Statistical Modelling xxxx; **xx(x)**: 1–27

Power logit regression for modeling bounded data

Francisco F. Queiroz[1] **and Silvia L. P. Ferrari**[1]

[1]Department of Statistics, University of São Paulo, Butantã, São Paulo, Brazil

Abstract: The main purpose of this article is to introduce a new class of regression models for bounded continuous data, commonly encountered in applied research. The models, named the power logit regression models, assume that the response variable follows a distribution in a wide, flexible class of distributions with three parameters, namely, the median, a dispersion parameter and a skewness parameter. The article offers a comprehensive set of tools for likelihood inference and diagnostic analysis, and introduces the new R package PLreg. Applications with real and simulated data show the merits of the proposed models, the statistical tools, and the computational package.

Key words: continuous proportions, fractional data, power logit distributions

Received January 2022; **revised** May 2022; **accepted** August 2022

1 Introduction

Bounded continuous data, particularly on the unit interval, appear in various research areas, including medicine, biology, sociology, psychology, economics, among many others. Some examples are the proportion of family income spent on health plans, mortality rate, percentage of body fat, fraction of the territory covered by treetops, loss given default, and efficiency scores calculated from data envelopment analysis (DEA). It is usually of interest to predict or explain the behavior of a continuous proportion from a set of other variables. Linear regression models may not be appropriate, as they may yield fitted values for the response variable that exceed its lower and upper bounds, below 0 and above 1. A natural strategy is to define a regression model where the dependent variable (Y) has a probability distribution function on the unit interval.

Different regression models for modeling data that are restricted to the interval $(0, 1)$ have been proposed. Beta regression models (Ferrari and Cribari–Neto, 2004; Smithson and Verkuilen, 2006), are widely used. Beta regression allows direct parameter interpretation, asymmetry and heteroscedasticity, while being reasonably flexible and having software available (Cribari–Neto and Zeileis, 2010). Several research papers have appeared in recent years regarding beta regression models for specific situations, such as errors-in-variables (Carrasco et al., 2014), longitudinal data (Wang et al., 1998; Di Brisco and Migliorati, 2020), high-dimensional data (Schmid et al., 2013; Weinhold et al., 2020), and time series (Rocha and Cribari-Neto, 2009; Pumi et al., 2021). Other distributions

Address for correspondence: Silvia L. P. Ferrari, Department of Statistics, University of São Paulo, Butantã, São Paulo 05508-090, Brazil.
E-mail: silviaferrari@usp.br
This article is based on part of the first author's PhD thesis which was supervised by the second author at the University of São Paulo, Brazil.

 10.1177/1471082X221140157

have been proposed to model variables with support on the unit interval, such as the rectangular beta distribution (Bayes et al., 2012), the log-Lindley distribution (Gómez-Déniz et al., 2014), the GJS class of distributions (Lemonte and Bazán, 2016), and the CDF-quantile distributions (Smithson and Shou, 2017). Regression models that accommodate observations at the extremes are found in Ospina and Ferrari (2012) and Queiroz and Lemonte (2021), among others.

Inference in beta regression models is usually based on maximum likelihood or Bayesian methods, for which the information from the data comes from the likelihood function. In either case, the inference can be highly influenced by atypical observations. To deal with this drawback, the inference procedure may be replaced by a robust method (Ribeiro and Ferrari, 2022). Alternatively, one may employ models based on distributions that are more flexible than the beta distribution. In this direction, Lemonte and Bazán (2016) propose the class of the GJS regression models, which is a generalization of the Johnson SB models (Johnson, 1949). The GJS distribution is defined from the transformation $X = [t(Y) - t(\mu)]/\sigma$, where $t(y) = \log[y/(1-y)]$, $y \in (0,1)$, $\mu \in (0,1)$, $\sigma > 0$, assuming that X has a standard symmetric distribution. Thus, the class of GJS distributions is constructed from symmetric distributions assigned to the logit of the variable that has support on $(0,1)$. However, it is known that the logit transformation is not able to bring common distributions of continuous fractions to symmetry; we illustrate this in Section 2.

In this work, we develop a very flexible class of regression models for continuous data on the unit interval, named the power logit models. These models employ a new class of distributions with three parameters that represent median, dispersion and skewness. This class of models has the GJS regression models as a particular case, with the advantage that the skewness parameter provides extra flexibility and accommodates highly skewed data. The regression parameters are interpretable in terms of the median and dispersion of the response variable. Since the interest lies in skewed data, the median may be a more appropriate measure of central tendency than the mean. We also define diagnostic and influence measures and present applications that show the usefulness of the proposed regression models in practice. Additionally, we develop a new `R` package, named `PLreg`, for fitting both power logit and GJS models.

The outline for the remainder of this article is as follows. Section 2 introduces the power logit class of distributions and some of its properties. Section 3 defines the class of power logit regression models. Section 4 presents some diagnostic tools and influence measures. Section 5 gives a brief presentation of the `PLreg` package. Section 6 presents and discusses real data applications. The article closes with some discussions and directions for future extensions.

2 The power logit distributions

We say that a random variable W has a distribution in the symmetric class of distributions if its probability density function is

$$g(w) = g(w; \mu, \sigma) = \frac{1}{\sigma} r\left(\frac{(w-\mu)^2}{\sigma^2}\right), \quad w \in \mathbb{R},$$

in which μ is the location parameter, $\sigma > 0$ is the scale parameter, and $r(z) > 0$, for $z \geq 0$, with $\int_0^\infty z^{-1/2} r(z) dz = 1$, is the density generator function (Fang and Anderson, 1990). We write $W \sim$

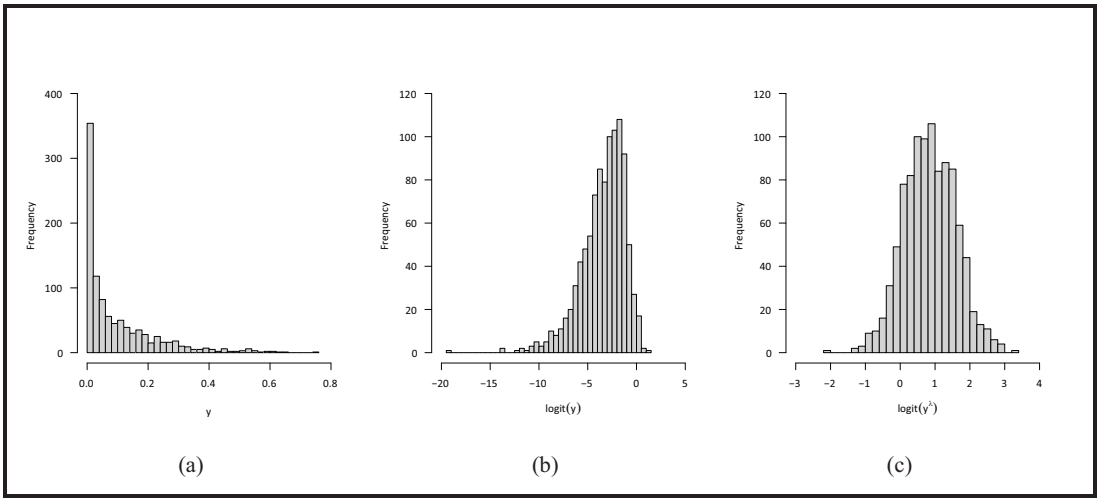

Figure 1 Histograms of y (original data), logit(y) and logit(y^{λ}); $\lambda = 0.11$

S($\mu, \sigma; r$). The class of the symmetric distributions has a number of well-known distributions as special cases depending on the choice of r. It includes the normal distribution as well as the Student-t, power exponential, type I and type II logistic, symmetric hyperbolic, sinh-normal and slash distributions, among others. Densities in this family have quite different tail behaviors, and some of them may have heavier or lighter tails than the normal distribution.

We now define a new class of distributions with support on the unit interval.

Definition 1 *Power logit distributions.* Let Y be a continuous random variable with support (0, 1) and let

$$Z = h(Y; \mu, \sigma, \lambda) = \frac{1}{\sigma}\left[\log\left(\frac{Y^{\lambda}}{1 - Y^{\lambda}}\right) - \log\left(\frac{\mu^{\lambda}}{1 - \mu^{\lambda}}\right)\right], \tag{2.1}$$

where $0 < \mu < 1$, $\sigma > 0$, and $\lambda > 0$. If Z has a standard symmetric distribution, that is, $Z \sim$ S(0, 1; r), we say that Y has a power logit distribution (PL) with parameters μ, σ and λ, and density generator function r. We write $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$.

The motivation for the power logit distributions is as follows. The logit transformation maps the interval (0, 1) to $(-\infty, +\infty)$ and, as such, is a candidate to define distributions with support in the unit interval from any distribution supported on the real line. In place of a particular distribution in $(-\infty, +\infty)$, we use the class of symmetric distributions. In addition, we add a power parameter in the logit transformation making it much more flexible. In fact, the power logit function, $\log[y^{\lambda}/(1 - y^{\lambda})]$, for $\lambda > 0$, is successful in achieving symmetry in situations where the logit transformation ($\lambda = 1$) fails. As an illustration, see Figure 1, which shows the histogram of a sample of 1 000 observations with a skewed distribution on (0, 1), along with the histograms of the observations transformed by the logit function and the power logit function with $\lambda = 0.11$. It is clear that the histogram of the logit transformed data is far from symmetry unlike the histogram of the power logit transformed

observations, which is nearly symmetric. The sample skewness coefficients for the original data and the transformed data are approximately 1.8 (original), −1.3 (logit), and zero (power logit). The capacity of the power logit function to transform for symmetry is further illustrated in Section 1 of the Supplementary Material. Finally, Definition 1 implies that $\log[Y^\lambda/(1-Y^\lambda)]$ has a symmetric distribution with location parameter $\log[\mu^\lambda/(1-\mu^\lambda)]$ and scale parameter σ. This parametrization is convenient because μ represents the median of Y, while σ and λ are dispersion and skewness parameters, respectively, as we will show later.

The power logit distribution may seem closely connected with the range-power transformation (Scrucca, 2019, eq. (6) with $l = 0$ and $u = 1$). However, Scrucca's approach is based on the Box-Cox transformation $\{[y/(1-y)]^\lambda - 1\}/\lambda$ for $\lambda \neq 0$ and $\log[y/(1-y)]$ for $\lambda = 0$, not a power logit transformation. Also, it postulates a distribution with support in the real line, namely, a normal finite mixture distribution, to the transformed data. It should be noted that the support of the Box-Cox transformed variable is not the entire real line unless $\lambda = 0$. Hence, Scrucca's proposed transformation is not the power logit transformation, and it does not define a proper distribution for continuous variables supported in the unit interval.

The probability density function (pdf) of $Y \sim \text{PL}(\mu, \sigma, \lambda; r)$ is

$$f_Y(y; \mu, \sigma, \lambda) = \frac{\lambda}{\sigma y(1-y^\lambda)} r(z^2), \quad y \in (0,1), \tag{2.2}$$

where

$$z = h(y; \mu, \sigma, \lambda) = \frac{1}{\sigma}\left[\log\left(\frac{y^\lambda}{1-y^\lambda}\right) - \log\left(\frac{\mu^\lambda}{1-\mu^\lambda}\right)\right]. \tag{2.3}$$

The density generator $r(\cdot)$ may involve an extra parameter (or an extra parameter vector), which is denoted by ζ. In addition, using the linear transformation $X = (b-a)Y + a$, we obtain the power logit distribution with support (a, b), $a < b$.

The power logit class of distributions reduces to the GJS class of distributions (Lemonte and Bazán, 2016) when $\lambda = 1$ and we write $Y \sim \text{GJS}(\mu, \sigma; r)$. Additionally, it leads to the logit normal distribution (Johnson, 1949), the L-Logistic distribution (da Paz et al., 2019), and the logit slash distribution (Korkmaz, 2020) by taking $\lambda = 1$ and Z as a standard normal, type II logistic, and slash random variable, respectively. The density generator function, $r(z)$, for $z \geq 0$, for the power logit normal (PL-N), power logit Student-t (PL-$\text{t}_{(\zeta)}$), power logit type I logistic (PL-LOI), power logit type II logistic (PL-LOII), power logit power exponential (PL-$\text{PE}_{(\zeta)}$), power logit slash (PL-$\text{slash}_{(\zeta)}$), power logit hyperbolic (PL-$\text{Hyp}_{(\zeta)}$), and power logit sinh-normal (PL-$\text{SN}_{(\zeta)}$) follow.

