Unstable Loss Development Factors

Gary Blumsohn, FCAS, Ph.D., and Michael Laufer, FCAS, MAAA

Abstract

Most actuaries learn loss development on the job and pick up whatever techniques are being used by those around them. The experienced actuary is exposed to many varieties of methods and techniques. In dealing with unstable triangles, actuaries will employ myriad assumptions, judgments and tools along the way to selecting loss development factors. The authors describe a recent survey demonstrating the variety of methods and variability of selections of loss development factors (prior to consideration of the tail) and the variability of the resulting reserve projections.

Keywords: Reinsurance Analysis, Trend and Loss Development, Reserving Methods, Reserve Variability, Uncertainty and Ranges.

1. INTRODUCTION

A major challenge in day-to-day reinsurance actuarial work is selecting loss development factors when triangles are unstable. This topic does not receive significant attention in exams and papers, and yet it's something reinsurance actuaries encounter regularly. Most actuaries learn how to select loss development factors on the job, picking up rules of thumb and helpful approaches along the way. Whether these ad hoc approaches are good or bad depend on the particulars of the underlying data.

In the early part of 2008, we asked a group of actuaries to select loss development factors for a 12-year triangle of umbrella business (disguised in various ways to avoid divulging proprietary information). The selection of the group was not random: it consisted of people signed up to attend the 2008 Casualty Actuarial Society Seminar on Reinsurance, as well as various acquaintances of the authors. There was nothing special about this triangle, other than that it was deemed to be sufficiently unstable for the purpose at hand.

The triangle was provided in an Excel spreadsheet, and the participants were asked to select age-to-age factors. To keep the topic focused on the triangle, participants were instructed to ignore the tail factor. They were also asked to describe how they selected their factors, including such items as what types of averages they relied on, how they dealt with outliers, and how they dealt with reversals in the pattern. The original request, including the triangle, is shown in Appendix A.

Originally, the project was intended to lead to work that would provide some guidance on the selection of factors. This paper does not provide such guidance, except indirectly. Rather, this paper reports the results of what we received and catalogs the high variation in people's responses.

While we provide some commentary along the way, we largely allow the results to speak for themselves.

We received 52 responses, although only 51 of them gave us selected factors. The other one, from senior actuary with many years of experience, gave us a list of questions that one would need to ask before even beginning to select factors. Though this actuary had a point, we continued with our project nevertheless.

Initial reactions from people varied from the positive

> "*Great and gutsy project!*"

to the horrified

> *"I believe the whole notion of 'picking factors' with no statistical guidance is something of a disgrace to the profession…."*

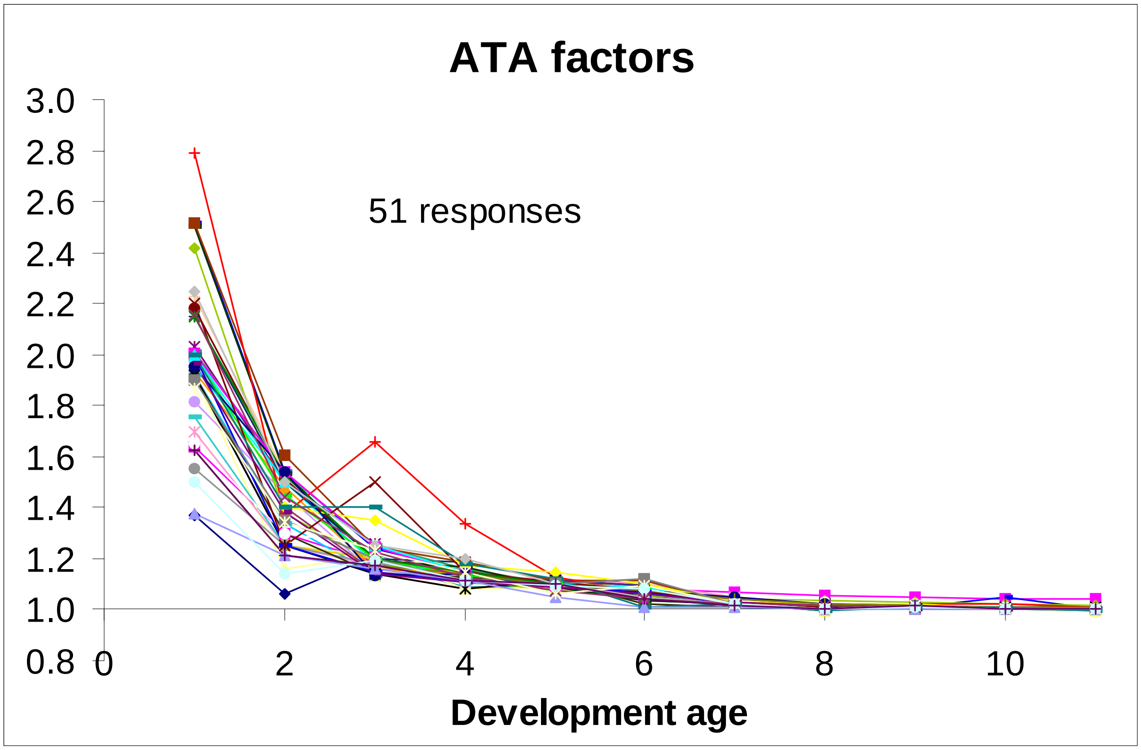
to the concerned

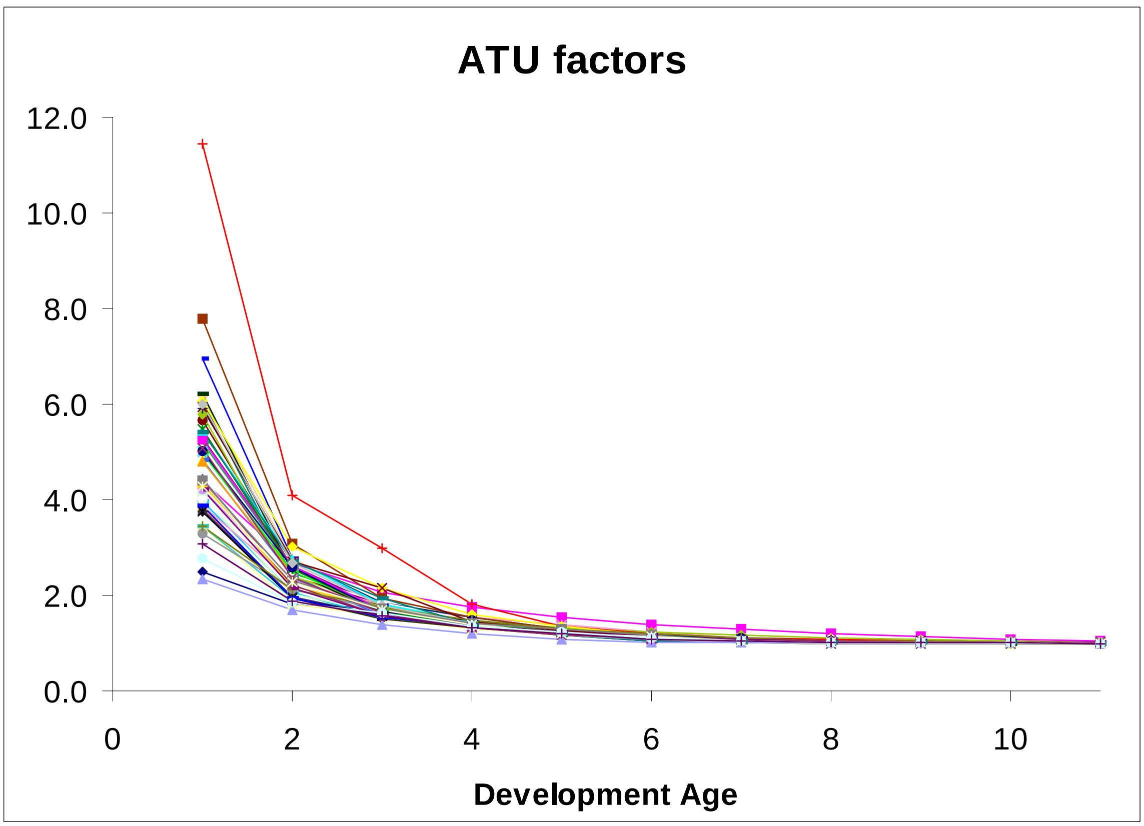
> *"While it may be helpful to share ideas on how to pick LDFs, it is vital that more information than just the triangle at hand be considered… I wouldn't make selections without other information such as individual claim information, changes in the underlying business, comparison to competitor or industry triangles if available, etc. Of course you can't always get the information you want…but I would hate to see people come to the seminar and learn some new selection techniques that don't look beyond the triangle."*

2. THE RESULTS

This section of the paper summarizes the results we received. For now, we content ourselves with describing the outputs, leaving for later people's explanations of their loss development factor picks.

The graphs below show the age-to-age (ATA) factors and the age-to-ultimate (ATU) factors selected by each of the 51 participants. While it isn't easy to draw detailed conclusions simply from looking at the graphs, one can quite easily see that the range of selections is wide.



