



Article Robust Statistical Inference in Generalized Linear Models Based on Minimum Renyi's Pseudodistance Estimators

María Jaenada [†] and Leandro Pardo ^{*,†}

Department of Statistics and Operation Research, Faculty of Mathematics, Complutense University of Madrid, Plaza Ciencias, 3, 28040 Madrid, Spain; mjaenada@ucm.es

* Correspondence: lpardo@mat.ucm.es

+ These authors contributed equally to this work.

Abstract: Minimum Renyi's pseudodistance estimators (MRPEs) enjoy good robustness properties without a significant loss of efficiency in general statistical models, and, in particular, for linear regression models (LRMs). In this line, Castilla et al. considered robust Wald-type test statistics in LRMs based on these MRPEs. In this paper, we extend the theory of MRPEs to Generalized Linear Models (GLMs) using independent and nonidentically distributed observations (INIDO). We derive asymptotic properties of the proposed estimators and analyze their influence function to asses their robustness properties. Additionally, we define robust Wald-type test statistics for testing linear hypothesis and theoretically study their asymptotic distribution, as well as their influence function. The performance of the proposed MRPEs and Wald-type test statistics are empirically examined for the Poisson Regression models through a simulation study, focusing on their robustness properties. We finally test the proposed methods in a real dataset related to the treatment of epilepsy, illustrating the superior performance of the robust MRPEs as well as Wald-type tests.

Keywords: generalized linear model; independent and nonidentically distributed observations; minimum Rényi's pseudodistance estimators; robust Wald-type test statistics for GLMs; influence function for GLMs; poisson regression model

MSC: 62F35; 62J12

1. Introduction

Generalized linear models (GLMs) were first introduced by Nelder and Wedderburn [1] and later expanded upon by McCullagh and Nelder [2]. The GLMs represent a natural extension of the standard linear regression models, which enclose a large variety of response variable distributions, including distributions of count, binary, or positive values. Let Y_1, \ldots, Y_n be independent response variables. The classical GLM assumes that the density function of each random variable Y_i belongs to the exponential family, having the form

$$f(y,\theta_i,\phi) = \exp\left\{\frac{y\theta_i - b(\theta_i)}{a(\phi)} + c(y,\phi)\right\},\tag{1}$$

for i = 1, ..., n, where the functions $a(\phi)$, $b(\theta_i)$ and $c(y, \phi)$ are known. Therefore, the observations are independent but not identically distributed, depending on a location parameter θ_i , i = 1, ..., n, and a nuisance parameter ϕ . Further, we denote by μ_i the expectation of the random variable Y_i and we assume that there exists a monotone differentiable function, so called link function g, verifying

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

with $\beta = (\beta_1, ..., \beta_k) \in \mathbb{R}^k$ (k < n) the regression parameter vector. The $k \times 1$ -vector of explanatory variables, x_i , is assumed to be nonrandom, i.e., the design matrix is fixed. Correspondingly, the location parameter depends on the explanatory variables $\theta = \theta(\mathbf{x}^T \beta)$



Citation: Jaenada, M.; Pardo, L. Robust Statistical Inference in Generalized Linear Models based on Minimum Renyi's Pseudodistance estimators. *Entropy* **2022**, *24*, 123. https://doi.org/10.3390/e24010123

Academic Editor: Philip Broadbridge

Received: 11 December 2021 Accepted: 11 January 2022 Published: 13 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the density function given in (1) can be written as $f_i(y, \beta, \phi)$, empathizing its dependence of β and x_i .

The maximum likelihood estimator (MLE) and the quasilikelihood estimators were well studied for the GLMs, and it is well known that they are asymptotically efficient but lack robustness in the presence of outliers, which can result in a significant estimation bias. Jaenada and Pardo [3] revised the different robust estimators in the statistical literature and studied the lack of robustness of the MLE as well. Among others, Stefanski et al. [4] studied optimally bounded score functions for the GLM and generalized the results obtained by Krasker and Welsch [5] for classical LRMs. Künsch et al. [6] introduced the so-called conditionally unbiased bounded-influence estimate, and Morgenthaler [7], Cantoni and Ronchetti [8], Bianco and Yohai [9], Croux and Hesbroeck [10], Bianco et al. [11], and Valdora and Yohai [12] continued the development of robust estimators for the GLMs based on general M-estimators. Later, Ghosh and Basu [13] proposed robust estimators for the GLM, based on the density power divergence (DPD) introduced in Basu et al. [14].

There are not many papers considering robust tests for GLMs. In this sense, Basu et al. [15] considered robust Wald-type tests based on the minimum DPD estimator, but assuming random explanatory variables for the GLM. The main purpose of this paper is to introduce new robust Wald-type tests based on the MRPE under fixed (not random) explanatory variables.

Broniatowski et al. [16] presented robust estimators for the parameters of the linear regression model (LRM) with random explanatory variables and Castilla et al. [17] considered Wald-type test statistics, based on MRPE, for the LRM. Toma and Leoni–Aubin [18] defined new robustness and efficient measures based on the RP and Toma et al. [19] considered the MRPE for general parametric models, and constructed a model selection criterion for regression models. The term "Rényi pseudodistance" (RP) was adopted in Broniatowski et al. [16] because of its similarity with the Rényi divergence (Rényi [20]), although this family of divergences was considered previously in Jones et al. [21]. Fujisawa and S. Eguchi [22] used the RP under the name of γ -cross entropy, introduced robust estimators obtained by minimizing the empirical estimate of the γ -cross entropy (or the γ -divergence associated to the γ -cross entropy) and studied their properties. Further, Hirose and Masuda [23] considered the γ likelihood function to find robust estimation. Using the γ -divergence, Kawashima and Fujisawa [24,25] presented robust estimators for sparse regression and sparse GLMs with random covariates. The robustness of all the previous estimators is based on density power weight, $f(y, \theta)^l$, which gives a small weight to outliers observations. This idea was also developed by Basu et al. [15] for the minimum DPD estimator and was considered some years ago by Windham [26]. More concretely, Basu et al. [14] considered the density power function multiplied by the score function.

The outline of the paper is as follows: in Section 2, some results in relation to the MRPEs for GLMs, previously obtained in Jaenada and Pardo [3], are presented. Section 3 introduces and studies Wald-type tests based on the MRPE for testing linear null hypothesis for the GLMs. In Section 4, the influence function of the MRPE as well as the influence functions of the Wald-type tests are derived. Finally, we empirically examine the performance of the proposed robust estimators and Wald-type test statistics for the Poisson regression model through a simulation study in Section 5, and we illustrate its applicability with real data sets for binomial and Poisson regression.

2. Asymptotic Distribution of the MRPEs for the GLMs

In this Section, we revise some of the results presented in Jaenada and Pardo [3] in relation to the MRPE. Let Y_1, \ldots, Y_n , be INIDO random variables with density functions with respect to some common dominating measure, g_1, \ldots, g_n respectively. The true densities g_i are modeled by the density functions given in (1), belonging to the exponential family. Such densities are denoted by $f_i(y, \beta, \phi)$ highlighting its dependence on the regression vector β , the nuisance parameter ϕ and the observation $i, i = 1, \ldots, n$. In the following, we assume that the explanatory variables x_i , are fixed, and therefore the response variables verify the INIDO set up studied in Castilla et al. [27]. For each of the response variables Y_i , the RP between the theoretical density function belonging to the exponential family, $f_i(y, \gamma)$, and the true density underlying the data, g_i , can be defined, for $\alpha > 0$ as

$$R_{\alpha}(f_i(y,\gamma),g_i) = \frac{1}{\alpha+1}\log\left(\int f_i(y,\gamma)^{\alpha+1}dy\right) - \frac{1}{\alpha}\log\left(\int f_i(y,\gamma)^{\alpha}g_i(y)dy\right) + k, \quad (2)$$

where

$$k = \frac{1}{\alpha(\alpha+1)} \log \left(\int g_i(y)^{\alpha+1} dy \right)$$

does not depend on $\gamma = (\boldsymbol{\beta}^T, \boldsymbol{\phi})^T$.

We consider (y_1, \ldots, y_n) a random sample of independent but nonhomogeneous observations of the response variables with fixed predictors (x_1, \ldots, x_n) . Since only one observation of each variable Y_i is available, a natural estimate of its true density g_i is the degenerate distribution at the the observation y_i . Consequently, in the following we denote \hat{g}_i the density function of the degenerate variable at the point y_i . Then, substituting in (2) the theoretical and empirical densities, yields to the loss

$$R_{\alpha}(f_i(y,\gamma),\widehat{g}_i) = \frac{1}{\alpha+1} \log\left(\int f_i(y,\gamma)^{\alpha+1} dy\right) - \frac{1}{\alpha} \log f_i(Y_i,\gamma)^{\alpha} + k.$$
(3)

If we consider the limit when α tends to zero we get

$$R_0(f_i(y,\gamma),\widehat{g}_i) = \lim_{\alpha \downarrow 0} R_\alpha(f_i(y,\gamma),\widehat{g}_i) = -\log f_i(Y_i,\gamma) + k.$$
(4)

Last expression coincides with the Kullback–Leibler divergence, except for the constant *k*. More details about Kullback–Leiber divergence can be seen in Pardo [28].

