Flexible Tweedie regression models for continuous data

Wagner Hugo Bonat & Célestin C. Kokonendji

View supplementary material

Published online: 23 Apr 2017.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Check for updates

Flexible Tweedie regression models for continuous data

Wagner Hugo Bonat [ID][a] and Célestin C. Kokonendji[b]

[a]Department of Statistics, Paraná Federal University, Curitiba, Brazil; [b]Laboratoire de Mathématiques de Besançon, Université Bourgogne Franche-Comté, Besançon, France

ABSTRACT

Tweedie regression models (TRMs) provide a flexible family of distributions to deal with non-negative right-skewed data and can handle continuous data with probability mass at zero. Estimation and inference of TRMs based on the maximum likelihood (ML) method are challenged by the presence of an infinity sum in the probability function and non-trivial restrictions on the power parameter space. In this paper, we propose two approaches for fitting TRMs, namely quasi-likelihood (QML) and pseudo-likelihood (PML). We discuss their asymptotic properties and perform simulation studies to compare our methods with the ML method. We show that the QML method provides asymptotically efficient estimation for regression parameters. Simulation studies showed that the QML and PML approaches present estimates, standard errors and coverage rates similar to the ML method. Furthermore, the second-moment assumptions required by the QML and PML methods enable us to extend the TRMs to the class of quasi-TRMs in Wedderburn's style. It allows to eliminate the non-trivial restriction on the power parameter space, and thus provides a flexible regression model to deal with continuous data. We provide an R implementation and illustrate the application of TRMs using three data sets.

1. Introduction

Statistical modelling is one of the most important areas of applied statistics with applications in many fields of scientific research, such as sociology, economy, ecology, agronomy, insurance and medicine, to cite but a few. There exists an infinity of statistical modelling frameworks, but the class of generalized linear models (GLMs) [1] is the most used in the last four decades. Special cases of the GLM class include the Gaussian, gamma, inverse Gaussian and Poisson regression models. These models are linked, since they belong to the class of exponential dispersion (ED) models and share the property to be described by their first two moments, mean and variance [2,3]. Furthermore, the variance function plays an important role in the context of ED models, since it describes the relationship between the mean and variance and characterizes the distribution [3].

Let Y denote the response variable and assume that the probability density function of Y belongs to the class of ED models. Furthermore, if we assume that $E(Y) = \mu$ and $\text{Var}(Y) = \phi V(\mu) = \phi \mu^p$ then $Y \sim \text{Tw}_p(\mu, \phi)$, where $\text{Tw}_p(\mu, \phi)$ denotes a Tweedie [3,4] random variable with mean μ and variance $\phi \mu^p$, such that $\phi > 0$ and $p \in (-\infty, 0] \cup [1, \infty)$ are the dispersion and power parameters, respectively. The support of the distribution depends on the value of the power parameter. For $p \geq 2, p = 1$, $1 < p < 2$ and $p = 0$, the support corresponds to the positive, count, non-negative and real values,

Supplemental data for this article can be accessed here. https://doi.org/10.1080/00949655.2017.1318876

respectively. In these cases $\mu \in \Omega$, where Ω is the steep convex support (i.e. the interior of the closed convex hull of the corresponding distribution support). Finally, for $p < 0$, the support corresponds to the real values, however the expectation μ is positive.

For practical data analysis, the Tweedie distribution is interesting, since it has the Gaussian ($p = 0$), Poisson ($p = 1$), non-central gamma ($p = 3/2$), gamma ($p = 2$) and inverse Gaussian ($p = 3$) distributions as special cases [2,3]. Another important case often applied in the context of insurance data [5,6] corresponds to the compound Poisson distribution, obtained when $1 < p < 2$. The compound Poisson distribution is a frequent choice for the modelling of non-negative data with probability mass at zero and highly right-skewed.

The power parameter plays an important role in the context of Tweedie models, since it is an index which distinguishes between some important continuous distributions. The algorithms we shall propose in Section 3 in contrast to current computational implementations of Tweedie regression models allow us to estimate the power parameter, which works as an automatic distribution selection. Although, the estimation of the regression parameters is less affected by the dispersion structure, the standard errors (SE) associated with the regression parameters are determined by the dispersion structure, which justifies dedicate attention to the estimation of the power and dispersion parameters.

The orthodox approach is based on the likelihood paradigm, which is an efficient estimation method. However, a particularity about the Tweedie distribution is that outside the special cases, its probability density has no closed-form expression and must be evaluated by numerical methods. Dunn and Smyth [7,8] proposed methods to evaluate the density function of the Tweedie distribution, but these methods are computationally demanding and show different levels of accuracy for different regions of the parameter space. Furthermore, the parameter space associated with the power parameter presents non-trivial restrictions and current software implementations [9] are restricted to deal with $p \geq 1$. These facts make the process of inference based on the likelihood paradigm difficult and sometimes slow.

The main goal of this paper is to propose alternative methods for estimation and inference of Tweedie regression models. In particular, we discuss the quasi-likelihood [10,11] and pseudo-likelihood [12] approaches. These methods are fast and computationally simple because they employ the first two moments, merely avoiding to evaluate the probability density function. Moreover, the second-moment assumptions required by the quasi- and pseudo-likelihood methods allow us to extend the Tweedie regression models to the class of quasi-Tweedie regression models in the style of Wedderburn [13]. The weaker assumptions of the second-moment specification eliminate the restrictions on the parameter space of the power parameter. Hence, it is possible to estimate negative and between zero and one values for the power parameter. In this way, we overcome the main restrictions of current software implementations and provide a flexible regression model to deal with continuous data.

Tweedie distributions are extensively used in statistical modelling, thereby motivating the study of their estimation in a more general framework. Applications include Vinogradov [14], Barndorff-Nielsen and Shephard [15] and Lee and Whitmore [16] who applied Tweedie distributions to describe the chaotic behaviour of stock price movements. Further applications include property and causality insurance, where Jørgensen and Paes De Souza [5] and Smyth and Jørgensen [6] fit the Tweedie family to auto-mobile insurance claims data. Tweedie distributions have also found applications in biology [17,18], fisheries research [19,20], genetics and medicine [21]. Chen and Tang [22] presented Bayesian semi-parametric models based on the reproductive form of ED models. Zhang [23] discussed the maximum likelihood and Bayesian estimation for Tweedie compound Poisson linear mixed models. For a recent application and further references, see Bonat and Jørgensen [10].

The rest of the paper is organized as follows. In the next section, we provide some background about Tweedie regression models. Section 3 discusses the approaches for estimation and inference. Section 4 presents the main results from our simulation study. Section 5 presents the application of Tweedie regression models to a data set concerning daily precipitation in Curitiba, Paraná State,

Brazil. This data set illustrates the analysis of positive continuous data with probability mass at zero. Section 6 reports some final remarks. In the supplementary material, we present an extra simulation study to show the flexibility of the Tweedie regression models to deal with heavy-tailed data as generated by the t-Student and slash distributions. Furthermore, we present two data analyses illustrating the flexibility of our model to analyse positive, highly right-skewed, as well as symmetric positive data, where current implementations have problems to deal with power parameter smaller than 1. The R implementation and some extra figures to illustrate the results of our simulation study are also presented in the supplementary material.

