1995 General
Insurance Convention

The Bootstrap Method

and some

Reserving applications

"Good simple ideas,...., are our most precious intellectual commodity, so there is no need to apologize for the easy mathematical level."
Bradley Efron

Presented at GISG 1995 by Christian Larsen

The Bootstrap method was introduced by Bradley Efron, Stanford University in 1979. Efron (1982) begins with some general statements, such as:

"Good simple ideas,...., are our most precious intellectual commodity, so there is no need to apologize for the easy mathematical level."

and

"An important theme of what follows, is the substitution of computational power for theoretical analysis. The payoff, off course, is the freedom from the constraints of traditional parametric theory, with its overreliance on a small set of standard models for which theoretical solutions are available. In the long run, ...(bootstrap) ..., should make clearer the virtues of parametric theory,..."

and

"From a traditional point of view, ... the methods ... are prodigious computational spendthrifts. We blithely ask the reader to consider techniques which require the usual statistical calculations to be multiplied a thousand times over. None of this would have been feasible twenty-five years ago, before the era of cheap and fast computation."

The situation

We consider a random sample $(X_1,...,X_N)$ of random size N. The distribution G of N is assumed to be known and the random variables are assumed to be independent and equally distributed with unknown distribution F:

$$(X_1,...,X_N) = (x_1,.....,x_n), \quad X_r \sim F, \; r=1,...,N \text{ and } N \sim G$$

It is further assumed that X_r is independent of the frequency N.

$X=(X_1,...,X_N)$ denotes the random sample and $x=(x_1,.....,x_n)$ the observed realisation.

The problem

Let $\mathbf{R}(X)$ be a function of X. Then $\mathbf{R}(X)$ is a stochastic variable with a distribution that is dependent on G and of the unknown F. The problem is to estimate the distribution of $\mathbf{R}$ on the basis of the observation x.

A solution - the Bootstrap Method

If the distribution of F was known then the distribution of **R** could, *in theory*, be calculated exactly or in practice be approximated with unlimited accuracy by the Monte Carlo method. The simple idea in Bootstrap is to substitute the distribution of F with the empirical distribution based on the observation x:

A. Construct the sample probability distribution E (i.e. the empirical distribution of X), putting mass 1/n at each point $x_1,..,x_n$.

B. Consider the random sample $Y=(Y_1,...,Y_N)$, $Y_r \sim E$, $r=1,...,N$, $N \sim G$

C. Approximate the distribution of **R**(X) by the (Bootstrap) distribution of **R**(Y)

The essential facts are that

> the distribution of **R**(Y) is only dependent on E and G and if E is a good approximation to F then one can expect that the distribution of **R**(Y) will be a good approximation to the distribution of **R**(X) and
>
> the distribution of (E,G) is known and therefore, the distribution of **R**(Y) can be calculated.

Application 1: Calculation of the uncertainty of reserving estimates

Let N denote the number of observed claims in a specific period and assume that N is Poisson distributed with a known mean. Let X_r, $r=1,...,N$ denote the information linked to the claims including information about accident period, development period and payments. Only information regarding the past is available and for some claims X_r, all payments and periods of payments are not necessarily included.

Let **RES** denote the stochastic total of the future payments, i.e. the stochastic reserve.

Let $\mathbf{R}_1$(X) denote a reserve estimator, e.g. the Chain Ladder estimator based on the claim information. $\mathbf{R}_1$(X) is a function of the claims X and therefore a stochastic variable.

The problem we wish to solve is to estimate the distribution of the Chain Ladder estimator $\mathbf{R}_1$(X) rather than just the Chain Ladder point estimate $\mathbf{R}_1$(x).

Example 1

Consider the accident periods 1985-1994 and assume that the number of claims with accident date in period i=1985,...,1994 and notification delay j=0,...,9 measured by years from accident period to notification period is Poisson($A_i\ p_j$) with the parameters outlined below in figure 1.

Assume further that each claim has only one payment which is Gamma distributed with mean $B_i=(1.05)^{(i-1985)}$ dependent on the accident period i and constant coefficient of variation equal to 2. The waiting time k =0,...,9 from year of notification to year of payment is assumed to be independent of the waiting time to notification and with distribution outlined in figure 1 below. The distribution of the waiting time r from accident year to year of payment, i.e. the convolution of p and q, is also calculated.

