ON THE ART OF SQUARING TRIANGLES

By PETER TER BERG

Posthuma Partners, The Hague, Netherlands

July 6, 2005

abstract: Run-off tables are analyzed as contingency tables. The entry-wise and inner matrix product enables the analysis to be organized without explicit recourse to generalized linear model methodology. The calendar effect is subject to two normalizing restrictions. After normalization this effect appears to fluctuate about a horizontal trend through the origin.

keywords: run-off, contingency table, entry-wise product, inner product, loglinear, Hankel matrix, calendar effect, normalization, optimization

1 | Introduction

Completing run-off tables is a well-known forecasting activity in applied actuarial work. The basic art of this has evolved from a triangular observational structure together with common sense. Later on it appeared that the structure could be embedded in a generalized linear model framework, which reshapes the basic cell data into a column vector and a design matrix embedding the row, column and calendar effects.

The motivation for writing this paper is simplifying the mathematics, which will more easily reveal the properties of the models. Notation, computation, interpretation and communication for such run-off table models become more swift and transparent. This applies in particular to the incorporation of an exposure matrix and the analysis of the calendar effect. It appears that the calendar effect is subject to two normalizing restrictions, which implies a logarithmic calendar effect that fluctuates about zero with a zero-trend. Furthermore the data structure will be kept in matrix format. A design matrix for the row, column and calendar effects does not show up.

The look and feel of the reformulated mathematics, mainly its matrix notation, will be quite different of what one is used to in run-off analysis until now. For instance we will deliberately use a Boolean matrix, which indicates the valid observations of the run-off table. Furthermore matters are easily arranged using the entry-wise matrix product $\circ$ and the inner (dot) product $\bullet$ which together with other entry-wise operations are reviewed in an appendix.
An obvious finding of this matrix framework is that the models are best seen as contingency table models. So it does not matter whether the run-off table is analyzed as arranged or in transposed form. Furthermore for the row-column models, permuting rows and/or columns does not matter either.

The mathematics and the models only focus on the specification for the expected value. We will not entertain on variances as such. It will become clear that improving these models for aspects of variance will destroy much of the simplicity and appeal for these contingency table models. We will close with a plea that further investment in these contingency table models will not be a promising avenue. Behind the scene of the discretized run-off table observations is a time-continuous process. Continuous in loss time, but also continuous in waiting time duration due to reporting and settlement. Smooth specifications on the continuous time level will dictate mathematically derived specifications for the discretized time observed level. This will make the contingency table approach obsolete.

2 | Structural Aspects of Run-Off

Typically a run-off table results as a cross tabulation of aggregate claim size where the vertical axis denotes loss periods and the horizontal axis accounting periods. The result is an upper triangular matrix with incremental data:

$$\begin{bmatrix} * & * & \cdots & * \\ & * & & \vdots \\ & & \ddots & \vdots \\ & & & * \end{bmatrix}$$

When we replace the accounting period by a development duration defined as:

$$\text{development duration} = 1 + \text{accounting period} - \text{loss period}$$

the upper triangular form switches to:

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & & \unicode{x22F0} & \\ \vdots & \unicode{x22F0} & & \\ y_{n1} & & & \end{bmatrix}$$

In this $n \times n$ matrix there are missing elements which need prediction. We introduce a (0,1)-matrix **S** where 1 denotes a valid observation and 0 a missing or invalid observation. Corresponding to **Y** we have:

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & & \unicode{x22F0} & 0 \\ \vdots & \unicode{x22F0} & \unicode{x22F0} & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \tag{2.1}$$

There is no particular need to confine to perfect triangular structures. Often data on old payments is missing such that trapezium-like structures result. So, we allow **Y** to have $m \neq n$ rows and redefine **S** likewise. The case where only future observations are missing will be denoted as the standard observational case. So, (2.1) is standard just as:

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & & & \vdots \\ 1 & 1 & & & & 1 \\ 1 & & & & \iddots & 0 \\ \vdots & & & 1 & \iddots & \vdots \\ 1 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix}$$

A trapezium-like structure will have a matrix **S** like:

$$\mathbf{S} = \begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 \\ \vdots & \iddots & \iddots & & \iddots & 0 \\ 0 & \iddots & & & \iddots & \vdots \\ 1 & & & & & \vdots \\ \vdots & \iddots & \iddots & & & \vdots \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

and even here we may have a hole in the inner part of the trapezium.

An often-neglected data aspect is formed by the exposure for **Y**, which we denote as a $m \times n$ matrix of volumes **V**. Often it will be a rank-1 matrix or it has all its elements put to 1. In what follows we allow it to be a general matrix with non-negative elements. This allows us to study whether and when an explicit exposure measure really matters. A possible quantification for **V** could follow from a cross tabulation of the number of occurrences, such as payment records, which generate **Y**. In this case **V** itself will be an incomplete matrix which is subject to predictive completion.

For computational purposes it will be convenient to replace the missing and non-valid entries of **Y** by zeros. We update **Y** using the entry-wise matrix product:

$$\mathbf{Y} \leftarrow \mathbf{S} \circ \mathbf{Y}$$

So, our data are stored through three $m \times n$ matrices:

$\mathbf{Y}$: incompletely observed incremental contingency table

$\mathbf{S}$: (0,1) – matrix which switches elements of $\mathbf{Y}$ on (1) or off (0)

$\mathbf{V}$: matrix of non-negative exposure volumes

From a notational and computational point of view, it will be convenient to introduce the auxiliary matrix:

$$\mathbf{W} = \mathbf{S} \circ \mathbf{V} \qquad \text{final weights for estimation}$$

We assume that the expected value of the entries of **Y** is proportional with the exposure weights:

$$E[\mathbf{Y} \mid \mathbf{V}] = \mathbf{V} \circ \mathbf{R}$$

As the case may be, this gives the elements of **R** the interpretation of claim frequency, mean claim amount, risk premium or loss ratio.

For parameter estimation we compare $\mathbf{Y}$ with its conditional expectation:

$$E\left[\mathbf{Y} \mid \mathbf{W}\right] = \mathbf{W} \circ \mathbf{R}$$

For prediction purposes, we have the conditional expectation:

$$E\left[\mathbf{Y} \mid \mathbf{V} - \mathbf{W}\right] = (\mathbf{V} - \mathbf{W}) \circ \mathbf{R}$$

Under stochastic independence, the completed contingency table follows as

$$\mathbf{Y} + E\left[\mathbf{Y} \mid \mathbf{V} - \mathbf{W}\right] = \mathbf{Y} + (\mathbf{V} - \mathbf{W}) \circ \mathbf{R}$$

The next steps are a model specification for **R**, normalization aspects and a criterion function for parameter estimation. We will start with the latter to emphasize the simplicity irrespective of model specifications. Then we will continue with an informal review on the classical row-column model and their solutions. After that, matters will be rewritten in loglinear format including a specification for the calendar effect.