2. Tweedie regression models

The Tweedie distribution belongs to the class of ED models [2,3]. Thus, for a random variable Y which follows an ED, the density function can be written as

$$f_Y(y; \mu, \phi, p) = a(y, \phi, p) \exp\{(y\psi - \kappa(\psi))/\phi\},$$

where $\mu = \mathrm{E}(Y) = \kappa'(\psi)$ is the mean, $\phi > 0$ is the dispersion parameter, ψ is the canonical parameter and $\kappa(\psi)$ is the cumulant function. The function $a(y, \phi, p)$ cannot be written in a closed form apart of the Gaussian, Poisson, gamma and inverse Gaussian cases. The variance is given by $\mathrm{Var}(Y) = \phi V(\mu)$ where $V(\mu) = \kappa''(\psi)$ is called the variance function. Tweedie densities are characterized by power variance functions of the form $V(\mu) = \mu^p$, where $p \in (-\infty, 0] \cup [1, \infty)$ is the index determining the distribution. Although, Tweedie densities are not known in closed form, their cumulant generating function is simple and given by

$$K(t) = \{\kappa(\psi + \phi t) - \kappa(\psi)\}/\phi,$$

where $\kappa(\psi)$ is the cumulant function

$$\psi = \begin{cases} \dfrac{\mu^{1-p}}{1-p}, & p \neq 1, \\ \log \mu, & p = 1, \end{cases} \quad \text{and} \quad \kappa(\psi) = \begin{cases} \dfrac{\mu^{2-p}}{2-p}, & p \neq 2, \\ \log \mu, & p = 2. \end{cases}$$

The remaining factor in the density, $a(y, \phi, p)$ needs to be evaluated numerically. Jørgensen [3] presents two series expressions for evaluating the density, for $1 < p < 2$ and for $p > 2$. In the first case can be shown that

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

and for $y > 0$ that

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

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

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

where $\alpha = (2-p)/(1-p)$.

A similar series expansion exists for $p > 2$ and it is given by

$$a(y, \phi, p) = \frac{1}{\pi y} V(y, \phi, p),$$

with $V = \sum_{k=1}^{\infty} V_k$ and

$$V_k = \frac{\Gamma(1 + \alpha k)\phi^{k(\alpha-1)}(p-1)^{\alpha k}}{\Gamma(1+k)(p-2)^k y^{\alpha k}}(-1)^k \sin(-k\pi\alpha).$$

Dunn and Smyth [7] presented detailed studies about these series and an algorithm to evaluate the Tweedie density function based on series expansions. The algorithm is implemented in the package `tweedie` [9] for the statistical software R [24] through the function `dtweedie.series`. Dunn and Smyth [8] also studied two alternative methods to evaluate the density function of the Tweedie distributions, one based on the inversion of cumulant generating function using the Fourier inversion and the sanddlepoint approximation, for more details see Dunn [9]. In this paper, we used only the approach described in this section, i.e. based on series expansions.

We now turn to Tweedie regression models. Consider a cross-sectional data set, $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, where y_i's are i.i.d. realizations of Y_i according to $Y_i \sim \text{Tw}_p(\mu_i, \phi)$ and $g(\mu_i) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, where $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$ are $(Q \times 1)$ vectors of known covariates and unknown regression parameters, respectively. It is straightforward to see that $E(Y_i) = \mu_i = g^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ and the $\text{Var}(Y_i) = C_i = \phi\mu_i^p$. Hence, the model is equivalently specified by its joint distribution and by its first two moments. The Tweedie regression model is parametrized by $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top = (\phi = \exp(\delta), p)^\top)^\top$. Note that, we introduce the reparametrization $\phi = \exp(\delta)$ for computational convenience. Finally, in this paper we adopt the orthodox logarithm link function. It is a natural choice because the exponential function (inverse of logarithm) maps the linear predictor to the positive real values. Furthermore, the logarithm link function is popular in the context of Poisson, gamma and inverse Gaussian regression models. Since these models are special cases of the Tweedie regression model, it is sensible to use this link function.

3. Estimation and inference

This section is devoted to estimation and inference of Tweedie regression models. In what follows, we shall discuss the maximum likelihood, quasi-likelihood and pseudo-likelihood methods.

3.1. Maximum likelihood estimation

The maximum likelihood estimator (MLE) for the parameter vector $\boldsymbol{\theta}$ denoted by $\hat{\boldsymbol{\theta}}_M$ is obtained by maximizing the following log-likelihood function:

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\{a(y_i; \boldsymbol{\lambda})\} + \frac{1}{\exp(\delta)}(y_i\psi_i - \kappa(\psi_i)). \tag{1}$$

As we shall show below, the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are orthogonal. Hence, it is sensible to discuss each of them separately. The score function for the regression parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_Q)$ is given by

$$\mathcal{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \left(\frac{\partial \mathcal{L}(\boldsymbol{\theta})^\top}{\partial \beta_1}, \ldots, \frac{\partial \mathcal{L}(\boldsymbol{\theta})^\top}{\partial \beta_Q} \right)^\top,$$

where

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \psi_i} \frac{\partial \psi_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$$

$$= \sum_{i=1}^{n} \mu_i x_{ij} \left[\frac{1}{\exp(\delta)\mu_i^p} \right] (y_i - \mu_i) \quad \text{for } j = 1, \dots, Q.$$

The entry (j, k) of the $Q \times Q$ Fisher information matrix $\mathcal{F}_{\boldsymbol{\beta}}$ for the regression coefficients is given by

$$\mathcal{F}_{\beta_{jk}} = -\mathrm{E}\left\{ \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_k} \right\} = \sum_{i=1}^{n} \mu_i x_{ij} \left[\frac{1}{\exp(\delta)\mu_i^p} \right] \mu_i x_{ik}. \tag{2}$$

Similarly, the score function for the dispersion parameters $\boldsymbol{\lambda} = (\exp(\delta), p)$ is given by

$$\mathcal{U}_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}, \boldsymbol{\beta}) = \left(\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \delta}^{\top}, \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial p}^{\top} \right)^{\top},$$

whose components are given by

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \delta} = \sum_{i=1}^{n} \frac{\partial}{\partial \delta} \log a(y_i; \boldsymbol{\lambda}) - \frac{1}{\exp(\delta)} (y_i \psi_i - \kappa(\psi_i)) \tag{3}$$

and

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial p} = \sum_{i=1}^{n} \frac{\partial}{\partial p} \log a(y_i; \boldsymbol{\lambda}) + \frac{1}{\exp(\delta)} \left[y_i \frac{\partial \psi_i}{\partial p} - \frac{\partial \kappa(\psi_i)}{\partial p} \right]. \tag{4}$$

The entry (j, k) of the 2×2 Fisher information matrix $\mathcal{F}_{\boldsymbol{\lambda}}$ for the dispersion parameters is given by

$$\mathcal{F}_{\lambda_{jk}} = -\mathrm{E}\left\{ \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \lambda_j \partial \lambda_k} \right\}. \tag{5}$$

