A Least Squares Method of Producing Bornhuetter-Ferguson Initial Loss Ratios

Paul Brehm, FCAS, MAAA

Actuaries have relied on the Bornhuetter-Ferguson methodology in loss reserving since the "The Actuary and IBNR" [2] was published in 1972. The methodology is an intuitively appealing, credibility-weighted compromise between link ratio and expected loss ratio methods, where 'credibility' is inversely proportional to the remainder of the loss development tail. However, for almost as long as this method has been in existence, practitioners have been asking, "What do I use for my expected loss ratios?" Answers to this question (often unsatisfying ones) include industry data, company data for comparable classes of business, loss ratio pricing targets, planned loss ratios, and more.

This paper addresses the above question by offering a methodology for producing underlying loss ratios for use in the Bornhuetter-Ferguson method that are derived from the data itself. In particular, this paper addresses how to determine the underlying loss ratio for the initial time period in the analysis using a least squares methodology. The initial loss ratio is then used as the seed value for all subsequent loss ratios.

1. Derivation

Let:

L_{ij} = loss ratio for accident period i (i=1,…,n) evaluated cumulatively at j (j=1,…,m)

F_j = development factor from age j to ultimate

U_i^{BF} = Bornhuetter-Ferguson estimate of ultimate for accident period i

U_i^* = underlying ultimate loss ratio for accident period i (used in the Bornhuetter-Ferguson formula)

T_i = trend from time (i-1) to i – accident year dimension

P_i = earned effect of pricing from time (i-1) to i

Then the Bornhuetter-Ferguson estimate of the ultimate loss ratio for accident period i, with cumulative losses evaluated at j is:

$$U_i^{BF} = L_{ij} + U_i^*\left(1 - \frac{1}{F_j}\right) \tag{1.1}$$

An alternative estimate for accident period i can be derived by using losses evaluated one period earlier with the appropriate development factor, F_{j-1}:

$$U_i^{BF} = L_{i,j-1} + U_i^*\left(1 - \frac{1}{F_{j-1}}\right) \tag{1.2}$$

If (1.2) is subtracted from (1.1), the difference in estimated ultimates should be zero, but for some estimation error:

$$U_i^{BF} - U_i^{BF} = L_{ij} + U_i^*\left(1 - \frac{1}{F_j}\right) - L_{i,j-1} - U_i^*\left(1 - \frac{1}{F_{j-1}}\right) = 0 + \varepsilon_{ij}$$

Alternatively, after some manipulation:

$$\left(L_{ij} - L_{i,j-1}\right) = U_i^*\left[\left(1 - \frac{1}{F_{j-1}}\right) - \left(1 - \frac{1}{F_j}\right)\right] + \varepsilon_{ij} \tag{1.3}$$

Note that the term on the left of equation (1.3) is simply the incremental loss ratio. The term on the right is its expectation, conditioned on the underlying or expected loss ratio and the selected development pattern. If j = 1, that is, the first evaluation, then the term $\frac{1}{F_{j-1}}$ is undefined. For this initial condition, let $\frac{1}{F_{j-1}} = 0$, and the bracketed term on the right becomes $\left[\frac{1}{F_j}\right]$.

Now assume that the accident period loss ratios can be linked together over time by periodic trend (T_i) and pricing (P_i) factors according to:

$$U_i^* = U_{i-1}^* \frac{\left(1 + T_i\right)}{\left(1 + P_i\right)} \tag{1.4}$$

By successive substitutions, all underlying ultimate loss ratios can be linked back to the initial underlying loss ratio:

$$U_i^* = U_1^* \prod_{k=2}^{i} \frac{(1+T_k)}{(1+P_k)} \tag{1.5}$$

Substituting (1.5) into (1.3) for U_i^* yields the general formula

$$(L_{ij} - L_{i,j-1}) = U_1^* \prod_{k=2}^{i} \frac{(1+T_k)}{(1+P_k)} \left[\left(1 - \frac{1}{F_{j-1}}\right) - \left(1 - \frac{1}{F_j}\right) \right] + \varepsilon_{ij} \tag{1.6}$$

Note that (1.6) is of the form $Y_{ij}=\beta X_{ij}$, where the Y_{ij} are incremental loss ratios, the X_{ij} are the 'independent variables,' and β is the initial underlying loss ratio seed for the Bornhuetter-Ferguson model, U^*_1. The independent variables are simply derived values constructed from trend and pricing factors in the accident period dimension and loss development factors in the development dimension. Formula (1.6), then, can be estimated as a simple linear regression through the origin.

