

RESEARCH ARTICLE

Process monitoring using inflated beta regression control chart

Luiz M. A. Lima-Filho^{1*}, Tarciana Liberal Pereira¹, Tatiene C. Souza¹, Fábio M. Bayer²

1 Departamento de Estatística, Universidade Federal da Paraíba, João Pessoa, Brazil, **2** Departamento de Estatística and LACESM, Universidade Federal de Santa Maria, Santa Maria, Brazil

* luizmalf@gmail.com

Abstract

This paper provides a general framework for controlling quality characteristics related to control variables and limited to the intervals (0, 1], [0, 1), or [0, 1]. The proposed control chart is based on the inflated beta regression model considering a reparametrization of the inflated beta distribution indexed by the response mean, which is useful for modeling fractions and proportions. The contribution of the paper is twofold. First, we extend the inflated beta regression model by allowing a regression structure for the precision parameter. We also present closed-form expressions for the score vector and Fisher's information matrix. Second, based on the proposed regression model, we introduce a new model-based control chart. The control limits are obtained considering the estimates of the inflated beta regression model parameters. We conduct a Monte Carlo simulation study to evaluate the performance of the proposed regression model estimators, and the performance of the proposed control chart is evaluated in terms of run length distribution. Finally, we present and discuss an empirical application to show the applicability of the proposed regression control chart.



OPEN ACCESS

Citation: Lima-Filho LMA, Pereira TL, Souza TC, Bayer FM (2020) Process monitoring using inflated beta regression control chart. PLoS ONE 15(7): e0236756. <https://doi.org/10.1371/journal.pone.0236756>

Editor: Feng Chen, Tongji University, CHINA

Received: April 8, 2020

Accepted: July 11, 2020

Published: July 30, 2020

Copyright: © 2020 Lima-Filho et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

1 Introduction

Standard control charts are directly applied to the output of a quality characteristic. However, the quality characteristic (process output) can be affected by external covariates (control variables), where we rather control a varying mean than a constant one. In these cases, the regression control chart [1] may be an effective statistical process control tool. Such method is widely used when the quality of a process or product is better characterized by a functional relationship between the response variable and one or more explanatory variables [2].

The standard regression control chart is based on the linear regression model, where the variable of interest is assumed to be normally distributed. However, in practice, several of these variables may not follow a normal distribution, leading to poor Gaussian-based inferences. Thus, several studies have been proposing non-Gaussian model-based control charts. [3] presented a model-based scheme for monitoring multiple gamma-distributed variables. By considering that robust methods can be effective in the presence of outlying observations, [4] explored the robust generalized linear model for a gamma-distributed response. [5] used

deviance residual for monitoring variables in a three-stage process assuming gamma, normal, and Poisson distributions.

Examples of a non-Gaussian process output are variables that assume values in the standard unit interval, such as fractions and proportions. In such instances, the usual regression control chart may be inappropriate since double bounded data are typically asymmetric and the Gaussian-based assumption is not suitable. In this sense, [6] proposed the beta regression control chart to monitor fractions and proportions related to control variables. The proposed control chart considers the beta regression model with varying dispersion [7], assuming that the mean and dispersion parameters of beta distributed variables are related to exogenous variables and modeled by regression structures. However, fractions and proportions may contain zeros and/or ones, leading to the unsuitable use of the beta distribution for data modeling [8].

Alternative regression models have been proposed to mend beta regression flaws in the presence of zeros and/or ones. [9] presented a unit inflated beta model for modeling efficiency scores as a function of exogenous variables. [10] proposed a zero inflated beta model to analyze data in corporate capital structures. [8] introduced a general class of zero or one inflated beta regression models, which is a natural extension of the beta regression model [11] to model variables that assume values in $(0, 1)$ or $[0, 1)$. [12] proposed an inflated beta regression model based on a reparametrization of the inflated beta distribution. This model accommodates mixed random variable responses, with non-negligible probabilities of assuming zeros and/or ones and continuous values in the interval $(0, 1)$ that follows a beta distribution. The inflated beta regression model introduced by [12] may be useful for developing model-based control charts for monitoring inflated beta distributed processes as it considers an interesting parametrization in terms of the response variable mean. However, the model proposed by [12] does not consider a regression structure for the precision parameter. The monitoring of the mean and precision (or dispersion) is relevant to the statistical process control [6, 13, 14]. In addition, incorrect modeling of the dispersion can generate a high number of false alarms or loss of detection power of special causes [6]. Moreover, dispersion modeling is necessary in regression models in order to obtain accurate inferences about the structure parameters of the mean regression [15].

Control chart is a dynamic tool that works under two different phases, namely Phase I and Phase II. In practical situations, the in-control parameters are unknown and have to be estimated from a Phase I data set. Different Phase I data sets lead to different control chart performance. Thus, it is important to study the practitioner-to-practitioner variability due to parameter estimation. The aim of Phase I analysis is to estimate the parameters, while quick detection of out-of-control state is conducted in Phase II [16]. The literature offers some studies related to Phase I and Phase II analyses in regression models. For example, [17] proposed Phase I profile monitoring schemes for binary responses that can be represented by logistic regression models. [17] developed several Hotelling T^2 -type Phase I control charts for monitoring the parameters of a logistic regression linking to a binary response and one or more predictor variables. [18] developed control charts by integrating an exponentially weighted moving average scheme with a likelihood ratio test based on logistic regression models in Phase II study. [19] proposed a new modeling and monitoring framework for Phase I analysis of multivariate profiles by incorporating the regression-adjustment technique into the functional principal components analysis. [20] proposed the monitoring of profiles using generalized linear models during Phase II in which the explanatory variables can be a fixed design or any random arbitrary design.

In this context, this paper introduces the inflated beta regression control chart (IBRCC) with varying dispersion, useful for monitoring double bounded variables when zeros or ones appear in the data along with the presence of control variables. The process output may

represent individual measures (e.g. efficiency score) or a ratio between continuous numbers (e.g. relative humidity). The contribution of the present paper is twofold. First, we extend the inflated beta regression model proposed by [12] by allowing a regression structure for the precision parameter. We also discuss likelihood inference of the model parameters. Second, we introduce the IBRCC based on the proposed inflated beta regression model with varying dispersion. Since in practice the parameters of the regression model are unknown, the proposed control chart is implemented into two phases. In Phase I, the parameters are estimated from an in-control sample, and in Phase II, we perform the monitoring scheme.

The remaining of the paper unfolds as follows. In Section 2, we describe the IBRCC and introduce the beta inflated mean regression model with varying dispersion. We also discuss likelihood inference and present the control limits estimation procedure. Section 3 presents a simulation study to evaluate (i) the inflated beta regression model with varying dispersion estimators and (ii) the performance of the proposed IBRCC and some competing control charts in the literature based on the run length (RL). In Section 4, we discuss and present an empirical application to show the applicability of the proposed IBRCC in real situations. Finally, some conclusions are presented in Section 5.

2 Inflated beta regression control chart

In this section, we introduce the IBRCC. Firstly, in Subsection 2.1 we present the inflated beta regression model with varying dispersion. The model we propose in this work is an extension of the model proposed by [12], where the authors used a reparametrization of the inflated beta distribution indexed by the response mean. In the Subsection 2.2 we present the model-based control limits for the proposed IBRCC. Secondly, in Subsubsection 2.2.1, we discuss the likelihood inference for the model parameters. Finally, in Subsubsection 2.2.2 we present the control limits estimation procedure.

2.1 Inflated beta regression model with varying dispersion

The inflated beta density function is given by [12]

$$f(y; \alpha_0, \alpha_1, \gamma, \phi) = \begin{cases} \alpha_0(1 - \gamma), & \text{if } y = 0, \\ c\mathcal{B}(y; \mu, \phi), & \text{if } y \in (0, 1), \\ \alpha_1\gamma, & \text{if } y = 1, \end{cases} \tag{1}$$

where $0 < \alpha_0 < 1, 0 < \alpha_1 < 1, 0 < \gamma < 1$, and $\phi > 0$ are the distribution parameters, $c = 1 - \alpha_0(1 - \gamma) - \alpha_1\gamma$, $\mu = \gamma(1 - \alpha_1)/c$, and $\mathcal{B}(y; \mu, \phi)$ is the beta density function given by [11]

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

where $0 < \mu < 1, \phi > 0$, and $\Gamma(\cdot)$ is the gamma function. Here, ϕ is a precision parameter (inversion of the dispersion), $E(y) = \gamma$, and $\text{Var}(y) = \frac{(1 + \alpha_1)\phi}{1 + \phi}\gamma + \left(\frac{(1 - \alpha_1)^2\phi}{(1 - \alpha_0(1 - \gamma) - \alpha_1\gamma)(1 + \phi)} - 1\right)\gamma^2$.

Density (1) is said to be zero and one inflated beta, i.e., $y \sim \mathcal{BIm}(\alpha_0, \alpha_1, \gamma, \phi)$. Note that $P(y = 0) = \alpha_0(1 - \gamma)$ and $P(y = 1) = \alpha_1\gamma$, thus if $\alpha_0 = 0$ and $\alpha_1 > 0$, the distribution in (1) is called one inflated beta distribution. Differently, if $\alpha_0 > 0$ and $\alpha_1 = 0$, the distribution given in (1) is called zero inflated beta distribution.

Let y_1, \dots, y_n be independent random variables where each y_t has the density in (1) for $t = 1, \dots, n$. The inflated beta regression model with varying dispersion, which is an extension of the regression model proposed by [12], is given by the following structures for modeling the

response y_t

$$g_1(\alpha_{0t}) = \tilde{x}_t^\top \omega, \tag{2}$$

$$g_2(\alpha_{1t}) = \tilde{x}_t^\top \kappa, \tag{3}$$

$$g_3(\gamma_t) = x_t^\top \beta, \tag{4}$$

$$g_4(\phi_t) = \ddot{x}_t^\top \zeta, \tag{5}$$

where $\omega = [\omega_0, \omega_1, \dots, \omega_{k_1}]^\top$, $\kappa = [\kappa_0, \kappa_1, \dots, \kappa_{k_2}]^\top$, $\beta = [\beta_0, \beta_1, \dots, \beta_{k_3}]^\top$, and $\zeta = [\zeta_0, \zeta_1, \dots, \zeta_{k_4}]^\top$ are vectors of unknown regression parameters, $\tilde{x}_t = [1, \tilde{x}_{t1}, \dots, \tilde{x}_{tk_1}]^\top$, $\tilde{x}_t = [1, \tilde{x}_{t1}, \dots, \tilde{x}_{tk_2}]^\top$, $x_t = [1, x_{t1}, \dots, x_{tk_3}]^\top$, and $\ddot{x}_t = [1, \ddot{x}_{t1}, \dots, \ddot{x}_{tk_4}]^\top$ are observations on k_1, k_2, k_3 , and k_4 covariates, respectively, and $g_1 : (0, 1) \rightarrow \mathbb{R}$, $g_2 : (0, 1) \rightarrow \mathbb{R}$, $g_3 : (0, 1) \rightarrow \mathbb{R}$, and $g_4 : (0, \infty) \rightarrow \mathbb{R}$ are real-valued link functions with continuous second derivatives. For $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$, several different link functions can be used, such as logit, probit, log-log, complementary log-log, or Cauchy. For $g_4(\cdot)$, the choices are log or square root link functions, for example. More details about link functions on the class of beta regression models can be found in [11] and [21].