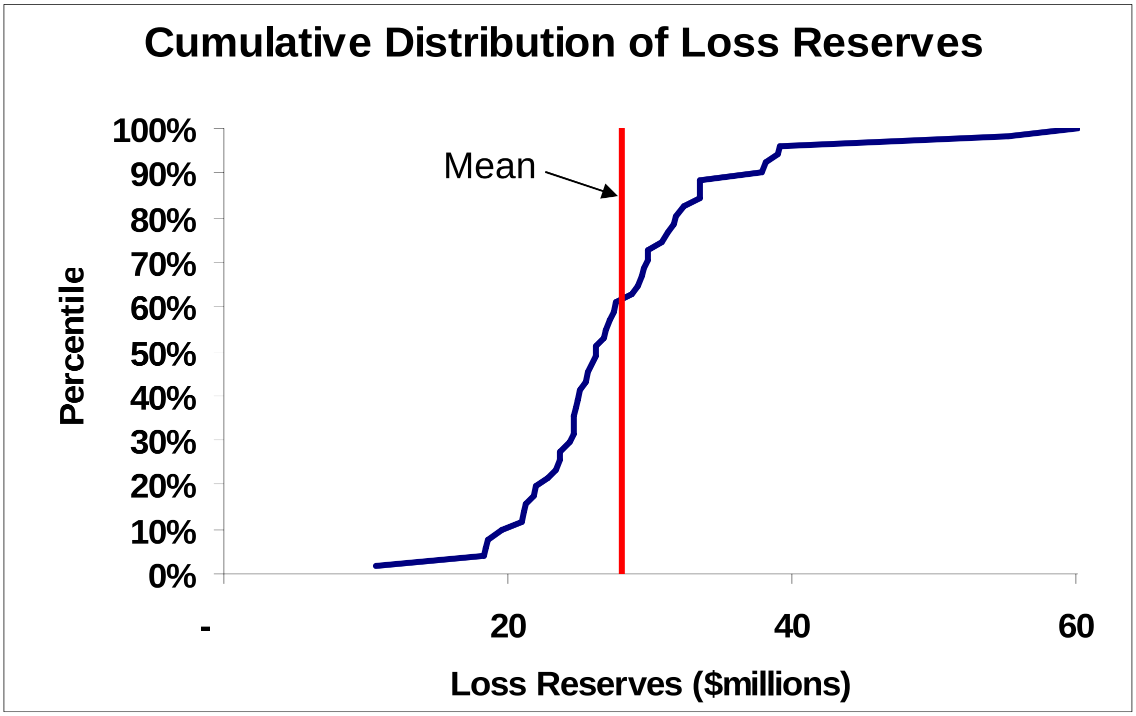


While one could measure the dispersion among selections by looking, for example, at the coefficient of variation of the various selections, we chose not to take that approach. In the spirit of a paper that is more psychology and rules of thumb than technical actuarial work, we are interested in the practical import of the dispersion among actuaries' selections. In that vein, our measure of dispersion will rely on looking at the expected reserve from a chain-ladder projection – in this case, the dispersion of the expected chain-ladder reserve that results from the different factor selections. We are, of course, aware that most actuaries would not use only one method to get the reserves, and would quite likely rely on a more stable method, like the Bornhuetter-Ferguson, to estimate the reserves for the more recent years. However, for our purposes we choose to ignore this because we are focusing on the development factors, rather than on the full range of reserving procedures.

The table below and the graph that follows show the implied reserves from the 51 respondents, ranging from a low of $10.7 million to a high of $60.2 million. The mean is $28.0 million, with a standard deviation of $8.3 million – a coefficient of variation of a whopping 30%.

Implied chain-ladder reserves, sorted from lowest to highest
(in $millions)

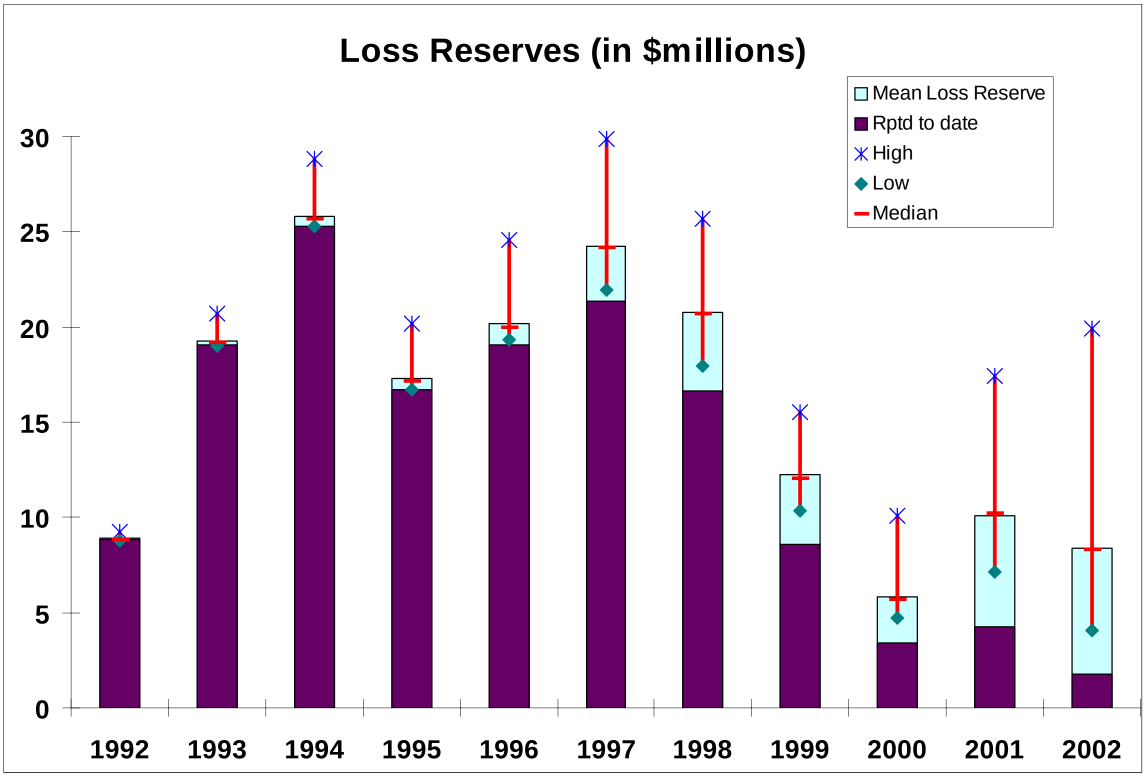
1	10.7		18	24.7		35	29.5
2	18.3		19	24.8		36	29.8
3	18.5		20	25.0		37	29.8
4	18.6		21	25.1		38	30.9
5	19.5		22	25.4		39	31.3
6	21.0		23	25.6		40	31.7
7	21.2		24	25.9		41	31.8
8	21.3		25	26.2		42	32.4
9	21.8		26	26.2		43	33.5
10	22.0		27	26.7		44	33.5
11	22.8		28	26.9		45	33.5
12	23.4		29	27.1		46	38.0
13	23.7		30	27.5		47	38.2
14	23.7		31	27.6		48	39.0
15	24.4		32	28.7		49	39.2
16	24.6		33	29.1		50	55.2
17	24.7		34	29.4		51	60.2
			Mean	28.0			
			Median	26.2			
			Std. deviation	8.3			
			Coefficient of variation	30%			



The range and the standard deviation of the reserves are greatly widened by three apparent outliers: one on the low end with reserves of $10.7 million, and two high-end selections of $60.2 million and $55.2 million.

Half the responses implied reserves in a fairly tight range between $23.7 million and $30.9 million. However, one cannot ignore that half the actuaries made picks that implied loss reserves outside of this range. In a world in which many employers, regulators, auditors, and investors seem to think actuaries can make loss picks that are within 10% of the "truth," this should be sobering. And it should not be forgotten that we deliberately ignored the tail factor in this exercise, which reduces the volatility.

The following graph shows the distribution of the reserves by accident year. It is not surprising that the greatest variability is in the most recent years.

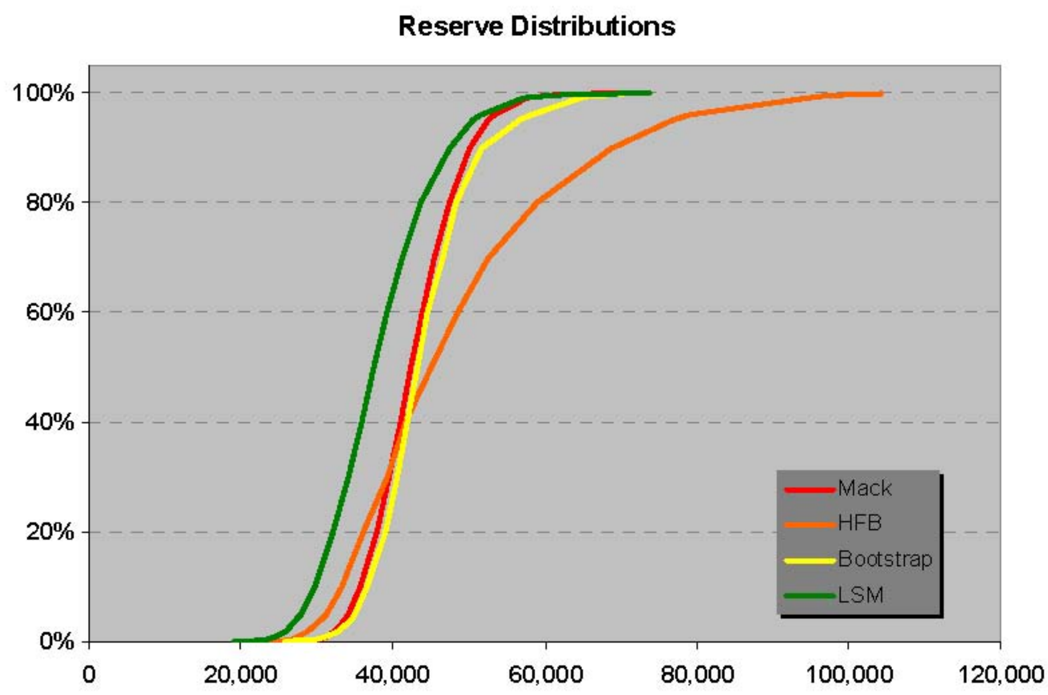


This graph has profound implications for pricing actuaries. It is not unusual for an actuary pricing a quota share to have 10 or fewer years of data. One would typically trend and develop losses and on-level the premiums, to bring everything to current level. One would then typically review the series of loss ratios in an attempt to divine the most likely loss ratio for the year being priced. Many pricing actuaries put more weight on the recent years than on the older years, on the assumption that more recent information is more valuable. However, the volatility in the estimates of the most recent years is extremely high, so that there is a high probability that the pricing estimate will differ significantly from the true mean.

In interpreting these results, it should not be forgotten that the distribution shown here is a distribution of *estimates of the mean* of the distribution of the unpaid losses. It is not a distribution of the unpaid losses. Typically, although not necessarily, the distribution of the unpaid losses will be considerably wider than distribution of estimates of the mean of the distribution of unpaid losses. For example, if we assume the estimate of the mean of the unpaid losses has a standard error of 30%, and the distribution of unpaid losses, given the expected reserve, has a coefficient of variation of 15%, then one would have to mix these two distributions to get an overall distribution of outcomes. The potential dispersion of loss reserve outcomes is extremely high.

2.1 A Side-Note on Reserve Variability and the Reputation of Casualty Actuaries

In March 2008, the CAS Task Force on Enhancing the Reputation of Casualty Actuaries produced a report ([2]) that discussed, among other things, the need for actuaries to communicate the uncertainty of their estimates to ensure that users understand it. They also called for an improvement in actuarial methodologies and terminology, all of which is to be applauded. What struck us, relative to our work in this paper, was a section titled "Comparison of Methods" on pp. 16 – 18 of that report, and, in particular, a graph on p. 17, reproduced below.