The derivatives in Equations (3)–(5) depend on the derivatives of the infinite sum $a(y_i; \boldsymbol{\lambda})$, and they cannot be expressed in closed form. Hence, numerical methods are required for approximating these derivatives. Let $\tilde{\mathcal{U}}_{\boldsymbol{\lambda}}$ and $\tilde{\mathcal{F}}_{\boldsymbol{\lambda}}$ denote the approximated score function and observed information matrix for the dispersion parameters, respectively. In this paper, we adopted the Richardson method [25], as implemented in the R package numDeriv [26] for computing these approximations. Furthermore, the cross-entries of the Fisher information matrix are given by

$$\mathcal{F}_{\beta_j \delta} = -\mathrm{E}\left\{ \frac{\partial \mathcal{U}_{\beta_j}(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \delta} \right\} = -\mathrm{E}\left\{ \mu_i x_{ij} \left[-\frac{1}{\exp(\delta)\mu_i^p} \right] (y_i - \mu_i) \right\} = 0$$

and

$$\mathcal{F}_{\beta_j p} = -\mathrm{E}\left\{ \frac{\partial \mathcal{U}_{\beta_j}(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial p} \right\} = -\mathrm{E}\left\{ \mu_i x_{ij} \left[\frac{\partial}{\partial p} \frac{1}{\exp(\delta)\mu_i^p} \right] (y_i - \mu_i) \right\} = 0.$$

Hence, the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are orthogonal. The joint Fisher information matrix for $\boldsymbol{\theta}$ is given by

$$\mathcal{F}_{\boldsymbol{\theta}} = \begin{pmatrix} \mathcal{F}_{\boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0} & \mathcal{F}_{\boldsymbol{\lambda}} \end{pmatrix},$$

whose entries are defined by Equations (2) and (5). Finally, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_M$ is $\mathrm{N}(\boldsymbol{\theta}, \mathcal{F}_{\boldsymbol{\theta}}^{-1})$, where $\mathcal{F}_{\boldsymbol{\theta}}^{-1}$ denotes the inverse of the Fisher information matrix. In practice the entry $\mathcal{F}_{\boldsymbol{\lambda}}$ is replaced by the approximation $\tilde{\mathcal{F}}_{\boldsymbol{\lambda}}$.

In order to solve the system of equations $\mathcal{U}_{\boldsymbol{\beta}} = \mathbf{0}$ and $\tilde{\mathcal{U}}_{\boldsymbol{\lambda}} = \mathbf{0}$, we employ the two-step Newton scoring algorithm, defined by

$$
\begin{aligned}
\boldsymbol{\beta}^{(i+1)} &= \boldsymbol{\beta}^{(i)} - \mathcal{F}_{\boldsymbol{\beta}}^{-1} \mathcal{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{(i)}, \boldsymbol{\lambda}^{(i)}), \\
\boldsymbol{\lambda}^{(i+1)} &= \boldsymbol{\lambda}^{(i)} - \tilde{\mathcal{F}}_{\boldsymbol{\lambda}}^{-1} \tilde{\mathcal{U}}_{\boldsymbol{\lambda}}(\boldsymbol{\beta}^{(i+1)}, \boldsymbol{\lambda}^{(i)}),
\end{aligned}
\tag{6}
$$

which in turn explicitly uses the orthogonality between $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$.

The numerical evaluation of the derivatives required in Equations (3)–(5) can be inaccurate, mainly for $p \approx 1$, i.e. the border of the parameter space. Thus, an alternative approach is to maximize directly the log-likelihood function in Equation (1) using a derivative-free algorithm as the Nelder–Mead method [27]. A more computationally efficient approach is to use the Nelder–Mead algorithm for maximizing only the profile log-likelihood for the dispersion parameters, which in turn is obtained by inserting the first equation of the two-step Newton scoring algorithm (6) in the log-likelihood function (1). Note that, by using this approach for each evaluation of the profile likelihood, we have a maximization problem for the regression parameters. We implemented these three approaches to obtain the MLE. The direct maximization of the log-likelihood function using the Nelder–Mead algorithm is slow, mainly for a large number of regression coefficients. The two-step Newton scoring algorithm presented many convergence problems for small values of the power parameter. Finally, the profile likelihood approach is the fast and stable implementation. However, the profile likelihood approach pointed out problems to compute the SE associated with the dispersion estimates for $p \approx 1$. In this paper, we used only the approach based on the profile log-likelihood, but we also provide R code for the other two approaches.

3.2. Quasi-likelihood estimation

We shall now introduce the quasi-likelihood estimation using terminology and results from Bonat and Jørgensen [10], Holst and Jørgensen [28] and Jørgensen and Knudsen [11]. The quasi-likelihood approach adopted in this paper combines the quasi-score and Pearson estimating functions for estimation of regression and dispersion parameters, respectively. The approach is also discussed in the context of estimating functions, see Liang and Zeger [29] and Jørgensen and Knudsen [11] for further details.

The quasi-score function for $\boldsymbol{\beta}$ has the following form:

$$
\mathcal{U}_{\boldsymbol{\beta}}^{q}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \left(\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_1} C_i^{-1} (y_i - \mu_i)^{\top}, \ldots, \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_Q} C_i^{-1} (y_i - \mu_i)^{\top} \right)^{\top},
$$

with $\partial \mu_i / \partial \beta_j = \mu_i x_{ij}$ for $j = 1, \ldots, Q$. The entry (j, k) of the $Q \times Q$ sensitivity matrix for $\mathcal{U}_{\boldsymbol{\beta}}^{q}$ is given by

$$
S_{\beta_{jk}} = E \left(\frac{\partial}{\partial \beta_k} \mathcal{U}_{\beta_j}^{q}(\boldsymbol{\beta}, \boldsymbol{\lambda}) \right) = - \sum_{i=1}^{n} \mu_i x_{ij} \left[\frac{1}{\exp(\delta) \mu_i^{p}} \right] x_{ik} \mu_i.
\tag{7}
$$

In a similar way, the entry (j, k) of the $Q \times Q$ variability matrix for $\mathcal{U}_{\boldsymbol{\beta}}^{q}$ is given by

$$
V_{\beta_{jk}} = \text{Var}(\mathcal{U}_{\boldsymbol{\beta}}^{q}(\boldsymbol{\beta}, \boldsymbol{\lambda})) = \sum_{i=1}^{n} \mu_i x_{ij} \left[\frac{1}{\exp(\delta) \mu_i^{p}} \right] x_{ik} \mu_i.
$$

Following Bonat and Jørgensen [10] and Jørgensen and Knudsen [11], the Pearson estimating function for the dispersion parameters has the form

$$\mathcal{U}_\lambda^q(\lambda, \boldsymbol{\beta}) = \left(\sum_{i=1}^n W_{i\delta}[(y_i - \mu_i)^2 - C_i]^\top, \sum_{i=1}^n W_{ip}[(y_i - \mu_i)^2 - C_i]^\top \right)^\top,$$

with $W_{i\delta} = -\partial C_i^{-1}/\partial \delta$ and $W_{ip} = -\partial C_i^{-1}/\partial p$. The Pearson estimating functions are unbiased estimating functions for λ based on the squared residuals $(y_i - \mu_i)^2$ with mean C_i. It is equivalent to treating the squared residual as a gamma variable, which is hence close in spirit to Perry's gamma regression method [30,31].