Notably, the model proposed by [12] assume that, for $t = 1, \dots, n$, $\phi_t = \phi$ (constant precision). In the present study, the precision parameter is allowed to vary across observations, making the proposed model more general than the original inflated beta regression model. The assumption of non-constant dispersion is natural in several production processes [22–24]. In practical situations, it is important to monitor the dispersion of the process because an increase in dispersion may indicate process deterioration, while a reduction in dispersion means an improvement in process capability [25]. In addition, it is possible to consider control variables for modeling the parameters α_0 and α_1 , which are related to the probabilities of zero and one, respectively.

2.2 Model-based control chart limits

The purpose of IBRCC is to monitor double bounded processes that contain values equal to zero or one, considering that the mean, precision, and parameters related to the probabilities of zero and one (α_0 and α_1) of the quality characteristic of interest are affected by control variables. Let $(1 - \alpha)$ be a control region where α is the type I error probability, the lower control limit (LCL), center line (CL), and upper control limit (UCL) of the proposed control chart are defined, respectively, by

$$\begin{aligned} \text{LCL}_t &= F^{-1}(\alpha/2; \alpha_{0t}, \alpha_{1t}, \gamma_t, \phi_t), \\ \text{CL}_t &= \gamma_t, \\ \text{UCL}_t &= F^{-1}(1 - \alpha/2; \alpha_{0t}, \alpha_{1t}, \gamma_t, \phi_t), \end{aligned}$$

where $F(y) = P(Y \leq y) = \int_0^y f(u, \alpha_0, \alpha_1, \gamma, \phi) du$ is the inflated beta cumulative distribution function and $F^{-1}(\cdot)$ is the quantile function of the inflated beta variable. The parameters α_{0t} , α_{1t} , γ_t , and ϕ_t are functions of ω , κ , β , and ζ , respectively, and through (2), (3), (4), and (5) we have $\alpha_{0t} = g_1^{-1}(\tilde{x}_t^\top \omega)$, $\alpha_{1t} = g_2^{-1}(\tilde{x}_t^\top \kappa)$, $\gamma_t = g_3^{-1}(x_t^\top \beta)$, and $\phi_t = g_4^{-1}(\ddot{x}_t^\top \zeta)$. In practice, the model parameters are unknown and estimation methods are necessary to estimate the in-control limits. Thus, we consider the likelihood theory [26–28], which we discuss in the following

subsection. We presented results of the log-likelihood and score functions, which are extensions of those developed for the inflated beta regression model proposed by [12].

2.2.1 Likelihood inference. We shall consider the maximum likelihood estimator (MLE) for the parameter vector $\theta = (\omega^\top, \kappa^\top, \beta^\top, \zeta^\top)^\top$. The log-likelihood function is given by

$$\ell(\theta) = \sum_{t=1}^n \ell_t(\alpha_{0t}, \alpha_{1t}, \gamma_t, \phi_t), \tag{6}$$

with

$$\begin{aligned} \ell(\theta) = & \sum_{t=1}^n \log [\alpha_{0t}(1 - \gamma_t)] \mathbb{1}_0(y_t) + \log (\alpha_{1t}\gamma_t) \mathbb{1}_1(y_t) \\ & + [\log (c_t) + \log (\mathcal{B}(y_t; \mu_t, \phi_t))][(1 - \mathbb{1}_0(y_t))(1 - \mathbb{1}_1(y_t))], \end{aligned}$$

where $\mathbb{1}_0(y_t)$ is an indicator function that equals 1 if $y = 0$ and 0 if $y \in (0, 1]$, $\mathbb{1}_1(y_t)$ is an indicator function that equals 1 if $y = 1$ and 0 if $y \in [0, 1)$, and

$$\log [\mathcal{B}(y_t; \mu_t, \phi_t)] = \log \Gamma(\phi_t) - \log \Gamma(\mu_t\phi_t) - \log \Gamma((1 - \mu_t)\phi_t) + (\mu_t\phi_t - 1)y_t^* + (\phi_t - 2)y_t^\dagger,$$

in which α_{0t} , α_{1t} , γ_t , and ϕ_t are given by the regression structures in (2), (3), (4), and (5), respectively. Additionally,

$$y_t^* = \begin{cases} \log \left(\frac{y_t}{1 - y_t} \right), & y_t \in (0, 1), \\ 0, & y_t = 0 \text{ or } y_t = 1, \end{cases}$$

and $y_t^\dagger = \begin{cases} \log (1 - y_t), & y_t \in (0, 1), \\ 0, & y_t = 0 \text{ or } y_t = 1. \end{cases}$

By deriving the log-likelihood function in (6) with respect to each element of the parameter vector θ , we obtain the score vector given by $U(\theta) = (U_\omega(\theta)^\top, U_\kappa(\theta)^\top, U_\beta(\theta)^\top, U_\zeta(\gamma)^\top)^\top$. The MLE of θ is obtained by solving the non-linear system $(U_\omega(\theta)^\top, U_\kappa(\theta)^\top, U_\beta(\theta)^\top, U_\zeta(\gamma)^\top)^\top = \mathbf{0}$, where $\mathbf{0}$ is a null vector of dimension $(k_1 + k_2 + k_3 + k_4)$. The MLEs cannot be expressed in closed-form, hence the maximization of the log-likelihood function needs to be numerically conducted through a Newton or quasi-Newton algorithm. In this work, we used the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [29] for computational implementation.

The Fisher information matrix $(K(\theta))$, which is useful for large sample inferences, requires the expectations of second order derivatives of the log-likelihood function. The score vector $U(\theta)$ and Fisher’s information matrix can be found in the Appendix.

To test hypotheses on the parameter $\theta_j, j = 1, \dots, (k_1 + k_2 + k_3 + k_4)$, we consider the null hypothesis $\mathcal{H}_0 : \theta_j = \theta_j^0$ versus $\mathcal{H}_1 : \theta_j \neq \theta_j^0$. The Wald test may be considered using the following z statistic [26]

$$z = \frac{\hat{\theta}_j - \theta_j^0}{\widehat{\text{se}}(\hat{\theta}_j)},$$

where $\hat{\theta}_j$ represents the MLE of θ_j and the standard error of $\hat{\theta}_j$ is given by $\widehat{\text{se}}(\hat{\theta}_j) = [\text{diag}(\widehat{\text{cov}}(\hat{\theta}))]_j^{1/2}$, in which $\widehat{\text{cov}}(\hat{\theta}) = K^{-1}(\hat{\theta})$ is the asymptotic variance and covariance matrix of $\hat{\theta}$. In large sample sizes and under \mathcal{H}_0 , the z statistic follows a standard

normal distribution [26]. The test is performed by comparing the computed z statistic with the usual quantiles of the standard normal distribution.

2.2.2 Control limits estimation. After obtaining the MLE of θ , $\hat{\theta} = (\hat{\omega}^\top, \hat{\kappa}^\top, \hat{\beta}^\top, \hat{\zeta}^\top)^\top$, and considering an in-control process, the estimated control limits are given by

$$\widehat{LCL}_t = F^{-1}(\alpha/2; \hat{\alpha}_{0t}, \hat{\alpha}_{1t}, \hat{\gamma}_t, \hat{\phi}_t), \tag{7}$$

$$\widehat{CL}_t = \hat{\gamma}_t,$$

$$\widehat{UCL}_t = F^{-1}(1 - \alpha/2; \hat{\alpha}_{0t}, \hat{\alpha}_{1t}, \hat{\gamma}_t, \hat{\phi}_t), \tag{8}$$

where $\hat{\alpha}_{0t} = g_1^{-1}(\tilde{x}_t^\top \hat{\omega})$, $\hat{\alpha}_{1t} = g_2^{-1}(\tilde{x}_t^\top \hat{\kappa})$, $\hat{\gamma}_t = g_3^{-1}(x_t^\top \hat{\beta})$, and $\hat{\phi}_t = g_4^{-1}(\tilde{x}_t^\top \hat{\zeta})$. Thus, we propose the following algorithm to implement the IBRCC.

1. Fit the inflated beta regression model with varying dispersion under Phase I and obtain the MLEs, namely $\hat{\omega}$, $\hat{\kappa}$, $\hat{\beta}$, and $\hat{\zeta}$.
2. Using covariates in Phase II, estimate α_{0t} , α_{1t} , γ_t , and ϕ_t such that $\hat{\alpha}_{0t} = g_1^{-1}(\tilde{x}_t^\top \hat{\omega})$, $\hat{\alpha}_{1t} = g_2^{-1}(\tilde{x}_t^\top \hat{\kappa})$, $\hat{\gamma}_t = g_3^{-1}(x_t^\top \hat{\beta})$, and $\hat{\phi}_t = g_4^{-1}(\tilde{x}_t^\top \hat{\zeta})$.
3. For a given type I error probability α and the estimates $\hat{\alpha}_{0t}$, $\hat{\alpha}_{1t}$, $\hat{\gamma}_t$, and $\hat{\phi}_t$, compute the estimated control limits using (7) and (8).
4. Plot each data point y_t together with the estimated control limits \widehat{UCL}_t and \widehat{LCL}_t , for $t = 1, \dots, n$.

The observation y_t that is out of the estimated control limits interval $(\widehat{UCL}_t, \widehat{LCL}_t)$ is considered out-of-control.

3 Simulation study

This section presents a Monte Carlo simulation study to evaluate the estimators of the introduced inflated beta regression model with varying dispersion and the performance of the IBRCC. The performance of the proposed control chart is compared with some alternatives in literature, namely: the usual linear regression control chart (RCC) [1], the beta regression control chart (BRCC) [6], and the inflated beta control chart (IBCC) [30]. Note that the RCC is a classical regression control chart that works under Gaussian assumptions. The other control charts are state-of-the-art alternatives, but the BRCC does not consider inflation in zeros and/or ones and the IBCC does not include covariates.