Period	A_i	B_i
1985	500	1.000
1986	500	1.050
1987	500	1.103
1988	500	1.158
1989	500	1.216
1990	500	1.276
1991	500	1.340
1992	500	1.407
1993	500	1.477
1994	500	1.551
All	5000	12.578

Delay	$j \sim$ p	$k \sim$ q	$j+k \sim$ r
0	0.30	0.35	0.1050
1	0.25	0.20	0.1475
2	0.20	0.20	0.1800
3	0.15	0.15	0.1875
4	0.07	0.10	0.1620
5	0.03	0	0.1095
6	0	0	0.0625
7	0	0	0.0315
8	0	0	0.0115
9	0	0	0.0030
Mean	$E(j)$=1.53	$E(k)$=1.45	$E(j+k)$=2.98

figure 1

On the basis of these distribution assumptions we get E(**RES**) = 2142 and the expected total claim amount is 500*12.578 = 6289. As a consequence, the expected amount paid already is therefore 4147.

Two data sets, PAST and FUTURE, are generated by simulation in The SAS® System on the basis of the distribution assumptions above. The PAST file consists of 4699 claims where $i+j<1995$ and of 3486 payments where $i+j+k<1995$. The Future file consists of the remaining data, i.e. claims with date of notification or date of payment in 1995 or later. The sum of future payments, i.e. the observed value of **RES**, is 2004, compared to the expected value 2142.

As an example, the claim information regarding three claims from the PAST file is shown below:

Observation	Claim identification	Accident period i	Transaction	Delay j or $j+k$	Payment
x_{108}	108	1985	Notification	0	
	108	1985	Payment	3	2251.89
x_{3086}	3086	1991	Notification	1	
	3086	1991	Payment	2	1466.06
x_{4649}	4649	1994	Notification	0	

figure 2

The PAST data is now triangulated and the Chain Ladder estimate $\mathbf{R}_J(x)$ for the reserve is calculated. The results are outlined in figure 3 below.

ACC. PERIOD	Development Period 1	2	3	4	5	6	7	8	9	10	Observed Payments	Total Reserve	Total Estimated Payment
1985	50	81	113	97	87	55	36	4	2	1	525	0	525
1986	59	64	103	85	95	51	43	11	5	1	516	1	516
1987	52	85	132	82	78	54	38	12	4	1	534	4	539
1988	77	66	94	79	127	44	27	9	3	1	515	14	528
1989	81	111	123	115	87	84	45	12	4	1	602	62	665
1990	79	116	84	96	118	64	42	11	4	1	494	122	616
1991	79	67	112	132	109	65	42	11	4	1	389	231	621
1992	74	121	138	125	128	76	50	13	5	1	333	397	730
1993	78	123	143	129	132	78	51	13	5	1	200	551	752
1994	112	148	185	167	171	101	67	17	6	2	112	864	976
ALL	740	983	1226	1107	1132	672	442	113	42	10	4220	2247	6467

figure 3

The estimated run off and the model run off are shown in figure 4. For example it is seen that on average 11.4% of the claim amount is estimated to be paid in the accident year, and that the model proportion r is 10.5%.

Run-off pattern

Est.	11.4%	15.2%	19.0%	17.1%	17.5%	10.4%	6.8%	1.7%	0.6%	0.2%	65.3%	34.7%	100%
Model	10.5%	14.8%	18.0%	18.8%	16.2%	11.0%	6.3%	3.2%	1.2%	0.3%	65.9%	34.1%	100%

figure 4

Figure 5 shows the observed and the estimated proportion of the estimated ultimate claim costs paid at different times. It is seen that at the end of 31 December 1994 the amount 4220 (65.3%) was paid and that the amount expected to be paid in the future, i.e. the outstanding claims reserve is 2247 (34.7%) .

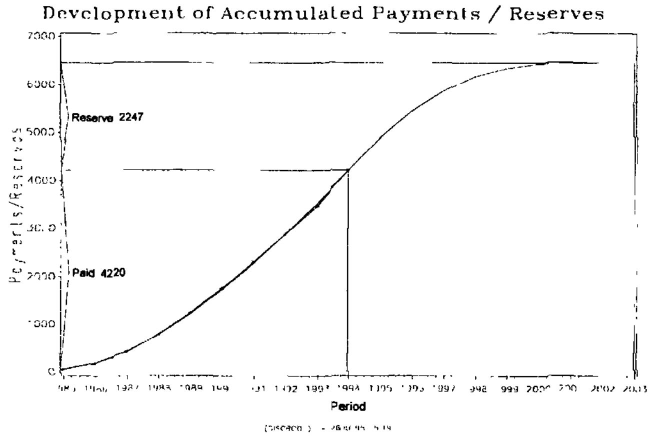

figure 5

In order to calculate the estimation uncertainty of the Chain Ladder estimate 2247, the Bootstrap method has been applied using Larsen & Partners' Actuarial Claims Reserving System. The estimated distribution of **R**(Y) (figure 6) is based on 200 Monte Carlo simulations from the empirical distribution. The procedure is:

1. Draw a random number m from G = Poisson(4699) .
2. Draw a random sample, $y_1,....,y_m$ from E
3. Calculate the Chain Ladder reserve $\mathbf{R}_j(y_1,...,y_m)$ based on $y_1,....,y_m$
4. Repeat 1-3 200 times.

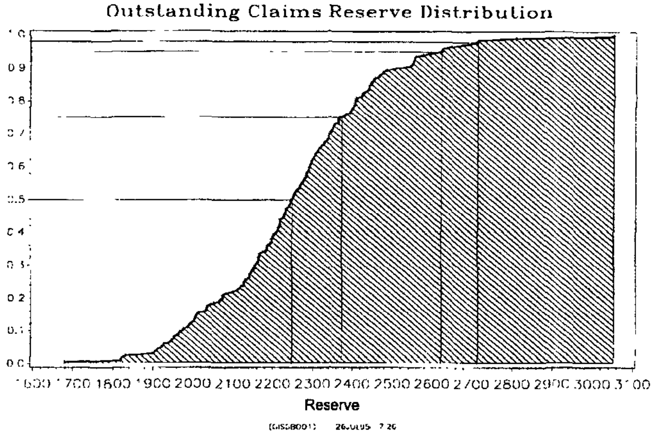

figure 6

Normally we do not know the distribution of F and a test of the quality of the method is difficult to define. However, in the situation above, where the distribution of F is known, we can easily calculate the distribution of **R**(X) by simulation:

1. Draw a random number m from G = Poisson(4699).
2. Draw a random sample, $x_1,....,x_m$ from F
3. Calculate the Chain Ladder reserve $\mathbf{R}_l(x_1,...,x_m)$ based on $x_1,....,x_m$
4. Repeat 1-3 200 times. The results are shown in figure 7.

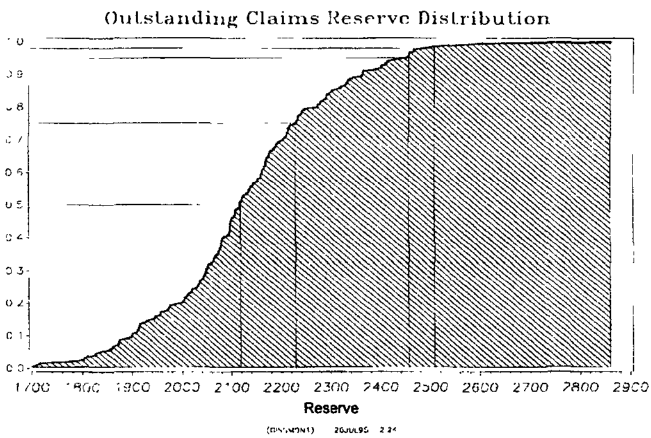

figure 7

The reserve estimate $\mathbf{R}_I(x)$ based on the observation x is exceeding the expected value E(**RES**) by approximately 105 and this error is inherited in the entire distribution of $\mathbf{R}_I(Y)$. However, the Bootstrap error distribution has no systematic error and it approximates to the actual error distribution reasonably well, see figure 8 below.

	$\mathbf{R}_I(Y)$	$\mathbf{R}_I(X)$	$\mathbf{R}_I(Y)-E(\mathbf{R}_I(Y))$	$\mathbf{R}_I(X)-E(\mathbf{R}_I(X))$
Mean	2253	2130	0	0
STD	214	179	214	179
Median	2247	2118	-6	-12
75% fraction	2373	2229	120	99
95% fraction	2622	2456	369	326
98% fraction	2713	2508	460	378

figure 8

Please note, that the distribution of $\mathbf{R}_I(Y)$ and of the error $\mathbf{R}_I(Y)-E(\mathbf{R}_I(Y))$ is calculated without using the knowledge of the underlying distribution *F*.

