Implications of Reinsurance and Reserves on Risk of Investment Asset Allocation

by Gary G. Venter, FCAS, MAAA,
John W. Gradwell, ACAS, MAAA,
Mohammed Q. Ashab, ACAS, MAAA, and
Alex Bushel

Implications of Reinsurance and Reserves on Risk of Investment Asset Allocation

Gary Venter, John Gradwell, Mohammed Ashab, Alex Bushel

Sedgwick Re Insurance Strategy, Inc.

Abstract

DFA makes possible a greater integration of asset management with underwriting management. This paper looks at how investment risk and reinsurance can affect the overall risk to the company, and how the two can be managed simultaneously. A significant underwriting variable is the risk of loss development, and models of the development risk are presented, with some methodology for determining which one is most appropriate given the data at hand. Term-structure models are key to asset risk modeling, and a test of these models is proposed.

Implications of Reinsurance and Reserves on Risk of Investment Asset Allocation

Asset-Liability Management

Property- liability insurers have traditionally managed investment and underwriting functions separately, except for some forays into duration matching and perhaps to set goals for their investment risk that recognize that they do have some underwriting exposure. Dynamic Financial Analysis (DFA), by jointly modeling asset and liability risks, provides a means to more closely integrate the management of investment and underwriting risk, and thereby directly manage the total risk of bottom line results. This paper will focus on modeling GAAP pre-tax surplus change, which includes the effect of unrealized gains and losses, but any income statement or balance sheet item could be modeled similarly.

The principal risk elements to pre-tax surplus change are asset risk, reserve development, and current year underwriting results. These each have separate modules in the model described below, but some common economic elements, such as inflation and interest rates, feed all the modules.

Looking at assets alone, higher yielding assets generally bear more risk of adverse deviation, with short-term treasury securities usually regarded as having the least risk and least expected return. However adding liabilities - even fixed liabilities - to this picture changes the risk profile. If liabilities are of medium term, then holding short-term assets could be of higher risk, as interest rates may decrease and generate less than enough investment income to cover the liabilities. Long-term investment also increases in risk in this case, as interest rates could go up, requiring liquidation of depressed assets to meet the liabilities. Long-term investments may still have higher expected returns than medium term, but the insurer with medium-term liabilities will be exposed to more risk than the asset-only investor for using those instruments. On the other hand, medium-term assets could be carried at a greater reduction in risk than for the usual investor in this case. This is the rationale for duration matching. Uncertain liabilities and payout timing complicates the matching process, and can render perfect matching impossible. Simulation of loss payment requirements against asset fluctuations can be used to estimate the risk of different investment strategies in this case.

But the real world keeps intruding: if a company with medium or long-term liabilities grows just with inflation, it tends to have positive cash flow. If positive cash flow were a certainty, assets would never have to be liquidated to pay liabilities; the risk-return situation reverts back to the asset-only situation. Add to this accounting for bonds at amortized values and long-term investments suddenly become low-risk high-return opportunities. In this paper bonds will be evaluated at market, which records more risk for long-term bonds, but the same approach could work with amortized costs - with different results expected.

It is when cash flow is also risky that the DFA approach to asset/liability management really shows its merits. Without the shield of reliable positive cash flows, the uncertainty about interest rates and loss payout requirements are back, complicated by the fact that cash flows will often but not always be positive. All of these elements can be simulated simultaneously to quantify their interactions. This would allow the measurement of the effect of different reinsurance strategies, through their impacts on cash flow, on the combined asset/liability risk. Inflation can affect both asset values, through the interest rates, as well as premium volume and loss payments, and so its impact is complex. Reserves may be inflation sensitive as well, which would add yet another impact on the surplus change. All of these effects can be captured using a DFA approach to asset-liability management.

MODELING ISSUES

SCENARIOS AND PROBABILITY

Prior to DFA modeling, risk was measured by scenario testing. A few scenarios were selected and the financial outcomes under those scenarios were computed. This enabled management to have some confidence that their strategies would bear up under various sorts of adverse developments. It did not, however, allow for an assessment of the probability of achieving various earnings targets. Without knowing the probabilities of the various scenarios arising, management could have been sacrificing overall profitability to guard against some exceedingly rare eventualities.

DFA can do more than merely increase the number of scenarios tested. With good models of the underlying processes it can generate a set of scenarios that in some sense reflects the probability of occurrence of the various outcomes. There are of course issues of how well the model represents the processes being modeled - there is both art and science to modeling. The criterion to which a model should be judged is not its ability to generate a wide variety of scenarios, but rather its ability to generate scenarios according to their likelihood of occurrence.

ASSET MODELS

The asset modeling approach adopted here is to first generate a series of treasury yield curves using diffusion models. This is detailed in Appendix 1. Many other economic variables, such as the inflation rate and security prices, have historically correlated to the current and past yield curves, so these variables can be modeled by regression and simulated from the regression models and the simulated yield curves. This builds in the correlations among these variables with appropriate levels of random fluctuation.

A portfolio of assets and liabilities is subject to risk from complex changes to the shape of the yield curve – not just simple upward and downward movements. Thus a yield curve model has to be able to generate curves of different shapes, and in accordance with the probability that they might arise. In Appendix 1 we introduce measures of yield curve shapes and compare some yield curve models and historical data as to the distribution of the shape of the yield curve conditional on the short term rate. It is shown there that some yield curve models, although they can generate yield curves of different shapes, tend to generate only very restricted shapes of yield curves for any given short-term rate. This is not consistent with historical data, and so those models could not be expected to produce scenarios in accord with occurrence probabilities.

Reserve Development Models

Many different assumptions can be made about the processes that generate loss development. In Appendix 2 a classification scheme is outlined that groups reserve development processes into 64 different classes. This is based on answering 6 yes-no questions about the development process. Empirical methods of answering these questions based on triangulated data are also discussed. Once a process is identified that plausibly could have generated the loss triangle in question, this process can be used to simulate scenarios of future development. Doing this study has implications for loss reserving as well, as each process of generating loss emergence implies a reserving methodology. The implied methodology is essentially the one that provides the best estimates of the parameters of the process that is generating the development, and is explored in Appendix 2.

In the examples below it is assumed that this study has been completed, and only two of these classes of processes are illustrated. The first starts by generating ultimate losses, and then uses random draws around expected percentages of payment to generate the paid losses at each age. This is essentially the process used by Stanard (1985) to testing development methods through loss simulation. It turns out that a parameterized form of the Bornheutter-Ferguson method is optimal for this process.

The second process is similar, but the paid losses at each age are then adjusted up or down by the difference between actual and expected reserve inflation. In this case the paid losses in each year will depend on the inflation for the year, and the final ultimate losses will end up different from the initial ultimate originally drawn. That sounds like a more realistic process for the generation of actual loss histories, but empirical tests of loss development do not always identify an effect of post-event inflation. If losses are sensitive to inflation after the loss date, the risks to holding a given set of assets will be different from what they would be otherwise. The optimal reserving method in this case involves estimating the impact of calendar-year inflation (i.e., diagonal trend) on the loss triangle.

Mack (1994) showed that the chain ladder is optimal for the process that generates each age's emerged loss as a factor times the cumulative emerged-to-date for the accident year, plus a random element. This process could be used to generate losses in a DFA model, but it is not illustrated here.

Underwriting Risk Models

Models of current year underwriting risk can be intricate, but are usually straightforward. The approach here is to simulate large individual losses from models of frequency, severity, and parameter uncertainty and smaller losses in the aggregate from a single aggregate distribution for each line. Then the difference between simulated and expected inflation is applied, followed by application of the reinsurance program.

SIMULATION ASSUMPTIONS

COMMON ASSUMPTIONS

A few simplifying assumptions will be made in all the simulations in order to highlight the essential elements being tested. These are not intrinsic to the model, however. First, it will be assumed that all cash flows take place at year-end or an infinitesimal time later at the beginning of the next year. Thus premiums are all written, expenses are paid out, and the remaining unearned premiums are invested at this instant. A year later the payments to be made for losses for that year and all previous accident years are paid out, any bonds mature, coupon payments are made, etc. All losses are assumed to pay out over a 10-year period with an average payout lag of three years after policy issuance, but the actual payout pattern may be stochastic. The following investment strategies will be tested: short term - everything is in one-year treasuries; medium term - all in three-year treasuries; long term - all in ten-year treasuries; and stocks plus - 50% in stocks and 50% in ten-year treasuries. Surplus is assumed to be one-fourth of assets.

COMPANY RISK FACTORS

Several different hypothetical companies will be simulated to test how various underwriting risks interact with the investment scenarios above. The first will be a what-if test of surplus only – the reserves and other assets are ignored. The second will assume the company has a fixed known payout pattern – i.e., no reserve risk. The third will be a company with stochastic reserves – there is a distribution around each payout – but with no inflation risk – the payouts have a random element but not correlated with inflation. Fourthly, the payouts will be assumed correlated with inflation. In this case the reserves will be adjusted at year-end by the ratio of the actual to expected inflation factor. All these tests will be based on a reinsurance program with a moderately high retention. The final test will repeat the fourth with a more conservative approach to reinsurance.

A Few Details

For each set of company risk assumptions and each investment strategy, the distribution of year-end pre-tax GAAP surplus is simulated. Comparisons are made of the mean, standard deviation, and 99^{th}, 90^{th}, 10^{th} and 1st percentile of each distribution. These percentiles correspond to the upper and lower 1-in-10 and 1-in-100 probability of exceeding levels.