We used the following structures for data generation

$$\begin{aligned} \text{logit}(\alpha_{0t}) &= \omega_0 + \omega_1 \tilde{x}_t, \\ \text{logit}(\alpha_{1t}) &= \eta_0 + \eta_1 \tilde{x}_t, \\ \text{logit}(\gamma_t) &= \beta_0 + \beta_1 x_t, \\ \log(\phi_t) &= \zeta_0 + \zeta_1 \tilde{x}_t, \end{aligned}$$

with $t = 1, \dots, n$. The values of \tilde{x}_t , \tilde{x}_t , and \tilde{x}_t were obtained from a Bernoulli distribution with parameter $p = 0.3$, and x_t was generated from a uniform distribution in the interval $(0, 1)$, thus considering discrete and continuous random variables. We considered 5,000 Monte Carlo replications and sample sizes $n = 100, 200, \text{ and } 500$. According to [31] and [32], this number of

Table 1. Different scenarios considered in the simulation study.

Scenario	Parameters							
	ω_0	ω_1	η_0	η_1	β_0	β_1	ζ_0	ζ_1
1	-1.00	-0.20	-1.00	-0.20	-0.30	0.60	2.00	1.00
2	-2.30	0.90	0.00	0.00	-2.40	0.80	4.50	-0.30
3	0.00	0.00	-2.50	0.50	3.50	-1.50	2.00	-0.70
4	-3.50	0.50	0.00	0.00	-4.40	1.20	5.50	-0.50
5	0.00	0.00	-3.50	0.30	2.50	-1.80	1.00	-0.20
6	-2.50	0.90	0.00	0.00	-2.70	1.00	3.00	-0.30
7	0.00	0.00	-1.00	-0.20	3.00	-1.20	1.50	-0.30

<https://doi.org/10.1371/journal.pone.0236756.t001>

replications is enough to obtain accurate results. All simulations were performed using the R programming language [33].

In the numerical evaluation, we considered several scenarios with different characteristics, namely: zero and one inflated beta regression model (Scenario 1), zero inflated beta regression model (Scenarios 2, 4, and 6), and one inflated beta regression model (Scenarios 3, 5, and 7). The parameter values are shown in Table 1. In Scenario 1, the mean is centered on the standard unit interval, $\gamma \in [0.43, 0.57]$, and the average percentage of zeros and ones in the sample is approximately equal to 13% for both. Scenarios 2, 4, and 6 consider the mean close to zero, with $\gamma \in [0.063, 0.154]$, $[0.012, 0.039]$, and $[0.039, 0.119]$, and average percentages of zeros in the sample equal to 9.3%, 3.4%, and 6.4%, respectively. For Scenarios 3, 5, and 7, the mean is close to one, with $\gamma \in [0.881, 0.971]$, $[0.668, 0.924]$, and $[0.858, 0.952]$, and average percentages of ones in the sample equal to 8.3%, 2.6%, and 23.5%, respectively.

3.1 Point estimation evaluation

For the point estimation evaluation, we computed the mean, percentage relative bias (RB), and mean square error (MSE) for each estimator in all Scenarios (see Table 1). For brevity and similarity of results, we only present results for Scenarios 1, 2, and 7 ($n = 100$ and $n = 500$) as shown in Table 2. The figures show that the mean of the estimators is close to the corresponding parameter values. The RB and MSE decrease when the sample size increases, indicating that the MLEs are consistent. For instance, for β_1 (γ submodel) in Scenario 1, the RB of the estimator is equal to 0.2257% for $n = 100$ and equal to -0.0548% for $n = 500$. Regarding MSE, considering ω_0 in Scenario 2 and $n \in \{100, 500\}$, the MSE is equal to 0.2291 and 0.0373, respectively. As in other studies related to beta regression [21, 34], it is noteworthy that the RB of MLEs corresponding to the precision covariate parameters is greater than those of that model the mean response. For instance, consider $\hat{\zeta}_1$ (ϕ submodel) in Scenario 7, we have $RB = -8.2035\%$ for $n = 100$ and $RB = -2.4411\%$ for $n = 500$. Regarding parameters related to the probabilities of zeros and ones, the bias also decreases considerably as sample size increases. For example, in Scenario 1 and $n = 100$, the estimator of ω_1 (α_0 submodel) yields $RB = 16.9130\%$ and the estimator of η_1 (α_1 submodel) yields $RB = 14.8922\%$. For $n = 500$, the bias for the same estimators reduces to 4.5210% and 3.4431%, respectively.

In practice, the regression model relating the output and covariates is rarely known and the parameters have to be estimated. Our simulation results show that the MLE in the proposed model perform well, presenting low MSE for the estimates in all situations. This way, the proposed control chart may also present good performance in practice. In the next section, we shall investigate the run length performance of the IBRCC with estimated parameters.

Table 2. Monte Carlo simulation results of point estimation evaluation.

Scenario 1									
$n \downarrow$	Param. \rightarrow	$\omega_0 = -1.00$	$\omega_1 = -0.20$	$\eta_0 = -1.00$	$\eta_1 = -0.20$	$\beta_0 = -0.30$	$\beta_1 = 0.60$	$\zeta_0 = 2.00$	$\zeta_1 = 1.00$
100	Mean	-1.0393	-0.2338	-1.0439	-0.2298	-0.3026	0.6013	2.0455	1.0717
	RB	3.9304	16.9130	4.3925	14.8922	0.8683	0.2257	2.2770	7.1753
	MSE	0.1396	0.3936	0.1425	0.4091	0.0279	0.0553	0.0403	0.1412
500	Mean	-1.0078	-0.2090	-1.0146	-0.2069	-0.3009	0.5997	2.0080	1.0142
	RB	0.7846	4.5210	1.4628	3.4431	0.3143	-0.0548	0.4141	1.4201
	MSE	0.0230	0.0695	0.0241	0.0686	0.0052	0.0101	0.0071	0.0250
Scenario 2									
$n \downarrow$	Param. \rightarrow	$\omega_0 = -2.30$	$\omega_1 = 0.90$	$\eta_0 = 0.0$	$\eta_1 = 0.0$	$\beta_0 = -2.40$	$\beta_1 = 0.80$	$\zeta_0 = 4.50$	$\zeta_1 = -0.30$
100	Mean	-2.3694	0.9302	—	—	-2.4034	0.80115	4.5469	-0.2616
	RB	3.0207	3.3572	—	—	0.1455	0.1445	1.0435	-12.7889
	MSE	0.2291	0.2588	—	—	0.0076	0.0161	0.0359	0.1254
500	Mean	-2.313	0.9030	—	—	2.4002	0.8005	4.5093	-0.2911
	RB	0.5780	0.3328	—	—	0.0096	0.0637	0.2072	-2.9518
	MSE	0.0373	0.0403	—	—	0.0013	0.0027	0.0062	0.0222
Scenario 7									
$n \downarrow$	Param. \rightarrow	$\omega_0 = 0.0$	$\omega_1 = 0.0$	$\eta_0 = -1.00$	$\eta_1 = -0.20$	$\beta_0 = 3.00$	$\beta_1 = -1.20$	$\zeta_0 = 1.50$	$\zeta_1 = -0.30$
100	Mean	—	—	-1.0201	-0.2602	3.0223	-1.2099	1.5475	-0.2753
	RB	—	—	2.0136	30.1092	0.7455	0.8254	3.1698	-8.2035
	MSE	—	—	0.0827	0.3659	0.0952	0.2023	0.0478	0.0990
500	Mean	—	—	-1.0007	-0.2116	3.0062	-1.2028	1.5094	-0.2927
	RB	—	—	0.0794	5.8189	0.2071	0.2380	0.6269	-2.4411
	RB	—	—	0.0156	0.0539	0.0155	0.0318	0.0085	0.0176

<https://doi.org/10.1371/journal.pone.0236756.t002>

3.2 Control charts performance

This section presents a run length analysis to evaluate the performance of the considered control charts. When the process is in-control, the run length (RL) distribution follows a geometric distribution with parameter α , which is the type I error probability [35]. The ability of a control chart to detect changes in the process is usually measured by the average number of observations until the detection of an out-of-control point (ARL) [36]. However, other measures can also be used for this purpose. We considered another location measure, the median (MRL), a dispersion measure, and the standard deviation (SDRL) of the RL distribution. Additionally, we computed the mean absolute percentage error (MAPE) for each measure for all evaluated control charts.

We compared the proposed IBRCC with the standard RCC [1], and the state-of-the-art charts, namely BRCC [6] and IBCC [30]. Since the BRCC does not consider values equal to zero or one, we replaced zeros by 0.0001 and ones by 0.9999 for its application. For all considered control charts, we examined two aspects of evaluation: in-control ($ARL_0 = \frac{1}{\alpha}$, $MRL_0 = \frac{\ln(0.5)}{\ln(1-\alpha)}$, $SDRL_0 = \sqrt{\frac{(1-\alpha)}{\alpha^2}}$) and out-of-control ($ARL_1 = \frac{1}{1-\beta}$, where β is the type II error probability) [30, 35]. For brevity, we do not present the MRL and SDRL results for the out-of-control process. The control charts were evaluated in Scenarios 2 to 7 (Table 2), considering inflation in 0 or 1. Scenario 1 was not covered in this section because it does not reflect real statistical process control situations, being possible to present perfect nonconforming and perfect conforming in the same process.

Table 3. Run length analysis to evaluate the IBRCC, BRCC, RCC, and IBCC with $\alpha = 0.01$.

