

Robust beta regression through the logit transformation

Yuri S. Maluf, Silvia L. P. Ferrari*, Francisco F. Queiroz

Department of Statistics, University of São Paulo, Brazil

Abstract

Beta regression models are employed to model continuous response variables in the unit interval, like rates, percentages, or proportions. Their applications rise in several areas, such as medicine, environment research, finance, and natural sciences. The maximum likelihood estimation is widely used to make inferences for the parameters. Nonetheless, it is well-known that the maximum likelihood-based inference suffers from the lack of robustness in the presence of outliers. Such a case can bring severe bias and misleading conclusions. Recently, robust estimators for beta regression models were presented in the literature. However, these estimators require non-trivial restrictions in the parameter space, which limit their application. This paper develops new robust estimators that overcome this drawback. Their asymptotic and robustness properties are studied, and robust Wald-type tests are introduced. Simulation results evidence the merits of the new robust estimators. Inference and diagnostics using the new estimators are illustrated in an application to health insurance coverage data.

Keywords. Beta regression, L_q -likelihood, Outliers, Proportional data, Robust estimators, Robust inference.

1 Introduction

Beta regression is a flexible and popular tool for modeling proportions, rates, and other continuous response variables restricted to the open unit interval. Beta regression models are employed to model the relationship between predictors and a continuous response variable that is assumed to follow a beta distribution. The beta regression model with constant precision was introduced by Ferrari and Cribari-Neto (2004) and was extended in various directions. For instance, in Smithson and Verkuilen (2006) and Simas et al (2010), both the mean and precision parameters are modeled using predictors. There are numerous applications involving beta regression in different areas such as medicine (Guolo and Varin, 2014; Swearingen et al, 2011), environment research (Silva et al, 2015), finance (Cook et al, 2008), and natural sciences (Geissinger et al, 2022).

The probability density function of the beta distribution in the mean-precision parame-

*Corresponding author: e-mail silviaferrari@usp.br.

terization is

$$f(y; \mu, \phi) = \frac{1}{B(\mu\phi, (1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad (1)$$

where $0 < \mu < 1$, $\phi > 0$, and $B(\cdot, \cdot)$ is the beta function, and we write $y \sim \text{Beta}(\mu, \phi)$. We have $\mathbb{E}(y) = \mu$ and $\text{Var}(y) = \mu(1-\mu)/(1+\phi)$, hence μ is the mean parameter and ϕ can be interpreted as a precision parameter. If $\mu\phi > 1$ and $(1-\mu)\phi > 1$, the beta density (1) is bounded, has a single mode in $(0, 1)$, and decreases to zero as $y \downarrow 0$ or $y \uparrow 1$. When $\mu\phi < 1$ or $(1-\mu)\phi < 1$, the beta density is unbounded at one or both boundaries.

The beta regression model considered here is defined as follows. Let y_1, \dots, y_n be independent random variables such that $y_i \sim \text{Beta}(\mu_i, \phi_i)$, for $i = 1, \dots, n$, with

$$\begin{aligned} g_\mu(\mu_i) &= \mathbf{X}_i^\top \boldsymbol{\beta}, & (\text{mean submodel}) \\ g_\phi(\phi_i) &= \mathbf{Z}_i^\top \boldsymbol{\gamma}, & (\text{precision submodel}) \end{aligned} \quad (2)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p_1})^\top$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{p_2})^\top$ are vectors of unknown regression coefficients ($p = p_1 + p_2 < n$); $\mathbf{X}_i = (x_{i1}, \dots, x_{ip_1})^\top$ and $\mathbf{Z}_i = (z_{i1}, \dots, z_{ip_2})^\top$ are vectors of the covariates, and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top \in \mathbb{R}^p$ is the unknown parameter vector. The link functions $g_\mu : (0, 1) \rightarrow \mathbb{R}$ and $g_\phi : (0, \infty) \rightarrow \mathbb{R}$ are strictly increasing and, at least, twice differentiable.

The maximum likelihood approach is usually employed for estimating $\boldsymbol{\theta}$. However, the maximum likelihood estimator (MLE) is highly sensitive to outliers. Recently, Ghosh (2019) and Ribeiro and Ferrari (2022) proposed robust estimators for the beta regression model (1)-(2). These estimators require suitable restrictions in the parameter space. If all the beta densities in model (1)-(2) are bounded, the robust estimators and their respective asymptotic covariance matrices are well-defined. Note that the boundedness of all the beta densities assumption requires implicit, non-trivial restrictions in the parameter space of the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Moreover, these restrictions depend on the covariate vectors \mathbf{X}_i and \mathbf{Z}_i for all $i = 1, \dots, n$. If the restrictions are not satisfied, relevant numerical problems may arise when employing the robust estimators in empirical applications.

This paper introduces two new robust estimators for the beta regression model (1)-(2). They are derived using methods similar to those employed by Ghosh (2019) and Ribeiro and Ferrari (2022), with the advantage of not requiring restrictions in the parameter space.

The remaining of this paper is organized as follows. Section 2 briefly describe the estimators developed by Ghosh (2019) and Ribeiro and Ferrari (2022). Section 3 presents two new robust estimators that overcome the limitations of the current robust estimators. Robustness and asymptotic properties of the new estimators are also presented in Section 3. Section 4 shows simulation results that evidence the merits of the new estimators over the MLE and the current robust estimators. An application of the proposed methods is discussed in Section 5. The paper closes with some remarks and directions for future works.

2 Current robust estimators

Ghosh (2019) proposed the minimum density power divergence estimator (MDPDE) for the beta regression model, a robust estimator based on the density power divergence that

involves a tuning constant, $\alpha \geq 0$ (Basu et al, 1998; Ghosh and Basu, 2013). It solves the estimating equation

$$\sum_{i=1}^n [\mathbf{U}(y_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)^\alpha - \mathcal{E}_{i,1-\alpha}(\boldsymbol{\theta})] = \mathbf{0},$$

where $f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)$ denotes the beta density (1) with μ_i and ϕ_i given by (2), $\mathbf{U}(y_i; \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \log(f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i))$, and $\mathcal{E}_{i,1-\alpha}(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{U}(y_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)^\alpha]$. The factor $f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)^\alpha$ acts as the weight of the i -th observation in the estimation procedure. If $\alpha = 0$, we have the maximum likelihood estimator. Choices of $\alpha \in (0, 1)$ leads to a robust procedure because observations that are inconsistent with the postulated model receive smaller weights. If $\alpha \geq 1$, the estimator is highly robust but severely inefficient. Hence, from now on, we will restrict $\alpha \in [0, 1)$. The role of $\mathcal{E}_{i,1-\alpha}(\boldsymbol{\theta})$ is to center the weighted score, ensuring Fisher-consistency. However, $\mathcal{E}_{i,1-\alpha}(\boldsymbol{\theta})$ is not well-defined unless $\mu_i \phi_i > \alpha/(1 + \alpha)$ and $(1 - \mu_i) \phi_i > \alpha/(1 + \alpha)$. Moreover, the asymptotic covariance matrix of the MDPDE is not well-defined unless $\mu_i \phi_i > 2\alpha/(1 + 2\alpha)$ and $(1 - \mu_i) \phi_i > 2\alpha/(1 + 2\alpha)$ (see Ribeiro and Ferrari (2022) for details).

Ribeiro and Ferrari (2022) proposed an estimator based on the maximization of a reparameterized L_q -likelihood. The L_q -likelihood (Ferrari and Yang, 2010) is

$$\ell_q(\boldsymbol{\theta}) = \sum_{i=1}^n L_q(f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)), \quad (3)$$

where $q = 1 - \alpha \in (0, 1]$ is the tuning constant and $L_q(u) = (u^{1-q} - 1)/(1 - q)$, for $q \in (0, 1)$, and $L_q(u) = \log(u)$, for $q = 1$. The estimator that comes from the maximization of (3) solves the estimating equation

$$\sum_{i=1}^n \mathbf{U}(y_i; \boldsymbol{\theta}) f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)^\alpha = \mathbf{0}.$$

Note that the estimating function is not unbiased unless $\alpha = 0$, hence the resulting estimator is not Fisher-consistent. In Ghosh (2019), the Fisher-consistency is achieved by centering the weighted score. Ribeiro and Ferrari (2022) obtained a Fisher-consistent estimator through a reparametrization of the L_q -likelihood, named surrogate maximum likelihood estimator (SMLE). The estimating equation is given in Ribeiro and Ferrari (2022, eq. (10)); it is not well-defined unless $\mu_i \phi_i > \alpha$ and $(1 - \mu_i) \phi_i > \alpha$. Also, the validity of its asymptotic covariance matrix requires that $\mu_i \phi_i > 2\alpha/(1 + \alpha)$ and $(1 - \mu_i) \phi_i > 2\alpha/(1 + \alpha)$.

A sufficient condition for the MDPDE and the SMLE and their respective asymptotic covariance matrices to be well-defined is that all the beta densities in model (1)-(2) are bounded. Under such assumption, the MDPDE and the SMLE have good properties such as B-robustness, V-robustness, and asymptotic normality. A crucial issue for the use of the proposed estimators is the choice of the tuning constant. Higher values of α increase robustness and decrease efficiency. Ribeiro and Ferrari (2022) developed an effective data-driven algorithm for selecting the optimal α . Simulation results and real data applications in Ribeiro and Ferrari (2022) evidence the superior performance of these estimators relative to the MLE for datasets containing outlier observations.

The findings in Ghosh (2019) and in Ribeiro and Ferrari (2022) are guaranteed for bounded beta densities but not necessarily otherwise. As we will show later, simulations for unbounded

beta densities reveal serious numerical problems of the MDPDE and the SMLE. In the next section we propose alternative robust estimators which have the advantage of being well-defined for all beta densities.