- normal: $r(z) = (2\pi)^{-1/2}\exp(-z/2)$;
- Student-t: $r(z) = \zeta^{\zeta/2}B(1/2, \zeta/2)^{-1}(\zeta + z)^{-(\zeta+1)/2}$, $\zeta > 0$ and $B(\cdot,\cdot)$ is the beta function;
- type I logistic: $r(z) = c\exp\{-z\}(1+\exp\{-z\})^2$, where $c \approx 1.484300029$ is the normalizing constant;
- type II logistic: $r(z) = \exp\{-z^{1/2}\}(1+\exp\{-z^{1/2}\})^{-2}$;
- power exponential: $r(z) = \zeta/[p(\zeta)2^{1+1/\zeta}\Gamma(1/\zeta)]\exp\left\{-z^{\zeta/2}/(2p(\zeta)^\zeta)\right\}$, $\zeta > 0$ and $p(\zeta)^2 = 2^{-2/\zeta}\Gamma(1/\zeta)/\Gamma(3/\zeta)$;

- slash: $r(z) = (\zeta/\sqrt{2\pi})\,(z/2)^{-(\zeta+1/2)}\,G\,(\zeta + 1/2, z/2)$, for $z > 0$, and $r(z) = 2\zeta/[(2\zeta + 1)\sqrt{2\pi}]$, for $z = 0$, where $\zeta > 0$ and $G(a, x) = \int_0^x t^{a-1}e^{-t}dt$ is the lower incomplete gamma function. When $\zeta = 1$ the slash distribution coincides with the canonical slash distribution;
- hyperbolic: $r(z) = \exp\left\{-\zeta\sqrt{1+z}\right\}/(2\zeta K_1(\zeta))$, with $K_s(\zeta) = \int_0^\infty \frac{x^{s-1}}{2}\exp\{-\frac{\zeta}{2}\left(x + \frac{1}{x}\right)\}\mathrm{d}x$, is the modified Bessel function of third-order and index s.
- sinh-normal: $r(z) = 1/(\zeta\sqrt{2\pi})\cosh(z^{1/2})\exp\left[-2/\zeta^2\sinh^2(z^{1/2})\right]$, where $\zeta > 0$ and $\sinh(\cdot)$ and $\cosh(\cdot)$ represent the hyperbolic sine and cosine functions, respectively.

The power logit density can assume different shapes depending on the combination of parameter values and density generator functions; see Figure 2 and Figure 2 of the Supplementary Material. Figures 2(a)–2(c) present the pdf of the PL-Hyp$_{(1.2)}$ distribution. Figure 2(a) suggests that μ is a parameter of central tendency. For fixed values of μ, λ, and ζ, the dispersion of the distribution increases as σ increases, suggesting that σ is a parameter that governs the dispersion of the distribution; see Figure 2(b). For fixed μ, σ, and ζ, the pdf becomes closer to symmetry around μ as λ increases, indicating that λ acts as a skewness parameter; see Figure 2(c). In Sections 3 and 4 of the Supplementary Material, we show that the parameter σ is in fact a dispersion parameter and that λ can be regarded as a skewness parameter. Finally, Figure 2(d) shows the pdf of the PL-N, PL-t$_{(4)}$, PL-PE$_{(1.5)}$, PL-slash$_{(1.5)}$ and PL-Hyp$_{(3)}$ distributions for a particular choice of the parameters. Note that the PL-t$_{(4)}$ and PL-slash$_{(1.5)}$ distributions display heavier tails than the other distributions.

We now give some properties of the power logit distributions. If $Y \sim \mathrm{PL}(\mu, \sigma, \lambda; r)$ then, the following properties hold, most of which are immediate consequences of Definition 1.

(P1) The cumulative distribution function (cdf) of Y is $F_Y(y; \mu, \sigma, \lambda) = R(z)$, where $R(\cdot)$ is the cdf of $Z \sim \mathrm{S}(0, 1; r)$ and z is given in (2.3).

(P2) The $100u\%$ quantile of Y is $y_u = \mu\left[e^{\sigma z_u}/(1 - \mu^\lambda(1 - e^{\sigma z_u}))\right]^{1/\lambda}$, for $u \in (0, 1)$, where z_u is the $100u\%$ quantile of $Z \sim \mathrm{S}(0, 1; r)$.

(P3) μ is the median of Y.

(P4) σ is a dispersion parameter in the sense of the quantile spread order (Townsend and Colonius, 2005).

(P5) For $\lambda = 1$,
1. $1 - Y \sim \mathrm{PL}(1 - \mu,\ \sigma,\ \lambda = 1;\ r)$;
2. $Y \sim \mathrm{GJS}(\mu, \sigma; r)$;
3. if $\mu = 0.5$, the power logit density function is symmetric around $y = 0.5$.

(P6) $Y^\lambda \sim \mathrm{GJS}(\mu^\lambda, \sigma; r)$.

(P7) $Y^c \sim \mathrm{PL}(\mu^c,\ \sigma,\ \lambda/c;\ r)$, for all $c > 0$.

(P8) Let $W \sim \mathrm{S}(-\log(-\log\mu), \sigma; r)$, then

$$-\log(-\log Y) \xrightarrow{\mathcal{D}} W, \quad \text{when} \quad \lambda \to 0^+,$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.

Property P5(3) states that, for $\lambda = 1$ and $\mu = 0.5$, the power logit distributions are (perfectly) symmetric. Nearly symmetric distributions arise when λ is large, as evidenced in Figure 2(c) and

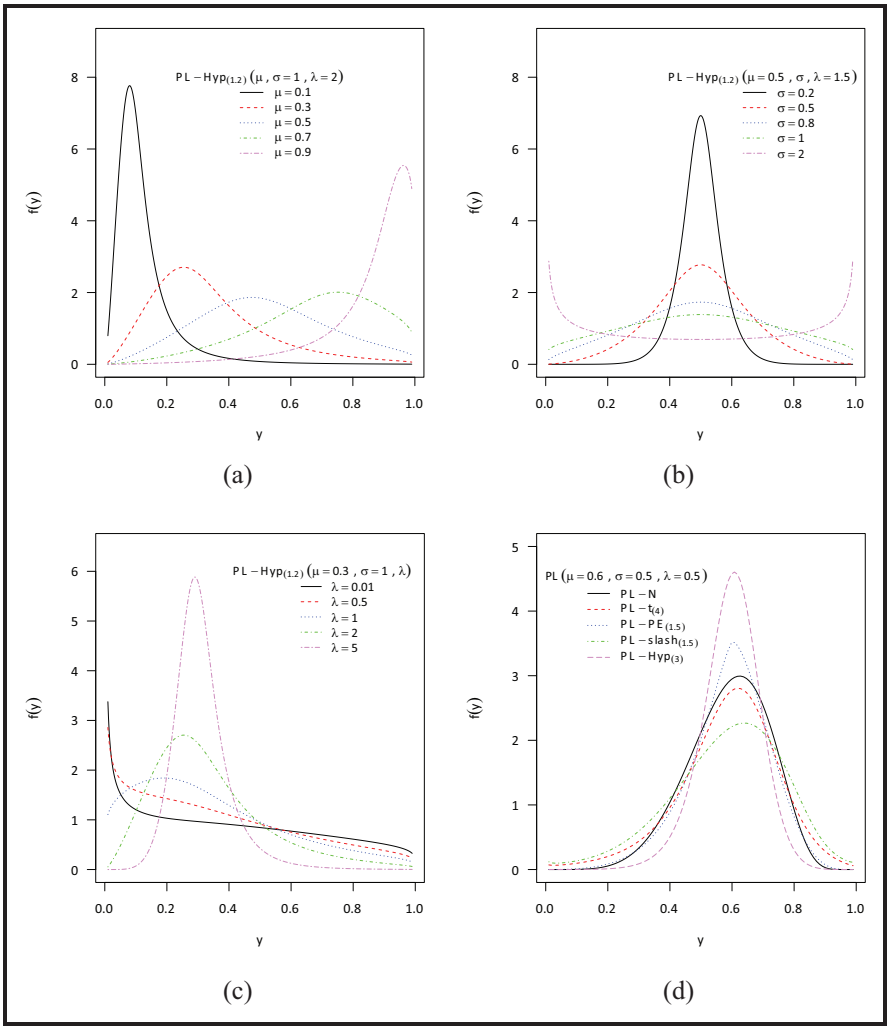

Figure 2 Plots of the pdf of some power logit distributions

Figures 2 and 3 of the Supplementary Material. Property P8 states that the power logit distributions have as a limiting case, when $\lambda \to 0^+$, the family of log-log distributions, as defined below.

Definition 2 *log-log distributions*. We say that a continuous random variable Y with support $(0, 1)$ has a log-log distribution with parameters $\mu \in (0, 1)$ and $\sigma > 0$ and density generator function r if

$$\frac{1}{\sigma}\{-\log(-\log Y) - [-\log(-\log \mu)]\} \sim \mathrm{S}(0, 1; r).$$

We write $Y \sim \text{log-log}(\mu, \sigma; r)$. The pdf of Y is

$$f(y; \mu, \sigma) = \frac{1}{\sigma y(-\log y)} r(z^2), \quad y \in (0, 1),$$

with $z = \sigma^{-1}\{-\log(-\log y) - [-\log(-\log \mu)]\}$.

This class of distributions has some interesting properties, for instance: the $100u\%$ quantile of Y is $y_u = \mu^{\exp\{-\sigma z_u\}}$, for $u \in (0, 1)$, where z_u is the $100u\%$ quantile of $Z \sim \text{S}(0, 1; r)$; μ and σ are the median and the dispersion of Y, respectively; if $Y \sim \text{log-log}(\mu, \sigma; r)$, then $Y^c \sim \text{log-log}(\mu^c, \sigma; r)$, for $c > 0$.

3 Power logit regression models

3.1 Definition

We define a class of power logit regression models as follows. Let $Y_1, \ldots, Y_n$ be n independent random variables, where $Y_i \sim \text{PL}(\mu_i, \sigma_i, \lambda; r)$, for $i = 1, \ldots, n$, and

$$\begin{aligned} d_1(\mu_i) &= \boldsymbol{x}_i^\top \boldsymbol{\beta} = \eta_{1i}, \\ d_2(\sigma_i) &= \boldsymbol{s}_i^\top \boldsymbol{\tau} = \eta_{2i}, \end{aligned} \tag{3.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_q)^\top \in \mathbb{R}^q$ and $\lambda > 0$ are the unknown parameters, which are assumed to be functionally independent and $p + q + 1 < n$; η_{1i} and η_{2i} are the linear predictors; $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ and $\boldsymbol{s}_i = (s_{i1}, \ldots, s_{iq})^\top$ are observations on p and q known independent variables. Intercepts are allowed in the median and dispersion submodels by setting $x_{i1} = 1$ and $s_{i1} = 1$, for $i = 1, \ldots, n$. We assume that the model matrices $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]^\top$ and $\mathbf{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n]^\top$ have column rank p and q, respectively. In addition, we assume that the link functions $d_1\colon (0, 1) \to \mathbb{R}$ and $d_2\colon (0, \infty) \to \mathbb{R}$ are strictly monotonic and twice differentiable. Some examples of link functions for the median submodel are: $d_1(\mu) = \log\{\mu/(1 - \mu)\}$ (logit); $d_1(\mu) = \Phi^{-1}(\mu)$ (probit), where $\Phi^{-1}(\cdot)$ is the cdf of a standard normal random variable; $d_1(\mu) = -\log\{-\log \mu\}$ (log-log); and $d_1(\mu) = \log\{-\log(1 - \mu)\}$ (complementary log-log). For the dispersion submodel, the log link, $d_2(\sigma) = \log \sigma$, is the natural choice.

The power logit regression models generalize some models: the GJS regression model (Lemonte and Bazán, 2016) is obtained by taking $\lambda = 1$; if $Y_i \sim \text{PL-LOII}(\mu_i, \sigma_i, 1)$ we have the L-logistic regression model (da Paz et al., 2019). Additionally, the model parameters are interpreted in terms of the median, dispersion and skewness of the response variable. Also, the introduction of the skewness parameter λ allows better fits for highly skewed data.

