Mathematical Statistics
Stockholm University

Estimation of Dispersion Parameters in GLMs with and without Random Effects

Meng Ruoyan

Examensarbete 2004:5

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.matematik.su.se/matstat

Mathematical Statistics
Stockholm University
Examensarbete **2004:5**,
http://www.matematik.su.se/matstat

Estimation of Dispersion Parameters in GLMs with and without Random Effects

Meng Ruoyan*

January 2004

Abstract

In this paper, we study estimation of dispersion parameters in GLMs with and without random effects with focus on its application in non-life insurance. Three different methods of estimating ϕ, the dispersion parameter in the GLMs, have been presented, namely maximum likelihood, Pearson and Deviance methods. We extended the GLMs by introducing a random effect representing so-called multi-level factor. This introduces an additional dispersion parameter α. We investigate two methods that have been suggested for estimating α: the maximum likelihood method and an alternative method proposed by Ohlsson and Johansson(2003b). We study and compared all the estimation methods partly by empirical studies with data from the non-life insurance and partly by simulations.

Keywords: Generalized linear models, dispersion parameters, multilevel factors, Tweedie models.

*E-mail: mengruoyan@mailcity.com
Supervisors: Esbjörn Ohlsson and Björn Johansson.

Contents

1 Non-life Insurance and Generalized Linear Models

The aim of this paper is to study estimation methods for dispersion parameters in Generalized Linear Models (GLMs) in the context of non-life insurance. Before we go into details, we need to give a brief picture of what GLMs are and how they are used in non-life insurance. In the first part of this section, we are going to introduce some important definitions concerning non-life insurance. In the second part an outline of the theory of GLMs will be presented. Finally, we present more details on how GLMs are applied in modelling the average claims in non-life insurance.

1.1 Mathematical Modelling for Non-life Insurance

Non-life insurance protects the insured against certain risks, such as financial loss or property damage liability. The insured is obliged to pay a premium to the insurer who undertakes to indemnify for the losses. The premium is the price to insure a specified risk for a specified period of time. An important task for actuaries is to model the premium so that it reflects differences between groups of customers. The most important part of the premium is the risk premium, which is defined as the average claim cost per insured. It can be expressed in terms of two other quantities, average claim amount and claim frequency:

$$\text{Risk premium} = \text{claim frequency} \times \text{average claim amount}$$

where *average claim amount* equals the total claims cost divided by the number of claims and *claim frequency* is the average number of claims per year. In this paper we shall only consider the problem of modelling the *average claim amount.*

We seek a model that describes the differences in average claim amount among various groups of customers. For instance, in the case of car insurance, large cars may have higher average claims than small ones. To deal with the problem, insurance companies often define several variables, which are related to properties of the customers or the insured objects. Such variables are called *rating factors*. In car insurance, ages of cars and ages of drivers are two commonly used rating factors. Rating factors are often discretized by forming group of adjacent values. For instance, car age can be discretized by letting cars aged 0-2 years form one group, 3-5 years another and so on. Customers belonging to the same group with respect to all the rating factors are said to belong to the same *tariff cell*. In the model, customers in the same tariff cell are considered to have the same average claim amount.

In a statistical model, we consider rating factors as regressors for the response variable, the average claim amount. In this circumstance, the classical regression models are no longer applicable for the following reasons:

1. Classical linear models are additive whilst we prefer a multiplicative model, because in the context of average claim amount we are more interested in the relative change against a 'standard level' rather than the change in absolute values.

2. In the classical linear models, it is assumed that the response variable has a Gaussian distribution. This is obviously unreasonable for average claim amount, which only takes non-negative values and is skewed to the right.

Hence classical linear models are not the most suitable in this context. We need to find other models which can solve these problems.

1.2 Introduction to Generalized Linear Models

Generalized linear models (GLMs) are an extension of classical linear models. Firstly, GLMs allow a more extensive distribution family for the response variable than just the Gaussian distribution. The probability density for the response variable can be expressed as:

$$f(y) = \exp\left\{\frac{y\theta - b(\theta)}{\phi/w} + c(y,\phi)\right\} \tag{1}$$

where w denotes the weight of the response variable. This distribution family is known as 'Exponential Dispersion Models' (EDM). Apart from the Gaussian distribution, there are many other distributions belonging to this family, such as the binomial, Poisson and gamma distributions.

GLMs are parameterized in terms of the parameters θ and ϕ. The latter is called the dispersion parameter. We are going to focus on estimation methods for this parameter in the next section.

The mean and variance of Y can be written as follows:

$$\mu = \mathbf{E}(\mathbf{Y}) = b'(\theta)$$

$$\mathbf{Var}(\mathbf{Y}) = \frac{b''(\theta)\phi}{w} = \frac{V(\mu)\phi}{w} \tag{2}$$

where $V(\mu)$ is called the variance function. The variance function is of much importance because it has one-to-one relationship to the distribution within EDM. In another words, the variance function determines exclusively the specific distribution within the EDM family (Jögensen, 1987). Table 1 summarizes some variance functions for different distributions in EDM:

Distribution	Normal	Poisson	Gamma	Binomial
$V(\mu)$	1	μ	μ^2	$\mu(1-\mu)$

Table 1: *Examples of the variance functions for the distributions in EDM.*

The other extension in GLM is that the relation between the mean and the regressors does not have to be linear. The so-called link function relates a linear predictor η to the expected value μ of response variable Y where η is defined by

$$\eta = \sum_{j=0}^{J} \beta_j \cdot x_j \tag{3}$$

where x_j are indicator variables determining whether to include a β_j or not. In classical linear models we define that $\eta = \mu$ whereas in the case of GLMs η and μ are related to each other through a monotonic differentiable link function $\eta = g(\mu)$. For instance $\eta = \log(\mu)$ and $\eta = \log(\frac{\mu}{1-\mu})$ are two of the most common link functions.

1.3 GLMs for the Average Claim Amount

As has been discussed in the previous section, GLMs have some advantages over classical linear models for modelling average claim amount. Before we apply GLMs to average claim amount, we first need to make some assumptions:

Assumption 1.1

a. The amount of different claims are independent of each other.

b. Within a tariff cell, all claims amounts have the same distribution.

There are mainly two ways to fit GLMs to data. One is to aggregate data in terms of the total cost of claims and total number of damages within every single tariff cell before fitting the GLM. The other way is to fit GLM directly to all the individual observations without aggregation. The results should be the same as concerns the regression parameters β_j, but we choose the latter one for the reason that non-aggregated data contains more information on the variance than the aggregated data does and thus should give a more precise estimate of ϕ.

We now define X_i as the i:th claim amount where $i = 1, 2, 3...M$ and M is the total number of claims. According to the definition, the response variable, average claim amount $Y_i = X_i/w_i$ has w_i equal to 1 in the non-aggregated data. The model for $\mathrm{E}(Y_i)$ can be described as by equation 3, where β_j in this case denotes the j:th rating factor. We use a log-link, $\eta_i = \log(\mu_i)$ to build up a multiplicative model for $\mathrm{E}(Y_i)$, which can be rewritten as follows:

$$E(Y_i) = \mu_i = \exp{(\eta_i)} = \exp{(\sum_{j=0}^{J} \beta_{j(i)} x_{j(i)})} = \gamma_0 \cdot \gamma_{1(i)} \cdot \gamma_{2(i)} \cdot ... \cdot \gamma_{j(i)} \qquad (4)$$

where $\gamma_{j(i)}$ is the relativity of rating factor j for claim number i. It is customary to choose as reference cell the tariff cell with the largest number of claims. γ_0 denotes the parameter for the reference cell. If claim i belongs to the reference cell, $\gamma_{j(i)} = 1$ for all the j and γ_0 is the expectation of claim i.

