Benchmark Rates for Excess of Loss Reinsurance Programs

A Generalised Non-Linear Quasi-Likelihood approach

Robert Verlaak[a,c,+], Werner Hürlimann[b] & Jan Beirlant[c]

[a] Aon Risk Consulting Brussels
[b] Aon Re and IRMG Switzerland
[c] Department of Mathematics KU-Leuven
[+] Corresponding author

September 5, 2005

Problem

- Summarise the Motor Third Party Liability excess of loss rates of different layers (and for different cedents) in one Market benchmark model
- Available data:
  - Belgium market: 4 years (2000-2004)
    172 layers
  - Swiss Market: 3 years (2001-2004)
    31 layers

Way of working

- Describe the rates with a (commercial) **Compound Model** (loading & discounts included)
- Describe the (transformed) severity with a **Pareto** and/or **Burr** model (hierarchical structure)
- Estimate & judge the unknown parameters within the framework of **Generalized (non-) Linear Models** (GLIM) by **quasi-likelihoods**
- Apply this approach on the rates per individual layer instead of the cumulative rates (independency condition – successive layers are not always know)
- Select the 'best' sub-model (some parameters constant over the years) with a F-test based on the ratio of the scaled deviance

Advantages

- Statistical framework to select sub-models (GLIM – Deviance - Hierarchical structure)
- Answer to the observed **heteroscedasticity** in the residuals (GLIM)
- No need to describe the underlying model for the residuals (quasi-likelihood)
- Theoretical framework versus simple & practical interpretation: **Deviance minimisation** versus **Weighted or Generalised Pearson Residuals minimisation**. (GLIM - quasi-likelihood)
- Rates are divided in 2 components that are related to **number of claims** and (transformed) **severity**.
  This helps to judge the outcome of the parameters; to make indirectly general conclusions about expected number of ceded claims and/or Market behaviour). (Commercial Compound Model)
- Better motivation to choose some hypothesis (e.g. motivation of Variance function to describe the observed heteroscedasticity; motivation of the Benchmark function: **fitting versus modelling**
- Describing **rates** instead of **Rate on Line** (more general) (Compound Model)
- …

Model

- Aggregate claims of a portfolio of insurance risks $S = \sum_{i=0}^{N} X_i$ (number of claims N and $X_i \sim \mathrm{X}$ for all i
- The excess of loss risk after a priority *R* and an upper limit L is equal to: $\aleph(R,L) = \sum_{i=0}^{N}\left[X_i \vee (R,L)\right] = \sum_{i=0}^{N}\left[(X_i - R)_+ \wedge (L-R)\right] = \sum_{i=0}^{N}\left[(X_i - R)_+ - (X_i - L)_+\right]$

  with $E[\aleph(R,L)] = E[N] \cdot \{E[X \wedge L] - E[X \wedge R]\}$
- Assume loading is partially incorporated in the transformed severity X' of X
- The market rate can be described as:

$$b(R,L) = v \cdot (1+l) \cdot E[N] \cdot \{E[X' \wedge L] - E[X' \wedge R]\} / PI$$
$$= d \cdot \{E[X' \wedge L] - E[X' \wedge R]\} \text{ with } 0 \le R < L \le \infty$$

Discount, prop. loading & underlying premium income v, l, PI

Model

- Transformed severity is assumed Pareto(2) or Burr

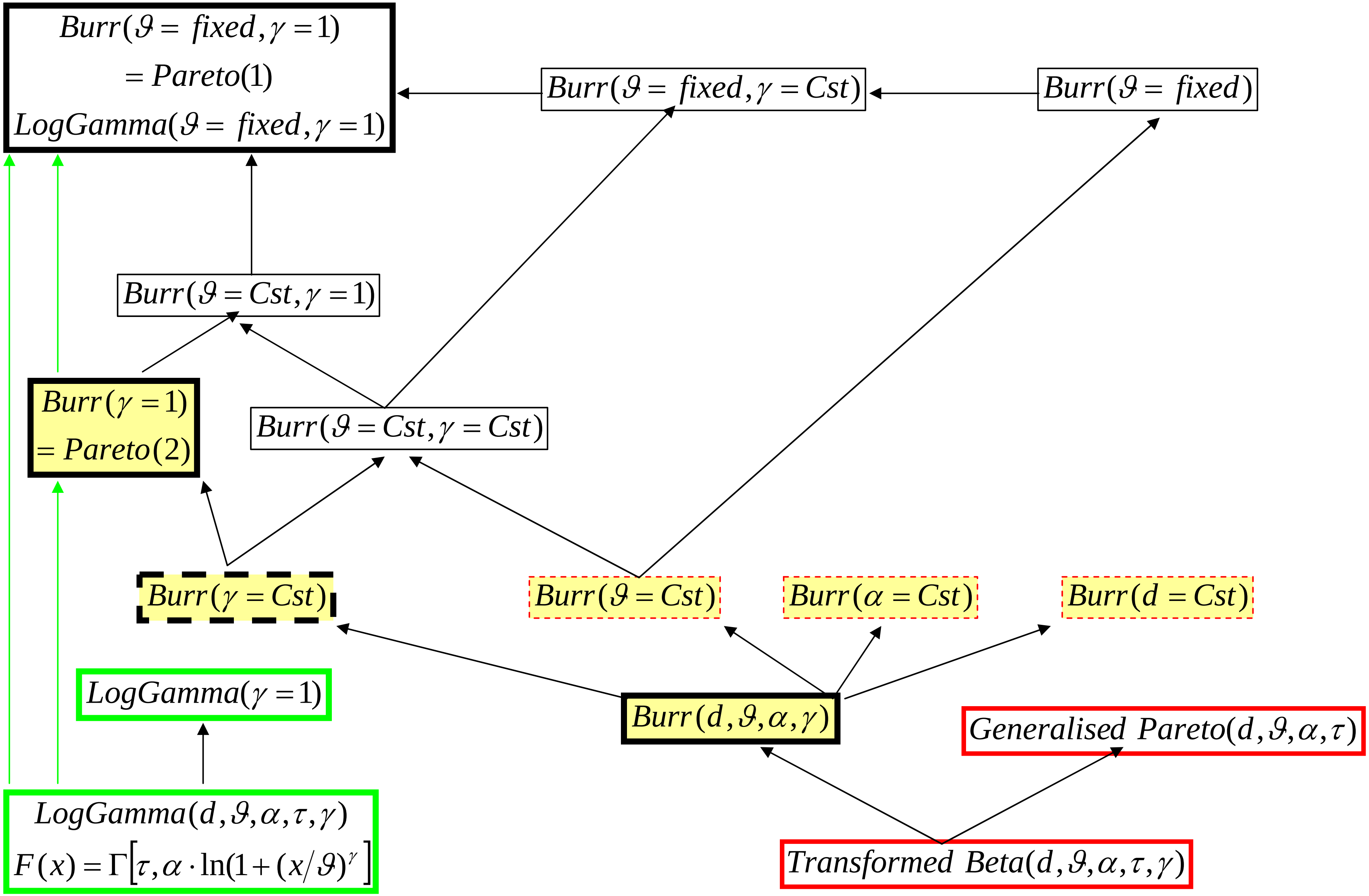
$$b_P(R,L) = \frac{d\cdot\vartheta}{\alpha-1}\cdot\left[\left(\frac{1}{1+R/\vartheta}\right)^{\alpha-1} - \left(\frac{1}{1+L/\vartheta}\right)^{\alpha-1}\right], \quad \alpha \neq 1$$

