Predictive Modeling
Where have we been?
Where are we going?

A Personal View

by

Glenn Meyers
ISO Innovative Analytics

CAGNY Meeting

May 29, 2008

Predictive Modeling is Not New!

- Traditional actuarial responsibility
- Predict the losses per unit of exposure for next year
- Involves trending, loss development and credibility

A CAS Midlife Example

- X_{ij} = Loss per unit of exposure
  - Construction class i
  - Protection class j
- Model $X_{ij} = a_i + b_j$
- Choose a_i and b_j so that

$$\sum_i \sum_j n_{ij} \left(X_{ij} - a_i - b_j \right)^2$$

is minimized.

Look at the Normal Equations

$$\frac{\partial\left(\sum_i\sum_j n_{ij}\left(X_{ij}-a_i-b_j\right)^2\right)}{a_i}=0\Rightarrow\sum_j n_{ij}X_{ij}=\sum_j n_{ij}\left(a_i+b_j\right)\ \forall i$$

$$\frac{\partial\left(\sum_i\sum_j n_{ij}\left(X_{ij}-a_i-b_j\right)^2\right)}{b_j}=0\Rightarrow\sum_i n_{ij}X_{ij}=\sum_i n_{ij}\left(a_i+b_j\right)\ \forall j$$

- "Unbiased in the Aggregate" - From Bailey "Insurance Rates with Minimum Bias" (PCAS 1963)
- Bailey solves for the a_i's and b_j's iteratively
- SAS Proc GLM (70's) solves with matrix algebra

Introduce "What's New" with an Example

- X ~ lognormal with $\mu = 5$ and $\sigma = 2$
- Two ways to estimate E[X] (= 1,097)
- Straight Average – $\hat{E}_N\left[X\right] = \frac{1}{n}\sum_{i=1}^{n} X_i$
- Lognormal Average – $\hat{E}_L\left[X\right] = e^{\hat{\mu}+\hat{\sigma}^2/2}$

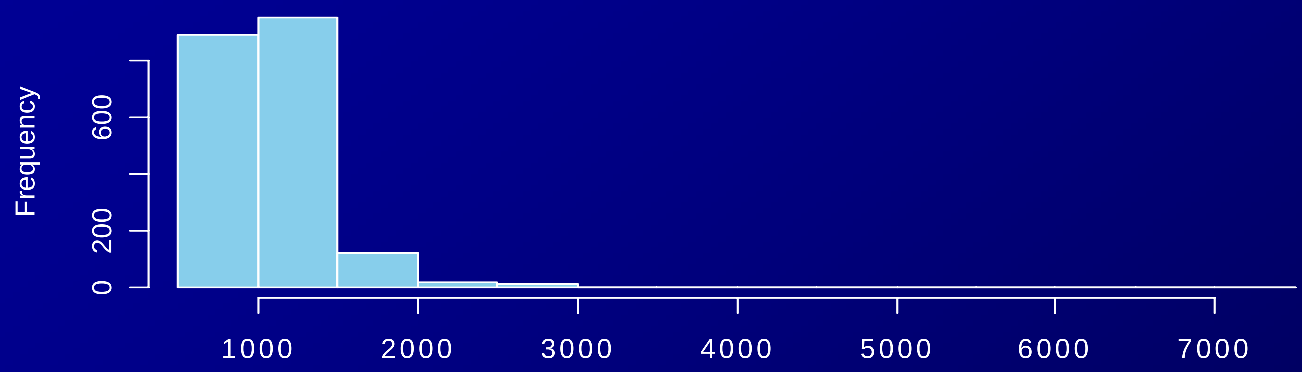
where $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\log\left(X_i\right),\ \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log\left(X_i\right)-\hat{\mu}\right)^2}$

Which Estimator is Better? $E_N[X]$ or $E_L[X]$?

- Straight Average, $E_N[X]$, is simple.
- Lognormal Average, $E_L[X]$ is complicated.
  - But derived from the maximum likelihood estimator for the lognormal distribution
- Evaluate by a simulation
  - Sample size of 500
  - 2,000 samples
- Look at the variability of each estimator

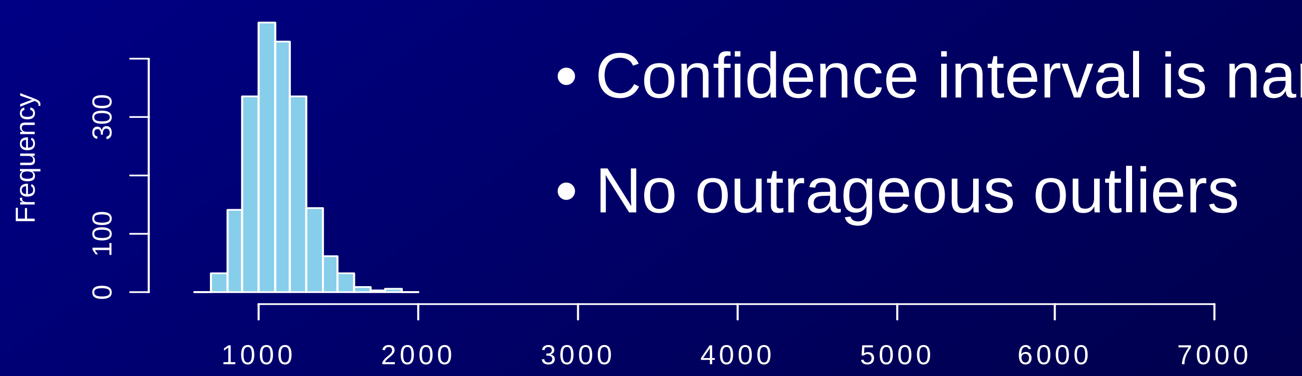
Results of Simulation

Straight Average



95% Confidence Interveal = (719.5 , 1821.3)
Maximum = 7320.5

Lognormal Average



95% Confidence Interveal = (823.4 , 1505.7)
Maximum = 1953.4

- Confidence interval is narrower
- No outrageous outliers

Lesson from Example 1

- *Knowing the distribution of the observations can lead to a better estimate of the mean!*
- Actuaries have long recognized this.
  - Longtime users of robust statistics
    - Calculate basic limit average severity
    - Fit distributions to get excess severity

Fitting Multivariate Models by Direct Maximum Likelihood Estimation

- Most statistical software packages have generic optimizers
  - Excel "Solver"
  - R "optim"
- Use to solve for maximum likelihood

