Institute of Actuaries of Australia

Best Practice in Reserving

Prepared by Glen Barnett

Presented to the Institute of Actuaries of Australia
XVth General Insurance Seminar 16-19 October 2005

The Institute of Actuaries of Australia
Level 7 Challis House 4 Martin Place
Sydney NSW Australia 2000
Telephone: +61 2 9233 3466 Facsimile: +61 2 9233 3446
Email: actuaries@actuaries.asn.au Website: www.actuaries.asn.au

Abstract

Reserving involves forecasting distributions. The distribution of interest incorporates several sources of variability and uncertainty. The new standards and requirements of DFA modeling have intensified the need for more statistically based reserving methodologies. We discuss the importance of model diagnostics in assessing two (often neglected) sources of uncertainty, model selection error and model specification error. Since good use of diagnostics can guard against egregious model error, reserving actuaries need the model assessment that diagnostic tools can provide.

We present a number of diagnostics relating to ratio models (including the chain ladder), which help to assess whether the cumulatives are suitable to use as a predictor of the next ones, and if they are, whether some modification of a basic ratio model may be required.

Many standard techniques (such as link ratio methods) can be expressed as special cases of statistical models. This has multiple advantages. Firstly, it makes available a variety of model diagnostics, allowing the actuary to assess the suitability of a given technique to the data. Further, it allows a more direct comparison of the performance of standard methods with more statistical, model-based approaches. Finally, where a standard technique is found to be appropriate to the data, distributions of outstanding claims (not just means and standard deviations) may often be readily computed.

A collection of standard techniques, including the standard chain ladder, are embedded in a more general statistical model, called the extended link ratio family (ELRF). A variety of diagnostic tests are discussed, and the suitability of ratio methods for several triangles (some well-known) is considered. This presentation will discuss and demonstrate ELRF models as applied to real data, using diagnostic statistics to measure their effectiveness. Examples will be given of the issues that have to faced when using ELRF models and the pitfalls that may arise. Common shortcomings of these methods point toward a somewhat different statistical model.

Many of these results have direct implications for complying with GPS210; when an appropriate model is found, statistical approaches can be used to estimate many of the required uncertainties, and compute both the 75th percentile and the mean and standard deviation of the predictive distribution of outstanding claims.

Keywords: General Insurance; Outstanding Claims Reserves; Models; Regression; Diagnostics; Chain Ladder; Ratio Methods; Model Uncertainty

1. Introduction

Modern reserving involves forecasting distributions – it is necessary to form meaningful notions of mean, coefficient of variation, 75th percentile and so on. While regulations do not require an explicit formulation of an entire distribution, making explicit the distribution which the above statistics summarize carries several major advantages. Firstly the model may be critically assessed – is lack of fit to the recent past ascribable to random variation, or model inadequacy? And what kinds of inadequacy? Secondly, full probabilistic models may be readily extended and generalised in various ways, as circumstances (such as model inadequacy) demand. This can be difficult with less formal structures. Thirdly, external information may be formally brought in and assigned appropriate weight in the description of the future.

2 Models, Variability and Uncertainty

2.1 Why we need models

Reserving includes prediction of means, but is much more. Reserving is a matter of forecasting distributions. Forecast of means alone may be quite uninformative about the required reserve, *even when they're correct* – two triangles with similar means but different variability will require different reserves.

However, without some form of stochastic model it's difficult to produce a meaningful predictive distribution. If a model isn't fully specified it becomes difficult to assess whether the model properly describes the patterns in the data - it's very hard to assess model inadequacies. It's very difficult to find what is happening with the variance (which is rarely constant on the dollar scale). But even more is required – even being able to estimate mean and variance does not tell you the 75th percentile of the aggregate of the outstanding, for example.

Explicit models can be formally challenged and assessed. Effects can be measured. Hypotheses can be tested. Opinions about the data may be checked and debates resolved. Good practice demands explicit models so that they can be directly challenged and assessed. Good practice demands explicit models so that judgement can be applied where it is most needed: deciding how the information about the past will inform our view of the future.

2.2 Sources of variability and uncertainty

Data usually provide a good basis on which to forecast the future, though often incorporating some external information as well. Occasionally, because of lack of relevant data (both within the organisation and in the industry) or because the future may be quite different from the past, it may not. Even in that case, the data will usually still provide important input into the prediction exercise.

Whatever the mode, variability (the size of the process noise in the data) affects the spread of the predictive distribution. Uncertainties, such as parameter uncertainty or model selection uncertainty *also* affect the spread of the predictive distribution.

Data problems aside, there are several sources of uncertainty and variability in the inference relating to the forecasting of outstanding claims. Different sources organise them in different ways (e.g. GGN210.1 paragraph 51, PS300 paragraph 41, Chatfield 1995). If the data does seem to provide an appropriate basis from which to forecast the future, how spread out the predictive distribution should be depends on three categories of variation and uncertainty. In rough order of ease of quantifying them, they are:
(i) process variability;
(ii) parameter uncertainty; and
(iii) model uncertainty.

Process noise

Process noise is how much the process varies about its (possibly changing) mean.; this is an essential component of the model. It may be different in different parts of the data. It is also an important component (as important as the description of the mean) of any stochastic model. It is important to measure it, and project its behaviour in the future. It is a major driver of the spread of possible outcomes of future claims.

Process variability is an inherent part of the insurer's business, it does not necessarily reflect variability from the industry as a whole – it is frequently observed to be different from insurer to insurer.

