Title

Estimating function analysis for a class of Tweedie regression models

Author

Wagner Hugo Bonat Departamento de Estatística - DEST, Laboratório de Estatística e Geoinformação - LEG, Universidade Federal do Paraná - UFPR, Curitiba, Paraná, Brasil, E-mail: *wbonat@ufpr.br* and
Departament of Mathematics and Computer science - IMADA, University of Southern Denmark - SDU, Odense, Denmark, E-mail: *wbonat@sdu.dk*

Abstract

We propose a new way to make inference on Tweedie regression models based on the estimating function approach. We adopted quasi-score function for regression parameters and Pearson estimating function for dispersion parameters. We perform a simulation study to compare our approach with the maximum likelihood method. The results show that both methods are similar, but estimating function approach is better to estimate small values of the power parameter. Some advantages to use estimating function are: i) avoid to evaluate the density function, ii) allow us estimate negative and between 0 and 1 values for the power parameter, iii) robust specification based on second-moments assumptions. We provide an `R` implementation for our new approach.

Keywords Tweedie regression, power variance function, estimating function, maximum likelihood.

Supplement materials http://www.leg.ufpr.br/doku.php/publications:papercompanions:tweedie

1 Introduction

Statistical modeling is one of the most significant fields of applied statistics with applications in many fields of scientific study, such as sociology, economy, agronomy, medicine and others. There exists an infinity of different statistical models, but the class of Generalized Linear Models (GLM) (Nelder and Wedderburn, 1972) is the most used in the last three decades. The success of this approach is due its ability to deal with different types of response variables, such as binary, count and continuous inside a general framework with a powerful scheme of inference based on the likelihood paradigm.

Some of the most important particular cases of GLM class are: a linear regression model based on the Gaussian distribution for real response variable, Gamma and inverse Gaussian regression models for positive real response variable, logistic regression based on the Binomial distribution for binary data and Poisson regression for count data. All these models are linked because they belong to the class of the exponential dispersion models (Jørgensen, 1997), and share an amazing characteristic: they are described by their first two moments (mean and variance). Furthermore, the variance function describes the relationship between the mean and variance of the response variable.

Let Y denote the response variable and assume that the probability function or density probability function of Y belongs to the class of exponential dispersion models and assume too that the $E(Y) = \mu$ and the $Var(Y) = \phi V(\mu) = \phi\mu^p$ then $Y \sim Tw_p(\mu, \phi)$, where $Tw_p(\mu, \phi)$ denotes a Tweedie (Tweedie, 1984), (Jørgensen, 1997) random variable with mean μ and variance $\phi\mu^p$ and $\phi > 0$ and $p \in (-\infty, 0] \cup [1, \infty)$ are parameters that describe the variance structure of Y.

Tweedie distribution has many interesting theoretical properties for a detailed description, see Jørgensen (1997). For practical situations in statistical modeling the Tweedie distribution is interesting because it delivers many important particular cases, and the parameter p identifies these cases. For example, $p = 0$ we have the Gaussian distribution, for $p = 1$ and $\phi = 1$ we have the Poisson distribution, for $p = 2$ and $p = 3$ correspond to the Gamma and inverse Gaussian distributions. Another important case is $1 < p < 2$ that corresponds to the Compound Poisson distribution.

Just by its particular cases the Tweedie distribution is already important for statistical modeling, but there exists an infinity of models, once the parameter p may be estimated based on a data set, making this simple relationship between mean and variance a rich class of statistical models.

For practical applications the estimation of the parameters that describe the variance structure (ϕ and p) is important and deserves the same attention devoted in the regression parameters. The orthodox approach is based on a likelihood paradigm, that is an efficient estimation method. A particularity about the Tweedie distribution is that outside the special cases, its probability density function cannot be written in a closed form, and requires any numerical method to evaluate the density function. Dunn and Smyth (2001) proposed some methods to evaluate the density function of the Tweedie distribution, but these methods are computationally demanding and shows a different level of accuracy for different regions of the parameter space. This fact makes the process of inference based on likelihood difficult and sometimes slow.

The main objective of the paper is to propose a new way to estimate the parameters (ϕ and p) based on Pearson estimating functions (Jørgensen and Knudsen, 2004). This method is very fast computationally, because it employs merely the first two moments (mean and variance) and in this way avoids evaluating the probability density function. Furthermore, we present an efficient and stable algorithm to obtain the point estimates. The inference is based on asymptotic results, and we show expressions for the sensitivity and the Godambe information matrix. The variance of the Pearson estimating function is approximated based on empirical, third and fourth moments. We run a simulation study to show the properties of our approach and compare with the maximum likelihood estimator in a finite sample scheme.

In the next Section we give some background about Tweedie distribution. In the Section 3 we present the Tweedie regression models and two approaches to make inference with respect the model parameters, maximum likelihood and estimating functions. Section 4 shows the main results from our simulation study and Section 5 reports some final remarks.

