Bootstrap-based inferential improvements to the simplex nonlinear regression model

Alisson de Oliveira Silva¹, Jonas Weverson de Ararú Silva², Patrícia L. Espinheira³*

¹ Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, Brazil, ² Centro de Ciências Agrárias, Departamento de Ciências Fundamentais e Sociais, Areia, Paraíba, Brazil, ³ Departamento de Estatística, Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil

* patespipa@de.ufpe.br

Abstract

In this paper we evaluate the performance of point and interval estimators based on the maximum likelihood (ML) method for the nonlinear simplex regression model. Inferences based on traditional maximum likelihood estimation have good asymptotic properties, but their performance in small samples may not be satisfactory. At outset we consider the maximum likelihood estimation for the parameters of the nonlinear simplex regression model, and so we introduced a bootstrap-based correction for such estimators of this model. We also develop the percentile and bootstrap t confidence intervals for those parameters as competitors to the traditional approximate confidence interval based on the asymptotic normality of the maximum likelihood estimators (MLEs). We then numerically evaluate the performance of these different methods for estimating the simplex regression model. The numerical evidence favors inference based on the bootstrap method, in special the bootstrap t interval, which was decisive in an application to real data.

1 Introduction

Normal linear regression models are widely used in the most diverse areas of knowledge. Currently, several proposals of regression models for doubly-constrained regression models for doubly-constrained response variables, which assume continuous values in (a, b), where a and b are known and −∞ < a < b < ∞, thus, such support can be easily transformed to the unit interval.

In this context, where y ∈ (0, 1) or (y ∈ (a, b)), the normal linear model is inadequate, because besides the possibility of occurrence of fitted values smaller than 0(a) or larger than 1 (b), in general, the data present asymmetry and heteroscedasticity, violating the usual assumptions of such model. Thus, it seems more appropriate to consider models based on distributions naturally supported on (0, 1) as is the case of the simplex regression model proposed by [1], for example.

The simplex distribution was developed from the generalized inverse Gaussian distribution and is part of the class of dispersion models defined by [2], which extend the [3] generalized linear models. Several papers have been conducted using this distribution. For example, [4]
used it to evaluate longitudinal data considering the constant dispersion parameter, using generalized estimating equations. [5] modified this approach with the assumption that the dispersion parameter varies across observations. Based on the dispersion models, Using the Bayesian approach and Monte Carlo simulations, [6] evaluates the estimators of the parameters of the simplex model with variable dispersion.

Other approaches for modeling limited data are the beta [7], Kumaraswamy [8], Johnson SB [9], unit gamma [10] regression models. Recently published papers show possible advantages of using the latter distribution over the beta distribution [11, 12]. Recently, [13] proposed to the class of non-linear simplex regression models, in which they estimate the model parameters using the maximum likelihood method and derive the local influence quantities. The authors showed that when data are concentrated at the extremes of the standard unit interval, the maximum likelihood estimation process of the simplex model is more stable than that of the beta regression model. [14] presented the zero-and-one-inflated simplex distribution for modeling proportion data. The authors introduced a new algorithm to compute maximum likelihood estimates of the parameters of the simplex distribution without covariates, and developed likelihood-based inference methods for the regression model using this new distribution.

The study of the behavior of asymptotic maximum likelihood estimators in small samples is an important area of research. These estimators can be biased when the sample size is small or even moderate. The bias is actually a measure of average risk. The average risk in replacing the true value of the parameter with a plausible estimated value. Bias can also be seen as how far the mean of an estimator is far from the true value of the parameter. Thus, it is desirable to obtain estimators with reduced bias in finite samples. When the sample size is large the bias tends to zero. In the literature, there are several ways to obtain less biased estimators in small samples. Here, we shall adopt a bias correction obtained from the bootstrap method [15].

In statistical inference it is of fundamental importance to associate reliability to the point estimates of the model, and one way to do this is through the construction of the interval estimators of the parameters in association with the probability that the estimators contain the true value of these parameters. Confidence intervals can be obtained through the assumption that the asymptotic distribution of the maximum likelihood estimators is the normal probability distribution, which may require large samples to ensure the validity of these approximation. In small samples, an alternative for constructing confidence intervals with good performance with respect to both the coverage rate of the true value of the parameter and the length of the interval is the bootstrap method [15]. Specifically we shall adopt two bootstrap-based confidence intervals, namely: the percentile and bootstrap. These two schemes typically have empirical coverage rates very close to the nominal ones [16].

Regarding modeling limited continuous data, several authors have already conducted improvements on inference based on the maximum likelihood estimation method. [17] propose both the nonlinear beta regression model and improvements for the maximum likelihood estimators. [18] present corrections to the generalized likelihood ratio statistic (LR) based on [19] for the class of beta regression models whereas [11] used the same strategy considering the unitary gamma distribution. [20] also evaluate the impact of model misspecification on empirical coverage of different prediction intervals, and investigate the impact of model misspecification on three bootstrap prediction intervals. [21] discuss test inference in small samples in the class of beta regression models. The authors consider the LR test and its bootstrap versions, show that the standard LR test tends to be quite liberal in small samples and that bootstrap-based tests provide more reliable inference even when the sample size is very small.

In this document our aim is twofold. At outset we shall developed bootstrap-based inferential improvements for the parameters that index the class of nonlinear simplex regression
models proposed by [13]. In which the mean of the response variable and the dispersion parameter are related to covariates by means of nonlinear predictors. In sequence, we shall jointly evaluate the performance of the competing estimators, namely: the MLEs and the bootstrap-based estimators introduced by us.

We evaluate several aspects of interval estimation by Monte Carlo simulations. The bootstrap method proved to be an important tool estimation on nonlinear simplex regression, because through it we can get around several inferential MLEs' problems in finite samples. Finally, we present an application whose data is from the Chemistry department of the National University of Colombia.

2 Nonlinear simplex regression model

In the literature there are several discrete and continuous distributions that belong to the class of dispersion models, among which we can mention the distributions: normal, inverse normal, gamma, Von Mises, Poisson, Binomial, negative Binomial, and others. In particular, if a random variable \( y \) follows the simplex distribution denoted by \( S(\mu, \sigma^2) \) with parameters \( 0 < \mu < 1 \) and \( \sigma^2 > 0 \), the density expression takes the following form:

\[
p(y; \mu, \sigma^2) = \left[2\pi\sigma^2 \{y(1-y)\}^{3/2}\right]^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}d(y; \mu)\right\}, \quad y \in (0, 1),
\]

where the deviance component \( d(y; \mu) \) is given by \( d(y; \mu) = \frac{(y-\mu)^2}{\mu(1-y)(1-\mu)} \). The variance function for the simplex distribution is expressed as \( V(\mu) = \mu^4(1-\mu)^3 \). The mean and variance of this distribution are given, respectively, by \( \mathbb{E}(y) = \mu \) and \( \text{Var}(y) = \mu(1-\mu) - \sqrt{\frac{\alpha}{2\pi}} \exp\left\{-\frac{1}{\alpha\mu(1-\mu)}\right\}\frac{1}{\Gamma\left(\frac{1}{2}, \frac{1}{2\alpha}\right)} \), where \( \Gamma(a, b) \) corresponds to the incomplete gamma function, defined by \( \Gamma(a, b) = \int_b^\infty x^{a-1}e^{-x}dx \). For more details on these properties, see [2].

The simplex distribution is quite flexible for modeling data in the continuous range \((0, 1)\), showing different shapes according to the values of the parameters that index the distribution. For examples, such as the \( J \) shape for \( S'(0.9, 36) \), the \( U \) shape for \( S'(0.5, 121) \) and the inverse \( J \) shape for \( S'(0.1, 36) \), in addition to the common shapes, namely left-symmetric, right-symmetric and symmetric. Also, unlike the beta distribution, the simplex model is very useful for accommodating data with bimodal distributions, example for \( S'(0.5, 20) \).

Let \( y_{1t}, \ldots, y_{nt} \) be independent random variables, where each \( y_{it} \), \( t = 1, \ldots, n \), follows a simplex distribution, whose probability density function is given by (1) with mean \( \beta \) and dispersion parameter \( \sigma^2_i \). The nonlinear simplex regression model proposed by [13] is defined by (1) and the systematic components given by

\[
g(\mu_i) = f_1(x_{i1}^T; \beta) = \eta_i \quad \text{and} \quad h(\sigma^2_i) = f_2(z_{i1}^T; \gamma) = \zeta_i, \quad t = 1, \ldots, n,
\]

where \( \beta = (\beta_1, \ldots, \beta_k)^T \) and \( \gamma = (\gamma_1, \ldots, \gamma_q)^T \) are unknown regression parameter vectors such that \( \beta \in \mathbb{R}^k \) and \( \gamma \in \mathbb{R}^q \), \( k + q < n \), \( \eta = (\eta_1, \ldots, \eta_n)^T \) and \( \zeta = (\zeta_1, \ldots, \zeta_n)^T \) are nonlinear predictors and \( x_{i1} = (x_{i11}, \ldots, x_{i1k}) \) and \( z_{i1} = (z_{i11}, \ldots, z_{i1q}) \) are, respectively, \( k \) and \( q \) observations of known covariates, which may coincide fully or partially such that \( k_1 \leq k \) and \( q_1 \leq q \). In the linear models \( k_1 = k \) and \( q_1 = q \), and therefore, \( x_{i1} = (x_{i11}, \ldots, x_{i1k}) \) and \( z_{i1} = (z_{i11}, \ldots, z_{i1q}) \) are, respectively, the \( t \)-th rows of the matrices \( X \) and \( Z \), for \( t = 1, \ldots, n \). Linear models are a particular case of nonlinear models. When we have nonlinearity in the parameters, at least one of the \( \partial f_j(\cdot; \beta)/\partial \beta_j, j = 1, \ldots, k \) depends on \( \beta^T = (\beta_1, \ldots, \beta_k) \) and, at least one of the \( \partial f_j(\cdot; \gamma)/\partial \gamma_l, l = 1, \ldots, q \) depends on \( \gamma^T = (\gamma_1, \ldots, \gamma_q) \). For linear simplex models, these derivatives depend
only on the covariates \( x_1, \ldots, x_k \) and \( z_1, \ldots, z_q \) respectively, and so \( \lambda = \partial \eta / \partial \beta = X \) and 
\( \tilde{Z} = \partial \zeta / \partial \gamma = Z \).

Moreover in (2) the link functions \( g : (0,1) \to \mathbb{R} \) and \( h : (0,\infty) \to \mathbb{R} \) are strictly monotone and at least twice differentiable. Different link functions can be chosen for \( g \) and \( h \). For example, for \( \mu \) we can use the logit function \( g(\mu) = \log(\mu/(1-\mu)) \), the probit function \( g(\mu) = \Phi^{-1}(\mu) \), where \( \Phi(\cdot) \) denotes the standard normal distribution function, the log-log function \( g(\mu) = \log(-\log(1-\mu)) \) and the log-log complementary function \( g(\mu) = \log(-\log(1-\mu)) \), among others. Since \( \sigma^2 > 0 \), we can use the log function \( h(\sigma^2) = \log(\sigma^2) \) and the identity function \( h(\sigma^2) = \sigma^2 \). However, one should be aware if the estimates resulting from the likelihood maximization process take on positive values. If, in fact, the identity link function is the most appropriate, negative estimates shall not occur for the \( \sigma^2_t, t = 1, \ldots, n \) and the diagnostic analysis shall corroborate for the model goodness-of-fit to the data when using such a link function. For more details, see [3, 22]. Finally, in (2) we have that \( g(\mu_t) = \eta_t \) and \( h(\sigma^2_t) = \zeta_t, t = 1, \ldots, n \), are the mean and the dispersion submodels, respectively.