Parameter uncertainty

Given a model, parameters are estimated from the data. Parameter uncertainty occurs because the fitted parameters are partly functions of the process noise. 'True' parameter values are not known. Even when the model is appropriate, slightly different values of the parameters would provide very good descriptions of the data.
Imagine we toss a coin 100 times and obtain 58 heads. We would be naïve to argue that the probability of a head is actually 0.58! A second sample may as easily obtain 52 heads or 63 heads. In the same manner, we pretend we know far more than we do if we project a fitted distribution and treat it as a predictive distribution. Parameter uncertainty projects out into the future, and uncertainty about (for example) claims inflation can be a major component of the spread of the predictive distribution.

Model uncertainty

Some actuaries assume that model uncertainty is not measurable and is therefore its assessment is purely a matter of judgement. This is not the case. Model uncertainty can be split up into two main types:

(i) *model selection uncertainty*
It is rare that anyone seriously entertains only a single possible model for describing the claims process. If there are several plausible models being contemplated, one cannot be certain that any particular one of them is the 'best' one to use. Model selection uncertainty becomes a consideration when several available models with different forecast distributions provide good descriptions of the data. For example, if two models provide similarly good fits, but lead to quite different descriptions of the future, then model selection uncertainty may be substantial.

Fortunately, the impact of this source of uncertainty on predictive distributions can be formally measured and incorporated into predictive distributions, for example, via a Bayesian approach. See section 2.3.

(ii) *model specification uncertainty*
Model specification uncertainty is more difficult to estimate – models not even considered might be as good or better as descriptions of the data. This includes 'parameter evolution error'. Even this category is not entirely the province of judgement.

Firstly, some part of model specification uncertainty may be shifted over to (measurable) model selection uncertainty by expanding the model class to a larger class of models. For example, if none of your models can deal with changing superimposed inflation, the set of models could be expanded to include models that allow this possibility; if only linear models have been considered, the set of models might be expanded to additive models. Expanded model classes, even where they aren't regarded as suitable models for prediction, can give useful information about the adequacy of the model classes they extend. You can see, for example, whether one predictor continues to be useful when another predictor is included in the model.

Secondly, a variety of diagnostics can be used to assess the adequacy of even an expanded model class. For example, residuals can be used to assess the suitability of assumptions about the mean, the variance, distributional shape, and to assess predictive ability of other potential predictors.

While it's important to consider other potential explanations of your data, by the same token you cannot simply continue to make larger and larger models – this would lead to serious overparameterisation. You *cannot* model everything. Indeed, better predictive power may come from a simpler model that appears to be a slightly worse fit to the data – parameter uncertainty can easily come to dominate the variance of a predictive distribution.

Neither of these approaches take account of potential predictors you have not yet imagined to be potentially related; ultimately there's some judgement involved in deciding when it is unlikely any other variables could improve the ability to predict.

Of course these aren't by any means the only possible sources of risk! These are simply the sources most directly related to the model and modelling issues, and are the ones amenable to detection or measurement. Other sources of risk lie outside the scope of this paper.

2.3 What does a range of answers from several methods reveal?

Some actuaries forecast using different methods and use the range of answers as a guide to the uncertainty in the reserve. However, reserve uncertainty is *not* the same thing as sensitivity to choice of method. It is not inclusive of the process variability and parameter uncertainty for the methods that make it up.

That it doesn't include process variability can be seen from a simple example. Imagine you generate a random triangle from some statistical model and produce forecasts from two different methods, say chain ladder and average development factor. Now increase the spread of the data around the statistical model, under the restriction that the a prediction (say the mean) from both methods be unchanged (there are enough degrees of freedom in the data to achieve all of these things). Clearly if the means are unchanged but the variance has increased the predictive distribution should be wider, but the difference in answers is unchanged. Consequently the 'range of different methods' approach doesn't contain process variability. In fact a similar argument shows it also doesn't contain parameter uncertainty. (It is possible to construct circumstances where it's *impacted* by both process variability and parameter uncertainty, but that doesn't imply it includes them in the appropriate amount.)

In fact, at best it's directly related to one component of an estimate of *model selection error*. In fact, it's not even a very good way to assess the contribution of model selection uncertainty – it doesn't weight by how well the models fit, or by number of parameters for example.

The spread of answers of different models is related to model selection error. Let's take a Bayesian perspective for a moment (though it is entirely possible to make a similar non-Bayesian argument). In a Bayesian framework, the posterior predictive distribution for the forecast (not conditional on a particular model) is a mixture distribution over all models under consideration. The predictive distribution *conditional on each model* are weighted by the posterior probability of each models:

$$f(y_f \mid \mathrm{Y}) = \textstyle\sum_k \; f(y_f \mid \mathrm{Y}, \mathrm{M}_k)\; p(\mathrm{M}_k \mid \mathrm{Y}) \,, \qquad (2.3.1)$$

where y_f is a vector of future values, Y is the observed data, and M_k is the kth model. Note that in this expression all parameters have been integrated out. If the models are *a priori* equally likely (which may not be the case for a variety of reasons, such as different numbers of parameters), the model weights will be proportional to the likelihoods of the models.

Consequently, when all of the models under consideration are about equally as good a fit to the data, have similar numbers of parameters and are otherwise regarded as equally reasonable before seeing the data, the variation in the distribution of answers is indeed related to model selection uncertainty. When some models have substantially higher posterior probability, those models should have a greater impact on that measurement.

2.4 The need for a model

To reserve under GPS210 requires a *distribution* of future losses. How does one construct such a forecast distribution?

