IBNR Reserve Under a Loglinear Location-Scale Regression Model

by Louis Doray

IBNR RESERVE UNDER A LOGLINEAR LOCATION-SCALE REGRESSION MODEL [1]

Abstract

In this paper, we develop models for known claims, when the data are grouped into the usual triangle and the goal is to predict IBNR claims. We assume that the payment for a certain accident and development year is composed of a deterministic part and a multiplicative random error. We use a loglinear location-scale regression model for the amount of claims. The parameters are estimated by maximum likelihood methods, so that their asymptotic properties are well known. The regression model presents many advantages over the chain ladder method: it has fewer parameters, and does not underestimate the reserve. Moreover, it will be possible with a simulation to establish a reserve with a certain level of confidence (for example 80%).

The logarithm of the error is assumed to follow certain known distributions (normal, extreme value, generalized loggamma, logistic and log inverse gaussian). We derive certain theoretical properties of these distributions and prove that the MLE's of the regression and scale parameters exist and are unique, when the error has a log-concave density.

In conclusion, we advocate the use of regression models over the chain ladder method, since they take into account both the error involved in the estimation of the parameters and the statistical error inherent in the prediction of future claims, the fit of the model can be tested statistically and confidence intervals for the reserve can be derived.

Keywords: Chain-ladder method; Weibull-extreme value regression; maximum likelihood; prediction; generalized loggamma; logistic; inverse gaussian; consistency.

[1] The author gratefully acknowledges the financial support of the CAFIR research fund and of the Natural Sciences and Engineering Research Council of Canada.

1 Introduction

1.1 IBNR claims

All insurance companies registered to do business in Canada are required by the regulatory authorities to set up reserves for claims which have been incurred but have not yet been reported as of their financial statement date, usually December 31. In determining the liabilities of the insurance company, the valuation actuary must also estimate the liabilities generated by claims incurred but not enough reserved (IBNER), (also called reported but not settled (RBNS)).

The distinction between these two parts of the loss reserve, the IBNR part and the IBNER part, is not always made in practice, especially when the data are aggregated. In this paper, by IBNR reserve, we will refer to both types of claims.

The primary purpose of those reserves is to ensure the protection of the policyholders: when the insurance company is notified of these claims, it will have the reserves, backed by sufficient assets, to pay those claims.

The delay in reporting the claim may depend on the type of claim (for example, asbestosis may take more than 10 years to manifest itself in a worker). The long delay observed in the settlement of certain claims is sometimes due to the fact that some of them are resisted by the insurance company, putting into motion a long judiciary process. In other cases, there will be a long delay before the ultimate cost of a claim can be determined exactly (in workers' compensation for example, the insurance company will have to wait for an annuity to terminate).

The 1987 Loss Development Study, undertaken by the Reinsurance Association of America, compares the development of losses for various lines of business. Automobile liability was the line where the claims got developed the fastest, while Workers' Compensation was slower to develop. General liability, excluding asbestos claims, had a development pattern similar to Workers' Compensation, but a little bit slower initially. Medical malpractice experienced the slowest development among those lines of business.

Due to this long reporting and settlement lag, it will be extremely important for the valuation actuary to develop adequate statistical models to project known losses to ultimate losses.

1.2 The chain ladder method and its deficiencies

By grouping the claims by accident year (year in which the accident giving rise to the claim occurred) and development year (number of years elapsed since this accident year), the data can be presented in a trapezoidal array.

In this paper, to illustrate the various models proposed, we will use the data in table 1 (taken from CIA Proceedings, Volume 20 no 1, p.183), which represents the liability claims in thousands of dollars incurred by a Canadian insurance company over the ten-year period 1978-1987. We will do the analysis with the incremental claims (in table 2), obtained by differencing successive cumulative amounts.

The problem of estimating IBNR claims consists in predicting, for each accident year,

Table 1: Claims Incurred

Accident year	Development year					
	1	2	3	4	5	6
1978	8489	9785	10709	11289	11535	11661
1979	12970	14766	16201	17060	17714	17979
1980	17522	20305	21774	22797	23220	23872
1981	21754	24338	25501	26284	27171	27526
1982	19208	21549	22769	23388	24229	24932
1983	19604	22073	23296	24543	25155	
1984	21922	24233	25374	26882		
1985	25038	28401	30545			
1986	32532	37006				
1987	39862					

the ultimate amount of claims incurred. The amount paid by the insurance company for those claims is then subtracted, leaving the reserve the insurer should hold for future payments. To calculate the reserve, all methods or models usually assume that the pattern of cumulative or incremental claims incurred or paid is stable across the development years, for each accident year. Since for the last accident year, only one amount will be available, the reserve will be highly sensitive to this amount. Moreover, because of growth experienced by the company, it will be bigger than any other amount in the data set, hence the importance of verifying that the development pattern of the claims has not changed over the years.

One of the earliest methods, and now most commonly used in the actuarial profession, is the chain ladder method. Assuming that for each accident year, the development pattern remains stable, development factors are calculated by dividing cumulative paid or incurred

Table 2: Incremental claims incurred

Accident year	Development year 1	2	3	4	5	6
1978	8489	1296	924	580	246	126
1979	12970	1796	1435	859	654	265
1980	17522	2783	1469	1023	423	652
1981	21754	2584	1163	783	887	355
1982	19208	2341	1220	619	841	703
1983	19604	2469	1223	1247	612	
1984	21922	2311	1141	1508		
1985	25038	3363	2144			
1986	32532	4474				
1987	39862					

claims after j periods since incurral by the cumulative amount after $j-1$ periods. These factors can be weighted by the amount each year. The year-to-year development factors are then applied to the most recent amount for each accident year, i.e. the amounts on the right-most diagonal.

Using the weighted approach with the cumulative claims of table 1, we obtain the development factors of table 3. Projecting the claims incurred to ultimate amounts with those development factors, we obtain a reserve estimate of 23,919.

Table 3: Loss Development Factors

Year	Development factors
1-2	1.13079
2-3	1.06479
3-4	1.04545
4-5	1.02922
5-6	1.02023

Many variations have been presented for the basic chain ladder method just introduced; a linear trend or an exponential growth may be assumed to be present among the development factors. Instead of taking their weighted average, they would be extrapolated into the future. The chain ladder method can also be adjusted for inflation.

However, the chain ladder method suffers from the following deficiencies:

1- it implicitly assumes too many parameters (one for each column).

2- it does not give any idea of the variability of the reserve estimate, or a confidence interval for the reserve.

3- as will be shown in section 2, it is negatively biased, which could lead to serious underreserving, a threat to the insurer's solvency.

We will therefore develop a stochastic model, which involves only 5 parameters. With this model, we will be able to calculate an amount such that there is an 80% probability that the reserve will be sufficient to cover the liabilities generated by the current block of business.

1.3 The general model

In this paper, we will consider loglinear location-scale regression models of the form

$$Z_i = \ln Y_i = X_i\beta + \sigma\epsilon_i, \quad Y_i > 0$$

where	Y_i	is the ith element of vector Y (the data), of dimension n,
	X	is the regression matrix, whose first column contains 1's,
		and whose ith row is the vector denoted by X_i
		and (i,j) element is denoted X_{ij},
	β	is the vector of unknown parameters to be estimated,
		of dimension p,
	$X_i\beta$	is the location parameter for Z_i,
	σ	is the scale parameter,
and	ϵ_i	is a random error with known density $f(\epsilon)$.

The loglinear location-scale model has been used extensively in reliability theory and in survival analysis (see for example, Kalbfleisch and Prentice (1980), Lawless (1982), Cohen and Whitten (1988), Bain and Engelhardt (1991)). It is easily shown that the random variable Z_i will have density

$$\frac{1}{\sigma}f(\frac{z_i - X_i\beta}{\sigma}), \quad -\infty < z_i < \infty.$$

As in Zenwirth (1990), for the location parameter, we will use $\alpha + \beta \ln j + \gamma j + \iota(i+j-2)$, where i is the accident year and j, the development year. Taylor (1986) cautions not to use cumulative claims amounts, but incremental claims in the analysis; otherwise, the estimates obtained would be biased, because of the non-independance of the cumulative amounts.

We will assume that $Y_i > 0$. To model negative values of Y_i, Cohen and Whitten (1988) use modified moment estimators and Cohen (1988), local maximum likelihood methods.

1.4 Outline of the paper

Section 2 considers the lognormal linear regression model and presents the results of a simulation study showing that the chain ladder estimate of the reserve is negatively biased. Other choices possible for the distribution of the random error are the extreme value distribution, leading to the Weibull-extreme value regression model (section 3), the generalized loggamma (section 4), the logistic (section 5), and the log inverse gaussian distribution (section 6). We derive certain theoretical properties of these distributions, such as their moment generating function and moments. We show how the actuary can establish a reserve with a certain level of confidence (for example 80%), with a simulation.

In section 7, we show that the MLE's of the regression and scale parameters exist and are unique when the error ϵ in the loglinear location-scale regression model has a log-concave density. Under misspecification of the error distribution in a linear location-scale model, the MLE's of the regression parameters are shown to be consistent (section 8), while we present a sufficient condition for the consistency of the MLE of the scale parameter, when the postulated model has lognormal errors. Finally, we present some remarks.

