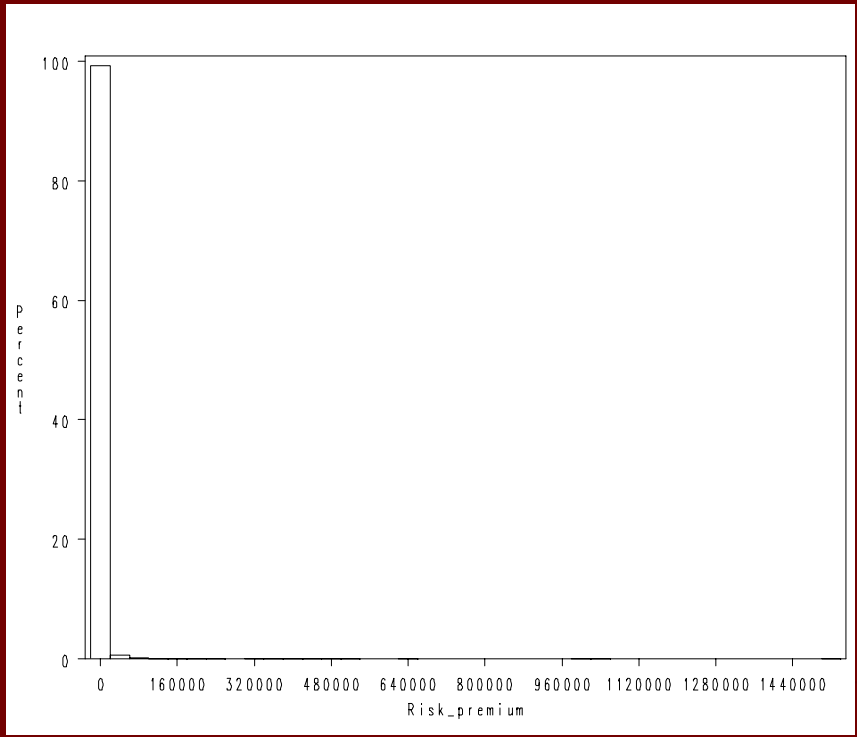
Society Of Actuaries in Ireland
13 November 2008
Derek Bain

Estimating the Predictive distribution for Risk Premiums using Bootstrapping

Direct Modelling of Risk Premiums

Direct modelling of Risk Premium (or claims cost per policy) using “traditional” severity distributions is problematic due to the large spike at zero. We need a distribution which accommodates a point mass at zero coupled with a wide spread of positive values.



Mean	1179.307	Std Deviation	18359
Median	0.000	Variance	337070333
Mode	0.000	Range	1527138

Quantile	Estimate
100% Max	1.52492E+06
99%	1.57563E+04
95%	2.43606E+03
90%	8.64781E+02
75% Q3	9.23712E+01
50% Median	0.00000E+00
25% Q1	0.00000E+00
10%	0.00000E+00
5%	0.00000E+00
1%	0.00000E+00
0% Min	-2.21898E+03

Tweedie Distribution

- Exponential Dispersion Models are (loosely) a family of densities characterised by the variance = [scale parameter].V(mean) where V() is called the variance function.
- The Tweedie distributions are a class of exponential dispersion models where the variance function is a power function $V(\mu) = \mu^p$.
- The case p=1 corresponds to a poisson distribution while p=2 corresponds to a gamma distribution. The case 1<p<2 correspond to compound poisson distributions.
- Tweedie distributions have a point mass at zero and a continuous density for values greater than zero (see appendix 1).
- The density function for the incurred claims cost and risk premium is an exponential density (so GLM theory applies!) with a variance function $V(\mu)=\sigma^2\mu^p$ with 1<p<2.

Model Form

- We are trying to fit a model of the form
  $Y = \exp(age+gender+licence+cover+group+NCD+area)$
  where y=(total claims)/exposure = observed claim rate per unit of exposure. $E(Y) = \mu$ and $Var(Y) = \sigma^2/w.(V(\mu)) = \sigma^2/w.\mu^p$ where w is the exposure.

  As Y has an exponential distribution the model can be fitted using standard GLM methods.
  Linear Predictor : age+gender+licence+cover+group+NCD+area
  Link Function: Ln()
  Variance Function: $V(\mu)=\mu^p$
  Weights : Exposure

  Main complications come from the form of the variance function (you need to specify the form of the deviance increment when using standard software) and the estimation of p.

Quasi (Log) Likelihood

- Define the quasi-likelihood function as

$$Q(\mu;y)=\int_{y}^{\mu}\{(y-t)/\sigma^{2}V(t)\}dt$$

- The function $Q(\mu,y)$ behaves like an ordinary likelihood. The main difference being that you are only specifying the relationship between the mean and the variance ($Var(Y)=\sigma^2V(Y)$) of the response variable Y. In a normal likelihood formulation you need to specify the complete density function for the response variable.

- Define the quasi-deviance as $D(y;\mu) = -2*\sigma^2*Q(\mu;y)$.

- Maximum (quasi) likelihood is equivalent to minimum (quasi) deviance.

- If $V(y) = c$ then solving then $D(y;\mu)= (y-\mu)^2/c$ - the familiar minimum sum of squares

- With $V(y) = \mu^p$ $p \neq 0,1,2$ then

$$D(y;\mu) = 2*\{(1/(1-p))*(y(y^{(1-p)}-\mu^{(1-p)}))-(1/(2-p))*(y^{(2-p)}-\mu^{(2-p)})\}$$

This is the form of the deviance increment we need.

