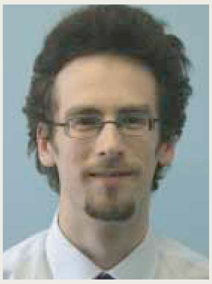
About the bootstrap

Dr David Odell and David Munroe, of Insureware, take us through the use of the bootstrap



David Munroe

Dr David Odell

Have you ever had this experience? One morning as you are about to leave the house you realise you have forgotten your keys. You look for them in the usual places but can't find them. You carefully retrace your steps of the night before, look beneath the couch cushions and under the bed and they still don't appear. By this time you are getting more than a little tense, so you make a conscious effort to calm down and think things through. Since you have no clue where you put the keys, you will try to reconstruct the process by which you mislaid them. You take some other object of a similar size, for example one of your child's toys, put it in the pocket where you normally keep the keys and walk out the front door, carefully wedging it so that it doesn't lock on you. Then re-enter your home, pretending to unlock the door with Thomas the Tank Engine. Making a conscious effort to be in the same oblivious state of mind you were in the night before, you casually throw the object anywhere. Then, quickly reverting to your anxious self, you go and search around the area where it landed. Odds are you'll find all sorts of interesting things there which might even include your keys, but just as likely it's the last place you look before going for the spare keys. This, in a nutshell, is the bootstrap method.

The bootstrap isn't a model. It's what you do when you have a question about some data and no model to help answer it. It says, just do something that's random in the same way (the bootstrap has a perfectly practical recipe for doing this) and look at the results. It isn't a very good way to pinpoint your lost keys but the statistical bootstrap can be a very effective way to do something else. The bootstrap doesn't produce a good estimate of the mean (that would be like the toy landing on the keys) but it can help you measure the variability in an estimate, in other words it might help you find what room the keys are in.

The method was named and popularised by Brad Efron in a 1979 paper in the Annals of Statistics. There are many variants but at the heart of them is the idea of supplementing a dataset with a number of pseudo-datasets formed by re-sampling from the dataset (with replacement). A statistic of interest, for example the mean, is calculated for the original data. With a small dataset and no knowledge of the underlying distribution, the question arises as to how to compute a distribution for this statistic. We might, for example, be required to estimate the 99th percentile of the mean. The bootstrap tells us to calculate the means for each of the pseudo-datasets and use the distribution of these to capture the distribution of the mean.

The mean of the pseudo-data (177,004) is different from the mean of the original data (154,867) because, in the re-sampling, we are permitted to pick the same number twice.

Here is a dataset - it is a paid loss array for a US Workers Compensation portfolio over the 11-year period 1977-1987:

Company ABC: Accident years vs Development years. 1 Unit = $1

	0	1	2	3	4	5	6	7	8	9	10
1977	153,638	188,412	134,534	87,456	60,348	42,404	31,238	21,252	16,622	14,440	12,200
1978	178,536	226,412	158,894	104,686	71,448	47,990	35,576	24,818	22,662	18,000	
1979	210,172	259,168	188,388	123,074	83,380	56,086	38,496	33,768	27,400		
1980	211,448	253,482	183,370	131,040	78,994	60,232	45,568	38,000			
1981	219,810	266,304	194,650	120,098	87,582	62,750	51,000				
1982	205,654	252,746	177,506	129,522	96,786	82,400					
1983	197,716	255,408	194,648	142,328	105,600						
1984	239,784	329,242	264,802	190,400							
1985	326,304	471,744	375,400								
1986	420,778	590,400									
1987	496,200										