$$= d\cdot\vartheta\left[\ln\left(\frac{1}{1+R/\vartheta}\right) - \ln\left(\frac{1}{1+L/\vartheta}\right)\right], \quad \alpha = 1$$

$$b_B(R,L) = d\cdot B\cdot\left[\beta\left(\frac{\gamma+1}{\gamma},\frac{\alpha\cdot\gamma-1}{\gamma},\frac{(L/\vartheta)^\gamma}{1+(L/\vartheta)^\gamma}\right) - \beta\left(\frac{\gamma+1}{\gamma},\frac{\alpha\cdot\gamma-1}{\gamma},\frac{(R/\vartheta)^\gamma}{1+(R/\vartheta)^\gamma}\right)\right] +$$

$$d\cdot\left[L\cdot\left(\frac{1}{1+(L/\vartheta)^\gamma}\right)^\alpha - R\cdot\left(\frac{1}{1+(R/\vartheta)^\gamma}\right)^\alpha\right], \text{ with } B = \frac{\theta\cdot\Gamma\left(\frac{\gamma+1}{\gamma}\right)\cdot\Gamma\left(\frac{\alpha\cdot\gamma-1}{\gamma}\right)}{\Gamma(\alpha)}$$

Model


$Burr(\vartheta = fixed, \gamma = 1)$
$= Pareto(1)$
$LogGamma(\vartheta = fixed, \gamma = 1)$
$Burr(\vartheta = fixed, \gamma = Cst)$
$Burr(\vartheta = fixed)$
$Burr(\vartheta = Cst, \gamma = 1)$
$Burr(\gamma = 1)$
$= Pareto(2)$
$Burr(\vartheta = Cst, \gamma = Cst)$
$Burr(\gamma = Cst)$
$Burr(\vartheta = Cst)$
$Burr(\alpha = Cst)$
$Burr(d = Cst)$
$LogGamma(\gamma = 1)$
$Burr(d, \vartheta, \alpha, \gamma)$
$Generalised\ Pareto(d, \vartheta, \alpha, \tau)$
$LogGamma(d, \vartheta, \alpha, \tau, \gamma)$
$F(x) = \Gamma\left[\tau, \alpha \cdot \ln(1 + (x/\vartheta)^{\gamma}\right]$
$Transformed\ Beta(d, \vartheta, \alpha, \tau, \gamma)$

GLIM

- refer to McCullagh & Nelder (1985), Hardin & Hilbe (2001), McCulloch & Searle (2001) or Dobson (2002)
- The rates $r_{j,i}(R_{j,i}, L_{j,i})$ of the XL contracts i of the year $j = 1..J$ priority (retention) $R_{j,i}$ upper limit $L_{j,i}$ assumed to be independent observations.
- To describe the **heteroscedasticity** we make use of a **Variance function** such that $Var\left[r_{j,i}(R_{j,i}, L_{j,i})\right] = \phi \cdot \mu_{j,i}{}^{k}$
- The expected value $\mu_{j,i} = E\left[r_{j,i}(R_{j,i}, L_{j,i})\right]$ is supposed to be equal to $b_P(R_{j,i}, L_{j,i}, d_j, \vartheta_j, \alpha_j)$ or $b_B(R_{j,i}, L_{j,i}, d_j, \theta_j, \alpha_j, \gamma_j)$
- We will make no assumption about the distribution function of the observed rates => **quasi-likelihood.**
  Maximising quasi-likelihood is equivalent with minimising the **Deviance** but also to **weighted least squares minimisation**

GLIM

- The quasi-likelihood involves a sum of n contributions $l^q(\boldsymbol{\mu};\mathbf{y}) = \sum_{i=1}^{n} l^q(\mu_i; y_i)$ and the individual components satisfy the differential equation $\partial l^q(\mu_i; y_i)/\partial\mu_i = (y_i - \mu_i)/V(\mu_i)$
  For $V(\mu) = \phi\cdot\mu^k, \phi > 0$ we have that:

$$
\begin{aligned}
l^q(\mu,\phi,y) &= (y-\mu)^2/(2\phi), && \text{Gaussian case } (\phi=\sigma^2) \quad \text{if } k=0 \\
&= \left(y\cdot\ln(\mu)-\mu\right)/\phi, && \text{Poisson case } (\phi=1) \quad \text{if } k=1 \\
&= \left(-y/\mu-\ln(\mu)\right)/\phi, && \text{Gamma case } (\phi=1/\text{scale parameter}) \quad \text{if } k=2 \\
&= \left(\mu\cdot y/(1-k)-\mu^2/(2-k)\right)/(\phi\mu^k), && \text{if } k\neq 1 \text{ or } 2
\end{aligned}
$$

- **Scaled deviance** is LogLikelihoods ratio of the model under investigation and the saturated model: $SD(\boldsymbol{\mu},\phi,\mathbf{y}) = -2\left[l^q(\boldsymbol{\mu},\phi,\mathbf{y}) - l^q(\mathbf{y},\phi,\mathbf{y})\right]$
- The **deviance** D is defined as $D(\boldsymbol{\mu},\mathbf{y}) = \phi\cdot SD(\boldsymbol{\mu},\phi,\mathbf{y})$
- **Maximising $l^q(\boldsymbol{\mu};\mathbf{y})$ is independent of ϕ & is equivalent with minimising D**

GLIM

- To examine the parameters $\widehat{\boldsymbol{\beta}}_{\mathbf{p}}, \widehat{\boldsymbol{\beta}}_{\mathbf{q}}$ $\quad p > q > 0$ of 2 nested models, each estimated in their corresponded models, one has that the distribution of the deviance fulfils the relation

$$D(\widehat{\boldsymbol{\beta}}_{\mathbf{p}}, \widehat{\boldsymbol{\beta}}_{\mathbf{q}}) = D(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}_{\mathbf{p}}), \mathbf{y}) - D(\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}_{\mathbf{q}}), \mathbf{y}) \sim \phi \cdot \chi^2_{p-q} + O_p(n^{-1/2})$$

- This leads to a chi-squared test: $\chi(1) = \dfrac{D(\widehat{\boldsymbol{\beta}}_{\mathbf{p}}, \widehat{\boldsymbol{\beta}}_{\mathbf{q}})}{\phi} \sim \chi^2_{p-q}(p-q)$

  if we estimate ϕ by $\widehat{\phi} = (\mathbf{y} - \widehat{\boldsymbol{\mu}})^T \cdot \mathbf{V}^-(\widehat{\boldsymbol{\mu}}) \cdot (\mathbf{y} - \widehat{\boldsymbol{\mu}}) / (N-p) = X^2 / (n-p)$