3 Robust estimators through the logit transformation

The limitation of the MDPDE and the SMLE discussed in the previous section comes from the fact that the beta densities are not closed under power transformations. Given a density v and a constant $\xi \in (0, \infty)$, the power transformations is

$$v^{(\xi)}(y) = \frac{v(y)^\xi}{\int v(y)^\xi dy} \propto v(y)^\xi, \quad \forall y \text{ in the support,}$$

provided that $\int v(y)^\xi dy < \infty$. For the beta density (1),

$$f(y; \mu, \phi)^\xi \propto y^{\xi(\mu\phi-1)}(1-y)^{\xi[(1-\mu)\phi-1]},$$

which is integrable for all $\xi \in (0, \infty)$ if and only if $\mu\phi \geq 1$ and $(1-\mu)\phi \geq 1$. Hence, the class of the bounded beta densities is closed under power transformations, unlike the complete class of the beta densities.

To overcome this problem, consider the logit transformation $y^* = \log[y/(1-y)]$. If $y \sim \text{Beta}(\mu, \phi)$, the density function of y^* is given by

$$h(y^*; \mu, \phi) = \frac{1}{B(\mu\phi, (1-\mu)\phi)} \frac{e^{-y^*(1-\mu)\phi}}{(1+e^{-y^*})^\phi}, \quad y^* \in \mathbb{R}.$$

The distribution of y^* is called exponential generalized beta of the second type (Kerman and McDonald, 2015) and we write $y^* \sim \text{EGB}(\mu, \phi)$. Note that $h(y^*; \mu, \phi)^\xi \propto h(y^*; \mu, \xi\phi)$, for all $y^* \in \mathbb{R}$, $\mu \in (0, 1)$, and $\phi, \xi > 0$. That is, the class of the EGB densities is closed under power transformations.

We will construct robust estimators for the parameters of the beta regression model (1)-(2) using the density function of the logit transformed response variable. These estimators are based on Ghosh (2019) and Ribeiro and Ferrari (2022) methods and will be described in the following.

Let $y_i^* = \log[y_i/(1-y_i)]$, where y_i , for $i = 1, \dots, n$, follow the postulated beta regression model (1)-(2). We denote the density function of y_i^* by $h_{\boldsymbol{\theta}}(\cdot; \mu_i, \phi_i)$. The first proposed robust estimator, named logit minimum density power divergence estimator (LMDPDE), minimizes the empirical version of the density power divergence given by

$$\mathcal{H}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathcal{V}_i(y_i^*; \boldsymbol{\theta}),$$

where

$$\mathcal{V}_i(y_i^*; \boldsymbol{\theta}) = \mathcal{K}_{i,1+\alpha}(\boldsymbol{\theta}) - \frac{1+\alpha}{\alpha} h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)^\alpha,$$

and

$$\mathcal{K}_{i,1+\alpha}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)^{1+\alpha} dy_i^* = \frac{B(\mu_i \phi_i (1+\alpha), (1-\mu_i) \phi_i (1+\alpha))}{B(\mu_i \phi_i, (1-\mu_i) \phi_i)^{1+\alpha}},$$

for $0 \leq \alpha < 1$. Note that the integral is finite for all $0 \leq \alpha < 1$. The estimating equation $\nabla_{\boldsymbol{\theta}} \mathcal{H}_n(\boldsymbol{\theta}) = \mathbf{0}$ is given by

$$\sum_{i=1}^n [\mathbf{U}(y_i; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)^{\alpha} - E_{i,1-\alpha}(\boldsymbol{\theta})] = \mathbf{0}, \quad (4)$$

in which

$$\begin{aligned} \mathbf{U}(y_i; \boldsymbol{\theta}) &= \left(\phi_i \frac{(y_i^* - \mu_i^*)}{g'_{\mu}(\mu_i)} \mathbf{X}_i^\top, \quad \frac{\mu_i(y_i^* - \mu_i^*) + (y_i^{\dagger} - \mu_i^{\dagger})}{g'_{\phi}(\phi_i)} \mathbf{Z}_i^\top \right)^\top, \\ E_{i,1-\alpha}(\boldsymbol{\theta}) &= \mathbb{E} [\mathbf{U}(y_i; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)^{\alpha}] = \left(\gamma_{1,i}^{(1+\alpha)} \mathbf{X}_i^\top, \gamma_{2,i}^{(1+\alpha)} \mathbf{Z}_i^\top \right)^\top, \end{aligned}$$

where $y_i^{\dagger} = \log(1 - y_i)$,

$$\gamma_{1,i}^{(\alpha)} = \frac{\phi_i \mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{g'_{\mu}(\mu_i)} (\mu_{i,\alpha}^* - \mu_i^*), \quad \gamma_{2,i}^{(\alpha)} = \frac{\mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{g'_{\phi}(\phi_i)} [\mu_i(\mu_{i,\alpha}^* - \mu_i^*) + (\mu_{i,\alpha}^{\dagger} - \mu_i^{\dagger})],$$

with $\mu_i^* = \mathbb{E}(y_i^*) = \psi(\mu_i \phi_i) - \psi((1 - \mu_i) \phi_i)$, $\mu_i^{\dagger} = \mathbb{E}(y_i^{\dagger}) = \psi((1 - \mu_i) \phi_i) - \psi(\phi_i)$, $\mu_{i,\alpha}^* = \psi(\mu_i \phi_{i,\alpha}) - \psi((1 - \mu_i) \phi_{i,\alpha})$, $\mu_{i,\alpha}^{\dagger} = \psi((1 - \mu_i) \phi_{i,\alpha}) - \psi(\phi_{i,\alpha})$, $\psi(\cdot)$ denoting the digamma function, and $\phi_{i,\alpha} = \phi_i \alpha$.

The weight of the i -th observation in the estimating equation (4) of the LMDPDE is $h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)^{\alpha}$. In contrast, the corresponding weight for the MDPDE is $f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)^{\alpha}$. If $\alpha = 0$, $h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)^{\alpha} = 1$ and $E_{i,1-\alpha}(\boldsymbol{\theta}) = 0$, for all $i = 1, \dots, n$; hence the LMDPDE coincides with the MLE. Unlike the estimating function of the MDPDE, that of the LMDPDE is well-defined for all $\alpha \in [0, 1)$ and $\boldsymbol{\theta} \in \mathbb{R}^p$.

The second robust estimator is based on Ribeiro and Ferrari (2022) method, and is named logit surrogate maximum likelihood estimator (LSMLE). The L_q -likelihood based on the density $h_{\boldsymbol{\theta}}(\cdot; \mu_i, \phi_i)$ is given by (3) with $f_{\boldsymbol{\theta}}(y_i; \mu_i, \phi_i)$ replaced by $h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_i)$. As expected, the estimator that comes from the maximization of the L_q -likelihood is not Fisher-consistent. In other words, the estimating function is biased. Since the class of the EGB densities is closed under power transformations, Fisher-consistency can be achieved by maximizing the L_q -likelihood in the parametrization $\tau_{1/(1-\alpha)}(\boldsymbol{\theta})$ (Ferrari and La Vecchia, 2012; La Vecchia et al, 2015), where $\tau_{\omega}(\boldsymbol{\theta}) : \Theta \mapsto \Theta$ is a continuous function satisfying $h_{\tau_{\omega}(\boldsymbol{\theta})}(y^*; \mu, \phi) = h_{\boldsymbol{\theta}}^{(\omega)}(y^*; \mu, \phi)$, for all $y^* \in \mathbb{R}$. The LSMLE is the maximizer of

$$\ell_{1-\alpha}^*(\boldsymbol{\theta}) = \sum_{i=1}^n L_{1-\alpha} \left(h_{\boldsymbol{\theta}}^{(\frac{1}{1-\alpha})}(y_i^*; \mu_i, \phi_i) \right),$$

where $h_{\boldsymbol{\theta}}^{(\frac{1}{1-\alpha})}(y_i^*; \mu_i, \phi_i) = h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_{i,(1-\alpha)^{-1}})$, with $0 \leq \alpha < 1$, and μ_i and ϕ_i satisfying (2). Note that $h_{\boldsymbol{\theta}}(y_i^*; \mu_i, \phi_{i,(1-\alpha)^{-1}})$ is the density function of the logit transformation of a

variable that follows a modified beta regression model with mean and precision submodels given respectively by

$$g_\mu^*(\mu_i) = g_\mu(\mu_i) = \mathbf{X}_i^\top \boldsymbol{\beta}, \quad g_\phi^*(\phi_i) = g_\phi(\phi_{i,1-\alpha}) = \mathbf{Z}_i^\top \boldsymbol{\gamma},$$

which will be denoted by $h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)$. Thus, the LSMLE is the maximizer of

$$\ell_{1-\alpha}^*(\boldsymbol{\theta}) = \sum_{i \in \wp} L_{1-\alpha}(h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)).$$

It solves the estimating equation

$$\sum_{i=1}^n \mathbf{U}^*(y_i^*; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)^\alpha = \mathbf{0}, \quad (5)$$

where $\mathbf{U}^*(y_i^*; \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \log h_{\boldsymbol{\theta}}^*(y_i^*; \mu_i, \phi_i)$ is the modified score vector for the i -th observation given by

$$\mathbf{U}^*(y_i^*; \boldsymbol{\theta}) = \left(\phi_i \frac{(y_i^* - \mu_i^*)}{g'_\mu(\mu_i)} \mathbf{X}_i^\top, \quad (1-\alpha)^{-1} \frac{\mu_i(y_i^* - \mu_i^*) + (y_i^* - \mu_i^*)}{g'_\phi(\phi_{i,1-\alpha})} \mathbf{Z}_i^\top \right)^\top.$$

In the Supplementary Material (Section 1) we show that the LSMLE is Fisher-consistent.