The graph shows the distribution of reserves that emerge from four approaches actuaries have used for getting a distribution of reserves, as applied to some unspecified general liability data.[1]

While we endorse the task force's suggestion that actuaries be concerned with the variability of reserves as demonstrated by a particular method, we want to stress the need to account for both the variability that results from a particular method of calculation as well as the variability among particular methods of calculation. At each step of the process, there is variability, and all the variability gets compounded. For example, if my analysis indicates that the mean loss reserve for a block of business is 100, and my approach to calculating the standard deviation of the reserve indicates that the standard deviation is 15, I could proceed to use those assumptions to estimate, say,

---

1 For Mack method, see [7]; for HFB method see [6]; for Bootstrap method see [3]; and for LSM method see [8].

the 95th percentile of the distribution of losses. However, I must recognize that if another actuary were to estimate the mean, that estimate could easily be, say, 92 or 112. Even if the other actuary's estimate of the standard deviation were the same as mine, the range of possible outcomes from these two actuaries will be considerably wider than the range of possible outcomes from my estimate alone. And if we recognize that the difficulties in estimating the standard deviation and the shape of the distribution are even greater than the difficulties in estimating the mean, we should recognize that our ability to accurately pin down the tail probabilities is very limited.

For example, in the GL example used by the CAS Task Force on Enhancing the Reputation of Casualty Actuaries, some of the methods gave a 95th percentile that is about 25% above the mean. If we assume, for convenience, that the reserves are lognormally distributed, this implies a coefficient of variation of about 15%. However, if we assume that the mean itself is lognormally distributed with a mean of 30%, then the mixture of the two distributions gives a 95th percentile that is more than 60% higher than the mean. To make things even harder, once one realizes that estimates of the coefficients of variation are themselves just estimates, and that the shape of the distribution is usually not much more than an educated guess, one realizes that one can place little faith in one's estimates of the tails of the distribution. If one is worried about the credibility of casualty actuaries, this is a key point that must be made.

2.2 How Did Participants Get Their Factors?

Most of the participants (42 of the 51) either wrote some explanation of how they had derived their factors or set up a spreadsheet so that it could be inferred how they thought. Broadly, these can be divided into those who picked factors based on reviewing the various averages of the factors (34 of the 42) and those who used some statistical approach to the problem (8 of the 42), though it must be noted that this classification is somewhat arbitrary, as there were a number of responses that were in a gray area, and where we used our judgment. A complete listing, edited for spelling and anonymity, together with the selected factors, is in Appendix B.

In comparing the factors from the various methods, we calculated the means and standard deviations of the implied loss reserves:

Method classification	# of Respondents	Mean implied reserve (in $millions)	Standard deviation of the implied reserve (in $millions)
Pick	34	27.1	5.9
Statistical	8	36.0	13.2
Not clear	9	24.6	5.0
Total	51	28.0	8.3

The statistical approaches gave a mean and standard deviation that were both much higher than from the pick approaches. However, given the small number of statistical approaches, and the possibly random fact that both of the high-end outliers are statistical approaches, this might be noise, rather than a signal that statistical methods generally lead to higher answers.

2.3 What Methods Did Participants Consider in Developing Their Factors?

We found broad categories in which participants described what they considered in making their factor selections, and they are summarized in the table below. Any single respondent might have given thought to more than one of these considerations, and some respondents did not tell us how they got their selections. Many respondents gave thought to why different methods were yielding different results, and they described why they believed a particular method to be appropriate to the case at hand.

The three responses well out of range mentioned above are removed from this table. The average estimate for the remaining 48 respondents is $27.2 million with a standard deviation of $5.3 million.

Table: Considerations in Selecting Factors for Unstable Triangles

	Average Reserve (in $millions)		Standard Deviation (in $millions)		# of Respondents	
Consideration	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
(1) Removed Outliers (including ex-high-low)	27.9	27.8	5.3	5.5	15	21
(2) Smoothed Links	28.7	24.0	8.1	3.3	23	13
(3) Would Use Only Factors > 1.00	27.5	25.0	5.7	1.0	12	5
(4-a) Used Volume Weighted Average	26.6	34.7	6.9	7.6	30	3
(4-b) Used Straight Average	25.2	27.6	11.3	5.1	8	24
(4-c) Used Ex-High-Low Average	27.1	28.0	5.6	5.0	12	17
(5) Used All (or Nearly All) Years	29.2	21.7	5.0	8.4	19	11
(6) Used Industry Data	32.4		4.8		3	
(7) Used Regression or Other Curve Fitting	28.0		4.9		9	
(8) No Risk Margin	24.7		7.3		2	
(9) Wanted More Information	26.2		8.3		17	

The decision whether or not to consider a certain method might have resulted in significantly different results; for example, whether to smooth the data, rely on all-year averages, use industry data, use curve-fitting techniques, or use volume-weighted averages.

We now address each category listed above in more detail.

2.3.1 Removing outliers (15 responses)

Some respondents began by checking the data for anomalies, and by excluding certain years or links; for example by excluding certain links when calculating link averages, by removing certain accident years, or by not using an average that contained the outlier. The questions here are (1) the extent to which outliers in the data can be identified, (2) the potential causes of the outliers, (3) how the outliers should be treated, and (4) the impact of removing or including the outliers on the result.

These questions are somewhat beyond the scope of this paper. Without resolving these issues, it seems we should consider the following:

- From a purely statistical standpoint, one might argue that removing outliers will tend to bias the results, especially seeing the distribution of development factors is probably positively skewed, so that outliers are more likely to be identified in the right-hand tail of the distribution. If these outliers are discarded, there will be a tendency for results to be biased downwards.[2]
- It must be recognized, though, that most triangles consist of a small number of points. A 12-year triangle, such as we are using as the basis for this paper, has only 66 loss-development points – hardly a large sample with which to apply fancy techniques. Outliers, particularly towards the tail of the triangle, can sharply change the results.
- If one is dealing with a large number of triangles, one might be most concerned with having an unbiased set of development factors. This might be the case if one is doing a reserve study, has divided the book of business into 20 segments, and is selecting development factors for each segment. By retaining the outliers, one acknowledges that the reserves on some segments will be high, and others will be low, but overall, one hopes to have an unbiased result. On the other hand, suppose one is a pricing actuary, separately pricing 20 reinsurance transactions. If one retains the outliers, the results on the 20 transactions may well be unbiased, but since clients are more likely to accept quotes that are based on low development factors than quotes based on high development factors (the well-known Winner's Curse – see [10], for example), the business that ends up on the books will more likely be where the loss estimates were biased downwards. In this situation, reducing the variance of the results by eliminating outliers may be more important than introducing a (hopefully) small amount of bias in the results. The uncomfortable upshot is that actuaries must recognize that there is no "best way to select development factors" isolated from the purpose for which those factors are being selected. Context matters.
- Until now, we've assumed the outlier question is a statistical one. Beyond the statistical question, there's the question of data errors. Reinsurance actuaries are familiar with the problems of data quality, and our well-honed intuition is alert for outliers that are not the result

[2] For a recent discussion of the impact of skewed distributions and small sample sizes in actuarial work, see [4].

of statistical fluctuation, but are the result of claim-department policy changes, changes in underwriting approach, changes in the mix of business, changes in the approach to settling claims or setting case reserves, coding errors, claim personnel errors, or simply errors without a cause known to the actuary. As a practical matter, it's virtually impossible for reinsurance actuaries to uncover the causes of many of the fluctuations in the data, or even to identify whether they are statistical fluctuations rather than data errors. We can speculate as to the causes, but without details of the underlying policies and claims, it's impossible to pin them down. Our assumption in eliminating outliers is that we are able to improve the quality of the data, and hence the quality of the answers, through "actuarial judgment," but one wonders whether it would be possible to somehow set up an experiment that would test this assumption.[3]

2.3.2 Link smoothing (23 responses)

Some respondents smoothed selected links across ages. A few believed it was appropriate to consider adjusting for reversals in the factors where links did not show a smooth pattern of decreasing with maturity; however others believed it unnecessary to smooth links at all.

Some respondents may have adjusted the data when they perceived takedowns followed later by upwards adjustments, or the reverse. In general, it probably isn't a bad a priori assumption that age-to-age factors decrease monotonically and smoothly until they reach 1, but it's only an assumption, and there are significant situations where the assumption has been wrong.

One of the authors worked on an instructive situation where this assumption proved wrong. It involved pricing working-layer excess workers compensation reinsurance. In reviewing the incurred-loss triangle, a surprising number of the 24-to-36 factors developed downwards, despite all the surrounding factors developing upwards. At first, we were tempted to smooth out this downward development. However, this was one of the relatively rare situations where we had complete access to the underlying data, and, once we examined the data, the downward development made perfect sense. The layer was high enough that most of the dollars of claims in

[3] The recent book *Super Crunchers* by Ian Ayres [1] devotes a chapter to "Experts Versus Equations" (pp. 112 – 139), which presents evidence that mechanically applied equations seem to come up with better predictions than "experts" in a wide variety of situations. Ranging from models of legal decision-making that were better at predicting Supreme Court decisions than a group of experts, to medical studies where models were better at predicting how patients with schizophrenia would respond to electroshock therapy, to studies showing that models made better purchasing decisions than professional purchasing managers, Ayres contends that models work better than experts. He attributes the failures by experts to the cognitive biases and overconfidence that are by now well known.

In the actuarial case, it would be intriguing to have an experiment to test whether, over a large number of triangles, one would get better answers from purely mechanical application of some algorithm for selecting development factors. We like to think we add value in our analysis of loss triangles, and no doubt some data errors would not be revealed by an algorithm. But one wonders whether the data "errors" we find, and the outliers we throw out, are too often valuable bits of data that we choose to ignore more frequently than we should – at our peril and at the peril of our employers.

the layer, and especially those that were reported early, were from very severe injuries, such as brain and spinal-cord injuries. When the case reserves were first put up, they covered an "average" claim amount for that injury type. However, a significant fraction of severely injured people die within a year or two of the injury, and when that happened, the case reserve dropped to zero, or perhaps a small amount paid to the claimant's survivors. Meanwhile, the case incurred amount on those claimants who survived was kept the same as when it had first been set, with no increase. This explained the downward case development observed in the 24-to-36 factor. The subsequent upward case development was explained by less obviously severe claims, such as back or knee injuries, bleeding into the excess layer, as well as eventual recognition of increased costs due to medical inflation and increased life expectancies for those who survived the first couple of years. As actuaries, we may have suggestions for better ways to set the case reserves so that the loss development is smoother, but the bottom line is that we were not the ones setting the case reserves, and the downward development was real. The real world isn't always smooth.