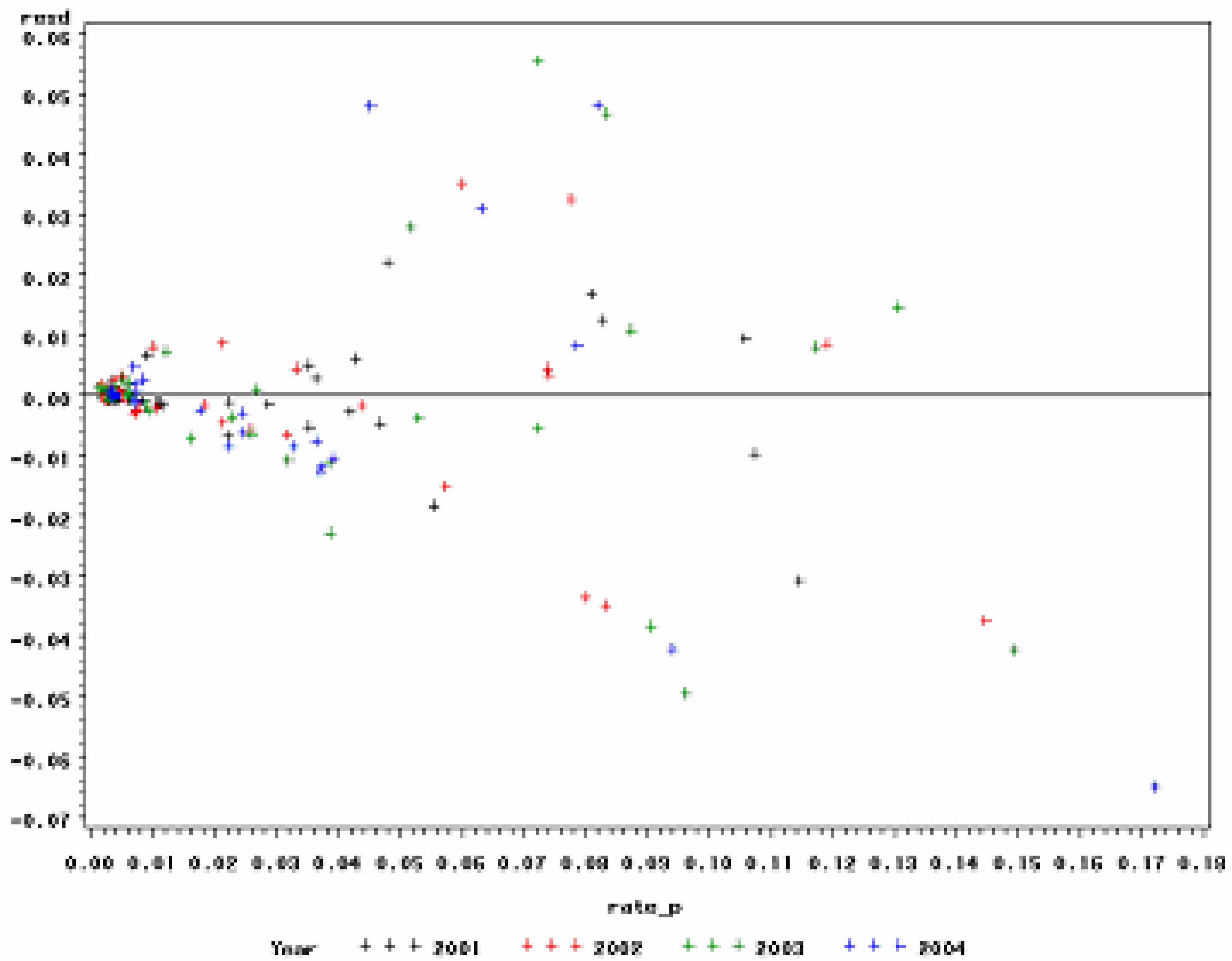
- Leads this to a first F-test $F(1) = \dfrac{D(\widehat{\boldsymbol{\beta}}_{\mathbf{p}}, \widehat{\boldsymbol{\beta}}_{\mathbf{q}})}{\hat{\phi}} \sim F(p-q, n-p)$

- A second F-test $F(2) = \dfrac{D(\widehat{\boldsymbol{\beta}}_{\mathbf{p}}, \widehat{\boldsymbol{\beta}}_{\mathbf{q}}) \big/ (p-q)}{D(\widehat{\boldsymbol{\beta}}_{n}, \widehat{\boldsymbol{\beta}}_{\mathbf{p}}) \big/ (n-p)} = \dfrac{SD(\widehat{\boldsymbol{\beta}}_{\mathbf{p}}, \widehat{\boldsymbol{\beta}}_{\mathbf{q}}) \big/ (p-q)}{S\,D(\widehat{\boldsymbol{\beta}}_{n}, \widehat{\boldsymbol{\beta}}_{\mathbf{p}}) \big/ (n-p)} \sim F(p-q, n-p)$

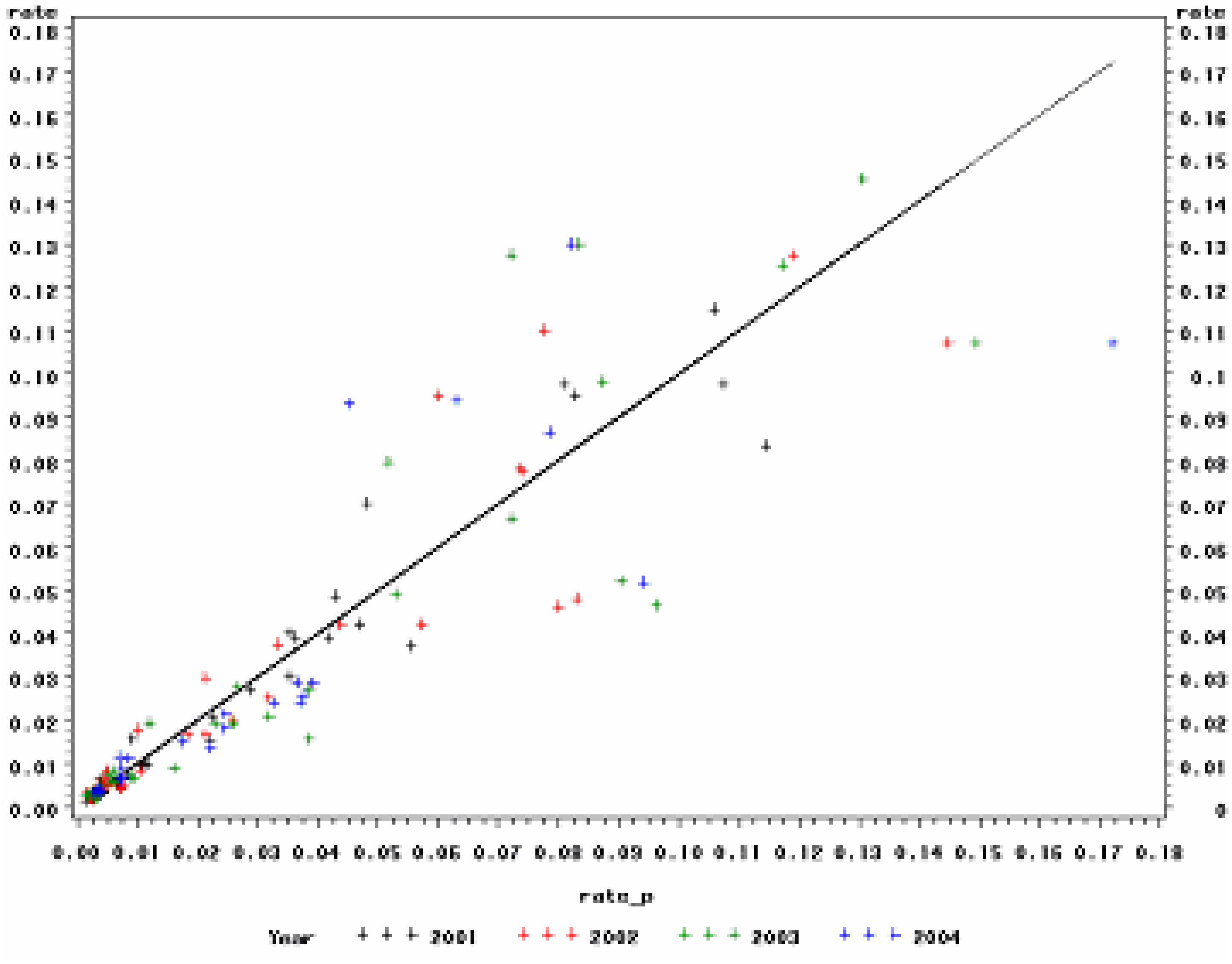
Belgian Motor Third Party Liability data