We shall now calculate the sensitivity matrix for the dispersion parameters. The entry (j, k) of the 2×2 sensitivity matrix is given by

$$S_{\lambda_{jk}} = \mathrm{E}\left(\frac{\partial}{\partial \lambda_k} \mathcal{U}_{\lambda_j}^q(\lambda, \boldsymbol{\beta}) \right) = -\sum_{i=1}^n W_{i\lambda_j} C_i W_{i\lambda_k} C_i,$$

where λ_1 and λ_2 denote either δ or p, giving

$$S_\lambda = \begin{pmatrix} -n & -\sum_{i=1}^n \log(\mu_i) \\ -\sum_{i=1}^n \log(\mu_i) & -\sum_{i=1}^n \log(\mu_i)^2 \end{pmatrix}. \tag{8}$$

Similarly, the cross-entries of the sensitivity matrix are given by

$$S_{\beta_j \lambda_k} = \mathrm{E}\left(\frac{\partial}{\partial \lambda_k} \mathcal{U}_{\beta_j}^q(\boldsymbol{\beta}, \lambda) \right) = 0 \tag{9}$$

and

$$S_{\lambda_j \beta_k} = \mathrm{E}\left(\frac{\partial}{\partial \beta_k} \mathcal{U}_{\lambda_j}^q(\lambda, \boldsymbol{\beta}) \right) = -\sum_{i=1}^n W_{i\lambda_j} C_i W_{i\beta_k} C_i, \tag{10}$$

with $W_{i\beta_k} = -\partial C_i^{-1}/\partial \beta_k$. Finally, the joint sensitivity matrix for the parameter vector $\boldsymbol{\theta}$ is given by

$$S_{\boldsymbol{\theta}} = \begin{pmatrix} S_{\boldsymbol{\beta}} & \mathbf{0} \\ S_{\lambda \boldsymbol{\beta}} & S_\lambda \end{pmatrix},$$

whose entries are defined by Equations (7)–(10).

We shall now calculate the asymptotic variance of the quasi-likelihood estimators denoted by $\hat{\boldsymbol{\theta}}_{QL}$, as obtained from the inverse Godambe information matrix, whose general form is $J_{\boldsymbol{\theta}}^{-1} = S_{\boldsymbol{\theta}}^{-1} V_{\boldsymbol{\theta}} S_{\boldsymbol{\theta}}^{-\top}$ for a vector of parameter $\boldsymbol{\theta}$, where $-\top$ denotes inverse transpose. The variability matrix for $\boldsymbol{\theta}$ has the form

$$V_{\boldsymbol{\theta}} = \begin{pmatrix} V_{\boldsymbol{\beta}} & V_{\boldsymbol{\beta}\lambda} \\ V_{\lambda\boldsymbol{\beta}} & V_\lambda \end{pmatrix}, \tag{11}$$

whereas $V_{\lambda\boldsymbol{\beta}} = V_{\boldsymbol{\beta}\lambda}^\top$ and V_λ depend on the third and fourth moments of Y_i, respectively. In order to avoid this dependence on high-order moments, we propose to use the empirical versions of V_λ and $V_{\lambda\boldsymbol{\beta}}$, whose entries are given by

$$\tilde{V}_{\lambda_{jk}} = \sum_{i=1}^n \mathcal{U}_{\lambda_j}^q(\lambda, \boldsymbol{\beta})_i \mathcal{U}_{\lambda_k}^q(\lambda, \boldsymbol{\beta})_i \quad \text{and} \quad \tilde{V}_{\lambda_j \beta_k} = \sum_{i=1}^n \mathcal{U}_{\lambda_j}^q(\lambda, \boldsymbol{\beta})_i \mathcal{U}_{\beta_k}^q(\lambda, \boldsymbol{\beta})_i.$$

Finally, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{QL}$ is $\mathrm{N}(\boldsymbol{\theta}, J_{\boldsymbol{\theta}}^{-1})$.

We may show by using standard results for inverse of partitioned matrix that

$$J_\theta^{-1} = \begin{pmatrix} S_\beta^{-1} V_\beta S_\beta^{-1} & S_\beta^{-1}(-V_\lambda S_\beta^{-1} S_{\lambda\beta}^\top + V_{\lambda\beta}^\top) S_\lambda^{-1} \\ S_\lambda^{-1}(-S_{\lambda\beta} S_\beta^{-1} V_\beta + V_{\lambda\beta}) S_\beta^{-1} & S_\lambda^{-1}(L + V_\lambda) S_\lambda^{-1} \end{pmatrix},$$

where $L = S_{\lambda\beta} S_\beta^{-1}(V_\beta S_\beta^{-1} S_{\lambda\beta}^\top - V_{\lambda\beta}^\top) - V_{\lambda\beta} S_\beta^{-1} S_{\lambda\beta}^\top$.

Moreover, note that $S_\beta^{-1} V_\beta S_\beta^{-1} = V_\beta^{-1}$, it shows that for known dispersion parameters, the asymptotic variance of the quasi-likelihood regression estimators reaches the Cramér-Rao lower bound, which in turn shows that the quasi-likelihood approach provides asymptotically efficient estimators for the regression coefficients.

Jørgensen and Knudsen [11] proposed the modified chaser algorithm to solve the system of equations $\mathcal{U}_\beta^q = 0$ and $\mathcal{U}_\lambda^q = 0$, defined by

$$\beta^{(i+1)} = \beta^{(i)} - S_\beta^{-1} \mathcal{U}_\beta^q(\beta^{(i)}, \lambda^{(i)}),$$

$$\lambda^{(i+1)} = \lambda^{(i)} - S_\lambda^{-1} \mathcal{U}_\lambda^q(\beta^{(i+1)}, \lambda^{(i)}).$$

The modified chaser algorithm uses the insensitivity property (9), which allows us to use two separate equations to update β and λ.

3.3. Pseudo-likelihood estimation

We shall now present the pseudo-likelihood approach using terminology and results from Gourieroux et al. [12]. The pseudo-likelihood approach considers the properties of estimators obtained by maximizing a likelihood function associated with a family of probability distributions, which does not necessarily contain the true distribution. In particular, in this paper, for estimation of Tweedie regression models, we adopted the Gaussian pseudo-likelihood, whose logarithm is given by

$$\mathcal{L}^p(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n\delta}{2} - \frac{p}{2}\sum_{i=1}^{n}\left(\log\mu_i - \frac{(y_i - \mu_i)^2}{2\exp(\delta)\mu_i^p}\right). \tag{12}$$

