A q-Exponential regression model

Alexandre G. Patriota

Abstract. In this paper we introduce a q-Exponential regression model for fitting data with discrepant observations. Maximum likelihood estimators for the model parameters and the (observed and expected) Fisher information matrix are derived. Moreover, we also present sufficient conditions to have consistent and asymptotically normally distributed estimators. Simulation studies show that the proposed estimators present good behaviour in the sense of decreasing bias, and symmetric distribution when the sample size increases.

Keywords. q-Exponential distribution, maximum-likelihood estimator, regression model.

1. q-Exponential distribution

The q-Exponential distribution emerges from the nonextensive statistical mechanics introduced by Tsallis [1]. This theory is a generalization of the classical Boltzmann-Gibbs (BG) statistical mechanics. The well-known BG entropy is $S_{BG} = -\int f(x) \ln f(x) dx$, where $f(x)$ is a density function (naturally, the entropy can analogously be defined for the discrete case). Under appropriated constraints, the normal and the exponential distributions maximize the classical entropy S_{BG} for distributions with support on $(-\infty, \infty)$ and $(0, \infty)$, respectively. The generalization proposed by [1] basically considers the q-entropy $S_q = \frac{1-\int f(x)^q dx}{1-q}$ instead of S_{BG}. One can readily see that, the classical BG entropy is recovered when $q \to 1$. For a detailed study of nonextensive statistical mechanics, we refer the reader to [2] and the references therein. An updated bibliography regarding the Tsallis's nonextensive statistical mechanics can be found in the following website http://tsallis.cat.cbpf.br/biblio.htm.

.

Under appropriated mean constrains (see Section 3.5 of [2]), the distribution with support on $(0,\infty)$ which maximizes the q-entropy S_q is of the following type

$$f(y) = \kappa_0(2-q)\bigg(1-(1-q)\kappa_0 y\bigg)^{\frac{1}{1-q}}, \tag{1.1}$$

where $1 < q < 3/2$ (constraint to have a support on $(0,\infty)$ and finite mean) and $\kappa_0 > 0$. There are many reparametrized versions of (1.1) and all of them are essentially the same distribution. For instance,

$$f(y) = \frac{1}{\kappa}\bigg(1-\frac{(1-q^*)}{\kappa}y\bigg)^{\frac{q^*}{1-q^*}}, \tag{1.2}$$

where $q^* = (2-q)^{-1} \in (1,2)$ and $\kappa = [\kappa_0(2-q)]^{-1} > 0$. Also, one can take $\gamma = (q^*-1)^{-1}$ and $\sigma = [(q^*-1)\kappa]^{-1}$, then the following reparametrized density arises

$$f(y) = \frac{\gamma}{\sigma}\bigg(1+\frac{1}{\sigma}y\bigg)^{-(\gamma+1)}, \tag{1.3}$$

where $\gamma > 1$ and $\sigma > 0$. We can work with either densities (1.1)–(1.3), since one is just a reparametrization of the other. The maximum likelihood estimators (MLEs) for the parameters of (1.3) was computed by [3]. By applying the Jacobian rule, we can attain the MLEs for the parameters of (1.1) and (1.2).

Applications for the nonextensive theory proposed by Tsallis have arisen in several fields. For example, applied physics (for modelling: time correlation function of water hydrogen bonds [4], dissipative optical lattices [5], trapped ion interacting with a classical buffer gas [6]), astrophysics (interstellar turbulence [7]), biology (multiple sclerosis magnetic resonance images [8]) and many others [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. More recently, the CMS Collaboration [20] uses the density (1.1) for fitting the charged hadron transverse-momentum. As can be seen, applications for the q-Exponential distribution abound in many fields and, for this reason, a regression model can also be useful for modelling the average behaviour of a random variable Y.

Suppose that Y follows a q-Exponential distribution, then its mean $\mathrm{E}(Y) = \int y f(y)dy$ may vary with other variables. In this context we can formulate a regression model for modelling more precisely the mean $\mathrm{E}(Y)$ taking into account such extra variables. When other variables have influence on the variation of the response variable Y and they are not considered in the model, the estimation of the parameters may be very affected since they produce an extra variability in the response variable that is not predicted by the model. The impact of this extra variability can be controlled by using a regression model.

With simple computations, we notice that the expectation of Y is $\mathrm{E}(Y) = \frac{1}{(3-2q)\kappa_0} = \frac{1}{2-q^*}\kappa = \frac{1}{\gamma-1}\sigma$. Then, taking other parametrization $\mu = \frac{1}{\gamma-1}\sigma$ and $\theta = \gamma - 1$ we obtain

$$f(y) = \frac{\theta+1}{\theta\mu}\bigg(1+\frac{y}{\theta\mu}\bigg)^{-(\theta+2)}, \tag{1.4}$$

where $\mu > 0$ and $\theta > 0$. If Y follows the density given in (1.4), we write in short that $Y \sim qExp(\mu, \theta)$. We will consider that $Y \sim qExp(\mu, \theta)$ throughout this paper.

The relationships among the parameters of models (1.1), (1.2) and (1.4) are:

$$\theta = \frac{2-q^*}{q^*-1} = \frac{3-2q}{q-1}, \quad \mu = \frac{\kappa}{2-q^*} = \frac{1}{(3-2q)\kappa_0} \quad \text{and} \quad \theta\mu = \frac{\kappa}{q^*-1} = \frac{1}{(q-1)\kappa_0}.$$

Thus, on the one hand, if $q \to 1$ ($q^* \to 1$), then $\theta \to \infty$, $\mu \to \kappa_0^{-1}$, $\mu\theta \to \infty$ and we will have the exponential distribution, $f(y) = \kappa_0 \exp(-y\kappa_0)$. On the other hand, if $q \to 3/2$ ($q^* \to 2$), then $\theta \to 0$, $\mu \to \infty$, $\theta\mu \to 2\kappa_0^{-1}$ and the following density emerges $f(y) = 2\kappa_0(2 + y\kappa_0)^{-2}$. Note that for this last situation our regression model are not applied, since $\int_0^\infty y f(y) = \infty$.

This paper is organized as follows. Section 2 introduces the regression model connecting the mean of Y, μ, with other variables. Section 2.1 presents the score functions, the (observed and expected) Fisher information and an iterative procedure to obtain the ML estimates. Section 2.2 gives some conditions and proofs for asymptotic normality of the MLEs. Finally, Section 2.3 applies the results of the paper on the issue of comparing two populations. We end the paper with conclusions and remarks in Section 3.

2. Regression model

The regression model allows to fit data when the mean varies with covariates. It is natural to explain the average behaviour of a random quantity (the mean μ) through other variables. Therefore, let $(Y_1, x_1), \ldots, (Y_n, x_n)$ be a sample, where n is the sample size, Y_i is a unidimensional response variable (the variable we are interest to explain) and x_i is the vector of **non-stochastic** covariates that may have influence on the average behavior of Y_i. All vectors in this paper will be column vectors and to represent a row vector we use the transpose symbol. Therefore, if $x_1 = (x_{11}, \ldots, x_{1p})^\top$ is a column vector, $x_1^\top$ is a row vector.

In this paper we define the following regression model

$$Y_i \overset{ind}{\sim} qExp(\mu_i, \theta) \tag{2.1}$$

for $i = 1, \ldots, n$, where "$\overset{ind}{\sim}$" means "independent distributed as", β is a vector with dimension p, $\mu_i = \mu_i(\beta, x_i)$ is a positive function with known shape which is assumed to be three times continuously differentiable with respect to each element of β. Notice that, when $\theta > 1$ (or equivalently $q < 4/3$), the mean and variance of Y exist and are given, respectively, by $\mathrm{E}(Y_i) = \mu_i$ and $\mathrm{Var}(Y_i) = \frac{(\theta+1)\mu_i^2}{\theta-1}$, which means that the regression model defined in (2.1) is heteroscedastic, that is, the variance varies with i. It is important to observe that when $0 < \theta < 1$ (or $4/3 < q < 3/2$) the mean exist but the variance does not. In this case, discrepant observations may also be modelling by using the q-Exponential regression model.

