A NOTE ON USING INFLATION-TRUNCATED DATA

Rodney Kreps

A NOTE ON USING INFLATION-TRUNCATED DATA

Rodney Kreps

Abstract:

When losses are reported excess of a fixed amount, the effect of inflation on the trended values is to eliminate information from the lower end of the data for the older years. Consequently, the corresponding low end of the recent years is not used in analyses. A simple maximum likelihood solution is proposed which uses all the data. The price paid is that the frequency and severity distribution analyses are then intertwined.

Introduction:

In pricing any insurance or reinsurance contract, it is always necessary to restate past loss data to current or future conditions. In doing this, the four elements are changes in exposure, development on known claims, IBNR claims, and trending for inflation. This note considers only the latter. When all claims are known from ground up, inflation is frequently represented by applying a common index to all claims from a given accident year; or, rarely, by different indices for different sizes of loss.

For certain contracts there is another complication induced by inflation. Loss data in reinsurance and excess pricing is frequently only reported when the loss amount is excess of some value, for example half of the attachment point. Inflation makes losses in the older years economically equivalent to larger losses in the more recent years. For example, with a reporting level of \$50,000, a \$40,000 1985 loss will not be reported, whereas the same physical loss in 1990 may cost \$60,000 and will be reported. With a constant reporting value, the net effect is that the on-level data is truncated from below by an increasing amount as one goes backward from the most recent year. In order to regard each year's data as a sample from the same population for statistical purposes, one must use economically equivalent data across the years. This implies that the lower values of more recent data are not used, thus losing information.

The solution using all data is approached starting from the most intuitive case of Poisson frequency and multinomial severity. There, the explicit maximum likelihood equations are given and solved. Next, the negative binomial is considered. Although its maximum likelihood equations can be written down, numerical solution of the minimization of the negative log-likelihood seems the way to go. From there, a heuristic argument leads to the form of the negative log-likelihood for a continuous severity distribution and either frequency distribution. A consequence of the form is that frequency and severity cannot be determined independently.

The simplest version: Poisson-multinomial

The typical problem is to estimate for a prospective year the frequency λ of events and the severity distribution, having exposure information and past losses reported excess of a fixed amount. The losses are brought to ultimate, including IBNR losses, and indexed to the year of interest. This is, or course, the actuarially problematical part.

For simplicity's sake, it is first assumed that a number of loss ranges ("bins") are defined, e.g. \$1001 to \$2000, \$2001 to \$5000, etc. The data is the number of events in each bin, by year. The information desired is the overall frequency of loss and the probability of a loss falling into each bin. This brings up a situation such as is pictured below:

probabilities	dollar bins	COUNTS			
p_5	5	n_{51}	n_{52}	n_{53}	N_5
p_4	4	n_{41}	n_{42}	n_{43}	N_4
p_3	3	n_{31}	n_{32}	n_{33}	N_3
p_2	2	**X**	n_{22}	n_{23}	N_2
p_1	1	**X**	**X**	n_{13}	N_1
	year:	1	2	3	
	exposure:	ε_1	ε_2	ε_3	
	Poisson parameter:	$\lambda\varepsilon_1$	$\lambda\varepsilon_2$	$\lambda\varepsilon_3$	

The dollar bins run vertically upward and the years run horizontally to the right. The n_{ik} are the number of event counts in each bin, by year. The underlying probability for an event to be in bin "i" is p_i and the total number of seen events in bin "i" is N_i. The exposure index relative to the year of interest for year "k" is ε_k. The process is taken to be Poisson, with parameters given by the product of the exposure index and the Poisson parameter of the year of interest. The problem is to estimate both lambda and the probabilities for each bin.

The complicating feature is the missing data (indicated by **X**) in bins 1 and 2 for year 1, and in bin 1 for year 2. Usually, in order to compare economically equivalent data it is necessary to disregard the lower two bins for all years. This has two unfortunate consequences: First, the lower end of the available data may be higher than we require for the problem at hand. Alternatively, in order to get data low enough, we may be limited in the number of past years that we could otherwise use. Second, we ignore perfectly good data (as much as any reinsurance data is perfectly good) which could add information. A *caveat* is appropriate here - the IBNR and development is more uncertain in the recent years, and this may temper one's desire to use the data. The other side of the coin is that the older years' data may also be suspect because of changes in the business mix and possible inappropriateness of the inflation indices.