2.3.3 Adjusting for factors less than 1.00

Akin to smoothing links, some respondents believed it was appropriate to adjust for reversals in the data where development factors were less than 1.00. Some expressed a tolerance for development factors that were marginally less than 1.00.

The workers compensation example given above applies to this situation as well. Sometimes downward development is real. Reinsurance actuaries sometimes learn about the case reserving habits of their clients – which clients tend to under-reserve and which tend to over-reserve. It is not altogether unusual to have downward development on incurred triangles (and occasionally on paid triangles). Most actuaries will concede this point, but when given a triangle where the case reserving practices are unknown, they are often resistant to allowing downward development, and perhaps for good reason. Downward case development is often a tipoff to bad data, or the existence of some unusual situation that is unlikely to be predictive of future downward development. While we cannot be sure that downward development is wrong, it may be correct to underweight it.

Another valid downward development situation occurs, of course, at mature ages in lines with significant salvage or subrogation, when data is given net of paid salvage and subrogation.

2.3.4 Type of average

Many respondents thought it was appropriate to calculate a variety of averages using varying weights and varying years. After doing so, respondents felt better equipped to discern trends, spot outliers, or raise other questions about the data.

Preferences for how to weight the averages ranged widely. (1) Some participants expressed a preference for volume-weighted averages on unstable triangles to avoid over-weighting erratic low-volume years. (2) Others were apprehensive about using volume-weighted averages without additional information such as large claim information, and preferred to use straight averages. (3) Some thought it was appropriate to use averages-excluding-high-low. (4) Some weighted their averages using time-sensitive weights.

Daniel Murphy's paper "Unbiased Loss Development Factors" [8] treats development factors as regression coefficients from the equation $y = bx$ + error, and shows the various assumptions that would make weighted or unweighted averages the best choice. Various authors have concluded that it generally appears that weighted averages are better than unweighted averages. See, for example, Struzzieri and Hussian's nice summary of these various results in their section "Best Link Ratio Averages" (pp. 384 – 388 of [9]). However, the interesting thing is that, with all of these papers by Murphy and Mack and others stressing that the key item is to check the assumptions, only one of the respondents to our survey appeared to have actually checked the assumptions (respondent number 32 – see Appendix B).

Though, as Struzzieri and Hussian note (p. 386), the literature generally supports weighted averages over unweighted averages, it is not perfectly settled in our minds. A counter to weighted averages is that if, say, high development factors correspond to years with high volume, the weighted average could bias the answer high. Another counter would be that in a time of high inflation, more recent years get more weight, even though they may be no different from the earlier years in real terms. While these counter-arguments are correct, our casual empiricism says that for the typically unstable triangles that we see in reinsurance, weighted averages are almost always superior.

2.3.5 Appropriate number of years to use in the average

Many participants preferred to use all (or nearly all) valuation years of data when reviewing unstable triangles. Others gave more weight to recent valuations, sometimes looking for patterns that they believed were different and discernible in the most recent valuations.

We are not aware of literature that answers the question of whether to put more weight onto recent information. Intuitively, it seems reasonable, since we believe the statistical process underlying the loss development changes over time, so more recent data is likely to be more representative of the future. However, the guidance on when to use each type of average seems to be fairly informal.

On the other hand, volatile data might be the result of claim situations that occur infrequently. For example, if a particular type of claim is encountered only once or twice a decade, and the development pattern for this type of claim is unique, it makes sense to include as many years as possible, to capture the data from this infrequent claim type. If this particular claim situation and its resulting development would not be captured in the most recent three or five-year average, using these short-term averages could yield inaccurate results; or if it is captured in the most recent years, it could be overweighted.

In determining the appropriate number of years to use, it is important for the actuary to assess the issues of (a) using later information, which presumably provides better insight into changes in the claims development process or changes in the claims environment, and (b) using more data points, which would contain development information on more (types of) claims.

2.3.6 Industry information (3 responses)

Some participants used industry link information culled from their own sources. These respondents generally weighted industry data with the data in the exercise for some or all of the links. These respondents may have considered adjusting their industry link set to be consistent with the exercise data.

While it's always useful to compare one's data to industry information, there was very little information provided about the triangle. Participants were told that it was "umbrella" data, but were not told whether it was personal umbrella or commercial umbrella, supported or unsupported, the limits being offered, and other important information. In the absence of this information, one must wonder how much weight one can reasonably give to industry information.

2.3.7 Regression or other curve fitting (9 responses)

Some respondents used limited regression techniques on the data to come up with selections or used regression techniques as a check on their selections. While each of these respondents used some regression technique, each one seemed to be using a different technique and applying it to different data, so that there wasn't a consistent method for us to describe.

Other respondents fit curves to the data, relying on a number of curves and fitting techniques. Curves mentioned include the inverse power/loglogistic, lognormal, exponential and Weibull. Again, there was no single, consistent method for us to describe.

2.3.8 Risk margin (2 responses)

Two respondents wondered whether to include a risk margin in their selections; they both ultimately decided not to include a risk margin.

We had not expected to encounter risk margins in this exercise, so it was interesting that more than one person raised the question. We aren't sure of the benefits of including a risk margin in development factors, rather than building risk margin in at the end of the procedure. For pricing, one would want a risk margin, though it's hard to see that the best place for it is in the development factors, rather than directly building it into the final price. A risk margin may be warranted in reserving, although, for example, US GAAP requires that companies carry the best estimate of the reserves – presumably precluding a risk margin.

2.3.9 Needs more information

We did not ask participants to tell us if they would have liked to see additional information; nevertheless, many participants described what additional information they felt would have been useful.

The type of information requested can be categorized into the following groups: (a) company loss data, (b) claims department information, (c) underwriting data, (d) industry data, and (e) prior selections.

A. Loss Data
    a. Paid triangle
    b. Claim count triangle
        i. Open claim count triangle
        ii. Paid closed claim count triangle
    c. Average claim size triangle
    d. Individual claim development

Participants thought the paid data could provide additional insight into the claims settlement and reserving process. Similarly, participants would have liked to have reviewed claim count data.

B. Qualitative claims information
    a. Changes in company case reserving philosophy
    b. Changes in company claim processing
    c. Claims audit report

C. Underwriting data
    a. Premium or other exposure proxy by year
    b. How long the company has been in the line
    c. Information on the underlying book

i. Retention and limit (and/or their changes over time)

ii. Type of umbrella

1. personal vs. commercial
2. supported vs. unsupported

d. Mix of business changes

Some respondents said they made assumptions with regard to exposures. Others said they would have liked to know more about the underlying book, with a view to guiding their selections.

D. Industry Information

a. Industry default development factors

b. Underwriting cycle information

c. Loss ratio benchmarks

d. Legal/legislative trends

Many participants said they would have liked to have had industry development factors, hoping to apply weights to company data vs. industry data. As mentioned above, some responders did assume an industry pattern and considered this industry pattern in their selection.

Some respondents said they would consider adjusting their factors by accident year based on where each year stood in terms of the underwriting cycle. This is an interesting notion that needs to be examined further. While some stable lines of business, like primary workers compensation, sometimes exhibit have cyclical loss development, we are not aware of anyone successfully applying cycles on top of unstable factors. It would be interesting to know whether the assumptions of cycles can be shown in fairly general cases to mute the instability of factors, or whether they might add to the noise.

Some respondents would have liked to have had company premium information and industry loss ratios in order to compare company loss ratio indications to industry loss ratio benchmarks. While loss ratios are always helpful, it isn't clear to us how to change development factors based on this information: if the company's loss ratios are much higher than industry loss ratios, does that mean higher or lower development factors?

Others mentioned they needed more information on recent industry trends. This request seems reasonable. An understanding of industry trends, and more generally, social and legal developments, might explain some apparently odd loss development, or might lead one to select development factors that are higher or lower than those suggested by the triangle.

E. Prior information

One respondent thought an appropriate method of validating selections would be to review prior selections that were based on this data. If, say, one were updating a reserve study, this would be a reasonable request, although one that obviously isn't always possible.

2.4 What lessons can we draw from the three respondents we removed from our review?

As mentioned above, three respondents were well out of range of the other participants. One of these participants was very low and two were very high. Each of these respondents discussed their methodology.

2.4.1 Second highest respondent

The respondent who produced the second highest loss reserve fit a single curve to all maturities.[4] Without knowing the details of the distribution that was fit, it's hard to comment definitively on this approach. However, the respondent produced low or moderate link ratios through the first six or so maturities and high factors beyond age six. In fitting the later maturities a reasonability check might have provided this respondent with additional insight into his fitting procedure. For example other respondents used a reasonability check on fitted results by comparing the fitted links to average links from the data over a multiple maturity period. In this case the respondent's fitted links from age 7:12 results in a total multi-age link of 1.29 while the data shows volume-weighted 7:12 link of about 1.03.

While fitting a distribution may seem more statistically sophisticated, there is the ticklish problem of selecting a distribution. Given the limited number of data points in a typical loss triangle, it's quite likely that many distributions will pass a goodness-of-fit test. It's not obvious that any of the "usual" distributions should be believed a priori to provide a good fit to loss development factors.

2.4.2 Lowest respondent

The respondent who produced very low results selected simple averages based on only the most recent valuations for the first four links.[5] For these earlier maturities, selecting only the most recent

---

[4] This respondent described his process as follows: "I model a distribution of the 'time until reported' random variable. I then fit the incremental amounts to what the distribution would imply as a fraction of the cumulative amount for the accident period. I then inspect the errors for consistent bias and make adjustments if appropriate."

[5] This respondent also noted that he generally prefers weighted averages, but "since I just came back from vacation and will soon be on the road again, I did mostly a straight average analysis." We are not sure what lessons should be drawn from this comment. If the comment is meant to say he was pressed for time, so that he took some short-cuts, then the lesson would be that short-cuts are dangerous, and one should be especially careful of making rapid selections when

valuations, of course, ignores the development of the earlier accident years. Many other respondents believed that for unstable triangles it was important to consider development beyond the most recent accident years.

Also after the first four links, this respondent selected factors based on judgment, believing the data to have little credibility beyond this point. This judgment resulted in factors that were much lower than what the data indicated. For example using the reasonability check described above we see that for this respondent the multi-age link over the period 5:12 is a factor of 1.08, while the data shows a volume-weighted 5:12 link of 1.21.

