

Regression Models and Loss Reserving

Leigh J. Halliwell, FCAS, MAAA
Consulting Actuary
leigh@lhalliwell.com

Casualty Loss Reserve Seminar
San Diego, GA
September 10-11, 2007

1

Outline

- Introductory Example
- Linear (or Regression) Models
- The Problem of Stochastic Regressors
- Reserving Methods as Linear Models
- Covariance

2

Introductory Example

A pilot is flying straight from X to Y. Halfway along (s)he realizes that (s)he's ten miles off course. What does (s)he do?

3

Linear (Regression) Models

- "Regression toward the mean" coined by Sir Francis Galton (1822-1911).
- The real problem: Finding the Best Linear Unbiased Estimator (BLUE) of vector y_2 , vector y_1 observed.
- $y = X\beta + e$. X is the design (regressor) matrix. β unknown; e unobserved, but (the shape of) its variance is known.
- For the proof of what follows see Halliwell [1997] 325-336.

4

The Formulation

$$\begin{bmatrix} y_1(t_1 \times 1) \\ y_2(t_2 \times 1) \end{bmatrix} = \begin{bmatrix} X_1(t_1 \times k) \\ X_2(t_2 \times k) \end{bmatrix} \beta_{(k \times 1)} + \begin{bmatrix} e_1(t_1 \times 1) \\ e_2(t_2 \times 1) \end{bmatrix},$$

$$\text{Var} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11}(t_1 \times t_1) & | & \Sigma_{12}(t_1 \times t_2) \\ \hline \Sigma_{21}(t_2 \times t_1) & | & \Sigma_{22}(t_2 \times t_2) \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} \Phi_{11}(t_1 \times t_1) & | & \Phi_{12}(t_1 \times t_2) \\ \hline \Phi_{21}(t_2 \times t_1) & | & \Phi_{22}(t_2 \times t_2) \end{bmatrix}$$

5

Trend Example

$$\begin{bmatrix} y_1(5 \times 1) \\ y_2(3 \times 1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & -6 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \beta_{(2 \times 1)} + \begin{bmatrix} e_1(5 \times 1) \\ e_2(3 \times 1) \end{bmatrix},$$

$$\text{Var} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \sigma^2 \begin{bmatrix} I_{(5 \times 5)} & | & 0_{(5 \times 3)} \\ \hline 0_{(3 \times 5)} & | & I_{(3 \times 3)} \end{bmatrix}$$

6

The BLUE Solution

$$\hat{y}_2 = X_2 \hat{\beta} + \Phi_{21} \Phi_{11}^{-1} (y_1 - X_1 \hat{\beta})$$

$$\hat{\beta} = (X_1' \Phi_{11}^{-1} X_1)^{-1} X_1' \Phi_{11}^{-1} y_1$$

$$\text{Var} [y_2 - \hat{y}_2] = \sigma^2 (\underbrace{\Phi_{22} - \Phi_{21} \Phi_{11}^{-1} \Phi_{12}}_{\text{process variance}} + \underbrace{(X_2 - \Phi_{21} \Phi_{11}^{-1} X_1) \text{Var} [\hat{\beta}] (X_2 - \Phi_{21} \Phi_{11}^{-1} X_1)'}_{\text{parameter variance}})$$

$$\text{Var} [\hat{\beta}] = \sigma^2 (X_1' \Phi_{11}^{-1} X_1)^{-1}$$

7

Special Case: $\Phi = I_t$

$$\hat{y}_2 = X_2 \hat{\beta}$$

$$\hat{\beta} = (X_1' X_1)^{-1} X_1' y_1$$

$$\text{Var} [y_2 - \hat{y}_2] = \sigma^2 I_{t_2} + X_2 \text{Var} [\hat{\beta}] X_2'$$

$$\text{Var} [\hat{\beta}] = \sigma^2 (X_1' X_1)^{-1}$$

8

Estimator of the Variance Scale

$$\hat{\sigma}^2 = \frac{(y_1 - X_1 \hat{\beta})' \Phi_{11}^{-1} (y_1 - X_1 \hat{\beta})}{t_1 - k}$$

9

Remarks on the Linear Model

- Actuaries need to learn the matrix algebra.
- Excel OK; but statistical software is desirable.
- X_1 is of full column rank, Σ_{11} non-singular.
- Linearity Theorem: $\widehat{A} y_2 = A \hat{y}_2$
- Model is versatile. My four papers (see References) describe complicated versions.

