Rapp

Tariffanalys with weak assumptions using Rapp

Lecture in Indonesia november 2014
Stig Rosenlund

The methods of this talk are implemented in the programming language Rapp. To find Rapp on the internet, search **free actuarial language.**

About me: PhD in Mathematical Statistics at the University of Göteborg in 1983. Actuary at the Swedish insurance company Länsförsäkringar 1983-2010. After retirement in 2010, I have resumed my research, now focusing on insurance mathematics. So far I have in the 2010's had three papers published - two in tariff analysis and one in reserving.

Please interrupt me with questions at any time!

You might be familiar with tariff analysis - pricing with GLM in a multiplicative model. The literature is large, but here I mention just the book by Ohlsson and Johansson - my former colleagues at Länsförsäkringar - and two papers by me. These are enough to follow my reasoning.

Ohlsson, E. & Johansson, B. (2010).
Non-Life Insurance Pricing with Generalized Linear Models.
174 pages + preface. Springer, Berlin.

Rosenlund, S. (2010).
Dispersion estimates for Poisson and Tweedie models.
ASTIN Bulletin **40**(1), 271-279.

Rosenlund, S. (2014).

Inference in multiplicative pricing.

Scandinavian Actuarial Journal **2014**(8), 690-713.

My papers are included in the manual as appendices.

The setup is that we have a number of arguments, each with a finite number of classes. For example car age, driver age, car make, geographical region. The risk premium or "pure premium" is the expected claim cost for a policy during one year.

I won't dwell much on general GLM theory. Ohlsson & Johansson give a good account.

The model is that the risk premium is expressible as a product of a base factor and a factor per argument determined by the class the insured object belongs to. In GLM you logarithm the factors to get an additive model. Depending on the assumptions for the distributionss of claim number and claim severity - the cost of one claim - you get some different solutions of the ML (maximum likelihood) equations, applied to a sample of claim statistics for eg five years.

Short description of models and methods for factor point estimates

Standard GLM

Claim numbers are Poisson distributed. Claim severities are Γ distributed with a constant CV (coefficient of variation). A fairly common name for the ML solution is the Standard GLM method. Popular with actuaries.

Tweedie

The whole claim cost, without separation of claim number and severity, follows a Tweedie model, depending on a parameter p between 1 and 2. I call its solution the Tweedie method.

MMT

The claim cost is Poisson distributed. This is the special Tweedie case p = 1. Following the terminology of Ohlsson & Johansson I call it the MMT (Method of Marginal Totals) method.

For MMT, you can also solve a system of equations defined by prescribing that the sum of multiplicatively computed estimated claim costs over any argument class - any marginal - be equal to the empirical claim cost of the argument class. This is the classical solution. It will be the same as the GLM Tweedie one. There are thus two different ways of numerical solving. The GLM-based one is solved by the Newton-Raphson method. The classical one is solved by an iterative method. Normally the Newton-Raphson method is best by being faster, but in some cases the classical one must be used.

All three methods are implemented in Rapp, with both point estimates and confidence intervals. For Standard GLM and Tweedie the intervals are by the Ohlsson & Johansson book. For MMT by my own method.

The assumption of Poisson distributed claim cost is clearly not realistic. So you might conclude that MMT should not be used. Nevertheless, I show in Rosenlund (2014) that it is still preferable in many cases - in fact most often - to Standard GLM. This is due to robustness considerations.

Very short summary of the findings of my two papers

<u>Rosenlund (2010)</u>

You can discard the concept of "Overdispersed Poisson" as something more than a mixed Poisson or Compound Poisson distribution. All Overdispersed Poisson processes are Compound Poisson.

<u>Rosenlund (2014)</u>

Point estimates

Contains a simulation study of mean square errors (MSEs) of point estimates by Standard GLM and MMT. Standard GLM gives mostly the smallest variances of point estimates. It is best when the multiplicative hypothesis is true. But when we assume more and more deviations from the hypothesis, the performance of MMT in relation to Standard GLM increases more and more.