As we will see, regression model (2.1) allows us to model several different populations with the same θ and it is also possible to test values for linear combinations of β. One immediate generalization of the regression model defined in (2.1) is to consider other covariates z's to explain a possible variation of the parameter θ, thus a natural generalization is to consider a function $\theta_i(\alpha, z_i)$ instead of θ for all $i = 1, \ldots, n$, where α is an unknown vector of parameters to be estimated. In this note, we will not consider this last generalization. If practical applications demand this generalization, it may be a subject for further work.

2.1. Maximum likelihood estimators

In this section we compute the score functions (the first derivatives of the log-likelihood function) and the (observed and expected) Fisher information (the negative of the second derivatives of the log-likelihood function). The estimates can be attained by using the iterative Newton-Raphson algorithm.

We start showing below the joint density function of $Y_1, \ldots, Y_n$ based on our reparametrization (1.4),

$$f(y_1, \ldots, y_n) = \prod_{i=1}^{n} \left\{ \frac{\theta+1}{\theta\mu_i} \left(1 + \frac{y_i}{\theta\mu_i}\right)^{-(\theta+2)} \right\}.$$

The log-likelihood function is then given by

$$\ell(\beta, \theta) = n \log\left(\frac{\theta+1}{\theta}\right) - \sum_{i=1}^{n} \log(\mu_i) - (\theta+2) \sum_{i=1}^{n} \log\left(1 + \frac{y_i}{\theta\mu_i}\right).$$

The MLEs are attained by taking the first derivatives of ℓ with respect to the parameters (i.e., computing the score functions) and set them equal to zero. Hence, the score functions are given by

$$U_\theta = \frac{\partial \ell}{\partial \theta} = \sum_{i=1}^{n} \frac{(\theta^2+3\theta+1)y_i - \mu_i\theta - (\theta^2+\theta)(y_i+\theta\mu_i)\ln\left(1+\frac{y_i}{\theta\mu_i}\right)}{(\mu_i\theta+y_i)(\theta+1)\theta} \tag{2.2}$$

and

$$U_{\beta_k} = \frac{\partial \ell}{\partial \beta_k} = -\sum_{i=1}^{n} \frac{\mu_i\theta - \theta y_i - y_i}{\mu_i(\mu_i\theta+y_i)} \frac{\partial \mu_i}{\partial \beta_k}$$

for $k = 1, \ldots, p$. Define $D_i \equiv D_i(\beta) = (\frac{\partial \mu_i}{\partial \beta_1}, \ldots, \frac{\partial \mu_i}{\partial \beta_p})^\top$, then a matrix version for $U_\beta = (U_{\beta_1}, \ldots, U_{\beta_p})^\top$ is

$$U_\beta = \frac{\partial \ell}{\partial \beta} = -\sum_{i=1}^{n} \frac{\mu_i\theta - \theta y_i - y_i}{\mu_i(\mu_i\theta+y_i)} D_i. \tag{2.3}$$

Next, we compute the (observed and expected) Fisher information, which depends on the second derivatives of ℓ, see Appendix A. The observed Fisher information is

$$J_{\theta,\beta} = -\begin{pmatrix} \frac{\partial^2 \ell}{\partial \theta^2} & \frac{\partial^2 \ell}{\partial \beta \partial \theta}^\top \\ \frac{\partial^2 \ell}{\partial \beta \partial \theta} & \frac{\partial^2 \ell}{\partial \beta \partial \beta^\top} \end{pmatrix}. \tag{2.4}$$

The expected Fisher information $K_{\theta,\beta} = \mathrm{E}(J_{\theta,\beta})$ is given by

$$K_{\theta,\beta} = \sum_{i=1}^{n} \begin{pmatrix} \dfrac{2+\theta^2+\theta}{\theta^2\left(\theta^4+7\theta^3+17(\theta^2+\theta)+6\right)} & \dfrac{2D_i^\top}{(5\theta+\theta^2+6)\,\theta\mu_i} \\ \\ \dfrac{2D_i}{(5\theta+\theta^2+6)\,\theta\mu_i} & \dfrac{(\theta+1)D_iD_i^\top}{(\theta+3)\mu_i^2} \end{pmatrix}, \tag{2.5}$$

see Eqs (A.1), (A.2) and (A.3) of Appendix A. Now, the iterative Newton-Raphson algorithm can be formulated to find the roots of (2.2) and (2.3)

$$\begin{pmatrix} \widehat{\theta}^{(v+1)} \\ \widehat{\beta}^{(v+1)} \end{pmatrix} = \begin{pmatrix} \widehat{\theta}^{(v)} \\ \widehat{\beta}^{(v)} \end{pmatrix} + J^{-1}_{\widehat{\theta}^{(v)},\widehat{\beta}^{(v)}} U_{\widehat{\theta}^{(v)},\widehat{\beta}^{(v)}}, \quad \text{for } v = 0, 1, \ldots, \tag{2.6}$$

where the quantities with the upper script "(v)" are estimates computed in the vth iteration. The observed Fisher information $J_{\theta,\beta}$ is given in (2.4) and $U_{\theta,\beta} = (U_\theta, U_\beta^\top)^\top$. The quatities $J_{\widehat{\theta}^{(v)},\widehat{\beta}^{(v)}}$ and $U_{\widehat{\theta}^{(v)},\widehat{\beta}^{(v)}}$ are $J_{\theta,\beta}$ and $U_{\theta,\beta}$, respectively, computed at $\theta = \widehat{\theta}^{(v)}$ and $\beta = \widehat{\beta}^{(v)}$. We can replace the observed Fisher information with the expected Fisher information $K_{\theta,\beta}$, given in (2.5), in the above Newton-Raphson algorithm. To start with the iterative process we must insert initial values $\widehat{\theta}^{(0)}$ and $\widehat{\beta}^{(0)}$. **Here, μ_i may be a complicated function of the parameter vector β and, then, it is difficult to suggest starting values for β and θ. Also, it is known that good initial estimates are required for highly nonlinear models, however, there is no standard procedure for attaining $\sqrt{n}$-consistent initial estimates.** In these cases, the user can define a starting value for β by looking at (x, y)-dispersion graphs and take θ as a big number. These starting values are obtained by approximating the q-Exponential distribution to the standard exponential one (i.e., when $q \to 1$). Other initial values can also be defined, if there exist a function $g(\cdot)$ that linearizes μ_i, i.e., $g(\mu_i) = \beta^\top x_i$, then we can also transform the data $g(Y_i)$ and apply the ordinary least square approach for estimating the vector of parameters β to get its initial value. If a generic distance between the estimates computed at the vth and $(v-1)$th iterative step is sufficiently small, we stop the iterative procedure.

In the case that the mean function is constant $\mu_i(\beta, x_i) = \mu$, then it is possible to derive initial values based on the method-of-moments approach. Noting that, under $\mu_i(\beta, x_i) = \mu$,

$$\mathrm{E}(Y_i^k) = \frac{\mu^k\theta^k\Gamma(k+1)\Gamma(\theta+1-k)}{\Gamma(\theta+1)},$$

by the method-of-moments approach, $\widehat{\theta}^{(0)}$ may be taken as the solution of

$$\frac{m_k}{m_1^k} = \frac{\theta^k\Gamma(k+1)\Gamma(\theta+1-k)}{\Gamma(\theta+1)}$$

for some $0 < k \leq 1$, where $m_k = n^{-1}\sum_i Y_i^k$. Notice that, for all $0 < k \leq 1$, the moments above are well defined for all $\theta > 0$. **We warn the reader that these moment estimators may be highly non-robust with significant bias and low**

precision, see for instance [21] for a detailed work on bias and accuracy of moment estimators. [22] presents a formal theory based on contraction mappings that guarantees convergence for the iterative process (2.6) when started even at non-consistent initial values.

