ANALELE ŞTIINŢIFICE ALE UNIVERSITĂŢII „ALEXANDRU IOAN CUZA" DIN IAŞI
Număr special Ştiinţe Economice 2010

STATISTICAL METHODS OF ESTIMATING LOSS RESERVES IN GENERAL INSURANCE

Ion PARTACHI
Academy of Economic Studies of Moldova
Chisinau, Moldova
ipartachi@gmail.com

Oleg VEREJAN
Academy of Economic Studies of Moldova
Chisinau, Moldova

Marcel BRADU
Academy of Economic Studies of Moldova
Chisinau, Moldova

Victoria VEREJAN
Academy of Economic Studies of Moldova
Chisinau, Moldova

Abstract

Relevant estimation of loss reserves related to general insurance activity was and is one of the most important issues for insurance companies. Maintenance of loss reserves at the right level represents the key condition of insurance monitoring authorities as the result of performance indicators of their activity depends on the value of these reserves. The forecasted value of future loss referred to prior events represents the loss reserve. In this paper, we present stochastic methods (Christofides method) of estimating the loss reserves, especially those of incurred but not reported reserves. The stochastic methods presented in the paper, in contrast to the determinist ones, adjust the result better and offer more information referring to the quality of data and exactness level of damage reserve forecast.

Keywords: loss reserves, stochastic models, deterministic models, incurred but not reported reserves(IBNR), the Chain-Ladder method, run-off triangle, predicting future payments, the total estimated loss reserve, the confidence interval for future payments.
JEL classification: C40, C73

1. INTRODUCTION

Almost all actuarial methods for estimating claims reserves have an underlying statistical model. Obtaining estimates of the parameters is not always carried out in a formal statistical framework and this can lead to estimates which are not statistically optimal. These traditional methods generally produce only point estimates.

Formal statistical models are used extensively in data analysis elsewhere to obtain a better understanding of the data, for smoothing and for prediction. Explicit assumptions are made and the parameters estimated via rigorous mathematics. Various tests can then be applied to test the goodness of fit of the model and, once a satisfactory fit has been obtained, the results can be used for prediction purposes.

Deterministic reserving models are, broadly, those which only make assumptions about the *expected* value of future payments. Stochastic models also model the *variation* of those future payments. By making assumptions about the random component of a model, stochastic models allow the validity of the assumptions to be tested statistically, and produce estimates not only of the expected value of the future payments, but also of the variation about that expected value.

A deterministic model simply makes a point estimate of the expected future payments in a given period. The one sure thing we can say about these expected payments, is that the actual payments will be different from expected. Deterministic models do not give us any idea whether this difference is significant. Stochastic models enable to produce a band within which the modeller expects payments to fall with a certain level of confidence, and can be used as an indication as to whether the assumptions of the model hold good.

Also a stochastic model allows to replace the individual data values by a summary that both describes the essential characteristics of the data by a limited number of parameters, and distinguishes between the systematic and random influences underlying the data.

All modelling, whether based on the traditional actuarial techniques such as the chain ladder or on more formal statistical models, requires a fair amount of skill and experience on the part of the modeller. All these models are attempting to describe the very complex claims process in relatively simple terms and often with very little data. The advantage of the more formal approach is that the appropriateness of the model can be tested and its shortcomings, if any, identified before any results are obtained.

The basic chain ladder assumptions can be stated as follows:

a) Each accident year has its own unique level.

b) Development factors for each period are independent of accident year or, equivalently, there is a constant payment pattern (there is a stable pattern to the way that claims have been settled in the past and that this pattern will continue into the future)

These assumptions can now be used to define the model more formally.

Let: X_{ij} – be the incremental paid claims for accident year i paid during development period j

S_i – be the ultimate claim payments for the i-th accident year

p_j – be the proportion of the amount paid during the j-th development period, and $\sum_j p_j = 1$

Thus the *chain ladder model* can be described by the following multiplicative model:

$$X_{ij} = p_j \times S_i \qquad (1)$$

2. ESTIMATING THE PARAMETERS OF THE FORMAL CHAIN LADDER MODEL

As the main set of relations involves products the usual approach is first to make these linear by taking logarithms and then use multiple regression to obtain estimates of the parameters in log-space. It will eventually be necessary to reverse this transformation to get back to the original data space.

