Robust Wald-type tests based on Minimum Rényi Pseudodistance Estimators for the Multiple Linear Regression Model

E. Castilla , N. Martin , S. Muñoz & L. Pardo

Abstract

We introduce a new family of Wald-type tests, based on minimum Rényi pseudodistance estimators, for testing general linear hypotheses and the variance of the residuals in the multiple regression model. The classical Wald test, based on the maximum likelihood estimator, can be seen as a particular case inside our family. Theoretical results, supported by an extensive simulation study, point out how some tests included in this family have a better behaviour, in the sense of robustness, than the Wald test. Finally, we provide a data-driven procedure for the choice of the optimal test given any data set.

Keywords and phrases: Influence function; Minimum density power divergence estimator; Multiple regression model; Rényi Pseudodistance; Robustness.

1 Introduction

Statistical inference based on the minimization of a suitable statistical pseudodistance or divergence is a useful and popular technique that has increased greatly over the last three or four decades. Minimum ϕ-divergence estimators ($M\phi E$) (Pardo, 2006), minimum density power divergence (DPD) estimators (Basu et al., 1998) or minimum Rényi pseudistance (RP) estimators (Jones et al., 2001) are good examples. These estimators have been presented in different statistical models as an alternative to the classical maximum likelihood estimator (MLE), which is known to have good efficiency properties, but not so good robustness properties. With this motivation, Durio and Isaia (2011) studied the minimum DPD estimators for the multiple regression model (MRM). In the cited paper, the robustness of DPD estimators was analyzed from a simulation study, with no theoretical support. Minimum DPD estimators have also been used in order to define Wald-type tests as, for example, in Basu et al. (2016) and Ghosh et al. (2016).

Broniatowski et al. (2012) considered the RP in order to give robust estimators, minimum RP estimators, for the MRM. In this paper we are going to introduce and study for the first time Wald-type tests, for testing linear hypotheses in relation to the regression coefficients in the MRM, based on minimum RP estimators.

A brief review on RP estimators and MRM will be done in Section 2. In Section 3 we present the definition as well as the asymptotic distribution of the Wald-type tests. Their influence function is introduced in Section 4. Section 5 is devoted to present a simulation study considering the models studied in Durio and Isaia (2011) to compare the results they obtained using the minimum DPD estimators and the results we have obtained based on minimum RP estimators. In Section 6 we present a data-driven procedure for the selection of the optimal tuning parameter. The paper ends with a brief concluding remark in Section 7. The proofs of the results stated in Sections 3 and 6 are given in Appendix A.

2 Minimum RP estimators

Let $X_1, \ldots, X_n$ be a random sample from a population having true density g which is being modeled by a parametric family of densities $f_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$. The RP between the densities g and $f_{\boldsymbol{\theta}}$ is given by

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

for $\alpha > 0$, whereas for $\alpha = 0$ it is given by

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

i.e., the Kullback-Leibler divergence, $D_{Kullback}(g, f_{\boldsymbol{\theta}})$, between g and $f_{\boldsymbol{\theta}}$ (see Pardo, 2006). In Broniatowski et al. (2012) it was established that $R_\alpha(g, f_{\boldsymbol{\theta}}) \geq 0$, with $R_\alpha(g, f_{\boldsymbol{\theta}}) = 0$ if and only if $f_{\boldsymbol{\theta}} = g$.

The minimum RP estimator is obtained by minimizing the RP, $R_\alpha(\widehat{g}, f_{\boldsymbol{\theta}})$, with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where $\widehat{g}$ is an empirical estimator of g based on the available data. The advantage of minimum RP estimators in relation with some minimum pseudodistances or divergence estimators is that the term

$$\frac{1}{\alpha(\alpha+1)} \log\left(\int \widehat{g}(x)^{\alpha+1} dx\right)$$

in (1) can be neglected because it does not depend on $\boldsymbol{\theta}$. Moreover, the integral $\int f_{\boldsymbol{\theta}}(x)^{\alpha} g(x) dx$, can be written as

$$\int f_{\boldsymbol{\theta}}(x)^{\alpha} dG(x),$$

being G the distribution function associated to the density function g. Therefore, it is only necessary to estimate G in order to get an estimator based on the minimization of $R_\alpha(g, f_{\boldsymbol{\theta}})$, and we can do this by using the empirical distribution function, G_n, associated to the random sample $X_1, \ldots, X_n$. This leads to the objective function given by

$$\frac{1}{\alpha+1} \log \int f_{\boldsymbol{\theta}}(x)^{\alpha+1} dx - \frac{1}{\alpha} \log \frac{1}{n} \sum_{i=1}^{n} f_{\boldsymbol{\theta}}(X_i)^{\alpha}.$$

The minimum RP estimator, $\widehat{\boldsymbol{\theta}}_\alpha$, can then be defined by

$$\widehat{\boldsymbol{\theta}}_\alpha = \begin{cases} \arg\sup_{\boldsymbol{\theta}} \left\{ -\frac{1}{\alpha+1} \log \int f_{\boldsymbol{\theta}}(x)^{\alpha+1} dx + \frac{1}{\alpha} \log \frac{1}{n} \sum\limits_{i=1}^{n} f_{\boldsymbol{\theta}}(X_i)^{\alpha} \right\} & \text{if} \quad \alpha > 0 \\ \arg\sup_{\boldsymbol{\theta}} \frac{1}{n} \sum\limits_{i=1}^{n} \log f_{\boldsymbol{\theta}}(X_i) & \text{if} \quad \alpha = 0 \end{cases}.$$

We can observe that for $\alpha = 0$ the MLE is obtained.

It is not difficult to establish that for $\alpha > 0$ the minimum RP estimators can be written as

$$\widehat{\boldsymbol{\theta}}_\alpha = \arg\sup_{\boldsymbol{\theta}} \sum_{i=1}^{n} \frac{f_{\boldsymbol{\theta}}(X_i)^{\alpha}}{C_\alpha(\boldsymbol{\theta})},$$

where

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

For more details see Jones et al. (2001) and Broniatowski et al. (2012).

It is interesting to note that in statistical information theory there is a family with the name "Rényi divergence" that we should not confuse with the RP considered in this paper. The Rényi divergence was introduced by Rényi (1961) depending on a tuning parameter $\alpha > 0$. Later, Liese and Vajda (1987) generalized the Rényi divergence to all the tuning parameters $\alpha \in \mathbb{R}$.

2.1 Minimum RP estimators in the Multiple Regression Model

In order to compute the minimum RP estimators for the MRM, let us follow the notation presented in Durio and Isaia (2011).

Let $(X_{i1}, ..., X_{ip}, Y_i)$, $\;i = 1, ..., n$, be $(p+1)$-dimensional independent and identically distributed (i.i.d.) random variables verifying the condition

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + \varepsilon_i, \tag{2}$$

with $\boldsymbol{X}_i^T = (X_{i1}, ..., X_{ip})$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ and $\varepsilon_i' s$ are i.i.d. normal random variables with mean zero and variance σ^2 and independent of the $\boldsymbol{X}_i$. The $n \times p$ matrix with elements X_{ij} will be denoted by $\mathbb{X}$, i.e., $\mathbb{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_n)^T$. We often use for (2) the matrix and vector notation

$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3}$$

with $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$.

Proposition 1 (Broniatowski et al., 2012) *Let us consider the MRM in (3). The minimum RP estimators, $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\sigma}_\alpha$, for $\boldsymbol{\beta}$ and σ are defined by*

$$\left(\widehat{\boldsymbol{\beta}}_\alpha, \widehat{\sigma}_\alpha\right) = \begin{cases} \arg\max_{\boldsymbol{\beta},\sigma} \sum\limits_{i=1}^{n} \sigma^{-\frac{\alpha}{\alpha+1}} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2\right) & \text{if } \alpha > 0 \\ \arg\max_{\boldsymbol{\beta},\sigma} \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}\|_2^2}{2\sigma^2}\right) & \text{if } \alpha = 0 \end{cases},$$

where $\|\boldsymbol{Y} - \mathbb{X}\boldsymbol{\beta}\|_2^2 = \sum_{i=1}^n (Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta})^2$ denotes the (squared) l_2-norm, and can be obtained by solving the system

$$\begin{cases} \sum\limits_{i=1}^{n} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2\right)\left(\frac{Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)\boldsymbol{X}_i = \boldsymbol{0}_p \\ \sum\limits_{i=1}^{n} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2\right)\left\{\left(\frac{Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2 - \frac{1}{1+\alpha}\right\} = 0 \end{cases}. \tag{4}$$

Remark 2 *If $\alpha = 0$ in (4), we get the system necessary to get the MLE of $\boldsymbol{\beta}$ and σ^2, whose well-known solution is given by*

$$\widehat{\boldsymbol{\beta}}_0 = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{Y} \quad \textit{and} \quad \widehat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_0\right)^2. \tag{5}$$

In the following, the MLE will be denoted by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$, respectively.

In Broniatowski et al. (2012) it was also established that

$$\sqrt{n}\left(\left(\widehat{\boldsymbol{\beta}}_\alpha, \widehat{\sigma}_\alpha\right)^T - (\boldsymbol{\beta}, \sigma)^T\right) \underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(\begin{pmatrix} \boldsymbol{0}_p \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_\alpha(\sigma) & 0 \\ \boldsymbol{0}_p^T & \frac{\sigma^2(\alpha+1)^3\left(3\alpha^2+4\alpha+2\right)}{4(2\alpha+1)^{5/2}} \end{pmatrix}\right), \tag{6}$$

where

$$\boldsymbol{\Sigma}_\alpha(\sigma) = \sigma^2 \frac{(\alpha+1)^3}{(2\alpha+1)^{\frac{3}{2}}} E\left[\mathbb{X}^T\mathbb{X}\right]^{-1}, \tag{7}$$

i.e., $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\sigma}_\alpha$ are asymptotically independent. A consistent estimator of $\boldsymbol{\Sigma}_\alpha(\sigma)$ is given by

$$\boldsymbol{\Sigma}_\alpha(\widehat{\sigma}_\alpha) = \widehat{\sigma}_\alpha^2 \frac{(\alpha+1)^3}{(2\alpha+1)^{\frac{3}{2}}} \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}. \tag{8}$$

In the next section we will introduce Wald-type test statistics for testing linear hypotheses about $\boldsymbol{\beta}$ based on $\widehat{\boldsymbol{\beta}}_\alpha$, as well as for testing a simple null hypothesis in relation to σ and based on $\widehat{\sigma}_\alpha$.

3 Wald-type test statistics based on minimum RP estimators

3.1 Wald-type tests for general linear hypotheses on $\boldsymbol{\beta}$

We are interested in testing the general linear hypotheses on $\boldsymbol{\beta}$ assuming σ^2 to be unknown,

$$H_0 : \boldsymbol{L}^T\boldsymbol{\beta} = \boldsymbol{c}, \tag{9}$$

where $\boldsymbol{L}$ is a $p \times r$ hypothesis matrix of full row-rank with $r \leq p$ and the right-hand-side $\boldsymbol{c}$ is a constant vector that can be equal to $\boldsymbol{c} = \boldsymbol{0}_r$ in many situations. In order to solve the testing problem given in (9) we are going to introduce a family of Wald-type test statistics based on minimum RP estimators.

Definition 3 *Let $\widehat{\boldsymbol{\beta}}_\alpha$ be the minimum RP estimator of $\boldsymbol{\beta}$. The family of Wald-type test statistics for testing the null hypothesis given in (9) given by:*

$$W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right) = \frac{(2\alpha+1)^{3/2}}{\widehat{\sigma}_\alpha^2\,(\alpha+1)^3}\left(\boldsymbol{L}^T\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{c}\right)^T\left(\boldsymbol{L}^T\left(\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\boldsymbol{L}\right)^{-1}\left(\boldsymbol{L}^T\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{c}\right). \tag{10}$$

Theorem 4 *The asymptotic distribution of the Wald-type test statistics $W_n(\widehat{\boldsymbol{\beta}}_\alpha)$, defined in (10), under the null hypothesis (9), is a chi-square distribution with r degrees of freedom, i.e.,*

$$W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right) \underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}} \chi_r^2,$$

where χ_r^2 represents the chi-square distribution with r degrees of freedom.

