The 2004 NCCI Excess Loss Factors

Dan Corro and Greg Engl*

October 17, 2005

1 Introduction

An in-depth review of the NCCI excess loss factors (ELFs) was recently completed and changes were implemented in the 2004 filing season. The most significant change was to incorporate the latest data, but the methodology was thoroughly reviewed and a number of methodological changes were made as well. Among the methodological items considered were:

1. Individual Claim Development
   Our intent here was to follow the method in Gillam and Couret [5] and merely update the parameters. However our treatment of reopened claims is new as is the way we implement individual claim development. This is covered in detail in section 2.

2. Organization of Data
   The prior procedure fit countrywide loss distributions by injury type and then adjusted the means of those distributions to be appropriate for each individual state. We extend this idea to match the first two moments. The prior procedure implicitly gives each state's data a weight proportional to the number of claims in the given state, and thus even the largest states do not get very much weight in the countrywide distributions. We give much more weight to individual states' own data and thus fit state specific loss distributions. For credibility reasons the

*We gratefully acknowledge the creative contributions of the many people involved in this project, including, but not limited to, NCCI staff and NCCI's Retrospective Rating Working Group.

prior loss distributions combined permanent total injuries with major permanent partial injuries, and minor permanent partial injuries with temporary total injuries. We fit fatal, permanent total (PT), permanent partial (PP), temporary total (TT), and medical only distributions separately. In order to do this we use data at third, fourth, and fifth report for fatal and permanent total injuries. Mahler [10] also uses data at third, fourth, and fifth report. For permanent partial, temporary total, and medical only injuries, where there is adequate data, we only use data at fifth report. This is covered in section 3.

3. Fitting Method
   We follow Mahler [10] and rely on the empirical data for the small claims and only fit a distribution to the tail. We fit a mixed exponential distribution to the tail. Keatinge [8] discusses the mixed exponential distribution. Rather than fitting with the traditional maximum likelihood method we choose to fit the excess ratio function of the mixed exponential to the empirical excess ratio function using a least squares approach. This yields an extremely good fit to the data. It should be noted that we do not fit the raw data, but rather the data adjusted to reflect individual claim development as described in section 2. This results in a data set that has already been smoothed significantly and so we were not concerned that the mixed exponential tail might drop off too rapidly. Mahler [10] noted that the excess ratios are not very sensitive to the splice point, i.e. the point where the empirical data ends and the tail fit begins. Thus we preferred to not attach too far out into the tail so that we could have some confidence in the tail probability, i.e. the probability of a claim being greater than the splice point. We generally chose splice points that resulted in a tail probability between 5% and 15%. This is covered in section 4.

4. Treatment of Occurrences
   We put a firmer foundation under the modeling of occurrences by basing it on a collective risk model. In the end we find that the difference between per claim excess ratios and per occurrence excess ratios is almost negligible. This is quite a sharp contrast with the past. Once, per occurrence excess ratios were assumed to be 10% higher than per claim excess ratios. This was later refined by Gillam [4] to the assumption that the cost of the average occurrence was 10% higher than the aver-

age claim. Gillam and Couret [5] then refined this even further to apply by injury type: 3.9% for fatal injuries, 6.6% for permanent total and major permanent partial injuries, and 0% for minor permanent partial and temporary total injuries. Our analysis shows that per occurrence excess ratios are less than .2% more than per claim excess ratios. This is covered in section 5.

In section 6 we discuss updating the loss distributions. The current procedure is to update the loss distributions annually by a scale transformation and to refit the loss distributions based on new data fairly infrequently. The scale transformation assumption is extremely convenient and is discussed by Venter [12]. What is needed is a method to decide when a scale transformation is adequate and when the loss distributions need to be refit. We conclude by reviewing the methodology changes. While the focus of this paper is on methodology, we also take the opportunity to briefly discuss the impact of the changes.

2 Individual Claim Development

When evaluating aggregate loss development it is not necessary to account for the different patterns that individual claims may follow as they mature to closure. In aggregate it does not matter whether ten claims of \$100 each all increase by \$10 or whether just one claim increases by \$100 to produce an ultimate loss of \$1,100 and an aggregate loss development factor (LDF) of 1.1. But if you are interested in the excess of \$110 per claim, it makes all the difference. Gillam and Couret [5] address the need to replace a single aggregate LDF with a distribution of LDFs in order to account for different possibilities for the ultimate loss of any immature claim. They refer to this as dispersion, and the name has stuck. Here, the term dispersion refers to a way of modelling ultimate losses that replaces each open claim with a loss distribution whose loss amounts correspond to the possibilities expected for that individual claim at closure.

The loss distribution used to determine the ELF should reflect the loss at claim closure. The calculation is done by injury type and uses incurred losses. It must reflect maturity in the incurred loss beyond its reporting maturity fully to closure, including any change in claim status (open/closed) and change in the incurred loss amount. Moreover, it must accommodate the reality that not all claims mature in the same way. Age to age aggregate incurred LDFs are determined from 1^{st} to 5^{th} report by state, injury type, and separately for indemnity and medical losses. The source is Workers Compensation Statistical Plan Data (WCSP), as adjusted for use in class ratemaking. As WCSP reporting ceases at 5^{th} report, 5^{th} to ultimate incurred LDFs, again separately for indemnity and medical losses, are determined from financial call data, typically in concert with the overall rate-level indication.

Individual claim WCSP data by injury type and report is the data source for the claim severity distributions. PP, TT, and medical only claims are included at a 5^{th} report basis. The far less frequent but often much larger Fatal and PT claims are included at 3^{rd}, 4^{th} and 5^{th} report basis. The WCSP data elements captured include state, injury type, report, incurred indemnity loss, incurred medical loss, and claim status. This detailed WCSP loss data is captured into a model for the empirical undeveloped loss distribution. That model consists of a discrete probability space to capture the probability of occurrence of individual claims together with two random variables for the claims' undeveloped medical and indemnity losses as well as four characteristic variables for state, injury type, report, and claim status. Eventually, this is refined into a model for the ultimate loss severity distribution that consists of a probability space together with one random variable for the claims' ultimate loss as well as two characteristic variables for state and injury type.

Because dispersion is exclusively focussed on open claims, without some accommodation, claims reported closed but that later reopen would not be correctly incorporated in the dispersion model. Accordingly, it is advisable to account for reopened claims prior to dispersing losses. The loss amounts considered are the total of the medical and indemnity losses for each claim. The methodology adjusts those loss amounts and probabilities by claim status and injury type, so as to model the impact of reopening claims. The details for the specific calculations used can be found in Appendix A and Appendix C. It is based on the observation that the few closed claims that reopen after a 5^{th} report (0.2%) are not typical, but are on average larger (by a factor of 8) and have a smaller CV (by a factor of 0.4). Appendix A shows quite generally how to calculate the resulting means and variances when a subset of claims have their status changed from closed to open.

The probability, mean, and variance of the three subsets of the loss model:

1. claims reported closed at 5^{th} report
2. claims reported open at 5^{th} report

3. claims that reopen subsequent to a 5^{th} report

completely determine the probability, mean, and variance of the complementary subsets:

1. claims 'truly closed' at 5^{th} report (those reported closed that do not reopen)

2. the complement set of 'truly open' claims.

That is, there is only one possibility for the probability, mean, and variance of the truly open and closed subsets, even though there are multiple possibilities for what particular claims reported closed at 5^{th} later reopen. In fact, those values can be explicitly determined from the formulas derived in Appendix A.

Knowing the probabilities of the truly open and closed subsets, we adjust the loss model by proportionally shifting the probabilities. The probability of each open claim is increased by a constant factor while the probability of each closed claim is correspondingly decreased by another factor. Knowing the mean and variance of the truly open subset lets us adjust the undeveloped combined medical and indemnity loss amounts of the open claims to match the two revised moments for open claims; this is done via a power transformation as described in Appendix C. The closed claim loss amounts are similarly adjusted. The result is a model of empirical undeveloped losses that reflects a trued up claim status as of a 5^{th} report, in the sense that no closed claims will reopen. That model, in turn, provides the input to the dispersion calculation. This approach is a refinement from that of Gillam and Couret [5] who account for the reopening of just a very few closed claims by dispersing all closed claims by just a very little. The idea here is to perform the adjustment prior to dispersion so that it is exactly the set of 'truly closed' claims whose losses are deemed to be at their ultimate cost and it is the complement set of 'truly open' claims that are dispersed.

In the resulting model for the empirical undeveloped loss distribution, the claim status variable is assumed to be correct in the sense that the loss amount for each closed claim is taken to be the known ultimate loss on the claim. Dispersion is applied only to open claims. Accordingly, the LDF applicable to all claims is adjusted to one appropriate for open claims only, and all development occurs on exactly the open claims. For each state, injury type, and report, one average LDF is determined from the medical and indemnity LDFs to apply to the sum of the medical and indemnity incurred losses of each claim. That combined incurred LDF is then modified to apply to just the open claims. More precisely, the relationship used to focus an aggregate LDF onto just the open claims is simply:

$$\begin{aligned} L_c &= \text{Aggregate undeveloped loss for closed claims} \\ L_o &= \text{Aggregate undeveloped loss for open claims} \\ \lambda &= \text{Aggregate LDF applicable to all claims} \\ \widetilde{\lambda} &= \text{Open only LDF} \\ \lambda\left(L_c + L_o\right) &= L_c + \widetilde{\lambda} L_o \Rightarrow \widetilde{\lambda} = \lambda + (\lambda - 1)\frac{L_c}{L_o} \end{aligned}$$

The adjusted to open only LDFs are determined and applied by state, injury type, and report.

Even though the adjusted LDFs are applied to all open claims independent of loss size, because the proportion of claims that remain open correlates with size of loss, the application of dispersion varies by the size of loss layer. Typically, larger losses are more likely to be open, and this application of development factors will have a greater impact in the higher loss layers. It follows that the application of loss development changes the shape of the severity distribution, making it better reflect the ultimate loss severity distribution.

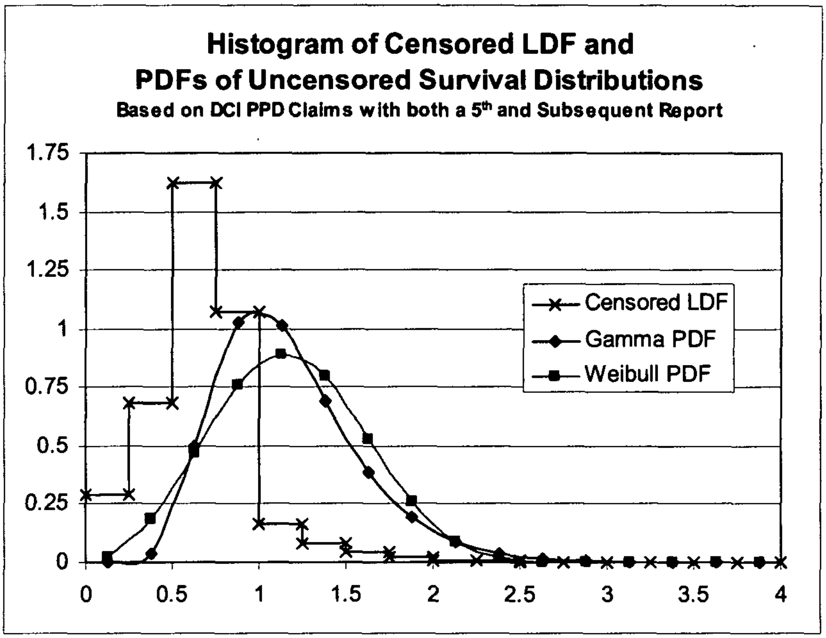
The next step is to apply dispersion to open claims. The technique used to disperse losses is formally equivalent to that used by Gillam and Couret [5]. The technique bears some similarity to kernel density estimation in which an assumed known density function (the kernel) is averaged across the observed data points so as to create a smoothed approximation. More precisely, the idea is to replace each open claim with a distribution of claims that reflect the various possibilities for the loss that is ultimately incurred on that claim. The expected loss at closure is just the applicable to ultimate LDF times the undeveloped loss. The LDF is varied according to an inverse transformed gamma distribution and multiplied by the undeveloped loss to model the possibilities for the ultimate loss.

The NCCI Detailed Claim Information (DCI) database was used to build a data set of observed LDFs beyond a 5^{th} report. We studied DCI claims open at 5^{th} report for which a subsequent DCI report was available. The observed LDF was determined as the ratio of the incurred loss at the latest available report divided by the incurred loss at 5^{th} report. If the claim remained open at that latest report, the observed LDF was considered "right censored." Censored regression of the kind used to study survival was used to fit this data. Open claims were identified as the censored observations, i.e. closed claims were deemed "dead" and open claims "alive" in the survival model. The survival model was used to determine an appropriate form to represent the distribution.

More precisely, the SAS PROC LIFEREG procedure was used to estimate accelerated failure time models from the LDF observations. Letting Y denote the observed LDF, the model was specified by the simplest possible equation $Y = \tilde{\lambda} + \varepsilon$, where $\tilde{\lambda}$ represents a constant and ε a variable error term. That is, the model specifies just an intercept term with no covariates at all. That model specification was selected because it corresponds to the application of a constant LDF ($\tilde{\lambda}$) to open claims. Moreover, the error term of the model corresponds precisely with dispersion, as that term is used here. Consequently, this application of survival analysis is somehat unconventional inasmuch as the issue is not the survival curve or the goodness of fit of the parameter estmate $\tilde{\lambda}$ that is key. Rather, the interest here is on the distribution of the error term ε. The SAS LIFEREG procedure is well suited to this because not only does it account for censored observations, it also allows for different structural forms to be assumed for the error term ε when estimating accelerated failure time models.

In this application, the estmated parameter for the intercept was not used since the LDF factors by state, injury type and report were taken from ratemaking data. What was of interest is the form and parameters that specify the error distribution. The Weibull, the Lognormal, the Gamma, and the generalized Gamma distribution were considered. In fact, the two-parameter Weibull, two-parameter Gamma, and the two-parameter Lognormal are all special cases of the three-parameter generalized Gamma (the Weibull and Gamma directly via parameter constraint, the Lognormal only asymptotically). The solutions for the generalized gamma implied that its three parameters enabled it to outperform the two parameter distributions. The three-parameter model guided the specification of the functional form and parameter values for the LDF distributions used in the dispersion calculation.

With the eventual goal to calculate excess ratios, it was important to assess whether the error term varies by size of loss. Gillam and Couret [5] assume that the CV of the dispersion distribution does not vary by size of loss. In addition to specifying different structural forms for the error term, models were fit to quintiles of the data, where by a quintile we mean that



the observations were divided into five equal volume groups according to claim size. It was observed that the CV of the error term did not show any significant variation by size of loss. This affirmed the prior assumption of a constant CV, and that assumption was again used in this dispersion calculation.

The LIFEREG procedure outputs the parameters that specify the dispersion pattern, by injury type, that relates a fifth report loss amount with the probable distribution of the incurred cost at "death" of the claim, i.e. at claim closure. Combining that with average LDFs from ratemaking, the uncensored distribution of the ultimate loss severity can be calculated. For any fixed open claim, the uncensored LDF distribution values times the (undeveloped) loss amount corresponds with the probable values for that claim at closure. It follows that the uncensored LDF distribution corresponds to age to ultimate LDFs applicable on a per open claim basis. The above chart illustrates how the survival model anticipates rightward movement of the reported empirical losses and fills out the right hand tail.

Because the mean LDF was already known, our primary focus was on the CV. This follows the approach of Gillam and Couret [5], whose decision to use a two-parameter gamma distribution for the reciprocal of the LDF was also followed. The use of the gamma to model the reciprocal amounts to the use of an inverse gamma for the LDF. That choice was reaffirmed by the DCI data and is illustrated somewhat in the above chart. We actually used a three-parameter inverse transformed gamma distribution, as the survival model suggested that would yield a better representation of the LDF distribution. The first two parameters, denoted α, τ in Klugman, et. al. [9] determine the CV of the distribution, which varies by report and injury type as indicated in the following table:

Injury	Report	α	τ	CV
Fatal & PT	3	5.7134	0.8	0.7
Fatal & PT	4	6.8664	0.8	0.6
Fatal & PT	5	8.7775	0.8	0.5
PP	5	8.7775	0.8	0.5
TT	5	12	3	0.1
Med Only	5	12	3	0.1

The third parameter, denoted β in Klugman, et. al. [9], determines the mean LDF and was directly solved for to make that mean equal the age to ultimate aggregate open claim LDF by state, report, and injury type. Even though open TT and Med only claims are not assumed to develop in aggregate (mean LDF = 1), the open TT and Med only claims are dispersed, but with a small CV.

Gillam and Couret [5] used a CV of 0.9 for the LDF on open claims; that selection was dictated to some degree by the need to account for potential unobserved large losses. The current ratemaking methodology makes separate provision for very large losses. This, in turn, enables this ELF revision to rely less on judgment and more on empirical data. The empirical data suggested the lower CVs used for the LDF distributions. All else equal, lowering the CV lowers the ELF at the largest attachment points. Much sensitivity analysis was done to assess the impact of this change in the assumed CV. It was determined that the selection did not represent an unreasonable reduction in the ELFs.