Scenario	n	Control Chart											
		IBRCC			BRCC			RCC			IBCC		
		\widehat{MRL}_0	\widehat{ARL}_0	\widehat{SDRL}_0	\widehat{MRL}_0	\widehat{ARL}_0	\widehat{SDRL}_0	\widehat{MRL}_0	\widehat{ARL}_0	\widehat{SDRL}_0	\widehat{MRL}_0	\widehat{ARL}_0	\widehat{SDRL}_0
2	100	69.48	99.43	97.75	7.54	10.65	10.12	43.53	62.19	61.17	77.08	109.87	107.91
	200	69.16	98.73	97.46	7.58	10.69	10.17	38.34	54.85	54.04	76.69	109.59	108.19
	500	69.42	99.28	97.67	7.57	10.69	10.16	37.96	54.16	53.32	77.53	110.57	108.69
3	100	74.27	106.22	104.76	7.68	10.86	10.34	23.62	33.78	33.09	87.16	124.41	122.26
	200	70.87	101.17	99.52	7.79	11.01	10.50	21.62	30.85	30.21	86.61	122.65	119.98
	500	70.24	100.49	98.76	7.87	11.10	10.57	21.42	30.58	29.93	88.00	125.35	122.89
4	100	63.05	90.08	88.78	19.24	27.48	26.92	35.35	50.50	49.66	97.63	138.71	135.74
	200	68.37	97.81	96.37	19.95	28.52	27.96	30.22	43.21	42.53	95.68	136.66	134.42
	500	69.22	98.90	97.20	20.33	29.04	28.42	29.84	42.51	41.75	97.63	139.37	137.04
5	100	66.68	94.64	92.66	19.00	27.16	26.56	32.33	46.20	45.30	93.19	132.47	129.93
	200	69.76	99.76	98.21	19.73	28.20	27.61	27.38	39.18	38.55	92.42	132.07	129.97
	500	69.79	100.31	100.08	20.24	28.91	28.30	26.02	37.16	36.47	92.98	132.53	129.88
6	100	69.48	99.03	97.10	10.04	14.24	13.71	27.49	39.24	38.47	79.40	113.00	110.72
	200	70.71	101.42	100.22	10.20	14.48	13.94	25.08	35.84	35.20	78.84	112.44	110.78
	500	70.12	100.28	98.59	10.29	14.61	14.06	24.84	35.49	34.80	80.02	114.04	112.06
7	100	75.14	105.57	103.76	3.00	4.17	3.64	21.89	31.30	30.71	82.76	118.03	116.25
	200	71.22	101.62	100.05	3.00	4.18	3.64	20.32	29.03	28.45	82.43	117.38	115.39
	500	69.88	99.88	98.28	3.00	4.19	3.65	19.97	28.52	27.90	82.88	118.20	115.92
MAPE	100	4.73	4.77	5.23	84.05	84.24	84.71	55.82	56.13	56.72	24.03	22.75	21.07
	200	1.45	1.32	1.30	83.63	83.82	84.28	60.92	61.17	61.64	22.94	21.80	20.39
	500	0.57	0.50	1.27	83.38	83.58	84.06	61.62	61.93	62.45	24.47	23.34	21.69

<https://doi.org/10.1371/journal.pone.0236756.t003>

To ensure that the comparisons between ARL_1 occur between control charts of same ARL_0 , we adjusted the chart limits to obtain ARL_0 equal to the specified nominal values of 100 and 370. This control chart calibration is suggested in the literature [30, 37–39]. After ARL_0 calibration, a δ change was induced in the mean and precision regression structures to generate out-of-control processes as the following: $\text{logit}(\gamma_t) = \delta + \beta_0 + \beta_1 x_t$ and $\text{log}(\phi_t) = \delta + \zeta_0 + \zeta_1 \ddot{x}_t$. By enabling the process to be out-of-control, we obtained the estimated ARL_1 for different values of δ . When $\delta = 0$, the process is in-control and the ARL_0 can be evaluated.

The \widehat{ARL}_0 , \widehat{MRL}_0 , and \widehat{SDRL}_0 evaluation results are shown in Tables 3 and 4. Consider an in-control process with α – values of 0.01 and 0.0027, from the geometric distribution of the RL we have ARL values equal to 100 and 370, nominal MRL values equal to 69.0 and 256.1, and values of $SDRL$ equal to 99.5 and 369.5, respectively. The IBRCC showed better performance than BRCC, RCC, and IBCC, reaching empirical values closer to the nominal levels in all evaluated scenarios. In Scenarios 2, 4, and 6, the IBRCC and IBCC obtained 0 as the lower control limit, thus no point exceeded this limit. Similarly, in Scenarios 3, 5, and 7, the upper control limit of the mentioned charts were 1. The fact that these scenarios present the 0 or 1 as control limits is related to the value of the probabilities of occurrence of 0 or 1. That is, the IBRCC and IBCC will present zero as control limit when $P(Y = 0) = \hat{\alpha}_0(1 - \hat{\gamma}_t) \geq \alpha/2$ and one as control limit when $P(Y = 1) = \hat{\alpha}_1\hat{\gamma}_t \geq \alpha/2$. The high probability of Y assuming values equal to one or zero means that these values are not atypical (out of control) but usual occurrences of the process.

Table 4. Run length analysis to evaluate the IBRCC, BRCC, RCC, and IBCC with $\alpha = 0.0027$.

Scenario	n	Control Chart											
		IBRCC			BRCC			RCC			IBCC		
		\widehat{MRL}_0	\widehat{ARL}_0	\widehat{SDRL}_0	\widehat{MRL}_0	\widehat{ARL}_0	\widehat{SDRL}_0	\widehat{MRL}_0	\widehat{ARL}_0	\widehat{SDRL}_0	\widehat{MRL}_0	\widehat{ARL}_0	\widehat{SDRL}_0
2	100	279.42	377.20	346.43	7.59	10.72	10.19	87.81	124.47	121.69	300.24	416.27	396.89
	200	266.36	368.23	350.79	7.61	10.73	10.20	73.39	105.73	104.39	293.99	411.46	396.08
	500	262.62	366.62	349.45	7.59	10.71	10.18	72.33	103.34	101.68	301.16	419.35	402.25
3	100	318.25	426.92	387.41	8.17	11.59	11.07	35.13	50.12	49.23	305.31	422.34	401.87
	200	279.92	385.53	364.98	8.23	11.68	11.16	31.33	44.70	43.90	299.12	415.33	395.29
	500	270.43	376.94	361.74	8.26	11.70	11.17	30.91	44.12	43.37	309.09	428.96	410.88
4	100	250.95	342.62	319.09	20.10	28.74	28.17	88.01	124.20	120.69	444.32	607.52	570.76
	200	261.78	364.01	348.01	20.42	29.19	28.62	71.42	101.93	100.16	434.44	596.99	564.95
	500	261.74	364.58	348.34	20.51	29.31	28.70	69.61	99.46	97.80	448.49	616.69	581.46
5	100	258.92	351.89	327.00	23.40	33.48	32.84	115.84	157.55	146.89	405.71	555.36	524.39
	200	278.17	382.32	358.68	24.04	34.35	33.72	73.29	103.69	100.90	399.52	551.36	523.64
	500	279.66	383.90	356.75	24.40	34.89	34.24	62.04	88.66	87.15	404.63	557.51	528.28
6	100	279.43	379.74	353.87	10.69	15.18	14.64	43.98	62.72	61.53	274.89	383.03	366.43
	200	280.08	386.16	365.02	10.78	15.29	14.75	39.10	55.86	55.04	270.35	378.39	364.49
	500	267.76	373.17	356.16	10.79	15.33	14.78	38.57	54.98	54.06	276.81	387.16	372.06
7	100	312.39	417.06	376.00	3.00	4.23	3.70	35.12	50.18	49.31	335.56	466.47	445.26
	200	286.76	390.91	365.49	3.00	4.24	3.70	31.35	44.82	44.10	333.98	463.49	443.27
	500	268.50	373.47	358.37	3.00	4.24	3.70	30.46	43.55	42.81	341.14	471.54	448.48
MAPE	100	11.26	7.50	7.04	95.25	95.32	95.46	73.59	74.36	75.22	34.45	28.42	22.32
	200	7.58	3.27	2.89	95.18	95.25	95.39	79.18	79.43	79.77	32.20	26.89	21.68
	500	4.82	1.63	3.89	95.15	95.22	95.36	80.22	80.45	80.75	35.45	29.78	23.74

<https://doi.org/10.1371/journal.pone.0236756.t004>

Tables 3 and 4 also show the MAPE results. The proposed control chart has the lowest values for the MAPE. For example, consider $\alpha = 0.01$, corresponding \widehat{ARL}_0 , and $n = 200$, the MAPE obtained for the IBRCC, BRCC, RCC, and IBCC were, 1.32, 83.82, 61.17, and 21.80, respectively. It is noteworthy that, for the IBRCC, the MAPE decreases considerably when the sample size increases.

Among the considered alternative control charts, the IBCC achieved better performance than the BRCC and RCC in all scenarios. In Scenario 2 (Table 4), for $n = 200$, the IBCC presented a false alarm after 411 samples, when, in fact, a false alarm was expected for every 370 samples. In the same scenario, the BRCC and the RCC presented a false alarm rate in approximately 11 and 106 samples, respectively. These results show the importance of considering an accurate model to reduce false alarms. We also note that the BRCC obtained the worst performance. In Table 3, consider Scenario 3 and $n = 100$, the RCC and BRCC presented a false alarm in approximately each 50 and 27 observations, respectively. It is important to note that BRCC performance worsened as the 0 or 1 percentage increased. Confirming this fact, the IBCC also presented lower MAPE than BRCC and RCC in all scenarios. Considering $\alpha = 0.0027$, $n = 500$, and the \widehat{MRL}_0 measure, the MAPE obtained for the IBRCC, BRCC, RCC, and IBCC were, respectively, 4.82, 95.15, 80.22, and 35.45.

Results of the \widehat{ARL}_1 evaluation are shown graphically in Figs 1 and 2. It was not possible to correct \widehat{ARL}_0 for the BRCC due to the poor in-control performance. Thus, the evaluation of \widehat{ARL}_1 was given only for the IBRCC, RCC, and IBCC. It is noteworthy that when several control charts are compared in terms of ARL, the one that presents the lowest \widehat{ARL}_1 among those with same \widehat{ARL}_0 is the control chart that outperforms the competitors [30]. By analyzing the

