

Estimation and testing on independent not identically distributed observations based on Rényi's pseudodistances

Elena Castilla¹, Maria Jaenada¹ and Leandro Pardo¹

¹Department of Statistics and O.R., Complutense University of Madrid, Spain

Abstract

In real life we often deal with independent but not identically distributed observations (i.n.i.d.o), for which the most well-known statistical model is the multiple linear regression model (MLRM) without random covariates. While the classical methods are based on the maximum likelihood estimator (MLE), it is well known its lack of robustness to small deviations from the assumed conditions. In this paper, and based on the Rényi's pseudodistance (RP), we introduce a new family of estimators in case our information about the unknown parameter is given for i.n.i.d.o.. This family of estimators, let say minimum RP estimators (as they are obtained by minimizing the RP between the assumed distribution and the empirical distribution of the data), contains the MLE as a particular case and can be applied, among others, to the MLRM without random covariates. Based on these estimators, we introduce Wald-type tests for testing simple and composite null hypotheses, as an extension of the classical MLE-based Wald test. Influence functions for the estimators and Wald-type tests are also obtained and analysed. Finally, a simulation study is developed in order to asses the performance of the proposed methods and some real-life data are analysed for illustrative purpose.

Keywords: asymptotic normality; consistency; independent not identically distributed observations; influence function; minimum Rényi's pseudodistance estimators, robustness; Wald-type tests.

1 Introduction

In parametric estimation the role of divergence measures is very intuitive: minimizing a suitable divergence measure between the data and the assumed model in order to estimate the unknown parameters. These estimators are called "minimum divergence estimators" (MDEs). There is a growing body of literature that recognizes the importance of MDEs because their robustness, without a significant loss of efficiency, in relation to the maximum likelihood estimator (MLE). See, for instance, Beran [1], Tamura and Boos [2], Simpson [3, 4], Lindsay [5], Pardo [6] and Basu et al. [7]. In the case of continuous models is convenient to consider families of divergence measures for which non-parametric

estimators of the unknown density function are not needed. From this perspective, the density power divergence (DPD) family, leading to the minimum DPD estimators, is the most important family of divergence measures. For more details see [7]. However, there is another important family of divergence measures which neither needs non-parametric estimators, the Rényi's pseudodistances (RP).

Let X_1, \dots, X_n be a random sample of size n from a population having true and unknown density function g , modelled by a parametric family of densities f_{θ} with $\theta \in \Theta \subset \mathbb{R}^p$. The RP between the densities f_{θ} and g is given, for $\alpha > 0$, by

$$R_{\alpha}(f_{\theta}, g) = \frac{1}{\alpha + 1} \log \left(\int f_{\theta}(x)^{\alpha+1} dx \right) + \frac{1}{\alpha(\alpha + 1)} \log \left(\int g(x)^{\alpha+1} dx \right) - \frac{1}{\alpha} \log \left(\int f_{\theta}(x)^{\alpha} g(x) dx \right). \quad (1)$$

The RP can be defined for $\alpha = 0$ taking continuous limits, yielding the expression

$$R_0(f_{\theta}, g) = \lim_{\alpha \downarrow 0} R_{\alpha}(f_{\theta}, g) = \int g(x) \log \frac{g(x)}{f_{\theta}(x)} dx,$$

i.e., the RP coincides with the Kullback-Leibler divergence between g and f_{θ} , $D_{Kullback}(g, f_{\theta})$, at $\alpha = 0$ (see [6]). The RP was considered for the first time in Jones et al. [8]; later Broniatowski et al. [9] established that the RP is positive for any two densities and for all values of the parameter α , $R_{\alpha}(f_{\theta}, g) \geq 0$ and further $R_{\alpha}(f_{\theta}, g) = 0$ if and only if $f_{\theta} = g$. This property embolden the definition of the minimum RP estimator as the minimizer of the RP between the assumed distribution and the empiric distribution of the data. Therefore, the minimum RP estimator based on the random sample X_1, \dots, X_n for the unknown parameter θ is given, for $\alpha > 0$, by

$$\hat{\theta}_{\alpha}^* = \arg \sup_{\theta \in \Theta} \sum_{i=1}^n \frac{f_{\theta}(X_i)^{\alpha}}{C_{\alpha}(\theta)}, \quad (2)$$

where

$$C_{\alpha}(\theta) = \left(\int f_{\theta}(x)^{\alpha+1} dx \right)^{\frac{\alpha}{\alpha+1}}.$$

Note that the value $\alpha = 0$ was defined as the KL divergence and hence, the minimum RP estimator coincides with the MLE at $\alpha = 0$. Besides, [9] studied the asymptotic properties and robustness of the minimum RP estimators and presented an application to the multiple linear regression model (MLRM) with random covariates. In the same vein, [10] introduced Wald-type tests based on the minimum RP estimators for the MLRM and [11] studied the minimum RP estimator for the linear regression model in the ultra-high dimensional set-up. Moreover, $\hat{\theta}_{\alpha}^*$ is a M-estimator and thus its asymptotic distribution and influence function (IF) can be obtained based on the asymptotic theory of the M-estimators.

However, far too little attention has been paid to the case of independent but non identically distributed observations (i.n.i.d.o.), for which the most well-known statistical model is the MLRM without random covariates. The nicest study for i.n.i.d.o. based on divergence measures, until now, is the paper of Ghosh and Basu [12] based on DPD measures. Some extensions are given in [13] and [14]. The main aim of this paper is to introduce and study the minimum RP estimator for i.n.i.d.o. We study their asymptotic

properties as well as we obtain its influence function in order to study its robustness. In Section 2 we introduce the minimum RP estimator for i.n.i.d.o. The consistency and asymptotic distribution is presented in Section 3. Section 4 is devoted to introduce and study Wald-type tests based on minimum RP estimators. The robustness of these estimators and Wald-type test is studied through its influence functions in Section 5. The special case of MLRM is considered in Section 6. Finally, an extensively study and two numerical examples of the MLRM are presented in Section 7 and 8, respectively.

2 The minimum RP estimator for independent but not identically distributed observations

Let Y_1, \dots, Y_n be i.n.i.d.o. being g_1, \dots, g_n the corresponding density functions with respect to some common dominating measure. We are interested in modeling g_i by the density function $f_i(y, \boldsymbol{\theta})$, $i = 1, \dots, n$, being $\boldsymbol{\theta}$ common for all the density functions $f_i(y, \boldsymbol{\theta})$. For each observation i , the ‘‘Rényi’s pseudodistance’’ between $f_i(y, \boldsymbol{\theta})$ and \widehat{g}_i , can be defined for positives values of α as

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

where

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

does not depend on $\boldsymbol{\theta}$. As we only have one observation of Y_i the best way to estimate \widehat{g}_i is assuming that the distribution is degenerate in y_i . Therefore, (3) yields to the loss

$$R_\alpha(f_i(y, \boldsymbol{\theta}), \widehat{g}_i) = \frac{1}{\alpha + 1} \log \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \right) - \frac{1}{\alpha} \log f_i(Y_i, \boldsymbol{\theta})^\alpha + k. \quad (4)$$

At $\alpha = 0$ the RP loss is given by

$$R_0(f_i(y, \boldsymbol{\theta}), \widehat{g}_i) = \lim_{\alpha \downarrow 0} R_\alpha(f_i(y, \boldsymbol{\theta}), \widehat{g}_i) = -\log f_i(Y_i, \boldsymbol{\theta}) + k. \quad (5)$$

Now, Expression (4) can be written as

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

and thus minimizing $R_\alpha(f_i(y, \boldsymbol{\theta}), \widehat{g}_i)$ in $\boldsymbol{\theta}$, for $\alpha > 0$, is equivalent to maximizing

$$V_i(Y_i, \boldsymbol{\theta}) = \frac{f_i(Y_i, \boldsymbol{\theta})^\alpha}{\left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \right)^{\frac{\alpha}{\alpha+1}}}.$$

In the following, we shall denote,

$$L_\alpha^i(\boldsymbol{\theta}) = \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \right)^{\frac{\alpha}{\alpha+1}}.$$

Based on this idea we are going to consider the objective function

$$H_n^\alpha(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \frac{f_i(Y_i, \boldsymbol{\theta})^\alpha}{\left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy\right)^{\frac{\alpha}{\alpha+1}}} = \frac{1}{n} \sum_{i=1}^n \frac{f_i(Y_i, \boldsymbol{\theta})^\alpha}{L_\alpha^i(\boldsymbol{\theta})} = \frac{1}{n} \sum_{i=1}^n V_i(Y_i, \boldsymbol{\theta}) \quad (6)$$

and then the minimum RP estimator, $\widehat{\boldsymbol{\theta}}_\alpha$, for the common $\boldsymbol{\theta}$, is given by

$$\widehat{\boldsymbol{\theta}}_\alpha = \arg \max_{\boldsymbol{\theta} \in \Theta} H_n^\alpha(\boldsymbol{\theta}), \quad (7)$$

with $H_n^\alpha(\boldsymbol{\theta})$ defined in (6) for $\alpha > 0$ and $H_n^0(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \log f_i(Y_i, \boldsymbol{\theta})$.

It is interesting to observe that when Y_1, \dots, Y_n are independent and identically distributed random variables, the estimator $\widehat{\boldsymbol{\theta}}_\alpha$ coincides with the estimator $\widehat{\boldsymbol{\theta}}_\alpha^*$ given in (2). In the following section we shall establish the consistency of $\widehat{\boldsymbol{\theta}}_\alpha$, as well as its asymptotic distribution.

3 Consistency and asymptotic distribution

We shall assume in the following that the true densities g_i $i = 1, \dots, n$, belong to the assumed model, i.e., $g_i \equiv f_i(y, \boldsymbol{\theta})$, $i = 1, \dots, n$, for some common parameter $\boldsymbol{\theta}$. We denote by $\boldsymbol{\theta}^*$ the true value of the unknown parameter. In the following we shall denote $E_{\boldsymbol{\theta}^*}[Y] = \int y f_i(y, \boldsymbol{\theta}^*) dy$, and we introduce the matrices

$$\boldsymbol{\Psi}_n = \frac{1}{n} \sum_{i=1}^n \boldsymbol{J}^{(i)}, \quad (8)$$

with

$$\boldsymbol{J}^{(i)} = \left(-E_{\boldsymbol{\theta}^*} \left[\frac{\partial^2 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right] \right)_{i,j=1, \dots, p}$$

and

$$\boldsymbol{\Omega}_n = \frac{1}{n} \sum_{i=1}^n \text{Var}_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial V_i(Y; \boldsymbol{\theta})}{\partial \theta_j} \right)_{j=1, \dots, p} \right]. \quad (9)$$

In order to get the asymptotic results we shall assume the following conditions:

- C1.** The support, \mathcal{X} , of the density functions $f_i(y, \boldsymbol{\theta})$ is the same for all i and does not depend on $\boldsymbol{\theta}$.
- C2.** There exists an open subset Θ^* of Θ containing the true value of the parameter $\boldsymbol{\theta}^*$ such that for almost all $y \in \mathcal{X}$ the density $f_i(y, \boldsymbol{\theta})$ admits all third derivatives with respect to $\boldsymbol{\theta}$ and $i = 1, \dots, n$.
- C3.** For $i = 1, 2, \dots$ the integrals

$$\int f_i(y, \boldsymbol{\theta})^{1+\alpha} dy$$

can be differentiated thrice with respect to $\boldsymbol{\theta}$ and we can interchange integration and differentiation.

C4. For $i = 1, 2, \dots$ the matrices $\mathbf{J}^{(i)}$ are positive definite. We denote by λ_n the minimum eigenvalue of $\mathbf{\Omega}_n$ and

$$\lambda_0 = \inf_n \lambda_n > 0.$$

C5. There exists functions $M_{jkl}^{(i)}$ such that

$$\left| \frac{\partial^3 V_i(y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| \leq M_{jkl}^{(i)}(y), \quad \forall \boldsymbol{\theta} \in \Theta^*, \forall j, k, l$$

and

$$E_{\boldsymbol{\theta}^*} \left[M_{jkl}^{(i)}(Y) \right] = m_{jkl} < \infty, \quad \forall j, k, l.$$

C6. For all j, k, l the sequences $\left\{ \frac{\partial V_i(Y; \boldsymbol{\theta})}{\partial \theta_j} \right\}_{j=1, \dots, p}$, $\left\{ \frac{\partial^2 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right\}_{j, k=1, \dots, p}$ and $\left\{ \frac{\partial^3 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right\}_{j, k, l=1, \dots, p}$ are uniformly integrable in the Cesàro sense, i.e.

$$\begin{aligned} \lim_{N \rightarrow \infty} \left(\sup_{n > 1} \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left| \frac{\partial V_i(Y; \boldsymbol{\theta})}{\partial \theta_j} \right| I_{\left\{ \frac{\partial V_i(Y; \boldsymbol{\theta})}{\partial \theta_j} > N \right\}}(Y) \right] \right) &= 0, \\ \lim_{N \rightarrow \infty} \left(\sup_{n > 1} \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left| \frac{\partial^2 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right| I_{\left\{ \frac{\partial^2 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} > N \right\}}(Y) \right] \right) &= 0, \\ \lim_{N \rightarrow \infty} \left(\sup_{n > 1} \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left| \frac{\partial^3 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| I_{\left\{ \frac{\partial^3 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} > N \right\}}(Y) \right] \right) &= 0. \end{aligned}$$

C7. For all $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left\| \mathbf{\Omega}_n^{-\frac{1}{2}} \frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|_2^2 I_{\left\{ \left\| \mathbf{\Omega}_n^{-\frac{1}{2}} \frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|_2 > \varepsilon \sqrt{n} \right\}} \right] \right\} = 0.$$

Note that **C6.** gives sufficient conditions for the weak law of large numbers with i.n.i.d.o. ([15]), while **C7.** is the assumption required for the multivariate central limit theorem for i.n.i.d.o. ([16]). The following theorem states the consistency of $\widehat{\boldsymbol{\theta}}_\alpha$ and the second one establishes its asymptotic distribution. Proof of both theorems are given in Appendix A.1 and Appendix A.2, respectively.

Theorem 3.1 *Let Y_1, \dots, Y_n be i.n.i.d.o. each with a density function $f_i(y, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. If conditions **C1.–C6.** holds, there exists a consistent sequence $\widehat{\boldsymbol{\theta}}_n$ of the system of equations*

$$\frac{\partial H_n^\alpha(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}_p. \quad (10)$$

Theorem 3.2 *Let Y_1, \dots, Y_n be i.n.i.d.o. each with a density function $f_i(y, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. If conditions **C1.–C7.** are satisfied the asymptotic distribution of the minimum RP estimator is given by*

$$\sqrt{n} \mathbf{\Omega}_n^{-\frac{1}{2}} \boldsymbol{\Psi}_n \left(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^* \right) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \mathbf{I}_p), \quad (11)$$

being \mathbf{I}_p the p -dimensional identity matrix and the matrices $\boldsymbol{\Psi}_n$ and $\mathbf{\Omega}_n$ were defined in (8) and (9), respectively.

4 Wald-type test for i.n.i.d.o.

4.1 Wald-type tests for simple null hypotheses

We define a family of Wald-type test statistics based on minimum RP estimators for testing the hypothesis

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}^0 \quad \text{against} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}^0, \quad (12)$$

for a given $\boldsymbol{\theta}^0 \in \Theta$.

Definition 4.1 Let $\widehat{\boldsymbol{\theta}}_\alpha$ be the minimum RP estimator of $\boldsymbol{\theta}$. The family of proposed Wald-type test statistics for testing the null hypothesis (12) is given by

$$\mathbf{W}_n^0(\boldsymbol{\theta}^0) = n(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^0)^T \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0)(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^0), \quad (13)$$

where $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^0) = \lim_{n \rightarrow \infty} \boldsymbol{\Psi}_n(\boldsymbol{\theta}^0) \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta}^0) \boldsymbol{\Psi}_n(\boldsymbol{\theta}^0)$.