For the seek of simplicity, let us denote

$$L^{i}_{\alpha}(\boldsymbol{\gamma}) = \left(\int f_{i}(y,\boldsymbol{\gamma})^{\alpha+1}dy\right)^{rac{lpha}{lpha+1}},$$

and

$$V_i(Y_i, \boldsymbol{\gamma}) = rac{f_i(Y_i, \boldsymbol{\gamma})^{lpha}}{L^i_{lpha}(\boldsymbol{\gamma})}.$$

The expression (3) can be rewritten as

$$R_{\alpha}(f_i(y,\gamma),\widehat{g}_i) = -\frac{1}{\alpha} \log \left(\frac{f_i(Y_i,\gamma)^{\alpha}}{\left(\int f_i(y,\gamma)^{\alpha+1} dy \right)^{\frac{\alpha}{\alpha+1}}} \right) + k = -\frac{1}{\alpha} \log V_i(Y_i,\gamma) + k.$$

Based on the previous idea, we shall define an objective function averaging the RP between all the the RPs. Since minimizing $R_{\alpha}(f_i(y, \gamma), \hat{g}_i)$ in γ is equivalent to maximizing log $V_i(Y_i, \gamma)$, we define a loss function averaging those quantities as

$$T_n^{\alpha}(\gamma) = \frac{1}{n} \sum_{i=1}^n \frac{f_i(Y_i, \gamma)^{\alpha}}{\left(\int f_i(y, \gamma)^{\alpha+1} dy\right)^{\frac{\alpha}{\alpha+1}}} = \frac{1}{n} \sum_{i=1}^n \frac{f_i(Y_i, \gamma)^{\alpha}}{L_{\alpha}^i(\gamma)} = \frac{1}{n} \sum_{i=1}^n V_i(Y_i, \gamma).$$
(5)

Based on (5), we can define the MRPE of the unknown parameter γ , $\hat{\gamma}_{\alpha}$, by

$$\widehat{\gamma}_{\alpha} = \arg \max_{\gamma \in \Gamma} T_n^{\alpha}(\gamma), \tag{6}$$

with $T_n^{\alpha}(\gamma)$ defined in (5)

$$T_n^0(\boldsymbol{\gamma}) = \frac{1}{n} \sum_{i=1}^n \log f_i(y_i, \boldsymbol{\gamma})$$

Now, since the MRPE is defined as a maximum, it must annul the first derivatives of the loss function given in (5). The estimating equations of the parameters β and ϕ are given by

$$\begin{cases} \frac{1}{n} \sum_{i=1}^{n} \frac{\partial V_i(Y_i, \gamma)}{\partial \beta} = \mathbf{0}_k \\ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial V_i(Y_i, \gamma)}{\partial \phi} = \mathbf{0}. \end{cases}$$
(7)

For the first equation, we have

$$\begin{aligned} \frac{\partial V_i(Y_i, \gamma)}{\partial \beta} &= \frac{1}{L^i_{\alpha}(\gamma)^2} \bigg\{ \alpha f_i(Y_i, \gamma)^{\alpha} \frac{\partial \log f_i(Y_i, \gamma)}{\partial \beta} L^i_{\alpha}(\gamma) \\ &- \bigg[\alpha \bigg(\int f_i(y, \gamma)^{\alpha+1} dy \bigg)^{\frac{\alpha}{\alpha+1}-1} \int f_i(y, \gamma)^{\alpha+1} \frac{\partial \log f_i(y, \gamma)}{\partial \beta} dy \bigg] f_i(Y_i, \gamma)^{\alpha} \bigg\}. \end{aligned}$$

The previous partial derivatives can be simplified as

$$\frac{\partial \log f_i(Y_i, \gamma)}{\partial \beta} = \frac{Y_i - \mu_i}{Var(Y_i)g'(\mu_i)} x_i = K_{1i}(Y_i, \gamma) x_i$$

and

$$\frac{\partial \log f_i(Y_i, \gamma)}{\partial \phi} = -\frac{(Y_i \theta_i - b(\theta_i))}{a(\phi)^2} a'(\phi) + \frac{\partial c(Y_i, \phi)}{\partial \phi} = K_{2i}(Y_i, \gamma).$$

See Ghosh and Basu [13] for more details. Now using the simplified expressions, we can write the estimating equation for β as

$$\sum_{i=1}^{n} \frac{\boldsymbol{x}_{i}}{L_{\alpha}^{i}(\boldsymbol{\gamma})} \{ M_{i}(Y_{i},\boldsymbol{\gamma}) - N_{i}(Y_{i},\boldsymbol{\gamma}) \} = \boldsymbol{0}_{k}$$
(8)

being

$$M_i(Y_i, \boldsymbol{\gamma}) = f_i(Y_i, \boldsymbol{\gamma})^{\alpha} K_{1i}(Y_i, \boldsymbol{\gamma})$$

and

$$N_i(Y_i, \gamma) = \frac{f_i(Y_i, \gamma)^{\alpha}}{\int f_i(y, \gamma)^{\alpha+1} dy} \int f_i(y, \gamma)^{\alpha+1} K_{1i}(y, \gamma) dy$$

Subsequently, the estimating equation for ϕ , is given by

$$\begin{split} \frac{\partial V_i(Y_i,\gamma)}{\partial \phi} &= \frac{1}{L_{\alpha}^i(\gamma)^2} \bigg\{ \alpha f_i(Y_i,\gamma)^{\alpha} \frac{\partial \log f_i(Y_i,\gamma)}{\partial \phi} L_{\alpha}^i(\gamma) \\ &- \bigg[\alpha \bigg(\int f_i(y,\gamma)^{\alpha+1} dy \bigg)^{\frac{\alpha}{\alpha+1}-1} \int f_i(y,\gamma)^{\alpha+1} \frac{\partial \log f_i(y,\gamma)}{\partial \phi} dy \bigg] f_i(Y_i,\gamma)^{\alpha} \bigg\} \\ &= \frac{1}{L_{\alpha}^i(\gamma)^2} \bigg\{ \alpha f_i(Y_i,\gamma)^{\alpha} \frac{\partial \log f_i(Y_i,\gamma)}{\partial \phi} L_{\alpha}^i(\gamma) \\ &- \bigg[\alpha \frac{L_{\alpha}^i(\gamma)}{\int f_i(y,\gamma)^{\alpha+1} dy} \int f_i(y,\gamma)^{\alpha+1} \frac{\partial \log f_i(y,\gamma)}{\partial \phi} dy \bigg] f_i(Y_i,\gamma)^{\alpha} \bigg\}. \end{split}$$

and thus, the estimating equation for ϕ is given by

$$\sum_{i=1}^{n} \frac{1}{L_{\alpha}^{i}(\gamma)} \{ M_{i}^{*}(Y_{i}, \gamma) - N_{i}^{*}(Y_{i}, \gamma) \} = 0$$
(9)

being

$$M_i^*(Y_i, \gamma) = f_i(Y_i, \gamma)^{\alpha} K_{2i}(Y_i, \gamma),$$

and

$$N_i^*(Y_i, \gamma) = \frac{f_i(Y_i, \gamma)^{\alpha}}{\int f_i(y, \gamma)^{\alpha+1} dy} \int f_i(y, \gamma)^{\alpha+1} K_{2i}(y, \gamma) dy$$

Under some regularity conditions, Castilla et al. [27] established the consistency and asymptotic normality of the MRPEs under the INIDO setup. Before stating the consistence and asymptotic distribution of the MRPEs for the GLM, let us introduce some useful notation. We define

$$S_{\alpha}^{i} = \int f_{i}(y, \boldsymbol{\beta}, \boldsymbol{\phi})^{\alpha+1} dy$$

$$m_{jli}(\boldsymbol{\gamma}) = \frac{1}{\int f_{i}(y, \boldsymbol{\gamma})^{\alpha+1} dy} \int f_{i}(y, \boldsymbol{\gamma})^{\alpha+1} K_{ji}(y, \boldsymbol{\gamma}) K_{li}(y, \boldsymbol{\gamma}) dy,$$

$$m_{ji}(\boldsymbol{\gamma}) = \frac{1}{\int f_{i}(y, \boldsymbol{\gamma})^{\alpha+1} dy} \int f_{i}(y, \boldsymbol{\beta}, \boldsymbol{\phi})^{\alpha+1} K_{ji}(y, \boldsymbol{\gamma}) dy,$$

$$l_{jli}(\boldsymbol{\gamma}) = \int \frac{f_{i}(y, \boldsymbol{\gamma})^{2\alpha+1}}{L_{\alpha}^{i}(\boldsymbol{\gamma})^{2}} (K_{ji}(y, \boldsymbol{\gamma}) - m_{ji}(\boldsymbol{\gamma})) (K_{li}(y, \boldsymbol{\gamma}) - m_{li}(\boldsymbol{\gamma})) dy,$$
(10)

for all j, l = 1, 2 and i = 1, ..., n.

