An Extension to the Cape Cod Method with Credibility Weighted Smoothing

Uri Korn, FCAS, MAAA

Abstract

The Cape Cod method is a commonly used technique where the a priori loss ratio is calculated as the weighted average of the chain ladder ultimate loss ratios across all years with the "used" premium as the weights. It applies the same a priori loss ratio estimate (on a trended, current rates level) across all years, without consideration for any possible changes that may have occurred. A difficulty arises when the loss ratios show improvement or deterioration, which is a fairly common scenario. When this occurs, the amount of credibility that should be given to the shift is mostly left to guesswork.

This paper uses the Kalman Filter to automatically smooth the loss ratios based on the amount of credibility inherent in the data in a manner that is robust and that is consistent with the Cape Cod method. It is shown how this method can be thought of as a credibility weighting between the Cape Cod and Chain Ladder techniques, each of which are possible at the two extremes. It is then shown how external predictive information, such as the state of the economy or the insurance cycle, can be incorporated to help produce more accurate results. Simulation results are presented that illustrate the error reduction this method can provide to both historical years and to the latest year.

Keywords. Loss Reserving, Credibility, Smoothing, Kalman Filter, Trend

1. INTRODUCTION

The Cape Cod or Stanard-Buhlmann (Stanard 1985) method is a commonly used technique where the a priori loss ratio is calculated as the weighted average of the chain ladder ultimate loss ratios across all years with the "used" premium as the weights. It applies the same a priori loss ratio estimate (on a trended, current rates level) across all years, without consideration for any possible changes that may have occurred. A difficulty arises when the loss ratios show improvement or deterioration, which is a fairly common scenario. This can happen as a result of using rate changes or trends that are not completely accurate, changes in policy wording, or temporary shifts in the exposure to loss caused by economic or other factors. When this occurs, the amount of credibility that should be given to the shift is mostly left to guesswork. Being too slow to give credit to improving experience can cause the company to miss out on profitable opportunities and also frustrate underwriting management, and detecting deterioration too late can cause declines in profitability and capital that could have been avoided. Being too slow to detect any type of change can also contribute to diminished confidence in the entire reserving process. On the other hand, reacting to noise too quickly will cause faulty decisions to be made with negative results as well.

This paper presents a method that automatically smooths loss ratios based on the amount of credibility inherent in the data and that is consistent with the Cape Cod approach. If no credibility is given to any changes, a single loss ratio will be indicated for all years, and the results will match the Cape Cod method. On the other extreme, if full credibility is given to the chain ladder indications from each year, the final results will match the Chain Ladder method. Anywhere in between can be thought of as a credibility weighting between these two methods. It is then shown how external predictive information, such as the state of the economy or the insurance cycle, can be incorporated to help produce more accurate results.

1.1 Research Context

Gluck (1997) improved upon the original Cape Cod technique by adding a decay factor which gives increased weight to the more local experience, effectively smoothing the data. But there is still little guidance as to how much credibility or smoothness should be used. (There are formulas in the appendix for approximating this, but they are difficult to follow and implement, and the iterative approach suggested to solve the equations is not guaranteed to converge to the optimal solution, and likely will not.)

The Kalman Filter is a very popular smoothing algorithm used in many econometric applications. De Jong and Zehnwirth (1983) were the first to introduce its use into reserving and used it to help smooth development patterns. Both Zehnwirth (1996) and Wuthrich and Merz (2008) use the Kalman Filter to smooth the actual reserving estimates, but their formulations are much more complicated than a simple Cape Cod approach and will not be discussed here. Evans and Schmid (2007) use the Kalman Filter to derive smooth trend estimates but their approach is not suitable, nor intended, to apply directly to loss ratio estimates. None of these approaches demonstrate a simple, easy to understand framework that is in line with traditional actuarial practice, as the Cape Cod method does. The Kalman Filter formulas can also seem non-intuitive and hard to understand, making implementation of such an algorithm in the reserving context challenging. Finally, and also very critical, the indicated smoothness derived from the Kalman Filter or similar methods can be very volatile and inaccurate, essentially precluding its use in practice. As mentioned in Schmid et al. (2013), even the time series used for NCCI ratemaking is too short to reliably estimate the variance of the year-to-year changes, which is essential to determining the credibility. Having a smaller amount of data than NCCI would compound this problem. If this issue is not properly handled, such as by using the strategies that will be discussed in this paper, the Kalman Filter results cannot be relied upon.

1.2 Objective

The goal of this paper is to present a simple, easy to understand, and yet powerful and robust framework of applying the Kalman Filter to smooth loss ratio estimates that is consistent with the

Cape Cod method. This smoothing algorithm is applied to the on-level, trended ultimate Chain Ladder loss ratios with weights equal to the premiums divided by the LDFs, or the "used premiums". The results of this algorithm are the a priori loss ratios to apply to each year via a Bornhuetter-Ferguson method. If the algorithm determines that no credibility or smoothness should be given, the result for each year will be the weighted average across all years, and the method will be equivalent to the Cape Cod. On the other extreme, if full credibility or maximum smoothness is indicated, the a priori loss ratios for each year will match those of the Chain Ladder method, and so the final results will be identical to the Chain Ladder as well. Anywhere in the middle, the method can be thought of as a credibility weighting between these two methods.

This paper will also discuss the intuition behind the Kalman Filter formulas relating them to basic credibility theory. Many of the approaches mentioned apply the Kalman Filter on the logarithm of loss ratios, making it inconsistent with the Cape Cod approach and hard to determine the relative weights by year and requiring a messy bias correction if not using Bayesian software for calculation. Taylor and McGuire (2003) show a solution to this problem via what they call an EDF Filter, but the math required to implement it is complex. This paper applies the Kalman Filter on the loss ratios themselves but modifies the algorithm in a similar but simpler fashion to be able to handle multiplicative innovations, that is, the changes from year to year, and non-normally distributed errors. Strategies are also shown to make it robust so that it can be used in practice even with sparse, volatile data, and this is illustrated via simulation testing.

1.3 Outline

Section 2 discusses the intuition behind the Kalman Filter and shows how to apply it to model loss ratios, and section 3 shows how to make the algorithm more robust. Incorporating external predictive information is discussed in section 4, and examining multiple lines simultaneously is discussed in section 5. Finally, section 6 shows the results of running simulations using the methods discussed.