Example 2 – Pareto Distribution

- Claim severity "data" taken from various cities over the years 2004-2007.
  - Simulated from known model

$$F(z) = 1 - \left(\frac{Scale}{Scale + z} \right)^{\alpha}$$

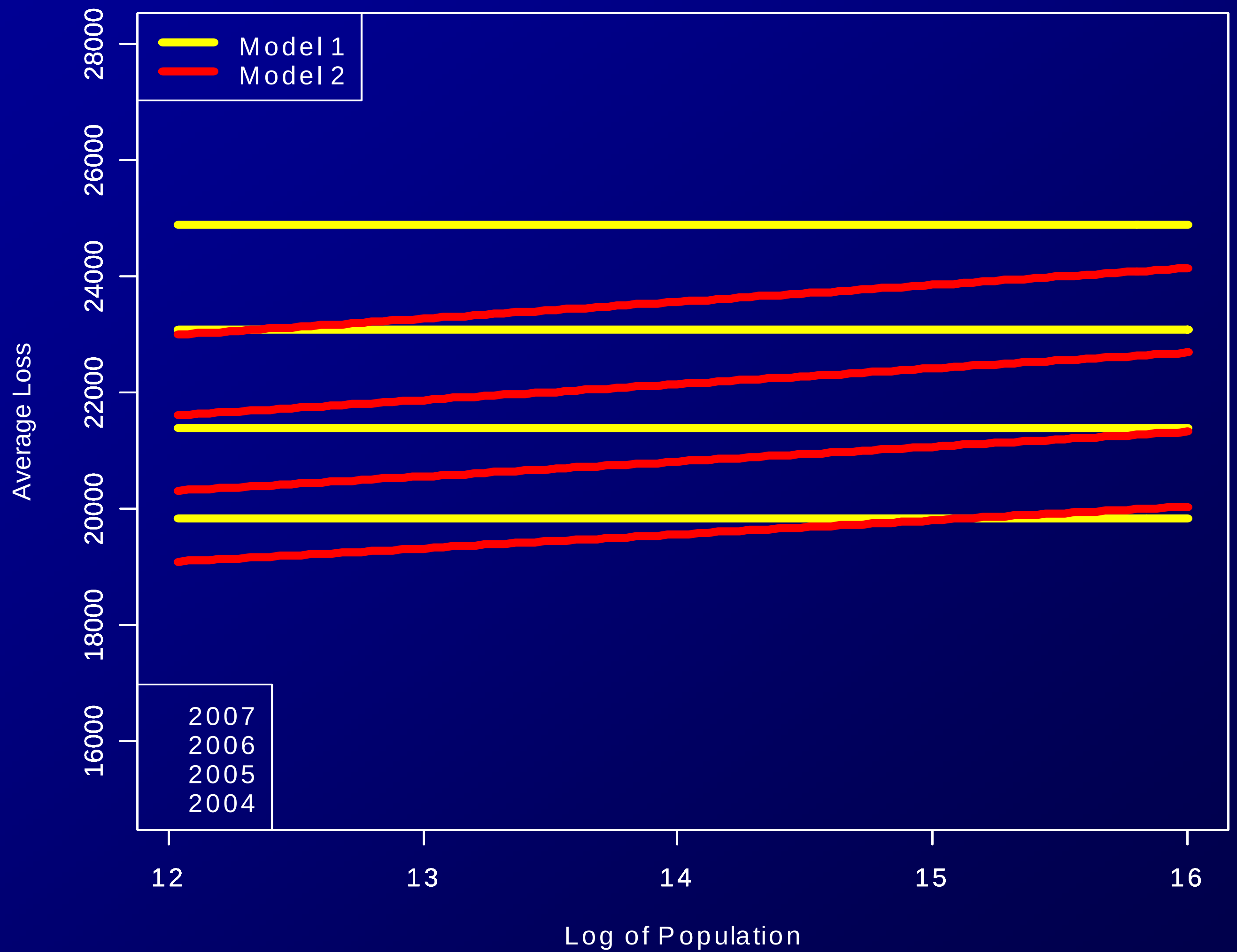
- Model 1 - Scale = $\theta \cdot Trend^{(Year - 2004)}$
- Model 2 - Scale = $(\theta + \gamma \cdot \log(Pop)) \cdot Trend^{(Year - 2004)}$
- Parameters to be estimated
  - Trend, α, θ, γ

Parameter Estimates

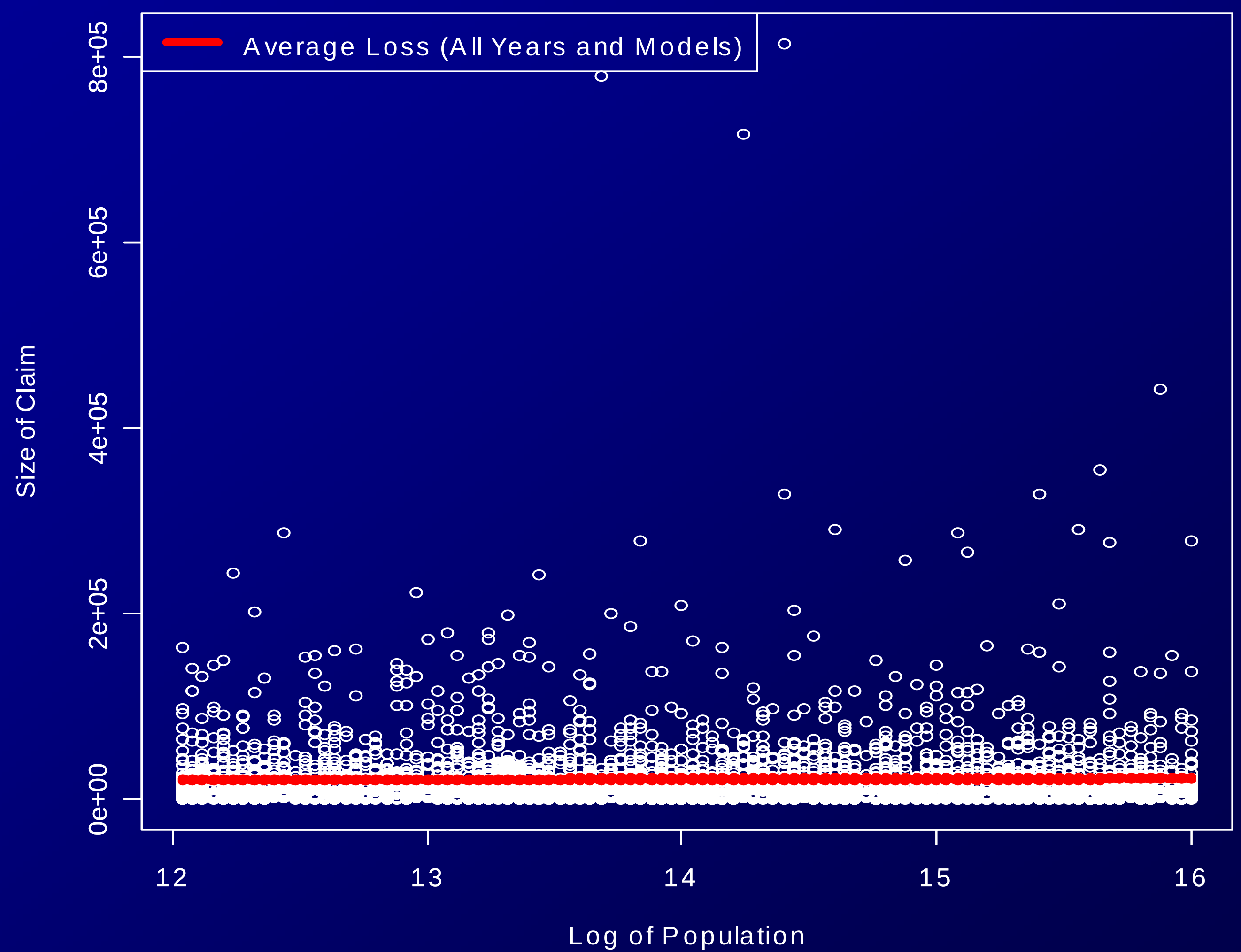
Parameter	True	Model 1	Model 2
θ	25,000	27,834	27,476
α	2.500	2.403	2.607
Trend	1.050	1.079	1.064
γ	500	---	408