10

The Problem of Stochastic Regressors

- See Judge [1988] 571ff; Pindyck and Rubinfeld [1998] 178ff.
- If X is stochastic, the BLUE of β may be biased:

$$\hat{\beta} = (X'X)^{-1} X'y$$

$$= (X'X)^{-1} X'(X\beta + e)$$

$$= \beta + (X'X)^{-1} X'e$$

$$E[\hat{\beta}] = \beta + E[(X'X)^{-1} X'e]$$

$$\neq \beta + E[(X'X)^{-1} X'] E[e] = \beta$$

11

The Clue: Regression toward the Mean

To intercept or not to intercept?

12

What to do?

- Ignore it.
- Add an intercept.
 - Barnett and Zehnwirth [1998] 10-13, notice that the significance of the slope suffers. The lagged loss may not be a good predictor.
 - Intercept should be proportional to exposure.
- Explain the torsion. Leads to a better model?

13

Galton's Explanation

- Children's heights regress toward the mean.
 - Tall fathers tend to have sons shorter than themselves.
 - Short fathers tend to have sons taller than themselves.
- Height = "genetic height" + environmental error
- A son inherits his father's *genetic* height:
 - ∴ Son's height = father's genetic height + error.
- A father's height proxies for his genetic height.
 - A tall father probably is less tall genetically.
 - A short father probably is less short genetically.
- Excellent discussion in Bulmer [1979] 218-221. Cf. also sportsoci.org/resource/stats under "Regression to Mean."

14

The Lesson for Actuaries

- Loss is a function of exposure.
- Losses in the design matrix, i.e., stochastic regressors (SR), are probably just proxies for exposures. Zero loss proxies zero exposure.
- The more a loss varies, the poorer it proxies.
- The torsion of the regression line is the clue.
- Reserving actuaries tend to ignore exposures – some even glad not to have to "bother" with them!
- SR may not even be significant.
- Covariance is an alternative to SR (see later).
- Stochastic regressors are nothing but trouble!

15

Reserving Methods as Linear Models

- The loss rectangle: AY_i at age j
- Often the upper left triangle is known; estimate lower right triangle.
- The earlier AYs lead the way for the later AYs.
- The time of each ij -cell is known – we can discount paid losses.
- Incremental or cumulative, no problem. (But variance structure of incrementals is simpler.)

16

The Basic Linear Model

$$y_{ij} = a_{ij} x_i f_j r + e_{ij} \quad \sum_j f_j = 1$$

- y_{ij} incremental loss of ij -cell
- a_{ij} adjustments (if needed, otherwise = 1)
- x_i exposure (relativity) of AY_i
- f_j incremental factor for age j (sum constrained)
- r pure premium
- e_{ij} error term of ij -cell

17

Familiar Reserving Methods

$$\mathbf{Y} = (\mathbf{X})(\boldsymbol{\beta}) + \mathbf{e}$$

$$y_{ij} = (f_j)(x_i r) + e_{ij} \quad \text{quasi Chain Ladder}$$

$$y_{ij} = (x_i f_j r)(1) + e_{ij} \quad \text{Bornhuetter - Ferguson}$$

$$y_{ij} = (x_i f_j)(r) + e_{ij} \quad \text{Stanard - Bühlmann}$$

$$y_{ij} = (x_i)(f_j r) + e_{ij} \quad \text{Additive}$$

- BF estimates zero parameters.
- BF, SB, and Additive constitute a progression.
- The four other permutations are less interesting.
- No stochastic regressors

18

Why not Log-Transform?

$$\ln y_{ij} = \ln x_i + \ln f_j + \ln r + e_{ij}$$

- Barnett and Zehnwirth [1998] favor it.
- Advantages:
 - Allows for skewed distribution of $\ln y_{ij}$.
 - Perhaps easier to see trends
- Disadvantages:
 - Linearity compromised, i.e., $\ln(Ay) \neq A \ln(y)$.
 - $\ln(x \leq 0)$ undefined.
- Something Better: Simulation with non-normal error terms (robust estimation, Judge [1998], ch. 22)

19

The Ultimate Question

- Last column of rectangle is ultimate increment.
- There may be no observation in last column:
 - Exogenous information for late parameters f_j or $f_j\beta$.
 - Forces the actuary to reveal hidden assumptions.
 - See Halliwell [1996b] 10-13 and [1998] 79.
- Risky to extrapolate a pattern. It is the *hiding*, not the *making*, of assumptions that ruins the actuary's credibility. Be aware and explicit.

