Double generalized linear compound poisson models to insurance claims data

Andersen, Daniel Arnfeldt; Bonat, Wagner Hugo

Published in:
Electronic Journal of Applied Statistical Analysis

DOI:
10.1285/i20705948v10n2p384

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Document license
CC BY-NC-ND

Citation for pulished version (APA):
Andersen, D. A., & Bonat, W. H. (2017). Double generalized linear compound poisson models to insurance claims data. Electronic Journal of Applied Statistical Analysis, 10(2), 384-407. https://doi.org/10.1285/i20705948v10n2p384

Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v10n2p384

Double Generalized Linear Compound Poisson Models to insurance claims data
By Andersen, Bonat

Published: 14 October 2017

Electronic Journal of Applied Statistical Analysis
Vol. 10, Issue 02, October 2017, 384-407
DOI: 10.1285/i20705948v10n2p384

Double Generalized Linear Compound Poisson Models to insurance claims data

Daniel Arnfeldt Andersen*[a] and Wagner Hugo Bonat[b]

[a] *University of Southern Denmark, Department of Mathematics and Computer Science, Odense, Denmark.*
[b] *University of Southern Denmark, Department of Mathematics and Computer Science and Paraná Federal University, Department of Statistics Curitiba, PR, Brazil.*

Published: 14 October 2017

This paper describes the specification, estimation and comparison of double generalized linear compound Poisson models based on the likelihood paradigm. The models are motivated by insurance applications, where the distribution of the response variable is composed by a degenerate distribution at the origin and a continuous distribution on the positive real line. We present maximum likelihood and restricted maximum likelihood algorithms for parameter estimation, with emphasis to the analysis of insurance data. Simulation studies are employed to evaluate the bias and consistency of the estimators in a finite sample framework. The simulation studies are also used to validate the fitting algorithms and check the computational implementation. Furthermore, we investigate the impact of an unsuitable choice for the response variable distribution on both mean and dispersion parameter estimates. We provide `R` implementation and illustrate the application of double generalized linear compound Poisson models using a data set about car insurances.

1 Introduction

This paper discusses the statistical analysis of data sets with exact zeros and continuous values in the response variable, with emphasis to the analysis of insurance data. Consider

*Corresponding author: daniel.arnfeldt@outlook.dk

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

as an example the price for a car insurance, which depends on several factors, including the age of the car, model of the car, age of the driver, and other covariates. This paper seeks to predict how expensive a particular car insurance will be on average to correctly decide the price. In real life insurances there will often be some and perhaps even a lot of the considered insurances where no damage has been caused, which means that the insurance company has no expenses for those insurances.

Standard statistical methods based on the Gaussian or gamma distributions are not suitable for the analysis of data with both exact zeros and continuous values in the response variable. To this end, consider the following standard model for the total expenses Z also known as the total claim size

$$Z = \sum_{i=1}^{N} X_i,$$

where N is the number of claims that the company received, and X_i is the size of the ith claim, where $i = 1, 2, \ldots, N$. If no damage has been caused, the number of claims N will be zero, and thus we define $Z = 0$ in this case. We assume that N and X_i are Poisson and gamma distributed random variables, respectively. Thus, Z follows a compound Poisson distribution (Jørgensen, 1994; Jørgensen and Smyth, 2002). The compound Poisson distribution is a special case of the Tweedie distribution with power parameter $1 \leq p \leq 2$, as we shall discuss in the Section 2. The Tweedie distribution offers a flexible family of models to deal with non-negative highly right-skewed data and can handle continuous data with probability mass at zero (Jørgensen, 1997; Bonat and Kokonendji, 2017).

In this paper, we discuss the estimation of Z by using the double generalized linear models framework constructed based on the compound Poisson distribution. Following the lines of Jørgensen (1994) and Jørgensen and Smyth (2002) we describe the specification, estimation and comparison of double generalized linear compound Poisson models based on the likelihood paradigm. We present maximum likelihood and restricted maximum likelihood estimation algorithms. Furthermore, we evaluate the properties of the estimators in terms of bias and consistency through simulation studies. In contrast to previous works (Jørgensen, 1994; Jørgensen and Smyth, 2002) and computational implementations (Dunn, 2013; Zhang, 2013), the algorithm we shall present in Section 3 allows for a joint estimation of the regression, dispersion and power parameters. Moreover, we provide `R` implementation and illustrate the application of double generalized linear compound Poisson models using a data set about car insurance in Sweden.

When dealing with real life datasets N might not be available, which means that the estimation can only be done based on the total claim size. This paper also develops and implements methods to estimate the expenses both for N known and N unknown. We expect to obtain more precise results when the number of claims N is available, since having more information available generally provides more precise estimates. Furthermore, the models can handle different exposures w on the observations. A standard example of exposures in the context of insurance data is the policy-years.

Modelling strategies to the analysis of data with clumping at zero include the tobit

models which treat the zero outcomes as censored observations below some cut point (Van de Ven and van Praag, 1981; Amemiya, 1984). A disadvantage of this approach is that it relies on a normality assumption of the latent response, which in turn can be restrictive for high right-skewed data as often found in insurance applications. Another alternative is the two-part models that use two equations to separate the modelling into two stages. The first part is a binary model for the dichotomous event of having zero or positive values, for which the logistic model is a frequent choice. Conditional on a positive value, the second part assumes a continuous distribution, such as the log-normal, gamma or inverse Gaussian (Min and Agresti, 2002). This approach generates what is called of a zero-inflated model. For a comparison between compound Poisson and gamma inflated models see Lecomte et al. (2013). While quite flexible the two-part approach has the disadvantage to increase the model complexity by adding an additional linear predictor to describe the excess of zeros. In particular for insurance applications the two aforementioned approaches share the disadvantage that the number of claims N when available is not taken into account on the parameter estimation and consequently on the estimation of the total claim size.

Tweedie regression models have been recently extended in different ways to the analysis of insurance data. Qian et al. (2016) proposed Tweedie compound Poisson models with elastic net for covariate selection. Yang et al. (2017) discussed insurance premium prediction via gradient tree-boosted Tweedie compound Poisson models. Furthermore, Boucher and Danail (2011) highlighted the importance to model both the mean and dispersion structures for claims reserving. For a general discussion about insurance analysis with extra zeros see Yip and Yau (2005).

The next section presents an overview of the Tweedie distribution with emphasis to the construction of the density function for the special case of the compound Poisson distribution. Section 3 describes the estimation algorithms. In the Section 4 we check the algorithm implementation and the properties of the maximum likelihood and restricted maximum likelihood estimators through simulation studies. In the Section 5 we analyse a data set concerns car insurances. Finally, Section 6 discusses the main results and directions for future works. The data that is analysed and the programs that were used to analyse it can be obtained from
`http://www.leg.ufpr.br/doku.php/publications:papercompanions:dglmtw`.

2 Tweedie Family

In this section we shall discuss the construction of Tweedie models as described in details by Jørgensen (1997). Tweedie models are a special case of exponential dispersion models whose general form is given by

$$f(y;\theta,\phi) = a(y,\phi)\exp\{(y\theta - \kappa(\theta))/\phi\}, \tag{1}$$

where $\mu = \mathrm{E}(Y) = \kappa'(\theta)$ is the mean, $\phi > 0$ is the dispersion parameter, θ is the canonical parameter and $\kappa(\theta)$ is the cumulant function. The variance is given by $\mathrm{Var}(Y) = \phi \mathrm{V}(\mu)$

where $\mathrm{V}(\mu) = \kappa''(\theta)$ is called the variance function. Tweedie densities are characterized by power variance functions of the form $\mathrm{V}(\mu) = \mu^p$, where $p \in (-\infty, 0] \cup [1, \infty)$ is the index determining the distribution. Some important special cases of the Tweedie distribution are the Gaussian ($p = 0$), Poisson ($p = 1$), gamma ($p = 2$) and inverse Gaussian ($p = 3$) distributions. The function $a(y, \phi)$ is a normalized constant to be discussed below.