As for the distribution of Y_i, there is no canonical choice. However in practice, it is often reasonable to assume that the standard deviation is approximately proportional to its mean value. Based on this fact, a often used assumption when using GLMs to model average claim amounts is that $V(\mu_i) = \mu_i^2$. From Table 1, we see that the corresponding distribution in EDM is the gamma distribution.

2 Estimation of the Dispersion Parameter ϕ

In this section we consider estimation of the dispersion parameter ϕ in GLMs in the context of non-life insurance. The parameter ϕ affects GLMs in the following aspects:

Firstly, for EDM, the variance of the response variable Y has the form: $Var(Y) = V(\mu)\phi/w$. This shows that the variance of Y is proportional to the value of ϕ, which implies that a good estimator of ϕ gives a good reflection of the variance of Y.

Secondly, for GLM we estimate the parameters β by maximizing the log-likelihood function, l for Y, i.e by setting $\partial l/\partial\beta = 0$. According to asymptotic theory the estimates of β follows approximatively a normal distribution: $\hat{\beta} \approx N(\beta; \mathbf{I}^{-1})$ where $\mathbf{I}$ is Fisher's information matrix, the negative expected value of $\mathbf{H}$, the matrix of the second order derivatives of the log-likelihood function l with respect to β:

$$I = -E\left(\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right) = -\sum_i x_{ij} \frac{w_i}{\phi v(\mu_i) g'(\mu_i)^2} x_{ik}$$

This expression tells us that although the value of $\hat{\beta}$ does not depend on ϕ, ϕ does affect the precision of β and a precise estimate $\hat{\phi}$ is important to get a good confidence interval for β.

In section 2.1, we introduce three different estimation methods for ϕ, and in section 2.2, we apply the methods to car insurance.

2.1 Three Estimation Methods of ϕ

We will assume that $V(\mu_i) = \mu_i^2$, corresponding to a gamma distribution. The density function can be expressed in the parameters μ and ϕ as (Ohlsson & Johansson, 2003):

$$f_Y(y_i) = exp\left(\frac{-y_i/\mu_i - \log(\mu_i)}{\phi/w_i} + c(y_i, \phi, w_i)\right) \tag{5}$$

where $c(y_i, \phi, w_i) = \log(w_i y_i/\phi) w_i/\phi - \log(y_i) - \log\Gamma(w_i/\phi)$

The expression above is a straight forward re-parametrization of the density function of gamma distribution in terms of α and β (See Appendix A1):

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$$

We are going to present three methods to estimate the parameter ϕ:

- Deviance method

- Pearson method
- Maximum Likelihood method

2.1.1 Deviance and Pearson Estimation of ϕ

We start with Deviance estimation of ϕ. Deviance, as a measure of goodness of fit, is defined as following:

$$D(\mathbf{y},\hat{\mu}) = \phi \cdot D^*(\mathbf{y},\hat{\mu}) = 2 \cdot \phi[l(\mathbf{y},\phi) - l(\hat{\mu},\phi)]$$

Where $l(\cdot)$ is log-likelihood function. D and D^* are called scaled and non-scaled Deviance respectively. We can estimate ϕ by the mean scaled Deviance, expressed in equation 6 (Ohlsson and Johansson, 2003a):

$$\hat{\phi}_d = \frac{D(\mathbf{y},\hat{\mu},\phi)}{n-r} = 2 \cdot \phi \frac{[l(\mathbf{y},\phi) - l(\hat{\mu},\phi)]}{n-r} \tag{6}$$

where r is the total number of unknown parameters. $D^* = D/\phi$ is approximately $\chi^2(n-r)$ distributed with expectation $(n-r)$. Thus $\hat{\phi}_d$ is approximately unbiased.

Now we assume that the response variable Y_i follows a gamma distribution, by which we can obtain the explicit expressions for $\hat{\phi}_d$:

$$\hat{\phi}_d = \frac{2}{n-r} \cdot \sum_i w_i \cdot \left(\log\left(\frac{\hat{\mu}_i}{y_i}\right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right) = \frac{2}{n-r} \cdot \sum_i w_i \cdot \log\left(\frac{\hat{\mu}_i}{y_i}\right) \tag{7}$$

where we use the fact that $\sum_i w_i \cdot (y-\hat{\mu})/\hat{\mu} = 0$, which can be derived from the ML equation for $\hat{\beta}$ in the case of gamma distribution (Ohlsson & Johansson, 2003).

Apart from the Deviance estimator of ϕ, we can also use the Pearson method to estimate ϕ, where the $\hat{\phi}_p$ can be expressed as:

$$\hat{\phi}_p = \frac{\phi \cdot X^2}{n-r} \tag{8}$$

where X^2 is the Pearson X^2 statistic with the form:

$$X^2 = \frac{1}{\phi} \sum_i \left(w_i \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)} \right)$$

.

$\hat{\phi}_p$ is approximately unbiased because Pearson's X^2 follows approximately $\chi^2(n-r)$ distribution with expectation $(n-r)$.

Under the assumption that Y_i is gamma distributed, we can derive the explicit expression of $\hat{\phi}_p$:

$$\hat{\phi}_p = \frac{1}{n-r} \cdot \sum_i w_i \cdot \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)} = \frac{1}{n-r} \cdot \sum_i w_i \cdot \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2} \tag{9}$$

2.1.2 Maximum Likelihood Estimation of ϕ

The maximum likelihood method (ML) estimate is the value that maximizes the likelihood or log-likelihood function. With the assumption that $V(\mu_i) = \mu_i^2$, Y_i follows a gamma distribution within EDM and its log-likelihood function can be expressed as:

$$l(\mu_i; \phi) = \sum_i^n \left(\frac{w_i}{\phi} \cdot \log\left(\frac{w_i}{\phi\mu_i}\right) + \left(\frac{w_i}{\phi} - 1\right) \cdot \log(y) - \frac{w_i y_i}{\phi\mu_i} - \log\left(\Gamma(\frac{w_i}{\phi})\right)\right)$$

where n is the total number of observations. To obtain the ML estimator of ϕ we first obtain the ML estimate of μ, which is independent of ϕ. Then the ML equation is obtained by setting the derivative of the log-likelihood function with respect to ϕ equal to zero. The ML estimate of ϕ is the solution of the ML equation.

The ML equation expressed in terms of the parameters μ_i and ϕ for the gamma distribution has the form (for the proof see Appendix A2):

$$D^* = 2 \cdot \sum_{i=1}^n w_i \left(\log\left(\frac{w_i}{\hat{\phi}}\right) - \Psi\left(\frac{w_i}{\hat{\phi}}\right)\right)$$

where D^* as defined before, the deviance for the gamma distribution and $\Psi(\cdot)$ is known as the digamma function:

$$D^* = 2 \cdot \sum_i^n w_i \cdot \log\left(\frac{\hat{\mu}_i}{y_i}\right); \quad \Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

Since Y_i here represents the i:th observation of loss, $w_i = 1$ for all the i. The equation above can be reduced to $\log(1/\hat{\phi}) - \Psi(1/\hat{\phi}) = D/2n$ and then an approximate solution for $\hat{\phi}$ is [Cordeiro and McCullagh 1991]:

$$\hat{\phi}_m \approx \frac{2D^*}{n(1 + (1 + 2D^*/3n)^{1/2})} \tag{10}$$

2.1.3 Comparison of the three Methods

This section applies only when the response variable Y is gamma distributed. An interesting statement made by Cordeiro and McCullagh (1991) is that from the inequality $1/2x < \log(x) - \Psi(x) < 1/x$, we can induce a relation between ϕ_d and ϕ_m:

$$D^*/2n < \hat{\phi}_m < D^*/n \Rightarrow \frac{\hat{\phi}_d \cdot (n-r)}{2n} < \hat{\phi}_m < \frac{\hat{\phi}_d \cdot (n-r)}{n}$$

This implies that when the total number of observations is large, $\hat{\phi}_d$ should always be larger than the $\hat{\phi}_m$ and the values of $\hat{\phi}_m$ is in between $\hat{\phi}_d/2$ and $\hat{\phi}_d$.

