Robust Inference under the Beta Regression Model with Application to Health Care Studies

Abhik Ghosh
Indian Statistical Institute, Kolkata, India
abhianik@gmail.com

Abstract

Data on rates, percentages or proportions arise frequently in many applied disciplines like medical biology, health care, psychology and several others. In this paper, we develop a robust inference procedure for the beta regression model which is used to describe such response variables taking values in $(0, 1)$ through some related explanatory variables. The existing maximum likelihood based inference has serious lack of robustness against outliers in data and generate drastically different (erroneous) inference in presence of data contamination. Here, we develop the robust minimum density power divergence estimator and a class of robust Wald-type tests for the beta regression model along with several applications. We derive their asymptotic properties and describe their robustness theoretically through the influence function analyses. Finite sample performances of the proposed estimators and tests are examined through suitable simulation studies and real data applications in the context of health care and psychology.

Keywords: Robustness; Beta Regression Model; Rates and Proportions Data; Minimum Density Power Divergence Estimator; Wald-Type Tests.

1 Introduction

In many biological experiments, medical researches including health care studies and psychology, survey researches in sociology and marketing, and several other applied sciences, we often come across data on rates, ratios, percentages or proportions, taking values in the unit interval $(0, 1)$. Examples of such data include the "body fat percentages" or any similar health condition measured in percentage, health assessment questionnaire (HAQ) data or similar rating data, accuracy percentage of a treatment in clinical trial, experimental scores measuring stress, depression, etc., in psychology, proportion of a certain group of patients (for some particular disease) in a region and many more. Such data can be modeled individually by a beta distribution having support $(0, 1)$. However, in order to better understand such variables and the underlying data-generating mechanism with more detailed inference, it is often required to relate its values with some other associated explanatory variables through a suitable regression structure will also enable us to do prediction. The beta regression model is designed to help in this situation, which model a response variable y taking values in $(0, 1)$ through any set of explanatory variables $\boldsymbol{x}$.

There are several recent specifications of the beta regression model; for example, see Paolino (2001), Kieschnick and McCullough (2003), Ferrari and Cribari-Neto (2004) and Vasconcellos and

Cribari-Neto (2005), among others. In this paper, we follow the most popular specification provided by Ferrari and Cribari-Neto (2004). This is because (i) it models the "mean" of such response variables on $(0, 1)$ to depend on a linear combination of available covariates through a suitable link function, (ii) the specification is closely related to the popular class of generalized linear model (McCullough and Nelder, 1989) and similar theories follow, (iii) it allows many different possible link functions to model various structures within the data, and (iv) the inference methodologies are well developed for this specification and are available in standard statistical software R (package '*betareg*') for practitioners.

Mathematically, suppose $y_1, \ldots, \ y_n$ are n independent responses each taking value in $(0, 1)$ and are associated with p-dimensional covariate values $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ respectively. Then, in the beta regression model (BRM) of Ferrari and Cribari-Neto (2004), each y_i is assumed to follow a beta distribution having density

$$f\left(y_i; \mu_i, \ \phi\right) = \frac{1}{B\left(\mu_i \phi, \ \left(1 - \mu_i\right) \phi\right)} \, y_i^{\mu_i \phi - 1} (1 - y_i)^{(1 - \mu_i)\phi - 1}, \qquad 0 < y_i < 1, \tag{1}$$

where $E\left(y_i\right) = \mu_i \in (0, 1)$ is related to the (given) value of the i-th explanatory variable $\boldsymbol{x}_i$ through a suitable link function g (defined on $(0, 1)$), ϕ is the precision parameter (since $Var\left(y_i\right) = \frac{\mu_i(1-\mu_i)}{1+\phi}$) and $B(\cdot, \cdot)$ denotes the (complete) beta function. Given g, we assume the regression structure $g\left(\mu_i\right) = \boldsymbol{x}_i^T \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the vector of unknown regression coefficients. Our objective then is to make inference about the parameter of interest $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ based on the available data $\{(y_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$.

This beta regression model has become very useful in many recent applications, since it can also be applied easily to the data within any finite intervals. If y takes values in any other open interval, say (a, b), we can apply the BRM in (1) with the transformed response $\frac{y-a}{b-a}$ which now takes values in $(0, 1)$. Further, if the response y also takes the value in the end-points 0 and 1, other than the use of sophisticated and complicated modifications, we can indeed apply the simpler BRM (1) easily with the widely used ad-hoc transformation $\frac{1}{n}[y(n-1) + 0.5]$, n being the sample size (Smithson and Verkuilen, 2006).

The existing inference procedures under the BRM (1) are primarily based on the classical maximum likelihood approach. The point estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ is obtained by maximizing the likelihood function with respect to $\boldsymbol{\theta}$, generating the maximum likelihood estimator (MLE), and any hypothesis testing problem can be solved by the likelihood ratio test or the Wald test based on the MLE; see Ferrari and Cribari-Neto (2004) for more details. The R package '*betareg*' also provides the inferential solution for the BRM (1) based on this standard maximum likelihood approach, which satisfies many asymptotic optimality properties. However, a serious problem with the maximum likelihood based inference is the high degree of sensitivity to potential outliers in the data. This lack of robustness often leads to drastically different (erroneous) inference in presence of even a small amount of data contamination. This non-robustness issue naturally affects the results also for the BRM while using the existing inference methodologies based on the likelihood principle in the presence of outliers. Since such outliers are not uncommon in practical datasets, we need to be very cautious before using the maximum likelihood based inference (and also using the R package '*betareg*'). To illustrate this issue, let us present a motivating example from an Australian health care study.

A Motivating Example (AIS Data):
Consider the data on health measurements of several athletes collected at the Australian Institute of Sport (AIS) which is publicly available in the R package "*sn*" (https://cran.r-project.org/web/packages/sn). Bayes et al. (2012) have recently studied the data corresponding to the 37 rowing athletes and tried to predict their body fat percentages (BFP) from their lean body masses (LBM) using Bayesian inference. Since the predictor variable BFP takes values within $(0, 1)$, a beta regression model can be used for this purpose.

Let us apply the BRM (1) with response y = BFP, covariate $\boldsymbol{x}_i = (1, \text{LBM})^T$ and a logit link function, $logit(E[\text{BFP}]) = \beta_1 + \beta_2 LBM$. Then, applying the existing maximum likelihood procedure using '*betareg*', the MLE of the parameter of interest $\boldsymbol{\theta} = (\beta_1, \beta_2, \phi)^T$ is given by $(0.097, -0.027, 95.472)^T$. Further, applying the existing Wald test based on this MLE, the p-values of the significance of two regression coefficients turn out to be 0.699 and 0 respectively, which indicates that the intercept component (β_1) is not significant in the model.

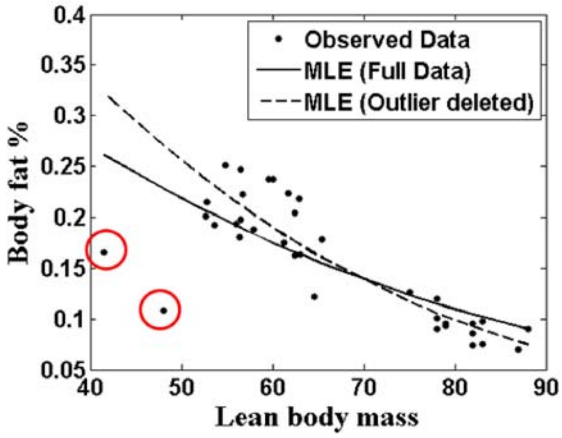


Figure 1: The AIS Data along with the fitted lines based on the MLE for the full data and the outlier deleted data. (The two outlying observations are marked with red circles).

However, by plotting the data (see Figure 1), one can clearly see that there are two outlying observations as also noticed by Bayes et al. (2012); the fitted line based on the above MLE does not yield a good fit to the bulk of the data in the presence of these two outliers. In fact, if we again compute the MLE of the parameter $\boldsymbol{\theta}$ after removing these two outliers, the resulting estimate becomes $(0.838, -0.038, 246.305)^T$, which drastically differs from the previous MLE based on the full data. The change in the fitted line is clearly visible in Figure 1 and the estimate of ϕ changes substantially! Further, after deleting these two outliers, both the p-values of the MLE based Wald test for testing the significance of the regression coefficients become 0. Thus, only these two outliers completely hide the significance of β_1 generating opposite inference. $\square$

As we have seen in the above example, few outliers in a dataset can lead to completely wrong inference through the existing likelihood procedures under the BRM. Several other authors have also recently noticed this non-robust behavior of the MLE and the instability of the related inferences against the outlying observations (Ferrari and Cribari-Neto, 2004; Espinheira et al., 2008a,b); they have developed some diagnostic tools to identify such influential observations or outliers in a BRM and suggested to delete them before doing maximum likelihood based inference. Although this solution with the prior outlier detection is feasible for the simple and small datasets, it is quite difficult and needs several additional analyses for most complicated datasets including the big or high-dimensional datasets of recent era. A robust inference procedure that can automatically take care of these outliers to successfully yield stable results are much more logical, efficient and useful in all such complicated cases. However, unlike other inferential set-ups, there exists no such robust inference procedure for the recently developed beta regression model. The only related work is the one by Bayes et al. (2012) who have proposed to solve this issue for the BRM under Bayesian paradigm through the use of a modified distribution in place of the simple beta distribution; but it does not really address the non-robustness problem of the MLE based inference with respect to the simpler specified model (1).

