Simultaneously modeling paid and incurred

B. Posthuma
Posthuma Partners
The Hague

E. Cator
Delft University of Technology
the Netherlands

CLRS Chicago, Concurrent Sessions ST-3

September 14, 2009

Outline

1. Introduction
    - IFRS and Basel Solvency II
    - Flexibility of the model
2. Model description
    - Incremental cells of the run-off tables
    - Structure of parameters
    - Data and estimation
3. Case study simultaneous modeling paid and incurred
    - Setup
    - Method
    - Results
4. Conclusions

Loss reserving

- Modern accounting (IFRS), capital management and global regulation rules put more stringent demands on loss reserving.
- Existing loss reserving methods are struggling to provide an adequate, yet sufficiently flexible, solution.
- We propose a combined stochastic model for paid and incurred run-off tables that provides an excellent tool for modern risk and capital management.

IFRS and Basel Solvency II

International Financial Reporting Standards (IFRS) and Solvency Regulation Rules are leading in the direction of Economic Capital, requiring market valuation of loss reserves:

- Stochastic reserving: a probability distribution for future cash flows.
- Discounting future cash flows.
- Prudence margin.

Flexibility of the model

In anticipation to Solvency II, our model allows:

- Stochastic loss reserving on a continuous time basis, including discounting.
- Adequate assessment of relevant percentiles.
- Flexibility in aggregating various data sets for different branches.
- Flexibility in handling different aggregation levels of the input data, and even in handling missing data.
- Projections of expired risk and risk in force.
- Proper modeling of negative adjustments to losses.
- Important management information: transparent loss ratios and duration claims settlement.

Paid and incurred

The datasets used often consist of two loss triangles, namely paid and reported incurred, together with a measure for exposure.

Many methods and models have been developed for analyzing a single loss triangle.

We model the paid and incurred *simultaneously*. In this talk we will present a case study of real data, showing that this leads to a significantly more accurate prediction of the future payments, when compared to the state-of-the-art model for a single triangle.

Initial definitions

We will use the following notations to define our model for the two run-off tables, the paid and the incurred.

- l indicates the loss period.
- k indicates the development period.
- $Y_{lk}^{(1)}$ indicates the incremental paid.
- $Y_{lk}^{(2)}$ indicates the incremental incurred.

Our goal is to model the vector $(Y^{(1)}, Y^{(2)})$, including all future values.

Auxiliary variables

We use the following auxiliary iid random variables:

$$Z_{lk}^{(1)} \sim N\left(\mu_{lk}^{(1)}, V_{lk}^{(1)}\right)$$

$$Z_{lk}^{(2)} \sim N\left(\mu_{lk}^{(2)}, V_{lk}^{(2)}\right)$$

Define the event

$$R = \left\{ \sum_k Z_{lk}^{(1)} = \sum_k Z_{lk}^{(2)} \quad (\forall l) \right\}.$$

This says that for each loss period, the total amount paid equals the total amount incurred.

Final step

Finally we define the incremental losses by

$$Y^{(1)} \sim Z^{(1)} \mid R$$
$$Y^{(2)} \sim Z^{(2)} \mid R$$

This means that $(Y^{(1)}, Y^{(2)})$ is normally distributed, and that the row sums of the two tables are always equal.

The parameters in this model are given by $\mu_{lk}^{(i)}$ and $V_{lk}^{(i)}$ for $i = 1, 2$. We should reduce the number of parameters.

Product structure of parameters

We choose the following product structure for our parameters:

$$\mu_{lk}^{(i)} = W_l \, e^{(X\alpha)_l} \, \Pi_k^{(i)} \quad (i = 1, 2)$$

$$V_{lk}^{(i)} = \sigma^{(i)} \, W_l \, e^{(X\alpha)_l} \, \tilde{\Pi}_k^{(i)} \quad (i = 1, 2)$$

- W_l is the exposure measure for loss period l.
- $X\alpha$ is a linear model for the loss ratios, with known matrix X and parameter vector α.
- $\Pi^{(i)}$ and $\tilde{\Pi}^{(i)}$ distribute the total expectation, respectively variation, over all development periods:

$$\sum_k \Pi_k^{(i)} = \sum_k \tilde{\Pi}_k^{(i)} = 1 \quad (i = 1, 2).$$

- $\sigma^{(i)}$ $(i = 1, 2)$ are parameters used to tune the total variation.