Theorem 4.2 The asymptotic distribution of the Wald-type test statistics $\mathbf{W}_n^0(\boldsymbol{\theta}^0)$, defined in (13), under the null hypothesis (12), is a chi-square distribution with p degrees of freedom.

Based on Theorem 4.2, we shall reject the null hypothesis in (12) if

$$\mathbf{W}_n^0(\boldsymbol{\theta}^0) > \chi_{p,\nu}^2, \quad (14)$$

being $\chi_{p,\nu}^2$ the upper ν -th quantile of χ_p^2 .

Theorem 4.3 Let $\boldsymbol{\theta}^*$ be the true value of $\boldsymbol{\theta}$, with $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}^0$, and let us denote

$$\ell(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\theta}^0)^T \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0)(\boldsymbol{\theta} - \boldsymbol{\theta}^0)$$

and $\sigma_{\mathbf{W}_n^0(\boldsymbol{\theta}^0)}^2(\boldsymbol{\theta}^*) = 4(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0)^T [\boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0) \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*) \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0)] (\boldsymbol{\theta}^* - \boldsymbol{\theta}^0)$. Then,

$$\sqrt{n} \left(\ell(\widehat{\boldsymbol{\theta}}_\alpha) - \ell(\boldsymbol{\theta}^*) \right) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}, \sigma_{\mathbf{W}_n^0(\boldsymbol{\theta}^0)}^2(\boldsymbol{\theta}^*)).$$

Corollary 4.4 Theorem 4.3, makes it possible to have an approximation of the power function for the test given in (14). This is given by

$$\begin{aligned} \pi_{\mathbf{W}_n^0(\boldsymbol{\theta}^0)}^\alpha(\boldsymbol{\theta}^*) &= P(\text{Rejecting } H_0 | \boldsymbol{\theta} = \boldsymbol{\theta}^*) = P(\mathbf{W}_n^0(\boldsymbol{\theta}^0) > \chi_{p,\nu}^2 | \boldsymbol{\theta} = \boldsymbol{\theta}^*) \\ &= P \left(\ell(\widehat{\boldsymbol{\theta}}_\alpha) - \ell(\boldsymbol{\theta}^*) > \frac{\chi_{p,\nu}^2}{n} - \ell(\boldsymbol{\theta}^*) \right) \\ &= 1 - \Phi_n \left(\frac{\sqrt{n}}{\sigma_{\mathbf{W}_n^0(\boldsymbol{\theta}^0)}(\boldsymbol{\theta}^*)} \left(\frac{\chi_{p,\nu}^2}{n} - \ell(\boldsymbol{\theta}^*) \right) \right), \end{aligned}$$

where $\Phi_n(\cdot)$ is a sequence of distribution functions which tends uniformly to the standard normal distribution function $\Phi(\cdot)$. We can observe that

$$\lim_{n \rightarrow \infty} \pi_{\mathbf{W}_n^0(\boldsymbol{\theta}^0)}^\alpha(\boldsymbol{\theta}^*) = 1 \quad \forall \alpha \geq 0,$$

so the Wald-type tests are consistent in the sense of Fraser.

From here, it can be deduced that the necessary sample size for the Wald-type tests to have a predetermined power, $\pi_{\mathbf{W}_n^0(\boldsymbol{\theta}^0)}^\alpha(\boldsymbol{\theta}^*) \approx \pi^*$, is given by

$$n = \left\lceil \frac{A + B + \sqrt{A(A + 2B)}}{2\ell^2(\boldsymbol{\theta}^*)} \right\rceil,$$

where $A = \sigma_{\mathbf{W}_n^0(\boldsymbol{\theta}^0)}^2(\boldsymbol{\theta}^*)(\Phi^{-1}(1 - \pi^*))^2$, and $B = 2\ell(\boldsymbol{\theta}^*)\chi_{p,\nu}^2$.

4.2 Wald-type tests for composite null hypotheses

We may be also interested on testing a set of $r < p$ redundant restrictions on the parameter vector $\boldsymbol{\theta}$. In this context, we are interested in testing

$$H_0 : \mathbf{M}^T \boldsymbol{\theta} = \mathbf{m} \quad \text{against} \quad H_1 : \mathbf{M}^T \boldsymbol{\theta} \neq \mathbf{m}, \quad (15)$$

where \mathbf{M} is $p \times r$ full rank matrix with $r < p$ and \mathbf{m} is a r -vector.

Definition 4.5 Let $\widehat{\boldsymbol{\theta}}_\alpha$ be the minimum RP estimator of $\boldsymbol{\theta}$. The family of proposed Wald-type test statistics for testing the null hypothesis (15) is given by

$$\mathbf{W}_n(\widehat{\boldsymbol{\theta}}_\alpha) = n(\mathbf{M}^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m})^T \left[\mathbf{M}^T \boldsymbol{\Sigma}_\alpha(\widehat{\boldsymbol{\theta}}_\alpha) \mathbf{M} \right]^{-1} (\mathbf{M}^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m}). \quad (16)$$

Theorem 4.6 The asymptotic distribution of the Wald-type test statistics $\mathbf{W}_n(\widehat{\boldsymbol{\theta}}_\alpha)$, defined in (16), under the null hypothesis (15), is a chi-square distribution with r degrees of freedom.

Based on Theorem 4.6, we shall reject the null hypothesis in (15) if

$$\mathbf{W}_n(\widehat{\boldsymbol{\theta}}_\alpha) > \chi_{r,\nu}^2. \quad (17)$$

We could generalize our results to a more general restricted space $\Theta_0 \subset \Theta$ defined by a set of $r < p$ non-redundant restrictions of the form $h(\boldsymbol{\theta}) = \mathbf{0}$, by substituting the matrix $\mathbf{M}^T \boldsymbol{\Sigma}_\alpha(\widehat{\boldsymbol{\theta}}_\alpha) \mathbf{M}$ by $\mathbf{H}^T \boldsymbol{\Sigma}_\alpha(\widehat{\boldsymbol{\theta}}_\alpha) \mathbf{H}$ with $\mathbf{H} = \frac{\partial h(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$ in (16). The asymptotic distribution stated in Theorem 4.6 still holds.

4.3 Contiguous alternatives hypothesis

The previous results provides an asymptotic approximation to the power function of the proposed Wald-type tests. We now consider the particular set of contiguous alternatives hypothesis of the form

$$H_1 : \boldsymbol{\theta}_n = \boldsymbol{\theta}^0 + n^{-1/2} \mathbf{d}, \quad (18)$$

where \mathbf{d} is a fixed vector in \mathbb{R}^p such that $\mathbf{d} \in \Theta$ and $\boldsymbol{\theta}^0$ is an element of Θ_0 . Note that the alternative hypothesis move towards $\boldsymbol{\theta}^0$ and it get closer when the sample size n increases.

Theorem 4.7 Under the contiguous alternative hypotheses given in (18), the asymptotic distribution of the Wald-type test statistics, $\mathbf{W}_n(\widehat{\boldsymbol{\theta}}_\alpha)$, defined in (15) is a non-central chi-square distribution with r degrees of freedom and non-centrality parameter

$$\delta = \mathbf{d}^T \mathbf{M} [\mathbf{M}^T \boldsymbol{\Sigma}_\alpha(\widehat{\boldsymbol{\theta}}_\alpha) \mathbf{M}]^{-1} \mathbf{M}^T \mathbf{d}.$$

5 Influence function analysis

In this section we shall obtain the influence function (IF) of the minimum RP functional for the non-homogeneous case. We shall denote by G_i the true distribution function associated to the observation Y_i whose density function is denoted by g_i and by $\mathbf{T}_\alpha(G_1, \dots, G_n)$ the minimum RP functional defined as the minimizer of

$$\sum_{i=1}^n H_n^{\alpha, (i)} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{1+\alpha} \log \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \right) - \frac{1}{\alpha} \log \left(\int f_i(y, \boldsymbol{\theta})^\alpha dG_i(y) \right) \right\} \quad (19)$$

or, fixed a value of α , under appropriate differentiability conditions, as the solution of the system of equations obtained after differentiating (19) and equalling to zero

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\int f_i(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_i(y, \boldsymbol{\theta}) g_i(y) dy}{\int f_i(y, \boldsymbol{\theta})^{\alpha+1} g_i(y) dy} - \frac{\int f_i(y, \boldsymbol{\theta})^\alpha \mathbf{u}_i(y, \boldsymbol{\theta}) g_i(y) dy}{\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) dy} \right\} = \mathbf{0}_p. \quad (20)$$

By $\mathbf{u}_i(y, \boldsymbol{\theta})$ we are denoting

$$\mathbf{u}_i(y, \boldsymbol{\theta}) = \frac{\partial \log(f_i(y, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}}$$

and by

$$g_{i,\varepsilon} = (1 - \varepsilon)g_i + \varepsilon\Delta_{t_i}, \quad i = 1, \dots, n$$

the contaminated density where Δ_{t_i} is the degenerated distribution at point t_i .

Let $\boldsymbol{\theta} = \mathbf{T}_\alpha(G_1, \dots, G_n)$ and we denote by

$$\boldsymbol{\theta}_\varepsilon^{i_0} = \mathbf{T}_\alpha(G_1, \dots, G_{i_0-1}, G_{i_0,\varepsilon}, G_{i_0+1}, \dots, G_n)$$

the minimum Renyi pseudodistance functional with contamination only in the i_0 -th direction, where $G_{i_0,\varepsilon}$ is the distribution function associated to the density function

$$g_{i_0,\varepsilon}(y) = (1 - \varepsilon)f_{i_0}(y, \boldsymbol{\theta}) + \varepsilon\Delta_{i_0}(y)$$

and

$$g_i(y) = \begin{cases} f_i(y, \boldsymbol{\theta}) & \text{if } i \neq i_0; \\ (1 - \varepsilon)f_i(y, \boldsymbol{\theta}) + \varepsilon\Delta_{i_0} & \text{if } i = i_0. \end{cases}$$

It is also possible to contaminate in all the directions and in this case we shall denote by

$$\boldsymbol{\theta}_\varepsilon = \mathbf{T}_\alpha(G_{1,\varepsilon}, \dots, G_{n,\varepsilon})$$

the minimum RP functional with contamination in all directions.

In the following theorem we present the expressions of the IF. See Appendix A.7 for details.

Theorem 5.1 *The influence function in the i_0 -th direction is given by*

$$IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n) = (\mathbf{M}_{n,\alpha}(\boldsymbol{\theta}))^{-1} \frac{-\ell_{i_0,\alpha}(\boldsymbol{\theta})}{\left(\int f_{i_0}(y, \boldsymbol{\theta})^\alpha g_{i_0}(y) dy\right)^2},$$

where

$$\ell_{i_0,\alpha}(\boldsymbol{\theta}) = f_{i_0}(y, \boldsymbol{\theta}) \int f_{i_0}(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_{i_0}(y, \boldsymbol{\theta}) dy - f_{i_0}(y, \boldsymbol{\theta}) \mathbf{u}_{i_0}(y, \boldsymbol{\theta}) \int f_{i_0}(y, \boldsymbol{\theta})^{\alpha+1} dy,$$

$$\mathbf{M}_{n,\alpha}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \left[\frac{\mathbf{A}_{i,\alpha}(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta})^{1+\alpha} dy\right)^2} - \frac{\mathbf{A}_{i,\alpha}^*(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) dy\right)^2} \right], \quad (21)$$

and

$$\begin{aligned} \mathbf{A}_{i,\alpha}(\boldsymbol{\theta}) &= \left[(1+\alpha) \int f_i(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_i^T(y, \boldsymbol{\theta}) \mathbf{u}_i(y, \boldsymbol{\theta}) dy + \int f_i(y, \boldsymbol{\theta})^{\alpha+1} \frac{\partial \mathbf{u}_i(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dy \right] \int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \\ &\quad - (1+\alpha) \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right) \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right)^T, \\ \mathbf{A}_{i,\alpha}^*(\boldsymbol{\theta}) &= \left[\alpha \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i^T(y, \boldsymbol{\theta}) \mathbf{u}_i(y, \boldsymbol{\theta}) dy + \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \frac{\partial \mathbf{u}_i(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dy \right] \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) dy \\ &\quad - \alpha \left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right) \left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right)^T. \end{aligned}$$

Similarly, the influence function in all directions is given by

$$IF(t_1, \dots, t_n, \mathbf{T}_\alpha, G_1, \dots, G_n) = (\mathbf{M}_{n,\alpha}(\boldsymbol{\theta}))^{-1} \sum_{i=1}^n \frac{-\ell_{i,\alpha}(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) dy\right)^2}.$$

Remark 5.2 *When the true distribution g_i belongs to the model so that $g_i(y) = f_i(y, \boldsymbol{\theta})$ for $i = 1, \dots, n$, then $\mathbf{M}_{n,\alpha}(\boldsymbol{\theta})$ given in (21) coincides with $\boldsymbol{\Psi}_n(\boldsymbol{\theta})$ given in (8) and*

$$IF(t_{i_0}, \mathbf{T}_\alpha, F_{1,\boldsymbol{\theta}}, \dots, F_{n,\boldsymbol{\theta}}) = \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\theta}) \mathbf{D}_{i_0,\alpha}(\boldsymbol{\theta}),$$

$$IF(t_1, \dots, t_n, \mathbf{T}_\alpha, F_{1,\boldsymbol{\theta}}, \dots, F_{n,\boldsymbol{\theta}}) = \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\theta}) \sum_{i=1}^n \mathbf{D}_{i,\alpha}(\boldsymbol{\theta}),$$

$$\text{with } \mathbf{D}_{i,\alpha}(\boldsymbol{\theta}) = \frac{-\ell_{i,\alpha}(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy\right)^2}.$$

Remark 5.3 *In particular, letting $t_i = t$, $G_i = G$, $f_i(y, \boldsymbol{\theta}) = f(y, \boldsymbol{\theta})$ for $i = 1, \dots, n$ (this situation corresponds to the case of independent and identically distributed, i.i.d., random variables) and $g(y) = f(y, \boldsymbol{\theta})$, we have*

$$IF(t, \mathbf{T}_\alpha, G) = (\mathbf{M}_\alpha(\boldsymbol{\theta}))^{-1} [f(y, \boldsymbol{\theta})^\alpha \mathbf{u}(y, \boldsymbol{\theta}) - \mathbf{c}_\alpha(\boldsymbol{\theta}) f(y, \boldsymbol{\theta})^\alpha],$$

where

$$\mathbf{c}_\alpha(\boldsymbol{\theta}) = \frac{\int f(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}(y, \boldsymbol{\theta}) dy}{\int f(y, \boldsymbol{\theta})^{\alpha+1} dy}$$

and

$$\begin{aligned} \mathbf{M}_\alpha(\boldsymbol{\theta}) &= \frac{1}{\int f(y, \boldsymbol{\theta})^{\alpha+1} dy} \left[\int f(y, \boldsymbol{\theta})^{\alpha+1} dy \int f(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}(y, \boldsymbol{\theta}) \mathbf{u}^T(y, \boldsymbol{\theta}) dy \right. \\ &\quad \left. - \left(\int f(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}(y, \boldsymbol{\theta}) dy \right) \left(\int f(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}(y, \boldsymbol{\theta}) dy \right)^T \right], \end{aligned}$$

as in [9].