The strategies and risk profiles tested below are not completely realistic. They are intended to illustrate the capabilities of DFA modeling in the asset-liability management arena, and the interaction of that with reserving and reinsurance. Because of this and for the sake of simplicity, the CIR (Cox, Ingersoll and Ross) model from Appendix 1 is used for the examples, but with different parameters. The initial short-term interest rate r is assumed to be 0.05, and its change is generated by the following process:

$$dr = 0.2(0.06 - r)dt + 0.075r^{1/2}dz.$$

The CPI and Wilshire 5000 stock index are simulated as measures of inflation and stock market performance. These are generated by regression on the yield curve and lags of the yield curve. The regressions were done on quarterly data, so for notational purposes the time periods will be expressed as quarters. Notation such as 3L40:12 will denote the third lag of the difference between the 40 quarter and 12 quarter interest rates, i.e., the 10 year rate less the 3 year rate seen 9 months ago. Without the colon 0L40 is just the 10 year rate for the current quarter.

The inflation variable estimated here, denoted qccpi, is the ratio of the CPI for a quarter to that for the previous quarter. The variables used in the fit along with indications of their significance are shown in the table below. The data used is from the fourth quarter of 1959 to first quarter 1997, as this was available from pointers within the CAS website.

Change in CPI

Variable	Estimate	T-statistic	Significance Level
1:4Lqccpi	0.9994	1649.4	<.01%
0L40:4	-0.2668	-5.3349	<.01%
2L40:20	0.8486	4.6411	<.01%
3L2:1	0.7182	3.4663	.07%

The most important indicator of inflation is recent inflation. The variable used to represent this, denoted 1:4Lqccpi, is the average of qccpi for the past four quarters. The coincident variable, 0L40:4 has a negative coefficient. This may be due to inflation influencing current interest rates, but with a greater impact on short term than long term rates, thus flattening the yield curve. At lag 2 quarters, the coefficient for 2L40:20 is positive and at lag 3 quarters that for 3L2:1 is positive. These indicate a general tendency for a steeper yield curve to anticipate future inflation. The r-squared, adjusted for degrees of freedom, is 65%. The standard error of the estimate is 0.0051. Thus the typical predicted quarterly change is accurate to about half a percentage point. The standard error is the standard deviation of a residual normally distribution around the predicted point, which can be used to draw the scenario actually simulated. The actual vs. fit is graphed in Appendix 3. The series can be seen to be fairly noisy, but the model does pick up the general movements over time. The residuals are graphed on a normal scale. Normality looks to be reasonably consistent with the observed residuals.

The stock market variable modeled, qcw5, is the ratio of the Wilshire 5000 index W5 at the end of a quarter to that at the previous quarter end. In this case the CPI percentage change variable qccpi was included in the regression as an explanatory variable. This allows creation of scenarios that have simulated values of W5 that are probabilistically consistent with the CPI value for the scenario.

The fitted equation for quarter ending data 1971 through first quarter 1997 is shown in the table below. In this regression only two variables were used, but they are composite series. The first, denoted 0-4Lqccpi, is the increase in qccpi over the last year, i.e., the current rate less the rate a year earlier. This variable has a negative coefficient, indicating that an increase in inflation is bad for equity returns. The other variable is denoted qcrelsprd. It represents the previous quarter's increase in the long-term spread less this quarter's increase in the short-term spread. Here the long-term spread is the difference between 10-year and 5-year rates, and the short-term spread is the difference between 6-month and 3-month rates. The increases noted are the quarter-to-quarter arithmetic increases in these spreads.

The coefficient on qcrelsprd is positive. This variable is positive if the increase in the short-term spread is less than the previous increase in the long-term spread, or if its decrease is greater. Either could suggest moderating inflation and interest rates, and thus be positive for equity returns.

Quarterly Change in Wilshire 5000

Variable	Estimate	T-statistic	Significance Level
0-4Lqccpi	-2.7113	-3.1936	0.2%
qcrelsprd	11.869	4.5273	<.01%
constant	1.02316	145.311	<.01%

The adjusted-r-squared is only 24% for this regression, indicating that the fit is not particularly good. The residual standard deviation is .0721, which allows a fairly wide deviation from the model. The fit is graphed in Appendix 3.

Results

The table below shows the mean surplus, the ratio of mean to standard deviation, and several percentiles of the surplus for the case in which there are no losses, just investment of surplus.

Surplus Only

	Mean	Mean/SD	1%	10%	90%	99%
Short	3048	-	3048	3048	3048	3049
Medium	3053	45.3	2867	2967	3125	3227
Long	3071	19.5	2706	2861	3284	3407
Stocks+	3136	13.1	2577	2829	3422	3760

The ratio of mean to standard deviation is chosen as a risk measure for which higher is better, as is the case with all the other figures in the table. This table is consistent with the idea that riskier investments have higher expected return, but could have more adverse developments as well. The one-year bonds have no risk in this case, as they are held a year and then mature.

The next table shows the results of adding fixed liabilities to the mix.

Fixed Liabilities

	Mean	Mean/SD	1%	10%	90%	99%
Short	3419	-	3419	3419	3419	3419
Medium	3434	14.8	2798	3104	3705	3953
Long	3492	7.3	2031	2848	4094	4409
Stocks+	3581	4.6	1951	2656	4630	5282

Here the mean surplus is higher, due to the expected profits from the insurance business. However, the risk is considerably greater, due to the larger investment portfolio compared to the same surplus. This works at both the low and high end of the probability distribution.

Adding variability to the liabilities further increases the risk, as shown below. Here the change in the extreme percentiles is greater for the short-term investments, showing that the increase in risk over fixed liabilities is greater when investing short.

Variable Liabilities -No Inflation on Reserves

	Mean	Mean/SD	1%	10%	90%	99%
Short	3422	21.0	3085	3220	3626	3821
Medium	3443	11.4	2600	3024	3786	4096
Long	3470	6.8	2182	2801	4117	4784
Stocks+	3540	4.1	1762	2287	4661	6120

If reserves are subject to post-event inflation, risk increases more:

With Post-Event Inflation

	Mean	Mean/SD	1%	10%	90%	99%
Short	3429	20.2	3021	3205	3635	3859
Medium	3438	10.6	2589	2972	3816	4289
Long	3538	6.3	1899	2848	4242	4879
Stocks+	3569	3.9	1358	2294	4613	6197

Stocks may pose too much of a risk at the 1% level in this case, where they may have been an acceptable risk without post-event inflation. This illustrates the value of understanding the reserve-generating process when setting investment strategy.

Finally, buying more reinsurance reduces the expected surplus but also the variability of surplus.

No Post-Event Inflation with More Reinsurance

	Mean	Mean/SD	1%	10%	90%	99%
Short	3227	55.3	3271	3351	3500	3618
Medium	3255	14.9	2884	3156	3749	3951
Long	3345	6.9	2197	2865	4202	4630
Stocks+	3473	4.4	1773	2564	4909	5642

For this company, buying more reinsurance with long-term investments has lower expected return and more downside risk than buying less reinsurance with medium term investments. This strategy would give up considerable upside potential, however.

CONCLUSION

The risks to the various investment strategies that an insurer may follow will change depending on underwriting risk and reserve development risk. To quantify this risk the process generating reserve development needs to be identified. Once that is done, the trade-offs between different investment strategies and different underwriting strategies - including alternative reinsurance programs - can be quantified by dynamic financial analysis.

APPENDIX 1 — SIMULATING ASSET PERFORMANCE

Most asset classes and many economic series have been found to correlate to the treasury yield curve. Realistic simulation of the yield curve is an involved undertaking, and a subject of ongoing research among academics and all sorts of financial practitioners. If any researchers have gotten this absolutely right, they're keeping it a secret, and probably getting wealthy. Some of the progress in this area is discussed below, along with some proposed tests of yield curve simulation methods for DFA modeling.

Once the yield curves have been generated, the other assets and economic values can be simulated by regressions against the yield curve and lags of the yield curve (and perhaps against the other economic variables already simulated). In each case, a random draw from the error term of the distribution should be added to the regression estimate in order to keep the correlations from being perfect (unless they happen to be, which is rare).

A good deal of the work in yield-curve simulation is done for the purpose of pricing or evaluating the pricing of interest-rate options. For this purpose it is important that the model captures the current yield curve and its short-term dynamics as precisely as possible. This would be important to insurers who are actively trading bond options. However, the usual emphasis in DFA modeling is a little different. The risks inherent in different investment strategies over a longer time frame are more of a concern. A wide variety of yield curves should be produced to test this, but the model producing the widest variety is not necessarily the most useful - the different yield curves should be produced in relative proportion to their probability of occurring. It would be nice if the short-term forecasts were very close to the current curve, but this is less important for DFA than it is for option trading.

Historical data on the distribution of yield curves can be used to test the reasonability of the distribution of curves being produced by any given model. However, it is not reasonable to expect that the probability of yield curves in a small given range showing up in the next two or three years is the same as their historical appearance. Some recognition needs to be given to the current situation and the speed at which changes in the curve tend to occur. Care also needs to be exercised in the selection of the historical period to which comparisons are to be made. The years 1979-81 exhibited dramatic changes in the yield curve, and the analyst needs to consider how prominent these years will be in the history selected. It seems reasonable that using a period beginning in the 1950's will give this unusual phase due recognition without over-emphasizing it.

