Estimation of

General Rating Models

for Insurance Tariffs

Paper presented at the 8th Symposium on Finance, Banking and Insurance
December 1999

W. A. Fels

Austrian Association of Insurance Companies, Economic Department
A-1030 Vienna, Schwarzenbergplatz 7
☎ +43 +1 +711 56 /242, 💻 fels@vvo.at

Abstract:

In general insurance tariffs consist of two components: First risks are classified according to a vector of basic criteria, say X_t, which - for example in automobile insurance - depend on the type or the power of the vehicles. Given the basic premiums, $X_t\alpha$, further extra charges and discounts are rated depending on usually more individual characteristics of the risk (say Z_t). All relevant tariffs can be represented by a Generalized Rating Model that determine the net expected premium by

$$E(y_t) = X_t\alpha.\exp\{Z_t\beta\}$$

While previous work focused either the estimation of linear ($X_t\alpha$) or multiplicative models ($\exp\{Z_t\beta\}$), we will estimate the parameters of the two linear kernel functions simultaneous by Maximum-Likelihood methods. This allows a valid selection and testing of rating criteria within the market-relevant tariff structure without misleading evidence due to divergent modelstructures. The efficiency of the suggested tariff can immediately be judged by ML-statistics and information criteria.

We illustrate the model and the estimation with examples from the Austrian motor liability market. Restricted net-tariff estimations provide premium structures which are consistent with exogenous fixed components like a Bonus-Malus scale or traditionally applied discounts.

Keywords: Insurance rating; tariff structures; model specification; Maximum Likelihood; automobile insurance

1 INTRODUCTION

The most noble goal of general insurance mathematics is the statistical determination of net risks premiums, say $\ddot{y}_t$, (t=1,…,T), due to observed benefits y_t and several criteria, X_t and Z_t that characterise that risk.

In general insurance typical tariffs consist of two parts: The base tariff consists of a sequence of risk criteria that defines the structure of the portfolio. These characteristics usually correspond to the structure of traditional risk statistics where groups are formed according to main criteria like the line of business in industrial fire insurance or the motor power in motor liability. This structure might be represented by a linear combination of dummy variables, say $X_t\alpha$. Beside this the premiums will depend on a system of additional charges and discounts which usually reflect more individual criteria like the age or gender in motor insurance. In general several risk groups of the basic tariff are rated by common discount factors.

As this structure is typical for insurance tariffs we will denote such a model as generalized rating model (GRM). Up to now we have not found a tariff structure which can not be represented by a GRM. However - as it will be discussed in chapter 2, GRMs can in general neither be transformed to a purely linear -, multiplicative - or to a GLM- model. Thus any inference derived from one of those models is misleading when a rating model should be specified.

Technical aspects of estimating GRMs are discussed in chapter 3. We suggest to apply ML-estimation. Once the Likelihood is explicitly expressed by the expected net premium and its variance the statistical aspects are trivial. However, to derive such a specification of the Likelihood might be challenging. As an example, chapter 3.3.3 of this paper demonstrates the solution for the Lognormal- Likelihood.

In chapter 4 we present estimations based on Austrian motor liability data. The examples demonstrate that the structure of the model influences the inference about relevant risk criteria. A reliable selection of variables is only possible within the relevant rating structure model.

2 GENERALIZED RATING MODELS

2.1 The structure of generalized rating models (GRM)

Let us denote the net-risk premium of the t^{th} cell by $\ddot{y}_t$. The premium function will depend on known exogenous rating factors, say $X_{t,i}$ and $Z_{t,j}$, i=0, …, k_X, j=1, …, k_Z. In general the market relevant tariff structure can be expressed by

(2.1) $$\ddot{y}_t = (\alpha_0 + \alpha_1 X_{t,1} + \alpha_2 X_{t,2} + \ldots + \alpha_{kx} X_{t,kx}) .\exp\{ \beta_1 Z_{t,1} + \beta_2 Z_{t,2} + \ldots + \beta_{kz} Z_{t,kz}\}$$

say Net-Premium$_t$ = [baseline tariff for subgroup t] * {net-excess and reduction factors}.

Applying vector notation and defining the k_x+1-dimensional vector $X_t = [1, X_{t,1}, X_{t,2}, \ldots, X_{t,kx}]$ and the k_z-dimensional Z_t -vector, which holds the discount and excess criteria $Z_{t,j}$, the rating model for risk group t is equivalent to

(2.2) $$\ddot{y}_t = X_t\alpha \, . \exp\{ Z_t\beta \}$$

Usually the covariates are $Z_{t,i}$ Dummy variables. As $\exp(ß_i Z_i) = \exp(ß_i)^{Z_i}$ the premiums of risks indicated by $Z_i = 1$ will be $\exp(ß_i)$-times higher than those of risks with $Z_i = 0$. But Z might also include quantitative variables like in example B.

Examples:

(A) The Austrian industrial **fire insurance** tariff is structured according to tariff books which include on average 30 risk groups. Given the line of business further additive premium factors are intended for security holdings and electronic equipment. This baseline tariff might be expressed by $X_t.\alpha$ where X indicates the specific sum insured and α is a vector of premium factors. For all branches of industry the tariffs apply common increases for the construction of building, and the roofing as well as reductions for fire-alarm and -fighting systems. Obviously these tariff components can be expressed by $\prod_i \exp(ß_i Z_{t,i})$.

(B) Usually the discount factors applied for a higher fixed **deductible**, say $m_{i,t}$, are common for several risk-groups and thus independent of the basic rates $\ddot{y}| m_{i,t}=0$. Thus the tariff function will be

$$(2.3) \qquad \ddot{y}_{i,t} = X_t\alpha.\exp\{ Z_t\beta + k.m_{i,t}\}.$$

This is obviously a generalized rating model. However this rating function will not represent the true data generating process because even under very simple distributional assumptions the expected burden will be $E(y_{i,t}) = X_t\alpha.\exp\{ Z_t\beta + k_t.m_{i,t}\}$ where k_t depends on the risk-distribution of class t. To illustrate this, suppose that the claim size distribution is locally exponential declining a the reference-value m_0. Then $E(y_{i,t}) = E(y_{0,t}).\exp\{- p^+_{0,t}/E(y_{0,t})\ (m_{i,t}-m_0)\ \}$ where $E(y_{0,t})$ is the expected burden of a contract with deductible m_0 and $p^+_{0,t}$ the claimfrequency for a type t - risk with deductible m_0 (W. Fels, 1997). If models for $E(y_{0,t}| X_t, Z_t)$ and $E(p^+_{0,t}| X_t, Z_t)$ were specified, the estimated k_t parameters will follow an erratic pattern, such that these parameters have to be smoothed in an second estimation step. Whenever for some groups a common percentage reduction for higher deductibles is required, the average of several estimated k_t has to be calculated. But the uncontrolled accumulation of estimation- and smoothing errors leads to inefficient estimators while estimating the required rating structure 2.3 directly by quasi-ML-methods guarantees efficiency even if 2.3 is only an approximation of the true data generating process.

2.2 Related rating structures and Generalized Linear Models

With the pioneer-work of Bailey and Simon (1960) the actuarian literature mentioned the problem of estimating pure multiplicative models:

$$(2.4) \qquad \ddot{y}_t = \alpha_0.\exp\{ Z_t\beta \}$$

In fact, the method was developed to fit a tariff structure for two-dimensional contingence tables, where Z_t classified the data according to two categorial variables. The oncomming development of this model focused alternative loss-functions: While Jan Jung's Method of Marginal Totals (Jung J., 1968) is appropriate for the Modified-Poisson distributed data (Mack, 1997, p. 167), D. T. Sant (1980) suggested a Least-Square approach which fits to the normal-distribution.

