UNDERSTANDING CORRELATIONS AND COMMON DRIVERS

Insureware

Innovative Statistical Solutions for P&C Insurance

Contents

1. Introduction

1.1. Correlation in the Industry

The word 'correlation' is one of the most misused statistical terms in the insurance industry.

We often come across statements like "the correlation between these two lines of business, CAL and PPA is 87%." What does this really mean?

We might suppose that correlation is a measure of the way the similarity between two lines creates parallel accounts, the paid losses in both lines tending to go up and down in sync.

It's easy to find examples of statistics that change in a similar way in time but which measure completely unrelated things. When annual chewing-gum sales tracks gun crime, or the science budget tracks the suicide rate we put this down to 'spurious correlation'.

What is spurious here is not in the statistics but in the way they are interpreted, and in fact a more thorough statistical analysis can easily separate spurious from non-spurious correlations.

As we will see this has everything to do with insurance. Where correlation is important in insurance is in calculating the risk margins for a portfolio of lines of business. Correlation measures impairment to risk diversification benefit.

It turns out that to get this right you need to very carefully separate out the true (non-spurious) correlations from what only look like correlation (spurious). It is only the true correlations that inflate the risk margins.

To understand why this is so, read on.

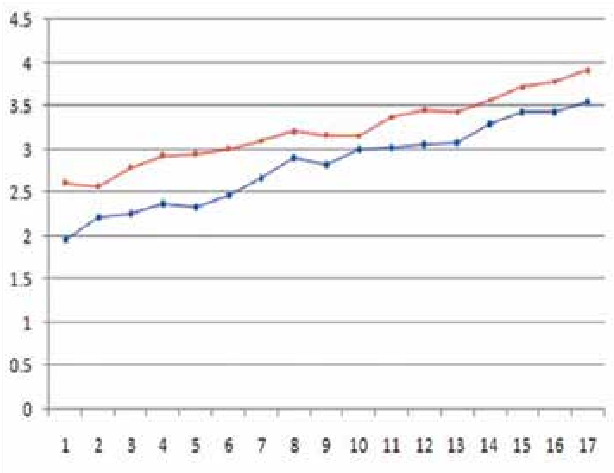
1.2. Correlation is model-dependent

The two realizations of time series in the graph below have a correlation of 97%.

However it is also clear that both are subject to an increasing trend. Are the two series correlated? Most likely they are not, in any meaningful way.

There is a multitude of phenomena that are completely unrelated but exhibit similar linear trends. In the example above, once the linear trends are removed the correlation measure falls to a statistically insignificant -1%.

Thus the answer is necessarily model-dependent.



Had the trends been ignored and instead the data were each modeled as a constant plus a "random factor", then the two "random factors" show a startling similarity, justifying a correlation assumption of 97%. Clearly this misrepresents the situation, mainly because it fails to account for the fact that each series in increasing on average. However, if the model is linear trend plus random factor then the variability in the observations is accounted for without the need to introduce a correlation.

This leads us to another correlation that is always important to consider whether or not it is statistically significant. We term it volatility correlation, the correlation between the unpredictable component of the data. This correlation is very important for understanding the diversification credit between multiple portfolios and it will be examined in detail below. However, first let us review correlation and its different types as found in the industry.

1.3. What is correlation?

When we speak of correlation, we may mean correlation between data or correlation between random variables that model the data. In both cases correlation measures how well can one variable can be modeled as a linear function of another. The closer the correlation coefficient is to 1, the better can one variable be represented as a linear function with positive slope of another. Similarly, the closer the correlation coefficient is to -1, the better can one variable be represented as a negatively sloping linear function of the other variable. A coefficient which is close to 0 indicates very weak linear relationship between the two variables.

1.4. Types of Correlation

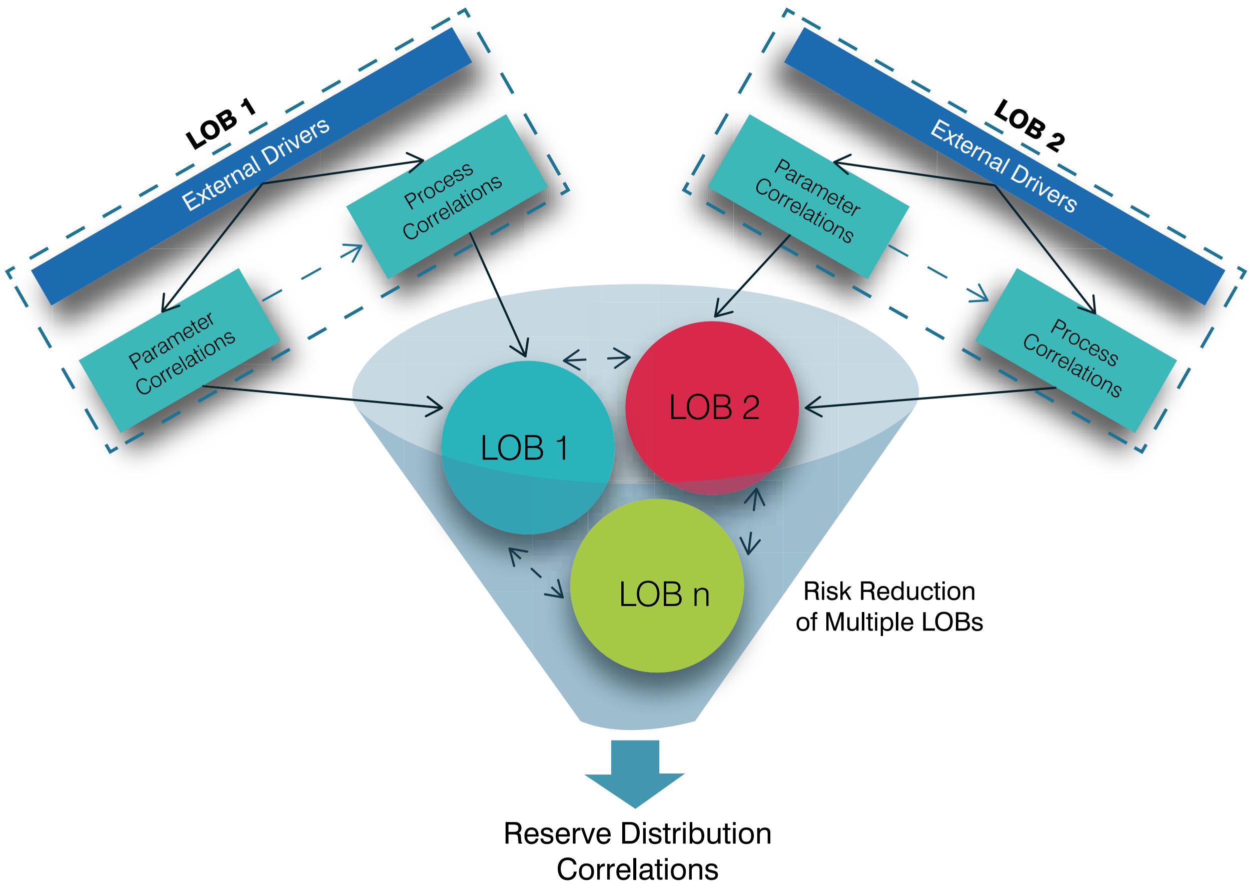
There are three types of correlations between LoBs: process (volatility) correlation, parameter correlation (which is related to common drivers as well as the volatility correlation), and reserve distribution correlation.

Perhaps the most important of the three, volatility correlation, is the correlation in the pure volatility component of the liabilities. This is measured from the data after all trends have been properly accounted for and it refers to losses in two lines of business tending to both, exceed or fall below the predicted values together.

This will directly impact the diversification credit of the business and thus one can easily see how this correlation is of significant concern. Indeed it is said that the correlation that matters is the correlation in the volatility component of forecasts.

The second type, parameter correlation is the correlation between the random variables representing the parameters of the model for losses runoff. This correlation can be influenced by common drivers but also by process correlation because estimation of model parameters depends on data subject to correlated random effects.

Reserve correlations are similarly correlations between the random variables that give the distribution of the reserves. As well as depending on process volatility, they depend on parameter correlations and parameter uncertainty: higher parameter uncertainty results in higher reserve correlations.



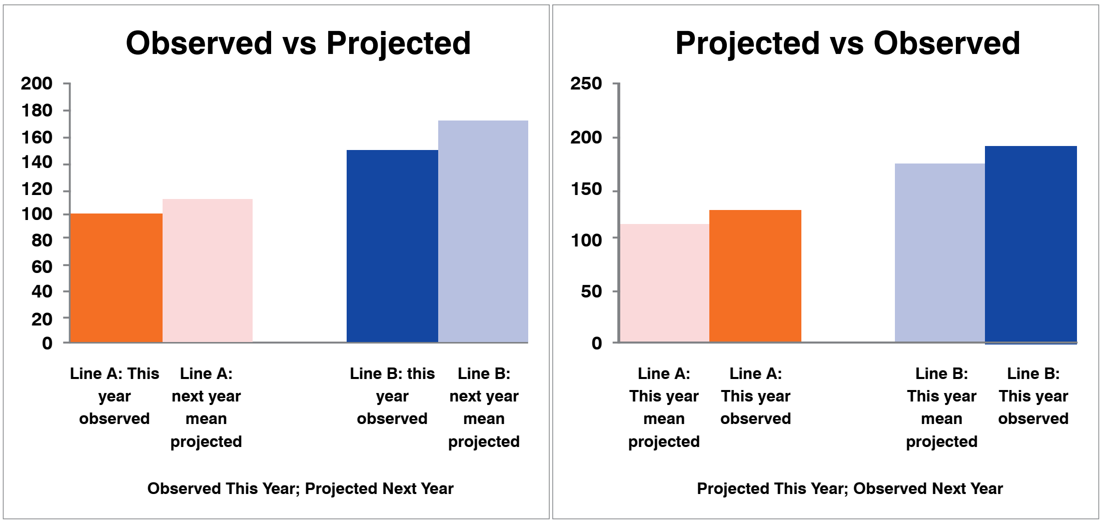
1.5. Example 1: Evidence of correlation?

Consider the example below. If our model for Line A predicts a 12% increase in mean loss for next year over the year just ended, and our model for Line B predicts a respective 14% increase, this is simply a statement about the similarity of the two models. The forecast mean losses are our best estimates, they are what we are planning on.

If the predictions are borne out, they may be taken as some evidence of a common driver of the increase in both lines of business or more importantly, evidence of the correlations between the parameters that describe both models, something that will need to be further tested.

Now suppose Line A has a forecast mean loss of 110 for next year and Line B a forecast mean loss of 150, and at the end of that year the actual losses for A and B are 120 and 180 respectively (Projected vs Observed below). This is evidence that lines A and B have positive volatility correlation with respect to the given model. To put it another way, in the presence of a positive volatility correlation, this kind of a result, where both A and B experience a shortfall (or an overrun) compared to the best prediction in the same year, is more common than not.

These two situations that look superficially similar, but only one of them provides evidence for correlation.



In the case illustrated on the left both lines project mean losses for the coming year as higher than the observed losses for this year.

Can this be taken as evidence for the correlations between projections?

First we must be clear that the projections here are given as means, i.e. single values which together with this year observed, can be plotted as two points on a plot showing the relation between these two lines. In this framework it is meaningless to speak of correlation.

If, however, we consider the projections as random variables, we need to look at the underlying data that the projections are based on. More specifically we need to look at the volatility around the projections, rather than whether the projected means themselves align. This leads to the concept of reserve correlation.

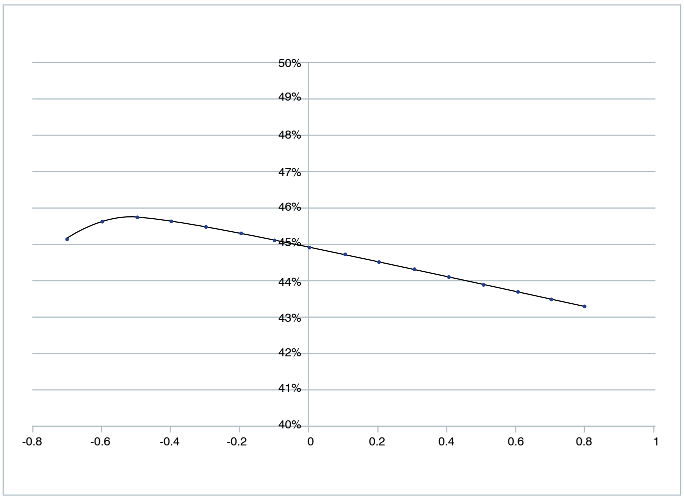
1.6. Volatility correlation

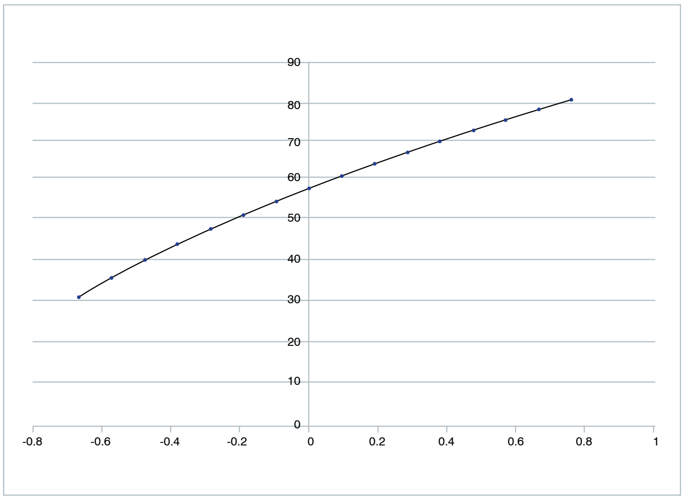
To illustrate this further, when at the end of the year we compare the reserves in two or more LoBs based on our previous projections with the observed losses, we are faced with either a shortfall or an overrun. Positive volatility correlation is the tendency for these unpredictable components to fall the same way, i.e. both are shortfalls or both are overruns.

The existence of a positive correlation in the forecast volatility (reserve correlation) between lines implies that the combined risk fund which covers for losses above the reserved amount, will, on average, experience larger draws than it would if the lines were uncorrelated. Presence of positive volatility correlation leads to a reduction in diversification credit.

1.7. Example 2: Effect of Volatility correlation

Projected mean losses for Line A are 100M with a coefficient of variation (CV) of 0.4, Line B mean losses are 150M with a CV of 0.3. Distributions are taken to be log normal.



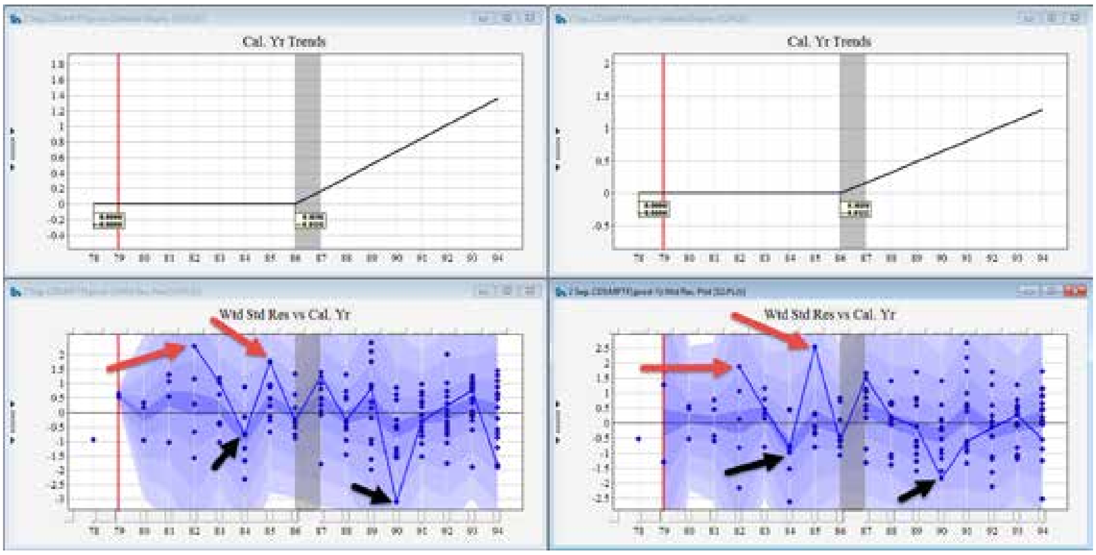


The probability of needing to draw on the risk fund is around 45% (it actually drops slightly for increasing correlations above -0.5), but given that there is such a draw, the average size of it increases sharply with correlation, from 58M when correlation = 0 to 80M when correlation = 0.8. Thus in the event of needing to draw on the risk fund, the positively correlated lines of business will require significantly greater capital than uncorrelated or negatively correlated ones.

