Reserve Ranges, Confidence Intervals and Prediction Intervals

Glen Barnett, Insureware
David Odell, Insureware
Ben Zehnwirth, Insureware

Insureware

Summary

- Uncertainty and variability are distinct concepts
- fundamental difference between confidence interval and prediction interval.
- When finding a CI or PI, assumptions should be explicit, interpretable, testable - related to volatility in the past.
- Differences between CI & PI explained via simple examples and then using loss triangles in the PTF modeling framework
- loss triangle regarded as sample path from fitted probabilistic model
- An identified optimal model in the PTF framework describes the trend structure and volatility about it succinctly – the "*four pictures*"
- model predicts lognormal distributions for each cell + their correlations, conditional on explicit, interpretable assumptions related to past volatility
- Immediate benefits include percentile and V@R tables for total reserve and aggregates, by calendar year and accident year.

Variability and Uncertainty

- different concepts; not interchangeable

"***Variability is a phenomenon in the physical world to be measured, analyzed and where appropriate explained. By contrast uncertainty is an aspect of knowledge***."

Sir David Cox

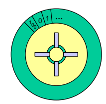
Variability and uncertainty

Process variability is a measure of how much the process varies about its mean – e.g. σ^2 (or σ)

Parameter uncertainty is how much uncertainty in some parameter estimate (e.g. var($\hat{\mu}$) or s.e.($\hat{\mu}$)) or function of parameter estimates (say for a forecast mean – *"uncertainty in the estimate"*)

Predictive variability is (for most models used) the sum of the process variance and parameter uncertainty

Example: Coin vs Roulette Wheel

Coin	***"Roulette Wheel"***
100 tosses *fair* coin (#*H*?) Mean = 50 Std Dev = 5 CI [50,50] In 95% of experiments with the coin the number of heads will be in interval [40,60].	No. 0,1, …, 100 Mean = 50 Std Dev = 29 CI [50,50]  In 95% of experiments with the wheel, observed number will be in interval [2, 97].

Where do you need more risk capital?

Introduce uncertainty into our knowledge - if coin or roulette wheel are mutilated then conclusions could be made only on the basis of observed data.

Example: Coin vs Roulette Wheel

Coin	***"Roulette Wheel"***
100 tosses Mean = ? Std Dev = ? CI [?,?] can toss coin 10 times first (*5 heads –> est. mean 50*)	No. 0,1, …, 100 Mean = ? Std Dev = ? CI [?,?] - similar thing with wheel (more complex)

Parameter uncertainty increases width of prediction interval

Process variability cannot be controlled but can be measured

A basic forecasting problem

Consider the following simple example –

n observations $Y_1 \ldots Y_n \overset{iid}{\sim} N(\mu,\sigma^2)$

$Y_i = \mu + \varepsilon_i \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Now want to forecast another observation...

(Actually, don't really need normality for most of the exposition, but it's a handy starting point.)

A basic forecasting problem

$Y_{n+1} = \mu + \varepsilon_{n+1}$

$\hat{Y}_{n+1} = \hat{\mu} + \hat{\varepsilon}_{n+1}$ ← *forecast of the error term*

μ *known*:

$\hat{Y}_{n+1} = \hat{\mu} + 0$ ← known

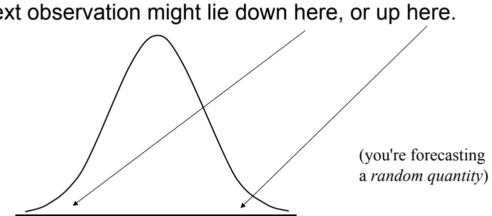
Variance of the forecast is

$\text{Var}(\mu) + \text{Var}(\varepsilon_{n+1}) = \sigma^2$

(Var(μ) ← 0)

A basic forecasting problem

Next observation might lie down here, or up here.



Similarly for future losses: may be high or low.

The *risk to your business* is not simply from the uncertainty in the mean – V@R is related to the amount you will *pay*, not its mean.