As is typical with kernel density models, Gillam and Couret [5] used a closed form integration formula to implement dispersion. However, in order to be able to perform the downstream data adjustments (in particular, adjusting to state conditions as discussed in the next section), we instead used the device of representing each open claim by 173 variants. The variants are determined by multiplying the undeveloped loss amount by 173 different LDFs. The variant LDFs have mean equal to the applicable overall LDF (as applicable to open claims only) and a CV of 0.5 for 5^{th} report Fatal, PT, and PP claims. The mean LDF applicable for medical only and TT cases is 1, as those cases are assumed not to develop in aggregate beyond a 5^{th} report. So even open medical only and TT claims are dispersed, albeit so as not to change the aggregate loss (and with a smaller CV of 0.1 for the LDF distribution). The choice of 173 points was done to enable the calculation to better capture the tail. Very small and very large LDFs are included in the model (corresponding to the 0.000001^{st} and 99.999999^{th} percentile of the inverse transformed gamma) albeit with a correspondingly very small weight (about 0.000001) being assigned to such variants. Dispersion does not change the contribution of any claim to the aggregate developed loss. It was determined that the use of 173 points provided a very close approximation to the continuous form. Additional details on that calculation can be found in Appendix B.

To summarize, the dispersion calculation starts with a finite probability space of claims together with a random variable giving the undeveloped claim values. Then both the probability measure and the random variable are adjusted to account for reopened claims. That gives a modified probability space of claims. Replacing each open claim with a distribution of 173 expected loss amounts at closure yields a developed dispersed probability space of claims with a random variable giving the ultimate claim value. This is done for each injury type and for all NCCI states. The next section describes how those random variables are adjusted to state specific conditions so as to yield the empirical distributions used in fitting the data to severity distributions.

3 Organization of Data

The idea of estimating excess ratios by injury type goes back at least to Uhthoff [11] and has been used as well by Harwayne [6], Gillam [4], and Gillam and Couret [5]. While we follow this approach as well, it should be noted that alternatives have recently been identified by Brooks [2] and Mahler [10].

Owing to the relatively few fatal and permanent total claims it is desirable to combine data across states. Differences between states preclude doing this without adjustment however. Gillam [4] addressed this by grouping states according to benefit structure. For an interesting recent approach incorporating benefit structure see Gleeson [7]. With the current dominance of medical costs this approach is less satisfactory. In the prior approach, Gillam and Couret [5] addressed the problem "by dividing each claim by the average cost per case for the appropriate state-injury-type combination." We refer to this data adjustment technique as mean normalization. This results in a countrywide database with mean of 1. Loss distributions were then fit to this normalized database. The countrywide loss distributions are then adjusted via a scale transformation (see Venter [12]) to be appropriate for each particular state. Thus the data for different states is adjusted to have the same mean. A natural variant of this would be median normalization, the thought being that the median might be more stable than the mean. A natural extension is to try and match more than one moment. We considered five data adjustment techniques altogether:

1. Mean Normalization
   As mentioned above, for a given injury type, each claim in state i, denoted by x_i (here x_i denotes the incurred loss on a claim from state i developed to ultimate), is transformed by $x_i \rightarrow x_i/\mu_i$, where μ_i denotes the mean of the x_i. The normalized claims for all states are now combined into a countrywide database. To get a database appropriate for state j, each normalized claim is then scaled up by the mean in state j, i.e. $x_i/\mu_i \rightarrow \mu_j \cdot x_i/\mu_i$

2. Median Normalization
   This is analogous to mean normalization, but claims are now normalized by the median rather than the mean.

3. Logarithmic Standardization
   A natural generalization of mean normalization would be to standardize claims, $x_i \rightarrow \frac{x_i - \mu_i}{\sigma_i}$. To avoid negative claim values when transforming the standardized database to a particular state we standardize the logged losses, $\log x_i \rightarrow \frac{\log x_i - \mu_i}{\sigma_i}$, where now μ_i, σ_i denote the mean and standard deviation of the logged losses. This results in a standardized countrywide database, which can then be adjusted to a given state j by $\frac{\log x_i - \mu_i}{\sigma_i} \rightarrow \sigma_j \cdot \frac{\log x_i - \mu_i}{\sigma_i} + \mu_j$. Appendix C discusses this in more detail.

4. Generalized Standardization
   This is analogous to logarithmic standardization except that instead of the mean and variance, percentiles can be used. For example, instead of the mean we could use the median and instead of the standard deviation we could use the 85^{th} percentile minus the median.

5. Power Transform
   Lastly, we considered a power transform, $x_i \rightarrow ax_i^b$, where the values of a and b are chosen so that the transformed values have the mean and variance of state j. That this is possible is shown in Appendix C. Thus for each state i there is a different power transform that takes the unadjusted state i claims and adjusts them to what they would be in state j, in the sense that the transformed claims from state i match the mean and variance in state j. Combining all of the adjusted claims results in an expanded state j specific database. Notice that the unadjusted state j claims appear in the expanded state j database and so the expanded state j database is indeed an expansion of the state j data. It should also be noted that the power transform generalizes both mean normalization and logarithmic standardization and the moments are matched in dollar space rather than in log space. This is discussed in more detail in Appendix C.

Extensive performance testing was conducted to decide which data adjustment techniques to use. The idea was to postulate realistic loss distributions for the states, based on realistic parameters, simulate data from the postulated loss distributions and see which techniques best recovered the postulated distributions. Initial tests showed that median normalization and generalized standardization performed poorly and so further tests concentrated on the remaining techniques. Based on our performance tests we chose to use logarithmic standardization for Fatal and Permanent Total (PT) claims and the power transform for Permanent Partial (PP), Temporary Total (TT), and Medical Only claims. It seemed that when there were only a limited number of claims and the difference in CVs between states was large the exponent in the power transform could occasionally be quite large, leading the power transform to underperform logarithmic standardization.

Gillam and Couret [5] call modeling PT and PP claims separately the "common sense approach." Owing to the scarcity of PT claims they have in the past been combined with Major PP claims. Due to our improved data adjustment techniques we are able to separate PT from PP. We also used data at 3^{rd}, 4^{th}, and 5^{th} report for Fatal and PT claims because of their relative scarcity, whereas we only used data at 5^{th} report for the other injury types.

In the prior approach, each state's weight in the countrywide database was proportional to the number of claims it contributed to the countrywide total. This seems implicitly like assigning a state's data a credibility of n/N, where n is the number of claims in the state and N is the countrywide total. Further, this implicit credibility did not vary by injury type. This makes sense when there is only one countrywide database. We however, use a different database for each state and give each state's data a weight of $\sqrt{n/N}$ in the state specific database, where n is the number of claims in the state and N is a standard based on actuarial judgment. Our view was that most states would have enough data to fit loss distributions for Medical Only, but that no state would have enough claims to fit a Fatal loss distribution and only the largest states would have enough PT claims. We thought it reasonable that three quarters of the states would have enough Medical Only claims, half of them would have enough TT claims and about a quarter of them would have enough PP claims. With this in mind, we chose N, the standard for full pooling weight, to be $2,000$ for Fatal claims, $1,500$ for PT claims, $7,000$ for PP claims, $8,500$ for TT claims and $20,000$ for Medical Only claims. It is intuitively sensible that the standard for Medical Only should be higher than for PT because excess ratios are driven by large claims and most PT claims are large whereas most Medical Only claims are typically small.

4 Fitting

Traditionally a parametric loss distribution would be fit to the entire data set by maximum likelihood. The first problem with this approach is that distributions which fit the tail well may not fit the small claims so well and thus there is a trade-off between fitting the tail well and fitting the small claims well. The need for a fitted loss distribution is really only in the tail as the number of small claims is quite large. Mahler [10] has recently used the empirical distribution for small claims and spliced a fitted loss distribution onto the tail. This is the approach we follow as well and we describe it in detail in Appendix E. Fitting the tail alone is of course much easier and the fits are much better than they have been in the past. The second problem with the traditional approach is that maximizing the likelihood function is somewhat indirect. While maximum likelihood fits typically result in loss distributions with excess ratio functions that do fit the data well, there is no intrinsic interest in the likelihood function itself. The primary objective is a loss distribution whose excess ratio function fits the data well and so instead of maximum likelihood we use least squares to fit the excess ratio function directly. Appendix D gives some general facts about excess ratio functions. In particular, Proposition 12 shows that a distribution is determined by its excess ratio function and so there is no loss of information in working with excess ratio functions rather than densities or distribution functions.

Mahler [10] uses a Pareto-exponential mixture to fit the tail. We use two to four term mixed exponentials. The mixed exponential distribution is described by Keatinge [8]. All things being equal, the mixed exponential is a thinner tailed distribution than has been used in the past. It has moments of all orders, whereas some loss distributions in use do not even have finite variances. However, the loss data used to fit the mixed exponential is driven by the inverse transformed gamma distribution of LDFs, as described in section 2, and the inverse transformed gamma is not a thin tailed distribution. This prevents the tail of the fitted loss distribution from being too thin. The mixed exponential also has an increasing mean residual life, and this is quite typical of Workers Compensation claim data. Fat tailed distributions may make sense in the presence of catastrophic loss potential, but recently NCCI has made a separate CAT filing so the new ELFs are for the first time explicitly non-CAT. From a geometrical perspective, the density function over the tail region should be decreasing and have no inflection points, as occurs where the first derivative of the density function is negative and its second derivative is positive. The mixed exponential class of distributions has alternating sign derivatives of all orders. And conversely any distribution with alternating sign derivatives of all orders can be approximated by a mixed exponential to within any desired degree of accuracy. Functions with this alternating derivative property are called completely monotone and this characterization of them follows from a theorem by Bernstein. (See Feller [3].) We initially considered using other distributions besides the mixed exponential, but the mixed exponential fits were so good that it was not necessary to consider other distributions further.

Mahler [10] noted that the excess ratios are not very sensitive to the splice point, i.e. the point where the empirical data ends and the tail fit begins. We found that to be the case as well. We were concerned with large losses being under represented in the data. Thus we preferred to not attach too far out into the tail so that we could have some confidence in the tail probability, i.e. the probability of a claim being greater than the splice point. So we generally chose splice points that resulted in a tail probability between 5% and 15%. While this gave us some confidence in the tail probability, we were still concerned about claims in the \$10 million to \$50 million range being under represented in the data. (Claims larger than \$50 million would be accounted for in the separate CAT filing.) The new excess ratios are based on one to three years of data, depending on the injury type, but the largest WC claims and events occur with return periods exceeding three years. WC catastrophe modeling indicates that claims and occurrences in the \$10 million to \$50 million range are underrepresented in the data used to fit the new curves. Because of this, we included an additional provision for individual claims and occurrences between \$10 million and \$50 million. This new provision is broadly grounded in the results of several WC catastrophe models, and known large WC occurrences. Previous excess ratio curves included a provision for anti-selection of 0.005, which has been eliminated in the new curves. The new provision, per-claim or per-occurrence, is .003 up to \$10 million, 0 for \$50 million or greater, and declines linearly from .003 to 0 between \$10 million and \$50 million. Thus the final adjusted excess ratio is 0.997 times the excess ratio before this adjustment, plus this adjustment. That is, if L is the loss limit and $R(L)$ is the unadjusted per claim or per occurrence excess ratio, then the adjusted excess ratio is given by

$$R'(L) = \begin{cases} .997R(L) + .003 & \text{if } \ L \leq \$10M \\ .997R(L) - \frac{.003}{\$40M}L + .00375 & \text{if } \ \$10M < L < \$50M \\ .997R(L) & \text{if } \ L \geq \$50M \end{cases} .$$

5 Modelling Occurrences

Data is typically collected on a per claim basis. This makes it a challenge to produce per occurrence excess ratios. The first attempt to address this was to merely increase the per claim excess ratios by 10% to account for occurrences. For low attachment points this could lead to excess ratios greater than 1. Gillam [4] improved this approach by assuming only that the average occurrence cost 10% more than the average claim. This affects the entry ratio used to compute the excess ratio. Gillam and Couret [5] then refined this approach still further by breaking down the 10% by injury type: 3.9% for fatal injuries, 6.6% for permanent total and major permanent injuries, and 0% for minor permanent partial and temporary total injuries. These approaches, while reasonable, rely heavily on actuarial judgment.

The first attempt to base per occurrence excess ratios more solidly on per occurrence data was by Mahler [10], who attempted to group claims into occurrences based on hazard group, accident date, and policy number. NCCI has a CAT[1] code which identifies claims in multiple claim occurrences. Singleton claims (occurrences with only one claim) have a CAT code of 00, all claims in the first multi-claim occurrence would have a CAT code of 01, claims in the second multi-claim occurrence would have a CAT code of 02, etc. Unfortunately there were several problems with the CAT code:

1. missing CAT codes
   For singleton claims it is permissible to report a blank field for the CAT code. This would then be converted to a 00. However there was no way of knowing whether a blank field was deliberately reported as a blank or inadvertently omitted.

2. orphans
   There were claims observed with nonzero CAT codes, but with no other claims with the same CAT code. One carrier, for example, appeared to have numbered the claims in a multiple claim occurrence sequentially.

3. variance in injury dates
   Claims were observed with the same CAT code, but with different injury dates. In one case the injury dates were 14 months apart.

4. grouping of CAT claims
   It is permissible to group small med only claims in reporting. This is not permissible however in the case of CAT claims. Nevertheless there was some evidence of grouped reporting for CAT claims.

Further complicating things was the fact that even with optimal reporting, multiple claim occurrences appear to be extremely rare. Based on an examination of data from carriers known to report their data well, it would appear that .2% is a reasonable estimate of the portion of all claims that

[1]Here a catastrophe is merely an occurrence with more than one claim. The term 'catastrophe' in this context has no implications as to the size of the occurrence.

occur as part of multi-claim occurrences. Based on the above problems, we decided not to try and build a per occurrence data base, but rather to use a collective risk model. From the per claim loss distributions we could easily get an overall per claim severity distribution. We estimated the frequency distribution for multiple claim occurrences from carriers thought to have recorded the CAT code correctly. The mean number of claims in a multiple claim occurrence is about 3, but most multiple claim occurrences consist of two claims.

Unfortunately the severity distribution of claims in multiple claim occurrences seemed to be different from the severity distribution of singleton claims. First, the mix of injury types in multiple claim occurrences was more severe than in singleton claims. Second, even when fixing an injury type, claims occurring as part of a multiple claim occurrence were more severe. We chose to address this issue by assuming that the severity distribution of claims in multiple claim occurrences differed from the distribution of singletons only by a scale transformation. This assumption goes at least as far back as Venter [12].

More formally, let X_i be the random variable giving the cost of a singleton claim of injury type i and let F_{X_i} be the distribution function of X_i. If S is the random variable giving the overall cost of a singleton occurrence then $F_S = \sum w_i F_{X_i}$, where w_i is the probability that a singleton claim is of injury type i. That is, the per claim severity distribution is a mixture of the injury type distributions. If Y_i is the random variable giving the cost of a claim of injury type i in a multiple claim occurrence then we assume that Y_i differs from X_i by a scale transform, i.e. $Y_i = a_i X_i$ for some constant a_i. If Z is the random variable giving the overall cost of a claim in a multiple claim occurrence then $F_Z = \sum w_i' F_{Y_i}$, where w_i' is the probability that a claim in a multiple claim occurrence is of injury type i. Then $M = Z_1 + \cdots + Z_N$ is the cost of a multiple claim occurrence, where N is the random variable giving the number of claims in a multiple claim occurrence and the Z_i are iid random variables with the same distribution as Z. Finally, the per occurrence severity distribution is given by $F = rF_S + (1-r)F_M$, where r is the probability that an occurrence consists of a single claim.

Because r is so close to 1 there is very little difference between per claim and per occurrence loss distributions. Per occurrence excess ratios are no more than .2% more than per claim excess ratios. This is a sharp contrast with the prior approaches.

6 Updating

Overall excess ratios are computed as a weighted average of the injury type excess ratios. Let $R(L)$ be the overall excess ratio at a loss limit of L, and let $R_i(r)$ be the excess ratio for injury type i at an entry ratio of r, then

$$R(L) = \sum_i w_i R_i(L/\mu_i),$$

where w_i is the percentage of losses of type i and μ_i is the mean loss of type i. The injury type weights, w_i, and average costs per case, μ_i, are updated annually, but the injury type excess ratio functions, R_i, are updated only infrequently. The idea is that the shape of the loss distributions changes much more slowly than the scale. The annual update thus involves adjusting the mix of injury types and adjusting the loss distributions by a scale transformation. Updating via a scale transformation is extremely convenient and is discussed by Venter [12].

