Casualty Actuarial Society
E-Forum, Spring 2011

The CAS *E-Forum,* Spring 2011

The Spring 2011 Edition of the CAS *E-Forum* is a cooperative effort between the Committee on Reinsurance Research and the CAS *E-Forum* and Committee.

This *E-Forum* includes five papers submitted in response to call for papers on reinsurance research. Some of these papers will be presented at the 2011 CAS Reinsurance Seminar, held in Philadelphia on June 5-7, 2011. This *E-Forum* also includes an additional paper.

CAS *E-Forum,* Spring 2011

Table of Contents

Reinsurance Research Call Papers

Additional Paper

E-Forum Committee

Windrie Wong, *Chairperson*

Mark A. Florenz
Karl Goring
Dennis L. Lange
Elizabeth A. Smith, *Staff Liaison*
John Sopkowicz
Zongli Sun
Yingjie Zhang

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

Some Considerations With Regard To Inflation

Matthew Ball, FIA, and Andy Staudt, FCAS, MAAA

Abstract

The following paper presents and uses a simplified framework to explore the impact of inflation across various aspects of loss reserving, pricing, and capital management. The primary intent is to highlight some general principles which can be used to understand where, how, and by how much inflation risk may affect various aspects of actuarial modeling; but its intent is also to encourage actuaries of the importance in adequately reflecting future expectations of inflation in their models.

Keywords. Inflation; insurance; reinsurance; collective risk model; reserving; ratemaking; capital modeling.

1. INTRODUCTION

Inflation has become a problem again. Relatively stable since the mid-90s, 2009 was officially deflationary for the first time in many countries' post-war histories and future trends are really anybody's guess with fears of hyper-inflation vs. deflation sharply divided across political lines. For hyper-inflation, we have the amount of credit many governments are pumping into the private sector, exchange rates of purchasing economies falling relative to producing economies, and investors hedging bets by purchasing large amounts of more traditional commodities. Also interesting is the theory that rising commodity prices in local currencies are contributing to the 2011 protests in the Near and Middle East as countries inflate their currencies in an attempt to maintain their pegs to the U.S. dollar. For deflation, we have high unemployment which should imply a lower than normal level of wage inflation (even wage deflation), an excess labor supply, property devaluation, decreased expenditure on "luxury" goods, and austerity. And this is just with regards to underlying index inflation—for insurance, material shifts in super-imposed claims inflation (judicial, social, labor, and otherwise) are already contributing to downward pressure on profits in many markets. Not to mention increased utilization of benefits as a result of a weak economy and all its associated problems. While these changing trends can wreak havoc on first-dollar insurance products, the effects can and have historically been catastrophic in higher layers where even the most subtle of shifts may be exacerbated. Given the increased importance of including inflation in actuarial models, the purpose of this paper is to explore and comment on the effect inflation can have on several key aspects of actuarial practice including loss development, dependence, credibility, decreased limits factors, loss distributions, and risk margins. Rather than merely considering the well-documented leveraged impact of inflation, our intention is to present practical results which will help actuaries understand where, how, and by how much inflation could affect their work and, as such, to stress the importance of modifying models to allow for inflation.

2. EXECUTIVE SUMMARY

The purpose of this paper is to explore the effect of inflation across several areas of actuarial practice so as to provide readers with the insight necessary to appropriately understand some of the general effects unexpected future inflation may have on different insured layers. Although the exact effect is totally dependent on each (re)insurer's unique situation—and as such the following results should be considered general under "nice" conditions and not universal in all, "nice" as well as "not-nice," circumstances—we will show the following broad rules-of-thumb:

- ***Leveraged impact of inflation***. The attachment point is the most important variable in determining the leveraged impact of inflation on reinsured layers (i.e., increasing the attachment will increase the impact of inflation) with the limit tempering the effect and co-insurance having no impact.
- ***Decreased limit factors***. Although the amount of experience in excess relative to ground-up or lower layers will typically increase quite quickly with inflation, the uncertainty around such relativities will typically decrease.
- ***Loss development.*** Inflation in excess of that observed historically slows down loss development; conversely, deflation speeds up development. This effect is exacerbated for higher layers of insurance, longer-tailed lines and lines which are characterized by single lump-sum payments rather than those characterized by periodic payments.
- ***Dependence.*** Volatile general inflation affecting all lines increases the dependence between risks whereas volatile super-imposed inflation affecting a single line decreases the dependence between lines.
- ***Loss distributions.*** Although it is obvious that inflation will change the distribution of losses in the insured layers, the key result is that for layers with a reasonable enough amount of experience, increasing the inflation rate will decrease the volatility of loss experience in that layer *relative to the mean*—the implication of this finding is explored in other sections.
- ***Credibility.*** Given a sufficient amount of historical experience by insured layer, inflation should increase the credibility of historical experience.
- ***Risk Margins.*** As inflation increases, the relative risk margin decreases across most insured layers suggesting that risk margins for excess layers are relatively less affected by inflation than the best estimate of total loss or outstanding reserves.

3. METHODOLOGY

3.1 Model

The following results are based on simulation using extensions of the collective risk model. See Klugman et al. [6] for a basic description of the collective risk model. Essentially, for each trial and accident year, we simulated the number of claims based on an assumed frequency distribution. For each claim, we simulated the ultimate loss amount based on an assumed severity distribution. We paid that claim out based on assumed loss development patterns and knowledge of how the claims for certain lines pay (e.g., for liability lines, we primarily relied on lump-sum settlements; for the indemnity portion of workers compensation, we primarily relied on steady payments and so forth). See Butsic [2] for example. We then applied calendar-year inflation trends to the incremental payments. We considered both deterministic as well as stochastic inflation to assess the impact of increasing amounts of inflation as well as increasing uncertainty in future inflation, respectively. As a result, many of the results are delineated as being from the "deterministic scenario" or from the "stochastic scenario."

The above is easily enough programmed into most computer languages and the Casualty Actuarial Society's (CAS) Public Loss Simulation Model contains much of the functionality required to explore these results further. In many situations, especially when trying to get reasonable parameter estimates in higher layers, this process can become quite time-consuming and the importance of efficient simulation and more complex techniques such as stratified sampling become necessary.

3.2 Parameterization

We present most of our results in terms of the following broad excess of loss (XOL) layers—ground-up, lower, working, and excess. To make the examples as consistent as possible, we set the lower layer, or retention, at a level where approximately 90% of the loss prior to inflation would fall. Similarly, we parameterized the working layer with the next 7.5% of the loss and then the excess layer with the final 2.5% of the loss. We also used several other splits, such as 75/15/10 and 95/3/2, to sensitivity test our results. Further, we parameterized our results using industry benchmark data and sensitivity tested the results using varying assumptions and parameters across several different lines of insurance.

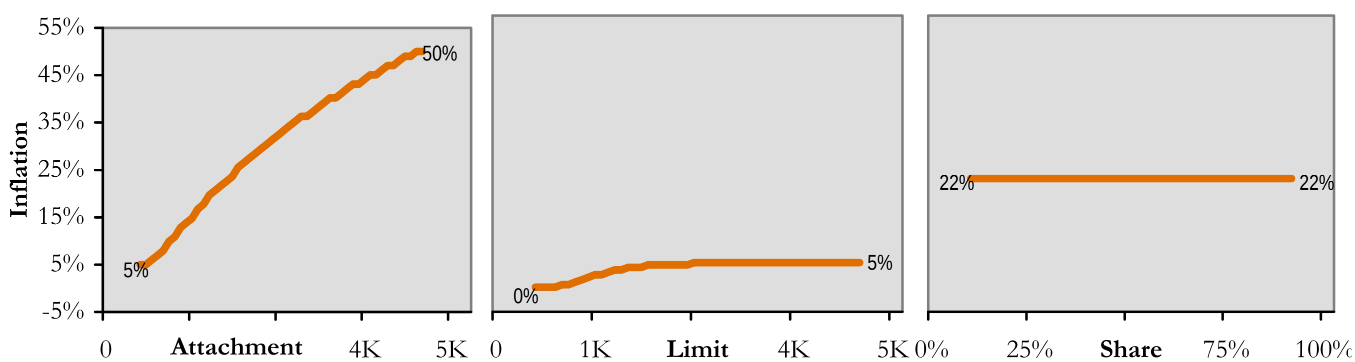
3.3 Caveats

We note that there are several caveats to our work. First, although we did attempt to sensitivity test our results to various parameterizations, lines of business, and reinsurance contracts, the infinite permutations make it impossible to present truly universal results using this type of empirical analysis. Therefore, it is important that these results only be considered general and used for reasonability checks or as an aid in assumption setting when the exact effect in the actuary's unique situation can not be determined. Further, we note that in addition to the above, our analysis tended to rely on "nice" situations—medium-sized books of business with homogenous claim profiles coupled with reasonable reinsurance contracts. It will often be the case that reinsurance contracts do not satisfy these nicety conditions and "kinks" will arise in the results. This further indicates the importance of modeling each unique situation. As a continuation of the above, we tended to use ”nice" continuous loss distributions without consideration of binary or CAT-type events which may further distort the results. Although, in any event, the effect of inflation on these types of events should be analyzed by event scenario rather than in aggregate.

4. DETAILED FINDINGS AND CONCLUSIONS

4.1 Leveraged impact of inflation

The leveraged impact of ground-up trends on higher layers is a well-documented phenomenon in actuarial literature. See for example Lange [6]. Essentially the theory shows that ground-up trends such as inflation are intensified in higher layers of insurance as small increases to ground-up losses result in relatively larger increases to losses within the layer. This effect is most pronounced for losses that were expected to fall below the attachment point and now trend into the excess layer as a result of inflation. For losses which have already or nearly exhausted the limit the effect is tempered and the impact of inflation on such losses can be minimal. Figure 1 below provides a simple illustration of this leveraged impact. Each panel shows the impact of 5% ground-up inflation on an excess layer reinsurance product by varying a single term of the contract.



(a) By Attachment (Unlimited) (b) By Limit (No Attachment) (c) By Share (1K xs 1.5K)

Figure 1. Leveraged impact of inflation (5%) for varying attachment, limit, and share.

Although shown in a very specific environment, these relationships will hold true in most situations and help illustrate four general principles with regard to (re)insurance and inflation. Namely, that (1) as the attachment increases from zero to unlimited, the expected layer inflation increases from the ground-up inflation rate to a theoretically unlimited amount (although in practice, for most reasonable excess layers, the leveraged inflation will appear to stabilize asymptotically at some large amount); (2) as the limit increases from zero to unlimited the layer inflation increases from 0% to the ground-up inflation rate; (3) the share does not affect inflation; and (4) the dominant determinant of the leveraged impact of inflation is the attachment point with the effect dampened by the limit. While the relative magnitude and exact impact are of course unique to any situation, these four principles can be used to help understand the general impact of inflation on most general (re)insurance contracts.

4.2 Decreased limits factors

The purpose of this section is to address the relationship between inflation and increased limits factors (ILFs)/decreased limits factors (DLFs). Specifically, we show that although, as would be expected from the prior section, deterministic inflation increases the DLF for excess layers, deterministic inflation decreases the volatility around that DLF. We also show that the effect of stochastic inflation on the DLF will often be negligible.

4.2.1 Some background on increased limits factors

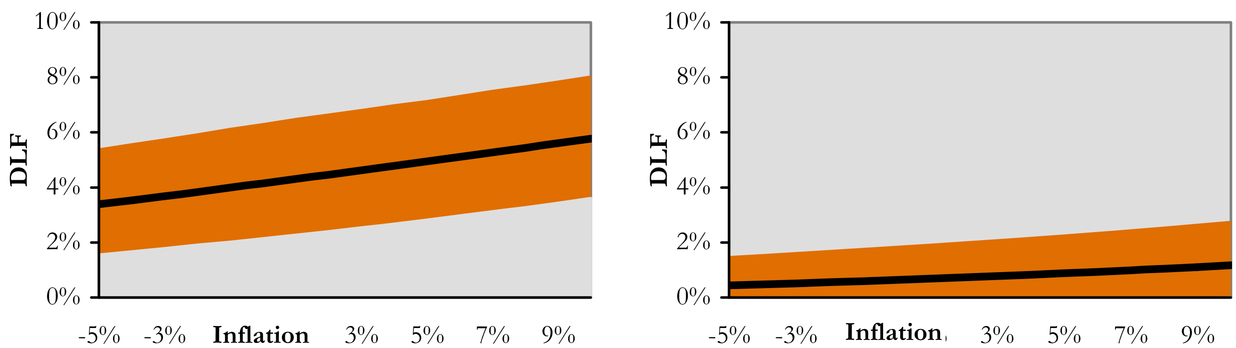
In his 1977 paper, "On the Theory of Increased Limits and Excess of Loss Pricing," Robert Miccolis does an excellent job of developing simple mathematical formulas, still widely used today, for setting increased limits factors (ILFs). His approach is "moment-based" whereby the ILF is set equal to the ratio of the expected value of losses in layer A to the expected value of losses in layer B. Unfortunately, this approach results in a deterministic ILF as the expectation of a random variable is a fixed rather than variable quantity. As such, to answer the question as to whether inflation affects the volatility of the ILF, we rely on a stochastic variant of this deterministic ILF, namely the ratio of losses in layer A to losses in layer B prior to expectation. Although this quantity is not as useful in practice, it does provide a reasonable approximation under nice conditions and will allow us to address whether or not there is a leveraged impact of *uncertainty* as well as a leveraged impact of inflation.[1]

4.2.2 The relationship between inflation and DLFs — deterministic scenario

Figure 2 plots the 95th percent confidence interval around the working and excess layer DLFs computed using increasing amounts of deterministic inflation. Note that as expected, the DLF for these upper layers increases with the inflation rate, i.e., more losses trend into the layer. However, as the DLFs of the lower, working, and excess layers must by definition sum to 100%, it will not always be the case that both the working and excess layers DLFs will increase. A more universal comparison would be the lower layer vs. a single upper layer where the lower layer DLF will always

[1] Namely, we conjecture that if aggregate losses are modeled using the collective risk model then the expectation of the ratio of losses in layer A to losses in layer B (i.e., E[A/B]), where A is a subset of the losses in B, will tend to the ratio of expectations (i.e., E[A]/E[B]) for sufficiently large number of independent and identically distributed insureds/claims.

decrease due to the effect of the upper limit and the upper layer DLF would always increase. In the case of multiple upper layers, it is possible that the DLF for one or more upper layers may actually decrease given the right relationship of reinsurance terms and loss distribution. That aside, more importantly note that increasing the underlying inflation rate does not appear to impact the volatility of the DLF. In fact, although not obvious from the graphs, the volatility actually decreases *relative* to the mean in these examples. This will commonly be the case as while inflation increases the losses in the layer, the reinsurance terms will "squeeze" the losses in the layer.

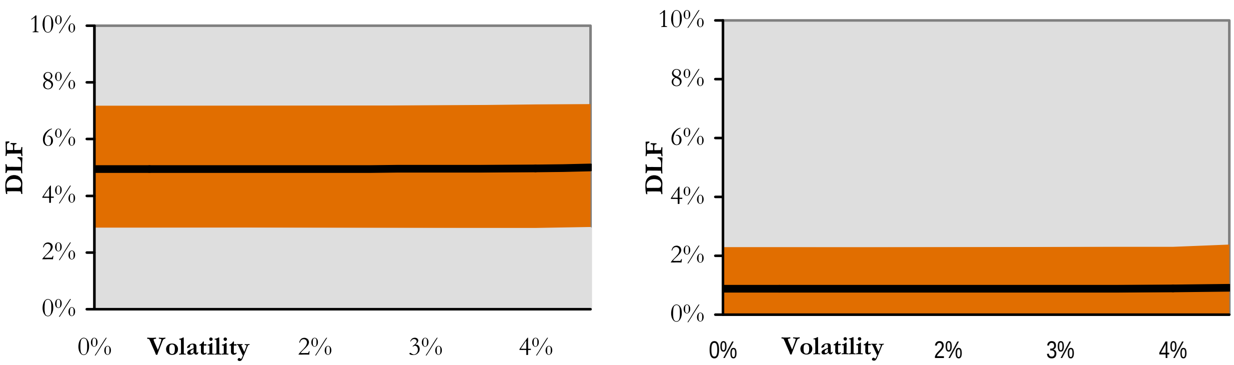


(a) Working layer (b) Excess layer

Figure 2. Confidence interval around estimates of the working and excess layer DLFs for increasing levels of deterministic inflation.

4.2.3 The relationship between inflation and DLFs — stochastic scenario

In the case of stochastic inflation, the effect of increasing the volatility of ground-up inflation on both the best estimate DLF as well as uncertainty around the best estimate is generally somewhat negligible. As will be discussed in more detail in the section on loss distributions, the exact direction and amount of the effect is dependent on several, often contra-directional, changes to the volatility and mean of the underlying frequency and severity distributions determined by the relationship between the reinsurance terms and the loss distribution before and after the application of stochastic inflation. And at the end of the day, the effect becomes somewhat immaterial. Figure 3 helps to demonstrate these results plotting the DLF and funnel of doubt for various levels of volatility.



(a) Working layer (b) Excess layer

Figure 3. Confidence interval around estimates of the working and excess layer DLFs for increasing levels of stochastic inflation volatility keeping expected inflation constant at 5%.

4.2.4 Going forward

The most interesting conclusion of this section is that while the leveraged impact of inflation can be significant, if we can develop an adequate expectation as to future inflation, and incorporate it into our models as such, then we can usually develop a good understanding of future reinsurance losses— even if they are considerably larger than they have been historically.

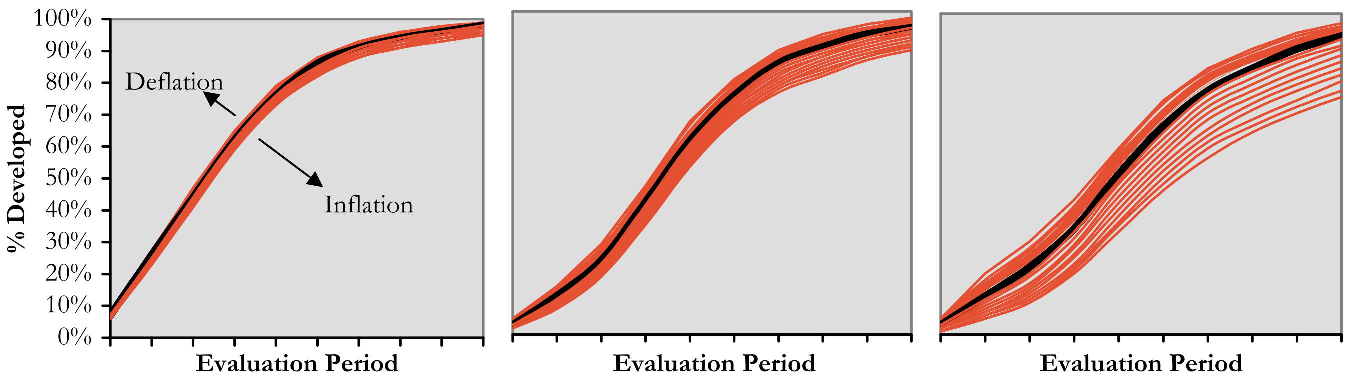
4.3 Loss development

The purpose of this section is to discuss the relationship between inflation and loss development. Primarily, we show that inflation in excess of that observed historically slows down loss development and, conversely, deflation speeds up development. This effect is exacerbated for higher layers of insurance, longer-tailed lines, and lines which are characterized by single lump-sum payments (i.e., liability lines) rather than those characterized by periodic payments.

4.3.1 The relationship between inflation and development— by reinsurance layer

Calendar year inflation which is consistent from year to year will not impact the accuracy of loss development methodologies (i.e., the chain-ladder method). This is shown more explicitly in Boles et al. [1], but can be understood by noting that loss development methodologies, by virtue of taking the ratio of losses from one period to the next, cancel out the impact of calendar year inflation in the numerator and denominator. That said, loss reserving methods which aren't 'development' methodologies all have to make some adjustment for inflation. Often this involves trending forward incremental amounts, adjusting the IELRs and so forth. But still, assuming the inflation is steady, and the adjustment is reasonable, the accuracy of these methods isn't affected.

However, it is when the inflation rate changes abruptly and materially, that our estimates of ultimate loss and unpaid claim liabilities will be distorted. Inflation in excess of that shown historically will slow down development patterns and inflation less than that shown historically will speed up development patterns. Like the leveraged impact of base inflation, these impacts are also considerably more leveraged in excess layers of insurance. This is illustrated in Figure 4 which plots the first 10 years of the cumulative paid pattern in each of the lower, working, and excess layers. The dark black reference line indicates the pattern with historical inflation removed. The lines above this reference line show the development pattern with increasing magnitudes of deflation; and the lines below show the development pattern with increasing magnitudes of inflation.



(a) Lower (b) Working (c) Excess

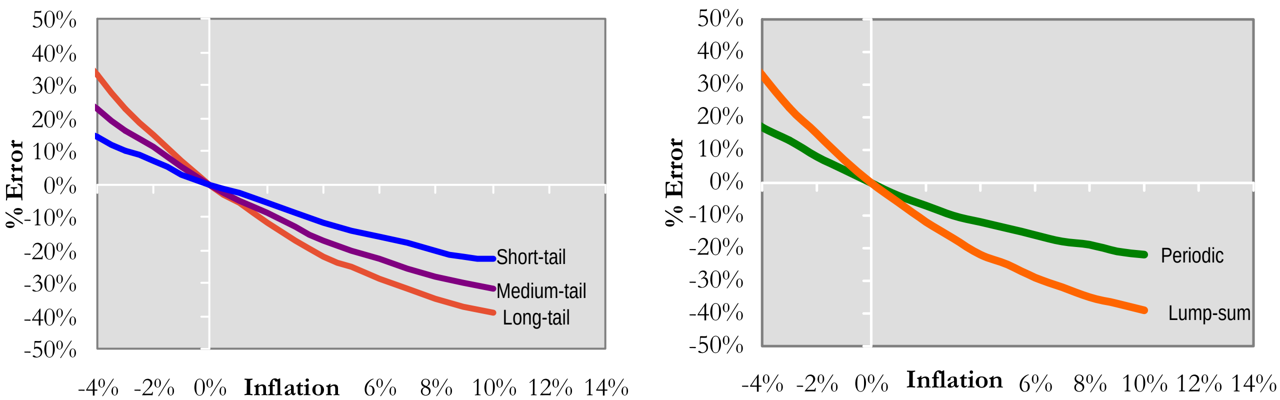
Figure 4. Effect of inflation on loss development for various insured layers.

4.3.2 The relationship between inflation and development—by type of pattern

Not only is there a leveraged distortion on development in higher layers, the amount by which the pattern is distorted very much depends on the time of pattern. Taking the working layer as an example, Figure 5(a) compares the error[2] in the chain-ladder method for a short-tailed, medium-tailed, and long-tailed line of business and various levels of inflation. As expected, the long-tailed line is significantly more affected than the shorter-tailed lines as the distortion compounds with inflation at the later evaluations. Figure 5(b) compares the error in the chain-ladder method for a line of business primarily characterized by periodic payments (i.e., workers compensation indemnity) vs. a line of business primarily characterized by single lump-sum payments (i.e., medical malpractice). Note that development for the periodic payment class is much less affected than development for

[2] Here, error specifically refers to the expected estimate of ultimate loss less the actual ultimate loss divided by the actual ultimate loss.

the lump-sum class as lump-sums at later maturities bear the full-brunt of inflation whereas with periodic payments only a portion of the total claim is adjusted for inflation at these later maturities minimizing the overall impact on development. Both of these results imply that as the duration increases, so does the distortion.



(a) Length of tail

(b) Primary type of payment

Figure 5. Effect of unexpected inflation on various types of patterns.

4.3.3 Going forward

Of all the results shown in the paper, loss development, and the projection of ultimate loss, is the one most sensitive to the exact conditions. The degree to which a pattern slows down (or speeds up in the case of deflation) is significantly dependent on whether the data is short-tailed or long-tailed, how losses are paid, the degree to which inflation is present in historic data, the exact policy limits, the size of the book, and so forth. To this end, and especially as inflation departs from its historical norm, it is necessary to utilize reserving methodologies which both make an implicit or explicit adjustment for historical inflation as well as allow you to incorporate your own actuarial judgment as to future inflation into the projections.

4.4 Dependence

This section considers the relationship between inflation and dependence. We show that with regard to general economic inflation affecting all lines simultaneously, increasing the volatility of inflation will increase the dependence between lines. Conversely, with regard to specific by-line inflation affecting only a single line, increasing the volatility of inflation will decrease the dependence between lines. Finally, we note that here the key driver of these results is the volatility of inflation rather than the actual inflation rate.

4.4.1 Inflation, dependence, and systemic vs. non-systemic risk

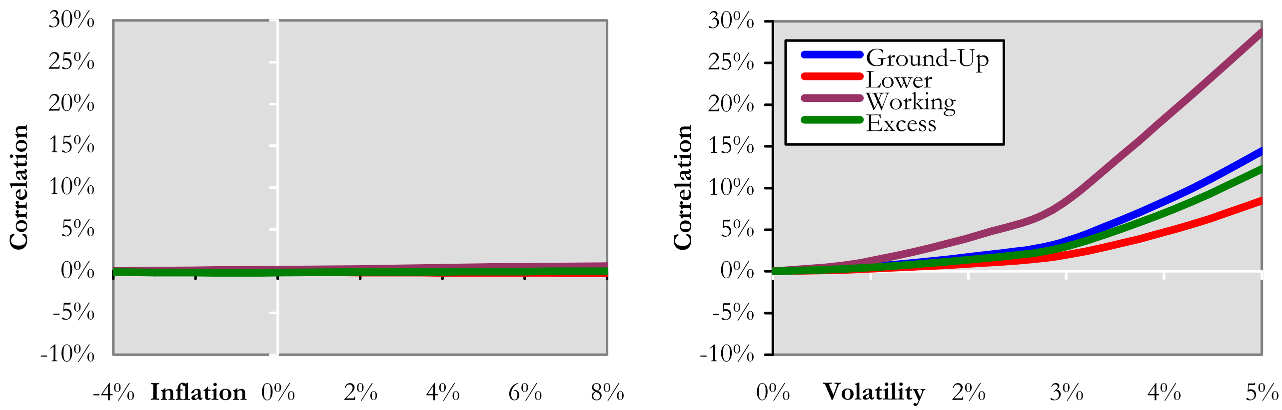
To understand these results, it is necessary to first take a step back and examine the relationship between inflation and dependence. Specifically, that it is not the magnitude of the expected inflation which matters, but rather it is the uncertainty around that expected magnitude. Without going into the mathematics, it is perhaps easiest to frame the problem by considering dependence as a function of the amount of systemic risk relative to the non-systemic risk. When the amount of systemic risk is substantially larger than the amount of non-systemic risk, dependence will generally be high as the systemic risk dominates and vice versa. Thus, by increasing the volatility of inflation which is exogenous to both lines of insurance, we are in turn increasing the amount of systemic risk relative to non-systemic risk and increasing the dependence between lines. On the other hand, by increasing the volatility of inflation for a single line, we are increasing the amount of non-systemic risk relative to systemic risk and as such decreasing the dependence between lines. While the direction is predictable, the rate at which the dependence changes depends on the initial ratio of systemic to non-systemic risk which is highly susceptible to the interrelationship between the reinsurance terms and the underlying frequency and severity distributions. As such, it will not always be the case that the dependence by layer changes in an ordered manner with predictable rates of change.

4.4.2 The relationship between inflation and dependence—general inflation

Figure 6 shows how inflation can change the dependence between lines by plotting the correlation[3] between two lines in the scenario where general monetary inflation affects both lines simultaneously. Figure 6(a) shows the effect of increasing the inflation rate in the deterministic scenario and Figure 6(b) shows the effect of increasing the volatility of inflation in the stochastic scenario. In order to emphasize our findings, we started with two lines which were independent of one another and then added inflation. First note that in the case of varying the degree of deterministic inflation, there is no effect as fixed inflation does not distort the degree of systemic vs. non-systemic risk. However, as we increase the volatility of general economic inflation, the correlation between lines increases as the systemic risk increases relative to the non-systemic risk. Although the direction of the change is fairly easy to assess, the magnitude of change is quite difficult to predict without actually modeling the specific scenario.

---

[3] As an aside, note that while we use correlation as our measure of dependence in this section, the correlation measure is not without its weaknesses and using such a measure to assess dependence may not always be appropriate.

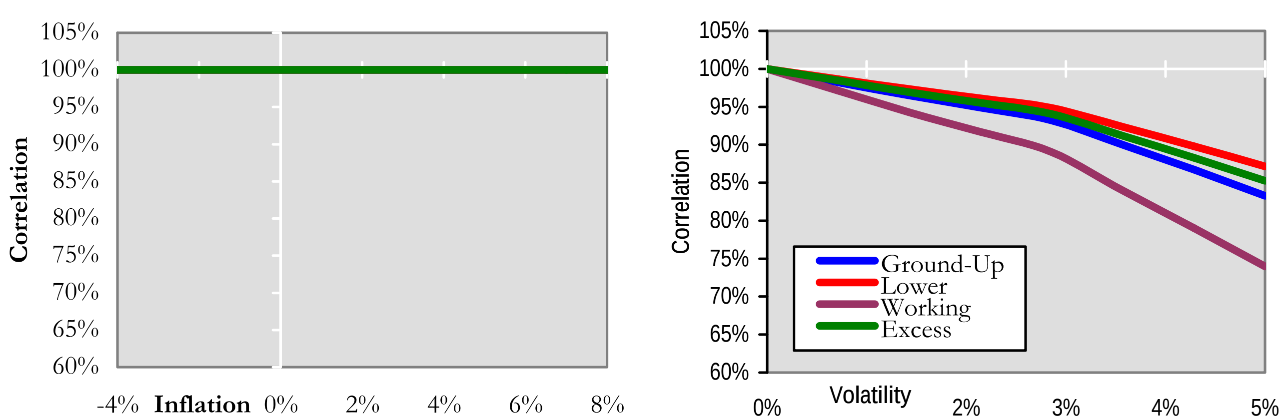


(a) Deterministic Scenario (b) Stochastic Scenario

Figure 6. Effect of general inflation on dependence by layer for various inflation rates and volatilities.

4.4.3 The relationship between inflation and dependence – specific inflation

Now, while the above results refer to the situation where inflation impacts several lines of insurance simultaneously, Figure 7 shows how specific by-line inflation affecting a single line, will decrease the dependence between lines. Figure 7(a) shows the effect of increasing the inflation rate in the deterministic scenario and Figure 7(b) shows the effect of increasing the volatility of inflation in the stochastic scenario. In order to emphasis our findings, we started with two lines which were perfectly correlated and then added inflation to one line. Again note that varying the level of deterministic inflation has little effect; but, with regard to varying the volatility of the inflation parameter, we see that the correlation between lines quickly decreases as the amount of non-systematic risk increases relative to the systemic risk.



(a) Deterministic Scenario (b) Stochastic Scenario

Figure 7. Effect of specific by-line inflation on dependence by layer for various inflation rates and volatilities.

4.4.4 Going forward

These results are nothing new, in fact they are the basis of the Marshall-Olkin copula structure and often used for introducing dependence among independent events in actuarial science through the form of a "contagion" parameter in the collective risk model. See Klinker et al. [6] for example. However, they are not always considered when setting correlation assumptions resulting in capital models which may mis-specify the amount of dependence and degree of risk between lines.

4.5 Loss distributions

Like loss development, the effect of inflation on loss distributions by layer is highly speculative and depends primarily on the interaction between the reinsurance terms and the ground-up distributions. However, there are a few obvious effects—inflation will cause both the mean frequency in the upper layers and the severity in all layers to increase. While inflation will typically cause the volatility of frequency in the upper layers to increase, the effect of inflation on the volatility of severity in the upper layers is less definite and depends primarily on how much room losses in the layer have to play (i.e., is it a tight or wide layer). With regard to ground-up experience, the volatility of the severity distribution will increase with inflation; and in the lower layer, the volatility of the severity distribution will decrease with inflation. More interesting though is the effect inflation has on the volatility *relative to the mean* by insured layer. With regard to both frequency and severity in the upper layers, the CV will typically decrease as inflation increases. Although this is by no means always the case, it will generally be the case when the upper layers insure a non-insignificant share of the loss (i.e., 5% is cutting it close, 10% is getting safer). Typically, the more loss experience there is in a layer, the more likely it is that the CV will decrease with inflation as the loss experience will be more stable and less likely to be affected by large loss "pops." Table 1 below summarizes these points. We have attempted to illustrate confidence in the result by using a scale of one to three arrows for not confident to very confident with a question mark indicating no confidence in making an assessment.

	Frequency			**Severity**		
Layer	**Mean**	**SD**	**CV**	**Mean**	**SD**	**CV**
Ground-up	No change	No change	No change	↑↑↑	↑↑↑	No change
Lower	No change	No change	No change	↑↑↑	↓↓↓	↓↓↓
Working	↑↑↑	↑↑	↓	↑↑↑	?	↓*
Excess	↑↑↑	↑↑	↓	↑↑↑	?	↓*

*Mostly depends on size of layer relative to loss.

Table 1. Effect of increasing **deterministic** inflation on underlying frequency and severity distributions by layer.

Figure 8 illustrates these same points graphically by focusing on the aggregate loss distribution and plotting the probability density for various amounts of inflation. With regards to the ground-up distribution, note that the "location" of the density changes substantially, while the "shape" of the density appears to change only slightly for the various levels of inflation. This makes sense considering the effect of inflation on the component frequency and severity as described above. While inflation does not affect the frequency distribution, it does increase the mean of the severity distribution (i.e., change in location) and although it does increase the standard deviation of the distribution it doesn't change the volatility relative the mean (i.e., similar shape). However, with regard to the upper layers (excess layer is pictured) note that the shape of the distribution, as well as location, changes substantially with inflation as would be indicated by the above.

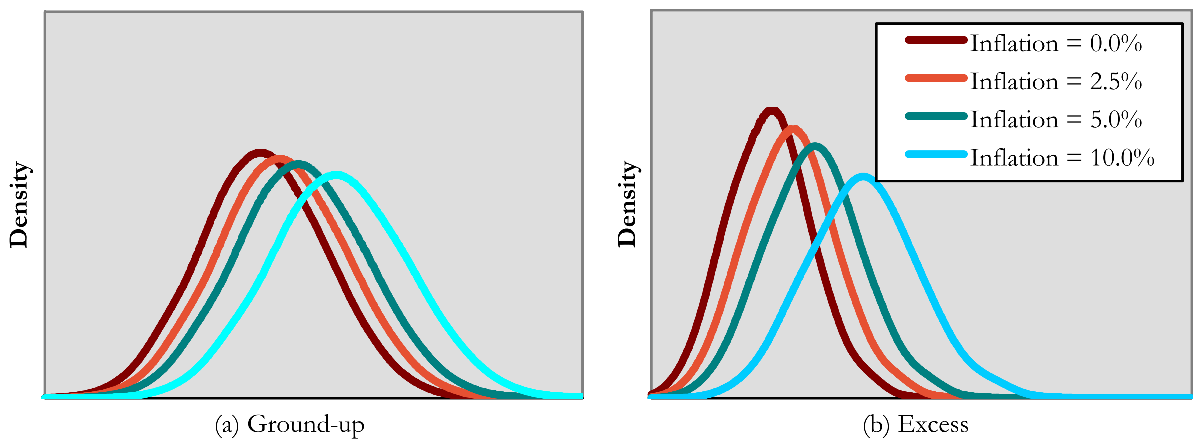


Figure 8. Shift in aggregate distribution due to various inflation scenarios.

These results are too broad for implementation in a specific situation, but are quite useful as an intermediate step for framing the following sections and so we have included them for completeness.

4.6 Credibility

The purpose of this section is to discuss the relationship between inflation and credibility as it relates to historical loss experience. Primarily, we show that (1) in the case of deterministic inflation, as the inflation rate increases, the credibility of ground-up experience will remain unchanged, but the credibility of experience by layer will increase; and that (2) in the case of stochastic inflation, as the amount of volatility increases, the credibility of ground-up experience will decrease although the change in credibility of experience by layer is typically minimal.

4.6.1 Some background on credibility

Without going into too much detail, credibility is the actuarial concept which refers to the amount of weight which should be assigned to historical experience. One of the most common credibility frameworks is Bühlmann credibility which gives the credibility weight as function of three different components—the amount of historical data, the variance of the hypothetical means (VHM) and the expected value of the process variance (EVPV). The relationship of the first component, the amount of data, on credibility is rather obvious in that as the amount of historical data increases so does the weight one should assign to it. For simplicity and without loss of generality, we consider just a single year of experience. The latter two components, the EVPV and the VHM, are rather more difficult to conceptualize, but are excellently illustrated in Steve Philbrick's 1981 paper. "An Examination of Credibility Concepts," in which he draws an analogy with marksmen shooting at targets. However, for our current purposes, it is more useful to think of these concepts within the framework of the collective risk model. Here, the EVPV is primarily driven by the variability in the size of losses (i.e., the coefficient of variation or CV for severity) and the VHM is primarily driven by the variability in the number of claims (i.e., the variance-to-mean ratio or VTM for frequency).

Consider first the EVPV. As the variability in the size of losses increases, the EVPV also increases, but the credibility of actual experience decreases. To understand this, consider the following example: if losses aren't variable and we observe a loss of $1,000, we can be 100% certain that all losses are $1,000. However, if losses are extremely variable and we observe a loss $1,000, we don't actually know if other losses are $1,000 or $1,000,000. In Philbrick's language, the more variability in losses, the more the targets overlap with one another and so with any one observation we are not very confident from which archer it came.

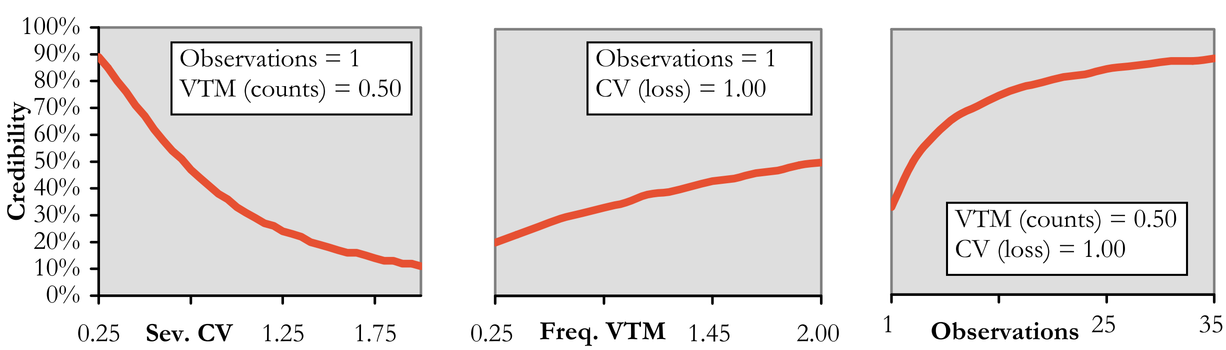
Consider next the VHM. Now, as the variability in the number of claims increases, the VHM also increases and so does the credibility of actual experience. This relationship is a little bit more difficult to understand as it is somewhat counterintuitive, but consider the following example: suppose there is either 1 claim or 100 claims and that the cost per claim is about $50. If we observe aggregate losses of about $50 we can be pretty sure that the number of claims is 1 and if we observe aggregate losses of about $5,000 we can be pretty sure that the number of claims was 100. However, if the number of claims is either 9 or 11 and we observe aggregate losses of $500, we really don't know whether the number of claims was 9 or 11. In Philbrick's language, the more variability in the number of claims there is, the further apart the targets are pushed so that with any one observation we can be more confident in which target the marksmen is aiming at.

Table 2 summarizes these relationships. Note that by re-framing the EVPV and VHM in terms of the key statistic—CV for severity and VTM for frequency—we can easily utilize the results from section 4.5 to explain how changes in the inflation rate might be expected to impact the credibility of historical data.

Driver	Credibility Component	Key Statistic	Credibility
Variability of losses increases	EVPV ↑	CV of severity	↓
Variability of the number of claims increases	VHM ↑	VTM of frequency	↑
Number of observations increases	N ↑	N/A	↑

Table 2. Drivers of credibility.

Figure 9 highlights these relationships graphically. Panel (a) shows that as the CV of the severity distribution increases, the credibility decreases. Panel (b) shows that as the VTM of the frequency distribution increases, so does the credibility and Panel (c) shows that as the number of observations increases, so does the credibility of historical experience. Note that the relationship between each of these components and the credibility is in no way linear and is different for each component. This implies that the exact credibility depends very much on the relationship between attachment and limit as well as the interaction between these terms and the underlying frequency and size-of-loss distributions.



(a) By CV of Loss (b) VTM of claim counts (c) By number of observations

Figure 9. Drivers of credibility.

4.6.2 The relationship between inflation and credibility—deterministic scenario

Figure 10 highlights the effect of increasing the rate of inflation on the credibility for the lower, working, and excess layers. Note that we have also included a black reference line on each graph to show the impact of credibility on the ground-up experience. With regard to ground-up experience, changing the rate of inflation does nothing to impact the credibility of historical data; whereas the credibility by layer increases with the rate of inflation. To some extent, these results make sense when considering layered (re)insurance contracts. The higher the inflation rate, the more losses we would expect to trend into upper layers and increase the amount of experience from which to project from. Further, this additional experience would act to stabilize the "attritional" component of losses in the layers relative to the "large" or "catastrophic" components. And finally, the higher the inflation, and without any indexation of limits and attachment, the more we would expect the reinsurance terms to come into play.

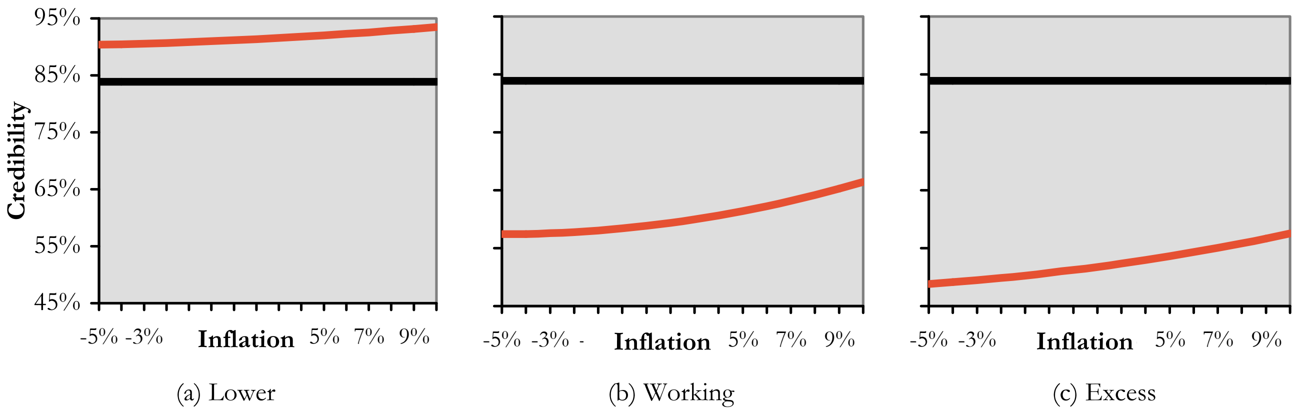


Figure 10. Effect of deterministic inflation on credibility by layer.

However, that said, it is also easy to understand these results with reference to how inflation changes the distribution of the frequency and severity components by layer, keeping both Table 1 and Table 2 in mind. Ground-up deterministic inflation will not change the underlying frequency and severity distribution and thus there is no impact on the credibility of experience. In the lower layer, although there is no change to the frequency distribution, the severity CV will decrease as inflation increases causing the credibility of experience to increase. In the upper layers, the effect is not so certain, but most typically, when these layers have a sufficient amount of experience, the severity CV will decrease with inflation causing the credibility of experience to increase. And although in the upper layers there is also a shift in the frequency distribution, this "frequency" effect is most often dominated by the "severity" effect and can be considered less material.

4.6.3 The relationship between credibility and inflation—stochastic scenario

With regards to stochastic inflation, the results by layer are not nearly as nice. It is first easiest to note that the credibility of ground-up experience will decrease with increased volatility in the inflation parameter as the underlying coefficient of variation for the severity distribution will increase. This result would be expected as the more variable historical experience is, the less we will rely on it. With regards to the insured layers, the effect of inflation volatility on credibility is harder to predict. In general, the credibility will decrease although the exact change is quite uncertain. *However*, as the magnitude of the change will usually be quite small, increased volatility of the inflation parameter is not as worrying.

4.6.4 Going forward

Perhaps the key implication from this section is that for lines with high inflation we should probably be giving more weight to more recent years' experience; whereas, where the claims inflation is largely uncertain, we should primarily rely on a longer history of data smoothing the results out based on long-term averages.

4.7 Risk Margins

The purpose of this section is to discuss the relationship between inflation and the risk margin. In a deterministic setting, we show that increasing the rate of inflation will actually lead to a lower relative risk margin[4] for most layers of reinsurance. With regard to stochastic inflation, we note that the effect of changing the volatility of inflation will most often be negligible.

4.7.1 Some background on risk margins

The risk margin is generally defined as the amount in excess of expected loss which is added as a load to reasonably compensate for the risk associated in an insurance contract. There are a many ways to measure the risk margin, each with their own properties and advocates, but some of the most common measures include the standard deviation (SD), semi-deviation (Semi-SD), value at risk (VaR), and conditional tail expectation (CTE). Because there is no best measure of risk, a standard of "coherence" is often ascribed to those risk measures which possess some desirable characteristics—namely monotonicity, sub-additivity, positive homogeneity, and translation invariance. Here we consider positive homogeneity in detail as it will help frame the effect of

[4] Here, risk margin refers to the percentage multiplicative load rather than a nominal additive load. While inflation will increase the nominal amount of risk margin, it will decrease the relative risk margin as a percentage of the mean.

inflation on risk margins. Simply put, positive homogeneity states that if the insured exposure grows by some percentage *q*, then our risk also grows by that same percentage *q*. Theoretically, this property makes sense—if we double the size of a portfolio, the risk should also double. Although, as we will discuss, this property, at least on the face of it, does not hold in the case of inflation.[5]

4.7.2 The relationship between inflation and risk margins—deterministic scenario

If we let *q* represent inflation, then for a given risk measure we might expect that 5% inflation would increase our estimate of risk by 5%. And if we were to apply a fixed inflation factor to aggregate losses by layer, this certainly would be true. However, with regard to non-proportional reinsurance, the policy terms absorb some of the inflation shock and in turn limit the increase in risk margin. In short, while the insurable exposure gets *q*% bigger, the risk associated with that exposure only gets *(q-ε)*% bigger. This implies that when considering the effect of inflation, the mean is substantially more leveraged than the risk associated with the mean. Figure 11 demonstrates this phenomenon. For comparison purposes, this figure normalizes the risk margin relative to the risk margin with no inflation in each of the layers as indicated by the cross-hairs. Here, the effect is most pronounced for the excess layer and almost negligible for the lower layer.

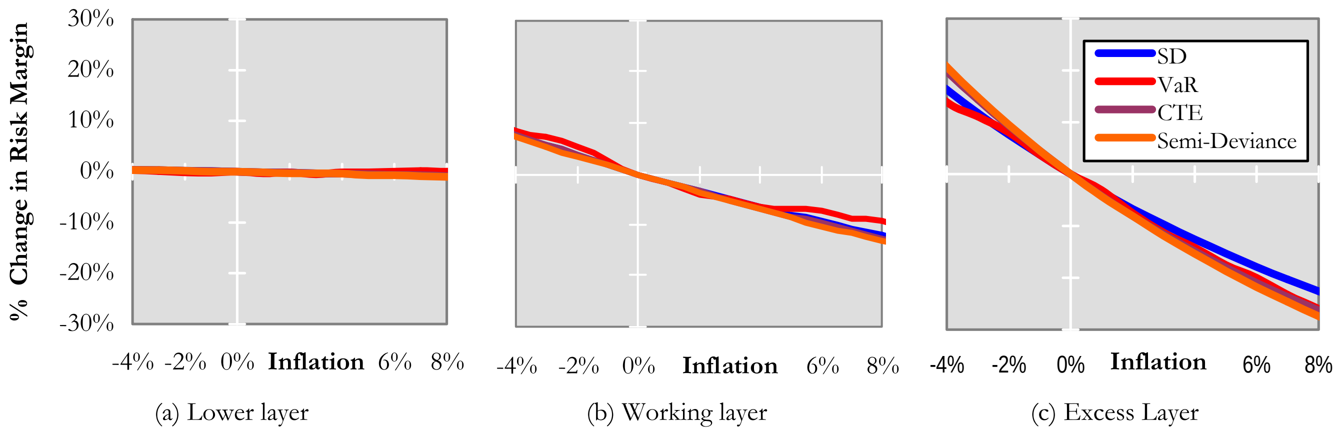


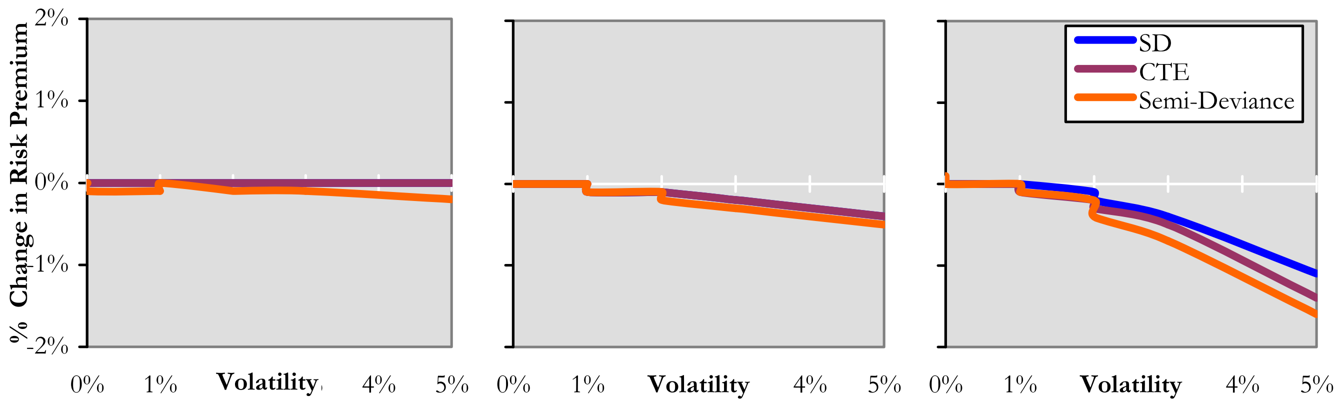
Figure 11. Percentage change in risk measure for increased levels of inflation (measured relative to no inflation).

4.7.3 The relationship between inflation and risk margins—stochastic scenario

The results with regard to stochastic inflation are less straightforward although the magnitude of results will not usually be that material. In this specific situation, Figure 12 shows that the risk

[5] Technically, positive homogeneity does hold in this situation; however, it only appears to not hold because by modelling inflation on a per-occurrence basis and subjecting each loss to the reinsurance terms of the layer we are in effect distorting the distribution rather than just "doubling the exposure base."

margin decreases as the volatility of inflation increases. This will not always be the case (i.e., the risk margin could increase) with the exact result depending on the relationship between the mean and volatility of losses in the layer relative to the reinsurance terms before and after the stochastic inflation is modeled. However, note that here as well as in most situations, the relative magnitude of the effect will be rather negligible both in relative and nominal terms.