Extended Quasi Likelihood

- The quasi-deviance function cannot be used to select a value for the p parameter in the variance function. To see this consider the quasi-deviance as a function of a function of p (note that D ≥ 0)

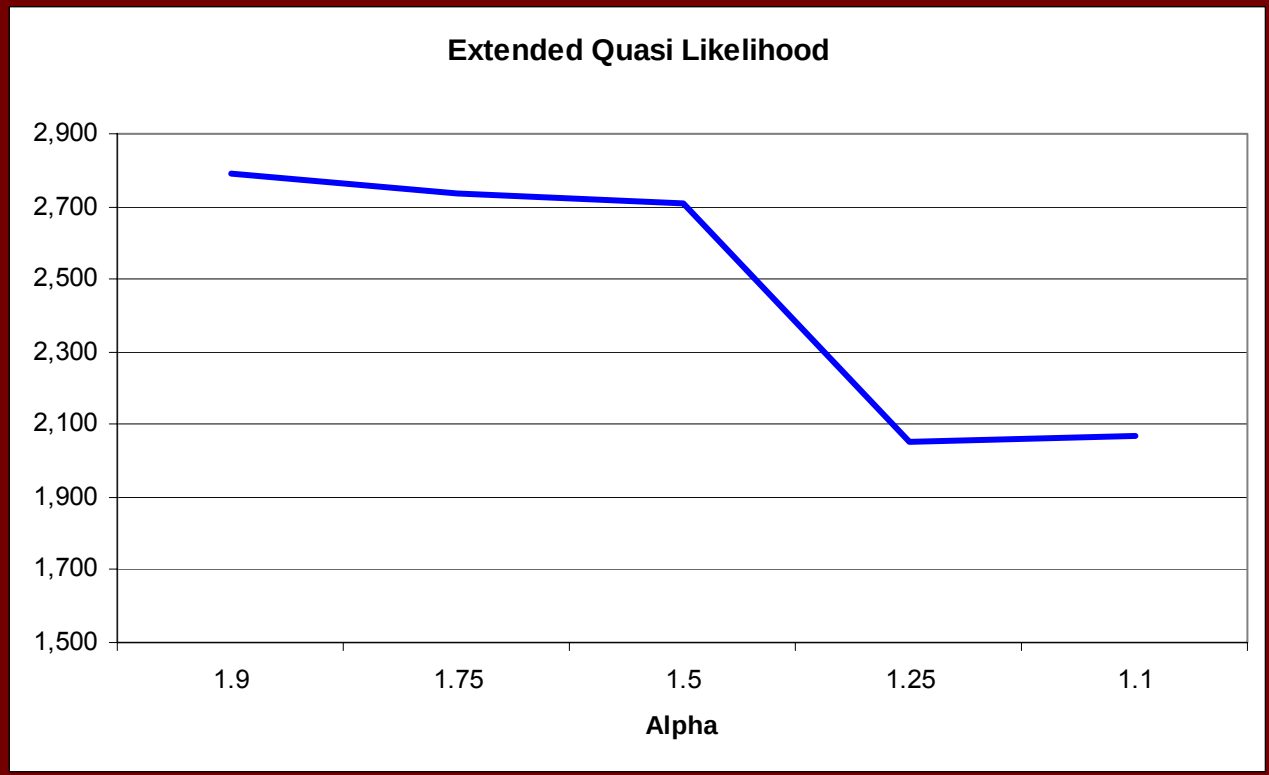
$$D(\mu; y; p) = 2\int_{\mu}^{y}\{(y-t)/\mu^{p}\}dt \to 0 \text{ as } p \to \infty \text{for } \mu > 1$$

- Fitting the p parameter using a minimum deviance criteria will result in a very large p value and the contribution of the other parameters will be swamped.
- Define the Extended Quasi Likelihood as

$$Q^{+} = -1/2\{\Sigma \ln(2\pi\sigma^{2}y^{p})\} - 1/2D(\mu;y;p)/\sigma^{2}$$

- Here the additional term is an increasing function of p and counterbalances the effect of the reducing deviance. Note that the deviance also enters Q+ through the estimated scale parameter σ^{2}.

Extended Quasi-Likelihood function



Final Model

Linear Predictor : age+gender+licence+cover+group+NCD+area
Link Function: Ln()
Variance Function: $V(\mu)=\mu^p$; P = 1.25
Weights : Exposure

