Residual Distributions

In Loss Triangles

Gary G Venter

Problem

- Standard methods may underestimate the volatility of residuals towards the right side of the triangle, thus understating runoff risk
- Theoretical
  – Aggregate distribution moment relationships likely to change across triangle
- Empirical
  – Supported by distributions in older completed runoff years
- Modeling
  – Models that can address the issue

Aggregate Moments

- X is severity variable, N is frequency, S is aggregate
- ES = EN EX
- Var S = EN VarX + $(EX)^2$VarN
- Usually VarN = aEN, VarX = $b(EX)^2$ for some a and b
- → Var S = $bEN(EX)^2 + a(EX)^2EN = (a+b)(EX)^2EN$
- Often larger claims pay later, so moving right, EX is going up, EN down.
  - EN goes down a lot, EX goes up a bit, EN EX goes down, $EN(EX)^2$ goes down less or even up, so:
  - Var S goes down slower than ES does.
  - If variance were proportional to a power of the mean, the power would be less than 1
- For Work Comp, there are big early payments, but also a lot of small claims that close fairly soon, so even there the later payments can be larger

Empirical Data

- Used completed 10x10 runoff squares of old years in CAS NAIC Triangle Database
- Looking at completed runoff can illustrate features of the runoff process
- Triangles there on net losses, so sometimes have pay pattern distorted by timing of reinsurance recoveries
- Computed incremental payout patterns as incremental losses divided by lag 10 cumulative losses by accident year; same for cumulative
- Looking for how mean and variance of payout percents relate to each other by column – is variance going down more slowly than mean?
- Using companies that had full ten years of history
- Results varied some by line, more by company

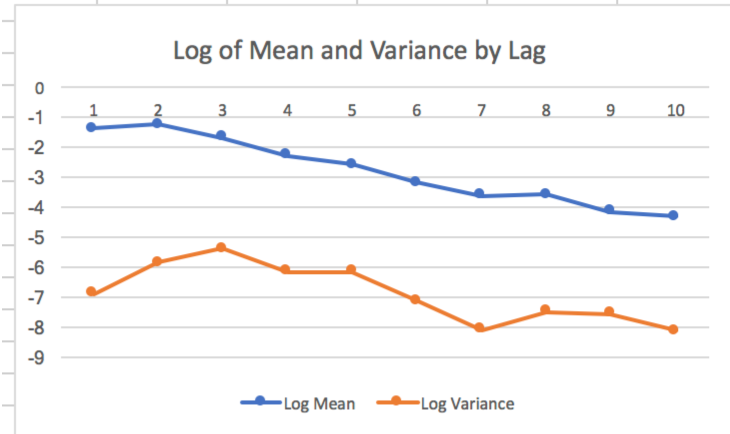
Mean-Variance Relationship Data

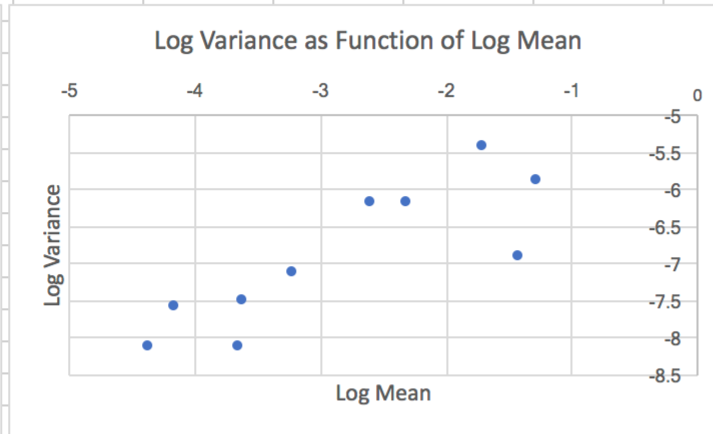
- Graph mean and variance of each column of triangle on log scale
- Graphically illustrates how mean and variance change together in later columns
- Patterns emerge by line

Work Comp Pattern

- Typically variance drops slower than mean at first, same rate as mean later
- That's pattern for about ¾ of the 50 companies that met sample criteria
- Since later payments are typically periodic, they could stabilize to some degree, reducing variance but not mean
- Powers of mean are not always < 1
- Other ¼ of companies have various patterns

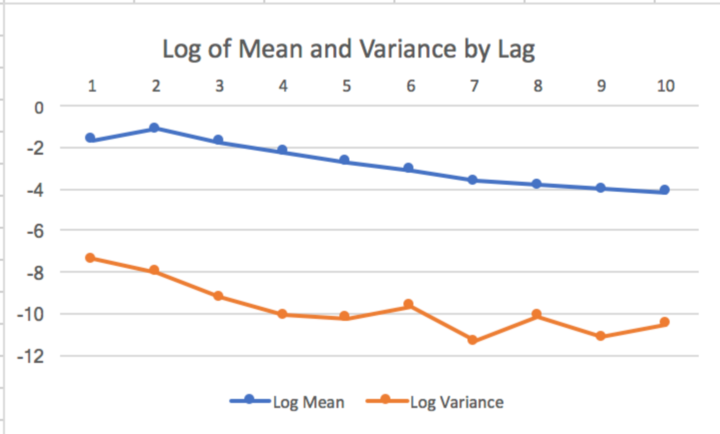
Lumbermens Underwriting Alliance

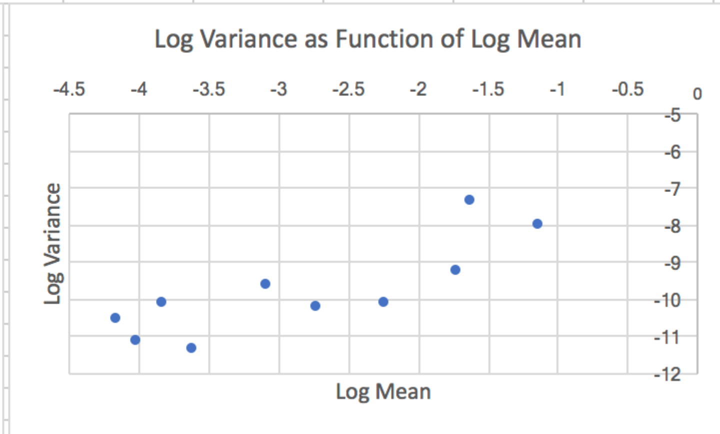




- Both mean and variance generally declining over the lags
- In scatterplot, latest means are lowest so they are smallest values on x-axis
- Slope of scatterplot is regression estimate of p in: Variance = s*mean^p
- Here that is 0.7
- But for first 3 lags, variance is increasing as payments slightly decrease
- Slope for rightmost 3 points in scatterplot is 0.19, vs. 0.98 for other 7 lags
  - In scatterplot, right is higher, which means earlier in triangle
- Later will estimate p by MLE – better and can be different than slope