Theorem 1. Let Y_1, \ldots, Y_n be a random sample from the GLM defined in (1). The MRPE $\hat{\gamma}_{\alpha} = (\hat{\beta}_{\alpha}^T, \hat{\phi}_{\alpha})^T$ is consistent and its asymptotic distribution is given by

$$\sqrt{n}\mathbf{\Omega}_n(\boldsymbol{\gamma})^{-\frac{1}{2}} \Psi_n(\boldsymbol{\gamma}) \left((\widehat{\boldsymbol{\beta}}_{\alpha}, \widehat{\boldsymbol{\phi}}_{\alpha}) - (\boldsymbol{\beta}, \boldsymbol{\phi}) \right) \stackrel{L}{\underset{n \to \infty}{\to}} N(\mathbf{0}_{k+1}, \boldsymbol{I}_{k+1}),$$

where **X** denotes the design matrix, I_k is the k-dimensional identity matrix and the matrices Ψ_n and Ω_n are defined by

$$\Omega_{n}(\gamma) = \frac{1}{n} \begin{pmatrix} X^{T} D_{11} X & X^{T} D_{12} \mathbf{1} \\ \mathbf{1}^{T} D_{12} X & \mathbf{1}^{T} D_{22} \mathbf{1} \end{pmatrix},$$

$$\Psi_{n}(\gamma) = \frac{1}{n} \begin{pmatrix} X^{T} \begin{pmatrix} D_{11}^{*} - (D_{1}^{*})^{T} D_{1}^{*} \end{pmatrix} X & X^{T} \begin{pmatrix} D_{12}^{*} - (D_{1}^{*})^{T} D_{2}^{*} \end{pmatrix} \mathbf{1} \\ \mathbf{1}^{T} \begin{pmatrix} D_{12}^{*} - (D_{1}^{*})^{T} D_{2}^{*} \end{pmatrix} X & \mathbf{1}^{T} \begin{pmatrix} D_{22}^{*} - (D_{2}^{*})^{T} D_{2}^{*} \end{pmatrix} \mathbf{1} \end{pmatrix},$$

with

$$\boldsymbol{D}_{jk} = diag \left(l_{jki}(\boldsymbol{\gamma}) \right)_{i=1,\dots,n,j,k=1,2} \qquad \boldsymbol{D}_{jk}^* = diag \left(m_{jki}(\boldsymbol{\gamma}) \right)_{i=1,\dots,n,j,k=1,2}$$

and

$$D_j^* = diag(m_{ji}(\gamma))_{i=1,\dots,n'}, j.k = 1, 2.$$

Proof. The consistency is proved for general statistical models in Castilla et al. [27] and the asymptotic distribution of the MRPEs for GLM is derived in Jaenada and Pardo [3]. \Box

3. Wald Type Tests for the GLMs

In this section, we define Wald-type tests for linear null hypothesis of the form

$$\mathbf{H}_0: \boldsymbol{M}^T \boldsymbol{\gamma} = \boldsymbol{m} \text{ vs } \mathbf{H}_1: \boldsymbol{M}^T \boldsymbol{\gamma} \neq \boldsymbol{m}$$
(11)

being $\gamma = (\boldsymbol{\beta}^T, \boldsymbol{\phi})^T$, \boldsymbol{M} a $(k+1) \times r$ full rank matrix and

$$\boldsymbol{m} = (m_1, \dots, m_r)^T \tag{12}$$

a *r*-dimensional vector ($r \le k + 1$). If the nuisance parameter ϕ is known, as with logistic and Poisson regression, the matrix $M = L_{k \times r}$. Additionally, choosing

$$\boldsymbol{M} = (\boldsymbol{L}_{k \times r}, \boldsymbol{O}_{1 \times r})$$

gives rise to a null hypothesis defined by a linear combination of the regression coefficients, β , with ϕ known or unknown. Further, the simple null hypothesis is a particular case when choosing *M* as the identity matrix of rank *k*,

$$H_0: \boldsymbol{\beta} = \boldsymbol{\beta}_0 \text{ vs } H_1: \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$$

with $\boldsymbol{m} = \boldsymbol{\beta}_0 = (\beta_1^0, \dots, \beta_k^0)^T$. In the following we assume that there exist a matrix $A_{\alpha}(\gamma)$ verifying

$$\lim_{n\to\infty} \Psi_n(\gamma) \mathbf{\Omega}_n(\gamma)^{-1} \Psi_n(\gamma) = A_{\alpha}(\gamma).$$

Definition 1. Let $\hat{\gamma}_{\alpha} = (\hat{\beta}_{\alpha}^T, \hat{\phi}_{\alpha})^T$ be the MRPE of $\gamma = (\beta^T, \phi)^T$ for the GLM. The Wald-type tests, based on the MRPE, for testing (11) are defined by

$$W_n(\widehat{\gamma}_{\alpha}) = n \Big(\boldsymbol{M}^T \widehat{\gamma}_{\alpha} - \boldsymbol{m} \Big)^T \Big(\boldsymbol{M}^T \boldsymbol{\Psi}_n(\widehat{\gamma}_{\alpha})^{-1} \boldsymbol{\Omega}_n(\widehat{\gamma}_{\alpha}) \boldsymbol{\Psi}_n(\widehat{\gamma}_{\alpha})^{-1} \boldsymbol{M} \Big)^{-1} \Big(\boldsymbol{M}^T \widehat{\gamma}_{\alpha} - \boldsymbol{m} \Big).$$
(13)

The following theorem presents the asymptotic distribution of the Wald-type test statistics, $W_n(\widehat{\gamma}_{\alpha})$.

Theorem 2. The Wald-type test $W_n(\hat{\gamma}_{\alpha})$ follows asymptotically, under the null hypothesis presented in (11), a chi-square distribution with degrees of freedom equal to the dimension of the vector *m* in (12)

Under the null hypothesis given in (11) the asymptotic distribution of the Wald-type test statistics is a chi-square distribution with r degrees of freedom.

Proof. We know that

$$\sqrt{n}\Big((\widehat{\boldsymbol{\beta}}_{\alpha}^{T},\widehat{\boldsymbol{\phi}}_{\alpha})^{T}-(\boldsymbol{\beta}^{T},\boldsymbol{\phi})^{T}\Big) \xrightarrow[n\to\infty]{L} N(\boldsymbol{0}_{k+1},\boldsymbol{A}_{\alpha}(\boldsymbol{\gamma})^{-1}).$$

Therefore,

$$\sqrt{n} \Big(\boldsymbol{M}^{\mathrm{T}} \widehat{\boldsymbol{\gamma}}_{\alpha} - \boldsymbol{m} \Big) = \sqrt{n} \boldsymbol{M} (\widehat{\boldsymbol{\gamma}}_{\alpha} - \boldsymbol{\gamma}) \xrightarrow[n \to \infty]{L} N(\boldsymbol{0}_{k+1}, \boldsymbol{M}^{\mathrm{T}} \boldsymbol{A}_{\alpha}(\boldsymbol{\gamma})^{-1} \boldsymbol{M}).$$

Now, the result follows taking into account that $\hat{\gamma}_{\alpha}$ is a consistent estimator of γ_0 . \Box

Based on the previous convergence, the null hypothesis in (11) is rejected, if

I

$$\mathcal{N}_n(\widehat{\gamma}_{\alpha}) > \chi^2_{r,\alpha} \tag{14}$$

being $\chi^2_{r,\alpha}$ the $100(1 - \alpha)$ percentile of a chi-square distribution with *r* degrees of freedom.

Finally, let γ_1 be a parameter point verifying $M^T \gamma_1 \neq m$, i.e., γ_1 is not on the null hypothesis. The next result establishes that the Wald-type tests given in (14) are consistent (see Fraser [29]).

Theorem 3. Let γ_1 be a parameter point verifying $M^T \gamma_1 \neq m$. Then the Wald-type tests given in (14) are consistent, i.e.,

$$\lim_{n\to\infty}P_{\gamma_1}\Big(W_n(\widehat{\gamma}_{\alpha})>\chi^2_{r,\alpha}\Big)=1.$$

Proof. See Appendix A. \Box

Remark 1. In the proof of the previous Theorem was established the approximate power function of the Wald-type tests defined in (13),

$$\pi_{W_n(\widehat{\gamma}_{\alpha})}(\gamma_1) \approx 1 - \phi_{\mathcal{N}(0,1)} \left(\frac{1}{\sigma(\gamma_1)} \left(\frac{\chi_{r,\alpha}^2}{\sqrt{n}} - \frac{W_n(\gamma_1)}{\sqrt{n}} \right) \right)$$

7 of 18

where

$$\sigma^{2}(\gamma_{1}) = \left(\frac{\partial l_{\widehat{\gamma}_{\alpha}}(\zeta)}{\partial \zeta^{T}}\right)_{\gamma=\gamma_{1}} A_{\alpha}(\gamma_{1})^{-1} \left(\frac{\partial l_{\widehat{\gamma}_{\alpha}}(\zeta)}{\partial \zeta}\right)_{\gamma=\gamma_{1}}$$

and

$$l_{\widehat{\gamma}_{\alpha}}(\boldsymbol{\zeta}) = \left(\boldsymbol{M}^{T}\widehat{\boldsymbol{\gamma}}_{\alpha} - \boldsymbol{m}\right)^{T} \left(\boldsymbol{M}^{T}\boldsymbol{A}_{\alpha}(\boldsymbol{\zeta})^{-1}\boldsymbol{M}\right)^{-1} \left(\boldsymbol{M}^{T}\widehat{\boldsymbol{\gamma}}_{\alpha} - \boldsymbol{m}\right).$$

From the above expression, the necessary sample size *n* for the Wald-type tests to have a predetermined power, π_0 , is given by $n = [n^*] + 1$, with

$$n^* = rac{A + B + \sqrt{A(A + 2B)}}{2l_{\gamma_1}^2(\gamma_1)}$$

being

$$A = \sigma^{2}(\gamma_{1}) \left(\phi^{-1}(1 - \pi_{0}) \right)^{2}, \quad B = 2\chi^{2}_{r,\alpha} l_{\gamma_{1}}(\gamma_{1})$$

and $[\cdot]$ the integer part.