To provide the quantities related to the estimation by the maximum likelihood procedure we shall consider the general case, with nonlinearity in the parameters. Thus, we must emphasize that, \( f_1(\cdot) \) and \( f_2(\cdot) \) are differentiable functions with Jacobian matrices. Based on (1) we have that the logarithm of the likelihood function is given by
\[
\ell(\beta, \gamma) = \sum_{t=1}^n \ell_t(\mu_t, \sigma^2_t),
\]
in which
\[
\ell_t(\mu_t, \sigma^2_t) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2_t - \frac{3}{2} \log \{y_t(1-y_t)\} - \frac{1}{2\sigma^2_t} d(y_t; \mu_t).
\]

The components of the score vector \((U_\beta(\beta, \gamma)^\top, U_\gamma(\beta, \gamma)^\top)^\top\) are given by \( U_\varphi(\beta, \gamma) = \hat{\lambda}^\top S U T(y - \mu) \) and \( U_\gamma(\beta, \gamma) = \hat{Z}^\top H \alpha \), with \( \hat{\lambda} = \frac{\partial m}{\partial \eta} \) and \( \hat{Z} = \frac{\partial \gamma}{\partial \eta} \) being derivative matrices of dimension \( n \times k \) and \( n \times q \), respectively, \( y = (y_1, \ldots, y_n)^\top, \mu = (\mu_1, \ldots, \mu_n)^\top \) and \( a = (a_1, \ldots, a_n)^\top \) are \( n \times 1 \) matrices and \( U = \text{diag}(u_t, \ldots, u_n) \) is a diagonal matrix in which the \( t \)-th component is defined as
\[
u_t = \frac{1}{\mu_t(1-\mu_t)} \left\{ d(y_t; \mu_t) + \frac{1}{\mu_t^2(1-\mu_t)^2} \right\} \quad \text{with} \quad a_t = -\frac{1}{2\sigma_t^2} + \frac{d(y_t; \mu_t)}{2(\sigma_t^2)}.
\]
Moreover, \( S = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) \)
\[
T = \text{diag} \left\{ \frac{1}{g'(\mu_1)}, \ldots, \frac{1}{g'(\mu_n)} \right\} \quad \text{and} \quad H = \text{diag} \left\{ \frac{1}{h'(\sigma_1^2)}, \ldots, \frac{1}{h'(\sigma_n^2)} \right\}.
\]

To obtain the Fisher information matrix for the parameter vectors \( \beta \) and \( \gamma \), we shall use the following results: \( E[d(y; \mu)] = \sigma^2 \), \( E[(y - \mu)d'(y; \mu)] = -2\sigma^2 \) and \( E[(y - \mu)d(y; \mu)] = 0 \) [4]; \( E[(y - \mu)d^2(y; \mu)] = 0 \), \( \frac{1}{2} E[d^2(y; \mu)] = \frac{\sigma^2}{\mu(1-\mu)} + \frac{1}{\mu^2(1-\mu)^2} \) and \( \text{Var}[d(y; \mu)] = 2(\sigma^2)^2 \) [23] and \( E[d'(y; \mu)] = 0 \) [5]. The Fisher information matrix for the parameter vector \( \theta = (\beta^\top, \gamma^\top)^\top \) so-called here by \( K(\beta, \gamma) \) is a diagonal matrix with two blocks of submatrices which are \( K_{\beta\beta} \) and \( K_{\gamma\gamma} \) defined as follows \( K_{\beta\beta} = \hat{\lambda}^\top SW\hat{\lambda} \) and \( K_{\gamma\gamma} = \hat{Z}^\top D\hat{Z} \). Here, \( W = \text{diag}(w_1, \ldots, w_n) \) and \( D = \text{diag}(d_1, \ldots, d_n) \) with
\[
w_t = \left( \frac{3\sigma_t^2}{\mu_t(1-\mu_t)} + \frac{1}{\mu_t^2(1-\mu_t)^2} \right) \frac{1}{g'(\mu_t)^2} \quad \text{and} \quad d_t = \frac{1}{2(\sigma_t^2)^2} \frac{1}{h'(\sigma_t^2)^2}.
\]
Since \( K(\beta, \gamma) \) is a diagonal block matrix, the vectors \( \beta \) and \( \gamma \) are globally orthogonal [24] so that
their MLEs $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically independent. For large samples and under regularity conditions the approximate distribution of the MLEs is given by

$$
\left( \begin{array}{c}
\hat{\beta} \\
\hat{\gamma}
\end{array} \right) \approx N_{k+q} \left( \left( \begin{array}{c}
\beta \\
\gamma
\end{array} \right), K^{-1} \right) \quad \text{and} \quad K^{-1} = \begin{pmatrix}
\left( \hat{\lambda}^T S W \hat{\lambda} \right)^{-1} & 0 \\
0 & \left( \hat{Z}^T D \hat{Z} \right)^{-1}
\end{pmatrix}
$$

(4)

To measure the degree of non-constant dispersion, we define $\lambda = \frac{s^2_{\text{max}}}{s^2_{\text{min}}}$, $t = 1, \ldots, n$. Note that the greater the $\lambda$ the further away the simplex regression model with varying dispersion is from the model in which the dispersion is supposed to be fixed, since the constant dispersion models hold that $\sigma^2_{t} = \ldots = \sigma^2_{n} = \sigma^2$, in either case $\lambda = 1$. Furthermore, this $\lambda$ definition measure actually as the increase of variance response effects the estimation process of the model. To become $\sigma^2$ variable it is necessary increases the $\sigma^2_{\text{max}}$, otherwise the $\sigma^2_{\text{min}}$ should be too small and, do not plausible. Thus, as greater is $\lambda$ as greater is the response variances, in the real problems. Here during the simulations we control the value of the maximum variance, because exactly what we want is to evaluate the properties of the estimators when the variance does not explode, but only grows slightly. When working with real data, the occurrence of large values of estimated $\lambda$ is substantial, i.e., $\lambda > 1$ (in particular, $n$ when it is large).

We still need to discuss the variances of the responses further. The first part of the expression in (4) implies that the vector $(\hat{\beta}^T, \hat{\gamma}^T)^T$ is asymptotically unbiased. Thus, as the sample size increases, $(\hat{\beta}^T, \hat{\gamma}^T)^T$ is approximately unbiased and its bias should be close to zero. In theory this fact is true only when $n$ approaches infinity, that is, asymptotically. In practice the better the approximation in (4), and this depends on the distribution, the faster the bias goes to zero, i.e. this can occur for sample sizes $n = 40, 50, \ldots$

However, this assumption is mostly valid for $\hat{\beta}^T$ due to its relationship with $\hat{\mu}$ which is theoretically unbiased, (exactly and not approximately, typically). On the other hand, the relationship of the $\hat{\gamma}^T$ vector is with $\hat{\sigma}^2$, which is theoretically biased in most distributions. Thus, it is already expected that the $\hat{\gamma}^T$ bias takes a long time to converge to zero and requires large sample sizes for this to occur.

This discussion reveals that we should be more aware of how the corrections act on the $\hat{\gamma}^T$. Note that biased $\hat{\gamma}^T$ shall induce biased response variances. As a consequence, hypothesis tests and confidence intervals should perform poorly and may lead to misleading conclusions about the model, such as the exclusion of important covariates.

### 3 Point estimation of the model parameters

The maximum likelihood estimators of $\beta$ and $\gamma$ are obtained by Fisher’s interactive scoring process in which the initial guess was proposed in [13], and are usually biased when the sample size is small or even moderate, particularly $\hat{\gamma}^T$. Nevertheless, the estimator’s bias can be corrected and one possibility is to use resampling methods, which are schemes that use repeated sampling within the same sample to calculate estimates. The bootstrap method is one of the most widely used resampling methods, and one that gives very satisfactory results for estimating a model. In this paper we adopt the parametric bootstrap where in the regression models context it assumes that the probability distribution of the response variable is known and indexed by unknown parameters. [15]. The steps for performing this method both for bias correction of the MLEs and for obtaining the confidence interval are described in the Algorithms (1), (2) and (3).
Algorithm 1: Parametric method
1: Suppose that \( y = (y_1, \ldots, y_n)^T \) is a random sample such that each \( y_t \), \( t = 1, \ldots, n \), follows a distribution \( F \) supposedly known and indexed by parameter vector \( \theta \);
2: From the original sample, obtain the \( \hat{\theta} \) estimate of \( \theta \);
3: Generate \( B \) bootstrap samples of size \( n \), namely \( y_b = (y_{1b}, \ldots, y_{nb}) \) from \( F(\hat{\theta}) \), \( b = 1, \ldots, B \);
4: For each bootstrap sample \( y_b \) compute \( \hat{\theta}^{(b)} \);
5: Repeat steps 3 and 4 a great number \( B \) of times, thus obtaining \( \hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)} \);
6: Use the estimates \( \hat{\theta}^{(b)} \), with \( b = 1, \ldots, B \) for compute the desired quantities, for instance: mean, variance, confidence interval, etc. regarding distribution of \( y \).

Once the estimate of the estimator’s bias is obtained we can construct the bias-corrected point estimators. Using the steps of the bootstrap method presented in Algorithm (1), a bootstrap estimate of the bias can be obtained by
\[
\hat{B}_{\text{boot}}(\hat{\theta}) = \hat{\theta}^* - \hat{\theta},
\]
where \( \hat{\theta}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{(b)} \), i.e., it is possible to approximate the expected value from the arithmetic mean of the bootstrap estimates of \( \theta \). Thus, we can obtain an estimator corrected up to second order by bootstrap [15, 25]:
\[
\tilde{\theta} = \hat{\theta} - \hat{B}_{\text{boot}}(\hat{\theta}) = \hat{\theta} - (\hat{\theta}^* - \hat{\theta}) = 2\hat{\theta} - \hat{\theta}^*.
\]
This estimator has the same asymptotic properties as the usual MLE and presents Lower bias in small samples [16]. A detailed discussion of the bootstrap second-order bias correction and its relation to the analytic correction can be found in [26].

4 Interval estimation of the model parameters

A set constructed on the basis of a point estimator in association with a probability that this set contains the true value of the parameter, defines a confidence interval estimator. The general form for approximate confidence intervals (CI) for \( \theta \) is: \( \mathbb{P}[l_1 \leq \theta \leq l_2] \approx 1 - \alpha \), \( 0 < \alpha < 1 \), where \( l_1 \) and \( l_2 \) are the lower and upper bounds of the confidence interval, respectively, and \( 1 - \alpha \) is the confidence level which converges to the probability of coverage. We should emphasize that \( l_1 \) and \( l_2 \) are quantiles of a distribution indexed by the parameter \( \theta \). Whether we assume that this distribution is known, it is possible to construct exact confidence intervals. However, defining the exact analytical distribution of a random variable is typically highly challenging.

Fortunately, there are diverse approaches to building approximate confidence intervals. The most widely used is the asymptotic confidence interval, which assumes asymptotic normality of the MLEs. According to (4) for the simplex model in large samples the distribution of \( \hat{\theta} = (\hat{\beta}^\top, \hat{\gamma}^\top)^\top \) is approximately normal with mean equal to \( \theta = (\beta^\top, \gamma^\top)^\top \) and the variance and covariance matrix given by (4). More precisely, we have that \( K^\beta = (\tilde{X}^\top SW\tilde{X})^{-1} \) evaluated at \( \theta \) is the \( k \times k \) matrix of variances and covariances of \( \beta \) and \( K^\gamma = (\tilde{Z}^\top D\tilde{Z})^{-1} \) evaluated at \( \theta \) is the \( q \times q \) matrix of variances and covariances of \( \gamma \).

Consider \( \beta_i \) and \( \gamma_j \), with \( i = 1, \ldots, k \) and \( j = 1, \ldots, q \), the \( i \)th and \( j \)th components of the vectors \( \beta \) and \( \gamma \), respectively. We shall denote \( K_{ii}^\beta \) and \( K_{jj}^\gamma \) as the \( i \)th and \( j \)th components of the main diagonal of the matrices \( K^\beta(\hat{\theta}) \) and \( K^\gamma(\hat{\theta}) \), respectively. Therefore, it follows that
\[ \left( \hat{\beta}_i - z_{\alpha/2} \left( K_{\beta}^{\alpha/2} \right), \hat{\beta}_i + z_{\alpha/2} \left( K_{\beta}^{\alpha/2} \right) \right) \] and \[ \left( \hat{\gamma}_j - z_{\alpha/2} \left( K_{\gamma}^{\alpha/2} \right), \hat{\gamma}_j + z_{\alpha/2} \left( K_{\gamma}^{\alpha/2} \right) \] are intervals with confidence approximately equal to \( 1 - \alpha \) for \( \beta_i \) and \( \gamma_j \), respectively, where \( z_{\alpha/2} \) is the \( 1 - \frac{\alpha}{2} \) quantile of the standard normal distribution. These intervals based on MLE may require large samples for the coverage to be close to the nominal ones. In small samples, they can have large coverage errors \([15, 25]\).

An workaround for reaching improvements to confidence intervals in small samples, without analytical complexities, is the bootstrap method. This approach typically provides confidence intervals that have coverage levels close to the true coverage probability. Here, we shall discuss two strategies bootstrap-based confidence intervals, namely: the percentile and bootstrap.

The percentile bootstrap confidence interval we shall denote by ‘\texttt{Boot}' \([16]\) is a bootstrap approach built based on a \( B \) finite replicates of the estimators of the parameters of interest. Furthermore, it displays the monotonic transformation invariance property. Let \( F(\theta) \) be the distribution of the response variable assumed to be known and indexed by the parameter vector \( \theta \). Moreover, let \( \hat{G} \) be the empirical distribution function of \( \hat{\theta} \) obtained from the \( B \) bootstrap replicas. We can construct the percentile confidence interval, with approximate coverage level \( 1 - \alpha \), by calculating \( 1 - \alpha/2 \) and \( \alpha/2 \) quantiles of \( \hat{G} \). The interval is given by \((\hat{G}^{-1}(\alpha/2), \hat{G}^{-1}(1 - \alpha/2))\). Defining \( \hat{G}^{-1}(\alpha/2) = \hat{G}^{\alpha/2} \) and \( \hat{G}^{-1}(1 - \alpha/2) = \hat{G}^{1-\alpha/2} \). The expressions of the percentile bootstrap confidence intervals for the parameters of the nonlinear simplex regression model are given by:

\[ \left( \hat{\beta}_i^{\alpha/2}, \hat{\beta}_i^{1-\alpha/2} \right) \quad \text{and} \quad \left( \hat{\gamma}_j^{\alpha/2}, \hat{\gamma}_j^{1-\alpha/2} \right). \] (5)

\( i = 1, \ldots, k \) and \( j = 1, \ldots, q \), the \( i \)-th and \( j \)-th components of the vectors \( \beta \) and \( \gamma \). The percentile interval is not necessarily symmetric about the value of \( \hat{\beta} \) and \( \hat{\gamma} \). Its construction ensures that improper values for the parameter of interest are not included in the confidence interval. The steps for its construction are described in Algorithm 2.

