. Boor, FCAS, CERA, Ph.D.

Abstract: The process of updating increased limits factors is very complex, data-intensive, analysis intensive, and labor intensive. So, may entities do not update their increased tables very often. While this article does not simplify that process, it does posit a simplified process for simply updating an increased limits table for trend. All it requires is the increased limits table (not the data behind it) and the trend rate, and consequently it may be executed in much less time and much more simply than the existing method. The core of this method is the use of interpolation along the curve to identify the relativities of limited loss costs at detrended policy limits. Those relativities then provide the basis for trend-adjusted increased limits factors.

Keywords: increased limits factors, interpolation along a curve, trend

---

1. INTRODUCTION

The current process for updating increased limits factors (ILFs) typically requires significant activity and effort. Losses must be trended and developed individually. Preferably the losses should not just be developed but rather developed using appropriate random development factors (per Boor 2017 and Couret 1997). Then, the losses must be capped at the various layers and summarized. Because of this, many increased limits tables are not reviewed very frequently. For example, at least one rating bureau does not typically review increased limits tables annually, even though many insurers use the tables. So, the amount of effort required results in the ILFs being less current.

While this paper does not simplify the process of performing a full review, it does provide a less labor-intensive method for updating ILFs for trend. The key is to replace the complex process of applying actuarial adjustments to individual claims, then collating and layering the results, with a more basic process where one simply uses the relative values of points on a curve.

Sometimes actuaries will simply fit a specific curve from some curve family by minimizing the error against the current ILFs, or choose a curve matching a characteristic or two of the current ILFs, then estimate the ILFs using values from that curve. No matter how the curve is selected, the resulting curves rarely match the ILFs exactly. Therefore, while a fitted curve is still prescribed, in this article the adjustments in the "interpolation along the curve" process from Boor 2014, are used to create an exact match to the current ILFs, while still providing the benefits of a fitted curve.

For the last step, a cumulative trend of T (trend factor of 1+T), between the effective date of the existing table and the prospective effective date of the revised table is used to detrend each limit. For example, if T=25%, and one is detrending the $500,000 limit, one would use a limit of $500,000/1.25 = $400,000 in prior uninflated dollars to match the $500,000 limit in costs in the

prospective period of the updated table. Then the relativities of the costs in those detrended limits to the losses capped at the detrended basic limit will form the new ILFs.

While that provides an overview of the process, there are important concerns and details for each step that must be discussed. That is done in the sections that follow.

2 THE INTERPOLATION ALGORITHM FOR UPDATING ILFS

If one can determine that the same loss trend applies across all loss sizes, then a shortcut process may be used to update increased limits factors for trend. Essentially, this approach involves using interpolation along a Pareto-based[1] curve in order to estimate the ILFs at detrended limits. Then, the results are rebalanced so that the detrended basic limit stands in for the basic limit. For example, if the loss trend is 10%, and the basic limit is $100,000, then losses of size $100,000÷1.10=$90,909 and above will be over $100,000 after trending. Since $90,909 is not in the limits table, either interpolation along the curve or a recalculation of increased limits factors beginning with raw data must be performed to update the increased limits factors. Similarly, the losses that will trend up to $250,000, were capped at $250,000÷1.10=$227,272 in the data. This merely uses interpolated values, albeit values from a generally very effective interpolation process, to estimate what the impacts (and revised ILFs) would be if the old data were adjusted using the trend between the effective date of the current table and the effective date of the updated table being developed.

This approach does not involve claim-by-claim detail. As noted earlier, it is presented because it involves significantly less labor than the recalculation from scratch. The table following shows the process.

[1] The Pareto is used here because it is common in the actuarial literature and known to have significant skew. If some other curve family fits the table, especially the upper ILFs, better, then it could conceivably be used instead.

Table 1: Trending of ILFs Using Interpolation Along the Curve

Part 1-Curve Fit to Current ILFs

First Step: Solver Setup/Curve

α **0.84** Lt Gray Values Solved

x_M=Truncation Pt. **200,000** to Minimize Dark Gray

$200K is upper restriction]

(1)	(2)	(3)	(4)
Table 1 c.1	Table 1 c.1 (Offset)	per new curve ****	((3)-(2))^2
Top of Layer	Current ILF	Fitted Pareto	Squared Error
0	0.000	0	
$250,000	1.000	1.000	0.00000
500,000	1.600	1.618	0.00033
1,000,000	2.500	2.308	0.03688
2,000,000	3.500	3.077	0.17867
$5,000,000	4.000	4.232	0.05401
			0.26989

Part 2- Detrended Limits/Trended ILFs

(5)	(6)	(7)	(8)
(100)/1.100	per curve	*****	(7)/[250K(7)]
Top of Layer Detrended at 10.0%	Fitted Pareto	Detrended ILFs Interpolated Along the Curve	Rebased to 250K Final Trended ILFs
-	0	0.000	0.000
227,273	0.920	0.920	1.000
454,545	1.529	1.514	1.645
909,091	2.209	2.370	2.576
1,818,182	2.966	3.356	3.647
4,545,455	4.104	3.945	4.287

Notes:

**** Fitted Pareto values are

{α - [(x_M/[Column (1)]) ^(α-1.0)]} /{α - [(x_M /250,000) ^(α-1.0)]}

"***** Values per Interpolation =

previous (2) + [(2)-previous (2)]*[(6) - previous (6)] /[(3)- previous(3)]

This interpolation process is identical to the process in Boor 2014. One begins with the ILFs in column 2). Then, one fits a limited expected loss curve corresponding to a Pareto distribution to those values. Given an α parameter for the general shape or decay of the distribution and a truncation point x_M, the limited expected mean would be α - [(x_M/Limit) ^(α **-1.0**)]. Then the formula for the basic limit is identical, excepting that the basic limit is used in the formula rather than the limit in question. In this case, initial but provisional values for α and x_M were selected arbitrarily, but the software[2] was directed to find the values of those that minimized the squared differences between the fitted curve and the ILFs. To insulate the basic limit from the truncation point, a maximum truncation point of $200,000 was selected judgmentally**.**

Once the curve is fit, it is easy to interpolate along it. Linear interpolation is the simplest. When one does linear (along a line) interpolation, one looks at how much "y" increases (from y_1 to y_2) as "x" goes from x_1 to x_2. Then, for x between x_1 to x_2, the estimate of the corresponding value y of begins with y_1. Next, one must consider how far the line increases between y_1 and y_2 (specifically, y_2 - y_1), how far the input x changes over the interval (x_2 - x_1) and how far in the input range one has moved. Combining those, the linear estimate is

$$est(y\ corresponding\ to\ x) = y_1 + (y_2 - y_1) \times \frac{(x - x_1)}{(x_2 - x_1)},$$

which means that the estimate changes proportionally to changes in x. In interpolation along a curve, the changes are proportional to changes in the fitted reference curve "f", or

$$est(y\ corresponding\ to\ x) = y_1 + (y_2 - y_1) \times \frac{(f(x) - f(x_1))}{(f(x_2) - f(x_1))}.$$

This procedure is designed to capture much of the shape in the curve inherent in the underlying increased limits/limited expected value function. Thus, as long as the curve family is an acceptable match to that function, it should generally provide reasonably accurate estimates of the ILF values[3].

In any event, that explains the interpolation process to column 7) of Table 1, where interpolation along the curve is applied to estimate the ILF values at the detrended limits in column 5). The last step is to simply rebalance the ILFs so that the ILF for the basic limit is unity (1.0). Those calculations provide the updated ILF table.

3. ADAPTING THE PROCESS FOR UNEVEN TREND

The procedure of the previous section assumes that the trend rate is the same for all loss sizes. However, that may not always be the case. In certain cases, there may be a rolling thunder of changes in the way courts interpret policies combined with attempts to offset that with law changes. Sometimes, those will, by their nature, affect average or "run-of-the-mill" claims more than large claims. Sometimes there may be some emerging issue (environmental impairment, for example) that

[2] The author's understanding is that many common spreadsheet programs will do this.

[3] The 2014 article by Boor provides some testing of this against various alternate approaches to interpolation, and generally suggests that it tends to provide more accurate estimates than the alternatives tested.

only generates large claims. So, some discussion of how to adapt the previous process may be helpful.

One way to resolve either scenario involves an additive load that would be added or subtracted from the various ILFs. Consider the situation where an external change adds additional costs, but only to the lowest (in this case conveniently the basic) layer. Then, if the change increases (or is expected to increase) the losses in the basic layer by 20%, one may simply add .2 to all the computed ILFs, then rebalance. For example, in the case of the basic and $5,000,000 final ILFs in Table 1, the basic limit would go to 1.2 and the $5,000,000 ILF would go to 4.487. Rebalancing the basic layer ILF to 1.000 would result in an ILF of 3.739=4.487/1.2 for the $5,000,000 limit. If the lowest limit is less than the basic limit, one would simply multiply the percentage by which losses in that lower layer change by the ILF for that layer to produce the additive load.

The process may be slightly more complicated when the large losses are disproportionately affected by new large claims. Say, for example that there is a new class of large claim that increases the number of large claims, uniformly insofar as the actuary can determine, by 50%. Then the results of Table 1 should be modified to reflect that. In this case the lowest limit nonetheless contains a portion of the costs for the large losses (the first $L of each loss, when the lowest limit is $L). So, it is necessary to determine the percentage of the losses at the $L limit that are actually just the lower portion of large losses. Say it is 10%. Note also that ILFs are unaffected by frequency. So, this may be regarded not as how much the large losses increased, but rather as how much the lower losses did not. This means that the additive correction for each ILF will now be a subtraction of the ILF(L)×(1.0-10%)×$\frac{\mathbf{50\%}}{\mathbf{150\%}}$, or the ILF for L times 30% from each ILF, followed by a rebalancing.

Those two examples illustrate a couple of key processes. All sorts of different situations might occur in practice, but these two illustrate key concepts that should form a starting point for an analysis.

4. SUMMARY

Interpolation along the curve offers a convenient way to update increased limits factors for trend. Hopefully this will allow actuaries to maintain more current increased limits tables.

REFERENCES

1) Boor, Joseph A., **'Interpolation Along a Curve',** *Variance,* Casualty Actuarial Society, Arlington, Virginia 2014: *Vol. 8 Issue 1 pp. 9-22*
2) Boor, Joseph A., **'Unbiased Development for Individual Claims - Taming the Wild Burning Cost',** *E-Forum,* Casualty Actuarial Society, Arlington, Virginia 2017: *Spring Issue, Vol. 2 pp. 1-28*
3) Couret, Jose R., **'Retrospective Rating: 1997 Excess Loss Factors',** *Proceedings of the Casualty Actuarial Society,* Casualty Actuarial Society, Arlington, Virginia 1997: *Vol. 84 pp. 450-481*

Over-Dispersion and Loss Reserving

Marco De Virgilis

Abstract: The intent of this paper is to underline the importance of testing the validity of the assumptions behind the common over-dispersed models.
Very often, these frameworks are applied blindly –without a proper assessment of the hypothesis behind the models. This could lead to inaccurate results or models being applied to data for which they are not suited.
Over-Dispersed stochastic frameworks are presented from a theoretical point of view, explaining all the assumptions and the hypothesis upon which they are built.
Practical examples are also presented and analyzed, showing the empirical consequences and implications of the different modeling choices.

Keywords. Generalized Linear Models; Over-Dispersion; Reserving; Tweedie; Stochastic Modeling; Bootstrap

1. INTRODUCTION

Stochastic reserving frameworks are, nowadays, very common and implemented in a fairly simple way. Several software solutions (either open source or proprietary) allow the user to perform such calculations without having the inconvenience of manually writing or designing the whole statistical procedure.

In the current market and regulation environment there is also considerable interest (COPLFR, 2018) in stochastic reserving and unpaid claim distributions. It follows that it is of paramount importance that claim liability estimates are produced by sound actuarial and statistical procedures. Failing in doing this could lead to inaccurate results that could affect company risk management or even the entire ability of being able to meet liability payments towards policyholders.

There is also a very wide and well-made selection of papers (Shapland, 2016) that describe the Over-Dispersed process. These, however, are either from a practical implementation perspective or focused on a narrow aspect of the process. In contrast, this paper, has the objective of providing a sound theoretical and statistical foundation to allow the reader to appreciate how an Over-Dispersed modeling framework works, and how to modify it and adapt it to each specific situation. The latter objective will be particularly stressed and a reproducible case study will be shown; in fact, given the ease of implementation provided by several statistical packages available, such modeling frameworks are often applied blindly, or without a thorough and proper review of the main underlying assumptions.

1.1 Research Context

In the actuarial context, there have been several studies and contributions towards stochastic claim

reserving and reserve uncertainty estimation (Renshaw et al., 1998). This paper falls into such category, but it has the objective to present results with a deeper focus on the theoretical foundations of the methodologies.

Statistical concepts will be presented and formally addressed, so the reader can also have an understanding of why so-called Over-Dispersed models work, and not only a practical guide on how to implement them.

1.2 Objective

The main objective of this paper is to present and prove statistical concepts behind the so-called Over-Dispersed Models. The reader will be introduced to formal definitions of the models and to tests that the hypothesis behind the model are met.

Moreover, having a deeper knowledge of the statistical structure of such models, the reader will be able to modify the main assumptions to adapt the framework to their specific needs.

1.3 Outline

The structure of this paper is as follows:

Section 2 presents the terminology and notations used throughout the paper.

Section 3 presents the theoretical background for Over-Dispersed models.

Section 4 is dedicated to the process of model definition.

Section 5 presents the process of model fitting.

Section 6 analyzes the issue of model validation.

Section 7 explores the implication of assumptions with regards to curve fitting.

Section 8 states some conclusions.

2. TERMINOLOGY AND NOTATION

This paper will make extensive use of specific terminology and notation. In this section we will present the main ideas from a theoretical point of view in order to fix the concepts, and then all the necessary mathematical notations will be introduced.

As already described, the paper is focused on producing distributions for unpaid claim reserves. Claim reserves are defined as the amount of money that the company set aside to pay policyholder claims. It follows that this definition is strictly linked to the concept of ultimate claim amounts.

Ultimate claim amounts are defined as the total amounts that the company is expecting to pay for all the claims relative to a specific cohort. Claims are, in fact, usually aggregated by accident years. Once ultimate claims have been produced, claim reserves can be calculated as the difference between ultimate claims and claims already paid.

To visualize these ideas, we can construct a run-off triangle:

		d				
		0	1	2	3	4
w	0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	$C_{0,3}$	$C_{0,4}$
	1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	
	2	$C_{2,0}$	$C_{2,1}$	$C_{2,2}$		
	3	$C_{3,0}$	$C_{3,1}$			
	4	$C_{4,0}$				

Table 1: Cumulative Run-off Triangle

This representation allows us to group the claims by origin periods, w, and summarize them by development periods, d. Each cell $C_{w,d}$ contains the total amount of claims from origin period w, evaluated as of development period d.

If we consider the incremental amounts $q_{w,d}$, what we need to estimate is the sum of the incremental transaction in the lower part of the triangle $\{q_{w,d} : w + d > n + 1\}$, where n represents the total numbers of accident years.[1]

		d				
		0	1	2	3	4
w	0					
	1					$q_{1,4}$
	2				$q_{2,3}$	$q_{2,4}$
	3			$q_{3,2}$	$q_{3,3}$	$q_{3,4}$
	4		$q_{4,1}$	$q_{4,2}$	$q_{4,3}$	$q_{1,4}$

Table 2: Incremental Lower Run-off Triangle

[1] We are assuming that we are dealing with a square triangle: $w = d = n$.

3. BACKGROUND AND METHODS

We will consider the amounts $\{q_{w,d} : w + d > n + 1\}$ as independent and identically distributed random variables and we will estimate them using the framework of Generalized Linear Models.

3.1 Generalized Linear Models

Generalized Linear Models (GLMs) are specific formulations of regression models and therefore consist of two main components: a stochastic component and a deterministic component.

The stochastic component addresses the randomness in the data, i.e. the probability distribution of the response variable, or the relation between the mean and the variance.

The deterministic component of the model indicates the connection between the predictors and the mean of the response variable.

3.1.1 The Stochastic Component: Exponential Dispersion Models

The GLM formulation assumes that the response variable belongs to a set of probability distributions called Exponential Dispersion Model (EDM) class. This class represents a generalization of the Natural Exponential Family.

The probability distribution function belonging to the EDM family takes the following form:

$$\pi(y;\theta,\phi) = a(y,\theta)\exp\left(\frac{y\theta - b(\theta)}{c(\phi)}\right), \tag{3.1}$$

Where:

y is the value of the response variable,

θ is the canonical parameter,

$\phi > 0$ is the dispersion parameter,

$a(.)$ is a normalizing function. This ensures that $\int \pi(y;\theta,\phi) = 1$,

$b(.)$ is the cumulant function,

$c(.)$ is an arbitrary function.

The mean and the variance of a random variable belonging to an EDM can be found as:

$$E[Y] = \mu = \frac{\partial b(\theta)}{\partial \theta}, \tag{3.2}$$

$$Var[Y] = c(\phi)\frac{\partial b^2(\theta)}{\partial \theta^2} = c(\phi)\frac{\partial}{\partial \theta}\left(\frac{\partial b(\theta)}{\partial \theta}\right) = c(\phi)\frac{\partial \mu}{\partial \theta} \tag{3.3}$$

It is then possible to define the variance function:

$$V(\mu) = \frac{\partial \mu}{\partial \theta}, \tag{3.4}$$

From which follows: [2]

$$Var[Y] = c(\phi)V(\mu) \tag{3.5}$$

3.1.2 The Deterministic Component: the Link Function

The GLM definition assumes that the mean and the linear predictors are linked together through a link function $g(.)$ such that $\mu = g^{-1}(\eta)$.

The linear combination of the predictors η, is defined as:

$$\eta = \beta_0 + \sum \beta_i x_i \tag{3.6}$$

The link function $g(.)$ is a monotonic, differentiable function that relates the fitted values μ to the combination of the linear predictors η.

The requirements of monotonicity ensures that the function between the values of η and μ is an injective function and therefore invertible, i.e. η and μ are in a one-to-one correspondence.

The requirements of differentiability ensures that the coefficients β_i can be estimated.

3.1.3 Parameter Estimation: Maximum Likelihood

The intuition of maximum likelihood arises from the concept of selecting the unknown parameters such that the probability density of the observed data is maximized.

Let's assume that we have a series of data points $x_1, x_2, \dots, x_n$ that are independent and identically distributed according to a specific distribution $P(x;\theta)$ and we need to estimate the parameter θ.

