QUASI-LIKELIHOOD ESTIMATION OF BENCHMARK RATES FOR EXCESS OF LOSS REINSURANCE PROGRAMS

BY

ROBERT VERLAAK, WERNER HÜRLIMANN AND JAN BEIRLANT

ABSTRACT

In this paper a method for determining benchmark rates for the excess of loss reinsurance of a Motor Third Party Liability insurance portfolio will be developed based on observed market rates. The benchmark rates are expressed as a percentage of the expected premium income that is available to cover the whole risk of the portfolio. The rates are assumed to be based on a compound process with a heavy tailed severity, such as Burr or Pareto distributions. In the absence of claim data these assumptions propagate the theoretical benchmark rate component of the regression model.

Given the whole set of excess of loss reinsurance rates in a given market, the unknown parameters are estimated within the framework of quasi-likelihood estimation. This framework makes it possible to select a theoretical benchmark rate model and to choose a parsimonious submodel for describing the observed market rates over a 4-years observation period. This method is applied to the Belgian Motor Third Party Liability excess of loss rates observed during the years 2001 till 2004.

KEYWORDS

Excess-of-loss reinsurance, Burr distribution, Pareto distribution, benchmark rate, generalised (non-) linear models, quasi-likelihood, weighted Pearson residuals.

1. INTRODUCTION

In the reinsurance market, the different companies compete in order to offer excess-of-loss contracts at the best possible prices given their own portfolio structure and the varying market conditions, as for example the cyclic transition from soft to hard markets and vice versa. It is therefore quite important to analyze market benchmarks.

It is natural to suppose that the market prices of excess-of-loss contracts are based on a compound modelling of the reinsured claims. The compound model is based on two components: the number of claims and the severity per

Astin Bulletin 39(2), 429-452. doi: 10.2143/AST.39.2.2044642

claim. The obtained expected values of the contracts are further loaded to a commercial rate and discounted with some rate depending on the long tail structure. Here the negotiation mechanisms of the reinsurance market are taken into account. The loading and discounting is not necessarily proportional to the expected claim cost. In most cases one may accept that the loadings relative to the expected claim cost increase for higher layers. This implies that one cannot expect to model the components of the underlying claim model but only a compound model of the commercial rates. This commercial compound model will also exist out of two components: one component combining the expected number of claims with the proportional loading factor (including the discount factor, not necessarily larger than 1), and secondly a transformed severity. Note that the distribution of the loading structure over the two components cannot be reconstituted from observed market rates. It also implies that the change of the excess of loss tariff over time should be analysed by comparing the commercial compound models. It is important to judge the significance of these differences.

In this paper we discuss the regression modelling of market rates on the basis of available rate data over a short observation period. In this way a summary of the market and the differences in the structure of the premium rates from one year to another are obtained. We stress here that most commonly in this setting the individual claim data are not available. Severity models can then only be used to propagate a rate model which will be considered here as a regression function. Once a commercial compound model is selected, one can easily estimate or predict the benchmark rate for any excess of loss cover. The influence of the priority and the cover of an excess of loss contract on the rates and the rate on lines (ROL's, i.e. the excess of loss premium expressed as a percentage of the excess of loss cover) are exactly known within the context of the selected model. This stands in contrast with other methods that are often used in practice that are only based on one covariate being a (weighted) average of the priority and (upper-) limit of the excess of loss contract.

We make use of quasi-likelihood regression theory as rooted in the theory of generalised non-linear models. This methodology allows for statistical inference in a regression context with heteroscedasticity, which is present in the excess of loss rate data. Moreover within the quasi-likelihood approach one does not need to specify a distribution function for the observed rates but only a model for the mean rates and a mean-variance relationship are needed. Nevertheless quasi-likelihood allows to retain nearly full efficiency compared to maximum likelihood estimation. The mean-variance relationship can be verified from the observed commercial rates, and up to some extent can also mathematically be motivated. References to quasi-likelihood estimation and generalised linear models are Agresti (2002), McCullagh & Nelder (1985), Hardin & Hilbe (2001), McCulloch & Searle (2001) or Dobson (2002).

Here, as an illustration, we approximate the benchmark rates with Pareto type models such as the Pareto and the Burr model. This choice essentially stems from the knowledge that the Pareto models are of the most favourite in

practice to describe the underlying claims (see e.g. Schmitter, 1978, Schmitter & Bütikofer, 1997, Doerr, 1980, Schmutz & Doerr, 1998). The analysis of the presented case study will show that this will lead to a fair estimate. The uses of the more general Burr model is natural in the loading structure context discussed above as a generalization of the Pareto models, and for describing heavy tailed data (see e.g. Beirlant et al., 2005). However other models could be suggested, such as the loggamma or the lognormal models (see e.g. Mack, 1997).

The quasi-likelihood theory supposes independent observations. This is the most important reason not to work with the cumulative rates of the combined excess of loss contracts. However, successive layers of the same underlying portfolio are related to each other. At the other hand, the rates of these successive layers in general are not defined by the same reinsurer. Also the highest excess of loss covers is strongly determined by more or less flat market rates. Both observations indicate that the independence assumption is not strongly violated. A further reason not to work with cumulative rates is the fact that the successive layers are not always specified in the data set such that the cumulative rate approach will lead to loss of information.

In section 2 we recollect some facts from excess of loss reinsurance. Section 3 deals with the quasi-likelihood approach as applied in the present context. Finally the proposed method is applied to Belgian Motor Third Party Liability data in section 4.

2. Excess of loss re-insurance

For a more detailed description of excess of loss reinsurance we refer to Carter et al. (2000), Gerathewohl (1980, 1982), Kiln (1982) and Verlaak and Beirlant (2003). Working in the framework of the classical *collective risk model*, we assume that the *aggregate claims* of a portfolio of insurance risks are described by the random variable S defined by a compound process $S = \sum_{i=0}^{N} X_i$ with N the random variable of the number of claims and the individual claims X_i being distributed as X for all i. The excess of loss risk after a priority R and an upper limit L is given by $\aleph(R,L) = \sum_{i=0}^{N} [X_i \vee (R,L)]$, with $X \wedge R = \min(X,R)$ and $X \vee (R,L) = (X \wedge L) - (X \wedge R)$. From classical risk theory we know that (Klugman et al. (1998)

$$E[\aleph(R,L)] = E[N] \cdot \{E[X \wedge L] - E[X \wedge R]\}. \tag{1}$$

We are considering market rates. This implies that some unknown loading and discounting structure is applied to the underlying compound process. As mentioned in the introduction, we assume that some part of the loading structure is incorporated in the transformed severity X.

Further we will assume that a market rate $b(R,L)$ can be described as

$$\begin{aligned} b(R,L) &= v \cdot (1+l) \cdot (E[N]/PI) \cdot \{E[X \wedge L] - E[X \wedge R]\} \\ &= d \cdot \{E[X \wedge L] - E[X \wedge R]\} \text{ with } 0 \le R < L \le \infty \end{aligned} \tag{2}$$

where PI, v, l denote respectively the premium income for the underlying insurance portfolio, the discount rate, and the (proportional) loading factor.

Due to the fact that the priority R and the upper limit L do not influence the expected number of claims $E[N]$, the factor $E[N]$ can be absorbed in the parameter d together with the discount v and loading factor l.

As discussed in the Introduction, we will further assume that X can be described by a Pareto or a Burr distribution. The Burr distribution being defined through the distribution function $F(x) = 1 - u^{\alpha}$, $u = 1/(1 + (x/\theta)^{\gamma})$. The special case $\gamma = 1$ yields the Pareto distribution. From Klugman et al. (1998) we find that in case of the Burr model the market rate equals

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

$$d \cdot \left[L \cdot \left(\frac{1}{1+(L/\theta)^{\gamma}}\right)^{\alpha} - R \cdot \left(\frac{1}{1+(R/\theta)^{\gamma}}\right)^{\alpha}\right], \text{ with } B = \frac{\theta \cdot \Gamma\left(\frac{\gamma+1}{\gamma}\right) \cdot \Gamma\left(\frac{\alpha\cdot\gamma-1}{\gamma}\right)}{\Gamma(\alpha)} \quad (3)$$

which specifies to the following expression in case of the Pareto distribution:

$$\begin{aligned} b_p(R,L) &= \frac{d\cdot\theta}{\alpha-1} \cdot \left[\left(\frac{1}{1+R/\theta}\right)^{\alpha-1} - \left(\frac{1}{1+L/\theta}\right)^{\alpha-1}\right], \quad \alpha \neq 1 \\ &= d\cdot\theta\left[\ln\left(\frac{1}{1+R/\theta}\right) - \ln\left(\frac{1}{1+L/\theta}\right)\right], \quad \alpha = 1. \end{aligned} \quad (4)$$

Here β refers to the beta distribution. Such assumption of course has to be confirmed by a residual analysis (see for example below in Figure 2 and Figure 3). The bench rates will further be denoted by $b(R,L,d,\theta,\alpha,\gamma)$ when the dependence on the model parameters is discussed.