Dealing with the main set of equations is relatively easy. Taking logarithms we obtain:

$$\ln(X_{ij}) \approx \ln(S_i) + \ln(p_j) \qquad (2)$$

For ease of reference the parameters are redefined: $\ln(X_{ij})= Y_{ij}$, $\ln(S_i)=a_i$, $\ln(p_j)=b_j$ and an error term introduced. Thus the model to be fitted is described by:

$$Y_{ij} = a_i + b_j + \varepsilon_{ij} \qquad (3)$$

with $b_0=0$ and $\varepsilon_{ij} \sim N(0,\sigma^2)$

We assume that the error values are identically and independently distributed with a normal distribution whose mean is zero and variance some fixed σ^2. If we assume that ε_{ij} has a normal distribution, thus the incremental paid claims X_{ij} has a log-normal distribution:

$$\varepsilon_{ij} \sim N(0,\sigma^2) \Rightarrow Y_{ij} = \ln(X_{ij}) \sim N\left(E[Y_{ij}], Var[Y_{ij}]\right) \Rightarrow$$
$$\Rightarrow X_{ij} = \exp(Y_{ij}) \sim \log N\left(E[Y_{ij}], Var[Y_{ij}]\right)$$
(4)

Thus we have to calculate **the expected value and the variance of Y_{ij}** and in the second step to revert back to the original space.

The expected value of Y_{ij} is: $E[Y_{ij}] = E[a_i + b_j + \varepsilon_{ij}] = a_i + b_j$ (5)

The variance of Y_{ij} is: $Var[Y_{ij}] = Var\left(E[Y_{ij}]\right) + \sigma^2$ (6)

First we have to compute the expected value of Y_{ij}: $E[Y_{ij}] = a_i + b_j$

Parameters a_i and b_j have to be estimated in a multiple regression framework

Suppose that we have a (4×4) data triangle X_{ij}:

	j			
i	**0**	**1**	**2**	**3**
0	x_{00}	x_{01}	x_{02}	x_{03}
1	x_{10}	x_{11}	x_{12}	x_{13}
2	x_{20}	x_{21}	x_{22}	x_{23}
3	x_{30}	x_{31}	x_{32}	x_{33}

ln ⇒

	j			
i	**0**	**1**	**2**	**3**
0	y_{00}	y_{01}	y_{02}	y_{03}
1	y_{10}	y_{11}	y_{12}	y_{13}
2	y_{20}	y_{21}	y_{22}	y_{23}
3	y_{30}	y_{31}	y_{32}	y_{33}

The following table is the form most convenient for the regression facility.

Table no. 1 The design matrix X

year of origin	development year	Y-variables (dependent variables)		Design matrix X (columns are the independent variables)						
i	**j**	X_{ij}	Y_{ij}	a_0	a_1	a_2	a_3	b_1	b_2	b_3
0	0	$x_{0,0}$	$\ln(x_{0,0})$	1	0	0	0	0	0	0
0	1	$x_{0,1}$	$\ln(x_{0,1})$	1	0	0	0	1	0	0
0	2	$x_{0,2}$	$\ln(x_{0,2})$	1	0	0	0	0	1	0
0	3	$x_{0,3}$	$\ln(x_{0,3})$	1	0	0	0	0	0	1
1	0	$x_{1,0}$	$\ln(x_{1,0})$	0	1	0	0	0	0	0
1	1	$x_{1,1}$	$\ln(x_{1,1})$	0	1	0	0	1	0	0
1	2	$x_{1,2}$	$\ln(x_{1,2})$	0	1	0	0	0	1	0
2	0	$x_{2,0}$	$\ln(x_{2,0})$	0	0	1	0	0	0	0
2	1	$x_{2,1}$	$\ln(x_{2,1})$	0	0	1	0	1	0	0
3	0	$x_{3,0}$	$\ln(x_{3,0})$	0	0	0	1	0	0	0

So we have a multiple regression model that in matrix form can be written as:

$$Y = X \cdot \beta + E \qquad (7)$$

with:
Y – the matrix of the dependent variables
X – the design matrix
β – the matrix of the parameters (the independent variables) to be estimated
E – matrix of the error terms

The regression takes the $\ln(X_{ij})$ or Y_{ij} values as the dependent variable and each of the columns of the matrix *X* as the independent variables.

Applying least square estimation, the parameters estimates are given by the following expression:

$$\beta = (X^T X)^{-1} \cdot X^T Y \qquad (8)$$

where:
X – the design matrix
X^T – the transpose of the design matrix
Y – the design of the dependent variables

The coefficients are the parameter estimates and are in the same order as the columns of the design matrix.

3. REGRESSION ANALYSIS: STATISTICAL TESTS

Statistics are necessary to test the appropriateness of the resulting model. In this case we can apply various diagnostic tests to test the significance of the overall model, to test the significance of the parameters, and to test the assumptions specific to error terms.