Finally, we observe that we may have a valid observation taking on the value 0. This takes place when one observes that no claims or payment records have occurred. So the appropriate exposure weight for such an observation in case of the Gamma and inverse Gaussian model as well as its loglinearized least squares approach should take the value 0 too. In order to cope with such matters we introduce a (0,1)-matrix that indicates whether the final weight matrix **W** is still zero irrespective of whether a valid observation applies

$$\mathbf{Z} = (z_{ij}) \qquad \begin{aligned} z_{ij} &= 1 \qquad \text{if } w_{ij} = 0 \\ &= 0 \qquad \text{if } w_{ij} > 0 \end{aligned}$$

Observe that $\mathbf{S} \circ \mathbf{Z} = \mathbf{W} \circ \mathbf{Z} = \mathbf{Y} \circ \mathbf{Z} = \mathbf{O}$. With the help of **Z**, irrelevant numerical problems caused by 0/0 and $0 \log 0$ are appropriately taken care of and empty cells with information of the type 0=0 get weight 0.

We will use $p = (m + n - 1)$ as a short way for the number of antidiagonals of a $m \times n$ matrix.

3 Estimation Criterion Functions

An obvious way to estimate parameters is minimizing the sum of squared errors or maximizing a likelihood function based on Poisson, Gamma or inverse Gaussian probability distributions. Irrespective of the model specification for the expected values arranged in the $m \times n$ matrix **R** these can be written as simple expressions using the entry-wise and inner matrix product. The simplest expression follows from the unweighted sum of squares:

$$f(\mathbf{R}) = (\mathbf{Y} - \mathbf{W} \circ \mathbf{R}) \bullet (\mathbf{Y} - \mathbf{W} \circ \mathbf{R})$$

However, some reflection on variances is in order. Should these be proportional to $\mathbf{W}, \mathbf{W} \circ \mathbf{R}$ or $\mathbf{W} \circ \mathbf{R}^{(2)}$? In case of simple proportionality on **W** we may write for the standardized squares

$$\begin{aligned} \frac{(\mathbf{Y} - \mathbf{W} \circ \mathbf{R})^{(2)}}{\mathbf{W} + \mathbf{Z}} &= \frac{\mathbf{Y}^{(2)} - 2\mathbf{Y} \circ \mathbf{W} \circ \mathbf{R} + \mathbf{W}^{(2)} \circ \mathbf{R}^{(2)}}{\mathbf{W} + \mathbf{Z}} \\ &= \frac{\mathbf{Y}^{(2)} - 2\mathbf{Y} \circ (\mathbf{W} + \mathbf{Z}) \circ \mathbf{R} + \mathbf{W} \circ (\mathbf{W} + \mathbf{Z}) \circ \mathbf{R}^{(2)}}{\mathbf{W} + \mathbf{Z}} \\ &= \frac{\mathbf{Y}^{(2)}}{\mathbf{W} + \mathbf{Z}} - 2\mathbf{Y} \circ \mathbf{R} + \mathbf{W} \circ \mathbf{R}^{(2)} \end{aligned}$$

where we have used $\mathbf{Y} \circ \mathbf{Z} = \mathbf{W} \circ \mathbf{Z} = \mathbf{O}$. Now, half the weighted sum of squares, deleting irrelevant constants, can be written as:

$$f(\mathbf{R}) = \tfrac{1}{2}\mathbf{W} \bullet \mathbf{R}^{(2)} - \mathbf{Y} \bullet \mathbf{R} \qquad (3.1)$$

In case of Poisson distributed **Y** we get in a similar way for minus the logarithm of the likelihood function, deleting irrelevant constants:

$$f(\mathbf{R}) = \mathbf{W} \bullet \mathbf{R} - \mathbf{Y} \bullet \log[\mathbf{R}] \qquad (3.2)$$

For the Gamma distribution we assume following TER BERG (1980a) a universal shape parameter φ which is like the mean **R** driven by **W**:

$$\mathbf{Y} \sim \text{Gamma entrywise with mean } \mathbf{W} \circ \mathbf{R} \text{ and shape } \varphi\mathbf{W}$$

A similar statement applies, following TER BERG (1980b), to the inverse Gaussian distribution. Both Gamma and inverse Gaussian imply variances entry-wise linear in **W** and quadratic in **R**:

$$V[\mathbf{Y}] = \varphi^{-1}\mathbf{W} \circ \mathbf{R}^{(2)} \qquad (3.3)$$

Denoting $\ln\Gamma$ for the natural logarithm of the Gamma function, we get for the minus loglikelihood for the Gamma case:

$$-\log L(\mathbf{R},\varphi) = \varphi \cdot f(\mathbf{R}) - \varphi \log \varphi \cdot \iota'\mathbf{W}\iota + \mathbf{S} \bullet \ln \Gamma[\mathbf{Z} + \varphi \mathbf{W}]$$

$$\text{where:} \quad f(\mathbf{R}) = \mathbf{Y} \bullet \mathbf{R}^{(-1)} + \mathbf{W} \bullet \log[\mathbf{R}] - \mathbf{W} \bullet \log[\mathbf{Y} + \mathbf{Z}] \tag{3.4}$$

For the Gamma case minimization of $f(\mathbf{R})$ does not depend on φ. Given the parameters for an optimal **R** an empirical counterpart of (3.3) is given by $(\mathbf{Y} - \mathbf{W} \circ \mathbf{R}) \circ (\mathbf{Y} - \mathbf{W} \circ \mathbf{R})$ from which we can derive an easy moment estimate

$$\varphi = \frac{\mathbf{W} \bullet \mathbf{R}^{(2)}}{(\mathbf{Y} - \mathbf{W} \circ \mathbf{R}) \bullet (\mathbf{Y} - \mathbf{W} \circ \mathbf{R})}$$

which can be used as a starting value in an additional line search to find the maximum likelihood estimate for φ.

For the inverse Gaussian case we introduce an auxiliary data matrix:

$$\mathbf{C} = \frac{\mathbf{W} \circ \mathbf{W}}{\mathbf{Y} + \mathbf{Z}}$$

We get for twice the minus loglikelihood, deleting the irrelevant constant:

$$\begin{aligned} -2\log L(\mathbf{R},\varphi) &= \varphi(\mathbf{Y} \bullet \mathbf{R}^{(-1)} + \mathbf{C} \bullet \mathbf{R} - 2\iota'\mathbf{W}\iota) \\ &\quad + (\iota'\mathbf{Z}\iota - mn)\log \varphi + (\mathbf{Z} - \iota\iota') \bullet \log[\mathbf{R}] \end{aligned} \tag{3.5}$$

Analytical optimization with respect to φ results in the estimator for φ conditional on **R**:

$$\hat{\varphi}(\mathbf{R}) = \frac{mn - \iota'\mathbf{Z}\iota}{\mathbf{Y} \bullet \mathbf{R}^{(-1)} + \mathbf{C} \bullet \mathbf{R} - 2\iota'\mathbf{W}\iota}$$

Substituting in (3.5) we get the criterion function:

$$f(\mathbf{R}) = (\iota'\mathbf{Z}\iota - mn)\log \hat{\varphi}(\mathbf{R}) + (\mathbf{Z} - \iota\iota') \bullet \log[\mathbf{R}] \tag{3.6}$$

which should be minimized with respect to the parameters entering **R**.