Happily, there is a maximum likelihood solution to the problem of using all the data. In order to provide it, begin by considering only year 2 (and drop the corresponding subscript to save typography). Given p_1 to p_5, the probability of observing n_1 to n_5 is the multinomial formula

$$M(n_1,..,n_5) = \Gamma(N+1) \prod_{i=1}^{5} \frac{(p_i)^{n_i}}{\Gamma(n_i+1)}, \quad N = \sum_{i=1}^{5} n_i \text{ and } \sum_{i=1}^{5} p_i = 1$$

The Poisson probability with parameter λ of observing N events is

$$P(N,\lambda) = \frac{\lambda^N e^{-\lambda}}{\Gamma(N+1)}$$

The key remark is that if the total is Poisson distributed with parameter λ, the probability of observing $n_2,..,n_5$ with no information on n_1 is the sum over the probabilities of observing no events in bin 1, one event, two events, etc.:

$$\text{prob} = \sum_{v=0}^{\infty} M(v,n_2,..,n_5)P(v+n_2+..+n_5,\lambda)$$

$$= \prod_{i=2}^{5} \frac{(p_i)^{n_i}}{\Gamma(n_i+1)} \sum_{v=0}^{\infty} \frac{(p_1)^v \lambda^{(v+n_2+..+n_5)} e^{-\lambda}}{\Gamma(v+1)}$$

$$= e^{-\lambda(1-p_1)} \prod_{i=2}^{5} \frac{(\lambda p_i)^{n_i}}{\Gamma(n_i+1)}$$

The effect is that of a multinomial in the observed counts times a factor which accounts for the reduced probability available to them.

For any year, a similar formula holds, which can be obtained by thinking of merging all the empty bins and using the preceding derivation. The probabilities have individual Poisson parameters $\varepsilon_k\lambda$, and the product of the probabilities is the overall likelihood. Ignoring terms which do not depend upon λ or p_i, the negative logarithm of the likelihood (NLL) is the sum of the NLLs for each year:

$$\text{NLL} = \varepsilon_1\lambda(1-p_1-p_2) - \sum_{i=3}^{5} n_{i1} \{ \ln[p_i]+\ln[\lambda] \}$$

$$+\varepsilon_2\lambda(1-p_1) - \sum_{i=2}^{5} n_{i2}\{\ln[p_i]+\ln[\lambda]\}$$

$$+\varepsilon_3\lambda - \sum_{i=1}^{5} n_{i3}\{\ln[p_i]+\ln[\lambda]\} + \gamma(\sum_{i=1}^{5} p_i - 1)$$

A Lagrange multiplier term γ has been added, to facilitate solution. To find the maximum likelihood, we set equal to zero the partial derivatives with respect to γ, λ, and all the p_i :

$$\frac{\partial(NLL)}{\partial\gamma} = 0 \Rightarrow \qquad \sum_{i=1}^{5} p_i = 1$$

$$\frac{\partial(NLL)}{\partial p_1} = 0 \Rightarrow \qquad \frac{n_{13}}{p_1} = \gamma - (\varepsilon_1+\varepsilon_2)\lambda$$

$$\frac{\partial(NLL)}{\partial p_2} = 0 \Rightarrow \qquad \frac{n_{22}+n_{23}}{p_2} = \gamma - \varepsilon_1\lambda$$

$$\frac{\partial(NLL)}{\partial p_3} = 0 \Rightarrow \qquad \frac{n_{31}+n_{32}+n_{33}}{p_3} = \gamma$$

$$\frac{\partial(NLL)}{\partial p_4} = 0 \Rightarrow \qquad \frac{n_{41}+n_{42}+n_{43}}{p_4} = \gamma$$

$$\frac{\partial(NLL)}{\partial p_5} = 0 \Rightarrow \qquad \frac{n_{51}+n_{52}+n_{53}}{p_5} = \gamma$$

$$\frac{\partial(NLL)}{\partial\lambda} = 0 \Rightarrow \qquad \lambda = \frac{\sum_{i=1}^{5} N_i}{\varepsilon_1(1-p_1-p_2)+\varepsilon_2(1-p_1)+\varepsilon_3}$$

Thus, we end up with a nonlinear system of seven equations in seven unknowns.