If we have a probabilistic model for losses, we can: assess the suitability of the model (Does it describe the past data? Does it reflect information we have about the future?); measure process variability (and other distributional features); measure the uncertainty in our estimates; and produce forecast distributions that take these into account.

Without a probabilistic model, how can you assess whether the *distribution* of the past data is described by the model?

2.5 What do we expect from a model?

A forecasting model must explain the important, relevant effects that are present in data. For example, superimposed inflation is often present in loss payments. *Changing* superimposed inflation is very difficult to include in some models, even though (as we see when proper diagnostics are used) it is often present in real data.

A forecasting model should capture the major features in the data, but it should *not* try to do everything, because a model which does attempt to capture all features of the data will be useless for forecasting.

3 Diagnostics

Diagnostics help protect us against serious model misspecification. We saw in section 2 that diagnostics are useful in assessing model selection and model specification uncertainties. Diagnostics are a crucial part of the assessment and challenge of models. Let us look at some diagnostics for ratio models.

3.1 Basic Diagnostics

It is important to see whether model is able to describe features in the data. For example, does it describe the mean? This leads us to consider ways in which the description might be inadequate. We would also want to know whether the model adequately describes the variance, and other features of the distribution.

Link ratio methods take a triangle of data – generally 'cumulative' (cumulative paid losses, cumulative claim numbers, or perhaps incurred losses), though in the US incremental numbers are also quite common, and for some computed (or even selected) ratios, project those out into the future. Let's assume we have a complete triangle of cumulative paid losses.

The mean in a ratio model

At least in respect of the predicted claims, any approach deserving of the name 'ratio method' will assume that the expected next number is proportional to the previous number. Since they generally work with cumulatives, if C_{ij} is the cumulative paid at accident year i development year j, given the previous cumulative, $C_{i,j-1}$, ratio methods assume:

$$\mathrm{E}(C_{ij}) = b_j\, C_{i,j-1} \quad . \qquad (3.1.1)$$

Let's restrict consideration to a single pair of development years (j and j–1), and let $y_i = C_{ij}$ and $x_i = C_{i,j-1}$, so that $\mathrm{E}(y_i|x_i) = b\, x_i$. That is, on average, the next cumulative is a multiple of its predecessor.

If we plot the y_i against the x_i, what would we see? If the ratio assumption is correct, we should see points scattered about a *line through the origin*. Even without a distributional assumption, or even a variance assumption, this is a diagnostic check of the basic ratio assumption portion of a ratio model. So if we think the mean of y increases linearly with x (the current cumulative and previous cumulative) and that line passes through origin, why not plot y vs x for this pair of years and see?

Let us examine the data from Mack (1994). The data are incurred losses for automatic facultative business in general liability, taken from the Reinsurance Association of America's Historical Loss Development Study (1991).

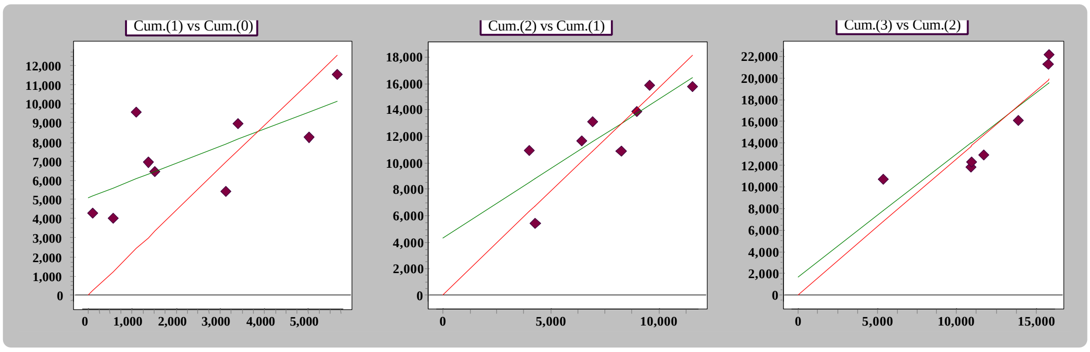

Figure 1. Plots of y vs x for the first 3 pairs of developments. The red line is the least squares line through the origin and the green line is the ordinary least squares line.

The leftmost plot (development year 1 vs development year 0) has a relationship that does not go through the origin. The centre plot also indicates the same thing, but it is somewhat harder to tell that it does not pass through the origin. With the third plot it is harder still.

This is not at all unexpected – it is often the case, especially in later developments, that for some ratios, b, y_i appears to be fairly close (in percentage terms) to $b\,x_i$. This may appear to be encouraging. But a large component of y is already *known*: note that $y = x+p$ (current cumulative = previous cumulative + current incremental). We're using x to predict $x+p$. Well, x predicts x just fine. Let's not congratulate ourselves on our ability to predict the known past. It's p we haven't observed. It is *only* p, the incremental, that is actually being *predicted*. Including x is actually getting in the way of assessing how well we predict what we don't already know. What we see as we move across Figure 1 is simply a result of the previous cumulative becoming a larger and larger proportion of the current cumulative. That is, as we go across the triangle, x becomes larger and larger compared to p, until the relationship between x and x completely overshadows the relationship between p and x.

Following Venter (1998), we see $E(y_i - x_i\,|x_i) = b\,x_i - x_i$, or $E(p_i\,|x_i) = (b-1)\;x_i = r\;x_i$, which has slope 1 less than the slope for y_i. That is, the ratio assumption for cumulatives is equally a ratio assumption about incrementals. If the assumption $E(p|x) = rx$ is tenable, p should also increase linearly with x (with slope 1 less than for the y plot), and pass through the origin. This gives a more powerful check of the basic ratio assumption, since it is much easier to see deficiencies in the ratio assumption.