(a) Lower layer (b) Working layer (c) Excess Layer

Figure 12. Percentage change in risk premium for increasing levels of inflation volatility (as measured relative to deterministic inflation).[6]

4.7.4 Going forward

The key implication of this section is that as the amount of ground-up inflation increases, insurers playing in higher layers need to worry more about their best estimate liability and less about their risk margin, relatively speaking.

[6] Note that the VaR amount is not shown here because, when dealing with relative amounts so small in magnitude, it is difficult to precisely estimate VaR using simulation techniques.

5. CONCLUSION

The primary intention of this paper was to explore the effect of inflation on several common areas of actuarial modeling. Where possible, we tried to present certain basic rules of thumb which might help provide general guidance to actuaries when working with inflation. However, we note that the exact magnitude and effect of inflation on losses is highly uncertain and will heavily depend on the actuary's unique situation; so we would hope that this paper will be used as proof that inflation should in a real, and non-trivial manner, be incorporated into most actuarial models.

Acknowledgments
We would like to thank the following individuals for their comments and helpful contribution to this paper: Yves Colomb, Ali Ishaq, George Maher, James Norris, Dom Yarnell, and Patrick Yu.

6. REFERENCES

[1] T. Boles and A. Staudt, "On the accuracy of loss reserving methodology," CAS *E-Forum*, Fall 2010, pp. 1-62.
[2] R. Butsic, "The effect of inflation on losses and premiums for property-liability insurers," CAS *Discussion Paper Program*, May 1981, pp. 58-102.
[3] S. D'Arcy, "An Illustration of the Impact of Inflation on Insurance Company Operations," CAS *Discussion Paper Program*, 1979, pp. 96-109.
[4] R. Durrett, "Probability: Theory and Examples," New York: Duxbury Press, 1995.
[5] T. Gault, L. Llaguno, and S. Lowe, "A Structural Simulation Model for Measuring General Insurance Risk," CAS *E-Forum*, Summer 2010, pp. 1-57.
[6] F. Klinker, G. Meyers, and D. Lalonde, "The aggregation and correlation of the insurance exposure," CAS *E-Forum*, Summer 2003, pp. 15-82.
[7] S. Klugman, H. Panjer, and G. Willmot, "Loss Models: From Data to Decisions," New York: Wiley, 2008.
[8] J. Lange, "The Interpretation of Liability Increased Statistics," *PCAS* LVI, 1969, pp. 163-173.
[9] S. Lowe and R. Warren, "Post-Recession Inflation: An Emerging Risk for P&C Insurers," *Emphasis* 3, 2010, pp. 24-29.
[10] T. Mack, "Credible Claims Reserve: The Benktander Method," *ASTIN,* 2000, pp. 425-441.
[11] R. Miccolis, "On the theory of increased limits and excess of loss pricing," *PCAS* 1977, Vol. LXIV, pp. 27-59.
[12] S. Philbrick, "An examination of credibility concepts," *PCAS* LXVIII, 1981, pp.195-212.
[13] R Development Core Team (2009). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Abbreviations

ALM, asset liability matching
CAS, Casualty Actuarial Society
CEIOPS, Committee of European Insurance and Occupational Pensions
CTE, conditional tail expectation
CV, coefficient of variation
DLF, decreased limit factor
ERM, enterprise risk management
ESG, economic scenario generator
EVPV, expected value of process variance
ILF, increased limit factor
QIS, qualitative impact study
Semi-SD, semi-deviation
SCR, Solvency Capital Requirement
SD, standard deviation
VaR, value-at-risk
VHM, variance of hypothetical means
VTM, variance-to-mean ratio
XOL, excess-of-loss

Biographies of the Authors

- **Matthew Ball** is a consultant with Towers Watson in Bermuda where is head of risk consulting. He is a Fellow in the Institute of Actuaries. He can be reached at: matthew.ball@towerswatson.com or +1 441 279 6706.
- **Andy Staudt** is a consultant with Towers Watson in London. He is a Fellow of the CAS, a member of the American Academy of Actuaries and has a Masters in Statistics from the University of California at Berkeley. He can be reached at: andy.staudt@towerswatson.com or +44 (0) 20 7170 3479.

Conditional Probability and the Collective Risk Model

Leigh J. Halliwell, FCAS, MAAA

Abstract. One of the most powerful and profound tools of casualty actuarial science is the collective risk model $S = X_1 + \ldots + X_N$. It is widely used by casualty actuaries, especially by those in the field of reinsurance. Nearly one hundred pages of one standard textbook (Klugman, [1998], Chapter 4) hardly suffice to survey the ingenuity with which actuaries and scholars have analyzed it. Much of their analysis proceeds from the application of conditional probability to the so-called individual risk model $S = X_1 + \ldots + X_n$. This paper penetrates deeper into both conditional probability and the collective risk model, deriving new insights into higher moments and their generating functions. Particular attention is devoted to the fourth moment of the collective risk model, for which no formula seems previously to have been published. An appendix extends conditional probability to a novel technique of loss development.

Keywords: conditional probability, moments, cumulants, collective risk model.

1. INTRODUCTION

This paper applies conditional probability to the moments of the collective risk model. In the next section we will set forth definitions of conditional moments and co-moments, and in third section will derive formulas in which unconditional moments are expressed in terms of conditional ones. Next, in the fourth section, after explaining why moments higher than the third are not additive, we will introduce an additivity-restoring adjustment known as a cumulant. In the fifth section we will apply conditional cumulant formulas to the collective risk model to seize the prize of a manageable formula for its fourth cumulant. Finally, in the sixth section we will explain the cumulant generating function, and show its usefulness in relating cumulants to moments and in deriving cumulants of the collective risk model.

2. DEFINITIONS

Let X be a random variable with finite mean $E[X]=\mu$. For positive integer *n*, define the n^{th} moment of X as $M_n[X]=E[(X-\mu)^n]$.[1] Here we have defined what probability theory calls the *central* moments of X. Of course, the first central moment is zero. The second moment is the variance, and the third is the skewness. We shall call the fourth moment the kurtosis.[2] Now let Θ denote an event which conditions the probability distribution of X. We can then speak of the conditional n^{th} moment $M_n[X|\Theta]=E[(X-\mu_\Theta)^n|\Theta]$, where $\mu_\Theta=E[X|\Theta]$.[3]

Furthermore, as a multivariate extension, we can define the n^{th} co-moment as $CM[X_1,\ldots,X_n]=E\left[\prod_{k=1}^{n}(X_k-\mu_k)\right]$. Since the order of the co-moment is the number of its arguments, it is superfluous to subscript the definition as CM_n.[4] The first co-moment is zero; the second is the covariance. We shall call the third and the fourth co-moments the co-skewness and the co-kurtosis. A co-moment of random variables is zero, if any of them is constant, since in that case one of the factors in the Π operator will always be zero. The same random variable may appear in the argument list more than once; as a special

---

1 The reader should not confuse $M_n[X]$ for the n^{th} moment with $M_X(t)$, the moment generating function of X.

2 Some define kurtosis as the fourth cumulant, $E[(X-\mu)^4]-3E[(X-\mu)^2]^2$, also known as excess kurtosis because the kurtosis of the normal distribution is three times the square of its variance. Sometimes (e.g., Daykin [1994], 24) skewness and kurtosis are defined as what we would call *coefficients* of skewness and kurtosis, i.e., the moments or cumulants stripped of dimension by dividing them by the third and fourth powers of the standard deviation.

3 In general, $E[(X-E[X])^n|\Theta]\neq E[(X-E[X|\Theta])^n|\Theta]$. Conditioning at one level of expectation should by default cascade into the next or nested level, and so on. The tendency to disregard this inequality may indicate a defect in the accepted notation. It helps (at least it helps this author) to regard unconditional expectation as conditional upon the universal event V: $E[X]=E[X|V]$.

4 One must be wary of such mistakes as equating $CM[X,X,Y]$ and $CM[X^2,Y]$, which confuses a third co-moment with a second.

case, $CM\left[\overbrace{X,\ldots,X}^{n\text{ times}}\right]=E\left[\prod_{k=1}^{n}(X-\mu)\right]=E\left[(X-\mu)^{n}\right]=M_{n}[X]$. The conditional co-moment is $CM\left[X_1,\ldots,X_n\middle|\Theta\right]=E\left[\prod_{k=1}^{n}\left(X_k-E\left[X_k\middle|\Theta\right]\right)\middle|\Theta\right]$.

3. UNCONDITIONAL MOMENTS IN TERMS OF CONDITIONAL

Our purpose here is to derive formulas that express unconditional moments in terms of moments and co-moments conditional upon Θ. This begins with the key sequence:

$$M_n[X]=E\left[(X-\mu)^n\right]=\underset{\Theta}{E}\left[E\left[(X-\mu)^n\middle|\Theta\right]\right]=\underset{\Theta}{E}\left[E\left[\{(X-\mu_\Theta)+(\mu_\Theta-\mu)\}^n\middle|\Theta\right]\right]$$

Next we expand this according to the binomial theorem:

$$\begin{aligned}
M_n[X]&=\underset{\Theta}{E}\left[E\left[\{(X-\mu_\Theta)+(\mu_\Theta-\mu)\}^n\middle|\Theta\right]\right]\\
&=\underset{\Theta}{E}\left[E\left[\sum_{k=0}^{n}\binom{n}{k}(X-\mu_\Theta)^{n-k}(\mu_\Theta-\mu)^k\middle|\Theta\right]\right]\\
&=\sum_{k=0}^{n}\binom{n}{k}\underset{\Theta}{E}\left[E\left[(X-\mu_\Theta)^{n-k}(\mu_\Theta-\mu)^k\middle|\Theta\right]\right]\\
&=\sum_{k=0}^{n}\binom{n}{k}\underset{\Theta}{E}\left[E\left[(X-\mu_\Theta)^{n-k}\middle|\Theta\right](\mu_\Theta-\mu)^k\right]\\
&=\sum_{k=0}^{n}\binom{n}{k}\underset{\Theta}{E}\left[M_{n-k}\left[X\middle|\Theta\right](\mu_\Theta-\mu)^k\right]
\end{aligned}$$

The fourth line follows from the third because $(\mu_\Theta-\mu)^k$ behaves as a constant within the nested expectation, and so can be taken outside it. Of these $n+1$ terms, the $(n–1)^{th}$ is zero, since $M_{n-(n-1)}\left[X\middle|\Theta\right]=M_1\left[X\middle|\Theta\right]=0$. Hence, in this binomial form, the n^{th} moment has n non-vanishing terms, namely:

$$M_n[X] = \sum_{k=0}^{n} \binom{n}{k} \underset{\Theta}{E}\left[M_{n-k}[X|\Theta](\mu_\Theta - \mu)^k\right]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[M_{n-k}[X|\Theta](\mu_\Theta - \mu)^k\right] + \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^n\right]$$

However, at this point we have not expressed the unconditional moment in terms of conditional moments and co-moments; we must express the expectation within the Σ operator as a co-moment. Letting $\zeta_{n-k} = \underset{\Theta}{E}\left[M_{n-k}[X|\Theta]\right]$, and remembering that first central moments are zero, we derive:

$$M_n[X] = \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[M_{n-k}[X|\Theta](\mu_\Theta - \mu)^k\right] + \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^n\right]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[\left\{M_{n-k}[X|\Theta] - \zeta_{n-k} + \zeta_{n-k}\right\}(\mu_\Theta - \mu)^k\right] + \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^n\right]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[\left\{M_{n-k}[X|\Theta] - \zeta_{n-k}\right\}(\mu_\Theta - \mu)^k\right]$$

$$+ \sum_{k=1}^{n-2} \binom{n}{k} \zeta_{n-k} \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^k\right] + \underset{\Theta}{E}\left[(\mu_\Theta - \mu)^n\right]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{CM}\left[M_{n-k}[X|\Theta], \overbrace{\mu_\Theta, \ldots, \mu_\Theta}^{k \text{ times}}\right]$$

$$+ \sum_{k=1}^{n-2} \binom{n}{k} \zeta_{n-k} \underset{\Theta}{M_k}[\mu_\Theta] + \underset{\Theta}{M_n}[\mu_\Theta]$$

$$= \underset{\Theta}{E}\left[M_n[X|\Theta]\right] + \sum_{k=1}^{n-2} \binom{n}{k} \underset{\Theta}{CM}\left[M_{n-k}[X|\Theta], \overbrace{E[X|\Theta], \ldots, E[X|\Theta]}^{k \text{ times}}\right]$$

$$+ \sum_{k=2}^{n-2} \binom{n}{k} \underset{\Theta}{E}\left[M_{n-k}[X|\Theta]\right] \underset{\Theta}{M_k}\left[E[X|\Theta]\right] + \underset{\Theta}{M_n}\left[E[X|\Theta]\right].$$

Hence, for $n \geq 2$, to express the n^{th} unconditional moment in the desired conditional form requires $n + \max(0, n-3)$ non-vanishing terms. The first Σ operator does not come into play until $n \geq 3$, and the second until $n \geq 4$. At least the complexity does not increase after the fourth moment.

The second, third, and fourth moments follow readily from the general formula:

$$\begin{aligned} Var[X] &= M_2[X] \\ &= \underset{\Theta}{E}[M_2[X|\Theta]] + \underset{\Theta}{M_2}[E[X|\Theta]] \\ &= \underset{\Theta}{E}[Var[X|\Theta]] + \underset{\Theta}{Var}[E[X|\Theta]] \qquad \text{cf. Klugman [1998], 393} \end{aligned}$$

$$\begin{aligned} Skew[X] &= M_3[X] \\ &= \underset{\Theta}{E}[M_3[X|\Theta]] + 3\underset{\Theta}{CM}[M_2[X|\Theta], E[X|\Theta]] + \underset{\Theta}{M_3}[E[X|\Theta]] \\ &= \underset{\Theta}{E}[Skew[X|\Theta]] + 3\underset{\Theta}{Cov}[Var[X|\Theta], E[X|\Theta]] + \underset{\Theta}{Skew}[E[X|\Theta]] \end{aligned}$$

$$\begin{aligned} Kurt[X] &= M_4[X] \\ &= \underset{\Theta}{E}[M_4[X|\Theta]] + 4\underset{\Theta}{CM}[M_3[X|\Theta], E[X|\Theta]] + 6\underset{\Theta}{CM}[M_2[X|\Theta], E[X|\Theta], E[X|\Theta]] \\ &\quad + 6\underset{\Theta}{E}[M_2[X|\Theta]]\underset{\Theta}{M_2}[E[X|\Theta]] + \underset{\Theta}{M_4}[E[X|\Theta]] \\ &= \underset{\Theta}{E}[Kurt[X|\Theta]] + 4\underset{\Theta}{Cov}[Skew[X|\Theta], E[X|\Theta]] + 6\underset{\Theta}{Coskew}[Var[X|\Theta], E[X|\Theta], E[X|\Theta]] \\ &\quad + 6\underset{\Theta}{E}[Var[X|\Theta]]\underset{\Theta}{Var}[E[X|\Theta]] + \underset{\Theta}{Kurt}[E[X|\Theta]] \end{aligned}$$

4. MOMENTS VERSUS CUMULANTS

That the conditional expression of the n^{th} moment requires $n + \max(0, n-3)$ terms indicates a "bend in the road" between $n = 3$ and $n = 4$. It is hardly coincidental that moments beyond the third are not additive. If X and Y are independent random variables with means μ and ν, the n^{th} moment of their sum is:

$$
\begin{aligned}
M_n[X+Y] &= E\left[\{(X+Y)-(\mu+\nu)\}^n\right] = E\left[\{(X-\mu)+(Y-\nu)\}^n\right] \\
&= \sum_{k=0}^{n}\binom{n}{k}E\left[(X-\mu)^{n-k}(Y-\nu)^k\right] \\
&= \sum_{k=0}^{n}\binom{n}{k}E\left[(X-\mu)^{n-k}\right]E\left[(Y-\nu)^k\right] \\
&= \sum_{k=0}^{n}\binom{n}{k}M_{n-k}[X]M_k[Y] \\
&= M_n[X]+M_n[Y]+\sum_{k=1}^{n-1}\binom{n}{k}M_{n-k}[X]M_k[Y] \\
&= M_n[X]+M_n[Y]+\sum_{k=2}^{n-2}\binom{n}{k}M_{n-k}[X]M_k[Y]
\end{aligned}
$$

The Σ operator disrupts the additivity when its range is non-empty, i.e., when $n-2 \geq 2$, or $n \geq 4$. So, with independence, the first three moments are additive; in the fourth moment the term $\binom{4}{2}M_{4-2}[X]M_2[Y] = 6M_2[X]M_2[Y] = 6Var[X]Var[Y]$ disrupts the additivity, a term analogous to $6\underset{\Theta}{E}\left[Var[X|\Theta]\right]\underset{\Theta}{Var}\left[E[X|\Theta]\right]$ in the kurtosis formula.

Nevertheless, adjustments to moments higher than the third can obviate the disruption and restore additivity. Such adjusted moments are known as cumulants. The fourth-order adjustment is:

$$M_4[X+Y]-3M_2[X+Y]^2 = M_4[X]+M_4[Y]+6M_2[X]M_2[Y]-3M_2[X+Y]^2$$
$$= M_4[X]+M_4[Y]+6M_2[X]M_2[Y]-3(M_2[X]+M_2[Y])^2$$
$$= M_4.[X]-3M_2[X]^2+M_4[Y]-3M_2[Y]^2.$$

Thus, the fourth cumulant, $\kappa_4[X]=M_4[X]-3M_2[X]^2 = Kurt_4[X]-3Var[X]^2$ is additive.[5]

Since the collective risk model involves sums of independent random variables, the fourth cumulant, which we shall call the excess kurtosis, will prove more useful than the fourth moment. Its conditional expression is:

$$XsKurt[X]=Kurt[X]-3Var[X]^2$$
$$= \underset{\Theta}{E}[Kurt[X|\Theta]]+4\underset{\Theta}{Cov}[Skew[X|\Theta],E[X|\Theta]]+6\underset{\Theta}{Coskew}[Var[X|\Theta],E[X|\Theta],E[X|\Theta]]$$
$$+6\underset{\Theta}{E}[Var[X|\Theta]]\underset{\Theta}{Var}[E[X|\Theta]]+\underset{\Theta}{Kurt}[E[X|\Theta]]-3\left(\underset{\Theta}{E}[Var[X|\Theta]]+\underset{\Theta}{Var}[E[X|\Theta]]\right)^2$$
$$= \underset{\Theta}{E}[Kurt[X|\Theta]]+4\underset{\Theta}{Cov}[Skew[X|\Theta],E[X|\Theta]]+6\underset{\Theta}{Coskew}[Var[X|\Theta],E[X|\Theta],E[X|\Theta]]$$
$$+\underset{\Theta}{Kurt}[E[X|\Theta]]-3\underset{\Theta}{E}[Var[X|\Theta]]^2-3\underset{\Theta}{Var}[E[X|\Theta]]^2$$
$$= \underset{\Theta}{E}[Kurt[X|\Theta]]-3\underset{\Theta}{E}\left[Var[X|\Theta]^2\right]+4\underset{\Theta}{Cov}[Skew[X|\Theta],E[X|\Theta]]$$
$$+6\underset{\Theta}{Coskew}[Var[X|\Theta],E[X|\Theta],E[X|\Theta]]+\underset{\Theta}{Kurt}[E[X|\Theta]]-3\underset{\Theta}{Var}[E[X|\Theta]]^2$$
$$+3\underset{\Theta}{E}\left[Var[X|\Theta]^2\right]-3\underset{\Theta}{E}[Var[X|\Theta]]^2$$
$$= \underset{\Theta}{E}[XsKurt[X|\Theta]]+4\underset{\Theta}{Cov}[Skew[X|\Theta],E[X|\Theta]]$$
$$+6\underset{\Theta}{Coskew}[Var[X|\Theta],E[X|\Theta],E[X|\Theta]]+\underset{\Theta}{XsKurt}[E[X|\Theta]]+3\underset{\Theta}{Var}[Var[X|\Theta]]$$

[5] The formulas for higher-order cumulants become increasingly more complicated. The reader can verify the additivity of the next two cumulants according to the definitions (cf. Section 6):

$$\kappa_5[X]=M_5[X]-10Var[X]Skew[X]$$

$$\kappa_6[X]=M_6[X]-15Var[X]\kappa_4[X]-10Skew[X]^2-15Var[X]^3.$$

5. MOMENTS OF THE COLLECTIVE RISK MODEL

The collective risk model, which casualty actuaries must study for their examinations, is stock-in-trade, especially in the field of reinsurance. It considers aggregate loss S as the sum of a random number N of independent, identically distributed claims: $S = X_1 + \ldots + X_N$. Here we will apply our conditional formulas[6] to derive the first four cumulants of S in terms of those of X and N. Because the X_i are independent and identically distributed, as well as due to the additivity of cumulants, $E[S|N] = N E[X]$, $Var[S|N] = NVar[X]$, $Skew[S|N] = N Skew[X]$, and $XsKurt[S|N] = N XsKurt[X]$. Because $E[X]$, $Var[X]$, $Skew[X]$, and $XsKurt[X]$ are constants, we may remove them from moments conditional upon N, being careful to raise them to the power of the conditional moments.

The first cumulant, the mean, is trivial: $E[S] = \underset{N}{E}[E[S|N]] = \underset{N}{E}[N E[X]] = E[N]E[X]$.

For the second, the variance:

$$\begin{aligned} Var[S] &= \underset{N}{E}[Var[S|N]] + \underset{N}{Var}[E[S|N]] \\ &= \underset{N}{E}[NVar[X]] + \underset{N}{Var}[N E[X]] \\ &= E[N]Var[X] + Var[N]E[X]^2. \end{aligned}$$

Every actuary at some time learned this formula; to many it remains familiar.

However, the third moment is not studied, and hence, not commonly known:

$$\begin{aligned} Skew[S] &= \underset{N}{E}[Skew[S|N]] + 3\underset{N}{Cov}[Var[S|N], E[S|N]] + \underset{N}{Skew}[E[S|N]] \\ &= \underset{N}{E}[N Skew[X]] + 3\underset{N}{Cov}[N\,Var[X], N E[X]] + \underset{N}{Skew}[N E[X]] \\ &= E[N]Skew[X] + 3Cov[N, N]Var[X]E[X] + Skew[N]E[X]^3 \\ &= E[N]Skew[X] + 3Var[N]Var[X]E[X] + Skew[N]E[X]^3 \end{aligned}$$

6 We will change the nomenclature of these formulas so as to agree with that of the collective risk model, i.e., S will appear instead of X, and N instead of Θ. In this section X will represent the severity of a claim.

Nonetheless, this formula appears in Patrik [1996, 377] and Klugman [1998, 298].

Last, we derive the excess kurtosis, whose formula we have not seen in print before:

$$\begin{aligned}
XsKurt[S] &= \underset{N}{E}\big[XsKurt[S|N]\big] + 4\underset{N}{Cov}\big[Skew[S|N], E[S|N]\big] \\
&\quad + 6\underset{N}{Coskew}\big[Var[S|N], E[S|N], E[S|N]\big] + \underset{N}{XsKurt}\big[E[S|N]\big] + 3\underset{N}{Var}\big[Var[S|N]\big] \\
&= \underset{N}{E}\big[N\,XsKurt[X]\big] + 4\underset{N}{Cov}\big[N\,Skew[X], N\,E[X]\big] \\
&\quad + 6\underset{N}{Coskew}\big[N\,Var[X], N\,E[X], N\,E[X]\big] + \underset{N}{XsKurt}\big[N\,E[X]\big] + 3\underset{N}{Var}\big[N\,Var[X]\big] \\
&= E[N]XsKurt[X] + 4Cov[N,N]Skew[X]E[X] \\
&\quad + 6Coskew[N,N,N]Var[X]E[X]^2 + XsKurt[N]E[X]^4 + 3Var[N]Var[X]^2 \\
&= E[N]XsKurt[X] + 4Var[N]Skew[X]E[X] \\
&\quad + 6Skew[N]Var[X]E[X]^2 + XsKurt[N]E[X]^4 + 3Var[N]Var[X]^2.
\end{aligned}$$

6. THE CUMULANT GENERATING FUNCTION

The moment generating function of a sum of independent random variables equals the product of their moment generating functions. Since logarithms convert multiplication into addition, it is natural to consider the logarithm of the moment generating function, which has come to be known as the cumulant generating function ψ (c.g.f.), i.e., $\psi_X(t) = \ln E\left[e^{tX}\right]$. Its derivatives at zero are called cumulants:[7] $\kappa_i[X] = \psi_X^{[i]}(0)$. If the X_i are independent of one another:

$$\psi_{\sum_i X_i}(t) = \ln E\left[e^{t\sum_i X_i}\right] = \ln E\left[\prod_i e^{tX_i}\right] = \ln \prod_i E\left[e^{tX_i}\right] = \sum_i \ln E\left[e^{tX_i}\right] = \sum_i \psi_{X_i}(t).$$

[7] In Section 4 we introduced cumulants as "moments adjusted to restore additivity." This hardly suffices for a definition, and we have not proven the existence and the uniqueness of the adjustment. The derivatives of the c.g.f. at zero constitute a proper definition of cumulant, and the Taylor-series argument of this section can be made into a rigorous proof of the uniqueness of the adjustment.

Since differentiation is a linear operator, the cumulant of a sum of independent random variables equals the sum of the cumulants of the random variables. The first three cumulants equal the mean, the variance, and the skewness. But equality ceases with the fourth cumulant: $\kappa_4[X] = M_4[X] - 3M_2[X]^2 = E\left[(X-\mu)^4\right] - 3Var[X]^2$ (Daykin [1994, 23] and Halliwell [2003, 65]). Here we will show the relevance of the c.g.f., to (1) the expression of cumulants in terms of moments and (2) the moments of the collective risk model.

First, the Taylor-series expansion of the c.g.f. embeds the cumulants:

$$\psi_X(t) = \psi_X(0) + \sum_{j=1}^{\infty} \psi_X^{[j]}(0) t^j / j! = 0 + \sum_{j=1}^{\infty} \kappa_j[X] t^j / j!.$$

The central moments of X, $M_n[X] = E\left[(X-\mu)^n\right]$, are similar coefficients in the Taylor-series expansion of the moment generating function of $X - \mu$:

$$E\left[e^{t(X-\mu)}\right] = 1 + \sum_{j=1}^{\infty} M_j[X] t^j / j! = 1 + 0 + \sum_{j=2}^{\infty} M_j[X] t^j / j!.$$

We can combine these two equations to relate the cumulants and the moments:

$$\sum_{j=1}^{\infty} \kappa_j[X] t^j / j! = \psi_X(t) = \ln E\left[e^{tX}\right] = \ln E\left[e^{t(X-\mu)+t\mu}\right] = t\mu + \ln E\left[e^{t(X-\mu)}\right]$$
$$= t\mu + \ln\left(1 + \sum_{j=2}^{\infty} M_j[X] t^j / j!\right).$$

But the logarithm has its own Taylor-series expansion for $-1 < x \le 1$, viz.:

$$\ln(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \ldots = \sum_{k=1}^{\infty} (-1)^{k-1} x^k / k.$$

So the relationship can be expressed as two polynomials in *t*:

$$\sum_{j=1}^{\infty}\kappa_j[X]t^j/j! = t\mu + \ln\left(1+\sum_{j=2}^{\infty}M_j[X]t^j/j!\right)$$

$$= t\mu + \sum_{k=1}^{\infty}(-1)^{k-1}\left(\sum_{j=2}^{\infty}M_j[X]t^j/j!\right)^k \Big/ k$$

$$= \mu t^1/1! + \sum_{j=2}^{\infty}M_j[X]t^j/j! + \sum_{k=2}^{\infty}(-1)^{k-1}\left\{\sum_{j=2}^{\infty}M_j[X]t^j/j!\right\}^k \Big/ k.$$

Matching coefficients of identical polynomials must be equal. It is the last expression on the right side of the final equation that complicates the matching; however, it is quartic and higher in *t*. Hence, the first three cumulants must be the mean, the variance, and the skewness. And the formula for higher cumulants begins as $\kappa_j[X] = M_j[X] + \dots$.

As an example of higher-order matching, we will derive the kurtosis formula. A fourth power of *t* arises in the last expression only from $k = 2$ powers of two, or as 2+2:

$$\kappa_4[X]t^4/4! = M_4[X]t^4/4! + (-1)^{2-1}\{(M_2[X]t^2/2!)(M_2[X]t^2/2!)\}/2$$

$$= M_4[X]t^4/4! - M_2[X]^2 t^4/8$$

$$\kappa_4[X] = M_4[X] - 3M_2[X]^2.$$

The fifth cumulant is a little more complicated, still involving $k = 2$, but obtained twice as 2+3 and 3+2. The sixth cumulant involves $k = 2$ as 2+4, 3+3, and 4+2, as well as $k = 3$ as 2+2+2. This c.g.f. technique is arguably the easiest way to derive the formulas of footnote 5.

Second, we will derive the c.g.f. of the collective risk model $S = X_1 + \dots X_N$, being mindful of the change in nomenclature (cf. footnote 6):

$$\psi_S(t) = \ln e^{\psi_S(t)} = \ln \underset{N}{E}\left[e^{\psi_{S|N}(t)}\right] = \ln \underset{N}{E}\left[e^{\psi_X(t)N}\right] = \psi_N(\psi_X(t)) = (\psi_N \circ \psi_X)(t).$$

So the c.g.f. of the aggregate loss is the composition of the cumulant generating functions of frequency and severity (Daykin [1994, 59]). This is the most elegant way to derive the aggregate cumulants, and it is more efficient than the conditional-moment technique of Section 5. To show this, we will derive the first two moments.

$$\psi_S(t) = \psi'_N(\psi_X(t))\psi'_X(t)$$
$$\therefore E[S] = \psi'_S(0) = \psi'_N(\psi_X(0))\psi'_X(0) = \psi'_N(0)\psi'_X(0) = E[N]E[X]$$

$$\psi''_S(t) = \psi'_N(\psi_X(t))\psi''_X(t) + \psi''_N(\psi_X(t))\psi'_X(t)^2$$
$$\therefore Var[S] = \psi''_S(0) = \psi'_N(0)\psi''_X(0) + \psi''_N(0)\psi'_X(0)^2 = E[N]Var[X] + Var[N]E[X]^2.$$

Curious and ambitious readers, performing the third and fourth derivatives, can verify the formulas in Section 5 for the aggregate skewness and excess kurtosis.

7. CONCLUSION

We have shown how unconditional moments can be expressed in terms of conditional moments and co-moments. Adjusting moments into cumulants allowed us to form fairly simple formulas for the skewness and the excess kurtosis of the collective risk model. These formulas can also be derived directly from the cumulant-generating function. Actuaries who have been reluctant to apply the method of moments to just the first two moments of the collective risk model can now with these formulas fit more versatile distributions to more than two moments. One ought to be more comfortable with extrapolations into the right tail of an aggregate loss distribution after having considered its skewness and kurtosis.

Aside from the collective risk model, a conditioning partition Θ can change the moments of a sum of independent random variables without changing their unconditional moments. The appendix shows how this can be done in loss reserving. Moreover, if some amount of capital or risk margin were allocated to a moment, conditioning would allow a sub-allocation to the partitions.

Acknowledgment

The author gratefully acknowledges the help of reviewers Jason Cook, FCAS, and Evan Glisson, FCAS, from whose comments this paper has much improved.

REFERENCES

[1.] Daykin, C. D., T. Pentikäinen, and M. Pesonen, *Practical Risk Theory for Actuaries*, London: Chapman & Hall, 1994.

[2.] Halliwell, Leigh J., "The Valuation of Stochastic Cash Flows," *CAS Forum* (Reinsurance Call Papers, Spring 2003), pp. 1-68, www.casact.org/pubs/forum/03spforum/03spf001.pdf.

[3.] Klugman, Stuart A., H. Panjer, and G. Willmot, *Loss Models: From Data to Decisions*, New York: John Wiley & Sons, 1998.

[4.] Patrik, Gary S., "Reinsurance," *Foundations of Casualty Actuarial Science*, Third Edition, Casualty Actuarial Society, 1996, pp. 283-380.

APPENDIX A

Conditional Probability and Claim Development

A recent assignment spurred our interest in the subject of this paper. We had a list of the case reserves of about 200 claims, and were satisfied that the total IBNER[8] for them was zero, i.e., $E[\text{IBNER}] = 0$. But in addition, we wanted some measure of the variance. The claims stemmed from an unusual exposure, and we deemed no other data sources appropriate. Since we assumed the average development to be zero, the claim list itself could serve as an empirical distribution f_X with moments $\mu \pm \sigma$. Regarding the *n* claims as independent, we might decide the moments of the total unpaid loss (i.e., case plus IBNER) to be $n\mu \pm \sqrt{n}\sigma$, or total IBNER to be $0 \pm \sqrt{n}\sigma$. But this ignores the likelihood of rank correlation, i.e., that after development large claims tend to stay large, and small claims tend to stay small. Hence, $\sqrt{n}\sigma$ is a maximal value.

Therefore, we decided to order the claims by their case reserves and to stratify them into 10 groups of approximately 20 claims. Belonging to a stratum is the event Θ that conditions a claim's probability density as $f_{X|\Theta}$. Since stratification provides no new information, $f_X(x) = \underset{\Theta}{E}\left[f_{X|\Theta}(x)\right]$. Then we assumed that each claim would develop as follows: with probability *p* its distribution would remain that of its stratum and with probability $q = 1 - p$ it would migrate randomly.

Consequently, the distribution of a developed claim is a mixture of distributions; with probability *p* the developed claim is distributed as $f_{X|\Theta}$ and with probability *q* as f_X. Let *Y* be the developed amount of claim *X*. Mixing is easy with moment generating functions.

[8] IBNER means "Incurred But Not Enough Reported (or Reserved)." Cf. Patrik [1996], 350.

The moment generating function of Y conditional upon the stratum of X is $M_{Y|\Theta}(t) = pM_{X|\Theta}(t) + qM_X(t)$. The overall, or unconditional, moment generating function of a developed claim is:

$$\begin{aligned} M_Y(t) &= \underset{\Theta}{E}\left[M_{Y|\Theta}(t)\right] \\ &= \underset{\Theta}{E}\left[pM_{X|\Theta}(t) + qM_X(t)\right] \\ &= p\underset{\Theta}{E}\left[M_{X|\Theta}(t)\right] + qM_X(t) \\ &= pM_X(t) + qM_X(t) \\ &= M_X(t). \end{aligned}$$

Since equality of moment generating functions implies identical distributions, Y is distributed as X. Since we have provided no new information, this "conservation of distribution" is fitting.

But the reader may now be wondering how the variance of total IBNER can change despite the conservation of the overall distribution. The paradox is resolved with a distinction: variance pertains to the *sum* of claims, whereas conservation pertains to their *mixture*, more accurately, to the mixture of their distributions. The overall or unconditional variance $Var[X]$ is conserved, but its apportionment between $\underset{\Theta}{E}\left[Var[X|\Theta]\right]$ and $\underset{\Theta}{Var}\left[E[X|\Theta]\right]$ depends on Θ. At the one extreme, a blunt or non-discriminating stratification Θ tells nothing about X: $Var[X|\Theta] = Var[X]$. In this case:

$$\begin{aligned} \underset{\Theta}{Var}\left[E[X|\Theta]\right] &= Var[X] - \underset{\Theta}{E}\left[Var[X|\Theta]\right] \\ &= Var[X] - \underset{\Theta}{E}\left[Var[X]\right] \\ &= Var[X] - Var[X] \\ &= 0. \end{aligned}$$

Conversely, if the variance of the conditional mean is zero, $E[X|\Theta]$ must be constant, or $E[X|\Theta] = E[X]$. So the conditional distributions of a blunt stratification tend to be indistinguishable as to their first two moments. In this case the variance of the sum tends toward the maximal $\sqrt{n}\sigma$. At the opposite extreme, Θ is so fine or discriminating that $Var[X|\Theta] = 0$. Then:

$$\begin{aligned} \underset{\Theta}{Var}[E[X|\Theta]] &= Var[X] - \underset{\Theta}{E}[Var[X|\Theta]] \\ &= Var[X] - \underset{\Theta}{E}[0] \\ &= Var[X]. \end{aligned}$$

This means that the all the variance is between the strata, no variance is within a stratum. In this case the variance of the sum tends toward the minimal value of zero. To borrow and mix notions from optics and credibility, the blunt stratification passes the white light of zero credibility; the fine stratification like a prism refracts light into the spectrum of full credibility.

Since we will be conditioning on migration M, we will drop Θ and speak of the i^{th} stratum. Let there be s strata, and let $\pi_i > 0$ be the probability for a claim to be in the i^{th} stratum, as determined by the actual portion of claims in that stratum. Though the strata need not to be balanced, or of equal population, $\sum_{i=1}^{s} \pi_i = 1$. We may model developed claim Y_i of the i^{th} stratum as follows. Randomly draw one undeveloped claim from each stratum's distribution; these $X_1, \ldots, X_s$ are independent. Then form an "unstratified" or average claim X as the choice of X_j with probability π_j. Finally, flip a "Bernoulli coin" with probability p of heads. If the coin lands heads, let Y_i equal X_i; otherwise, let it equal X.

The mean of the developed claim is $E[Y_i] = \underset{\text{M}}{E}\left[E[Y_i|\text{M}]\right] = pE[X_i] + qE[X]$. According to the formula of Section 3, the variance is:

$$\begin{aligned} Var[Y_i] &= \underset{\text{M}}{E}\left[Var[Y_i|\text{M}]\right] + \underset{\text{M}}{Var}\left[E[Y_i|\text{M}]\right] \\ &= pVar[X_i] + qVar[X] \\ &\quad + p(E[X_i] - E[Y_i])^2 + q(E[X] - E[Y_i])^2 \\ &= pVar[X_i] + qVar[X] \\ &\quad + p(qE[X_i] - qE[X])^2 + q(pE[X] - pE[X_i])^2 \\ &= pVar[X_i] + qVar[X] + pq(E[X_i] - E[X])^2. \end{aligned}$$

Since considerations of rank correlation drew us to this model, we should also determine the covariance of Y_i with X_i. Taking the next formula without proof,[9] we have:

$$Cov[Y_i, X_i] = \underset{\text{M}}{E}\left[Cov[Y_i, X_i|\text{M}]\right] + \underset{\text{M}}{Cov}\left[E[Y_i|\text{M}], E[X_i|\text{M}]\right].$$

The second term on the right side of the equation is zero. For the migration M does not affect the expectation of X_i, and the covariance of something with a constant is zero. Hence, $Cov[Y_i, X_i] = \underset{\text{M}}{E}\left[Cov[Y_i, X_i|\text{M}]\right]$. In the following reduction, we must consider that the random migration can return (P) with probability π_i to the i^{th} stratum. Again, a covariance term becomes zero due to the immunity of X_i to **P**:

9 The proof hinges on $Cov[X, Y|\Theta] = \underset{\Theta}{E}\left[E\left[\{(X - \mu_\Theta) - (\mu_\Theta - \mu)\}\{(Y - \nu_\Theta) - (\nu_\Theta - \nu)\}|\Theta\right]\right]$.

$$\begin{aligned}
Cov[Y_i, X_i] &= \underset{\mathrm{M}}{E}\left[Cov[Y_i, X_i|\mathrm{M}]\right] \\
&= pCov[X_i, X_i] + qCov[X, X_i] \\
&= pCov[X_i, X_i] + q\left\{\underset{\mathrm{P}}{E}\left[Cov[X, X_i|\mathrm{P}]\right] + \underset{\mathrm{P}}{Cov}\left[E[X|\mathrm{P}], E[X_i|\mathrm{P}]\right]\right\} \\
&= pCov[X_i, X_i] + q\underset{\mathrm{P}}{E}\left[Cov[X, X_i|\mathrm{P}]\right] \\
&= pCov[X_i, X_i] + q\sum_{j=1}^{s}\pi_j Cov[X_j, X_i] \\
&= pCov[X_i, X_i] + q\pi_i Cov[X_i, X_i] \\
&= (p + q\pi_i)Var[X_i].
\end{aligned}$$

The covariance of the developed claim amount with the undeveloped is positive, but the correlation coefficient is more informative:

$$Corr[Y_i, X_i] = \frac{Cov[Y_i, X_i]}{\sqrt{Var[Y_i]Var[X_i]}} = \frac{(p + q\pi_i)Var[X_i]}{\sqrt{Var[Y_i]Var[X_i]}} = (p + q\pi_i)\sqrt{\frac{Var[X_i]}{Var[Y_i]}}.$$

The correlation increases with respect to *p*, the probability that the distribution of a developed claim remains that of its stratum, from a minimum of $\pi_i\sqrt{\frac{Var[X_i]}{Var[X]}}$ for $p = 0$ to a maximum of 1 for $p = 1$. It seems that $Corr[Y_i, X_i] \approx p$. However, this is Pearson correlation, whereas we are concerned with rank, or Spearman, correlation.

Because an analytic answer eluded us, we resorted to simulation. Keeping with the assignment that spurred our interest, we simulated 1,000 iterations of the "development" of the integers from 1 to 200 in $s = 10$ groups of 20 consecutive integers over a range of non-migration probabilities *p* from 0% to 100% in steps of 5%. We randomly permuted the integers within each group – this alone would suffice if $p = 1$ and inter-group migrations were impossible. But then we flipped the Bernoulli coin for each integer, marked which places were the migrating "tails," and randomly permuted among those places their integers.

Then we calculated the rank correlation for that iteration, and averaged it over all the iterations. The table below contains the result:

#Iter	#Goups	#InGrp	*p*	RankCorr
1,000	10	20	0%	0.000
1,000	10	20	5%	0.046
1,000	10	20	10%	0.100
1,000	10	20	15%	0.146
1,000	10	20	20%	0.201
1,000	10	20	25%	0.250
1,000	10	20	30%	0.302
1,000	10	20	35%	0.349
1,000	10	20	40%	0.396
1,000	10	20	45%	0.446
1,000	10	20	50%	0.497
1,000	10	20	55%	0.550
1,000	10	20	60%	0.601
1,000	10	20	65%	0.648
1,000	10	20	70%	0.696
1,000	10	20	75%	0.747
1,000	10	20	80%	0.797
1,000	10	20	85%	0.844
1,000	10	20	90%	0.895
1,000	10	20	95%	0.944
1,000	10	20	100%	0.990

Indeed, it seems that the rank correlation approximates *p*. Nonetheless, it cannot exactly equal *p*. For at a near 100% probability of not migrating, permutation within each group still disrupts a perfect correlation. Therefore, we suspect RankCorr(*p*) to start out at zero with a slope of unity, but to be slightly concave (i.e., to have a negative second derivative) so that it loses ground to *p* as *p* increases to one.

In sum, as *p*, the probability of not migrating (i.e., the probability for the distribution of a claim to remain that of its stratum) approaches zero, stratification becomes irrelevant. Regardless of how the claims are stratified, they will all develop according to the overall distribution. This will produce an aggregate standard deviation approaching the maximal $\sqrt{n}\sigma$. And if there were only one stratum, migration would be from overall to overall, and the aggregate standard deviation again would be $\sqrt{n}\sigma$. But as *p* increases, and as the strata become narrower, the aggregate standard deviation decreases. In the extreme, with one claim per stratum (better, with zero variance within each stratum) and $p = 1$, the aggregate standard deviation is zero.

Pondering these relations with two moments led us to the idea of adding higher moments to the conditional distributions, and thence to treating the higher moments of the collective risk model. Although we do not intend for this to be a paper on a new development method, the reader can see how this claim-by-claim method can be employed to apportion moments of loss that mesh with any desired aggregate moments, as well as to obtain useful subtotals, e.g., by accident year.

A Method for Efficient Simulation of the Collective Risk Model

DAVID L. HOMER AND RICHARD A. ROSENGARTEN

Abstract

The Collective Risk Model (CRM) constructs aggregate losses from a claim count distribution and a claim size distribution. The aggregate losses are $Z = X_1 + ... + X_N$, where the X_i are independent and identically distributed as well as independent from the claim counts N.

Simulating individual claims can be a lengthy process when the expected number of claims is large. Often it is sufficient to collect only individual claims greater than some threshold τ together with the aggregate smaller claims. This is the case when modeling the effects of excess of loss reinsurance.

The simulation run time can be significantly reduced, therefore, by simulating large losses individually and small losses in aggregate. The challenge in doing this is to preserve the risk characteristics of the original CRM, because the small losses and the large losses are not generally independent.

This paper shows how to do this by first simulating the total claim counts and then conditionally simulating both the individual large losses and an approximation to the aggregate small losses. In the case where the claim count distribution is a mixed Poisson, it is shown that the distribution of losses simulated from this method converges to the CRM distribution. This result is a generalization of the principle that the limiting behavior of a mixed Poisson CRM is controlled by the mixing distribution.

1 Introduction

The Collective Risk Model (CRM) constructs aggregate losses from a claim count distribution and a claim size distribution. The aggregate losses are $Z = X_1 + ... + X_N$, where the X_i are independent and identically distributed as well as independent from the claim counts N.

Simulating individual claims can be a lengthy process when the expected number of claims is large. Often it is sufficient to collect only individual claims greater than some threshold τ together with the aggregate smaller claims. This is the case when modeling the effects of excess of loss reinsurance, for example.

The simulation run time can be significantly reduced, therefore, by simulating large losses individually and small losses in aggregate. The challenge in doing this is to preserve the risk characteristics of the original CRM, because the small losses and the large losses are not generally independent. This paper shows how to do this by first simulating the total claim counts and then conditionally simulating both the individual large losses and an approximation to the aggregate small losses. The small losses are drawn from a *Conditional Aggregate Distribution (CAD)* so this method is referred to as the CAD method.

Section 2 provides a brief review of other methods of reflecting the dependence between large and small losses.

After providing some notation, definitions, and basic facts, Section 3 describes the CAD method for generating large and small losses in the CRM. An illustrative example shows that the method can be highly accurate.

Section 4 discusses mixed Poisson claim count distributions and proves a theorem that shows the distribution simulated from the CAD method converges to the CRM distribution when the claim counts arise from a mixed Poisson distribution. This provides theoretical support for the practical observation that the CAD method seems to work. Additionally, the theorem supports two other practical observations: (1) the particular choice of the conditional aggregate distribution used to approximate the small losses is to some extent immaterial and (2) the mixing distribution seems to control the overall aggregate distribution. These are related to ideas presented by Mildenhall [12] and their connections are discussed.

Section 5 provides a reinsurance application that uses only the total aggregate loss mean and variance together with large the claim size and count distributions.

Section 6 illustrates a multi-line example.

2 Brief Review of Methods for Reflecting Large-Small Dependence

Dependence between large and small losses as well as more general methods of reflecting dependencies have been discussed by several authors. The methods include: recursion, Fourier Transform, numerical integration, and simulation with copulas, as well as the Iman-Connover method [5].

Using two-dimensional Panjer recursions, Walhin [17] illustrates how different results are obtained when small and large losses are modelled independently as opposed to the dependence structure implicit in the CRM. Homer and Clark [3] perform similar calculations using two-dimensional Fourier Transforms. These methods are powerful and convenient when the expected claim counts are relatively small.

Other techniques discuss more generally the modeling of dependencies between random variates, but not specifically between the large and small losses of the CRM. Homer [4] shows how to extend Heckman and Meyers' [2] numerical integration to two dimensions. Numerical integration works effectively when the claim counts are high but requires extensive programming and lacks the flexibility of simulation.

Dependencies can be imposed in simulation exercises with tools like copulas or the Iman-Connover method. Wang [22] and Venter [16] discuss the use of copulas and Mildenhall [12] generalizes the Iman-Connover method to provide additional dependence structures.

3 The Conditional Aggregate Distribution (CAD) Method

The basic idea is to simulate the total claim count N and then conditionally simulate the large claim count N_L. The small claim count N_S follows as $N - N_L$. Large claims are simulated individually. Small claims are conditionally simulated in the aggregate from an approximating distribution, the *conditional aggregate distribution*.

It will be helpful to establish some notation and recall some basic facts of the CRM in order to describe the CAD method and show how the losses from the CAD method reproduce various moments of the CRM losses as well as the correlation between large and small losses.

3.1 Notation

The CRM losses are $Z = X_1 + ... + X_N$ where the X_i are independent, identically distributed (iid) severities with common distribution $F_X(x)$, N is the random claim count with distribution $Q_N(n)$, and independent of the X_i.

The losses X_i are partitioned into losses smaller than some threshold τ and losses greater than or equal to τ. The small claim count is N_S and the large count N_L with $N = N_S + N_L$. The aggregate large losses are the sum of the individual large losses $Z_L = X_{L,1} + ... + X_{L,N_L}$ and similarly for small losses Z_S, with $Z = Z_S + Z_L$.

The distributions of the individual small and large claim sizes respectively are

$$F_{X_S}(x) = \frac{F_X(x)}{F_X(\tau)}, \quad x \in (0, \tau), \tag{1}$$

and

$$F_{X_L}(x) = \frac{F_X(x) - F_X(\tau)}{1 - F_X(\tau)}, \quad x \in [\tau, \infty). \tag{2}$$

The large claim count distribution conditional on N total claims is a Binomial distribution because the claim sizes are iid and independent from the claim counts:

$$\Pr(N_L = m|N = n) = B(n, m, q) = \binom{n}{m} q^m (1 - q)^{(n-m)}, \tag{3}$$

where $q = 1 - F_X(\tau)$ is the probability of a large loss.

Correlation of large and small losses: Large and small losses are correlated through the claim count random variable (r.v.). The value of the correlation coefficient [15] is given by

$$\rho(Z_S, Z_L) = \frac{q(1-q)\mathrm{E}[X_S]\mathrm{E}[X_L](\sigma^2(N) - \mathrm{E}[N])}{\sigma(Z_S)\sigma(Z_L)}, \quad (4)$$

where $\sigma(Y)$ denotes the standard deviation of the r.v. Y.

3.2 The CAD$_k$ Algorithm

The pseudo-code for a single trial is as follows:

1. Draw N the number of total claims from the total claim count distribution Q_N.
2. Draw N_L the number of large claims from the large claim count distribution conditional on N total claims using equation (3).
3. Set the small claims $N_S = N - N_L$.
4. Draw the individual large claims $\{X_1, ..., X_{N_L}\}$ from the claim size distribution conditional on $X_i > \tau$, given by equation (2).
5. Draw the aggregate small claims from a distribution parameterized by matching the first k moments of $Z_S|N_S$.

3.2.1 Preservation of Means, Variances and Correlations

To see how means, variances and large-small correlations are preserved consider how the large and small losses are constructed. The simulated losses in steps 1-4 are completely consistent with the CRM. In the last step an approximation is used: the small aggregate claims Z_S are simulated from an aggregate distribution with the matching k conditional moments. Denote this method with k matching moments by CAD$_k$. Further, let $\mathcal{F}$ represent the distributional family used in step 5, and set

$$\widehat{Z} := CAD_k(N, X, \mathcal{F})$$

to mean the total loss r.v. generated by CAD$_k$. Similarly, $\widehat{Z}_S$ is the small loss r.v. generated by CAD$_k$. The notation $\widehat{Z}_L$ is not needed since, by construction, $\widehat{Z}_L = Z_L$.