1.8. How to determine volatility correlation

How can we know that two long-tail lines are correlated? The first step is de-trending the data. This can be understood as smoothing the data down to a pattern of statistically significant trends and then subtracting these trends from the data, the differences being known as “residuals” – the random deviations from the trend, the volatility we have been speaking about.

This is accomplished in the Probabilistic Trend Family (PTF) modeling framework by the placement of parameters at identified change points along the development, accident or calendar axes. Once this is done, the residuals for each of the two lines should appear to be randomly scattered around zero. We can then carry out a standard statistical test for correlations in the residuals.



The plots above show residuals by calendar year for two correlated segments, S1 and S2, beneath their respective calendar year trends. In this case both segments have the same trend pattern, zero from 1978 to 1986, followed by a non-zero trend from 1986 to 1984. This common pattern suggests the presence of common drivers.

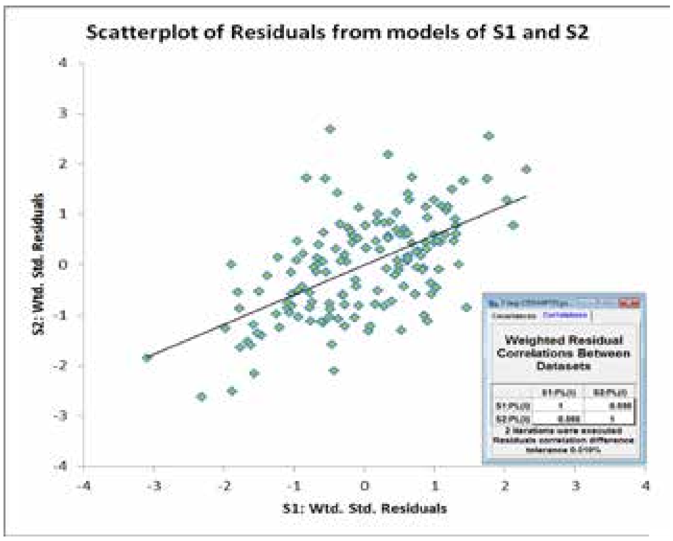
We will look more closely at this later, but in any case, this effect on each line of business is accounted for by the models. What remains to be accounted for is the potential presence of volatility correlation, i.e. a pattern of common deviation from the models.

The blue line joins the residuals (observed minus expected) corresponding to the observations occurring in accident year 1982. The losses in calendar years 1984 and 1990 (black arrows) are lower than expected for this accident year in both lines of business. In calendar years 1982 and 1985 (red arrows), the losses are higher than expected in both lines. The blue line trace shows that the residuals for the two segments are more likely to be both, positive or both negative, rather than for one to be positive and the other negative.

In other words, relative to our best models for the two lines, there is a tendency for both segments to either fall short of or to exceed expectations in the same years.

A scatter plot of the residuals (right) shows a linear relationship. As S1 residuals increase so do S2 residuals.

The volatility in the lines is positively correlated r=0.598.



The remainder of this document consists of a series of case studies illustrating the commonly found example of spurious correlation, the various types of common drivers as well as volatility correlation, and their impact on reserve risk, underwriting risk, and the combination of reserve and underwriting risk.

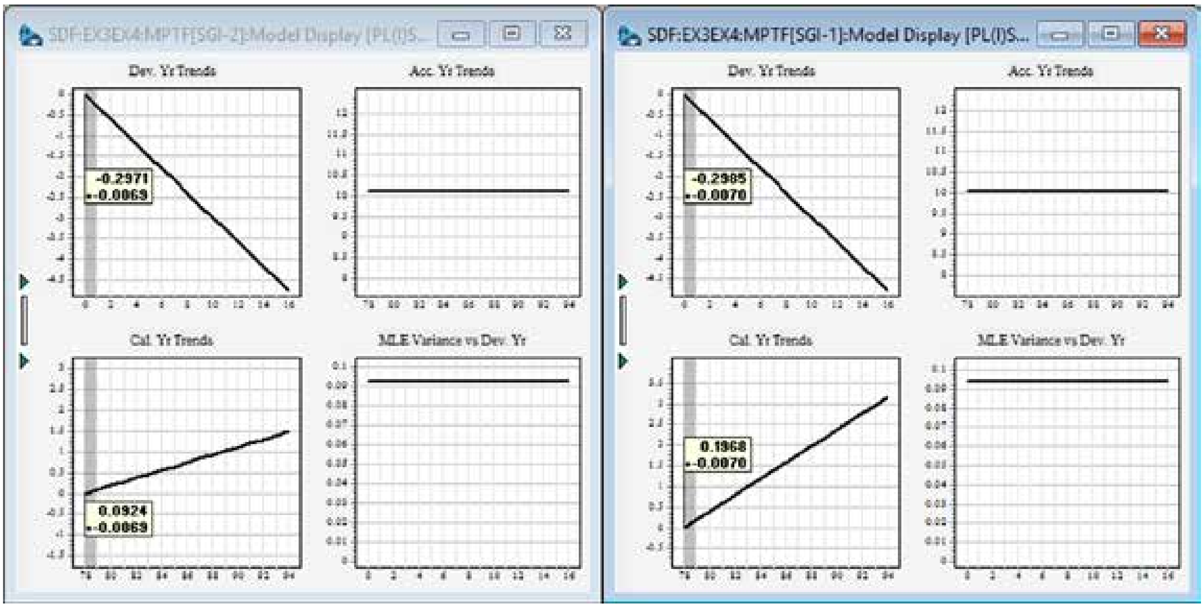
2. Case study: Spurious correlation - an artefact of the wrong model

This section discusses spurious correlation as a result of an incorrect model for the data.

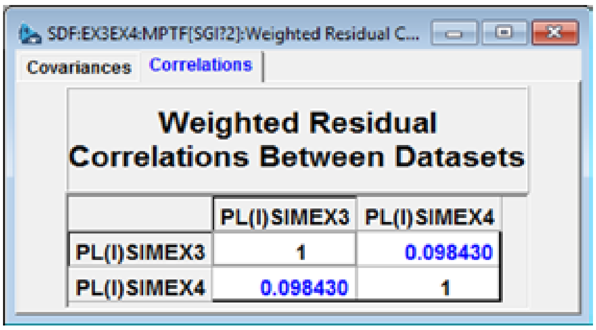
To illustrate this point, two LoBs are simulated independently so that one LoB has a calendar year trend of 10% and the other of 20%. Since the processes are simulated independently, one does not expect any relationship in the volatility component of the data. And indeed given the correct models for the underlying data process, this correlation should not be statistically significant.

However, if an incorrect model is used, one that, for instance, does not fit the calendar year trends, then spurious correlation between the residuals can be observed. This correlation is meaningless since it arises as a result of calendar year trends being present in both LoBs which were not accounted for by either model. The correlation is an artefact of the models which do not fit all the trends in the data.

The two models correctly fitted to the data are shown below - the parameters estimated from the data are very close to the true parameters (as we would expect for simulated data).



In the event that the trends are described accurately (as above), the volatility correlation between the two segments is expected to be insignificant and close to zero. If we estimate this correlation, this is exactly what we find (note the blue font which indicates statistical insignificance).

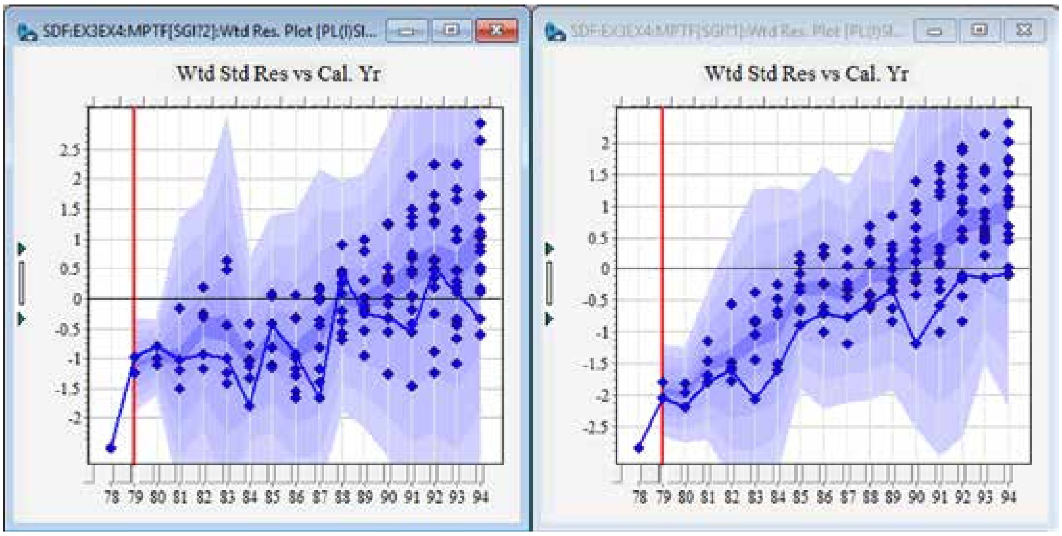

SDF:EX3EX4:MPTF[SGI?2]:Weighted Residual C...

Covariances | Correlations

Weighted Residual Correlations Between Datasets

	PL(I)SIMEX3	PL(I)SIMEX4
PL(I)SIMEX3	1	0.098430
PL(I)SIMEX4	0.098430	1

In contrast, consider if the respective calendar year trends are not fitted to the data. The calendar year residual displays are shown below to emphasize the correlation metrics (first accident year marked).



SDF:EX3EX4:MPTF[SGI?2]:Weighted Residual C...

Covariances | Correlations

Weighted Residual Correlations Between Datasets

	PL(I)SIMEX3	PL(I)SIMEX4
PL(I)SIMEX3	1	0.699335
PL(I)SIMEX4	0.699335	1

The residuals are clearly highly correlated (but they do not come from the same distributions across the calendar years). The residual is clearly a function of time in that early calendar years are highly likely to contain negative residuals (both LoBs) and more recent calendar years (post 1988) positive residuals.

If we measure this correlation we find it to be 0.699. The correlation is both high and statistically significant, however, this result is purely a reflection of calendar trends being present in both datasets which are not described by either model. Recall that the data were simulated independently. In this case the correlation is spurious and simply measures trend structure which is not captured.

3. Case study: Common calendar year drivers

Three examples of common calendar year drivers are considered.

The first example describes Gross data and Net of Reinsurance data for an E&O D&O line. The trend structure is almost the same, especially along calendar years. That is, the trend changes occur in the same periods, indicating potential common drivers. The resulting volatility (process) correlation and parameter correlation are high.

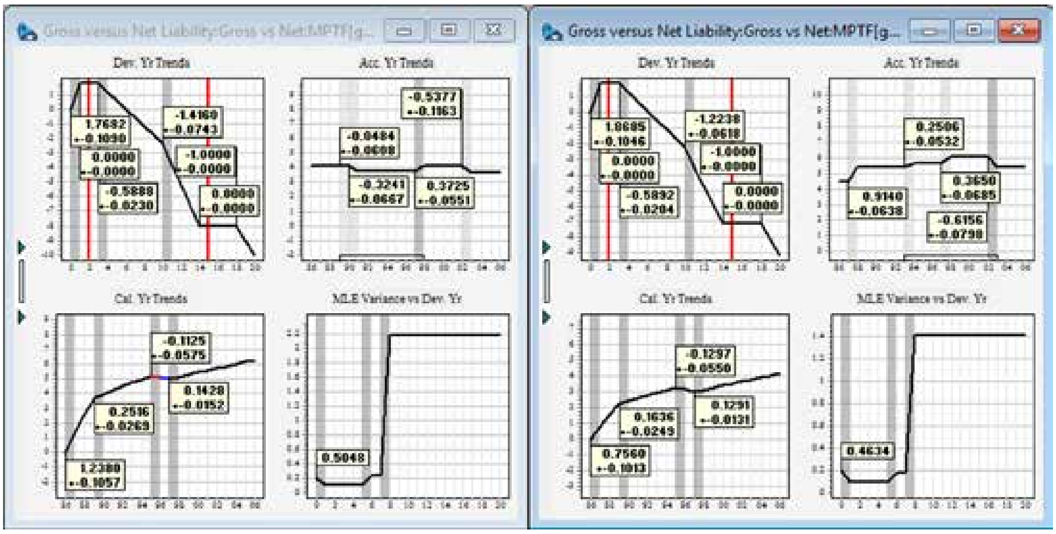
The second example illustrates the common trend structure identified in layered data - again, especially calendar year trends. Volatility correlations and parameter correlations are high.

Finally, the last study in this section demonstrates common calendar year drivers for two LoBs (same line, different states) where volatility correlation is also statistically significant but not high.

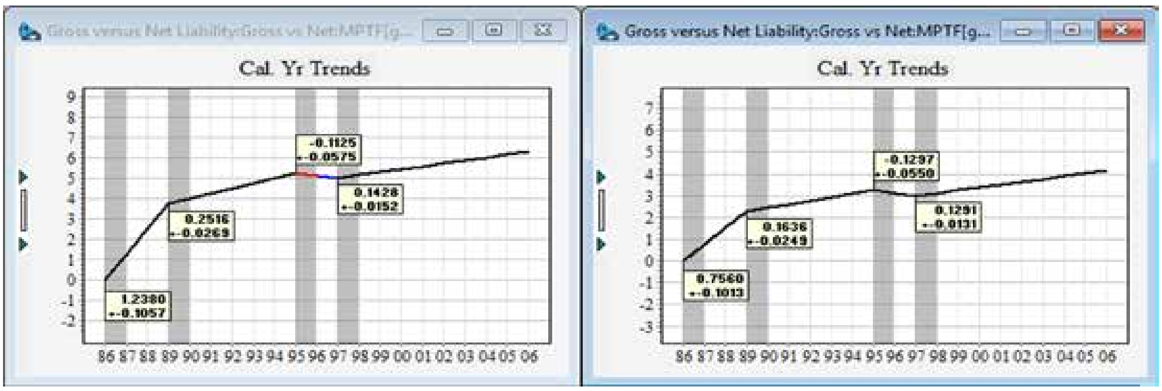
Although not comprehensive, the above list serves as a solid basis to the concept of discussing volatility risk, common calendar year driver risk, and understanding the difference between volatility risk (as a result of process correlations) and common calendar year driver risk.

3.1. Gross versus Net of Reinsurance

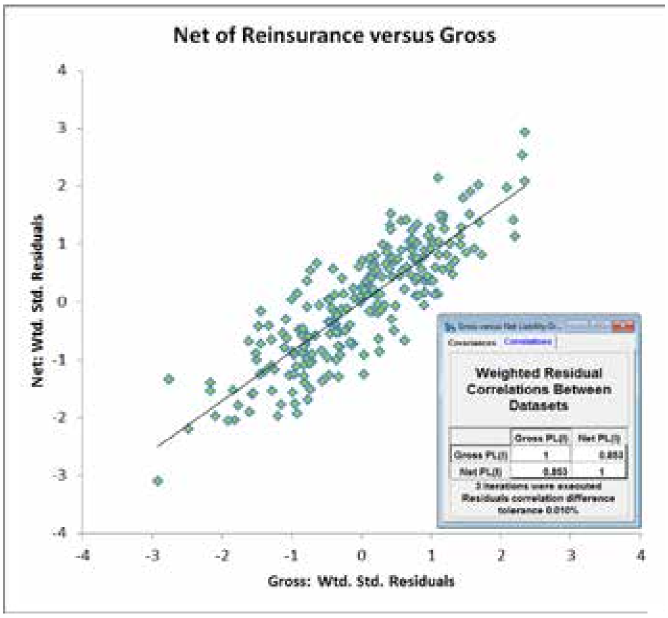
In this first example of detecting and quantifying common calendar year drivers, Gross versus Net of Reinsurance for E&O and D&O data, common calendar year drivers are expected to be found. Net of Reinsurance is a subset of Gross and therefore common features are to be expected, but are not always found. Trends, especially calendar and accident, are closely related. The comparable models for Gross (left) and Net of Reinsurance (right) are shown below.