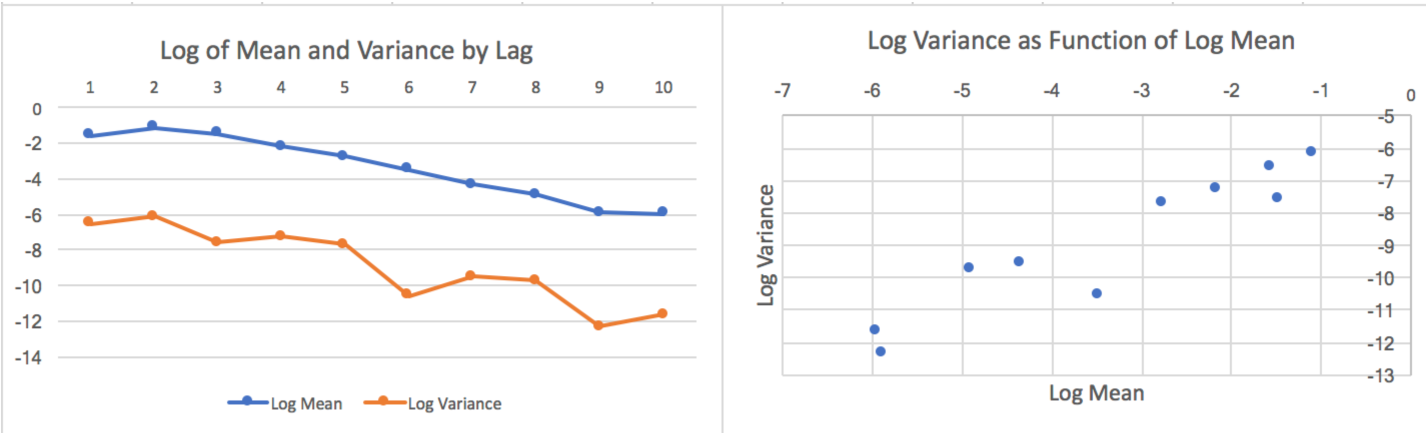
State Farm





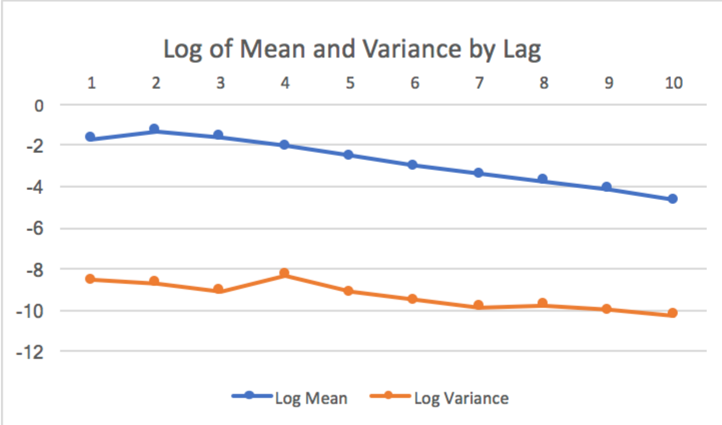
- Variance dropping with mean, or a bit faster, at first
- Levels off later
- Opposite of typical pattern
- For 1st six lags, power is 1.2, for last 4 is -0.85.
- Negative is because variance increasing while mean is decreasing for these points
- Based on right 6 and left 4 points on scatterplot

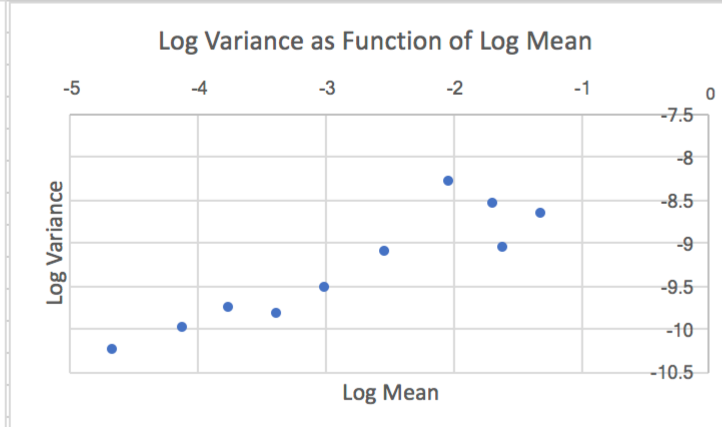
Island Insurance



- Variance generally declining with mean
- Power is 1.11
- Yasuda Fire & Marine looks similar but power is 0.74
- Church Mutual as well, power = 0.79

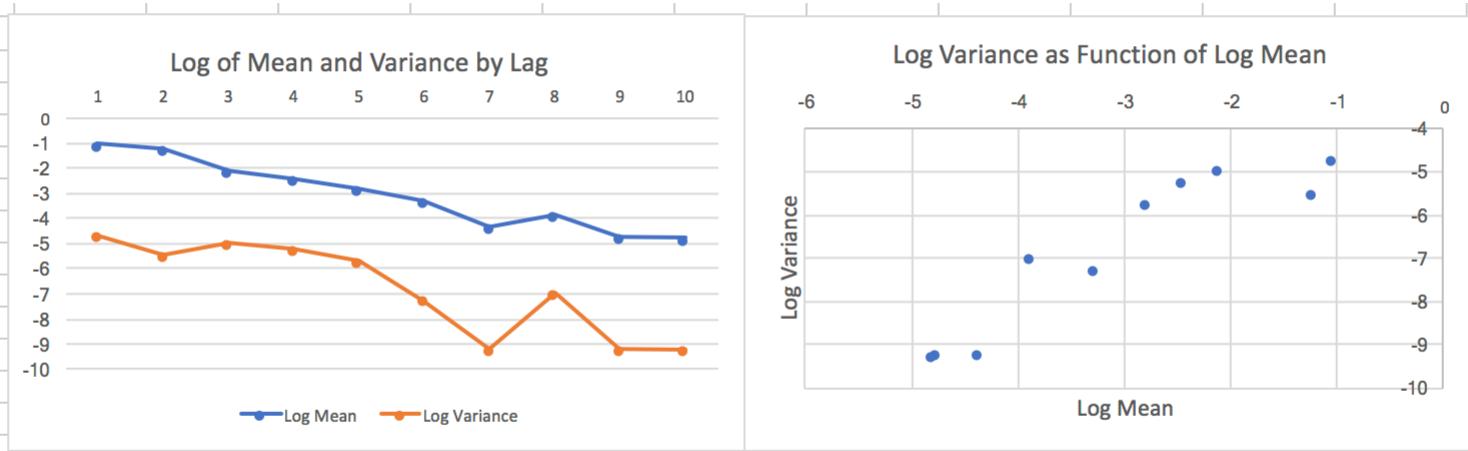
California Casualty





- Typical pattern – variance slightly increases over 1st 4 lags as mean decreases
- Power for 1st 4 lags is -0.6 and is 0.49 for last 4 lags
- Maybe ok to use overall power of 0.53
- Split is supported mainly by fact it is typical but probably not worth modeling in small triangles