In this paper, we are particularly interested in the case where $1 < p < 2$ which corresponds to the compound Poisson distribution. We shall now obtain the form of the cumulant function associated with the compound Poisson distribution and discuss how to compute the normalized constant for the cases where N is known and unknown. To ease the notation, we define a new parameter α by

$$\alpha = \frac{p-2}{p-1} \quad \Longleftrightarrow \quad p = \frac{\alpha - 2}{\alpha - 1},$$

and we let $\alpha = -\infty$ corresponds to the case where $p = 1$. Another useful relation is that

$$(p-1)(\alpha - 1) = -1.$$

In order to obtain the exponential dispersion model corresponding to the power variance function, we should obtain the expression for the cumulant function $\kappa(\theta)$. Let $\tau(\mu) = \kappa'(\theta)$ denote the mean value mapping. Furthermore, note that

$$\begin{aligned} \tau^{-1}(\mu) &= \int \mu^{-p} \, \mathrm{d}\mu \\ &= \frac{1}{(1-p)} \mu^{1-p}, \qquad p \neq 1 \\ &= (\alpha - 1)\mu^{1/(\alpha-1)}, \end{aligned}$$

which in turn gives us that

$$\begin{aligned} \mu = \tau(\theta) &= \{\theta(1-p)\}^{\frac{1}{1-p}}, \qquad p \neq 1 \\ &= \left(\frac{\theta}{\alpha - 1} \right)^{\alpha - 1}. \end{aligned} \tag{2}$$

We can show from Eq. (2) that

$$\theta = \frac{\mu^{1-p}}{1-p}, \quad p \neq 1. \tag{3}$$

Finally, by combining Eq. (3) and the fact that $\kappa'(\theta) = \tau(\mu)$, gives us the following cumulant generating function,

$$\kappa(\theta) = \frac{\mu^{2-p}}{2-p}, \quad p \neq 2.$$

which in turn characterizes the Tweedie distribution for $1 < p < 2$.

We shall now discuss the construction of the probability function of the compound Poisson distribution for the case of N known. Let $N, X_1, X_2, \ldots$ be a sequence of independent random variables, with N being Poisson distributed and the X_i's being gamma and identically distributed. As described in the Section 1 the compound Poisson distribution is defined by

$$Z(w) = \sum_{i=1}^{N(w)} X_i,$$

where $Z = 0$ by definition if $N = 0$ and now we introduce the exposure denoted by w. Let $\tilde{n}$ be the realization of N to avoid confusion with the sample size, that we shall denote by n. Let $Y(w) = Z(w)/w$ be the claim rate per unit of exposure. The joint density function for $N(w)$ and $Y(w)$ is given by

$$f_{Y(w),N(w)}(y, \tilde{n}; \mu, \phi, p) = a(y, \tilde{n}; \phi) \exp\left(\frac{w}{\phi}\left[\frac{y\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right]\right), \tag{4}$$

where

$$a(y, \tilde{n}; \phi) = \begin{cases} \frac{\left\{\left(\frac{w}{\phi}\right)^{1-\alpha}\kappa(-1/y)\right\}^{\tilde{n}}}{\Gamma(-\tilde{n}\alpha)\tilde{n}!y} & \text{if } y > 0, \\ 1 & \text{if } y = 0. \end{cases}$$

For further details about the density function, see Jørgensen (1994, 1997) and Smyth (1996). The marginal density of $Y(w)$ can be found by summing out $N(w)$ in Eq.(4), which gives

$$f_{Y(w)}(y; \mu, \phi, p) = \sum_{\tilde{n}=1}^{\infty} a(y, \tilde{n}; \phi) \exp\left(\frac{w}{\phi}\left[\frac{y\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right]\right). \tag{5}$$

Jørgensen (1997) showed that for the general case, where $N(w)$ is unknown the normalized constant $a(y, \phi)$ in Eq. (1) is given for $y > 0$ by

$$a(y, \phi) = \frac{1}{y} W(y, \phi),$$

with $W(y, \phi, p) = \sum_{k=1}^{\infty} W_k$ and

$$W_k = \frac{y^{-k\delta}(p-1)^{\delta k}}{\phi^{k(1-\delta)}(2-p)^k k!\Gamma(-k\delta)},$$

where $\delta = (2-p)/(1-p)$. Dunn and Smyth (2005) presented detailed studies about this series and an algorithm to evaluate the Tweedie density function based on series expansions. The algorithm is implemented in the package `tweedie` (Dunn, 2013) for the statistical software `R`(R Core Team, 2016) through the function `dtweedie.series`.

For both cases $N(w)$ known or unknown the probability of exact zero is given by

$$P(Y(w) = 0) = \exp\left(-\frac{w\mu^{2-p}}{\phi(2-p)}\right).$$

The compound Poisson distribution provides a convenient and interpretable probability model to deal with insurance data. Since, the number of claims $N(w)$ is modelled by a Poisson distribution and the claim size X_i is modelled by a gamma distribution. Furthermore, the compound Poisson distribution can deal with exact zeros as often found in insurance applications.

3 Double Generalized Linear Models

Generalized linear models (GLMs) (Nelder and Wedderburn, 1972) assume a constant dispersion parameter. Double generalized linear models (DGLMs) are an extension of GLMs obtained by estimating the dispersion parameter separately for each observation, allowing for more flexibility in the models. In this section, we shall describe the double generalized linear models framework and how it is used for carrying out the parameter estimation. Estimation of the compound Poisson distribution using the DGLM framework, has previously been considered by Jørgensen and Smyth (2002), but assuming that the power parameter is known.

We assume a generalized linear model with a logarithmic link function for the mean cost by

$$g_\mu(\mu_i) = \log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta},$$

where $\mathbf{x_i}$ and $\boldsymbol{\beta}$ are $(q \times 1)$ vectors of known covariates and unknown regression parameters, respectively. We refer to this as the mean submodel. Simultaneously to the mean submodel, we assume a dispersion submodel denoted by

$$g_\phi(\phi_i) = \log(\phi_i) = \mathbf{z}_i^\top \boldsymbol{\gamma},$$

where $\mathbf{z_i}$ and $\boldsymbol{\gamma}$ are $(r \times 1)$ vectors of known covariates and unknown regression parameters, respectively. In this paper, the logarithmic link function was employed for the modelling of both mean and dispersion submodels. Further details about DGLMs can be found in Nelder and Pregibon (1987); Smyth (1989) and Smyth and Verbyla (1999). The estimation of the mean submodel is identical for unknown or known number of claims. Thus, we shall discuss the estimation of the mean submodel in this section, while the estimation of the dispersion submodels will be described in the next subsections.

In actuarial studies it is often convenient to use weights on the observations to improve estimation on valuable objects or objects with many policy-years. To this end, we consider the convolution formula for exponential dispersion models (Jørgensen, 1997).

Consider an additive exponential dispersion model $\mathrm{ED}^*(\theta, \lambda)$, and assume that $Z_1, \ldots, Z_n$ are independent and

$$Z_i \sim \mathrm{ED}^*(\theta, \lambda_i), \quad i = 1, \ldots, n.$$

Then the distribution of $Z_+ = Z_1 + \cdots + Z_n$ is

$$Z_+ \sim \mathrm{ED}^*(\theta, \lambda_1 + \ldots + \lambda_n). \tag{6}$$

This is the additive form of the convolution formula. The proof can be found in Jørgensen (1997). For the corresponding reproductive property, we assume that the random variables $Y_1, \ldots, Y_n$ are independent. Furthermore, we assume that

$$Y_i \sim \text{ED}\left(\mu, \frac{\phi}{w_i}\right), \qquad i = 1, \ldots, n, \tag{7}$$

where w is the exposure. If we let $w_+ = w_1 + \cdots + w_n$, the reproductive form of the convolution formula is

$$\frac{1}{w_+}\sum_{i=1}^{n} w_i Y_i \sim \text{ED}\left(\mu, \frac{\phi}{w_+}\right).$$

The result follows from the duality transformation, and a special case of Eq. (6) with $\lambda_i = w_i/\phi$ for $i = 1, \ldots, n$. The Eq. (7) is important, since it shows that the reproductive exponential models are closed when adding observations together that have the same claim rate, where the weights are given by the inverse dispersion parameters. This gives a very convenient model for actuarial studies, such that we can use the exposure w in the parameter estimation procedure.