A basic forecasting problem

- Even when mean is *known exactly*, still underlying process uncertainty (- with 100 tosses of a *fair* coin, might get 46 heads or 57 heads etc).

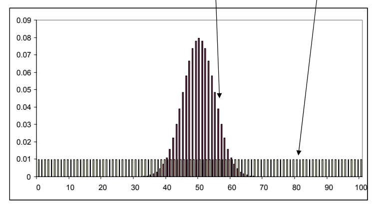
-Except with modeling losses you design a model to describe what's going on with the data.

-It doesn't really make sense to talk about a mean (or any other aspect of the distribution) in the absence of a probabilistic model.

[If you don't have a distribution, what distribution is this "mean" the mean of?]

- assumptions need to be explicit so you can *test* distribution is related to what's going on in the data

- *don't want to use coin model if your data is actually coming from a roulette wheel!*



even worse if you don't get the mean right

Of course, in practice we don't know the mean – we only have a sample to tell us about it.

Let's make a different assumption to before – *now we don't know μ*

- we'll have an estimate of the future mean based – through our model – on past values

- but estimate of the future mean is not exactly the actual mean, even with a great model (*random variation*)

- the *estimate* is uncertain (can estimate the uncertainty - if the model is a good description)

- So we will have an estimate of the mean and we'll also have a *confidence interval* for the mean.

- interval designed so that if we were able to rerun history (retoss our coin, respin our roulette wheel), many times, the intervals we generate will include the unknown mean a given fraction of the time

But that probability relies on the model...

if the model doesn't describe the data, confidence interval is useless (won't have close to required probability coverage)

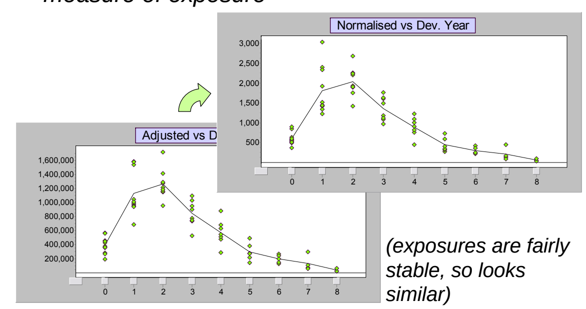
Confidence Interval for Mean of Coin Model (100 tosses) with a small sample (10)

10 tosses, 5 heads. => $\hat{p} = ½$ $\hat{\mu} = 100p$

$\mathrm{Var}(\mu) = 100^2\,\mathrm{Var}(\hat{p}) = 100^2\ ½\ ½\ /10 = 15.8^2$

95% CI for μ $= 50 \pm 1.96 \times 15.8$ $\sim= (19,81)$

Let's look at some real long-tail data. Has been inflation-adjusted and then normalized for a measure of exposure



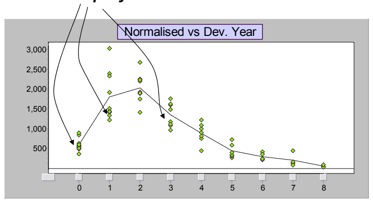
(exposures are fairly stable, so looks similar)

- No trends in the accident year direction

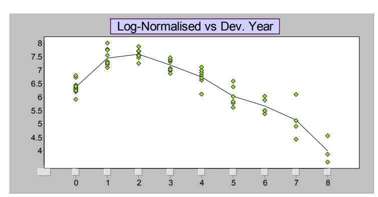
- Calendar years are sufficiently stable for our current purpose (one year is a bit low – it could be omitted if preferred, but we will keep it)

Note a tendency to "clump" just below the mean

– more values below mean: skewness



On the log-scale this disappears, and the variance is pretty stable across years:



Skewness is removed – values much more symmetric about center

(many other reasons to take logs)

Consider a single development - say DY 3:

Normalized	Logs
1,489	7.306
1,606	7.381
1,087	6.991
1,628	7.395
1,178	7.072
1,118	7.019
1,761	7.474
972	6.879

ML estimates of μ, σ: 7.190 and 0.210

(note that the MLE of σ is s_n, with the n denominator, not the more common s_{n-1}).