Celina Mutual



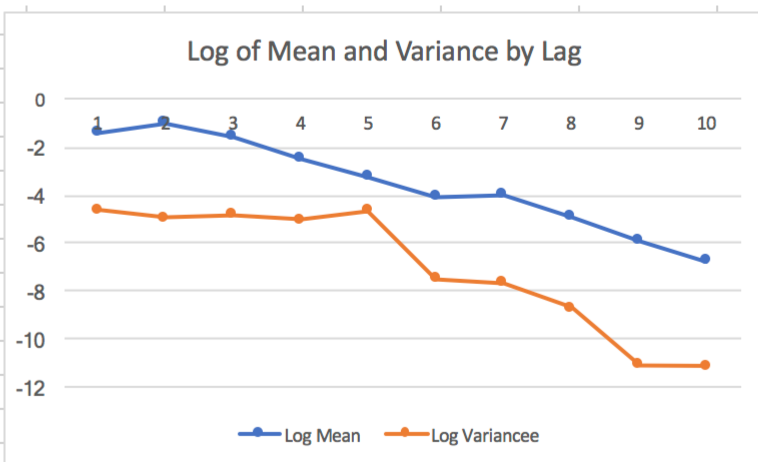
- Typical pattern but steeper
- Power for 1st 5 lags is 0.28, for last 5 it is 2.0

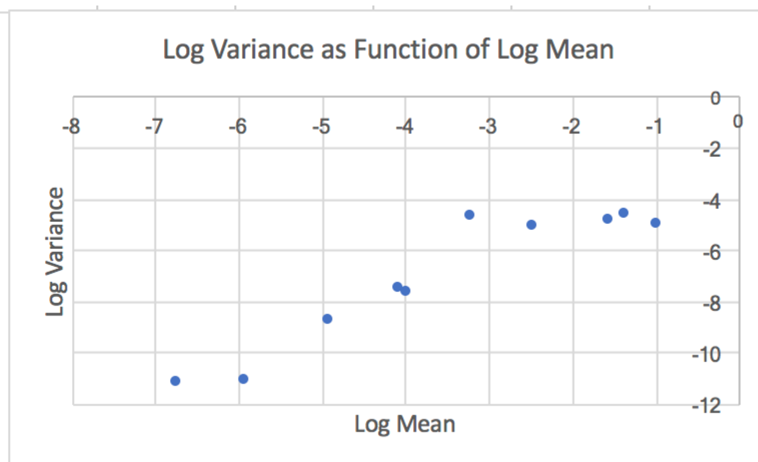
Commercial Auto

- Typical pattern among the 50 companies was variance increasing or declining slowly at first, then dropping sharply
  - A lot of those companies had virtually completed payments by lag 8
- 2nd most common pattern was gradual decline in variance at near constant power of mean, usually in range 0.7 to 1.4
  - Such companies tended to have continuing payments through lag 10

Island Insurance

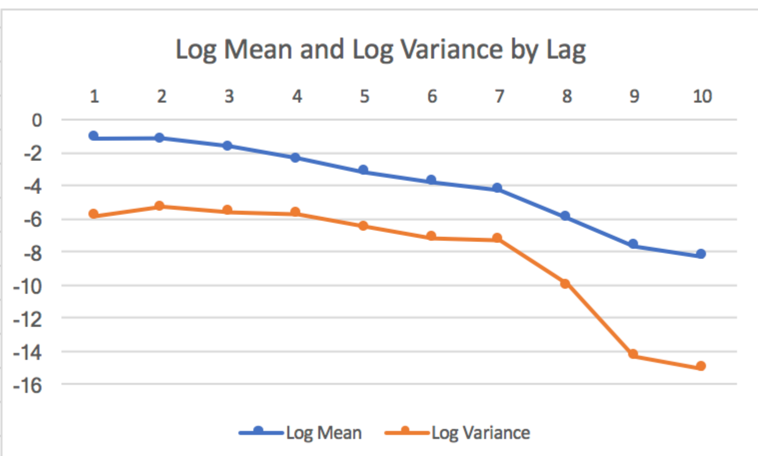
- Typical pattern
- Variance increasing slightly at first then falling off
- Power of mean -0.04 for first 5 lags 1.6 for last 5
- Protective Insurance, THE Insurance, Celina Mutual very similar