In this paper, we develop a robust inference procedure for the BRM (1) without changing its original distributional form. Among several approaches of robust inference, we follow the minimum divergence approach where we quantify the discrepancy between data and the parametric model through a statistical divergence measure and minimize it to estimate the unknown parameters. In particular, we consider the density power divergence (DPD) of Basu et al. (1998), because the resulting estimator has become very popular in recent times due to its high asymptotic efficiency along with strong robustness properties. It has also been applied to many real life inference problems; see Section 2.1 and Basu et al. (2011) for additional details. We develop the robust minimum density power divergence estimator (MDPDE) for the BRM (1) along with its asymptotic properties in Section 2. Based on the proposed MDPDE, we develop a robust Wald-type hypothesis testing procedure in Section 3 and derive its asymptotic properties. We also theoretically illustrate the robustness of both the proposed estimator and the testing procedure through suitable influence function analyses. In Section 4, our proposed methods are applied to reanalyze the motivating example along with two additional real data examples from different health-care studies (including psychology). Finite sample performances of the proposed inference are examined through suitable simulation studies in Section 5. Finally the paper ends with some concluding remarks in Section 6.

2 Robust Minimum Density Power Divergence Estimators

2.1 Background

The density power divergence (DPD) measure between two densities f_1 and f_2 (with respect to some common dominating measure) is defined in terms of a tuning parameter $\alpha \geq 0$ (Basu et al., 1998) as

$$d_\alpha(f_1, f_2) = \int f_2^{1+\alpha} - \frac{1+\alpha}{\alpha}\int f_1 f_2^\alpha + \frac{1}{\alpha}\int f_1^{1+\alpha}, \qquad \text{if } \alpha > 0; \tag{2}$$

$$d_0(f_1, f_2) = \lim_{\alpha \to 0} d_\alpha(f_1, f_2) = \int f_1 \log\left(\frac{f_1}{f_2}\right). \tag{3}$$

Note that the DPD measure at $\alpha = 0$ coincides with the famous likelihood disparity, minimization of which is known to be equivalent to the maximum likelihood approach. The DPD family connects the Likelihood disparity (at $\alpha = 0$) to the L_2-Divergence (at $\alpha = 1$) smoothly through the tuning parameter α. For the sake of completeness and a better understanding, let us start by recalling the minimum DPD estimation under the independent and identically distributed (iid) set-up.

For n iid observations $Y_1, \ldots, Y_n$ modeled by a parametric family $\mathcal{F}_{\boldsymbol{\theta}} = \left\{ f_{\boldsymbol{\theta}} \, : \, \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k \right\}$, the minimum DPD estimator (MDPDE) is obtained by minimizing the estimated DPD measure (2) between the observed data (at f_1) and the model density $f_{\boldsymbol{\theta}}$ (at f_2), or, equivalently by minimizing the quantity

$$\int f_{\boldsymbol{\theta}}^{1+\alpha} - \frac{1+\alpha}{\alpha} \int f_{\boldsymbol{\theta}}^{\alpha} dG_n = \int f_{\boldsymbol{\theta}}^{1+\alpha} - \frac{1+\alpha}{\alpha} \frac{1}{n} \sum_{i=1}^{n} f_{\boldsymbol{\theta}}^{\alpha} \left(Y_i \right),$$

with G_n being the empirical distribution function based on the observed data (Basu et al., 1998, 2011). Under suitable differentiability assumptions, the estimating equation of $\boldsymbol{\theta}$ is given by

$$\frac{1}{n} \sum_{i=1}^{n} u_{\boldsymbol{\theta}} \left(Y_i \right) f_{\boldsymbol{\theta}}^{\alpha} \left(Y_i \right) - \int u_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}^{1+\alpha} = 0.$$

Note that, at $\alpha = 0$, this MDPDE estimating equation coincides with the estimating (score) equation of the MLE and so MDPDE at $\alpha = 0$ is nothing but the MLE. The MDPDE at $\alpha > 0$ yields a generalization of the MLE which down-weights the effect of the outlying observations in the estimating equation by α-th power of the model density and hence is expected to be more robust. This MDPDE has become very popular in recent days, because (i) it does not need non-parametric kernel estimation unlike many other divergences, (ii) it generates robust estimator having high asymptotic efficiency at properly chosen α, and (ii) it can be obtained from a simple unbiased estimating equation along with an underlying objective function which helps to avoid the problem of multiple roots.

However, in general, our data for the BRM (1) are NOT iid, and hence the above approach cannot be applied directly. This is because we generally do not make any distributional assumptions on the covariates $\boldsymbol{x}_i$s and treat them as fixed (given) so that, for each i, $y_i \sim Beta\left(\mu_i \phi, \ (1-\mu_i)\,\phi\right)$ with density given by (1). Thus, each y_i is independent but not identically distributed. Recently, Ghosh and Basu (2013) have proposed an extension of the MDPDE for the general independent but non-homogeneous set-up by considering the average DPD measure over different distributions. Ghosh and Basu (2016) applied this extended approach to develop robust inferences for a simple class of canonical generalized linear models (GLMs) with fixed designs including normal, Poisson and logistic regressions; Ghosh (2016) has also applied it to an exponential regression model to propose a robust estimator of the tail index. However, unfortunately, the class of GLMs considered in Ghosh and Basu (2016) does not directly cover our BRM (1). So, in this paper, we further extend this approach to develop a robust estimator for the BRM (1) with fixed covariates (design).

2.2 The MDPDE for the Beta regression Model

Consider the beta regression model (BRM) set-up as decried in Section 1. Let us assume that the responses $y_1, \ldots, y_n$ are independent but $y_i \sim g_i$ for each $i = 1, \ldots, n$, where g_is are potentially different true densities of y_is depending on $\boldsymbol{x}_i$s. We model g_i by the BRM given by (1), i.e., by the model density $f_i\left(\cdot, \boldsymbol{\theta}\right) \equiv Beta\left(\mu_i \phi, (1-\mu_i)\,\phi\right)$ density. The unknown parameter of interest is $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^T, \phi\right)^T$ which is common across the densities. Following Ghosh and Basu (2013), we define the Minimum DPD Estimator (MDPDE) of $\boldsymbol{\theta}$ under the BRM (1) as the minimizer of the average DPD measure with tuning parameter $\alpha \geq 0$ given by

$$n^{-1} \sum_{i=1}^{n} d_\alpha \left(\widehat{g}_i \left(\cdot \right), f_i \left(\cdot, \boldsymbol{\theta} \right) \right), \tag{4}$$

where $\widehat{g}_i$ is an estimate of g_i based on the given data. Since the DPD measure is a proper statistical divergence, the resulting minimizer is clearly Fisher consistent for $\boldsymbol{\theta}$. For the present case of BRM (1), since we have only one observation from each g_i, a simple estimate of it is given by the degenerate distribution at y_i for any $i = 1, \ldots, n$. Hence, after some simplification, the minimizer of (4) is seen to be the minimizer of the simpler objective function

$$H_{n,\alpha}(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \left[K_{i,\alpha}(\boldsymbol{\theta}) - \frac{1+\alpha}{\alpha} f_i(y_i, \boldsymbol{\theta})^\alpha \right],$$

with $K_{i,\alpha}(\boldsymbol{\theta}) = \frac{B((1+\alpha)\mu_i\phi - \alpha, (1+\alpha)(1-\mu_i)\phi - \alpha)}{B(\mu_i\phi, (1-\mu_i)\phi)^\alpha}$. We need to minimize this objective function $H_{n,\alpha}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ to obtain its MDPDE with tuning parameter α, say $\widehat{\boldsymbol{\theta}}_{n,\alpha} = (\widehat{\boldsymbol{\beta}}_{n,\alpha}^T, \widehat{\phi}_{n,\alpha})^T$. Note that, the above objective function $H_{n,\alpha}(\boldsymbol{\theta})$ becomes [1 − log-likelihood] as $\alpha \to 0$ and hence the proposed MDPDE at $\alpha = 0$ coincides with the usual MLE of Ferrari and Cribari-Neto (2004) which is known to be non-robust but fully efficient. Further, it can be seen that the MDPDE at $\alpha = 1$ coincides with the minimum L_2-distance estimator which is known to be highly robust but inefficient under any general model. Hence the tuning parameter α in the proposed MDPDE under the BRM is expected to yield a trade-off between robustness and efficiency of the estimator.