For a higher dimensional classification, the additive model

$$(2.5) \qquad \ddot{y}_t = X_t\alpha$$

seemed more appropriate. Applications of the linear model can be found in Johnson and Hay (1971), F.A.Ruygt (1982), J. Lemaire (1985, Part II) and others.

More recent analyses, like Stroinski and Currie (1989) apply Generalized Linear Models (GLiM) which are associated with the work of Nelder and Wedderburn (1972) respectively McCullarg and Nelder (1989). These models generalise the distributional assumption of the linear model to allow efficient estimation for distributions of the exponential family. Although the class of generalized rating model (2.2) overlaps with models of the GLiM-class, in order to avoid confusion, generalized rating models should not be denoted as GLiMs. In GLiMs only one linear kernel $X_t\alpha$ is specified, which is related to the expected burden by means of the (inverse) link-function h():

(2.6) $$\ddot{y}_t = h(X_t\alpha)$$

Due to h() GLiMs will estimate reasonable rating stuctures only in very special cases. Purely linear (2.5) and the purely multiplicative tariffs (2.4) can be specified when $h(x) = x$ respectively $h(x)=\exp(x)$. These cases correspond to the canonical link functions of Normal- and Poisson-models. In all other cases the nonlinear inverse link function h(x) will lead to dubious tariff models.

Among others A. E. Renshaw (1994) illustrated this for a motor insurance portfolio: Assuming Gamma distributed claim sizes $y_{c,t}$ implies $h(X_{c,t}\alpha) = -1/X_{c,t}\alpha$. For a given claim frequency the net premiums of all risks caracterised by a dummy, say $X_{c,2}$, should be $-100.\alpha_2/E(y_{c,t}|X_{c,2,t}=1)$% larger than the premiums of risks with $X_{c,2}=0$. Thus the optimal discount for one risk criteria depends on all other risk criteria ($X_{c,1,t}$, $X_{c,3,t}$, …). Lady- or regional discounts will vary for different cars. As $E(y_{c,k}-y_{c,t}) = 1/X_{c,t}\alpha - 1/X_{c,k}\alpha$, the same holds for absolute premium-differences. Up to now we have not seen any tariff following that structure.

Another difference between GRM and purely linear or multiplicative models is the treatment of multicolinearity. If V_1 and V_2 denote two highly correlated or dependent variables an estimation of the models $\ddot{y} = \alpha_0+\alpha_1 V_1+\alpha_2 V_2$ and $\ddot{y} = \alpha_0.\exp(\alpha_1 V_1+\alpha_2 V_2)$ will collapse due to colinearity while the models $\ddot{y} = (\alpha_0+\alpha_1 V_1).\exp(\alpha_2 V_2)$ and $(\alpha_0+\alpha_2 V_2).\exp(\alpha_1 V_1)$ can still be correctly estimated.

2.3 What is estimated

Recall that the most noble goal of rating models is the statistical estimation of the net premium $\ddot{y}$. Nevertheless a broad line of actuarial literature focuses the estimation of frequency (f_t) - and claim severity (s_t) - models. According to thecollective model the expected premiums are calculated by

(2.7) $$\mathbf{E}_{ind}(y_t) = \hat{f}_t \,.\, \hat{s}_t$$

We might view this as a classical structural model of the risk process. However, even if $\hat{f}_t$ and $\hat{s}_t$ were estimated efficiently, $\hat{y}_t$ will not be a satisfying estimator, neither statistically nor economically: On one hand, correlations between the frequencies and the size distribution, which induce a bias of order 2 $Cov(f_t,s_t)$ are strictly ignored. Furthermore $\hat{f}_t$ and $\hat{s}_t$ might be based on variables that are relevant for the submodels but need not influence y_t significantly. Consider variables indicating urban regions, with higher claim frequencies and lower average claims. The costs of motor liability insurance might still be similar to non-urban regions. However testing the influence of such a dummy within the final $\hat{y}_t$ - specification will usually be omitted.

But the major drawback of the structural approach is, that the initial problem, the estimation of a workable tariff-function $\ddot{y}_t$ is not even touched yet. Given $\mathbf{E}_{ind}(y_t)$ a tariff-function that could be applied in practical situations will in general only be found after further approximations. In the case of purely linear $\hat{f}_t$ - and $\hat{s}_t$ - specifications $\mathbf{E}_{ind}(y_t)$ will become a fairly improper polynomial risk model. A multiplicative tariff structure requires the estimation of purely multiplicative submodels, but this is the only way to gain a rating model from the individual model. And obviously, this tariff structure is not always the required one.

2.4 Observational equivalent representations

It might be argued, that multiplicative models are wide enough to cover almost all relevant tariff structures and that additve components could after a proper transformation of variables be as well be estimated within a multiplicative framework. The first argument is exhausted by the empirical paractice, where almost everywhere additive baseline-tariffs are applied. The second argument will be relativated by the following considerations:

Suppose that two GRMs which are based on the same set of criteria lead to the same premium level for all risk groups although they are obviously differently structured. We will denote such tariffs as observational equivalent representations. For example, if criteria of the basic-tariff - let us denote them by D_t - were taken into account as discount factor and the former and the latter tariff specified by the left and the right side of the equation

$$(2.8) \qquad (X_t\alpha + D_t\delta).\exp\{Z_t\beta\} = X_t\alpha.\exp\{Z_t\beta + D_t\delta\}$$

lead to the same premium level for all t, then the two tariffs are observational equivalent.

In the following we will assume that the parameters of $\ddot{y}_t = (X_t\alpha + D_t\delta).\exp\{Z_t\beta\}$ can be identified, such that there exists no other α, δ and β leading to the same premium distribution. Thus we will not consider cases where the columns of X and D respectively those of Z are linear dependent.

Theorem: Assume that the criteria X_t, D_t and Z_t are linear independent.
Given an identified rating model a displacement of criteria between the basic tariff and the multiplicative components leads to an observational equivalent representation if and only if the manipulation can be traced back to separate shifts of single variables D_t following the form (2.8) where

(a) the basic tariff $X_t\alpha + D_t\delta$ defines at most k_X+2 different premia or

(b) D_t is a scalar dummy and either $D_t.X_t\alpha$ or $(1-D_t).X_t\alpha$ is constant for all t.

Proof:

Notice, that the k_x+1 - dimensional vector $X_t = (1, X_{t,1}, \dots, X_{t,kx})$ includes at least a constant term.

Given α and δ, the equation system $X_t\alpha + D_t\delta = X_t\alpha.\exp\{D_t\delta\}$ can be explicitly solved in the k_x+2 parameters α and δ if it has to be fulfilled for at most k_x+2 different riskgroups. This case, corresponding to part (a) of the theorem. An analytical solution for (2.8) will also be aviable if the complete system defines at most k_x+2+k_z different risk premia. However, for linear independet Z_t this requires, that condition (a) is fullfilled.

Let us turn to part (b):

For D_t=0 and arbitrary X_t (2.8) implies $\alpha = \alpha$ and $\beta = \beta$. Now suppose D_t=1. In this case the reduced system $(X_t\alpha + \delta) = X_t\alpha.\exp\{\delta\}$ can only be solved for δ if all X_t are constant, say $X_t = X_c$ whenever D_t=1. In that case the solution is:

$$\text{If } D_t=1 \Rightarrow X_t=X_c \text{ then } \quad \alpha=\alpha,\ \beta=\beta \text{ and } \delta=X_c\alpha.(\exp\{\delta\}-1) \text{ respectively } \delta=\log\left(\frac{X_c\alpha+\delta}{X_c\alpha}\right)$$

If D_t took another value beside 0 and 1 the parameters can no longer be identified. Thus observational equivalent transformations have to be based on thedisplacement of dummy variables.