It is possible to compute the joint probability distribution as:

$$P(x_1, x_2, \dots, x_n; \theta) = \prod P(x_i; \theta) \tag{3.7}$$

The parameter θ can be found by maximizing the previous function, either by differentiation or by any other numeric algorithm (e.g. Newton-Raphson or Fisher Scoring algorithm).[3]

If we now take into account that $y_i \sim EDM(\mu_i, \phi/\omega_i)$,[4] $\mu = g^{-1}(\eta)$ and the equation (3.6) for η,

[2] For the remainder of the paper we will assume that $c(.)$ is the identity function, i.e. $c(\phi) = \phi$.
[3] Typically, the function that is maximized is the *log-likelihood* not the standard *likelihood*.
[4] Equation (3.1) has 2 parameters, θ and ϕ, however, here, we have μ and ϕ, because μ can be expressed as a function of the canonical parameter θ through equation (3.2).

we can obtain all the derivatives $\frac{\partial \log P(y;\, \mu,\phi/\omega)}{\partial\, \beta_j}$ and find the estimates of the model coefficients β_j. Here ω_i represent the known weight assigned to observation i.

3.2 Quasi-distributions

The previous dissertation is based on the assumption that the data $x_1, x_2, \dots, x_n$ is distributed according to an EDM. Sometimes, however, it is not possible to make this assumption, and we do not have a specific shape for the data distribution.

In order to overcome this lack of information we can still make some assumptions regarding the data that will allow us to specify a model. For our purposes, we will specify a relation between the mean of the data and the variance.

In particular we will assume that:

$$Var[Y] = \phi V(\mu) = \phi \mu^{\xi}\,, \tag{3.8}$$

Where:

$\phi > 0$ is the dispersion parameter,

ξ is a power parameter that specify the relation between the mean and the variance.

3.2.1 Quasi-likelihood

In this context we cannot estimate the parameter β_j following the method of maximizing the likelihood function. This is because we do not have a complete probability distribution but just a relation between the mean and the variance of the data.

In order to overcome this problem we then define the quasi likelihood function as:

$$Q(y, \mu) = \int_x^{\mu} \frac{y - t}{V(t)} dt\,, \tag{3.9}$$

Where:

y represents the actual observation,

μ represents the expectation of the observation y,

$V(t)$ is the variance function, equation (3.4).

Defining the quasi-likelihood in this way leads to properties similar to those of the standard log-likelihood functions. It can therefore be used to estimate the parameters of a linear model β_j.

This can be achieved by re-writing the term μ considering the link function $g(.)$ such that $\mu = g^{-1}(\eta)$ and the set of parameters $\beta_1, \beta_2, \dots, \beta_n$ that define η as shown in equation (3.6).

3.2.1 Distribution Examples

Setting the power parameter ξ equal to 0, 1, 2 or, 3 and solving equation (3.8) will lead to quasi-likelihood functions that are equal to standard log-likelihood functions.

In particular, we have the following results:

$\xi = 0$ leads to the Normal Distribution,

$\xi = 1$ leads to the Poisson Distribution,

$\xi = 2$ leads to the Gamma Distribution,

$\xi = 3$ leads to the Inverse Gaussian Distribution.

Here we will prove such relationship for the Poisson distribution:

Let's start by setting $\xi = 1$ and therefore $V(\mu) = \mu$.
If we now solve the equation (3.8) we will find that:

$$Q(y,\mu) = \int_y^{\mu} \frac{y-t}{V(t)} dt = y log\mu - \mu \qquad (3.10)$$

We recognize that the expression $ylog\mu - \mu$ is the log-likelihhod function of a standard Poisson distribution.

The main difference is that for a standard Poisson distribution we have that $Var[Y] = E[Y]$, whereas here we have that $Var[Y] = \phi E[Y]$, which gives us the possibility to allow for more variability.

The parameter ϕ, in this context, is often called the over-dispersion parameter, hence Over-Dispersed Poisson distribution.[5]

[5] In a standard context we would have $\phi = 1$, which leads to $Var[Y] = E[Y]$.

Here a summary table:

$V(\mu)$	$Q(x,\mu)$	**Distribution Name**
1	$-\frac{(y-\mu)^2}{2}$	Normal
μ	$ylog\mu-\mu$	Poisson
μ^2	$-\frac{y}{\mu}-log\mu$	Gamma
μ^3	$-\frac{y}{2\mu^2}+\frac{1}{\mu}$	Inverse Gaussian
μ^{ξ}	$\mu^{-\xi}\left(\frac{\mu y}{1-\xi}-\frac{\mu^2}{2-\xi}\right)$	—

Table 3 Quasi-likelihoods

3.2.1 Limitations of quasi-likelihood

We have seen that by introducing the concept of quasi-likelihood, we can relax the assumption of EDM and fit GLMs to data that do not necessarily follow a formally defined probability distribution.

This practice, however, has some limitations. In fact, it won't be possible to compute any metrics that are directly linked to the log-likelihood value or carry out any analysis based on the probability distribution of the data.

An example from the first context would be the Akaike's Information Criterion (AIC). This metric is an estimator of the prediction error and the quality of a statistical model. It can be calculated as: $2k-2l$, where k represents the number of parameters in the model and l is the log-likelihood.

A drawback of not having the probability distribution of the response variable formally defined will not allow us to simulate responses from the model. This process if often useful to investigate and further analyze the distribution of each individual points in the dataset.

3.3 Tweedie distributions

The Tweedie distributions, called after M. Tweedie, are a family of distributions, belonging to the class of the EDM, with a defined and interesting relationship between the mean and the variance.

A random variable Y follows a Tweedie distribution if $y_i \sim EDM(\mu_i, \phi/\omega_i)$ with $E[Y]=\mu$ and $Var[Y]=\phi\mu^{\xi}$. From this relation it also follows that $V(\mu)=\mu^{\xi}$.

The Tweedie distributions are defined for $\xi \in \mathbb{R}$, with exclusion of $0 < \xi < 1$.[6]

We recognize that the Tweedie distributions are defined in a very similar way in which we specified the mean-variance relationships for quasi-distributions at equation (3.8), having, however, a formally stated probability distribution and likelihood function.

For the stated reasons, Tweedie densities are often a very good alternative in situations where we might have otherwise used quasi-distributions.

Another interesting aspect of the Tweedie distributions is that by changing the parameter ξ, sometimes called Tweedie index parameter, we can find the shape of familiar distributions.

Here a summary table:

ξ	**Tweedie Distribution**
$\xi = 0$	Normal
$\xi = 1$	Poisson
$1 < \xi < 2$	Compound Poisson - Gamma
$\xi = 2$	Gamma
$\xi = 3$	Inverse Gaussian

Table 4 Tweedie distributions

3.3.1 Tweedie GLM

The Tweedie distributions can be written in the terms of an EDM as specified by equation (3.1), and following the procedure of maximum likelihood estimation, section 3.1.3, it is possible to derive the coefficients of a GLM defined in terms of a Tweedie distribution (Dunn et al., 2001).

A big advantage is that we could overcome the issues explained at section 3.2.1. It is of particular interest, especially in the actuarial practice, the second point, i.e. it is possible to simulate target response variables.

It is also important to note that in order to define a Tweedie GLM model we need the EDM structure to be fully defined, i.e. the parameter ξ has to be known a priori. This, however, is not always

[6] For the remainder of the paper we will not cover the case in which $\xi < 0$.

the case as the parameter ξ has to be estimated from the data.

The process could be done through a grid search by checking the value of the likelihood for each value of ξ. Various index parameters are chosen, the GLM is fitted and the likelihood is computed: the value of ξ that leads to the highest value of the likelihood is the final estimate and it is the value that should be chosen to fit the final GLM.

A similar procedure cannot be followed when we define a model in terms of a quasi-distribution, in this case the parameter ξ has to be chosen prior to fitting the model.

4. MODEL DEFINITION

We will describe in this section how to set up the two modeling frameworks described, one based on quasi-distributions and one based on Tweedie GLM. The target, as explained in section 2, is to predict and estimate the future incremental claims.

We can therefore set up a model that takes the following form:

$$E[q_{ij}] = \mu_{ij} = exp(\eta_{ij}), \tag{4.1}$$

$$Var[q_{ij}] = \phi\mu_{ij}^{\xi} = \phi V(\mu_{ij}), \tag{4.2}$$

$$\eta_{ij} = \alpha_i + \beta_j, \tag{4.3}$$

Where:

$i = 1, \dots, n$ represents the Accident Year (indicated with w in section 2),

$j = 2, \dots, n$ represents the Development Period (indicated with d in section 2),[7]

$\phi > 0$ is the dispersion parameter,

ξ is the index parameter that specify the relation between the mean and the variance,

α_i are the GLM coefficients for each individual Accident Year i,

β_j are the GLM coefficients for each individual Development Period j,

exp is the exponential function i.e. the link function $g(.)$ is the natural logarithmic function.

In order to fully define the GLM model we also need to specify a probability distribution for the random variable q_{ij}, i.e. the stochastic component of the model.

According to the framework presented in section 3.2 we will define q_{ij} as belonging to a quasi-distribution with parameters μ_{ij}, ϕ and ξ, whereas according to the framework presented in section

[7] The indices j for the Development Periods start at $j = 2$ to avoid the problem of multicollinearity.

3.3 we will define $q_{ij} \sim Tweedie(\mu_{ij}, \phi, \xi)$.

It is important to note that in both of the frameworks, the parameters that define the stochastic component of the model are exactly the same. The only difference is the definition of the probability distribution and hence the technique to estimate such parameters.

4.1 Model Output

The output expected from the model is the series of the GLM coefficients α_i and β_j, such that $E[q_{ij}] = exp(\alpha_i + \beta_j)$, alongside the dispersion parameter ϕ. The power index ξ is an input to the model.

Once we have such coefficients, we could estimate the lower triangle and then, by summing the individual estimates, we could obtain the amount of $IBNR$.[8] We are, in fact, dealing with incremental amounts.

A more interesting application, rather than the point estimates themselves, is that we could build a probability distribution of the model output and, therefore, a distribution of the $IBNR$.

This process can be carried out in two different ways according to the definition of the model at the beginning of this section –either based on quasi-distributions or based on the Tweedie distributions.

4.1.1 Bootstrap

If we have defined the model based on quasi-distribution, we will follow a bootstrap procedure in order to compute the probability distribution of the $IBNR$. Several texts have been written on this subject (Shapland, 2016), so we will not further investigate or describe this procedure.

We will only point out that, since the whole procedure is based on the model residuals, various practical issues could arise and adjustments are often needed, e.g. heteroscedasticity or non-zero sum of residuals.

The main reason why this is the case is because we do not have a formally defined probability distribution from which we could sample with a standard Monte Carlo methodology.

4.1.2 Tweedie sampling

If, instead, we have defined the model based on the Tweedie distribution, we have a formally defined probability distribution from which we could sample.

Recall that each individual observation $q_{ij} \sim Tweedie(\mu_{ij}, \phi, \xi)$ and that we have the estimates

[8] For the remainder of the paper we will be using the term $IBNR$ or Reserve interchangeably.

μ_{ij} obtained through the GLM process. At this point, in order to compute the samples $\hat{q}_{ij}$ it will be sufficient to sample from the Tweedie distribution with the estimated parameters. For each value q_{ij} we will have the full sampled probability distribution that will allow us to compute the probability distribution of the $IBNR$.

This procedure has the advantages of not having to deal with the model residuals and therefore overcome several of the issues of the standard bootstrap procedure.

4.2 Estimating ξ

In all our previous discussion, for either the model based on the quasi-distributions or the model based on the Tweedie distributions, we assumed that the parameter ξ is known and used as an input for the model.

This in not necessarily true all the time. However, setting a priori the value of the parameter ξ can lead to interesting results. A very important result in the actuarial field is observed when we set the parameter equal to 1.

In this case we will have that $Var[q_{ij}] = \phi\mu_{ij}$, i.e. the variance of the estimate is linearly proportional to the mean through the coefficient ϕ.

The advantage of having this relation is that the estimates μ_{ij} will be the same as the estimates obtained with a standard Chain-Ladder technique (Renshaw, 1998).

These results can simplify the calculations of the GLM and can extend the bootstrap procedure, as described in section 4.1.1, to the standard Chain-Ladder calculations (Shapland, 2018).

A more rigorous approach, however, will be to estimate the parameter ξ from the data and modify the model accordingly. This could be achieved as explained in section 3.3.1.

Such procedure will allow us to define a Tweedie distribution that is more representative of the data under investigation and, therefore, having a more accurate simulation process, section 4.1.2.

5. MODEL FITTING

In this section, we focus our attention on the actual process of choosing an appropriate value for the parameter ξ and fitting the model.

We will build two models: one assuming that the value for the parameter ξ is equal to 1 and one choosing the best ξ maximizing the likelihood as described in section 4.2.

We want to underline one more time that setting the parameter $\xi = 1$, has the advantages of producing estimates μ_{ij} that resemble those of the standard Chain-Ladder model. We can identify,

therefore, a stochastic model that underlies a common and well-known actuarial technique.

In our analysis, however, we are more interested in finding the best model according to the data we have; we do not want to introduce any not verified pre-assumptions or create any kind of data leakage.[9]

5.1 Analysis Objective

The ultimate objective of the analysis is to build a probability distribution for the INBR. This will allow to construct confidence intervals, calculate moments, or risk measures such as Value at Risk.

The estimated IBNR will be defined as the sum of the estimated incremental values for the lower part of the triangle:

$$IBNR = \sum \hat{\mu}_{ij} \qquad (5.1)$$

In this expression, we used the notation $\hat{\mu}_{ij}$ to indicate that we are referring to the mean estimates of the lower part of the triangle, i.e. $i + j > n + 1$.

Moreover, since the values μ_{ij} will be the mean estimates for the observed values q_{ij} obtained from the GLM, we can build a probability distribution for each of them, and, under the assumptions of independence, the probability distributions of the $IBNR$ will be the sum of the probability distributions of each μ_{ij}.

[9] Data leakage is defined as the creation of unexpected additional information in the training data. In this case it will be forcing the parameter $\xi = 1$ without supporting evidence.

5.2 Building the Models

In this section we will build the two aforementioned models for a sample loss triangle. We will use the following loss data, shown in incremental form:[10]

	1	2	3	4	5	6	7	8	9	10	11
1977	153,638	188,412	134,534	87,456	60,348	42,404	31,238	21,252	16,622	14,440	12,200
1978	178,536	226,412	158,894	104,686	71,448	47,990	35,576	24,818	22,662	18,000	
1979	210,172	259,168	188,388	123,074	83,380	56,086	38,496	33,768	27,400		
1980	211,448	253,482	183,370	131,040	78,994	60,232	45,568	38,000			
1981	219,810	266,304	194,650	120,098	87,582	62,750	51,000				
1982	205,654	252,746	177,506	129,522	96,786	82,400					
1983	197,716	255,408	194,648	142,328	105,600						
1984	239,784	329,242	264,802	190,400							
1985	326,304	471,744	375,400								
1986	420,778	590,400									
1987	496,200										

Table 5 Incremental Loss Triangle

The procedure to build the models will be very similar for the two procedures. There is only one caveat: when following the first procedure we will fix $\xi = 1$. Whereas when following the second approach, alongside the model coefficients, we will also estimate ξ by maximizing the likelihood function.

We will look now at the individual steps that will allow us to construct the full distribution of the $IBNR$.

Given the incremental loss data from Table 9, we can fit the initial GLMs. These models will take the form specified at the equations (4.1), (4.2) and (4.3). As already specified, we will construct two different models: one that replicates the Chain-Ladder estimates and one estimating the best parameter ξ according to the data.

Once the model coefficients have been found, we will use them to predict the lower part of the triangle. The sum of this individual values, as specified in equation (5.1), will be the mean of the expected $IBNR$.[11]

The solution of GLM models, however, will also give us the estimate of the Tweedie parameter ϕ, and for the second model ξ.[12]

---

[10] B. Zehnwirth and G. Barnett. Best Estimates for Reserves. Proceedings of the CAS. Volume LXXXVII. Number 167. November 2000.

[11] We are predicting incremental values, so the value of the $IBNR$ can be calculated by summing them up.

[12] In the first model we made the assumption $\xi = 1$.

Under the assumption that each observed value $q_{ij} \sim Tweedie(\mu_{ij}, \phi, \xi)$ we could sample random variates from each respective distribution to create simulated loss values q_{ij}^{*}.

At this point we have obtained a set of simulated loss values that we can use to fit a GLM and predict the lower part of this simulated triangle, which will be a simulated value of the $IBNR$. We can repeat this process a sufficient number of times in order to create a distribution of the expected $IBNR$.

In case we are following the second approach, the methodology will be exactly the same with one difference: before fitting the GLM on the simulated data, we will always estimate the Tweedie parameter ξ. This procedure can be achieved with many available statistical packages. However, it will be very slow and computationally heavy.

5.3 Model Results

In this section we will look at the distribution of the $IBNR$ obtained with both of the procedures described.