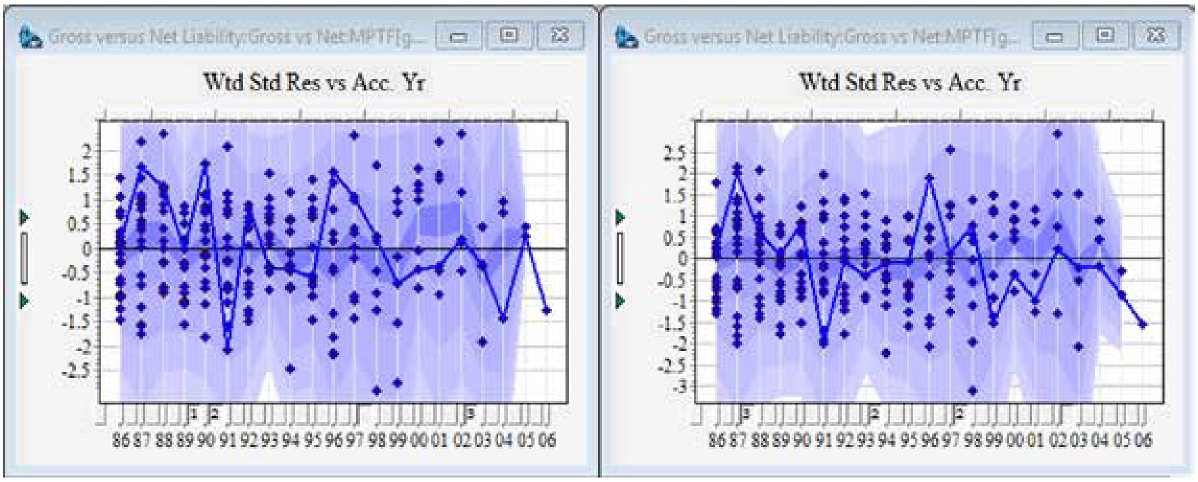
The model trends are very similar; trend and volatility changes usually coincide. The critical trends in common are the calendar year trends (below) and accident year level changes. Common calendar year drivers are clearly visible as the trend changes occur at the same point.



Similarly, the process volatility is closely related. A scatter plot of the residuals, from the respective Gross and Net of Reinsurance models, exhibits a clear (linear) relationship; a correlation of 0.854.



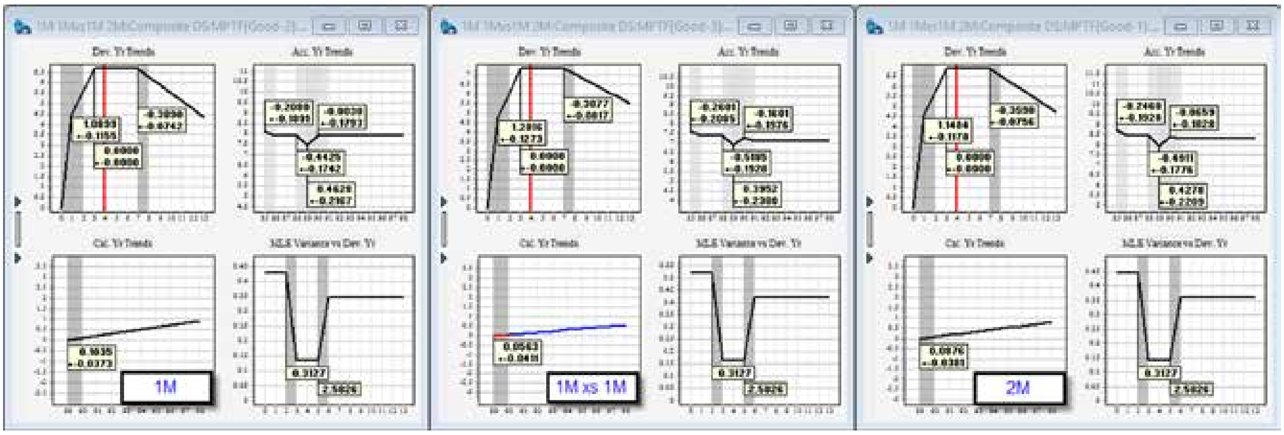
The residuals by accident year traced for the last calendar year are clearly correlated; when a value in a year is low/high in one segment it is usually low/high in the other segment also at the same time.



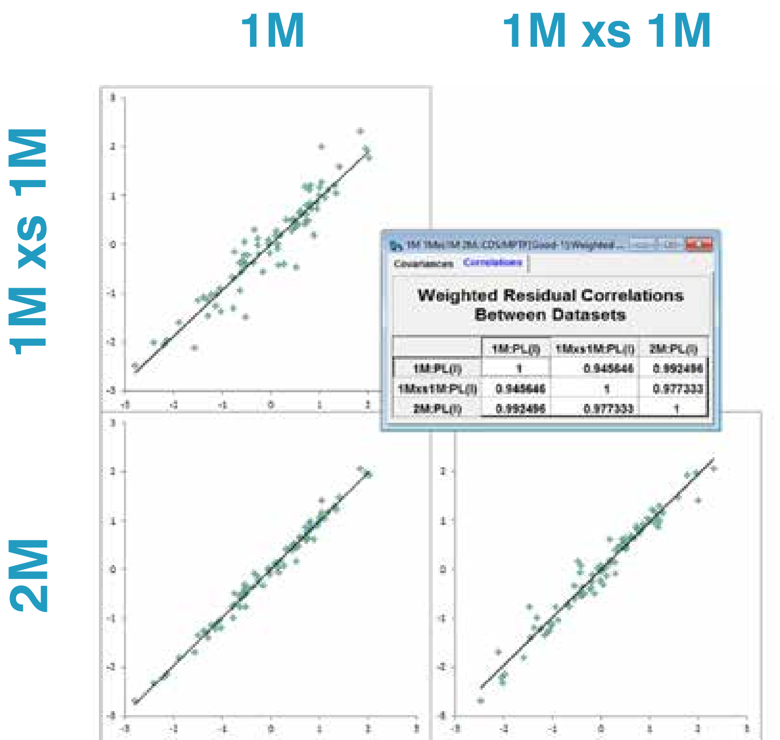
This case study illustrates the worst possible relationship between two Lines of Business (if in fact they were separate lines); namely common drivers (accident year and calendar year) and volatility correlation. In almost all cases, this proximity of relationship is only expected when the LoBs analyzed are in fact subsets of one another.

3.2. Layers: Limited to 1M, 1Mxs1M, and Limited to 2M

In this next example, data are split into three layers - paid losses with each individual loss limited to 1M, paid losses with individual losses in excess of 1M with the excess limited to 1M (1Mxs1M), and paid losses limited to 2M. Similar trend structure and common drivers are expected since 1M + 1Mxs1M = 2M.

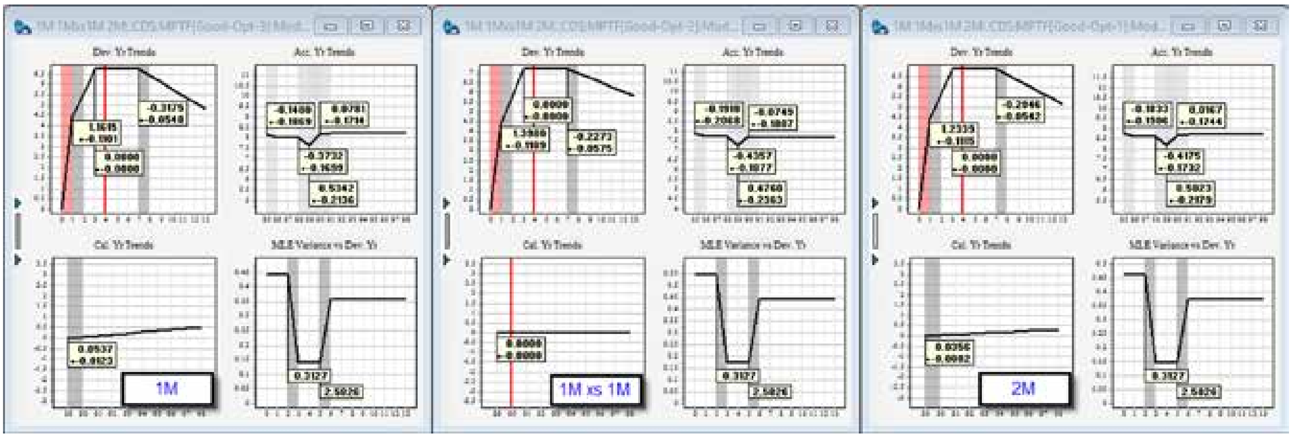


The Layer 1M has a higher inflation rate than 2M, and 1Mxs1M has inflation rate that is statistically insignificant. If the only available array is 1Mxs1M then it would be prudent not to set the inflation to zero, as process volatility is high. One could argue that positive inflation is present, and we have a very uncertain estimate of it (5.63%+-4.11%). If any one of the other two arrays is available the very high process (volatility) correlation between the layers reduces parameter uncertainty in the composite model. In this case there is convincing evidence that inflation for 1Mxs1M is zero.



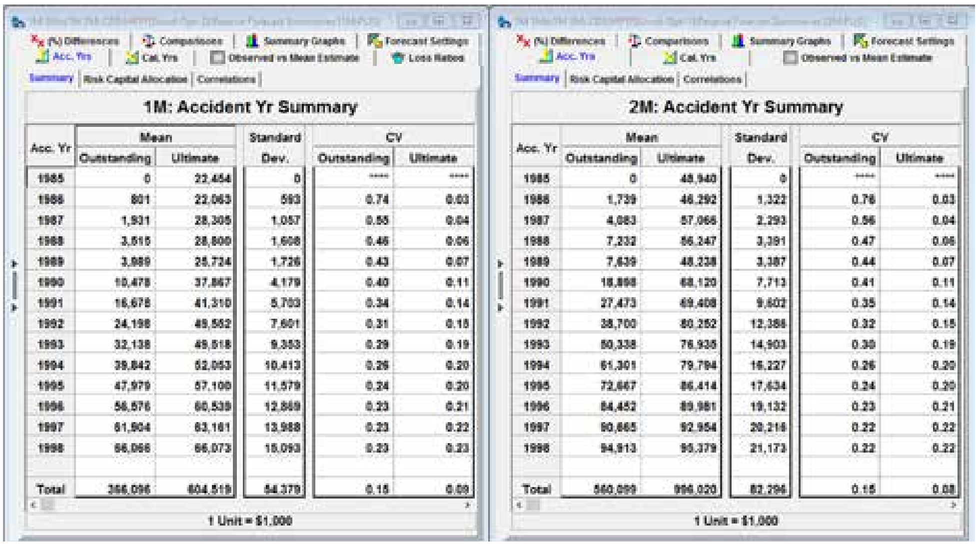
The three residual displays by calendar year for the layers exhibit very high process correlation.

When the composite model is optimized some trends in the data are found to be common (red bars / lines indicate common parameters) between the layers and for 1Mxs1M the calendar year trend is zero.



Having recognized the same trend structure (and common drivers) in the three layers, the efficiency of the reinsurance program (in terms of reducing risk capital as a proportion of the mean reserve) can be assessed.

Indeed, the CV of the aggregate reserves for 1M and 2M are the same (0.15). That means that both reinsurance (ceding) programs are equally capital efficient!



1M: Accident Yr Summary

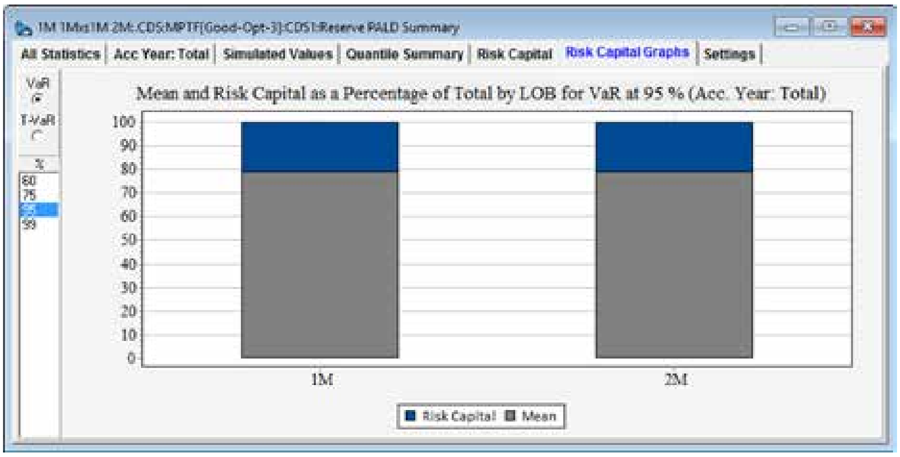
Acc. Yr	Mean		Standard	CV	
	Outstanding	Ultimate	Dev.	Outstanding	Ultimate
1985	0	22,454	0	****	****
1986	801	22,063	593	0.74	0.03
1987	1,931	28,305	1,057	0.55	0.04
1988	3,515	28,800	1,608	0.46	0.06
1989	3,989	25,724	1,726	0.43	0.07
1990	10,478	37,867	4,179	0.40	0.11
1991	16,678	41,310	5,703	0.34	0.14
1992	24,198	49,552	7,601	0.31	0.15
1993	32,138	49,518	9,353	0.29	0.19
1994	39,842	52,053	10,413	0.26	0.20
1995	47,979	57,100	11,579	0.24	0.20
1996	56,576	60,539	12,869	0.23	0.21
1997	61,904	63,161	13,988	0.23	0.22
1998	66,066	66,073	15,093	0.23	0.23
Total	366,096	604,519	54,379	0.15	0.09

1 Unit = $1,000

2M: Accident Yr Summary

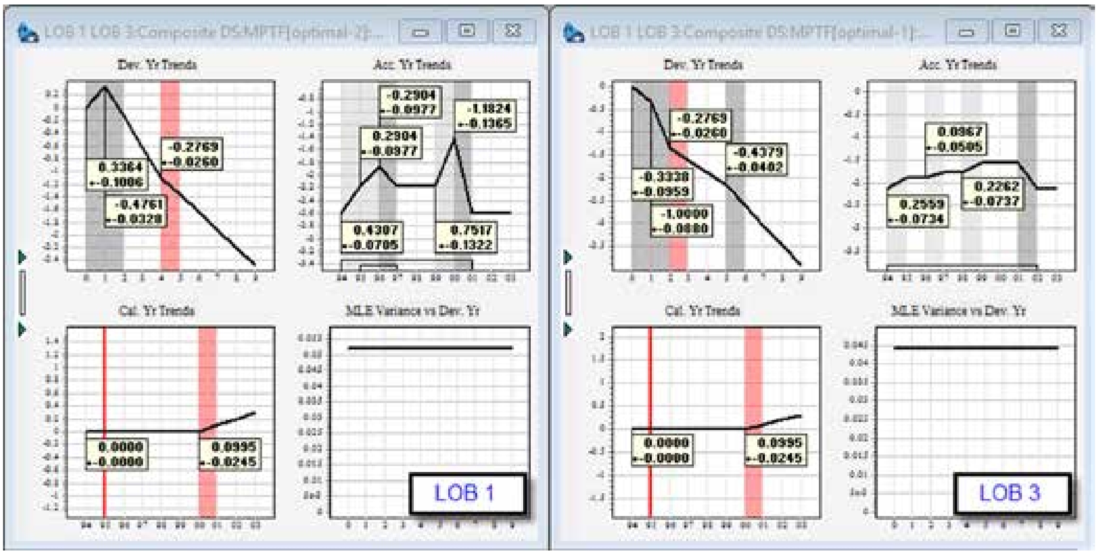
Acc. Yr	Mean		Standard	CV	
	Outstanding	Ultimate	Dev.	Outstanding	Ultimate
1985	0	48,940	0	****	****
1986	1,739	46,292	1,322	0.76	0.03
1987	4,083	57,066	2,293	0.56	0.04
1988	7,232	56,247	3,391	0.47	0.06
1989	7,639	48,238	3,387	0.44	0.07
1990	18,898	68,120	7,713	0.41	0.11
1991	27,473	69,408	9,602	0.35	0.14
1992	38,700	80,252	12,386	0.32	0.15
1993	50,338	76,935	14,903	0.30	0.19
1994	61,301	79,794	16,227	0.26	0.20
1995	72,667	86,414	17,634	0.24	0.20
1996	84,452	89,981	19,132	0.23	0.21
1997	90,665	92,954	20,216	0.22	0.22
1998	94,913	95,379	21,173	0.22	0.22
Total	560,099	996,020	82,296	0.15	0.08

1 Unit = $1,000



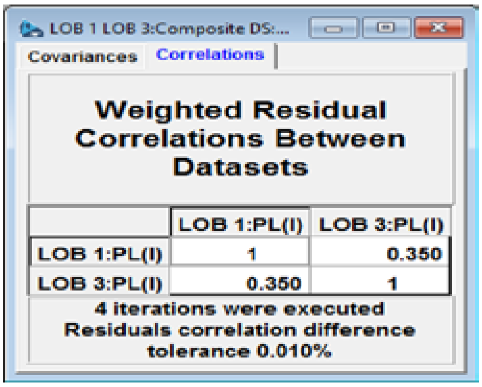
3.3. LoB 1 and LoB 3

The optimal model for the two LoBs, LoB 1 and LoB 3, is shown below.



