ELSEVIER

Modelling losses using an exponential-inverse Gaussian distribution

Nikolaos Frangos*, Dimitris Karlis

Department of Statistics, Athens University of Economics and Business, 76 Patission Street, 10434 Athens, Greece

Received July 2002; received in revised form March 2004; accepted 6 April 2004

Abstract

An exponential-inverse Gaussian distribution is used to model the claim size distribution. The distribution has shorter tails than the Pareto distribution and it is considered as a plausible model for data without large tails. We present the model allowing for covariates. Properties of the model are discussed. An EM algorithm is provided to fit the model. The algorithm is quite simple and programmable without need for any special functions. The model can be seen as a random effect model for exponential survival times regression. A real data application using a car-insurance company portfolio data is provided. © 2004 Elsevier B.V. All rights reserved.

1. Introduction

Generalized linear models (GLM) have been used for actuarial purposes by several authors (see, e.g. Haberman and Renshaw (1996) and references therein). In the framework of actuarial applications, there are several attempts to use GLM in order to describe the claim frequency, the claim size or other characteristics of a portfolio. As far as the claim frequency modelling is concerned, Poisson and negative binomial regression models have been used by many authors (see, e.g. Apostolakis (1998) and references therein). Beirlant et al. (1991) examined via a logistic regression model the probability that at least one claim will occur for a contract with given characteristics. In addition they fitted an accelerated failure time model to examine the size of the claim. Loss distributions are described in depth in the book of Klugman et al. (1998). Parametric regression models are available for a range of distributions, like the exponential, the Pareto and the Burr family. Many of them are popular in survival analysis, too. Beirlant et al. (1998) used a Burr regression model for portfolio segmentation purposes.

---

* Corresponding author.
E-mail addresses: nef@aueb.gr (N. Frangos), karlis@aueb.gr (D. Karlis).

The aim of the present paper is to propose a new model which is similar in nature to the Pareto model. In fact the model differs from the standard Pareto one by using an inverse Gaussian mixing distribution for the parameter of the exponential density instead of the gamma one, as the derivation of the Pareto distribution is based on. It is important to note that different parameterizations of the exponential distribution may lead to other models. The model is fitted to data from a large Greek insurance company, concerning the size of car accident claims. An EM type algorithm is provided that facilitates the estimation procedure. Model diagnostics are provided.

To motivate further the new model, note that it can be considered as candidate model for data with moderate tails. In order to have a picture for the tail of the distribution proposed, we follow Wang (1998) and his right tail index. Comparing with other densities (see Table 3 of Wang (1998)), the proposed model has an index 2.31. The E-IG model of that table refers to a different model than ours. This value shows that the tail is not as large as the one for the Pareto model, but larger than that of the gamma model for example. In addition, the proposed model offers simple estimation procedures, it is always J-shaped and it has a decreasing failure rate. Interestingly, it offers an interpretation based on mixed effects exponential regression model, having potential applications for survival analysis models.

The remaining of the paper proceeds as follows. In Section 2, we derive the new model, while in Section 3 we fully describe ML estimation through an EM algorithm. Section 4 contains an application to a dataset concerning car-insurance claims. Diagnostics for the model, including goodness of fit, residual and influence diagnostics are provided in Section 5. Some further issues are examined in Section 6, while concluding remarks can be found in Section 7.

2. The model

Consider the amount y_i paid for the ith contract and $\mathbf{x}_i$ a vector of covariates related to the ith contract ($i = 1, \ldots, n$), that may contain the characteristics of the insured person or the car. The model assumes that

$$y_i | \theta_i t_i \sim f(y_i | \theta_i t_i), \quad \theta_i \sim \text{IG}(\gamma, \delta), \quad t_i = \exp(\mathbf{x}_i' \beta), \tag{1}$$

where β is a vector of regression coefficients, $f(y|\theta) = \theta^{-1} \exp(-y/\theta)$, $y, \theta > 0$, i.e. the density of the exponential distribution with mean θ and $\text{IG}(\gamma, \delta)$ denotes the density of the inverse Gaussian distribution given by

$$f(z) = \frac{\delta}{\sqrt{2\pi}} \exp(\delta\gamma) z^{-3/2} \exp\left(-\frac{1}{2}\left(\frac{\delta^2}{z} + \gamma^2 z\right)\right), \quad \gamma, \delta, z > 0,$$

with mean and variance given as $E(Z) = \delta/\gamma$ and $\text{Var}(Z) = \delta/\gamma^3$ respectively. Note that there are several different parameterizations of the inverse Gaussian distribution (see, e.g. Seshadri, 1993).

It can be verified that the resulting unconditional density for y_i is given by

$$f(y_i) = \frac{\delta}{t_i} \frac{\exp(-\gamma(\phi(y_i, x_i, \beta, \delta) - \delta))}{\phi(y_i, x_i, \beta, \delta)^3} (\gamma\phi(y_i, x_i, \beta, \delta) + 1), \quad \delta, \gamma > 0, \quad y_i > 0, \tag{2}$$

where $\phi(y_i, x_i, \beta, \delta) = (\delta^2 + 2y_i/t_i)^{1/2}$, $i = 1, \ldots, n$. In the sequel for brevity we will denote $\phi_i = \phi(y_i, x_i, \beta, \delta)$ but keep in mind that ϕ_i is a function of the unknown parameters. This distribution will be referred to as the exponential-inverse Gaussian distribution (denoted as E-IG). It is easy to see that

$$F(y_i) = 1 - \delta \frac{\exp(-\gamma(\phi_i - \delta))}{\phi_i}, \quad \delta, \gamma > 0, \quad y_i > 0. \tag{3}$$

The density has always a mode at 0. The mean and the variance are given by

$$E(y_i) = \mu_i = \frac{\delta t_i}{\gamma}, \qquad \text{Var}(y_i) = t_i^2\left(\frac{2\delta}{\gamma^3} + \frac{\delta^2}{\gamma^2}\right),$$

respectively. The hazard function $h(y_i)$, assuming $t_i = 1$ and dropping subscripts, takes the form

$$h(y_i) = \frac{\gamma \phi_i + 1}{\phi_i^2},$$

which is clearly decreasing (as expected for a mixture of the exponential density).

The E-IG distribution is not well-known among actuaries. An early derivation of the distribution for reliability purposes can be found in Bhattacharya and Kumar (1986) but since then there are quite a few references of it and we are not aware of any application of it. Note, however, a misprint in that paper leading to a different distribution function. Willmot (1993) used it as a mixing density to construct mixed Poisson distributions.

Hesselager et al. (1998) refers to an exponential-inverse Gaussian model which is different than ours since it assumes an inverse Gaussian density for θ^{-1} and not θ as in the present paper. This density is also cited in Wang (1998) and Genest et al. (2002). Bivariate and multivariate versions of the E-IG distribution have been described by Whitmore and Lee (1991) and Al-Mutairi (1997).