4 | Review on the Row-Column Model

4.1 Expected value

As a simple model for **R** we consider first the column vector of row totals:

$$\rho = \mathbf{R}\iota$$

This gives ρ the interpretation of ultimate loss ratios or claim frequencies. A rank-1 row-column model for **R** follows from the outer product:

$$\mathbf{R} = \rho\pi' = \begin{bmatrix} \rho_1\pi_1 & \cdots & \rho_1\pi_n \\ \vdots & & \vdots \\ \rho_m\pi_1 & \cdots & \rho_m\pi_n \end{bmatrix} \qquad (4.1)$$

This implies a normalization restriction on the proportionality factors in π:

$$\pi'\iota = 1 \qquad (4.2)$$

If the exposure matrix **V** is also a rank-1 matrix, $\mathbf{V} = \mathbf{uv}'$ say, we have:

$$E[\mathbf{Y} \mid \mathbf{V}] = \mathbf{V} \circ \mathbf{R} = (\mathbf{uv}') \circ (\rho\pi') = (\mathbf{u} \circ \rho)(\mathbf{v} \circ \pi)'$$

which is again a rank-1 matrix. For the Poisson model this creates an identification problem, which is removable by putting the elements of **V** equal to 1.

Next we need a criterion for the estimation of the parameters in ρ and π. Well-known variants are based on the likelihood function for the Poisson and Gamma distribution as well as a nonlinear least squares procedure.

4.2 Poisson

Substituting (4.1) into (3.2) gives:

$$f(\rho, \pi) = \rho'\mathbf{W}\pi - (\mathbf{Y}\iota) \bullet \log[\rho] - (\mathbf{Y}'\iota) \bullet \log[\pi]$$

The first order derivatives with respect to log[ρ] become:

$$\frac{\partial f}{\partial \log[\rho]} = \rho \circ \mathbf{W}\pi - \mathbf{Y}\iota$$

Conditionally on π we may solve this first order condition as:

$$\rho = \frac{\mathbf{Y}\iota}{\mathbf{W}\pi} \qquad (4.3)$$

Next the first order derivatives with respect to log[π] become:

$$\frac{\partial f}{\partial \log[\pi]} = \pi \circ \mathbf{W}'\rho - \mathbf{Y}'\iota$$

and now conditionally on ρ we may solve this first order condition as:

$$\pi = \frac{\mathbf{Y}'\iota}{\mathbf{W}'\rho} \tag{4.4}$$

This suggests an easy optimization procedure. Conditionally on π we determine ρ by (4.3) and continue by determining π by (4.4) conditionally on ρ. Next it may be fruitful to normalize π

$$\pi \leftarrow \frac{\pi}{\iota'\pi}$$

before returning to (4.3) and iterating until convergence. In the contingency table literature such zigzag procedures are known as iterative proportional fitting and in the numerical optimization literature as (block) cyclic coordinate descent or alternating variables methods. Generally their convergence is slow, but in this particular case of two parameter blocks, matters are quite acceptable.

The standard observational case allows a well-known finite step recursive algorithm for completing the run-off triangle: the chain ladder.
For non-standard cases we cannot put the incremental **Y** in cumulative form and the chain ladder breaks down. The iterative proportional fitting procedure still applies, however. Observe that during the iterations and in the optimum we have $\rho'\mathbf{W}\pi = \iota'\mathbf{Y}\iota$.

4.3 Gamma

Substituting (4.1) into (3.4) gives:

$$f(\rho,\pi) = (\mathbf{W}\iota) \bullet \log[\rho] + (\mathbf{W}'\iota) \bullet \log[\pi] + \rho^{(-1)} \bullet \mathbf{Y}\pi^{(-1)}$$

The first order derivatives with respect to $\rho^{(-1)}$ and $\pi^{(-1)}$ become:

$$\frac{\partial f}{\partial \rho^{(-1)}} = \mathbf{Y}\pi^{(-1)} - \rho \circ \mathbf{W}\iota$$

$$\frac{\partial f}{\partial \pi^{(-1)}} = \mathbf{Y}'\rho^{(-1)} - \pi \circ \mathbf{W}'\iota$$

This gives rise to a zigzag solution scheme:

$$\rho = \frac{\mathbf{Y}\pi^{(-1)}}{\mathbf{W}\iota} \qquad \pi = \frac{\mathbf{Y}'\rho^{(-1)}}{\mathbf{W}'\iota}$$

For run-off analysis this procedure was put forward by MACK (1991).

4.4 Nonlinear least squares

Substituting (4.1) into (3.1) gives:

$$f(\rho,\pi) = \tfrac{1}{2}\rho^{(2)} \bullet \mathbf{W}\pi^{(2)} - \rho'\mathbf{Y}\pi$$

The first order derivatives with respect to ρ and π become:

$$\frac{\partial f}{\partial \rho} = \rho \circ \mathbf{W}\pi^{(2)} - \mathbf{Y}\pi$$

$$\frac{\partial f}{\partial \pi} = \pi \circ \mathbf{W}'\rho^{(2)} - \mathbf{Y}'\rho$$

This gives rise to a zigzag solution scheme:

$$\rho = \frac{\mathbf{Y}\pi}{\mathbf{W}\pi^{(2)}} \qquad \pi = \frac{\mathbf{Y}'\rho}{\mathbf{W}'\rho^{(2)}}$$

This algorithm for run-off analysis originates with DE VYLDER (1978). If we apply it to a complete matrix **Y** with no missing values, it generates the first principal component of a matrix, which is related to the singular value decomposition.

4.5 A zigzag family

Introducing a control parameter *c* taking the values −1, 0 and 1 for Gamma, Poisson and nonlinear least squares, the zigzag schemes can be taken together as:

$$\rho = \frac{\mathbf{Y}\pi^{(c)}}{\mathbf{W}\pi^{(c+1)}}$$

$$\pi = \frac{\mathbf{Y}'\rho^{(c)}}{\mathbf{W}'\rho^{(c+1)}}$$

$$\pi \leftarrow \frac{\pi}{\iota'\pi}$$

Although the derivation of this zigzag family is based on integer values of *c* being 0 and ±1, we may apply the algorithm for other real values of *c* too. So, we can display the total expected value of the completion as a function of *c* and see whether its level is sensitive for changes of *c*.