Based on Theorem 4, the null hypothesis given in (9) will be rejected if we have that

$$W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right) > \chi_{r,\tau}^2, \tag{11}$$

with $\chi_{r,\tau}^2$ verifying $\Pr\left(\chi_r^2 > \chi_{r,\tau}^2\right) = \tau$.

Remark 5 *If we consider*

$$\boldsymbol{L}^T = \boldsymbol{I}_{p\times p}, \tag{12}$$

we have

$$\boldsymbol{L}^T\boldsymbol{\beta} = \boldsymbol{0}_p$$

if and only if $\beta_i = 0$, $i = 1, \ldots, p$. Therefore, for testing

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$$

we can consider the Wald-type test statistics with $\boldsymbol{L}^T$ defined in (12). In this case the asymptotic distribution of the Wald type test statistics is a chi-square distribution with p degrees of freedom. If we consider $\boldsymbol{L}^T = (0, .., 0, 1^{(i}, 0, .., 0)$, we can test

$$H_0 : \beta_i = 0.$$

Now we consider $\boldsymbol{\beta}^* \in \Theta$ verifying that $\boldsymbol{L}^T\boldsymbol{\beta}^* \neq \boldsymbol{c}$, i.e., $\boldsymbol{\beta}^*$ does not belong to the null hypothesis. We denote

$$l_{\boldsymbol{\beta}}(\sigma) = \left(\boldsymbol{L}^T\boldsymbol{\beta} - \boldsymbol{c}\right)^T\left(\boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha(\sigma)\,\boldsymbol{L}\right)^{-1}\left(\boldsymbol{L}^T\boldsymbol{\beta} - \boldsymbol{c}\right)$$

and

$$l_{\boldsymbol{\beta}}(\widehat{\sigma}_\alpha) = \left(\boldsymbol{L}^T\boldsymbol{\beta} - \boldsymbol{c}\right)^T\left(\boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha(\widehat{\sigma}_\alpha)\,\boldsymbol{L}\right)^{-1}\left(\boldsymbol{L}^T\boldsymbol{\beta} - \boldsymbol{c}\right).$$

Theorem 6 provides an approximation to the power function for the Wald-type test statistics given in (11).

Theorem 6 *Let $\boldsymbol{\beta}^* \in \Theta$ such that $\boldsymbol{L}^T\boldsymbol{\beta}^* \neq \boldsymbol{c}$ be the true value of the parameter, so that $\widehat{\boldsymbol{\beta}}_\alpha \underset{n\to\infty}{\overset{P}{\longrightarrow}} \boldsymbol{\beta}^*$. The power function of the Wald-type test statistics given in (11) in $\boldsymbol{\beta}^*$, is given by*

$$\pi_{W_{n,\beta}^{\alpha}}\left(\boldsymbol{\beta}^*\right) = 1 - \phi_n\left(\frac{1}{\sigma\left(\boldsymbol{\beta}^*\right)}\left(\frac{\chi^2_{r,\tau}}{\sqrt{n}} - \sqrt{n}l_{\boldsymbol{\beta}^*}(\sigma)\right)\right), \tag{13}$$

where $\phi_n(x)$ tends uniformly to the standard normal distribution $\phi(x)$ and $\sigma(\boldsymbol{\beta}^*)$ is given by

$$\sigma^2\left(\boldsymbol{\beta}^*\right) = \left.\frac{\partial l_{\boldsymbol{\beta}}(\sigma)}{\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \boldsymbol{\Sigma}_\alpha(\sigma) \left.\frac{\partial l_{\boldsymbol{\beta}}(\sigma)}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}.$$

It is clear that

$$\lim_{n\to\infty} \pi_{W_{n,\beta}^{\alpha}}\left(\boldsymbol{\beta}^*\right) = 1$$

for all $\alpha \in (0,1)$. Therefore, the Wald-type test statistics are consistent in the sense of Fraser (1957).

Remark 7 *Based on the previous theorem we can obtain the sample size necessary to get a fix power π_0, i.e, $\pi_{W_{n,\beta}^{\alpha}}(\boldsymbol{\beta}^*) = \pi_0$. By (13) we must solve the equation*

$$1 - \pi_0 = \phi\left(\frac{1}{\sigma\left(\boldsymbol{\beta}^*\right)}\left(\frac{\chi^2_{r,\tau}}{\sqrt{n}} - \sqrt{n}l_{\boldsymbol{\beta}^*}(\sigma)\right)\right),$$

and we get that $n = [n^*] + 1$, where

$$n^* = \frac{A + B + \sqrt{A(A+2B)}}{2\, l_{\boldsymbol{\beta}^*}(\sigma)^2},$$

$$A = \sigma^2\left(\boldsymbol{\beta}^*\right)\left(\phi^{-1}\left(1-\pi_0\right)\right)^2 \text{ and } B = \frac{1}{2}\chi^2_{r,\tau} l_{\boldsymbol{\beta}^*}(\sigma).$$

We can also find an approximation of the power of the Wald-type test statistics, $W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right)$, given in (11) at a contiguous alternative hypothesis close to the null hypothesis. Let $\boldsymbol{\beta}_n$ such that $\boldsymbol{L}^T\boldsymbol{\beta_n} \neq \boldsymbol{c}$ be a given alternative and let $\boldsymbol{\beta}_0$ be the closest element to $\boldsymbol{\beta}_n$ in the Euclidean distance sense such that $\boldsymbol{L}^T\boldsymbol{\beta_0} = \boldsymbol{c}$. A first possibility to introduce contiguous alternative hypotheses is to consider a fixed $\boldsymbol{d} \in \mathbb{R}^d$ and to permit $\boldsymbol{\beta}_n$ moving towards $\boldsymbol{\beta}_0$ as n increases, in the following way

$$H_{1,n} : \boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{d}. \tag{14}$$

A second approach is to relax the condition $\boldsymbol{L}^T\boldsymbol{\beta_0} = \boldsymbol{c}$ defining the null hypothesis. Let $\boldsymbol{\delta} \in \mathbb{R}^r$ and consider the following sequence, $\boldsymbol{\beta}_n$, of parameters moving towards $\boldsymbol{\beta}_0$ according to

$$H^*_{1,n} : \boldsymbol{L}^T\boldsymbol{\beta_n} - \boldsymbol{c} = n^{-1/2}\boldsymbol{\delta}. \tag{15}$$

Note that

$$\boldsymbol{L}^T\boldsymbol{\beta}_n - \boldsymbol{c} = \boldsymbol{L}^T\boldsymbol{\beta}_0 + \boldsymbol{L}^T n^{-1/2}\boldsymbol{d} - \boldsymbol{c} = n^{-1/2}\boldsymbol{L}^T\boldsymbol{d}.$$

Then the equivalence between the two hypotheses, given in (14) and (15), is given by

$$\boldsymbol{L}^T\boldsymbol{d} = \boldsymbol{\delta}.$$

If we denote by $\chi^2_r(\Delta)$ the non central chi-square distribution with r degrees of freedom and noncentrality parameter Δ, we can state the following theorem:

Theorem 8 *The following properties hold:*

i) $W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right) \underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}} \chi_r^2(\Delta_1)$*under* $H_{1,n}$ *given in (14), with*

$$\Delta_1 = \boldsymbol{d}^T\boldsymbol{L}\left[\boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha(\sigma)\boldsymbol{L}\right]^{-1}\boldsymbol{L}^T\boldsymbol{d}.$$

ii) $W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right) \underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}} \chi_r^2(\Delta_2)$ *under* $H_{1,n}^*$ *given in (15), with*

$$\Delta_2 = \boldsymbol{\delta}^T\left(\boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha(\sigma)\boldsymbol{L}\right)^{-1}\boldsymbol{\delta}.$$

3.2 Wald type tests for testing a simple null hypothesis about σ

We can also consider the problem of testing

$$H_0 : \sigma = \sigma_0 \text{ versus } H_1 : \sigma \neq \sigma_0 \tag{16}$$

Definition 9 *The Wald-type test statistics,* $W_n(\widehat{\sigma}_\alpha)$*, based on the minimum RP estimator,* $\widehat{\sigma}_\alpha$*, for testing (16) is given by*

$$W_n(\widehat{\sigma}_\alpha) = \frac{n}{l(\alpha)}\left(\frac{\widehat{\sigma}_\alpha - \sigma_0}{\sigma_0}\right)^2,$$

being

$$l(\alpha) = \frac{(\alpha+1)^3\left(3\alpha^2+4\alpha+2\right)}{4(2\alpha+1)^{5/2}}.$$

Theorem 10 *The asymptotic distribution of the Wald-type test statistics,* $W_n(\widehat{\sigma}_\alpha)$, *is a chi-square distribution with one degree of freedom.*

Corollary 11 *Based on Theorem 10 we reject the null hypothesis* $H_0 : \sigma = \sigma_0$*, if*

$$W_n(\widehat{\sigma}_\alpha) > \chi^2_{1,\tau}. \tag{17}$$

An approximation to the power function in σ^* is presented in Theorem 12.

Theorem 12 *Let* $\sigma^* \neq \sigma_0$ *such that* $\widehat{\sigma}_\alpha \underset{n\longrightarrow\infty}{\overset{\mathcal{P}}{\longrightarrow}} \sigma^*$*. The power function of the Wald-type test statistics given in (17) in* σ^**, has the expression*

$$\pi_{W^\alpha_{n,\sigma}}(\sigma^*) = 1 - \Phi_n\left(\frac{\sigma_0}{\sigma^*}\left(\chi^2_{1,\tau} - \frac{\sqrt{n}}{\sqrt{l(\alpha)}}\left(\frac{\sigma^*-\sigma_0}{\sigma_0}\right)\right)\right), \tag{18}$$

where $\Phi_n(x)$ is a sequence of distributions functions that converge uniformly to the standard normal distribution. We have $\lim_{n\longrightarrow\infty}\pi_{W^\alpha_{n,\sigma}}(\sigma^*) = 1$. Therefore, the Wald-type test statistics for testing (16) are consistent in the sense of Fraser (1957).

Remark 13 *If we consider* $\alpha = 0$ *we get the MLE,* $\widehat{\sigma}$*, of* σ *and the classical Wald-type test is given by*

$$W_n(\widehat{\sigma}) = 2n\left(\frac{\widehat{\sigma}-\sigma_0}{\sigma_0}\right)^2.$$

In a similar way to the previous section we can obtain an approximation to the power functions of the Wald-type test statistics given in (17).

We can also consider the Wald-type test statistics based on the minimum DPD estimators.

Remark 14 *The minimum DPD estimators are obtained (see Basu et al., 1998) by minimizing the DPD given by*

$$d_\gamma\left(g, f_{\boldsymbol{\theta}}\right)=\int f_{\boldsymbol{\theta}}(x)^{\gamma+1} dx-\left(1+\frac{1}{\gamma}\right) \int f_{\boldsymbol{\theta}}(x)^\gamma g(x) dx+\frac{1}{\gamma} \int g(x)^{\gamma+1} dx .$$

In the case of the MRM, Durio and Isaia (2011) obtained that the minimum DPD estimators, $\widehat{\boldsymbol{\beta}}_\gamma$ and $\widehat{\sigma}_\gamma$, for the parameters $\boldsymbol{\beta}$ and σ^2 respectively, are given by

$$\left(\widehat{\boldsymbol{\beta}}_\gamma, \widehat{\sigma}_\gamma\right)=\begin{cases}\arg\max_{\boldsymbol{\beta},\sigma}\left(\sum\limits_{i=1}^n \frac{\gamma+1}{n\gamma} \frac{\sigma^{-\gamma}}{(2\pi)^{\frac{\gamma}{2}}} \exp\left(-\frac{\gamma}{2}\left(\frac{Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2\right)-\frac{1}{\sigma^\alpha\sqrt{(2\pi)^\gamma(1+\gamma)}}\right) & \text{if } \gamma>0 \\ \arg\max_{\boldsymbol{\beta},\sigma} \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|\boldsymbol{Y}-\mathbb{X}\boldsymbol{\beta}\|_2^2}{2\sigma^2}\right) & \text{if } \gamma=0\end{cases},$$

and they verify (see Ghosh and Basu, 2016) that

$$\begin{cases}\sum\limits_{i=1}^n \exp\left(-\frac{\gamma}{2}\left(\frac{Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2\right)\left(\frac{Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)\boldsymbol{X}_i=\mathbf{0}_p \\ \sum\limits_{i=1}^n \exp\left(-\frac{\gamma}{2}\left(\frac{Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2\right)\left\{1-\left(\frac{Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2\right\}=\frac{\gamma}{(1+\gamma)^{3/2}}\end{cases}.$$

For $\gamma=0$, we get the MLE presented in (5).