Plot of Expected Loss

$$E[Loss] = \frac{(\theta + \gamma \cdot \log(Pop)) \cdot Trend^{Year-2005}}{\alpha - 1}$$



Plot with Actual Loss



Is the log(*Pop*) Term Statistically Significant?

- Use the likelihood ratio test
  - $L(\theta_2,\alpha_2,\mathit{Trend}_2,\gamma_2)$ = Log Likelihood for Model 2
  - $L(\theta_1,\alpha_1,\mathit{Trend}_1)$ = Log Likelihood for Model 1

- $2\cdot(L(\theta_2,\alpha_2,\mathit{Trend}_2,\gamma_2) - L(\theta_1,\alpha_1,\mathit{Trend}_1)) \sim \chi^2(1)$

- P-Value for test = 0.034
  - Significant at 0.05 level, but not at 0.01 level

Test Goodness of Fit with P-P Plots

- Calculate percentile, p_i, of each data point

$$p_i = 1 - \left(\frac{b_i}{b_i + z_i} \right)^{\alpha}, b_i = \left(\theta + \gamma \cdot \log\left(Pop_i \right) \right)^{Year_i - 2004}$$

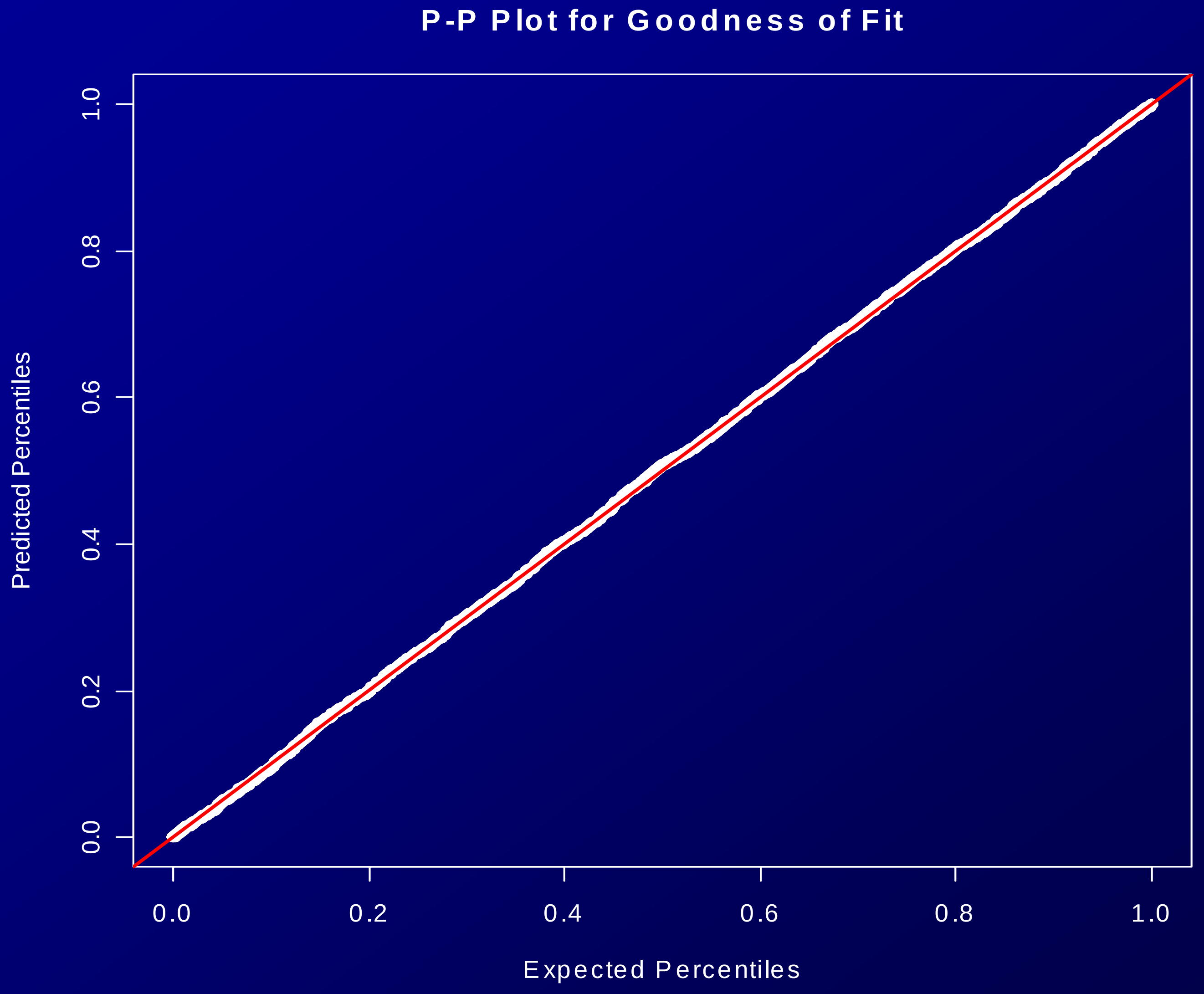
- Plot against expected percentiles

$$\left(p_i, \frac{i}{n+1} \right)_{i=1}^{n}$$

- Straight 45º line indicates a good fit

P-P Plot for Example 2

Fit should be good – I knew the model



Lesson From Example 2

- *Maximum likelihood is practical for multivariate models with today's PCs with the right software installed!*
  - SAS, R and others.
  - Personal best – 20 parameters on a loss reserve model