2 Lognormal linear regression model

When it is assumed that ϵ_i are independent and identically distributed $N(0,1)$ random variables, we obtain the lognormal linear regression model. Doray (1992) has studied

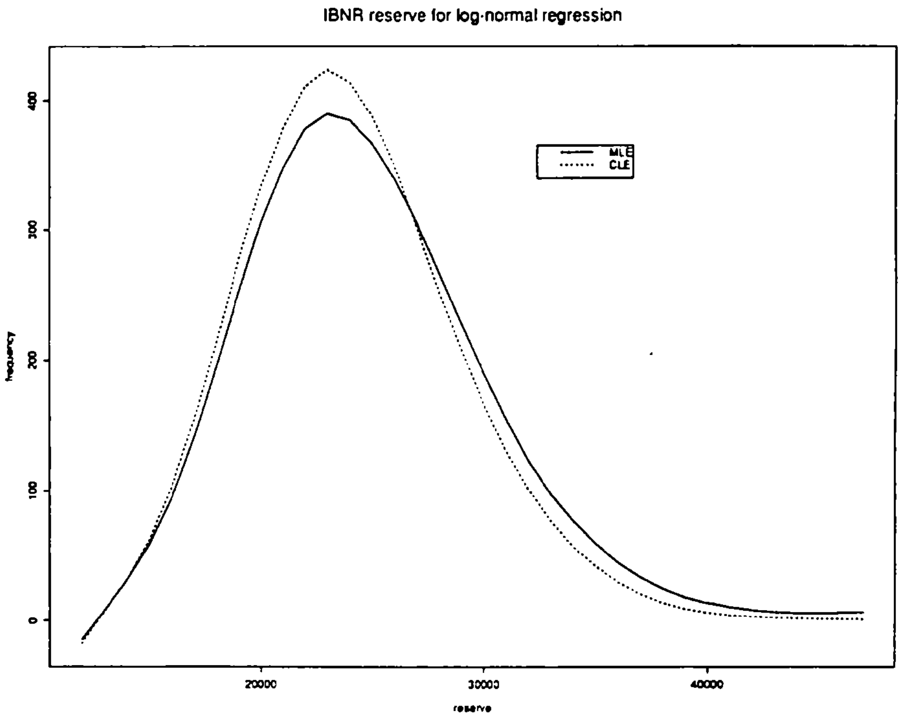
Table 4: Frequency distribution of the IBNR reserve under the normal error assumption

Amount	MLE	CLE	Amount	MLE	CLE
< 13000	0	0	30000-31000	165	152
13000-14000	4	2	31000-32000	150	126
14000-15000	12	11	32000-33000	103	80
15000-16000	33	30	33000-34000	96	68
16000-17000	62	72	34000-35000	76	47
17000-18000	126	131	35000-36000	50	40
18000-19000	191	199	36000-37000	36	26
19000-20000	253	301	37000-38000	28	16
20000-21000	323	376	38000-39000	20	5
21000-22000	372	391	39000-40000	14	2
22000-23000	449	441	40000-41000	13	10
23000-24000	449	498	41000-42000	8	2
24000-25000	393	443	42000-43000	7	3
25000-26000	366	436	43000-44000	7	0
26000-27000	342	375	44000-45000	2	2
27000-28000	334	274	45000-46000	2	1
28000-29000	285	231	46000-47000	6	0
29000-30000	214	207	$\geq$ 47000	9	2

extensively this model, taking into account the estimation error on the parameters and the statistical prediction error in the model. He has derived various estimators for the IBNR reserve, among them the maximum likelihood estimator and the uniformly minimum variance unbiased estimator (UMVUE), as well as an expression for the variance of the latter estimator. The variance of the IBNR reserve is also calculated. The joint distribution of the amounts in each cell of the lower triangle is shown to follow a multivariate lognormal (MLN) distribution.

To compare the traditional chain ladder estimator of the reserve with the MLE, a simulation was performed, assuming the model $\ln Y_{ij} = \alpha_i + \beta_j + \epsilon_{ij}$ is the true model.

Figure 1


IBNR reserve for log-normal regression
MLE
CLE
400
300
200
100
0
frequency
20000
30000
40000
reserve

Five thousand sets of realizations of Y_{ij} in the trapezium were randomly generated, where each Y_{ij} is independent $LN(\hat{\alpha}_i + \hat{\beta}_j, \tilde{\sigma^2})$, where the values of $\hat{\beta}$ and $\tilde{\sigma^2}$ are the MLE's of the parameters. For each set, we calculated the chain ladder estimate (CLE) and the MLE of the predicted value of IBNR claims using the multivariate lognormal distribution (see appendix 10.1 for the algorithm used for the simulation). The results of the simulation are summarized in table 4 and figure 1. We see from those results that the reserve has a distribution skewed to the right, which comes from the lognormal assumption. The reason why the chain ladder estimate, generally used by actuaries to determine insurance company reserves, underestimates the expected liability, is that it does not capture this long-tail behaviour, as is apparent from table 4.

The MLE of the reserve gives 25,262, while the CLE gives 23,919. The reserve for IBNR claims the insurance company will hold could be set at, for example, the 80-th percentile of the predicted distribution of IBNR claims, that is at 29,019 in our example. The actuary could then state, that in his or her opinion, there is an 80% probability that the reserve will be sufficient to meet the liabilities of the current block of business.

Asymptotically (i.e. as the upper trapezium of data gets larger), the various variables to be predicted will become independent, and from that perspective, we can consider an asymptotic confidence interval for the reserve, using the central limit theorem. The lower bound for the 80% asymptotic confidence interval of the reserve is 29,514, which can be compared with the amount of 29,019 obtained in the simulation.

A provision for adverse deviation could also be defined as equal to the 80-th percentile of the predicted distribution of IBNR claims minus the UMVUE of the reserve (24,403). This gives 4616 as the PAD for the claims of section 1.2.

3 Weibull-extreme value regression model

In this section, we examine the Weibull-extreme value regression model. Let us assume that ϵ follows a standard type I extreme value (or Gumbel) distribution with

probability density function (pdf)	$f(\epsilon) = \exp(\epsilon - e^{\epsilon})$, $\quad -\infty < \epsilon < \infty$,
cumulative distribution function (cdf)	$F(\epsilon) = 1 - \exp(-e^{\epsilon})$,
moment generating function (mgf)	$M_{\epsilon}(t) = \Gamma(1+t)$, $\quad t > -1$,
mean	$E(\epsilon) = -\gamma = -0.5772156649015329\ldots$,
	where γ is Euler's constant
and variance	$Var(\epsilon) = \pi^2/6$.

The extreme value density is skewed to the left. The probability that a standard normal random variable take a value greater than 1.96 is 0.025, while the corresponding probability for the standard extreme value is only 0.0008256. Lawless (1982, p. 17-19) and Johnson and Kotz (1970) discuss the properties of the extreme value distribution.

Under this assumption for the density of ϵ, Y_i has the pdf

$$\frac{1}{\sigma e^{X_i\beta}}\left(\frac{y_i}{e^{X_i\beta}}\right)^{\frac{1}{\sigma}-1}\exp\left[-\left(\frac{y_i}{e^{X_i\beta}}\right)^{\frac{1}{\sigma}}\right], \quad y_i > 0,$$

which will be recognized as that of a Weibull random variable (Hogg and Klugman (1984)). Under this parametrization, the shape parameter is equal to $1/\sigma$ and the scale parameter to $e^{X_i\beta}$. The hazard rate will be increasing if $\sigma < 1$, decreasing if $\sigma > 1$ and constant if $\sigma = 1$, in which case the Weibull distribution reduces to the exponential distribution. The mean and variance of Y_i are:

$$\begin{aligned} E(Y_i) &= e^{X_i\beta}\Gamma(1+\sigma) \\ Var(Y_i) &= e^{2X_i\beta}[\Gamma(1+2\sigma) - \Gamma(1+\sigma)^2]. \end{aligned}$$

A proof of those results is contained in Lawless (1982).

The likelihood function based on the data $z_i = \ln y_i$, is

$$L(\beta,\sigma) = \prod_{i=1}^{n} \frac{1}{\sigma} \exp\left[\frac{z_i - X_i\beta}{\sigma} - \exp\left(\frac{z_i - X_i\beta}{\sigma}\right)\right],$$

and the log likelihood is

$$l(\beta,\sigma) = \sum_{i=1}^{n}\left[-\ln\sigma + \frac{z_i - X_i\beta}{\sigma} - \exp\left(\frac{z_i - X_i\beta}{\sigma}\right)\right].$$

Let us define $w_i = (z_i - X_i\beta)/\sigma$.

The first and second partial derivatives of l with respect to β_j and σ are

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= -\frac{1}{\sigma}\sum_{i=1}^{n} X_{ij} + \frac{1}{\sigma}\sum_{i=1}^{n} X_{ij}e^{w_i}, \quad j = 1,\ldots,p. \\ \frac{\partial l}{\partial \sigma} &= -\frac{n}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n} w_i + \frac{1}{\sigma}\sum_{i=1}^{n} w_i e^{w_i}. \\ \frac{\partial^2 l}{\partial \beta_j \partial \beta_k} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} X_{ij}X_{ik}e^{w_i}, \quad j,k = 1,\ldots,p. \\ \frac{\partial^2 l}{\partial \sigma^2} &= \frac{n}{\sigma^2} + \frac{2}{\sigma^2}\sum_{i=1}^{n} w_i - \frac{2}{\sigma^2}\sum_{i=1}^{n} w_i e^{w_i} - \frac{1}{\sigma^2}\sum_{i=1}^{n} w_i^2 e^{w_i}. \\ \frac{\partial^2 l}{\partial \beta_j \partial \sigma} &= \frac{1}{\sigma^2}\sum_{i=1}^{n} X_{ij} - \frac{1}{\sigma^2}\sum_{i=1}^{n} X_{ij}e^{w_i} - \frac{1}{\sigma^2}\sum_{i=1}^{n} X_{ij}w_i e^{w_i}, \quad j = 1,\ldots,p. \end{aligned}$$

In appendix 10.2, we have listed some asymptotic properties of MLE's. The terms in the observed information matrix can be simplified by using the fact that the MLE's for

β_j and σ satisfy the equations $\frac{\partial l}{\partial \beta_j} = \frac{\partial l}{\partial \sigma} = 0$. The observed information matrix I_0 then becomes

$$\frac{1}{\hat\sigma^2}\begin{pmatrix} n & \sum \ln j & \sum j & \sum i+j-2 & n+\sum \hat w_i \\ \sum \ln j & \sum (\ln j)^2 e^{\hat w_i} & \sum j(\ln j)e^{\hat w_i} & \sum (i+j-2)(\ln j)e^{\hat w_i} & \sum (\ln j)\hat w_i e^{\hat w_i} \\ \sum j & \sum j(\ln j)e^{\hat w_i} & \sum j^2 e^{\hat w_i} & \sum j(i+j-2)e^{\hat w_i} & \sum j\hat w_i e^{\hat w_i} \\ \sum i+j-2 & \sum (i+j-2)(\ln j)e^{\hat w_i} & \sum j(i+j-2)e^{\hat w_i} & \sum (i+j-2)^2 e^{\hat w_i} & \sum (i+j-2)\hat w_i e^{\hat w_i} \\ n+\sum \hat w_i & \sum (\ln j)\hat w_i e^{\hat w_i} & \sum j\hat w_j e^{\hat w_i} & \sum (i+j-2)\hat w_i e^{\hat w_i} & n+\sum \hat w_i^2 e^{\hat w_i} \end{pmatrix}$$

where $\hat w_i = (z_i - X_i\hat\beta)/\hat\sigma$.