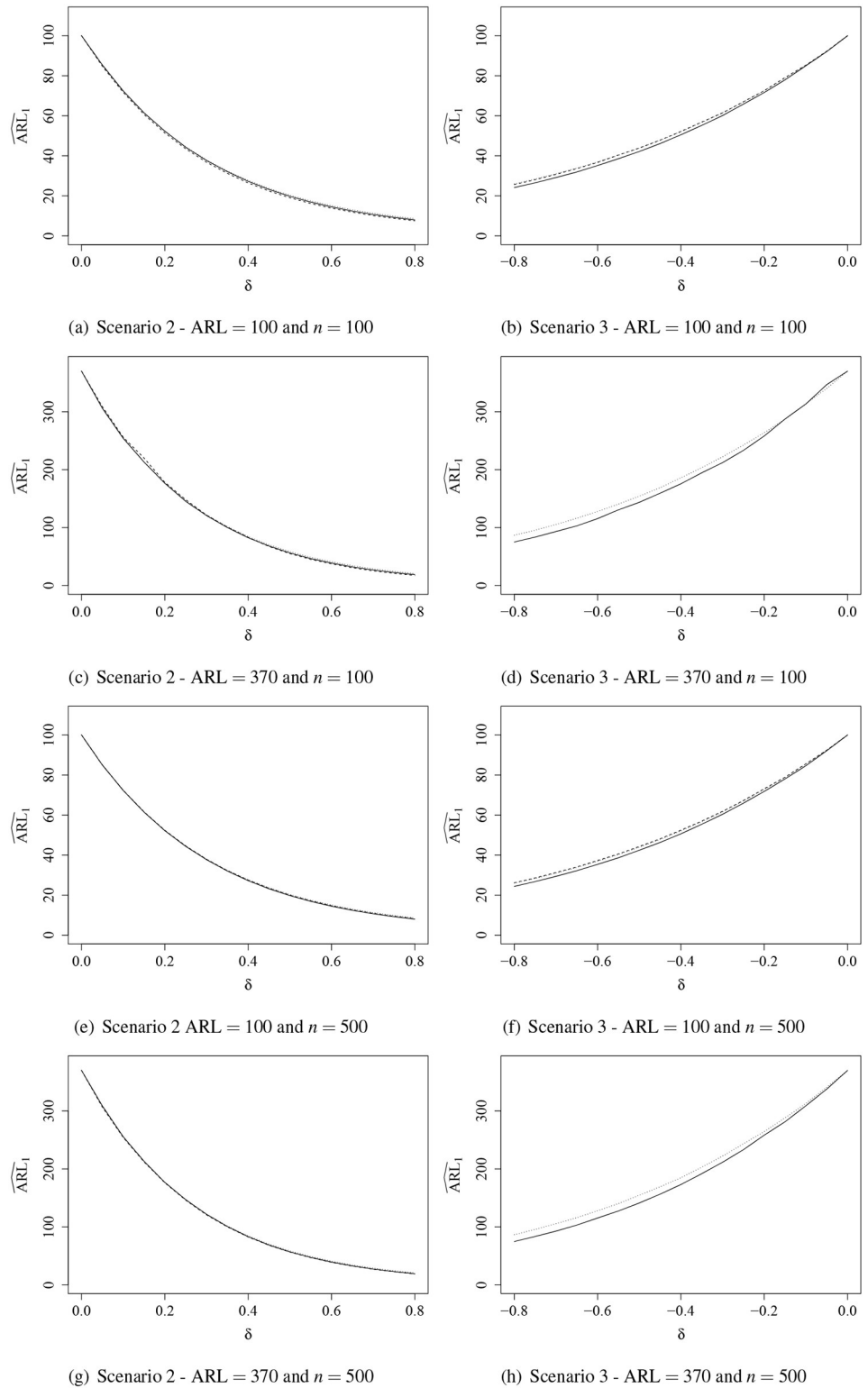


Fig 1. \widehat{ARL}_1 curves evaluation for the inflated beta regression control chart (solid line), regression control chart (dashed line), and inflated beta control chart (dotted line) when the mean is out-of-control.

<https://doi.org/10.1371/journal.pone.0236756.g001>

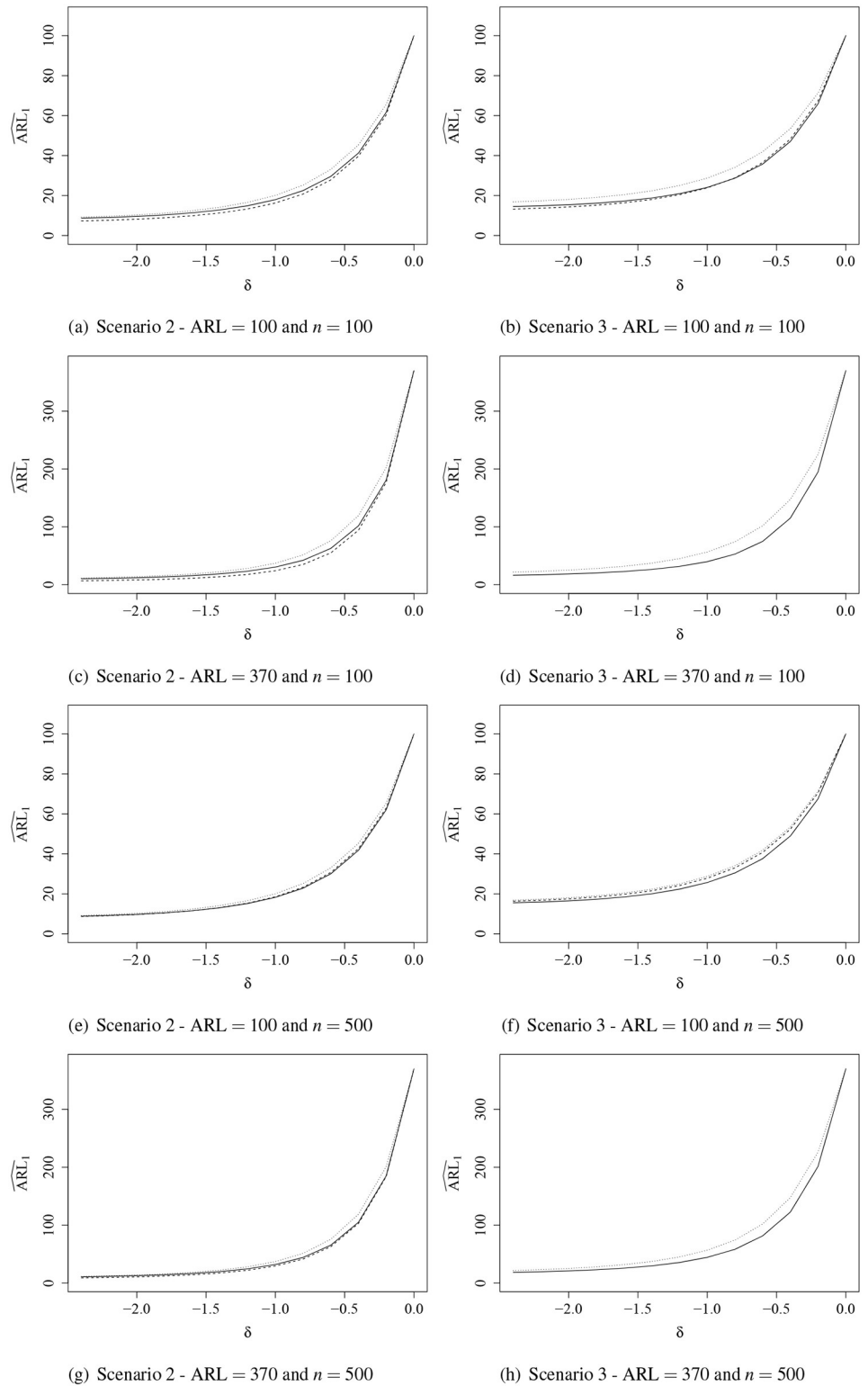


Fig 2. \widehat{ARL}_1 curves evaluation for the inflated beta regression control chart (solid line), regression control chart (dashed line), and inflated beta control chart (dotted line) when the precision is out-of-control.

<https://doi.org/10.1371/journal.pone.0236756.g002>

ARL₁ results, when the perturbation was introduced in the mean of the process (Fig 1), we observe that in Scenario 3 the IBRCC performs better than the RCC and IBCC, and in Scenario 2 the performance of the control charts are similar. We note that the IBRCC detects more quickly the out-of-control process. For example, in Scenario 3, $ARL = 370$, $n = 100$, and $\delta = -0.4$, the IBRCC takes 176 samples on average to detect a change in the process, while the IBCC takes 186 and the RCC takes 192 to detect a change of same magnitude. The simulation results showed similar behavior when a perturbation in the precision of the process occurs (Fig 2). The control charts detect process changes more quickly as the precision increases (dispersion decreases).

By considering the results obtained in the simulation, we see a necessity of using a control chart based on an appropriate regression model, such as the IBRCC, when the variable of interest is restricted to the intervals $[0, 1)$ or $(0, 1]$. The use of the linear regression-based control chart is inappropriate for data of this type since the support of the usual regression model is the whole real space. Interestingly, the BRCC proved to be more inadequate in the presence of values equal to zero or one than the traditional RCC or the IBCC that uses inflated beta distribution but does not consider a regression structure. Since the BRCC does not accommodate values equal to zero or one, by substituting zeros for 0.0001 and ones for 0.9999, an inflation in these values is induced. That is, the probability mass at 0.0001 and/or 0.9999 exceeds what is allowed by the beta distribution, which is an absolutely continuous distribution. This reflects on the estimates of the parameters of the regression structures and, automatically, the estimates of the control limits are impaired.

4 Real data application

This section contains an empirical application in which the proposed control chart (IBRCC) and three other competing control charts are analyzed: the RCC, BRCC, and IBCC. The data evaluated in this section refer to the public administrative efficiency of the municipalities in the state of São Paulo, Brazil. The data are a subset of those analyzed by [40], who considered all Brazilian municipalities. The dataset we used contains 427 municipalities for the year 2000 and it is available at http://www.de.ufpb.br/~luiz/datasets/Dataset_plosone.txt. The covariates are from Secretaria do Tesouro Nacional (<http://www.tesouro.fazenda.gov.br/>), Instituto Brasileiro de Geografia e Estatística (IBGE) (<https://www.ibge.gov.br/>), and Instituto de Pesquisa Econômica Aplicada (IPEA) (<https://www.ipea.gov.br/portal/>), Brazil. The quality characteristic, y , is introduced by [40] and represents individual observations of an efficiency index, assuming values in $(0, 1]$ and measuring how well mayors spend taxpayer money in order to provide them with public services. The efficiency index is equal to one when there is full efficiency. There are 32 units that are fully efficient (i.e., about 7.5% of the observations are equal to one). A brief description of the variables used in the analysis is presented in Table 5. Variables *CONS*, *R2*, and *MT* are dummies, i.e., they are equal to 0 or 1. The covariate *CONS* equals 1 if the municipality participates in the inter-municipal consortia, the covariate *R2* equals 1 whenever the municipality receives more than 10% of its tax revenue to *royalty*, and the covariate *MT* equals 1 whenever the municipality is tourist, 0 otherwise for the three dummies covariates. It is important to mention that 100 municipalities were sorted to estimate the model parameters (Phase I), while the remaining observations were used for monitoring (Phase II).

At the outset, the inflated (at one) beta mean regression model, the beta regression model substituting 1 for 0.9999, and a linear regression model were selected and fitted. We used the logit link for γ and α_1 and the log link for ϕ . For the beta regression, we considered logit for μ and log link for ϕ . The maximum likelihood estimates of the models parameters are displayed in Table 6. All covariates were significant at the nominal level of 5%. In order to compare the

Table 5. Description of the variables for efficiency data.

Variable	Description
<i>EFFIC</i>	Efficiency scores
<i>EXP</i>	Personnel expenses in Reais
<i>INC</i>	Average income in Reais
<i>CONS</i>	Participation in inter-municipal consortia
<i>URB</i>	Urbanization rate
<i>R2</i>	Royalties
<i>E20</i>	Demographic density
<i>MT</i>	Tourist municipality

<https://doi.org/10.1371/journal.pone.0236756.t005>

fitted regression models, we considered the MAPE and MSE between the observed and fitted values. According to these criteria the inflated beta regression model outperforms the other ones, with MAPE = 26.8835 and MSE = 0.0387, while the beta regression model obtains MAPE = 29.4101 and MSE = 0.0454, and linear regression model achieves MAPE = 29.3617 and MSE = 0.0389.