2. THE KALMAN FILTER

The method presented in this paper uses the Kalman Filter to determine the amount of smoothness or credibility that should be given to each year. The Kalman Filter was originally developed in 1960 for use in signal processing (Kalman 1960) but has become very common for solving time series econometric models. It is able to handle more complex types of models than are illustrated in this paper. For ease of understanding and implementation, a simplified version that contains only the needed components is discussed instead.

2.1 Intuition Behind the Kalman Filter

To understand how this algorithm works, assume that rate and trend are both flat and that we are attempting to predict the expected loss ratio for year 2 where we know (for certain) that the loss ratio for year 1 was 70%. Before observing any experience from the second year, our prediction would be 70%, the same as year 1. Assume that now we observe a (projected) loss ratio of 80% in the second year, which is still incomplete, and we want to estimate the expected loss ratio to be used in a Bornhuetter-Ferguson method to estimate the IBNR for the remainder of the year. If there was no loss volatility, we would assume that the 80% loss ratio will continue for the remainder of the year and this would be our estimate. On the other hand, if the loss volatility was extremely high such that the 80% prediction for this year had a large degree of uncertainty, we would give it almost no credibility, and our estimate would be the year 1 estimate, which is 70%. More practically, our estimate should fall somewhere in between these two extremes and take into account both the volatility of the losses and the volatility of the year-to-year changes. If these two variances were equal, we would select the midpoint, 75%. More generally, the optimal credibility to give to the second year's experience equals the variance of the year-to-year changes divided by the sum of the two variances, since this would produce the result with the lowest variance. Venter (2003) derives this result and shows that it is the basis for Buhlmann credibility. The variance of this estimate cannot be greater than each of the individual predictors; otherwise, we would just select one of them instead. The inverse of the variance equals the sum of the inverses of each of the variances. (Bolstad 2007)

If we now want to estimate the expected loss ratio for year 3, similar logic would apply, except that now the variance of the year 2 estimate needs to be taken into account as well. The total variance of using the year 2 estimate for year 3 would equal the variance of this estimate plus the variance of the year-to-year changes. This variance would then be compared to the loss volatility to calculate the optimal credibility to give to the third year's experience in the year 3 estimate. Once we have observed and predicted the loss ratio for the third year, this estimate can now be used to improve the prediction for the second year. To determine the amount of credibility to give to the year 3 estimate for the year 2 result, a similar formula is used except that the variance of this predictor is compared against the variance of the year-to-year changes instead of the variance of the losses.

This is essentially what the Kalman Filter does (the part that we are using, at least); the actual formulas are shown in the next section.

2.2 Kalman Filter Formulas

Similar logic is used to run the Kalman Filter. A first iteration is performed looking at the years (or quarters, etc.) going forwards. Then, once an initial estimate has been determined for each year, another iteration is performed, this time, starting at the end and traveling backwards by year. This is done to back-smooth the results and modify the earlier estimates taking into account what is known about the later years, since the first iteration only considers the reverse.

As alluded to in the previous section, two values are used in the first iteration for each prediction estimate and variance. The first represents the prediction for a particular year before observing the experience of that year, and the second represents a revised prediction that also takes into account the experience of that year.

There are three unknown parameters that are needed to run this algorithm: the starting loss ratio for the first year, the volatility of the experience, and the volatility of the year-to-year changes. Maximum likelihood is used to determine these values. Note that the likelihood is calculated using the initial loss ratio estimates, that is, the estimates before considering the experience for each year. This is done because otherwise, if the estimates after considering each year's experience were used, the algorithm would seek to minimize the differences between these actual and fitted loss ratios, which would result in indications that were completely smoothed to all of the noise in the experience. Then, after this forward iteration has been performed and after the values of all of the unknown parameters have been determined, another back-smoothing iteration is performed to calculate the final results.

The amount of credibility each new year is given in the rolling forward predictor is known as the Kalman gain and is equivalent to the credibility discussed in the previous section. This is shown as K in the formulas below. The formulas below show the predictor of year t before considering that year's experience as $X_{t|t-1}$, the predictor after considering the year's experience as $X_{t|t}$, and the final back-smoothed predictor as $X_{t|T}$. Similar notation is used for the variance. Note that these are <u>not</u> the final formulas, as some changes are needed to make the algorithm more suitable for loss ratios, which are shown later in section 2.3. Explanations are given by the formulas to relate it to the concepts discussed in the previous section. For the notation, Y are the observed loss ratios, X are the predicted loss ratios, P are the variances of the predictors, n is the forecast error, R is the loss volatility, Q is the volatility of the year-to-year changes, K is the Kalman gain, f is the total variance of the predictor including the volatility of the losses, and *loglik* is the log-likelihood. *Norm(a, b, c)* is used here to represent the log-likelihood of the normal distribution at *a,* with mean of *b,* and variance of *c.* (Kim and Nelson 1999)

The best estimate for the next year before observing the experience is the previous year's prediction. The variance of this prediction is the same as the previous year's variance plus the volatility of the year-to-year changes.

$$X_{t|t-1} = X_{t-1|t-1} \tag{2.1}$$

$$P_{t|t-1} = P_{t-1|t-1} + Q \tag{2.2}$$

The total error for the amount a prediction can differ from actual equals the prediction error (from the "true" value) plus the loss volatility.

$$f_t = P_{t|t-1} + \mathrm{R} \tag{2.3}$$

To determine the amount of credibility to give to a year's experience, the variance of the rolling forward prediction is compared against the loss volatility. This is shown as *K,* and is called the Kalman gain.

$$K_t = P_{t|t-1} \,/\, f_{t,} = P_{t|t-1} \,/\, (\, P_{t|t-1} + \mathrm{R}\,) \tag{2.4}$$

$$n_t = Y_t - X_{t|t-1} \tag{2.5}$$

$$X_{t|t} = X_{t|t-1} + K_t n_t = (1 - K_{t)})\, X_{t|t-1} + K_t\, Y_t \tag{2.6}$$

The variance of this rolling forward predictor decreases after observing and incorporating the experience, based on the formula mentioned that the inverse of the variance equals the sum of the inverses of the two variances. Simple algebra can show that this is equivalent to the below.

