

Generalized Wald-type Tests based on Minimum Density Power Divergence Estimators

Basu, A.¹; Mandal, A.¹; Martin, N.² and Pardo, L.³

¹Indian Statistical Institute, Kolkata 700108, India

²Department of Statistics, Carlos III University of Madrid, 28038 Getafe (Madrid), Spain

³Department of Statistics and O.R. Complutense University of Madrid, 28040 Madrid, Spain

March 10, 2015

Abstract

In testing of hypothesis the robustness of the tests is an important concern. Generally, the maximum likelihood based tests are most efficient under standard regularity conditions, but they are highly non-robust even under small deviations from the assumed conditions. In this paper we have proposed generalized Wald-type tests based on minimum density power divergence estimators for parametric hypotheses. This method avoids the use of nonparametric density estimation and the bandwidth selection. The trade-off between efficiency and robustness is controlled by a tuning parameter β . The asymptotic distributions of the test statistics are chi-square with appropriate degrees of freedom. The performance of the proposed tests are explored through simulations and real data analysis.

AMS 2010 subject classification: Primary 62F35; Secondary 62F03.

Keywords: Density Power Divergence, Robustness, Tests of Hypotheses.

1 Introduction

Basu et al. (1998) have introduced the minimum density power divergence estimator (MDPDE) that minimizes the density power divergence measure. The robustness properties of these estimators have been studied in detail by several authors. However, the problem of hypothesis testing based on the density power divergence measures has only recently been explored (Basu et al., 2013, 2014). The results indicate that the test statistics based on the density power divergence measures, when the parameters are estimated by the MDPDEs, have substantially superior performance compared to the likelihood ratio test in the presence of outliers. On the other hand, in pure data these tests are often competitive to the likelihood ratio tests. So the tests based on the density power divergence are very useful practical tools in robust statistics.

However, sometimes it is not easy to get the expression of the density power divergence measure between the population densities estimated under the null hypothesis and under the unrestricted parameter space. Another difficulty is that in many situations the asymptotic distributions of the above test statistics are linear combinations of independent chi-square variables. While none of these problems are insurmountable, they lead to some reduction in the appeal of these otherwise useful tests.

To overcome these problems we propose a class of generalized Wald-type test statistics based on minimum density power divergence estimators of the model parameters. Our aim is to manipulate the statistics in a manner that allows us to exploit the nice properties of the MDPDEs in constructing the test statistics, and yet come up with asymptotic distributions which are pure chi-squares rather than linear combinations of independent chi-squares. On the whole, we expect to make the tests theoretically and operationally far more simple, without compromising the efficiency and the robustness properties of the tests.

In Section 2 we have introduced the necessary notations and described the existing results. We have proposed Wald-type test statistics in the context of the simple null hypothesis in Section 3. The

problem of the composite null hypothesis is considered in Section 4. Some simulation studies are reported in Section 5 to explore the behavior of the proposed families of test statistics. Section 6 provides a brief description on the choice of the tuning parameter β associated with the proposed Wald-type test. Section 7 has some concluding remarks, and the proofs are given in the Appendix.

Throughout this paper, we will refer to the usual Wald test statistic based on the maximum likelihood estimator as the “classical Wald statistic”, whereas the new robust statistics are referred to as “proposed Wald-type test statistics”.

2 Notational Set up and Existing Concepts

Let \mathcal{G} denote the set of all distributions having densities with respect to a dominating measure (generally the Lebesgue measure or the counting measure). Given any two densities g and f in \mathcal{G} , the density power divergence between them is defined, as the function of a nonnegative tuning parameter β , as

$$d_\beta(g, f) = \begin{cases} \int \left\{ f^{1+\beta}(x) - \left(1 + \frac{1}{\beta}\right) f^\beta(x)g(x) + \frac{1}{\beta}g^{1+\beta}(x) \right\} dx, & \text{for } \beta > 0, \\ \int g(x) \log \left(\frac{g(x)}{f(x)} \right) dx, & \text{for } \beta = 0. \end{cases} \quad (1)$$

The case corresponding to $\beta = 0$ may be derived from the general case by taking the continuous limit as $\beta \rightarrow 0$, and in this case $d_0(g, f)$ is the classical Kullback-Leibler divergence. The quantities defined in equation (1) are genuine divergences in the sense $d_\beta(g, f) \geq 0$ for all $g, f \in \mathcal{G}$ and all $\beta \geq 0$, and $d_\beta(g, f)$ is equal to zero if and only if the densities g and f are identically equal. More details about inference based on divergence measures can be found in Pardo (2006) and Basu et al. (2011).

We consider a parametric model of densities $\{f_\theta : \theta \in \Theta \subset \mathbb{R}^p\}$ and suppose that we are interested in the estimation of θ . Let G represent the distribution function corresponding to the density g . The minimum density power divergence functional $T_\beta(G)$ at G is defined by the requirement $d_\beta(g, f_{T_\beta(G)}) = \min_{\theta \in \Theta} d_\beta(g, f_\theta)$. Clearly the term $\int g^{1+\beta}(x)dx$ in (1) has no role in the minimization of $d_\beta(g, f_\theta)$ over $\theta \in \Theta$. Thus the essential objective function to be minimized in the computation of the minimum density power divergence functional $T_\beta(G)$ reduces to

$$\int f_\theta^{1+\beta}(x)dx - \left(1 + \frac{1}{\beta}\right) \int f_\theta^\beta(x)g(x)dx = \int f_\theta^{1+\beta}(x)dx - \left(1 + \frac{1}{\beta}\right) \int f_\theta^\beta(x)dG(x). \quad (2)$$

Notice that in the above objective function the density g appears only as a linear term (unlike, say, the objective function of the minimum Hellinger distance functional where the square root of the density g is the relevant quantity). Given a random sample X_1, \dots, X_n from the distribution G , we can therefore approximate the objective function in (2) by replacing G with its empirical estimate G_n . For a given tuning parameter β , the MDPDE $\hat{\theta}_\beta$ of θ can be obtained by minimizing

$$\begin{aligned} & \int f_\theta^{1+\beta}(x)dx - \left(1 + \frac{1}{\beta}\right) \int f_\theta^\beta(x)dG_n(x) \\ &= \int f_\theta^{1+\beta}(x)dx - \left(1 + \frac{1}{\beta}\right) \frac{1}{n} \sum_{i=1}^n f_\theta^\beta(X_i) = \frac{1}{n} \sum_{i=1}^n V_\theta(X_i) \end{aligned} \quad (3)$$

over $\theta \in \Theta$, where $V_\theta(x) = \int f_\theta^{1+\beta}(y)dy - \left(1 + \frac{1}{\beta}\right) f_\theta^\beta(x)$. In the special case $\beta = 0$, we must minimize the expression $-\frac{1}{n} \sum_{i=1}^n \log f_\theta(X_i)$; the corresponding minimizer turns out to be the maximum likelihood estimator (MLE) of θ . The minimization of the expression in (3) over θ does not require the use of a nonparametric density estimate of the true unknown distribution G . Existing theory (e.g. De Angelis and Young, 1992) shows that in general there is hardly any advantage in introducing smoothing for such functionals which may be empirically estimated using the empirical distribution function alone, except in very special cases. It is therefore natural to substitute G with G_n in such situations.

Let $\mathbf{u}_\theta(x) = \frac{\partial}{\partial \theta} \log f_\theta(x)$ be the score function of the model. Under differentiability of the model the minimization of the objective function in equation (3) leads to an estimating equation of the form

$$\frac{1}{n} \sum_{i=1}^n \mathbf{u}_\theta(X_i) f_\theta^\beta(X_i) - \int \mathbf{u}_\theta(x) f_\theta^{1+\beta}(x) dx = \mathbf{0}_p, \quad (4)$$

which is an unbiased estimating equation under the model. Here $\mathbf{0}_p$ denotes the null vector of dimension p . Since the corresponding estimating equation weights the score $\mathbf{u}_\theta(X_i)$ with the power of the density $f_\theta^\beta(X_i)$, the outlier resistant behavior of the estimator is intuitively apparent. See Basu et al. (1998) and Jones et al. (2001) for more details.

The functional $T_\beta(\cdot)$ is Fisher consistent; it takes the value θ_0 when the true density $g = f_{\theta_0}$ is in the model. When it is not, $\theta_\beta^g = T_\beta(G)$ represents the best fitting parameter. For brevity we will suppress the g superscript in the notation for θ_β^g ; f_{θ_β} is the model element closest to the density g in the minimum density power divergence sense corresponding to the tuning parameter β .

Let g be the true data generating density, and $\theta_\beta = T_\beta(G)$ be the best fitting parameter. To set up the notation we define the quantities

$$\begin{aligned} \mathbf{J}_\beta(\theta) &= \int \mathbf{u}_\theta(x) \mathbf{u}_\theta^T(x) f_\theta^{1+\beta}(x) dx \\ &+ \int \{\mathbf{I}_\theta(x) - \beta \mathbf{u}_\theta(x) \mathbf{u}_\theta^T(x)\} \{g(x) - f_\theta(x)\} f_\theta^\beta(x) dx, \end{aligned} \quad (5)$$

$$\mathbf{K}_\beta(\theta) = \int \mathbf{u}_\theta(x) \mathbf{u}_\theta^T(x) f_\theta^{2\beta}(x) g(x) dx - \boldsymbol{\xi}_\beta(\theta) \boldsymbol{\xi}_\beta^T(\theta), \quad (6)$$

where $\boldsymbol{\xi}_\beta(\theta) = \int \mathbf{u}_\theta(x) f_\theta^\beta(x) g(x) dx$, and $\mathbf{I}_\theta(x) = -\frac{\partial}{\partial \theta^T} \mathbf{u}_\theta(x)$ is the so-called Fisher information function at the model. The following results, proved in Basu et al. (2011), form the basis of our subsequent developments. We assume that the conditions D1–D5 of Basu et al. (2011, p. 304) are true. Then the following results hold.

1. The minimum density power divergence estimating equation (4) has a consistent sequence of roots $\hat{\theta}_\beta$, i.e. $\hat{\theta}_\beta \xrightarrow[n \rightarrow \infty]{p} \theta_\beta$.
2. $n^{1/2}(\hat{\theta}_\beta - \theta_\beta)$ has an asymptotic multivariate normal distribution with mean vector zero and covariance matrix $\mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1}$, where $\mathbf{J} = \mathbf{J}_\beta(\theta_\beta)$, $\mathbf{K} = \mathbf{K}_\beta(\theta_\beta)$ are as in (5) and (6).

In Sections 3 and 4 we describe the proposed Wald-type statistics for the simple and composite null hypothesis respectively.

3 The Simple Null Hypothesis

Let X_1, \dots, X_n be a random sample of size n from a distribution modeled by the probability density function $f_\theta(x)$, where $\theta \in \Theta \subset \mathbb{R}^p$ and $x \in \mathcal{X}$, the sample space. In this section we define a family of Wald-type test statistics based on MDPDEs for testing the null hypothesis

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta \neq \theta_0, \quad (7)$$

which will henceforth be referred to as the proposed Wald-type statistics.

Definition 1. Let $\hat{\theta}_\beta$ be the MDPDE of θ . The family of proposed Wald-type test statistics for testing the null hypothesis in (7) is given by

$$W_n = n \left(\hat{\theta}_\beta - \theta_0 \right)^T \left(\mathbf{J}_\beta^{-1}(\theta_0) \mathbf{K}_\beta(\theta_0) \mathbf{J}_\beta^{-1}(\theta_0) \right)^{-1} \left(\hat{\theta}_\beta - \theta_0 \right), \quad (8)$$

where the matrices $\mathbf{J}_\beta(\theta_0)$ and $\mathbf{K}_\beta(\theta_0)$ are as defined in (5) and (6) respectively.