Asymptotic normality. Let $\widehat{\boldsymbol{\theta}}_\alpha = (\widehat{\boldsymbol{\beta}}_\alpha^\top, \widehat{\boldsymbol{\gamma}}_\alpha^\top)^\top$ and $\widetilde{\boldsymbol{\theta}}_\alpha = (\widetilde{\boldsymbol{\beta}}_\alpha^\top, \widetilde{\boldsymbol{\gamma}}_\alpha^\top)^\top$ be the LMDPDE and the LSMLE, respectively, for fixed $\alpha \in [0, 1)$. Since they are M-estimators, we have that $\widehat{\boldsymbol{\theta}}_\alpha \xrightarrow{a} N(\boldsymbol{\theta}, \mathbf{V}_{1,\alpha}(\boldsymbol{\theta}))$ and $\widetilde{\boldsymbol{\theta}}_\alpha \xrightarrow{a} N(\boldsymbol{\theta}, \mathbf{V}_{2,\alpha}(\boldsymbol{\theta}))$, where \xrightarrow{a} denotes asymptotic distribution,

$$\mathbf{V}_{1,\alpha}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_{1,\alpha}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_{1,\alpha}(\boldsymbol{\theta}) \boldsymbol{\Lambda}_{1,\alpha}^{-1}(\boldsymbol{\theta}), \quad \mathbf{V}_{2,\alpha}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_{2,\alpha}^{-1}(\boldsymbol{\theta}) \boldsymbol{\Sigma}_{2,\alpha}(\boldsymbol{\theta}) \boldsymbol{\Lambda}_{2,\alpha}^{-1}(\boldsymbol{\theta}),$$

and the expressions for $\boldsymbol{\Lambda}_{j,\alpha}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_{j,\alpha}(\boldsymbol{\theta})$, $j = 1, 2$, are given in the Appendix; see the Supplementary Material (Section 2) for details. The covariance matrices $\mathbf{V}_{1,\alpha}(\boldsymbol{\theta})$ and $\mathbf{V}_{2,\alpha}(\boldsymbol{\theta})$ are well-defined for all $\alpha \in [0, 1)$ and $\boldsymbol{\theta} \in \mathbb{R}^p$ unlike those of the SMLE and MDPDE (Ribeiro and Ferrari, 2022). In addition, the asymptotic covariance matrices are equal to the asymptotic covariance matrix of the MLE for $\alpha = 0$.

Robustness properties. In the context of robust estimators, the influence function plays an important role. Introduced by Hampel (1974), the influence function represents the first-order measure of the effect on the asymptotic bias caused by a slight contamination in a data point. Since the LMDPDE and the LSMLE are M-estimators, their influence functions are respectively given by

$$IF(y^*; \widehat{\boldsymbol{\theta}}_\alpha) = \boldsymbol{\Lambda}_{1,\alpha}^{-1}(\boldsymbol{\theta}) [\mathbf{U}(y; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}(y^*; \mu, \phi)^\alpha - E_{1-\alpha}(\boldsymbol{\theta})],$$

$$IF(y^*; \widetilde{\boldsymbol{\theta}}_\alpha) = \boldsymbol{\Lambda}_{2,\alpha}^{-1}(\boldsymbol{\theta}) \mathbf{U}^*(y^*; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}^*(y^*; \mu, \phi)^\alpha,$$

where $E_{1-\alpha}(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{U}(y; \boldsymbol{\theta}) h_{\boldsymbol{\theta}}(y^*; \mu, \phi)^\alpha]$. The influence functions of the LMDPDE and the LSMLE are bounded, that is they are B-robust. We also extend the robustness analysis to

the change-of-variance function, which measures the bias on the covariance matrix due to an infinitesimal contamination in a data point. We show that the change-of-variance functions of the LMDPDE and the LSMLE are bounded, that is they are V-robust (Hampel et al, 2011, Section 2.5); see the Supplementary Material, Section 3, for details. These robustness properties do not hold for the MLE and are guaranteed for the MDPDE and SMLE for bounded beta densities; see Ribeiro and Ferrari (2022).

Robust Wald-type tests. Let $m : \mathbb{R}^p \rightarrow \mathbb{R}^d$, with $d \leq p$, be a continuously differentiable function of $\boldsymbol{\theta}$. Assume that its Jacobian $d \times p$ matrix, $\mathbf{J}_m(\boldsymbol{\theta})$, has rank d . Consider the null hypothesis $m(\boldsymbol{\theta}) = \boldsymbol{\eta}_0$, for a fixed $\boldsymbol{\eta}_0 \in \mathbb{R}^d$, to be tested against a two sided alternative. Let

$$W_{j,\alpha}(\boldsymbol{\theta}) = (m(\boldsymbol{\theta}) - \boldsymbol{\eta}_0)^\top [\mathbf{J}_m(\boldsymbol{\theta}) \mathbf{V}_{j,\alpha}(\boldsymbol{\theta}) \mathbf{J}_m(\boldsymbol{\theta})^\top]^{-1} (m(\boldsymbol{\theta}) - \boldsymbol{\eta}_0),$$

for $j = 1, 2$. The Wald-type test statistics that use the LMDPDE and the LSMLE are, respectively, given by $W_{1,\alpha}(\boldsymbol{\theta}_\alpha)$ and $W_{2,\alpha}(\boldsymbol{\theta}_\alpha)$. Under the null hypothesis, both statistics are asymptotically χ_d^2 -distributed. They coincide with the usual Wald test statistic if $\alpha = 0$.

Selecting the tuning constant. Ribeiro and Ferrari (2022, Section 3) proposed a data-driven algorithm to select the tuning constant for the MDPDE and the SMLE. The idea is to select α , in an ordered grid from $\alpha = 0$ to $\alpha = \alpha_{\max}$, that is closest to zero such that the estimates of the parameters are sufficiently stable, ensuring full efficiency for non-contaminated data. If the algorithm does not reach stability up to α_{\max} , it returns the MLE ($\alpha = 0$). The authors suggest setting $\alpha_{\max} = 0.5$. Here, the algorithm is implemented for selecting α for the LMDPDE and the LSMLE.

4 Simulation studies

In this section, we evaluate the performance of the robust estimators and the MLE for the beta regression model (1)-(2). We employ the logit and the logarithmic link functions for the mean and precision submodels, respectively. Both submodels include an intercept, i.e., $x_{i,1} = z_{i,1} = 1$, $i = 1, \dots, n$. The sample sizes are set at $n = 40, 80, 160$, and 320 . The covariate values for the mean submodel are set for the sample size $n = 40$ as random draws from a standard uniform distribution and replicated twice, four times and eight times for the other values of n . For the non-constant precision scenario, the covariate values for the mean submodel are used in the precision submodel. All the covariate values are kept constant over all the simulated samples. We consider non-contaminated and contaminated samples with a fixed contamination rate, namely 5%. All simulations were carried out using the R software (R Core Team, 2022), and the results are based on 1000 Monte Carlo replications.

We consider three different scenarios. Figure 1 shows scatter plots of a sample generated under each scenario for $n = 40$.

Scenario A: bounded beta densities; constant precision. The parameters are set at $\beta_1 = -1$, $\beta_2 = -2$ and $\gamma_1 = 5$. The possible values for μ range in $(0.05, 0.27)$ and $\phi = \exp(5) \cong 148$. For the contaminated samples, we replace the observations generated with the 5% smallest means by observations generated with mean $\mu'_i = (1 + \mu_i)/2$. All the beta densities in this scenario

are bounded.

Scenario B: unbounded beta densities; constant precision. The parameters are set at $\beta_1 = -1$, $\beta_2 = -5.5$ and $\gamma_1 = 5$. The possible values for μ range in $(0.001, 0.27)$ and $\phi = \exp(5) \cong 148$. For the contaminated samples, the observations generated with the 5% highest means are replaced by observations generated with mean $\mu'_i = 0.002$. Some beta densities in this scenario are unbounded.

Scenario C: unbounded beta densities; varying precision. The parameters are set at $\beta_1 = -3$, $\beta_2 = 7.5$, $\gamma_1 = 1$, and $\gamma_2 = 2$. The possible values for μ and ϕ range in $(0.05, 0.98)$ and $\phi \in (2.7, 20.1)$, respectively. For the contaminated samples, the observations generated with the 5% highest means are replaced by observations generated with $\mu'_i = \exp(\beta_1)/(1 + \exp(\beta_1)) \cong 0.05$ and $\phi'_i = \exp(\gamma_1 + \gamma_2) \cong 20.1$. Some beta densities in this scenario are unbounded.

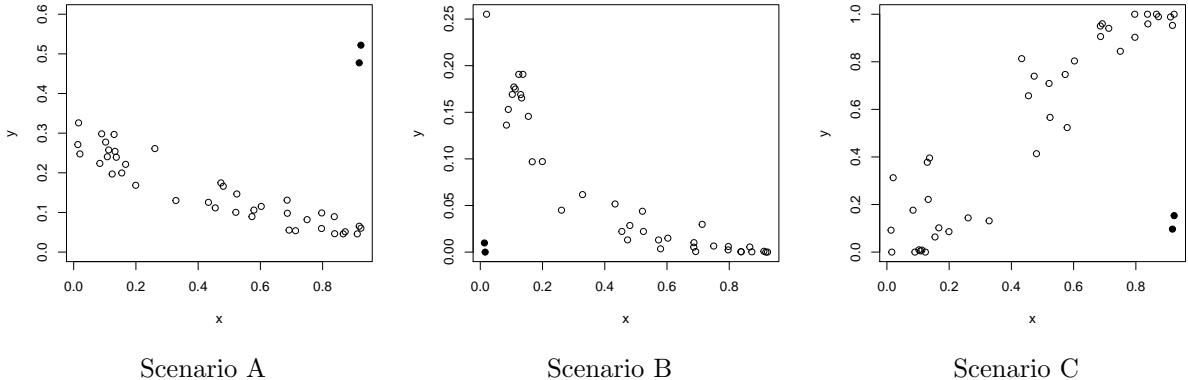


Figure 1: Scatter plots of contaminated samples generated as in Scenarios A, B, and C with $n = 40$. The contaminated observations are represented by solid black circles.

First, we run simulations for the three scenarios with $n = 40$ for fixed values of the tuning constant α ranging from 0 to 0.05 incremented by 0.05. For each α value, we compute the failure rate over the 1000 simulated samples for the robust estimators: the MDPDE, the SMLE, the LMDPDE, and the LSMLE. We consider a failure whenever the optimization algorithm for computing the estimate does not reach convergence or the asymptotic standard error can not be calculated. In Scenario A, no sample resulted in failure for any of the estimators. Recall that all the beta densities in this scenario are bounded. Figure 2 displays plots of the failure rates for Scenarios B and C. The failure rate of SMLE and the MDPDE tend to increase as α grows, more so in Scenario C. In contrast, for the new robust estimators, the failure rate is equal (or close) to zero for all values of α , both under non-contaminated or contaminated data. Simulations for the other sample sizes reveal a similar pattern.