The first model, the one with $\xi = 1$, led to the following results for the total $IBNR$ value:

1st Quartile	Median	Mean	3rd Quartile	Standard Deviation	CV = SD/Mean
5,170,246	5,277,257	5,279,430	5,386,454	160,713	0.03

Table 6 First Model Results

And the following distribution:

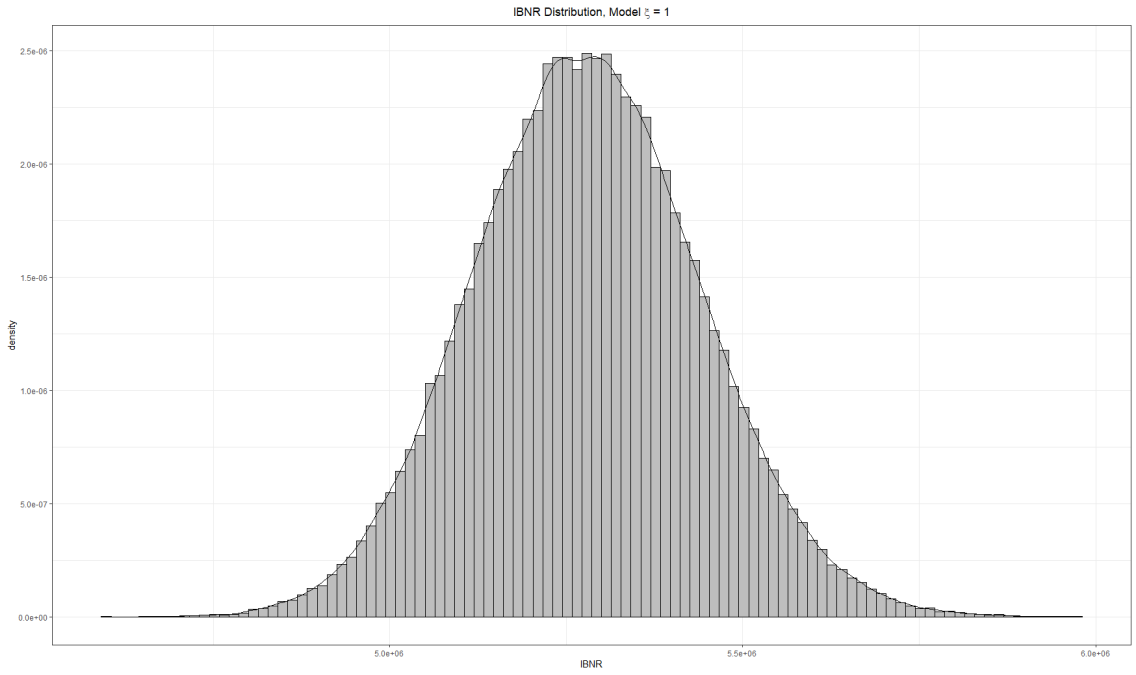


Figure 1 IBNR Distribution, Model $\xi = 1$

We can also look at a 95% confidence interval:

2.5%	97.5%
4,970,226	5,600,093

Table 7 IBNR Confidence Interval

The second model, the one with the parameter $\xi = \bar{\xi}$ obtained by maximizing the likelihood function, led to the following results for the total $IBNR$ value:

1st Quartile	Median	Mean	3rd Quartile	Standard Deviation	CV = SD/Mean
5,052,754	5,225,989	5,223,112	5,409,020	264,261	0.05

Table 8 Second Model Result

And the following distribution:

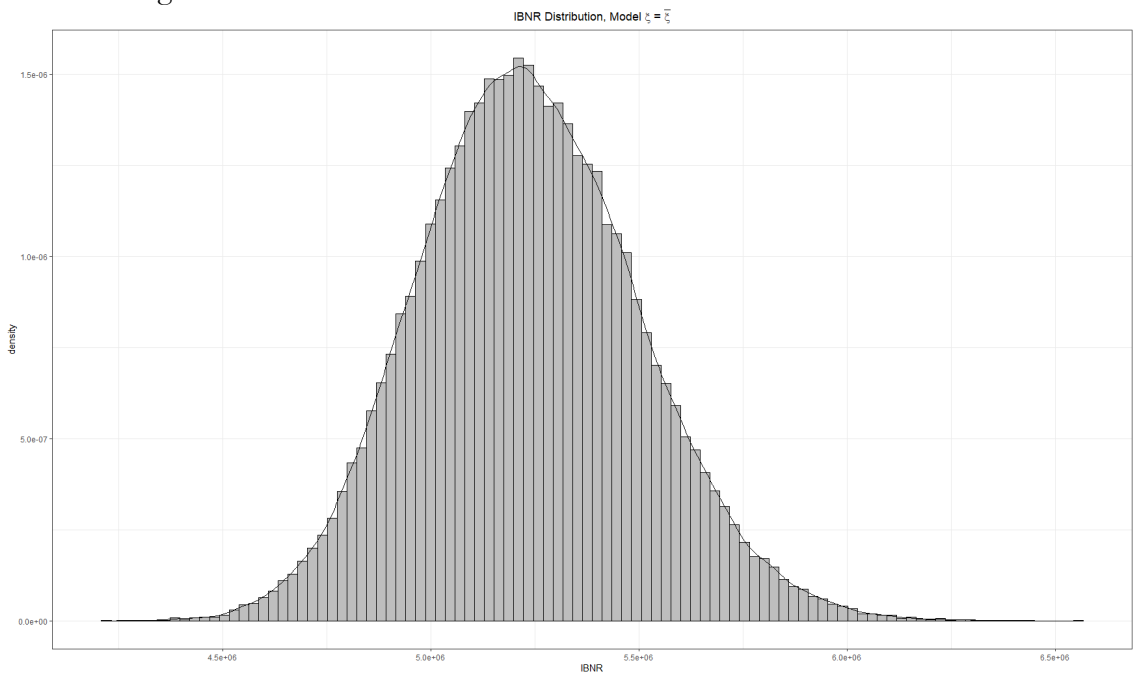


Figure 2 IBNR Distribution, Model $\xi = \bar{\xi}$

With the relative 95% confidence interval:

2.5%	97.5%
4,741,983	5,778,540

Table 9 IBNR Confidence Interval

Here a comparison of the two distributions side by side:

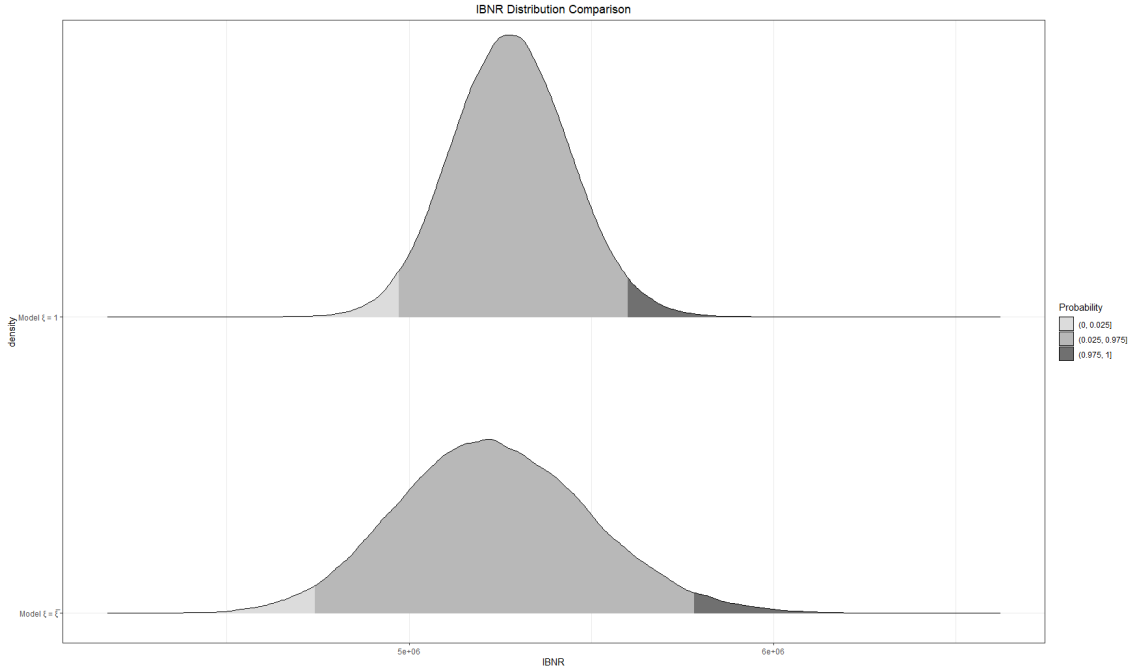


Figure 3 IBNR Distribution Comparison

5.4 Results Analysis

As we can observe from the previous descriptive statistics, the two models produce very close results for the mean value of the $IBNR$ estimates. However the standard deviation, and therefore the overall shape, of the two distributions is very different.

This behavior depends, as expected, on the parameter ξ, in one case fixed to one, in the other estimated to be equal to 2.143.[13]

Since the variance function is proportional to the mean, as specified by equation (4.2), we are not surprised that the second model leads to a higher level of variability around the mean: the CV of the first distribution is 0.03 versus a value of 0.05 for the second.

If we are only interested in a mean estimate, this phenomena will not impact our estimates as much. However if we are trying to build confidence intervals around the mean, as it usually happens, the difference in the two models could be substantial.

This aspect is even more evident when dealing with losses that are very high from an absolute amount perspective.

[13] More detail on this process in the next section.

6. MODEL VALIDATION

We will look at the topic of model validation having emphasis on the numerical example previously presented. Again, all the details and the code used for the analysis can be found in the appendix.

Model validation is a very important part of any analytical analysis, the outputs of a model and the model assumptions have to be validated against the real-data in order to ensure that the model is performing as expected and the analysis objectives can be achieved.

We will validate that the GLM distribution assumptions are met, i.e. that the data is actually following a Tweedie distribution with a specified parameter ξ and we will also check the quality of fit of the two models.

6.1 Validating the parameter ξ

In order to check the robustness of the value of the parameter ξ, it is useful to look at the profile likelihood of the model. This process consists in fitting several GLMs to the data with different input values of ξ and then computing the log-likelihood. The value that will lead to the maximum log-likelihood should be the value to use when fitting the final GLM.

These are the results for the triangle presented in Table 5:[14]

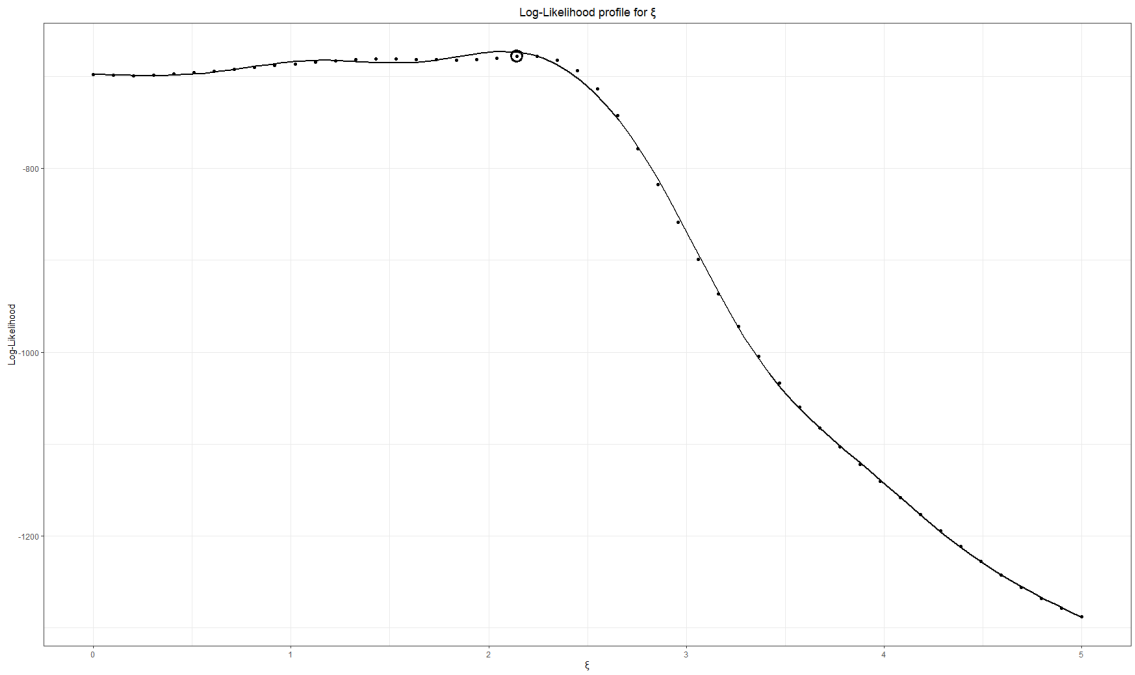


Figure 4 Log-likelihood profile

[14] The Tweedie distribution is not defined for values of ξ between 0 and 1. In the plot the values of such points have been interpolated.

We can also look at the individual values for the log-likelihood:

$\xi = 1$	$\xi = 2.143$
-686.161	-677.909

Table 10 Log-likelihood values

6.2 Validating the Model

There several techniques to validate the output of a GLM, such as looking at the Person or deviance residuals. These two techniques in some contexts, however, could be flawed (Dunn et al., 1996) because they will not lead to residual values that are exactly normally distributed.

In order to overcome this issue, we will look at the values of the quantile residuals (Dunn et al., 1996), which are always normally distributed.

We can then look at the QQ-plot for the two models to check which one leads to better results.

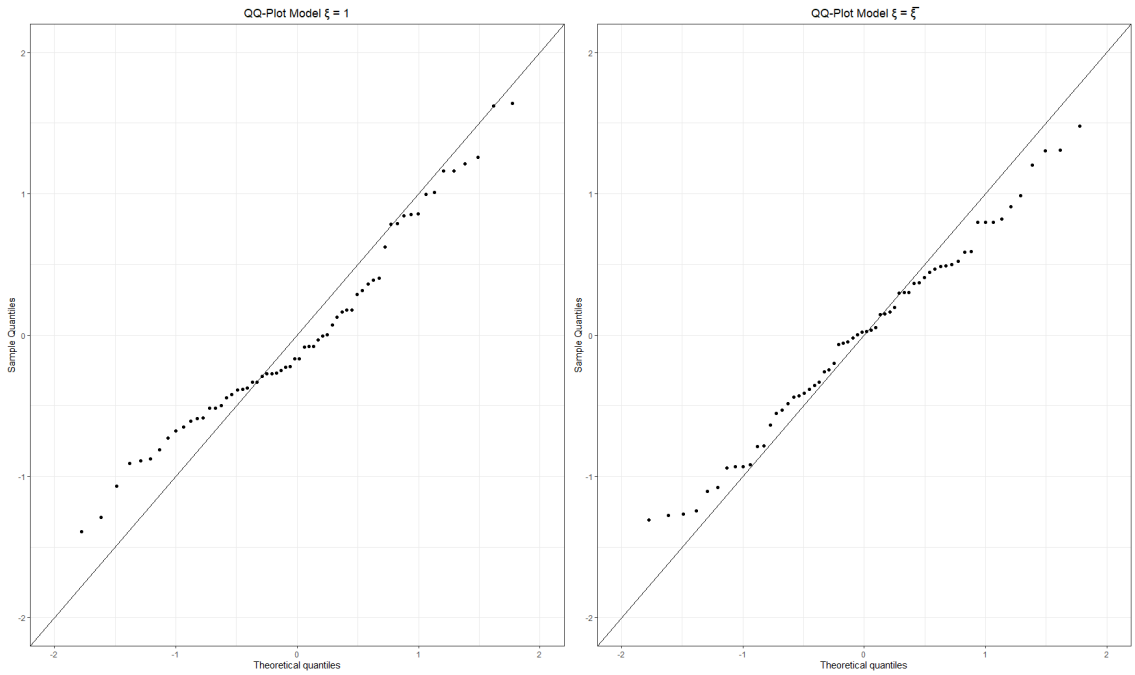


Figure 5 QQ-Plot Comparison for the Models

Again, we can see that the second Model led to a better fit.

For completeness we can also look at the value of the residual deviance achieved by the two models:

$\xi = 1$	$\xi = 2.143$
36,594	0.0267

Table 11 Residual deviance values

7. DEVELOPMENT CURVE FITTING

The concept of fitting a development curve to triangle data is justified by several advantages that will either make the whole reserving process easier and faster or stabilize results and make them less susceptible to outliers (Clark, 2003).

As described by (Clark, 2003), the procedure of fitting a parametrized curve to loss data is based on the concept of maximum likelihood and hence share a lot of assumptions with the topics treated in this paper.

7.2 Theoretical Definition

The target of fitting a development curve to loss data is to find a curve, or more specifically, the parameters that define the curve shape that describe the process of loss emergence. Such curve will be a monotonically increasing function that goes from 0 to 1 as time goes from 0 to infinity. At each point in time such function, $f(t|\theta)$, will describe the percentage of paid (or incurred) claims for each specific origin period.

The expected value μ_{ij} of each incremental claim amount q_{ij} can therefore be defined as:

$$\mu_{ij} = ULT_i[f(j|\theta) - f(j-1|\theta)], \tag{7.1}$$

Where:

ULT_i is the Ultimate Claim amount for origin period i,

$f(j|\theta)$ is the selected emergence function evaluated at time j subject to parameters θ.

According to the formulation above, it is clear that the numbers of parameters that need to be estimated are the sum of how many origin periods we have plus the number of how many parameters the emergence function has. If we look at the triangle in Table 5, we will have 11, the number of accident years, plus how many parameters the chosen emergence function has.

From a calculation standpoint this could be very difficult to treat and optimize. To overcome this issue we could rewrite equation (7.1) in order to reduce the number of parameters to find. If we consider that $ULT_i = \frac{\sum_0^j q_{ij}}{f(j|\theta)}$, we can then set:

$$\mu_{ij} = \frac{\sum_0^j q_{ij}}{f(j|\theta)}[f(j|\theta) - f(j-1|\theta)] = \frac{c_{ij}}{f(j|\theta)}[f(j|\theta) - f(k|\theta)], \tag{7.2}$$

Thus reducing the number of parameters that need to be estimated equal to the number of the parameters in the function $f(t|\theta)$.

If we now consider the quantities q_{ij} distributed according to a specific probability distribution, we can rewrite the likelihood function considering the equation (7.2), and then maximizing it to obtain

the vector of parameters θ. It follows that the choice of the probability distribution we select for q_{ij} will have great impact on the solution θ, their respective standard errors, and ultimately on the shape of the emergence function, and hence, on the estimated ultimate claim amounts and its variability.

7.2 Fitting the curves

In this section we will take a look at an actual curve fitting process on the data presented in Table 5. We will fit two types of curves commonly used for these purposes: the Loglogistic cumulative distribution function (cdf) and the Weibull cdf. Each of these curves ~~we~~ will be fitted by the process of maximizing the likelihood as described in the previous section; the first time assuming that the data comes from a Tweedie distribution with parameter $\xi = 1$ and the second time with the estimated parameter $\xi = 2.143$. All the details and technical steps, alongside the code used, are made available in the appendix.

Both of the curves are described by two parameters, α a shape parameter, and β, a scale parameter. Here the values found for the data analyzed:

	$\xi = 1$		$\xi = 2.143$	
Curve	α	β	α	β
Loglogistic	1.639	2.634	1.603	2.622
Weibull	1.252	3.280	1.096	3.445

Table 12 Development curve paramters

and the graphical representation:

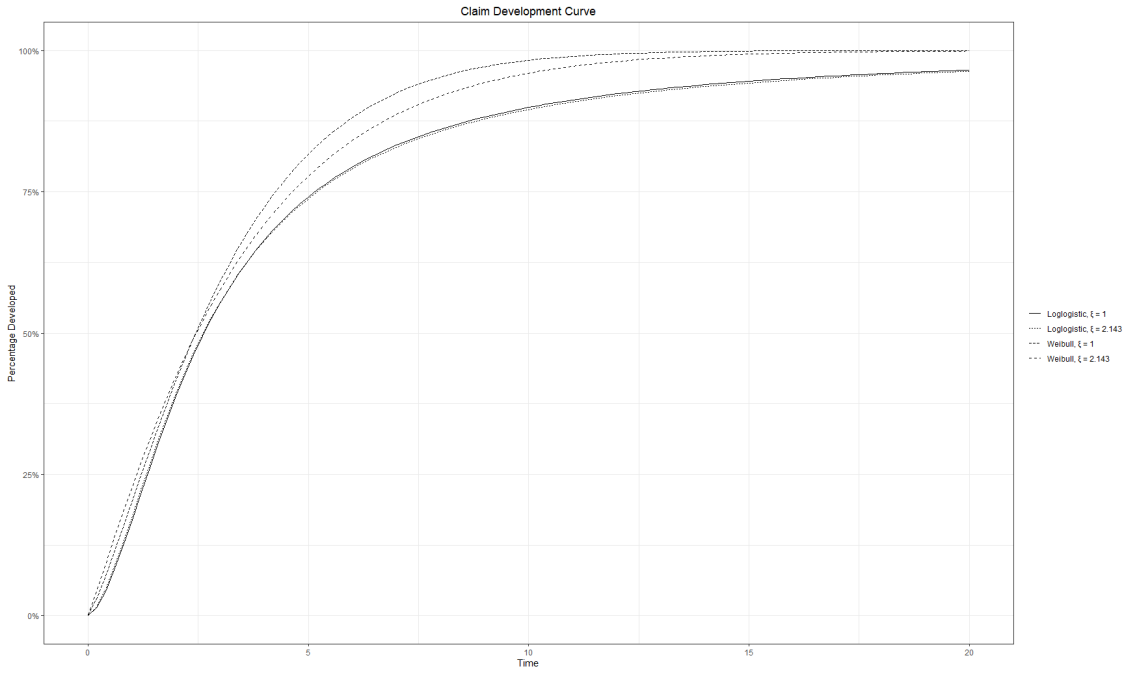


Figure 6 Claim Development Pattern

As we can see, the curves are very close to each other. However, we observe a very different situation regarding the Fisher Information Matrix, which is used to calculate the covariance matrix, and hence, the standard errors of the parameters.