From the convolution formula, see Jørgensen (1994) and using terminology and results from Nelder and Pregibon (1987), we have that the prior weights are given by

$$w_{\mu_i} = \frac{w_i}{\phi_i},$$

and recall that the power variance function is given by

$$\text{V}_p(\mu_i) = \mu_i^p.$$

Thus, the diagonal matrix of working weights $\boldsymbol{W}_\mu$ has the following form

$$\begin{aligned} \boldsymbol{W}_\mu &= \text{diag}\left\{\left(\frac{\partial g_\mu(\mu_i)}{\partial \mu}\right)^{-2} \frac{w_{\mu_i}}{\text{V}_p(\mu_i)}\right\} \\ &= \text{diag}\left\{\frac{w_i \mu_i^{2-p}}{\phi_i}\right\}. \end{aligned} \tag{8}$$

Furthermore, the working response vector $\mathbf{z}_\mu$ is such that its elements are given by

$$\begin{aligned} z_{\mu_i} &= \frac{\partial g_\mu(\mu_i)}{\partial \mu}(y_i - \mu_i) + g_\mu(\mu_i) \\ &= \frac{y_i - \mu_i}{\mu_i} + \log(\mu_i). \end{aligned}$$

Finally, the scoring update for the regression coefficients is given by

$$\boldsymbol{\beta}^{k+1} = (\boldsymbol{X}^\top \boldsymbol{W}_\mu \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W}_\mu \mathbf{z}_\mu, \tag{9}$$

where k is the number of iterations and $\boldsymbol{X} = (x_1^\top, \ldots, x_n^\top)^\top$ is an $n \times q$ design matrix. As initial values we use $y = \mu$. All terms on the right hand side of Eq. (9) are obtained from the previous iteration and note that the parameters in the dispersion structure $(\boldsymbol{\gamma}, p)$ are assumed known.

3.1 Estimation of dispersion parameters for unknown number of claims

For the case of unknown number of claims, we perform the estimation of the dispersion parameters based on the marginal density of $Y(w)$ presented in Eq. (5). The log-likelihood function is given by

$$\ell(\mu, \phi, p) = \sum_{i=1}^{n} \log f_{Y(w)}(y_i; \mu_i, \phi_i, p).$$

We can use the iterative weighted least squares algorithm to estimate both the mean and the dispersion regression parameters. Since, the Tweedie distribution is a member of the exponential dispersion family, we know that the mean and dispersion parameters are orthogonal, which implies that the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be estimated separately. For a detailed discussion about maximum likelihood estimation for Tweedie regression models, see Bonat and Kokonendji (2017).

From Eq. (5) the unit deviance is given by

$$\mathrm{d}(y_i; \mu_i) = 2w_i \left(\frac{y_i \mu_i^{1-p}}{1-p} - \frac{\mu_i^{2-p}}{2-p} \right).$$

To simplify the notation, we use d_i to denote the ith element of the unit deviance vector. We use the unit deviance as the response vector of the dispersion submodel, since it does not depend on ϕ_i. By using the saddlepoint approximation in the dispersion submodel, we can show that

$$\mathrm{d}_i \sim \phi_i \chi^2_1, \quad \phi_i \downarrow 0,$$

which means that the unit deviance approximately follows a gamma generalized linear model, see Smyth and Verbyla (1999). Moreover, a gamma model corresponds to a Tweedie model with $p = 2$, so the variance function for ϕ is

$$\mathrm{V}_p(\phi) = \phi^2.$$

Thus, the diagonal matrix of working weights $\boldsymbol{W}_\phi$ is given by

$$\begin{aligned} \boldsymbol{W}_\phi &= \mathrm{diag} \left\{ \left[\frac{\partial g_\phi(\phi_i)}{\partial \phi} \right]^{-2} \frac{1}{2\mathrm{V}_p(\phi_i)} \right\} \\ &= \frac{1}{2}. \end{aligned}$$

Consequently, the elements of the working response vector $\mathbf{z}_\phi$ are given by

$$\begin{aligned} z_{\phi_i} &= \frac{\partial g_\phi(\phi_i)}{\partial \phi}(\mathrm{d}_i - \phi_i) + g_\phi(\phi_i) \\ &= \frac{\mathrm{d}_i - \phi_i}{\phi_i} + \log(\phi_i). \end{aligned}$$

Finally, the updated value of $\boldsymbol{\gamma}$ is obtained by

$$\boldsymbol{\gamma}^{k+1} = (\boldsymbol{Z}^\top \boldsymbol{W}_\phi \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{W}_\phi \mathbf{z}_\phi. \tag{10}$$

All terms in Eq. (10) are from the previous iteration. Iterating between the mean submodel Eq. (9) and the dispersion submodel Eq.(10) until the increase in the log-likelihood is smaller than some desired threshold, gives estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for a fixed value of the power parameter.

It is well known that maximum likelihood estimators for dispersion parameters are biased downwards for small sample sizes. This can be corrected by restricted maximum likelihood (REML). We can apply this correction in double generalized linear models as well. Consider the hat matrix for the mean submodel

$$\boldsymbol{H} = \boldsymbol{W}_\mu^{1/2} \boldsymbol{X} \left(\boldsymbol{X}^\top \boldsymbol{W}_\mu \boldsymbol{X} \right)^{-1} \boldsymbol{X}^\top \boldsymbol{W}_\mu^{1/2},$$

and let h_i be the diagonal elements of $\boldsymbol{H}$, known as the leverage. Modifying the diagonal matrix of working weights by

$$\begin{aligned} \boldsymbol{W}_\phi^* &= \operatorname{diag} \left\{ \left[\frac{\partial g_\phi(\phi_i)}{\partial \phi} \right]^{-2} \frac{1 - \mathrm{h}_i}{2\mathrm{V}_p(\phi_i)} \right\} \\ &= \frac{1 - \mathrm{h}_i}{2}, \end{aligned}$$

and the elements of the working response vector $\mathbf{z}_\phi^*$ as

$$\begin{aligned} z_{\phi_i}^* &= \frac{\partial g_\phi(\phi_i)}{\partial \phi} \left[\mathrm{d}_i^* - \phi_i \right] + g_\phi(\phi_i) \\ &= \frac{\mathrm{d}_i^* - \phi_i}{\phi_i} + \log(\phi_i), \end{aligned}$$

with

$$\mathrm{d}_i^* = \frac{\mathrm{d}_i}{1 - \mathrm{h}_i},$$

gives approximately unbiased estimators for ϕ_i when

$$\boldsymbol{\gamma}^{k+1} = (\boldsymbol{Z}^\top \boldsymbol{W}_\phi^* \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{W}_\phi^* \mathbf{z}_\phi^*,$$

is used as the scoring update for $\boldsymbol{\gamma}$. Further information about the leverage adjustment can be found in Smyth et al. (2001) and Lee and Nelder (1998). The component $1 - \mathrm{h}_i$ is justified by the fact that an observation with leverage 1 provides no information about ϕ_i. The mean submodel does not change in the restricted maximum likelihood case and the model maximizes by alternating between the mean submodel and dispersion submodel exactly as before. Convergence is decided by the penalized profile log-likelihood

$$\ell^*(\mu, \phi, p) = \ell(\mu, \phi, p) + \frac{1}{2} \log |\boldsymbol{X}^\top \boldsymbol{W}_\mu \boldsymbol{X}|. \tag{11}$$

Finally, the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ is given by

$$\begin{pmatrix}\hat{\boldsymbol{\beta}}\\ \hat{\boldsymbol{\gamma}}\end{pmatrix} \sim \mathrm{N}\left\{\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}; \begin{pmatrix}\mathrm{I}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) & \mathbf{0}\\ \mathbf{0} & \mathrm{I}_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}})\end{pmatrix}^{-1}\right\},$$

where

$$\mathrm{I}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}^\top \boldsymbol{W}_\mu \boldsymbol{X}\right) \quad \text{and} \quad \mathrm{I}_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}}) = \left(\boldsymbol{Z}^\top \boldsymbol{W}_\phi \boldsymbol{X}\right)$$

are the Fisher information matrices for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. The asymptotic standard errors for the regression parameters for both mean and dispersion submodels are obtained by computing the square-root of the inverse of the diagonal terms of the Fisher information matrix. Since we do not have orthogonality between $\boldsymbol{\gamma}$ and p, we cannot directly invert the Fisher information to obtain the asymptotic variance for the power parameter. Consequently, the standard error for the power parameter is obtained, based on the profile likelihood function.