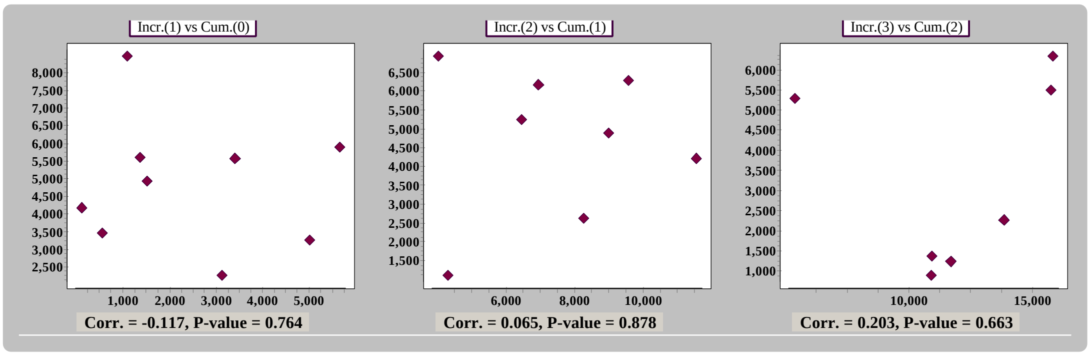

Figure 2. Plots of *p* vs *x* for the first 3 pairs of developments

In order to see more detail in Figure 2, the origins are some distance below and to the left of each plot. Clearly the incremental values for each development year are *not* a multiple of the values in the previous development year. They require an intercept. For the third pair of development years it may be that the relationship isn't linear (or possibly it just has an outlier).

Brosius (1992) gives theoretical reasons why a ratio model might require an intercept and Murphy (1994) notices that intercepts may be required, and derives calculations for the predictive variance when adding intercepts to various kinds of ratio. Note that if the relationship between y and x requires an intercept, then so does the relationship between p and x. One of the main uses of a plot of *p* vs *x* is that it allows checking for an intercept.

Correlation

If an intercept is required, a natural next question is 'does the previous cumulative have any predictive power for the next incremental?'. If *p* and *x* are not correlated, *x* isn't really adding anything to the prediction over just the intercept.

Looking back at Figure 2, the correlations are all fairly small. In fact, all of them are closer to 0 than would be expected for unrelated variables! There is no indication whatever that the previous cumulative has any predictive power for this data.

Let us look at some further examples. Paid loss triangles for four long tail lines of business have been selected from a large database. The arrays were selected haphazardly, without regard for the data, from a list of triangles that didn't have strong payment inflation (we will see the reason for that condition shortly – judgements about inflation were made by a different person from the one selecting the triangles).

Then the plots of *p* vs *x* for the first pair of developments were generated (incremental for development year 1 against development year 0) for each triangle and the correlations between *p* and *x* computed. These are presented in Figure 3 below.

Plainly none of the relationships pass through the origin. Further, the correlations indicate that payments in development year zero are of little value in predicting the paid loss in the next development year. For these data, a ratio model is not adequate, and further, once the intercept is included, the value development year 0 doesn't help to predict the next incremental.

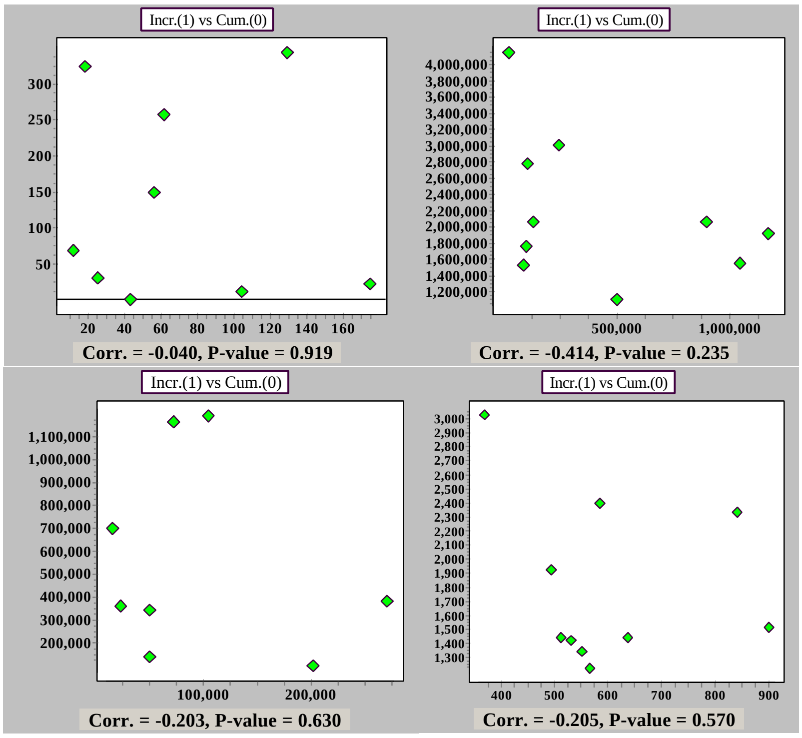

Figure 3. Plots of incremental vs previous cumulative for four triangles

Note that if corr(x, p) = 0, then corr(rx, p) = 0, for any r. That is, if x and p are uncorrelated, *no* ratio has predictive power, *no matter how it was obtained*. Ratio selection by actuarial judgement *can't* overcome zero correlation. If you find that the absolute value of the correlation between incremental and previous cumulative is not stronger than you'd expect for unrelated variables, you have no choice but to abandon the previous cumulative as a predictor (as an explanation for the movements in the data) – *because it does no better than a set of random numbers*.