In general, the parameter q has an interpretation in statistical mechanics. It is known as the distortion parameter, since if $q \neq 1$ the q-entropy is not additive, see [2] for further details. Therefore, the user might be interested in an estimator for this distortion parameter. By using the invariance property of the MLEs, the MLE for q is easily attained through the MLE of θ. Suppose that we have the MLE for θ, $\widehat{\theta}$ say, then the MLE for q is $\widehat{q} = \frac{\widehat{\theta}+3}{\widehat{\theta}+2}$ which implies that $\widehat{q} \in (1, 3/2)$ for all $\widehat{\theta} \in (0, \infty)$.

2.2. Asymptotic normality for the MLEs

The asymptotic normality of the estimators attained equating the score functions to zero is assured under some regularity conditions. In this section, we establish some regular conditions on the functions μ_i for $i = 1, \ldots, n$. Define $\gamma = (\theta, \beta^\top)^\top$.

- C1 (Identifiability condition) The covariates x_i and the functions $\mu_i(\beta, x_i) > 0$ for $i = 1, \ldots, n$ are such that, if $\gamma_1 \neq \gamma_2$ then $\ell(\gamma_1) \neq \ell(\gamma_2)$ for $\gamma_1, \gamma_2 \in (0, \infty) \times \mathbb{R}^p$.
- C2 (Differentiation condition) The functions $\mu_i(\beta, x_i)$, for $i = 1, \ldots, n$, are three times continuously differentiable.
- C3 (Finite asymptotic Fisher information) The functions $\mu_i(\beta, x_i)$, for $i = 1, \ldots, n$, are such that the limiting matrices $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \frac{D_i}{\mu_i}$ and $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \frac{D_i D_i^\top}{\mu_i^2}$ exist.
- C4 (Finite dimensional condition) The dimension of the vector β does not increase with the sample size.
- C5 The following matrix has rank p for all $n > p$,

$$D_\beta = \left(\frac{D_1}{\mu_1}, \ldots, \frac{D_n}{\mu_n}\right)^\top.$$

Proposition 2.1. *Under C5, the expected total Fisher information $K_{\theta,\beta}$ is positive definite for all $n \geq 1$, i.e, $w^\top(K_{\theta,\beta})w > 0$ for all $w \in \mathbb{R}^{p+1}$ and $w \neq 0$.*

Proof. Let 1_n be a vector of ones and I_n be the $n \times n$ identity matrix, define

$$H_\theta = \begin{pmatrix} \dfrac{2+\theta^2+\theta}{\theta^2\left(\theta^4+7\theta^3+17\theta^2+17\theta+6\right)} & \dfrac{2}{(5\theta+\theta^2+6)\,\theta} \\[2ex] \dfrac{2}{(5\theta+\theta^2+6)\,\theta} & \dfrac{\theta+1}{(\theta+3)} \end{pmatrix}$$

and $F_\beta = \mathrm{diag}(1_n,\ D_\beta)$. Then, the expected Fisher information can be shortly written as

$$K_{\theta,\beta} = F_\beta^\top (H_\theta \otimes I_n) F_\beta \qquad (2.7)$$

where and "$\otimes$" is the kronecker product of matrices. The proof of Proposition 2.1 is given by noting that the diagonal elements of H_θ are positive and that

$$\det(H_\theta) = \frac{1}{(\theta+2)\,(5\,\theta+\theta^2+6)\,(\theta+1)} > 0,$$

then the matrix $H_\theta \otimes I_n$ has full rank and is positive definite. Provided that $\text{rank}(D_\beta) = p$, we have that $\text{rank}(F_\beta) = p+1$ and then $K_{\theta,\beta}$ is positive definite for all $n \geq 1$. □

Proposition 2.2. *Let $f(y)$ be as defined in (1.4). Then, for all $\mu > 0$ and $\theta > 0$, there exist integrable functions g_j, $j = 1,2,3,4,5$ such that $\left|\frac{\partial f(y)}{\partial \theta}\right| < g_1(y)$, $\left|\frac{\partial f(y)}{\partial \mu}\right| < g_2(y)$, $\left|\frac{\partial^2 f(y)}{\partial \theta^2}\right| < g_3(y)$, $\left|\frac{\partial^2 f(y)}{\partial \mu \partial \theta}\right| < g_4(y)$ and $\left|\frac{\partial^2 f(y)}{\partial \mu^2}\right| < g_5(y)$ where $\int_0^\infty g_{1k}(y)dy < \infty$ and $\int_0^\infty g_{2j}(y)dy < \infty$ for $k = 1,2$ and $j = 1,2,3$.*

Proposition 2.3. *Let $\widehat{\gamma}$ be the MLE of $\gamma = (\theta, \beta^\top)^\top$. Under C1–C5, (i) $\widehat{\gamma}$ is consistent and asymptotically it is the unique maximizer of $\ell(\gamma)$ and*

$$(ii) \quad \sqrt{n}(\widehat{\gamma} - \gamma) \xrightarrow{D} \mathcal{N}_{p+1}(0, K_\gamma^{*-1}),$$

where "$\xrightarrow{D}$" means "converges in distribution to", $K_\gamma^* = \lim_{n\to\infty} \bar{K}_\gamma$, $\bar{K}_\gamma = n^{-1}K_\gamma$ and $\mathcal{N}_m(0,\Sigma)$ is a m-variate normal distribution with zero mean and covariance matrix Σ.

Proof. By the Proposition 2.2 and condition C2, we can use the Fubini's theorem to change the order of integrations and we can also change the order of integration and differentiation. This implies that the expectations of the score functions equal zero and that the inverse of the Fisher information is the variance of the score function. Under C3 and C5, the limit of the Fisher information exist and it is positive definite then, by C1, we have that $\widehat{\gamma}$ is consistent and, asymptotically, it is the unique point that maximizes $\ell(\gamma)$. The proof for the second part of the Proposition 2.3 follows by noting that C1–C5 are the required conditions stated in [23], Ch.9., to have MLEs asymptotically normally distributed. The MLE for the other reparametrized versions have also asymptotic normal distributions, the proof follows by using the delta method, see [24]. □

Notice that, the asymptotic distribution of the MLEs does not depend if the variance of Y_i exist for all $i = 1,\ldots,n$ or if $\lim_{n\to\infty} n^{-1}\sum_i \text{Var}(Y_i) < \infty$. This happens because the score function and the observed Fisher information are functions of $y_i^k/(\theta\mu_i + y_i)^j$ for $k = 0,1,2$ and $j = 1,2$ rather than a function of y_i^k. As $y_i^k/(\theta\mu_i + y_i)^j$ for $k = 0,1,2$ and $j = 1,2$ have finite expectations we have to impose conditions on the limiting average of these expectations (weighed by D_i).

We can test the hypothesis $\mathcal{H}_0 : C\gamma = d$ by using the following Wald statistic

$$\xi = n(C\widehat{\gamma} - d)^\top \left[C\bar{K}_{\widehat{\gamma}}^{-1}C^\top\right]^{-1}(C\widehat{\gamma} - d) \tag{2.8}$$

which converges in distribution to a chisquare distribution with $c = rank(C)$ degrees of freedom. The statistic (2.8) can be used to test linear combinations of

β. It is very useful when several treatments are considered and we want to verify if all treatment effects (or some of them) are equal to the innocuous one.

We remark that the null hypothesis must not be on the border of the parametric space. For example, we cannot test if $\theta = 0$ or $\theta = \infty$ (these tests are equivalent to test if $q = 3/2$ or $q = 1$, respectively). The likelihood ratio (LR) statistic can also be formulated for testing more general hypothesis such as $\mathcal{H}_0 : \gamma \in \Gamma_0$. Then, if Γ_0 is a smooth subset of $\Gamma = (0, \infty) \times \mathbb{R}^p$ and the true parameter γ_0 is an interior point of Γ_0, the following LR statistic

$$LR = 2\Big(\ell(\widehat{\gamma}) - \ell(\widehat{\gamma}_0)\Big). \tag{2.9}$$

is asymptotically chisquared distributed with $c = p + 1 - \dim(\Gamma_0)$ degrees of freedom, where $\widehat{\gamma}_0 = \arg\max_{\gamma\in\Gamma_0} \ell(\gamma)$, this result and others (when Γ_0 is a semi-algebraic set with singularities) can be studied in [25] and the references therein.