3.2 Estimation of dispersion parameters for known number of claims

When the number of claims $N(w)$ is available, the joint density function from Eq. (4) is used as the basis of the estimation. The log-likelihood function corresponding to the joint density is given by

$$\begin{aligned}\ell(\mu,\phi,p) =& \sum_{i=1}^{n} \log f_{Y(w),N(w)}(y_i,\tilde{n}_i;\mu_i,\phi_i,p)\\ =&(1-\alpha)\sum_{i=1}^{n}\tilde{n}_i\log(w_i/\phi_i) + \sum_{i=1}^{n} 1_{\{\tilde{n}_i\neq 0\}}\tilde{n}_i \log\kappa_p(-1/y_i)\\ &- \sum_{i=1}^{n} 1_{\{\tilde{n}_i\neq 0\}}\log\Gamma(-\tilde{n}_i\alpha)\\ &+ \frac{1}{\phi_i}\sum_{i=1}^{n} w_i\left\{\frac{y_i\mu_i^{1-p}}{1-p} - \frac{\mu_i^{2-p}}{2-p}\right\},\end{aligned} \tag{12}$$

We note by passing that y_i is sufficient for $\boldsymbol{\beta}$, hence the estimation of $\boldsymbol{\beta}$ is identical to the case of unknown number of claims, see Section 3.1. For the estimation of ϕ, we modify the working responses and working weights, since a joint density function does not fit into the framework of the generalized linear models. Notice that

$$\begin{aligned}\frac{\partial \log f}{\partial\phi} &= \frac{(\alpha-1)\tilde{n}}{\phi} - \frac{w}{\phi^2}\left[\frac{y\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right]\\ &= \frac{\tilde{n}}{(1-p)\phi} - \frac{w}{\phi^2}\left[\frac{y\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right],\end{aligned}$$

and

$$\frac{\partial^2 \log f}{\partial \phi^2} = \frac{\tilde{n}}{(p-1)\phi^2} + \frac{2w}{\phi^3}\left[\frac{y\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right].$$

We have that,

$$\mathrm{E}\left[\frac{y\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right] = \frac{\mu^{2-p}}{(1-p)(2-p)},$$

since $\mathrm{E}(Y) = \mu$, and we know from earlier that

$$\mathrm{E}(N(w)) = \frac{w\mu^{2-p}}{\phi(2-p)}.$$

The prior weights for the dispersion submodel are given by

$$w_\phi = \frac{2w\mu^{2-p}}{\phi(2-p)(p-1)},$$

gives us the Fisher information for ϕ as

$$\begin{aligned}
\mathrm{E}\left(-\frac{\partial^2 \log f}{\partial \phi^2}\right) &= -\left(\frac{w\mu^{2-p}}{\phi^3(2-p)(p-1)} + \frac{2w\mu^{2-p}}{\phi^3(1-p)(2-p)}\right) \\
&= -\left(\frac{w\mu^{2-p} - 2w\mu^{2-p}}{\phi^3(2-p)(p-1)}\right) \\
&= \frac{w\mu^{2-p}}{\phi^3(2-p)(p-1)} \\
&= \frac{w_\phi}{2\mathrm{V}_p(\phi)},
\end{aligned}$$

where $\mathrm{V}_p(\phi) = \phi^2$. By defining dispersion responses as

$$\begin{aligned}
\mathrm{d}(y;\mu) &= \frac{2\mathrm{V}_p(\phi)}{w_\phi}\frac{\partial \log f}{\partial \phi} + \phi \\
&= \frac{2}{w_\phi}\left(\frac{\tilde{n}\phi}{1-p} - w\left[\frac{y\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right]\right) + \phi.
\end{aligned}$$

We can now rewrite the first derivative of the log-likelihood as

$$\frac{\partial \ell(\beta,\gamma,p)}{\partial \phi} = \frac{w_\phi(\mathrm{d}-\phi)}{2\mathrm{V}_p(\phi)}. \tag{13}$$

By the definitions of the dispersion prior weights and the dispersion responses, we have transformed Eq.(13) so it fits into the concept of generalized linear models.

We have the diagonal matrix of working weights $\boldsymbol{W}_\phi$ as

$$\begin{aligned}
\boldsymbol{W}_\phi &= \mathrm{diag}\left\{\left[\frac{\partial g_\phi(\phi_i)}{\partial \phi}\right]^{-2}\frac{w_{\phi_i}}{2\mathrm{V}_p(\phi_i)}\right\} \\
&= \frac{w_{\phi_i}}{2},
\end{aligned}$$

where the division of 2 is to match the model developed in Section 3.1. We have the working response vector $\mathbf{z}_\phi$ with components

$$\begin{aligned} z_{\phi_i} &= \frac{\partial g_\phi(\phi_i)}{\partial \phi}[\mathrm{d}_i - \phi_i] + g_\phi(\phi_i) \\ &= \frac{\mathrm{d}_i - \phi_i}{\phi_i} + \log(\phi_i). \end{aligned}$$

The scoring update for $\boldsymbol{\gamma}$ is then found as

$$\boldsymbol{\gamma}^{k+1} = (\boldsymbol{Z}^\top \boldsymbol{W}_\phi \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{W}_\phi \mathbf{z}_\phi.$$

Alternating between the mean submodel and the dispersion submodel until the increase in the log-likelihood function (12) is smaller than a predefined threshold, gives the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for a fixed value of p.

When $N(w)$ is observed, we have more information available to estimation of ϕ, meaning there is less need for the restricted maximum likelihood adjustment of $\boldsymbol{\gamma}$. Nonetheless there might be cases where the adjustment is still useful. Since the estimation of $\boldsymbol{\beta}$ has not changed, we can use the same penalized profile log-likelihood as in (11), as the basis for the estimation. Adjusting the working weight matrix to

$$\begin{aligned} \boldsymbol{W}_\phi^* &= \mathrm{diag}\left\{ \left[\frac{\partial g_\phi(\phi_i)}{\partial \phi}\right]^{-2} \frac{\max(w_{\phi_i} - \mathrm{h}_i, 0)}{2\mathrm{V}_p(\phi_i)} \right\} \\ &= \frac{\max(w_{\phi_i} - \mathrm{h}_i, 0)}{2}, \end{aligned}$$

where h_i is the leverages as before. The elements of the working response vector change to

$$\begin{aligned} z_{\phi_i}^* &= \frac{\partial g_\phi(\phi_i)}{\partial \phi}\,[\mathrm{d}_i^* - \phi_i] + g_\phi(\phi_i) \\ &= \frac{\mathrm{d}_i^* - \phi_i}{\phi_i} + \log(\phi_i), \end{aligned}$$

where

$$\mathrm{d}_i^* = \frac{w_\phi \mathrm{d}_i}{w_\phi - \mathrm{h}_i},$$

and

$$\boldsymbol{\gamma}^{k+1} = (\boldsymbol{Z}^\top \boldsymbol{W}_\phi^* \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{W}_\phi^* \mathbf{z}_\phi^*,$$

is the new scoring iteration for g_ϕ. Standard errors of the parameters are found exactly as for $N(w)$ unknown.

