GUY CARPENTER

Refining Reserve Runoff Ranges

GARY G. VENTER

Marsh & McLennan Companies

Reducing Variance of Reserves

Find better fitting models

- Will reduce process variance
- Also improves goodness of fit

Use fewer parameters

- Will reduce parameter variance
- Also reduces MLE parameter penalties

Include exposure information

- Possibly can eliminate some parameters
- May give better fit as well

Goodness of Fit

Can base on variance of runoff

- Parameter variance
- Process variance
- Sum is total variance

Information criteria

- Penalizes likelihood function
- AIC original
- Some other versions will be discussed

Poisson – Constant Severity PCS Model

- Assumes aggregate losses in each incremental cell of development triangle all have a Poisson frequency and the same constant severity b.
- Cell means are modeled as products of row and column parameters – such as estimated ultimate for row and fraction to emerge in column

Poisson – Constant Severity Distribution

- For loss X, X/b is Poisson in λ
- EX = bλ
- Variance = $b^2\lambda$
- Variance/Mean = b
- Also called over-dispersed Poisson: ODP
- But that makes it sound like frequency
- Could be under-dispersed if b < 1
- Really requires all cells to be integer multiples of b

Modeling with Row and Column Factors

- Assume there are n+1 rows and columns in a triangle, numbered 0, … , n
- For accident year w and delay d, denote incremental losses by $q_{w,d}$
- Let U_w be the estimated ultimate for accident year w and g_d be the fraction of losses that go to delay d
- Model: $b\lambda_{d,w} = U_w g_d$

Estimating Parameters by MLE

- Poisson distribution can be used to build likelihood function
- Answer:
  - b is not determinable by MLE – higher b always gives higher likelihood – but other parameters do not depend on b
  - U_w is last diagonal grossed up by chain-ladder development factor
  - g_d is emergence pattern from development factors

PCS Estimate is Chain Ladder Result

- Proven by Hachemeister and Stanard (1975) for Poisson case
- Good write-up in Clark (2003) CAS Forum
- But uses (apparently) more parameters
- Fitted values are $U_w g_d$ so fraction of ultimate for that column (like in BF)
- Basically back down from last diagonal by development factors
- So not chain ladder fitted values

What is Variance of PCS Estimate?

- Process variance for ultimate is just b times ultimate mean using moment-type estimator

$$\widehat{b} = \frac{1}{N-p}\sum_{w,d}\frac{\left(q_{w,d} - U_w g_d\right)^2}{U_w g_d}$$

- N observations, p parameters
- Clark suggests parameter variance from delta method

Delta Method (review of Part 4)

- Information matrix is negative of matrix of 2nd derivatives of log-likelihood wrt all the parameters
- Matrix inverse of information matrix is covariance matrix of parameters
- Variance of function of the parameters takes vector of 1st derivatives of function and right and left multiplies covariance matrix by it

Sample Triangle from Taylor-Ashe

357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
443,160	693,190	991,983	769,488	504,851	470,639				
396,132	937,085	847,498	805,037	705,960					
440,832	847,631	1,131,398	1,063,269						
359,480	1,061,648	1,443,370							
376,686	986,608								
344,014									