Reinsurance Years	2001	2002	2003	2004	Total
Number of layers	34	48	53	37	172
Number of unlimited layers	15	19	20	16	70
Number of cedent	15	19	20	16	70
Number of layers per program	**2001**	**2002**	**2003**	**2004**	**Total**
4 layer program	4	8	4		16
3 layer program	12	24	39	27	102
2 layer program	16	14	8	6	44
1 layer program	2	2	2	4	10
Total	34	48	53	37	172
Number of cedent	**2001**	**2002**	**2003**	**2004**	**Total**
4 layer program	1	2	1		4
3 layer program	4	8	13	9	34
2 layer program	8	7	4	3	22
1 layer program	2	2	2	4	10
Total	15	19	20	16	70

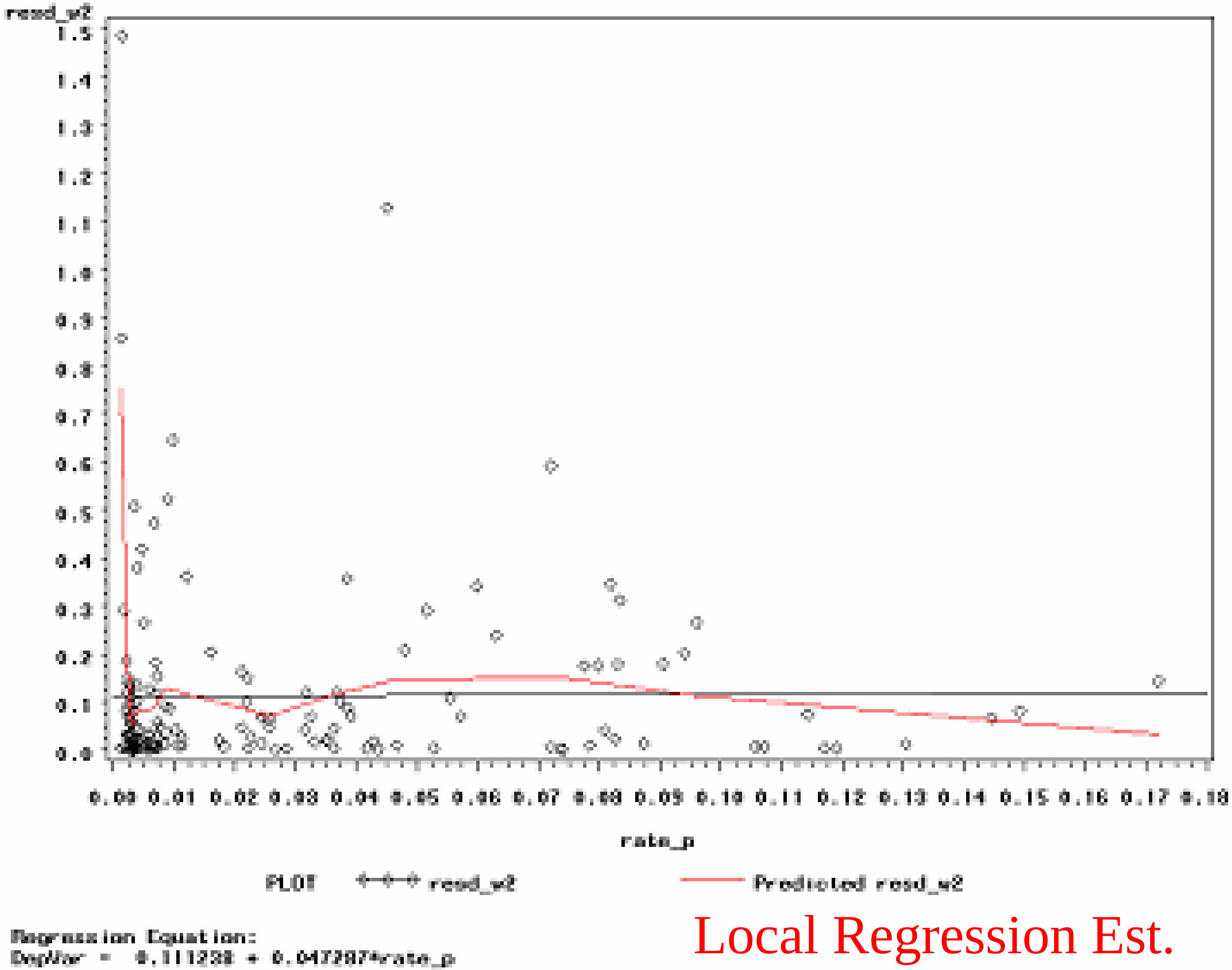
Burr full model & heteroscedasticity



Burr full model & heteroscedasticity

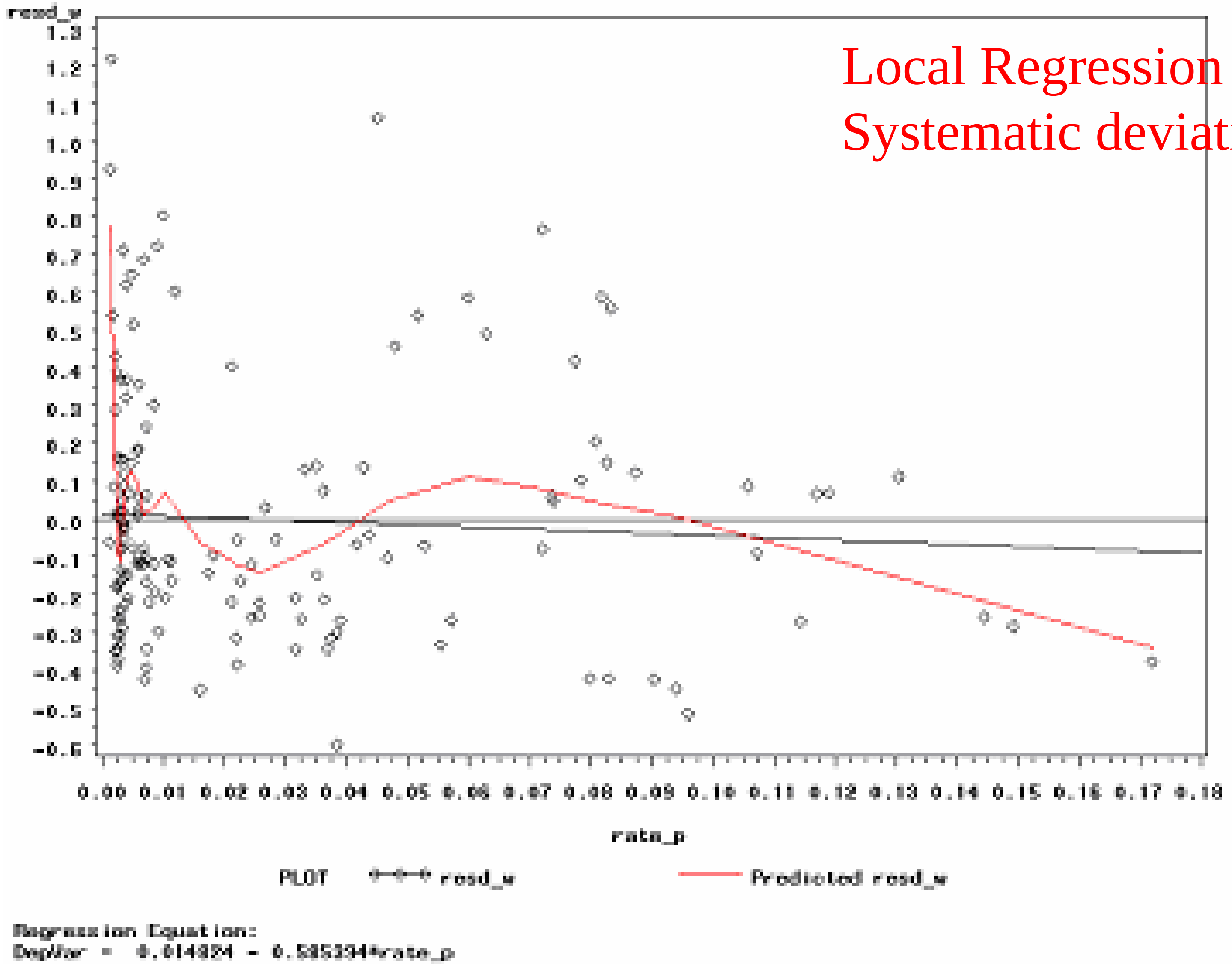


Heteroscedasticity – Variance function k=2



Local Regression Est.

Heteroscedasticity – Variance function k=2



Local Regression Est.
Systematic deviation?

GLIM = weighted least squares regression

- Parameter estimation: minimising the deviance or a weighted least squares regression

$$\sum_{j=1}^{J}\sum_{i=1}^{N_j} w_{j,i}\cdot(b_{j,i}-r_{j,i})^2=\sum_{j=1}^{J}\sum_{i=1}^{N_j} w_{j,i}\cdot\left[b(R_{j,i},L_{j,i},d_j,\vartheta_j,\alpha_j,\gamma_j)-r_{j,i}(R_{j,i},L_{j,i})\right]^2$$