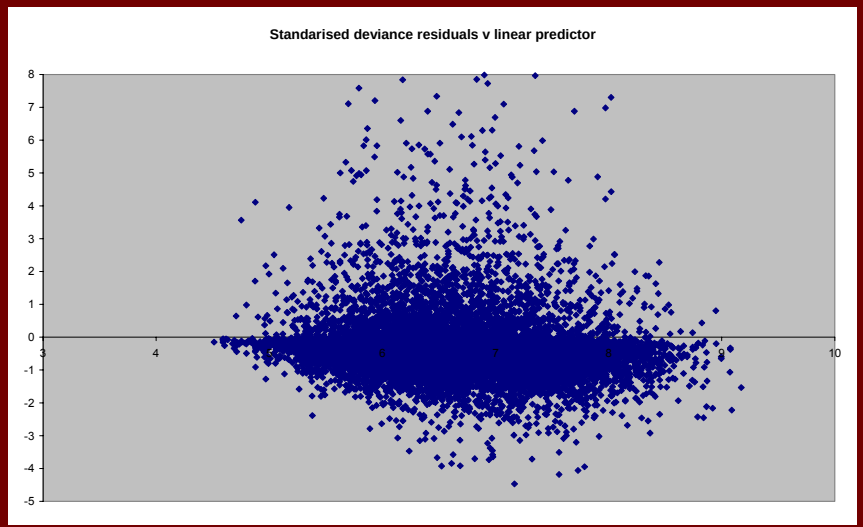
SAS Code used in fit

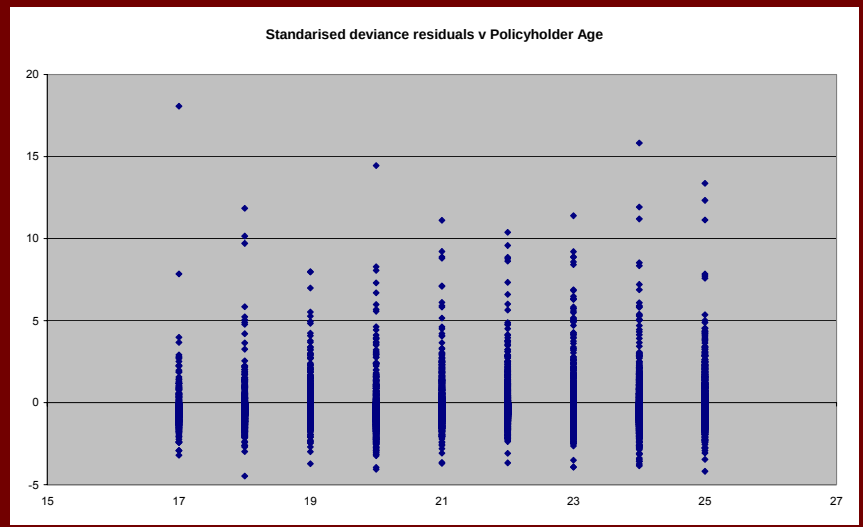
```
proc genmod data = summary_input2;

class group age cover gender licence ncd area;
        a=_MEAN_;
        y=_RESP_;
        p=1.25;
        d=0;
        if y GE 0 then d=2*((y**(2-p)-((2-p)*y*a**(1-p))+((1-p)*a**(2-p)))/((1-p)*(2-p)));
        variance var = a**p;
        deviance dev = d;
        model risk_premium=age gender licence cover group ncd area/noint scale=deviance link=log;
        weight exposure;
        output out=genout pred=rp stdresdev=resid stdreschi=resid2 resdev=r;
run;
```

Results from fit

•While the model fitted has been kept simple – so as to facilitate the bootstrap procedure the fit still seems reasonably good.





Bootstrapping regression models

The simplest regression model is $y_j = x^T_j \beta + e_j$ where (y_j, x_j) are the response and the px1 vector of covariates for the jth case. There are two approaches to bootstrapping this model.

1. Resample residuals

   Set $r_j = y_j - x^T_j \beta'$ and $rs_j = (r_j - ave(r_j))/sqrt(1-h_j))$ where h_j is the leverage of the jth observation. Note that rs_j has zero mean and constant variance.

   Then set $y_j^* = x^T_j \beta' + \varepsilon_j$ where ε_j is taken randomly from the set of standardised residuals $\{rs_1, rs_2, \ldots rs_n\}$. Refit the regression model and obtain a new set of parameter estimates. Repeat this 10,000 times and you generate an empirical distribution for the parameter estimates.