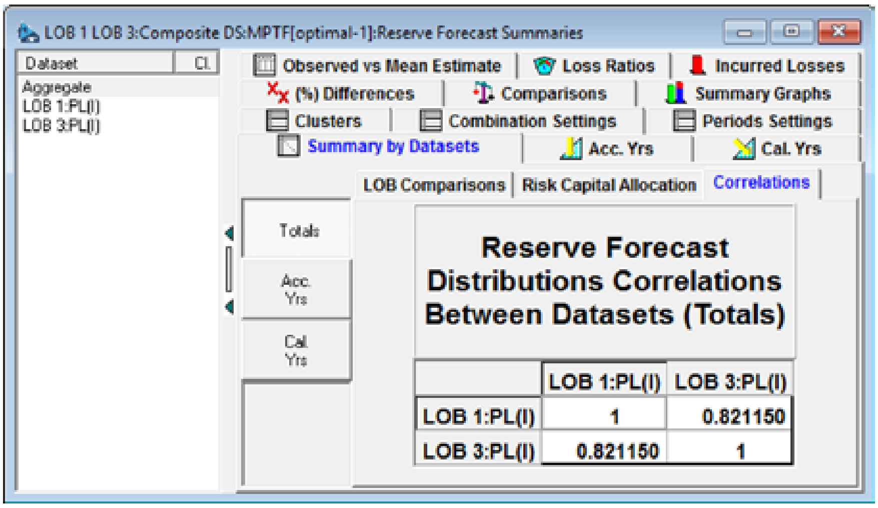
As with the Layers example, this model shows common calendar year drivers affecting both LoBs since the changes in calendar year trend occur at the same time. Synchronous changes in trend are a key indicator of common drivers.

Furthermore, the LoBs not only have common drivers, but the process volatility between the LoBs is also correlated as illustrated below.



	LOB 1:PL(I)	LOB 3:PL(I)
LOB 1:PL(I)	1	0.350
LOB 3:PL(I)	0.350	1

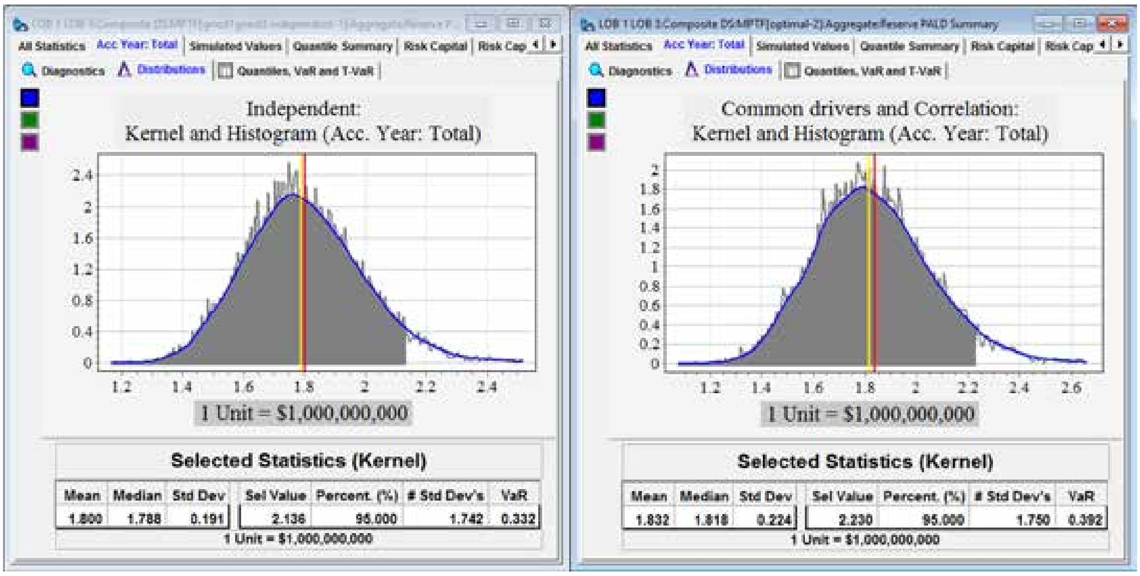
There is a reduction in risk diversification credit from writing these two lines by way of the common parameters and process correlation. The reserve correlations (0.821) are much higher than the process correlation. This unusual case is a result of the most recent calendar year trends for the two segments LoB 1 and LoB 3 being set to be the same for each line in the future.



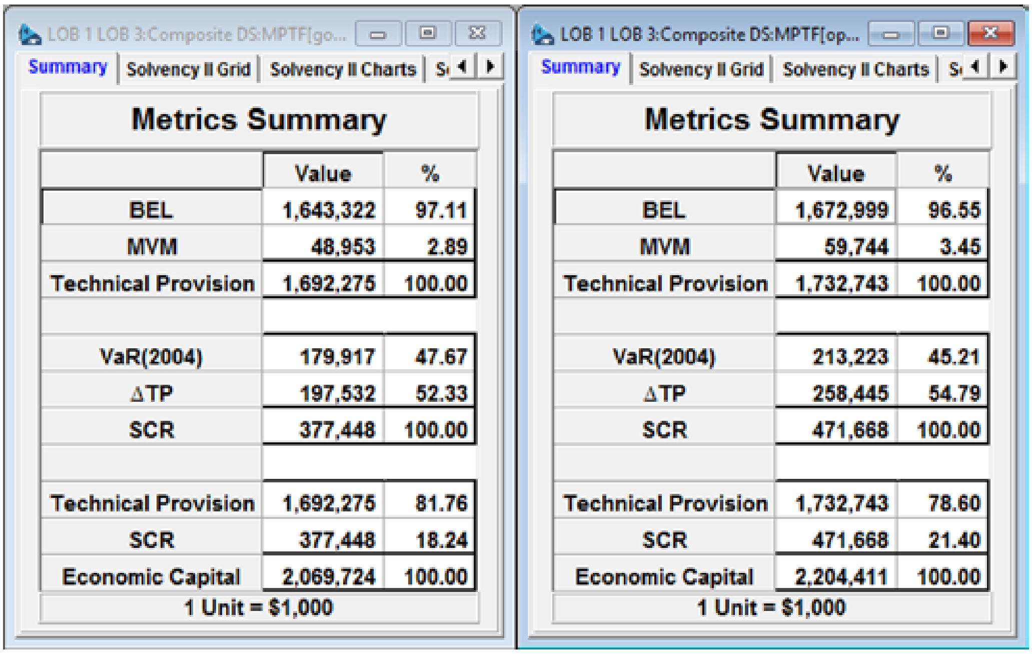
	LOB 1:PL(I)	LOB 3:PL(I)
LOB 1:PL(I)	1	0.821150
LOB 3:PL(I)	0.821150	1

To see the effect of the inclusion of the correlation and common drivers on the risk capital, the risk capital is calculated for two models: a model where the correlation is explicitly set to zero and no common trends are applied - that is, only the data in each LoB is used to estimate the trends (development and calendar) - and a model where the common trends and process volatility is included. The risk capital requirement assuming independence is then compared with the risk capital calculation where the process correlation and common drivers are applied.

For comparison, the value-at-risk at the 95th quantile (percentile) is calculated for both models: independent (left) and incorporating the common drivers and process correlation (right). Comment on figures eg: 1.8B vs 1.83B.



Similarly, the Solvency II one year risk horizon metrics (see brochure: Solvency II - One-year and ultimate year horizons for long tail liabilities) are calculated for the model for the two lines assuming independence (left) and the optimal model with both common calendar year, development year trends, and process correlation (right).



Metrics Summary (left)

	Value	%
BEL	1,643,322	97.11
MVM	48,953	2.89
Technical Provision	1,692,275	100.00
VaR(2004)	179,917	47.67
ΔTP	197,532	52.33
SCR	377,448	100.00
Technical Provision	1,692,275	81.76
SCR	377,448	18.24
Economic Capital	2,069,724	100.00

1 Unit = $1,000

Metrics Summary (right)

	Value	%
BEL	1,672,999	96.55
MVM	59,744	3.45
Technical Provision	1,732,743	100.00
VaR(2004)	213,223	45.21
ΔTP	258,445	54.79
SCR	471,668	100.00
Technical Provision	1,732,743	78.60
SCR	471,668	21.40
Economic Capital	2,204,411	100.00

1 Unit = $1,000

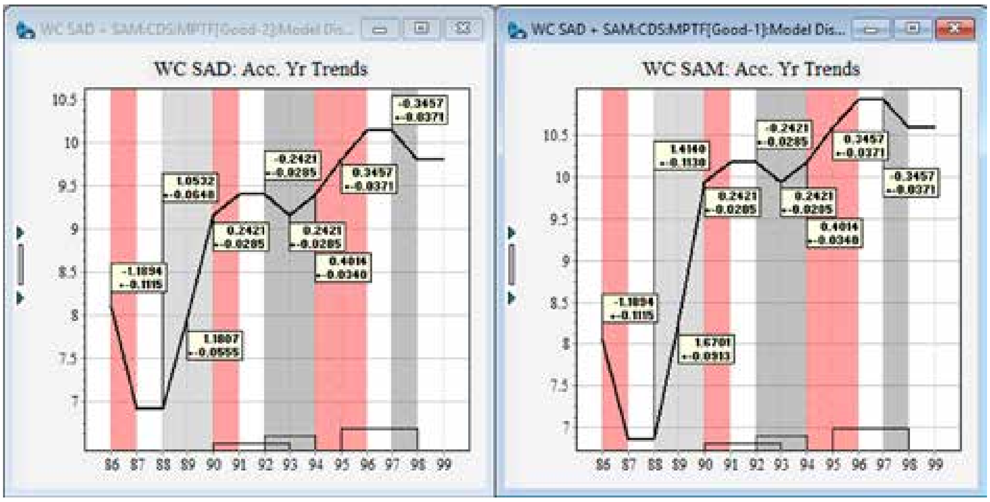
The required technical provision has increased from 1.692B (assuming independence: left) to 1.733B (common parameters and process correlation: right). The additional 41M (approximately 2.4%) is the penalty for the lack of risk diversification.

4. Case study: Common accident year drivers

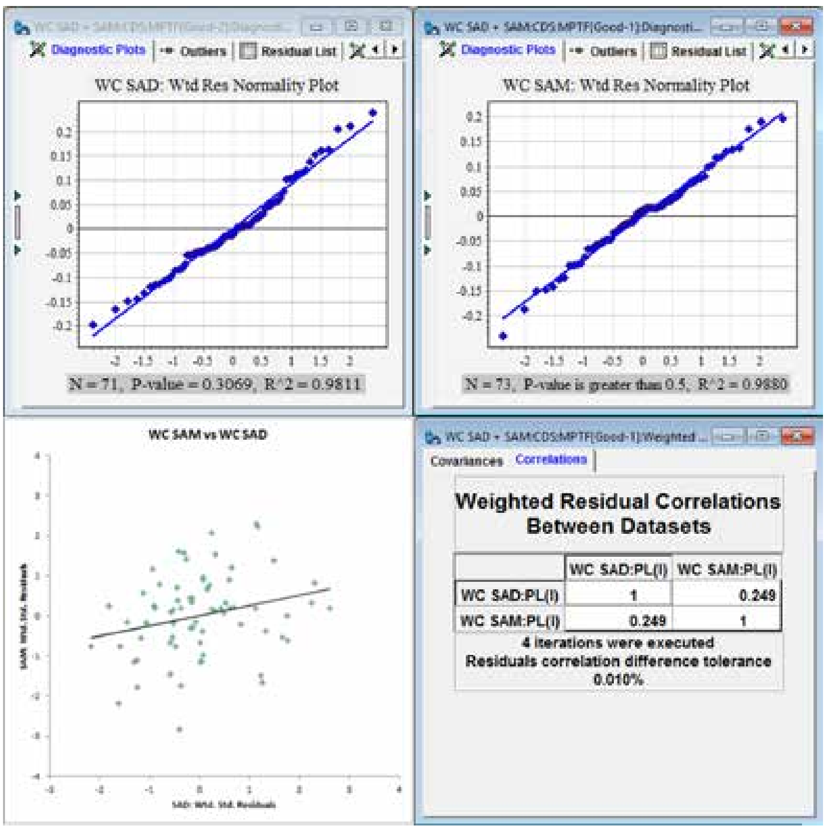
In this case study, common accident year drivers are demonstrated in the context of two segments of Worker's Compensation: SAD and SAM. The two segments have changes in accident level in common and also demonstrate synchronous changes in level. The synchronous changes in parameters are a critical component of identifying common drivers whether by accident or calendar year.

4.1. Worker's Compensation Segments: WC SAD and WC SAM

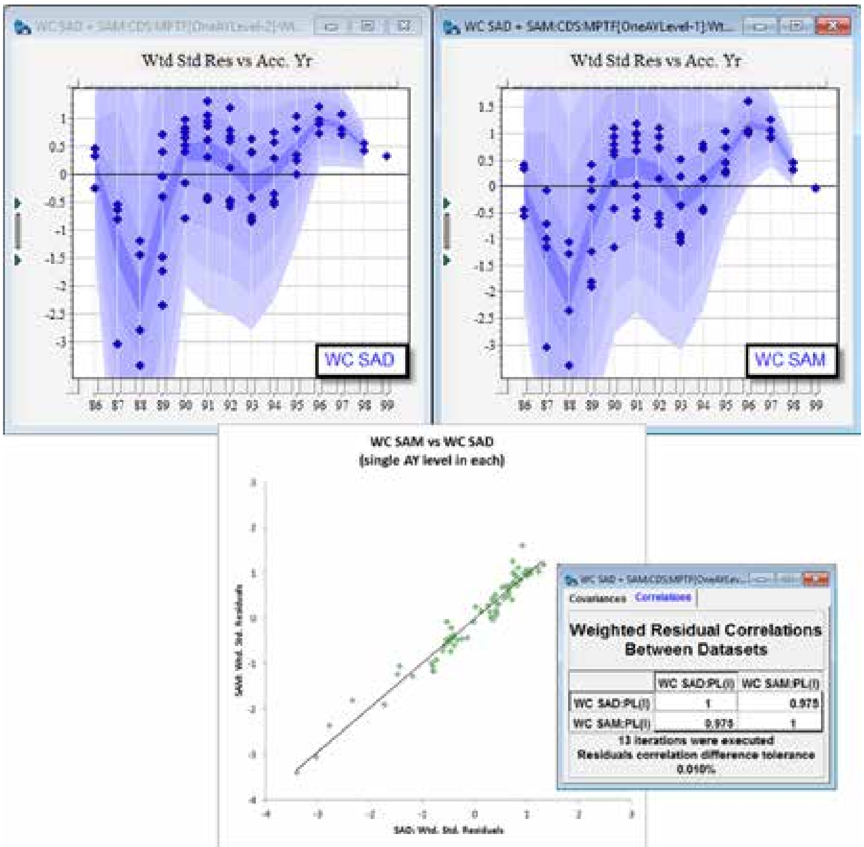
Consider the following two segments of Worker's Compensation written in California: SAD (left) and SAM (right). The red bars indicate common parameters between the segments. Although the calendar and development year parameters vary slightly, the accident year parameters move synchronously thus making the mean ultimates vary synchronously (but this is not correlation).



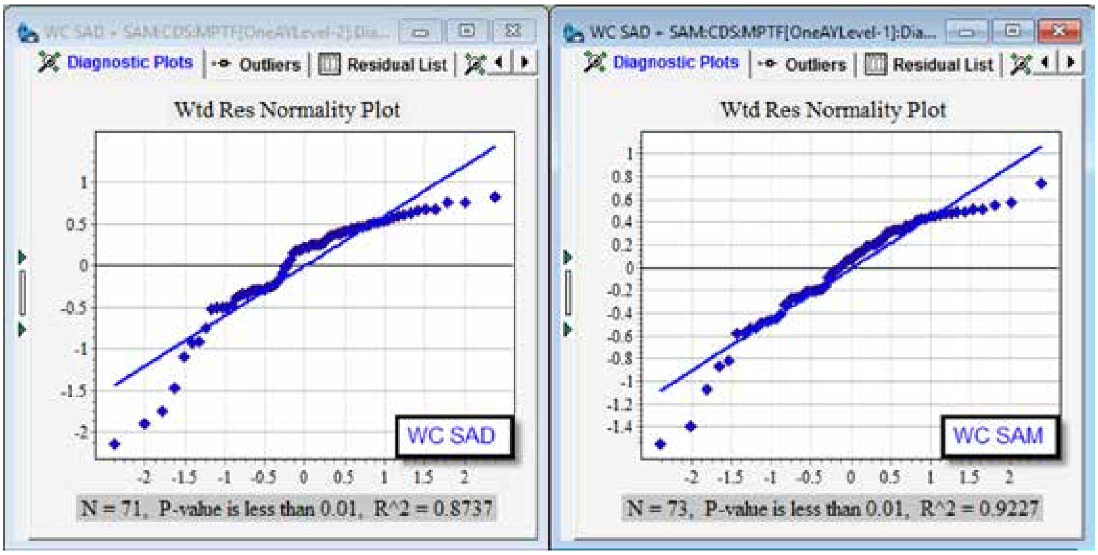
Both sets of residuals can be assumed to originate from a normal distribution, so the process correlation (0.249) below is the volatility correlation between two normal distributions.