  with the weights $w_{j,i}=b^{-k}(R_{j,i},L_{j,i},d^*_j,\vartheta^*_j,\alpha^*_j,\gamma^*_j)$
  evaluated at the (unknown) optimal parameter values $\{d^*_j,\vartheta^*_j,\alpha^*_j,\gamma^*_j\}$

- Not the same solution minimising $\displaystyle\sum_{j=1}^{J}\sum_{i=1}^{N_j}\frac{\left[b(R_{j,i},L_{j,i},d_j,\vartheta_j,\alpha_j,\gamma_j)-r_{j,i}(R_{j,i},L_{j,i})\right]^2}{b^k(R_{j,i},L_{j,i},d_j,\vartheta_j,\alpha_j,\gamma_j)}$
- Lower rates (= higher XL-covers) become more weight (fitting from the **right** instead of **left**)
- The first minimisation will give the same solution as obtained by minimisation of the corresponding **deviance**
- The second is focussed on the minimisation of the **Generalised Pearson Residuals** (direct practical interpretation)

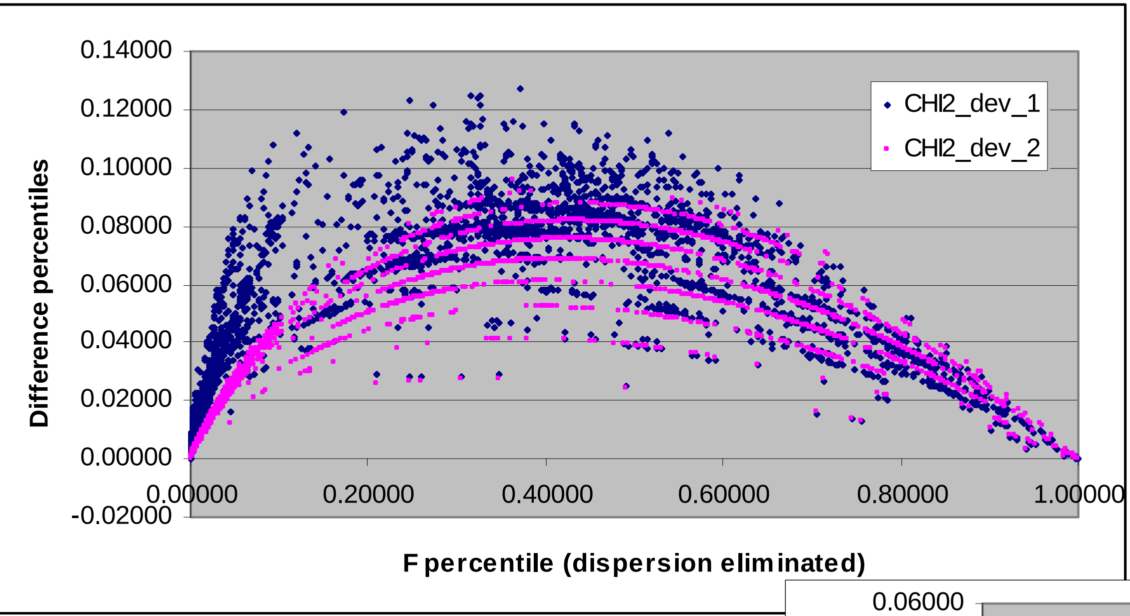
Independency of the rates $r_{j,i}(R_{j,i}, L_{j,i})$?

- Estimation of the correlation between the residuals of the rates of successive layers of the reinsurance program for a cedent per year.

  The residuals are obtained by correcting with the full Burr model.