We don't know the true mean.

Assuming random sample from process with constant mean, predict mean for next value as 7.190

- without some indication of its accuracy, not very helpful:

95% confidence interval for μ: (7.034, 7.345)

Now we want to look at a *prediction*.

Want to predict a random outcome where we don't know the mean
(assuming variance known here, but generally doesn't change prediction intervals a great deal)

Note with coin tossing/roulette wheel experiment, almost never pay the mean (50).

- don't pay mean, pay the random outcome

- risk posed to the business is NOT just the uncertainty in the mean.

To understand your business you need to understand the actual risk of the process, not just the risk in the estimate of the mean.

Want to forecast another observation, so let's revisit the simple model:

$$Y_{n+1} = \mu + \varepsilon_{n+1}$$

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\varepsilon}_{n+1}$$

forecast of the error term

μ UNKNOWN:

So

$$\hat{Y}_{n+1} = \hat{\mu} + 0$$

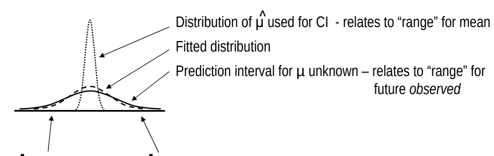
Variance of forecast

$$= \mathrm{Var}(\hat{\mu}) + \mathrm{Var}(\varepsilon_{n+1})$$
$$= \sigma^2/n \quad + \quad \sigma^2$$

(in practice you replace σ^2 by its estimate)

So again, imagine that the distributions are normal.

The next observation might lie –



down here, or up here. (implications for risk capital)

predictive variance

$$= \mathrm{Var}(\widehat{\mu + \varepsilon_{n+1}})$$
$$= \mathrm{Var}(\hat{\mu}) + \mathrm{Var}(\varepsilon_{n+1})$$

(= parameter uncertainty + process var.)

(NB variances add if independent, additive errors)

Alternative way to look at it:

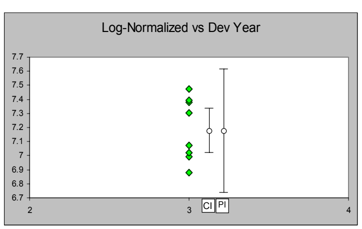
prediction error = $Y_{n+1} - \hat{Y}_{n+1}$.

Predictive variance = var(prediction error)

$= \mathrm{var}(Y_{n+1} - \hat{Y}_{n+1}) = \sigma^2 + \mathrm{var}(\hat{\mu})$

Returning to example:

95% *prediction interval for Y_{n+1}* is (6.75,7.63):



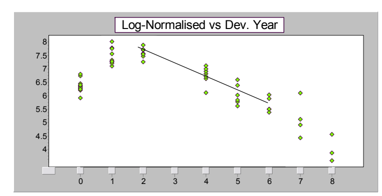
- Parameter uncertainty can be reduced
  - more data reduces parameter uncertainty (more than 10 tosses of the coin in pre-trial).

  In some cases you can go back and get more loss data (e.g. dig up an older year)

  – or eventually you'll have another year of data

- But process variability doesn't reduce with more data
  - an aspect of the process, not knowledge

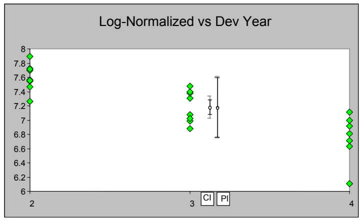
- Note that nearby developments are related:
- If DY 3 was all missing, you could take a fair guess at where it was!



- So in this case we *do* have much more data!

- Need a model to take full advantage of this.