If the common accident year trends are ignored and the average accident year level fitted to both segments, then a very high spurious correlation measure of 0.975 is obtained.



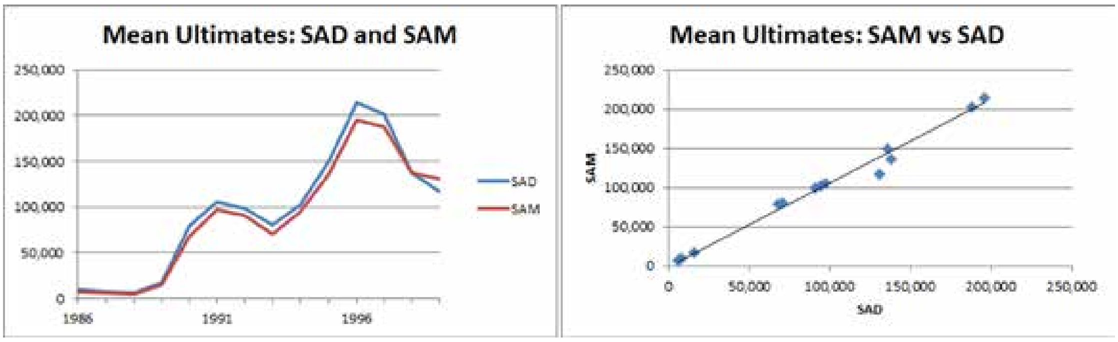
Above, the residual displays with scatter plot for SAD and SAM are shown for a model which does not describe the accident year changes. The spurious correlation (0.897) is very high, but given the two residual plots above it showing patterns in residual movements, it is clear that what the correlation is largely picking up is common under and over-fitting by the model (indicated by the red arrows) rather than genuine process correlation.



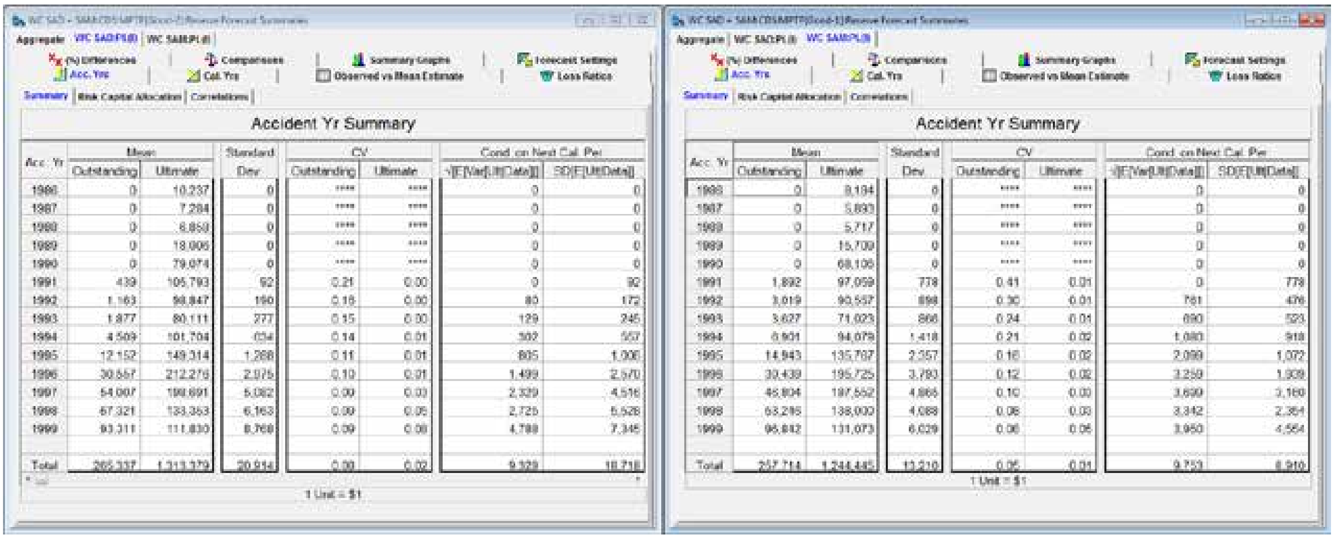
The high correlation is an artefact of a poor model which does not fit all the trends in the data. Instead, the correlation reflects the commonality of the trends rather than process (volatility) correlation.

However, the correlation is no longer the correlation measured between two (normal) distributions; the means vary over time in both sets of residuals. For instance, residuals for accident years 90-93 have a positive mean, whereas 87-88 have a negative mean.

In the correct model (page 16), with the accident year levels correctly fitted, the correlation between the segments is predominantly the volatility correlation. The result of the fitted accident year levels is that the mean ultimate losses (by accident year) move synchronously (common drivers), however the risk factors arising from volatility are not as severe (volatility correlation is only 0.25). The accident year levels moving together result in a much stronger relationship than volatility correlation.



These features are also illustrated in the accident year summaries for each segment displayed below.



Accident Yr Summary

Acc. Yr	Mean		Standard	CV		Cond. on Next Cal. Per.	
	Outstanding	Ultimate	Dev	Outstanding	Ultimate	√[E[Var[Ult|Data]]]	SD[E[Ult|Data]]
1986	0	10,237	0	****	****	0	0
1987	0	7,284	0	****	****	0	0
1988	0	6,850	0	****	****	0	0
1989	0	18,906	0	****	****	0	0
1990	0	79,074	0	****	****	0	0
1991	439	105,793	92	0.21	0.00	0	92
1992	1,163	98,847	190	0.16	0.00	80	172
1993	1,877	80,111	277	0.15	0.00	129	245
1994	4,509	101,704	634	0.14	0.01	302	557
1995	12,152	149,314	1,288	0.11	0.01	805	1,006
1996	30,567	212,276	2,975	0.10	0.01	1,499	2,570
1997	54,007	198,691	5,082	0.09	0.03	2,329	4,516
1998	67,321	133,353	6,163	0.09	0.05	2,725	5,528
1999	93,311	111,830	8,768	0.09	0.08	4,788	7,345
Total	265,337	1,313,379	20,914	0.08	0.02	9,329	10,718

1 Unit = $1

Accident Yr Summary

Acc. Yr	Mean		Standard	CV		Cond. on Next Cal. Per.	
	Outstanding	Ultimate	Dev	Outstanding	Ultimate	√[E[Var[Ult|Data]]]	SD[E[Ult|Data]]
1986	0	8,184	0	****	****	0	0
1987	0	5,890	0	****	****	0	0
1988	0	5,717	0	****	****	0	0
1989	0	15,709	0	****	****	0	0
1990	0	68,106	0	****	****	0	0
1991	1,892	97,059	778	0.41	0.01	0	778
1992	3,019	90,557	898	0.30	0.01	761	476
1993	3,627	71,023	866	0.24	0.01	690	523
1994	6,901	94,079	1,418	0.21	0.02	1,080	918
1995	14,943	135,787	2,357	0.16	0.02	2,098	1,072
1996	30,438	195,725	3,793	0.12	0.02	3,259	1,939
1997	46,804	187,562	4,866	0.10	0.03	3,699	3,160
1998	63,286	138,030	4,088	0.06	0.03	3,342	2,364
1999	96,842	131,073	6,029	0.06	0.05	3,950	4,554
Total	257,714	1,244,445	13,210	0.05	0.01	9,753	6,910

1 Unit = $1

The mean ultimates move synchronously (left) and a graph of the mean ultimates of SAM versus the mean ultimates of SAD (right) shows an almost perfect linear relationship.

However, the reserve distribution correlation is only 0.086! The reserve correlation, which is calculated from the model, can be pictured as the correlation in the predicted losses not explained by the means – and therefore is the critical measure when evaluating risk diversification. Low reserve correlation is good news for risk diversification as it tells us that there is no evidence that the deviations from predicted ultimates (the means) will move in the same direction. Had the correlation been significantly positive, the risk of both lines exceeding the predicting means together would have given cause for concern.

As we have seen models that do not capture the trends in the three directions in the data may indicate spurious correlations and erroneous conclusions. It is also important that the weighted standardized residuals of each model can be regarded as a random sample from a (normal) distribution. This way, the process (volatility) correlation can be measured correctly.

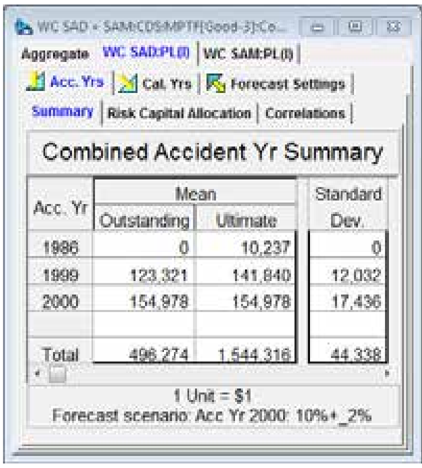
5. Case Study: Common accident year drivers and pricing future accident years

In this example we continue with the two segments of Worker's Compensation SAD and SAM.

These two lines demonstrate common accident year drivers. The impact of common accident year drivers must be considered when pricing future accident (or underwriting) years. The close relationship of the accident year parameters are considered in respect of future forecast assumptions.

The linear relationship in mean ultimates is important when forecasting future underwriting (accident) years. If the accident year level for one segment is expected to increase by 10%$\pm$ 2%, then the other segment is also likely to increase by 10% $\pm$ 2% in the same accident year. The relationship in the mean parameter estimates is not volatility (risk) correlation and does not indicate lack of diversification. The movement in means may be related to internal or external drivers - and risk exposure can be managed. Whereas the volatility correlation, if not specifically measured and accounted for, is not readily able to be connected to any internal or external drivers and not considered by the separate models of the LoBs.

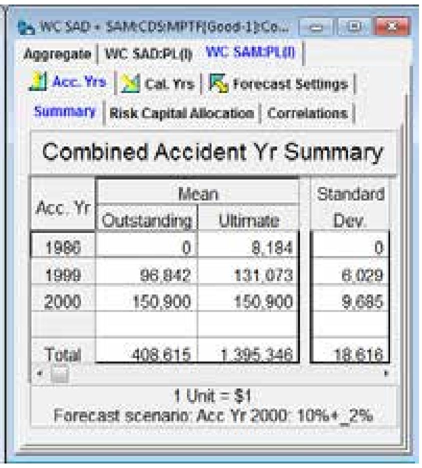
The synchronous movement in the accident year trends is readily observed in the model displays shown previously (page 16). The correlation between the mean accident year level parameters provides an idea of the closeness of the relationship and is measured at 0.995. This measure provides support that if a level change is expected to occur in one segment, then a corresponding level change is expected to occur in the other. This correlation measure is not able to be identified prior to analysis (identification of trends), nor does it necessarily imply the magnitude of the change in parameter levels are the same (although in this example the changes in mean level are essentially identical as a result of the constraints between the segments - it is a feature of the model). It does, however, emphasize the importance of adjusting accident year levels for both segments simultaneously.



Combined Accident Yr Summary

Acc. Yr	Mean		Standard
	Outstanding	Ultimate	Dev.
1986	0	10,237	0
1999	123,321	141,840	12,032
2000	154,978	154,978	17,436
Total	496,274	1,544,316	44,338

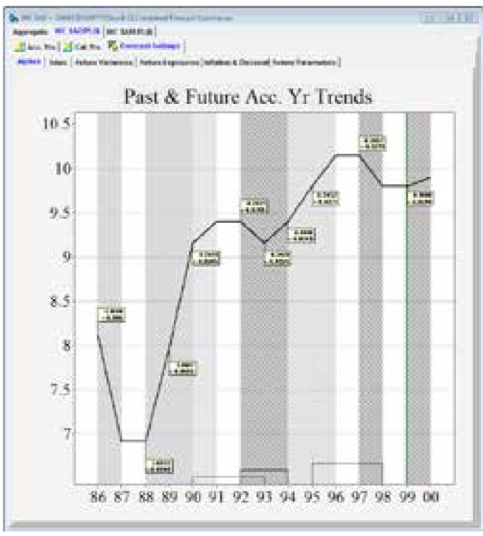
1 Unit = $1
Forecast scenario: Acc Yr 2000: 10%+_2%

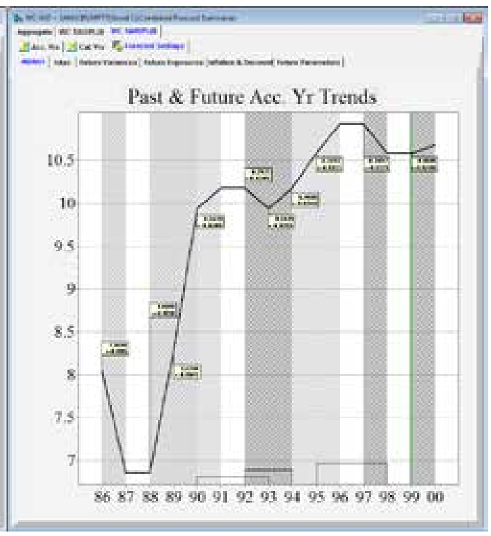


Combined Accident Yr Summary

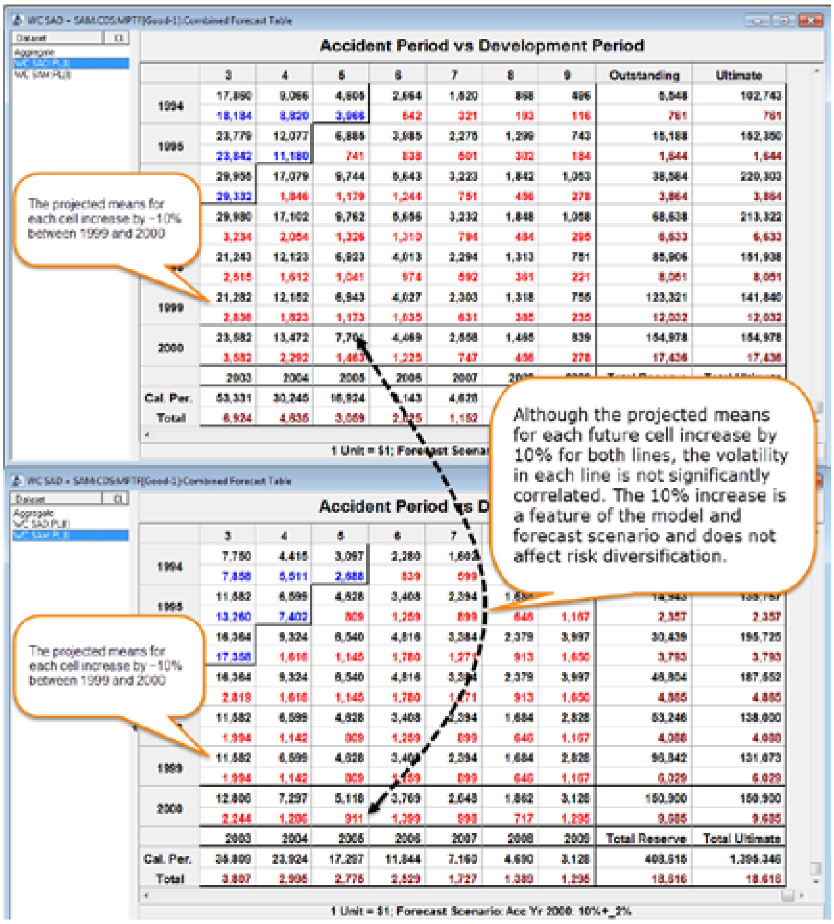
Acc. Yr	Mean		Standard
	Outstanding	Ultimate	Dev.
1986	0	8,184	0
1999	96,842	131,073	6,029
2000	150,900	150,900	9,685
Total	408,615	1,395,346	18,616

1 Unit = $1
Forecast scenario: Acc Yr 2000: 10%+_2%



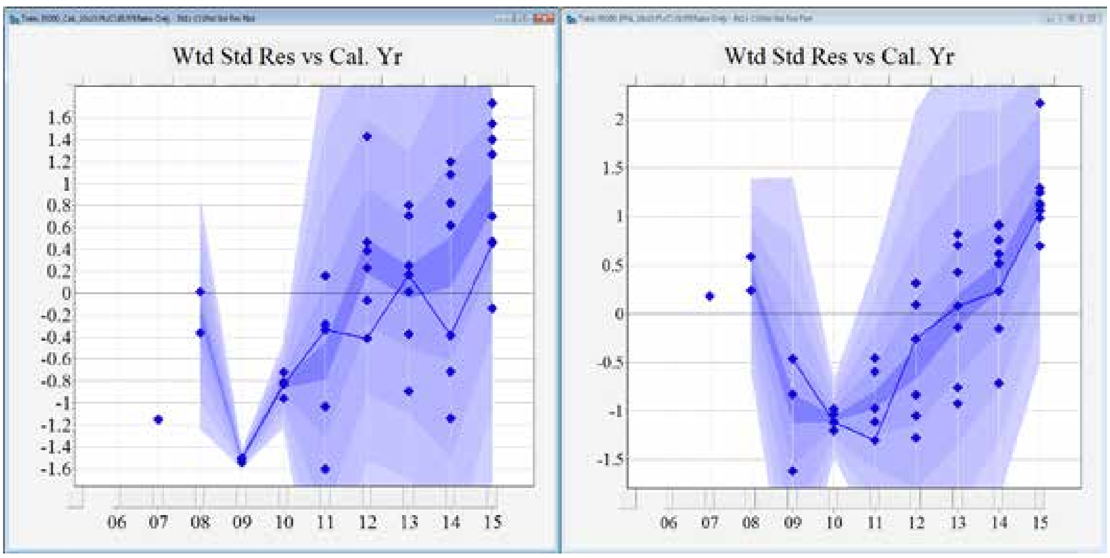


The relationship in the mean parameter estimates is not volatility (risk) may be related to internal or external drivers. They are explicitly incorporated in the model and risk exposure can be managed. The close relationship between the two segments does not eliminate the risk diversification credit for combining the analysis of the reserve distribution with the future accident (underwriting) year (see Modeling multiple lines of business brochure and Pricing: Segments, Layers, and Reinsurance brochure). The joint increase in parameters (with the associated uncertainty) is still accompanied by the increase of the overall risk diversification as the uncertainty in the parameter estimates is not highly correlated between the segments.