$$P_{t|t} = P_{t|t-1}\,(1 - K_t) = P_{t|t-1}\,\mathrm{R} \,/\, (\, P_{t|t-1} + \mathrm{R}\,) \tag{2.7}$$

$$1 \,/\, P_{t|t} = 1 \,/\, \mathrm{R} + 1 \,/\, P_{t|t-1} = (\, P_{t|t-1} + \mathrm{R}\,) \,/\, P_{t|t-1}\,\mathrm{R}\,;\; P_{t|t} = P_{t|t-1}\,\mathrm{R} \,/\, (P_{t|t-1} + \mathrm{R}\,)$$

The likelihood is calculated on the prediction error before observing that year's experience using the variance calculated for the rolling forward predictor.

$$loglik_t = Norm(\, n_t\,,\, 0,\, f_t) \tag{2.8}$$

After the initial prediction of the last year has been calculated, the results are back-smoothed. This matches the result mentioned in the previous section.

$$X_{t|T} = X_{t|t} + (\, P_{t|t} \,/\, P_{t+1|t}\,)(\, X_{t+1|T} - X_{t|t}\,) = Z\, X_{t+1|T} + (1 - Z)X_{t|t}\,,\; where\; Z = P_{t|t} \,/ (P_{t|t} + \mathrm{R}\,) \tag{2.9}$$

Even though a few modifications will be made to these formulas to apply more to loss ratios, an illustration is shown below using the numbers from the previous section. The *R* (loss volatility) and *Q* (volatility of year-to-year changes) parameters, which are determined via maximum likelihood, are assumed to be 1 and 0.5, respectively.

$X_{1|0} = 70\%$

$X_{2|1} = X_{1|0} = 70\%$

$P_{1|0} = 0$

$P_{2|1} = P_{1|0} + Q = 0 + 0.5 = 0.5$

$f_2 = P_{2|1} + \mathrm{R} = 0.5 + 1 = 1.5$

$K_2 = P_{2|1} \,/\, f_2 = 0.5 \,/\, 1.5 = 0.333,$

$n_2 = Y_2 - X_{2|1} = 80\% - 70\% = 10\%$

$X_{2|2} = X_{2|1} + K_2 n_2 = 70\% + 0.333 \times 10\% = 73.33\%$

$P_{2|2} = P_{2|1}\,(1 - K_2) = 0.5 \times (1 - 0.333) = 0.333$

2.3 Modifications for Loss Ratios

As mentioned, this smoothing algorithm will be applied to determine the a priori loss ratios for use in a Bornhuetter-Ferguson method. The inputs are the chain ladder loss ratios, since these are the loss ratios that have been observed for incomplete years at the current point in time. The "used premiums" are used as the weights, since this represents the volume for the losses observed thus far. If no smoothness is indicated, the a priori loss ratios will match that of the Cape Cod technique. If, on the other hand, maximum smoothness is given, they will match the chain ladder estimates, and using these in a Bornhuetter-Ferguson method will yield identical results as this method. Anywhere in between can be thought of as a credibility weighting between these two methods as the IBNR predicted for the remainder of each year will only consider each year's experience to the extent that it is credible.

To apply this algorithm on loss ratio data, a couple of modifications are necessary. The first is to deal with years that have different premium volumes, and thus different expected loss volatility, since the original formulas assume that this is constant per year. To allow for different variances, a variance factor can be used as one of the parameters instead of the actual variance. Assuming that the variance of each year is inversely proportional to the premium volume, which is a good assumption if all policies are homogenous in terms of severity, the variance for each year is equal to this variance factor divided by the premium. For incomplete years, the "used" premium is used instead, as discussed.

Ideally, the factor applied to the premiums of incomplete years should reflect the additional variance of these years, which includes both the decreased volume as well as any uncertainty in the loss development patterns. Performing some algebra, it can be seen that the factor relating to the decreased volume is actually the claim count development factor and not the loss development factor, as used in the Cape Cod method. (The derivation is shown in Appendix A.) However, using the claim count development factor would be ignoring any uncertainty in the loss development pattern, so using the loss development factor, which is usually slightly higher than the claim count development factor, is recommended to account for this additional variance as an approximation. This will also make it consistent with the Cape Cod method, which is a desirable property. Alternatively, it is also possible to use the claim count development factors and estimate the uncertainty in the development patterns more exactly if desired.

Another modification is needed to handle non-normally distributed errors. Instead of calculating the likelihood using a normal distribution as the original algorithm does, a gamma or negative binomial distribution can be used instead. (A gamma distribution is appropriate for modeling on severity data and a negative binomial for modeling on frequency data.) The mean and variance resulting from the Kalman Filter algorithm can be used to solve for the two parameters of the appropriate distribution. If using a gamma distribution, for example, the variances calculated in the Kalman Filter algorithm will really be the variances divided by the means squared, and so it is assumed that the variance is proportional to the square of the mean. A negative binomial is not appropriate for modeling loss ratios, since this data often has a variance-to-mean ratio less than one, which this distribution does not allow. A Poisson distribution cannot be used since it does not have an additional parameter for the variance. An overdispersed Poisson has another parameter for the variance but is more difficult to implement. Similarly, implementing a Tweedie distribution, which is often used to model on loss ratios, is difficult as well.

But both a Poisson and Tweedie can be approximated fairly well. Calculating the log-likelihood as the average of the log-likelihoods of the normal and gamma distributions produces results that are very close to using a Generalized Linear Model with a Poisson distribution. Taking a weighted average between these two log-likelihoods with the weight to the gamma distribution equal to half the desired power of a Tweedie distribution also comes very close to using a Generalized Linear Model with a Tweedie distribution. So, for example, applying a weight of 1.67 / 2 = 0.835 to the gamma log-likelihood and a weight of 0.165 to the normal log-likelihood comes very close to using a Tweedie with a power of 1.67. When this is done, another parameter is needed as the constant factor to convert the variance to the coefficient of variation, which is needed to solve for the gamma parameters. (If only a gamma is used, this parameter is not needed, since the variance variable in the Kalman Filter formulas will already represent the variance divided by the mean squared.) Conducting a simulation[1] and comparing the results to a similar GLM when no smoothness resulted (which was about half the time) produced results that were very close. The gamma results matched the GLM results almost exactly. The Poisson and Tweedie results were within 0.05 percentage points of the GLM indications 89% and 98% of the time, respectively, and were within 0.1 percentage points 100% of the time. The results show that this method produces a fairly decent approximation.