Fortunately, the solution is both intuitive and easily generalized. The values $\varepsilon_k\lambda$ are the mean total number of events, <u>including</u> the unseen events, in year "k". Remembering that N_i is the total seen events in bin "i", the solution can be expressed as

$$p_1 = \frac{N_1}{\varepsilon_3\lambda}$$

$$p_2 = \frac{N_2}{(\varepsilon_2+\varepsilon_3)\lambda}$$

$$p_3 = \frac{N_3}{(\varepsilon_1+\varepsilon_2+\varepsilon_3)\lambda}$$

$$p_4 = \frac{N_4}{(\varepsilon_1+\varepsilon_2+\varepsilon_3)\lambda}$$

$$p_5 = \frac{N_5}{(\varepsilon_1+\varepsilon_2+\varepsilon_3)\lambda}$$

That is, the probability for each bin is the total number of events seen in it divided by the expected total number of events that could have contributed.

The quantity γ is the mean total number of events

$$\gamma = (\varepsilon_1+\varepsilon_2+\varepsilon_3)\lambda$$

and finally, the frequency parameter λ is

$$\lambda = \frac{N_1}{\varepsilon_3} + \frac{N_2}{(\varepsilon_2+\varepsilon_3)} + \frac{N_3+N_4+N_5}{(\varepsilon_1+\varepsilon_2+\varepsilon_3)}$$

The expected frequency is a sum over bins of the exposure-leveled number of seen events.

These rules seem quite intuitive. The generalizations to more complicated bin and/or date structures are fairly self-evident, as the same rules will still hold. Variable reporting levels by year would be one way the structure could be more complex.

Negative Binomial:

If the distribution is taken to be negative binomial instead of Poisson, when we go back to the discussion of year "2" the lemma is still straight-forward. The negative binomial probability with parameters (α, p) of observing N events is

$$NB(N,\alpha,p) = \frac{p^N \Gamma(N+\alpha)(1-p)^\alpha}{\Gamma(N+1)\Gamma(\alpha)}$$

$$\text{with} \quad \lambda \equiv \text{mean} = \frac{\alpha p}{1-p} \quad \text{and} \quad \frac{\text{variance}}{\text{mean}} = \frac{1}{1-p}$$

The probability of observing $n_2,..,n_5$ with no information on n_1 becomes

$$\text{prob} = \frac{(1-p)^\alpha \Gamma(n_2+..+n_5+\alpha)}{(1-pp_1)^{n_2+..+n_5+\alpha}\Gamma(\alpha)} \prod_{i=2}^{5} \frac{(pp_i)^{n_i}}{\Gamma(n_i+1)}$$

This has a similar form to the Poisson case, but with a different modifying function. The Poisson form is recovered in the limit p -> 0 with λ held constant.

The NLL for the three years has the corresponding changes. It is assumed that p, which governs the ratio of variance to mean, is held fixed, so that the exposure changes manifest (proportional to the mean values) through the $\alpha_k = \varepsilon_k \alpha$.

Apart from irrelevant quantities, the NLL is

$$\begin{aligned}
NLL = {} & \varepsilon_1 \alpha \ln\left[\frac{1 - p(p_1+p_2)}{1 - p}\right] + (n_{31}+..+n_{51})\ln[1 - p(p_1+p_2)] \\
& - \sum_{v=0}^{(n_{31}+..+n_{51}-1)} \ln[\varepsilon_1 \alpha + v] - \sum_{i=3}^{5} n_{i1} \{ \ln[p_i] + \ln[p] \} \\
& + \varepsilon_2 \alpha \ln\left[\frac{1 - pp_1}{1 - p}\right] + (n_{22}+..+n_{52})\ln[1 - pp_1] \\
& - \sum_{v=0}^{(n_{22}+..+n_{52}-1)} \ln[\varepsilon_2 \alpha + v] - \sum_{i=2}^{5} n_{i2} \{ \ln[p_i] + \ln[p] \} \\
& - \sum_{v=0}^{(n_{13}+..+n_{53}-1)} \ln[\varepsilon_3 \alpha + v] - \sum_{i=1}^{5} n_{i3} \{ \ln[p_i] + \ln[p] \}
\end{aligned}$$

Again, the extensions to more complicated date or bin structures follow the same form. The partial derivative equations here are far more complex than in the Poisson case. At this point it is easier just to work directly with the NLL and do the minimization numerically, rather than trying for analytic solutions (this is why the Lagrange term has been omitted).