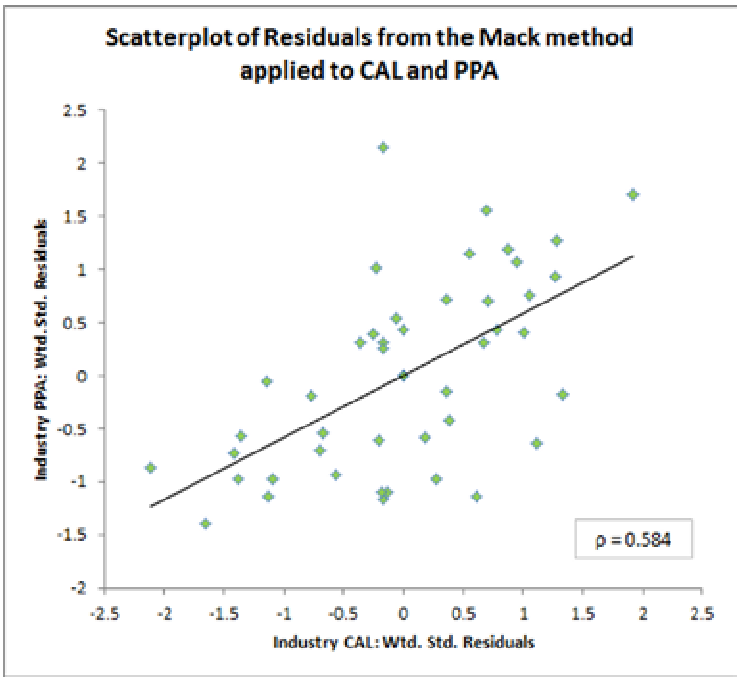


5.1. Spurious correlation between Industry PPA and CAL data

In this second example Paid Losses for the Industry PPA and CAL data from AM Best (2015) are modeled using the Mack method. The residuals are shown by Calendar year for CAL (left) and PPA (right) below with the trace line for accident year 2004 highlighted.

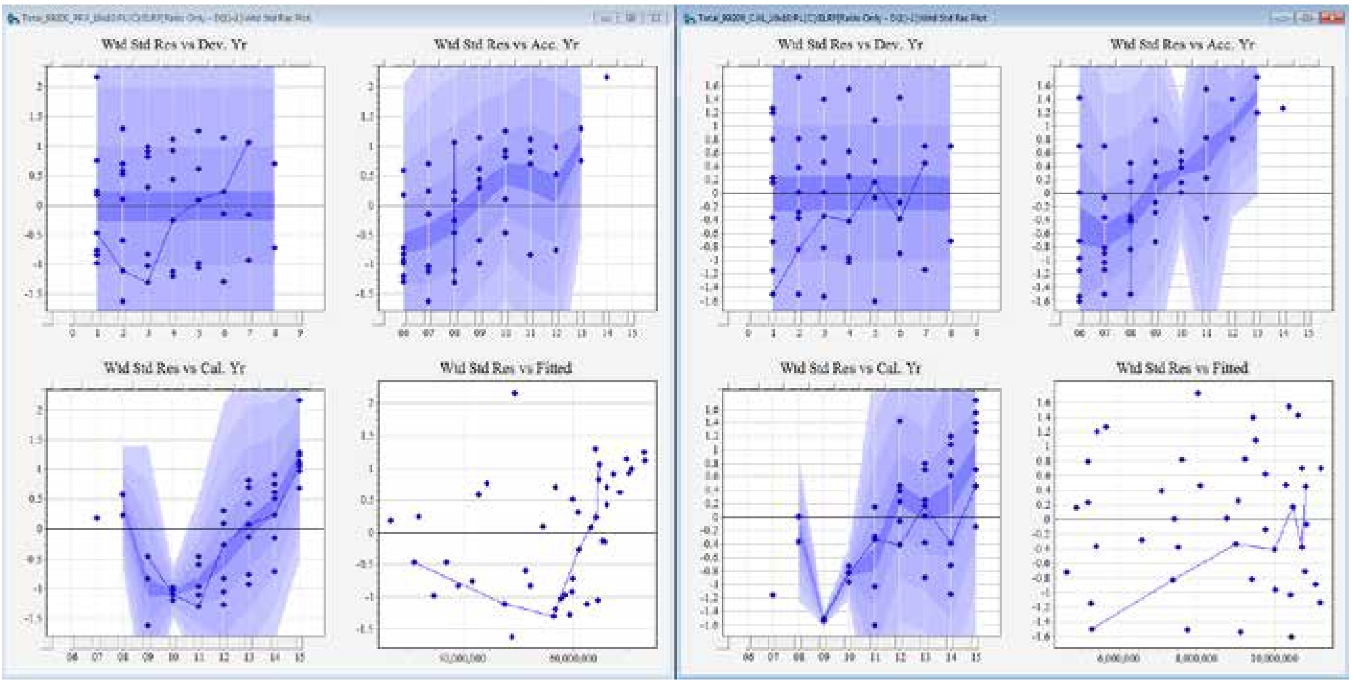


The marked residuals for the Mack method exhibit correlation (by eye). This correlation is then measured and shown in the residual scatter plot of PPA vs CAL below.



Although the correlation is strong in the residuals, this correlation is spurious. The calendar year residuals show the Mack method is over fitting the data more recently - a clear negative trend is evident in both residual displays (though it is much stronger for PPA). The volume weighted average link ratios (of which the Mack method is the regression formulation) do not describe the salient features of the data and, as a result, there is correlation found between the lines which would not be present in a correct model for the data (see Section 6.2).

The method has not described the trends in the data in either the calendar year direction or the accident year direction - see the full residual displays against each trend direction below. The Mack method (and in fact all link ratio methods - including bootstrapping from link ratio models) are inappropriate for both these LoBs. Link ratio methods cannot describe changing calendar year trends yet, as seen in Section 6.2, changing calendar year trends are found in these data.

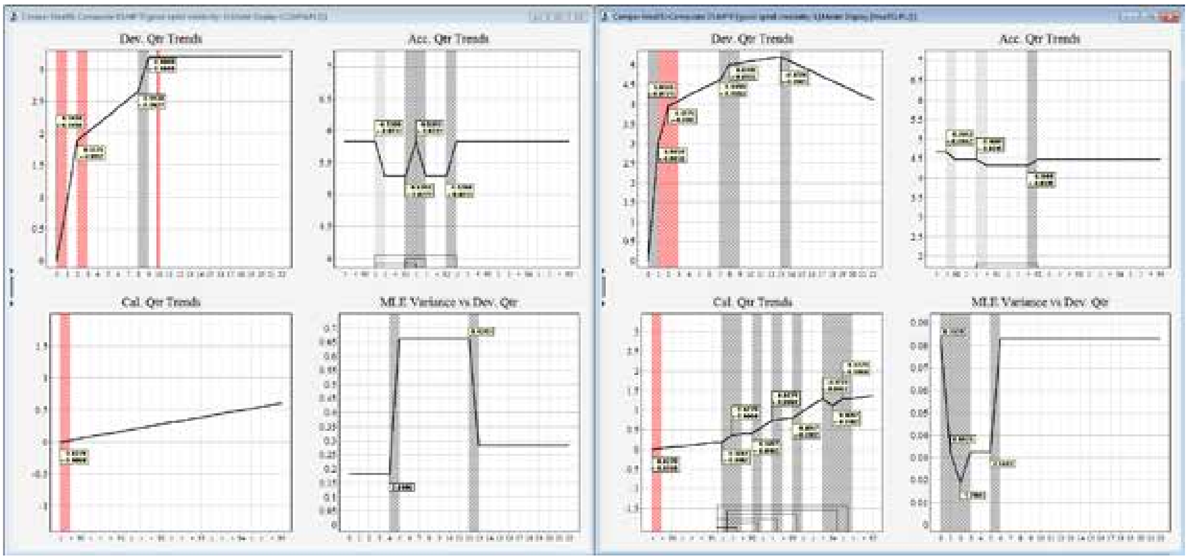


6. Case study: Companies versus the Industry

Industry-wide correlations between LoBs are not a reliable guide to the correlations between an individual company's LoBs which should be modeled directly in all cases. In fact individual and aggregate correlations can even be inconsistent (see Simpson's Paradox). However, industry correlations can be used as a guide especially in the presence of sparse data where the model for an individual company can be credibility adjusted to take industry-wide trends into account. Two examples are considered and for each example we find different trends in the companies compared to the trends found in the industry.

6.1. Company A versus the Industry

The following discussion relates to Auto BI written by Company A (representing about 3% of the industry) and the Auto BI industry data: MAA951. The industry data have high unstable calendar year trends and a final negative development year trend, whereas Company A, on its own, has an insignificant (zero) calendar year trend and an insignificant (zero) final development period trend.



The optimal model for the company is on the left; the optimal model for the industry on the right. Note the difference in both development and calendar period trends. The industry has clear, volatile periods of inflation whereas there is no evidence of this in Company A. Furthermore, the industry data shows a definite decrease in payments after development period 14 whereas there is no evidence (yet) of decreasing levels of payments in Company A.

Process correlation between the company and the industry data is statistically significant but low at 0.240.

Compa+Maa951:Composite DS:MPTF[good1 optall-1]:Weighted Residua...

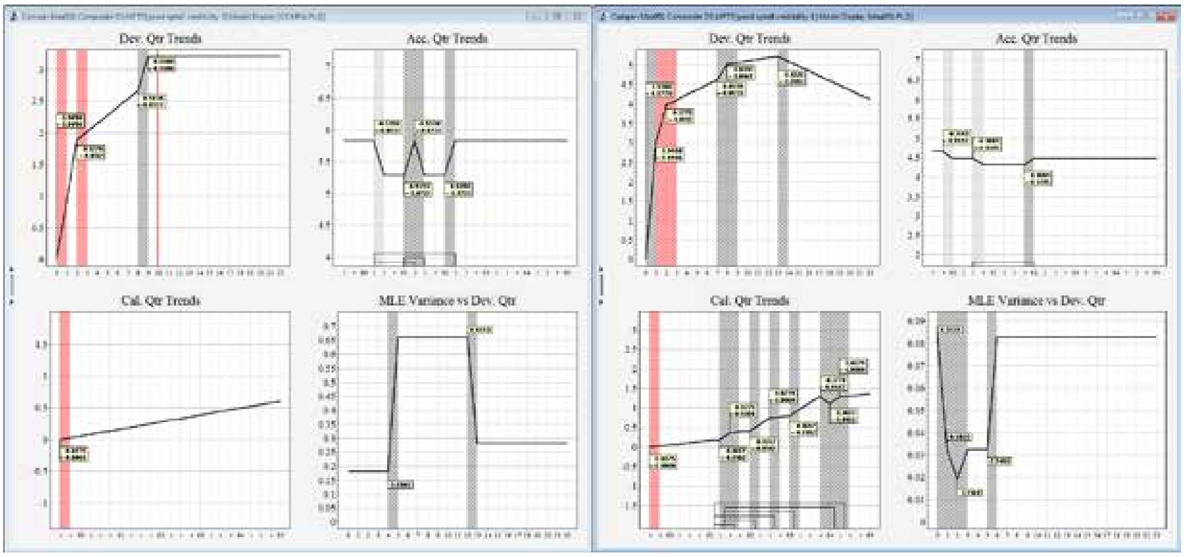
Covariances | Correlations | Final Covariances | Final Correlations

Final Weighted Residual Correlations Between Datasets

	COMPA:PL(I)	Maa951:PL(I)
COMPA:PL(I)	1	0.240
Maa951:PL(I)	0.240	1

3 iterations were executed
Residuals correlation difference tolerance
0.010%

Industry data can be used to credibility adjust the trends in the model for Company A. For instance, the calendar year trend in Company A can be set to be the same as the base trend in the industry.



Some parameters in the development direction are also found to be the same, but the trends in the company's data are not the same as in the industry. The zero development period trend (in the tail) for the company is not credibility adjusted to the industry as there is no statistical evidence to do so. Given the low process correlation between the company and the industry it is unlikely that the same process correlation effects measured from the industry between multiple lines would apply to Company A's equivalent lines.

6.2. Companies LMI and TG, versus the Industry for CAL, PPA

A.M. Best Schedule P data (2015) are used to compare CAL and PPA for two companies, LMI and TG, with each other and the industry. It may be expected that CAL and PPA are highly correlated, however, what do the data say?

As mentioned at the start of this brochure, correlation only has meaning relative to mean predictions (that is, correlation in volatility). If we calculate spurious correlation between the paid losses in the two LoBs (PPA and CAL) for the Industry, Company LMI, and Company TG, then we obtain the following matrix.

		LMI		TG		Total	
		CAL	PPA	CAL	PPA	CAL	PPA
LMI	CAL	1.0	0.684	0.952	0.734	0.942	0.643
	PPA	0.684	1.0	0.755	0.967	0.858	0.993
TG	CAL	0.952	0.755	1.0	0.785	0.959	0.709
	PPA	0.734	0.967	0.785	1.0	0.852	0.976
Total	CAL	0.942	0.858	0.959	0.852	1.0	0.818
	PPA	0.643	0.993	0.709	0.976	0.818	1.0

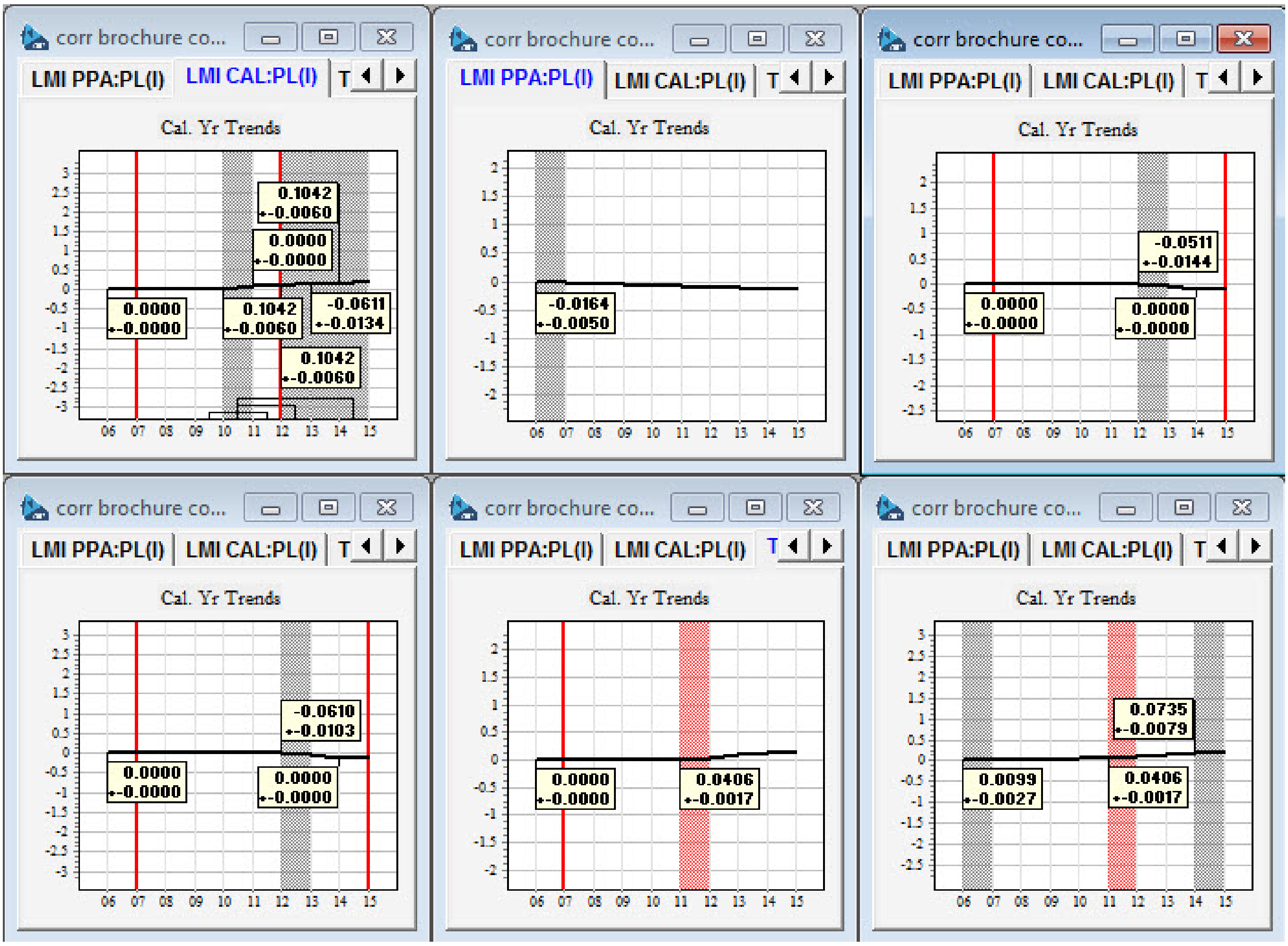
The paid loss data, with no adjustment for trends, is showing the high spurious correlation of about 0.99 between PPA and CAL in the Industry. However, when the trends in each LoB have been fully adjusted for, the resulting correlation matrix is very different.