Equivalently, we can also obtain the MDPDE $\widehat{\boldsymbol{\theta}}_{n,\alpha} = (\widehat{\boldsymbol{\beta}}_{n,\alpha}^T, \widehat{\phi}_{n,\alpha})^T$ by solving the estimating equations obtained by differentiating the objective function $H_{n,\alpha}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$. For the BRM, these estimating equations simplify to

$$\sum_{i=1}^{n} \left[\gamma_{1,i}^{(\alpha)}(\boldsymbol{\theta}) - \left(y_{1,i}^* - \mu_{1,i}^* \right) \frac{\phi}{g'(\mu_i)} f_i(y_i, \boldsymbol{\theta})^\alpha \right] \boldsymbol{x}_i = 0, \tag{5}$$

$$\sum_{i=1}^{n} \left[\gamma_{2,i}^{(\alpha)}(\boldsymbol{\theta}) - \left\{ \mu_i \left(y_{1,i}^* - \mu_{1,i}^* \right) + \left(y_{2,i}^* - \mu_{2,i}^* \right) \right\} \frac{\phi}{g'(\mu_i)} f_i(y_i, \boldsymbol{\theta})^\alpha \right] = 0, \tag{6}$$

where g' denotes the derivative of g,

$$y_{1,i}^* = \log \frac{y_i}{1-y_i}, \qquad \mu_{1,i}^* = E\left(y_{1,i}^*\right) = \psi\left(\mu_i\phi\right) - \psi\left(\left(1-\mu_i\right)\phi\right)$$

$$y_{2,i}^* = \log\left(1-y_i\right), \qquad \mu_{2,i}^* = E\left(y_{2,i}^*\right) = \psi\left(\left(1-\mu_i\right)\phi\right) - \psi\left(\phi\right)$$

and

$$\begin{aligned}
\gamma_{1,i}^{(\alpha)}(\boldsymbol{\theta}) &= \left(\psi\left(a_{i,\alpha}\right) - \psi\left(b_{i,\alpha}\right) - \mu_{1,i}^*\right) \frac{\phi K_{i,\alpha}(\boldsymbol{\theta})}{g'(\mu_i)} \\
\gamma_{2,i}^{(\alpha)}(\boldsymbol{\theta}) &= \left[\mu_i \left(\psi\left(a_{i,\alpha}\right) - \psi\left(b_{i,\alpha}\right) - \mu_{1,i}^*\right) + \left(\psi\left(b_{i,\alpha}\right) - \psi\left(a_{i,\alpha} + b_{i,\alpha}\right) - \mu_{2,i}^*\right)\right] K_{i,\alpha}(\boldsymbol{\theta})
\end{aligned}$$

with $a_{i,\alpha} = (1+\alpha)\,\mu_i\phi - \alpha$, $b_{i,\alpha} = (1+\alpha)(1-\mu_i)\phi - \alpha$ and $\psi(\cdot)$ being the digamma function. Clearly the estimating equations are unbiased at the model for any $\alpha \geq 0$. Also, at $\alpha = 0$, we have $\gamma_{1,i}^{(\alpha)}(\boldsymbol{\theta}) = 0 = \gamma_{2,i}^{(\alpha)}(\boldsymbol{\theta})$ for all $i = 1, \ldots, n$ and these MDPDE estimating equations then coincide with the MLE estimating (score) equations as expected.

The asymptotic distribution of this proposed MDPDE can be derived from a general results of Ghosh and Basu (2013) under Assumptions (A1)–(A7) of their paper. In particular, whenever the model assumption (1) holds with true parameter value $\boldsymbol{\theta}_0$, i.e., $g_i(\cdot) = f_i(\cdot, \boldsymbol{\theta}_0)$ for all i, we have the following:

1. There exists a consistent sequence $\widehat{\boldsymbol{\theta}}_{n,\alpha}$ of roots to the estimating equations (5) and (6) of the MDPDE.
2. Asymptotically $\boldsymbol{\Omega}_n(\boldsymbol{\theta}_0)^{-1/2}\boldsymbol{\Psi}_n(\boldsymbol{\theta}_0)\left[\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{n,\alpha} - \boldsymbol{\theta}_0\right)\right] \sim N_{p+1}\left(\mathbf{0}_{p+1}, \boldsymbol{I}_{p+1}\right)$,

where $\mathbf{0}_{p+1}$ is the zero vector of length (0), $\boldsymbol{I}_{p+1}$ is identity matrix of order $*(p+1)$ and

$$\begin{aligned}
\boldsymbol{\Psi}_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \gamma_{11,i}^{(\alpha)}(\boldsymbol{\theta})\,\boldsymbol{x}_i\boldsymbol{x}_i^T & \gamma_{12,i}^{(\alpha)}(\boldsymbol{\theta})\,\boldsymbol{x}_i \\ \gamma_{12,i}^{(\alpha)}(\boldsymbol{\theta})\,\boldsymbol{x}_i^T & \gamma_{22,i}^{(\alpha)}(\boldsymbol{\theta}) \end{bmatrix}, \\
\boldsymbol{\Omega}_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \left\{\gamma_{11,i}^{(2\alpha)}(\boldsymbol{\theta}) - \gamma_{1,i_0}^{(\alpha)}(\boldsymbol{\theta})^2\right\}\boldsymbol{x}_i\boldsymbol{x}_i^T & \left\{\gamma_{12,i}^{(2\alpha)}(\boldsymbol{\theta}) - \gamma_{1,i_0}^{(\alpha)}(\boldsymbol{\theta})\,\gamma_{2,i_0}^{(\alpha)}(\boldsymbol{\theta})\right\}\boldsymbol{x}_i \\ \left\{\gamma_{12,i}^{(2\alpha)}(\boldsymbol{\theta}) - \gamma_{1,i_0}^{(\alpha)}(\boldsymbol{\theta})\,\gamma_{2,i_0}^{(\alpha)}(\boldsymbol{\theta})\right\}\boldsymbol{x}_i^T & \left\{\gamma_{22,i}^{(2\alpha)}(\boldsymbol{\theta}) - \gamma_{2,i_0}^{(\alpha)}(\boldsymbol{\theta})^2\right\} \end{bmatrix},
\end{aligned}$$

with explicit forms of $\gamma_{jk,i}^{(\alpha)}(\boldsymbol{\theta})$ being given by

$$\begin{aligned}
\gamma_{11,i}^{(\alpha)}(\boldsymbol{\theta}) &= \frac{\phi^2 K_{i,\alpha}(\boldsymbol{\theta})}{g'(\mu_i)^2}\left[\psi_1(a_{i,\alpha}) + \psi_1(b_{i,\alpha}) + (\psi(a_{i,\alpha}) - \psi(b_{i,\alpha}) - \mu_{1,i}^*)^2\right] \\
\gamma_{12,i}^{(\alpha)}(\boldsymbol{\theta}) &= \frac{\phi K_{i,\alpha}(\boldsymbol{\theta})}{g'(\mu_i)}\left[\mu_i\left\{\psi_1(a_{i,\alpha}) + \psi_1(b_{i,\alpha}) + (\psi(a_{i,\alpha}) - \psi(b_{i,\alpha}) - \mu_{1,i}^*)^2\right\}\right. \\
&\qquad \left. + \left\{-\psi_1(b_{i,\alpha}) + (\psi(a_{i,\alpha}) - \psi(b_{i,\alpha}) - \mu_{1,i}^*)(\psi(b_{i,\alpha}) - \psi(a_{i,\alpha} + b_{i,\alpha}) - \mu_{2,i}^*)\right\}\right] \\
\gamma_{22,i}^{(\alpha)}(\boldsymbol{\theta}) &= K_{i,\alpha}(\boldsymbol{\theta})\left[\mu_i^2\left\{\psi_1(a_{i,\alpha}) + \psi_1(b_{i,\alpha}) + (\psi(a_{i,\alpha}) - \psi(b_{i,\alpha}) - \mu_{1,i}^*)^2\right\}\right. \\
&\qquad + 2\mu_i\left\{-\psi_1(b_{i,\alpha}) + (\psi(a_{i,\alpha}) - \psi(b_{i,\alpha}) - \mu_{1,i}^*)(\psi(b_{i,\alpha}) - \psi(a_{i,\alpha} + b_{i,\alpha}) - \mu_{2,i}^*)\right\} \\
&\qquad \left. + \left\{\psi_1(b_{i,\alpha}) - \psi_1(a_{i,\alpha} + b_{i,\alpha}) + (\psi(b_{i,\alpha}) - \psi(a_{i,\alpha} + b_{i,\alpha}) - \mu_{2,i}^*)^2\right\}\right],
\end{aligned}$$

and ψ_1 being the trigamma function. The required conditions (A1)–(A7) of Ghosh and Basu (2013) can be verified to hold under mild boundedness conditions on the given covariate values (fixed design). However, the form of the above asymptotic variance matrix indicates that, given any fixed design, the asymptotic relative efficiency of the proposed MDPDE decreases as α increases but this loss in efficiency is not significant at small positive values of α. We will verify this property empirically again in Section 5; but this small loss in efficiency leads to increased robustness of the proposed estimator over the non-robust MLE which we justify through the influence function analysis in the next subsection.