The asymptotic variance-covariance matrix of the parameters is equal to the inverse of I_0, and could be found using a symbolic computational language like MAPLE, or evaluated numerically. The expected information matrix can also easily be obtained (ref. Lawless (1982), p. 301-302).

Maximizing the log likelihood with the data of section 1.2 by using the Newton-Raphson algorithm or the SAS (1985) LIFEREG procedure, we find the MLE's, estimated standard errors and correlation matrix appearing in table 5. In section 7, we show that for certain location-scale models, the MLE's exist and are unique; this is true in particular for the Weibull-extreme value regression model.

All parameters are highly significant (at the 0.0001 level). It should also be noticed that the scale parameter estimator $\hat\sigma$ is not independent of the location parameter estimator, as is the case in normal regression. This complicates somewhat the estimation of the IBNR reserve.

Table 5: Weibull-extreme value regression

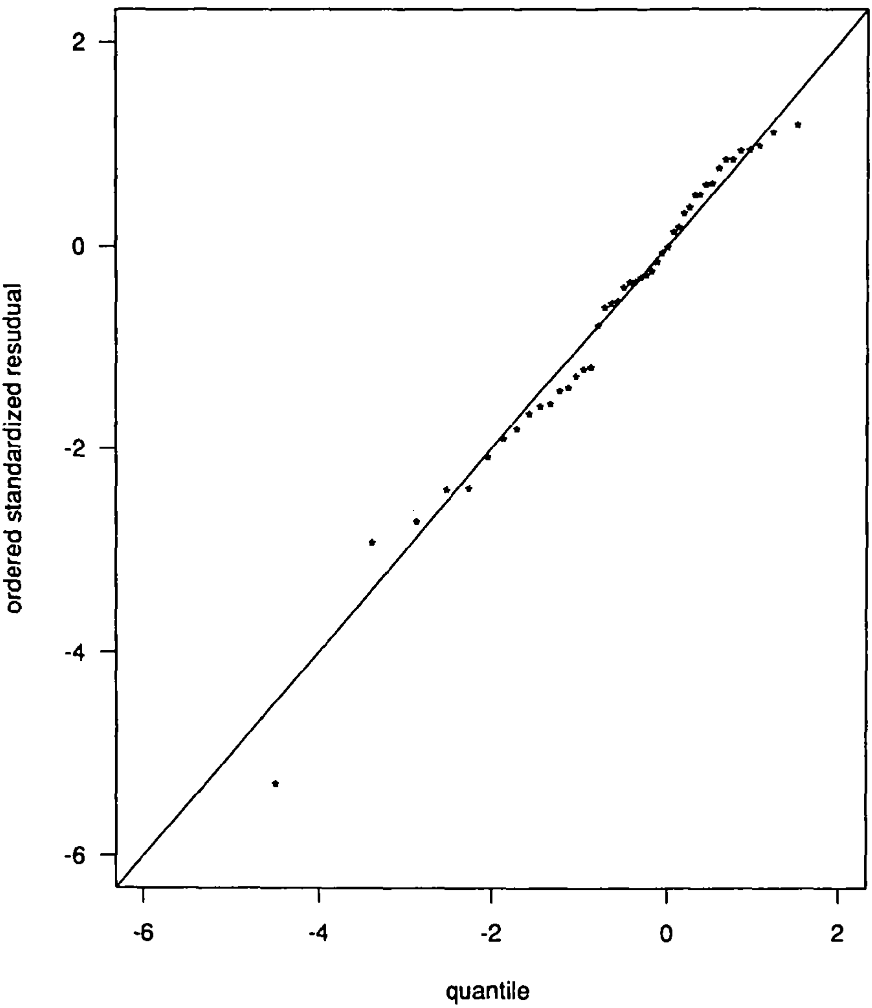
parameter	MLE	std. error	correlation matrix				
α	9.02897	0.11505	1	0.429	-0.515	-0.461	-0.017
β	-3.26637	0.25407	0.429	1	-0.972	0.214	0.0004
γ	0.40378	0.10372	-0.515	-0.972	1	-0.280	-0.006
ι	0.10811	0.01641	-0.461	0.214	-0.280	1	0.011
σ	0.02459	0.00642	-0.017	0.0004	-0.006	0.011	1

A Q-Q plot of the residuals appears in figure 2. It shows no evident departure from the extreme value distribution. It should be noted that the above standard errors and correlation matrix of the parameters are based on the joint asymptotic multivariate normal distribution of the MLE's. This approximation will be appropriate only when the number of cells in the trapezium of data is large enough (in our example, we have 45 cells).

How large is large enough? Bain and Engelhardt (1991) considered this problem for the Weibull distribution, but without covariates in the location parameter. They provide a table giving the bias of the MLE of the shape parameter of the Weibull distribution for different sample sizes. With a sample size of 40, the MLE overestimates the shape parameter by only 3.5%. If the sample size is only 10, care should be taken, since the bias is then around 15%. Those factors were obtained by a simulation study. We will not correct for the bias in our analysis, but we should remember that this might be a good idea for small sample sizes.

To test for $\sigma = 1$ (test of exponentiality of Y_i), we can use the asymptotic normality of the MLE's; unless the sample size is large, Lawless (1982) cautions that the normal approximation might not be very good. A likelihood ratio test can also be performed

Figure 2: Extreme value Q-Q plot of residuals



using the test statistic

$$\Lambda = -2 \log \frac{L(\tilde{\beta}, 1)}{L(\hat{\beta}, \hat{\sigma})},$$

where $\tilde{\beta}$ is the MLE of β under $H_0 : \sigma = 1$; the likelihood ratio statistic Λ has an asymptotic $\chi^2_{(1)}$ distribution. Performing a simple normal test leads us to reject the hypothesis $H_0 : \sigma = 1$. A Weibull distribution is therefore more appropriate for the data than an exponential distribution.

We now turn our attention to the problem of predicting the IBNR reserve. In a log-linear location-scale model, the total error in the log predicted amount Z_{kl} is composed of two parts: an estimation error on the parameters and a statistical prediction error. We saw earlier that in the Weibull-extreme value regression model, the estimators of the parameters have an asymptotic multivariate normal distribution, while the process error has an independent extreme value distribution.

Let Y_{kl} denote the random variable for the amount to be predicted in accident year k and development year l, and let us define $Z_{kl} = \ln Y_{kl}$. The random variable Z_{kl} being equal to $Z_{kl} = \hat{\alpha} + \hat{\beta} \ln k + \hat{\gamma} k + \hat{\iota}(k + l - 2) + \hat{\sigma}\epsilon$, we can appreciate the difficulty involved in trying to get its exact distribution. For this, we would need to find the distribution of the product of a normal and an extreme value random variable ($\hat{\sigma}$ and ϵ) and convolute this with a non-independent normal random variable. To get the distribution of Y_{kl}, the distribution of Z_{kl} is then exponentiated. It is highly doubtful that such a distribution would have a simple density. Instead of trying to accomplish this task, we will perform a simulation study to evaluate IBNR reserves. This will make it possible to find a confidence interval for the reserve.

Table 6: Frequency distribution of the IBNR reserve under the extreme value error assumption

Amount	Frequency
< 15000	0
15000-16000	1
16000-17000	12
17000-18000	54
18000-19000	144
19000-20000	357
20000-21000	664
21000-22000	904
22000-23000	982
23000-24000	791
24000-25000	605
25000-26000	285
26000-27000	142
27000-28000	46
28000-29000	8
29000-30000	4
30000-31000	1
> 31000	0