4.6 Numerical illustration

We use the run-off data in DE VYLDER (1978) which is a trapezium structure with $m = 10$ and $n = 6$:

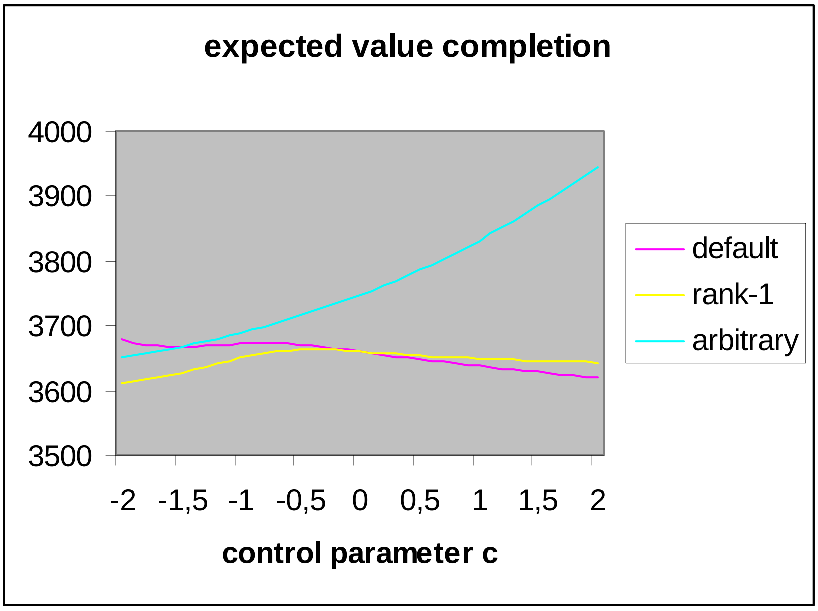
$$\mathbf{Y} = \begin{bmatrix} & & & & & 4627 \\ & & & & 15140 & 13343 \\ & & & 43465 & 19018 & 12476 \\ & & 116531 & 42390 & 23505 & 14371 \\ & 346807 & 118035 & 43784 & 12750 & 12284 \\ 308580 & 407117 & 132247 & 37086 & 27744 & \\ 358211 & 426329 & 157415 & 68219 & & \\ 327996 & 436744 & 147154 & & & \\ 377369 & 561699 & & & & \\ 333827 & & & & & \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We do not have an exposure matrix, so by default $\mathbf{V} = \iota\iota'$. Just to compare this default choice for **V** with something different, we consider a rank-1 matrix which reflects somewhat the first column of **Y** and an arbitrary one:

$$\mathbf{V} = \begin{bmatrix} 88 & 88 & 88 & 88 & 88 & 88 \\ 222 & 222 & 222 & 222 & 222 & 222 \\ 255 & 255 & 255 & 255 & 255 & 255 \\ 265 & 265 & 265 & 265 & 265 & 265 \\ 256 & 256 & 256 & 256 & 256 & 256 \\ 301 & 301 & 301 & 301 & 301 & 301 \\ 342 & 342 & 342 & 342 & 342 & 342 \\ 327 & 327 & 327 & 327 & 327 & 327 \\ 402 & 402 & 402 & 402 & 402 & 402 \\ 334 & 334 & 334 & 334 & 334 & 334 \end{bmatrix} \qquad \text{rank-1} \qquad (4.2)$$

$$\mathbf{V} = \begin{bmatrix} 21 & 31 & 8 & 4 & 2 & 1 \\ 15 & 23 & 6 & 3 & 1 & 1 \\ 35 & 51 & 13 & 7 & 3 & 1 \\ 34 & 51 & 11 & 7 & 4 & 2 \\ 43 & 68 & 17 & 6 & 1 & 2 \\ 47 & 76 & 23 & 6 & 6 & 2 \\ 62 & 67 & 20 & 14 & 5 & 3 \\ 51 & 108 & 23 & 12 & 5 & 3 \\ 73 & 92 & 26 & 12 & 6 & 3 \\ 65 & 97 & 25 & 12 & 6 & 3 \end{bmatrix} \quad \text{arbitrary} \qquad (4.3)$$

For the range $c \in [-2,2]$ we get the following graphical display:



In this numerical illustration, the rank-1 curve starts at the lower end at $c = -2$, reaches a maximum at $c \approx -0.4$ and crosses the default curve at $c = 0$, in agreement with theory. The arbitrary curve is monotone increasing and shows that the quantification for the exposure **V** does matter.

Beware that a flat pattern does not necessarily imply small standard deviations!

5 | Loglinear Mean and Normalization

5.1 Row-Column

This is basically the same as the earlier row-column model but now with the parameters entering in an additive way after a logarithmic transformation:

$$\begin{aligned} &\rho = \exp[\alpha] && \pi = \exp[\beta] \\ &\mathbf{R} = \exp[\alpha\iota' + \iota\beta'] && \log[\mathbf{R}] = \alpha\iota' + \iota\beta' \end{aligned} \tag{5.1}$$

Now a normalizing restriction should be put on β and a symmetric one is:

$$\iota'\beta = 0 \tag{5.2}$$

which is a different from the earlier normalization $\iota'\pi = 1.$ Alternatively we could fix a specific component of β equal to 0.

5.2 Row-Column-Diagonal

A third parameter direction is obtained along the diagonal of the run-off table. In order to study these accounting period effects, calendar effects for short, it will be useful to introduce the Hankel matrix. Any $m \times n$ matrix has $p = (m + n - 1)$ antidiagonals. If each antidiagonal is determined by a single value, the matrix is of the Hankel-type:

$$H(\mathbf{c}) = \begin{bmatrix} c_1 & c_2 & \cdots & c_{n-1} & c_n \\ c_2 & c_3 & \cdots & c_n & c_{n+1} \\ \vdots & \vdots & & \vdots & \vdots \\ c_{m-1} & c_m & \cdots & c_{p-2} & c_{p-1} \\ c_m & c_{m+1} & \cdots & c_{p-1} & c_p \end{bmatrix} = (c_{i+j-1}) \qquad \mathbf{c} \in \Re^p$$

The following linear properties will be convenient for our later purposes:

$$\begin{aligned} &H(a\mathbf{v}) = aH(\mathbf{v}) && a \in \Re, \mathbf{v} \in \Re^p \\ &H(\mathbf{u} + \mathbf{v}) = H(\mathbf{u}) + H(\mathbf{v}) && \mathbf{u}, \mathbf{v} \in \Re^p \end{aligned}$$

Defining a trend vector **t** by

$$\mathbf{t}'_q = [1 \quad \cdots \quad q] \qquad q = m, n, p$$

we have a crucial property:

$$H(\mathbf{t}_p) = \mathbf{t}_m\iota'_n + \iota_m(\mathbf{t}_n - \iota_n)'$$

This will reveal two homogeneous linear restrictions on the parameters of the calendar effect.