We can also compare the Pearson estimator with the deviance estimator by taking advantage of Taylor expansion:

$$f(a) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 \ldots \frac{f^{(n)}(a)}{n!}(x-a)^n \ldots$$

We apply the above to the expression of $\hat{\phi}_d$ and obtain (for a proof see Appendix A3):

$$\hat{\phi}_d \approx \frac{1}{n-r} \cdot \sum_i w_i \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)^2 - \frac{1}{n-r} \cdot \sum_i \frac{2w_i}{3} \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)^3$$

We see that $\hat{\phi}_p$ is the first term of the expression above. The last term, which can take both positive and negative values can be considered as a measure of the difference between these two estimators. Thus the $\hat{\phi}_p$ can be either larger or smaller than the $\hat{\phi}_d$. These two estimators are approximately the same only when the values of y_i are close to the predicted values, i.e. the GLM has a good fit to the data.

What deserves notice here is that even though the Pearson and Deviance estimators can be considered as measures of deviations of Y_i from $\hat{\mu}_i$, they can come to quite different results. The Pearson estimator to some extent might exaggerate the deviations by taking the square of the difference between y_i and $\hat{\mu}_i$ while the Deviance estimator on the opposite might under-estimates the deviations because it diminishes with the logarithm of the ratio of $\hat{\mu}_i$ and y_i.

2.2 Numerical Examples from Car Insurance

In this section we study the ϕ estimators for the data from car insurance. In order to gain more insights into the properties of the ϕ estimators, in the next section we study the empirical distributions of the ϕ estimators obtained by simulation. Finally our empirical outcomes will be compared with the theoretical discussion in the previous section.

2.2.1 Analysis of car insurance data

The data that we use consists of totally 896,321 car insurances claims from year 1995 to 2000. We divide all the claims into 4 groups corresponding to the insurance covers, namely *Third part liability* (TPL), *Hull*, *Partial kasko* and *Mer*. The cover *TPL* protects the insured against financial loss arising out of legal liability imposed upon him/her in connection with bodily injury, property damage, medical payments etc. *Hull* covers damage made to the insured's vehicle that results from a collision with another vehicle or object or damage incurred when vehicle is transported by boat or train etc. *Partial kasko* is a comprehensive cover that protects against losses resulting from glass damage (*Gls*), theft (*Sto*), fire (*Bra*), salvage (*Rad*) and machine (*Mas*). *Mer* covers some risks that are not defined in the other risk covers, e.g. intentional damage.

We are going to fit a GLM to the empirical data for each cover and each subcover of *Partial kasko* and thereby obtain the corresponding ϕ estimates. Each observation corresponds to a specific claim identified by an exclusive claim number. The data set includes only claims with positive amount, i.e. we exclude zero claims. The following table lists the number of observations and the claim years for each cover/subcover:

Risk cover	Nmb of obs	Duration
TPL	51,880	1995-1998
P.kasko	342,105	1995-2000
Mer	51,013	1995-2000
Hull	109,797	1995-2000
Gls	198,817	1995-2000
Sto	97,005	1995-2000
Bra	6,257	1996-2000
Rad	32,045	1995-2000
Map	7,402	1996-2000

Table 2: *Car insurance: Number of observations of average claim amounts from 1995 to 2000 in terms of different covers.*

Using the SAS procedure *Proc Genmod*, we estimate ϕ by the three different methods for the different covers. The estimation results are summarized in table 3:

2.2.2 Analysis of Estimators

In this section, we are going to discuss the estimation results in two different ways. The first way is to compare the estimates for different covers. Secondly

Average claim			
Cover	Pearson	Deviance	ML
TPL	34,054	2,623	2,054
P.kasko	6,477	1,317	1,124
Mer	1,957	1,202	1,035
Hull	1,285	1,023	0.897
Gls	8,143	0,384	0,362
Sto	2,363	1,146	0,993
Bra	1,764	1,330	1,125
Rad	3,723	1,257	1,077
Mas	1,150	1,026	0,893

Table 3: *Estimation results of parameter ϕ in the GLM for average claims in car insurance with three different methods: Pearson, Deviance and ML.*

we are going to make a comparison of estimation results among the three methods.

First of all, comparing the ϕ estimates for the different covers we see that the ϕ estimates for *TPL* are much larger than the other three risk covers, no matter which estimation method is used. The reason is that large claims are far more dominating in *TPL* than the other covers. Claims above 1 million SEK are quite common for *TPL* while they seldom occur for the other covers. Hence it is reasonable to expect that values of ϕ are larger in this case. *Partial kasko* is also worth noticing because although the ϕ estimates with ML and Deviance methods are close to those of *Mer* and *Hull*, it has an usually large estimate with the Pearson method. If we study furthermore the ϕ estimates for the subcover of *Partial kasko* we find that *Gls* is the main cause to the large value of the ϕ estimate in the whole group. *Gls* protects against the risk of damages to car windows and the value of car windows are not expected to vary significantly. Thus we suspect that there might exist abnormal observations or errors in the data.

The existence of outliers in the data can be examined by plotting residuals. There are several ways to define residuals but we choose the *Anscombe* residual (P.McCullagh and J.A.Nelder, 1997). An advantage of using anscombe residuals is that with some mathematical manipulations we can obtain the adjusted residual, which is normal-like distributed. The anscombe residual is given by the following expression for gamma distributed Y_i:

$$r_i = \frac{3(y^{1/3} - \mu_i^{1/3})}{\mu_i^{1/3}}$$

where r_i are identically distributed with mean zero if the assumption that Y_i is gamma distributed holds. Means as well as variances of r_i for the four covers are showed in the following table:

Risk cover	Mean	Std.Dev of residual	Skewness
TPL	-0,843	1,224	3,977
P.kasko	-0,423	1,008	1,996
Mer	-0,361	0,993	0,816
Hull	-0,305	0,927	0,602

Table 4: *Means, standard deviations and skewness of the anscombe residuals of the GLM for the average claims in terms of different covers.*

For *TPL*, the variance, as expected is larger than the other three covers. The residuals of *Partial kasko* although with a smaller variance than in the *TPL*, are quite skewly distributed. The skewness of the residuals indicates the existence of extreme values in both *TPL* and *Partial kasko*. We have also checked the distributions of residuals for all the covers and find that all of the distributions are a little skewed to the right, which implies that the GLM with gamma distribution does not fit perfectly to our data. However we still regard them as acceptable and leave the problem to the later sections, where we are going to study what happens to the estimators of ϕ when the distribution assumption is questionable.

We next compare the estimation results obtained with different methods within the different covers. It is shown clearly that the values of $\hat{\phi}$ are rather close for ML and Deviance. The Deviance estimates are a little bit larger than ML for all the covers. This results is in accordance with our theoretical analysis in section 2.1.3. However, the Pearson estimates take the largest value compared with the other two, probably due to the existence of extreme values.