	$\boldsymbol{\xi = 1}$		$\boldsymbol{\xi = 2.143}$	
Curve	$\boldsymbol{se(\alpha)}$	$\boldsymbol{se(\beta)}$	$\boldsymbol{se(\alpha)}$	$\boldsymbol{se(\beta)}$
Loglogistic	0.0010	0.0017	0.7598	0.7649
Weibull	0.0005	0.0016	0.3685	0.8822

Table 13 Parameter standard Errors

Again, it is not surprising that due to the higher variance assumed when fitting the curves considering the parameter $\xi = 2.143$, we have higher standard errors. This will ultimately lead to higher uncertainty in the estimate of the ultimate claim amounts.

8. CONCLUSIONS

This paper has analyzed the concept of Over-Dispersion and assumptions testing in the context of loss reserving. We focused on the main theoretical details behind the so-called quasi-distributions and how they relate to the Tweedie distribution. A formal definition of the mathematical structure has

been provided, alongside the practical implications that arise from the process of parameter selection.

The paper also presented how a linear model can be defined and built in order to predict expected claim amounts. We analyzed how the resulting predictions differ, both in terms of expected mean and overall distribution, when we change the underlying assumptions. This concept has been particularly stressed to underlining the importance of hypothesis testing in order to build the most accurate model for the data.

Another application that has been investigated is fitting a development curve to loss data. As shown, this shares several assumptions with the concepts previously discussed, and we analyzed how predictions and prediction errors vary with varying assumptions. Even in this case we underline the importance of hypothesis validation and assumption verification.

Overall, we carried out several tests in order to demonstrate how important assumptions testing is when applying common and widespread models. The main outcome is that predictions obtained with models that have been fully calibrated on the data always perform better, leading to the most accurate and stable predictions.

Supplementary Material

The full code used to produce the numeric results presented in the paper is available. Moreover the simulated IBNR vectors are also available. Producing these results could be very heavy from a computational point of view and therefore we decided to make them available as well.

9. REFERENCES

[1] Committee on Property and Liability Financial Reporting, American Academy of Actuaries, "A Public Policy Practice Note, Statements of Actuarial Opinion on Property & Casualty Loss Reserves," **2018**.

[2] Clark, D.R., "LDF Curve Fitting and Stochastic Reserving: A Maximum Likelihood Approach," *Casualty Actuarial Society Forum*, Fall **2003**.

[3] Dunn Peter K., Smyth Gordon K., "Generalized Linear Models With Examples In R," *Springer*, **2018**

[4] Dunn Peter K., Smyth Gordon K., "Randomized Quantum Residuals," *Journal of Computational and Graphical Statistics*, **1996**, *Vol. 5, No. 3*.

[5] Dunn, Peter K. Smyth, Gordon K, "Tweedie family densities: methods of evaluation," *Proceedings of the 16th International Workshop on Statistical Modelling*, **2001**.

[6] McCullagh P., "Quasi-Likelihood Functions," *Annals of Statistics,* **1983**, *Vol. 11, Issue 1*.

[7] Peters G. W., Shevchenko P. V. and Wüthrich M. V., "Model Uncertainty in Claims Reserving within Tweedie's Compound Poisson Models," *ASTIN Bulletin,* **2009**, *Vol. 39, Issue 1*.

[8] Renshaw A.E., Verrall R.J., "A Stochastic Model Underlying the ChainLadder Technique", *British Actuarial Journal,* **1998**, *Vol. 4, Issue 4*.

[9] Shapland M. R., "Using the ODP Bootstrap Model: A Practitioner's Guide," CAS Monographs, **2016**, *No. 4*.

[10] Wedderburn R. W. M., "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," *Biometrika,* **1974.**

Abbreviations and notations

AIC, Akaike's Information Criterion
AY, Accident Year
CDF, Cumulative Development Factors
DP, Development Period
EDM, Exponential Distribution Model
GLM, Generalized Linear Model
IBNR, Incurred But Not Reported
MLE, Maximum Likelihood Estimator
ODP, Over-Dispersed Poisson

Biography of the Author

Marco De Virgilis is a Senior Actuarial Analyst working for Allianz Global Corporate & Specialty in the Chicago Office. After achieving his Msc in Actuarial Sciences he worked as a reserving actuary for Direct Line Group in the London office reporting on both personal and commercial lines.
Following this experience, he worked as a consultant for Deloitte. During this time he developed a more thorough knowledge of the actuarial market and industry practices.
He then worked in the R&D department as a Data Scientist for Unipol, one of the biggest insurance Italian players. During this experience he developed analytics skills linking standard actuarial practices and modern frameworks.

Contact Information
- devirgilis.marco@gmail.com

The Chase — An Actuarial Memoir

Glenn Meyers, FCAS, MAAA, CERA, Ph.D.

June 7, 2020

Abstract

One of the more interesting problems that actuaries have to deal with at some point in their careers is that of assigning a price to risk. Like many others, I have been chasing this problem, off and on, throughout my entire actuarial career. My most recent attempt at a solution is published in my monograph, Meyers (2019), on stochastic loss reserving. I am writing this paper as a historical memoir, summarizing the path I took to arrive at this particular solution. Over the years I have worked on projects related to the two distinct parts to this problem. The first part is to obtain a prediction of the distribution of the outcomes. The second part is to assign a financial value to that distribution in such a way that it can compared to other risky investments. In retirement I was able to direct my experiences toward a solution to the problem for loss reserve liabilities. I hope the reader will find this paper both interesting and informative.

1 Introduction

With the release of my recent major publication, Meyers (2019) on stochastic loss reserving, I can see that my actuarial career is winding down.

The monograph deals with the quantification and valuation of risky liabilities — a problem that I have been chasing, off and on, over my entire actuarial career. Looking back, I thought that the twists and turns of this Chase had the elements of an interesting actuarial story. The goal of this paper is to tell that story.

To provide some context, it will help to know about my employers and the kind of work I was doing. Let me start with a quick overview of my career.

- After receiving my Ph.D. in pure mathematics from SUNY Albany in 1972, my intent was to become a college math professor. Unfortunately the academic market was flooded with brand new Ph.D.s and the best position I could get was a temporary position at the University of Rhode Island. While there, I decided to drop the subject of my graduate work and develop skills that were more marketable. I chose computing and numerical methods. I also passed the first two actuarial exams. When my contract

with URI was renewed for a second year, I taught numerical methods to engineering students.

- I took my first job as an actuary at the Hartford Insurance Group in 1974. My main responsibility was to prepare rate filings for private passenger auto insurance.

- In 1976 I took a very different job at CNA Insurance. For most of the eight years I was there, I worked on a series of special projects. For four of those years, my boss was Yakov Avichai[1], a statistician and not an actuary. Many of my projects had to do with what we now call predictive modeling using SAS. Others, which I will discuss below in some detail, involved large account pricing. In my final two years at CNA, I was the actuarial manager for large account pricing.

  Another part of my responsibilities was to represent CNA on several ISO and NCCI research committees. These committees worked on topics in credibility, individual risk rating and increased limits ratemaking.

- In 1984, armed with both a Ph.D. and an FCAS, I gave academia another try at the Department of Statistics and Actuarial Science at the University of Iowa. In my four years there, I got an overview of the statistical landscape and established some good academic contacts. But in the end, I concluded that my professional future lay elsewhere in the insurance industry.

- In 1988 I joined ISO where I remained until I retired in 2011. As I interviewed for the job, it was clear that they wanted to use their comprehensive insurance data to develop new insurance products. This was at the time when ISO was transitioning from a rating bureau that was wholly owned and controlled by the insurance companies to an independent, for profit, corporation. My job there was to do research projects that would lead to new products. I will defer discussing those projects that relate to the Chase until later. The other projects that I worked on included:

  - Classification ratemaking in response to California Proposition 103
  - Special studies published in the *ISO Insurance Issues Series*
  - Implementing catastrophe models in the ISO loss costs
  - Securitization of catastrophe risk
  - Producing claim severity distributions for reinsurers
  - Develop procedures for identifying fraudulent claims with ISO's ClaimSearch database.
  - Highly refined risk classification for the ISO Risk Analyzer Suite

  Some of my projects led to successful ISO products, and others didn't. In exploring the Verisk website I see that many of the products that I worked on have been significantly expanded. I feel proud of my participation on the teams that worked these products.

[1] Scroll down a bit to see his obituary

- In retirement, I continued writing, attending CAS and ASTIN meetings and serving on a few of their committees. Also, I continued the Chase.

The objective of the Chase was to find a solution to two related problems.

1. Obtaining the distribution of financially relevant outcomes.
2. Assigning a financial value to that distribution.

The remainder of this paper will describe the path I took to get to my current solution of the problem.

2 First Exposure

I took my first actuarial job in 1974 preparing private passenger auto rate filings. At that time, the standard template for ratemaking in that line of insurance was based on the *PCAS* paper by Philipp K. Stern [1956]. Summarized at a high level, the Stern approach was to estimate the costs and expenses for the future policy period and add a 5% (???) profit margin.

The market conditions at the time allowed for significant price increases, subject to approval from state regulators. This led to an adversarial relationship between the company actuaries and the regulators where various estimates were often challenged. One of the more significant challenges to the Stern methodology was the omission of investment income generated by policyholder premiums. Many, but not all, states required insurers to include consideration of investment income in their ratemaking.

My first few rate filings used the Stern methodology. As I gained experience, I would work on states that required us to consider investment income. Then I started getting special projects, two of which were relevant to the Chase.

For auto physical damage, application of the Stern methodology would often indicate that the rates for a $50 deductible coverage would be less than the rate for a $100 deductible coverage. The explanation given for this was that those who purchased the lower deductible were better drivers.

Since we did not want to charge a lower rate for the lower deductible (otherwise nobody would purchase the higher deductible) we had an algorithm to artificially raise the $50 deductible rate and lower the $100 deductible rate. When I examined the algorithm, I noticed that the total premium generated after application of the algorithm was less than the total premium generated before application of the algorithm. I then went to my boss and offered to rewrite the algorithm so that it would balance the overall premium.

The next day he came back to me with a paper, Cahill (1936), that described how to use claim severity distributions for pricing coverages with deductibles. My project was to construct an empirical claim severity distribution. Then I developed the code to adjust the distribution for inflation and calculate loss elimination ratios by deductible.

Another project I got was to critique an investment income methodology required by the Massachusetts Division of Insurance. A high-level description of the methodology was to calculate the present value of the loss after consideration of the loss payout pattern. The profit margin was determined assuming that the insurer investment was determined by the so-called Kenney Ratio. The profit margin was then determined by set so that an insurer received a 15% return on its investment.

I saw nothing structurally wrong with the approach. My analysis consisted of sensitivity testing the final rates to the various input parameters. Afterwards, I asked my boss how he would critique the method, he said simply — We think the return on investment should be 18% (???).

Looking back, these two projects were my introduction to the Chase.

1. I was introduced to loss distributions.
2. I was exposed to the problem of determining an appropriate return for an insurer's investment. At the same time, I was exposed to Bayesian analysis and utility theory by studying operations research, specifically Raiffa (1970), and Willett (1901) on what was then Part 4 on the actuarial exams.

3 Retrospective Rating

My early projects at CNA involved data processing. One of these projects was to build a database of individual claims from which we could construct claim severity distributions. Working with the raw claim audit files, I built a database that allowed one to track the development of individual claims for both paid and incurred losses over time.

When that project was completed, my next project involved retrospective rating. I talked to an underwriter who lectured me saying (as I recall): "Retrospective rating is not for everyone. The accounts we want are high frequency/low severity risks with relatively stable loss ratios. The ones we don't want are the high severity/low frequency risks. We will collect the minimum premium for several years running. Then the big claim will blow through the maximum and we end up losing money."

The question to be addressed was one of pricing. How do we vary the insurance charge by estimated claim severity? To address this I needed claim severity distributions from the database that I had been working on earlier, and the collective risk model.[2] This model can be thought of as a computer simulation where we:

1. Select a random number of claims, N, from a claim count distribution.
2. for $n = 1, \ldots, N$, select a claim size, Z_n, from a claim severity distribution.
3. Then the loss for the simulated account loss $X = \sum_{n=1}^{N} Z_n$.

[2]My introduction to the collective risk model came from Beard, Pentikäinen and Pesonen,1977.

Given a large set of simulated account losses, $\{X_K\}_{K=1}^{10,000}$, I calculated the insurance charge that would yield an adequate average retrospective premium, using numerical methods. Meyers (1980) describes the method in detail. It is worth noting that the simulations originally ran overnight on a mainframe computer taking about 45 minutes of CPU time.

I presented this paper at my first CAS meeting, where I received my ACAS, as part of a call paper program in May of 1980. Another paper in the call, written by Shaw Mong (1980), attracted my attention. What it did was calculate the cumulative probabilities of a collective risk model by numerical methods — not by simulation. A critical assumption in Mong's model is that the claim severity distribution was a gamma distribution[3]. At that time, I was using empirically tabulated claim severity distributions. But I thought Mong's approach was important enough to investigate in detail.

Back at the office I worked with a summer intern, Nathaniel Schenker who was then a graduate student in statistics at the University of Chicago, to duplicate Mong's model. When completed we compared it with our simulation model by approximating our empirical severity distributions with a gamma distribution. Unfortunately it did not work well when we approximated our severity distributions with a gamma distribution.

A high-level summary of Mong's algorithm is that it expresses the Fourier transform (FT) of the aggregate loss distribution in terms of the FT of the gamma distribution and the probability generating function of the claim count distribution. The aggregate loss distribution was then obtained by a numerical inversion of the FT.

Upon reflection, the key property of the gamma distribution that made Mong's algorithm work was that there was a closed form expression of its FT. Another distribution with a closed form FT was the uniform distribution, and the piecewise uniform distribution – otherwise known as a histogram. And histograms can approximate any distribution (with finite support) to any desired degree of accuracy. It was a fairly simple task to replace the gamma distribution in our code with a histogram. Thus we were able to calculate aggregate loss distributions without resorting to simulation. The bad news was that the numerical integration formulas in Mong's algorithm were not efficient for this problem, and the simulations actually ran faster.

At this point I teamed up with Phil Heckman and together we came up with an efficient numerical integration formula for the Fourier inversion and we derived a formula for the excess pure premium, useful for retrospective rating. The final algorithm ran in a matter of seconds on a mainframe. We described the algorithm in Heckman and Meyers (1983).

A feature of the collective risk model is that care must be taken to counteract the "law of large numbers." For example, if we choose the λ parameter (the mean of the claim count distribution) in a Poisson distribution, the coefficient of variance of the aggregate loss distribution will approach zero as the size of the account, as measured by λ, increases. We did not accept this conclusion. After all, variances of loss ratios for entire lines of insurance were noticeably larger than zero.

[3]Several years later I found out that Mong's model was a generalization of the Tweedie distribution. See Meyers (2009).

Also, I was serving on an NCCI task force to update their Table M.[4] The data underlying this table consisted of tens of thousands of account losses and their associated premiums. Here, we also saw that the "law of large numbers" was violated.

We attributed this phenomenon to parameter uncertainty. To account for this in our collective risk model we chose our claim count distribution to correspond to the following simulation algorithm:[5]

1. Select a random variable, χ, from a gamma distribution with mean 1, and variance, c.
2. Given an overall mean claim count, λ, select a random claim count from a Poisson distribution with mean $\chi \cdot \lambda$.

It can be shown (Heckman and Meyers (1983) Eq 3.4) that a consequence of including parameter uncertainty in the claim count of a random loss, X, is that:[6]

$$Var[X] = a \cdot E[X] + c \cdot E[X]^2 \tag{1}$$

A consequence of Equation 1 is that the coefficient of variation of X is equal to

$$\sqrt{\frac{a}{E[X]} + c}$$

which approaches $\sqrt{c}$ as the expected loss, $E[X]$, increases.

As Phil and I were completing our paper, Nat Schenker came back for another summer and together we made a first attempt at estimating the parameter, c, in our model from empirical data. Our results were written up in our paper, Meyers and Schenker (1983).

The early 1980's were a period of high inflation, high interest rates and a competitive insurance market. One way insurers were competing in the large account market was to offer paid loss retros, where the retrospective premium was determined by the paid, rather than the incurred losses. A question our underwriters wanted us to address was "How much higher should the insurance charge be when we offered a paid loss retro?"

Given that we had the data to construct claim severity distributions by claim maturity (see the first paragraph of this section) we evaluated the expected retrospective premium for paid losses for each adjustment period using Heckman Meyers algorithm. Then using numerical methods we were able to calculate the insurance charge that yielded the desired present value of the expected retrospective premium. I described these calculations in Meyers (1986).

When I became the actuarial manager for large account pricing, I implemented this retrospective rating tool as part of our large account pricing workup. I know for a fact that the tool continued to be used for (at least) several years after I left CNA.

[4] Table M was a table of excess pure premium ratios, tabulated by account size, used for retrospective rating in the US. My understanding at the time was that it was the standard used for all lines of insurance.

[5] Heckman and Meyers (1983) allowed for parameter uncertainty in both claim count and claim severity. The salient points can be made here with the simpler model.

[6] In the terminology of Heckman and Meyers, Eq 3.4, $a = E[z^2]/E[z]$. Typically, this is a large number.

In closing this section, I would like to describe an interesting failed project. The risk manager for one of our large workers' compensation accounts had a problem. For several years running, their incurred loss retrospective rating plan had a fairly large maximum. Typically, the retrospective premium was well below the maximum premium. While the account was prepared to pay this maximum premium for a given year, they were worried that the retrospective premium would increase simultaneously for several prior years. I was asked to develop a methodology to price a contract adjustment that lowered the maximum premium after the first retrospective adjustment.

If this project were to be done today, we would apply our favorite stochastic loss reserve model to the account's current losses and estimate the expected cost of lowering the maximum premium. But in the early 1980's, before Mack (1994), Barnett and Zehnwirth (1998) or England and Verrall (2002), we did not have such a model. The only research on this topic that I knew of at the time was a paper by Hachemeister (1976) that treated loss development as a Markov chain. I tried to apply the Hachemeister methodology to this account and came up with a negligible charge for lowering the maximum premium. Neither I nor my colleagues believed this result and we did not offer the adjustment.

As it turned out, the worse case scenario actually happened. We got a new claims manager in the branch office that dealt with this account. In examining the branch's claims on file, the manager reassigned several claims to the workers' compensation "tabular" program. This had an adverse affect of the account's retrospective premium going back several years. When I became a manager in the large account pricing unit, I had to write a letter to the risk manager explaining these loss adjustments.