Remark that the rate is also influenced by an annual aggregate deductible, an annual aggregate limit or a reinstatement (payable pro rata time or pro rata amount or both). To calculate this influence one needs to know the claim number process N (for instance Poisson or Negative Binomial). Analogous to the severity X, one can assume that, for this kind of clauses, a loading structure is also incorporated in a transformed claim number process. But this implies that the expected number of claims can be isolated from the discount and loading factor. This will not be pursued here, as we do not use the claim number process explicitly.

Modelling the rates with a compound process allows to deduce or verify indirectly the claim severity as appreciated by the market. Moreover, the expected number of claims (relative to a proportional factor) ceded to the excess of loss contract can also be estimated. Let $E[N_R]$ denote the expected number of claims in excess of the priority R and $\bar{F}(R) = 1 - F(R)$ the survival function of the severity. Then we have that $E[N_R] = E[N] \cdot \bar{F}(R) = \frac{d}{v\cdot(1+l)} \cdot PI \cdot \bar{F}(R)$. This

means that, once the parameters of $b(R,L,d,\vartheta,\alpha,\gamma)$ are estimated and once one has an acceptable judgment of the discounted loading factor, an estimate of $E[N_R]$ can be deduced. This will be illustrated with the Belgian case study.

Further from Klugman et al. (1998) we find that the coefficient of variation in case of a Burr respectively Pareto severity equals to $\sqrt{\frac{\Gamma(\alpha)\cdot\Gamma(1+2/\gamma)\cdot\Gamma(\alpha-2/\gamma)}{\Gamma^2(1+1/\gamma)\cdot\Gamma^2(\alpha-1/\gamma)}-1}$ and $\sqrt{\frac{\alpha}{\alpha-2}}$. Both are independent of θ and make the γ constant sub-model from this point of view very attractive. Afterwards we will see that for the Belgium Motor Third Party Liability this sub-model is one of the best choices.

The benchmark rates as described before gave a summary of what the market has done after the reinsurance renewal. However it is helpful in practice to have a good reference before starting a new negotiation such that one can easily compare the new proposed rates.

Supposing that the only difference with the subsequent year will be a tariff increase t, a claim inflation c and a homogenous portfolio increase p. Then we have from (2) that the benchmark rate $b'(R,L)$ of the following year can be described by a severity $X' = X\cdot(1+c)$, applied to a portfolio with premium income $PI' = PI\cdot(1+t)\cdot(1+p)$ and an expected claim frequency $E[N'] = E[N]\cdot(1+p)$. Without any additional modelling assumption the next year reference benchmark is then easily calculated from the current benchmark:

$$\begin{aligned} b'(R,L) &= v\cdot(1+l)\cdot(E[N']/PI')\cdot\{E[X'\wedge L]-E[X'\wedge R]\} \\ &= \frac{1+c}{1+t}\cdot d\cdot\left\{E\left[X\wedge\frac{L}{1+c}\right]-E\left[X\wedge\frac{R}{1+c}\right]\right\} = \frac{1+c}{1+t}\cdot b\left(\frac{R}{1+c},\frac{L}{1+c}\right) \end{aligned} \tag{5}$$

Note that the coefficient of variation of the severity is not affected while the expected excess of loss rates will be influenced, even if the tariff change is equal to the claims inflation.

In the context of the Burr (and Pareto) modelling we additionally have that the next year reference severity X' is still Burr (or Pareto) such that the reference benchmark can be described through a set of new parameters:

$$b'_B(R,L) = b(R,L,d',\theta',\alpha',\gamma') = b(R,L,d/(1+t),\,\theta\cdot(1+c),\alpha,\gamma) \tag{6}$$

If one has additional information concerning other expected changes, one can incorporate these also in the parameters. The parameter d is further affected through the expected change in reinsurance loading and the expected change in frequency. If one holds back the γ constant sub-model, one can increase the variability of the severity by increasing the coefficient of variation through the parameter α. This will essentially influence the tail and thus the prices of the higher layers, the price of the lower layers can be kept more or less constant by correcting additionally the parameter d.

Finally remark that a rate of a limited cover is described by the difference of two unlimited covers. Consequently, no constant factor or intercept will be used to describe the market rates. This implies that the rates are supposed to

be additive: the rate of one larger layer can be subdivided in the sum of non overlapping sub layers with the same total coverage. Note that clauses like annual aggregate limits or reinstatements do not satisfy the additive property. However, in practice their influence is (in most cases) very limited.

Further, the benchmark approach as defined above and formulated in (2) implies that for all cedents the claim severity for large claims is supposed to be common for the whole market. Also the expected number of claims (from ground up and biased with the discounted loading factor of the excess contract) is supposed to be constant per unit of premium income. This implies that the number of expected claims is proportional to the premium income. This seems to be a reasonable approximation for Motor Third Party Liability.

However for other branches one could absorb the premium income in the severity, such that the severity as a percentage of the premium income is common for the market. The expected number of claims should then be fixed for the market, which could be a better approximation for windstorm, earthquake or flood.

3. Maximum Quasi-likelihood estimation

When the response distribution Y in a regression setting does not exhibit normal residuals with a constant variance, as it will turn out to be the case in our setting, quasi-likelihood estimation provides an inferential method which works as well or almost as well as maximum likelihood but without having to make specific distributional assumptions. The idea behind the quasi-likelihood is to derive a likelihood-like quantity whose construction requires few assumptions. The derivative of the log-likelihood of Y has expected value 0 and, in case of the exponential family of distributions, possesses a variance which is inversely proportional to the variance function $V = V(\mu)$ that describes *Variance(Y)* as a function of the expected value μ. Then it is straightforward to verify that for some factor ϕ (the dispersion parameter), the ratio $q = \left\{\frac{Y-\mu}{\phi \cdot V(\mu)}\right\}$ satisfies the same conditions as the derivative of the log-likelihood. One then defines the log quasi-likelihood via the contributions $q_i = \int_{y_i}^{\mu_i} \frac{(y_i - t)}{\phi \cdot V(t)}\, dt$ of y_i, $i = 1, \ldots, N$. Here we can refer to McCullagh and Nelder (1985).

The quasi-likelihood function is now considered as a function of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$ and is defined by the system of partial differential equations $\partial l^q(\boldsymbol{\mu}; \mathbf{y})/\partial \boldsymbol{\mu} = \mathbf{V}^{-}(\boldsymbol{\mu}) \cdot (\mathbf{y} - \boldsymbol{\mu})$ where the vector of responses $\mathbf{y}$ of length N has mean $\boldsymbol{\mu}$ and covariance matrix $\phi \cdot \mathbf{V}(\boldsymbol{\mu})$ with the dispersion ϕ being strictly positive and $\mathbf{V}(\boldsymbol{\mu})$ a positive semi-definite matrix whose elements are known function of $\boldsymbol{\mu}$. In addition it is necessary to assume that the systematic part of the model is specified in terms of the mean-value parameter. The systematic part is described in general by $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is a vector of p unknown regression parameters to be estimated. Remark that this relationship is not necessarily linear as it will be the case in our example.

The maximum quasi-likelihood equations for estimating the regression coefficient $\boldsymbol{\beta}$ are given by $\partial l^q(\boldsymbol{\mu}(\boldsymbol{\beta}); \mathbf{y})/\partial \boldsymbol{\beta} = 0$ and can be written as $\mathbf{D}^T \cdot \mathbf{V}^{-}(\boldsymbol{\mu}) \cdot (\mathbf{y} - \boldsymbol{\mu}(\bar{\boldsymbol{\beta}})) = 0$, where the derivative matrix $\mathbf{D} = d\boldsymbol{\mu}/d\boldsymbol{\beta}$ is assumed to have rank p for all $\boldsymbol{\beta}$. This corresponds to a weighted least squares minimisation where the weights depend only on the current estimates of the regression parameters. In our case study we assume that $\mathbf{V}(\boldsymbol{\mu}) = \text{diag}[V(\mu_1), \ldots, V(\mu_N)]$ so that 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 $\partial l^q(\mu_i; y_i)/\partial \mu_i = (y_i - \mu_i)/V(\mu_i)$. Furthermore the quantities $(y_i - \mu_i)/\sqrt{V(\mu_i)}$ are referred to as Pearson residuals. Remark that they possess constant variances under the given assumptions.