2 Background

The Tweedie distribution belongs to the class of exponential dispersion models (EDM) (Jørgensen, 1997). Thus, for a random variable Y which follows an EDM, the density function can be written as:

$$p_Y(y;\mu,\phi) = a(y,\phi)\exp\{(y\theta - k(\theta))/\phi\} \tag{1}$$

where $\mu = E(Y) = k'(\theta)$ is the mean, $\phi > 0$ is the dispersion parameter, θ is the canonical parameter, and $k(\theta)$ is the cumulant function. The function $a(y,\phi)$ cannot be written in closed form apart the particular cases cited. The variance is given by $Var(Y) = \phi V(\mu)$ where $V(\mu) = k''(\theta)$ 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 (cgf) is simple. The cgf is given by

$$K(t) = \{k(\theta + \phi t) - k(\theta)\}/\phi$$

where $k(\theta)$ is the cumulant function,

$$\theta = \begin{cases} \frac{\mu^{1-p}}{1-p} & p \neq 1 \\ \log\mu & p = 1 \end{cases}$$

and

$$k(\theta) = \begin{cases} \frac{\mu^{2-p}}{2-p} & p \neq 2 \\ \log\mu & p = 2. \end{cases}$$

The remaining factor in the density, $a(y,\phi)$ needs to be evaluated numerically. Jørgensen (1997) presents two series expressions for evaluating the density: one for $1 < p < 2$ and one 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

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

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

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

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

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

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

with $V = \sum_{i=1}^{\infty} V_k$ 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 (2001) presents a detailed study about these series and an algorithm to evaluate the Tweedie density function based on these series expansion. The algorithm is implemented in the package `tweedie` (Dunn, 2013) for the statistical software `R`(R Core Team, 2014) through the function `dtweedie.series`. Dunn and Smyth (2005) and Dunn and Smyth (2008) studied two more 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 sandlepoint approximation, for more details see (Dunn, 2013). In this paper we use only the approach described above.

3 Tweedie regression models

The Tweedie regression models were presented by Jørgensen and Paes De Souza (1994), Dunn and Smyth (2005), Hasan and Dunn (2011) and others. Consider independent responses $Y_1, Y_2, \ldots, Y_n$ are observed such that

$$Y_i \sim Tw_p(\mu_i, \phi)$$

where the mean μ_i is linked to linear predictor through a known link function g,

$$g(\mu_i) = x_i^T \beta$$

where x_i is a vector of covariates and β is a vector of unknown regression parameters. Let q be the dimension of β. On an equivalent way we can define the model using a matricial notation. Let $\mathbf{Y}$ a vector of response variable, then the Tweedie regression model can be defined by

$$\mathbf{Y} \sim Tw_p(\boldsymbol{\mu}, \phi I) \tag{2}$$

where I is a $n \times n$ dimensional identity matrix. In this case is easy to see that $E(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(X\beta)$ and the $Var(\mathbf{Y}) = C = diag(\phi \boldsymbol{\mu}^p)$. In this paper we define the link function g as the logarithm function. Note that the model is equivalently defined by its joint distribution defined in (2) or by its first two moments (mean and variance).

Denote the vector of parameters by $\boldsymbol{\theta} = (\beta, \boldsymbol{\lambda} = (\phi, p))$. The parameter vector can be divided in two sets, the first are the regression parameters and the second are parameters that describe the variance structure. In this paper we are interested to make inference about the second set. The orthodox method is based on likelihood function and it will be describe in the next section.

3.1 Maximum likelihood

Let $p_Y(y; \beta, \boldsymbol{\lambda})$ denote the Tweedie probability or density probability function as given in equation (1) and evaluated as described in Section 2. Then, the log-likelihood function for a sample of size n is given by

$$l(\phi, p) = \sum_{i=1}^{n} \log p_Y(y_i; \beta, \boldsymbol{\lambda}). \tag{3}$$

Maximizing the equation (3) with respect to β and $\boldsymbol{\lambda}$ we have the maximum likelihood estimator denoted by $\hat{\beta}_M$ and $\hat{\boldsymbol{\lambda}}_M$. The maximization process can be done by different ways, Dunn and Smyth (2005) proposed a method based on the BFGS algorithm and Jørgensen and Paes De Souza (1994) proposed a different scheme based on profile likelihood. In this paper we propose to use the Nelder-Mead (Nelder and Mead, 1965) method as implemented in the function `optim` of the `R` statistical software (R Core Team, 2014). In our simulation studies Nelder-Mead method shows stable and efficient results. To make inference about $\hat{\theta} = (\hat{\beta}, \hat{\boldsymbol{\lambda}})^T$ we use the well known asymptotic distribution of the maximum likelihood estimator,

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, I_o(\hat{\boldsymbol{\theta}})^{-1})$$

where $I_o(\boldsymbol{\theta})$ denote the observed information of $\boldsymbol{\theta}$. Note that, in the Tweedie regression models we cannot compute the Fisher information, because the second derivatives of log-likelihood function are not available in a closed form. In this way, we use the observed information matrix computed numerically using the Richardson method (Soetaert and Herman, 2009), on the point $\hat{\boldsymbol{\theta}}$.