This project was my first brush with stochastic loss reserving and the problem of dependencies. It was at the forefront of my thinking when I wrote Meyers (2007) titled "The Common Shock Model for Correlated Insurance Losses."

4 Increased Limits Ratemaking

I joined ISO in 1988 as a researcher in what was then called the Actuarial Development Department. My first project was to work on a revision to their increased limits ratemaking methodology. Before describing the new methodology, I should say something about my background that indicated I could help on this project.

First, having represented CNA on the ISO Increased Limits Subcommittee, I was familiar with their original methodology. Prior to my joining this subcommittee, they had developed a methodology that used a five-parameter Pareto distribution described by Patrik (1980) and a risk load methodology described by Miccolis (1977). One of Miccolis' sources for risk loads was Bühlmann (1970), which I was also familiar with.

In 1984, I left CNA to join the faculty in the Department of Statistics and Actuarial Science at the University of Iowa. While there I got to know Stuart Klugman and and Bob Hogg, the authors of *Loss Distributions*, a textbook that had been recently added to the CAS Syllabus of Examinations. I taught courses in compound interest, numerical analysis,

life contingencies, survival models and risk theory. Also, I developed and taught their first course in casualty insurance covering credibility theory and loss distributions.

During the summers I worked for ISO doing research on fitting loss distributions. As I approached my third summer, ISO suggested that I apply for a full time position there, and the next year I did.

There are two aspects of increased limits ratemaking that should be treated separately: (1) Fitting the claim severity distribution; and (2) Determining a risk load. Let's treat these two aspects of the problem in separate subsections.

4.1 Claim Severity Distributions

The Hogg and Klugman text first describes a few basic distributions such as the normal and gamma distributions, and then generates other distributions by transforming the random variables with either a logarithmic or a power transform. The text also describes mixtures of distributions. As examples they gave a mixture of a gamma and a loggamma distributions, and then moved on to compound distributions where the parameters of the distributions are also random variables. The examples in the text were mainly using a gamma distribution for mixing the parameters.

My reading of the Hogg and Klugman text suggested that something in this very rich family of transformed and mixed distributions would fit the ISO data. This is the problem that I worked on for ISO during the summers while at the University of Iowa. After this work, I was favoring a discrete mixture of Pareto distributions when I started at ISO in 1988.

When I joined ISO, the increased limits ratemaking formula was unstable, producing unacceptable swings from year to year. The main problem ISO was having was not in their choice of claim severity model, but in the way they "developed" the claims to their ultimate value. To sidestep this, I advocated fitting the model to settled claims arranged by how long it took the claims to settle. The counterargument was that we were discarding information. My counter to that was that with tens of thousands of claims, we had enough information. In the end, we used only settled claims.

After thorough exploration, we ended up with a model consisting of a mixture of several Pareto distributions. I originally named it the "Pareto Soup" model.[7]

As the project gathered momentum, we decided to add another mathematical actuary to our staff. We chose Clive Keatinge, FCAS. When Clive came on, we were beginning to grapple with fitting the Pareto Soup model. Clive took the initiative to fit the model by multivariate Newton-Raphson iteration. His first several weeks on the job were spent calculating the many partial derivatives in the Jacobian matrix needed for the iterations. When finished, the fitting procedure ran in a matter of seconds.

When we finished the model, Clive was instrumental in transferring the entire increased limit ratemaking methodology to our Increased Limits Division.

[7]When we came to file the model with insurance regulators, our management changed its name to the more formal "Mixed Pareto" model.

This was not the end of the claim severity distribution story. The next year, ISO issued a call for individual claim data from excess and umbrella policies. Our job was to see if we could use it to enhance our increased limits model. Spoiler alert — We ended up changing our mixed Pareto to a mixed exponential model.[8] The path that led to the mixed exponential distribution makes for an interesting story.

Our fits on the data that included the excess and umbrella data into our Pareto Soup model were not particularly good. A more flexible model seemed to be called for. Clive suggested that we needed a model with alternating signs in its successive derivatives. Consider the density function, $f(x)$, for the exponential distribution.

$$f(x) = e^{-x}$$

$$f'(x) = -e^{-x} \text{ negative}$$
$$f''(x) = e^{-x} \text{ positive}$$
$$f'''(x) = -e^{-x} \text{ negative}$$
$$\dots$$

Now let's back up a bit. While I was an academic, I joined the Risk Theory Society, a small but active subgroup of the American Risk and Insurance Association. The group consisted mainly of business and economics professors with an interest in insurance. I maintained my membership in that group until I retired at the end of 2011, as it provided me with a view of insurance that was quite different from what one normally gets at actuarial society meetings.

Many members of this society were interested in utility theory. This theory proposes that a utility function, $u(x)$, can be used to describe how a person makes decisions under uncertainty. Conditions imposed on $u(x)$ typically include $u'(x) > 0$, i.e. more is better and $u''(x) < 0$, i.e. risk aversion. Further conditions proposed at society seminars included $u'''(x) > 0$ and $u''''(x) < 0$.

A society member, Pat Brockett, see Brockett and Golden (1987), demonstrated that any utility function with alternating signs in successive derivatives could be expressed as a mixture of exponential functions. When Clive suggested the alternating signs condition, I pointed out Brockett's result and thought it could be applied to exponential distributions as well. Clive implemented it into our increased limits ratemaking methodology. Keatinge (1999) describes his methodology for fitting mixed exponential distributions.

It is worth pointing out that several years later, Lee and Lin (2010) gave a more general result that drops the alternating signs condition and demonstrates that any positive loss distribution can be expressed as a mixture of Erlang distributions.

[8] Note that the Pareto distribution is special case of the mixed exponential distribution. See p. 54 of the Hogg and Klugman text

4.2 Risk Load

I had gained familiarity with the original ISO risk load formula when I represented CNA on the ISO Increased Limits Subcommittee. A common complaint at that time was that the risk load was too high at the higher limits. In fact, another actuary at CNA asked me to show her how to remove the risk load from the increased limits factors.

While in academia, I fell out of the loop with regards to the continuing risk load saga. When I joined ISO I found out that they had changed the formula from where the risk load had been proportional to the pure premium variance, to a formula where the risk load was proportional to the pure premium standard deviation. When applied across different lines of insurance, this risk load produced some counterintuitive reversals. My job was to fix this.

But first, let's step back a little bit. While in academia, I had begun to do research on the price of risk. It started when I joined the CAS Committee on the Theory of Risk (COTOR).

The US Internal Revenue Service had decided to require insurers to deduct their discounted, rather than their statutory (undiscounted) reserves when computing their income tax. Insurers argued that casualty insurance reserves were highly uncertain and that the undiscounted reserves provided a risk margin so that funds would be available in case of an underestimate.

COTOR was investigating more explicit ways of providing a risk margin for loss reserves. We produced a white paper COTOR (1987) titled "Risk Theoretic Issues in the Discounting of Loss Reserves." I thought the report did a nice job of scoping out the problem. It called for the development of a stochastic loss reserve model (nonexistent at the time for casualty insurance) and for a way to evaluate risk with either utility theory or ruin theory. It specifically rejected the Capital Asset Pricing Model (CAPM) which held that public companies should not be concerned with company-specific risk. See for example Butsic (1979).

What was lacking was an example showing how all this would fit together. I was teaching a course in life contingencies at the time. The Society of Actuaries textbook *Actuarial Mathematics* by Bowers *et. al.* provided a good part of what we were missing, only it was for life insurance, not casualty insurance. It provided (1) a long-tailed line of insurance; (2) a stochastic model; and (3) discounting. Missing was a way to treat parameter uncertainty.

In addition to my work at CNA on parameter uncertainty, I drew on what I consider to be one of the best readings I encountered on the actuarial exams — the book *Decision Analysis* (1970) by Howard Raiffa.[9] It was an excellent combination of Bayesian analysis and utility theory.

My intent in writing the paper, Meyers (1989) titled "Risk Theoretic Issues in Loss Reserving: The Case of Workers' Compensation Pension Reserves" was to provide a template for calculating the risk margin along the lines of the COTOR white paper. I fully expected the template to evolve over time.

[9]Operations Research was one of the topics on Part 4 when I was taking the exams.

Features of this paper include:

- Using a simple 3-parameter mortality model.
- Using standard life contingency formulas for process risk.
- Using Bayesian analysis to quantify parameter uncertainty and to get the predictive distributions of pension reserves.
- Using utility theory to calculate the risk margin.
- Using a market-based profit margin to determine parameters of the utility function.

For this particular example, the amount of the discount and the risk margins were very different.

Just as I was beginning my work at ISO I presented this paper at Second International Conference on Insurer Solvency. At the conference, Neal Doherty, who I knew from the Risk Theory Society, challenged my use of utility theory for an insurance company and suggested that the CAPM was more a appropriate measure of risk for an insurance company. The debate was short, and we did not agree. I "knew" full well that insurers care about company-specific risk. They buy reinsurance. As one who had just used market prices to determine the parameters of a utility function, I was open to alternatives to using utility theory to describe a company's risk aversion, but did not have a viable alternative at the time.

It was at this time that I decided to do some "opposition research" and dig deeply into the mathematics of the CAPM.

Back at ISO, I got to know a new colleague, John Cozzolino.[10] As one can see from a search of his papers on the CAS website, we shared a common interest in utility theory and risk loads. In my quest to better understand the mathematics underlying the CAPM, John steered me to an alternative derivation of the CAPM that is in the Appendix of Chapter 7 of Copeland and Weston (1980). Here is a high-level summary of that derivation.

First, here are the CAPM behavioral assumptions:

- Investor chooses a portfolio of securities that will:
  1. Maximize total expected return.
  2. Be subject to a total variance constraint.
- For a given set of securities, and their associated returns, an investor calculates how many shares to buy in each security using the method of Lagrange multipliers.
- The total number of shares bought may not match the total number of shares in the market.

[10] John and I worked together in the Actuarial Development division at ISO for a few years. He later went on to be a professor of risk management at Pace University.

- The CAPM calculates the return of the security that forces a match of the total number of shares bought with the total number of shares in the market.

Once I understood the mathematical underpinnings of the CAPM, I began to think about how I might adapt these ideas in an insurance setting. Let's now consider the difference between securities and lines of insurance making the following analogies.

- An insurance line of business is analogous to a company issuing a security.
- An individual insurance contract is analogous to an individual security, e.g. a share of the company's stock.

The analogy breaks down when we compare the losses on an individual insurance contract, with the return on a company's security. The losses can be different for two different insurance policies, but the return on an individual security is the same for all securities issued by the company. The consequences of this are:

- Let R represent the random return of a single security. If an investor buys n securities, the variance of their total return is

$$Var[R \cdot n] = Var[R] \cdot n^2 \tag{2}$$

  That is to say, the variance of the total return on that security is proportional to the square of the volume, n, bought of that security.

- Let X_E represent the random loss paid on an insurance contract written with exposure, E. Let $E[X_E]$ represents the volume of the insurance contract. Then according to Equation 1 above

$$Var[X_E] = a \cdot E[X_E] + c \cdot E[X_E]^2 \tag{3}$$

  That is to say, the variance of the loss paid on an insurance contract is a linear combination its volume, $E[X_E]$, and its volume squared, $E[X_E]^2$.

As I dug into the underlying math of the CAPM it became clear that one could use Equation 3 in the model in place of Equation 2. That model was used in our revised risk load formula. Here is an outline of the model.

- Insurer chooses an insurance portfolio that will:
  1. Maximize total expected return.
  2. Be subject to a total variance constraint.
- For each line of business, the insurer calculates how much exposure to write in each line of insurance using the method of Lagrange multipliers.
- The total insured exposure may not match the total amount of exposure desired by the market.

- The formula calculates the risk load of the contract that forces a match of the amount of exposure sold with the total exposure desired by the market.

The original paper describing this methodology is in Meyers (1991) titled "The Competitive Market Equilibrium (CME) Formula for Increased Limits Ratemaking." That paper described the theoretical underpinnings of the ISO formula for risk loads. Before elaborating on how we implemented the risk load, I want to describe a second derivation of the same risk load formula that we came up with a few years later, described in Meyers (1996). I believe this second derivation is more intuitive and it better lines up with the way actuaries currently think about risk management.[11]

As one looks through CAS publications of the early 1990's they will find several papers that address the subject of risk loading. A paper that attracted my attention was Kreps (1990) titled "Reinsurance Company Risk Loads from Marginal Surplus Requirements."

It also attracted the attention of Phil Heckman, who wrote Heckman (1992) titled "Some Unifying Remarks on Risk Loads" where he showed that the Marginal Capital and the CME approaches are equivalent. I followed up and derived the CME risk load by the marginal capital approach, in Meyers (1996) titled "The Competitive Market Equilibrium Risk Load Formula for Catastrophe Ratemaking."

Let's look at the behavioral assumptions underlying the marginal capital approach. Let $C[X]$ denote the amount of capital needed by an insurer with loss portfolio, X.

- The insurer will prefer to add Y_1 with expected return $R[Y_1]$ instead of adding Y_2 with expected return $R[Y_2]$ if

$$\frac{R[Y_1]}{C[X+Y_1]-C[X]} > \frac{R[Y_2]}{C[X+Y_2]-C[X]}$$

- Over time, the search for better risks, Y, will reach a point of diminishing returns. Thus this behavior leads to a constant K where

$$\frac{R[Y]}{C[X]-C[X-Y]} = K \tag{4}$$

  for all Y in the insurer's portfolio.

If we make $C[X]$ an increasing function of the variance, i.e. $C[X] = f(Var[X])$, the above behavior leads to a solution of the optimization problem — Maximize total return subject to a variance constraint. To see this, suppose we have contracts i and j with the same marginal capital. If $R[Y_i] > R[Y_j]$, then a portfolio with contract i will have a greater total return than a portfolio with contract j.

[11] In the 30 years since I wrote Meyers (1991), I have yet to hear of an insurance company setting their line of business acquisition targets using Lagrange multipliers.

Now let's derive some expressions for the risk load.

$$\begin{aligned}\frac{R[Y]}{C[X+Y]-C[X]} &= \frac{R[Y]}{f(Var[X+Y])-f(Var[X])} \\ &\approx \frac{R[Y]}{Var[X+Y]-Var[X]} \cdot \frac{1}{f'(Var[X])} \\ &\approx K \text{ the insurer's cost of raising capital.}\end{aligned}$$

This implies that

$$\begin{aligned}R[Y] &\approx K \cdot f'(Var[X]) \cdot (Var[X+Y]-Var[X]) \\ &\equiv \lambda \cdot (Var[X+Y]-Var[X])\end{aligned}$$

We call

$$\lambda = K \cdot f'(Var[X]) \tag{5}$$

the risk load multiplier.

Note that the approximation that we used in deriving Equation (5),

$$f(Var[X+Y])-f(Var[X] \approx f'(Var[X]) \cdot (Var[X+Y]-Var[X])$$

is best when Y is a small addition to the insurer's existing portfolio, X.

Let's break up the insurer's existing portfolio into its individual insurance contracts. Let

$$X = \sum_i X_i$$

Then $Var[X]$ is equal to the sum of the elements in the covariance matrix

$$\{Cov[X_i, X_j]\}$$

If we were to add Y to the insurer's portfolio we would add a row and column consisting of $\{Cov[X_i, Y\}$ to the covariance matrix, with $Cov[Y,Y] = Var[Y]$ in the lower right corner. To get the marginal variance we sum the covariances $\{Cov[X_i, Y]\}$ in the new row and column. With this we can write

$$R[Y] = \lambda \cdot (Var[Y] + 2 \cdot \sum_i Cov[X_i, Y]) \tag{6}$$

The parameter uncertainty underlying the loss model in Equation 3 creeps into the risk load formula by:

$$\begin{aligned}Cov[X_i, Y] &= E_\chi[Cov[X_i, Y|\chi] + Cov_\chi[E[X_i], E[Y]|\chi] \\ &= 0 + E[X_i] \cdot E[Y] \cdot Cov[\chi, \chi] \\ &= c \cdot E[X_i] \cdot E[Y]\end{aligned}$$

when X_i and Y are in the same line of insurance. It is also worth noting that if $c = 0$, Equation 6 reduces to the risk load in Miccolis (1977) — Equation 26.

In the original ISO risk load formula, the risk load multiplier was set so that the total risk load was a judgmentally selected percentage of the total premium. We decided to continue that practice. Then the question arose, what should the selected percentage be? With an explicit C[X] and a given cost of raising capital, K, Equations 5 and 6 allow for a ballpark calculation of the total risk load. Let's suppose that the required capital is a multiple, T, of its portfolio standard deviation. This is $f(Var[X]) = T \cdot \sqrt{Var[X]}$. Then:

$$\lambda = K \cdot \frac{T}{2 \cdot \sqrt{Var[X]}} = \frac{T^2 \cdot K}{2 \cdot C[X]}$$

For an insurer with random loss portfolio, $\{X_i\}$, we have:

$$\begin{aligned}
\text{Total Risk Load} &= \lambda \cdot \left(\sum_j Var[X_j] + 2 \cdot \sum_{j \neq i} Cov[X_i, X_j] \right) \\
&= \lambda \cdot \left(2 \cdot Var[X] - \sum_j Var[X_j] \right) \\
&< \lambda \cdot 2 \cdot Var[X] \\
&= \frac{T^2 \cdot K}{2 \cdot C[X]} \cdot \frac{2 \cdot C[X]^2}{T^2} \\
&= K \cdot C[X]
\end{aligned}$$

This equation puts the upper bound on the total risk load equal to the amount of its capital times its cost of capital. If the off-diagonal entries of the $\{Cov[X_i, X_j]\}$ matrix sums are large compared to the sum of the on-diagonal entries, as should be case with many insureds within each line of business[12], the total risk load should be close to its upper bound.

For a numerical example, suppose $K = 12\%$ and we assume a 2:1 Premium to Capital ratio.[13] Then the total risk load is equal to $K \cdot \text{Premium}/2$=6% of the total premium.

[12] If there are n diagonal entries in a line of insurance, there will be $n^2 - n$ positive off-diagonal entries in that line.

[13] A 2:1 Premium to Capital ratio was common at the time. See Feldblum (1993). Lately, premium to capital ratios have inched below 1:1. Feldblum also reports that the return on capital was a bit below 12%.

Now let's look at an insurance market where there is perfect competition and there are m insurers, indexed by the subscript j, each with their own books of business $\{X_{ij}\}$. When competing for the contract, Y, we have according to Equation 6:

$$\frac{R[Y]}{\lambda_j} = Var[Y] + 2 \cdot \sum_i Cov[X_{ij}, Y]$$

Then summing both sides of this equation over the m insurers and dividing by m yields:

$$R[Y] = \overline{\lambda} \cdot \left(Var[Y] + 2 \cdot \sum_{i=1}^{n} Cov[\overline{X}_i, Y] \right) \tag{7}$$

$$\text{where} \quad \overline{\lambda} = \frac{1}{\frac{1}{m}\sum_{j=1}^{m} \frac{1}{\lambda_j}} \quad \text{and} \quad \overline{X}_i = \frac{1}{m} \sum_{j=1}^{m} X_{ij}$$

Equation 7 is the competitive market equilibrium risk load formula. This is the methodology that we filed with state regulators for the commercial lines. The risk load multiplier was set so that the total risk load was 6% of the total premium.