We now report simulation results using the data-driven algorithm for selecting the optimum value of α proposed by Ribeiro and Ferrari (2022). Figures 3, 4, and 5-6 display the boxplots of the parameter estimates using the MLE, the SMLE, the MDPDE, the LSMLE, and the LMDPDE under Scenarios A, B, and C, respectively.

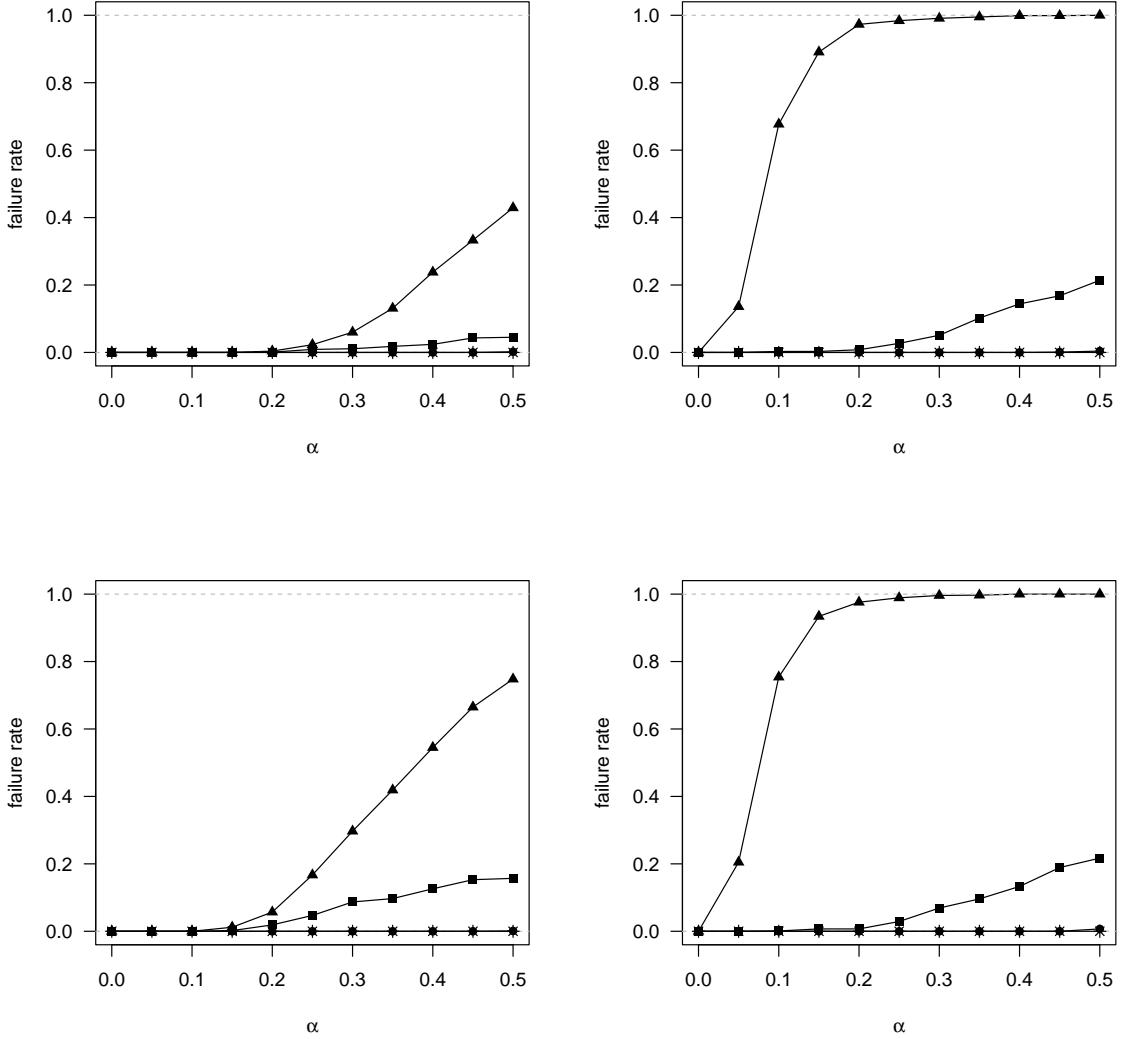


Figure 2: Plots of the failure rate versus the tuning constant α for Scenario B (left) and Scenario C (right) for non-contaminated samples (first row) and contaminated samples (second row) : SMLE (triangle), MDPDE (square), LSMLE (circle), and LMDPDE (star).

The MLE is highly affected by contaminated observations for all the scenarios and presents a severe bias. For instance, the maximum likelihood estimates of γ_1 in Scenario B are around its true value, 5, for the non-contaminated data and around 2 for the contaminated data. In Scenario A, all the robust estimators present good performances for both contaminated and non-contaminated data. They behave similarly to the MLE under non-contaminated samples. Recall that all the beta densities in this scenario are bounded. In Scenarios B and C, which include unbounded beta densities, the MDPDE and the SMLE do not behave well for the non-contaminated data. In these cases, for almost all the samples, the selected optimum tuning constants are zero, resulting in non-robust estimates. It happens because the optimization fails; hence, the algorithm does not reach stability in the estimates and returns the MLE. In contrast, the LMDPDE and the LSMLE perform well in the presence

and absence of contamination. Also, these estimators have similar behavior. For samples of moderate and large sizes ($n = 80, 160, 360$), the performances of the LMDPDE and LSMLE are excellent. The simulation results indicate that the proposed estimators (LMDPDE and LSMLE) are robust in the presence of outliers, unlike the MLE. Also, the MDPDE and SMLE may not be useful in scenarios involving unbounded beta densities. In general, in the presence of contamination, the robustness obtained through the new proposed estimators comes at the cost of a slight increase in variability, especially in small sample sizes ($n = 40$).

The data-driven algorithm to select the tuning constant α had an excellent performance for the LMDPDE and the LSMLE (see Figure 7). The selected optimum α is zero for the LMDPDE and the LSMLE for non-contaminated data, except for a few samples when $n = 40$. Recall that $\alpha = 0$ corresponds to the MLE. For the contaminated data, the selected optimum values of α are around 0.1 for Scenarios A and B and around 0.15 for Scenario C. Hence, the algorithm can identify the need to use a robust procedure. In Scenario A, the MDPDE and the SMLE behave like the LMDPDE and the LSMLE. On the other hand, for Scenarios B and C, the selected optimum α values for the MDPDE and the SMLE are close to zero for both non-contaminated and contaminated data. This happens because the algorithm does not achieve stability under contaminated data for most of the samples.

We now report the empirical levels of the Wald test (that uses the MLE) and the robust Wald-type test based on the LMDPDE and the LSMLE. The considered nominal level is 5%. The null hypotheses considered for Scenarios A and B are $H_0^1 : \beta_2 = \beta_2^0$, $H_0^2 : (\beta_1, \beta_2) = (\beta_1^0, \beta_2^0)$, and $H_0^3 : (\beta_1, \beta_2, \gamma_1) = (\beta_1^0, \beta_2^0, \gamma_1^0)$. For Scenario C, we set $H_0^4 : (\beta_2, \gamma_2) = (\beta_2^0, \gamma_2^0)$, $H_0^5 : (\beta_1, \beta_2, \gamma_2) = (\beta_1^0, \beta_2^0, \gamma_2^0)$, and $H_0^6 : \gamma_2 = \gamma_2^0$. The values of the parameters fixed at the null hypotheses are those used in the simulations above. The results are shown in Table 1. For non-contaminated data, the empirical levels of all the tests are close to the nominal levels. For contaminated data, the usual Wald test presents a type I error close to 100%, being highly unreliable. In contrast, the robust Wald-type tests show to be reasonably reliable, with only slight inflation in the type I error relatively to non-contaminated situations.

Overall, our simulations suggest that the new robust estimators proposed in this paper, namely the LSMLE and the LMDPDE, exhibited the same performance as the SMLE and the MDPDE for bounded beta densities (Scenario A). For unbounded beta densities, as in Scenarios B and C, the MDPDE and the SMLE are unreliable, presenting severe bias for contaminated data. For all the scenarios, the new estimators behave as the MLE for non-contaminated data and prove to be robust in the presence of contamination. Hence, practitioners should employ the new proposed estimators in real data applications.