For $k \geq 2$, CAD$_k$ preserves the mean, variance, and correlation of large and small losses:

Claim 3.1 *For $j \leq k$, $E[\widehat{Z}_S^j] = E[Z_S^j]$, and for $k \geq 2$,*

$$\rho(\widehat{Z}_S, Z_L) = \rho(Z_S, Z_L). \quad (5)$$

Proof

$$\mathrm{E}[\widehat{Z}_S^j] = \mathop{\mathrm{E}}_{N,N_L}[\mathrm{E}[\widehat{Z}_S^j|N,N_L]] = \mathop{\mathrm{E}}_{N,N_L}[\mathrm{E}[Z_S^j|N,N_L]] = \mathrm{E}[Z_S^j], \tag{6}$$

by construction. To see that correlation is preserved, it suffices to show that $\mathrm{E}[\widehat{Z}_S Z_L] = \mathrm{E}[Z_S Z_L]$. This follows as above since $\widehat{Z}_S,\ Z_L$ are independent given $N,\ N_L$. □

3.2.2 Selecting a Conditional Aggregate Distribution

The central limit theorem promises that the conditional small losses are asymptotically normal, but in fairly typical insurance situations, the r.v. $Z_S|N_S$ will carry significant skewness. It seems natural, then, to consider non-normal two-parameter families as well as three-parameter families to match the conditional moments of the aggregate small claims; i.e., consider CAD_2 and CAD_3 models.

The statistics used for fitting are generally the mean, variance, and skewness. The mean, variance, and skewness of conditional small claims are given by:

$$\mathrm{E}[Z_S|N_S] = N_S \mathrm{E}[X_S], \tag{7}$$

$$\sigma^2(Z_S|N_S) = N_S \sigma^2(X_S), \tag{8}$$

$$\gamma(Z_S|N_S) = \gamma(X_S)/\sqrt{N_S}. \tag{9}$$

Table 10 of Appendix A shows the parameterizations and method of moment fits for various distributions. In several instances, a shift is used to provide an extra parameter. Section 4 develops some theory showing that the form of the conditional aggregate distribution is in some sense immaterial.

3.3 Basic Example

The following example provides a comparison between direct simulation of the CRM and simulation using the CAD.

The severity distribution is a 1ognormal ($\mu = 9$ and $\sigma = 2$) censored at \$1,000,000. The frequency distribution is a negative binomial (mean=526.99 and variance=17884). These are the same parameters used by Mildenhall in [12], section 4.1.

The conditional aggregate distribution is a lognormal. (See formulae in Appendix A.)

Tables 1 and 2 summarize the claim size and claim count distributions.

Table 1: Claim Size Distribution

Claim Size	Incremental Probability	Cumulative Probability
0	0.0%	0.0%
10,000	54.2%	54.2%
20,000	13.2%	67.4%
30,000	6.9%	74.4%
40,000	4.4%	78.8%
50,000	3.1%	81.9%
60,000	2.3%	84.2%
70,000	1.8%	86.0%
80,000	1.4%	87.4%
90,000	1.2%	88.6%
100,000	1.0%	89.6%
200,000	5.0%	94.6%
300,000	1.9%	96.5%
400,000	1.0%	97.4%
500,000	0.6%	98.0%
600,000	0.4%	98.4%
700,000	0.3%	98.7%
800,000	0.2%	98.9%
900,000	0.2%	99.1%
1,000,000	0.9%	100.0%

Table 2: Negative Binomial Parameters

Mean	526.99
Variance	17,885

Table 3 provides a comparison of percentiles and statistics for the aggregate small and large losses, while Table 4 compares the total losses. CRM large and CAD large losses are drawn from the same distribution so they only differ due to different simulations. CRM small and CAD small losses look equally close; the CAD approximation seems to work well. The correspondence in Table 4 suggests that the dependence structure is preserved and this is further supported by Table 5 which shows the simulated and theoretical correlation for large and small losses. Table 6 shows the improved run-time using methods programmed in R [14].

Table 3: CRM and CAD Simulated Losses

Cumulative Probability	CRM Small Losses	CAD Small Losses	CRM Large Losses	CAD Large Losses
1.0%	8.0	8.0	1.9	2.0
2.0%	8.7	8.8	2.4	2.5
3.0%	9.3	9.3	2.8	2.8
4.0%	9.7	9.7	3.1	3.1
5.0%	9.9	10.0	3.4	3.4
10.0%	11.1	11.2	4.3	4.3
20.0%	12.7	12.7	5.5	5.4
30.0%	13.9	13.9	6.4	6.4
40.0%	15.0	15.0	7.3	7.3
50.0%	16.1	16.1	8.2	8.1
60.0%	17.3	17.3	9.0	9.0
70.0%	18.5	18.5	10.0	9.9
80.0%	20.2	20.0	11.2	11.1
90.0%	22.5	22.4	13.1	13.0
95.0%	24.5	24.6	14.6	14.5
99.0%	28.7	28.8	18.0	17.8
99.9%	33.5	33.4	22.3	21.5
Mean	16.5	16.5	8.5	8.4
Std	4.5	4.5	3.5	3.4

Table 4: CRM and CAD Simulated Losses

Cumulative Probability	CRM Total Losses	CAD Total Losses
1.0%	11.3	11.2
2.0%	12.5	12.4
3.0%	13.3	13.3
4.0%	14.0	13.9
5.0%	14.6	14.6
10.0%	16.5	16.5
20.0%	19.0	18.9
30.0%	20.9	20.8
40.0%	22.7	22.6
50.0%	24.3	24.3
60.0%	26.2	26.1
70.0%	28.3	28.1
80.0%	30.7	30.6
90.0%	34.3	34.3
95.0%	37.5	37.3
99.0%	44.0	43.9
99.9%	53.1	51.5
Mean	25.0	24.9
Std	7.1	7.0

Table 5: Theoretical, CRM, and CAD Small-Large Linear Correlation

	Correlation
Theoretical	57.3%
CRM	58.4%
CAD	57.0%

Table 6: CRM and CAD Simulation Run-Times

Trial Count	CRM	CAD	x Faster
5,000	1.08	0.13	8.31
10,000	2.15	0.22	9.77
20,000	4.33	0.44	9.84

Before moving on to some underlying theory, we note several properties of the CAD method for loss simulation modeling:

1. It captures individual large losses.

2. It is easy to program (wtih Excel\@Risk, or in R, for example) with fast run times.

3. It works well no matter the size of $\lambda = \mathrm{E}[N]$ (as long as $\lambda_L = \mathrm{E}[N_L]$ is manageable.)

4. It reflects the joint distribution of large and small losses.

5. It can be adapted to situations with incomplete knowledge (specifically when the severity distribution is not known or assumed; see the example in Section 5).

6. It is easy to incorporate into complex models (For example, CAD can be used for multiple lines of business correlated via the claim count r.v.; see the example in Section 6).

4 CAD with the Mixed Poisson Claim Count

The losses simulated from the CAD method can be shown to converge to the losses in the CRM when the claim count is a mixed Poisson. The particular conditional aggregate distribution used is somewhat immaterial while the mixing distribution of the Poisson controls the unconditional aggregate shape.

This section discusses mixed Poisson distributions and then proves a convergence theorem for the losses simulated with the CAD method.

4.1 Mixed Poisson Claim Counts

A Mixed Poisson distribution is just a Poisson distribution with a random parameter. Formally,

Definition: N is a mixed Poisson r.v. (Q_N is a mixed Poisson distribution) if $N \sim$ Poisson(λG) for $\lambda = \mathrm{E}[N]$ and non-negative G such that $\mathrm{E}[G] = 1$ and $\sigma^2(G) = c$. In this case we write $N = MP(\lambda, G)$.

The r.v. G is referred to as the *mixing distribution*, and c the *contagion parameter*. Note that for $N = MP(\lambda, G)$,

$$\sigma^2(N) = \lambda(1 + c\lambda) \tag{10}$$

and

$$\gamma(N) = \frac{1 + c\lambda(3 + \lambda\sqrt{c}\gamma(G))}{\sqrt{\lambda}(1 + c\lambda)^{3/2}} . \tag{11}$$

Thus mixed Poisson claim counts carry positive contagion in the sense that $c \geq 0$ and the *variance-to-mean ratio* $d = (1 + c\lambda) \geq 1$.

A convenient aspect of the mixed Poisson for ground-up claims is that large and small claim counts are also mixed Poisson with the same mixing distribution. Using $CRM(N, X) = Z = X_1 + ... + X_N$ as notation for the CRM losses and abbreviating the *coefficicent of variation* (c.v.) as $\nu(Y) = \sigma(Y)/\mathrm{E}[Y]$,

Claim 4.1 *If $Z = CRM(MP(\lambda, G), X)$, then*

$$Z_S = CRM(MP((1-q)\lambda, G), X_S),\ \textit{and}$$

$$Z_L = CRM(MP(q\lambda, G), X_L),$$

where q is the probability of a large loss. Furthermore,

$$\rho(Z_S, Z_L) = c/[\nu(Z_S)\nu(Z_L)]. \tag{12}$$

Proof See Mildenhall [12]. Equation (12) follows from equation (4). □

Recall that for $Z = CRM(N, X)$,

$$\begin{aligned}
\mathrm{E}[Z] &= \lambda\mu(X) && (13)\\
\sigma^2(Z) &= \lambda\sigma^2(X) + \mu^2(X)^2\sigma^2(N) && (14)\\
\gamma(Z) &= \big[\mu^3(X)\gamma(N)\sigma^3(N) + 3\mu(X)\sigma^2(X)\sigma^2(N) \\
&\quad +\lambda\gamma(X)\sigma^3(X)\big]/\sigma^3(Z) && (15)
\end{aligned}$$

Here and later it is convenient, in particular, to have $\lambda = \mathrm{E}[N]$ and, in general, to have $\mu(Y)$ denote $\mathrm{E}[Y]$ and $\mu'_j(Y)$ denote $\mathrm{E}[Y^j]$ for a r.v. Y.

We may now use equations (10) and (11) and (13)–(15) to derive expressions for the c.v. and skewness of $Z = CRM(MP(\lambda, G), X)$:

$$\nu(Z) = \sqrt{c + \frac{1+\nu^2(X)}{\lambda}} \tag{16}$$

$$\gamma(Z) = \frac{\mu'_3(X)/(\mu^3(X)\sqrt{\lambda}) + 3c\sqrt{\lambda}(1+\nu^2(X)) + (c\lambda)^{3/2}\gamma(G)}{(1+\nu^2(X)+c\lambda)^{3/2}}\ . \tag{17}$$

It follows that as long as G and X do not depend on λ, $\nu(Z) \to \nu(G) = \sqrt{c}$, and $\gamma(Z) \to \gamma(G)$ as $\lambda \to \infty$. We may thus infer that the choice of G wields critical influence on the properties of a mixed Poisson CRM. This intuition is confirmed by the convergence theorem and examples in section 4.4 (as well as by Proposition 1 of [12]).

4.2 Negative Binonial

The most common example of a mixed Poisson is the negaive binomial, arising from $G \sim$ gamma. The gamma mixing distribution has parameters $\alpha = 1/c$ and $\beta = c$. We specify the negative binomial in terms of the mean and variance-to-mean ratio, and write $N \sim NB[\lambda, d]$. Its pdf is given by

$$\Pr(N = n) = \frac{\Gamma(n + \lambda/(d-1))}{n!\Gamma(\lambda/(d-1))} d^{-\lambda/(d-1)} \left(\frac{d-1}{d}\right)^n .$$

In the mixed Poisson formulation ($d = 1 + c\lambda$) the Negative Binomial pdf becomes

$$\Pr(N = n) = \frac{\Gamma(n + 1/c)}{n!\Gamma(1/c)} (1 + c\lambda)^{-1/c} \left(\frac{c\lambda}{1 + c\lambda}\right)^n .$$

This is the parameterization given in [10]. In [12], Mildenhall notes two types of negative binomial models, distinguished by their behavior as λ varies. In the *over-dispersed Poisson* (ODP) model, the variance-to-mean ratio is independent of λ. This forces the c parameter to depend on λ as $c = c_\lambda = (d-1)/\lambda$. In this case the c.v. $\nu(N) = \sqrt{c_\lambda + 1/\lambda} \to 0$ as $\lambda \to \infty$ (and $G = G_\lambda \xrightarrow{D} 1$). The *contagion* model, on the other hand, holds c fixed so that $d = d_\lambda \to \infty$ and $\nu(N) \to \sqrt{c}$ as $\lambda \to \infty$.

4.3 Other Mixing Distributions

Tables 11–13 in Appendix B show various choices for the mixing distribution G. A twist is that Tables 11–12 add shift and slope parameters s and m. So, the general form for G is $G = s + mH$, where H is the named distribution. Refer to the appendices of [6] for the standard parameterizations of the H-distributions. The parameters of H are then expressed in terms of the contagion c, and the (optional) parameters s and m. The parameters m and s are constrained by $0 <= s < 1$ and $m > 0$. They may be redundant or determined by the conditions $\mu(G) = 1$ and $\sigma^2(G) = c$.

Table 13 shows various ways to construct G from components G_i. In this case, c is expressed in terms of the contagions c_i of the components.

The second columns of Tables 11–13 show the skewness of G. Note the relationship $\mu_3'(G) = 1 + 3c + c^{3/2}\gamma(G)$ so that the symmetric distributions have third moment equal to $1 + 3c$. The skewness $\gamma(G)$ for a component distribution is expressed in terms of the $\gamma_i = \gamma(G_i)$

See the notes after Table 13 for a more detailed discussion.

Returning to our main context, the practitioner may have trustworthy estimates for the mean and c.v. of Z_S. This will rarely, if ever, be the case for the skewness $\gamma(Z_S)$. By equation (17), and Claim 4.1, the choice of G affords the opportunity to "take a view" of $\gamma(Z_S)$ in the limit $\lambda \to \infty$. For example, if one believes that the skewness will diversify away, then the continuous or discrete uniform might be the proper choice for G. Otherwise, consideration could be given to the ratio $\kappa(G) = \gamma(G)/\nu(G) = \gamma(G)/\sqrt{c}$ (the "skew-nu" ratio). For the unshifted Poisson, gamma, and inverse Gaussian, κ is constant ($\kappa = 1, 2, 3$, respectively). For the lognormal, $\kappa = 3 + c$. Choosing the shifted exponential or Pareto will result in much higher skewness for ordinarily encountered values of c. Adding the shift parameter allows for higher skewness with the more traditional choices. For example, the shifted gamma allows any skew-nu ratio ≥ 2. Another reason to add a shift is to reflect an assumption on the effective mimimum value of Z_S. That is, adding a shift to G will tend to increase the effective minimum of N_S and, therefore, of Z_S (Compare the simulated minimum values in Appendix C, Exhibit 5 to those in Exhibit 2).

4.4 Convergence Theorem

For the convergence theorem, we need the notions of characteristic function and weak convergence of distributions:

Definition:

1. The *characteristic function* of the r.v. Y is the complex-valued $\phi_Y(t) = \mathrm{E}[e^{itY}]$, $t >$

$0, i = \sqrt{-1}$.

2. A sequence of distribution functions is said to *converge weakly* to a limit F (written $F_n \xrightarrow{D} F$) if $F_n(y) \to F(y)$ for all y that are continuity points of F. A sequence of random variables Y_n is said to converge weakly or *converge in distribution* to a limit Y ($Y_n \xrightarrow{D} Y$) if their distribution functions $F_{Y_n}(y)$ converge weakly.

Theorem 4.2 *Suppose we are given $N_\lambda = MP(\lambda, G)$, and r.v.'s Y_n such that $\mu(Y_n) = nm$, $\sigma^2(Y_n) <= n^j s^2$ for some j, $0 <= j < 2$, and fixed s. Define Y_{N_λ} by $Y_{N_\lambda}|(N_\lambda = n) = Y_n$. Then*

$$Y_{N_\lambda}/(\lambda m) \xrightarrow{D} G \text{ as } \lambda \to \infty.$$

Proof Without loss of generality we may assume $m = 1$, so that $\mu(Y_n) = n$. Set

$$\bar{Y}_\lambda = Y_{N_\lambda}/\lambda.$$

Applying the Continuity theorem (see Durrett, Theorem 3.4 [1], for example), which states that convergence of characteristic functions implies convergence in distribution, we need to show

$$L := \lim_{\lambda\to\infty} \phi_{\bar{Y}_\lambda}(t) = \phi_G(t).$$

Note that $\phi_{\bar{Y}_\lambda(t)} = \phi_{Y_\lambda}(\bar{t})$, where $\bar{t} = t/\lambda$. Define N_λ^G and L_λ^G by

$$N_\lambda^G = N_\lambda | G \; (\sim \text{Poisson}(\lambda G)),$$

$$L_\lambda^G = \underset{N_\lambda^G}{\mathrm{E}}[\phi_{Y_n}(\bar{t})|G, N_\lambda^G = n].$$

Then $L = \lim_{\lambda\to\infty} \underset{G}{\mathrm{E}}[L_\lambda^G]$, and $|L_\lambda^G| \leq 1$ so by the Bounded Convergence Theorem it suffices to show that

$$\lim_{\lambda\to\infty} L_\lambda^G = e^{iGt}.$$

Now, if $Z_n = Y_n - n$ then $\mu(Z_n) = 0$ and $\mu_2'(Z_n) = \sigma^2(Y_n) = n^j s^2$. So, by Durrett, Theorem 3.8 [1],

$$\begin{aligned}
\lim_{\lambda\to\infty} L_\lambda^G &= \lim_{\lambda\to\infty} \underset{N_\lambda^G}{\mathrm{E}}[e^{i\bar{t}n}\phi_{Z_n}(\bar{t})|G, N_\lambda^G = n] \\
&= \lim_{\lambda\to\infty} \underset{N_\lambda^G}{\mathrm{E}}[e^{i\bar{t}n}(1 + n^j O(\bar{t}^2))|G, N_\lambda^G = n] \\
&= \lim_{\lambda\to\infty} \underset{N_\lambda^G}{\mathrm{E}}[e^{i\bar{t}n}|G, N_\lambda^G = n] \\
&\quad + \lim_{\lambda\to\infty} \underset{N_\lambda^G}{\mathrm{E}}[e^{i\bar{t}n} n^j O(\bar{t}^2)|G, N_\lambda^G = n]. \qquad (18)
\end{aligned}$$

Note that $N_\lambda^G \sim \text{Poisson}(\lambda G)$ implies that $\text{E}[(N_\lambda^G)^r] = O((\lambda G)^r)$, for all $r \geq 0$. With a second application of Durrett, Theorem 3.8 [1] to $e^{i\bar{t}n}$, we can evaluate the second term in 18 as

$$\begin{aligned} L^* &= \lim_{\lambda\to\infty} \mathop{\text{E}}_{N_\lambda^G}[e^{i\bar{t}n} n^j O(\bar{t}^2) | G, N_\lambda^G = n] \\ &= \lim_{\lambda\to\infty} \mathop{\text{E}}_{N_\lambda^G}[(1 + i\bar{t}n + n^2 O(\bar{t}^2)) n^j O(\bar{t}^2) | G, N_\lambda^G = n] \\ &= \lim_{\lambda\to\infty} [O((\lambda G)^j) O(\bar{t}^2) + iO((\lambda G)^{1+j}) O(\bar{t}^3) + O((\lambda G)^{2+j}) O(\bar{t}^4)] \\ &= 0, \text{ as } 0 \leq j < 2. \end{aligned}$$

Finally, the Poisson characteristic function $\phi(t) = e^{\lambda(e^{it}-1)}$ and one more application of Durrett, Theorem 3.8 [1] show that

$$\begin{aligned} \lim_{\lambda\to\infty} L_\lambda^G &= \lim_{\lambda\to\infty} \mathop{\text{E}}_{N_\lambda^G}[e^{i\bar{t}n} | G, N_\lambda^G = n] \\ &= \lim_{\lambda\to\infty} e^{\lambda G(e^{i\bar{t}}-1)} \\ &= \lim_{\lambda\to\infty} e^{\lambda G(i\bar{t}+O(\bar{t}^2))} \\ &= e^{iGt}. \qquad \square \end{aligned}$$

4.4.1 Convergence of CAD and CRM

If we set $Y_n = \sum_{i=1}^n X_i,\ X_i$ iid, then $\sigma^2(Y_n) = n\sigma^2(X)$ and we have Proposition 1 of [12], i.e., for $Z = CRM(MP(\lambda, G), X)$,

$$Z/\mu(Z) \to G,$$

no matter the choice of X ("severity is irrelevent" [1]). In our context, setting $Y_n = \widehat{Z_S} | N_S = n$ shows that for $k >= 2$ and $\widehat{Z_S} = CAD_k(MP(\lambda(N_S), G), X_S, \mathcal{F})$,

$$\widehat{Z_S}/\mu(\widehat{Z_S}) \to G$$

no matter the choice of X or $\mathcal{F}$ (severity and conditional aggregate distribution are irrelevant). Putting the two cases together supports $\widehat{Z_S}$ as a good approximation for Z_S as each of these r.v.'s converge to G when normalized by the mean. The theorem equally applies to the CAD total losses $\widehat{Z}$ by setting $Y_n = \widehat{Z_S} + Z_L | (N_S = n - B, N_L = B)$, where $B \sim Bin(n, q)$. Thus, the CAD small, large (by construction), and total losses converge to those of the CRM.

[1] Mildenhall [12] explains in the context of a CRM that, "in some cases the actual form of the severity distribution is essentially irrelevant to the shape of the aggregate distribution."

4.4.2 Convergence to G - Examples

Of course, the theorem also applies to Z_L, but this is irrelevant to most insurance situations, due to the relatively small expected claim count. In this case, severity may be quite relevant. On the other hand, Z_S will take on the characteristics of G for moderately sized insurance portfolios. The top chart of Appendix C, Exhibit 1 shows the pdf of Z_S for a portfolio similar to the one in the Basic Example of Section 3.3 - with $\mu(Z) = \$25,000,000$ and large loss threshold of \$200,000 (solid area). The mixing distribution G is the three-point Hermite (Appendix B, notes). Overlaid is the pdf of $\widehat{Z_S}$ where $\widehat{Z_S}|N_S \sim$ shifted exponential (as in Appendix A, Table 10). It's interesting that the highly skewed, monotonic exponential distribution diversifies away to the symmetric, tri-modal Hermite. In fact the Table 10 shifted exponential, as a CAD_2 model, satisfies the convergence theorem with $j = 1$. If we match only the mean (i.e., use a CAD_1 model) we may reparameterize the shifted exponential as

$$N_s\mu(X_S) - \sqrt{N_S^j}\sigma(X_S) + \text{Exp}[\sqrt{N_S^j}\sigma(X_S)],$$

and this also satisfies the convergence theorem as long as $j < 2$. The bottom chart of Appendix C, Exhibit 1 shows the case $j = 1.5$ converging to G, but more slowly. Of course, a (CAD_1) model with $j = 0$ would converge to G too quickly to be useful in approximating the actual CRM. For example, such a model would have $\nu^2(\widehat{Z_S}) = \nu^2(X_S)/\lambda^2 + c + 1/\lambda$, so that the severity component $\rightarrow 0$ as $1/\lambda^2$ rather than $1/\lambda$ as in equation (16).

Exhibits 2-5 in Appendix C expand on the Basic Example in Section 3 in light of the convergence theorem. The claim count distribution in this example was a negative binomial wtih mean $\lambda = 527$ and and variance-to-mean ratio $d = 33.94$. Equivalently, this is a mixed Poisson with gamma mixing distribution and contagion $c = 0.0625$. This is the subject of Appendix C, Exhibit 2. We ran the CAD algorithm using the @Risk software with 30,000 iterations. We also simulated the small losses directly from the assumed claim count and lognormal severity distributions as a basis for comparison.

The top chart of Exhibit 2.1 shows the simulated pdf of the "true" losses (solid region) versus six different choices for the CAD distributional family $\mathcal{F}$. These include both CAD_2 and CAD_3 models. Visually the fits are excellent, even for exotic choices such as the shifted exponential and the (CAD_3) distribution on two points. The table at the bottom of Exhibit 2.1 is adapted from the standard @Risk "Detailed Statistcs" output. It shows moment and percentile statistics for each distribution. Convergence to the mixing distribution is evidenced by considering the ratio of skewness to the c.v. (the skew-nu ratio). For a gamma distribution, this ratio is equal to 2.

Exhibit 2.2 shows scatterplots of simulated large versus small losses. The top chart shows the true small losses (Z_L vs. Z_S), while the bottom chart generates small losses via tha CAD algorithm (Z_L vs. $\widehat{Z_S}$). The close similarity of the two plots indicates that CAD does a good job of reflecting the overall dependence of large and small losses, as well as matching the numerical correlation per Claim 3.1.

Exhibits 3-5 repeat Exhibit 2 for different choices of the mixing distribution. A lognormal mixing distribution is used in Exhibit 3 with similar results. Here, convergence to G is evidenced by a skew-nu ratio in the 3-ish range. Exhibits 4 and 5 reflect more unusual choices for the mixing distribution - a uniform and a three-point shifted binomial, respectively. The shifted binomial is parameterized to match the skewness of the gamma mixing distribution, i.e., $\gamma(G) = 0.5$. In these cases, due to the distinctive shapes of the pdf graphs, visual inspection serves as evidence of convergence to G. Once again, the large vs. small loss scatterplots match up extremely well. The scatterplot for the shifted binomial has three distinct regions, corresponding to the three possible values of G. Each region appears very nearly symmetric, reflecting the fact that $\rho(Z_S|G, Z_L|G) = 0$ by equation (4).

In [12] Mildenhall uses the Iman Conover (IC) method to model the dependence of large and small losses. This is a rank-order correlation method that has the advantage of being easy to use in spreadsheets and simulations. To apply IC, Mildenhall uses simulated output from method of moments fitted curves for both small and large losses. The curve used is a shifted gamma, i.e., a fit to the first three moments of the unconditional losses. In Appendix C, Exhibit 6-7, the IC method is applied with the shifted gamma fitted curve for small losses, but the actual CRM simulated output for large losses. For the gamma mixing distribution, IC appears to do a good job matching the pdf graphs and scatterplots from Exhibit 2. Note, however, that as long as the first three moments are kept constant, the small loss curve fit will not vary with a change in the mixing distribution. The result is a poor fit to the small loss pdf for the shifted binomial mixing distribution (Exhibit 7.1). The IC method also will not reproduce the three distinct regions of the large vs. small loss scatterplots in Exhibit 5.2. If we "cheat" by applying IC to CRM simulated output for *both* large and small losses, the resulting scatterplot will show three distinct regions (Exhibit 7.2). However, the rank-order construction will not replicate $\rho(Z_S|G, Z_L|G) = 0$, as can be seen by noting the positive slope within each region. That is, the CAD method reflects the conditional small/large indepence correctly, but the IC method does not.

5 CAD with Limited Information - A Reinsurance Example

The example considered in this section is typical of a reinsurance pricing exercise requiring simultaneous modeling of large and small losses. It is a reinsurance coverage with two sections - (1) a stop-loss on the cedant's "net" losses and (2) excess-of-loss (XoL) coverage. Here, net losses are losses limited to the large loss threshold τ. Excess losses include all amounts exceeding τ and limited to the policy limit. Aggregate net and excess loss are thus given by:

$$Z_{Net} = Z_S + N_L \tau$$

and

$$Z_{XoL} = Z_L - N_L \tau.$$

The stop-loss covers net losses excess of an annual aggregate deductible (AAD) and limited to the annual aggregate limit (AAL), that is

$$Z_{SL} = \min(AAL, \max(0, Z_{Net} - AAD)).$$

Finally, the reinsurance coverage will reimburse the total of the two coverage sections:

$$Z_{Re} = Z_{SL} + Z_{XoL}.$$

To evaluate and price such a reinsurance contract, it is clearly important to accurately reflect the dependence of large and small losses. For example, the large-small dependence may significantly impact downside risk measures such as Tail Value-at-Risk (TVaR). The CAD methodology is thus an excellent candidate for the loss modeling. We will continue to assume the underlying losses follow a mixed Poisson CRM, with contagion parameter c=0.0625. Various choices for the mixing distribution G will be considered.

To this point, the CAD method as presented requires the full (ground-up) severity and claim count distributions. In reinsurance applications, however, the available data may be insufficient to reasonably parameterize these distributions. We will demonstrate how to apply CAD with more limited input information.

For this example, the input data is limited to the mean and c.v. of total aggregate losses $(\mu(Z),\ \nu(Z))$, the mean $\lambda(N_L)$ of the large loss claim count, and the large loss severity distribution F_{X_L}. This information set-up is fairly typical in reinsurance pricing. The parameters $\mu(Z)$, $\nu(Z)$ may have been estimated using aggregated data such as loss development triangles and historical loss ratios. The distribution F_{X_L} may have been derived by fitting a curve to the supplied large loss listing, with $\lambda(N_L)$ based on historical excess claim counts.

Alternatively, F_{X_L} may be an empirical distribution developed to replicate selected loss costs for several XoL layers. In this example, we do assume an empirical distribution for F_{X_L}, with the large loss threshold $\tau = \$200,000$. The large loss distribution and other parameter values are shown in Table 7.

Table 7: Initial Parameters for Reinsurance Example

Parameter		Value
τ		\$200,000
AAD		\$25,000,000
AAL		\$20,000,000
$\mu(Z)$		\$25,000,000
$\nu(Z)$		0.28
$\lambda(N_L)$		21.5
Contagion c		0.0625
Large Loss Severity($F(X_L)$)		
Claim Size	Incremental Probability	Cumulative Probability
200,000	19.6%	19.6%
300,000	25.2%	44.8%
400,000	14.1%	58.9%
500,000	8.9%	67.8%
600,000	6.1%	73.9%
700,000	4.4%	78.3%
800,000	3.3%	81.6%
900,000	2.6%	84.2%
1,000,000	15.8%	100.0%
Implied Large Loss Statistics		
$\mu(X_L)$		\$490,900
$\nu(X_L)$		0.5691
$\mu(Z_L)$		\$10,554,350
$\nu(Z_L)$		0.3522

The large loss values in the bottom portion of Table 7 are easily computed from $F(X_L)$, $\lambda(N_L)$, and c. The c.v. $\nu(Z_L)$ is derived with equation (16) for Z_L and X_L, noting that Z_L is also a mixed Poisson CRM with contagion c.

The CAD algorithm also requires a value for the mean total claim count λ. It may be that sufficient historical data is available for a reliable estimate of λ. If this is not the case, we *posit* a value for λ. For this example, we set $\lambda = 500$.

Given a choice for the mixing distribution G, CAD steps 1-4 may now be executed. This will generate simulated values for Z_L, N_L, and N_S. To simulate values for Z_S in step 5, we need to derive expressions for the mean and c.v. of $Z_S|N_S$. Note that $\mu(Z_S) = \mu(Z) - \mu(Z_L)$, $\lambda(N_S) = \lambda(N) - \lambda(N_L)$, and

$$\mu(Z_S|N_S) = N_S\mu(X_S) = N_S\mu(Z_S)/\lambda(N_S). \tag{19}$$

By equations (8) and (9), $\nu(Z_S|N_S) = \nu(X_S)/\sqrt{N_S}$. Equation (16) applied to Z_S can be used with equations (12) and (19) and the fact that $\sigma^2(Z_S) = \sigma^2(Z) - \sigma^2(Z_L) - 2\rho(Z_S, Z_L)\sigma(Z)\sigma(Z_L)$ to eliminate $\nu(X_S)$ from the expression for $\nu(Z_S|N_S)$. After some algebra, the formula for $\nu(Z_S|N_S)$ becomes:

$$\nu(Z_S|N_S) = \sqrt{\frac{\lambda(N_S)\left[\mu^2(Z)(\nu^2(Z) - c) - \mu^2(Z_L)(\nu^2(Z_L) - c)\right] - \mu^2(Z_S)}{N_S\mu^2(Z_S)}}. \tag{20}$$

Equations (19) and (20) now allow for the method of moments fit in step 5 without referring to the small loss severity r.v. X_S. This limited information version of the algorithm is strictly a CAD_2 excercise. To derive an expression for $\gamma(Z_S|N_S)$, say, would involve an *a priori* estimate of the skewness $\gamma(Z)$ - rarely, if ever, available. Table 8 substitutes the known values from Table 7 into equations (19) and (20).

Table 8: Small Loss Model

$\mu(Z_S)$	\$14,455,650
$\lambda(N_S)$	478.50
$\mu(Z_S|N_s)$	$= 30,189.45N_S$
$\nu(Z_S|N_S)$	$= 2.46/\sqrt{N_S}$

We may now run the CAD algorithm to determine an appropriate premium for the coverage of Z_{Re}. The premium P is set as $P = \mu(Z_{Re}) + u\Phi$, where Φ is a downside risk measure and u is the load factor. For this exercise, $\Phi = TVaR(Z_{Re}, 0.99)$, the Tail Value-at-Risk of Z_{Re} at the 99th percentile. The load factor is set equal to 10%. Table 9 shows the results of running the CAD algorithm with 30,000 iterations and various choices of the mixing distribution G. There is some variation in $\mu(Z_{SL})$ and significant variation in the TVaR values as G varies. This results in a smaller, but still significant variation in indicated premium.

Care should be taken in applying the limited information CAD method. The choice of the parameters λ, and c, along with the input information will impute values for some of the other loss statistics. The preceding example imputes values for the small loss statistics $\mu(X_S)$, $\sigma^2(X_S)$, $\mu(Z_S)$, $\sigma^2(Z_S)$, and also for $\sigma^2(Z_L)$ (through the choice of c). However, there is no a priori guarantee that, say, $\sigma^2(X_S) > 0$. There may also be a more subtle inconsistency, such as $\mu(X_S) > \mu(X_L)$. The practitioner should include these types of consistency checks when applying the limited information CAD.

It is possible that input information such as found in Table 7 is internally consistent but inconsistent with the mixed Poisson CRM. Informally, we say that the input information *admits a mixed Poisson CRM* if there is a choice of λ and c resulting in no inconsistencies.

Table 9: Simulation Results from Different Mixing Distributions

Mixing Distribution	Uniform	Gamma	Log-normal	Shifted Gamma	S. Log-normal	Expo-nential	Pareto	Beta	S. Bi-nomial
$\mu(Z_L)$	10.5	10.6	10.6	10.6	10.5	10.6	10.5	10.6	10.5
$\mu(Z_S)$	14.4	14.4	14.4	14.4	14.4	14.4	14.4	14.4	14.4
$\mu(Z_{Net})$	18.7	18.7	18.7	18.7	18.7	18.7	18.7	18.7	18.7
$\mu(Z_{XoL})$	6.2	6.3	6.3	6.3	6.2	6.3	6.2	6.3	6.2
$\mu(Z_{SL})$	1.6	1.5	1.5	1.4	1.2	1.4	1.1	1.5	1.7
$\mu(Z_{Re})$	7.8	7.7	7.7	7.7	7.5	7.7	7.4	7.7	7.9
TVaR(ZRe, 99)	22.3	27.3	28.5	33.0	34.6	33.1	34.3	31.8	26.5
Premium	10.1	10.5	10.6	11.0	10.9	11.0	10.8	10.9	10.6

6 CAD with Multiple Lines of Business

This section adapts the CAD method to model multiple lines of business and impose correlation between lines. In this context, let Z_i, $i = 1 \ldots m$, be the aggregatee loss r.v. for the ith line, and τ_i the large loss threshhold. All other notations ($Z_{i,S}$, $Z_{i,L}$, etc.) carry through. As in the previous section we allow for limited information, but say that Z_i admits a mixed Poisson CRM with parameters λ_i and c_i. Note that by equation (16), $c_i < \min(\nu^2(Z_i) - 1/\lambda(N_i), \nu^2(Z_{i,S}) - 1/\lambda(N_{i,S}), \nu^2(Z_{i,L}) - 1/\lambda(N_{i,L}))$.

6.1 Common Shock CAD

Of course, one can extend the CAD method to m lines of business simply by iterating m times. For the multi-line mixed Poisson CRM, it's natural to impose correlation via a common shock component on the mixing distributions G_i [11]. As noted in Appendix B, the *twisted product* construction is well-suited to this purpose.

With notation as above set $c_{\min} = \min\{c_i, i = 1 \ldots m\}$, and take w such that $0 \le w \le 1$. The parameter w is the weight given to the common shock component. We now assume that the mixing distribution G_i has the form

$$G_i[c_i] = G_1 \bullet G_{2,i} = G_1[wc_{\min}]G_{2,i}[(c_i - wc_{\min})/G_1].$$

Here, G_1 is the common (or industry) component and $G_{2,i}$ is the line-specific component, with contagion parameter "distorted" by G_1. By the discussion in Appendix B, $\sigma^2[G_i] = wc_{\min} + c_i - wc_{\min} = c_i$, as required.

Programatically, step 1 of the CAD algorithm becomes

Step 1^{CS}: Draw G_1 from $G_1[wc_{\min}]$. Then, for each i, draw N_i from $MP(\lambda_i G_1, G_{2,i}[(c_i - wc_{\min})/G_1])$.

Steps 2-5 then proceed unchanged for each line. By analogy with equation (12), the common shock CAD results in the following correlations for $i \ne j$:

$$\rho(\widehat{Z_{i,S}}, \widehat{Z_{j,S}}) = wc_{\min}/(\nu(Z_{i,S})\nu(Z_{j,S}))$$
$$\rho(\widehat{Z_{i,S}}, Z_{j,L}) = wc_{\min}/(\nu(Z_{i,S})\nu(Z_{j,L}))$$
$$\rho(Z_{i,L}, Z_{j,L}) = wc_{\min}/(\nu(Z_{i,L})\nu(Z_{j,L})).$$

6.2 Common Shock CAD with Conditional Correlation

In [11], Meyers employs a common shock model acting on the severity distributions, in addition to a claim count model similar to that described above. The CAD method suppresses reference to the small loss severity, especially in the case of limited information. To generate a second source of between-line correlation, we specify a fixed correlation matrix to be applied to the $Z_{i,S}|N_{i,S}$ in step 5 of the CAD algorithm. Step 5 is then replaced by

Step 5^{Corr}: Draw aggregate small losses for each line from a joint distribution $[\widehat{Z_{1,S}}|N_{1,s} \ldots \widehat{Z_{m,S}}|N_{m,s}]$ with correlation matrix $\Gamma = [r_{ij}]$ and such that the marginals $\widehat{Z_{i,S}}|N_{i,s}$ are parameterized by matching the first k moments of $Z_{i,S}|N_{i,S}$.

For $i \neq j$, Step 5^{Corr} implies that

$$\text{Cov}(\widehat{Z_{i,S}}|N_{i,S}, \widehat{Z_{j,S}}|N_{j,S}) = r_{ij}\sqrt{N_{i,S}N_{j,S}}\sigma(X_{i,S})\sigma(X_{j,S}).$$

It follows that for common shock CAD with conditional correlation:

$$\underset{N_{i,S},N_{j,S}}{\text{E}}[\text{Cov}(\widehat{Z_{i,S}}|N_{i,S}, \widehat{Z_{j,S}}|N_{j,S})] \approx h_{ij}r_{ij}\sqrt{\lambda(N_{i,S})\lambda(N_{j,S})}\sigma(X_{i,S})\sigma(X_{j,S}),$$

where $h_{ij} = \text{E}[\sqrt{G_iG_j}] = \underset{G_1}{\text{E}}[\sqrt{G_{2,i}G_{2,j}}|G_1]$, using $\text{E}[\sqrt{N}] \approx \sqrt{\lambda}$ for N Poisson. Furthermore,

$$\begin{aligned}\text{Cov}[\text{E}[\widehat{Z_{i,S}}|N_{i,S}], \text{E}[\widehat{Z_{j,S}}|N_{j,S}]] &= \mu(X_{i,S})\mu(X_{j,S})\text{Cov}[N_{i,s}, N_{j,s}] \\ &= \mu(X_{i,S})\mu(X_{j,S})wc_{\min}\lambda(N_{i,S})\lambda(N_{j,S}) \\ &= wc_{\min}\mu(Z_{i,S})\mu(Z_{j,S}).\end{aligned}$$

Using equation (16) to eliminate the small loss severity we find;

$$\begin{aligned}\rho(\widehat{Z_{i,S}}, \widehat{Z_{j,S}}) &= \frac{\underset{N_{i,S},N_{j,S}}{\text{E}}[\text{Cov}(\widehat{Z_{i,S}}|N_{i,S}, \widehat{Z_{j,S}}|N_{j,S})] + \text{Cov}[\text{E}[\widehat{Z_{i,S}}|N_{i,S}], \text{E}[\widehat{Z_{j,S}}|N_{j,S}]]}{\sigma(Z_{i,s})\sigma(Z_{i,s})} \\ &\approx \frac{wc_{\min} + h_{ij}r_{ij}\prod_{\iota=i,j}\sqrt{\nu^2(Z_{\iota,S}) - c_\iota - 1/\lambda(N_{\iota,s})}}{\nu(Z_{i,S})\nu(Z_{j,S})}. \qquad (21)\end{aligned}$$

Note that $h_{ij} = 1$ if $w = 1$ and $c_i = c_j = c_{\min}$. In particular, if Z_i and Z_j are identical, then write $c_i = c_j = t(\nu^2(Z_{i,S}) - 1/\lambda(N_{i,S}))$, and (21) reduces to

$$\begin{aligned} \rho(\widehat{Z_{i,S}}, \widehat{Z_{j,S}}) &\approx (t + r_{ij}(1-t))[1 - 1/(\nu^2(Z_{i,S})\lambda(N_{i,S}))] \\ &\approx (t + r_{ij}(1-t)), \end{aligned}$$

if $\lambda(N_{i,S}) >> 1/\nu^2(Z_{i,S})$.

7 Conclusion

The CAD method provides a way to efficiently simulate the CRM while preserving the inherent dependencies between large and small losses. These dependencies are fundamentally driven by the claim counts and the theorem presented herein shows how the mixed Poisson CRM and CAD method model will converge as the expected claim count grows. This provides theoretical support for the practical oservation that the CAD method does a good job approximating the CRM.

8 Acknowledgements

The authors would like to thank the reviewers for many helpful comments and suggestions.

9 Biography of Authors

Richard Rosengarten is a Vice President in Reinsurance with Towers Watson. He is a Fellow of the Casualty Actuarial Society, a Member of the American Academy of Actuaries, and holds a Ph.D in mathematics. He can be reached via email at richard.rosengarten@towerswatson.com.

David Homer is a Vice President in Reinsurance with Towers Watson. He is a Fellow of the Casualty Actuarial Society and a Member of the American Academy of Actuaries. He can be reached via email at david.homer@towerswatson.com.

References

[1] Durrett, R., *Probability: Theory and Examples*, Duxbury Press, 2004.

[2] Heckman, Philip E. and Glenn G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *PCAS* LXX, 1983, pp. 22–61.

[3] Homer, David L. and David R. Clark, "Insurance Applications of Bivariate Distributions," *PCAS* XC, 2003, pp. 274–307.

[4] Homer, David L., "Part 2—Aggregating Bivariate Claim Severities With Numerical Fourier Inversion," *CAS Forum*, Winter 2006, pp. 205–230.

[5] Iman, R., and W. Conover. "A distribution-free approach to inducing rank correlation among input variables," *Commun. Statist. Simula. Computa.*, 1982, 11(3), pp. 311-334.

[6] Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot, *Loss Models: From Data to Decisions*, New York, John Wiley & Sons, Inc., 1998.

[7] Heckman, Philip E., and Glenn G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *PCAS* LXX, 1983, pp. 22–61.

[8] Meyers, Glenn G., "The Competitive Market Equilibrium Risk Load Formula for Increased Limits Ratemaking," *PCAS* LXXVIII, 1991, pp. 163–200.

[9] Meyers, Glenn G., "Discussion of Shaun Wang's Aggregation of Correlated Risk Portfolios: Models and Algorithms," *PCAS* LXXXV, 1999, pp. 781–805.

[10] Meyers, Glenn G., "Estimating Between Line Correlations Generated by Parameter Uncertainty," *CAS Forum*, Summer 1999, pp. 197–222.

[11] Meyers, Glenn G., "The Common Shock Model for Correlated Insurance Losses," *Variance* 1(1), 2007, pp. 197–222

[12] Mildenhall, Stephen J., "Part 1—Correlation and Aggregate Loss Distributions With An Emphasis On The Iman-Conover Method," *CAS Forum* Winter 2006, pp. 103–204.

[13] Panjer, Harry H., "Recursive Evaluation of a Family of Compound Distributions," *ASTIN Bulletin*, Vol. 12, No. 1, 1981, pp. 22–26.

[14] R Development Core Team, "R: A Language and Environment for Statistical Computing," Vienna, Austria, R Foundation for Statistical Computing, 2008, ISBN 3-900051-07-0, URL http://www.R-project.org.

[15] Sundt, Bjørn, "On Multivariate Panjer Recursions," *ASTIN Bulletin*, Vol. 29, No. 1, 1999, pp. 29–45.

[16] Venter, Gary G.,"Tails of Copulas," *PCAS* LXXXIX, 2002, pp. 68–113.

[17] Walhin, Jean-François, "On the Optimality of Multiline Excess of Loss Covers," *CAS Forum*, Spring 2003, pp. 231-243.

[18] Walhin, Jean-François, and Jose Paris, "Using Mixed Poisson Processes in Connection with Bonus-Malus Systems," *ASTIN Bulletin*, Vol. 29, No. 1, 1999, pp. 81–100.

[19] Walhin, Jean-François, and Jose Paris, "The Mixed Bivariate Hoffmann Distribution," *ASTIN Bulletin*, Vol. 31, 2001, pp. 123–131.

[20] Tremblay, L., "Using the Poisson Inverse Gaussian Distribution in Bonus-Malus Systems," *ASTIN Bulletin*, Vol. 22, 1992, pp. 97–106.

[21] Simar, L, Jose, "Maxmlum Likelihood Estimation of a Compound Polsson Process," *The Annals of Statistics*, Vol. 4, 1976, pp. 1200–1209.

[22] Wang, Shaun S., "Aggregation of Correlated Risk Portfolios: Models and Algorithms," *PCAS* LXXXV, 1998, pp. 848–939.

A Conditional Aggregate Distributions

Table 10: CAD_2 and CAD_3 Fits to $Z_S|N_S$

Distribution	Statistics	Fit
Normal $(\widehat{\mu}, \widehat{\sigma})$	$\mu = \widehat{\mu}$ $\sigma^2 = \widehat{\sigma}^2$	$\widehat{\mu} = N_S \mu(X_S)$ $\widehat{\sigma} = \sqrt{N_S}\sigma(X_S)$
Uniform on $(\widehat{\mu} - \widehat{r}, \widehat{\mu} + \widehat{r})$	$\mu = \widehat{\mu}$ $\sigma^2 = \widehat{r}^2/3$	$\widehat{\mu} = N_S \mu(X_S)$ $\widehat{r} = \sqrt{3N_S}\sigma(X_S)$
Lognormal $(\widehat{\mu}, \widehat{\sigma})$	$\mu = e^{\widehat{\mu} + \widehat{\sigma}^2/2}$ $\sigma^2 = \mu^2(e^{\widehat{\sigma}^2} - 1)$	$\widehat{\mu} = \ln[N_S \mu(X_S)] - \widehat{\sigma}^2/2$ $\widehat{\sigma} = \sqrt{\ln[1 + \sigma^2(X_S)/(N_S \mu^2(X_S))]}$
Gamma $(\widehat{\alpha}, \widehat{\beta})$	$\mu = \widehat{\alpha}\widehat{\beta}$ $\sigma^2 = \widehat{\alpha}\widehat{\beta}^2$	$\widehat{\alpha} = N_S/\nu^2(X_S)$ $\widehat{\beta} = \mu(X_S)\nu^2(X_S)$
Shifted Exponential $(\widehat{s}, \widehat{\theta})$	$\mu = \widehat{\theta} + \widehat{s}$ $\sigma^2 = \widehat{\theta}^2$	$\widehat{\theta} = \sqrt{N_S}\sigma(X_S)$ $\widehat{s} = N_S \mu(X_S) - \widehat{\theta}$
2-Point (CAD_3) $(\mathrm{P}(\widehat{\mu} - \widehat{a}) = p$ $\mathrm{P}(\widehat{\mu} + \widehat{b}) = 1 - p\ )$	$\mu = \widehat{\mu}$ $\sigma^2 = p\widehat{a}^2 + (1-p)\widehat{b}^2$ $\gamma = \dfrac{(1-p)\widehat{b}^3 - p\widehat{a}^3}{\sigma^{3/2}}$	$\widehat{\mu} = N_S \mu(X_S)$ $\widehat{s} = \sqrt{N_S}\sigma(X_S)$ $p = \left(1 + \gamma(X_S)\sqrt{\dfrac{1}{4N_S + \gamma^2(X_S)}}\right) \Big/ 2$ $\widehat{a} = \widehat{s}\sqrt{(1-p)/p}$ $\widehat{b} = \widehat{a}p/(1-p)$
Shifted Lognormal $(\widehat{\mu}, \widehat{\sigma}, \widehat{s})$	$\mu = \widehat{s} + e^{\widehat{\mu} + \widehat{\sigma}^2/2}$ $\sigma^2 = (e^{\widehat{\sigma}^2} - 1)(\mu - \widehat{s})^2$ $\gamma = \eta(\eta^2 + 3)$, where $\eta = \sqrt{e^{\widehat{\sigma}^2} - 1}$	$\widehat{\mu} = \ln(N_S \mu(X_S) - \widehat{s}) - \widehat{\sigma}^2/2$ $\widehat{\sigma} = \sqrt{\ln\left[1 + \dfrac{N_S \sigma^2(X_S)}{(N_S \mu(X_S) - \widehat{s})^2}\right]}$ $\widehat{s} = N_S \mu(X_S) - \sqrt{N_S}\sigma(N_S)/(\zeta - 1/\zeta)$, where $\zeta = [\sqrt{4 + \gamma^2(X_S)/N_S} + \gamma(X_S)/(2\sqrt{N_S})]^{1/3}$
Shifted Gamma $(\widehat{\alpha}, \widehat{\beta}, \widehat{s})$	$\mu = \widehat{s} + \widehat{\alpha}\widehat{\beta}$ $\sigma^2 = \widehat{\alpha}\widehat{\beta}^2$ $\gamma = 2/\sqrt{\widehat{\alpha}}$	$\widehat{\alpha} = 4N_S/\gamma^2(X_S)$ $\widehat{\beta} = \gamma(X_S)\sigma(X_S)/2$ $\widehat{s} = N_S \mu(X_S) - \widehat{\alpha}\widehat{\beta}$
Generalized Beta $(\widehat{\alpha}, \widehat{\beta}, \widehat{m}(=\max))$ $(\min=0)$	$\mu = \widehat{\alpha}\widehat{m}/(\widehat{\alpha} + \widehat{\beta})$ $\sigma^2 = \mu^3\widehat{\beta}/[\widehat{\alpha}(\mu + \widehat{\alpha}\widehat{\beta})]$ $\gamma = 2\mu\sigma(\widehat{\alpha} - \widehat{\beta})/\eta,$ $\eta = \sigma^2\widehat{\alpha} + \mu^2\widehat{\beta}$	$\widehat{\alpha} = (1 - 1/\zeta)N_S/\nu^2(X_S) - 1/\zeta$ $\widehat{\beta} = \widehat{\alpha}(\zeta - 1)$ $\widehat{m} = \zeta N_S \mu(X_S)$, where $\zeta = 1 + \nu(X_S)\dfrac{\gamma(X_S)\nu(X_S) + 2N_S}{2\nu(X_S) - \gamma(X_S)}$

B Poisson Mixing Distributions

B.1 Tables of distributions

Table 11: Continuous Mixing Distributions

Family and Equation	Skewness
Gamma: $G = s + \text{Gamma}\left[\frac{(1-s)^2}{c}, \frac{c}{(1-s)}\right]$	$\frac{2\sqrt{c}}{1-s}$
Lognormal (Logn): $G =$ $s + \text{Logn}\left[\ln\left(\frac{(1-s)^2}{\sqrt{(1-s)^2+c}}\right), \sqrt{\ln\left(1+\frac{c}{(1-s)^2}\right)}\right]$	$\frac{\sqrt{c}}{1-s}\left(3+\frac{c}{(1-s)^2}\right)$
Exponential (Exp): $G = 1-\sqrt{c}+\text{Exp}[\sqrt{c}],\ c<1$	2
Inverse Gaussian (IG): $G = s + \text{IG}\left[(1-s), \frac{(1-s)^3}{c}\right]$	$\frac{3\sqrt{c}}{1-s}$
Pareto (Par): $G = 1-\sqrt{\frac{c}{k}}+\text{Par}\left[\sqrt{\frac{c}{k}}\left(\frac{k+1}{k-1}\right), \frac{2k}{k-1}\right]$ where $max(1,c) < k < 3$	$\frac{2}{\sqrt{k}}\left(\frac{3k-1}{3-k}\right)$
Uniform (U): $G = \text{U}\left[1-\sqrt{3c}, 1+\sqrt{3c}\right],\ c<1/3$	0
Generalized Beta on (s,M+s) (GB): $G = \text{GB}\left[\alpha, (M-1+s)\alpha/(1-s), s, M+s\right]$, where $\alpha = (1-s)[(1-s)(M-1+s)/c-1]/M$	$\frac{2\sqrt{c}(M-2(1-s))}{((1-s)(M-1+s)+c)}$

Table 12: Discrete Mixing Distributions

Family and Equation	Skewness
Discrete Uniform on 2m+1 points: $G = D[\Delta, p, m]$, defined by $\mathrm{P}(1) = p,\ \mathrm{P}(1 \pm j\Delta) = \dfrac{1-p}{2m},\ j \leq m$ $\Delta = \sqrt{\dfrac{6c/(1-p)}{(m+1)(2m+1)}},\ 1 - m\Delta > 0$	0
Poisson (Psn): $G = s + \dfrac{c}{(1-s)}\mathrm{Psn}[(1-s)^2/c]$	$\dfrac{\sqrt{c}}{1-s}$
Negative Binomial (NB[λ,d]): $G = s + \dfrac{c}{d(1-s)}\mathrm{NB}[d(1-s)^2/c, d]$ M an integer ≥ 1	$\dfrac{(2-1/d)\sqrt{c}}{1-s}$
Binomial (Bin): $G = s + \dfrac{(1-s)^2 + cM}{M(1-s)}\mathrm{Bin}\left[M, \dfrac{(1-s)^2}{(1-s)^2 + cM}\right]$ M an integer ≥ 1	$\dfrac{\sqrt{c}}{1-s} - \dfrac{1-s}{M\sqrt{c}}$

Table 13: Component Mixing Distributions

Family and Equation	Skewness
Weighted Sum: $G[c] = pG_1[c_1] + (1-p)G_2[c_2]$ $c = p^2 c_1 + (1-p)^2 c_2$	$\dfrac{pc_1^{3/2}\gamma_1 + (1-p)c_2^{3/2}\gamma_2}{c^{3/2}}$
Straight Product: $G[c] = G_1[c_1]G_2[c_2]$, G_1, G_2 independent. $c = c_1 + c_2 + c_1 c_2$	$\dfrac{c_1 c_2[6 + 3(\sqrt{c_1}\gamma_1 + \sqrt{c_2}\gamma_2) + \sqrt{c_1 c_2}\gamma_1\gamma_2]}{c^{3/2}}$
Twisted Product: $G[c] = G_1[c_1]G_2[c_2/G_1]$ $c = c_1 + c_2$	$\dfrac{\mu_3'(G_1)f(G_1, G_2) - 1 - 3c}{c^{3/2}}$, where $f(G_1, G_2) = \mathop{\mathrm{E}}_{G_1}(\mu_3'(G_2[c_2/G_1]\|G_1))$

B.2 Additional Notes

1. **Products of Mixing Distributions**. In several papers ([8],[10], for example), Meyers presents count r.v.'s of the form $N = N^*[G_1[c_1]\lambda, d(G_1))]$, where G_1 is a mixing distribution, and N^* is a family depending on λ, and d (i.e., $N^* \sim NB[\lambda, d]$). We consider the case $N^* \sim MP(\lambda, G_2[c_2])$, with $d = d_2 = 1 + c_2\lambda$. Then N is also mixed Poisson, with $N \sim MP(\lambda, G_1G_2)$. If G_1 and G_2 are independent then we call $G = G_1G_2$ a *straight product.* In this case the contagion parameter for G is $c = c_1 + c_2 + c_1c_2$. The conditional r.v. $N|G_1$ has variance-to-mean ratio $d(G_1) = 1 + c_2G_1\lambda$. Should we wish to hold $d(G_1)$ constant, we may drop the independence of G_1, G_2, and assume that G_2 depends on G_1 as $G_2 = G_2^*[c_2/G_1]$ where G_2^* is a family of mixing distributions. With a slight abuse of notation, we drop the * and define the *twisted product* as $G_1 \bullet G_2 = G_1G_2[c_2/G_1]$. For a twisted product, $c = c_1 + c_2$, and $d|G_1 = d_2 = 1 + c_2\lambda$.