The value of the power parameter for all aforementioned cases was considered fixed. However, in practice is important to estimate the power parameter. The easiest approach would be to estimate the model for some different predefined values of p and choice the best one, for example using the maximised value of the log-likelihood function. A better

approach is to use an optimization algorithm named Golden Section Search (Sauer, 2006) to estimate p. The algorithm is relatively slow, but it will always converges and it is still better than the naive approach of looping through a vector of predefined p values. Even though the algorithm in itself is rather slow, it is in fact faster than using standard algorithms as for example the Newton-Raphson algorithm. This is because the calculation of the derivatives is slow, making zero-order optimization algorithms better. The algorithm works on functions with one variable, where the function is continuous on the interval $[a, b]$, where we want to find the maximum. The algorithm is applied to the log-likelihood function ℓ, as a function of p. The maximum likelihood estimate for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are found for each value of p considered in a profile likelihood manner. The algorithms described in this paper were implemented in `R` and the code is available as a supplementary material. It is important to highlight that, although the flexibility of the `tweedie` package it does not allow to model the dispersion structure and a joint estimation of regression, dispersion and power parameters.

4 Simulation studies

In this section we present a simulation study that was conducted to check the properties of the maximum likelihood and restricted maximum likelihood estimators in a finite sample scenario. The simulations have also been used to validate the fitting algorithm and the computational implementation. We simulated 1500 data sets and evaluated the bias and consistency for the full set of parameters, i.e. $\boldsymbol{\beta}, \boldsymbol{\gamma}$, and p. We considered the cases of an unknown and known number of claims, combined with different values of the power parameter $p = (1.1, 1.3, 1.5, 1.7, 1.9)$. We considered samples of size 50, 100, 250, 500 and 1000. We have used one covariate in both the mean and dispersion structures and the data were simulated using the `tweedie` package Dunn (2013). The covariate values were generated from a standard Gaussian distribution. The regression parameters in the mean sub-model were fixed at $\beta_0 = 0.5$ and $\beta_1 = 1.5$. Similarly, the regression parameters in the dispersion sub-model were fixed at $\gamma_0 = -1$ and $\gamma_1 = 0.5$. These values were chosen in order to have enough mean and dispersion variation and also keep the mean small enough to have a substantial amount of exact zeros. The models were fitted by using the maximum likelihood and restricted maximum likelihood methods.

We start by considering the case of unknown number of claims. Figure 1 shows the average bias plus and minus the average standard errors by sample sizes, estimation methods and values of the power parameter.

The dispersion coefficients are slightly underestimated for small sample sizes, but for large sample sizes the simulation does not show any bias. The power parameter shows a small bias in the extreme cases, i.e. $p = 1.1$ and $p = 1.9$ for a sample of sizes 50 and 100. Simulations for the mean structure does not show any bias in any of the cases. The standard errors decrease while the sample size increases, as expected showing the consistency of the estimators. In general, the difference between maximum likelihood and restricted maximum likelihood estimators are of little magnitude. However, for

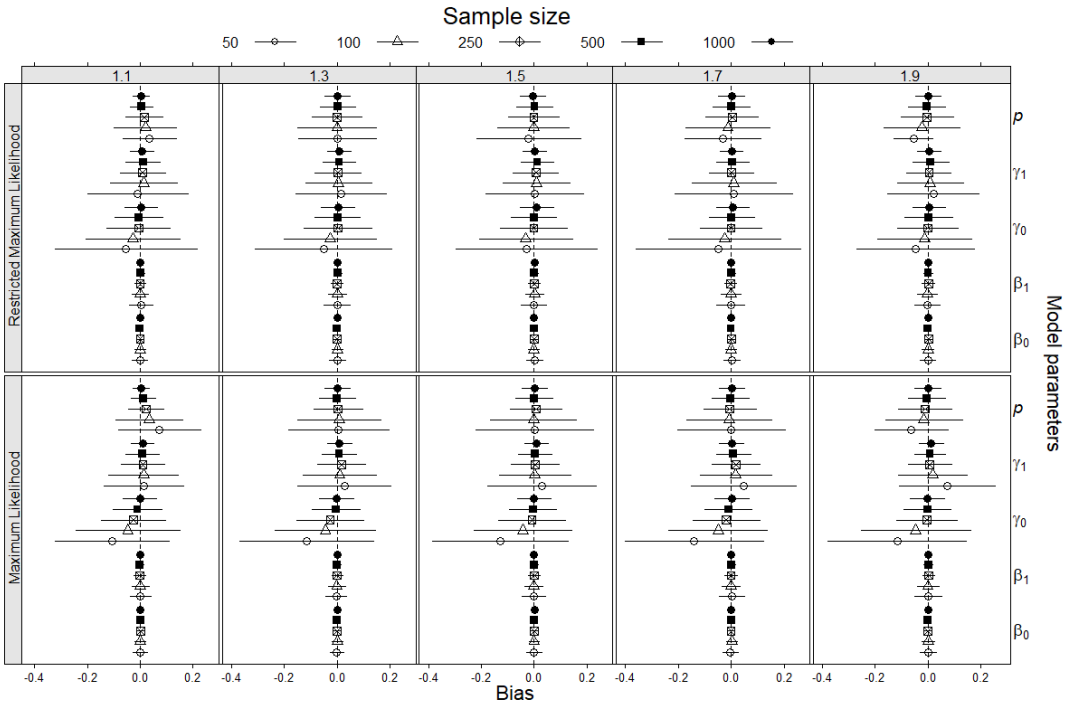

Figure 1: Average bias plus and minus average standard error for the parameters β_0, β_1, γ_0, γ_1, and p, by sample sizes, estimation methods and values of the power parameter - Case $N(w)$ unknown.

small sample sizes the REML estimators present slightly smaller bias than the MLE. The estimation procedure however, was slower in the REML case.

The results for the case of a known number of claims has been done in the same way and are illustrated in Figure 2.

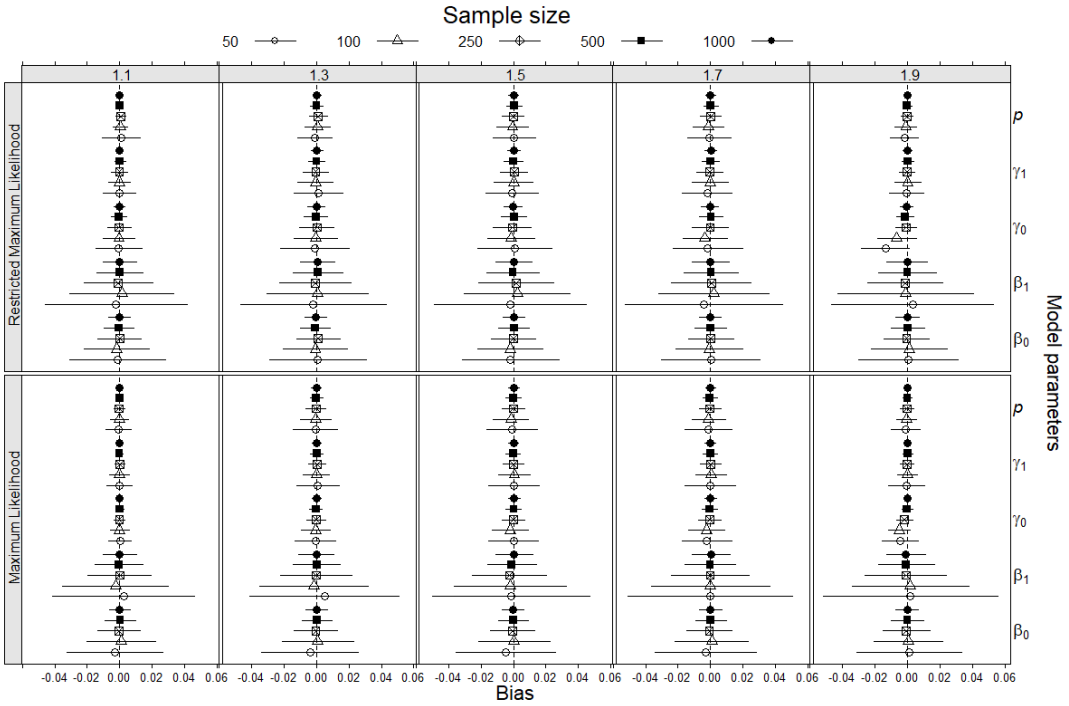

Figure 2: Average bias plus and minus average standard error for the parameters β_0, β_1, γ_0, γ_1, and p, by sample sizes, estimation methods and values of the power parameter - Case $N(w)$ known.