Finally, the power logit regression models have the log-log regression models as a limiting case, when $\lambda \to 0^+$, which are defined by setting $Y_i \sim \text{log-log}(\mu_i, \sigma_i; r)$, for $i = 1, \ldots, n$, with μ_i and σ_i

given in (3.1). In practice, the log-log regression models may be regarded as a parsimonious alternative to the power logit regression models when the estimated λ is close to zero.[1]

3.2 Parameter estimation

Let $y_1, \ldots, y_n$ be n observed responses from a power logit regression model. The log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top, \lambda)^\top$ is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\mu_i, \sigma_i, \lambda), \tag{3.2}$$

where $\ell_i = \ell_i(\mu_i, \sigma_i, \lambda) = \log\lambda - \log\sigma_i - \log\{1 - y_i^\lambda\} + \log\{r(z_i^2)\} + c$, $z_i = h(y_i; \mu_i, \sigma_i, \lambda)$, and c is constant with respect to $\boldsymbol{\theta}$. Note that, from (3.1), μ_i and σ_i are defined as functions of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, respectively; that is, $\mu_i = d_1^{-1}(\eta_{1i})$, $\sigma_i = d_2^{-1}(\eta_{2i})$, for $i = 1, \ldots, n$. The score function (see Section 5 of the Supplementary Material) is given by the $(p + q + 1)$-vector $\mathbf{U}_n(\boldsymbol{\theta}) = (\mathbf{U}_{\boldsymbol{\beta}}^\top, \mathbf{U}_{\boldsymbol{\tau}}^\top, \mathrm{U}_\lambda)^\top$, with

$$\mathbf{U}_\beta = \mathbf{X}^\top\mathbf{W}\mathbf{T}_1\boldsymbol{\mu}^*, \quad \mathbf{U}_\tau = \mathbf{S}^\top\mathbf{T}_2\boldsymbol{\sigma}^*, \quad \mathrm{U}_\lambda = \mathbf{1}_n^\top\boldsymbol{\lambda}^*,$$

where $\mathbf{T}_1 = \mathrm{diag}\{1/\dot{d}_1(\mu_1), \ldots, 1/\dot{d}_1(\mu_n)\}$, $\mathbf{T}_2 = \mathrm{diag}\{1/\dot{d}_2(\sigma_1), \ldots, 1/\dot{d}_2(\sigma_n)\}$, $\mathbf{W} = \mathrm{diag}\{z_1 v(z_1), \ldots, z_n v(z_n)\}$, $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_n^*)^\top$, $\boldsymbol{\sigma}^* = (\sigma_1^*, \ldots, \sigma_n^*)^\top$, $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_n^*)^\top$, $\mathbf{1}_n$ denotes a n-dimensional vector of ones, $\dot{d}_j(t) = \mathrm{d}d_j(t)/\mathrm{d}t$, for $j = 1, 2$, $v(t) = -2r'(t^2)/r(t^2)$, $r'(u) = \mathrm{d}r(u)/\mathrm{d}u$, $\mu_i^* = \lambda/[\sigma_i\mu_i(1 - \mu_i^\lambda)]$, $\sigma_i^* = \sigma^{-1}[z_i^2 v(z_i) - 1]$, and

$$\lambda_i^* = \frac{1}{\lambda} + \frac{y_i^\lambda \log y_i}{1 - y_i^\lambda} - \frac{1}{\sigma_i} z_i v(z_i)\left[\frac{\log y_i}{(1 - y_i^\lambda)} - \frac{\log\mu_i}{(1 - \mu_i^\lambda)}\right].$$

The negative of the Hessian matrix for $\boldsymbol{\theta}$ is given by

$$\mathbf{J}_n(\boldsymbol{\theta}) = -\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\tau}} & \mathbf{J}_{\boldsymbol{\beta}\lambda} \\ \mathbf{J}_{\boldsymbol{\beta}\boldsymbol{\tau}}^\top & \mathbf{J}_{\boldsymbol{\tau},\boldsymbol{\tau}} & \mathbf{J}_{\boldsymbol{\tau}\lambda} \\ \mathbf{J}_{\boldsymbol{\beta}\lambda}^\top & \mathbf{J}_{\boldsymbol{\tau}\lambda}^\top & \mathrm{J}_{\lambda\lambda} \end{bmatrix}, \tag{3.3}$$

[1] For brevity, the log-log regression models will not be studied in the following sections, but to obtain the associated quantities, take the limit when $\lambda \to 0^+$ of those defined for the power logit regression models.

where $\mathbf{J}_{\boldsymbol{\beta\beta}} = \mathbf{X}^\top \mathbf{W}_1 \mathbf{T}_1 \mathbf{X}$, $\mathbf{J}_{\boldsymbol{\tau\tau}} = \mathbf{S}^\top \mathbf{W}_2 \mathbf{T}_2 \mathbf{S}$, $\mathrm{J}_{\lambda\lambda} = \mathbf{1}_n^\top \mathbf{W}_3 \mathbf{1}_n$, $\mathbf{J}_{\boldsymbol{\beta\tau}} = \mathbf{X}^\top \mathbf{W}_4 \mathbf{T}_1 \mathbf{T}_2 \mathbf{S}$, $\mathbf{J}_{\boldsymbol{\beta}\lambda} = \mathbf{X}^\top \mathbf{W}_5 \mathbf{T}_1 \mathbf{1}_n$, $\mathbf{J}_{\boldsymbol{\tau}\lambda} = \mathbf{S}^\top \mathbf{W}_6 \mathbf{T}_1 \mathbf{1}_n$, with $\mathbf{W}_j = \mathrm{diag}\{w_1^{(j)}, \ldots, w_n^{(j)}\}$,

$$w_i^{(1)} = \frac{\lambda}{[\sigma_i \mu_i (1-\mu_i^\lambda)]^2} \left\{\sigma_i[1-\mu_i^\lambda(1+\lambda)] z_i v(z_i) + \lambda[v(z_i) + z_i v'(z_i)]\right\} \dot{d}_1(\xi_i)^{-1}$$

$$+ \frac{\lambda}{\sigma_i \mu_i (1-\mu_i^\lambda)} z_i v(z_i) \frac{\ddot{d}_1(\mu_i)}{\dot{d}_1(\mu_i)^2},$$

$$w_i^{(2)} = -\left\{\frac{1}{\sigma_i^2} - \frac{1}{\sigma_i^2} z_i^2 [3v(z_i) + z_i v'(z_i)]\right\} \dot{d}_2(\sigma_i)^{-1} + \frac{1}{\sigma_i}[z_i^2 v(z_i) - 1] \frac{\ddot{d}_2(\sigma_i)}{\dot{d}_2(\sigma_i)^2},$$

$$w_i^{(3)} = \frac{1}{\lambda^2} - \frac{y_i^\lambda \log^2 y_i}{(1-y_i^\lambda)^2} + \frac{1}{\sigma_i^2}[v(z_i) + z_i v'(z_i)] \left(\frac{\log y_i}{1-y_i^\lambda} - \frac{\log \mu_i}{1-\mu_i^\lambda}\right)^2$$

$$+ \frac{1}{\sigma_i} z_i v(z_i) \left[\frac{y_i^\lambda \log^2 y_i}{(1-y_i^\lambda)^2} - \frac{\mu_i^\lambda \log^2 \mu_i}{(1-\mu_i^\lambda)^2}\right],$$

$$w_i^{(4)} = \frac{\lambda}{\sigma_i^2 \mu_i (1-\mu_i^\lambda)} z_i [2v(z_i) + z_i v'(z_i)],$$

$$w_i^{(5)} = \frac{-\lambda}{\sigma_i \mu_i (1-\mu_i^\lambda)} \left\{\frac{1-\mu_i^\lambda(1-\lambda \log \mu_i)}{\lambda(1-\mu_i^\lambda)} z_i v(z_i) + \frac{1}{\sigma_i}\left(\frac{\log y_i}{1-y_i^\lambda} - \frac{\log \mu_i}{1-\mu_i^\lambda}\right)[v(z_i) + z_i v'(z_i)]\right\},$$

$$w_i^{(6)} = -\frac{1}{\sigma_i^2}\left(\frac{\log y_i}{1-y_i^\lambda} - \frac{\log \mu_i}{1-\mu_i^\lambda}\right) z_i [2v(z_i) + z_i v'(z_i)].$$

where $\ddot{d}_j(z) = \mathrm{d}\dot{d}_j(z)/dz$, for $j = 1, 2$.

The maximum likelihood estimate (mle) of $\boldsymbol{\theta}$ can be obtained by solving simultaneously the nonlinear system of equations $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}_{p+q+1}$, which does not have closed form, and $\mathbf{0}_{p+q+1}$ denotes a $(p+q+1)$-dimensional vector of zeros. As we can see from the score function $\mathbf{U}_\beta$, $v(\cdot)$ acts as a weighting function, that is, observations with small value for $v(\cdot)$ are downweighted for estimating $\boldsymbol{\beta}$.

The choice of $r(\cdot)$ may induce a function $v(z)$ that decreases as z departs from zero, and hence some power logit distributions produce robust estimation against outliers. For the PL-N, PL-$\mathrm{t}_{(\zeta)}$, PL-$\mathrm{PE}_{(\zeta)}$ and PL-$\mathrm{Hyp}_{(\zeta)}$ distributions one has, respectively, $v(z) = 1$, $v(z) = (\zeta+1)/(\zeta+z^2)$, $v(z) = \zeta(z^2)^{\zeta/2-1}/(2p(\zeta)^\zeta)$, and $v(z) = \zeta/\sqrt{1+z^2}$. For the PL-$\mathrm{t}_{(\zeta)}$, PL-$\mathrm{PE}_{(\zeta)}$ (with $0 < \zeta < 2$) and PL-$\mathrm{Hyp}_{(\zeta)}$ distributions, $v(z)$ is decreasing in z^2 and hence extreme observations tend to have small weights in the estimation process.

In applications with simulated data of a small sample size, we noted that the profile log-likelihood for λ may display weak concavity. We propose the use of a penalized log-likelihood function for λ as it will be described below in detail. We anticipate the expected effect of the penalization in Figure 3,

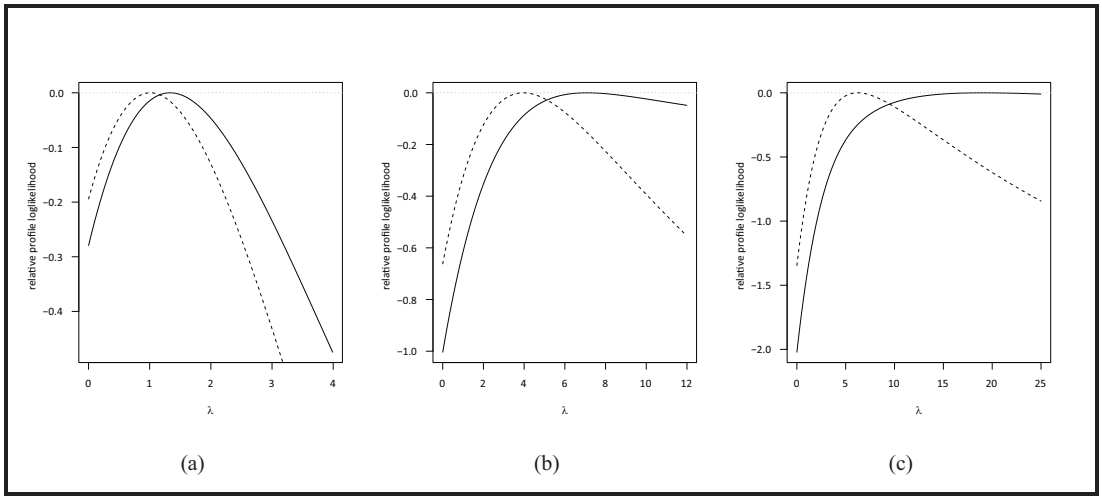

Figure 3 Relative versions of the usual (solid) and penalized (dashed) profile log-likelihood functions of λ for three samples

which presents the plots of the relative profile log-likelihood function of λ and the corresponding penalized version, obtained from three random samples of 25 observations taken from the PL-PE$_{(1.3)}$ distribution with $\mu = 0.7$, $\sigma = 0.5$, and $\lambda = 1$. The profile log-likelihood function for λ (solid line) for the first sample (Figure 3(a)) is well behaved, but the other two samples give rise to profile log-likelihoods with weak concavity regions and estimates for λ far from its true value, that is, $\lambda = 1$ (Figures 3(b)–3(c)). The corresponding relative profile penalized log-likelihood functions (dashed lines) do not display flat or nearly flat regions for any of the three samples. The estimates of λ become closer to the true value when the penalized log-likelihood function is employed.

In the statistical literature there are several reports of monotone likelihoods for different models: Cox regression model (Bryson and Johnson, 1981); logistic regression (Albert and Anderson, 1984); skew normal and skew t distributions (Sartori, 2006; Azzalini and Arellano-Valle, 2013); modified extended Weibull distribution (Lima and Cribari-Neto, 2019). To deal with monotone likelihoods, one may consider a modification in the log-likelihood function in order to obtain estimators with better properties.

As in Sartori (2006), we modify the profile log-likelihood of λ instead of the log-likelihood of $\boldsymbol{\theta}$, requiring less computational effort. We consider a penalization based on the Jeffreys' prior (Jeffreys, 1946), with the observed information matrix in place of Fisher's information matrix. This approach performed well in simulation experiments. The penalized profile log-likelihood for λ is

$$\ell_p^*(\lambda) = \ell^*(\lambda) + \frac{1}{2}\log\left(\frac{\mathrm{J}_{\lambda\lambda}^*}{n}\right), \tag{3.4}$$

in which $\ell^*(\lambda) = \ell(\widehat{\boldsymbol{\beta}}_\lambda, \widehat{\boldsymbol{\tau}}_\lambda, \lambda)$ is the profile log-likelihood for λ and $\mathrm{J}_{\lambda\lambda}^* = \mathrm{J}_{\lambda\lambda}(\widehat{\boldsymbol{\beta}}_\lambda, \widehat{\boldsymbol{\tau}}_\lambda, \lambda)$, where $\widehat{\boldsymbol{\beta}}_\lambda$ and $\widehat{\boldsymbol{\tau}}_\lambda$ are the mle for $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, respectively, with fixed λ. The penalized maximum likelihood estimate (pmle) of $\boldsymbol{\theta}$, that is, $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\tau}}, \widetilde{\lambda})^\top$, can be computed through numerical optimization following two steps:

i. Compute $\widetilde{\lambda}$ such that

$$\widetilde{\lambda} = \underset{\lambda>0}{\operatorname{argmax}}\, \ell_p^*(\lambda).$$

ii. Find $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\tau}}$ by maximizing $\ell(\boldsymbol{\beta}, \boldsymbol{\tau}, \widetilde{\lambda})$.