The following are proposed as general criteria that a model of the yield curve should meet:

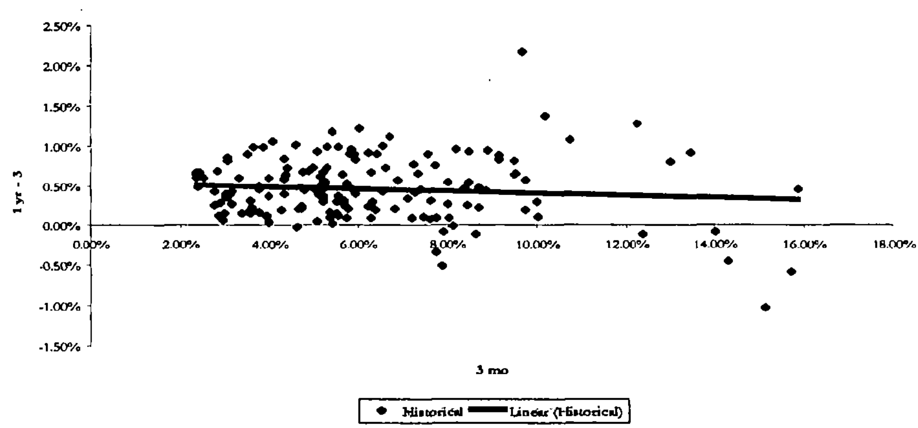
- It should closely approximate the current yield curve.
- It should produce patterns of changes in the short-term rate that match those produced historically.
- Over longer simulations, the ultimate distributions of yield curve shapes it produces, given any short-term rate, should match historical results.

This last criterion looks at the contingent distributions of yield curve shapes given the short-term rate. Thus it allows for the possibility that the distribution of short-term rates simulated even after several years will not match the diversity of historical rates. But it does require that for any given short-term rate the distribution of yield curves should be as varied as seen historically for that short-term rate. It could be argued that somewhat less variability would be appropriate, and this may be so. How much less would be a matter of judgment, but too little variation in this conditional distribution would seem ill-advised when generating scenarios to test investment strategies against.

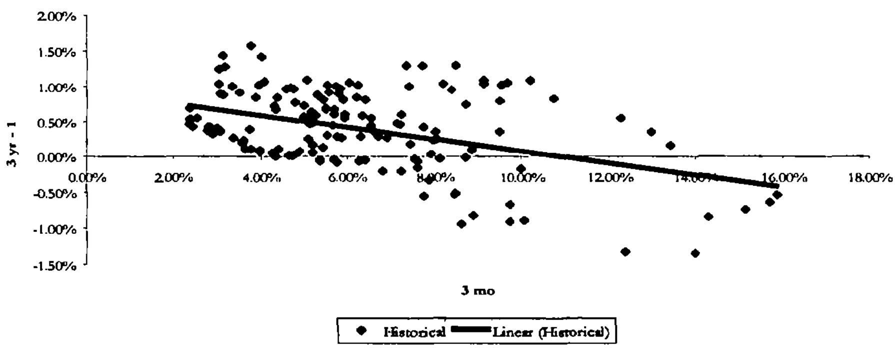
To measure the distribution of yield curve shapes, some shape descriptors are needed. The ones used here are based on differences of interest rates of different maturities. The first measures are just the successive differences in yield rates for 3-month, 1-year, 3-year, and 10-year instruments. Then the differences in these differences are taken, and finally the differences of those second differences. The first differences quantify the steepness of different parts of the yield curve. These would be zero for a flat curve. The second differences quantify the rate of change in the steepness as you move up the curve. These would be zero for a linearly rising curve. The third difference would be zero for a quadratic curve, and so quantifies the degree to which the curve is not quadratic.

These shape measures will be reviewed historically as a function of the 3-month rate. The patterns for these six measures are graphed below along with the regression lines against the 3-month rate. It is interesting to note that the 1-year / 3-month yield spread appears to be independent of the 3-month rate, but the longer-term spreads appear to decline slightly with higher 3-month rates. At least in the US economy, when the short-term rates are high, the long-term rates tend to show less response, perhaps because investors expect the short-term rates to come down, and so the yield curve flattens out or even shows reversals (i.e., short-term rates higher than long-term). It might be argued that the slopes of the regression lines are small enough compared to the noise that they should not be considered significant. It turns out, however, that in testing models against this data the non-significance of the slope is a most significant issue - most models tend to produce more steeply falling slopes than the data shows.

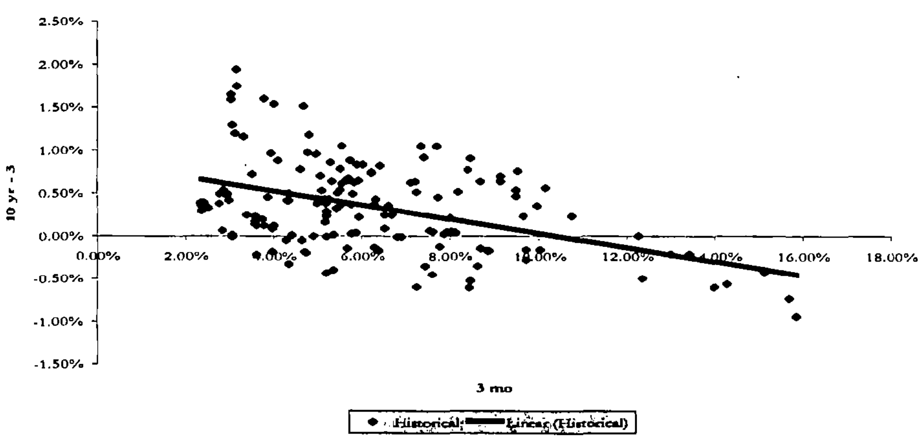
Historical 1 Year - 3 Month

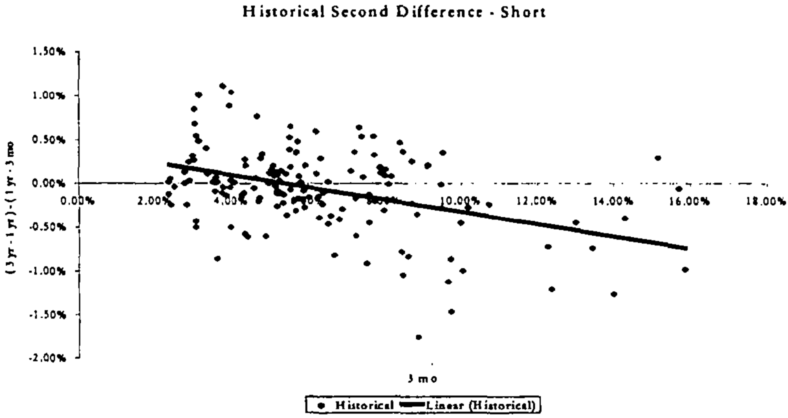


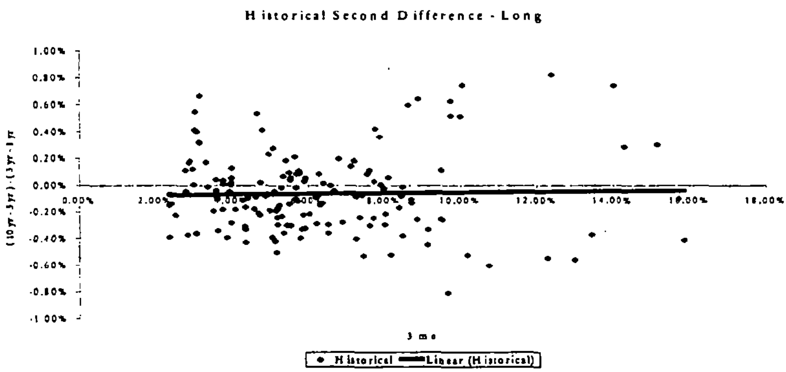
Historical 3 Year - 1 Year

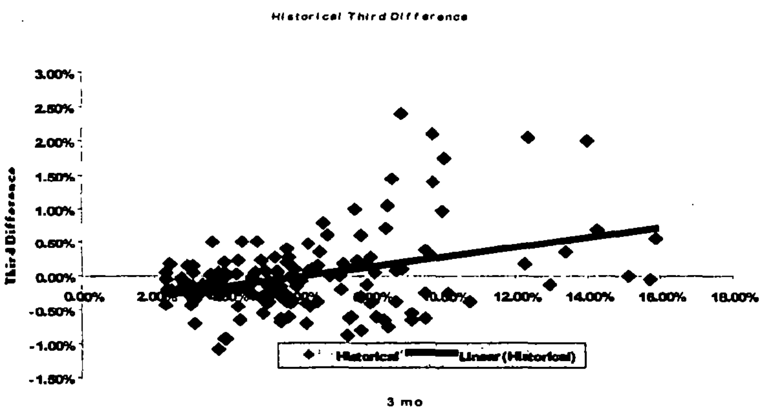


Historical 10 Year - 3 Year




Historical Second Difference - Short
(3 yr - 1 yr) - (1 yr - 3 mo)
3 mo
Historical
Linear (Historical)


Historical Second Difference - Long
(10 yr - 3 yr) - (3 yr - 1 yr)
3 mo
Historical
Linear (Historical)


Historical Third Difference
Third Difference
3 mo
Historical
Linear (Historical)

Typically the short-term interest rate, denoted as r, is modeled directly, and longer-term rates are inferred from the implied behavior of r, along with market considerations. The modeling of r is usually done as a continuously fluctuating diffusion process. This is based on Brownian motion. A continuously moving process is hard to track, and processes with random elements do not follow a simple formula. These processes are usually described by the probability distribution for their outcomes at any point in time. A Brownian motion has a simple definition for the probabilities of outcomes: the change from the current position between time zero and time t is normally distributed with mean zero and variance $\sigma^2 t$ for some σ. If r is the short-term interest rate and it follows such a Brownian motion, it is customary to express the instantaneous change in r by $dr = \sigma dz$. Here z represents a Brownian motion with $\sigma=1$. If r also has a trend of bt during time t, this could be expressed as $dr = bdt + \sigma dz$.