Table 1

Company pseudo-ABC # 1: Accident years vs Development years. 1 Unit = $1

	0	1	2	3	4	5	6	7	8	9	10
1977	129,522	42,404	51,000	83,380	129,522	205,654	190,400	194,650	87,582	259,168	142,328
1978	496,200	21,252	27,400	590,400	253,482	496,200	123,074	219,810	496,200	158,894	
1979	326,304	178,536	259,168	38,496	120,098	264,802	104,686	60,232	131,040		
1980	178,536	24,818	259,168	123,074	326,304	259,168	255,408	42,404			
1981	471,744	190,400	14,440	105,600	35,576	16,622	12,200				
1982	375,400	590,400	16,622	264,802	177,506	71,448					
1983	105,600	21,252	329,242	22,662	197,716						
1984	24,818	226,412	27,400	188,388							
1985	47,990	60,232	205,654								
1986	211,448	129,522									
1987	190,400										

Table 2

And here is the same dataset re-sampled (with replacement), a bootstrap pseudo-dataset.

Bootstrapping this one million times I find that the 99th percentile of the mean is 192,310. This is very nice; unfortunately it doesn't answer any of the questions I'd like to put to these data. In fact it is completely irrelevant because this is not a small sample from the same population; each value in the triangle is from a different population (distribution).

I decided to randomise my keys-finding procedure by starting from any person and any house ..., not just in my street, but in the world!

If these 66 figures were, say, a random sample of share transactions at a brokerage in the course of a year and I wanted the 99th percentile of the total annual commission, then it would be relevant. In this case, however, my problem is to do with forecasting on the basis of a specifically structured set of data. The pseudo-data above are of no use since the structure (of real data) has been lost. To return to the opening analogy, using this pseudo-data to help understand the variability in a forecast would be as if I had decided to randomise my keys-finding procedure by starting from any person and any house (with the same number of rooms), not just in my street, but in the world!

Before we leave this example, it is worth pointing out that the bootstrap exercise we've just done is about the parameter uncertainty and not the process variability, the parameter in this case being the mean. If I was interested in the process variability I might try bootstrapping the 75th percentile of the data (since there are fewer than 100 data points it would be meaningless to bootstrap the 99th percentile).

Here is another view of the original data:

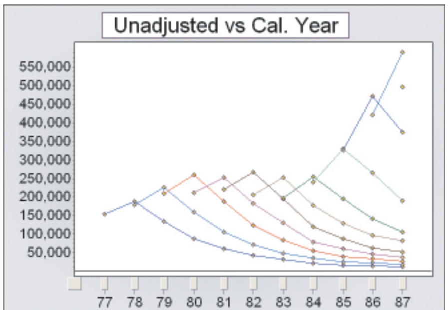


Figure 1

The coloured lines trace the paid losses for a common underwriting year. Here we can clearly see that there is a common development pattern and that the level of payments in the tail increases with each successive underwriting year. We can also see that the size of the payments in the first two development periods is not a good indicator for the level in the tail.

So, let's be clear on the problem before we see whether and how the bootstrap can help us. Our forecasting (reserving) problem is to estimate the entries in the empty cells of the data square above. We can assume that we already have a method for doing this, but that this method only produces means (best estimates) and possibly standard deviations for the cells. What we want the bootstrap for is to get the full distributions of these numbers, or at least of their aggregate. We don't want the parameter uncertainties for the means but the actual variability in the amounts we are going to pay because only this, which should incorporate both process variability and parameter uncertainty, can give an indication of the reserve risk.

What we want from the bootstrap is a way of producing thousands of pseudo-datasets that have the same basic structure as the original but that are random in the same way as the original. The bootstrap should enable us to do this if we can separate the structure in the data from the attendant process variability. We then re-sample (with replacement) from the process variability component and use this to create a pseudo-dataset. Furthermore, we also re-sample the process variability for the cells in the forecast portion of the array so that our resulting distribution is for the actual risk and not just the parameter uncertainty. Simple!

There are a number of technical problems to do with scaling for different parts of the array which we will ignore in order to focus on the key issue; that is, decomposing the data into structure and process variability. If our forecasting method is based on a model for the data, we are bound to equate structure with the fitted values created by the model.