Basically, our algorithm obtains the maximum likelihood estimates using the Nelder-Mead algorithm and compute the asymptotic variance of $\hat{\boldsymbol{\theta}}$ based on the inverse of negative of the Hessian matrix computed numerically by Richardson method. Note that this approach is computationally expensive, because we need evaluate the probability or density probability function of the Tweedie distribution many times inside the process of maximization. In the next section we shall present a new way to make inference about β and $\boldsymbol{\lambda}$ based on estimating functions.

3.2 Estimating functions

In this Section we describe estimating functions approach to estimate $\boldsymbol{\theta} = (\beta, \boldsymbol{\lambda})$. We adopted the quasi-score function for regression parameters and Pearson estimating function for dispersion parameters. Jørgensen and Knudsen (2004) describes the approach of estimating function, as well its properties. The quasi-score function is defined by,

$$\psi_\beta(\beta, \boldsymbol{\lambda}) = D^T C^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \tag{4}$$

where $D^T = \nabla_\beta \boldsymbol{\mu}$. The $q \times q$ matrix

$$S_\beta = E(\nabla_\beta \psi_\beta) = -D^T C^{-1} D \tag{5}$$

is called the *sensitivity matrix* of ψ_β and the $q \times q$ matrix

$$V_\beta = Var(\psi_\beta) = D^T C^{-1} D \tag{6}$$

is called the *variability matrix* of ψ_β.

In a similar way the Pearson estimating function is defined by,

$$\psi_{\boldsymbol{\lambda}_i}(\beta, \boldsymbol{\lambda}) = \mathbf{r}^T W_{\boldsymbol{\lambda} i} \mathbf{r} - tr(W_{\boldsymbol{\lambda} i} C^{-1}) \tag{7}$$

where $W_{\boldsymbol{\lambda} i} = C^{-1} \frac{\partial C}{\partial \boldsymbol{\lambda}_i} C^{-1}$ and $\mathbf{r} = (\mathbf{Y} - \boldsymbol{\mu})$. Note that everything we need to evaluate these equations are the derivatives with respect to $\boldsymbol{\lambda}_1 = \phi$ and $\boldsymbol{\lambda}_2 = p$. Is easy to show that,

$$\frac{\partial C}{\partial \phi} = \text{diag}(\boldsymbol{\mu}^p) \quad \text{and} \quad \frac{\partial C}{\partial p} = \text{diag}(\phi \log(\boldsymbol{\mu})\boldsymbol{\mu}^p). \tag{8}$$

The entries (i, j) of the 2×2 *sensitivity matrix* of $\psi_{\boldsymbol{\lambda}}$ are given by,

$$S_{\boldsymbol{\lambda}_{ij}} = E\left(\frac{\partial}{\partial \boldsymbol{\lambda}_i}\psi_{\boldsymbol{\lambda}_j}\right) = -tr\left(C^{-1}\frac{\partial C}{\partial \boldsymbol{\lambda}_i}C^{-1}\frac{\partial C}{\partial \boldsymbol{\lambda}_j}\right). \tag{9}$$

We can show using results about characteristic function of linear and quadratic forms of Non-Normal variables (Knight, 1985), that the entries of *variability matrix* of $\psi_{\boldsymbol{\lambda}}$ are given by,

$$V_{\boldsymbol{\lambda}_{ij}} = \text{Cov}(\psi_{\boldsymbol{\lambda}_i}; \psi_{\boldsymbol{\lambda}_j}) = 2\text{tr}(W_{\boldsymbol{\lambda}i}CW_{\boldsymbol{\lambda}j}C) + \sum_k k_l^{(4)}(W_{\boldsymbol{\lambda}i})_{ll}(W_{\boldsymbol{\lambda}j})_{ll} \tag{10}$$

where $k^{(4)}$ denote the fourth cumulant of $\mathbf{Y}$. To take into account the covariance between the vectors β and $\boldsymbol{\lambda}$, we need to compute the cross *sensitivity* and *variability matrix*. The entries of the cross sensitivity matrix between β and $\boldsymbol{\lambda}$ are given by,

$$S_{\boldsymbol{\beta_i\lambda_j}} = E\left(\frac{\partial}{\partial \boldsymbol{\lambda_j}}\psi_{\beta_i}\right) = 0. \tag{11}$$

In a similar way the entries of the cross *sensitivity matrix* between $\boldsymbol{\lambda}$ and β are given by,

$$S_{\boldsymbol{\lambda_i\beta_j}} = E\left(\frac{\partial}{\partial \beta_j}\psi_{\boldsymbol{\lambda}_i}\right) = -\text{tr}\left(C^{-1}\frac{\partial C}{\partial \boldsymbol{\lambda}_i}C^{-1}\frac{\partial C}{\partial \beta_j}\right). \tag{12}$$