2.4.3 Highest respondent

The respondent who produced the highest results described used a regression analysis on the first four links. In particular, noting a negative correlation between the age-to-age factor and the dollars of reported loss, he fit a regression line with the age-to-age factor (say the 12-24 factor) as the dependent variable and the dollars of reported loss (at age 12, in this example) as the independent variable. This regression was then used to fill out succeeding points within the lower half-triangle. After the first four years, he selected age-to-age factors judgmentally, "with the consideration of weighted averages, simple averages, smoothness, conservatism, and historical LDF range." In fact, it was in large part due to the judgmental factors after age 4 that led to the high result. If he had simply used the all-year weighted averages, the 5:12 factor would have been 1.21, and the reserve would have been $45.3 million, which would still have been high, but much lower than the selected 5:12 factor of 1.35 that gave reserves of $60.2 million.

The point regarding the negative correlation for the first four factors is correct, and is not addressed by most other participants. It is notable, however, that the dollars of loss could be low either because i) there is light reporting at, say, age 12, which is then offset by "more normal" reporting at age 24, and hence the negative correlation that was noted, or ii) the exposure is lower. Exposures were not given in our example, so it is not possible for the actuary to distinguish between them. This lack of exposure data was noted by a few of the respondents.

2.4.4 Lessons from the outliers?

It's hard to draw firm lessons from the three outliers. One apparent lesson is for the need to use reasonableness checks when heavily relying on judgment or models. If the selected factors or the fitted model differ significantly from the mean of the data, one must be sure there is a good reason

under pressure. On the other hand, if the comment was meant to say that he had just returned from vacation, and was thus in a good mood, then it is reminiscent of some of the literature on cognitive biases that has found that traders are more optimistic on sunny days, and that their trades reflect it. (See [5].) Perhaps the lesson is that actuaries should work in windowless offices.

for the discrepancy. There's an ever-present danger of getting so engrossed in the details of one's favorite approach that the big picture of fitting to data can be missed. A well-constructed graph will often visually draw one back to the underlying data, and provide some reasonableness to the work.

3. WHAT CONCLUSIONS CAN WE DRAW FROM THIS EXERCISE?

As we noted at the start of the paper, this exercise began with a hope that we could draw on the results that we received to provide some guidance to actuaries faced with selecting development factors from unstable loss triangles. This paper has turned out to be more of a description of the results we received, together with some commentary. The number of approaches varied widely – and so did the selected factors. It has proved difficult to find many general themes in the approaches; rather, it has provided much food for thought, and a guide to the many things that actuaries need to keep in mind when working with development factors.

We found that actuaries who used what sounded like similar methods when they were being described, sometimes ended up with results that proved to be significantly different. Conversely, much of the time, actuaries using what sounded like very different methods, came up with results that were very similar. Some actuaries saw patterns where others saw only noise, and there was a fair amount of disagreement as to what constituted reasonable or appropriate assumptions.

The guidance we hoped to provide is for another paper. Perhaps the best advice that we can provide for now is for actuaries to keep open minds, and to approach unstable triangles from a variety of perspectives. Having many tools in one's toolkit, and tempering all of these tools with good common sense and careful judgment, is more likely to give consistently reasonable answers over the variety of situations in which we find ourselves.

Acknowledgment

Many thanks to the 50+ participants for sharing their techniques and thoughts on this exercise. The authors wish to acknowledge David Homer and Joel Vaag for their insightful review of the paper and suggestions for improvements.

Disclaimer:

The opinions expressed in this paper are those of the authors, not of their employers.

Appendix A

Here is the e-mail that was sent to people attending the 2008 CAS Reinsurance Seminar:

From: Vincent Edwards [mailto:vedwards@casact.org]

Sent: Mon 5/12/2008 4:23 PM

To: cas@lists.casact.org

Subject: 2008 Reinsurance Seminar Research Opportunity! (File Attached)

Dear CAS Reinsurance Seminar Attendee,

The CAS Reinsurance Research Committee is conducting research to look at how practicing actuaries select loss development factors when dealing with unstable triangles. Some of the early results of this work will be presented at the "Loss Triangle Philosophy" session at the Reinsurance Seminar. To increase our sample size, we want to extend an invitation for attendees to participate in this study. In addition, session attendees may find it interesting to participate before they see what others have done.

Here's what we're trying to do:

One of the biggest challenges in day-to-day reinsurance actuarial work is selecting loss development factors when triangles are unstable. This topic does not receive significant attention in exams and papers, and yet it's something reinsurance actuaries do on a regular basis. Most actuaries learn how to select loss development factors on the job, pick up rules of thumb and helpful approaches along the way. Whether these approaches are good or bad probably depends on the particulars of the underlying data.

Taking a look at how different actuaries do things can be beneficial, even if it simply shows how many different approaches there are, and how widely the ultimate results vary. The attached Excel spreadsheet has an umbrella triangle (disguised in various ways to avoid divulging anything proprietary), but there is no special meaning to the choice, other than that it's unstable.

To participate, here are your instructions:

Select age-to-age factors from the triangle, and insert them in row 33 of the spreadsheet, where indicated. To keep the topic focused, we're ignoring the tail factor. Also, give a few sentences on how you selected your factors: all year averages? 5 year averages? Weighted or un-weighted? How many different averages did you look at before making your selection?

How did you deal with apparent outliers or with reversals in the nice smooth pattern you might have expected?

If you would like to participate and have your results included at the Seminar, please respond by emailing the spreadsheet to LDFs@comcast.net by May 17. Responses will be kept anonymous.

Thanks.

Gary Blumsohn

Chair

Committee on Reinsurance Research

The attached Excel file contained the following

Umbrella incurred loss triangle

Accident Year	1	2	3	4	5	6	7	8	9	10	11	12
1991	1,782	3,000	6,924	10,167	12,369	14,047	13,577	14,289	13,831	14,419	14,563	14,484
1992	430	2,814	3,557	5,745	9,033	7,884	8,715	8,982	9,048	8,934	8,856	
1993	2,234	3,902	10,841	14,262	17,666	19,154	19,411	19,021	18,854	19,085		
1994	3,335	12,937	23,694	20,477	19,715	23,689	23,955	25,066	25,269			
1995	2,006	5,406	9,802	8,949	10,611	10,623	16,633	16,699				
1996	7,640	8,485	12,085	13,515	15,418	18,894	19,029					
1997	6,643	13,184	18,530	17,782	20,867	21,358						
1998	2,474	9,684	10,636	16,266	16,649							
1999	4,229	6,135	5,972	8,613								
2000	2,065	2,982	3,384									
2001	3,448	4,240										
2002	1,736											

Age-to-age

1991	1.684	2.308	1.468	1.217	1.136	0.967	1.052	0.968	1.043	1.010	0.995
1992	6.544	1.264	1.615	1.572	0.873	1.105	1.031	1.007	0.987	0.991	
1993	1.747	2.778	1.316	1.239	1.084	1.013	0.980	0.991	1.012		
1994	3.879	1.831	0.864	0.963	1.202	1.011	1.046	1.008			
1995	2.695	1.813	0.913	1.186	1.001	1.566	1.004				
1996	1.111	1.424	1.118	1.141	1.225	1.007					
1997	1.985	1.405	0.960	1.173	1.024						
1998	3.914	1.098	1.529	1.024							
1999	1.451	0.973	1.442								
2000	1.444	1.135									
2001	1.230										

Insert your selected ATA factors below. Ignore the tail factor.

Selected ATA	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Appendix B

The table below shows the selected factors that the various participants provided, sorted in increasing order of the implied reserve from the factors.

		Selected age-to-age factors										
Reserve ranking	Implied reserve	1	2	3	4	5	6	7	8	9	10	11
1	10,707	1.375	1.213	1.163	1.113	1.050	1.010	1.007	1.005	1.003	1.001	1.000
2	18,323	1.921	1.300	1.142	1.124	1.112	1.021	1.013	1.008	1.005	1.003	1.002
3	18,512	1.626	1.213	1.173	1.113	1.103	1.043	1.013	1.002	1.017	1.003	1.000
4	18,561	1.921	1.157	1.214	1.081	1.085	1.050	1.021	1.002	1.017	1.000	1.000
5	19,512	1.499	1.138	1.194	1.113	1.094	1.082	1.020	1.002	1.010	1.002	1.000
6	21,028	2.000	1.250	1.150	1.110	1.090	1.070	1.021	1.010	1.005	1.002	1.001
7	21,163	1.921	1.250	1.142	1.113	1.112	1.093	1.021	1.000	1.000	1.000	1.000
8	21,304	1.758	1.250	1.142	1.113	1.094	1.075	1.021	1.013	1.007	1.003	1.000
9	21,807	1.371	1.063	1.214	1.113	1.085	1.120	1.013	1.008	1.017	1.003	1.000
10	21,958	2.000	1.250	1.210	1.100	1.100	1.075	1.020	1.005	1.005	1.003	1.000
11	22,827	1.921	1.340	1.142	1.112	1.093	1.060	1.021	1.014	1.008	1.005	1.003
12	23,398	2.000	1.400	1.150	1.120	1.100	1.080	1.020	1.005	1.003	1.002	1.001
13	23,703	2.005	1.383	1.142	1.134	1.112	1.093	1.021	1.000	1.000	1.000	1.000
14	23,729	2.014	1.535	1.249	1.163	1.089	1.034	1.023	0.994	1.014	1.001	0.995
15	24,433	1.550	1.250	1.182	1.163	1.100	1.049	1.030	1.018	1.011	1.007	1.004
16	24,636	1.962	1.380	1.138	1.111	1.090	1.065	1.030	1.015	1.010	1.005	1.002
17	24,706	2.005	1.538	1.135	1.142	1.094	1.075	1.021	0.995	1.017	1.003	0.995
18	24,706	2.005	1.538	1.135	1.142	1.094	1.075	1.021	0.995	1.017	1.003	0.995
19	24,835	1.866	1.343	1.232	1.146	1.062	1.093	1.021	1.002	1.014	1.003	1.000
20	25,002	1.921	1.250	1.142	1.081	1.112	1.093	1.021	1.000	1.017	1.010	1.005
21	25,056	1.921	1.250	1.142	1.081	1.112	1.050	1.025	1.020	1.015	1.010	1.010
22	25,439	2.153	1.443	1.227	1.127	1.074	1.043	1.025	1.014	1.008	1.004	1.005
23	25,619	1.816	1.457	1.140	1.112	1.104	1.094	1.020	1.014	1.008	1.002	1.000
24	25,856	2.500	1.535	1.195	1.165	1.090	1.035	1.023	0.995	1.014	1.000	1.000
25	26,235	2.000	1.400	1.400	1.180	1.120	1.010	1.020	0.995	1.012	1.010	0.995
26	26,247	2.005	1.538	1.247	1.142	1.094	1.050	1.021	1.010	1.005	1.003	1.000
27	26,730	2.513	1.538	1.188	1.090	1.094	1.075	1.011	1.005	1.007	1.005	1.002
28	26,880	2.000	1.250	1.200	1.140	1.100	1.100	1.025	1.010	1.010	1.005	1.000
29	27,137	2.250	1.500	1.250	1.200	1.100	1.050	1.025	1.000	1.000	1.000	1.000
30	27,468	1.900	1.350	1.200	1.115	1.100	1.120	1.020	1.005	1.015	1.001	1.000
31	27,599	2.150	1.525	1.200	1.150	1.100	1.050	1.025	1.015	1.005	1.003	1.002
32	28,712	2.005	1.538	1.135	1.142	1.094	1.075	1.021	1.018	1.017	1.003	1.002
33	29,107	1.626	1.213	1.173	1.142	1.094	1.075	1.044	1.021	1.016	1.012	1.009
34	29,444	2.000	1.450	1.200	1.135	1.100	1.075	1.025	1.015	1.010	1.008	1.005
35	29,528	2.179	1.515	1.197	1.156	1.094	1.075	1.021	1.017	1.010	1.003	1.000
36	29,805	1.934	1.482	1.158	1.141	1.095	1.073	1.026	1.021	1.017	1.006	1.004
37	29,837	2.033	1.500	1.256	1.138	1.078	1.047	1.029	1.019	1.013	1.009	1.006