The parameter estimate, $\hat{\beta}$, is the initial loss ratio we are solving for. Think of the result as the initial loss ratio that is the least squares best estimate based on the data and conditioned on all the assumptions concerning pricing, trend, and loss development.

If it is assumed that trend is constant over the experience period, i.e., $T_i = T$ for all i, the formula (1.6) simplifies to:

$$(L_{ij} - L_{i,j-1}) = U_1^* (1+T)^{i-1} \left[\prod_{k=2}^{i} \frac{1}{(1+P_k)} \right] \left[\left(\frac{1}{1-F_{j-1}}\right) - \left(\frac{1}{1-F_j}\right) \right] + \varepsilon_{ij} \tag{1.7}$$

The functional form of (1.7) is particularly useful. Given a set of loss development factors and an earned price index, the above regression can be iterated over a range of annual trend assumptions. The final model can be chosen based on the underlying trend that maximizes R^2. (I know, it's data mining.)

Once U_1^* has been estimated, subsequent underlying loss ratios can be estimated as

$$U_i^* = U_1^* \prod_{k=2}^{i} \frac{(1+T_k)}{(1+P_k)} \tag{1.8}$$

or

$$U_i^* = U_1^* (1+T)^{i-1} \prod_{k=2}^{i} \frac{1}{(1+P_k)} \tag{1.9}$$

for the constant trend case.

Since this is a regression model, the estimate $\hat{\beta}$ of U_1^* has an associated standard error, and a confidence interval can be established. The variance of $\hat{\beta}$ is

$$\sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum x_{ij}^2} \tag{1.10}$$

where x_{ij} are the independent variables.

An unbiased estimate of σ^2 is

$$S^2 = \frac{\sum \varepsilon_{ij}^2}{(n-1)} \tag{1.11}$$

where the ε_{ij} are the residuals from the regression, and n is the number of terms in the regression. There are n-1 degrees of freedom, as we are only estimating one parameter. The standard error of the coefficient -- the square root of the variance -- can be calculated as

$$S_{\hat{\beta}} = \left[\frac{\sum \varepsilon_{ij}^2}{(n-1)\sum x_{ij}^2} \right]^{1/2} \tag{1.12}$$

The 100-α% confidence interval around $\hat{\beta}$ is

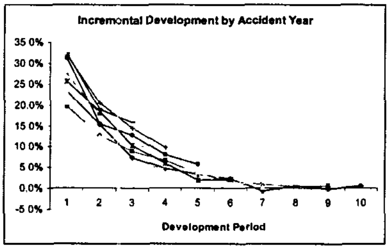
$$\hat{\beta} \pm t_{\alpha/2} S_{\hat{\beta}} \tag{1.13}$$

The confidence interval (1.13) can be used to establish a range of estimates and thereby gauge the sensitivity of the reserve indication. For example, in the case where a trend factor is also estimated by ordinary least squares, a confidence interval can be estimated for the trend factor, as well. If a low estimate of trend is paired with the lower bound of β in formula (1.9) and a high estimate of trend is likewise paired with the upper bound of the confidence interval for β, both a low and a high loss ratio pattern can be traced over accident periods and used in the Bornhuetter-Ferguson estimation to derive low and high reserve estimates.

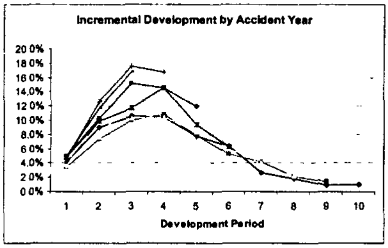
2. Example

Following is an example based on general liability data. The graphs below show the incremental loss ratios by accident period over time (development period) – case incurred on the left, paid data on the right.