\textbf{Algorithm 2:} Bootstrap confidence interval—Percentile

1: Generate \( B \) bootstrap samples \( y^b = (y_1^b, y_2^b, \ldots, y_n^b) \) based on \( F(\hat{\theta}) \), for \( b = 1, \ldots, B \);
2: Let \( \hat{\theta} = s(\hat{y}) \), in which, \( \hat{y} = (y_1, \ldots, y_n) \) is the original sample. Thus, the respective bootstrap estimate of de \( \hat{\theta} \) is computed as follows: \( \hat{\theta}^b = s(y^b), b = 1, \ldots, B \);
3: The \( B \) replicas of \( \hat{\theta} \) must be ordered.
4: The lower and upper limits of the percentile interval are provide by the replicas of \( \hat{\theta} \) of order \( B \times (\alpha/2) \) and \( B \times (1 - \alpha/2) \), respectively, by assuming that \( B \times (\alpha/2) \) and \( B \times (1 - \alpha/2) \) are integers and \( 0 < \alpha < 1 \); Meaning \( \hat{G}^{-1}(\alpha/2) = \hat{G}^{\alpha/2} \) and \( \hat{G}^{-1}(1 - \alpha/2) = \hat{G}^{1-\alpha/2} \).

4.1: Whether \( B \times (1 - \alpha/2) \) and \( B \times (1 - \alpha/2) \) are not integers, we can use the following procedure:

4.1.1: Assuming \( 0 < \alpha < 1 \), let \( p = \lfloor (B + 1)\alpha/2 \rfloor \) be the largest integer less than or equal to the number \( (B + 1)\alpha/2 \); then, we define the lower and upper bounds of the percentile interval by the \( p \)-th and \( (B + 1 - p) \)-th ordered elements of the \( B \) bootstrap replicas of \( \hat{\theta} \), respectively.

The bootstrap, confidence interval, here so-called as ‘\texttt{Boot}' \([16]\) is a pivotal method to construct confidence intervals that rely on the traditional t-Student confidence interval. This
interval is based on the bootstrap estimate of the $T$ distribution, where $T$ is given by

$$T = \frac{\hat{\theta} - \theta}{\hat{e}(\hat{\theta})},$$

where $\hat{e}(\hat{\theta})$ is the standard error of $\hat{\theta}$. The construction of the bootstrap confidence interval is given by Algorithm 3.

**Algorithm 3**: Bootstrap confidence interval—Bootstrap

1. Generate $B$ bootstrap samples $y^b = (y_1^b, y_2^b, \ldots, y_n^b)$ from $F(\hat{\theta})$;
2. For each bootstrap sample, it is compute

$$T^b = \frac{\hat{\theta}^b - \hat{\theta}}{\hat{e}(\hat{\theta})},$$

with $b = 1, 2, \ldots, B$, where $\hat{\theta} = s(y)$ is the estimated value of $\theta$ from the original sample $y$, $\hat{\theta}^b = s(y^b)$ is the estimated value of $\theta$ for the bootstrap sample $y^b$ and $\hat{e}(\hat{\theta})$ is the standard error of $\hat{\theta}$ for the bootstrap sample $y^b$. Note that $\hat{e}(\theta)$ is the standard error of $\theta$ for the bootstrap sample $y^b$.

3. The $\alpha/2$ and $1 - \alpha/2$ percentiles of $T^b$ are obtained by the values $\hat{\beta}^{(\alpha/2)}$ and $\hat{\beta}^{(1-\alpha/2)}$, respectively, as follows

$$\hat{\beta}^{(\alpha/2)} = \left\{ \frac{T^b \leq \hat{\beta}^{(\alpha/2)}}{B} \right\}$$

and

$$\hat{\beta}^{(1-\alpha/2)} = \left\{ \frac{T^b \leq \hat{\beta}^{(1-\alpha/2)}}{B} \right\} = 1 - \frac{\alpha}{2}.$$

Thus, the bootstrap confidence interval is given by

$$\left( \hat{\theta} - \hat{\beta}^{(1-\alpha/2)} \hat{e}(\hat{\theta}), \hat{\theta} - \hat{\beta}^{(\alpha/2)} \hat{e}(\hat{\theta}) \right),$$

in which $\hat{\beta}^{(\alpha/2)} = \hat{e}(\hat{\theta})$. The amounts $\hat{\beta}^{(\alpha/2)}$ and $\hat{\beta}^{(1-\alpha/2)}$ can be obtained as follows:

1. Sort the $B$ bootstrap replicates $T^b$;
2. The quantiles $\hat{\beta}^{(\alpha/2)}$ and $\hat{\beta}^{(1-\alpha/2)}$ are, respectively, the replicates corresponding to the integer parts of $B \times (\alpha/2)$ and $B \times (1 - \alpha/2)$;

2.1. If $B \times (\alpha/2)$ and $B \times (1 - \alpha/2)$ are not integers, we can use the following procedure:

Assuming $0 < \alpha < 1$, is $k = [(B + 1)\alpha/2]$ the largest integer less than or equal to the number $(B + 1)\alpha/2$. Thus, the quantiles bootstrap $\hat{\beta}^{(\alpha/2)}$ and $\hat{\beta}^{(1-\alpha/2)}$ are given, respectively, by the $k$-th and $(B + 1 - k)$-th ordered elements of $T^b$. Therefore, the bootstrap confidence intervals for the parameters of the simplex nonlinear regression model are given by the following expressions

$$(\hat{\beta}_1 - \hat{\beta}^{(1-\alpha/2)} \hat{e}(\hat{\theta}), \hat{\beta}_1 - \hat{\beta}^{(\alpha/2)} \hat{e}(\hat{\theta}))$$

and

$$(\hat{\gamma}_j - \hat{\gamma}^{(1-\alpha/2)} \hat{e}(\hat{\theta}), \hat{\gamma}_j - \hat{\gamma}^{(\alpha/2)} \hat{e}(\hat{\theta})),$$

with $i = 1, \ldots, k$ and $j = 1, \ldots, q$, the $i$-th and $j$-th components of the vectors $\beta$ and $\gamma$.

According to [16], the bootstrap intervals outperform the asymptotic interval displaying empirical coverages closer to the exact nominal levels, but tend not to be accurate in actual practice. Percentile intervals are more accurate, but display less satisfactory coverage performances. An outstanding discussion on bootstrap-based confidence intervals can be found in [27]. In what follows we shall evaluate the finite-sample performances of the confidence intervals introduced in this section.

### 5 Numerical results on point estimation

In this section we present the Monte Carlo simulations results, carried out to evaluate the performances of the maximum likelihood estimators of the nonlinear simplex regression model.
and the bootstrap versions on small samples. In what follows we shall assuming the following nonlinear simplex regression model:

\[ g(\mu_t) = \beta_1 + x_{1t}^2 + \beta_2 x_{2t} + \beta_3 x_{3t} \quad \text{and} \quad h(\sigma_t^2) = \gamma_1 + \gamma_2, \quad t = 1, \ldots, n, \]

where \( g(\cdot) \) and \( h(\cdot) \) are the logit and logarithmic link functions, respectively. The realizations of the covariates were generated using the uniform distribution as follows: \( x_{1t} \sim U(0.5, 1.5), x_{2t} \sim U(0, 1), x_{3t} \sim U(-0.5, 0.5) \) and \( z_t \sim U(0.5, 1.5) \) which are retained fixed for each Monte Carlo replication. Three different scenarios were considered for the mean response, namely: \( \mu_t \in (0.02, 0.32) \) with \( \beta = (-2.4, 1.2, -1.5, -1.7)^T \); \( \mu_t \in (0.19, 0.86) \) with \( \beta = (-1.7, -1.8, 1.2, -1.3)^T \) and \( \mu_t \in (0.78, 0.98) \) with \( \beta = (2.1, -1.5, -1.6, -1.2)^T \). Furthermore, concerning the degree of non-constant dispersion, we report here the results for \( \lambda \approx 12 \) with \( \gamma = (-1.3, -1.6)^T \); \( \lambda \approx 45 \) with \( \gamma = (-1.3, -2.1)^T \) and \( \lambda \approx 128 \) with \( \gamma = (-1.3, -2.4)^T \). The sample sizes chosen were \( n = 40, 80 \) and 120. For the last two cases we initially generated \( n = 40 \) covariates observations and these were replicated twice and three times, respectively to obtain the sample sizes \( n = 80 \) and \( n = 120 \). This was done to ensure that the non-constant dispersion intensity was the same for all sample sizes. The number of Monte Carlo and bootstrap replications were \( R = 10000 \) and \( B = 500 \), respectively. The parameter estimates in (6) were obtained by maximizing the log-likelihood function using the Fisher’s nonlinear optimization method.

For each both Monte Carlo replicate and the maximum likelihood estimate of the model parameters, \( B \) bootstrap replicate estimates were generated. Thus, at the end of the bootstrap some quantities regarding the parameters are estimated, namely: the corrected bootstrap estimates and the bootstrap confidence intervals, percentile and bootstrap, respectively. Finally, outside the bootstrap, the asymptotic intervals of the parameters are also computed based on the quantiles of the standard normal distribution.

Aiming to evaluate the performance of the point estimation of the parameters, the relative bias and the square root of the mean square error were calculated for each sample size. Additionally, we introduce a measure suggested during the review of the article, which we shall so-call Unified Quadratic Bias (UQB) define as \( \sqrt{bias^2_{\mu} + \ldots + bias^2_{\gamma}} \). In Tables 1–3 we consider, respectively, the scenarios where \( \mu_t \in (0.02, 0.32), (\mu_t \approx 0), (\mu_t \approx 0.19, 0.86), (\mu_t \approx 0.5) \) and \( \mu_t \in (0.78, 0.98), (\mu_t \approx 1), t = 1, \ldots, n \). In these tables are reported the relative biases and the square roots of the mean square errors (RMSEs) of the parameter estimators for \( n = 40, 80 \) and 120 and \( \lambda \approx 12, 45 \) and 128. We observe that in modulo the estimates of the relative bias of the bootstrap corrected estimators are smaller than those of the maximum likelihood estimators, evidencing the efficacy of the bootstrap scheme in bias correction.

For example, the relative bias estimate of the \( \hat{\beta}_1 \) (BOOT) estimator is equal to 0.0003, while that of the \( \hat{\beta}_1 \) (MLE-asymptotic) is 0.001. For \( \mu_t \in (0.19, 0.86), n = 120 \) and \( \lambda \approx 12 \), the estimated biased is equal to 0.001 for \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \). In fact, it is noteworthy the high performance of the bootstrap correction when \( \mu_t \in (0.19, 0.86) \), since its estimators exhibit lower biases than the MLEs-asymptotic for all model parameters, for the different levels of non-constant dispersion and the sample sizes. In all scenarios considered, we note that the RMSEs of the estimators decrease when the sample size increases.