		Selected age-to-age factors										
Reserve ranking	**Implied reserve**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
38	30,883	1.950	1.500	1.180	1.140	1.100	1.075	1.040	1.020	1.010	1.005	1.003
39	31,336	1.650	1.300	1.250	1.150	1.100	1.060	1.040	1.030	1.020	1.010	1.005
40	31,715	2.000	1.500	1.250	1.150	1.100	1.090	1.025	1.010	1.010	1.010	1.000
41	31,819	1.951	1.538	1.135	1.142	1.094	1.075	1.050	1.025	1.017	1.003	1.000
42	32,416	2.200	1.250	1.500	1.150	1.110	1.075	1.025	1.010	1.010	1.005	1.000
43	33,487	2.225	1.535	1.249	1.163	1.094	1.050	1.029	1.020	1.015	1.010	1.005
44	33,520	2.150	1.500	1.200	1.185	1.115	1.095	1.025	1.015	1.010	1.005	1.000
45	33,539	1.700	1.250	1.175	1.130	1.120	1.100	1.050	1.025	1.015	1.010	1.010
46	37,958	2.517	1.538	1.241	1.142	1.094	1.055	1.021	1.015	1.010	1.050	1.000
47	38,201	2.419	1.376	1.188	1.115	1.078	1.057	1.043	1.034	1.027	1.023	1.019
48	38,973	2.517	1.603	1.247	1.189	1.112	1.112	1.023	1.014	1.014	1.001	1.000
49	39,156	2.000	1.400	1.350	1.175	1.150	1.100	1.030	1.025	1.020	1.000	1.000
50	55,197	1.641	1.297	1.187	1.134	1.103	1.083	1.069	1.058	1.051	1.045	1.040
51	60,152	2.788	1.378	1.656	1.336	1.130	1.070	1.040	1.020	1.020	1.020	1.010

The next table shows an edited version of the narratives from the respondents, sorted in the same order as the table above. We have tried to edit lightly when the narrative wasn't clear. We have also edited spelling and egregious grammatical errors, as well as editing to retain the anonymity of the participants. While we have done our best to avoid changing the meaning of what was intended by the participants, it is quite possible that we have misunderstood something and have misrepresented some of what was intended. We apologize for any such errors. We also freely admit that we do not fully understand all of the explanations. In general, participants dashed off a quick few sentences – rather than a detailed documentation of what they had done. Where the explanation is blank, the respondent didn't provide one.

1	I generally prefer weighted averages, 3-5 years for incurred and 5-7 for paid, but since I just came back from vacation and will soon be on the road again I did mostly a straight average analysis. With an unstable pattern, one technique I frequently use is a 5 or 7 year average excluding hi-lo. For year 1, I did a 3 year average, seeing that the numbers were reasonably close and showing a downward trend, possibly suggesting that claims are being settled more quickly. Year 2, 4: 5 year ex-hi/lo Year 3: No clue what to do here, just made it halfway between 2 and 4. As I typically do in later years where the observations are less credible, I just choose round numbers that slowly go down to 1. I've seen others use something like a Bondi curve for similar purposes. At the end of the day, everything we do in reserving is "wrong" according to management; reserving actuaries take the hit when adverse development rears its ugly head. That's probably why it receives so little attention on the syllabus.
2	
3	For 12-24 and 24-36 months, average ATA factors are lower in the more recent years (1996-2001). Considering this, I selected 5-point ex hi-lo averages for 12-24 through 72-84. For 84-96 I selected a 3-year weighted average; beyond 96 I selected all year weighted averages. I looked at simple and weighted averages for 1 to 7 years plus all years. I also looked at 12 to 36 months, 36 to 84 months, and 84 months to current as reasonability checks of my selections.
4	Apparently something going on in the latest diagonal. Hard to tell without paid triangle or claims audit report. Would like to use some sort of industry default as a test to see how the pattern here compares. No idea as to how exposures have increased/decrease => can't tell whether the 2002 year is uncharacteristically low, as expected, or higher than expected. I went with the standard 3/5/All Year Averages. I'm not convinced that alternative methods for selecting LDFs do a better job than standard averages of capturing the full gamut of frequency/severity scenarios which could impact loss development. Perhaps this is just ignorance on my part.
5	For the most part I rely upon an average of 3-year loss-weighted and 5-year ex-hi-lo, with some regression for factors in the tail. I might be inclined to add part of a cumulative standard deviation for a volatile line such as umbrella.
6	
7	
8	My first pass was to select the median of the 3, 5, 7 year weighted averages. I smoothed the tail, using 7-8 and 10-11 as anchors. Then I graphed the individual ATA factors and tweaked/smoothed the 6-7 selection. Then I graphed the whole range and tweaked 7-8 from median to higher 5-weighted average.
9	

10	
11	Generally used the 5 year weighted average. Exception 2-3 ATA factor is the average of the 5-year and 7-year. The 5-year point on its own was a bit low and uncomfortable so it was judgmentally increased. The selections were smoothed to decrease with maturity. The 5-year 4-5 ATA was reduced from 1.081 to 1.060 --- the difference was used with the more mature points (e.g., 8-9 and 11-12) for additional smoothing. In situations with unstable data, we like using dollar weighted averages versus straight averages so years with small volume/high development do not receive undo weight. And we're more likely to use more years rather than fewer years. With more stable data we're more likely to rely on shorter/more recent averages; and also use straight averages and not weighted averages. Like to use more recent information rather then older diagonals. Legal and legislative trends or changes in company claim processing. Company maybe starting out and the earlier stuff they are "just cutting their teeth." In fact when there is a lot of data, we will show/use the latest 10 or 12 diagonals and not use the diagonals prior to that. I do realize this point contradicts the prior point -- I guess there is a balance that needs to be struck --- and in fact we like to have two selections one based on a longer term average and one based on shorter term averages. In this case we went with mid- to longer-term averages given the "instability" of the data. We like to see decreasing factors --- so we smooth. Will borrow from one ATA to smooth another ATA. Would like more information: Have the retentions and limits been changing over time? Can we get the individual claims at each valuation date? If so we would trend losses (perhaps even retentions and limits) to current date and then recast development triangle. Are there any changes in payments or case reserving philosophy? Important to look at the analysis every year (or more frequently) and adjust selections with latest information. Test prior assumptions and adjust when necessary. This point is the most important point here --- perhaps even more important than the ATA selections themselves (assuming that they were selected within reason).

12	I dislike using straight averages to make LDF selections for umbrella, or excess in general. I calculated an all-year and a 5-year weighted average, and made selections for smoothing purposes, trying to ensure my selections kept the cumulative factors somewhat consistent with the experience. If benchmark data had been made available, I would have looked at that as well. Also, I spent very little time with this. True it's just a sample, but I try not to overanalyze loss triangles. I will never know the "true" LDFs (assuming they even exist) regardless of the amount of time I spend analyzing the data, and I don't want to fool myself that my work is more predictive than it really is.
13	Early years: I like all-year weighted averages for highly variable books like umbrella. In this triangle, early years straight average > weighted average, implies larger years develop less. I would like to see premium normally, as well as previous selections and actual vs. expected to see how well previous selections are holding up. I also look at the year of the development - for example, I will put more or less weight on a year if it's clearly a soft market year whose development won't be repeated in the hard market. In this example, it really looks like a lot of losses got put up, then brought down, then developed upward. I tried to pick factors for 2:5 whose product was generally close to the weighted average development I saw on 94-98 for 2:5 - though I built in some conservatism. One other comment: This looks pretty short tailed for umbrella business. Is it personal umbrella? Based on what I see here, it's hard to justify development beyond 8 years. With more time, I'd poke into the whys and wherefores around that.
14	For the first evaluation, I excluded what I considered an outlier, the AY 1992 point and then took the average of the remaining ratios, excluding the remaining high and low values. For the next 5 evaluations, I took the all point average, excluding the high and low points. For the last 5 evaluations, the ratios were more stable and I used an all point average.
15	This is a formulaic approach giving some extra weight to more recent experience. Some judgmental smoothing is then applied.
16	Looked at various averages

17	I am basically selecting the "volume weighted all" factors. The logic behind the selection is based on the recent paper http://www.casact.org/pubs/forum/08fforum/1Bardis_Majidi_Murphy.pdf. The main idea is to build a statistical framework which would help test the underlying assumptions of the chain-ladder method. Given a set of selected development factors you can find a linear regression (with "good" assumptions about the variance of its error terms) that produces as best linear unbiased estimators (BLUE) the selected development factors. The significance of the previous statement is that a practitioner can use the robust statistical regression framework to check the reasonability of his/her selections. So residuals in the AY/DY/CY dimensions will provide the underlying trends and normality tests and the AIC/BIC fitting criteria will provide a statistical evidence of how well the chain ladder method "fits" the historical data. The important point is that all these statistics are within the confines of the chain ladder method that practitioners are comfortable with. I looked at the volume weighted, straight average (all vs. 5 years) and judgmental selections (with and without outliers) and the visual inspection suggested that the "volume weighted all" model performs as well as anybody. This should not come as a total surprise given the change in the volume of losses by accident year. Surprisingly also, it does not produce any outliers (within 1.5 interquartile distance) and the normality graph is pretty "tight" around the 45 degrees line. The data exhibits some decreasing trends on the accident year dimension which could suggest that an exposure adjustment is in order. Both the calendar year trend and standardized residuals vs. fitted values suggests that the selections understate the low historical losses (not surprisingly since some of the low volume years exhibit higher development than the average).
18	For the age to age factors I selected the all year weighted averages since the individual ATA factors are volatile. I did not make adjustments for reversals (ages 8 to 9 and 11 and 12) since they were minimal. If the AGA factor in that case was 0.800 instead of 0.995, I would make an adjustment.
19	I will often look at the triangle and if there appear to be some big takedowns following big increases, I will "smooth out" the triangle by removing the reversals. Then I'll tend to look at the last 3, 5 and all years both weighted and unweighted.