So, in summary:

Data = Structure + Process Variability; or, equivalently: Data = Fitted Model + Residual

The process variability (residuals) can be bootstrapped (re-sampled) without any distributional assumptions. We combine re-sampled residuals with fitted values to produce each bootstrap pseudo-dataset.

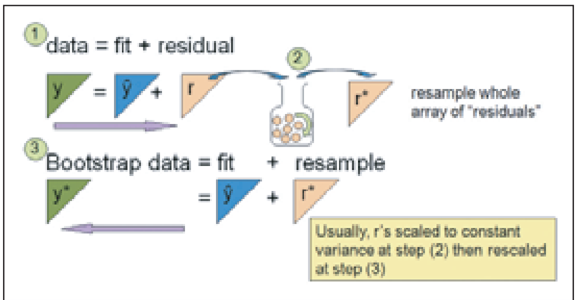


Figure 2

By bootstrapping the original triangle to the left we erroneously assumed that all the values came from the same population - that is no structure in the data. The real triangle and the bootstrap triangle are then plotted on a log-scale versus development year below:

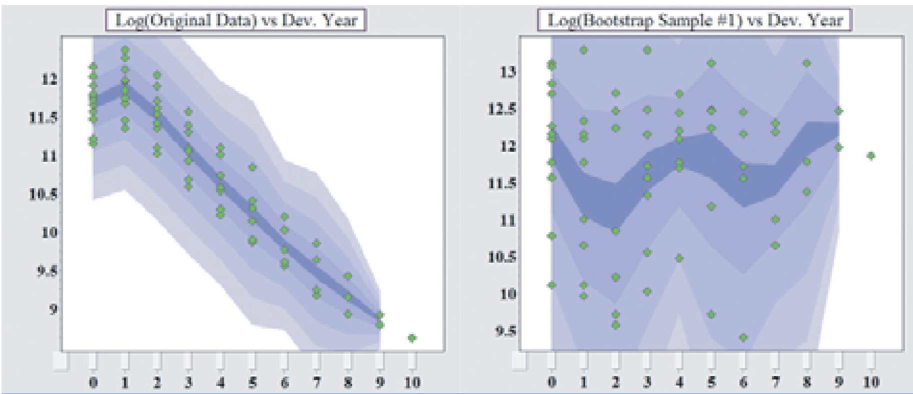


Figure 3

As we did not correctly separate out the structure from the process variability, the bootstrap sample pseudo-dataset (right) is almost structureless compared to the real data (left).

Everything hinges on getting the cut between structure and variability right. This needs the precision of a good sushi chef.

Everything hinges on getting the cut between structure and variability right. This needs the precision of a good sushi chef. If we cut too far into the variability, we'll have incorrect means and too little process variation and if we cut too far into the structure, we'll have incorrect means and too much process variation. Either way we'll miss what we want, which is the precise distribution(s) for the forecast(s). The bootstrap technique is only meaningful if the structure has been correctly estimated (fitted) - the model has to be right.

Look at the two pseudo-datasets below. Each is derived by bootstrapping from a different model. One of them is obviously failing to capture the variability we are interested in. Can we discover the error in the way the associated model has cut the data?

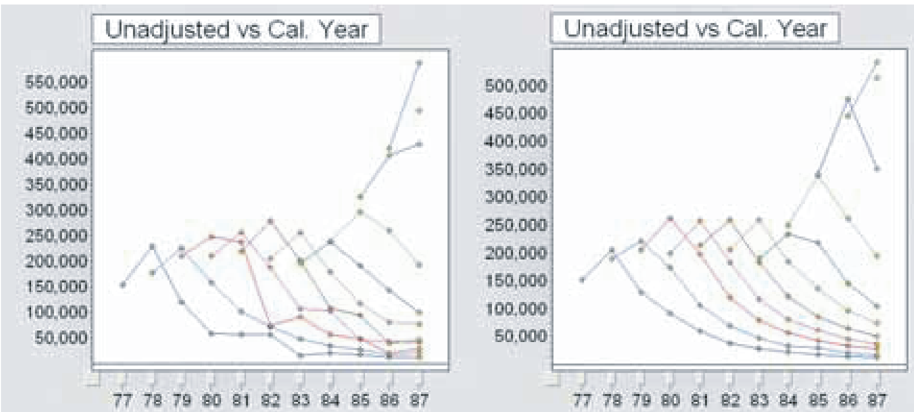


Figure 4

The pseudo-data on the left is so wobbly that it is hard to see it and the original data as both being typical members of the same class.