The key question is how to determine when a simple scale transformation update is adequate and when the loss distributions need to be refit. If X is the random variable corresponding to last year's loss distribution and Y is the random variable corresponding to this year's loss distribution, then the scale transformation updating assumption is that there is some constant, c, such that Y and cX have the same distribution. Then the normalized distribution, Y/μ_Y has the same distribution as $cX/c\mu_X = X/\mu_X$ and thus $\mathrm{Var}(Y/\mu_Y) = \mathrm{Var}(X/\mu_X) = \sigma_X^2/\mu_X^2 = CV_X^2$. So if successive year's loss distributions really did differ only by a scale transform then the CV would remain constant over time. Thus monitoring the CV over time might give a criterion for when it is necessary to update the underlying loss distributions and not just the injury type weights and average costs per case.

Since the injury type loss distributions are normalized to have mean 1, applying a uniform trend factor would have no impact. Thus the losses used for fitting are typically not trended to a future effective date. This is extremely convenient in that it does not require us to decide in advance when the loss distributions need to be updated. However, if the trend is not uniform, then it could result in a change in the shape of the loss distributions. This could for instance happen if there was a persistent difference in medical and indemnity trends and the percentage of loss due to medical costs varied by claim size, as it typically does, even after controlling for injury type. How significant this phenomenon is remains an open question. It is in some sense limited as medical trends cannot exceed inflation forever without the medical sector consuming an unacceptably large fraction of GDP. Nevertheless, this does suggest that monitoring the difference in cumulative medical and indemnity trends might provide a guide as to when the shape of the loss distributions needs to be updated.

7 Conclusion

With the present revision we have implemented several changes to the methodology as summarized in the table below. We retained the general approach to dispersion of individual claim development due to Gillam and Couret [5], using an inverse transformed gamma for the distribution of LDFs, but lowering the CV from .9 to .5. Instead of fitting a loss distribution to all of the claims, we followed Mahler [10] and fit only the tail, using the empirical distribution for the small claims. For the tail we used a mixed exponential as compared to the prior transformed betas fit to the entire distribution. Instead of combining PT with Major PP claims, we fit PT and PP claims separately, using data at 3^{rd}, 4^{th}, and 5^{th} report for Fatal and PT claims. The prior approach used only data at 5^{th} report. To adjust the data from one state to be comparable with another state we used logarithimic standardization for Fatal and PT claims and power transforms for PP, TT, and Med Only. The prior approach was to use mean normalization for all injury types. We then fit state specific loss distributions rather than the countywide ones used before. Finally, to go from per claim data to per occurrence ELFs we used a collective risk model of occurrences. This contrasts sharply with prior approaches based on estimates of how much the mean occurrence cost exceeded the mean claim cost. The prior approach implicitly assumed a 3.9% load for Fatal claims, a 6.6% load for PT/Major PP claims, and a 0% load for TT and Med Only claims.

	new approach	prior approach
dispersion	CV = .5	CV = .9
fitting	fit tail only	fit whole distribution
form of distribution	empirical/mixed exponential	transformed beta
injury types	PT, PP separate	PT, Major PP combined
data	3^{rd}, 4^{th}, 5^{th} report for F, PT	5^{th} report
data adjustment	logarithmic standardization, power transform	mean normalization
applicability of distributions	state specific	countrywide
per occurrence	collective risk	3.9% F, 6.6% PT/Maj PP

While the changes made to the ELF methodology were significant, they were more evolutionary than revolutionary. Nevertheless, the new ELFs are quite a bit lower than the old ones at the larger limits in many states. We examined carefully the impact of the change in the dispersion CV and the use of mixed exponential rather than transformed beta distributions. Had we used a dispersion CV of 0.9 rather than 0.5, the ELFs would have been higher than the new ones. But at the higher limits, where the decrease was most pronounced, ELFs based on a CV of 0.9 would still be much closer to the new ELFs than the old. We also refit the old transformed beta distributions to the new data and found that even with the old distributional forms, fit to the entire distribution, the result is a much thinner tail than in the distributions underlying the old ELFs. We thus concluded that changes in the empirical loss distributions underlying the prior and the revised ELFs are what drive the reduction in ELFs. The prior review of ELFs relied on data that preceded the decline of WC claim frequency that so dominated WC experience in the 1990s, and beyond. There are solid theoretical reasons to suggest that this is just the sort of dynamic that can significantly change the shape of the loss distributions in a fashion that may not be captured by scale adjustments and as such require the development of new ELFs.

APPENDIX

A Adjusting for Reopened Claims

This appendix details some calculations referenced in section 2 on developing individual claims, in particular on the treatment of reopened claims. We consider a set of observed individual claims grouped by their open/closed claim status and determine how the first two moments of the open and closed subsets change when some claims are 'reopened,' i.e. when some claims are reclassified from the closed to the open subset. The discussion applies quite generally to show how the first two moments are impacted by a change in a characteristic, like claim status, to a selected subset of observations. The mean and variance of a finite set of observed values have natural generalizations to vector valued observations. It is convenient to express the findings as they apply in a multi-dimensional context, even though the specific application in this paper requires only the one-dimensional case.

Suppose we have a finite set of claims C and that a vector $x_c \in \mathbb{R}^n$ is associated with each $c \in C$. Suppose each $c \in C$ is also assigned a probability of occurrence $\omega_c > 0$ For any nonempty subset $A \subseteq C$, we make the following definitions

$$\begin{aligned}
\text{Probability of the set } A &= |A|_\omega = \sum_{a\in A} \omega_a \\
\text{Mean of } A &= \mu_A = \frac{1}{|A|_\omega} \sum_{a\in A} \omega_a x_a \in \mathbb{R}^n \\
\text{Variance of } A &= \sigma_A^2 = \frac{1}{|A|_\omega} \sum_{a\in A} \omega_a \left\| x_a - \mu_A \right\|^2 \geq 0
\end{aligned}$$

and we make the usual convention that for the empty set $|\phi|_\omega = \sigma_\phi^2 = 0$ and $\mu_\phi = \overrightarrow{0}$ is the 0-vector.

Observe that the mean is a vector and the variance a scalar and that for $n = 1$ this defines the mean and variance associated with the probability density function $f(a) = \frac{\omega_a}{|A|_\omega}$ on A when we view the subset A as a probability space in its own right. A natural WC application of multi-dimensionality is the case $n = 2$ in which the first coordinate measures the indemnity loss amount and the second component the medical loss of a claim $c \in C$. Note that we have the usual relationship between the mean, the variance and the second moment:

$$\begin{aligned}
\sigma_A^2 &= \frac{1}{|A|_\omega}\sum_{a\in A}\omega_a \|x_a-\mu_A\|^2 = \frac{1}{|A|_\omega}\sum_{a\in A}\omega_a (x_a-\mu_A)\cdot(x_a-\mu_A) \\
&= \frac{1}{|A|_\omega}\sum_{a\in A}\omega_a (x_a\cdot x_a - 2\mu_A\cdot x_a + \mu_A\cdot\mu_A) \\
&= \frac{1}{|A|_\omega}\sum_{a\in A}\omega_a \|x_a\|^2 - 2\left(\mu_A\cdot\frac{1}{|A|_\omega}\sum_{a\in A}\omega_a x_a\right) + \frac{|A|_\omega}{|A|_\omega}\|\mu_A\|^2 \\
&= \frac{1}{|A|_\omega}\sum_{a\in A}\omega_a \|x_a\|^2 - 2(\mu_A\cdot\mu_A) + \|\mu_A\|^2 \\
&= \frac{1}{|A|_\omega}\sum_{a\in A}\omega_a \|x_a\|^2 - 2\|\mu_A\|^2 + \|\mu_A\|^2 = \frac{1}{|A|_\omega}\sum_{a\in A}\omega_a \|x_a\|^2 - \|\mu_A\|^2
\end{aligned}$$

And thus

$$\|\mu_A\|^2 + \sigma_A^2 = \frac{1}{|A|_\omega}\sum_{a\in A}\omega_a \|x_a\|^2 .$$

There are the evident relationships with the union and intersection of subsets $A, B \subseteq C$; for the mean we have:

$$\begin{aligned}
\mu_{A\cup B} &= \frac{1}{|A\cup B|_\omega}\sum_{c\in A\cup B}\omega_c x_c = \frac{1}{|A\cup B|_\omega}\left(\sum_{a\in A}\omega_a x_a + \sum_{b\in B}\omega_b x_b - \sum_{c\in A\cap B}\omega_c x_c\right) \\
&= \frac{1}{|A\cup B|_\omega}\left(|A|_\omega\mu_A + |B|_\omega\mu_B - |A\cap B|_\omega\mu_{A\cap B}\right)
\end{aligned}$$

And thus

$$\mu_{A\cup B} + \frac{|A\cap B|_\omega}{|A\cup B|_\omega}\mu_{A\cap B} = \frac{|A|_\omega}{|A\cup B|_\omega}\mu_A + \frac{|B|_\omega}{|A\cup B|_\omega}\mu_B.$$

and similarly for the variance:

$$\begin{aligned}
|A\cup B|_\omega\left(\|\mu_{A\cup B}\|^2 + \sigma_{A\cup B}^2\right) &= \sum_{c\in A\cup B}\omega_c\|x_c\|^2 \\
&= \sum_{a\in A}\omega_a\|x_a\|^2 + \sum_{b\in B}\omega_b\|x_b\|^2 - \sum_{c\in A\cap B}\omega_c\|x_c\|^2 \\
&= |A|_\omega\left(\|\mu_A\|^2+\sigma_A^2\right) + |B|_\omega\left(\|\mu_B\|^2+\sigma_B^2\right) \\
&\quad - |A\cap B|_\omega\left(\|\mu_{A\cap B}\|^2 + \sigma_{A\cap B}^2\right)
\end{aligned}$$

And thus

$$\begin{aligned}
\sigma^2_{A\cup B} + \frac{|A\cap B|_\omega}{|A\cup B|_\omega}\sigma^2_{A\cap B} &= \frac{|A|_\omega}{|A\cup B|_\omega}\sigma^2_A + \frac{|B|_\omega}{|A\cup B|_\omega}\sigma^2_B \\
&+\frac{1}{|A\cup B|_\omega}\left(|A|_\omega \|\mu_A\|^2 + |B|_\omega \|\mu_B\|^2 - |A\cap B|_\omega \|\mu_{A\cap B}\|^2\right) \\
&- \|\mu_{A\cup B}\|^2
\end{aligned}$$

We are especially interested in the case when C is a disjoint union, so we make the assumption:

$$C = A\cup B \qquad A\cap B = \phi \qquad A \neq \phi$$

Think of the decomposition as reflecting a two-valued claim status, like open and closed. The goal is to determine how the mean and variance change after "moving" a subset D from A to B. The example of this paper is when the claim decomposition reflects claim closure status as of a 5^{th} report, ($A =$ closed and B=open) and D is a set of closed claims that reopen after a 5^{th} report.

In this case of a disjoint union, it is especially easy to express μ_C and σ^2_C in terms of the corresponding statistics for A and B. From the above formula for the mean of a union:

$$\begin{aligned}
\mu_C &= \mu_{A\cup B} + \overrightarrow{0} = \mu_{A\cup B} + \frac{|A\cap B|_\omega}{|A\cup B|_\omega}\mu_{A\cap B} \\
&= \frac{|A|_\omega}{|A\cup B|_\omega}\mu_A + \frac{|B|_\omega}{|A\cup B|_\omega}\mu_B. \\
&= w\mu_A + (1-w)\,\mu_B \quad \text{where } w = \frac{|A|_\omega}{|C|_\omega} \in (0,1].
\end{aligned}$$

The second moments are similarly weighted averages, with the same subset weights w and $1-w$. From what we just saw for the mean of a disjoint union combined with the above formula for the variance of a union:

$$\begin{aligned}
\sigma^2_C &= \sigma^2_{A\cup B} + 0 = \sigma^2_{A\cup B} + \frac{|A\cap B|_\omega}{|A\cup B|_\omega}\sigma^2_{A\cap B} \\
&= w\sigma^2_A + (1-w)\,\sigma^2_B + w\,\|\mu_A\|^2 + (1-w)\,\|\mu_B\|^2 - \|w\mu_A + (1-w)\,\mu_B\|^2 \\
&= w\sigma^2_A + (1-w)\,\sigma^2_B + w\,\|\mu_A\|^2 + (1-w)\,\|\mu_B\|^2 \\
&\quad -w^2\,\|\mu_A\|^2 - 2w\,(1-w)\,\mu_A\cdot\mu_B - (1-w)^2\,\|\mu_B\|^2 \\
&= w\sigma^2_A + (1-w)\,\sigma^2_B + w\,(1-w)\left(\|\mu_A\|^2 - 2\mu_A\cdot\mu_B + \|\mu_B\|^2\right) \\
&= w\sigma^2_A + (1-w)\,\sigma^2_B + w\,(1-w)\,\|\mu_A - \mu_B\|^2
\end{aligned}$$

This expresses the variance of a disjoint union in terms of the means and variances of the subsets.

Notice that these formulas for μ_C and σ_C^2 show how the mean and variance of the subset A are constrained by those of the superset C. For the remainder of this appendix we assume $\sigma_C > 0$ and so we have:

$$\sigma_C^2 = w\sigma_A^2 + (1-w)\,\sigma_B^2 + w\,(1-w)\,\|\mu_A - \mu_B\|^2 \geq w\sigma_A^2 \Rightarrow w\left(\frac{\sigma_A}{\sigma_C}\right)^2 \leq 1$$

Observe that assigning the difference vector δ and scalar ratio r as:

$$\delta = \mu_A - \mu_C \qquad\qquad r = \frac{\sigma_A}{\sigma_C}$$

then we also have:

$$\begin{aligned}
\mu_C &= w\,(\mu_C + \delta) + (1-w)\,\mu_B \Rightarrow \mu_B = \mu_C - \frac{w\delta}{1-w} \\
&\Rightarrow \mu_A - \mu_B = \mu_C + \delta - \left(\mu_C - \frac{w\delta}{1-w}\right) = \frac{\delta\,(1-w) + w\delta}{1-w} = \frac{\delta}{1-w}
\end{aligned}$$

But then:

$$\begin{aligned}
\sigma_C^2 &= w\sigma_A^2 + (1-w)\,\sigma_B^2 + w\,(1-w)\left\|\frac{\delta}{1-w}\right\|^2 \geq w\sigma_A^2 + \frac{w\,\|\delta\|^2}{1-w} = wr^2\sigma_C^2 + \frac{w\,\|\delta\|^2}{1-w} \\
&\Rightarrow \left(1 - wr^2\right)\sigma_C^2 \geq \frac{w\,\|\delta\|^2}{1-w} \Rightarrow r \leq \sqrt{\frac{1}{w}} \text{ and } \|\delta\| \leq \sigma_C\sqrt{\frac{(1-w)\,(1-wr^2)}{w}}
\end{aligned}$$

and we see how, for any nonempty subset A, the mean difference vector δ is constrained by the probability allocation together with the deviation ratio r and the standard deviation of C.

Now suppose we have "local information" on how the proper subset $D \subset A$ fits within A, captured in the two numbers p, r and the difference vector δ:

$$\begin{aligned}
p &= \frac{|D|_\omega}{|A|_\omega} \\
r\sigma_A &= \sigma_D \\
\delta &= \mu_D - \mu_A
\end{aligned}$$

in which we specify that $r = 1$ should $\sigma_A = 0$. From what we've just seen, applying the above to any nonempty subset $D \subset A$, the following two inequalities must hold:

$$r \leq \sqrt{\frac{1}{p}} \qquad \|\delta\| \leq \sigma_A \sqrt{\frac{(1-p)(1-pr^2)}{p}}$$

Define the sets:

$$\begin{aligned} \widehat{A} &= A \backslash D = \{a \in A | a \notin D\} \\ \widehat{B} &= B \cup D \\ &\Rightarrow C = \widehat{A} \cup \widehat{B} \qquad \widehat{A} \cap \widehat{B} = \phi \qquad \widehat{A} \neq \phi \neq \widehat{B} \end{aligned}$$

In terms of the above open/closed claim example, this second decomposition represents the "truly closed" verses the "truly open" claims, as of a 5^{th} report.