On the other hand, any structure based on a dummy D_t can as well be represented by an indicator of the complementary group, say by $\bar{D}_t$=1-D_t. By using the notation $\alpha_r = (\alpha_1, \alpha_2, \ldots, \alpha_{kx})'$ the l.h.s. of (2.8) can be expressed by $(\alpha_0+\delta+X_{r,t}\alpha_r - \delta.\bar{D}_t).\exp\{Z_t\beta\}$. Similar to the arguments above, one will find the solution for cases where X_t is constant whenever $\bar{D}_t$=1, e.g. D_t=0:

$$\text{If } D_t=0 \Rightarrow X_t=X_c \text{ then } \quad \beta=\beta$$

$$\text{and} \quad \delta=(X_c\alpha)\,[\exp(\delta)-1],\quad \alpha_0=\alpha_0\exp(\delta)+X_c\alpha.[1-\exp(\delta)],\quad \alpha_r=\alpha_r\exp(\delta),$$

$$\text{respectively} \quad \delta=\log\left(1+\frac{\delta}{X_c\alpha}\right),\qquad \alpha_0=\frac{\alpha_0+\delta}{1+\delta/(X_c\alpha)},\qquad \alpha_r=\alpha_r\exp(-\delta) \quad \blacksquare$$

Remarks:

- For practical purposes case (a) is almost irrelevant. It might occure when very few of the possible combinations of X and D - criteria are relevant for the tariff. When X_t includes only the constant term, e.g. k_X =0, case (a) correspondents to basic tariffs $\alpha_0 + D_t\delta$ which defines at most two risk groups, e.g. D_t denotes a dummy. Thus given a pure multiplicative model any single dummy can be transformed to become a baseline tariff criterion.
- Case (b) is relevant because rating factors are often cathegorial and ordinal criteria, which are expressed by dummy-variables.
- Due to (b) it is obviously that given a pure multiplicative model any single cathegorial risk criteria can be transformed to become a baseline tariff criterion.

In practice it can be easily verified whether two estimations represent equivalent models. As equivalent representations generate the same expected values $\ddot{y}_t$ for all t, all overall-model criteria of the two estimations will coincide. Thus identical Likelihood-statistics indicate the existence of observational equivalent representations with high probability.

3 ESTIMATION

3.1 Distributional assumptions

To develop efficient estimators we introduce distributional assumptions for the average benefit y_t.

Suppose that individual - possibly aggregated - data (y_t, w_t; X_t, Z_t) are observed. Here w_t stands for the natural weight. Usually y_t will be the average risk burden of w_t independent observations that depend on the criteria X_t and Z_t.

We will require that $y_t| X_t, Z_t$ is independently distributed with mean μ_t and variance proportional to $1/w_t$:

A1: $\mathbf{E}(y_t| X_t, Z_t) = \mu_t$

A2: $\mathbf{Var}(y_t| X_t, Z_t) \propto 1/w_t$

Notice that these assumptions have statistical and economical implications: Obviously all risks belonging to one cell of a tariff table that is characterized by X_t and Z_t should be be rated with a common premia $\ddot{y}_t$. Of course even within the same cell $\mathbf{E}(y_t)$ might vary according to further factors not included in the information set $\{X_t, Z_t\}$. Nevertheless the insurance will rate the risks according to the conditional expected value $\mu_t = \mathbf{E}(y_t| X_t, Z_t)$ although it is not the true model and can be „only" interpreted as quasi-ML-approach.

3.2 A unified framework for ML-estimators

ML-estimators for GRMs can be easily constructed, when the Likelihood $\mathcal{L}_t$ is explicitly parameterizised in the expected values μ_t and the dispersion parameter v. When $\partial\mathcal{L}_t/\partial\mu_t$ is known, the partial derivatives with respect to the regression parameters at $\ddot{y}_t = X_t\,\alpha.\ \exp\{Z_t\beta\}$ can be calculated by the following formulars:

(3.1)
$$\frac{\partial\mathcal{L}_t}{\partial\alpha_i} = X_{t,i}\,.\exp\{Z_t\beta\}.\frac{\partial\mathcal{L}_t}{\partial\,\mu_t} \qquad i=0,\ldots,k_X$$
$$\frac{\partial\mathcal{L}_t}{\partial\beta_j} = Z_{t,j}\,.\ddot{y}_t.\frac{\partial\mathcal{L}_t}{\partial\,\mu_t} \qquad j=1,\ldots,k_Z$$

(3.2)
$$\frac{\partial^2\mathcal{L}_t}{\partial\alpha_i\,\partial\alpha_j} = X_{t,i}\,.X_{t,j}\,.\exp(Z_t\,\beta)^2\,\frac{\partial^2\mathcal{L}_t}{\partial\,\mu_t^2}$$
$$\frac{\partial^2\mathcal{L}_t}{\partial\alpha_i\,\partial\beta_j} = X_{t,i}\,.Z_{t,j}\,.\exp(Z_t\,\beta)\,.\left(\frac{\partial\mathcal{L}_t}{\partial\,\mu_t}+\ddot{y}_t\,\frac{\partial^2\mathcal{L}_t}{\partial\,\mu_t^2}\right),\ \frac{\partial^2\mathcal{L}_t}{\partial\beta_i\,\partial\beta_j} = Z_{t,i}\,.Z_{t,j}\,.\ddot{y}_t.\left(\frac{\partial\mathcal{L}_t}{\partial\,\mu_t}+\ddot{y}_t.\,\frac{\partial^2\mathcal{L}_t}{\partial\,\mu_t^2}\right)$$

(3.3)
$$\frac{\partial^2\mathcal{L}_t}{\partial\alpha_i\,\partial v} = X_{t,i}\,.\exp\{Z_t\beta\}\,\frac{\partial^2\mathcal{L}_t}{\partial\mu_t\,\partial v}, \qquad \frac{\partial^2\mathcal{L}_t}{\partial\beta_j\,\partial v} = Z_{t,j}\,.\ddot{y}_t.\,\frac{\partial^2\mathcal{L}_t}{\partial\mu_t\,\partial v}$$

For the most relevant distribution applied in insurance models, the partial derivatives are summarized in table 1 to 3.