In Ghosh and Basu (2016) it was established that

$$\sqrt{n}\left(\left(\widehat{\boldsymbol{\beta}}_\gamma, \widehat{\sigma}_\gamma\right)^T-(\boldsymbol{\beta},\sigma)^T\right) \underset{n\longrightarrow\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}\left(\begin{pmatrix}\mathbf{0}_p \\ 0\end{pmatrix},\begin{pmatrix}\boldsymbol{\Sigma}_\gamma(\sigma) & \mathbf{0}_p^T \\ \mathbf{0}_p & \frac{\sigma^2(1+\gamma)^2}{(2+\gamma^2)^2}\left(\frac{2\left(1+2\gamma^2\right)(1+\gamma)^3}{(1+2\gamma)^{\frac{5}{2}}}-\gamma^2\right)\end{pmatrix}\right),$$

being

$$\boldsymbol{\Sigma}_\gamma(\sigma)=\boldsymbol{E}\left[\mathbb{X}\mathbb{X}^T\right]^{-1}\sigma^2\frac{(1+\gamma)^3}{(1+2\gamma)^{\frac{3}{2}}}.$$

The Wald-type test statistics for testing (9) and (16), based on the minimum DPD estimators are given, respectively, by

$$W_n(\widehat{\boldsymbol{\beta}}_\gamma)=\frac{1}{\widehat{\sigma}_\gamma^2}\frac{(1+2\gamma)^{\frac{3}{2}}}{(1+\gamma)^3}\left(\boldsymbol{L}^T\widehat{\boldsymbol{\beta}}_\gamma-\boldsymbol{c}\right)^T\left(\boldsymbol{L}^T\left(\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}\boldsymbol{L}\right)^{-1}\left(\boldsymbol{L}^T\widehat{\boldsymbol{\beta}}_\gamma-\boldsymbol{c}\right) \tag{19}$$

$$W_n\left(\widehat{\sigma}_\gamma\right)=\frac{n}{l(\gamma)}\left(\frac{\widehat{\sigma}_\gamma-\sigma_0}{\sigma_0}\right)^2, \tag{20}$$

being

$$l(\gamma)=\frac{(1+\gamma)^2}{(2+\gamma^2)^2}\left(\frac{2\left(1+2\gamma^2\right)(1+\gamma)^3}{(1+2\gamma)^{\frac{5}{2}}}-\gamma^2\right).$$

4 Robustness of the Wald-type tests based on the minimum RP estimators

Let $\boldsymbol{T}$ and S be the statistical functionals corresponding to the minimum RP estimators estimators $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\sigma}_\alpha$. For a given density function $g(\boldsymbol{x},y)$ corresponding to the random vector $(\boldsymbol{X},Y)$ with $g(y/\boldsymbol{x})\equiv N(\boldsymbol{x}^T\boldsymbol{\beta},\sigma^2)$. The functional $\boldsymbol{T}$ and S are defined through the solution of the systems of equations

$$\int \phi_1\left(\frac{y-\boldsymbol{x}^T\boldsymbol{T}(G)}{S(G)}\right)\boldsymbol{x}^T dG(\boldsymbol{x},y)=\mathbf{0}_p$$
$$\int \phi_2\left(\frac{y-\boldsymbol{x}^T\boldsymbol{T}(G)}{S(G)}\right)\boldsymbol{x}^T dG(\boldsymbol{x},y)=0,$$

being $G(\boldsymbol{x}, y)$ the distribution function associated to the density function $g(\boldsymbol{x},y)$ and

$$\begin{aligned}\phi_1(u) &= \exp\left(-\tfrac{\alpha}{2}u^2\right)u\\ \phi_2(u) &= (u^2 - \tfrac{1}{\alpha+1})\exp\left(-\tfrac{\alpha}{2}u^2\right).\end{aligned}$$

Considering a contaminated model

$$G_{\varepsilon,\boldsymbol{x}_0,y_0}(\boldsymbol{x},y) = (1-\varepsilon)G(\boldsymbol{x},y) + \varepsilon\Delta_{(\boldsymbol{x}_0,y_0)},$$

where $(\boldsymbol{x}_0, y_0)$ is an arbitrary point in $\mathbb{R}^p \times \mathbb{R}$, in Broniatoski et al. (2012) it was established that the influence functions of the functional associated to the minimum RP estimators of $\boldsymbol{\beta}$ and σ are

$$IF(\boldsymbol{x}_0,y_0;\mathbf{T},G) = \sigma(\alpha+1)^{3/2}\exp\left(-\frac{\alpha}{2}\left(\frac{y_0-\boldsymbol{x}_0^T\boldsymbol{\beta}}{\sigma}\right)^2\right)\left(\frac{y_0-\boldsymbol{x}_0^T\boldsymbol{\beta}}{\sigma}\right)E\left[\mathbb{X}_0\mathbb{X}_0^T\right]^{-1}\boldsymbol{x}_0^T$$

and

$$IF(\boldsymbol{x}_0,y_0;S,G) = \frac{(\alpha+1)^{5/2}}{2}\exp\left(-\frac{\alpha}{2}\left(\frac{y_0-\boldsymbol{x}_0^T\boldsymbol{\beta}}{\sigma}\right)^2\right)\left[\left(\frac{y_0-\boldsymbol{x}_0^T\boldsymbol{\beta}}{\sigma}\right)^2 - \frac{1}{\alpha+1}\right].$$

Let us consider the Wald-type tests $W_n(\widehat{\sigma}_\alpha)$ for testing $H_0: \sigma = \sigma_0$. The functional associated with the Wald-type test statistics, $W_n(\widehat{\sigma}_\alpha)$, evaluated at G is given (ignoring the multiplier n) by

$$W_n(G) = l_\alpha(\sigma_0)(S(G)-\sigma_0)^2,$$

being

$$l_\alpha(\sigma_0) = \frac{4(2\alpha+1)^{5/2}}{\sigma_0^2(\alpha+1)^3(3\alpha^2+4\alpha+2)}.$$

Let

$$G_{\varepsilon,\boldsymbol{x}_0,y_0}(\boldsymbol{x},y) = (1-\varepsilon)G(\boldsymbol{x},y) + \varepsilon\Delta_{(\boldsymbol{x}_0,y_0)}$$

be the ε-contaminated distribution of G with respect to the point mass distribution $\Delta_{(\boldsymbol{x}_0,y_0)}$. The influence function of $W_n^\alpha(\widehat{\sigma}_\alpha)$ is defined as

$$IF(\boldsymbol{x}_0,y_0;W_n^\alpha,G) = \left(\frac{\partial W_n^\alpha(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon}\right)_{\varepsilon=0}.$$

We have

$$\frac{\partial W_n^\alpha(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon} = l_\alpha(\sigma_0)\,2\,(S(G_{\varepsilon,\boldsymbol{x}_0,y_0})-\sigma_0)\frac{\partial S(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon}$$

and

$$\left(\frac{\partial W_n^\alpha(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon}\right)_{\varepsilon=0} = l_\alpha(\sigma_0)\,2\,(S(G)-\sigma_0)\left(\frac{\partial S(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon}\right)_{\varepsilon=0}.$$

But under the null hypothesis $H_0: \sigma = \sigma_0$, we have $S(G) = \sigma_0$. Therefore, $IF(\boldsymbol{x}_0,y_0;W_n^\alpha,G) = 0$, which shows that the influence function analysis based on the first derivative of $W_n^\alpha(G_{\varepsilon,\boldsymbol{x}_0,y_0})$ is not adequate to quantify the robustness of the Wald-type tests for testing $H_0: \sigma = \sigma_0$ and we must use the second order influence function. The second order influence function of the Wald-type test statistics $W_n^\alpha(\widehat{\sigma}_\alpha)$ is given by

$$IF_2(\boldsymbol{x}_0,y_0;W_n^\alpha,G) = \left(\frac{\partial^2 W_n^\alpha(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon^2}\right)_{\varepsilon=0}.$$

We have

$$\frac{\partial^2 W_n^\alpha(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon^2} = 2l_\alpha(\sigma_0)\left(\left(\frac{\partial S(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon}\right)^2 + (S(G)-\sigma_0)\frac{\partial^2 S(G_{\varepsilon,\boldsymbol{x}_0,y_0})}{\partial\varepsilon^2}\right)$$

and

$$IF_2(\boldsymbol{x}_0,y_0;W_n^\alpha,G) = 2\,l_\alpha(\sigma_0)\,IF(\boldsymbol{x}_0,y_0;S,G)^2.$$

The functional associated with the Wald-type test statistics $W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right)$ evaluated at G, is given (ignoring the multiplier n) by

$$W_n(G) = (\boldsymbol{L}^T\boldsymbol{T}(G) - \boldsymbol{c})^T(\boldsymbol{L}^t\boldsymbol{\Sigma}\,(\boldsymbol{T}(G)\boldsymbol{L})^{-1}\,(\boldsymbol{L}^T\boldsymbol{T}(G) - \boldsymbol{c}).$$

For $G_\varepsilon = (1-\varepsilon)\,G + \varepsilon\Delta_{(\boldsymbol{x},y)}$, we have

$$\begin{aligned}\frac{\partial W_n(G_\varepsilon)}{\partial\varepsilon} =& 2\left(\boldsymbol{L}^T(G_\varepsilon\right) - \boldsymbol{c})\left(\boldsymbol{L}^t\boldsymbol{\Sigma}\left(\boldsymbol{T}(G_\varepsilon)\right)\boldsymbol{L}\right)^{-1}\boldsymbol{L}^T\frac{\partial\boldsymbol{T}(G_\varepsilon)}{\partial\varepsilon}\\ &+\left(\boldsymbol{L}^T(G_\varepsilon\right) - \boldsymbol{c})^T\frac{\partial\left(\boldsymbol{L}^t\boldsymbol{\Sigma}\left(\boldsymbol{T}(G_\epsilon)\right)\boldsymbol{L}\right)^{-1}}{\partial\varepsilon}\left(\boldsymbol{L}^T(G_\varepsilon\right) - \boldsymbol{c}).\end{aligned}$$

For $\varepsilon = 0$ we have $G_\varepsilon = G$, $\boldsymbol{T}(G) = \boldsymbol{\beta}$ y $\boldsymbol{L}^T\boldsymbol{T}(G) = \boldsymbol{c}$ and $\boldsymbol{L}^T\boldsymbol{T}(G) - \boldsymbol{c} = \boldsymbol{0}.$ Therefore,

$$\left(\frac{\partial W_n(G_\varepsilon)}{\partial\varepsilon}\right)_{\varepsilon=0} = 0$$

and in this case it is necessary to consider the second order influence function,

$$IF_2\left(\boldsymbol{x}_0,y_0;W_n^\alpha,G\right) = \left(\frac{\partial^2 W_n^\alpha\left(G_{\varepsilon,\boldsymbol{x}_0,y_0}\right)}{\partial\varepsilon^2}\right)_{\varepsilon=0} = IF\left(\boldsymbol{x}_0,y_0;T,G\right)^T\left(\boldsymbol{L}^t\boldsymbol{\Sigma}\left(\boldsymbol{T}(G)\right)\boldsymbol{L}\right)^{-1}IF\left(\boldsymbol{x}_0,y_0;\boldsymbol{T},G\right).$$

In Figure 1 we present the plots of the influence functions associated to the minimum RP estimators $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\sigma}_\alpha$ (top left and medium left panel) and the influence functions associated to their corresponding Wald-type test statistics (top right and medium right panel). Here, $u = y_0 - \boldsymbol{x}_0^T\boldsymbol{\beta}$ represents the perturbation and $\sigma_0 = 1$. For the influence functions of the parameters $\widehat{\boldsymbol{\beta}}_\alpha$ we have considered a simple linear regression model with $x_0 = 1$ and $V_x = 1$. The non-robustness of the MLE is clearly reflected in the plots. Gross error sensitivities (measured as the maximum of the corresponding influence functions) are plotted in bottom panel of Figure 1. While the gross error sensitivity of the influence function functional associated to σ attains in this case its minimum value at $\alpha \simeq 0.8$ (Broniatoski et al., 2012); the corresponding gross error sensitivity associated to its Wald-type test decreases with α, implying that the robustness increases.