5.1 Influence function of the Wald-type test statistics

Once we have computed the IF for the minimum RP estimators, we can define and study the IF for the Wald-type test statistics. First, we define the associated statistical functional, evaluated at G_1, \dots, G_n as

$$W_\alpha^0(G_1, \dots, G_n) = (\mathbf{T}_\alpha(G_1, \dots, G_n) - \boldsymbol{\theta}^0)^T \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0) (\mathbf{T}_\alpha(G_1, \dots, G_n) - \boldsymbol{\theta}^0), \quad (22)$$

corresponding to (12) for the simple null hypothesis. Let us consider first the contamination only in one direction, say i_0 -th direction. The corresponding IF is then defined as

$$\begin{aligned} IF(t_{i_0}, W_\alpha^0, G_1, \dots, G_n) &= \frac{\partial}{\partial \varepsilon} W_\alpha^0(G_1, \dots, G_{i_0-1}, G_{i_0, \varepsilon}, G_{i_0+1}, \dots, G_n) \Big|_{\varepsilon=0} \\ &= 2\mathbf{T}_\alpha(G_1, \dots, G_n) - \boldsymbol{\theta}^0)^T \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0) IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n). \end{aligned} \quad (23)$$

However, if we evaluate (23) at the null distribution $G_i = F_{i, \boldsymbol{\theta}^0}$, it becomes identically zero. Therefore, it becomes necessary to consider the second order IF of the proposed Wald-type test functional

$$\begin{aligned} IF^{(2)}(t_{i_0}, W_\alpha^0, G_1, \dots, G_n) &= \frac{\partial^2}{\partial \varepsilon^2} W_\alpha^0(G_1, \dots, G_{i_0-1}, G_{i_0, \varepsilon}, G_{i_0+1}, \dots, G_n) \Big|_{\varepsilon=0} \\ &= 2IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n) \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0) IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n). \end{aligned}$$

Similarly, we can consider contamination in all directions, obtaining that the second order influence function of the proposed Wald-type tests functional for testing simple null hypothesis is given by

$$\begin{aligned} IF^{(2)}(t_1, \dots, t_n, W_\alpha^0, G_1, \dots, G_n) &= \frac{\partial^2}{\partial \varepsilon^2} W_\alpha^0(G_1, \dots, G_{i_0-1}, G_{i_0, \varepsilon}, G_{i_0+1}, \dots, G_n) \Big|_{\varepsilon=0} \\ &= 2IF(t_1, \dots, t_n, \mathbf{T}_\alpha, G_1, \dots, G_n) \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0) IF(t_1, \dots, t_n, \mathbf{T}_\alpha, G_1, \dots, G_n). \end{aligned}$$

Remark 5.4 When the true distribution belongs to the model, then the second order influence functions of the proposed Wald-type tests functional for testing simple null hypothesis in (12) are given by

$$\begin{aligned} IF^{(2)}(t_{i_0}, W_\alpha, F_{1, \boldsymbol{\theta}^0}, \dots, F_{n, \boldsymbol{\theta}^0}) &= 2 [\mathbf{D}_{i_0, \alpha}(\boldsymbol{\theta}^0)]^T [\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\theta}^0) \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0) \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\theta}^0)] [\mathbf{D}_{i_0, \alpha}(\boldsymbol{\theta}^0)], \\ IF^{(2)}(t_1, \dots, t_n, W_\alpha, F_{1, \boldsymbol{\theta}^0}, \dots, F_{n, \boldsymbol{\theta}^0}) &= 2 \left[\sum_{i=1}^n \mathbf{D}_{i, \alpha}(\boldsymbol{\theta}^0) \right]^T [\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\theta}^0) \boldsymbol{\Sigma}_\alpha^{-1}(\boldsymbol{\theta}^0) \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\theta}^0)] \left[\sum_{i=1}^n \mathbf{D}_{i, \alpha}(\boldsymbol{\theta}^0) \right]. \end{aligned}$$

Remark 5.5 In a similar manner, when the true distribution belongs to the model, the second order influence functions of the proposed Wald-type tests functionals for testing

composite null hypothesis in (15) are given by

$$\begin{aligned}
& IF^{(2)}(t_{i_0}, W_\alpha^0, F_{1, \theta^0}, \dots, F_{n, \theta^0}) \\
& \quad = 2 [\Psi_n^{-1}(\theta^0) \mathbf{D}_{i_0, \alpha}(\theta^0)]^T \mathbf{M} [\mathbf{M}^T \Sigma_\alpha(\theta^0) \mathbf{M}]^{-1} \mathbf{M}^T [\Psi_n^{-1}(\theta^0) \mathbf{D}_{i_0, \alpha}(\theta^0)], \\
& IF^{(2)}(t_1, \dots, t_n, W_\alpha^0, F_{1, \theta^0}, \dots, F_{n, \theta^0}) \\
& \quad = 2 \left[\Psi_n^{-1}(\theta^0) \sum_{i=1}^n \mathbf{D}_{i_0, \alpha}(\theta^0) \right]^T \mathbf{M} [\mathbf{M}^T \Sigma_\alpha(\theta^0) \mathbf{M}]^{-1} \mathbf{M}^T \left[\Psi_n^{-1}(\theta^0) \sum_{i=1}^n \mathbf{D}_{i_0, \alpha}(\theta^0) \right].
\end{aligned}$$

6 Multiple linear regression model

Consider the MLRM

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (24)$$

where the errors ε_i 's are i.i.d. normal random variables with mean zero and variance σ^2 , $\mathbf{X}_i^T = (X_{i1}, \dots, X_{ip})$ is the vector of independent variables corresponding to the i -th condition and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the vector of regression coefficients to be estimated. We will consider that, for each i , \mathbf{X}_i is fixed, yielding to independent but not identically distributed Y_i 's (i.n.i.d.o.), with $Y_i \sim \mathcal{N}(\mathbf{X}_i^T \boldsymbol{\beta}, \sigma^2)$. Under the previous notation, with $f_i \equiv \mathcal{N}(\mathbf{X}_i^T \boldsymbol{\beta}, \sigma^2)$, we have, for $\alpha > 0$,

$$\begin{aligned}
V_i(Y_i; \boldsymbol{\theta}, \mathbf{X}_i) &= \frac{1}{(2\pi)^{\alpha/2} \sigma^\alpha} \exp\left(\frac{-\alpha(Y_i - \mathbf{X}_i^T \boldsymbol{\beta})^2}{2\sigma^2}\right) \\
&\quad \frac{1}{((2\pi)^{\alpha/2} \sigma^\alpha \sqrt{1+\alpha})^{-\frac{\alpha}{\alpha+1}}} \\
&= \left(\frac{1+\alpha}{2\pi}\right)^{\frac{\alpha}{2(\alpha+1)}} \sigma^{-\frac{\alpha}{\alpha+1}} \exp\left(-\frac{\alpha}{2} \left(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right).
\end{aligned}$$

Thus, our objective function to be minimized becomes

$$\frac{1}{n} \sum_{i=1}^n V_i(Y_i; \boldsymbol{\theta}, \mathbf{X}_i) = \left(\frac{1+\alpha}{2\pi}\right)^{\frac{\alpha}{2(\alpha+1)}} \frac{1}{n} \sum_{i=1}^n \sigma^{-\frac{\alpha}{\alpha+1}} \exp\left(-\frac{\alpha}{2} \left(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right).$$

Taking into account that the term $\left(\frac{1+\alpha}{2\pi}\right)^{\frac{\alpha}{2(\alpha+1)}}$ does not depend on the model parameters, we have that, for $\alpha > 0$

$$(\hat{\boldsymbol{\beta}}_\alpha, \hat{\sigma}_\alpha) = \arg \max_{\boldsymbol{\beta}, \sigma} \sum_{i=1}^n \sigma^{-\frac{\alpha}{\alpha+1}} \exp\left(-\frac{\alpha}{2} \left(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right). \quad (25)$$

Derivating with respect to $\boldsymbol{\beta}$ and σ we see that the estimators $\hat{\boldsymbol{\beta}}_\alpha$ and $\hat{\sigma}_\alpha$ are solutions of the system

$$\begin{aligned}
& \sum_{i=1}^n \exp\left(-\frac{\alpha}{2} \left(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right) \left(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}}{\sigma}\right) \mathbf{X}_i = \mathbf{0}_p \\
& \sum_{i=1}^n \exp\left(-\frac{\alpha}{2} \left(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right) \left\{ \left(\frac{Y_i - \mathbf{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2 - \frac{1}{1+\alpha} \right\} = 0,
\end{aligned} \quad (26)$$

which is exactly the same as the one suggested by Castilla et al. (2020) for the case of homogeneous data. If $\alpha = 0$, we have

$$\left(\widehat{\boldsymbol{\beta}}_{\alpha=0}, \widehat{\sigma}_{\alpha=0}\right) = \arg \max_{\boldsymbol{\beta}, \sigma} \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|\mathbf{Y} - \mathbb{X}\boldsymbol{\beta}\|_2^2}{2\sigma^2}\right) \quad (27)$$

and we get the system necessary to get the MLE of $\boldsymbol{\beta}$ and σ , whose well-known solution is given by

$$\widehat{\boldsymbol{\beta}}_0 = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \mathbf{Y} \quad \text{and} \quad \widehat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \mathbf{X}_i^T \widehat{\boldsymbol{\beta}}_0\right)^2,$$

where $\mathbb{X}^T = (\mathbf{X}_1, \dots, \mathbf{X}_n)_{p \times n}$ is the matrix of explanatory variables.

Lemma 6.1 *Consider the set-up of the MLRM with i.n.i.d.o. defined in (24) and assume that the true data generating density belongs to the model family. If the following mild conditions about the explanatory variables hold*

M1. *The values of the explanatory variables are such that, for all j, k and l*

$$\sup_{n>1} \max_{1 \leq i \leq n} |X_{ij}| = O(1), \quad \text{and} \quad \sup_{n>1} \max_{1 \leq i \leq n} |X_{ij} X_{ik}| = O(1)$$

and

$$\frac{1}{n} \sum_{i=1}^n |x_{ij} x_{ik} x_{il}| = O(1).$$

M2. *The matrix \mathbb{X}^T satisfies*

$$\inf_n \left[\min \text{ eigenvalue of } \frac{1}{n} \mathbb{X}^T \mathbb{X} \right] > 0,$$

$$n \times \max_{1 \leq i \leq n} \left[\mathbf{X}_i^T (\mathbb{X}^T \mathbb{X})^{-1} \mathbf{X}_i \right] = O(1),$$

then C1.–C7. are satisfied.

On the other hand, after some heavy computations we follow that expressions (8) and (9) are given by

$$\boldsymbol{\Psi}_n = \begin{bmatrix} \frac{-1}{\sigma^{2(\alpha+1)^{3/2}}} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T\right) & 0 \\ 0 & \frac{-2}{\sigma^{2(\alpha+1)^{5/2}}} \end{bmatrix}, \quad (28)$$

and

$$\boldsymbol{\Omega}_n = \begin{bmatrix} \frac{\left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T\right)}{(2\alpha+1)^{3/2}} & 0 \\ 0 & \frac{(3\alpha^2+4\alpha+2)}{\sigma^{2(\alpha+1)^2(2\alpha+1)^{5/2}}} \end{bmatrix}. \quad (29)$$

Theorem 6.2 *Consider the set-up of the MLRM with i.n.i.d.o. defined in (24) and assume that the true data generating density belongs to the model family and the observed explanatory variables satisfy conditions M1. and M2.. Then,*

1. *There exists a consistent sequence as $\widehat{\boldsymbol{\theta}}_\alpha = (\widehat{\boldsymbol{\beta}}_\alpha, \widehat{\sigma}_\alpha)$ of solutions to the minimum Rényi estimating equations (26).*

2. $\widehat{\beta}_\alpha$ and $\widehat{\sigma}_\alpha$ are asymptotically independent and their asymptotic distribution is given by

$$\sqrt{n} \left(\widehat{\theta}_\alpha - \theta^* \right) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}, \Sigma_\alpha),$$

with

$$\Sigma_\alpha = \lim_{n \rightarrow \infty} \Sigma_n, \\ \Sigma_n = \begin{bmatrix} \frac{\sigma^2(\alpha+1)^3}{(2\alpha+1)^{3/2}} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{\sigma^2(\alpha+1)^3(3\alpha^2+4\alpha+2)}{4(2\alpha+1)^{5/2}} \end{bmatrix}.$$

We could now apply the theory stated to test any simple or composite hypothesis on the linear regression parameters. The asymptotic distribution under the null hypothesis of the Wald type tests defined in (16) is given in Theorem 6.2 and the asymptotic distribution of the Wald-type test statistics under contiguous alternative hypothesis is given in Theorem 4.7. The non-centrality parameter in Theorem 4.7 can be expressed as

$$\delta = \mathbf{d}^{*T} [\mathbf{M}^T \Sigma_n \mathbf{M}]^{-1} \mathbf{d}^*,$$

with $\mathbf{d}^* = \mathbf{M}^T \mathbf{d}$. If we consider the composite null hypothesis (15), then

$$\delta = \frac{(2\alpha+1)^{3/2}}{\sigma(\alpha+1)^3} \mathbf{d}^{*T} \left[\mathbf{M}^T \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right) \mathbf{M} \right]^{-1} \mathbf{d}^*.$$

Now, based on Remark 5.2 we can obtain the IF of the functional associated to the minimum RP estimator of θ . These are given by

$$\begin{aligned} IF(t_{i_0}, \mathbf{T}_\alpha, F_{1,\theta}, \dots, F_{n,\theta}) &= \Psi_n^{-1}(\theta) (\mathbf{D}_{i_0,\alpha}^T(\beta), \mathbf{D}_{i_0,\alpha}(\sigma))^T, \\ IF^{(2)}(t_{i_0}, W_\alpha, F_{1,\theta}, \dots, F_{n,\theta}) &= 2 (\mathbf{D}_{i_0,\alpha}^T(\beta), \mathbf{D}_{i_0,\alpha}(\sigma)) [\Psi_n^{-1}(\theta) \Sigma_n^{-1}(\theta) \Psi_n^{-1}(\theta)] \\ &\quad \times (\mathbf{D}_{i_0,\alpha}^T(\beta), \mathbf{D}_{i_0,\alpha}(\sigma))^T, \\ IF^{(2)}(t_{i_0}, W_\alpha^0, F_{1,\theta}, \dots, F_{n,\theta}) &= 2 \left[\Psi_n^{-1}(\theta) (\mathbf{D}_{i_0,\alpha}^T(\beta), \mathbf{D}_{i_0,\alpha}(\sigma))^T \right]^T \mathbf{M} [\mathbf{M}^T \Sigma_n(\theta) \mathbf{M}]^{-1} \\ &\quad \times \mathbf{M}^T [\Psi_n^{-1}(\theta) (\mathbf{D}_{i_0,\alpha}^T(\beta), \mathbf{D}_{i_0,\alpha}(\sigma))], \end{aligned}$$

with

$$\begin{aligned} \mathbf{D}_{i_0,\alpha}(\beta) &= \frac{-1}{\sigma} \exp \left(-\frac{\alpha}{2} \left(\frac{t_{i_0} - \mathbf{x}_{i_0}^T \beta}{\sigma} \right)^2 \right) \left(\frac{t_{i_0} - \mathbf{x}_{i_0}^T \beta}{\sigma} \right) \mathbf{x}_{i_0}, \\ \mathbf{D}_{i_0,\alpha}(\sigma) &= \frac{-1}{\sigma^2} \exp \left(-\frac{\alpha}{2} \left(\frac{t_{i_0} - \mathbf{x}_{i_0}^T \beta}{\sigma} \right)^2 \right) \left[\left(\frac{t_{i_0} - \mathbf{x}_{i_0}^T \beta}{\sigma} \right)^2 - \frac{1}{\alpha+1} \right]. \end{aligned}$$

Note that, since $\Psi_n(\theta)$ and $\Sigma_n(\theta)$ are diagonal matrices, we could express separately the IF of the functional $\mathbf{T}_\alpha(\beta)$ and $\mathbf{T}_\alpha(\sigma)$ associated to the minimum RP estimator, $\widehat{\beta}_\alpha$ and $\widehat{\sigma}_\alpha$ respectively. Following [14], we consider two different fixed design matrices for the univariate linear regression model:

Design 1 Two-points design. We fix $\mathbf{x}_i = (1, x_{i1})^T$, with $x_{i1} = a, i = 1, \dots, n/2$ and $x_{i1} = b, i = n/2 + 1, \dots, n$.