In accordance with Maronna et al. [30], the breakdown point of the estimators $\hat{\gamma}_{\alpha}$ of a parameter γ is the largest amount of contamination that the data may contain such that $\hat{\gamma}_{\alpha}$ still gives enough information about γ . The derivation of a general breakdown points it is in general not easy, so it may deserve a separate paper where it may be jointly considered the replacement finite-sample breakdown point introduced by Donoho and Huber [31]. Although breakdown point is an important theoretical concept in robust statistics, perhaps is more useful the definition of breakdown point associated to a finite sample: replacement finite-sample break down point. More details can be seen in Section 3.2.5 of Maronna et al. [30].

4. Influence Function

We derive in this section the IF of the MRPEs of the parameters $\gamma = (\beta^T, \phi)^T$ and Wald-type statistics based on these MRPEs, $W_n(\hat{\gamma}_\alpha)$. The influence function (IF) of an estimator quantifies the impact of an infinitesimal perturbation in the true distribution of the data on the asymptotic value of the resulting parameter estimate (in terms of the corresponding statistical functional). An estimator is said to be robust if its IF is bounded. If we denote $G = (G_1, \ldots, G_n)$ the true distributions underlying the data, the functional $T_\alpha(G)$ and associated to the MRPE for the parameters γ is such that

$$\frac{1}{n}\sum_{i=1}^{n}R_{\alpha}(f_i(y,\boldsymbol{T}_{\alpha}(\boldsymbol{G})),g_i(y))=\min_{\boldsymbol{\gamma}}\frac{1}{n}\sum_{i=1}^{n}R_{\alpha}(f_i(y,\boldsymbol{\gamma}),g_i(y)).$$

The IF of a estimator is defined as the limiting standardized bias due to infinitesimal contamination. That is, given a contaminated distribution at the point (y_t, x_t) , $G_{\varepsilon} = (1 - \varepsilon)G + \varepsilon \Delta_{(y_t, x_t)}$ with $\Delta_{(y_t, x_t)}$ the degenerate distribution at (y_t, x_t) , the IF of the estimator $\hat{\gamma}_{\alpha}$ in terms of its associated functional $T_{\alpha}(G)$ is computed as

$$\operatorname{IF}((y_t, \boldsymbol{x}_t), \boldsymbol{T}_{\alpha}(\boldsymbol{G})) = \lim_{\varepsilon \to 0} \frac{\boldsymbol{T}_{\alpha}(\boldsymbol{G}_{\varepsilon}) - \boldsymbol{T}_{\alpha}(\boldsymbol{G})}{\varepsilon}.$$

In the following, let us denote $T_{\alpha}(G) = (T_{\alpha}^{\beta}(G), T_{\alpha}^{\phi}(G))$, where $T_{\alpha}^{\beta}(G)$ and $T_{\alpha}^{\phi}(G)$ are the functionals associated the parameters β and ϕ , respectively. Then, they must satisfy the estimating equations of the MRPE given by

$$\sum_{i=1}^{n} \frac{x_{i}}{L_{\alpha}^{i}\left(\left(T_{\alpha}^{\beta}(G), T_{\alpha}^{\phi}(G)\right)\right)} \left\{ M_{i}\left(y_{i}, \left(T_{\alpha}^{\beta}(G), T_{\alpha}^{\phi}(G)\right)\right) - N_{i}\left(y_{i}, \left(T_{\alpha}^{\beta}(G), T_{\alpha}^{\phi}(G)\right)\right) \right\} = \mathbf{0}_{k}$$

$$\sum_{i=1}^{n} \frac{1}{L_{\alpha}^{i}\left(\left(T_{\alpha}^{\beta}(G), T_{\alpha}^{\phi}(G)\right)\right)} \left\{ M_{i}^{*}\left(y_{i}, \left(T_{\alpha}^{\beta}(G), T_{\alpha}^{\phi}(G)\right)\right) - N_{i}^{*}\left(y_{i}, \left(T_{\alpha}^{\beta}(G), T_{\alpha}^{\phi}(G)\right)\right) \right\} = 0$$
(15)

where the quantities $L_{\alpha^i(\gamma)}$, $M_i(y_i, \gamma)$, $N_i(y_i, \gamma)$, $M_i^*(y_i, \gamma)$ and $N_i^*(y_i, gamma)$ are defined in Section 2. Now, evaluating the previous equation at the contaminated distribution G_{ε} , implicitly differentiating the estimating equations in ε and evaluating them at $\varepsilon = 0$, we can obtain the expression of the IF for the GLM.

We first derive the expression IF of MRPEs at the $i_0 - th$ direction. For this purpose, we consider the contaminated distributions

$$G_{i_0,\varepsilon} = (G_1, \ldots, G_{i_0-1}, G_{i_0,\varepsilon}, G_{i_0+1}, \ldots, G_n),$$

with $G_{i_0,\varepsilon} = (1 - \varepsilon)G_{i_0} + \varepsilon \Delta_{(y_{i_0}, x_{i_0})}$. Here, only the i_0 -th component of the vector of distributions is contaminated. If the true density function g_i of each variable belongs to the exponential model, we have that

$$g_i(y) = \begin{cases} f_i(y, \gamma) & i \neq i_0\\ (1-\varepsilon)f_i(y, \gamma) + \varepsilon \Delta_{(y_{i_0}, \mathbf{x}_{i_0})}(y) & i = i_0. \end{cases}$$

Accordingly, we define

$$\gamma_{\varepsilon}^{i_0} = T_{\alpha}(G_1,\ldots,G_{i_0-1},G_{i_0,\varepsilon},G_{i_0+1},\ldots,G_n)$$

the MRPE when the true distribution underlying the data is $G_{i_0,\varepsilon}$. Based on Remark 5.2 in Castilla et al. [27] the IF of the MRPE at the $i_0 - th$ direction with (y_{i_0}, x_{i_0}) the point of contamination is given by

$$IF((y_{i_0}, \mathbf{x}_{i_0}), \mathbf{T}_{\alpha}, \mathbf{G}) = \left(\frac{\partial \mathbf{T}_{\alpha}(\mathbf{G}_{i_0, \gamma_{\varepsilon}})}{\partial \varepsilon}\right)_{\varepsilon=0}$$

= $\mathbf{\Psi}_n(\gamma)^{-1} \frac{f_{i_0}(y_{i_0}, \gamma)^{\alpha}}{\int f_{i_0}(y, \gamma)^{\alpha+1} dy} \begin{pmatrix} K_{1i}(y_{i_0}, \gamma) - f_{i_0}(y_{i_0}, \gamma)^{-\alpha} N_{i_0}(y_{i_0}, \gamma) \\ K_{2i}(y_{i_0}, \gamma) - f_{i_0}(y_{i_0}, \gamma)^{-\alpha} N_{i_0}^*(y_{i_0}, \gamma) \end{pmatrix} \begin{pmatrix} \mathbf{x}_{i_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}.$

In a similar manner, the IF in all directions (i.e., all components of the vector of distributions are contaminated) has the following expression

$$IF((y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n), \mathbf{T}_{\alpha}, \mathbf{G}) = \left(\frac{\partial T_{\alpha}(\mathbf{G}_{\gamma_{\varepsilon}})}{\partial \varepsilon}\right)_{\varepsilon=0}$$

= $\mathbf{\Psi}_n(\gamma)^{-1} \sum_{i=1}^n \left(\frac{f_i(y_i, \gamma)^{\alpha}}{\int f_i(y, \gamma)^{\alpha+1} dy} \begin{pmatrix} K_{1i}(y_i, \gamma) - f_i(y_i, \gamma)^{-\alpha} N_i(y_i, \gamma) \\ K_{2i}(y_i, \gamma) - f_i(y_i, \gamma)^{-\alpha} N_i^*(y_i, \gamma) \end{pmatrix} \begin{pmatrix} \mathbf{x}_i & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix},$

with $(y_1, x_1), \ldots, (y_n, x_n)$ the point of contamination. We next derive the expression of the IF for the Wald-type tests presented in Section 3. The statistical functional associated with the Wald-type tests for the linear null hypothesis (11) at the distributions $G = (G_1, \ldots, G_n)$, ignoring the constant *n*, is given by

$$W_{\alpha}(G) = \left(\boldsymbol{M}^{T}\boldsymbol{T}_{\alpha}(G) - \boldsymbol{m}\right)^{T} \left(\boldsymbol{M}^{T}\boldsymbol{A}_{\alpha}(\boldsymbol{T}_{\alpha}(G))^{-1}\boldsymbol{M}\right)^{-1} \left(\boldsymbol{M}^{T}\boldsymbol{T}_{\alpha}(G) - \boldsymbol{m}\right).$$
(16)