When $\beta = 0$, $\widehat{\boldsymbol{\theta}}_\beta$ coincides with the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, and $\mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0)\mathbf{K}_\beta(\boldsymbol{\theta}_0)\mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0)$ coincides with the inverse of the Fisher information matrix, and thus we recover the classical Wald statistic. Many interesting applications of the classical Wald test statistic can be seen in the statistical literature. For example, [Lemonte \(2013\)](#) has presented an application on generalized linear models with dispersion covariates.

The asymptotic distribution of W_n is presented in the next theorem. The result follows easily from the asymptotic distribution of MDPDEs considered in [Section 2](#), so we skip the proof.

Theorem 1. *The asymptotic null distribution of the proposed Wald-type test statistic given in [\(8\)](#) is a chi-square distribution with p degrees of freedom.*

In many cases the power function of this testing procedure cannot be calculated explicitly. In the following theorem we present a useful asymptotic result for approximating the power function of the proposed Wald-type test statistics given in [\(8\)](#).

Theorem 2. *Let $\boldsymbol{\theta}^*$ be the true value of parameter with $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$. Then the convergence*

$$n^{1/2} \left(l(\widehat{\boldsymbol{\theta}}_\beta) - l(\boldsymbol{\theta}^*) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p(\mathbf{0}_p, \sigma_W^2(\boldsymbol{\theta}^*))$$

holds, where

$$l(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \left(\mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \right)^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0),$$

and

$$\sigma_W^2(\boldsymbol{\theta}^*) = 4(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)^T \left(\mathbf{J}_\beta^{-1}(\boldsymbol{\theta}^*) \mathbf{K}_\beta(\boldsymbol{\theta}^*) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}^*) \right)^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\theta}_0). \quad (9)$$

Proof. See the Appendix. □

Remark. We approximate the power function β_{W_n} of the proposed Wald-type test statistics at $\boldsymbol{\theta}^*$ by

$$\begin{aligned} \beta_{W_n}(\boldsymbol{\theta}^*) &= P_{\boldsymbol{\theta}^*}(W_n > \chi_{p,\alpha}^2) \\ &= P_{\boldsymbol{\theta}^*} \left(l(\widehat{\boldsymbol{\theta}}_\beta) - l(\boldsymbol{\theta}^*) > \frac{\chi_{p,\alpha}^2}{n} - l(\boldsymbol{\theta}^*) \right) \\ &= P_{\boldsymbol{\theta}^*} \left(n^{1/2} \left(l(\widehat{\boldsymbol{\theta}}_\beta) - l(\boldsymbol{\theta}^*) \right) > n^{1/2} \left(\frac{\chi_{p,\alpha}^2}{n} - l(\boldsymbol{\theta}^*) \right) \right) \\ &= P_{\boldsymbol{\theta}^*} \left(n^{1/2} \frac{l(\widehat{\boldsymbol{\theta}}_\beta) - l(\boldsymbol{\theta}^*)}{\sigma_W(\boldsymbol{\theta}^*)} > \frac{n^{1/2}}{\sigma_W(\boldsymbol{\theta}^*)} \left(\frac{\chi_{p,\alpha}^2}{n} - l(\boldsymbol{\theta}^*) \right) \right) \\ &= 1 - \Phi_n \left(\frac{n^{1/2}}{\sigma_W(\boldsymbol{\theta}^*)} \left(\frac{\chi_{p,\alpha}^2}{n} - l(\boldsymbol{\theta}^*) \right) \right), \end{aligned} \quad (10)$$

where $\Phi_n(\cdot)$ is a sequence of distribution functions tending uniformly to the standard normal distribution function $\Phi(\cdot)$. Here $\chi_{p,\alpha}^2$ is the $100(1-\alpha)$ percentile point of chi-square distribution with p degrees of freedom, i.e. $P(\chi_p^2 > \chi_{p,\alpha}^2) = \alpha$. It is clear that

$$\lim_{n \rightarrow \infty} \beta_{W_n}(\boldsymbol{\theta}^*) = 1,$$

for all $\alpha \in (0, 1)$. Therefore, the test is consistent in the sense of [Fraser \(1957\)](#).

Based on this result given in [\(10\)](#) we can calculate the required sample size in order to get a given value for the power. If we want to get a power of $\beta^* = \beta_{W_n}(\boldsymbol{\theta}^*)$, we need a sample size n given by

$$n = [n^*] + 1, \quad n^* = \frac{A + B + \sqrt{A(A + 2B)}}{2l^2(\boldsymbol{\theta}^*)}, \quad (11)$$

where $[x]$ is the largest integer less than or equal to x , $A = \sigma_W^2(\boldsymbol{\theta}^*) (\Phi^{-1}(1 - \beta^*))^2$ and $B = \frac{1}{2} \chi_{p,\alpha}^2 l(\boldsymbol{\theta}^*)$.

Now we are going to derive the asymptotic distribution of W_n under contiguous alternative hypotheses described by

$$H_{1,n} : \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2} \mathbf{d}, \quad (12)$$

where \mathbf{d} is a fixed vector in \mathbb{R}^p such that $\boldsymbol{\theta}_n \in \Theta \subset \mathbb{R}^p$ for all n .

Theorem 3. *Under the contiguous alternative hypotheses given in (12) the asymptotic distribution of the proposed Wald-type test statistic W_n is a non-central chi-square with p degrees of freedom and non-centrality parameter*

$$\delta = \mathbf{d}^T \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{d}. \quad (13)$$

Proof. See the Appendix. □

Remark. Using the last theorem we can get an approximation to the power function at the contiguous alternative $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2} \mathbf{d}$ through the relation

$$\beta_{W_n}(\boldsymbol{\theta}_n) = 1 - G_{\chi_p^2(\delta)}(\chi_{p,\alpha}^2),$$

where $G_{\chi_p^2(\delta)}(\chi_{p,\alpha}^2)$ is the distribution function of a non-central chi-square, with p degrees of freedom and non-centrality parameter δ given in equation (13), evaluated at the point $\chi_{p,\alpha}^2$.

It may be observed that this expression permits us to obtain an approximation of the power function at a generic point $\boldsymbol{\theta}^*$, because we can consider $\mathbf{d} = n^{1/2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)$ and then $\boldsymbol{\theta}_n = \boldsymbol{\theta}^*$.

Example 1. Let X_1, \dots, X_n be a random sample of size n from $\mathcal{Exp}(\theta)$, the exponential distribution with mean θ . Suppose we want to test the hypothesis $H_0 : \theta = \theta_0$, against $H_1 : \theta \neq \theta_0$. It is easy to show that the minimum density power divergence estimator $\hat{\theta}_\beta$ of θ can be obtained by iteratively solving the equation

$$\hat{\theta}_\beta = \frac{\sum_{i=1}^n X_i \exp\left\{-\frac{\beta X_i}{\theta_\beta}\right\}}{\sum_{i=1}^n \exp\left\{-\frac{\beta X_i}{\theta_\beta}\right\} - \frac{n\beta}{(1+\beta)^2}}. \quad (14)$$

Note that, putting $\beta = 0$ in equation (14), we get an explicit expression for the MLE as $\hat{\theta}_0 = n^{-1} \sum_{i=1}^n X_i$. Standard calculation with the matrices $\mathbf{J}_\beta(\boldsymbol{\theta}_0)$ and $\mathbf{K}_\beta(\boldsymbol{\theta}_0)$, as defined in (5) and (6) respectively, gives us

$$n^{1/2} \left(\hat{\theta}_\beta - \theta_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, h(\beta) \theta_0^2),$$

where

$$h(\beta) = \frac{(1 + \beta)^2 P(\beta)}{(1 + \beta^2)^2 (1 + 2\beta)^3}, \text{ and } P(\beta) = 1 + 4\beta + 9\beta^2 + 14\beta^3 + 13\beta^4 + 8\beta^5 + 4\beta^6. \quad (15)$$

Therefore, for testing hypothesis $H_0 : \theta = \theta_0$, against $H_1 : \theta \neq \theta_0$, the proposed Wald-type test statistic is given by

$$W_n = \frac{n(\hat{\theta}_\beta - \theta_0)^2}{h(\beta) \theta_0^2}, \quad (16)$$

whose asymptotic distribution, under the null hypothesis, is a chi-square with one degree of freedom.

4 The Composite Null Hypothesis

We will now consider the problem of testing composite null hypothesis. In testing of a composite null hypothesis, the restricted parameter space $\Theta_0 \subset \Theta$ is often defined by a set of r restrictions of the form

$$\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}_r, \quad (17)$$

where $\mathbf{m} : \mathbb{R}^p \rightarrow \mathbb{R}^r$ (see [Serfling, 1980](#)). Assume that the $p \times r$ matrix

$$\mathbf{M}(\boldsymbol{\theta}) = \frac{\partial \mathbf{m}^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (18)$$

exists and is continuous in $\boldsymbol{\theta}$, and $\text{rank}(\mathbf{M}(\boldsymbol{\theta})) = r$, where $r \leq p$. Let X_1, \dots, X_n be a random sample of size n from a distribution modeled by probability density function $f_{\boldsymbol{\theta}}(x)$, where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ and $x \in \mathcal{X}$, the sample space. Our interest is in testing the hypothesis

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \text{ against } H_1 : \boldsymbol{\theta} \notin \Theta_0, \quad (19)$$

where Θ_0 is a subset of the parameter space Θ .

Definition 2. Let $\widehat{\boldsymbol{\theta}}_{\beta}$ be the MDPDE of $\boldsymbol{\theta}$. The family of proposed Wald-type test statistics for testing the null hypothesis in (19) is given by

$$W_n = n \mathbf{m}^T(\widehat{\boldsymbol{\theta}}_{\beta}) \left[\mathbf{M}^T(\widehat{\boldsymbol{\theta}}_{\beta}) \mathbf{J}_{\beta}^{-1}(\widehat{\boldsymbol{\theta}}_{\beta}) \mathbf{K}_{\beta}(\widehat{\boldsymbol{\theta}}_{\beta}) \mathbf{J}_{\beta}^{-1}(\widehat{\boldsymbol{\theta}}_{\beta}) \mathbf{M}(\widehat{\boldsymbol{\theta}}_{\beta}) \right]^{-1} \mathbf{m}(\widehat{\boldsymbol{\theta}}_{\beta}), \quad (20)$$

where the matrices \mathbf{J}_{β} , \mathbf{K}_{β} and \mathbf{M} are defined in equations (5), (6) and (18) respectively, and the function \mathbf{m} is defined in (17).

As in the simple null hypotheses case, when $\beta = 0$, the Wald-type test statistic reduces to the classical Wald test. In the next theorem we present the asymptotic distribution of W_n .

Theorem 4. *The asymptotic null distribution of the proposed Wald-type test statistics given in (20) is chi-square with r degrees of freedom.*

Proof. See the Appendix. □

It should be noted that the validity of the results presented are based on a crucial assumption that the unknown parameter vector $\boldsymbol{\theta}$ is an interior point in Θ . A common violation occurs when the null hypothesis constrains some of the parameters to lie on the boundary of the parameter space. Asymptotic theory under nonstandard conditions, in the case of MLE and likelihood ratio test, has been developed in [Chernoff \(1954\)](#), [Chant \(1974\)](#), [Self and Liang \(1987\)](#), [Susko \(2013\)](#) and references therein. One could, along similar lines, develop a theory for testing null hypothesis using proposed Wald-type statistics involving MDPDEs that cover parameters in the boundary of the parameter space. This will require a detailed investigation appropriate for a separate paper, and has not been carried out in the present work. For the rest of the paper we assume that the unknown parameter lies in the interior of the parameter space.