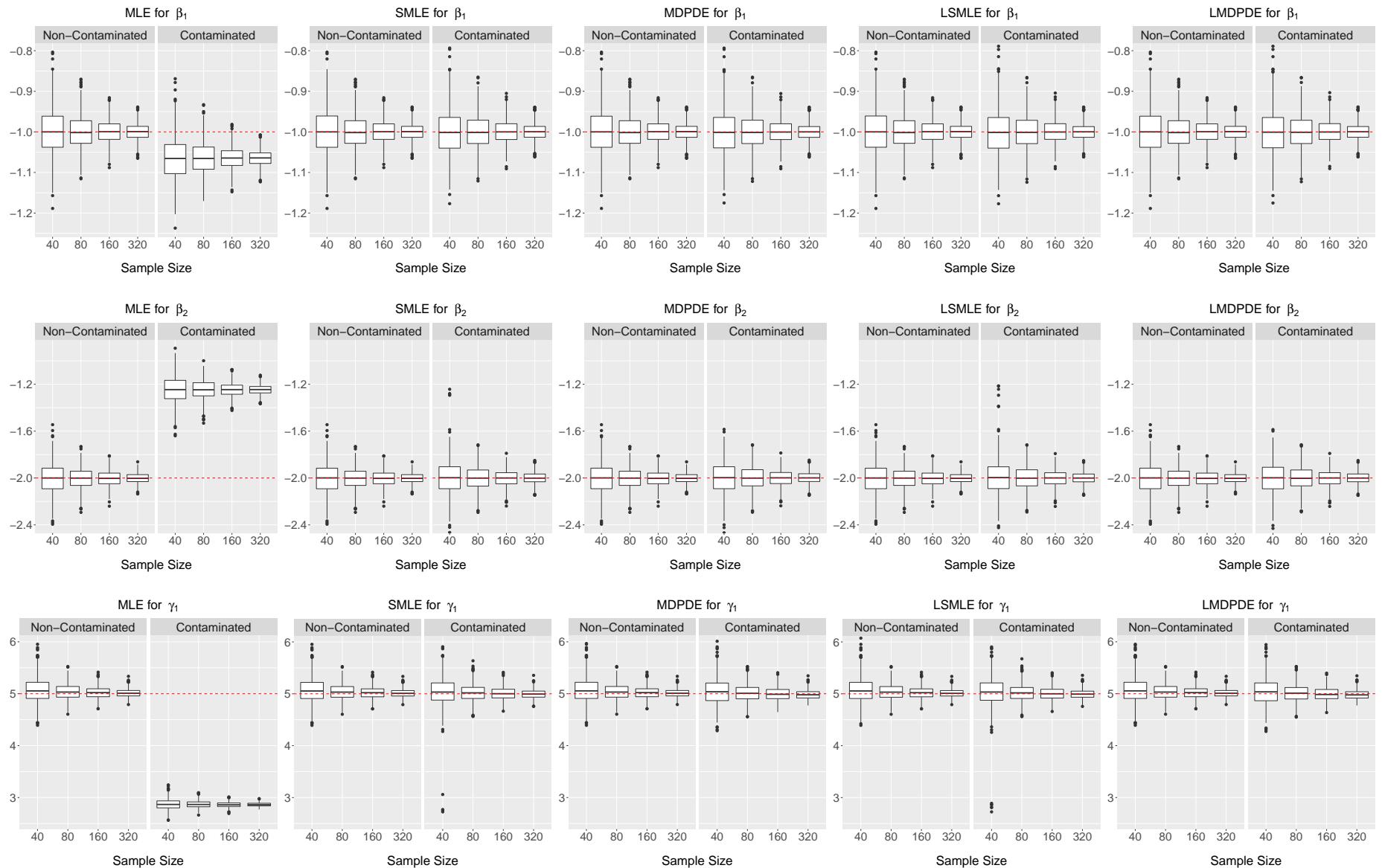


Figure 3: Boxplots of estimates of β_1 (first row), β_2 (second row), and γ_1 (third row) for the MLE and the robust estimators. The red dashed line represents the true parameter value.

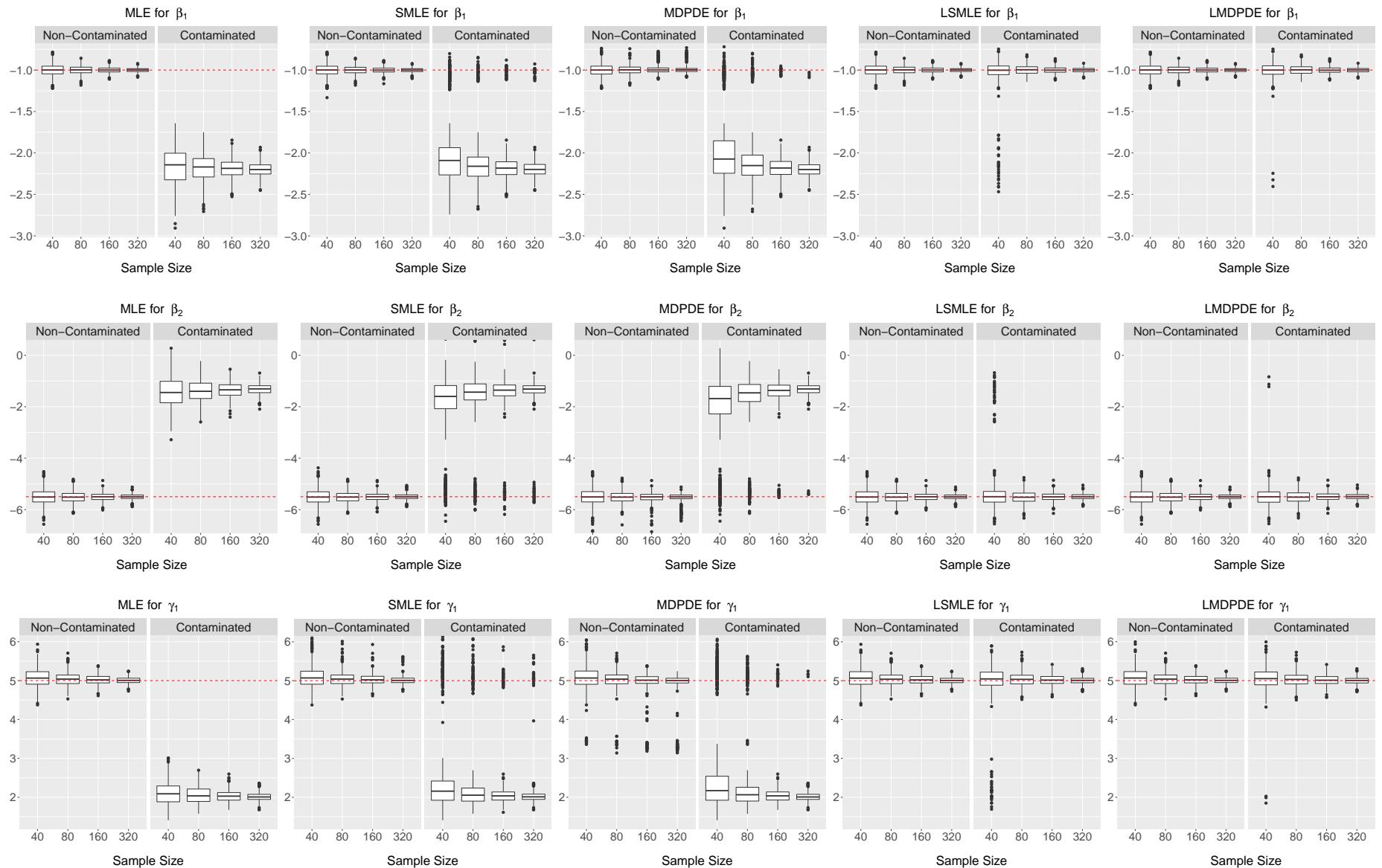


Figure 4: Boxplots of estimates of β_1 (first row), β_2 (second row), and γ_1 (third row) for the MLE and the robust estimators. The red dashed line represents the true parameter value.

C1

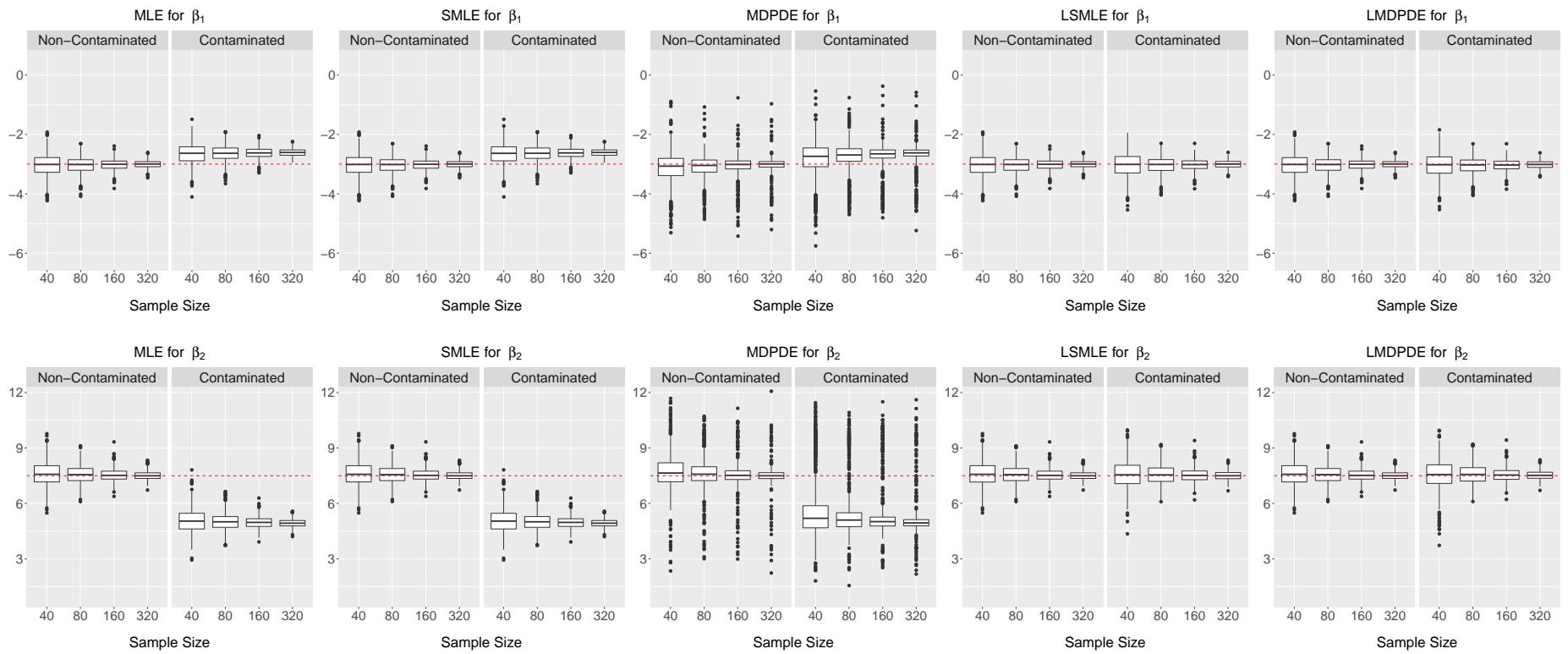


Figure 5: Boxplots of estimates of β_1 (first row) and β_2 (second row) for the MLE and the robust estimators. The red dashed line represents the true parameter value.