Let's look at smoothed histograms of the variability which was re-sampled to create these:

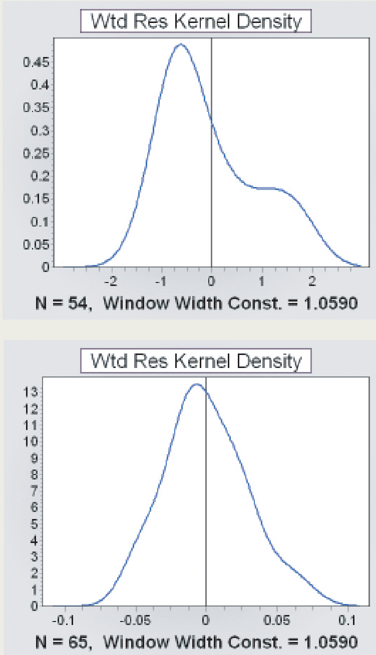


Figure 5

(The difference in scale of these graphs is a result of the different methods used and not otherwise relevant.) The shape on the left with its two unbalanced modes is cause for concern. But then isn't the point of the bootstrap that it is non-parametric, we don't expect to have nice normal distributions?
This is true, but what if the negative and positive residual values are clustered in distinct parts of the array? That would destroy the credibility of the bootstrap by showing that the cut had been made too far into the structure side of the data.

When we plot the residuals against calendar year, we see that in fact this is exactly what has occurred. The model used on the left is the Mack model (equivalently, volume weighted average link ratios) and the one on the right is a trend-and-volatility-based model from the Probabilistic Trend Family (PTF) modelling framework.

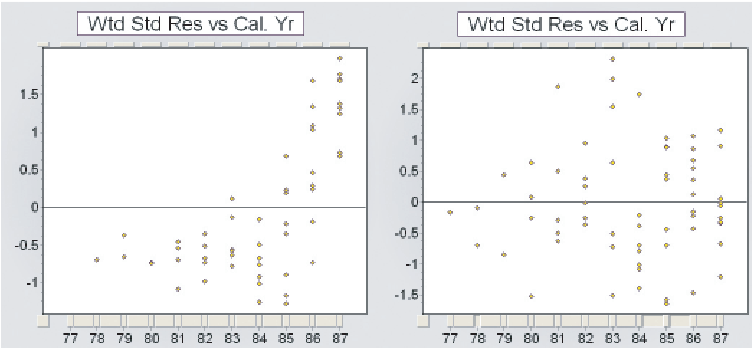


Figure 6

The identified (optimal) model in the Probabilistic Trend Family (PTF) modelling framework fits a parsimonious set of parameters to all three directions, development, accident, and calendar periods, as well as modelling the process variance. The modelling framework, allows us to cut the data into the two parts, structure and process variability, correctly. The trend structure in the data in the three directions is depicted below. [The bottom right graph represents the variance of the process variability (residuals) that have been tested to come from normal distribution].