Consider the null hypothesis $H_0 : \boldsymbol{\theta} \in \Theta_0 \subset \Theta$. By Theorem 4, the null hypothesis is rejected if $W_n \geq \chi_{r, \alpha}^2$. The following theorem can be used to approximate the power function. Assume that $\boldsymbol{\theta}^* \notin \Theta_0$ is the true value of the parameter so that the unrestricted estimator $\widehat{\boldsymbol{\theta}}_{\beta} \xrightarrow[n \rightarrow \infty]{p} \boldsymbol{\theta}^*$.

Theorem 5. *Suppose*

$$l^*(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{m}^T(\boldsymbol{\theta}_1) \left[\mathbf{M}^T(\boldsymbol{\theta}_2) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_2) \mathbf{K}_{\beta}(\boldsymbol{\theta}_2) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_2) \mathbf{M}(\boldsymbol{\theta}_2) \right]^{-1} \mathbf{m}(\boldsymbol{\theta}_1).$$

Then

$$n^{1/2} \left(l^*(\widehat{\boldsymbol{\theta}}_{\beta}, \widehat{\boldsymbol{\theta}}_{\beta}) - l^*(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, \sigma_W^2(\boldsymbol{\theta}^*)),$$

where

$$\sigma_W^2(\boldsymbol{\theta}^*) = \left(\frac{\partial l^*(\boldsymbol{\theta}, \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^T \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}^*) \mathbf{K}_{\beta}(\boldsymbol{\theta}^*) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}^*) \left(\frac{\partial l^*(\boldsymbol{\theta}, \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}. \quad (21)$$

Proof. See the Appendix. □

We may also find an approximation of the power of W_n at an alternative close to the null hypothesis. Let $\boldsymbol{\theta}_n \in \Theta - \Theta_0$ be a given alternative, and let $\boldsymbol{\theta}_0$ be the element in Θ_0 closest to $\boldsymbol{\theta}_n$ in terms of the Euclidean distance. One possibility to introduce contiguous alternative hypotheses in this context is to consider a fixed $\mathbf{d} \in \mathbb{R}^p$ and to permit $\boldsymbol{\theta}_n$ move towards $\boldsymbol{\theta}_0$ as n increases through the relation

$$H_{1,n} : \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\mathbf{d}. \quad (22)$$

A second approach is to relax the condition $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}_r$ defining Θ_0 . Let $\boldsymbol{\delta} \in \mathbb{R}^r$ and consider the following sequence of parameters $\{\boldsymbol{\theta}_n\}$ moving towards $\boldsymbol{\theta}_0$ according to the set up

$$H_{1,n}^* : \mathbf{m}(\boldsymbol{\theta}_n) = n^{-1/2}\boldsymbol{\delta}. \quad (23)$$

Note that a Taylor series expansion of $\mathbf{m}(\boldsymbol{\theta}_n)$ around $\boldsymbol{\theta}_0$ yields

$$\mathbf{m}(\boldsymbol{\theta}_n) = \mathbf{m}(\boldsymbol{\theta}_0) + \mathbf{M}^T(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) + o(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|). \quad (24)$$

By substituting $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\mathbf{d}$ in (24) and taking into account that $\mathbf{m}(\boldsymbol{\theta}_0) = \mathbf{0}_r$, we get

$$\mathbf{m}(\boldsymbol{\theta}_n) = n^{-1/2}\mathbf{M}^T(\boldsymbol{\theta}_0)\mathbf{d} + o(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|). \quad (25)$$

So the equivalence relationship between the hypotheses $H_{1,n}$ and $H_{1,n}^*$ is

$$\boldsymbol{\delta} = \mathbf{M}^T(\boldsymbol{\theta}_0)\mathbf{d} \text{ as } n \rightarrow \infty. \quad (26)$$

In the following theorem we show the asymptotic distributions of the test statistics under the alternative hypotheses $H_{1,n}$ and $H_{1,n}^*$ as given in (22) and (23) respectively.

Theorem 6. *The asymptotic distribution of W_n is given by*

- i) $W_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_r^2 \left(\mathbf{d}^T \mathbf{M}(\boldsymbol{\theta}_0) \left[\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0) \right]^{-1} \mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{d} \right)$ under $H_{1,n}$ given in (22).
- ii) $W_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_r^2 \left(\boldsymbol{\delta}^T \left[\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0) \right]^{-1} \boldsymbol{\delta} \right)$ under $H_{1,n}^*$ given in (23).

Proof. See the Appendix. □

While our theory is perfectly general, we discuss two special cases in Sections 4.1 and 4.2 – those of the Normal and Weibull models – where immediate applications our theory can be envisaged.

4.1 The Normal Distribution

Under the $N(\mu, \sigma^2)$ model, consider the problem of testing

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0, \quad (27)$$

where σ^2 is an unknown nuisance parameter. In this case the parameter space is given by $\Theta = \{(\mu, \sigma) \in \mathbb{R}^2 | \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$, and the parameter space under the null distribution is $\Theta_0 = \{(\mu, \sigma) \in \mathbb{R}^2 | \mu = \mu_0, \sigma \in \mathbb{R}^+\}$. If we consider the function $m(\boldsymbol{\theta}) = \mu - \mu_0$, where $\boldsymbol{\theta} = (\mu, \sigma)^T$, the null hypothesis H_0 can be written as

$$H_0 : m(\boldsymbol{\theta}) = 0.$$

We observe that in our case $\mathbf{M}(\boldsymbol{\theta}) = (1, 0)^T$. Based on (4) and taking into account the fact that $f_{\boldsymbol{\theta}}(x)$ is the normal density with mean μ and variance σ^2 , the estimator $\hat{\boldsymbol{\theta}}_{\beta} = (\hat{\mu}_{\beta}, \hat{\sigma}_{\beta})^T$ of $\boldsymbol{\theta}$ is the solution of the system of nonlinear equations

$$\begin{cases} \frac{\partial}{\partial \mu} \frac{1}{\sigma^{\beta} (2\pi)^{\frac{\beta}{2}}} \left(\frac{1}{n\beta} \sum_{i=1}^n \exp \left\{ -\frac{\beta}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right\} - \frac{1}{(1+\beta)^{3/2}} \right) = 0, \\ \frac{\partial}{\partial \sigma} \frac{1}{\sigma^{\beta} (2\pi)^{\frac{\beta}{2}}} \left(\frac{1}{n\beta} \sum_{i=1}^n \exp \left\{ -\frac{\beta}{2} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right\} - \frac{1}{(1+\beta)^{3/2}} \right) = 0. \end{cases}$$

Simple calculations yield the expressions

$$\mathbf{J}_{\beta}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{1+\beta} (2\pi)^{\beta/2} \sigma^{2+\beta}} \begin{pmatrix} \frac{1}{1+\beta} & 0 \\ 0 & \frac{\beta^2+2}{(1+\beta)^2} \end{pmatrix},$$

and

$$\mathbf{K}_{\beta}(\boldsymbol{\theta}) = \frac{1}{\sigma^{2+2\beta} (2\pi)^{\beta}} \left(\frac{1}{(1+2\beta)^{3/2}} \begin{pmatrix} 1 & 0 \\ 0 & \frac{4\beta^2+2}{1+2\beta} \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & \frac{\beta^2}{(1+\beta)^3} \end{pmatrix} \right).$$

Therefore,

$$\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0) = \sigma^2 \frac{(\beta+1)^3}{(2\beta+1)^{\frac{3}{2}}},$$

and

$$\begin{aligned} W_n &= nm^T(\hat{\boldsymbol{\theta}}_{\beta}) \left(\mathbf{M}^T(\hat{\boldsymbol{\theta}}_{\beta}) \mathbf{J}_{\beta}^{-1}(\hat{\boldsymbol{\theta}}_{\beta}) \mathbf{K}_{\beta}(\hat{\boldsymbol{\theta}}_{\beta}) \mathbf{J}_{\beta}^{-1}(\hat{\boldsymbol{\theta}}_{\beta}) \mathbf{M}(\hat{\boldsymbol{\theta}}_{\beta}) \right)^{-1} m(\hat{\boldsymbol{\theta}}_{\beta}) \\ &= n \frac{(\hat{\mu}_{\beta} - \mu_0)^2 (2\beta+1)^{\frac{3}{2}}}{\hat{\sigma}_{\beta}^2 (\beta+1)^3}, \end{aligned}$$

and on the basis of Theorem 4, we have

$$W_n = n \frac{(\hat{\mu}_{\beta} - \mu_0)^2 (2\beta+1)^{\frac{3}{2}}}{\hat{\sigma}_{\beta}^2 (\beta+1)^3} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_1^2. \quad (28)$$

We observe that for $\beta = 0$ we get the classical Wald test statistic for testing the hypothesis mentioned in (27).

4.2 The Weibull Distribution

While the normal model is the most important model where the test statistic described in Section 4.1 would be useful, it is also important to explore the applicability of these tests in other models to demonstrate the general nature of the method. In testing composite hypotheses, therefore, we have included numerical results based on the Weibull distribution in our subsequent numerical study, together with the results on the normal model. Here we describe the proposed Wald-type test statistics for the Weibull case. The probability density function of $\mathcal{W}(\sigma, p)$, a two parameter Weibull distribution, is given by

$$f_{\boldsymbol{\theta}}(x) = \frac{p}{\sigma} \left(\frac{x}{\sigma} \right)^{p-1} \exp \left\{ - \left(\frac{x}{\sigma} \right)^p \right\}, \quad x > 0,$$

where $\boldsymbol{\theta} = (\sigma, p)^T$, and the parameter space is given by $\Theta = \{(\sigma, p) | \sigma \in \mathbb{R}^+, p \in \mathbb{R}^+\}$. A comprehensive comparison of different estimation methods for the Weibull distribution is given in Teimouri et al. (2013). We are interested in testing

$$H_0 : \sigma = \sigma_0 \text{ against } H_1 : \sigma \neq \sigma_0, \quad (29)$$

where p is a nuisance parameter. Let us consider the function $m(\boldsymbol{\theta}) = \sigma - \sigma_0$. Then, as in the normal case which was considered in Section 4.1, the null hypothesis H_0 can be written as

$$H_0 : m(\boldsymbol{\theta}) = 0,$$

and $\mathbf{M}(\boldsymbol{\theta}) = (1, 0)^T$.