5 Simulation Study

For the simulation study performed throughout this section we will consider $R = 1,000$ simulated samples.

5.1 Efficiency and robustness of the Minimum RP and DPD estimators

We provide a simulation study on the behaviour of the minimum RP estimators, $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\sigma}_\alpha$, in the MRM. Minimum DPD estimators, $\widehat{\boldsymbol{\beta}}_\gamma$ and $\widehat{\sigma}_\gamma$, studied in Basu et al. (1998), are also considered. The robustness is evaluated, for both RP and DPD estimators, in the same theoretical models presented by Durio and Isaia (2011).

Scheme 1

Consider the MRM model

$$Y_i = \boldsymbol{X}_i^T\boldsymbol{\beta} + \varepsilon_i,$$

where $i \in \{1,\ldots,n\}$, $n = 600$, the errors $\varepsilon_i's$ are assumed to be i.i.d. normal with mean zero and variance $\sigma^2 = 0.01$, $\boldsymbol{\beta} = (\beta_1^0, \beta_2^0)^T = (0.5, 0.5)^T$ are the regression coefficients and $\boldsymbol{X}_i^T = (X_{i1}, X_{i2})^T$ denotes the *ith* observation for the covariates. Here, we assume X_{1i}, X_{2i} to be random uniform variables in the interval $(0,1)$ so that $Y_i \sim \mathcal{N}(0.5X_{1i} + 0.5X_{2i}, 0.01)$ for each i.

The root of the mean square error (RMSE) of the corresponding RP and DPD estimators are computed based on the R such simulated samples for different values of the tuning parameters α and γ, as can be seen in Tables 1 and 2. A more detailed explanation is given in the pseudo-code of Algorithm 1 for RP estimators (see Remark 14 for details about DPD estimators).

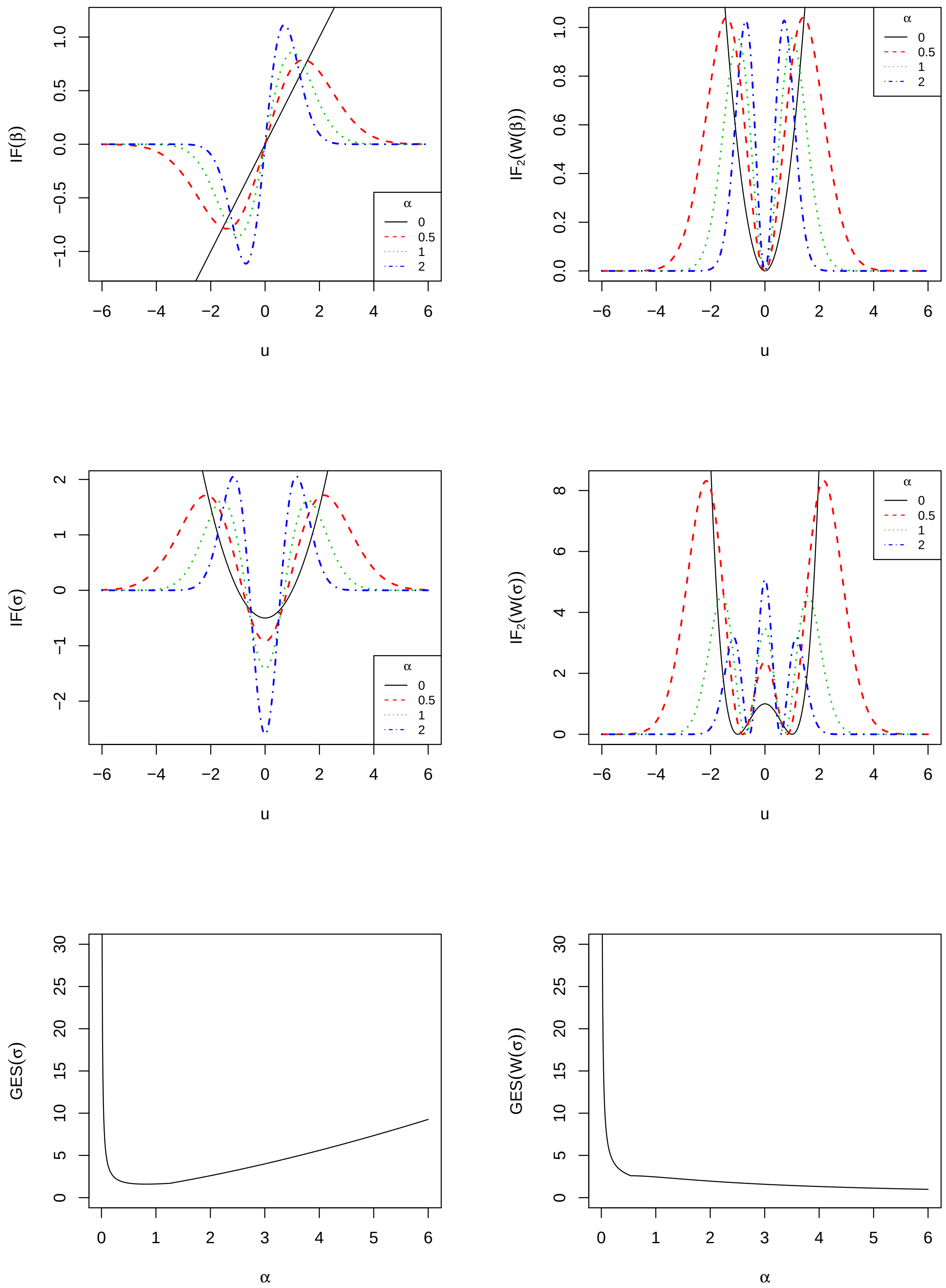

Figure 1: Influence functions and gross error sensitivities for RP-estimators and their corresponding Wald-type tests. Simple regression model with $\sigma_0 = 1$, $x_0 = 1$ and $V_x = 1$.

Algorithm 1 Computation of RMSEs of $\boldsymbol{\beta}$ and σ for the tuning parameter α in Scheme 1

Input: R, n, $\boldsymbol{\beta} = (\beta_1^0, \beta_2^0)^T$, σ, α .

for $j \leftarrow 1$ to R **do**

  for $i \leftarrow 1$ to n **do**

    Generate $X_{i1}^{(j)}, X_{i2}^{(j)} \sim \mathcal{U}(0,1)$, $\varepsilon_i^{(j)} \sim \mathcal{N}(0, \sigma^2)$

    $Y_i^{(j)} \leftarrow \beta_1^0 X_{i1}^{(j)} + \beta_2^0 X_{i2}^{(j)} + \varepsilon_i^{(j)}$

  end for

  Compute $\widehat{\boldsymbol{\beta}}_\alpha^{(j)} = (\widehat{\beta}_{\alpha,1}^{(j)}, \widehat{\beta}_{\alpha,2}^{(j)})^T$ and $\widehat{\sigma}_\alpha^{(j)}$ as defined in Proposition 1 or Remark 2.

end for

$RMSE_\alpha(\sigma) \leftarrow \left[\frac{1}{R}\sum_{j=1}^R (\widehat{\sigma}_\alpha^{(j)} - \sigma)^2\right]^{1/2}$

$RMSE_\alpha(\boldsymbol{\beta}) \leftarrow \left[\frac{1}{2R}\sum_{j=1}^R \{(\widehat{\beta}_{\alpha,1}^{(j)} - \beta_1^0)^2 + (\widehat{\beta}_{\alpha,2}^{(j)} - \beta_2^0)^2\}\right]^{1/2}$

return $RMSE_\alpha(\boldsymbol{\beta})$, $RMSE_\alpha(\sigma)$

Scheme 2

In order to study and compare the robustness of the RP and DPD estimators, $n_{out} \in \{30, 60, 120\}$ over the total $n = 600$ observations in each simulated data set are generated through the regression coefficients $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_1^0, \tilde{\beta}_2^0)^T = (0.7, 0.7)^T$, i.e., there is a 5%, 10% and 20% of outliers, respectively. RMSEs of estimators are also presented in Tables 1 and 2.

		outliers					outliers		
α	pure data	5%	10%	20%	γ	pure data	5%	10%	20%
0.0	0.0104	0.0154	0.0236	0.0424	0.0	0.0104	0.0154	0.0236	0.0424
0.1	0.0104	0.0139	0.0207	0.0389	0.1	0.0104	0.0139	0.0207	0.0389
0.2	0.0105	0.0130	0.0186	0.0356	0.2	0.0105	0.0130	0.0186	0.0356
0.3	0.0107	0.0126	0.0171	0.0327	0.3	0.0107	0.0126	0.0171	0.0327
0.4	0.0109	0.0125	0.0161	0.0301	0.4	0.0109	0.0125	0.0161	0.0302
0.5	0.0112	0.0125	0.0155	0.0280	0.5	0.0112	0.0125	0.0155	0.0281
0.6	0.0114	0.0126	0.0151	0.0262	0.6	0.0114	0.0126	0.0152	0.0263
0.7	0.0117	0.0127	0.0150	0.0248	0.7	0.0117	0.0127	0.0150	0.0250
0.8	0.0120	0.0130	0.0149	0.0237	0.8	0.0120	0.0129	0.0149	0.0239
0.9	0.0123	0.0132	0.0150	0.0228	0.9	0.0123	0.0132	0.0150	0.0231
1.0	0.0126	0.0135	0.0151	0.0222	1.0	0.0126	0.0134	0.0150	0.0225
1.2	0.0133	0.0141	0.0155	0.0215	1.2	0.0133	0.0139	0.0153	0.0217
1.5	0.0143	0.0150	0.0163	0.0212	1.5	0.0143	0.0148	0.0160	0.0212
2.0	0.0161	0.0168	0.0179	0.0219	2.0	0.0159	0.0163	0.0173	0.0214

Table 1: RMSEs for $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\boldsymbol{\beta}}_\gamma$

5.2 Efficiency and robustness of the Wald-type test statistics

In this section, we empirically study the behaviour of the RP-based Wald-type tests for the MRM, for testing hypotheses both on $\boldsymbol{\beta}$ and σ. As in the previous section, we also carry out the study with DPD-based Wald-type tests in order to make a general comparison of both methods.