Table 1 : ML-Estimation for the Normal- and the Modified Poisson Model

	Normaldistibution	**Modified Poisson**
Log-Likelihood $\mathcal{L}_{t(\,;yt)}$	$-\dfrac{w_t\,(y_t-\mu_t)^2}{2\sigma^2}$ $-\tfrac{1}{2}\,\mathrm{Ln}(2\pi\sigma^2/w_t)$	$\dfrac{w_t y_t}{v}.\mathrm{Ln}\left(\dfrac{w_t\,\mu_t}{v}\right)-\dfrac{w_t\,\mu_t}{v}-$ $\mathrm{Ln}(y_t.\Gamma(w_t y_t/v))$
$\mathrm{Var}(y_t)$	$\sigma^2/w_t = v/w_t$	$v\,\mu_t/w_t$
Auxiliary terms given $E(y_t)=\ddot{y}_t$	$\mathcal{D}_t = w_t\,.\,(y_t-\ddot{y}_t)^2$	$z_t = w_t.\ddot{y}_t/v$ $q_t = w_t\,.\,(y_t-\ddot{y}_t)/v$
$\partial\mathcal{L}_t/\partial\mu_t$	$w_t\;(y_t-\ddot{y}_t)/v$	$w_t\dfrac{y_t-\ddot{y}_t}{v\,\ddot{y}_t}=q_t/\ddot{y}_t$
$\partial\mathcal{L}_t/\partial v$	$(\mathcal{D}_t/v-1).\dfrac{1}{2v}$	$\dfrac{w_t y_t\,\mathrm{Ln}(y_t/\ddot{y}_t)}{v^2}-\dfrac{q_t}{v}-\mathbf{LD}(w_t y_t/v)/v$
$\partial^2\mathcal{L}_t/\partial\mu_t^2$	$-\dfrac{w_t}{v}$	$-\dfrac{w_t y_t}{v\,\ddot{y}_t^2}$
$\partial^2\mathcal{L}_t/\partial\mu_t\partial v$	$-\dfrac{w_t(y_t-\ddot{y}_t)}{v^2}$	$-q_t/(v\,\ddot{y}_t)$
$\partial^2\mathcal{L}_t/\partial v^2$	$[\tfrac{1}{2}-\mathcal{D}_t/v].\dfrac{1}{v^2}$	$\left\{\tfrac{1}{4}+q_t+\dfrac{w_t y_t}{v}\ln\left(\dfrac{\ddot{y}_t}{y_t}\right)\right\}.\dfrac{2}{v^2}+\dfrac{\mathbf{R}(w_t y_t/v)}{v^2}$

Remarks: Ψ denotes the Digamma-Function , **LD**(x) ≡ x (Ln(x)-Ψ(x)),
R(x) ≡- ½ + x +2 x Ln(x) - 2 x Ψ(x) - x² Ψ'(x).

3.3 Examples

3.3.1 Gamma Distribution

Following Mack (1997, p. 71) a proper volume depending parametrisation of the Gamma distribution is

$$f(y_t) = \left(\frac{w_t\,y_t}{v\,\mu_t}\right)^{w_t/v} .\exp\left(-\frac{w_t\,y_t}{v\,\mu_t}\right) / \,(y_t\,\Gamma(w_t/v)) .$$

$\Gamma(\cdot)$ denotes the Gamma function, wt the natural weight of the t^{th} observation. In this specification the mean and the variance of y_t is immediately μ_t and $v.\mu_t^2/w_t$ respectively. Its scewness is $2.\sqrt{v/w_t}$. The special case of an exponential distribution occurs with $v = w_t$. With increasing v the modus of y_t will be strict positive and lies above $\mu_t(w_t-v)/w_t$. For Austrian motor liability data v was estimated in the region of 135% to 142% of $E(w_t)$ such that the modal value lies 65% to 58% below the mean (Fels, 1998).

The Likelihood function is

$$\mathcal{L}_{t(\mu t,\,v;\,y_t)} = \frac{w_t}{v}\,\mathrm{Ln}\left(\frac{w_t\,y_t}{v\,\mu_t}\right)-\frac{w_t\,y_t}{v\,\mu_t}-\mathrm{Ln}(y_t\,\Gamma(w_t/v))$$

The relevant derivatives can be found in table 2 (page 9).

Table 2: ML-Estimation for the Gamma- and Inverse Gaussian Model

	Gamma	Inverse Gaussian
Log-Likelihood $\mathcal{L}_{t(;yt)}$	$-\frac{w_t\, y_t}{v\, \mu_t} + \frac{w_t}{v} \text{ Ln } \left(\frac{w_t y_t}{v \mu_t}\right)$ $-\text{Ln}(y_t\, \Gamma(w_t/v))$	$-\frac{w_t}{2v}\left(\frac{y_t}{\mu_t^2} - \frac{2}{\mu_t} + \frac{1}{y_t}\right)$ $-\frac{1}{2} \text{Ln}(2\pi\, y_t^3\, v/w_t)$
$\text{Var}(y_t)$	$v\, \mu_t^2 / w_t$	$v\, \mu_t^3 / w_t$
Auxiliary terms given $E(y_t) = \ddot{y}_t$	$q_t = w_t\, .\, [y_t/\ddot{y}_t - 1]/v$ $\mathcal{D}_t = w_t\, .\, [y_t/\ddot{y}_t - 1]^2$	$\mathcal{D}_t = w_t\, .\, (y_t^{-1} - \ddot{y}_t^{-1})$
$\partial\mathcal{L}_t/\partial\mu_t$	$\frac{w_t}{v} \cdot \frac{y_t - \ddot{y}_t}{\ddot{y}_t^2} = q_t/\ddot{y}_t$	$-\mathcal{D}_t\, y_t / (v\, \ddot{y}_t^2)$
$\partial\mathcal{L}_t/\partial v$	$\frac{q_t}{v} - \frac{w_t}{v^2}\left(\text{Ln}\left(\frac{w_t\, y_t}{v\, \ddot{y}_t}\right) - \Psi\left(\frac{w_t}{v}\right)\right)$	$\frac{y_t\, \mathcal{D}_t^2}{2\, w_t\, v^2} - \frac{1}{2v}$
$\partial^2\mathcal{L}_t/\partial\mu_t^2$	$\frac{w_t}{v\, \ddot{y}_t^2} \cdot \left(1 - 2\frac{y_t}{\ddot{y}_t}\right)$	$\frac{y_t\, (2\, \mathcal{D}_t - w_t/\ddot{y}_t)}{v\, \ddot{y}_t^3}$
$\partial^2\mathcal{L}_t/\partial\mu_t\, \partial v$	$-\frac{w_t\, (y_t - \ddot{y}_t)}{v^2\, \ddot{y}_t^2} = -\, q_t\, /(\ddot{y}_t\, v)$	$y_t \mathcal{D}_t / (v\, \ddot{y}_t)^2$
$\partial^2\mathcal{L}_t/\partial v^2$	$\frac{2}{v^2} \cdot \left(\frac{1}{4} + \frac{w_t}{v} \text{Ln}\left(\frac{y_t}{\ddot{y}_t}\right) - q_t\right) + \frac{\mathbf{R}(w_t/v)}{v^2}$	$\frac{1}{2\, v^2} - \frac{y_t \mathcal{D}_t^2}{w_t\, v^3}$

3.3.2 Inverse Gaussian Distribution

For an Inverse Gaussian distribution with mean $\mu_t = X_t\alpha\, .\exp(Z_t\beta)$ and variance $v.\mu_t^3/w_t$ the log-likelihood

$$\mathcal{L} = -\Sigma_{t=1}^{T} \frac{w_t}{2v}\left(\frac{y_t}{\mu_t^2} - \frac{2}{\mu_t} + \frac{1}{y_t}\right) - \Sigma_{t=1}^{T}\ \frac{1}{2} \text{Ln}(2\pi\, \mu_t^3\, v/w_t)$$
$$= -\Sigma_{t=1}^{T} \frac{w_t\, y_t}{2v}\left(\frac{1}{\mu_t} - \frac{1}{y_t}\right)^2 - \Sigma_{t=1}^{T}\ \frac{1}{2} \text{Ln}(2\pi\, y_t^3\, v/w_t)$$

has to be maximised. For each v this coincides with the minimal deviance, e.g. the minimum of

(3.4) $$\mathcal{D} = \Sigma_t\, w_t\, \frac{(y_t - \ddot{y}_t)^2}{\ddot{y}_t^2\, y_t}$$

Following T. Mack (1997, p. 51) given estimates $\ddot{y}_t$ for $E(y_t)$ the parameter v can be initialised respectively estimated by $\hat{v} = \Sigma_{t1}\ w_t(1/y_t\ - 1/\ddot{y}_t) / T$.