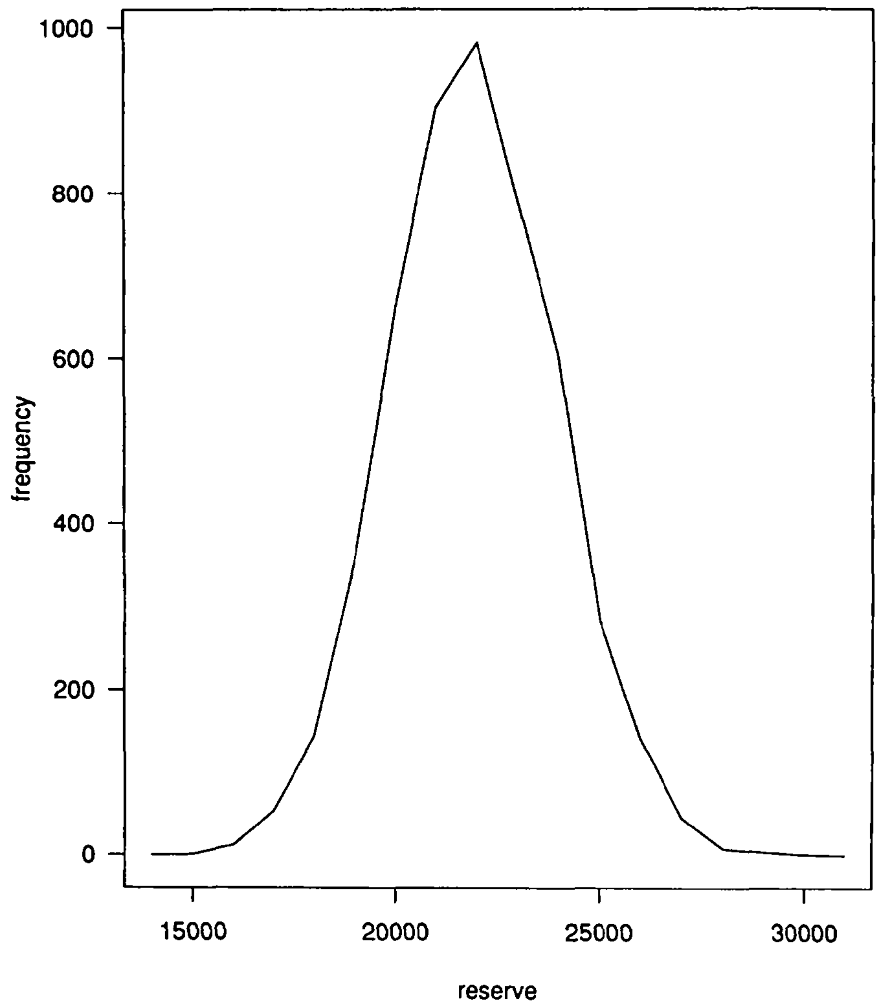
In appendix 10.1, we show how to generate a multivariate normal distribution, using the Choleski decomposition method. To be able to simulate the random variable Y_{kl}, we just need to show how to generate a standard extreme value random variable ϵ, with cdf

$$P[\epsilon \leq \epsilon_0] = 1 - \exp(-e^{\epsilon_0}), \quad -\infty < \epsilon_0 < \infty.$$

This cdf is easily inverted, yielding

$$\epsilon = \ln(-\ln(1 - U)), \quad 0 < U < 1,$$

Figure 3: IBNR reserve for Weibull-extreme value regression



where U is a uniform random variable on the interval $[0,1]$. Note that $1-U$ is also uniform on $[0,1]$, simplifying the algorithm.

Table 6 and figure 3 contain the results of a simulation of 5000 values for the IBNR reserve. The mean of the IBNR claims is 22,402 and the standard deviation of this estimate is 2011. The 80-th percentile for the simulated distribution of the IBNR reserve is 23,980.

Comparison of the extreme value and the normal distributions shows that the former has a heavier left tail and a lighter right tail than the latter. The estimation error on the regression parameters is of the same order in both models, while the stochastic error is smaller in the extreme value case.

4 Generalized loggamma regression model

The regression model used in this section will be the following

$$Z_i = \ln Y_i = X_i\beta + \sigma\epsilon_i,$$

where ϵ_i has a loggamma distribution with pdf

$$f(\epsilon;q) = \frac{|\,q\,|}{\Gamma(q^{-2})} q^{-2q^{-2}} \exp[q^{-2}(q\epsilon - e^{q\epsilon})], \quad -\infty < \epsilon < \infty,$$

and the shape parameter q can take any non-zero value (ref. Lawless (1982), p. 322-328). Under this parametrization, as q tends to 0, we obtain the normal distribution with pdf

$$f(\epsilon) = \frac{1}{\sqrt{2\pi}} \exp(-\epsilon^2/2), \quad -\infty < \epsilon < \infty.$$

The following special cases for the random variable Y_i can be obtained for certain values of the parameters q and σ: Weibull ($q = 1$), exponential ($q = \sigma = 1$), lognormal

($q = 0$) and reciprocal Weibull ($q = -1$). The density is negatively skewed for $q > 0$, with absolute skewness and kurtosis increasing as q increases; it is positively skewed for $q > 0$. A likelihood ratio test can be performed to test for the appropriateness of a certain member of the family.

Prentice (1974) and Farewell and Prentice (1977) have studied the properties of this generalized distribution. If we define the parameter $k = q^{-2}$, then it has moment generating function $\Gamma(k + t)$, $t > -k$, mean $\psi(k)$ and variance $\psi'(k)$, where $\psi(\cdot)$ and $\psi'(\cdot)$ are respectively the digamma and trigamma functions, the first and second derivatives of the gamma function. The series expansion for these two functions are:

$$\begin{aligned} \psi(n) &= -\gamma + \sum_{k=1}^{n-1} \frac{1}{k}, \qquad \text{for an integer } n \geq 2 \\ \psi'(z) &= \sum_{k=0}^{\infty} (z + k)^{-2}, \qquad z \neq 0, -1, -2, \ldots \end{aligned}$$

The log likelihood function gives

$$l(\beta, \sigma, q) = \sum_{i=1}^{n} \ln f(w_i; q) - \ln \sigma,$$

where $w_i = (z_i - X_i\beta)/\sigma$ and

$$\ln f(w_i; q) = \ln \mid q \mid -2q^{-2} \ln q - \ln \Gamma(q^{-2}) + q^{-2}(qw_i - e^{qw_i}).$$

The first and second partial derivatives of l with respect to β and σ gives

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{X_{ij}}{q\sigma} [\exp(qw_i) - 1], \quad j = 1, \ldots, p. \\ \frac{\partial l}{\partial \sigma} &= \sum_{i=1}^{n} \{ \frac{w_i}{q\sigma} [\exp(qw_i) - 1] - \frac{1}{\sigma} \} \end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} &= \sum_{i=1}^{n} X_{ij} X_{ik} (\frac{-1}{\sigma^2}) \exp(q w_i) \\
\frac{\partial^2 l}{\partial \sigma^2} &= \sum_{i=1}^{n} \frac{1}{\sigma^2}[1 - w_i^2 \exp(q w_i)] - \frac{2 w_i}{q \sigma^2}[\exp(q w_i) - 1] \\
\frac{\partial^2 l}{\partial \beta_j \partial \sigma} &= \sum_{i=1}^{n} X_{ij} (\frac{-1}{\sigma^2})[w_i \exp(q w_i) + \frac{1}{q}(\exp(q w_i) - 1].
\end{aligned}
$$

Again, using the fact that the MLE's satisfy $\frac{\partial l}{\partial \sigma} = \frac{\partial l}{\partial \beta_j} = 0$, we can simplify the last two partial derivatives and obtain

$$
\frac{\partial^2 l}{\partial \sigma^2} \Big|_{(\hat{\beta}, \hat{\sigma})} = -\frac{1}{\hat{\sigma}^2}[n + \sum \hat{w}_i^2 \exp(q \hat{w}_i)]
$$

and

$$
\frac{\partial^2 l}{\partial \beta_j \partial \sigma} \Big|_{(\hat{\beta}, \hat{\sigma})} = -\frac{1}{\hat{\sigma}^2} \sum X_{ij} \hat{w}_i \exp(q \hat{w}_i).
$$

To find the MLE's of the parameters, we can use the approach suggested by Farewell and Prentice (1977). The parameter q is fixed at a value q_0 and the profile log likelihood is maximized using the Newton-Raphson algorithm over the regression parameters β and the scale parameter σ. This gives the estimates $(\hat{\beta}(q_0), \hat{\sigma}(q_0))$. This procedure of maximizing the profile log likelihood is repeated for many values of q_0, until an overall maximum of the log likelihood over q_0 is attained. This value gives the MLE $\hat{q}$.

The SAS package fits generalized loggamma regression models. Using the SAS LIFEREG procedure for complete data, we find the results appearing in Table 7.

The default convergence criterion used by SAS is that a maximum is assumed to have occurred if the relative change in the parameters is less than 0.001. However, as can be seen from table 8, the likelihood keeps increasing beyond this value of $\hat{q}$. The convergence criterion we used is that the score statistic with respect to each parameter should be of

Table 7: Generalized loggamma regression (SAS program)

parameter	MLE	std. error	correlation matrix					
α	9.32243	0.02789	1	0.469	-0.521	-0.160	-0.497	0.497
β	-3.12566	0.07028	0.469	1	-0.991	0.645	-0.150	0.150
γ	0.35670	0.02969	-0.521	-0.991	1	-0.626	0.124	-0.123
ι	0.10058	0.00357	-0.160	0.645	-0.626	1	-0.087	0.086
σ	0.04035	0.03187	-0.497	-0.150	0.124	-0.087	1	-0.981
q	9.99342	7.63421	0.497	0.150	-0.123	0.086	-0.981	1

the order of 10^{-6}. Past the value of $q_0 = 31.623$ (corresponding to $k = q_0^{-2} = 0.001$), some elements of the information matrix become so large that it cannot be inverted and the standard Newton-Raphson algorithm fails.

Table 8: Generalized loggamma regression for various values of q_0

q_0	$\hat{\alpha}(q_0)$	$\hat{\beta}(q_0)$	$\hat{\gamma}(q_0)$	$\hat{\iota}(q_0)$	$\hat{\sigma}(q_0)$	$l(q_0)$
0	8.97986	-3.14641	0.30881	0.12298	0.31380	-11.70862
1	9.02897	-3.26637	0.40378	0.10811	0.24588	-8.66845
2	9.15105	-3.19165	0.38375	0.10369	0.17552	-7.82173
3	9.24020	-3.13178	0.35787	0.10264	0.12742	-7.23110
4	9.27974	-3.12132	0.35336	0.10188	0.09803	-6.64823
6	9.30818	-3.12572	0.35608	0.10088	0.06590	-5.68347
8	9.31835	-3.12611	0.35666	0.10061	0.04950	-5.03186
10	9.32308	-3.12419	0.35609	0.10063	0.03964	-4.62194
20	9.33019	-3.11565	0.35296	0.10088	0.01986	-3.87515
30	9.33340	-3.11023	0.35061	0.10092	0.01324	-3.68571

A few remarks should be made here.