Design 2 Fixed-Normal design. We fix $\mathbf{x}_i = (1, x_{i1})^T$, where $x_{i1}, i = 1, \dots, n$ are prefixed independent and identically distributed observations from a $\mathcal{N}(\mu_x = 0, \sigma_x = 1)$.

Figure 2 presents the ℓ_2 -norm of the first order IF of the minimum RP estimator and second order IF of Wald type test estimators for testing (12) with $\boldsymbol{\theta}_0 = (1, 1, 1)^T$ with both fixed designs and contamination in one direction for different values of α . Clearly, the IF is bounded for positives values of the parameter α and is unbounded at the MLE, highlighting its lack of robustness. Moreover, the supremum of the ℓ_2 -norm of the IF indicates the robustness of the estimator. Hence, we could study the optimal parameter of α through the gross error sensitivity function. We define the gross error sensitivity of the functional \mathbf{T}_α considering contamination in the i_0 direction as

$$\gamma^*(\mathbf{T}_\alpha, F_{1,\boldsymbol{\theta}}, \dots, F_{n,\boldsymbol{\theta}}) = \sup_{t_{i_0}} \{ \|IF_{i_0}(t_{i_0}, \mathbf{T}_\alpha, F_{1,\boldsymbol{\theta}}, \dots, F_{n,\boldsymbol{\theta}})\| \}. \quad (30)$$

Considering separately the influence function of the functionals $\mathbf{T}_\alpha(\boldsymbol{\beta})$ and $\mathbf{T}_\alpha(\sigma)$, it is easy to show that

$$\begin{aligned} \gamma^*(\mathbf{T}_\alpha(\boldsymbol{\beta}), F_{1,\boldsymbol{\theta}}, \dots, F_{n,\boldsymbol{\theta}}) &= \sigma \frac{(\alpha + 1)^{3/2}}{\alpha^{1/2}} \exp\left(-\frac{1}{2}\right) \left\| \left(\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right)^{-1} \mathbf{x}_{i_0} \right\|, \\ \gamma^*(\mathbf{T}_\alpha(\sigma), F_{1,\boldsymbol{\theta}}, \dots, F_{n,\boldsymbol{\theta}}) &= \frac{(\alpha + 1)^{5/2}}{\alpha} \exp\left(-\frac{3\alpha + 2}{2(\alpha + 1)}\right). \end{aligned} \quad (31)$$

Figure 1 represents the gross error sensitivity functions depending on the parameter α , using Design 1 and fixing the true standard error $\sigma = 1$. The optimal value of α depends on the functional, being $\alpha = 1/2$ and $\alpha = \sqrt{2/3}$ for $\mathbf{T}_\alpha(\boldsymbol{\beta})$ and $\mathbf{T}_\alpha(\sigma)$ respectively. Therefore, a global optimal value of α , in terms of robustness, should vary between values $\alpha = 0.5$ to $\alpha = 0.82$ if the true standard error is $\sigma = 1$.

Finally, we study the Asymptotic Relative Efficiency (ARE) of the proposed minimum RP estimators with respect to the MLE, which is B.A.N. (Best Asymptotically Normal). The ARE is computed as the ratio of their asymptotic variances. Note that this ratio does not depend on the regression parameters, but is only determined by α .

$$\begin{aligned} \text{ARE}(\widehat{\beta}_\alpha) &= \frac{(2\alpha + 1)^{3/2}}{(\alpha + 1)^3}, \\ \text{ARE}(\widehat{\sigma}_\alpha) &= \frac{2(2\alpha + 1)^{5/2}}{(\alpha + 1)^3(3\alpha^2 + 4\alpha + 2)}. \end{aligned} \quad (32)$$

Table 1 represents the ARE of the minimum RP estimator, $(\widehat{\beta}_\alpha, \widehat{\sigma}_\alpha)$. As shown, the increment of α leads to an efficiency loss, which is heightened for the standard error estimator. Therefore, to ensure sufficient efficiency, the parameter α should be chosen from low values. However, the efficiency reduction might be worth in contrast with the robustness advantage. In view of the error sensitivity function study, values above $\alpha = 0.82$ are not advocated.

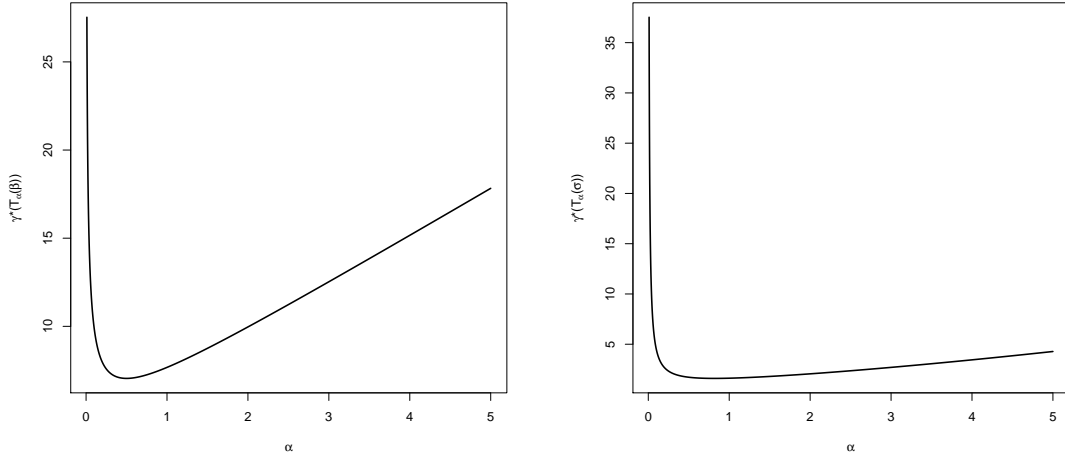


Figure 1: Gross error function for $T_\alpha(\beta)$ (left) and $T_\alpha(\sigma)$ (right).

Table 1: ARE of the minimum RP estimator with respect to the MLE for the multiple linear regression model for different values of α .

α	0	0.1	0.2	0.3	0.4	0.5	0.8	1	1.5
$\text{ARE}(\widehat{\beta}_\alpha) (\times 100)$	100.00	98.76	95.86	92.12	88.01	83.81	71.89	64.95	51.20
$\text{ARE}(\widehat{\sigma}_\alpha) (\times 100)$	100.00	97.54	91.92	84.95	77.65	70.57	52.50	43.30	27.77

7 Numerical results

We empirically evaluate the performance of the proposed Wald-type test statistics based on minimum RP estimator for MLRM through an extensive simulation study. We consider the univariate regression model with fixed design matrix

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i, \quad i = 1, \dots, n$$

and the two different design matrices presented in Section 6. We generate the response variable from the linear regression model (24) with regression parameters $\beta^0 = (1, 1)$ and $\sigma^0 = 1$. To introduce contamination on the data, we swap the true regression vector to $\beta^0 = (1.5, 2)$ for a 10% of the sample size. We analyse the performance of Wald-type tests for simple null hypothesis on both regression parameters, β and σ , at different values of the tuning parameter α . Note that, for the proposed design matrices, the matrix $\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T$ is finitely defined and is positive definite.

We consider two different null hypotheses

$$H_0 : \beta_1 = 1, \tag{33}$$

$$H_0 : \sigma = 1, \tag{34}$$

corresponding with the composite null hypothesis

$$H_0 : \mathbf{M}^T \beta = \mathbf{m},$$

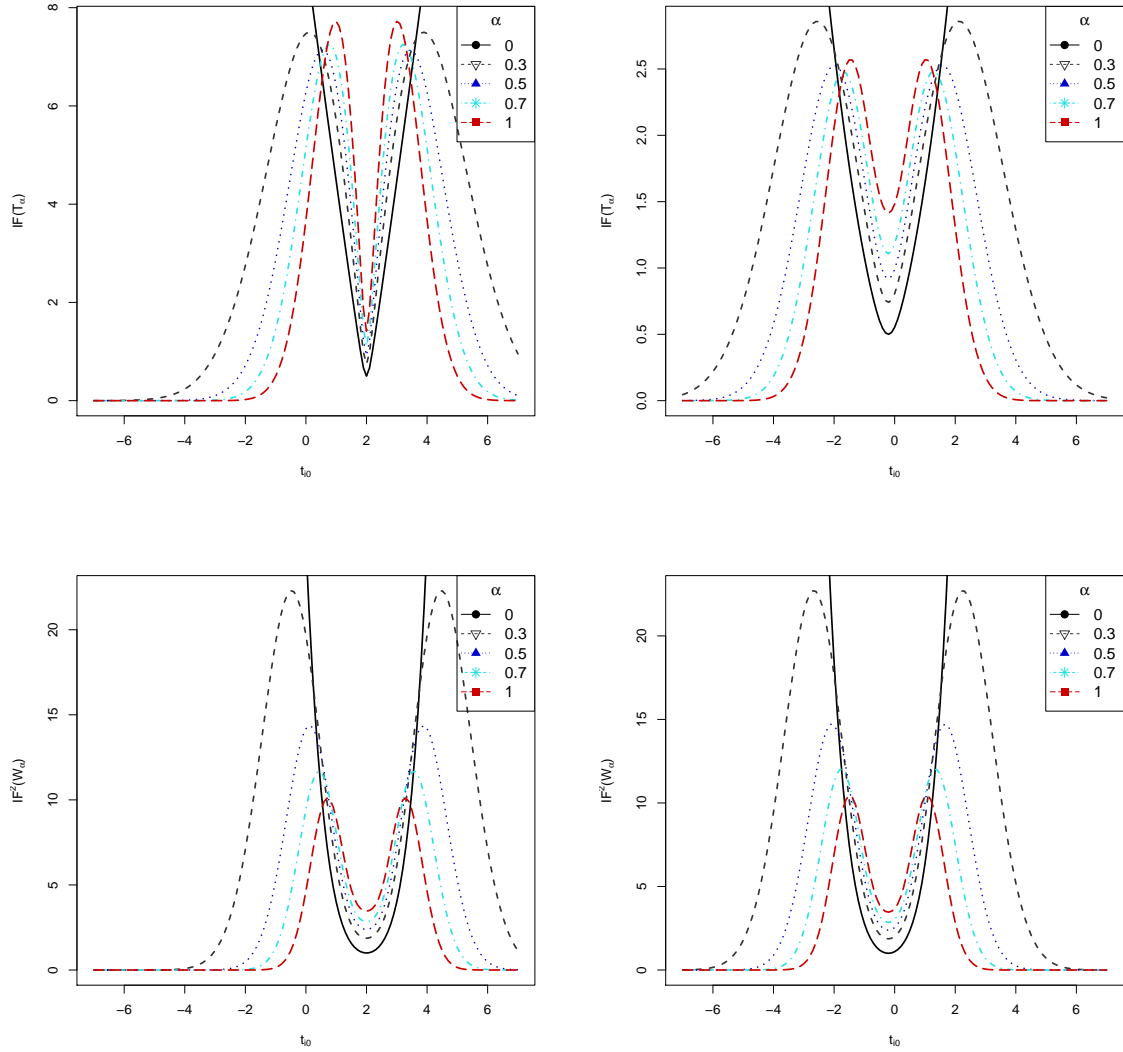


Figure 2: ℓ_2 -norm of the first order IF of the minimum RP estimator (top) and second order IF of Wald type test estimators for testing (12) with $\theta_0 = (1, 1, 1)^T$ (bottom) with fixed Design 1 (left) and Design 2 (right), and contamination in the direction $i_0 = 1$.

with

$$\begin{aligned} \mathbf{M}_{\beta_1} &= [0, 1, 0] \quad \text{and} \quad \mathbf{m}_{\beta_1} = 1, \\ \mathbf{M}_{\sigma} &= [0, 0, 1] \quad \text{and} \quad \mathbf{m}_{\sigma} = 1, \end{aligned}$$

respectively. The Wald-type test statistics for testing (33)-(34), which we will denote $W_n(\widehat{\beta}_{1\alpha})$ and $W_n(\widehat{\sigma}_\alpha)$, are given in (16) by substituting the corresponding matrices. So as to investigate the trade-off between efficiency and robustness depending on the tuning parameter α , we compute the empirical levels for the proposed Wald-type tests and powers when the true parameter values are $\beta_1^0 = 0.45$ and $\sigma^0 = 0.8$, respectively. These levels and powers are computed as the number of times that the null hypothesis is rejected over the total simulated samples $R = 1000$. Figures 3-6 contain the root mean square error (RMSE), empirical level and power results for the null hypothesis tests (33) and (34), for a 5% significance level. The results show the clear improvement in robustness when α increases, in detriment to the efficiency. The MLE produces the best performance with pure data, showing its major efficiency, and the behaviour of the minimum RP estimator improves when α decreases, i.e., estimators based in low values of the parameter enjoy greater efficiency. However, in presence of data contamination, the RMSE and empirical level of the Wald-type test statistics rise for low values of α , highlighting its lack of robustness. The most revealing setting is Design 1, at which the empirical level and power of the Wald-type tests based on the MLE reaches their worst results, but the proposed Wald-type test based on RP loss statistics continues to perform adequately for sufficiently high values of α .

For the first hypothesis test (33), we could apply Theorem 4.7 to obtain the power under the contiguous alternative hypothesis (18). The distribution of the Wald-type tests $W_n(\widehat{\beta}_1)$ is given by a chi-squared with 1 degree of freedom and non-centrality parameter

$$\delta = \frac{\sigma(\alpha + 1)^3}{(2\alpha + 1)^{3/2}} \mathbf{d}^{*T} \left(\frac{1}{n} \sum_{i=1}^n X_{1i}^2 \right)^{-1} \mathbf{d}^*,$$

depending on the standard deviation error σ , the tuning parameter α and the fixed value $d_x = \mathbf{d}^{*T} \left(\frac{1}{n} \sum_{i=1}^n X_{1i}^2 \right)^{-1} \mathbf{d}^*$. The choice $d_x = 0$ corresponds with the level of the test. Table 2 summarizes the empirical power results over different values of α and d_x , with $\sigma = 1$.

Table 2: Empirical power values of the null hypothesis (33) under contiguous hypothesis.