Table 3: Lognormal Distributions with 2 and 3 parameters

$$y_t \approx \mathbf{LnN(}\ Ln(\mu_t) - ½\, Ln(1+ \mu^{s-2}.v/w_t),\ \ Ln(1+\mu^{s-2}.v/w_t)^{½}\ \mathbf{)}$$

	3-parametric case	**2-parametric** (s=2)
Log-Likelihood $\mathcal{L}_{t(;yt)}$	$-\frac{\{½ zt\ +\ Ln(y_t\ /\ \mu_t\}^2}{2\, z_t} - ½Ln[2\,\pi\, y_t^2\, z_t\,]$	
with	$z_t = Ln[1+v\,\mu^{s-2}/w_t]$	$z_t = Ln[1+v/w_t]$
$Var(y_t)$	$v\,\mu_t^{s}/w_t$	$v\,\mu_t^2/w_t$
Auxiliary terms given $E(y_t) = \ddot{y}_t$	$q_t = ¼ + \frac{1}{z_t} - \left(\frac{Ln(y_t/\ddot{y}_t)}{z_t}\right)^2;\quad h_t = \frac{v\,\mu_t^{s-2}}{w_t + v\,\mu_t^{s-2}}$	$q_t = ¼ + \frac{1}{z_t} - \left(\frac{Ln(y_t/\ddot{y}_t)}{z_t}\right)$
$\partial\mathcal{L}_t/\partial\mu_t$	$\frac{1}{2\,\ddot{y}_t}\left(\frac{2\,Ln(y_t/\ddot{y}_t)}{z_t} + 1 - (s-2)\,h_t\,q_t\right)$	$\frac{½ + Ln(y_t/\ddot{y}_t)/z_t}{\ddot{y}_t}$
$\partial\mathcal{L}_t/\partial v$	$-½\,h_t\,q_t\,/\,v$	$-½\,q_t\,/(v+w_t)$
$\partial\mathcal{L}_t/\partial s$	$-½\,Ln(\ddot{y}_t)\,h_t\,q_t$	
$\partial^2\mathcal{L}_t/\partial\mu_t^2$	$-\frac{1}{\ddot{y}_t^2}\left(½ + \frac{Ln(y_t/\ddot{y}_t)+1}{z_t}\right) + \frac{s-2}{2\,\ddot{y}_t^2}\,h_t\left(-q_t\,(s-3)+(s-2)\,h_t\,\frac{2+z_t}{z_t}\,[q_t - 1/(2\,z_t)] - 4\frac{Ln(y_t/\ddot{y}_t)}{z_t^2}\right)$	$-\frac{1}{\ddot{y}_t^2}\left(½ + \frac{Ln(y_t/\ddot{y}_t)+1}{z_t}\right)$
$\partial^2\mathcal{L}_t/\partial v^2$	$\frac{h_t^2}{2\,v^2}\left(q_t + \frac{1}{z_t^2} - \frac{2\,Ln(y_t/\ddot{y}_t)^2}{z_t^3}\right)$	$\frac{1}{2\,(v+w_t)^2}\left(\frac{1}{z_t^2} + q_t - \frac{2\,Ln(y_t/\ddot{y}_t)^2}{z_t^3}\right)$
$\partial^2\mathcal{L}_t/\partial s^2$	$½\,Ln(\ddot{y}_t)^2\,q_t\ \ h_t\,(h_t-1) - ½\left(\frac{Ln(\ddot{y}_t)\,h_t}{z_t}\right)^2 . \left(\frac{2\,Ln(y_t/\ddot{y}_t)^2}{z_t} - 1\right)$	
$\partial^2\mathcal{L}_t/\partial\mu_t\partial v$	$-\frac{h_t}{v\,\ddot{y}_t}\,\frac{Ln(y_t/\ddot{y}_t)}{z_t^2} + \frac{s-2}{2}\,\frac{h_t}{v\,\ddot{y}_t}\ (-q_t + h_t\,(1+2/z_t)\,[q_t - 1/(2\,z_t)\,]\,)$	$\frac{-\,Ln(y_t/\ddot{y}_t)}{z_t^2\ \ \ddot{y}_t\ \ (v+w_t)}$
$\partial^2\mathcal{L}_t/\partial\mu_t\partial s$	$-\frac{h_t\,q_t}{2\,\ddot{y}_t} - \frac{h_t\,Ln(\ddot{y}_t)}{2\,\ddot{y}_t}\,.\left(\frac{2\,Ln(y_t/\ddot{y}_t) + (s-2)\,h_t}{z_t^2} + \frac{(s-2)\,(½ - 2q_t)\,h_t}{z_t} - (s-2)\,q_t\,(h_t-1)\right)$	
$\partial^2\mathcal{L}_t/\partial v\,\partial s$	$-\frac{h_t\,Ln(\ddot{y}_t)}{v}\left(\frac{q_t\,(1-h_t)}{2} + \frac{h_t}{z_t^2}\left(\frac{Ln(y_t/\ddot{y}_t)^2}{z_t} - \frac{1}{2}\right)\right)$	

3.3.3 Lognormal Distribution

It is well known, that when the random variable x is distributed lognormal, say $x \approx LnN(\mu_x,\sigma_x)$, then $exp(x) \approx N(\mu_x,\sigma_x)$. This relation suggests to estimate a Gaussian model for $x_t \equiv exp(y_t)$ whenever the risk burden y_t is Lognormal. Although such estimations have often been applied in practical situations, they are quite misleading. Because in this case the moments of y_t would become

(3.5) $$E(y) = e^{\mu_x + ½\sigma_x^2} \qquad Var(y) = e^{2\mu_x + \sigma_x^2}(e^{\sigma_x^2} - 1)$$

Estimating the mean of y_t by $\exp E(\text{Log}(y)) = \exp(\mu_x)$ will underestimate E(y) systematically $e^{\frac{1}{2}\sigma_x^2}$ - times.

Furthermore this approach will neither allow to specify $E(y_t)$ independent of the variance components σ_x and w_t nor lead to an uniformly in w_t decreasing variance. To fullfill the fundamental moment assumptions A1 and A2 the lognormal distribution has to be specified by

$$(3.6) \qquad y_t \approx \text{LnN}\left(\text{Ln}(\mu_t) - \tfrac{1}{2}\,\text{Ln}(1+v\mu^{s-2} w_t^{-1}),\ \text{Ln}(1+v\mu^{s-2} w_t^{-1})^{1/2}\right)$$

$$\text{s.t.} \qquad \text{Prob}(y_t = y) = \frac{\text{Exp}\left(-\frac{1}{2}\left(\text{Ln}\frac{y}{\mu_t} + \frac{1}{2}\text{Ln}\left(1+\frac{v\,\mu_t^{s-2}}{w_t}\right)\right)^2 \Big/ \text{Ln}\left(1+\frac{v\,\mu_t^{s-2}}{w_t}\right)\right)}{y\sqrt{2\pi\,\text{Ln}\left(1+\frac{v\,\mu_t^{s-2}}{w_t}\right)}}$$

This lognormal specification, which was developed by Fels (1999 a) yields

$$(3.7) \qquad E(y_t) = \mu_t, \qquad \text{Var}(y_t) = v\,\mu_t^s / w_t$$

For given μ the variance depends on two parameters, v and s. Although the simultaneous estimation of all parameters is possible, the estimation becomes more complex and especially with bad initial values quite unpleasant.