Let us define

$$\xi_{\alpha,\beta}(\boldsymbol{\theta}) = \int_0^\infty \left(\frac{x}{\sigma}\right)^\alpha f_{\boldsymbol{\theta}}^\beta(x) dx,$$

and

$$\eta_{\alpha,\beta,\gamma}(\boldsymbol{\theta}) = \int_0^\infty \left(\frac{x}{\sigma}\right)^\alpha \left[\log\left(\frac{x}{\sigma}\right)\right]^\beta f_{\boldsymbol{\theta}}^\gamma(x) dx.$$

It can be shown that

$$\xi_{\alpha,\beta}(\boldsymbol{\theta}) = \left(\frac{p}{\sigma}\right)^{\beta-1} \beta^{-\frac{\beta p - \beta + \alpha + 1}{p}} \Gamma\left(\frac{\beta p - \beta + \alpha + 1}{p}\right), \quad (30)$$

and

$$\eta_{\alpha,\beta,\gamma}(\boldsymbol{\theta}) = \sigma \left(\frac{p}{\sigma}\right)^\gamma \int_0^\infty y^{\alpha+\gamma p - \gamma} (\log y)^\beta \exp(-\gamma y^p) dy, \quad (31)$$

where $\Gamma(\cdot)$ denote the gamma function. Note that $\xi_{\alpha,\gamma}(\boldsymbol{\theta}) = \eta_{\alpha,0,\gamma}(\boldsymbol{\theta})$. For $\beta \neq 0$ the value of $\eta_{\alpha,\beta,\gamma}(\boldsymbol{\theta})$ is calculated using numerical integration. Let us define

$$\mathbf{R}_\gamma(\boldsymbol{\theta}) = \int_0^\infty \mathbf{u}_\theta(x) \mathbf{u}_\theta^T(x) f_{\boldsymbol{\theta}}^\gamma(x) dx = \begin{pmatrix} r_{11} & r_{12} \\ r_{12} & r_{21} \end{pmatrix},$$

where $\mathbf{u}_\theta(x)$, the score function of the Weibull distribution, is given by

$$\mathbf{u}_\theta(x) = \frac{\partial \log f_{\boldsymbol{\theta}}(x)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -\frac{p}{\sigma} + \frac{p}{\sigma} \left(\frac{x}{\sigma}\right)^p \\ \frac{1}{p} + \log\left(\frac{x}{\sigma}\right) - \left(\frac{x}{\sigma}\right)^p \log\left(\frac{x}{\sigma}\right) \end{pmatrix}.$$

Then it can be shown that

$$r_{11} = \left(\frac{p}{\sigma}\right)^2 \{\xi_{0,\gamma}(\boldsymbol{\theta}) - 2\xi_{p,\gamma}(\boldsymbol{\theta}) + \xi_{2p,\gamma}(\boldsymbol{\theta})\},$$

$$r_{12} = \frac{p}{\sigma} \left\{ -\frac{1}{p} \xi_{0,\gamma}(\boldsymbol{\theta}) - \eta_{0,1,\gamma}(\boldsymbol{\theta}) + 2\eta_{p,1,\gamma}(\boldsymbol{\theta}) + \frac{1}{p} \xi_{p,\gamma}(\boldsymbol{\theta}) - \eta_{2p,1,\gamma}(\boldsymbol{\theta}) \right\},$$

and

$$r_{22} = \frac{1}{p^2} \xi_{0,\gamma}(\boldsymbol{\theta}) + \eta_{0,2,\gamma}(\boldsymbol{\theta}) + \eta_{2p,2,\gamma}(\boldsymbol{\theta}) + \frac{2}{p} \eta_{0,1,\gamma}(\boldsymbol{\theta}) - 2\eta_{p,2,\gamma}(\boldsymbol{\theta}) - \frac{2}{p} \eta_{p,1,\gamma}(\boldsymbol{\theta}).$$

Now

$$\mathbf{J}_\gamma(\boldsymbol{\theta}) = \int_0^\infty \mathbf{u}_\theta(x) \mathbf{u}_\theta^T(x) f_{\boldsymbol{\theta}}^{1+\gamma}(x) dx = \mathbf{R}_{1+\gamma}(\boldsymbol{\theta}), \quad (32)$$

$$\mathbf{K}_\gamma(\boldsymbol{\theta}) = \int_0^\infty \mathbf{u}_\theta(x) \mathbf{u}_\theta^T(x) f_{\boldsymbol{\theta}}^{1+2\gamma}(x) dx = \mathbf{R}_{1+2\gamma}(\boldsymbol{\theta}). \quad (33)$$

Then some little algebra shows that the proposed Wald-type test statistic is given by

$$W_n = n(\hat{\sigma}_\beta - \sigma_0)^2 \left[\mathbf{M}^T(\boldsymbol{\theta}) \mathbf{J}_\beta^{-1}(\hat{\sigma}_\beta, \hat{p}_\beta) \mathbf{K}_\beta(\hat{\sigma}_\beta, \hat{p}_\beta) \mathbf{J}_\beta^{-1}(\hat{\sigma}_\beta, \hat{p}_\beta) \mathbf{M}(\boldsymbol{\theta}) \right]^{-1}$$

$$= \frac{n(\hat{\sigma}_\beta - \sigma_0)^2 \left[r_{\beta+1}^{(1,1)}(\hat{\sigma}_\beta, \hat{p}_\beta) r_{\beta+1}^{(2,2)}(\hat{\sigma}_\beta, \hat{p}_\beta) - \left(r_{\beta+1}^{(1,2)}(\hat{\sigma}_\beta, \hat{p}_\beta) \right)^2 \right]^2}{\begin{pmatrix} -r_{\beta+1}^{(2,2)}(\hat{\sigma}_\beta, \hat{p}_\beta) & r_{\beta+1}^{(1,2)}(\hat{\sigma}_\beta, \hat{p}_\beta) \\ r_{\beta+1}^{(1,2)}(\hat{\sigma}_\beta, \hat{p}_\beta) & r_{\beta+1}^{(2,2)}(\hat{\sigma}_\beta, \hat{p}_\beta) \end{pmatrix} \begin{pmatrix} -r_{\beta+1}^{(2,2)}(\hat{\sigma}_\beta, \hat{p}_\beta) \\ r_{\beta+1}^{(1,2)}(\hat{\sigma}_\beta, \hat{p}_\beta) \end{pmatrix}}$$

$$= \frac{n \left\{ (\hat{\sigma}_\beta - \sigma_0) \left(\frac{\hat{p}_\beta}{\hat{\sigma}_\beta} \right)^\beta \left[\tilde{r}_{\beta+1}^{(1,1)}(\hat{p}_\beta) \tilde{r}_{\beta+1}^{(2,2)}(\hat{p}_\beta) - \left(\tilde{r}_{\beta+1}^{(1,2)}(\hat{p}_\beta) \right)^2 \right] \right\}^2}{\begin{pmatrix} -\tilde{r}_{\beta+1}^{(2,2)}(\hat{p}_\beta) & \tilde{r}_{\beta+1}^{(1,2)}(\hat{p}_\beta) \\ \tilde{r}_{\beta+1}^{(1,2)}(\hat{p}_\beta) & \tilde{r}_{\beta+1}^{(2,2)}(\hat{p}_\beta) \end{pmatrix} \begin{pmatrix} -\tilde{r}_{\beta+1}^{(2,2)}(\hat{p}_\beta) \\ \tilde{r}_{\beta+1}^{(1,2)}(\hat{p}_\beta) \end{pmatrix}},$$

where

$$\begin{aligned}
\tilde{r}_\gamma^{(1,1)}(\hat{p}_\beta) &= \varepsilon_{0,\gamma}(\hat{p}_\beta) - 2\varepsilon_{\hat{p}_\beta,\gamma}(\hat{p}_\beta) + \varepsilon_{2\hat{p}_\beta,\gamma}(\hat{p}_\beta), \\
\tilde{r}_\gamma^{(1,2)}(\hat{p}_\beta) &= -\frac{1}{\hat{p}_\beta}\varepsilon_{0,\gamma}(\hat{p}_\beta) + \left(\log \hat{p}_\beta + \frac{1}{\hat{p}_\beta}\right)\varepsilon_{\hat{p}_\beta,\gamma}(\hat{p}_\beta) - \log \hat{p}_\beta \varepsilon_{2\hat{p}_\beta,\gamma}(\hat{p}_\beta) - \kappa_{0,1,\gamma}(\hat{p}_\beta) + \kappa_{\hat{p}_\beta,1,\gamma}(\hat{p}_\beta), \\
\tilde{r}_\gamma^{(2,2)}(\hat{p}_\beta) &= \frac{1}{\hat{p}_\beta}\varepsilon_{0,\gamma}(\hat{p}_\beta) - \frac{2}{\hat{p}_\beta}\log \hat{p}_\beta \varepsilon_{\hat{p}_\beta,\gamma}(\hat{p}_\beta) + (\log \hat{p}_\beta)^2 \varepsilon_{2\hat{p}_\beta,\gamma}(\hat{p}_\beta) \\
&\quad + \frac{2}{\hat{p}_\beta}\kappa_{0,1,\gamma}(\hat{p}_\beta) + \kappa_{0,2,\gamma}(\hat{p}_\beta) - 2\log \hat{p}_\beta \kappa_{\hat{p}_\beta,1,\gamma}(\hat{p}_\beta), \\
\varepsilon_{\alpha,\gamma}(p) &= \gamma^{\frac{(p-1)\gamma+\alpha+1}{p}} \Gamma\left(\frac{(p-1)\gamma+\alpha+1}{p}\right), \\
\kappa_{\alpha,\delta,\gamma}(\hat{p}_\beta) &= \hat{p}_\beta \int_0^\infty (\log y)^\delta y^{(\hat{p}_\beta-1)\gamma+\alpha} \exp\{-\gamma y^{\hat{p}_\beta}\} dy.
\end{aligned}$$

5 Simulation Study

In this section we present a simulation study to analyze the behavior of the proposed Wald-type test statistics introduced in this paper with some classical procedures for the same problems. We pay special attention to the robustness issue.

5.1 The Case of the Simple Null Hypothesis

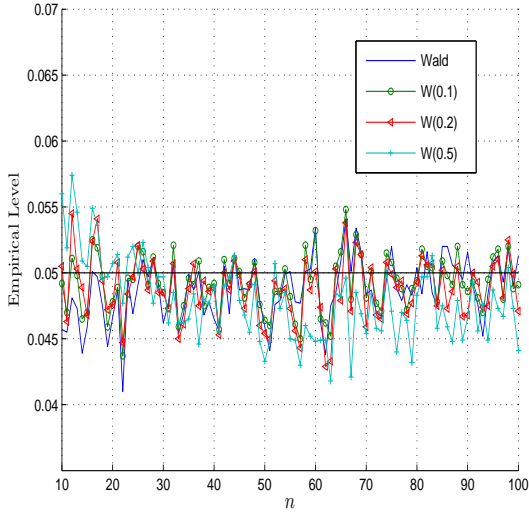
We have proposed Wald-type test statistics for the simple null hypothesis in (8). We will now study the performance of the test through simulation in case of the exponential model. The expression for the test statistic is simplified in (16). We want to test the hypothesis $H_0 : \theta = 2$ against the alternative $H_1 : \theta \neq 2$. In the first case we have generated data from the $\mathcal{Exp}(2)$ distribution, and the observed level (measured as the proportion of test statistics exceeding the chi-square critical value in 10,000 replications) are presented in Figure 1(a). We have taken three proposed Wald-type test statistics for $\beta = 0.1, 0.2$ and 0.5 , denoted by $W(\beta)$, and compared with the classical Wald test statistic. It is clear that the observed levels of all these tests are very close to the nominal level of 0.05 . Note that the classical Wald test statistic is a special case of the proposed family of test statistics corresponding to $\beta = 0$.

Next, the same hypotheses were tested when the data were generated from the $\mathcal{Exp}(1)$ distribution. The observed power (obtained in a similar manner as above) of the test is presented in Figure 1(b) under the alternative hypothesis. It is noticed that the classical Wald statistic performs best in terms of the power of the test, and for large β the test statistics show slightly poor performance. However, this discrepancy decreases rapidly with the sample size, and by the time $n \geq 50$, all the observed powers are practically equal to one. In any case for $\beta = 0.1$ or 0.2 the tests are almost as powerful as the classical Wald test.