α	d_x							
	0	2	5	10	15	20	25	30
0	0.05	0.28	0.59	0.88	0.97	0.99	1.00	1.00
0.2	0.05	0.27	0.58	0.86	0.97	0.99	1.00	1.00
0.5	0.05	0.25	0.52	0.81	0.94	0.98	1.00	1.00
0.8	0.05	0.22	0.44	0.75	0.90	0.97	0.99	1.00
1	0.05	0.21	0.41	0.71	0.87	0.95	0.98	0.99
1.5	0.05	0.17	0.35	0.60	0.78	0.89	0.95	0.97

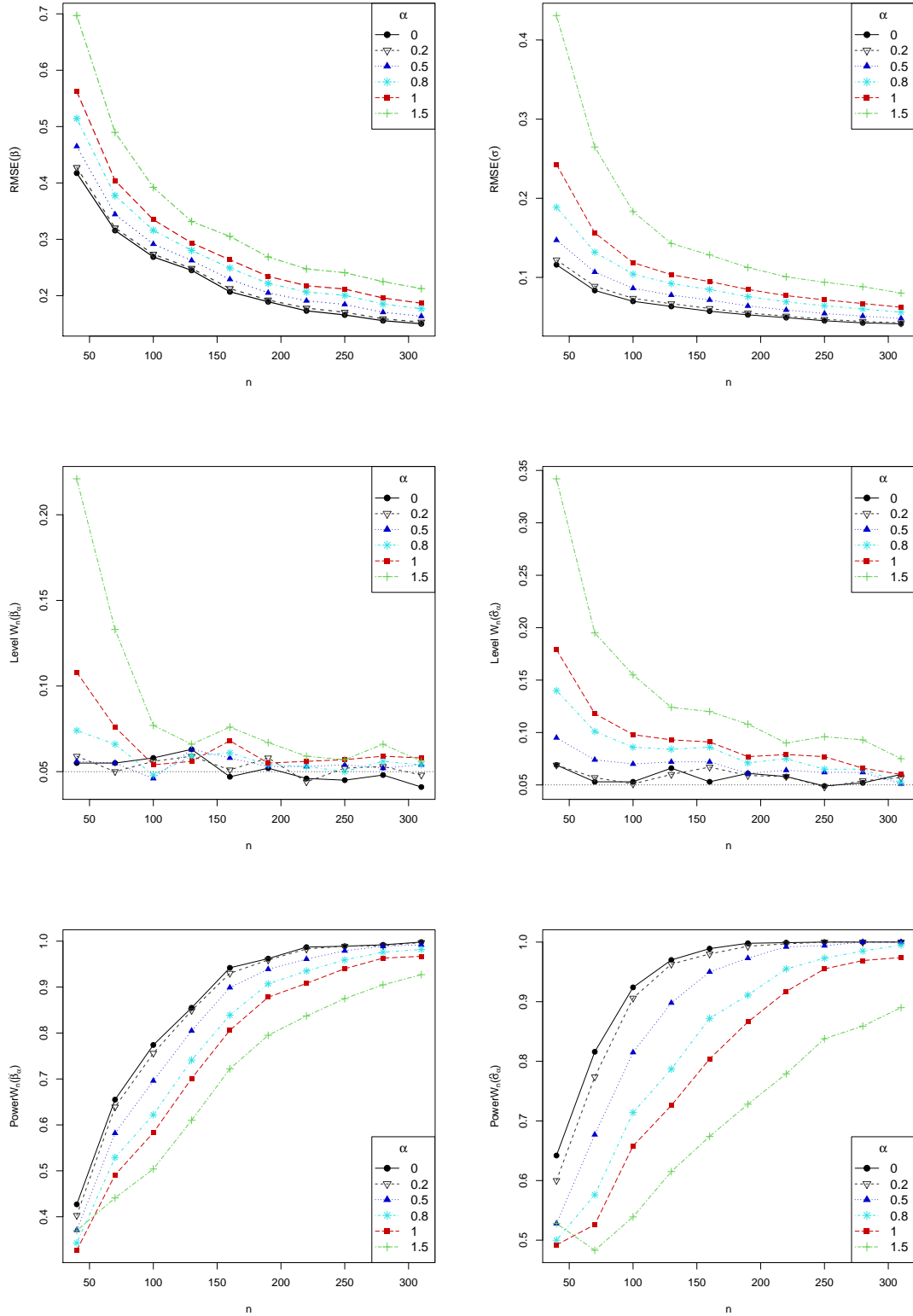


Figure 3: RMSE (top) empirical level (middle) and empirical power (bottom) against sample size for the null hypothesis (33) (left) and (34) (right) for the corresponding Wald type tests with pure data and Design 1.

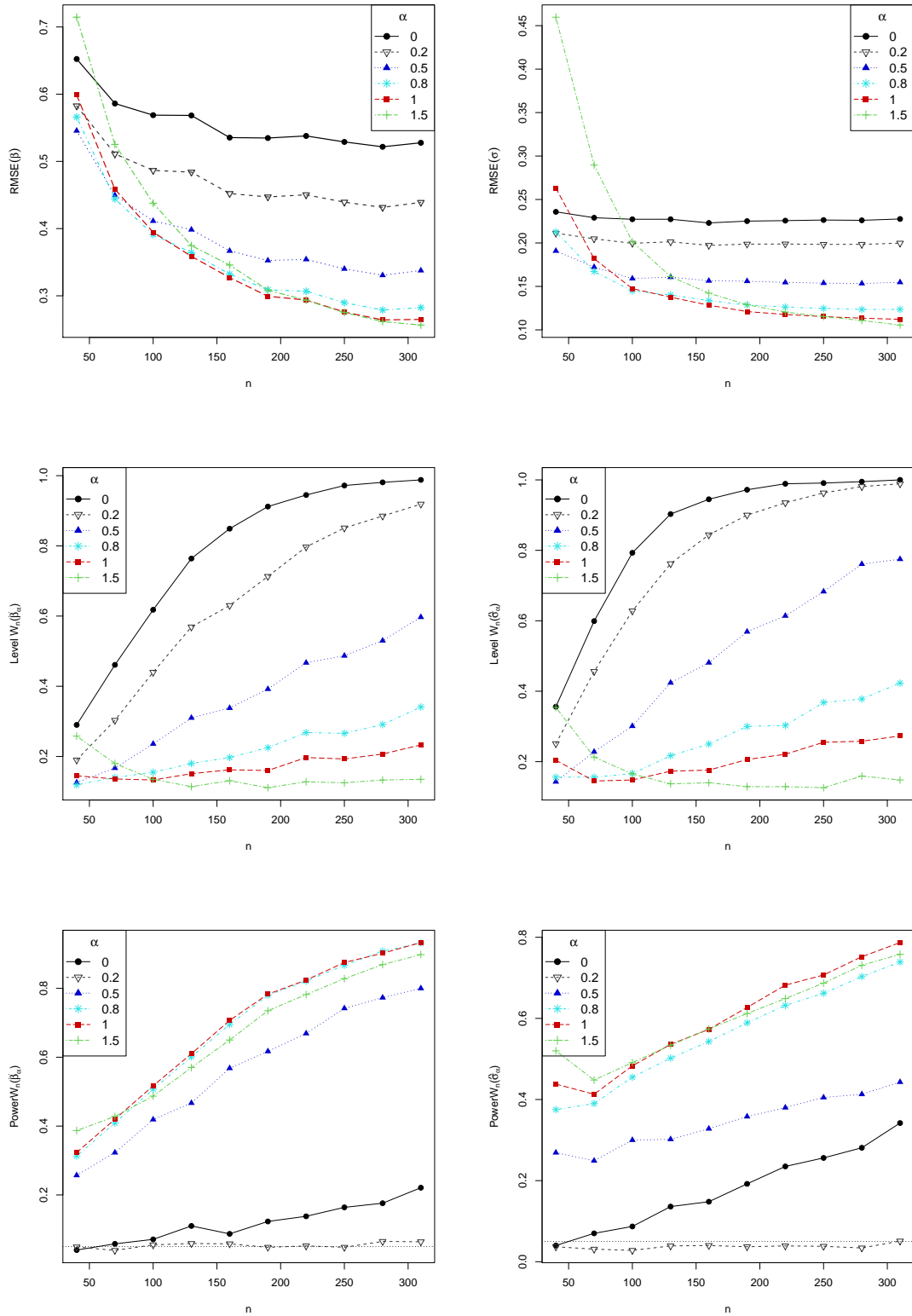


Figure 4: RMSE (top) empirical level (middle) and empirical power (bottom) against sample size for the null hypothesis (33) (left) and (34) (right) for the corresponding Wald-type tests with 10% of outliers and Design 2.

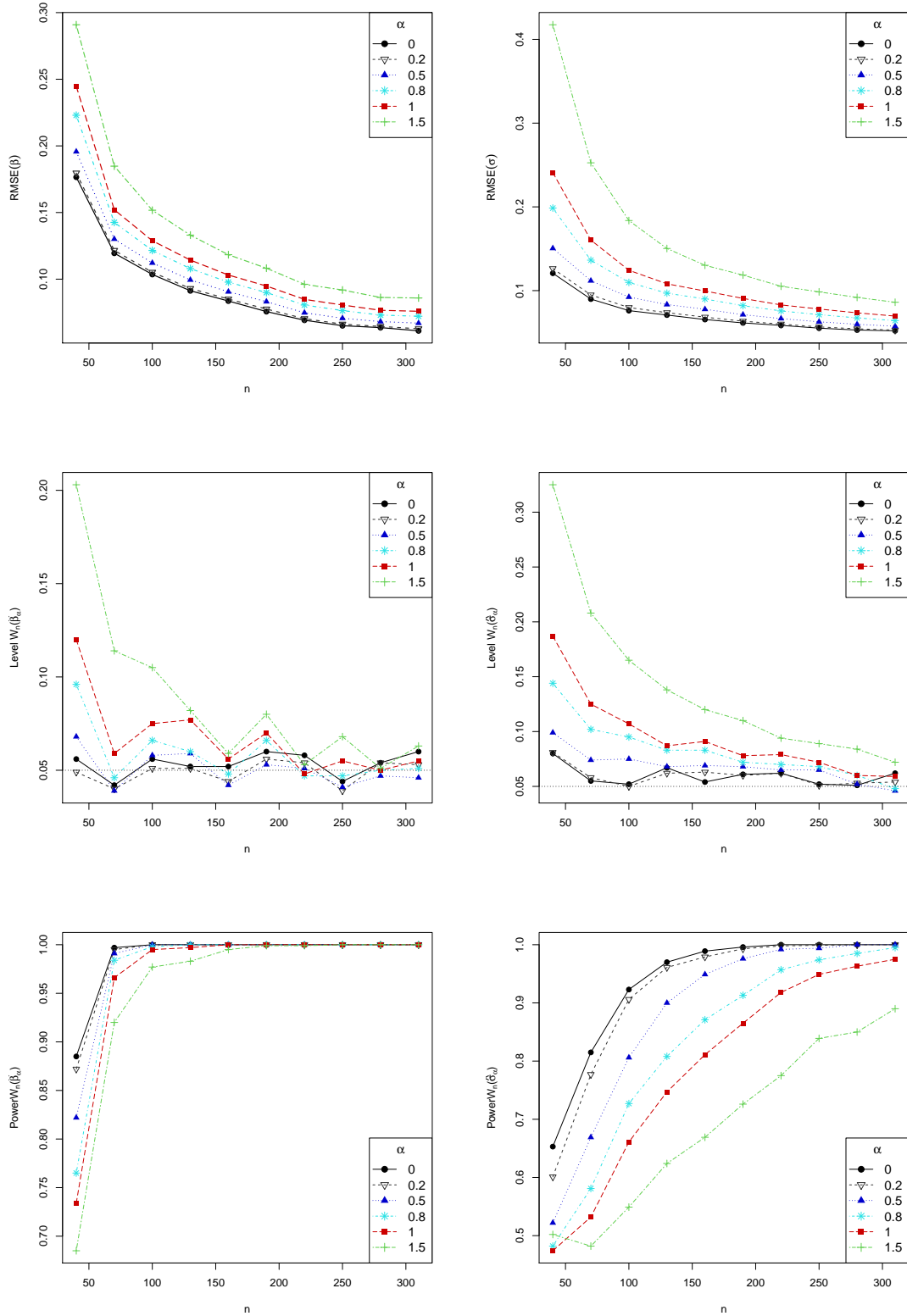


Figure 5: RMSE (top) empirical level (middle) and empirical power (bottom) against sample size for the null hypothesis (33) (left) and (34) (right) for the corresponding Wald-type tests with pure data and Design 2.

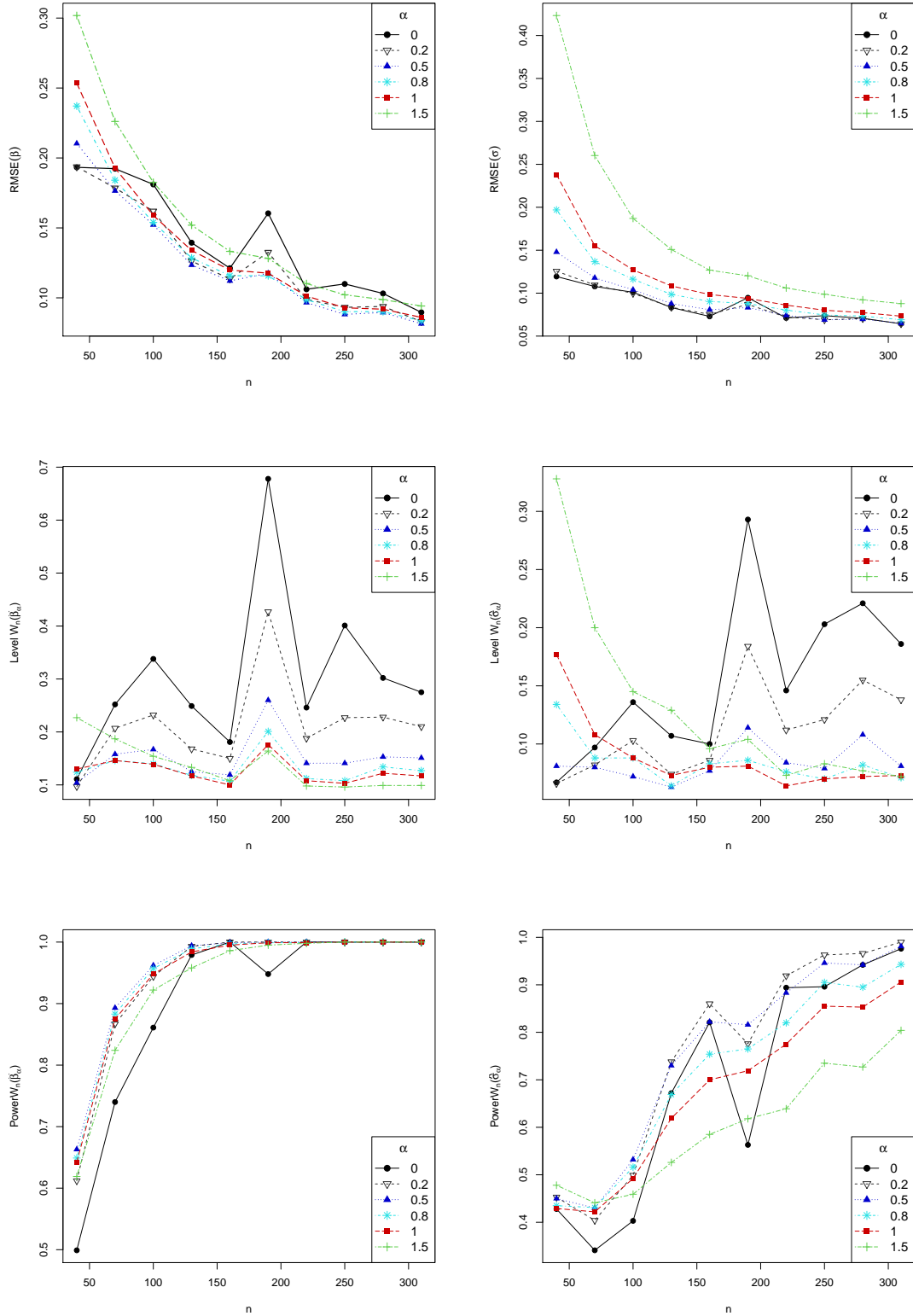


Figure 6: RMSE (top) empirical level (middle) and empirical power (bottom) against sample size for the null hypothesis (33) (left) and (34) (right) for the corresponding Wald-type tests with 10% of outliers and Design 2.

Note that greater values of d_x produces greater power values as expected, and empirical power decreases with α . However, the efficiency loss is not very significant in comparison to the robustness advantage.

8 Real Data examples

8.1 Brain and Weight Data

These data, adapted from a larger data set in [18], were presented in Rousseeuw and Leroy ([19], pp. 57) as an example of the unrobustness of the classical MLE in simple linear regression. In this sample, the body weight (in kilograms) and the brain weight (in grams) of $n = 28$ animals are compared, to investigate if a larger brain is required to govern a heavier body. As suggested in in [19], a transformation should be done to clearly represent either the larger or smaller measurements. In this case, we take the Napierian logarithm of both brain and body weights. Observations 6, 16 and 25, those corresponding to dinosaurs, possess an unusual small brain as compared with a heavy body, which clearly affects to the slope of the classical estimation method ($\alpha = 0$) as can be seen in Figure 7 and Table 3. The estimates of the regression coefficients and the error variance obtained from the minimum RP estimation for various α are also presented here, observing how the estimation based on $\alpha > 0$ is more robust to the presence of these outliers.