Again, evaluating the Wald-type test functionals at the contaminated distribution G_{ε} and implicitly differentiating the expression, we can get the expression of it IF. In particular, the IF of the Wald-type test statistics at the i_0 -th direction and the contamination point (y_{i_0}, x_0) is given by

$$IF_1((y_{i_0}, \mathbf{x}_0), \mathbf{W}_{\alpha}, \mathbf{G}) = \left(\frac{\partial W_{\alpha}(\mathbf{G}_{i_0, \varepsilon})}{\partial \varepsilon}\right)_{\varepsilon = 0}$$

= $2\left(\mathbf{M}^T \mathbf{T}_{\alpha}(\mathbf{G}) - \mathbf{m}\right)^T \left(\mathbf{M}^T \mathbf{A}_{\alpha}(\mathbf{T}_{\alpha}(\mathbf{G}))^{-1}\mathbf{M}\right)^{-1} \mathbf{M}^T IF((y_{i_0}, \mathbf{x}_0), \mathbf{T}_{\alpha}, \mathbf{G}).$

Evaluating the previous expression at the null hypothesis, $M^T T_{\alpha}(G) = m$, the IF becomes identically zero,

$$F_1((y_{i_0}, \boldsymbol{x}_{i_0})\boldsymbol{W}_{\alpha}, \boldsymbol{G}) = \boldsymbol{0}_{k+1}.$$

Therefore, it is necessary to consider the second order IF of the proposed Wald-type tests. Twice differentiating in $W_{\alpha}(G_{\varepsilon})$, we get

$$IF_{2}((y_{i_{0}}, \boldsymbol{x}_{i_{0}}), \boldsymbol{W}_{\alpha}, \boldsymbol{G}) = \left(\frac{\partial^{2} \boldsymbol{W}_{\alpha}(\boldsymbol{G}_{i_{0},\varepsilon})}{\partial \varepsilon^{2}}\right)_{\varepsilon=0}$$

= 2IF((y_{i_{0}}, \boldsymbol{x}_{i_{0}}), \boldsymbol{T}_{\alpha}, \boldsymbol{F}_{\beta})^{T} \boldsymbol{M} \left(\boldsymbol{M}^{T} \boldsymbol{A}_{\alpha}(\boldsymbol{T}_{\alpha}(\boldsymbol{G}))^{-1} \boldsymbol{M}\right)^{-1} \boldsymbol{M}^{T} IF((y_{i_{0}}, \boldsymbol{x}_{i_{0}}), \boldsymbol{T}_{\alpha}, \boldsymbol{G}).

Finally, the second order IF of the Wald-type tests in all directions is given by

$$IF_{2}((y_{1}, \boldsymbol{x}_{1}), \dots, (y_{n}, \boldsymbol{x}_{n}), \boldsymbol{W}_{\alpha}, \boldsymbol{G}) = \left(\frac{\partial^{2} \boldsymbol{W}_{\alpha}(\boldsymbol{G}_{\varepsilon})}{\partial \varepsilon^{2}}\right)_{\varepsilon=0}$$

= $2IF((y_{1}, \boldsymbol{x}_{1}), \dots, (y_{n}, \boldsymbol{x}_{n}), \boldsymbol{T}_{\alpha}, \boldsymbol{G})^{T} \boldsymbol{M} \left(\boldsymbol{M}^{T} \boldsymbol{A}_{\alpha}(\boldsymbol{T}_{\alpha}(\boldsymbol{G}))^{-1} \boldsymbol{M}\right)^{-1} \boldsymbol{M}^{T}$
 $\cdot IF((y_{1}, \boldsymbol{x}_{1}), \dots, (y_{n}, \boldsymbol{x}_{n}), \boldsymbol{T}_{\alpha}, \boldsymbol{G}).$

To asses the robustness of the MRPEs and Wald-type test statistics we must discuss the boundedness of the corresponding IF. The boundedness of the second order IF of the Wald-type test statistics is determined by the boundedness of the IF of the MRPEs. Further, the matrix $\Psi_n(\gamma)$ is assumed to be bounded, so the robustness of the estimators only depend on the second factor of the IF. Most standard GLMs enjoy such properties for positives values of α , but the influence function is unbounded at $\alpha = 0$, corresponding with the MLE. As an illustrative example, Figure 1 plots the IF of the MRPEs for the Poisson regression model with different values of $\alpha = 0, 0.5$ at one direction. The model is fitted with only one covariate, the parameter ϕ is known for Poisson regression ($\phi = 1$) and the true regression vector is fixed $\beta = 1$. As shown, the IF of the MRPEs with positives values of α are bounded, whereas the IF of the MLE is not, indicating it lack of robustness.



Figure 1. IF of MRPEs with $\alpha = 0$ (**left**) and $\alpha = 0.5$ (**right**) of Poisson regression model.

5. Numerical Analysis: Poisson Regression Model

We illustrate the proposed robust method for the Poisson regression model. As pointed out in Section 1 the Poisson regression model belongs to the GLM with known shape parameter $\phi = 1$, location parameter $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ and known functions $b(\theta_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ and $c(y_i) = -\log(y_i!)$. Since the nuisance parameter is known, for the seek of simplicity in the following we only use $\boldsymbol{\beta} = \gamma$. In Poisson regression, the mean of the response variable is linked to the linear predictor through the natural logarithm, i.e., $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$. Thus, we can apply the previous proposed method to estimate the vector of regression parameters $\boldsymbol{\beta}$ with objective function given in Equation (5).

The results provided are computed in the software R. The minimization of the objective function is performed using the implemented *optim()* function, which applies the Nelder-Mead iterative algorithm (Nelder and Mead [32]). Nelder-Mead optimization algorithm is robust although relatively slow. The corresponding objective function $T_n^{\alpha}(\gamma)$ given in (5) is highly nonlinear and requires the evaluation of nontrivial quantities. Further, the computation of the Wald-type test statistics defined in (13) requires to evaluate the covariance matrix of the MRPEs, involving nontrivial integrals. Some simplified expressions of the main quantities defined throughout the paper for the Poisson regression model, such as $L^{i}_{\alpha}(\beta)$, $K_{1i}(y,\beta)$, $N_{i}(y,\beta)$, $m_{1i}(\beta)$, $m_{11i}(\beta)$ or $l_{11i}(\beta)$, are given in the Appendix B. There is no closed expression for these quantities, and they need to be approximated numerically. Since the minimization is iteratively performed, computing such expressions at each step of the algorithm and for each observation may entail an increased computational burden. Nonetheless, the complexity is not significant for low-dimensional data. On the other hand, the optimum in (5) need not to be uniquely defined, since the objective function may have several local minima. Then, the choice of the initial value of the iterative algorithm is crucial. Ideally, a good initial point should be consistent and robust. In our results the MLE is used as initial estimate for the algorithm.

We analyze the performance of the proposed methods in Poisson regression through a simulation study. We asses the behavior of the MRPE under the sparse Poisson regression model with k = 12 covariates but only 3 significant variables. We set the 12-dimensional regression parameter $\beta = (1.8, 1, 0, 0, 1.5, 0, ... 0)$ and we generate the explanatory variables, x_i , from the standard uniform distribution with variance-covariance matrix having Toeplitz structure, with the (j, l)-th element being $0.5^{|j-l|}$, j, l = 1, ..., p. The response variables are generated from the Poisson regression model with mean $\mu_i = x_i^T \beta$, $Y_i \sim \mathcal{P}(\mu_i)$. To evaluate the robustness of the proposed estimators, we contaminate the responses using a perturbed distribution of the form $(1 - b)\mathcal{P}(\mu_i) + b\mathcal{P}(2\mu_i)$, where *b* is a realization of a Bernoulli variable with parameter ε so called the contamination level. That is, the distribution of the contaminated responses lies in a small neighbourhood of the assumed model. We repeat the process R = 1000 for each value of α .

Figure 2 presents the mean squared error of the estimate (MSE), MSE = $||\hat{\beta}_{\alpha} - \beta||_2$, (left) and the MSE on the prediction (right) against contamination level on data for different values of $\alpha = 0, 0.1, 0.3, 0.5$ and 0.7. The sample size is fixed at n = 200 and the MSE on the prediction is calculated using n = 200 new observations following the true model. As shown, greater values of α correspond to more robust estimators, revealing the role of the tuning parameter on the robustness gain. Most strikingly, the MSE grows linearly for the MLE, while the proposed estimators manage to maintain a low error in all contaminated scenarios.