If a gamma distribution is used, the yearly innovations are assumed to be multiplicative since it assumes that the variance is proportional to the square of the response, which works well with handling multiplicative relationships, similar to its use in Generalized Linear Models. If a normal distribution is used, the yearly innovations are assumed to be additive. If the approximation of the Poisson distribution is used, as described, the yearly innovations are assumed to be in between additive and multiplicative. It is difficult to say what the appropriate form these innovations should take[2], but if it is desired to have multiplicative innovations, the formulas can be modified to use the product of Q and the loss ratio for a Poisson distribution. For a Tweedie distribution with power p, the product of Q and the loss ratio to the power of two minus p is used instead. This change will cause the variance of the innovations to be related to the square of the mean.

The final formulas that take these modifications into account are shown below. *epow* is the exponential power used (0 for normal, 1 for Poisson, between 1 and 2 for Tweedie, and 2 for gamma), *EP* is the used premium, *Gamma(x, alpha, beta)* is the gamma log-likelihood at x with parameters *alpha* and *beta,* and *NB(x, n, p)* is the negative binomial log-likelihood at x with parameters n and p. These formulas assume that the year-to-year changes are multiplicative, although this may or may not be the case.

$$X_{1|0} = \langle \textit{Set from a parameter} \rangle \tag{2.10}$$

$$P_{1|0} = 0 \tag{2.11}$$

$$X_{t|t-1} = X_{t-1|t-1} \tag{2.12}$$

$$P_{t|t-1} = P_{t-1|t-1} + Q\, X_{t|t-1}{}^{2\,-\,epow} \tag{2.13}$$

---

1 Frequency was simulated using a negative binomial with a mean of 50 and a variance-to-mean ratio of 2.5. Severity was simulated from a lognormal distribution with mu and sigma parameters of 10 and 2, respectively, a retention of $100 thousand and limit of two million. Trend per year was 5%, autocorrelation was 10%, and variance of the year-to-year changes was 0.0001. Premium was set so that the expected loss ratio for the first year would be 70%. 500 simulations were run.

2 Looking at industry data using a Box-Cox test (which is out of scope of this paper produced conflicting results with a very large confidence interval depending on the line and the time period looked at.

$$f_t = P_{t|t-1} + R \; / \; EP_t \tag{2.14}$$

$$K_t = P_{t|t-1} \; / \; f_t, \tag{2.15}$$

$$n_t = Y_t - X_{t|t-1} \tag{2.16}$$

$$X_{t|t} = X_{t|t-1} + K_t n_t \tag{2.17}$$

$$P_{t|t} = P_{t|t-1} \; (1 - K_t) \tag{2.18}$$

$$loglik\text{-}norm_t = Norm(\, n_t \, , 0, f_t \,) \tag{2.19}$$

$$alpha = X^2_{t|t-1} \; / \; (f_t \times <Parameter>) \tag{2.20}$$

$$beta = X_{t|t-1} \; / \; (f_t \times <Parameter>) \tag{2.21}$$

$$loglik\text{-}gamma_t = Gamma(\, Y_t \, , alpha, beta \,) \tag{2.22}$$

$$loglik_t = (epow/2) \; loglik\text{-}gamma_t + (1 - epow/2) \; loglik\text{-}norm_t \tag{2.23}$$

Back-Smoothing:

$$X_{t|T} = X_{t|t} + (\, P_{t|t} \; / \; P_{t+1|t} \,) \, (\, X_{t+1|T} - X_{t|t} \,) \tag{2.24}$$

For a negative binomial:

$$n = X_{t|t-1} \; / \; (\, f_t \times <Parameter> - 1 \,) \tag{2.25}$$

$$p = 1 \; / \; (\, f_t \times <Parameter> \,) \tag{2.26}$$

$$loglik_t = NB(\, Y_t \, , n, p \,) \tag{2.27}$$

Since a gamma distribution is used which does not have any likelihood at zero, any zero loss ratios should be set to a very small number slightly above zero.

As general advice, when solving for the two variance parameters, it is recommended to use one parameter for the total variance and another parameter for the percentage of the total variance that is attributable to the year-to-year changes (a logit function can be used to ensure that this value is between zero and one). The noise variance parameter can then be set to the total variance parameter multiplied by one minus this percentage, and then multiplied by the average premium volume, or something similar, to make this parameter relative to the premium volume. If this strategy is not used, care should be taken as solving for these variance parameters directly can sometimes cause difficulty with optimization routines.

With volatile data, it is often helpful to cap losses at an appropriate point to make the data more stable. If there have been changes in retentions or policy limits, the premium should be adjusted appropriately as well. It is also possible to use this same algorithm on claim frequency and/or severity separately. For frequency, the premium should be adjusted if there have been changes in the retentions or policy limits by dividing out the average expected (conditional) severity. When looking at frequency, it is possible to include all claims, or to only include significant claims greater than a certain threshold.

3. ROBUSTIFYING THE METHOD

As mentioned, the indicated smoothness of the Kalman Filter can be unreliable with relatively few data points. It also struggles with data as volatile as loss ratios. Without addressing these issues, the algorithm cannot be used in practice.

