# Regression Models and Loss Reserving

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

**Casualty Loss Reserve Seminar**
**Atlanta, GA**
**September 12, 2006**



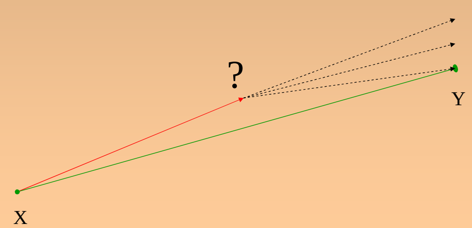
## Outline

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



## 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?






## Linear (Regression) Models

- "Regression toward the mean" coined by Sir Francis Galton (1822-1911).
- The real problem: Finding the <u>B</u>est <u>L</u>inear <u>U</u>nbiased <u>E</u>stimator (BLUE) of vector $\mathbf{y}_2$, vector $\mathbf{y}_1$ observed.
- $\mathbf{y} = X\beta + \mathbf{e}$. X is the design (regressor) matrix. $\beta$ unknown; $\mathbf{e}$ unobserved, but (the shape of) its variance is known.
- For the proof of what follows see Halliwell [1997] 325-336.



## The Formulation

$$\begin{bmatrix} \mathbf{y}_{1(t_1\times 1)} \\ \hline \mathbf{y}_{2(t_2\times 1)} \end{bmatrix} = \begin{bmatrix} X_{1(t_1\times k)} \\ \hline X_{2(t_2\times k)} \end{bmatrix} \beta_{(k\times 1)} + \begin{bmatrix} \mathbf{e}_{1(t_1\times 1)} \\ \hline \mathbf{e}_{2(t_2\times 1)} \end{bmatrix},$$

$$Var\begin{bmatrix} \mathbf{e}_1 \\ \hline \mathbf{e}_2 \end{bmatrix} = \left[\begin{array}{c|c} \Sigma_{11(t_1\times t_1)} & \Sigma_{12(t_1\times t_2)} \\ \hline \Sigma_{21(t_2\times t_1)} & \Sigma_{22(t_2\times t_2)} \end{array}\right]$$

$$= \sigma^2 \left[\begin{array}{c|c} \Phi_{11(t_1\times t_1)} & \Phi_{12(t_1\times t_2)} \\ \hline \Phi_{21(t_2\times t_1)} & \Phi_{22(t_2\times t_2)} \end{array}\right]$$



## Trend Example

$$\begin{bmatrix} \mathbf{y}_{1(5\times 1)} \\ \hline \mathbf{y}_{2(3\times 1)} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ \hline 1 & 6 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \beta_{(2\times 1)} + \begin{bmatrix} \mathbf{e}_{1(5\times 1)} \\ \hline \mathbf{e}_{2(3\times 1)} \end{bmatrix},$$

$$Var\begin{bmatrix} \mathbf{e}_1 \\ \hline \mathbf{e}_2 \end{bmatrix} = \sigma^2 \left[\begin{array}{c|c} I_{(5\times 5)} & 0_{(5\times 3)} \\ \hline 0_{(3\times 5)} & I_{(3\times 3)} \end{array}\right]$$

## The BLUE Solution

$$\hat{\mathbf{y}}_2 = X_2\hat{\beta} + \Phi_{21}\Phi_{11}^{-1}\left(\mathbf{y}_1 - X_1\hat{\beta}\right)$$

$$\hat{\beta} = \left(X_1'\Phi_{11}^{-1}X_1\right)^{-1}X_1'\Phi_{11}^{-1}\mathbf{y}_1$$

$$Var\left[\mathbf{y}_2 - \hat{\mathbf{y}}_2\right] = \underbrace{\sigma^2\left(\Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{12}\right)}_{\text{process variance}}$$

$$+ \underbrace{\left(X_2 - \Phi_{21}\Phi_{11}^{-1}X_1\right)Var\left[\hat{\beta}\right]\left(X_2 - \Phi_{21}\Phi_{11}^{-1}X_1\right)'}_{\text{parameter variance}}$$

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

7

## Special Case: $\Phi = I_t$

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

$$\hat{\beta} = \left(X_1'X_1\right)^{-1}X_1'\mathbf{y}_1$$

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

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

8

## Estimator of the Variance Scale

$$\hat{\sigma}^2 = \frac{\left(\mathbf{y}_1 - X_1\hat{\beta}\right)'\Phi_{11}^{-1}\left(\mathbf{y}_1 - X_1\hat{\beta}\right)}{t_1 - k}$$