The Pearson estimate for *TLP* deserves extra attention since it is abnormally larger than the deviance and the ML estimates. It possibly indicates the sensitivity of the Pearson estimator to large amounts. For the cover *Mer* however, all these three estimates are fairly close to each other. We may summarize the results as follows:

1. The covers *TLP* and *Partial kasko* represent two different phenomena. *TLP* has both large variance and skewness in the residuals. In practice the distribution of *TLP* claims has fat tails due to the large amounts in some claims. In this circumstance, the assumption of a gamma distribution makes less sense than for the other covers. For *Partial kasko* variance of the residuals is not very large but the skewness, caused by *outliers*, is large. To deal with this kind of data, it is important to find out the causes for the outliers. It turned out that in this case, the outliers were caused by defective recording procedures.

2. The Pearson estimates differ greatly from the estimates with the other two methods especially for the *TPL* and *Partial kasko*. It implies that the Pearson estimator of ϕ is sensitive to the variance of data and extreme values, while such values do not have much effect on ML and deviance estimators.

2.2.3 Sensitivity of estimates to extreme values

In the previous section, we found that estimation results of ϕ were quite dependent on which estimation method we employed, particularly for the covers where extremely large losses occur. This observation motivates us to study more closely how the ϕ estimates react to extreme values. In order to reduce the load of work, we choose the covers *TLP* as well as *Gls* as our study objects, where *Gls* is one of the subgroups of *Partial Kasko* and contributes most to the variation in the group.

There are mainly two alternatives when dealing with extreme values. One is truncation where the claim cost above a certain level is cut away. Truncation is often used to deal with the problem of large claims in non-life insurance. The other alternative is to delete all observations defined as 'extreme'. When the outliers are assumed to be caused by errors, it is natural to use the second method. After truncation or deletion we obtain a modified data set which will be used to fit GLMs and estimate the parameter ϕ. In order to study the sensitivity of $\hat{\phi}$ to the extreme values we are going to set several levels at which the data will be truncated or deleted so that we can observe how ϕ estimates change with different levels. We use truncation for risk cover *TLP* and deletion for *Gls*. The results are showed in table 5 and table 6 respectively:

Method	Original estimate	level1	level2	level3	level4
Pearson	34,054	10,194	2,953	1,782	0,640
Deviance	2,623	2,124	1,588	1,333	0,801
ML	2,054	1,711	1,326	1,136	0,718

Table 5: *Estimations of the parameter ϕ in the GLM for the cover TLP with truncations on four different levels (Unit: SEK): level 1 = 1 000 000; level 2 = 200 000; level 3 = 100 000; level 4 = 30 000.*

Method	Original estimate	level1	level2	level3	level4
Pearson	8.143	1,444	0,286	0,235	0,261
Deviance	0.384	0,352	0,335	0,322	0,309
ML	0.362	0,333	0,318	0,306	0,295

Table 6: *Estimations of the parameter ϕ in the GLM for the cover Gls with truncations on four different levels (Unit: SEK): level 1 = 500 000; level 2 = 50 000; level 3 = 4 686 (corresponding to 99% quantile) level 4 = 3 415 (corresponding to 95% quantile).*

From the table 5 and 6, we find that the Pearson estimates are quite sensitive to large values while the Deviance and ML estimates are much less affected by truncation or deletion. Besides, we also notice that after truncation or deletion of the extreme values on certain levels, the Pearson estimates decrease dramatically and become less than the values of the deviance estimate.

2.3 Simulation

From the previous section, we found that in practice, there can be large differences between the ϕ estimates. In this section, we are going to study the properties of the ϕ estimators when the underlying distribution is known and try to find out which estimator is the most reliable by a simulation study. The estimators will be compared in terms of unbiasedness and efficiency. In order to limit the study, we will only choose two risk covers: *TPL* and *Hull*.

Two problems will be of interest in this section. First we are going to study how the ϕ estimates behave when the assumption of a gamma distribution is true. The other interesting problem is what the empirical distributions of $\hat{\phi}$ look like when the response variable is no longer gamma distributed but the relationship $V(\mu_i) = \mu_i^2$ still holds. There are many non-EDM distributions which satisfy the condition above, we choose the Generalized Pareto distribution which has a fat tail. We simulate data out of this distribution and fit GLMs to it under the false assumption that Y is gamma distributed. We are interested in studying whether estimators are robust to variation of the Y distribution.

2.3.1 Simulations with Gamma Distributed Claims

We assume that the claims Y_i follow the Gamma distribution where $E(Y_i) = \mu_i$ and $Var(Y_i) = \phi \cdot \mu_i^2$. For the parameter μ_i, we use the estimated values for *TPL* and *Hull* from the previous sections. We choose the true value of ϕ also based on the earlier estimating results so as to make the simulated data as realistic as possible. Besides, the number of claims per tariff cell is chosen in accordance with the empirical data. We fitted GLMs to the simulated data with 10 000 repetitions, each of which contained the same number of observations as in the real data. We obtain the corresponding 10 000 ϕ estimates and so their empirical distribution. The mean of the empirical distribution will be compared to the true value of ϕ. Results of the simulations with the different estimation methods are presented in table 7 and 8:

Estimation method	Mean of $\hat{\phi}$	Std.Dev of $\hat{\phi}$
ML	1,9999	0,010
deviance	2,5426	0,014
Pearson	2,0015	0,020

Table 7: *Means and standard deviations of the three estimators of ϕ for TPL with gamma distributed average claims. The results are obtained by 10 000 simulations and the true value of ϕ is 2.*

Estimation method	Mean of $\hat{\phi}$	Std.Dev of $\hat{\phi}$
ML	1,0497	0,005
deviance	1,2194	0,006
Pearson	1,0496	0,006

Table 8: *Means and standard deviations of the three estimators of ϕ for Hull with gamma distributed average claims. The results are obtained by 10 000 simulations and the true value of ϕ is 1.05.*

For the Pearson and ML estimates, the means of the estimates are very close to the true value whilst the means of the deviance estimates deviates quite much from the true value for both covers. It is also shown that the standard deviations of all the estimates are quite small, indicating good efficiency of the estimators.

2.3.2 Simulations with Pareto Distributed Claims

In this section we are trying to answer at least partly the question of how the distributional assumption affects the estimation of ϕ with the three methods. In practice, one often uses a quadratic variance function, corresponding to a gamma distribution. However, even if a quadratic variance function may be realistic, the gamma assumption may not always be suitable, especially in the presence of extreme values. In many cases, distributions with fatter tails fit better the empirical data, e.g. the log-normal distribution or the generalized Pareto distribution etc. Neither of them belongs to the EDM family. Therefore we cannot apply these distributions directly in the frame of GLM. We have studied what happens to the ϕ estimates if we still assume gamma distribution though the real data follows a generalized Pareto distribution. It may be noted that McCullagh and Nelder (1997) recommends using the Pearson estimator in situations where the gamma assumption is in doubt.

We simulated data from a generalized Pareto distribution with the same true value of ϕ as before. Fitting GLMs to the data with the assumption of a gamma distribution, after 10 000 repetitions, we obtained the empirical distribution of the ϕ estimator. Table 9 and 10 show the means and variances of $\hat{\phi}$:

Estimation method	Mean of $\hat{\phi}$	Std.Dev of $\hat{\phi}$
ML	0,807	0,003
deviance	1,469	0,006
Pearson	1,989	0,069

Table 9: *Means and standard deviations of the three estimators of ϕ for TPL with generalized pareto distributed average claims. The results are obtained by 10 000 simulations and the true value of ϕ is 2.*

We see that only the means of the Pearson estimates are close to the true value. Both the ML and deviance estimators are biased. Another point worth noticing is that the variance of the Pearson estimator is much greater than the corresponding value when the response variable is gamma distributed.