Generalized Linear Models

- Generalization of the “General Linear Model”
  - The General Linear Model
    - Least-squares analysis of continuous and categorical variables.
    - I first encountered it in SAS in late 70’s.
- First book – 1989, McCullagh and Nelder
- Latest book – Good introduction for actuaries
  - Generalized Linear Models for Insurance Data
    - De Jong and Heller

Properties of GLM's

- Efficient maximum likelihood estimation for a specific (but broad) class of distributions.
- For most common problems
  - Convergence takes a single digit # of iterations
  - For generic maximum likelihood optimizers it takes a triple digit number of iterations

Properties of GLM's

- Link function - g (Monotonic and smooth)
  – Let μ be the mean of the independent variable

$$g(\mu) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \cdot x_i$$

Some Common Links

Identity	$g(\mu) = \mu$
Inverse	$g(\mu) = 1/\mu$
Inverse squared	$g(\mu) = 1/\mu^2$
log	$g(\mu) = \log(\mu)$
logit	$g(\mu) = \log(\mu/(1-\mu))$

Properties of GLM's

- Distribution Function (with mean μ)
  - Variance of response distribution is a function of μ
  - Variance function is ***determined by the distribution***

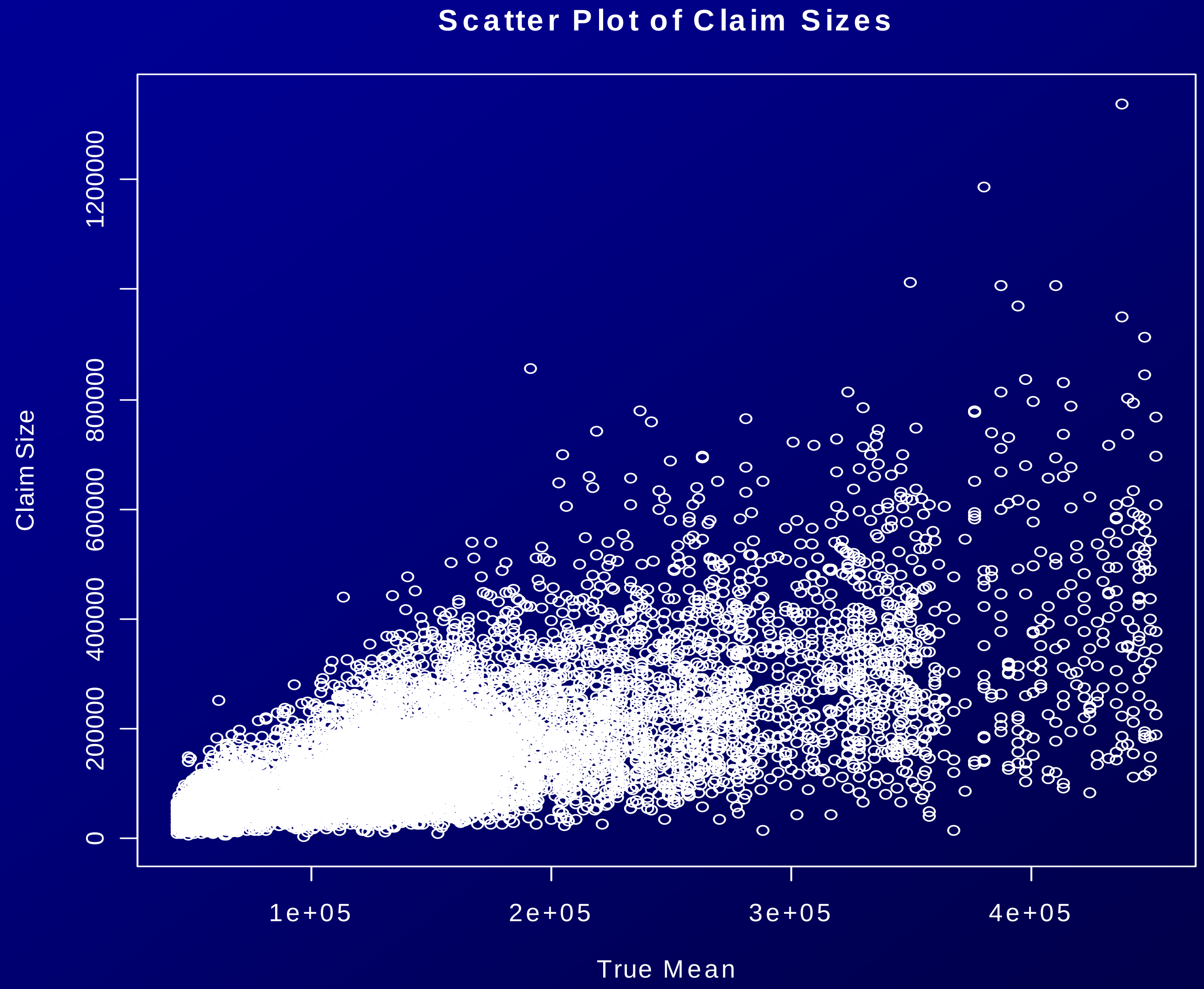
Some Common Distributions

Distribution	Variance
Normal	$1/\sigma^2$
Poisson	μ
Gamma	μ^2/ν
Inverse Gaussian	μ^3/σ^2
Negative Binomial	$\mu(1+\kappa\mu)$

Example 3 – Property Claim Size

- Construction
  - Frame, Masonry, and Fire Resistive
- Protection
  - 1,2, …, 10 with 1 being the best protection
- Amount of Insurance

Properties of Simulated Data



Model 1

$$\log(\mu) = \alpha_0 + \mathit{class}_i + \alpha_1 \cdot \log(\mathit{prot}) + \alpha_2 \cdot \log(\mathit{aoi})$$

```
Call:
glm(formula = z~cons+log(prot)+log(aoi),family = Gamma(link="log"))


Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.674263   0.155404   55.82   <2e-16 ***
consMasonry   -0.204571   0.013600  -15.04   <2e-16 ***
consResistive -0.913219   0.013648  -66.91   <2e-16 ***
log(prot)      0.380237   0.007967   47.73   <2e-16 ***
log(aoi)       0.235316   0.012365   19.03   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01
```