9

## Remarks on the Linear Model

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

10

## The Problem of Stochastic Regressors

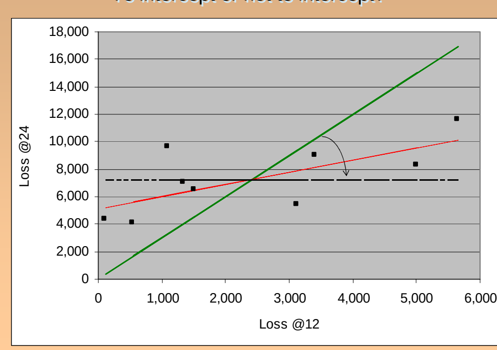
- See Judge [1988] 571ff; Pindyck and Rubinfeld [1998] 178ff.
- If X is stochastic, the BLUE of β may be biased:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'\mathbf{y} \\ &= (X'X)^{-1}X'(X\beta + \mathbf{e}) \\ &= \beta + (X'X)^{-1}X'\mathbf{e} \\ E\left[\hat{\beta}\right] &= \beta + E\left[(X'X)^{-1}X'\mathbf{e}\right] \\ &\neq \beta + E\left[(X'X)^{-1}X'\right]E\left[\mathbf{e}\right] = \beta \end{aligned}$$

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 sportsci.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

$$\boldsymbol{y}_{ij} = a_{ij} x_i f_j r + \boldsymbol{e}_{ij} \qquad \sum_j f_j = 1$$

- $\boldsymbol{y}_{ij}$ 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
- $\boldsymbol{e}_{ij}$ error term of $ij$-cell

17

## Familiar Reserving Methods

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

| | |
|---|---|
| $\boldsymbol{y}_{ij} = (f_j)(x_i r) + \boldsymbol{e}_{ij}$ | quasi Chain Ladder |
| $\boldsymbol{y}_{ij} = (x_i f_j r)(1) + \boldsymbol{e}_{ij}$ | Bornhuetter - Ferguson |
| $\boldsymbol{y}_{ij} = (x_i f_j)(r) + \boldsymbol{e}_{ij}$ | Stanard - Bühlmann |
| $\boldsymbol{y}_{ij} = (x_i)(f_j r) + \boldsymbol{e}_{ij}$ | 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?

$$\ln \boldsymbol{y}_{ij} = \ln x_i + \ln f_j + \ln r + \boldsymbol{e}_{ij}$$

- Barnett and Zehnwirth [1998] favor it.
- Advantages:
  - Allows for skewed distribution of $\ln \boldsymbol{y}_{ij}$.
  - Perhaps easier to see trends
- Disadvantages:
  - Linearity compromised, i.e., $\ln(\mathrm{A}\mathbf{y}) \neq \mathrm{A}\ln(\mathbf{y})$.
  - $\ln(x \leq 0)$ 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: $\hat{\mathbf{y}}_2$ and $Var\left[\mathbf{y}_2 - \hat{\mathbf{y}}_2\right]$
- Interesting quantities are normally linear:
  - AY totals and grand totals
  - Present values
- Powerful theorems (Halliwell [1997] 303f):

$$E\left[\mathrm{A}\hat{\mathbf{y}}_2\right] = \mathrm{A}E\left[\hat{\mathbf{y}}_2\right]$$

$$Var\left[\mathrm{A}\mathbf{y}_2 - \mathrm{A}\hat{\mathbf{y}}_2\right] = \mathrm{A}\,Var\left[\mathbf{y}_2 - \hat{\mathbf{y}}_2\right]\mathrm{A}'$$

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

21

## Transformed Observations

$$\begin{bmatrix} \mathrm{A}\mathbf{y}_1 \\ \hline \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathrm{AX}_1 \\ \hline \mathrm{X}_2 \end{bmatrix}\beta + \begin{bmatrix} \mathrm{A}\mathbf{e}_1 \\ \hline \mathbf{e}_2 \end{bmatrix},$$

$$Var\begin{bmatrix} \mathrm{A}\mathbf{e}_1 \\ \hline \mathbf{e}_2 \end{bmatrix} = \left[\begin{array}{c|c} \mathrm{A}\Sigma_{11}\mathrm{A}' & \mathrm{A}\Sigma_{12} \\ \hline \Sigma_{21}\mathrm{A}' & \Sigma_{22} \end{array}\right]$$

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

22

## Example in Excel

WC Example.lnk

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

Off-diagonal element

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

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

$$= 50 + 10\rho$$

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

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