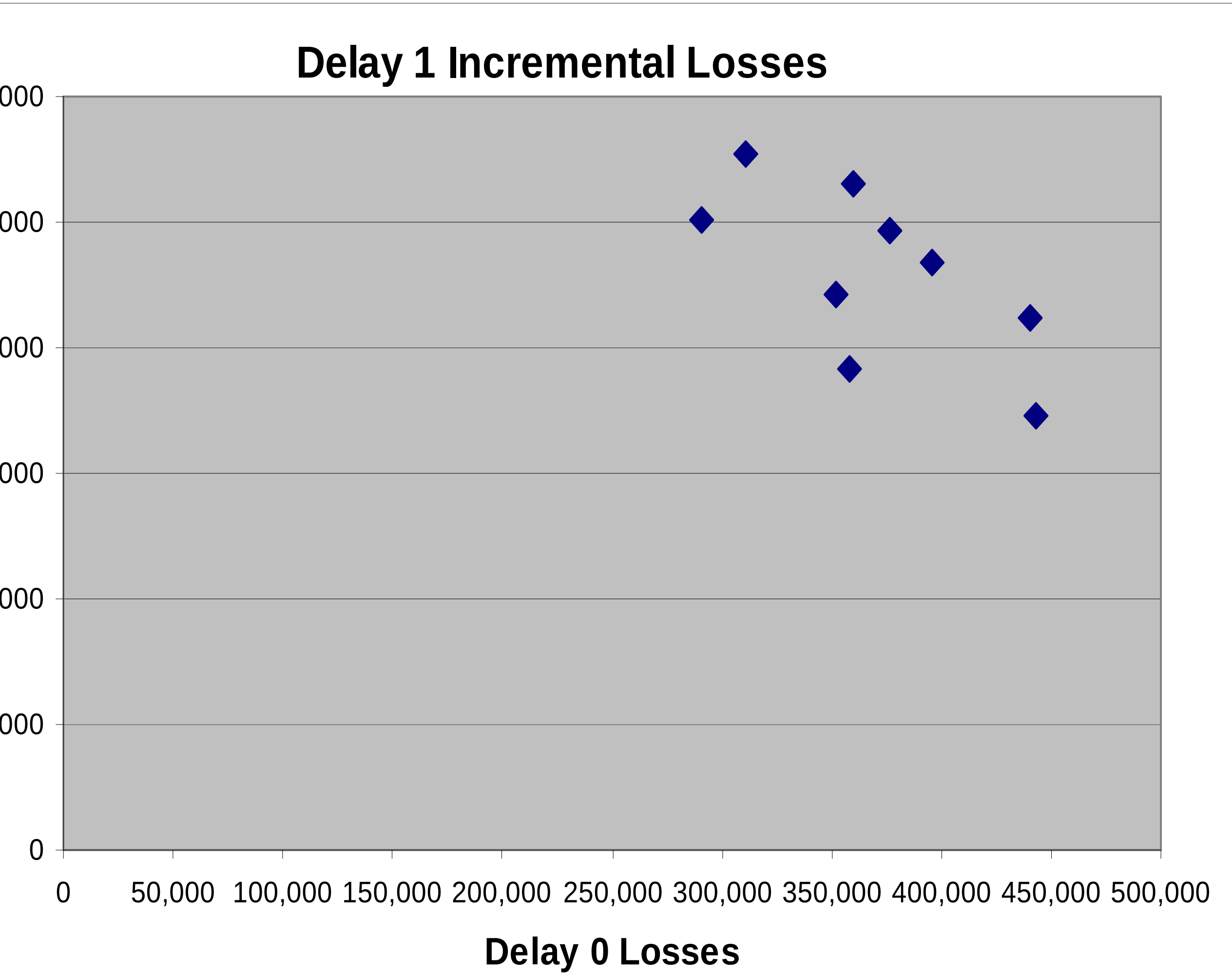
Estimates from Triangle

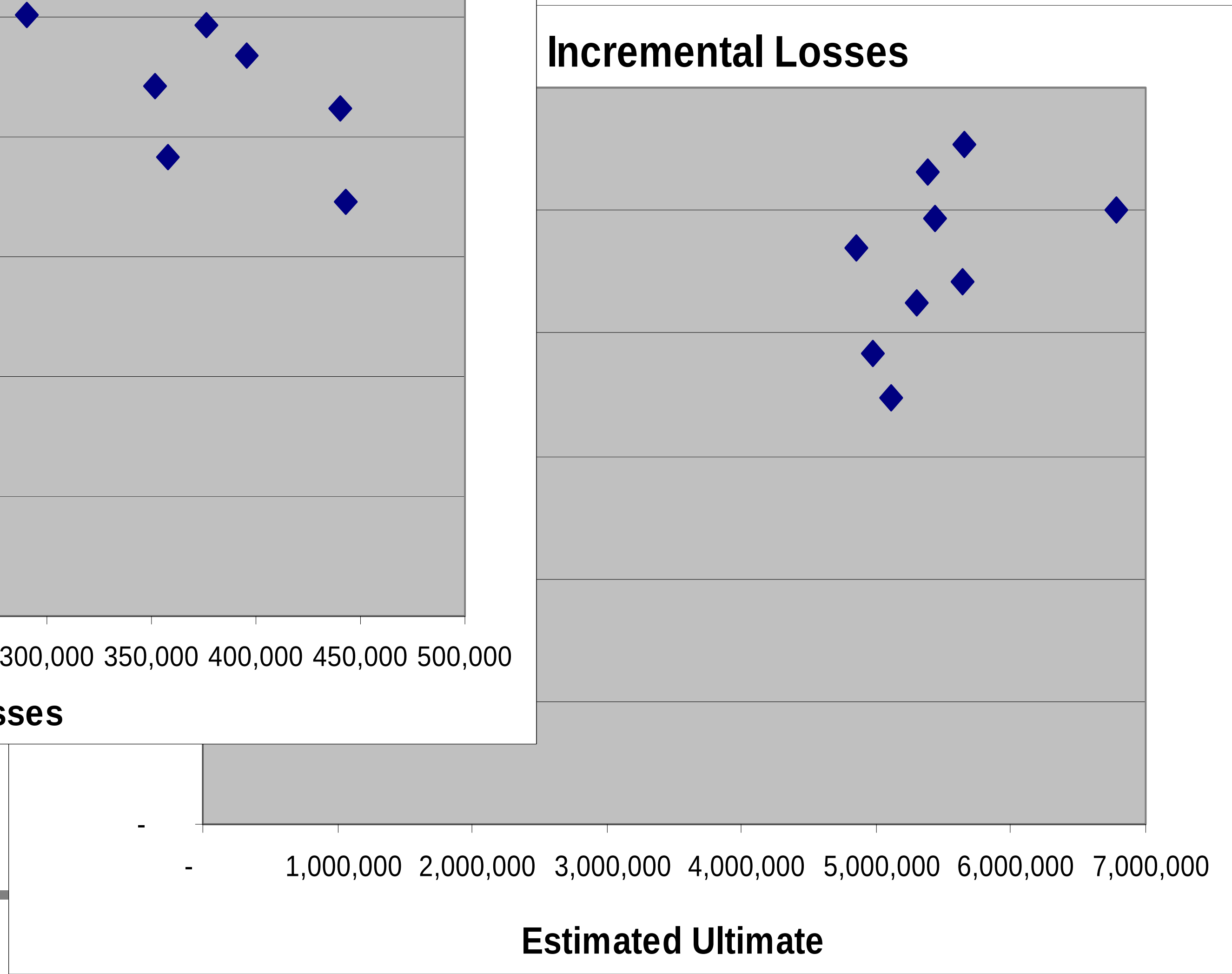
- Reserve by chain ladder or PCS is 18,681,000
- Standard deviation of prediction:

Chain ladder	PCS
2,447,000	2,827,000

- PCS a much better fit – process standard deviation lower by half
- But parameter standard deviation 70% higher

Fitting Example: 0-1 Incremental





Testing Diagonals – Some Are Suspicious

Diagonal	Average Residual	Fraction Positive
0	87,787	1 of 1
1	35,158	1 of 2
2	(76,176)	0 of 3
3	(74,853)	1 of 4
4	100,127	4 of 5
5	(26,379)	2 of 6
6	103,695	5 of 7
7	(115,163)	1 of 8
8	(17,945)	3 of 9
9	38,442	6 of 10

Diagonal Effects in PCS

- **Include factor h_j for j^{th} diagonal**
- **Model: $b\lambda_{d,w} = U_w g_d h_{w+d}$**
- **Iterative MLE estimation of parameters:**

$$U_w = \sum_{d=0}^{n-w} q_{w,d} \Big/ \sum_{d=0}^{n-w} g_d h_{w+d} \qquad g_d = \sum_{w=0}^{n-d} q_{w,d} \Big/ \sum_{w=0}^{n-d} U_w h_{w+d}$$

$$h_j = \sum_{w+d=j} q_{w,d} \Big/ \sum_{w+d=j} U_w g_d$$

But First, How to Compare MLE Fits

- Use information criteria
- Compare loglikelihoods penalized for number of parameters
- AIC – penalize by 1 for each parameter
- HQIC – penalize by ln ln N (N=55→1.388)
- Small sample AIC (AIC_c), per-parameter penalty is N/[N – p – 1]

Loglikelihood with Diagonal Factors

- Tried original model with no diagonals, that plus 7th diagonal, and that plus 6th
- 19, 20 and 21 parameters
- Loglikelihoods: -149.11; -145.92; -145.03
- @ 1.388 ppp 6th and 7th together better than none, but worse than 7th alone
- AIC_c penalty for 20th and 21st parameters is 2.5 and 2.65 so both together worse than none, but 7th better than none