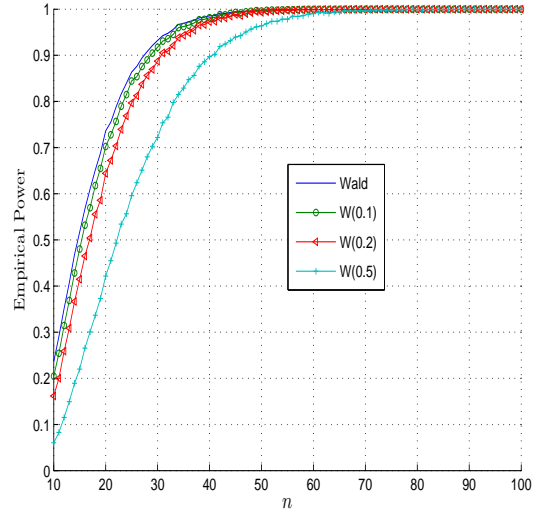
To evaluate the stability of the level and the power of the tests under contamination, we repeated the simulations for $H_0 : \theta = 2$ against $H_0 : \theta \neq 2$ with data generated from the exponential mixture $0.95\mathcal{Exp}(2) + 0.05\mathcal{Exp}(10)$, and also tested the same hypotheses with data generated from $0.95\mathcal{Exp}(1) + 0.05\mathcal{Exp}(10)$ distribution. The results are given in Figures 1(c) and 1(d) respectively.

In this case there is a huge inflation in the observed level of the classical Wald test statistic and somewhat smaller inflation in the tests with small values of β . But as β increases the resistant nature of the tests are clearly apparent. For larger values of β the levels turn out to be more acceptable. The opposite behavior is seen in case of power. There appears to be a complete breakdown in power for the classical Wald test, but the power remains much more stable for the proposed Wald-type test statistics for moderate and large values of β .

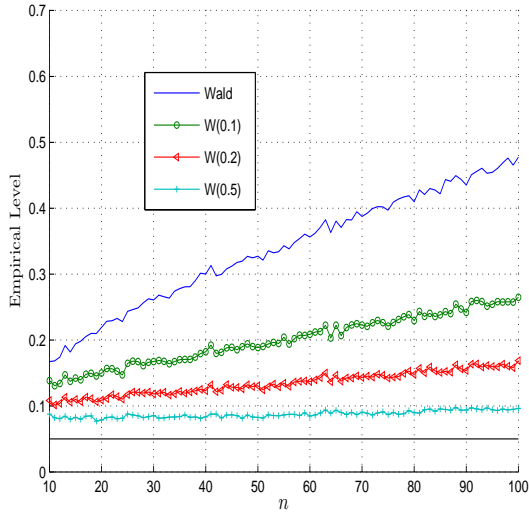
On the whole it appears to be fair to claim that for sample sizes equal to or larger than 40 the efficiency of many of our proposed Wald-type tests are very close to the efficiency of the classical Wald test, at least in the example studied here, but the robustness properties of our tests are often



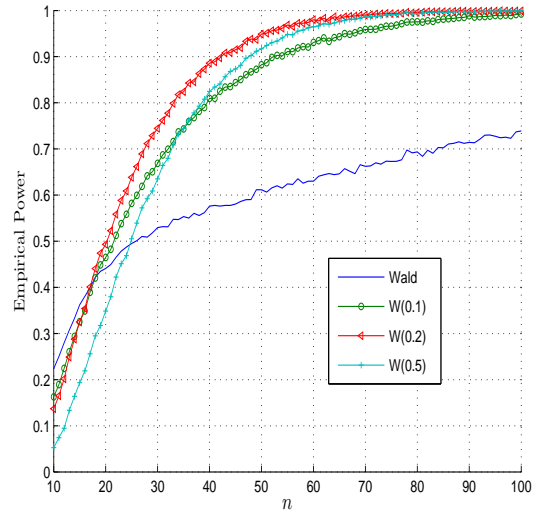
(a)



(b)



(c)



(d)

Figure 1: (a) Simulated levels of different tests for pure exponential data; (b) simulated powers of different tests for pure exponential data; (c) simulated levels of different tests for contaminated exponential data; (d) simulated powers of different tests for contaminated exponential data.

significantly better than the classical Wald test in terms of maintaining the stability of both the level and power.

5.2 The Case of the Composite Null Hypothesis

5.2.1 The Normal Case

To explore the performance of our proposed Wald-type test statistics in case of the composite null hypothesis, we have performed a simulation study for the family of test statistics given in Example 4.1. We considered the hypothesis $H_0 : \mu = 0$ against the alternative $H_1 : \mu \neq 0$ with σ^2 unknown when data were generated from the $\mathcal{N}(0, 1)$ distribution. Subsequently, the same hypotheses were tested when the data were generated from the $\mathcal{N}(-1, 1)$ distribution. The results are given in Figures 2(a) and 2(b). In either case the nominal level was 0.1, and all tests were replicated 10,000 times.

It may be noticed that all the tests with large values of β are slightly liberal for very small sample sizes and lead to somewhat inflated observed levels. However, the observed levels settle down reasonably with increasing sample size. The observed powers of the tests as given in Figure 2(b) are, in fact, extremely close. In very small sample sizes our proposed test statistics have slightly higher power than the Wald test, but this must be a consequence of the observed levels of these tests being higher than the latter for such sample sizes. On the whole the proposed Wald-type tests appear to be quite competitive to the classical Wald test for pure normal data.

Now, we show the performance of the proposed tests under contaminated data. So, we have tested the same hypothesis, but the data have been generated from the $0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(10, 1)$ distribution. The observed levels are shown in Figure 2(c). We notice that there is a drastic and severe inflation in the observed level of the classical Wald test. As β increases, however, the resistant nature of the tests are clearly apparent. By the time $\beta = 0.2$, the levels have already settled down to acceptable values.

Finally, we have generated data from the normal mixture $0.9\mathcal{N}(-1, 1) + 0.1\mathcal{N}(10, 1)$, and the power functions are plotted in Figure 2(d). There is a complete breakdown in power for the classical Wald test and the proposed Wald-type tests corresponding to the small values of β , but the power remains quite stable for values of β equal to 0.2 or greater.

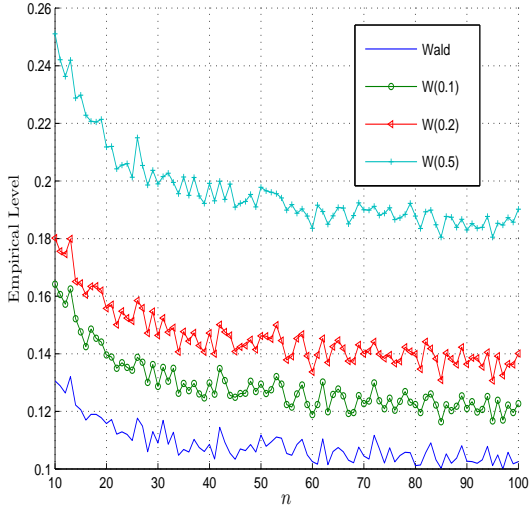
A similar conclusion as the previous simulation study can be drawn in this case. For sample sizes equal to or larger than 30 the efficiency of the proposed Wald-type tests for β greater than 0.2 are very close to the efficiency of the classical Wald test, but the robustness properties of those tests are much better than the Wald test in terms of both the level and power.

In this simulation we have taken the 10% significance level instead of the usual 5%. Moreover, the contamination proportion is also different from the previous section. This is just to demonstrate that our tests produce analogous results in a variety of simulation set ups. The findings for the 5% simulation results in this case, not presented here, are along similar lines.

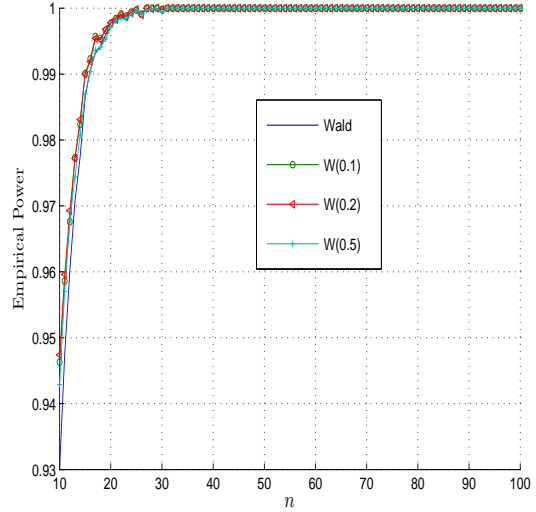
5.2.2 The Weibull Case

As we have mentioned before, it is important to demonstrate the properties of the proposed method in models other than the normal so that one has a better idea about the scope of the method. Accordingly we performed tests of composite hypotheses under the Weibull model in the spirit of Section 5.2.1. Let us consider the hypothesis defined in (29), where σ_0 is taken to be 1.5. In the first study we have generated data from the $\mathcal{W}(1.5, 1.5)$ distribution. The plot for the observed level for the hypothesis $H_0 : \sigma = 1.5$ against the two sided alternative is given in Figure 3(a), where we have used 10,000 replications. Next the same hypotheses were tested when the data were generated from the $\mathcal{W}(1, 1.5)$ distribution. The observed power function is plotted in Figure 3(b) for different values of β . The powers are remarkably close. In all cases the nominal level was 0.05.

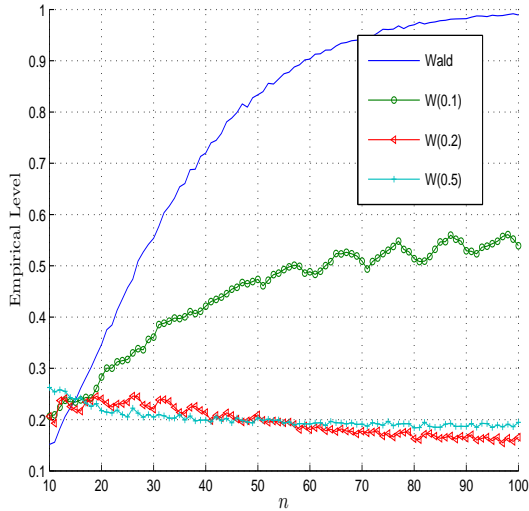
To evaluate the stability of the level and the power of the tests under contamination, we repeated the tests with data generated from the Weibull mixture $0.95\mathcal{W}(1.5, 1.5) + 0.05\mathcal{W}(10, 1.5)$, and then from $0.95\mathcal{W}(1, 1.5) + 0.05\mathcal{W}(10, 1.5)$. In either case the first, larger component is our target. In Figure 3(c), the levels of the statistics under the contamination of first type are presented indicating the stability of levels for moderately large values of β . Figure 3(d) demonstrates the stability of powers under contaminated data of the second type for the same values of β .