2.3. Example for two populations

Let $Z_1, \ldots, Z_{n_1}$ and $W_1, \ldots, W_{n_2}$ be two observed samples taken from two different populations Z and W, respectively. Assume that $Z \sim qExp(\mu_z, \theta)$ and $W \sim qExp(\mu_w, \theta)$. Therefore, we can define $Y_1, \ldots, Y_n$, with $n = n_1 + n_2$, where $Y_i = Z_i$ for $i = 1, \ldots, n_1$ and $Y_i = W_i$ for $i = n_1 + 1, \ldots, n$. Let x_i be a covariate which is 0 for $i = 1, \ldots, n_1$ and 1 for $i = n_1 + 1, \ldots, n$. In this case, the covariate is just indicating where population the observation come from.

Define $\mu_i = \exp(\beta_0 + \beta_1 x_i)$, thus, we have that, $\exp(\beta_0) = \mu_z$ for $i = 1, \ldots, n_1$ and $\exp(\beta_0 + \beta_1) = \mu_w$ for $i = n_1 + 1, \ldots, n$. Hence, we can estimate β_0, β_1 and θ via our regression model and then return to the original parameters μ_z, μ_w and θ, if necessary. One may be interested in verifying if these two samples come from the same underlying population. This hypothesis is equivalent to test if $\beta_1 = 0$. The score functions for this model are

$$U_\theta = \sum_{i=1}^{n_1} \frac{(\theta^2 + 3\theta + 1)z_i - \mu_z\theta - (\theta^2 + \theta)(\theta\mu_z + z_i)\ln\Big(1 + \frac{z_i}{\theta\mu_z}\Big)}{(\mu_z\theta + z_i)\,(\theta + 1)\,\theta} +$$
$$+ \sum_{i=1}^{n_2} \frac{(\theta^2 + 3\theta + 1)w_i - \mu_w\theta - (\theta^2 + \theta)(\theta\mu_w + w_i)\ln\Big(1 + \frac{w_i}{\theta\mu_w}\Big)}{(\mu_w\theta + w_i)\,(\theta + 1)\,\theta},$$

$$U_{\beta_0} = -\sum_{i=1}^{n_1} \frac{\mu_z\theta - \theta z_i - z_i}{\mu_z\theta + z_i} \quad \text{and} \quad U_{\beta_1} = -\sum_{i=n_1+1}^{n} \frac{\mu_w\theta - \theta w_i - w_i}{\mu_w\theta + w_i},$$

where $\mu_z = \exp(\beta_0)$ and $\mu_w = \exp(\beta_0 + \beta_1)$. We have also that,

$$D_\beta = \begin{pmatrix} 1_{n_1} & 0 \\ 1_{n_2} & 1_{n_2} \end{pmatrix} \quad \text{and} \quad F_\beta = \mathrm{diag}(1_n, D_\beta)$$

and then the expected Fisher information is

$$K_{\theta,\beta} = \begin{pmatrix} \dfrac{(2+\theta^2+\theta)n}{\theta^2\left(\theta^4+7\theta^3+17\theta^2+17\theta+6\right)} & \dfrac{2n}{\left(5\theta+\theta^2+6\right)\theta} & \dfrac{2n_2}{\left(5\theta+\theta^2+6\right)\theta} \\ \dfrac{2n}{\left(5\theta+\theta^2+6\right)\theta} & \dfrac{(\theta+1)n}{(\theta+3)} & \dfrac{(\theta+1)n_2}{(\theta+3)} \\ \dfrac{2n_2}{\left(5\theta+\theta^2+6\right)\theta} & \dfrac{(\theta+1)n_2}{(\theta+3)} & \dfrac{(\theta+1)n_2}{(\theta+3)} \end{pmatrix}$$

and its inverse is

$$K^{-1}_{\theta,\beta} = \begin{pmatrix} \dfrac{(\theta+1)^2(\theta+2)^2}{n} & -\dfrac{2(\theta+1)(\theta+2)}{n\theta} & 0 \\ -\dfrac{2(\theta+1)(\theta+2)}{n\theta} & \dfrac{n_2\theta^2(\theta+3)+n_1(\theta+2)(2+\theta^2+\theta)}{nn_1\theta^2(\theta+1)} & -\dfrac{\theta+3}{n_1(\theta+1)} \\ 0 & -\dfrac{\theta+3}{n_1(\theta+1)} & \dfrac{(\theta+3)n}{n_2n_1(\theta+1)} \end{pmatrix},$$

where $\beta = (\beta_0, \beta_1)^\top$. Therefore, the diagonal elements of the inverse of $K_{\theta,\beta}$ are the asymptotic variances of the ML estimators $\widehat{\theta}$, $\widehat{\beta}_0$ and $\widehat{\beta}_1$ which are given by

$$\sigma^2_{\widehat{\theta}} = \frac{(\theta+1)^2(\theta+2)^2}{n}, \qquad \sigma^2_{\widehat{\beta}_0} = \frac{n_2\theta^2(\theta+3)+n_1(\theta+2)(2+\theta^2+\theta)}{nn_1\theta^2(\theta+1)}$$

and

$$\sigma^2_{\widehat{\beta}_1} = \frac{(\theta+3)n}{n_2n_1(\theta+1)}.$$

These variances can be estimated replacing the parameters by their MLEs. For instance, to test if $\beta_1 = 0$ we define $C = (0, 0, 1)$, $\gamma = (\theta, \beta_0, \beta_1)^\top$ and $d = 0$, then we have $C\gamma = \beta_1$. Replacing these quantities in the Wald statistic (2.8), we arrive at the following statistic

$$\xi^2 = \frac{\widehat{\beta}_1^2}{\widehat{\sigma}^2_{\widehat{\beta}_1}}, \tag{2.10}$$

which is asymptotically chisquared distributed, where rank of C is 1. One must compute the p-value, $P(\chi^2(1) > \xi^2) =$ p-value, where $\chi^2(1)$ is a chisquare distribution with one degree of freedom. If p-value is lesser than $\alpha\%$, we say that $\beta_1 \neq 0$ with level of significance equal to $\alpha\%$.

Note that, the asymptotic variance of $\widehat{\beta}_1$ is not much affected by the values of θ, since $(\theta+3)/(\theta+1) \in (1, 3)$ for all $\theta > 0$. That is, the asymptotic behaviour of the MLE $\widehat{\beta}_1$ is well behaved for any values of θ. Therefore, the hypothesis testing with respect to β_1 is sufficiently reliable (this fact is also verified by simulations in Section 2.3.1). However, the asymptotic variances $\sigma^2_{\widehat{\theta}}$ and $\sigma^2_{\widehat{\beta}_0}$ are affected by the values of θ. On the one hand, the asymptotic standard deviation of $\sqrt{n}(\widehat{\theta} - \theta)$ is of order θ^2, hence, the larger θ the more imprecise is its MLE. For instance, if $\theta = 100$ ($q \approx 1.0098$) then the asymptotic standard error of $\sqrt{n}(\widehat{\theta} - \theta)$ is of order around 10 000. On the other hand, the asymptotic standard deviation of $\sqrt{n}(\widehat{\beta}_0 - \beta_0)$ is of order θ^{-2}, hence, the smaller θ the more imprecise is its MLE. For instance, if $\theta = 0.001$ ($q \approx 1.49975$) then the asymptotic standard error of $\sqrt{n}(\widehat{\beta}_0 - \beta_0)$ is of order around 10 000. To overcome such imprecisions, one must increase the sample size. At last, if the interest lies in testing β_1, our regression model can be used without any worries about θ. However, if the interest lies in θ and β_0, we suggest using our results with caution, since the estimates $\widehat{\theta}$ and $\widehat{\beta}_0$ may be very unstable. One can use other procedure to consistently estimate θ when θ is near of the border of the parametric space.