The claim count presented in [8] is concisely described as $N = NB[G_1\lambda, d]$. As d is fixed with respect to G_1, this is equivalent to $N = MP(\lambda, G_1 \bullet G_2)$, with $G_2 \sim$ Gamma and $c_2 = (d-1)/\lambda$. Now, its also the case that d is fixed with respect to λ, and thus the underlying negative binomial model (i.e., $N|G_1 = 1$) is of the ODP type. On the other hand, if $G_1 \sim$ gamma, then $N_1 = MP(\lambda, G_1)$ is a negative binomial model of the contagion type. If we set $c_1 = wc$, for some $0 <= w <= 1$, then $c = c_1 + c_2$ implies that $c_2 = (1-w)c$. Thus N can be considered a sort of credibility weighting between the ODP and contagion models.

The straight product formulation is seen in the "common shock" method for modeling correlation over several lines of business. This method assigns to the ith line of business the claim count $N_i = MP(\lambda_i, G_1G_{2,i})$. Here, G_1 is the common ("industy-based" in [10]) component and the $G_{2,i}$ are the line-specific components. As in equation (12), this generates a correlation of $\rho_{ij} = c_1/(\nu_i\nu_j)$ between lines i and j, $i \neq j$. A twisted product is also well-suited to this purpose, and produces the same correlations. As above, $c = c_1 + c_2$ allows us to consider the model as a credibility weighting, now between the common and line-specific components.

We do not have a closed-form formula for the skewness of $G = G_1 \bullet G_2$. However, suppose $\mu_3'(G_2) = \sum_{i=0}^{3} a_i c_2^i$. This is the case for $G_2 \sim$ gamma, and several others, but not for $G_2 \sim$ exponential. (The exponential is not a special case of gamma–unless $c = 1$–as the shift $s = 1 - \sqrt{c}$ is forced.) Then $\mu_3'(G) = a_0\mu_3'(G_1) + a_1(1 + c_1)c_2 + a_2c_2^2 + a_3c_2^3$, from which $\gamma(G)$ can be computed.

2. **Discrete Mixing Distributions** - The three-point "Hermite" distribution given by $Pr(1 + k\sqrt{3c}) = 2/3 - |k|/6,\ k = -1, 0, 1$ is used in [10]. This is an instance of the general discrete uniform with a mass at $G = 1$. A Poisson *mixing* distribution is an important limiting case of the framework presented in [19] and [18]. This is one example of *infinitely*

divisible mixing distributions, in which case the claim count can also be represented as a compound Poisson in the sense of [6].

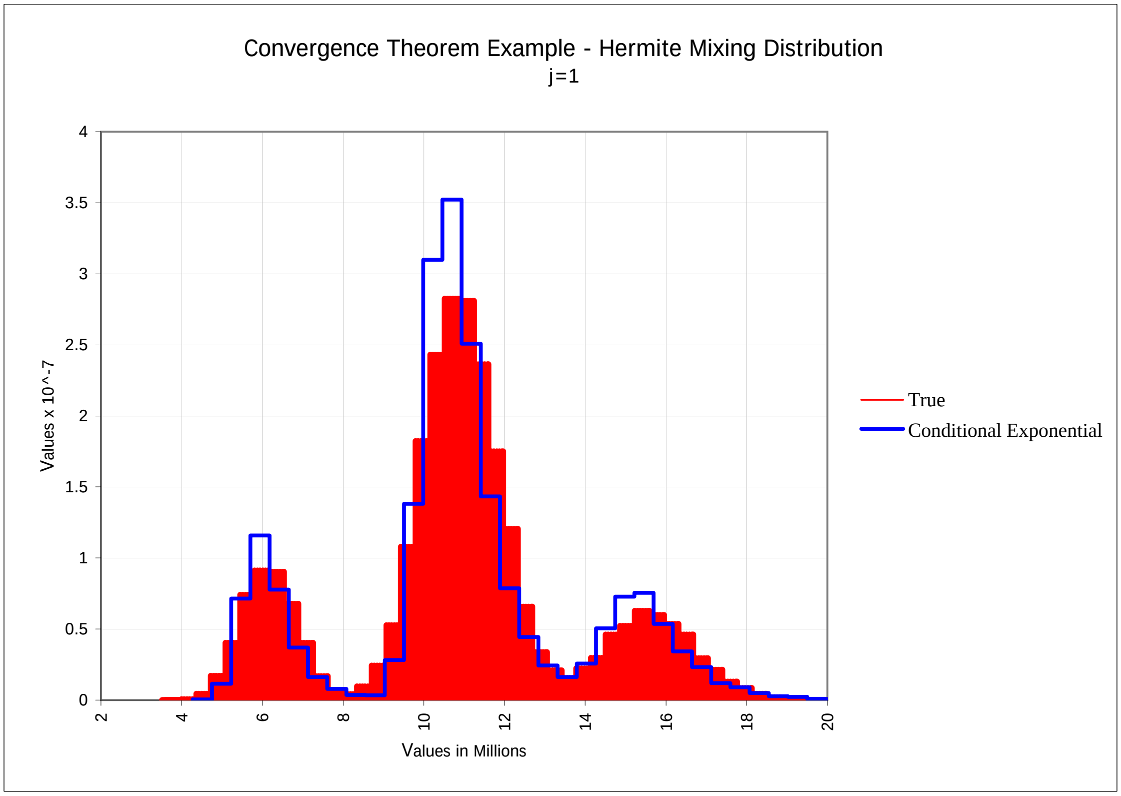
The shifted binomial is a very flexible choice for G. It converges to a Poisson as the the integer parameter $M \to \infty$, s fixed. For a given value of M, with $M \leq 1/c$, setting $s = 1 - \sqrt{Mc}$ results in a symmetric distribution different from the discrete uniform. In fact, $G \to$ normal as $c \to 0$ with $M = [1/c]$, and $s = 1 - \sqrt{Mc}$. In general, for any skewness value $\gamma > 0$, there is s such that $\gamma(G) = \gamma$, as long as $M\sqrt{c}(\sqrt{c} - \gamma) < 1$. (Note that this condition is satisfied trivially for $\gamma \geq \sqrt{c}$.)

In [21], Simar gives an algorithm for constructing a non-parametric maximum likelihood estimator (NPMLE) based on claim count observations. The NPMLE is then a finite mixing distribution whose size depends on the number of observations.

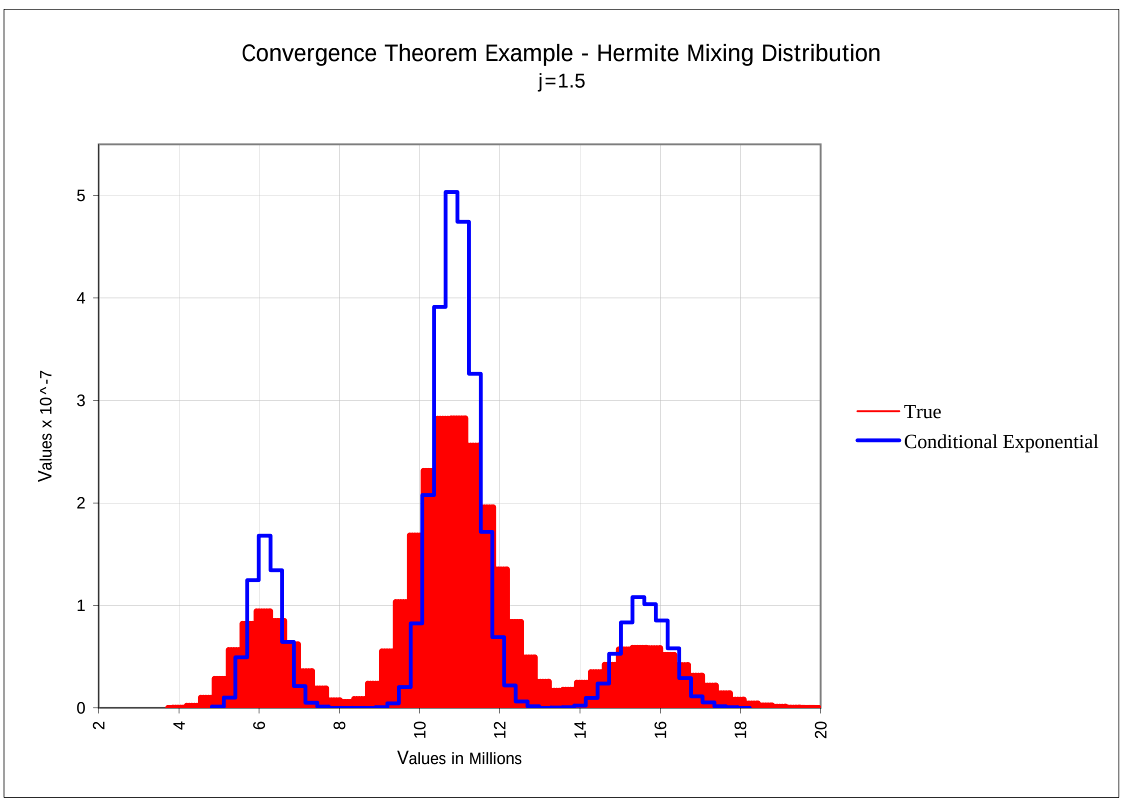
3. **Other Continuous Mixing Distributions** - The inverse Gaussian as a mixing distribution is the subject of [20] and is mentioned in [22], [18], and [12]. The resulting claim count is the Poisson-inverse Gaussian, or PIG. Given its populartiy as a model for aggregate distributions the lognormal is also a natural candidate as a mixing distribution.

C CAD Examples

Exhibit 1

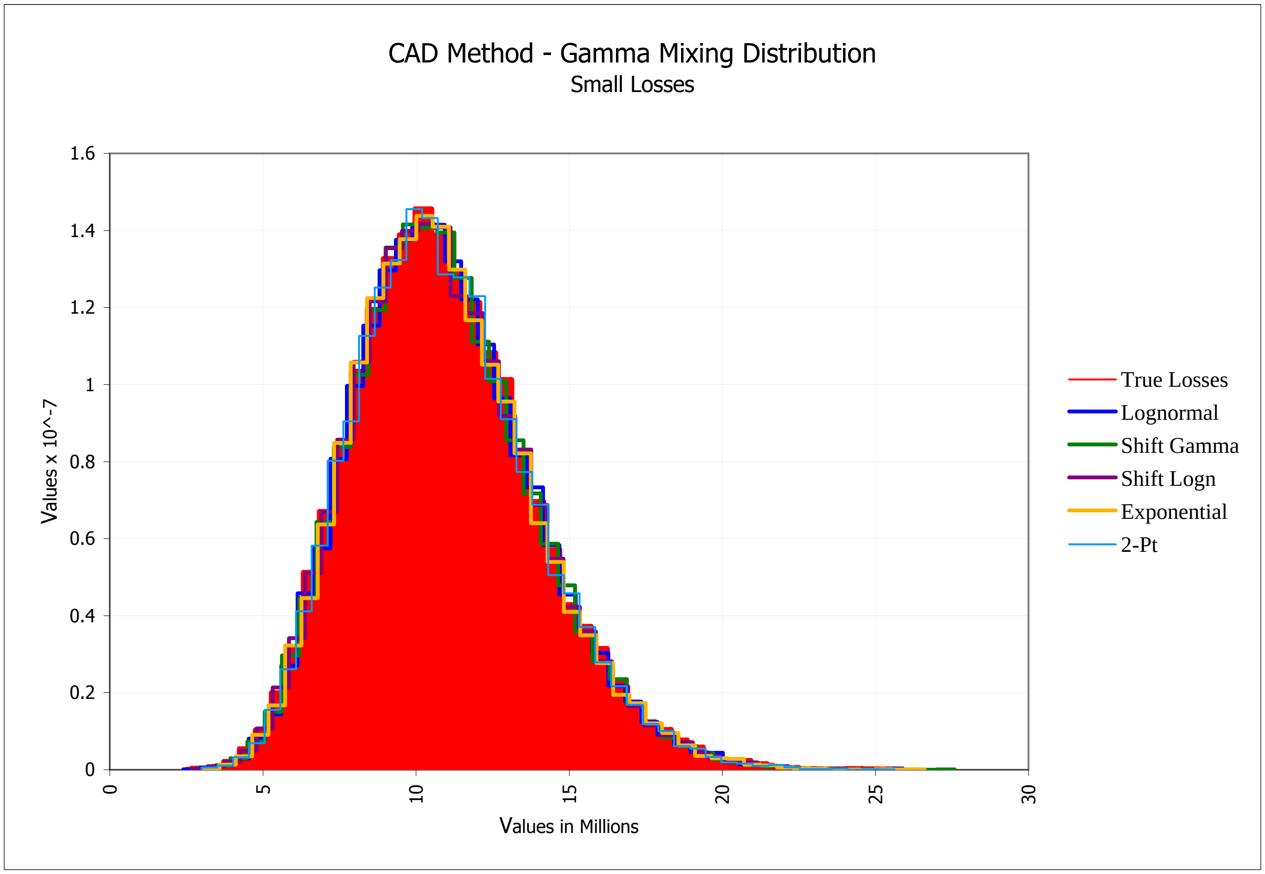


Convergence Theorem Example - Hermite Mixing Distribution
j=1
Values x 10^-7
4
3.5
3
2.5
2
1.5
1
0.5
0
2
4
6
8
10
12
14
16
18
20
Values in Millions
True
Conditional Exponential



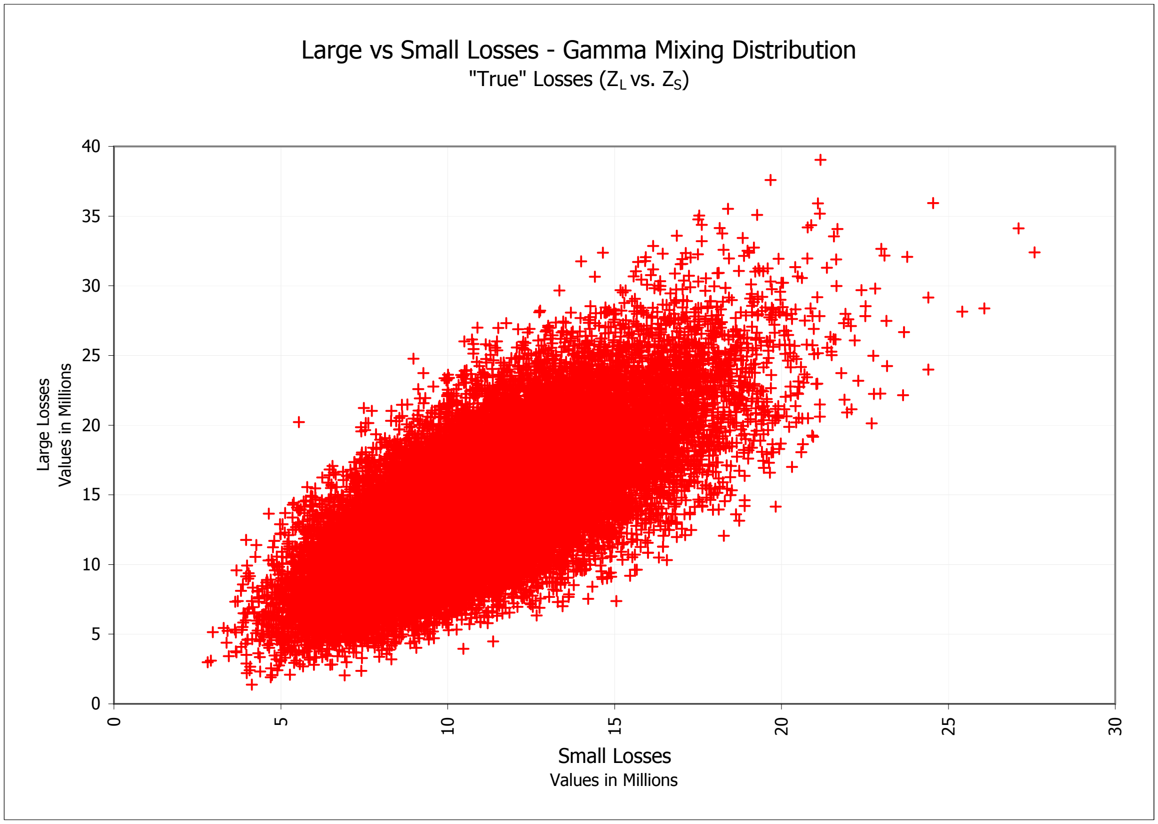
Convergence Theorem Example - Hermite Mixing Distribution
j=1.5
Values x 10^-7
5
4
3
2
1
0
2
4
6
8
10
12
14
16
18
20
Values in Millions
True
Conditional Exponential

Exhibit 2.1



Detail Stats - Gamma Mixing							
Loss Type	Large	Small	Small	Small	Small	Small	Small
Method	"True" (CRM Sim)	CAD Logn.	CAD S. Gamma	CAD S. Logn	CAD Exponential	CAD 2-pt.	"True" (CRM Sim)
Minimum	1,385,975	2,398,609	2,808,190	2,685,245	2,998,957	3,045,139	2,682,782
Maximum	39,042,540	25,898,300	27,588,630	25,894,870	25,626,020	26,605,610	25,627,990
Mean	14,140,170	10,938,800	10,937,300	10,943,990	10,939,010	10,937,040	10,931,920
Std Deviation	4,652,660	2,898,627	2,893,873	2,895,765	2,894,886	2,891,107	2,889,024
Variance	2.16473E+13	8.40204E+12	8.3745E+12	8.38546E+12	8.38037E+12	8.3585E+12	8.34646E+12
Skewness	**0.532**	**0.5071**	**0.5040**	**0.4819**	**0.4904**	**0.5250**	**0.4949**
CV	**0.329**	**0.2650**	**0.2646**	**0.2646**	**0.2646**	**0.2643**	**0.2643**
Skew-Nu	**1.618**	**1.9139**	**1.9050**	**1.8213**	**1.8531**	**1.9860**	**1.8725**
Mode	11,839,840	10,443,110	10,444,170	9,741,620	9,137,559	10,569,540	10,023,330
5% Perc	7,280,964	6,617,527	6,616,314	6,616,352	6,625,957	6,646,187	6,628,636
10% Perc	8,478,519	7,405,594	7,398,611	7,405,873	7,405,128	7,443,271	7,402,793
15% Perc	9,427,957	7,977,141	7,977,279	7,996,510	7,976,839	7,995,743	7,983,588
20% Perc	10,153,280	8,455,025	8,456,465	8,461,667	8,455,437	8,454,707	8,465,613
25% Perc	10,815,190	8,869,101	8,878,454	8,888,060	8,855,242	8,867,610	8,873,186
30% Perc	11,444,160	9,251,367	9,269,360	9,269,123	9,255,394	9,250,547	9,259,721
35% Perc	12,006,930	9,625,078	9,627,173	9,644,261	9,634,785	9,618,922	9,630,174
40% Perc	12,584,930	9,981,641	9,984,575	9,983,185	9,993,351	9,983,671	9,988,756
45% Perc	13,161,940	10,341,280	10,341,520	10,339,860	10,323,750	10,333,640	10,332,560
50% Perc	13,720,500	10,687,030	10,689,310	10,694,390	10,687,300	10,673,790	10,677,960
55% Perc	14,293,300	11,059,730	11,045,600	11,041,930	11,070,400	11,032,790	11,041,920
60% Perc	14,890,360	11,433,110	11,409,440	11,445,670	11,436,400	11,406,070	11,427,120
65% Perc	15,557,820	11,838,890	11,807,440	11,862,800	11,844,620	11,814,010	11,830,620
70% Perc	16,231,410	12,259,280	12,251,480	12,282,940	12,255,120	12,252,290	12,260,860
75% Perc	17,038,840	12,749,390	12,735,450	12,767,710	12,742,260	12,740,690	12,741,150
80% Perc	17,900,140	13,276,090	13,296,010	13,312,140	13,294,120	13,272,500	13,256,540
85% Perc	18,937,040	13,942,170	13,935,960	13,938,120	13,950,790	13,907,760	13,903,870
90% Perc	20,365,650	14,781,560	14,794,910	14,778,790	14,808,720	14,764,230	14,745,370
95% Perc	22,465,010	16,101,710	16,124,400	16,093,960	16,095,310	16,130,030	16,089,070

Exhibit 2.2


Large vs Small Losses - Gamma Mixing Distribution
"True" Losses (Z_L vs. Z_S)
Large Losses
Values in Millions
0
5
10
15
20
25
30
35
40
Small Losses
Values in Millions

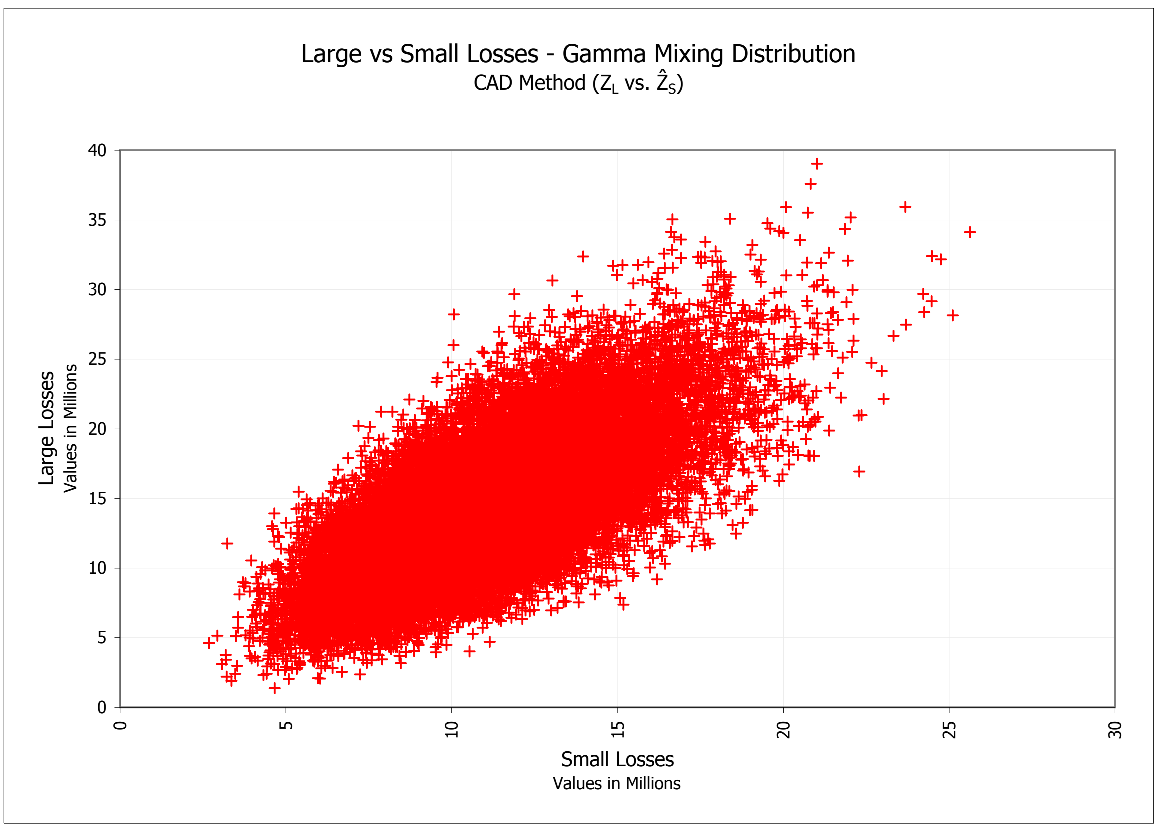
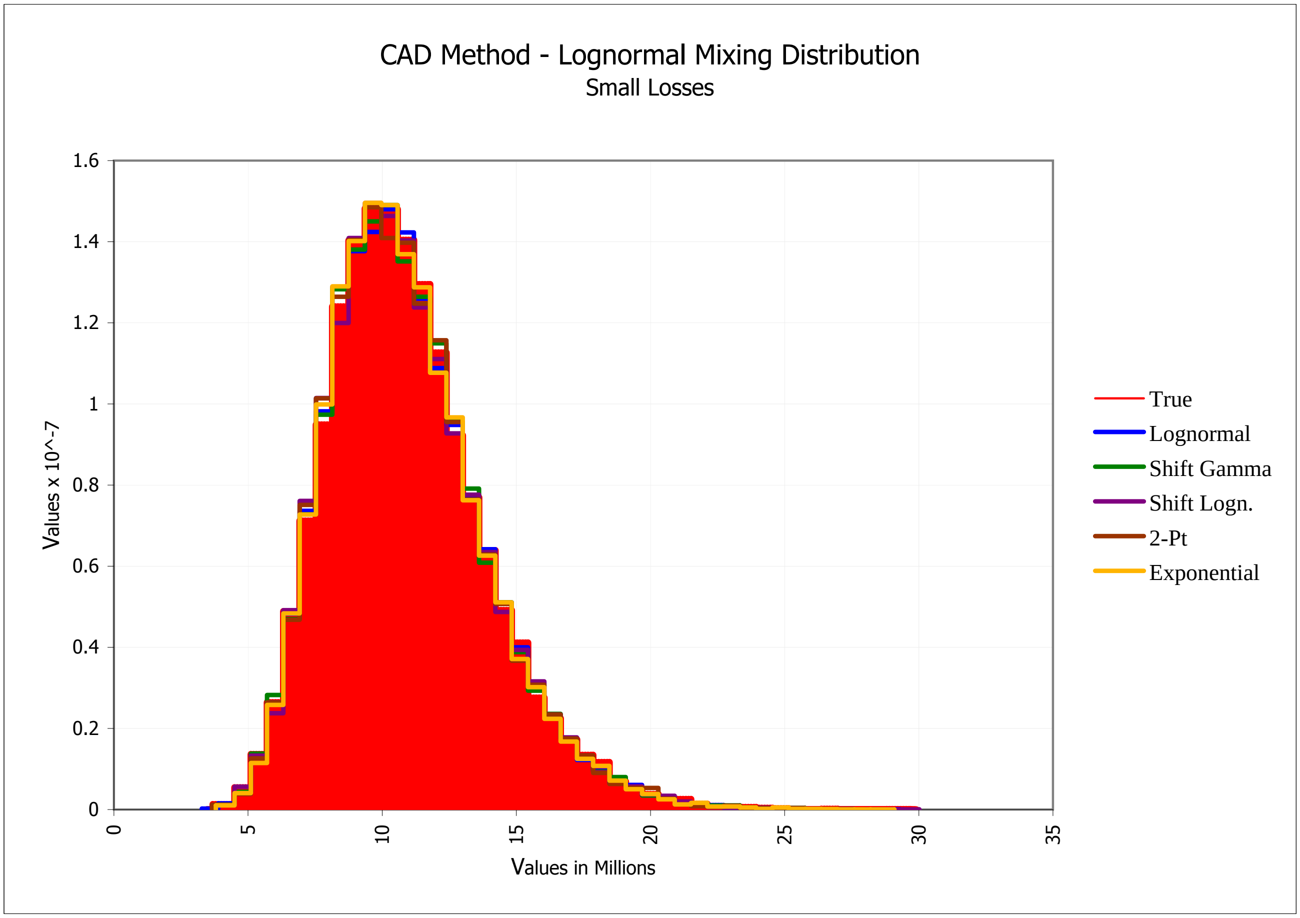
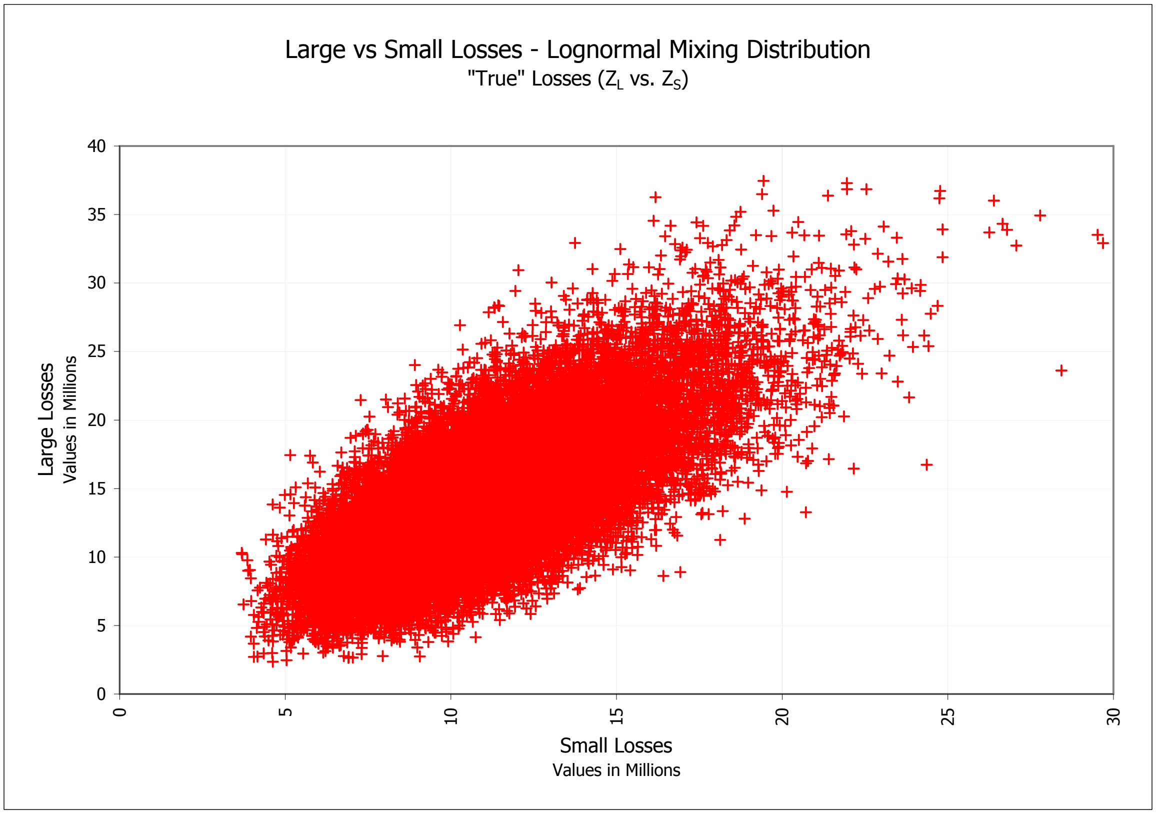

Large vs Small Losses - Gamma Mixing Distribution
CAD Method (Z_L vs. $\hat{Z}_S$)
Large Losses
Values in Millions
0
5
10
15
20
25
30
35
40
Small Losses
Values in Millions

Exhibit 3.1



Detail Stats - Lognormal Mixing							
Loss Type	Large	Small	Small	Small	Small	Small	Small
Method	"True" (CRM Sim)	CAD Logn.	CAD S. Gamma	CAD S. Logn	CAD Exponential	CAD 2-pt.	"True" (CRM Sim)
Minimum	2,354,636	3,279,025	3,784,197	3,671,143	3,656,671	3,801,383	3,686,909
Maximum	37,452,880	29,297,280	28,652,320	30,017,500	27,943,100	29,091,590	29,686,280
Mean	14,086,670	10,905,680	10,904,890	10,901,090	10,899,620	10,897,280	10,906,430
Std Deviation	4,663,644	2,896,549	2,898,968	2,905,355	2,888,524	2,884,288	2,899,240
Variance	2.17496E+13	8.39E+12	8.40402E+12	8.44109E+12	8.34357E+12	8.31912E+12	8.40559E+12
Skewness	**0.657**	**0.7345**	**0.7532**	**0.7511**	**0.7555**	**0.7820**	**0.7548**
CV	**0.331**	**0.2656**	**0.2658**	**0.2665**	**0.2650**	**0.2647**	**0.2658**
Skew-Nu	**1.985**	**2.7654**	**2.8332**	**2.8182**	**2.8509**	**2.9546**	**2.8393**
Mode	13,312,870	9,229,467	10,256,040	10,247,740	9,900,903	9,298,220	9,593,592
5% Perc	7,379,982	6,807,512	6,786,931	6,814,347	6,839,257	6,830,828	6,783,801
10% Perc	8,522,062	7,512,230	7,506,557	7,503,218	7,531,037	7,537,819	7,511,076
15% Perc	9,390,949	8,032,948	8,046,047	8,007,588	8,034,989	8,046,408	8,052,007
20% Perc	10,102,340	8,449,887	8,447,854	8,445,220	8,455,437	8,468,788	8,484,241
25% Perc	10,756,510	8,834,509	8,828,217	8,834,182	8,813,141	8,837,245	8,866,154
30% Perc	11,354,550	9,198,359	9,198,009	9,201,309	9,186,013	9,198,724	9,204,506
35% Perc	11,937,640	9,551,106	9,542,255	9,528,743	9,529,370	9,527,572	9,562,121
40% Perc	12,484,700	9,891,079	9,887,307	9,866,483	9,868,970	9,867,238	9,890,388
45% Perc	13,034,730	10,224,090	10,223,390	10,216,780	10,210,100	10,201,090	10,234,600
50% Perc	13,589,010	10,570,880	10,557,930	10,549,800	10,563,310	10,539,160	10,571,090
55% Perc	14,173,760	10,913,890	10,924,320	10,898,780	10,914,400	10,887,340	10,918,910
60% Perc	14,742,170	11,276,170	11,304,320	11,270,320	11,281,810	11,266,530	11,280,100
65% Perc	15,378,140	11,680,180	11,694,860	11,669,240	11,685,900	11,647,070	11,666,960
70% Perc	16,102,160	12,113,640	12,118,480	12,111,900	12,106,920	12,094,540	12,089,420
75% Perc	16,888,410	12,602,800	12,576,840	12,575,800	12,572,790	12,572,670	12,563,000
80% Perc	17,763,110	13,162,890	13,152,030	13,162,970	13,135,720	13,120,340	13,138,160
85% Perc	18,831,010	13,835,790	13,811,970	13,845,860	13,791,750	13,804,000	13,818,980
90% Perc	20,265,510	14,747,880	14,734,090	14,753,650	14,695,780	14,693,850	14,735,270
95% Perc	22,529,300	16,170,290	16,180,220	16,216,460	16,162,980	16,163,700	16,199,130

Exhibit 3.2


Large vs Small Losses - Lognormal Mixing Distribution
"True" Losses (Z_L vs. Z_S)
Large Losses
Values in Millions
Small Losses
Values in Millions
0
5
10
15
20
25
30
35
40

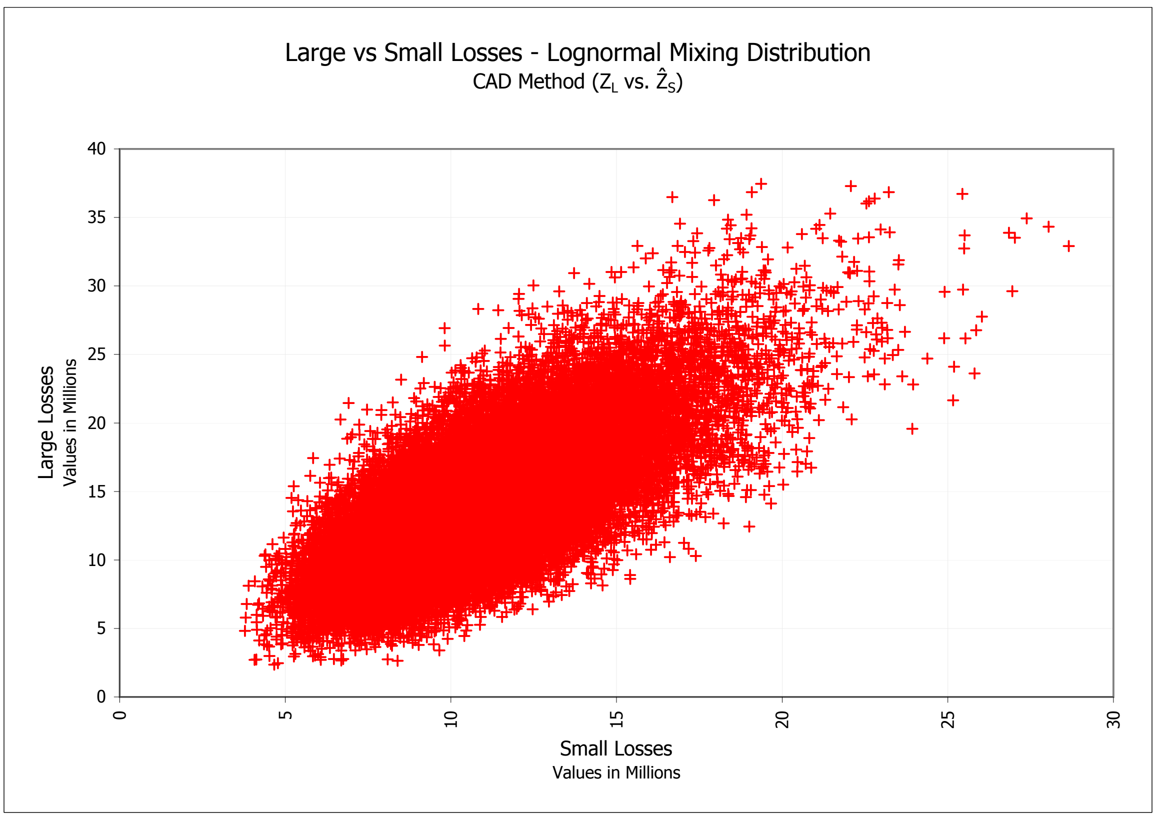
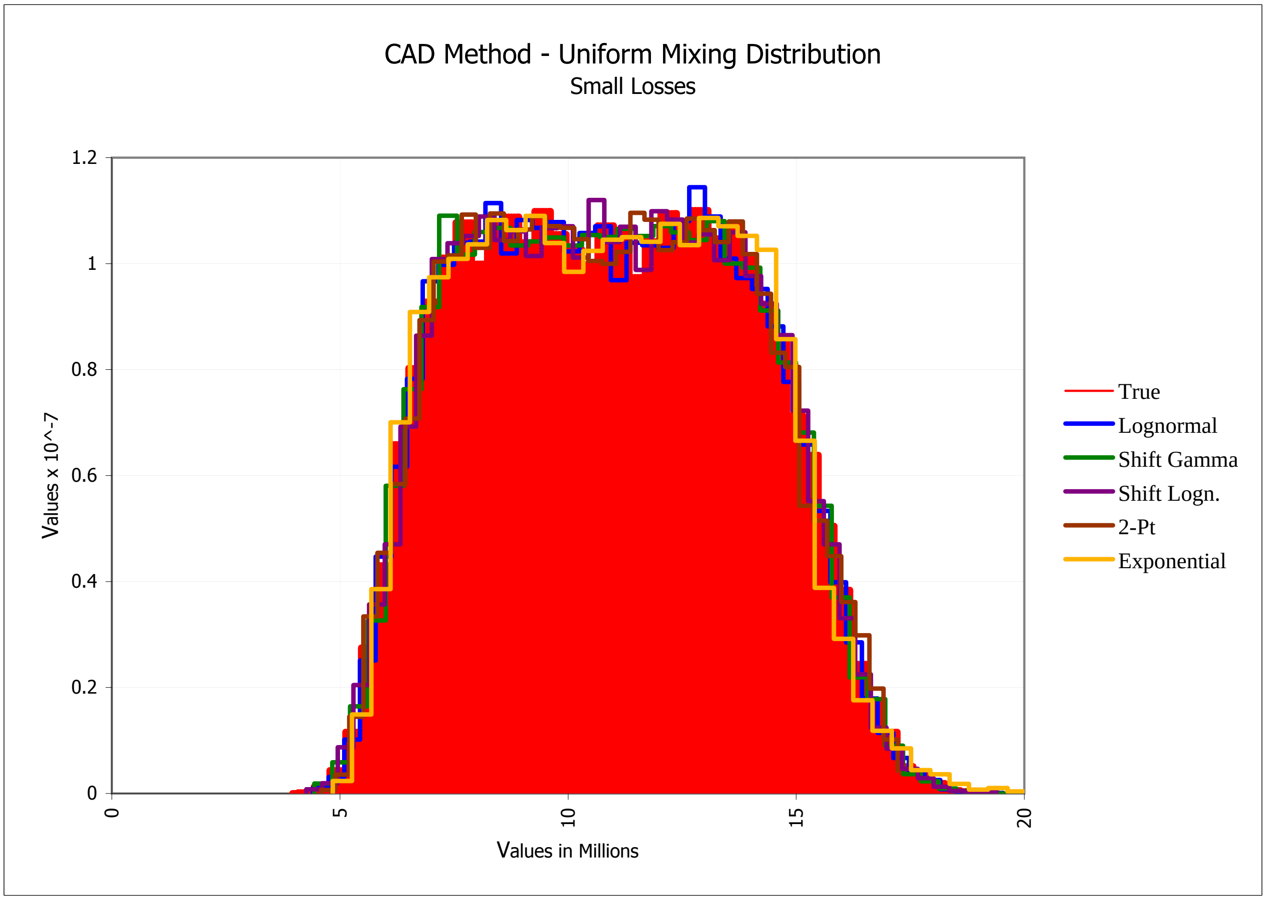
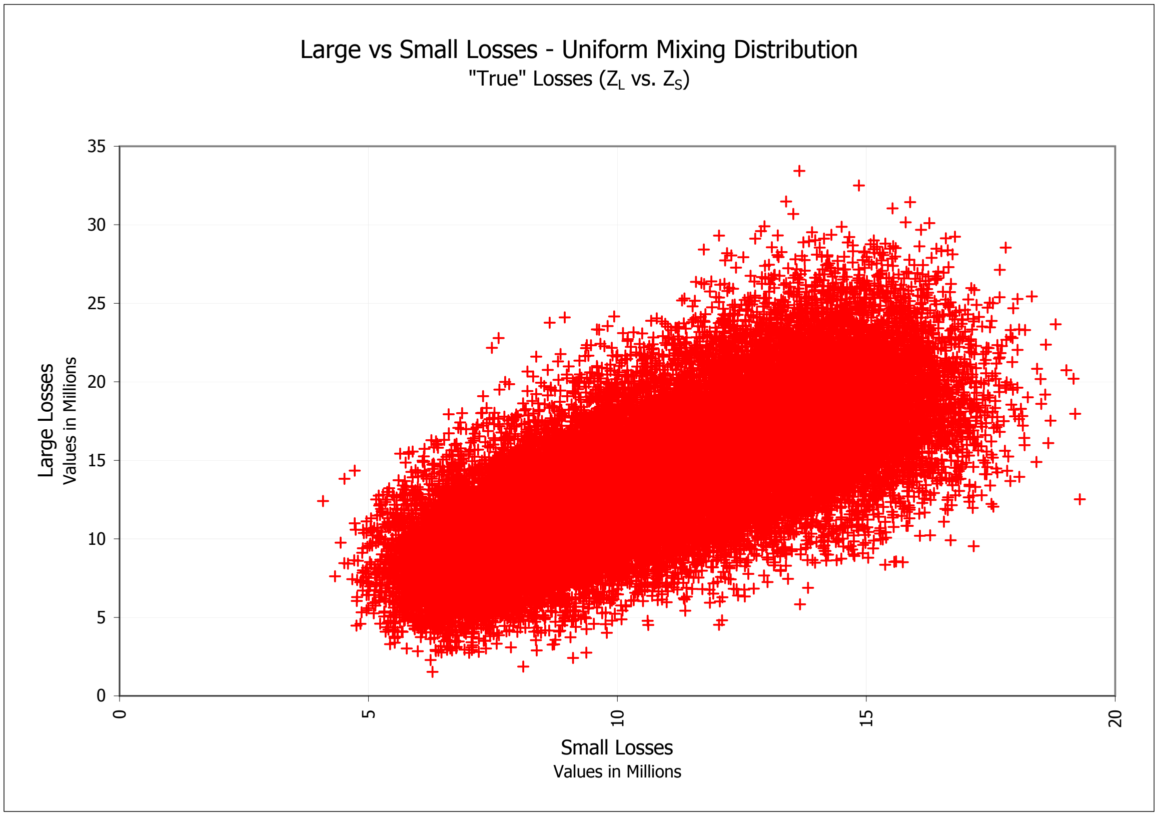

Large vs Small Losses - Lognormal Mixing Distribution
CAD Method (Z_L vs. $\hat{Z}_S$)
Large Losses
Values in Millions
Small Losses
Values in Millions
0
5
10
15
20
25
30
35
40

Exhibit 4.1



Detail Stats - Uniform Mixing							
Loss Type	Large	Small	Small	Small	Small	Small	Small
Method	"True" (CRM Sim)	CAD Logn.	CAD S. Gamma	CAD S. Logn	CAD Exponential	CAD 2-pt.	"True" (CRM Sim)
Minimum	2,622,778	4,238,870	4,287,148	4,468,165	4,670,314	4,702,374	4,487,489
Maximum	32,122,340	18,258,020	18,495,580	19,186,070	17,875,910	22,198,040	18,737,170
Mean	14,053,220	10,889,670	10,907,560	10,898,670	10,890,100	10,895,280	10,893,480
Std Deviation	4,635,183	2,851,964	2,868,312	2,871,454	2,853,996	2,866,448	2,870,310
Variance	2.14849E+13	8.1337E+12	8.22722E+12	8.24525E+12	8.14529E+12	8.21653E+12	8.23868E+12
Skewness	**0.317**	**0.0601**	**0.0552**	**0.0624**	**0.0553**	**0.1008**	**0.0575**
CV	**0.330**	**0.2619**	**0.2630**	**0.2635**	**0.2621**	**0.2631**	**0.2635**
Skew-Nu	**0.960**	**0.2295**	**0.2099**	**0.2367**	**0.2108**	**0.3832**	**0.2182**
Mode	13,955,570	11,017,940	6,980,460	10,483,140	8,609,236	14,108,960	12,211,590
5% Perc	7,048,361	6,452,205	6,455,139	6,397,590	6,378,900	6,456,264	6,419,141
10% Perc	8,138,355	7,055,768	7,013,859	7,003,973	7,048,737	7,014,243	7,021,224
15% Perc	9,061,992	7,556,159	7,537,291	7,552,863	7,552,022	7,506,401	7,519,121
20% Perc	9,832,303	8,057,059	8,037,449	8,029,881	8,068,227	7,994,043	8,006,109
25% Perc	10,508,630	8,513,653	8,513,914	8,529,380	8,570,468	8,496,666	8,504,713
30% Perc	11,221,460	9,005,443	9,001,551	8,979,073	9,023,686	8,992,163	9,005,018
35% Perc	11,867,710	9,487,384	9,519,624	9,481,191	9,508,290	9,477,224	9,504,768
40% Perc	12,520,510	9,952,474	9,959,408	9,980,229	9,972,253	9,973,597	9,957,272
45% Perc	13,151,170	10,395,700	10,469,080	10,447,890	10,437,390	10,457,010	10,423,910
50% Perc	13,770,240	10,893,210	10,916,910	10,885,080	10,868,990	10,902,820	10,891,530
55% Perc	14,387,300	11,351,630	11,342,420	11,337,470	11,302,950	11,362,770	11,371,680
60% Perc	15,052,900	11,818,450	11,799,030	11,827,650	11,747,140	11,827,170	11,831,270
65% Perc	15,751,770	12,266,250	12,275,180	12,301,380	12,233,950	12,246,880	12,247,990
70% Perc	16,481,540	12,716,240	12,744,370	12,728,420	12,705,610	12,714,500	12,729,500
75% Perc	17,263,490	13,187,060	13,228,660	13,177,340	13,203,610	13,171,750	13,199,210
80% Perc	18,107,500	13,637,550	13,672,250	13,665,560	13,675,010	13,672,670	13,665,310
85% Perc	19,074,990	14,168,400	14,169,310	14,175,580	14,130,870	14,144,830	14,148,740
90% Perc	20,253,470	14,693,240	14,774,350	14,737,180	14,695,780	14,665,920	14,733,240
95% Perc	22,030,010	15,450,310	15,511,660	15,494,220	15,508,650	15,398,370	15,511,520