Table 7 presents some descriptive statistics of the estimated control limits. Note that the proposed control chart and the IBCC are the only ones that have an upper control limit constant and equal to one. Differently, when beta regression control chart is used, the control limits were restricted to the open interval (0, 1) and thus, in this case, fully efficient municipalities are considered out-of-control. In addition, we verify that, by using the standard RCC, the limits assume values below zero and above one, not being restricted to the interval (0, 1], where the data are distributed. The interpretation of the limits, in this case, makes no practical sense and leads to loss of detection power of out-of-control points.

Fig 3 graphically presents the control limits of the (a) IBRCC, (b) BRCC, (c) RCC, and (d) IBCC together with the observed values of efficiency considering $ARL_0 = 100$. Considering the fact that the efficiency index assumes values in (0, 1], the proposed model-based control chart

Table 6. Adjusted models for efficiency data.

The fitted inflated beta regression model										
	Submodel for γ				Submodel for ϕ		Submodel for α_0			
	Intercept	<i>E20</i>	<i>INC</i>	<i>URB</i>	Intercept	<i>MT</i>	Intercept	<i>EXP</i>	<i>INC</i>	<i>R2</i>
Estimate	0.6552	0.0001	-0.0015	0.0124	2.0339	0.7744	-2.2551	-1.0427	0.0014	0.5052
Std. error	0.4348	<0.0001	0.0005	0.0059	0.1527	0.3822	0.0884	0.0071	0.0001	0.0987
<i>p</i> -value	0.1319	<0.0001	0.0009	0.0371	<0.0001	0.0428	<0.0001	<0.0001	<0.0001	<0.0001
The fitted beta regression model										
	Submodel for μ					Submodel for ϕ				
	Intercept	<i>EXP</i>	<i>E20</i>	<i>URB</i>	<i>R2</i>	Intercept	<i>CONS</i>	<i>URB</i>		
Estimate	-0.7151	-0.1436	0.0002	0.0171	0.7391	4.2200	-2.2783	-0.0292		
Std. error	0.4272	0.0704	0.0001	0.0056	0.3100	0.8236	0.3102	0.0095		
<i>p</i> -value	0.0941	0.0415	0.0280	0.0023	0.0173	<0.0001	<0.0001	0.0022		
The fitted linear regression model										
	Model for μ									
	Intercept	<i>CONS</i>	<i>URB</i>							
Estimate	0.3829	0.1282	0.0031							
Std. error	0.1126	0.0497	0.0013							
<i>p</i> -value	0.0010	0.0114	0.0196							

<https://doi.org/10.1371/journal.pone.0236756.t006>

Table 7. Descriptive statistics—minimum (min), first quartile ($Q_{1/4}$), median, mean, third quartile ($Q_{3/4}$), and maximum (max)—control limits for efficiency data.

Limits	min	$Q_{1/4}$	median	mean	$Q_{3/4}$	max
LCL_{IBRCC}	0.0042	0.2176	0.2500	0.2546	0.2832	0.5936
UCL_{IBRCC}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
LCL_{BRCC}	0.0001	0.0998	0.1656	0.1429	0.2006	0.4323
UCL_{BRCC}	0.0001	0.9582	0.9864	0.9594	0.9974	0.9999
LCL_{RCC}	-0.0070	0.1731	0.2064	0.2040	0.2305	0.3640
UCL_{RCC}	0.9057	1.0858	1.1191	1.1167	1.1432	1.2767
LCL_{IBCC}	0.2424	0.2424	0.2424	0.2424	0.2424	0.2424
UCL_{IBCC}	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

<https://doi.org/10.1371/journal.pone.0236756.t007>

(IBRCC) presents limits with a smaller range. Interestingly, the BRCC does not accommodate values equal to one by substituting values equal to one for 0.9999, an inflation in these values is induced, therefore the BRCC is less adequate in the presence of values equal to one than the traditional RCC. The use of the linear regression-based control chart is inappropriate for data of this type since the support of the usual regression model is the whole real space. Finally, the IBCC that uses inflated beta distribution but does not consider regression structure presents constant limits that are not appropriate in situations where we have control variables (covariates). It is worth mentioning that IBRCC detected 7 out-of-control points, while BRCC detected 36 out-of-control points. Lastly, we carried out the RESET misspecification test [41],

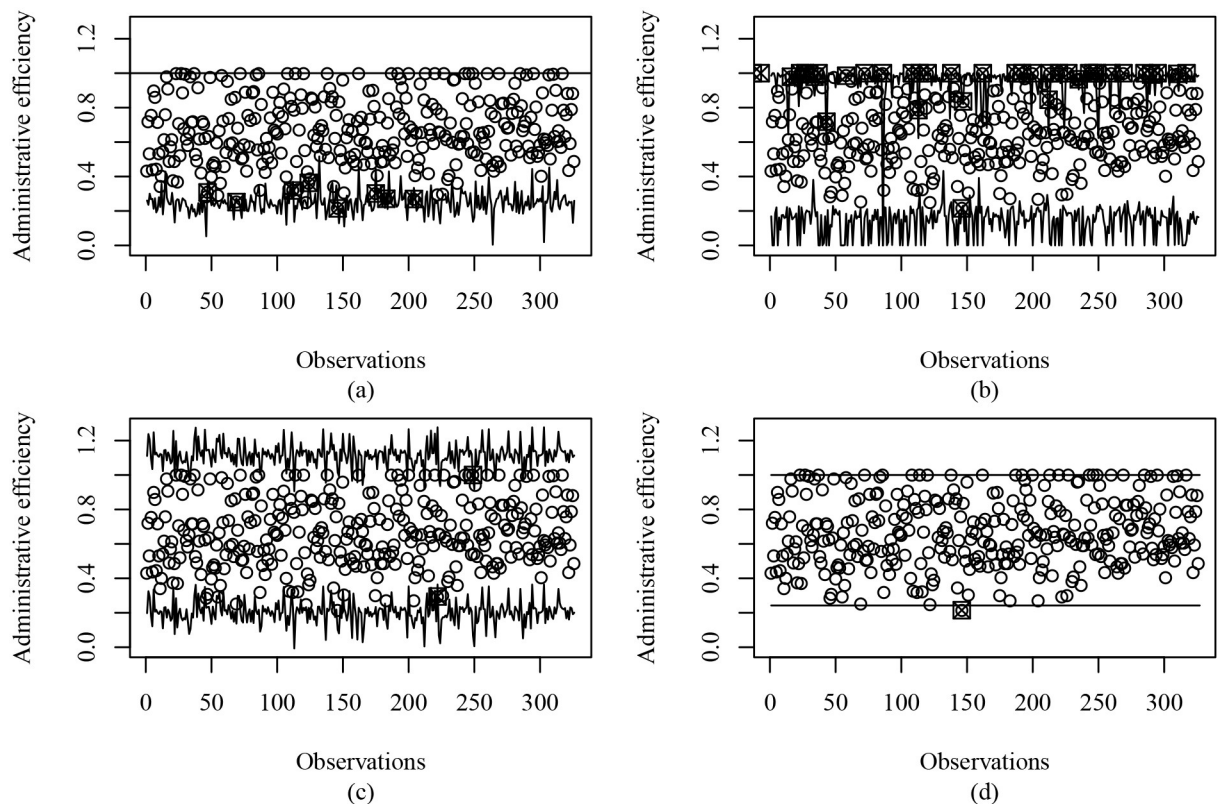


Fig 3. Plot of the control limits based on (a) inflated beta regression control chart, (b) beta regression control chart, (c) regression control chart, and (d) inflated beta control chart for monitoring the efficiency indexes for municipalities in the state of São Paulo, Brazil, considering $ARL_0 = 100$.

<https://doi.org/10.1371/journal.pone.0236756.g003>

where the null hypothesis is that the fitted model is correctly specified and the alternative hypothesis is that there is model misspecification. We perform the test using the second power of the estimated mean linear predictor as testing variables. We do not reject the null hypothesis at the 1% nominal level, thus suggesting that our model is correctly specified.

5 Conclusions

In this paper, we proposed a new model-based control chart for controlling quality characteristics limited to the intervals (0, 1] or [0, 1) using the inflated beta regression model. For this purpose, we extended the inflated beta regression model proposed by [12] by allowing a regression structure for the precision parameter. In this way, it is possible to model the mean response, the data precision, and functions of the probability of a given observation assuming zero or one through a regression framework. Our simulation study showed that the relative bias and mean square error decrease when the sample size increases. With regard to the sensitivity analysis in terms of run length (RL), the proposed IBRCC showed the best performance in all considered cases. In addition, the results indicated that it is better to ignore the explanatory variables and use the inflated beta control chart (IBCC) than to use a control chart based on an inappropriate regression model. We also considered an application to real data and highlight the practical importance of the proposed chart when the response is distributed in unit intervals containing ones. Finally, we suggest the use of the inflated beta regression control chart to monitor output quality characteristics, which is better characterized by a functional relation between the response variable, double bounded in unit intervals containing zeros or ones along with one or more explanatory variables.

A Score function and Fisher’s information matrix

In this appendix we obtain the score function and presented a closed-form expression for Fisher’s information matrix for all parameters of the inflated beta regression model with varying dispersion. We assume that the observed values of the dependent fractional variable are sorted according to the 0, 1, and (0, 1)-values with n_0 , n_1 , and $n - m$ terms, respectively, where $m = n_0 + n_1$. Furthermore, $\mu_t^* = E(y_t^*) = \psi(\mu_t \phi_t) - \psi((1 - \mu_t) \phi_t)$ and $\mu_t^\dagger = E(y_t^\dagger) = \psi((1 - \mu_t) \phi_t) - \psi(\phi_t)$, where $\psi(\cdot)$ is the digamma function.