Cox, Ingersoll and Ross (1985) provided a model of the motion of the short-term rate that has become widely studied. In the CIR model r follows the following process:

$$dr = a(b - r)dt + sr^{1/2}dz.$$

Here b is the level of mean reversion. If r is above b, then the trend component is negative, and if r is below b it is positive. Thus the trend is always towards b. The speed of mean reversion is expressed by a, which is sometimes called the half-life of the reversion. Note that the volatility depends on r itself, so higher short-term rates would be associated with higher volatility. The period 1979-81 had high rates and high volatility, and studies that emphasize this period have found that the power of ½ on r is too low. It appears to be about right in longer studies however.

Nonetheless, the CIR model fails to capture other elements of the movement of short-term rates. There have been periods of high volatility with low interest rates, and the rates sometimes seem to gravitate towards a temporary mean for a while, then shift and go towards some other. One way to account for these features is to let the volatility parameter s and the reversion mean b both be stochastic themselves.

Andersen and Lund (Working Paper No. 214, Northwestern University Department of Finance) give one such model:

$$dr = a(b - r)dt + sr^k dz_1 \qquad k>0$$
$$d\ln s^2 = c(p - \ln s^2)dt + v dz_2$$
$$db = j(q - b)dt + wb^{1/2} dz_3$$

Here there are three standard Brownian motion processes, z_1, z_2, and z_3. The rate r moves subject to different processes at different times. It always follows a mean-reverting process, with mean b. But that mean itself changes over time, following a mean-reverting process defined by k, q, and w. The volatility parameter s^2 also varies over time via a mean reverting geometric Brownian motion process (i.e., Brownian motion on the log). In total there are eight parameters: a, c, j, k, p, q, v, and w and three varying factors r, b, and s.

Models of the short-term rate can lead to models of the whole yield curve. This is done by modeling the prices of zero-coupon bonds with different maturities all paying \$1. If P(T) is the current price of such a bond for maturity T, the implied continuously compounding interest rate can be shown to be -ln[P(T)]/T. P(T) itself is calculated as the risk adjusted discounted expected value of \$1. Here "discounted" means continuously discounted by the evolving interest rate r, and "expected value" means that the mean discount is calculated over all possible paths for r. This can be expressed as:

$$P(T) = E^*[\exp(-\int r_t dt)]$$

Where r_t is the interest rate at time t, the integral is over the time period 0 to T, and E^* is the risk-adjusted expected value of the results of all such discounting processes.

If E were not risk adjusted, P(T) could be estimated by many instances of simulating the r process to time T over small increments and then discounting back over each increment. The risk-adjusted expected value is obtained by using a risk-adjusted process to simulate the r's. This process is like the original process except that it tends to produce higher r's over time. These higher rates provide a reward for bearing the longer-term interest rate risk. Increasing the trend portion of the diffusion process produces the adjusted process. In the CIR model it is increased by λr, where λ is called "the market price of risk." Andersen and Lund add λrs, and also add a similar risk element to the b diffusion.

However, in the case of the CIR model a closed form solution exists which simplifies the calculation. The yield rate for a zero coupon bond of maturity T is given by $Y(T) = A(T) + rB(T)$ where:

$A(T) = -2(ab/s^2T)\ln C(T) - 2aby/s^2$

$B(T) = [1 - C(T)]/yT$

$C(T) = (1 + xye^{T/x} - xy)^{-1}$

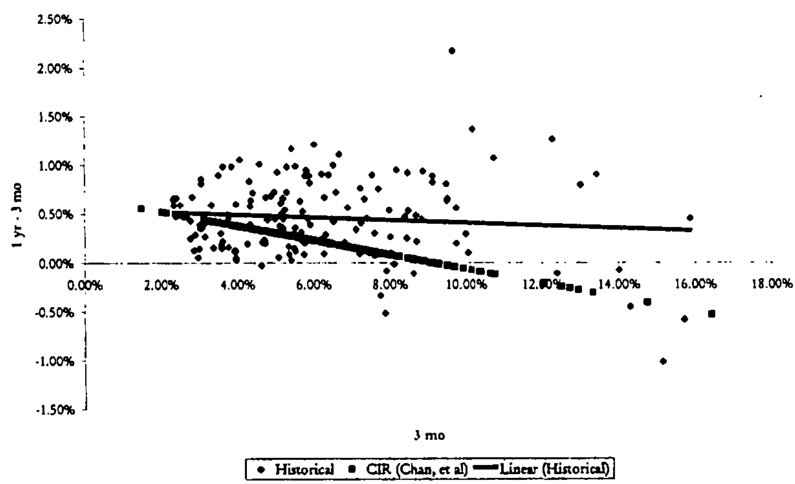
$x = [(a - \lambda)^2 + 2s^2]^{-1/2}$

$y = (a - \lambda + 1/x)/2.$

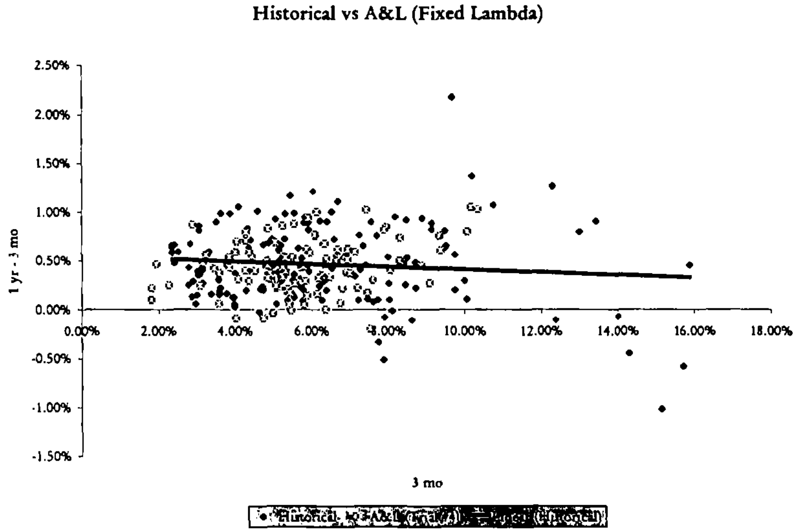
Note that neither A nor B is a function of r, so Y is a linear function of r (but not of T of course). Thus for the CIR model, all the yield curve shape measures defined above are linear functions of r, and as the three-month rate is as well, the shape measures are strictly linear in the three-month rate. This is in contrast to the historical data, which shows a dispersion of the shape measures around a perhaps linear relationship. The graph below as an example shows the historical and CIR implied 1 year less 3 month spread as a function of the 3-month rate, along with the historical trend line.

The parameters used here for the CIR model, from Chan etal. (1992) are: a=.2339; b=.0808; s=.0854, with λ set to .03. Different parameter values could possibly get the slope closer to that of the historical data, but the dispersion around the line cannot be achieved with this model. Experimentation with different parameter values suggests that even getting the slopes to match historical for all three of the first-difference measures may be difficult as well.

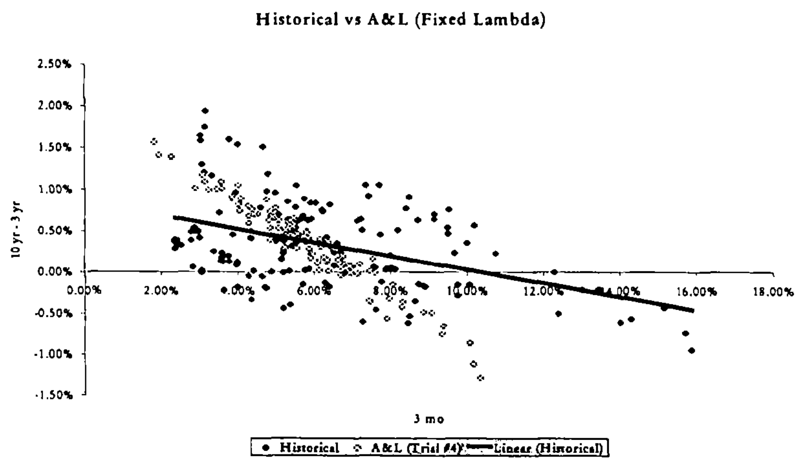
Historical vs CIR (Fixed Lambda)



Another potential problem with the CIR model is that the very long-term rates do not vary with r at all, but it's not clear how long the rates have to be for this.