The pseudo-score function for θ is given by

$$\mathcal{U}_\theta^p(\beta, \lambda) = \left(\frac{\partial\mathcal{L}^p(\theta)}{\partial\beta_0}^\top, \ldots, \frac{\partial\mathcal{L}^p(\theta)}{\partial\beta_Q}^\top, \frac{\partial\mathcal{L}^p(\theta)}{\partial\delta}^\top, \frac{\partial\mathcal{L}^p(\theta)}{\partial p}^\top\right)^\top,$$

whose components have the following form:

$$\frac{\partial\mathcal{L}^p(\theta)}{\partial\beta_j} = -\frac{p}{2}\sum_{i=1}^{n}x_{ij} + \sum_{i=1}^{n}\frac{p(y_i - \mu_i)^2}{2\exp(\delta)\mu_i^p}x_{ij} + \sum_{i=1}^{n}\frac{(y_i - \mu_i)}{\exp(\delta)\mu_i^{p-1}}x_{ij}, \tag{13}$$

$$\frac{\partial\mathcal{L}^p(\theta)}{\partial\delta} = -\frac{n}{2} + \frac{1}{2\exp(\delta)}\sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{\mu_i^p} \tag{14}$$

and

$$\frac{\partial\mathcal{L}^p(\theta)}{\partial p} = -\frac{1}{2}\sum_{i=1}^{n}\log(\mu_i) + \frac{1}{2\exp(\delta)}\sum_{i=1}^{n}\frac{\log(\mu_i)}{\mu_i^p}(y_i - \mu_i)^2. \tag{15}$$

We note in passing that Equation (13) is an unbiased estimating function for β_j based on the linear and squared residuals. Similarly, note that Equations (14) and (15) are unbiased estimating functions for δ and p based on the squared residuals.

Gourieroux et al. [12] showed under classical assumptions that the pseudo-likelihood estimator denoted by $\hat{\boldsymbol{\theta}}_{PL}$ and obtained by maximizing Equation (12) converges almost surely to $\boldsymbol{\theta}$. Furthermore, $\hat{\boldsymbol{\theta}}_{PL}$ converges in distribution to $N(\boldsymbol{\theta}, \mathcal{S}_{\boldsymbol{\theta}}^{-1} \mathcal{V}_{\boldsymbol{\theta}} \mathcal{S}_{\boldsymbol{\theta}}^{-1})$ with

$$\mathcal{S}_{\boldsymbol{\theta}} = E\left(-\frac{\partial^2 \mathcal{L}^p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right) \quad \text{and} \quad \mathcal{V}_{\boldsymbol{\theta}} = E(\mathcal{U}_{\boldsymbol{\theta}}^p(\boldsymbol{\beta}, \boldsymbol{\lambda})\mathcal{U}_{\boldsymbol{\theta}}^p(\boldsymbol{\beta}, \boldsymbol{\lambda})^\top).$$

Similarly, the variability matrix (11) in the context of quasi-likelihood estimation, the matrix $\mathcal{V}_{\boldsymbol{\theta}}$ depends on third and fourth moments. Hence, we propose to use the empirical version of $\mathcal{V}_{\boldsymbol{\theta}}$, which is given by

$$\tilde{\mathcal{V}}_{\boldsymbol{\theta}} = \sum_{i=1}^n \mathcal{U}_{\boldsymbol{\theta}}^p(\boldsymbol{\theta})_i \mathcal{U}_{\boldsymbol{\theta}}^p(\boldsymbol{\theta})_i,$$

where the sum is understood to be element-wise. We shall now compute the components of the $\mathcal{S}_{\boldsymbol{\theta}}$. First, note that the matrix $\mathcal{S}_{\boldsymbol{\theta}}$ can be partitioned as

$$\mathcal{S}_{\boldsymbol{\theta}} = \begin{pmatrix} \mathcal{S}_{\boldsymbol{\beta}} & \mathcal{S}_{\boldsymbol{\beta}\delta} & \mathcal{S}_{\boldsymbol{\beta}p} \\ \mathcal{S}_{\delta\boldsymbol{\beta}} & \mathcal{S}_\delta & \mathcal{S}_{\phi p} \\ \mathcal{S}_{p\boldsymbol{\beta}} & \mathcal{S}_{p\phi} & \mathcal{S}_p \end{pmatrix}.$$

The entry (j, k) of the $Q \times Q$ matrix $\mathcal{S}_{\boldsymbol{\beta}}$ is given by

$$\mathcal{S}_{\beta_{jk}} = \sum_{i=1}^n \left(\frac{p^2 x_{ij} x_{ik}}{2} + \frac{x_{ij} x_{ik}}{\exp(\delta)\mu_i^{p-2}}\right).$$

Similarly, the entries $\mathcal{S}_\delta$ and $\mathcal{S}_p$ are, respectively, given by

$$\mathcal{S}_\delta = \frac{n}{2} \quad \text{and} \quad \mathcal{S}_p = \sum_{i=1}^n \frac{\log(\mu_i)^2}{2}.$$

Furthermore, the cross-entries have the form

$$\mathcal{S}_{\beta_j\delta} = \sum_{i=1}^n \frac{p x_{ij}}{2}, \quad \mathcal{S}_{\beta_j p} = \sum_{i=1}^n \frac{\log(\mu_i) x_{ij} - p}{2} \quad \text{and} \quad \mathcal{S}_{\delta p} = \sum_{i=1}^n \frac{\log(\mu_i)}{2}.$$

Finally, we propose the Newton scoring algorithm to solve the system of equations $\mathcal{U}_{\boldsymbol{\theta}}^p(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \mathbf{0}$, defined by

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \mathcal{S}_{\boldsymbol{\theta}}^{-1} \mathcal{U}_{\boldsymbol{\theta}}^p(\boldsymbol{\beta}^{(i)}, \boldsymbol{\lambda}^{(i)}).$$

In that case, we have to update $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ together, since the cross-entries of $\mathcal{S}_{\boldsymbol{\theta}}$ are not zeroes.

4. Simulation study

In this section, we present a simulation study that was conducted to compare the properties of the estimation methods. We evaluated the bias, consistency, coverage rate and efficiency of the MLE, quasi-likelihood estimator (QMLE) and pseudo-likelihood estimator (PMLE) . We simulated cross-sectional data sets, $(y_i, \boldsymbol{x}_i)$, $i = 1, \dots, n$, where y_i's are i.i.d. realizations of Y_i according to $Y_i \sim \text{Tw}_p(\mu_i, \phi)$ and $g(\mu_i) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$. We generated 1000 data sets considering the four sample sizes 100, 250, 500 and 1000. We considered five values of the power parameter (0, 1.01, 1.5, 2 and 3) combined with three amounts of variation. We used the average coefficient of variation to measure

the amount of variation introduced in the data. We defined, small, medium and large amount of variation data sets generated using coefficient of variation equaling to 15%, 50% and 80%, respectively. The values of the power parameter were chosen to have non-standard situations, as the cases of $p = 0$ and $p = 1.01$ where we expect the MLE does not work. The case of $p = 2$ is also difficult for maximum likelihood estimation, since the probability density function should be evaluated using two different infinity sums, for $p < 2$ and $p > 2$. The cases $p = 1.5$ and $p = 3$ represent the standard compound Poisson (called non-central gamma) and inverse Gaussian distributions, respectively. In these cases, we expect that the MLE works well, so we have safe results to compare with our two alternative approaches.