Exhibit 4.2


Large vs Small Losses - Uniform Mixing Distribution
"True" Losses (Z_L vs. Z_S)
Large Losses
Values in Millions
Small Losses
Values in Millions

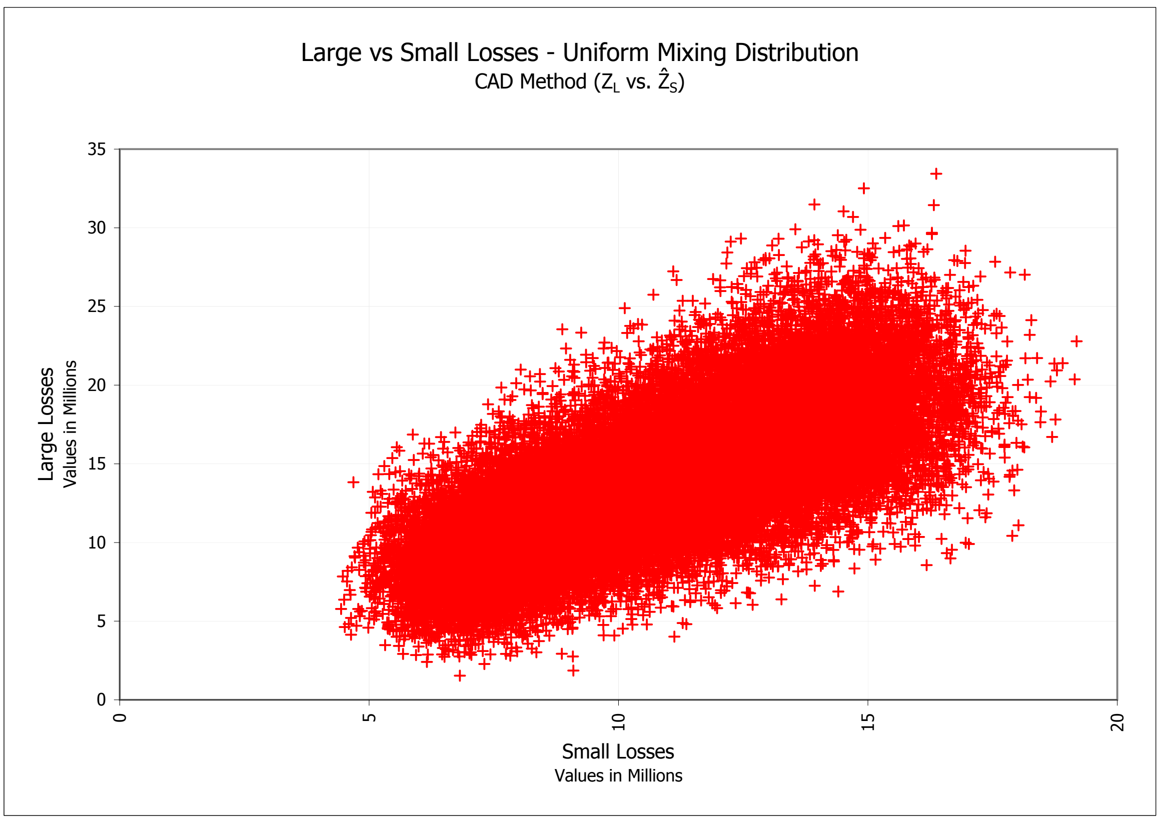
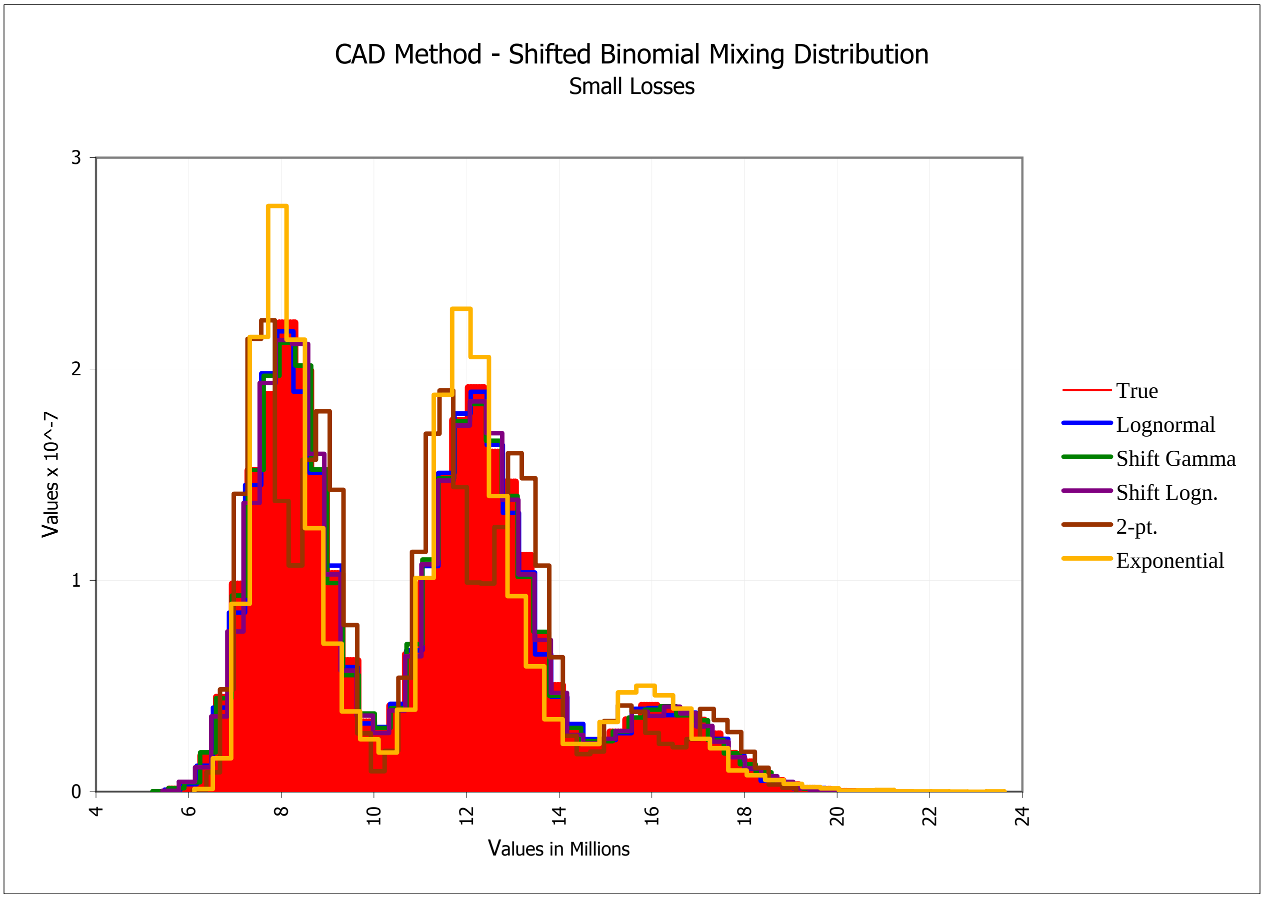

Large vs Small Losses - Uniform Mixing Distribution
CAD Method (Z_L vs. $\hat{Z}_S$)
Large Losses
Values in Millions
Small Losses
Values in Millions

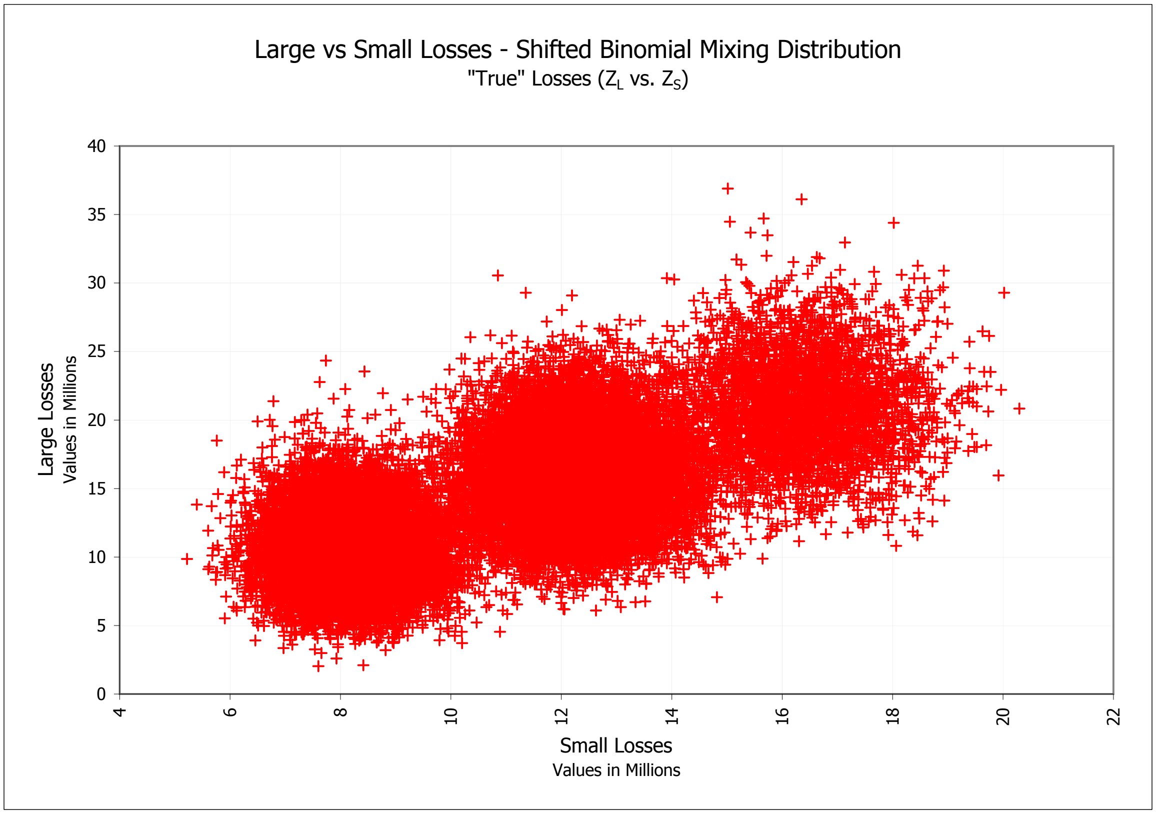
Exhibit 5.1



Detail Stats - Shifted Biniomial

Loss Type Method	Large "True" (CRM Sim)	Small CAD Logn.	Small CAD S. Gamma	Small CAD S. Logn	Small CAD Exponential	Small CAD 2-pt.	Small "True" (CRM Sim)
Minimum	2,029,372	5,480,892	5,220,685	5,438,516	6,086,327	6,125,506	5,578,668
Maximum	36,899,030	20,782,580	20,297,310	20,802,180	19,114,040	23,614,350	20,532,800
Mean	14,126,680	10,956,630	10,947,030	10,951,330	10,953,510	10,946,910	10,949,270
Std Deviation	4,631,549	2,889,085	2,882,983	2,889,601	2,896,795	2,889,096	2,880,868
Variance	2.14513E+13	8.34681E+12	8.31159E+12	8.3498E+12	8.39142E+12	8.34688E+12	8.2994E+12
Skewness	**0.515**	**0.4845**	**0.4885**	**0.4872**	**0.4863**	**0.5301**	**0.4920**
CV	**0.328**	**0.2637**	**0.2634**	**0.2639**	**0.2645**	**0.2639**	**0.2631**
Skew-Nu	**1.571**	**1.8374**	**1.8550**	**1.8465**	**1.8386**	**2.0086**	**1.8700**
Mode	12,964,350	8,087,866	8,148,018	8,147,161	7,299,558	8,102,896	8,147,660
5% Perc	7,483,617	7,227,355	7,225,197	7,235,228	7,216,417	7,361,047	7,229,344
10% Perc	8,535,075	7,570,712	7,577,371	7,579,443	7,468,091	7,599,695	7,574,011
15% Perc	9,306,652	7,828,013	7,841,232	7,836,373	7,677,954	7,797,241	7,843,340
20% Perc	9,965,355	8,066,383	8,081,884	8,077,192	7,929,945	7,978,476	8,080,636
25% Perc	10,604,400	8,292,712	8,314,742	8,302,414	8,350,285	8,161,311	8,308,913
30% Perc	11,226,030	8,556,097	8,553,878	8,543,078	8,707,333	8,385,700	8,549,825
35% Perc	11,833,320	8,867,372	8,849,983	8,827,609	8,980,929	8,692,694	8,862,549
40% Perc	12,426,840	9,308,444	9,295,946	9,276,051	9,299,908	9,299,985	9,303,648
45% Perc	13,021,220	10,554,790	10,493,120	10,543,550	10,690,050	10,919,380	10,509,370
50% Perc	13,661,470	11,250,190	11,224,580	11,240,710	11,176,100	11,368,280	11,244,800
55% Perc	14,291,840	11,625,320	11,602,130	11,615,890	11,450,980	11,632,780	11,611,090
60% Perc	14,945,040	11,923,280	11,906,950	11,918,820	11,725,980	11,850,840	11,900,920
65% Perc	15,622,390	12,183,350	12,177,440	12,192,390	12,085,760	12,071,050	12,163,850
70% Perc	16,370,580	12,450,640	12,452,890	12,462,020	12,593,970	12,294,140	12,441,100
75% Perc	17,150,500	12,752,340	12,755,160	12,758,750	12,977,290	12,571,500	12,758,940
80% Perc	17,999,210	13,127,510	13,112,140	13,122,760	13,294,120	12,960,450	13,111,080
85% Perc	19,036,000	13,675,010	13,649,720	13,648,090	13,678,680	13,653,300	13,627,890
90% Perc	20,423,720	15,070,390	15,084,120	15,098,890	15,033,470	15,360,840	15,105,940
95% Perc	22,467,470	16,539,630	16,510,020	16,539,500	16,659,090	16,393,870	16,502,230

Exhibit 5.2


Large vs Small Losses - Shifted Binomial Mixing Distribution
"True" Losses (Z_L vs. Z_S)
Large Losses
Values in Millions
Small Losses
Values in Millions
0
5
10
15
20
25
30
35
40
4
6
8
10
12
14
16
18
20
22

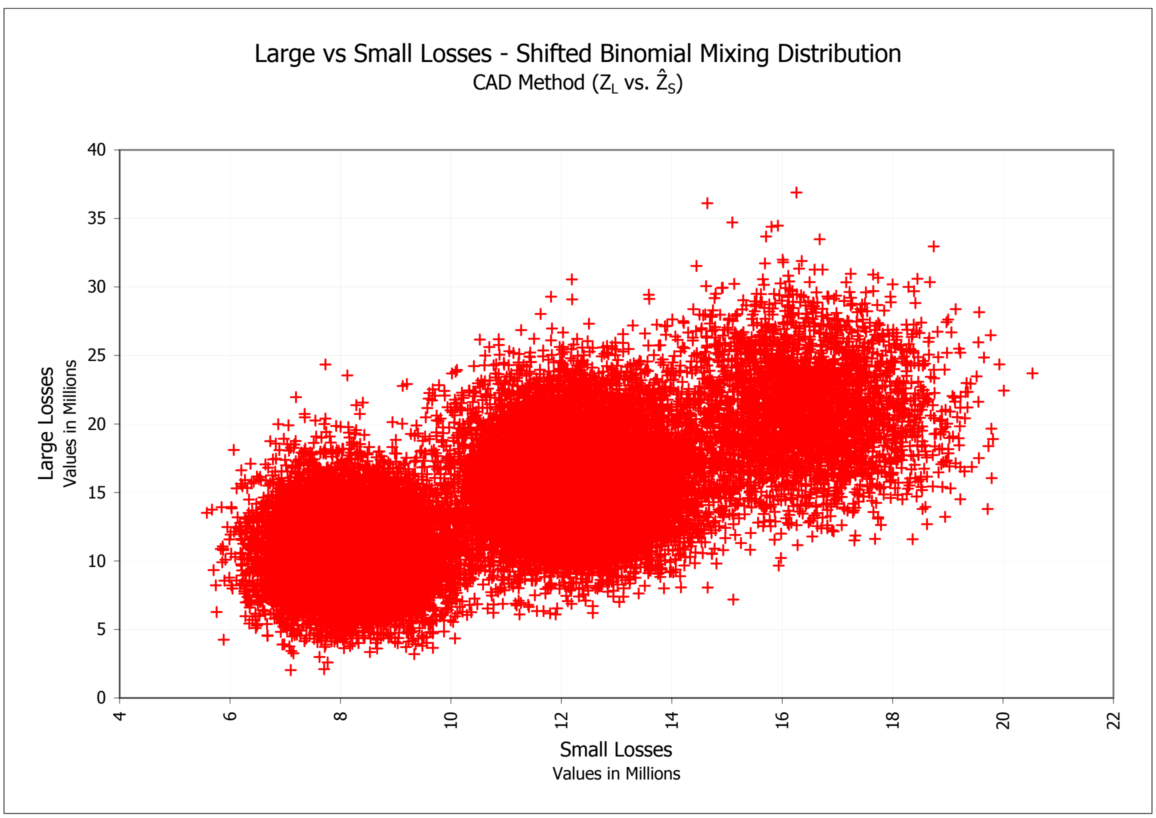
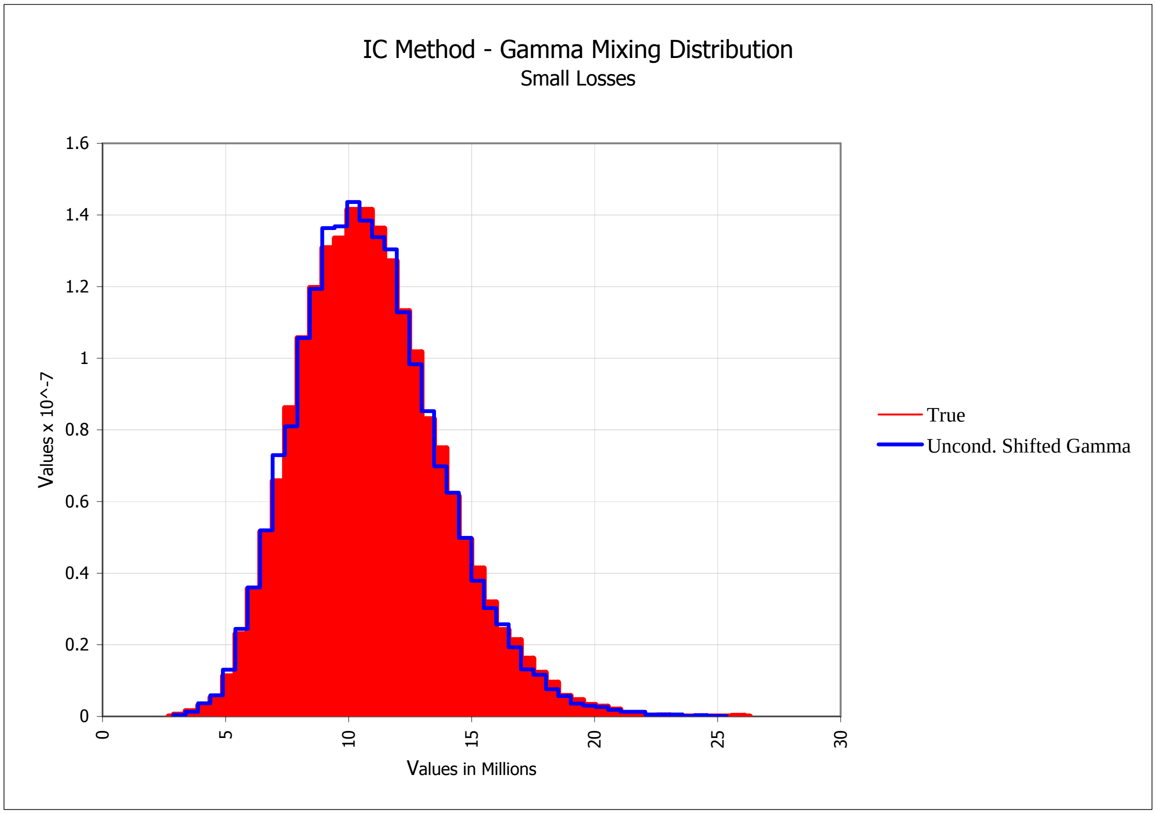

Large vs Small Losses - Shifted Binomial Mixing Distribution
CAD Method (Z_L vs. $\hat{Z}_S$)
Large Losses
Values in Millions
Small Losses
Values in Millions
0
5
10
15
20
25
30
35
40
4
6
8
10
12
14
16
18
20
22

Exhibit 6


IC Method - Gamma Mixing Distribution
Small Losses
Values x 10^-7
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
Values in Millions
True
Uncond. Shifted Gamma

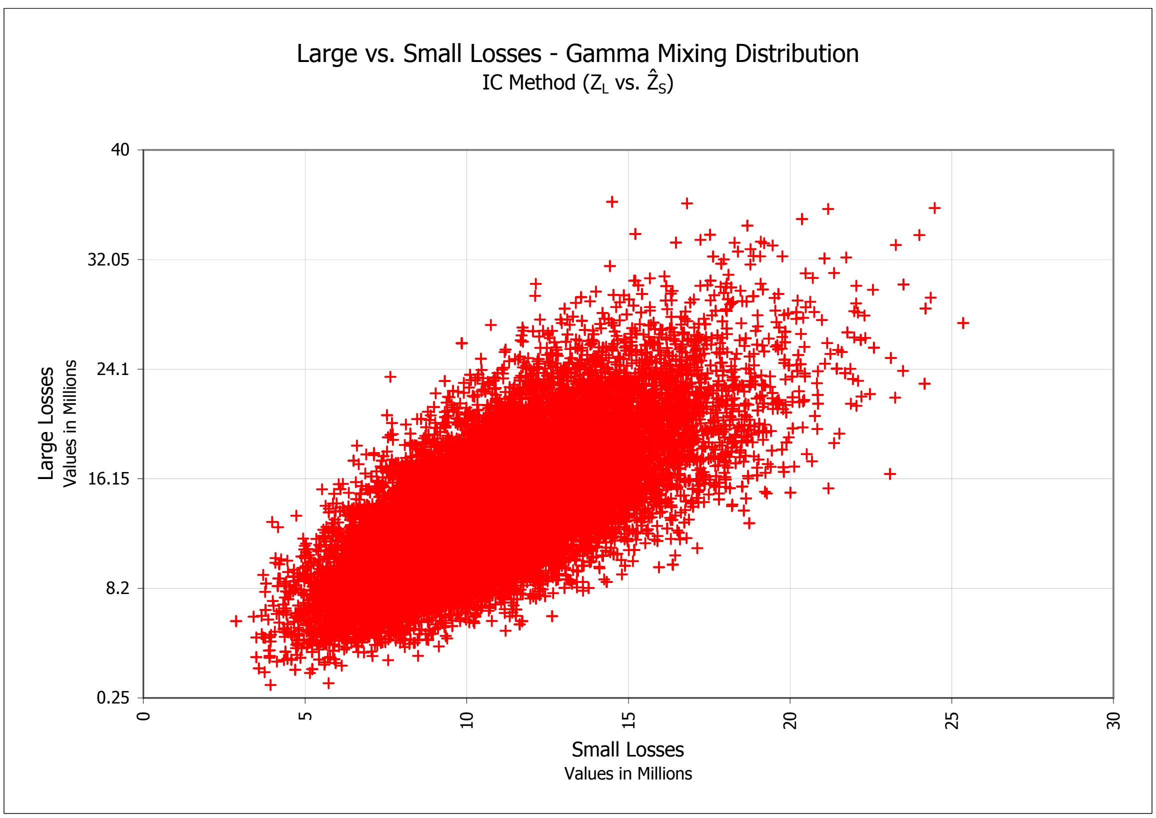
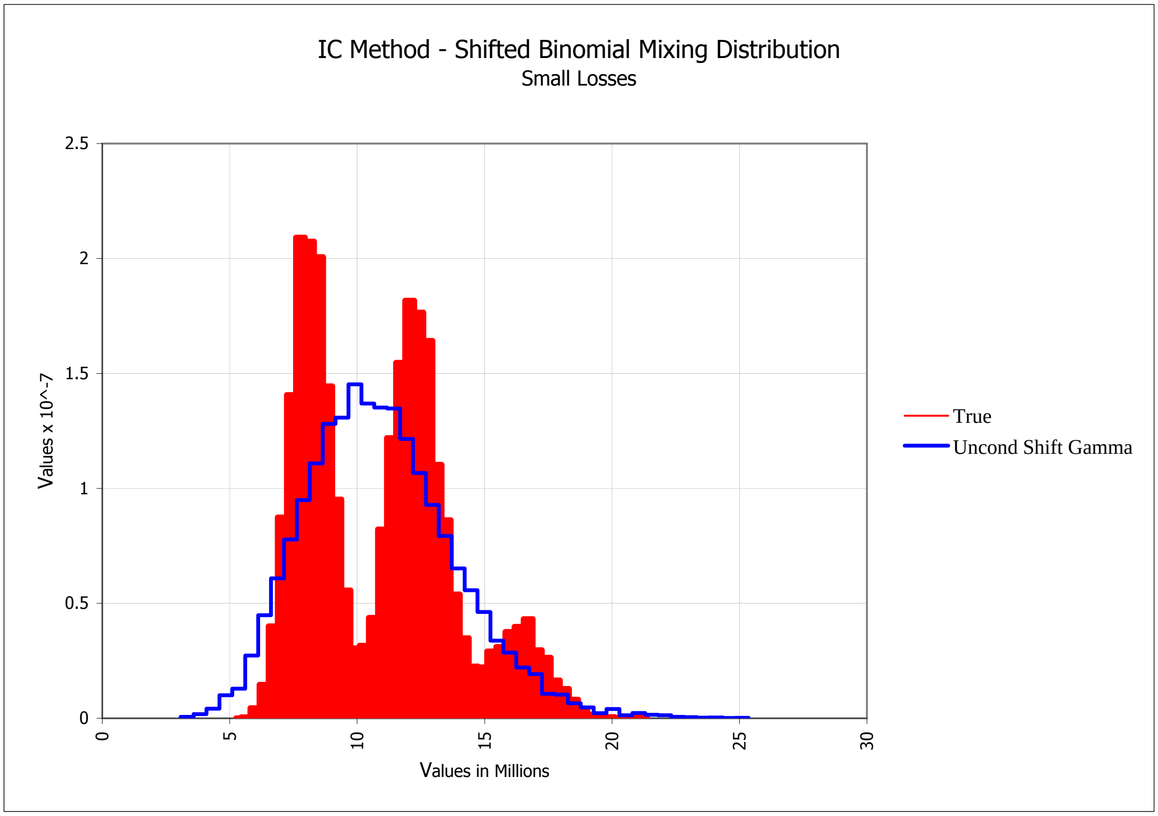

Large vs. Small Losses - Gamma Mixing Distribution
IC Method (Z_L vs. $\hat{Z}_S$)
Large Losses
Values in Millions
40
32.05
24.1
16.15
8.2
0.25
0
5
10
15
20
25
30
Small Losses
Values in Millions

Exhibit 7.1


IC Method - Shifted Binomial Mixing Distribution
Small Losses
Values x 10^-7
Values in Millions
True
Uncond Shift Gamma
0
0.5
1
1.5
2
2.5
0
5
10
15
20
25
30

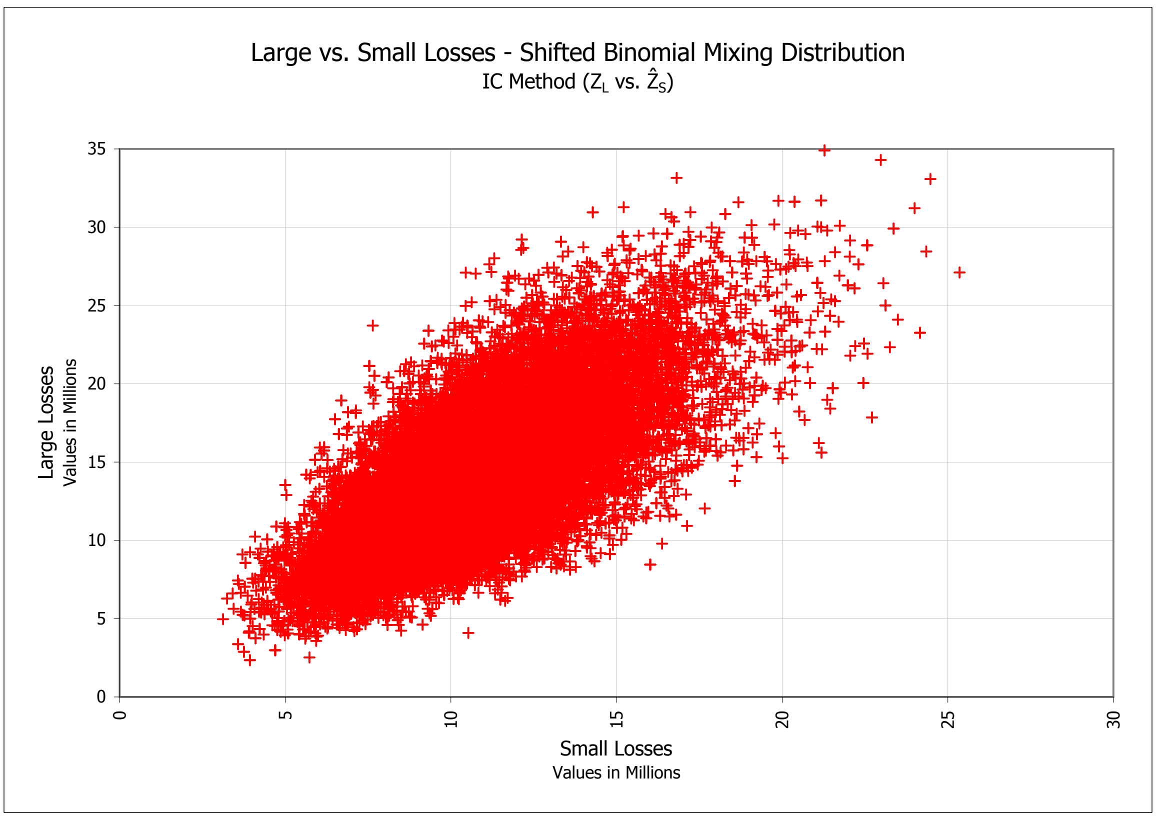
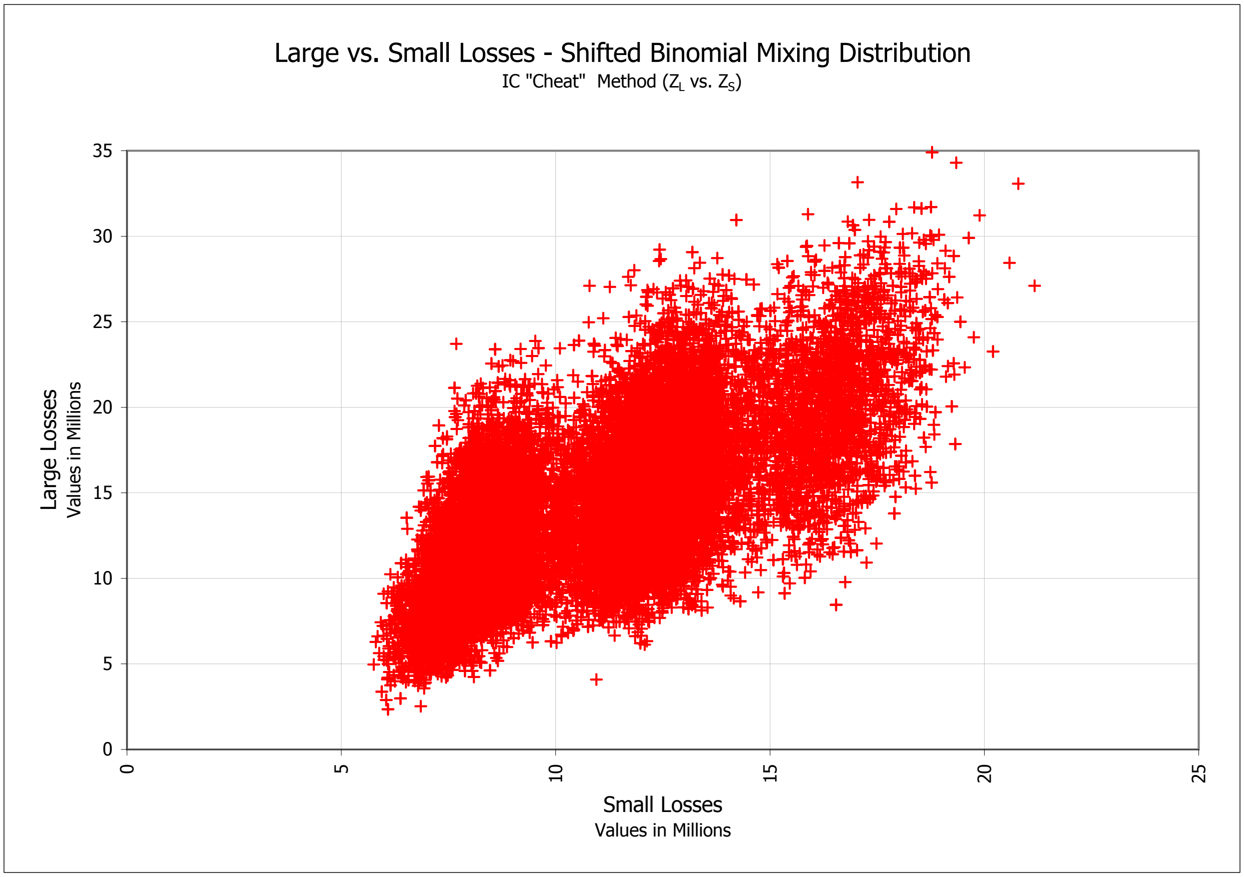

Large vs. Small Losses - Shifted Binomial Mixing Distribution
IC Method (Z_L vs. $\hat{Z}_S$)
Large Losses
Values in Millions
Small Losses
Values in Millions
0
5
10
15
20
25
30
35
0
5
10
15
20
25
30

Exhibit 7.2


Large vs. Small Losses - Shifted Binomial Mixing Distribution
IC "Cheat" Method (Z_L vs. Z_S)
Large Losses
Values in Millions
0
5
10
15
20
25
30
35
Small Losses
Values in Millions
0
5
10
15
20
25

International Evidence on Medical Spending

Robert D. Lieberthal

Abstract

U.S. medical spending is high by measures including the level of spending, level of spending per capita, and level of spending as a share of GDP. U.S. medical spending growth is average by measures including the annual growth rate, annual growth rate per capita, and annual growth in spending as a percent of GDP. The volatility of U.S. medical spending growth is low by measures including the standard deviation, skew, and excess kurtosis.

Foreign healthcare systems, with a much larger government involvement, have not been able to control medical spending growth better than the U.S. with its mixed system. Foreign cost curves start at a lower level, but increase as quickly or even faster. In many countries, the variance around the trend is high, or a single trend over time does not exist. The implication is that it is difficult to find a foreign solution to the U.S.'s problems with high medical spending, and that the U.S. may be a world leader in terms of minimizing medical spending volatility.

If the U.S. healthcare cost curve comes to resemble that of other countries, the risk of long-tailed lines of insurance linked to the cost of medical care will increase. The healthcare cost curve is a macroeconomic process, so there may be no ways for insurers to bend their cost curve. Insurers may be able to use market solutions, such as prediction markets, inflation-indexed bonds, and futures contracts, to improve prediction and hedging of long-term medical spending growth. My recommendations for insurers are cognizance and caution when writing long-tailed lines of insurance linked to medical spending.

INTERNATIONAL MEDICAL SPENDING DATA

Unique aspects of U.S. medical spending

U.S. medical spending is high relative to spending in other countries. The healthcare sector is a larger share of the U.S. economy than any other economy. Since the U.S. economy is the largest, with one of the highest levels of income per capita, that means that the U.S. also spends the most per capita on medical care.

As U.S. medical spending continues to rise, it is unclear whether the rest of the world will follow. U.S. medical spending could continue to be idiosyncratic. The U.S. has the highest spending, so as long as the U.S. spending growth rate equals or exceeds that of other developed countries, the U.S. will always have the highest level of medical spending. The U.S.'s unusually large private healthcare sector could generate a highly volatile cost curve. If the U.S. truly is idiosyncratic, then the U.S. will have limited success in applying foreign solutions for controlling medical spending to the U.S. economy.

One difference between U.S. and foreign healthcare systems that could inform forecasts of cost control is the lack of universal health insurance. Other developed countries have universal or near universal healthcare. Studying the effect of universal coverage on healthcare cost curves in other countries could help us understand the effect of health insurance expansion in the U.S. Identification of the effect of the level of insurance on medical costs would require variation over time in the level of healthcare coverage in several countries. My data does not include such variation (see Table 1). Even in countries that changed to universal insurance relatively recently, other contemporaneous health policy changes and general issues of low numbers of observations make it almost impossible to observe how changes in coverage affect the growth rates for medical spending.

Country	Year of implementation	Notes
Austria	1965 (92% insurance)	Big jump in reported percentage in 1965, numbers continued to rise to 100%
Canada	Entire sample	
Finland	1964	One time jump to 100% in 1964
Iceland	Entire sample	
Ireland	1980	One time jump to 100% in 1980
Japan	Entire sample	
Norway	Entire sample	
Spain	1987 (97% insurance)	Spotty reporting; 1987 is the first year with near universal coverage reported
Switzerland	Continuous	Rises from 74% to 99% by 1987
U.K.	Entire sample	
U.S.	N/A	Flat at 84-85% from 1997-2008
Min	Pre-1960	
Max	N/A	

Table 1: Introduction of national healthcare system

One problem that the U.S. shares with other developed countries is that medical spending seems to be growing at a high rate. The notion of bending the cost curve implies a smooth function that generates future spending as a multiple of current spending (Orszag 2009). The terminology of medical trend is similar, implying a given rate of growth that actuaries must factor into the calculation of long-term liabilities that will continue until we reach a resistance point (Getzen 2007). It is difficult to find an example of a country that has hit the resistance point for healthcare as a share of the economy. Any decline in medical spending growth in the U.S. and other countries seems to be temporary or is associated with outside factors, such as economic contraction, rather than through efficiencies or cost cutting measures.

OECD data

The Organization for Economic Cooperation and Development (OECD) provides data on medical spending, demographics, and population health variables starting in 1960[1] (OECD 2010). I focus on 11 countries that have reported annual spending data since 1960. The countries are Austria, Canada, Finland, Iceland, Ireland, Japan, Norway, Spain, Switzerland, the United Kingdom, and the United States.

The OECD breaks down medical spending data by currency and by source. The spending data is available in the national currency unit of each country, on a dollar basis using annual exchange rates, on a dollar basis using the exchange rate in 2000, and on a purchasing power parity basis. The OECD reports data split out by funding source, including a breakdown of public and private medical spending. The OECD also provides spending per capita and as a share of GDP.

The main drawback of the OECD data is the variability of reporting across countries and change in reporting standards over time. The OECD takes the data as given by member countries, and then reports it without an extensive set of edits and checks. In addition, the OECD does not standardize what constitutes medical spending, nor does it promulgate mandatory reporting standards for the way different countries choose to break down the data (Ward 2005).

I difference out any time invariant differences between different countries' reporting standards in my rate of change calculations. For example, say that observed medical spending $\hat{S}$ is overestimated by 3% at time 0 and time 1 compared to the true level of spending S. In this case, the observed medical spending growth rate $\hat{\tau}_1$ is equal to the true growth rate τ_1 because the error is time invariant. See Equation 1 below.

$$\hat{S}_0 = 1.03S_0$$
$$\hat{S}_1 = 1.03S_1$$
$$\hat{\tau}_1 = \ln\frac{1.03S_1}{1.03S_0} = \ln\frac{S_1}{S_0} = \tau_1$$

Equation 1: Time invariant errors in spending

I also assume that errors or differences in reporting are not correlated with the reported rates of change or the distributions of the rate of change. This assumption is more difficult to justify, since it is entirely possible that countries that experience a large run-up in medical spending will change their reporting systems to collect and disseminate data that are more detailed. Given the small number of data points that I have, I do not have the degrees of freedom to make additional modifications to the data.

[1] Not all variables are available for the entire period 1960-2009, not all countries provide data going back to 1960 for each variable, and not every variable is reported annually since 1960—for instance, some are reported quadrennially.

High U.S. medical spending

Per capita spending data shows that over time there is no single trend for many countries. Figure 1 shows the spending of four countries, Iceland, Switzerland, the U.K, and the U.S., in each country's national currency unit. I graphed the spending on a log scale in order to make exponential rates of growth appear linear rather than curved. I chose the four countries as representative of different trends over time. Switzerland has the most even growth in medical spending. The U.K. has the median rate of spending growth. The experience of Iceland is like an S-curve with slower growth followed by faster growth and then slower growth again. The U.S. is just behind the U.K. in compound annual growth in medical spending. The graph does obscure the variation in trend rates that occurs during flat periods of exponential growth because of the scaling by national currency units.

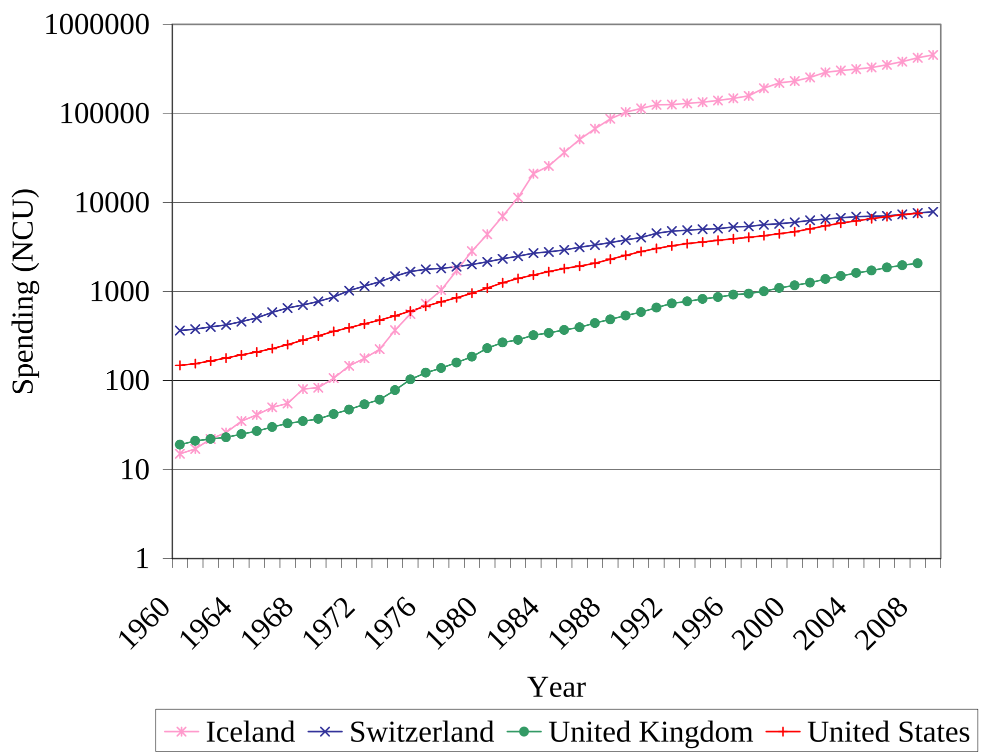


Figure 1: Medical spending in four OECD countries

Absolute spending may be less important for health policy than healthcare's share of the overall economy. Normalizing medical spending by nominal GDP also removes one of the problems with the spending data, which is that the use of national currency units can conflate medical spending changes with other macroeconomic changes (see Figure 2). On the other hand, the percent of GDP measure conflates GDP growth and medical spending growth, and the two variables have a complicated causal relationship (Amiri & Ventelou 2010). Insurance companies write policies based on medical spending rather than spending as a proportion of GDP, so predicting the share of GDP may not help in writing insurance.

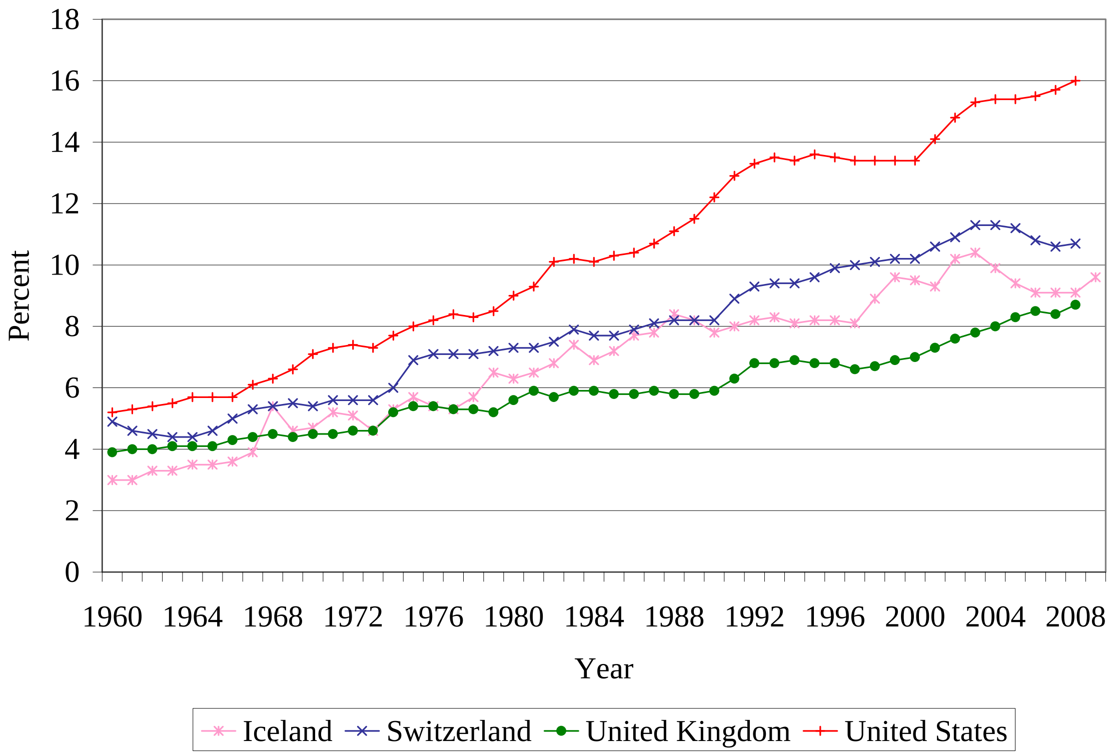


Figure 2: Medical spending as a share of GDP in four OECD countries

It is also unclear from a health policy point of view which variable to target. There is one view that rising medical spending as a share of GDP does not matter as long as GDP is rising. Medical care is a superior good, so we should expect it to rise as a share of rising incomes (Pauly 2003). The level of medical spending may be easier to target, for example with a "global budgeting" system where the amount of spending in a given year is fixed (Long & Marquis 1994). The problem with global budgeting is that unknowable factors, whether political, economic, or demographic, can induce a great deal of year-to-year volatility in spending. The volatility will in turn hamper planning as well as adversely affect insurance lines linked to medical spending. Higher growth rates may be acceptable to insurers if the trade-off is lower volatility in the spending growth rate.

INTERNATIONAL MEDICAL SPENDING TRENDS

Moderate U.S. spending growth

U.S. spending growth rates are average and they display a low volatility. The mean and standard deviation of U.S. spending are equal to or below the median over the 1961-2007 and the 1982-2007 periods. The skew and excess kurtosis of U.S. spending growth are below the median for the 1961-2007 and the 1982-2007 periods. The mean and standard deviation of spending growth for the median country have been more moderate recently than over the entire 1961-2007 period. The skew and excess kurtosis of spending growth for the median country have been more moderate recently than over the entire 1961-2007 period, so medical spending growth has continued to display low volatility, at least up until the beginning of the Great Recession (see Figure 3).

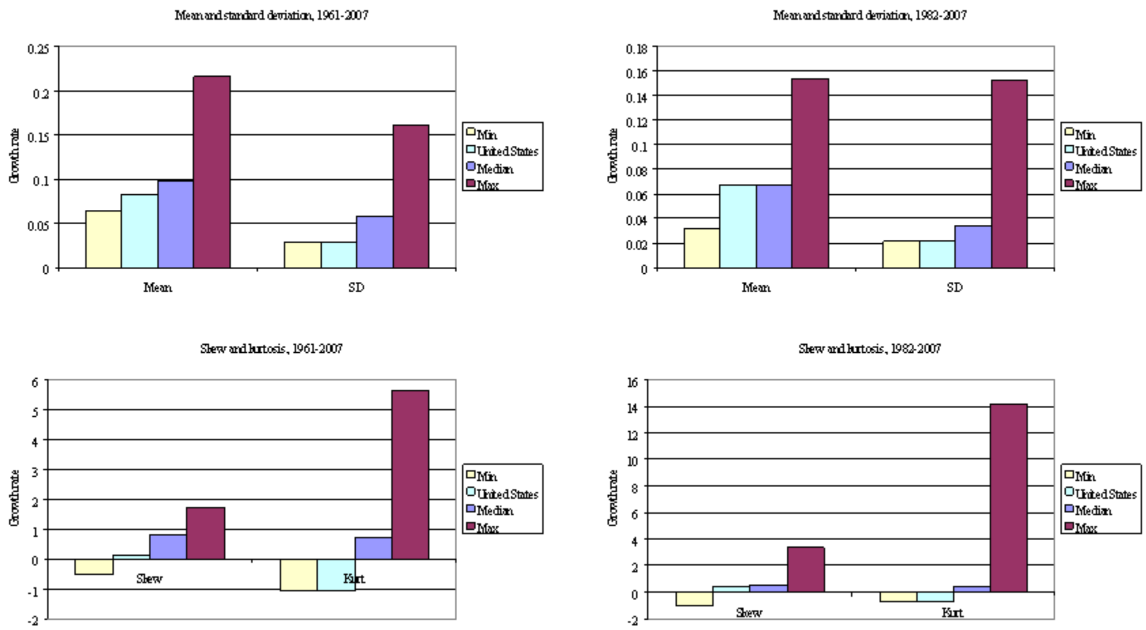


Figure 3: Summary statistics of spending growth

In Table 2, I summarize the mean, standard deviation, skewness, and excess kurtosis of spending growth rates on a continuous, logarithmic basis in all countries in my sample. The spending growth rates in the U.S. are within international norms. The U.S. rate of spending growth is low relative to other countries (tied for 8th overall with Austria and Canada), the standard deviation of the U.S. is the lowest of 11 countries, the skew is ranked 10th out of 11 countries (and lowest overall in absolute value terms), and the excess kurtosis is the lowest out of 11 countries. From an insurer's point of view, the relative risk associated with fluctuations in U.S. medical spending appears to be lower, especially given the relatively thin tails of the distribution.