Figure 2. Mean Squared Error (MSE) on estimation (**left**) and prediction (**right**) against contamination level on data.

Furthermore, it is to be expected that the error of the estimate decreases with larger samples sizes. In this regard, Figure 3 shows the MSE for different values of $\alpha = 0, 0.1, 0.3, 0.5$ and 0.7, against the sample size in the absence of contamination (left) and under 5% of contamination. Our proposed estimators are more robust than the classical MLE with almost all contaminated scenarios, since the MSE committed is lower for all positives values of α than for $\alpha = 0$ (corresponding to the MLE), except for too small sample sizes. Conversely, the MLE is, as expected, the most efficient estimator in absence of contamination, closely to our proposed estimators with $\alpha = 0.1, 0.3$, highlighting the importance of α in controlling the trade-off between efficiency and robustness. In this regard, values of α about 0.3 perform the best taking into account the low loss of efficiency and the gain in robustness. Finally, note that small sample sizes adversely affect to greater values of α .





On the other hand, one could be interested on testing the significance of the selected variables. For this purpose, we simplify the true model and we examine the performance of the proposed Wald-type test statistics under different true coefficients values. In particular, let us consider a Poisson regression model with only two covariates, generated from the uniform distribution as before, and the linear null hypothesis

$$H_0: \beta_2 = 0. (17)$$

That is, we are interested in assessing the significance of the second variable. The sample size if fixed at n = 200 and the true value of the component of the regression vector is set $\beta_1 = 1$. We study the power of the tests under increasing signal of the second parameter β_2 and increasing contamination level. Here, the model is contaminated by perturbing the true distribution with $(1 - b)\mathcal{P}(\mu_i) + b\mathcal{P}(\tilde{\mu}_i)$, where $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ is the mean of the Poisson variable in the absence of contamination, $\tilde{\mu}_i = \mathbf{x}_i^T \boldsymbol{\beta}$ is the contaminated mean, with $\boldsymbol{\beta} = (1, 0)$, and *b* is a realization of a Bernoulli variable with probability of success ε . Table 1 presents the rejection rate of the Wald-type test statistics for different true values of β_2 under different contaminated scenarios. As expected, stronger signals produce higher power for all Wald-type test. Moreover, the power of the Wald-type test statistics based on the MRPEs with positives values of α keeps sufficiently high. Then, our proposed robust estimators are able to detect the significance of the variable even in heavily contaminated scenarios.

β2	α	Contamination Level					
		0	5%	10%	15%	20%	25%
0.3	0	0.332	0.264	0.227	0.187	0.157	0.141
	0.1	0.435	0.376	0.328	0.285	0.251	0.223
	0.3	0.557	0.511	0.483	0.416	0.390	0.360
	0.5	0.617	0.563	0.533	0.493	0.467	0.427
	0.7	0.638	0.590	0.568	0.536	0.513	0.476
	0	0.756	0.730	0.683	0.621	0.551	0.493
	0.1	0.833	0.798	0.775	0.736	0.681	0.622
0.5	0.3	0.885	0.870	0.864	0.829	0.792	0.752
	0.5	0.895	0.891	0.886	0.867	0.842	0.814
	0.7	0.901	0.897	0.893	0.879	0.854	0.832
	0	0.971	0.979	0.968	0.948	0.915	0.862
	0.1	0.980	0.988	0.983	0.973	0.962	0.932
0.7	0.3	0.988	0.995	0.992	0.987	0.985	0.969
	0.5	0.989	0.995	0.995	0.992	0.992	0.977
	0.7	0.989	0.995	0.993	0.995	0.990	0.983

Table 1. Rejection rate of Wald-type test statistics based on MRPEs with different true values of β_2 and contamination levels.

6. Real Data Applications

6.1. Example I: Poisson Regression Regression

We finally apply our proposed estimators in a real dataset arising from Crohn's disease. The data were first studied in Lô and Ronchetti [33] to asses the adverse events of a drug. The clinical study included 117 patients affected by the disease, for whom information was recorded for 7 explanatory variables: BMI (body mass index), HEIGHT, COUNTRY (one of the two countries where the patient lives), SEX, AGE, WEIGHT, and TREAT (the drug taken by the patient in factor form: placebo, Dose 1, Dose 2), in addition to the response variable AE (number of adverse events). Lô and Ronchetti [33] considered a Poisson regression model for the Crohn data and determined that only variables Dose 1, BMI, HEIGHT, SEX, AGE, and COUNTRY may be essentially significant. Further, they flagged observations 23rd, 49th, and 51st to be highly influential on the classical analysis. Table 2 presents the estimated coefficient of the explanatory variable when fitting the Poisson regression model. Robust methods suggest higher coefficients for the variables BMI and AGE, whereas fewer values for the coefficients of the categorical variables COUNTRY, SEX, Dose 1.

Table 2. Estimated coefficients for Crohn's disease data for different values of α with original data and clean data (after removing influential observations).

	Intercept	BMI	Height	Age	Country	Sex	Dose 1
Original Data							
MLE ($\alpha = 0$)	6.261	0.026	-0.037	0.012	-0.394	-0.646	-0.533
$\alpha = 0.1$	5.197	0.037	-0.033	0.014	-0.489	-0.800	-0.469
$\alpha = 0.3$	4.798	0.058	-0.036	0.021	-0.545	-1.284	-0.832
$\alpha = 0.5$	4.391	0.067	-0.037	0.028	-0.557	-1.535	-1.036
$\alpha = 0.7$	5.699	0.067	-0.047	0.036	-0.737	-1.759	-1.157

Following the discussion in Lô and Ronchetti [33], classical tests may not select variable AGE to be significant. Then, we propose testing the significance of that variable using Wald-type test statics based on different values α . Table 3 shows the p-values of the corresponding tests with null hypothesis H₀:AGE = 0, with the original data and after removing the outlying observations.

Table 3. *p*-values of test with null hypothesis H_0 :AGE = 0 with original and clean data (after removing influential observations).

	Original Data	Clean Data
MLE ($\alpha = 0$)	0.059	0.011
lpha=0.1	0.018	0.004
$\alpha = 0.3$	0.001	0.000
lpha=0.5	0.000	0.000
lpha=0.7	0.000	0.000

The MLE rejects the significance of the variable AGE when the original data are used, whereas the Wald-type test statistics with positives values of α indicate strong evidence against the null hypothesis. In contrast, if the influential observations are removed, all Wald-type test statistics agree in the significance of the variable. This example illustrates the robustness of the proposed statistics.

6.2. Example II: Binomial Regression

We finally illustrate the applicability of the MRPE for robust inference in the binomial regression model. We examine the damaged carrots dataset, first studied in Phelps [34] and later discussed by Cantoni and Ronchetti [8] and Ghosh and Basu [13] to illustrate robust procedures for binomial regression. The data contain 24 samples, among which the 14th observation was flagged as an outlier in the y-space but not a leverage point. The data are issued from a soil experiment and give the proportion of carrots showing insect damage in a trial with three blocks and eight dose levels of insecticide. The explanatory variables are the logarithm transform of the dose (Logdose) and two dummy variables for Blocks 1 and 2.

Binomial regression is a natural extension of the logistic regression when the response variable *Y* does not follow a Bernoulli distribution but a Binomial distribution counting the number of successes in a series of *m* independent Bernoulli trials. Binomial regression model belongs to the GLM with known shape parameter $\phi = 1$, location parameter $\theta_i = x_i^T \beta$ and functions $b(\theta_i) = m \log(1 + \exp(x_i^T \beta))$ and $c(y_i) = \log(\binom{m}{y_i})$. The mean of the response variable is then linked to the linear predictor through the logit function, i.e.,

$$\log\left(\frac{\mu_i}{m-\mu_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta}$$

Table 4 presents the estimated coefficients of the regression vector for the carrots data using the MLE and robust MRPEs when the model is fitted with the original data and the model fitted without the outlying observation. The results provided are computed in the same manner as in Section 5, adapting the corresponding quantities in Equation (5) for the binomial model. All integrals involved were numerically approximated, and the MLE is used as initial estimate for the optimization algorithm. The influence of observation 14 stands out when using the MLE; the estimated coefficients are remarkably different when fitting the model with and without observation 14. In contrast, all methods estimate similar coefficients after removing the outlying observation, coinciding with the robust estimates for moderately high values of the tuning parameter α .

	Intercept	Logdose	B1	B2			
	Original Data						
MLE ($\alpha = 0$)	1.480	-1.817	0.542	0.843			
lpha = 0.1	1.729	-1.949	0.527	0.755			
$\alpha = 0.3$	2.017	-2.100	0.479	0.652			
$\alpha = 0.5$	2.090	-2.134	0.386	0.625			
lpha=0.7	2.150	-2.161	0.258	0.615			
Clean Data							
MLE ($\alpha = 0$)	2.141	-2.179	0.546	0.636			
lpha = 0.1	2.126	-2.167	0.529	0.633			
$\alpha = 0.3$	2.105	-2.149	0.479	0.627			
$\alpha = 0.5$	2.108	-2.144	0.385	0.621			
lpha=0.7	2.154	-2.163	0.257	0.614			

Table 4. Estimated coefficients for damaged carrots data for different values of α with original data and clean data (after outliers removal).