As we had expected the MLEs-asymptotic of the parameters of the dispersion submodel tend to be more biased than those of the mean submodel, especially regarding \( \gamma_1 \). For instance, for \( \mu_t \in (0.02, 0.32), n = 120 \) and \( \lambda \approx 45 \), the relative bias estimate of \( \hat{\gamma}_1 \) is equal to 0.043, while that of the \( \hat{\gamma}_1 \) is \( < 0.0001 \). More expressive are the biases of \( \hat{\gamma}_1 \) and \( \hat{\gamma}_2 \) which drop from \((0.136, -0.010) \) to \((-0.001, 0.001) \) after the bootstrap correction, respectively, when \( n = 40 \) and \( \lambda \approx 45 \).
Table 1. Relative biases and root mean square errors of the Maximum Likelihood Estimators (MLEs-asymptotic) and bootstrap corrected MLEs of the model parameters: Logit($\mu_i / (1 - \mu_i)$) = $\beta_1 x_1^{\lambda_i} + \beta_2 x_2 + \beta_3 x_3$, and log($\sigma_i^2$) = $\gamma_1 z_i^2 + \gamma_2 x_4$, $t = 1, \ldots, n$, $\beta = (-2.4, 1.2, -1.5, -1.7)$, $\mu_i \in (0.02, 0.32)$, $t = 1, \ldots, n$.

| n  | $\theta$ | $\lambda = 12$ | $\lambda = 45$ | $\lambda = 128$ |
|----|----------|----------------|---------------|---------------|
|    |          | MLE | BOOT | MLE | BOOT | MLE | BOOT | MLE | BOOT |
|    | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ | Bias | $\sqrt{\text{MSE}}$ |
| 40 | $\beta_1$ | 0.0003 | 0.082 | -0.001 | 0.082 | 0.001 | 0.085 | 0.0002 | 0.0854 | 0.002 | 0.089 | 0.001 | 0.089 |
|    | $\beta_2$ | -0.001 | 0.131 | 0.001 | 0.131 | 0.004 | 0.134 | 0.004 | 0.135 | 0.004 | 0.143 | 0.002 | 0.142 |
|    | $\beta_3$ | 0.003 | 0.137 | 0.002 | 0.138 | 0.001 | 0.143 | 0.003 | 0.143 | 0.001 | 0.153 | -0.0002 | 0.153 |
|    | $\beta_4$ | -0.001 | 0.115 | -0.001 | 0.115 | -0.001 | 0.116 | -0.001 | 0.116 | 0.001 | 0.123 | 0.001 | 0.123 |
|    | $\gamma_1$ | 0.136 | 0.274 | -0.001 | 0.239 | 0.136 | 0.273 | -0.001 | 0.237 | 0.138 | 0.271 | 0.003 | 0.233 |
|    | $\gamma_2$ | -0.010 | 0.323 | 0.001 | 0.311 | <0.0001 | 0.233 | 0.001 | 0.229 | 0.001 | 0.191 | 0.001 | 0.188 |

| UQB | 0.136 | 0.003 | 0.136 | 0.004 | 0.138 | 0.004 |
|-----|-------|-------|-------|-------|-------|-------|
| 80  | $\beta_1$ | 0.001 | 0.057 | 0.001 | 0.057 | <0.0001 | 0.059 | -0.001 | 0.059 | 0.001 | 0.064 | 0.003 | 0.064 |
|    | $\beta_2$ | -0.001 | 0.091 | <0.0001 | 0.091 | 0.002 | 0.094 | 0.001 | 0.095 | 0.003 | 0.099 | -0.001 | 0.098 |
|    | $\beta_3$ | -0.001 | 0.093 | -0.001 | 0.093 | 0.002 | 0.098 | 0.001 | 0.098 | <0.0001 | 0.108 | -0.001 | 0.108 |
|    | $\beta_4$ | -0.0004 | 0.079 | -0.0004 | 0.079 | -0.0002 | 0.081 | <0.0001 | 0.081 | -0.001 | 0.085 | -0.001 | 0.085 |
|    | $\gamma_1$ | 0.067 | 0.173 | 0.002 | 0.160 | 0.065 | 0.173 | 0.001 | 0.162 | 0.060 | 0.168 | -0.001 | 0.158 |
|    | $\gamma_2$ | -0.0004 | 0.209 | 0.003 | 0.206 | -0.0004 | 0.154 | 0.001 | 0.154 | -0.003 | 0.128 | -0.0003 | 0.127 |

| UQB | 0.067 | 0.004 | 0.065 | 0.002 | 0.066 | 0.004 |
|-----|-------|-------|-------|-------|-------|-------|
| 120 | $\beta_1$ | 0.001 | 0.046 | 0.0003 | 0.046 | 0.001 | 0.049 | <0.0001 | 0.049 | 0.0002 | 0.052 | -0.0002 | 0.052 |
|    | $\beta_2$ | -0.001 | 0.075 | 0.0003 | 0.075 | 0.001 | 0.077 | -0.0002 | 0.078 | 0.006 | 0.081 | 0.002 | 0.081 |
|    | $\beta_3$ | -0.0003 | 0.077 | -0.001 | 0.077 | -0.0003 | 0.081 | -0.001 | 0.081 | 0.001 | 0.089 | 0.001 | 0.089 |
|    | $\beta_4$ | -0.0002 | 0.065 | -0.0002 | 0.065 | -0.0001 | 0.066 | -0.0008 | 0.066 | -0.0003 | 0.070 | 0.0002 | 0.071 |
|    | $\gamma_1$ | 0.043 | 0.134 | <0.000 | 0.128 | 0.043 | 0.135 | 0.001 | 0.128 | 0.039 | 0.133 | <0.0001 | 0.127 |
|    | $\gamma_2$ | -0.001 | 0.165 | 0.0004 | 0.164 | -0.001 | 0.123 | <0.0001 | 0.124 | -0.003 | 0.105 | -0.001 | 0.104 |

For $\hat{\gamma}_1$, in particular, the bootstrap correction provides a substantial reduction of the estimated bias. This is important since the correct estimation of the of the dispersion submodel parameters, directly interferes with the estimates of the response variances, which, when corrected, produce Z-tests that lead to truer decisions. Even so, the corrections were also effective for $\hat{\beta}_i$, $i = 1, 2, 3, 4$ because the goal is for the bias values to be as close to zero as possible, and for these parameters, through correction, the estimated bias became some times $< 0.0001$, i.e. the goal was achieved.

It is important to note that the estimated biases of the usual and corrected maximum likelihood estimators are notably smaller when the mean of the response variable is close to the upper limit of the unit interval than for the two other scenarios considered (Tables 1 and 3). Based on the Unified Quadratic Bias measure it becomes more evident how effective the bias correction we propose is. Let shall evaluated the results on $\mu_i \in (0.78, 0.98)$ and $n = 40$, for $\lambda = 12, 45$ and 128, we have that the values of the UQB for the original MLEs are equal to 0.132, 0.137 and 0.138, whereas for the corrected version these values became 0.005, 0.001 and 0.002, respectively (Table 3).

6 Numerical results on confidence intervals

Concerning interval estimation, we computing the empirical coverage of the intervals (%), obtained from the relative frequencies in which the intervals contained the true value of the...
Table 2. Relative biases and root mean square errors of the Maximum Likelihood Estimators (MLEs) and bootstrap corrected MLEs of the model parameters: Logit(μ/1−μ) = β_1 + x_{11}^2 + β_2 x_{12} + β_3 x_{13} and log(σ^2) = γ_1 + z_{14}^2, β = (−1.7, −1.8, 1.2, −1.3)’, μ ∈ (0.19, 0.86), t = 1, . . . , n.

| n   | θ     | λ ≈ 12 Bias | λ ≈ 12 √MSE | λ ≈ 45 Bias | λ ≈ 45 √MSE | λ ≈ 128 Bias | λ ≈ 128 √MSE |
|-----|-------|-------------|--------------|-------------|--------------|--------------|--------------|
| 40  | β_1   | 0.002       | 0.119        | 0.002       | 0.125        | <0.0001      | 0.124        |
|     | β_2   | 0.001       | 0.098        | −0.0004     | 0.097        | <0.0001      | 0.101        |
|     | β_3   | 0.007       | 0.220        | 0.004       | 0.219        | 0.001        | 0.232        |
|     | β_4   | 0.005       | 0.177        | 0.001       | 0.177        | 0.008        | 0.176        |
|     | γ_1   | 0.130       | 0.274        | −0.004      | 0.241        | 0.139        | 0.275        |
|     | γ_2   | −0.020      | 0.333        | 0.0003      | 0.309        | 0.002        | 0.235        |
| UQB |       | 0.133       | 0.006        | 0.139       | 0.004        | 0.137        | 0.004        |

| n   | θ     | λ ≈ 12 Bias | λ ≈ 12 √MSE | λ ≈ 45 Bias | λ ≈ 45 √MSE | λ ≈ 128 Bias | λ ≈ 128 √MSE |
|-----|-------|-------------|--------------|-------------|--------------|--------------|--------------|
| 80  | β_1   | 0.001       | 0.083        | 0.0002      | 0.083        | 0.001        | 0.087        |
|     | β_2   | 0.001       | 0.068        | −0.0004     | 0.068        | 0.0003       | 0.071        |
|     | β_3   | 0.001       | 0.154        | −0.01       | 0.154        | 0.002        | 0.162        |
|     | β_4   | 0.002       | 0.124        | 0.0002      | 0.123        | 0.003        | 0.127        |
|     | γ_1   | 0.064       | 0.171        | <0.0001     | 0.159        | 0.065        | 0.171        |
|     | γ_2   | −0.005      | 0.209        | 0.003       | 0.202        | 0.004        | 0.154        |
| UQB |       | 0.064       | 0.003        | 0.065       | 0.002        | 0.066        | 0.001        |

| n   | θ     | λ ≈ 12 Bias | λ ≈ 12 √MSE | λ ≈ 45 Bias | λ ≈ 45 √MSE | λ ≈ 128 Bias | λ ≈ 128 √MSE |
|-----|-------|-------------|--------------|-------------|--------------|--------------|--------------|
| 120 | β_1   | 0.001       | 0.068        | <0.0001     | 0.068        | 0.001        | 0.070        |
|     | β_2   | 0.001       | 0.055        | <0.0001     | 0.055        | 0.001        | 0.058        |
|     | β_3   | 0.001       | 0.126        | −0.001      | 0.126        | 0.0002       | 0.132        |
|     | β_4   | 0.002       | 0.100        | 0.001       | 0.100        | 0.002        | 0.102        |
|     | γ_1   | 0.042       | 0.134        | <0.0001     | 0.127        | 0.043        | 0.134        |
|     | γ_2   | −0.004      | 0.168        | 0.0004      | 0.165        | −0.0002      | 0.124        |
| UQB |       | 0.042       | 0.001        | 0.043       | 0.001        | 0.043        | 0.001        |

The lower and upper bounds were also estimated (via the average after the end of the Monte Carlo process), thus we were able to estimate the average length of the intervals and left and right non-coverage rates. The left rate is computed whenever the interval upper limit is less than the true value of the parameter and right rate is computed whenever the interval lower limit is greater than true value of the parameter.

In what follows we report the results of Monte Carlo simulations on interval estimation. We shall just take the nominal levels 0.90 and 0.95 concerning to Tables 4 and 5, respectively. These tables display the coverage rates of the following competing interval estimators: the asymptotic ML-like or ML interval approximation (ML-I), bootstrap (Boot), and percentile (Bootp) for the model parameters in (6).

Regarding coverage rates the interval that performs best is the bootstrapp, with empirical coverage substantially closest to the nominal levels, for all parameters model. The asymptotic confidence interval displayed considerable undercoverage and the percentile confidence interval Bootp overall outperforms the ML-I, only for γ_1 the Boot displays a poor performance. For instance, consider n = 40, μ ∈ (0.19, 0.86), 1 − α = 0.90 and all λ’s values, the Bootp coverage ratios for this parameter is approximately equal to 0.66. Whereas those of the ML-I and Boot, are approximately equal to 0.80 and 0.90, respectively. These behavior are similar for all scenarios of the mean response. The conclusions about coverage rates are quite similar whatever the nominal level is. To exemplify, for 1 − α = 0.95 and for the same other settings when
Table 3. Relative biases and root mean square errors of the Maximum Likelihood Estimators (MLEs-asymptotic) and bootstrap corrected MLEs of the model parameters: \( \text{Logit}(\mu_t/1-\mu_t) = \beta_1 + \beta_2 x_{1t} + \beta_3 x_{3t} + \beta_4 x_{4t} \) and \( \log(\sigma_t^2) = \gamma_1 + \gamma_2 t, \beta = (2.1, -1.5, -1.6, -1.2), \mu_t \in (0.78, 0.98), t = 1, \ldots, n. \)