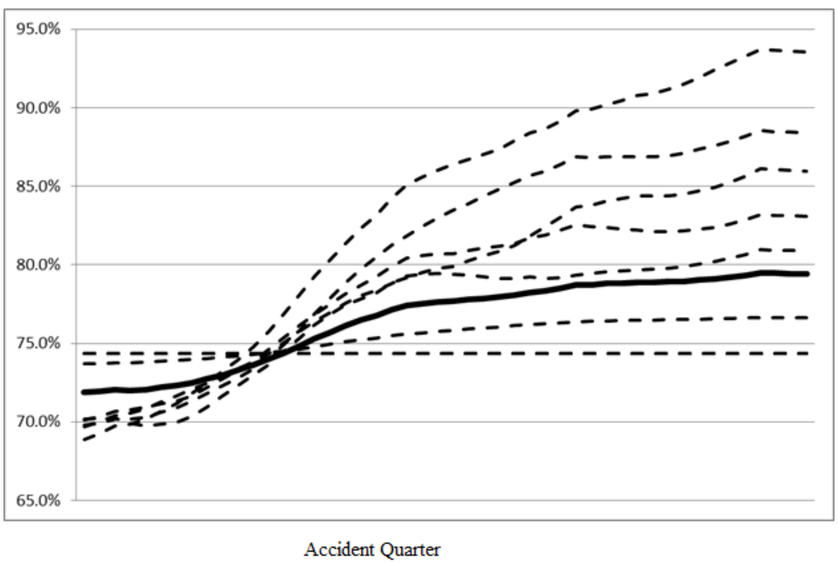
The number of available data points depends on how long the company's history is with the segment being analyzed. It also depends on how consistent processes and practices have been since this determines the relevant data that can be used. Even though the purpose of this algorithm is to address gradual shifts, it may still be beneficial to discard older information that is deemed less relevant and that does not add any value for prediction of the more recent data. If less than twenty years or so of data are available for analysis, it is strongly recommended to use quarterly data instead, which will increase the number of data points four-fold. Even with twenty years of data or more, using quarterly data can still greatly increase the accuracy of the method since it enables better estimation of the variance. If different loss ratios are expected in each quarter due to the effects of seasonality, this can be addressed similarly to the incorporation of external data, as described in section 4.1. (Credibility can be incorporated as well, as described in section 4.2.)

Another technique to make the algorithm more robust is to use bootstrapped aggregation, or "bagging", where multiple iterations of the algorithm are performed, each time on only a fraction of the years or quarters. The final indicated a priori loss ratios are then calculated as the average across all iterations. Each iteration will receive a varying amount of smoothness based upon which years/quarters are included, and averaging across all of these produces a much more stable and reliable result. (Just to be clear, the average of each indicated loss ratio should be used, and not the average of the smoothness parameters, since the former produces much more reliable results than the latter.) Using fifty iterations with selecting two thirds of the data each time seems to perform quite well both in simulation tests and on actual data. (When implementing, it is important to either explicitly set the random number generator seed or to ensure that the same bootstrapped simulations are used each time to avoid having the indications change slightly when rerun.)

To implement, if a data point is skipped, the Kalman gain should be set to zero to give it no credibility. This will result in the predictor variance being increased by the year-to-year variance reflecting the fact that the prediction interval is being extended by skipping this point. So even though the Kalman gain is artificially decreased at one point, this will cause it to be increased for the following point. The likelihood of this point should still be included in the overall likelihood so that it affects the average, however, since the bootstrapping is only needed for the amount of smoothness, and bootstrapping on this will only decrease stability slightly.

An example where the Kalman Filter was run fifty times from simulated data is shown in Figure 1. The first ten individual runs are shown as well as the run that resulted in the most smoothness (dotted lines). The average is shown as the thick solid line. Note how volatile the amount of smoothness can be from single runs, ranging from far too much credibility given to none at all, which occurred in 17 out of the 50 runs. The average incorporates all of these indications and results in a much more stable and reasonable result.

Figure 1



4. ADDING PREDICTIVE VARIABLES

4.1 Formulas

Predictive variables, such as the state of the economy or of the market cycle, can be incorporated to improve the accuracy of the predictions. The following formulas can be used, where V is the total impact of the predictive variables at each period, v are the predictive variables, and *coef* are fitted coefficients for each of these variables[3]:

$$b_{1|0} = \text{<Set From a Parameter>} \tag{4.1}$$

$$P_{1|0} = 0 \tag{4.2}$$

$$V_t = exp(\textstyle\sum_i \ coef_i \times v_i \ ) \tag{4.3}$$

$$b_{t|t-1} = b_{t-1|t-1} \tag{4.4}$$

$$X_{t|t-1} = b_{t|t-1} \ V_t \tag{4.5}$$

$$P_{t|t-1} = P_{t-1|t-1} + Q \ X_{t|t-1}{}^{2 \ - \ epow} \tag{4.6}$$

$$f_t = P_{t|t-1} \ V_t^2 + \mathrm{R} \ / \ EP_t \tag{4.7}$$

$$K_t = P_{t|t-1} \ V_t \ / \ f_t, \tag{4.8}$$

$$n_t = Y_t \ - \ X_{t|t-1} \tag{4.9}$$

$$b_{t|t} = b_{t|t-1} + K_t n_t \tag{4.10}$$

$$P_{t|t} = P_{t|t-1} \ (1 \ - \ K_t \ V_t) \tag{4.11}$$

$$loglik\text{-}norm_t = Norm(\ n_t \ , \ 0, f_t) \tag{4.12}$$

$$alpha = X^2{}_{t|t-1} \ / \ (f_t \times \text{<Parameter>}) \tag{4.13}$$

$$beta = X_{t|t-1} \ / \ (f_t \times \text{<Parameter>}) \tag{4.14}$$

$$loglik\text{-}gamma_t = Gamma(\ Y_t \ , \ alpha, \ beta \ ) \tag{4.15}$$

$$loglik_t = (epow/2) \ loglik\text{-}gamma_t + (1 \ - \ epow/2) \ loglik\text{-}norm_t \tag{4.16}$$

Back-Smoothing:

$$b_{t|T} = b_{t|t} + (\ P_{t|t} \ / \ P_{t+1|t} \ ) \ (\ b_{t+1|T} \ - \ b_{t|t} \ ) \tag{4.17}$$

$$X_{t|T} = b_{t|T} \ V_t \tag{4.18}$$

An exponential function was used to calculate the impact of the predictive variables, similar to a log-link GLM, but other alternatives are possible as well. *b* is an intermediate variable similar to an intercept. Using this method is similar to using a GLM where the intercept can vary over time.

The predictive variables here function similarly to a GLM, in that their effect is calculated cumulatively, as opposed to being incremented by an additional amount for each year. This means

[3] These formulas are obtained by replacing the *H* matrix from the original formulas with the result of the predictive variables.

that if, for example, the change in GDP is judged to affect loss ratios, then the actual GDP should be used as a variable, and not the change in the GDP. This way, the incremental effect to each year will be the change in this variable. Similarly, if the change in the GDP growth rate is desired instead, then the GDP change should be used as a variable.