To continue to use ratios when the previous cumulative is no better at prediction than random numbers cannot be justified. Consequently, failure to check for this circumstance must be similarly unacceptable.

The inflation issue

Why did we exclude cases where inflation is present? An inflationary trend (or, for that matter, an increasing or decreasing trend in exposure) will cause incrementals (and hence cumulatives) in adjacent developments to move up or down together – and hence to be correlated. The explanation for the relationship is a missing predictor (the common inflation or exposure trend) and the previous cumulative is acting as a proxy for it.

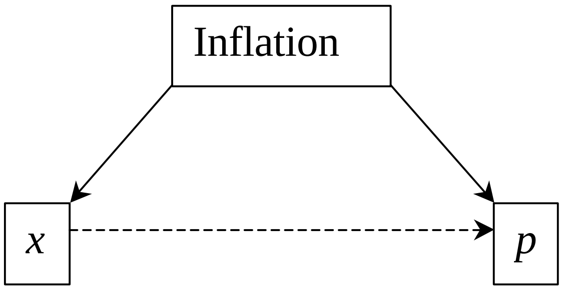

Figure 4. Inflation (or any other source of related trend down the accident years) induces a correlation between *x* and *p*.

It might seem like it would make sense to use a ratio in this circumstance, but if the trend changes partway through the accident periods, the predictive power does as well; it's dangerous to use the previous cumulative as a proxy for the shared trend. Better, if possible, to adjust for the obvious sources of shared trend and see if the ratio has any explanatory power above that. We can do a simple (if rough) adjustment for shared accident period trend by adjusting *x* and *p* for a linear relationship with accident period and graphing the adjusted values to see if any predictive power remains.

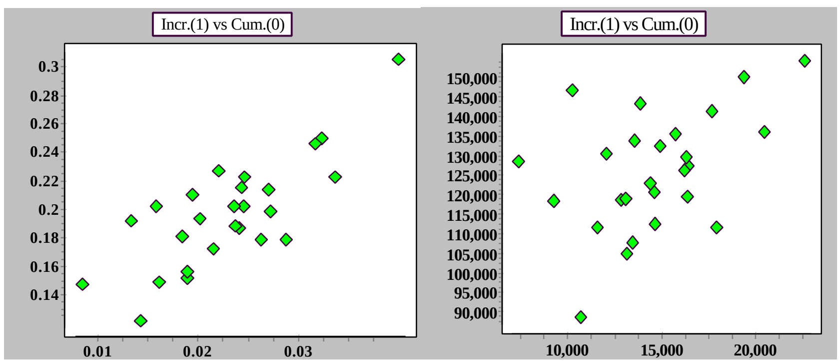

Figure 5 Plot of *p* vs *x* for data with inflation and unadjusted for increasing exposure (left); and after adjusting for linear trend with accident year (right).

We see from Figure 5 that when there are trends down the accident years, the previous cumulative may indeed be correlated with the next incremental (though there is still the need for an intercept!), but after a simple adjustment for the increasing trend, the relationship becomes weak.

It is also possible to look at plots to see if the linear adjustment was a sufficient approximation to the common trend.

However, these adjustments already require regression to compute. At this point it is in fact easier to deal with a formal model, to use statistical estimation to do all the adjustments and use residual plots to do the job we have been doing by less formal means.

In the context of a statistical model, we have a variety of diagnostic tools available, including tests on coefficients, residual plots, and checks on distributional and variance assumptions. If the assumptions are not close to appropriate, prediction intervals will be misleading at best.

3.2 The Extended Link Ratio Family (ELRF)

Consider the regression model

$$y_i = bx_i + e_i,$$

or, equivalently,

$$p_i = (b-1)\, x_i + e_i, \qquad \text{(where } p_i \text{ is the incremental, } y_i - x_i \text{)}$$

in each case with $\mathrm{Var}[e_i] = \sigma^2 x_i^\delta$. Note that the corresponding model for the incrementals is also a ratio model. Note that this isn't yet a complete stochastic model – it's just the mean and variance function. For a complete model we'd need to add that the errors are, for example, independent (or specify something about how they are dependent) and specify a distribution.

Let's make it a complete model, for simplicity taking independent Gaussian errors (these make the usual chain ladder estimates optimal). This assumption may not be reasonable, but it can be *checked*:

$$p_i = (b-1)\, x_i + e_i, \qquad e_i \sim \text{iid } \mathrm{N}(0, \sigma^2 x_i^\delta). \tag{3.2.1}$$

This is now (conditionally on the previous cumulative), a weighted regression through the origin. But in order to more readily assess both the need for an intercept suggested from the plots we've already seen, and assess the usefulness of the previous cumulative. That suggests the following model:

$$p_i = a + (b-1)\, x_i + e_i, \quad e_i \sim \text{iid } \mathrm{N}(0, \sigma^2 x_i^\delta). \tag{3.2.2}$$

Diagnostic information

Consider two possible situations for the parameters:

(i) $b > 1 \quad a = 0$;

Here, the message is that link-ratios are suitable for projection. An intercept that is not significantly different from 0 is consistent with the ratio assumption.

(ii) $b = 1 \quad a \neq 0$;

This means that $x(i)$ has no predictive power in forecasting $y(i) - x(i)$. The estimate of a is a weighted average of the incrementals in development period j and we would forecast a future incremental by averaging the incrementals down its development period. The standard link ratio technique is abandoned in favour of averaging incrementals for each development period down the accident periods.