20

Linear Transformations

- Results: \hat{y}_2 and $Var [y_2 - \hat{y}_2]$
- Interesting quantities are normally linear:
 - AY totals and grand totals
 - Present values

- Powerful theorems (Halliwell [1997] 303f):

$$E[A\hat{y}_2] = AE[\hat{y}_2]$$

$$Var[Ay_2 - A\hat{y}_2] = AVar[y_2 - \hat{y}_2]A'$$

- The present-value matrix is diagonal in the discount factors.

21

Transformed Observations

$$\begin{bmatrix} Ay_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} AX_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} Ae_1 \\ e_2 \end{bmatrix},$$

$$Var \begin{bmatrix} Ae_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} A\Sigma_{11}A' & A\Sigma_{12} \\ \Sigma_{21}A' & \Sigma_{22} \end{bmatrix}$$

If A^{-1} exists, then the estimation is unaffected. Use the BLUE formulas on slide 7.

22

Example in Excel



WC Example.Ink

23

Covariance

- An example like the introductory one:
 - From Halliwell [1996a], 436f and 446f.
 - Prior expected loss is \$100; reaches ultimate at age 2. Incremental losses have same mean and variance.
 - The loss at age 1 has been observed as \$60.
 - Ultimate loss: \$120 CL, \$110 BF, \$100 Prior Hypothesis.
- Use covariance, not the loss at age 1, to do what the CL method purports to do.

24

Generalized Linear Model

$$\begin{bmatrix} 60 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} 0.5 \cdot 100 \\ 0.5 \cdot 100 \end{bmatrix} (1) + \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

Off-diagonal element

$$\hat{\mathbf{y}}_2 = (0.5 \cdot 100)(1) + (\rho \sigma^2)(1 \sigma^2)^{-1} (60 - (0.5 \cdot 100)(1))$$

$$= 50 + 10\rho$$

$$\text{Var}[\mathbf{y}_2 - \hat{\mathbf{y}}_2] = (1 - \rho^2) \sigma^2$$

Result: $\rho = 1$ CL, $\rho = 0$ BF, $\rho = -1$ Prior Hypothesis

25

Conclusion

- Typical loss reserving methods:
 - are primitive linear statistical models
 - originated in a bygone deterministic era
 - underutilize the data
- Linear statistical models:
 - are BLUE
 - obviate stochastic regressors with covariance
 - have desirable linear properties, especially for present-valuing
 - fully utilize the data
 - are versatile, of limitless form
 - force the actuary to clarify assumptions

26

References

- Barnett, Glen, and Ben Zehnwirth, "Best Estimates for Reserves," *PCAS LXXXVII* (2000), 245-321.
- Bulmer, M.G., *Principles of Statistics*, Dover, 1979.
- Halliwell, Leigh J., "Loss Prediction by Generalized Least Squares," *PCAS LXXXIII* (1996), 436-489.
- "", "Statistical and Financial Aspects of Self-Insurance Funding," *Alternative Markets / Self Insurance*, 1996, 1-46.
- "", "Conjoint Prediction of Paid and Incurred Losses," *Summer 1997 Forum*, 241-379.
- "", "Statistical Models and Credibility," *Winter 1998 Forum*, 61-152.
- Judge, George G., et al., *Introduction to the Theory and Practice of Econometrics*, Second Edition, Wiley, 1988.
- Pindyck, Robert S., and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, Fourth Edition, Irwin/McGraw-Hill, 1998.
- Venter, Gary G., "Testing the Assumptions of Age-to-Age Factors," *PCAS LXXXV* (1998), 807-847.

27