The optimization algorithms require the specification of initial values to be used in the iterative scheme. Our suggestion is to use as an initial point estimate for $\boldsymbol{\beta}$ the ordinary least squares estimate of this parameter vector obtained from a linear regression of the transformed responses $d_1(y_1), \ldots, d_1(y_n)$ on $\boldsymbol{X}$, that is, $\boldsymbol{\beta}^{(0)} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{v}$, where $\boldsymbol{v} = (d_1(y_1), \ldots, d_1(y_n))^\top$. For $\boldsymbol{\tau}$, we suggest $\boldsymbol{\tau}^{(0)} = (\tau_1^{(0)}, 0, \ldots, 0)^\top$, where $\tau_1^{(0)}$ is the sample standard deviation of $\log[y/(1-y)]$. Finally, an initial value for λ is $\lambda^{(0)} = 1$. In addition, quantities evaluated at $\widetilde{\boldsymbol{\theta}}$ ($\widehat{\boldsymbol{\theta}}$) will be written with a tilde (circumflex).

The penalization used in the profile log-likelihood for λ is $\mathcal{O}_p(1)$ as $n \to \infty$. Then, both $\ell^*(\lambda)$ and $\ell_p^*(\lambda)$ are $\mathcal{O}_p(n)$ and they differ only by the penalization term, which is $\mathcal{O}_p(1)$. Thus, the first-order asymptotic distribution of $\widetilde{\boldsymbol{\theta}}$ coincides that of $\widehat{\boldsymbol{\theta}}$; for more details, see Azzalini and Arellano-Valle (2013). Then, under suitable regularity conditions, $\widetilde{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$ and

$$\widetilde{\boldsymbol{\theta}} \overset{a}{\sim} \mathcal{N}_{p+q+1}\left(\boldsymbol{\theta},\ \mathbf{K}_n(\boldsymbol{\theta})^{-1}\right),$$

where $\overset{a}{\sim}$ denotes approximately distributed for large n, and $\mathbf{K}_n(\boldsymbol{\theta})$ is the total Fisher's information matrix. There is no closed-form expression for $\mathbf{K}_n(\boldsymbol{\theta})$, but the asymptotic behavior remains valid if $\mathbf{K}_n(\boldsymbol{\theta})^{-1}$ is approximated by $\mathbf{J}_n(\boldsymbol{\theta})^{-1}$ or by the sandwich covariance matrix $\mathbf{J}_n(\boldsymbol{\theta})^{-1}\mathbf{B}_n(\boldsymbol{\theta})\mathbf{J}_n(\boldsymbol{\theta})^{-1}$, where $\mathbf{B}_n(\boldsymbol{\theta}) = \sum_{i=1}^n (\partial \ell_i/\partial \boldsymbol{\theta})(\partial \ell_i/\partial \boldsymbol{\theta}^\top)$. We employ the sandwich matrix to obtain standard errors in the applications of Section 6. The asymptotic normal distribution can be used to construct approximate confidence intervals and confidence regions for the parameters. Also, the usual asymptotic properties of the large sample tests, such as the likelihood ratio and Wald tests, remain valid. Although confidence intervals for λ may be based on the asymptotic distribution of $\widetilde{\lambda}$, we recommend using the asymptotic χ_1^2 distribution of the profile penalized likelihood ratio statistic $W_p^*(\lambda) = 2\{\ell_p^*(\widetilde{\lambda}) - \ell_p^*(\lambda)\}$ (Sartori, 2006). Confidence intervals constructed through $W_p^*(\lambda)$ are not symmetric around $\widetilde{\lambda}$ and do not include values outside the parameter space.

In order to evaluate the performance of the penalized maximum likelihood estimator in the power logit models, we conducted a simulation study for power logit regression models with $\log(\mu_i/(1-\mu_i) = \beta_1 + \beta_2 x_i$ and $\log \sigma_i = \tau_1 + \tau_2 s_i$, where $\beta_1 = 0.5, \beta_2 = 1.5, \tau_1 = -1, \tau_2 = 0.5$, and $\lambda = 1$. The covariates x_i and s_i were generated as independent random draws from a uniform distribution on the unit interval and were kept fixed for all the replicates. We considered four different power logit distributions: PL-$t_{(5)}$, PL-$PE_{(1.5)}$, PL-$Hyp_{(1.2)}$, and PL-$slash_{(1.4)}$. We generated 3 000 Monte Carlo replicates with $n = 40$ and $n = 120$. All simulations were performed using the R software (R Core Team, 2021) with the BFGS algorithm (Press et al., 1992). The empirical bias and the root mean squared error (RMSE) of the maximum likelihood estimates with and without penalization are presented in Table 1.

Inspection of Table 1 shows that the maximum likelihood estimators for the β's, both usual and penalized, present bias close to zero and small mean squared error for all the investigated models;

Table 1 Bias and root mean squared error (RMSE) for the mle and pmle

		$n = 40$				$n = 120$			
		mle		pmle		mle		pmle	
		bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
PL-$t_{(5)}$	β_1	−0.00	0.14	−0.00	0.14	−0.00	0.09	−0.00	0.09
	β_2	−0.00	0.28	−0.00	0.28	−0.00	0.18	−0.00	0.18
	τ_1	0.21	0.65	0.05	0.48	0.08	0.34	0.04	0.31
	τ_2	−0.14	0.61	−0.04	0.55	−0.06	0.36	−0.03	0.35
	λ	2.07	5.69	0.79	2.67	0.60	1.79	0.36	1.49
PL-$PE_{(1.5)}$	β_1	−0.00	0.12	−0.00	0.12	−0.00	0.08	−0.00	0.08
	β_2	−0.00	0.24	−0.00	0.24	−0.00	0.15	−0.00	0.15
	τ_1	0.20	0.66	0.05	0.48	0.08	0.35	0.04	0.32
	τ_2	−0.15	0.60	−0.06	0.54	−0.06	0.36	−0.04	0.34
	λ	2.08	5.71	0.80	2.60	0.61	1.91	0.41	1.60
PL-$Hyp_{(1.2)}$	β_1	−0.00	0.16	0.00	0.16	−0.00	0.10	−0.00	0.10
	β_2	−0.01	0.33	−0.01	0.33	−0.00	0.20	−0.00	0.20
	τ_1	0.15	0.54	0.04	0.43	0.05	0.30	0.02	0.28
	τ_2	−0.11	0.57	−0.04	0.53	−0.05	0.35	−0.03	0.34
	λ	1.49	3.88	0.72	2.25	0.43	1.37	0.29	1.20
PL-$slash_{(1.4)}$	β_1	−0.02	0.19	−0.02	0.18	−0.01	0.12	−0.00	0.12
	β_2	−0.01	0.37	−0.01	0.37	−0.00	0.23	−0.00	0.23
	τ_1	−0.01	0.49	−0.10	0.43	−0.01	0.29	−0.04	0.28
	τ_2	0.01	0.57	0.06	0.55	0.01	0.35	0.04	0.34
	λ	0.50	3.79	−0.08	1.70	0.14	1.24	−0.04	1.10

for the PL-$t_{(5)}$ regression model, the bias of $\widehat{\beta}_2$ is zero up to two decimal places for $n = 40$. The penalization seems not to interfere in the estimation of the β's. The estimates of the parameters associated with the dispersion submodel, that is, τ_1 and τ_2, present small bias. The penalized maximum likelihood estimators have smaller bias than the usual maximum likelihood estimators; in the PL-$t_{(5)}$ regression model, for example, the bias of $\widehat{\tau}_1$, when $n = 40$, is 0.21, while the bias of $\widetilde{\tau}_1$ is 0.05. The corresponding RMSE are also smaller when the penalization is used; for the PL-$PE_{(1.5)}$ regression model and $n = 40$, the RMSE of $\widehat{\tau}_1$ is 0.66 and 0.48 for $\widetilde{\tau}_1$. The penalization is effective in the estimation of λ. In all the scenarios, when the sample size is small ($n = 40$), the bias of the $\widehat{\lambda}$ is considerably large, as well as the mean squared error. The penalized maximum likelihood estimator has much smaller bias and mean squared error. For instance, in the PL-$PE_{(1.5)}$ regression model with $n = 40$, the bias of $\widehat{\lambda}$ is 2.08 and RMSE = 5.71, while the bias and RMSE of $\widetilde{\lambda}$ are, respectively, 0.80 and 2.60. As expected, the bias and RMSE of the estimators decrease with increasing sample size. Simulation studies with other power logit models were carried out and similar results were observed. We recommend using the penalized maximum likelihood estimator when the sample size is small. For large or moderate samples sizes, the usual and penalized maximum likelihood estimators have similar performances.

3.3 Choosing the extra parameter ζ

As mentioned before, the density generator function $r(\cdot)$ may involve an extra parameter, denoted by ζ. This parameter is considered fixed in the estimation process. To select a suitable value for ζ,

we suggest to choose $\widehat{\zeta}$ such that

$$\widehat{\zeta} = \underset{\zeta \in \Theta^{\zeta}}{\operatorname{argmin}} \Upsilon_{\zeta},$$

with

$$\Upsilon_{\zeta} = n^{-1} \sum_{i=1}^{n} |\Phi^{-1}[R(\widetilde{z}^{(i)})] - \upsilon^{(i)}|, \tag{3.5}$$

where Θ^{ζ} represents the parameter space of ζ, $\widetilde{z}^{(i)}$ is the ith order statistic of $\widetilde{z}$, $\upsilon^{(i)}$ is the mean of the ith order statistic in a random sample of size n of the standard normal distribution and $\Phi(\cdot)$ is the cdf of the standard normal distribution. Note that $\Phi^{-1}[R(Z)]$ has a standard normal distribution. This measure was proposed by Vanegas and Paula (2015) and the idea is to choose the value of ζ such that $(\Phi^{-1}[R(\widetilde{z}^{(1)})], \ldots, \Phi^{-1}[R(\widetilde{z}^{(n)})])$ be as close as possible to an ordered sample of the standard normal distribution. Numerical results presented in Section 6 of the Supplementary Material show a good performance of this criterion for the choice of ζ.

4 Diagnostic tools

In this section we present some diagnostic tools for the power logit regression models. First, we define two overall goodness-of-fit measures: the pseudo R^2 (R_p^2) and the Υ_{ζ} measure. Following Ferrari and Cribari–Neto (2004), the pseudo R^2 is defined as the square of the sample correlation coefficient between the estimated linear predictor $d_1(\widetilde{\mu})$ and $d_1(y)$. Note that $0 \leq R_p^2 \leq 1$ and a perfect agreement between $\widetilde{\mu}$ and y yields $R_p^2 = 1$. The Υ_{ζ} measure is defined in (3.5). Small values of Υ_{ζ} suggest that the fitted model suitably describes the data.

Some residuals for the power logit regression models, as well as influence and leverage measures, are presented below.

4.1 Residuals

We propose three residuals for the power logit regression models: the quantile residual, the deviance residual, and the standardized residual.

Quantile residual: Dunn and Smyth (1996) defined the quantile residual, which has a standard normal distribution asymptotically if the model is correctly specified and the parameter estimators are consistent. For the power logit regression models, the quantile residual is defined as

$$r_i^q = \Phi^{-1}[R(\widetilde{z}_i)], \quad i = 1, \ldots, n,$$

where $R(\cdot)$ is the cdf of $Z \sim \mathrm{S}(0, 1; r)$.

Deviance residual: The deviance residual is defined as

$$\text{sign}(y_i - \widetilde{\mu}_i)\{2[\ell_i(\bar{\mu}_i, \widetilde{\sigma}_i, \widetilde{\lambda}) - \ell_i(\widetilde{\mu}_i, \widetilde{\sigma}_i, \widetilde{\lambda})]\}^{1/2}, \quad i = 1, \ldots, n,$$

where $\bar{\mu}_i$ is the mle of μ_i under the saturated model (McCullagh and Nelder, 1989). This residual measures the discrepancy of the fitted model and the data as twice the difference between the maximum log-likelihood achievable and that achieved by the postulated model. For the power logit regression models, we have $\bar{\mu}_i = y_i$ and the deviance residual can be written as

$$r_i^d = \text{sign}(\widetilde{z}_i)\left\{2\log\left[\frac{r(0)}{r(\widetilde{z}_i^2)}\right]\right\}^{1/2}, \quad i = 1, \ldots, n.$$

For the PL-N models, the quantile and deviance residuals coincide and $r_i^q = r_i^d = \widetilde{z}_i$.