The quasi-likelihood approach is naturally imbedded in the theory of generalized linear and non-linear models where the classical linear regression model is extended by linking the systematic regression component to the expected value of the responses by a general link function g. Moreover the distribution of the responses is assumed to belong to the exponential family of distributions. The standard reference to generalized linear models is McCullagh & Nelder (1985).

Remark that another possible estimation approach in the present heteroscedastic setting would consist of using a variance-stabilizing transformation (see for instance chapter 7 in Seber (1977)) of a response Y. The generalized linear model and quasi-likelihood approach however have gained more recognition by now as they provide a richer toolbox for data analysis.

Estimation of the market rates

In the present setting we need the following assumptions.

- The rates $r_{j,i}(R_{j,i}, L_{j,i})$ of the XL contracts $i = 1, \ldots, n_j$ of the year $j = 1, \ldots, J$, depending on the priority $R_{j,i}$ and upper limit $L_{j,i}$, are assumed to be statistically independent.
- The rates $r_{j,i}(R_{j,i}, L_{j,i})$ are observations from a distribution that depends on $R_{j,i}$ and $L_{j,i}$, for which the variance $\text{Var}\lfloor r_{j,i}(R_{j,i}, L_{j,i})\rfloor$ and the expected value $\mu_{j,i} = E\lfloor r_{j,i}(R_{j,i}, L_{j,i})\rfloor > 0$ are related by $\text{Var}[r_{j,i}(R_{j,i}, L_{j,i})] = \phi \cdot \mu_{j,i}^k$, with $\phi > 0$ assumed to be independent of the year j. The parameter k will be assumed to be constant over time. This parameter will be chosen appropriately as discussed below.

The expected value $\mu_{j,i} = E\lfloor r_{j,i}(R_{j,i}, L_{j,i})\rfloor$ is assumed to be equal to $b_B(R_{j,i}, L_{j,i}, d_j, \theta_j, \alpha_j, \gamma_j)$, while the link function g is taken equal to the identity function.

As mentioned above the parameters can be estimated by a weighted least squares regression minimizing $\sum_{j=1}^{J}\sum_{i=1}^{n_j} w_{j,i} \cdot (b_{j,i} - r_{j,i})^2 = \sum_{j=1}^{J}\sum_{i=1}^{n_j} w_{j,i} \cdot [b(R_{j,i}, L_{j,i}, d_j, \theta_j, \alpha_j, \gamma_j) - r_{j,i}(R_{j,i}, L_{j,i})]^2$ over the feasible set of the parameters $\{d_j, \theta_j, \alpha_j, \gamma_j\}$ with the weight $w_{j,i} = b^{-k}(R_{j,i}, L_{j,i}, d_j^*, \theta_j^*, \alpha_j^*, \gamma_j^*)$ evaluated at the (unknown) optimal parameter values $\{d_j^*, \theta_j^*, \alpha_j^*, \gamma_j^*\}$. This implies that the

numerical procedure to solve the minimisation problem has to be re-weighted for each iteration with the weights evaluated at the parameter values of the last iteration. It is important to note that the direct minimisation of

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j}\frac{\left[b\left(R_{j,i},L_{j,i},d_j,\theta_j,\alpha_j,\gamma_j\right)-r_{j,i}\left(R_{j,i},L_{j,i}\right)\right]^2}{b^k\left(R_{j,i},L_{j,i},d_j,\theta_j,\alpha_j,\gamma_j\right)}$$

will not give the same solution due to the fact that the weight function is also incorporated in the objective function, the weight being a function of the parameters, and the derivatives of this function now being explicitly incorporated in the derivative equations. However, only the first minimisation will give the same solution as obtained by quasi-likelihood minimisation. This has the important advantage that one can measure the significance of the difference between several sub-models as will be discussed below. On the other hand if one is less interested in the selection of an appropriate sub-model, but only in the optimal set of parameters given a fixed model structure, one could solve the second minimisation problem. This also has the advantage that one can give an easy interpretation in the sense that one minimizes directly the sum of the squares of the Pearson residuals, assuming that $b_{j,i}^k$ is a good estimate of the variance function.

REMARK 3.1. In the previous formulation we looked only at rates and did not take into account the scale of the insurance company. One could argue that a larger company should have more impact on the benchmark. To achieve this, one could minimize the nominal reinsurance premiums instead of the rates. This would result in multiplying the rates with the corresponding premium incomes $PI_{j,i}$ resulting in the minimisation of $\sum_{j=1}^{J}\sum_{i=1}^{n_j} w'_{j,i}\cdot(b_{j,i}-r_{j,i})^2$ with $w'_{j,i}=\frac{PI_{j,i}^2}{PI_{j,i}^k\cdot b_{j,i}^k}$. Note that for $k=2$ it follows that $w'_{j,i}=w_{j,i}$, and the premium weight will disappear. But the choice $k=1$ will result in $w'_{j,i}=w_{j,i}\cdot PI_{j,i}$, so that the premium is introduced as an additional volume or importance weight.

The premium income plays then the role of a classical weight, in the sense of being proportional to the volume of the observation. This choice could be motivated by interpreting the premium income as a measure that is more or less proportional with the number of insured risks, which seems to be acceptable as long as all the reinsured portfolios are comparable with each other and apply an equivalent tariff. However the heterogeneity of the underlying portfolios could also be an argument to work with weights and to obtain a benchmark that corresponds better with the relative importance within the heterogeneity of the market.

This leads to a natural generalisation using weights $w''_{j,i}=\frac{PI_{j,i}^{(2-l)}}{b_{j,i}^k}$, where $PI_{j,i}^{(2-l)}$ implicitly plays the role of an importance weight of the corresponding observation, which also includes the nominal minimisation problem. Here l and k can be chosen independently of each other.

Instead of using the premium income of the underlying portfolio as a weight, one could also use the premium ceded to the reinsurance contract. Arguments in favour of this would be that this weight is more or less proportional with the number of claims ceded to the reinsurance contract and also that the larger companies will start from higher layers and should have less impact. Here an important drawback is that the information of two successive layers would produce the same importance as the information contained in a layer that is the sum of two layers.

For the Belgium data below we preferred to work with the premium income of the underlying portfolio (instead of the reinsurance contract) and choose $l = 2$, such that the weights disappear. However, if no reasonable k could be found to obtain more or less constant Pearson residuals, $l \neq 2$ needs to be explored.

REMARK 3.2. One of the important hypothesis for a generalized (non-)linear approach is that the variance of the observed XL-rates is only function of the expected value. In general this assumption will not be true, but there are reasons that it could be a good approximation. Nevertheless, it must always be carefully checked based on the analysis of the residuals.

Assuming a compound process to describe the excess of loss claims and assuming that the rates in practice are estimated based on a simple statistic of n years, then a formula of the variance of the XL-rates can easily be derived.

To avoid an overload of notation we will not keep into account a loading structure or a transformed severity as discussed above, and assume that the risk is not changed over the years. The rate for an excess of loss layer could be estimated by $\hat{r}(R,L) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{PI_j}\sum_{i=0}^{N}\left[X_{i,j} \vee (R,L)\right]\right)$, assuming that $X_{i,j} \sim$ X for all i and j and $PI_j = PI$ the corresponding premium income for the underlying insurance portfolio in year j. For $S = \sum_{i=0}^{N} X_i$ one has that $Var(S) = E(N) \cdot [E(X^2) + A \cdot E^2(X)]$ with $A = \frac{Var(N) - E(N)}{E(N)} \in (-1, \infty)$ (see Bühlmann, 1970), such that the variance of the estimated rate $\hat{r}(R,L)$ is equal to $\frac{A + 1 + CoV^2(X \vee (R,L))}{n \cdot E(N)} \cdot r^2(R,L) = [\beta_1 + \beta_2 \cdot CoV^2(X \vee (R,L))] \cdot r^2(R,L)$ with $\beta_i \geq 0$, where $CoV(X \vee (R,L))$ denotes the coefficient of variation of the excess of loss severity $X \vee (R,L)$ and $r(R,L)$ is the expected excess of loss rate. Note that n, A, $E(N)$ and thus also β_1 and β_2, are independent of R and L, but the variance of the XL-rate is not a fixed proportion of $r^2(R,L)$. However, a variance function of the type $r^k(R,L)$ with $k = 2$ seems to be worthwhile to check, if one may accept that the coefficient of variation of the severity is more or less constant. In practice one may expect that this coefficient will decrease for increasing rates, because increasing rates will correspond with lower and smaller layers. This implies that from a practical point of view one can expect that k lies in the interval $[0,2]$. Further, one has to keep in mind that the variance in the rates is not only due to random fluctuations, but also depends of more or less structural differences that are not (yet) modelled.