2.3 Influence Function of the MDPDE under the BRM

The influence function is a classical tool to measure the theoretical robustness property of any estimator under the iid set-up (Hampel et al., 1986). It measures the asymptotic bias due to infinitesimal contamination in the data. The concept has been suitably extended and applied to the cases of non-homogeneous observations by Huber (1983), Ghosh and Basu (2013, 2016) and Aerts and Haesbroeck (2016), where the corresponding statistical functional and the influence function both depend on the sample size n (unlike the iid case). Note that, for such non-homogeneous cases the contamination can be in any of the distributions indexed by i or in all of them. We use this concept to illustrate the robustness of our proposed MDPDE under the BRM.

Assuming G_i to be the true distribution function of y_i corresponding to the density g_i for each i, the statistical functional corresponding to the MDPDE of $\boldsymbol{\theta}$ under the BRM (1) is defined as

$$\boldsymbol{T}_\alpha(G_1, \ldots, G_n) = \arg\min_{\boldsymbol{\theta}} n^{-1} \sum_{i=1}^{n} d_\alpha \left(g_i(\cdot), f_i\left(\cdot, \boldsymbol{\theta}\right)\right), \tag{7}$$

whenever the minimum exists. This is a Fisher consistent functional at the assumed BRM by the definition of the DPD measure. Suppose first, for simplicity, the contamination is in only the i_0-th distribution through $G_{i_0,\epsilon} = (1-\epsilon)\, G_{i_0} + \epsilon \wedge_{t_{i_0}}$, where ϵ is the contamination proportion and $\wedge_{t_{i_0}}$ is the degenerate distribution at the contamination point t_{i_0}. The corresponding (first order) influence function (IF) of the proposed MDPDE functional $\boldsymbol{T}_\alpha$ is defined as

$$\begin{aligned} \mathcal{IF}\left(t_{i_0}, \boldsymbol{T}_\alpha; G_1, \ldots, G_n\right) &= \left| \frac{\partial \boldsymbol{T}_\alpha(G_1, \ldots, G_{i_0,\epsilon}, \ldots, G_n)}{\partial \epsilon} \right|_{\epsilon=0} \\ &= \lim_{\epsilon \downarrow 0} \frac{\boldsymbol{T}_\alpha(G_1, \ldots, G_{i_0,\epsilon}, \ldots, G_n) - \boldsymbol{T}_\alpha(G_1, \ldots, G_n)}{\epsilon}. \end{aligned}$$

Note that, whenever this IF is bounded in t_{i_0} the asymptotic bias due to infinitesimal contamination at G_{i_0} remains bounded implying the robustness of the corresponding estimator. On the other hand, if this IF is unbounded in t_{i_0}, then the same bias may tend to infinity for distant contaminations implying the non-robust nature of the estimator.

For our beta regression model with $g_i(\cdot) = f_i\left(\cdot, \boldsymbol{\theta}\right)$ for all i, some calculations yield the simplified form of the above IF as given by

$$\mathcal{IF}\left(t_{i_0}, \boldsymbol{T}_\alpha; F_1, \ldots, F_n\right) = \boldsymbol{\Psi}_n\left(\boldsymbol{\theta}\right)^{-1} \begin{bmatrix} \left(t^*_{1,i_0} - \mu^*_{1,i_0}\right) \frac{\phi}{g'(\mu_{i_0})} f_{i_0}(t_{i_0}, \boldsymbol{\theta})^\alpha - \gamma^{(\alpha)}_{1,i_0}\left(\boldsymbol{\theta}\right) \\ \left\{\mu_i \left(t^*_{1,i_0} - \mu^*_{1,i_0}\right) + \left(t^*_{2,i_0} - \mu^*_{2,i_0}\right)\right\} f_{i_0}(t_{i_0}, \boldsymbol{\theta})^\alpha - \gamma^{(\alpha)}_{2,i_0}\left(\boldsymbol{\theta}\right) \end{bmatrix},$$

where $t^*_{1,i_0} = \log \frac{t_{i_0}}{1-t_{i_0}}$, $t^*_{2,i} = \log\left(1 - t_{i_0}\right)$ and F_i is the distribution function of $f_i(\cdot, \boldsymbol{\theta})$ for each $i = 1, \ldots, n$. Clearly this IF of the proposed MDPDE is bounded for all $\alpha > 0$ but unbounded at $\alpha = 0$. This implies that the proposed MDPDE with $\alpha > 0$ is robust against contamination in data, whereas that at $\alpha = 0$ (existing MLE) is clearly non-robust. Further, it can also be verified that, given any fixed design, the supremum of this IF decreases as α increases, which in turn implies the increase in their robustness. This fact will be further seconded through empirical illustrations in Section 5.

Similar results can also be obtained if there are contaminations in all the G_is. The resulting influence function is then the sum of the previous IFs for individual component-wise contaminations and hence the implication is again the same indicating robustness at $\alpha > 0$ and non-robustness at $\alpha = 0$.

3 Robust Hypothesis Testing: A Wald-Type Test Statistics

Let us now consider the second important aspect of statistical inference, namely the testing of statistical hypothesis. As noted previously, the existing MLE based likelihood ratio tests or Wald tests are highly non-robust against data contamination in any general non-homogeneous set-up including the BRM. Suitable robust hypothesis testing procedures under the general non-homogeneous set-up have been developed in Ghosh and Basu (2017) and Basu et al. (2017) by extending the likelihood ratio and the Wald-type tests respectively. In this section, we develop a robust hypothesis testing procedure based on the proposed MDPDE for the BRM; here we restrict ourselves only to the Wald-type tests which are easy to implement in practice.

Consider the BRM (1) with the set-up as discussed in the previous sections. Consider the most common class of general linear hypotheses given by

$$H_0 : \boldsymbol{M}\boldsymbol{\beta} = \boldsymbol{m}_0 \quad \text{against} \qquad H_1 : \boldsymbol{M}\boldsymbol{\beta} \neq \boldsymbol{m}_0, \tag{8}$$

where $\boldsymbol{M}$ is a known matrix of order $r \times p$ and $\boldsymbol{m}_0$ is a known r-vector of reals. We make the standard assumption that $rank\,(\boldsymbol{M}) = r$ so that there exists a true null parameter value $\boldsymbol{\beta}_0 \neq \boldsymbol{0}_p$ (say) satisfying $\boldsymbol{M}\boldsymbol{\beta}_0 = \boldsymbol{m}_0$. Suppose $\widehat{\boldsymbol{\theta}}_{n,\alpha} = \left(\widehat{\boldsymbol{\beta}}_{n,\alpha}^T, \widehat{\phi}_{n,\alpha}\right)^T$ denotes the MDPDE of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ under the BRM (1). We define the Wald-Type test statistic for testing hypothesis (8) as

$$W_{n,\alpha} = n\left(\boldsymbol{M}\widehat{\boldsymbol{\beta}}_{n,\alpha} - \boldsymbol{m}_0\right)^T \left[\boldsymbol{M}\boldsymbol{\Psi}_n^{11}(\widehat{\boldsymbol{\theta}}_{n,\alpha})^{-1}\boldsymbol{\Omega}_n^{11}(\widehat{\boldsymbol{\theta}}_{n,\alpha})\boldsymbol{\Psi}_n^{11}(\widehat{\boldsymbol{\theta}}_{n,\alpha})^{-1}\boldsymbol{M}^T\right]^{-1} \left(\boldsymbol{M}\widehat{\boldsymbol{\beta}}_{n,\alpha} - \boldsymbol{m}_0\right), \tag{9}$$

where $\boldsymbol{\Psi}_n^{11}$ and $\boldsymbol{\Omega}_n^{11}$ are the $p \times p$ principle sub-matrix of the matrices $\boldsymbol{\Psi}_n$ and $\boldsymbol{\Omega}_n$ respectively and are given by $\boldsymbol{\Psi}_n^{11}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \gamma_{11,i}^{(\alpha)}(\boldsymbol{\theta})\,\boldsymbol{x}_i\boldsymbol{x}_i^T$ and $\boldsymbol{\Omega}_n^{11}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \left\{\gamma_{11,i}^{(2\alpha)}(\boldsymbol{\theta}) - \gamma_{1,i_0}^{(\alpha)}(\boldsymbol{\theta})^2\right\}\boldsymbol{x}_i\boldsymbol{x}_i^T$. Note that, since the MDPDE at $\alpha = 0$ coincides with the MLE, the test statistic $W_{n,0}$ is noting but the non-robust classical MLE based Wald test. So, the proposed test statistics $W_{n,\alpha}$ are the robust generalization of the Wald tests and hence referred to as the Wald-type tests.