5.2.1 Wald-type tests for $\boldsymbol{\beta}$

We are now interested in testing

$$H_0 : \beta_1 = 0.5 \tag{21}$$

		outliers					outliers		
α	pure data	5%	10%	20%	γ	pure data	5%	10%	20%
0.0	0.0029	0.0108	0.0192	0.0321	0.0	0.0029	0.0108	0.0192	0.0321
0.1	0.0029	0.0089	0.0169	0.0308	0.1	0.0029	0.0089	0.0169	0.0308
0.2	0.0030	0.0075	0.0147	0.0293	0.2	0.0030	0.0075	0.0148	0.0293
0.3	0.0032	0.0066	0.0130	0.0276	0.3	0.0031	0.0067	0.0131	0.0277
0.4	0.0033	0.0060	0.0116	0.0258	0.4	0.0033	0.0061	0.0118	0.0261
0.5	0.0035	0.0056	0.0105	0.0241	0.5	0.0034	0.0058	0.0109	0.0247
0.6	0.0037	0.0054	0.0096	0.0224	0.6	0.0035	0.0056	0.0103	0.0234
0.7	0.0039	0.0053	0.0090	0.0210	0.7	0.0037	0.0055	0.0099	0.0224
0.8	0.0041	0.0053	0.0086	0.0197	0.8	0.0038	0.0055	0.0096	0.0216
0.9	0.0043	0.0053	0.0082	0.0185	0.9	0.0039	0.0055	0.0094	0.0210
1.0	0.0045	0.0054	0.0080	0.0176	1.0	0.0040	0.0055	0.0093	0.0205
1.2	0.0050	0.0057	0.0078	0.0161	1.2	0.0042	0.0056	0.0093	0.0200
1.5	0.0057	0.0062	0.0078	0.0146	1.5	0.0044	0.0057	0.0093	0.0196
2.0	0.0071	0.0075	0.0085	0.0135	2.0	0.0045	0.0058	0.0094	0.0190

Table 2: RMSEs for $\widehat{\sigma}_\alpha$ and $\widehat{\sigma}_\gamma$

In this case $\boldsymbol{L}^T = (1, 0)$. Therefore, hypothesis (21) can be expressed as

$$H_0 : \boldsymbol{L}^T \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = 0.5.$$

Let us consider the Wald-type tests based on minimum RP estimators, $W_n(\widehat{\boldsymbol{\beta}}_\alpha)$, given in (10), as well as the Wald-type tests based on minimum DPD estimators, $W_n(\widehat{\boldsymbol{\beta}}_\gamma)$ given in (19). In both cases we have,

$$\boldsymbol{X}_i \boldsymbol{X}_i^T = \begin{pmatrix} X_{i,1}^2 & X_{i,1}X_{i,2} \\ X_{i,1}X_{i,2} & X_{i,2}^2 \end{pmatrix}.$$

We denote

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \left(\sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^T \right)^{-1},$$

and we have

$$W_n(\widehat{\boldsymbol{\beta}}_\alpha) = \frac{(2\alpha+1)^{3/2}}{\widehat{\sigma}_\alpha^2 \, (\alpha+1)^3} \frac{(\widehat{\beta}_{\alpha,1} - 0.5)^2}{a_{11}} \tag{22}$$

and

$$W_n(\widehat{\boldsymbol{\beta}}_\gamma) = \frac{(2\gamma+1)^{3/2}}{\widehat{\sigma}_\gamma^2 \, (\gamma+1)^3} \frac{(\widehat{\beta}_{\gamma,1} - 0.5)^2}{a_{11}}. \tag{23}$$

At the 0.05 level of significance, hypothesis (21) will be rejected for RP-based Wald-type tests (equivalently for DPD Wald-type tests) if

$$W_n(\widehat{\boldsymbol{\beta}}_\alpha) > \chi^2_{1,0.05}. \tag{24}$$

Schemes on the study of the level and the power

As in Scheme 1 of Section 5.1, for each $i \in \{1, \ldots, n\}$, where $n = 600$, we generate a value Y_i from a $\mathcal{N}(0.5X_{1i} + 0.5X_{2i}, 0.01)$, and we evaluate expressions (22) and (23) under R simulated samples, obtaining the empirical levels, $\widehat{\alpha}_{\widehat{\boldsymbol{\beta}}_\alpha}$ and $\widehat{\alpha}_{\widehat{\boldsymbol{\beta}}_\gamma}$, presented in Table 3. These levels are computed as the number of times the null hypothesis is rejected over the total simulated samples R. To study the robustness we consider again Scheme 2 of Section 5.1, where $n_{out} \in \{30, 60, 120\}$ over the total n observations in each simulated data set are generated from a $\mathcal{N}(0.7X_{1i} + 0.7X_{2i}, 0.01)$.

To investigate the power robustness of these tests, we change the true data generating parameter value to be $\beta_1 = 0.45$. Thus, for each $i \in \{1, \ldots, n\}$, where $n = 600$, we generate a value Y_i from

		outliers					outliers		
α	pure data	5%	10%	20%	γ	pure data	5%	10%	20%
0.0	0.050	0.144	0.358	0.825	0.0	0.05	0.144	0.358	0.825
0.1	0.043	0.099	0.269	0.747	0.1	0.043	0.099	0.269	0.747
0.2	0.039	0.078	0.203	0.659	0.2	0.039	0.078	0.203	0.658
0.3	0.040	0.066	0.166	0.569	0.3	0.040	0.066	0.166	0.569
0.4	0.040	0.057	0.131	0.485	0.4	0.039	0.057	0.130	0.485
0.5	0.042	0.052	0.114	0.392	0.5	0.042	0.052	0.112	0.391
0.6	0.043	0.051	0.097	0.318	0.6	0.043	0.051	0.095	0.315
0.7	0.043	0.047	0.088	0.275	0.7	0.043	0.046	0.085	0.268
0.8	0.042	0.047	0.078	0.238	0.8	0.041	0.044	0.074	0.233
0.9	0.043	0.046	0.071	0.208	0.9	0.043	0.044	0.066	0.201
1.0	0.042	0.047	0.062	0.187	1.0	0.040	0.043	0.058	0.179
1.2	0.045	0.047	0.062	0.158	1.2	0.043	0.038	0.049	0.147
1.5	0.049	0.046	0.057	0.133	1.5	0.047	0.038	0.042	0.110
2.0	0.051	0.054	0.066	0.114	2.0	0.041	0.036	0.041	0.073

Table 3: Levels for $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\boldsymbol{\beta}}_\gamma$

		outliers					outliers		
α	pure data	5%	10%	20%	γ	pure data	5%	10%	20%
0.0	0.998	0.865	0.468	0.056	0.0	0.998	0.865	0.468	0.056
0.1	0.998	0.938	0.632	0.069	0.1	0.998	0.937	0.632	0.069
0.2	0.997	0.962	0.772	0.121	0.2	0.997	0.962	0.771	0.121
0.3	0.996	0.971	0.848	0.199	0.3	0.996	0.971	0.846	0.199
0.4	0.996	0.975	0.891	0.290	0.4	0.995	0.973	0.886	0.286
0.5	0.992	0.972	0.905	0.406	0.5	0.992	0.972	0.901	0.383
0.6	0.991	0.970	0.910	0.483	0.6	0.990	0.970	0.909	0.469
0.7	0.988	0.970	0.911	0.575	0.7	0.988	0.970	0.909	0.543
0.8	0.986	0.967	0.913	0.623	0.8	0.986	0.967	0.905	0.587
0.9	0.979	0.959	0.904	0.652	0.9	0.980	0.957	0.895	0.612
1.0	0.974	0.953	0.896	0.681	1.0	0.974	0.949	0.888	0.629
1.2	0.954	0.932	0.879	0.706	1.2	0.955	0.929	0.865	0.641
1.5	0.930	0.896	0.848	0.701	1.5	0.931	0.891	0.837	0.629
2.0	0.857	0.831	0.791	0.660	2.0	0.855	0.822	0.771	0.572

Table 4: Powers for $\widehat{\boldsymbol{\beta}}_\alpha$ and $\widehat{\boldsymbol{\beta}}_\gamma$

a $\mathcal{N}(0.45X_{1i} + 0.5X_{2i}, 0.01)$. To study the robustness, a 5%, 10% and 20% of observations in each simulated data set are generated from a $\mathcal{N}(0.7X_{1i} + 0.7X_{2i}, 0.01)$. Empirical powers, $\widehat{\pi}_{\widehat{\boldsymbol{\beta}}_\alpha}$ and $\widehat{\pi}_{\widehat{\boldsymbol{\beta}}_\gamma}$, obtained after R simulated samples, are presented in Table 4.

We also study the behaviour of RP-based Wald-type tests for different samples sizes $n \in \{50, 100, \dots, 650, 700\}$, as shown in Figure 2, for a 10% of contamination and tuning parameter $\alpha \in \{0.0, 0.5, 1.0, 1.5, 2.0\}$.

5.2.2 Wald-type tests for σ

We now consider the problem of testing

$$H_0 : \sigma = 0.1 \tag{25}$$

The Wald-type test statistics $W_n(\widehat{\sigma}_\alpha)$, based on RP estimators, for testing (25) is given by

$$W_n(\widehat{\sigma}_\alpha) = \frac{n}{l(\alpha)} \left(\frac{\widehat{\sigma}_\alpha - 0.1}{0.1} \right)^2, \tag{26}$$

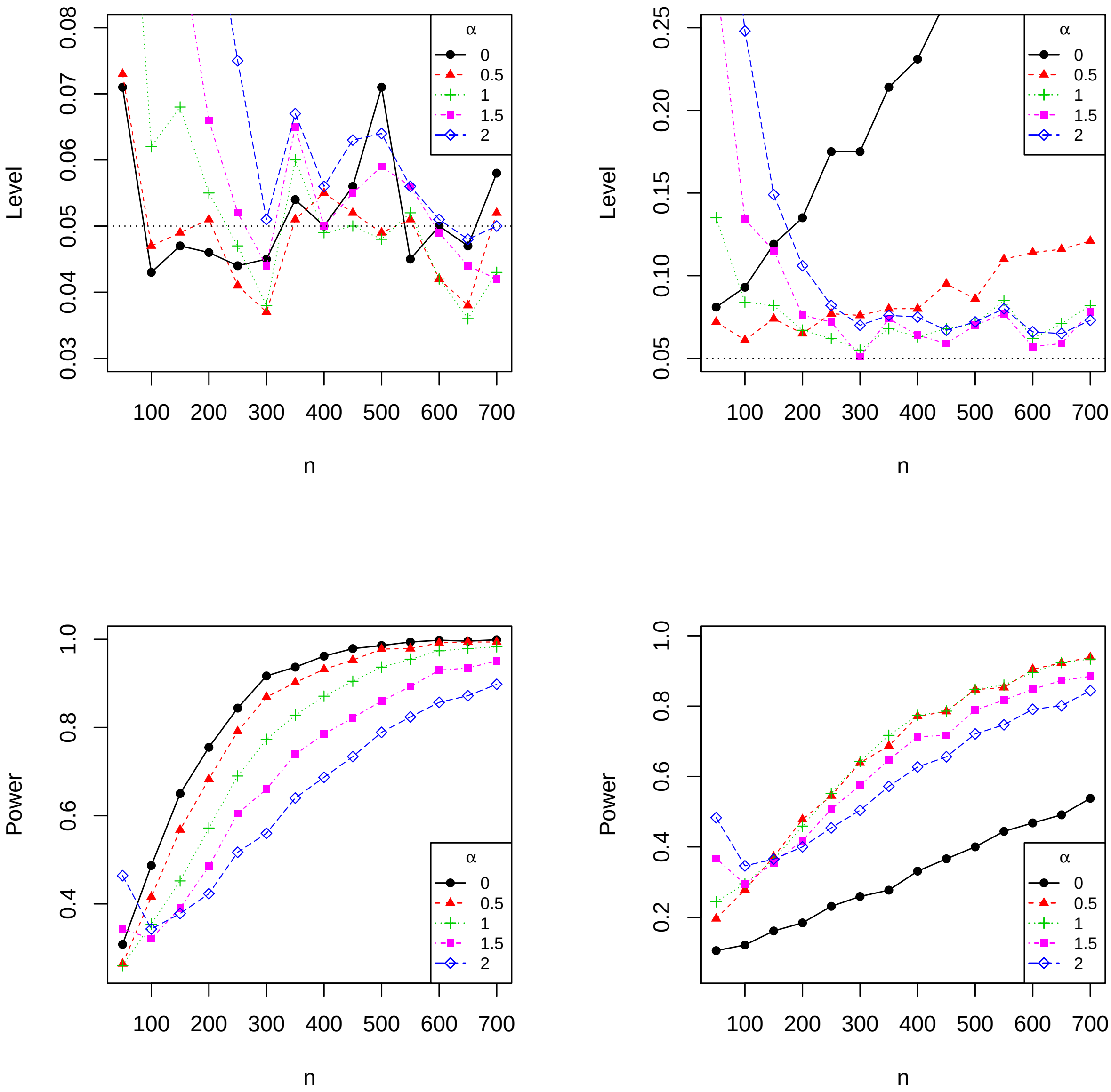

Figure 2: Levels (top) and powers (bottom) of the pure model (left) and the model contaminated with a 10% of outliers (right) for different sample sizes. Wald-type tests for $\boldsymbol{\beta}$.

being $l(\alpha) = \frac{(\alpha+1)^3(3\alpha^2+4\alpha+2)}{4(2\alpha+1)^{5/2}}$, rejecting the null hypothesis at level 0.05 if $W_n(\widehat{\sigma}_\alpha) > \chi^2_{1,0.05}$.