Country	Mean	SD	Skew	Excess Kurtosis
Austria	0.08	0.06	1.57	5.81
Canada	0.08	0.04	0.24	0.26
Finland	0.10	0.06	-0.45	0.98
Iceland	0.21	0.16	0.74	-0.45
Ireland	0.12	0.07	0.93	0.81
Japan	0.09	0.08	1.25	1.34
Norway	0.10	0.06	0.80	0.93
Spain	0.15	0.09	0.86	0.65
Switzerland	0.06	0.04	0.88	0.18
U.K.	0.10	0.05	1.74	3.66
U.S.	0.08	0.03	0.10	-1.06
Min	0.06	0.03	-0.45	-1.06
Max	0.21	0.16	1.74	5.81

Table 2: Summary statistics of medical spending growth (national currency unit basis)

I adjust spending for a U.S. dollar exchange weighted value to show how much of the U.S.'s low medical spending volatility could be driven by the reserve status of the U.S. dollar (Krugman 1984). The currency adjustment reduces the range of statistics, but not the U.S.'s ranks in spending growth and volatility. The U.S. continues to have the lowest standard deviation and excess kurtosis, as well as one of the lowest average annual rates of growth in spending when dollar adjusted.[2]

One concern I have is that blending public and private spending drives the apparent low variability in the U.S. data. For insurers, private spending is more important than the rate of change in public spending. Public spending can still be important, especially if it causes the changes in private spending. For policymakers, the rate of change in public spending is more important than private spending. Private spending can still be important to policymakers, especially if it leads to political pressure to change the system.

In every system I studied, public growth rates exceed private growth rates.[3] It also appears that public spending is driving a great deal of the volatility in overall spending, or at least that public spending is more variable on several measures. The standard deviation of spending is higher for private than public spending in six out of nine countries, whereas in the U.S. and Finland, it is lower and the statistics are equal for Iceland. The skew and excess kurtosis in public systems is higher in many countries despite public spending being much larger. For the U.S., every statistic is higher for the public portion of spending, suggesting that public spending may be driving much of the volatility in U.S. medical trend (see Table 3).

---

[2] Table available from the author on request.

[3] Canada did not report public/private breakouts of spending until 1971, while Switzerland did not report public/private breakouts of spending until 1986, so I excluded both from the analysis.

	Mean		**SD**		**Skew**		**Excess Kurtosis**	
Country	Public	Private	Public	Private	Public	Private	Public	Private
Austria	0.08	0.07	0.06	0.09	2.82	-1.26	11.38	7.43
Finland	0.11	0.09	0.08	0.06	-0.09	0.74	1.55	0.48
Iceland	0.22	0.20	0.17	0.17	0.59	-0.37	-0.69	2.51
Ireland	0.12	0.12	0.07	0.14	0.51	1.03	-0.16	6.11
Japan	0.09	0.07	0.08	0.14	1.12	0.90	0.84	3.19
Norway	0.11	0.10	0.06	0.34	0.70	2.94	0.38	12.28
Spain	0.15	0.13	0.13	0.15	1.91	0.98	5.89	3.44
U.K.	0.10	0.10	0.06	0.08	1.41	0.39	3.44	-0.62
U.S.	0.10	0.07	0.06	0.03	2.64	-0.05	9.41	-0.67
Min	0.08	0.07	0.06	0.03	-0.09	-1.26	-0.69	-0.67
Max	0.22	0.20	0.17	0.34	2.82	2.94	11.38	12.28

Table 3: Summary statistics of public and private medical spending growth (national currency unit basis)

It is also possible that medical spending volatility is merely a reflection of general volatility in the macroeconomy. One way to look at aggregate economic fluctuations is nominal GDP growth. The U.S. is one of the lowest volatility countries, with a mean growth rate on the lower end and the lowest standard deviation of growth rate. In addition, the skewness is low and the excess kurtosis is strongly negative, surpassed only by Japan. The U.S. could be a low volatility country, in which case there is no additional lesson for reducing the volatility in medical spending through health policy (see Table 4).

Country	**Mean**	**SD**	**Skew**	**Excess Kurtosis**
Austria	0.06	0.03	0.40	0.40
Canada	0.08	0.04	0.58	0.23
Finland	0.08	0.06	-0.34	1.23
Iceland	0.20	0.14	0.65	-0.48
Ireland	0.11	0.06	-0.71	2.55
Japan	0.07	0.06	0.33	-1.04
Norway	0.08	0.04	-0.59	1.43
Spain	0.11	0.05	-0.06	0.73
Switzerland	0.05	0.04	0.64	-0.12
U.K.	0.08	0.04	0.93	2.47
U.S.	0.07	0.02	0.40	-0.53
Min	0.05	0.02	-0.71	-1.04
Max	0.20	0.14	0.93	2.55

Table 4: Summary statistics of nominal GDP growth (national currency unit basis)

Another explanation for the volatility in spending could be demography. Unlike fiscal statistics, demography changes too slowly to explain the volatility in spending growth (White 2007). Most countries have had a slowly aging population as measured by the percent of population above age 65. Japan is the most rapidly aging country. All countries have a small standard deviation of rate of aging. The skews vary much more widely between countries, with the U.S. and Canadian skews closest to zero. There are also high excess kurtoses for a few countries, especially Ireland, reflecting the fact that most years of data contain zero change for the 65 and older population, with a few years of positive (or negative) growth. Measured by excess kurtosis, the U.S. is again a low volatility country from the perspective of aging (see Table 5).

Country	Mean	SD	Skew	Excess Kurtosis
Austria	0.01	0.01	-0.77	2.18
Canada	0.01	0.01	-0.09	-0.48
Finland	0.02	0.01	0.62	0.52
Iceland	0.01	0.01	-0.26	0.32
Ireland	0.00	0.01	3.44	17.80
Japan	0.03	0.01	-0.29	-0.56
Norway	0.01	0.01	-0.37	-1.27
Spain	0.02	0.01	-0.70	-0.19
Switzerland	0.01	0.01	1.50	4.37
U.K.	0.01	0.01	-1.33	4.54
U.S.	0.01	0.01	-0.13	-0.32
Min	0.00	0.01	-1.33	-1.27
Max	0.03	0.01	3.44	17.80

Table 5: Summary statistics of growth in population 65 and over

Modeling spending growth over time

Medical spending growth is hard to predict year-to-year using simple linear regression. For example, the R^2 on a regression of current year spending growth on prior spending growth is only 50%.[4] Regressions using data for a more recent period show less predictability, with an R^2 below 40%.

It is also unclear how far back we should be looking in trying to model medical spending growth. There are so few data points at our disposal, so it seems that more data is better. On the other hand, the U.S. healthcare system today is vastly different from the healthcare system of 50 years ago, so it is not clear whether we should use older data at all. Time series econometrics can help, but ultimately the decision of which data and model to use is a judgment call.

One problem with the time horizon I chose is the possibility that the time series of medical spending, or even medical spending growth, is nonstationary. Even accounting for a linear time trend, it would be difficult to believe that the average medical spending, and trends in spending, would not change over time because healthcare has changed so radically in all countries. In the U.S., the last 50 years have coincided with major medical innovation, the rise of private health insurance, the rise of public health insurance, and major demographic changes (Folland, Goodman, & Stano

[4] Calculated using U.S. data over the past 50 years.

2010). Similar developments have occurred in other developed countries, some of which have also experienced other major macroeconomic disruptions that may have affected medical spending.[5]

I test the stationarity of spending $\hat{S}$, spending growth $\hat{\tau}$, and one difference in spending growth $\Delta\hat{\tau} = \hat{\tau}_t - \hat{\tau}_{t-1}$. By inspection, it seems that no country has had a stationary time series in spending over the past 50 years (see Figure 1; other countries available by request). The plot of spending growth rates is tighter, although there are some outlier countries such as Iceland (see Figure 4). The plot of once differenced spending growth appears stationary for all countries, although it is more volatile for Iceland and the U.K. than Switzerland and the U.S. This suggests that, even accounting for nonstationarity, some countries have higher volatility in medical spending than others (see Figure 5).

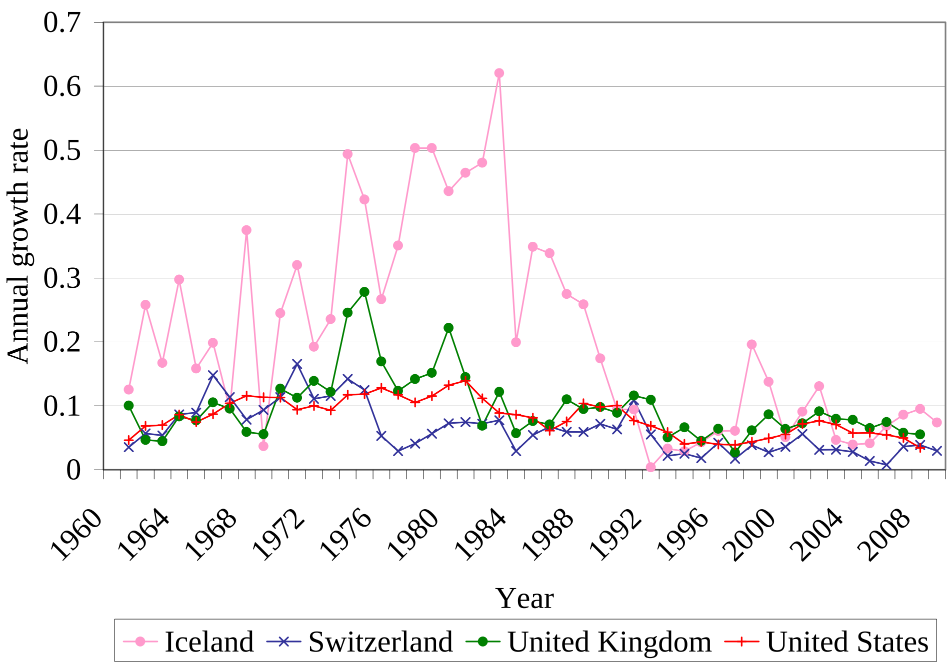


Figure 4: Medical spending growth in four OECD countries

[5] A consideration of financial crises is beyond the scope of this paper, but one example would be the IMF bailout of the U.K. in the 1970s.

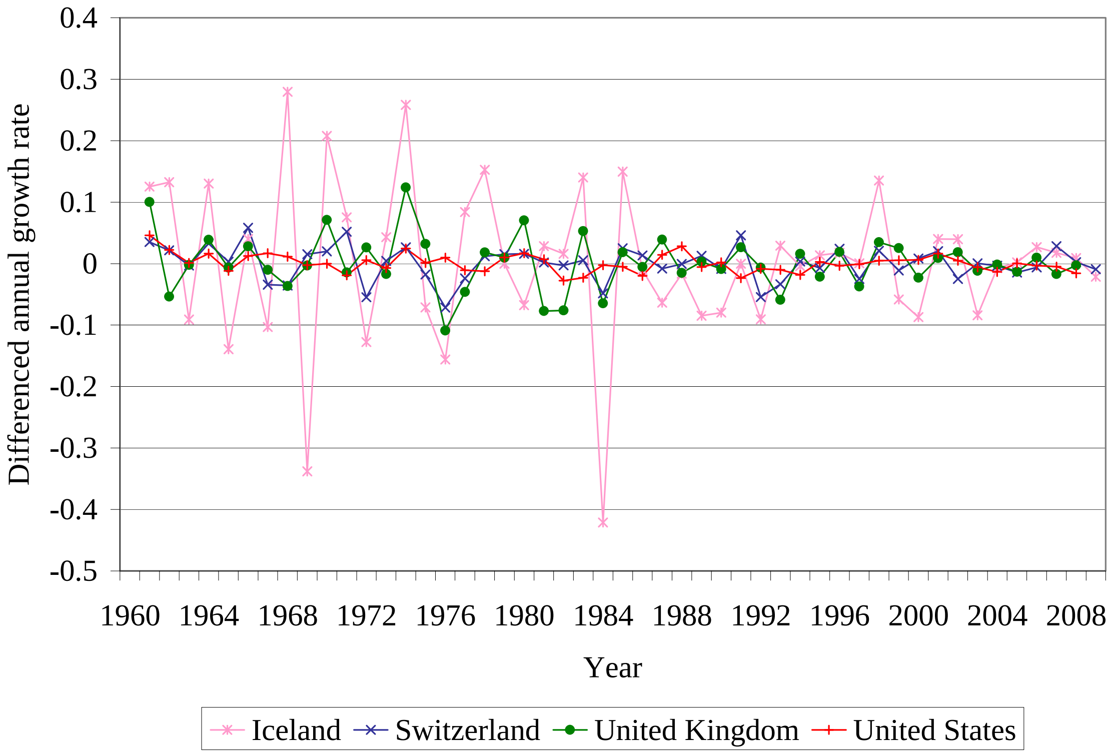


Figure 5: Once differenced medical spending growth in four OECD countries

A Dickey-Fuller test for unit roots suggests that five of the countries have stationary spending growth rates, while the other six have a unit root in spending growth. Among those countries where I fail to reject the null of a unit root is the U.S (see Table 6). Therefore, the observed high correlation between last year and current year spending growth may be an artifact of the nonstationarity of the time series. Japan is the only country with a stationary spending growth time series and high autoregressive element in spending growth. Japan gets close to the ideal of a country where past data is useful for forecasting future spending growth, and the correlation year-to-year is high.

Country	Spending	Spending growth	Δ Spending growth
Austria	1.00	<0.01	<0.01
Canada	1.00	0.30	<0.01
Finland	1.00	0.21	<0.01
Iceland	1.00	0.05	<0.01
Ireland	1.00	<0.01	<0.01
Japan	0.98	<0.01	<0.01
Norway	1.00	<0.01	<0.01
Spain	1.00	<0.01	<0.01
Switzerland	1.00	0.10	<0.01
U.K.	1.00	0.04	<0.01
U.S.	1.00	0.57	<0.01

Table 6: Dickey-Fuller test for unit root

Summary statistics for medical spending point to overall low relative volatility in the U.S. Summary statistics do not account for structural breaks, macroeconomic changes in other variables, demography, and the organization and financing of healthcare. The next step is to account for these changes within the framework of a time series regression. I decided to fit an autoregressive (AR) model because of the high degree of persistence in U.S. spending growth year-on-year. I chose an AR(1) model because of the small amount of data I have.

The main problem with fitting an AR(1) model to all the countries' data is the nonstationarity I observed. The time series for the level of spending in every country likely has a unit root. The graphs of spending for almost every country contain at least two trend lines if not more; to the extent that they do not, it is because the scaling of the graph obscures so much of the variation. The AR(1) parameter will be close to 1 not because spending is highly autoregressive and predictable but because the time series is nonstationary. Nonstationarity biases upward the statistical tests for the strength of correlations over time, such as t-tests. The result is that not only is the time series for spending misestimated, but any explanatory variables added to the AR(1) models are likely to show power to explain the growth in spending whether or not the relationship truly exists (Ferson, Sarkissian, & Simin 2003).

One way to deal with the problem of nonstationarity is through differencing. I fit the medical spending growth series directly for Austria, Ireland, Japan, Norway, and Spain, and I fit the once differenced spending growth rate for Canada, Finland, Iceland, Switzerland, the U.K., and the U.S.[6] The AR(1) parameter is 0.42 for Austria, 0.56 for Ireland, 0.72 for Japan, 0.30 for Norway, and 0.47 for Spain. In addition to the relatively low AR(1) coefficient, the p-value for Norway is 0.05, leading me to the conclusion that the summary statistics may be as good as, or better than, the AR(1) model when forecasting the growth rate in medical spending in Norway. More complicated time series

[6] Full AR(1) results are available upon request.

models are not appropriate in all cases.

Of all these countries, Japan is the one where forecasts of future growth using current and prior growth rates are the most informative. Planning for future medical spending budgets, and writing insurance based on medical claims, could be easier in Japan than in Austria, Ireland, Norway, and Spain, where future spending growth rates are less persistent.

The AR(1) results for spending growth and differenced spending growth series for countries with a unit root in spending growth show the importance of using the correct time series model. The AR(1) parameter for a regression of spending growth is 0.82 for Canada, 0.79 for Finland, 0.69 for Iceland, 0.74 for Switzerland, 0.66 for the U.K., and 0.91 for the U.S. The AR(1) parameter for once differenced medical spending growth is 0.14 for Canada, 0.14 for Finland, -0.48 for Iceland, -0.13 for Switzerland, -0.15 for the U.K., and 0.17 for the U.S. The p-value is high for Canada, Finland, Switzerland, and the U.K. so the AR(1) model may be inferior to simple summary statistics. The coefficient is negative for Iceland, Switzerland, and the U.K., so the time series of interest (medical spending growth) may return to a given level over time. The U.S. coefficient is small but significant, so the model is valid but may not help with long-term spending growth forecasts.

IMPLICATIONS FOR LONG-TAILED MEDICAL INSURANCE

Examples of long-tailed lines with medical exposure

Guaranteed renewable health insurance is one long-tailed insurance line specifically based on the risk of medical spending growth. Guaranteed renewable insurance includes a given medical trend factor for the life of the policy. Forecast errors that lead to higher than expected medical spending must come out of reserves, and can expose an insurer to ruin if reserves are not high enough. The guarantee of class average underwriting can be for 10 years or more at the time that the insurer forms the pool of insured lives, so a single year of forecast error can be very costly if it occurs in the early years of the contract (Lieberthal ND).

Workers compensation is a long-tailed line of property and casualty insurance that is exposed to the risk of medical spending growth. In the case of workers compensation insurance, the insurer must pay for the medical care arising from on-the-job injuries, potentially for a long period. As the standard of care changes, and becomes more expensive, the workers compensation insurer may have to provide benefits that meet the current standard of medical care, even if it is much better, and much more expensive, than care that was available at the time the policy was written, an effect called "social inflation" (Feldblum 1993).

Medical spending growth can have an even greater effect on excess casualty reinsurance. The reason is the "leveraged effect of limits on severity trend" (Werner and Modlin 2010, pp.117-118). Leverage comes from the fact that, for claims below the limit, the trend on excess claims is unobserved. When losses are high enough to trigger excess claims, losses jump from zero to a positive number, and the trend is undefined. Then, excess trend can either be above trend on total losses, or dampened below the trend on total losses, depending on what portion of the risk the reinsurer takes and whether there is an upper limit to the exposure (Keatinge 1989).

Example of a 10-year tail of claims

Take as an example an insurer who receives $12,000 in premiums up front for an insurance liability that is worth $10,005 in present value terms. The insurer expects to pay out the liability over 10 years (see Table 7). The insurer expects to pay out $863 at the end of year 1, rising an expected 7% per year on an annual compound basis in future years due to medical spending growth. The insurer calculated the expected claims in year 1 by observing that claims for a similar exposure were $807 in the prior year, and then inflating prior experience by the 7% forecast trend. The insurer uses a 3% discount rate and expects to earn 3% on invested reserves. The gross loading factor is 20%, which is worth $1,995 at time 0 and $2,681 (10 years compounding at 3%) at the end of year 10. The $2,681 is the final expected gross surplus before accounting for costs associated with the policy.

(A)	(B)	(C)	(D)	(E)
Year	Nominal claims	Discounted claims	Reserves	Surplus
0	0	0	10005	1995
1	863	838	9442	2055
2	923	870	8802	2117
3	988	904	8078	2181
4	1057	939	7263	2246
5	1131	976	6350	2313
6	1210	1013	5331	2382
7	1295	1053	4196	2453
8	1386	1094	2936	2527
9	1483	1137	1541	2603
10	1587	1181	0	2681
Totals	11923	10005		

Table 7: Base case insurance

Next, I assume that the insurer forecast year 1 incorrectly but the forecast claims in dollar terms for every other year remains the same. The trend for year 1 is 9%, which is 7% average U.S. trend plus 2%, one standard deviation of U.S. trend in the summary statistics. Instead of $863, the claims for year 1 are now $880, $17 higher. The claims for year 2 are still $923, and the claims for every other year after year 2 do not change (see Table 8).

In this case, the trend rates year-by-year are 9% in year 1, 5% in year 2, and then 7% thereafter. The trend is mean-reverting, and the only loss is because of the single year's higher payment and absence of investment returns on the deviation from year 1 experience. The adverse experience hardly makes a dent in the final gross surplus of the contract, which is $2,659, only $22 less than originally expected. The equivalent is the $17 deviation from expectations compounded nine years at 3%, which is $22. If the insurer had known ahead of time what the future would be, it would see that at the beginning of the contract the gross load was $1,979 rather than $1,995.

(A)	(B)	(C)	(D)	(E)
Year	Nominal claims	Discounted claims	Reserves	Surplus
0	0	0	10021	1979
1	880	854	9442	2038
2	923	870	8802	2099
3	988	904	8078	2162
4	1057	939	7263	2227
5	1131	976	6350	2294
6	1210	1013	5331	2363
7	1295	1053	4196	2434
8	1386	1094	2936	2507
9	1483	1137	1541	2582
10	1587	1181	0	2659
Totals	11940	10021		

Table 8: One bad year of 9% trend followed by base case claims thereafter

Next, I assume that changes in the level of spending are permanent. The year-by-year rates of medical trend are 9% in year 1, and then 7% in year 2 and thereafter. At the end of year 10, the gross surplus is only $2,400, which means that had the insurer known the future at time 0, it would have seen that the gross load was only 17% rather than 20%, and that a single-year deviation from experience knocked three percentage points off the anticipated loading factor. The true gross surplus at the beginning of the contract is $1,785 rather than $1,995 (see Table 9).

(A) Year	(B) Nominal claims	(C) Discounted claims	(D) Reserves	(E) Surplus
0	0	0	10215	1785
1	880	854	9641	1839
2	942	888	8988	1894
3	1008	922	8250	1951
4	1079	959	7419	2010
5	1155	996	6487	2070
6	1236	1035	5446	2132
7	1323	1076	4286	2196
8	1416	1118	2999	2262
9	1515	1161	1574	2330
10	1621	1206	0	2400
Totals	12175	10215		

Table 9: One bad year of 9% trend followed inflated by 7% trend thereafter

Finally, I assume that the spending growth rate is autoregressive, so that a single year's increase in spending growth rates leads to a permanent change in the level of spending growth. The year-by-year rates of medical trend are 9% in year 1, and then 7.5% in year 2 and thereafter. This could be because of an innovation that was very costly upfront and leads to continuing increases in costs that are above expectations.

At the end of year 10, the gross surplus is only $2,101, which means that had the insurer known the future at time 0, it would have seen that the gross load was only 15% rather than 20%. A single-year deviation from experience knocked five percentage points off the anticipated loading factor because the growth rate settles at a level that is 0.5 percentage points higher. The true gross surplus at the beginning of the contract is $1,564 (see Table 10). The persistence of higher trend takes an additional three percentage points off the anticipated gross loading factor beyond the effect of the single-year claims mistake compounded over the life of the insurance policy. If the time 0 expected gross load, including the reserves and the costs of underwriting, marketing, and capital, had been less than 15%, the insurer might have defaulted on the contract.

(A) Year	(B) Nominal claims	(C) Discounted claims	(D) Reserves	(E) Surplus
0	0	0	10436	1564
1	880	854	9869	1611
2	946	892	9219	1659
3	1017	931	8479	1709
4	1093	971	7640	1760
5	1175	1014	6694	1813
6	1263	1058	5632	1867
7	1358	1104	4443	1923
8	1460	1153	3116	1981
9	1570	1203	1639	2040
10	1688	1256	0	2101
Totals	12450	10436		

Table 10: One bad year of 9% trend followed by autoregressive spending growth of 7.5% trend thereafter

The forecast error problem is much worse if the insurer is writing excess reinsurance. Say that a reinsurer agrees to pay claims in excess of $9,266. This translates into expected payments of $2,657: $1,070 in year 9 and $1,587 in year 10. In present value terms, the reinsurer expects to pay out $2,001, and the reinsurer accepts $2,400 as the premium, which represents a 20% gross load.

Leverage comes in to the base case as year 10 nominal payments are 48% higher than year 9 payments. In the one bad year scenario, excess claims start in year 9, and are $13 more than expected on a present value basis, and the gross load falls to 19%. Year 10 nominal payments are only 46% higher than year 9 payments because year 10 payments cannot exceed $1,587 in year 10. In the one bad year followed by 7% trend example, excess claims start in year 9, and are $192 more than expected on a present value basis. Leverage causes the gross load to fall to 9%. Year 10 nominal payments are only 26% higher than year 9 payments because year 10 payments cannot exceed $1,621 in year 10.

The autoregressive trend example ends with a small deficit. In the autoregressive trend example, payments start in year 9 and are $403 more than expected on a present value basis. Year 10 nominal payments are 13% higher than year 9 payments. The expected 20% gross load with a 7% trend at the beginning of the contract was only $399, so the unexpected claims cause the contract to end with a 0% final gross load ($3 deficit). I summarize the full insurance case and present summarized results for the reinsurance case in Table 11.

Scenario	Naive premium	Total discounted claims	Final gross surplus	Time 0 gross surplus	Time 0 gross load
Full insurance					
Base case insurance	12000	10005	2681	1995	20%
One bad year of 9% trend followed by base case claims thereafter	12000	10021	2659	1979	20%
One bad year of 9% trend followed inflated by 7% trend thereafter	12000	10215	2400	1785	17%
One bad year of 9% trend followed by autoregressive spending growth of 7.5% trend thereafter	12000	10436	2101	1564	15%
20% excess reinsurance					
Base case insurance	2400	2001	536	399	20%
One bad year of 9% trend followed by base case claims thereafter	2400	2014	519	386	19%
One bad year of 9% trend followed inflated by 7% trend thereafter	2400	2193	278	207	9%
One bad year of 9% trend followed by autoregressive spending growth of 7.5% trend thereafter	2400	2403	-3	-3	0%

Table 11: Summary of full insurance and 20% reinsurance cases

Solutions to forecast errors in medical spending

Prediction markets have the potential to help insurance companies with long-tailed lines linked to medical spending. Currently, Intrade.com has contracts covering macroeconomic indicators including the unemployment rate, the U.S. federal budget deficit, and health reform indicators including such market prediction polls as "Individual mandate to be ruled unconstitutional by U.S. Supreme Court." There are three contracts on the individual mandate with different expiration dates: October 2011, December 2012, and December 2013 (Intrade.com 2011).

Prediction market contracts might be more appropriate for the less severe examples I have given of forecast errors. For a single year of adverse experience, the insurer could make a hedging bet each year and could factor the annual costs into the single premium through the reserve calculation. The main obstacles would be deciding how contract outcomes map onto forecast errors, and making sure that the prediction market is thick enough to allow for as much hedging as the insurer needs.

Dealing with the situation where spending is autoregressive would be much more difficult with prediction markets. The insurer would have to make a much larger bet on medical spending rising in the earlier rather than the later years of the contract in order to deal with compound investment losses. Dealing with the situation where spending growth is autoregressive is even more difficult using short-term prediction contracts. It would take an extreme amount of financial engineering to figure out how to use the single year bets to hedge multiple years of higher trend.

Asset markets could be a more appropriate way to handle the trend risk. Financial products that would be helpful include insurance-indexed futures and macro markets based on trading national income shares (Cox & Schwebach 1992; Shiller 1993). Indexed futures on health insurance would give any insurer with an exposure to rising medical claims the kind of hedge that they need. Unfortunately, health insurance futures have foundered along with most private futures markets for macroeconomic variables. Macro markets have also not succeeded widely yet,[7] but given the need for improved forecasting, there is reason to try again. It may be that the failure of markets to insure macroeconomic risks is due to market failure.

Inflation hedging and forecasting took a major leap forward in many countries after the introduction of inflation-indexed bonds called TIPS (Treasury Inflation-Protected Securities) in the U.S. (Chen & Terrien 2001). One proposal is for the government to reduce the basis risk between overall inflation and medical inflation by issuing TIPS indexed to specific portions of inflation, in this case medical care (Jennings 2006). Insurer's claims are generally linked to overall medical spending, not prices; insurers could not hedge increases in claims due to changes in the quantity with a TIPS bond. That said, TIPS give a market-based forecast of inflation that improves upon expert opinion alone, because markets can often aggregate information for forecasting better than individuals do.

If it is the case that only the government can backstop medical spending growth, it is not necessarily true that TIPS is the only or best way for the government to be involved. One possibility is that the government should focus on health policies and other interventions that would tame medical spending growth. The OECD data suggests that that effort has not been successful in the U.S. or other countries. The federal government could also provide reinsurance, say on health insurance claims for any individual that exceed $36,000 in a year (Antos, King, Muse, Wildsmith, & Xanthopoulos 2004). The reinsurance policy would not help other non-health insurers with medical claims, and might actually harm them if government reinsurance fuels greater medical spending growth.

Acknowledging unsolved problems

In 2009, negative GDP growth reduced the growth of medical spending to one of the lowest rates ever, but the percent of the economy devoted to medical care still grew (Martin, Lassman, Whittle, Catlin, & The National Health Expenditure Accounts Team 2011). In other words, even

[7] The Case-Shiller indices for real estate are an exception.

though the economy shrank, medical care continued to increase its share of the pie, which goes against the idea that medical care was increasing its share of the pie because it is a superior good. While one year does not prove or disprove any model, it shows the difficulty of forecasting spending growth based on so little data.

The good news is that over the long term, the growth rates in U.S. medical spending are not outrageous. Table 12 shows that medical spending growth is generally explainable by the combination of economic growth and population aging. The combination does not always add up perfectly. In addition, the evidence on the nonstationarity of many countries' medical spending growth indicates that differenced medical spending growth, rather than spending growth itself, may be the correct dependent variable. The point is that the cost curve is not an outrageous fact of the economy to be explained, but rather part of the macroeconomy of developed countries and the organization of their healthcare systems.

Country	Growth in medical spending per capita (%) (A)	Growth in GDP per capita (%) (B)	Growth in population 65 and older (%) (C)	Annual excess growth (%) (A–B–C)
Austria	0.08	0.06	0.01	0.01
Canada	0.08	0.08	0.01	-0.01
Finland	0.10	0.08	0.02	0.00
Iceland	0.21	0.20	0.01	0.00
Ireland	0.12	0.11	0.00	0.01
Japan	0.09	0.07	0.03	-0.01
Norway	0.10	0.08	0.01	0.01
Spain	0.15	0.11	0.02	0.02
Switzerland	0.06	0.05	0.01	0.00
U.K.	0.10	0.08	0.01	0.01
U.S.	0.08	0.07	0.01	0.00

Table 12: Components of medical spending growth[8]

The problem for newly created and existing insurers and reinsurers under the latest change to the U.S. healthcare system is daunting. The Patient Protection and Affordable Care Act will bring a new form of health insurer, the Accountable Care Organization, into existence. Accountable Care Organizations are integrated delivery systems that share risk with payers, and so will face the risk that unforeseen increases in medical spending will severely deplete or wipe out their capital base. The previous managed care explosion of the 1990s led to some reinsurance of capitated physician practices, but many practices were too small or too specialized to utilize reinsurance arrangements (Simon & Emmons 1997). Hospitals, with large and sophisticated risk management departments, may be more interested in utilizing reinsurance as part of their Accountable Care Organization

[8] Table style adapted from (White, 2007)

efforts, but may also have the market power to drive harder bargains. As a result, reinsurers will face an important opportunity to sell risk management to new healthcare organizations, but will have little prior data to rely on.

In the short term, it is important to keep spending levels, growth, volatility, and difficulties in forecasting future growth rates in mind when writing insurance. We should be upfront about the true duration of the liabilities of the policies we write, not only if things go well but also if they go badly. This may make more risks uninsurable than we previously thought, but it is better not to take on an uninsurable risk than expose policyholders to the possibility of firm ruin. The risks are especially great in the U.S., which will be undertaking massive healthcare system changes in the next three years with unforeseeable effects on the cost curve.

It is harder to determine what actions insurers should take over the medium and long term to manage the rate and volatility of spending growth. No country in my data was able to arrest the rate of growth in spending through the direct application of health policy. The best insurers can do is to accept the cost curve as it is and try to deal with their modeling challenges. It may be that the best that policymakers can work towards is economic growth, which can at least make medical spending affordable. Making changes to health policy does not always reduce spending, and could cause volatility that disrupts many types of insurance markets.

Acknowledgements

I received feedback from a review committee provided by the Casualty Actuarial Society consisting of James Heer, the chair of the CAS Committee on Reinsurance Research, and Avi Adler and Jeff Dollinger. I also benefited from discussions with Indraneel Chakraborty. This paper is an outgrowth of my dissertation work, so I am grateful for input from my advisor, Mark Pauly, and my thesis committee, Scott Harrington, Greg Nini, and Jessica Wachter. I received support for this work through an Agency for Healthcare Research and Quality dissertation research grant (R36 HS18835-01) and The Wharton School.

BIBLIOGRAPHY

[1.] Amiri, A., and B. Ventelou, "Causality Test Between Health Care Expenditure and GDP in U.S.: Comparing Periods," Working Paper for Groupement de Recherche en Economie Quantitative d'Aix-Marseille, July 2010, available at http://halshs.archives-ouvertes.fr/docs/00/52/03/26/PDF/DTGREQAM2010-29.pdf.

[2.] Antos, Joseph, et al. "Analyzing the Kerry and Bush Health Proposals: Estimates of Cost and Impact," American Enterprise Institute, 2004, available at http://www.cahi.org/cahi_contents/resources/pdf/AEIanalyzingkerryandbusyhealthproposals.pdf.

[3.] Chen, P., and M. Terrien, "TIPS as an Asset Class." *The Journal of Investing* 10(2), 2001, pp. 73-81.

[4.] Cox, S.H., and R.G. Schwebach, "Insurance Futures and Hedging Insurance Price Risk," *Journal of Risk and Insurance* 54(4), 1992, pp. 628-644.

[5.] Feldblum, Sholom, "Workers' Compensation Ratemaking," *CAS Exam Study Note, Exam 5,* Arlington, VA: Casualty Actuarial Society, 1993, available at http://www.casact.org/library/studynotes/feldblum5.pdf.

[6.] Ferson, W.E., S. Sarkissian, and T.T. Simin, "Spurious Regressions in Financial Economics?" *Journal of Finance* 58(4), 2003, pp. 1393-1414.

[7.] Folland, S., A.C. Goodman, and M. Stano, *The Economics of Health and Health Care,* 6th ed., Boston: Prentice Hal, 2010.

[8.] Getzen, T.E., "Modeling Long Term Healthcare Cost Trends," 2007, project available from the Society of Actuaries at http://www.soa.org/files/pdf/research-getzen-health-cost.pdf.

[9.] Intrade.com, "Individual Mandate to be Ruled Unconstitutional by U.S. Supreme Court," Legal–U.S. Healthcare Reform Contracts, 2011, accessed February 11, 2011, at http://www.intrade.com.

[10.] Jennings, W.W. "Disaggregated TIPS: The Case for Disaggregating Inflation-Linked Bonds into Bonds Linked to Narrower CPI Components Slicing and Dicing TIPS," *Journal of Pension Economics and Finance* 5(3), 2006, pp. 325-343.

[11.] Keatinge, Clive L., "The Effect of Trend on Excess of Loss Coverage," *Proceedings of the Casualty Actuarial Society* LXXVI, 1989, pp. 126-46, available at http://www.casact.org/pubs/proceed/proceed89/89126.pdf.

[12.] Krugman, P., "The International Role of the Dollar: Theory and Prospect," In *Exchange Rate Theory and Practice,* J.F.O. Bilson and R.C. Marston (Eds.), Chicago: The University of Chicago Press, 1984, pp. 261-278.

[13.] Lieberthal, R.D., "The Impact of Medical Spending Growth on Guaranteed Renewable Health Insurance," unpublished dissertation.

[14.] Long, S.H., and M.S. Marquis, "Toward a Global Budget for the U.S. Health System: Implementation Issues and Information Needs," RAND Corporation, 1994, available at http://www.rand.org/pubs/issue_papers/IP143/index2.html.

[15.] Martin, A., D. Lassman, L. Whittle, A. Catlin, and The National Health Expenditure Accounts Team, "Recession Contributes to Slowest Annual Rate of Increase in Health Spending in Five Decades," *Health Affairs* 30(1), 2011, pp. 11-22.

[16.] OECD, "OECD Health Data 2010: Statistics and Indicators," 2010.

[17.] Orszag, P., "IMAC, UBend," Blog entry posted on the Office of Management and Budget Web Site, July 17 2009, available at http://www.whitehouse.gov/omb/blog/09/07/17/IMACUBend/.

[18.] Pauly, M.V., "Should We be Worried about High Real Medical Spending Growth in the United States?" *Health Affairs,* 2003, Supplemental Web Exclusives, W3-15-27.

[19.] Shiller, R.J., *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks,* Oxford, U.K: Clarendon Press, 1993.

[20.] Simon, C.J., and D.W. Emmons, "Physician Earnings at Risk: An Examination of Capitated Contracts," *Health Affairs* 16(3), 1997, p. 120.

[21.] Ward, D., *OECD Short-Term Economic Statistics Expert Group (STESEG): Data and Metadata Reporting and Presentation Handbook,* 2005.

[22.] Werner, Geoff, and Claudine Modlin, "Losses and LAE," Chapter 6 in *Basic Ratemaking,* 4th ed., Arlington, VA: Casualty Actuarial Society, 2010.

[23.] White, C., "Health Care Spending Growth: How Different is the United States from the rest of the OECD?" *Health Affairs* 26(1), 2007, pp. 154-161.

Biography of the Author

Robert D. Lieberthal, PhD, is Assistant Professor of Health Policy at the Jefferson School of Population Health at Thomas Jefferson University in Philadelphia. He can be reached at robert.lieberthal@jefferson.edu.

Index Clause for Aggregate Deductibles and Limits in Non-Proportional Reinsurance

Ka Chun Yeung, FCAS, FIAA, CCRA

Abstract
Index clauses currently in place in the market do not specify how Annual Aggregate Deductibles (AAD) and Annual Aggregate Limits (AAL) should be indexed, which result in inconsistency when indexed deductibles and limits are in place.
In this paper, concepts of indexed deductible and limit will be revisited for developing indexing methods for AAD and AAL. Formal mathematical proofs and numerical examples will be presented. The introduced AAL indexing methods enable determination of paid reinstatement premium when index clause for per-claim deductible and limit is in place.
The choice of appropriate method for indexing AAD and AAL shall take into account both theoretical and practical soundness.

Keywords. Reinsurance; Excess (Non-Proportional); Deductibles, Retentions and Limits.

1. INTRODUCTION

Index clause (or "Stability clause") have become a standard clause in non-proportional (NP) reinsurance contracts for long-tail classes in many international markets. Index clause handles the leveraged effect of inflation on excess layer loss cost by adjusting the per-claim deductible and limit so that effect of inflation is shared between the primary insurer and NP reinsurer equally.

However, most index clauses do not specify how Annual Aggregate Deductibles (AAD) and Annual Aggregate Limits (AAL) should be adjusted for inflation. In practice many NP reinsurance contracts try to mitigate this inherent problem by

(a) Specifying unlimited reinstatements or an AAL that is much greater than the per-claim limit, or

(b) Simply endorsing that the AAL is "un-indexed", although per-claim deductible and limit are still subject to index clause adjustment

While AAD's are becoming more common for long-tail NP reinsurance (for various purposes, e.g., reinsurance premium saving or fulfilling sufficient risk transfer), the above mitigating measures do not provide good solutions for AAD's.

1.1 Research Context

The method for indexing a per-claim deductible and limit has been explained in Ferguson [1]. It has become a standard calculation method specified in the index clause of long-tail excess of loss reinsurance contracts in many markets. Implementation of index clause, wording, and pricing has been discussed in that paper as well.

Feldblum [6] and Feldblum [8] suggested a different method for indexing per-claim deductibles and limits, by making use of internal rate of return concept. However this method has not been widely adopted. Further, calculation procedures with this method can be complicated in a varying inflation environment.

1.2 Objective

This paper will propose two methods for indexing AAD's and AAL's, both based on achieving the goals of "equitable share of inflation effect" and "equitable share of deflated payments and actual payments" between primary insurer and reinsurer. The current per-claim deductibles and limit indexation method will be briefly revisited. The proposed methods for indexing AAD's and AAL's will be developed in a manner consistent with the per-claim indexing approach.

1.3 Outline

The remainder of the paper proceeds as follows. In section 2 the concept of index clause currently in place in the market will be revisited. In section 3 two methods for indexing AAD's and AAL's will be introduced, first through intuitive arguments from a retrocessionaire's point of view, then numerical examples and formal mathematical proofs will be presented. In section 0 practical issues will be discussed, including implementing an AAD and AAL index clause, pricing approaches, and calculating paid-reinstatement premium.

2. INDEX CLAUSE – REVISITING THE CONCEPTS

The following is an example of index clause wording that explains how per-claim deductibles and per-claim limits are indexed:

Each loss payment shall be brought back separately to its respective value at the base date according to the indices prevailing on the date the loss payments are made, by means of the following formula:

$$\frac{\textit{actual amount of payment} \times \textit{index at base date}}{\textit{index at date of payment}} = \textit{adjusted payment value}$$

All actual payments and adjusted payment values shall then be separately totaled and deductible and limit shall be multiplied by the following fraction:

$$\frac{\textit{Total of actual payments}}{\textit{Total of adjusted payment values}}$$

in order to determine the overall indexed deductible and, where applicable, limit of indemnity, and thus the amount recoverable in accordance with the provisions of this clause.

2.1 Index Clause in Practice – An Example

An excess of loss reinsurance program has a per-claim deductible of $3 million and per-claim limit of $5 million, both subject to an index clause. Let time 0 denote the base date for index clause calculation. Two incremental payments have been made for a claim at time 1 and 2 (in years), as shown in the following table.

Incremental Actual Payment ($000s)

payment time	0	1	2	row sum
claim 1	$0.0	$3,180.0	$1,308.0	$4,488.0

Next, adjusted payments are calculated:

Incremental Adjusted Payment ($000s)

payment time	0	1	2	row sum
claim 1	$0.0	$3,000.0	$1,200.0	$4,200.0
Index	100	106	109	

For example, adjusted payments for time 1 = $3.18 million × 100/106 = $3 million, which can be interpreted as the deflated value at time 0 of an actual payment $3.18 million paid at time 1 according to the specified index. Hereafter in this paper, the author will use the term “deflated value” or “deflated payment” which take the same technical calculation steps as “adjusted payment” presented above, but the author believes the term “deflated” better represents inflation measurement and sharing concepts underlying index clause calculations.

In the third step, the indexed deductible and indexed limit are calculated.

Indexed deductible = $3 million × 4,488/4,200 = $3.206 million.

Indexed limit = $5 million × 4,488/4,200 = $5.343 million.

In the final step, the amount to be paid by the NP reinsurer to the primary insurer for claim 1 = ($4.488 – $3.206) million = $1.282 million.

In practice, at time 2 when the primary insurer notifies claim 1 to the NP reinsurer, the indexed deductible and limit will be calculated immediately, whether the claim is fully settled at time 2 or not. The NP reinsurer then needs to make a payment to the primary insurer if the total of all actual payments exceeds the indexed deductible calculated at time 2.

At time 3 another payment is made:

Incremental Actual Payment ($000s)

payment time	0	1	2	3	row sum
claim 1	$0.0	$3,180.0	$1,308.0	$2,808.0	$7,296.0

Similarly, adjusted payments are calculated:

Incremental Adjusted Payment ($000s)

payment time	0	1	2	3	row sum
claim 1	$0.0	$3,000.0	$1,200.0	$2,400.0	$6,600.0
Index	100	106	109	117	

Indexed deductible = $3 million × 7,296/6,600 = $3.316 million.

Indexed limit = $5 million × 7,296/6,600 = $5.527 million.

The cumulative amount to be paid by NP reinsurer to primary insurer for claim 1 = ($7.296 – $3.316) million = $3.980 million. Therefore NP reinsurer pays ($3.980 – $1.282) million = $2.698 million at time 3 to primary insurer.

To check whether effect of inflation is shared between the primary insurer and NP reinsurer equally, first consider inflation for the gross claim:

$$\frac{\text{Total of all actual payments}}{\text{Total of all deflated payments}} \tag{2.1}$$

Inflation of gross claim = 7.296M / 6.600M – 1 = 10.5%

Similarly, consider inflation of NP reinsurer's excess of loss payments:

$$\frac{\text{Total of all actual payments} - \text{indexed deductible}}{\text{Total of all deflated payments} - \text{unindexed deductible}} \tag{2.2}$$

Finally, inflation of primary insurer's retained claim:

$$\frac{\text{indexed deductible}}{\text{unindexed deductible}} \tag{2.3}$$

Inflation of the NP reinsurer's payment = 3.980M / 3.600M – 1 = 10.5%

Inflation of the primary insurer's retained claim = 3.316M / 3.000M – 1 = 10.5%

2.2 Generalizing the Principles of Index Clause

In the numerical example in section 2.1, inflation for the gross claim, inflation for the NP reinsurer's payment, and inflation for the primary insurer's retained claim are the same. It can be verified that the three inflation measurements in equations (2.1), (2.2), and (2.3) are equal in general.

Notations:

i = claim identifier, $i = 1, 2, 3, \ldots$

t = time of payment, $t = 0, 1, 2, 3, \ldots$ base date is denoted by $t = 0$

v_t = deflating factor for payment made at time t

= value of index clause index at time 0 ÷ value of index clause index at time t

$X_{i,t}$ = Actual dollar payment of i^{th} claim at time t

$X_i^T = \sum_{t=1}^{T} X_{i,t} \cdot v_t$ = Total of all deflated payments made between time 0 and time T

d = un-indexed deductible per claim

l = un-indexed limit per claim

2.2.1 Indexed Deductible and Indexed Limit

According to equation (2.1), given incremental payment information up to time T for the i^{th} claim, indexed deductible is calculated as:

$$d'_{i,T} = d \times \frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} \tag{2.4}$$

Similarly, indexed limit for the i^{th} claim is calculated as:

$$l'_{i,T} = l \times \frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} \tag{2.5}$$

2.2.2 Cumulative Payments Paid by NP Reinsurer and Incremental Payments Paid by Primary Insurer

Cumulative payments paid by the NP reinsurer to the primary insurer at time T for the i^{th} claim is:

$$Y_{i,T} = \min\{\max\{(\sum_{t=1}^{T} X_{i,t}) - d'_{i,T}, 0\}, l'_{i,T}\} \tag{2.6}$$

It can be proved that, under the conditions $X_{i,T+1} \geq 0$ and $v_{t+1} \leq v_t \quad \forall t \leq T$, then $d'_{i,T+1} \geq d'_{i,T}$, $l'_{i,T+1} \geq l'_{i,T}$ and $Y_{i,T+1} \geq Y_{i,T}$. Proof of the third inequality is shown in Appendix A. The third inequality means that if the following two conditions are fulfilled:

(1) The primary insurer's incremental payment for the next period is a net outflow for any claim, and

(2) There is no deflation along the claim payment time horizon

then in the next period, incremental payments made by the NP reinsurer to the primary insurer is net outflow as well, meaning that the NP reinsurer would not request a payback from primary insurer. It is desirable to observe the third inequality because the primary insurer may be concerned that an increase in the indexed deductible over time could offset or exceed the increase in cumulative gross payment. That would not happen as indicated by the inequality $Y_{i,T+1} \geq Y_{i,T}$.

Next, consider incremental payments paid by primary insurer net of recoveries from the NP reinsurer at time T+1 for the i^{th} claim:

$$X_{i,T+1} - Y_{i,T+1} + Y_{i,T} \tag{2.7}$$

Under the conditions $X_{i,T+1} \geq 0$ and $v_{t+1} \leq v_t \quad \forall t \leq T$, then $(X_{i,T+1} - Y_{i,T+1} + Y_{i,T}) \geq 0$. That means at time T+1 primary insurer's gross incremental payment is always greater than the incremental recovery from the NP reinsurer. The proof is shown in Appendix A.

Sections 3.2 and 3.3 will discuss whether indexed AAD and AAL demonstrate similar desirable properties as well.

2.2.3 Inflation of Gross Claims, NP Reinsurer's Payments and Primary Insurer's Retained Claims

In Ferguson [1], "equitable share of inflation effect" means that applying the indexed deductible and indexed limit on gross claims will result in equal inflations for the primary insurer's retained claim and the NP reinsurer's claim payments.

According to equation (2.1), at time T, inflation for the i^{th} gross claim can be rewritten as:

$$\frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} = \frac{1}{w_{i,T}} \tag{2.8}$$

The notation $w_{i,T}$ represents the reciprocal of inflation at time T for the i^{th} gross claim.

Inflation for NP reinsurer's excess of loss payments is as follows:

$$\frac{\min\{\max\{(\sum_{t=1}^{T} X_{i,t}) - d'_{i,T}, 0\}, l'_{i,T}\}}{\min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}} \tag{2.9}$$

Inflation for the primary insurer's retained claim is as follows:

$$\frac{\min\{(\sum_{t=1}^{T} X_{i,t}), d'_{i,T}\}}{\min\{\sum_{t=1}^{T} X_{i,t} \cdot v_t), d\}} \tag{2.10}$$

It can be proved that, at time T for the i^{th} claim, the gross claim's inflation equals inflation for the NP reinsurer's excess of loss payments and also equals inflation for the primary insurer's retained claim. That means:

$$\frac{1}{w_{i,T}} = \frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} = \frac{\min\{\max\{(\sum_{t=1}^{T} X_{i,t}) - d'_{i,T}, 0\}, l'_{i,T}\}}{\min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}} = \frac{\min\{(\sum_{t=1}^{T} X_{i,t}), d'_{i,T}\}}{\min\{\sum_{t=1}^{T} X_{i,t} \cdot v_t), d\}} \tag{2.11}$$

The proof is shown in Appendix B.

2.2.4 An Alternative View: Indexing Deductibles and Limits by Principle of Equitable Sharing of Deflated Payments and Actual Payments

In Ferguson [1], "equitable share of deflated payments and actual payments" means that the ratio of the NP reinsurer's actual claim payment to actual gross claim equals the ratio of the NP reinsurer's deflated claim payment to deflated gross claim. This concept can be applied to explain equations (2.4) and (2.5).

If the index clause's selected index correctly reflects claims inflation at each payment time, then the following expression represents the value of the i^{th} claim as if all its future partial payments were paid at time 0.

$$\sum_{t=1}^{T} X_{i,t} \cdot v_t \tag{2.12}$$

Similarly, the following expression represents the NP reinsurer's payment to primary insurer as if all future partial payments of the i^{th} claim were paid at time 0:

$$\min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\} \tag{2.13}$$

That means that the un-indexed deductible and limit are directly applied to the total of all deflated payments for calculating excess layer loss.

What should be the NP reinsurer's share in the total actual payment of the i^{th} claim ($\sum_{t=1}^{T} X_{i,t}$)? The NP reinsurer should pay the proportion of $\sum_{t=1}^{T} X_{i,t}$, which is the same as the ratio of expression in (2.13) to expression in (2.12). That means NP reinsurer's share in the total actual payment is as follows:

$$\sum_{t=1}^{T} X_{i,t} \times \frac{\min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} \tag{2.14}$$

It can be verified that the above expression is identical to the right-hand side of equation (2.6), and therefore results in the same formulas for indexed deductible and indexed limit in equations (2.4) and (2.5).