Finally we can show that the entries of the *cross variability matrix* between β and $\boldsymbol{\lambda}$, are given by,

$$V_{\boldsymbol{\lambda_i}\beta_j} = E\left(\sum_{l=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} W_{\boldsymbol{\lambda_i}}^{(ij)} A_k^{(j)} r_l r_j r_k\right), \tag{13}$$

where $A = D^T C^{-1}$ and $A^{(j)}$ denote the j^{th} collumn of A. In a similar way $W_{\boldsymbol{\lambda_i}}^{(ij)}$ denote the i^{th} and j^{th} entries of the matrix $W_{\boldsymbol{\lambda_i}}$. Furthermore, the joint *sensitivity matrix* of ψ_β and $\psi_{\boldsymbol{\lambda}}$ is given by

$$S_{\boldsymbol{\theta}} = \begin{pmatrix} S_\beta & S_{\beta\boldsymbol{\lambda}} \\ S_{\boldsymbol{\lambda}\beta} & S_{\boldsymbol{\lambda}} \end{pmatrix}, \tag{14}$$

whose entries are defined in equations (5), (9), (12) and (11). Likewise, the joint *variability matrix* of ψ_β and $\psi_{\boldsymbol{\lambda}}$ is given by

$$V_{\boldsymbol{\theta}} = \begin{pmatrix} V_\beta & V_{\beta\boldsymbol{\lambda}} \\ V_{\boldsymbol{\lambda}\beta} & V_{\boldsymbol{\lambda}} \end{pmatrix}, \tag{15}$$

whose entries are defined in equations (6), (10) and (13).

Denote $\hat{\boldsymbol{\theta}}_e = (\hat{\beta}_e, \hat{\boldsymbol{\lambda}}_e)$ the estimate of $\boldsymbol{\theta}$, then the asymptotic distribution of $\hat{\boldsymbol{\theta}}_e$ is

$$\hat{\boldsymbol{\theta}}_e \sim N(\boldsymbol{\theta}, J_{\boldsymbol{\theta}}^{-1}) \tag{16}$$

where $J_{\boldsymbol{\theta}}^{-1}$ is the inverse of Godambe information matrix,

$$J_{\boldsymbol{\theta}}^{-1} = S_{\boldsymbol{\theta}}^{-1} V_{\boldsymbol{\theta}} S_{\boldsymbol{\theta}}^{-T}, \tag{17}$$

where $S_{\boldsymbol{\theta}}^{-T} = (S_{\boldsymbol{\theta}}^{-1})^T$.

Jørgensen and Knudsen (2004) proposed the *chaser* algorithm to solve the system of equations $\psi_\beta = 0$ and $\psi_{\boldsymbol{\lambda}} = 0$.

$$\begin{aligned} \beta^{(i+1)} &= \beta^{(i)} - S_\beta^{-1}\psi_\beta(\beta^{(i)}, \boldsymbol{\lambda^{(i)}}) \\ \boldsymbol{\lambda}^{(i+1)} &= \boldsymbol{\lambda}^{(i)} - S_{\boldsymbol{\lambda}}^{-1}\psi_{\boldsymbol{\lambda}}(\beta^{(i+1)}, \boldsymbol{\lambda^{(i)}}) \end{aligned}$$

The *chaser* algorithm uses the insensibility property, which allow us to use two separate equations to update β and $\boldsymbol{\lambda}$, for details see Jørgensen and Knudsen (2004). The described procedure was implemented in `R` and a generic function called `glm.tw()` is made available on the supplement material web page.

To compute the variance of dispersion parameters we need information about the third central moment and fourth cumulant. In the case of the Tweedie regression models we can compute these quantities based on the equations presented in the Section 2. An alternative approach is compute the empirical versions, in this way we avoid the supposition of multivariate Tweedie distribution for the data. The empirical fourth cumulant may be computed based on the data by the following equation:

$$k_l^{(4)} = (y_l - \hat{y}_l)^4 - 3(\hat{\phi}\hat{\mu}_l^{\hat{p}})^2.$$

The empirical third central moment may be computed based on equation (13) ignoring the expectation. The main overhead about to use empirical cumulants instead of the theorical cumulants, is that the variance should be overestimated, in this way the confidence interval based on this approach should be a little bigger than its should be.

4 Simulation study

4.1 Design of the study

We made a simulation study to evaluate the properties of the estimator based on the estimating function approach and compare its properties with the maximum likelihood estimator in finite sample. Our focus here is about the parameters that describe the variance structure of Tweedie regression models (ϕ,p). We use five different sample sizes (n = 50, 100, 250, 500 and 1000), and compare two measures of estimator quality (bias and coverage rate). In this manner, we have a quality measure based on point estimates and other based on confidence intervals.