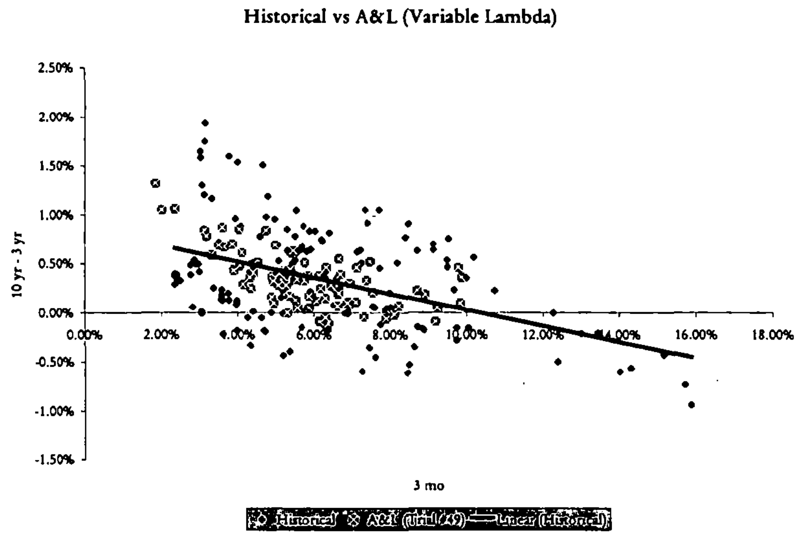


The Andersen-Lund model does provide more dispersion around the trend line, and also has about the right slope for the 3-month to 1-year spread, as the graph above shows. It does not do as well with the 3-year to 10-year spread in either



slope or dispersion, as shown here.

One approach that seems to give a degree of improvement is to let the market price of risk vary as well, through its own stochastic process. This would allow the same short-rate process to generate different yield curves at different times due to different market situations. This approach is capable of fixing the slope and dispersion problem for the long spread, as shown below.



Allowing stochastic market price of risk may improve the CIR model's performance on these tests as well, but it's not clear how to do this and still maintain a closed-form yield curve, which is the main advantage of CIR.

The table below summarizes some of the comparisons of model and historical results discussed above. For each of the models and each of the yield spreads, the linear relationship between the yield spread and the three-month rate is summarized by three statistics: the slope of the regression line of the spread on the three-month rate, the value on that line for r = .06, and the standard deviation of the points around the line. The value at r = .06 was compared instead of the intercept of the line to show how the model matched historical values for a typical interest rate.

The values were based on simulations of rates about three years beyond the initial values. Thus perhaps less variability of the residuals might be justifiable than in the historical data, which were quarterly values from 1959 through 1997. The whole variety of yield curve shapes from this nearly forty-year period may not be likely in just three years. A longer simulation period would thus give a better test of these models, and a somewhat lower residual standard deviation than historical may be acceptable for the test actually performed.

1 yr - 3 mo	**Historical**	**CIR**	**AL Fixed**	**AL Variable**
Slope	1.17%	-7.31%	2.49%	1.66%
Predicted @ 6%	0.48%	0.23%	0.44%	0.43%
Std Dev of Residuals	0.35%	0.00%	0.25%	0.43%
3 yr - 1 yr				
Slope	-8.41%	-16.58%	-5.68%	-2.57%
Predicted @ 6%	0.42%	0.49%	0.39%	0.40%
Std Dev of Residuals	0.52%	0.00%	0.12%	0.36%
10 yr - 3 yr				
Slope	-8.17%	-34.23%	-29.22%	-9.86%
Predicted @ 6%	0.35%	0.89%	0.29%	0.32%
Std Dev of Residuals	0.48%	0.00%	0.12%	0.50%

All the models tested had a lower residual standard deviation for the 3-year to 1-year spread than seen historically, but not unreasonably so for the variable price of risk model. The slopes of the 10-year to 3-year spread were all steeper than historical, but again the variable model was best.

This methodology gives an indication of a method of testing interest rate generators. There are quite a few of these in the finance literature, so none of the generators tested above can be considered optimal. In addition some refinement of the testing methodology may be able to tighten the conclusions discussed above.

Appendix 2 - Simulating Loss Development

The principal task in simulating a company's loss development is identifying the stochastic process that generates that development. Testing different processes against the historical development data is a way to approach this task. The second task is to model how the company's carried reserves respond to the loss emergence scenarios generated. One assumption for this may be that the company knows the process that produces its development, and uses a reserving methodology appropriate for that process. The simulation would proceed by generating loss emergence scenarios stochastically and then applying the selected reserving method to produce the carried reserves for each scenario. On the other hand, if the company has a fixed reserve methodology that it is going to use no matter what, then that methodology can be used to produce the carried reserves from the simulated emergence.

For this discussion, "emergence" could either mean case emergence or paid emergence, or both. The main concern here is simulating the emerging losses by period. This may or may not involve simulating the ultimate losses. For instance, one way to generate the losses to emerge in a period is to multiply simulated ultimate losses times a factor drawn from a percentage emerged distribution. This is appropriate when the process producing the losses for each period works by taking a randomized percent of ultimate losses. This method might involve some quite complicated methods of simulating ultimates, but all those that take period emergence as a percentage of ultimate will be considered to be using the same type of emergence pattern. Several other emergence patterns will be considered below, and the reserving methods appropriate for each will be discussed. Then methods for identifying the emergence patterns from the data triangles will be explored.

Types of Emergence Patterns

Six characteristics of emergence patterns will be considered here. Each will be treated as a binary choice, thus producing 64 types of emergence patterns. However there will be sub-categories within the 64, as not all of the choices are actually binary. The six basic choices for defining loss emergence processes are:

Do the losses that emerge in a period depend on the losses already emerged? Mack has shown that the chain ladder method assumes an emergence pattern in which the emerged loss for a period is a constant factor times the previous emerged, plus a random disturbance. Other methods, however, might apply factors only to ultimate losses, and then add a random disturbance. The latter is the emergence pattern assumed by the Bornheutter-Ferguson (BF) method, for example.

Is all loss emergence proportional? Both the chain ladder and BF methods use factors to predict emergence, and so are based on processes where emergence is proportional to something - either ultimate losses in the BF case or previously emerged in the chain ladder. However, the expected loss emergence for a period could be constant - not proportional to anything. Or it could be a factor times something plus a constant. If this is the emergence pattern used, then the reserving methodology should also incorporate additive elements.

Is emergence independent of calendar year events? Losses to emerge in a period may depend on the inflation rate for the period. This is an example of a calendar year or diagonal effect. Another example is strong or weak development due to a change in claim handling methods. Thus this is not a purely binary question - if there are diagonal effects there will be sub-choices relating to what type of effect is included. The Taylor separation method is an example of a development method that recognizes calendar year inflation. In many cases of diagonal effects, the ultimate losses will not be determined until all the development periods have been simulated.

Are the parameters stable? For instance a parameter might be a loss development factor. A stable factor could lead to variable losses due to randomness of the development pattern, but the factor itself would remain constant. The alternative is that the factor changes over time. There are sub-cases of this, depending on how they change.

Are the disturbance terms generated from a normal distribution? The typical alternative is lognormal, but the possibilities are endless. Clearly the loss development method will need to respond to this choice.

Are the disturbance terms homoskedastic? Some regression methods of development assume that the random disturbances all have the same variance, at least by development age. Link ratios are often calculated as the ratio of losses at age j+1 divided by losses at age j, which assumes that the variance of the disturbance term is proportional to the mean loss emerged. Another alternative is for the standard deviation to be proportional to the mean. The variance assumption used to generate the emerging losses can be employed in the loss reserving process as well.

Notation

Losses for accident year w evaluated at the end of that year will be denoted as being as of age 0, and the first accident year in the triangle is year 0. The notation below will be used to specify the models.

$c_{w,d}$: cumulative loss from accident year w as of age d

$c_{w,\infty}$: ultimate loss from accident year w

$q_{w,d}$: incremental loss for accident year w to emerge in period d

f_d: factor used in emergence for age d

h_w : factor (dollar amount) used in emergence for year w

g_{w+d}: factor used in emergence for calendar year w+d

a_d: additive term used in emergence for age d

QUESTION 1

The stochastic processes specified by answering the six questions above can be numbered in binary by considering yes=1 and no=0. Then process 111111 (all answers yes) can be specified as follows:

$$q_{w,d} = c_{w,d-1} f_d + e_{w,d} \qquad (1)$$

where $e_{w,d}$ is normally distributed with mean zero. Here f_d is a development factor applied to the cumulative losses simulated at age d-1. A starting value for the accident year is needed which could be called $c_{w,-1}$. For each d it might be reasonable to assume that $e_{w,d}$ has a different variance. Note that for this process, ultimate losses are generated only as the sum of the separately generated emerged losses for each age.

Mack has shown that for process 111111 the chain ladder is the optimal reserve estimation method. The factors f_d would be estimated by a no-constant linear regression. In process 111110 (heteroskedastic) the chain ladder would also be optimal, but the method of estimating the factors would be different. Essentially these would use weighted least squares for the estimation, where the weights are inversely proportional to the variance of $e_{w,d}$. If the variances are proportional to $c_{w,d-1}$, the resulting factor is the ratio of the sum of losses from the two relevant columns of the development triangle.

In all the processes 1111xx Mack showed that some form of the chain ladder is the best linear estimate, but when the disturbance term is not normal, linear estimation is not necessarily optimal.

Processes of type 0111xx do not generate emerged losses from those previously emerged. A simple example of this type of process is:

$$q_{w,d} = h_w f_d + e_{w,d} \qquad (2)$$

Here h_w can be interpreted as the ultimate losses for year w, with the factors f_d summing to unity. For this process, reserving would require estimation of the f's and h's. I call this method of reserving the parameterized BF, as Bornheutter and Ferguson estimated emergence as a percentage of expected ultimate. The method of estimating the parameters would depend on the distribution of the disturbance term $e_{w,d}$. If it is normal and homoskedastic, a regression method can be used iteratively by fixing the f's and regressing for the h's, then taking those h's to find the best f's, etc. until both f's and h's converge. If heteroskedastic, weighted regressions would be needed. If a lognormal disturbance is indicated, the parameters could be estimated in logs, which is a linear model in the logs.