In (1) we need to impose the constraint $\gamma = \delta$, in order the model to be estimable, i.e. we set the mean of the mixing distribution equals 1, since, otherwise, identifiability problems exist. Note that the model (1) is in fact an exponential regression model with random effects. Namely it can be written in the equivalent form

$$y_i | \theta_i t_i \sim f(y_i | \theta_i t_i), \quad \log(t_i) = \mathbf{x}_i' \beta + \theta_i,$$

where the random effects θ_i are related to the inverse Gaussian distribution since $\exp(\theta_i) \sim \text{IG}(\gamma, \delta)$ (note that this density appear in Kanefuji and Iwase (1996) under the name exponential-inverse Gaussian distribution). Under the restriction $E(\theta_i) = 1$, the overdispersion relative to the simple exponential distribution is δ^{-2}. Hence, if $\delta \to \infty$ the distribution tends to the simple exponential distribution.

Another interesting connection is the following. It can be seen that both models (the E-IG and the Pareto distribution) are special cases of the general model with a generalized inverse Gaussian mixing distribution. The probability density function of the generalized inverse Gaussian (GIG) distribution is given by

$$f(z; \lambda, \delta, \gamma) = \left(\frac{\gamma}{\delta} \right)^\lambda \frac{z^{\lambda-1}}{2K_\lambda(\delta\gamma)} \exp\left(-\frac{1}{2} \left(\frac{\delta^2}{z} + \gamma^2 z \right) \right), \quad z > 0, \ \lambda \in \mathcal{R},$$

where $\gamma, \delta \in \Theta_\lambda$, and

$$\Theta_\lambda = \begin{cases} \gamma \geq 0, \delta > 0 & \text{if } \lambda > 0, \\ \gamma > 0, \delta > 0 & \text{if } \lambda = 0, \\ \gamma > 0, \delta \geq 0 & \text{if } \lambda < 0. \end{cases}$$

We will denote this distribution as $\text{GIG}(\lambda, \gamma, \delta)$. More details about the GIG distribution can be found in Jorgensen (1982). The moments around the origin of the $\text{GIG}(\lambda, \gamma, \delta)$ are given by

$$E(z^r) = \left(\frac{\delta}{\gamma} \right)^r \frac{K_{\lambda+r}(\delta\gamma)}{K_\lambda(\delta\gamma)}, \tag{4}$$

and this formula holds for negative values of r, i.e. for inverse moments, too. The inverse Gaussian distribution is a special case of the GIG distribution for $\lambda = -1/2$, while the gamma distribution arises for $\lambda > 0$ and $\gamma = 0$.

Assuming that there are not covariates in the model, i.e. setting $t_i = 1$ then a GIG mixture of the exponential distribution has density function given by

$$f(y) = \frac{K_{\lambda-1}(\gamma\sqrt{\delta^2 + 2y})}{K_\lambda(\delta\gamma)} \frac{\gamma}{\delta^\lambda} (\delta^2 + 2y)^{(\lambda-1)/2}, \quad \gamma, \delta, y > 0, \ \lambda \in \mathcal{R}.$$

This distribution contains as special cases both the E-IG distribution and the Pareto distribution. It also contains the model of Hesselager et al. (1998) since the reciprocal inverse Gaussian distribution is also a special case of the GIG distribution.

3. Estimation for the E-IG model

Consider a sample $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ of independent observations, where y_i is the response and $\mathbf{x}_i$ a vector of covariates. Assume that the data are produced according to model (1) and hence, their density is given in (2) with the additional restriction $E(\theta_i) = 1$. Then, the log-likelihood can be written as

$$L = n \log \delta - \sum_{i=1}^{n} \log t_i - \delta \sum_{i=1}^{n} \phi_i + n\delta^2 - 3 \sum_{i=1}^{n} \log \phi_i + \sum_{i=1}^{n} \log(\delta\phi_i + 1). \tag{5}$$

Maximization of the above function with respect to the vector of parameters $w = (\delta, \beta)$ is not easy. Fortunately the mixture derivation of the model can be used for maximum likelihood (ML) estimation via an EM type algorithm. The EM algorithm is a powerful algorithm for ML estimation for data arising from mixtures, since the mixing operation can be considered as producing missing data. Thus in our case, the missing data are simply the realizations of the unobserved mixing parameter θ_i for each data point. If one augments the unobserved data θ_i to the observed data $(y_i, \mathbf{x}_i)$, for $i = 1, \ldots, n$, then the complete data log-likelihood takes the form

$$L_c = - \sum_{i=1}^{n} \log \theta_i - \sum_{i=1}^{n} \log t_i - \sum_{i=1}^{n} \frac{y_i}{\theta_i t_i} + \sum_{i=1}^{n} \log f(\theta_i|\delta),$$

where $f(\theta_i|\delta)$ denotes the mixing distribution, the inverse Gaussian with mean equal to 1 in our case. It is easy to see that the log-likelihood factorizes in two terms. Parameter δ is involved only in the second term which in fact corresponds to the likelihood of the inverse Gaussian component. Thus, one needs to obtain the sufficient statistics for the parameters of the inverse Gaussian distribution, which for our parametrization are given as $\sum \theta_i$ and $\sum \theta_i^{-1}$. On the other hand, the regression coefficients are involved only in the first part, which is equivalent to fitting an exponential GLM. Care must be paid on the fact that the observations are now of the form y_i/θ_i and thus we need to obtain the expectations of θ_i^{-1}.

A quite useful result is the following. It can be shown that the GIG distribution is conjugate distribution in the Bayesian sense for the parameter θ of the exponential distribution. The same is true for the corresponding parameter of the gamma distribution. Note that the GIG distribution has been shown to be conjugate also for the parameter of the Poisson distribution and the variance parameter of the normal distribution. This result is summarized in the following lemma.

Lemma. *Let $f(x|\theta) = \theta^{-1} \exp(-x/\theta)$ and let the prior distribution of the parameter θ be the $\mathrm{GIG}(\lambda, \gamma, \delta)$ distribution. The posterior distribution $\theta|x$ is a $\mathrm{GIG}(\lambda - 1, \gamma, \sqrt{\delta^2 + 2x})$ distribution.*

Proof. To see this, consider that the prior distribution $g(\theta)$ is the $\mathrm{GIG}(\lambda, \gamma, \delta)$ distribution. Then the posterior $g(\theta|x)$ can be obtained as

$$g(\theta|x) \propto \theta^{\lambda-1} \exp\left[-\frac{1}{2}\left(\frac{\delta^2}{\theta} + \gamma^2\theta\right)\right] \theta^{-1} \exp\left(-\frac{x}{\theta}\right),$$

and hence

$$g(\theta|x) \propto \theta^{\lambda-2} \exp\left[-\frac{1}{2}\left(\frac{\delta^2 + 2x}{\theta} + \gamma^2\theta\right)\right],$$

which is a $\mathrm{GIG}(\lambda - 1, \gamma, \sqrt{\delta^2 + 2x})$ distribution.