- The significance of the overall model is tested with *ANOVA* applying and *F* test and calculate the coefficient of determination R^2

$$F = \frac{SSR/p}{SSE/n-p} \tag{9}$$

where: SST – total deviation

SSR – explained deviation

SSE – unexplained deviation

n – number of observation (known payments), p – number of parameters

If the $F \geq F_{\alpha}$ the null hypothesis Ho (all slope coefficient are simultaneously zero) is rejected and the model is accepted.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{10}$$

the closer to 1, the higher the explanatory value of the model

- The significance of the parameters is tested by *t- Student* test. The null hypothesis is Ho: parameter is not significantly different from 0. If *t-stat* is larger than critical value, ($t \geq t_{\alpha}$) Ho is rejected - the parameter is significantly different from 0 and the dependent variable is explained by independent variable.
- To estimate the parameters of the regression model we use a least square method that is by minimize the sum of the squares of the error terms ε_{ij}. We assume that ε_{ij} has a normal distribution $\varepsilon_{ij} \sim N(0,\sigma^2)$. This assumption can be tested applying *Q-Q plots*. Each residual is plotted against its expected value. If the plot is linear normality assumption is satisfied.

The independence of the residual terms can be tested using *the Durbin Watson test*.

4. PREDICTING FUTURE PAYMENTS AND THEIR STANDARD ERRORS

In order to derive estimates of the model parameters it was convenient to take logarithms and work in log-space. To obtain results in the original space it is necessary to reverse this transformation.

Obtaining the parameter estimates in log-space is relatively straightforward. To revert back to the original space is not so simple and it is necessary to use the relationships between the parameters of the log-normal distribution and the underlying normal distribution.

$$E(X_{ij}) = \exp\left(E[Y_{ij}] + 0{,}5 \cdot Var[Y_{ij}]\right) \tag{11}$$

$$Var(X_{ij}) = \exp\left(2 \cdot E[Y_{ij}] + 2 \cdot Var[Y_{ij}]\right) - \exp\left(2 \cdot E[Y_{ij}] + Var[Y_{ij}]\right) \tag{12}$$

So the first step is to derive the predicted values and their variations in log-space. **The predicted values** in log-space are obtained from the estimates of the parameters produced by the regression applying relation (5). For example, the first future value to be predicted is for accident year 1 development year 3 and this is given by $E(Y_{1,3})=a_1+b_3$

To obtain **the variance of Y_{ij}** according to (6), we have to calculate each component:

a) $Var\left(E[Y_{ij}]\right)$ – the variance of $E[Y_{ij}]$

b) σ^2 – the model variance

$Var\left(E[Y_{ij}]\right)$ is obtained from its symmetric variance-covariance matrix:

$$Var\left(E[Y_{ij}]\right) = \sigma^2 \cdot X_F \cdot (X^T X)^{-1} \cdot X_F^T \qquad (13)$$

with:

σ^2 – the model variance

X_F – the design matrix of future value (future design matrix)

X_F^T – the transpose of the future design matrix

$(X^T X)^{-1}$ – the model information matrix

The variance-covariance matrix is a square and symmetric with each side equal to the number of future values to be projected. The diagonal elements contain the variances of each of each of these values and are in the same order as the future design matrix elements.

The design matrix of future values X_F following the same format as the original design matrix:

Table no. 2 The future design matrix X^F

year of origin	development year	Y-variables (dependent variables)	The future design matrix X (columns are the independent variables)						
i	**j**	Y_{ij}	a_0	a_1	a_2	a_3	b_1	b_2	b_3
1	3	$Y_{1,3}$	0	1	0	0	0	0	1
2	2	$Y_{2,2}$	0	0	1	0	0	1	0
2	3	$Y_{2,3}$	0	0	1	0	0	0	1
3	1	$Y_{3,1}$	0	0	0	1	1	0	0
3	2	$Y_{3,2}$	0	0	0	1	0	1	0
3	3	$Y_{3,3}$	0	0	0	1	0	0	1

Thus the variances for the future values in log-space are the sum of the variance-covariance matrix values (values situated on the principals diagonal) and the model variance σ^2

$$Var[Y_{ij}] = Var\left(E[Y_{ij}]\right) + \sigma^2 \qquad (6)$$

When we apply a regression method, the model variance σ^2 is a not known value and we use an unbiased estimator s^2

$$s^2 = \frac{E^T E}{n-p} \qquad (14)$$

with: E – *the matrix of the error terms,* n-p – *the degrees of freedom*

Finally the future values $E(X_{ij})$ and their variation $Var(X_{ij})$ are calculated from the estimates obtained in the log-space according to formulas (11) and (12)

The following table shows the projected values of payments, their variances and standard errors