All scenarios consider models with an intercept ($\beta_0 = 2$) and slopes ($\beta_1 = 0.8$, $\beta_2 = -1.5$). The covariates are a sequence from -1 to 1, representing a continuous covariate, a factor with two levels (0 and 1) and length equaling the sample size. For $p = 0$ the dispersion parameter values are $\phi = (75, 850, 2100)$ corresponding, respectively, to small (15%), medium (50%) and large (80%) variation. Similarly, for $p = 1.01$, $p = 1.5$, $p = 2$ and $p = 3$ the dispersion parameter values are $\phi = (1.5, 15, 40)$, $\phi = (0.2, 2, 5.3)$, $\phi = (0.023, 0.25, 0.65)$ and $\phi = (0.0003, 0.0034, 0.0083)$, respectively. Figure 1 shows the bias plus and minus the SE for the parameters on each model and scenario. The scales are standardized for each parameter dividing the bias and the limits of the confidence intervals by the SE obtained on the sample of size 100.

The results in Figure 1 show that for the quasi- and pseudo-likelihood methods and all simulation scenarios, both the bias and SE tend to 0 as the sample size is increased. It shows the consistency and unbiasedness of our estimators. As expected the maximum likelihood method did not work for $p = 0$ and $p = 1.01$ in the medium and large variation scenarios. In these cases, the algorithm failed for all simulated data sets. In the cases of small variation the algorithm converged for 132 and 326 data sets for $p = 0$ and $p = 1.01$, respectively. In these scenarios, although the large bias for the dispersion parameters, the regression coefficients were consistently estimated.
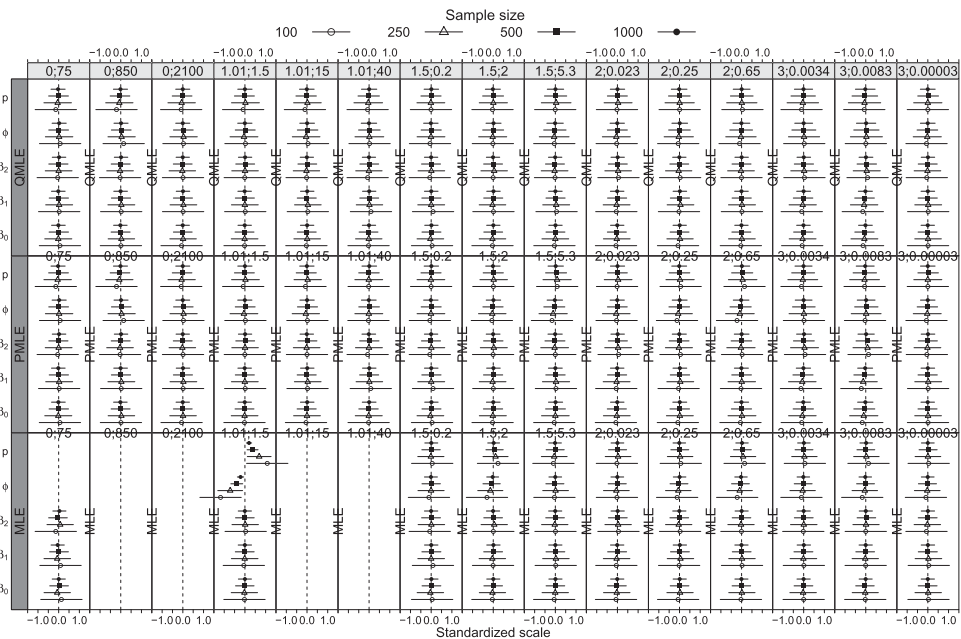

Figure 1. Bias and confidence interval on a standardized scale by estimation methods (MLE, PMLE and QMLE), sample size and different values of the power and dispersion parameters ($p; \phi$). The standardized scale (y-axis) is obtained for each parameter by dividing the bias and the limits of the confidence intervals by the SE obtained on the sample of size 100.

In general the coverage rates are close to the nominal level (0.95) for all parameters and simulation scenarios (see, Figure S4). The MLE presented coverage rate zero for the dispersion parameters, when $p = 0$ and $p = 1.01$ in all simulation scenarios (not shown). The quasi-likelihood method presented coverage rate closer to the nominal level than the pseudo-likelihood method, mainly for dispersion parameters and large values of the power parameter ($p \geq 1.5$). Regarding the estimation methods, the MLE presented a coverage rate close to the nominal level for large values of the power parameter as expected. The alternative approaches worked well in all simulation scenarios, including the cases where the MLE did not work. Finally, the empirical efficiency was computed as the ratio between the variance of the MLE and the variance obtained by the alternative approaches. We computed the efficiency only for the cases where $p \geq 1.5$, since for the other cases the MLE presented no reliable results.

The results in Figure S5 show that for the regression coefficients both QMLE and PMLE approaches presented efficiency close to 1 in all simulation scenarios. Concerns the dispersion parameters, for the small variation scenario the QMLE and PMLE presented efficiency close to 1. However, when the variation increased these estimators loss efficiency, the worst scenario appears for $p = 1.5$ and large variation, where the efficiency presented values around 20%. In general the PMLE is more efficient than the QMLE for the dispersion and power parameters.

5. Smoothing time series of rainfall in Curitiba, Paraná, Brazil

In this section, we present a real application of Tweedie regression models. The data set and R code can be obtained in the supplementary material.

This example concerns daily rainfall data in Curitiba, Paraná State, Brazil. The data were collected for the period from 2010 to 2015 corresponding to 2191 days. The main goal is to smooth the time series to help us better see patterns or trends. The analysis of rainfall data is in general challenged by the presence of many zeroes and the highly right-skewed distribution of the data. The plots shown in Figure 2 illustrate some of these features for the Curitiba rainfall data. In particular, Figure 2(B) highlights the right-skewed distribution and the considerable proportion of exact 0s (51%).
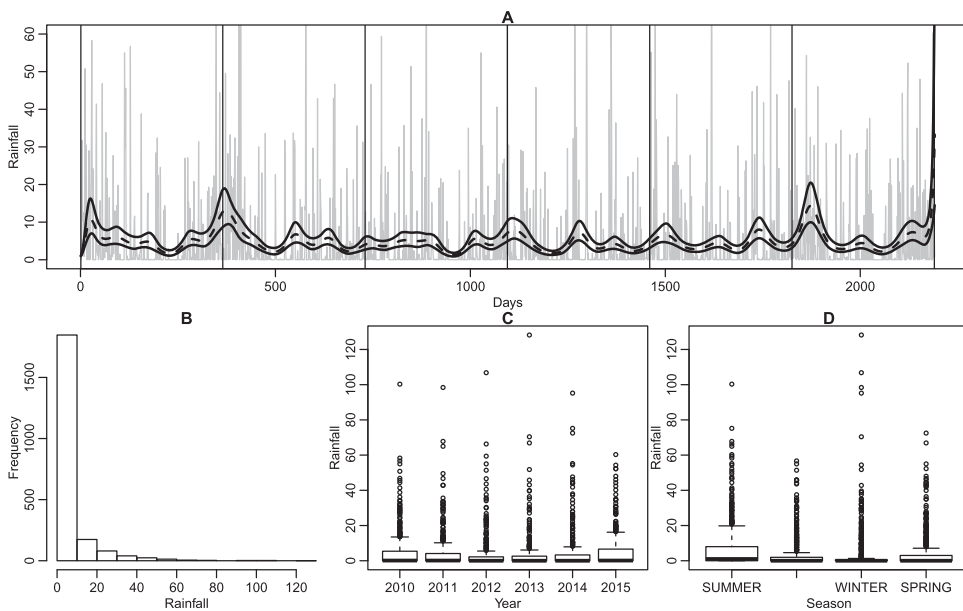

Figure 2. Time series plot for Curitiba rainfall data with fitted values (A). Vertical black lines indicate January 1st. Histogram of daily rainfall for the whole period (B). Boxplots for year (C) and season (D).