Denoting the *p* calendar effects on a logarithmic scale by $\gamma \in \Re^p$ we can display the entry-wise logarithm of the matrix of population means using the Hankel matrix by:

$$\log[\mathbf{R}] = \alpha\iota' + \iota\beta' + H(\gamma) \tag{5.3}$$

This specification has a deeper normalization aspect. To see this we introduce a linear trend regressor matrix $\mathbf{T}$:

$$\mathbf{T} = [\iota \quad \mathbf{t}] = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & q \end{bmatrix} \qquad q = m, n, p \tag{5.4}$$

and do symbolically a least squares regression of γ on $\mathbf{T}$:

$$\begin{aligned}
&\gamma = \mathbf{Ta} + \varepsilon \\
&\hat{\mathbf{a}} = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\gamma && \text{least squares estimator } \mathbf{a} \\
&\hat{\gamma} = \mathbf{T}\hat{\mathbf{a}} && \text{least squares predictor } \gamma \\
&\hat{\varepsilon} = \gamma - \hat{\gamma} && \text{least squares prediction error} \\
&\mathbf{T}'\hat{\varepsilon} = \mathbf{0}_2 && \text{orthogonality restriction}
\end{aligned} \tag{5.5}$$

Using the decomposition $\gamma = \hat{\gamma} + \hat{\varepsilon}$ we rewrite $H(\gamma)$ as:

$$\begin{aligned}
H(\gamma) &= H(\hat{\gamma} + \hat{\varepsilon}) = H(\hat{\gamma}) + H(\hat{\varepsilon}) = H(\mathbf{T}\hat{\mathbf{a}}) + H(\hat{\varepsilon}) \\
&= H(\hat{a}_1\iota + \hat{a}_2\mathbf{t}) + H(\hat{\varepsilon}) = H(\hat{a}_1\iota) + H(\hat{a}_2\mathbf{t}) + H(\hat{\varepsilon}) \\
&= \hat{a}_1 H(\iota) + \hat{a}_2 H(\mathbf{t}) + H(\hat{\varepsilon}) \\
&= \hat{a}_1\iota\iota' + \hat{a}_2\mathbf{t}\iota' + \hat{a}_2\iota(\mathbf{t} - \iota)' + H(\hat{\varepsilon}) \\
&= (\mathbf{T}\hat{\mathbf{a}})\iota' + \hat{a}_2\iota(\mathbf{t} - \iota)' + H(\hat{\varepsilon})
\end{aligned} \tag{5.6}$$

Now we can rewrite the specification for $\log[\mathbf{R}]$ as:

$$\begin{aligned}
\log[\mathbf{R}] &= (\alpha + \mathbf{T}\hat{\mathbf{a}})\iota' + \iota(\beta + \hat{a}_2(\mathbf{t} - \iota))' + H(\hat{\varepsilon}) \\
&= \tilde{\alpha}\iota' + \iota\tilde{\beta}' + H(\tilde{\gamma})
\end{aligned} \tag{5.7}$$

Deleting the tildes, this is exactly in the form of (5.3) which shows that we may start with three linear restrictions on these parameters. Besides a restriction on β like (5.2), we have two restrictions on γ

$$\mathbf{T}'\gamma = \mathbf{0}_2 \tag{5.8}$$

This can be interpreted as that a least squares regression of γ against time gives a horizontal line through the origin. As a result the components of γ fluctuate about 0.

5.3 Column-Diagonal

This is a special case of (5.3) with the row effects $\alpha = \mathbf{0}$. The origin of this specification can be traced to VERBEEK (1972) for Poisson claim count data in an excess-of loss reinsurance environment. For the standard observational case, he derived a finite step algorithm for obtaining maximum likelihood estimates. TAYLOR (1977) gave a heuristic interpretation to Verbeek's algorithm and adopted it for *average* claim amount data under the name of separation technique. There, on page 224, we presumably also find the first appearance of the row-column-diagonal model.

Putting $\alpha = \mathbf{0}$ in (5.3) we get:

$$\log[\mathbf{R}] = \mathbf{Ta}\iota' + \iota\beta' + H(\gamma) \tag{5.9}$$

which is in the format (5.3) with $\alpha = \mathbf{Ta}$ and where β and γ are subject to the normalizing restrictions (5.2-8). If we migrate the constant term of the row effects in $\mathbf{Ta}$ to the column effects, we get:

$$\log[\mathbf{R}] = a\mathbf{t}\iota' + \iota\beta' + H(\gamma) \tag{5.10}$$

where the normalizing restrictions on β don't apply anymore and we only are left with the restrictions on γ given by (5.8).

6 Linearized Least Squares

Adopting the specification for the matrix of means and variances, as these apply for the Gamma and inverse Gaussian based criterion functions, it is natural to consider a linearizing logarithmic transformation. Taking care of nonessential zeros we define:

$$\tilde{\mathbf{Y}} = \log[\mathbf{Y} + \mathbf{Z}] - \log[\mathbf{W} + \mathbf{Z}]$$

An approximation for the mean and variance of the logarithm of a Gamma or inverse Gaussian variate was derived in TER BERG (1980ab):

$$E[\tilde{\mathbf{Y}}] \approx \log[\mathbf{R}] - (2\varphi\mathbf{W})^{(-1)} \approx \log[\mathbf{R}] \qquad \text{for large } \varphi\mathbf{V}$$
$$V[\tilde{\mathbf{Y}}] \approx (\varphi\mathbf{W})^{(-1)}$$

These approximations should properly incorporate $\mathbf{Z}$ for coping with nonessential zeros, but in the weighted sum of squares these cancel out anyway:

$$f(\mathbf{R}) = \mathbf{W} \bullet (\tilde{\mathbf{Y}} - \log[\mathbf{R}])^{(2)}$$

In case of the loglinear row-column model the normal equations boil down to the linear system, denoting $\mathbf{U} = \mathbf{W} \circ \tilde{\mathbf{Y}}$:

$$\begin{bmatrix} (\mathbf{W}\iota)_\Delta & \mathbf{W} \\ \mathbf{W}' & (\mathbf{W}'\iota)_\Delta \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \mathbf{U}\iota \\ \mathbf{U}'\iota \end{bmatrix}$$

Here the suffix Δ denotes, following page 34 in THEIL (1983), the reshaping of a column vector into a diagonal matrix. We extend this to an $m \times n$ matrix **W** to be diagonally reshaped into an $m \times p$ matrix as follows:

$$\mathbf{W}_\Delta = \begin{bmatrix} w_{11} & \cdots & \cdots & \cdots & w_{1n} & 0 & \cdots & 0 \\ 0 & w_{21} & & & & w_{2n} & & \vdots \\ \vdots & & \ddots & & & & \ddots & 0 \\ 0 & \cdots & 0 & w_{m1} & \cdots & \cdots & \cdots & w_{mn} \end{bmatrix}$$

When $n = 1$ this coincides with a column vector transformed into a diagonal matrix. When $m = 1$ a row vector remains unaltered. With this extension of the meaning of Δ we can display the normal equations for the loglinear row-column-diagonal model as:

$$\begin{bmatrix} (\mathbf{W}\iota)_\Delta & \mathbf{W} & \mathbf{W}_\Delta \\ \mathbf{W}' & (\mathbf{W}'\iota)_\Delta & (\mathbf{W}')_\Delta \\ (\mathbf{W}_\Delta)' & ((\mathbf{W}')_\Delta)' & ((\mathbf{W}_\Delta)'\iota)_\Delta \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \mathbf{U}\iota \\ \mathbf{U}'\iota \\ (\mathbf{U}_\Delta)'\iota \end{bmatrix}$$

or using an ad hoc suffix $\Diamond$ which implicitely defines the matrix operation

$$\mathbf{W}_{\Diamond}\xi = (\mathbf{I} \circ \mathbf{U}_{\Diamond})\iota \qquad \text{where} \quad \xi' = [\alpha' \quad \beta' \quad \gamma'] \tag{6.1}$$