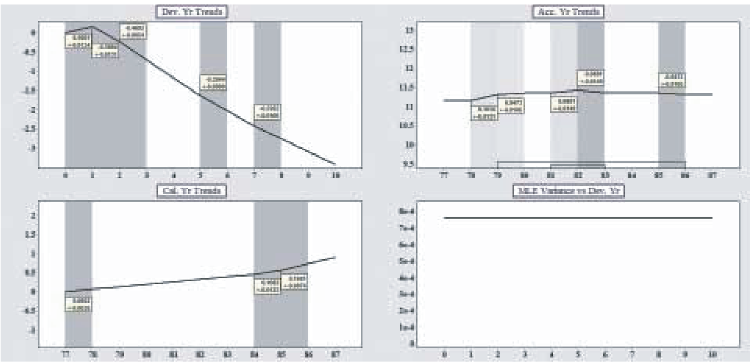


Figure 7

It is no surprise that the identified parsimonious PTF model produces bootstrap samples that resemble the original data since it captures the calendar year trend structure, as well the structure in the other two directions. The weighted standardised residuals are random (no structure) and all come from the same distribution. Note major calendar year trend shifts in 1985-1986-1987.

We now give a second counter-example of cutting the structure incorrectly. We saw previously that the Mack method did not remove all the structure by calendar year. What if we use the bootstrap on the remaining noise? Will not the pseudo-samples differ from the real data as before?

Again, the clearest way to answer this is a plot of the residuals versus calendar year of the Mack method applied to the real data, and the bootstrap triangles.

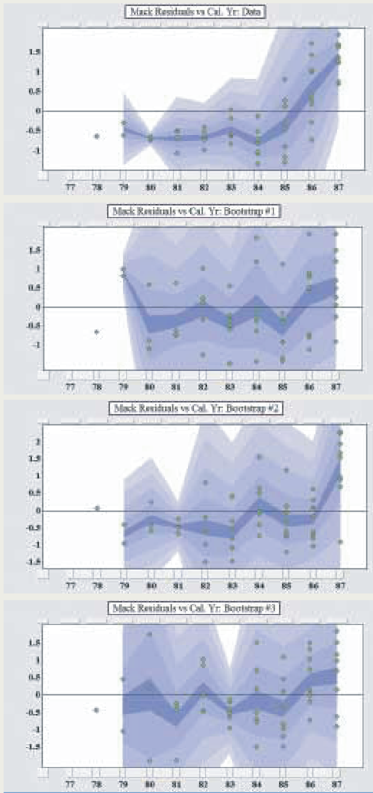


Figure 8

The uppermost residual plot is that of the Mack method applied to the real data. The other three residual plots are those of the Mack method applied to three bootstrap triangles generated from the Mack method (to the real data). It is obvious that the structure versus calendar observed in the original data (uppermost), has been lost to a great degree in the bootstrap samples. A bootstrap pseudo-sample will remove any structure still remaining in the process variability (noise) as measured by the Mack method. The high positive residuals in the latest calendar years are randomly assigned to any calendar year!