Table 1. Dispersion and power parameter estimates and SE by estimation methods for the Curitiba rainfall data.

| | Estimation methods | | | | | |
| | MLE | | QMLE | | PMLE | |
Parameter	Estimate	SE	Estimate	SE	Estimate	SE
δ	2.0284	0.0292	2.2791	0.2194	2.8543	0.4355
p	1.6774	0.0089	1.4721	0.1455	1.2652	0.2492

In order to smooth the Curitiba rainfall time series, we fitted a Tweedie regression model with linear predictor expressed in terms of B-splines [32]. The natural basis regression smoothing framework was used to select the degree of smoothness [33]. In this case, we found that 14 degrees of freedom were enough to smooth the times series. The models were fitted by using the three estimation methods, namely, MLE, QMLE and PMLE. Table 1 presents estimates and SE for the dispersion and power parameters.

The results in Table 1 show slightly different estimates for the dispersion and power parameters, depending on the estimation method used. However, the confidence intervals obtained by the QMLE and PMLE approaches contain the MLE. The SE obtained by the alternative approaches are larger than the ones obtained by the MLE.

In general, the QMLE method presented regression coefficients and confidence intervals more similar to the MLE than the PMLE method. The relative average difference between the MLE and QMLE estimates was 3.36%. On the other hand, the relative average difference between the MLE and PMLE estimates was 14.58%. Similarly, the confidence intervals obtained by the QMLE method were on average 3.33% wider than the corresponding MLE intervals. On the other hand, the confidence intervals obtained by the PMLE approach were 39.98% wider than the MLE intervals.

For all estimation methods, the power parameter estimates are in the interval $1 < p < 2$, suggesting a compound Poisson distribution, as expected, since the response variable is continuous with exact 0s. The fitted values and 95% confidence interval obtained by the quasi-likelihood method are shown in Figure 2. The fitted values obtained by the MLE and PMLE approaches were similar to the ones obtained by the QMLE (not shown). The smooth function captures the swing in the data and highlights the seasonal behaviour with dry and wet months around the winter and summer seasons, respectively.

In order to compare the computational times required by each approach for fitting the Tweedie regression model to this data set, we used the package `rbenchmark` [34]. The computations were done by a standard personal computer at 2.90 GHz with 8 G RAM by using the R software version 3.2.2 for 10 replications. The results showed that the QMLE approach is 37 and 0.22 times faster than the MLE and PMLE approaches, respectively.

6. Discussion

In this paper, we adopted the quasi- and pseudo-likelihood approaches to estimation and inference of Tweedie regression models. These approaches employ merely second-moments assumptions, allowing to extend the Tweedie regression models to the class of quasi-Tweedie regression models, which in turn offers robust and flexible models to deal with continuous data. Characteristics such as symmetry or asymmetry, heavy-tailed and excess 0s are easily handled because of the flexibility of the model class. These features indicate that the Tweedie model is a potential useful tool for the modelling of continuous data. The main advantage in practical terms is that we have one model for virtually all kinds of continuous data. Thus, model selection is done automatically when fitting the model.

The main advantages of the alternative estimation approaches in relation to the orthodox maximum likelihood method are their easy implementation and computational speed. Furthermore, by

employing only second-moment assumptions, we eliminated the non-trivial restriction on the parameter space of the power parameter, making the fitting algorithm simple and efficient. It also allows us to apply the Tweedie regression models for symmetric and heavy-tailed data, as the cases of Gaussian and t-Student data, where in general the power parameter presents negative and close to 0 values. Another potential application of Tweedie regression model is for the analysis of left-skewed data, where we also expect negative values for the power parameter.

The theoretical development in Section 3 showed that the quasi-likelihood approach has much in common with the orthodox maximum likelihood method. The quasi-score function employed in the context of quasi-likelihood estimation coincides with the score function for Tweedie distributions, which also implies that it will coincide for all ED models. The asymptotic variance of the quasi-likelihood estimators for the regression parameters coincides with the asymptotic variance of the MLE, in the case of known power and dispersion parameters. Hence, the quasi-likelihood approach provides asymptotic efficient estimation for the regression parameters. Furthermore, the quasi-likelihood approach as used in this paper combining the quasi-score and Pearson estimating functions, presents the insensitivity property (see Equation (9)) which is analogue to the orthogonality property in the context of maximum likelihood estimation. The insensitive property allows us to apply the two-step Newton scoring algorithm, using two separate equations to update the regression and dispersion parameters. A similar procedure can be used in the maximum likelihood framework, since the vectors β and λ are orthogonal. In the context of quasi-likelihood estimation, in this paper, we used the unbiased Pearson estimating function to estimation of the power and dispersion parameters. The discussion about efficiency in this case is difficult, since we cannot obtain a closed-form expression for the Fisher information matrix. The fact that the sensitivity and variability matrices associated with the dispersion parameters do not coincide indicate that the Pearson estimating functions are not optimum. Furthermore, the use of empirical high-order moments for the calculation of the Godambe information matrix must imply some efficiency loss. Thus, the SE associated with the power and dispersion parameters obtained by the quasi-likelihood method will be larger than the ones provided by the maximum likelihood method. On the other hand, it also makes the model robust against misspecification.

Concerning the pseudo-likelihood approach, it is a well-known result that when $\phi \to 0$ the ED models converge to the Gaussian distribution. Thus, at least for the small variation scenario the Gaussian pseudo-likelihood should provide descent estimators for both regression and dispersion parameters. Furthermore, since the estimators are obtained based on unbiased estimating functions, we also expect asymptotic unbiased and consistent estimators. The discussion about efficiency in the context of pseudo-likelihood estimation is difficult, because of the fact that the regression and dispersion parameters are not orthogonal. Hence, the asymptotic variance of the regression parameters also depends on high-order moments. In this paper, we used empirical high-order moments for the calculation of the asymptotic variance of the pseudo-likelihood estimators. Thus, we expect some efficiency loss for both regression and dispersion parameters, which as discussed for the quasi-likelihood method implies that the SE for the power and dispersion parameters computed by the pseudo-likelihood method will be larger than the ones obtained by the maximum likelihood method.