| \( n \) | \( \theta \) | \( \lambda \approx 12 \) | \( \lambda \approx 45 \) | \( \lambda \approx 128 \) |
|---|---|---|---|---|
|   | MLE | BOOT | MLE | BOOT | MLE | BOOT | MLE | BOOT |
|   | Bias | \( \sqrt{MSE} \) | Bias | \( \sqrt{MSE} \) | Bias | \( \sqrt{MSE} \) | Bias | \( \sqrt{MSE} \) |
| 40 | \( \beta_1 \) | 0.001 | 0.063 | 0.0003 | 0.063 | 0.001 | 0.067 | <0.0001 | 0.067 | 0.001 | 0.067 | <0.0001 | 0.067 |
|   | \( \beta_2 \) | -0.0003 | 0.056 | <0.0001 | 0.056 | -0.0003 | 0.059 | <0.0001 | 0.059 | -0.001 | 0.062 | -0.0002 | 0.061 |
|   | \( \beta_3 \) | 0.0004 | 0.123 | 0.001 | 0.123 | <0.0001 | 0.129 | 0.0003 | 0.129 | -0.001 | 0.132 | -0.001 | 0.132 |
|   | \( \beta_4 \) | -0.002 | 0.097 | -0.001 | 0.097 | 0.0004 | 0.101 | 0.001 | 0.101 | 0.0003 | 0.101 | 0.001 | 0.101 |
|   | \( \gamma_1 \) | 0.130 | 0.271 | -0.004 | 0.238 | 0.137 | 0.273 | -0.001 | 0.236 | 0.138 | 0.274 | -0.001 | 0.237 |
|   | \( \gamma_2 \) | -0.023 | 0.336 | 0.002 | 0.310 | -0.0004 | 0.239 | -0.0001 | 0.228 | 0.004 | 0.191 | 0.0004 | 0.186 |
|   | UQB | 0.132 | 0.005 | 0.137 | 0.001 | 0.138 | 0.002 |
| 80 | \( \beta_1 \) | 0.0004 | 0.044 | <0.0001 | 0.044 | 0.0004 | 0.047 | <0.0001 | 0.047 | 0.001 | 0.048 | 0.0004 | 0.048 |
|   | \( \beta_2 \) | <0.0001 | 0.039 | <0.0001 | 0.039 | -0.001 | 0.041 | -0.0004 | 0.041 | -0.001 | 0.043 | -0.001 | 0.043 |
|   | \( \beta_3 \) | 0.001 | 0.086 | 0.001 | 0.086 | -0.001 | 0.091 | -0.001 | 0.091 | -0.0003 | 0.093 | -0.0002 | 0.093 |
|   | \( \beta_4 \) | 0.001 | 0.069 | 0.001 | 0.070 | 0.0004 | 0.071 | 0.001 | 0.071 | <0.0001 | 0.072 | 0.001 | 0.072 |
|   | \( \gamma_1 \) | 0.064 | 0.171 | 0.001 | 0.159 | 0.067 | 0.172 | 0.001 | 0.158 | 0.067 | 0.172 | 0.000 | 0.159 |
|   | \( \gamma_2 \) | -0.008 | 0.215 | 0.001 | 0.207 | 0.001 | 0.156 | 0.001 | 0.153 | 0.003 | 0.129 | 0.001 | 0.128 |
|   | UQB | 0.065 | 0.002 | 0.067 | 0.002 | 0.067 | 0.002 |
| 120 | \( \beta_1 \) | 0.0002 | 0.037 | <0.0001 | 0.037 | 0.001 | 0.038 | 0.0002 | 0.038 | 0.0003 | 0.039 | <0.0001 | 0.039 |
|   | \( \beta_2 \) | 0.0003 | 0.031 | <0.0001 | 0.031 | <0.0001 | 0.034 | <0.0001 | 0.034 | <0.0001 | 0.035 | <0.0001 | 0.035 |
|   | \( \beta_3 \) | <0.0001 | 0.071 | <0.0001 | 0.071 | 0.0003 | 0.075 | 0.0004 | 0.075 | <0.0001 | 0.075 | <0.0001 | 0.075 |
|   | \( \beta_4 \) | -0.001 | 0.056 | -0.001 | 0.056 | -0.001 | 0.057 | -0.0004 | 0.057 | -0.0003 | 0.059 | <0.0001 | 0.059 |
|   | \( \gamma_1 \) | 0.042 | 0.134 | 0.001 | 0.128 | 0.046 | 0.136 | 0.003 | 0.128 | 0.045 | 0.134 | 0.001 | 0.128 |
|   | \( \gamma_2 \) | -0.004 | 0.168 | 0.001 | 0.164 | 0.001 | 0.123 | 0.001 | 0.121 | 0.002 | 0.102 | 0.001 | 0.102 |
|   | UQB | 0.042 | 0.002 | 0.046 | 0.003 | 0.045 | 0.001 |

https://doi.org/10.1371/journal.pone.0272512.t003

1 - \( \alpha \) = 0.90, the ML-I_{\text{Boot}}, Boot, and Bootp coverage rates are around 0.87, 0.95 and 0.74, respectively (Table 5). Now consider \( n = 120, 1 - \alpha = 0.95, \lambda = 45, \mu_t \in (0.19, 0.86) \), as to \( \beta_3 \), those values are equal to 0.933, 0.942 and 0.934 (Table 8). Meaning, even when the size of the sample increases the empirical coverage of the Boot, interval is closest to the nominal level.

Our interest hereafter shall lie in evaluate some interval properties only for the nonlinearity parameters of the mean and the dispersion submodels, meaning \( \beta_1 \) and \( \gamma_2 \). Tables 6 through 9 present the mean lower (Lower) and upper (Upper) bounds, mean lengths (Size), the empirical probability of coverage (Coverage), as well as the left and right coverage rates (%) of the interval estimators of the previously mentioned parameters. These last two quantities evaluate the balancing of the interval. Perfect balancing occurs when these two percentages (%) are identical.

In Tables 6–8 are presented the results for \( \beta_2, n = 40, n = 80 \) and \( n = 120 \), and only for \( \mu_t \in (0.02, 0.32), t = 1, \ldots, n \), where \( \beta_2 = 1.2 \). Note that if \( \beta_2 = 1 \) the mean submodel becomes linear. For \( n = 40, \lambda \approx 12 \) and \( \lambda \approx 45 \) only the Boot, interval and for \( 1 - \alpha = 0.99 \) considers this possibility.

Its empirical probability of coverage is the closest to the true one, however, the its mean length interval is longer than those of the other two intervals, which allows the inclusion of \( \beta_2 = 1.0 \), i.e., linearity (Table 6). This behavior of the Boot, interval holds for \( n = 80 \) (Table 7). When \( \lambda \approx 128 \) all intervals include the possibility of linearity when \( n = 40 \), i.e., \( \beta_2 = 1.0 \) (Table 6), for all confidence nominal levels. This result is interesting, as it shows how the
### Table 4. Coverage rates of the interval estimators: ML-I, Boot, and Bootp for $\theta$, the model parameters: Logit$(\mu_i/1 - \mu_i) = \beta_1 + x_{i1}^\beta + \beta_2 x_{i2} + \beta_3 x_{i3}$ and $\log(\sigma_i^2) = \gamma_1 + \lambda x_{i4}$, $t = 1, \ldots, n, 1 - \alpha = 0.90$. 

| $n$ | $\theta$ | $\mu_i \in (0.02, 0.32)$ | $\mu_i \in (0.19, 0.86)$ | $\mu_i \in (0.78, 0.98)$ |
|-----|----------|--------------------------|--------------------------|--------------------------|
|     |          | $\lambda \approx 12$     | $\lambda \approx 45$     | $\lambda \approx 128$     |
|     |          | ML-I Boot Bootp           | ML-I Boot Bootp           | ML-I Boot Bootp           |
| 40  | $\beta_1$ | 0.852 0.898 0.858 0.853 0.897 0.858 | 0.859 0.897 0.862 0.856 0.899 0.855 | 0.859 0.897 0.862 0.856 0.899 0.855 |
|     | $\beta_2$ | 0.859 0.902 0.868 0.858 0.896 0.862 | 0.884 0.900 0.894 0.886 0.900 0.884 | 0.884 0.900 0.894 0.886 0.900 0.884 |
|     | $\beta_3$ | 0.848 0.893 0.855 0.849 0.894 0.855 | 0.852 0.892 0.857 0.853 0.892 0.856 | 0.852 0.892 0.857 0.853 0.892 0.856 |
|     | $\beta_4$ | 0.853 0.892 0.857 0.853 0.892 0.856 | 0.857 0.896 0.882 0.886 0.900 0.884 | 0.857 0.896 0.882 0.886 0.900 0.884 |
|     | $\gamma_1$ | 0.800 0.894 0.663 0.804 0.902 0.663 | 0.808 0.905 0.658 0.808 0.905 0.658 | 0.808 0.905 0.658 0.808 0.905 0.658 |
|     | $\gamma_2$ | 0.853 0.894 0.899 0.861 0.891 0.902 | 0.863 0.892 0.897 | 0.863 0.892 0.897 |
| 80  | $\beta_1$ | 0.881 0.900 0.881 0.880 0.899 0.885 | 0.882 0.899 0.881 0.880 0.899 0.885 | 0.882 0.899 0.881 0.880 0.899 0.885 |
|     | $\beta_2$ | 0.886 0.901 0.886 0.882 0.895 0.885 | 0.884 0.903 0.884 0.882 0.900 0.884 | 0.884 0.903 0.884 0.882 0.900 0.884 |
|     | $\beta_3$ | 0.883 0.902 0.885 0.883 0.901 0.884 | 0.883 0.900 0.884 0.882 0.900 0.884 | 0.883 0.900 0.884 0.882 0.900 0.884 |
|     | $\beta_4$ | 0.880 0.895 0.880 0.881 0.898 0.882 | 0.882 0.897 0.881 0.880 0.896 0.882 | 0.882 0.897 0.881 0.880 0.896 0.882 |
|     | $\gamma_1$ | 0.852 0.896 0.768 0.847 0.892 0.770 | 0.861 0.897 0.791 | 0.861 0.897 0.791 |
|     | $\gamma_2$ | 0.879 0.899 0.899 0.881 0.887 0.902 | 0.884 0.893 0.897 | 0.884 0.893 0.897 |
| 120 | $\beta_1$ | 0.883 0.899 0.883 0.876 0.886 0.876 | 0.887 0.898 0.876 0.875 0.898 0.876 | 0.887 0.898 0.876 0.875 0.898 0.876 |
|     | $\beta_2$ | 0.890 0.900 0.889 0.900 0.896 0.892 | 0.888 0.900 0.885 0.886 0.900 0.885 | 0.888 0.900 0.885 0.886 0.900 0.885 |
|     | $\beta_3$ | 0.886 0.896 0.884 0.877 0.889 0.877 | 0.886 0.898 0.886 | 0.886 0.898 0.886 |
|     | $\beta_4$ | 0.885 0.897 0.885 0.890 0.900 0.888 | 0.889 0.899 0.890 | 0.889 0.899 0.890 |
|     | $\gamma_1$ | 0.872 0.898 0.815 0.868 0.896 0.815 | 0.875 0.900 0.827 | 0.875 0.900 0.827 |
|     | $\gamma_2$ | 0.885 0.893 0.899 0.891 0.893 0.904 | 0.884 0.890 0.892 | 0.884 0.890 0.892 |
intensity of non-constant dispersion negatively affects the performances of the three confidence intervals considered.

As the sample size increases the problem is smoothed. For instance, when \( n = 80 \), \( 1 - \alpha = 0.90 \), \( \lambda \approx 128 \), only the bootstrap \( t \) interval considers the possibility of linearity, namely: \( \text{ML-Interval: } (1.006, 1.401), \text{Boot-Interval: } (0.993, 1.410) \) and \( \text{Boot-p: } (1.010, 1.406) \). Nevertheless, if we would use only one decimal approximation those intervals would become \( \text{ML-Interval: } (1.0, 1.4), \text{Boot-Interval: } (1.0, 1.4) \) and \( \text{Boot-p: } (1.0, 1.4) \), therefore, admittedly equivalents. We should also point out that the average lengths (Size) of all intervals decrease as the sample size increases.

Our spotlight hereafter is show how the Boot interval considerably outperforms the accuracy and balance of its competitors, concerning for \( \beta_2 \) interval. We shall fix \( 1 - \alpha = 0.95 \) and consider \( n = 40 \) and three \( \lambda \)'s values. We shall compose the following set consisting of: the empirical coverage and left and right non-coverages rates of the interval estimators, expressed as \([ (\cdot), [\%], [\%] ]\). For \( \lambda \approx 12 \), the sets of the ML-Interval, Boot, and Bootp estimators are equal to \([ (0.919), [3.61\%], [4.49\%] ]\), \([ (0.950), [2.40\%], [2.61\%] ]\) and \([ (0.923), [3.20\%], [4.47\%] ]\). For \( \lambda \approx 45 \) those sets have become \([ (0.919), [2.65\%], [3.21\%] ]\), \([ (0.950), [2.55\%], [2.67\%] ]\) and \([ (0.922), [2.97\%], [3.41\%] ]\). Finally, when \( \lambda \approx 128 \) the respective sets are \([ (0.917), [3.94\%], [4.40\%] ]\), \([ (0.949), [2.66\%], [2.39\%] ]\) and \([ (0.916), [3.98\%], [4.48\%] ]\).

Table 9 present the simulation results for \( \beta_2 \) (\( \mu_t \approx 0.5 \) and \( \mu_t \approx 1 \)), when \( n = 40, \lambda \approx 128 \) and \( \beta_2 \) equal to \( -1.8 \) and \( -1.5 \), respectively. We note that for the three nominal levels and the different scenarios, the asymptotic type confidence interval has the shortest average length. We also note that the bootstrap confidence interval presents the best empirical coverage and balance proprieties, followed by the Boot interval which had very similar values to the ML-Interval.

Figs 1 and 2 contain histograms constructed from the 10000 maximum likelihood estimates of the parameter \( \beta_2 \) and \( \gamma_2 \), respectively, for \( n = 40, \lambda \approx 128 \) and the different scenarios for \( \mu_t \), \( t = 1, \ldots, n \). The distinct lines represent the different confidence intervals under evaluation, and their lengths correspond to the respective average lengths. The values below and above of the vertical lines are the non-coverage rates, meaning the percentages of replicates in which
Table 5. Coverage rates of the interval estimators: ML-I, Boot, and Bootp for \( \theta \), the model parameters: \( \text{Logit}(\mu_t/1-\mu_t) = \beta_1 + x_{1t}^\beta + \beta_2 x_{2t} + \beta_3 x_{3t}, \) and \( \log(\sigma_t^2) = \gamma_t + z_{1t}^\gamma, t = 1, \ldots, n, 1 - \alpha = 0.95 \).