20	I generally went with 5 year weighted averages, until the factors close to the tail where I ignored the one age with negative development and selected a small positive factor. I looked at 3, 5, and all-year weighted and simple and hi-lo out averages. In real life I would look at the paid triangles to see what kind of ultimates that was producing as compared to incurred. Also might look at some diagnostics such as change in average claim size, etc if I was feeling really crazy. I would also get industry development factors. I also noticed that for the first 2 ages, the recent diagonals appear to be lower than the older ones - would have to investigate if this was random or a trend. Since I went with 5 year averages, I am giving it some, but not full credit, as my selections are still higher than the recent 3 diagonals for these ages. This isn't the case for the later evaluations - recent diagonals are not generally lower. For the 91 year it looks like the 0.967 and 0.968 get 'erased' by the factor after them, reserve takedown gets put back up the next year, so I judgmentally selected 1.0 for the 8-9 year even though the 5 year avg was <1, although I didn't completely throw out that year.
21	
22	My usual technique is to compare Weighted-All with Weighted for a certain number of years. I believe that straight Average is too subject to biases caused by years with very little loss reported. Also, I believe that removing the Min-Max is inherently biased since LDFs are capped from below by 0, but uncapped from above, so removing the Max will tend to have a larger effect than removing the min. I often use the Brosius Linear method as a "sanity check", but in this case it is prone to distortion for two reasons: It is most accurate when applied to trended loss ratios to on-leveled earned premium to best remove effects of loss trends and changes in exposures, and the linear statistical model assumes that variations from the "best fit" are due to random noise (independent, identically normally distributed for that matter) that is not always the case. For the purposes of this exercise, I also ran Markus Gesman's package in 'R' to calculate the bootstrap ultimates based on England and Verrall. However, this too is based on a generalized linear model that less accounts for changes in underlying factors as opposed to finding the best generalized linear fit to the data supplied. I will often choose a final "raw" selection and then fit a smoothed weibull and lognormal to that data. If the fitted curve fits he raw data "well," then the argument can be made that the fluctuations are the "noise" and the fit is the signal. Certain lines of business, however, which are known to over-reserve and then take down, are not suited to this kind of smoothing. In this case, the distribution of each age's ATA seems random, and the all-year weighted is a decent selection for each age. It also is a good candidate for smoothing, so the final selected ATA's would be the weibull-smoothed version of the weighted all-year ATU.

23	I calculated weighted 3, 5, and all-year averages; simple 3, 5, and all-year averages; and 5, 7 excluding hi-lo averages. I calculated the average of averages. I also calculate the median of these averages. Then I ran regression on losses at evaluation 1 through 6, assuming zero intercept. The final selection is basically a weighted average of these three methods for ATA from 1 to 5, and using average of averages for the rest, with adjustments and linear interpolation to smooth out the LDF curve.
24	Looked at volume weighted averages 3 years, 5 years and all years. Then simple averages 3 years, 5 years and all years. Finally simple average excluding the highest and lowest data points. In the end I selected RTRs according to the following: 1-2: simple average all years 2-3: simple average ex high/low 3-4: simple average 5 years 4-5: simple average ex high/low 5-6: simple average ex high/low 6-7: simple average ex high/low 7-8: simple average all 8-9: simple average all 9-10: simple average all 10-11: simple average all 11-12: simple average all I ended up not focusing on the volume weighted averages. Generally I would use these when information is available on premium volume by year, or number of claims by year. Large dollars of loss do not necessarily increase predictive power, as this might be the result of one or two unusually large claims, the development of which might not be representative of the average.

25	Assumption: I assumed the policy count and class/limit profile of the book remained relatively stable during over the history provided. Methods Employed: 1) Evaluated magnitude of losses for each year at each maturity and ranked as "H" high, "M" medium and "L" low. With low volume data I've found that the link ratio methods are very sensitive to the actual magnitude of losses at a point in time. Determined separate average link ratios for L, M, and H. 2) All years weighted average 3) All years straight average 4) All years straight average excluding minimum and maximum 5) All years weighted average excluding 1992. After review the 1992 year seemed very different from the others due to the low magnitude of losses and the up and down development pattern. Selection for older years looked primarily at the all year weighted average excluding 1992. Selection for more recent years looked at "L, M, H" averages based upon the magnitude of losses along the diagonal. Other data would have been helpful: Premium / Policy Counts for each calendar/accident year and perhaps a limit profile Paid closed claim count triangle Open claim count triangle Separate paid and outstanding loss triangles Any kind of class of business distribution (SIC, NAICS, GL Class Code, CMP Program)
26	

27	1st Pass 2.005 1.538 1.135 1.142 1.094 1.075 1.021 0.995 1.017 1.003 0.995 1.140. All-years weighted averages were selected because they behaved reasonably in a 48 month model of early development which attempts to spot unreasonable relationships between the early ATA factors. The 3 & 5 year weighteds did not pass this test. On volatile triangles like this, I also want to use as many years as I can. I never look at simple averages of ATA factors – only weighted. 2nd Pass 2.005 1.538 1.188 1.090 1.094 1.075 1.011 1.005 1.007 1.005 1.002 1.140. The modest reversals were eliminated using exponential smoothing. I generally do this as I find reversals to be abhorrent. When possible, I develop gross losses and salvage and subrogation separately in order to avoid analyzing downward development. The fact that I left 1 small reversal shows that I am not a fanatic! 3rd Pass 2.513 1.538 1.188 1.090 1.094 1.075 1.011 1.005 1.007 1.005 1.002 1.140. The last revision of the 12-24 ATA factor is based on the early development model. As a last comment, when I develop losses I want both Paid and Reported as I develop them together. This helps dramatically in estimating tail factors and the general quality of the data. I often hear, especially regarding volatile groups, that the paid data is just not useful. When developed on its own that may be true but when developed alongside the reported losses, I usually find it very helpful in making judgment calls at the very least.
28	I used simple and weighted averages for all years, most recent 3, and most recent 5 as summary references. While I see some volatility in the age-age factors, I didn't see many that were so freakish as to cause me to censor them out. I generally do not select to allow reversals unless the evidence is clear, and that they aren't re-reversed back later.
29	Generally I look at the average ex-hi/lo and the weighted average, and pick something close. I prefer to pick round numbers, as they won't fluctuate as much from year to year. In certain cases I will ignore the weighted average if one observation is influencing it too much (e.g. the 0.963 4-5 factor in 1994). I selected 1.00 for the last 4 factors, as I would normally cover these developments in a tail factor.
30	I considered all averages (weighted and unweighted) that end with the factor in the last diagonal. No ex-ante averages were considered. Most selections rely on the long term average (all-years or close to all-years) with tempered movements in the direction of the latest calendar years' link ratios. There is no smoothing between consecutive link ratios. Smoothing seems appropriate when there is an outlier. With erratic data, it's not too clear what points are outliers and to what degree they are outliers.
31	I selected mostly the volume weighted ex hi/lo with some judgment applied.
32	Weighted average all years, with a bias against factors < 1. No obvious calendar-year correlation effects so Mack gives the answer. I am not a reserving actuary and I don't get called upon to pick factors very often (I can recall no instances in the last three years, in fact) so you should underweight my choices.

33	As part of the selection process, I looked at the all-year weighted averages, the weighted averages of the last 5 years, of the last 4 years and the last 3 years, as well as the mean of the last 5 excluding high and low. For the 1-2 ATA factor, there appears to be a downward trend, but the 2001 factor seems to be very low. Rather than allowing that factor or the 1998 factor (which is very high) unduly influence the selection, I selected the mean of the last 5 excluding the lowest (2001) and highest (1998). I selected the next 3 ATA's in the same manner (i.e. last 5 ex high-low).For the 5-6 and 6-7 ATA's, the all-year weighted averages appeared to be reasonable. After that age, the limited number of data points were indicating even more erratic patterns. The 7-12 ATA is indicated as 1.031, whereas an indicated decay rate off the 6-7 ATA of 1.075 would generate 1.056. I decided to weight 50-50 the ATA's indicated by the data with the ATA's generated by a decay rate. In general, I look to select ATA's that are monotonically decreasing.
34	I also would include information that I have regarding the industry. It was not clear if the umbrella was supported or unsupported, but looking at the information I would guess supported. My selections are based on looking at all the averages and my general expectations regarding the line of business. I will typically select a smooth pattern, with the overall pattern based on the overall data.
35	I looked at 5 different averages: weighted and unweighted and for all years and 5 years, as well as weighted average ex-hi-lo. I did not calculate an average for less than 5 years, since the line is umbrella and due to the long tailed nature with volatility more experience is best. I prefer weighted averages as opposed to straight because it smoothes out the data. The weighted average of all and excluding high and low are very close. This gave me comfort in the selections, but I am still skeptical about the difference in the older years prior to 1996. I selected the weighted average ex hi-lo for 12-60 months and then went to all year weighted average for 72 and 84 months. If you kick out the high and low in the 72 and 84 months, you are only left with 3 and 4 factors, so I went with the all year weighted average for these selections. I smoothed out the 96 ATA factor and put the 108 factor in its place. For the 108 factor, I averaged the 108 and 120 factor. For 132, I assumed one based on the data, but given this is umbrella there will be a tail factor. I still am not sure about the data, since it does seem like it changed starting in 1996. The factors starting dropping off dramatically in 1996. There also seem to be reserve takedowns at 12/31/96 for the older years. It also seems like they are setting higher initial reserves at 12 months starting for 1996 and 1997. They strengthened reserves at 12/31/00 for '94-'97, but not so high as to match the cumulative products for the early 1990's. The danger of my selections would be if reserves are weaker now, since my selected factors more reflect the more recent years because I selected the weighted average and the hi-lo which often kicks out the older years. I would want to see more info such as: large loss listing, paid data, claim counts, premium volume, mix of business changes.