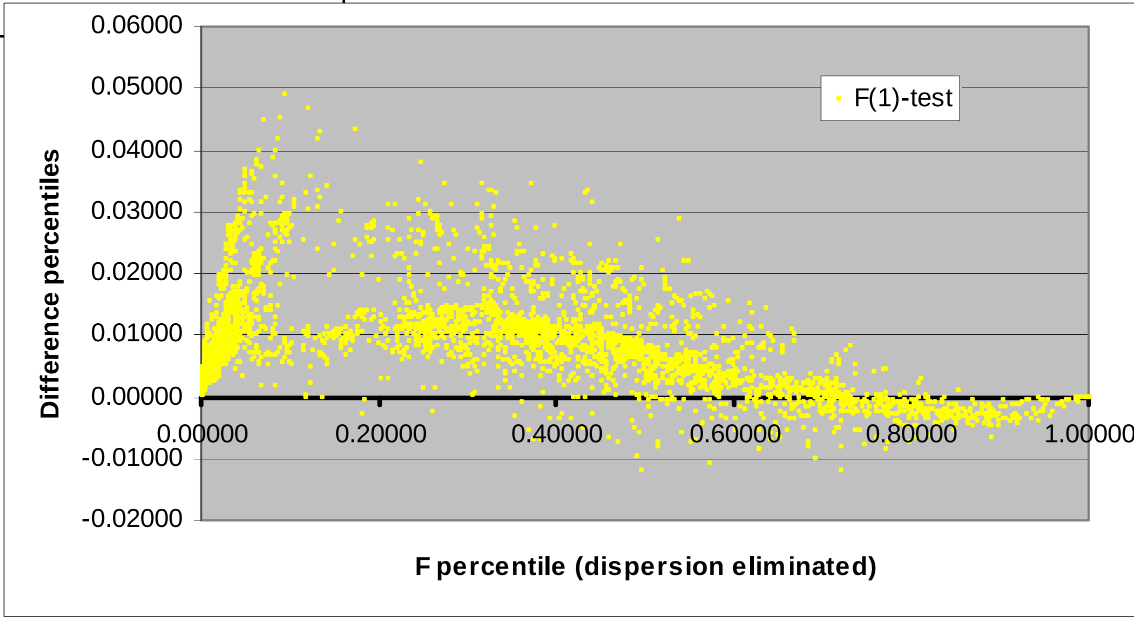
  The estimated correlation is equal to **"–0.0196"** and the 95% confidence interval is equal to **(-0.213260 , 0.175546)**. (estimation is based on 102 couples).

Model selection F(2) test

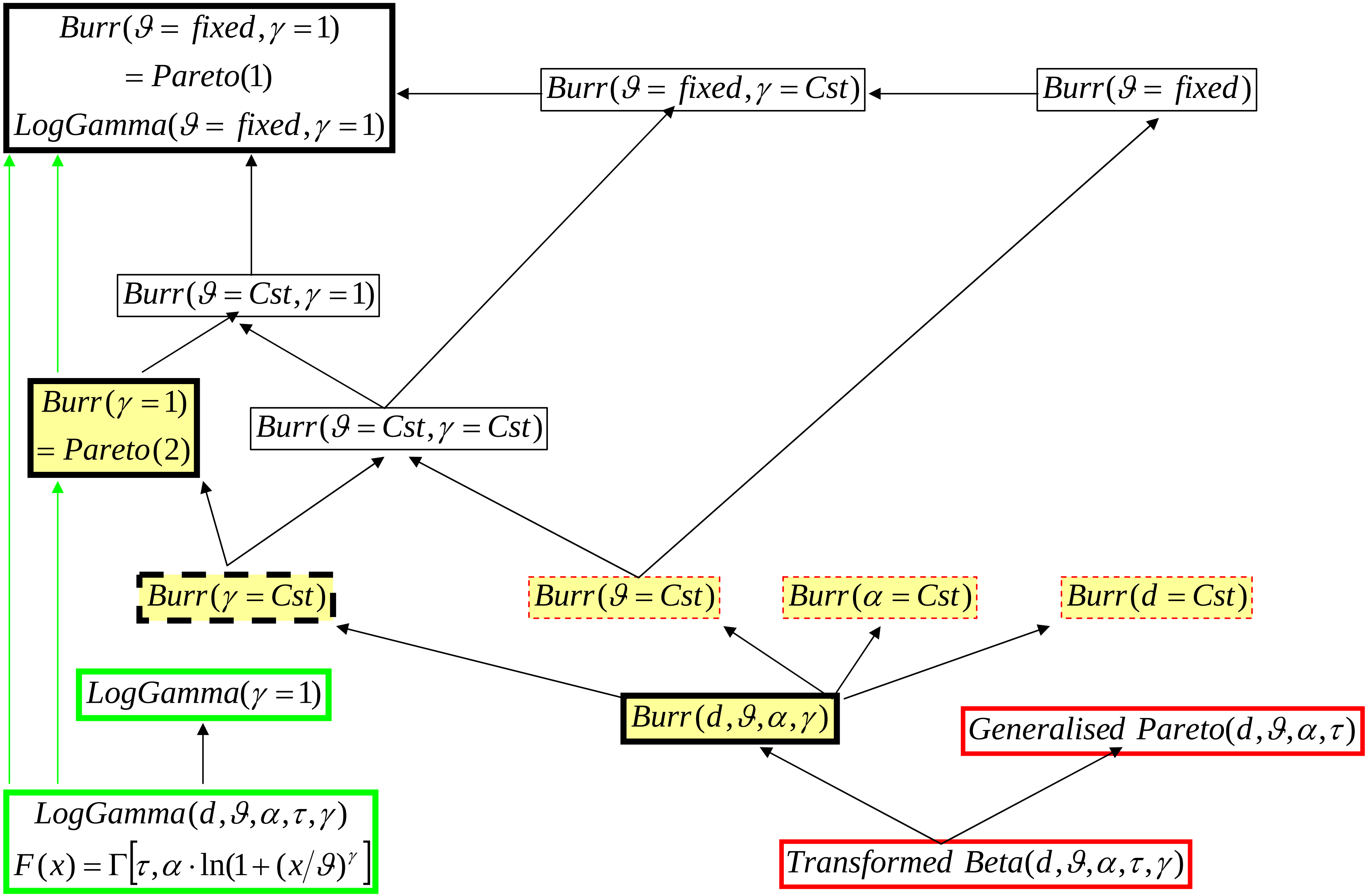


Look only for sub-models with one or more parameters equal for successive years (4,608 – 54,000)

Look for sub-models that are nearly as good as the full Burr model =>
Look for highest percentiles and smallest sub-models