Usually apriori knowledge of the relation between μ_t and $\text{Var}(y_t)$ is available. In this case it is sufficient to estimate the model for fixed s. For example, if a quadratic mean-variance relation similar to the Gamma model is considered, e.g. $\text{Var}(y_t) = v\,\mu_t^2/w_t$, we can set s=2 and estimate only v and μ_t. This will reduce the complexity of the estimation and lead to stable estimators even in the case when the basic assumption A2 is violated. The latter problem was discussed by Fels (1999, b).

3.3.4 Further distributions

Maximum-likelihood estimators for general rating models under the assumption of normal and the modified Poisson distributed data are developed and discussed in W. Fels (1999, a). These results are summarised in table 1.

It should be noted that the structure of general rating models can also be applied for discrete claim frequency specifications. Fels (1999, a) presented also the estimators for the negative binomial and the Lagrange-Poisson distribution.

3.4 The ML-estimator

This approach is purely based on Maximum-Likelihood (ML) estimation, a method, that is sufficiently discussed in literature. We have no new contribution on ML. The remaining chapter serves only to summarize the most relevant aspects for the implementation of the estimators.

3.4.1 The Newton-Raphson algorithm

Let us denote the relevant parameter vector of the GRM (2.2) by

$$\theta \equiv (\alpha', \beta', v)' = (\alpha_0 , \alpha_1 , \alpha_2 , ..., \alpha_{kx}, \beta_1, \beta_2, ..., \beta_{kz} , v)'$$

The dimension of this rowvector is equivalent to the number of repressors ($1+k_x+k_z$) and k_v, the dimension of the dispersion parameter v, which is usually scalar. Let us denote the number of parameters by k_θ.

The Newton-Raphson algorithm for the ML-estimator requires a sequential updating of the parameters. Given consistent initial values θ_0, then estimation might be updating the parameters in the n^{th} step due to

(3.8) $$\theta_n = \theta_{n-1} + \mathbf{H} \cdot \frac{\partial \mathcal{L}}{\partial \theta}\Big|_{\theta n-1} \qquad \text{with } \mathbf{H} \text{ is } \mathbf{F}^{-1} \text{ and } \mathbf{F} \equiv -\Sigma_t \frac{\partial^2 \mathcal{L}_t}{\partial \theta\, \partial \theta'} .$$

Notice that **H** and the Fisher information matrix **F** are of dimension $k_\theta \times k_\theta$.

Given $\sqrt{T}$ - consistent initial values θ_0 the Newton-Raphson iteration ensures the existence of an efficient estimator for regular models (see L. Le Cam, 1990). Unfortunately $\sqrt{T}$-consistent starting values for α can be estimated by linear regression only under the condition that β=0. As in general no consistent initial values for the whole parameter vector can be found, iterations might lead to irregular **F**-estimators. In the context of rating models where the Likelihood is defined only for $\ddot{y}_t \geq 0$ this will typically be associated with observations where $X_t\, \hat{\alpha} \leq 0$, s.t. $E(y_t)$ becomes nonpositive. In this case one might specify other starting values or switch to the Bernd-Hall-Hall-Hausmann algorithm where in (3.8) the matrix **H** is replaced by $-\mathbf{J}^{-1}$ with

(3.9) $$\mathbf{J} = \Sigma_t \frac{\partial \mathcal{L}_t}{\partial \theta}' \frac{\partial \mathcal{L}_t}{\partial \theta} .$$

3.4.2 Estimation for exponential distributions

It is well known that for exponential distributions the derivatives of the Likelihood with respect to the location parameters μ_t can be factored in $\partial \mathcal{L}_t / \partial \mu_t = L'_t / \mathcal{D}$ respectively $\partial^2 \mathcal{L}_t / \partial \mu_t^2 = L''_t / \mathcal{D}$ where L'_t and L''_t are independent of the scale parameter $\mathcal{D}$.

To estimate the concentrated parameter vector $\theta_c = (\alpha_0 , \alpha_1 , \alpha_2 , ..., \alpha_{kx}, \beta_1, \beta_2, ..., \beta_{kz})$ the Newton-Raphson-algorithm (3.8) can be applied for the lower dimensional system without knowledge of the dispersion parameter v. Taking L instead of $\mathcal{L}$ only the derivatives (3.1) and (3.2) have to be applied. The required expression for L'_t and L''_t result immediately from $\partial\mathcal{L}_t / \partial\mu_t$ and $\partial^2\mathcal{L}_t/\partial\mu_t^2$ reported in table 1 and 2 when σ^2 and v are set to one for the Gaussian- respectively the Gamma and the Wald-distibution.

3.4.3 Inference

Within the ML-framework the basic-criteria for GRMs is the Likelihood statistic $\mathcal{L}(\theta) = \Sigma_t \mathcal{L}_t(\theta;y_t)$.

For example, the test statistic for the hypothesis that d variables within a k-dimension model have no effect, results from a comparing the Likelihood of the k -dimension model, say $\mathcal{L}(\theta_k)$, with that of the reduced estimation $\mathcal{L}(\theta_{k-d})$. If the eliminated variables have no effect, the Likelihood-ratio statistic

$$(3.10) \qquad \boldsymbol{LR}(k, k\text{-}d) = 2\big(\mathcal{L}(\theta_k) - \mathcal{L}(\theta_{k-d}) \big)$$

will be asymptotically $\chi^2(d)$-distributed.

Another approach for testing the validity of parameter restrictions results from the fact, that the asymptotic distribution of the ML estimate $\hat{\theta}$ is unbiased normal with $\mathrm{Var}(\hat{\theta}) = \mathbf{H}|_{\hat{\theta}}$. Here $\mathbf{H}^{-1}$ can either be the Fisher matrix **F** defined in 3.8 or **J** (see 3.9). Based on this normal approximation a 100 α% confidence ellipsoid is given by

$$(3.11) \qquad \chi^2(k_\theta) \geq (\hat{\theta} - \theta).\, \mathbf{H}.(\theta - \hat{\theta})'$$

where $\chi^2(k_\theta)$ is the α% fractile of the chi-square distribution with k_θ degrees of freedom.

Taking into account the possibility of misspecified distributional assumptions, Halbert White's (1982) Quasi-ML covariance estimator with $E(\hat{\theta} - \theta).(\theta - \hat{\theta})' = [\mathbf{F}\,\mathbf{J}^{-1}\,\mathbf{F}]^{-1}/T$ should be applied.

Generalized Linear Models are often judged according to the scaled deviance, which can be interpreted as LR-Statistic that compares the Likelihood of the current model $\mathcal{L}(\theta_k)$ with that of a full model $\mathcal{L}_f(\theta_T)$, where the $\mathcal{L}_{f,t}$ is evaluated at $\ddot{y}_t = y_t\ (k_\theta)$ with $v=\hat{v}_k$:

$$(3.12) \qquad \boldsymbol{S}(\theta_k) = 2\big(\mathcal{L}_f(\theta_T) - \mathcal{L}(\theta_k) \big) = 2\ \Sigma_t\, \mathcal{L}_{f,t}(\theta_T;\, \ddot{y}_t) - \mathcal{L}_t(\theta_k;\, y_t)$$

Information criteria are usually applied to evaluate the parsimonity of a model. The widely applied Schwarz-Bayes information criteria (Schwarz, 1978) judges a tariff structure with k_θ-parameters by BIC $= \mathcal{L}(\theta_k) - ½\, k_\theta \ln(T)$. Tariff stuctures with higher BIC statistics should be favoured.