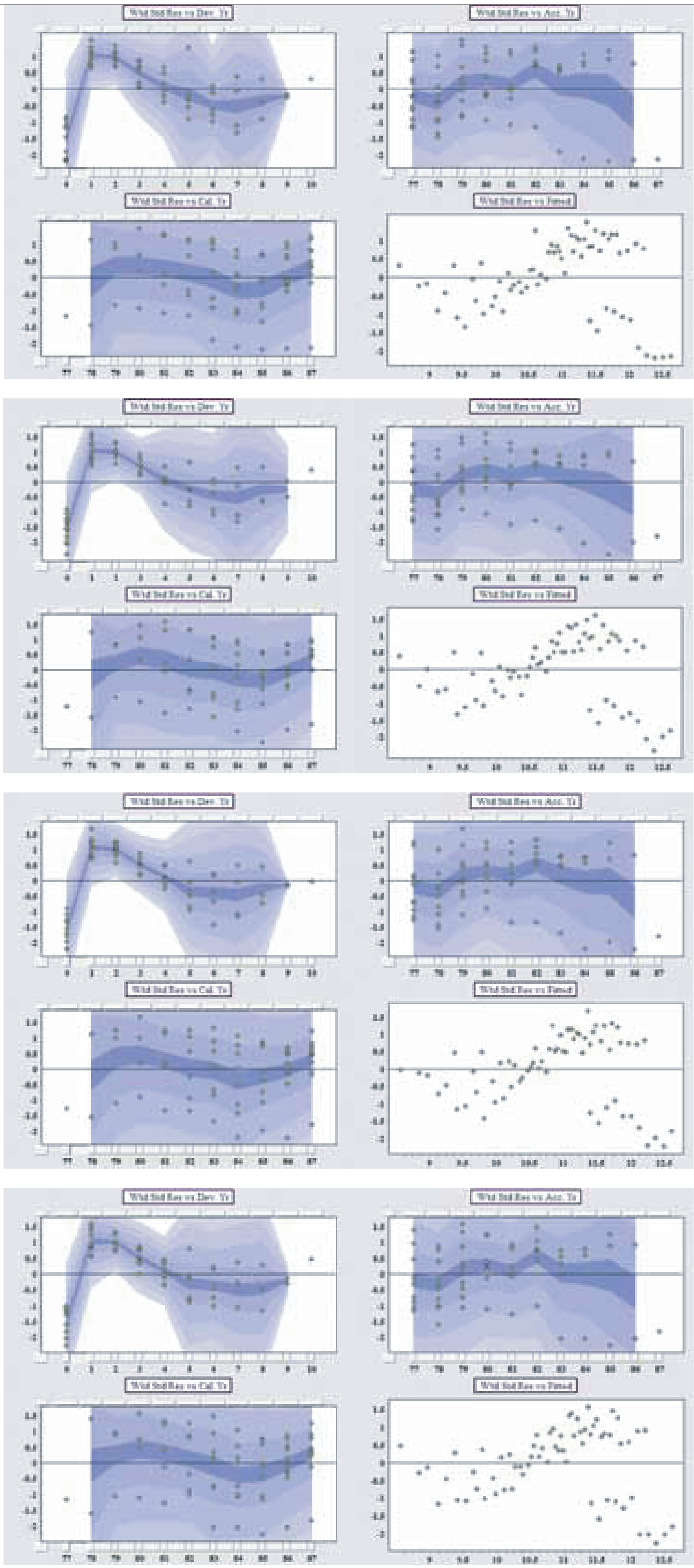


Figure 9

We illustrate this by comparing residual graphs of three bootstrap samples from the identified optimal PTF model with the residual graphs of the real data. The model fitted to each triangle has a single parameter in each direction.

Note the four sets of residual graphs are indistinguishable.

This confirms that the identified PTF model has separated the trend structure from the random variation accurately.

The four residual graphs below are based on fitting a two-away ANOVA model on a log scale that removes (estimates) trends between every two contiguous development years and between every two contiguous accident years, leaving any remaining structure visible by calendar year. This model is a diagnostic model for testing the instability of calendar year trends (and should not be used for forecasting).