36	I'm not sure whether you were specifically looking for implicit or explicit inclusion of some risk margin in the selections (to anticipate for when things invariably go wrong, e.g., for reserving) or whether you were looking for "best estimate," without regard to a margin. I took the latter approach. In answer to your additional information request, I used the following 4 averages: Unweighted all year and weighted all year, 6 year and 3 year. In my selection, I assumed that weighted averages are more representative than unweighted. That is, I assumed more volume in one AY would indicate more credibility than an AY with less volume. This assumption would not be true if larger AY volumes were only due to larger shares in that AY, e.g. larger reinsurance shares but same underlying book. I gave a majority of the weight to the all year volume-weighted average. I did not include a tail factor, per your instructions. To help smooth selections and adjust for reversals, I selected an "industry" pattern based upon excess layer, LOB, and lagging assumptions for this umbrella book. To do this right, would naturally need more information on the underlying book, but nonetheless I took a stab at it. I first scaled the industry factors based on the volume weighted averages underlying my assumptions above, and comparing actual vs. expected development at each maturity. I then generated a scaled industry factor set, in this case using a factor of 0.9. For purposes of illustration for this survey, I then credibility weighted this scaled industry factor set using a claim count based formulaic approach and an estimate of excess claim counts underlying the umbrella triangle. Of course, would ideally have all this information including all individual claims and their histories in the data and not have to make these assumptions. The credibility formula uses actual claim counts (vs. e.g. expected claim counts using premiums, expected loss ratios, claim count reporting patterns, etc.) and a $z=n/(n+K)$ form. I adjusted the K from a default based upon the extra variability in the LDFs. These selected factors were slightly overridden to produce the ATA factors that I put into your spreadsheet. I did not include a specific risk margin in the selections, for the "things that invariably go wrong." To do this, I might not have scaled back the industry factors, nor included weights for the more recent (in general lower) averages. May also have selected factors higher than the averages and of course also included a tail factor.
37	(1) I assumed the data is on-level with a constant level of exposure. (2) I fit an inverse power function utilizing David Clark's maximum likelihood approach. http://www.casact.org/pubs/forum/03fforum/03ff041.pdf (3) I examined the total variance (parameter and process risk) for the two year weighted average, three year weighted average, etc… and selected the weighting with the lowest variance. This happened to be an all year weighted average. (4) I did not exclude any "outliers."

38	Attached are my picks. However, let me add that given paid losses and claim counts, I might have done something completely different. Due to the volatility, I tended to use the long term weighted averages as my guide but looked at all years, last five years and last three years weighted averages. I also did some smoothing of some of the points by moving dollars between generally contiguous evaluations (but not necessarily contiguous), recomputing LDFs and looking at the same statistics. After selecting, I compared the product of my selections through the first 7 points to the product of the all weighted average through the first 7 points and adjusted my selections slightly. The points after 7 were based on a judgmentally selected decay in the first 7 selections and the same set of six statistics noted above. The latter LDFs were selected more conservatively given it is umbrella (although did not have attachment point and limits) and given additional information on paid losses and open claims would probably have selected differently. I also compared my results to the fitted values from inverse power, exponential and Weibull curves. This has not been subjected to our quality control process and should be considered my thoughts.
39	I looked at 3 and 4 year averages, both weighted and unweighted, the 5 year excluding the high and low, the median, and an all year simple average. I tend to ignore outliers and reversals in the pattern. Loss data that comprises few losses, driven by severity rather than frequency, which this appears to be, cannot be relied upon for reasonable patterns – smoothing helps this out.
40	I mainly looked at weighted averages and also looked at how close these methods were. I may also give some credibility to an industry or default pattern.
41	In general, I used an all-year weighted average, but also looked at 3, 5, and 8 year averages. For unstable triangles, I tend to use as many years as I can. For the 1-2 ATA factor, I excluded the 6.544 because it was such an outlier, although it had little impact since the dollar amount was small. Beginning with the 7-8 ATA factor, I started smoothing the pattern out judgmentally - no real scientific reason as to why, just because there were so few points.
42	
43	The selections primarily use all year link ratios excluding high and low. Links are further smoothed.
44	I looked at straight, weighted, truncated, and geometric averages over a wide range of time frames (from 1 to 11 years) to see what type of pattern emerged. Without knowing the underlying business, it's tough to select factors. The use of the geometric average gave a quick look at the impact of outliers. For those older ages where it appeared to reverse, I judgmentally smoothed out the pattern.
45	

46	In the absence of an obvious pattern in any of the averages, will initially select overall weighted average, with the following exceptions: Age 1 - weighted average looks more reasonable relative to Age 2 Age 3 - weighted 75%/25% with Ages 4 & 2 to smooth the reversal (those are assumed more reasonable relative to industry/company default patterns, so Age 3 is replaced). Also concern over several negative points in the data. Age 6 - appears to be driven up by an outlier, however if I exclude the outlier it looks artificially low, so I gave some credit off the weighted average with an eye on the overall progression. Ages 5 & 7 appear reasonable relative to industry/company default XS pattern. (Age 5 highs are balanced by a low "outlier.") Ages 8-10: set to industry/company default (#'s changed slightly and rounded for proprietary reasons).
47	I looked at five weighted averages, but based my selections on an inverse power fit of the entire triangle of incremental reported losses, using the method described in Dave Clark's 2003 CAS Forum paper on "LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach." I also considered fitting only the last few diagonals, but ultimately decided that it made sense to fit all the data. I ignored loss trend in this example. Graphs of the factors (actual all years versus fitted all years, columns of actuals by development period compared to fitted, standardized residuals) increased my confidence that the maximum-likelihood approach was doing a good job fitting the entire triangle with all its volatility.
48	I used unweighted and no less than following age-to-age. For this case, it looks like better experience in recent years. If I find that there is a solid reason for the better experience, I will apply an adjustment factor for all ATA factors.
49	I didn't look at any averages formally. I eye-balled each column and struck repeated compromises between what the column suggested and my feeling that the age-to-age factors should decrease monotonically.
50	I model a distribution of the "time until reported" random variable. I then fit the incremental amounts to what the distribution would imply as a fraction of the cumulative amount for the accident period. I then inspect the errors for consistent bias and make adjustments if appropriate. I find this method to deal well with volatile development data. Because it is looking across the entire curve for the parameterization, the impact of an outlier is generally smoothed across development periods. In this example, my factors are larger for the later development periods than what history would indicate. However with a 6-7 factor on the 2nd most recent diagonal of 1.5, clearly the potential for late period development is there, and I feel this should impact the selection for other periods. This method is only appropriate when negative incremental amounts are observed sporadically, not consistently. I prefer to use a method that models case reserve balances and incremental paid losses within a single model framework, which therefore does not suffer this problem.

51	I looked at two kinds of averages, weighted averages and simple averages. If I had some extra knowledge on this book of business, I might have done some time weighted averages. I usually prefer the weighted average to simple average because the simple average has inherent bias. However, I didn't use either of them for LDFs of age 1-4. I found them to be disturbingly low and feel uncomfortable to go with any of them (5-year weighted average etc). I found the age 1-4 LDFs have a wider range then those of age 5 and on. I also noticed that there is a negative correlation between the LDF factors and reported amount at the same age. This is very typical for LOBs of high variance. For age 1-4, I used a linear regression to predict the LDF factors. For example: 2002 factor age 1 factor (3.14) was calculated using age one reported amounts and age 1-2 ATA factors. 2002 age 2 factor (1.65) was calculated using age two reported amounts and age 2-3 ATA factors. For age 5 and on, I judgmentally picked LDF with the consideration of weighted averages, simple averages, smoothness, conservatism, and historical LDF range. I did this for each year and each age. I did the age to ultimate factors by using ultimate divided by reported. ATA factors were derived from age to ultimate factors. There is a bump in the ATA curve, which is mostly caused by the low reported amount of year 2000.
52	*Authors' note: This participant didn't provide selected factors, and was thus not included in the earlier tables. However, we felt that the thought process was worth including.* Being the pain in the (you know where) actuary I am, I would first ask the underwriters/claim handlers a series of questions (either my own or via an underwriting/claim audit): What are the attachment and limit profiles of the underlying umbrella business and how have these changed over time? Is the data net or gross of reinsurance and how has the company's retention changed over time? What is the company's case reserving policy and how has this policy changed over time? What is the nature of the underlying umbrella business (i.e. commercial or personal)? Absent any answers to my questions, then I would probably take a series of averages (3 to 5 year simple average, 3 to 5 year weighted average, 5 year excluding high & low simple average). I would then probably take an average of the averages (yes I've done this) and then make a selection. In other words smooth the data to the point where a selection is easier and more formulaic and less arbitrary. While making selections, I would also note where in the underwriting cycle I am (helps to select a calibration period for the averages) - or perhaps use loss ratio benchmarks to see if ultimate loss projections take sense.

5. REFERENCES

[1] Ayres, Ian, *Super Crunchers: Why Thinking-by-Numbers Is the New Way To Be Smart*, 2007, Bantam.
[2] CAS Task Force for Enhancing the Reputation of Casualty Actuaries, Unpublished report, 2008, presented to the CAS Board in March 2008 and distributed to CAS Research Committees in April 2008.
[3] England, Peter and Richard Verrall, "Analytic and Bootstrap Estimates of Prediction Errors in Claim Reserving," *Insurance: Mathematics & Economics* 25, 1999, pp. 281.
[4] Fleming, Kirk G., "Yep, We're Skewed," *Variance* 2:2, 2008, pp. 179-183.
[5] Hirshleifer, David A. and Tyler Shumway, "Good Day Sunshine: Stock Returns and the Weather" *Dice Center Working Paper No. 2001-3*, 2001. Available at SSRN: http://ssrn.com/abstract=265674 or DOI: 10.2139/ssrn.265674.)
[6] Hodes, Douglas, Sholom Feldblum, and Gary Blumsohn, "Workers Compensation Reserve Uncertainty," *PCAS* LXXXVI, 1999, pp. 263-392.
[7] Mack, Thomas, "Measuring the Variability of Chain Ladder Reserve Estimates," *CAS Forum*, Spring 1994, Vol. 1, p. 247.
[8] Murphy, Daniel M., "Unbiased Loss Development Factors," *CAS Forum*, Spring 1994, Vol. 1, p. 183.
[9] Struzzieri, Paul J. and Paul R. Hussian, "Using Best Practices to Determine a Best Reserve Estimate," *CAS Forum*, Fall 1998, pp. 353-413
[10] Thaler, Richard H., *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, 1994, Princeton University Press.