Get Rid of Insignificant Parameters

- Parameters not significantly different from 0 or 1 sometimes defaulted to those values
- Parameters not significantly different from each other can be set equal
- When changes are systematic, a parameter for a year or delay could be set to the average of the parameters before and after it
- More generally, several parameters in a row could be expressed as a linear or geometric trend
- A few additive constants can replace several small factors

Example Model

3 accident year parameters

- Most years at level U_a
- U_0 lower
- U_7 higher
- U_6 average of U_a and U_7

2 payout % parameters – high or low

- Delay 0: low
- 1 – 3: high
- 5 – 8: low
- 4: average of high and low
- 9 = 1 – sum of others

1 diagonal parameter c with all factors either 1, 1+c, or 1–c

- 7th low
- 4th & 6th high
- All others 1.0

Fit of Example Model

Parameter	U_0	U_7	U_a	g_a	g_b	c
Estimate	3,810,000	7,113,775	5,151,180	0.0679	0.1740	0.1985
Std Error	372,849	698,091	220,508	0.0034	0.0056	0.0569

- Likelihood of -146.66 worse than 20 parameter model at -145.92
- But AIC_c penalty lower by 25.5
- HQIC penalty lower by 19.4
- Standard deviation 1,350,000 compared to 2,447,000 of chain ladder and 2,827,000 of original PCS
- Reserve about 5% higher with all models that include diagonal effects

Variance Comparison

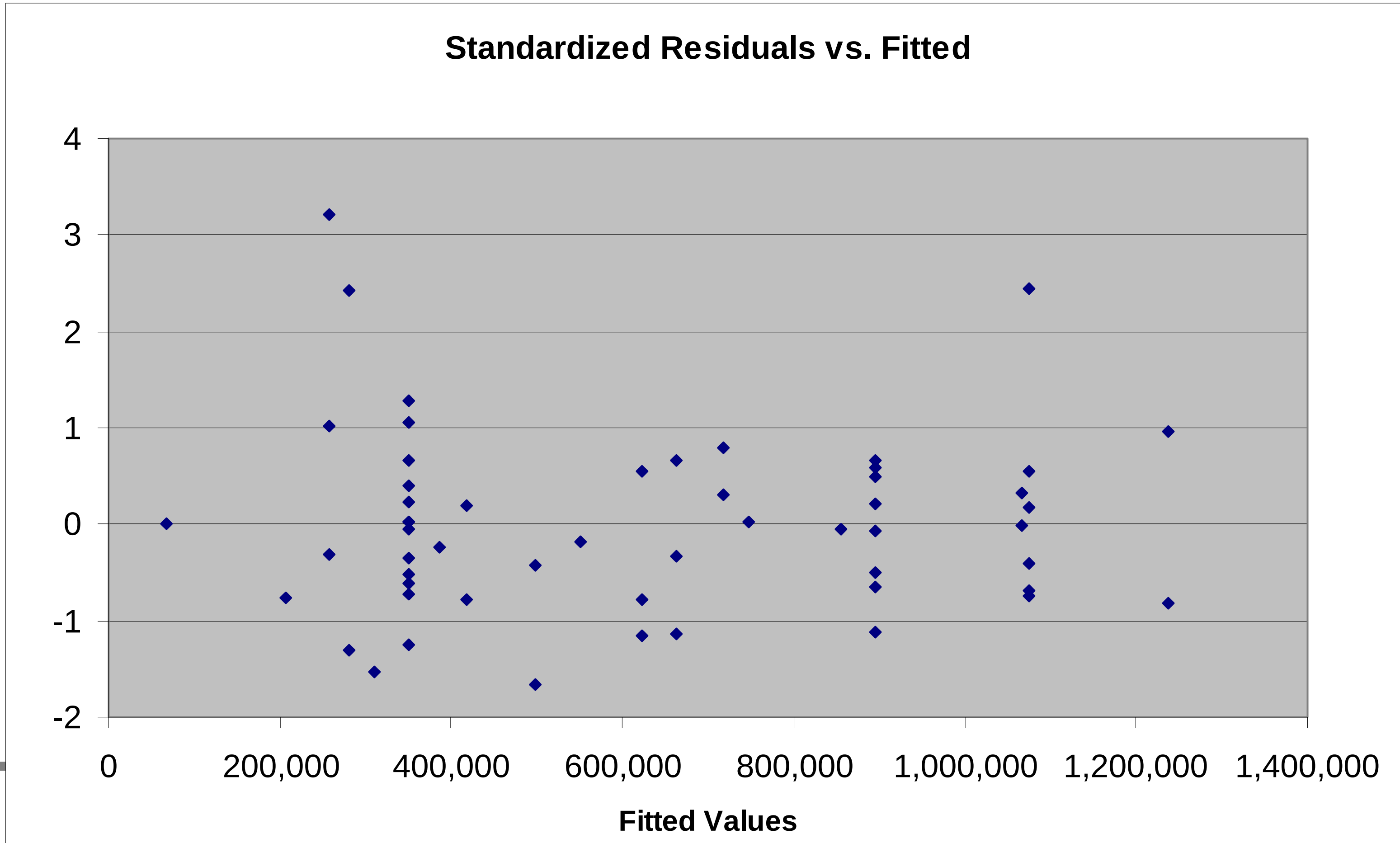
Model....	Original 19 Parameter	6 Parameter
Parameter Variance	7,009,527,908,811	1,103,569,529,544
Process Variance	982,638,439,386	718,924,545,072
Total Variance	7,992,166,348,198	1,822,494,074,616
Parameter Std Dev	2,647,551	1,050,509
Process Std Dev	991,281	847,894
Standard Deviation	2,827,042	1,349,998