As expected, the standard errors of the parameters have decreased now when more information is available. At the same time, the estimation procedure now gives an unbiased result for the dispersion structure for $p = 1.1$, but shows the same sign of a slight bias for small sample sizes with $p = 1.9$ in the dispersion structure. The results do not show any bias for the power and regression parameters associated with the mean structure.

In order to investigate the impact of the power parameter in the fitted models Figure 3(A) illustrates the change in the parameters estimates for the dispersion sub-model for different values of the power parameter. Similarly, Figure 3(B) illustrates the change in the standard errors associated with the regression parameters ($\boldsymbol{\beta}$) for different values of the power parameter. In the supplementary material we present the Figures 1 and 2 using the x axis in a standardized scale. It provides a better visualizations of the estimators properties. On the other hand, by using an standardized x-axis we loose the perspective about how the

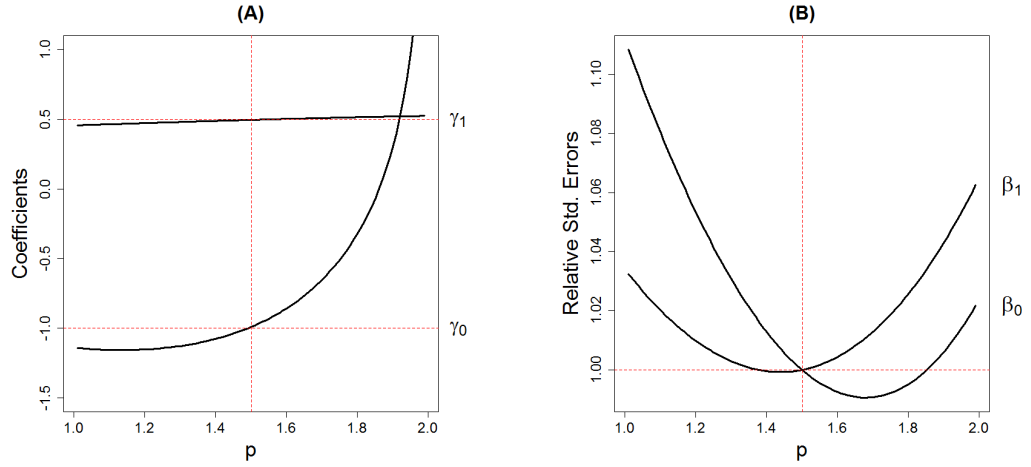

Figure 3: Illustration of the change in the parameters in the dispersion submodel for different values of p (A). Illustration of the change in the relative standard errors for $\boldsymbol{\beta}$ for different values of p (B). The red lines correspond to the true values.

The results presented in Figure 3 show that the value of p used, has a large impact on the dispersion parameters. In particular on the base dispersion, causing the standard errors of the parameter estimators in the mean structure to change a lot. Figure 3(B) should be interpreted such that if $p = 1.1$ were to be used, the estimated standard error of β_0 would be approximately 10% larger than the true value. It highlights the importance of a joint estimation of the regression, dispersion and power parameters when fitting double generalized linear compound Poisson models.

We also conducted another simulation study using 10 covariates in the mean and dispersion structures. The values of the covariates, regression and dispersion parameters were generated randomly from a standard Gaussian distribution. The results were really similar to the one covariate case presented in this Section. Thus, for a matter of space, we opted to include only the one covariate simulation study.

5 Analysis of Third Party Motor Insurance for Sweden, 1977

In this section we analyse the data set collected by Andrews (1985) which has previously been analysed by Jørgensen and Smyth (2002). In Sweden all risk classifications are standardized among all insurance companies, giving the opportunity to analyse them combined. In this data set we have 2182 observations, for which 385 are exact zeros. In order to compare the estimates and standard errors, we fitted the models by using the four approaches presented, i.e. maximum likelihood and restricted maximum likelihood combined with $N(w)$ known and unknown. Table 1 provides a description of the data

Table 1: Variable description - Third Party Motor Insurance for Sweden, 1997.

Variables	Description
Kilometres	Kilometres travelled per year.
	$1 : < 1000$.
	$2 : 1000 - 15000$.
	$3 : 15000 - 20000$.
	$4 : 20000 - 25000$.
	$5 : > 25000$.
Zone	Geographical zone.
	1 : Stockholm, Goteborg, Malmo with surroundings.
	2 : Other large cities with surroundings.
	3 : Smaller cities with surroundings in southern Sweden.
	4 : Rural areas in southern Sweden.
	5 : Smaller cities with surroundings in northern Sweden.
	6 : Rural areas in northern Sweden.
	7 : Gotland.
Bonus	No claims bonus. Equal to the number of years, plus one, since last claim.
Make	$1 - 8$ represent eight different common car models.
	All other models are combined in class 9.
Insured w	Number of insured in policy-years.
Claims $N(w)$	Number of claims.
Payment $Z(w)$	Total value of payments in Swedish krona.

set.

Make 4 is the Volkswagen 1200, which went out of production shortly after 1977. The other models are unknown, due to risk of impacting the sale of those cars. The exposure w will entering the estimation as weights. We start by comparing the estimates of the mean effects. To this end, we have combined Bonus level 5 and 6, and Kilometres 2 and 3, which does not cause any lose of goodness-of-fit. This has been verified by using the *Akaike* Information Criterion (AIC) and likelihood ratio test based on the method of a known number of claims. This reduction of factor levels has thus been applied for the data analysis. Table 2 shows the regression parameter estimates along with standard errors for the mean submodel for each method.

The estimates do not differ among the methods used. The value of p is somewhat similar. Increasing mileage does indeed increase the claim rate per unit of exposure, and that a higher bonus group decreases the claim rate per unit of exposure. Table 3 presents the estimates and standard errors associated with the dispersion parameters obtained by using the different approaches.

The change from maximum likelihood to restricted maximum likelihood has close to no effect on the value of p that maximizes the log-likelihood. The base dispersion is higher when $N(w)$ is not available. The standard errors in the mean structure are somewhat identical for the different methods - even though the dispersion coefficients are quite different. To explain this behaviour, recall the weight matrix for the mean submodel

Table 2: Mean submodel estimates for each method considered using the logarithmic link function. The numbers in the brackets denote the standard error.