		LMI		TG		Total	
		CAL	PPA	CAL	PPA	CAL	PPA
LMI	CAL	1	0	0	0	0	0
	PPA	0	1	0	0.251	0	0
TG	CAL	0	0	1	0	0	0
	PPA	0	0.251	0	1	0	0
Total	CAL	0	0	0	0	1	0
	PPA	0	0	0	0	0	1

All the correlations between PPA and CAL are statistically insignificant (thus grey zeroes), and only correlation between the two company's PPA remain.

Note that reserve distribution correlations are typically much lower than volatility correlations.

Calendar year trends are shown below.



There is some similarity in trend structure and parameter values. However, the industry trends are not a replacement for modeling the company data, though they may be used to credibility adjust the individual models, especially in the presence of very sparse data.

7. Case study: Risk capital allocation

Reserve correlation is an important component of the allocation of risk capital and as a measure of risk diversification (reserve correlation between LoBs). Reserve correlations within LoBs are primarily a function of parameter uncertainty; the higher the parameter uncertainty, the higher the correlation between cells.

As the probabilistic trend family modeling frameworks comprise a clear structure for relating cells by trends with associated uncertainty, it is no surprise that the correlations between cells and across LoBs are included in a natural way within this framework.

Risk capital allocation according to variability can be calculated directly using the variance-covariance formula. This formula can be used to allocate capital across LoBs, across calendar/ accident periods, or both.

7.1. Risk capital allocation across six LoBs using the variance-covariance formula

Risk capital can be allocated between LoBs and across calendar years by the same variance-covariance formula which estimates the relative uncertainty or risk of the specific LoB. Percentage allocation to the *i*th line, L, is:

$$A_i = \frac{\sum_j C_{ij}}{\sum_{ij} C_{ij}}$$

where Cij, is the covariance of Li and Lj. The formula can be extended to include time (either calendar or accident period), by summation of the covariances across the relevant time period. Similarly, allocation across time periods for a single LoB can be readily considered by treating i,j as time indices rather than LoB indices – the formula still holds.

All Lines:CDS:MPTF[GoodS2-7]:Reserve Forecast Summaries

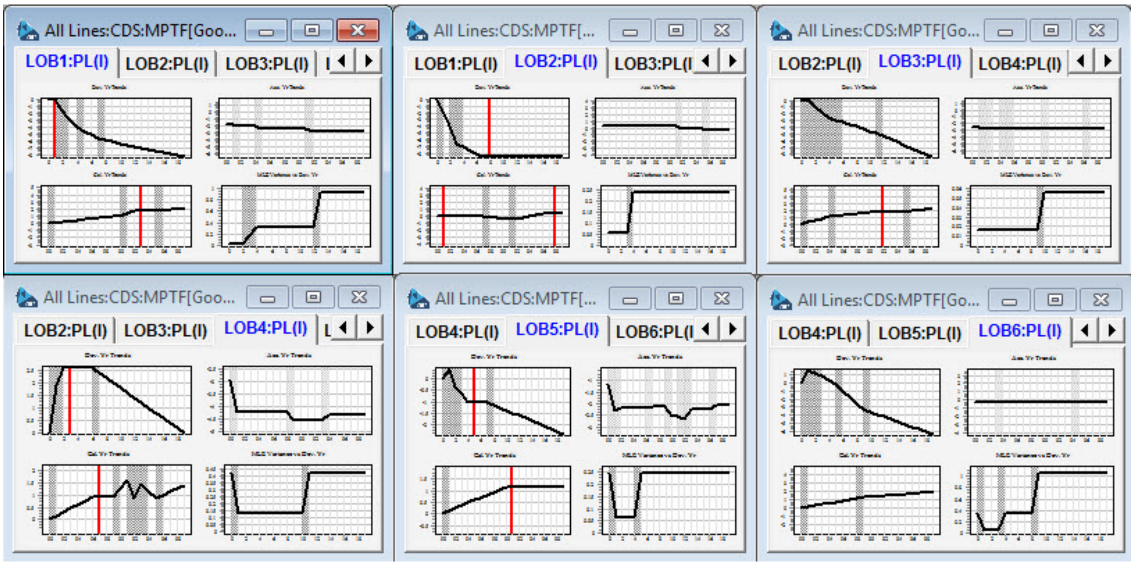
Reserve Breakdown by LOB

	Premium	Paid To 2009	Incurred To 2009	CRE 2009	Mean: Outstanding	Mean: Ultimate	Std Dev	Loss Ratios (%): Mean	Loss Ratios (%): Std Dev
LOB1:PL(I)	2,264,874	1,032,476	1,116,869	84,394	89,030	1,121,505	13,148	49.517	0.581
LOB2:PL(I)	369,472	136,891	137,630	739	1,579	138,470	153	37.478	0.041
LOB3:PL(I)	6,206,950	3,190,598	3,477,378	286,780	460,482	3,651,080	26,649	58.822	0.429
LOB4:PL(I)	55,983	47,449	76,316	28,867	78,261	125,710	38,276	224.551	68.371
LOB5:PL(I)	153,747	48,524	73,701	25,177	45,486	94,009	11,151	61.146	7.253
LOB6:PL(I)	429,305	157,467	169,994	12,527	22,488	179,955	2,567	41.918	0.598
Total	9,480,331	4,613,404	5,051,889	438,485	697,327	5,310,731	49,796	56.018	0.525

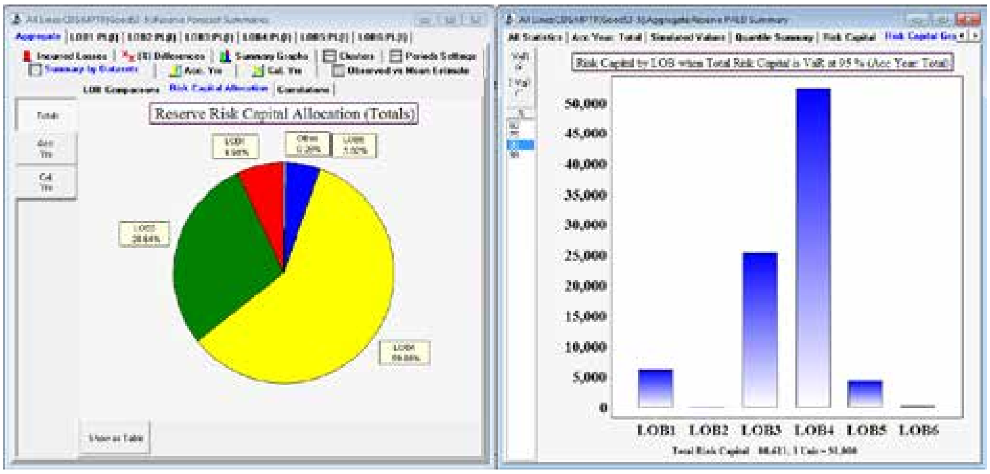
1 Unit = $1,000
Forecast scenario: Conservative

The above forecast summary is for six LoBs. Clearly, LoB 4 is expected to take the most risk capital followed by LoB 3 – just based on the relative standard deviations.

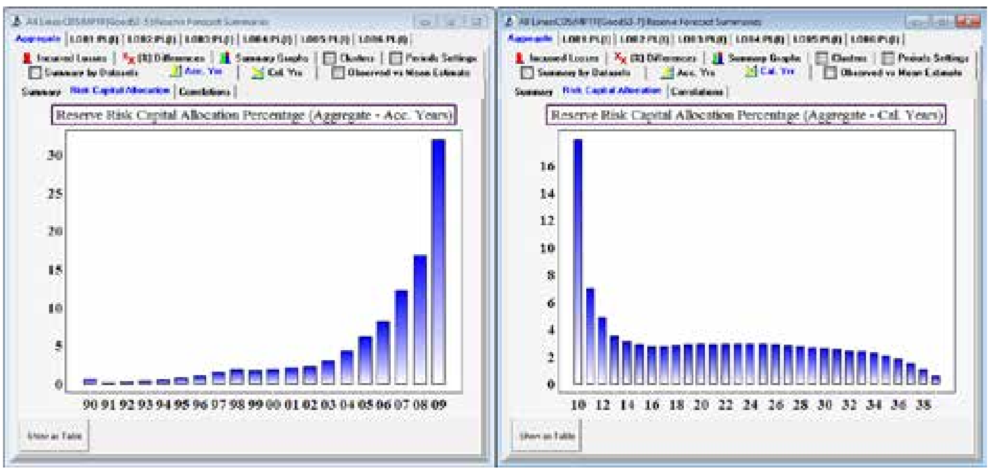
As detailed in Insureware's pricing brochure, the model for these Lines of Business show distinct trend and volatility metrics.



Risk capital allocated across LoBs can be calculated in general (forecast summary) or using specified value-at-risk (V@R) levels using the Predictive Aggregate Loss Distributions (PALD) simulations. Since there are no analytical distributions for the aggregate of log normals, simulations from the correlated lognormals in all the cells are produced in the PALD module to obtain distributions of reserves by accident year, calendar year, and the total. These simulations can then be used to calculate percentiles, V@R, or other statistics.



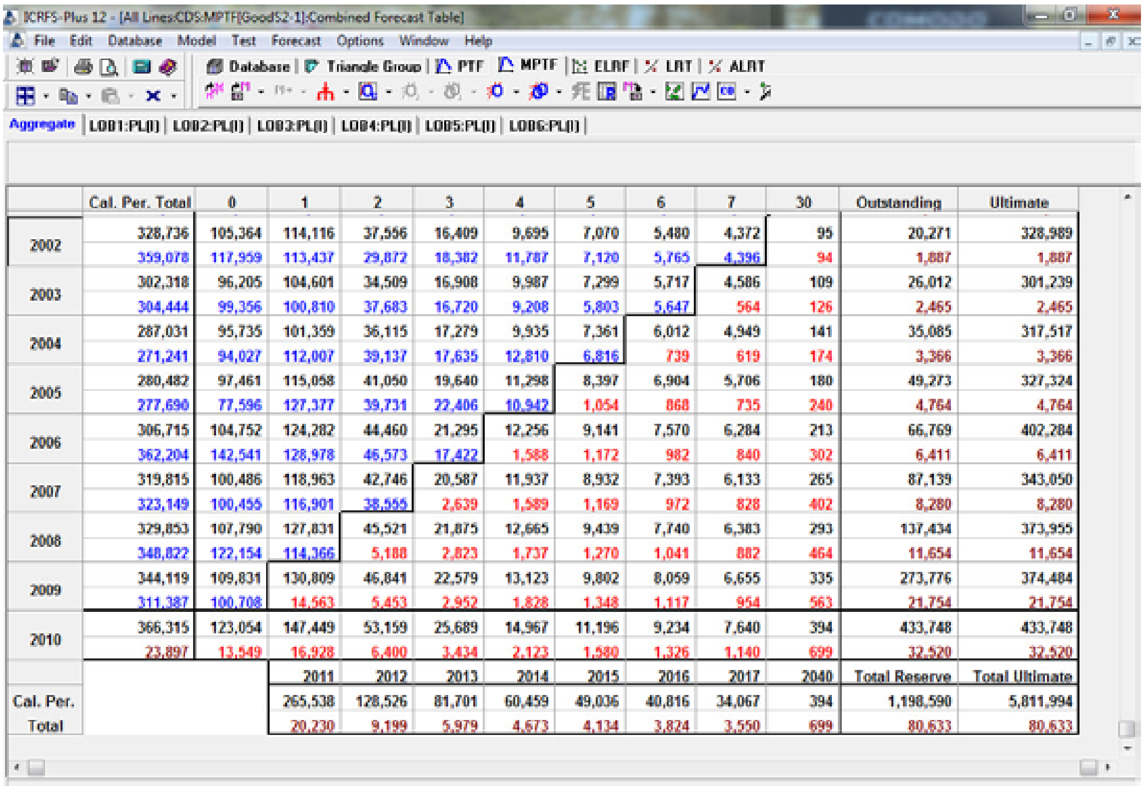
The allocation by accident period and calendar period for the aggregate of the six LoBs is as shown above. Similarly breakdowns within each LoB can also be calculated (not shown) where the allocation within an LoB follows the risk characteristics of that LoB.



8. Case study: Reserve, underwriting, and combined risk

A single composite model measures the reserve, underwriting and combined risks for each LoB and the aggregate. Reserve risk and underwriting risk are not treated as two separate analyses; rather, the same model can be applied for both, along with the analysis of the combined risk. Any correlations between future and reserve periods are driven by common parameters. That these parameters are common is another reason not to separate the reserve and underwriting calculations.

Combined risk is less than the sum of reserve risk and underwriting risk due to the diversification credit since the underwriting years typically have low correlation with reserve years. This is an important result typically ignored when considering reserving versus underwriting risk even when the majority of business underwritten in the next underwriting period is renewal business. The mix of risks in the underwriting period is usually the same as the reserving period.



	Cal. Per. Total	0	1	2	3	4	5	6	7	30	Outstanding	Ultimate
2002	328,736	105,364	114,116	37,556	16,409	9,695	7,070	5,480	4,372	95	20,271	328,989
	359,078	117,959	113,437	29,872	18,382	11,787	7,120	5,765	4,396	94	1,887	1,887
2003	302,318	96,205	104,601	34,509	16,908	9,987	7,299	5,717	4,586	109	26,012	301,239
	304,444	99,356	100,810	37,683	16,720	9,208	5,803	5,647	564	126	2,465	2,465
2004	287,031	95,735	101,359	36,115	17,279	9,935	7,361	6,012	4,949	141	35,085	317,517
	271,241	94,027	112,007	39,137	17,635	12,810	6,816	739	619	174	3,366	3,366
2005	280,482	97,461	115,058	41,050	19,640	11,298	8,397	6,904	5,706	180	49,273	327,324
	277,690	77,596	127,377	39,731	22,406	10,942	1,054	868	735	240	4,764	4,764
2006	306,715	104,752	124,282	44,460	21,295	12,256	9,141	7,570	6,284	213	66,769	402,284
	362,204	142,541	128,978	46,573	17,422	1,588	1,172	982	840	302	6,411	6,411
2007	319,815	100,486	118,963	42,746	20,587	11,937	8,932	7,393	6,133	265	87,139	343,050
	323,149	100,455	116,901	38,555	2,639	1,589	1,169	972	828	402	8,280	8,280
2008	329,853	107,790	127,831	45,521	21,875	12,665	9,439	7,740	6,383	293	137,434	373,955
	348,822	122,154	114,366	5,188	2,823	1,737	1,270	1,041	882	464	11,654	11,654
2009	344,119	109,831	130,809	46,841	22,579	13,123	9,802	8,059	6,655	335	273,776	374,484
	311,387	100,708	14,563	5,453	2,952	1,828	1,348	1,117	954	563	21,754	21,754
2010	366,315	123,054	147,449	53,159	25,689	14,967	11,196	9,234	7,640	394	433,748	433,748
	23,897	13,549	16,928	6,400	3,434	2,123	1,580	1,326	1,140	699	32,520	32,520
			2011	2012	2013	2014	2015	2016	2017	2040	Total Reserve	Total Ultimate
Cal. Per. Total			265,538	128,526	81,701	60,459	49,036	40,816	34,067	394	1,198,590	5,811,994
			20,230	9,199	5,979	4,673	4,134	3,824	3,550	699	80,633	80,633

The forecast table excerpt above corresponds to the six LoBs presented previously but where one future underwriting period has been added to the forecast scenario. The reserve and underwriting distributions are forecast jointly to calculate the total reserve for the combined reserve and underwriting periods. In this way, risk diversification by writing multiple underwriting periods is correctly included in the analysis.