Now let's turn to the implementation of our new increased limits methodology. The first thing to note is that this methodology was developed as ISO was transitioning from a company that was wholly owned and controlled by the insurance industry to an independent for profit corporation. Before the transition, ISO filed advisory rates. After the transition, ISO filed loss costs, i.e. the expected losses without the expenses and profit.

The filed increased limits factors did contain the risk load, but our circulars were designed to make it easy to modify the risk load multiplier, or remove the risk load entirely. We also produced software that allowed for easy modification of the increased limits factors.

Before the transition, all rates and rating methodologies were dictated by insurer committees. After the transition, all these decisions were made by ISO staff. What governed our decisions was the fact that insurers were not required to license, i.e. pay for the use of, our products. Our products had to be useful to insurers.

My boss, John Kollar and his boss, Phil Miller guided the development and the implementation of this product through the minefield of ISO management, insurer committees, the ISO legal department and the regulators. They hired Clive Keatinge, an actuary with very strong math skills, to work with me on the project. As we neared completion of the project, they had the product reviewed by two external actuaries with Ph.D.s.

This increased limits methodology was one of our first products to roll out under this new environment. As it was a high profile event, it was important that the new methodology be plausible, logically consistent and defensible.

5 Dynamic Financial Analysis

Hurricane Andrew (1992) and the Northridge Earthquake (1994) triggered a major rethinking about how the insurance industry should deal with natural catastrophes. Before these events, insurers typically tried to extrapolate the expected loss historical experience spanning many years. Typically 30 years as I recall.

These events brought to prominence some catastrophe modeling firms such as Applied Insurance Research (AIR) and Risk Management Solutions (RMS). These firms examined long-term weather patterns and geological information on the frequency and intensity of events by location. This combined with engineering information provided insurers with long-term estimates of damages based on the locations of their current insurance portfolio.

ISO decided to replace its existing hurricane ratemaking methodology with one based on a commercially available hurricane model. As catastrophes are risky, I used the output from this model to calculate risk loads. The methodology for these calculations is described in Meyers (1996).

The risk loads I calculated for the hurricane coverage were shockingly high. It was clear that we could not include such a high risk load in our loss costs as we had done for commercial liability. However, the cost of capital needed to support the hurricane exposure was prohibitive, so other ways of managing the hurricane risks needed to be developed. In Meyers (1996) I explored alternatives such as reinsurance and geographic diversification. Each alternative holds promise, but in the mid-1990's the private markets were unable to provide the needed coverage. So government stepped in with facilities such as Florida Hurricane Catastrophe Fund. My sense is that the private markets are gradually getting more involved, but exploring that is beyond the scope of this paper.

The magnitude of catastrophe risk loads initially caused me to doubt the idea that the risk load multiplier should be the same for all risks in an insurer's portfolio. At some point I recognized that a crucial difference between commercial liability and property catastrophe coverages was that an insurer ties up its capital for a longer period of time for the liability lines of insurance, and that this tying up of capital has a cost. In Meyers (1996, p. 574) I attempted to use this fact to justify varying the risk load multiplier. But as the idea of insurers selecting their portfolio according to getting the greatest return on marginal capital was central to the justification of this risk load methodology, I did not like this idea!

Eventually it dawned on me that as losses got paid, the amount of capital needed to support the portfolio could be returned to the company owners as the need for capital diminishes. One can calculate the actuarial present value of the returned capital in a way similar to that used in Meyers (1986) to calculate the present value of the retrospective premium.[14]

[14]Usually, the returned capital will be reinvested in the insurance company enabling it to write new business.

Here is a model that describes the cash flow for the return of capital.

- Let:
  - i be the risk-free rate of return in investments.
  - r be the risky rate of return that investors demand from the insurer.
  - C_t be the amount of capital required to write the risk at the end of t years.
- At time $t = 0$ the investors provide C_0.
- At time $t = 1$ the insurer calculates C_1. Then $C_0 \cdot (1 + i) - C_1$ is returned to the investors.
- $\cdots$
- At time t the insurer calculates C_t. Then $C_{t-1} \cdot (1+i) - C_t$ is returned to the investors.
- $\cdots$
- The present value of this cash flow to investors is:

$$C_0 - \sum_{t=1}^{\infty} \frac{C_{t-1} \cdot (1+i) - C_t}{(1+r)^t} = (r-i) \cdot \sum_{t=0}^{\infty} \frac{C_t}{(1+r)^{t+1}} \tag{8}$$

Similar to reasoning for Equation 4, over time insurers will settle on a risky rate of return, r, that will satisfy investors and attract business. Equation 8 is the risk load that results from considering how long an insurer holds capital.

Now let's turn to a development in the measurement of risk that occurred in the late 1990's. Actuaries have often criticized the standard deviation of the losses as a measure of risk. The objection was that it penalized favorable and unfavorable outcomes equally. While recognizing the problems, my own view was that the standard deviation was the best of the various alternatives, perhaps influenced by the fact that the math, as illustrated above, works out nicely.

So when I heard about the paper, Artzner, *et. al.*(1999) titled "Coherent Measures of Risk," I decided to investigate. As the paper was highly technical, I thought it would be helpful to summarize it for actuaries in Meyers (2000).

After reading this paper, I started using the "Tail Value-At-Risk" (TVaR) as my preferred measure of risk.

$$TVaR_{\alpha}[X] = E[X|X > F^{-1}(\alpha)] \tag{9}$$

where $F(x)$ is the cumulative distribution function for X.

I usually select the threshold, α, to be somewhere in the high 90's as a percent. The choice is tuned to be consistent with existing rules of thumb for adequate capitalization. Over the years, I have developed a healthy respect for the various rules of thumb that have withstood the test of time.

Another development in the late 1990's was that of "Dynamic Financial Analysis" (DFA). Generally stated, DFA aspires to manage the entirety of an insurer's risk and returns. See Szkoda (1997) for an introduction to the topic.

The CAS sponsored DFA call paper programs starting in 1997 and ending in 2003. I submitted papers to most of them. I suggest Meyers (2001) as one to read. Let $APV[\cdot]$ denote the actuarial present value operator. The goal of this paper was to calculate a target combined ratio with the premium defined by:

$$\begin{aligned} \text{Target Premium} &= APV[\text{Loss}] \\ &+ APV[\text{Expenses}] \\ &+ \text{Risk Load (as defined by Equation 8)} \\ &+ \text{Reinsurance Premium} - APV[\text{Reinsurance Recovery}] \end{aligned}$$

Then the target combined ratio is defined by

$$\frac{\text{Nominal Expected Loss} + \text{Nominal Expected Expenses}}{\text{Target Premium}}$$

The paper envisioned using the collective risk model as implemented by the Heckman and Meyers (1983) algorithm to calculate the various C_t^{LOB}s using the $TVaR$ risk measure. It would then allocate the total capital to each C_t^{LOB} in proportion to its marginal capital.

I envisioned insurers using the ideas in this paper to help decide which lines of business to grow in. For example:

- They might ask if the premium they could charge yielded a better than target combined ratio.
- They could also use the ideas in the paper to decide on an appropriate reinsurance strategy. One could compare the net cost of reinsurance with the cost of capital needed to support that line of business.[15]

ISO developed a product based on these ideas. We called it the "Underwriting Risk Model" (URM). The user would supply estimates of the expected loss by line of business and settlement lag. ISO would provide the claim severity and count distributions and the software to produce the output. The product failed to catch on.

While there are probably many reasons for its failure, my personal view was the combination of unfamiliar (and not universally agreed to) concepts and the fact that the product demanded a lot of care (inputs and updates) by the user made the URM a hard sell. Also, it didn't help that some reinsurance brokers were providing similar services on a "pro-bono" basis. I viewed this pursuit as one being done best by staff within an insurance company. I thought it was necessary to understand the nuances of an insurer's decision making process. This was not going to be done by someone from an external organization like ISO.

[15]Based on the exercises I did, I recall finding that for all but the very small insurers, raising capital was generally less expensive than reinsurance for the liability coverages. And reinsurance was almost always less expensive than raising capital for the catastrophe coverages.

Being in my late 50's, liking my work environment, and having a steady stream of other interesting projects, I did not want to jump to another employer. I was ready to give up the Chase.

Or so I thought. What kept me involved was my participation in various CAS/IAA activities.

6 Back to Risk Margins for Loss Reserves

Somewhat surprisingly, to me anyway, the risk margins for loss reserves had reemerged as a hot topic. This section describes the path I took to arrive at a solution to the risk margin problem. This section consists of four subsections, the first three subsections describe the developments that encouraged me to press on to a solution, with the fourth subsection describing this solution.

6.1 A Clear and Public Description of the Problem

As I was busily writing DFA papers and trying to market the URM, I was invited to join the Insurer Solvency Assessment Working Party, sponsored by the International Actuarial Association (IAA). The European Union (EU) was revising their solvency standards[16]and asked the IAA for input. The working party produced the book, IAA (2004), titled *A Global Framework for Insurer Solvency Assessment* which came to be known as the IAA Blue Book.

Upon completion of that book many members of the working party, including myself, went on to represent their actuarial organization on the newly created IAA Solvency Subcommittee, which reported to the IAA Insurance Regulation Committee.[17]

The participating actuarial organizations of the IAA usually sent representatives to the committee meetings who were well-established in their organization's leadership, e.g. presidents (present, past and future). The main body of the IAA was not a research organization. IAA publications were attempts to develop an international consensus on a variety of actuarial issues, and failing that they would non-judgmentally recognize the major points of view. As a researcher, I thought it was good to hear what the actuarial leadership was thinking.

The committee meetings I attended got reports from various groups that had a stake in insurer solvency standards, such as the International Association of Insurance Supervisors (IAIS), the European Insurance and Occupational Pensions Authority (EIOPA), the International Accounting Standards Board (IASB) and the Chief Risk Officers (CRO) Forum.

One of the topics I followed closely was the development of the insurer's liability for unpaid losses, otherwise referred to as the "technical provisions" in the European Solvency

[16]The revised standards, Solvency II, went into effect on January 1, 2016

[17]Individual actuaries do not "belong" to the IAA. Actuarial organizations belong and send representatives to the various IAA committees.

II directive.[18] This directive was first published in 2009, and after a number of amendments, was finally put into effect on January 1, 2016. These provisions are defined as:

1. "The value of the technical provisions shall be equal to the sum of a best estimate and a risk margin."

2. "The best estimate shall correspond to the probability-weighted average of future cash flows, taking account of the time value of money using the relevant risk-free interest rate term structure."

3. "The risk margin shall be calculated by determining the cost of providing an amount of eligible own funds equal to the Solvency Capital Requirement necessary to support the insurance obligations over the lifetime thereof."

4. "Insurance undertakings shall segment their insurance obligations into homogeneous risk groups, and as a minimum by lines of business, when calculating the technical provisions."

The IAA has published two books, IAA (2009) and IAA (2018) to aid actuaries in the implementation of the technical provisions and IFRS 17, with the second book written after IFRS 17 was "finalized."[19]

The IAA, the IAIS, and the IASB have indicated that there are five key desirable characteristics of risk margins:[20]

1. The less that is known about the current estimate and its trend, the higher the risk margins should be.

2. Risks with low frequency and high severity will have higher risk margins than risks with high frequency and low severity.

3. For similar risks, contracts that persist over a longer timeframe will have higher risk margins than those of shorter duration.

4. Risks with a wide probability distribution will have higher risk margins than those risks with a narrower distribution.

5. To the extent that emerging experience reduces uncertainty, risk margins will decrease, and vice versa.

[18]The provisions quoted here are stated in Section 2, Article 77 and Article 80, of Chapter VI of the act, p 222.

[19]Currently, IFRS 17 is scheduled to be implemented on January 1, 2023.

[20]IAA (2009) Executive Summary

Approaches for determining risk margins have been grouped into the following four families of approaches that meet the IASB's current view to have an explicit risk margin:[21]

1. Quantile methods $\cdots$
2. Cost of capital methods based on the amount of return, in addition to the amount earned by the insurer from its investment of capital, that is required for the total return on the insurance enterprise to be adequate.
3. Discount related methods $\cdots$
4. Explicit assumptions $\cdots$

In their conclusion — "The cost of capital method (without simplification) is the most risk sensitive and is the method most closely related to pricing risk in other industries. However, in part as a result, it is also more challenging to implement than the other methods."

Digging deeper, Appendix C of IAA (2009) describes two similar formulas for the cost of capital risk margin.

- The Swiss Solvency Test risk margin — also used in Solvency II

$$M_{SST} = (r - i) \cdot \sum_{t=0}^{\infty} \frac{C_t}{(1+i)^{t+1}} \tag{10}$$

- The Capital Cash Flow risk margin

$$M_{CCF} = (r - i) \cdot \sum_{t=0}^{\infty} \frac{C_t}{(1+r)^{t+1}} \tag{11}$$

where

i = risk-free rate of return on investments.
r = total rate of return demanded by investors for taking on insurance risk.
C_t = amount of capital required to support an insurance portfolio at time $t = 0, 1, \cdots$.

Noting that Equation 11 is the same as Equation 8, my preference is for M_{CFF}.[22] Not wanting to dwell on this difference, I thought it was more important to focus on the bigger issue — the C_ts. It turned out that how to calculate the C_ts was not settled. The key phrase in the IAA's conclusion is "without simplification." In talking with attendees at various IAA/ASTIN meetings, I sensed that most, if not all, EU insurers were using a simplified calculation of the C_ts.[23]

[21] IAA (2009) Executive Summary

[22] For r sufficiently large, M_{SST} will be larger than C_0.

[23] There is always the possibility that insurers were calculating the C_t internally and not revealing their results in public.

Here is an indication of what they were doing. In preparation for the implementation of Solvency II, EIOPA issued a series of "Quantitative Impact Surveys" (QIS) where EU insurers would complete test versions of the Solvency II requirements. The instructions for QIS 5 (2010, p. 59) provides a hierarchy of simplifications for calculating the C_ts.

1. Make a full calculation of all future C_ts without using simplifications.
2. Approximate the individual risks or sub-risks within some or all modules and sub-modules to be used for the calculation of future C_ts.
3. Approximate the whole C_ts for each future year, e.g. by using a proportional approach.
4. Estimate all future C_ts "at once", e.g. by using an approximation based on the duration approach.
5. Approximate the risk margin by calculating it as a percentage of the best estimate.

Sensing that most EU insurers were using some form of the above simplifications, I was attracted to #1. Here we had an actuarial research problem publicly stated with unusual clarity — Given a loss triangle, calculate a cost of capital risk margin. And I had a good idea of what was expected in a solution.

6.2 Stochastic Loss Reserving

Accepting either of the risk margin formulas in Equations 10 or 11, the problem boiled down to calculating the C_ts. When examining the stochastic loss reserve models that I considered to be the publicly available state of the art at that time (early 2000's), e.g. Mack (1994) along with England and Verrall (2002), I didn't find anything that dealt with calculating C_t for $t > 0$. So I decided that at the very least, I would have to dig deeply into stochastic loss reserving.

While I may have been short on experience with stochastic loss reserve models, I had a lot of experience in other areas of actuarial modeling. Drawing on this experience, I approached the project with the following considerations.

- I think of myself to be a Bayesian. Actuaries who specialize in loss reserving stress the importance of judgment. This was a good fit and so I focused on Bayesian models.
- Actuaries have long recognized that loss development patterns often change over time. It has been a standard practice, e.g. Berquist and Sherman (1977), to adjust the data. What is really being done here is changing the model with some hand-selected fixed parameters. Instead, whenever possible, I will choose to change the model to account for systematic changes in the loss environment and allow all parameters to be uncertain.

- One should always test their model on holdout data. A number of papers I had read on stochastic loss reserving followed the pattern where they: (1) proposed a model; (2) did some math to predict a statistic such as a standard deviation or a percentile; and (3) illustrated the calculation on at most a handful of loss triangles. Nowhere did I see any large-scale organized attempt to test predictions of these statistics.

My first attempt at a stochastic loss reserve model was Meyers (2007b) Features of the model include:

- A set of loss development scenarios obtained by maximum likelihood on a compound negative binomial model for 40 large insurers.
- The model then used Bayes' theorem on Schedule P data to get posterior probability weights for each scenario.
- The final model was a posterior probability weighted mixture of the scenarios.
- To validate the model on holdout data, this paper used the model's predictive distributions to calculate the percentile of a sum of losses taken from holdout data. The calculation was repeated for several Schedule P loss triangles. It then performed tests to see if the percentiles were uniformly distributed.[24]

Now let's back up a bit. Starting in 2004, the CAS Committee on the Theory of Risk, issued a series of competitions to fit claim severity distributions. The "data" for the computation was simulated from an unknown model created by Stuart Klugman. I fared well in these competitions with a model consisting of a mixture of preselected grid of vector-valued parameters with the weights determined by Bayes' theorem. Then the estimate for a "statistic of interest," e.g. the expected cost for an excess layer of insurance, would be a posterior probability weighted average of of the conditional expected cost of the layer given the parameters.

At the 2006 CAS Ratemaking Seminar, Klugman showed how he would have solved these problems if he were permitted to enter the competition using Bayesian Markov Chain Monte Carlo (MCMC).

Up to this point I had implemented Bayesian models using either numeric multiple integration, e.g. Meyers (1989), or by posterior probability weighting a parameter grid as in my submissions to the COTOR Challenges. This approach worked well for models with no more than three or four parameters. But stochastic loss reserve models could easily have well over a dozen dimensional parameters and such high dimensional multiple integration or grid searches were (and still are) unworkable in any reasonable amount of time. My way of getting around the multi-dimensional problem in Meyers (2007b) was to fit a model to 40 large insurers by maximum likelihood, then posterior probability weighting the parameters of those fitted models.

[24]Similar tests are also in the two editions of my monograph, Meyers (2015) and Meyers (2019).

Bayesian MCMC is a completely different approach to the multi-dimensional problem. It defines a Markov chain in terms of a prior distribution of parameters, $p(\theta)$ and a conditional distribution $f(x|\theta)$. If one runs the Markov chain for a sufficiently large number of steps, the Markov chain converges to the posterior distribution $\{f(\theta|x)\}$. This means that after a sufficient number of steps, say N_1, the values

$$\{f(\theta_i|x)\}_{i=N_1}^{N_2}$$

consist of a sample of size $N_2 - N_1 + 1$ from the posterior distribution of θ. Given that sample from the posterior distribution of θ one can calculate the posterior distribution of any statistic of interest that depends upon θ. Note that θ can be a vector of any dimension.