As discussed in section 3.1, if the incrementals possess a trend down the accident periods, the estimate of the parameter b in equation (14.7) will be significant. We should therefore incorporate an accident period trend parameter for the incremental data.

Accident year trends

In order to get full value out of this diagnostic model (we ***don't*** advocate it as a predictive model), we would also like to be able to remove the bulk of any inflation trend. We can do that in a simple way, simply removing a linear trend. That leads to the following:

$$p_i = a + (b-1)\, x_i + l\, z_i + e_i, \qquad e_i \sim \text{iid } N(0, \sigma^2 x_i^{\delta}). \qquad (3.2.3)$$

where z_i is a variable containing the accident year that p_i is in, to allow us to remove at least the linear part of the accident year trend (of course there are far better ways to model trends such as superimposed inflation; this again emphasises that the purpose of this model is as a diagnostic tool, not a predictive model). This is the Extended Link Ratio Family of models.

These models (as regression models between consecutive pairs of developments) condition on the previous cumulative. At first glance this appears to be treating as fixed something that in the equation for the previous pair of years is random. This seeming inconsistency is in fact this is not an issue. The likelihood for all years (after development year 0) combined can be decomposed into the products of likelihoods that come from these pairwise conditionals. This means that it is in fact quite consistent to have each variable being a random variable in one equation and being conditioned on in the next. Details are in Barnett and Zehnwirth (2000).

3.3 Model Diagnostics and Residual Plots

In Figure 2 and the subsequent discussion on correlation, we looked at correlations for the Mack data and saw that the ratio had little additional explanatory power over the intercept. With the ELRF model we can obtain similar information – the large standard error of the slope term relative to the estimate (i.e. small parameter t-ratios) similarly suggest there's little predictive power for the slope term for this data.

Which residual plots should we look at? Plots vs the predictor, x_i will show somewhat similar patterns to the *p* vs *x* plots; in practice we often just rely on p vs x plots for this purpose. We can also plot residuals against fitted values. These can be combined across years. This gives us much more power to pick up lack of intercept problems than looking at each year individually. Even though the intercepts vary, the basic trend when intercepts are required is the same:

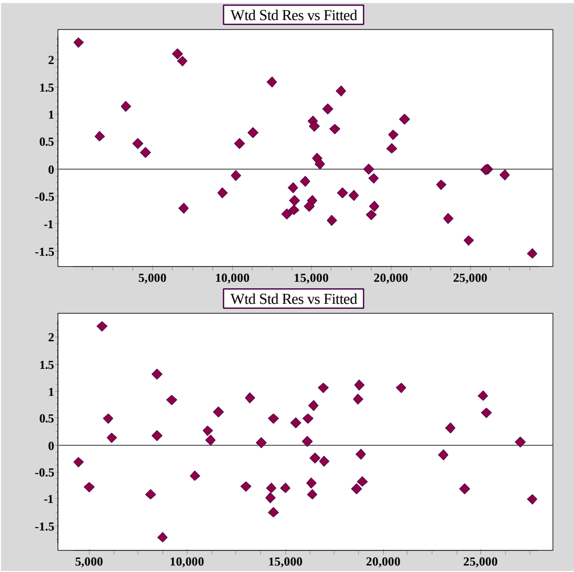

Figure 6. Standardised residuals vs fitted values for a chain ladder model fitted to the Mack data, and for a ratios + intercepts model to the same data.

Residuals against accident and development period may be done, but these won't show noticeable lack of fit in the mean, because standard ratio models are fully parameterised in both of those directions (though the accident year parameters are hidden in the conditioning on development year 0). To check for changing superimposed inflation, residuals can be plotted against calendar (payment) years. For example, Figure 7 below shows a plot of residuals against calendar years for the ABC data presented in Barnett and Zehnwirth (2000); the residuals are from the full ELRF model. Even though the ratios are very smooth and apparently quite stable for this data, there is a serious deficiency in the ratio model!

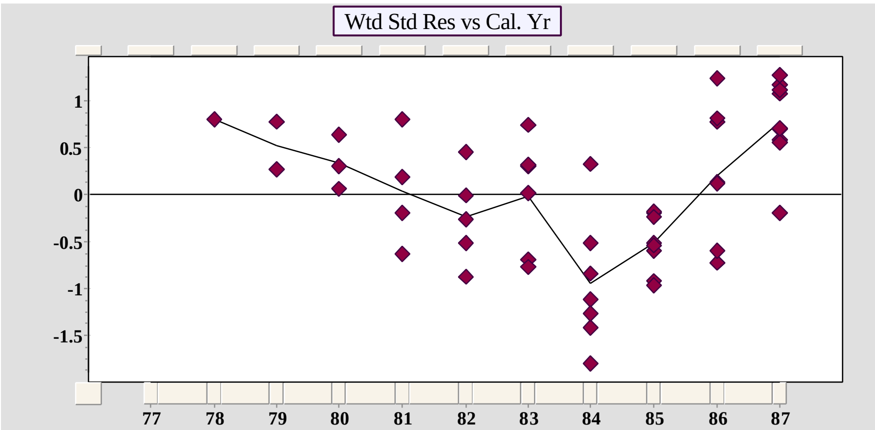

Figure 7. Standardised residuals vs calendar years for the ABC data

Without a model to adjust for the trends in the other directions, this kind of diagnostic is very difficult to perform.