1- the likelihood is so flat that it makes the standard error of $\hat{q}$ (7.63421), calculated assuming asymptotic normality, totally unreliable. Bain and Engelhardt (1991, p. 393) report that the asymptotic normal distribution for $\hat{k}$ will not be very accurate unless the sample size is greater than 200 or 400. Farewell and Prentice (1977) note that the skewness in the $\hat{q}$ distribution is related to an asymptotic variance that increases rapidly as $|q|$ increases. To get a confidence interval for $\hat{q}$, a likelihood ratio test would be preferable. This interval for $\hat{q}$ would include all the values q_0 satisfying

$$-2[\ln l(\hat{q},\hat{\beta},\hat{\sigma}) - \ln l(q_0,\hat{\beta}(q_0),\hat{\sigma}(q_0))] \leq 3.841.$$

2- the correlation between $\hat{\sigma}$ and $\hat{q}$ almost equal to -1 should be noted. From table 8, we can see that as q_0 increases, $\hat{\sigma}(q_0)$ decreases. Cox and Hinkley (1968) have shown that in the general regression model $Z = \alpha + X\beta + \sigma\epsilon(q)$, $(\hat{\alpha},\hat{\sigma},\hat{q})$ are asymptotically independent of $\hat{\beta}$, if the columns of X add to zero.

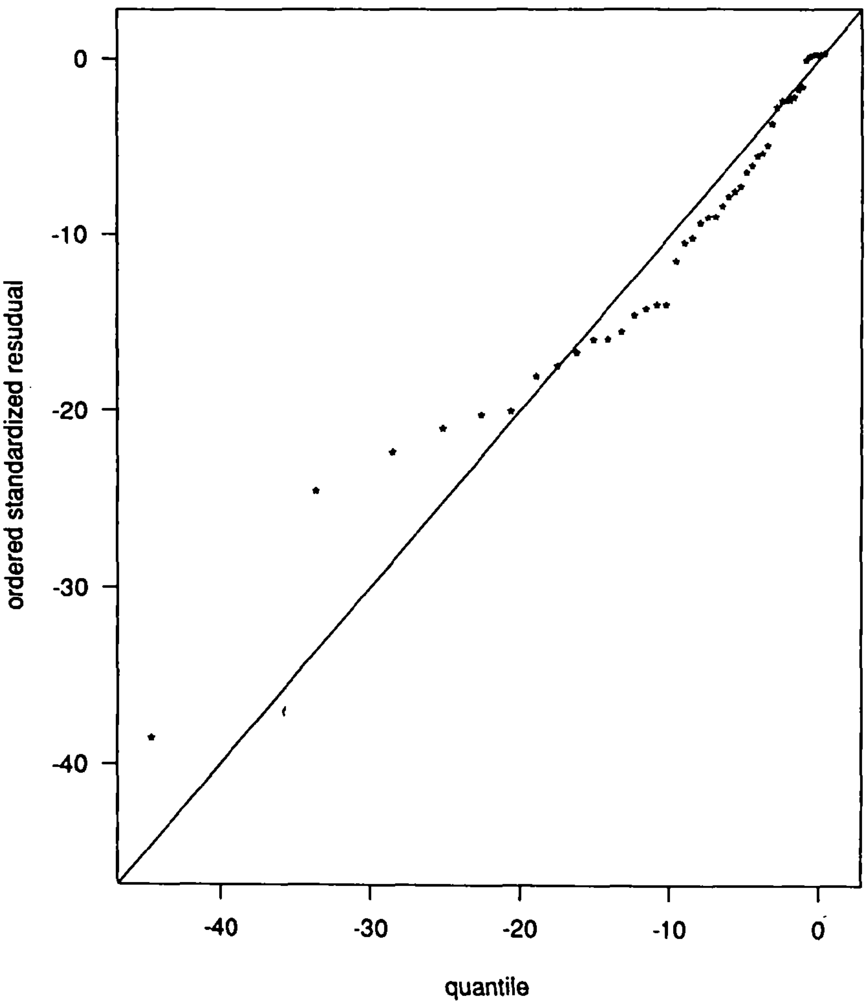
3- The regression parameters $(\alpha,\beta,\gamma,\iota)$ for any fixed value of q_0 are very close to those obtained in the normal and extreme value regression, and so is their standard error and their correlation matrix.

It should be remembered however that, although the MLE $\hat{q}$ cannot be found accurately, we know that it exists and is unique, because of the log-concavity of the loggamma distribution (see section 7).

If the exact value of $\hat{q}$, was available, this would make the estimation of E(IBNR claims) much more complicated than in the normal or extreme value cases, because of the non-independence of $\hat{q}$ with $\hat{\beta}$ and $\hat{\sigma}$. In this model, Y_{kl} is equal to

$$Y_{kl} = e^{\hat{\alpha}+\hat{\beta}\ln k+\hat{\gamma}k+\hat{\iota}(k+l-2)+\hat{\sigma}\epsilon(\hat{q})},$$

Figure 4: Loggamma (q=10) Q-Q plot of residuals



and we can see that the estimation error on the parameters is not independent of the process error $\epsilon(\hat{q})$, since $\hat{\beta}$, $\hat{\sigma}$ are estimated using the same set of past data which is used in estimating $\hat{q}$.

To assess the adequacy of the loggamma regression model, we fitted that model with a fixed q value, $q = 10$. Figure 4 presents the corresponding Q-Q plot. Since the left tail of the distribution is too short, we will not simulate the IBNR reserve; however, Devroye (1986) presents many algorithms to generate gamma random variables.

5 Logistic regression model

The logistic linear model is

$$Z_i = \ln Y_i = X_i\beta + \sigma\epsilon_i,$$

where ϵ has a standard logistic distribution with (see Lawless (1982), p. 46)

pdf	$f(\epsilon) = \frac{e^\epsilon}{(1+e^\epsilon)^2}$, $-\infty < \epsilon < \infty$,
cdf	$F(\epsilon) = 1 - (1 + e^\epsilon)^{-1}$,
mgf	$\Gamma(1 + t)\Gamma(1 - t)$, $\mid t \mid < 1$,
mean	$E(\epsilon) = 0$,
variance	$Var(\epsilon) = \pi^2/3$.

The density of the logistic distribution somewhat looks like the standard normal density. The symmetry of the pdf around $\epsilon = 0$ implies that there is probability 1/2 that the amount Y_i be understated or overstated. The probability that a standard logistic random variable exceeds 1.96 is 0.12347. The logistic distribution has thick tails, which behave like that of the exponential distribution. The loglogistic is a special case of the Burr distribution, with the parameter α equal to 1 (ref. Panjer and Willmot (1992), p. 120).

The random variable Z_i has density

$$f_{Z_i}(z_i) = \frac{1}{\sigma}\frac{\exp[\frac{z_i - X_i\beta}{\sigma}]}{[1+\exp(\frac{z_i - X_i\beta}{\sigma})]^2}, \quad -\infty < z_i < \infty,$$

and Y_i has the loglogistic density

$$\frac{1}{\sigma e^{X_i\beta}}\left(\frac{y_i}{e^{X_i\beta}}\right)^{\frac{1}{\sigma}-1}\left[1+\left(\frac{y_i}{e^{X_i\beta}}\right)^{\frac{1}{\sigma}}\right]^{-2}, \quad y_i > 0, \tag{5.1}$$

where again $e^{X_i\beta}$ is the scale parameter and $1/\sigma$ the shape parameter. In proposition 5.1, we derive the moments of order k of a loglogistic random variable with density 5.1 and show that its moment generating function does not exist.

Proposition 5.1: If Y has density

$$f_Y(y) = \frac{\delta^{1/\sigma}}{\sigma}\frac{y^{1/\sigma - 1}}{[1+\delta^{1/\sigma}y^{1/\sigma}]^2}, \quad y > 0,$$

then

$$E(Y^k) = \delta^{\frac{1}{\sigma}-(k+1)}[1-\sigma(k+1)]\pi \operatorname{cosec}[\pi\sigma(k+1)],$$

for all k such that $\frac{2}{\sigma} - 1 < k < \frac{4}{\sigma} - 1$, and the moment generating function of Y does not exist.

Proof: $E(Y^k) = \int_0^\infty y^k \frac{\delta^{1/\sigma}}{\sigma}\frac{y^{1/\sigma-1}}{[1+\delta^{1/\sigma}y^{1/\sigma}]^2}dy.$

By letting $y^{1/\sigma} = v$, we obtain

$$E(Y^k) = \delta^{1/\sigma}\int_0^\infty \frac{v^{\sigma(k+1)-1}}{[1+\delta^{1/\sigma}v]^2}dv.$$

Using the formula

$$\int_0^\infty \frac{x^{\mu-1}}{(1+\beta x)^2}dx = \frac{1-\mu}{\beta^\mu}\pi \text{ cosec } \mu\pi,$$

the result is easily obtained. The integral will have a finite value iff

$$-1 < (k+1)\sigma - 3 < 1$$

or

$$\frac{2}{\sigma} - 1 < k < \frac{4}{\sigma} - 1.$$

The moments of all positive orders do not exist; therefore, the moment generating function of Y does not exist. □

The likelihood function is

$$L(\beta,\sigma) = \prod_{i=1}^{n} \frac{1}{\sigma} \frac{\exp(w_i)}{[1+\exp(w_i)]^2}, \quad -\infty < w_i < \infty,$$

where $w_i = \frac{z_i - X_i\beta}{\sigma}$, from which we get the log likelihood

$$l(\beta,\sigma) = \sum_{i=1}^{n}[w_i - 2\ln(1+e^{w_i}) - \ln\sigma].$$

For first and second order partial derivatives with respect to the parameters, see Kalbfleisch and Prentice (1980; p. 54-57). The SAS procedure LIFEREG was used to fit a logistic regression model to the data of section 1.3. The MLE's of the parameters, their estimated standard error and the estimated correlation matrix appear in table 3.5.