4 MODEL CHARACTERISTICS

4.1 Database

The following analysis are based on Austrian motor liability data. The Austrian Association of Insurance Companies collects all relevant information on an individual base, accumulates the data according to several quite general multivariate combinations of risk criteria and returns it to its members on a CD-ROM.

Table 4: Variable List

Code	Criteria	Variable	Description
Y_t	Benefits	VLTOTL	Average total benefits of an annual liability contract, including total payment, direct regulation costs and reserves
w_t	Nat. Weight	JE	Years insured
kW	Motor Power	kWC8	Ordinal indicator of the kilowatt-power 2 = up to 26 kW, 3 = up to 30 kW, 4 = up to 40 kW, 5= up to 55 kW, 6 = up to 67 kW, 7 = up to 89 kW, 8 = up to 111 kW, 9 = more than 111 kW
		kWC82	Variable defined by KWC8² For example: KWC82=25 indicates cars with 40 to 55 kW power
		kW67b89	Dummy variable for cars with 67kw to 89 kW power
		kW89b111	Dummy variable for cars with 89kw to 111 kW power
		kWgt111	Dummy variable for cars with more than 111 kW power
Age	Age of the policy holder	b24J	Dummy indicating young policy holders up to 24years
		J25b29	Dummy indicating policy holders between 25 and 29 years
Fem	Gender	Fem	Dummy indicating female policy holders
YouM	Gender/Age	YouM	Dummy for young male policy holders up to 24 years
BM	Bonus /Malus	BMRaba	Applied bonus - malus rating factor. For example: risks of bonus class 04 are rated with a 30% discount s.t. BMRABA = 0.7.

For 1996 information from about 1.6 million privately used passenger cars is available. It covers a volume of 1.2 millions of years insured. Taking into account only data from cars for less than 5 passengers where the gender and the age of the policy holder is reported there are still left about 1.2 millions contracts that represent about 900,000 years insured. The adequate dataset that includes all relevant cross-classifications of the mentioned criteria condensed the primary data to 6195 cells[1]. The relevant variables and tariff criteria are summarized in table 4.

Other risk criteria will not be analyzed within this paper, although it has been shown that some available variables like the age of the car have a relevant discriminating power (Fels, 1998). It is also well known,

[1]) The original dataset including also missing values is the file K_6_214.CSV published by the Austrian Association of Insurance Companies (see VVO, 1998).

that the fit of motor rating models can be dramatically be increased, when more differenced scales for young drivers were applied. However, rating young men between 18 and 20 years separately required a technical extra charge of about 150% for this subgroup. As insurance companies would not accept such extreme increases, we prefer to rate the youngest risks together with those up to 24 years with a common average extra charge.

4.2 Insurance Benefits

Let us first analyze criteria that influence the total insurance benefits. Before discussing the final results two remarks are necessary:

First, it is not necessary to specify seven dummies for the eight kW-classes. As the risk burden increases almost linear with the motorpower indicator KWC8, a regression on that variable could explain the majority of the variance between the power classes. However cars between 40 and 55 kW have almost the same requirements as cars with 55 to 67 kW and the progression for the strongest cars is degressive. To map these derivatives form linearity it is sufficient to specify the progression with the five kW-variables listed in table 4, e.g. the counting number kWC8, its square and three dummies for strong cars.

We will estimate the models under the Gamma-distribution assumption. Once it can be shown that the squared residuals of simple estimations are proportional to $E(y_t)^{1.9}$ which indicates an almost gamma-like relation. In comparison several rating specifications with the Lognormal-, the Modified Poisson- and the Wald-Distribution the Gamma specification yields the highest Likelihood. These estimations will not be discussed in the following.

Models for insurance benefits

Mod.	Base Tariff (X-Variables)	Extra Charge (Z-Variables)	Likelihood	t(Fem)	Lik exkl. Fem	LRStat
M51	kW, Age, Fem		-55420.86	-0.55	-55421.00	0.298
M52	kW	Age, Fem	-55403.41	-2.20	-55405.78	4.745
M53	Age, Fem	kW	-55402.97	-0.10	-55402.97	0.011
M54	Age	kW, Fem	-55400.72	-2.14	-55402.97	4.495
M54		kW, Age, Fem	-55400.72	-2.14	-55402.97	4.495
M55	Age, YouM	kW, Fem	-55369.82	0.44	-55369.92	0.190
M55		kW, Age, Fem, YouM	-55369.82	0.44	-55369.92	0.190
M56	Age, Fem, YouM	kW	-55369.50	0.90	-55369.92	0.843

Table 5 Remarks:

t(Fem) t-statistic for the hypothesis that the coefficient of the gender dummy Fem is zero. Within these specifications this statistic is not exact t-distributed. But the critical values -1,6 and -2,3 might still serve as approximate limits for a 95% respectively 99% significance test.

LRStat : Likelihood Ratio comparing the models with and without Fem. Under the null hypothesis LRStat is asymptotically $\chi^2(1)$ distributed. Thus for LRStat <3.8 and <6.6 an overall gender effect can be rejected at the 95% respectively 99% level of significance.

Let us take a look at the linear model M51. Taking into account the motor power and the age of the policy holder, the benefits of women are estimated to be 24 ATS, e.g. about 2 € less than these of men. The

Likelihood Ratio test that compares this specification to one without the variable FEM rejects gender-effects very clearly (LRStat=0,3).

On the other hand, the purely linear model M51 has a relative poor Likelihood compared with the following specifications of table 5. Model M54, where based on the two age groups extra charges for the car power and the gender are multiplicativly added will perform much better. In this specification women are about 4,5% better then men ($ß_{FEM}$ = -0.04037). In models M52 and M54 multiplicative gender effects are evident at the 95% but not at the 99% level of significance. Nevertheless model M53 demonstrates that gender has no effect if a base tariff beside an age classification.

However the data could still be fitted better when a specific gender-discrimination for younger risks were specified. Adding the variable YouM to Model M54 yields Model M55 with an 30.1 increase of the Likelihood. The $\chi^2(1)$ - distributed LR-statistic of H0: YouM=0 equals 61,8 and is highly significant. On the other hand testing the influence of FEM within Model M55 and M56 suggests that a gender discrimination for people older than 24 is not justified. With other words: The higher benefits of men result from gender effects of younger policy holders. Men and women older than 24 years have a similar loss-pattern.

This examples illustrate, that inference about the influence of a risk criterion depends as well on other criteria mentioned as on the suggested model structure. Inference from a purely linear or purely multiplicative model can be misleading if a GRM will be applied in practice.

Notice that Model M54 and M55 are reported in two observational equivalent representations. Within a purely multiplicative structure a dichotomised criterion can also be specified as base-tariff criteria without changing the net rates of any tariff cell. However, moving a second criterion to the base tariff will in general change the model specification. For example M55 and M56 are not equivalent.

4.3 Tariff Specification

The typical Austrian car liability tariff falls under the following structure:

(4.1) $\text{Premium} = [\alpha_0 + \Sigma_{i=3}\, \alpha_{i-3}\, I(kWC8_t = i)]\ .\exp\{\beta_1\, Fem_t + \ldots\}.\ \ BMRaba_t$

The basic rating level depends only on the power of the car. Usually a about 10% discounts for women is reckoned up, e.g. β_1 = -0.105. Further criteria, for example a rough classification according to the occupation, are always rated by additional discounts.

