REVSTAT – Statistical Journal
Volume 22, Number 4, October 2024, 503–526
https://doi.org/10.57805/revstat.v22i4.524

A Simple Mean-Parameterized Maxwell Regression Model for Positive Response Variables

Author: Artur J. Lemonte
– Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Brazil
arturlemonte@gmail.com

Received: May 2021 Revised: December 2022 Accepted: December 2022

Abstract:

- We study a quite simple parametric regression model that may be very useful to model positive response variables in practice. The frequentist approach is considered to perform inferences, and the traditional maximum likelihood method is employed to estimate the unknown parameters. Monte Carlo simulation results indicate that the maximum likelihood approach is quite effective to estimate the model parameters. We also derive a closed-form expression for the second-order bias of the maximum likelihood estimator, which is easily computed as an ordinary linear regression and is then used to define bias-corrected maximum likelihood estimates. We consider the normalized quantile residuals for the new parametric regression model to assess departures from model assumptions, and global and local influence methods are also discussed. Applications to real data are considered to illustrate the new regression model in practice, and comparisons with two of the most popular existing regression models are made.

Keywords:

- *maximum likelihood estimation; Maxwell distribution; Maxwell-Boltzmann distribution; parametric inference.*

AMS Subject Classification:

- 62F10, 62J05, 62J20.

✉ Corresponding author

1. INTRODUCTION

The Maxwell distribution, also known in the statistic and physic literatures as Maxwell–Boltzmann distribution, has probability density function (PDF) in the form

$$f(y;\alpha) = \sqrt{\frac{2}{\pi}}\,\frac{y^2 \mathrm{e}^{-y^2/(2\alpha^2)}}{\alpha^3}, \quad y > 0,$$

where $\alpha > 0$ is the scale parameter. The mean and variance of the Maxwell distribution reduce to

$$\mathbb{E}(Y) = 2\alpha\sqrt{\frac{2}{\pi}}, \quad \text{and} \quad \mathbb{VAR}(Y) = \frac{\alpha^2(3\pi-8)}{\pi}.$$

There are, of course, some works related specifically to the one-parameter Maxwell distribution in the statistic literature. The reader is referred to Tyagi and Bhattacharya [31], Bekker and Roux [3], Dey and Maiti [13], Dey *et al.* [12], Al-Baldawi [1], Li [24], Fan [15], Dar *et al.* [11] and Hossain *et al.* [20], among others. It is evident that the one-parameter Maxwell distribution has noticeable scientific importance and, of course, it leaves open quite a number of new directions of research. In this paper, we provide a complete study regarding this one-parameter family of distributions in a parametric regression setup on the basis of a mean-parameterized Maxwell distribution.

In a parametric regression framework, it is typically more useful to model directly the mean (mode or median) of the response variable. In the last few years, several works have been published and so contributed to the regression literature on parameterizations based on the mean, mode, or median. To mention a few, but not limited to, we refer the reader to Yao and Li [33], Lemonte and Bazan [23], Chen *et al.* [6], Castellares *et al.* [5], Bourguignon *et al.* [4], Gallardo *et al.* [17], Gómez *et al.* [19], Leão *et al.* [22] and Menezes *et al.* [26]. In this paper, in order to obtain a regression structure for the mean of the Maxwell distribution, we shall work with a different parameterization of the Maxwell PDF. Let $\mu = 2\alpha(2/\pi)^{1/2}$ and, hence, $\alpha = (1/2)\mu(2/\pi)^{-1/2}$. In this case, substituting this expression in the Maxwell PDF, a reparameterization for the PDF is obtained; that is, the mean-parameterized Maxwell PDF is given by

$$f(y;\mu) = \left(\frac{2}{\pi}\right)^2 \frac{8y^2}{\mu^3} \exp\left(-\frac{4y^2}{\pi\mu^2}\right), \quad y > 0, \tag{1.1}$$

so that $\mathbb{E}(Y) = \mu > 0$ is the mean of the Maxwell distribution. Additionally, we have that $\mathbb{VAR}(Y) = 0.178\mu^2 \propto \mu^2$. The cumulative distribution function (CDF) of the mean-parameterized Maxwell takes the form

$$F(y,\mu) = \frac{2\gamma(3/2, 4y^2/(\pi\mu^2))}{\sqrt{\pi}}, \quad y > 0,$$

where $\gamma(a,x) = \int_0^x t^{a-1}\mathrm{e}^{-t}dt$ is the lower incomplete gamma function. We shall use the notation $\mathrm{Mw}(\mu)$ to refer to this distribution. We have that $\lim_{y\to 0} f(y) = \lim_{y\to\infty} f(y) = 0$ and, in addition, the mode is simply given by $\mu\sqrt{\pi}/2$. The Maxwell failure rate function is given by

$$r(y) = \left(\frac{2}{\pi}\right)^2 \frac{8y^2}{\mu^3}\left[1 - \frac{2\gamma(3/2, 4y^2/(\pi\mu^2))}{\sqrt{\pi}}\right]^{-1} \exp\left(-\frac{4y^2}{\pi\mu^2}\right), \quad y > 0.$$

Figure 1 displays some plots of the PDF and failure rate function for some values of μ.

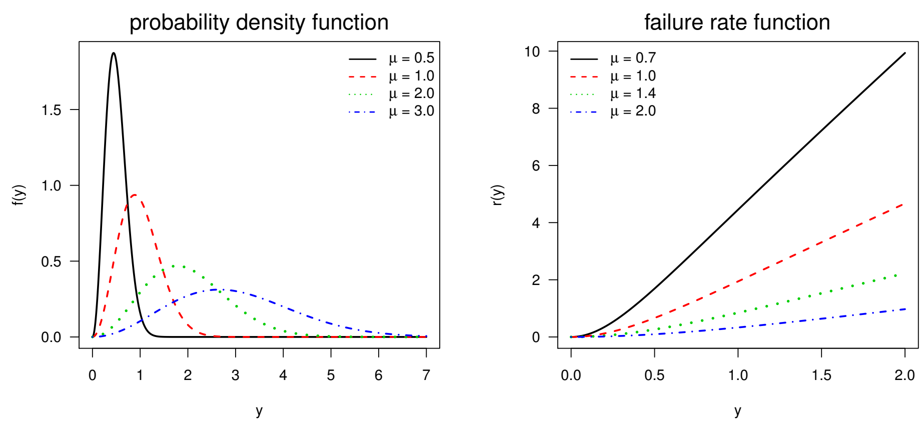

Figure 1: Density and failure rate functions.

We have the following propositions.

Proposition 1. The Maxwell PDF is log-concave for all values of $\mu > 0$.

Proof: The result follows by noting that the second derivative of $\log(f(y;\mu))$ is given by

$$\frac{\mathrm{d}^2 \log(f(y;\mu))}{\mathrm{d}y^2} = -\left(\frac{2}{y^2} + \frac{8}{\pi\mu^2}\right) < 0. \qquad \square$$

Proposition 2. For any $\mu > 0$, the Maxwell failure rate function is monotone increasing.

Proof: The result holds by using the log-concavity of the Maxwell PDF. $\square$

Remark 1. It is rather easy to generate random variates from the mean-parameterized Maxwell distribution. If U follows a gamma distribution with shape parameter $3/2$ and scale parameter 1, then $Y = \mu\sqrt{\pi U/4} \sim \mathrm{Mw}(\mu)$.

In this paper, we shall provide a parametric regression structure for the Maxwell distribution parameter, which involves covariates (explanatory variables) and unknown regression parameters. Furthermore, some quantities (e.g., score function, Fisher information matrix, etc.) related to the mean-parameterized Maxwell regression model are simple and compact, which makes the frequentist approach very easy to implement. Obviously the Bayesian approach has its merits and could also be considered and, in addition, these methodologies could be compared and contrasted. However, the comparison of these two methodologies is beyond the scope of this paper and hence can be considered in a future work. Also, it is quite common in practice, after modeling the real data at hand, to check the regression model assumptions and conduct diagnostic studies in order to detect possible atypical observations that may distort the results of the analysis. A first way to perform sensitivity analysis is by means of global influence starting from the case deletion proposed by Cook [7]. In addition, Cook [8] introduced a general framework to detect atypical observations under small perturbations on the data or in the model. In this paper, global and local influence are also considered to detect atypical observations in the class of Maxwell regression models. Throughout this paper, an *atypical* observation means that it can be an outlier[1], or observation with a large residual in absolute value, or an influential observation in the sense of global or local influences. Finally, it is well-known the residuals carry important information concerning the appropriateness of assumptions that underlie statistical models, and thereby play an important role in checking model adequacy identifying discrepancies between models and data. Hence, we propose the normalized quantile residual introduced by Dunn and Smyth [14] for the Maxwell regression model to study discrepancies between the model and data. In summary, the main contributions of this paper are as follows:

- We propose a Maxwell distribution parameterized in terms of its mean, allowing easy interpretation of the distribution parameter.
- Based on the mean-parameterized Maxwell distribution, we propose a novel parametric regression model for positive response variables, which is quite simple and may be very useful in practice, allowing for parameter interpretation in terms of the response in the original scale; that is, the regression parameters are interpretable in terms of the mean of the variable of interest.
- The direct modeling of the mean parameter in the mean-parameterized Maxwell regression model will promote its wider use in practice, putting it on the same level of interpretability and parsimony of some well-known regression models for positive response variables.
- The simulation and data analysis examples in this article reinforce that the proposed framework is a quite simple yet flexible way to model positive response variables.

The rest of this paper is organized as follows. The mean-parameterized Maxwell regression model is introduced in Section 2, and likelihood-based inference, as well as Monte Carlo simulation experiments are also performed. In Section 3, we propose diagnostic measures (i.e., global and local influence) for the mean-parameterized Maxwell regression model and, in particular, the normal curvature of local influence is derived under a specific perturbation scheme, namely: case weighting perturbation. Additionally, we also consider the normalized quantile residual to assess departures from the underlying distribution. Section 4 contains real data applications of the mean-parameterized Maxwell regression model for illustrative purposes. The paper ends up with some concluding remarks in Section 5.