7. Conclussions

In this paper, we presented the MRPE and Wald-type test statistics for GLMs. The proposed MRPEs and statistics have appealing robustness properties where the data are contaminated due to outliers or leverage points. MRPEs are consistent and asymptotically normal and represent an attractive alternative to the classical nonrobust methods. Additionally, robust Wald-type test statistics, based on the MRPEs, were developed. Through the study of the IFs and the development of an extensive simulation study, we proved their robustness from a theoretical and practical point of view, respectively. In particular, we illustrated the superior performance of the MRPEs and the corresponding Wald-type tests for the Poisson regression model.

Author Contributions: Conceptualization, M.J. and L.P.; methodology, M.J. and L.P.; software, M.J. and L.P.; validation, M.J. and L.P.; formal analysis, M.J. and L.P.; investigation, M.J. and L.P.; resources, M.J. and L.P.; data curation, M.J. and L.P.; writing—original draft preparation, M.J. and L.P.; writing—review and editing, M.J. and L.P.; visualization, M.J. and L.P.; supervision, M.J. and L.P.; project administration, M.J. and L.P.; funding acquisition, M.J. and L.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Spanish Grants PGC2018-095194-B-100 (L. Pardo and M. Jaenada) and FPU/018240 (M. Jaenada).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The real datasets are publicly available on the R package *robustbase* in CRAN under the names of *CrohnD* (Poisson regression example) and *carrots* (binomial regression example).

Acknowledgments: We are very grateful to the referees and associate editor for their helpful comments and suggestions. This research is supported by the Spanish Grants PGC2018-095194-B-100 (L. Pardo and M. Jaenada) and FPU/018240 (M. Jaenada). M.Jaenada and L. Pardo are members of the Instituto de Matematica Interdisciplinar, Complutense University of Madrid.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

DPD	Density Power Divergence
IF	Influence Function
GLM	Genelarized Linear Model
LRM	Linear Regression Model
MLE	Maximum Likelihood Estimator
MRPE	Minimum Rényi Pseudodistance Estimator
RP	Rényi Pseudodistance

Appendix A. Proof of Theorem 3

Let us define

$$l_{\boldsymbol{\eta}}(\boldsymbol{\zeta}) = \left(\boldsymbol{M}^{T}\boldsymbol{\eta} - \boldsymbol{m}\right)^{T} \left(\boldsymbol{M}^{T}\boldsymbol{A}_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})^{-1}\boldsymbol{M}\right)^{-1} \left(\boldsymbol{M}^{T}\boldsymbol{\eta} - \boldsymbol{m}\right)$$

so the Wald-type test statistic is such that

$$nl_{\widehat{\gamma}_{\alpha}}(\widehat{\gamma}_{\alpha}) = W_n(\widehat{\gamma}_{\alpha})$$

We know that $\widehat{\gamma}_{\alpha} \xrightarrow[n \to \infty]{p} \gamma_1$ and therefore $l_{\widehat{\gamma}_{\alpha}}(\gamma_1)$ and $l_{\gamma_1}(\gamma_1)$ have the same asymptotic distribution. A first order Taylor expansion of $g(\zeta) = l_{\widehat{\gamma}_{\alpha}}(\zeta)$ at $\widehat{\gamma}_{\alpha}$ around γ_1 gives,

$$l_{\widehat{\gamma}_{\alpha}}(\widehat{\gamma}_{\alpha}) = l_{\widehat{\gamma}_{\alpha}}(\gamma_{1}) + \left(\frac{\partial l_{\widehat{\gamma}_{\alpha}}(\zeta)}{\partial \zeta^{T}}\right)_{\gamma=\gamma_{1}}(\widehat{\gamma}_{\alpha}-\gamma_{1}) + o_{p}(\|\widehat{\gamma}_{\alpha}-\gamma_{1}\|).$$

Based on the asymptotic distribution of $\widehat{\gamma}_{\alpha}$ we have

$$\sqrt{n}o_p(\|\widehat{\gamma}_{\alpha} - \gamma_1\|) = o_p(1)$$

therefore

$$\sqrt{n} \Big(l_{\widehat{\gamma}_{\alpha}}(\widehat{\gamma}_{\alpha}) - l_{\gamma_1}(\gamma_1) \Big) \text{ and } \sqrt{n} \Big(\frac{\partial l_{\widehat{\gamma}_{\alpha}}(\zeta)}{\partial \zeta^T} \Big)_{\gamma = \gamma_1} (\widehat{\gamma}_{\alpha} - \gamma_1)$$

have asymptotically the same distribution, i.e.,

$$\sqrt{n}\Big(l_{\widehat{\gamma}_{\alpha}}(\widehat{\gamma}_{\alpha})-l_{\gamma_{1}}(\boldsymbol{\theta}_{1})\Big) \xrightarrow[n \to \infty]{L} N\left(\boldsymbol{0}_{k+1},\left(\frac{\partial l_{\widehat{\gamma}_{\alpha}}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{T}}\right)_{\gamma=\gamma_{1}}\boldsymbol{A}_{\alpha}(\gamma_{1})^{-1}\left(\frac{\partial l_{\widehat{\gamma}_{\alpha}}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}}\right)_{\gamma=\gamma_{1}}\right).$$

Now, we shall denote,

$$\sigma^{2}(\boldsymbol{\gamma}_{1}) = \left(\frac{\partial l_{\widehat{\boldsymbol{\gamma}}_{\alpha}}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}^{T}}\right)_{\boldsymbol{\gamma} = \boldsymbol{\gamma}_{1}} \boldsymbol{A}_{\alpha}(\boldsymbol{\gamma}_{1})^{-1} \left(\frac{\partial l_{\widehat{\boldsymbol{\gamma}}_{\alpha}}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}}\right)_{\boldsymbol{\gamma} = \boldsymbol{\gamma}_{1}}.$$

Then, we have,

$$\begin{split} P_{\gamma_1}\Big(W_n(\widehat{\gamma}_{\alpha}) > \chi^2_{r,\alpha}\Big) &= P_{\gamma_1}\Big(W_n(\widehat{\gamma}_{\alpha}) - nl_{\gamma_1}(\gamma_1) > \chi^2_{r,\alpha} - nl_{\gamma_1}(\gamma_1)\Big) \\ &= P_{\gamma_1}\left(\frac{\sqrt{n}}{\sigma(\gamma_1)}\Big(l_{\widehat{\gamma}_{\alpha}}(\widehat{\gamma}_{\alpha}) - l_{\gamma_1}(\gamma_1)\Big) > \frac{1}{\sigma(\gamma_1)}\left(\frac{\chi^2_{r,\alpha}}{\sqrt{n}} - \sqrt{n}l_{\gamma_1}(\gamma_1)\right)\Big) \\ &\approx 1 - \phi_{N(0,1)}\left(\frac{1}{\sigma(\gamma_1)}\left(\frac{\chi^2_{r,\alpha}}{\sqrt{n}} - \sqrt{n}l_{\gamma_1}(\gamma_1)\right)\right), \end{split}$$

where $\phi_{N(0,1)}(t)$ represents the distribution function of a standard normal distribution evaluated at *t*. Finally,

$$\lim_{n\to\infty}P_{\gamma_1}\Big(W_n(\widehat{\gamma}_{\alpha})>\chi^2_{r,\alpha}\Big)=1.$$

Appendix B. Poisson Regression Model

We derive here some explicit expression for the particular case of the Poisson regression. Following the discussion in Section 5, we denote here $\gamma = \beta$ since the nuisance parameter is known, $\phi = 1$. The Poisson distribution with parameter $e^{x_i^T \beta}$ is given by

$$f_i(y,\boldsymbol{\beta}) = \frac{1}{y!} e^{-e^{x_i^T\boldsymbol{\beta}}} e^{yx_i^T\boldsymbol{\beta}}, y = 0, 1, \dots$$

Differentiating its logarithm with respect to the regression vector, we get

 $\frac{\partial \log f_i(y, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = (y - e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}) \boldsymbol{x}_i^T.$

 $K_{1i}(y,\boldsymbol{\beta}) = y - e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}.$

so we can write

Further, we have that

$$N_i(y,\boldsymbol{\beta}) = \frac{f_i(y,\boldsymbol{\beta})^{\alpha}}{\sum\limits_{y=0}^{\infty} f_i(y,\boldsymbol{\beta})^{\alpha+1}} \sum\limits_{y=0}^{\infty} f_i(y,\boldsymbol{\beta})^{\alpha+1} \Big(y - e^{\boldsymbol{x}_i^T \boldsymbol{\beta}} \Big).$$

so the estimating equations of the Poisson regression model are given by

$$\sum_{i=1}^{n} \frac{1}{L_{\alpha}^{i}(\boldsymbol{\beta})} \left(f_{i}(\boldsymbol{y}_{i}, \boldsymbol{\beta})^{\alpha} \left(\boldsymbol{y}_{i} - \boldsymbol{e}^{\boldsymbol{x}_{i}^{T} \boldsymbol{\beta}} \right) - N_{i}(\boldsymbol{y}_{i}, \boldsymbol{\beta}) \right) \boldsymbol{x}_{i} = \boldsymbol{0}_{k}.$$
 (A1)

For $\alpha = 0$, we have

$$N_i(y_i, \boldsymbol{\beta}) = 0$$
 and $L^i_{\alpha}(\boldsymbol{\beta}) = 1$

so the estimating equations are given by

$$\sum_{i=1}^{n} \left(y_i - e^{\mathbf{x}_i^T \mathbf{\beta}} \right) \mathbf{x}_i = \mathbf{0}_k,$$

yielding to the maximum likelihood estimating equations.