To decide about which values of ϕ and p we take into account in the simulation study, we first plot graphics of the likelihood contours for $\phi = 0.5$ and $p = 1.1$, near the Poisson distribution, $p = 2$ the Gama and $p = 3$ Inverse Gaussian distributions. The Figure 1 shows these graphics.

The graphics presented in Figure 1 show that the likelihood contours are similar a quadratic function for $p = 2$ and $p = 3$, indicating that for these values of p the asymptotic distribution is well-behaved and near the Gaussian multivariate distribution. However, for $p = 1.1$ the likelihood behavior is non-quadratic, it shows that small values of p indicate more challenging setup to make inference. Thus, we choose the values of p = 1.1, 1.3, 1.5, 1.7 and 1.9 for the simulation study. The parameter ϕ measures the variability, so bigger values of ϕ indicate more challenging setup to make inference. We choose the values of ϕ = 0.5, 1, 1.5, 2 and 2.5. Combining five samples sizes, five values of p and five values of ϕ we have 125 different scenarios for our simulation study. All simulations was done using the `R` software and the package `tweedie` (Dunn, 2013).

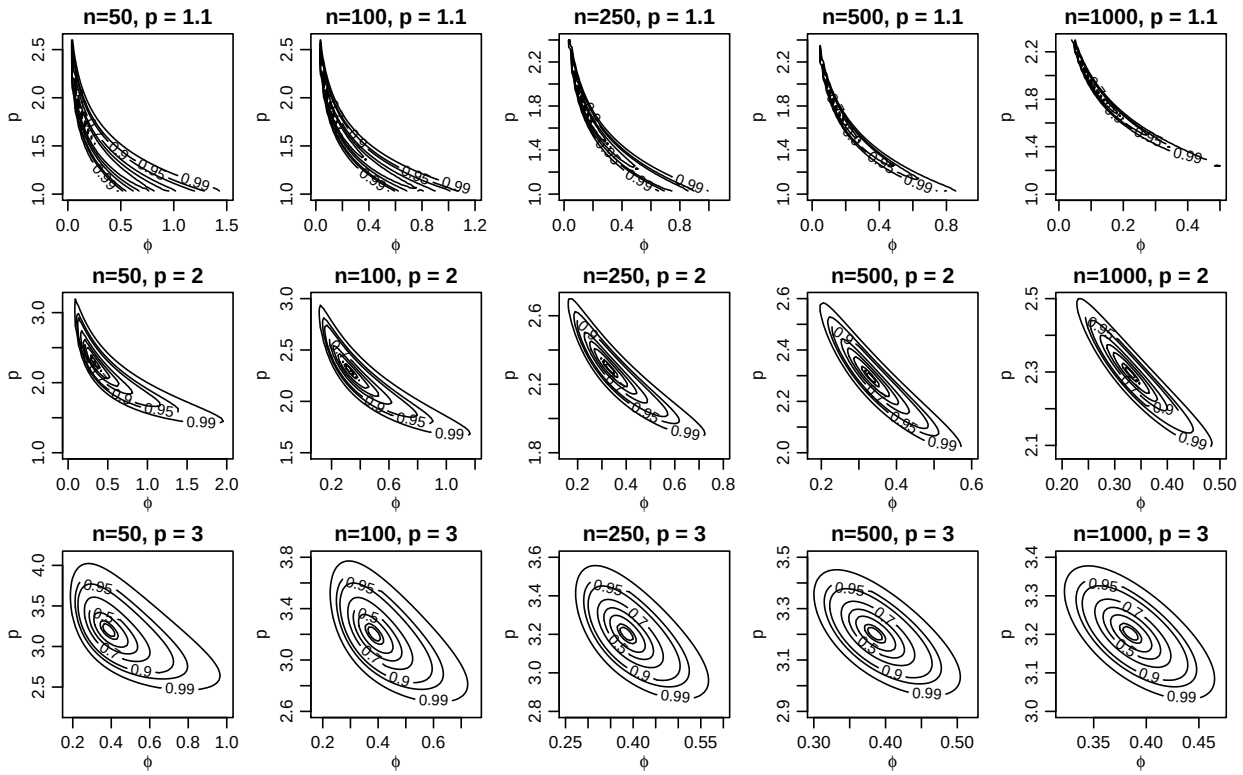

Figure 1: Likelihood contours for different values of p and sample sizes.

4.2 Results

We perform simulations to compare with the performance of estimators based on Pearson estimating functions against maximum likelihood estimators. We used two measures of quality estimator: the bias $b = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ and the coverage rate.

Our simulations consist of 1000 realizations from the Tweedie regression model (Section 3). We used a regression structure with $\beta_0 = 0.5$ and $\beta_1 = 1$, our model has one covariate, that was generated as a sequence from 0 to 2 and length depending on sample size. We used five sample sizes $n = 50$, 100, 250, 500 and 1000 and different combinations between the parameters ϕ and p, see Section 4.1. We choose to introduce the results through graphics. The Figure 2 presents the expected bias of $\hat{\phi}$ for different sample sizes, values of ϕ and p and estimation method, PEF (Pearson Estimating Function) and MLE (Maximum Likelihood Estimator).