A Q-Q plot of the residuals in figure 5 shows that the logistic distribution does not provide a very good fit for the right tail. We will therefore not attempt to predict the IBNR reserve, but just indicate how it could easily be done by simulation, if it was appropriate to do so.

Figure 5: Logistic Q-Q plot of residuals

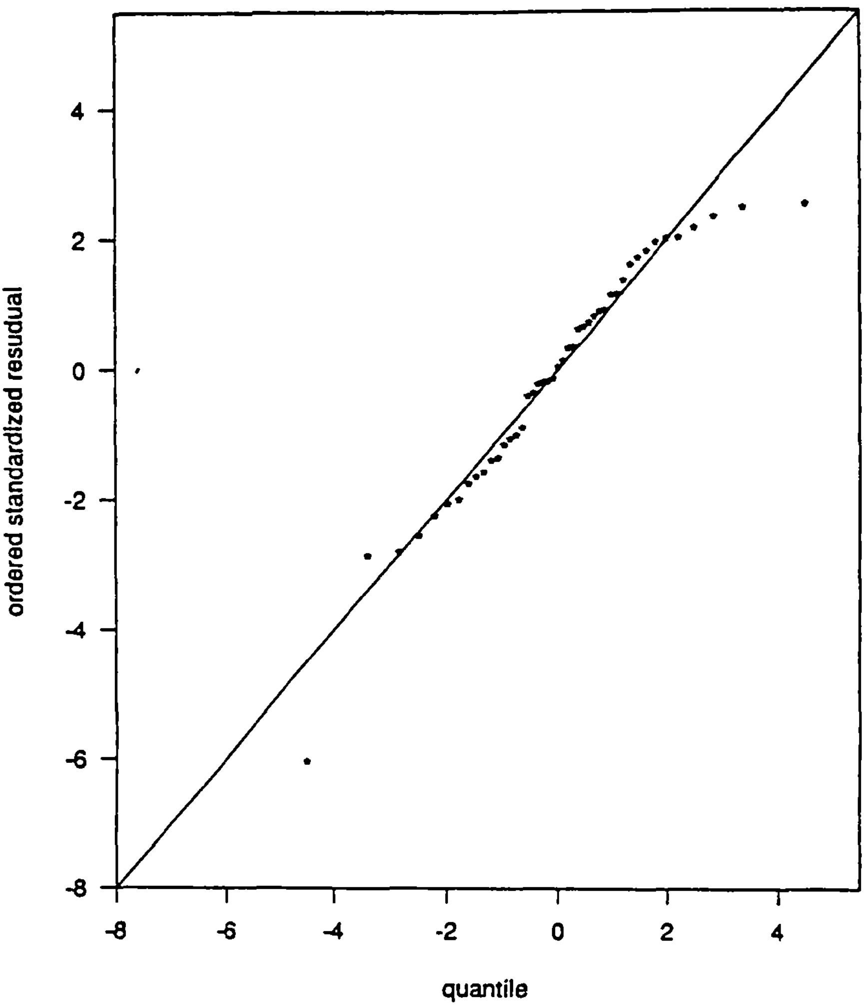


Table 9: Logistic regression

parameter	MLE	std. error	correlation matrix				
α	8.94023	0.13799	1	0.437	-0.516	-0.540	0.039
β	-3.31681	0.30143	0.437	1	-0.964	0.078	0.072
γ	0.38904	0.12058	-0.516	-0.964	1	-0.169	-0.083
ι	0.11789	0.02004	-0.540	0.078	-0.169	1	0.025
σ	0.17957	0.02203	0.039	0.072	-0.083	0.025	1

The loglogistic model for Y_{kl} is $Y_{kl} = e^{\hat{\alpha}+\hat{\beta}\ln k+\hat{\gamma}k+\hat{\iota}(k+l-2)+\hat{\sigma}\epsilon}$. The joint asymptotic distribution for $(\hat{\beta}, \hat{\sigma})$ is multivariate normal with parameter estimates given in table 9 and can be easily simulated (see Appendix 10.1). Inverting the cdf of the logistic random variable ϵ yields

$$\epsilon = \ln(\frac{1-U}{U}), \quad \text{where } U \text{ is uniform } [0,1].$$

The value of is then exponentiated to give Y_{kl}.

6 Log Inverse Gaussian regression model

The inverse gaussian regression model for Y_i is $Y_i = e^{X_i\beta+\epsilon_i}$, where the multiplicative error e^{ϵ} is assumed to have a standard inverse gaussian (IG), or Wald distribution, with density

$$f_V(v) = (2\pi\lambda v^3)^{-1/2}\exp\{-\frac{(v-1)^2}{2\lambda v}\}, \quad v > 0, \quad \lambda > 0.$$

This long-tail positively skewed distribution with exponential tails has a shape similar to that of the lognormal distribution (ref. Cohen and Whitten (1988), p. 77) and is located between the gamma and lognormal in Pearson's system of distributions, which shows possible regions of variation of the skewness and kurtosis (Jorgensen (1982), p. 19). To learn more about the inverse gaussian distribution, see Chhikara and Folks (1989) and Jorgensen (1982). Here are some of its important properties. The mean equals 1 and the variance λ. It is unimodal and a member of the exponential family. If V is $IG(1,\lambda)$, and $a > 0$ is a constant, aV is $IG(a, a\lambda)$. The sum of n independent $IG(1,\lambda)$ is $IG(n,\lambda)$.

Taking the log of Y_i, we obtain the loglinear model

$$Z_i = \ln Y_i = X_i\beta + \epsilon_i,$$

where ϵ has a log inverse gaussian (LIG) distribution. The pdf of ϵ is now derived.

Let $\epsilon = \ln V$, where V is $IG(1,\lambda)$. Then $V = e^{\epsilon}$ and $dV/d\epsilon = e^{\epsilon}$. It follows that

$$\begin{aligned} f(\epsilon) &= e^{\epsilon}(2\pi\lambda e^{3\epsilon})^{-1/2}\exp[-\frac{(e^{\epsilon}-1)^2}{2\lambda e^{\epsilon}}] \\ &= (2\pi\lambda e^{\epsilon})^{-1/2}\exp[-\frac{(e^{\epsilon}-2+e^{-\epsilon})}{2\lambda}] \\ &= (2\pi\lambda)^{-1/2}e^{-\epsilon/2}e^{1/\lambda}\exp[-\frac{1}{\lambda}\cosh\epsilon], \end{aligned} \tag{6.1}$$

where $\cosh\epsilon = (e^{\epsilon}+e^{-\epsilon})/2$.

In the next two propositions, we derive the moment generating function and the mean of the LIG distribution.

Proposition 6.1: The mgf of the LIG distribution with pdf (6.1) is

$$M_{\epsilon}(t) = (2\pi\lambda)^{-1/2}e^{1/\lambda}2K_{1/2-t}(1/\lambda).$$

Proof: Let the constant $C = (2\pi\lambda)^{-1/2}e^{1/\lambda}$. Then

$$\begin{aligned} M_{\epsilon}(t) &= E(e^{t\epsilon}) = \int_{-\infty}^{\infty} e^{t\epsilon}f(\epsilon)d\epsilon \\ &= C\int_{-\infty}^{\infty} e^{\epsilon(t-1/2)}\exp[-\frac{1}{\lambda}\cosh\epsilon]d. \end{aligned}$$

Using the formula

$$\int_{-\infty}^{\infty} \exp[-\alpha x - \frac{1}{\lambda} \cosh x] dx = 2K_{\alpha}(1/\lambda),$$

on page 309 of Gradshteyn and Ryzhik (1980), we get

$$M_{\epsilon}(t) = (2\pi\lambda)^{-1/2} e^{1/\lambda} 2K_{1/2-t}(1/\lambda),$$

for $t\epsilon[-\infty, 1/2]$, where $K_{\alpha}(\cdot)$ denotes the Bessel function of the third kind of order α. □

Proposition 6.2

$$E(\epsilon) = e^{2/\lambda}\{-\gamma - \ln(2/\lambda) - \sum_{n=1}^{\infty} \frac{(-1)^n (2/\lambda)^n}{n \cdot n!}\}$$

Proof: We know that $E(\epsilon) = M'_{\epsilon}(t) \mid_{t=0}$.

The reader will appreciate the difficulty involved in taking the derivative of $M_{\epsilon}(t)$ with respect to t, since we need to differentiate with respect to the order of the Bessel function. From Abramowitz and Stegun (1972), p. 445, we get

$$\frac{\partial}{\partial \alpha} K_{\alpha}(x) \mid_{\alpha=1/2} = -\sqrt{\frac{\pi}{2x}} E_i(-2x) e^x,$$

where $-E_i(-x) = E_1(x) = \int_x^{\infty} \frac{e^{-t}}{t} dt$. So

$$\begin{aligned} E(\epsilon) &= (2\pi\lambda)^{-1/2} e^{1/\lambda} \cdot 2\sqrt{\pi\lambda/2} E_1(2/\lambda) e^{1/\lambda} \\ &= e^{2/\lambda} E_1(2/\lambda), \end{aligned}$$

where the series expansion for $E_1(x)$ is

$$E_1(x) = -\gamma - \ln x - \sum_{n=1}^{\infty} \frac{(-1)^n x^n}{n \cdot n!}$$

□

Let us now consider the estimation of the parameters λ and β. Y_i has an inverse gaussian distribution with parameters $(e^{X_i\beta}, \lambda e^{X_i\beta})$. The likelihood function is

$$L(\beta,\lambda) = \prod_{i=1}^{n} e^{X_i\beta}(2\pi\lambda e^{X_i\beta} y_i^3)^{-1/2} \cdot \exp\left\{-\frac{(y_i - e^{X_i\beta})^2}{2\lambda e^{X_i\beta} y_i}\right\},$$

and the log likelihood is

$$l(\beta,\lambda) = \sum_{i=1}^{n} X_i\beta - \frac{1}{2}\ln\lambda - X_i\beta/2 - \frac{3}{2}\ln y_i - \frac{(y_i - e^{X_i\beta})^2}{2\lambda e^{X_i\beta} y_i}.$$