With transparent notation, we seek to determine the subset probability and the moments $\widehat{w}$, $\mu_{\widehat{A}}$, $\mu_{\widehat{B}}$, $\sigma_{\widehat{A}}$, $\sigma_{\widehat{B}}$ in terms of the original subset probability and moments w, μ_A, μ_B, σ_A, σ_B together with the local information p, r and δ. The calculations only require some persistence:

$$\begin{aligned} p &= \frac{|D|_\omega}{|A|_\omega} \Rightarrow |D|_\omega = p\,|A|_\omega \Rightarrow \left|\widehat{A}\right|_\omega = |A|_\omega - |D|_\omega = |A|_\omega - p\,|A|_\omega = (1-p)\,|A \\ &\Rightarrow \quad \widehat{w} = \frac{\left|\widehat{A}\right|_\omega}{|C|_\omega} = \frac{\left|\widehat{A}\right|_\omega}{|A|_\omega}\frac{|A|_\omega}{|C|_\omega} = (1-p)\,w \end{aligned}$$

Continuing in turn, we have:

$$\begin{aligned} \mu_A &= p\mu_D + (1-p)\,\mu_{\widehat{A}} = p\,(\mu_A + \delta) + (1-p)\,\mu_{\widehat{A}} \\ &\Rightarrow \quad (1-p)\,\mu_{\widehat{A}} = \mu_A - p\mu_A - p\delta = (1-p)\,\mu_A - p\delta \\ &\Rightarrow \quad \mu_{\widehat{A}} = \mu_A - \left(\frac{p}{1-p}\right)\delta. \end{aligned}$$

And since we now know $\widehat{w}$ and $\mu_{\widehat{A}}$, we determine $\mu_{\widehat{B}}$ from:

$$\mu_C = \widehat{w}\mu_{\widehat{A}} + (1 - \widehat{w})\,\mu_{\widehat{B}} \Rightarrow \mu_{\widehat{B}} = \frac{\mu_C - \widehat{w}\mu_{\widehat{A}}}{1 - \widehat{w}}$$

And we get $\sigma_{\widehat{A}}$ from:

$$\begin{aligned} \sigma_A^2 &= p\sigma_D^2 + (1-p)\,\sigma_{\widehat{A}}^2 + p\,(1-p)\left\|\mu_{\widehat{A}} - \mu_D\right\|^2 \\ &\Rightarrow \quad \sigma_{\widehat{A}}^2 = \frac{\sigma_A^2 - p\sigma_D^2}{1-p} - p\left\|\mu_{\widehat{A}} - \mu_D\right\|^2 \end{aligned}$$

And finally, we can obtain $\sigma_{\widehat{B}}$ from:

$$\begin{aligned}
\sigma_C^2 &= \widehat{w}\sigma_{\widehat{A}}^2 + (1-\widehat{w})\,\sigma_{\widehat{B}}^2 + \widehat{w}\,(1-\widehat{w})\left\|\mu_{\widehat{A}} - \mu_{\widehat{B}}\right\|^2 \\
\Rightarrow \quad \sigma_{\widehat{B}}^2 &= \frac{\sigma_C^2 - \widehat{w}\sigma_{\widehat{A}}^2}{1-\widehat{w}} - \widehat{w}\left\|\mu_{\widehat{A}} - \mu_{\widehat{B}}\right\|^2
\end{aligned}$$

The requisite formulas for the adjusted moments and subset probabilities are summarized in the following proposition:

Proposition 1 *Let $C = A \cup B$ be a decomposition of C into mutually exclusive subsets, as above, and suppose D is a proper subset of A and set*

$$\begin{aligned}
w &= \frac{|A|_\omega}{|C|_\omega} \\
p &= \frac{|D|_\omega}{|A|_\omega} \\
\delta &= \mu_D - \mu_A.
\end{aligned}$$

Then for the alternative decomposition $C = \widehat{A} \cup \widehat{B}$ where

$$\begin{aligned}
\widehat{A} &= A \backslash D = \{a \in A | a \notin D\} \\
\widehat{B} &= B \cup D \\
\widehat{w} &= \frac{\left|\widehat{A}\right|_\omega}{|C|_\omega}
\end{aligned}$$

we have:

$$\begin{aligned}
\phi &= \widehat{A} \cap \widehat{B} \\
\widehat{A} &\neq \phi \neq \widehat{B} \\
\widehat{w} &= (1-p)\,w \\
\mu_{\widehat{A}} &= \mu_A - \left(\frac{p}{1-p}\right)\delta \\
\mu_{\widehat{B}} &= \frac{\mu_C - \widehat{w}\mu_{\widehat{A}}}{1-\widehat{w}} \\
\sigma_{\widehat{A}}^2 &= \frac{\sigma_A^2 - p\sigma_D^2}{1-p} - p\left\|\mu_{\widehat{A}} - \mu_D\right\|^2 \\
\sigma_{\widehat{B}}^2 &= \frac{\sigma_C^2 - \widehat{w}\sigma_{\widehat{A}}^2}{1-\widehat{w}} - \widehat{w}\left\|\mu_{\widehat{A}} - \mu_{\widehat{B}}\right\|^2
\end{aligned}$$

Proof. Clear from the above. ■

It is straightforward to generalize the formulas that express the mean and variance of a disjoint union of two sets to apply to partitions of more than two sets. The formula for the mean is immediate:

$$
\begin{aligned}
C &= \bigcup_{i=1}^{m} A_i \qquad\qquad A_i \cap A_j = \phi \text{ for } i \neq j \qquad\qquad w_i = \frac{|A_i|_\omega}{|C|_\omega} > 0 \\
\mu_C &= \frac{1}{|C|_\omega} \sum_{c \in C} \omega_c x_c = \frac{1}{|C|_\omega} \sum_{i=1}^{m} \sum_{a \in A_i} \omega_a x_a = \frac{1}{|C|_\omega} \sum_{i=1}^{m} \frac{|A_i|_\omega}{|A_i|_\omega} \sum_{a \in A_i} \omega_a x_a \\
&= \frac{1}{|C|_\omega} \sum_{i=1}^{m} |A_i|_\omega \, \mu_{A_i} = \sum_{i=1}^{m} w_i \mu_{A_i}
\end{aligned}
$$

and for the variance we first consider the expression for the second moment:

$$
\begin{aligned}
\|\mu_C\|^2 + \sigma_C^2 &= \frac{1}{|C|_\omega} \sum_{c \in C} \omega_c \|x_c\|^2 = \frac{1}{|C|_\omega} \sum_{i=1}^{m} \sum_{a \in A_i} \omega_a \|x_a\|^2 \\
&= \frac{1}{|C|_\omega} \sum_{i=1}^{m} |A_i|_\omega \left(\|\mu_{A_i}\|^2 + \sigma_{A_i}^2 \right) \\
&= \sum_{i=1}^{m} w_i \left(\|\mu_{A_i}\|^2 + \sigma_{A_i}^2 \right) = \sum_{i=1}^{m} w_i \|\mu_{A_i}\|^2 + \sum_{i=1}^{m} w_i \sigma_{A_i}^2
\end{aligned}
$$

and we find that:

$$\begin{aligned}
\sigma_C^2 &= \sum_{i=1}^{m} w_i \left\|\mu_{A_i}\right\|^2 + \sum_{i=1}^{m} w_i\sigma_{A_i}^2 - \left\|\mu_C\right\|^2 \\
&= \sum_{i=1}^{m} w_i \left(\mu_{A_i} \cdot \mu_{A_i}\right) + \sum_{i=1}^{m} w_i\sigma_{A_i}^2 - \left(\sum_{i=1}^{m} w_i\mu_{A_i}\right) \cdot \left(\sum_{i=1}^{m} w_i\mu_{A_i}\right) \\
&= \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i=1}^{m} w_i \left(\mu_{A_i} \cdot \mu_{A_i}\right) - \left(\sum_{i=1}^{m} w_i^2 \left(\mu_{A_i} \cdot \mu_{A_i}\right) + 2\sum_{i<j} w_i w_j \left(\mu_{A_i} \cdot \mu_{A_j}\right)\right) \\
&= \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i=1}^{m} \left(w_i - w_i^2\right) \left(\mu_{A_i} \cdot \mu_{A_i}\right) - 2\sum_{i<j} w_i w_j \left(\mu_{A_i} \cdot \mu_{A_j}\right) \\
&= \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i=1}^{m} w_i \left(1 - w_i\right) \left(\mu_{A_i} \cdot \mu_{A_i}\right) - 2\sum_{i<j} w_i w_j \left(\mu_{A_i} \cdot \mu_{A_j}\right) \\
&= \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i=1}^{m} w_i \left(\sum_{j\neq i} w_j\right) \left(\mu_{A_i} \cdot \mu_{A_i}\right) - 2\sum_{i<j} w_i w_j \left(\mu_{A_i} \cdot \mu_{A_j}\right) \\
&= \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i<j} w_i w_j \left(\mu_{A_i} \cdot \mu_{A_i} + \mu_{A_j} \cdot \mu_{A_j}\right) - 2\sum_{i<j} w_i w_j \left(\mu_{A_i} \cdot \mu_{A_j}\right) \\
&= \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i<j} w_i w_j \left(\mu_{A_i} \cdot \mu_{A_i} + \mu_{A_j} \cdot \mu_{A_j} - 2\mu_{A_i} \cdot \mu_{A_j}\right) \\
&= \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i<j} w_i w_j \left(\mu_{A_i} - \mu_{A_j}\right) \cdot \left(\mu_{A_i} - \mu_{A_j}\right) \\
&= \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i<j} w_i w_j \left\|\mu_{A_i} - \mu_{A_j}\right\|^2
\end{aligned}$$

and the generalization of the formula for the variance of a partition is:

$$\sigma_C^2 = \sum_{i=1}^{m} w_i\sigma_{A_i}^2 + \sum_{i<j} w_i w_j \left\|\mu_{A_i} - \mu_{A_j}\right\|^2.$$

Consider the special case of the set of m mean vectors $M = \left\{\mu_{A_i}\right\}$ expressed as a disjoint union of singleton subsets in which the vector μ_{A_i} is assigned the probability w_i. Then the formula gives:

$$\begin{aligned}
\sigma_M^2 &= \sum_{i=1}^{m} w_i \sigma^2_{\{\mu_{A_i}\}} + \sum_{i<j} w_i w_j \left\| \mu_{A_i} - \mu_{A_j} \right\|^2 \\
&= \sum_{i=1}^{m} w_i (0) + \sum_{i<j} w_i w_j \left\| \mu_{A_i} - \mu_{A_j} \right\|^2 \\
&= \sum_{i<j} w_i w_j \left\| \mu_{A_i} - \mu_{A_j} \right\|^2
\end{aligned}$$

But this is just the second term in the earlier expression for σ_C^2 and we find that

$$\sigma_C^2 = \sum_{i=1}^{m} w_i \sigma_{A_i}^2 + \sigma_M^2$$

which generalizes the usual decomposition of the variance into the sum of the within and the between variance. This has application to cluster analysis, where it affords a useful geometrical interpretation. In cluster analysis it is common to work with vectors so as to capture the influence of multiple data fields. So as above assume each claim $c \in C$ is assigned a vector of values that captures information about the claim that we seek to organize into a classification scheme. Viewing the m subsets $A_i \subseteq C$ as defining clusters of vectors, the set of m mean vectors $M = \{\mu_{A_i}\}$ is the set of 'centroids' of those clusters. The goal of cluster analysis is to separate the data into like clusters, but there is both a local and a global perspective to that classification problem: selecting like data in each cluster (minimize the within clusters variance) and separating the clusters (maximize the between clusters variance). The above shows that the two are one and the same when the Euclidean metric is used to measure the distance between observations. Indeed, decreasing the within clusters variance is the same as increasing the between centroids variance, as the two sum to the constant σ_C^2.

B Discrete Individual Claim Development

We want to populate the tails of the LDF distribution so that the dispersion model contemplates a claim developing quite dramatically. Accordingly, we seek a finite set of probabilities

$$0 < p_1 < p_2 < \cdots < p_n < 1$$

that cover $(0, 1)$ with an emphasis on populating the right and left hand tails near 0 and 1. We are confronted with a practical working limit of no more than 200 points. We have also observed that 100 equally spaced points will result in the dispersion reflecting too confined a range, about 1/3 to 3-fold for the full range of dispersion. To cover a wider range, we use 171 non-uniform probabilities, and focus on the tails. Then treating the probabilities p_i as defining percentiles, we determine the corresponding percentile values u_i from a gamma distribution. That finite sequence $\{u_i\}$ of values is the starting point to capture a gamma density. But this representation is then refined, replacing the percentiles with the means over the 172 intervals $[0, u_1), [u_1, u_2), \ldots, [u_{170}, u_{171})$ and $[u_{171}, \infty)$. The new sequence of values, again denoted as $\{u_i\}$, is an optimized discrete approximation to a gamma. It is "weighted" in the sense that mean value u_i has associated with it the frequency weight v_i, where

$$v_1 = p_1, v_2 = p_2 - p_1, ..., v_{171} = p_{171} - p_{170}, v_{172} = 1 - p_{171}.$$

The interval width provides the weight assigned to the corresponding percentile value and is selected to be at most $\frac{1}{100}$ so that the usual "percentiles" are "covered." By definition, inverting and transforming those observations produces a discrete approximation to values from an inverse transformed gamma distribution. These are the candidates for the set of loss development factors used for dispersion. Parameters were selected so as to achieve a target mean LDF as well as a target CV for the LDFs. In order to assure the correct mean, one more observation is added, forcing the weighted mean of the sequence $\{u_i | 1 \leq i \leq 173\}$ to be exactly the appropriate open claim only LDF. There is the concern that if that final observation is allotted too little weight, it will have the potential for becoming an outlier. So the added observation has weight $\frac{1}{100}$ and the other weights are adjusted by a factor of $\frac{99}{100}$, making the 173 weights $\{v_i | 1 \leq i \leq 173\}$ again total to 1. From this construction, it is expected that the $\{u_i | 1 \leq i \leq 173\}$ will exhibit a slightly smaller variance than the theoretical inverse transformed gamma, and that is indeed observed to be the case in the calculations. For example, when targeting a CV of 0.500, the model yielded a CV of 0.495.

This discussion does not describe the (comparatively minor) adjustment for reopened claims. The reopened claim adjustment is achieved by first using the results of Appendix A to determine means and variances after reclassifying some closed claims as open, and then matching two moments using the power transform as detailed in Appendix C.2. In this way, the "truly open" claims are dispersed.

We now fill in the details of the algorithm used to build the dispersion model. The first step is to specify the α and τ parameters, by injury type and report, for the inverse transformed gamma. The parameters were selected from an analysis of LDF distributions as presented in section 2. The parameterizations follow that of the Appendix of Klugman et al. [9].

Recall that the α and τ parameters determine the CV and once they are set, the θ parameter dictates the mean.

The next step is to build a discrete approximation to a Gamma distribution with parameters α and θ=1. This is captured in two finite sequences, u and v. The u sequence captures the values while the v sequence stores the corresponding probability of occurrence "weights." We identify the "percentile" u-value of the distribution function associated with the following list of probabilities p_i, $1 \leq i \leq 171$:

$$\begin{aligned}
p_0 &= 0 \\
p_i &= p_{i-1} + 10^{-6} & 1 \leq i \leq 10 \\
p_i &= p_{i-1} + 10^{-5} & 11 \leq i \leq 19 \\
p_i &= p_{i-1} + 10^{-4} & 20 \leq i \leq 28 \\
p_i &= p_{i-1} + 10^{-3} & 29 \leq i \leq 37 \\
p_i &= p_{i-1} + 10^{-2} & 38 \leq i \leq 86 \\
p_{86+i} &= 1 - p_{86-i} & 1 \leq i \leq 85.
\end{aligned}$$

These probabilities were selected to give greater granularity to the right and left tails. This corresponds to 171 finite intervals: $[u_0 = 0, u_1), \ldots, [u_i, u_{i+1})$ for $0 \leq i \leq 170$ and the right hand tail interval $[u_{171}, \infty)$. We let $\Gamma(\alpha; u)$ denote the incomplete gamma function as formally defined in the Appendix of Klugman et al. [9], where that function is also noted to be the distribution function of a gamma distribution with parameters α and $\theta = 1$ (and for the transformed gamma with parameters α , $\theta = 1$, and $\tau = 1$). A binary search routine is used to associate the value u_i with the probability p_i, finding u_i that satisfies:

$$|\Gamma(\alpha; u_i) - p_i| < 0.0000000001 \quad 1 \leq i \leq 171.$$

The first difference of the p_i gives the frequency probability v_i of an observation falling within the interval $[u_{i-1}, u_i)$, i.e. between percentile p_{i-1} and

p_i. The mean value over each of the 172 intervals is readily determined from the observation that given $f(\alpha, 1; x) = \frac{x^\alpha e^{-x}}{x\Gamma(\alpha)}$, we get

$$xf(\alpha, 1; x) = \frac{x^{\alpha+1}e^{-x}}{x\Gamma(\alpha)} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \cdot \frac{x^{\alpha+1}e^{-x}}{x\Gamma(\alpha+1)} = \alpha\frac{x^{\alpha+1}e^{-x}}{x\Gamma(\alpha+1)} = \alpha f(\alpha+1, 1; x)$$

and thus

$$\int_0^z xf(\alpha, 1; x)dx = \int_0^z \alpha f(\alpha+1, 1; x)dx = \alpha\Gamma(\alpha+1; z).$$

This lets us specify the sequence u of length 172 whose components are the mean value of the inverse transformed gamma over the 172 intervals. The sequence v, also of length 172, with components equaling the corresponding frequency provides a sequence of weights to apply to the corresponding LDF values captured in the sequence u.