Covariates	N(w) known		N(w) unknown	
	MLE	REML	MLE	REML
Base cost	6.543 (0.084)	6.549 (0.086)	6.544 (0.097)	6.538 (0.103)
Kilometres2-3	0.249 (0.064)	0.249 (0.064)	0.239 (0.069)	0.237 (0.073)
Kilometres4	0.336 (0.109)	0.337 (0.107)	0.300 (0.115)	0.291 (0.121)
Kilometres5	0.509 (0.120)	0.530 (0.122)	0.471 (0.119)	0.478 (0.124)
Bonus2	−0.309 (0.107)	−0.317 (0.113)	−0.313 (0.122)	−0.314 (0.127)
Bonus3	−0.378 (0.115)	−0.383 (0.114)	−0.390 (0.116)	−0.390 (0.121)
Bonus4	−0.693 (0.127)	−0.697 (0.133)	−0.655 (0.137)	−0.655 (0.142)
Bonus5-6	−0.872 (0.087)	−0.874 (0.089)	−0.898 (0.098)	−0.882 (0.102)
Bonus7	−1.316 (0.076)	−1.323 (0.078)	−1.316 (0.091)	−1.317 (0.096)
Make2	0.231 (0.088)	0.219 (0.088)	0.256 (0.086)	0.261 (0.091)
Make3	−0.041 (0.099)	−0.055 (0.095)	0.018 (0.130)	0.031 (0.137)
Make4	−0.623 (0.100)	−0.627 (0.102)	−0.617 (0.110)	−0.611 (0.115)
Make5	0.005 (0.093)	0.008 (0.095)	0.030 (0.108)	0.039 (0.114)
Make6	−0.379 (0.086)	−0.375 (0.086)	−0.376 (0.090)	−0.378 (0.095)
Make7	−0.264 (0.113)	−0.269 (0.116)	−0.213 (0.124)	−0.202 (0.130)
Make8	0.443 (0.167)	0.430 (0.167)	0.506 (0.195)	0.525 (0.201)
p	1.725 (0.016)	1.732 (0.015)	1.675 (0.050)	1.671 (0.047)

Table 3: Dispersion submodel estimates for each method considered using the logarithmic link function. The numbers in the brackets denote the standard error.

Covariates	N(w) known		N(w) unknown	
	MLE	REML	MLE	REML
Base dispersion	4.736 (0.038)	4.768 (0.011)	5.418 (0.359)	5.676 (0.255)
Kilometres2-3	−0.112 (0.028)	−0.068 (0.007)	0.096 (0.231)	0.063 (0.164)
Kilometres4	−0.211 (0.048)	−0.204 (0.008)	−0.129 (0.267)	−0.211 (0.190)
Kilometres5	−0.335 (0.053)	−0.227 (0.008)	−0.420 (0.267)	−0.482 (0.190)
Bonus2	0.493 (0.048)	0.507 (0.010)	0.241 (0.316)	0.183 (0.224)
Bonus3	0.618 (0.051)	0.448 (0.010)	−0.013 (0.316)	−0.046 (0.224)
Bonus4	0.788 (0.057)	0.795 (0.010)	0.417 (0.316)	0.392 (0.224)
Bonus5-6	0.864 (0.039)	0.804 (0.008)	0.436 (0.274)	0.385 (0.194)
Bonus7	1.203 (0.034)	1.079 (0.010)	0.995 (0.316)	1.059 (0.225)
Make2	−0.083 (0.039)	−0.141 (0.010)	−0.451 (0.338)	−0.535 (0.240)
Make3	0.188 (0.044)	0.039 (0.010)	0.568 (0.338)	0.345 (0.240)
Make4	0.398 (0.045)	0.415 (0.010)	0.328 (0.338)	0.188 (0.240)
Make5	−0.104 (0.041)	−0.074 (0.010)	−0.018 (0.338)	−0.111 (0.240)
Make6	0.303 (0.038)	0.283 (0.010)	0.094 (0.338)	−0.048 (0.240)
Make7	0.070 (0.050)	0.125 (0.010)	0.022 (0.338)	−0.098 (0.240)
Make8	0.023 (0.074)	−0.016 (0.010)	0.199 (0.338)	0.049 (0.240)

that we have found in Eq. (8):

$$\boldsymbol{W}_{\mu} = \text{diag}\left\{\frac{w_i \mu^{2-p}}{\phi_i}\right\}.$$

Differences in the standard errors for $\boldsymbol{\beta}$ among the different methods should be found in differences in the weight matrix. The reason that the values do not change much for the different methods, are explained by the fact that when the dispersion estimates are higher, the estimated value of p is lower, which means that the nominator and denominator relative to each other are the same among the different methods. When $N(w)$ is observed and more information is available, the standard error of p is lower as expected. The 95% confidence intervals obtained based on the profile likelihood for the Tweedie power parameter p are presented in Table 4.

Table 4: Confidence intervals (95%) for Tweedie power parameter, for each method considered.

Method	$\hat{p}_L$	$\hat{p}$	$\hat{p}_U$
$N(w)$ known ML	1.693	1.725	1.757
$N(w)$ known REML	1.701	1.732	1.762
$N(w)$ unknown ML	1.572	1.675	1.767
$N(w)$ unknown REML	1.575	1.671	1.761

The intervals are wider when $N(w)$ is not observed, which is an expected result. We also see that all methods have overlapping intervals, suggesting that the difference is not significant. Standard errors for the dispersion sub-models, are larger when $N(w)$ is unknown, which is expected since $N(w)$ provides a lot of information on the dispersion.

In terms of covariates effects the results in Table 2 show clearly that the three considered covariates are statistically significant for explaining the expected claim rate per unit of exposure. Increasing the kilometres travelled per year increase the claim rate per unit of exposure. For example, for kilometres $>$ 25000 we expect an increasing of 66.36% in the claim rate per unit of exposure. On the other hand, the covariate Bonus has a negative effect in the claim rate per unit of exposure. For the more extreme level of the Bonus covariate, we expect a decreasing of 3.72 times in the claim rate per unit of exposure in relation to the base cost. The covariate Make indicates that the claim rate per unit of exposure varies among the car models.

The dispersion sub-model can be interpreted in a similar way. The results in Table 3 show that the covariate Kilometres decreases the dispersion parameter estimates. This result is expected, since this covariate increases the expected value of the response variable. Thus, for the more extreme level of this covariate, we expect a decreasing of 39.79% in the dispersion parameter estimates. On the other hand, the number of bonus increases 3.33 times the dispersion parameter estimates. Again, it is expected since the covariate Bonus decreases the expected claim rate per unit of exposure. In general the

car model also changes the dispersion structure, however the effect is weaker than the Bonus and Kilometres covariates.

We have considered diagnostic measures on the restricted maximum likelihood methods, and the constant dispersion method. These are illustrated in Figure 4.