With moderate deviations from parameter multiplicativity, MMT is typically better in the MSE sense when there are many arguments or many claims, ie for mass consumer insurance.

The reason for this is that MMT has normally smaller bias. It has zero bias on the marginals by virtue of the marginal totals equations. This spills over also on the tariff combinations. Since

$$\text{MSE} = \text{bias}^2 + \text{variance}$$

we get smaller MSE for MMT for sufficiently many claims. When the number of claims increases to infinity the variances decrease to zero, but the biases remain approximately the same.

Variance estimates

For MMT I develop a method called

MVW = MMT Variance estimates under Weak assumptions

for confidence intervals, using only the Compound Poisson model. GLM is used for claim frequency, but not for mean claim. Start with the raw method of taking the simple variance estimates you obtain from the Compound Poisson model applied to univariate data. That is, as if the tariff analysis was made with just one argument. Adjust the estimates using GLM in an adhoc way. Usually the differences between the raw method's variance estimates and the adjusted estimates are very small. The advantage of MVW over the S-GLM variance estimates is that it requires much fewer assumptions, and hence is more realistic.

For simulated and fairly realistic cases, these confidence intervals are compared with the ones of Standard GLM, that follow from its model, and the Tweedie method for risk premiums. MVW confidence intervals are found to be mostly the best. The study reports both cover probabilities, which should be close to 0.95 for 95 % confidence intervals, and interval lengths, which should be small.

Overall conclusion

With more than a few arguments and a few claims MMT should be used with confidence intervals by MVW. The Γ form of the claim severity distribution is not important for the Standard GLM confidence intervals to be valid, only the assumption of constant CV.

The Tweedie method should not be used at all.

I can give a very simplified rule for point estimates, under realistic departures from multiplicativity.

How many claims are "Few", "Moderate" and "Many" depends on claim severity sizes. If large claims contribute much to the total claim cost, it has the same effect as few claims.

Total number of claims	Best method: Number of arguments: Few	Moderate	Many
Few	Standard GLM	?	MMT
Moderate	?	MMT	MMT
Many	MMT	MMT	MMT