  Even just fitting line through DY2-4 has a big effect on the width of the *confidence interval*:



- Only changes the *prediction interval* by ~2%. So calculated V@R hardly changes

But that prediction interval is on *log* scale.

- To take a *prediction* interval back to the normalized-$ scale, just back-transform the endpoints of the PI
- To produce a *confidence interval* for the *mean on the normalized-$ scale* is harder

  (*can't* just backtransform limits on CI

  – that's an interval for the *median*)

- not a particularly enlightening bit of algebra, so we're leaving the derivation out here

There are some companies around for whom (for some lines of business) the process variance is *very large*.

- some have a coefficient of variation near 0.6.

[so standard deviation is > 60% of the mean]

That's just a feature of the data.

- May not be able to control it, but you sure need to *know* it.

Why take logs?

- tends to stabilize variance
- multiplicative effects (including economic effects, such as inflation) become additive

  (percentage increases or decreases are multiplicative)

- exponential growth or decay → linear
- skewness often eliminated

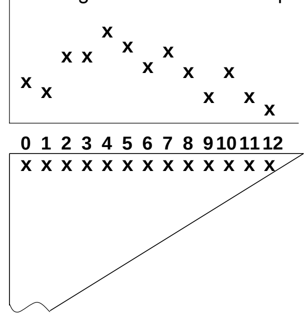
  distributions tend to look near normal

- Using logs a familiar way of dealing with many of these issues (standard in finance)

NB for these to work have to take logs of *incremental* paid, *not* cumulative paid.

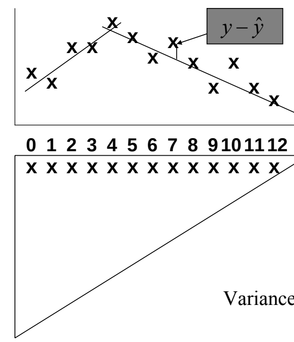
Probabilistic Modelling of trends

e.g. trends in the development year direction



If we graph the data for an accident year against development year, we can see two trends.

Probabilistic Modelling



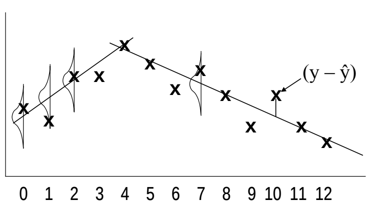
Could put a line through the points, using a ruler.

Or could do something formally, using regression.

$$\text{Variance} = \frac{\sum (y - \hat{y})^2}{13 - 3}$$

Introduction to Probabilistic Modelling

Models Include More Than The Trends



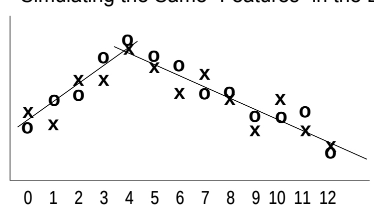
- The model is not just the trends in the mean, but the distribution about the mean

  (*Data = Trends + Random Fluctuations*)

$$\sigma^2 = \frac{\sum (y-\hat{y})^2}{(13-3)}$$

Introduction to Probabilistic Modelling

Simulating the Same "Features" in the Data



- Simulate "new" observations based in the trends and standard errors
- Simulated data should be indistinguishable from the real data

What does it mean to say a model gives a good fit?

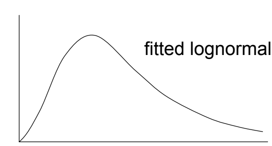
e.g. lognormal fit to claim size distribution