Standardized residual: The power logit regression models with constant dispersion are equivalent to

$$y_i^\dagger = \mu_i^\dagger(\boldsymbol{\beta}; \boldsymbol{x_i}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{4.1}$$

where $\varepsilon_i \sim \text{S}(0, \sigma^2; r)$, $y_i^\dagger = \log[y_i^\lambda/(1 - y_i^\lambda)]$, and $\mu_i^\dagger(\boldsymbol{\beta}; \boldsymbol{x_i}) = \mu_i^\dagger = \log[\mu_i^\lambda/(1 - \mu_i^\lambda)]$ is a nonlinear function of $\boldsymbol{\beta}$ with matrix of derivatives $\mathbf{D} = \partial\boldsymbol{\mu}^\dagger/\partial\boldsymbol{\beta} = \sigma\boldsymbol{\mu}^*\mathbf{T}_1\mathbf{X}$, with $\boldsymbol{\mu}^\dagger = (\mu_1^\dagger, \ldots, \mu_n^\dagger)^\top$. Assuming that λ is fixed, model (4.1) is a symmetric nonlinear regression model (Galea et al., 2005; Cysneiros and Vanegas, 2008). Then, an ordinary residual may be defined as $r_i = y_i^\dagger - \widetilde{\mu}_i^\dagger$, for $i = 1, \ldots, n$. Using expansions up to order n^{-1} from Cox and Snell (1968), when they exist, $\mathbb{E}(\boldsymbol{r}) \approx (\mathbf{I}_n - \mathbf{H})\boldsymbol{\eta}$ and $\text{Var}(\boldsymbol{r}) \approx \sigma^2\xi_r\{\mathbf{I}_n - (d_r\xi_r)^{-1}\mathbf{H}\}$ where $\boldsymbol{r} = (r_1, \ldots, r_n)^\top$, $d_r = d_r(\zeta) = \mathbb{E}[Z^2v^2(Z)]$, ξ_r is such that $\text{Var}(Z) = \xi_r$, $\mathbf{H} = \mathbf{D}(\mathbf{D}^\top\mathbf{D})^{-1}\mathbf{D}^\top$, $\mathbf{I}_n$ is the identity matrix of order n and $\boldsymbol{\eta}$ is the difference between the linear and quadratic approximations of $\mu_i^\dagger(\widetilde{\boldsymbol{\beta}}; \boldsymbol{x_i})$. Therefore, we define a standardized residual for the power logit regression models with constant dispersion based on r_i as

$$r_i^p = \frac{\widetilde{y}_i^\dagger - \widetilde{\mu}_i^\dagger}{\widetilde{\sigma}\sqrt{\widetilde{\xi}_r\{1 - (\widetilde{d}_r\widetilde{\xi}_r)^{-1}\widetilde{h}_{ii}\}}}, \quad i = 1, \ldots, n, \tag{4.2}$$

where $\widetilde{y}_i^\dagger = \log(y_i^{\widetilde{\lambda}}/(1 - y_i^{\widetilde{\lambda}}))$, $\widetilde{h}_{ii}$ is the ith diagonal element of $\mathbf{H}$ evaluated in $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\tau}}, \widetilde{\lambda})$. For power logit regression models with varying dispersion, the standardized residual takes the form (4.2), with $\widetilde{\sigma}$ replaced by $\widetilde{\sigma}_i$ and $\widetilde{h}_{ii}$ being the ith diagonal element of $\mathbf{H} = \boldsymbol{\Sigma}^{-1/2}\mathbf{D}(\mathbf{D}^\top\boldsymbol{\Sigma}^{-1}\mathbf{D})^{-1}\mathbf{D}^\top\boldsymbol{\Sigma}^{-1/2}$ evaluated at $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\tau}}, \widetilde{\lambda})$, where $\boldsymbol{\Sigma} = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$.

Table 2 presents ξ_r and d_r for some power logit models. In Table 2, $K_2(\cdot)$ denotes the modified Bessel function of third order and index 2, $h_1(\zeta) = \int_1^\infty(\sqrt{x^2 - 1}/x)\exp\{-\zeta x\}\text{d}x$, $\text{erf}(x) = (2/\sqrt{\pi})\int_0^z \exp\{-t^2\}\text{d}t$ is the error function, and

$$q(\zeta) = \begin{cases} (\zeta^2/4)\left(1 - \zeta^2/4\right), & \text{for } 0 < \zeta < 2, \\ 2.197543451 - 1.963510026\log(\zeta\sqrt{2}) + [\log(\zeta\sqrt{2})]^2, & \text{for } \zeta \geq 2. \end{cases}$$

Table 2 ξ_r and d_r for some power logit models

Model	ξ_r	d_r
PL-N	1	1
PL-t$_{(\zeta)}$	$\zeta/(\zeta-2)$, $\zeta>2$	$(\zeta+1)/(\zeta+3)$
PL-PE$_{(\zeta)}$	1	$\zeta^2\Gamma(1/\zeta)^{-2}\,\Gamma(3/\zeta)\,\Gamma\big((2\zeta-1)/\zeta\big)$, $\zeta>1/2$
PL-LOI	≈ 0.79569	≈ 1.47724
PL-LOII	$\pi^2/3$	$1/3$
PL-slash$_{(\zeta)}$	$\dfrac{\zeta}{\zeta-1}$, $\zeta>1$	$\approx \dfrac{4\zeta(\zeta+1/2)[(\zeta+3/2)(\zeta+5/2)+\zeta+1]}{(\zeta+1)(\zeta+3/2)^2(\zeta+5/2)}$
PL-Hyp$_{(\zeta)}$	$K_2(\zeta)/[\zeta K_1(\zeta)]$	$\zeta^2 h_1(\zeta)/K_1(\zeta)$
PL-SN$_{(\zeta)}$	$\approx q(\zeta)$	$\approx 2+4/\zeta^2-(\sqrt{2\pi}/\zeta)(1-\mathrm{erf}(\sqrt{2}/\zeta))\exp\{2/\zeta^2\}$

In order to study the performance of the proposed residuals in detecting outlier observations, we conducted an application in simulated data. We generated a sample of size $n=40$ of a power logit normal regression model with constant dispersion and $\log(\mu_i/(1-\mu_i))=3.5-3.5x_i$, for $i=1,\ldots,40$, where x_i is taken from a uniform distribution on the unit interval, for $i=1,\ldots,38$, and $x_{39}=0.8$, $x_{40}=1.2$, $\log\sigma=-1.5$ and $\lambda=0.5$. We contaminated the data as follows: we replaced y_{39} and y_{40} by $y^*_{39}=0.9$ and $y^*_{40}=\max y_i$. These datasets, contaminated and uncontaminated, were analysed using the PL-N and PL-t$_{(5)}$ regression models. The results are presented in Figure 4. The plots suggest that the proposed residuals are efficient in identifying atypical observations. For the PL-t$_{(5)}$ regression model, the standardized residual highlights more clearly observation #40, that is a leverage point. Furthermore, we observe that the PL-t$_{(5)}$ regression model is less sensitive to outliers than the PL-N regression model, as expected. This is because $v(z)$ returns small weights for observations with large residuals in the PL-t$_{(\zeta)}$ models, when ζ is not large. This experiment was performed for other power logit regression models and similar results were observed. Other applications in simulated data are presented in Section 7 of the Supplementary Material in order to verify different types of perturbations in the model.

4.2 Local influence

The local influence analysis is a diagnostic method proposed by Cook (1986) that evaluates the effect of small perturbations in the data or the model. For the power logit regression models, we are interested in evaluating the influence of small perturbations on the estimation of the parameters associated with the median and dispersion. Thus, we consider λ as a fixed constant. In practice, this parameter is replaced by its (penalized) maximum likelihood estimate. Therefore, let $\boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\boldsymbol{\tau}^\top)^\top$. To assess the local influence, we use the likelihood displacement $LD_{\boldsymbol{\omega}}=2[\ell(\widehat{\boldsymbol{\theta}})-\ell(\widehat{\boldsymbol{\theta}}_\omega)]$, where $\widehat{\boldsymbol{\theta}}_\omega$ denotes the mle under the perturbed model. Thus, the normal curvature for $\boldsymbol{\theta}$ at the direction $\boldsymbol{h}$, $\|\boldsymbol{h}\|=1$, is given by $C_{\boldsymbol{h}}=2|\boldsymbol{h}^\top\boldsymbol{\Delta}^\top\ddot{\boldsymbol{\ell}}^{-1}_{\widehat{\theta}\widehat{\theta}}\boldsymbol{\Delta}\boldsymbol{h}|$, where $\ddot{\boldsymbol{\ell}}_{\theta\theta}=-\mathbf{J}_n(\boldsymbol{\beta},\boldsymbol{\tau},\lambda)$ and $\boldsymbol{\Delta}$ is a $(p+q)\times n$ matrix that depends on the perturbation scheme and is defined as $\boldsymbol{\Delta}=\partial^2\ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top$ evaluated at $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$, the no perturbation vector. The index graph of $\boldsymbol{h}_{\max}$, the eigenvector corresponding to the highest absolute eigenvalue of $-\boldsymbol{\Delta}^\top\ddot{\boldsymbol{\ell}}^{-1}_{\widehat{\theta}\widehat{\theta}}\boldsymbol{\Delta}$, can reveal the most influential observations in $\widehat{\boldsymbol{\theta}}$. Another possibility was proposed by Lesaffre and Verbeke (1998), named total local influence, which consists of the construction of the index plot of $C_i=2|\boldsymbol{\Delta}_i^\top\ddot{\boldsymbol{\ell}}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{\Delta}_i|$, where $\boldsymbol{\Delta}_i$ is the ith column of $\boldsymbol{\Delta}$.

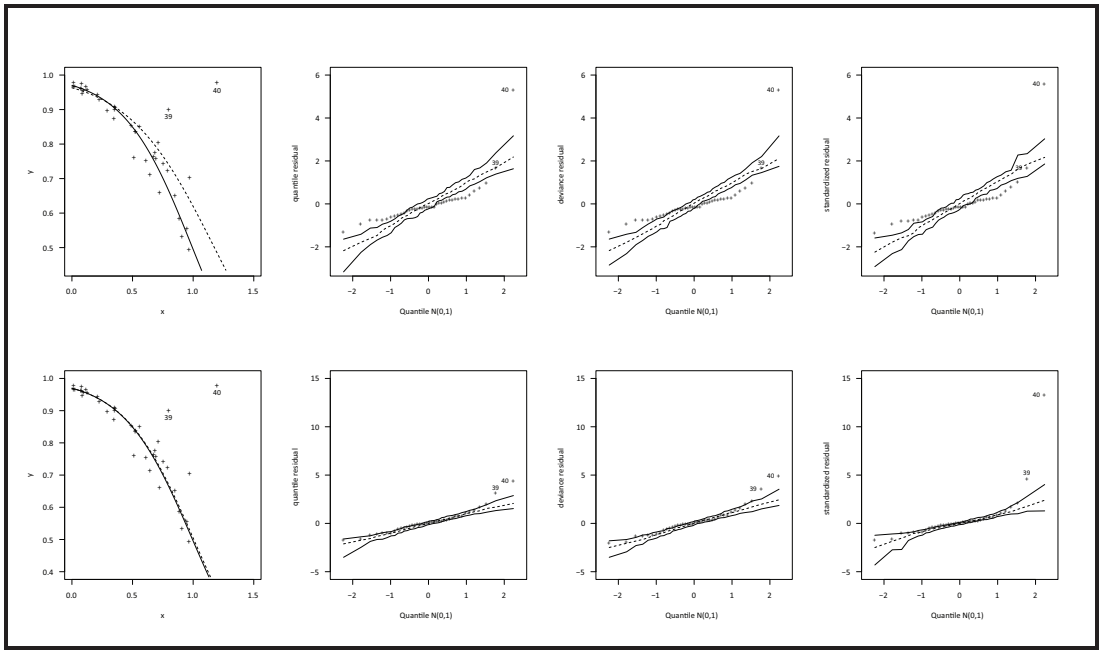

Figure 4 Scatter plots with the fitted lines for the uncontaminated (solid line) and contaminated (dashed line) data and normal probability plots of the quantile, deviance and standardized residuals with simulated envelopes for the contaminated data; PL-N (top line) and PL-$t_{(5)}$ (bottom line)

In this work, we consider two perturbation schemes: case-weights and covariate perturbation. The structure of $\boldsymbol{\Delta}$ for each perturbation scheme is given in the following.

Case-weights perturbation: Here, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ is an $n \times 1$ vector of weights, with $0 \leq w_i \leq 1$, for $i = 1, \ldots, n$, and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. The perturbed log-likelihood function is $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \ell_i(\mu_i, \sigma_i, \lambda)$. For this perturbation scheme, $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\boldsymbol{\beta}}^\top, \boldsymbol{\Delta}_{\boldsymbol{\tau}}^\top)^\top$, where

$$\boldsymbol{\Delta}_{\boldsymbol{\beta}} = \mathbf{X}^\top \widehat{\mathbf{W}} \widehat{\mathbf{T}}_1 \widehat{\mathbf{D}}_\beta \quad \text{and} \quad \boldsymbol{\Delta}_{\boldsymbol{\tau}} = \mathbf{S}^\top \widehat{\mathbf{T}}_2 \widehat{\mathbf{D}}_\tau,$$

where $\mathbf{D}_\beta = \text{diag}\{\mu_1^*, \ldots, \mu_n^*\}$ and $\mathbf{D}_\tau = \text{diag}\{\sigma_1^*, \ldots, \sigma_n^*\}$.