Model 1

$\log(\mu) = \alpha_0 + class_i + \alpha_1 \cdot \log(prot) + \alpha_2 \cdot \log(aoi)$

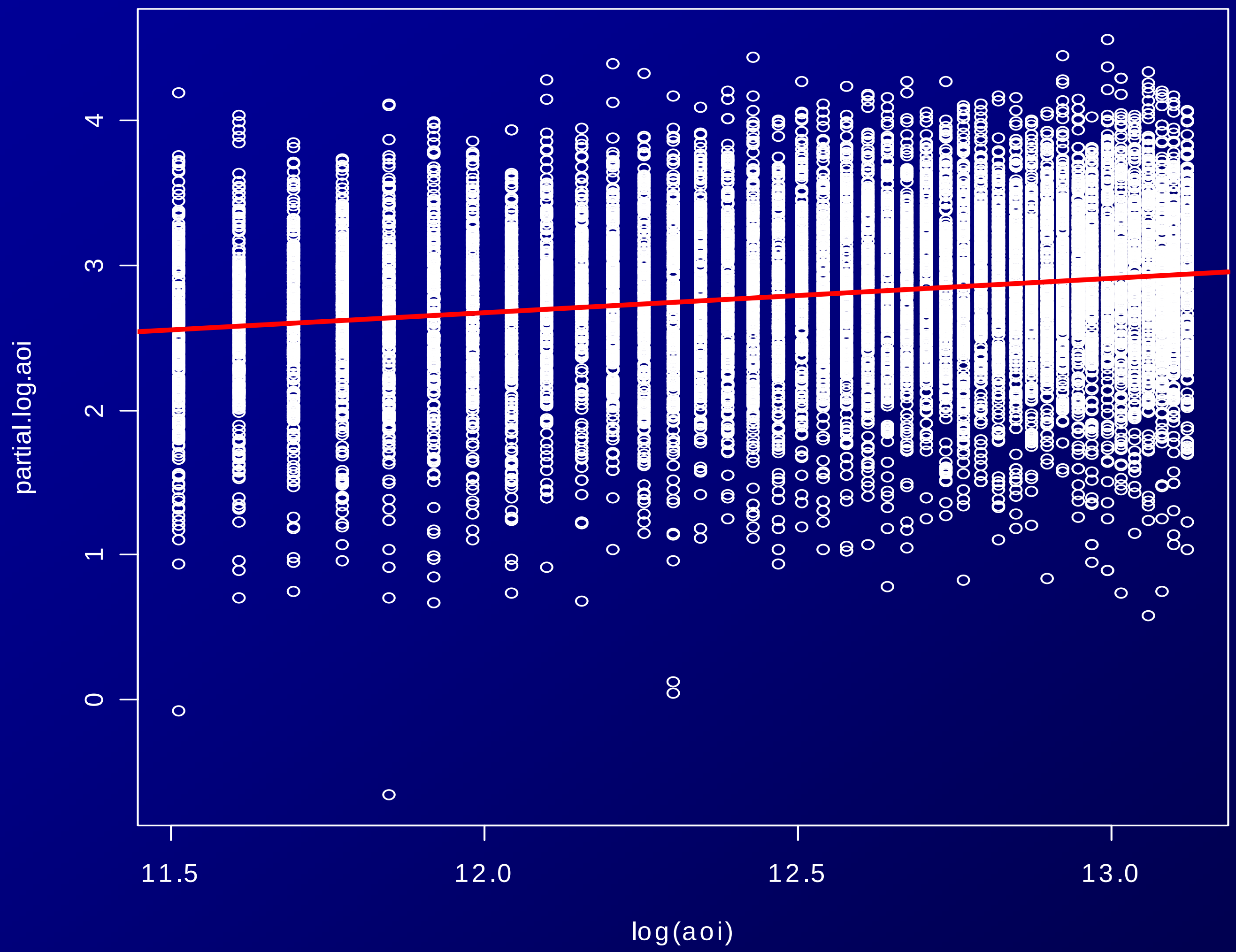
- Is the model linear in log(*prot*) and log(*aoi*)?
- Test with ***Partial Residual Plots***

$$\left(\log(prot), \log(z) - \log(\hat{\mu}) + \alpha_1 \cdot \log(prot)\right)$$

$$\left(\log(aoi), \log(z) - \log(\hat{\mu}) + \alpha_2 \cdot \log(aoi)\right)$$

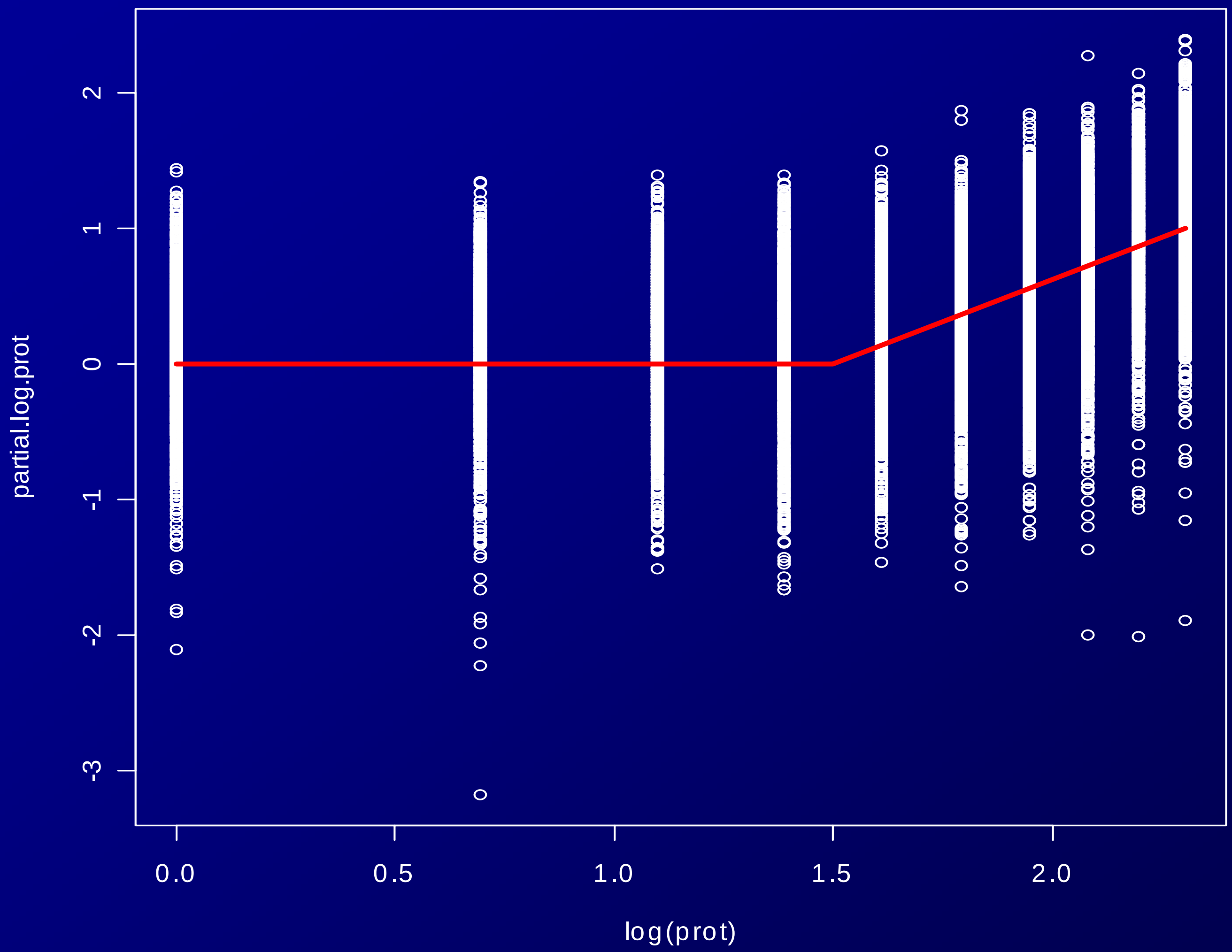
- The plots should be distributed about a straight line with slope α_i