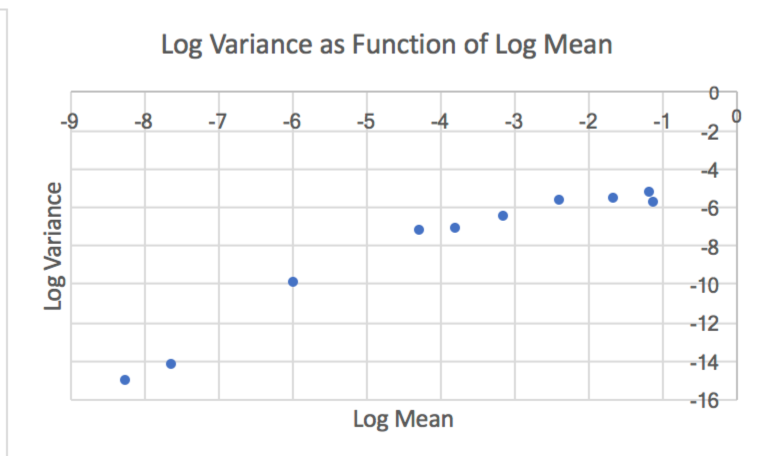




Grange Insurance

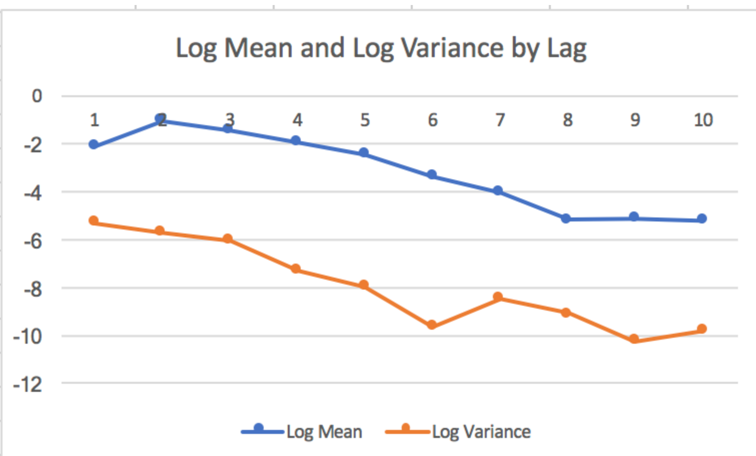
- Version of typical pattern with variance slowly decreasing then quickly decreasing
- For first 7 lags power is 0.77, then 2.3 for last 3 lags

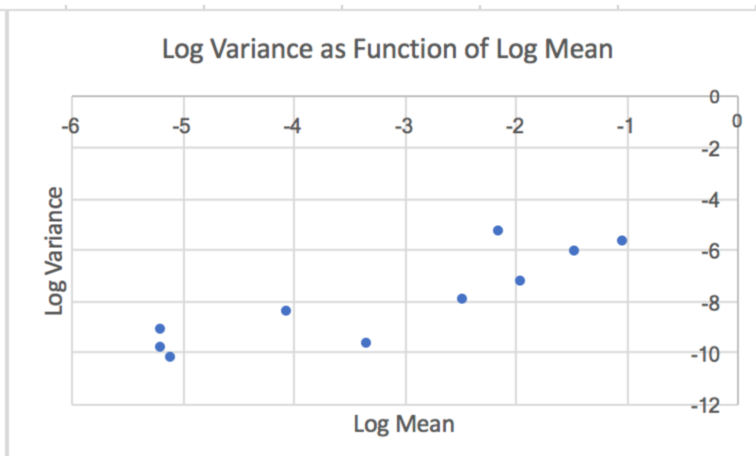




USAA

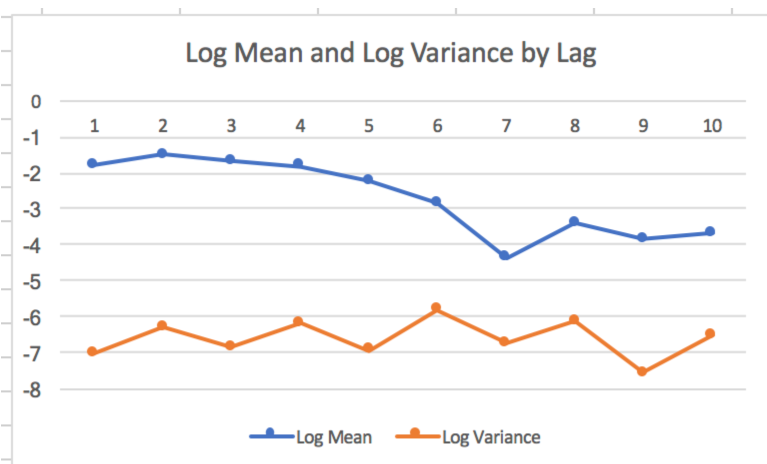
- Example of gradual steady decline, with power = 0.98. Others:
  - Federal Insurance, power = 1.36
  - State Farm, power = 0.71
  - Erie Insurance Exchange, power = 0.98
  - Florida Farm Bureau, power = 1.05
  - Grinnell, power = 1.10
  - Lumber Insurance, power = 0.98
  - Eveready Insurance, power = 0.70
  - Federated Rural Electric, power = 1.2
  - Brotherhood Mutual, power = 1.11
  - Interboro Mutual, power = 1.2

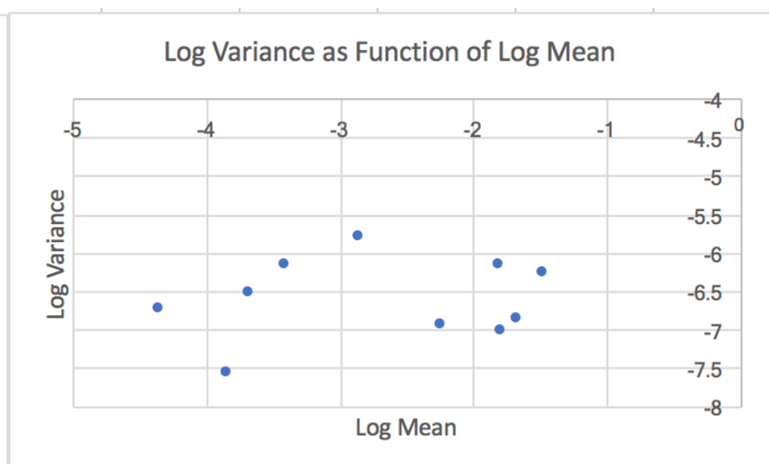




Philadelphia IND INS

- One of various other patterns
- Payout pattern declines slowly
- Variance basically flat
- Power of 0.08
- Mean/variance like normal distribution – however this one is fairly skewed





Other Liability

- This showed a lot of the same patterns as Commercial Auto
- However drop in variance was less at late lags, with generally longer payout pattern

Products Liability

- Some companies had pattern like Commercial Auto, Other Liability
- Quite a few had power around 2
- Possible influence of claims made coverage
- Possible higher volatility frequency distribution

How to Model This?

- Will assume model is one of those that specifies a density function for each observation, as in MLE estimation, Bayesian estimation, GLM etc.
- Usually model specifies mean as a function of the parameters, then specifies a density function, like lognormal, for the residuals, with one parameter for each cell (like μ_j) determined by the mean of the cell, then with all the other parameters of the density (like σ) constant for all cells, to be estimated.
- But more flexible if you let σ_j vary across the triangle as well.
- Tweedie distribution has parameters μ_j, λ and p with variance = $\lambda\mu_j^p$. Often make λ and p constant, but we will consider varying λ_j across the cells.