In particular, for testing the significance of individual regression coefficient β_j, i.e., testing

$$H_0 : \beta_j = 0 \quad \text{against} \qquad H_1 : \beta_j \neq 0, \tag{10}$$

for any $j = 1, \ldots, p$, the proposed test statistic (9) simplifies to $W_{n,\alpha} = \frac{n\widehat{\beta}_{n,\alpha,j}^2}{\sigma_j(\widehat{\boldsymbol{\beta}}_{n,\alpha})}$, where $\widehat{\beta}_{n,\alpha,j}$ is the MDPDE of β_j and $\sigma_j(\boldsymbol{\theta})$ is the asymptotic variance of $\sqrt{n}\widehat{\beta}_{n,\alpha,j}$.

3.1 Asymptotic Properties

The first property that we need for any proposed test statistic is its null distribution to find out the critical region of the test. Although exact null distribution is not easy to obtain in general, the asymptotic distribution of our proposed test statistic $W_{n,\alpha}$ can be derived directly from that of the MDPDE. We assume that the matrices involved in the asymptotic variance of the MDPDE of $\boldsymbol{\beta}$, namely $\boldsymbol{\Psi}_n^{11}$ and $\boldsymbol{\Omega}_n^{11}$ are continuous in $\boldsymbol{\theta}$. Then, it is straightforward from the results of Section 2.2 that the asymptotic null distribution of $W_{n,\alpha}$ for hypothesis (8) is χ_r^2, the chi-square distribution with r degrees of freedom. So, the critical region of the proposed testing procedure at α_0-level of significance is given by

$$\left\{W_{n,\alpha} > \chi^2_{r,\alpha_0}\right\},$$

where χ^2_{r,α_0} is the $(1-\alpha_0)$-th quantile of the χ_r^2 distribution. For the particular case of the hypothesis (10), the corresponding null asymptotic distribution of $W_{n,\alpha}$ is χ_1^2. So, we can also perform the one-sided testing for the significance of β_j by considering the tests statistics $W_{n,\alpha}^+ = \frac{\sqrt{n}\widehat{\beta}_{n,\alpha,j}}{\sqrt{\sigma_j(\widehat{\boldsymbol{\beta}}_{n,\alpha})}}$, which has an asymptotic standard normal distribution at the null hypothesis in (10).

Further we can apply suitable results from Basu et al. (2017) on the Wald-type tests for the general non-homogeneous set-up to obtain useful power approximations for our proposal in the BRM. In particular, one can verify that the tests based on $W_{n,\alpha}$ are consistent at any fixed alternative for every $\alpha \geq 0$; this fact also follows from the Fisher consistency of the MDPDE used in the construction of test statistics and we leave the details for the reader.

So, for the comparison purpose, we need to compute the asymptotic power under the contiguous sequence of alternatives $H_{1,n} : \boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \frac{\boldsymbol{d}}{\sqrt{n}}$ for $\boldsymbol{d} \in \mathbb{R}^p - \{\mathbf{0}_p\}$, where $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^T, \phi_0)^T$ is the null parameter value satisfying $\boldsymbol{M}\boldsymbol{\beta}_0 = \boldsymbol{m}_0$. However, using the asymptotic distribution of the MDPDE from Section 2.2, one can obtain the asymptotic distribution of our test statistics $W_{n,\alpha}$ under the hypothesis $H_{1,n}$ to be $\chi_r^2(\delta)$, the non-central χ^2 with degrees of freedom r and non-centrality parameter $\delta = \boldsymbol{d}^T\boldsymbol{M}\left[\boldsymbol{M}\boldsymbol{\Psi}_n^{11}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Omega}_n^{11}(\boldsymbol{\theta}_0)\boldsymbol{\Psi}_n^{11}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{M}^T\right]^{-1}\boldsymbol{M}^T\boldsymbol{d}$. The asymptotic contiguous power of the proposed testing procedure can then be computed as $\left[1 - F_{\chi_r^2(\delta)}(\chi^2_{r,\alpha_0})\right]$, where $F_{\chi_r^2(\delta)}$ denotes the distribution function of $\chi_r^2(\delta)$. In particular, the pitman's asymptotic relative efficiencies of the $W_{n,\alpha}$ based Wald-type tests at $\alpha > 0$ with respect to the most powerful (but non-robust) classical Wald test (at $\alpha = 0$) depend on the non-centrality parameter δ and is directly proportional to the ratio of the inverse variance matrix of the MDPDE with that of the MLE. Hence it is directly proportional to the asymptotic efficiency of the MDPDE itself. So, for any given fixed design, the asymptotic power under contiguous alternative decreases slightly with increasing α but the loss is not quite significant at small positive α as in the case of efficiency of the MDPDEs; see Section 5 for corresponding empirical illustrations.

3.2 Robustness Analysis

We theoretically study the robustness of the proposed Wald-type tests through the corresponding influence function analysis (Hampel et al., 1986). Considering the set-up of Section 2.3, we define the statistical functional corresponding to the proposed test statistics $W_{n,\alpha}$ (ignoring the multiplier n) as

$$W_\alpha(G_1,\ldots,G_n) = \left(\boldsymbol{M}\boldsymbol{T}_\alpha(G_1,\ldots,G_n) - \boldsymbol{m}_0\right)^T \boldsymbol{\Sigma}(G_1,\ldots,G_n)^{-1}\left(\boldsymbol{M}\boldsymbol{T}_\alpha(G_1,\ldots,G_n) - \boldsymbol{m}_0\right), \quad (11)$$

where $\boldsymbol{\Sigma}(G_1,\ldots,G_n) = \left[\boldsymbol{M}\boldsymbol{\Psi}_n^{11}(\boldsymbol{T}_\alpha(G_1,\ldots,G_n))^{-1}\boldsymbol{\Omega}_n^{11}(\boldsymbol{T}_\alpha(G_1,\ldots,G_n))\boldsymbol{\Psi}_n^{11}(\boldsymbol{T}_\alpha(G_1,\ldots,G_n))^{-1}\boldsymbol{M}^T\right]$ and $\boldsymbol{T}_\alpha(G_1,\ldots,G_n)$ is the functional for the MDPDE as defined in (7). We can define its influence function as in the case of estimation by assuming contamination in any fixed distribution or in all distributions.

Let us again consider the contamination only in the distribution G_{i_0} at the contamination point t_{i_0}. Then, using the Fisher consistency of $\boldsymbol{T}_\alpha$, a routine differentiation yields the (first order) influence function of the test functional W_α to be identically zero at the model, i.e.,

$$\mathcal{IF}\left(t_{i_0}, W_\alpha; F_1, \ldots, F_n\right) = 0.$$

Therefore, this first order influence function cannot indicate the robustness of our proposed Wald-type tests, which is expected from the literature of similar quadratic tests (Heritier and Ronchetti, 1994; Toma and Broniatowski, 2010; Ghosh and Basu, 2017; Basu et al., 2017). So, we need to consider the second order influence function for W_α defined analogously with the second order partial derivative as

$$\mathcal{IF}_2\left(t_{i_0}, W_\alpha; G_1, \ldots, G_n\right) = \left|\frac{\partial^2 W_\alpha(G_1, \ldots, G_{i_0,\epsilon}, \ldots, G_n)}{\partial^2 \epsilon}\right|_{\epsilon=0}.$$

It indicates a second order approximation to the asymptotic bias due to infinitesimal contamination in contrast to the first order approximation provided by the first order IF. For the present BRM, some calculations yield the form of this second order IF for the proposed Wald-type test functional W_α at the model as given by

$$\mathcal{IF}_2\left(t_{i_0}, W_\alpha; F_1, \ldots, F_n\right) = \mathcal{IF}\left(t_{i_0}, \boldsymbol{T}_\alpha; F_1, \ldots, F_n\right) \boldsymbol{M}^T \boldsymbol{\Sigma}(F_1, \ldots, F_n)^{-1} \boldsymbol{M} \mathcal{IF}\left(t_{i_0}, \boldsymbol{T}_\alpha; F_1, \ldots, F_n\right).$$

Therefore, this influence function is bounded if and only if the IF of the MDPDE $\boldsymbol{T}_\alpha$, derived in Section 2.3, is bounded and this holds only for all $\alpha > 0$. Hence, the proposed Wald-type test statistics are expected to be robust for $\alpha > 0$ but non-robust at $\alpha = 0$ (which is the classical MLE based Wald test); further numerical illustrations are given in Section 5.