The simulation study presented in Section 3 showed that in general the quasi- and pseudo-likelihood estimators are unbiased and consistent for large sample, as suggest the asymptotic results presented in Section 3. In general the coverage rate presented values close to the nominal level for both methods and simulation scenarios. The main disadvantage of the quasi- and pseudo-likelihood estimators in relation to the maximum likelihood is the loss of efficiency on the estimation of the dispersion parameters, mainly on the high variation simulation scenario. However, it is important to highlight that the loss of efficiency on the estimation of the dispersion parameter does not affect the efficiency of the regression parameters that in general present values close to 1. As expected the maximum likelihood approach did not work well for small values of the power parameter. The algorithm presented many convergence problems, mainly when dealing with large sample size. The simulation study presented in the supplementary material showed that at least to some extent the Tweedie

regression model can handle heavy-tailed data as generated by the t-Student and slash distributions. However, for the cases of high variation data, some loss of efficiency on the estimation of the regression parameters is expected.

We illustrated the application of Tweedie regression models through the analysis of three data sets (two data analyses are presented in the supplementary material). The data sets were chosen to cover different types of continuous data. The first data set illustrates the case of right-skewed and zero inflation. As expected the three estimation methods estimated the power parameter in the interval 1 and 2, which in turn indicates a compound Poisson distribution. The second data analysis deals with right-skewed data, but without zero inflation, in that case we expected $p \geq 2$. The results of this data analysis confirmed our expectations. Finally, the third example considered symmetric data. In that case, we expected power parameter close to 0 indicating the Gaussian distribution. The two alternative methods confirmed our expectations. The maximum likelihood method for this data set converged, since the sample size is small, but offers a non-optimum fit. Regarding the estimation in general the quasi-likelihood estimates were more similar to the maximum likelihood estimates than the pseudo-likelihood estimates. In all data analyses the SE associated with the power and dispersion parameters obtained by the alternative methods were larger than the ones obtained by the maximum likelihood method. It shows the efficiency loss of these approaches and agrees with the results of our simulation study and theoretical development.

Possible topics for further investigation and extensions include extending the Tweedie regression models to the class of double Tweedie regression models, where the dispersion parameter is also described as a function of covariates [35]. It is also possible to incorporate penalized splines in the mean structure and to use regularization techniques as LASSO and SCAD for high dimensional data, with important applications in genetics. The current version of the fitting algorithms (which is available in the supplementary material) is a preliminary implementation of the Tweedie regression models. We plan to develop an R package with a GLM style interface to facilitate and propagate the use of Tweedie regression models. The package should also include residual analysis and influence measures.

ORCID

Wagner Hugo Bonat ⬥ http://orcid.org/0000-0002-0349-7054

References

[1] Nelder JA, Wedderburn RWM. Generalized linear models. J R Stat Soc Ser A. 1972;135:370–384.
[2] Jørgensen B. Exponential dispersion models. J R Stat Soc Ser B (Methodol). 1987;49:127–162.
[3] Jørgensen B. The theory of dispersion models. London: Chapman and Hall; 1997.
[4] Tweedie MCK. An index which distinguishes between some important exponential families. In: Ghosh JK, Roy J, editors. Statistics: applications and new directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference. Calcutta: Indian Statistical Institute; 1984.
[5] Jørgensen B, Paes De Souza MC. Fitting Tweedie's compound Poisson model to insurance claims data. Scand Actuar J. 1994;1994:69–93.

[6] Smyth GK, Jørgensen B. Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. ASTIN Bull. 2002;32:143–157.

[7] Dunn PK, Smyth GK. Series evaluation of Tweedie exponential dispersion model densities. Stat Comput. 2005;15:267–280.

[8] Dunn PK, Smyth GK. Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. Stat Comput. 2008;18:73–86.

[9] Dunn PK. tweedie: Tweedie exponential family models. R package version 2.1.7. Vienna: R Core Team; 2013.

[10] Bonat WH, Jørgensen B. Multivariate covariance generalized linear models. J R Stat Soc: Ser C (Appl Stat). 2016;65:649–675.

[11] Jørgensen B, Knudsen SJ. Parameter orthogonality and bias adjustment for estimating functions. Scand J Stat. 2004;31:93–114.

[12] Gourieroux C, Monfort A, Trognon A. Pseudo maximum likelihood methods: theory. Econometrica. 1984;52:681–700.

[13] Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika. 1974;61:439–447.

[14] Vinogradov V. On the power-variance family of probability distributions. Commun Stat – Theory Methods. 2004;33:1007–1029.

[15] Barndorff-Nielsen OE, Shephard N. Normal modified stable processes. Theory Probab Math Stat. 2001;65:1–19.

[16] Lee MT, Whitmore GA. Stochastic processes directed by randomized time. J Appl Probab. 1993;30:302–314.

[17] Kendall WS. A scale invariant clustering of genes on human chromosome 7. BMC Evol Biol. 2004;4:1–10.

[18] Kendall WS. Scale invariant correlations between genes and SNPs on human chromosome 1 reveal potential evolutionary mechanisms. J Theor Biol. 2007;245:329–340.

[19] Foster SD, Bravington MV. A Poisson-Gamma model for analysis of ecological non-negative continuous data. Environ Ecol Stat. 2013;20:533–552.

[20] Shono H. Application of the Tweedie distribution to zero-catch data in CPUE analysis. Fish Res. 2008;93:154–162.

[21] Kendall WS, Lagerwaard FJ, Agboola O. Characterization of the frequency distribution for human hematogenous metastases: evidence for clustering and a power variance function. Clin Exp Metastasis. 2000;18:219–229.

[22] Chen X-D, Tang N-S. Bayesian analysis of semiparametric reproductive dispersion mixed-effects models. Comput Stat Data Anal. 2010;54:2145–2158.

[23] Zhang Y. Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. Stat Comput. 2013;23:743–757.

[24] R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016.

[25] Fornberg B, Sloan DM. A review of pseudospectral methods for solving partial differential equations. Acta Numerica. 1994;3:203–267.

[26] Gilbert P, Varadhan R. numDeriv: accurate numerical derivatives. R package version 2014.2-1. Vienna: R Core Team; 2015.

[27] Nelder JA, Mead R. A simplex method for function minimization. Comp J. 1965;7:308–313.

[28] Holst R, Jørgensen B. Generalized linear longitudinal mixed models with linear covariance structure and multiplicative random effects. Chilean J Stat. 2015;6:15–36.

[29] Liang KY, Zeger SL. Inference based on estimating functions in the presence of nuisance parameters. Stat Sci. 1995;10:158–173.

[30] Jørgensen B, Demétrio CGB, Kristensen E, et al. Bias-corrected Pearson estimating functions for Taylor's power law applied to benthic macrofauna data. Stat Probab Lett. 2011;81:749–758.

[31] Park H, Cho K. Use of covariates in Taylor's power law for sequential sampling in pest management. J Agric Biol Environ Stat. 2004;9:462–478.

[32] de Boor C. On calculating with B-splines. J Approx Theory. 1972;6:50–62.

[33] Wood SN. Generalized additive models: an introduction with R. Chapman Hall/CRC: Texts in Statistical Science; 2006.

[34] Kusnierczyk W. rbenchmark: Benchmarking routine for R. R package version 1.0.0. Vienna: R Core Team; 2012.

[35] Wu KYK, Li WK. Double generalized threshold models with constraint on the dispersion by the mean. Comput Stat Data Anal. 2015;82:59–73.