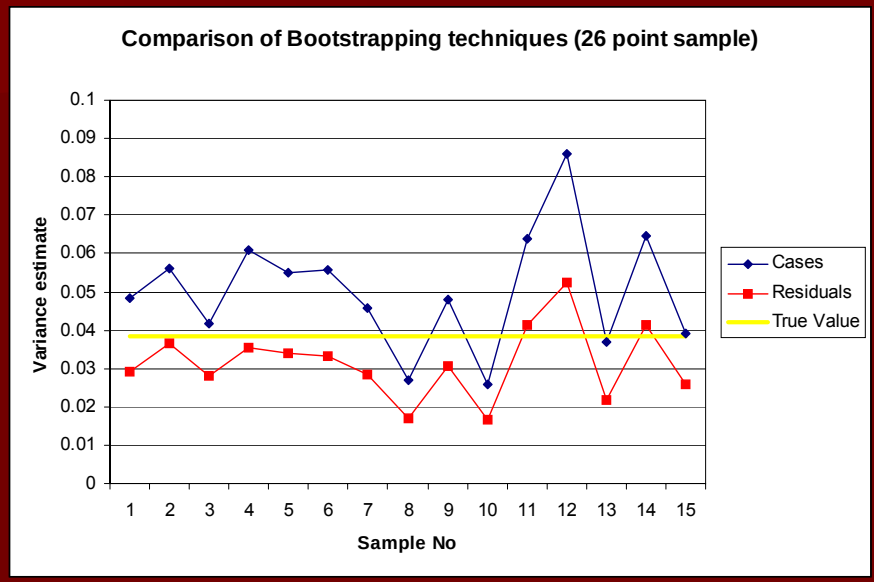
Bootstrapping regression models

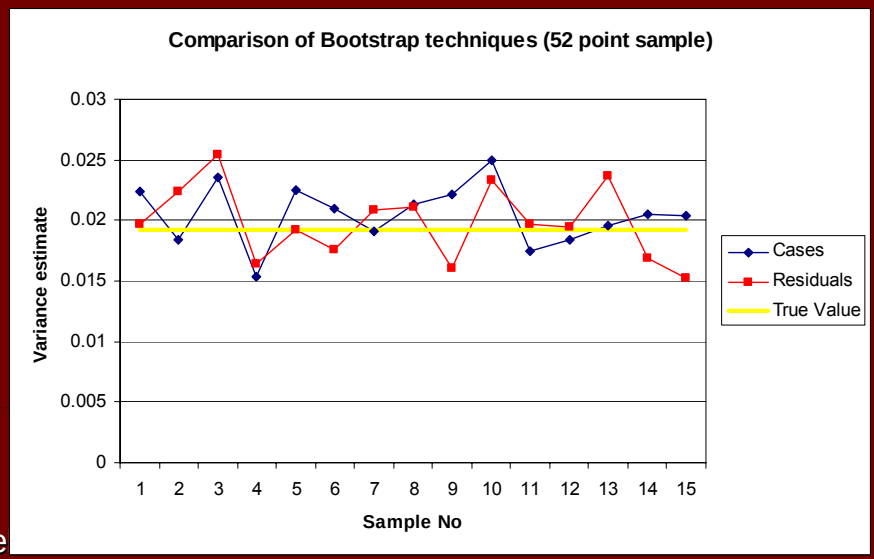
2. Resample cases $\{(y_j,x_j) \ldots (y_n,x_n)\}$

   Under this approach you resample (with replacement) the original cases and generate a bootstrap data set $\{(y_1,x_1)^* \ldots (y_n,x_n)^*\}$ to which you fit the regression model and generate a new set of parameter estimates. Repeat this 10,000 times and you generate an empirical distribution for the parameter estimates. (As an aside this approach to bootstrapping seems to offer an alternative method for generating pseudo cumulative incurred claims triangles without incurring the negative incremental claims problems associated with re-sampling residuals.)

Bootstrapping regression models

- The approach of re-sampling cases to generate pseudo data is the more usual form of bootstrapping. The approach is robust in that if an incorrect model is fitted an appropriate measure of parameter uncertainty is still obtained. However re-sampling residuals is more efficient if the correct model has been fitted.
- The graphs shows both approaches in estimating the variance of a 26 point data sample mean and a 52 point sample mean. In the larger sample the two approaches are equivalent.