Exponential Family

- Family of distributions defined by form of density: parameters appear only in exponential function with a restricted interaction of parameters and data
- That form leads to simplified methods for MLE
  - Useful before modern computers – from 1950s
- Normal, Poisson, Gamma, Negative Binomial, Tweedie all in it
- But only some restricted forms of these distributions are. E.g., reparametrizing NB gives a useful form not in it.
- For all of these but NB, variance = $\theta\mu^p$ for a certain p.
- For normal, Poisson, common form of Tweedie, Gamma, p = 0, 1, $1 < p < 2$, 2.
- For the Tweedie, p is a parameter and can be > 1
- None has $0 < p < 1$

In Application

- p would be the same for every cell of a fitted model by choice of distribution or Tweedie p.
- If all the other parameters are constant across the triangle, then variance = $\theta\mu^p$ for all the cells.
- But if θ varies by cell there might not be any constant mean-variance relationship across the triangle, or it could be some other relationship depending on how θ varies.
- In exponential family, if variance = $\theta\mu^p$, then skewness / CV = p.
- This happens whether or not θ is constant across the cells and still holds if density reparameterized.

Variance a Lower Power of Mean

- Say you are ok with normal skewness of zero but want the variance proportional to the 0.7 power of the mean. Reparameterize normal to do this.
- Set μ_j= f(parameters) and $\sigma_j^2 = s\mu_j^{0.7}$, with s a parameter, constant for all observations j, to be estimated by MLE, Bayes estimation, etc.
  – Instead of σ you are estimating s to be constant across the cells.
- Or with one more parameter k, you can set $\sigma_j^2 = s\mu_j^k$, and estimate the power as part of the model. Usual regression just assumes k = 0.

What about Other Distributions?

- Now let μ and σ denote the mean and standard deviation of a cell, not the normal parameters
- Denote the parameters for the j^{th} cell by a_j, b_j.
- Suppose you want to model μ_j = f(parameters) and $\sigma_j^2 = s\mu_j^k$. Say you want skewness / CV = 2.
- Then use the gamma, $\mu_j = a_j b_j$ and $\sigma_j^2 = a_j b_j^2$.
- Solve then for: $a_j = \mu_j^2/\sigma_j^2$ and $b_j = \sigma_j^2/\mu_j$, so in terms of the cell moments, $a_j = \mu_j^{2-k}/s$ and $b_j = s\mu_j^{k-1}$
- This makes each cell gamma distributed with mean and variance functions of the parameters, and makes the variance proportional to μ_j^k across all the cells. Still $\sigma_j^2 = b_j\mu_j^2$ in every cell.

More Skewed Distributions

- For an Inverse Gaussian with $\sigma_j^2 = \mu_j^3/a_j$ take $a_j = \mu_j^3/\sigma_j^2 = \mu_j^{3-k}/s$. Then $\sigma_j^2 = s\mu_j^k$. Skew / CV = 3.
- For lognormal parameterized by $\mu_j = \exp(b_j + ½a_j^2)$
  - $CV^2 = \sigma_j^2/\mu_j^2 = \exp(a_j^2) - 1$. Then:
  - $a_j^2 = \log(1+\sigma_j^2/\mu_j^2)$ and $b_j = \log(\mu_j) - ½\log(1+\sigma_j^2/\mu_j^2)$
  - $a_j^2 = \log(1+s\mu_j^{k-2})$ and $b_j = \log(\mu_j) - ½\log(1+s\mu_j^{k-2})$
  - Skew / CV > 3
- For Tweedie with $\sigma_j^2 = \lambda_j\mu_j^p$, let $\lambda_j = s\mu_j^{k-p}$
  - Every cell is then Tweedie with $\sigma_j^2 = \lambda_j\mu_j^p = s\mu_j^k$
  - There is no single Tweedie across the triangle anyway, as each cell has at least a different μ_j

What Distribution to Use?

- Can try a few and check likelihood function
- With variance modeled, choose distribution based on other shape characteristics
- Almost all companies and lines in the sample had skewness/CV < 2.5

Distribution	Skewness / CV
Normal	0
Poisson, ODP	1
Negative binomial	p = 2 – mean/variance; 2 > p >1
Tweedie	2 > p > 1; variance ~ $mean^p$
Gamma	2
Inverse Gaussian	3
Lognormal	$3+CV^2$
Inverse Gamma	$4/(1–CV^2)$ for CV<1 and infinite otherwise

GIG Distribution

- Gaussian - Inverse Gaussian weighted average. (Allows negative observations)
- Give both distributions variance = $s\mu^k$, but have one more parameter, a, for percent Gaussian
- Then skewness/CV can range from 0 to 3
- Vs. choice of distribution to get right shape, using an implicit parameter, plus usually one more for the scale
- GIG with 3 parameters for shape instead of 2 can get realistic variance-mean relationship and tail shape without having to try various distributions
- Often have a lot of parameters already for cell means (like row, column factors) so one more not a big deal
- Normal-gamma or normal-lognormal possible instead
- Used single k for entire triangle but could split

Completed Δ Fits – GIG MLE

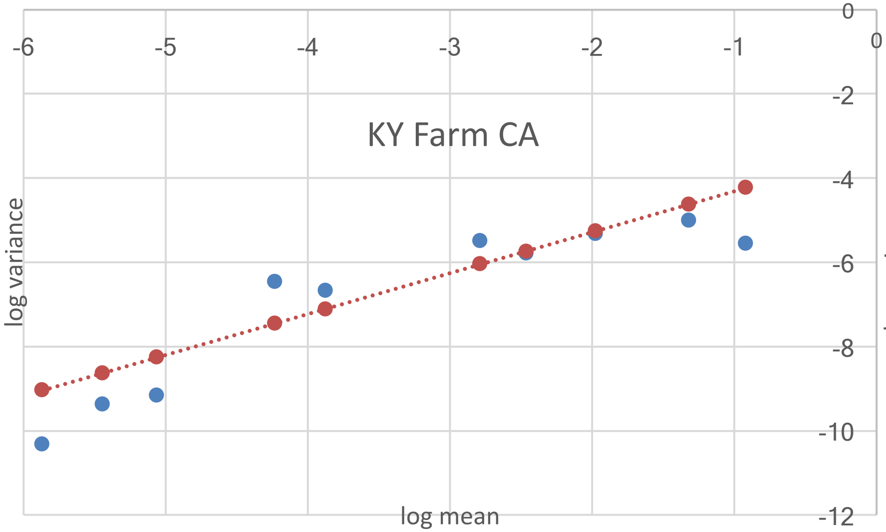
Shape like normal but variance decreasing slowly as mean decreases