Estimation method	Mean of $\hat{\phi}$	Std.Dev of $\hat{\phi}$
ML	0,982	0,005
deviance	1,179	0,007
Pearson	1,049	0,010

Table 10: *Means and standard deviations of the three estimators of ϕ for Hull with generalized pareto distributed average claims. The results are obtained by 10 000 simulations and the true value of ϕ is 1.05.*

3 Estimation of Dispersion parameters in GLMs with Random Effect

In practice, one is often encountered with the problem of multi-level factors. A multi-level factor has a large number of classes, many of which with insufficient data for reliable estimates. Car model is a typical example of a multi-level factor. Nelder & Verral (1997) and Ohlsson & Johansson (2003b) suggested that multi-level factors can be modelled as random effects in GLMs. These models include an additional dispersion parameter α and two different methods of estimating α will be introduced. An empirical study of the three methods is presented with the same data as we used in the previous section. The section is ended with a simulation study.

3.1 Introduction to GLMs with Random Effect

According to Ohlsson and Johansson (2003b), GLMs can be extended by introducing a random effect U in the way that $\mu_i(u_k) = E(Y_i|U_k = u_k) = \mu_i u_k$ for each level k of the random effect and $\mu_i = \gamma_0\gamma_{1(i)}\gamma_{2(i)}\cdots\gamma_{j(i)}$, the product of all the relativities. It is natural to assume that $E(U_k) = 1$ since U_k can be interpreted as a random deviation from the mean value given by other rating factors. This corresponds to assuming mean zero for the random effects in additive models. The main problem is how we find a suitable predictor for U_k. According to Ohlsson and Johansson (2003b), the optimal linear predictor can be expressed as $g(Y_i) = E(U_k|Y_i)$, which minimizes the mean square error. In the first part of this section, we are going to give a brief introduction to Jewell's Theorem, providing an idea of how to obtain $g(Y_i) = E(U_k|Y_i)$ when Y_i is an EDM and μ_i depends on a random effect. Then we will discuss so-called *Tweedie models*, where we can obtain an explicit expression for $E(U_k|Y_i)$.

3.1.1 Jewell's Theorem

We assume that conditioning on parameter Θ, the distribution of the response variable Y is an EDM with density:

$$f_{Y_i|\Theta}(y_i|\theta) = \exp\left(\frac{y_i\theta - b(\theta)}{\phi/w_i} + c(y_i, \phi, w_i)\right) \tag{11}$$

where we have suppressed the index k and defined $\mu(\Theta) = E(Y|\Theta)$. We assume also that:

Assumption [2]

- Θ_k are independent and identically distributed random variables for $k = 1, 2...K$.
- For $k = 1, 2...K$, the pairs (Y_{ik}, Θ_k) are independent.
- Conditional on Θ_k the random variables Y_{1k},Y_{2k},...,$Y_{I_k,k}$ are independent.

Given the distribution of the parameter Θ, which functions as prior distribution, we can get the posterior distribution $F_{\Theta|Y}(\theta|y)$ with the help of Bayes' theorem. In Jewell's theorem, it is assumed that distribution of Θ is of the form:

$$f_\Theta(\theta) = \exp\left(\frac{\theta\delta - b(\theta)}{1/\alpha} + d(\delta, \alpha)\right) \tag{12}$$

where:

$$\delta = E(b'(\Theta)) = E(\mu(\Theta)) \tag{13}$$

and

$$\alpha = \frac{E(b''(\Theta))}{Var(b'(\Theta))} = \frac{E(V(\mu(\Theta)))}{Var(\mu(\Theta))} \tag{14}$$

α is the additional dispersion parameter in the GLMs with random effect. Equation 14 shows how the dispersion parameter α is related to the variance of $\mu(\Theta)$. It follows from Bayes's theorem that:

$$f_{\Theta|Y}(\theta|y) \propto \exp\left(\frac{\theta\tilde{\delta} - b(\theta)}{1/\tilde{\alpha}}\right) \tag{15}$$

where $\tilde{\alpha} = \alpha + (w./\phi)$ and $\tilde{\delta} = E(\mu(\Theta)|Y) = (\alpha\delta + (w./\phi)\bar{y})/\tilde{\alpha}$. We see from the above that the posterior distribution belongs to the same family as the prior distribution of Θ. In such a case, we call the distribution a natural conjugate distribution of $Y_i|\Theta$.

3.1.2 Tweedie Models and Random Effects in GLMs

In Jewell's theorem, only the random effect is considered. In this section we are going to introduce a sub-family of EDM, for which it is possible to extend Jewell's theorem to the situation with fixed effects apart from the random effect.

An EDM is called a Tweedie model if the variance function can be written as:

$$V(\mu) = \mu^p \tag{16}$$

where p is a positive integer.

For the Tweedie models, under certain conditions (Ohlsson and Johansson, 2003b):

$$\alpha = \frac{E(\mu(\Theta_k)^p)}{Var(\mu(\Theta_k)}$$

where in the case of $p = 2$, α can expressed as:

$$\alpha = 1 + \frac{1}{Var(\mu(\Theta_k))} \tag{17}$$

This indicates that the dispersion parameter has a negative relationship with the variance of $\mu(\Theta_k)$.

In GLMs with a random effect, the mean can be expressed in the form:

$$\mu_i(U_k) = \mu_i U_k$$

where μ_i is given by the ordinary rating factors and U_k denotes the the random effect on k:th level. We denote the so called canonical link function by $h(\cdot)$. Based on the results in the previous section, we can obtain the following (Ohlsson and Johansson, 2003b): We define the response variable Y_{ik} and $E(Y_{ik}|U_k) = \mu_i U_k$. Conditioning on $U_k = u_k$, the Y_{ik} follow a Tweedie model with $1 \leq p \leq 2$ and $\Theta = h(U_k)$ follows the natural conjugate distribution where $\alpha > 0$ and $\delta > 0$, then the optimal predictor of U_k is of the form: $\hat{u}_k = E(U_k|Y_{ik})$, and is given by

$$\hat{u}_k = \frac{\sum_i w_{ik} y_{ik}/\mu_i^{p-1} + \phi\alpha}{\sum_i w_{ik}\mu_i^{2-p} + \phi\alpha} \tag{18}$$

3.2 Estimation of the α

We are interested in the estimation of the dispersion parameters in GLMs and now focus on the additional dispersion parameter α. In this section, two methods of estimating α will be presented, starting with maximum likelihood estimations in the marginal distribution of Y_{ik}. Then an alternative method proposed by Ohlsson and Johansson(2003b) is introduced. We assume that the response variable $Y_{ik}|U_k$ follows a Tweedie model.

3.2.1 ML Estimation of α

For $p = 1$ or $p = 2$ in the Tweedie model, it is possible to derive an explicit expression for the marginal density of Y. The ML estimator can then be obtained as the solution to the ML equations.

For p=2, i.e. $V(\mu) = \mu^2$, according to Ohlsson & Johansson (2002, unpublished), the ML equation is:

$$\frac{1}{K}\sum_k \left[\frac{\alpha+1}{\alpha} - \frac{\nu w_{.k} + \alpha + 1}{\nu w_{.k}\bar{y}_k^* + \alpha}\right] - \frac{1}{K}\sum_k \log\left[\frac{\nu w_{.k}\bar{y}_k^* + \alpha}{\alpha}\right]$$

$$+\frac{1}{K}\sum_k [\Psi(\nu w_{.k} + \alpha + 1) - \Psi(\alpha + 1)] = 0$$

where K denotes the total number of classes of the multi-level factor, $\nu = 1/\phi$ and $y_k* = y_{ik}/\mu_k$. Using the estimated α we can then predict the random effect U_k.

The estimation results of α depends partly on $\hat{\phi}$, as we can tell from the expression above. Thus the estimate of ϕ will affect the quality of the estimate of α when we use the ML estimator.