Fit better with diagonals; Many fewer parameters

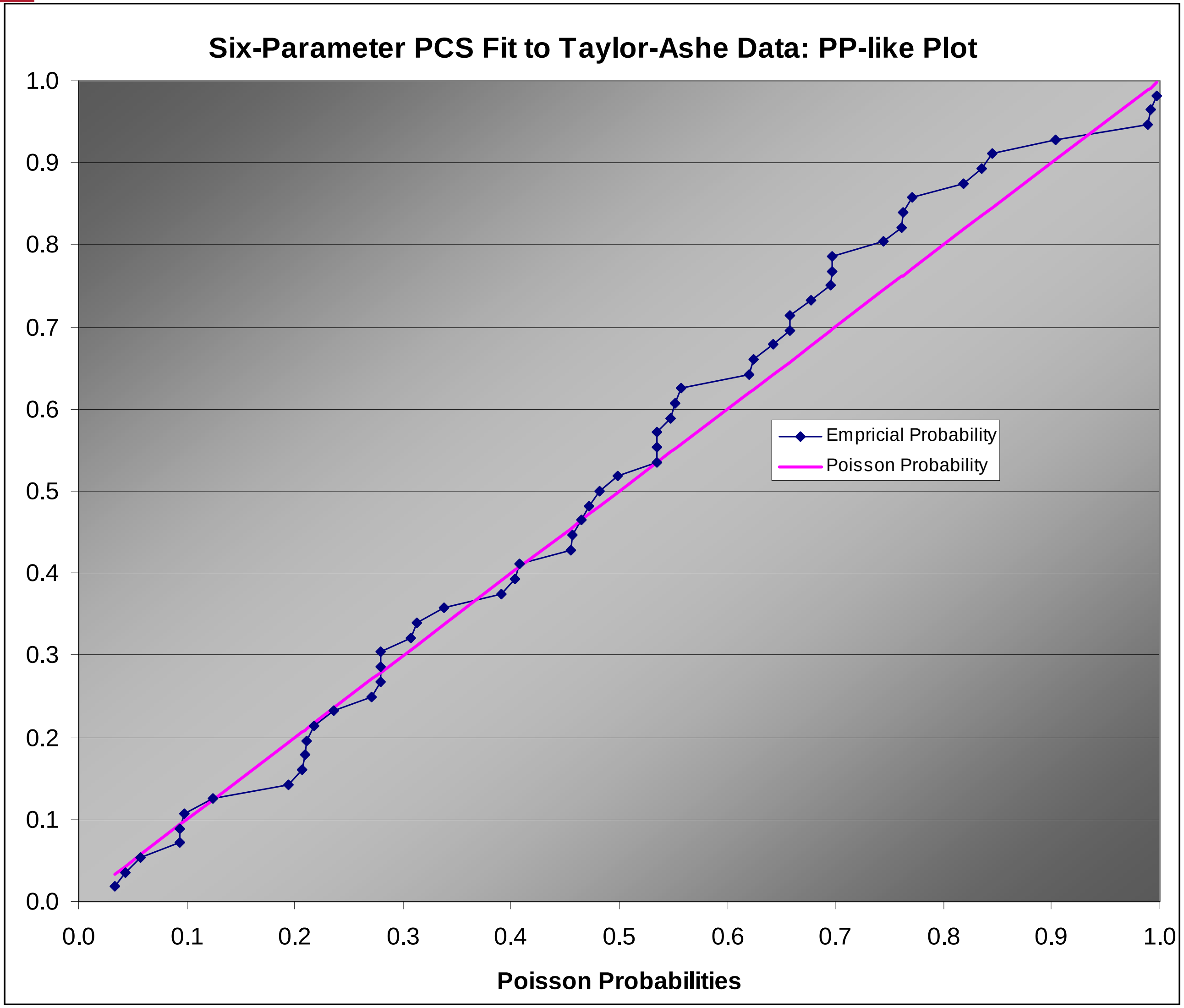
Testing Variance Assumption

- If variance ∝ mean^2, standardized residuals (divided by $b\lambda^{1/2}$) would tend to increase with fitted

Probably ok



Testing Poisson Assumption



Computed Poisson probability of $q_{w,d}/b$ for each point.

Sorted these and gave empirical probability of 1/56 to each.

Comparison reasonable

Issues

- No claim that this is the best model
- Can try other distribution assumptions and compare loglikelihoods
- Ignoring issue of model risk, which is probably large
- But getting better fits with diagonal effects and eliminating unneeded parameters can reduce variance

Modeling Diagonal Effects in Chain Ladder

- Make one big regression, with incrementals all in one column as dependent variable, instead of separate regression for each column
- Each previous cumulative is an independent variable = 0 except for corresponding points of incrementals
- Dummy variables pick out each diagonal
- More variance and independence issues – whole triangle not just each column
- Need heteroscedasticity adjustment to compute variance of runoff
  - Later lags with smaller increments hopefully have smaller variances
  - So you want heteroscedasticity but must adjust for it
  - Paper uses method of Long and Ervin

Example 2 – Long Haul Trucking Triangle

Part of Design Matrix for Example 2

- Diagonals added to chain ladder
- Parameter reduction for late insignificant factors done by adding constant to regression