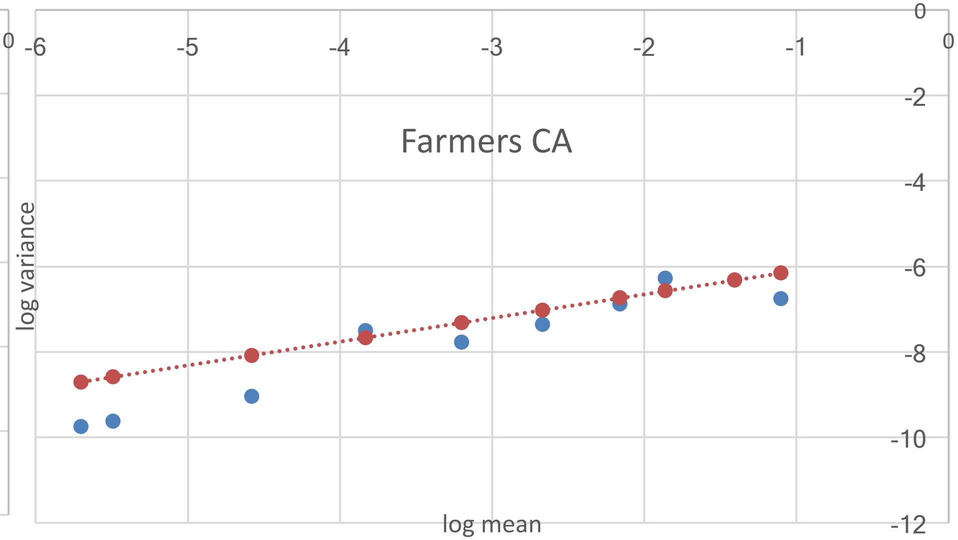
- Also assumed a cell normal if it or the column mean was zero or negative
- CA = commercial auto
- PR = products liability
- MLE pretty easy in R
- MLE better way to estimate power – can make a difference
- Some strange data won't converge for this distribution – e.g., one with 73% of payments at lag 10

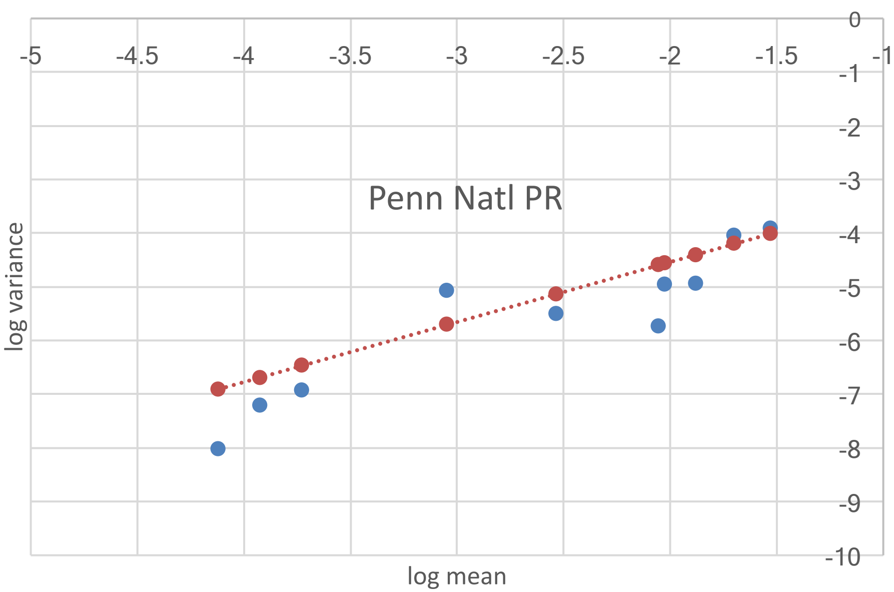
Company/Line	Power	% Normal
State Farm CA	0.750	100%
Farmers CA	0.553	21%
KY Farm CA	0.972	100%
Penn Natl PR	1.118	0%
Federal PR	0.591	100%
Employers PR	1.168	19%

