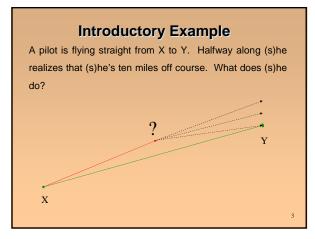
Regression Models and Loss Reserving

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

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

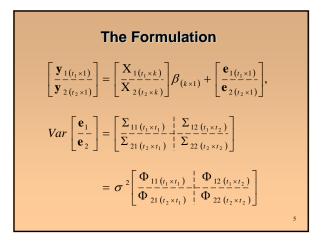
Outline

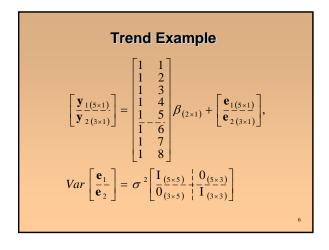
- Introductory Example
- Linear (or Regression) Models
- The Problem of Stochastic Regressors
- Reserving Methods as Linear Models
- Covariance



Linear (Regression) Models

- "Regression toward the mean" coined by Sir Francis Galton (1822-1911).
- The real problem: Finding the <u>Best Linear Unbiased</u> <u>Estimator (BLUE) of vector y₂, vector y₁ observed.</u>
- $y = X\beta + e$. 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.



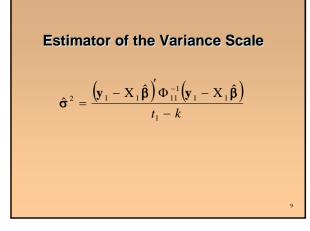


10

The BLUE Solution $\hat{\mathbf{y}}_{2} = X_{2}\hat{\boldsymbol{\beta}} + \Phi_{21}\Phi_{11}^{-1}(\mathbf{y}_{1} - X_{1}\hat{\boldsymbol{\beta}})$ $\hat{\boldsymbol{\beta}} = (X_{1}'\Phi_{11}^{-1}X_{1})^{-1}X_{1}'\Phi_{11}^{-1}\mathbf{y}_{1}$ $Var[\mathbf{y}_{2} - \hat{\mathbf{y}}_{2}] = \sigma^{2}(\Phi_{22} - \Phi_{21}\Phi_{11}^{-1}\Phi_{12})$ $+ (X_{2} - \Phi_{21}\Phi_{11}^{-1}X_{1})'ar[\hat{\boldsymbol{\beta}}](X_{2} - \Phi_{21}\Phi_{11}^{-1}X_{1})'$ $Var[\hat{\boldsymbol{\beta}}] = \sigma^{2}(X_{1}'\Phi_{11}^{-1}X_{1})^{-1}$

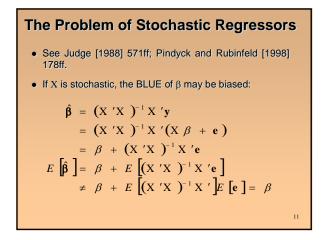
Special Case:
$$\Phi = \mathbf{I}_{t}$$

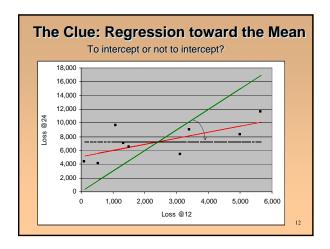
 $\hat{\mathbf{y}}_{2} = \mathbf{X}_{2}\hat{\boldsymbol{\beta}}$
 $\hat{\boldsymbol{\beta}} = (\mathbf{X}_{1}'\mathbf{X}_{1})^{-1}\mathbf{X}_{1}'\mathbf{y}_{1}$
 $Var [\mathbf{y}_{2} - \hat{\mathbf{y}}_{2}] = \sigma^{2}\mathbf{I}_{t_{2}} + \mathbf{X}_{2}Var [\hat{\boldsymbol{\beta}}]\mathbf{X}_{2}'$
 $Var [\hat{\boldsymbol{\beta}}] = \sigma^{2} (\mathbf{X}_{1}'\mathbf{X}_{1})^{-1}$



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, Σ_{11} non-singular.
- Linearity Theorem: $\mathbf{A} \mathbf{y}_2 = \mathbf{A} \hat{\mathbf{y}}_2$
- Model is versatile. My four papers (see References) describe complicated versions.





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.

13

- Intercept should be proportional to exposure.
- Explain the torsion. Leads to a better model?

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."

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!

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.)

The Basic Linear Model

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

- y_{ii} 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

Example Reserving Methods $Y = (X)(\beta) + e$ $y_{ij} = (f_j)(x_i r) + e_{ij} \quad \text{quasi Chain Ladder}$ $y_{ij} = (x_i f_j r)(1) + e_{ij} \quad \text{Bornhuetter - Ferguson}$ $y_{ij} = (x_i f_j)(r) + e_{ij} \quad \text{Stanard - Bühlmann}$ $y_{ij} = (x_i)(f_j r) + e_{ij} \quad \text{Additive}$ $P_{ij} = (x_i)(f_j r) + e_{ij} \quad \text{$

20

Why not Log-Transform?

 $\ln \mathbf{y}_{ii} = \ln x_i + \ln f_i + \ln r + \boldsymbol{e}_{ii}$

- 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., $\ln(Ay) \neq A \, \ln(y).$
 - $\ln(x \le 0)$ undefined.
- Something Better: Simulation with non-normal error terms (robust estimation, Judge [1998], ch. 22)

The Ultimate Question

- Last column of rectangle is ultimate increment.
- There may be no observation in last column:
 - Exogenous information for late parameters f_i or $f_i\beta$.
 - 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.

Linear Transformations

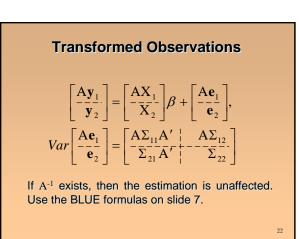
- Results: $\hat{\mathbf{y}}_2$ and $Var \left[\mathbf{y}_2 \hat{\mathbf{y}}_2 \right]$
- Interesting quantities are normally linear:
 AY totals and grand totals
 - Present values
- Powerful theorems (Halliwell [1997] 303f):

$$E[\mathbf{A}\hat{\mathbf{y}}_{2}] = \mathbf{A}E[\hat{\mathbf{y}}_{2}]$$
$$Var[\mathbf{A}\mathbf{y}_{2} - \mathbf{A}\hat{\mathbf{y}}_{2}] = \mathbf{A}Var[\mathbf{y}_{2} - \hat{\mathbf{y}}_{2}]$$

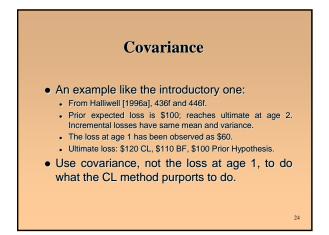
• The present-value matrix is diagonal in the discount factors.

ΙA΄

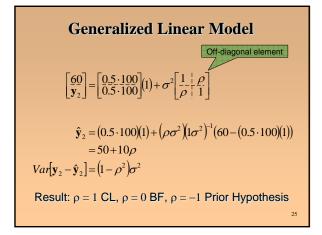
21

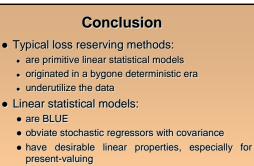






26





- fully utilize the data
- are versatile, of limitless form
- force the actuary to clarify assumptions

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.
 - 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
- 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 LXXV (1998), 807-847.

27