3.2.2 An Alternative Estimation Method

According to Ohlsson and Johansson (2003b), under the assumption of Tweedie models, one can prove that

$$\alpha = \frac{E(\mu(\Theta)^p)}{Var(\mu(\Theta))}$$

where $\Theta = h(U)$. We now consider separate estimation of $\sigma^2 = \phi E(\mu(\Theta)^p)$ and $\sigma_\Theta^2 = Var(\mu(\Theta))$, whose ratio is the term $\phi\alpha$. From equation 18 we see that with the $\widehat{\alpha\phi}$ we can predict U_k directly. In another words, we do not need to estimate α and ϕ separately to obtain $\hat{U}_k$.

An estimator of σ^2 can be obtained as follows (Ohlsson and Johansson, 2003b)

$$\hat{\sigma}_k^2 = \frac{1}{I_k - 1} \sum_i w_{ik} \mu_i^{2-p} \left(\frac{Y_{ik}}{\mu_i} - \bar{u}_k \right)^2$$

where I_k is the number of observations in class k and $\bar{u}_k$ is the weighted average of y_{ik}/μ_i:

$$\bar{u}_k = \frac{\sum_i (w_{ik}\mu_i^{2-p}) y_{ik}/\mu_i}{\sum_i w_{ik}\mu_i^{2-p}}$$

Weighing $\hat{\sigma}_k^2$ together with weights $I_k - 1$ we get the estimator as:

$$\hat{\sigma}^2 = \frac{\sum_k (I_k - 1)\hat{\sigma}_k^2}{\sum_k (I_k - 1)}$$

An estimator of σ_U^2 is given by by the following expression:

$$\hat{\sigma}_U^2 = \frac{\sum_k \sum_i w_{ik}\mu_i^{2-p}(\bar{u}_k - 1)^2 - K\hat{\sigma}^2}{\sum_k \sum_i w_{ik}\mu_i^{2-p}}$$

The estimator of $\alpha\phi$ can be expressed as: $\widehat{\alpha\phi} = \hat{\sigma}^2/\hat{\sigma}_U^2$. It can be shown that given μ_i, $\hat{\sigma}^2$ and $\hat{\sigma}_U^2$ are unbiased estimators of σ^2 and σ_U^2 respectively.

3.2.3 Estimation Algorithm

The estimation algorithm used for estimating U_k can summarized as the follows (Ohlsson and Johansson, 2003b):

0. Set the initial value of u_k as 1 for all k;
1. Run SAS Proc Genmod, with $\log(U_k)$ as offset variable, to get $\hat{\mu}_i$;
2. Estimate α using the $\hat{\mu}_i$ from step 1;
3. Compute $\hat{u}_k$ using the estimates of step 1 and step 2;
4. Repeat step 1-3 until convergence;

3.3 Application to Car Insurance

The application of GLMs to car insurance will be extended by including a random effect. The point of interest here is the estimation of α. We consider basically two estimation methods. The first is to estimate ϕ as before and α using the ML estimator in the marginal distribution of Y. The other method is to use the alternative estimator in section 3.2.2. The application to real insurance data will be complemented by a simulation study.

3.3.1 Empirical Study

We use the same data as before with the difference that we introduce a new variable *car model* as a multi-level factor U_k and choose different rating factors from the previous model. There are around 2000 different car models, i.e. U_k has about 2000 classes. The new model can be expressed as follows:

$$\mu_i(u_k) = E(Y_i|U_k = u_k) = \gamma_0\gamma_{1(i)}\gamma_{2(i)}...\gamma_{j(i)}u_k$$

where the parameter $\gamma_{j(i)}$ denotes the relativity for the i:th level with respect to the j:th rating factor.

A study is made for each of the four different covers. It is worth noticing that the ML estimator of α depends partly on the value of $\hat{\phi}$, which can be obtained with three different methods: ML, Pearson and deviance. We are going to use all of these methods to estimate ϕ and study how they affect the estimation of α. However, with the alternative methods of estimating U_k, we consider $\alpha\phi$ as a whole, which means that the results will be the same no matter which method we use to estimate ϕ. The estimation results for four covers are shown in the tables 11-14:

	Without U_k	With U_k–ML		With U_k–Alt
	$\hat{\phi}$	$\hat{\phi}$	$\widehat{\alpha\phi}$	$\widehat{\alpha\phi}$
ML	0,919	0,875	9,338	17,913
Deviance	1,051	0,997	11,797	17,913
Pearson	1,449	1,194	16,397	17,913

Table 11: *Estimation of $\alpha\phi$ for the cover Hull in the GLM with the random effect. The results are obtained by two methods, ML and the alternative method, where ML estimator produces three different estimates due to the different methods of estimating corresponding ϕ and is obtained by multiplying $\hat{\phi}$ with $\hat{\alpha} - ML$. The alternative estimator does not depend on $\hat{\phi}$.*

	Without U_k	With U_k–ML		With U_k–Alt
	$\hat{\phi}$	$\widehat{\phi}$	$\widehat{\alpha\phi}$	$\widehat{\alpha\phi}$
ML	1,055	1,003	15,054	29,435
Deviance	1,227	1,163	19,645	29,435
Pearson	2,002	1,647	37,687	29,435

Table 12: *Estimation of $\alpha\phi$ for the cover Mer in the GLM with random effect.*

First of all, we study the ML estimator of α. Since we have obtained three different estimates for each $\hat{\alpha}$ due to different methods of estimating ϕ, we are interested in finding out whether these three estimates get along with

	Without U_k	With U_k–ML		With U_k–Alt
	$\hat{\phi}$	$\hat{\phi}$	$\widehat{\alpha\phi}$	$\widehat{\alpha\phi}$
ML	1,193	1,075	3,971	22,524
Deviance	1,408	1,252	4,863	22,524
Pearson	3,942	5,778	22,774	22,524

Table 13: *Estimation of $\alpha\phi$ for the cover Partial kasko in the GLM with random effect where the extreme values have already been deleted.*

	Without U_k	With U_k–ML		With U_k–Alt
	$\hat{\phi}$	$\hat{\phi}$	$\widehat{\alpha\phi}$	$\widehat{\alpha\phi}$
ML	1,371	1,332	16,098	162,738
Deviance	1,649	1,605	29,710	162,738
Pearson	3,508	3,475	1062,704	162,738

Table 14: *Estimations of the $\alpha\phi$ in the GLM for the cover TPL with truncations on the level 200 000 SEK.*

each other. The tables 11-14 show that the $\hat{\alpha\phi}$ with Pearson $\hat{\phi}$ deviates quite much from the other two estimates and values of $\hat{\phi}_d$ and $\hat{\phi}_m$ are relatively close to each other. This can be seen even more clearly in the covers *Partial Kasko* and *TPL*.

Comparing estimation results between the ML and the alternative method, we see that the estimates of $\widehat{\alpha\phi}$ obtained by alternative method are more close to the ML-estimates with $\hat{\phi}_p$ for the covers *Hull*, *Mer* as well as *Partial kasko*. For *TPL*, the estimates with alternative method differs quite much from all the ML estimates. It seems that the *TPL* data are so skewed that random effect prediction is not possible in that case. Also with $\hat{\alpha\phi} = 163$ all $\hat{U}_k$ will be close to 1.

From section 2.2.2, we have learned that cover *TPL* has large variance and skewness. Besides $\hat{\phi}_p$ is very sensitive to extreme values. This might be an explanation for the strange estimation results for $\alpha\phi$ for *TPL*.