Based on Example 2.3, it is easy to implement the same procedure for $k \geq 2$ populations. We remark that the covariate x_i can be continuous or discrete quantities.

2.3.1. Small simulation study. In this section we conduct a small simulation study to verify the results of this paper. We consider the two-populations model as defined in the previous section and we generate 10 000 Monte Carlo samples to study the behaviour of the maximum likelihood estimates and the Wald statistic for moderate sample sizes.

The simulation was done considering the following configuration: $n = 20$, 30, 50, 150, 500, $\theta = 9^{-1},\ 1,\ 9$, $\beta_0 = \log(3)$ and $\beta_1 = 0$. This parameter values imply that $q = 1.09, 1.33, 1.47$ (when $\theta = 9, 1, 9^{-1}$, respectively) and $\mu_z = \mu_w = 3$. Then, for each combination $(n, \theta, \beta_0, \beta_1)$ we generate $N = 10\ 000$ Monte Carlo samples and for each sample and parameter setting, we compute the MLEs $(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_N)$ and the p-value associated with the hypothesis $\mathcal{H}_0 : \beta_1 = 0$. For each parameter setup we compute the median of the maximum likelihood estimates. Since for small samples the distribution of the MLEs may be very asymmetric or bimodal, the median is a good measure of the location of the true distribution of the ML estimates. Here, we present the MLE for q instead of θ. We remark that when $\theta = 9^{-1}$ and $\theta = 1$ the data contains discrepant observations.

The p-value is computed as $P(\chi^2(1) > \xi_j^2) = \text{p-value}_j$, for $j = 1, \ldots, N$, where ξ_j^2 is the statistic (2.10) computed for the jth Monte Carlo sample. If the Wald and LR statistics are chisquared distributed, the distributions of the respective p-values under the null hypothesis must be close to the uniform distribution on $(0, 1)$. From the theory, we known that it happens when $n \to \infty$, but for finite sample sizes, the Wald and LR statistics may behave very different from the chisquare distribution.

Tables 1 and 2 present the median of the maximum likelihood estimates for each setup. As can be seen, the median values are closer to the true values. Moreover, the larger the sample size, the closer to the true values are the median values. Table 3 depicts the rejection rates under the null hypothesis considering a nominal level of 5%. As expected, all empirical values are around 5%.

TABLE 1. Median of the maximum likelihood estimators based on a 10 000 Monte Carlo samples when $n = 20,\ 30,\ 50$. The empirial standard errors are in paranteses.

	n		
	20	30	50
$q = 1.47$	1.4169 (0.16)	1.4392 (0.12)	1.4532 (0.08)
β_0	0.2936 (2.48)	0.4871 (2.34)	0.6684 (2.16)
β_1	0.0048 (0.76)	-0.0108 (0.62)	-0.0030 (0.47)
$q = 1.33$	1.2320 (0.17)	1.2768 (0.15)	1.3018 (0.12)
β_0	0.9987 (1.16)	1.0265 (0.87)	1.0501 (0.60)
β_1	-0.0148 (0.65)	-0.0096 (0.53)	0.0016 (0.41)
$q = 1.09$	1.0000 (0.11)	1.0000 (0.10)	1.0191 (0.09)
β_0	1.0556 (0.39)	1.0705 (0.29)	1.0830 (0.22)
β_1	0.0043 (0.51)	0.0092 (0.41)	-0.0013 (0.31)

TABLE 2. Median of the maximum likelihood estimators based on a 10 000 Monte Carlo samples when $n = 100,\ 150,\ 500$. The empirial standard errors are in paranteses.

	n		
	50	150	500
$q = 1.47$	1.4650 (0.05)	1.4677 (0.04)	1.4719 (0.02)
β_0	0.8840 (1.90)	0.9288 (1.72)	1.0464 (1.20)
β_1	-0.0041 (0.33)	-0.0037 (0.27)	0.0049 (0.15)
$q = 1.33$	1.3191 (0.08)	1.3234 (0.06)	1.3309 (0.03)
β_0	1.0786 (0.33)	1.0824 (0.25)	1.0930 (0.13)
β_1	-0.0023 (0.28)	-0.0019 (0.23)	0.0020 (0.13)
$q = 1.09$	1.0605 (0.07)	1.0699 (0.06)	1.0858 (0.04)
β_0	1.0860 (0.16)	1.0924 (0.13)	1.0969 (0.07)
β_1	0.0058 (0.22)	0.0000 (0.18)	0.0004 (0.10)

Figures 1–2, 3–4 and 5–4 present the histogram for $\widehat{q}$, $\widehat{\beta}_0$ and $\widehat{\beta}_1$, respectively. In each Figure there are nine graphs considering different values for (n, q). As can be seen, when n increases all estimates seem to converge to symmetric distributions.

Figures 1 and 2 show that the MLE of $q = \frac{\theta+3}{\theta+2}$ has a bad performance when $q = 1.09$ or 1.47 and the sample size is small ($n = 20, 30$) and moderate ($n = 50, 100, 150$), i.e., the distributions of the estimates are not symmetric and unimodal. But, when $q = 1.33$, the MLE $\widehat{q}$ is symmetric and well behaved even for small values of n. In Figures 3 and 4 we can see that for small values of q (or large values of θ), the MLE for β_0 is symmetric around $\log(3)$ for all sample

Table 3. Rejection rates (at 1%, 5% and 10% of nominal levels) under the null hypothesis $\beta_1 = 0$ for the Wald statistic.

	Sample size					
	20	30	50	100	150	500
1%						
$q = 1.09$	0.97	0.95	0.84	0.93	0.83	1.01
$q = 1.33$	0.96	0.93	0.93	0.89	1.10	0.92
$q = 1.47$	0.95	0.82	0.85	0.87	1.00	0.97
5%						
$q = 1.09$	4.36	4.45	4.56	4.76	4.96	5.19
$q = 1.33$	4.27	4.63	4.74	4.53	5.12	5.20
$q = 1.47$	4.64	4.74	4.64	4.58	4.71	4.74
10%						
$q = 1.09$	8.70	9.28	9.42	9.61	9.59	10.47
$q = 1.33$	8.99	9.18	9.29	9.35	10.28	10.22
$q = 1.47$	9.37	9.69	9.60	9.49	9.72	9.75

Table 4. Rejection rates (at 1%, 5% and 10% of nominal levels) under the null hypothesis $\beta_1 = 0$ for the LR statistic.

	Sample size					
	20	30	50	100	150	500
1%						
$q = 1.09$	1.69	1.64	1.88	2.27	2.18	2.03
$q = 1.33$	1.12	0.95	1.01	0.88	1.11	0.95
$q = 1.47$	0.97	0.84	0.82	0.89	1.05	0.93
5%						
$q = 1.09$	4.69	4.85	5.27	5.48	5.64	5.75
$q = 1.33$	4.40	4.73	4.78	4.57	5.18	5.23
$q = 1.47$	4.82	4.80	4.78	4.61	4.68	4.83
10%						
$q = 1.09$	9.18	9.54	9.73	10.14	9.98	10.62
$q = 1.33$	9.25	9.45	9.38	9.39	10.30	10.29
$q = 1.47$	9.39	9.91	9.70	9.51	9.75	9.74

sizes. However, when q increases the performance of $\widehat{\beta}_0$ becomes problematic (its distribution is bimodal for small values of n). Finally, Figures 5 and 6 show that the MLE for β_1 is always symmetric around zero for all chosen values of q and n.

3. Conclusions and Remarks

In this paper we proposed a q-Exponential regression model. Some sufficient conditions to have normality for the proposed estimators are established. We specialized our regression model to the issue of modelling two populations and we also presented a test to verify if the data came from the same population. The simulation studies showed that the proposed estimators present good behaviour in the sense of decreasing bias, and symmetric distribution when n increases. The rejection rates (at 1%, 5% and 10% of nominal levels) of the Wald and LR statistics under the null hypothesis ($\beta_1 = 0$) are all close to the adopted nominal levels.