Let

$$\frac{\partial \log [\mathcal{B}(y_t; \mu_t, \phi_t)]}{\partial \mu_t} = \phi_t (y_t^* - \mu_t^*).$$

The score function for ω is given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \omega_j} = \sum_{t=1}^{n_0} \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \alpha_{0t}} \frac{\partial \alpha_{0t}}{\partial \omega_j} + \sum_{t=m+1}^n \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \alpha_{0t}} \frac{\partial \alpha_{0t}}{\partial \omega_j} + \sum_{t=m+1}^n \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \mu_t} \frac{\partial \mu_t}{\partial \alpha_{0t}} \frac{\partial \alpha_{0t}}{\partial \omega_j},$$

where $\partial \alpha_{0t} / \partial \omega_j = \tilde{x}_{ij} / g'_1(\alpha_{0t})$. Therefore,

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \omega_j} = \sum_{t=1}^n \left\{ \frac{1}{\alpha_{0t} g'_1(\alpha_{0t})} \tilde{x}_{ij} \mathbb{1}_0(y_t) - \left[\frac{1 - \gamma_t}{c_t g'_1(\alpha_{0t})} \tilde{x}_{ij} \right. \right. \\ \left. \left. - \phi_t \frac{\gamma_t (1 - \gamma_t) (1 - \alpha_{1t}) (y_t^* - \mu_t^*)}{c_t^2 g'_1(\alpha_{0t})} \tilde{x}_{ij} \right] (1 - \mathbb{1}_0(y_t)) (1 - \mathbb{1}_1(y_t)) \right\}. \end{aligned}$$

The score function for κ is given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa_j} = \sum_{t=n_0+1}^m \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \alpha_{1t}} \frac{\partial \alpha_{1t}}{\partial \kappa_j} + \sum_{t=m+1}^n \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \alpha_{1t}} \frac{\partial \alpha_{1t}}{\partial \kappa_j} + \sum_{t=m+1}^n \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \mu_t} \frac{\partial \mu_t}{\partial \alpha_{1t}} \frac{\partial \alpha_{1t}}{\partial \kappa_j},$$

where $\partial \alpha_{1t} / \partial \kappa_j = \check{x}_{ij} / g'_2(\alpha_{1t})$. Thus,

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \kappa_j} = \sum_{t=1}^n \left\{ \frac{1}{\alpha_{1t} g'_2(\alpha_{1t})} \check{x}_{ij} \mathbb{1}_1(y_t) - \left[\frac{\gamma_t}{c_t g'_2(\alpha_{1t})} \check{x}_{ij} \right. \right. \\ \left. \left. + \phi_t \frac{\gamma_t(1-\gamma_t)(1-\alpha_{0t})(y_t^* - \mu_t^*)}{c_t^2 g'_2(\alpha_{1t})} \check{x}_{ij} \right] (1 - \mathbb{1}_0(y_t))(1 - \mathbb{1}_1(y_t)) \right\}. \end{aligned}$$

For β , the score function is given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{t=1}^{n_0} \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \gamma_t} \frac{\partial \gamma_t}{\partial \beta_j} + \sum_{t=n_0+1}^m \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \gamma_t} \frac{\partial \gamma_t}{\partial \beta_j} + \sum_{t=m+1}^n \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \gamma_t} \frac{\partial \gamma_t}{\partial \beta_j} + \sum_{t=1}^n \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \mu_t} \frac{\partial \mu_t}{\partial \gamma_t} \frac{\partial \gamma_t}{\partial \beta_j},$$

where $\partial \gamma_t / \partial \beta_j = x_{ij} / g'_3(\gamma_t)$ and $\partial \mu_t / \partial \beta_j = \frac{(1-\alpha_{0t})(1-\alpha_{1t})}{c_t^2 g'_3(\gamma_t)} x_{ij}$. Then we have

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{t=1}^n \left\{ -\frac{1}{(1-\gamma_t) g'_3(\gamma_t)} x_{ij} \mathbb{1}_0(y_t) + \frac{1}{\gamma_t g'_3(\gamma_t)} x_{ij} \mathbb{1}_1(y_t) \right. \\ \left. + \left[\frac{(\alpha_{0t} - \alpha_{1t})}{c_t g'_3(\gamma_t)} x_{ij} + \phi_t \frac{(1-\alpha_{0t})(1-\alpha_{1t})(y_t^* - \mu_t^*)}{c_t^2 g'_3(\gamma_t)} x_{ij} \right] (1 - \mathbb{1}_0(y_t))(1 - \mathbb{1}_1(y_t)) \right\}. \end{aligned}$$

The score function for ζ is given by

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \zeta_j} = \sum_{t=1}^n \frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \phi_t} \frac{\partial \phi_t}{\partial \zeta_j},$$

where $\partial \phi_t / \partial \zeta_j = \check{x}_{ij} / g'_4(\phi_t)$. Therefore,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \zeta_j} = \sum_{t=1}^n \left[\frac{\mu_t(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger)}{g'_4(\phi_t)} \check{x}_{ij} \right] (1 - \mathbb{1}_0(y_t))(1 - \mathbb{1}_1(y_t)).$$

In matrix form, each term of the score vector is given by

$$\begin{aligned} U_\omega(\boldsymbol{\theta}) &= \check{X}^\top T_1 \{J_1 I_0 - [G_2 - RA_2(Y^* - \mathcal{M}^*)] I_2\}, \\ U_\kappa(\boldsymbol{\theta}) &= \check{X}^\top T_2 \{J_2 I_1 - [G_1 + RA_1(Y^* - \mathcal{M}^*)] I_2\}, \\ U_\beta(\boldsymbol{\theta}) &= X^\top T_3 \{-B_2 I_0 + B_1 I_1 + [S_3 + W(Y^* - \mathcal{M}^*)] I_2\}, \\ U_\zeta(\boldsymbol{\theta}) &= \check{X}^\top T_4 M I_2, \end{aligned}$$

where $T_1 = \text{diag}\left\{\frac{1}{g'_1(\alpha_{01})}, \dots, \frac{1}{g'_1(\alpha_{0n})}\right\}$, $T_2 = \text{diag}\left\{\frac{1}{g'_2(\alpha_{11})}, \dots, \frac{1}{g'_2(\alpha_{1n})}\right\}$, $T_3 = \text{diag}\left\{\frac{1}{g'_3(\gamma_1)}, \dots, \frac{1}{g'_3(\gamma_n)}\right\}$, $T_4 = \text{diag}\left\{\frac{1}{g'_4(\phi_1)}, \dots, \frac{1}{g'_4(\phi_n)}\right\}$, $Y^* = \text{diag}(y_1^*, \dots, y_n^*)$, $\mathcal{M}^* = \text{diag}(\mu_1^*, \dots, \mu_n^*)$, $M = \text{diag}\{m_1, \dots, m_n\}$, $m_t = \mu_t(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger)$, $A_1 = \text{diag}\{(1 - \alpha_{01}), \dots, (1 - \alpha_{0n})\}$, $A_2 = \text{diag}\{(1 - \alpha_{11}), \dots, (1 - \alpha_{1n})\}$, $B_1 = \text{diag}\left\{\frac{1}{\gamma_1}, \dots, \frac{1}{\gamma_n}\right\}$, $B_2 = \text{diag}\left\{\frac{1}{1-\gamma_1}, \dots, \frac{1}{1-\gamma_n}\right\}$, $R = \text{diag}\left\{\frac{\phi_1 \gamma_1 (1-\gamma_1)}{c_1^2}, \dots, \frac{\phi_n \gamma_n (1-\gamma_n)}{c_n^2}\right\}$, $J_1 = \text{diag}\left\{\frac{1}{\alpha_{01}}, \dots, \frac{1}{\alpha_{0n}}\right\}$, $J_2 = \text{diag}\left\{\frac{1}{\alpha_{11}}, \dots, \frac{1}{\alpha_{1n}}\right\}$, $W = \text{diag}\left\{\frac{\phi_1 (1-\alpha_{01})(1-\alpha_{11})}{c_1^2}, \dots, \frac{\phi_n (1-\alpha_{0n})(1-\alpha_{1n})}{c_n^2}\right\}$, $G_1 = \text{diag}\left\{\frac{\gamma_1}{c_1}, \dots, \frac{\gamma_n}{c_n}\right\}$,

$G_2 = \text{diag}\left\{\frac{1-\gamma_1}{c_1}, \dots, \frac{1-\gamma_n}{c_n}\right\}$, $S_3 = \text{diag}\left\{\frac{z_{01}-z_{11}}{c_1}, \dots, \frac{z_{0n}-z_{1n}}{c_n}\right\}$, \tilde{X} is a $n \times k_1$ matrix whose t -th row is \tilde{x}_t^\top , \check{X} is a $n \times k_2$ matrix whose t -th row is \check{x}_t^\top , X is a $n \times k_3$ matrix whose t -th row is x_t^\top , \ddot{X} is a $n \times k_4$ matrix whose t -th row is \ddot{x}_t^\top , $I_0 = (\mathbb{1}_0(y_1), \dots, \mathbb{1}_0(y_n))^\top$, $I_1 = (\mathbb{1}_1(y_1), \dots, \mathbb{1}_1(y_n))^\top$, and $I_2 = ((1 - \mathbb{1}_0(y_1))(1 - \mathbb{1}_1(y_1)), \dots, (1 - \mathbb{1}_0(y_n))(1 - \mathbb{1}_1(y_n)))^\top$.

The joint information matrix for the parameter vector $\theta = (\omega^\top, \kappa^\top, \beta^\top, \zeta^\top)^\top$ is given by

$$K(\theta) = \begin{pmatrix} K_{\omega\omega} & K_{\omega\kappa} & K_{\omega\beta} & K_{\omega\zeta} \\ K_{\kappa\omega} & K_{\kappa\kappa} & K_{\kappa\beta} & K_{\kappa\zeta} \\ K_{\beta\omega} & K_{\beta\kappa} & K_{\beta\beta} & K_{\beta\zeta} \\ K_{\zeta\omega} & K_{\zeta\kappa} & K_{\zeta\beta} & K_{\zeta\zeta} \end{pmatrix},$$

where $K_{\omega\omega} = \tilde{X}^\top T_1^2 \left\{ \frac{I_1}{B_2} + [G_2^2 + R^2 A_2^2 V^*] C \right\} \tilde{X}$, $K_{\omega\kappa} = K_{\kappa\omega}^\top = \tilde{X}^\top T_1 T_2 [G_1 G_2 - R^2 A_1 A_2 V^*] C \check{X}$, $K_{\omega\beta} = K_{\beta\omega}^\top = \tilde{X}^\top T_1 T_3 [-S_2 + W A_2 R V^*] C X$, $K_{\omega\zeta} = K_{\zeta\omega}^\top = \tilde{X}^\top T_1 T_4 R A_2 H C \ddot{X}$, $K_{\kappa\kappa} = \check{X}^\top T_2^2 \left\{ \frac{I_2}{B_1} + [G_1^2 + R^2 A_1^2 V^*] C \right\} \check{X}$, $K_{\kappa\beta} = K_{\beta\kappa}^\top = \check{X}^\top T_2 T_3 [S_1 - W A_1 R V^*] C X$, $K_{\kappa\zeta} = K_{\zeta\kappa}^\top = -\check{X}^\top T_2 T_4 R A_1 H C \ddot{X}$, $K_{\beta\beta} = X^\top T_3^2 \left\{ \frac{B_2}{I_1} + \frac{B_1}{I_2} + [S_3^2 + W^2 V^*] C \right\} X$, $K_{\beta\zeta} = K_{\zeta\beta}^\top = X^\top T_3 T_4 W H C \ddot{X}$, $K_{\zeta\zeta} = \ddot{X}^\top T_4^2 D C \ddot{X}$, $S_1 = \text{diag}\left\{\frac{1-z_{01}}{c_1}, \dots, \frac{1-z_{0n}}{c_n}\right\}$, $S_2 = \text{diag}\left\{\frac{1-\alpha_{11}}{c_1}, \dots, \frac{1-\alpha_{1n}}{c_n}\right\}$, $V^* = \text{diag}\{v_1^*, \dots, v_n^*\}$, $v_t^* = \psi'(\mu_t \phi_t) + \psi'((1 - \mu_t) \phi_t)$, $H = \text{diag}\{\mu_1 v_1^* + f_1, \dots, \mu_n v_n^* + f_n\}$, $f_t = -\psi'((1 - \mu_t) \phi_t)$, $D = \text{diag}\{d_1, \dots, d_n\}$, $d_t = \psi'(\phi_t) - \psi'(\mu_t \phi_t) \mu_t^2 - \psi'((1 - \mu_t) \phi_t) (1 - \mu_t)^2$, $C = \text{diag}\{c_1, \dots, c_n\}$, and $c_t = 1 - \alpha_{0t}(1 - \gamma_t) - \alpha_{1t} \gamma_t$.