2. THE MAXWELL REGRESSION MODEL

The model. Let $Y_1, ..., Y_n$ be n independent random variables, where each Y_i $(i = 1, ..., n)$ is Maxwell distributed and has PDF (1.1) with mean parameter μ_i; that is, $Y_i \sim \mathrm{Mw}(\mu_i)$ for $i = 1, ..., n$. In this work, we assume the following functional relation:

$$\log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \tag{2.1}$$

[1] An outlying observation, or "outlier," is one that appears to deviate markedly from other members of the sample in which it occurs.

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ is a vector of unknown regression coefficients, $\boldsymbol{\beta} \in \mathbb{R}^p$ with $p < n$, and $\boldsymbol{x}_i^\top = (x_{i1}, ..., x_{ip})$ are observations on p known covariates (or independent variables, or regressors). Generally, we have $x_{i1} = 1$ (for $i = 1, ..., n$) in practice and, hence, β_1 corresponds to the intercept parameter. It is worth emphasizing that other other links for the mean parameter in (2.1) could be considered, namely: identity ($\mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$), and square root ($\sqrt{\mu_i} = \boldsymbol{x}_i^\top \boldsymbol{\beta}$). However, the logarithm function is the most common and useful in such a case; that is, the main advantage of the exponential form $\mu_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ is that the requirement $\mu_i > 0$ is automatically satisfied for all $i = 1, ..., n$, whereas the identity and square root do not ensure such a requirement for all $i = 1, ..., n$. Note that the variance $\mathbb{VAR}(Y_i) = 0.178\mu_i^2 \propto \mu_i^2$ is a function of μ_i and, as a consequence, of the covariate values. Hence, non-constant response variances are naturally accommodated into the regression model. Moreover, we assume that the model matrix $\boldsymbol{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]^\top$ has column rank p.

Remark 2. Let ξ be the mode of the mean-parameterized Maxwell distribution, and so we have that $\xi = \mu\sqrt{\pi}/2$. The mean-parameterized Maxwell regression model is defined by the link function $\log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, for $i = 1, ..., n$, where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$, and $\boldsymbol{x}_i^\top = (x_{i1}, ..., x_{ip})$. Let $x_{i1} = 1$ (for $i = 1, ..., n$) and, hence, β_1 corresponds to the intercept parameter. In this case, $\log(\mu_i) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$. Note that

$$\log(\xi_i) = \beta_1^* + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \quad i = 1, ..., n,$$

where $\beta_1^* = \beta_1 + \log(\sqrt{\pi}/2)$ corresponds to the 'adjusted' intercept. Therefore, we can easily obtain the Maxwell modal regression model from the mean-parameterized Maxwell regression model.

Parameter estimation. Let $\boldsymbol{y} = (y_1, ..., y_n)^\top$ be the n-vector of the observed responses. We have that the parameter vector $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ represents the effects of the covariates on the mean parameter of the Maxwell regression model and, hence, we are interested in estimating this regression parameter vector. To do so, we shall consider the traditional maximum likelihood (ML) method. The log-likelihood function for this class of regression models, except for an unimportant constant term, has the form

$$\ell(\boldsymbol{\beta}) = -3\sum_{i=1}^{n} \log(\mu_i) - \frac{4}{\pi}\sum_{i=1}^{n} \frac{y_i^2}{\mu_i^2},$$

where $\mu_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$ for $i = 1, ..., n$. The ML estimate $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, ..., \widehat{\beta}_p)^\top$ of $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ is obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. The maximization can be performed, for example, in the R software [28] by using the `optim(...)` function. The score function, obtained by differentiating the log-likelihood function $\ell(\boldsymbol{\beta})$ with respect to the unknown parameters, is given by the p-vector $\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{X}^\top \boldsymbol{s}$, where $\boldsymbol{s} = (s_1, ..., s_n)^\top$ with $s_i = 8y_i^2/(\pi\mu_i^2) - 3$. After some algebra, the expected (Fisher) information matrix for $\boldsymbol{\beta}$ takes the form $\boldsymbol{K} = 6\boldsymbol{X}^\top \boldsymbol{X}$.

The ML estimate $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, ..., \widehat{\beta}_p)^\top$ can also be obtained by solving the nonlinear system of equations $\boldsymbol{U}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{0}_p$, where $\boldsymbol{0}_p$ denotes a p-dimensional vector of zeros. There is no closed-form expression for the ML estimate $\widehat{\boldsymbol{\beta}}$ and its computation has to be performed numerically using a nonlinear optimization algorithm. For example, the Newton–Raphson iterative technique (or the Gauss–Newton and Quasi-Newton methods) could be applied to solve these equations and obtain $\widehat{\boldsymbol{\beta}}$ numerically. On the other hand, one can use the Fisher scoring method to estimate $\boldsymbol{\beta}$ by iteratively solving the equation

$$\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{z}^{(m)}, \tag{2.2}$$

where $\boldsymbol{z} = (z_1, ..., z_n)^\top = \boldsymbol{X}\boldsymbol{\beta} + (1/6)\boldsymbol{s}$ acts as an adjusted dependent variable, and $m = 0, 1, ...$ is the iteration counter. The cycles through the scheme (2.2) consists of an iterative ordinary least squares algorithm to optimize the log-likelihood function, and the iterations go on until convergence is achieved (a stopping criterion must be defined). Equation (2.2) reveals that the calculation of the ML estimate $\widehat{\boldsymbol{\beta}}$ can be carried out using any software with a matrix algebra library as, for example, the R software. The optimization algorithms require the specification of initial values to be used in the iterative scheme.

In the following, we make some assumptions on the behavior of $\ell(\boldsymbol{\beta})$ as the sample size n approaches infinity, such as the regularity of the first three derivatives of $\ell(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, and the existence and uniqueness of the ML estimate of $\boldsymbol{\beta}$; see, for example, Cox and Hinkley [9]. When n is large and under standard regularity conditions, the ML estimators of the Maxwell regression parameters are asymptotically normal, asymptotically unbiased and have asymptotic variance-covariance matrix given by the inverse of the expected Fisher information matrix: $\widehat{\boldsymbol{\beta}} \overset{a}{\sim} N_p(\boldsymbol{\beta}, \boldsymbol{K}^{-1})$. This asymptotic normal distribution can be used to construct approximate confidence intervals for the Maxwell regression parameters. Let β_r $(r = 1, ..., p)$ be r-th component of $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$. The asymptotic confidence interval for β_r is simply given by $\widehat{\beta}_r \pm \Phi^{-1}(1 - \vartheta/2)\,\mathrm{se}(\widehat{\beta}_r)$, for $r = 1, ..., p$, with asymptotic coverage of $100(1 - \vartheta)\%$. Here, $\mathrm{se}(\cdot)$ is the square root of the diagonal element of $\boldsymbol{K}(\widehat{\boldsymbol{\beta}})^{-1}$ corresponding to each parameter (i.e., the asymptotic standard error), and $\Phi^{-1}(\cdot)$ is the standard normal quantile function.

Finite sample bias of the ML estimator. It is well-known that ML estimators are asymptotically unbiased and efficient, but for small samples, the ML estimators may not be unbiased. Here, we shall provide a general closed-form expression for the second-order biases of the ML estimators of the Maxwell regression parameters. To that end, we shall use the general expression given by Cox and Snell [10, Eq. (20)]. The closed-form expression will, in turn, allow us to obtain bias-corrected estimates of the unknown parameters. We shall use the following notation: $\kappa_{rs} = \mathbb{E}(\partial^2\ell(\boldsymbol{\beta})/\partial\beta_r\partial\beta_s)$, $\kappa_{rst} = \mathbb{E}(\partial^3\ell(\boldsymbol{\beta})/\partial\beta_r\partial\beta_s\partial\beta_t)$ and $\kappa_{rs}^{(t)} = \partial\kappa_{rs}/\partial\beta_t$, for $r, s, t = 1, ..., p$. After some algebra, we obtain

$$\kappa_{rs} = -6\sum_{i=1}^{n} x_{ir}x_{is}, \quad \kappa_{rst} = 12\sum_{i=1}^{n} x_{ir}x_{is}x_{it}, \quad \text{and} \quad \kappa_{rs}^{(t)} = 0.$$

Let B_a denote the second-order bias of $\widehat{\beta}_a$ $(a = 1, ..., p)$. From Cox and Snell [10], we can express B_a in the form

$$B_a = {\sum_{s,t,u}}' \kappa^{a,s}\kappa^{t,u}\left(\kappa_{st}^{(u)} - \frac{1}{2}\kappa_{stu}\right),$$

where $\kappa^{r,s}$ is the (r, s)-th element of $\boldsymbol{K}^{-1}$, and $\sum'$ denotes the summation over all combinations of parameters $\beta_1, ..., \beta_p$. Plugging the cumulants given before into this expression, we can obtain the bias of $\widehat{\boldsymbol{\beta}}$, say $\boldsymbol{B}$, in matrix form. We can show after some algebra that the $p \times 1$ bias vector $\boldsymbol{B}$ reduces to

$$\boldsymbol{B} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\delta}, \tag{2.3}$$

where $\boldsymbol{\delta}$ is the n-vector containing the elements of the main diagonal of the matrix $-(6\pi)^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. Note that the second-order bias vector $\boldsymbol{B}$ is simply the set coefficients from a simple ordinary least squares regression of $\boldsymbol{\delta}$ on the columns of the model matrix $\boldsymbol{X}$. As expression (2.3) makes clear, it is possible to express the bias vector of $\widehat{\boldsymbol{\beta}}$ as the solution of an ordinary least squares regression. Additionally, the bias vector $\boldsymbol{B}$ involves simple operations on matrices and vectors, and we can calculate it numerically via software with numerical linear algebra facilities such as R with minimal effort. It is worth emphasizing that the bias vector $\boldsymbol{B}$ will be small when $\boldsymbol{\delta}$ is orthogonal to the columns of $\boldsymbol{X}$. However, the second-order bias vector $\boldsymbol{B}$ may be large in small and moderate sized samples. From (2.3), we define the bias-corrected ML estimate $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - \boldsymbol{B}$. We say that $\widetilde{\boldsymbol{\beta}} = (\widetilde{\beta}_1, ..., \widetilde{\beta}_p)^\top$ is bias-adjusted ML estimate to order n^{-1}, since its bias is of order n^{-2}. It is expected that $\widetilde{\boldsymbol{\beta}}$ has superior finite-sample behavior relative to $\widehat{\boldsymbol{\beta}}$, whose bias is of order n^{-1}. It is not difficult to show that $\widetilde{\boldsymbol{\beta}} \overset{a}{\sim} N_p(\boldsymbol{\beta}, \boldsymbol{K}^{-1})$.