3.3.2 Simulation Study

In order to give more insight into the properties of the estimators of α, we performed a simulation study. The main purpose of simulations is to compare the estimated α with the true value, in which way we can get some idea of how well different estimation methods work. For the sake of simplicity, we choose to create a data set with only one fixed effect and one random effect. To begin with, we assume that

$$Y_i|U_k \sim Gamma(\mu_i U_k, \phi)$$

where $E(Y_i|U_k) = \mu_i U_k$. The corresponding natural conjugate distribution $f_\Theta(\theta)$ is of the form

$$f_\Theta(\theta) = \exp\left(\frac{\theta\delta + \log(-\theta)}{1/\alpha} + d\right)$$

where the canonical link function can be expressed as $\Theta = -1/U_k$. It can be shown that $U_k \sim InverseGamma(\alpha+1, \alpha)$ (for a proof see the Appendix A4). To generate the data Y_i, we take the following steps:

- Set the true values of the fixed factors μ_i (See table 15)
- Set the true values of $\phi = 2$ and $\alpha = 12$
- Simulate variable $U_k \sim InverseGamma(\alpha+1, \alpha)$
- Simulate variable $Y_i|U_k \sim Gamma(\mu_i U_k, \phi)$

We suppose that the fixed effect has only five levels, μ_i for $i = 1, 2, 3, 4, 5$ that take the following values:

We assume also that the multi-level factors has totally 1110 levels with the number of observations for each level given by the following table:

Thus we have 6 000 observations on each level i and totally 30 000 observations. We fit the extended GLMs with the random effect U_k into the

level i=	1	2	3	4	5
μ_i	1000	1240	1130	1080	1020

Table 15: *Values of μ_i on level 1, 2, 3, 4 and 5. The true values of μ_i are chosen based on the estimation results in the models without random effect in section 2.*

level k=	1-10	11-110	111-1110
n_{ik}	200	20	2

Table 16: *The number of observations for each k on the i:th level where $k = 1, 2, ...1100$ and $i = 1...5$.*

simulated data for 5 000 times and estimate the parameter α (or $\alpha\phi$) as well as the corresponding ϕ with the methods described in the previous sections. We then obtain the means and variances of estimates of both α (or $\alpha\phi$) and ϕ and then compare the means with the true values. The estimation results are showed in the tables 17 and 18:

ϕ estimation	Mean of $\hat{\phi}$	Std.Dev of $\hat{\phi}$	Mean of $\widehat{\alpha\phi}$	Std.Dev of $\widehat{\alpha\phi}$
ML	1,981	0,014	23,265	1,969
Deviance	2,532	0,020	45,092	4,314
Pearson	1,929	0,028	25,482	2,669

Table 17: *Means and standard deviations of ML estimator of $\widehat{\alpha\phi}$ with different methods of estimating the parameter ϕ in the GLM where the response variable average claim amount is gamma distributed. The true value of $\alpha\phi$ is 24.*

Comparing the three ML estimators we see that the best result is obtained when we use $\hat{\phi}_m$ where the mean value of $\widehat{\alpha\phi}$ is most close to the true value while the standard deviation is the smallest. With the alternative method, the mean value of the $\hat{\alpha\phi}$ is also around the true value but with a larger standard deviation. Therefore we can conclude that the ML estimation of α with $\hat{\phi}_m$ is the most preferable in this case.

In practice, the distribution of response variable is often skewly distributed, as the case for covers *Partial Kasko* and *TPL*. In the following, we are going

Mean of $\widehat{\alpha\phi}$	Std.Dev of $\widehat{\alpha\phi}$
26,630	2,866

Table 18: *Means and standard deviations of the alternative estimator of $\alpha\phi$ in the GLM with random effect where the response variable average claim is gamma distributed. The true value of $\alpha\phi$ is 24.*

to generate a data which follows the Generalized Pareto distribution though we still assume in the GLMs that it is gamma distributed. The random effect U_k is still inverse gamma distributed. We are going to compare all the estimators of α when the assumption of distribution is not proper. We choose the same true values of α and ϕ. The results are shown in the tables 19 and 20:

ϕ estimation	Mean of $\hat{\phi}$	Std.Dev of $\hat{\phi}$	Mean of $\widehat{\alpha\phi}$	Std.Dev of $\widehat{\alpha\phi}$
ML	1,209	0,009	11,502	0,937
Deviance	1,434	0,013	14,856	1,205
Pearson	1,790	0,078	21,820	4,363

Table 19: *Means and standard deviations of ML estimator of $\widehat{\alpha\phi}$ with different methods of estimating the parameter ϕ in the GLM where the response variable average claim amount is generalized pareto distributed. The true value of $\alpha\phi$ is 24.*

Mean of $\widehat{\alpha\phi}$	Std.Dev of $\widehat{\alpha\phi}$
22,175	4,458

Table 20: *Means and standard deviations of the alternative estimator of $\alpha\phi$ in the GLM with random effect where the response variable average claim is generalized pareto distributed. The true value of $\alpha\phi$ is 24.*

When the distribution assumption has been changed, ML estimate of α with $\hat{\phi}_m$ becomes much worse while ML estimate of α with $\hat{\phi}_p$ is still close to the alternative estimate and the true value.

4 Conclusion

In the GLMs without random effect, there is only one dispersion parameter ϕ. We have discussed three methods of estimating ϕ in two different situations. In the first case, where the GLM fits well to our empirical data, the ML estimator of ϕ turns out to be unbiased with 95% confidence. Both Pearson and Deviance estimators are biased. Meanwhile, the Deviance estimates are always larger than the ML estimates, which has been shown both theoretically and empirically. The Pearson estimator is pretty good in the sense that the value of the estimate is quite close to the true value. In the second case, when the distributional assumption is doubtful, we find that the ML estimator is no longer unbiased. However, the Pearson estimator is robust because the estimate does not change much even when the distribution assumption has been changed. Another finding is that the Pearson estimator is very sensitive to extreme values. When we delete or truncate the extreme values, the Pearson estimate decreases dramatically as it should. On the other side, the Deviance and ML estimates do not change much after deletion or truncation.

In the GLMs with random effect, there is an additional dispersion parameter, α. In this paper, two estimation methods, ML and the alternative method have been introduced to estimate the quantity $\alpha\phi$, which can be used directly to calculate the random effect U_k. The ML estimator is dependent of the ϕ estimate. It has been shown that when the distributional assumption is sensible, all of the ML estimators are biased with 95% confidence. However, the ML estimates with $\hat{\phi}_m$ and $\hat{\phi}_p$ are quite close to the true value while the ML estimates with $\hat{\phi}_d$ deviates much from the true value. After we change the assumption, the estimate with $\hat{\phi}_m$ becomes worse and its value is close to the one of ML estimates with $\hat{\phi}_d$. The ML estimates with $\hat{\phi}_p$ is affected comparatively less and the value of the estimate is still relatively close to the true value. This is in accordance with the results in the GLMs without random effects, where the $\hat{\phi}_p$ is the most robust. Comparing the ML with the alternative method, we find that except for the cover TPL, the value of the alternative estimate gets along with the ML estimates with $\hat{\phi}_p$ no matter whether the distributional assumption is proper or not. This is the case in both empirical and simulation studies. However, a disadvantage of these two estimators is that the standard deviations of the estimates are large.

When the value of $\hat{\alpha\phi}$ is large, the $\hat{u_k}$ is convergent to 1, which implies that the random effect does not affect the model substantially and even can be ignored. This could possibly explain the abnormal estimations of $\hat{\alpha\phi}$ for *TPL* where the random effect car model does not have much importance on the response variable, average claim amount. In such cases, neither ML nor the alternative estimators yield sensible estimations.