In the special case when $\lambda = -1/2$, i.e. when the simple inverse Gaussian is the prior distribution, the posterior distribution is a $GIG(-3/2, \gamma, \sqrt{\delta^2 + 2x})$ distribution. Thus, posterior expectations can be computed quite easily. Making use of the relationships between the modified Bessel functions of different orders (see, Abramowitz and Stegun, 1974):

$$K_{1/2}(x) = K_{-1/2}(x) = \left(\frac{\pi}{2x}\right)^{1/2} \exp(-x), \qquad K_{-n}(x) = K_n(x),$$

$$K_{n+1}(x) = \frac{2n}{x} K_n(x) + K_{n-1}(x) \quad \text{and} \quad K_{3/2}(x) = K_{1/2}(x)(1 + x^{-1}),$$

one can obtain that

$$E(\theta_i | y_i, x_i) = \frac{\phi_i^2}{\gamma \phi_i + 1}, \tag{6}$$

$$E(\theta_i^{-1} | y_i, x_i) = \frac{3}{\phi_i^2} + \frac{\gamma^2}{1 + \gamma \phi_i}. \tag{7}$$

So, an EM type algorithm for the E-IG distribution can be described as follows:

- *E-step*: Let $w = (\delta, \beta)$ be the vector of the parameters to be estimated. Given the values of the parameters after the kth iteration, say $w^{(k)}$, calculate for the ith observation the expectations $s_i = E(\theta_i | x_i, w^{(k)})$ from (6) and $m_i = E(\theta_i^{-1} | y_i, x_i, w^{(k)})$ from (7) for $i = 1, \ldots, n$.
- *M-step*: Update δ using

$$\delta^{(k+1)} = \left(\frac{n}{\sum_{i=1}^{n} s_i + \sum_{i=1}^{n} m_i - 2n}\right)^{1/2}.$$

  Update β's by fitting an exponential GLM model using as response $y_i m_i$ and the vectors $\mathbf{x}_i$ as covariate information.
- If some convergence criterion is satisfied then we stop iterating, otherwise we go back to the E-step for one more iteration.   □

Standard errors for the parameters can be obtained either through the second derivatives of the log-likelihood (see Appendix B) or through resampling techniques.

4. Application

The data were kindly provided by a major insurance company in Greece and concern an automobile insurance portfolio. The data refer to the policyholders at the end of the year 2001, and specifically they describe the amounts paid for claims at that year in hundred thousand drachmas (1 euro = 340.75 drachmas).

Our interest lies in identifying factors that affect the amount paid for each claim and specifically the factors that correspond to the insured person and his/her characteristics, including the characteristics of the car. For this purpose, the data consist of the amount paid for each policy (this was the response variable) and several characteristics of the insured person, as the age, the gender, the marital status and the bonus–malus level and characteristics of the car, like the cubic capacity, the age of the car and other characteristics related to the vehicle (like the existence of antiblocking system (ABS), the existence of extinguisher, etc.). Only policyholders with complete records, i.e. with availability of all the variables under consideration were considered. There were 2131 observations that met our criteria. Variable selection techniques were applied in order to find the variables that are considered as better predictors. The model presented below was the best fitted model. We forced some variables to be included in the model despite the fact that they are not statistically significant, like the gender and the cubic capacity. The variables
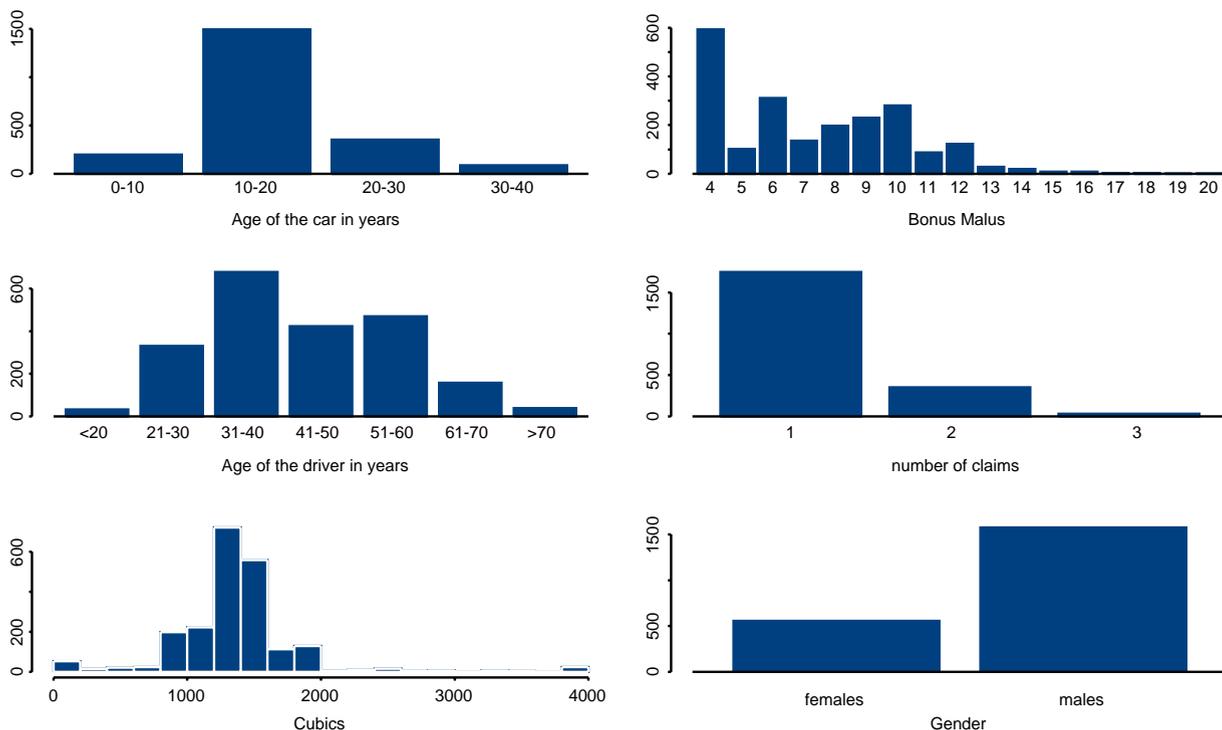
Fig. 1. Descriptive histograms for the data used in the example.

considered are similar to those used in Beirlant et al. (1991). Recall that only policyholders with at least one reported accident are used for the analysis.