Also, equation (2.14) shows the following relationships between the NP reinsurer's actual cumulative payment ($Y_{i,T}$) and the total deflated gross payments ($\sum_{t=1}^{T} X_{i,t} \cdot v_t$) for the i^{th} claim:

(1) $Y_{i,T} = 0$ when $\sum_{t=1}^{T} X_{i,t} \cdot v_t \le d$

That is, the NP reinsurer makes no payment if the total of deflated gross payments is below the un-indexed deductible.

(2) $Y_{i,T} > 0$ when $\sum_{t=1}^{T} X_{i,t} \cdot v_t > d$

That is, the NP reinsurer makes payment if the total of deflated gross payments is greater than the un-indexed deductible.

(3) $\sum_{t=1}^{T} X_{i,t} = d'_{i,T}$ if and only if $\sum_{t=1}^{T} X_{i,t} \cdot v_t = d$

That is, total deflated gross payments equal the un-indexed deductible if and only if the total actual gross payments equal the indexed deductible. Once reaching this condition for a particular claim, the NP reinsurer will start paying immediately after the primary insurer makes another payment for that claim in the future.

The third relationship is particularly useful for understanding reasonableness of the indexed deductible and limit formula. In section 3, in order to verify the formulas for indexing AAD's and AAL's, it will be checked whether total actual payments and total deflated payments on aggregate basis hold similar relationships.

3. INDEX CLAUSE FOR AGGREGATE DEDUCTIBLES AND LIMITS

AAD's are becoming more common for long-tail NP reinsurance. Without a proper index clause for AAD's, many NP reinsurance contracts simply endorse an "un-indexed" AAD, however, the per-claim deductible and limit are still subject to index clause adjustment. Such un-indexed AAD's in practice are simple to implement, but there are two problems. First, an un-indexed AAD may provide a misleading picture of how the NP reinsurer's expected loss will be reduced relative to "no-

AAD". Second, while an index clause for per-claim deductible and limit is used to share effect of inflation equitably between primary insurer and reinsurer, the goal cannot be achieved without an indexed AAD.

Consider this example: a NP reinsurance contract has per-claim limit of $5 million and per-claim deductible of $3 million both subject to an index clause adjustment, but with an un-indexed AAD of $5 million in place. Inflation is 4% per annum. If a claim is settled at $10 million by single payment in year 5, this claim is a total loss to the excess of loss layer. The per-claim limit and per-claim deductible are indexed to become $6.083 million and $3.650 million, respectively. If there was no AAD in place, the NP reinsurer would have paid $6.083 million to the primary insurer for this claim. With the $5 million un-indexed AAD, the NP reinsurer now pays $1.083 million. In this example, the primary insurer's additional retention under the un-indexed AAD provision is less than the occurrence of first total loss to the excess layer. This is not the expected outcome if one simply and carelessly interprets the structure to be $5mil xs $3mil xs $5mil, ignoring the gap between an indexed per-claim limit and un-indexed AAD.

3.1 Intuitive Arguments: Retrocessionaire's Point of View

Consider this example: a primary insurer purchases NP reinsurance $5 million xs $3 million with unlimited free reinstatements. An index clause will be applied to both per-claim deductible and limit.

The reinsurer wants to limit its potential frequency risk arising from this NP reinsurance contract, and decides to purchase a retrocession that caps the aggregate loss amount to the NP contract at an AAL equivalent to four times the per-claim-limit, which is $20 million. From the retrocessionaire's point of view, there should be an index clause for AAL as well, in order to share the effect of inflation between the reinsurer and retrocessionaire equitably.

The retrocessionaire now considers what factors shall and shall not enter into the AAL indexing formula. To start with, consider how the original index clause affects the transactions between the NP reinsurer, the primary insurer, and the original policyholder(s). If the original policy has a $5,000 policyholder retention, and the policyholder incurs one loss of $5,000,000, then the primary insurer will only pay $4,995,000. From the NP reinsurer's point of view, the actual amount paid by the primary insurer, $4,995,000, should be used for calculating the indexed per-claim deductible instead of the original policyholder's incurred loss of $5,000,000. Similarly, the retrocessionaire will only use the actual amount paid by the NP reinsurer (i.e., the difference between $4,995,000 and the indexed per-claim deductible) for calculating the indexed AAL, not the ground-up claim size of $4,995,000 paid by the primary insurer. Therefore, all claims below the indexed per-claim deductible should not be used for calculating the indexed AAL.

Following similar logics, the retrocessionaire makes a comparison between the original excess of loss program and the retrocession program:

	Original Excess of Loss program	Retrocession program
Deductible and Limit	Deductible and limit applied per claim.	"Aggregate" means that AAL is applied to sum of all excess layer losses. The sum of all excess layer losses is determined with index-clause-adjusted deductibles and limits applied to each claim separately
Basis of Inflation Measurement	Measured separately for each "claim", depending on the coverage basis. For example, if it is a per accident excess of loss, a "claim" may involve multiple claimants from the same accident and the sum of all claimant's claim amounts is used to determine loss to layer for each accident.	Measured for all losses to the excess of loss program combined together.
Measuring Inflation	Ratio of sum of all actual payments paid by the primary insurer to the policyholder(s) that belong to a "claim", to sum of all deflated payments that belong to the "claim"	Ratio of sum of all actual payments paid by the reinsurer to primary insurer according to the excess of loss program, to sum of all deflated payments to the excess of loss program

Conclusively, the AAL will be indexed by the formula:

$$\text{unindexed AAL} \times \frac{\text{Total of all actual payments to excess of loss layer}}{\text{Total of all deflated payments to excess of loss layer}} \tag{3.1}$$

Note that potential claim payouts by the above retrocessionaire are identical to the situation where a reinsurer sells a NP reinsurance \$5 million xs \$3 million with AAD \$20 million all subject to index clause. The method for indexing AAL can be applied for indexing AAD as well.

3.2 Indexing AAD and AAL Method 1: Matching Deflated Excess Loss with Deflated Gross Loss Per Claim

Additional notations are introduced, along with the notations in section 2.2:

D = un-indexed AAD

L = un-indexed AAL

D_T' = indexed AAD, given payment information up to time T for all claims

L_T' = indexed AAL, given payment information up to time T for all claims

The numerator of the fraction in equation (3.1), total of all actual payments to excess of loss layer, equals:

$$\sum_i Y_{i,T} \tag{3.2}$$

Next, consider the denominator of the fraction in equation (3.1), total of all deflated payments to excess layer. If gross, excess layer and retained claims are matched together, then deflated excess layer loss equals the difference between deflated gross loss and primary insurer's retention (un-indexed). For the i^{th} claim, deflated excess layer loss equals $\min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}$, which is identical to $Y_{i,T} \cdot w_{i,T}$. As a result the denominator is:

$$\sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\} = \sum_i Y_{i,T} \cdot w_{i,T}. \tag{3.3}$$

The formula for indexed AAL is as follows:

$$L_T' = L \times \frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}}. \tag{3.4}$$

The formula for indexed AAD is as follows:

$$D_T' = D \times \frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}}. \tag{3.5}$$

3.2.1 Inflation of Claims Before and After Application of Indexed AAD and AAL

Taking the retrocessionaire's point of view as described in section 3.1, the objective is to show that the following three programs have equal average inflation:

(1) Average inflation of total payments made by the NP reinsurer underlying the original excess

of loss contract, assuming unlimited reinstatements, equals:

$$\frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}}. \tag{3.6}$$

(2) Average inflation of total payments of the retrocession program that indemnifies the NP reinsurer portion of aggregate loss exceeding the indexed AAL equals:

$$\frac{\max\{(\sum_i Y_{i,T}) - L_T^{'}, 0\}}{\max\{(\sum_i Y_{i,T} \cdot w_{i,T}) - L, 0\}}. \tag{3.7}$$

(3) Average inflation of total payments made by the NP reinsurer underlying the original excess of loss contract, with the aggregate payments capped by the indexed AAL, equals:

$$\frac{\min\{(\sum_i Y_{i,T}), L_T^{'}\}}{\min\{(\sum_i Y_{i,T} \cdot w_{i,T}), L\}}. \tag{3.8}$$

As illustrated in section 3.1, inflation is measured for all claims to the excess of loss program combined, not measured for each claim separately. The three expressions in equations (3.6), (3.7), and (3.8) have very similar forms compared to the expressions in equations (2.8), (2.9), and (2.10) respectively.

Proof of the following equality:

$$\frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}} = \frac{\max\{(\sum_i Y_{i,T}) - L_T^{'}, 0\}}{\max\{(\sum_i Y_{i,T} \cdot w_{i,T}) - L, 0\}} = \frac{\min\{(\sum_i Y_{i,T}), L_T^{'}\}}{\min\{(\sum_i Y_{i,T} \cdot w_{i,T}), L\}} \tag{3.9}$$

is outlined below. A detailed proof is shown in Appendix C.

First consider equation (3.8). It can be shown that $(\sum_i Y_{i,T}) \le L_T^{'}$ if and only if $(\sum_i Y_{i,T} \cdot w_{i,T}) \le L$. Therefore, when the expression in (3.8) equals $L_T^{'} \div L$, by using definition of $L_T^{'}$ in equation (3.4), it can be shown that $L_T^{'} \div L$ equals $(\sum_i Y_{i,T}) \div (\sum_i Y_{i,T} \cdot w_{i,T})$, which is equal to the expression in (3.6). Otherwise, the expression in (3.8) equals $(\sum_i Y_{i,T}) \div (\sum_i Y_{i,T} \cdot w_{i,T})$. Again, this equals the expression in (3.6).

After proving equality of expressions in (3.6) and (3.8), it can be noted that the numerator in (3.6) equals the sum of the numerators in (3.7) and (3.8). Similarly, the denominator in (3.6) equals the sum of the denominators in (3.7) and (3.8) as well. Based on these facts, the expressions in (3.7) must equal the expressions in (3.6).

Conclusively, the equality in (3.9) holds. Inflation of the NP reinsurer's claims before and after application of indexed AAL (and AAD) are the same.

3.2.2 An Alternative View: Indexing AAD and AAL by Principle of Equitable Sharing of Deflated Payments and Actual Payments

First, note that the NP reinsurer's deflated aggregate excess layer payments without any AAL equal the total of the deflated gross partial payments with un-indexed deductibles and un-indexed limits applied to each claim separately. This can be represented by equation (3.3), $\sum_i Y_{i,T} \cdot w_{i,T} = \sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}$.

Similarly, the following expression represents the retrocessionaire's payment to the NP reinsurer if future partial payments of all claims were paid at time 0:

$$\max\{(\sum_i Y_{i,T} \cdot w_{i,T}) - L, 0\} = \max\{(\sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}) - L, 0\}. \tag{3.10}$$

What should be the retrocessionaire's share in the total actual excess layer payment $(= \sum_i Y_{i,T})$? The retrocessionaire should pay the proportion of $\sum_i Y_{i,T}$ that is same as the ratio of expression in (3.10) to expression in (3.3). That means that the retrocessionaire's share in the total actual payment is:

$$\sum_i Y_{i,T} \times \frac{\max\{(\sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}) - L, 0\}}{\sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}}. \tag{3.11}$$

It can be verified that the above expression is identical to the numerator of the expression in (3.7) and therefore results in the same formulas for indexed AAL in equation (3.4).

Also, equation (3.11) shows the following relationships between the retrocessionaire's actual cumulative payment $(\max\{(\sum_i Y_{i,T}) - L'_T, 0\})$ and the NP reinsurer's deflated aggregate excess layer loss before applying AAL $(\sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\})$:

(1) $\max\{(\sum_i Y_{i,T}) - L'_T, 0\} = 0$ when $\sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\} \leq L$.

That is, the retrocessionaire makes no payment if the NP reinsurer's aggregate deflated payments (before applying AAL) is below the un-indexed AAL.

(2) $\max\{(\sum_i Y_{i,T}) - L'_T, 0\} > 0$ when $\sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\} > L$.

That is, the retrocessionaire makes payment if the NP reinsurer's aggregate deflated payments (before applying AAL) is greater than the un-indexed AAL.

(3) $\sum_i Y_{i,T} = L'_T$ if and only if $\sum_i \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\} = L$.

That is, the NP reinsurer's aggregate deflated payments (before applying AAL) equals the un-indexed AAL if and only if the NP reinsurer's aggregate actual payments (before applying AAL) equals the indexed AAL. Once reaching this condition, the retrocessionaire will start paying immediately after the NP reinsurer makes another payment in the future.

3.2.3 Monotonicity Property of Retrocessionaire's Cumulative Payments

Incremental payments that the retrocessionaire makes to the NP reinsurer in the next period are considered net outflow and the retrocessionaire will not request a payback from the NP reinsurer, as long as the following two conditions are fulfilled:

(1) No deflation occurs along the claim payment time horizon.

(2) The total of all actual claims payments exceed the indexed deductible during current payment period.

The notation S'_T represents the retrocessionaire's cumulative actual payments to NP reinsurer at time T:

$$S'_T = \max\{(\sum_i Y_{i,T}) - L'_T, 0\}. \tag{3.12}$$

Therefore, the proposition means that, $S'_{T+1} \geq S'_T$ under the conditions $v_{t+1} \leq v_t \quad \forall t \leq T$ and $X_{i,T+1} \geq 0 \quad \forall i$ (implying that $(\sum_i Y_{i,T+1}) - (\sum_i Y_{i,T}) \geq 0$). The retrocessionaire's cumulative payment with indexed AAD and AAL is monotonically increasing over time. The proof is shown in Appendix D.

3.2.4 Monotonicity Property of Indexed AAD and AAL

In section 2.2.2, it was indicated that $d'_{i,T+1} \geq d'_{i,T}$ and $l'_{i,T+1} \geq l'_{i,T}$ under the conditions $v_{t+1} \leq v_t \quad \forall t \leq T$ and $X_{i,T+1} \geq 0 \quad \forall i$. The proof is shown in Appendix E.

Indexed AAD and AAL calculated using equations (3.4) and (3.5), however, are not monotonically increasing over time, even given the conditions $v_{t+1} \leq v_t \; \forall t \leq T$ and $X_{i,T+1} \geq 0 \; \forall i$. Indexed AAD and AAL are neither monotonically increasing nor decreasing over time. This is an undesirable property under practical considerations, which will be illustrated with a numerical example in section 3.4.4.

3.3 Indexing AAD and AAL Method 2: Deflating Incremental Excess Loss According to Payment Time

Another method to calculate deflated excess of loss payment is to multiply the deflating factor v_t with incremental actual payments of the NP reinsurer $(Y_{i,t} - Y_{i,t-1})$ and use as denominator of equation (3.1). Therefore:

$$\text{Total of all deflated payments to excess of loss layer} = \sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t . \tag{3.13}$$

The formula for indexed AAL and AAD becomes:

$$L_T'' = L \times \frac{\sum_i Y_{i,T}}{\sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t} . \tag{3.14}$$

$$D_T'' = D \times \frac{\sum_i Y_{i,T}}{\sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t} . \tag{3.15}$$

By rewriting the numerator $\sum_i Y_{i,T}$ as $\sum_{t=1}^{T} \sum_i (Y_{i,t} - Y_{i,t-1})$, equation (3.14) can be compared with equation (2.5):

	Equation (2.5): Indexed per-claim limit	Equation (3.14): Indexed AAL
Indexed limit	$l_{i,T}'$	L_T''
Numerator	$\sum_{t=1}^{T} X_{i,t}$	$\sum_{t=1}^{T} [\sum_i (Y_{i,t} - Y_{i,t-1})]$
Denominator	$\sum_{t=1}^{T} X_{i,t} \cdot v_t$	$\sum_{t=1}^{T} [\sum_i (Y_{i,t} - Y_{i,t-1}) \cdot v_t]$

Therefore, the concepts in sections 2.2.2, 2.2.3, and 2.2.4 can be applied to verify properties of the retrocessionaire's payment underlying the formula for indexed AAL in equation (3.14).

3.3.1 Inflation of Gross Claims, NP Reinsurer's Payments and Primary Insurer's Retained Claims

The following three programs have equal average inflation:

(1) Average inflation of the total payments made by the NP reinsurer underlying the original excess of loss contract, assuming unlimited reinstatements:

$$\frac{\sum_i Y_{i,T}}{\sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t}. \tag{3.16}$$

(2) Average inflation of total payments of the retrocession program that indemnifies the NP reinsurer portion of aggregate loss exceeding the indexed AAL:

$$\frac{\max\{(\sum_i Y_{i,T}) - L_T'', 0\}}{\max\{(\sum_{t=1}^{T} \sum_i (Y_{i,t} - Y_{i,t-1}) \cdot v_t) - L, 0\}}. \tag{3.17}$$

(3) Average inflation of total payments made by the NP reinsurer underlying the original excess of loss contract, with the aggregate payments capped by the indexed AAL:

$$\frac{\min\{(\sum_i Y_{i,T}), L_T''\}}{\min\{(\sum_{t=1}^{T} \sum_i (Y_{i,t} - Y_{i,t-1}) \cdot v_t), L\}}. \tag{3.18}$$

3.3.2 Indexing AAD and AAL by Principle of Equitable Sharing of Deflated Payments and Actual Payments

The following expression represents the retrocessionaire's payment to the NP reinsurer if future partial payments of all claims were paid at time 0:

$$\max\{(\sum_{t=1}^{T} \sum_i (Y_{i,t} - Y_{i,t-1}) \cdot v_t) - L, 0\}. \tag{3.19}$$

The retrocessionaire should pay the proportion of $\sum_i Y_{i,T}$, which is the same as the ratio of the expression in (3.19) to the expression in (3.13). That means that the retrocessionaire's share in the total actual payment is:

$$\sum_i Y_{i,T} \times \frac{\max\{(\sum_{t=1}^{T} \sum_i (Y_{i,t} - Y_{i,t-1}) \cdot v_t) - L, 0\}}{\sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t}. \tag{3.20}$$

It can be verified that the above equation agrees with the formula for indexed AAL in equation (3.14), therefore showing the following relationships between the retrocessionaire's actual cumulative payment $\max\{(\sum_i Y_{i,T}) - L_T'', 0\}$ and NP reinsurer's aggregate deflated payments according to the original excess of loss program before applying AAL ($\sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t$):

(1) $\max\{(\sum_i Y_{i,T}) - L_T'', 0\} = 0$ when $\sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t \le L.$

That is, the retrocessionaire does not make any payment if the NP reinsurer's aggregate deflated payments (before applying AAL) is below the un-indexed AAL.

(2) $\max\{(\sum_i Y_{i,T}) - L''_T, 0\} > 0$ when $\sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t > L$.

That is, the retrocessionaire makes payment if NP reinsurer's aggregate deflated payments (before applying AAL) is greater than the un-indexed AAL.

(3) $\sum_i Y_{i,T} = L''_T$ if and only if $\sum_i \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t = L$.

That is, the NP reinsurer's aggregate deflated payments (before applying AAL) equals the un-indexed AAL if and only if the NP reinsurer's aggregate actual payments (before applying AAL) equals the indexed AAL. Once reaching this condition, the retrocessionaire will start paying immediately after the NP reinsurer makes another payment in the future

3.3.3 Monotonicity Properties of Retrocessionaire's Cumulative Payments, Indexed AAD and AAL

The notation S''_T represents the retrocessionaire's cumulative actual payments to the NP reinsurer at time *T*:

$$S''_T = \max\{(\sum_i Y_{i,T}) - L''_T, 0\}. \tag{3.21}$$

It can be proved that $S''_{T+1} \ge S''_T$ under the conditions $v_{t+1} \le v_t \quad \forall t \le T$ and $X_{i,T+1} \ge 0 \quad \forall i$. Retrocessionaire's cumulative payment with indexed AAD and AAL [using equations (3.14) and (3.15)] is monotonically increasing over time. The proof is similar to the proof of $Y_{i,T+1} \ge Y_{i,T}$, shown in Appendix A.

Also under the conditions $v_{t+1} \le v_t \quad \forall t \le T$ and $X_{i,t} \ge 0 \quad \forall i$ and $\forall t$, the two sequences $\{L''_t\}$ and $\{D''_t\}$ are both monotonically increasing on *t*.

3.3.4 Incremental Payments Paid by the NP Reinsurer Net of Recoveries from the Retrocessionaire

Consider incremental aggregate payments paid by the NP reinsurer net of recoveries from the retrocessionaire at time *T*+1:

$$(\sum_i Y_{i,T+1} - \sum_i Y_{i,T}) - S''_{T+1} + S''_T. \tag{3.22}$$

Under the conditions $X_{i,T+1} \ge 0 \quad \forall i$ and $v_{t+1} \le v_t \quad \forall t \le T$, then $(\sum_i Y_{i,T+1} - \sum_i Y_{i,T} - S''_{T+1} + S''_T) \ge 0$. That is, at time *T*+1 the NP reinsurer's incremental payment to the primary insurer is always greater than the incremental recovery from the retrocessionaire.

Practically, if the NP reinsurer is not making a recovery from the primary insurer, then any claims emerging from the original excess of loss program will not result in a net cash-inflow for the NP reinsurer with the retrocession program in place.

3.4 A Numerical Example

The original excess of loss reinsurance program has a per-claim deductible of $3 million and per-claim limit of $1 million, both subject to the index clause. The original program has unlimited free reinstatements.

At time $T = 4$, there are three large claims, as shown in the following table.

Incremental Actual Gross Payment ($000s)

payment time	0	1	2	3	4	row sum
claim 1	$0.0	$2,120.0	$1,090.0	$0.0	$1,230.0	$4,440.0
claim 2	$0.0	$2,120.0	$2,180.0	$0.0	$0.0	$4,300.0
claim 3	$0.0	$0.0	$0.0	$4,680.0	$0.0	$4,680.0

Adjusted payments (or deflated payments) are calculated as follows:

Incremental Adjusted Gross Payment ($000s)

payment time	0	1	2	3	4	row sum
claim 1	$0.0	$2,000.0	$1,000.0	$0.0	$1,000.0	$4,000.0
claim 2	$0.0	$2,000.0	$2,000.0	$0.0	$0.0	$4,000.0
claim 3	$0.0	$0.0	$0.0	$4,000.0	$0.0	$4,000.0
Index	100	106	109	117	123	

All three claims have deflated values equal to the un-indexed ceiling (sum of un-indexed deductible and limit). According to the three relationships among the NP reinsurer's actual cumulative payment per claim and total deflated gross payments per claim illustrated in section 2.2.4, all three claims are total losses to the excess of loss program after the deductible and limit are indexed. However, the NP reinsurer's cumulative actual payments at time $T = 4$ are different for these three claims, as shown in the following table:

NP Reinsurer's Cumulative Actual Payments ($000s) at time $T = 4$

	Cumulative Actual Payments	Indexed Deductible	Indexed Limit	NP Reinsurer's Cum. Payment
claim 1	$4,440.0	$3,330.0	$1,110.0	$1,110.0
claim 2	$4,300.0	$3,225.0	$1,075.0	$1,075.0
claim 3	$4,680.0	$3,510.0	$1,170.0	$1,170.0
Total	$13,420.0	-	-	$3,355.0

3.4.1 Indexed AAL with Method 1: Matching Deflated Excess Loss with Deflated Gross Loss Per Claim

Assume that the NP reinsurer purchases a retrocession capping its potential aggregate payments at $3 million AAL subject to indexed clause. From the above table, it seems straight forward that indexed AAL should be $3.355 million at time 4, that is, the total of the losses to excess layer of the three claims.

For time periods before $T = 4$, what are the values of indexed AAL at each stage? And what if other un-indexed AAL ($2 million, $1 million) were chosen instead?

Table 3-1: Indexed AAL with Method 1 at Each Payment Time ($000s)

payment time	1	2	3	4
un-indexed AAL = $3mil	$3,000.0	$3,225.0	$3,367.5	$3,355.0
un-indexed AAL = $2mil	$2,000.0	$2,150.0	$2,245.0	$2,236.7
un-indexed AAL = $1mil	$1,000.0	$1,075.0	$1,122.5	$1,118.3

The above table is compared with the NP reinsurer's cumulative actual payments and cumulative deflated payments at each time. Deflated payment to excess of loss layer for the i^{th} claim is calculated using equation (3.3).

Table 3-2: NP Reinsurer's Cumulative Actual Payment ($000s)

payment time	0	1	2	3	4
claim 1	$0.0	$0.0	$0.0	$0.0	$1,110.0
claim 2	$0.0	$0.0	$1,075.0	$1,075.0	$1,075.0
claim 3	$0.0	$0.0	$0.0	$1,170.0	$1,170.0
Total	$0.0	$0.0	$1,075.0	$2,245.0	$3,355.0

Table 3-3: NP Reinsurer's Cumulative Deflated Payment with Method 1 ($000s)

payment time	0	1	2	3	4
claim 1	$0.0	$0.0	$0.0	$0.0	$1,000.0
claim 2	$0.0	$0.0	$1,000.0	$1,000.0	$1,000.0
claim 3	$0.0	$0.0	$0.0	$1,000.0	$1,000.0
Total	$0.0	$0.0	$1,000.0	$2,000.0	$3,000.0

3.4.2 Observations: Why Indexed AAL Changes over Time upon New Claims

From Table 3-3, the NP reinsurer's aggregate cumulative deflated payment equals $1 million at time 2. If an un-indexed AAL = $1 million was chosen, then indexed AAL at time 2 equals $1.075 million according to Table 3-1.

At time 3, the indexed AAL increases to $1.1225 million from $1.075 million at time 2. The increase in index AAL may sound intuitively incorrect, because if the purpose of AAL is to "limit" the NP reinsurer's aggregate payment, then it seems contradictory to observe an increase in the indexed AAL, even after one total loss has been observed at time 2. The phenomenon can be explained by considering the concepts of indexed per-claim deductible and limit as follows:

- Recall that $d'_{i,T+1} \geq d'_{i,T}$ and $l'_{i,T+1} \geq l'_{i,T}$ under the conditions $v_{t+1} \leq v_t \quad \forall t \leq T$ and $X_{i,T+1} \geq 0 \quad \forall i$. Therefore, indexed per-claim deductible and limit can increase over time, even if a claim is already a "total loss to excess layer" at a certain stage.

- To measure the effect of inflation, the formulas for indexed per-claim deductible and limit consider all gross payments known for a claim, irrespective of whether the claim has already become a total loss to excess layer in the past.

Similar arguments can be used to explain why indexed AAL can increase upon new payments made by the NP reinsurer at time 3.

3.4.3 Observations: Conditions when Indexed AAL by Method 1 Decreases upon New Claims

Now if un-indexed AAL = \$2 million was chosen, then indexed AAL at time 3 equals \$2.245 million according to Table 3-1. This is consistent with Table 3-3, indicating that two total losses are observed and that the NP reinsurer's aggregate cumulative actual payment equals \$2.245 million at time 3. However, at time 4, indexed AAL drops down to \$2.2367 million.

In general, if one can accept that per-claim indexed deductible and limit can increase over time when there is no deflation and that there is no gross claims recovery, it is reasonable to expect that similar monotonicity property shall be observed for indexed AAL. However, the above numerical example disproves any monotonicity property. It can be explained from two angles:

(1) In order to analyze why the indexed AAL at time 4 decreases, consider the change in average inflation. At time 3, claim 1 is still not observed as a loss to excess layer, and average inflation for claim 2 and claim 3 combined is 12.25% (= \$2.245mil ÷ \$2.000mil – 1). At time 4, claim 1 is observed as a total loss to the excess layer with average inflation 11.0%, which is lower than 12.25%. Average inflation for claims 1, 2, and 3 combined decreases to 11.83%, and therefore indexed AAL decreases accordingly.

(2) From Table 3-3, two total losses to excess layer are observed at time 3, therefore the NP reinsurer retains all cumulative actual payments to the excess of loss program (= \$2.245mil for claim 2 and claim 3 combined). At time 4, the NP reinsurer shall retain two-thirds of the aggregate cumulative actual payments to the excess of loss program when three total losses to excess layer are observed. Since claim 1's actual payment (\$1.110 million) is less than claims 2 and 3 combined average (\$1.1225 million), therefore two-thirds of the NP reinsurer's cumulative aggregate actual payment is only \$2.2367 million (= \$3.355 million × 2/3) and indexed AAL is adjusted downward accordingly.

Conclusively, from time T to T+1, indexed AAL decreases when the average inflation of aggregate claims to the original excess of loss program paid at time T+1 (e.g., consider claim 3, claim 2 and claim 1 combined) is lower than average inflation of aggregate claims to the original excess of loss program up to time *T* (e.g. consider claim 3 and claim 2 combined).

Although Method 1 for indexing AAL has no monotonicity property, the indexed AAL are at each stage correctly reflecting the split between the NP reinsurer's and the retrocessionaire's payments, according to principles of equitable sharing of inflation and equitable sharing of deflated payments, assuming that deflated value of gross, excess layer, and retained claims should be matched together.

3.4.4 Indexed AAL with Method 2: Deflating Incremental Excess Loss According to Payment Time

A deflating factor v_t is multiplied with incremental excess loss according to payment time. Resulting in tables of indexed AAL's and NP reinsurer's cumulative deflated payments that are different from the corresponding tables in section 3.4.1.

Table 3-4: Indexed AAL with Method 2 at Each Payment Time ($000s)

payment time	1	2	3	4
un-indexed AAL = $3mil	$3,000.0	$3,270.0	$3,390.8	$3,484.3
un-indexed AAL = $2mil	$2,000.0	$2,180.0	$2,260.6	$2,322.9
un-indexed AAL = $1mil	$1,000.0	$1,090.0	$1,130.3	$1,161.4

Table 3-5: NP Reinsurer's Cumulative Deflated Payment with Method 2 ($000s)

payment time	0	1	2	3	4
claim 1	$0.0	$0.0	$0.0	$0.0	$902.4
claim 2	$0.0	$0.0	$986.2	$986.2	$986.2
claim 3	$0.0	$0.0	$0.0	$1,000.0	$1,000.0
Total	$0.0	$0.0	$986.2	$1,986.2	$2,888.7

To compare Method 2 with Method 1, consider if the un-indexed AAL = $2 million was chosen.

Under Method 2, the NP reinsurer's cumulative aggregate deflated payment at time 3 equals $1.9862 million, which is less than the un-indexed AAL of $2 million. Therefore, at time 3, indexed AAL ($2.2606 million) should be greater than the NP reinsurer's cumulative aggregate actual payment ($2.245 million).

However, Table 3-2 indicates that the NP reinsurer's actual payments for claim 2 and claim 3 are both total loss to excess layer, since both payments equal their corresponding indexed per-claim limit. Ideally, indexed AAL should equal $2.245 million at this stage, such that the retrocessionaire will start to pay immediately after another excess layer claim is observed. Comparing the two methods for indexing AAL, Method 1 can always satisfy such a requirement, because when deflating excess layer loss, Method 1 takes into account matching of gross, retained, and excess layer payments. Method 2, however, generally cannot satisfy such a requirement as it ignores the link between gross and excess layer payments.

Despite the above advantage, Method 1 has a major shortcoming. At time 4, the NP reinsurer pays $1.110 million to the primary insurer but receives $1.1183 million (= $3.355mil – $2.2367mil) from the retrocessionaire, therefore resulting in net cash-inflow for the NP reinsurer despite claim emergence. Scenarios similar to this are problematic because often the primary insurer practically takes up the role of the retrocessionaire: that means the NP reinsurer is only liable up to the indexed AAL and then the primary insurer will be responsible for the portion of aggregate claims above. In this numerical example, the primary insurer makes a payment of $1.230 million to its policyholder for claim 1, and also makes a net payment of $8,300 (= $1.1183mil – $1.110mil) to the NP reinsurer. Practically, the primary insurers may not be convinced to make the payment to NP reinsurer under an indexed AAL, especially since they will not need to do so if the AAL is simply un-indexed. Under Method 2, however, the situation becomes different: at time 4, the NP reinsurer makes a net payment of $77,900 to the primary insurer (= $1.110mil – [$3.355mil – $2.3229mil]). As illustrated in section 3.3.4, the NP reinsurer's incremental payment to the primary insurer is always greater than the incremental recovery from the retrocessionaire.

4. PRACTICAL ISSUES

4.1 Which Method to Use for Indexing AAD and AAL: Method 1 or Method 2?

It is not straightforward to decide whether Method 1 or Method 2 is the correct method simply by relying on principles of equitable sharing of inflation and equitable sharing of deflated payments. Each method uses its own way to determine deflated excess loss, therefore, equitable sharing can be "achieved by definition". A comparison from both a theoretical and practical point of view is shown below:

	Method 1: Matching Deflated Excess Loss with Deflated Gross Loss Per Claim	Method 2: Deflating Incremental Excess Loss According to Payment Time
Advantages	(1) Theoretical: Indexed AAL match with Indexed per-claim limit (2) Practical: If un-indexed AAL is chosen to be *k* times un-indexed per-claim limit, then occurrence of *k* total losses, but not more, will be exactly covered under indexed AAL	(1) Practical: As long as no deflation, indexed AAD and AAL increase over time when claims emerge (2) Practical: NP reinsurer always has net cash-outflow when claims emerge which sounds reasonable
Disadvantages	(1) Practical: Indexed AAD and AAL may decrease over time, which can be difficult to explain to primary insurers (2) Practical: Decreasing AAL may require the primary insurer (who takes the retrocessionaire's role) to make extra payment to the NP reinsurer besides paying the gross claim	(1) Theoretical: Indexed AAL mismatch with Indexed per-claim limit (2) Practical: If un-indexed AAL is chosen to be *k* times un-indexed per-claim limit, occurrence of *k* total losses generally result in indexed AAL greater than total of the indexed per-claim limit of the *k* total losses

Although Method 1 appears to be more appropriate from a theoretical point of view by matching both actual and deflated excess loss with gross loss and retained loss, in practice the importance of such theoretical advantage is not easily observable. Generally, when AAL is exhausted it is more likely to observe a mix of partial losses and total losses to the excess layer rather than purely total losses. The theoretical advantage only has more meaning in terms of coverage interpretation: un-indexed AAL equals k times un-indexed per-claim limit implies that exactly k total losses will be covered.

Practically Method 2 will likely receive higher level of acceptance by the market. It is because under Method 2 indexed AAL retains most of the desirable properties that are observed in indexed per-claim deductible and limit, including:

- Equitable sharing of inflation and equitable sharing of deflated payments (although "equitable sharing" depends on excess loss deflating method assumption).
- Indexed AAL "increases with claims inflation" (indexed AAL increases over time when claims emerge and inflation is positive).
- All parties (primary insurer, NP reinsurer, retrocessionaire) have net cash-outflow when claims emerge (and inflation is positive).

Indexed AAD and AAL under Method 2 are generally greater than that under Method 1. Therefore primary insurer may prefer to use Method 2 for indexed AAL, and the NP reinsurer may prefer to use Method 2 for indexed AAD.

4.2 Pricing Excess of Loss Reinsurance with Indexed AAD and AAL

The objective of pricing is to estimate expected loss cost for the prospective quotation year, and express the estimated value as a percentage of Gross Net Premium Income (GNPI) for the quotation year. This percentage is often called risk rate of the reinsurance program.

In this section, the view of the retrocessionaire as illustrated in section 3 will be taken. In taking this view, the objective is to estimate the expected value of aggregate loss cost to the original excess layer program that exceed the "indexed AAL". Using the notations in section 3, the expected value of the random variable S_T'' (or S_T' if Method 1 for indexing AAL is chosen) will be calculated. In the following discussion of various pricing approaches, it is assumed that Method 2 for indexing the AAL is chosen. Nevertheless most procedures and observations are appropriate for both Method 1 and Method 2.

Additional assumptions and notations are as follows:

T = time when all claims to the original excess of loss program are settled, assuming that T is not a random variable (e.g., one can choose T to be 50 years or even 100 years if the line of business has an extremely long tail, but practically 20 years or 25 years shall be reasonable choices)

$X_{i,t}$ = random variables for incremental loss payment, from losses that occur during the prospective quotation year (but revalued as if the occurrence date is the average accident date). As a result, $Y_{i,T}$ (loss to excess layer) and S_T'' (aggregate loss excess of indexed AAL) are random variables too. In addition, $d_{i,t}'$, $l_{i,t}'$ (indexed deductible and limit) and L_T'' are random variables as well.

In practice the distribution of $\sum_{t=1}^{T} X_{i,t}$ (ultimate ground-up loss random variable) is often modeled first, then the payment pattern at time t ($X_{i,t} \div \sum_{t=1}^{T} X_{i,t}$) is estimated.

N = number of loss random variable for the prospective quotation year. The definition of "loss occurrence" needs to match with the distribution of $\sum_{t=1}^{T} X_{i,t}$. For modeling convenience, loss occurrence can be defined as the event when $\sum_{t=1}^{T} X_{i,t}$ exceeds the indexed deductible, therefore it is then only necessary to model the severity of large losses that hit the excess of loss program.

p = GNPI for the prospective quotation year. Assume that p can be forecasted accurately at inception.

S_T'' = $\max\{(\sum_{i=1}^{N} Y_{i,T}) - L_T'', 0\}$ = the random variable of aggregate loss cost to the original excess layer program that exceeds the indexed AAL (i.e., aggregate loss cost to the retrocession program)

$\frac{\mathrm{E}[S_T'']}{p}$ = risk rate of the retrocession program = ratio of expected value of S_T'' to GNPI of the prospective quotation year

4.2.1 Empirical Approach

The empirical approach (also called burning cost approach) uses claims and GNPI in historical observation year(s):

- Step 1: historical ground-up claim sizes are revalued for claims inflation. For long-tail classes, claim payments for future development years and Pure IBNR need to be forecasted.
- Step 2: by using the deductible and limit for the prospective quotation year, indexed deductibles and limits are determined for calculating excess layer loss for each claim.
- Step 3: aggregate (as-if) actual excess loss is determined for each payment time, and therefore

aggregate (as-if) deflated excess loss can be determined for each payment time as well, in order to determined indexed AAL (L_T'') at final settlement time T.

- Step 4: risk rate of the retrocession program is estimated as the ratio of aggregate (as-if) actual excess loss exceeding L_T'' to on-level GNPI of a historical year. If more than one historical year is available, weighted average of the ratios is taken as the risk rate of the retrocession program.

Notations

N^h = number of loss random variable for a historical observation year

$Y_{i,T}^h$ = random variable for as-if loss to the excess layer, by revaluing historical ground-up loss random variable in an observation year for claims inflation

p^h = on-level GNPI for a historical observation year

Underlying the empirical approach, it is assumed that if on-level GNPI and claim sizes are revalued appropriately, then expected historical loss frequency (= number of claims per on-level GNPI) equals prospective quotation year's expected loss frequency:

$$\frac{\mathrm{E}[N^h]}{p^h} = \frac{\mathrm{E}[N]}{p}. \tag{4.1}$$

Generally $p^h \neq p$. For example, when portfolio growth is not due to rate increases, then $p^h < p$ and $\mathrm{E}[N^h] < \mathrm{E}[N]$. For this reason, if one uses the following expression to estimate risk rate of the retrocession program:

$$\frac{1}{p^h} \times \mathrm{E}[\max\{(\sum_{i=1}^{N^h} Y_{i,T}^h) - L \times \frac{\sum_{i=1}^{N^h} Y_{i,T}^h}{\sum_{i=1}^{N^h}\sum_{t=1}^{T}(Y_{i,t}^h - Y_{i,t-1}^h)\cdot v_t}, 0\}]. \tag{4.2}$$

Then risk rate will likely be underestimated since $\mathrm{E}[\sum_{i=1}^{N^h} Y_{i,T}^h]$ is less than $\mathrm{E}[\sum_{i=1}^{N} Y_{i,T}]$ but the same L (= un-indexed AAL) is used.

Often a conventional solution is to modify L by multiplying with p^h/p. As a result risk rate of the retrocession program is estimated by the expression:

$$\frac{1}{p^h} \times \mathrm{E}[\max\{(\sum_{i=1}^{N^h} Y_{i,T}^h) - \frac{L \cdot p^h}{p} \times \frac{\sum_{i=1}^{N^h} Y_{i,T}^h}{\sum_{i=1}^{N^h}\sum_{t=1}^{T}(Y_{i,t}^h - Y_{i,t-1}^h)\cdot v_t}, 0\}]. \tag{4.3}$$

However, the expression in (4.3) is still a biased estimator of the risk rate. The proof is straightforward by considering the short-tail case, which means per-claim deductible, limit, and AAL will not be indexed.

There are other shortcomings with the empirical approach. For example, when pricing an excess of loss layer without any AAD or AAL, often more than one claim is observed in each accident year on average. Observing 10-years experience can generally provide a reasonably large sample size. However, under the empirical approach, each observation year is only considered to be one sample. Overall, the empirical approach is not highly accurate for estimating the risk rate for the retrocession program.

4.2.2 Simulation Approach

One option is to apply a "historical simulation" approach:

- Step 1: realized values of pairs of $Y_{i,T}^{h}$ (revalued ultimate actual excess loss) and $\sum_{t=1}^{T}(Y_{i,t}^{h} - Y_{i,t-1}^{h}) \cdot v_t$ (revalued ultimate deflated excess loss) from all observation years are collected to form a pool of sample losses. Forecast of future claim payment development may be needed. Equal weights can be assigned to each realized pair.
- Step 2: on-level GNPI (p^h) and realized values of N^h are used to estimate E[N] and/or other parameters for distribution of N. An allowance for Pure IBNR may be needed.
- Step 3: in each simulated scenario, the number of losses are simulated from distribution of N. Then loss sizes are sampled randomly from the pool of actual and deflated loss pairs, which would then allow calculation of simulated values of $\sum_{i=1}^{N} Y_{i,T}$ (aggregate actual loss cost to the original excess layer program), $\sum_{i=1}^{N^h}\sum_{t=1}^{T}(Y_{i,t} - Y_{i,t-1}) \cdot v_t$ (aggregate deflated loss cost to the original excess layer program), and L_T'' (indexed AAL) and finally S_T'' (aggregate loss cost to the retrocession program).
- Step 4: repeat scenario generations in Step 3 until sufficiently large number of scenarios are generated. Then take the average of the simulated S_T'' divided by p as the risk rate for the retrocession program.

Historical simulation approach can be viewed as a refinement of empirical approach, by making use of empirical distribution of historical loss sizes (actual and deflated) while matching with prospective quotation year's loss frequency through the simulation procedure. Historical simulation approach is an appropriate choice when a large reliable sample of historical losses is available.

Another simulation approach alternative is to model the severity distribution of $\sum_{t=1}^{T} X_{i,t}$ (ultimate ground-up loss random variable) as well as the payment pattern at time t

($X_{i,t} \div \sum_{t=1}^{T} X_{i,t}$). After ground-up severity and payment pattern are simulated, $\sum_{i=1}^{N} Y_{i,T}$ and $\sum_{i=1}^{N^b} \sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t$ can be calculated as well.

Options for modeling payment pattern include:

(1) Deterministic payment pattern: every simulated ground-up loss has the same payment pattern between time $t = 1$ and T.

(2) Stochastic payment pattern that is independent of $\sum_{t=1}^{T} X_{i,t}$

(3) Stochastic payment pattern that varies with $\sum_{t=1}^{T} X_{i,t}$. For example, large claims generally take a longer time to reach full settlement than small claims. However, the modeler should judge the strength of dependency between claim size and payment pattern for claims that penetrate the excess layer, and thus whether it is necessary to insert such extra complexity in the simulation procedure.

If it is decided to model the payment pattern stochastically, one simplification is to model "average settlement time". It is assumed that each claim is settled fully with a single payment at some time between 1 and T. However, remember that with this simplification, the indexed AAL calculated under Method 2 will always be the same as the indexed AAL calculated under Method 1. Therefore, it is not recommended to use such simplification if Method 2 for indexing AAL is chosen.

Even if the same deterministic payment pattern is applied for all ground-up claims, different excess layer payment patterns will still be observed for claims of different sizes: larger claims will have shorter average excess layer payment patterns. The implications are very different for indexing AAL with Method 1 or Method 2. If Method 1 for indexing AAL is chosen, then for each claim the ratio of actual excess loss to deflated excess loss equals $Y_{i,T} \div (Y_{i,T} \cdot w_{i,T})$ and is the same for all ground-up claim sizes. However, if Method 2 for indexing AAL's is chosen, then for each claim the ratio of actual excess loss to deflated excess loss equals $Y_{i,T} \div (\sum_{t=1}^{T} (Y_{i,t} - Y_{i,t-1}) \cdot v_t)$, and the ratio is higher for smaller claims to the excess layer. The above observations do not add extra complexity to the simulation approach if Method 2 for indexing AAL is chosen, but it is necessary to consider whether the implications reasonably reflect the reality.

4.2.3 Collective Risk Model

For short-tail classes, in order to estimate expected aggregate loss cost to the original excess layer program that exceed an un-indexed AAL, it is often convenient to adopt a collective risk model approach as follows:

- Step 1: the distribution of actual loss to excess layer random variable $Y_{i,T}$ is approximated by a discrete distribution.
- Step 2: some choice of distribution for number of loss random variable N (e.g., any (a,b,0) class distribution) allows a recursive formula to be used for determining distribution of aggregate loss cost to the original excess layer program $\sum_{i=1}^{N} Y_{i,T}$.
- Step 3: since un-indexed AAL is a constant, it is straightforward to calculate expected value of $\max\{(\sum_{i=1}^{N} Y_{i,T}) - L, 0\}$.

For long-tail classes, however, it is not that straightforward to calculate the expected value of $S_T'' = \max\{(\sum_{i=1}^{N} Y_{i,T}) - L_T'', 0\}$ because L_T'' is a random variable dependent on $Y_{i,t}$'s. Similar to section 4.2.2, there are several options to model payment patterns such that the distribution of L_T'' can be simplified as follows:

(1) Modeling L_T'' stochastically: assume L_T'' equals L multiplied by a random variable M. The expected value of M shall equal the average ratio of actual aggregate excess loss to deflated aggregate excess loss, and M is assumed to be independent of $Y_{i,t}$'s. It is reasonable to choose M to be lower bounded by 1 and to have an upper bound. Then the expected value of $\mathrm{E}[S_T'']$ can be calculated using conditional expectation:

$$\mathrm{E}[S_T''] = \mathrm{E}_M[\,\mathrm{E}[S_T'' \mid M]\,] = \mathrm{E}_M[\,\mathrm{E}[\max\{(\sum_{i=1}^{N} Y_{i,T}) - L \cdot M, 0\} \mid M]\,]. \tag{4.4}$$

(2) Deterministic payment pattern (Excess): every excess layer loss has the same payment pattern between time t = 1 and T. This is appropriate when Method 2 is chosen for indexing AAL, because L_T'' is then no longer a random variable. L_T'' is calculated as L multiplied by the reciprocal of deflated value of $1 using the selected deterministic payment pattern. Expected value of S_T'' can then be calculated easily like in the short-tail case.

(3) Deterministic payment pattern (Ground-up): every ground-up loss has the same payment pattern between time t = 1 and T. This is appropriate when Method 1 is chosen for indexing AAL, because L_T' is then no longer a random variable.

4.2.4 Allowance for Investment Income

Most loss payments of the retrocession program are paid long after the quotation year. Risk premium of the retrocession program calculated by any of the pricing approaches in section 4.2.1 to 4.2.3 shall be reduced by investment income that can be earned by the retrocessionaire between the time of premium installments and the time of loss payments.

One option is to determine an average payment pattern of the retrocession program's loss payments. Then a discount for investment income can be calculated deterministically, and multiplied with $\mathrm{E}[S_T^{''}]$ determined from the selected pricing approach.

A second option is to incorporate investment income allowance directly into stochastic modeling of loss to the retrocession program. Recall that $S_T^{''} = \max\{(\sum_{i=1}^{N} Y_{i,T}) - L_T^{''}, 0\}$ is the random variable of aggregate loss cost to the retrocession program before including allowance for investment income, then $\mathrm{E}[S_T^{''} - S_{T-1}^{''}]$ equals the expected loss payment of the retrocession program at time T. Therefore, assuming all retrocession premiums are received on the base date, risk premium for the retrocession program including investment income allowance equals:

$$\sum_{t=1}^{T} \frac{\mathrm{E}[S_t^{''}] - \mathrm{E}[S_{t-1}^{''}]}{(1+r_t)^t}. \tag{4.5}$$

Where r_t denotes the annualized investment return from time 0 to t.

The second option is more practical if simulation pricing approach is used, for which not much extra modeling complexity will be added to the simulation procedures.

If a collective risk model pricing approach is used with $L_T^{''}$ modeled stochastically, much effort is needed in determining $\mathrm{E}[S_t^{''}] - \mathrm{E}[S_{t-1}^{''}]$ for all t between 1 and T, since it is necessary to define distributions of $L_t^{''} = L \cdot M_t$ for all t between 1 and T.

4.3 Limited Reinstatement and Calculating Paid Reinstatement Premium with Indexed AAL

4.3.1 Revision: Calculating Paid Reinstatement Premium for Short-Tail Classes

In short-tail classes, paid reinstatement premium are most often paid "at 100% additional premium as to time but pro rata as to amount reinstated only" (also called "100% pro-rata capita"). It means that upon occurrence of any claim to the excess layer with ground-up size X, irrespective of the time of loss occurrence or loss payment, the primary insurer pays an additional reinstatement premium to the NP reinsurer of the following amount:

$$\text{GNPI} \times \text{reinsurance premium rate} \times \frac{\min\{X-d, l\}}{l}. \tag{4.6}$$

Paid reinstatement provision is often associated with limited number of reinstatements. For example, if "two full reinstatements" are offered, it is identical to state that the excess layer has an AAL that equals three times the per-claim limit. In general, relationship between annual aggregate limit L, per-claim limit l and number of reinstatements k can be represented by the equation:

$$\text{number of reinstatements} = k = \frac{L}{l} - 1. \tag{4.7}$$

The maximum possible amount of total reinstatement premium paid by the primary insurer equals:

$$\text{GNPI} \times \text{reinsurance premium rate} \times \frac{L-l}{l}. \tag{4.8}$$

Therefore, the primary insurer is not required to pay reinstatement premium for the portion of aggregate excess layer loss that exceeds $(L - l)$. Here the author introduces the term "Annual Aggregate Reinstatement Limit", or AARL, to describe the value $(L - l)$.

To generalize, if N claims are observed each with ground-up size X_i, then total reinstatement premium paid by primary insurer equals:

$$\text{GNPI} \times \text{reinsurance premium rate} \times \min\{(\sum_{i=1}^{N} \frac{\min\{\max\{X_i - d, 0\}, l\}}{l}), k\}. \tag{4.9}$$

4.3.2 Calculating Paid Reinstatement Premium for Long-Tail Classes with Indexed Per-Claim Deductible, Limit and Method 1 for Indexing AAL

Recall that in equation (2.6), $Y_{i,T} = \min\{\max\{(\sum_{t=1}^{T} X_{i,t}) - d'_{i,T}, 0\}, l'_{i,T}\}$ represents the cumulative excess layer loss at time T for the i^{th} claim, and that $l'_{i,T}$ represents the indexed limit for the i^{th} claim. Modifying equation (4.9) so as to fit into long-tail environment implies that, at time T, the cumulative total reinstatement premium paid by the primary insurer equals:

$$\text{GNPI} \times \text{reinsurance premium rate} \times \min\{(\sum_{i=1}^{N} \frac{Y_{i,T}}{l'_{i,T}}), k\}. \tag{4.10}$$

It can be easily verified that equation (4.10) is identical to the following:

$$\text{GNPI} \times \text{reinsurance premium rate} \times \min\{(\sum_{i=1}^{N} \frac{Y_{i,T}}{L'_T \times \frac{l}{L}}), \frac{L-l}{l}\}. \tag{4.11}$$

Equation (4.11) indicates that the primary insurer is not required to pay reinstatement premium for the portion of aggregate excess layer loss that exceeds $(L-l) \times \frac{L'_T}{L}$ (= indexed value of AARL).