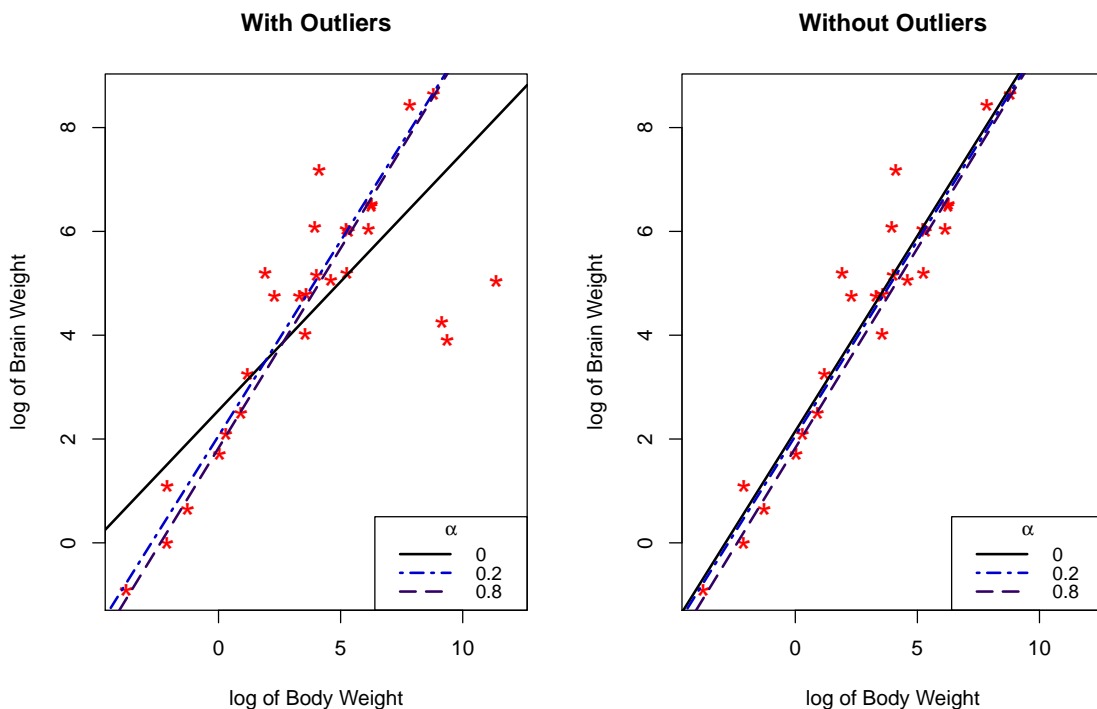


Figure 7: Plots of the data-points and fitted regression lines for the Brain and Weight Data using several minimum RP estimators before and after deleting the outliers.

Table 3: The parameter estimates of the linear regression model for the Brain and Weight Data using several minimum RP estimators.

α	With outliers			Without outliers		
	σ	β_0	β_1	σ	β_0	β_1
0	1.4714	2.5523	0.4958	0.6962	2.1504	0.7522
0.2	0.6410	2.0617	0.7509	0.6309	2.0580	0.7519
0.4	0.4929	1.9378	0.7560	0.4929	1.9378	0.7560
0.6	0.4092	1.8616	0.7634	0.4092	1.8616	0.7634
0.8	0.3640	1.8265	0.7694	0.3640	1.8265	0.7694
1	0.3378	1.8142	0.7731	0.3378	1.8142	0.7731

In order to compare the performance of the Wald-type test with different values of the tuning parameter α , we consider the following tests

$$H_0 : \beta_0 = 1.98, \quad (35)$$

$$H_0 : \beta_1 = 0.73, \quad (36)$$

$$H_0 : (\beta_0, \beta_1) = (1.98, 0.73), \quad (37)$$

where the values 1.98 and 0.73 are respectively the mean value of the estimated coefficients β_0 and β_1 with the different values of α and using the original data (with outliers) listed in Table 3. Table 8.1 shows the p-values obtained by using the corresponding Wald-type test statistics, $W_n(\hat{\beta}_0)$, $W_n(\hat{\beta}_1)$ and $W_n(\hat{\beta})$.

Table 4: p-value obtained for the tests (35)-(37) using the corresponding Wald-type test statistics.

α	With outliers			Without outliers		
	$W_n(\hat{\beta}_0)$	$W_n(\hat{\beta}_1)$	$W_n(\hat{\beta})$	$W_n(\hat{\beta}_0)$	$W_n(\hat{\beta}_1)$	$W_n(\hat{\beta})$
0	0.080	0.000	0.000	0.452	0.542	0.072
0.2	0.713	0.556	0.204	0.723	0.537	0.197
0.4	0.833	0.437	0.358	0.833	0.437	0.358
0.6	0.539	0.310	0.305	0.539	0.310	0.305
0.8	0.423	0.236	0.235	0.423	0.236	0.235
1	0.393	0.203	0.203	0.393	0.203	0.203

As shown, the robustness of the test increases with α , showing the robustness improvement of the proposed Wald-type test statistics. Note that the major difference between the p-value using clean data and data with outliers is obtained with the value $\alpha = 0$ corresponding to the classical MLE.

8.2 First Word Data

These data, originally presented in Mickey et al. [20], consist on $n = 21$ observations and relate the age in which children speak their first word to their Gesell adaptative score, a measure of mental ability. By means of a sequential approach to detect outliers via

stepwise regression, [20] concluded that observation 18 was an outlier. While estimates of the regression coefficients obtained with the MLE do not change excessively when omitting this outlier (Figure 8), we do observe a greater change in the error variance estimation (Table 5). As expected, minimum RP estimates for $\alpha > 0$ remain more robust.

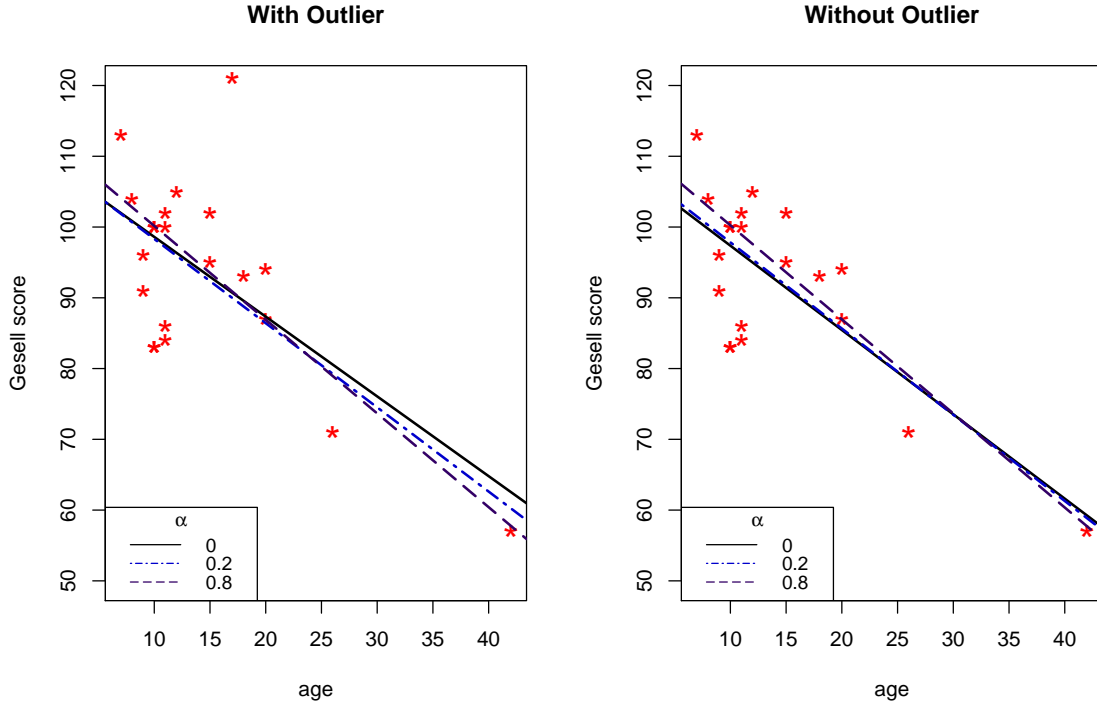


Figure 8: Plots of the data-points and fitted regression lines for the First Word Data using several minimum RP estimators before and after deleting the outliers.

Table 5: The parameter estimates of the linear regression model for the First Word Data using several minimum RP estimators.

α	With outliers			Without outliers		
	σ	β_0	β_1	σ	β_0	β_1
0	10.4845	109.8730	-1.1269	8.1976	109.2816	-1.1916
0.2	9.7860	110.2068	-1.1897	8.5501	110.0225	-1.2183
0.4	9.2980	110.8118	-1.2338	8.7780	110.8276	-1.2451
0.6	9.0319	111.7370	-1.2710	8.8019	111.8168	-1.2767
0.8	8.3349	113.4011	-1.3246	8.1972	113.5345	-1.3292
1	4.4187	116.6086	-1.4065	4.4187	116.6086	-1.4065

As in the previous example, we consider the following tests

$$H_0 : \beta_0 = 112.56, \tag{38}$$

$$H_0 : \beta_1 = -1.28, \tag{39}$$

$$H_0 : (\beta_0, \beta_1) = (112.56, -1.28), \tag{40}$$

where the values 112.56 and 0.73 are respectively the mean value of the estimated coefficients β_0 and β_1 with the different values of α and using the original data (with outliers) listed in Table 5. Table 8.2 shows the p-values obtained by using the corresponding Wald-type test statistics, $W_n(\widehat{\beta}_0)$, $W_n(\widehat{\beta}_1)$ and $W_n(\widehat{\beta})$.

Table 6: p-value obtained for the tests (38)-(40) using the corresponding Wald-type test statistics.

α	With outliers			Without outliers		
	$W_n(\widehat{\beta}_0)$	$W_n(\widehat{\beta}_1)$	$W_n(\widehat{\beta})$	$W_n(\widehat{\beta}_0)$	$W_n(\widehat{\beta}_1)$	$W_n(\widehat{\beta})$
0	0.072	0.098	0.071	0.013	0.293	0.001
0.2	0.110	0.331	0.065	0.065	0.485	0.007
0.4	0.243	0.636	0.098	0.234	0.719	0.061
0.6	0.596	0.949	0.318	0.628	0.997	0.311
0.8	0.590	0.620	0.588	0.529	0.583	0.529
1	0.001	0.078	0.000	0.001	0.078	0.000

The results highlight again the gain in robustness.

9 Concluding remarks

In this paper we have presented the minimum RP estimators for the case of i.n.i.d.o. Wald-type tests based on them are also developed. Classical MLE and Wald-test are obtained as a particular case of these new estimators and tests. In particular, we have studied the case of MLRM. Through the study of the influence functions and the development of an extensive simulation study we prove their robustness from a theoretical and practical point of view, respectively. Application to different models is a problem that will be of interest for further consideration.

Acknowledgements: This research is supported by the Spanish Grants no. PGC2018-095194-B-100, no. FPU19/01824 and no. FPU16/03104.

A Proof of Results

A.1 Proof of Theorem 3.1

The proof follows similar steps that the proof presented in [12] for the minimum DPD estimators for i.n.i.d.o and the proof presented in [17] for the MLE with i.n.i.d.o.

To prove the existence, with probability tending to 1, of a consistent sequence of solutions of the system of equations (10), we study the behaviour of the objective function in (6), $H_n^\alpha(\boldsymbol{\theta})$, on a neighbourhood of the true parameter value. We consider the sphere Q_a with center at the true value of the parameter $\boldsymbol{\theta}^*$ and radius a . We will show that for any sufficiently small a

$$H_n^\alpha(\boldsymbol{\theta}) < H_n^\alpha(\boldsymbol{\theta}_0)$$

with probability tending to 1 at all points $\boldsymbol{\theta}$ on the surface of Q_a . This inequality ensures that the objective function $H_n^\alpha(\boldsymbol{\theta})$ has a local maximum in the interior of Q_a . Since $H_n^\alpha(\boldsymbol{\theta})$ is differentiable the system of equations (10) must be satisfied at a local maximum. Therefore, for any $a > 0$, the system of equations (10) has a solution $\widehat{\boldsymbol{\theta}}_n(a)$ within Q_a verifying

$$\lim_{n \rightarrow \infty} P_{\boldsymbol{\theta}^*} \left(\left\| \widehat{\boldsymbol{\theta}}_n(a) - \boldsymbol{\theta}^* \right\| \right) = 1.$$

We consider a Taylor series expansion of $H_n^\alpha(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0 = (\theta_0^1, \dots, \theta_0^p)$,

$$\begin{aligned} H_n^\alpha(\boldsymbol{\theta}) - H_n^\alpha(\boldsymbol{\theta}^*) &= \sum_{j=1}^p \left(\frac{\partial H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\theta_j - \theta_j^0) \\ &+ \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \left(\frac{\partial^2 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\theta_j - \theta_j^0) (\theta_k - \theta_k^0) \\ &+ \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \left(\frac{\partial^3 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\theta_j - \theta_j^0) (\theta_k - \theta_k^0) (\theta_l - \theta_l^0) \\ &= \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\theta_j - \theta_j^0) \\ &+ \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\theta_j - \theta_j^0) (\theta_k - \theta_k^0) \\ &+ \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^3 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\theta_j - \theta_j^0) (\theta_k - \theta_k^0) (\theta_l - \theta_l^0) \\ &= L_1 + L_2 + L_3, \end{aligned} \tag{41}$$

where $\boldsymbol{\theta}^*$ belong to the interior of the ball centred on $\boldsymbol{\theta}^*$ and radius a . We study separately right-hand terms L_1, L_2 and L_3 in (41).