- Diagonal 3 gets a factor, then 4, 7 and 9 get the same factor 1+c and 10 gets 1 – c
- Process variance a little less than chain ladder due to diagonals
- Parameter variance quite a bit less due to eliminating insignificant parameters
- Normal distribution assumption

Incrment	L0	L1	L2	L3	L4	D3
18,904	11,305	-	-	-	-	-
13,953	8,828	-	-	-	-	-
15,324	8,271	-	-	-	-	-
11,942	7,888	-	-	-	-	1
15,306	8,529	-	-	-	-	-
16,873	10,459	-	-	-	-	-
12,027	8,178	-	-	-	-	-
17,515	10,364	-	-	-	-	-
20,650	11,855	-	-	-	-	-
28,759	17,133	-	-	-	-	-
31,091	19,373	-	-	-	-	-
29,131	18,433	-	-	-	-	-
17,474	-	30,210	-	-	-	-
11,505	-	22,781	-	-	-	-
9,373	-	23,595	-	-	-	1
11,799	-	19,830	-	-	-	-
11,943	-	23,835	-	-	-	-
12,668	-	27,331	-	-	-	-

Example 3

- Looks at when it could help to leave out some data
  - Basically when you can show a change in the process
- Uses probability distribution with 2 parameters to fit emergence pattern (lags)
- Uses exposure information to make all accident years the same level

Conclusions

Fixed effects vs. development factors

- Fixed effects model often fits better
- Usually has more parameters
- Combination can give higher or lower variance
- A lot of potential for reducing # of parameters

Diagonal effects

- Exist in many triangles
- Test for them
- Adjust for them

Reduce parameters

- Set some the same
- Put in trends – linear or nonlinear
- Use additive constants
- Use lag distributions
- Use exposure information
    - Triangles of on-level loss ratios