Denote the applicable development factor for open claims as $\widetilde{\lambda}$. The next step in building the dispersion model is to specify a sequence Λ of length 173 whose component values (properly weighted) are distributed as an inverse transformed gamma distribution of mean $\widetilde{\lambda}$ (and CV determined from the corresponding α and τ parameters). The formula for the expectation of an inverse transformed gamma random variable, X, allows us to calculate the θ parameter:

$$E[X] = \frac{\theta \cdot \Gamma(\alpha - \frac{1}{\tau})}{\Gamma(\alpha)} \Rightarrow \theta = \widetilde{\lambda}\left(\frac{\Gamma(\alpha)}{\Gamma(\alpha - \frac{1}{\tau})}\right).$$

The dispersion model uses the inverse transformed gamma as the LDF distribution. By definition, a distribution is inverse transformed gamma exactly if, when transformed and inverted, it conforms to a gamma distribution like that approximated by the sequences u and v of discrete values and weights, respectively. Following the parametrization of the Appendix of Klugman et al. [9], to make the finite sequence Λ contain values distributed as the inverse transform gamma, we just use the equivalence:

$$\left(\frac{\theta}{\Lambda_i}\right)^\tau = u_i \Leftrightarrow \frac{\Lambda_i}{\theta} = u_i^{-\frac{1}{\tau}} = \frac{1}{u_i^{\frac{1}{\tau}}} \Leftrightarrow \Lambda_i = \frac{\theta}{u_i^{\frac{1}{\tau}}}.$$

Since we are using a discrete approximation, and to assure we do get the correct expected developed loss, we augment the Λ sequence by an additional value in order to force the weighted mean to be $\widetilde{\lambda}$. More precisely we set $v_{173} = \frac{1}{100}$, rescale the other weights by setting $v_j = \frac{99}{100} v_j$ for $1 \le j \le 172$, and set $\Lambda_{173} = 100\left(\widetilde{\lambda} - \sum_{j=1}^{172} v_j \Lambda_j\right)$, which assures that:

$$\sum_{j=1}^{173} v_j = 1 \qquad \sum_{j=1}^{173} v_j \Lambda_j = \widetilde{\lambda}.$$

Having the Λ and v sequences in hand, completing the dispersion loss severity model is then very straightforward. Individual claim data are captured from WCSP data into observations that include state, injury type, claim status, a weight w, and a loss amount l, as described in section 2. Closed and open claims are separated into two subsets of observations, L_c and L_o respectively. Then for each open claim of weight w and undeveloped loss amount equal to l in L_o, 173 "dispersed" observations are captured into the data set $\widetilde{L_o}$ using the sequences Λ and v to assign the observations with weights equal to the product $w \times v_i$ and developed loss amounts equal to the product $\Lambda_i \times l$, $1 \le i \le 173$. Losses in $\widetilde{L_o}$ are adjusted to be at least \$1. Finally, forming the union $L_c \cup \widetilde{L_o}$ of two sets, each consisting of observations of individual claim data at closure, results in the dispersion model for ultimate claim severity.

C Data Adjustment Techniques

Let $x_{i1}, x_{i2}, \ldots, x_{in_i}$ be the incurred loss amounts on the claims (of a given injury type) in state i and let $\mu_i = \frac{1}{n_i} \sum_j x_{ij}$ be the sample mean. Under mean normalization we divide each claim amount by the state sample mean to get x_{ij}/μ_i. Pooling all the mean normalized claims for all states gives us a countrywide mean normalized database, $\{x_{ij}/\mu_i\}$. This database has mean 1 of course. If we fix a state k and multiply each mean normalized claim amount in the countrywide database by μ_k we get a database, $\{\mu_k x_{ij}/\mu_i\}$, that has mean μ_k. This database augments the claims in state k with out of state claims that have been adjusted to the state k level. We now generalize this simple idea to the case of standardization as well as the power transform.

C.1 Logarithmic Standardization

A natural way to generalize mean normalization would be to standardize claims, i.e. to subtract the state sample mean from every claim and divide by the standard deviation, $x_i \to \frac{x_i-\mu_i}{\sigma_i}$. Pooling all of the standardized claims would result in a countrywide standardized database with mean 0 and standard deviation 1. Then for a given state k, we might multiply each standardized claim by σ_k and and add μ_k to get a database, $\left\{\sigma_k \frac{x_i-\mu_i}{\sigma_i} + \mu_k\right\}$, appropriate for state k. Unfortunately, this can result in negative claim amounts so we prefer to work with logged losses and standardize them by mapping $\log x_i \to \frac{\log x_i-\mu_i}{\sigma_i}$, where now μ_i, σ_i denote the sample mean and standard deviation of the logged losses. This results in a standardized database of logged losses, $\left\{\frac{\log x_i-\mu_i}{\sigma_i}\right\}$. To get a database appropriate for a given state k it is natural to multiply each standardized logged loss by σ_k, add μ_k, and then exponentiate to get a database, $\left\{\exp(\sigma_k \frac{\log x_i-\mu_i}{\sigma_i} + \mu_k)\right\}$. The linear transformation, $\frac{\log x_i-\mu_i}{\sigma_i} \to \sigma_k \frac{\log x_i-\mu_i}{\sigma_i} + \mu_k$, results in a database that matches the mean and variance of the logged losses in state k, but upon exponentiation we lose this property. That is, the database, $\left\{\exp(\sigma_k \frac{\log x_i-\mu_i}{\sigma_i} + \mu_k)\right\}$, may not have the same mean and variance as the claims in state k. However, under reasonable conditions we can find μ, σ such that the database, $\left\{\exp(\sigma \frac{\log x_i-\mu_i}{\sigma_i} + \mu)\right\}$, will have the mean and variance in state k. We proceed now to establish this. We begin with a lemma.

Lemma 2 *Let $x_1, x_2, \ldots, x_n$ be a finite sequence of real numbers, not all equal, and let $\varphi : (0, \infty) \to \mathbb{R}$ by*

$$\varphi(t) = \frac{\left(\sum_{i=1}^{n} t^{x_i}\right)^2}{n \sum_{i=1}^{n} t^{2x_i}},$$

then

1. $\varphi(1) = 1$
2. φ is strictly increasing on $(0, 1)$ and strictly decreasing on $(1, \infty)$

3. $\lim_{t\to\infty}\varphi(t) = k/n$, where k is the number of i such that $x_i = \max\{x_j | 1 \le j \le n\}$.

Proof. We have $\varphi(1) = \frac{n^2}{n\cdot n} = 1$, thus proving item 1.

To prove item 2, first note that

$$\begin{aligned}\frac{d\varphi}{dt} &= \frac{\left(\sum_{i=1}^n t^{2x_i}\right)\left(2\sum_{i=1}^n t^{x_i}\right)\left(\sum_{i=1}^n x_i t^{x_i-1}\right) - \left(\sum_{i=1}^n t^{x_i}\right)^2\left(\sum_{i=1}^n 2x_i t^{2x_i-1}\right)}{n\left(\sum_{i=1}^n t^{2x_i}\right)^2} \\ &= \frac{2\sum_{i=1}^n t^{x_i}}{n\left(\sum_{i=1}^n t^{2x_i}\right)^2}\left[\left(\sum_{i=1}^n t^{2x_i}\right)\left(\sum_{i=1}^n x_i t^{x_i-1}\right) - \left(\sum_{i=1}^n t^{x_i}\right)\left(\sum_{i=1}^n x_i t^{2x_i-1}\right)\right]\end{aligned}$$

As the term $2\sum_{i=1}^n t^{x_i}/n\left(\sum_{i=1}^n t^{2x_i}\right)^2$ is positive, we note that, after relabelling indices for convenience, $\frac{d\varphi}{dt}$ has the same sign as

$$\begin{aligned}\gamma(t) &= \left(\sum_{i=1}^n t^{2x_i}\right)\left(\sum_{j=1}^n x_j t^{x_j-1}\right) - \left(\sum_{j=1}^n t^{x_j}\right)\left(\sum_{i=1}^n x_i t^{2x_i-1}\right) \\ &= \sum_{1\le i,j\le n} x_j t^{2x_i+x_j-1} - \sum_{1\le i,j\le n} x_i t^{2x_i+x_j-1} \\ &= \sum_{1\le i,j\le n} (x_j - x_i) t^{2x_i+x_j-1} \\ &= \sum_{1\le i<j\le n} (x_j - x_i) t^{2x_i+x_j-1} + \sum_{1\le j<i\le n} (x_j - x_i) t^{2x_i+x_j-1} \\ &= \sum_{1\le i<j\le n} (x_j - x_i) t^{x_i} t^{x_i+x_j-1} + \sum_{1\le j<i\le n} (x_j - x_i) t^{x_i} t^{x_i+x_j-1} \\ &= \sum_{1\le i<j\le n} (x_j - x_i) t^{x_i} t^{x_i+x_j-1} + \sum_{1\le i<j\le n} (x_i - x_j) t^{x_j} t^{x_i+x_j-1} \\ &= \sum_{1\le i<j\le n} (x_j - x_i) t^{x_i} t^{x_i+x_j-1} - \sum_{1\le i<j\le n} (x_j - x_i) t^{x_j} t^{x_i+x_j-1} \\ &= \sum_{1\le i<j\le n} (x_j - x_i)(t^{x_i} - t^{x_j}) t^{x_i+x_j-1}.\end{aligned}$$

Observe that for $t < 1$, the differences $x_j - x_i$ and $t^{x_i} - t^{x_j}$, not all of which are 0, have the same sign, which implies that $\gamma(t) > 0$. Similarly, for $t > 1$, those differences have opposite signs, hence $\gamma(t) < 0$, thus proving item 2.

To prove item 3, we sort and relabel the x_i as necessary so that $x_1 = \cdots = x_k = \max\{x_i\}$ and $x_i < x_1$ for $i > k$. We then find that

$$\varphi(t) = \frac{\left(\sum_{i=1}^{n} t^{x_i}\right)^2}{n\sum_{i=1}^{n} t^{2x_i}} = \frac{t^{2x_1}\left(\sum_{i=1}^{k} 1 + \sum_{i=k+1}^{n} t^{x_i - x_1}\right)^2}{nt^{2x_1}\left(\sum_{i=1}^{k} 1 + \sum_{i=k+1}^{n} t^{2(x_i - x_1)}\right)} = \frac{\left(k + \sum_{i=k+1}^{n} t^{x_i - x_1}\right)^2}{n\left(k + \sum_{i=k+1}^{n} t^{2(x_i - x_1)}\right)}.$$

Since $x_i - x_1 < 0$ for $i > k$ it follows that

$$\lim_{t\to\infty} \varphi(t) = \lim_{t\to\infty} \frac{\left(k + \sum_{i=k+1}^{n} t^{x_i - x_1}\right)^2}{n\left(k + \sum_{i=k+1}^{n} t^{2(x_i - x_1)}\right)} = \frac{\left(k + \sum_{i=k+1}^{n} 0\right)^2}{n\left(k + \sum_{i=k+1}^{n} 0\right)} = \frac{k^2}{nk} = \frac{k}{n}$$

as claimed. This completes the proof of item 3 and the lemma. ■

Now interpreting $x_1, x_2, \ldots, x_n$ to be the standardized logged losses this lemma allows us to prove the following proposition which shows that standardization of logged losses, followed by a linear transformation and re-exponentiation does what we want under reasonable conditions.

Proposition 3 *Let $x_1, x_2, \ldots, x_n$ be a finite sequence of real numbers, not all equal, and let k be the number of i such that $x_i = \max\{x_j | 1 \le j \le n\}$. Then for any pair of positive real numbers, μ, σ, such that $\mu^2/(\mu^2+\sigma^2) > k/n$, there exists a unique pair of real numbers, m, s, with $s > 0$ such that the finite sequence, $e^{m+sx_1}, e^{m+sx_2}, \ldots, e^{m+sx_n}$, has mean μ and standard deviation σ. More precisely, if $y_i = e^{m+sx_i}$, then*

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \textit{and} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2.$$

Proof. From the lemma, there exists a unique $t > 1$ with $\varphi(t) = \frac{\mu^2}{\mu^2+\sigma^2}$. Observe that since $\sum_{i=1}^{n} t^{x_i} > 0$ we can define

$$s = \ln t > 0 \quad \text{and} \quad m = \ln\left(\frac{n\mu}{\sum_{i=1}^{n} t^{x_i}}\right).$$

Then setting $y_i = e^{m+sx_i}$ for $1 \le i \le n$, we have

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{e^m}{n}\sum_{i=1}^{n}(e^s)^{x_i} = \frac{e^m}{n}\sum_{i=1}^{n} t^{x_i} = \left(\frac{n\mu}{n\sum_{i=1}^{n} t^{x_i}}\right)\sum_{i=1}^{n} t^{x_i} = \mu.$$

We also have

$$\begin{aligned}
\frac{n\mu^2}{\sum_{i=1}^{n} y_i^2} &= \frac{n\mu^2}{\sum_{i=1}^{n}(e^m t^{x_i})^2} = \frac{(n\mu)^2}{ne^{2m}\sum_{i=1}^{n} t^{2x_i}} \\
&= \frac{\left(e^m\sum_{i=1}^{n} t^{x_i}\right)^2}{ne^{2m}\sum_{i=1}^{n} t^{2x_i}} = \frac{\left(\sum_{i=1}^{n} t^{x_i}\right)^2}{n\sum_{i=1}^{n} t^{2x_i}} \\
&= \varphi(t) = \frac{\mu^2}{\mu^2+\sigma^2}
\end{aligned}$$

which implies $\frac{1}{n}\sum_{i=1}^{n} y_i^2 = \mu^2 + \sigma^2$ and thus

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i^2 - 2\mu y_i + \mu^2) \\
&= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - 2\mu\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) + \mu^2 \\
&= \mu^2 + \sigma^2 - 2\mu\mu + \mu^2 \\
&= \sigma^2.
\end{aligned}$$

To prove uniqueness, let $\hat{m}, \hat{s}$ be another such pair, and set $\hat{y}_i = e^{\hat{m}+\hat{s}x_i}$ for $1 \le i \le n$. It follows that

$$\frac{\mu^2}{\mu^2+\sigma^2} = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i^2} = \frac{\left(e^{\hat{m}}\sum_{i=1}^{n}(e^{\hat{s}})^{x_i}\right)^2}{ne^{2\hat{m}}\sum_{i=1}^{n}(e^{\hat{s}})^{2x_i}} = \frac{\left(\sum_{i=1}^{n}(e^{\hat{s}})^{x_i}\right)^2}{n\sum_{i=1}^{n}(e^{\hat{s}})^{2x_i}} = \varphi(e^{\hat{s}}).$$

Since $\hat{s} > 0$ implies $e^{\hat{s}} > 1$, it follows that $e^{\hat{s}} = t = e^s$ and thus $\hat{s} = s$. Finally, we have

$$\hat{y}_i = e^{\hat{m}+\hat{s}x_i} = e^{\hat{m}+sx_i} = e^{\hat{m}-m+m+sx_i} = e^{\hat{m}-m}e^{m+sx_i} = e^{\hat{m}-m}y_i$$

for $1 \le i \le n$, which implies $\mu = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \frac{e^{\hat{m}-m}}{n}\sum_{i=1}^{n} y_i = e^{\hat{m}-m}\mu$. Since $\mu \neq 0$, it follows that $\hat{m} = m$ and the proof is complete. ∎

It is possible to generalize the previous result from a finite sample, $x_1, x_2, \ldots, x_r$ to a distribution with finite support. The argument mirrors that for the discrete case. As before, we begin with a lemma.

Lemma 4 *Let f be a continuous probability density on the finite interval, $[a,b]$, and let $\varphi : (0,\infty) \to \mathbb{R}$ by*

$$\varphi(t) = \frac{\left(\int_a^b t^x f(x)dx\right)^2}{\int_a^b t^{2x} f(x)dx},$$

then

1. $\varphi(1) = 1$
2. φ *is strictly increasing on* $(0,1)$ *and strictly decreasing on* $(1,\infty)$
3. $\lim_{t\to\infty} \varphi(t) = 0$.

Proof. We have $\varphi(1) = \frac{1^2}{1} = 1$, thus proving item 1.