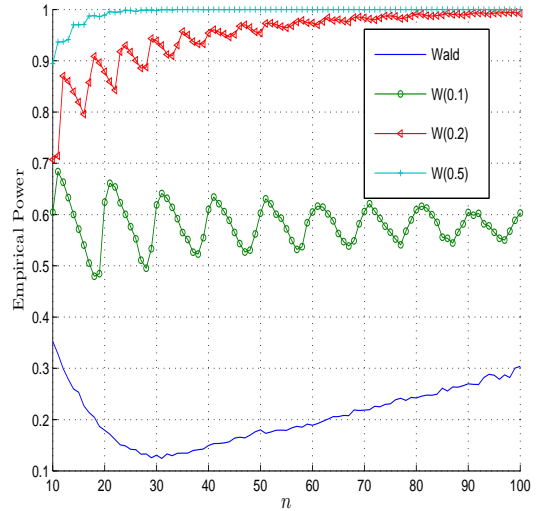
(a)



(b)

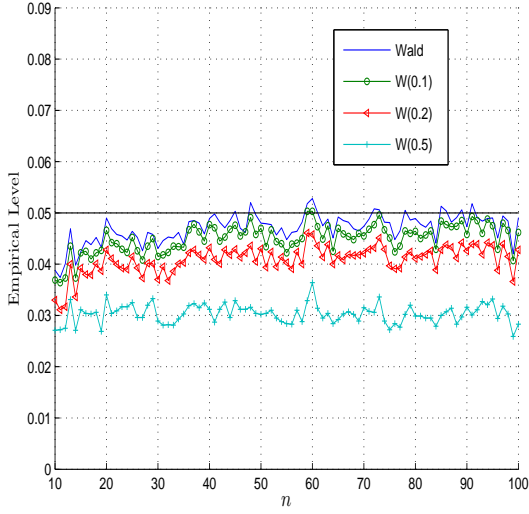


(c)

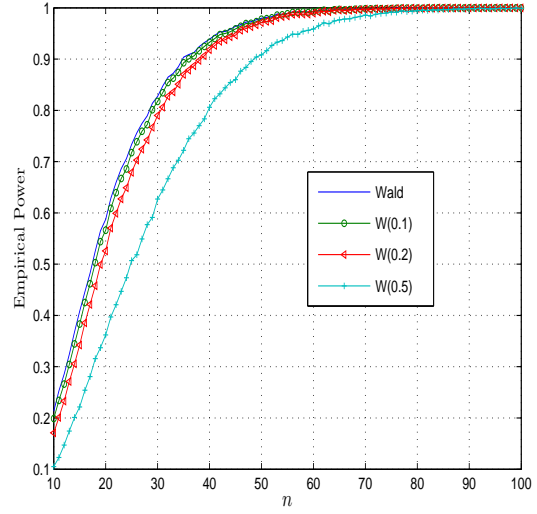


(d)

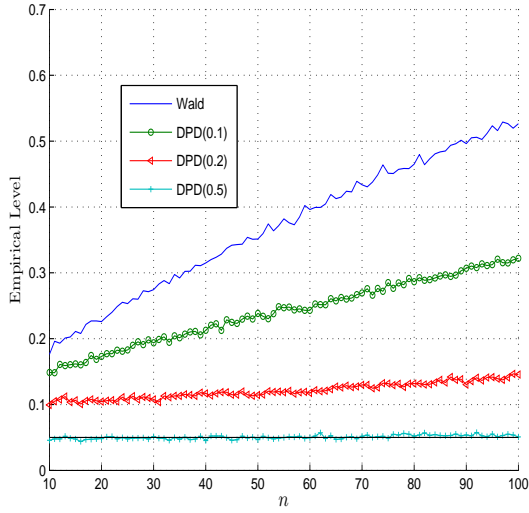
Figure 2: (a) Simulated levels of different tests for pure normal data; (b) simulated powers of different tests for pure normal data; (c) simulated levels of different tests for contaminated normal data; (d) simulated powers of different tests for contaminated normal data.



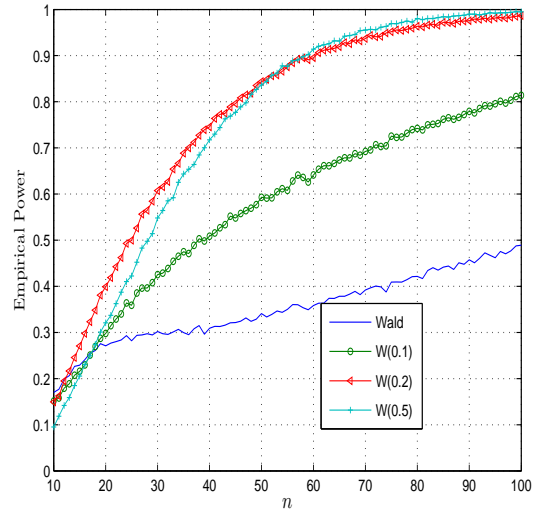
(a)



(b)



(c)



(d)

Figure 3: (a) Simulated levels of different tests for pure Weibull data; (b) simulated powers of different tests for pure Weibull data; (c) simulated levels of different tests for contaminated Weibull data; (d) simulated powers of different tests for contaminated Weibull data.

5.2.3 Comparison with Other Robust Tests

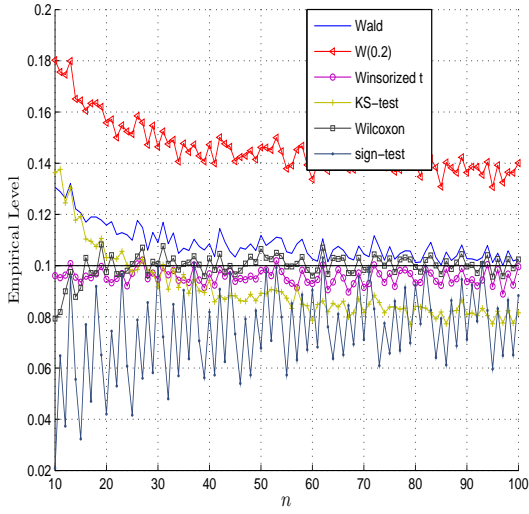
In this section we compare the proposed Wald-type test with some other popular robust tests. For comparison we have used a parametric test – the Winsorized t -test proposed by [Dixon and Tukey \(1968\)](#), as well as three non-parametric tests – the one-sample Kolmogorov-Smirnov test (KS-test), the two-sided Wilcoxon signed rank test and the two-sided sign test. The set up, the parameters taken for the simulation and the level of contamination are exactly the same as in [Section 5.2.1](#). For the Winsorized t -test we have winsorized 15% extreme observations from each side of the data set. It should be mentioned here that the null hypothesis is slightly different for the non-parametric tests. For the KS-test we first standardized the data using robust statistics. Then we test whether the corresponding distribution is a standard normal. The data are standardized using the transformation $Z = (X - \mu_0)/MadN$, where μ_0 is the value of μ under H_0 , and $MadN = 1.4826 \times$ median absolute deviation about the median. In case of Wilcoxon test and the sign test we test the null hypothesis that the median is zero, without making any assumption on the model distribution. For comparison we have used only one proposed Wald-type test in this case, that corresponding to tuning parameter 0.2. To emphasize the robustness properties of these tests we have also included the classical Wald test in this investigation. The results are presented in [Figure 4](#).

[Figure 4\(a\)](#) shows that the observed levels of the Winsorized t -test, the KS-test and Wilcoxon test are very close to the nominal level of 0.1 for the pure normal data. For small sample sizes the proposed Wald-type test is liberal, whereas the sign test is a little bit conservative. The observed powers of all the tests in [Figure 4\(b\)](#) rapidly converge to unity in very small sample sizes. The result in [Figure 4\(c\)](#) is very interesting; it shows, for the contaminated data, all the tests except the proposed Wald-type test for $\beta = 0.2$ fail to maintain the nominal level of 0.1. The observed level of the sign test is close to the nominal level for small sample sizes, but eventually as sample size increases it also breaks down. The powers of the tests for the contaminated data are plotted in [Figure 4\(d\)](#); all the robust tests show stable power. For the small sample sizes the proposed Wald-type test shows better power, however, it may be due to its inflated level in the pure data. Therefore, for the contaminated data, the proposed Wald-type test does significantly better than the others in holding on to its nominal level as well as power. On the whole, the proposed Wald-type tests are not only superior to the classical Wald test under contamination, but they also appear to be competitive or better than the other popular robust tests as far as this simulation study is concerned.

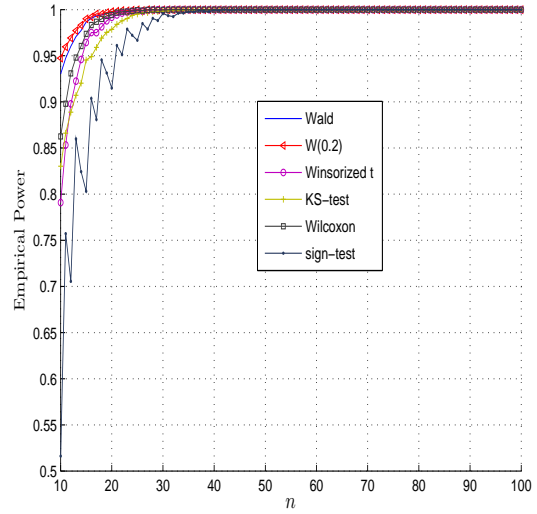
5.3 Real Data Examples

5.3.1 Leukemia Data

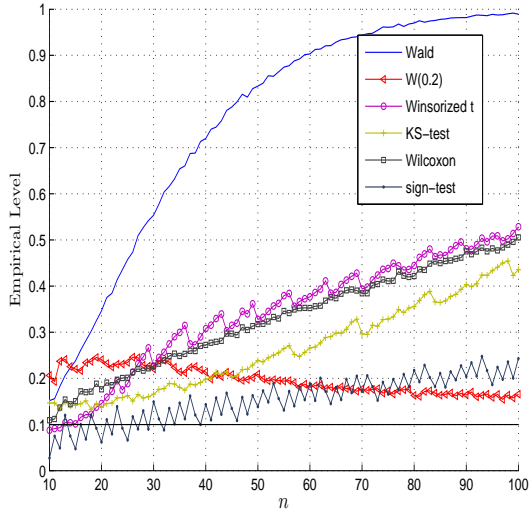
Let us consider the leukemia data set given in [Tiku et al. \(1986\)](#), and [Gross and Clark \(1975\)](#). [Table 1](#) gives (100 times) the white blood cell counts of patients who had acute myelogenous leukemia. We assume an exponential distribution with parameter θ models this data. One fact that is immediately noticeable is that there are two huge outlying observations at 1000 with respect to the exponential model, whereas other observations appear to be reasonable with respect to the same. The maximum likelihood estimator of θ for the full data is 246.41, but if we delete the two outliers it comes out to be 138.75. We consider testing the null hypothesis $H_0 : \theta = 140$ against $H_1 : \theta \neq 140$ for the full data. The p -value of the classical Wald test is 0.0024, so in the presence of the outliers the classical Wald test rejects the null hypothesis. For the outlier deleted data the p -value of the classical Wald test becomes 0.9733. This extreme turnaround illustrates that the classical Wald test is highly influenced by very small proportion of outlying observations. On the other hand, the proposed Wald-type tests for large values of β lead to similar inference in all situations. [Figure 5\(a\)](#) shows the p -values of the proposed Wald-type tests for the full data as well as for the outlier deleted data. This stable behavior of the test statistics based on MDPDEs for the full data approximately coincides with the stability of the minimum density power divergence estimates of θ itself, obtained under an exponential model, which is presented in [Figure 5\(b\)](#). The minimum density power divergence estimators for the full data and the outlier deleted data are practically identical for $\beta > 0.45$. So the robustness of the test statistic is clearly linked to the robustness of the estimator.