Simulation study. In what follows, we report Monte Carlo simulation experiments for the mean-parameterized Maxwell regression model. To explore the performance of the ML method in estimating the regression parameter vector $\boldsymbol{\beta}$, we report the results of simulations designed to evaluate the accuracy of the ML estimators of $\boldsymbol{\beta}$. The bias-adjusted ML estimate is also considered in the Monte Carlo simulations. The Monte Carlo experiments were carried out using $\log(\mu_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}$, where $x_{i1} = 1$ $(i = 1, ..., n)$, and $n = 10, 20, 30, 50, 80$ and 150. The true values of the regression parameters were taken as $\beta_1 = 1.0$, $\beta_2 = 0.5$ and $\beta_3 = 1.5$. The values of x_{i2} were obtained as random draws of the standard normal distribution, and the values of x_{i3} were obtained as random draws of the exponential distribution with mean equals 1. The covariate values were held constant throughout the simulations. We evaluate the point estimates by considering the following quantities: the mean, the relative bias[2] (RB), and the mean square error (MSE). These quantities are computed from 15,000 Monte Carlo replications. The numerical results are presented in Table 1. Note that the performance of the ML estimator of $\boldsymbol{\beta}$ is good, exhibiting small bias in all cases considered. It is noteworthy that the bias-adjusted estimator is better than the usual ML estimator for estimating the Maxwell regression parameters, mainly in very small sample sizes. However, for large sample sizes, the bias-corrected ML estimator becomes less justifiable. As expected, the MSE decreases as the sample size increases. In short, the numerical results reveal that the ML method can be used quite effectively to estimate the Maxwell regression parameters, and the bias-corrected ML estimator becomes a good alternative when the sample size is very small.

We now consider a Monte Carlo simulation study in the following way. First, we simulate data from the mean-parameterized Maxwell regression model and analyse the simulated data using the following models: mean-parameterized Maxwell, gamma, and inverse-Gaussian regression models. Next, we simulate data from a gamma model and analyse the simulated data using all three models (mean-parameterized Maxwell, gamma, and inverse-Gaussian regression models). Finally, we simulate data from an inverse Gaussian model and analyse the simulated data using all three models (mean-parameterized Maxwell, gamma, and inverse-Gaussian regression models). The gamma and inverse Gaussian regression models are very useful models for continuous positive response variables [see, for example, 25]. The Monte Carlo experiments were carried out using $\log(\mu_i) = \beta_1 + \beta_2 x_i$, for $i = 1, ..., n$, and $n = 50, 90$ and 150.

[2] The relative bias of an estimate $\widehat{\theta}$, defined as $[\mathbb{E}(\widehat{\theta}) - \theta]/\theta$, is obtained by estimating $\mathbb{E}(\widehat{\theta})$ by Monte Carlo.

Table 1: Simulation results regarding the point estimates of the Maxwell model parameters.

		ML estimator			Bias-corrected ML estimator		
		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widetilde{\beta}_1$	$\widetilde{\beta}_2$	$\widetilde{\beta}_3$
$n = 10$	Mean	0.913	0.550	1.516	0.939	0.536	1.512
	RB	−0.087	0.101	0.010	−0.061	0.072	0.008
	MSE	0.297	0.533	0.665	0.293	0.546	0.661
$n = 20$	Mean	0.971	0.510	1.499	0.981	0.506	1.498
	RB	−0.029	0.019	0.000	−0.019	0.012	−0.002
	MSE	0.245	0.500	0.517	0.244	0.503	0.515
$n = 30$	Mean	0.988	0.495	1.496	0.991	0.497	1.498
	RB	−0.012	−0.010	−0.003	−0.009	−0.006	−0.002
	MSE	0.199	0.489	0.524	0.199	0.487	0.526
$n = 50$	Mean	0.991	0.494	1.502	0.994	0.495	1.503
	RB	−0.009	−0.013	0.002	−0.006	−0.009	0.002
	MSE	0.187	0.470	0.455	0.187	0.468	0.455
$n = 80$	Mean	0.994	0.500	1.500	0.996	0.500	1.500
	RB	−0.006	0.001	0.000	−0.004	0.000	0.000
	MSE	0.182	0.439	0.444	0.182	0.439	0.444
$n = 150$	Mean	0.996	0.501	1.500	0.997	0.501	1.500
	RB	−0.004	0.002	0.000	−0.003	0.002	0.000
	MSE	0.175	0.430	0.432	0.175	0.430	0.432

The values of covariate x_i were obtained as random draws of the uniform distribution on the unit interval $(0, 1)$, and the covariate values were held constant throughout the simulations. We set $\beta_1 = 1.0$ and $\beta_2 = 0.8$. For the gamma and inverse Gaussian models, we consider the precision parameter, say ϕ, equals to $\phi = 4$ and $\phi = 5$, respectively. Tables 2 and 3 list the simulation results based on 10,000 Monte Carlo replications for the true data generating process (DGP) under three different scenarios: the Maxwell model as the true DGP, the gamma model as the true DGP, and the inverse Gaussian model as the true DGP. In Table 2 we present the point estimates, standard deviation (SD) between parentheses, and the values of Akaike information criterion (AIC) and Bayesian information criterion (BIC), whereas in Table 3 we present the coverage probability (CP) of the confidence intervals for β_1 and β_2 at the nominal levels 90% and 95%.

From Table 2, as expected, note that the Maxwell model yields the best fit under the Maxwell DGP, as well as the gamma and inverse models when these models correspond to the true DGPs; see the AIC and BIC values for the fitted models. It is also interesting to note that under the gamma DGP, the Maxwell model outperforms the inverse Gaussian model based on the AIC and BIC values. It is worth mentioning that under the inverse Gaussian DGP, the SDs of the ML estimates of the model parameters become larger than in the other two DGPs (Maxwell and gamma models). On the other hand, the ML estimates are close to the true values of the regression parameters, which indicates the 'robustness' of each model when estimating the regression parameters under model misspecification. From the numerical results in Table 3, we have that under the Maxwell DGP, the coverage rates of the confidence intervals are close to the nominal significance levels for all regression models, being the Maxwell regression model with the best performance, as expected. However, it is noteworthy that the coverage rates of the confidence intervals of the Maxwell regression parameters under the gamma and inverse Gaussian DGPs are not near the nominal levels, mainly under the gamma DGP. Finally, it should be mentioned that much more numerical work is needed to come to any general conclusion about the 'robustness' of the Maxwell regression model under model misspecification and, hence, future research regarding this issue can be conducted in a separate paper elsewhere.

Table 2: Simulation results considering three different data generating process.

n	Model	Maxwell DGP			
		β_1	β_2	AIC	BIC
50	Maxwell	1.005(0.098)	0.780(0.179)	233.618	237.807
	Gamma	1.007(0.101)	0.780(0.191)	236.195	244.478
	Inverse Gaussian	1.005(0.104)	0.783(0.198)	245.476	253.759
90	Maxwell	1.002(0.083)	0.785(0.151)	349.574	354.574
	Gamma	1.004(0.086)	0.783(0.160)	352.967	362.466
	Inverse Gaussian	1.003(0.090)	0.785(0.169)	367.312	376.812
150	Maxwell	1.000(0.057)	0.790(0.108)	582.682	588.703
	Gamma	1.002(0.059)	0.788(0.113)	587.642	598.674
	Inverse Gaussian	1.001(0.062)	0.790(0.120)	611.891	622.923

n	Model	Gamma DGP			
		β_1	β_2	AIC	BIC
50	Maxwell	1.033(0.145)	0.770(0.238)	249.305	253.494
	Gamma	1.008(0.137)	0.767(0.227)	247.040	255.324
	Inverse Gaussian	1.003(0.141)	0.778(0.239)	252.414	260.697
90	Maxwell	1.038(0.112)	0.777(0.183)	375.505	380.505
	Gamma	1.012(0.104)	0.773(0.169)	371.079	380.578
	Inverse Gaussian	1.010(0.105)	0.778(0.172)	379.219	388.719
150	Maxwell	1.029(0.092)	0.799(0.148)	625.968	631.989
	Gamma	1.001(0.086)	0.796(0.138)	617.266	628.297
	Inverse Gaussian	1.001(0.085)	0.796(0.139)	631.973	643.005

n	Model	Inverse Gaussian DGP			
		β_1	β_2	AIC	BIC
50	Maxwell	1.121(0.225)	0.938(0.483)	351.457	355.646
	Gamma	1.006(0.172)	0.778(0.358)	281.375	289.658
	Inverse Gaussian	1.002(0.176)	0.788(0.368)	272.734	281.017
90	Maxwell	1.128(0.203)	0.905(0.445)	525.594	530.594
	Gamma	1.009(0.168)	0.748(0.350)	417.941	427.441
	Inverse Gaussian	1.010(0.161)	0.744(0.332)	405.532	415.032
150	Maxwell	1.133(0.164)	0.950(0.362)	891.968	897.989
	Gamma	1.012(0.129)	0.784(0.269)	702.632	713.664
	Inverse Gaussian	1.009(0.127)	0.789(0.265)	681.196	692.228

Table 3: Coverage rates of confidence intervals considering three different data generating process.