\( \mu_t \in (0.02, 0.32) \)

| n  | \( \theta \) | \( \lambda \approx 12 \) | \( \lambda \approx 45 \) | \( \lambda \approx 128 \) |
|----|-------------|------------------|------------------|------------------|
|    |             | ML-I Bootp       | ML-I Bootp       | ML-I Bootp       |
| 40 | \( \beta_1 \) | 0.913 0.948 0.919 | 0.915 0.949 0.918 | 0.916 0.949 0.920 |
|    | \( \beta_2 \) | 0.919 0.950 0.923 | 0.919 0.950 0.922 | 0.917 0.949 0.916 |
|    | \( \beta_3 \) | 0.909 0.944 0.914 | 0.914 0.947 0.916 | 0.917 0.948 0.918 |
|    | \( \beta_4 \) | 0.911 0.938 0.912 | 0.914 0.942 0.914 | 0.918 0.945 0.916 |
|    | \( \gamma_1 \) | 0.872 0.945 0.744 | 0.872 0.951 0.744 | 0.876 0.951 0.745 |
|    | \( \gamma_2 \) | 0.915 0.954 0.944 | 0.921 0.939 0.950 | 0.923 0.940 0.948 |
| 80 | \( \beta_1 \) | 0.934 0.949 0.934 | 0.933 0.948 0.934 | 0.937 0.951 0.937 |
|    | \( \beta_2 \) | 0.937 0.951 0.938 | 0.935 0.946 0.937 | 0.935 0.952 0.934 |
|    | \( \beta_3 \) | 0.936 0.951 0.937 | 0.936 0.951 0.937 | 0.937 0.948 0.937 |
|    | \( \beta_4 \) | 0.933 0.947 0.934 | 0.935 0.944 0.933 | 0.937 0.948 0.936 |
|    | \( \gamma_1 \) | 0.914 0.948 0.840 | 0.907 0.944 0.838 | 0.917 0.947 0.860 |
|    | \( \gamma_2 \) | 0.936 0.942 0.950 | 0.937 0.940 0.948 | 0.941 0.942 0.946 |
| 120| \( \beta_1 \) | 0.939 0.946 0.937 | 0.931 0.941 0.932 | 0.942 0.950 0.941 |
|    | \( \beta_2 \) | 0.941 0.948 0.941 | 0.942 0.946 0.942 | 0.940 0.951 0.939 |
|    | \( \beta_3 \) | 0.938 0.947 0.936 | 0.934 0.944 0.934 | 0.942 0.950 0.941 |
|    | \( \beta_4 \) | 0.941 0.949 0.940 | 0.942 0.947 0.939 | 0.941 0.947 0.941 |
|    | \( \gamma_1 \) | 0.928 0.946 0.879 | 0.923 0.948 0.878 | 0.931 0.950 0.889 |
|    | \( \gamma_2 \) | 0.939 0.943 0.948 | 0.942 0.944 0.950 | 0.940 0.942 0.946 |

\( \mu_t \in (0.19, 0.86) \)

| n  | \( \theta \) | \( \lambda \approx 12 \) | \( \lambda \approx 45 \) | \( \lambda \approx 128 \) |
|----|-------------|------------------|------------------|------------------|
|    |             | ML-I Bootp       | ML-I Bootp       | ML-I Bootp       |
| 40 | \( \beta_1 \) | 0.913 0.948 0.918 | 0.914 0.947 0.917 | 0.916 0.950 0.917 |
|    | \( \beta_2 \) | 0.909 0.950 0.915 | 0.914 0.951 0.918 | 0.921 0.951 0.923 |
|    | \( \beta_3 \) | 0.914 0.951 0.918 | 0.913 0.948 0.914 | 0.918 0.951 0.920 |
|    | \( \beta_4 \) | 0.915 0.946 0.919 | 0.919 0.953 0.921 | 0.917 0.950 0.920 |
|    | \( \gamma_1 \) | 0.872 0.946 0.754 | 0.867 0.951 0.743 | 0.871 0.949 0.740 |
|    | \( \gamma_2 \) | 0.915 0.964 0.943 | 0.915 0.943 0.946 | 0.920 0.943 0.948 |
| 80 | \( \beta_1 \) | 0.934 0.950 0.934 | 0.934 0.950 0.935 | 0.933 0.950 0.935 |
|    | \( \beta_2 \) | 0.930 0.948 0.932 | 0.932 0.948 0.932 | 0.936 0.952 0.937 |
|    | \( \beta_3 \) | 0.930 0.948 0.931 | 0.930 0.945 0.933 | 0.937 0.950 0.936 |
|    | \( \beta_4 \) | 0.934 0.948 0.934 | 0.933 0.948 0.933 | 0.937 0.950 0.935 |
|    | \( \gamma_1 \) | 0.917 0.948 0.847 | 0.914 0.948 0.845 | 0.916 0.950 0.846 |
|    | \( \gamma_2 \) | 0.937 0.947 0.949 | 0.939 0.946 0.949 | 0.942 0.949 0.953 |
| 120| \( \beta_1 \) | 0.940 0.949 0.942 | 0.940 0.947 0.939 | 0.941 0.949 0.942 |
|    | \( \beta_2 \) | 0.937 0.947 0.938 | 0.937 0.947 0.937 | 0.939 0.950 0.939 |
|    | \( \beta_3 \) | 0.940 0.948 0.938 | 0.939 0.950 0.937 | 0.939 0.947 0.939 |
|    | \( \beta_4 \) | 0.941 0.950 0.941 | 0.941 0.950 0.941 | 0.943 0.952 0.942 |
|    | \( \gamma_1 \) | 0.927 0.951 0.882 | 0.927 0.950 0.881 | 0.930 0.948 0.880 |
|    | \( \gamma_2 \) | 0.937 0.942 0.946 | 0.939 0.945 0.947 | 0.940 0.946 0.949 |

\( \mu_t \in (0.78, 0.98) \)

(Continued)
the true value of the parameter was smaller than the lower limit of the interval (below) and larger than the upper limit of the range (above).

These graphics were designed according to [28]. Through them it is possible to verify that for the different $\mu_t$ scenarios, the analyzed intervals are approximately symmetrical around the true value of $\beta$. We further note that for $\mu_t \in (0.02, 0.32)$, the intervals were better balanced when compared to the scenarios where $\mu_t \in (0.19, 0.86)$ and $\mu_t \in (0.78, 0.98)$. Overall, the bootstrap confidence interval stands out as better balanced. Fig 2 show that only the asymptotic confidence interval is approximately symmetric around the true value's $\gamma$. The bootstrap confidence interval is slightly asymmetric around $\gamma$. Regarding the bootstrap percentile confidence interval it exhibits very strong asymmetry, especially for the nominal 99% level. We also observe that the asymptotic confidence interval exhibits strong unbalancing, as the rates (% Right) are markedly higher than the observed rates (% Left). However, the bootstrap, and percentile confidence intervals are approximately balanced for all nominal levels and scenarios. Therefore, based on the results presented, we suggest using the bootstrap confidence interval which showed better coverage and balance performances.

7 Application: Fluid Catalytic Cracking Data (FCC)

In this application the data are from the Chemistry Department of the National University of Colombia [29] and concerns a process regarding the volume and quality of gasoline produced in a refinery. The fluid catalytic cracking process known as Fluid Catalytic Cracking (FCC) is used to convert high molecular weight hydrocarbons into small molecules of higher commercial value by contacting them with a catalyst. This process is often described as the heart of the refinery, as it allows production to be tailored for a higher demand and especially high profit products [29]. The process catalyst consists of fine particles of 10 to 150 microns, easily fluidizable having the zeolite $Y$ [29] as the main component. Another important substance that participates in the catalysis process is the vanadium. This chemical component is known to participate in catalyst destruction, reducing the active surface, selectivity and crystallinity of
the zeolite Y especially in the presence of steam. Every 1000 ppm of vanadium in the catalyst is known to reduce gasoline yield by about 2.3%. The process also depends on the temperature, which must be close to 720°C [29]. The data set consists of 28 observations.

Aiming to fit a model to these data [13] chose the following candidate to covariates: steam ($x_2$), temperature ($x_3$) and vanadium concentration ($x_4$). Moreover, the authors defined a linear predictor relating these covariates to unknown parameters. However, the residual analysis highlighted the possibility that the predictor is non-linear in some of the parameters. To build the nonlinear model, the authors follow several steps that are carefully detailed in their article. The model chosen uses probit and logarithmic link functions for the mean and dispersion submodels, respectively, and was defined as follows:

$$F_1(m_t) = \beta_1 + \beta_2 x_2^t + \beta_3 x_3^t + \beta_4 x_4^t$$ and $$\log(\sigma_i^2) = \gamma_1 + \gamma_2 x_4^t$$ with $t = 1, \ldots, 28$. 

Table 6. Lower and upper bounds, size, empirical coverage (Coverage) and percentages of Lower (%Left) and upper (%Right) non-coverage of the ML-1, Boot, and Bootp intervals for $\beta_2$, in the model: Logit($\mu_t/1-\mu_t$) = $\beta_1 + \beta_2 x_2^t + \beta_3 x_3^t + \beta_4 x_4^t$ and $\log(\sigma_i^2) = \gamma_1 + \gamma_2 x_4^t$, $t = 1, \ldots, n$. $\mu_t \in (0.02, 0.32), \beta_2 = 1.2, n = 40.$

| $\alpha$ | Intervals | Lower | Upper | Size | Coverage | %Left | %Right |
|----------|-----------|-------|-------|------|----------|-------|--------|
| 10%      | ML-1      | 1.153 | 1.644 | 0.490| 0.859    | 6.61  | 7.46   |
|          | Boot      | 1.127 | 1.683 | 0.556| 0.902    | 4.74  | 5.10   |
|          | Bootp     | 1.147 | 1.644 | 0.497| 0.868    | 5.97  | 7.26   |
| 5%       | ML-1      | 1.107 | 1.691 | 0.584| 0.919    | 3.61  | 4.49   |
|          | Boot      | 1.074 | 1.744 | 0.670| 0.950    | 2.40  | 2.61   |
|          | Bootp     | 1.097 | 1.689 | 0.592| 0.923    | 3.20  | 4.47   |
| 1%       | ML-1      | 1.015 | 1.783 | 0.768| 0.976    | 0.83  | 1.60   |
|          | Boot      | 0.964 | 1.869 | 0.905| 0.991    | 0.34  | 0.57   |
|          | Bootp     | 1.002 | 1.777 | 0.775| 0.976    | 0.83  | 1.56   |

| $\alpha$ | Intervals | Lower | Upper | Size | Coverage | %Left | %Right |
|----------|-----------|-------|-------|------|----------|-------|--------|
| 10%      | ML-1      | 1.154 | 1.657 | 0.503| 0.858    | 7.23  | 6.95   |
|          | Boot      | 1.127 | 1.692 | 0.565| 0.896    | 5.63  | 4.75   |
|          | Bootp     | 1.152 | 1.657 | 0.506| 0.862    | 7.00  | 7.30   |
| 5%       | ML-1      | 1.106 | 1.705 | 0.599| 0.919    | 4.10  | 5.03   |
|          | Boot      | 1.074 | 1.755 | 0.681| 0.950    | 2.78  | 2.24   |
|          | Bootp     | 1.101 | 1.702 | 0.602| 0.922    | 3.77  | 4.07   |
| 1%       | ML-1      | 1.012 | 1.799 | 0.788| 0.977    | 1.10  | 1.23   |
|          | Boot      | 0.962 | 1.883 | 0.922| 0.988    | 0.70  | 0.50   |
|          | Bootp     | 1.004 | 1.792 | 0.789| 0.977    | 1.01  | 1.33   |

| $\alpha$ | Intervals | Lower | Upper | Size | Coverage | %Left | %Right |
|----------|-----------|-------|-------|------|----------|-------|--------|
| 10%      | ML-1      | 0.938 | 1.471 | 0.533| 0.856    | 7.01  | 7.41   |
|          | Boot      | 0.907 | 1.508 | 0.600| 0.899    | 5.21  | 4.92   |
|          | Bootp     | 0.937 | 1.474 | 0.538| 0.855    | 6.20  | 7.25   |
| 5%       | ML-1      | 0.887 | 1.522 | 0.635| 0.917    | 3.94  | 4.40   |
|          | Boot      | 0.850 | 1.575 | 0.725| 0.949    | 2.66  | 2.39   |
|          | Bootp     | 0.883 | 1.522 | 0.639| 0.916    | 3.98  | 4.43   |
| 1%       | ML-1      | 0.787 | 1.621 | 0.834| 0.973    | 1.10  | 1.58   |
|          | Boot      | 0.731 | 1.712 | 0.981| 0.988    | 0.61  | 0.54   |
|          | Bootp     | 0.781 | 1.617 | 0.836| 0.973    | 1.09  | 1.58   |

https://doi.org/10.1371/journal.pone.0272512.t006
Hereafter we shall so-called this model as ‘Model-I’. We emphasize that this simplex model outperformed a competing beta model [13]. Table 10 displays the maximum likelihood estimates, the bootstrap bias-corrected estimates, their respective standard errors (SE), and the p-values associated with the Z-tests for the significance of the model parameters. The ML estimates and their bootstrap corrected versions are quite similar with regards to the parameters of the mean submodel. Whereas the corrected estimates have lower standard errors than the ML estimates, with the exception only for the parameter $\beta_1$. Concerning the parameters of the dispersion submodel, we notice that the maximum likelihood estimates and their corrected version present slightly different values. Additionally, their respective standard errors are quite similar.
Table 8. Lower and upper bounds, size, empirical coverage (Coverage) and percentages of lower (%Left) and upper (%Right) non-coverage of the ML-I, Boot, and Bootp intervals for \( \beta_2 \), in the model: \[ \text{Logit}(\mu_t/1-\mu_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} \text{ and } \log(\sigma^2_t) = \gamma_1 + \gamma_2 t, \] \( t = 1, \ldots, n \), \( \mu_t \in (0.02, 0.32) \), \( \beta_2 = 1.2, n = 120 \).