Table no. 3 The projected values of payments and their variances

i	**j**	**a_i**	**b_j**	***$E(Y_{ij})$***	***$Var[Y_{ij}]$***	***$E(X_{ij})$***	***$Var(X_{ij})$***	**Se(X_{ij})**
1	3	a_1	b_3	$E(Y_{1,3})$	$Var[Y_{1,3}]$	$E(X_{1,3})$	$Var(X_{1,3})$	Se($X_{1,3}$)
2	2	a_2	b_2	$E(Y_{2,2})$	$Var[Y_{2,2}]$	$E(X_{2,2})$	$Var(X_{2,2})$	Se($X_{2,2}$)
2	3	a_2	b_3	$E(Y_{2,3})$	$Var[Y_{2,3}]$	$E(X_{2,3})$	$Var(X_{2,3})$	Se($X_{2,3}$)
3	1	a_3	b_1	$E(Y_{3,1})$	$Var[Y_{3,1}]$	$E(X_{3,1})$	$Var(X_{3,1})$	Se($X_{3,1}$)
3	2	a_3	b_2	$E(Y_{3,2})$	$Var[Y_{3,2}]$	$E(X_{3,2})$	$Var(X_{3,2})$	Se($X_{3,2}$)
3	3	a_3	b_3	$E(Y_{3,3})$	$Var[Y_{3,3}]$	$E(X_{3,3})$	$Var(X_{3,3})$	Se($X_{3,3}$)
Total								

In the table above the future payments $E(X_{ij})$ is the *estimate outstanding loss reserves* for each accident year and the sum of all the projected values $\sum E(X_{ij})$ indicate **the total estimated reserve**.

5. ACCIDENT YEAR AND OVERALL STANDARD ERRORS

Calculating the variances or standard errors across accident years and in total requires one further step involving the covariances. The information needed is in the last matrix $\sigma^2 \cdot X_F \cdot (X^T X)^{-1} \cdot X_F^T$ together with the values calculated for ***$E(X_{ij})$*** and their variances. The variance of the sum of two values X_{ij} and X_{kl} is given by:

$$Var(X_{ij} + X_{kl}) = Var(X_{ij}) + Var(X_{kl}) + 2\operatorname{cov}(X_{ij}, X_{kl}) \qquad (15)$$

and this extends to sums of more than two values by including all pairs of covariances.

In the case of log-linear models the covariances can be calculated in the original space by the following convenient formula:

$$\operatorname{cov}(X_{ij}, X_{kl}) = E(X_{ij}) \cdot E(X_{kl}) \cdot \left(\exp(\operatorname{cov}(Y_{ij}, Y_{kl})) - 1\right) \quad (16)$$

This process can be applied to obtain the variance and then the standard errors for any combination of values, for instance, for each accident year or each payment year and more interestingly for the overall total reserve estimate.

So the variance of the total reserves estimates will include all possible combinations of covariances (of pairs) of values involved in the calculation.

The calculations are as in the previous example and can be tabulated easily to produce the following matrix of covariances.

Table no. 4 The matrix of covariances of estimate reserves

(i,j)	(1,3)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
(1,3)	–	$\mathrm{cov}(X_{1,3},X_{2,2})$	$\mathrm{cov}(X_{1,3},X_{2,3})$	$\mathrm{cov}(X_{1,3},X_{3,1})$	$\mathrm{cov}(X_{1,3},X_{3,2})$	$\mathrm{cov}(X_{1,3},X_{3,3})$
(2,2)	$\mathrm{cov}(X_{1,3},X_{2,2})$	–	$\mathrm{cov}(X_{2,2},X_{2,3})$	$\mathrm{cov}(X_{2,2},X_{3,1})$	$\mathrm{cov}(X_{2,2},X_{3,2})$	$\mathrm{cov}(X_{2,2},X_{3,3})$
(2,3)	$\mathrm{cov}(X_{1,3},X_{2,3})$	$\mathrm{cov}(X_{2,2},X_{2,3})$	–	$\mathrm{cov}(X_{2,3},X_{3,1})$	$\mathrm{cov}(X_{2,3},X_{3,2})$	$\mathrm{cov}(X_{2,3},X_{3,3})$
(3,1)	$\mathrm{cov}(X_{1,3},X_{3,1})$	$\mathrm{cov}(X_{2,2},X_{3,1})$	$\mathrm{cov}(X_{2,3},X_{3,1})$	–	$\mathrm{cov}(X_{3,1},X_{3,2})$	$\mathrm{cov}(X_{3,1},X_{3,3})$
(3,2)	$\mathrm{cov}(X_{1,3},X_{3,2})$	$\mathrm{cov}(X_{2,2},X_{3,2})$	$\mathrm{cov}(X_{2,3},X_{3,2})$	$\mathrm{cov}(X_{3,1},X_{3,2})$	–	$\mathrm{cov}(X_{3,2},X_{3,3})$
(3,3)	$\mathrm{cov}(X_{1,3},X_{3,3})$	$\mathrm{cov}(X_{2,2},X_{3,3}$	$\mathrm{cov}(X_{2,3},X_{3,3})$	$\mathrm{cov}(X_{3,1},X_{3,3})$	$\mathrm{cov}(X_{3,2},X_{3,3})$	–

Total = $\sum \mathrm{cov}(X_{ij}, X_{kl})$

The diagonal elements are left blank as the values here should be the variances which were estimated previously. The matrix is symmetric, as is to be expected, and so summing the range produces the sum of covariances of all possible pairs of values.