n	Model	Maxwell DGP			
		CP(90%)		CP(95%)	
		β_1	β_2	β_1	β_2
50	Maxwell	0.892	0.899	0.955	0.941
	Gamma	0.890	0.900	0.941	0.945
	Inverse Gaussian	0.855	0.903	0.917	0.941
90	Maxwell	0.917	0.897	0.952	0.948
	Gamma	0.900	0.862	0.941	0.945
	Inverse Gaussian	0.848	0.879	0.921	0.945
150	Maxwell	0.893	0.896	0.952	0.941
	Gamma	0.897	0.876	0.962	0.945
	Inverse Gaussian	0.855	0.866	0.928	0.928

n	Model	Gamma DGP			
		CP(90%)		CP(95%)	
		β_1	β_2	β_1	β_2
50	Maxwell	0.776	0.800	0.841	0.883
	Gamma	0.883	0.893	0.934	0.948
	Inverse Gaussian	0.824	0.869	0.893	0.931
90	Maxwell	0.759	0.814	0.841	0.879
	Gamma	0.890	0.890	0.938	0.948
	Inverse Gaussian	0.828	0.900	0.917	0.945
150	Maxwell	0.759	0.790	0.828	0.855
	Gamma	0.879	0.883	0.934	0.945
	Inverse Gaussian	0.852	0.883	0.900	0.934

n	Model	Inverse Gaussian DGP			
		CP(90%)		CP(95%)	
		β_1	β_2	β_1	β_2
50	Maxwell	0.886	0.828	0.955	0.897
	Gamma	0.941	0.921	0.969	0.948
	Inverse Gaussian	0.907	0.903	0.948	0.955
90	Maxwell	0.872	0.790	0.938	0.886
	Gamma	0.948	0.900	0.983	0.972
	Inverse Gaussian	0.910	0.897	0.966	0.962
150	Maxwell	0.834	0.731	0.872	0.831
	Gamma	0.914	0.872	0.962	0.934
	Inverse Gaussian	0.879	0.872	0.928	0.941

3. DIAGNOSTIC MEASURES

It is well-known that regression models are sensitive to the underlying model assumptions and hence a sensitivity analysis is strongly advisable after fitting regression models to a dataset. In order to assess the sensitivity of the ML estimates of the mean-parameterized Maxwell model parameters in the presence of atypical observations, we shall consider the global and local influence methods [7, 8]. Additionally, the normalized quantile residual will be considered to assess departures from the underlying distribution.

Global influence. A first way to perform sensitivity analysis is by means of global influence starting from the case deletion proposed by Cook [7], which is a common approach to study the effect of dropping the i-th case from the dataset. Let $\widehat{\boldsymbol{\beta}}_{(-i)}$ be the ML estimate of $\boldsymbol{\beta}$ without the i-th observation in the sample. To assess the influence of the i-th case on the ML estimate $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, ..., \widehat{\beta}_p)^\top$, the basic idea is to compare the difference between $\widehat{\boldsymbol{\beta}}_{(-i)}$ and $\widehat{\boldsymbol{\beta}}$. If the deletion of an observation seriously influences an estimate, more attention should be paid to that particular observation. Hence, if $\widehat{\boldsymbol{\beta}}_{(-i)}$ is far from $\widehat{\boldsymbol{\beta}}$, then this case is regarded as an influential observation. To measure the global influence, the generalized Cook distance is defined as the standardized norm of $\widehat{\boldsymbol{\beta}}_{(-i)} - \widehat{\boldsymbol{\beta}}$ in the form $GD_i = (\widehat{\boldsymbol{\beta}}_{(-i)} - \widehat{\boldsymbol{\beta}})^\top \boldsymbol{J}_n(\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}_{(-i)} - \widehat{\boldsymbol{\beta}})$ for $i = 1, ..., n$, where $\boldsymbol{J}_n(\boldsymbol{\beta}) = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$ is the observed (Fisher) information matrix, and $\boldsymbol{W} = \text{diag}\{w_1, ..., w_n\}$ with $w_i = 16y_i^2/(\pi\mu_i^2)$. Note that we have to compute $\widehat{\boldsymbol{\beta}}_{(-i)}$ for all $i = 1, ..., n$. To avoid employing the direct model estimation for all observations, we can use the following one-step approximation to reduce the number of models to be fitted: $\widehat{\boldsymbol{\beta}}_{(-i)} \doteq \widehat{\boldsymbol{\beta}} - \boldsymbol{J}_n(\widehat{\boldsymbol{\beta}})^{-1} \dot{\boldsymbol{L}}_i(\widehat{\boldsymbol{\beta}})$, where $\dot{\boldsymbol{L}}_i(\boldsymbol{\beta}) = \partial \ell_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$, and $\ell_i(\boldsymbol{\beta}) = -3\log(\mu_i) - 4y_i^2/(\pi\mu_i^2)$. It follows that $\widehat{\boldsymbol{\beta}}_{(-i)} - \widehat{\boldsymbol{\beta}} \doteq -\boldsymbol{J}_n(\widehat{\boldsymbol{\beta}})^{-1} \boldsymbol{x}_i \, \widehat{s}_i$, where $\widehat{s}_i := s_i(\widehat{\boldsymbol{\beta}})$. Hence, the generalized Cook distance reduces to $GD_i = \widehat{s}_i^2 \, \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \widehat{\boldsymbol{W}} \boldsymbol{X})^{-1} \boldsymbol{x}_i$, for $i = 1, ..., n$, where $\widehat{\boldsymbol{W}} := \boldsymbol{W}(\widehat{\boldsymbol{\beta}})$. The index plot of GD_i may reveal those influential observations on the ML estimates of the Maxwell regression parameters.

Local influence. In the following, the local influence method under a specific perturbation scheme (case weighting perturbation) is carried out in order to assess the sensitivity of the ML estimates of the Maxwell regression parameters. Let $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ be a k-dimensional vector of perturbations, where $\boldsymbol{\Omega} \subset \mathbb{R}^k$ is an open set. The perturbed log-likelihood function is denoted by $\ell(\boldsymbol{\beta}|\boldsymbol{\omega})$. The vector of no perturbation is $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ such that $\ell(\boldsymbol{\beta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\beta})$. The Cook's idea for assessing local influence is essentially analyzing the local behavior of the log-likelihood displacement $LD_{\boldsymbol{\omega}} = 2[\ell(\widehat{\boldsymbol{\beta}}) - \ell(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})]$, where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}$ denotes the ML estimate under $\ell(\boldsymbol{\beta}|\boldsymbol{\omega})$, around $\boldsymbol{\omega}_0$ by evaluating the curvature of the plot of $LD_{\boldsymbol{\omega}_0 + a\boldsymbol{d}}$ against a, where $a \in \mathbb{R}$ and $\boldsymbol{d}$ is a unit norm direction. One of the measures of particular interest is the direction $\boldsymbol{d}_{\max}$ corresponding to the largest curvature $C_{\boldsymbol{d}_{\max}}$. Cook [8] proved that the normal curvature at the direction $\boldsymbol{d}$ is given by $C_{\boldsymbol{d}}(\boldsymbol{\beta}) = 2|\boldsymbol{d}^\top \boldsymbol{\Delta}^\top \boldsymbol{J}_n(\boldsymbol{\beta})^{-1} \boldsymbol{\Delta} \boldsymbol{d}|$, where $\boldsymbol{\Delta} = \partial^2 \ell(\boldsymbol{\beta}|\boldsymbol{\omega})/\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top$ and $\boldsymbol{J}_n(\boldsymbol{\beta})$ are evaluated at $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\omega}_0$. We have that $\boldsymbol{J}_n(\boldsymbol{\beta}) = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$ and, after some algebra, we can show that $\boldsymbol{\Delta} = \boldsymbol{X}^\top \boldsymbol{S}$, where $\boldsymbol{S} = \text{diag}\{s_1, ..., s_n\}$. Let $(1/2)C_{\boldsymbol{d}_{\max}}$ be the largest eigenvalue of $\boldsymbol{L} = -\boldsymbol{\Delta}^\top \boldsymbol{J}_n(\boldsymbol{\beta})^{-1} \boldsymbol{\Delta}$, and $\boldsymbol{d}_{\max}$ be the corresponding unit norm eigenvector ($||\boldsymbol{d}_{\max}|| = 1$). The index plot of the largest eigenvector ($\boldsymbol{d}_{\max}$) of $\boldsymbol{L}$ may reveal those influential observations on the ML estimate $\widehat{\boldsymbol{\beta}}$.

Residuals. Usually, the residuals are defined in order to study departures from the response distribution assumptions. More precisely, the residuals carry important information concerning the appropriateness of assumptions that underlie statistical models, and thereby play an important role in checking model adequacy. The use of residuals for assessing the adequacy of fitted regression models is nowadays commonplace due to the widespread availability of statistical software, many of which are capable of displaying residuals and diagnostic plots, at least for the more commonly used models. We shall consider the normalized quantile residuals proposed in Dunn and Smyth [14] to check the adequacy of the Maxwell regression model fitted to a dataset, which is simply defined as

$$R_i = \Phi^{-1}\left(\frac{2\gamma(3/2, 4y_i^2/(\pi\widehat{\mu}_i^2))}{\sqrt{\pi}}\right), \quad i = 1, ..., n, \tag{3.1}$$

where $\widehat{\mu}_i = \exp(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}})$. The normalized quantile residuals in (3.1) have a standard normal distribution asymptotically [14, 16]. Since the exact distribution of the above residual is not known, it is usual to add envelopes as suggested by Atkinson [2, § 4.2] into the normal quantile-quantile plot (QQ-plot) for R_i to decide whether the observed residuals are consistent with the fitted regression model. Thus, observations corresponding to absolute residuals outside the limits provided by the simulated envelope are worthy of further investigation. Additionally, if a considerable proportion of points falls outside the envelope, then one has evidence against the adequacy of the fitted model.

Remark 3. The simple closed-form expression for the bias vector of the ML estimators of the Maxwell regression parameters in (2.3) can be used to define improved Pearson residuals [see, for example, 10] for the mean-parameterized Maxwell regression model. Hence, future research can be done to compare through Monte Carlo simulations the improved Pearson residuals and the normalized quantile residuals.