| \( \lambda \approx 12 \) | \( \alpha \) | Intervals | Lower | Upper | Size | Coverage | %Left | %Right |
|----------------|---------|-----------|-------|-------|------|---------|-------|-------|
| 10%            | ML-I    | 1.249     | 1.549 | 0.300 | 0.890 | 5.21    | 5.80  |
|                | Boot    | 1.246     | 1.557 | 0.311 | 0.900 | 4.90    | 5.13  |
|                | Bootp   | 1.248     | 1.549 | 0.300 | 0.889 | 5.18    | 5.90  |
| 5%             | ML-I    | 1.221     | 1.578 | 0.357 | 0.941 | 2.65    | 3.21  |
|                | Boot    | 1.217     | 1.590 | 0.372 | 0.948 | 2.55    | 2.67  |
|                | Bootp   | 1.218     | 1.576 | 0.358 | 0.940 | 2.55    | 3.41  |
| 1%             | ML-I    | 1.164     | 1.634 | 0.469 | 0.986 | 0.54    | 0.83  |
|                | Boot    | 1.160     | 1.652 | 0.492 | 0.989 | 0.51    | 0.62  |
|                | Bootp   | 1.161     | 1.630 | 0.469 | 0.986 | 0.42    | 0.99  |

| \( \lambda \approx 45 \) | \( \alpha \) | Intervals | Lower | Upper | Size | Coverage | %Left | %Right |
|----------------|---------|-----------|-------|-------|------|---------|-------|-------|
| 10%            | ML-I    | 1.247     | 1.554 | 0.307 | 0.890 | 5.30    | 5.71  |
|                | Boot    | 1.242     | 1.559 | 0.316 | 0.896 | 4.89    | 5.55  |
|                | Bootp   | 1.249     | 1.555 | 0.306 | 0.892 | 5.19    | 5.60  |
| 5%             | ML-I    | 1.218     | 1.584 | 0.366 | 0.942 | 2.67    | 3.11  |
|                | Boot    | 1.213     | 1.592 | 0.378 | 0.946 | 2.64    | 2.73  |
|                | Bootp   | 1.219     | 1.583 | 0.364 | 0.942 | 2.64    | 3.18  |
| 1%             | ML-I    | 1.161     | 1.641 | 0.481 | 0.986 | 0.56    | 0.79  |
|                | Boot    | 1.155     | 1.655 | 0.501 | 0.988 | 0.62    | 0.56  |
|                | Bootp   | 1.160     | 1.638 | 0.477 | 0.985 | 0.58    | 0.94  |

| \( \lambda \approx 128 \) | \( \alpha \) | Intervals | Lower | Upper | Size | Coverage | %Left | %Right |
|----------------|---------|-----------|-------|-------|------|---------|-------|-------|
| 10%            | ML-I    | 1.044     | 1.370 | 0.326 | 0.888 | 5.86    | 5.37  |
|                | Boot    | 1.035     | 1.372 | 0.337 | 0.900 | 5.13    | 4.89  |
|                | Bootp   | 1.049     | 1.376 | 0.327 | 0.885 | 6.63    | 4.83  |
| 5%             | ML-I    | 1.013     | 1.402 | 0.389 | 0.940 | 3.19    | 2.79  |
|                | Boot    | 1.004     | 1.408 | 0.404 | 0.951 | 2.77    | 2.10  |
|                | Bootp   | 1.016     | 1.405 | 0.389 | 0.939 | 3.57    | 2.51  |
| 1%             | ML-I    | 0.952     | 1.463 | 0.511 | 0.986 | 0.64    | 0.72  |
|                | Boot    | 0.942     | 1.476 | 0.534 | 0.988 | 0.65    | 0.51  |
|                | Bootp   | 0.954     | 1.464 | 0.510 | 0.984 | 0.82    | 0.73  |

In Table 11 are reported the interval estimates of the Model-I parameters assuming nominal levels equal to 90%, 95% and 99%. Mindful the three estimation methods, ML-I, Boot, and Bootp it is notice that the interval estimates for \( \beta_1, \beta_2, \beta_4 \) and \( \beta_5 \) are quite similar. Whereas for the \( \beta_3 \) parameter the bootstrap scheme estimates display lengths substantially longer than that of its competitors, for all nominal levels. Exemplify, for \( 1 - \alpha = 0.95 \), the ML-I, Boot, and Bootp interval estimates are, respectively, \((-35.445; -20.240), (-40.929; -9.544) \) and \((-34.745; -20.474) \). Another feature concerning the bootstrap interval estimator is that some of its estimates include the value zero for the parameters. This fact occurs for the \( \beta_2 \) (99%), \((-0.466; 0.142) \) and for the \( \beta_4 \) (95%) and (99%). Nevertheless, \( \beta_2 = 0 \) implies both in the exclusion of steam, which is a covariate recognized as important to the process and in the assumption of a linear predictor for the mean submodel. The most important information that the
Table 9. Lower and upper bounds, size, empirical coverage (Coverage) and percentages of lower (%Left) and upper (%Right) non-coverage of the ML-I, Boot, and Bootp intervals. For $\beta_2$ in the model: Logit($\mu$/$1-\mu$) = $\beta_1 + \beta_2 x_2^1 + \beta_3 x_3 + \beta_4 x_4$ and log($\sigma^2_\tau$) = $\gamma_1^t + \gamma_2^t$, $t = 1, \ldots, n, n = 40, \lambda = 128$.

| $\alpha$ | Intervals | Lower | Upper | Size | Coverage | %Left | %Right |
|----------|-----------|-------|-------|------|----------|-------|--------|
| 10%      | ML-I     | −1.994 | −1.609 | 0.385 | 0.861     | 5.02  | 8.83   |
|          | Bootp    | −1.998 | −1.560 | 0.439 | 0.906     | 4.57  | 4.80   |
| 5%       | ML-I     | −2.031 | −1.573 | 0.459 | 0.921     | 2.46  | 5.47   |
|          | Bootp    | −2.035 | −1.505 | 0.530 | 0.951     | 2.27  | 2.58   |
| 1%       | ML-I     | −2.103 | −1.500 | 0.603 | 0.971     | 0.49  | 2.41   |
|          | Bootp    | −2.111 | −1.392 | 0.718 | 0.989     | 0.48  | 0.63   |
|          | Bootp    | −2.095 | −1.480 | 0.615 | 0.973     | 0.65  | 2.03   |

$\mu \in (0.19, 0.86), \beta_2 = -1.8$.

| $\alpha$ | Intervals | Lower | Upper | Size | Coverage | %Left | %Right |
|----------|-----------|-------|-------|------|----------|-------|--------|
| 10%      | ML-I     | −1.614 | −1.384 | 0.230 | 0.859     | 6.13  | 8.00   |
|          | Bootp    | −1.624 | −1.364 | 0.261 | 0.901     | 4.60  | 5.32   |
| 5%       | ML-I     | −1.636 | −1.362 | 0.274 | 0.917     | 3.27  | 5.00   |
|          | Bootp    | −1.648 | −1.334 | 0.314 | 0.947     | 2.47  | 2.78   |
| 1%       | ML-I     | −1.634 | −1.356 | 0.279 | 0.918     | 3.68  | 4.54   |
|          | Bootp    | −1.679 | −1.319 | 0.361 | 0.972     | 1.01  | 1.82   |
|          | Bootp    | −1.698 | −1.274 | 0.424 | 0.987     | 0.60  | 0.65   |
|          | Bootp    | −1.675 | −1.309 | 0.366 | 0.973     | 1.19  | 1.55   |

$\mu \in (0.78, 0.98), \beta_2 = -1.5$.

https://doi.org/10.1371/journal.pone.0272512.t009

...figures in Table 11 reveal, though, is that only the bootstrap interval considers the possibility that both $\gamma_1$ and $\gamma_2$ are simultaneously equal to zero, both to 95% and 99%. Bootp reaches this conclusion for the 99% level, whereas for ML-I estimator it is only possible that $\gamma_1 = 0$ and when $1 - \alpha = 0.99$.

Therefore, we shall evaluate a nonlinear simplex model with constant dispersion. Among the competing models, the one that presented the best goodness-of-fit uses log-log complementary and logarithmic link functions for the mean and dispersion submodel, as we shall describe in the following: log($-\log(1.0 - \mu_{i})$) = $\beta_1 + \beta_2 \frac{x_2^1}{\exp(\beta_3)} + \beta_4 x_3 + \beta_5 \sqrt{x_4}$ and log($\sigma^2_\tau$) = $\gamma_1^t$, with $t = 1, \ldots, 28$.

In what follows we shall appoint this model as ‘Model-II’ and provide some quantities about its parameters, namely: the maximum likelihood estimates, their bootstrap corrected versions: (·) and the standard errors of the estimates ML: ‘[-]’. Thus, $\beta_2$: 0.9112 (0.9051) [0.1081], $\beta_2$: −1.4166 (−1.5062) [0.0016], $\beta_3$: −26.1208 (−25.7744) [3.6902], $\beta_4$: −0.2615 (−0.2571) [0.0635], $\beta_5$: −0.3366 (−0.3431) [0.0607] and $\gamma_1$: 0.0455 (0.2782) [0.2673]. Concerning Model-II it is possible to notice differences between the ML estimates and their bootstrap corrected versions. For example, $\beta_2$ is equal to −1.42 while its corrected version becomes −1.51. A further important issue regarding Model-II is that the correct modeling of the dispersion, considerably reduced the standard errors of the estimates. The SE of $\beta_2$ was 0.023 for Model-I whereas becomes 0.0016 for Model-II. Table 12 report the interval estimates of the ‘Model-II’ model parameters assuming nominal levels equal to 90%, 95% and 99%. It is noteworthy that the correct dispersion modeling improves the interval estimators performances. The accuracy of the boot, and ML-Ia interval estimators regarding to the $\beta_2$ parameter is
Fig 1. Interval estimation for $\beta_2$, $n = 40$, $\lambda \approx 128$. (a) $\beta_2 = 1.2$, (b) $\beta_2 = -1.8$ and (c) $\beta_2 = -1.5$.

https://doi.org/10.1371/journal.pone.0272512.g001
Fig 2. Interval estimation for $\gamma_2$, $n = 40$. $\gamma_2 = -2.4$.

https://doi.org/10.1371/journal.pone.0272512.g002
especially noteworthy. Let interval estimators be ML-\( \hat{\omega} \), Boot, and Bootp, respectively, the interval estimates are \((-1.419; -1.413), (-1.419; -1.414)\) and \((-1.904; -0.757)\), for 99%. For 95%, \((-1.419; -1.413), (-1.419; -1.413)\) and \((-2.032; -0.631)\). Finally, \((-1.420; -1.412), (-1.4212; -1.4116)\) and \((-2.3049; -0.3229)\), for 90% Here, we note that the Bootp displays a poor performance. It should be reminded that \(\beta_2\) is the parameter associated with the nonlinearity of the model, as well as \(\beta_1\). In fact, after dispersion was assumed constant the Boot,

Table 10. Maximum likelihood estimates (\(\hat{\theta}\)), bootstrap bias-corrected estimates (\(\hat{\theta}\)), standard errors (SE) and \(p\)-values associated with Z-tests for the parameters of the [13]. Fluid Catalytic Cracking (FCC).

| Parameters | \(\beta_1\) | \(\beta_2\) | \(\beta_3\) | \(\beta_4\) | \(\beta_5\) | \(\gamma_1\) | \(\gamma_2\) |
|------------|------------|------------|------------|------------|------------|------------|------------|
| \(\hat{\theta}\) | 1.399 | -0.061 | -27.843 | -0.182 | -0.420 | 0.756 | -1.150 |
| \(\hat{\theta}\) | 1.411 | -0.069 | -27.782 | -0.180 | -0.428 | 0.898 | -0.944 |
| \(EP(\hat{\theta})\) | 0.077 | 0.023 | 3.879 | 0.063 | 0.075 | 0.370 | 0.314 |
| \(EP(\hat{\theta})\) | 0.082 | 0.018 | 3.185 | 0.044 | 0.066 | 0.370 | 0.314 |
| \(P\)-value | <0.0001 | 0.007 | <0.0001 | 0.004 | <0.0001 | 0.041 | 0.0003 |

https://doi.org/10.1371/journal.pone.0272512.t010

Table 11. ML-\( \hat{\omega} \), Boot, and Bootp interval estimates for the parameters of the Model 'Model-1'. Fluid catalytic cracking (FCC) data.