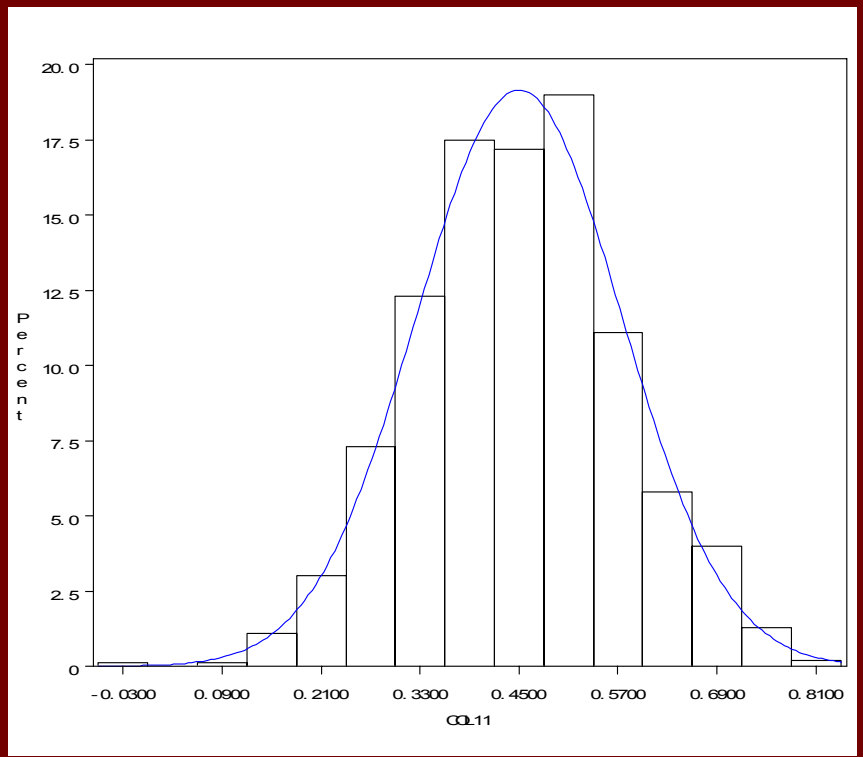


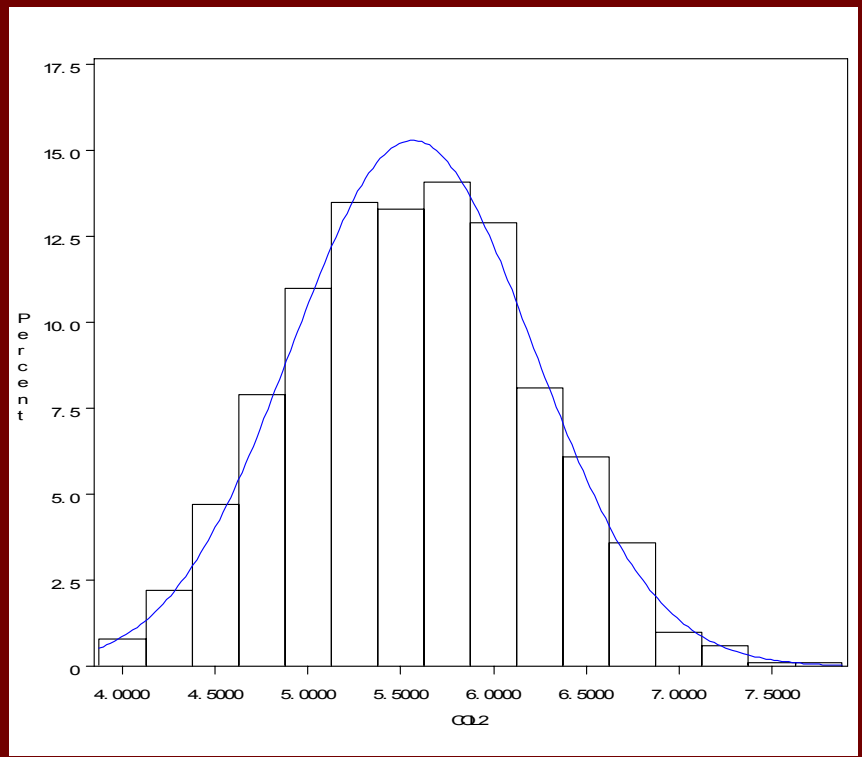
Output from bootstrap

- Bootstrap Algorithm
  1. Fit GLM to original data set and store parameter estimates.
  2. Generate pseudo data set (same size as original data set) from original data set by re-sampling with replacement from original data set (I did this by generating a file of random numbers and merging the original file onto the new file where the original row number equalled the random number).
  3. Refit the GLM to the pseudo data and store the parameters.
  4. Repeat stages 2 and 3 N times.
  5. Generate required statistics from parameter estimates $\beta_1,\beta_2,...\beta_N$ where β_i is the vector of parameters output during iteration i.

Output from bootstrap

- The following graphs display the distribution of 2 parameter estimates based on 1,000 iterations of the bootstrap.





Empirical Parameter distributions

- Risk Premium(1,1,1,1,1,1,1) is given by

$$\exp(\text{age}(1)+\text{gender}(1)+\text{licence}(1)+\text{cover}(1)+\text{group}(1)+\text{NCD}(1)+\text{area}(1))$$

Where age(1) is parameter estimate for level one of the age variable, gender(1) is the parameter estimate for level one of the gender variable etc.

- For the parameters estimated in the model we have an multivariate empirical distribution (via the bootstrap).
- In order to produce an empirical distribution for a risk premium we

1. Randomly draw a vector of parameter estimates from the multivariate empirical parameter distribution and calculate a risk premium. Note that the re-sampling must preserve the dependence structure between parameters.
2. Repeat step one N times to produce an empirical distribution.

Risk Premium Distribution

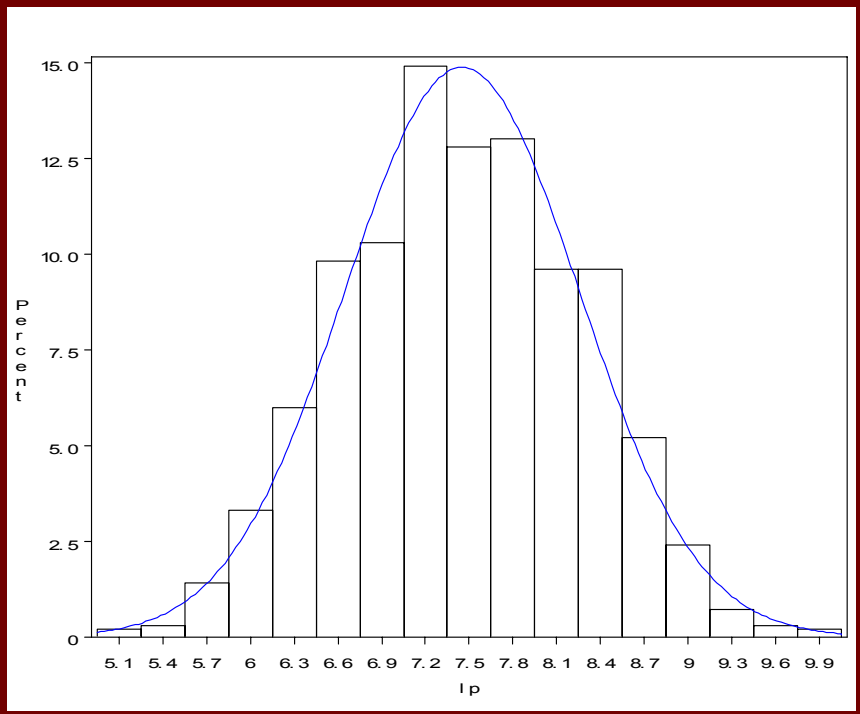
- Extract from Parameter empirical distributions