Partial Residual Plot for log(*aoi*)



Looks straight to me

Partial Residual Plot for log(*prot*)



Not straight

Dealing with Nonlinear Effects

- Generalized additive model (GAM)
- Allows a spline to replace the linear term

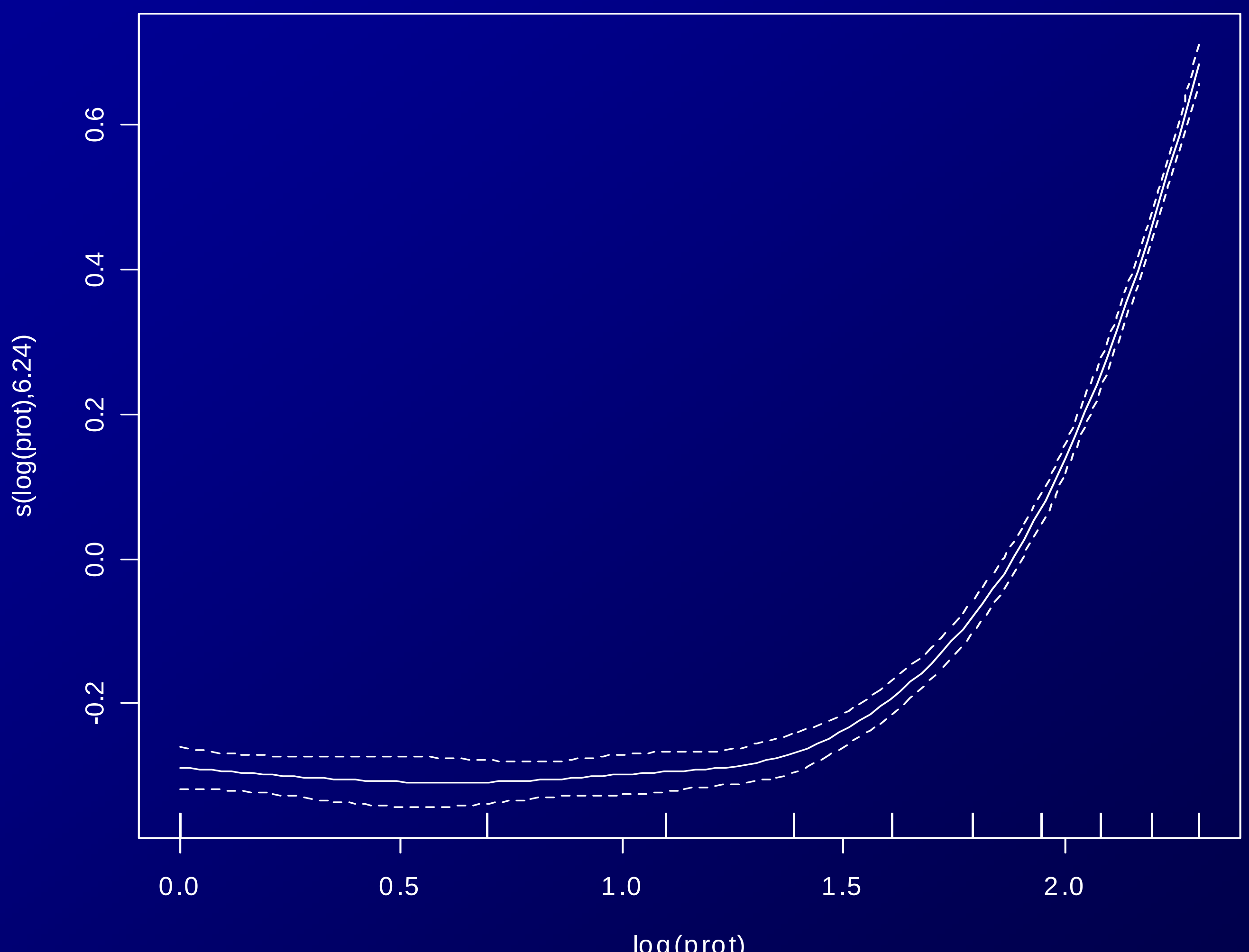
```
Family: Gamma
Link function: log

Formula:
z ~ cons + s(log(prot)) + log(aoi)

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.09303    0.14133   64.34   <2e-16 ***
consMasonry   -0.20656    0.01240  -16.66   <2e-16 ***
consResistive -0.91135    0.01244  -73.23   <2e-16 ***
log(aoi)       0.24568    0.01128   21.79   <2e-16 ***

Approximate significance of smooth terms:
               edf Est.rank     F p-value
s(log(prot)) 6.238        8 229.9  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plot of the Spline



Commentary on GLM

- GLM's represent a significant advance over the normal/least squares paradigm.
  - Based on maximum likelihood estimation