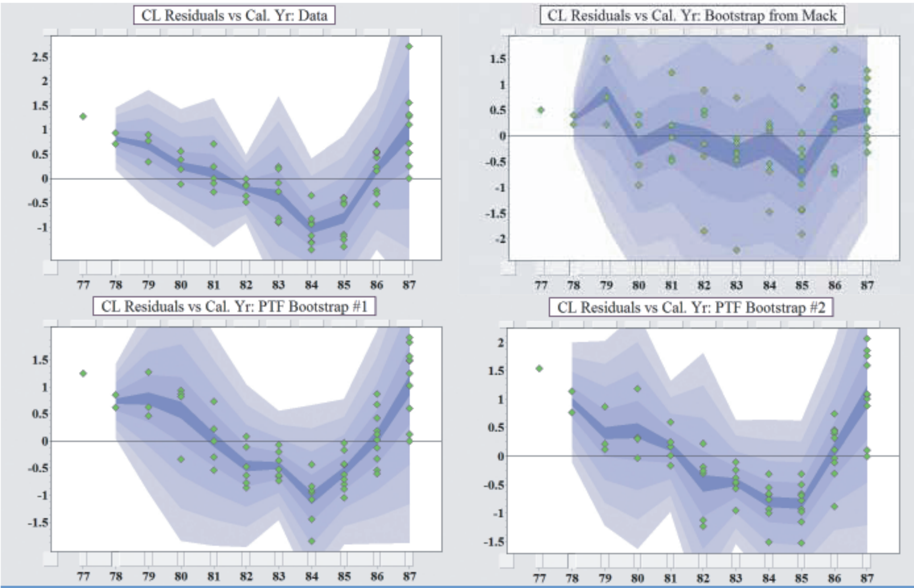


Figure 10

Spot the odd one out? The bootstrap sample (top right) from the Mack method fails to replicate the calendar year trend changes - it has failed to act with the precision of a good sushi chef. The real data is top left and the other two are two of the bootstrap samples generated from the identified PTF model (above), and are indistinguishable from the real data.

The diagnostic use of the bootstrap: create some bootstrap pseudo-datasets and see if they replicate the features in the real data.

We have just used the bootstrap as a diagnostic tool for testing a model. To do this we create one or a small number of bootstrap samples from a model and then see whether these and the original data share a family resemblance. If not, the model has failed in some obvious way to separate structure from process variability (randomness).

What about the distributional use of the bootstrap? For the identified PTF model above this is not needed since the weighted standardized residuals are (tested) from a normal distribution and the model itself produces distributions for all forecast cells.

Most actuaries that use the Mack method and bootstrapping do not in fact bootstrap the Mack method! They actually bootstrap the residuals of the Log-Linear Poisson model, and the noise component used in the forecast portion of the table is drawn from the parametric Inverse Gamma distribution.

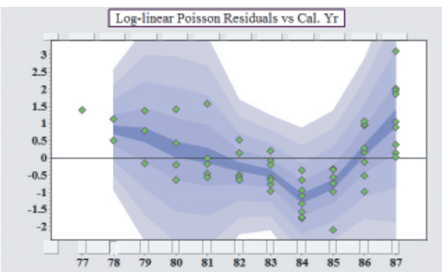


Figure 11

Left: The Log-Linear Poisson residuals for the ABC data also show obvious structure in the calendar direction. As with the Mack method, this is because calendar trends were not correctly accounted for in the model.

The additional components have been carefully tweaked to produce a standard deviation compatible with that found in the basic Mack method.

The Log-Linear Poisson model does not have the same residuals as the Mack method and it is not the same model, though there are some interesting relationships.

There is no a priori reason to assume that the distribution found in this way has any connection to the data. To return to the sushi chef analogy, it doesn't matter how you cut up the fish if what ends up on the plate is mostly surimi, a.k.a. seafood extender.

Parameter uncertainty and process variability

We decompose data into trend structure (in the three directions) and process variability (randomness). These are two different sources of variability for which both must be handled correctly and with the correct terminology.

The parsimonious estimated trend parameters have associated uncertainties, and the remaining (random) variation in the outcomes is due to the variation in the underlying process that cannot be eliminated or reduced, since it is a component of the generating process. All we can do is measure its distribution(s), and make decisions accordingly.

Predictions of variability (probability distributions) in the future incorporate both parameter uncertainty and process variability.

Furthermore, it should be clear that we must be interested in more than just 'an answer', such as reserves or ultimates by accident period. In respect of Market Value Margins, cost of capital calculations, T-VaRs and VaRs we also need the probability distributions of the liability stream and their correlations by calendar period. These are conditional on an explicit set of assumptions that are transparent, auditable and can be monitored in a probabilistic framework.

To find out more about bootstrap techniques, the Mack and related methods included in the Extended Link Ratio Family (ELRF) modelling framework, the PTF and MPTF modelling frameworks, and all aspects of managing long tail liability risks that are relevant to Solvency II, please visit the Insureware website: www.insureware.com.