Application 2: Estimation of the total uncertainty

We consider a reserving method **R**. The total uncertainty consists of the estimation uncertainty (related to the past) plus the uncertainty related to the future payments **RES**-E(**RES**). We assume that the selected model is 'correct' so that the reserve estimate is unbiased, i.e. E(**R**(X)) = E(**RES**).

In order to estimate reserving margins we would estimate the distribution of the stochastic variable **R** + (**RES**-E(**RES**)).

Example 2

Again we consider the distribution outlined in example 1 and the reserving method $\mathbf{R}_I$ defined above. Since the claims are independent it follows that **R**(X) and (**RES**-E(**RES**)) are independent and we therefore only have to calculate the convolution of the distributions of $\mathbf{R}_I(X)$ and (**RES**-E(**RES**)), for example by simulation.

The distribution of $\mathbf{R}_I(X)$ is approximated by the Bootstrap distribution $\mathbf{R}_I(Y)$. The distribution of **RES**-E(**RES**) could have been estimated by simulation on the basis of the estimated parameters and the model assumptions. However, the distribution has been simulated on the basis of the original parameters. The results are outlined in figure 9 below:

	$R_1(Y)$-Bootstrap	RES - E(RES) estimated	Total
Mean	2253	0	2253
STD	214	77	230
Median	2247	-4	2248
75% fraction	2373	39	2380
95% fraction	2622	120	2632
99% fraction	2713	166	2758

figure 9

It is seen, that the main contribution to the uncertainty of 'Total' rises from the estimation uncertainty. For example, the 95% fraction is only increased from 2622 to 2632 when the future randomness is included. It is often seen, as in this example, that focus should be on the randomness in the past rather than in the future when reserving margins are estimated.

Application 3: Selection of reserving method

We consider two different reserving methods $\mathbf{R}_1$ and $\mathbf{R}_2$. Assume that both methods are unbiased estimators i.e.

$$E(\mathbf{R}_1(X)) = E(\mathbf{R}_2(X)) = E(\mathbf{RES})$$

We would then prefer to use $\mathbf{R}_2$ rather than $\mathbf{R}_1$ if

$$\mathrm{Var}(\mathbf{R}_2(X)) < \mathrm{Var}(\mathbf{R}_1(X)).$$

The problem is that the distributions of $\mathbf{R}_1(X)$ and $\mathbf{R}_2(X)$ are unknown. However, using Bootstrap approximation, we can easily estimate the variances of $\mathbf{R}_1(Y)$ and $\mathbf{R}_2(Y)$ and base the selection on these.

Example 3

We consider the distribution outlined in example 1 and two different reserving methods:

$\mathbf{R}_1$: Deterministic Chain Ladder method based on the payments (as above)

$\mathbf{R}_2$: A stochastic model with unknown but constant claim inflation.

Both methods are reasonably central/unbiased, $E(\mathbf{R}_1(X))= 2130$, $E(\mathbf{R}_2(X))=2145$ and $E(\mathbf{RES})=2143$.

It is seen from figure 10 that the stochastic model $\mathbf{R}_2$ gives a more precise reserve estimate but further more, the standard deviation of $\mathbf{R}_2$ is less than that of the Chain Ladder method $\mathbf{R}_1$:

	$\mathbf{R}_1$(Y)- Bootstrap	$\mathbf{R}_1$(X) - Real dist.	$\mathbf{R}_2$(Y) - Bootstrap	$\mathbf{R}_2$(X) - Real dist.
Mean	2253	2130	2134	2145
STD	214	179	177	153
Median	2247	2118	2137	2144
75% fraction	2373	2229	2257	2244
95% fraction	2622	2456	2414	2418
98% fraction	2713	2508	2505	2487

figure 10

Conclusion

It is concluded that if the individual claim data are available the Bootstrap is an effective method to calculate approximations to the uncertainty distributions of claims reserves. Only distribution assumptions regarding the claim frequency are required. The application is not dependent on the reserving method and even for complicated reserving methods the uncertainty can be estimated. It can be used to select between different reserving methods to obtain the most robust method and to calculate reserving margins.

References

Efron, B. (1979), Bootstrap methods: Another look at the jacknife. *Ann. Stat*, 7.p. 1-26.

Efron B. (1982) The Jacknife, The Bootstrap and Other Resampling Plans. *Society for Industrial and Applied Mathematics*., Regional Conference series in Applied Mathematics.