I was starting to examine the possibility of using Bayesian MCMC for loss reserving when I looked at[25] Verrall (2007) which showed how to fit the chain ladder loss reserve model using Bayesian MCMC. It was clear to me that Bayesian MCMC was destined to be a prominent modeling tool for stochastic loss reserving. I then applied Bayesian MCMC to the compound negative binomial model in Meyers (2009).

6.3 Dependency Modeling and The CAS Loss Reserve Database

It was about this time that I was invited to join a joint CAS/Australian Institute of Actuaries task force to study correlations between lines of business for stochastic loss reserve models. Another member of the task force was Edward W. (Jed) Frees, who helped introduce the concept of copulas to actuaries, Frees and Valdez (1998).[26]

As the task force was discussing data sources, Jed volunteered the use of the NAIC dabase that the NAIC had been providing to the University of Wisconsin for several years, free of charge. This database contained Schedule P loss triangles for hundreds of American insurers, spanning a period of several years. Recalling my work in Meyers (2007b), I proposed linking together several years of successive Schedule P data to complete the loss triangle, i.e. fill out the lower triangle with holdout data.

This involved a lot of tedious work. Because Schedule P data was compiled net of reinsurance and the companies composing an insurance group had various within group reinsurance arrangements, we thought it was best to compile the data at the insurance group level. When doing this, it was necessary to check the overlapping Schedule P accident years to see if the companies that made up an insurance group changed. With the financial help of the CAS, we hired Peng Shi, then one of Jed's doctoral students, to do the job.

When the job was completed, the CAS was able to persuade the NAIC to allow us to post the data, called the CAS Loss Reserve Database, on the CAS website.

[25] Verrall presented this paper and I presented my Meyers (2007b) paper at the same concurrent session at the 2007 CAS Annual Meeting.

[26] Jed also served as a part-time consultant to ISO as we were developing our Risk Analyzer suite of products.

The task force did not produce a report, but some papers eventually emerged from the task force's work.

- Meyers and Shi (2011) served to notify the actuarial profession that the CAS Loss Reserve Database was available. It illustrated how we thought the database should be used to test proposed stochastic loss reserve models. The two models we tested performed well on the upper triangle (training) data, and performed poorly on the lower triangle (testing) data. As I saw it, the actuarial profession had a challenge on its hands.

- Zhang and Dukic (2013) showed how to fit a bivariate stochastic loss reserve model with the dependency modeled by a copula. They used a Bayesian MCMC model.

- Avanzi, Taylor and Wong (2015) point out that artificial or "illusory" correlations can be generated by poor modeling. To quote their abstract: "1. In any attempt to measure cross-LoB correlations, careful modeling of the data needs to be the order of the day. The exercise will not be well served by rough modeling, such as the use of simple chain ladders, and may indeed result in the prescription of excessive risk margins and/or capital margins. 2. Such empirical evidence as examined in the paper reveals cross-LoB correlations that vary only in the range zero to very modest."

- Meyers (2017) and Meyers (2019) show that given a Bayesian MCMC stochastic loss reserve model for two separate lines of insurance, one can fit a bivariate stochastic model that captures the dependencies between the two lines of insurance. Statistical tests comparing the performance of a bivariate model assuming independence with one that allows for dependence strongly favor the independence assumption, confirming the conclusions of Avanzi, Taylor and Wong (2015).

6.4 A Cost of Capital Risk Margin Formula for Loss Reserves

After writing my Meyers (2009) Bayesian MCMC paper, most of my paper publishing activities were derived from my predictive modeling day job at ISO. My early efforts in Bayesian MCMC were done by programming the MCMC algorithms directly into R. While this helped me to better understand the algorithm, I was advised that the specialized MCMC software was orders of magnitude faster, and so on the side I started using the JAGS software package. And yes, it was a lot faster.

I retired from ISO at the end of 2011. I had no need or desire to go into consulting, teaching or any other employment. Having a well-defined problem, the CAS Loss Reserve Database and familiarity with a new technology applicable to stochastic loss reserving, I continued the Chase, weaving it in with other more traditional retirement activities.

While my ultimate goal was to develop a cost of capital risk margin formula, I sensed an interest in the actuarial community for stochastic loss reserving that was independent of the risk margin problem. And my desire to get the best model that I could before attacking that problem led me to the first edition of my monograph, Meyers (2015).

This monograph examined 200 relatively well behaved loss triangles taken from the CAS Loss Reserve Database. It considered a number of Bayesian MCMC models as well as the Mack chain ladder model and the bootstrap over-dispersed Poisson model as implemented by the R ChainLadder package. For each model and loss triangle, the model predicted the percentile of the sum of losses in the lower triangle test data. A model was deemed "successful" if the 200 percentiles could pass a test for being drawn from a uniform distribution.

The most successful models in that monograph were the MCMC models that allowed for:

- Correlation between accident years for incurred data
- Changing claim settlement rates for paid data

I had planned to address the risk margin problem after finishing the monograph. But fairly quickly, the following considerations intervened.

1. Reading Zhang and Dukic (2013) piqued my interest in addressing the dependency problem with Bayesian MCMC. This led to Meyers (2017).
2. The Bayesian MCMC software of choice among actuaries was shifting to a package called Stan. As I considered it important to distribute easy to run software with my publications, I decided to start using Stan.
3. While the monograph evaluated models by their performance on the 200 loss triangles on the lower triangle test data, I was frequently asked how to compare the fit of Bayesian MCMC models for single loss triangles — without waiting for the lower triangle test data to come in. It turned out that Stan also included a cross-validation package that could compare Bayesian MCMC models with only the training data.

These considerations led me to write a second edition of the monograph — to be completed after I addressed the risk margin problem.

As mentioned in Section 6.1 above, calculating the risk margin boils down to calculating the C_ts in Equations 10 or 11. The problem is that the value of a given C_t depends on the losses that were reported at time $t-1$.

Whatever the model, the output of one of my Bayesian MCMC analyses is a set of 10,000 equally likely lognormal distribution parameters $\{\mu^i_{wd}, \sigma^i_d\}$ for $w = 1, \ldots, 10$, $d = 1, \ldots, 10$ and $i = 1 \ldots, 10,000$. After 10 years, let's define the "ultimate" loss, U_i for each i as the sum of the lognormal means

$$U_i = \sum_{w=1}^{10} e^{\mu^i_{w,10} + (\sigma^i_{10})^2/2}$$

At the end of the current year, $t = 0$, all the U_is are equally likely. Then the capital at the end of the current year is given by

$$C_0 = TVaR_\alpha[\{U_i\}] - E[\{U_i\}]$$

To get C_t for $t > 0$ you have somebody pick a random I from in the range of 1 to 10,000. Then they reveal a set of randomly selected losses, $\{x_{wd}\}$, from the lognormal distributions with parameters μ^I_{wd} and σ^I_d for the next t calendar year diagonals. But they do not reveal the I.

As a thought experiment, you could append the simulated calendar year losses to your existing loss triangle and refit your model using MCMC. You could then calculate the C_t in the same manner as you calculated C_0 above. As this would have to be repeated many times, it would not be practical to actually do this, as fitting MCMC models took several seconds. There is a much faster way. Recall that in Meyers (2007b) I used 40 equally likely sets of parameters taken from other insurers as my prior distribution. Similarly, we can use the 10,000 equally likely parameter sets from the original MCMC sample. Here are more details.

While you may not know the I, you can calculate the posterior probability, P_{it}, of each parameter set, i, given the first t diagonals $\{x_{wd}\}$. You then calculate

$$C_t = TVaR_\alpha[\{U_i\}] - E[\{U_i\}]$$

for $t = 1, \ldots, 9$, where the $TVaR_\alpha$ and the expected value E are calculated with the probabilities $\{P_{it}\}$. You then calculate the risk margin, M, using either Equation 10 or 11.

Since I is selected at random, you should repeat this sequence of steps many (say 10,000) times to obtain a set of risk margins, $\{M\}$. Then the final risk margin is the arithmetic average, $E[\{M\}]$.

In addition to the three considerations mentioned above, the second edition of the monograph, Meyers (2019), shows how to calculate the cost of capital risk margins. In addition, it shows a way to aggregate the risk margins for the various lines of insurance.

And so for me, the Chase ends up with a solution to the problem posed by white paper, COTOR (1987). This solution:

- As called for in the white paper, it makes use of a stochastic loss reserve model.
- Replaces the call for the use of utility/ruin theory with a cost of capital risk margin. Unlike utility theory (at least as I understood it back then), the cost of capital risk margin considers how long it takes the claims to settle.

The white paper rejected the use of CAPM, mainly because of its failure to recognize firm specific risk. But as does the CAPM, this cost of capital risk margin recognizes the role of the market in pricing risk. We see this in the choice of the return on risky investments, r. The MCMC output can be used to provide a set of stochastic cash flows that enable a comparison with other investments. After such a comparison, one may want to change r.

No solution to problems such as this is perfect. The Chase should continue. While this current solution may be temporary, I do hope is that this solution will add to our general understanding of the problem.

7 My Work Environment

When writing the previous sections, I tried to describe my progress in terms of what I knew and was thinking at the time. I think I did a good job of recognizing the sources of the ideas that came together for the Chase. But I don't think that tells the entire story. I consider my work environment to be a major contributor to my success in actuarial research. In this final section, I would like to say something about this environment.

First of all, I got off to a good start in my actuarial career. My stint doing private passenger rate filings, along with the actuarial exams, gave me a good understanding of what actuaries do. My next job, much of which was spent under the supervision of a statistician who was not an actuary, forced me to look at actuarial problems from a different perspective. Success in my early projects led to more interesting projects, assignments and ultimately, jobs.

A second point is that most, if not all, of my research was about problems originating from the industrial (as opposed to academic) insurance community. That being said, I found it to be a good practice to scan what was happening in the academic world and seize upon it when I found something useful. This paper provides some examples.

Working on a variety of projects has a cumulative effect. As I was working on a current project, I often found myself drawing from my experience from earlier, and sometimes related, projects.

For example, the Chase was about assigning a financial value to uncertain insurance losses. Over time that evolved into finding the cost of capital to support an uncertain loss. As I was engaged in the Chase, I was also working on projects involving capital substitutes.

- In the late 1990's ISO hired Nolan Asch, who came to us as the former chief actuary for a reinsurance company. Nolan and I teamed up to produce a suite of reinsurance projects including: (1) A reinsurance exposure rating tool based on the ISO increased limits factors; (2) Fitted property claim severity distributions for property excess of loss reinsurance.

- ISO investigated ways to use our data to provide indices for catastrophe securitization. Our most successful product for this was the PCS Catastrophe Index.

  My role on the team was to quantify basis risk, which involved passing scenarios from a catastrophe model through sample insurer portfolios and/or through a proposed index and then evaluate the effect on the need for capital. There are two publicly available excerpts from the work I did on this: (1) A paper Meyers (1998); and (2) a report put out by the American Academy of Actuaries, Index Securitization Task Force (1999).

For most of the last decade of my employment at ISO, predictive modeling projects on (what was then called) big data were dominating my workdays. These projects introduced me to a number of different statistical tools such as neural nets, boosted trees, principal components, cross validation, generalized additive models, the Gini index, the R programming language, unsupervised learning and .... In this environment I learned to quickly evaluate and use new modeling tools. This included Bayesian MCMC.

Another aspect of my work environment was traveling to conferences and committee meetings. While at ISO, each year I typically attended one or both of the CAS Spring/Annual meetings, one or two CAS specialty seminars, the ARIA Risk Theory Society meeting, IAA/ASTIN meetings and usually two-plus professional committee meetings. I often had speaking roles at the conferences I attended. ISO was very tolerant and supportive of my travel schedule. I like to think that they benefitted from the increased visibility of ISO and the new insights I brought back to my various projects.

While writing papers and speaking at conferences gave me professional visibility, an equally important benefit of these activities is that they forced me to think more thoroughly and clearly about what I was doing. Before writing a paper, or giving a presentation, I talk a lot to people who are interested in the topic. I usually began writing a paper by first starting to write a slide deck for a presentation. Usually about two thirds of the way through the slide deck, a pattern emerged and I began writing the paper itself. I would really have liked to give some presentations on a topic before writing a paper, but the speaking invitations usually came after I wrote the paper.

The person I talked to most about my projects was John Kollar — my boss for 17 of the 23 years I worked at ISO. We would usually meet, one on one, in the mornings two or three times a week. Our conversations would often run for a half hour or more. Rarely did we get technical — at least in my view. Our talks focused on identifying the key underlying drivers of the project, who else had to get involved, and the next steps. I found John to be an excellent sounding board and collaborator.[27]

The benefits of a good work environment do not come immediately. As I hope this paper illustrates, some of the key elements of the Chase grew out of my work environment dating back over several years. On the flip side, I have found that the benefits of a good work environment continue into retirement. When I officially retired at the end of 2011, my "work" environment changed radically. Fortunately, I had a tankful of benefits from my former environment that lasted through the two editions of my monograph and a few papers. The benefits become stale over time, but they are replaced by the other benefits of retirement.

Finally, I want to give my heartfelt thanks to the many members of the actuarial profession who I have worked with over the years. It has been a great ride.

[27] As President of the CAS, John was the driving force behind the creation of the Certified Enterprise Risk Analyst (CERA) designation.

References

American Academy of Actuaries (AAA) Index Securitization Task Force,"Evaluating the Effectiveness of Index-Based Insurance Derivatives in Hedging Property/Casualty Insurance Transactions," 1999.

Artzner, Phillipe ,Freddy Delbaen,Jean-Marc Eber and David Heath, "Coherent Measures of Risk", *Mathematical Finance* 9 (1999),no. 3, 203-228.

Avanzi, Benjamin, Greg Taylor and Bernard Wong, "Correlations between Insurance Lines of Business: An Illusion or a Real Phenomenon? Some Methodological Considerations," UNSW Business School Research Paper No. 2015ACTL11 and *ASTIN Bulletin* 46:2, 2016, pp. 225?263. doi: 10.1017/asb.2015.31.

Barnett, Glen and Ben Zehnwirth, "Best Estimates for Reserves," Casualty Actuarial Society E-Forum Casualty Actuarial Society - Arlington, Virginia, 1998: Fall, 1-54.

Beard, R.E, T. Pentikäinen, and E. Pesonen, *Risk Theory*, 2nd Edition, Chapman and Hall, 1977.

Berquist, James R. and Richard E. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive Systematic Approach," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society*, Arlington, Virginia, 1977: LXVII, 123-184.

Brockett, Patrick L and Linda L. Golden, 1987. A Class of Utility Functions Containing All the Common Utility Functions," Management Science 33, 955-964.

Bühlmann, Hans, *Mathematical Models in Risk Theory*, Springer-Verlag (1970)

Butsic, Robert P., "Risk and Return for Property-Casualty Insurers," CLRS Transcripts Casualty Actuarial Society - Arlington, Virginia,1979: May 52-83

Cahill, James M., "Deductible and Excess Coverages, Liability and Property Damage Lines, Other than Automobile," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia 1936: XXIII.

CAS Committee on the Theory or Risk (COTOR), "Risk Theoretic Issues in the Discounting of Loss Reserves", *Casualty Actuarial Society Fall Forum*, 1987.

Copeland, Thomas E. and J. Fred Weston, *Financial Theory and Corporate Policy*, Addison-Wesley Publishing Company, Third Printing, May 1980.

England, P.D., and R.J. Verrall, "Stochastic Claims Reserving in General Insurance," *Institute of Actuaries and Faculty of Actuaries*, 28 January 2002.

Feldblum, Sholom, "Statutory Returns on Surplus and the Cost of Equity Capital", *Casualty Actuarial Society Summer Forum*, 1993.

Frees, Edward W. and Emiliano A. Valdez, "Understanding Relationships Using Copulas",*North American Actuarial Journal*, 1998, Vol 2, No 1, pp. 1-25.

Hachemeister, Charles A.,"Breaking Down the Loss Reserve Process," Casualty Loss Reserve Seminar Transcript, *CLRS Transcripts Casualty Actuarial Society - Arlington, Virginia*, 1976

Heckman, Philip E. and Glenn G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Count and Claim Severity Distributions"*Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia. 1983: LXX, 22-61

Heckman, Philip E. "Some Unifying Remarks on Risk Loads" *Casualty Actuarial Society Spring Forum*, 1992

Hogg, Robert V. and Stuart A. Klugman, *Loss Distributions*, John Wiley and Sons, 1984.

International Actuarial Association, *A Global Framework for Insurer Solvency Assessment*, 2004.

International Actuarial Association, *Measurement of Liabilities for Insurance Contracts:Current Estimates and Risk Margins*, 2009.

International Actuarial Association, *Risk Adjustments for Insurance Contracts under IFRS 17*, 2018.

Keatinge, Clive L., "Modeling Losses with the Mixed Exponential Distribution," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia, 1999: LXXXVI, 654-698

Klugman, Stuart A. "Bayesian Estimation of Parameters: Advantages and Practical Examples," Presentation at the 2006 CAS Ratemaking Seminar.

Kreps, Rodney, "Reinsurer Risk Loads From Marginal Surplus Requirements" *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia, 1990: LXXVII, 196-203

Lee, S. C. and Lin, X. S., "Modeling and evaluating insurance losses via mixtures of Erlang distributions," North American Actuarial Journal, 2010,14(1):107-130.

Mack, Thomas, "Measuring the Variability of Chain Ladder Reserve Estimates" Casualty Actuarial Society E-Forum Casualty Actuarial Society - Arlington, Virginia, 1994: Spring, Vol 1,101-182.

Meyers, Glenn G., "An Analysis of Retrospective Rating," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia,1980: LXVII,110-143

Meyers, Glenn and Nathaniel Schenker, "Parameter Uncertainty in the Collective Risk Model"*Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia. 1983: LXX, 111-143

Meyers, Glenn G., "The Cash Flow of a Retrospective Rating Plan,"*Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia, 1986: LXXIII, 113-128

Meyers, Glenn G., "Risk Theoretical Issues in Loss Reserving: The Case of Workers' Compensation Loss Reserves,"*Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia, 1989: LXXVI, 171-192.

Meyers, Glenn G., "The Competitive Market Equilibrium Formula for Increased Limits Ratemaking,"*Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia, 1991: LXXVIII, 163-200.

Meyers, Glenn G., "The Competitive Market Equilibrium Formula for Catastrophe Ratemaking,"*Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia, 1996: LXXXIII, 563-600.

Meyers, Glenn G., "A Buyer's Guide to Options and Futures on a Catastrophe Index," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia,1998: LXXXV 187.

Meyers, Glenn G., "Coherent Measures of Risk: An Exposition for the Lay Actuary" *Casualty Actuarial Society E-Forum*, Casualty Actuarial Society - Arlington, Virginia, 2000: Summer. (Note, this paper is not referenced directly in the E-Forum table of contents. One can access it through link in my web-based "paper" listed there.)

Meyers, Glenn G., "The Cost of Financing Insurance" Casualty Actuarial Society E-Forum Casualty Actuarial Society - Arlington, Virginia, 2001: Spring, Vol 1, 221-264.

Meyers, Glenn G., "The Common Shock Model for Correlated Insurance Losses," Variance 1:1, 2007a, pp. 40-52.