This set of $(m+n+p)$ equations in $(m+n+p)$ parameters has zero rows and columns corresponding with components of γ which do not yet correspond with valid observations in the run-off table **Y**. These are visible through the zero columns of $\mathbf{W}_{\Delta}$. Inspired by the normalizing restriction $\mathbf{T}'\gamma = \mathbf{0}$, a natural aliasing value for these components of γ is 0. We design a permutation matrix **P** by permuting the columns of $\mathbf{I}_p$:

$$\mathbf{P} = [\mathbf{P}_0 \quad \mathbf{P}_1] \qquad \mathbf{P}^{-1} = \mathbf{P}'$$

such that $\mathbf{P}_0'\gamma = \mathbf{0}$ form the aliasing restrictions and the components of $\mathbf{P}_1'\gamma$ are subject to parameter estimation. We form a matrix **A**:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{O} & \mathbf{O} \\ \iota & \mathbf{O} & \mathbf{O} \\ \mathbf{0} & \mathbf{T} & \mathbf{P}_0 \end{bmatrix}$$

such that the normalizing and aliasing restrictions on β and γ can be written as $\mathbf{A}'\xi = \mathbf{0}$. Premultiplication with **A** gives a symmetric system

$$\mathbf{A}\mathbf{A}'\xi = \mathbf{0} \tag{6.2}$$

Combining (6.1-2) in an additive way, we get our final invertible set of $(m+n+p)$ equations in $(m+n+p)$ unknown parameters:

$$(\mathbf{W}_{\Diamond} + c\mathbf{A}\mathbf{A}')\xi = (\mathbf{I} \circ \mathbf{U}_{\Diamond})\iota \qquad \forall\, c > 0$$

Solving for ξ we can evaluate **R** and derive a calibration factor:

$$\chi = \frac{\mathbf{S} \bullet \mathbf{Y}}{\mathbf{W} \bullet \mathbf{R}}$$

The use of the calibration factor should undo distortions induced by the linearization and exponential transformation. For the other criterion functions this calibration factor may be useful too. Its value should be about 1 and if it differs firmly from 1, further data analysis seems in order. For the loglinear Poisson model this calibration factor always equals 1.

The loglinearizing approach to run-off analysis is due to KREMER (1982). Observe that in the foregoing nowhere have there been used properties of lognormality. Indeed, the use of the calibration factor χ was deliberately chosen in order to avoid the risk of a specification error in such a lognormal assumption, which is sensitive for the exponential back-transformation. A correction factor under lognormality could be

$$\exp\left(\frac{1}{2}\frac{\mathbf{W}\bullet(\tilde{\mathbf{Y}}-\log[\mathbf{R}])^{(2)}}{\iota'\mathbf{S}\iota}\right) \qquad \text{or} \qquad \exp\left(\frac{1}{2}\frac{\mathbf{W}\bullet(\tilde{\mathbf{Y}}-\log[\mathbf{R}])^{(2)}}{\iota'\mathbf{S}\iota+\mathrm{rk}(\mathbf{A})-(m+n+p)}\right)$$

where r(an)k(**A**) denotes the number of aliasing and normalizing restrictions. This correction factor exceeds 1 whereas the calibration factors can take on values smaller than 1 too.
Furthermore, actuarial ruin theory contains statements of the appropriateness of the Gamma distribution for the aggregate compound claim size, especially the uppertail. Now, the Gamma density is skewed to the right and the logarithm of a Gamma variate has a density which is skewed to the left! As far as a verdict on these matters is possible it seems that a lognormal specification is biased to prudent-prone with a risk for exaggeration. Of course, when such an odd prediction occurs, this will be noticed by a guarding actuary. The need for such a form of safeguard is not a virtue of the model, however.

This linearized least squares estimate could serve as a starting value for the numerical optimization procedure for the loglinear Gamma and inverse Gaussian models.

7 | Computing Loglinear Parameter Estimates

Without the normalization restrictions, the minimization of the criterion function $f(\xi)$ does not result in a unique stationary minimum point. However, the level of this minimum is unique. There are several strategies to perform the optimization. A first approach is to reduce the dimension of the numerical problem to a minimum by first solving parameters in an analytical way. Observe that for the Poisson, Gamma and nonlinear least squares case it will be easy to write the optimal value for α in terms of β and γ. Next normalizing β by fixing a certain element of β to 0, instead of $\iota'\beta = 0$, leaves (n–1) free parameters for this row effect. Finally we have:

$$\gamma = \mathbf{PP}'\gamma = (\mathbf{P}_0\mathbf{P}_0' + \mathbf{P}_1\mathbf{P}_1')\gamma = \mathbf{P}_0(\mathbf{P}_0'\gamma) + \mathbf{P}_1(\mathbf{P}_1'\gamma) = \mathbf{P}_1\tilde{\gamma}$$

Now we are only left with $\mathbf{T}'\mathbf{P}_1\tilde{\gamma} = \mathbf{0}$ as a restriction on $\tilde{\gamma}$. Yet, appealing as this may be, programming code will be easier when stated in the original parameters. Observe that a unique normalized point results by minimizing a combination of the criterion function with a quadratic form of the normalizing restrictions:

$$\tilde{f}(\xi) = f(\xi) + \tfrac{1}{2}c\xi'\mathbf{AA}'\xi \qquad \forall c > 0 \tag{7.1}$$

The investments in matrix manipulation for the display of the normal equations for the linearized least squares approach is also useful here for the display of the gradient and Hessian which makes the Newton-Raphson method an option for optimization. We need specific auxiliary matrices **Q** and **U** for the Poisson, Gamma and nonlinear least squares case:

Poisson	$\mathbf{Q} = \mathbf{W} \circ \mathbf{R}$	$\mathbf{U} = \mathbf{Q} - \mathbf{Y}$
Gamma	$\mathbf{Q} = \mathbf{Y} \circ \mathbf{R}^{(-1)}$	$\mathbf{U} = \mathbf{W} - \mathbf{Q}$
least squares	$\mathbf{Q} = (\mathbf{W} \circ \mathbf{R})^{(2)}$	$\mathbf{U} = \mathbf{Q} - \mathbf{W} \circ \mathbf{Y} \circ \mathbf{R}$

The gradient boils down to

$$\mathbf{g} = \frac{\partial \tilde{f}}{\partial \xi} = (\mathbf{I} \circ \mathbf{U}_{\Diamond})\iota + c\mathbf{AA}'\xi$$

That the Hessian matrix **H** is positive definite in the case of Poisson and Gamma is well-known. It is given by

$$\mathbf{H} = \frac{\partial^2 \tilde{f}}{\partial \xi \partial \xi'} = \mathbf{Q}_{\Diamond} + c\mathbf{AA}'$$

The Hessian for nonlinear least squares has a slight adjustment:

$$\mathbf{H} = (\mathbf{Q} + \mathbf{U})_{\Diamond} + c\mathbf{A}\mathbf{A}' = \mathbf{Q}_{\Diamond} + c\mathbf{A}\mathbf{A}' + \mathbf{U}_{\Diamond}$$

At a stationary point we have $\mathbf{g} = \mathbf{0}$ but **U** will differ from a zero matrix. This may cause an indefinite Hessian. One can doubt whether this forms a real danger in applied work in case of non-pathological run-off tables.