Does not mean we think the model generated the data

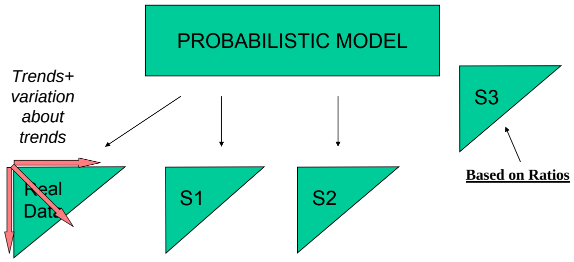
- Real Sample: $x_1,\ldots,x_n$

- Fitted Distribution

- Random Sample from fitted distribution: $y_1,\ldots,y_n$



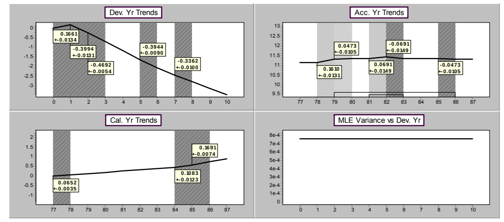
y's look like x's: — Model has probabilistic mechanisms that can reproduce the data



Simulated triangles cannot be distinguished from real data – similar trends, trend changes in same periods, same amount of random variation about trends

Models project past volatility into the future

Trends in three directions, plus volatility



"picture" of the model

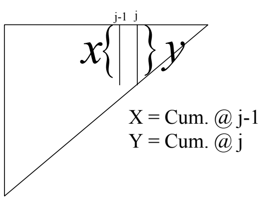
Testing a ratio model

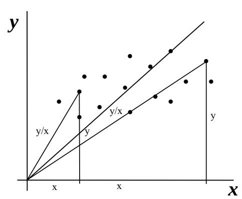
- Since the correctness of an interval ("range") depends on the model, it's necessary to test a model's appropriateness
- Simple diagnostic model checks exist
- Ratio models can be embedded in regression models, and so we can do more rigorous testing – extend this to a *diagnostic model*

ELRF
(Extended Link Ratio Family)

x is cumu. at dev. j-1 and y is cum. at dev. j

- Link Ratios are a comparison of columns
- We can graph the ratios of Y:X - *line through O?*





Using ratios => $E(Y|x) = \beta x$

Mack (1993)

$$y = bx + \varepsilon \quad : \quad V(\varepsilon) = \sigma^2 x^\delta$$

Minimize $\sum w (y - bx)^2$

where $w = 1 / x^\delta$

1. $\delta = 1$, $\hat{b} = \dfrac{\sum x \frac{y}{x}}{\sum x} = \dfrac{\sum y}{\sum x}$

Chain Ladder Ratio (***Volume Weighted Average***)

2. $\delta = 2$, $\hat{b} = \dfrac{1}{n} \sum \dfrac{y}{x}$

Arithmetic Average

Intercept (Murphy (1994))

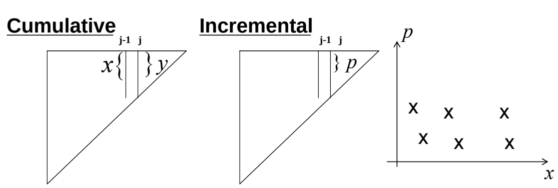
$$y = a + bx + \varepsilon \quad : \quad V(\varepsilon) = \sigma^2 x^\delta$$

Since ***y*** already includes ***x***: ***y = x + p***

$$p = a + (b-1)x + \varepsilon \quad : \quad V(\varepsilon) = \sigma^2 x^\delta$$

Incremental at j — Cumulative at j -1

Is b -1 significant ? Venter (1996)



$$p = a + (b-1)x + \varepsilon \quad : \quad V(\varepsilon) = \sigma^2 x^\delta$$

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

Use link-ratios for projection

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

$\hat{a} = \text{Ave(Incrementals)}$

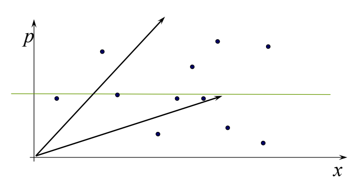
Abandon Ratios - No predictive power

Is assumption $E(p|x) = a + (b-1)x$ *tenable?*

Note: If corr(x, p) = 0, then corr($(b-1)x$, p) = 0

If x, p uncorrelated, *no* ratio has predictive power

Ratio selection by actuarial judgement can't overcome zero correlation.