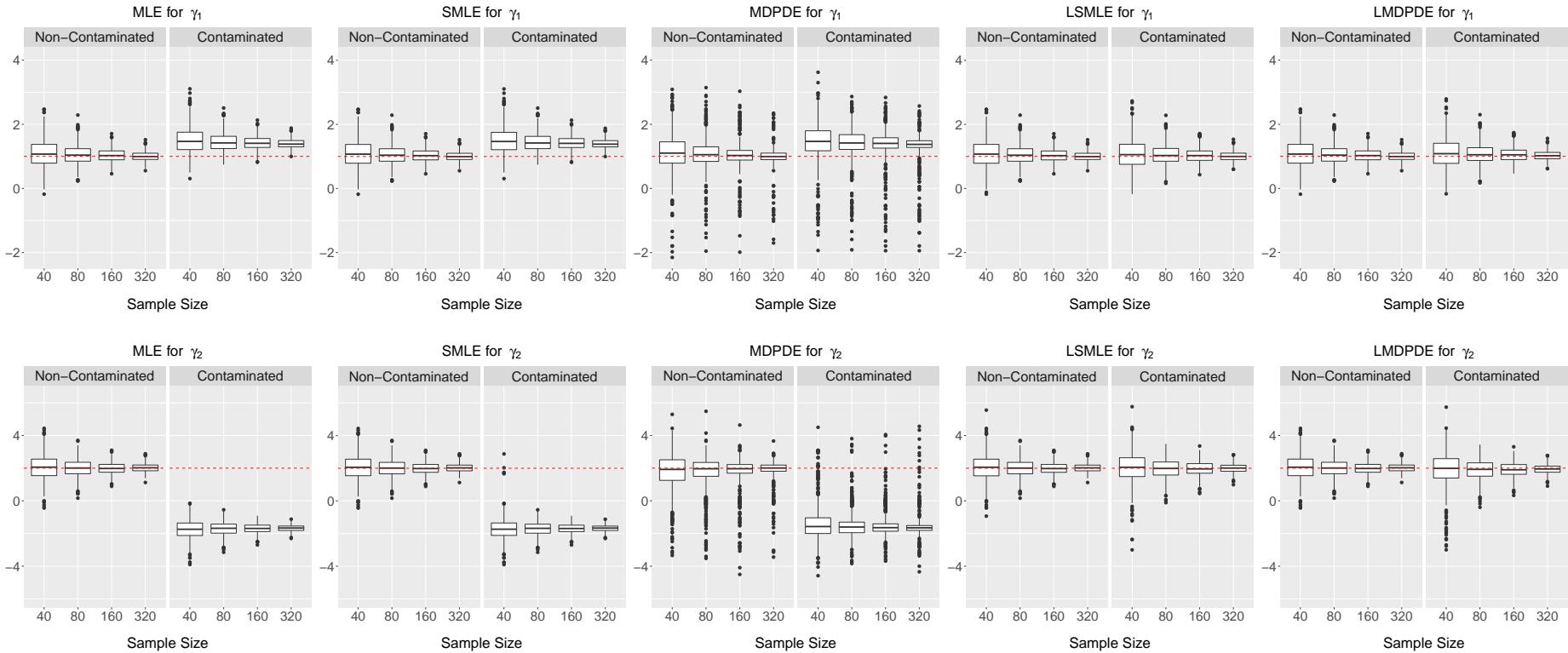


Figure 6: Boxplots of estimates of γ_1 (first row) and γ_2 (second row) for the MLE and the robust estimators. The red dashed line represents the true parameter value.

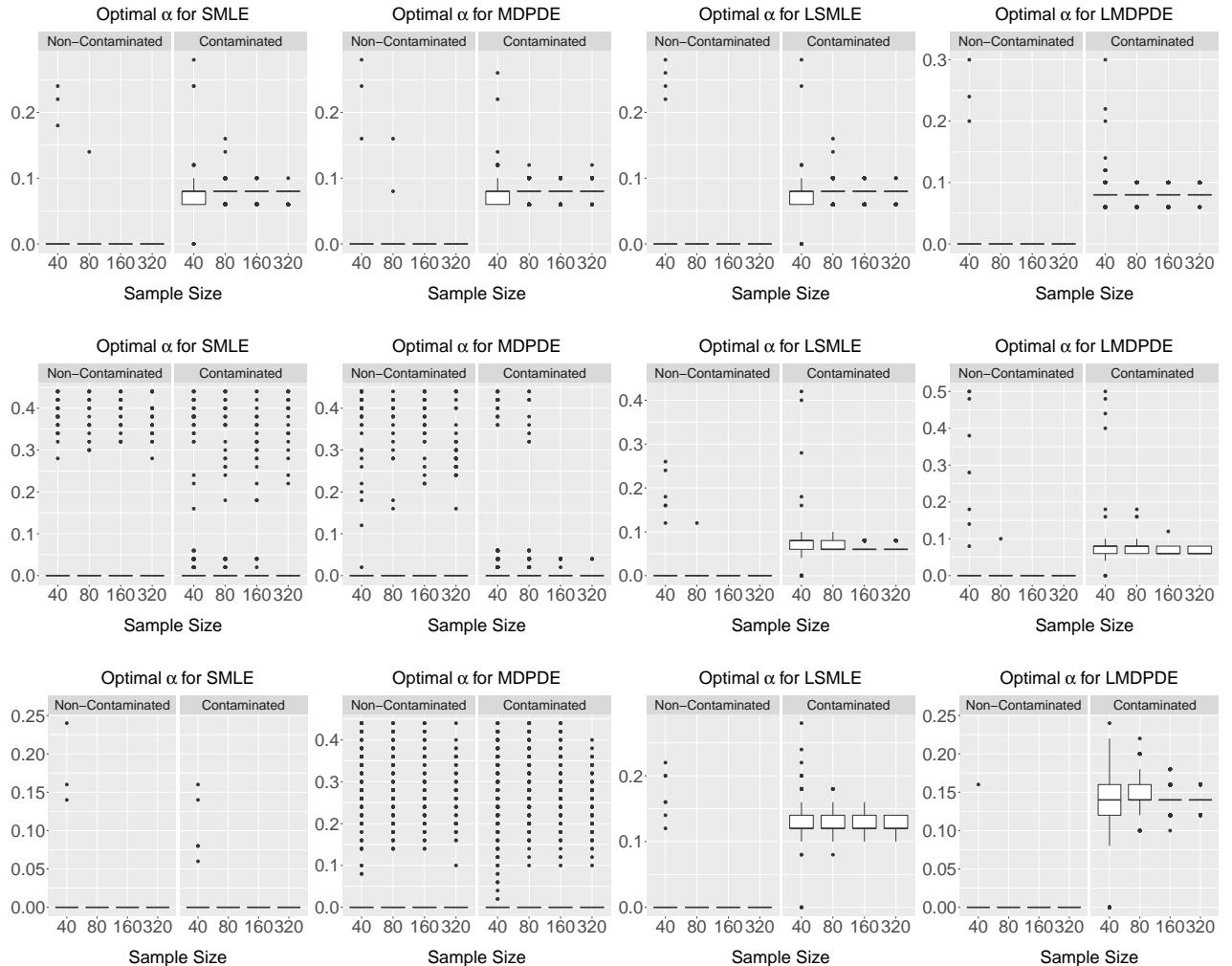


Figure 7: Boxplots of the optimal values for the tuning parameter α for the robust estimators under Scenario A (first row), B (second row), and C (third row).

5 An application to health insurance coverage data

We shall now present and discuss an application of the new robust estimators to health insurance coverage data collected by the Institute of Applied Economic Research (Instituto de Pesquisa Econômica Aplicada, IPEA). The dataset includes information on 80 cities in the state of São Paulo, Brazil, in 2010. This application's dataset and R codes are available at <https://github.com/ffqueiroz/RobustBetareg>.

The response variable (y) is the health insurance coverage index (HIC). The covariates are the percentage of the total population who lives in the city's urban zone (Urb) and the per capita gross domestic product (GDP). We consider the beta regression model (1)-(2) with $g_\mu(\cdot)$ and $g_\phi(\cdot)$ being the logit and the log functions, respectively. First, both covariates and an intercept are included in the mean and precision submodels. We fit the model using the MLE and the robust estimators. The Wald-type tests based on all the estimators agree that the covariate Urb is not significant for the precision submodel (p -value greater than 0.3 for all

Table 1: Empirical null levels of Wald-type tests under Scenarios A, B, and C at the 5% nominal level.

		Scenario A					
Estimator	<i>n</i>	Non-cont.			Cont.		
		H0 ¹	H0 ²	H0 ³	H0 ¹	H0 ²	H0 ³
MLE	40	0.05	0.06	0.07	0.88	0.98	1.00
	80	0.06	0.06	0.06	1.00	1.00	1.00
	160	0.05	0.06	0.06	1.00	1.00	1.00
	320	0.05	0.05	0.05	1.00	1.00	1.00
LMDPDE	40	0.05	0.06	0.07	0.07	0.07	0.08
	80	0.06	0.06	0.06	0.07	0.07	0.07
	160	0.05	0.06	0.06	0.06	0.06	0.07
	320	0.05	0.05	0.05	0.06	0.06	0.07
LSMLE	40	0.05	0.06	0.07	0.08	0.08	0.09
	80	0.06	0.06	0.06	0.07	0.07	0.07
	160	0.05	0.06	0.06	0.06	0.06	0.07
	320	0.05	0.05	0.05	0.06	0.06	0.06
Scenario B							
Estimator	<i>n</i>	Non-cont.			Cont.		
		H0 ¹	H0 ²	H0 ³	H0 ¹	H0 ²	H0 ³
MLE	40	0.07	0.08	0.08	1.00	1.00	1.00
	80	0.05	0.06	0.06	1.00	1.00	1.00
	160	0.05	0.06	0.07	1.00	1.00	1.00
	320	0.06	0.07	0.08	1.00	1.00	1.00
LMDPDE	40	0.07	0.07	0.08	0.07	0.10	0.10
	80	0.06	0.06	0.06	0.07	0.09	0.09
	160	0.05	0.06	0.07	0.06	0.07	0.08
	320	0.06	0.07	0.08	0.08	0.10	0.10
LSMLE	40	0.07	0.08	0.08	0.09	0.12	0.01
	80	0.06	0.06	0.06	0.08	0.09	0.01
	160	0.05	0.06	0.07	0.06	0.07	0.09
	320	0.06	0.07	0.08	0.08	0.10	0.10
Scenario C							
Estimator	<i>n</i>	Non-cont.			Cont.		
		H0 ⁴	H0 ⁵	H0 ⁶	H0 ⁴	H0 ⁵	H0 ⁶
MLE	40	0.08	0.08	0.08	1.00	1.00	1.00
	80	0.08	0.07	0.06	1.00	1.00	1.00
	160	0.06	0.06	0.05	1.00	1.00	1.00
	320	0.05	0.05	0.06	1.00	1.00	1.00
LMDPDE	40	0.08	0.08	0.08	0.13	0.13	0.12
	80	0.08	0.07	0.06	0.12	0.10	0.10
	160	0.05	0.06	0.05	0.11	0.12	0.09
	320	0.04	0.05	0.06	0.07	0.08	0.08
LSMLE	40	0.08	0.08	0.08	0.12	0.12	0.11
	80	0.08	0.07	0.06	0.11	0.10	0.09
	160	0.06	0.06	0.05	0.10	0.09	0.09
	320	0.05	0.05	0.05	0.07	0.07	0.07

the tests); see the Supplementary Material, Section 4. The postulated reduced model is the beta regression model with