Using this method, it also is possible to fit a constant trend to the data by including the year as a predictive variable. This example is used to help illustrate this method. Loss ratios with a constant frequency trend per year were simulated[4]. Three methods were compared: a (Tweedie) GLM, the Kalman Filter model with no trend and the Kalman Filter model with the year as a predictive variable to represent the trend (both using the approximation for the Tweedie distribution that was discussed).

It is interesting to see the results of the Kalman Filter without trend model. Sometimes this model can do a fairly decent job of following the trend in the data, although it often needs to adapt too much to the data in order to do so, and as a result, produces some overfitting as in Figure 2. In this example, the Kalman Filter with trend model indicated no smoothness and so the result is very close to the Tweedie GLM. The dotted, "actual" line here is the "true" value for each year before volatility is added in the simulation, and the solid, "observed" line is the result with added volatility.

[4] Frequency was simulated using a negative binomial with a mean of 25 and a variance-to-mean ratio of 3. Severity was simulated from a lognormal distribution with mu and sigma parameters of 10 and 2, respectively, a retention of $100 thousand and limit of two million. Trend per year was 3%, autocorrelation was 40%, and variance of the year-to-year changes was 0.0025. Premium was set so that the expected loss ratio for the first year would be 70%. For the bagging, 25 iterations were used using ⅔ of the data on each iteration. 200 simulations were run. The models were fit using the approximation for the Tweedie family mentioned earlier.

Figure 2

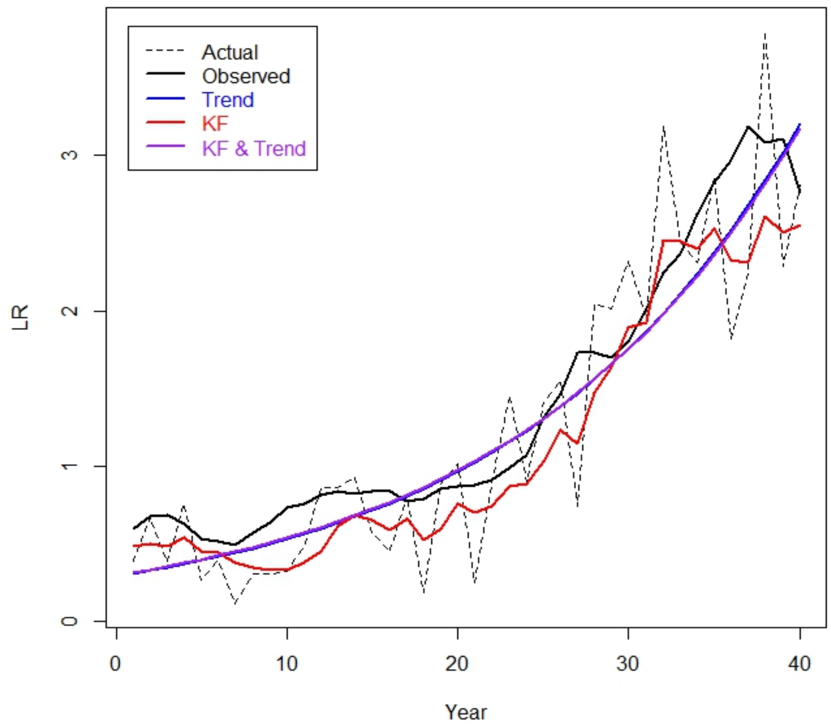
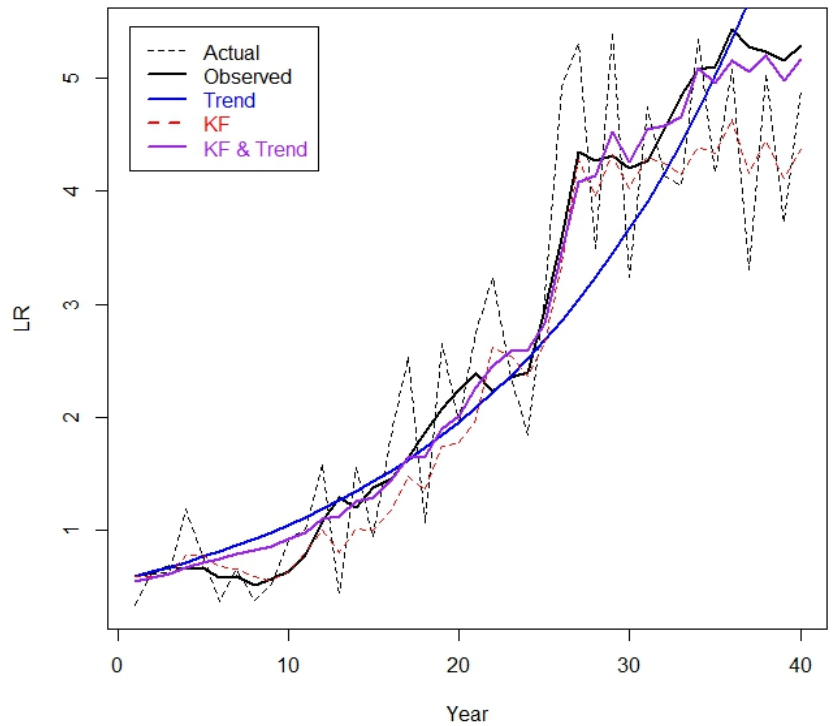


Figure 3 is another example where the Kalman Filter with trend model differs from the GLM and also smooths to the data. In this example, this model does a very good job of adapting to the changing loss ratios per year as well as to the trend in the data, much better than both the simpler trend model and Kalman Filter model (although, of course, this will not always be the case). (The Kalman Filter line is shown with a thinner line in the below graph, as it is not relevant in this example.)

Figure 3



The results of running many simulations are shown in Figure 4. As expected, the Kalman Filter with trend model outperforms both the Kalman Filter without trend and the GLM models.

Figure 4

Method	RMSE[5]
Kalman Filter Without Trend	0.201
GLM	0.167
Kalman Filter With Trend	0.157

[5] RMSE stands for Root mean squared error. It is the square root of the average error squared.