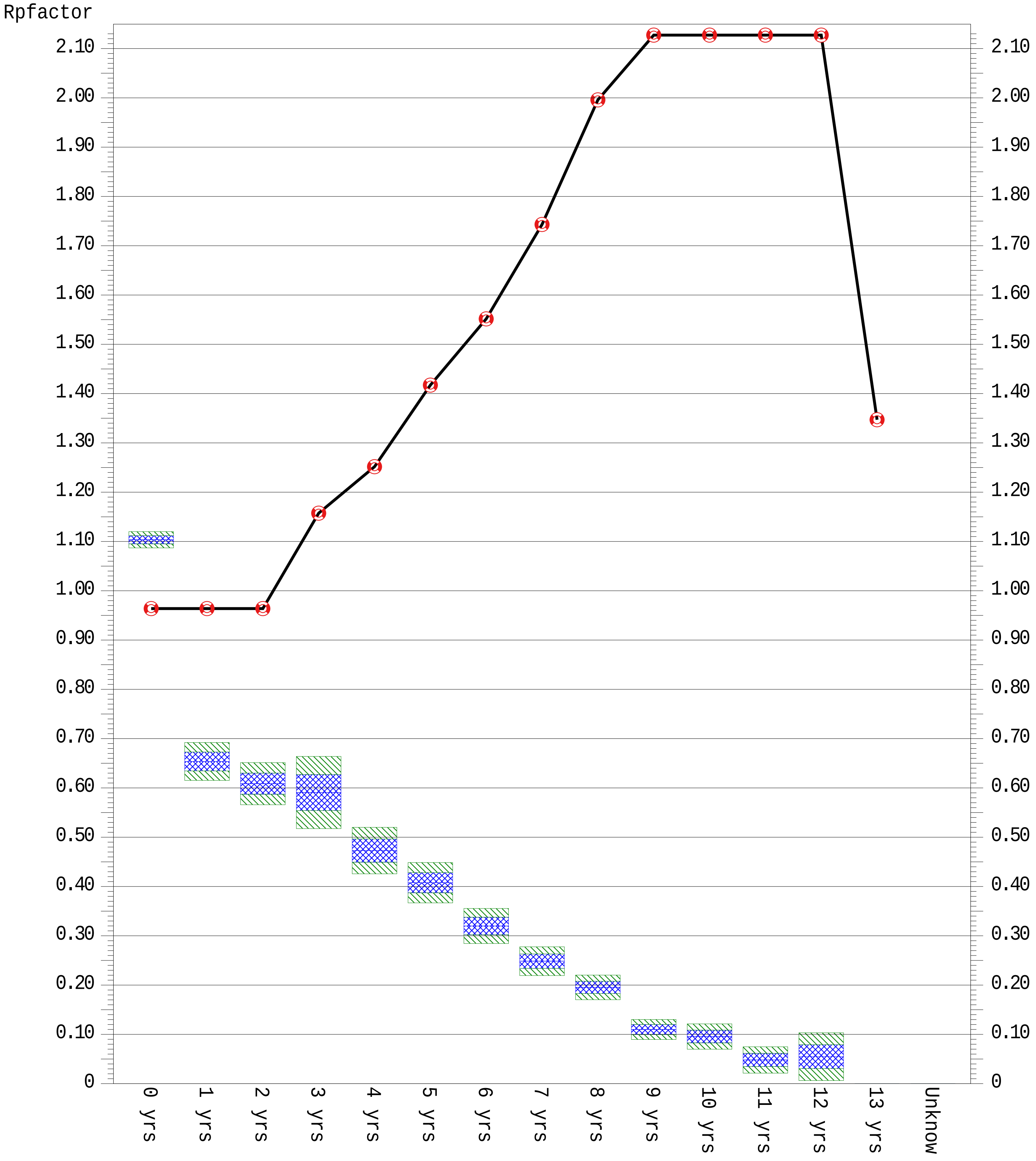
A demo run of Taran-demo.Rpp follows, with all kinds of graphs. It might take place in a separate session.

The input data are from a real assignment at Länsförsäkringar. There are about 2 million insurance records with computed exposure and about 120 000 claims.

The last two pictures here are graphs over risk premium factors, tariff factors, and univariate risk premiums. The tariff was set before the actuary (me) was consulted - it reflects the univariate factor ladder.