Testing and model selection procedures

The full or global model provides a separate set of parameters for each reinsurance year. If one keeps one or more parameters constant over a group of years, then these models define a so called sub-model. With the help of a deviance analysis between the full model and the sub-model, one can decide if the sub-model explains the data as well as the full model does.

The scaled deviance SD is defined as the logarithm of a ratio of likelihoods of the model under investigation and the full or saturated model: $SD(\hat{\boldsymbol{\mu}}, \hat{\phi}, \mathbf{y}) = -2\left[l^q(\hat{\boldsymbol{\mu}}, \hat{\phi}, \mathbf{y}) - l^q(\mathbf{y}, \hat{\phi}, \mathbf{y})\right]$. The deviance D is defined as $D(\hat{\boldsymbol{\mu}}, \mathbf{y}) = \hat{\phi} \cdot SD(\hat{\boldsymbol{\mu}}, \hat{\phi}, \mathbf{y})$, which means that the deviance D is equal to the scaled deviance SD with the dispersion $\hat{\phi} = 1$. Note also that maximising the quasi likelihood $l^q(\boldsymbol{\mu}, \mathbf{y})$ is independent of the dispersion ϕ and is equivalent with minimising the deviance D.

To examine the parameters $\boldsymbol{\beta}_p, \bar{\boldsymbol{\beta}}_q (p > q > 0)$ of 2 nested models with parameter sets of dimension p respectively q, each estimated in the corresponding models, one has that the distribution of the deviance satisfies the relation $D(\bar{\boldsymbol{\beta}}_\mathbf{p}, \bar{\boldsymbol{\beta}}_\mathbf{q}) = D(\boldsymbol{\mu}(\bar{\boldsymbol{\beta}}_\mathbf{p}), \mathbf{y}) - D(\boldsymbol{\mu}(\bar{\boldsymbol{\beta}}_\mathbf{q}), \mathbf{y}) \sim \phi \cdot \chi^2_{p-q} + O_p(N^{-1/2})$, with χ^2_{p-q} the chi-squared distribution with $p-q$ degrees of freedom.

This leads to a first chi-squared test $\chi^2_{(1)}$ to make inference about the submodel q in relation to model p, supposing that the dispersion is known such that it is independent of the choice of the submodel:

$$\chi^2_{(1)} = \frac{D\left(\bar{\boldsymbol{\beta}}_\mathbf{p}, \bar{\boldsymbol{\beta}}_\mathbf{q}\right)}{\phi} \sim \chi^2_{p-q}(p-q) \tag{7}$$

The dispersion parameter ϕ can be estimated by $\bar{\phi} = (\mathbf{y} - \bar{\boldsymbol{\mu}})^T \cdot \mathbf{V}^-(\bar{\boldsymbol{\mu}}) \cdot (\mathbf{y} - \bar{\boldsymbol{\mu}}) / (N-p) = X^2/(N-p)$, where X^2 is a generalisation of the Pearson statistic. If the dispersion parameter is estimated by $\bar{\phi}(2) = X^2/(N-p)$ a second chi-squared test $\chi^2_{(2)}$ is defined with the dispersion depending on the submodel.

Supposing that the dispersion is fixed but unknown and that it is independent of the choice of the submodel leads to a first F test to make inference about the submodel q in relation to model p:

$$F(1) = \frac{D\left(\bar{\boldsymbol{\beta}}_\mathbf{p}, \bar{\boldsymbol{\beta}}_\mathbf{q}\right)}{\hat{\phi}} \sim F(p-q, N-p) \tag{8}$$

Further, to avoid the estimate of parameter ϕ, one can make use of a second F-test:

$$F(2) = \frac{D\left(\bar{\boldsymbol{\beta}}_\mathbf{p}, \bar{\boldsymbol{\beta}}_\mathbf{q}\right)/(p-q)}{D\left(\bar{\boldsymbol{\beta}}_\mathbf{N}, \bar{\boldsymbol{\beta}}_\mathbf{p}\right)/(N-p)} = \frac{SD\left(\bar{\boldsymbol{\beta}}_\mathbf{p}, \bar{\boldsymbol{\beta}}_\mathbf{q}\right)/(p-q)}{SD\left(\bar{\boldsymbol{\beta}}_\mathbf{N}, \bar{\boldsymbol{\beta}}_\mathbf{p}\right)/(N-p)} \sim F(p-q, N-p). \tag{9}$$

In the Belgian Motor Third Party example we will concentrate on $F(2)$.

Another approach to selecting appropriate models is based on model selection criteria some of which are based on information criteria such as the Akaike information criteria defined as

$$AIC = -2l^q(\hat{\boldsymbol{\mu}}; \mathbf{y}) / \hat{\phi} + 2p = -2\sum_{i=1}^{N} l^q(\hat{\mu}_i; y_i) / \hat{\phi} + 2p \tag{10}$$

and

$$AICC = -2l^q(\hat{\boldsymbol{\mu}}; \mathbf{y}) / \hat{\phi} + 2p \frac{N}{N-p} = -2\sum_{i=1}^{N} l^q(\hat{\mu}_i; y_i) / \hat{\phi} + 2p \frac{N}{N-p} \tag{11}$$

Using this approach models with smaller *AIC(C)* are selected attempting to strike a balance between simplicity and predictive power (see for instance in Burham and Anderson (2004)).

4. Belgian Motor Third Party data

4.1. Quasi-likelihood estimation

The Belgian Motor Third Party data set contains the quotations of 172 excess of loss layers for Motor Third Party for the reinsurance years 2001 till 2004 (General Third Party is almost always included). All the selected layers are without any Annual Aggregate Deductible clause.

Note that Figure 1 indicates that the low rates, which correspond with the unlimited layers, increased strongly over the years. For the higher rates the evolution over the years is less clear but point in the same direction. The final year seems to be more or less comparable with the preceding year.

The need for a quasi-likelihood approach to fit the quoted rates can clearly be observed in Figure 2. Based on the estimates by a full Burr model, which corresponds with a Burr market rate model per reinsurance year with 4 parameters for every year (16 in total), we have that the residuals $r_{j,i}(R_{j,i}, L_{j,i}) - \hat{\mu}_{j,i}$ diverge when the predicted rates increase. In the right hand side panel of Figure 2 it is shown that the observed and the predicted rates correspond rather well.

As mentioned before, the choice of the parameter $k = 2$ for the variance function is motivated by a graphical presentation of the Pearson residuals (corrected with the variance weight μ^k) versus the predicted rates through the full Burr model. Figure 3 gives the classical regression line in combination with a Local Regression estimate (Proc Loess in SAS) for $k = 2$. The top graph gives the result for the squared Pearson residuals and indicates that the choice for $k = 2$ is fairly good. One could try to optimise this choice, but we prefer a more or less rounded value that could be used for several years. The fact that for $k = 2$ the analysis is independent of the nominal premium values makes this choice also more attractive.

TABLE 1

DATA DESCRIPTION

Reinsurance Years	**2001**	**2002**	**2003**	**2004**	**Total**
# layers	34	48	53	37	172
# unlimited layers	15	19	20	16	70
# cedent	**15**	**19**	**20**	**16**	**70**
# 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**
# 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**

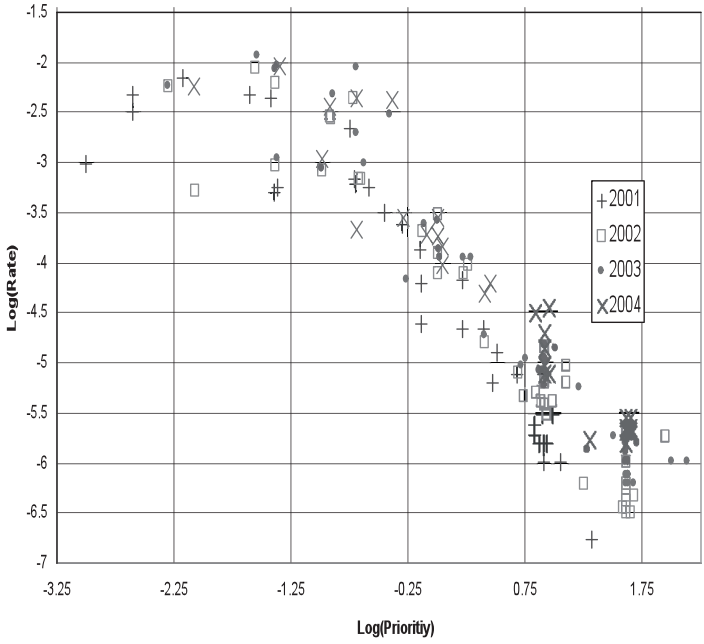

FIGURE 1: Rate per year & priority.