4. ILLUSTRATIVE EXAMPLES

In what follows, we shall consider real data examples to illustrate the Maxwell regression model in practice. All computations regarding the mean-parameterized Maxwell regression model were carried out using the R program. The R code to compute the ML estimates of the mean-parameterized Maxwell regression model parameters is provided in the Appendix.

Life of metal pieces data. Here, we consider the biaxial fatigue data on the life (in cycles to failure) of metal pieces reported by Rieck and Nedelman [29]. The response variable (Y) is the life (in number of cycles to failure) of $n = 46$ metal pieces, and the explanatory variable (x) is the work per cycle (mJ/m^3). We assume that $Y_i \sim \mathrm{Mw}(\mu_i)$, for $i = 1, \ldots, 46$, where $\log(\mu_i) = \beta_1 + \beta_2 \log(x_i)$. The ML estimates, asymptotic standard errors (SE) and the 95% asymptotic confidence intervals (CI) of the Maxwell regression parameters are listed in Table 4. Figure 2 displays the normalized quantile residuals for the Maxwell regression model. We have in this figure the quantile residuals against the index, and the normal QQ-plot (with generated envelopes), respectively. Note that the residuals appear satisfactory (random) and, more important, there is no observation falling outside the envelope. Therefore, the mean-parameterized Maxwell regression model provides a good fit to the biaxial fatigue data. Figure 2 shows the index plot of the generalized Cook distance, as well as the index plot of $|\boldsymbol{d}_{\max}|$. The generalized Cook distance identifies the cases #4 and #46 as possible influential observations on the ML estimates of the Maxwell regression parameters. We remove each of these observations individually from the dataset and, after that, we fit the mean-parameterized Maxwell regression model. We observe that there is no inferential change regarding the regression parameters when removing the cases #4 and #46 from the dataset and, hence, these observations have no influence on the ML estimates of the Maxwell regression parameters. The estimated Maxwell regression model is

$$\log(\widehat{\mu}_i) = 12.4733 - 1.706 \log(x_i), \quad i = 1, \ldots, 46.$$

The coefficients of the mean-parameterized Maxwell regression model can be interpreted as follows. The expected life (in cycles to failure) of a metal piece should decrease approximately 81.84% $[(1 - \mathrm{e}^{-1.7060}) \times 100\%]$ as the logarithm of work per cycle increases one unity;

that is, there is a decrease in the expected rate of life (in cycles to failure) by a factor of (approximately) 0.1816 [$\exp(-1.7060) = 0.1816$].

Advertising media data. Next, we shall consider data corresponding to the impact of newspapers on sales. These data are the advertising budget (in thousands of dollars) along with sales. The advertising experiment has $n = 200$ observations, and they are available in the R package `datarium` [21]. The response variable (Y) corresponds to the sales (in thousands of dollars), while the covariate (x) corresponds to the advertising budget on newspapers (in thousands of dollars). We assume that $Y_i \sim \text{Mw}(\mu_i)$, for $i = 1, ..., 200$, where $\log(\mu_i) = \beta_1 + \beta_2\, x_i$. The mean-parameterized Maxwell regression estimates are provided in Table 4. In addition, Figure 3 confirms that the Maxwell regression model is suitable to model the data, since there are no observations falling outside the envelope. The index plots of GD_i (generalized Cook distance) and $|\boldsymbol{d}_{\max}|$ (local influence) are presented in Figure 3. It is identified the cases #37 and #129 as possible influential observations on the ML estimates of the mean-parameterized Maxwell regression parameters. We remove each of these observations individually from the dataset and, after that, we fit the Maxwell regression model. There is no inferential change regarding the regression parameters when removing these cases from the dataset, so these observations have no influence on the ML estimates of the Maxwell regression parameters. The estimated Maxwell regression model is

$$\log(\widehat{\mu}_i) = 2.6879 + 0.003\, x_i, \quad i = 1, ..., 200,$$

and the ML estimates of the mean-parameterized Maxwell regression parameters deliver the following interpretation. The expected sale (in thousands of dollars) should increase (approximately) 0.301% [$(\mathrm{e}^{0.003} - 1) \times 100\%$] as the advertising budget on newspapers increases one thousand dollars; that is, there is an increase in the expected sale by a factor of (approximately) 1.003 [$\exp(0.003) = 1.003$].

Radioimmunoassay data. Now, we consider the radioimmunoassay data, reported in Tiede and Pagano [30]. These data were obtained from the Nuclear Medicine Department of the Veteran's Administration Hospital, Buffalo, New York. The variable of interest (Y) is the radioactivity count rate, and the covariate (x) corresponds to the dose concentration (measured in micro-international units per milliliter). We assume that $Y_i \sim \text{Mw}(\mu_i)$, for $i = 1, ..., 14$, where $\log(\mu_i) = \beta_1 + \beta_2\, x_i$. Table 4 lists the ML estimates, asymptotic SEs and the 95% asymptotic CIs of the Maxwell regression parameters. Residuals plots are displayed in Figure 4, which confirms that the mean-parameterized Maxwell regression model is suitable to model the data, since there are no observations falling outside the envelope. Figure 4 displays the index plot of the generalized Cook distance, as well as the index plot of $|\boldsymbol{d}_{\max}|$. It is identified the cases #1, #2 and #14 as possible influential observations on the ML estimates of the Maxwell regression parameters. We remove each of these observations individually from the dataset and, after that, we fit the mean-parameterized Maxwell regression model. It is noteworthy that there is no inferential change regarding the regression parameters when removing the cases #1, #2 and #14 from the dataset, revealing that these observations have no influence on the ML estimates of the Maxwell regression parameters. The estimated Maxwell regression model is

$$\log(\widehat{\mu}_i) = 8.6091 - 0.0190\, x_i, \quad i = 1, ..., 14.$$

The mean-parameterized Maxwell parameter estimates deliver interesting interpretation. The expected radioactivity count rate should decrease approximately 1.88% $[(1-e^{-0.0190})\times 100\%]$ as the dose concentration increases one unity; that is, there is a decrease in the expected radioactivity count rate by a factor of (approximately) 0.98 $[\exp(-0.0190) = 0.98]$.

Table 4: Parameter estimates.

Parameter	Life of metal pieces data		
	Estimate	SE	95% CI
β_1	12.4733	0.4007	$(11.688; 13.259)$
β_2	-1.7060	0.1114	$(-1.924; -1.488)$

Parameter	Advertising media data		
	Estimate	SE	95% CI
β_1	2.6879	0.0498	$(2.590; 2.785)$
β_2	0.0030	0.0011	$(0.001; 0.005)$

Parameter	Radioimmunoassay data		
	Estimate	SE	95% CI
β_1	8.6091	0.1390	$(8.337; 8.881)$
β_2	-0.0190	0.0032	$(-0.025; -0.013)$

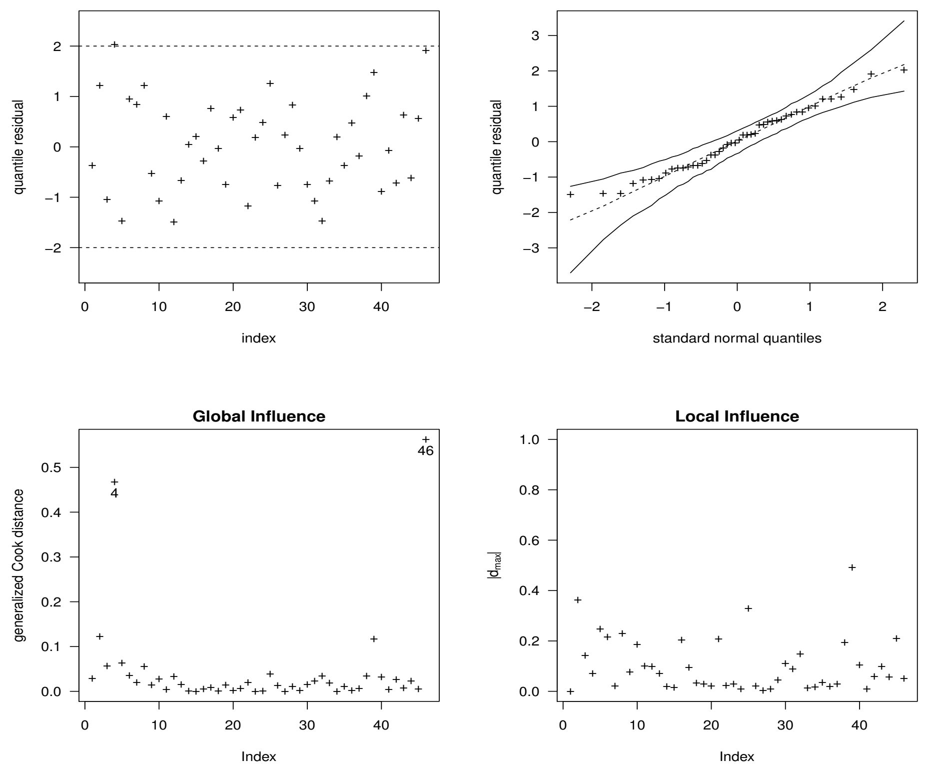

Figure 2: Residuals plots (top), and influence plots (bottom); life of metal pieces data.