On the other hand, the corresponding Wald-type test statistic $W_n(\widehat{\sigma}_\gamma)$, based on DPD estimators, is given by

$$W_n(\widehat{\sigma}_\gamma) = \frac{n}{l(\gamma)} \left(\frac{\widehat{\sigma}_\gamma - 0.1}{0.1} \right)^2, \tag{27}$$

being $l(\gamma) = \frac{(1+\gamma)^2}{(2+\gamma^2)^2} \left(\frac{2(1+2\gamma^2)(1+\gamma)^3}{(1+2\gamma)^{5/2}} - \gamma^2 \right)$. The null hypothesis is rejected if $W_n(\widehat{\sigma}_\gamma) > \chi^2_{1,0.05}$.

Schemes on the study of the level and the power

As in previous sections, for each $i \in \{1, \ldots, n\}$, where $n = 600$, we generate a value Y_i from a $\mathcal{N}(0.5X_{1i} + 0.5X_{2i}, 0.01)$ and we evaluate expressions (26) and (27) under R simulated samples, obtaining the empirical levels $\widehat{\alpha}_{\widehat{\sigma}_\alpha}$ and $\widehat{\alpha}_{\widehat{\sigma}_\gamma}$, presented in Table 5. The same is done for $n_{out} \in \{30, 60, 120\}$ observations generated with parameters $(0.7, 0.7)$.

The power is empirically estimated by changing the true data generating parameter value to be $\sigma = 0.08$. Empirical powers, $\widehat{\pi}_{\widehat{\sigma}_\alpha}$ and $\widehat{\pi}_{\widehat{\sigma}_\gamma}$, obtained after R simulated samples, under pure and contaminated schemes, are presented in Table 6.

We study again the behaviour of RP-based Wald-type tests for different samples sizes $n \in \{50, 100, \ldots, 650, 700\}$ as shown in Figure 3, for a 10% of contamination and different values for the tuning parameter $\alpha \in \{0.0, 0.5, 1.0, 1.5, 2.0\}$.

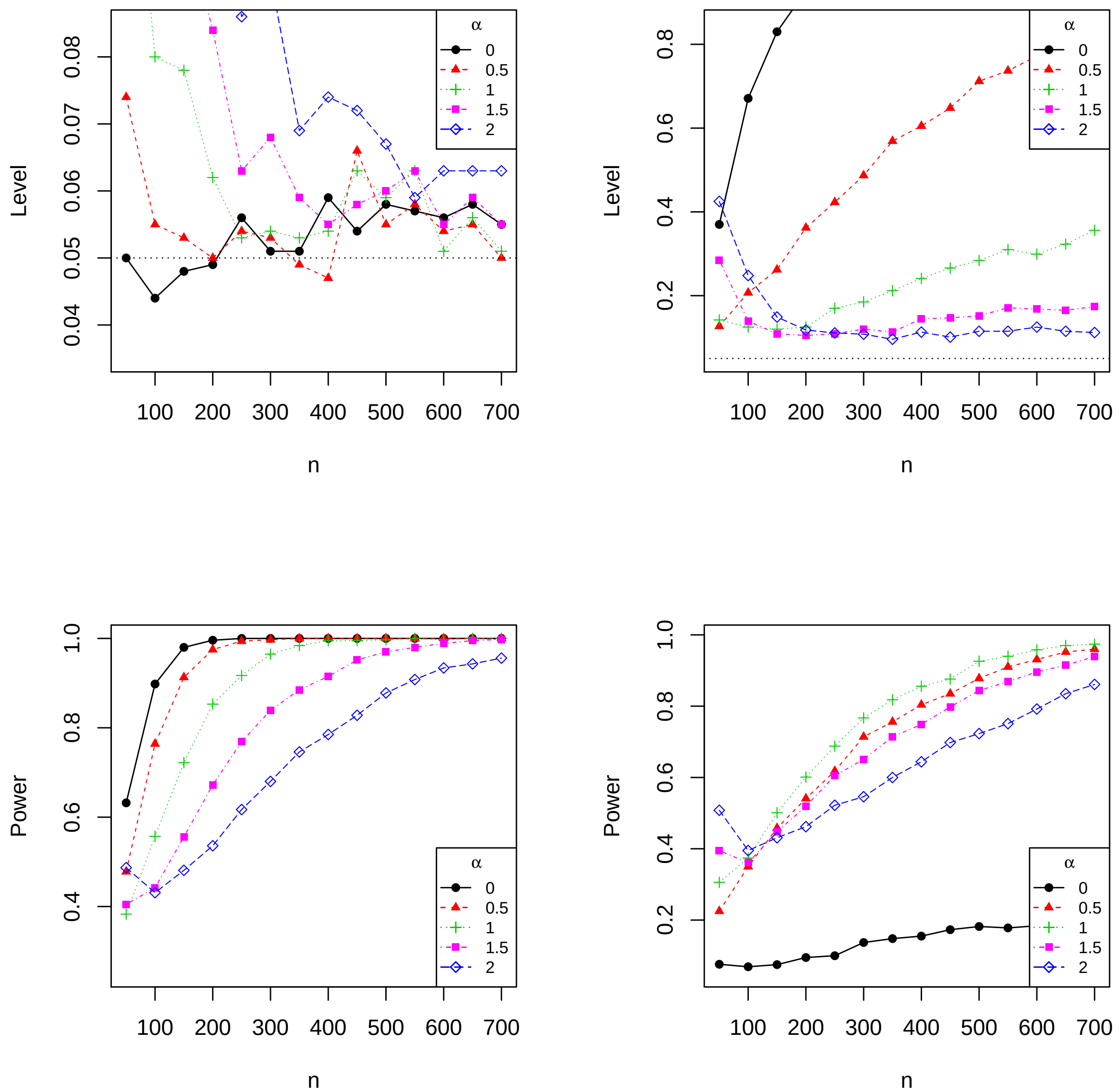

Figure 3: Levels (top) and powers (bottom) of the pure model (left) and the model contaminated with a 10% of outliers (right) for different sample sizes. Wald-type tests for σ.

α	pure data	outliers 5%	10%	20%	γ	pure data	outliers 5%	10%	20%
0.0	0.056	0.910	1.000	1.000	0.0	0.056	0.910	1.000	1.000
0.1	0.055	0.785	1.000	1.000	0.1	0.055	0.785	1.000	1.000
0.2	0.058	0.616	0.993	1.000	0.2	0.058	0.621	0.994	1.000
0.3	0.055	0.450	0.959	1.000	0.3	0.056	0.470	0.962	1.000
0.4	0.053	0.324	0.886	1.000	0.4	0.053	0.344	0.909	1.000
0.5	0.054	0.246	0.772	1.000	0.5	0.054	0.278	0.827	1.000
0.6	0.051	0.199	0.646	1.000	0.6	0.053	0.238	0.741	1.000
0.7	0.049	0.168	0.528	0.997	0.7	0.050	0.211	0.675	1.000
0.8	0.052	0.146	0.434	0.973	0.8	0.051	0.194	0.629	1.000
0.9	0.052	0.128	0.358	0.934	0.9	0.050	0.181	0.574	1.000
1.0	0.051	0.119	0.299	0.898	1.0	0.050	0.175	0.543	0.995
1.2	0.049	0.096	0.238	0.793	1.2	0.054	0.155	0.497	0.988
1.5	0.055	0.083	0.168	0.600	1.5	0.053	0.150	0.462	0.972
2.0	0.063	0.076	0.125	0.380	2.0	0.052	0.139	0.434	0.965

Table 5: Levels for $\widehat{\sigma}_\alpha$ and $\widehat{\sigma}_\gamma$

α	pure data	outliers 5%	10%	20%	γ	pure data	outliers 5%	10%	20%
0.0	1.000	0.722	0.184	1.000	0.0	1.000	0.722	0.184	1.000
0.1	1.000	0.967	0.067	1.000	0.1	1.000	0.967	0.067	1.000
0.2	1.000	0.994	0.359	0.982	0.2	1.000	0.994	0.349	0.984
0.3	1.000	0.998	0.720	0.838	0.3	1.000	0.998	0.689	0.862
0.4	1.000	0.998	0.872	0.486	0.4	1.000	0.998	0.849	0.575
0.5	1.000	0.997	0.931	0.181	0.5	1.000	0.997	0.905	0.262
0.6	1.000	0.997	0.953	0.125	0.6	1.000	0.997	0.928	0.127
0.7	1.000	0.997	0.962	0.178	0.7	1.000	0.997	0.936	0.094
0.8	1.000	0.996	0.964	0.275	0.8	1.000	0.997	0.937	0.107
0.9	0.999	0.996	0.961	0.377	0.9	1.000	0.996	0.931	0.122
1.0	0.997	0.993	0.958	0.456	1.0	1.000	0.995	0.927	0.127
1.2	0.997	0.987	0.940	0.523	1.2	1.000	0.992	0.909	0.130
1.5	0.989	0.966	0.896	0.561	1.5	0.999	0.990	0.880	0.124
2.0	0.934	0.875	0.792	0.539	2.0	0.997	0.984	0.849	0.110

Table 6: Powers for $\widehat{\sigma}_\alpha$ and $\widehat{\sigma}_\gamma$

5.3 Discussion of Results

In terms of efficiency, as expected, the MLE shows the best behaviour when pure data are considered, both for the estimation of $\boldsymbol{\beta}$ and σ. Robustness properties of the proposed alternative estimators are seen even when a low contamination is considered. The robustness of the tuning parameters α and γ increases at the same time as the contamination. In this sense, when a low degree of contamination is considered (5% of outliers) the best behaviour is obtained for $\alpha, \gamma \simeq 0.5$, whereas with a moderate or high contamination (10% and 20% of outliers), $\alpha, \gamma \simeq 1$ and $\alpha, \gamma \simeq 1.5$ are the best choices, respectively. RP and DPD estimators show a very similar behaviour for the estimation of $\boldsymbol{\beta}$, but RP estimators seem more precise when estimating σ (Tables 1 and 2).

When studying Wald-type tests for $\boldsymbol{\beta}$, and sample size $n = 600$, we have to distinguish two cases: the one in which pure data are considered and the case of contaminated data. In the first case, all the estimators present accurate results, with the exception of high tuning parameter estimators, which seem less appropriate for measuring the power. When contaminated data are considered, the MLE shows again its important lack of robustness, in contrast to RP and DPD estimators. RP and DPD have a very similar behaviour (Tables 3 and 4). The consideration of different samples sizes in Figure 2 reflects the poor behaviour of RP estimators with high tuning parameter for low samples sizes when estimating the level. Similar conclusions are obtained when studying Wald-type tests for σ, but here RP estimators improve DPD estimators for contaminated data, in concordance with the results on efficiency (Tables 5 and 6). In the graphs of Figure 3, we observe again the bad behaviour of the classical Wald test under the presence of outliers, while some Wald-type tests present a very good behaviour in robustness with a small lost of efficiency.

It seems then, that a moderate value of the tuning parameter $\alpha \simeq 1$ offers the best trade-off between efficiency and robustness, and that RP offers a better behaviour, when estimating the unknown variance or testing hypotheses about it, than DPD estimators.

6 On the choice of the optimal tuning parameter

In the previous sections, we have theoretically and empirically showed the efficiency but lack of robustness of the MLE as the solution of $\alpha = 0$, in the MRM. When increasing α the robustness of the RP estimators and associated Wald-type tests tends to increase with an associated loss in efficiency. From our analyses, it seems that a low-medium value of α is expected to provide the best trade-off for possibly moderately contaminated data. Although this ad-hoc choice of α should work good in practice, the question that arises in applying the method to practical real data scenarios is how to choose α appropriately, based only on the data given at hand. In this section we provide a data-driven procedure for the choice of the tuning parameter given any data set. We adapt the approach of Warwick and Jones (2005) to determine the optimal value of the tuning parameter for RP estimators in the normal MRM.