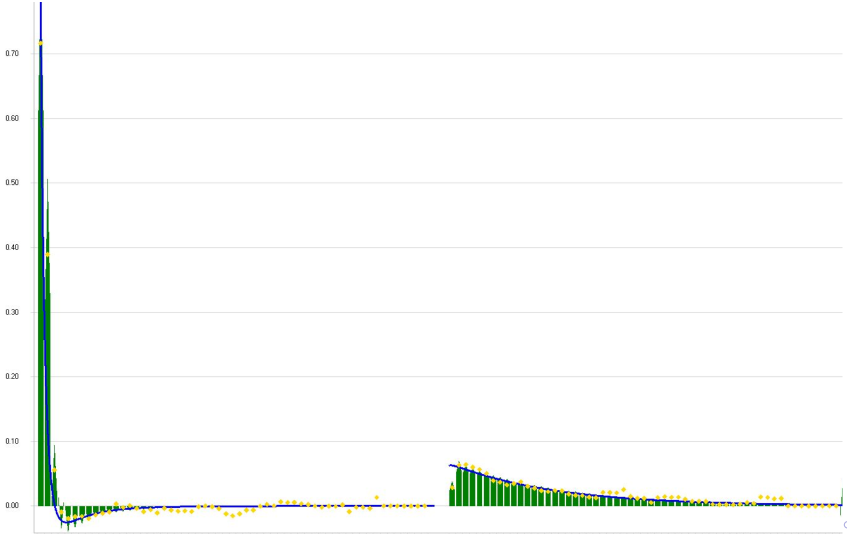
Structure of development curves

We use an extra parameterization of the development curves $\Pi^{(i)}$ and $\tilde{\Pi}^{(i)}$, since otherwise the number of parameters would still be rather large.

This is done by introducing a parametric family of *development functions* $f^{(i)}(s;\theta)$. Intuitively, one could say that the fraction of the total claim amount in a small loss period that is paid (or incurred) in the time interval $[s, s+ds]$ after that loss period, equals $f^{(i)}(s;\theta)\,ds$.
If development period k equals $[s_1, s_2]$ and a loss period has length T, then

$$\Pi_k^{(i)} = \int_0^T \left[\int_{\max(s_1-t,0)}^{\max(s_2-t,0)} f^{(i)}(s;\theta)\,ds \right] dt.$$

Example of development curves



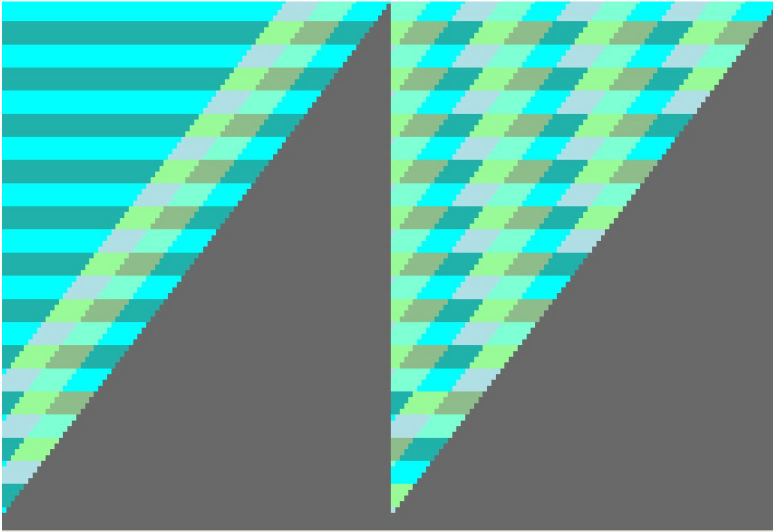
Data and estimation

The data usually consists of observations of aggregated cells of the two run-off tables. We allow for any kind of aggregation, given by some selection-matrix S. So if $Y = (Y^{(1)}, Y^{(2)})$, our data is given by SY.

Note that SY is still normally distributed, with a known distribution, given the parameters. Therefore we can use Maximum Likelihood to estimate all the parameters.

Also prediction is relatively easy, since $Y \mid SY$ is again normally distributed. This structure makes our model very flexible. For a more detailed description, see “Combined analysis of Paid and Incurred Losses”, CAS e-forum, Fall 2008.

Example of aggregation



Goal and data description

Most methods and models used in practice consider only one loss triangle. We would like to show that considering both paid and incurred simultaneously significantly improves future predictions, based on a case study using real data.

We compared our model to a similar model using only the paid run-off table, but with a similar structure as described before (we define $Y^{(1)} = Z^{(1)}$, so there is no conditioning). We consider this a state-of-the-art model for one triangle.

We used 15 recent loss triangles (both paid and incurred) from a wide range of insurance products, varying from liability to car insurance. The time covered by the tables varies from 8 to 22 loss years.