On the other hand, the asymptotic distribution of $\hat{\beta}_{\alpha}$ is given by

$$\sqrt{n} \left(\boldsymbol{X}^T \boldsymbol{D}_{11} \boldsymbol{X} \right)^{-\frac{1}{2}} \frac{1}{n} \boldsymbol{X}^T \left(\boldsymbol{D}_{11}^* - \left(\boldsymbol{D}_1^* \right)^T \boldsymbol{D}_1^* \right) \boldsymbol{X} \left(\widehat{\boldsymbol{\beta}}_{\alpha} - \boldsymbol{\beta} \right) \xrightarrow[n \to \infty]{L} N(\boldsymbol{0}_k, \boldsymbol{I}_k)$$

being

$$D_{11} = diag(l_{11i}(\boldsymbol{\beta})),$$

with

$$l_{11i}(\boldsymbol{\beta}) = \frac{1}{L_{\alpha}^{i}(\boldsymbol{\beta})^{2}} \sum_{y=0}^{\infty} f_{i}(y, \boldsymbol{\beta})^{2\alpha+1} (K_{1i}(y, \boldsymbol{\beta}) - m_{1i}(\boldsymbol{\beta}))$$

and

$$m_{1i}(\boldsymbol{\beta}) = \frac{1}{\sum_{y=0}^{\infty} f_i(y, \boldsymbol{\beta})^{\alpha+1}} \sum_{y=0}^{\infty} f_i(y, \boldsymbol{\beta})^{\alpha+1} \left(y - e^{\mathbf{x}_i^T \boldsymbol{\beta}} \right).$$

Finally

$$\boldsymbol{D}_{11}^* = diag(m_{11i}((\boldsymbol{\beta}))) = \frac{1}{\sum\limits_{y=0}^{\infty} f_i(y, \boldsymbol{\beta})^{\alpha+1}} \sum\limits_{y=0}^{\infty} f_i(y, \boldsymbol{\beta})^{\alpha+1} \left(y - e^{\boldsymbol{x}_i^T \boldsymbol{\beta}}\right)^2$$

References

- 1. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. J. R. Stat. Soc. 1972, 135, 370–384. [CrossRef]
- McCullagh, P.; Nelder, J.A. Generalized Linear Models. In Monographs on Statistics and Applied Probability; Chapman and Hall: London, UK, 1983.
- 3. Jaenada, M.; Pardo, L. The minimum Renyi's Pseudodistances estimators for Generalized Linear Models. In *Data Analysis and Related Applications: Theory and Practice*; Proceeding of the ASMDA; Wiley: Athens, Greece, 2021.
- 4. Stefanski, L.A.; Carroll, R.J.; Ruppert, D. Optimally bounded score functions for generalized linear models with applications to logistic regression. *Biometrika* **1986**, *73*, 413–424. [CrossRef]
- 5. Krasker, W.S.; Welsch, R.E. Efficient bounded-influence regression estimation. J. Am. Stat. Assoc. 1982, 77, 595–604. [CrossRef]
- 6. Künsch, H.R.; Stefanski, L.A.; Carroll, R.J. Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *J. Am. Stat. Assoc.* **1989**, *84*, 460–466.
- 7. Morgenthaler, S. Least-absolute-deviations fits for generalized linear models. *Biometrika* 1992, 79, 747–754. [CrossRef]
- 8. Cantoni, E.; Ronchetti, E. Robust inference for generalized linear models. J. Am. Stat. Assoc. 2001, 96, 1022–1030. [CrossRef]
- Bianco, A.M.; Yohai, V.J. Robust estimation in the logistic regression model. In *Robust Statistics, Data Analysis, and Computer Intensive Methods*; Springer: New York, NY, USA, 1996; pp. 17–34.
- 10. Croux, C.; Haesbroeck, G. Implementing the Bianco and Yohai estimator for logistic regression. *Comput. Stat. Data Anal.* 2003, 44, 273–295. [CrossRef]
- 11. Bianco, A.M.; Boent, G.; Rodrigues, I.M. Robust tests in generalized linear models with missing responses. *Comput. Stat. Data Anal.* **2013**, *65*, 80–97. [CrossRef]
- 12. Valdora, M.; Yohai, V.J. Robust estimators for generalized linear models. J. Stat. Plan. Inference 2014, 146, 31–48. [CrossRef]
- 13. Ghosh, A.; Basu, A. Robust estimation in generalized linear models: the density power divergence approach. *Test* **2016** *25*, 269–290. [CrossRef]
- 14. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimising a density power divergence. *Biometrika* **1998**, *85*, 549–559. [CrossRef]
- 15. Basu, A.; Ghosh, A.; Mandal, A.; Martin, N.; Pardo, L. Robust Wald-type tests in GLM with random design based on minimum density power divergence estimators. *Stat. Method Appl.* **2021**, *3*, 933–1005. [CrossRef]
- 16. Broniatowski, M.; Toma, A.; Vajda, I. Decomposable pseudodistances and applications in statistical estimation. *J. Stat. Plan. Inference* **2012**, 142, 2574–2585. [CrossRef]
- 17. Castilla, E.; Martín, N.; Muñoz, S.; Pardo, L. Robust Wald-type tests based on Minimum Rényi Pseudodistance Estimators for the Multiple Regression Model. *J. Stat. Comput. Simul.* **2020**, *14*, 2592–2613. [CrossRef]
- Toma, A.; Leoni-Aubin, S. Optimal robust M-estimators using Rényi pseudodistances. J. Multivar. Anal. 2013, 115, 259-273. [CrossRef]
- 19. Toma, A., Karagrigoriou, A.; Trentou, P. Robust model selection criteria based on pseudodistances. *Entropy* **2020**, *22*, 304. [CrossRef] [PubMed]
- 20. Rényi, A. On measures of entropy and information. In *Proceeding of the 4th Symposium on Probability and Statistics*; University of California Press: Berkely, CA, USA, 1961; pp. 547–561.
- Jones, M.C.; Hjort, N.L.; Harris, I.R.; Basu, A. A comparison of related density-based minimum divergence estimators. *Biometrika* 2001. 88, 865–873. [CrossRef]
- 22. Fujisawa, H.; Eguchi, S. Robust parameter estimation with a small bias against heavy contamination. *J. Multivar. Anal.* 2008, 99, 2053-2081. [CrossRef]
- 23. Hirose, K.; Masuda, H. Robust relative error estimation. Entropy 2018, 20, 632. [CrossRef]
- 24. Kawashima, T.; Fujisawa, H. Robust and sparse regression via γ -divergence. *Entropy* **2017**, *19*, 608. [CrossRef]
- 25. Kawashima, T.; Fujisawa, H. Robust and sparse regression in generalized linear model by stochastic optimization. *Jpn. J. Stat. Data Sci.* **2019**, *2*, 465–489. [CrossRef]
- 26. Windham, M.P. Robustifying model fitting. J. R. Stat. Soc. Ser. B 1995, 57, 599–609. [CrossRef]
- 27. Castilla, E.; Jaenada, M.; Pardo, L. Estimation and testing on independent not identically distributed observations based on Rényi's pseudodistances. *arXiv* 2021, arXiv:2102.12282.
- 28. Pardo, L. Statistical Inference Based on Divergence Measures; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018.
- 29. Fraser, D.A.S. *Non parametric Methods in Statistics;* John Wiley & Sons: New York, NY, USA, 1957.
- 30. Maronna, R.A.; Martin, R.D.; Yohai, V.J. Robust Statistics Theory and Methods; John Wiley & Sons. Inc.: Hoboken, NJ, USA, 2006.
- 31. Donoho, D.L.; Huber, P.J. The notion of breakdown point. In *A Festschrift for Erich L. Lehmann*; CRC Press: Boca Raton, FL, USA, 1983.

- 32. Nelder, J.A.; Mead, R. A simplex method for function minimization. Comput. J. 1965, 7, 308–313. [CrossRef]
- 33. L.ô; S.N.; Ronchetti, E. Robust and accurate inference for generalized linear models. J. Multivar. Anal. 2009, 100, 2126–2136. [CrossRef]
- 34. Phelps, K. Use of the Complementary Log-Log Function to Describe Dose Response Relationships in Insecticide Evaluation Field Trials. In *Lecture Notes in Statistics, No. 14.: Proceedings of the International Conference on Generalized Linear Models*; Gilchrist, R., Ed.; Springer: Berlin, Germany, 1982.