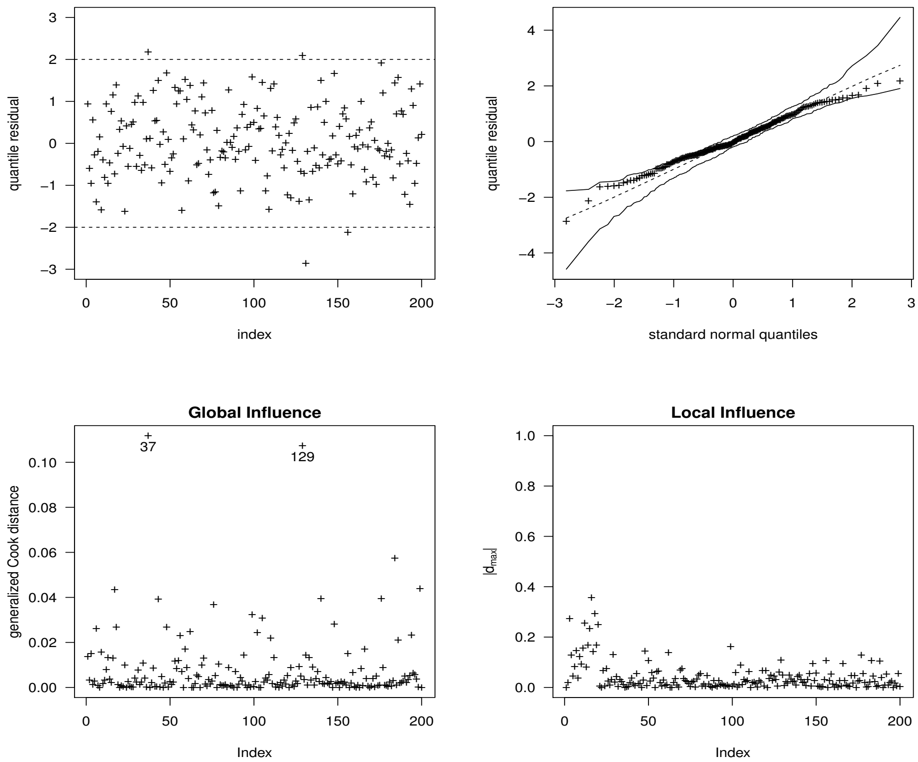

Figure 3: Residuals plots (top), and influence plots (bottom); advertising media data.

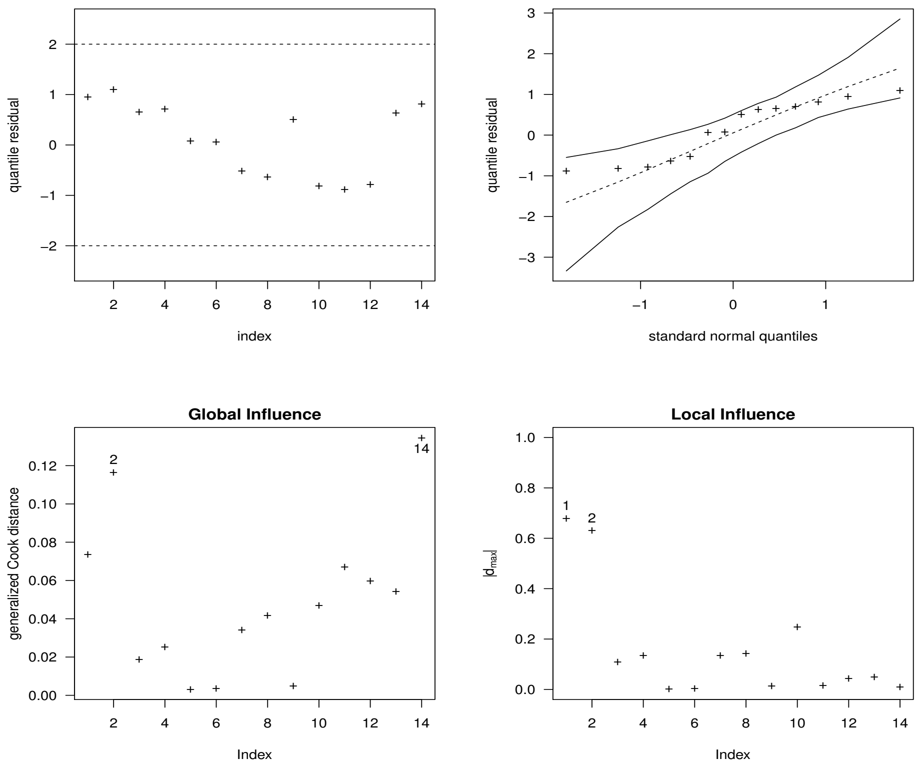

Figure 4: Residuals plots (top), and influence plots (bottom); radioimmunoassay data.

4.1. Competing models

Obviously, there are plenty of regression models in the statistic literature that can be used to model continuous positive response variables. Perhaps the most useful (and simple as well) regression models to deal with positive response variables are the gamma and inverse Gaussian generalized linear models [25]. Beyond the very simple form of these models, the gamma and inverse Gaussian regression models have been quite used in practice mainly because of the well-developed R function `glm()`. The gamma PDF is

$$f(y) = \frac{\phi^{\phi} y^{\phi-1}}{\Gamma(\phi)\mu^{\phi}} \exp\left(-\frac{\phi y}{\mu}\right), \quad y > 0,$$

where $\mu > 0$ is the mean, and $\phi > 0$ is the precision parameter. In the generalized linear model terminology, $\phi^{-1} > 0$ corresponds to the dispersion parameter. We shall use the notation $\mathrm{Ga}(\mu, \phi)$ to refer to this distribution. The gamma distribution reduces to the exponential distribution when $\phi = 1$. If $Y \sim \mathrm{Ga}(\mu, \phi)$, the variance is $\mathbb{VAR}(Y) = \phi^{-1}\mu^2 \propto \mu^2$. Remembering that the variance of the mean-parameterized Maxwell distribution is $0.178\mu^2 \propto \mu^2$ and, hence, the heteroscedastic form based on the gamma and Maxwell distributions are similar, thus modeling the variance of the response variable in a quadratic form. The inverse Gaussian PDF is

$$f(y) = \left(\frac{\phi}{2\pi y^3}\right)^{1/2} \exp\left(-\frac{\phi(y-\mu)^2}{2\mu^2 y}\right), \quad y > 0,$$

where $\mu > 0$ is the mean, and $\phi > 0$ is the precision parameter. We shall use the notation $\mathrm{IG}(\mu, \phi)$ to refer to this distribution. If $Y \sim \mathrm{IG}(\mu, \phi)$, the variance is $\mathbb{VAR}(Y) = \phi^{-1}\mu^3 \propto \mu^3$.

In the following, we fit the gamma and inverse Gaussian regression models to the data previously analyzed using the mean-parameterized Maxwell regression model. For each of the three datasets previously analyzed, we consider the same regression structures for the mean parameter of the gamma and inverse Gaussian models that were considered for the mean of the Maxwell model. The parameter estimates of the gamma and inverse Gaussian parameters are listed in Tables 5 and 6, respectively. Residuals plots are displayed in Figures 5 and 6 for the gamma and inverse Gaussian regression models, respectively. We consider the deviance residual for these regression models, which appear to be a very good choice in the generalized linear model framework [27]. Similar to the mean-parameterized Maxwell regression model, Figure 5 also reveals that the gamma regression model seems to be appropriate to fit these real datasets, once none observation is outside the envelope. On the other hand, the inverse Gaussian regression model appears not suitable to model the advertising media data (some observations are outside the envelope), but it appears suitable to model the other datasets (see Figure 6). At this moment, the natural question is which one is the best in modeling these datasets. The next section addresses this question.

Table 5: Parameter estimates; gamma regression.

Parameter	Life of metal pieces data: $\widehat{\phi}^{-1} = 0.1554$		
	Estimate	SE	95% CI
β_1	12.4449	0.3869	$(11.686; 13.203)$
β_2	-1.6945	0.1076	$(-1.905; -1.484)$

Parameter	Advertising media data: $\widehat{\phi}^{-1} = 0.1318$		
	Estimate	SE	95% CI
β_1	2.7047	0.0443	$(2.618; 2.791)$
β_2	0.0031	0.0010	$(0.001; 0.005)$

Parameter	Radioimmunoassay data: $\widehat{\phi}^{-1} = 0.0990$		
	Estimate	SE	95% CI
β_1	8.6514	0.1071	$(8.441; 8.861)$
β_2	-0.0191	0.0025	$(-0.024; -0.014)$

Table 6: Parameter estimates; inverse Gaussian regression.

Parameter	Life of metal pieces data: $\widehat{\phi}^{-1} = 0.00034$		
	Estimate	SE	95% CI
β_1	11.7484	0.5068	$(10.755; 12.742)$
β_2	-1.5103	0.1280	$(-1.761; -1.260)$

Parameter	Advertising media data: $\widehat{\phi}^{-1} = 0.0079$		
	Estimate	SE	95% CI
β_1	2.7057	0.0438	$(2.620; 2.792)$
β_2	0.0031	0.0010	$(0.001; 0.005)$

Parameter	Radioimmunoassay data: $\widehat{\phi}^{-1} = 3.15\text{e-}05$		
	Estimate	SE	95% CI
β_1	8.5721	0.1259	$(8.325; 8.819)$
β_2	-0.0170	0.0019	$(-0.021; -0.013)$