Supporting information

S1 Dataset.

(TXT)

Author Contributions

Conceptualization: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza, Fábio M. Bayer.

Data curation: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza, Fábio M. Bayer.

Formal analysis: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza, Fábio M. Bayer.

Investigation: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza, Fábio M. Bayer.

Methodology: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza, Fábio M. Bayer.

Resources: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza.

Software: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza.

Writing – original draft: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza, Fábio M. Bayer.

Writing – review & editing: Luiz M. A. Lima-Filho, Tarciana Liberal Pereira, Tatiene C. Souza, Fábio M. Bayer.

References

1. Mandel B. J., (1969). The regression control chart. *Journal of Quality Technology* 1 (1), 1–9.
2. Shu L., Tsung F., Tsui K.-L., (2004). Run-length performance of regression control charts with estimated parameters. *Journal of Quality Technology* 36 (3), 280–292. <https://doi.org/10.1080/00224065.2004.11980274>
3. Jearekpaorn D., Montgomery D., Runger G., Borror C., (2003). Process monitoring for correlated gamma-distributed data using generalized-linear-model-based control charts. *Quality and Reliability Engineering International* 19, 477–491. <https://doi.org/10.1002/qre.521>
4. Jearekpaorn D., Montgomery D., Runger G., Borror C., (2005). Model-based process monitoring using robust generalized linear models. *International Journal of Production Research* 43 (7), 1337–1354. <https://doi.org/10.1080/00207540412331299693>
5. Jearekpaorn D., Borror C., Runger G., Montgomery D., (2007). Process monitoring for mean shifts for multiple stage processes. *International Journal of Production Research* 45, 5547–5570. <https://doi.org/10.1080/00207540701325371>
6. Bayer F., Tondolo C., Müller F., (2018). Beta regression control chart for monitoring fractions and proportions. *Computers & Industrial Engineering* 119 (1), 416–426. <https://doi.org/10.1016/j.cie.2018.04.006>
7. Bayer F. M., Cribari-Neto F., (2017). Model selection criteria in beta regression with varying dispersion. *Communications in Statistics—Simulation and Computation* 46 (1), 729–746. <https://doi.org/10.1080/03610918.2014.977918>
8. Ospina R., Ferrari S. L. P., (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis* 56, 1609–1623. <https://doi.org/10.1016/j.csda.2011.10.005>
9. Hoof A., (2007). Second stage DEA: comparison of approaches for modelling the DEA score. *European Journal of Operational Research* 181, 425–435. <https://doi.org/10.1016/j.ejor.2006.05.019>
10. Cook D. O., Kieschnick R., McCullough B. D., (2008). Regression analysis of proportions in finance with self selection. *Journal of Empirical Finance* 15 (5), 860–867. <https://doi.org/10.1016/j.jempfin.2008.02.001>
11. Ferrari S., Cribari-Neto F., (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31 (7), 799–815. <https://doi.org/10.1080/0266476042000214501>
12. Bayes C. L., Valdivieso L., (2016). A beta inflated mean regression model for fractional response variables. *Journal of Applied Statistics* 43 (10), 1814–1830. <https://doi.org/10.1080/02664763.2015.1120711>
13. Teyarachakul S., Chand S., Tang J., (2007). Estimating the limits for statistical process control charts: A direct method improving upon the bootstrap. *European Journal of Operational Research* 178, 472–481. <https://doi.org/10.1016/j.ejor.2006.01.038>
14. Capizzi G., Masarotto G., (2011). A least angle regression control chart for multidimensional data. *Technometrics* 53 (3), 285–296. <https://doi.org/10.1198/TECH.2011.10027>
15. Smyth G. K., Verbyla A. P., (1999). Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics* 10 (6), 695–709. [https://doi.org/10.1002/\(SICI\)1099-095X\(199911/12\)10:6%3C695::AID-ENV385%3E3.0.CO;2-M](https://doi.org/10.1002/(SICI)1099-095X(199911/12)10:6%3C695::AID-ENV385%3E3.0.CO;2-M)
16. Rahimi S. B., Amiri A., Ghashghaei R., (2019). Simultaneous monitoring of mean vector and covariance matrix of multivariate simple linear profiles in the presence of within profile autocorrelation. *Communications in Statistics - Simulation and Computation* Accepted.
17. Yeh A. B., Huwang L., Li Y.-M., (2009). Profile monitoring for a binary response. *IIE Transactions* 41 (11), 931–941. <https://doi.org/10.1080/07408170902735400>
18. Shang Y., Tsung F., Zou C., 7 2011. Profile monitoring with binary data and random predictors. *Journal of Quality Technology* 43, 196–208. <https://doi.org/10.1080/00224065.2011.11917857>
19. Zhang J., Ren H., Yao R., Zou C., Wang Z., (2015). Phase I analysis of multivariate profiles based on regression adjustment. *Computers & Industrial Engineering* 85, 132–144. <https://doi.org/10.1016/j.cie.2015.02.025>

20. Qi D., Wang Z., Zi X., Li Z., (2016). Phase ii monitoring of generalized linear profiles using weighted likelihood ratio charts. *Computers & Industrial Engineering* 94, 178–187. <https://doi.org/10.1016/j.cie.2016.01.022>
21. Canterle D. R., Bayer F. M., (2019). Variable dispersion beta regressions with parametric link functions. *Statistical Papers* 60, 1541–1567. <https://doi.org/10.1007/s00362-017-0885-9>
22. Gan F. F., (1995). Joint monitoring of process mean and variance using exponentially weighted moving average control charts. *Technometrics* 37 (4), 446–453. <https://doi.org/10.1080/00401706.1995.10484377>
23. Riaz M., (2008). Monitoring process variability using auxiliary information. *Computational Statistics* 23 (2), 253–276. <https://doi.org/10.1007/s00180-007-0084-6>
24. Sheu S.-H., Tai S.-H., Hsieh Y.-T., Lin T.-C., (2009). Monitoring process mean and variability with generally weighted moving average control charts. *Computers & Industrial Engineering* 57 (1), 401–407. <https://doi.org/10.1016/j.cie.2008.12.010>
25. Huwang L., Huang C.-J., Wang Y.-H.T., (2010). New EWMA control charts for monitoring process dispersion. *Computational Statistics & Data Analysis* 54 (10), 2328–2342. <https://doi.org/10.1016/j.csda.2010.03.011>
26. Pawitan Y., (2001). In *All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford Science publications.
27. Casella G., Berger R. L., (2002). *Statistical Inference*, 2nd Edition. Thomson Learning.
28. Chen F., Chen S., Ma X., (2018). Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. *Journal of Safety Research* 65, 153–159. <https://doi.org/10.1016/j.jsr.2018.02.010> PMID: 29776524
29. Press W., Teukolsky S., Vetterling W., Flannery B., (1992). *Numerical recipes in C: The art of scientific computing*, 2nd Edition. Cambridge University Press.
30. Lima-Filho L. M. A., Pereira T. L., de Souza T. C., Bayer F. M., (2019). Inflated beta control chart for monitoring double bounded processes. *Computers & Industrial Engineering* 136, 265–276. <https://doi.org/10.1016/j.cie.2019.07.017>
31. Boon Chong M., 1 2004. Performance Measures for the Shewhart Control Chart. *Quality Engineering* 16 (4), 585–590. <https://doi.org/10.1081/QEN-120038020>
32. Schaffer J. R., Kim M.-J., (2007). Number of replications required in control chart Monte Carlo simulation studies. *Communications in Statistics - Simulation and Computation* 36 (5), 1075–1087. <https://doi.org/10.1080/03610910701539963>
33. R Core Team, (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
34. Simas A. B., Souza W. B., Rocha A. V., (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis* 54 (2).
35. Ho L. L., Fernandes F. H., Bourguignon M., (2019). Control charts to monitor rates and proportions. *Quality and Reliability Engineering International* 35 (1), 74–83. <https://doi.org/10.1002/qre.2381>
36. Montgomery D. C., (2009). *Introduction to Statistical Quality Control*, 6th Edition. John Wiley & Sons.
37. Moraes D. A. O., Oliveira F., Quinino R., Duczmal L., (2014). Self-oriented control charts for efficient monitoring of mean vectors. *Computers & Industrial Engineering* 75, 102–115. <https://doi.org/10.1016/j.cie.2014.06.008>
38. Paroissin C., Penalva L., Pétrau A., Verdier G., (2016). New control chart for monitoring and classification of environmental data. *Environmetrics* 27, 182–193. <https://doi.org/10.1002/env.2382>
39. Lima-Filho L. M. A., Bayer F. M., (2019). Kumaraswamy control chart for monitoring double bounded environmental data. *Communications in Statistics - Simulation and Computation* Accepted.
40. Sousa M. C. S., Cribari-Neto F., Stosic B. D., Nov. 2005. Explaining DEA Technical Efficiency Scores in an Outlier Corrected Environment: The Case of Public Services in Brazilian Municipalities. *Brazilian Review of Econometrics* 25 (2), 287–313. <https://doi.org/10.12660/bre.v25n22005.2507>
41. Pereira T. L., Cribari-Neto F., (2014). Detecting model misspecification in inflated beta regressions. *Communications in Statistics—Simulation and Computation* 43 (3), 631–656. <https://doi.org/10.1080/03610918.2012.712183>