Thus the **overall variance** of the estimate outstanding loss reserves is the sum of two values: the sum of the variances of the predicting values from table 3 - $\sum Var(X_{ij})$ and the sum of all pairs of covariances from table 4 - $\sum \mathrm{cov}(X_{ij}, X_{kl})$

After we estimated the future payments for each accident year *i* we can compute **the confidence interval** for these future payments otherwise for estimated reserves:

$$X_{ij} \in [E(X_{ij}) \pm t_{\alpha} \cdot Se(X_{ij})] \quad (17)$$

6. PRACTICAL EXAMPLE

Consider the following example in order to illustrate the calculation of the Loss reserves using stochastical methods, named "Christofides method".

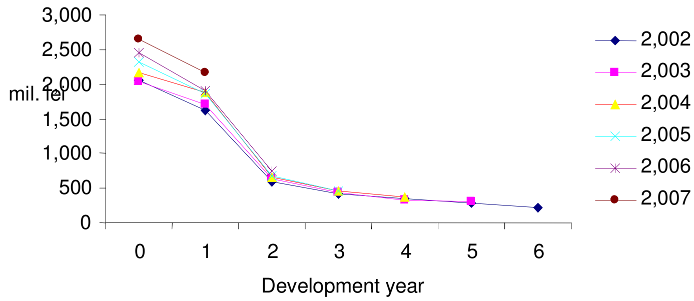
The example considers the incremental data of the Motor TPL insurance for an insurance company from Republic of Moldova. The information that is needed is simply the triangle of the paid loss (the data are in '000 MDL).

Incremental paid loss ('000 MDL)

Incremental

Year of origin (i)		Development year (j)						
		0	1	2	3	4	5	6
2002	**0**	2062	1629	583	421	341	276	228
2003	**1**	2031	1706	643	448	335	307	
2004	**2**	2164	1887	667	454	369		
2005	**3**	2320	1860	671	463			
2006	**4**	2462	1909	736				
2007	**5**	2651	2158					
2008	**6**	3084						

Figure no. 1 Evolution of the loss (incremental data)



In accordance with the format development in the table, the origin year (i) are represented by the rows, and the development years (j) by the columns. Origin is taken as accident year, and these years are listed down the left hand side from 0 to 6 (the current year). The development years from 0 to 6 a listed along the top of the table – year 0 being the accident year itself in each case.

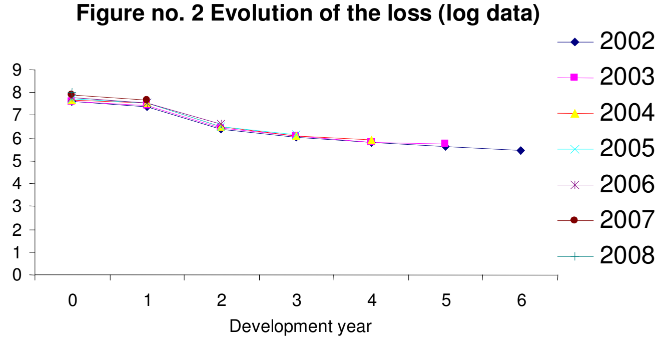
We are standing at the end of the origin year 6, and using the Christofides method for establishing the reserves at this date.

According the methods we shall to calculate the logarithm of the data from the previous table. Taking logarithms (natural logarithms will be assumed) gives:

Log incremental paid loss

Year of origin (i)		Development year (j)						
		0	1	2	3	4	5	6
2002	0	7.63	7.40	6.37	6.04	5.83	5.62	5.43
2003	1	7.62	7.44	6.47	6.10	5.81	5.73	
2004	2	7.68	7.54	6.50	6.12	5.91		
2005	3	7.75	7.53	6.51	6.14			
2006	4	7.81	7.55	6.60				
2007	5	7.88	7.68					
2008	6	8.03						

Figure no. 2 Evolution of the loss (log data)



The following table is in the form most convenient for the regression facility of any of the popular spreadsheet packages.