Using assumption C6., we have that

$$A_j^{(n)} = \left(\frac{\partial H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial V_i(Y; \boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] = 0.$$

We are going to establish the last equality,

$$\frac{\partial V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j} = \frac{1}{L_\alpha^i(\boldsymbol{\theta})^2} \left(\alpha f_i(Y, \boldsymbol{\theta})^\alpha u_j(Y, \boldsymbol{\theta}) L_\alpha^i(\boldsymbol{\theta}) - \frac{\partial L_\alpha^i(\boldsymbol{\theta})}{\partial \theta_j} f_i(Y, \boldsymbol{\theta})^\alpha \right).$$

with $u_j(y, \boldsymbol{\theta}) = \frac{\partial \log(f_i(Y, \boldsymbol{\theta}))}{\partial \theta_j}$. But

$$\begin{aligned} \frac{\partial L_\alpha^i(\boldsymbol{\theta})}{\partial \theta_j} &= \frac{\alpha}{\alpha + 1} \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \right)^{\frac{\alpha}{\alpha+1}-1} (\alpha + 1) \int f_i(y, \boldsymbol{\theta})^{\alpha+1} u_j(y, \boldsymbol{\theta}) dy \\ &= \alpha \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \right)^{\frac{\alpha}{\alpha+1}-1} \int f_i(y, \boldsymbol{\theta})^{\alpha+1} u_j(y, \boldsymbol{\theta}) dy. \end{aligned}$$

Therefore,

$$\begin{aligned} L_\alpha^i(\boldsymbol{\theta})^2 \frac{\partial V_i(Y; \boldsymbol{\theta})}{\partial \theta_j} &= \alpha f_i(Y, \boldsymbol{\theta})^\alpha u_j(Y, \boldsymbol{\theta}) L_\alpha^i(\boldsymbol{\theta}) \\ &\quad \left[\alpha \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \right)^{\frac{\alpha}{\alpha+1}-1} \int f_i(y, \boldsymbol{\theta})^{\alpha+1} u_j(y, \boldsymbol{\theta}) dy \right] f_i(Y, \boldsymbol{\theta})^\alpha \end{aligned}$$

and

$$\begin{aligned} E_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] &= \int (\alpha f_i(y, \boldsymbol{\theta}^*)^\alpha u_j(y, \boldsymbol{\theta}^*) L_\alpha^i(\boldsymbol{\theta}_0)) f_i(y, \boldsymbol{\theta}^*) dy \\ &\quad - \int \left[\alpha \left(\int f_i(y, \boldsymbol{\theta}^*)^{\alpha+1} dy \right)^{\frac{\alpha}{\alpha+1}-1} \int f_i(y, \boldsymbol{\theta}_0)^{\alpha+1} u_j(y, \boldsymbol{\theta}^*) dy \right] f_i(y, \boldsymbol{\theta}^*)^{\alpha+1} dy \\ &= 0. \end{aligned}$$

On the other hand, we denote

$$B_{jk}^{(n)} = \left(\frac{\partial^2 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

and applying again condition C6., we obtain the convergence

$$B_{jk}^{(n)} = \left(\frac{\partial^2 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^2 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial^2 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] = (-\boldsymbol{\Psi}_n)_{jk}.$$

Finally, applying condition C6. to the third derivative, we have

$$\left(\frac{\partial^3 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^3 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial^3 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right].$$

Assumption C5. ensures the existence of $M_{jkl}, j, k, l = 1, \dots, p$ s.t.

$$\left| \left(\frac{\partial^3 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right| \leq M_{jkl}^{(i)}(y),$$

and therefore there exists $\gamma_{jkl}^{(i)}(y)$ verifying

$$0 \leq |\gamma_{jkl}^{(i)}(y)| \leq 1 \tag{42}$$

in such a way that

$$\left(\frac{\partial^3 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = M_{jkl}^{(i)}(Y_i) \gamma_{jkl}^{(i)}(Y_i).$$

and

$$E_{\boldsymbol{\theta}^*} \left[M_{jkl}^{(i)}(Y_i) \right] = m_{jkl},$$

with

$$\left| \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial^3 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right] \right| < m_{jkl}.$$

The previous convergence provide that for all a and for all ε there exists n_0 such that for all $n > n_0$ we have

$$\begin{aligned} P \left(\left| A_j^{(n)} \right| > a^2 \right) &< \frac{\varepsilon}{p+p^2+p^3} \\ P \left(\left| B_{jk}^{(n)} - (\boldsymbol{\Psi}_n)_{jk} \right| \geq a \right) &< \frac{\varepsilon}{p+p^2+p^3} \\ P \left(\left| \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^3 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right| \geq 2m_{jkl} \right) &\leq \frac{\varepsilon}{p+p^2+p^3}. \end{aligned}$$

We shall denote now by S^* the event containing the $p + p^2 + p^3$ inequalities,

$$\left\{ \left| A_j^{(n)} \right| > a^2; \left| B_{jk}^{(n)} - (-\Psi_n)_{jk} \right| \geq a; \left| \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^3 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right| \geq 2m_{jkl} \right\}.$$

It is clear that $P(S^*) < \varepsilon$ and $P((S^*)^C) \geq 1 - \varepsilon$. In the following we denote $S = (S^*)^C$. We finally study the sign of $H_n^\alpha(\boldsymbol{\theta}) - H_n^\alpha(\boldsymbol{\theta}^*)$ under the event S and for $\boldsymbol{\theta} \in Q_a$.

Since $\boldsymbol{\theta} \in Q_a$ in S holds

$$|L_1| = \left| \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^n \left(\frac{\partial V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\theta_j - \theta_j^0) \right| \leq p a a^2 \quad (43)$$

and

$$\left| \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \left\{ \left(B_{jk}^{(n)} - (-\Psi_n)_{jk} \right) \right\} (\theta_j - \theta_j^0) (\theta_k - \theta_k^0) \right| \leq \frac{1}{2} p^2 a^2 a.$$

We now consider the negative quadratic form

$$\mathbf{A} = -\frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p (\theta_j - \theta_j^0) (\theta_k - \theta_k^0) \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial^2 V_i(Y; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right].$$

An orthogonal transformation can reduce the quadratic form \mathbf{A} to its diagonal form $\mathbf{A} = \sum_{i=1}^p \lambda_i \xi_i^2$ with $\sum_{i=1}^p \xi_i^2 = \sum_{i=1}^p (\theta_i - \theta_i^0)^2 = a^2$. Shorting the negatives eigenvalues λ_i we get

$$\sum_{i=1}^p \lambda_i \xi_i^2 \leq -\lambda_0 a^2 < 0.$$

A study of the sign of the function $\frac{1}{2} p^2 a^3 - \lambda_0 a^2$ proves that we can find $c > 0, a_0 > 0$ so that for $a < a_0$

$$\begin{aligned} |L_2| &= \left| \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \left\{ \left(B_{jk}^{(n)} - (-\Psi_n)_{jk} \right) \right\} (\theta_j - \theta_j^p) (\theta_k - \theta_k^p) \right. \\ &\quad \left. + \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p (\theta_j - \theta_j^p) (\theta_k - \theta_k^p) (-\Psi_n)_{jk} \right| < -ca^2. \end{aligned}$$

Lastly,

$$\begin{aligned} |L_3| &= \left| \frac{1}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial^3 V_i(Y_i; \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\theta_j - \theta_j^p) (\theta_k - \theta_k^p) (\theta_l - \theta_l^p) \right| \\ &< \frac{2}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p m_{jkl} a^3 = a^3 b, \end{aligned}$$

being

$$b = \frac{2}{6} \sum_{j=1}^p \sum_{k=1}^p \sum_{l=1}^p m_{jkl}.$$

Therefore,

$$H_n^\alpha(\boldsymbol{\theta}) - H_n^\alpha(\boldsymbol{\theta}_0) < pa^3 - ca^2 + ba^3$$

and $pa^3 - ca^2 + ba^3 < 0$ if and only if $a < \frac{c}{b+p}$. Therefore assuming $a < \frac{c}{b+p}$ we get that in the event S

$$H_n^\alpha(\boldsymbol{\theta}) - H_n^\alpha(\boldsymbol{\theta}_0) < 0 \quad \forall \boldsymbol{\theta} \in Q_a.$$

Thus the event C involving all $\boldsymbol{\theta} \in Q_a$ s.t. $H_n^\alpha(\boldsymbol{\theta}) - H_n^\alpha(\boldsymbol{\theta}^*) < 0$, is contained in S , i.e. $P(C) \geq P(S) > 1 - \varepsilon$. Choosing a lower than $\min\left(a_0, \frac{c}{b+p}\right)$, we have

$$\lim_{n \rightarrow \infty} P(\forall \boldsymbol{\theta} \in Q_a / H_n^\alpha(\boldsymbol{\theta}) - H_n^\alpha(\boldsymbol{\theta}_0) < 0) = 1.$$

Thus, there exists $\widehat{\boldsymbol{\theta}}_n(a) \in Q_a$, i.e., $\|\widehat{\boldsymbol{\theta}}_n(a) - \boldsymbol{\theta}_0\| < a$ such that $H_n^\alpha(\boldsymbol{\theta})$ has a local maximum in $\widehat{\boldsymbol{\theta}}_n(a)$, i.e.,

$$\forall a \leq \min\left(a_0, \frac{c}{b+p}\right),$$

we obtain the required convergence

$$\lim_{n \rightarrow \infty} P\left(\|\widehat{\boldsymbol{\theta}}_n(a) - \boldsymbol{\theta}^*\| < a\right) = 1.$$

A.2 Proof of Theorem 3.2

We denote

$$H_{n,j}^\alpha(\boldsymbol{\theta}) = \frac{\partial H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j}$$

with $H_n^\alpha(\boldsymbol{\theta})$ defined in (6). A Taylor expansion of $H_{n,j}^\alpha(\boldsymbol{\theta})$ around $\boldsymbol{\theta}^*$, gives,

$$\begin{aligned} H_{n,j}^\alpha(\boldsymbol{\theta}) &= H_{n,j}^\alpha(\boldsymbol{\theta}^*) + \sum_{k=1}^p \left(\frac{\partial^2 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\theta_k - \theta_k^0) \\ &\quad + \frac{1}{2} \sum_{k=1}^p \sum_{l=1}^p \left(\frac{\partial^3 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\theta_k - \theta_k^0) (\theta_l - \theta_l^0) \end{aligned}$$

with $\boldsymbol{\theta}^*$ in the segment connecting $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$. It is clear that at the minimum RP estimator the function $H_{n,j}^\alpha$ vanishes, $H_{n,j}^\alpha(\widehat{\boldsymbol{\theta}}_\alpha) = 0$. Therefore,

$$\begin{aligned} H_{n,j}^\alpha(\boldsymbol{\theta}^*) &= - \sum_{k=1}^p \left(\frac{\partial^2 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} (\widehat{\theta}_{\alpha,k} - \theta_k^0) \\ &\quad - \frac{1}{2} \sum_{k=1}^p \sum_{l=1}^p \left(\frac{\partial^3 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\widehat{\theta}_{\alpha,k} - \theta_k^0) (\widehat{\theta}_{\alpha,l} - \theta_l^0). \end{aligned}$$

Using that

$$H_{n,j}^\alpha(\boldsymbol{\theta}^*) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

it holds

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \theta_j} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= \sqrt{n} \sum_{k=1}^p (\widehat{\theta}_{\alpha,k} - \theta_k^0) \left\{ - \left(\frac{\partial^2 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right. \\ &\quad \left. - \frac{1}{2} \sum_{k=1}^p \sum_{l=1}^p \left(\frac{\partial^3 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\widehat{\theta}_{\alpha,l} - \theta_l^0) \right\}. \end{aligned}$$

If we denote,

$$\begin{aligned} Z_{kn} &= \sqrt{n} \sum_{k=1}^p \left(\hat{\theta}_{\alpha,k} - \theta_k^0 \right), \\ A_{jkn} &= - \left(\frac{\partial^2 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} - \frac{1}{2} \sum_{k=1}^p \sum_{l=1}^p \left(\frac{\partial^3 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \left(\hat{\theta}_{\alpha,l} - \theta_l^0 \right), \\ T_{jn} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \end{aligned}$$

we can write

$$T_{jn} = \sum_{k=1}^p A_{jkn} Z_{kn}.$$

Finally, we define the following vectors

$$\begin{aligned} \mathbf{Z}_n &= (Z_{1n}, \dots, Z_{pn})^T, \\ \mathbf{T}_n &= (T_{1n}, \dots, T_{pn})^T, \\ \mathbf{A}_n &= (A_{jkn})_{j=1, \dots, p; k=1, \dots, p}. \end{aligned}$$

It is clear that

$$\begin{aligned} \mathbf{T}_n &= \mathbf{A}_n \mathbf{Z}_n \\ &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \theta_1} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \dots, \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \theta_p} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)^T \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \end{aligned}$$

and it is a simple exercise to verify that $V_i(Y, \boldsymbol{\theta})$, $i = 1, \dots, n$, are independent with

$$E_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] = 0$$

and

$$Var_{\boldsymbol{\theta}^*} \left[\left(\frac{\partial V_i(Y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right] < \infty.$$

By Assumption C7. and applying the multivariate extension of Lindeberg-Levy central limit theorem we get

$$\sqrt{n} \boldsymbol{\Omega}_n^{-\frac{1}{2}} \mathbf{T}_n \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \mathbf{I}_p)$$

or equivalently

$$\sqrt{n} \boldsymbol{\Omega}_n^{-\frac{1}{2}} \mathbf{A}_n \mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \mathbf{I}_p).$$

By assumption C5.,

$$\left(\frac{\partial^3 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

is bounded with probability tending to one. Therefore based on the consistency of $\hat{\boldsymbol{\theta}}_\alpha$ we have that the second term of A_{jkn} converges to zero in probability. Moreover

$$- \left(\frac{\partial^2 H_n^\alpha(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow[n \rightarrow \infty]{P} (\Psi_n)_{jk}$$

and hence

$$\Omega_n^{-\frac{1}{2}} (\mathbf{A}_n - \Psi_n) \mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{P} \mathbf{0}_p.$$

Therefore,

$$\Omega_n^{-\frac{1}{2}} \Psi_n \mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \mathbf{I}_p)$$

and finally

$$\Omega_n^{-\frac{1}{2}} \Psi_n \left(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^* \right) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \mathbf{I}_p).$$

A.3 Proof of Theorem 4.2

We have by (11) that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^*) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*)),$$

where $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*) = \lim_{n \rightarrow \infty} \Psi_n(\boldsymbol{\theta}^*) \Omega_n^{-1}(\boldsymbol{\theta}^*) \Psi_n(\boldsymbol{\theta}^*)$. Therefore,

$$\sqrt{n}(\mathbf{M}^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m}) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \mathbf{M}^T \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*) \mathbf{M}).$$

As $\text{rank}(\mathbf{M}) = p$, we have that

$$n(\mathbf{M}^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m})^T (\mathbf{M}^T \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*) \mathbf{M})^{-1} (\mathbf{M}^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m})$$

converges in law to a chi-square distribution with p degrees of freedom. But under H_0 , $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^0) = \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*)$, and thus $\mathbf{W}_n^0(\boldsymbol{\theta}^0)$ converges to a chi-square distribution with p degrees of freedom.

A.4 Proof of Theorem 4.3

A first-order Taylor expansion of $\ell(\boldsymbol{\theta})$ around $\boldsymbol{\theta}^*$ at $\widehat{\boldsymbol{\theta}}_\alpha$ is given by

$$\ell(\widehat{\boldsymbol{\theta}}_\alpha) - \ell(\boldsymbol{\theta}^*) = \left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^*) + o_p(n^{-1/2}).$$

Then the asymptotic distribution of the random variable $\sqrt{n}(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^*)$ matches the asymptotic distribution of the random variable $\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \sqrt{n}(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^*)$ and the result follows.

A.5 Proof of Theorem 4.6

We have by (11) that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^*) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*)),$$

where $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*) = \lim_{n \rightarrow \infty} \Psi_n(\boldsymbol{\theta}^*) \Omega_n^{-1}(\boldsymbol{\theta}^*) \Psi_n(\boldsymbol{\theta}^*)$. Therefore,

$$\sqrt{n}(\mathbf{M}^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m}) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}_p, \mathbf{M}^T \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*) \mathbf{M}).$$

As $\text{rank}(\mathbf{M}) = r$, we have that

$$n(\mathbf{M}^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m})^T (\mathbf{M}^T \boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*) \mathbf{M})^{-1} (\mathbf{M}^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m})$$

converges in law to a chi-square distribution with r degrees of freedom. But $\boldsymbol{\Sigma}_\alpha(\widehat{\boldsymbol{\theta}}_\alpha)$ is a consistent estimator of $\boldsymbol{\Sigma}_\alpha(\boldsymbol{\theta}^*)$, and thus $\mathbf{W}_n^0(\widehat{\boldsymbol{\theta}}_\alpha)$ converges to a chi-square distribution with r degrees of freedom.