Loss reconciliation

We compared the prediction accuracy of both models using loss reconciliation:

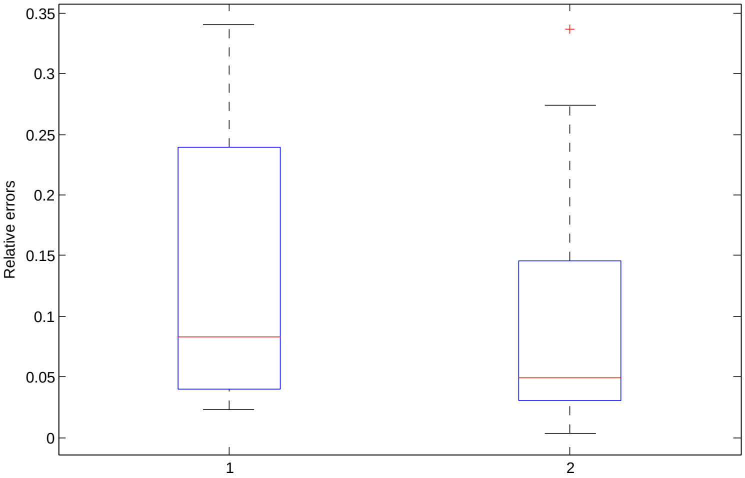
1. Delete the last two or three years of data in the triangles.
2. Fit both models to the pruned triangles.
3. Predict, using the fitted models, the loss of the deleted years.
4. Calculate the relative error of the two predictions, compared to the actual loss.

Note that we can make these predictions for every loss year in every product, greatly increasing the number of pairs of relative errors that we can compare.

Totals

We started by predicting the total loss over the deleted two years for each product. Clearly, these totals can be predicted with greater accuracy than losses for individual loss years.

Boxplot totals



Totals

Average absolute prediction error using 1 triangle:	14%
Average absolute prediction error using 2 triangles:	10%

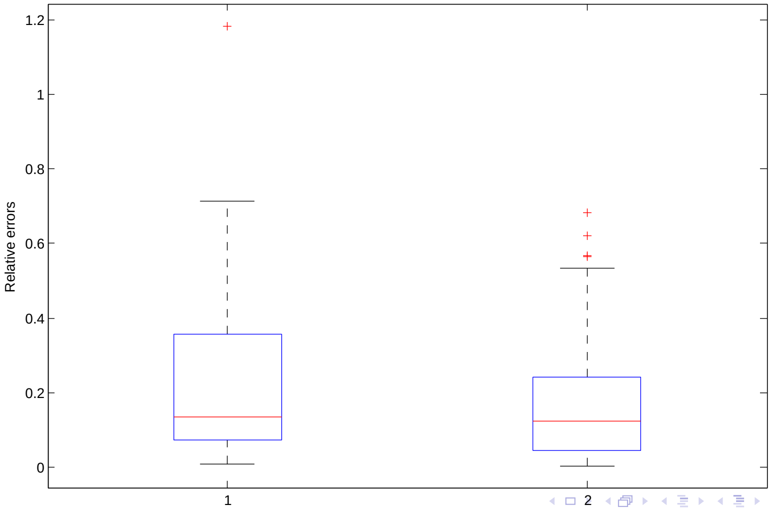
A paired t-test to show that using 2 triangles has a significantly lower average absolute prediction error gives a p-value of 15%. The reason that this is not significant is the relatively low number of totals, namely 15.

Each loss year separately

To increase the number of predictions, we considered the predictions for every loss year separately. Problem: small predicted (or actual) loss corresponds to a high relative prediction error. Since there are many loss years that had a small loss in the years 2007-2009, the high relative errors of these “small years” distort the outcome of our test.

To handle this problem, note that it is more important to accurately predict high losses. We therefore deleted *three* years from our data, 2006-2009, so that the predicted loss over 3 years would be bigger. From this data set, we selected loss years with a predicted loss > 1.5 million or > 3 million. The total loss over 3 years ranged from 6 to 140 million.

Boxplot years



Numerical results

$>$ 1.5 million:

paired t-test p-value $= 0.003$ ($n = 70$)

	1 triangle	2 triangles
mean abs. rel. error	23%	17%
median abs. rel. error	14%	12%

$>$ 3 million:

paired t-test p-value $= 0.025$ ($n = 48$)

	1 triangle	2 triangles
mean abs. rel. error	17%	13%
median abs. rel. error	11%	8%

Conclusions

Our simultaneous model of paid and incurred proves to be a flexible and accurate tool for loss reserving, providing results for real data that are superior to the state-of-the-art one triangle method.