The Figure 2 shows that in general the PEF estimator overestimate and MLE estimator underestimate ϕ for small sample size, but the bias decrease when the sample size increase as required. The bias of PEF estimator increase when the value of ϕ increase. The bias of MLE estimator is similar for all values of ϕ. In general the values of p do not affect the bias of $\hat{\phi}$. The bias of MLE estimator is lesser than PEF estimator, but for sample size around 100, the bias of PEF estimator is small enough to be useful for practical situations. In a similar way the Figure 3 presents the expected bias for $\hat{p}$ for different sample sizes, values of ϕ and p and estimation methods.

The results presented in Figure 3 show that the bias of PEF estimator is small for all sample sizes and parameter combinations. In fact the parameter p is well estimated using the PEF estimator for any configuration. The MLE estimator is less accurate to estimate small values of p using a small sample size, in this case MLE estimator overestimate p, but again the bias decreases fast

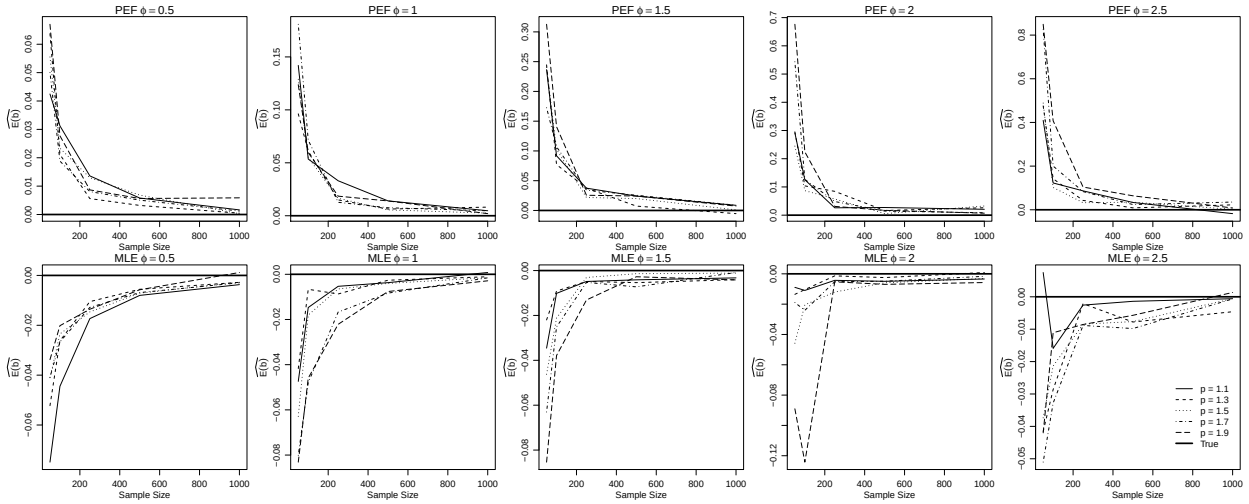

Figure 2: Expected bias of $\hat{\phi}$ for different methods, sample sizes and parameter combinations.

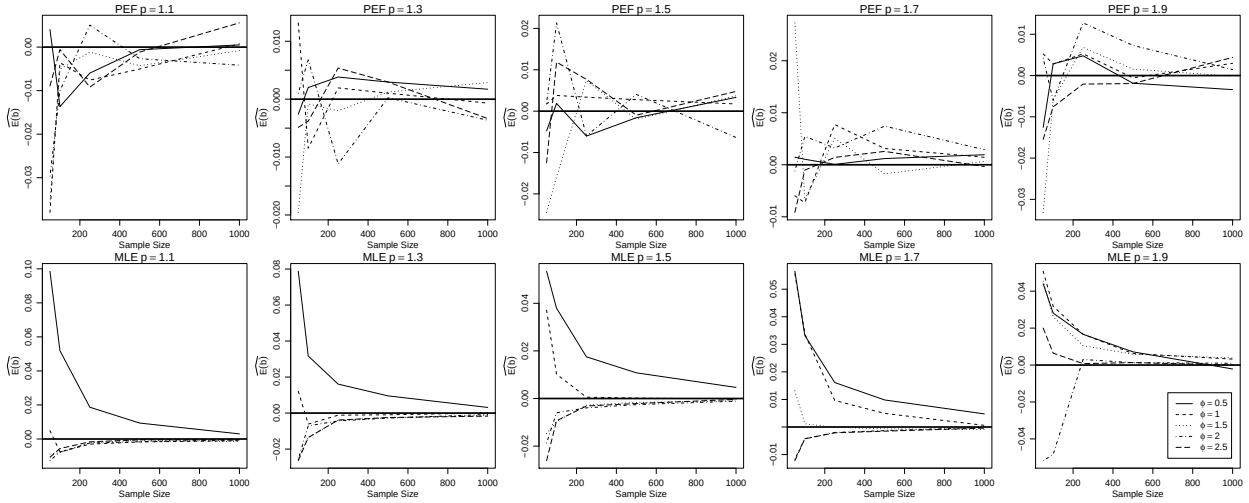

Figure 3: Expected bias of $\hat{p}$ for different methods, sample sizes and parameter combinations.

when the sample size increase and for sample size around 100 the bias is small enough for practical applications.