- Since it has been around for over a decade, there is a lot of supporting software.
  - e. g. GAM
- Restricts the choice of response distributions.
  - Too restrictive ??? Debatable.
- Links can be supplied by the user.

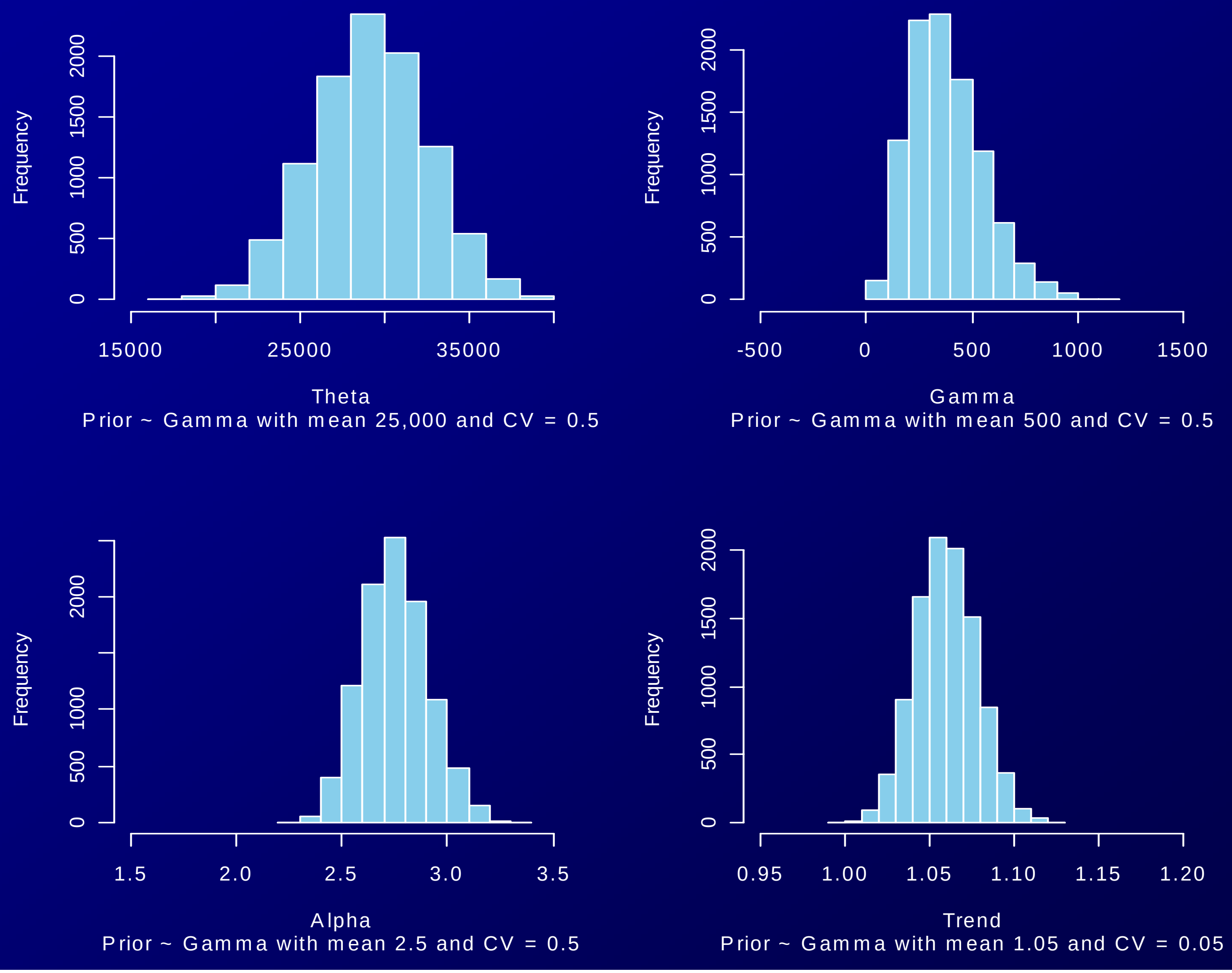
The Future - Predicting Ranges

- Anybody can predict the future
- It is harder to make the right prediction
- How much prediction error should be tolerate?
- Determined by well thought out estimates of the prediction error.
  - Verified by back testing with P-P plots

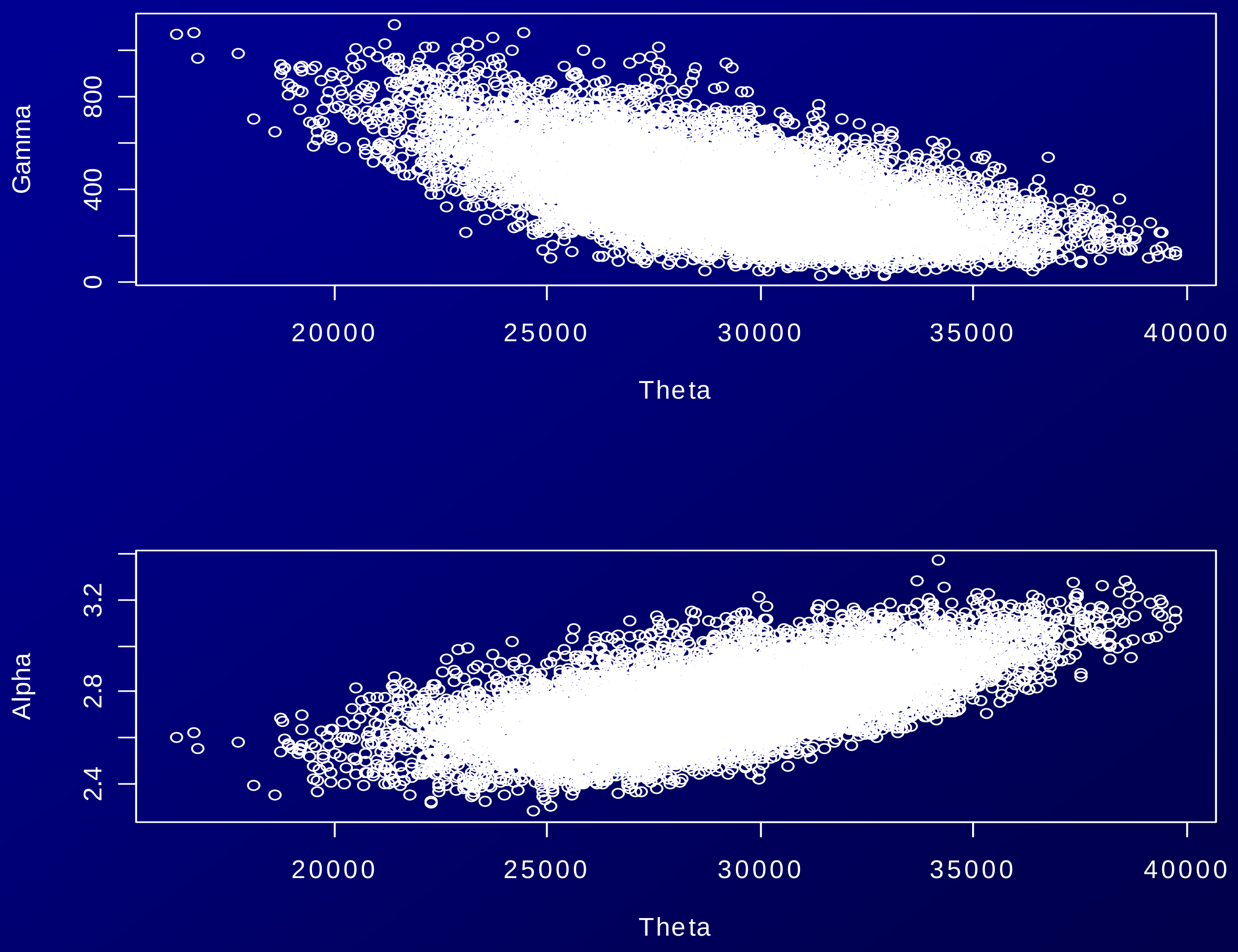
Back to Example 2 Parameter Uncertainty and the Gibbs Sampler

- Gibbs sampler is often used for Bayesian analyses.
- It randomly generates parameters in proportion to posterior probabilities.
- Parameters randomly fed into the sampler in proportion to prior probabilities.
- Accepted in proportion to $\dfrac{\text{Likelihood}}{\text{Maximum Likelihood}}$
- Results in the posterior distribution.

Posterior Distribution of Parameters