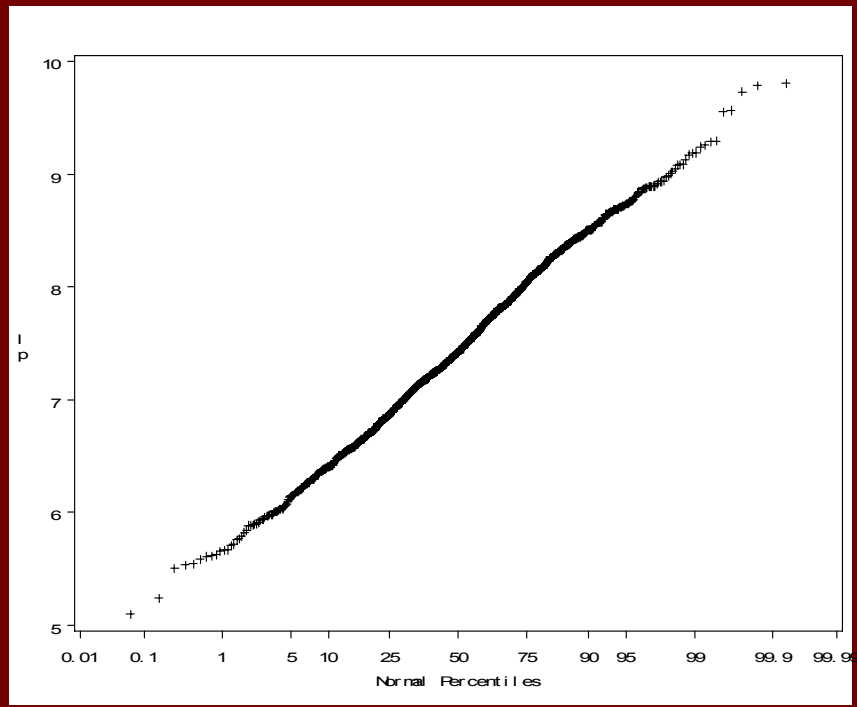
Index	P1(1)	P1(2)	P1(3)	P1(4)	P1(5)	P1(6)	P1(7)	P1(8)	P1(9)	P2(1)	P2(1)	P3(1)	P3(2)	
1	5.681112	5.445491	4.920464	5.526682	4.825199	4.723221	4.695506	4.794742	4.698815	0.451008	0	0.582003		0
2	5.456503	5.193568	4.858869	4.918304	4.694945	4.471013	4.524467	4.780155	4.706512	0.411565	0	0.364002		0
3	6.571027	5.581127	5.264005	5.10033	5.222713	5.024645	4.971457	5.007636	4.861884	0.463174	0	0.411377		0
4	5.239579	5.292666	4.795608	5.297541	4.726769	4.697402	4.6389	4.697812	4.639287	0.461228	0	0.543496		0
5	5.173393	5.30584	4.855483	5.46932	4.683211	4.510123	4.556009	4.662907	4.481004	0.301002	0	0.571876		0
6	6.288829	5.551724	5.055563	6.002004	4.896509	5.018913	4.973691	5.018677	5.001962	0.517063	0	0.584115		0
7	4.285431	4.650733	4.118431	5.288863	4.02891	4.005287	3.998739	4.11361	4.222197	0.370499	0	0.962577		0

- Extract from Risk Premium empirical distributions

index	P1(1)	P2(1)	P4(1)	P5(3)	P6(1)	P7(1)	LP	rp_para_error
1	5.6811	0.4612	0.0530	0.2548	1.0693	0.0047	7.5242	1345
2	5.6811	0.4612	0.0530	0.2548	1.0693	0.0047	7.5242	1345
4	6.5710	0.3705	0.0913	0.6565	1.0721	0.0569	8.8073	4853
5	5.1734	0.4560	-0.0011	0.0438	1.1910	0.0486	6.9116	729
6	5.1734	0.4560	-0.0011	0.0438	1.1910	0.0486	6.9116	729
7	6.2888	0.4997	0.2177	0.3145	1.0165	0.0387	8.3858	3184

Bias Adjustment - Linear Predictor Distribution





Bias Adjustment - Linear Predictor Distribution

- Parameters for Normal Distribution

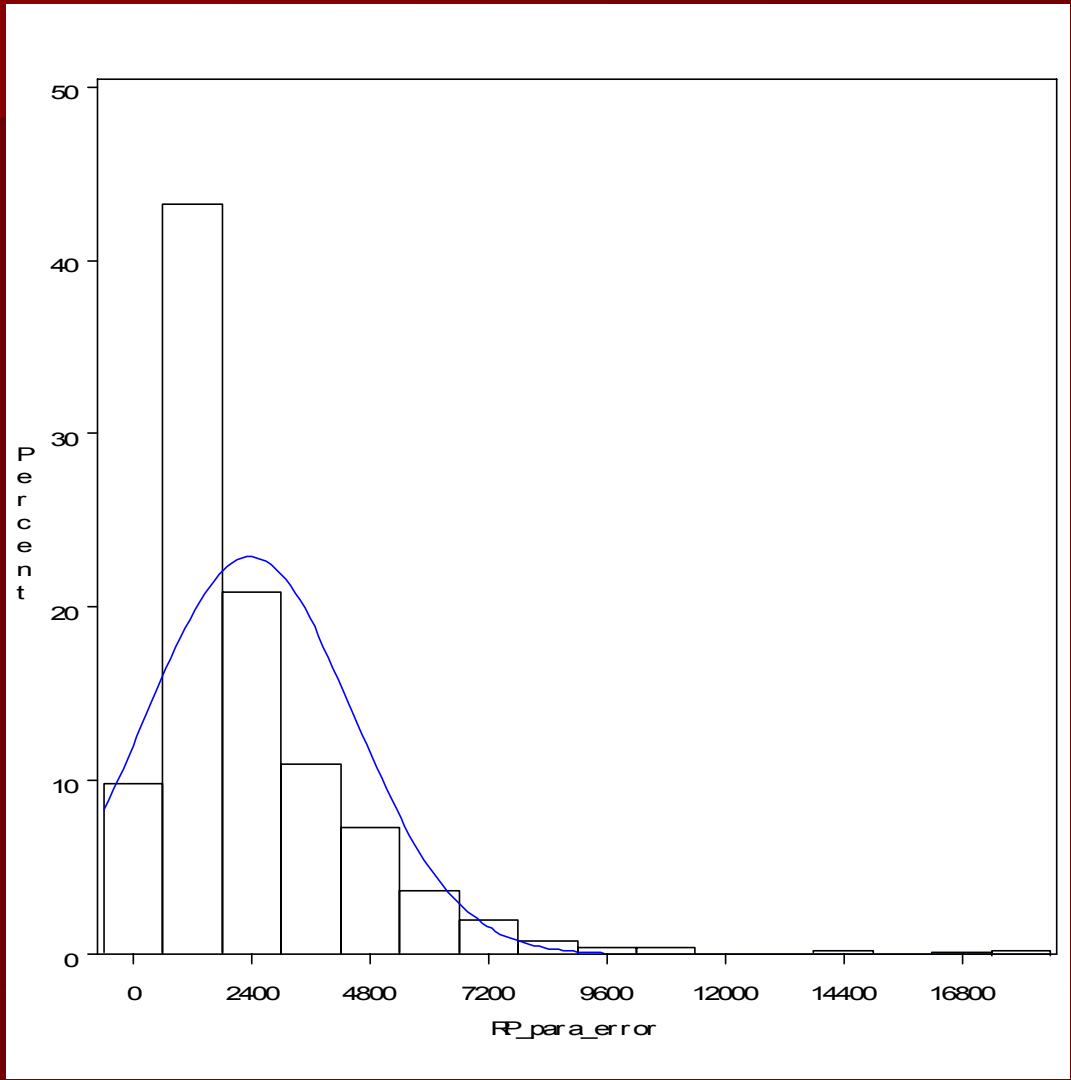
Parameter	Symbol	Estimate
Mean	μ	7.451092
Std Dev	σ	0.804347