We can also examine the influence of the contamination on the level and power of the proposed Wald-type tests through the level and power influence function analysis (Hampel et al., 1986; Ghosh and Basu, 2017). For this purpose, we can directly apply the corresponding results for the general non-homogeneous cases from Basu et al. (2017) to conclude that the power influence function is indeed a matrix multiple of the IF of the MDPDE $\boldsymbol{T}_\alpha$. Therefore, the proposed test is robust in asymptotic contiguous power whenever the IF of $\boldsymbol{T}_\alpha$ is bounded, i.e., for all $\alpha > 0$, but is non-robust at $\alpha = 0$. However, following Basu et al. (2017), the level influence function of this type of tests under non-homogeneous data is always zero indicating the robustness of asymptotic level for all $\alpha \geq 0$ against infinitesimal contiguous contamination at the null hypothesis.

4 Applications to Real-life Data

4.1 Example 1: AIS Data

Let us start our illustration with reanalyzing the motivating AIS Dataset described in Section 1. We compute the proposed MDPDEs of the parameter $\boldsymbol{\theta} = (\beta_1, \beta_2, \phi)^T$ of the fitted BRM for different values of the tuning parameters α based on the full data and the outlier deleted data. The resulting estimates are reported in Table 1 along with the MLE (at $\alpha = 0$). Clearly, unlike the MLE, the proposed MDPDEs with $\alpha \geq 0.3$ change very little in the presence of two outlying observations. Further, the MDPDEs obtained based on the full data are themselves very close to the outlier deleted MLE (See Figure 2) and so they can be used safely without bothering about the outliers.

Table 1: MDPDEs of $(\beta_1, \beta_2, \phi)^T$ for the AIS data, along with the p-values for testing $H_0 : \beta_1 = 0$ using the Wald-type tests

	Full Data				Outlier deleted data			
α	β_1	β_2	ϕ	p-value	β_1	β_2	ϕ	p-value
0 (MLE)	0.098	-0.027	96.616	0.699	0.838	-0.038	246.305	0
0.1	0.328	-0.031	116.026	0.158	0.832	-0.038	238.036	0
0.2	0.765	-0.037	206.180	0	0.824	-0.038	231.658	0
0.3	0.807	-0.038	219.286	0	0.815	-0.038	227.072	0
0.4	0.804	-0.038	218.032	0	0.804	-0.038	224.270	0
0.5	0.794	-0.038	216.333	0	0.790	-0.038	223.383	0

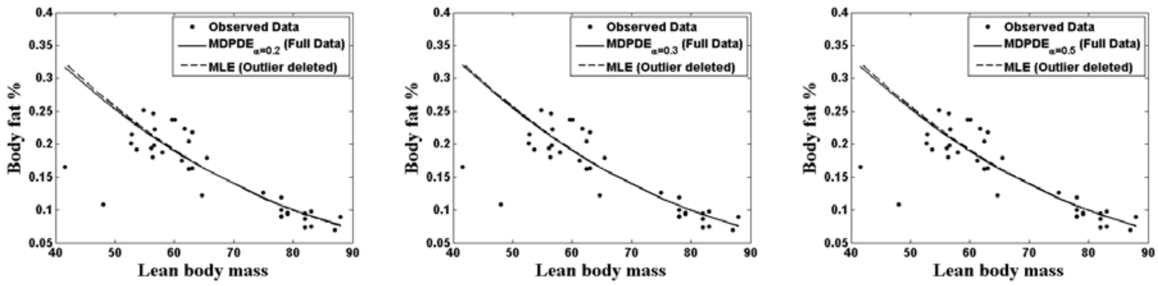
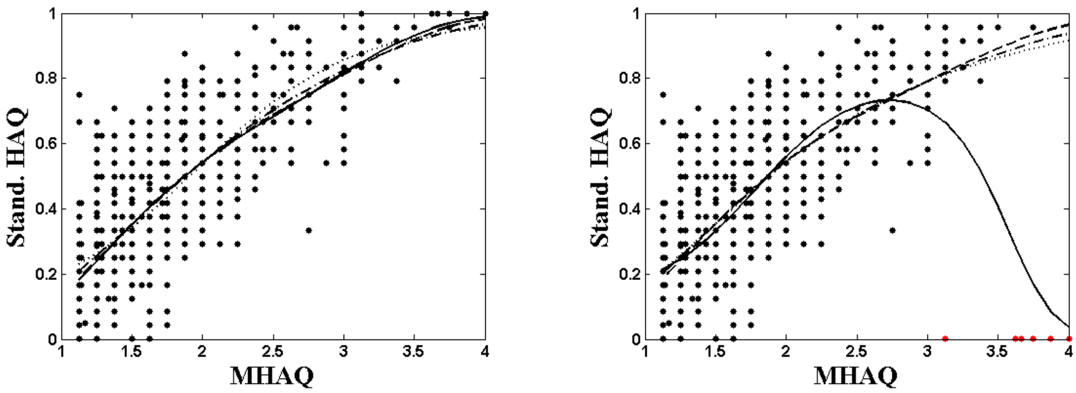


Figure 2: The BRM fitted lines for the AIS data based on the MDPDEs at $\alpha = 0.2, 0.3, 0.5$ for the full data, along with that based on the outlier deleted MLE.

Next, let us consider the problem of testing significance of the intercept term, namely $H_0 : \beta_1 = 0$, which changes drastically due to the presence of two outliers while using the existing MLE based Wald-tests. We apply our proposed Wald-type tests based on the MDPDEs for this testing problem and resulting p-values are reported in Table 1. Again, the proposed tests with $\alpha \geq 0.2$ generate stable p-values (which is zero) even in the full data with outliers. Therefore, the use of the proposed MDPDE and corresponding Wald-type tests with slightly larger $\alpha > 0$ can successfully tackle the effect of two outliers in the dataset yielding robust estimators and inference even without separately finding and removing these outliers.

4.2 Application 2: HAQ Dataset

In this example, we consider data on certain standardized health assessment questionnaires (HAQ) from the Division for Women and Children at the Oslo University Hospital at Ulleval, Oslo, Norway. The data, obtained from Prof. Nils L. Hjort of University of Oslo through personal communication, contain the original (elaborative) HAQ scores along with an easy-to-use modified version (MHAQ) for 1018 patients. These data have been used by Claeskens and Hjort (2008) to predict the original HAQ score from the simpler MHAQ scores, after suitable standardization, through a beta regression model. They have argued that the most healthy 219 patients with MHAQ = 1 need to be treated separately, but the remaining 799 patients' data can be modelled well by a polynomial BRM with covariates $\boldsymbol{x} = (1, \text{MHAQ}, \text{MHAQ}^2, \text{MHAQ}^3)^T$ and the logit link function. The corresponding fitted line based on the MLE is plotted in Figure 3a; clearly there is no outlier in the data. Here, the response variable HAQ takes the values in $[0, 3]$ inclusive of the end-points and, to get it within the open interval $(0, 1)$, we use the popular ad-hoc transformation $y = ((\text{HAQ}/3).(n-1) + 0.5)/n$, where $n = 799$ is the total sample size (Smithson and Verkuilen, 2006; Melo et al., 2015). Now, let us apply the proposed MDPDEs for this clean dataset to illustrate the behavior of our proposal in pure data. The resulting estimators in fact turn out to be very close to the MLE which can clearly be seen from the fitted lines in Figure 3a.



(a) Clean Data

(b) Data with Outliers (red)

Figure 3: The BRM fitted lines for the HAQ data based on the MDPDEs with different α [Solid line: $\alpha = 0$ (MLE); Dashed line: $\alpha = 0.1$; Dash-dotted line: $\alpha = 0.4$; Dotted line: $\alpha = 0.7$].