4.2 Further Robustifying This Method

Accidentally including a variable that has no true predictive value can degrade performance. Significance tests can also be unreliable. One way to address this issue and also to increase the accuracy even for truly predictive variables is to use penalized regression. This applies a penalty to large coefficient values, which helps to stabilize the model. With categorical/dummy variables, the effect is similar to credibility weighting, but this method can be used for all types of variables. Ridge regression, a type of penalized regression, will be illustrated. To implement, the logarithm of a normal probability density function with a mean of 0 is evaluated at each coefficient value (just for the predictive variables, that is), and this sum is added to the total log-likelihood. The variance of this normal distribution can be estimated using cross validation.

One simple way to perform cross validation is to test various candidate variance values and fit the model on only a fraction of the data. The remaining data is then used to calculate the mean squared error divided by the mean to the appropriate power (one for Poisson, two for gamma, etc.), multiplied by the used premium. This process should be repeated several times to gather a more reliable estimate. It also helps reduce the number of iterations needed if the same samplings are used for each value being testing, although this is not required. A graph of the average mean square errors can show whether enough iterations have been performed.

The same variance is usually used for all coefficients. Non-dummy variables should be standardized to all be on the same scale so that their variances are comparable; this can be done by subtracting out the mean and dividing by the standard deviation, or if dummy variables are being used as well, by dividing by two times the standard deviation (Gelman 2008). Using this method lessens the negative effect of noise variables and also improves the performance of predictive variables. There are other methods of performing cross validation that will not be discussed here.

5. MULTIPLE LINES

Multiple lines can be evaluated together using the same variance parameters, R and Q, but allowing different initial loss ratio parameters for each line. This will leverage the volatility estimation across all of the lines together.

Going one step further, it is possible to do the same, but have the initial loss ratios related to each other via credibility weighing. This can be done using Bayesian credibility, and this method can be implemented simply, without the use of specialized Bayesian software, as will be explained. If a normal distribution is used as the prior distribution for the initial loss ratios, this is a conjugate prior since a normal distribution is also being used for the loss ratios, and so, the posterior distribution will be normally distributed. This means that maximum likelihood estimation, which returns the mode, can be used to estimate the mean, since the mean is identical to the mode for a normal distribution. Performing credibility in this fashion will also match the Buhlmann-Straub credibility results (Herzog 1989). To implement, another parameter should be added for the complement of credibility. Then, the log-likelihood of a normal probability density function evaluated at each initial loss ratio with a mean of the credibility complement should be calculated for each line. Adding the sum of these log-likelihoods to the total log-likelihood will cause the loss ratios to shift towards the overall mean and credibility weighting will be performed. The variance of this normal prior distribution is equivalent to the between variance used in the Buhlmann-Straub method. One way to estimate it is to use the Buhlmann-Straub formulas (as described in Korn 2015 to apply to loss ratios, for example).

Using this approach to calculating the between variances, however, does not consider the loss ratio changes by year as calculated by the Kalman Filter and so is slightly inconsistent. As an alternative, cross validation can be used instead, similar to ridge regression, which was described earlier. Different between variances can be tested where the loss ratios are fit using only a fraction of the data and the remainder of the data is used to calculate the mean square error divided by the mean to the appropriate power, multiplied by the used premium. Using this will be consistent with the loss ratio changes by year.

There is still an issue, however, since credibility weighting the initial loss ratios towards the mean but then allowing the remaining ones to vary freely sometimes produces results that deviate away from the mean with time, even if this is not the case, especially if the between variance chosen is relatively small. Bayesian credibility was used to credibility weight the initial loss ratios, which has the formula:

$$f(Posterior \mid Data, Parameters) = f(Likelihood \mid Data, Parameters) \times f(Prior \mid Parameters).$$

Credibility weighting is performed since the prior component, *f(Prior | Parameters)*, applies a penalty to the parameters as they deviate away from the mean. This prior needs to be a function of the model parameters.

However, it is also possible to reparameterize the model so that instead of using the initial loss ratios as the parameters, the ending loss ratios are used instead. Note that it is possible to solve for the ending loss ratios given all of the Kalman Filter parameters including the initial loss ratios. Because of this, it is also possible to invert the equations and to solve for the initial loss ratios given the ending loss ratios. So, the ending loss ratios can be used as the parameters of the model, the initial loss ratios can be solved for, and then the Kalman Filter can be run as normal. To make the process simpler, instead of actually performing all of these calculations, we can run the Kalman Filter as normal using the initial loss ratios as the parameters, but still "pretend" that the ending loss ratios are the parameters and calculate the prior distribution credibility penalty using the ending loss ratios, since the result would be exactly the same. So, in summary, nothing needs to be changed, and the ending loss ratios can be used for credibility weighting.

This method produces better behaving models that do not artificially deviate either towards or away from the mean. Because the Kalman Filter iterates forwards through all of the loss ratios, and then conducts another iteration backwards to smooth the results, the ending loss ratio can almost be thought of as the midpoint of the iteration. Therefore, it is recommended to use the ending loss ratios for calculating the log-likelihoods of the normal prior distribution.

6. SIMULATION RESULTS

A simulation was run[6] to help illustrate the benefits this method can provide, although, of course, the exact benefit will vary from case to case. In this scenario, two random variables were combined to simulate the frequency per year, and it was assumed that one of these was known. This was done to simulate a scenario where a predictive variable is known that affects the frequency per year, such as the state of the economy, but that not everything about how the frequency changes is known.

The summary of the results are shown in Figure 5.

[6] Frequency was simulated using a negative binomial with a mean of 50 for complete years and a variance-to-mean ratio of 2.5. Severity was simulated from a lognormal distribution with mu and sigma parameters of 10 and 2, respectively, a retention of $100 thousand and limit of two million. Autocorrelation was 30% for each of the frequency variables and for the severity variable, variance of the year-to-year changes was 0.005 for each of the frequency variables and 0.00025 for the severity variable. Development factors were used that affected the frequency that decreased by 0.05 starting at the 22nd period. Premium was set so that the expected loss ratio for the first year would be 70%. For the methods that used bagging, 25 iterations were used using ⅔ of the data on each iteration. 500 simulations were run. The models were fit using the approximation for the Tweedie family mentioned earlier.