Median covariate perturbation: Now, we perturb a continuous covariate in the median submodel, say $\boldsymbol{x}_j$. Following Thomas and Cook (1989), we replace x_{ij} by $x_{ij\omega} = x_{ij} + \sigma_x \omega_i$, where σ_x is the sample standard deviation of $\boldsymbol{x}_j$. The perturbed linear predictor for the median submodel is

$$\eta_{1i\omega} = \beta_1 x_{i1} + \cdots + \beta_j(x_{ij} + \sigma_x \omega_i) + \cdots + \beta_p x_{ip} = \boldsymbol{x}_{i\omega}^\top \boldsymbol{\beta},$$

where $\boldsymbol{x}_{i\omega} = (x_{i1}, \ldots, x_{ij\omega}, \ldots, x_{ip})^\top$. The perturbed log-likelihood function is $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\mu_{i\omega}, \sigma_i, \lambda)$, where $\mu_{i\omega} = d_1^{-1}(\eta_{1i\omega})$ and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$. Assuming that $\mathbf{X}$ and $\mathbf{S}$ are

functionally independent, the components of $\mathbf{\Delta}$ are

$$\mathbf{\Delta}_{\boldsymbol{\beta}} = -\sigma_x\widehat{\beta}_j\mathbf{X}^\top\widehat{\mathbf{W}}_1\widehat{\mathbf{T}}_1 + \sigma_x\boldsymbol{c}_{\boldsymbol{\beta}}^j\widehat{\boldsymbol{\mu}}^{*\top}\widehat{\mathbf{W}}\widehat{\mathbf{T}}_1 \quad \text{and} \quad \mathbf{\Delta}_{\boldsymbol{\tau}} = -\sigma_x\widehat{\beta}_j\mathbf{S}^\top\widehat{\mathbf{W}}_4\widehat{\mathbf{T}}_1\widehat{\mathbf{T}}_2,$$

where $\boldsymbol{c}_{\boldsymbol{\beta}}^j$ denotes a $p \times 1$ vector with one at the jth position and zero elsewhere, and $\widehat{\beta}_j$ denotes the jth element of $\widehat{\boldsymbol{\beta}}$.

Dispersion covariate perturbation: Here, we perturb a continuous covariate in the dispersion submodel, say $\boldsymbol{s}_k$. We replace s_{ik} by $s_{ik\omega} = s_{ik} + \sigma_s\omega_i$, where σ_s is the sample standard deviation of $\boldsymbol{s}_k$. The perturbed linear predictor for the dispersion submodel is

$$\eta_{2i\omega} = \tau_1 s_{i1} + \cdots + \tau_k(s_{ik} + \sigma_s\omega_i) + \cdots + \tau_q s_{iq} = \boldsymbol{s}_{i\omega}^\top\boldsymbol{\tau},$$

where $\boldsymbol{s}_{i\omega} = (s_{i1}, \ldots, s_{ik\omega}, \ldots, s_{iq})^\top$. The perturbed log-likelihood function is $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\mu_i, \sigma_{i\omega}, \lambda)$, where $\sigma_{i\omega} = d_2^{-1}(\eta_{2i\omega})$, and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$. Assuming that $\mathbf{X}$ and $\mathbf{S}$ are functionally independent, the components of $\mathbf{\Delta}$ are

$$\mathbf{\Delta}_{\boldsymbol{\beta}} = -\sigma_s\widehat{\tau}_k\mathbf{X}^\top\widehat{\mathbf{W}}_4\widehat{\mathbf{T}}_1\widehat{\mathbf{T}}_2 \quad \text{and} \quad \mathbf{\Delta}_{\boldsymbol{\tau}} = -\sigma_s\widehat{\tau}_k\mathbf{S}^\top\widehat{\mathbf{W}}_2\widehat{\mathbf{T}}_2 + \sigma_s\boldsymbol{c}_{\boldsymbol{\tau}}^k\widehat{\boldsymbol{\sigma}}^{*\top}\widehat{\mathbf{T}}_2,$$

where $\boldsymbol{c}_{\boldsymbol{\tau}}^k$ denotes a $q \times 1$ vector with one at the kth position and zero elsewhere, and $\widehat{\tau}_k$ denotes the kth element of $\widehat{\boldsymbol{\tau}}$.

Simultaneous (median and dispersion) covariate perturbation: Now, we perturb a continuous covariate in the median and in the dispersion submodels, say $\boldsymbol{x}_j$ and $\boldsymbol{s}_k$. We replace x_{ij} and s_{ik} by $x_{ij\omega} = x_{ij} + \sigma_x\omega_i$ and $s_{ik\omega} = s_{ik} + \sigma_s\omega_i$, respectively. Then

$$\eta_{1i\omega} = \beta_1 x_{i1} + \cdots + \beta_j(x_{ij} + \sigma_x\omega_i) + \cdots + \beta_p x_{ip} = \boldsymbol{x}_{i\omega}^\top\boldsymbol{\beta},$$
$$\eta_{2i\omega} = \tau_1 s_{i1} + \cdots + \tau_k(s_{ik} + \sigma_s\omega_i) + \cdots + \tau_q s_{iq} = \boldsymbol{s}_{i\omega}^\top\boldsymbol{\tau},$$

The perturbed log-likelihood function is $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\mu_{i\omega}, \sigma_{i\omega}, \lambda)$, and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$ and

$$\mathbf{\Delta} = \begin{bmatrix} -\sigma_x\widehat{\beta}_j\mathbf{X}^\top\widehat{\mathbf{W}}_1\widehat{\mathbf{T}}_1 + \sigma_x\boldsymbol{c}_{\boldsymbol{\beta}}^j\widehat{\boldsymbol{\mu}}^{*\top}\widehat{\mathbf{W}}\widehat{\mathbf{T}}_1 - \sigma_s\widehat{\tau}_k\mathbf{X}^\top\widehat{\mathbf{W}}_4\widehat{\mathbf{T}}_1\widehat{\mathbf{T}}_2 \\ -\sigma_s\widehat{\tau}_k\mathbf{S}^\top\widehat{\mathbf{W}}_2\widehat{\mathbf{T}}_2 + \sigma_s\boldsymbol{c}_{\boldsymbol{\tau}}^k\widehat{\boldsymbol{\sigma}}^{*\top}\widehat{\mathbf{T}}_2 - \sigma_x\widehat{\beta}_j\mathbf{S}^\top\widehat{\mathbf{W}}_4\widehat{\mathbf{T}}_1\widehat{\mathbf{T}}_2. \end{bmatrix}.$$

4.3 Generalized leverage

The generalized leverage measure is defined by $\mathrm{GL}_{ij} = \partial\widehat{y}_i/\partial y_j$ and reflects the rate of instantaneous change in the ith predicted value, when the jth response variable is increased by an infinitesimal amount (Wei et al., 1998). The idea is to use GL_{ii} to evaluate the influence of y_i on its own predicted value. A high leverage value suggests that the observation may be an outlier on the covariates. Wei et al. (1998) showed that the generalized leverage matrix can be expressed as

$$\mathrm{GL}(\boldsymbol{\theta}) = \dot{\mathbf{L}}_{\boldsymbol{\theta}}\mathbf{J}_n(\boldsymbol{\theta})^{-1}\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{y}},$$

where $\dot{\mathbf{L}}_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}$ and $\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{y}} = \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{y}^\top$, with $\boldsymbol{\theta}$ evaluated at $\widehat{\boldsymbol{\theta}}$. For the power logit regression models, we consider the penalized maximum likelihood estimator, which has better small sample properties. After some algebra, we can show that $\dot{\mathbf{L}}_{\boldsymbol{\theta}} = [\mathbf{T}_1\mathbf{X}, \ \mathbf{0}_{n,q}, \ \mathbf{0}_{n,1}]$ and

$$\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{y}} = \begin{bmatrix} \mathbf{X}^\top\mathbf{T}_1\mathbf{D}_\beta\mathbf{D}_y\mathbf{W}_\beta \\ \mathbf{S}^\top\mathbf{T}_2\mathbf{D}_y\mathbf{W}_\tau \\ \mathbf{1}_n^\top\mathbf{D}_y\mathbf{W}_\lambda \end{bmatrix},$$

where $\mathbf{0}_{a,b}$ is an $a \times b$ matrix of zeros, $\mathbf{D}_\beta = \text{diag}\{\mu_1^*, \ldots, \mu_n^*\}$, $\mathbf{D}_y = \text{diag}\{y_1^*, \ldots, y_n^*\}$, $\mathbf{W}_\beta = \text{diag}\{v(z_1) + z_1v'(z_1), \ldots, v(z_n) + z_nv'(z_n)\}$, $\mathbf{W}_\tau = \text{diag}\{\sigma_1^{-1}z_1[2v(z_1) + z_1v'(z_1)], \ldots, \sigma_n^{-1}z_n[2v(z_n) + z_nv'(z_n)]\}$, $\mathbf{W}_\lambda = \text{diag}\{w_1^\lambda, \ldots, w_n^\lambda\}$, $y_i^* = \lambda/[\sigma_i y_i(1 - y_i^\lambda)]$, and

$$w_i^\lambda = \frac{\sigma_j y_i^\lambda(1 + \lambda\log y_j - y_j^\lambda)}{\lambda(1 - y_j^\lambda)} - \frac{1}{\sigma_j}\left(\frac{\log y_j}{1 - y_j^\lambda} - \frac{\log\mu_j}{1 - \mu_j^\lambda}\right)[v(z_j) + z_jv'(z_j)]$$
$$- z_jv(z_j)\frac{y_i^\lambda(\lambda\log y_j - 1) + 1}{\lambda(1 - y_j^\lambda)},$$

for $i = 1, \ldots, n$. The index plot of the diagonal elements of $\text{GL}(\widetilde{\boldsymbol{\theta}})$ may be used to assess observations with high influence on their own predicted values.

5 R implementation: `PLreg` package

To facilitate the practical use of the proposed models, we developed the new R package `PLreg` available at `CRAN` (`https://CRAN.R-project.org/package=PLreg`). It allows fitting power logit regression models, in which the density generator may be normal, Student-t, power exponential, slash, hyperbolic, sinh-normal, or type II logistic. Diagnostic tools associated with the fitted model, such as the residuals, local influence measures, leverage measures, and goodness-of-fit statistics, are implemented. The estimation process follows the maximum likelihood approach and, currently, the package supports two types of estimators: the usual maximum likelihood estimator and the penalized maximum likelihood estimator. The skewness parameter λ may be fixed, so the package also allows fitting GJS ($\lambda = 1$) and log-log ($\lambda = 0$) regression models.

The main function of the `PLreg` package is `PLreg()`, which follows the standard approach for implementing regression models in `R`. Once the fitting process has been accomplished, an object of the `S3` class '`PLreg`' is produced for which several methods are available. The arguments of the `PLreg()` function are similar to those of functions in other packages for regression models in `R`, such as the `betareg()` function.

The codes for the applications in the next section are available at the GitHub repository `https://github.com/ffqueiroz/PowerLogitRegression`.

Table 3 Estimates, standard errors, asymptotic 95% confidence intervals for μ, Υ measure, and AIC for the beta, GJS and PL models - Employment in non-agricultural sectors data

	Est. (*SE*)					
	μ	σ	λ	CI	Υ	AIC
beta	0.82 (0.07)	5.43 (0.54)		(0.80, 0.84)	0.15	−286.00
GJS-N	0.88 (0.01)	1.42 (0.07)		(0.86, 0.90)	0.22	−255.20
GJS-$t_{(4.56)}$	0.85 (0.01)	1.19 (0.08)		(0.83, 0.88)	0.21	−248.63
GJS-$PE_{(1.44)}$	0.85 (0.01)	1.46 (0.09)		(0.83, 0.88)	0.21	−249.88
GJS-$SN_{(0.53)}$	0.88 (0.01)	5.54 (0.26)		(0.86, 0.90)	0.23	−255.05
PL-N	0.84 (0.01)	2.43 (0.12)	9.41 (0.75)	(0.81, 0.86)	0.07	−306.54
PL-$t_{(100)}$	0.84 (0.01)	2.41 (0.12)	9.44 (0.76)	(0.81, 0.86)	0.07	−305.98
PL-$PE_{(2.69)}$	0.84 (0.01)	2.38 (0.10)	9.16 (0.65)	(0.81, 0.86)	0.05	−310.81
PL-$SN_{(0.97)}$	0.84 (0.01)	5.31 (0.23)	9.06 (0.63)	(0.81, 0.86)	0.06	−310.63

6 Applications

6.1 Employment in non-agricultural sectors

The dataset refers to the employment in non-agricultural sectors in 200 randomly selected Brazilian municipal districts of the state of São Paulo in the year 2010. The data were extracted from the Atlas of Brazil Human Development database, available at `https://www.pnud.org.br/atlas/`. The response variable is the proportion of people aged 18 or over who are employed in non-agricultural activities (y).