Warwick and Jones (2005) proposed to minimize an asymptotic approximation of the summed mean-squared error (MSE) of the estimators $\widehat{\boldsymbol{\theta}}_\alpha = (\widehat{\boldsymbol{\beta}}_\alpha, \widehat{\sigma}_\alpha)$ to choose the optimal α. This approximation is given by the sum of the estimated asymptotic bias and variance of $\widehat{\boldsymbol{\theta}}_\alpha$, given by $(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^*)(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}^*)^T + \frac{1}{n}Trace(\widehat{\boldsymbol{V}}_n(\widehat{\boldsymbol{\theta}}_\alpha))$, where $\boldsymbol{\theta}^*$ is the true target parameter value. Warwick and Jones (2005) suggested to use a suitable pilot estimator $\boldsymbol{\theta}^P$ in place of $\boldsymbol{\theta}^*$. If we take $\boldsymbol{\theta}^P = \widehat{\boldsymbol{\theta}}_\alpha$ as the pilot estimator, the approach coincides with that of Hong and Kim (2001) but does not take into account the model missespecification. As a potential choice of $\boldsymbol{\theta}^P$, we propose the RP estimator with parameter $\alpha^P = 1$. Thus, we minimize the estimated quantity

$$\widehat{MSE}(\alpha) = (\widehat{\boldsymbol{\theta}}_\alpha - \widehat{\boldsymbol{\theta}}_1)(\widehat{\boldsymbol{\theta}}_\alpha - \widehat{\boldsymbol{\theta}}_1)^T + \frac{1}{n}Trace(\widehat{\boldsymbol{V}}_n(\widehat{\boldsymbol{\theta}}_\alpha)). \tag{28}$$

As pointed out in Ghosh and Basu (2015), the estimation of the variance component in (28) should not assume the model to be true for a better robustness trade-off. So, instead of using (7) and (8), and after some heavy manipulations, a model robust estimate of the variance matrix is given in Theorem 15.

Theorem 15 *Let us consider the MRM in (3). Let $\widehat{\boldsymbol{\theta}}_\alpha$ be the minimum RP estimator for tuning parameter α. A model robust estimate of the variance matrix is given by*

$$\widehat{\boldsymbol{V}}_n(\widehat{\boldsymbol{\theta}}_\alpha) = \widehat{\boldsymbol{J}}_n^{-1}(\widehat{\boldsymbol{\theta}}_\alpha)\widehat{\boldsymbol{K}}_n(\widehat{\boldsymbol{\theta}}_\alpha)(\widehat{\boldsymbol{J}}_n^{-1}(\widehat{\boldsymbol{\theta}}_\alpha))^T,$$

where

$$\widehat{\boldsymbol{J}}_n(\widehat{\boldsymbol{\theta}}_\alpha) = \begin{bmatrix} J_{11,\alpha} & J_{12,\alpha} \\ J_{21,\alpha} & J_{22,\alpha} \end{bmatrix}, \qquad \widehat{\boldsymbol{K}}_n(\widehat{\boldsymbol{\theta}}_\alpha) = \begin{bmatrix} K_{11,\alpha} & \mathbf{0}_{k+1} \\ \mathbf{0}_{k+1}^T & K_{22,\alpha} \end{bmatrix}.$$

For $\alpha > 0$ we have

$$K_{11,\alpha} = \widehat{\sigma}_\alpha^{-\frac{2\alpha}{\alpha+1}-2} \frac{\alpha^2}{(2\alpha+1)^{3/2}}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1}$$

$$K_{22,\alpha} = \widehat{\sigma}_\alpha^{-\frac{2\alpha}{\alpha+1}-2} \frac{\alpha^2(3\alpha^2+4\alpha+2)}{(\alpha+1)^2(2\alpha+1)^{5/2}}$$

$$
\begin{aligned}
J_{11,\alpha} =& \frac{1}{n}\sum_{i=1}^{n} \widehat{\sigma}_{\alpha}^{-\frac{\alpha}{\alpha+1}-2} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right)\left\{\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2 \alpha^2 - \alpha\right\}\left(\frac{1}{n}\boldsymbol{X}_i\boldsymbol{X}_i^T\right),\\
J_{12,\alpha} =& \frac{1}{n}\sum_{i=1}^{n}\left\{-\frac{\alpha^2}{\alpha+1}\widehat{\sigma}_{\alpha}^{1-\frac{4\alpha+3}{\alpha+1}}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)\left[1-(\alpha+1)\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right]\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right)\boldsymbol{X}_i\right.\\
&\left.-2\alpha\widehat{\sigma}_{\alpha}^{1-\frac{4\alpha+3}{\alpha+1}}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right)\boldsymbol{X}_i\right\} = J_{21,\alpha}^T,\\
J_{22,\alpha} =& \frac{1}{n}\sum_{i=1}^{n}\left\{-\frac{2\alpha}{\alpha+1}\widehat{\sigma}_{\alpha}^{1-\frac{4\alpha+3}{\alpha+1}}\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right)\right.\\
&+\frac{\alpha(4\alpha+3)}{(\alpha+1)^2}\widehat{\sigma}_{\alpha}^{1-\frac{4\alpha+3}{\alpha+1}}\left[1-(\alpha+1)\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right]\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right)\\
&\left.-\frac{\alpha^2}{\alpha+1}\widehat{\sigma}_{\alpha}^{1-\frac{4\alpha+3}{\alpha+1}}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\left[1-(\alpha+1)\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right]\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_{\alpha}}{\widehat{\sigma}_{\alpha}}\right)^2\right)\right\},
\end{aligned}
$$

whereas for $\alpha = 0$ we have

$$
\begin{aligned}
K_{11,0} =& \frac{1}{\widehat{\sigma}^2}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^T\right)^{-1},\\
K_{22,0} =& \frac{2}{\widehat{\sigma}^2},\\
J_{11,0} =& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widehat{\sigma}^2}\left(\frac{1}{n}\boldsymbol{X}_i\boldsymbol{X}_i^T\right),\\
J_{12,0} =& \frac{1}{n}\sum_{i=1}^{n}\frac{-2}{\widehat{\sigma}^2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)\boldsymbol{X}_i = J_{21,0}^T,\\
J_{22,0} =& \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widehat{\sigma}^2}\left[1-3\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)^2\right].
\end{aligned}
$$

Remark 16 *Note that, under the true model, $\widehat{\boldsymbol{V}}_n(\widehat{\boldsymbol{\theta}}_{\alpha})$ asymptotically converges to the variance matrix given in (6).*

Optimal α will then be obtained by minimizing (28) through a grid search over the preferable interval, advisedly $\alpha \in [0, 2]$. This method has also been extended and applied in some directions in Ghosh and Basu (2016), Basu et al. (2017) and Balakrishnan et al. (2019), all of them for DPD-based estimators. In Castilla et al. (2018), it was applied to minimum phi-divergence estimators in multinomial logistic regression with complex survey.

7 Concluding Remarks

In this paper we have introduced a new family of Wald-type tests for the MRM. This family is based on the minimum RP estimators (Broniatowski et al., 2012) which depend on a tuning parameter α, with the MLE as a particular case ($\alpha = 0$). Through the study of its influence function, this family of tests is shown to present a much more robust behaviour than the classical test for the MRM based on the MLE. These results are empirically tested through an extensive simulation study, in which RP-based Wald type tests are compared with DPD-based Wald type tests (Basu et al., 1998) too, showing a better behaviour when estimating the unknown variance. Finally, we have provided a data-driven procedure for the choice of the optimal α given any data set.

A Appendix: Proof of the Results

Proof of Theorem 4

Proof. We have $\boldsymbol{L}^T\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{c} = \boldsymbol{L}^T(\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0)$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_0) \underset{n\to\infty}{\overset{L}{\longrightarrow}} N\left(\boldsymbol{0},\boldsymbol{\Sigma}_\alpha\left(\sigma\right)\right)$. Therefore

$$\sqrt{n}\left(\boldsymbol{L}^T\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{c}\right) \underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(\boldsymbol{0},\ \boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha\left(\sigma\right)\boldsymbol{L}\right)$$

and the asymptotic distribution of $W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right)$ is a chi-square distribution with r degrees of freedom because $\boldsymbol{\Sigma}\left(\widehat{\sigma}_\alpha\right)$ is a consistent estimator of $\boldsymbol{\Sigma}_\alpha\left(\sigma\right)$

$$\boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha\left(\sigma\right)\boldsymbol{L}\left(\boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha\left(\sigma\right)\boldsymbol{L}\right)^{-1} = \boldsymbol{I}_{p\times p}.$$

■

Proof of Theorem 6

Proof. We have

$$\begin{aligned}\pi_n^\alpha\left(\boldsymbol{\beta}^*\right) &= \Pr\left(W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right) > \chi^2_{r,\tau}\right) = \Pr\left(n\left(l_{\widehat{\boldsymbol{\beta}}_\alpha}(\widehat{\sigma}_\alpha) - l_{\boldsymbol{\beta}^*}(\sigma)\right) > \chi^2_{r,\tau} - nl_{\boldsymbol{\beta}^*}(\sigma)\right)\\ &= \Pr\left(\sqrt{n}\left(l_{\widehat{\boldsymbol{\beta}}_\alpha}(\widehat{\sigma}_\alpha) - l_{\boldsymbol{\beta}^*}(\sigma)\right) > \frac{\chi^2_{r,\tau}}{\sqrt{n}} - \sqrt{n}l_{\boldsymbol{\beta}^*}(\sigma)\right).\end{aligned}$$

Now we are going to get the asymptotic distribution of the random variable $\sqrt{n}\left(l_{\widehat{\boldsymbol{\beta}}_\alpha}(\widehat{\sigma}_\alpha) - l_{\boldsymbol{\beta}^*}(\sigma)\right)$.

It is clear that $l_{\widehat{\boldsymbol{\beta}}_\alpha}(\widehat{\sigma}_\alpha)$ and $l_{\widehat{\boldsymbol{\beta}}_\alpha}(\sigma)$ have the same asymptotic distribution because $\boldsymbol{\Sigma}_\alpha\left(\widehat{\sigma}_\alpha\right) \underset{n\to\infty}{\overset{P}{\longrightarrow}} \boldsymbol{\Sigma}_\alpha\left(\sigma\right)$. A first Taylor expansion of $l_{\widehat{\boldsymbol{\beta}}_\alpha}(\sigma)$ at $\widehat{\boldsymbol{\beta}}_\alpha$ around $\boldsymbol{\beta}^*$ gives

$$l_{\widehat{\boldsymbol{\beta}}_\alpha}(\sigma) - l_{\boldsymbol{\beta}^*}(\sigma) = \left.\frac{\partial l_{\boldsymbol{\beta}}(\sigma)}{\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}(\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^*) + o_p\left(\left\|\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^*\right\|\right).$$

Therefore,

$$\sqrt{n}\left(l_{\widehat{\boldsymbol{\beta}}_\alpha}(\widehat{\sigma}_\alpha) - l_{\boldsymbol{\beta}^*}(\sigma)\right) \underset{n\to\infty}{\overset{L}{\longrightarrow}} N\left(0, \sigma^2\left(\boldsymbol{\beta}^*\right)\right),$$

where

$$\sigma^2\left(\boldsymbol{\beta}^*\right) = \left.\frac{\partial l_{\boldsymbol{\beta}}(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^T}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\left(\sigma\right)\left.\frac{\partial l_{\boldsymbol{\beta}}(\boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}.$$

Now the result follows. ■

Proof of Theorem 8

Proof. We have

$$\begin{aligned}\boldsymbol{L}^T\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{c} &= \boldsymbol{L}^T\ \boldsymbol{\beta}_n - c + \boldsymbol{L}^T\left(\widehat{\boldsymbol{\beta}}_\alpha -\ \boldsymbol{\beta}_n\right)\\ &= \boldsymbol{L}^T\ \boldsymbol{\beta}_0 + \boldsymbol{L}^{\boldsymbol{T}}n^{-1/2}\boldsymbol{d} - c + \boldsymbol{L}^{\boldsymbol{T}}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\alpha}} -\ \boldsymbol{\beta}_{\boldsymbol{n}}\right)\\ &= \boldsymbol{L}^{\ \boldsymbol{T}}n^{-1/2}\boldsymbol{d} + \boldsymbol{L}^{\boldsymbol{T}}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\alpha}} -\ \boldsymbol{\beta}_{\boldsymbol{n}}\right).\end{aligned}$$

Therefore,

$$\boldsymbol{L}^T\ \widehat{\boldsymbol{\beta}}_\alpha - c = \boldsymbol{L}^{\boldsymbol{T}}\ n^{-1/2}\boldsymbol{d} + \boldsymbol{L}^{\boldsymbol{T}}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\alpha}} -\ \boldsymbol{\beta}_{\boldsymbol{n}}\right).$$