In Fig. 1 one can see the histograms for the variables considered in the model to take a flavor of the range of their values. It must be pointed out that the portfolio is quite heterogeneous containing a variety of different policyholders with respect to all the combinations of the characteristics.

The derived model is presented in Table 1. The standard errors of the parameters were calculated using the asymptotic variance covariance matrix as described in Appendices A and B. For each variable we also present the t-values for the hypothesis that the associated coefficient is zero, together with the p-value of this test based on asymptotic normality. For the categorical variables we employed a likelihood ratio test, namely the values under the t-value label are the log-likelihood ratios when the variable is present or absent in the model. From the results one can see the factors that affect the size of the claim. As expected the number of claims has a positive effect, while the gender is not so important. This is perhaps due to the fact that the gender considered was that of the person that signed the contract but usually more than one persons drive the car. Bonus–malus has a positive effect. For driver age the base category is the elderly people. The most dangerous age is between 31 and 40 years, while small ages are less dangerous. For the car age the new cars were the base category. As the age of the car increases the expected amount increases, too. Unexpectedly, cars with ABS cause statistically significant larger amounts. Finally, the overdispersion parameter δ was found 1.23 indicating the deviation from the simple exponential model.

An interesting byproduct of the EM algorithm is that at the E-step we calculate for each observation the expected risk. Since the mixing parameter θ_i can be considered as a risk factor, at the E-step we obtain the expected risk for each individual. This is the posterior expected amount for a person with the specific characteristics and, hence, it can be used to predict the future amount that the individual will cost to the company. This value represents the expected claims of the individual and hence this quantity can be helpful for pricing and bonus–malus purposes.

Table 1
Results of the fitted E-IG regression model

Variable	Estimate	S.E.	t-Value/Wald	P-value
Constant	−5.2826	0.1986	−26.5971	0.0000
Bonus–Malus	0.0446	0.0073	6.1509	0.0000
Driver age			45.832	0.0000
>70 years	0	–		
61–70 years	−0.2224	0.1631	−1.3642	0.1725
51–60 years	−0.0683	0.1513	−0.4512	0.6518
41–50 years	−0.1896	0.1522	−1.2458	0.2128
31–40 years	0.1405	0.1499	0.9373	0.3486
21–30 years	−0.3929	0.1557	−2.5232	0.0116
<20 years	−0.5757	0.2223	−2.5900	0.0096
Car age			21.716	0.0000
<10 years	0	–		
11–20 years	−0.4008	0.1116	−3.5907	0.0003
21–30 years	−0.2192	0.1038	−2.1112	0.0348
31–40 years	−0.6121	0.1226	−4.9913	0.0000
Number of claims	2.1084	0.0494	42.6870	0.0000
ABS				
No ABS	0	–		
With ABS	0.3196	0.0605	5.2835	0.0000
Gender				
Female	0	–	–	–
Male	0.0544	0.0465	1.1702	0.2419
Cubic capacity (in 1000 cc)	0.0048	0.0034	1.4317	0.1522
δ	1.2382	0.0431	28.6966	0.0000

In Fig. 2 one can see the expected amounts for several categories. In fact we have plotted the expected amount, obtained after the last iteration of the EM algorithm, for several variables used for creating the predicting equation. These quantities reflect the expected amount for each policyholder given his/her characteristics and is based on the derived model. The plots are very informative and allow for easy interpretation of the results. These expectations can be used for further research, as for example for deriving bonus–malus models based on the expected amount of claims instead of merely using the expected number of claims (see, Frangos and Vrontos, 2001).

5. Diagnostics

5.1. Goodness of fit

It is easy to see from the definition of the model in (1) that the quantities $r_i = y_i/t_i$ are free from the covariates and given the restriction $\gamma = \delta$ they follow a special case of the E-IG model with density

$$f(r_i) = \delta \frac{\exp(-\delta\phi_i + \delta^2)}{\phi_i^3} (\delta\phi_i + 1),$$

where now $\phi_i = \sqrt{\delta^2 + 2r_i}$. Thus a QQ plot of the observed and the theoretical values of the r_i's can be a diagnostic about the adequacy of the model. Such a plot can be seen in Fig. 3 together with the similar plot of the
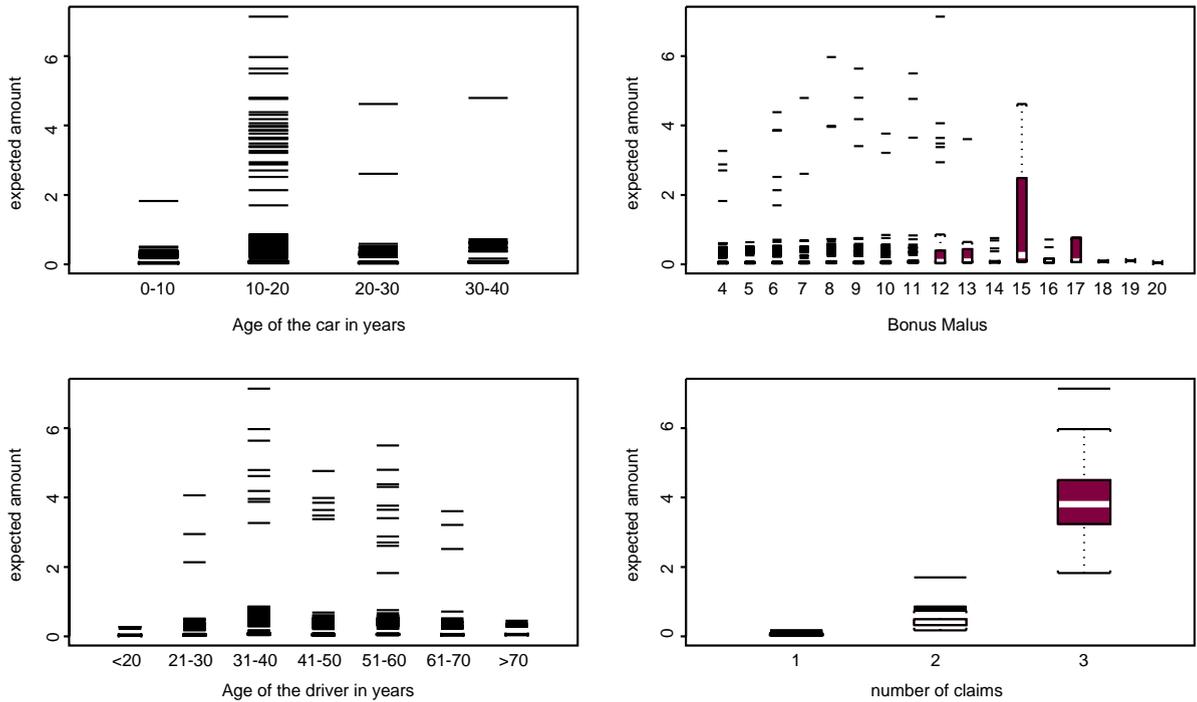
Fig. 2. Expected amount for a claim with respect several of the explanatory variables.

simple exponential density model. The E-IG model improves very much with respect to the exponential model (the likelihood ratio statistic was 123.43) and the fit is satisfactory. Especially at the right tail of the density the PP plot is almost a straight line indicating a good fit in the tail.