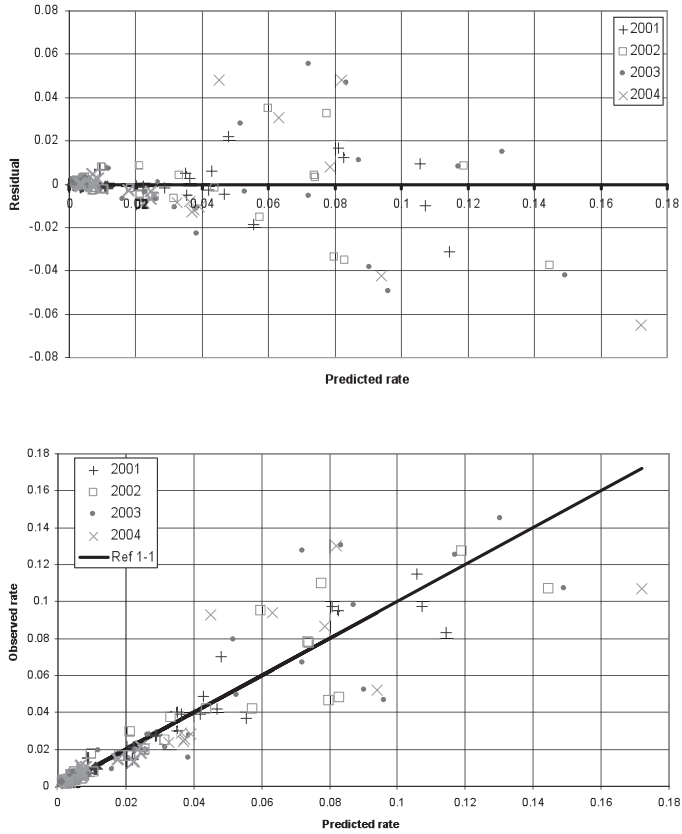

FIGURE 2: Residuals (top) & rates (bottom) as a function of the predicted rates (full Burr model).

Figure 2 shows that the fitting is fairly good, but the Local Regression Curve in the bottom side plot of Figure 3 suggest that the Burr model oscillates around the observed rates in some systematic way. The lowest rates (essentially corresponding with the unlimited upper layers) seem to be well estimated on average. But the model seems to underestimate the next group of rates and the higher rates, which essentially correspond with the first layer, are somewhat overestimated.

An explanation could be that the lower layers are basically quoted on the claim statistic of the corresponding cedent, the highest (unlimited) layer is essentially based on market rates. The quoted rates of these two parts of the cover are brought in line with each other through the negotiation with the reinsurance market. But this does not imply that they can be described through a relative simple global quoting model (only two covariates and an analytical description of the severity). Nevertheless the proposed model is fairly good and can further be reduced to a much smaller model as shown below.

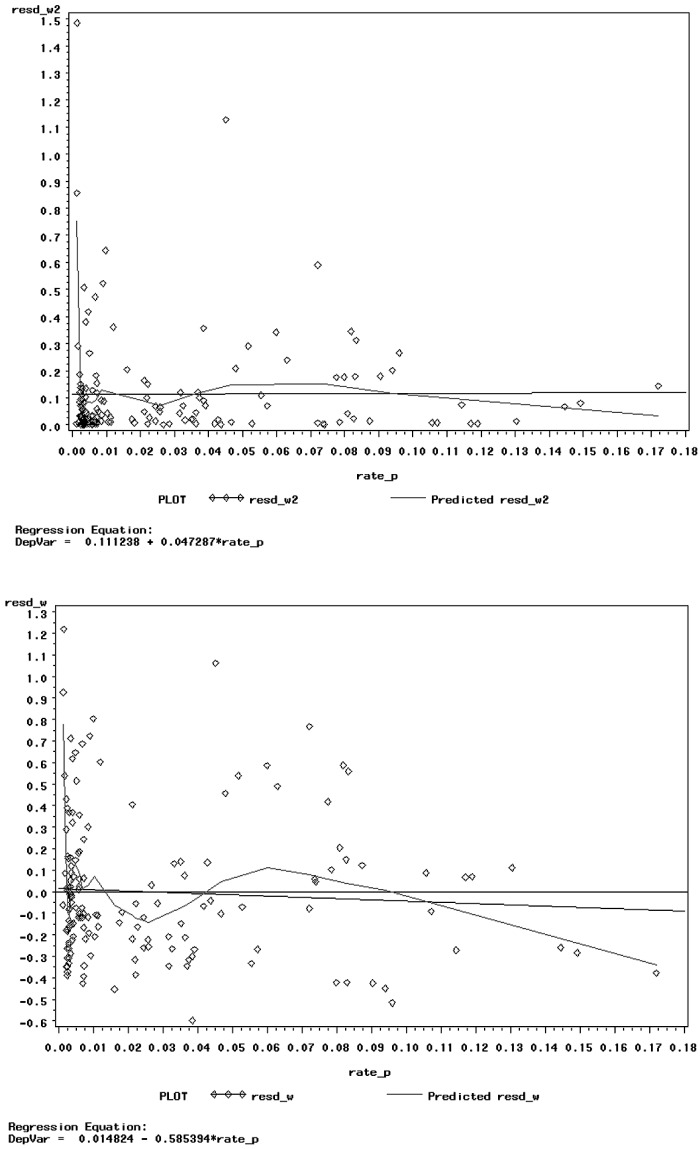

FIGURE 3: Squared (top panel) Pearson (bottom panel) residuals from the full Burr model.

One of the assumptions of GLIM is the independency of the observations. Although GLIM can be extended to correlated data by Generalized Estimating Equations (GEEs) (Godambe & Kale 1991, McCulloch & Searle 2001, Verbeke & Molenberghs 2000), we limit ourselves in this study to independence. This choice can be motivated by the 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). The estimation is based on 102 couples (= number of layers minus the number of cedents). So we conclude that the independency assumption is acceptable as a working hypothesis.

4.2. The model selection procedure

Here we try to answer several questions. Firstly, could we limit ourselves to a Pareto model or is a Burr model preferred? What is the smallest sub-model, which can be used to describe the 4 years of excess of loss rates adequately? What is the highest number of successive years with equal parameters?

The parameter d is strongly linked to the reinsurance loading and the claim frequency. At least the reinsurance loading cannot be supposed to be constant over the years and makes the d constant sub-model less attractive. For the parameter α one does not expect important changes every year, but rate corrections mainly related to the higher layers cannot be excluded over a longer period. Only if one has reasons to exclude important rate corrections for the higher layers, one can keep α constant. However an important correction in 2002 will appear. Important to remark is that it is not natural to keep α a priori constant and to let γ vary over the years, because the parameter γ acts as an generalisation through a transformation of the form $X^{1/\gamma}$ with X Pareto distributed. Furthermore both parameters are strongly related with each other in the sense that the distribution function for larger claims only depends of the product $\alpha \cdot \gamma$.This implies that it seems much more reasonable to keep only γ constant or to keep α and γ together constant.

The parameter θ leads to a generalisation of the 1-parameter Pareto model to the 2-parameter Pareto model (see Klugman et al., 1998). This parameter is strongly related with the parameter d and interacts also with α (see (3) and (4)), which makes it worthwhile to keep it constant to stabilize the estimate of the other parameters. On the other hand θ is linked to the lowest excess of loss priorities (see 1-parameter Pareto model), which are not constant over the years due to claim inflation and market corrections.

So the γ constant sub-model a priori seems to be the most attractive sub-model between the full Burr and Pareto model compared to the other constant sub-models. But to stabilize the parameters estimation the θ constant sub-model is also important and this in combination with or without the γ constant. However, if one has enough observations then we prefer to explore the observed rates in detail before concluding.

There are many sub-models possible. To avoid an overflow of sub-models (54,000) we will limit ourselves to sub-models with one or more parameters equal for successive years, assuming that only neighbouring years will make a chance to be comparable (4,608).