The regression model introduced here can be generalized in some aspects. For instance, a new model must be studied if the covariates are subject to measurement errors or if the responses have a particular structure of dependence (longitudinal studies). Also, to overcome the problem in the estimation of the parameter θ (or q), one can propose a bayesian methodology.

When $q = 1$, the q-Exponential distribution becomes the usual exponential one which is a particular case of the generalized linear models. Hence, the regression model for $q = 1$ was intensively studied in the statistical literature, see for instance the two classical references [27] and [28]. When $q \to 3/2$, the density that emerges is $f(y) = 2\kappa_0(2 + y\kappa_0)^{-2}$, for this case $yf(y)$ is not integrable.

It is noteworthy that, the distribution with support on $\mathbb{R}$ that maximizes the q-entropy is

$$f(y) = K_q\Big(1 - \alpha(1-q)(y-\mu)^2\Big)^{\frac{1}{1-q}}, \tag{3.1}$$

where K_q is a normalizing constant, $1 \le q < 3$ and $\mathrm{E}(Y) = \mu$ is the location parameter usually takes equal to zero. We remark that, taking $v = \frac{3-q}{q-1}$ and $\sigma^2 = \frac{1}{\alpha(3-q)}$ the density (3.1) becomes the well-known Student-t distribution

$$f(y) = K_v\left(1 + \frac{(y-\mu)^2}{v\sigma^2}\right)^{-\frac{v+1}{2}},$$

where K_v is a normalizing constant. The above distribution lies in the elliptical class of distributions that has been extensively studied in the statistical literature. In fact, some important references for elliptical regression models in the multivariate context (that are sufficient general to hold many practical applications) are [29], [30], [31], [32], [33], [34], [35], [36], [37] and more recently [38, 39]. For this reason, this paper did not consider the distribution (3.1). One can use those very general results already derived in the statistical literature for modelling $\mathrm{E}(Y_i) = \mu_i(\beta, x_i)$ and return to the main parametrization taking $q = \frac{v+3}{v+1}$ and $\alpha = \frac{v+1}{2\sigma^2 v}$.

Appendix A. Second derivatives of the log-likelihood function

The second derivatives are given by

$$\frac{\partial^2 \ell}{\partial \theta^2} = \sum_{i=1}^{n} \frac{-{y_i}^2 + {\mu_i}^2\theta^2 + 2{\mu_i}^2\theta^3 - 2\theta\mu_i y_i - 4\,\theta^2\mu_i y_i - \theta {y_i}^2 - 4\,\theta^3\mu_i y_i + \theta^3 {y_i}^2}{\left(\mu_i\theta + y_i\right)^2\left(\theta+1\right)^2\theta^2},$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \theta} = -\sum_{i=1}^{n} \frac{\left(2\mu_i - y_i\right) y_i}{\mu_i\left(\mu_i\theta + y_i\right)^2} D_i \quad \text{and} \quad \frac{\partial^2 \ell}{\partial \beta \partial \beta^\top} = \sum_{i=1}^{n} w_{1i} D_i D_i^\top - \sum_{i=1}^{n} w_{2i} V_i$$

where $V_i = \partial D_i / \partial \beta^\top$,

$$w_{1i} = \frac{1}{\mu_i^2} + \frac{\theta+2}{\mu_i^2}\left(\frac{y_i^2}{\left(\theta\mu_i + y_i\right)^2} - 2\frac{y_i}{\theta\mu_i + y_i}\right) \quad \text{and} \quad w_{2i} = \frac{\mu_i\theta - \theta y_i - y_i}{\mu_i\left(\mu_i\theta + y_i\right)}.$$

In order to find the expected Fisher information, we now must take the following expectations.

$$\mathrm{E}\left(\frac{y_i}{\theta\mu_i + y_i}\right) = \frac{1}{\theta+2}, \quad \mathrm{E}\left(\frac{y_i}{(\theta\mu_i + y_i)^2}\right) = \frac{\theta+1}{(\theta+2)(\theta+3)\theta\mu_i} \tag{A.1}$$

and

$$\mathrm{E}\left(\frac{y_i^2}{(\theta\mu_i + y_i)^2}\right) = \frac{2}{(\theta+2)(\theta+3)}. \tag{A.2}$$

Hence,

$$\mathrm{E}(w_{1i}) = -\frac{\theta+1}{(\theta+3)\mu_i^2} \quad \text{and} \quad \mathrm{E}(w_{2i}) = 0. \tag{A.3}$$

References

[1] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics*, **52**, 479–487.

[2] Tsallis, C. (2009). *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World.* Springer.

[3] Shalizi, C.R. (2007). Maximum likelihood estimation for q-exponential (Tsallis) distributions, http://arxiv.org/abs/math/0701854

[4] Campo, M.G., Ferri, G.L., Roston, G.B. (2009). *q*-exponential distribution in time correlation function of water hydrogen bonds, *Brazilian Journal of Physics*, 39, 439–443.

[5] Douglas, P., Bergamini, D., Renzoni, F. (2006). Tunable Tsallis distributions in dissipative optical lattices, *Physical Review Letters*, 96, 110601.

[6] DeVoe, R.G. (2009). Power-law distributions for a trapped ion interacting with a classical buffer gas, *Physical Review Letters*, 102, 063001.

[7] Esquivel, A., Lazarian, A. (2010). Tsallis statistics as a tool for studying interstellar turbulence, *The Astrophysical Journal*, 710, 125–132.

[8] Diniz, P.R.B., Murta-Junior, L.O., Brum, D.G, Arajo, D.B., Santos, A.C. (2010). Brain tissue segmentation using q-entropy in multiple sclerosis magnetic resonance images *Brazilian Journal of Medical and Biological Research*, 43, 77–84.

[9] Anastasiadis, A.D., Albuquerque, M.P., Albuquerque, M.P., Mussi, D.B. (2009). Tsallis q-exponential describes the distribution of scientific citationsa new characterization of the impact, *Scientometrics*, 83, 205-218.

[10] Malacarne, L.C., Mende, R.S., Lenzi, E.K. (2001). q-exponential distribution in urban agglomeration *Physical Review E*, 65, 017106.

[11] Shimada,T., Yukawa, S. Ito, N. (2003). Life-span of families in fossil data forms q-exponential distribution, *International Journal of Modern Physics C*, 14, 1267–1271.

[12] Burlaga, L.F., Viñas, A.F. (2005). Triangle for the entropic index q of non-extensive statistical mechanics observed by Voyager 1 in the distant heliosphere, *Physica A*, 356, 375–384.

[13] Burlaga, L.F., Ness, N.F. (2009). Compressible turbulence observed in the heliosheath by Voyager 2, *The Astrophysical Journal*, 703, 311324.

[14] Carvalho, J.C., Silva, R., Nascimento, J.D. , Medeiros, J.R. (2008). Power law statistics and stellar rotational velocities in the Pleiades, *Europhysics Letters*, 84, 59001.

[15] Daniels, K.E., Beck, C., Bodenschatz, E. (2004). Defect turbulence and generalized statistical mechanics, *Physica D*, 193, 208–217.

[16] Ferri, G.L., Savio, M.F.R., Plastino, A. (2010). Tsallis' q-triplet and the ozone layer *Physica A*, 389, 1829–1833.

[17] Liu, B. Goree, J. (2008). Superdiffusion and non-Gaussian statistics in a driven-dissipative 2D dusty plasma, *Physical Review Letters*, 100, 055003.

[18] Pickup, R.M., Cywinski, R., Pappas, C., Farago, B., Fouquet, P. (2009). Generalized spin-glass relaxation, *Physical Review Letters*, 102, 097202.

[19] Upadhyaya, A., Rieu, J.P., Glazier, J.A., Sawada, Y. (2001). Anomalous diffusion and non-Gaussian velocity distribution of hydra cells in cellular aggregates, *Physica A*, 293, 549–558.

[20] CMS Collaboration (2010). Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at ?s = 0.9 and 2.36TeV., *Journal of High Energy Physics*, 02, 041

[21] Robertson, CA, Fryer, JG (1970). The bias and accuracy of moment estimators, *Biometrika*, **57**, 57–65.