Now, to illustrate the robustness aspect, let us change only 6 largest HAQ values to $(1-\text{HAQ})$ values and again derive the MLE and the MDPDEs; the fitted lines are shown in Figure 3b (the artificial outliers are marked as red points). Note that, only due to these 6 outliers, which is about only 0.75% of the total number of observations, the MLE changes to a drastically different fit which clearly gives an erroneous inference. In fact the MLE based Wald test for testing significance of intercept term now gives the p-value of 0.64 (implying non-significance) with these outliers, which was zero (significant) in the original clean data. However, the MDPDE based fits remain very stable for all $\alpha \geq 0.1$ even in the presence of these outliers as seen from Figure 3b. Also, the MDPDE based Wald-type test yields correct p-value of zero for testing the significance of intercept term both in the clean data and with these outliers.

4.3 Application 3: Stress-Anxiety Data (Psychology)

Our final example is from a psychological trial among 166 nonclinical women in Australia measuring the scores on suitable tests of their anxiety, depression and stress symptoms. The details of the data can be found in Smithson and Verkuilen (2006) who have analyzed it with a beta regression model with response as anxiety scores and the covariates being the intercept and the stress scores along with the logit link function. Chien (2013) has studied these data to illustrate that there are several groups of highly influential observations affecting the MLE. We consider a set of 5 such outliers with higher anxiety scores and compute the MLE of the BRM parameters based on the full data and after deleting these outliers. The corresponding fitted lines are shown in Figure 4 which clearly indicates the non-robust nature of the MLE against the outlying observations.

We have applied our proposed MDPDE for these data and, as before, the MDPDEs with $\alpha > 0.2$ yields robust estimators. For brevity, we only present the fitted lines corresponding to the MDPDE with $\alpha = 0.3$ based on the full data in Figure 4; clearly the result is very close to that of the outlier deleted MLE indicating the robustness of our proposal.

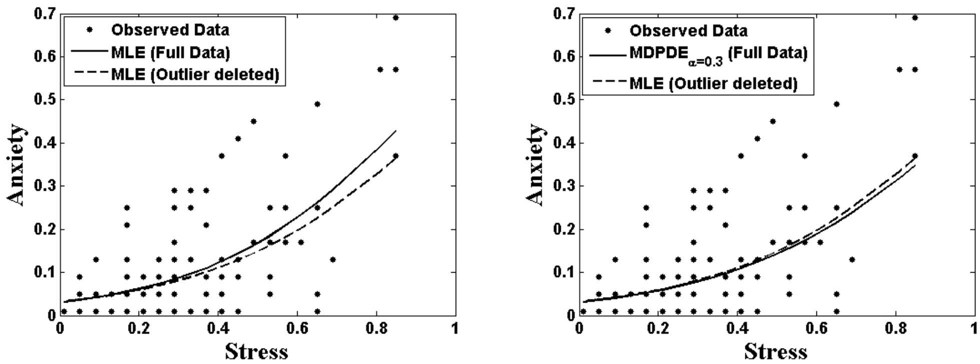


Figure 4: The BRM fitted lines for the Stress-Anxiety data based on the MDPDEs at $\alpha = 0$ (MLE) and $\alpha = 0.3$ for the full data, along with that of the outlier deleted MLE.

5 Simulation Studies

5.1 Performance of the MDPDE

Let us now study the finite-sample behavior of the proposed estimator, MDPDE, through suitable simulation study and compare with theoretical (asymptotic) results. Consider the sample size $n = 50$ and fix n covariate values $x_1, \ldots, x_n$ being independent observations from $U(0, 1)$. We generate 1000 samples from the BRM (1) with $p = 2$, one intercept (β_1) and one slope (β_2) corresponding to the covariates x_i, along with the logit link function. The true value of the parameter $\boldsymbol{\theta} = (\beta_1, \beta_2, \phi)^T$ is taken as $(-1, 1, 5)^T$. For each of the samples, we compute the MDPDEs with different α and derive their empirical bias and MSE over these 1000 replications (without any outlier) which are reported in Table 2. Clearly, MLE has the least bias and MSE under pure data as expected and the bias and MSE of the proposed MDPDE increase slightly with increasing α. But this increase in bias or MSE is not quite significant at small positive α like 0.3, 0.4, which is consistent with the asymptotic efficiency described in Section 2.2.

Table 2: Empirical Bias and MSE of the MDPDEs with different α under pure data

	Bias			MSE		
α	β_1	β_2	ϕ	β_1	β_2	ϕ
0 (MLE)	-0.010	0.011	0.332	0.058	0.172	1.124
0.1	-0.010	0.011	0.325	0.058	0.174	1.123
0.2	-0.012	0.013	0.338	0.059	0.178	1.180
0.3	-0.014	0.015	0.367	0.062	0.185	1.293
0.4	-0.017	0.018	0.410	0.064	0.193	1.464
0.5	-0.020	0.022	0.464	0.067	0.203	1.696
0.6	-0.024	0.026	0.526	0.071	0.214	1.990
0.7	-0.028	0.031	0.593	0.074	0.225	2.347

Table 3: Empirical Bias and MSE of the MDPDEs with different α under contaminated data

	Bias			MSE		
α	β_1	β_2	ϕ	β_1	β_2	ϕ
0 (MLE)	0.232	-0.240	-0.332	0.141	0.283	1.066
0.1	0.216	-0.218	-0.300	0.132	0.270	1.037
0.2	0.199	-0.197	-0.247	0.125	0.262	1.050
0.3	0.184	-0.177	-0.182	0.121	0.259	1.100
0.4	0.170	-0.160	-0.111	0.119	0.260	1.196
0.5	0.158	-0.145	-0.036	0.118	0.264	1.343
0.6	0.147	-0.133	0.043	0.119	0.269	1.550
0.7	0.138	-0.122	0.120	0.120	0.276	1.790

Next, to study the finite-sample robustness behavior of the proposed MDPDEs, we repeat the previous simulation study, but after contaminating each sample at 10% level. For this purpose, we have randomly changed 10% of the response values y to $(1-y)$ and recalculated the MDPDEs based on the contaminated samples. The corresponding bias and MSE are reported in Table 3. It can be observed that the absolute Bias and MSE of the MLE are the worst, since it is the most non-robust one. As α increases, both the absolute bias and MSE decrease significantly providing more accurate results. Thus, the robustness of the proposed MDPDE under contamination significantly improves with increasing values of α; this is again consistent with the theoretical influence function analysis discussed in Section 2.3.

Similar results are observed in several simulation studies with different contamination scheme and proportions and different true parameter values; so those are not repeated here for brevity.

5.2 Performance of the Wald-type tests

In this section, we illustrate the empirical levels and powers of the proposed Wald-type tests based on the MDPDEs through simulation studies. For the sake of consistency, let us consider the same simulation set-up as described in the previous section; with each simulated sample of size $n = 50$, both with and without contamination as before, we apply the proposed testing procedure for different hypotheses. In particular, we perform the Wald-type tests with different α for six null hypotheses given by

$$\begin{array}{llllr} & H_0^{L1}: \beta_1 = -1; & H_0^{L2}: \beta_2 = 1; & H_0^{L3}: (\beta_1, \beta_2)^T = (-1, 1)^T; & \text{for studying level,} \\ \text{and} & H_0^{P1}: \beta_1 = 0; & H_0^{P2}: \beta_2 = 0; & H_0^{P3}: (\beta_1, \beta_2)^T = (0, 0)^T; & \text{for studying power,} \end{array}$$

against their respective omnibus alternatives. Note that, all these hypotheses belong to the class of general linear hypotheses (8) considered in Section 3. Based on 1000 replications, we compute the empirical sizes and powers at the 5% level of significance for testing these hypotheses under pure data as well as under contaminated data; the results are reported in Tables 4 and 5 respectively.

Table 4: Empirical sizes and powers for the MDPDE based Wald-type tests for different null hypotheses and different α under pure data

	Size			Power		
α	H_0^{L1}	H_0^{L2}	H_0^{L3}	H_0^{P1}	H_0^{P2}	H_0^{P3}
0 (Wald)	0.060	0.058	0.055	0.996	0.719	0.995
0.1	0.058	0.057	0.057	0.995	0.712	0.996
0.2	0.058	0.054	0.059	0.994	0.711	0.995
0.3	0.060	0.058	0.066	0.994	0.697	0.995
0.4	0.066	0.063	0.069	0.993	0.691	0.995
0.5	0.071	0.066	0.074	0.992	0.679	0.993
0.6	0.076	0.070	0.082	0.989	0.667	0.993
0.7	0.076	0.069	0.086	0.986	0.657	0.991

Table 5: Empirical sizes and powers for the MDPDE based Wald-type tests for different null hypotheses and different α under contaminated data

	Size			Power		
α	H_0^{L1}	H_0^{L2}	H_0^{L3}	H_0^{P1}	H_0^{P2}	H_0^{P3}
0 (Wald)	0.172	0.100	0.163	0.784	0.416	0.831
0.1	0.155	0.098	0.153	0.798	0.428	0.844
0.2	0.144	0.092	0.146	0.806	0.444	0.859
0.3	0.140	0.082	0.137	0.819	0.454	0.863
0.4	0.137	0.078	0.129	0.824	0.462	0.859
0.5	0.133	0.078	0.124	0.828	0.461	0.860
0.6	0.125	0.079	0.120	0.828	0.460	0.856
0.7	0.120	0.077	0.120	0.821	0.453	0.854

It can be observed that the sizes of the MDPDE based Wald-type tests increase slightly under pure data for any hypothesis. In fact, all the empirical sizes are slightly inflated due to the use of asymptotic critical values for testing with $n = 50$. Also, as α increases, the powers under pure data decrease very little for all three hypotheses. The changes at small $\alpha > 0$ under pure data with respect to the classical Wald test at $\alpha = 0$ are clearly not quite significant. On the other hand, under contaminated data, the sizes and powers of the classical Wald-test ($\alpha = 0$) change drastically for all hypotheses. But those for the proposed Wald-type tests at small positive α remain more stable under contamination.