The partial derivatives are

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{n} \frac{-1}{2\lambda} + \frac{(\text{[illegible]}^{X_i\beta})^2}{\text{[illegible]}^{X_i\beta} y_i},$$

so that $\hat{\lambda} = \sum \frac{(y_i - e^{X_i\beta})^2}{n e^{X_i\beta} y_i}$.

$$\begin{aligned}
\frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{X_{ij}}{2\lambda}[\lambda + y_i e^{-X_i\beta} - e^{X_i\beta}] \\
\frac{\partial^2 l}{\partial \lambda^2} &= \sum_{i=1}^{n} \frac{1}{2\lambda^2} - \frac{(y_i - e^{X_i\beta})^2}{\lambda^3 e^{X_i\beta} y_i} \\
\frac{\partial^2 l}{\partial \lambda \partial \beta_j} &= \sum_{i=1}^{n} \frac{-X_{ij}}{2\lambda^2}[y_i e^{-X_i\beta} - e^{X_i\beta}/y_i] \\
\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} &= \sum_{i=1}^{n} \frac{X_{ij}X_{ik}}{2\lambda}[-y_i e^{-X_i\beta} - e^{X_i\beta}/y_i]
\end{aligned}$$

To find the MLE's of β and λ, one could use the Newton-Raphson algorithm. The log-concavity of the LIG distribution will guarantee the existence of unique MLE's (see section 7).

The quantiles of this distribution could be obtained from the IG distribution, since

$$P[\epsilon \le \epsilon_0] = P[e^{\epsilon} \le e^{\epsilon_0}] = P[Y \le e^{\epsilon_0}],$$

where $Y \sim IG$. Therefore the q quantile of the LIG distribution is equal to the log of the q quantile of the IG distribution. Those can be calculated or obtained from a table, e.g. Koziol (1989). If an inverse gaussian regression model was found to be appropriate, to simulate $Y_{kl} = e^{\hat{\alpha}+\hat{\beta}\ln k+\hat{\gamma}k+\hat{\iota}(k+l-2)+\epsilon}$, we would need to simulate e^{ϵ}, which is $IG(1,\lambda)$. Michael, Schucany and Haas (1976) developed an algorithm to simulate such a distribution.

7 Existence and uniqueness of MLE's

In this section, we show that all the distributions used in this chapter for the error ϵ are log-concave. A consequence of this fact is that the MLE's will exist and be unique, although they need not be finite (ref. Burridge (1981)). When convergence is achieved in the Newton-Raphson algorithm, this implies that we found a global maximum, not just a local maximum.

Let us consider the loglinear location-scale model

$$Z_i = \ln Y_i = X_i\beta + \sigma\epsilon_i.$$

If we reparametrize to $\phi = 1/\sigma$, the log-likelihood of the data becomes

$$l(\sigma,\beta) = n\ln\phi + \sum_{i=1}^{n} \ln f(w_i)$$

where $w_i = (z_i - X_i\beta)\phi$ and $f(\cdot)$ is the density function of the error ϵ_i. Since w_i is a linear function of each of the parameters β and ϕ and is therefore concave, and the function ln is concave, l will be concave provided $\ln f(\cdot)$ is concave (ref. Burride (1981)). We have therefore shown the remarkable property that, in a loglinear location-scale regression model, the existence of the MLE's does not depend on the data but only on the log-concavity of the density of the error ϵ. We now show this is indeed the case for the five distributions used so far.

1- If $\epsilon \sim N(0,1)$, $f(\epsilon) = \frac{1}{\sqrt{2\pi}} \exp(-\epsilon^2/2)$, and $\ln f(\epsilon) = K - \epsilon^2/2$; so $\frac{\partial^2}{\partial\epsilon^2} \ln f(\epsilon) = -1 < 0\ \forall\epsilon$.

2- If $\epsilon \sim$ extreme value, $f(\epsilon) = \exp(\epsilon - e^{\epsilon})$, and $\ln f(\epsilon) = \epsilon - e^{\epsilon}$; so $\frac{\partial^2}{\partial\epsilon^2} \ln f(\epsilon) = -e^{\epsilon} < 0\ \forall\epsilon$.

3- If $\epsilon \sim$ generalized loggamma,

$$f(\epsilon; q) = \frac{|\, q \,|}{\Gamma(q^{-2})} q^{-2q^{-2}} \exp[q^{-2}(\epsilon q - e^{q\epsilon})],$$

and $\ln f(\epsilon; q) = K + q^{-2}(\epsilon q - e^{q\epsilon})$; then $\frac{\partial^2}{\partial\epsilon^2} \ln f(\epsilon; q) = -e^{q\epsilon} < 0,\ \ \forall\epsilon$.

4- If $\epsilon \sim$ logistic, $f(\epsilon) = \frac{e^{\epsilon}}{(1+e^{\epsilon})^2}$; then $\ln f(\epsilon) = \epsilon - 2\ln(1+e^{\epsilon})$ and $\frac{\partial^2}{\partial\epsilon^2} \ln f(\epsilon) = \frac{-2e^{-\epsilon}}{(1+e^{-\epsilon})^2} < 0\ \forall\epsilon$.

5- If $\epsilon \sim LIG$, $f(\epsilon) = (2\pi\beta e^{\epsilon})^{-\frac{1}{2}} \exp[\frac{-(e^{\epsilon/2} - e^{-\epsilon/2})^2}{2\beta}]$; so $\ln f(\epsilon) = K - \frac{\epsilon}{2} - \frac{(e^{\epsilon/2} - e^{-\epsilon/2})^2}{2\beta}$,

$$\frac{\partial \ln f(\epsilon)}{2\epsilon} = -\frac{1}{2} - \frac{e^{\epsilon} - e^{-\epsilon}}{2\beta}$$

and $\frac{\partial^2}{\partial\epsilon^2} \ln f(\epsilon) = -(\frac{e^{\epsilon} + e^{-\epsilon}}{2\beta}) < 0\ \forall\epsilon$.

An example of a distribution for ϵ which does not have the property of log-concavity for all ϵ is the Student's t distribution with n degrees of freedom, and density

$$f(\epsilon) = \frac{(1 + \epsilon^2/2n)^{-(n+1)/2}}{\sqrt{n}\beta(1/2, n/2)}.$$

Then $\ln f(\epsilon) = K - \frac{1}{2}(n+1)\ln(1+\epsilon^2/n)$,

$$\frac{\partial}{\partial\epsilon}\ln f(\epsilon) = -(n+1)\epsilon/(\epsilon^2+n),$$

and $\frac{\partial}{\partial\epsilon^2}\ln f(\epsilon) = -(n+1)\frac{n-\epsilon^2}{(n+\epsilon^2)^2}$, which is positive for $\epsilon > \sqrt{n}$ or $\epsilon < -\sqrt{n}$.

8 Consistency of the parameters under error misspecificati‹

Gould and Lawless (1988) investigated the consistency of the maximum likelihood estimators of the regression parameters under misspecification of the error distribution in a linear location-scale model.

The postulated model is

$$Z = \alpha + X\beta + \sigma\epsilon, \quad -\infty < \epsilon < \infty, \tag{8.1}$$

where σ is a scale parameter and ϵ has a specified distribution with density $f(\epsilon)$. They assume that the true unknown model is given by

$$Z = \mu_0 + X\mu + \tau w, \quad -\infty < w < \infty, \tag{8.2}$$

where w has density $g(w)$. The location-scale structure of the postulated model has the correct form; only the error distribution is misspecified.

If the following three assumptions are satisfied,

1- the covariates are centered;

2- all the expectations below exist and

3- $n^{-1}(X'X)$ is bounded as $n \to \infty$,

White (1982) proves that the MLE's of (α, β, σ) converge in probability to a unique limit $(\alpha^*, \beta^*, \sigma^*)$. Gould and Lawless (1988) then show that $\hat{\beta} = \mu^*$ and $\hat{\beta}$ is therefore a consistent estimator of μ. In addition, for $\hat{\alpha}$ and $\hat{\sigma}$ to be consistent estimators of μ_0 and τ, they must satisfy the two equations

$$E_T(\frac{\partial}{\partial W} \log W) = 0$$

and

$$E_T(W \cdot \frac{\partial}{\partial W} \log(W) + 1) = 0 \tag{8.3}$$

where $W = (\tau w + \mu_0 - \alpha^*)/\sigma^*$ and E_T indicates that the expectation is taken with respect to the true error distribution $g(w)$.

Gould and Lawless (1988) also analyze the asymptotic efficiency of the MLE based on the correct model. We will derive conditions that $g(w)$ must satisfy in order for $\hat{\alpha}$ and $\hat{\sigma}$ to be consistent estimators of μ_0 and τ, when the error ϵ in the postulated model (8.1) has a normal $N(0,1)$ distribution.

Lemma 8.1: Under the assumption of standard normal errors in model (8.1), a sufficient condition for $\hat{\alpha}$ and $\hat{\sigma}$ to be consistent estimators of μ_0 and τ is that $E(w) = 0$ and $Var(w) = 1$.

Proof: If $f(\epsilon) = \frac{1}{\sqrt{2\pi}} e^{-\epsilon^2/2}$, then $\frac{\partial}{\partial \epsilon} \log f(\epsilon) = -\epsilon$, and the equations (8.3) become $E_T(W) = 0$ and $E_T(W^2) = 1$.

Since $W = (\tau w + \mu_0 - \alpha^*)/\sigma^*$, the condition $E_T(W) = 0$ implies that $\mu_0 = \alpha^*$ i.e. $\hat{\alpha}$ is a consistent estimator of μ_0. If $E_T(W) = 0$, then $E_T(W^2) = Var_T(W) = (\tau/\sigma^*)^2 Var(w) =$

1. The condition $Var(w) = 1$ will imply that $\tau = \sigma^*$, i.e. that $\hat{\sigma}$ is a consistent estimator of τ. □

The consistency of $\hat{\alpha}$ and $\hat{\sigma}$ therefore depends only on the first two moments of the distribution of w, when the postulated model is lognormal linear.