We know, under $H_{1,n}$, that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\boldsymbol{n}}\right) \underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(\boldsymbol{0},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\left(\sigma\right)\right)$$

and

$$\sqrt{n}\left(\boldsymbol{L}^T\ \widehat{\boldsymbol{\beta}}_\alpha-\boldsymbol{c}\right)\underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}}\mathcal{N}\left(\boldsymbol{L^T d},\ \boldsymbol{L^T\Sigma_\alpha}\left(\sigma\right)\boldsymbol{L}\right).$$

Moreover, we know that if $\boldsymbol{Z}\in\mathcal{N}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$, $\boldsymbol{\Sigma}$ is a symmetric projection of rank k and $\boldsymbol{\Sigma\mu}=\boldsymbol{\mu},$ then $\boldsymbol{Z}^T\boldsymbol{Z}$ is a chi-square distribution with k degrees of freedom and noncentrality parameter $\boldsymbol{\mu}^T\boldsymbol{\mu}.$

The quadratic form is

$$W_n\left(\widehat{\boldsymbol{\beta}}_\alpha\right)=\boldsymbol{Z}^T\boldsymbol{Z}$$

with

$$\boldsymbol{Z}=\sqrt{n}\left[\boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha\left(\sigma\right)\boldsymbol{L}\right]^{-1/2}\left(\boldsymbol{L}^T\widehat{\boldsymbol{\beta}}_\alpha-\boldsymbol{c}\right)$$

and

$$\boldsymbol{Z}\underset{n\longrightarrow\infty}{\overset{\mathcal{L}}{\longrightarrow}}\mathcal{N}\left(\left[\boldsymbol{L}^T\boldsymbol{\Sigma}_\alpha\left(\sigma\right)\boldsymbol{L}\right]^{-1/2}\boldsymbol{L}^T\boldsymbol{d},\boldsymbol{I}_{r\times r}\right),$$

where $\boldsymbol{I}$ is the identity $r\times r$ matrix. Hence the application of the result is immediate. The noncentrality parameter is

$$\boldsymbol{d}^T\boldsymbol{L}\left[\boldsymbol{L^T\Sigma_\alpha}\left(\sigma\right)\boldsymbol{L}\right]^{-1}\boldsymbol{L^T d}.$$

■

Proof of Theorem 10

Proof. The proof is obtained by taking into account that

$$\sqrt{n}\left(\widehat{\sigma}_\alpha-\sigma_0\right)\underset{n\rightarrow\infty}{\overset{L}{\longrightarrow}}N(0,\frac{\sigma_0^2\left(\alpha+1\right)^3\left(3\alpha^2+4\alpha+2\right)}{4\left(2\alpha+1\right)^{5/2}}).$$

■

Proof of Theorem 12

Proof.

$$\begin{aligned}\pi_{W^\alpha_{n,\sigma}}\left(\sigma^*\right)&=\Pr\left(\frac{\sqrt{n}}{\sqrt{l\left(\alpha\right)}}\left(\frac{\widehat{\sigma}_\alpha-\sigma_0}{\sigma_0}\right)>\sqrt{\chi^2_{1,\tau}}\right)=\Pr\left(\frac{\sqrt{n}}{\sqrt{l\left(\alpha\right)}}\left(\left(\frac{\widehat{\sigma}_\alpha-\sigma^*}{\sigma_0}\right)+\left(\frac{\sigma^*-\sigma_0}{\sigma_0}\right)\right)>\chi^2_{1,\tau}\right)\\&=\Pr\left(\frac{\sqrt{n}}{\sqrt{l\left(\alpha\right)}}\left(\left(\frac{\widehat{\sigma}_\alpha-\sigma^*}{\sigma^*}\right)\right)>\frac{\sigma_0}{\sigma^*}\left(\chi^2_{1,\tau}-\frac{\sqrt{n}}{\sqrt{l\left(\alpha\right)}}\left(\frac{\sigma^*-\sigma_0}{\sigma_0}\right)\right)\right)\\&=1-\Phi_n\left(\frac{\sigma_0}{\sigma^*}\left(\chi^2_{1,\tau}-\frac{\sqrt{n}}{\sqrt{l\left(\alpha\right)}}\left(\frac{\sigma^*-\sigma_0}{\sigma_0}\right)\right)\right).\end{aligned}$$

■

Proof of Theorem 15

Proof. Let us consider the MRM in (3). The minimum RP estimator, $\widehat{\boldsymbol{\theta}}_\sigma$, is obtained by minimizing the function $\frac{1}{n}\sum_{i=1}^n h(Y_i,\boldsymbol{\theta})$, where

$$h(Y_i,\boldsymbol{\theta})=\begin{cases}log(\sigma)+\frac{1}{2}\left(\frac{Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2 & \text{if } \alpha=0\\ \sigma^{-\frac{\alpha}{\alpha+1}}exp\left(\frac{\alpha}{2}\left(\frac{Y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}}{\sigma}\right)^2\right) & \text{if } \alpha>0\end{cases}.$$

Following Theorem 4 in Broniawtoski et al. (2012), $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ converges in distribution to the matrix

$$\boldsymbol{V}_n(\boldsymbol{\theta}_0)=\boldsymbol{J}_n^{-1}(\boldsymbol{\theta}_0)\boldsymbol{K}_n(\boldsymbol{\theta}_0)(\boldsymbol{J}_n^{-1}(\boldsymbol{\theta}_0))^T,$$

where

$$\boldsymbol{J}_n(\boldsymbol{\theta}_0) = -\int \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} h(y, \boldsymbol{\theta}_0) dF_{\boldsymbol{\theta_0}}(y)$$

and

$$\begin{aligned}
\boldsymbol{K}_n(\boldsymbol{\theta}_0) &= \int \frac{\partial}{\partial\boldsymbol{\theta}} h(y, \boldsymbol{\theta}_0) \frac{\partial}{\partial\boldsymbol{\theta}^T} h(y, \boldsymbol{\theta}_0) dF_{\boldsymbol{\theta_0}}(y) - \left(\int \frac{\partial}{\partial\boldsymbol{\theta}} h(y, \boldsymbol{\theta}_0) dF_{\boldsymbol{\theta_0}}(y)\right) \left(\int \frac{\partial}{\partial\boldsymbol{\theta}} h(y, \boldsymbol{\theta}_0) dF_{\boldsymbol{\theta_0}}(y)\right)^T \\
&= \int \frac{\partial}{\partial\boldsymbol{\theta}} h(y, \boldsymbol{\theta}_0) \frac{\partial}{\partial\boldsymbol{\theta}^T} h(y, \boldsymbol{\theta}_0) dF_{\boldsymbol{\theta_0}}(y).
\end{aligned}$$

The estimation of $\boldsymbol{J}_n(\boldsymbol{\theta}_0)$ can be done in a natural manner by replacing $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}}_\alpha$ and the distribution function by the corresponding empirical sample values, so that we have

$$\widehat{\boldsymbol{J}}_n(\widehat{\boldsymbol{\theta}}_\alpha) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} h(Y_i, \widehat{\boldsymbol{\theta}}_\alpha).$$

However, we will normally have only one observation from each $F_{\boldsymbol{\theta}}(Y_i)$, and for such situations it is not possible to estimate the variance based on only one observation. Therefore, we replace $\boldsymbol{\theta}_0$ by $\widehat{\boldsymbol{\theta}}_\alpha$ and $F_{\boldsymbol{\theta}}(y)$ by the model normal-densities $f_{\boldsymbol{\theta}}(y)$ (see, for example, Ghosh and Basu, 2015). Thus, we use the approximation

$$\widehat{\boldsymbol{K}}_n(\widehat{\boldsymbol{\theta}}_\alpha) = \left.\left(\int \frac{\partial}{\partial\boldsymbol{\theta}} h(y, \boldsymbol{\theta}_0) \frac{\partial}{\partial\boldsymbol{\theta}^T} h(y, \boldsymbol{\theta}_0) f_{\boldsymbol{\theta_0}}(y) dy\right)\right|_{\widehat{\boldsymbol{\theta}}_\alpha}.$$

The result follows taking into account that

$$\begin{aligned}
\left.\frac{\partial}{\partial\boldsymbol{\beta}} h(Y_i, \widehat{\boldsymbol{\theta}}_\alpha)\right|_{\alpha=0} &= \left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)\boldsymbol{X}_i, \\
\left.\frac{\partial}{\partial\boldsymbol{\beta}} h(Y_i, \widehat{\boldsymbol{\theta}}_\alpha)\right|_{\alpha>0} &= \alpha\widehat{\sigma}_\alpha^{-\frac{\alpha}{\alpha+1}-1} \exp\left(\frac{-\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_\alpha}{\widehat{\sigma}_\alpha}\right)^2\right)\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_\alpha}{\widehat{\sigma}_\alpha}\right)\boldsymbol{X}_i, \\
\left.\frac{\partial}{\partial\sigma} h(Y_i, \widehat{\boldsymbol{\theta}}_\alpha)\right|_{\alpha=0} &= \frac{1}{\sqrt{2\pi}\widehat{\sigma}^2}\left[\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)^2 - 1\right], \\
\left.\frac{\partial}{\partial\sigma} h(Y_i, \widehat{\boldsymbol{\theta}}_\alpha)\right|_{\alpha>0} &= \alpha\widehat{\sigma}_\alpha^{-\frac{\alpha}{\alpha+1}-1} \exp\left(\frac{-2\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_\alpha}{\widehat{\sigma}_\alpha}\right)^2\right)\left[\left(\frac{Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_\alpha}{\widehat{\sigma}_\alpha}\right)^2 - \frac{1}{\alpha+1}\right].
\end{aligned}$$

■

References

[1] Balakrishnan, N., Castilla, E., Martn N. and Pardo, L. (2019).Robust estimators and test-statistics for one-shot device testing under the exponential distribution. *IEEE transactions on Information Theory*, **65**(5), 3080–3096

[2] Basu, A., Ghosh, A. Mandal, Martín, N. and Pardo, L. (2017). A Wald-type test statistic for testing linear hypothesis in logistic regression models based on minimum density power divergence estimator. *Electonic Journal of Statistics*, **11**, 2741–2772.

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

[4] Basu, A., Mandal, A., Martín, N. and Pardo, L. (2016). Generalized Wald-type tests based on minimum density power divergence estimators. *Statistics*, **50** (1), 1–26.

[5] Broniatowski, M., Toma, A. and Vajda, I. (2012). Decomposable pseudodistances and applications in statistical estimation, *Journal of Statistical Planning and Inference*, **142**, 2574–2585.

[6] Castilla, E, Martín, N. and Pardo, L. (2018). Pseudo minimum phi-divergence estimator for the multinomial logistic regression model with complex sample design. *AStA Advanced Statistical Analysis*, **102**(3), 381–411.

[7] Durio, A., Isaia, E. D. (2011). The Minimum Density Power Divergence Approach in Building Robust Regression Models. *Informatica,* **22**(1), 43-56.

[8] Fraser, D.A.S. (1957). *Nonparametric Methods in Statistics.* John Wiley & Sons, New York.

[9] Ghosh, A. and Basu, A. (2015). Robust estimation for non- homogeneous data and the selection of the optimal tuning parameter: The density power divergence approach. *Journal of Applied Statitsics*, **42**, 2056-2072.

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

[11] Ghosh, A., Mandal, A., Martín, N. and Pardo, L. (2016). Influence analysis of robust Wald-type tests. *Journal of Multivariate Analysis,* **147**, 102-126.

[12] Hong, C. and Kim, Y. (2001). Automatic selection of the tuning parameter in the minimum density power divergence estimation. *Journal of the Korean Statistical Society*, **30**, 453-465.

[13] Jones, M.C., Hjort, N.L., Harris, I.R. and Basu, A. (2001). A comparison of related density-based minimum divergence estimators. *Biometrika*, **88**, 865-873.

[14] Liese, F. and Vajda, I. (1987). *Convex Statistical Distances.* Teubner, Leipzig.

[15] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures.* Chapman & Hall/CRC, Boca de Raton.

[16] Rényi, A. (1961). On measures of entropy and information. In: Neyman, J. (ed.) *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, Berkeley, 547-561.

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