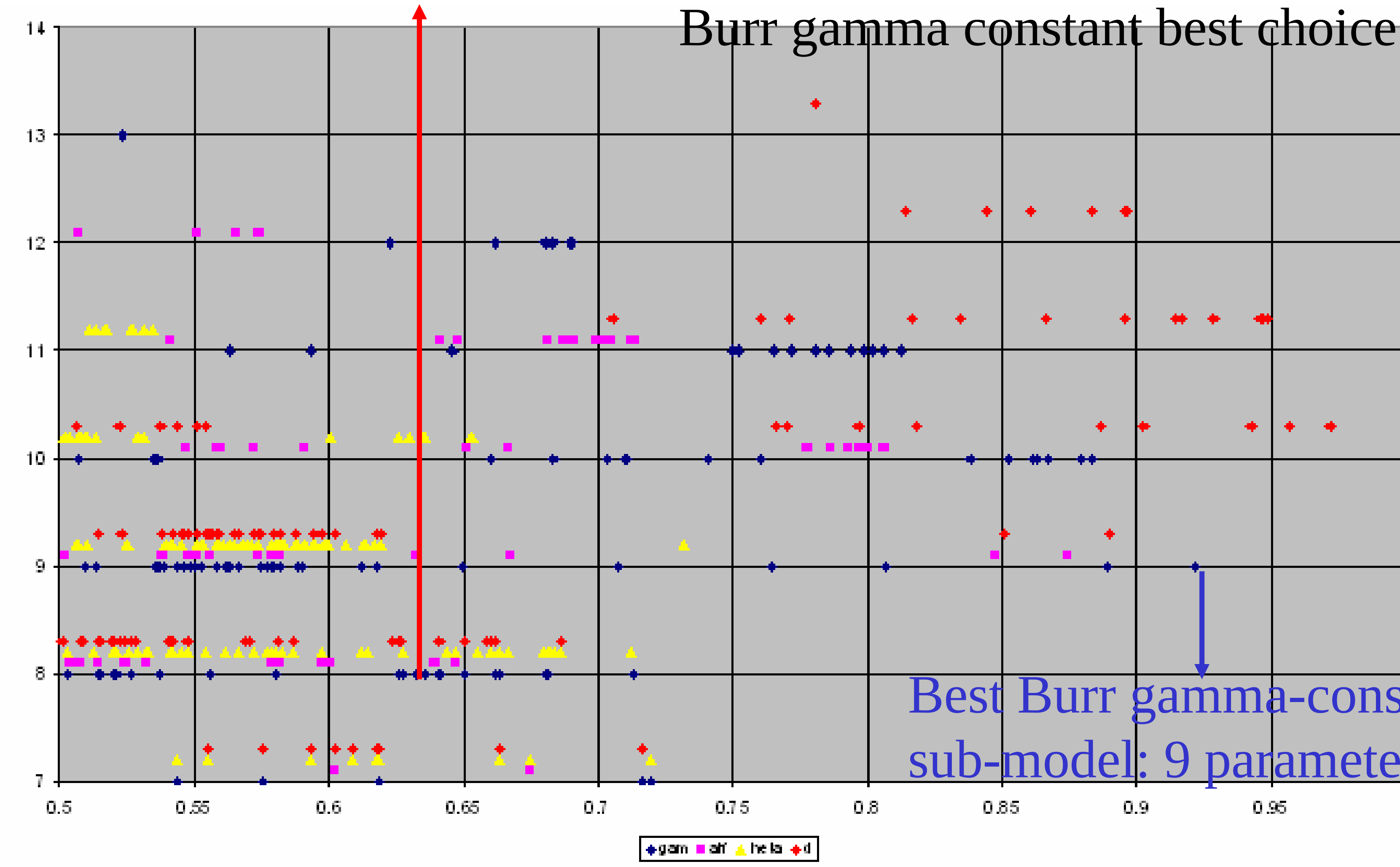
Model selection



Model selection F(2) test

Best Pareto(2) sub-model
8 parameters

Full Burr overestimated?
Pareto(2) (very) good
Burr gamma constant best choice

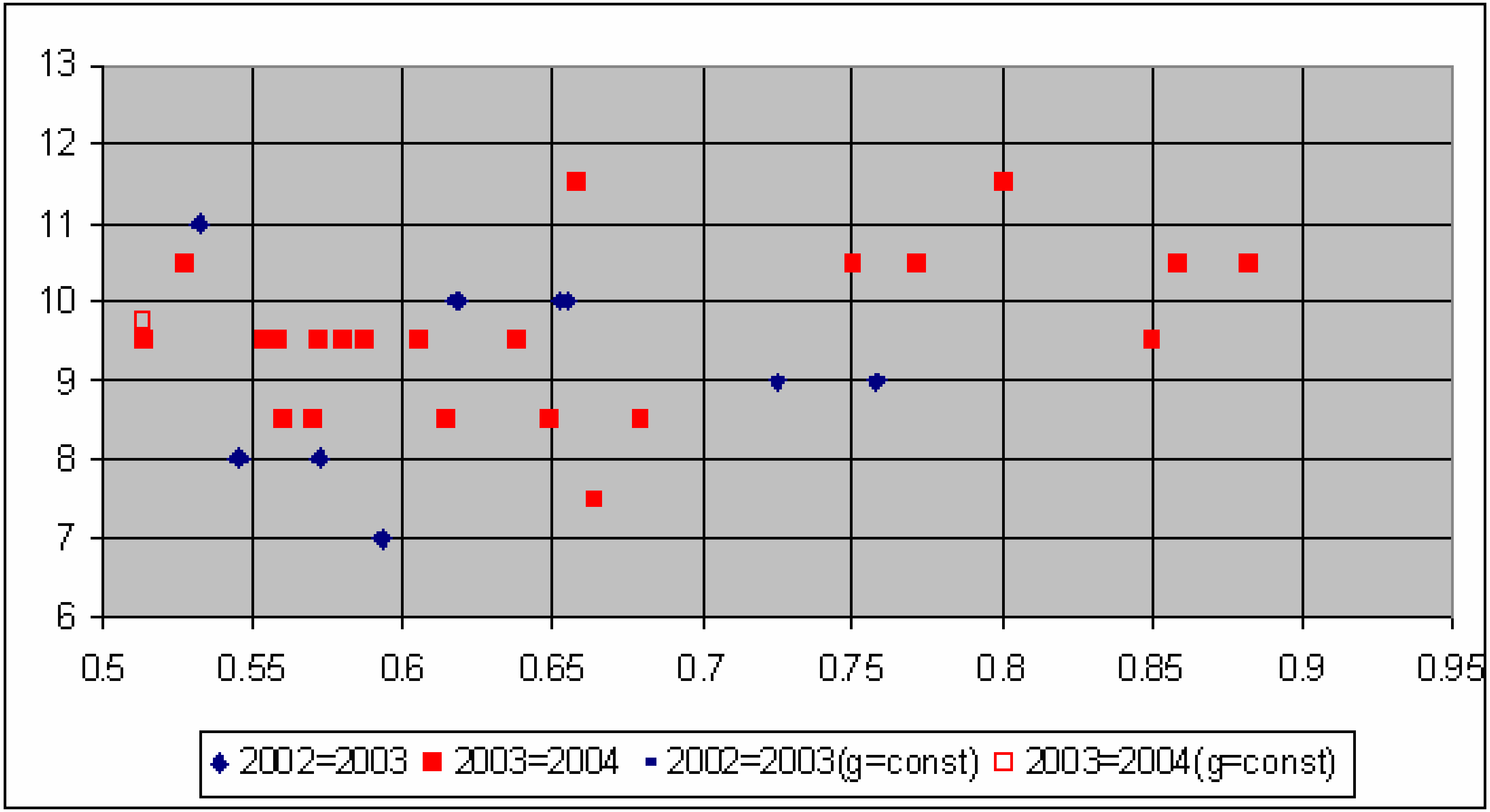


Best Burr gamma-constant
sub-model: 9 parameters

Model selection F(2) test

Best Pareto(2) sub-model
8 parameters

Best Burr γ -constant sub-model: 9 parameters



Annex: the XL-rate variance hypothesis

- A simple estimate for the rate for an excess of loss layer (without loading)

$$\hat{r}(R,L) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\sum_{i=0}^{N}\left[X_{i,j} \vee (R,L)\right]}{PI_j}\right)$$

  for i & j $X_{i,j} \sim \mathrm{X}$ and $PI_j = PI$

- The variance of the estimated rate is then equal to:

$$\left[\beta_1 + \beta_2 \cdot CoV^2(X \vee (R,L))\right] \cdot r^2(R,L)$$

- One can expect that k lies in the interval [0,2]
- Variance not only comming from random fluctuations
- Analysis of the residuals has to confirm the final choise