QUESTION 2

Additive terms can be added to either of the above processes. Thus an example of a 0011xx process would be:

$$q_{w,d} = a_d + h_w f_d + e_{w,d} \qquad (3)$$

If the f's are zero, this would be a purely additive model. A test for additive effects can be made by adding them to the estimation and seeing if significantly better fits result.

QUESTION 3

Diagonal effects can be added similarly. A 0001xx model might be:

$$q_{w,d} = a_d + h_w f_d g_{w+d} + e_{w,d} \qquad (4)$$

Again this can be tested by goodness of fit. There may be too many parameters here. It will usually be possible to reasonably simulate losses without using so many distinct parameters. Specifying relationships among the parameters can lead to reduced parameter versions of these processes. For instance, some of the parameters might be set equal, such as $h_w=h$ for all w. Note that the 0111xx process $q_{w,d}= hf_d + e_{w,d}$ is the same as the 0011xx process $q_{w,d}= a_d + e_{w,d}$, as a_d can be set to hf_d. The resulting reserve estimation method is an additive version of the chain ladder, and is sometimes called the Cape Cod method.

Another way to reduce the number of parameters is to set up trend relationships. For example, constant calendar year inflation can be specified by setting $g_{w+d}=(1+j)^{w+d}$. Similar trend relationships can be specified among the h's and f's. If that is too much parameter reduction to adequately model a given data triangle, a trend can be established for a few periods and then some other trend can be used in other periods.

QUESTION 4

Rather than trending, the parameters in the loss emergence models could evolve according to some more general stochastic process. This could be a smooth process or one with jumps. The state-space model is often used to describe parameter variability. This model assumes that observations fluctuate around an expected value that itself changes over time as its parameters evolve. The degree of random fluctuation is measured by the variance of the observations around the mean, and the movement of the parameters is quantified by their variances over time. The interplay of these two variances determines the weights to apply, as in credibility theory.

To be more concrete, a formal definition of the model follows where the parameter is the 2nd to 3rd development factor. Let:

β_i=2nd to 3rd factor for ith accident year

y_i=3rd report losses for ith accident year

x_i=2nd report losses for ith accident year

The model is then:

$y_i=x_i\beta_i+\varepsilon_i$. (5)

The error term ε_i is assumed to have mean 0 and variance σ_i^2.

$\beta_i=\beta_{i-1}+\delta_i$. (6)

The fluctuation δ_i is assumed to have mean 0 and variance ν_i^2, and to be independent of the ε's.

In this general case the variances could change with each period i. Usually some simplification is applied, such as constant variances over time, or constant with occasional jumps in the parameter - i.e., occasional large ν_i's.

If this model is adopted for simulating loss emergence, the estimation of the factors from the data can be done using the Kalman filter.

Questions 5 and 6

The error structure can be studied and usually reasonably understood from the data triangles. The loss estimation method associated with a given error structure will be assumed to be maximum likelihood estimation from that structure. Thus for normal distributions this is weighted least squares, where the weights are the inverses of the variances. For lognormal this is the same, but in logs.

Identifying Emergence Patterns

Given a data triangle, what is the process that is generating it? This is useful to know for loss reserving purposes, as then reserve estimation is reduced to estimation of the parameters of the generating process. It is even more critical for simulation of company results, as the whole process is needed for simulation purposes.

Identifying emergence patterns can be approached by fitting different ones to the data and then testing the significance of the parameters and the goodness of fit. As more parameters often appear to give a better fit, but reduce predictive value, a method of penalizing over-parameterization is needed when comparing competing models. The method proposed here is to compare models based on sum of squared residuals divided by the square of the degrees of freedom, i.e., divided by the square of observations less parameters.

This measure gives impetus to trying to reduce the number of parameters in a given model, e.g., by setting some parameters the same or by identifying a trend in the parameters. This seems to be a legitimate exercise in the effort of identifying emergence patterns, as there are likely to be some regularities in the pattern, and simplifying the model is a way to uncover them.

Fitting the above models is a straightforward exercise, but reducing the number of parameters may be more of an art than a science. Two approaches may make sense: top down, where the full model is fit and then regularities among the parameters sought; and bottom up, where the most simplified version is estimated, and then parameters added to compensate for areas of poor fit.

To illustrate this approach, the data triangle of reinsurance loss data first introduced by Thomas Mack will be the basis of model estimation.

QUESTIONS 1 & 2 – FACTORS AND CONSTANT TERMS

Table 1 shows incremental incurred losses by age for some excess casualty reinsurance. As an initial step, the statistical significance of link ratios and additive constants was tested by regressing incremental losses against the previous cumulative losses. In the regression the constant is denoted by a and the factor by b. This provides a test of question 1 - dependence of emergence on previous emerged, and also one of question 2 – proportional emergence. Here they are being tested by looking at whether or not the factors and the constants are significantly different from zero, rather than by any goodness-of-fit measure.

Table 1 - Incremental Incurred Losses

0	1	2	3	4	5	6	7	8	9
5012	3257	2638	898	1734	2642	1828	599	54	172
106	4179	1111	5270	3116	1817	-103	673	535	
3410	5582	4881	2268	2594	3479	649	603		
5655	5900	4211	5500	2159	2658	984			
1092	8473	6271	6333	3786	225				
1513	4932	5257	1233	2917					
557	3463	6926	1368						
1351	5596	6165							
3133	2262								
2063									

Table 2 - Statistical Significance of Link Ratios and Constants

	0 to 1	1 to 2	2 to 3	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8
'a'	5113	4311	1687	2061	4064	620	777	3724
Std a	1066	2440	3543	1165	2242	2301	145	0
'b'	-0.109	0.049	0.131	0.041	-0.100	0.011	-0.008	-0.197
std b	0.349	0.309	0.283	0.071	0.114	0.112	0.008	0

Table 2 shows the estimated parameters and their standard deviations. As can be seen, the constants are usually statistically significant (parameter nearly double its standard deviation, or more), but the factors never are. The lack of significance of the factors shows that the losses to emerge at any age d+1 are not proportional to the cumulative losses through age d. The assumptions underlying the chain ladder model are thus not met by this data. A constant amount emerging for each age usually appears to be a reasonable estimator, however.

Figure 1 illustrates this. A factor by itself would be a straight line through the origin with slope equal to the development factor, whereas a constant would give a horizontal line at the height of the constant.

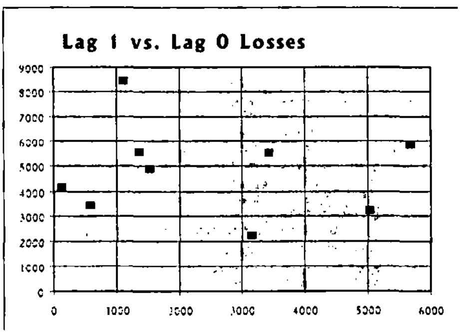


Figure 1

Although emerged losses are not proportional to previous emerged, they could be proportional to ultimate incurred. To test this, the parameterized BF model (2) was fit to the triangle. As this is a non-linear model, fitting is a little more involved. A method of fitting the parameters will be discussed, followed by an analysis of the resulting fit.

To do the fitting, an iterative method can be used to minimize the sum of the squared residuals, where the w,d residual is $[q_{w,d}-f_d h_w]$. Weighted least squares could also be used if the variances of the residuals are not constant over the triangle. For instance, the variances could be proportional to $f_d{}^p h_w{}^q$, in which case the regression weights would be $1/f_d{}^p h_w{}^q$.

A starting point for the f's or the h's is needed to begin the iteration. While almost any reasonable values could be used, such as all f's equal to 1/n, convergence will be faster with values likely to be in the ballpark of the final factors. A natural starting point thus might be the implied f_d's from the chain ladder method. For ages greater than 0, these are the incremental age-to-age factors divided by the cumulative-to-ultimate factors. To get a starting value for age 0, subtract the sum of the other factors from unity. Starting with these values for f_d, regressions were performed to find the h_w's that minimize the sum of squared residuals (one regression for each w). These give the best h's for that initial set of f's. The standard linear regression formula for these h's simplifies to:

$$h_w = \sum_d f_d q_{w,d} \,/\, \sum_d f_d^2 \qquad (7)$$

Even though that gives the best h's for those f's, another regression is needed to find the best f's for those h's. For this step the usual regression formula gives:

$$f_d = \sum_w h_w q_{w,d} \,/\, \sum_w h_w^2 \qquad (8)$$

Now the h regression can be repeated with the new f's, etc. This process continues until convergence occurs, i.e., until the f's and h's no longer change with subsequent iterations. Ten iterations were used in this case, but substantial onvergence occurred earlier. The first round of f's and h's and those at convergence are in Table 3. Note that the h's are not the final estimates of the ultimate losses, but are used with the estimated factors to estimate future emergence. In this case, in fact, h(0) is less than the emerged to date. A statistical package that includes nonlinear regression could ease the estimation.