We fitted different distributions in the power logit class, including some GJS distributions. The extra parameter, if any, was chosen as described in Section 3.3. For the sake of comparison we also fitted the beta distribution because it is the most employed distribution to model continuous proportions. The parameterization of the beta law uses the mean and a precision parameter (Ferrari and Cribari–Neto, 2004), denoted here by μ and σ. The results are presented in Table 3 [2]. The fitted models with the smallest values of Υ and AIC are those of the power logit distributions; for instance, the Υ value of the GJS-$SN_{(0.53)}$ model is almost four times that of the PL-$SN_{(0.97)}$ model. Based on these measures, all the PL models fit the data similarly and better than the beta and the GJS distributions. In addition, the estimates of λ are large (around 9), indicating that the GJS distributions ($\lambda = 1$) may not be suitable to represent the distribution of the data.

Figure 5 displays some diagnostic plots. Since the Υ measure and the AIC are similar for all of the power logit distributions, we chose the PL-N distribution to summarize the data, because of its simplicity. Figures 5(a)–5(c) present the normal probability plots with simulated envelope of the quantile residual for the beta, GJS-N and PL-N. These plots show that the beta and GJS-N models do not fit the data well, but the PL-N distribution seems to be suitable for modeling the data. This is confirmed in Figures 5(d)–5(f)[3]. Diagnostic plots for the other GJS distributions were done and showed similar behavior. Based on the PL-N model, the estimated median proportion of

[2]The Υ measure for the beta distribution is computed similarly to (3.5).

[3]The relative quantile discrepancy in Figure 5(d) is the difference between estimated quantiles and empirical quantiles divided by the latter.

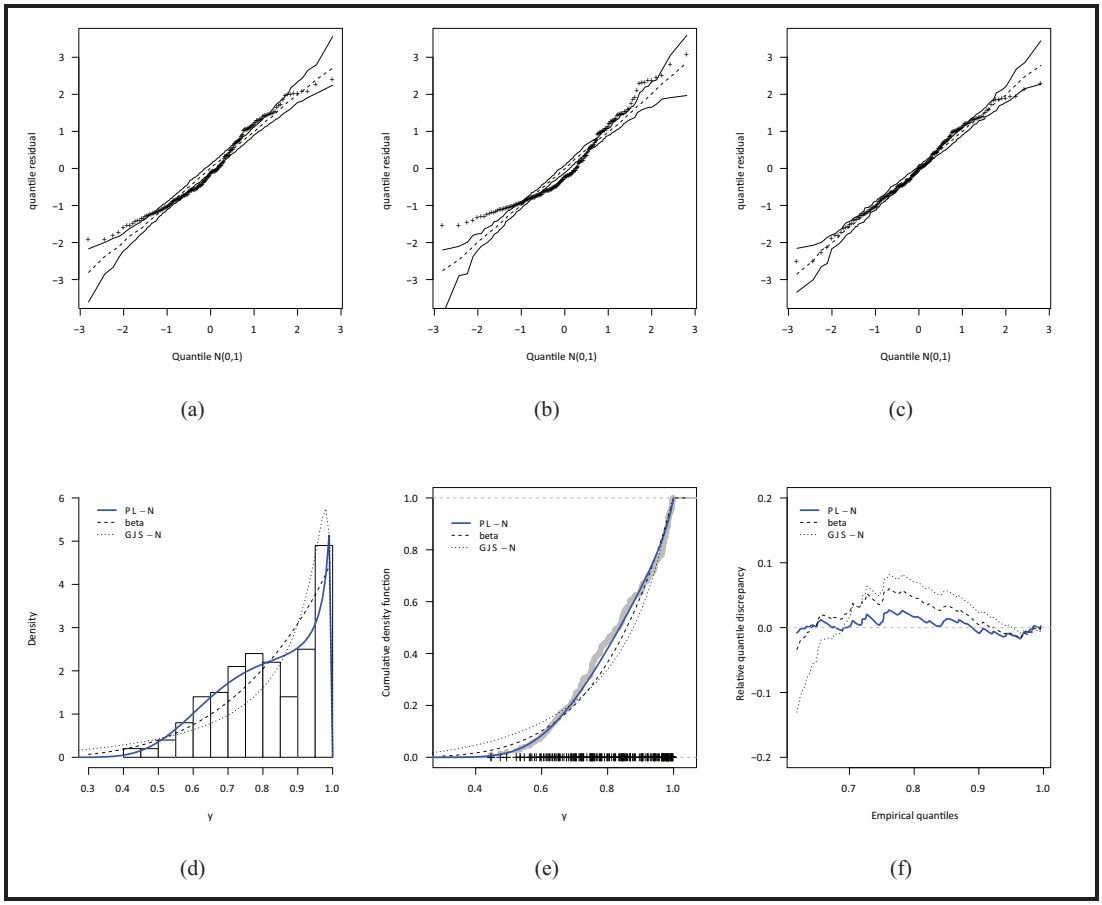

Figure 5 Normal probability plots of the quantile residual for the beta (a), GJS-N (b) and PL-N (c) models, histogram of y with fitted pdfs (d), fitted cdfs (e) and quantile relative discrepancies (f) for the beta, GJS-N and PL-N models - Employment in non-agricultural sectors data

adults who are employed in non-agricultural sectors in the cities of the state of São Paulo is 0.84. The 95% confidence interval is (0.81, 0.86).

Except for the beta distribution, the models considered in the analysis above belong to the power logit class. Other distributions could possibly provide a suitable fit for the data. A Poisson model with offset could be a natural candidate if the total number of employed people aged 18 or over in each district were available, which is not the case here. We then fitted the following four-parameter continuous distributions: generalized beta type 1 (GB1, Rigby et al., 2020, Section 6.2), logit-sinh-arcsinh (logitSHASH) and logit-skew-t type 3 (logitST3), which are obtained as the distribution of the inverse logit transformation of a sinh-arcsinh (Jones and Pewsey, 2009) and a skew-Student-t type 3 (Rigby et al., 2020, Section 18.4.11) random variable respectively, and flexible beta (FB, Migliorati et al., 2018), that is a mixture of two beta distributions with different means and the same precision parameter. In addition, we fitted the three-parameter logit-skew-normal type 2

distribution (logitSN2), obtained as the distribution of the inverse logit transformation of a skew-normal type 2 random variable (Rigby et al., 2020, Section 13.5.2.1). The parameters of the FB distribution can be interpreted as the mean, a precision parameter, the distance between the two components and the mixture parameter. The parameters of the GB1, logitSHASH, logitSN2, and logitST3 distributions are not directly expressed as moments or quantiles, which is inconvenient for modeling and interpretation purposes. The GB1, logitSHASH, logitSN2, and logitST3 distributions were fitted using the package `gamlss` in R. We implemented our own code for fitting an FB distribution. Details are given in Section 8 of the Supplementary Material. Overall, the four distributions provided fitted models comparable with the fitted PL-N model, with a slight advantage for the FB distribution and an apparent disadvantage for the logitSN2 and logitST3 distributions. Considering parsimony, simplicity, and ease of parameter interpretation, the PL-N model still proves to be an appropriate distribution to represent the data.

6.2 Firm cost data

This application is from a questionnaire sent to risk managers of large corporations in the USA. The dataset was introduced by Schmit and Roth (1990) (see also Gómez-Déniz et al., 2014; Ribeiro and Ferrari, 2022) and is available in the personal web page of Professor E. Frees (Wisconsin School of Business Research) [4]. The response variable is `firmcost`, defined as premiums plus uninsured losses as a percentage of the total assets and is a measure of the firm's risk management cost effectiveness. A good risk management performance corresponds to smaller values of `firmcost`. Following Ribeiro and Ferrari (2022), we considered two covariates: `sizelog`, the logarithm of total assets, and `indcost`, a measure of the firm's industry risk. The dataset contains information on 73 firms. The postulated initial model is a power logit slash regression model with

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 \times \texttt{sizelog}_i + \beta_3 \times \texttt{indcost}_i,$$
$$\log \sigma_i = \tau_1 + \tau_2 \times \texttt{sizelog}_i + \tau_3 \times \texttt{indcost}_i,$$

for $i = 1, \ldots, 73$. The results are presented in Table 4.

Note that all the covariates are statistically significant for the median submodel. For the dispersion submodel, both covariates are not significant. Removing each covariate at a time, the other covariate remains non significant, indicating that a constant dispersion model should be considered; see Table 5. The fitted model indicates that the median `firmcost` is positively related with `indcost` and negatively related with `sizelog`. The estimate of λ is 1.79 and the asymptotic 95% confidence interval based on the profile penalized likelihood ratio statistic is (0.59, 8.50), indicating that the GJS-slash regression model ($\lambda = 1$) might be suitable (see Figure 6(a); the horizontal dashed line gives the 95% confidence interval). Diagnostic plots for the GJS-slash and PL-slash models are presented in Figure 6.

[4] Available at: https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/CSVData/RiskSurvey.csv.

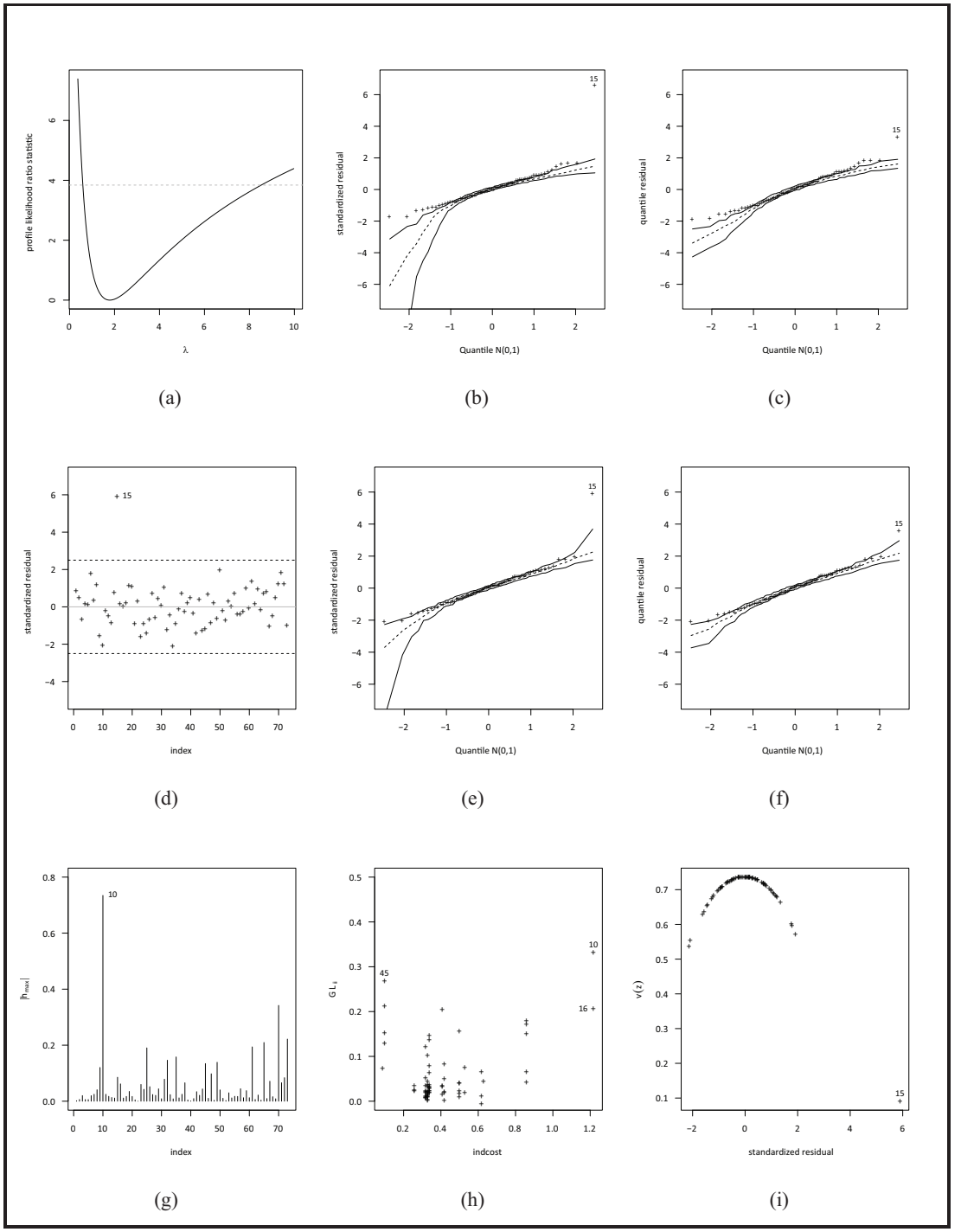

Figure 6 Plot of $W_p^*(\lambda)$ (a), normal probability plots of the standardized residual (b) and quantile residual (c) for the GJS-slash model, standardized residual against index of the observations (d), normal probability plots of the standardized residual, (e) and quantile residual (f), $|h_{\max}|$ under case-weight perturbation (g), GL_{ii} (h), and $v(z)$ against the standardized residual (i) for the PL-slash regression model with constant dispersion - Firm cost data

Table 4 Estimates, standard errors, 95% confidence intervals, and *p*-values for the PL-slash regression model with varying dispersion - Firm cost data

	Est.	*SE*	CI	*p*-value
$\log[\mu/(1-\mu)]$				
`intercept`	3.822	1.304	(1.267, 6.378)	0.003
`indcost`	2.312	1.050	(0.253, 4.370)	0.028
`sizelog`	−0.908	0.164	(−1.230, −0.586)	<0.001
$\log\sigma$				
`intercept`	−0.569	0.546	(−1.639, 0.500)	0.297
`indcost`	0.366	0.595	(−0.801, 1.533)	0.538
`sizelog`	0.074	0.088	(−0.097, 0.246)	0.395
λ	2.04	0.91		
ζ	1.88			
AIC	−232.1			

Table 5 Estimates, standard errors, 95% confidence intervals, and *p*-values for the PL-slash regression model with constant dispersion - Firm cost data

	Est.	*SE*	CI	*p*-value
$\log[\mu/(1-\mu)]$				
`intercept`	3.867	1.165	(1.584, 6.149)	0.001
`indcost`	2.133	0.786	(0.592, 3.674)	<0.007
`sizelog`	−0.905	0.128	(−1.156, −0.654)	0.001
$\log\sigma$	0.133	0.497	(−0.840, 1.107)	0.788
λ	1.79	0.90		
ζ	2.29			
AIC	−234.0			

The PL-slash model fits the data better than the GJS-slash model (compare Figures 6(b)–6(c) with 6(e)–6(f)). In fact, Figures 6(d)–6(f) indicate that the postulated PL-slash regression model suitably fits the data. Case #15 is highlighted in almost all the graphics; this observation corresponds to a firm with the highest value of `firmcost`. On the other hand, this case does not appear in Figure 6(g) as an influential observation; in fact, the weight for this case in the estimation process is small (see Figure 6(i)). Additionally, case #10 is highlighted in the influence plot (Figure 6(g)) and in the generalized leverage plot (Figure 6(h)), but the exclusion of this case from the dataset does not substantially change the fitted model. Overall, we conclude that the PL-slash regression model fits the data well.