To prove item 2, first note that

$$\begin{aligned}
\frac{d\varphi}{dt} &= \left[\left(\int_a^b t^{2x} f(x)dx\right)\left(2\int_a^b t^x f(x)dx\right)\left(\frac{d}{dt}\int_a^b t^x f(x)dx\right)\right. \\
&\quad \left. - \left(\int_a^b t^x f(x)dx\right)^2 \left(\frac{d}{dt}\int_a^b t^{2x} f(x)dx\right)\right] / \left(\int_a^b t^{2x} f(x)dx\right)^2 \\
&= \left[\left(\int_a^b t^{2x} f(x)dx\right)\left(2\int_a^b t^x f(x)dx\right)\left(\int_a^b x t^{x-1} f(x)dx\right)\right. \\
&\quad \left. - \left(\int_a^b t^x f(x)dx\right)^2 \left(\int_a^b 2x t^{2x-1} f(x)dx\right)\right] / \left(\int_a^b t^{2x} f(x)dx\right)^2 \\
&= \frac{2\int_a^b t^x f(x)dx}{\left(\int_a^b t^{2x} f(x)dx\right)^2}\left[\left(\int_a^b t^{2x} f(x)dx\right)\left(\int_a^b x t^{x-1} f(x)dx\right)\right. \\
&\quad \left. - \left(\int_a^b t^x f(x)dx\right)\left(\int_a^b x t^{2x-1} f(x)dx\right)\right]
\end{aligned}$$

As the term $2\int_a^b t^x f(x)dx / \left(\int_a^b t^{2x} f(x)dx\right)^2$ is positive, we note that $\frac{d\varphi}{dt}$ has the same sign as

$$\gamma(t) = \left(\int_a^b t^{2x} f(x)dx\right)\left(\int_a^b x t^{x-1} f(x)dx\right) - \left(\int_a^b t^x f(x)dx\right)\left(\int_a^b x t^{2x-1} f(x)dx\right).$$

Relabelling dummy variables for convenience, we get

$$\begin{aligned}
\gamma(t) &= \left(\int_a^b t^{2x} f(x)dx\right)\left(\int_a^b y t^{y-1} f(y)dy\right) - \left(\int_a^b t^{y} f(y)dy\right)\left(\int_a^b x t^{2x-1} f(x)dx\right) \\
&= \int_a^b \int_a^b y t^{2x+y-1} f(x)f(y)dxdy - \int_a^b \int_a^b x t^{2x+y-1} f(x)f(y)dxdy \\
&= \int_a^b \int_a^b (y-x) t^{2x+y-1} f(x)f(y)dxdy \\
&= \int_a^b \int_a^y (y-x) t^{2x+y-1} f(x)f(y)dxdy + \int_a^b \int_a^x (y-x) t^{2x+y-1} f(x)f(y)dydx \\
&= \int_a^b \int_a^y (y-x) t^{2x+y-1} f(x)f(y)dxdy - \int_a^b \int_a^y (y-x) t^{2y+x-1} f(x)f(y)dxdy \\
&= \int_a^b \int_a^y (y-x)(t^{2x+y-1} - t^{2y+x-1}) f(x)f(y)dxdy \\
&= \int_a^b \int_a^y (y-x)(t^{x} - t^{y}) t^{x+y-1} f(x)f(y)dxdy.
\end{aligned}$$

Observe that for $t < 1$, the differences $y - x$ and $t^x - t^y$ have the same sign, which implies that $\gamma(t) > 0$. Similarly, for $t > 1$, those differences have opposite signs, hence $\gamma(t) < 0$, thus proving item 2.

To prove item 3, first consider the case when $f(x) > 0$ for all $x \in [a, b]$. Since f is continuous on $[a, b]$, it is uniformly continuous on $[a, b]$. Thus, for any $\epsilon > 0$, there is a partition

$$[a,b] = \bigcup_{i=1}^{n} [a_i, b_i] \text{ with } a = a_1, a_i < b_i = a_{i+1}, b_n = b$$

such that

$$x_1, x_2 \in [a_i, b_i] \Longrightarrow |f(x_1) - f(x_2)| \leq \epsilon.$$

Let $\alpha = \min\{f(x)|x \in [a,b]\} > 0$ and let $\alpha_i = \min\{f(x)|x \in [a_i, b_i]\}$, then $\{f(x)|x \in [a_i, b_i]\} \subseteq [\alpha_i, \alpha_i + \epsilon]$. We claim that

$$\lim_{t\to\infty} \frac{\int_{a_i}^{b_i} t^x dx \int_{a_j}^{b_j} t^x dx}{\int_a^b t^{2x} dx} = 0.$$

To see this assume that $b_i \le a_j$, then

$$\begin{aligned}
\frac{\int_{a_i}^{b_i} t^x dx \int_{a_j}^{b_j} t^x dx}{\int_a^b t^{2x} dx} &= \frac{\frac{t^x}{\ln t}\Big|_{a_i}^{b_i} \frac{t^x}{\ln t}\Big|_{a_j}^{b_j}}{\frac{t^{2x}}{2\ln t}\Big|_a^b} \\
&= \frac{2}{\ln t}\left(\frac{\left(t^{b_i} - t^{a_i}\right)\left(t^{b_j} - t^{a_j}\right)}{\left(t^{2b} - t^{2a}\right)}\right) \\
&= \frac{2}{\ln t}\left(\frac{\left(t^{b_i - b_j} - t^{a_i - b_j}\right)\left(1 - t^{a_j - b_j}\right)}{\left(t^{2(b-b_j)} - t^{2(a-b_j)}\right)}\right).
\end{aligned}$$

Thus

$$\lim_{t\to\infty} \frac{\int_{a_i}^{b_i} t^x dx \int_{a_j}^{b_j} t^x dx}{\int_a^b t^{2x} dx} = \lim_{t\to\infty} \frac{2}{\ln t}\left(\frac{(0-0)(1-0)}{\left(t^{2(b-b_j)} - 0\right)}\right) = 0$$

as claimed.

From what we've just claimed, for every i and j, there exists a $t_{i,j}$ such that for all $t > t_{i,j}$ we have

$$\frac{\int_{a_i}^{b_i} t^x dx \int_{a_j}^{b_j} t^x dx}{\int_a^b t^{2x} dx} \le \frac{\alpha\epsilon}{n^2(\alpha_i + \epsilon)(\alpha_j + \epsilon)}.$$

Then for all $t > t_{i,j}$ we have

$$\begin{aligned}
\frac{\int_{a_i}^{b_i} t^x f(x) dx \int_{a_j}^{b_j} t^x f(x) dx}{\int_a^b t^{2x} f(x) dx} &\le \frac{(\alpha_i + \epsilon)(\alpha_j + \epsilon)\int_{a_i}^{b_i} t^x dx \int_{a_j}^{b_j} t^x dx}{\alpha \int_a^b t^{2x} dx} \\
&\le \frac{(\alpha_i + \epsilon)(\alpha_j + \epsilon)}{\alpha} \frac{\alpha\epsilon}{n^2(\alpha_i + \epsilon)(\alpha_j + \epsilon)} = \frac{\epsilon}{n^2}.
\end{aligned}$$

Thus for $t > \max\{t_{i,j}\}$, it follows that

$$\begin{aligned}
\varphi(t) &= \frac{\left(\int_a^b t^x f(x) dx\right)^2}{\int_a^b t^{2x} f(x) dx} = \frac{\left(\sum_{i=1}^n \int_{a_i}^{b_i} t^x f(x) dx\right)^2}{\int_a^b t^{2x} f(x) dx} \\
&= \sum_{1 \le i,j \le n} \frac{\int_{a_i}^{b_i} t^x f(x) dx \int_{a_j}^{b_j} t^x f(x) dx}{\int_a^b t^{2x} f(x) dx} \le \sum_{1 \le i,j \le n} \frac{\epsilon}{n^2} = \epsilon.
\end{aligned}$$

Thus, $\lim_{t\to\infty}\varphi(t) = 0$ in the case when $f(x) > 0$ for all $x \in [a,b]$. Finally, if we set $g(x) = \frac{f(x)+1}{b-a+1}$ then g is a positive, continuous probability density function on $[a,b]$ and we have

$$
\begin{aligned}
0 &= (b-a+1)\lim_{t\to\infty}\frac{\left(\int_a^b t^x g(x)dx\right)^2}{\int_a^b t^{2x} g(x)dx} \\
&= (b-a+1)\lim_{t\to\infty}\frac{\left(\int_a^b t^x (\frac{f(x)+1}{b-a+1})dx\right)^2}{\int_a^b t^{2x} (\frac{f(x)+1}{b-a+1})dx} \\
&= \lim_{t\to\infty}\frac{\left(\int_a^b t^x (f(x)+1)dx\right)^2}{\int_a^b t^{2x}(f(x)+1)dx} = \lim_{t\to\infty}\frac{\left(\int_a^b t^x f(x)dx + \int_a^b t^x dx\right)^2}{\int_a^b t^{2x} f(x)dx + \int_a^b t^{2x}dx} \\
&= \lim_{t\to\infty}\frac{\left(\int_a^b t^x f(x)dx + \left(\frac{t^b - t^a}{\ln t}\right)\right)^2}{\int_a^b t^{2x} f(x)dx + \left(\frac{t^{2b}-t^{2a}}{\ln t}\right)} \\
&= \lim_{t\to\infty}\frac{\left(\int_a^b t^{x-b} f(x)dx + \left(\frac{1-t^{(a-b)}}{\ln t}\right)\right)^2}{\int_a^b t^{2(x-b)} f(x)dx + \left(\frac{1-t^{2(a-b)}}{\ln t}\right)} \\
&= \lim_{t\to\infty}\frac{\left(\int_a^b t^{x-b} f(x)dx\right)^2}{\int_a^b t^{2(x-b)} f(x)dx} = \lim_{t\to\infty}\frac{\left(\int_a^b t^x f(x)dx\right)^2}{\int_a^b t^{2x} f(x)dx} \\
&= \lim_{t\to\infty}\varphi(t).
\end{aligned}
$$

This completes the proof of the lemma. ∎

Now interpreting f to be the density of the standardized logged losses the lemma allows us to prove the analog of Proposition 3 in the continuous case, namely that there is a linear transformation of the standardized logged losses such that after re-exponentiation we get the desired mean and variance.

Proposition 5 *Let f be a continuous probability density on the finite interval, $[a,b]$. Then for any pair of positive real numbers, μ,σ, there exists a unique pair of real numbers, m,s, with $s>0$ such that*

$$g(y)=\frac{1}{sy}f\left(\frac{\ln y-m}{s}\right)$$

is a continuous probability density on $[e^{m+as},e^{m+bs}]$ with mean μ and standard deviation σ.

Proof. From the lemma, there exists a unique $t>1$ with $\varphi(t)=\frac{\mu^2}{\mu^2+\sigma^2}$. Observe that $\varphi(t)>0$ implies $\int_a^b t^x f(x)dx>0$, thus we can define

$$s=\ln t>0 \quad\text{and}\quad m=\ln\left(\frac{\mu}{\int_a^b t^x f(x)dx}\right).$$

Let $c=e^{m+as}$ and $d=e^{m+bs}$. We also introduce the change of variable $x=\frac{\ln y-m}{s}\Leftrightarrow y=e^{m+sx}$, hence $\frac{dy}{dx}=ys$, which implies $dy=ysdx$. Then

$$\int_c^d g(y)dy=\int_a^b\frac{1}{sy}f(x)ysdx=\int_a^b f(x)dx=1.$$

Further, we have

$$\begin{aligned}\int_c^d yg(y)dy &= \int_a^b y\frac{1}{sy}f(x)ysdx=\int_a^b yf(x)dx\\ &= \int_a^b e^{m+sx}f(x)dx=e^m\int_a^b(e^s)^x f(x)dx\\ &= e^m\int_a^b t^x f(x)dx=\left(\frac{\mu}{\int_a^b t^x f(x)dx}\right)\int_a^b t^x f(x)dx=\mu.\end{aligned}$$

Since f is continuous, g is continuous as well and we have shown that g is a continuous probability density function on $[c,d]=[e^{m+as},e^{m+bs}]$ with mean

μ. As in the discrete case, we note that

$$\begin{aligned}
\frac{\mu^2}{\int_c^d y^2 g(y)dy} &= \frac{\left(\int_c^d yg(y)dy\right)^2}{\int_a^b (e^{m+sx})^2 f(x)dx} = \frac{\left(\int_a^b e^{m+sx} f(x)dx\right)^2}{\int_a^b (e^m t^x)^2 f(x)dx} \\
&= \frac{\left(e^m \int_a^b t^x f(x)dx\right)^2}{e^{2m} \int_a^b t^{2x} f(x)dx} = \frac{\left(\int_a^b t^x f(x)dx\right)^2}{\int_a^b t^{2x} f(x)dx} \\
&= \varphi(t) = \frac{\mu^2}{\mu^2 + \sigma^2},
\end{aligned}$$

which implies $\int_c^d y^2 g(y)dy = \mu^2 + \sigma^2$. Thus

$$\begin{aligned}
\int_c^d (y-\mu)^2 g(y)dy &= \int_c^d y^2 g(y)dy - 2\mu \int_c^d yg(y)dy + \mu^2 \int_c^d g(y)dy \\
&= \mu^2 + \sigma^2 - 2\mu^2 + \mu^2 \\
&= \sigma^2.
\end{aligned}$$

To prove uniqueness, let $\hat{m}, \hat{s}$ be another such pair, and let

$$\hat{g}(y) = \frac{1}{\hat{s}y} f\left(\frac{\ln y - \hat{m}}{\hat{s}}\right) \text{ for } y \in \left[\hat{c}, \hat{d}\right] = \left[e^{\hat{m}+a\hat{s}}, e^{\hat{m}+b\hat{s}}\right].$$

From a similar change of variable as above, it follows that

$$\begin{aligned}
\frac{\mu^2}{\mu^2+\sigma^2} &= \frac{\left(\int_{\hat{c}}^{\hat{d}} y\hat{g}(y)dy\right)^2}{\int_{\hat{c}}^{\hat{d}} y^2 \hat{g}(y)dy} = \frac{\left(\int_a^b e^{\hat{m}+\hat{s}x} f(x)dx\right)^2}{\int_a^b e^{2(\hat{m}+\hat{s}x)} f(x)dx} \\
&= \frac{\left(e^{\hat{m}} \int_a^b e^{\hat{s}x} f(x)dx\right)^2}{e^{2\hat{m}} \int_a^b e^{2\hat{s}x} f(x)dx} = \frac{\left(\int_a^b \left(e^{\hat{s}}\right)^x f(x)dx\right)^2}{\int_a^b \left(e^{\hat{s}}\right)^{2x} f(x)dx} = \varphi(e^{\hat{s}}).
\end{aligned}$$

Since $\hat{s} > 0$, it follows that $e^{\hat{s}} > 1$, implying that $e^{\hat{s}} = t = e^s$ and thus $\hat{s} = s$. Finally, we have

$$\mu = \int_a^b e^{\hat{m}+\hat{s}x} f(x)dx = \int_a^b e^{\hat{m}+sx} f(x)dx = e^{\hat{m}-m} \int_a^b e^{m+sx} f(x)dx = e^{\hat{m}-m}\mu.$$

Since $\mu > 0$, it follows that $\hat{m} = m$, and the proof is complete. ∎

Remark 6 *What holds for the continuous case with infinite support is not so straightforward. For example, letting* $f(x) = \frac{e^{-\frac{x}{\theta}}}{\theta}$ *be the exponential density on* $[0, \infty)$, *the interested reader can readily verify that the condition* $0 < \frac{\sigma}{\mu} < \sqrt{1+\sqrt{2}}$ *is both necessary and sufficient for the existence of positive numbers* m *and* s *as in the proposition. More generally, if* $f(x)$ *is a probability density on* $[a, \infty)$ *with moment generating function* $M_X(t) = \int_a^\infty e^{tx} f(x)dx$ *and we are given a target mean* μ *and standard deviation* σ, *then one suggestion is to first try to determine* s *by solving the following implicit equation for the target coefficient of variation:*

$$\frac{\sigma}{\mu} = \frac{\sqrt{M_X(2s) - M_X(s)^2}}{M_X(s)}$$

and, if successful, determine m *from:*

$$m = \ln\left(\frac{\mu}{M_X(s)}\right).$$

C.2 The Power Transform

A more subtle way to transform claims is with a power transform, $x \longrightarrow ax^b$. With $a = 1/\mu$ and $b = 1$ we can see that the power transform generalizes mean normalization. With logarithmic standardization we first log the data, then standardize, and then re-exponentiate: $x \longrightarrow \log x \longrightarrow \frac{\log x - \mu}{\sigma} \longrightarrow \exp(\frac{\log x - \mu}{\sigma})$. But $\exp(\frac{\log x - \mu}{\sigma}) = e^{-\mu/\sigma}x^{1/\sigma}$ and so the power transform generalizes logarithmic standardization as well. Thus the power transform could potentially outperform both mean normalization and logarithmic standardization. In addition, with the power transform there is no need to log the losses and then re-exponentiate. The moments are matched in dollar space rather than in log space. The idea is to choose a and b so that the transformed losses from one state match the mean and variance of the losses from another state. In this way we can use the out of state losses to build an expanded database for each state. We now prove, under reasonable conditions, that it is possible to choose a and b in the power transform so that the transformed losses from one state do indeed match the mean and variance of another state.