I	j	ln(Yij)	a0	a1	a2	a3	a4	a5	a6	b1	b2	b3	b4	b5	b6
0	0	7.63	1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	7.40	1	0	0	0	0	0	0	1	0	0	0	0	0
0	2	6.37	1	0	0	0	0	0	0	0	1	0	0	0	0
0	3	6.04	1	0	0	0	0	0	0	0	0	1	0	0	0
0	4	5.83	1	0	0	0	0	0	0	0	0	0	1	0	0
0	5	5.62	1	0	0	0	0	0	0	0	0	0	0	1	0
0	6	5.43	1	0	0	0	0	0	0	0	0	0	0	0	1
1	0	7.62	0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	7.44	0	1	0	0	0	0	0	1	0	0	0	0	0
1	2	6.47	0	1	0	0	0	0	0	0	1	0	0	0	0
1	3	6.10	0	1	0	0	0	0	0	0	0	1	0	0	0
1	4	5.81	0	1	0	0	0	0	0	0	0	0	1	0	0
1	5	5.73	0	1	0	0	0	0	0	0	0	0	0	1	0
2	0	7.68	0	0	1	0	0	0	0	0	0	0	0	0	0
2	1	7.54	0	0	1	0	0	0	0	1	0	0	0	0	0
2	2	6.50	0	0	1	0	0	0	0	0	1	0	0	0	0
2	3	6.12	0	0	1	0	0	0	0	0	0	1	0	0	0
2	4	5.91	0	0	1	0	0	0	0	0	0	0	1	0	0
3	0	7.75	0	0	0	1	0	0	0	0	0	0	0	0	0
3	1	7.53	0	0	0	1	0	0	0	1	0	0	0	0	0
3	2	6.51	0	0	0	1	0	0	0	0	1	0	0	0	0
3	3	6.14	0	0	0	1	0	0	0	0	0	1	0	0	0
4	0	7.81	0	0	0	0	1	0	0	0	0	0	0	0	0
4	1	7.55	0	0	0	0	1	0	0	1	0	0	0	0	0
4	2	6.60	0	0	0	0	1	0	0	0	1	0	0	0	0
5	0	7.88	0	0	0	0	0	1	0	0	0	0	0	0	0
5	1	7.68	0	0	0	0	0	1	0	1	0	0	0	0	0
6	0	8.03	0	0	0	0	0	0	1	0	0	0	0	0	0

Each row corresponds to a data value and its representation by the model parameters.

Within the class of log-linear models changing the model just involves changing the design matrix. The spreadsheet regression command, which requires a column for the dependent values and a range for the independent values (i.e. the design matrix) is then used to carry out the regression and output the result. It is necessary to specify that the fit is without a constant and to define results or output range. This is quite straightforward in practice and the results are produced almost instantly.

The spreadsheet output in this case will be:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9997
R Square	0.9993
Adjusted R Square	0.9321
Standard Error	0.0297
Observations	28

ANOVA

	df	*SS*	*MS*	*F*	*Sign. F*
Regression	13	19.51	1.5006	1699.68	0.0000
Residual	15	0.01	0.0009		
Total	28	19.52			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0			
a0	7.61	0.017	440.60	2.9268E-32
a1	7.65	0.017	443.31	2.6703E-32
a2	7.71	0.018	435.93	3.4348E-32
a3	7.73	0.018	417.77	6.5008E-32
a4	7.79	0.020	390.94	1.7598E-31
a5	7.88	0.023	347.32	1.0374E-30
a6	8.03	0.030	270.38	4.4364E-29
b1	-0.20	0.017	-11.93	4.6698E-09
b2	-1.21	0.018	-65.65	7.211E-20
b3	-1.57	0.020	-78.90	4.6128E-21
b4	-1.80	0.022	-81.52	2.827E-21
b5	-1.96	0.026	-75.78	8.4385E-21
b6	-2.18	0.034	-63.33	1.2348E-19

The variances for the future values in log-space are the sum of the variance-covariance matrix values obtained above and the model variance σ^2 (or s^2).

The following table shows the various values, especially the loss reserves and their variances and standard errors.