It is sometimes assumed that the reserve and underwriting cycles will result in correlation between the two.

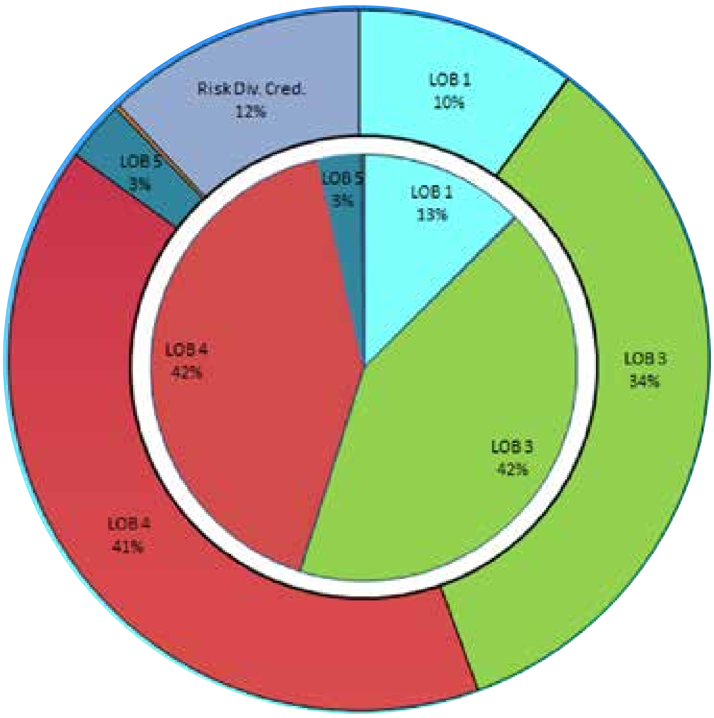
However, since these cycles are independent of the data and are rather imposed by the nature of the reserving and underwriting methodology as demonstrated in the next section.

The risk capital allocation table for the reserve, underwriting, and combined estimates of reserve mean and risk capital (V@R at 95%) is shown below. While the means are additive, the risk capital clearly is not. Furthermore, if the reserve and future underwriting periods were highly correlated then risk capital for the combined forecast would be close to the sum of the risk capital for the individual pieces.

In this example, a 12% discount in risk capital is obtainable as a result of risk diversification credit between both writing the multiple lines and by combining reserve and underwriting risk. Note that for LoB 4 – the line with the greatest risk capital requirement, minimal diversification credit is obtained for combining reserve with underwriting risk.

	Reserve (1990~2009)		Underwriting (2010)		Combined (1990~2010)		
	Mean Outstanding	Risk Capital	Mean Outstanding	Risk Capital	Mean Outstanding	Risk Capital	Discount effect (%)
LOB 1	90,130	6,295	92,995	11,090	183,129	14,116	19
LOB 2	1,226	11	4,640	159	5,866	89	48
LOB 3	460,482	25,123	274,773	33,601	735,255	47,700	19
LOB 4	78,261	51,828	15,726	6,096	93,988	56,346	3
LOB 5	45,486	4,399	6,638	177	51,853	3,871	15
LOB 6	22,488	233	9,641	134	32,129	292	20
Total	698,077	87,890	404,142	51,257	1,102,219	122,414	12

Combined (Reserve + Underwriting) Risk vs Reserve Risk + Underwriting Risk at the 95% quantile



Inner pie chart: Reserve+Future risk; Outer pie chart: Combined risk with risk diversification credit (purple).

The above display highlights the risk diversification credit gained by recognizing the nature of the reserving and underwriting problem. The inner pie chart shows the risk capital allocation percentages should reserve and future underwriting risk be calculated independently. The outer ring shows the assessment of the combined risk along with the diversification credit (12%) arising from the reduced risk capital requirement when considering the diversification between reserve risk and future risk.

9. Case study: Common accident year drivers and the reserving cycle

This case study considers the evidence of a reserving cycle and demonstrates that this cycle is not a feature of the long-tail liabilities but rather is a result of market pressure and methodology. The rationale behind the reserving cycle is described. The A.M. Best (2011) Commercial Multi-Peril data are then examined for evidence of this cycle. While the booked reserves do provide some evidence of a reserving cycle, the long-tail liabilities do not show any evidence of common accident year drivers. The apparent correlation in the behavior of the booked reserves is methodological and not a feature of the data.

A common fallacy in the industry is a belief that long-tail liability losses exhibit a reserving cycle. The actual losses from long-tail liabilities do not follow a reserving cycle. Booked reserves and premiums, however, may.

Premiums are set based on the demand for retaining market share and competitiveness amongst economic conditions. If the market under-prices the risk, individual companies will also under-price risk in order to maintain market presence. As a result of common commercial interests, there is a definite element of industry-wide dynamic. Booked reserves follow this cycle as management are pressured to select the lower actuarial reserve estimates in times of underpricing risk to remain competitive. Similarly, booked reserves and prices rise as the market responds to catastrophes (and management is under pressure to be conservative).

If this booked reserve estimate pressure was not bad enough, actuaries who use Bornhuetter-Ferguson methodology are even more at risk as this method introduces spurious correlation between premiums and booked reserves before further management influence.

The cycle is described as follows:

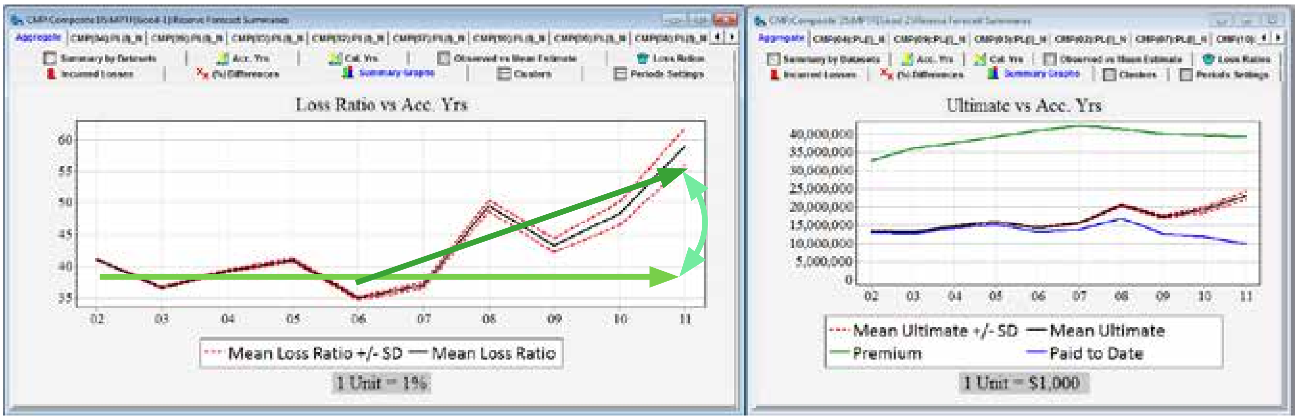
- Competition is low due to insurers leaving the market due to catastrophes (whether on the loss or asset side), prices rise, booked reserves are high.
- As prices rise, profits increase, more players enter the market.
- More players result in competition resulting in decreasing prices, lower reserves are booked.
- A catastrophe occurs resulting in players leaving the market.
- And the cycle starts again…

True best estimates of long-tail liabilities do not respect the market's business cycle but rather reflect the true risk of the business written. Typically, most companies write the same mix of risk from year to year. The prudent management team realizes this and both sets prices and reserves according to the level of risk taken.

9.1. Illustrative example: A.M. Best Schedule P Commercial Multi-Peril (CMP)

In 2011, ten company groups wrote over 50% of the total reserves of US Commercial Multi-Peril (CMP) based on reserves held (where reserves held are defined as the sum of Case Reserve Estimates and Bulk & IBNR).

Although the loss ratios (for Ultimate Earned Premium) for the Industry are still healthy, it is clear that conditions are worsening more recently. This could be a result of a number of factors arising from market and economic conditions (businesses folding, reduced value, etc).

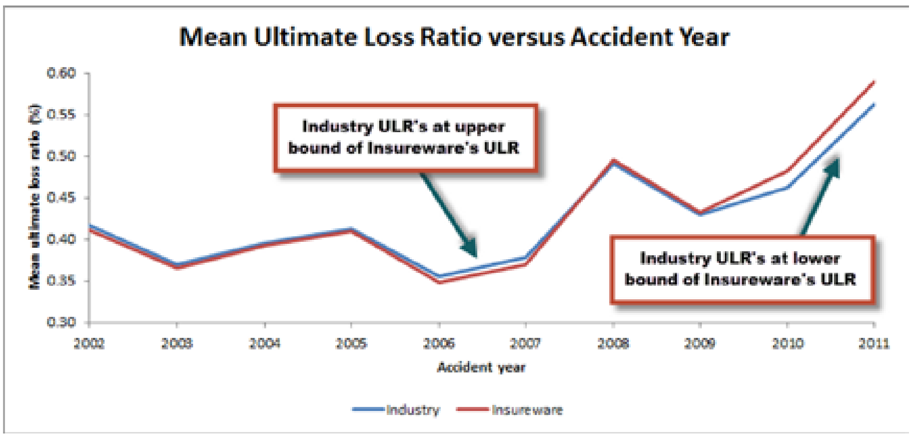


The position in the cycle is probably between the competition for market presence amongst difficult conditions and a catastrophe occurring. Conditions have worsened, loss costs are increasing while total premiums are decreasing (see above right), but the industry (as a whole) is still profitable (after allowing for ALAE). What is happening to booked reserves versus premium? Are the companies pricing their risk accurately?

9.2. Industry mean ultimate loss ratios booked reserves versus independent estimates

The mean ultimate loss ratios for ultimates held in the Industry are compared with the mean ultimates estimated from the Probabilistic Trend Family (PTF) modeling framework for an optimal model and future forecast scenario. The key element here is that Insureware's estimates are based only on trends and volatility found in the Industry data and future expectations are thus independent of both market pressure and other commercial considerations.

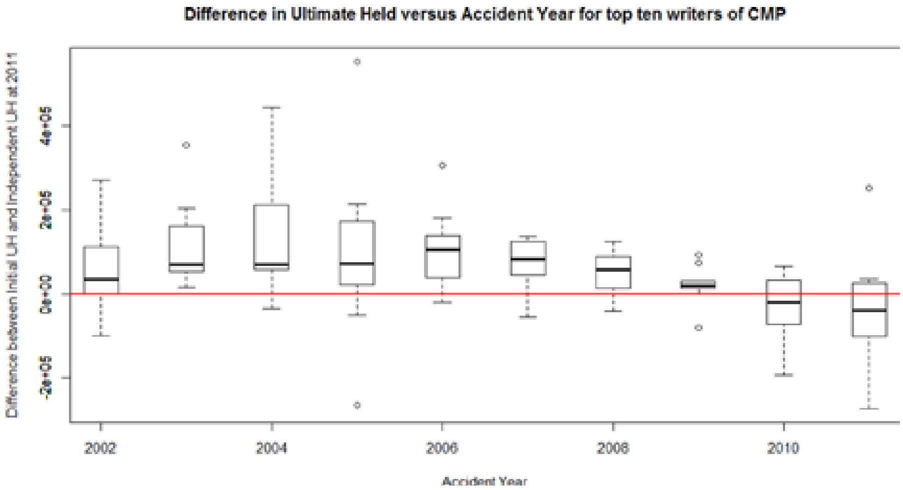
What we expect to see is that as the conditions worsen from 2007 onward, the mean loss ratios do not increase as greatly for the booked reserves versus the Insureware estimates of the mean ultimate loss ratios. That is, we expect the Industry to be more optimistic about the mean ultimate loss ratios due to collective decreasing of booked reserves in connection with the lower premium raised.



The above graph illustrates that the industry is behaving exactly as expected. Insureware's estimates of mean ultimate loss ratios are not biased by management or other external commercial pressures and are more optimistic during the good years (2006~2007) and significantly more pessimistic during more difficult market conditions. This collective response to changing market conditions further reinforces the belief that the risk in the industry and individual companies are highly correlated.

9.3. The reserving cycle

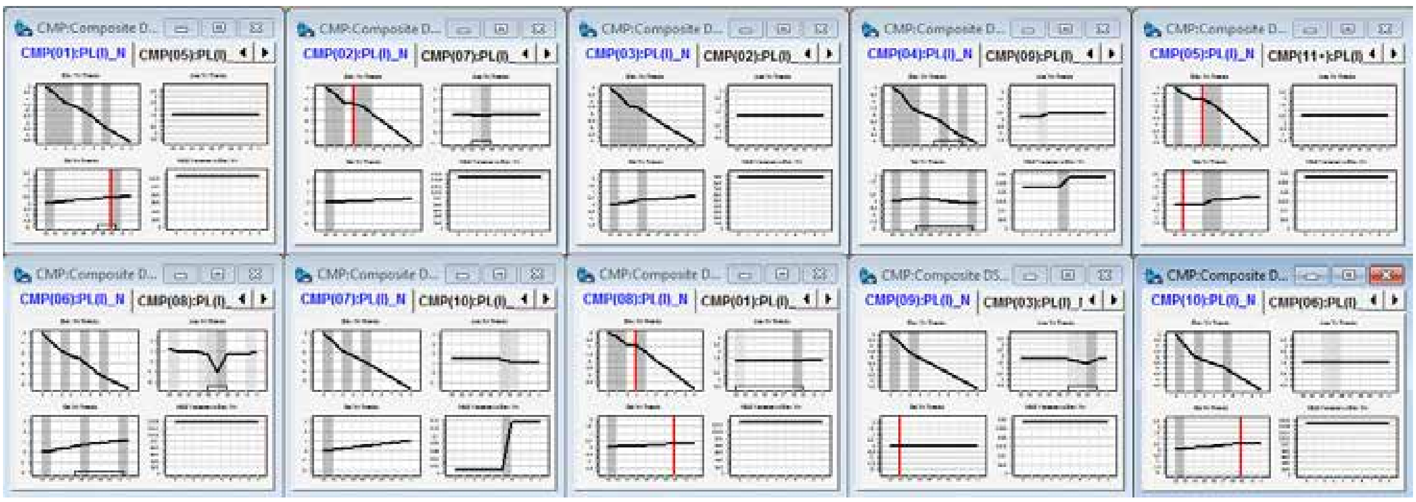
In order to determine the effect of the reserving cycle, we compare the estimated ultimate at the start of the policy period versus the projected ultimate as at year end 2011. If the cycle exists then this will be illustrated in the difference between the two ultimates responding to market conditions.



The early accident years (up to 2006) are consistently conservative. That is, for the largest ten writers of CMP, the ultimates are estimated very conservatively with the result that by year end 2011, the estimates of the ultimates have been revised downward. For 2007~2009, the ultimates are still being estimated conservatively (relative to the independent mean ultimate as at year 2011), but with decreasing conservatism. In the most recent two accident years, the company ultimates are considerably more optimistic – reflecting the effect of the reserving cycle at the time of greater market pressures.

9.4. Probabilistic Trend Family models for the largest ten writers of CMP do not demonstrate common accident year drivers

Below are the model displays for the ten largest writers of CMP by reserves held. The trends in the three directions and volatility are displayed (left to right: development year trends, accident year trends, calendar year trends, volatility by development period). The key components of note are that: a) the trends in the three directions are unique to each company, and b) there are no indications of common accident year drivers. The latter is expected should the reserving cycle be a feature of the long-tail liabilities.



The lack of common accident year level changes between the top ten writers of CMP (despite loss ratios behaving similarly), emphasizes the conclusion that the reserving cycle is not a feature of the long-tail liability losses but rather reflects management's selection of booked reserves from the range of actuarial estimates. This common management dynamics does not constitute correlation.

info@insureware.com
Suite 6 & 7 40-44 St Kilda Road
St Kilda VIC 3182 Australia
Tel: +61 3 9533 6333 Fax: +61 3 9533 6033

Insureware
Innovative Statistical Solutions for P&C Insurance