5.2. Right tail index

Another check of the model was based on the right index tail proposed by Wang (1998). This index was proposed in order to describe the tail behavior of different densities and it can discriminate between competing models. The right tail index is defined as

$$d(X) = \frac{\int_0^\infty \sqrt{S_X(t)}\,dt}{E(X)} - 1,$$

where $S_X(t) = P(X \geq t)$.

Empirical estimation for this index is discussed in Jones and Zitikis (2003). An estimate of the right tail index is obtained as

$$\hat{d}(X) = \sum_{i=1}^{n} c_i X_{(i)} / \bar{x},$$

where $X_{(i)}$ is the ith ordered observation and

$$c_i = \left(\frac{n-i}{n}\right)^{1/2} - \left(\frac{n-i+1}{n}\right)^{1/2} - \frac{1}{n}, \quad i = 1, \ldots, n.$$

PP-plot for the Exp model   PP-plot for the E-IG model
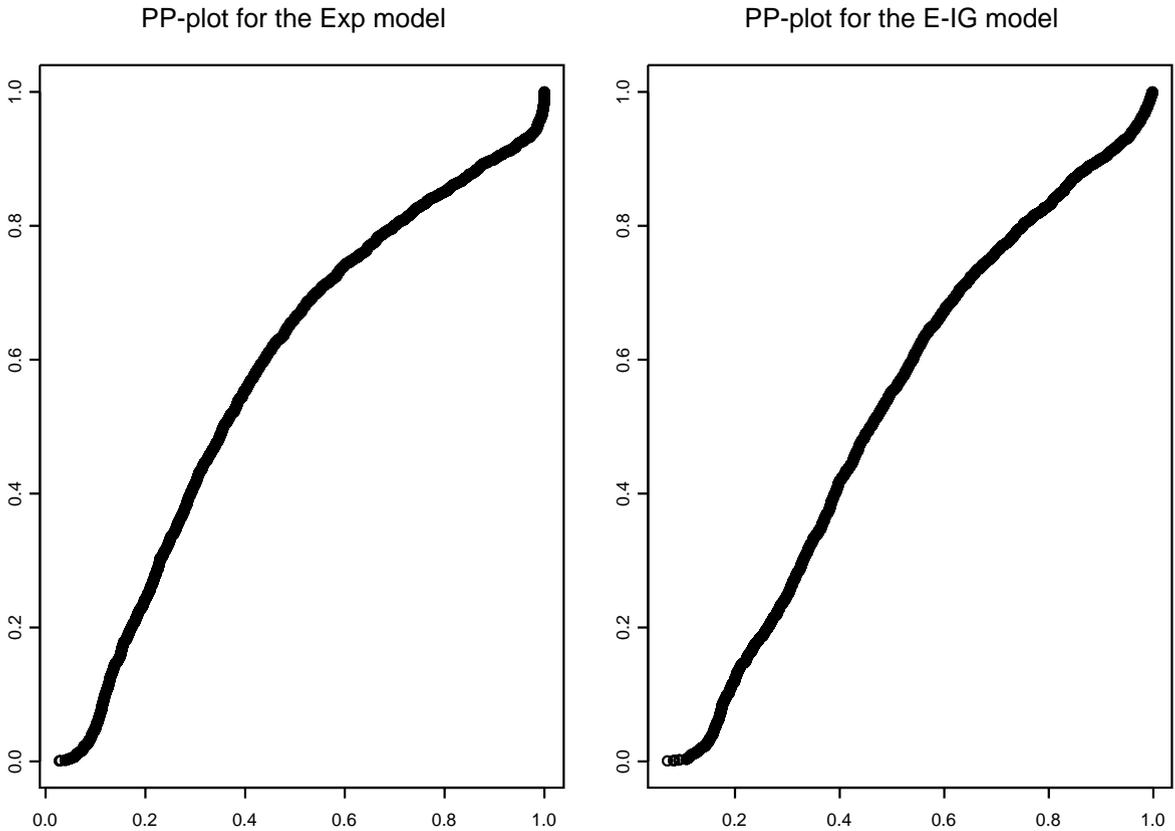
Fig. 3. PP plots for the exponential model and the E-IG model. Clearly the E-IG model improves considerably with respect to the simple exponential model. Note also the good fit of the E-IG model at the tails of the data.

According to this approach the tail index for r_i's was calculated to be 1.740, while the theoretical value is 1.93. However for small sample sizes the empirical counterpart underestimates this index. For this reason we built parametric bootstrap confidence intervals for the empirical right tail index. The 95% bootstrap confidence intervals for the index is given by $(1.65, 2.13)$ based on 50 000 bootstrap replications. The interval contains the value we obtained for our data. This is an indication that our model is adequate for modelling the tail of the data. On the contrary, Pareto distribution with the same first two moments has an index which is much larger (see Fig. 4). In Fig. 4 we have plotted the right tail index for the Pareto and E-IG distributions with mean equal to 1 and varying variance. There is a clear difference between the two densities.

5.3. Residuals

Another useful approach is to consider the common Pearson type residuals for assessing the goodness of fit of the proposed model. It is easy to see that for our model, the variance of each observation is given as a function of the mean t_i by $V(t_i) = t_i^2(1 + 2/\delta^2)$. The quantity in the parenthesis, is in fact, an overdispersion factor relative to the simple exponential model. Thus we can construct Pearson type residuals as