Continuous distributions:

Often parameterization of the loss distribution - for example by a Pareto family - is of interest. Heuristically, this may be thought of as the limit where the bins become very small. All the n_i are zero or one (except for the case of identical losses), and the probabilities p_i are not independent, but given by the underlying distribution. Let us denote the lowest observable loss value for the year "k" by L_k; the underlying cumulative distribution function by F(x); and the corresponding probability density function by f(x), where we have suppressed the explicit parameter dependence in the severity distribution.

The parallel to the discussion of year "2" is that there are n events $x_1,..,x_n$ observed above the value L and the overall frequency is Poisson distributed with parameter λ. By a similar development to the earlier discussion, the probability of seeing these n events with no information below L is essentially

$$prob = e^{-\lambda(1-F(L))} \lambda^n \prod_{i=1}^{n} f(x_i)$$

The overall likelihood is the product of these for each year, as before:

$$\text{likelihood} = \prod_{k=1}^{3} e^{-\varepsilon_k\lambda[1-F(L_k)]}(\varepsilon_k\lambda)^{n_k}\prod_{i=1}^{n_k} f(x_{ik})$$

The corresponding NLL is, ignoring irrelevant terms,

$$\text{NLL} = \sum_{k=1}^{3}\left\{ \varepsilon_k\lambda[1-F(L_k)] - n_k\ln(\lambda) - \sum_{i=1}^{n_k} \ln[f(x_{ik})] \right\}$$

Letting N be the total number of seen events, this achieves the conceptually simpler and perhaps more familiar form

$$\text{NLL} = \sum_{k=1}^{3} \varepsilon_k\lambda[1-F(L_k)] - N\ln(\lambda) - \sum_{i=1}^{N} \ln[f(x_i)]$$

Equating to zero the partial derivative with respect to λ gives

$$\lambda = \frac{N}{\sum_{k=1}^{3} \varepsilon_k[1-F(L_k)]}$$

This equation is completely parallel to that of the last partial derivative in the multinomial case. It gives λ as a function of the data and the parameters of the distribution. The parallel solution for λ would be

$$\lambda = \frac{M_1}{(\varepsilon_1+\varepsilon_2+\varepsilon_3)} + \frac{M_2}{(\varepsilon_2+\varepsilon_3)} + \frac{M_3}{\varepsilon_3}$$

where M_1 is the total number of events greater than L_1, M_2 is the total number of events greater than L_2 and less than L_1, and M_3 is the total number of events greater than L_3 and less than L_2. Since there are many fewer degrees of freedom in this case than in the multinomial, this value of λ is unlikely to be the actual solution. However, it should provide a good starting point for the minimization of the NLL.

If we denote the parameters in the severity function collectively by the vector β, the partial derivative equations have the form

$$0 = \sum_{k=1}^{3}\left\{ -\varepsilon_k\lambda\frac{\partial F}{\partial\beta}(L_k,\beta) + \sum_{i=1}^{n_k} \frac{1}{f(x_{ik})}\frac{\partial f}{\partial\beta}(x_{ik},\beta)\right\}$$

Once more, numerical minimization is probably easier than trying to solve these equations.