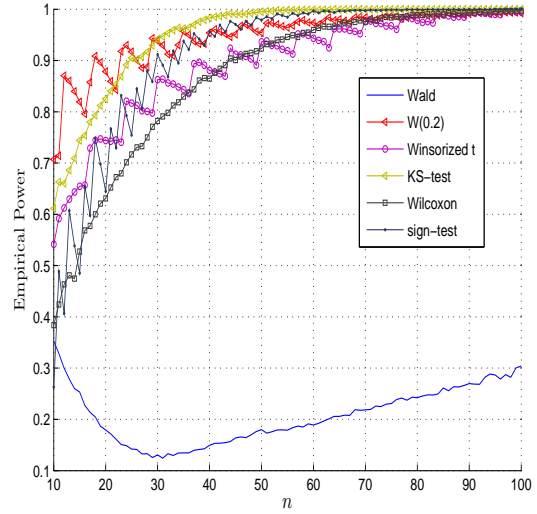
(a)



(b)



(c)



(d)

Figure 4: Simulated levels and powers of different robust tests for pure and contaminated data in case of the normal distribution.

Table 1: Leukemia Data

Patient	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Count	23	7.5	43	26	60	105	100	170	54	70	94	320	350	1000	520	1000

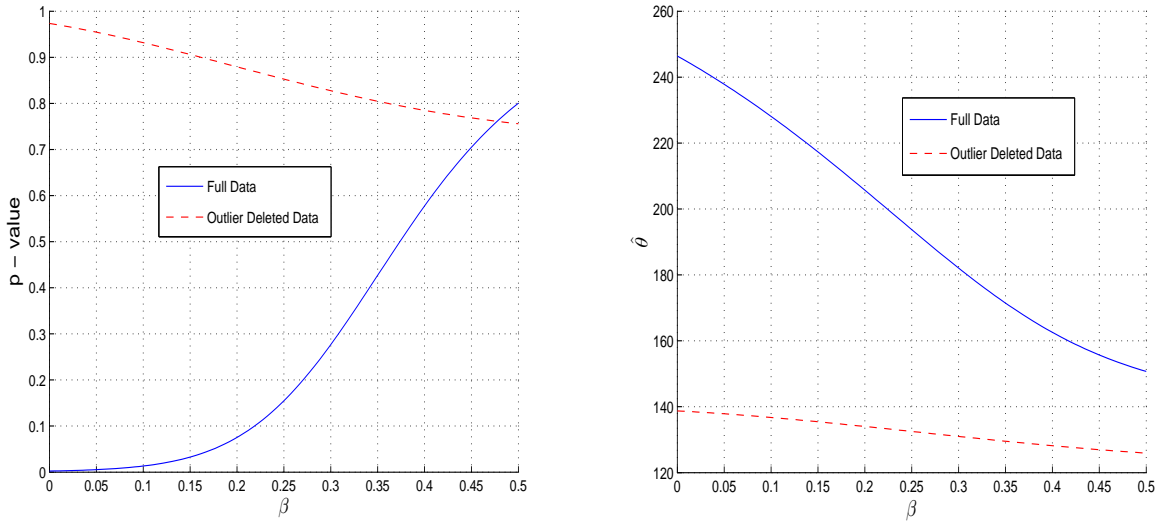


Figure 5: (a) Two sided p -values of the proposed Wald-type tests, and (b) estimates of θ for different values of β in case of the Leukemia data.

5.3.2 Telephone-Fault Data

Welch (1987) and Simpson (1989) have presented and analyzed data from an experiment to test a method of reducing faults on telephone lines. The data consist of the ordered differences between the inverse test rates and the inverse control rates in 14 matched pairs of areas (see Table 2). For simplicity these data are modeled as a random sample from a normal distribution with mean μ and standard deviation σ . Figure 6 presents a kernel density estimate for these data, the normal model fits using the estimates of μ and σ based on the maximum likelihood estimators, as well as the minimum density power divergence estimators (with tuning parameter 0.15), and the minimum Hellinger distance estimators (Simpson, 1989). The first observation of this dataset produces a small bump to the extreme left of the figure and is obviously a huge outlier with respect to the normal model; the remaining 13 observations appear to be reasonable with respect to the same. Therefore, if we ignore the first observation, these data would have a nice unimodal structure which could be well modeled by an appropriate normal density. Figure 6 shows that the minimum Hellinger distance estimates and the minimum density power divergence estimates automatically ignore the effect of the outlier and give good fits for the main part of the data. On the other hand, the maximum likelihood estimates try to be inclusive and generate a result which neither models the outlier deleted data, nor provides a fit to the outlier component.

Let us now consider testing of the null hypothesis $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$, where σ is unknown. For the full data, the classical Wald test fails to reject the null hypothesis due to the presence of the large outlier (two sided p -value is 0.6584); however the robust Hellinger deviance test (Simpson, 1989) comfortably rejects the null (two sided p -value based on the chi-square null distribution is 0.0061), as does the classical Wald test based on the cleaned data after the removal of the large outlier (two sided p -value is 0.0076).

Table 2: Telephone-Fault Data

Pair	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Difference	-988	-135	-78	3	59	83	93	110	189	197	204	229	289	310

Under the normal model, the usual estimates of μ (and σ) are highly inflated due to the presence

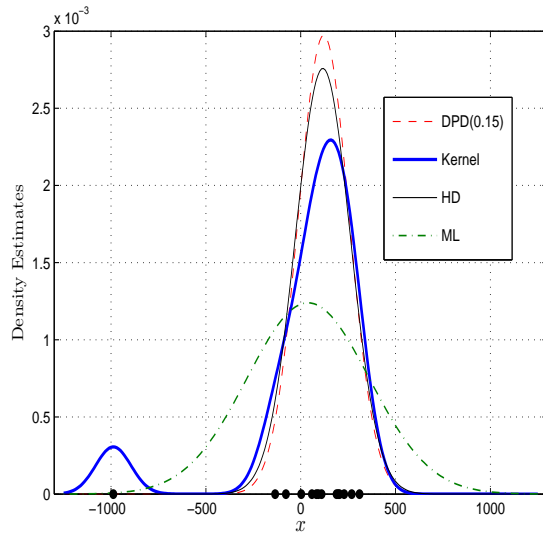


Figure 6: Kernel density estimate and different normal fits for the Telephone-Fault data.

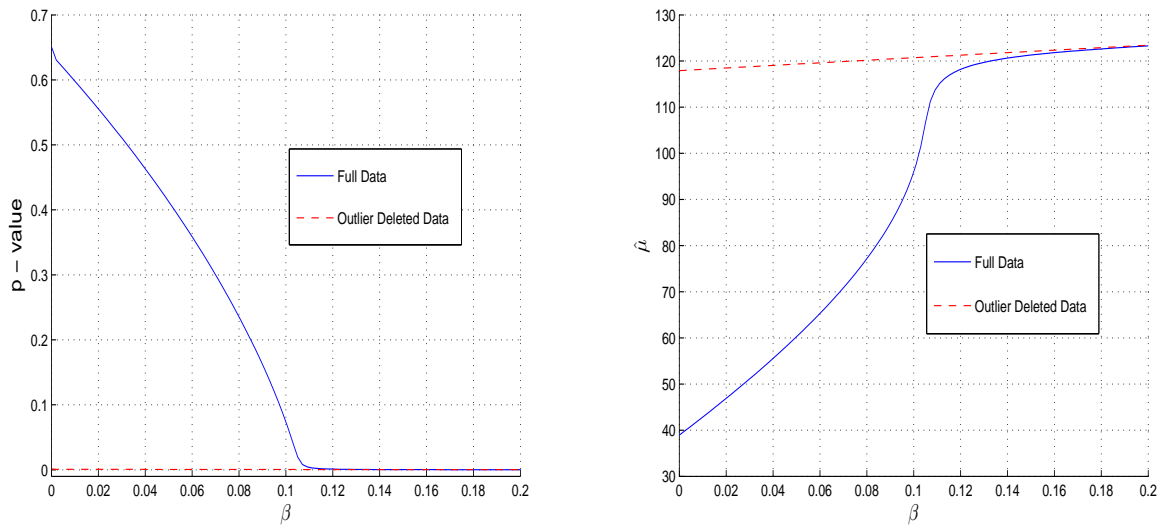


Figure 7: (a) Two sided p -values of the proposed Wald-type tests, and (b) estimates of μ for different values of β in case of the Telephone-Fault data.

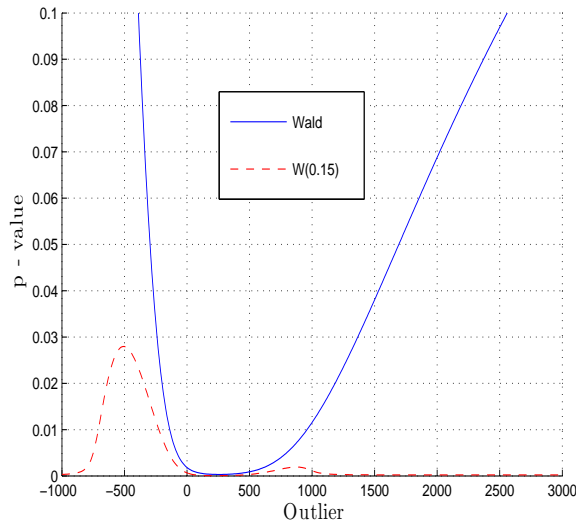


Figure 8: Two sided p -values for the tests for the mean for the Telephone-Fault data under the normal model against the first outlying observation.

of the large outlier, and as a result the likelihood ratio test under the normal model fails to reject the null hypothesis. From the robustness perspective, this is precisely what we like to avoid, and here we demonstrate that proper choices of the tuning parameter within the class of tests developed in this paper achieve this goal. Here we analyze the performance of the proposed Wald-type tests for different values of β . Figure 7(a) represents the p -values of the tests for different values of β in a region of interest. While it is clearly seen that the tests fail to reject the null hypothesis for these data at very small values of β , the decision turns around sharply, as β crosses and goes beyond 0.1. On the other hand, the p -values of the same test based on the outlier deleted data remain stable, supporting rejection, at all values of β as seen in Figure 7(a). The stable behavior of the test statistic based on the MDPDE for the full data approximately coincides with the stability of the MDPDE of μ itself, obtained under a two-parameter normal model, which is presented in Figure 7(b). The minimum density power divergence estimators for the full data and the outlier deleted data are practically identical for $\beta > 0.12$. At least in this example, the robustness of the test statistic is clearly linked to the robustness of the estimator.

To further explore the robustness properties of the proposed Wald-type test statistics we look at the two sided p -values for different values of the outlier. For this purpose we vary the first outlying observation in the range from -1000 to 3000 by keeping the remaining 13 observations fixed. Figure 8 shows the corresponding p -values of the density power divergence tests with $\beta = 0.15$ as well as the classical Wald test. It shows that initially the p -value of the density power divergence test with $\beta = 0.15$ increases as the first observation moves away from the center of the data set, but after a certain limit the test gradually nullifies the effect of the outlier. On the other hand, the p -values of the classical Wald test keep on increasing with the outlier on either tail.