$$R_i = \frac{\delta(y_i - t_i)}{t_i\sqrt{(2 + \delta^2)}}.$$
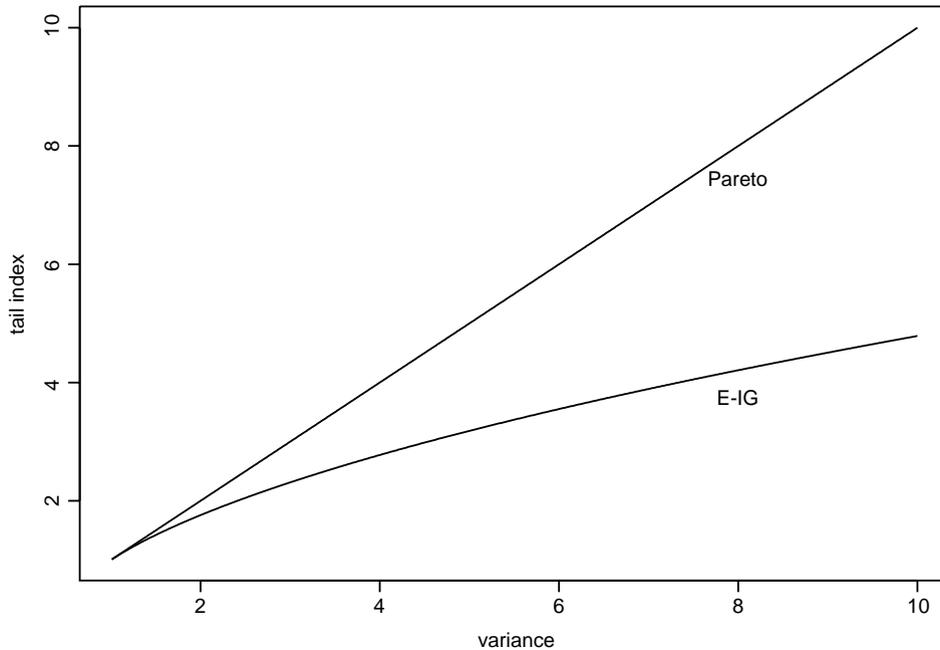
Fig. 4. Plot of the right tail index as a function of the variance for the E-IG and the Pareto models. The mean equals 1 for all the cases.

In the usual way we can use as a diagnostic the sum of squared Pearson's residuals. This sum equals 1828.43 which is much lower than the number of observations and thus it implies a satisfactory goodness of fit of the model. In Fig. 5 one can see the absolute residuals. There are few of them with large values. Detailed examination of the residuals with respect the explanatory variables did not show any particular pattern.

Other type of residuals can be also used as diagnostics, like deviance residuals, etc.

5.4. Influence diagnostics

In order to assess the influence of the observations for the estimated parameters we made use of a statistic similar to the Cook distance for simple linear models. Our statistic has the form

$$D_i = (w_{(i)} - w)' \hat{S}_w^{-1} (w_{(i)} - w),$$

where w is the vector of all the parameters of the model, i.e. $w = (\delta, \beta)$ and $w_{(i)}$ is the same vector when removing the ith observation. Moreover $\hat{S}_w^{-1}$ is the estimated covariance matrix for w obtained when considering all the data. The statistic mimics the Cook distance and differs form that in McCullagh and Nelder (1989) since we use directly the estimated covariance matrix. Plot of D_i's may reveal observations with high influence for the parameters. We advocate not to use a limit above which the observation is considered as influential but mainly to check from the plot comparatively the influence of each observation

Fig. 6 presents the D_i's for our data set. We can see that there are some observations that have larger statistics. However, the values of the statistics do not indicate large departures but rather they indicate that those points have an effect in the estimates. Examinations of the observations revealed that they have rather extreme values for some explanatory variables (e.g. observation 655 refers to a car with large cubic capacity).
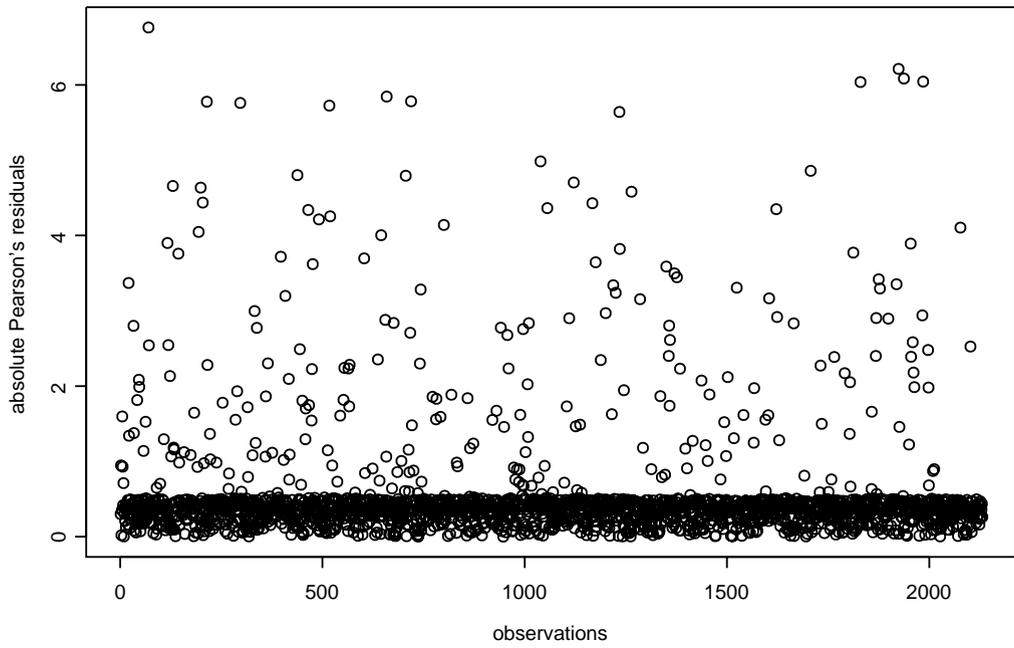
Fig. 5. Absolute Pearson's residuals for our dataset. There are few observations with high residual.
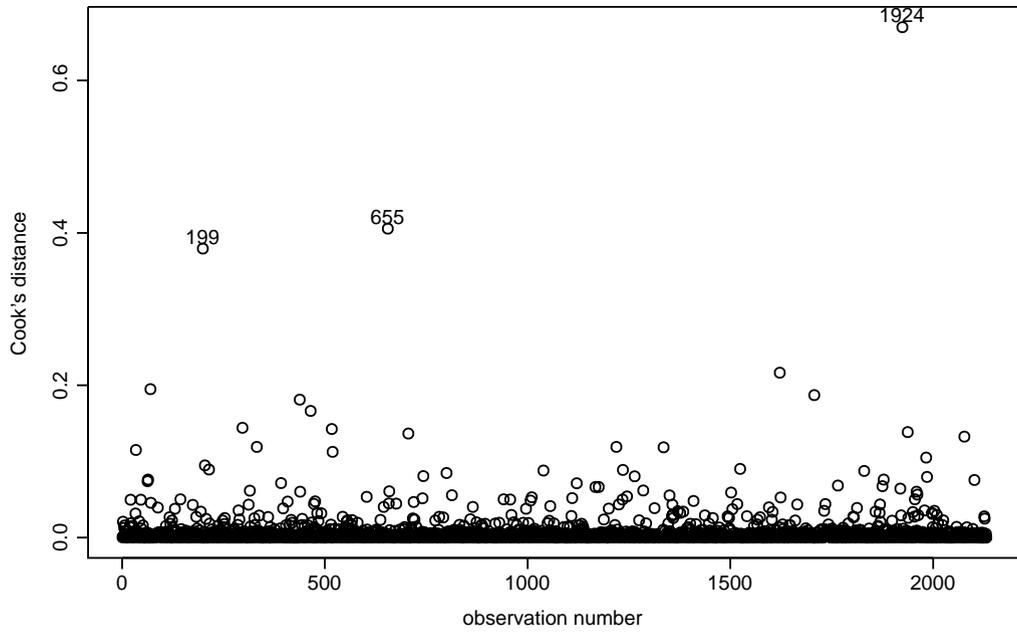
Fig. 6. Cook distances for the data used in the example.