The negative binomial case has a completely parallel development, with the probability of observing n events with no information below L being

$$\text{prob} = \frac{(1-p)^{\alpha}\Gamma(n+\alpha)}{\{1-p[1-F(L)]\}^{n+\alpha}\Gamma(\alpha)} p^n \prod_{i=1}^{n} f(x_i)$$

The likelihood and NLL for the three years have the corresponding changes. Again letting N be the total number of seen values,

$$\begin{aligned} \text{NLL} = & -\alpha \ln[1-p] \sum_{k=1}^{3} \varepsilon_k - N\ln(p) - \sum_{i=1}^{N} \ln[f(x_i)] \\ & + \sum_{k=1}^{3} \left\{ (n_k+\varepsilon_k\alpha)\ln[1-p(1-F(L_k))] - \sum_{v=0}^{(n_k-1)} \ln[\varepsilon_k\alpha+v] \right\} \end{aligned}$$

Conclusion:

The price we pay for being able to use more data is that the frequency and severity maximum likelihood calculations are now interdependent. This will induce correlations between the frequency and severity parameters, which will manifest in the variance-covariance matrix[1] resulting from the numerical minimization. In doing any model which allows for the uncertainty of the parameters, these correlations must be taken into account aw well as the parameter variance.

We lose, except in the simplest case, the possibility of finding analytic solutions. Fortunately, we usually want numbers anyway, and the explicit construction for the NLL allows (relatively) straightforward computation.

Addendum:

Since we have the NLL, we can also put in the possibility of trend by making λ or α an explicit function of time, in an obvious extension. Then for a given severity distribution family, there will be at least four possible frequency distributions: trended and untrended Poisson and negative binomial. The decision between them can be made on the basis of the smallest NLL, with appropriate allowance for the different numbers of parameters. One way of doing this is to use the Akaike[2] criterion: add to the minimized NLLs the number of parameters in the fit, and choose the lowest value.

---

[1]The derivation of the variance-covariance matrix from the mixed partial derivatives of the NLL is given in, for example, *Loss Distributions* by Hogg and Klugman, John Wiley and Sons (1984) page 81 and following.

[2]See the discussion in any good econometrics book, or go to Akaike, H. (1973), "Information Theory and the Extension of the Maximum Likelihood Principle," in B.N. Petrov and F. Csaki, eds., *2nd International "Symposium on Information Theory*, Akailseoniai-Kuido, Budapest, pp. 267-281 and the subsequent work, especially Akaike, H (1978), "On the Likelihood of a Time Series Model," Paper

Appendix - a formal derivation

Although a heuristic derivation of the continuous case was given earlier, the following is a formal derivation due to Ed Weissner which holds for either case.

Let A = a random sample was observed
B = of size N
C = with precisely M observations $\geq$ L
D = and the observations (no particular order) are $x_1, \ldots, x_M$

The likelihood function is given by

$$L(\lambda) = P[ACD] = \sum_{N=M}^{\infty} P[ABCD] \qquad \text{(law of total probabilities)}$$

$$= \sum_{N=M}^{\infty} P[AB]\ P[C|AB]\ P[D|ABC]$$

Now, AB obeys a Poisson law

C|AB obeys a Binomial Law with n = N, # of successes = M, and probability of success p defined by

$$p = 1 - F(L) \qquad \text{for continuous}$$

$$= \sum_{i \geq i(L)}^{\infty} p_i \qquad \text{for discrete}$$

D|CAB obeys a likelihood function that accounts for "no particular order" and draws each observation from the truncated distribution

$$\frac{f(x)}{1-F(L)} \qquad \text{for continuous}$$

$$\frac{p_i}{\sum_{i \geq i(L)}^{\infty} p_i} \qquad \text{for discrete}$$

Applying these to the likelihood above, it follows that for the continuous case

---

presented at the Institute of Statisticians 1978 Conference on Time Series Analysis, Cambridge University, Cambridge, England, July 1978.

$$L(\lambda) = \sum_{N=M}^{\infty} \left[\frac{e^{-\lambda}\lambda^N}{N!}\right] \left[\frac{N!}{M!(N-M)!} [1-F(L)]^M [F(L)]^{N-M}\right] \left[M! \frac{\prod f(x_i)}{[1-F(L)]^M}\right]$$

and in the discrete case

$$L(\lambda) = \sum_{N=M}^{\infty} [.....] [....] \left[\frac{M!}{\prod n_i!} \frac{\prod p_i^{n_i}}{[1-F(L)]^M}\right]$$

where the products $\prod_{i \geq i(L)}$ are i ≥ 2 for bin 1 missing, etc. Note that the combination of the binomial and "truncated multinomial" give the multinomial used in the text.