Proposition 7 *Let $x_1, x_2, \ldots, x_n$ be a finite sequence of positive real numbers, not all equal and and let k be the number of i such that $x_i = \max\{x_j | 1 \leq j \leq n\}$. Then given $\mu > 0$ and $\gamma \in \left[0, \sqrt{\frac{n-k}{k}}\right)$, there exist unique constants $a > 0$ and $b \geq 0$ such that the database $\{ax_i^b\}$ has mean μ and CV γ.*

Proof. If $\gamma = 0$ then we must take $b = 0$ and $a = \mu$, and the result holds. So assume $\gamma > 0$. Let $\sigma = \gamma\mu > 0$ and set

$$z_i = \ln x_i \text{ for } 1 \leq j \leq n.$$

Then clearly k is the number of i such that $z_i = \max\{z_j | 1 \leq j \leq n\}$. We have:

$$\begin{aligned}
\gamma &\in \left(0, \sqrt{\frac{n-k}{k}}\right) \\
&\Leftrightarrow \frac{\sigma^2}{\mu^2} < \frac{n}{k} - 1 \\
&\Leftrightarrow \frac{\mu^2 + \sigma^2}{\mu^2} < \frac{n}{k} \\
&\Leftrightarrow \mu^2/(\mu^2 + \sigma^2) > k/n
\end{aligned}$$

and so by Proposition 3 there is a unique pair of real numbers m, s with $s > 0$ such that the finite sequence, $e^{m+sz_1}, e^{m+sz_2}, \ldots, e^{m+sz_n}$, has mean μ and standard deviation σ. Letting $a = e^m$ and $b = s$ we have:

$$e^{m+sz_i} = e^m \left(e^{z_i}\right)^s = e^m \left(e^{\ln x_i}\right)^s = ax_i^b \text{ for } 1 \leq j \leq n$$

and the existence of the constants a and b is proved. Uniqueness of $a = e^m$ and $b = s$ follows from the uniqueness of m and s, and the proof is complete.

■

D Excess Ratio Functions

We collect here some facts about excess ratio functions. We show how to recover the distribution function from the excess ratio function, give a characterization of excess ratio functions, and discuss the mixed exponential case. We start with some basic definitions and results.

Definition 8 *A random variable X is a loss variable if it is nonnegative valued, has finite nonzero mean, and has a density f that is continuous when restricted to $[0, +\infty)$. We denote by F the distribution function of X. The survival function of X is $S = 1 - F$. The excess ratio function of X is given by $R(r) = \int_r^\infty (x - r) f(x) dx / E[X]$ for $r \geq 0$. We denote by $\widehat{F}$ the function given by $\widehat{F}(r) = \int_0^r x f(x) dx / E[X]$. We use subscripts on F, $\widehat{F}$, S, and R when necessary to indicate dependence on X.*

The following proposition expresses the excess ratio function in terms of F and $\widehat{F}$.

Proposition 9 *Let X be a loss variable with mean μ, then*

$$R(r) = 1 - \widehat{F}(r) - \frac{r}{\mu}\left[1 - F(r)\right].$$

Proof. From the definition of $R(r)$ we have

$$\begin{aligned}
R(r) &= \frac{1}{\mu}\int_r^\infty (x - r) f(x) dx \\
&= \frac{1}{\mu}\left[\int_r^\infty x f(x) dx - r\int_r^\infty f(x) dx\right] \\
&= \frac{1}{\mu}\left[\mu - \int_0^r x f(x) dx - rS(r)\right] \\
&= 1 - \frac{1}{\mu}\int_0^r x f(x) dx - \frac{r}{\mu}S(r) \\
&= 1 - \widehat{F}(r) - \frac{r}{\mu}[1 - F(r)].
\end{aligned}$$

■

It is well known (see, for example, Billingsley [1], page 282) that the mean of a nonnegative random variable, X, can be expressed in terms of its survival function as $E[X] = \int_0^\infty S(x) dx$. It is easy to see that a similar result also holds for excess ratio functions.

Proposition 10 *Let X be a loss variable with survival function S and excess ratio function R, then*

$$R(r) = \frac{\int_r^\infty S(x) dx}{\int_0^\infty S(x) dx}.$$

Proof. Let X have density f, then noting that $S'(x) = -f(x)$ and using integration by parts, we have

$$\begin{aligned}
\int_r^\infty S(x)dx &= xS(x)|_r^\infty + \int_r^\infty xf(x)dx \\
&= -rS(r) + \int_r^\infty xf(x)dx \\
&= -r\int_r^\infty f(x)dx + \int_r^\infty xf(x)dx \\
&= \int_r^\infty (x-r)f(x)dx,
\end{aligned}$$

where the second equality follows as $xS(x) = x\int_x^\infty f(y)dy \leq \int_x^\infty yf(y)dy \to 0$ as $x \to \infty$ since X has finite mean. Thus $R(r) = \int_r^\infty (x-r)f(x)dx/E[X] = \int_r^\infty S(x)dx/\int_0^\infty S(x)dx$. ■

Survival functions and excess ratio functions share several elementary properties given in the next proposition.

Proposition 11 *If g is a survival function of a loss variable or an excess ratio function then*

1. $g(0) = 1$ *(and $g(x) = 1$ for $x < 0$ if g is a survival function)*
2. *g is non increasing*
3. $\lim_{x\to\infty} g(x) = 0$

The following proposition shows how to recover the distribution function from the excess ratio function. Thus the excess ratio function characterizes a loss distribution and so there is no loss of information in considering excess ratio functions rather than densities or distribution functions.

Proposition 12 *Let X be a loss variable with survival function S, and excess ratio function R, then $\frac{d}{dr}R(r) = -S(r)/E[X]$. Further, if we set $g(x) = S(x)/E[X]$ then $R(r) = \int_r^\infty g(x)dx$.*

Proof. For the first assertion, we have

$$\begin{aligned}\frac{d}{dr}R(r) &= \frac{1}{E(X)}\frac{d}{dr}\int_r^\infty (x-r)f(x)dx \\ &= \frac{1}{E(X)}\left[\frac{d}{dr}\int_r^\infty xf(x)dx - \frac{d}{dr}r\int_r^\infty f(x)dx\right].\end{aligned}$$

By the Fundamental Theorem of Calculus, we have $\frac{d}{dr}\int_r^\infty xf(x)dx = -rf(r)$ and $\frac{d}{dr}\int_r^\infty f(x) = -f(r)$, thus

$$\frac{d}{dr}R(r) = \frac{1}{E(X)}\left[-rf(r) + rf(r) - \int_r^\infty f(x)dx\right] = \frac{-S(r)}{E(X)}$$

For the second assertion, by Proposition 10 we have

$$R(r) = \int_r^\infty S(x)dx \Big/ \int_0^\infty S(x)dx = \int_r^\infty \frac{S(x)}{E(X)}dx = \int_r^\infty g(x)dx$$

■

This proposition also shows that the excess ratio function of a loss variable X is also the survival function of another random variable with density $S(x)/E[X]$. We next give characterizations of survival functions and excess ratio functions.

Proposition 13 *Let $g : [0, +\infty) \to \mathbb{R}$ be differentiable with g' continuous, $g(0) = 1$, and $\lim_{x\to\infty} g(x) = 0$, and let*

$$\widetilde{g}(x) = \begin{cases} 1 & \text{if } x < 0 \\ g(x) & \text{if } x \geq 0 \end{cases},$$

then $\widetilde{g}$ is the survival function of some nonnegative random variable X with density, f, that is continuous when restricted to $[0, +\infty)$ if and only if $g' \leq 0$.

Proof. Suppose $\widetilde{g} = S_X$ for some nonnegative random variable X with density, f, that is continuous when restricted to $[0, +\infty)$. Then for $x \geq 0$

$$g(x) = \widetilde{g}(x) = S_X(x) = \int_x^\infty f(y)dy,$$

and so by the Fundamental Theorem of Calculus $g'(x) = -f(x) \leq 0$.

Conversely, suppose $g' \leq 0$ and define

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ -g'(x) & \text{if } x \geq 0 \end{cases},$$

then f restricted to $[0, +\infty)$ is continuous and

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} -g'(x)dx = -g(x)|_0^\infty = -\lim_{x\to\infty} g(x) + g(0) = -0 + 1 = 1,$$

so f is a probability density function of some nonnegative random variable X. For $x < 0$ we have $S_X(x) = 1 = \widetilde{g}(x)$ and for $x \geq 0$

$$S_X(x) = \int_{x}^{\infty} f(y)dy = \int_{x}^{\infty} -g'(y)dy = -g(y)|_x^\infty = -\lim_{y\to\infty} g(y) + g(x) = g(x) = \widetilde{g}(x).$$

Thus $\widetilde{g} = S_X$. ■

Proposition 14 *Let $g : [0, +\infty) \to \mathbb{R}$ be twice differentiable with g'' continuous, $g(0) = 1$, and $\lim_{x\to\infty} g(x) = 0$, then g is the excess ratio function of some loss variable if and only if $g' \leq 0$ and $g'' \geq 0$.*

Proof. Suppose $g = R_X$ for some loss variable X with density f, survival function S, and mean μ. Then by Proposition 12,

$$g' = -S/\mu \leq 0 \text{ and } g'' = -S'/\mu = f/\mu \geq 0.$$

Conversely, suppose $g' \leq 0$ and $g'' \geq 0$. Since $g'' \geq 0$ we know that g' is non decreasing. So if $g'(0) = 0$ then $g'(x) = 0$ for all x as $g' \leq 0$. This would imply that g is constant and so $g(x) = g(0) = 1$ for all x, but this contradicts our hypothesis that $\lim_{x\to\infty} g(x) = 0$. Thus we must have $g'(0) < 0$. Observe also that

$$\int_0^\infty |g'(x)|\, dx = -\int_0^\infty g'(x)dx = -g(x)|_0^\infty = -0 + g(0) = 1,$$

and so $\lim_{x\to\infty} g'(x) = 0$. If we let

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ -\frac{1}{g'(0)} g''(x) & \text{if } x \geq 0 \end{cases}$$

then $f \geq 0$ and f is continuous when restricted to $[0, +\infty)$. Further

$$\int_0^\infty f(x)dx = \left(-\frac{1}{g'(0)}\right) g'(x)|_0^\infty = \left(-\frac{1}{g'(0)}\right) \lim_{x\to\infty} g'(x) + \frac{g'(0)}{g'(0)} = -0+1 = 1.$$

Thus f is a density function for some nonnegative random variable X. Since for $t \geq 0$

$$S_X(t) = \int_t^\infty f(x)dx = \left(-\frac{1}{g'(0)}\right) g'(x)|_t^\infty = \left(-\frac{1}{g'(0)}\right) \lim_{x\to\infty} g'(x) + \frac{g'(t)}{g'(0)} = \frac{g'(t)}{g'(0)},$$

it follows that

$$\begin{aligned} E(X) &= \int_0^\infty S_X(t)dt = \int_0^\infty \frac{g'(t)}{g'(0)}dt = \left(\frac{1}{g'(0)}\right) g(x)|_0^\infty \\ &= \left(\frac{1}{g'(0)}\right) \lim_{x\to\infty} g(x) - \frac{g(0)}{g'(0)} = 0 - \frac{1}{g'(0)} = -\frac{1}{g'(0)}. \end{aligned}$$

Thus, $0 < E(X) = -\frac{1}{g'(0)} < \infty$ and so X is a loss variable. Finally, by Proposition 10 we have

$$R_X(t) = \frac{\int_t^\infty S_X(x)dx}{\int_0^\infty S_X(x)dx} = \frac{\int_t^\infty g'(x)dx}{\int_0^\infty g'(x)dx} = \frac{g(x)|_t^\infty}{g(x)|_0^\infty} = \frac{0 - g(t)}{0 - g(0)} = g(t).$$

■

We can now characterize excess ratio functions in terms of survival functions.

Proposition 15 *Excess ratio functions are exactly the restrictions to $[0, +\infty)$ of survival functions of nonnegative random variables with densities that when restricted to $[0, +\infty)$ have nonpositive, continuous derivatives.*

Proof. Let $g = R_X$ be an excess ratio function of a loss variable X. Then by Proposition 14, $g' \leq 0$ and $g'' \geq 0$. Proposition 13 then implies there is a nonnegative random variable Y such that

$$S_Y(x) = \begin{cases} 1 & \text{if } x < 0 \\ g(x) & \text{if } x \geq 0 \end{cases},$$

and Y has a density function, f, that is continuous when restricted to $[0, +\infty)$. For $x \geq 0$ we have $g(x) = \int_x^\infty f(y)dy$ and so $g'(x) = -f(x)$, which implies that $f' = -g'' \leq 0$ and f' is continuous as g'' is continuous.

Conversely, let X be a nonnegative random variable with a density function, f, that when restricted to $[0, +\infty)$ has a continuous derivative and $f' \leq 0$. Let $g : [0, +\infty) \to \mathbb{R}$ by

$$g(x) = \int_x^\infty f(y)dy.$$

Then $g' = -f \leq 0$, which implies that $g'' = -f' \geq 0$. Then by Proposition 14, g is the excess ratio function of some loss variable. ∎

In the exponential case things are particularly simple as the next proposition shows.

Proposition 16 *Let $f(x) = \frac{1}{m}e^{-x/m}$ be an exponential density, then $R(x) = S(x) = e^{-x/m}$. That is, for an exponential distribution the excess ratio function is the same as the survival function.*

Proof. This follows directly from applying Definition 8 and using integration by parts. ∎

For finite mixtures we have the following proposition.

Proposition 17 *Let $f_1, f_2, \ldots, f_n$ be densities with corresponding excess ratio functions $R_1, R_2, \ldots, R_n$ and means $\mu_1, \mu_2, \ldots, \mu_n$. Then given weights $w_i \in (0, 1)$ with $\sum w_i = 1$, the mixed density $f = w_1 f_1 + w_2 f_2 + \cdots + w_n f_n$ has excess ratio function*

$$R = \widehat{w}_1 R_1 + \widehat{w}_2 R_2 + \cdots + \widehat{w}_n R_n,$$

where $\widehat{w}_i = w_i \mu_i / \mu$ and μ is the mean of the mixed distribution.

Proof. From the definition of the excess ratio function, we have

$$\begin{aligned}
R(r) &= \frac{1}{\mu}\int_r^\infty (x - r)f(x)dx \\
&= \frac{1}{\mu}\int_r^\infty (x - r)\left[\sum\nolimits_{i=1}^n w_i f_i(x)\right]dx \\
&= \sum\nolimits_{i=1}^n \frac{w_i}{\mu}\int_r^\infty (x - r)f_i(x)dx \\
&= \sum\nolimits_{i=1}^n \left(\frac{w_i \mu_i}{\mu}\right)\frac{1}{\mu_i}\int_r^\infty (x - r)f_i(x)dx \\
&= \sum\nolimits_{i=1}^n \widehat{w}_i R_i(r).
\end{aligned}$$

∎

Corollary 18 *If* $f(x) = \sum_{i=1}^{n} w_i \frac{1}{m_i} e^{-x/m_i}$ *is a finite mixed exponential density, then its excess ratio function is given by*

$$R(x) = \frac{\sum w_i m_i e^{-x/m_i}}{\sum w_i m_i}.$$

E Splicing Loss Distributions

We start with a loss variable, X (see Definition 8). The interpretation is that this represents the empirical losses. We then choose a point $l > 0$, such that $\Pr(X > l) > 0$ and $\Pr(X = l) = 0$. The point l, is called the splice point because we want to rely on X for claims less than l, but we want to splice on a distribution for claims larger than l. We let $Y = X - l$ conditional on $X \geq l$. That is, we truncate and shift X. More formally, if $X : \Omega \to [0, +\infty)$ then let $\Omega_0 = \{\omega \in \Omega | X(\omega) \geq l\}$ and define $Y : \Omega_0 \to [0, +\infty)$ by $Y(\omega) = X(\omega) - l$. The following proposition expresses the survival function, the density, and expected value of Y in terms of X.

Proposition 19 *Let* X *be a loss variable and let* l *be the splice point as above, then*

1. $S_Y(r - l) = \frac{1 - F_X(r)}{1 - F_X(l)}$ for $r \geq l$,

2. $f_Y(r - l) = \frac{f_X(r)}{1 - F_X(l)}$ for $r \geq l$, and

3. $E[Y] = \frac{E[X] R_X(l)}{1 - F_X(l)}$.

Proof. To prove item 1 we note first that $\Pr(X \geq l) = \Pr(X > l)$, then for $r \geq l$ we have

$$\begin{aligned} S_Y(r - l) &= \Pr(Y > r - l) = \Pr(X - l > r - l | X \geq l) \\ &= \Pr(X > r | X \geq l) = \frac{S_X(r)}{S_X(l)} = \frac{1 - F_X(r)}{1 - F_X(l)}. \end{aligned}$$

For item 2 we note that

$$F_Y(r - l) = 1 - S_Y(r - l) = 1 - \frac{1 - F_X(r)}{1 - F_X(l)} = \frac{F_X(r) - F_X(l)}{1 - F_X(l)}.$$

Then

$$f_Y(r-l) = \frac{d}{dr}\left[\frac{F_X(r)-F_X(l)}{1-F_X(l)}\right] = \frac{f_X(r)}{1-F_X(l)}.$$

For item 3 we have,

$$\begin{aligned} E[Y] &= \int_0^\infty y f_Y(y)dy \\ &= \int_l^\infty (r-l) f_Y(r-l)dr \\ &= \int_l^\infty (r-l)\left(\frac{f_X(r)}{1-F_X(l)}\right)dr \\ &= \frac{E[X]}{1-F_X(l)} \frac{\int_l^\infty (r-l) f_X(r)dr}{E[X]} \\ &= \frac{E[X]R_X(l)}{1-F_X(l)}, \end{aligned}$$

completing the proof. ∎

We want to fit an excess ratio function (see Definition 8), R_0, from a mixed exponential distribution to R_Y. More precisely, we want to replace the empirical loss variable, X, with a loss variable $\widetilde{X}$ such that if $\widetilde{Y} = \widetilde{X} - l$ conditional on $\widetilde{X} \geq l$ then

1. $f_{\widetilde{X}}(x) = f_X(x)$ for $x \leq l$
2. $R_{\widetilde{Y}} = R_0$.