Note that, computationally, the calculation of the statistics is rather easy via the EM algorithm since we have quite close initial values and usually the algorithm converges after a few iterations. More refined influence diagnostic can be also calculated but we do not pursue further about this in the present paper.

6. Other topics

6.1. Computational aspects

The EM algorithm described in Section 3 was used to estimate the parameters. As initial values we used the regression parameters of a simple exponential distribution GLM. The initial value for the parameter δ was set equal to the overdispersion of the exponential model. The model converged after quite a few iterations using a rather strict criterion. We stopped iterating when the relative change of the log-likelihood between two successive iterations was smaller than 10^{-12}. To ensure that the global maximum has been obtained and the algorithm had not been trapped in a local maximum we ran the algorithm with random initial values but for all the cases the same solution was obtained.

Standard errors were obtained as described in Appendices A and B. Alternatively one could use the property that the likelihood ratio and the Wald tests for the significance of an individual parameter are equivalent for large samples and, thus, the log-likelihood change on omitting the variable can be related to the t-statistic and hence, its standard error (Aitkin, 1999). Note that reduced models can be fitted easily and quickly by the proposed algorithm, since very good initial values are available from the full model. For the same reason jackknife or bootstrap standard errors can be obtained relatively easy.

6.2. An extended model

The present model fitted the claim amount for all the policyholders with at least one claim at the period of study. This implies that there are not 0 claims at the database used for fitting the model. A more elaborated model could consist of two parts, the first part trying to model those without claim for the period of study while the other part modelling those with claims. Let c_i be the number of claims for the ith individual and z_i be an indicator variable that takes the value 1 if $c_i > 0$ and 0 if $c_i = 0$. Then the distribution of the amount paid for an individual is given as

$$f(y_i) = f(y_i|z_i = 0)P(z_i = 0) + f(y_i|z_i = 1)P(z_i = 1), \tag{8}$$

but since $f(y_i|z_i = 0)$ is a degenerate distribution at 0 (since if the individual has no claim then the amount paid is necessarily 0) and thus the model consists of two parts. The former fits a logistic regression model to the characteristics of the policyholder in order to decide whether the policyholder will have a claim, and the latter given that there is at least one claim tries to model the claim amount. Our model is applicable for this part.

The above model can be seen as a model with exact zeros and is useful for the calculation of the aggregate claim size of the entire portfolio. Being, in fact, a finite mixture model, one can fit the general model without special effort using standard techniques for finite mixtures (see, Bohning, 1999).

7. Concluding remarks

Mixtures of the exponential distribution are very important models for describing survival data. The reason is that all the distributions with decreasing failure rate can be retrieved as mixtures of the exponential distribution (see, e.g. Proschan, 1963; Barlow and Proschan, 1975). Some known distributions belong to this family like the Pareto distribution, via a gamma mixing density as well as the gamma distribution for shape parameter <1, as shown in Gleser (1989).

In the present paper another member of this family, the exponential-inverse Gaussian distribution was proposed. The model has moderate tails, ML estimation is relatively easy via an EM type algorithm and the model can be seen as a mixed effects exponential regression model.

The EM algorithm described is easily extended to cover other cases. For example, the algorithm is quite general and it may be used for other mixing distributions. Such an example is the Pareto distribution. A similar algorithm for mixed effects Poisson regression has been proposed in Karlis (2002). In addition, the algorithm can be easily extended to cover the case of multivariate exponential-inverse Gaussian distributions, like those described in Whitmore and Lee (1991). Furthermore, the data augmentation used in the paper to derive the EM algorithm can be the basis for constructing Bayesian estimation procedures through an MCMC scheme.

Finally, the exponential-generalized inverse Gaussian model defined at the end of Section 2 contains both the Pareto and the E-IG models as subcases. Thus, it would be interesting to fit the more general model in order to select between those two and other models included in this general model. However the estimation task is not any more simple. Creating a similar EM algorithm is not simple as the M-step involves ML estimation for the parameters of the GIG distribution and their behavior is not simple (see Jorgensen, 1992).

Acknowledgements

Appendix A

Suppose that $x|\theta$ follows an exponential distribution and that θ follows an GIG$(\lambda, \delta, \gamma)$ density given in (2). Then the resulting mixed distribution has probability density function

$$
\begin{aligned}
f(x) &= \left(\frac{\gamma}{\delta}\right)^\lambda \frac{1}{2K_\lambda(\delta\gamma)} \int_0^\infty \theta^{-1} \exp\left(-\frac{x}{\theta}\right) \theta^{\lambda-1} \exp\left(-\frac{1}{2}\left(\frac{\delta^2}{\theta} + \gamma^2\theta\right)\right) \, \mathrm{d}\theta \\
&= \left(\frac{\gamma}{\delta}\right)^\lambda \frac{1}{2K_\lambda(\delta\gamma)} \int_0^\infty \theta^{\lambda-2} \exp\left(-\frac{1}{2}\left(\frac{\delta^2 + x}{\theta} + \gamma^2\theta\right)\right) \, \mathrm{d}\theta.
\end{aligned}
$$

Using the integral representation of the modified Bessel function as verified from the density of the GIG distribution one obtains that

$$
f(x) = \frac{K_{\lambda-1}(\gamma\sqrt{\delta^2 + 2x})}{K_\lambda(\delta\gamma)} \frac{\gamma}{\delta^\lambda} (\delta^2 + 2x)^{(\lambda-1)/2}, \quad \gamma, \delta, x > 0, \quad \lambda \in \mathcal{R}.
$$

For $\lambda = -1/2$ and the properties of the modified Bessel function one can obtain the density of the E-IG distribution given in (2).

Appendix B

Asymptotic standard errors can be obtained by standard procedures based on the information matrix. Let w denote the vector of the parameters for the case of p covariates including the constant, namely $w = (\delta, \beta_1, \ldots, \beta_p)$.