- Linear Predictor (lp) ~ $N(\mu,\sigma^2)$
- With a GLM $E(Y) = Exp(\mu)$.
- What we have calculated is $Exp(lp \sim N(\mu,\sigma^2)) = Exp(\mu).exp(\frac{1}{2}.\ \sigma^2)$
- For an unbiased estimator we need to consider the random variable $\Gamma = lp - \frac{1}{2}.\ \sigma^2$.

  $\Gamma \sim N(\mu - \frac{1}{2}.\ \sigma^2, \sigma^2)$

  $E(exp(\Gamma)) = exp(\mu - \frac{1}{2}.\ \sigma^2).exp(\frac{1}{2}.\ \sigma^2) = exp(lp)/exp(\frac{1}{2}.\ \sigma^2) = exp(\mu)$

Risk Premium Distribution

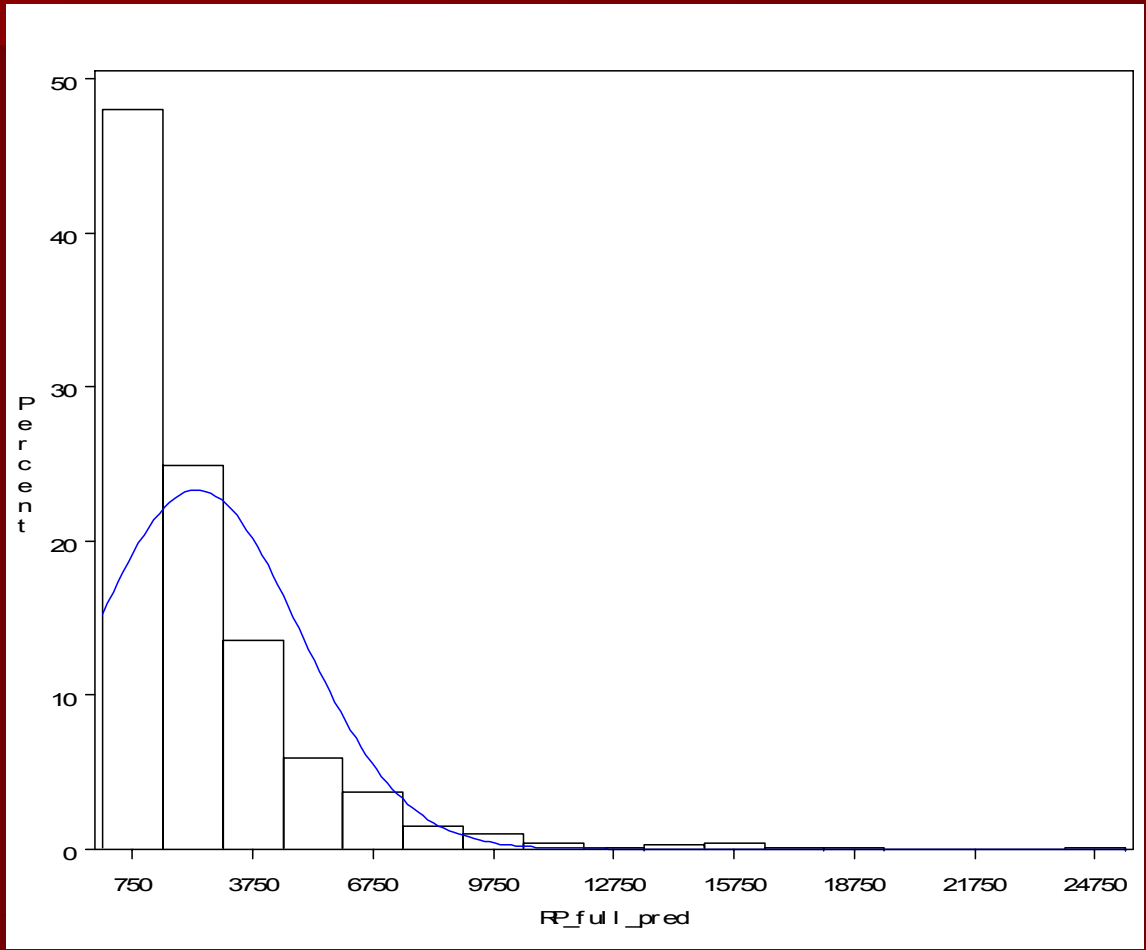


Full Predictive Distribution

- In order to examine the predictive distribution we need to add in the process variance. Each element of the generated empirical distribution is mean (μ) of a Tweedie distribution with variance $V(\mu) = \sigma^2\mu^{1.25}$.
- In order to generate an observation from the predictive distribution we need to generate an observation from this Tweedie distribution.
- A simplified approach is to assume that the observation comes from a Gamma distribution with mean $\alpha.\beta=\mu$ and variance $\alpha.\beta^2=\sigma^2\mu^{1.25}$. (solving gives a $\Gamma(1/\sigma^2\mu^{0.75}, \sigma^2\mu^{0.25})$ distribution)