Tariff- and risk premium factors + confidence intervals

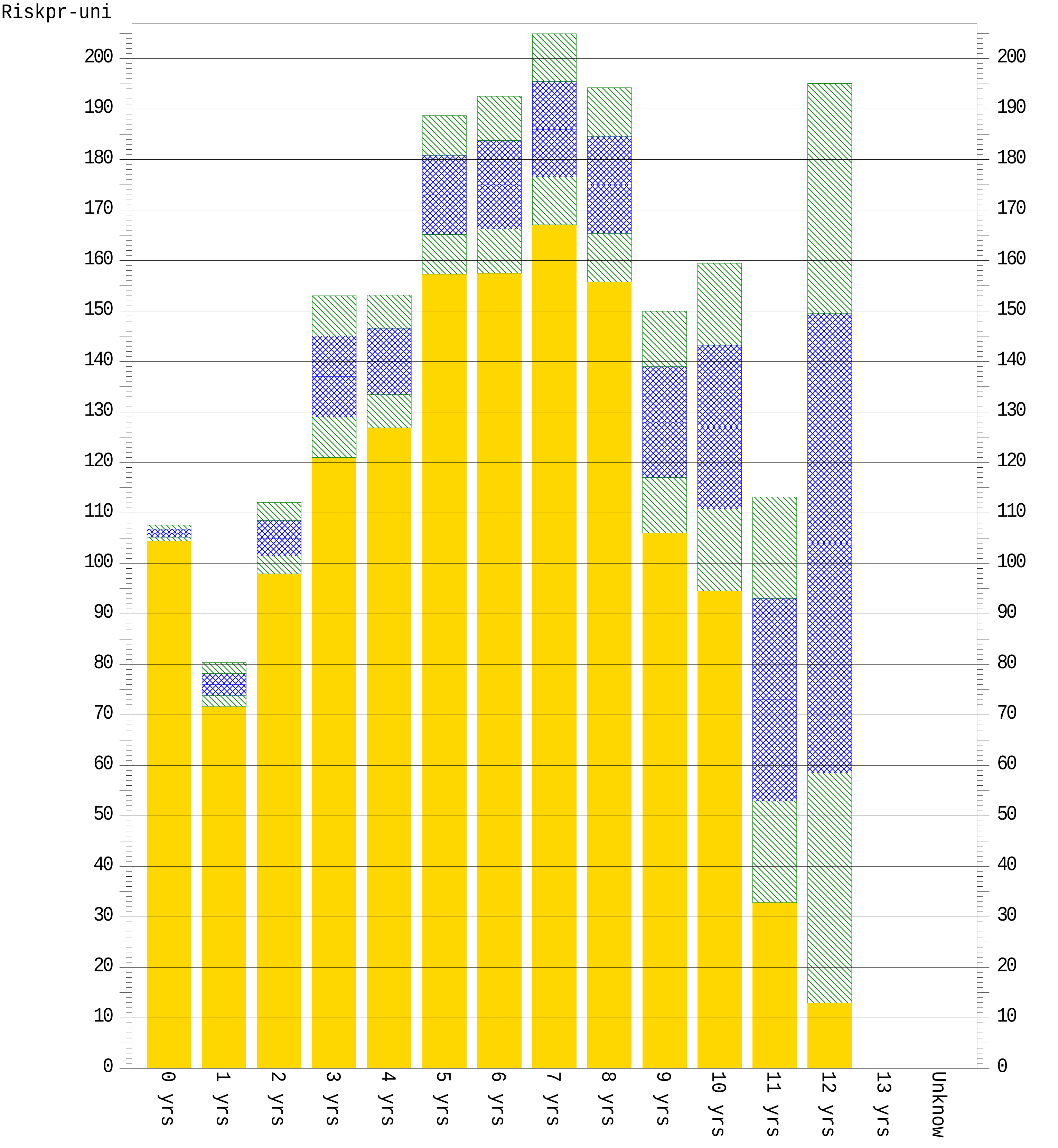
Black=tariff, single lines=90% conf, cross lines=60% conf



Class:	0 yrs	1 yrs	2 yrs	3 yrs	4 yrs	5 yrs	6 yrs	7 yrs	8 yrs	9 yrs	10 yrs	11 yrs	12 yrs	13 yrs	Unknow
Tarf :	0.964	0.964	0.964	1.157	1.252	1.417	1.552	1.743	1.996	2.127	2.127	2.127	2.127	1.347	0.000
Lo90%:	1.087	0.615	0.566	0.517	0.426	0.367	0.284	0.219	0.170	0.090	0.070	0.021	0.006	0.000	0.000
Point:	1.103	0.654	0.609	0.591	0.473	0.408	0.320	0.248	0.195	0.110	0.096	0.048	0.055	0.000	0.000
Up90%:	1.120	0.692	0.651	0.664	0.520	0.449	0.355	0.278	0.220	0.130	0.121	0.075	0.103	0.000	0.000

Risk premium univariate with confidence intervals

Single lines=90% confidence, cross lines=60% confidence



Class:	0 yrs	1 yrs	2 yrs	3 yrs	4 yrs	5 yrs	6 yrs	7 yrs	8 yrs	9 yrs	10 yrs	11 yrs	12 yrs	13 yrs	Unknow
Lo90%:	104.409	71.643	97.929	120.973	126.861	157.306	157.489	167.080	155.769	106.054	94.561	32.839	12.940	0.000	0.000
Point:	106.000	76.000	105.000	137.000	140.000	173.000	175.000	186.000	175.000	128.000	127.000	73.000	104.000	0.000	0.000
Up90%:	107.591	80.357	112.071	153.027	153.139	188.694	192.511	204.920	194.231	149.946	159.439	113.161	195.060	0.000	0.000