Calculating the variances or standard errors across accident years and in total requires one further step involving the covariances.

i	j	ai	bi	E(Yij)	Var[E(Yij)]	s^2	Var(Yij)	Var(Yij)/2	Xij	Var(Xij)	SE(Xij)	SE(Xij),%
1	2	3	4	5	6	7	8	9	10	11	12	13
1	6	7.65	-2.18	5.4760	0.0021	0.0009	0.0029	0.00147	239.2	168.7	13.0	5.4%
2	5	7.71	-1.96	5.7531	0.0016	0.0009	0.0025	0.00124	315.6	246.5	15.7	5.0%
2	6	7.71	-2.18	5.5322	0.0021	0.0009	0.0030	0.00149	253.1	191.6	13.8	5.5%
3	4	7.73	-1.80	5.9241	0.0015	0.0009	0.0024	0.00118	374.4	330.4	18.2	4.9%
3	5	7.73	-1.96	5.7720	0.0016	0.0009	0.0025	0.00127	321.6	262.1	16.2	5.0%
3	6	7.73	-2.18	5.5510	0.0022	0.0009	0.0030	0.00152	257.9	202.9	14.2	5.5%
4	3	7.79	-1.57	6.2196	0.0015	0.0009	0.0024	0.00118	503.1	596.6	24.4	4.9%
4	4	7.79	-1.80	5.9890	0.0016	0.0009	0.0024	0.00122	399.5	391.0	19.8	4.9%
4	5	7.79	-1.96	5.8369	0.0017	0.0009	0.0026	0.00131	343.2	309.3	17.6	5.1%
4	6	7.79	-2.18	5.6159	0.0023	0.0009	0.0031	0.00157	275.2	238.0	15.4	5.6%
5	2	7.88	-1.21	6.6744	0.0016	0.0009	0.0025	0.00124	792.8	1,555.9	39.4	5.0%
5	3	7.88	-1.57	6.3095	0.0016	0.0009	0.0025	0.00127	550.5	767.9	27.7	5.0%
5	4	7.88	-1.80	6.0790	0.0017	0.0009	0.0026	0.00131	437.1	501.9	22.4	5.1%
5	5	7.88	-1.96	5.9268	0.0019	0.0009	0.0028	0.00140	375.5	395.2	19.9	5.3%
5	6	7.88	-2.18	5.7059	0.0024	0.0009	0.0033	0.00166	301.1	301.0	17.4	5.8%
6	1	8.03	-0.20	7.8293	0.0021	0.0009	0.0029	0.00147	2,516.8	18,669.3	136.6	5.4%
6	2	8.03	-1.21	6.8262	0.0021	0.0009	0.0030	0.00149	923.1	2,548.9	50.5	5.5%
6	3	8.03	-1.57	6.4614	0.0022	0.0009	0.0030	0.00152	640.9	1,253.1	35.4	5.5%
6	4	8.03	-1.80	6.2308	0.0023	0.0009	0.0031	0.00157	508.9	814.1	28.5	5.6%
6	5	8.03	-1.96	6.0786	0.0024	0.0009	0.0033	0.00166	437.2	634.5	25.2	5.8%
6	6	8.03	-2.18	5.8577	0.0029	0.0009	0.0038	0.00191	350.6	471.6	21.7	6.2%
TOTAL									**11,117.3**	**30,850**	175.6	**1.6%**
										Var 1		

The calculations are as in the previous example and can be tabulated easily to produce the following matrix of covariances.

I	j	1	2	2	3	3	3	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6
		6	5	6	4	5	6	3	4	5	6	2	3	4	5	6	1	2	3	4	5	6
1	6		0	62	0	0	64	0	0	0	68	0	0	0	0	74	0	0	0	0	0	86
2	5	0		21	0	54	7	0	0	57	8	0	0	0	63	8	0	0	0	0	73	10
2	6	62	21		0	7	68	0	0	8	73	0	0	0	8	80	0	0	0	0	10	93
3	4	0	0	0		35	28	0	55	9	8	0	0	60	10	8	0	0	0	70	12	10
3	5	0	54	7	35		27	0	9	60	9	0	0	10	66	10	0	0	0	12	77	12
3	6	64	7	68	28	27		0	8	9	75	0	0	8	10	82	0	0	0	10	12	96
4	3	0	0	0	0	0	0		74	64	51	0	82	16	14	11	0	0	95	19	16	13
4	4	0	0	0	55	9	8	74		53	42	0	16	67	14	11	0	0	19	79	16	13
4	5	0	57	8	9	60	9	64	53		39	0	14	14	73	13	0	0	16	16	84	15
4	6	68	8	73	8	9	75	51	42	39		0	11	11	13	89	0	0	13	13	15	104
5	2	0	0	0	0	0	0	0	0	0	0		231	184	158	127	0	194	45	36	31	25
5	3	0	0	0	0	0	0	82	16	14	11	231		131	113	90	0	45	114	29	25	20
5	4	0	0	0	60	10	8	16	67	14	11	184	131		92	74	0	36	29	93	23	19
5	5	0	63	8	10	66	10	14	14	73	13	158	113	92		67	0	31	25	23	97	20
5	6	74	8	80	8	10	82	11	11	13	89	127	90	74	67		0	25	20	19	20	117
6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		2,394	1,662	1,320	1,134	909
6	2	0	0	0	0	0	0	0	0	0	0	194	45	36	31	25	2,394		618	491	422	338
6	3	0	0	0	0	0	0	95	19	16	13	45	114	29	25	20	1,662	618		346	297	238
6	4	0	0	0	70	12	10	19	79	16	13	36	29	93	23	19	1,320	491	346		240	192
6	5	0	73	10	12	77	12	16	16	84	15	31	25	23	97	20	1,134	422	297	240		170
6	6	86	10	93	10	12	96	13	13	15	104	25	20	19	20	117	909	338	238	192	170	
																		Total Var2 =			**32,884**	