index	P1(1)	P2(1)	P4(1)	P5(3)	P6(1)	P7(1)	LP	rp_para_error	Bias Adj	alpha	beta	RP_full_pred
1	5.6811	0.4612	0.0530	0.2548	1.0693	0.0047	7.5242	1345	1.38	2.22	605.60	336
2	5.6811	0.4612	0.0530	0.2548	1.0693	0.0047	7.5242	1345	1.38	2.22	605.60	37
4	6.5710	0.3705	0.0913	0.6565	1.0721	0.0569	8.8073	4853	1.38	5.81	834.63	2345
5	5.1734	0.4560	-0.0011	0.0438	1.1910	0.0486	6.9116	729	1.38	1.40	519.61	827
6	5.1734	0.4560	-0.0011	0.0438	1.1910	0.0486	6.9116	729	1.38	1.40	519.61	1156
7	6.2888	0.4997	0.2177	0.3145	1.0165	0.0387	8.3858	3184	1.38	4.24	751.15	3664

Full Predictive Distribution



Selected References

- Brent Jorgensen and Marta C. Paes Souza (1994). Fitting Tweedie's Compound Poisson Model to Insurance Claims Data. Scand. Actuarial Journal.
- Brent Jorgensen (1987). Exponential Dispersion Models. JRSS Series B.
- A.C.Davison (2008). Bootstrap Methods and their Applications. University College Dublin Study Note.
- Rob Kass (2005). Compound Poisson Distribution and GLM's – Tweedie's Distribution. Royal Flemish Academy of Belgium for Science and Arts.
- Peter England (2002) Addendum to "Analytic and Bootstrap Estimates of Prediction Error in Claims Reserving". Insurance, Mathematics and Economics 31.
- Peter England and Richard Verrall (2006). Predicitive Distributions of Outstanding Liabilities in General Insurance. Annals of Actuarial Science Vol1.
- Mike Brockman and Tom Wright (1992). Statistical Motor Rating : Making Effective Use of Your Data. JIA Vol. 119.
- S Margets . Private Correspondence.

Appendix 1

- Exponential Dispersion Densities – For full details see Jorgensen (1987) or Jorgensen and Paes Souza (1994)

Exponential Dispersion Models

- $Y \sim ED(\mu,\sigma^2) \Rightarrow P(y;\theta\ \lambda) = a(\lambda,y).\exp[\lambda\{y\theta - k(\theta)\}]$

  Mean: $\mu = k'(\theta) = \tau(\theta)$

  Dispersion parameter: $\sigma^2 = 1/\lambda$

  Variance : $\sigma^2 V(\mu)$ where $V(\mu) = \tau'(\tau^{-1}(\mu)) = \tau'(\theta) = k''(\theta)$

- Tweedie Distribution

  $Y \sim ED^p(\mu,\sigma^2)$

  Mean: μ

  Variance : $\sigma^2 V(\mu) = \sigma^2.\mu^p$ $P \in (-\infty,0] \cup [1, \infty)$

- Scale Property for Tweedie distribution

  $C.ED^p(\mu,\sigma^2) = ED^p(c\mu,c^{2-p}\sigma^2)$

Exponential Convolution Model

$Y \sim ED(\tau(\theta),\lambda^{-1}) \Leftrightarrow \lambda Y \sim ED^{*}(\theta,\lambda)$

$Z=\lambda y \sim p(z;\theta\ \lambda)=a^{*}(\lambda;z).\exp\{z\theta-\lambda k(\theta)\}$

Model $ED^{*}(\theta,\lambda)$ has mean $m=\lambda\mu$ and variance $\lambda V(\mu) = \lambda V(m/\lambda)$

Model satisfies convolution formula

$ED^{*}(\theta,\lambda_1) * \ldots * ED^{*}(\theta,\lambda_k) = ED^{*}(\theta,(\lambda_1+\ldots+ \lambda_k))$
where * denotes convolution.

Tweedie Model

$N(t) = \text{Poisson}(mt)$ {number of claims}

$Z_t = \Gamma(-\theta,-\alpha)\ z>0\ \alpha,\theta<0$ {individual claim size}

$Z(w) = \Sigma Z_i\ i=1 \text{ to } N(w)$ {aggregate claims – w is exposure}

$Z(w) \sim ED^*(\theta,\lambda^{1-\alpha}w) \Leftrightarrow Z(w)/\lambda^{1-\alpha}w \sim ED^p(\lambda^{\alpha-1}\mu,\lambda^{\alpha-1}/w)$

$\lambda^{1-\alpha}Z(w)/\ \lambda^{1-\alpha}w = Y(w) \sim ED^p(\mu,\sigma^2/w)$ by scale property ($\sigma^2=1/\lambda$)

Variance $(Y(w)) = \sigma^2/w.V(\mu) = \sigma^2/w.\mu^p$ $p = (\alpha-1)/(\alpha-2) \in (1,2)$