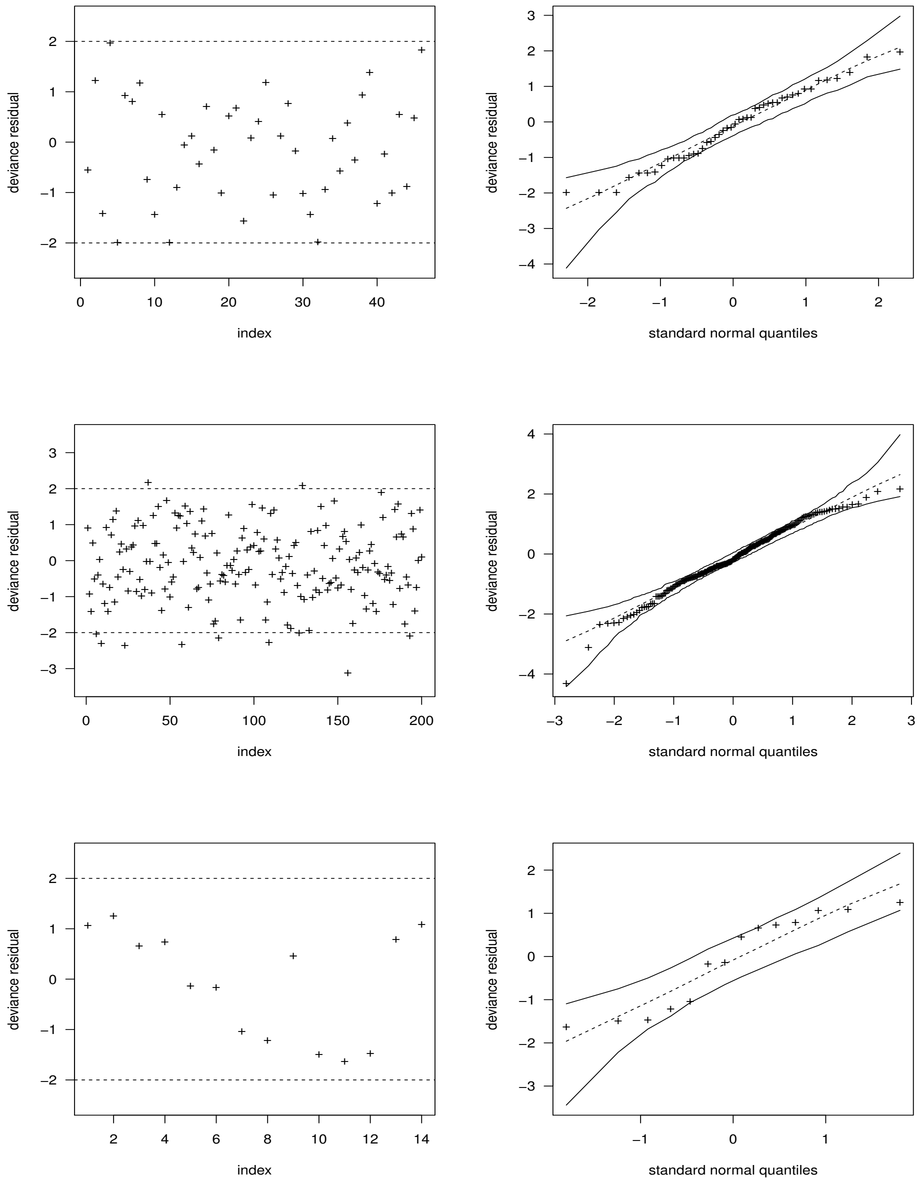

Figure 5: Residuals plots for the gamma regression: life of metal pieces data (top), advertising media data (middle), and radioimmunoassay data (bottom).

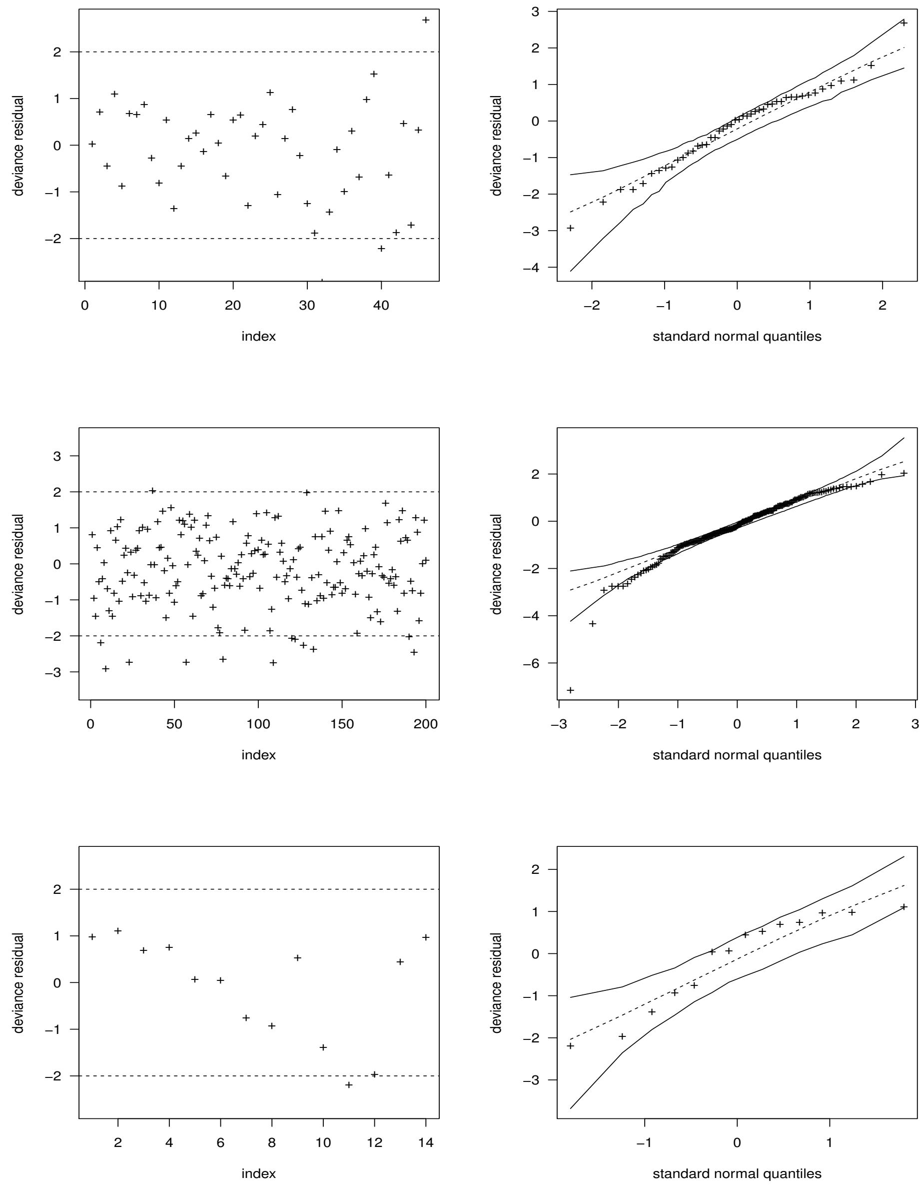

Figure 6: Residuals plots for the inverse Gaussian regression: life of metal pieces (top), advertising media data (middle), and radioimmunoassay data (bottom).

4.2. Choosing the best model

Here, we try to put some light on the following natural question: what is the best regression model to fit the data in the previous sections among the mean-parameterized Maxwell, gamma and inverse Gaussian regression models. It is worth stressing that this question is not easy to be answered in generality. The values of AIC and BIC of all fitted regression models are listed in Table 7. The Maxwell and gamma regressions outperform the inverse Gaussian regression to model the life of metal pieces data as well as the advertising media data, while these three regression models can be considered equivalent to model the radioimmunoassay data. On the basis of AIC and BIC values, it seems that the Maxwell and gamma regression models should be chosen as the best regression models to fit these three datasets. In terms of parsimony, the mean-parameterized Maxwell regression model should be preferable, once it has the advantage of having fewer parameters to be estimated than the gamma regression model. Remembering that in the gamma regression model is necessary to estimate a precision parameter, while in the Maxwell regression model is not.

Table 7: AIC and BIC values.

Model	Life of metal pieces		Advertising media		Radioimmunoassay	
	AIC	BIC	AIC	BIC	AIC	BIC
Maxwell	635.14	638.80	1295.17	1301.77	240.21	241.50
Gamma	635.73	641.22	1292.80	1302.70	239.49	241.41
Inverse Gaussian	647.68	653.17	1322.40	1332.30	241.38	243.30

From now on, we only consider the mean-parameterized Maxwell and gamma regression models. By following with the analysis in order to select the best regression model, we shall consider the generalized likelihood ratio test statistic (V_{LR}) proposed by Vuong [32]. The statistic V_{LR} measures the distance between two models in terms of the Kullback–Leibler information criterion. The test statistic can be expressed as $V_{\mathrm{LR}} = \Lambda\Psi^{-1/2}$, and

$$\Lambda = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\log\left(\frac{\mathrm{Mw}(\widehat{\mu}_i)}{\mathrm{Ga}(\widehat{\mu}_i,\widehat{\phi})}\right),$$

$$\Psi = \frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{\mathrm{Mw}(\widehat{\mu}_i)}{\mathrm{Ga}(\widehat{\mu}_i,\widehat{\phi})}\right)\right]^2 - \left[\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\mathrm{Mw}(\widehat{\mu}_i)}{\mathrm{Ga}(\widehat{\mu}_i,\widehat{\phi})}\right)\right]^2.$$

The statistic V_{LR} converges in distribution to a standard normal distribution under the null hypothesis of equivalence of the models. The null hypothesis is not rejected if $|V_{\mathrm{LR}}| \leq \Phi^{-1}(1-\alpha/2)$, where $\Phi^{-1}(\cdot)$ is the standard normal quantile function, and α is the significance level. On the other hand, we reject at significance level α the null hypothesis in favor of the Maxwell model being better (worse) than the gamma model if $V_{\mathrm{LR}} > \Phi^{-1}(1-\alpha)$ ($V_{\mathrm{LR}} < -\Phi^{-1}(1-\alpha)$). Table 8 lists the observed values of V_{LR} (and the corresponding p-values), indicating that the mean-parameterized Maxwell and gamma regression models are equivalent to fit these datasets. However, in terms of parsimony, the mean-parameterized Maxwell regression model should be preferable as mentioned early. In summary, the results in this section reveal that the mean-parameterized Maxwell regression model can be a good (and simple as well) alternative to the well-developed gamma regression model in practice.

Table 8: Generalized likelihood ratio statistic.