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = \beta_1 + \beta_2 \text{Urb}_i + \beta_3 \text{GDP}_i,$$

$$\log(\phi_i) = \gamma_1 + \gamma_2 \text{GDP}_i,$$

for $i = 1, \dots, 80$. For some values of α , the SMLE and MDPDE could not be computed and the data-driven algorithm for selecting α did not reach stability and returned $\alpha = 0$ (MLE). The estimates and standard errors for the LSMLE and LMDPDE are similar. Here, we present the results for the LSMLE and those for the LMDPDE are shown in the Supplementary Material (Section 4). Table 2 presents the estimates, asymptotic standard errors, z -statistics (estimate

divided by the asymptotic standard error), and asymptotic p -values of the Wald-type tests of nullity of coefficients. It also reports the results without observation #1, which is the most evident outlier. This observation corresponds to a city with an atypical value for HIC, around 0.98.

Table 2: Estimates, asymptotic standard errors (Std. error), z -stat, and asymptotic p -values for the full and reduced data.

	MLE for the full data				LSMLE for the full data			
	Estimate	Std. error	z -stat	p -value	Estimate	Std. error	z -stat	p -value
<i>mean submodel</i>								
Intercept	-4.429	0.629	-7.037	0.000	-5.992	0.518	-11.562	0.000
Urb	3.429	0.706	4.857	0.000	4.884	0.581	8.410	0.000
GDP	0.010	0.003	3.535	0.000	0.013	0.003	4.652	0.000
<i>precision submodel</i>								
Intercept	2.086	0.229	9.115	0.000	3.332	0.229	14.573	0.000
GDP	-0.002	0.005	-0.456	0.648	-0.013	0.004	-2.934	0.003
MLE without observation #1								
	MLE without observation #1				LSMLE without observation #1			
	Estimate	Std. error	z -stat	p -value	Estimate	Std. error	z -stat	p -value
<i>mean submodel</i>								
Intercept	-5.978	0.508	-11.777	0.000	-5.978	0.508	-11.777	0.000
Urb	4.854	0.569	8.538	0.000	4.854	0.569	8.538	0.000
GDP	0.013	0.003	4.721	0.000	0.013	0.003	4.721	0.000
<i>precision submodel</i>								
Intercept	3.391	0.229	14.816	0.000	3.391	0.229	14.816	0.000
GDP	-0.013	0.004	-3.003	0.003	-0.013	0.004	-3.003	0.003

For the full data, the data-driven algorithm for selecting the optimum α returned $\alpha = 0.06$ for both LMDPDE and LSMLE, indicating that a robust fit is needed. For the data without observation #1, the algorithm returned $\alpha = 0$ (MLE) for both the new robust estimators. As we observe in Table 2, the MLE is highly influenced by observation #1. For instance, the estimated coefficient for Urb in the mean submodel moves from 3.429 (full data) to 4.854 (reduced data). Additionally, the covariate GDP in the precision submodel is non-significant (p -value equal to 0.648) for the full data and highly significant (p -value equal to 0.003) for the reduced data. In contrast, the results based on the LSMLE are not impacted by the exclusion of the outlier observation. The results for the LSMLE for the full data are close to those for the MLE for the reduced data.

Following Ribeiro and Ferrari (2022), Figure 8 presents the normal probability plots with simulated envelopes of residuals for the MLE and the LSMLE and the plot of estimated weights against residuals for the LSMLE. We consider the ‘standardized weighted residual 2’ proposed by Espinheira et al (2008). The residual plots of the MLE clearly evidence the lack of fit of the maximum likelihood estimation. As expected, observation #1 is highlighted as an outlier for both the MLE and the LSMLE fits. The residual plots for the LSMLE suggest a suitable fit for all the observations except for case #1. In fact, this observation receives a weight close to zero for the LSMLE fit.

6 Concluding remarks

This paper introduces two new robust estimators for the beta regression models: the LSMLE and the LMDPDE. The proposed estimators overcome the limitations of the current

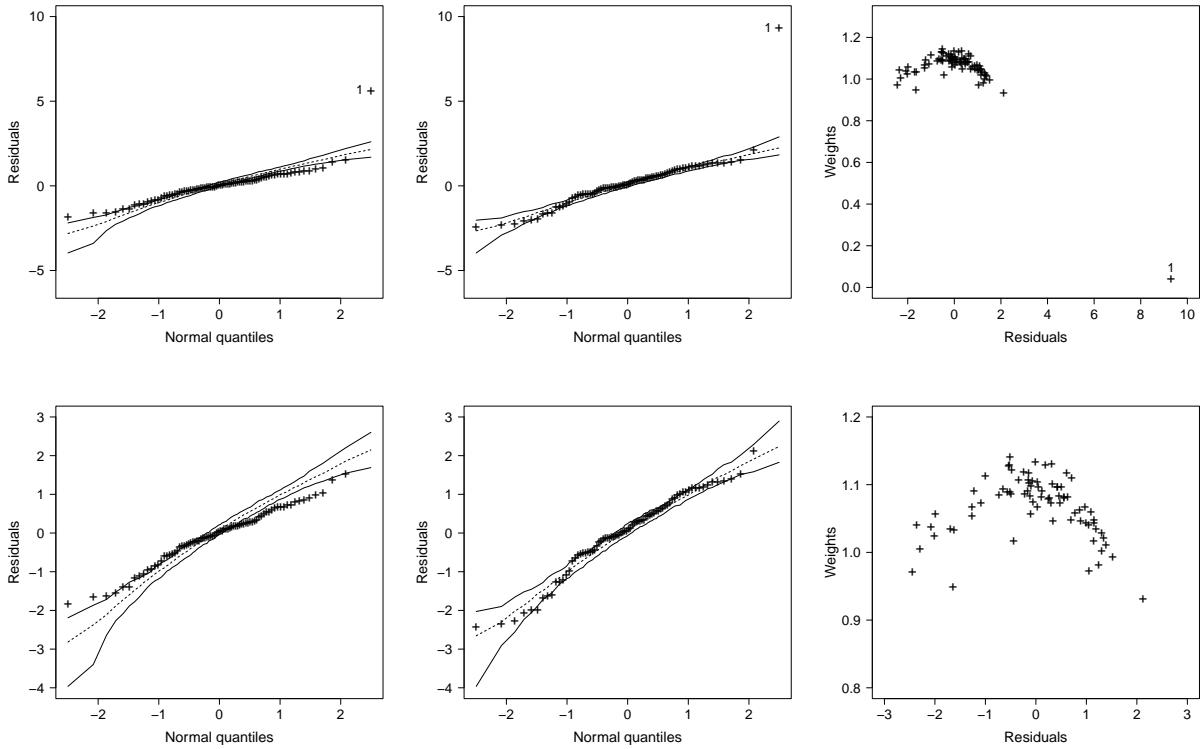


Figure 8: Normal probability plots with simulated envelope of the residuals for MLE (first column) and LSMLE (second column) and plot of estimated weights for LSMLE (third column). The plots in the second row are zoomed versions of those in the first row.

robust estimators, the SMLE and the MDPDE. Simulation results and a real data application evidence the excellent performance of the new estimators even in situations where the existing estimators fail. The new robust estimators present similar behavior and are easily implemented. Practitioners may employ our R codes for their own analyses. An R package for robust beta regression inference and diagnostics is under development and will soon be available.

The development of robust estimators for inflated beta regression models (Ospina and Ferrari, 2012) is a natural, although non-trivial, extension of our work. The second and third authors have been working on this topic. The findings will be reported elsewhere.

Acknowledgments

This work 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 the funding provided by CNPq (Grant No. 305963-2018-0). A special thank goes to Terezinha T.K.A. Ribeiro for helpful discussion and for sharing her R code on which we based our implementation of the estimators developed in this paper.