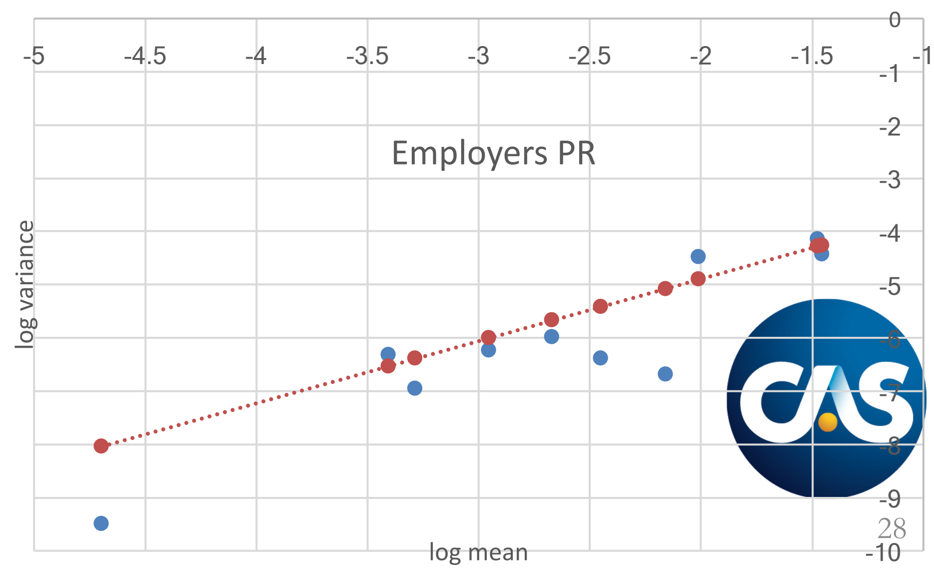
Fairly highly skewed so all IG

Actual and Fitted Log Variance

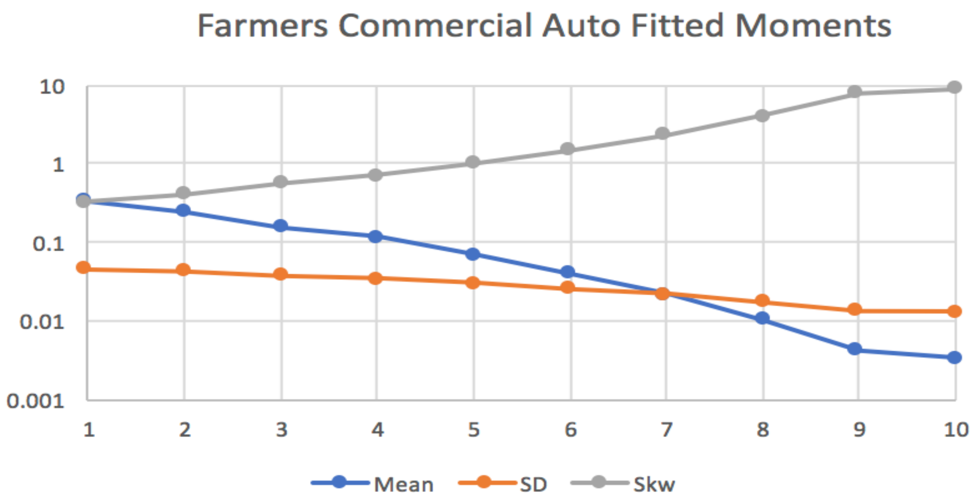


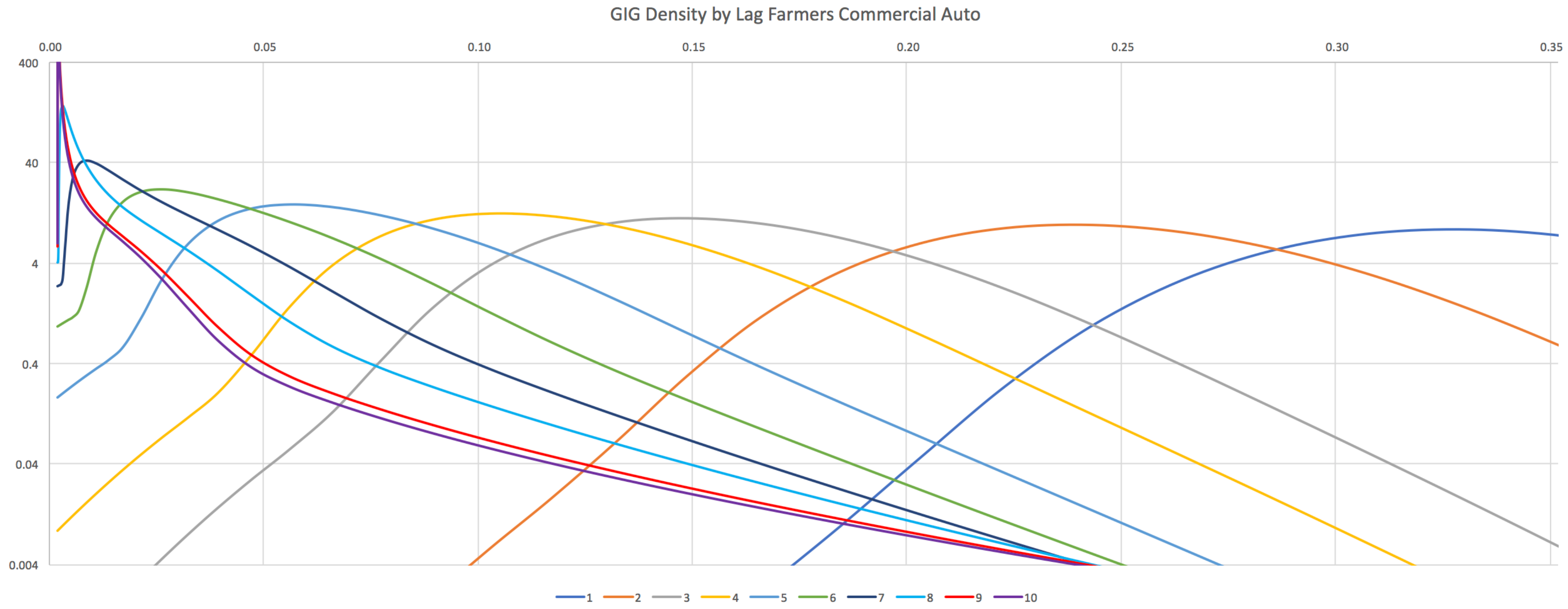






Farmers CA Fitted Distributions





GIG MLE for 10x10 Square in R

```
library(tweedie) # not needed here – used for Tweedie MLE which is very similar
library(optimx)
library(statmod)
y = read.table('farmers_ca.txt', header = FALSE)  # 10 x 10 txt file of payouts – each row sums to 1
nll.gig = function(v) {# NLL function for GIG
s = v[1]                 # scale, want > 0 but usually comes out that way without constraint
k = v[2]                 # power, can be any real number
a = 1/(1+exp(v[3]))   # fraction normal ; v[3] can be any real but 0 < a < 1; easier to optimize that way
mu = colMeans(y)    # could estimate means by MLE but here jut taking column means
sd = s*abs(mu)^k     # absolute value in case mu is negative
ll = 0
for (j in 1:10) {    for (i in 1:10) {
if (y[i,j] > 0 & mu[j] > 0)
ll = ll + log(a*dnorm(y[i,j], mu[j], sd[j], log = FALSE) + (1-a)*dinvgauss(y[i,j], mu[j], mu[j]^3/sd[j]^2))
 else ll = ll + dnorm(y[i,j], mu[j], sd[j], log = TRUE)}}   # zero or neg values get normal dist
  -ll
}
ans = optimx(c(.01, 0.3, -8), nll.gig, method = "Nelder-Mead", control = list(parscale = c(1, 10, 100)) )
#optimx is R function that consolidates several optimization methods; better if scaled
```

Inverse Gaussian vs. Gamma

- Skewness indicative of right tail shape but negative moments, like $E(X^{-2})$ tell about the left tail
- Gamma can have some negative moments infinite , which means that probability concentrated near zero
- Density graph asymptotic to y-axis
- Inverse Gaussian has all moments existing, so near x=0, density is asymptotic to x-axis
- Lognormal similar and more skewed, so a possible alternative
- Usually parameterization of inverse Gaussian has variance = μ^3 / λ. λ not a scale parameter – but alternative with scale parameter exists

