Using Generalized Linear Models to Develop Loss Triangles in Reserving

BY ZHIHAN JENNIFER ZHANG, DR. JELENA MILOVANOVIC, MELISSA TOMITA, DR. JOHN ZICARELLI

ARIZONA STATE UNIVERSITY

OCTOBER 1, 2018

1. Notation

Notation used to describe reserving methods vary from paper to paper, but, for the remainder of the article, the notation in table 1 will be used:

Table 1: *Notation*

Notation	Meaning
w	Accident year
d	Development year (age)
t	Calendar year
$c(w, d)$	Cumulative loss from accident year w at age d
$q(w, d)$	Incremental loss from accident year w at age d
α_w	Base value for accident year w
ι_t	Trend for calendar year t
γ_d	Trend for development age d

2. The Probabilistic Trend Family (PTF)

2.1. Barnett and Zehnwirth's Idea

The use of generalized linear models in loss reserving is not new; many statistical models have been developed to fit the loss data gathered by various insurance companies. The most popular models belong to what Glen Barnett and Ben Zehnwirth in "Best Estimates for Reserves" call the "extended link ratio family (ELRF)," as they are developed from the chain ladder algorithm used by actuaries to estimate unpaid claims.

Although these models are intuitive and easy to implement, they are nevertheless flawed because many of the assumptions behind the models do not hold true when fitted with real-world data. Even more problematic is that the ELRF cannot account for environmental changes like inflation that are often observed in the status quo. Barnett and Zehnwirth conclude that a new set of models that contain parameters for not only accident year and development period trends but also payment year trends would be a more accurate predictor of loss development.

Called the "probabilistic trend family" in their paper, these models are designed to account for trends in not only the accident year and development year directions, but also the calendar/payment year direction. The general form of the model is as follows:

$$\log q(w,d)=a_w+\sum_{j=1}^{t-1}\iota_j+\sum_{k=1}^{d-1}\gamma_k \tag{1}$$

Recall that $q(w,d)$ denotes the incremental payment in accident year w and development age d, α_w gives a "base value" for accident year w, ι_j represent calendar year trends, and γ_k stand for development year trends.

This undergraduate thesis project applies the paper's ideas to data gathered by Company XYZ. The data was fitted with an adapted version of Barnett and Zehnwirth's new model in R, and a trend selection algorithm was developed to accompany the regression code. The final forecasts were compared to Company XYZ's booked reserves to evaluate the predictive power of the model.

2.2. Simple Example

To illustrate the process of estimating parameters for a model in the PTF family, we generated a simple example where the trends are easy to identify. Suppose we had an incremental loss triangle that had the following values on a log-scale:

Table 2: *Simulated log-transformed incremental loss triangle*

	Months 12	Months 24	Months 36	Months 48
2015	1	2	3	6
2016	2	3	6	
2017	3	6		
2018	6			

From the triangle, we can observe two calendar year trends. The first starts in calendar year 2015 and continues to calendar year 2017, increasing by one each year. The second starts in calendar year 2017 and ends in calendar year 2018, increasing by three each year.

Using the notation in Equation 1, the loss triangle would look as such:

Table 3: *Simulated log-transformed incremental loss triangle with notation*

	Months 12	Months 24	Months 36	Months 48
2015	$1=\alpha$	$2=\alpha+\iota_1$	$3=\alpha+2\iota_1$	$6=\alpha+2\iota_1+\iota_2$
2016	$2=\alpha+\iota_1$	$3=\alpha+2\iota_1$	$6=\alpha+2\iota_1+\iota_2$	
2017	$3=\alpha+2\iota_1$	$6=\alpha+2\iota_1+\iota_2$		
2018	$6=\alpha+2\iota_1+\iota_2$			

We then can represent the value in each cell as follows:

Table 4: *Predictors and response variables for simulated example*

Months	# α	# ι_1	# ι_2	$log\ q(w, d)$
12	1	0	0	1
	1	1	0	2
	1	2	0	3
	1	2	1	6
24	1	1	0	2
	1	2	0	3
	1	2	1	6
36	1	2	0	3
	1	2	1	6
48	1	2	1	6

From this table, we can see that this is a regression problem with three predictors and response. In other words, we can fit the data with the following equation:

$$\hat{q}=\alpha+\iota_1 x_1+\iota_2 x_2, \tag{2}$$

where x_1 and x_2 denote the number of ι_1 and ι_2, respectively. Performing the regression confirms that $\iota_1 = 1$ and $\iota_2 = 3$.

2.3. Preliminary Problems and Potential Solutions

The previous example, while simple, illustrates the rationale behind the PTF as well as the process an analyst might take to estimate the parameters for a model in the family. However, when faced with real-world data, several complications arise.

2.3.1. Incremental Values are Linear on a Log Scale

As Barnett and Zehnwirth note, "trends in the data on the original dollar scale are hard to deal with, since trends on that scale are not generally linear ... it is the logarithms of the incremental data that show linear trends."[1] Thus, we would need to log-transform our incremental loss triangles before attempting to the fit the model.

However, while cumulative payments are always positive, incremental payments can occasionally be negative values (especially near the tail). Since we cannot log-transform negative values, we would have to adjust our data to accommodate these values. Shapland describes three potential ways of doing so:

- "Zero out" negative values. That is, if the incremental payment is negative, we assume that value is 0 after log-transforming the remaining values.
- Replace the value with - log(-$q(w, d)$) instead of log($q(w, d)$).
- Shift all the incremental values so that no negatives remain before taking the natural logarithm. After analysis, these values would need to be shifted back.