Then the matrix $\mathbf{H} = h_{ij}$, $i, j = 1, \ldots, p + 1$ of the second derivatives of the log-likelihood with respect to the parameter vector w is given as

$$h_{1,1} = -\frac{n}{\delta^2} - 3\delta \sum_{i=1}^{n} \phi_i^{-1} + \delta^3 \sum_{i=1}^{n} \phi_i^{-3} + 2n + 6\delta^2 \sum_{i=1}^{n} \phi_i^{-4} - 3 \sum_{i=1}^{n} \phi_i^{-2}$$

$$+ \sum_{i=1}^{n} \frac{3\delta/\phi_i - \delta^3/\phi_i^3}{\delta\phi_i + 1} - \sum_{i=1}^{n} \frac{(\phi_i + \delta^2/\phi_i)^2}{(\delta\phi_i + 1)^2},$$

$$h_{1,j+1} = \sum_{i=1}^{n} \frac{y_i x_{ji}}{\phi_i t_i} - \delta^2 \sum_{i=1}^{n} \frac{x_{ji}}{\phi_i^3} \frac{y_i}{t_i} - 6\delta \sum_{i=1}^{n} \frac{x_{ji} y_i}{\phi_i^4 t_i} + \sum_{i=1}^{n} \frac{-x_{ji} y_i/\phi_i t_i + \delta^2 (y_i x_{ji}/\phi_i^3 t_i)}{\delta\phi_i + 1}$$

$$+ \delta \sum_{i=1}^{n} \frac{\phi_i + \delta^2/\phi_i}{(\delta\phi_i + 1)^2} \frac{x_{ji} y_i}{\phi_i t_i}, \quad j = 1, \dots, p,$$

$$h_{j+1,k+1} = \delta \sum_{i=1}^{n} \frac{y_i^2 x_{ji} x_{ki}}{\phi_i^3 t_i^2} - \delta \sum_{i=1}^{n} \frac{x_{ji} x_{ki} y_i}{\phi_i t_i} + 6 \sum_{i=1}^{n} \frac{x_{ji} x_{ki} y_i^2}{\phi_i^4 t_i^2} - 3 \sum_{i=1}^{n} \frac{x_{ji} x_{ki} y_i}{\phi_i^2 t_i} - \delta \sum_{i=1}^{n} \frac{x_{ki}}{\phi_i^3} \frac{y_i^2}{t_i^2} \frac{x_{ji}}{\delta\phi_i + 1}$$

$$+ \delta \sum_{i=1}^{n} \frac{x_{ki}}{\phi_i} \frac{y_i}{t_i} \frac{x_{ji}}{\delta\phi_i + 1} - \delta^2 \sum_{i=1}^{n} \frac{x_{ki}}{\phi_i^2} \frac{y_i^2}{t_i^2} \frac{x_{ji}}{(\delta\phi_i + 1)^2}, \quad j, k = 1, \dots, p.$$

The entries of the information matrix $\mathbf{I} = I_{ij}$ are given as $I_{ij} = E(h_{ij})$, $i, j, = 1, \dots, p + 1$. Since the expectations are hard to be evaluated, one may use asymptotic Taylor expansions keeping the first few terms.

Confidence intervals and tests about the parameters can be obtained by using familiar asymptotic χ^2 approximations for likelihood ratio statistics. Alternatively one may use asymptotic normality, treating μ as approximately normally distributed with mean vector $\mathbf{0}$ and covariance matrix given by $\mathbf{I}^{-1}$.

References

Abramowitz, M., Stegun, I.A., 1974. Handbook of Mathematical Functions. Dover, New York.

Aitkin, M., 1999. A general maximum likelihood analysis of variance components in generalized linear models. Biometrics 55, 117–128.

Al-Mutairi, D.K., 1997. Properties of an inverse Gaussian mixture of bivariate exponential distribution and its generalization. Statistics and Probability Letters 33, 359–365.

Apostolakis, N., 1998. Tariff classification in auto insurance. M.Sc. Thesis. Department of Statistics, Athens University of Economics. ISBN: 960-7929-00-4.

Barlow, R.E., Proschan, F., 1975. Statistical Theory of Reliability and Life Testing: Probability Models. Holt, Rinehart and Winston, New York.

Beirlant, J., Derveaux, V., De Meyer, A.M., Goovaerts, M.J., Labie, E., Maenhoudt, B., 1991. Statistical risk evaluation applied to (Belgian) car insurance. Insurance: Mathematics and Economics 10, 289–302.

Beirlant, J., Goegebeur, Y., Verlaak, R., Vynckier, P., 1998. Burr regression and portfolio segmentation. Insurance: Mathematics and Economics 23, 231–250.

Bhattacharya, S.K., Kumar, S., 1986. E-IG model in life testing. Calcutta Statistical Association Bulletin 35, 85–90.

Bohning, D., 1999. Computer Assisted Analysis of Mixtures and Applications in Meta-analysis, Disease Mapping and Others. CRC Press, New York.

Frangos, N.E., Vrontos, S.D., 2001. Design of optimal bonus–malus systems with a frequency and a severity component on an individual basis in automobile insurance. ASTIN Bulletin 31, 1–22.

Genest, C., Marceau, E., Mesfioui, M., 2002. Upper stop-loss bounds for sums of possibly dependent risks with given means and variances. Statistics and Probability Letters 57, 33–41.

Gleser, L.J., 1989. The gamma distribution as a mixture of exponential distributions. American Statistician 43, 115–117.

Haberman, S., Renshaw, A.E., 1996. Generalized linear models and the actuarial science. The Statistician 45, 407–436.

Hesselager, O., Wang, S., Willmot, G.E., 1998. Exponential and scale mixtures and equilibrium distributions. Scandinavian Actuarial Journal, 125–142.

Jones, B.L., Zitikis, R., 2003. Empirical estimation of risk measures and related quantities. North American Actuarial Journal 7, 44–54.

Jorgensen, B., 1982. The Generalized Inverse-Gaussian Distribution, Lecture Notes in Statistics, vol. 9. Springer-Verlag.

Kanefuji, K., Iwase, K., 1996. Exponential inverse Gaussian distribution. Computational Statistics 11, 315–326.

Karlis, D., 2002. Exact ML estimation for mixed Poisson regression models. Statistical Modelling: An International Journal 1, 305–318.

Klugman, S.A., Panjer, H.H., Willmot, G.E., 1998. Loss Models: From Data to Decisions. Wiley, New York.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed. Chapman & Hall, New York.

Proschan, F., 1963. Theoretical explanation of observed decreasing failure rate. Technometrics 5, 375–383.

Seshardi, V., 1993. The Inverse Gaussian Distribution. Oxford Science Publications.

Wang, S.S., 1998. An actuarial index of right-tail risk. North American Actuarial Journal 2, 88–101.

Whitmore, G.A., Lee, M.L.T., 1991. A multivariate survival distribution generated by an inverse Gaussian mixture of exponentials. Technometrics 33, 39–50.

Willmot, G.E., 1993. On recursive evaluation of mixed Poisson probabilities and related quantities. Scandinavian Actuarial Journal, 114-133.