

EXPLORATIONS BY DAVE CLARK

## Estimation of Inverse Power Parameters via GLM

Richard Sherman's 1984 paper "Extrapolating, Smoothing and Interpolating Development Factors" provided a number of useful ideas for working with development patterns and remains a useful resource. The CAS's 2013 Tail Factor Working Party found that the Sherman curve fit "enjoys fairly broad acceptance both with consulting firms and insurance companies."

In *Variance*, Volume 9, Number 2, Jon Evans has two new papers that extend the ideas of Sherman's original paper and show its continued relevance.

The great value of the original Sherman paper is in identifying a curve form that closely fits the sequence of age-to-age (ATA) factors for long-tailed casualty lines. The most basic form is the "inverse power" curve of the following form:

$$ATA_t = 1+a \cdot t^b.$$

In this form, the  $t$  represents the development time such that, for example,  $ATA_{12}$  represents the age-to-age factor or link ratio from age 12 months to age 24 months. A modest expansion of this formula allows a shift term,  $c$ , to be added to the time index, though we will ignore this for the present discussion:

$$ATA_t = 1+a \cdot (t+c)^b.$$

The parameters of the inverse power curve are most frequently estimated by rearranging the formula into a (log) linear form and then applying ordinary least squares formulas for the intercept and slope:

$$\ln(ATA_t - 1) = \ln(a) + b \cdot \ln(1/t).$$

The attraction of this log-linear

form is that simple, closed-form solutions can produce the estimated model parameters. Anyone with a spreadsheet can apply the method with little technical knowledge.

Further, the inverse power curve can easily be compared with alternative fitted curves. Sherman gives several examples, with the exponential decay formula being perhaps most familiar:

$$ATA_t = 1+a \cdot e^{-b \cdot t}.$$

The exponential decay formula can be calculated in a similar log-linear form, so that we quickly have alternative fitted curves to compare to our development data:

$$\ln(ATA_t - 1) = \ln(a) - b \cdot t.$$

While the mathematical simplicity of the log-linear form is appealing, it creates difficulties in practice. The difficulties were noted in the discussion of Sherman's paper by Lowe and Mohrman (1985). The first difficulty is that the log-transform  $\ln(ATA_t - 1)$  requires that every ATA factor used in the fit be strictly greater than 1.000. There can be no "negative development" in the actual data, and even factors that are only slightly greater than 1.000 can cause distortions in the fit.

A second problem is that the log-transformed data is a bit more difficult to interpret or explain to the audience receiving the results of the analysis. An age-to-age factor of 1.010 is easily interpreted as a 1 percent increase in loss dollars, but what does  $\ln(.01) = -4.605$  represent? How do we interpret the -4.605 for our client or explain why we

want a fitted line that closely matches this value?

Both of these difficulties are overcome when we instead approach the parameter estimation using generalized linear models (GLM). We can still use the "inverse power" form that fits the insurance patterns so well, but make use of a better technique for the parameter estimation.

The key idea in GLM is that we include a "link function"  $g()$  but apply it in inverse form  $g^{-1}()$  to the linear combination of the predictor variable(s). Rather than apply a log-transform to the quantity  $(ATA_t - 1)$ , we use an exponential transform on the linear function.

$$ATA_t - 1 = \exp(\beta_0 + \beta_1 \cdot \ln(t)) = \mu_t$$

$$a = \exp(\beta_0) \quad b = -\beta_1$$

Using this "log-link" on the right side of the equation rather than applied to the response variable, we avoid any problem with actual negative development. Expected development must still be positive but the actual values being fit need not be. In short, a log-link GLM can handle negative development in the data where a log-linear regression cannot. The GLM approach is more robust.

With the log-link, the "canonical" variance structure is the quasi-Poisson or over-dispersed Poisson (ODP) model. The ODP model assumes that the variance is proportional to the expected value.

The GLM application follows a Poisson quasi-loglikelihood (QLL). The prefix *quasi* means that we are not explicitly assuming a distribution but

rather only assuming that the variance is proportional to the variance of the Poisson distribution.