Graph 2.1
Incurred Data
Accident Years over (Development) time

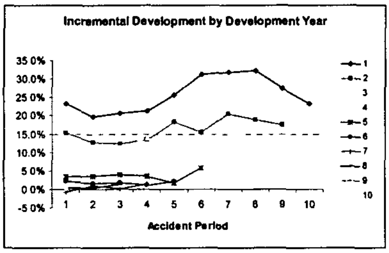


Graph 2.2
Paid Data
Accident Years over (Development) time

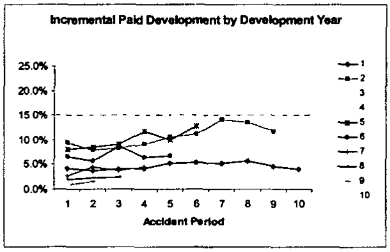


When viewed by development period over accident period below, the incremental loss ratios by evaluation would ideally behave like random pattern of points about a smooth trend line, if a constant trend and on-level factors truly picked up all the sources of systematic change over time. However, the data shows a departure in the pattern over accident periods starting in accident period 5 (see Graphs 2.3 and 2.4, below). This suggests a non-constant trend parameter or, alternatively, something affecting the loss ratios other than trend, e.g., underwriting or mix changes. In reality the departure associated with accident period 5 may well be better characterized as a calendar period distortion. Barnett and Zehnwirth's model [1] may be a good alternative in this case.

Graph 2.3
Incurred Data
Development Periods over (Accident) time

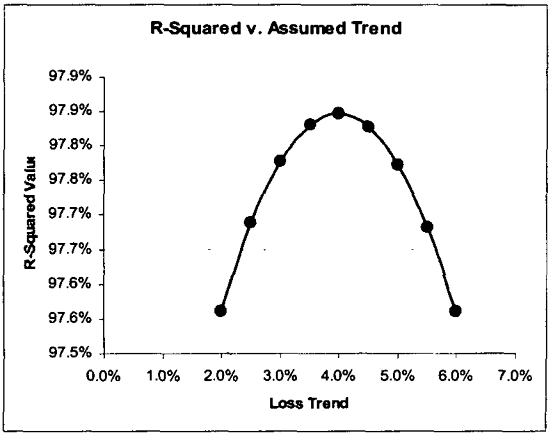


Graph 2.4
Paid Data
Development Periods over (Accident) time



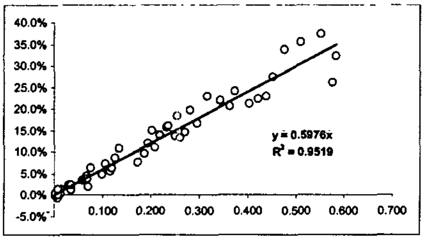
In this example, the constant trend case was modeled first for illustration purposes. In the example data, the least squares trend estimate using an exponential trend fit to pure premium was 3.5% (with an associated standard error of 0.016). However, R^2 was maximized using a trend of 4%:

Graph 2.5

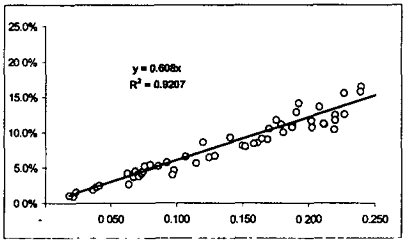


The resulting regressions can be seen below.

Graph 2.6
Incurred Data Regression



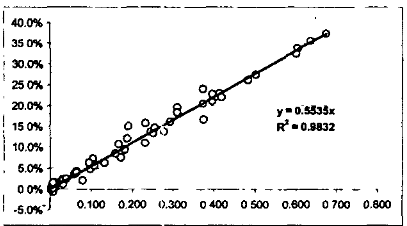
Graph 2.7
Paid Data Regression



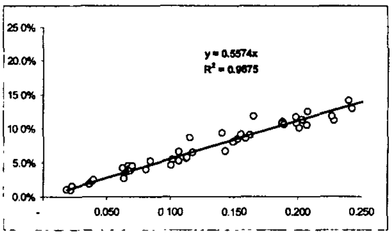
It can be seen from the above regressions that the paid and incurred data yield consistent results (initial loss ratios of 59.8% and 60.8%) from models with a strong goodness of fit (R^2 values of 96% and 92%, respectively). It would be appropriate, and more thorough, at this point to examine the residuals for serial correlation and non-constant error variance (heteroscedasticity). If either was a problem, the regressions could be adjusted accordingly.

To continue this example, trend was next assumed to vary over time. Underlying annual trend was set to 2% (rather than 4.0% overall), with additional period-on-period changes added to accident periods 5 through 9 to account for the calendar period distortion or "surprise" trend[1] (much like the industry observed in liability coverages in the late 90's). The resulting regressions are shown below.

Graph 2.8
Incurred Data Regression 2



Graph 2.9
Paid Data Regression 2



[1] I've never tried it, but it occurs to me that the regression could simply be augmented with 'distortion dummies' to automatically estimate the degree of departure from an underlying trend. This will be a subject of future research.