In general the results indicate that both methods given good point estimates for the parameters ϕ and p. In this manner, we evaluated the quality of point estimates. Now, we need to evaluate the quality of confidence intervals. For this task, we computed the coverage rate of the confidence interval of $\hat{\phi}$ and $\hat{p}$ for different sample sizes, combinations of ϕ and p and estimation methods. The coverage rate for the confidence interval of $\hat{\phi}$ is shown in Figure 4.

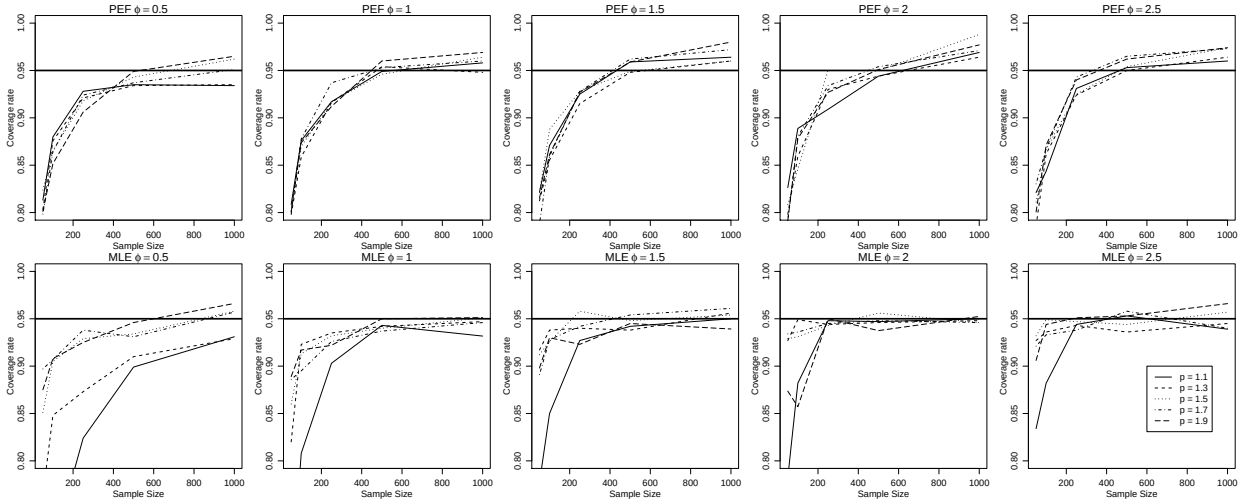

Figure 4: Coverage rate of ϕ for different methods, sample sizes and parameter combinations.

The results presented in Figure 4 show that for both methods the coverage rate is lesser than the nominal level for small sample size. The confidence intervals based on PEF approach achieve the nominal level for sample size around 250 for all configurations. Although, for big values of ϕ the coverage rate is slightly bigger than the nominal level.

The confidence interval based on MLE approach is not realistic for small sample size and small values of p and ϕ. For example, for $\phi = 0.5$ and $p = 1.1$ the confidence interval based on MLE approach does not achieve the nominal level same with sample size equal the 1000. For bigger values of ϕ the situation is better and for sample size around 250 the confidence intervals show coverage rate near the nominal level.

In general the results demonstrate that confidence intervals based on PEF approach does not depend on the combinations between ϕ and p values, for all configurations the results evidenced that for sample size around 250 the confidence intervals are well estimated. On the other hand, MLE approach has difficult to estimate confidence interval for small values of ϕ and p. Similar analysis is presented in Figure 5 for parameter p.

Figure 5 shows that the coverage rate for confidence intervals based on PEF approach are near the nominal level for all configurations considered, but again for big values of ϕ the coverage rate is slightly bigger than the nominal level. These results indicate that in general this approach presents confidence intervals bigger than should be. These results were expected, because we are using empirical third and fourth moments to compute the variance of $\hat{\phi}$ and $\hat{p}$. We argue that for practical data analysis these confidence intervals are enough accurate.

The confidence intervals based on MLE approach are not realistic for small values of ϕ and p, for example for $p = 1.1$ the confidence interval based on MLE approach does not achieve the nominal level, same using sample size equal to 1000. When the values of ϕ and p increase the results improve and are near the nominal level for all sample size.