One should also check for correct variance specification – each of the residual plots should not show substantial changes in spread. The most important ones to check are generally residuals against development year (while it's really required for the mean, it's valuable for checking the variance). And the distributional assumption should be checked. We assumed normality for convenience. In some cases the distribution is distinctly right skewed. We find that log-transformed incrementals generally seem be very close to normal – conditional on a suitable model of course (this also makes proper modelling of superimposed inflation convenient).

A general diagnostic tool

A more general diagnostic tool is to simulate multiple samples from your model. If someone can pick the real data from the simulated (other than by knowing the real numbers), then either the data has structure not in the model or the model has structure not in the data. In either case the model doesn't describe the data. This diagnostic tool answers the question 'could this real triangle be generated by this model?'.

4. Modelling the trend

The problems we find in real data suggest that it is necessary to adjust for changes in the three time directions (ratio models already adjust for trends in two directions – development and accident years; it is necessary to be able to adjust for the third). Analysis of many triangles further suggests that ratios often have little predictive power.

The occurrence of superimposed inflation suggests looking at log-incrementals (changes in inflation become changes in slope). Indeed, the log scale is the natural scale for all the ways in which percentage changes occur in claim payments. Further, if the coefficient of variation of incrementals is constant (or nearly constant), then log-incrementals will have nearly-constant standard deviation. The exponential tail in the incremental development of most triangles becomes a linear tail.

One should also note that the reserving exercise is to predict future loss *payments*. Many actuaries use incurred losses rather than paid losses to make these predictions; this raises the obvious question – *do incurred losses contain substantive information not in the past payments about future payments?* This is a question that can be answered scientifically (i.e. by actually examining the data). Firstly, the past paids must be removed, to see if the extra information (the case estimates) tell us anything further. If the calendar period trends in the (increments of) case reserves don't match those in the paids, then predictions based on incurreds will be wrong. In practice it has often been the case that case reserve calendar trends are *much* lower than paid trends. Then again, more recently it is occasionally the case that case reserve trends are much higher (as those setting case estimates overreact to past underestimation). If either of these conditions are true, incurreds might contain some information, but a model relating the trends in the two triangles would have to be built. The next question is also a simple one that may be addressed scientifically – *do case reserve trend changes lead or lag the trend changes in the paid?* Clearly if case reserve changes lag trend changes in the paid, the information they contain is taken from the paids, but later; this would suggest using the paid losses alone would be more informative. In most cases, even when trend sizes correspond, case reserves lag paids. Case estimates occasionally do yield insights, and merit careful analysis alongside the paid losses, but it is frequently the case that the insights that are obtained suggest that case estimates are not suitable for predicting future paid losses.

For identifiability reasons it is not possible to model linear trends in all three directions; we choose to model linear trends in the development and calendar year directions, and model accident year changes as changing levels (possibly after adjusting for exposure).

That is, the model framework is

$$y_{i,j} = \ln(p_{i,j}) = a_i + \sum_{k=1}^{j} g_k + \sum_{t=1}^{i+j} i_t + e_{i,j} \qquad \text{with } e_{i,j} \sim \mathrm{N}(0,\ s_j^2)\ , \qquad (4.1)$$

where $p_{i,j}$ is the incremental, $y_{i,j}$ is here the log-incremental, a_i is the level of accident year i, g_j s a parameter that measures the development trend from j–1 to j, i_t measures inflation (calendar period) trend from calendar period t–1 to t and s_j^2 is the variance in development period j. We refer to this as the Probabilistic Trend Family of models (PTF).

Note that y in this section is defined differently from what it was for ELRF models; this is to allow y to represent a response variable in both cases, so that the regression notation remains familiar in both contexts.

Equation 4.1 is referred to as a model framework. It is not appropriate to use the above expression as a model; it has too many parameters. Instead it defines the *potential* parameters that are used in a modelling exercise. In practice only a few gammas (γ, which describe the percentage trends in the development) are used; the later gammas are normally constant. Similarly, there are often few changes in the accident year levels and the inflation parameters. Further, the variance of the logs is often constant for many development periods.

Let us model the Mack data. The very simple model depicted below gives a remarkably good description of the data:

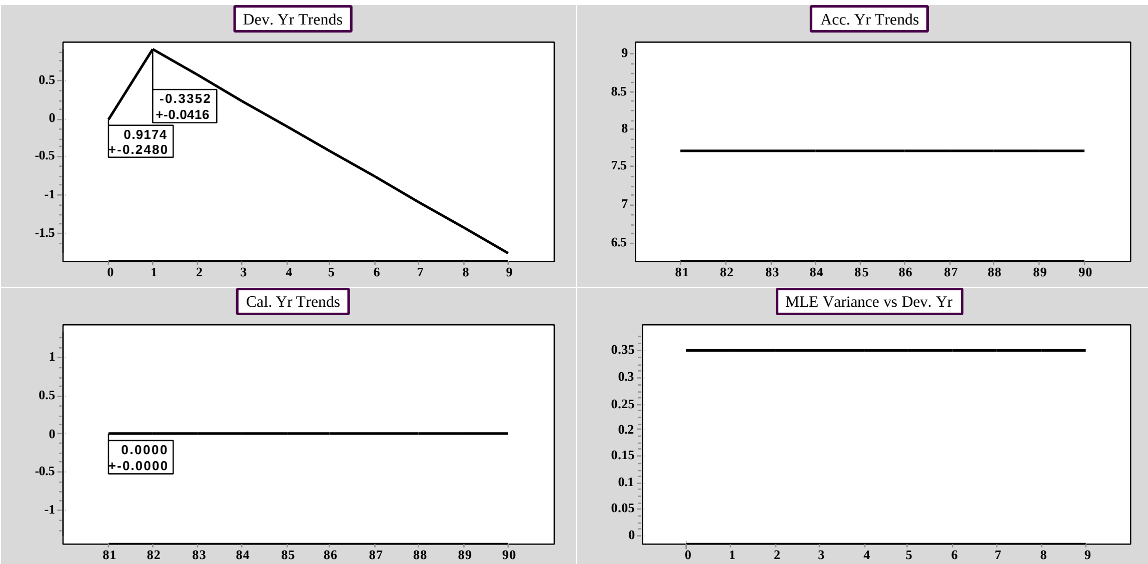

Figure 8. Model display of a PTF model for the Mack data.