5.3 On the choice of the tuning parameter α

The proposed DPD based robust estimators and Wald-type tests depend on a tuning parameter α. We have seen, both theoretically and empirically, that the efficiency of the proposed MDPDE under pure data decreases slightly as α increases, but their robustness under contamination increases significantly. Thus, the tuning parameter α yields a trade-off between efficiency and robustness of the proposed estimator. For hypothesis testing also, the asymptotic contiguous power decreases slightly with increasing α, but the robustness of its level and power improves significantly under contamination; here α trades off the contiguous power under pure data with robustness against outliers. Therefore, in either case, this tuning parameter α needs to be chosen appropriately for a given dataset.

As observed from various simulations and real data analyses, an $\alpha \approx 0.3,\ 0.4$ gives sufficiently robust estimator without significant loss in efficiency under pure data and also provides a desired trade-off for the corresponding Wald-type test. So, the empirical suggested value of α is to be taken around 0.3 to 0.4 which is expected to work well in most of the applications.

However, for a better trade-off based on the amount of contamination in the given dataset, a data-driven choice of this tuning parameter α could be useful. There are only a few such approaches for the DPD based inference. We propose to follow the approach presented by Warwick and Jones (2005) and Ghosh and Basu (2015) for the iid and the non-homogeneous data respectively. Their approach is mainly based on choosing α by minimizing an appropriate estimate of the MSE given by

$$E\left[(\widehat{\boldsymbol{\theta}}_{n,\alpha} - \boldsymbol{\theta}^*)^T(\widehat{\boldsymbol{\theta}}_{n,\alpha} - \boldsymbol{\theta}^*)\right] = (\boldsymbol{\theta}_\alpha - \boldsymbol{\theta}^*)^T(\boldsymbol{\theta}_\alpha - \boldsymbol{\theta}^*) + \frac{1}{n} Trace\left[\boldsymbol{\Psi}_n^{-1}\boldsymbol{\Omega}_n\boldsymbol{\Psi}_n^{-1}\right],$$

where $\boldsymbol{\theta}^*$ is the target parameter value, $\boldsymbol{\theta}_\alpha = \boldsymbol{T}_\alpha(G_a, \ldots, G_n)$ and $\widehat{\boldsymbol{\theta}}_{n,\alpha}$ is the MDPDE with tuning parameter α. For the present case of beta regression models, we can estimate this MSE by plugging in the MDPDE $\widehat{\boldsymbol{\theta}}_{n,\alpha}$ for $\boldsymbol{\theta}_\alpha$ and also in the variance part, but need to use different pilot estimators for θ^*. Ghosh and Basu (2015) have suggested that the use of the MDPDE with $\alpha = 0.5$ serves well as the pilot estimator in case of linear regression model. This suggestion may be followed in the present case of BRM also, but it surely needs substantial further investigation which we hope to do in our future research.

6 Concluding remarks

In this paper, we have developed a robust statistical inference procedure under the beta regression model for modeling responses on $(0, 1)$. We have proposed the minimum DPD estimator for estimating the parameters in the BRM and developed a class of Wald-type tests based on them for testing general linear hypotheses in regression coefficients. Beside discussing their asymptotic properties, we have also justified the robustness of the proposed methodology through appropriate influence function analyses. Suitable numerical illustrations have been provided along with three important real data applications from health-care studies.

This work also leads to many possible important future extensions. For example, it will be useful to further extend the proposals to develop robust inference for more general beta regression models, like the inflated zero or one (or both) beta regression models for datasets containing 0 or 1 or both values, or the variable dispersion beta regression models for non-homogeneous precision parameter ϕ etc. Also, the proposed scheme for selection of a data-driven choice of the tuning parameter α needs more investigation. We plan to pursue some of these extensions in our future works.

Acknowledgments:
The author wants to express his sincere thanks to Prof. Nils L. Hjort of University of Oslo for the HAQ dataset and Prof. Ayanendranath Basu of Indian Statistical Institute for several constructive suggestions and comments about the work. This work is supported by the INSPIRE Faculty research grant from the Department of Science and Technology, Govt. of India.

References

Aerts, S. and Haesbroeck, G. (2016). Robust asymptotic tests for the equality of multivariate coefficients of variation *Test*, doi: 10.1007/s11749-016-0504-4.

Basu, A., Ghosh, A., Martin, N. and Pardo, L. (2017) Robust Wald-type tests for non-homogeneous observations based on minimum density power divergence estimator. Preprint.

Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, *85*, 549–559.

Basu, A., Shioya, H. and Park, C. (2011). *Statistical Inference: The Minimum Distance Approach.* Chapman & Hall/CRC. Boca Raton, Florida.

Bayes, C. L., Bazan, J. L., and Garcia, C. (2012). A New Robust Regression Model for Proportions. *Bayesian analysis*, *7(4)*, 841—866.

Chien, L. (2013). Multiple deletion diagnostics in beta regression models. *Computational Statistics*, *28*, 1639—1661.

Claeskens, G., and Hjort, N. L. (2008). *Model selection and model averaging.* Cambridge University Press.

Cook and Weisberg (1994). *An Introduction to Regression Graphics.* John Wiley and Sons, New York.

Espinheira, P., Ferrari, S., and Cribari-Neto, F. (2008a). Influence diagnostics in beta regression. *Computational Statistics and Data Analysis*, *52(9)*, 4417—4431.

Espinheira, P., Ferrari, S., and Cribari-Neto, F. (2008b). On Beta Regression Residuals. *Journal of Applied Statistics*, *35(4)*, 407—419.

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

Ghosh, A. (2016). Divergence based robust estimation of the tail index through an exponential regression model. *Statistical Methods and Applications*, doi:10.1007/s10260-016-0364-9.

Ghosh, A., and Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electronic Journal of Statistics, 7*, 2420—2456.

Ghosh, A., and Basu, A. (2015). Robust Estimation for Non-Homogeneous Data and the Selection of the Optimal Tuning Parameter: The DPD Approach. *Journal of Applied Statistics, 42(9)*, 2056—2072.

Ghosh, A., and Basu, A. (2016). Robust Estimation in Generalized Linear Models : The Density Power Divergence Approach. *TEST, 25(2)*, 269—290.

Ghosh, A., and Basu, A. (2017). Robust Bounded Influence Tests for Independent but Non-Homogeneous Observations. *Statistica Sinica*, to appear.

Hampel, F. R., Ronchetti, E., Rousseeuw, P. J., and Stahel W.(1986). *Robust Statistics: The Approach Based on Influence Functions*. New York, USA: John Wiley & Sons.

Heritier, S., and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc. 89*, 897–904.

Huber, P. J. (1983). Minimax aspects of bounded-influence regression (with discussion). *Journal of the American Statistical Association*, **69**, 383-393.

Kieschnick, R., and McCullough, B. D. (2003). Regression analysis of variates observed on $(0, 1)$: percentages, proportions and fractions. *Statistical Modelling, 3*, 193–213.

Melo, O. O., Melo, C. E., and Mateu, J. (2015). Distance-based beta regression for prediction of mutual funds. *AStA Advances in Statistical Analysis, 99*, 83–106.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman & Hall, London.

Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis*, *9*, 325—346.

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

Toma, A., and Broniatowski, M. (2010). Dual divergence estimators and tests: robustness results. *J. Multivariate Anal. 102*, 20–36.

Vasconcellos, K.L.P., and Cribari-Neto, F., (2005). Improved maximum likelihood estimation in a new class of beta regression models. *Brazilian Journal of Probability and Statistics, 19*, 13–31.

Warwick, J., and Jones, M. C. (2005). Choosing a robustness tuning parameter. *Journal of Statistical Computation and Simulation, 75*, 581—588.