Annex: relation size company & XL retention

- From practice one knows that:
  retained priority $R(x)\uparrow$ if $PI = x\uparrow$ (non-proportional)
  rule of thumb: $1 \le E\lfloor N_{R(x)} \rfloor \le 4$

$$E\left[N_{R(x)}\right] = E[N]\cdot \overline{F}[R(x)] = d\cdot x\cdot \overline{F}[R(x)] \stackrel{?}{=} c$$

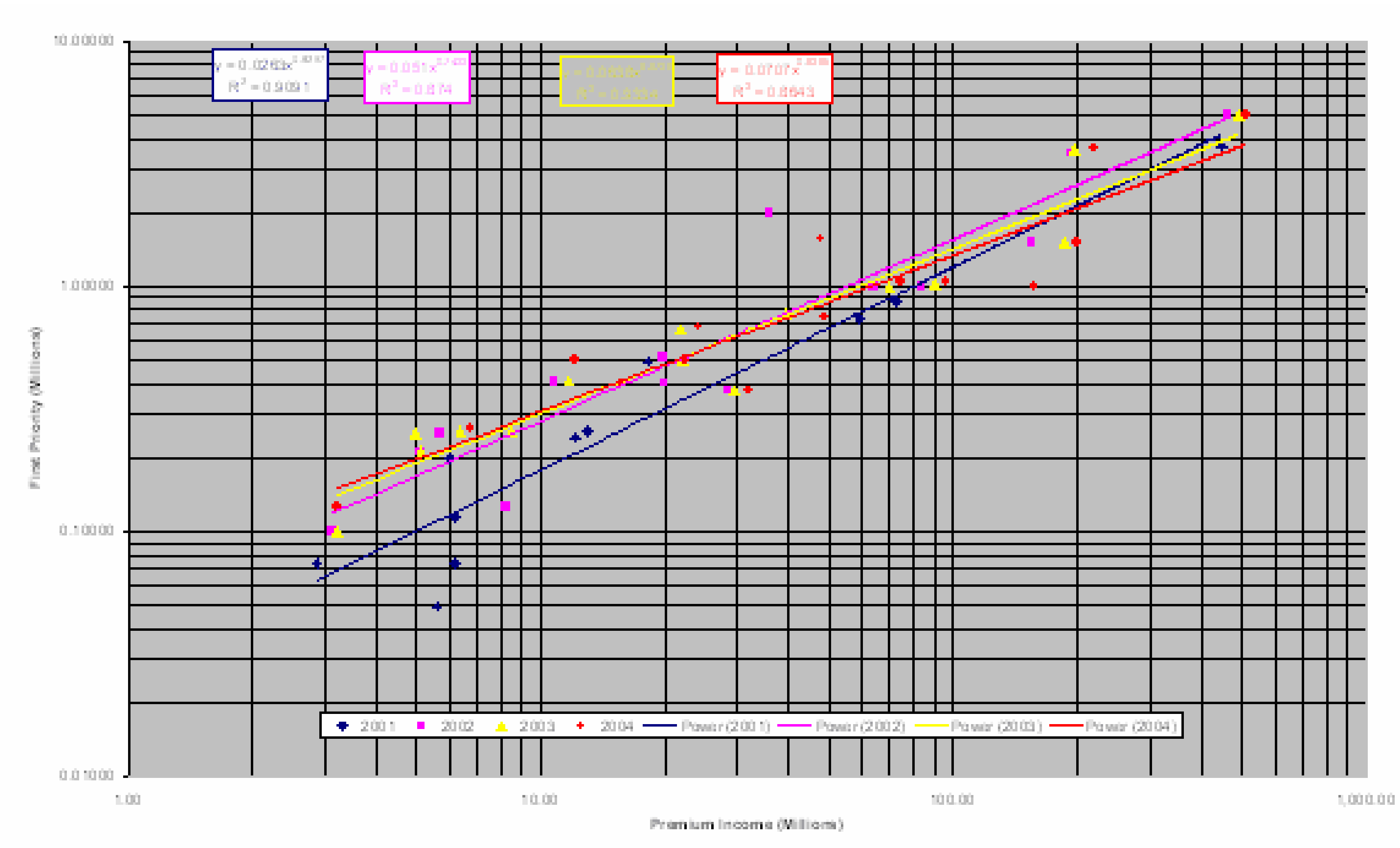
$$\Leftrightarrow$$

$$\overline{F}[R(x)] \stackrel{?}{=} \frac{c}{x\cdot d}$$

- For Burr: $R(x) = \vartheta\left[\left(\frac{x\cdot d}{c}\right)^{\beta} - 1\right]^{1/\tau} = \left[a\cdot x^{\beta} + b\right]^{\delta}$
- For Pareto(2) $R(x) = a\cdot x^{\beta} + b$
- For Pareto(1) $R(x) = a\cdot x^{\beta}$ = standard power function

- Graphical check for the Belgian MTPL data, with a standard power function => seems to confirm the rule of thum

Annex: relation size company & XL retention



Annex: relation size company & XL retention

- Indirect estimate of the average number of claims (biased with the discounted loading factor) ceded to the reinsurance market

	2004		2003
(a) Estimated number of ceded claims with best γ model			
Mean	2.615		2.330
Median	2.576		2.047
b) Estimated number of ceded claims with best γ model & pro rata corrected with individual rate in respect of Market rate			
Mean	2.211		2.053
Median	2.397		2.127
Average of (a) & (b)			
Mean	2.413		2.191

Annex: Rate on Line method

- Rate on Line $RoL(R,L) = a \cdot \left\{ \frac{E[X' \wedge L] - E[X' \wedge R]}{L-R} \right\}$ $a = \nu \cdot (1+l) \cdot E[N]$

  Rate on line intensity $RoL(R) \overset{d}{=} \lim_{L \downarrow R} RoL(R,L) = a \cdot \frac{\partial E[X' \wedge R]}{\partial R} = a \cdot \overline{F}(R)$

  For Burr $RoL(R) = B \cdot (R^{\gamma} + \theta^{\gamma})^{(-\alpha)}$ $B = a \cdot \theta^{\alpha \cdot \gamma}$

  For Pareto(2) $RoL(R) = B \cdot (R + \theta)^{(-\alpha)}$ $B = a \cdot \theta^{\alpha}$

  For Pareto(1) $RoL(R) = B \cdot R^{(-\alpha)}$

- Approximation in practice

  $RoL(R,L) = RoL(P)$ with $P = \sqrt{R \cdot L}$ or $P = \beta \cdot R + (1-\beta) \cdot L, \beta \in (0,1)$ $\beta = 1/2$ or $= 1/3$

  In practice one will use a standard power function for RoL(P) => Pareto(1)