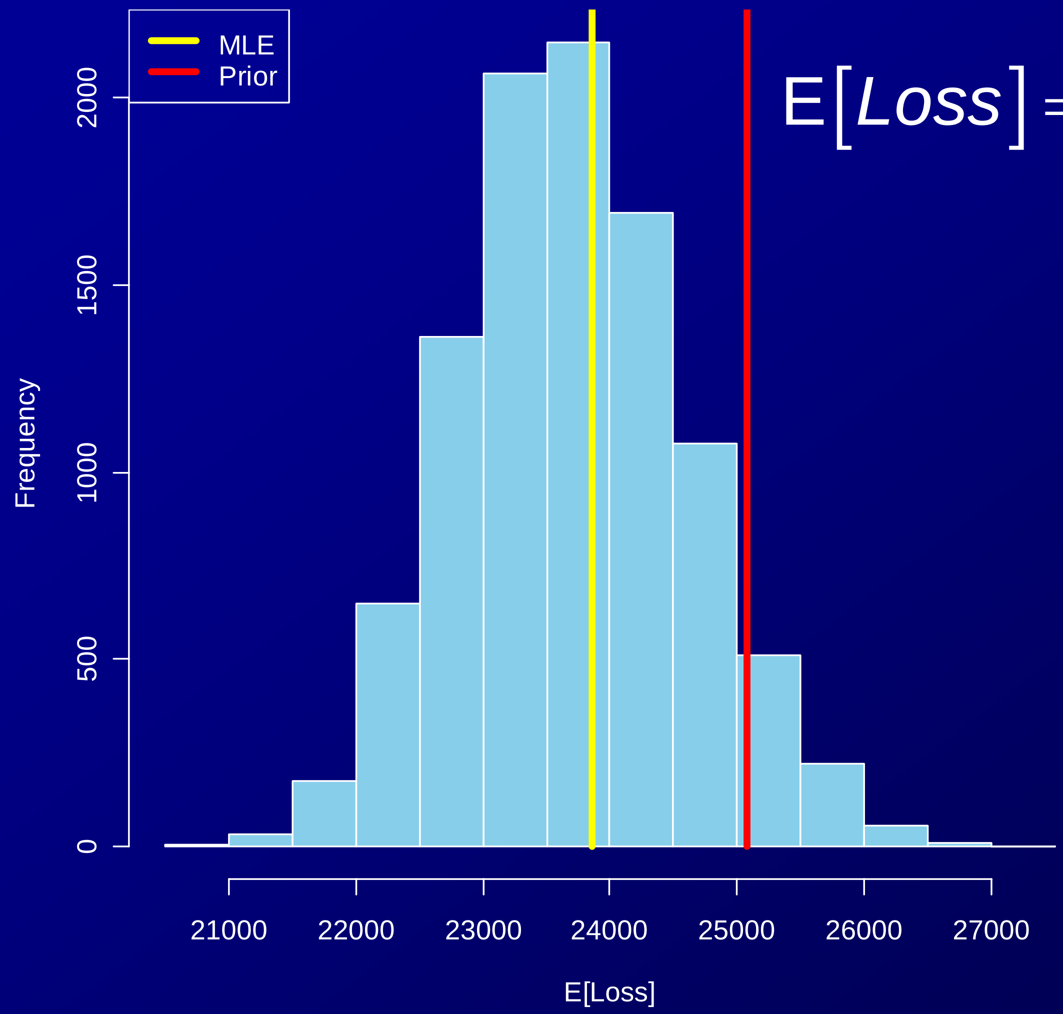


Posterior Distribution of Parameters



Posterior Distribution of E[*Loss*] for 2007 with log(*Pop*)=15

$$E[Loss] = \frac{(\theta + \gamma \cdot \log(15)) \cdot Trend^3}{\alpha - 1}$$



Methods are New, but What Else? The Data!

- Large data sets and more variables
- More variables are statistically significant ***in sample***!
- Statistical significance does not mean “practical significance.”
- Practical significance is best tested by graphical methods.
- Need to test “out of sample.”

Software – An incomplete list

- PC SAS
- SAS Enterprise Miner (JMP for Graphics)
- R, the examples and graphics for this talk were done using R.
- S-Plus (similar to R)
- Statistica
- SPSS

Concluding Remarks

- Most of the buzz in predictive modeling has to do with pricing applications.
- Other insurance applications
  - Loss Reserving
  - Fraud detection
  - Premium Audit
- What to do with ranges of estimates?
  - Accounting issues e.g. loss reserve risk margins