The reader is referred to the 1974 Wedderburn paper for a more complete description of quasi-likelihoods.

For our application the Poisson QLL is given below:

$$QLL = \sum w_i \cdot [(ATA_i - 1) \cdot \ln(\mu_i) - \mu_i]$$

The function allows weights  $w_i$  to be included as part of the fitting procedure. Since we typically use dollar-weighted average ATA factors, the weights are naturally set as the sum of the dollars in the column used in the denominator of the ATA calculation

$$ATA_i = \frac{\sum_{t=1}^{n-i} C_{i,t+1}}{\sum_{t=1}^{n-i} C_{i,t}} w_i \sum_{t=1}^{n-i} C_{i,t}$$

The QLL can be maximized with the “best” parameters  $\beta_0$  and  $\beta_1$  using available software. The **glimmix** procedure in SAS will perform the calculation. The **glm.fit** function in R can also be used but requires a fix to allow negative values (see the code by David Firth in the references). More conveniently, a simple iterative routine can be built into a VB function within an Excel spreadsheet (or even — gasp — using Excel’s “Solver”).

The estimating equations for finding the best model parameters are easily derived:

$$\sum w_i \cdot (ATA_i - 1) = \sum w_i \cdot (\widehat{ATA}_i - 1)$$

$$\sum w_i \cdot (ATA_i - 1) \cdot \ln(t) = \sum w_i \cdot (\widehat{ATA}_i - 1) \cdot \ln(t)$$

From these estimating equations, we see that GLM estimation is working with the original dollars from the development triangle, and that the fitted values balance to the actual dollars. There is no difficulty when some actual development is negative and no difficulty in interpreting what is being fit.

The GLM can also be expanded for other transforms of the development time index. Instead of the logarithmic transform that creates the inverse power curve, we can use the time index directly to be equivalent to the exponential decay curve.

If the inverse power curve is too thick-tailed and the exponential decay is too thin-tailed, then other transforms are possible. An intermediate form is to use the square root of the development time.

As with the original Sherman paper, these various transforms of the time index represent variations on the same basic model. Using the log-link GLM form simply gives us a more robust method for estimating parameters for the model. ●

---

*David R. Clark, FCAS, MAAA, works for Munich Reinsurance as part of the actuarial research and modeling team in Princeton, New Jersey.*

## References

- Evans, Jonathan Palmer, “A Continuous Version of Sherman’s Inverse Power Curve Model with Simple Cumulative Development Factor Formulas,” *Variance* 9:2, 2015, pp. 187-195.
- Evans, Jonathan Palmer, “Tail Factor Convergence in Sherman’s Inverse Power Curve Loss Development Factor Model,” *Variance* 9:2, 2015, pp. 227-233.
- Firth, D (2003). An amended quasi-poisson function for R. Available at <http://bit.ly/2paRUXA>.
- Lowe, S.P. and D.F. Mohrman, “Discussion of ‘Extrapolating, Smoothing and Interpolating Development Factors,’” *Proceedings of the Casualty Actuarial Society* 72, 1985, pp. 182-189.
- McCullagh, P. and J.A. Nelder, *Generalized Linear Models*, Chapman & Hall/CRC, second edition 1989.
- Sherman, R.E., “Extrapolating, Smoothing and Interpolating Development Factors,” *Proceedings of the Casualty Actuarial Society* 71, 1984, pp. 122-155.
- “The Estimation of Loss Development Tail Factors: A Summary Report,” *Casualty Actuarial Society Forum*, 2013: Fall, Vol. 1, pp. 1-111.
- Wedderburn, R. (1974) “Quasi-Likelihood Functions, Generalized Models and the Gauss-Newton Method,” *Biometrika* 61(3), pp. 439-443.

Inverse Power	Exponential Decay	Square Root
$\mu_i = \exp(\beta_0 + \beta_1 \cdot \ln(t))$	$\mu_i = \exp(\beta_0 + \beta_1 \cdot t)$	$\mu_i = \exp(\beta_0 + \beta_1 \cdot \sqrt{t})$