Standard regression assumes each observation q has the same variance, which is to say the variance is proportional to $f_d^p h_w^q$, with p=q=0. If p=q=1 the weighted regression formulas become:

$$h_w^2 = \sum_d [q_{w,d}^2/f_d] \,/\, \sum_d f_d \text{ and}$$

$$f_d^2 = \sum_w [q_{w,d}^2/h_w] \,/\, \sum_w h_w$$

Table 3 - BF Parameters

Age d	0	1	2	3	4	5	6	7	8	9
f_d 1st	0.106	0.231	0.209	0.155	0.117	0.083	0.038	0.032	0.018	0.011
f_d ult	0.162	0.197	0.204	0.147	0.115	0.082	0.037	0.030	0.015	0.009
Year w	0	1	2	3	4	5	6	7	8	9
h_w 1st	17401	15729	23942	26365	30390	19813	18592	24154	14639	12733
h_w ult	15982	16501	23562	27269	31587	20081	19032	25155	13219	19413

For comparison, the development factors from the chain ladder are shown in Table 4. The incremental factors are the ratios of incremental to previous cumulative. The ultimate ratios are cumulative to ultimate. Below them are the ratios of these ratios, which represent the portion of ultimate losses to emerge in each period. The zeroth period shown is unity less the sum of the other ratios. These factors were the initial iteration for the f_d's shown above.

Table 4 - Development Factors

	0 to 1	1 to 2	2 to 3	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8	8 to 9
Incremental	1.22	0.57	0.26	0.16	0.10	0.04	0.03	0.02	0.01
	0 to 9	1 to 9	2 to 9	3 to 9	4 to 9	5 to 9	6 to 9	7 to 9	8 to 9
Ultimate	6.17	2.78	1.77	1.41	1.21	1.10	1.06	1.03	1.01
0.162	0.197	0.204	0.147	0.115	0.082	0.037	0.030	0.015	0.009

Having now estimated the BF parameters, how can they be used to test what the emergence pattern of the losses is?

A comparison of this fit to that from the chain ladder can be made by looking at how well each method predicts the incremental losses for each age after the initial one. The sum of squared errors adjusted for number of parameters is the comparison measure, where the parameter adjustment is made by dividing the sum of squared errors by the square of [the number of observations less the number of parameters], as discussed earlier. Here there are 45 observations, as only the predicted points count as observations. The adjusted sum of squared residuals is 81,169 for the BF, and 157,902 for the chain ladder. This shows that the emergence pattern for the BF (emergence proportional to ultimate) is much more consistent with this data than is the chain ladder emergence pattern (emergence proportional to previous emerged).

The Cape Cod (CC) method was also tried for this data. The iteration proceeded similarly to that for the BF, but only a single h parameter was fit for all accident years. Now:

$$h = \sum_{w,d} f_d q_{w,d} \,/\, \sum_{w,d} f_d^2 \qquad (9)$$

The estimated h is 22,001, and the final factors f are shown in Table 5. The adjusted sum of squared errors for this fit is 75,409. Since the CC is a special case of the BF, the unadjusted fit is of course worse than that of the BF method, but with fewer parameters in the CC, the adjustment makes them similar. This formula for h is the same as the formula for h_w except the sum is taken over all w.

Intermediate special cases could be fit similarly. If, for instance, a single factor were sought to apply to just two accident years, the sum would be taken over those years to estimate that factor, etc.

Table 5 - Factors in CC Method

0	1	2	3	4	5	6	7	8	9
0.109	0.220	0.213	0.148	0.124	0.098	0.038	0.028	0.013	0.008

This is a case where the BF has too many parameters for prediction purposes. More parameters fit the data better, but use up information. The penalization in the fit measure adjusts for this problem, and shows the CC to be a somewhat better model. Thus the data is consistent with random emergence around an expected value that is constant over the accident years.

The CC method would probably work even better for loss ratio triangles than for loss triangles, as then a single target ultimate value makes more sense. Adjusting loss ratios for trend and rate level could increase this homogeneity.

In addition, a purely additive development was tried, as suggested by the fact that the constant terms were significant in the original chain ladder, even though the factors were not. The development terms are shown in Table 6. These are just the average loss emerged at each age. The adjusted sum of squared residuals is 75,409. This is much better than the chain ladder, which might be expected, as the constant terms were significant in the original significance-test regressions while the factors were not. The additive factors in Table 6 differ from those in Table 2 because there is no multiplicative factor in Table 6.

Table 6 - Terms in Additive Chain Ladder

1	2	3	4	5	6	7	8	9
4849.3	4682.5	3267.1	2717.7	2164.2	839.5	625	294.5	172

As discussed above, the additive chain ladder is the same as the Cape Cod method, although it is parameterized differently. The exact same goodness of fit is thus not surprising.

Finally, an intermediate BF-CC pattern was fit as an example of reduced parameter BF's. In this case ages 1 and 2 are assumed to have the same factor, as are ages 6 and 7 and ages 8 and 9. This reduces the number of f parameters from 9 to 6. The number of accident year parameters was also reduced: years 0 and 1 have a single parameter, as do years 5 through 9. Year 2 has its own parameter, as does year 4, but year 3 is the average of those two. Thus there are 4 accident year parameters, and so 10 parameters in total. Any one of these can be set arbitrarily, with the remainder adjusted by a factor, so there are really just 9. The selections were based on consideration of which parameters were likely not to be significantly different from each other.

The estimated factors are shown in Table 7. The accident year factor for the last 5 years was set to 20,000. The other factors were estimated by the same iterative regression procedure as for the BF, but the factor constraints change the simplified regression formula. The adjusted sum of squared residuals is 52,360, which makes it the best approach tried. This further supports the idea that claims emerge as a percent of ultimate for this data. It also indicates that the various accident years and ages are not all at different levels, but that the CC is too much of a simplification. The actual and fitted values from this, the chain ladder, and CC are in Exhibit 1. The fitted values in Exhibit 1 were calculated as follows. For the chain ladder, the factors from Table 4 were applied to the cumulative losses implied from Table 1. For the CC the fitted values are just the terms in Table 6. For the BF-CC they are the products of the appropriate f and h factors from Table 7.

Table 7 - BF-CC Parameters

Age d	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
f_d	*	**0.230**	**0.230**	**0.160**	**0.123**	**0.086**	**0.040**	**0.040**	**0.017**	**0.017**
Year w	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
h_w	**14829**	**14829**	**20962**	**25895**	**30828**	**20000**	**20000**	**20000**	**20000**	**20000**

Calendar Year Impacts – Testing Question 3

One type of calendar year impact is high or low diagonals in the loss triangle. Mack suggested a high-low diagonal test which counts the number of high and low factors on each diagonal, and tests whether or not that is likely to be due to chance. Here another high-low test is proposed: use regression to see if any diagonal dummy variables are significant. An actuary will often have information about changes in company operations that may have created a diagonal effect. If so, this information could lead to choices of modeling methods - e.g., whether to assume the effect is permanent or temporary. The diagonal dummies can be used to measure the effect in any case, but knowledge of company operations will help determine how to use this effect. This is particularly so if the effect occurs in the last few diagonals.

A diagonal in the loss development triangle is defined by w+d = constant. Suppose for some given data triangle, the diagonal w+d=7 is found to be 10% higher than normal. Then an adjusted BF estimate of a cell might be:

$$q_{w,d}=1.1f_dh_w \text{ if } w+d=7, \text{ and } q_{w,d}=f_dh_w \text{ otherwise} \quad (10)$$

1	2	5	4
3	8	9	
7	10		
7			

The small sample triangle of incremental losses here will be used as an example of how to set up diagonal dummies in a chain ladder model. The goal is to get a matrix of data in the form needed to do a multiple regression. First the triangle (except the first column) is strung out into a column vector. This is the dependent variable. Then columns for the independent variables are added. The second column is the cumulative losses at age 0 for the loss entries that are at age 1, and zero for the other loss entries. The regression coefficient for this column would be the 0 to 1 cumulative-to-incremental factor. The next two columns are the same for the 1 to 2 and 2 to 3 factors. The last two columns are the diagonal dummies. They pick out the elements of the last two diagonals. The coefficients for these columns would be additive adjustments for those diagonals, if significant.

2	1	0	0	0	0
8	3	0	0	1	0
10	7	0	0	0	1
5	0	3	0	1	0
9	0	11	0	0	1
4	0	0	8	0	1

This method of testing for diagonal effects is applicable to many of the emergence models. In fact, if diagonal effects are found significant in chain ladder models, they probably are needed in the BF models of the same data, so goodness-of-fit tests should be done with those diagonal elements included. Some examples are given in Appendix 2.

Another popular modeling approach is to consider diagonal effects to be a measure of inflation (e.g., see Taylor 1977). In a payment triangle this would be a natural interpretation, but a similar phenomenon could occur in an incurred triangle. In this case the latest diagonal effects might be projected ahead as estimates of future inflation. An understanding of what in company operations is driving the diagonal effects would help address these issues.

As with the BF model, the parameters of the model with inflation effects, $q_{w,d} = h_w f_d g_{w+d} + e_{w,d}$, can be estimated iteratively. With reasonable starting values, fix two of the three sets of parameters, fit the third by least squares, and rotate until convergence is reached. Alternatively, a non-linear search procedure could be utilized. As an example of the simplest of these models, modeling $q_{w,d}$ as just $6756(0.7785)^d$ gives an adjusted sum of squares of 57,527 for the reinsurance triangle above. This is not the best fitting model, but is better than some, and has only two parameters. Adding more parameters to this would be an example of the bottom up fitting approach.

Testing Question 4 - Stability of parameters

If a pattern of sequences of high and low residuals is found when plotted against time, instability of the parameters may be indicated. This can be studied and a randomness in the parameters incorporated into the simulation process, e.g., through the state-space model.