Inverse Gaussian Alternative

- Scale parameter b, with mean = ab and variance = ab^2.
- Density is: $f(x) = (2\pi)^{-\frac{1}{2}} \frac{a}{b} \left(\frac{x}{b}\right)^{-\frac{3}{2}} exp\left(-\frac{\left({}^{x}/_{b}-a\right)^2}{2a^2\left({}^{x}/_{b}\right)}\right)$
- The CDF uses the standard normal Φ:
- $F(x) = \Phi\left(a\sqrt{\frac{b}{x}}\left[\frac{x}{ab}-1\right]\right) + e^{2a}\Phi\left(-a\sqrt{\frac{b}{x}}\left[\frac{x}{ab}+1\right]\right)$
- To simulate x, take two random draws and set:
- y = ½normsinv(rand()$_1$)^2 and z = 1/rand()$_2$
- w = $b\left(a + y - \sqrt{y(2a+y)}\right)$
- If w > ab(z – 1) then x = $(ab)^2$/w, otherwise x = w.
- CV is $a^{-½}$ and skewness is 3CV.
- Sum of IG variates all with same b is IG in b and sum of a's.
- Can simulate sum of claims with one IG draw, like gamma.
- Can do that with standard parameterization, but the IG parameters for the sum very awkward – see Wikipedia for how bad this can be

Weibull Distribution

- $f(x) = \frac{\tau}{x}\left(\frac{x}{\theta}\right)^{\tau} e^{-\left({}^{x}/_{\theta}\right)^{\tau}}$
- $E(X) = \theta\Gamma(1 + {}^{1}/_{\tau})$. $1 + CV^2 = \frac{\Gamma(1+{}^{2}/_{\tau})}{\Gamma(1+{}^{1}/_{\tau})^2}$
- Skewness < 0 for τ > 3 or so, ⇑ as τ ⇓
- Usually no single τ works for whole triangle
- But if make variance = $s\mu^k$, get changing τ
- Skewness can then vary across the triangle like it does for GIG
- Need all observations > 0

Problem is all those gammas

- Makes it harder to force variance = $s\mu^k$
- But then $CV^2 = s\mu^{k-2}$, so solve for τ using R function uniroot inside of fitting program from

$$1 + CV^2 = \frac{\Gamma(1+{}^2/_\tau)}{\Gamma(1+{}^1/_\tau)^2}$$

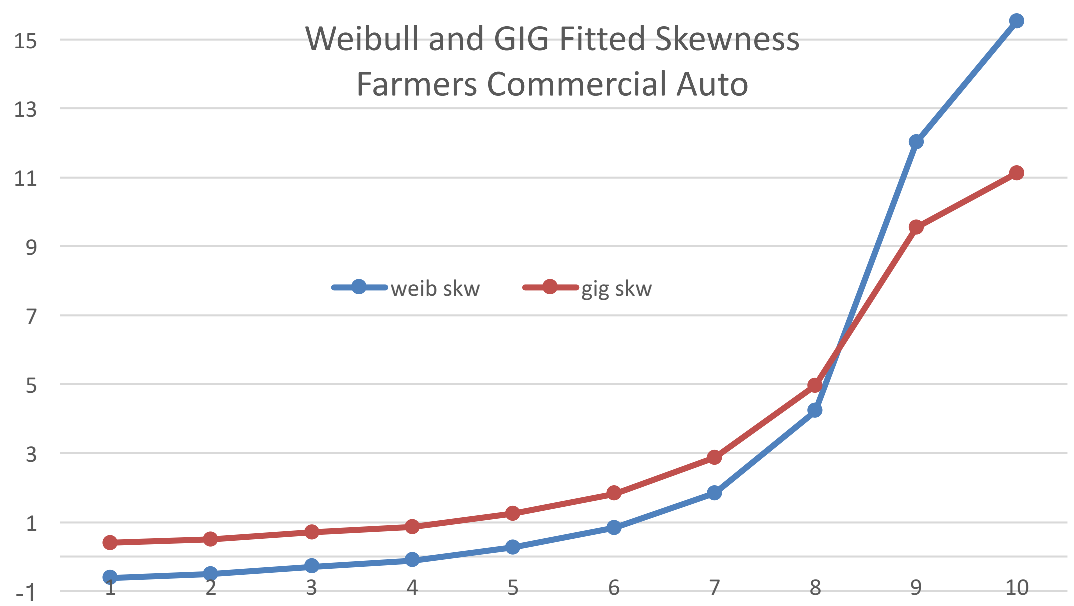
```
weib = function (t, u) gamma(1+2/t)/gamma(1+1/t)^2 – 1 – u
tau = as.numeric(uniroot(weib, c(0.1, 10000), u=s*m^(k–2))[1])
```

Tried for Farmers CA

- Only 2 very slight negatives and 2 zeros in triangle – made them all 0.000001 for e.g.
- Fit actually better than GIG by NLL
- Power 0.503 instead of 0.553 but mean and variances by column all very close
- Skewnesses close to GIG fit but lower at right side and higher on left of triangle
- Good alternative when all positive

Weibull Fit

	1	2	3	4	5	6	7	8	9	10
τ	9.71	7.62	5.27	4.12	2.71	1.74	1.05	0.63	0.39	0.36
θ	0.350	0.261	0.169	0.127	0.078	0.046	0.022	0.0072	0.0012	0.00074
weib skw	-0.62	-0.51	-0.29	-0.11	0.27	0.83	1.85	4.22	12.02	15.54
gig skw	0.40	0.50	0.69	0.86	1.24	1.82	2.88	4.95	9.56	11.13



Incremental or Cumulative?

- Build model of incremental or cumulative losses?
- Want residuals to be independent
- Cumulative might be positively correlated
- Incremental might be negatively correlated due to catch up after slow period

Testing Independence of Residuals

- Dividing by row totals effectively models loss levels by AY – row total is the parameter
- Results show remaining payout pattern differences among accident years – so are proxies for residuals
- Some models try to adjust for trends in payout patterns over time
- So looked at correlations between adjacent columns cumulative and incremental with and without detrending each column

State Farm – Fairly Typical

		1 to 2	2 to 3	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8	8 to 9	9 to 10	Average
Cumulative	Correlation detrended	56%	59%	87%	41%	54%	91%	77%	65%		66%
	Correlation	88%	88%	86%	67%	85%	97%	90%	75%		84%
Incremental	Correlation detrended	-22%	-13%	23%	-34%	10%	-16%	-7%	0%	-40%	-11%
	Correlation	-17%	-11%	29%	-17%	10%	-14%	-3%	31%	-1%	1%

- In 10 x 10 triangle can have spurious correlations
- Average correlation probably more reasonable
- Cumulative showing a lot more correlation
- Most companies look like this
- Conclusion: model incremental losses
- But if you do model cumulative, t-test for significance of development factors should be difference from 1, not 0

Conclusions

- Model incremental losses
- Variance often decreases slower than mean does for smaller cells
- Can model as variance proportional to power less than 1 of mean
- Can do that with any distribution by setting the parameters appropriately
- GIG does that and matches other shape characteristics like higher moments as well
- Weibull good too when you can control variance – mean relationship