These adjustments can produce slightly different results and can be implemented in R.

2.3.2 Selection of Trends Can Be Difficult

In the example above, we could determine by observing the original triangle that there were two calendar year trends affecting the payments. However, with real-world data, the location of these trends may be difficult to identify, especially if trends are present in all three directions. Again, there are several methods by which we can determine the trends:

- **By inspection.** In Barnett and Zehnwirth's example, the data was first fitted with a basic model in which the analyst assumed there was one trend in each direction. The residuals of this model were plotted against the development year, accident year and calendar year indices, and trends were identified through these residual plots. Because the trends are found by inspecting these plots, this method can produce different results depending on the analyst.
- **By performing best subset selection (i.e., trying every combination).** We could hypothetically fit the data with every possible combination of trends. This can be computationally difficult, however, especially for large triangles — for a triangle with m accident years and n development periods, there would be $2^{(m-1)+(n-1)+(\max(n,m)-1)}$ such combinations. In our example, $m = n = 20$, so we would have to test

[1] Barnett, Glen, and Ben Zehnwirth, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society*, 2000, Vol. 87, pp. 245–321.

$2^{3(20-1)} = 2^{57}$ combinations.

- **By performing stepwise selection.** This method would choose the combination of predictors that minimizes the Akaike Information Criterion (AIC). The AIC statistic rewards goodness of fit, but has a penalty for increasing the number of parameters. Thus, using the AIC to select trends can prevent overfitting.

In the example below, the third method is used.

2.3.3. Projecting Calendar Year Trends

Finally, the purpose of this process is to arrive at estimates for ultimate losses. However, this requires developing estimates for losses in future calendar years, which may involve trends that we have not and cannot observe in the data. There are two ways to account for these trends:

- Extend the most recent trend into the future. We can make the broad assumption that calendar year trends will remain unchanged and extend the most recent trend to apply to future calendar years.
- Assign future calendar year trends based on external research. This can be complicated, however, as some of the calendar year trends may be absorbed by development year and accident year trends.

This project uses the first method to account for future calendar year trends.

3. Fitting Real-World Data to PTF Models

3.1. The Data

With this methodology in mind, we then proceeded to use the probabilistic trend family to estimate ultimate losses for a line of business. We were given a cumulative incurred losses triangle from Company XYZ for a long-tailed line, and we also were given their booked reserves as of December 31, 2016, and December 31, 2017. This information not only allows us to use regression to arrive at ultimate losses but also gives us an example to compare our final results against as of December 31, 2016, and December 31, 2017.

The dataset we were given was particularly interesting because the company had experienced what was functionally a change in claims handling procedures in 2015. Assuming that this calendar year effect was significant, our model should be able to account for the effect it had on loss development.

Because the procedure described in section 2.2 above can be time- and labor-intensive, we developed two functions in R that can perform the analysis automatically. These two functions are not shown here, but they automate the aforementioned process, performing stepwise selection to choose cutpoints for the trends and selecting the trends with a generalized linear model.

3.2. Evaluating the Results

Figure 1 shows the resulting comparisons. Because we know what the booked reserves for Company XYZ were as of both December 31, 2016, and December 31, 2017, we were able to compare our ultimate losses to both estimates and see how the differences changed over time. In both figures, numbers are given in thousands, and differences of greater than three million are highlighted. As the figure shows, our model generally predicts higher ultimate losses than the booked reserves.

Notice that the differences between the ultimate loss

Figure 1: *Comparison of the model's predicted ultimate losses using a full triangle against Company XYZ's booked reserves as of year-end 2016 and 2017. Numbers shown in thousands. Differences of over 3 million are highlighted.*

Full Triangle						
Incurred Year	**Model Ult**	**2016 Ult**	**Diff**	**2017 Ult**	**Diff**	**Change in Diff**
2005	50,776	49,397	1,379	49,256	1,520	141
2006	52,697	52,081	616	51,902	795	179
2007	55,679	57,900	(2,221)	57,565	(1,886)	335
2008	53,914	52,840	1,074	52,781	1,133	59
2009	59,402	58,449	953	58,876	526	(427)
2010	48,218	46,260	1,958	45,090	3,128	1,170
2011	40,125	37,598	2,527	37,305	2,820	293
2012	45,703	42,798	2,905	40,916	4,787	1,882
2013	53,135	49,801	3,334	47,895	5,240	1,906
2014	71,277	62,001	9,276	66,575	4,702	(4,574)
2015	61,729	54,329	7,400	59,405	2,324	(5,076)
2016	59,027	56,704	2,323	56,378	2,649	326
		Tot Abs Diff	35,967	**Tot Abs Diff**	31,511	(4,456)
		Avg Abs Diff	2,997	**Avg Abs Diff**	2,626	(371)
	DY =	2, 3, 7, 8, 9		CY =	NA	
	AY =	9, 10, 11		AIC =	307.848	

estimates for accident years 2013-2016 were particularly large. This coincides with the period in which the company was experiencing changes in claims handling procedures, and may indicate that Company XYZ's method of compensating for those changes could be improved.

4. Conclusion

Using models from the probabilistic trend family (PTF) to predict ultimate losses is an alternative method for reserving that bears exploring. The probabilistic trend family improves upon traditional reserving methods by not only overcoming issues with models in the extended link ratio family but also offering a statistically rigorous way to select trends.

The method described in this paper is one way by which a company can generate a model from the PTF to fit its loss data. Further testing over time would be necessary to judge the predictive power of the model, but the ultimate losses predicted by the model can nevertheless offer insights about what the booked reserves of a particular line of business should look like.