Data	V_{LR}	p-value
Life of metal pieces	−0.9320	0.3513
Advertising media	−0.5243	0.6001
Radioimmunoassay	−1.2102	0.2262

5. CONCLUDING REMARKS

In this paper, based on the mean-parameterized Maxwell distribution, a parametric class of regression models to deal with positive response variables was studied. By employing the frequentist approach, the estimation of the Maxwell regression parameters is conducted by the maximum likelihood method. We also provide a closed-form expression for the expected Fisher information matrix. Monte Carlo simulation experiments reveal that the maximum likelihood method is quite effective to estimate the Maxwell model parameters, and that the initial guesses we recommend for the Maxwell regression parameters worked perfectly well in the Monte Carlo simulations as well as real data applications. We also give a simple formula for calculating bias-corrected maximum likelihood estimates of the mean-parameterized Maxwell regression parameters. We discuss diagnostic techniques (global and local influence, and residuals analysis) for the mean-parameterized Maxwell regression model. Diagnostic methods have been an important tool in regression analysis to detect anomalies with the fitted model, such as departures from the model assumptions, presence of outliers and presence of influential observations. In particular, an appropriate matrix for assessing local influence on the Maxwell parameter estimates under a specific perturbation scheme is obtained. Additionally, we illustrate the methodology developed in this paper by means of applications to real data. We verify through the real data applications that the mean-parameterized Maxwell regression model was superior to the well-known inverse Gaussian regression model, and was very similar to the gamma regression model, which is, probably, the most used regression model to deal with positive response variables in practice. Finally, it is worth stressing that the formulas related with the mean-parameterized Maxwell regression model are manageable (such as log-likelihood function, score function, expected Fisher information matrix, etc.) and with the use of modern computer resources and its numerical capabilities, this regression model may prove to be an useful addition to the arsenal of applied statisticians.

The previous developments regarding the mean-parameterized Maxwell regression model indicate that this model can be indeed very useful in practice. Therefore, we would like to point out that the current work opens new possibilities for future works. In particular, an interesting extension of the mean-parameterized Maxwell regression model which allows for explanatory variables to be measured with error may be developed. Also, one may study the mean-parameterized Maxwell regression model under random effects. Additionally, due to recent advances in computational technology, one may explore other estimation methods for the mean-parameterized Maxwell regression model such as the Bayesian approach. In addition, Bayesian influence diagnostics can also be treated via the Kullback–Leibler divergence and, hence, atypical observations can also be identified in a Bayesian context. A very interesting extension of the developments considered in this paper would be to study the mean-parameterized Maxwell regression model in a semiparametric context. Obviously an in-depth investigation of such studies is beyond the scope of the current paper, but certainly are very interesting topics for future works.

APPENDIX. The R code

```
## R function to estimate the mean-parameterized
## Maxwell parameters (link function = "log")
Maxwell.reg <- function(formula, data){
    cl <- match.call()
    if (missing(data))
      data <- environment(formula)
    mf <- match.call(expand.dots = FALSE)
    m <- match(c("formula", "data"), names(mf), 0L)
    mf <- mf[c(1L, m)]
    mf$drop.unused.levels <- TRUE
    oformula <- as.formula(formula)
    mf$formula <- formula
    mf[[1L]] <- as.name("model.frame")
    mf <- eval(mf, parent.frame())
    mt <- terms(formula, data = data)
    Y <- model.response(mf, "numeric")
    X <- model.matrix(mf)
    if (length(Y) < 1)
      stop("empty model")
    if (!(min(Y) >= 0))
      stop("invalid dependent variable")
    floglikMax <- function(vPar){
      veta <- X%*%vPar
      vmu <- exp(veta)
      loglik <- sum( -3*log(vmu) - (4/pi)*(Y^2/vmu^2) )
      loglik
    }
    fscoreMax <- function(vPar){
      veta <- X%*%vPar
      vmu <- exp(veta)
      vt <- (8/pi)*(Y^2/vmu^2) - 3
      vt <- as.vector(vt)
      score <- t(X)%*%vt
      score
    }
    fFisherMax <- function(){
      6*(t(X)%*%X)
    }
    start <- c( solve(t(X)%*%X)%*%t(X)%*%log(Y+0.1) )
    opt <- optim(start, fn = floglikMax, gr = fscoreMax, method = "BFGS",
                 control=list(fnscale=-1), hessian=FALSE)
    if (opt$conv != 0)
      stop("algorithm did not converge")
    beta <- opt$par
    se <- sqrt(diag(solve(fFisherMax())))
    z.value <- beta/se
    p.value <- 2*(1 - pnorm(abs(z.value)))
    names(beta) <- colnames(X)
    rval <- cbind( round(beta, 6), round(se, 6),
                   round(z.value, 6), round(p.value, 6) )
    colnames(rval) <- c("Estimate", "Std. Error",
                         "z value", "Pr(>|z|)")
    return(rval)
}

## Example: Life of metal pieces data
## y = "number of cycles to failure" and x = "work  per cycle"
data(Biaxial, package="ssym")
attach(Biaxial)
y <- Life
x <- log(Work)
Maxwell.fit <- Maxwell.reg(y ~ x)
Maxwell.fit
             Estimate Std. Error   z value Pr(>|z|)
(Intercept) 12.473253   0.400698  31.12885        0
x           -1.706004   0.111374 -15.31775        0
```

ACKNOWLEDGMENTS

The author would like to thank two anonymous reviewers for their insightful comments and suggestions. The author also acknowledges the financial support of the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq: grant 304776/2019–0).

REFERENCES

[1] Al-Baldawi, T.H.K. (2015). Some Bayes estimators for Maxwell distribution with conjugate informative priors, *Al-Mustansiriyah Journal of Science*, **26**, 64–69.

[2] Atkinson, A.C. (1985). *Plots, Transformation and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Oxford University Press, New York.

[3] Bekker, A. and Roux, J.J.J. (2005). Reliability characteristics of the Maxwell distribution: A Bayes estimation study, *Communications in Statistics – Theory and Methods*, **34**, 2169–2178.

[4] Bourguignon, M.; Leão, J. and Gallardo, D.I. (2019). Parametric modal regression with varying precision, *Biometrical Journal*, **61**, 1–19.

[5] Castellares, F.; Ferrari, S.L.P. and Lemonte, A.J. (2018). On the Bell distribution and its associated regression model for count data, *Applied Mathematical Modelling*, **6**, 172–185.

[6] Chen, Y.; Genovese, C.R.; Tibshirani, R.J. and Wasserman, L. (2016). Nonparametric modal regression, *The Annals of Statistics*, **44**, 489–514.

[7] Cook, R.D. (1977). Detection of influential observation in linear regression, *Technometrics*, **19**, 15–18.

[8] Cook, R.D. (1986). Assessment of local influence, *Journal of the Royal Statistical Society B*, **48**, 133–169.

[9] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

[10] Cox, D.R. and Snell, E.J. (1968). A general definition of residuals (with discussion), *Journal of the Royal Statistical Society B*, **30**, 248–275.

[11] Dar, A.A.; Ahmed, A. and Reshi, J.A. (2017). Bayesian analysis of Maxwell–Boltzmann distribution under different loss functions and prior distributions, *Pakistan Journal of Statistics*, **33**, 419–440.

[12] Dey, S.; Dey, T. and Maiti, S.S. (2013). Bayesian inference for Maxwell distribution under conjugate prior, *Model Assisted Statistics and Applications*, **8**, 193–203.

[13] Dey, S. and Maiti, S.S. (2013). Estimation of the parameter of Maxwell distribution under different loss functions, *Journal of Statistical Theory and Practice*, **4**, 279–287.

[14] Dunn, P.K. and Smyth, G.K. (1996). Randomised quantile residuals, *Journal of Computational and Graphical Statistics*, **5**, 236–244.

[15] Fan, G. (2016). Estimation of the loss and risk functions of parameter of Maxwell distribution, *Science Journal of Applied Mathematics and Statistics*, **4**, 129–133.

[16] Feng, C.; Sadeghpour, A. and Li, L. (2017). Randomized quantile residuals: an omnibus model diagnostic tool with unified reference distribution, *arXiv* preprint arXiv:1708.08527.

[17] Gallardo, D.I.; Gómez-Déniz, E.; Leão, J. and Gómez, H.W. (2020). Estimation and diagnostic tools in reparameterized slashed Rayleigh regression model. An application to chemical data, *Chemometrics and Intelligent Laboratory Systems*, **207**, 104–189.

[18] Green, P.J. and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall, Boca Raton.

[19] Gómez, Y.M.; Gallardo, D.I.; Leão, J. and Gómez, H.W. (2020). Extended exponential regression model: diagnostics and application to mineral data, *Symmetry*, **12**, 2042.

[20] Hossain, M.P.; Sanusi, R.A.; Omar, M.H. and Riaz, M. (2019). On designing Maxwell CUSUM control chart: an efficient way to monitor failure rates in boring processes, *The International Journal of Advanced Manufacturing Technology*, **100**, 1923–1930.

[21] Kassambara, A. (2019). *datarium: Data Bank for Statistical Analysis and Visualization*, R package version 0.1.0.

[22] Leão, J.; Bourguignon, M.; Saulo, H.; Santos-Neto, M. and Calsavara, V. (2021). The negative binomial beta prime regression model with cure rate: application with a aelanoma dataset, *Journal of Statistical Theory and Practice*, **15**, p63.

[23] Lemonte, A.J. and Bazan, J. (2016). New class of Johnson S_B distributions and its associated regression model for rates and proportions, *Biometrical Journal*, **58**, 727–746.

[24] Li, L. (2016). Minimaxestimation of the parameter of Maxwell distribution under different loss functions, *American Journal of Theoretical and Applied Statistics*, **5**, 202–207.

[25] McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, 2nd ed., Chapman & Hall, London.

[26] Menezes, A.F.B.; Mazucheli, J. and Chakraborty, S. (2021). A collection of parametric modal regression models for bounded data, *Journal of Biopharmaceutical Statistics*, **31**, 490–506.

[27] Pierce, D.A. and Schafer, D.W. (1986). Residuals in generalized linear models, *Journal of the American Statistical Association*, **81**, 977–986.

[28] R Core Team (2019). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[29] Rieck, J.R. and Nedelman, J.R. (1991). A log-linear model for the Birnbaum–Saunders distribution, *Technometrics*, **33**, 51–60.

[30] Tiede, J.J. and Pagano, M. (1979). Application of robust calibration to radioimmunoassay, *Biometrics*, **35**, 567–574.

[31] Tyagi, R.K. and Bhattacharya, S.K. (1989). Bayes estimator of the Maxwell's velocity distribution function, *Statistica*, **49**, 563–566.

[32] Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, **57**, 307–333.

[33] Yao, W. and Li, L. (2014). A new regression model: modal linear regression, *Scandinavian Journal of Statistics*, **41**, 656–671.