90%

| Parameters | ML-\( \hat{\omega} \) | Boot\( \omega \) | Bootp |
|------------|------------|------------|--------|
| \(\beta_1\) | (1.273; 1.526) | (1.235; 1.547) | (1.250; 1.531) |
| \(\beta_2\) | (-0.098; -0.023) | (-0.117; -0.020) | (-0.086; -0.011) |
| \(\beta_3\) | (-34.223; -21.463) | (-38.522; -15.849) | (-34.253; -21.156) |
| \(\beta_4\) | (-0.285; -0.079) | (-0.346; -0.003) | (-0.308; -0.069) |
| \(\beta_5\) | (-0.146; -0.297) | (-0.554; -0.268) | (-0.544; -0.284) |
| \(\gamma_1\) | (0.148; 1.364) | (0.111; 1.790) | (-0.278; 1.401) |
| \(\gamma_2\) | (-1.667; -0.633) | (-1.756; -0.191) | (-2.110; -0.545) |

95%

| Parameters | ML-\( \hat{\omega} \) | Boot\( \omega \) | Bootp |
|------------|------------|------------|--------|
| \(\beta_1\) | (1.249; 1.550) | (1.187; 1.580) | (1.186; 1.567) |
| \(\beta_2\) | (-0.105; -0.016) | (-0.146; -0.008) | (-0.094; -0.004) |
| \(\beta_3\) | (-35.445; -20.240) | (-40.929; -9.544) | (-34.745; -20.474) |
| \(\beta_4\) | (-0.305; -0.059) | (-0.382; 0.061) | (-0.331; 0.038) |
| \(\beta_5\) | (-0.567; -0.274) | (-0.579; -0.218) | (-0.566; -0.261) |
| \(\gamma_1\) | (0.031; 1.481) | (-0.042; 1.992) | (-0.480; 1.554) |
| \(\gamma_2\) | (-1.766; -0.534) | (-1.986; 0.001) | (-2.301; 0.315) |

99%

| Parameters | ML-\( \hat{\omega} \) | Boot\( \omega \) | Bootp |
|------------|------------|------------|--------|
| \(\beta_1\) | (1.201; 1.597) | (1.093; 1.613) | (1.142; 1.607) |
| \(\beta_2\) | (-0.119; -0.002) | (-0.226; 0.018) | (-0.104; 0.010) |
| \(\beta_3\) | (-37.834; -17.852) | (-46.445; -6.613) | (-37.023; 18.382) |
| \(\beta_4\) | (-0.344; -0.020) | (-0.466; 0.148) | (-0.390; 0.013) |
| \(\beta_5\) | (-0.613; -0.227) | (-0.660; -0.173) | (-0.620; -0.219) |
| \(\gamma_1\) | (-0.196; 1.709) | (-0.244; 2.213) | (-0.701; 1.756) |
| \(\gamma_2\) | (-1.960; -0.341) | (-2.316; 0.338) | (-2.639; 0.015) |

https://doi.org/10.1371/journal.pone.0272512.t011
scheme provided intervals for $\beta_3$ with considerably shorter lengths compared to those of the ‘Model-I’ model.

One area of research that we have been working on intensively regards model selection criteria for nonlinear models. The $R^2_{FC}$ criterion proposed by [7] for the beta regression model, defined as the square of the correlation between $g(y)$ and $\hat{g}$, has proven quite effective in assessing the goodness-of-fit of models to data in the different applications we have performed on nonlinear models. The $R^2_{FC}$ correct was proposed by [30] and is defined as:

$$R^2_{FC} = 1 - (1 - R^2_{FC})(n - 1)/(n - (k + q))$$

Models I and II display $R^2_{FC}$; $R^2_{FC}$ measures equal to {0.6506, 0.5508} and {0.6818, 0.6095}, respectively. Thus, the choice of Model II is adequate and this model was inferred based on the bootstrap interval estimator.

### 8 Conclusion

In this paper we evaluate the point and interval estimation for the parameters indexing the nonlinear simplex regression model [13] in small samples. Additionally, we propose inferential improvements based on the bootstrap method.

Often MLEs can be biased when the sample size is small or even moderate. Thus, we consider comparing the point MLE performances of the model parameters and their corrected versions through a bootstrap scheme. The results of Monte Carlo simulations showed that, in general, the corrected estimators presented lower biases than the maximum likelihood estimators, evidencing the efficacy of the bootstrap scheme in bias correction. The MLEs of the parameters of the dispersion submodel are strongly biased, and the bootstrap corrected estimator provides a substantial reduction of these bias. Thus reinforcing the importance of using

---

**Table 12. ML-I, Boot, and Bootp interval estimates for the parameters of Model-II.** Fluid catalytic cracking (FCC) data.

| Parameters | ML-I | Boot | Bootp |
|------------|------|------|-------|
| $\beta_1$  | (0.733;1.089) | (0.678;1.067) | (0.836;1.001) |
| $\beta_2$  | (−1.419; −1.413) | (−1.419; −1.414) | (−1.904; −0.757) |
| $\beta_3$  | (−32.190; −20.050) | (31.595; −18.054) | (−31.100; −22.146) |
| $\beta_4$  | (−0.365; −0.157) | (−0.383; −0.136) | (−0.376; −0.160) |
| $\beta_5$  | (−0.436; −0.236) | (−0.473; −0.221) | (−0.440; −0.216) |

### 90% Confidence Intervals

| Parameters | ML-I | Boot | Bootp |
|------------|------|------|-------|
| $\beta_1$  | (0.699;1.123) | (0.641;1.113) | (0.817;1.017) |
| $\beta_2$  | (−1.419; −1.413) | (−1.419; −1.413) | (−2.032; −0.631) |
| $\beta_3$  | (−33.353; −18.888) | (−33.045; −15.755) | (−32.268; −21.152) |
| $\beta_4$  | (−0.385; −0.137) | (−0.410; −0.103) | (−0.410; −0.140) |
| $\beta_5$  | (−0.455; −0.217) | (−0.495; −0.197) | (−0.462; −0.200) |

### 95% Confidence Intervals

| Parameters | ML-I | Boot | Bootp |
|------------|------|------|-------|
| $\beta_1$  | (0.632;1.189) | (0.4698;1.1882) | (0.7729;1.0600) |
| $\beta_2$  | (−1.420; −1.412) | (−1.4212; −1.4116) | (−2.3049; −0.3229) |
| $\beta_3$  | (−35.626; −16.615) | (−35.5726; −10.3756) | (−33.4200; −20.0357) |
| $\beta_4$  | (−0.425; −0.098) | (−0.4488; −0.0422) | (−0.4652; −0.1085) |
| $\beta_5$  | (−0.492; −0.180) | (−0.5163; −0.1802) | (−0.4823; −0.1633) |

---

https://doi.org/10.1371/journal.pone.0272512.t012
the proposed scheme in the bias correction of estimators of the nonlinear simplex regression model.

Usually the asymptotic confidence intervals based on MLE’s require large samples in order the coverage rates to be close to the nominal ones. An alternative to constructing adequate confidence intervals on small samples is through the bootstrap method. Thus, we consider three competing interval estimators, namely: the MLE-asymptotic, percentile and bootstrap, estimator intervals. Regarding coverage rate in every simulation’s scenarios the bootstrap, confidence interval outperformed the two others competitors. Furthermore, in almost all experiments it was the best balanced interval.

As a penalty for providing this outperformance, the bootstrap, is typically larger than that of its competitors. Overall, however, the bootstrap interval proved to be the most appropriate estimator interval for nonlinear simplex regression. Not only from the simulation results, but it was also decisive in the application. In a scenario with only \( n = 28 \) observations, it was able to point the misspecification of the dispersion model which yield to a new and best fitted model.

Supporting information

S1 File.
(OX)

S1 Text.
(TXT)

Author Contributions

Conceptualization: Alisson de Oliveira Silva, Patrícia L. Espinheira.

Data curation: Patrícia L. Espinheira.

Formal analysis: Alisson de Oliveira Silva, Jonas Weverson de Ararújo Silva, Patrícia L. Espinheira.

Investigation: Alisson de Oliveira Silva, Jonas Weverson de Ararújo Silva, Patrícia L. Espinheira.

Methodology: Alisson de Oliveira Silva, Jonas Weverson de Ararújo Silva, Patrícia L. Espinheira.

Project administration: Alisson de Oliveira Silva.

Resources: Alisson de Oliveira Silva, Jonas Weverson de Ararújo Silva, Patrícia L. Espinheira.

Software: Alisson de Oliveira Silva, Jonas Weverson de Ararújo Silva, Patrícia L. Espinheira.

Supervision: Alisson de Oliveira Silva, Patrícia L. Espinheira.

Validation: Alisson de Oliveira Silva, Jonas Weverson de Ararújo Silva, Patrícia L. Espinheira.

Visualization: Alisson de Oliveira Silva, Jonas Weverson de Ararújo Silva, Patrícia L. Espinheira.

Writing – original draft: Alisson de Oliveira Silva.

Writing – review & editing: Alisson de Oliveira Silva, Jonas Weverson de Ararújo Silva, Patrícia L. Espinheira.
References

1. Barndorff-Nielsen OE, Jørgensen B. Some parametric models on the simplex. Journal of Multivariate Analysis. 1991; 39(1):106–116. https://doi.org/10.1016/0047-259X(91)90008-P

2. Jørgensen B. The theory of dispersion models. London: Chapman and Hall; 1997.

3. McCullagh P, Nelder JA. Generalized linear models. London: Chapman and Hall; 1989.

4. Song PXK, Tan M. Marginal models for longitudinal continuous proportional data. Biometrics. 2000; 56(2):496–502. https://doi.org/10.1111/j.0006-341X.2000.00496.x PMID: 10877309

5. Song PXK, Qiu Z, Tan M. Modelling heterogeneous dispersion in marginal models for longitudinal proportional data. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 2004; 46(5):540–553. https://doi.org/10.1002/bimj.200110052

6. López FO. A bayesian approach to parameter estimation in simplex regression model: a comparison with beta regression. Revista Colombiana de Estadística. 2013; 36(1):1–21.

7. Ferrari SLP, Cribari-Neto F. Beta regression for modelling rates and proportions. Journal of Applied Statistics. 2004; 31(7):799–815. https://doi.org/10.1080/0266476042000214501

8. Mitnik PA, Baek S. The Kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. Statistical Papers. 2013; 54(1):177–192. https://doi.org/10.1007/s00362-011-0417-y

9. Guedes AC, Cribari-Neto F, Espinheira PL. Modified likelihood ratio tests for unit gamma regressions. Journal of Applied Statistics. 2020; p. 1–25. https://doi.org/10.1080/02664763.2019.1683152 PMID: 35707584

10. Mousa AM, El-Sheikh AA, Abdel-Fattah MA. A gamma regression for bounded continuous variables. Advances and Applications in Statistics. 2016; 49(4):305–326. https://doi.org/10.17654/AS049040305

11. Simas AB, Barreto-Souza W, Rocha AV. Improved estimators for a general class of beta regression models. Computational Statistics & Data Analysis. 2010; 54(2):348–366. https://doi.org/10.1016/j.csda.2009.08.017

12. Lima FP, Cribari-Neto F. Bootstrap-based inferential improvements to the simplex nonlinear regression model. PLOS ONE | https://doi.org/10.1371/journal.pone.0272512 August 9, 2022 26 / 27
26. Ferrari SL, Cribari-Neto F. On bootstrap and analytical bias corrections. Economics Letters. 1998; 58(1):7–15. https://doi.org/10.1016/S0165-1765(97)00276-0

27. Hall P. Theoretical comparison of bootstrap confidence intervals. The Annals of Statistics. 1988; p. 927–953.

28. Ospina R, Cribari-Neto F, Vasconcellos KL. Improved point and interval estimation for a beta regression model. Computational Statistics & Data Analysis. 2006; 51(2):960–981. https://doi.org/10.1016/j.csda.2005.10.002

29. Salazar SMG. Contribución al estudio de la reacción de decomposición de la Zeolita Y en presencia de vapor de agua y vanadio. Universidad Nacional de Colombia; 2005.

30. Bayer FM, Cribari-Neto F. Model selection criteria in beta regression with varying dispersion. Communications in Statistics-Simulation and Computation. 2017; 46(1):729–746. https://doi.org/10.1080/03610918.2014.977918