A.6 Proof of Theorem 4.7

A Taylor series expansion of $M^T \boldsymbol{\theta} - \mathbf{m}$ around $\boldsymbol{\theta}_n$ yields

$$\begin{aligned} M^T \widehat{\boldsymbol{\theta}}_\alpha - \mathbf{m} &= M^T \boldsymbol{\theta}_n - \mathbf{m} + M^T (\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_n) + o(\|\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_n\|_1) \\ &= n^{-1/2} M^T \mathbf{d} - \mathbf{m} + M^T (\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_n) + o(\|\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_n\|_1). \end{aligned}$$

Using Theorem 6.2, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_n) \xrightarrow[n \rightarrow \infty]{L} N(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$ and $\sqrt{n}o(\|\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_n\|_1) = o_p(1)$, we get the asymptotic convergence

$$\sqrt{n} \left(M^T \widehat{\boldsymbol{\theta}}_\alpha \right) \xrightarrow[n \rightarrow \infty]{L} N(M^T \mathbf{d}, M^T \boldsymbol{\Sigma}_\alpha M).$$

We now consider the random variable $\mathbf{Z} = \sqrt{n} M^T \widehat{\boldsymbol{\theta}}_\alpha (M^T \boldsymbol{\Sigma}_\alpha M)^{-1/2}$ satisfying

$$\mathbf{Z} \xrightarrow[n \rightarrow \infty]{L} N((M^T \boldsymbol{\Sigma}_\alpha M)^{-1/2} M^T \mathbf{d}, \mathbf{I}_{r \times r}).$$

Hence, the asymptotic distribution of the quadratic form $\mathbf{W} = \mathbf{Z}^T \mathbf{Z}$ is given by a non-central chi-square distribution with r degrees of freedom and non-centrality parameter

$$\delta = \mathbf{d}^T M [M^T \boldsymbol{\Sigma}_\alpha (\widehat{\boldsymbol{\theta}}_\alpha) M]^{-1} M^T \mathbf{d}.$$

A.7 Proof of Theorem 5.1

The IF of the functional $\mathbf{T}_\alpha(G_1, \dots, G_n)$ with contamination in the i_0 -th direction will be obtained replacing $\boldsymbol{\theta}_\varepsilon^{i_0}$ and $g_{i_0, \varepsilon}$ in $\boldsymbol{\theta}$ and g_{i_0} respectively in the equality (20), differentiating with respect to ε and evaluating the corresponding equality in $\varepsilon = 0$.

In (20) we replace $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_\varepsilon^{i_0}$ and $g_{i_0}(y)$ by

$$g_{i_0, \varepsilon} = (1 - \varepsilon)g_{i_0}(y) + \varepsilon \Delta_{t_{i_0}}(y),$$

and for $i \neq i_0$ we consider the original $g_i(y)$. We get

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \frac{\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} \mathbf{u}_i(y, \boldsymbol{\theta}_\varepsilon^{i_0}) dy}{\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} dy} - \frac{1}{n} \sum_{i \neq i_0}^n \frac{\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) \mathbf{u}_i(y, \boldsymbol{\theta}_\varepsilon^{i_0}) dy}{\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) dy} \\ & - \frac{\int f_{i_0}(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_{i_0, \varepsilon}(y) \mathbf{u}_{i_0}(y, \boldsymbol{\theta}_\varepsilon^{i_0}) dy}{\int f_{i_0}(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_{i_0, \varepsilon}(y) dy} = \mathbf{0}_p. \end{aligned} \quad (44)$$

Now, we denote

$$\begin{aligned} \zeta_{i, \alpha}(\boldsymbol{\theta}_\varepsilon^{i_0}) &= \frac{\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} \mathbf{u}_i(y, \boldsymbol{\theta}_\varepsilon^{i_0}) dy}{\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} dy}, \\ \zeta_{i, \alpha}^*(\boldsymbol{\theta}_\varepsilon^{i_0}) &= \frac{\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) \mathbf{u}_i(y, \boldsymbol{\theta}_\varepsilon^{i_0}) dy}{\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) dy}, \\ \zeta_{i, \alpha}^{**}(\boldsymbol{\theta}_\varepsilon^{i_0}) &= \frac{\int f_{i_0}(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_{i_0, \varepsilon}(y) \mathbf{u}_{i_0}(y, \boldsymbol{\theta}_\varepsilon^{i_0}) dy}{\int f_{i_0}(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_{i_0, \varepsilon}(y) dy}. \end{aligned}$$

Therefore, (44) can be written as,

$$\frac{1}{n} \sum_{i=1}^n \zeta_{i,\alpha}(\boldsymbol{\theta}_\varepsilon^{i_0}) - \frac{1}{n} \sum_{i \neq i_0}^n \zeta_{i,\alpha}^*(\boldsymbol{\theta}_\varepsilon^{i_0}) - \zeta_{i_0,\alpha}^{**}(\boldsymbol{\theta}_\varepsilon^{i_0}) = \mathbf{0}. \quad (45)$$

Now, we have

$$\begin{aligned} \frac{\partial \zeta_{i,\alpha}(\boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \varepsilon} &= \left(\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} dy \right)^{-2} \left\{ \left[(1+\alpha) \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha \frac{\partial f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \boldsymbol{\theta}_\varepsilon^{i_0}} \frac{\partial \boldsymbol{\theta}_\varepsilon^{i_0}}{\partial \varepsilon} \mathbf{u}_i(y, \boldsymbol{\theta}_\varepsilon^{i_0}) dy \right. \right. \\ &\quad \left. \left. + \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} \frac{\partial \mathbf{u}_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \boldsymbol{\theta}_\varepsilon^{i_0}} \frac{\partial \boldsymbol{\theta}_\varepsilon^{i_0}}{\partial \varepsilon} dy \right] \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} dy \right. \\ &\quad \left. - \left[(1+\alpha) \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha \frac{\partial f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \boldsymbol{\theta}_\varepsilon^{i_0}} \frac{\partial \boldsymbol{\theta}_\varepsilon^{i_0}}{\partial \varepsilon} dy \right] \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha dy \right\}, \end{aligned}$$

and

$$\begin{aligned} \left. \frac{\partial \zeta_{i,\alpha}(\boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \varepsilon} \right|_{\varepsilon=0} &= \frac{IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n)}{\left(\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} dy \right)^2} \\ &\times \left\{ \left[(1+\alpha) \int f_i(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_i^T(y, \boldsymbol{\theta}) \mathbf{u}_i(y, \boldsymbol{\theta}) dy + \int f_i(y, \boldsymbol{\theta})^{\alpha+1} \frac{\partial \mathbf{u}_i(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dy \right] \int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy \right. \\ &\quad \left. - (1+\alpha) \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right) \left(\int f_i(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right)^T \right\} \\ &= \frac{IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n)}{\left(\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} dy \right)^2} \mathbf{A}_{i,\alpha}(\boldsymbol{\theta}), \end{aligned}$$

with

$$\mathbf{A}_{i,\alpha}(\boldsymbol{\theta}) = \left[(1+\alpha) \int f_i(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_i^T(y, \boldsymbol{\theta}) \mathbf{u}_i(y, \boldsymbol{\theta}) dy + \int f_i(y, \boldsymbol{\theta})^{\alpha+1} \frac{\partial \mathbf{u}_i(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dy \right] \int f_i(y, \boldsymbol{\theta})^{\alpha+1} dy.$$

Therefore,

$$\begin{aligned} \left. \frac{\partial}{\partial \varepsilon} \sum_{i=1}^n \zeta_{i,\alpha}(\boldsymbol{\theta}_\varepsilon^{i_0}) \right|_{\varepsilon=0} &= \sum_{i=1}^n \left. \frac{\partial \zeta_{i,\alpha}(\boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \varepsilon} \right|_{\varepsilon=0} \\ &= IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n) \sum_{i=1}^n \frac{\mathbf{A}_{i,\alpha}(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{1+\alpha} dy \right)^2}. \end{aligned}$$

Now,

$$\begin{aligned} \frac{\partial \zeta_{i,\alpha}^*(\boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \varepsilon} &= \left(\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) dy \right)^{-2} \left\{ \left[\alpha \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{\alpha-1} g_i(y) \frac{\partial f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \boldsymbol{\theta}_\varepsilon^{i_0}} \frac{\partial \boldsymbol{\theta}_\varepsilon^{i_0}}{\partial \varepsilon} \mathbf{u}_i(y, \boldsymbol{\theta}_\varepsilon^{i_0}) dy \right. \right. \\ &\quad \left. \left. + \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) \frac{\partial \mathbf{u}_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \boldsymbol{\theta}_\varepsilon^{i_0}} \frac{\partial \boldsymbol{\theta}_\varepsilon^{i_0}}{\partial \varepsilon} dy \right] \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) dy \right. \\ &\quad \left. - \left[\alpha \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^{\alpha-1} g_i(y) \frac{\partial f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \boldsymbol{\theta}_\varepsilon^{i_0}} \frac{\partial \boldsymbol{\theta}_\varepsilon^{i_0}}{\partial \varepsilon} dy \right] \int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha dy \right\}. \end{aligned}$$

This is,

$$\begin{aligned}
& \left. \frac{\partial \zeta_{i,\alpha}^*(\boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \varepsilon} \right|_{\varepsilon=0} = \frac{IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n)}{\left(\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) dy \right)^2} \\
& \times \left\{ \left[\alpha \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i^T(y, \boldsymbol{\theta}) \mathbf{u}_i(y, \boldsymbol{\theta}) dy + \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \frac{\partial \mathbf{u}_i(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dy \right] \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) dy \right. \\
& \left. - \alpha \left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right) \left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right)^T \right\} \\
& = \frac{IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n)}{\left(\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) dy \right)^2} \mathbf{A}_{i,\alpha}^*(\boldsymbol{\theta}),
\end{aligned}$$

with

$$\begin{aligned}
\mathbf{A}_{i,\alpha}^*(\boldsymbol{\theta}) &= \left[\alpha \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i^T(y, \boldsymbol{\theta}) \mathbf{u}_i(y, \boldsymbol{\theta}) dy + \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \frac{\partial \mathbf{u}_i(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dy \right] \int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) dy \\
& \quad - \alpha \left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right) \left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) \mathbf{u}_i(y, \boldsymbol{\theta}) dy \right)^T.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left. \frac{\partial}{\partial \varepsilon} \sum_{i=1}^n \zeta_{i,\alpha}^*(\boldsymbol{\theta}_\varepsilon^{i_0}) \right|_{\varepsilon=0} &= \sum_{i=1}^n \left. \frac{\partial \zeta_{i,\alpha}^*(\boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \varepsilon} \right|_{\varepsilon=0} \\
&= IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n) \sum_{i=1}^n \frac{\mathbf{A}_{i,\alpha}^*(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta}_\varepsilon^{i_0})^\alpha g_i(y) dy \right)^2}.
\end{aligned}$$

In a similar manner,

$$\left. \frac{\partial}{\partial \varepsilon} \zeta_{i_0,\alpha}^{**}(\boldsymbol{\theta}_\varepsilon^{i_0}) \right|_{\varepsilon=0} = \left. \frac{\partial \zeta_{i_0,\alpha}^{**}(\boldsymbol{\theta}_\varepsilon^{i_0})}{\partial \varepsilon} \right|_{\varepsilon=0} = \frac{\boldsymbol{\ell}_{i_0,\alpha}(\boldsymbol{\theta})}{\left(\int f_{i_0}(y, \boldsymbol{\theta})^\alpha g_{i_0}(y) dy \right)^2},$$

with

$$\boldsymbol{\ell}_{i_0,\alpha}(\boldsymbol{\theta}) = f_{i_0}(y, \boldsymbol{\theta}) \int f_{i_0}(y, \boldsymbol{\theta})^{\alpha+1} \mathbf{u}_{i_0}(y, \boldsymbol{\theta}) dy - f_{i_0}(y, \boldsymbol{\theta}) \mathbf{u}_{i_0}(y, \boldsymbol{\theta}) \int f_{i_0}(y, \boldsymbol{\theta})^{\alpha+1} dy.$$

Therefore, equality (45) can be written as

$$\begin{aligned}
& IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n) \left\{ \frac{1}{n} \sum_{i=1}^n \left[\frac{\mathbf{A}_{i,\alpha}(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta})^{1+\alpha} dy \right)^2} - \frac{\mathbf{A}_{i,\alpha}^*(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) dy \right)^2} \right] \right\} \\
& + \frac{\boldsymbol{\ell}_{i_0,\alpha}(\boldsymbol{\theta})}{\left(\int f_{i_0}(y, \boldsymbol{\theta})^\alpha g_{i_0}(y) dy \right)^2} = \mathbf{0}.
\end{aligned}$$

Finally,

$$IF(t_{i_0}, \mathbf{T}_\alpha, G_1, \dots, G_n) = (\mathbf{M}_{n,\alpha}(\boldsymbol{\theta}))^{-1} \frac{-\boldsymbol{\ell}_{i_0,\alpha}(\boldsymbol{\theta})}{\left(\int f_{i_0}(y, \boldsymbol{\theta})^\alpha g_{i_0}(y) dy \right)^2},$$

where

$$\mathbf{M}_{n,\alpha}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \left[\frac{\mathbf{A}_{i,\alpha}(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta})^{1+\alpha} dy \right)^2} - \frac{\mathbf{A}_{i,\alpha}^*(\boldsymbol{\theta})}{\left(\int f_i(y, \boldsymbol{\theta})^\alpha g_i(y) dy \right)^2} \right].$$

A.8 Proof of Lemma 6.1

This proof is very similar to that of [12] (Lemma 6.1).

A.9 Proof of Theorem 6.2

The consistence follows directly from Lemma 6.1 and Theorem 3.1, while the asymptotic distribution is obtained applying Lemma 6.1 and Theorem 3.2 to the MLRM.

References

- [1] Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, **5**(3), 445–463.
- [2] Tamura, R. N., & Boos, D. D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association*, **81**(393), 223–229.
- [3] Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82**(399), 802–807.
- [4] Simpson, D. G. (1989). Hellinger deviance tests: efficiency, breakdown points, and examples. *Journal of the American Statistical Association*, **84**(405), 107–113.
- [5] Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *The Annals of Statistics*, **22**(2), 1081–1114.
- [6] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*. Chapman & Hall/CRC, Boca de Raton.
- [7] Basu, A., Shioya, H., & Park, C. (2011). *Statistical inference: the minimum distance approach*. Chapman & Hall/CRC, Boca de Raton.
- [8] Jones, M.C., Hjort, N.L., Harris, I.R. & Basu, A. (2001). A comparison of related density-based minimum divergence estimators. *Biometrika*, **88**, 865–873.
- [9] Broniatowski, M., Toma, A. & Vajda, I. (2012). Decomposable pseudodistances and applications in statistical estimation, *Journal of Statistical Planning and Inference*, **142**, 2574–2585.
- [10] Castilla, E., Martín N., Muñoz S. & Pardo, L. (2020). Robust Wald-type tests based on Minimum Rényi Pseudodistance Estimators for the Multiple Regression Model. *Journal of Statistical Computation and Simulation*. **90**(14), 2655–2680.
- [11] Castilla, E., Ghosh, A., Jaenada, M. & Pardo, L. (2020). On regularization methods based on Rényi’s pseudodistances for sparse high-dimensional linear regression models. <https://arxiv.org/abs/2007.15929>.

- [12] Ghosh, A., & 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.
- [13] Ghosh, A., & Basu, A. (2018). Robust bounded influence tests for independent non-homogeneous observations. *Statistica Sinica*, **28**(3), 1133–1155.
- [14] Basu, A. , Ghosh, A. , Martin, N. & Pardo, L. (2018). Robust Wald-type tests for non-homogeneous observations based on the minimum density power divergence estimator. *Metrika*, **81**(5) 493–522.
- [15] Chandra, T. K. (1989). Uniform integrability in the Cesàro sense and the weak law of large numbers. *Sankya, Serie A*, **51**, 309–317.
- [16] Feller, W. (1971). *An Introduction to Probability Theory and its applications*. Volume II, 2nd edition. John Wiley & Sons.
- [17] Fanny Leroy, F., Dauxois, J.Y. & Tubert-Bitter, P. (2016). On the Parametric Maximum Likelihood Estimator for Independent but Non-identically Distributed Observations with Application to Truncated Data. *Journal of Statistical Theory and Applications*, **15**(1), 96–107.
- [18] Weisberg, S. *Applied linear regression*. Vol. 528. John Wiley & Sons, 2005.
- [19] Rousseeuw, P. J., & Annick M. L. (2005). *Robust regression and outlier detection*. Vol. 589. John Wiley & Sons.
- [20] Mickey, M. R, Dunn, O. J. & Clark, V. (1967). Note on the use of stepwise regression in detecting outliers. *Computers and Biomedical Research*, **1**, 105–111.