Note that the diagonal elements are left blank as the values here should be the variances which were estimated previously. The matrix is symmetric, as is to be expected, and so summing the range produces the sum of covariances of all possible pairs of values. This sum of all pairs of covariances is 32 884.

The sum of the variances of the projected values obtained earlier was 30 850 and so the overall variance, which is the sum of these two values, is 63734. The overall standard error, which is the square root of this value, is therefore estimated as 252 or just 2,3% of the overall reserve estimate of **11 117 000 MDL.**

The table below summarizes the results.

Year of origin (i)		**Development year (j)**						
		0	**1**	**2**	**3**	**4**	**5**	**6**
2002	**0**							
2003	**1**							239
2004	**2**						316	253
2005	**3**					374	322	258
2006	**4**				503	400	343	275
2007	**5**			793	550	437	375	301
2008	**6**		2517	923	641	509	437	351
						Oustanding Loss		**11,117**
						Stahdard Error		**252**
						Var, %		**2.3%**
						Loss Reserves probability, 95%		**11,533**

For comparison, in the table bellow show the chain ladder overall loss reserves estimation.

Cumulative

Year of origin (i)		Development year 0	1	2	3	4	5	6
2002	0	2062	3691	4274	4695	5036	5312	5540
2003	1	2031	3737	4380	4828	5163	5470	
2004	2	2164	4051	4718	5172	5541		
2005	3	2320	4180	4851	5314			
2006	4	2462	4371	5107				
2007	5	2651	4809					
2008	6	3084						
TOTAL j		16,774	24,839	23,330	20,009	15,740	10,782	5,540
TOTAL j-1			13,690	20,030	18,223	14,695	10,199	5,312
Factors			1.81	1.16	1.10	1.07	1.06	1.04
Cumul. Factors			2.74	1.51	1.30	1.18	1.10	1.04
Proportion of paid			55%	86%	91%	93%	95%	96%

Year of origin	Paid to date	Cumulative factors	Ultimate Loss	Loss Reserves
2003	5,470	1.04	5,705	235
2004	5,541	1.10	6,109	568
2005	5,314	1.18	6,276	962
2006	5,107	1.30	6,622	1,515
2007	4,809	1.51	7,263	2,454
2008	3,084	2.74	8,451	5,367
TOTAL				11,101

Thus, the chain ladder overall estimate was 11 101 000 MDL.

7. CONCLUSION

The individual values obtained by the two methods are also close but the advantages of a stochastic model is that the basic chain ladder estimates are point estimates whereas the regression based estimates are statistical estimates with both a mean and a standard error estimate.

All the usual information that can be produced from the traditional chain ladder can be derived from the regression chain ladder including estimates of development factors.

The stochastic approach as shown above can produce additional information, based on the model assumptions, such as standard errors of parameters and reserve estimates, that the traditional approach does not. The statistical estimates obtained by the regression approach also facilitate stability comparisons across companies and classes.

This completes our consideration of the regression chain ladder. The technique does not require that we have a complete triangle of data and can work with almost any shape data as long as there are sufficient points from which to obtain estimates of the parameters.

References

[1] Christofides S (1997, September), *Regression models based on log-incremental payments,* Claims Reserving Manual, v 2.
[2] International Financial Reporting Standard 4.(2004), Insurance Contracts. International Accounting Standards Committee Foundation (IASCF).
[3] Late claims reserves in reinsurance (2000), Swiss Reinsurance Company, Zurich
[4] Micus Cathy (2004, September), *The importance of precisely assessing IBNR reserves*. Gecalux Group, Quarterly Newsletter, Issue 6.
[5] Teylor Greg. (2003, September), *Loss Reserving: Past, Present and Future*. University of Melbourne, Research paper number 109, Australia.
[6] Verejan O., Pârțachi I. (2004), *Statistica actuarială în asigurări.* Editura Economica, Bucureşti.
[7] Verejan, O., Pârțachi, I. (2004), *Statistica actuarială în asigurări comerciale*, Editura ASEM, Chişinău.