Note that the model has changing trends only in the development period direction. This model is consistent with the information we obtain from a full diagnostic analysis of the chain ladder model on this data – that intercepts are required (the accident year level is the log-intercept for the first development; adding the development trend gives the intercepts for the later years), that ratios don't help, that there are no changing calendar trends. This model takes advantage of the fact that the intercepts are themselves related: that the tail of the runoff becomes exponential. The residuals below indicate no major problems with lack of it, though there are some mild indications of remaining trend in all three directions; if this was chosen as a suitable model for forecasting, those suggestions of trend would need to be carefully evaluated again the following year.

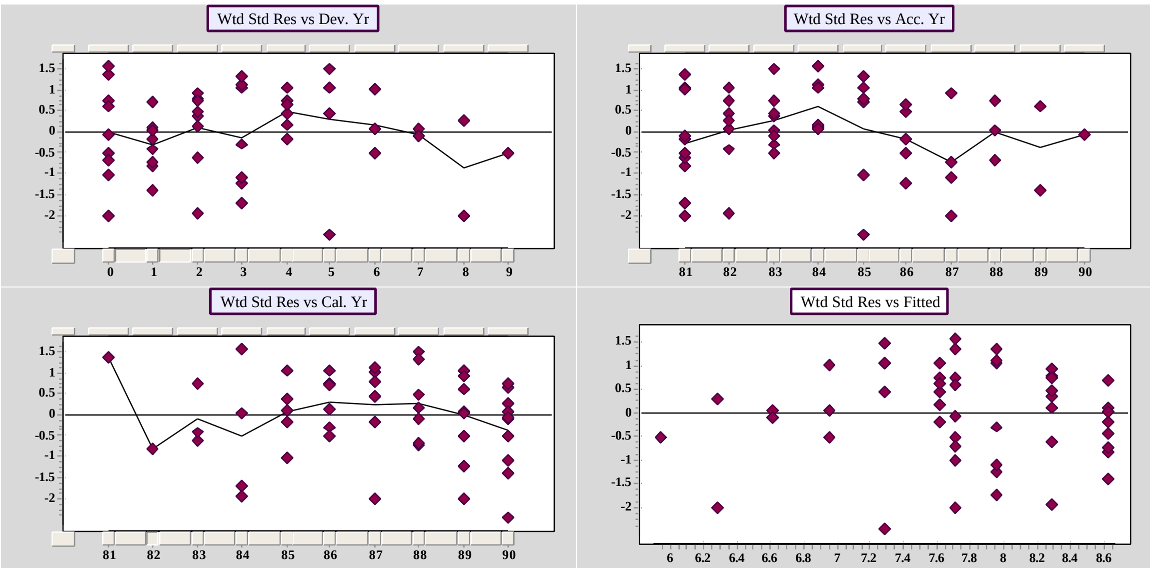

Figure 9. Residual display of the PTF model displayed in Figure 8 for the Mack data.

These and other diagnostics (such as examination of the shape of the distribution of residuals) form an important part of the reserving process; models for past payments must be evaluated critically before they can be regarded as suitable starting points for building a predictive model for future payments. Models must be challenged, or they will often be badly wrong.

REFERENCES

Barnett, Glen, and Ben Zehnwirth, 2000, ‘Best Estimates for Reserves,’ *Proceedings of the Casualty Actuarial Society*, LXXVII. pp. 245-321

Brosius, Eric, 1992, *Loss Development Using Credibility*, Casualty Actuarial Society Part 7 Exam Study Kit.

Chatfield, Christopher, 1995, ‘Model Uncertainty, Data Mining and Statistical Inference,’ *Journal of the Royal Statistical Society*, Series A, 158, Part 3, 419-466.

Cook, R. Dennis and Sanford Weisberg, 1982, *Residuals and Influence in Regression*, New York, NY: Chapman and Hall.

Mack, Thomas, 1993, ‘Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates,’ *ASTIN Bulletin*, 23, 2, 1993, pp. 213–225.

Mack, Thomas, 1994, ‘Which Stochastic Model is Underlying the Chain Ladder Method?’ *Insurance Mathematics and Economics*, 15, 2/3, 1994, pp. 133–138.

Murphy, Daniel M., 1994, ‘Unbiased Loss Development Factors’, *Proceedings of the Casualty Actuarial Society*, LXXXI, 144–155, pp. 154–222.

Reinsurance Association of America, 1991, *Historical Loss Development Study*, 1991 Edition, Washington, D.C.: R.A.A.

Venter, Gary G., 1998, ‘Testing the Assumptions of Age-to-Age Factors,’ *Proceedings of the Casualty Actuarial Society*, LXXXV, pp. 807–847.

Zehnwirth, Ben, 1994, ‘Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals and Risk Based Capital,’ *Casualty Actuarial Society Forum*, Spring 1994, 2, pp. 447–605.