The fact that we work in the context of a statistical model implies that we can make inference about the significance of sub-models, which leads to a better interpretation of the results.

First of all we look to the 10 best sub-models based on the Akaike information criteria AIC and AICC (see Table 3). Both criterions indicate that the choice for 7 free parameters is a good compromise to describe the data sufficiently well.

The Akaike information criterion has the disadvantage that one needs an estimate of the dispersion factor and that it is less clear how much two different models deviate from the full model, especially when the numbers of parameters are different. To overcome these drawbacks we will make use of the F(2)-test p-values as described in (9). The higher the F(2) p-values are for a certain sub-model, the better this sub-model correspond with the description of the data through the full model. F(2) can be seen as a simple criterion without explicit estimate of the dispersion factor.

The simplest way to illustrate the F(2) approach is to look at Figure 4. Here all remaining 4,608 sub-models are brought together and ordered through the F(2) p-values and the number of free parameters. Further all models are classified in 5 (overlapping) groups. One for each sub-model with a constant parameter (the so called d, θ, α and γ constant models, if two parameters are constant they belong to both groups) and one group with all models that have the same number of parameters. One can conclude that:

- More then 9 free parameters are not needed. There are enough sub-models with very large F(2) values (larger then 0.8). Similarly, one needs at least 6 free parameters (sub-model with less then 6 free parameters are not shown but have always a F(2) value smaller then 0.05). The maximum F(2) value for 6 free parameters is around 0.363 and is not low enough to exclude this choice, but smaller models are unacceptable. The fact that one has a view on a relevant range of free parameters instead of just one optimal model will give much more and useful information.
- There are enough sub-models with only 7 free parameters that perform as well as these with 8 free parameters. Choosing between 7 or 8 free parameters is in favour of 7 free parameters.
- The loss of goodness of fit between 6 and 7 free parameters is relative high (more then 0.35). Between 9 and 7 (or 8) the loss is limited (around 0.20) and there are enough sub-models for 7 free parameters with relative high F(2) values (more then 0.6). Choosing for 7 free parameters seems to be a good compromise between goodness of fit and number of free parameters. The 'best F(2)' sub-model with 7 free parameters corresponds also with the AIC(C) selection.
- For all sub-models between 6 and 9 free parameters there exist always a γ constant model with the highest F(2) value, suggesting that for this sample the γ constant model is the most interesting sub-model between (full) Burr and Pareto.

Based on the AIC(C) and the F(2) analysis we decide to select the first model (= B-3029) in Table 3. This model has the nice property that the difference between each year is explained by only one parameter. For year 2002 the parameter α decreases significantly, implying an important increase of the XL-rate for the higher layers. For year 2003 an increase of 19.5% of the parameter d occurs, implying an overall increase for all layers with 19.5%. For year 2004 again we find a small decrease of α, which implies a small correction essentially for the higher layers.

Further the model belongs to the sub-classes with θ and γ constant over the 4 years and makes this model extremely attractive to describe and interpret the evolution of the sample.

The second model B-3923 has the property that the parameter d and γ stay constant during the 4 years. Together with the property that also the difference of each year is explained by one parameter (same as previous model but the role of d is taken over by θ with an increase of 8%). This model will keep the frequency and the proportional loading constant, but takes the claims inflation explicitly into account which is estimated to be equal to 8% in 2003.

As mentioned before, due to the fact that the coefficient of variation for the γ constant sub-model is independent of θ makes this sub-model also from this point of view very attractive.

The overall conclusion is that we prefer to describe the Belgium MTPL sample with one of the first two models. We do not have any special preference between both, but we decided to select the first one for illustration.

To give a simple view on the differences between the years we calculate the benchmark rates for several unlimited layers based on the first model B-3029 in Table 3. Table 2 illustrates clearly that the differences are mostly not constant for different layers, and that there is a large difference between 2001 and 2002, especially for the higher layers.

TABLE 2

BENCHMARK RATES FOR UNLIMITED LAYERS (B-3029)

Priority x 10^6	Bench 2004	Bench 2003	Bench 2002	Bench 2001	2004 vs 2003	2003 vs 2002	2002 vs 2001
0,50	8,29%	7,81%	6,53%	4,95%	**6,2%**	**19,5%**	**32,0%**
0,75	5,48%	5,07%	4,24%	2,95%	**8,2%**	**19,5%**	**44,1%**
1,00	3,85%	3,51%	2,94%	1,88%	**9,9%**	**19,5%**	**55,8%**
1,50	2,21%	1,96%	1,64%	0,93%	**12,7%**	**19,5%**	**76,7%**
2,00	1,45%	1,27%	1,06%	0,54%	**14,9%**	**19,5%**	**94,6%**
2,50	1,04%	0,89%	0,75%	0,36%	**16,7%**	**19,5%**	**110,2%**
3,00	0,79%	0,67%	0,56%	0,25%	**18,2%**	**19,5%**	**124,1%**
4,00	0,51%	0,42%	0,35%	0,14%	**20,6%**	**19,5%**	**148,2%**
5,00	0,36%	0,29%	0,25%	0,09%	**22,6%**	**19,5%**	**168,8%**

TABLE 3

OVERVIEW OF MOST INTERESTING (SUB-)MODELS (empty cells equal the value above)

Model ID	**B-3029**	**B-3923**	**B-3064**	**B-4057**	**B-4035**	**B-4056**	**B-4042**	**B-3958**	**B-4050**	**B-3285**	**P-0348**	**P-0418**	**P-0489**	**B-0001**
Model Type	**Burr**	**Burr**	**Burr**	**Burr**	**Burr**	**Burr**	**Burr**	**Burr**	**Burr**	**Burr**	**Pareto**	**Pareto**	**Pareto**	**Burr**
$\#d$	2	1	2	1	1	1	1	1	1	2	2	3	4	4
$\#\theta$	1	2	1	1	1	1	1	2	1	3	3	2	4	4
$\#\alpha$	3	3	1	3	4	3	3	1	3	3	3	3	4	4
$\#\gamma$	1	1	3	2	1	2	2	3	2	1	0	0	0	4
# Par	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**9**	**8**	**8**	**12**	**16**
AIC	157.93	157.96	158.33	158.43	158.82	158.83	158.90	158.96	159.04	158.71	159.86	159.89	167.45	170.44
rank	**1**	**2**	**3**	**4**	**6**	**7**	**8**	**9**	**10**	**5**	**48**	**49**	**2511**	**3039**
AICC	158.53	158.56	158.92	159.02	159.41	159.42	159.50	159.55	159.63	159.70	160.65	160.67	169.25	173.73
rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**42**	**44**	**2572**	**3477**
F(2)	0.7197	0.7163	0.6750	0.6636	0.6185	0.6177	0.6091	0.6023	0.5933	0.9216	0.6354	0.6327	0.2333	1.0000
d_1	0.2426	0.2525	0.2455	0.2644	0.2630	0.2635	0.2645	0.2515	0.2592	0.2113	0.3373	0.3373	0.3373	0.2982
d_2										0.3735	1.0059	0.9088	0.9682	0.3030
d_3	0.2899		0.2992									1.1028	1.1504	0.2200
d_4													0.8762	3.1584
θ_1	0.5686	0.5583	0.5699	0.5824	0.5825	0.5829	0.5818	0.5572	0.5822	0.7902	1.7419	1.7419	1.7419	1.3159
θ_2										0.4889	0.4498	0.4674	0.4537	0.4957
θ_3		0.6037						0.6065		0.5306	0.4885		0.4632	0.6025
θ_4													0.5200	0.6941
α_1	1.4080	1.3838	1.3203	1.5275	1.5062	1.5051	1.4582	1.2503	1.4866	1.9967	5.1709	5.1709	5.1709	4.0215
α_2	1.2375	1.2187		1.3384	1.3190	1.3178			1.2598	1.5215	2.8499	2.8448	2.8422	1.1819
α_3				1.2591	1.2757	1.2745	1.2342						2.8512	0.8732
α_4	1.2033	1.1860			1.2419		1.2016		1.2258	1.4844	2.7933	2.7866	2.7992	5.3211
γ_1	2.1216	2.1566	2.2782	2.0112	2.0399	2.0431	2.1111	2.3915	2.0615	1.7347	1.0000	1.0000	1.0000	1.1280
γ_2			1.9956				1.8411	2.1015	2.1317					2.1838
γ_3				2.0702										2.9499
γ_4			1.9411			1.9878		2.0453						0.6170

4.3. An application to priority setting

In practice, as a rule of thumb, one keeps the ceded expected number of claims more or less constant (e.g. between 0.5 and 4). If we would define the retained priority such that the expected number of claims is independent of the portfolio size, then one could check this relationship to verify if the above-mentioned rule of thumb has some common market sense.