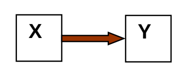


Corr. often close to 0

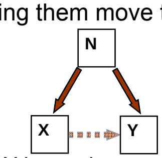
-Sometimes not.
Does this imply ratios are a good model?

- ranges?

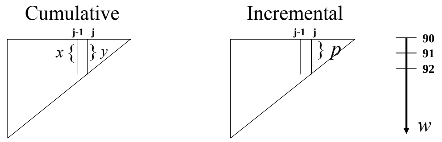
- With two associated variables, tempting to think X causes changes in Y.



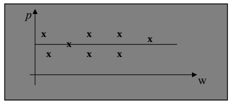
- However, may have *another* variable impacting both X and Y in a similar way, causing them move together:



Here X is a noisy proxy for N;
N is a better predictor of Y (sup. inflation, exposures)

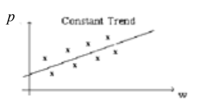


Condition 1:

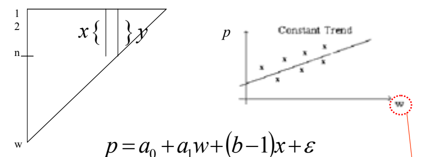


$$p = a + (b-1)x + \varepsilon \;:\; V(\varepsilon) = \sigma^2 x^\delta$$

Condition 2:



Now Introduce Trend Parameter For Incrementals



$$p = a_0 + a_1 w + (b-1)x + \varepsilon$$

a_0 = Intercept

a_1 = Trend

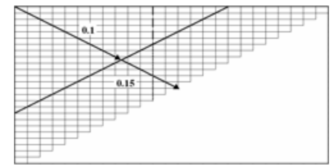
b = Ratio

p vs acci. yr, not previous cumulative

NB: diagnostic model, not predictive model

Condition 3:

Incremental



Review 3 conditions:

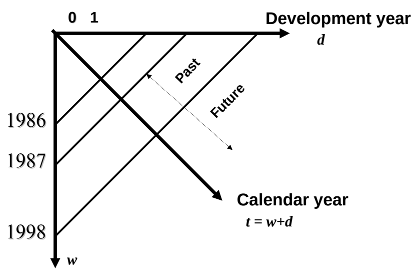
Condition 1: Zero trend

Condition 2: Constant trend, positive or negative

Condition 3: Non-constant trend

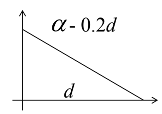
Probabilistic Modelling

Trends occur in three directions:

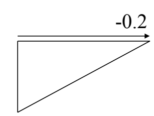


M3IR5 Data

0	1	2	3	4	5	6	7	8	9	10	11	12	13
100000	81873	67032	54881	44933	36788	30119	24660	20190	16530	13534	11080	9072	7427
100000	81873	67032	54881	44933	36788	30119	24660	20190	16530	13534	11080	9072	
100000	81873	67032	54881	44933	36788	30119	24660	20190	16530	13534	11080		
100000	81873	67032	54881	44933	36788	30119	24660	20190	16530	13534			
100000	81873	67032	54881	44933	36788	30119	24660	20190	16530				
100000	81873	67032	54881	44933	36788	30119	24660	20190					
100000	81873	67032	54881	44933	36788	30119	24660						
100000	81873	67032	54881	44933	36788	30119							
100000	81873	67032	54881	44933	36788								
100000	81873	67032	54881	44933									
100000	81873	67032	54881										
100000	81873	67032											
100000	81873												
100000													