[22] Dominitz, J, Sherman, RP (2005). Some Convergence Theory for Iterative Estimation Procedures with an Application to Semiparametric Estimation , *Econometric Theory*, **21**, 838–863.

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

[24] Lehmann, E.L., Casella, G. (1998). *Theory of Point Estimation*, 2nd ed., Springer-Verlag, New York.

[25] Drton, M. (2009). Likelihood ration tests and singularities, *The Annals of Statistics*, **25** 573–578.

[26] R Development Core Team (2008). R: A Language and Environment for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. `http://www.R-project.org.`

[27] Nelder, J., Wedderburn, R. (1972). Generalized Linear Models, *Journal of the Royal Statistical Society, Series A*, **135**: 370-384.

[28] McCullagh, P., Nelder, J. (1989). *Generalized Linear Models*, Second Edition, Chapman and Hall/CRC.Theoretical Statistics.

[29] Lange, K.L., Little, R.J.A., Taylor, J.M.G. (1989). Robust statistical modeling using the t distribution, *Journal of the Americam Statistical Association*, **84**:881–896.

[30] Welsh, A.H., Richardson, A.M. (1997). Approaches to the robust estimation of mixed models, In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, vol. 15. Elsevier Science, Amsterdam, pp. 343–384.

[31] Kowalski, J., Mendoza–Blanco, J.R., Tu, X.M., Gleser, L.J. (1999). On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regressions, *Communication in Statistics, Theory and Methods*, **28**:2119–2140.

[32] Fernández, C., Steel, M.F.J. (1999). Multivariate Student-t regression models: pitfalls and inference, *Biometrika*, **86**:153–167.

[33] Galea, M., Riquelme, M., Paula, G.A. (2000). Diagnostic methods in elliptical linear regression models, *Brazilian Journal of Probability and Statistics*, **14**:167–184.

[34] Liu, S. (2002). Local influence in multivariate elliptical linear regression models, *Linear Algebra Application*, **354**:159–174.

[35] Díaz–García, J.A., Galea, M., Leiva, V. (2003). Influence diagnostics for elliptical multivariate linear regression models, *Communication in Statistics, Theory and Methods*, **32**:625–641.

[36] Díaz–García, J.A., Gutiérrez–Jáimez, R. (2007). The distribution of residuals from a general elliptical distribution, *Journal of Statistics planning and inference*, **137**:2347–2354.

[37] Osorio, F., Paula, G.A., Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure, *Computational Statistics and Data Analysis*, **51**:4354–4368.

[38] Russo, C.M., Paula, G.A., Aoki, R. (2009). Influence diagnostics in nonlinear mixed-effects elliptical models, *Computational Statistics and Data Analysis*, **53**, 4143–4156.

[39] Patriota, A.G. (2011). A note on influence diagnostics in nonlinear mixed-effects elliptical models, *Computational Statistics and Data Analysis*, **55**, 218–225.

Acknowledgment

This work was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) – Brazil (10/50014-0). The author also acknowledge the helpful comments and suggestions from an anonymous referee.

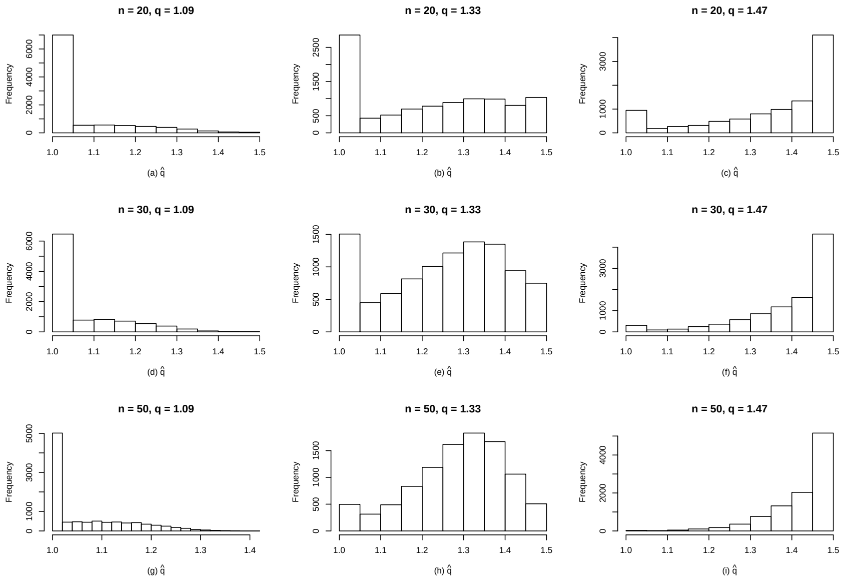

FIGURE 1. Histogram of the 10 000 ML estimates $\widehat{q}$. (a) $n = 20$ and $q = 1.09$, (b) $n = 20$ and $q = 1.33$, (c) $n = 20$ and $q = 1.47$, (d) $n = 30$ and $q = 1.09$, (e) $n = 30$ and $q = 1.33$, (f) $n = 30$ and $q = 1.47$, (g) $n = 50$ and $q = 1.09$, (h) $n = 50$ and $q = 1.33$, (i) $n = 50$ and $q = 1.47$.

Alexandre G. Patriota
Rua do Matão, 1010
Instituto de Matemática e Estatística,
Universidade de São Paulo,
Brazil

e-mail: `patriota@ime.usp.br`

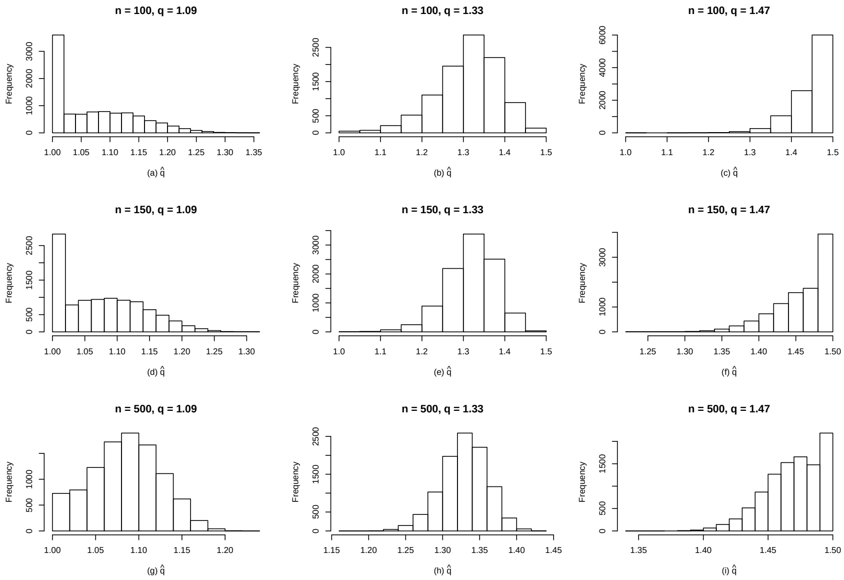

FIGURE 2. Histogram of the 10 000 ML estimates $\widehat{q}$. (a) $n = 100$ and $q = 1.09$, (b) $n = 100$ and $q = 1.33$, (c) $n = 100$ and $q = 1.47$, (d) $n = 150$ and $q = 1.09$, (e) $n = 150$ and $q = 1.33$, (f) $n = 150$ and $q = 1.47$, (g) $n = 500$ and $q = 1.09$, (h) $n = 500$ and $q = 1.33$, (i) $n = 500$ and $q = 1.47$.