Although since 1994 insurance companies are free to develop individual bonus-malus systems, still almost all companies apply the traditional ordered scheme from 1977 with small variations. Even when new tariffs are developed the bonus-malus scale is apriori fixed. This implies that in the estimation of the rating scheme

(4.2) $VLTOTL_t = [\alpha_0 + \Sigma_{i=3}\, \alpha_{i-3}\, I(kWC8_t = i)]\ .\exp\{\beta_1\, FEM_t + \ldots + \ln(BMRaba_t)\}$

no free parameter for ln(BMRaba) can be specified, respectively the parameter of ln(BMRaba) is intrinsically aliased to be one.

Apart from this last exogenous term the tariff estimation will be similar to those of the loss-models discussed in chapter 4.2. Nevertheless the results are quiet different:

The rows in table 6 are ordered according to the Likelihood of the different specifications. Model M64 follows the typical structure of the Austrian rating schemes. It performs much better than a purely linear specification but is inferior to the multiplicative specification M66.

9-variables Gamma-Rating-Models				(without gender-effect for young risks)		
Mod.	Base Tariff (X-Variables)	Extra Charge (Z-Variables)	Likelihood	t(Fem)	Lik exkl. Fem	LRStat
M61	kW, Fem, Age		-54357.33	-4.72	-54368.21	21.762
M62	kW, Age	Fem	-54352.53	-5.64	-54368.21	31.362
M63	Fem, Age	kW	-54345.73	-3.02	-54354.60	17.741
M64	kW	Fem, Age	-54344.08	-5.43	-54358.61	29.068
M65	kW, Fem	Age	-54342.23	-5.80	-54358.61	32.768
M66	Fem	kW, Age	-54340.35			
M66	Age	kW, Fem	-54340.35	-5.37	-54354.60	28.501
M66		kW, Fem, Age	-54340.35			

Table 6

Notice that in all these specifications a gender discrimination can be accepted at the 99% level of significance.

Especially young male drivers are worse risks than young women. Adding an interaction term for men younger than 24 years to the original specifications reported in table 6 will increase the Likelihood in all cases significantly (table 7). The gender discrimination for young risks is not only statistical but also economical relevant: According to model M77 men up to 24 years should be rated with an 89% extra charge compared to men older than 30 years. Young women require only an 28% extra charge while women older than 30 years could be rated with an 4.7% lady bonus compared to similar men.

10-Variables Gamma-Rating Models				(with gender-effect for young risks)		
Mod.	Base Tariff (X-Variables)	Extra Charge (Z-Variables)	Likelihood	t(Fem)	Lik exkl. Fem	LRStat
M71	kW, Age, YouMJ	Fem	-54317.93	-2.67	-54321.46	7.069
M72	kW, Fem, Age, YouM		-54317.50	-2.84	-54321.46	7.929
M73	kW, Age	Fem, YouM	-54316.96	-2.70	-54320.57	7.228
M74	kW	Fem, Age, YouM	-54315.79	-2.79	-54319.65	7.727
M75	kW, Fem	Age, YouM	-54314.93	-3.10	-54319.65	9.447
M76	Fem, Age, YouM	kW	-54312.26	-2.23	-54315.63	6.738
M77	Age, YouM	kW, Fem	-54311.94			
M77	Fem, YouM	kW, Age	-54311.94			
M77	Age	kW, Fem, YouM	-54311.94			
M77		kW, Fem, Age, YouM	-54311.94	-2.73	-54315.63	7.377

Table 7

Even after young men and women are rated at different levels, a gender discrimination of older risks remains significant for all rating models. Since it had already been shown in chapter 4.2 that older men and women create similar costs this result seems curious. The solution of the puzzle lies in the construction of the bonus-system. It could be demonstrated by an Lagrange-Poisson GRM-estimation for claim numbers that women and men older than 24 years have a similar claim frequency patterns. But women receive less bonus due to shorter observation periods in the system. Thus a lady bonus of about 4,7% is justified to compensate the bonus favor that men receive due to longer continuous driving - and insurance - periods.

REFERENCES

LEMAIRE, J. (1985) *Automobile Insurance, Actuarial Models*. Kluwer-Nijhoff Publishing, Boston.

RENSHAW, A. R. (1994) *Modelling the claim process in the presence of covariates*. ASTIN Bulletin **24**. S. 265-285.

MACK T. (1997) *Schadenversicherungsmathematik.* DGVM-Schriftenreihe Angewandte Versicherungs-mathematik **28**. VVW Karlsruhe.

FELS, W., (1998) *Geschlecht und Alter in der Kfz-Haftpflichtversicherung.* Verband der Versicherungsunternehmen Österreichs. Vienna.

FELS, W., (1999, a) *Maximum - Likelihood Schätzung allgemeiner Tarifmodelle.* Verband der Versicherungsunternehmen Österreichs. Vienna.

FELS, W., (1999, b) *Zur Varianzspezifikation durchschnittlicher Bedarfswerte.* Verband der Versicherungsunternehmen Österreichs. Vienna.

WHITE H. (1982) *Maximum Likelihood Estimation of Misspecified Models*. Econometrica **50**, S. 1-25.

LE CAM L. (1970) *On the assumptions used to prove asymptotic normality of maximum likelihood estimates.* Ann. Math. Statistics **41**, 802-828.

BAILEY, R. A. and SIMON, L. J. (1960) *Two studies in automobile insurance ratemaking.* ASTIN Bulletin **1**. 192-217.

MCCULLARG P. and NELDER J.A. (1989) *Generalized Linear Models.* 2nd Edition.. Chapman and Hall.

RUYGT, F. A. (1982) *Vehicle dependent Rating Factors.* In: *New Motor rating structure in the Netherlands.* (ed. G.W. de Wit) 58-71. Astin-group Nederland.

SCHWARZ, G. (1978) *Estimating the Dimension of a Model*, Annals of Statistics, 6, S. 461-464.

SANT, D. T. (1980) *Estimating expected Losses in Automobile Insurance*, Journal of Risk and Insurance 47. 133-151.

STROINSKI K.J. and CURRIE I.D. (1989) *Selection of variables for automobile insurance rating*, IME **8**, 35-46.

JOHNSON, P.D. and HAY G.B. (1971) *Statistical Studies in motor insurance*. Journal of the Institute of Actuaries **97**, 119-249.

JUNG, J. (1968) *On Automobile Insurance Rate Making*, ASTIN Bulletin **5**. 41-48.

NELDER J. A. and WEDDERBURN R. W. M. (1972) *Generalized linear models*, Journal of the Royal Statistical Society, Series A, **135**, 370-384.

VVO (1998) *Handbuch zur CD-Rom "Zukunftsorientierte Kfz-Schadenbedarfstatistik", Version 1996/1997*, Austrian Association of Insurance Companies, Vienna.

1 Introduction **2**

2 Generalized rating Models **2**

2.1 The structure of generalized rating models (GRM) *2*

2.2 Related rating structures and Generalized Linear Models *3*

2.3 What is estimated *4*

2.4 Observational equivalent representations *5*

3 Estimation **7**

3.1 Distributional assumptions *7*

3.2 A unified framework for ML-estimators *7*

3.3 Examples *8*

3.3.1 Gamma Distribution 8

3.3.2 Inverse Gaussian Distribution 9

3.3.3 Lognormal Distribution 10

3.3.4 Further distributions 11

3.4 The ML-estimator *12*

3.4.1 The Newton-Raphson algorithm 12

3.4.2 Estimation for exponential distributions 12

3.4.3 Inference 13

4 Model characteristics **14**

4.1 Database *14*

4.2 Insurance Benefits *15*

4.3 Tariff Specification *16*

5 Alte Textstellen Fehler! Textmarke nicht definiert.

5.1 Observational equivialent spezifications, Ver.1 *20*