We now derive the distribution function, the probability density function, and the excess ratio function of the spliced distribution $\widetilde{X}$.

Proposition 20 *The distribution function of the spliced random variable $\widetilde{X}$ is given by*

$$F_{\widetilde{X}}(r) = \begin{cases} F_X(r) & \text{if } r \leq l \\ 1 - [1-F_X(l)]S_{\widetilde{Y}}(r-l) & \text{if } r > l \end{cases}.$$

Proof. For $r \leq l$, we have $f_{\widetilde{X}}(x) = f_X(x)$. Thus

$$F_{\widetilde{X}}(r) = \int_0^r f_{\widetilde{X}}(x)dx = \int_0^r f_X(x)dx = F_X(r).$$

From this and Proposition 19, we see that for $r > l$

$$S_{\tilde{Y}}(r-l) = \frac{1 - F_{\tilde{X}}(r)}{1 - F_{\tilde{X}}(l)} = \frac{1 - F_{\tilde{X}}(r)}{1 - F_X(l)}$$

and therefore

$$F_{\tilde{X}}(x) = 1 - [1 - F_X(l)]S_{\tilde{Y}}(r-l).$$

∎

This allows us to determine the distribution function of $\widetilde{X}$ since we know the empirical distribution F_X and our assumption that $R_{\widetilde{Y}} = R_0$ determines the distribution of $\widetilde{Y}$ as well by Proposition 12. We have thus shown that the following two conditions

1. $f_{\widetilde{X}}(x) = f_X(x)$ for $x \leq l$
2. $R_{\widetilde{Y}} = R_0$,

uniquely determine a random variable $\widetilde{X}$. What we have not shown is that the above two conditions are consistent, i.e. that there exists a random variable, $\widetilde{X}$, that satisfies them. We do this by working with the density of $\widetilde{X}$ and show that $\widetilde{X}$ is a loss variable as well.

Proposition 21 *The density of $\widetilde{X}$ is given by*

$$f_{\widetilde{X}}(r) = \begin{cases} f_X(r) & \text{if } r \leq l \\ [1 - F_X(l)]f_{\widetilde{Y}}(r-l) & \text{if } r > l \end{cases}.$$

and this defines a valid density function of a loss variable with mean given by

$$E[\widetilde{X}] = E[X]\widehat{F}_X(l) + S_X(l)\left(E[\widetilde{Y}] + l\right).$$

Proof. Item 1 of the definition of $\tilde{X}$ ensures that $f_{\tilde{X}}(r) = f_X(r)$ for $r \leq l$. For $r > l$, we have from Proposition 20 that $F_{\widetilde{X}}(r) = 1 - [1 - F_X(l)]S_{\widetilde{Y}}(r-l)$ and thus

$$f_{\tilde{X}}(r) = [1 - F_X(l)]f_{\widetilde{Y}}(r-l).$$

It remains to show that $f_{\tilde{X}}$ is a valid density function. To show this, we compute

$$\begin{aligned}\int_0^\infty f_{\tilde{X}}(r)dr &= \int_0^l f_X(r)dr + [1 - F_X(l)] \int_l^\infty f_{\tilde{Y}}(r-l)dr \\ &= \int_0^l f_X(r)dr + [1 - F_X(l)] \int_0^\infty f_{\tilde{Y}}(r)dr \\ &= F_X(l) + [1 - F_X(l)] = 1.\end{aligned}$$

From $f_{\tilde{X}}$ we can compute the mean of $\widetilde{X}$.

$$\begin{aligned}E[\widetilde{X}] &= \int_0^\infty x f_{\tilde{X}}(x)dx = \int_0^l x f_X(x)dx + S_X(l) \int_l^\infty x f_{\tilde{Y}}(x-l)dx \\ &= \int_0^l x f_X(x)dx + S_X(l) \int_0^\infty (x+l) f_{\tilde{Y}}(x)dx \\ &= \int_0^l x f_X(x)dx + S_X(l) \left(\int_0^\infty x f_{\tilde{Y}}(x)dx + l \right) \\ &= E[X]\widehat{F}_X(l) + S_X(l) \left(E[\widetilde{Y}] + l \right).\end{aligned}$$

This shows that $\widetilde{X}$ is a loss variable because by assumption $\widetilde{Y}$ is. ■

Now we turn to the excess ratio function of $\widetilde{X}$.

Proposition 22 *The excess ratio function of the spliced random variable $\widetilde{X}$ is given by*

$$R_{\tilde{X}}(r) = \begin{cases} 1 - \frac{E[X]}{E[\widetilde{X}]}\left[1 - R_X(r)\right] & \text{if } r \le l \\ R_{\tilde{X}}(l) R_{\tilde{Y}}(r-l) & \text{if } r > l \end{cases}.$$

Proof. Using Definition 8 we first note that for $r \le l$ we have

$$\hat{F}_{\tilde{X}}(r) = \int_0^r x f_{\tilde{X}}(x)dx / E[\tilde{X}] = \frac{E[X]}{E[\tilde{X}]} \int_0^r x f_X(x)dx / E[X] = \frac{E[X]}{E[\tilde{X}]} \hat{F}_X(r).$$

Then using this relation and Propositions 9 and 20 we have for $r \le l$

$$
\begin{aligned}
R_{\tilde{X}}(r) &= 1 - \hat{F}_{\tilde{X}}(r) - \frac{r}{E[\tilde{X}]}\left[1 - F_{\tilde{X}}(r)\right] \\
&= 1 - \frac{E[X]}{E[\tilde{X}]}\hat{F}_X(r) - \frac{r}{E[\tilde{X}]}\left[1 - F_X(r)\right] \\
&= 1 - \frac{E[X]}{E[\tilde{X}]}\left[\hat{F}_X(r) + \frac{r}{E[X]}(1 - F_X(r))\right] \\
&= 1 - \frac{E[X]}{E[\tilde{X}]}[1 - R_X(r)].
\end{aligned}
$$

Now for $r > l$, using Propositions 19, 20, and 21 we get

$$
\begin{aligned}
R_{\tilde{X}}(r) &= \frac{1}{E[\tilde{X}]}\int_r^\infty (x-r) f_{\tilde{X}}(x)dx \\
&= \frac{1}{E[\tilde{X}]}\left[\int_r^\infty (x-r)[1 - F_X(l)] f_{\tilde{Y}}(x-l)dx\right] \\
&= \frac{1}{E[\tilde{X}]}\left[\int_{r-l}^\infty (x-(r-l)) S_X(l) f_{\tilde{Y}}(x)dx\right] \\
&= \frac{S_X(l)E[\tilde{Y}]}{E[\tilde{X}]}\left[\frac{\int_{r-l}^\infty (x-(r-l)) f_{\tilde{Y}}(x)dx}{E[\tilde{Y}]}\right] \\
&= \frac{S_{\tilde{X}}(l)}{E[\tilde{X}]}\frac{E[\tilde{X}]R_{\tilde{X}}(l)}{1 - F_{\tilde{X}}(l)} R_{\tilde{Y}}(r-l) \\
&= R_{\tilde{X}}(l) R_{\tilde{Y}}(r-l).
\end{aligned}
$$

∎

We would typically start with a distribution X that has mean 1 and so we would naturally normalize $\widetilde{X}$ and work with $\widetilde{X}/\widetilde{\mu}$ where $\widetilde{\mu} = E[\widetilde{X}]$. We use the following slightly more general proposition.

Proposition 23 *Let X be a random variable with density f_X and distribution function F_X, and let $\alpha > 0$, then*

1. $f_{X/\alpha}(x) = \alpha f_X(\alpha x)$
2. $F_{X/\alpha}(x) = F_X(\alpha x)$

Proof. We note that

$$\begin{aligned} F_{X/\alpha}(x) &= \Pr(X/\alpha \le x) = \Pr(X \le \alpha x) = F_X(\alpha x) \\ &= \int_{-\infty}^{\alpha x} f(y)dy = \int_{-\infty}^{x} \alpha f(\alpha y)dy. \end{aligned}$$

Thus $f_{X/\alpha}(x) = \alpha f_X(\alpha x)$ and $F_{X/\alpha}(x) = F_X(\alpha x)$. ■

From this and Proposition 20 we get the following.

Proposition 24 *The distribution function of the normalized spliced random variable $\widetilde{X}/\widetilde{\mu}$, where $\widetilde{\mu} = E[\widetilde{X}]$, is given by*

$$F_{\widetilde{X}/\widetilde{\mu}}(r) = \begin{cases} F_X(\widetilde{\mu}r) & \text{if } r \le l/\widetilde{\mu} \\ 1 - [1 - F_X(l)]S_{\widetilde{Y}}(\widetilde{\mu}r - l) & \text{if } r > l/\widetilde{\mu} \end{cases}.$$

We can similarly recast Proposition 22.

Proposition 25 *Let $E[X] = 1$, then the excess ratio function of the normalized spliced random variable $\widetilde{X}/\widetilde{\mu}$,where $\widetilde{\mu} = E[\widetilde{X}]$, is given by*

$$R_{\widetilde{X}/\widetilde{\mu}}(r) = \begin{cases} 1 - \frac{1}{\widetilde{\mu}}[1 - R_X(\widetilde{\mu}r)] & \text{if } r \le l/\widetilde{\mu} \\ R_{\widetilde{X}}(l)R_{\widetilde{Y}}(\widetilde{\mu}r - l) & \text{if } r > l/\widetilde{\mu} \end{cases}.$$

Proof. By Proposition 23 and the change of variables $y = \tilde{\mu}x$, we have

$$\begin{aligned} R_{\widetilde{X}/\widetilde{\mu}}(r) &= \frac{\int_r^\infty (x-r) f_{\widetilde{X}/\widetilde{\mu}}(x)dx}{E[\widetilde{X}/\widetilde{\mu}]} = \frac{\int_r^\infty (\tilde{\mu}x - \tilde{\mu}r) f_{\tilde{X}}(\tilde{\mu}x)dx}{E[\tilde{X}]/\tilde{\mu}} \\ &= \frac{\frac{1}{\tilde{\mu}}\int_{\tilde{\mu}r}^\infty (y - \tilde{\mu}r) f_{\tilde{X}}(y)dy}{E[\tilde{X}]/\tilde{\mu}} = R_{\tilde{X}}(\tilde{\mu}r). \end{aligned}$$

Then by application of Proposition 22 we have

$$R_{\tilde{X}/\tilde{\mu}}(r) = R_{\tilde{X}}(\tilde{\mu}r) = \begin{cases} 1 - \frac{1}{E[\tilde{X}]}[1 - R_X(\tilde{\mu}r)] & \text{if } r \le l/\tilde{\mu} \\ R_{\tilde{X}}(l)R_{\tilde{Y}}(\tilde{\mu}r - l) & \text{if } r > l/\tilde{\mu} \end{cases}$$

■

In our case we fit a mixed exponential to the tail of the empirical random variable X. More precisely, we assume that $\widetilde{Y}$ is a mixed exponential. That is, using the parameterization in Klugman, et. al. [9] (see page 43 on mixture models as well), we assume

$$f_{\widetilde{Y}}(x) = \sum_i w_i \frac{1}{m_i} e^{-x/m_i}$$

and thus

$$F_{\widetilde{Y}}(x) = \sum_i w_i(1 - e^{-x/m_i}) = 1 - \sum_i w_i e^{-x/m_i}.$$

Then by Corollary 18,

$$R_{\widetilde{Y}}(r) = \frac{\sum w_i m_i e^{-r/m_i}}{\sum w_i m_i}.$$

We now state Propositions 24 and 25 in the mixed exponential case.

Proposition 26 *If $\widetilde{Y}$ has a mixed exponential distribution as above then the distribution function of the normalized spliced random variable $\widetilde{X}/\widetilde{\mu}$ is given by*

$$F_{\widetilde{X}/\widetilde{\mu}}(r) = \begin{cases} F_X(\widetilde{\mu}r) & \text{if } r \le l/\widetilde{\mu} \\ F_{\widetilde{X}/\widetilde{\mu}}(l/\widetilde{\mu}) + \left[1 - F_{\widetilde{X}/\widetilde{\mu}}(l/\widetilde{\mu})\right]\left[1 - \sum w_i e^{-(\widetilde{\mu}r - l)/m_i}\right] & \text{if } r > l/\widetilde{\mu} \end{cases}.$$

Proof. From Proposition 24 for $r > l/\widetilde{\mu}$ we get

$$\begin{aligned}
F_{\widetilde{X}/\widetilde{\mu}}(r) &= 1 - [1 - F_X(l)]S_{\widetilde{Y}}(\widetilde{\mu}r - l) \\
&= 1 - [1 - F_X(l)][1 - F_{\widetilde{Y}}(\widetilde{\mu}r - l)] \\
&= 1 - [1 - F_X(l) - F_{\widetilde{Y}}(\widetilde{\mu}r - l) + F_X(l)F_{\widetilde{Y}}(\widetilde{\mu}r - l)] \\
&= F_X(l) + F_{\widetilde{Y}}(\widetilde{\mu}r - l) - F_X(l)F_{\widetilde{Y}}(\widetilde{\mu}r - l) \\
&= F_X(l) + [1 - F_X(l)]\, F_{\widetilde{Y}}(\widetilde{\mu}r - l) \\
&= F_{\widetilde{X}}(l) + \left[1 - F_{\widetilde{X}}(l)\right] F_{\widetilde{Y}}(\widetilde{\mu}r - l) \\
&= F_{\widetilde{X}/\widetilde{\mu}}(l/\widetilde{\mu}) + \left[1 - F_{\widetilde{X}/\widetilde{\mu}}(l/\widetilde{\mu})\right]\left[1 - \sum w_i e^{-(\widetilde{\mu}r - l)/m_i}\right]
\end{aligned}$$

■

Proposition 27 *If $\widetilde{Y}$ has a mixed exponential distribution as above and $E[X] = 1$, then the excess ratio function of the normalized spliced random variable $\widetilde{X}/\widetilde{\mu}$ is given by*

$$R_{\widetilde{X}/\widetilde{\mu}}(r) = \begin{cases} 1 - \frac{1}{\widetilde{\mu}}[1 - R_X(\widetilde{\mu}r)] & \text{if } r \le l/\widetilde{\mu} \\ R_{\widetilde{X}}(l)\frac{\sum w_i m_i e^{-(\widetilde{\mu}r - l)/m_i}}{\sum w_i m_i} & \text{if } r > l/\widetilde{\mu} \end{cases}.$$

References

[1] Billingsley, Patrick, *Probability and Measure*, 2nd edition, Wiley, 1986.

[2] Brooks, Ward, "The 2001 California Retrospective Rating Plan–Developing Losses to Ultimate," unpublished manuscript.

[3] Feller, William, *An Introduction to Probability Theory and Its Applications,* Volume II, Wiley, 1966.

[4] Gillam, William R., "Retrospective Rating: Excess Loss Factors," *PCAS* LXXXVIII, 1991, pp. 1–40.

[5] Gillam, W.R., and Couret, J.R., "Retrospective Rating: 1997 Excess Loss Factors," *PCAS* LXXXIV, 1997, pp. 450–481.

[6] Harwayne, Frank, "Accident Limitations for Retrospective Rating," *PCAS* LXIII, 1976, pp. 1–31.

[7] Gleeson, Owen M., "Ratemaking for Excess Workers Compensation," *CAS Forum*, Winter 2001, pp. 61-114.

[8] Keatinge, C.L., "Modeling Losses with the Mixed Exponential Distribution," *PCAS* LXXXVI, 1999, pp. 654–698.

[9] Klugman, Stuart A., Panjer, Harry H., Willmot, Gordon E., *Loss Models: From Data to Decisions,* 2nd edition, Wiley, 2004.

[10] Mahler, H.C., "Workers Compensation Excess Ratios: An Alternative Method of Estimation," *PCAS* LXXXV, 1998, pp. 132–156.

[11] Uhthoff, Dunbar R., "Excess Loss Ratios via Loss Distributions," *PCAS* XXXVII, 1950, pp. 82–93.

[12] Venter, G. G., "Scale Adjustments to Excess Expected Losses," *PCAS*, LXIX, 1982, pp. 1–14.