In the revised regressions, the case incurred estimate of the initial loss ratio is 55.3% (R^2 of 98.3%) and the paid loss estimate of the initial loss ratio is 55.7% (R^2 of 96.8%). $S_{\hat{\beta}}$ was 0.7% for both the paid and case incurred data, yielding a 95% confidence interval of roughly +/-1.5 loss ratio points at time 1.

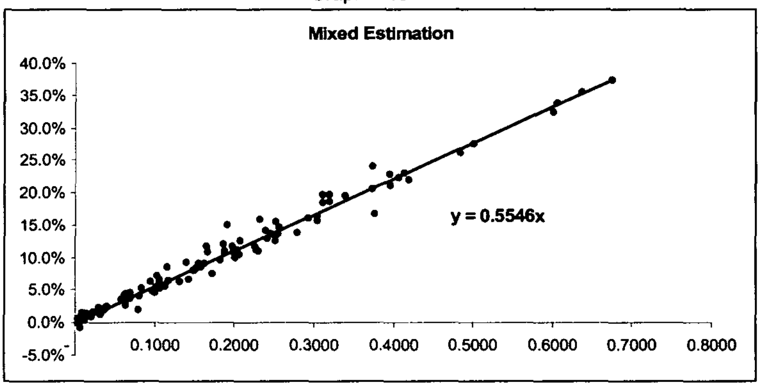
Given two estimates of the seed loss ratio, along with their respective error variances, we can credibility weight the two together to get one estimate. The formula for the paid data credibility parameter is:

$$Z_{Paid} = \frac{\left(\frac{1}{S^2_{Paid}}\right)}{\left(\left(\frac{1}{S^2_{Paid}}\right)+\left(\frac{1}{S^2_{Incurred}}\right)\right)} \tag{2.1}$$

where S^2 is shown above in (1.11).

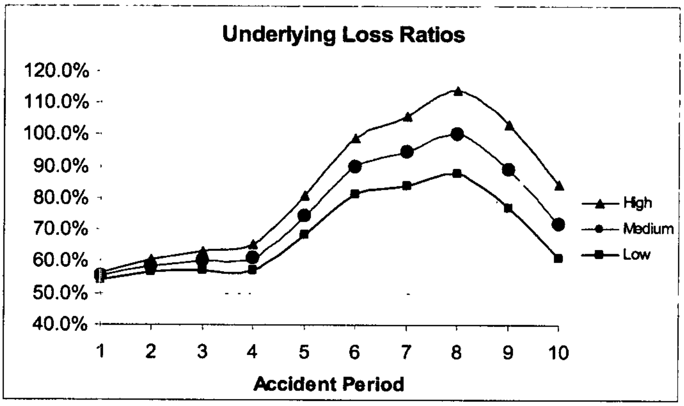
In practice, the credibility weighted solution can be derived directly by combining the paid and incurred regression matrices and doing a single, mixed regression [3]. The mixed estimate for this example is shown graphically in Graph 2.10. The mixed estimate of the initial loss ratio is 55.5% with an R^2 of 99.2%. $S_{\hat{\beta}}$ is 0.46% for both the paid and case incurred data, yielding a 95% confidence interval of roughly +/-1.1 loss ratio points at time 1.

Graph 2.10



The mixed estimate initial loss ratio, U^*_1, and the trend assumptions applied in the regression model substituted into formula (1.8) yields a pattern of underlying loss ratios as shown below. For the sake of this graph, the low and high loss ratios were calculated according to formula (1.13) for accident period 1 using a 95% confidence level. Subsequent accident period ultimate loss ratios were calculated with the selected trend plus and minus 1.6% respectively – one standard deviation around the least squares annual loss cost trend.

Graph 2.11



3. Conclusion

The above method has a strong appeal. Its strengths include utilizing all readily available data (dollars, counts, trends, premiums, exposures, pricing) and utilizing paid and incurred losses simultaneously to produce a 'best' (least squares) answer, in a computationally tractable manner, while still allowing the flexibility for ample actuarial judgment.

This method has always served me well, even with 'misbehaved' or sparse data. I hope it fills a need in your actuarial tool box.

References

[1] Barnett, G., and Zehnwirth, B. "Best Estimates for Reserves," 2000, Proceedings of the Casualty Actuarial Society, Vol. LXXXVII, p. 245

[2] Bornhuetter, R., and Ferguson, R. "The Actuary and IBNR," PCAS LIX, 1972, p. 181

[3] Brehm, P., and Guenthner D., "The Econometric Method of Mixed Estimation, An Application to the Credibility of Trend," CAS Discussion Paper Program, 1990.