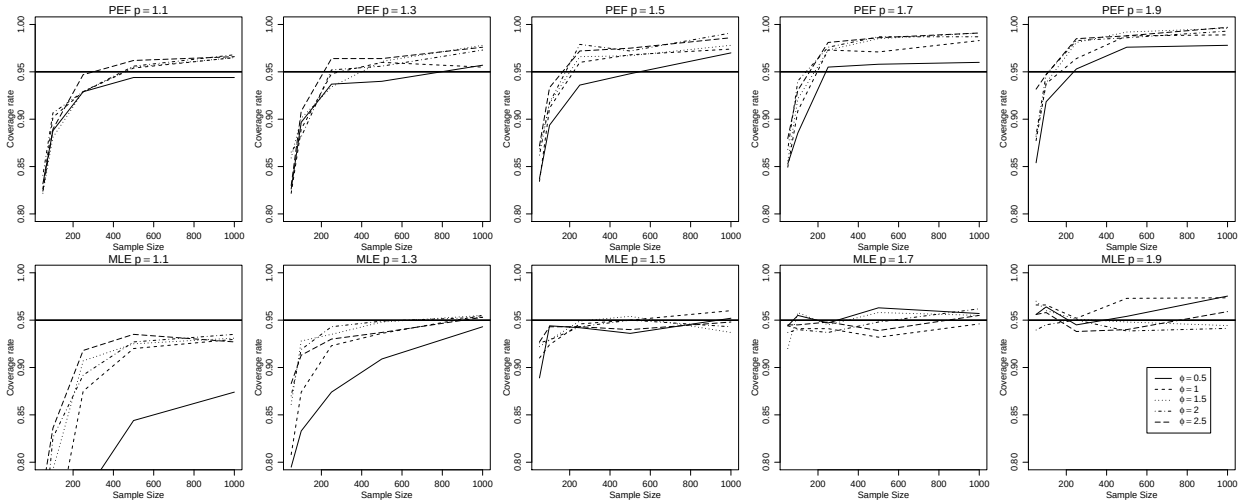

Figure 5: Coverage rate of p for different methods, sample sizes and parameter combinations.

In general way both methods show that are able to compute interval confidence, with a coverage rate near the nominal level. Of course, the results improve when the sample size increase, which is expected because our inferential methods are based on asymptotic results.

5 Conclusion

In this paper, we presented a new approach to make inferences with respect parameters of Tweedie regression models. Our approach is based on the quasi-score function for regression parameters and the Pearson estimating function for dispersion parameters. It is a well known result that quasi-score function yields the same estimator that maximum likelihood approach for regression parameters. Thus, we focus on dispersion parameters or the parameters that describe the variance structure of the Tweedie regression models.

We perform a simulation study to evaluate the quality of our estimator and compare with the maximum likelihood approach. The results show that both methods are similar, but the results based on Pearson estimating function are robust in the sense that for all combinations between parameters considered in the simulation study the PEF approach shows good results. On the other hand, the MLE approach showed difficult to estimate small values of ϕ and p.

Furthermore, we have many advantages to use estimating function approach. First, we do not need to evaluate the density function, that is a hard computational task. Second, we do not need hope about negative or near 1 values of p, once our approach deals with this situation naturally. Moreover, we can estimate values of p between 0 and 1, because our approach is based on second-moments assumptions, in this way we do not need to suppose that the response variable is distributed as the Tweedie distribution, it becomes our approach robust to misspecification.

A suggestion for future work with estimating function approach and Tweedie regression models may well be extend the Tweedie models for non-independent data, for example in longitudinal data analysis or repeat measures experiments. Tweedie models may be good models to deal with rainfall data, in this case is important to be able to analyze data with spatial and space-time structures, so extend Tweedie models to deal with dependent data is a promising approach and the use of estimation function become possible to do it in an elegant way.

References

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

Dunn, P. K. and Smyth, G. K. (2001). Tweedie family densities:methods of evaluation., *Proceedings of the 16th International Workshop on Statistical Modelling, Odense, Denmark*.

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

Dunn, P. and Smyth, G. (2008). Evaluation of tweedie exponential dispersion model densities by fourier inversion, *Statistics and Computing* **18**(1): 73–86.

Hasan, M. M. and Dunn, P. K. (2011). Two tweedie distributions that are near-optimal for modelling monthly rainfall in australia, *International Journal of Climatology* **31**(9): 1389–1397.

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

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 Paes De Souza, M. C. (1994). Fitting tweedies compound poisson model to insurance claims data, *Scandinavian Actuarial Journal* **1994**(1): 69–93.

Knight, J. L. (1985). The joint characteristic function of linear and quadratic forms of non-normal variables, *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* **47**(2): pp. 231–238.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization, *The Computer Journal* **7**(4): 308–313.

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

R Core Team (2014). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Soetaert, K. and Herman, P. M. (2009). *A Practical Guide to Ecological Modelling. Using R as a Simulation Platform*, Springer. ISBN 978-1-4020-8623-6.

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families, *in* J. K. Ghosh and J. Roy (eds), *Statistics: Applications and New Directions*, Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Calcutta: Indian Statistical Institute.