As derived in paragraph 2 we know that $E[N_{R(x)}] = E[N] \cdot \bar{F}[R(x)] = d(x) \cdot x \cdot \bar{F}[R(x)]$ with $R(x)$ the priority corresponding with a premium income $PI = x$

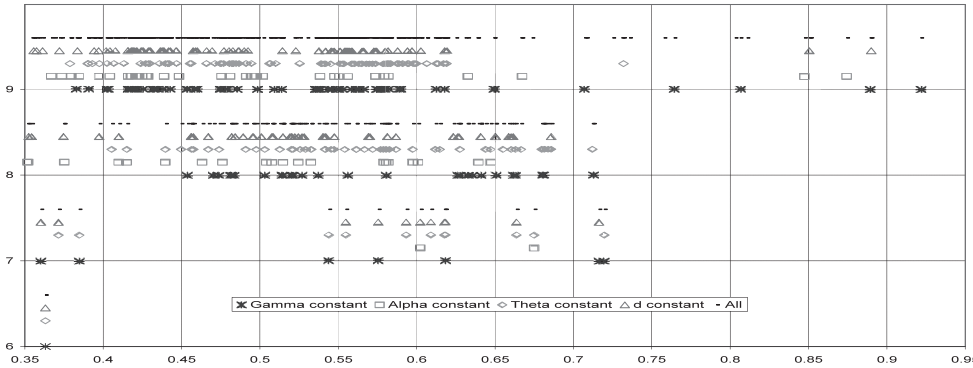

FIGURE 4: F(2) p-values versus # parameters per constant Burr sub-model.

and $d(x)$ the expected number of claims from ground up per unit of premium income and biased with the discounted loading factor of the excess contract. We suppose that $d(x)$ depends of the size of the underlying portfolio through the premium income. This could for example be the case if the reinsurance loading depends of the portfolio size. We have that $E[N_{R(x)}] = c \Leftrightarrow \bar{F}[R(x)] = \frac{c}{x \cdot d(x)}$, with $c > 0$.

Under the Burr model we get that $R(x) = \vartheta\left[\left(\frac{x \cdot d(x)}{c}\right)^{1/\alpha} - 1\right]^{1/\gamma}$. Thus the retained priority $R(x)$ is of the form $[\alpha \cdot (x \cdot d(x))^{\beta} + b]^{(1/\gamma)}$ or $[\alpha \cdot x^{\beta} + b]^{(1/\gamma)}$ if $d(x)$ is constant.

For a Pareto model we have that $R(x)$ is of the form $\alpha \cdot (x \cdot d(x))^{\beta} + b$ or $\alpha \cdot x^{\beta} + b$ if $d(x)$ is constant.

Note that for the one parameter Pareto the term b is equal to zero, which means that the relationship reduces to a standard power function. Note also that in all other cases the term b is equal to $-\theta^{\gamma} < 0$, which implies that for small values of x, $R(x)$ does not exist. The interpretation here is that the portfolio is too small to generate the expected number of claims.

To verify this relationship for the Belgian Motor Third Party Liability we confine ourselves to 2003. A first analysis indicates that the relation between premium income and first priority fairly well can be well described by a standard power function $R^*(x) = \alpha \cdot x^{\beta}$ (see Figure 5 top panel). But this relationship will not result in a constant number of ceded claims.

To obtain a better view on the ceded number of claims (which are biased by loading and discount), we will make use of the benchmark model B-3029 (see Table 3), but will correct each reinsurance program proportionally with the individual cumulative observed rate and the corresponding benchmark rate. This means that one will keep the severity constant over all reinsurance programs and absorb the relative difference between the observed rates and the benchmark rates fully in the frequency (related with the parameter d) $E\left[N^{cor}_{R(x_i)}\right] = d \cdot x_i \cdot \bar{F}[R(x_i)] \cdot r_i(R_1,\infty)/b(R_1,\infty)$.

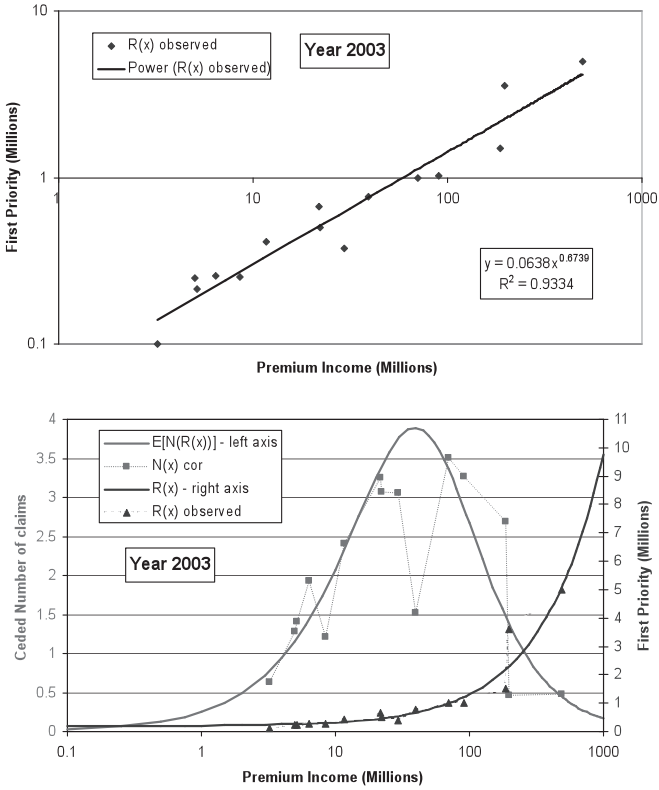

FIGURE 5: Relationship between Premium volume, retained priority & estimation of ceded number of claims.

These estimates are plotted in Figure 5 (bottom panel – left axis) together with the first priority relationship $R^{**}(x) = a \cdot x^{\beta} + b = 0.0151 \cdot x^{0.934} + 0.1975$ where the parameters are estimated by a simple minimisation of the quadratic logarithmic difference between the corrected ceded numbers of claims related to the observed priorities and the ceded numbers based on the relationship $R^{**}(x)$. This figure indicates clearly that the number of ceded claims is situated between 0.5 and 4 and confirms the basic rule of thumb. We also observe that the rule of thumb is too simple in the sense that the corrected ceded numbers of claims follow a pattern that depends of the portfolio size. The pattern is based on $R^{**}(x)$, which fits also very well with the observed first priorities.

In the figure one point deviates strongly from the ceded pattern. This has essentially to do with the fact that for that particular insurance company the rate was set too low. In fact that rate was corrected in 2004.

It is remarkable that the relationships deduced from a constant ceded frequency via a Burr severity fits well in practice although that cession is portfolio size dependent. We also verified the relationship $R^{***}(x) = (a \cdot x^{\beta} + b)^{g}$, but the difference was not significant.

TABLE 4

ESTIMATED CEDED FREQUENCY (B-3029 - BIASED BY DISCOUNT & LOADING)

	2004		2003	
	Without correction	**Corrected**	**Without correction**	**Corrected**
Mean	2.643	2.192	2.322	2.013
Median	2.613	2.411	2.043	1.935

Figure 5 (bottom panel – left axis) suggest that the average ceded frequency for the 2003 sample is around 2.

Finally we would like to remark that the corrected frequency has the advantage that the benchmark model is translated in a simple way to a market conform individual model. Given a sample of large claims for a particular insurance company, one can analogically correct the parameter d with the estimated expected frequency that exceeds some (smaller) priority. This is one of the simplest ways to combine individual claim experience with benchmark severity, a severity which is measured only through the sample of excess of loss rates negotiated on the reinsurance market.

5. CONCLUSIONS

- The excess of loss commercial rates for Motor Third Party Liability in Belgium are fitted with the help of an underlying Compound Burr model, which include the Compound Pareto as a special case.
  The motivation of this choice is mainly based on the common use in practice of the Compound Pareto model for reinsurance quotation for this line of business.
  More important however is the observation that the commercial rates are fitted well with this model. This makes it possible to derive the claim severity, as appreciated and quoted by the reinsurance market, from the excess of loss rates without the analysis of the underlying claims. This claim severity should be interpreted as a reinsurance price per claim for the real underlying claim severity, because loading structures related to the risk appetite of the reinsurance market are in some sense (partially) included.
  Although only a relative small number of observed reinsurance programs and rates are observed per year, makes the severity model it possible to summarise accurately the observed rates over the whole range of reinsurance covers (low and high layers) in a relative simple parametric model.
  This parametric model (Compound Burr) keeps account with the priority and limit of each observed layer as covariates. This has the advantage that also non-successive layers may belong to the sample of analysed observations

and that once the model is calibrated one can calculate the benchmark rate for any combination of priorities and limits.