But the most simple approach is to minimize the adjusted function (7.1) with a general purpose optimisation procedure, such as the BFGS quasi-Newton procedure, which is a built-in solver routine for most spread-sheet facilities. Whenever the normalizing and aliasing restrictions are not sufficiently satisfied at the optimum a restart with increased value of *c* will do. Basically, this is all we need. Even the choice of the starting point is not really important and all the matrix manipulation, necessary for gradient and Hessian, can be dropped.

Finally, we remark that "loglinear" does not necessarily mean membership of generalized linear models. Although the Gamma, inverse Gaussian and Poisson distribution are members of the exponential family, this fact does not play a role in the foregoing numerical mathematics. So, it may apply to other distributions not being members of this exponential family.

8 Illustration Calendar Effect

Using **Y** and the default **V** as displayed in paragraph 4.6 the completion for the row-column model and the various criterion functions is as follows:

Table 1: expected value completion row-column (x 1000)					
	linearized				nonlinear
	least		inverse		least
loss year	squares	Gamma	Gaussian	Poisson	squares
1	266	264	268	267	266
2	677	670	681	657	636
3	711	700	715	712	715
4	651	645	655	619	603
5	235	236	237	256	258
6	16	16	16	16	16
7	45	46	46	44	43
8	94	96	95	96	97
9	293	296	295	299	304
10	700	703	704	695	700
total	3688	3672	3711	3660	3638
calibration factor	0,99886	0,99659	0,99281	1,00000	0,99992
calibrated total	3684	3659	3685	3660	3637

In this table the completion for the loss years 6 through 10 corresponds with future payments, whereas the completion for the loss years 1 through 5 concerns a reconstruction.

Using a row-column-diagonal model the following results are obtained:

Table 2: expected value completion row-colum-diagonal (x 1000)					
	linearized				nonlinear
	least		inverse		least
loss year	squares	Gamma	Gaussian	Poisson	squares
1	264	263	265	256	254
2	663	662	665	633	609
3	699	696	702	701	697
4	665	664	667	614	598
5	238	239	239	256	257
6	16	16	16	16	16
7	46	46	46	45	46
8	96	96	96	100	102
9	297	298	299	295	296
10	635	638	637	650	649
total	3618	3619	3632	3566	3525
calibration factor	0,99671	0,99418	0,99271	1,00000	1,00004
calibrated total	3606	3598	3606	3566	3525

In this numerical illustration the level of the completion appears to be somewhat lower than in case of the row-column models.

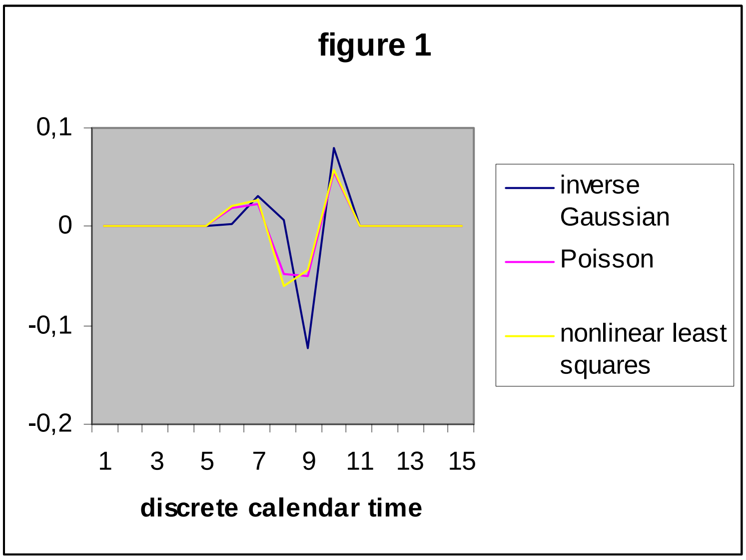
The results for the various models are quite similar and the calibration factors are close to 1.

The logarithmic calendar effects as estimated by the normalized γ-parameters are as follows:

Table 3: logarithmic calendar effects

calendar year	linearized least squares	Gamma	inverse Gaussian	Poisson	nonlinear least squares
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0,0035	0,0036	0,0035	0,0190	0,0219
7	0,0310	0,0301	0,0311	0,0227	0,0261
8	0,0075	0,0075	0,0076	-0,0472	-0,0607
9	-0,1222	-0,1197	-0,1224	-0,0495	-0,0445
10	0,0801	0,0785	0,0802	0,0551	0,0572
11	0	0	0	0	0
12	0	0	0	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	0	0	0	0	0

The estimates for linearized least squares, Gamma and inverse Gaussian are quite similar. This does not come as a surprise as the linearized least squares approach can be viewed as an approximation for Gamma and inverse Gaussian for estimation purposes. A graphical display of these effects for inverse Gaussian, Poisson and nonlinear least squares gives:



9 | Properties and reflections

9.1 Exposure matrix

The specification of the exposure matrix **V** influences the result for the final outcome of the completed run-off matrix. However, differences may be small just as that differences between criterion functions may be small. This should be interpreted that all approaches use the same expected value specification (first moment) and that they only differ as regards the weighing of the evidence (second moments), which is of secondary importance compared with the means. This will particularly apply when we may confine **V** to be (close to) a rank-1 matrix. For the display of prediction intervals or the quantile based choice of prudent provisions the true content of **V** will matter, however.

9.2 Permutation invariance

It is well-known that the chain ladder algorithm produces the same results whether applied to **Y** or to its transpose. This in fact applies to all row-column models whether with standard or nonstandard observational cases. This is due to the fact that the choice of the horizontal axis as development duration or as loss time does not enter the likelihood based criterion functions. When viewed as a contingency table this is obvious.

What may come as a surprise is that permuting rows and/or columns does not have an effect either on the results for the completion. However, again this is obvious from its contingency table nature. If the various values for the loss periods and the development durations were recoded as nominal levels without a chronological order, say the Latin and Greek alphabet, the particular display of the contingency table does not matter.

Of course, a permutation will violate a chronological order which is present in the run-off table. Such a chronological order does not enter in the likelihood based functions, however. This is an instance where good mathematics does not necessarily imply a good model.

9.3 Calendar effects

In this paper calendar effects were treated as Fixed Effects (FE) which were part of the expected value specification and were subject to parameter estimation. A drawback of this approach is the aliasing of the future calendar effects. A more natural treatment would be a Random

Effects (RE) approach where the calendar effects form a latent stochastic process on their own. This requires a careful reconsideration of variances and covariances between the cells of the run-off table in addition to the variances of the calendar effect process. In case a calendar effect is driven by inflationary forces, the cells along a single antidiagonal will have positive correlation. If however a calendar effect is caused by an abundance of claims to be reported, this will create a competition between claims to be reported for the several loss periods. Such a mechanism will induce negative correlations between the cells along the same antidiagonal. So, a RE modelling of the calendar effect creates correlation along the elements of an antidiagonal and will reduce the expected value specification for log[**R**] to a loglinear row-column model.