Meyers, Glenn G., "Estimating Predictive Distributions for Loss Reserve Models,"Variance 1:2, 2007b, pp. 248-272.

Meyers, Glenn G., "Stochastic Loss Reserving with the Collective Risk Model,"Variance 3:2, 2009, pp. 239-269.

Meyers, Glenn G. "Pure Premium Regression with the Tweedie Model," Presentation at the November 2009 CAS Annual Meeting.

Meyers, Glenn G. and Peng Shi, "The Retrospective Testing of Stochastic Loss Reserve Models," *Casualty Actuarial Society E-Forum*, Summer 2011.

Meyers, Glenn G. 2015. "Stochastic Loss Reserving Using Bayesian MCMC Models," *CAS Monograph Series*, Number 1.

Meyers, Glenn G. "Dependencies in Stochastic Loss Reserve Models," *Variance*, Casualty Actuarial Society - Arlington, Virginia 2017: Volume 11, Issue 01, 74-94.

Meyers, Glenn G. 2019. "Stochastic Loss Reserving Using Bayesian MCMC Models (2nd Edition)," *CAS Monograph Series*, Number 8.

Miccolis, Robert S., "On the Theory of Increased Limits and Excess of Loss Pricing," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society - Arlington, Virginia, 1977: LXIV, 27-59*

Mong, Shaw, "Estimating Aggregate Loss Probability and Increased Limits Factor" *Casualty Actuarial Society Discussion Paper Program*, Casualty Actuarial Society - Arlington, Virginia, 1980: May 358-393

Raiffa, Howard, *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison Wesley Publishing Company, 1970.

Patrik, Gary, "Estimating Casualty Loss Amount Distributions," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia 1980: LXVII, 57

Scollnik, D. P. M., "Actuarial Modeling with MCMC and BUGS," *North American Actuarial Journal*, 5:2, 2001, pp. 96 - 124.

Stern, Philipp K., "Current Ratemaking Procedures for Private Passenger Liability Insurance," *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society* - Arlington, Virginia,1956: XLIII,112-165.

Szkoda, Susan, "How DFA Can Help the Property/CasualtyIndustry," *The Actuarial Review*,Volume 24,No. 1,1997.

Verrall, Richard,"Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion," *Variance* 1:1, 2007, pp. 53-80.

Willett, Allan H., *The Economic Theory of Risk and Insurance*, Ph.D. Thesis, Columbia 1901. Published for the S.S. Huebner Foundation for Insurance Education in 1951 by Richard D. Irwin, Inc. Homewood, Illinois.

Estimating Working Life Expectancy from Cohort Change Ratios: An Example using Major League Pitchers

David A. Swanson, Jack Baker, Jeff Tayman and Lucky M. Tedrow

Abstract. Census survival methods are the oldest and most widely applicable methods of estimating adult mortality and for populations with negligible migration they can provide excellent results. In the form of cohort change ratios, this approach can be used to estimate working life expectancy without constructing a life table. We describe this approach and illustrate it with an application to the career length of major league baseball pitchers, a group for which little information on working life expectancy is available. This lack of information is surprising given the detailed historical data on baseball players and the large number of working life tables for various occupations. This deficit may be due at least in part to an assumption that the process needed to construct MLB working life tables is such a demanding task that it discourages many from the attempt, a situation that applies to other career groups, especially in sports. If so, we believe this paper shows that such an assumption needs to be re-examined by showing how easy it is to use the cohort change ratio method to estimate working life.

Keywords. Career length, professional sports, cohort change ratios

1. INTRODUCTION

As noted *in Methods for Estimating Adult Mortality from Census Data (United Nations* (2002: 5), "Census survival methods are the oldest and most widely applicable methods of estimating adult mortality…(and can) provide excellent results (for) populations that experience negligible migration…" The reason for the ubiquity of this approach is threefold: (1) data requirements are minimal in that only two successive age distributions are needed; (2) the two successive age distributions are usually easily obtained from census counts; and (3) the method is straightforward in that it requires neither a great deal of judgment nor "data-fitting" techniques to implement. This ubiquity is in contrast to other methods, such as "Model Life Tables", which require more data, as well as judgment and, often, data fitting (United Nations, 1982: 16-27). Our purpose in this paper, however, is not to debate the relative merits of these and other approaches. Our purpose here is to simply demonstrate how to calculate life expectancy from census survival rates, or in the more general form, which we prefer, "cohort change ratio" (CCR)

In the paper that follows, we first describe A CCR in general, describe how it can be used to generate population projections and then describe how CCRs can be used to construct life tables. Following these methodological discussions, we construct and discuss a working life table for MLB pitchers and conclude the paper with remarks on the results and suggestions for

future research and applications.

2. Cohort Change Ratios

As shown by Baker et al. (2017) CCRs have a wide range of applications. When migration is negligible, they can be used to construct life tables and calculate life expectancy (Baker et al., 2017: 165-171). These same CCRs also can be used to generate forecasts (Baker et al., 2017: 45-58).

A cohort change ratio (CCR) is typically computed from age-related data in the two most recent censuses (Baker et al. 2017: 2):

$$ {}_nCCR_{x,\,t} = {}_nP_{x,t} \,/\, {}_nP_{x-k,\,t-k.} \qquad [1] $$

where,

${}_nP_{x,\,t}$ is the population aged x at the most recent census (t),

${}_nP^{x-k,\,t-k}$ is the population aged x-k at the 2nd most recent census (t-k), and

k is the number of years between the most recent census at time t and the one preceding it at time t-k.

As implied by Eq. [1], a cohort change ratio is not typically computed for a single cohort, but for all of the cohorts found in two successive census counts.

Given the nature of the CCR, 10-14 is the youngest five-year age group for which CCRs as defined in Eq. [1] can be made if there are 10 years between censuses. To analyze age groups younger than ten in a given application, a Child-Adult Ratio (CAR) can be used. This ratio, computed separately for ages 0-4 and ages 5-9, relates young children to adults in the age groups most likely to be their parents (Baker et al. 2017: 3, Smith et al. 2013: 178).

The open-ended age group uses the same approach found in life table construction (Baker et al. 2017: 3), and its CCR differs slightly from those for the age groups beyond age 10 up to the oldest open-ended age group. If for example the final closed age group is aged 70-74, with persons aged 75+ as the terminal open-ended age group, then calculation for the CCRx+,t requires the summation of the three oldest age groups to get the population age 65+ at time t-k:

$$ {}_\infty CCR_{75,t} = {}_\infty P_{75,t} \,/\, {}_\infty P_{65,t-k}. \qquad [2] $$

3. Using CCRs to Estimate Life Expectancy

Baker et al. (2017: 165) show that when migration is negligible a CCR can be interpreted as a census survival ratio, which means the expectation of life at age x can be computed as:

$$e_x = (T_x / l_{(n/2)}) / (l_x / l_{(n/2)}) = T_x / l_x \quad [3]$$

where,

x is age,

n is the width of the age groups (up to, but not including the terminal, open-ended age group),

e_x is the life expectancy (average years remaining) at age x,

T_x is the total person years remaining to persons age x,

l_x is the number of persons reaching age x,

$l_{(n/2)}$ = persons aged x to $x+n$ are assumed to be concentrated at the mid-point of the age group, and

$$l_{(x+n/2)} / l_{(x-n/2)} = P2_{(x,n)} / P1_{(x-n,n)} \quad [4]$$

where,

$P2_{(x,n)}$ is the number of persons counted in the second census in age group x to $x+n$, and

$P1_{(x-n,n)}$ is the number of persons counted in the first census in age group $x-n$ to x.

In general, the life-table probability of surviving from the mid-point of one age group to the next ($l_{(x+n/2)} / l_{(x-n/2)}$) is approximated by the census survival ratio ($P2_{(x,n)} / P1_{(x-n,n)}$). Continuing, the cumulative multiplication of the probabilities shown in [4] gives the conditional survival schedule ($l_x / l_{(n/2)}$). From the conditional l_x values given by [4] the conditional estimates of the number of person years lived in each age group ($_nL_x$) can be calculated as:

$$_nL_x / l_{(n/2)} = (n / 2) \times [(l_x / l_{(n/2)} + l_{(x+n)} / l_{(n/2)}] \quad [5]$$

where,

$_nL_x$ is the number of person years lived in each age group.

Given a value of $T_x / l_{(n/2)}$ for some initial age x, total remaining years expected at age x (T_x) values can be calculated as:

$$T_{(x-n)} / l_{(n/2)} = T_x / l_{(n/2)} + {}_nL_{(x-n)} / l_{(n/2)}. \quad [6]$$

This leads us back to equation [3], so that the expectation of life at age x is:

$$e_x = (T_x / l_{(n/2)}) / (l_x / l_{(n/2)}) = T_x / l_x.$$

Extending this approach, we note that when the radix of a life table is equal to 1 (l_0 = 1.00), then life expectancy at birth can be computed directly from the expression:

$$e_0 = S_0 + (S_0 \times S_1) + (S_0 \times S_1 \times S_2) +,..., + (S_0 \times S_1 \times S_2,..., \times S_x) \quad [7]$$

where,

e_0 is life expectancy at birth,

S_0 is the survivorship from $t = 0$ (e.g., birth) to $t = 1$(e.g., age 1),

S_1 is the survivorship from $t = 1$ (e.g., age 1) to $t = 2$(e.g., age 2), and so on through S_x, and

S_x is $_1L_x / {}_1L_{(x-n)}$.

Equation [7] is set up for single year age groups. However, we can generalize it to other age groups: ${}_nS_x = {}_nL_x / {}_nL_{(x-n)}$, so that:

$$e_0 = {}_nS_0 + ({}_nS_0 \times {}_nS_1) + ({}_nS_0 \times {}_nS_1 \times {}_nS_2) +,...,+ ({}_nS_0 \times {}_nS_1 \times {}_nS_2,..., \times {}_nS_x) \qquad [7.a]$$

As equations [7] and [7.a] both imply, the fundamental life table function is inherent in our method; that is, via the ${}_nS_x$ values, we have ${}_nq_x$ values. In summary, our approach is the result of combining either equation [7] or [7.a] for computing life expectancy with equation [1] to estimate e_x. Broadly speaking, the method can be applied to any population subject to renewal through a single increment (entry into the major leagues) and extinction through a single decrement (exit from the major leagues), where there are at least two successive counts that provide the population by some measure of time (consecutive years in the major leagues).

4. A CCR-Based Working Life Table: Career Length of MLB Pitchers

Although baseball is the "national sport," only a handful of studies have examined its career prospects in terms of working life. Using data from 1902 to 1993, Witnauer et al. (2007) found that non-pitching rookie position players can expect to play 5.6 years. Abel and Kruger (2005) examined the overall life expectancy of MLB players by position for the period 1909-1919 and found that pitchers could expected to live to age 67.7, which exceeded the life expectancy of males (63.4) for this same period. While they did not construct a life table, Hardy et al. (2017) used a regression-based approach to assess the determinants of career length among major league pitchers playing between 1989 and 1992 and found that mean career length was 10.97 years. Truncated careers can be attributed to several factors, including injury, poor performance, and scandals (Gutman 1992, Hardy et al. 2017).

Using the method described in the preceding section and 1980 and 1981 data found in the eighth edition of *Total Baseball* (Thorn 2004), Table 1 shows the working life expectancy of major league pitchers who entered the major leagues in 1980.[2] This table shows the number of pitchers who "survive" (l_x) by season (year x, where $x = 0$ to $10+$) starting with the 296 pitchers who first entered MLB in 1980 (l_0). Of these 296 pitchers, 247 completed the first year of play and only 109 (44%) completed five consecutive years. The (S_x) column shows the proportion surviving through each year, which can be interpreted as the probability of making it through the entire season. The probability of making it through the first year is (S_0) = 0.83333 (247 / 296).

Corresponding to S_0, the number of "exits" (d_0) in the first year among the 296 initial pitchers is 49, which corresponds the probability of "exiting" in the first year is (q_0) = 0.166667 (49 /296). At the start of each year, the expected remaining consecutive years of pitching is provided. For the initial 296 pitchers, their expected working life is

3.990 years. At the completion of the first year, those still pitching can expect to do so for 3.157 more years. Those who complete five consecutive years can expect to pitch for 1.172 more years.

Table 1. Major league pitcher working years life table, 1980-1981

Consecutive years in the majors	Probability of pitching next year S_x	Probability of not pitching next year q_x	Number of pitchers l_x	Pitchers leaving the majors d_x	Pitching years expectancy e_x
0	0.83333	0.16667	296	49	3.990
1	0.70723	0.29277	247	72	3.157
2	0.76923	0.23077	175	40	2.568
3	0.93939	0.06061	135	8	2.114
4	0.85714	0.14286	127	18	1.688
5	0.85714	0.14286	109	16	1.172
6	0.90000	0.10000	93	9	0.859
7	0.83333	0.16667	84	14	0.578
8	0.81818	0.18182	70	13	0.343
9	0.78571	0.21429	57	12	0.151
10+	0.00000	1.00000	45	45	0

Source: Thorn (2004)

Note, the number of teams (26) was constant between 1980 and 1982

5. DISCUSSION

The life table (Table 1) shows that the expected working life of MLB pitchers is short and becomes shorter at the completion of each year. Witnauer et al. (2007) found that, in general, a (non-pitching) player who enters MLB can expect to play 5.6 years. Moreover, they found that those who made it through three years could expect to play an additional six years. MLB pitchers, however, who complete three years can expect to play for only 2.114 years more.

The data shown in Table 1 also indicate that there is a high level of volatility in the major league careers of pitchers, especially in the initial years. The factors causing this likely include: (1) injuries that lead to one or more missed seasons; (2) being sent down to the minors for one or more seasons to gain more experience, (one common example is that the initial listing represents a pitcher who is "called up" for a couple of games at the end of the season to "have a cup of coffee," followed by a return to the minors); and (3) either outright release or a decision to quit professional baseball made by the player. Given this volatility, this example

can be viewed as a rather strenuous test of how well the CCR method can perform in subject areas where there is less stability year to year than found in large populations. These areas would include any highly competitive activity such as professional sports. In this regard, while there is substantial variation in the careers of major league such variation can be smoothed by fitting a model to the data. Given that the life table of major league pitchers used survival rates, the Weibull model may be a good candidate (Namboodiri and Suchindran 1987).

Even without smoothing, the CCR method may work reasonably well as a method to estimate the working life expectancy of players in other professional sports by position, such as football (e.g., quarterbacks), basketball (e.g., point guards), hockey (e.g., goalies), and soccer (e.g., strikers). Continuing along this line of reasoning, it may be worthwhile to examine the CCR method in terms of occupational specialties in organizations (e.g., corporations, governments, the armed forces).

Forgetting positions or occupational specialties, the method also may be worthwhile examining in terms of participants classified by years played in a number of activities, including tennis, golf, and NASCAR racing. Keep in mind that at the professional level, some of these sports are affected by luxury taxes, salary caps and other financial restrictions, while others are not (Dietl et al. 2010). As such, the history of the implementation of these measures may be important in terms of constructing a working life table or otherwise estimating expected career length.

In conclusion, we observe that the aim of this paper was to demonstrate that reasonable estimates of MLB working life expectancies can be easily constructed by using the CCR method. If it does change assumptions about these being demanding tasks, perhaps the knowledge gaps concerning working MLB life expectancies can be filled in, an outcome that may serve to help address similar knowledge gaps that exist in other occupations, and in particular, sports, both at the professional and amateur levels.

6. ENDNOTES

1. Although Green and Armstrong (2015) discuss simple vs. complex methods in terms of forecasting, their discussion applied here in that the CCR approach falls into the simple methodological category rather than the complex category. Adapting their discussion to methods in general, the work of Green and Armstrong (2015) suggests that while there is no evidence that shows complexity improves accuracy, complexity remains popular among: (1) researchers, because they are rewarded for publishing in highly ranked journals, which favor complexity; (2) methodologists, because complex methods can be used to provide information

that support decision makers' plans; and (3) clients, who may be reassured by incomprehensibility.

2. The addition of two teams (Colorado Rockies and Florida Marlins) to the National League in 1993 may have extended slightly the career of some of the pitchers who had played consecutively since 1981. Because we used ten as the terminal, open-ended consecutive years played, there is, however, no effect on the working life table by this expansion. Similarly, there is no effect on the working life table by subsequent expansions.

7. REFERENCES

Abel, E. and M. Kruger. (2005). Longevity of major league baseball players. *Research in Sports Medicine* 13: 1-5.

Baker, J., Swanson, D., Tayman, J., and Tedrow, L. (2017). *Cohort change ratios and their applications*. (2017). Springer: Dordrecht, The Netherlands.

Dietl, H., M. Lang, and S. Werner. (2010). The effect of luxury taxes on competitive balance, club profits, and social welfare in sports leagues. *International Journal of Sport Finance* 5(1), 41-51.

Green, K. and J.S. Armstrong (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research* (68): 1678-1685.

Gutman, D. (1992). *Baseball Babylon*. New York: Penguin Books.

Hardy, R., T, Ajibewa, R. Bowman, and J. Brand. (2017) Determinants of major league baseball pitchers' Career Length. *The Journal of Arthroscopic and Related Surgery* 33, 445-449.

Namboodiri, K. and C. Suchindran. (1987). *Life table techniques and their applications*. Academic Press: San Diego, CA.

Saint Onge, J., R. Rogers, and P. Krueger. (2008). Major league baseball players' life expectancies. *Social Science Quarterly* 89(3), 817-830

Sarna, S. and J. Kaprio (1994). Life expectancy of former elite athletes. *Sports Medicine* 17, 149–151.

Sarna S.,T. Sahi, M. Koskenvuo, and J. Kaprio. (1993). Increased life expectancy of world class male athletes. *Medicine and Science in Sports and Exercise* 25, 491–495.

Schnohr, P. (1971) Longevity and causes of death in male athletic champions. *Lancet* 2, 1364–1365.

Smith, S. K., J. Tayman, J., & D. Swanson. (2013). *A practitioner's guide to state and local population projections*. Springer: Dordrecht, The Netherlands.

Thorn, J. (2004). *Total baseball: The ultimate baseball encyclopedia (8th edition)*. Sport Media Publishing.

United Nations. 2002. Methods for Estimating Adult Mortality. New York, NY: Population Division, United Nations.

United Nations. 1982. Model Life tables for Developing Countries. New York, NY: Population Division, United Nations.

Witnauer, W., R. Rogers, and J. Saint Onge (2007). Major league baseball career length in the 20th century. *Population Research and Policy Review* 26, 371-386.

AUTHORS

David A. Swanson
Department of Sociology
University of California Riverside
Riverside, California USA 92521
dswanson@ucr.edu

Jack Baker
HealthFitness Corporation
Minneapolis, Minnesota USA 55431
quantmod@gmail.com

Jeff Tayman
University of California San Diego
San Diego, California USA 92093
jtayman@ucsd.edu

Lucky M. Tedrow
Department of Sociology
Western Washington University
Bellingham, Washington USA 98225
Lucky.tedrow@wwu.edu