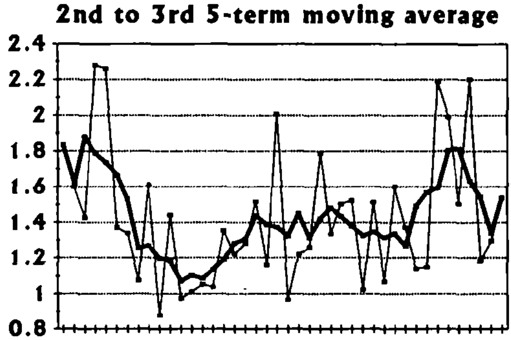


Figure 2

Figure 2 shows the 2nd to 3rd factor by accident year from a large development triangle (data in Exhibit 2) along with its five-term moving average. The moving average is the more stable of the two lines, and is sometimes in practice called "the average of the last five diagonals." There is apparent movement of the mean factor over time as well as a good deal of random fluctuation around it. There is a period of time in which the moving average is as low as 1.1 and other times it is as high as 1.8.

The state-space model assumes that observations fluctuate around a mean that itself changes over time. The degree of random fluctuation is measured by variance around the mean, and the movement of the mean by its variance over time. The interplay of these two variances determines the weights to apply, as in credibility theory.

The state-space model thus provides underlying assumptions about the process by which development changes over time. With such a model, estimation techniques that minimize prediction errors can be developed for the changing development case. This can result in estimators that are better than either using all data, or taking the average of the last few diagonals. For more details on the state space models see the Verrall and Zehnwirth references.

QUESTIONS 5 & 6: VARIANCE ASSUMPTIONS

Parameter estimation changes depending on the form of the variance. Usually in the chain ladder model the variance will plausibly be either a constant or proportional to the previous cumulative or its square. Plotting or fitting the squared residuals as a function of the previous cumulative will usually help decide which of these three alternatives fits better. If the squared residuals tend to be larger when the explanatory variable is larger, this is evidence that the variance is larger as well.

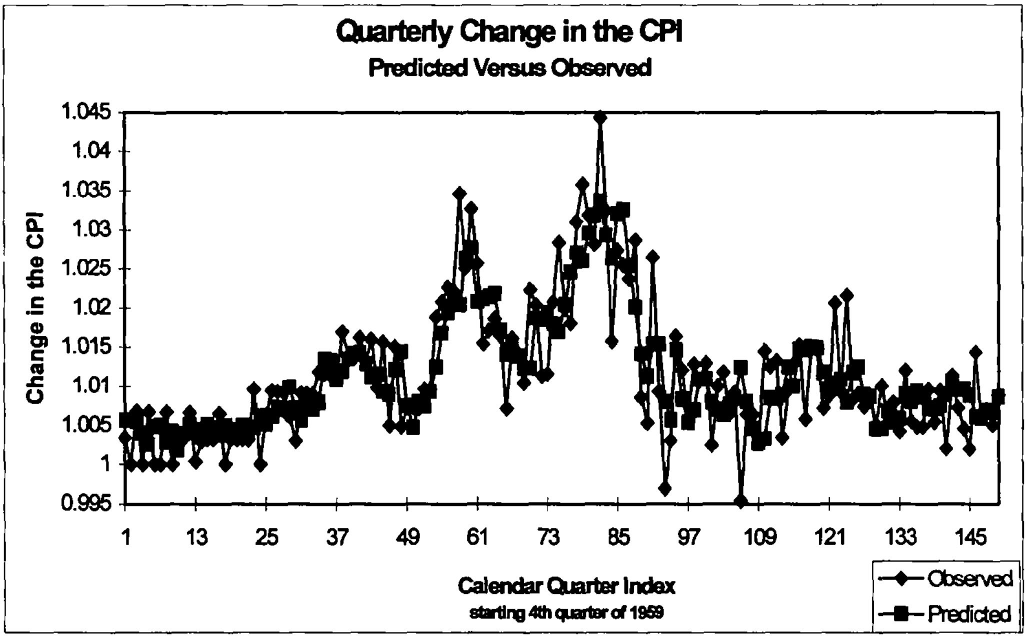
Another variance test would be for normality of the residuals. Normality is often tested by plotting the residuals on a normal scale, and looking for linearity. This is not a formal test, but it is often considered a useful procedure. If the residuals are somewhat positively skewed, a lognormal distribution may be reasonable. The non-linear models discussed are all linear in logs, and so could be much easier to estimate in that form. However, if some increments are negative, a lognormal model becomes awkward. The right distribution for the residuals of loss reserving models seems an area in which further research would be helpful.

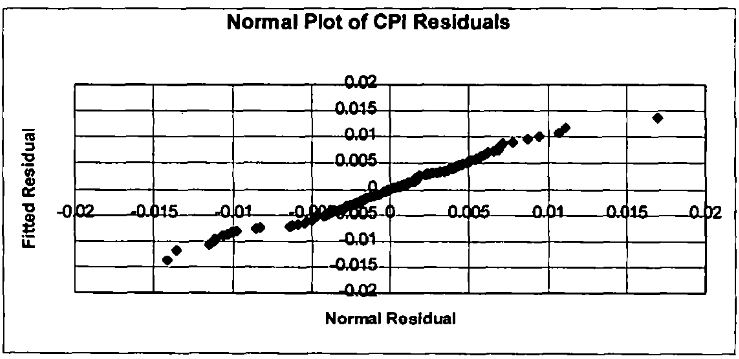
CONCLUSION

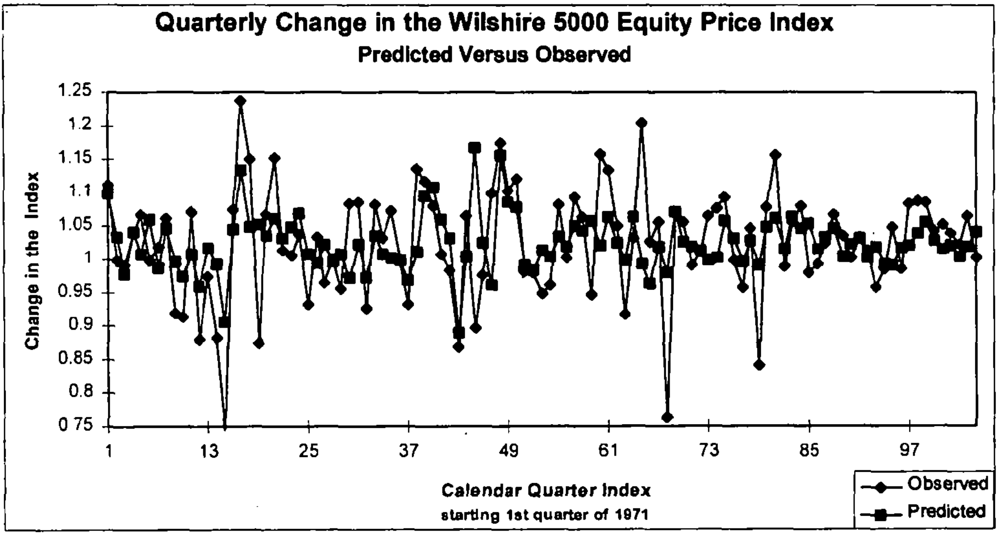
The first test that will quickly indicate the general type of emergence pattern faced is the test of significance of the cumulative-to-incremental factors at each age. This is equivalent to testing if the cumulative-to-cumulative factors are significantly different from unity. When this test fails, the future emergence is not proportional to past emergence. It may be a constant amount, it may be proportional to ultimate losses, as in the BF pattern, or it may depend on future inflation.

The addition of an additive component may give an even better fit. Reduced parameter models could also give better performance, as they will be less responsive to random variation. If an additive component is significant, converting the triangle to on-level loss ratios may improve the model. Tests of stability and for calendar-year effects may lead to further improvements.

APPENDIX 3 – REGRESSION GRAPHS






Quarterly Change in the Wilshire 5000 Equity Price Index
Predicted Versus Observed
Change in the Index
1.25
1.2
1.15
1.1
1.05
1
0.95
0.9
0.85
0.8
0.75
1
13
25
37
49
61
73
85
97
Calendar Quarter Index
starting 1st quarter of 1971
Observed
Predicted

References

Berquist and Sherman *Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach*, PCAS 1977

Bornheutter and Ferguson *The Actuary and IBNR*, PCAS 1972

Chan etal. *An Empirical Comparison of Alternative Models of the Short-Term Interest Rate* Journal of Finance 47 (1992).

Cox, Ingersoll and Ross *A Theory of the Term Structure of Interest Rates* Econometrica 53 (March 1985)

Gerber and Jones *Credibility Formulas with Geometric Weights*, Society of Actuaries Transactions 1975

de Jong and Zehnwirth *Claims Reserving, State-space Models and the Kalman Filter*, Journal of the Institute of Actuaries 1983

Mack *Measuring the Variability of Chain Ladder Reserve Estimates*, CAS Forum Spring 1994

Murphy *Unbiased Loss Development Factors*, PCAS 1994

Popper *Conjectures and Refutations*, Poutledge 1969

Stanard *A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques*, PCAS 1985

Taylor *Separation of Inflation and Other Effects from the Distribution of Non-Life Insurance Claim Delays,* ASTIN Bulletin 1977

Verrall *A State Space Representation of the Chain Ladder Linear Model,* Journal of the Institute of Actuaries, December 1989

Zehnwirth *Linear Filtering and Recursive Credibility Estimation,* ASTIN Bulletin, April 1985

Zehnwirth *Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital,* CAS Forum Spring 1994