When analyzing a run-off table with a sudden shock in an observed antidiagonal, it looks obvious to take this into account using a dummy variable. For parameter estimation this practice may make sense. For prediction purposes it is a void affair, however. We do not know when and how sudden shocks in the future will occur. So, we are bound to use a simple model using observational knowledge which is also available when prediction is at stake.

9.4 Continuous time

This paper has reviewed the existing contingency table procedures for run-off table completion. Basically these appear to be models with time entering in a discrete or better discretized manner. Indeed, the true nature of time for the original run-off risk process is continuous. Modeling due to KARLSSON (1974,1976), JEWELL (1989,1990) and HESSELAGER (1995) has a continuous time perspective for the original (unobserved) individual risk process with mathematical implications for the derived (observed) collective risk process. Such a deeper structural approach has much potential to exploit such matters as chronological order giving rise to smooth development patterns. Alternatively, we could add a smoothness criterion on the chronological development parameters β and loss time parameters α. But should we not prefer quality to be built in, not added on? Furthermore a continuous time approach has a natural answer for extrapolation of the uppertail of the development duration, where smoothed contingency table procedures remain arbitrary.

So, discretized run-off table analysis along the lines of contingency table procedures is easy, but future improvements, including variances, should profit from a continuous time approach.

A | Matrix Aspects

A.1 General remarks

We will denote matrices by boldface uppercase and column vectors by boldface lowercase letters. Transposition will be indicated by a prime. A column vector with all components equal to 1 is denoted by ι and the identity matrix by $\mathbf{I}$. Whenever useful a suffix indicates the order. In this paper, most formulations using matrices will use the entry-wise and inner product as compared to standard matrix multiplication. Aspects of standard matrix algebra, such as rank and inverse, are eminently displayed in STRANG (1988). Econometric model formulation using matrices is reviewed by THEIL (1983).

A.2 Entry-wise and inner product

Consider matrices $\mathbf{A}$ and $\mathbf{B}$, both with *m* columns and *n* rows:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$$

The entry-wise matrix product, also known as Hadamard or Schur product, is denoted by $\circ$ and defined as:

$$\mathbf{A} \circ \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & \cdots & a_{1n}b_{1n} \\ \vdots & & \vdots \\ a_{m1}b_{m1} & \cdots & a_{mn}b_{mn} \end{bmatrix} = \mathbf{B} \circ \mathbf{A}$$

For square **A**, $\mathbf{I} \circ \mathbf{A}$ is easy notation for a diagonal matrix with the diagonal of **A** as diagonal.

The inner (dot) product is denoted by $\bullet$ and defined as:

$$\mathbf{A} \bullet \mathbf{B} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij} = \mathbf{B} \bullet \mathbf{A}$$

We have the expressions:

$$\mathbf{A} \bullet \mathbf{B} = \iota'(\mathbf{A} \circ \mathbf{B})\iota = \mathrm{tr}\mathbf{A}'\mathbf{B}$$

where tr(ace) denotes the sum of the diagonal elements of a square matrix. Clearly, $\mathrm{tr}\mathbf{A}'\mathbf{B}$ is more easily evaluated as $\mathbf{A} \bullet \mathbf{B}$.

Observe that both the entry-wise product and the inner product are easily evaluated in electronic spreadsheet facilities.

A.3 Other entry-wise operations

Just like entry-wise multiplication, we can define entry-wise powering:

$$\mathbf{A}^{(k)} = \begin{bmatrix} a_{11}^k & \cdots & a_{1n}^k \\ \vdots & & \vdots \\ a_{m1}^k & \cdots & a_{mn}^k \end{bmatrix} \qquad k \in \Re$$

For an arbitrary function f we will indicate the entry-wise nature by using brackets instead of parentheses:

$$f[\mathbf{A}] = \begin{bmatrix} f(a_{11}) & \cdots & f(a_{1n}) \\ \vdots & & \vdots \\ f(a_{m1}) & \cdots & f(a_{mn}) \end{bmatrix}$$

In this paper this will be used for the exponential exp, the natural logarithm log and the natural logarithm of the Gamma function $\ln\Gamma$.

Occasionally, it will also be convenient to express the entry-wise ratio of two conformable matrices (vectors) by:

$$\frac{\mathbf{A}}{\mathbf{B}} = \begin{bmatrix} \frac{a_{11}}{b_{11}} & \cdots & \frac{a_{1n}}{b_{1n}} \\ \vdots & & \vdots \\ \frac{a_{m1}}{b_{m1}} & \cdots & \frac{a_{mn}}{b_{mn}} \end{bmatrix}$$

instead of $\mathbf{A} \circ \mathbf{B}^{(-1)}$.

B | References

DE VYLDER, F. (1978) Estimation of IBNR Claims by Least Squares. *Bulletin of the Association of Swiss Actuaries*, 249-254.
GRILICHES, Z. and M.D. INTRILIGATOR (1983) *Handbook of Econometrics Volume* I. Amsterdam: North-Holland Publishing Company.
HESSELAGER, O. (1995) Modelling of Discretized Claim Numbers in Loss Reserving. *Astin Bulletin* **25** 119-135.
JEWELL, W.S. (1989) Predicting IBNYR Events and Delays I. Continuous Time. *Astin Bulletin* **19** 25-56.
JEWELL, W.S. (1990) Predicting IBNYR Events and Delays II. Discrete Time. *Astin Bulletin* **20** 93-111.
KARLSSON, J.E. (1974) A Stochastic Model for Time Lag in Reporting of Claims. *Journal of Applied Probability* **11** 382-387.
KARLSSON, J.E. (1976) The Expected Value of IBNR-Claims. *Scandinavian Actuarial Journal* **1976** 108-110.
KREMER, E. (1982) IBNR-Claims and the Two-Way Model of ANOVA. *Scandinavian Actuarial Journal* **1982** 47-55.
MACK, T. (1991) A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves. *Astin Bulletin* **21** 93-109.
STRANG, G. (1988) *Linear Algebra and its Applications*. 3rd edition. San Diego: Harcourt Brace Jovanovich, Inc.
TAYLOR, G.C. (1977) Separation of Inflation and Other Effects from the Distribution of Non-Life Insurance Claim Delays. *Astin Bulletin* **9** 219-230.
TER BERG, P. (1980a) On the Loglinear Poisson and Gamma Model. *Astin Bulletin* **11** 35-40.
TER BERG, P. (1980b) Two Pragmatic Approaches to Loglinear Claim Cost Analysis. *Astin Bulletin* **11** 77-90.
THEIL, H. (1983) Linear Algebra and Matrix Methods in Econometrics. Chapter 1 in GRILICHES and INTRILIGATOR (1983).
VERBEEK, H. (1972) An Approach to the Analysis of Claims Experience in Motor Liability Excess of Loss Reinsurance. *Astin Bulletin* **6** 195-202.