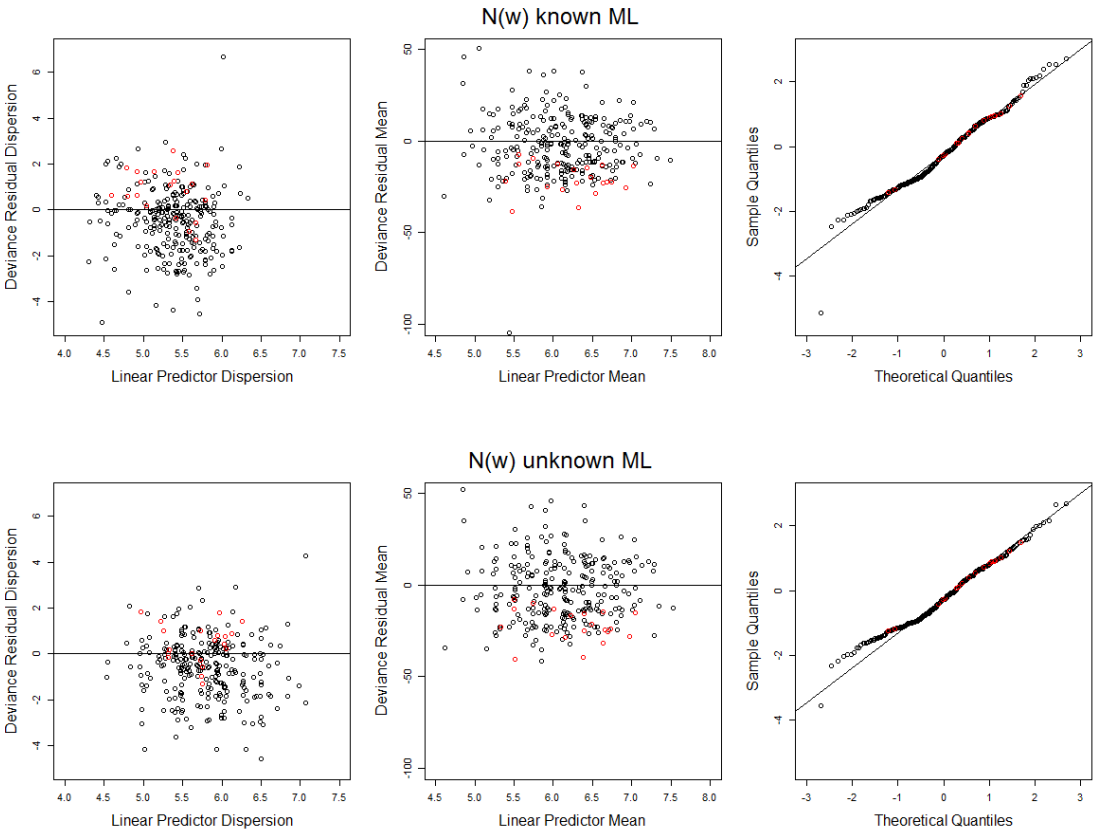

Figure 4: Deviance residual plots for the dispersion and mean submodel along with quantile plots. Red dots represents the exact zeros observations.

Both methods considered in the plots do not show any sign of significant outliers and we conclude that they are both providing a good fit to the data.

6 Discussion

We developed and implemented methods to handle the analysis of datasets with a combination between zeros and continuous values in the response variable, which is often the case for actuarial studies. One major issue we encountered was the derivation of the density function, when the number of claims $N(w)$ is unavailable, since it requires calculations of infinite sums. There have not been any findings of cases, where the number of

terms in the sum that contributes significantly to the sum is not a finite number. This number may however, be extremely large, and the terms that contribute significantly to the sum may also occur far away from index 1, making the approach of starting at 1, and continue adding terms until the terms are smaller than some threshold, very time dependent. We have used the method developed by Dunn and Smyth (2005) that starts by identifying which index number that corresponds to the largest term in the sum, and then continues to add terms on both sides of the maximum, until machine accuracy is reached.

Standard errors for the regression coefficients can be found in the usual way. The standard error of the Tweedie power parameter p have been found using profile likelihood. The confidence interval is in most cases close to symmetric around the optimal value of p, which makes the asymptotic approximation performs decently. The estimation of p is done by the Golden Section Search algorithm in double generalized linear models, which has the advantage of being derivative free, and thus, avoids to evaluate a large number of infinite sums.

Simulation studies verified that the methods are indeed asymptotically unbiased, but the sample size may need to be large for p close to 1. Simulation studies have also shown that estimation of the costs of the insurance claims does not change no matter what value of p is used. In fact, any distribution that is a member of the Tweedie family of distributions could be used to achieve unbiased estimators. Simulations have shown that the value of p used, may cause the standard errors of the parameters, for modelling the claim rate per unit of exposure to change, causing different results in hypothesis tests and discriminant analysis. We have found that the standard errors of the parameters can easily change by more than 10%, making the correct estimation of p important for actuarial studies, where the interest often lies in a particular group according to the insured item.

Data analysis has shown that all methods considered, provide similar estimators. However, if the number of claims is observed, one should always use the methods including this information, because of faster convergence and potentially (but not necessarily) more accurate results. We have also shown that including covariates in the estimation of the dispersion parameters may be necessary to achieve unbiased standard errors of the mean response, meaning that one should always start by fitting the model using covariates to model both the mean and dispersion structures.

The difference found between using maximum likelihood and restricted maximum likelihood, were small in its magnitude in either simulation studies and data analysis. To this end, we recommend to use the methods based on maximum likelihood, because they converge faster. However, we highlight that for small sample size as $n = 50$ the restricted maximum likelihood method presents a better performance in terms of bias and consistency than the maximum likelihood method, mainly for the case of $N(w)$ unknown.

The estimation methods presented in this paper, are all based on maximum likelihood. An extension of the methods developed in this paper, could be to develop methods using estimating functions Bonat and Kokonendji (2017); Jørgensen and Knudsen (2004). The standard error for the dispersion parameter ϕ in the constant dispersion method has

not been found, but could potentially be found using the Fisher information matrix. Similarly for the Tweedie power parameter p in the double generalized linear model methods, the standard error could also be found by the Fisher information matrix. In both cases one would have to be careful, since we do not have orthogonality between ϕ and p.

The methods have been implemented in the statistical software `R`, by taking advantage of the `tweedie` package (Dunn, 2013). The package provides a function to calculate the marginal density of $Y(w)$ and the `lm` function has also been used to perform the weighted least squares fit. The package `rootSolve` Soetaert (2015) has been used for the calculations of the approximate confidence interval for p. Generally, `R` also provides a lot of practical functions for statistical analysis.

References

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, 24(1):3–61.

Andrews, D.F.and Herzberg, A. M. (1985). *Data*. Springer, New York.

Bonat, W. H. and Kokonendji, C. C. (2017). Flexible Tweedie regression models for continuous data. *Journal of Statistical Computation and Simulation*, 87(11):2138–2152.

Boucher, J. and Danail, D. (2011). On the importance of dispersion modeling for claims reserving: An application with the Tweedie distribution. *Variance*, 5(2):158–172.

Dunn, P. K. (2013). *tweedie: Tweedie exponential family models*. R package version 2.1.7.

Dunn, P. K. and Smyth, G. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15(4):267–280.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall, London.

Jørgensen, B. and Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, 31(1):93–114.

Jørgensen, B. and Smyth, G. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *Astin Bulletin*, 32(1):143–157.

Jørgensen, B.and de Souza, M. (1994). Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal*, 1:69–93.

Lecomte, J. B., Benoit, H. P., Ancelet, S., Etienne, M., Bel, L., and Parent, E. (2013). Compound Poisson-gamma vs. delta-gamma to handle zero-inflated continuous data under a variable sampling volume. *Methods in Ecology and Evolution*, 4(12):1159–1166.

Lee, Y. and Nelder, J. A. (1998). Generalized linear models for the analysis of quality-improvement experiments. *The Canadian Journal of Statistics*, 26(1):95–105.

Min, Y. and Agresti, A. (2002). Modeling nonnegative data with clumping at zero: A survey. *Journal of Iranian Statistical Society*, 1(2):7–33.

Nelder, J. A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, 74(2):221–232.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A*, 135(3):370–384.

Qian, W., Yang, Y., and Zou, H. (2016). Tweedie's compound Poisson model with grouped elastic net. *Journal of Computational and Graphical Statistics*, 25(2):606–625.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sauer, T. (2006). *Numerical analysis*. Pearson, Boston.

Smyth, G. (1996). Regression analysis of quantity data with exact zeros. In *Proceedings of the Second Australia-Japan Workshop on Stochastic Models in Engineering, Technology and Management, Gold Coast, Australia*, pages 17–19.

Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(1):47–60.

Smyth, G. K., Huele, A. F., and Verbyla, A. P. (2001). Exact and approximate REML for heteroscedastic regression. *Statistical modelling*, 1(3):161–175.

Smyth, G. K. and Verbyla, A. P. (1999). Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics*, 10(6):695–709.

Soetaert, K. (2015). *rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations*. R package version 1.6.6.

Van de Ven, W. P. M. M. and van Praag, B. (1981). The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics*, 17(2):229–252.

Yang, Y., Qian, W., and Zou, H. (2017). Insurance premium prediction via gradient tree-boosted Tweedie compound Poisson models. *Journal of Business & Economic Statistics*, pages 1–15.

Yip, K. C. and Yau, K. K. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, 36(2):153–163.

Zhang, Y. (2013). Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Statistics and Computing*, 23(6):743–757.