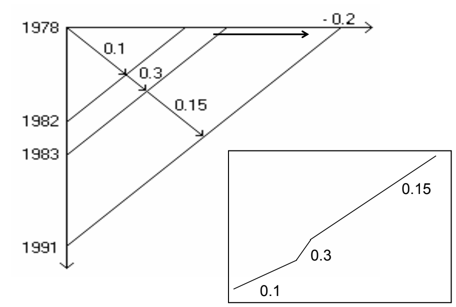
alpha = 11.513



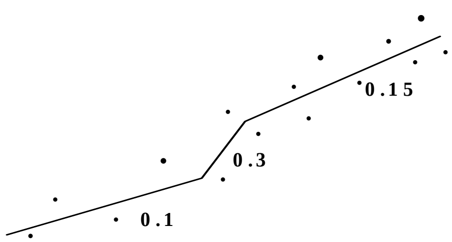
PAID LOSS = EXP(alpha - 0.2d)

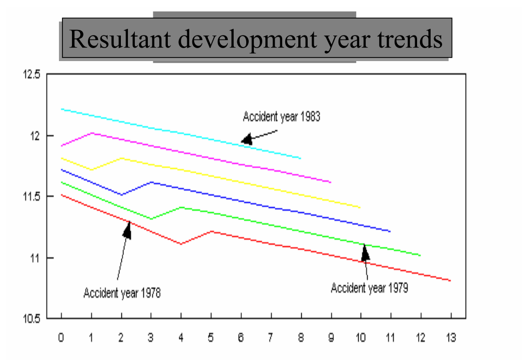
Probabilistic Modelling

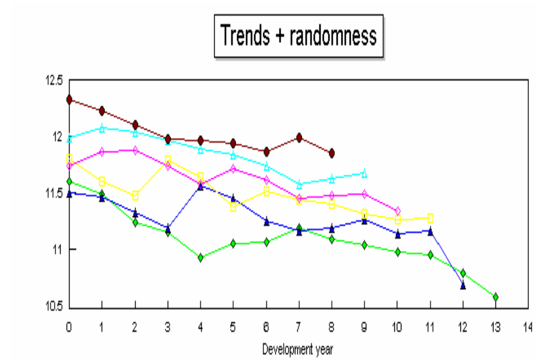
Axiomatic Properties of Trends



Sales Figures







Discussion of some interactive analysis examples (ELRF vs PTF)

Summary of highlights:

Two triangles

- ABC
- LRHigh

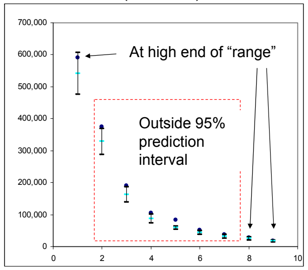
ABC: summary

– chain ladder PIs for next year seem far too low

– doesn't even predict *last* diagonal:

ABC: one-step-ahead prediction of last diagonal

(DYs 1-9)



All values well above mean pred.

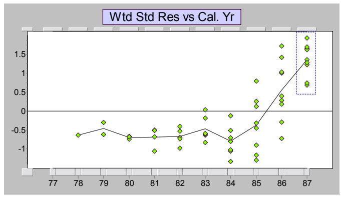
- most above 97½%ile

Why is this?

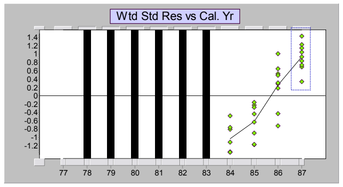
ABC

Chain ladder fit – residuals vs. calendar year

change in cal. trends! (missing feature)



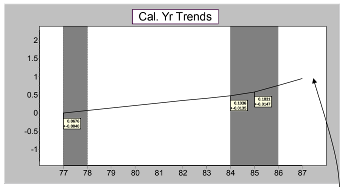
Can we use ratios from last few years?



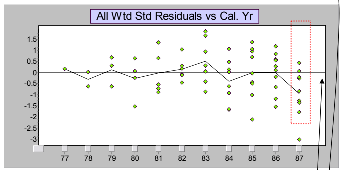
No good – recent ratios still include old incrementals

ABC

PTF fit: Model calendar year trends



Can we predict last year? (not fitted)



(what inflation next year?)

Little high but most points well inside P.I.

---

further example (ELRF vs PTF)

- LR High
  (chain ladder PIs much too *high*)