Figure 5

Method	RMSE All Years	RMSE Latest Year	RMSE All Years - Compared to Cape Cod	RMSE Latest Year - Compared to Cape Cod
Cape Cod	1.211	0.312	0%	0%
Kalman Filter	0.538	0.132	-55.5%	-57.6%
Kalman Filter with Bagging	0.508	0.126	-58.1%	-59.5%
Kalman Filter with Predictive Variable	0.485	0.120	-59.9%	-61.4%
Kalman Filter with Predictive Variable and Bagging	0.462	0.114	-61.9%	-63.3%
Kalman Filter with Predictive Variable, Penalized Regression and Bagging[7]	0.453	0.105	-62.6%	-66.3%
Tweedie GLM with Predictive Variable (Weighted by Used Premium per Year)	0.704	0.165	-41.8%	-47.2%

The main conclusion is the amount of benefit this method is capable of providing over the Cape Cod, which does not adapt to changing conditions and cannot include predictive variables. Each of these individually is also able to provide significant benefit.

7. CONCLUSIONS

The goal of this paper was to present a relatively simple method that can be implemented in spreadsheets to extend the Cape Cod and is capable of accounting for changes indicated in the data and from external predictive variables. Estimating expected loss ratios per year with volatile data can often be a confusing and difficult task, subject to a large degree of judgement. It is our hope to improve this process by adding some guidance from modern statistical techniques without losing the simple and intuitive nature of the Cape Cod method.

[7] Only 100 iterations were performed for this method because of its longer running time. Also, only 10 iterations of bootstrapping were performed. Using a higher number is expected to further improve the performance of this method.

Acknowledgment

The author wishes to acknowledge the contributions of the Reserving committee who helped make clarifications and corrections to this paper, in particular, Kelly Moore, Steve Lafser, Ryan Royce, and Denise Ambrogio. All remaining errors are the authors.

APPENDIX A

The appropriate development factor to relate to the increased volatility of incomplete experience can be derived using the formula for the variance of aggregate losses, where A are the ultimate aggregate losses, F is the ultimate frequency, S is the ultimate severity, VTM_F is the variance-to-mean ratio for the frequency, CV_S is the coefficient of variation for the severity, RPT are the reported losses, and ULT are the ultimate losses:

$$V(A) = V(F)\,E(S^2) + E(F)\,V(S) = VTM_F\,F\,S^2 + F\,CV_s\,S^2 = F\,S^2\,(\,VTM_F + CV_S^2\,)$$

The variance of the reported losses is equal to the below, since the observed frequency is F / $CCDF$ (where $CCDF$ is the claim count development factor), and the observed severity is S / SDF (where SDF is the severity development factor, which is equal to the LDF divided by the $CCDF$):

$$V(RPT) = \frac{F}{CCDF} \times \frac{S^2}{SDF^2} \times (\,VTM_F + {CV_S}^2\,)$$

The variance of ultimate losses is then equal to:

$$\begin{aligned} V(ULT) &= V(RPT) \times LDF^2 = V(RPT) \times CCDF^2 \times SDF^2 \\ &= \frac{F}{CCDF} \times \frac{S^2}{SDF^2} \times (\,VTM_F + {CV_S}^2\,) \times CCDF^2 \times SDF^2 \\ &= F \times S^2 \times (VTM_F + {CV_S}^2\,) \times CCDF \\ &= ULT \times S \times (VTM_F + {CV_S}^2\,) \times CCDF \end{aligned}$$

Note how all SDF terms cancel out and the only development term remaining is the claim count development factor.

8. REFERENCES

[1] Bolstad, W. 2007. Introduction to Bayesian Statistics (Second Edition). Wiley. 2007, p.106-108, 207-209

[2] De Jong, P. and Zehnwirth, B. 1983. Claims reserving, state-space models and the Kalman filter. Journal of the Institute of Actuaries, 110, pp 157-181.

[3] Evans, Jonathan P., and Frank Schmid. 2007. Forecasting Workers Compensation Severities and Frequency Using the Kalman Filter, Casualty Actuarial Society Forum, 2007: Winter, pp 43–66

[4] Gelman, A. 2008. Scaling regression inputs by dividing by two standard deviations. Stat. Med. 27: 2865–2873.

[5] Gluck, S. M. 1997. Balancing Development and Trend in Loss Reserve Analysis. PCAS 84, pp 482–532.

[6] Herzog, T. 1989. Credibility: The Bayesian Model Versus Buhlmann Model. Transactions of the Society of Actuaries, 1989, Volume 41.

[7] Kalman RE. A New Approach to Linear Filtering and Prediction Problems. ASME. J. Basic Eng. 1960;82 pp. 35-45.

[8] Kim C. and Nelson C. 1999. State-Space Models With Regime Switching. MIT Press, 1999. pp. 29-30, 36-37

[9] Korn, U. 2015. Credibility for Pricing Loss Ratios and Loss Costs, Casualty Actuarial Society E-Forum, 2015: Fall, pp 327–347

[10] Schmid, F., Laws, C., & Montero, M. 2013. Bayesian Trend Selection. In Casualty Actuarial Society E-Forum, Spring 2013.

[11] Stanard, James N. 1985 A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques. PCAS LXXII, 1985, pp.124–148.

[12] Taylor, G., & McGuire, G. 2007. Adaptive Reserving using Bayesian revision for the Exponential Dispersion Family. Centre for Actuarial Studies, Department of Economics, University of Melbourne.

[13] Venter, Gary G. 2003. Credibility Theory for Dummies. Casualty Actuarial Society Forum, 2003: Winter, pp 221-267

[14] Wüthrich, M.V. and Merz, M. 2008. Stochastic Claims Reserving Methods in Insurance. Wiley.

[15] Zehnwirth B. 1996. Kalman filters with applications to loss reserving. Research Paper Number 35, Centre for Actuarial Studies, The University of Melbourne.

Biography of the Author

Uri Korn is an AVP & Actuary at Axis Insurance serving as the Research and Development support for all commercial lines of insurance. Prior to that, he was a Supervising Actuary at AIG in the Casualty pricing department. His work and research experience includes practical applications of credibility, trend estimation, increased limit factors, non-aggregated loss development methods, and Bayesian models. Uri Korn is a Fellow of the Casualty Actuarial Society and a member of the American Academy of Actuaries.