A Appendix

The matrices $\Lambda_{1,\alpha}(\boldsymbol{\theta})$ and $\Sigma_{1,\alpha}(\boldsymbol{\theta})$ used in the covariance matrix of the LMDPDE are

$$\Lambda_{1,\alpha}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{X}^\top \gamma_{11}^{(1+\alpha)} \mathbf{X} & \mathbf{X}^\top \gamma_{12}^{(1+\alpha)} \mathbf{Z} \\ \mathbf{Z}^\top \gamma_{12}^{(1+\alpha)} \mathbf{X} & \mathbf{Z}^\top \gamma_{22}^{(1+\alpha)} \mathbf{Z} \end{bmatrix}$$

and

$$\Sigma_{1,\alpha}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{X}^\top \left[\gamma_{11}^{(1+2\alpha)} - \gamma_1^{(1+\alpha)2} \right] \mathbf{X} & \mathbf{X}^\top \left[\gamma_{12}^{(1+2\alpha)} - \gamma_1^{(1+\alpha)} \gamma_2^{(1+\alpha)} \right] \mathbf{Z} \\ \mathbf{Z}^\top \left[\gamma_{12}^{(1+2\alpha)} - \gamma_1^{(1+\alpha)} \gamma_2^{(1+\alpha)} \right] \mathbf{X} & \mathbf{Z}^\top \left[\gamma_{22}^{(1+2\alpha)} - \gamma_2^{(1+\alpha)2} \right] \mathbf{Z} \end{bmatrix},$$

where $\gamma_j^{(\alpha)} = \text{diag}\{\gamma_{j,i}^{(\alpha)}; i = 1, \dots, n\}$, for $j = 1, 2$, $\gamma_{11}^{(\alpha)} = \text{diag}\{\gamma_{11,i}^{(\alpha)}; i = 1, \dots, n\}$, $\gamma_{12}^{(\alpha)} = \text{diag}\{\gamma_{12,i}^{(\alpha)}; i = 1, \dots, n\}$, $\gamma_{22}^{(\alpha)} = \text{diag}\{\gamma_{22,i}^{(\alpha)}; i = 1, \dots, n\}$, with

$$\begin{aligned} \gamma_{11,i}^{(\alpha)} &= \frac{\phi_i^2 \mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{[g'_\mu(\mu_i)]^2} [v_{i,\alpha} + (\mu_{i,\alpha}^* - \mu_i^*)^2], \\ \gamma_{12,i}^{(\alpha)} &= \frac{\phi_i \mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{g'_\mu(\mu_i) g'_\phi(\phi_i)} \left\{ \mu_i [v_{i,\alpha} + (\mu_{i,\alpha}^* - \mu_i^*)^2] - \psi'((1 - \mu_i)\phi_{i,\alpha}) + (\mu_{i,\alpha}^* - \mu_i^*)(\mu_{i,\alpha}^\dagger - \mu_i^\dagger) \right\}, \\ \gamma_{22,i}^{(\alpha)} &= \frac{\mathcal{K}_{i,\alpha}(\boldsymbol{\theta})}{[g'_\phi(\phi_i)]^2} \left\{ \mu_i^2 \psi'(\mu_i \phi_{i,\alpha}) + (1 - \mu_i)^2 \psi'((1 - \mu_i)\phi_{i,\alpha}) - \psi'(\phi_{i,\alpha}) \right. \\ &\quad \left. + [\mu_i(\mu_{i,\alpha}^* - \mu_i^*) + (\mu_{i,\alpha}^\dagger - \mu_i^\dagger)]^2 \right\}, \end{aligned}$$

and $v_{i,\alpha} = \psi'(\mu_i \phi_{i,\alpha}) + \psi'((1 - \mu_i)\phi_{i,\alpha})$.

For the LSMLE, the matrices $\Lambda_{2,\alpha}(\boldsymbol{\theta})$ and $\Sigma_{2,\alpha}(\boldsymbol{\theta})$ are given by

$$\Lambda_{2,\alpha}(\boldsymbol{\theta}) = - \begin{bmatrix} (1 - \alpha) \mathbf{X}^\top \mathbf{B}_1 \mathbf{T}_\mu^2 \Phi^2 \mathbf{V} \mathbf{X} & \mathbf{X}^\top \mathbf{B}_1 \mathbf{T}_\mu \mathbf{T}_\phi^* \mathbf{C} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{B}_1 \mathbf{T}_\mu \mathbf{T}_\phi^* \mathbf{C} \mathbf{X} & (1 - \alpha)^{-1} \mathbf{Z}^\top \mathbf{B}_1 \mathbf{T}_\phi^{*2} \mathbf{D} \mathbf{Z} \end{bmatrix}$$

and

$$\Sigma_{2,\alpha}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{X}^\top \mathbf{B}_2 \mathbf{T}_\mu^2 \Phi^2 \mathbf{V}_{1+\alpha} \mathbf{X} & (1 - \alpha)^{-1} \mathbf{X}^\top \mathbf{B}_2 \mathbf{T}_\mu \mathbf{T}_\phi^* \mathbf{C}_{1+\alpha} \mathbf{Z} \\ (1 - \alpha)^{-1} \mathbf{Z}^\top \mathbf{B}_2 \mathbf{T}_\mu \mathbf{T}_\phi^* \mathbf{C}_{1+\alpha} \mathbf{X} & (1 - \alpha)^{-2} \mathbf{Z}^\top \mathbf{B}_2 \mathbf{T}_\phi^{*2} \mathbf{D}_{1+\alpha} \mathbf{Z} \end{bmatrix},$$

where $\mathbf{B}_j = \text{diag}\{b_{i,j}; i = 1, \dots, n\}$, $j = 1, 2$,

$$\begin{aligned} b_{i,1} &= \frac{B(\mu_i \phi_i, (1 - \mu_i) \phi_i)^{1-\alpha}}{B(\mu_i \phi_{i,1-\alpha}, (1 - \mu_i) \phi_{i,1-\alpha})}, \\ b_{i,2} &= \frac{B(\mu_i \phi_{i,1+\alpha}, (1 - \mu_i) \phi_{i,1+\alpha})}{B(\mu_i \phi_i, (1 - \mu_i) \phi_i)^{2\alpha} B(\mu_i \phi_{i,1-\alpha}, (1 - \mu_i) \phi_{i,1-\alpha})}, \end{aligned}$$

$\mathbf{T}_\mu = \text{diag}\{t_{i,\mu}; i = 1, \dots, n\}$, $\mathbf{T}_\phi^* = \text{diag}\{t_{i,\phi}; i = 1, \dots, n\}$,

$$t_{i,\mu} = [g'_\mu(\mu_i)]^{-1}, \quad t_{i,\phi} = [g'_\phi(\phi_{i,1-\alpha})]^{-1},$$

$\Phi = \text{diag}\{\phi_i; i = 1, \dots, n\}$, $\mathbf{V} = \text{diag}\{v_i; i = 1, \dots, n\}$, $\mathbf{V}_{1+\alpha} = \text{diag}\{v_{i,1+\alpha}; i = 1, \dots, n\}$,
 $\mathbf{C} = \text{diag}\{c_i; i = 1, \dots, n\}$, $\mathbf{C}_{1+\alpha} = \text{diag}\{c_{i,1+\alpha}; i = 1, \dots, n\}$,

$$c_i = \phi_i [\mu_i \psi'(\mu_i \phi_i) - (1 - \mu_i) \psi'((1 - \mu_i) \phi_i)], \\ c_{i,1+\alpha} = \phi_i [\mu_i \psi'(\mu_i \phi_{i,1+\alpha}) - (1 - \mu_i) \psi'((1 - \mu_i) \phi_{i,1+\alpha})],$$

$\mathbf{D} = \text{diag}\{d_i; i = 1, \dots, n\}$, $\mathbf{D}_{1+\alpha} = \text{diag}\{d_{i,1+\alpha}; i = 1, \dots, n\}$,

$$d_i = \mu_i^2 \psi'(\mu_i \phi_i) + (1 - \mu_i)^2 \psi'((1 - \mu_i) \phi_i) - \psi'(\phi_i), \\ d_{i,1+\alpha} = \mu_i^2 \psi'(\mu_i \phi_{i,1+\alpha}) + (1 - \mu_i)^2 \psi'((1 - \mu_i) \phi_{i,1+\alpha}) - \psi'(\phi_{i,1+\alpha}).$$

References

- Basu A, Harris I, Hjort N, Jones M (1998) Robust and efficient estimation by minimising a density power divergence. *Biometrika* 85:549–559
- Cook OD, Kieschnick R, McCullough B (2008) Regression analysis of proportions in finance with self selection. *Journal of Empirical Finance* 15:860–867
- Espinheira PL, Ferrari SLP, Cribari Neto F (2008) On beta regression residuals. *Journal of Applied Statistics* 35:407–419
- Ferrari D, La Vecchia D (2012) On robust estimation via pseudo-additive information. *Biometrika* 99:238–244
- Ferrari D, Yang Y (2010) Maximum L_q -likelihood estimation. *The Annals of Statistics* 38:753–783
- Ferrari SLP, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31:799–815
- Geissinger EA, Khoo CL, Richmond IC, Faulkner SJ, Schneider DC (2022) A case for beta regression in the natural sciences. *Ecosphere* 13, DOI 10.1002/ecs2.3940
- Ghosh A (2019) Robust inference under the beta regression model with application to health care studies. *Statistical Methods in Medical Research* 28:871–888
- Ghosh A, Basu A (2013) Robust estimation for independent non-homogeneous observations using density power divergence with application to linear regression. *Electronic Journal of Statistics* 32:2420–2456
- Guolo A, Varin C (2014) Beta regression for time series analysis of bounded data, with application to canada google flu trends. *The Annals of Applied Statistics* 8:74–88
- Hampel F, Ronchetti EM, Rousseeuw P, Stahel W (2011) Robust Statistics: The Approach Based on Influence Functions. John Wiley and Sons, New York
- Hampel FR (1974) Influence curve and its role in robust estimation. *Journal of the American Statistical Association* 69:383–393

- Kerman S, McDonald JB (2015) Skewness-kurtosis bounds for EGB1, EGB2, and special cases. *Communications in Statistics - Theory and Methods* 44:3857–3864
- La Vecchia D, Camponovo L, Ferrari D (2015) Robust heart rate variability analysis by generalized entropy minimization. *Computational Statistics and Data Analysis* 82:137–151
- Ospina R, Ferrari SLP (2012) A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis* 56:1609–1623
- R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>
- Ribeiro TKA, Ferrari SLP (2022) Robust estimation in beta regression via maximum L_q -likelihood. *Statistical Papers* DOI 10.1007/s00362-022-01320-0
- Silva CC, Madruga MR, Tavares HR, Oliveira TF, Saraiva ACF (2015) Application of the beta regression on the neutralization index of power equipment insulating oil. *International Journal of Power and Energy Systems* 35:52–57
- Simas AB, Barreto-Souza W, Rocha AV (2010) Improved estimators for a general class of beta regression models. *Computational Statistic and Data Analysis* 54:348–366
- Smithson M, Verkuilen J (2006) A better lemon squeeizer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods* 11:55–71
- Swearingen CJ, Tilley CB, Adams RJ, Rumboldt Z, Nicholas SJ, Bandyopadhyay D, Woolson FR (2011) Application of beta regression to analyze ischemic stroke volume in NINDS rt-PA clinical trials. *Neuroepidemiology* 37:73–82