- The explicit isolation of a common (reinsurance) claim severity in the benchmark model has the additional advantage that, through the premium income of the underlying portfolio, one also obtains an estimate for the expected number of claims ceded to the reinsurance market. These estimated numbers of claims are biased with a discounted loading factor and the heterogeneity of the underlying portfolios.
  Due to the compound modelling one will generate a benchmark summarisation of the rates but also of frequency and severity. Instead of one benchmark one will have three benchmarks.
  Further by correction through the frequency component one generate in an easy way an individual benchmark for each cedent, which has the advantage of a common severity for the whole market.
  Important to note is that no claim statistics are needed for these benchmark analysis. However if available one can verify them with the experience in practice, although one has to keep the discounted loading in mind. Nevertheless a common benchmark severity is a good starting point when analyzing a claim statistic, especially if the sample is (too) small such that some parameters from the common Burr model can be taken over.
- Moreover, the decomposition gives the additionally ability of making interpretations, deductions and verifications. A nice example of an indirect deduction is the verification of the rule of thumb for the Belgium market: "A more or less fixed number of claims for each company are ceded to the reinsurance market", which is done without analysing of any claim statistic and leads to a more general relationship between priority setting and the size of the underlying portfolio.
  Further becomes it a handy toolkit to make considerations about the expected next renewal excess of loss pricing.
- The analysis is done within the framework of Generalised Non Linear Models (GLIM). The motivation of this choice is mainly based on the observed heteroscedasticity of the residuals. A residual analysis for the Belgian data indicates that a square root variance function describes the heteroscedasticity well. A choice that also can be motivated from an approximation of the variance of the estimated excess of loss rates through experience rating.
  This in combination with the fact that a square root variance function makes no difference between a fit of the nominal excess of loss premiums (which corresponds with the importance scaling by premium weights) or fit of the excess of loss rates (without importance scaling) did lead to the final choice.
  Within the GLIM framework we analysed the rates with the quasi likelihood approach, which has the advantages that only the variance function has to be chosen to describe the residuals. But, maybe more important in practice is that the estimation of the parameters can be explained by a weighted regression, where the weights are inversely proportional to the variance function (or estimated variance of the residuals). This way of working is much easier to interpret and to explain to management than GLIM.

The conclusion of this approach is that much more weight is given to the rate residuals of the higher layers (= lower rates) then these of the lower layers (= higher rates), which is the opposite to what one often thinks.

- The over several years observed reinsurance rates are described in one large model. The fact that one works in a well know statistical framework makes it possible to judge how parsimonious a sub-model may be without significant loss of describing well the observed rates. This analysis was done with the help of the Akaike information criteria AIC and AICC, but has the disadvantage that one need an estimate of the dispersion factor. To overcome this drawback we used also an F-test as an ordering criterion through the corresponded p-values. Both methods will come to the same conclusion: for the Belgium MTPL-market a γ-constant Burr model is preferred over a general Burr model and a general Pareto model.
- The GLIM assumptions were carefully checked for the Belgium case. All were acceptable fulfilled, implying that this approach makes sense in practice. However, the most crucial assumption is the independence of the observations (commercial non-cumulative rates, not necessarily successive layers). Nevertheless the correlation is very low and not significant different of zero should there be some dependency between the layers of the same cedent. Further is the error on the estimated benchmark rate not only due to randomness but is also influenced by some structural elements. The underlying portfolios can be different (more or less commercial vehicles, inclusion of different perils and coverage, influences of tariff segmentations, …) and the clauses in the observed excess of loss contracts are not always the same (different stabilisation and interest clauses, …), leading to some unexplained parts of the residuals. If the heterogeneity is relatively large, one may need to incorporate some additional covariates or random effects. This is not taken into account in this analysis, further investigation in that direction seems to be interesting. A mixed or Bayesian approach could be helpful to derive a more sophisticated individual Benchmark model per cedent.

6. Acknowledgment

The authors would like to thank the referee for his careful reading of this manuscript, which leads to an improved presentation. Also are greatly appreciated the suggestions and recommendations from Filip Jacobs and Yves Samain. The authors acknowledge the financial support of the Onderzoeksfonds K.U. Leuven (GOA/07: Risk Modeling and Valuation of Insurance and Financial Cash Flows, with Applications to Pricing, Provisioning and Solvency).

7. References

Agresti, A. (2002) *Categorial Data Analysis*, John Wiley & Sons, New York.

Bühlmann, H. (1970) *Mathematical Methods in Risk Theory*, Springer-Verlag, Berlin.

Burham, K.P. and Anderson, D.R. (2004) *Model Selection and Multi-Model Inference – A Practical Information – Theoretic Approach*, Springer Verlag New York.

Carter, R.L., Lucas L.D. and Ralph N. (2000) *Reinsurance*, Reaction Publishing Group (in association with Guy Carpenter & Company).

DOBSON, A.J. (2002) *An Introduction to Generalized Linear Models*, Chapmann and Hall.
DOERR, R. (1980) *Property Excess of Loss: Pareto Rating*. Swiss Re publications.
GERATHEWOHL, K. (1980) *Reinsurance Principles and Practice*, Volume I, Verlag Versicherungswirtschaft e.V., Karlsruhe.
GERATHEWOHL, K. (1982) *Reinsurance Principles and Practice*, Volume II, Verlag Versicherungswirtschaft e.V., Karlsruhe.
GODAMBE, V.P. and KALE, B.K. (1991) *Estimating Equations*. Oxford Science Series 7.
HARDIN, J. and HILBE J. (2001) *Generalized Linear Models and Extensions*, Stata Press college station Texas (Tex.).
KILN, R. (1982) *Reinsurance in Practice*, Witherby & Co., London.
KLUGMAN, S.A., PANJER, H. and WILLMOT, G.E. (1998) *Loss Models, From Data to Decisions*, John Wiley & Sons, New York.
MACK, Th. (1997) *Schadenversicherungsmathematik*. Schriftenreihe Angewandte Versicherungsmathematik, Heft 28, Verlag Versicherungswirtschaft.
MCCULLAGH, P. and Nelder, J.A. (1985) *Generalized Linear Models*, Chapmann and Hall.
MCCULLOCH, C.E. and SEARLE, S.R. (2001) *Generalized, Linear and Mixed Models*, Wiley New York (N.Y.).
SCHMITTER, H. (1978) *Quotierung von Sach-Schadenexzedenten mit Hilfe des Paretomodells*. Swiss Re publications.
SCHMITTER, H. and BÜTIKOFER, P. (1997) *Abschätzung von Risikoprämien für Sach-Schadenexzedenten mit Hilfe des Paretomodells*. Swiss Re publications.
SCHMUTZ, M. and DOERR, R. (1998) *Das Paretomodell in der Sach-Rückversicherung*. Swiss Re publications.
SEBER, G.A.F. (1977) *Linear Regression Analysis*, Wiley New York.
STRAUB, E. (1988) *Non-life Insurance Mathematics*, Springer Verlag.
VERBEKE G. and MOLENBERGHS G. (2000) *Linear Mixed Models for Longitudinal Data*, Springer Verlag New York.
VERLAAK R. and BEIRLANT J. (2003) Optimal Reinsurance Programs: An optimal Combination of several Reinsurance Protections on an heterogeneous Insurance Portfolio, *Insurance: Mathematics and Economics*, **33**: 381-403.
WANG S. (2002) A universal framework for pricing financial and insurance risks, *Astin Bulletin*, **32(2)**, 213-234.
WANG S. (1996) Premium calculation by transforming the layer premium density, *Astin Bulletin*, **26(1)**, 71-92.
WANG S. (1998) Implementation of Proportional Hazards Transforms in Ratemaking, *Proceedings of the Casualty Actuarial Society* **LXXV**, 940-979.

ROBERT VERLAAK
Aon Benfield, Brussels
Faculty of Business and Economics,
Katholieke Universiteit Leuven
Belgium
E-Mail: robert.verlaak@skynet.be

WERNER HÜRLIMANN
IRIS integrated risk management ag
Zürich
Swiss

JAN BEIRLANT
Department of Mathematics and Leuven Statistics Research Centre
Katholieke Universiteit Leuven
Belgium