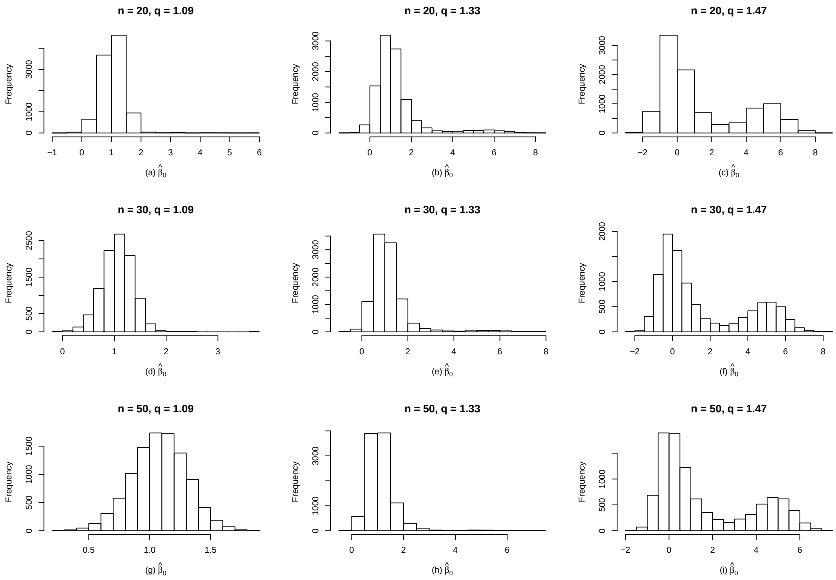

FIGURE 3. Histogram of the 10 000 ML estimates $\widehat{\beta}_0$. (a) $n = 20$ and $q = 1.09$, (b) $n = 20$ and $q = 1.33$, (c) $n = 20$ and $q = 1.47$, (d) $n = 30$ and $q = 1.09$, (e) $n = 30$ and $q = 1.33$, (f) $n = 30$ and $q = 1.47$, (g) $n = 50$ and $q = 1.09$, (h) $n = 50$ and $q = 1.33$, (i) $n = 50$ and $q = 1.47$.

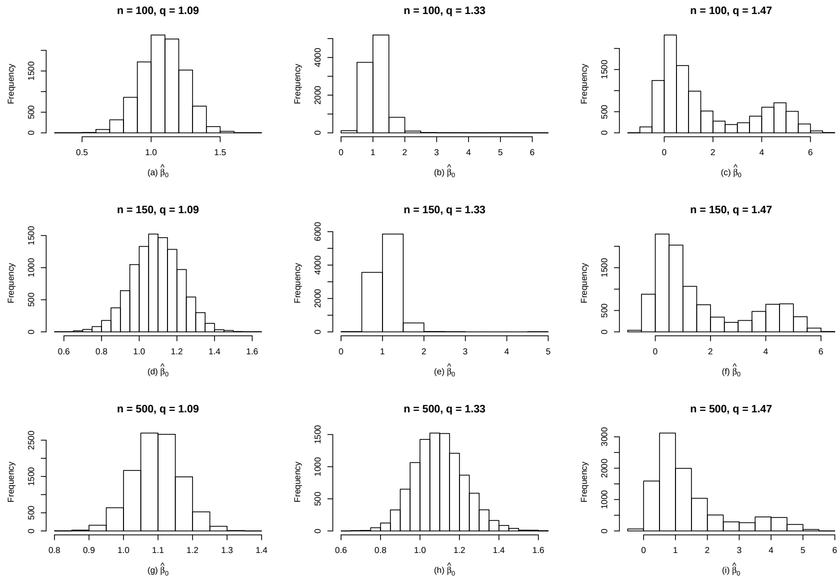

FIGURE 4. Histogram of the 10 000 ML estimates $\widehat{\beta}_0$. (a) $n = 100$ and $q = 1.09$, (b) $n = 100$ and $q = 1.33$, (c) $n = 100$ and $q = 1.47$, (d) $n = 150$ and $q = 1.09$, (e) $n = 150$ and $q = 1.33$, (f) $n = 150$ and $q = 1.47$, (g) $n = 500$ and $q = 1.09$, (h) $n = 500$ and $q = 1.33$, (i) $n = 500$ and $q = 1.47$.

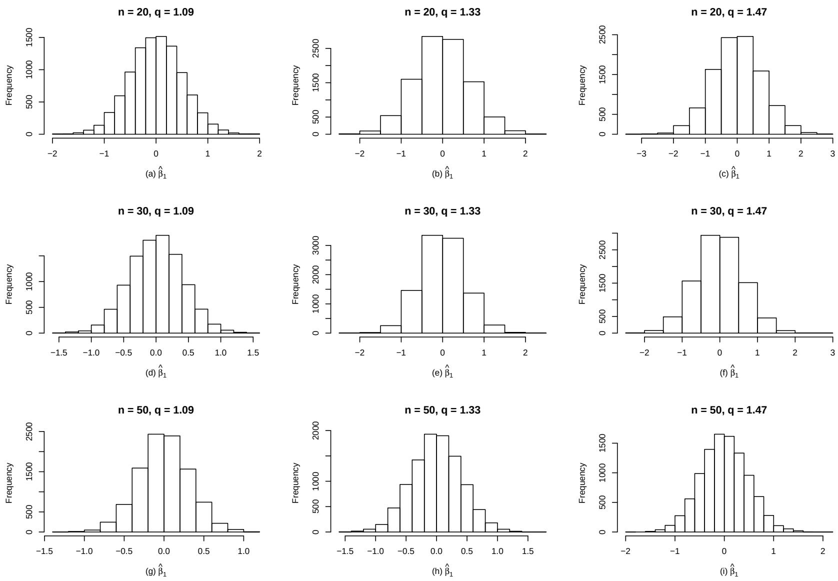

FIGURE 5. Histogram of the 10 000 ML estimates $\widehat{\beta}_1$. (a) $n = 20$ and $q = 1.09$, (b) $n = 20$ and $q = 1.33$, (c) $n = 20$ and $q = 1.47$, (d) $n = 30$ and $q = 1.09$, (e) $n = 30$ and $q = 1.33$, (f) $n = 30$ and $q = 1.47$, (g) $n = 50$ and $q = 1.09$, (h) $n = 50$ and $q = 1.33$, (i) $n = 50$ and $q = 1.47$.

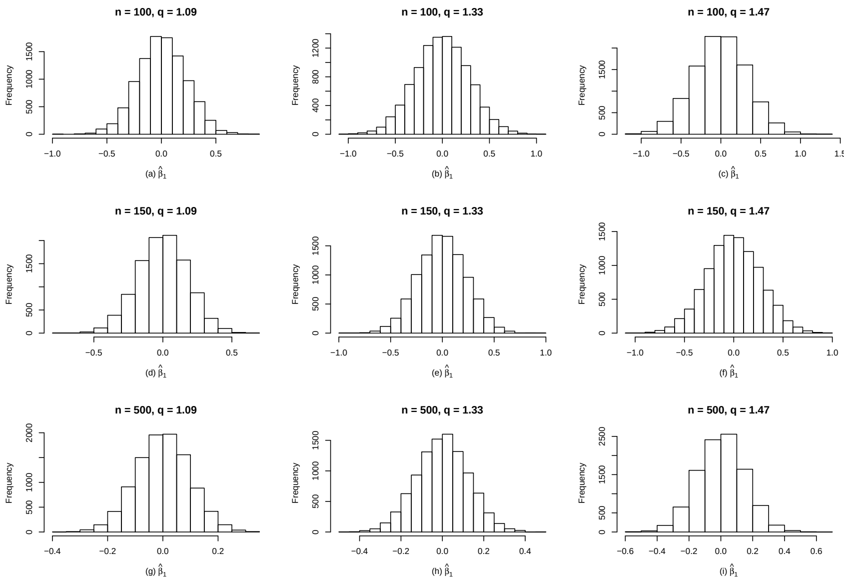

FIGURE 6. Histogram of the 10 000 ML estimates $\widehat{\beta}_1$. (a) $n = 100$ and $q = 1.09$, (b) $n = 100$ and $q = 1.33$, (c) $n = 100$ and $q = 1.47$, (d) $n = 150$ and $q = 1.09$, (e) $n = 150$ and $q = 1.33$, (f) $n = 150$ and $q = 1.47$, (g) $n = 500$ and $q = 1.09$, (h) $n = 500$ and $q = 1.33$, (i) $n = 500$ and $q = 1.47$.