Ribeiro and Ferrari (2022) fittted a beta regression model with varying precision (with covariates `indcost` and `sizelog`) to this dataset using the maximum likelihood approach. Observations #15, #16 and #72 were detected as atypical observations. Figure 7(a) displays the scatter plot of `firmcost` versus `indcost` with the fitted lines based on the beta regression model for the full data and the data without outliers; the scatter plots were produced by setting the value of `sizelog` at its sample median. The exclusion of the outliers changes substantially the fitted lines. Case #15 causes the

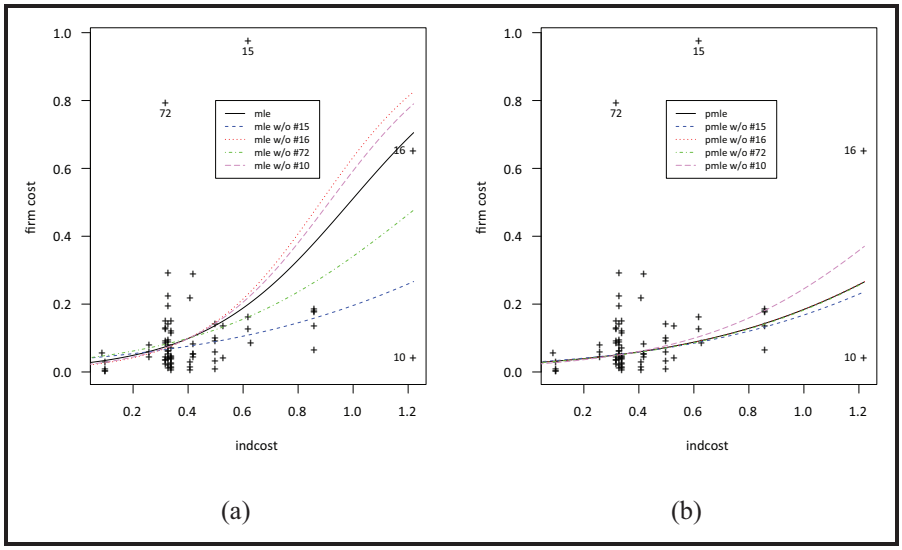

Figure 7 Scatter plots of firm cost versus `indcost` with the fitted lines based on the beta regression model with varying precision (a) and PL-slash with constant dispersion (b) for the full data and the data without outliers - Firm cost data

largest change. In comparison, the fitted lines change much less for the PL-slash regression model (Figure 7(b)), suggesting a robust fit. In Ribeiro and Ferrari (2022) the lack of robustness of the maximum likelihood estimation in beta regression is remedied by a modification in the estimation procedure. Here, we replaced the beta distribution by the PL-slash law, a highly flexible distribution. This application suggests that the PL-slash model induced robustness in the likelihood-based estimation process.

7 Concluding remarks

The power logit distributions generalize the GJS distributions at the cost of introducing an additional parameter. The application in Section 6.1 reveals that the GJS distributions are not flexible enough to produce an adequate fit to the data. The use of the corresponding power logit distributions improves the fit considerably. An interesting feature of the power logit distributions is that their three parameters are interpretable as the median, dispersion and skewness parameters. Additionally, they tend to be naturally more flexible than two-parameter distributions such as the beta, simplex, and Kumaraswamy distributions. They may also depend on an extra parameter that index the underlying symmetric distribution. The extra parameter, if any, may be seen as a tuning constant that is chosen to promote even more flexibility to suitably fit different configurations of data, particularly with outlying observations.

The article introduces regression models in which the response variable is assumed to follow a distribution in the power logit class. The real data applications in Section 6 suggest that the proposed models are useful for modeling continuous bounded data. We emphasize that the article offers a broad set of tools for likelihood inference and diagnostics. Three types of residuals are defined. We

draw attention to the standardized residual, which can more clearly highlight leverage observations than the deviance and quantile residuals. Applications in simulated data in Section 4 and the Supplementary Material (Section 7) show that the proposed residuals can identify different deviations from the postulated model and detect the presence of atypical observations. Goodness-of-fit measures, such as AIC and Υ, coupled with residual plots with simulated envelopes may be employed to select the density generator function $r(\cdot)$.

The new `R` package `PLreg` enables practitioners to employ the power logit regression models in their own data analyses. The package is introduced in this article and provides a comprehensive range of facilities for fitting both PL and GJS regression models and performing diagnostic analysis. Moreover, `PLreg` provides procedures for choosing the extra parameter, when needed.

We conclude this article by outlining some interesting directions for further research. The power logit regression models may be extended to accommodate situations in which the data include observations at the boundaries. The authors have been working on zero-and/or-one inflated power logit regression models and the findings will be reported elsewhere. Different regression structures, such as nonlinear, spatial, time series, mixed components, and random forest regressions, also deserve investigation and computing implementation. Finally, non-nested hypothesis tests (Vuong, 1989) may be developed to perform model selection within the PL regression models.

Acknowledgement

We thank the associate editor and the reviewers for their constructive comments on an earlier version of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001 and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq). The second author gratefully acknowledges funding provided by CNPq (Grant No. 305963-2018-0).

Supplementary Material

Supplementary files are available online.

References

Albert A and Anderson J (1984) On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, **71**, 1–10.

Azzalini A and Arellano-Valle RB (2013) Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, **143**, 419–433.

Bayes CL, Bazán JL and García C (2012) A new robust regression model for proportions. *Bayesian Analysis*, **7**, 841–866.

Bryson M and Johnson M (1981) The incidence of monotone likelihood in the Cox model. *Technometrics*, **23**, 381–383.

Carrasco JM, Ferrari SLP and Arellano-Valle RB (2014) Errors-in-variables beta regression models. *Journal of Applied Statistics*, **41**, 1530–1547.

Cook RD (1986) Assessment of local influence. *Journal of the Royal Statistical Society B*, **48**, 133–169.

Cox DR and Snell EJ (1968) A general definition of residuals. *Journal of the Royal Statistical Society B*, **30**, 248–265.

Cribari–Neto F and Zeileis A (2010) Beta regression in R. *Journal of Statistical Software*, **34**, 1–24.

Cysneiros FJA and Vanegas LH (2008) Residuals and their statistical properties in symmetrical nonlinear models. *Statistics and Probability Letters*, **78**, 3269–3273.

da Paz RF, Balakrishnan N and Bazán JL (2019) L-logistic regression models: Prior sensitivity analysis, robustness to outliers and applications. *Brazilian Journal of Probability and Statistics*, **33**, 455–479.

Di Brisco AM and Migliorati S (2020) A new mixed-effects mixture model for constrained longitudinal data. *Statistics in Medicine*, **39**, 129–145.

Dunn PK and Smyth GK (1996) Randomised quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236–244.

Fang KT and Anderson TW (1990) *Statistical Inference in Elliptical Contoured and Related Distributions*. New York: Allerton Press.

Ferrari SLP and Cribari–Neto F (2004) Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31**, 799–815.

Galea M, Paula GA and Cysneiros FJA (2005) On diagnostics in symmetrical nonlinear models. *Statistics and Probability Letters*, **73**, 459–467.

Gómez-Déniz E, Sordo MA and Calderín-Ojeda E (2014) The log-Lindley distribution as an alternative to the beta regression model with applications in insurance. *Insurance: Mathematics and Economics*, **54**, 49–57.

Jeffreys H (1946) An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **186**, 453–461.

Johnson NL (1949) Systems of frequency curves generated by the methods of translation. *Biometrika*, **36**, 149–176.

Jones MC and Pewsey A (2009) Sinh-arcsinh distributions. *Biometrika*, **96**, 761–780.

Korkmaz M (2020) A new heavy-tailed distribution defined on the bounded interval: the logit slash distribution and its application. *Journal of Applied Statistics*, **47**, 2097–2119.

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

Lesaffre E and Verbeke G (1998) Local influence in linear mixed model. *Biometrics*, **54**, 570–583.

Lima VM and Cribari-Neto F (2019) Penalized maximum likelihood estimation in the modified extended Weibull distribution. *Communications in Statistics–Simulation and Computation*, **48**, 334–349.

McCullagh P and Nelder J (1989) *Generalized Linear Models*. 2nd edition. London: Chapman & Hall.

Migliorati S, Di Brisco AM and Ongaro A (2018) A new regression model for bounded responses. *Bayesian Analysis*, **13**, 845–872.

Ospina R and Ferrari SLP (2012) A general class of zero-or-one inflated beta regression models.

Computational Statistics and Data Analysis, **56**, 1609–1623.

Press WH, Teukolosky SA, Vetterling WT and Flannery BP (1992) *Numerical Recipes in C: The Art of Scientific Computing*. 2nd edition. Cambridge: Cambridge University Press.

Pumi G, Prass TS and Souza RR (2021) A dynamic model for double-bounded time series with chaotic-driven conditional averages. *Scandinavian Journal of Statistics*, **48**, 68–86.

Queiroz FF and Lemonte AJ (2021) A broad class of zero-or-one inflated regression models for rates and proportions. *Canadian Journal of Statistics*, **49**, 566–590.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Austria: Vienna.

Ribeiro TK and Ferrari SLP (2022) Robust estimation in beta regression via maximum Lq-likelihood. *Statistical Papers*. doi: 10.1007/s00362-022-01320-0.

Rigby RA, Stasinopoulos DM, Heller GZ and De Bastiani F (2020) *Distributions for Modeling Location, Scale, and Shape Using GAMLSS in R*. Boca Raton: CRC Press.

Rocha AV and Cribari-Neto F (2009) Beta autoregressive moving average models. *Test*, **18**, 529–545.

Sartori N (2006) Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *Journal of Statistical Planning and Inference*, **136**, 4259–4275.

Schmid M, Wickler F, Maloney KO, Mitchell R, Fenske N and Mayr A (2013) Boosted beta regression. *Plos One*, **8**, 1–15.

Schmit JT and Roth K (1990) Cost effectiveness of risk management practices. *Journal of Risk and Insurance*, **57**, 455–470.

Scrucca L (2019) A transformation-based approach to Gaussian mixture density estimation for bounded data. *Biometrical Journal*, **61**, 873–888.

Smithson M and Shou Y (2017) CDF-quantile distributions for modelling random variables on the unit interval. *British Journal of Mathematical and Statistical Psychology*, **70**, 412–438.

Smithson M and Verkuilen J (2006) A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, **11**, 54–71.

Thomas W and Cook RD (1989) Assessing influence on regression coefficients in generalized linear models. *Biometrika*, **76**, 741–749.

Townsend J and Colonius H (2005) Variability of the max and min statistic: a theory of the quantile spread as a function of sample size. *Psychometrika*, **70**, 759–772.

Vanegas LH and Paula GA (2015) A semiparametric approach for joint modeling of median and skewness. *Test*, **24**, 110–135.

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**, 307–333. doi:10.2307/1912557.

Wang XF, Hu B, Wang B and Fang K (1998) Bayesian generalized varying coefficient models for longitudinal proportional data with errors-in-covariates. *Journal of Applied Statistics*, **41**, 1342–1357.

Wei BC, Hu YQ and Fung WK (1998) Generalized leverage and its applications. *Scandinavian Journal of Statistics*, **25**, 25–37.

Weinhold L, Schmid M, Mitchell R, Maloney KO, Wright MN and Berger M (2020) A random forest approach for bounded outcome variables. *Journal of Computational and Graphical Statistics*, **29**, 639–658.