In summary, we recommend the Pearson estimator for estimating ϕ with the reason that it is the most robust against the distributional assumption since in practice it is often difficult to find a canonical distribution that fits the data well, particularly when there are large values. Concerning the α estimation in the model with a random effect, it turns out that the ML estimator with $\hat{\phi}_p$ and the alternative estimator are superior to the others, though both are biased, the estimates are robust and do not deviate much from the true value.

5 Reference

- P.McCullagh and J.A.Nelder (1997): "Generalized Linear Models."

- Esbjörn Ohlsson and Björn Johansson (2003a): "Prissättning inom sakförsäkring med Generaliserade linjära modeller."

- Esbjörn Ohlsson and Björn Johansson (2003b): "Credibility theory and GLM revised"

- Gauss M. Cordeiro and P.McCullagh (1991): "Bias Correction in Generalized Linear Models" Journal of the Royal Statistical Society. Series B, Volume 53, Issue 3 629-643.

- Nelder.J.A and Verrall, R.J. (1997): "Credibility Theory and Generalized linear models." ASTIN Bulletin, Vol 27:1, 71-82.

6 Appendix A

Appendix A1 : Re-parametrization of gamma distribution with GLM

The gamma distribution belongs to the GLMs and can be expressed in terms of two parameters μ and ϕ. Though it takes different look, it is completely consistent with the density function of gamma distribution in terms of α and β, which we are more familiar with. By some simple manipulations, we can reparametrize the density function by α and β:

$$f_Y(y_i) = exp\left(\frac{-y_i/\mu_i - \log(\mu_i)}{\phi/w_i} + \log(w_i y_i/\phi)w_i/\phi - \log(y_i) - \log\Gamma(w_i/\phi)\right)$$
$$= e^{-w\beta y} \cdot \left(\frac{\beta}{\alpha}\right)^{\alpha w} \cdot (\alpha w y)^{\alpha w} \cdot y^{-1} \cdot \frac{1}{\Gamma(\alpha w)}$$
$$= \frac{(w\beta)^{\alpha w}}{\Gamma(\alpha w)} \cdot y^{\alpha w - 1} \cdot e^{-\beta w y}$$

The last expression reminds us of the density function of gamma distribution:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where we have $\alpha' = \alpha w$ and $\beta' = \beta w$ in this case. The relation between these two parametrization can be expressed as $\mu = \alpha/\beta$ and $\phi = 1/\alpha$

Appendix A2 : Proof of the ML equation for the gamma distribution

$$l(\mu_i;\phi) = \sum_i^n \frac{w_i}{\phi}\log\left(\frac{w_i}{\phi\mu_i}\right) + \left(\frac{w_i}{\phi} - 1\right)\log(y) - \frac{w_i y_i}{\phi\mu_i} - \log\Gamma\left(\frac{w_i}{\phi}\right)$$

$$\Longrightarrow \frac{\partial l}{\partial\phi} = \sum_i^n -\frac{w_i}{\phi^2}\log\left((\frac{w_i}{\phi\mu_i}\right) - \frac{w_i}{\phi^2} - \frac{w_i}{\phi^2}\log(y) + \frac{w_i y_i}{\phi^2\mu_i} - \Psi\left(\frac{w_i}{\phi}\right)\left(-\frac{w_i}{\phi^2}\right)$$

Setting $\partial l/\partial\phi = 0$

$$\Longrightarrow 2\cdot\sum_i^n w_i\left(\log\left(\frac{w_i}{\phi}\right) - \Psi\left(\frac{w_i}{\phi}\right)\right) = 2\cdot\sum_i^n w_i\left(\log\left(\frac{\mu_i}{y_i}\right) + \frac{y_i-\mu_i}{\mu_i}\right)$$

Where the right side of the equation equals the deviance for the gamma distribution.

Appendix A3 : An Approximate Form of ϕ_d with Taylor Expansion

Taylor expansion:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Applying Taylor expansion to the function $f(y_i) = w_i \cdot \log(\hat{\mu}_i / y_i)$ with $x = y_i$ and $a = \hat{\mu}_i$

$$\begin{aligned}
f'(y_i) &= -\frac{w_i}{y_i} \\
f''(y_i) &= \frac{w_i}{y_i^2} \\
f^{(3)}(y_i) &= -\frac{2 \cdot w_i}{y_i^3}
\end{aligned}$$

$$\begin{aligned}
&\Longrightarrow f(y_i) = w_i \log(\hat{\mu}_i / y_i) \approx \frac{w_i}{2\mu_i^2}(y_i - \hat{\mu}_i)^2 - \frac{2\ w_i}{6\hat{\mu}_i^3}(y_i - \hat{\mu}_i)^3 \\
&\Longrightarrow \hat{\phi}_d = \frac{2}{n-r} \sum_i f(y_i) = \frac{2}{n-r} \sum_i w_i \log\left(\frac{\hat{\mu}_i}{y_i}\right) \\
&\approx \frac{w_i}{n-r} \cdot \log(\hat{\mu}_i / \hat{\mu}_i) - \frac{1}{n-r} \cdot \sum_i w_i \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right) \\
&+ \frac{1}{n-r} \cdot \sum_i w_i \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)^2 - \frac{1}{n-r} \cdot \sum_i \frac{2w_i}{3} \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)^3 \\
&= \frac{1}{n-r} \cdot \sum_i w_i \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)^2 - \frac{1}{n-r} \cdot \sum_i \frac{2w_i}{3} \left(\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}\right)^3
\end{aligned}$$

where we use the fact that $\sum_i w_i \cdot (y_i - \hat{\mu}_i)/(\hat{\mu}_i) = 0$

Appendix A4 : Proof of Inversely Gamma Distribution

If we have a variable X following a gamma distribution, the variable Y=1/X is inverse gamma distributed. In other words, to prove that U_k follows a inverse gamma distribution, we only need to prove $\Theta' = -\Theta = 1/U_k$ is gamma distributed.

We have density function of Θ with the form:

$$f_\Theta(\theta) = \exp\left(\frac{\theta + \log(-\theta)}{1/\alpha} + d\right)$$

$$F_{\Theta'}(\theta') = P(\Theta' \leq \theta') = P(-\Theta \leq \theta') = 1 - P(\Theta \leq -\theta') = 1 - F_\Theta(-\theta')$$

$$\Longrightarrow f_{\Theta'}(\theta') = f_\Theta(-\theta') \propto \exp\left(\frac{-\theta' + \log(\theta')}{1/\alpha}\right) = \theta'^{\alpha} e^{-\alpha\theta'}$$

This indicates that Θ' ~Gamma $(\alpha + 1, \alpha)$ and hence $U_k \sim$ InverseGamma $(\alpha + 1, \alpha)$.

7 Acknowledgement

I would like to extend my sincere gratitude and appreciation to many people who make this master thesis possible.

First of all, great thanks are due to my supervisors, Esbjörn Ohlsson and Björn Johansson, who have been providing me with a great many inspiring ideas and instructions on both content and structure of this thesis. Thanks to their participance, it has become a fully enjoyable learning process to write the thesis.

I am highly indebted to Actuary Department at Länsförskringar, Sak, whose assistance is vital for this thesis work. A sincere thank to the head of Actuary Department, Eva Blomberg for providing me such a great opportunity and also special thanks to Gunde Brandin and Fredrik Johansson who have been generously sharing knowledge and experience with me.

I would also like to acknowledge with great appreciation faculties of Mathematical Statistical Institution at Stockholm University. Thanks a lot for leading and guiding me the journey of exploration of the mathematical and statistical world, which is extraordinary interesting and challenging.

Last but not the least, I would like to send my most sincere gratitude to my family: mother Ming Meng, father Yan Hou, sister Nannan Lundin and my husband Yaohua Yu, who are always supportive to me.