We must point out here that one of the assumptions for the above development to be valid is that $n^{-1}(X'X)$ be bounded as $n \to \infty$. This condition is not verified in the model

$$Y_{ij} = \alpha + \beta \ln j + \gamma j + \iota(i + j - 2) + \epsilon_{ij}.$$

The covariate i would need to be removed from the model, for example by normalizing the amounts Y_{ij}, in order for $n^{-1}(X'X)$ to be bounded as $n \to \infty$.

9 Conclusion

In this paper, we have presented an anthology of models differing between them only in the distribution assumed for the error ϵ. To discriminate between the normal, extreme value, logistic and loggamma distribution for ϵ, we can assume that ϵ belongs to the generalized $\log F$ distribution (Prentice (1974)), with pdf

$$f(\epsilon) = (m_1/m_2)^{m_1} e^{wm_1}[1 + m_1 e^{w}/m_2]^{-(m_1/m_2)}.$$

After finding the MLE's $(\hat{m}_1, \hat{m}_2)$, we can perform a likelihood ratio test for

$$\begin{array}{ll} (m_1, m_2) = (1,1): & \text{logistic distribution} \\ (m_1, m_2) = (1,\infty): & \text{extreme value distribution} \\ m_2 = \infty: & \text{generalized loggamma distribution} \\ (m_1, m_2) \to (\infty,\infty): & \text{normal distribution,} \end{array}$$

to select one particular member of the family. Gould (1986) did an extensive study of the location-scale model with the error ϵ following the $\log F$ distribution. Her conclusions are that if one tries to estimate two shape parameters as in the $\log F$ family, the precision of the estimates may be so low as to make them virtually uninformative. However, as we have also observed, the MLE $\hat{\beta}$ of the regression parameters is quite robust with respect to misspecification of the distribution of ϵ.

Numerous other researchers have in the past also encountered difficulty when trying to estimate the shape parameter of the generalized loggamma distribution. Lawless (1982, p. 237), observed that, even with sample sizes of 200 or 300, it is not uncommon for the Newton-Raphson algorithm not to converge to the MLE's. Because in usual insurance situations, the trapezium of data contains a small number of cells (in our case, 45 observations with 5 parameters to estimate), the actuary might encounter problems with this distribution. According to Prentice (1974), two distributions in the loggamma family with very different values of the shape parameter k, will look very similar, creating estimation problems. The extreme value distribution ($q = 1$) is difficult to discriminate from the normal distribution ($q = 0$), when the sample size is small.

In view of these facts, we therefore recommend that a simple distribution be assumed for ϵ, like the extreme value or the normal. After comparing the log likelihood, fit can be assessed by a Q-Q plot. If a symmetric distribution is needed, the normal distribution should be assumed for ϵ, since it is the only symmetric member of the generalized loggamma family. Fitting the normal model is useful for finding initial parameter estimates for the extreme value model. The estimated IBNR reserve can then be easily calculated under both assumptions.

The assumption of a normal distribution for ϵ presents one advantage over that of the extreme value distribution. When reserves are to be discounted for interest, we can still find the distribution of the present value of the future payments. If the force of interest δ is constant over a year, it follows from a property of the lognormal distribution that the joint distribution of the discounted value of the future payments is also multivariate lognormal. Stochastic interest rates could also be built into the model and the reserve estimated by simulation.

In conclusion, regression models present many advantages over the chain ladder method: they have fewer parameters and do not underestimate the reserve; the properties of the estimators of the parameters have been well studied; they take into account both the error involved in the estimation of the parameters and the statistical error inherent in the prediction of future claims; the fit of the model can be tested statistically by a $Q-Q$ plot; and confidence intervals for the reserve can be calculated with a simulation. We therefore strongly advocate the use of regression models.

10 Appendices

10.1 Algorithm to generate a multinormal random variable

To simulate the distribution of the IBNR reserve, we need to generate a $MLN(\mu, \sum)$ random variable. The following algorithm was used.

1. Generate $Z \sim MN(0, I)$, using the Box-Muller transformation

$$Z_1 = (-2 \ln U_1) \cos(2\pi U_2)$$

$$Z_2 = (-2 \ln U_2) \cos(2\pi U_2),$$

   where U_1 and U_2 are i.i.d., uniform on $(0, 1)$.

2. Transform Z to Y, a $MN(\mu, \sum)$ distribution:

$$Y = \mu + CZ,$$

   where $\sum = CC'$ and C is calculated from the Choleski factorization algorithm (ref. Kellison (1975)):

$$\begin{aligned} c_{11} &= \sqrt{\sigma_{11}} \\ c_{ij} &= \frac{1}{c_{jj}}(\sigma_{ij} - \sum_{k=1}^{j-1} c_{ik}c_{kj}) \\ c_{ii} &= \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} c_{ik}^2} \end{aligned}$$

3. Exponentiate each component of Y

$$e^Y = (e^{Y_{kl}}) \sim MLN(\mu, \sum).$$

10.2 Asymptotic properties of MLE's

If $X_1, \ldots, X_n$ is a random sample of size n from the density $f(x; \underline{\theta})$, where $\underline{\theta} = (\theta_1, \ldots, \theta_{p+1})$ contains the regression parameter vector β and the scale parameter σ, then under certain regularity conditions, the following results hold.

1- The MLE $\hat{\underline{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ exists.

2- It is a consistent estimator of θ.

3- $\hat{\theta}_1, \ldots, \hat{\theta}_{p+1}$ are asymptotically efficient,

$$\text{i.e.} \quad \lim_{n \to \infty} \frac{Var(\hat{\theta}_j)}{CRLB(\hat{\theta}_j)} = 1,$$

where $CRLB(\hat{\theta}_j)$ is the Cramér-Rao lower bound, obtained as $1/nE[\frac{\partial \log f}{\partial \theta_j}]^2$.

4- $\sqrt{n}(\hat{\theta} - \theta)$ has an asymptotically multivariate normal $MN(\underline{0}, I_0^{-1})$ distribution where I_0 is the observed information matrix, with element

$$I_{ij} = -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L(\underline{\theta}; x_1, \ldots, x_n) \mid_{\underline{\theta} = \hat{\underline{\theta}}} .$$

11 References

Abramowitz, M. and Stegun, I.A. (1972). Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, U.S. Department of Commerce, Washington.

Bain, L.J. and Engelhardt, M. (1991). Statistical Analysis of Reliability and Life-Testing Models, Theory and Methods, Marcel Dekker Inc., New York.

Burridge, J. (1981). A note on maximum likelihood estimation for regression models using grouped data. JRSS B, 43, 41-45.

Chhikara, R.S. and Folks, J.L. (1989). The Inverse Gaussian Distribution: Theory, Methodology and Applications. Marcel Dekker Inc., New York.

Cohen, C. (1988). Three-parameter estimation. In Lognormal Distributions, Theory and Applications, edited by Crow, E.L. and Shimizu, K., Marcel Dekker, Inc., New York.

Cohen, A.C. and Whitten, B.J. (1988). Parameter estimation in reliability and life span models. Marcel Dekker, Inc., New York.

Cox, D.R. and Hinkley, D.V. (1968). A note on the efficiency of least-squares estimates. JRSS B, 30, 284-289.

Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag, New York.

Doray, L.G. (1992). UMVUE of the IBNR in a Lognormal Linear Regression Model, submitted to Insurance: Mathematics and Economics.

Farewell, V.T. and Prentice, R.L. (1977). A study of distribution shape in life testing. Technometrics, 19, 69-75.

Gould, A. (1986). Some issues in the regression analysis of survival data. Ph.D. thesis, University of Waterloo.

Gould, A. and Lawless, J.F. (1988). Consistency and efficiency of regression coefficient estimates in location-scale models. Biometrika, 75, 535-540.

Gradshteyn, I.S. and Ryzhik, I.M. (1980). Tables of Integrals, Series and Products. Academic Press, New York.

Hogg, R.V. and Klugman, S.A. (1984). Loss Distributions. John Wiley and Sons, Inc., New York.

Johnson, N.L. and Kotz, S. (1970). Distributions in Statistics: Continuous Univariate Distributions, Vol. 1 and 2, Houghton Mifflin Company, Boston.

Jörgensen, B. (1982). Statistical Properties of the Generalized Inverse Gaussian Distribution. Springer-Verlag, New York.

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.

Kellison, S.G. (1975). Fundamentals of Numerical Analysis. Richard D. Irwin, Inc., Illinois

Koziol, J.A. (1989). Handbook of Percentage Points of the Inverse Gaussian Distribution. CRC Press, Inc., Boca Raton, Florida.

Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data. Wiley, New York.

Michael, J.R., Schucany, W.R. and Haas, R.W. (1976). Generating random variates using transformations with multiple roots. The American Statistician, 30, 88-90.

Panjer, H.H. and Willmot G.E. (1992). Insurance risk models. Society of Actuaries, Schaumburg.

Prentice, R.L. (1974). A loggamma model and its maximum likelihood estimation. Biometrika, 61, 539-544.

Reinsurance Association of America (1987). Loss Development Study. In Proceedings of the XXIth Astin Colloquium, New York, 1989.

SAS (1985) User's Guide: Statistics, Version 5 Edition, SAS Insitute Inc. Cory, NC.

Taylor, G.C. (1986). Claim reserving in non-life insurance. North-Holland, Amsterdam.

White, H. (1982). Maximum likelihood estimation of misspecified models. Econometrica, 50, 1-25.

Zenwirth, B. (1990). Interactive Claims Reserving Forecasting System, Version 5.3, Manual Volume 1, Benhar Nominees Pty. Ltd., Australia.