In practice, equation (4.10) is the easier method to represent how paid reinstatement should be calculated, but it brings up several issues:

(1) It is possible and reasonable that under some circumstances the NP reinsurer is required to pay the primary insurer for excess claim, but the primary insurer is not required to pay any

reinstatement premium at the same time.

To demonstrate this, use the numerical example in section 3.4. For example, if four full reinstatements each at 100% pro-rata capita is offered. At time 4, three total excess layer losses are observed, therefore $\sum_{i=1}^{N}(Y_{i,T} \div l'_{i,T})$ equals 300%. Assume at time 5, no other claims are reported, but the primary makes another payment for claim 1. As a result, indexed limit for claim 1 at time 5 (= $l'_{1,5}$) is greater than that at time 4 (= $l'_{1,4}$), and the NP reinsurer is required to pay primary insurer the difference between $l'_{1,5}$ and $l'_{1,4}$. However, $(Y_{1,5} \div l'_{1,5}) = (Y_{1,4} \div l'_{1,4}) = 100\%$ and therefore $\sum_{i=1}^{N}(Y_{i,T} \div l'_{i,T})$ at time 5 is unchanged at 300%, which means that primary insurer is not required to pay additional reinstatement premium at time 5.

The above observation sounds contradictory to the reinstatement premium calculation performed in the short-tail case, where reinstatement premium is received by the NP reinsurer every time an excess claim is paid until the AARL is used up.

The author suggests, however, that a broader view should be taken to interpret the reinstatement premium calculation if it is to compare with the short-tail case. The reinstatement premium is received by the NP reinsurer every time a per-claim limit needs to be reinstated. In the numerical example, the difference between $l'_{1,5}$ and $l'_{1,4}$ simply reflects an adjustment of claim 1's indexed limit due to the updated average inflation information for this claim, but does not involve any portion of the limit being used from time 4 to 5. Therefore no limit needs to be reinstated. The portion of per-claim limit is used and needs to be reinstated if and only if $(Y_{i,t} \div l'_{i,t}) > (Y_{i,t-1} \div l'_{i,t-1})$ but not just under the condition $Y_{i,t} > Y_{i,t-1}$.

(2) As a result, the NP reinsurer's loss payment should not be constrained by whether the condition $\sum_{i=1}^{N}(Y_{i,T} \div l'_{i,T}) \geq k$ has been met at a particular point of time, but should only be capped by the AAL.

(3) Method 1 for indexing AAL is the method that is consistent with equation (4.10). This means that when $\sum_{i=1}^{N}(Y_{i,T} \div l'_{i,T})$ (= total of ratios of actual excess claim to indexed per-claim limit) exactly equals the number of full reinstatements offered, then the remaining "unused AAL" will be sufficient to pay exactly one more total loss to the excess layer (or equivalent) that will emerge in the future.

(4) When an index clause applies to per-claim deductible and limit only but not AAL, it can be problematic if paid reinstatement provision is in place. For example, it is possible that aggregate excess layer loss exceeds the un-indexed AAL, but cumulative total reinstatement premium has not yet reached the maximum according to equation (4.10). Further, if aggregate excess layer loss is less than the un-indexed AAL at time *T*–1 but exceeds the un-indexed AAL at time *T*, then what should be the amount of reinstatement premium to be paid at time *T*?

4.3.3 Calculating Paid Reinstatement Premium for Long-Tail Classes with Indexed Per-Claim Deductible, Limit and Method 2 for Indexing AAL

When Method 2 for indexing AAL (that means L_T'' calculated using equation (3.14)) is chosen, then $\sum_{i=1}^{N}(Y_{i,T} \div l'_{i,T})$ is not a measure of "used limit" that is consistent with L_T''. Consider an example, when the number of total excess layer losses occurred equals *k* (= number of full reinstatements offered), then the remaining "unused AAL" is sufficient to pay future occurrences of one more total excess layer loss plus another partial loss to the excess layer.

Attempting to correct the inconsistency, modifying equation (4.11) can result in the following formula for cumulative total reinstatement premium paid by the primary insurer at time *T*:

$$\text{GNPI} \times \text{reinsurance premium rate} \times \min\{(\sum_{i=1}^{N} \frac{Y_{i,T}}{L_T'' \times \frac{l}{L}}), k\}. \tag{4.12}$$

However, the above formula only corrects the inconsistency partially. Further comparing with equation (4.10), it is much more difficult to explain the concept and reasonableness of equation (4.12) when the calculation of the reinstatement premium in short-tail case had already been widely accepted in the market.

Conclusively, it is still reasonable in practice to use equation (4.10) to calculate reinstatement premium even if Method 2 for indexing AAL is chosen.

5. CONCLUSIONS

Two methods for indexing AAD and AAL are presented in this paper: Method 1 matches deflated excess loss with deflated gross loss per claim, and Method 2 deflates incremental excess loss according to payment time. The two methods are developed with concepts that are closely linked to the concepts underlying indexation of per-claim deductible and limit.

In comparing the advantages and disadvantages of the two methods from a practical point of view, indexed AAL's with Method 2 retain most of the desirable properties that are observed in the indexed per-claim deductible and limit. Method 2 will likely receive a higher level of acceptance by the market.

For the various proposed pricing approaches, the empirical approach (burning cost approach) is less preferable than the simulation or collective risk model approaches. In fact, the accuracy of the empirical approach is questionable even in the short-tail case with un-indexed AAD and AAL.

Finally, the method for calculating reinstatement premium is applicable whether Method 1 or Method 2 for indexing AAD and AAL is chosen.

Acknowledgment

The author acknowledges Paul Vendetti and Gary Blumsohn for their comments on the draft of this paper.

Appendix A : Monotonicity Properties of NP Reinsurer's Cumulative Payment and Primary Insurer's Net Cash-flow under Indexed Per-Claim Deductible and Limit

Proposition 1: NP reinsurer's cumulative payment made to primary insurer is monotonically increasing, that is, $Y_{i,T+1} \geq Y_{i,T}$, under the conditions $X_{i,T+1} \geq 0$ and $v_{t+1} \leq v_t \; \forall t \leq T$.

Proof:

By using equations (2.4), (2.5) , and (2.6), express $Y_{i,T}$ in terms of $X_{i,t}$, v_t, d, and l:

$$Y_{i,T} = (\sum_{t=1}^{T} X_{i,t} \div \sum_{t=1}^{T} X_{i,t} \cdot v_t) \times \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}$$

And similarly $Y_{i,T+1} = (\sum_{t=1}^{T+1} X_{i,t} \div \sum_{t=1}^{T+1} X_{i,t} \cdot v_t) \times \min\{\max\{(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t) - d, 0\}, l\}$

Next, consider three cases of $\sum_{t=1}^{T} X_{i,t}$. (I) $\sum_{t=1}^{T} X_{i,t} > l'_{i,T} + d'_{i,T}$; (II) $\sum_{t=1}^{T} X_{i,t} \leq d'_{i,T}$; and (III) $d'_{i,T} < \sum_{t=1}^{T} X_{i,t} \leq l'_{i,T} + d'_{i,T}$;

Case (I): when $\sum_{t=1}^{T} X_{i,t} > l'_{i,T} + d'_{i,T}$.

From equations (2.4) and (2.5) $\Rightarrow \sum_{t=1}^{T} X_{i,t} > (l+d)\times(\sum_{t=1}^{T} X_{i,t} \div \sum_{t=1}^{T} X_{i,t}\cdot v_t)$

$\Rightarrow \sum_{t=1}^{T} X_{i,t}\cdot v_t > l + d$

$$
\begin{aligned}
\therefore Y_{i,T+1} - Y_{i,T} &= (\sum_{t=1}^{T+1} X_{i,t} \div \sum_{t=1}^{T+1} X_{i,t}\cdot v_t)\times l - (\sum_{t=1}^{T} X_{i,t} \div \sum_{t=1}^{T} X_{i,t}\cdot v_t)\times l \\
&= l \times \frac{(\sum_{t=1}^{T+1} X_{i,t})\times(\sum_{t=1}^{T} X_{i,t}\cdot v_t) - (\sum_{t=1}^{T} X_{i,t})\times(\sum_{t=1}^{T+1} X_{i,t}\cdot v_t)}{(\sum_{t=1}^{T+1} X_{i,t}\cdot v_t)\times(\sum_{t=1}^{T} X_{i,t}\cdot v_t)} \\
&= l \times X_{i,T+1} \times \frac{(\sum_{t=1}^{T} X_{i,t}\cdot v_t) - (\sum_{t=1}^{T} X_{i,t}\cdot v_{T+1})}{(\sum_{t=1}^{T+1} X_{i,t}\cdot v_t)\times(\sum_{t=1}^{T} X_{i,t}\cdot v_t)} \\
&\geq l \times X_{i,T+1} \times \frac{(\sum_{t=1}^{T} X_{i,t}\cdot v_t) - (\sum_{t=1}^{T} X_{i,t}\cdot v_t)}{(\sum_{t=1}^{T+1} X_{i,t}\cdot v_t)\times(\sum_{t=1}^{T} X_{i,t}\cdot v_t)} = 0
\end{aligned}
$$

Case (II): when $\sum_{t=1}^{T} X_{i,t} \leq d'_{i,T}$.

From equations (2.4) and (2.5) $\Rightarrow \sum_{t=1}^{T} X_{i,t}\cdot v_t \leq d \;\Rightarrow\; Y_{i,T} = 0$.

$\therefore\; Y_{i,T+1} \geq Y_{i,T}$.

Case (III): when $d'_{i,T} < \sum_{t=1}^{T} X_{i,t} \leq l'_{i,T} + d'_{i,T}$.

From equations (2.4) and (2.5) $\Rightarrow d < \sum_{t=1}^{T} X_{i,t}\cdot v_t \leq d + l$.

$$
\begin{aligned}
\therefore Y_{i,T+1} - Y_{i,T} &= (\sum_{t=1}^{T+1} X_{i,t} \div \sum_{t=1}^{T+1} X_{i,t}\cdot v_t)\times \min\{(\sum_{t=1}^{T+1} X_{i,t}\cdot v_t) - d, l\} \\
&\quad - (\sum_{t=1}^{T} X_{i,t} \div \sum_{t=1}^{T} X_{i,t}\cdot v_t)\times[(\sum_{t=1}^{T} X_{i,t}\cdot v_t) - d] \\
&\geq (\sum_{t=1}^{T+1} X_{i,t} \div \sum_{t=1}^{T+1} X_{i,t}\cdot v_t)\times \min\{(\sum_{t=1}^{T} X_{i,t}\cdot v_t) - d, l\} \\
&\quad - (\sum_{t=1}^{T} X_{i,t} \div \sum_{t=1}^{T} X_{i,t}\cdot v_t)\times[(\sum_{t=1}^{T} X_{i,t}\cdot v_t) - d] \\
&\geq (\sum_{t=1}^{T+1} X_{i,t} \div \sum_{t=1}^{T+1} X_{i,t}\cdot v_t)\times[(\sum_{t=1}^{T} X_{i,t}\cdot v_t) - d] \\
&\quad - (\sum_{t=1}^{T} X_{i,t} \div \sum_{t=1}^{T} X_{i,t}\cdot v_t)\times[(\sum_{t=1}^{T} X_{i,t}\cdot v_t) - d] \\
&\geq 0.
\end{aligned}
$$

Conclusion:

Combining cases (I), (II), and (III), under all situations $Y_{i,T+1} \geq Y_{i,T}$ holds, when the conditions $X_{i,T+1} \geq 0$ and $v_{t+1} \leq v_t \;\forall t \leq T$ can be fulfilled.

Proposition 2: $X_{i,T+1} - Y_{i,T+1} + Y_{i,T} \geq 0$ under the conditions $X_{i,T+1} \geq 0$ and $v_{t+1} \leq v_t \ \forall t \leq T$.

Proof:

First, by using equations (2.4), (2.5), and (2.6):

$$X_{i,T+1} - Y_{i,T+1} + Y_{i,T} = X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times \min\{\max\{(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t) - d, 0\}, l\} + \frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} \times \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}.$$

Next, consider five cases:

Case (I): when $\sum_{t=1}^{T} X_{i,t} \cdot v_t - d < 0$ and $\sum_{t=1}^{T+1} X_{i,t} \cdot v_t - d < l$

$$\begin{aligned} X_{i,T+1} - Y_{i,T+1} + Y_{i,T} &= X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times \max\{(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t) - d, 0\} + 0 \\ &\geq X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times [(\sum_{t=1}^{T} X_{i,t} \cdot v_t - d) + X_{i,T+1} \cdot v_{T+1}] \\ &\geq X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times [X_{i,T+1} \cdot v_{T+1}] \quad (\because \sum_{t=1}^{T} X_{i,t} \cdot v_t - d < 0) \\ &\geq X_{i,T+1} - 1 \times [X_{i,T+1}] \quad (\because v_{t+1} < v_t \quad \forall t) \\ &= 0 \end{aligned}.$$

Case (II): when $\sum_{t=1}^{T} X_{i,t} \cdot v_t - d < 0$ and $\sum_{t=1}^{T+1} X_{i,t} \cdot v_t - d \geq l$.

$$\begin{aligned} X_{i,T+1} - Y_{i,T+1} + Y_{i,T} &= X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times \min\{\max\{(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t) - d, 0\}, l\} + 0 \\ &\geq X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times \max\{(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t) - d, 0\} + 0 \\ &\geq 0 \end{aligned}$$

Case (III): when $\sum_{t=1}^{T} X_{i,t} \cdot v_t - d \geq 0$ and $\sum_{t=1}^{T+1} X_{i,t} \cdot v_t - d < l$.

$$
\begin{aligned}
X_{i,T+1} - Y_{i,T+1} + Y_{i,T} &= X_{i,T+1} - \left(\sum_{t=1}^{T+1} X_{i,t} \div \sum_{t=1}^{T+1} X_{i,t} \cdot v_t\right) \times \left[\left(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t\right) - d\right] \\
&\quad + \left(\sum_{t=1}^{T} X_{i,t} \div \sum_{t=1}^{T} X_{i,t} \cdot v_t\right) \times \left[\left(\sum_{t=1}^{T} X_{i,t} \cdot v_t\right) - d\right] \\
&= X_{i,T+1} - \sum_{t=1}^{T+1} X_{i,t} + \sum_{t=1}^{T} X_{i,t} + d'_{i,T+1} - d'_{i,T} \\
&\geq 0 \quad (\because d'_{i,T+1} \geq d'_{i,T})
\end{aligned}
.
$$

Case (IV): when $0 \leq \sum_{t=1}^{T} X_{i,t} \cdot v_t - d < l$ and $\sum_{t=1}^{T+1} X_{i,t} \cdot v_t - d \geq l$.

$$
\begin{aligned}
X_{i,T+1} - Y_{i,T+1} + Y_{i,T} &= X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times \min\left\{\left(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t\right) - d, l\right\} \\
&\quad + \frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} \times \left[\left(\sum_{t=1}^{T} X_{i,t} \cdot v_t\right) - d\right] \\
&\geq X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times \left[\left(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t\right) - d\right] \\
&\quad + \frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} \times \left[\left(\sum_{t=1}^{T} X_{i,t} \cdot v_t\right) - d\right] \\
&\geq 0
\end{aligned}
.
$$

Case (V): when $\sum_{t=1}^{T} X_{i,t} \cdot v_t - d \geq l$ and $\sum_{t=1}^{T+1} X_{i,t} \cdot v_t - d \geq l$.

$$
\begin{aligned}
X_{i,T+1} - Y_{i,T+1} + Y_{i,T} &= X_{i,T+1} - \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} \times l + \frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} \times l \\
&= X_{i,T+1} - l \times X_{i,T+1} \frac{\sum_{t=1}^{T} X_{i,t} \cdot v_t - v_{T+1} \times \sum_{t=1}^{T} X_{i,t}}{\left(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t\right) \times \left(\sum_{t=1}^{T} X_{i,t} \cdot v_t\right)} \\
&= X_{i,T+1} \times \frac{\left[\left(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t\right) - l\right] \times \left(\sum_{t=1}^{T} X_{i,t} \cdot v_t\right) + l \times v_{T+1} \times \sum_{t=1}^{T} X_{i,t}}{\left(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t\right) \times \left(\sum_{t=1}^{T} X_{i,t} \cdot v_t\right)} \\
&\geq 0 \quad \left(\because \sum_{t=1}^{T+1} X_{i,t} \cdot v_t \geq l + d\right)
\end{aligned}
.
$$

Conclusion:

Combining the five cases: $X_{i,T+1} - Y_{i,T+1} + Y_{i,T} \geq 0$ under the conditions $X_{i,T+1} \geq 0$ and $v_{t+1} \leq v_t \ \forall t \leq T$.

Appendix B : Equal Inflation of Gross Claims, NP Reinsurer's Payments under Indexed Per-Claim Deductible and Limit, and Primary Insurer's Retained Claim

Proposition: at time T for the i^{th} claim, gross claim's inflation equals inflation for NP reinsurer's excess of loss payments and also equals inflation for the primary insurer's retained claim, which is represented by equation (2.11):

$$\frac{1}{w_{i,T}}=\frac{\sum_{t=1}^{T}X_{i,t}}{\sum_{t=1}^{T}X_{i,t}\cdot v_t}=\frac{\min\{\max\{(\sum_{t=1}^{T}X_{i,t})-d'_{i,T},0\},l'_{i,T}\}}{\min\{\max\{(\sum_{t=1}^{T}X_{i,t}\cdot v_t)-d,0\},l\}}=\frac{\min\{(\sum_{t=1}^{T}X_{i,t}),d'_{i,T}\}}{\min\{\sum_{t=1}^{T}X_{i,t}\cdot v_t),d\}}.$$

Proof:

Firstly, the case of $\sum_{t=1}^{T}X_{i,t}\le d'_{i,T}$ (which also implies $\sum_{t=1}^{T}X_{i,t}\cdot v_t\le d$) can be ignored.

Next, when $d'_{i,T}<\sum_{t=1}^{T}X_{i,t}\le l'_{i,T}+d'_{i,T}$ (which also implies $d<\sum_{t=1}^{T}X_{i,t}\cdot v_t\le l+d$.)

Inflation for primary insurer's retained claim = $\dfrac{\min\{(\sum_{t=1}^{T}X_{i,t}),d'_{i,T}\}}{\min\{\sum_{t=1}^{T}X_{i,t}\cdot v_t),d\}}$

$$=\frac{d'_{i,T}}{d}=d\times\frac{\sum_{t=1}^{T}X_{i,t}}{\sum_{t=1}^{T}X_{i,t}\cdot v_t}\times\frac{1}{d}=\frac{1}{w_{i,T}}.$$

Inflation for NP reinsurer's excess of loss payments = $\dfrac{(\sum_{t=1}^{T}X_{i,t})-d'_{i,T}}{(\sum_{t=1}^{T}X_{i,t}\cdot v_t)-d}$

$$=\frac{(\sum_{t=1}^{T}X_{i,t})-d\times(\sum_{t=1}^{T}X_{i,t})\div(\sum_{t=1}^{T}X_{i,t}\cdot v_t)}{(\sum_{t=1}^{T}X_{i,t}\cdot v_t)-d}=\frac{\sum_{t=1}^{T}X_{i,t}}{\sum_{t=1}^{T}X_{i,t}\cdot v_t}=\frac{1}{w_{i,T}}.$$

Finally, when $\sum_{t=1}^{T}X_{i,t}>l'_{i,T}+d'_{i,T}$ (which also implies $\sum_{t=1}^{T}X_{i,t}\cdot v_t>d+l$,)

Inflation for NP reinsurer's excess of loss payments = $\dfrac{l'_{i,T}}{l}=\dfrac{d'_{i,T}}{d}=\dfrac{1}{w_{i,T}}$.

Conclusion:

$$\frac{1}{w_{i,T}}=\frac{\sum_{t=1}^{T}X_{i,t}}{\sum_{t=1}^{T}X_{i,t}\cdot v_t}=\frac{\min\{\max\{(\sum_{t=1}^{T}X_{i,t})-d'_{i,T},0\},l'_{i,T}\}}{\min\{\max\{(\sum_{t=1}^{T}X_{i,t}\cdot v_t)-d,0\},l\}}=\frac{\min\{(\sum_{t=1}^{T}X_{i,t}),d'_{i,T}\}}{\min\{\sum_{t=1}^{T}X_{i,t}\cdot v_t),d\}}.$$

Appendix C: Equal Inflation of Claims Before and After Application of Indexed AAD and AAL with Method 1

Proposition: equal inflation of claims before and after application of indexed AAD and AAL can be represented by the equation in (3.9):

$$\frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}} = \frac{\max\{(\sum_i Y_{i,T}) - L_T',0\}}{\max\{(\sum_i Y_{i,T} \cdot w_{i,T}) - L,0\}} = \frac{\min\{(\sum_i Y_{i,T}), L_T'\}}{\min\{(\sum_i Y_{i,T} \cdot w_{i,T}), L\}}.$$

Proof:

Note that $(\sum_i Y_{i,T}) \leq L_T' \quad \Leftrightarrow \quad (\sum_i Y_{i,T}) \leq L \times \frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}} \quad \Leftrightarrow \quad (\sum_i Y_{i,T} \cdot w_{i,T}) \leq L$.

Case (I): when $(\sum_i Y_{i,T}) \leq L_T'$.

This case can be ignored because the retrocessionaire makes no payment.

Case (II): when $(\sum_i Y_{i,T}) > L_T'$.

From equation (3.8),

$$\frac{\min\{(\sum_i Y_{i,T}), L_T'\}}{\min\{(\sum_i Y_{i,T} \cdot w_{i,T}), L\}} = \frac{L_T'}{L} = L \times \frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}} \times \frac{1}{L} = \frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}}.$$

From equation (3.7),

$$\frac{\max\{(\sum_i Y_{i,T}) - L_T',0\}}{\max\{(\sum_i Y_{i,T} \cdot w_{i,T}) - L,0\}} = \frac{(\sum_i Y_{i,T}) - L \times \frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}}}{(\sum_i Y_{i,T} \cdot w_{i,T}) - L} = \frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}}.$$

Conclusion:

The equality in (3.9) holds:

$$\frac{\sum_i Y_{i,T}}{\sum_i Y_{i,T} \cdot w_{i,T}} = \frac{\max\{(\sum_i Y_{i,T}) - L_T',0\}}{\max\{(\sum_i Y_{i,T} \cdot w_{i,T}) - L,0\}} = \frac{\min\{(\sum_i Y_{i,T}), L_T'\}}{\min\{(\sum_i Y_{i,T} \cdot w_{i,T}), L\}}.$$

Appendix D: Monotonicity Property of Retrocessionaire's Cumulative Payment with Indexed AAD and AAL with Method 1

Proposition: According to section 3.2.2, under two conditions:

(1) $X_{i,T+1} \geq 0 \; \forall i$, and

(2) $v_{t+1} \leq v_t \;\; \forall t \leq T,$

then $S'_{T+1} \geq S'_T$, where $S'_T = \max\{(\sum_i Y_{i,T}) - L'_T, 0\}$ as defined in equation (3.12). S'_T represents retrocessionaire's cumulative payment at time T.

Proof:

First it is to prove three inequalities:

$$\sum_i Y_{i,T+1} \geq \sum_i Y_{i,T} . \quad \text{(D.1)}$$

$$\sum_i Y_{i,T+1} \cdot w_{i,T+1} \geq \sum_i Y_{i,T} \cdot w_{i,T} . \quad \text{(D.2)}$$

$$w_{i,T} \geq w_{i,T+1} \quad \forall i . \quad \text{(D.3)}$$

For (D.1), it has been proved in Appendix A that $Y_{i,T+1} \geq Y_{i,T}$.

Therefore $\sum_i Y_{i,T+1} \geq \sum_i Y_{i,T}$ is trivial.

For (D.2), consider $Y_{i,T+1} \cdot w_{i,T+1} - Y_{i,T} \cdot w_{i,T}$ and refer to equation (3.3):

$$Y_{i,T+1} \cdot w_{i,T+1} - Y_{i,T} \cdot w_{i,T}$$

$$= \min\{\max\{(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t) - d, 0\}, l\} - \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\}$$

$$\geq \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\} - \min\{\max\{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - d, 0\}, l\} = 0.$$

For (D.3), consider $w_{i,T} - w_{i,T+1}$ and refer to equation (2.8):

$$w_{i,T} - w_{i,T+1} = \frac{\sum_{t=1}^{T} X_{i,t} \cdot v_t}{\sum_{t=1}^{T} X_{i,t}} - \frac{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t}{\sum_{t=1}^{T+1} X_{i,t}}$$

$$= \frac{(\sum_{t=1}^{T+1} X_{i,t})(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - (\sum_{t=1}^{T} X_{i,t})(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t)}{(\sum_{t=1}^{T} X_{i,t})(\sum_{t=1}^{T+1} X_{i,t})} .$$

$$= X_{i,T+1} \times \frac{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - (\sum_{t=1}^{T} X_{i,t} \cdot v_{T+1})}{(\sum_{t=1}^{T} X_{i,t})(\sum_{t=1}^{T+1} X_{i,t})}$$

$$\geq X_{i,T+1} \times \frac{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - (\sum_{t=1}^{T} X_{i,t} \cdot v_t)}{(\sum_{t=1}^{T} X_{i,t})(\sum_{t=1}^{T+1} X_{i,t})} = 0.$$

Next for proving $S'_{T+1} \geq S'_T$, consider both cases of $L'_{T+1} \leq L'_T$ and $L'_{T+1} > L'_T$:

Case (I): when $L'_{T+1} \leq L'_T$.

$S'_{T+1} - S'_T = \max\{(\sum_i Y_{i,T+1}) - L'_{T+1}, 0\} - \max\{(\sum_i Y_{i,T}) - L'_T, 0\}$ from equation (3.12)

$\geq \max\{(\sum_i Y_{i,T+1}) - L'_T, 0\} - \max\{(\sum_i Y_{i,T}) - L'_T, 0\}$

$\geq \max\{(\sum_i Y_{i,T}) - L'_T, 0\} - \max\{(\sum_i Y_{i,T}) - L'_T, 0\}$ from equation (D.1)

$= 0$

Case (II): when $L'_{T+1} > L'_T$ and $S'_T = 0$

$S'_{T+1} - S'_T \geq 0$ is trivial to prove

Case (III): when $L'_{T+1} > L'_T$ and $S'_T > 0$

$S'_{T+1} - S'_T = \max\{(\sum_i Y_{i,T+1}) - L'_{T+1}, 0\} - [(\sum_i Y_{i,T}) - L'_T]$ from equation (3.12)

$\geq (\sum_i Y_{i,T+1}) - L'_{T+1} - (\sum_i Y_{i,T}) + L'_T$

$= (\frac{L'_{T+1}}{L}) \times (\sum_i Y_{i,T+1} \cdot w_{i,T+1} - L) - (\frac{L'_T}{L}) \times (\sum_i Y_{i,T} \cdot w_{i,T} - L)$ from equation (3.4)

$\geq (\frac{L'_{T+1}}{L}) \times (\sum_i Y_{i,T} \cdot w_{i,T} - L) - (\frac{L'_T}{L}) \times (\sum_i Y_{i,T} \cdot w_{i,T} - L)$ from equation (D.2)

≥ 0 since $L'_{T+1} > L'_T$.

Conclusion:

Combining cases (I), (II), and (III), under all situations $S'_{T+1} \geq S'_T$ holds, when the conditions $v_{t+1} \leq v_t$ $\forall t \leq T$ and $X_{i,T+1} \geq 0$ $\forall i$ can be fulfilled.

Appendix E: Monotonicity Properties of Indexed Per-Claim Deductible and Limit

Proposition: According to section 3.2.4, under two conditions:

(1) $X_{i,T+1} \geq 0 \; \forall i$, and

(2) $v_{t+1} \leq v_t \; \forall t \leq T$

then $d'_{i,T+1} \geq d'_{i,T}$ and $l'_{i,T+1} \geq l'_{i,T}$ with $d'_{i,T}$ and $l'_{i,T}$ as defined in equations (2.4), (2.5)

Proof of: $d'_{i,T+1} \geq d'_{i,T}$ by considering $(d'_{i,T+1} - d'_{i,T}) \div d$

$$\frac{d'_{i,T+1} - d'_{i,T}}{d} = \frac{\sum_{t=1}^{T+1} X_{i,t}}{\sum_{t=1}^{T+1} X_{i,t} \cdot v_t} - \frac{\sum_{t=1}^{T} X_{i,t}}{\sum_{t=1}^{T} X_{i,t} \cdot v_t} \qquad \text{from equation (2.4)}$$

$$= X_{i,T+1} \times \frac{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - (\sum_{t=1}^{T} X_{i,t} \cdot v_{T+1})}{(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t) \times (\sum_{t=1}^{T} X_{i,t} \cdot v_t)}$$

$$\geq X_{i,T+1} \times \frac{(\sum_{t=1}^{T} X_{i,t} \cdot v_t) - (\sum_{t=1}^{T} X_{i,t} \cdot v_t)}{(\sum_{t=1}^{T+1} X_{i,t} \cdot v_t) \times (\sum_{t=1}^{T} X_{i,t} \cdot v_t)} = 0.$$

Similar logic can be applied for proving $l'_{i,T+1} \geq l'_{i,T}$.

REFERENCES

[1] Ferguson, R., "Nonproportional Reinsurance and the Index Clause," *PCAS* **1974**, Vol. LXI, pp. 141-164.
[2] Cook, C., "Discussion of Ferguson: 'Nonproportional Reinsurance and the Index Clause,'" *PCAS* **1974**, Vol. LXI, pp. 165-169
[3] Khury, C.K., "Discussion of Ferguson: 'Nonproportional Reinsurance and the Index Clause,'" *PCAS* **1975**, Vol. LXII, pp. 9-10.
[4] Levin, J., "Discussion of Ferguson: 'Nonproportional Reinsurance and the Index Clause,'" *PCAS* **1975**, Vol. LXII, pp. 1-4.
[5] Rodermund, M., "Discussion of Ferguson: 'Nonproportional Reinsurance and the Index Clause,'" *PCAS* **1975**, Vol. LXII, pp. 5-8.
[6] Feldblum, S., "Indexed Retentions and Multiple Claimant Losses in Nonproportional Reinsurance," *CAS Forum*, **1987** Fall, pp. 3-6.
[7] Fisher, R., "Letter to Mr. Feldblum from Mr. Fisher Dated 4/21/87," *CAS Forum*, **1987** Fall, pp. 7-10.
[8] Feldblum, S., "Letter to Mr. Fisher from Mr. Feldblum Dated 4/27/87," *CAS Forum*, **1987** Fall, pp. 11-15.

Abbreviations and notations

AAD, annual aggregate deductible
AAL, annual aggregate limit
AARL, annual aggregate reinstatement limit
NP, non proportional
XS, excess of
GNPI, gross net premium income

Biography of the Author

Ka Chun Yeung is senior pricing actuary at Gen Re in Sydney. He has worked in Gen Re's Hong Kong and Cologne offices before taking up his role in the Sydney P&C treaty team. He received his BS in Actuarial Science from the University of Hong Kong. He is a Fellow of the Casualty Actuarial Society, Fellow of the Institute of Actuaries of Australia, and RMS Certified Catastrophe Risk Analyst. He is a member of the CAS Committee on Reinsurance Research. The opinions expressed here are those of the author and not necessarily Gen Re.

Ka Chun Yeung
Senior Pricing Actuary
Gen Re
Level 24, 123 Pitt Street, Sydney NSW 2000 Australia
Email: kyeung@genre.com
Email: kacyeung@graduate.hku.hk

The Conditional Validity of Risk-Adjusted Discounting

Leigh J. Halliwell, FCAS, MAAA

Abstract: The constellation of the initiatives of ERM, Solvency II, and International Accounting traces back through capital management to modern finance and portfolio theory. These supposedly dynamic and market-oriented initiatives will eventually disappoint the (re)insurance industry, if they uncritically endorse risk-adjusted discounting. One's job is rendered more difficult, if not impossible, without the right tools. Building on earlier papers, the author will here show how a seminal academic paper from the 1960s contains the seeds of the downfall of risk-adjusted discounting. It is too much to expect a retraction, but hopefully, the emerging standards for these initiatives at least will not force risk-adjusted discounting upon the practitioners.

Keywords: present value, risk-adjusted discounting, stochastic cash flow.

1. INTRODUCTION

One standard textbook begins with this clear pronouncement about risk-adjusted discounting:

> To calculate present value, we discount expected payoffs by the rate of return offered by the equivalent investment alternative in the capital market. The rate of return is often referred to as the discount rate, hurdle rate, or opportunity cost of capital. [Brealey and Myers, 2002, p 15]

Having published several critiques of this principle,[1] we have challenged others to show where in the academic literature it has been rigorously derived. At length, someone directed our attention to Robichek and Myers [1966], whose co-author, Stewart C. Myers, is the same as the co-author of the textbook just cited.[2] To our surprise, this brief paper, far from deriving the principle, actually points out its "conceptual problems," as its title reveals. We are wholly in accord with its second paragraph:

> Since time and risk are logically separate variables, summing up their effects in the one number k requires a particular assumption about the actual relationship between the effects of time and risk on present value. The main purpose of this communication is to uncover this assumption and to point out that valuation errors may result if the risk-adjusted discount rate is used when this assumption does not hold.

After treating a simple example, its authors conclude:

> … the general conclusion [is] that the rate at which income is expected to be realized over time depends on the rate at which uncertainty is expected to be resolved over time. If uncertainly is

[1] See the author's publications in the References, especially Halliwell [2003], Appendix A.

[2] He is also the co-inventor of the Myers-Cohn insurance-pricing model, which has been used in Massachusetts rate hearings. Introductions to this model may be found in D'Arcy and Doherty [1988], Mahler [1998], and D'Arcy [1999].

expected to be resolved at a constant rate over time, then the required rate of return k predicts accurately the rate at which income is expected to be realized. But this need not always be the case.

To this conclusion Philbrick [1994] agrees. But if risk-adjusted discounted depends on a certain resolution of uncertainty, how should one value a project (i.e., a stochastic cash flow) whose uncertainty resolves otherwise?[3] But putting this aside for now, we will test our theory, deriving risk-adjusted discounting from it in the case of a dividend-paying stock on the assumption of continuous risk resolution. First, we value the stock according to the prevailing theory.

2. THE DIVIDEND-GROWTH MODEL

The dividend-growth model (also called the Gordon, or Gordon-Shapiro, model) is the commonly accepted method of valuing a dividend-paying stock (Bowlin [1990, pp 96f], Brealey and Myers [2002, Chapter 4]). Because we will deal with continuous risk resolution, we will formulate a continuous version of this model. At time *t* the stock is *expected* to pay out a dividend at the rate:

$$dC(t) = \mu_0 e^{\gamma t} dt$$

We use '*C*' for cash instead of '*D*' for dividend, to avoid confusion with the differential '*d*' and the force of interest 'δ'. μ_0 is the instantaneous dividend flow (in units of currency per time) at time zero, from which it is expected continuously to grow at rate γ (in units of time^{-1}). We will discount this expected dividend stream at the cost of capital κ, so the discount function is $e^{-\kappa t}$. Hence, according to the principle of risk-adjusted discounting, the value of the stock at time *t* is:

$$V(t) = \int_{u=t}^{\infty} e^{-\kappa(u-t)} dC(u) = \int_{u=t}^{\infty} e^{-\kappa(u-t)} \mu_0 e^{\gamma u} du$$

Then we work out the integral:

[3] Halliwell [2001, Appendix D] shows that an asset should appreciate at a risk-free rate while uncertainly is not resolving, or more accurately "the price of an asset whose uncertainty is not changing remains proportional to the price of an asset whose future payment is certain."

$$V(t) = \int_{u=t}^{\infty} e^{-\kappa(u-t)} \mu_0 e^{\gamma u} du = \mu_0 \int_{u=t}^{\infty} e^{-\kappa(u-t)} e^{\gamma(u-t)} du \cdot e^{\gamma t}$$

$$= \mu_0 \int_{u=t}^{\infty} e^{-(\kappa-\gamma)(u-t)} d(u-t) \cdot e^{\gamma t} = \mu_0 \int_{v=0}^{\infty} e^{-(\kappa-\gamma)v} dv \cdot e^{\gamma t}$$

$$= \frac{\mu_0}{\kappa-\gamma} e^{\gamma t}$$

Of course, for the integral to converge, the cost of capital must exceed the dividend growth rate, or $\kappa > \gamma$.

What is the instantaneous (expected) total return on the stock at time *t*, which we will call $\rho(t)$? It must consider both the dividend and the price appreciation. Therefore:

$$\rho(t) = \frac{1}{V(t)} \lim_{\Delta t \to 0} \frac{\Delta C(t) + \Delta V(t)}{\Delta t} = \frac{1}{V(t)} \frac{dC(t)}{dt} + \frac{1}{V(t)} \frac{dV(t)}{dt}$$

The first term is the instantaneous dividend yield, which we will call *y*, and the second is the instantaneous price appreciation. Simplifying to the utmost, we have:

$$\rho(t) = \frac{1}{V(t)} \frac{dC(t)}{dt} + \frac{1}{V(t)} \frac{dV(t)}{dt}$$

$$= \frac{\mu_0 e^{\gamma t}}{\left(\frac{\mu_0}{\kappa-\gamma} e^{\gamma t}\right)} + \frac{\frac{\mu_0}{\kappa-\gamma} e^{\gamma t} \gamma}{\left(\frac{\mu_0}{\kappa-\gamma} e^{\gamma t}\right)}$$

$$= (\kappa-\gamma) + \gamma$$

$$= \kappa$$

Accordingly, the total return is the dividend yield ($y = \kappa - \gamma$) plus the rate of appreciation γ.[4]

[4] Halliwell [1999, p 412]: '[A cash flow] is always earning its cost of capital ϱ, or working at "ϱ-power."'

3. THE ALTERNATIVE MODEL: STOCHASTIC CASH FLOWS

Next, we value the stock with our stochastic-cash-flow theory. The dividend stream we now regard as the Wiener process $dC(t) = \mu_0 e^{\gamma t} dt + \sigma_0 e^{\gamma t} dX(t)$, whose *mean* equals the $dC(t)$ of the dividend-growth model. [5] Although the drift and volatility functions are exponential in t, this equation represents arithmetic Brownian motion, rather than geometric. The arithmetic form allows for us to discount the increments and for their sum to be normally distributed.

Now $dC(t)$, the *actual* dividend received during interval $[t, t+dt]$, is normally distributed with mean $\mu_0 e^{\gamma t} dt$ and variance $\sigma_0^2 e^{2\gamma t} dt$, where the dimension of σ_0^2 is currency squared per time. If we let δ represent a flat and persistent risk-free force of interest, the discount function is $v(t) = e^{-\delta t}$. Hence, the present value of the stock at time t, which, we maintain [Halliwell, 2003, Section 3], should be considered as a random variable, is:

$$PV[C(t)] = \int_{u=t}^{\infty} v(u-t) dC(u) = \int_{u=t}^{\infty} e^{-\delta(u-t)} dC(u)$$

The discounted dividend received during interval $[u, u+du]$ from the standpoint of time t is normally distributed with mean $e^{-\delta(u-t)} \mu_0 e^{\gamma u} du$ and variance $e^{-2\delta(u-t)} \sigma_0^2 e^{2\gamma u} du$. Therefore, PV is normally distributed with mean:

$$\begin{aligned} E[PV[C(t)]] &= \int_{u=t}^{\infty} e^{-\delta(u-t)} \mu_0 e^{\gamma u} du = \mu_0 \int_{u=t}^{\infty} e^{-\delta(u-t)} e^{\gamma(u-t)} du \cdot e^{\gamma t} \\ &= \mu_0 \int_{u=t}^{\infty} e^{-(\delta-\gamma)(u-t)} d(u-t) \cdot e^{\gamma t} = \mu_0 \int_{v=0}^{\infty} e^{-(\delta-\gamma)v} dv \cdot e^{\gamma t} \\ &= \frac{\mu_0}{\delta - \gamma} e^{\gamma t} \end{aligned}$$

This agrees with the dividend-growth formula, except that δ takes the place of κ. Again, for

[5] We will not need the stochastic calculus, but introductions to it can be found in Wilmott [1995, pp 20-29] and Panjer [1998, Section 10.13].

convergence, $\delta > \gamma$.[6] Similarly, and due to the independence of the $dC(t)$, the variance of *PV* is:

$$Var[PV[C(t)]] = \int_{u=t}^{\infty} e^{-2\delta(u-t)} \sigma_0^2 e^{2\gamma u} du = \sigma_0^2 \int_{u=t}^{\infty} e^{-2\delta(u-t)} e^{2\gamma(u-t)} du \cdot e^{2\gamma t}$$

$$= \sigma_0^2 \int_{u=t}^{\infty} e^{-2(\delta-\gamma)(u-t)} d(u-t) \cdot e^{2\gamma t} = \sigma_0^2 \int_{v=0}^{\infty} e^{-2(\delta-\gamma)v} dv \cdot e^{2\gamma t}$$

$$= \frac{\sigma_0^2}{2(\delta-\gamma)} e^{2\gamma t}$$

So finally, the present value of the stochastic dividend flow at time *t* is a normal random variable with mean $\frac{\mu_0}{\delta-\gamma} e^{\gamma t}$ and variance $\frac{\sigma_0^2}{2(\delta-\gamma)} e^{2\gamma t}$.

We showed (Halliwell [2003, p. 66]) that the price of a quantum, or stand-alone, $N(\mu, \sigma^2)$ present-valued stochastic cash flow *X* is $q_X = \mu_X - a\sigma_X^2$ for an economic agent whose risk-aversion parameter is *a*. Accordingly, at time *t* such an agent will value the stock as:

$$V(t) = \frac{\mu_0}{\delta-\gamma} e^{\gamma t} - a(t) \frac{\sigma_0^2}{2(\delta-\gamma)} e^{2\gamma t}$$

[6] It augurs well for our treatment of "time and risk [as] logically separate variables," as quoted above from Robichek and Myers, that it places a realistic constraint on growth, viz., that perpetual growth must be less than risk-free growth, a constraint that risk-adjusted discounting does not impose.

Furthermore, we argued elsewhere [Halliwell, 2001, Section 5 and Appendix D] that the product of one's risk aversion and expected wealth should remain constant. In this stand-alone realm, in which expected wealth is increasing by a factor of $e^{\gamma t}$, $a(t) = a_0 e^{-\gamma t}$. Therefore:

$$V(t) = \frac{\mu_0}{\delta - \gamma} e^{\gamma t} - a_0 e^{-\gamma t} \frac{\sigma_0^2}{2(\delta - \gamma)} e^{2\gamma t} = \left(\frac{\mu_0}{\delta - \gamma} - a_0 \frac{\sigma_0^2}{2(\delta - \gamma)} \right) e^{\gamma t} = V_0 e^{\gamma t}$$

The agent, in addition to receiving the dividend, will receive price appreciation at rate γ, as happens also according to the dividend-growth model.[7]

For the dividend-growth model and our theory to agree, the valuations must be equal, i.e., $\frac{\mu_0}{\kappa - \gamma} e^{\gamma t} = \left(\frac{\mu_0}{\delta - \gamma} - a_0 \frac{\sigma_0^2}{2(\delta - \gamma)} \right) e^{\gamma t}$, or $\frac{\mu_0}{\kappa - \gamma} = \frac{\mu_0}{\delta - \gamma} - a_0 \frac{\sigma_0^2}{2(\delta - \gamma)}$. Hence:

$$\begin{aligned} \kappa &= (\kappa - \gamma) + \gamma \\ &= \frac{\mu_0}{\left(\frac{\mu_0}{\kappa - \gamma} \right)} + \gamma \\ &= \frac{\mu_0}{\left(\frac{\mu_0}{\delta - \gamma} - a_0 \frac{\sigma_0^2}{2(\delta - \gamma)} \right)} + \gamma \end{aligned}$$

Therefore, there is a number κ defined in terms of the parameters of the stochastic-cash-flow model (i.e., in terms of μ_0, a_0, σ_0, δ, and γ) at which one can discount the expected dividend stream and arrive at the "correct" value. Furthermore, we can give a simple and meaningful interpretation to the right side of the last equation. The expected instantaneous dividend yield of the stock at time t, according to the stochastic theory, is:

[7] As a check, $V(t)$ decreases with increasing a. Since dividends can be negative in arithmetic Brownian motion, sufficient risk aversion will make $V(t)$ negative. In the case of risk-neutrality, when $a = 0$, $V(t) = E[PV[C(t)]]$.

$$
\begin{aligned}
E[y(t)] &= \frac{1}{V(t)} E\left[\frac{dC(t)}{dt}\right] \\
&= \frac{1}{V(t)} \frac{dE[C(t)]}{dt} \\
&= \frac{E\left[\mu_0 e^{\gamma t} dt + \sigma_0 e^{\gamma t} dX(t)\right] / dt}{\left(\frac{\mu_0}{\delta - \gamma} - a_0 \frac{\sigma_0^2}{2(\delta - \gamma)}\right) e^{\gamma t}} \\
&= \frac{\mu_0 e^{\gamma t} dt / dt}{\left(\frac{\mu_0}{\delta - \gamma} - a_0 \frac{\sigma_0^2}{2(\delta - \gamma)}\right) e^{\gamma t}} \\
&= \frac{\mu_0}{\left(\frac{\mu_0}{\delta - \gamma} - a_0 \frac{\sigma_0^2}{2(\delta - \gamma)}\right)}
\end{aligned}
$$

Obviously, since both the expected dividend and the price are growing at rate γ, the expected dividend rate $E[y\{t)]$ is constant, or just $E[y]$. And this allows us to see that the expected total return, which Brealey and Myers call *inter alia* the cost of capital, is:

$$
\begin{aligned}
\kappa &= \frac{\mu_0}{\left(\frac{\mu_0}{\delta - \gamma} - a_0 \frac{\sigma_0^2}{2(\delta - \gamma)}\right)} + \gamma \\
&= E[y] + \gamma
\end{aligned}
$$

Therefore, risk-adjusted discounting is a special case of our theory. It is approximately correct, even as Newtonian mechanics is approximately correct vis-à-vis Special Relativity (and would be exactly correct, if the speed of light were infinite). The trouble is that the approximation is taken for the truth.

4. CONCLUSION

Therefore, risk-adjusted discounting is conditionally valid, sc., valid on the condition that the stochastic cash flow is continuously replicating itself on an exponentially-increasing scale.[8] To the reader we will leave to decide how often this condition applies to actual financial decisions, particularly to underwriting decisions. For some to counter that in the grand scheme every risk is but a drop in the ocean is as specious as for actuaries to argue from the central limit theorem that every distribution may be deemed normal. Robichek and Myers correctly state that "time and risk are logically separate variables."[9] We've all heard of a distinction without a difference. However, their claim that "valuation errors may result if … this assumption does not hold"[10] implies that this is one distinction that does make a difference.

[8] One might also add as ancillary conditions the independence of the flow from the (rest of) the agent's stochastic wealth, and the flatness of the yield curve.

[9] Even here, the adverb 'logically' is timid; time and risk are truly separate variables. So far as we know, the separation, or distinction, of time and risk is a basic principle only in Van Slyke [1995 and 1999] and Schnapp [2001]. Mango [2003] is ambiguous. However, agreement on this principle does not ensure agreement *in toto*. Van Slyke, in particular, urges that capital markets can and do synthesize the views of their participants into higher truths, a belief to which we do not subscribe. Nevertheless, Van Slyke [1995, p 587] is correct in rating the effect on finance of this principle as nothing short of revolutionary. Terms such as 'radical' and 'revolutionary' are bandied and overused; however, we regard this as much a revolution as the Copernican, which took a century finally to be settled. As the evidence mounted for heliocentrism, the old guard must have resorted to ploys like "Geo or helio, what's the difference? The day looks the same, anyway." So it is today in financial theory. But when the camel's nose gets under the tent, soon it will be overturned.

[10] Here again (see previous footnote), the auxiliary verb 'may' tones down. More accurately, valuation errors *will* result if the assumption does not hold, and it's just a matter of how serious these errors *may* be."

REFERENCES

[1.] Bowlin, Oswald D., Martin, John D., and Scott, David F., *Financial Analysis*, Second Edition, New York, McGraw-Hill, 1990.

[2.] Brealey, Richard A., and Myers, Stewart, *Principles of Corporate Finance*, Seventh Edition, New York, McGraw-Hill, 2002.

[3.] D'Arcy, Stephen P., and Doherty, Neil A., *The Financial Theory of Pricing Property-Liability Insurance Contracts*, University of Pennsylvania, 1988.

[4.] D'Arcy, Stephen P., "Introduction to the Discounted Cash Flow Approach," *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, Casualty Actuarial Society, 1999, 19-25.

[5.] Halliwell, Leigh J., "Insights into Present Value and Duration," *CAS Forum* (Spring 1999), 407-437, www.casact.org/pubs/forum/99spforum/99spf409.pdf.

[6.] "A Critique of Risk-Adjusted Discounting," *2001 ASTIN Colloquium*, www.actuaries.org/ASTIN/Colloquia/Washington/Halliwell.pdf.

[7.] "The Valuation of Stochastic Cash Flows," *CAS Forum* (Reinsurance Call Papers, Spring 2003), 1-68, www.casact.org/pubs/forum/03spforum/03spf001.pdf.

[8.] "Valuing Stochastic Cash Flows: A Thought Experiment," *CAS Forum* (Winter 2004), 291-294, www.casact.org/pubs/forum/04wforum/04wf291.pdf.

[9.] Mahler, Howard C., "The Myers-Cohn Profit Model, A Practical Application," PCAS LXXXV (1998), 698-774, www.casact.org/pubs/proceed/proceed98/980689.pdf.

[10.] Mango, Donald, "Capital Consumption: An Alternative Methodology for Pricing Reinsurance," *CAS Forum* (Winter 2003), 351-378, www.casact.org/pubs/forum/03wforum/03wf351.pdf.

[11.] Panjer, Harry H., editor, *Financial Economics: with Applications to Investments, Insurance and Pensions*, Schaumburg, IL, Actuarial Foundation, 1998.

[12.] Philbrick, Stephen W., "Accounting for Risk Margins," *CAS Forum* (Spring 1994), 1-87, www.casact.org/pubs/forum/94spforum/94spf001.pdf.

[13.] Robichek, Alexander A., and Myers, Stewart C., "Conceptual Problems in the Use of Risk-Adjusted Discount Rates," *Journal of Finance*, Vol. 21, No. 4 (December 1966), 727-730.

[14.] Schnapp, Frank, "The Diversification Property," *Presented to the 11th AFIR Colloquium*, September 2001, www.actuaries.ca/meetings/afir/handouts.

[15.] Van Slyke, Oakley E., "A General Theory of Finance," *Proceedings of the 5th AFIR Colloquium*, September 1995, vol. II, 585-635, www.actuaries.org/AFIR/colloquia/Brussels/VanSlyke.pdf.

[16.] Van Slyke, Oakley E., "The Cost of Capital: An Axiomatic Approach," *Actuarial Considerations Regarding Risk and Return in Property-Casualty Insurance Pricing*, Casualty Actuarial Society, 1999, 135-164.

[17.] Wilmott, Paul, Howison, Sam, and Dewynne, Jeff, *The Mathematics of Financial Derivatives: A Student Introduction*, Cambridge University Press, 1995.