5.3.3 Darwin's Plant Fertilization Data

Charles Darwin had performed an experiment which may be used to determine whether self-fertilized plants and cross-fertilized plants have different growth rates. In this experiment pairs of *Zea mays* plants, one self and the other cross-fertilized, were planted in pots, and after a specific time period the height of each plant was measured. A particular sample of 15 such pairs of plants led to the paired differences (cross-fertilized minus self fertilized) presented in increasing order in Table 3 (see Darwin,

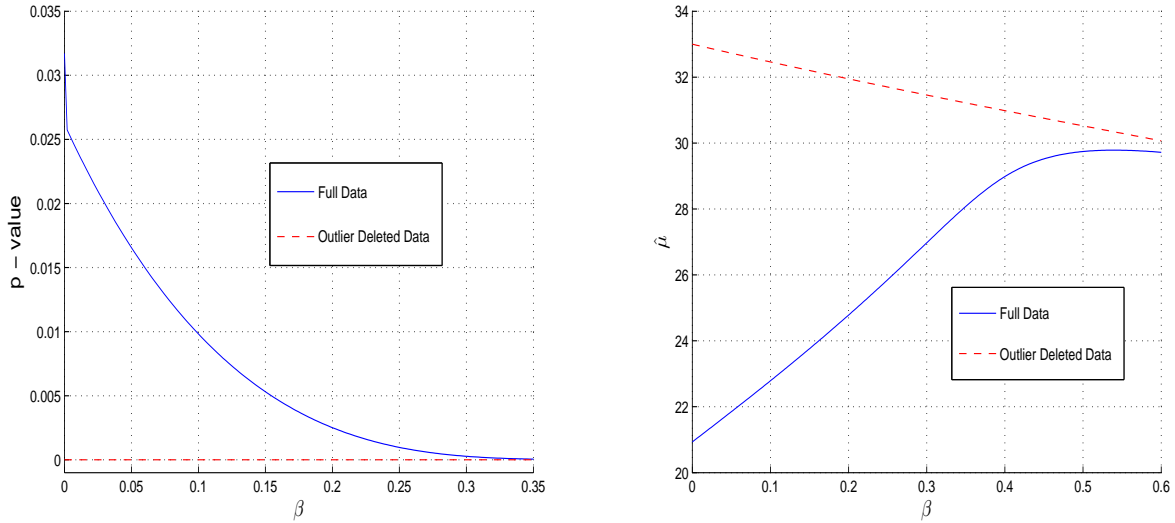


Figure 9: (a) Two sided p -values of the proposed Wald-type tests, and (b) estimates of μ for different values of β in case of Darwin's fertilization data.

1878).

Table 3: Darwin's Plant Fertilization Data

Pair	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Difference	-67	-48	6	8	14	16	23	24	28	29	41	49	56	60	75

As in the previous example, we assume a normal model for the paired differences and test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$, i.e. we test whether the mean of the paired differences is different from zero. The unconstrained MDPDEs of μ under the normal model corresponding to different values of the tuning parameter β are presented in Figure 9(b). The two negative paired differences appear to be geometrically well separated from the rest of the data, though they are perhaps not as huge outliers as the first observation in the Telephone-Fault data. These two observations do have a substantial impact on the parameter estimates and the test statistic for testing H_0 using the proposed Wald-type test statistic with very small values of β , and it is instructive to compare the analysis to the case where these two outliers have been removed from the data. For small values of β , the two sided p -values of the test statistics are drastically different for the full data and outlier deleted cases (Figure 9(a)), but they get closer with increasing β , and they essentially coincide for $\beta \geq 0.3$. Once again this seems to be directly linked to the robustness of the parameter estimates; Figure 9(b), which also depicts the progression of the parameter estimates for the outlier deleted data, clearly demonstrates that.

5.3.4 Air-conditioning System Failure Data

Proschan (1963) gives 213 values of successive failure time of air-conditioning system of each member of a fleet of Boeing 720 jet airplanes. After roughly 2000 hours of service the planes received a major overhaul; the failure interval containing major overhaul is omitted from the listing since the length of that failure interval may have been affected by major overhaul. The purpose of the data collection was to obtain information as to the distribution of failure intervals for the air conditioning systems of all planes pooled together. Obvious applications of such information would be in predicting reliability, scheduling maintenance, and providing spare parts. We have modeled the lifetime data as a random sample from a Weibull distribution with the shape parameter p and the scale parameter σ . However,

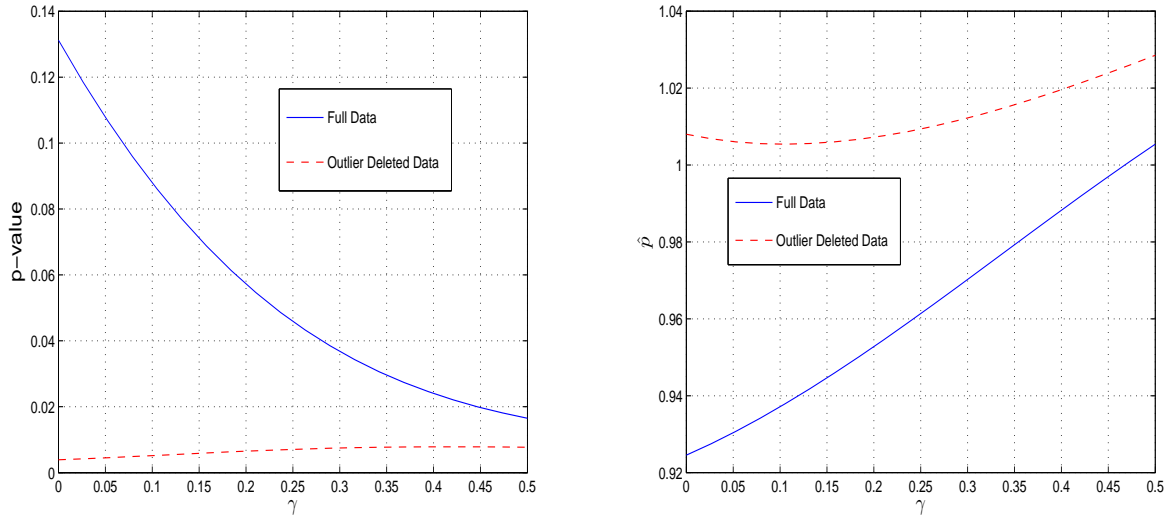


Figure 10: (a) Two sided p -values of the proposed Wald-type tests, and (b) estimates of p for different values of β in case of Air-conditioning Equipment Failure data.

there are few large observations in the data, which may be regarded as outliers with respect to the Weibull model. If we treat the observations which are greater than 400 (total count of such observations is only 7) as outliers we notice a clear difference in MLEs for the full data and the reduced data. The MDPDEs of p for different values of β are plotted in Figure 10(b). Now we expect that a similar phenomenon should be reflected in testing a hypothesis based on these estimators. We chose to test $H_0 : p = 0.85$ against $H_1 : p \neq 0.85$, where σ is unknown. Figure 10(a) gives the p -values of different proposed Wald-type tests for different values of β . For the outlier deleted data all the tests clearly reject the null hypothesis, but in case of full data the classical Wald test as well as the proposed Wald-type tests with small tuning parameter β (approximately $\beta \leq 0.2$) fail to reject the null hypothesis at 5% level of significance.

5.3.5 One Sided Tests

In many parametric hypothesis testing problems including testing for the mean under the normal model the case of real interest involves a one sided alternative. For the Telephone-Fault data problem, interest could lie in determining whether the mean fault rate is higher than zero (rather than simply whether it is different from zero). Darwin's interest in the fertilization problem was, presumably, to determine whether cross fertilization leads to a higher growth rate compared to self-fertilization; indeed the test performed on Darwin's data by R. A. Fisher (reported in Fisher, 1966) considers the one sided alternative. In this subsection we consider the relevant one sided alternatives for the two real data examples presented earlier in this section under the normal model. For this purpose we consider the signed Wald-type test statistic (the signed square root of the statistic presented in (28)) as in Simpson (1989), and determine the relevant (conservative) one sided p -value using the appropriate t -distribution. For a scalar parameter θ , given the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$, the proposed signed Wald-type statistic essentially turns out to be $W_n = \frac{n^{1/2}(\hat{\theta}_\beta - \theta_0)}{\hat{\sigma}_\beta}$, where $\hat{\sigma}_\beta^2$ is the estimated variance of $n^{1/2}\hat{\theta}_\beta$. The test rejects at level α when W_n exceeds the $100(1 - \alpha)$ th quantile of the t -distribution with $n - 1$ degrees of freedom (or the corresponding quantile of the standard normal distribution when n is large). Similarly, one gets a critical region in the lower tail for the less than type alternative.

Telephone-Fault data: Simpson (1989) has reported the one sided p -value (for the greater than alternative) in case of the Hellinger deviance test for the mean under the normal model to be 0.0085

for the full data and 0.0093 for the outlier deleted data. The one sided p -values for proposed signed Wald-type test statistics corresponding to $\beta = 0.15$ and $\beta = 0.30$ are 0.0033 and 0.0020 respectively for the full data, and 0.0040 and 0.0030 respectively for the outlier deleted data, computed under the $t(13)$ distribution. On the other hand, the p -values for the full data and the outlier deleted data are 0.33 and 0.004 for the classical one sided matched pairs Wald test. Clearly the outlier adversely affects the classical Wald test, but not the proposed robust Wald-type tests. The mean of the ordered differences between the inverse test rates and inverse control rates does appear to be actually greater than zero.

Darwin’s plant fertilization data: In this case we want to test whether the mean of the paired differences (cross-fertilized minus self-fertilized) is zero against the greater than alternative. The classical Wald test performed by Fisher (1966) gives a statistic of 2.15, with a one sided p -value of 0.025 computed under the $t(14)$ distribution; the corresponding outlier deleted statistic has a one sided p -value of 0.00007. On the other hand, the one sided p -values for the proposed signed Wald-type test statistics for $\beta = 0.15$ and $\beta = 0.30$ are 0.0145, 0.0027 for the full data, and 0.0001, 0.0001 for the outlier deleted data. Clearly the two large outliers are influencing the decision of the classical Wald test, or the proposed Wald-type tests for very small values of β , to the extent that the decision is reversed from what would have been obtained with the outlier cleaned data at 1% level of significance; however larger values of β lead to a more consistent behavior of the tests. This data set requires stronger downweighting compared to the Telephone-Fault data, as the outliers here are less extreme, and therefore more difficult to identify. Under a suitable robust test, it does appear that the mean growth of cross-fertilized plants is higher than that of the self-fertilized plants.

6 On the Choice of Tuning Parameter β

Finally, we have to give some guidance to the practitioner as to what value of the tuning parameter β is to be used in a particular case involving a sample of real data. Ideally we would use the parameter $\beta = 0$ if the data were always pure, which in practice is hardly ever the case. Several authors (see Basu et al., 1998 and Basu et al., 2011) have demonstrated that a small positive value of β often provides highly robust solutions with only a slight loss in efficiency. To analyze a particular set of real data we need a data driven choice of the tuning parameter which provides the necessary stability with as little loss in efficiency as possible.

Our experience in repeated simulations and real data examples is that one generally does not require a choice of β larger than 0.6 in most practical situations; quite often, in fact, $\beta = 0.5$ or lower is sufficient. Thus $\beta = 0.6$ can be viewed as a suitable conservative choice of the tuning parameter. However, at times this may lead to a greater loss in efficiency than is necessary, so a more refined data driven choice of the tuning parameter would be helpful. In this context we prefer to begin with the MDPDE corresponding to $\beta = 0.5$ as the pilot estimator, and determine the data driven choice of the tuning parameter β following the approach of Warwick and Jones (2005). This approach minimizes an empirical measure of the mean square error of the parameter estimate to determine the “optimal” tuning parameter. The pilot estimator is necessary to get the empirical measure of the bias. We note that Ghosh and Basu (2013) also used the estimator corresponding to $\beta = 0.5$ as the pilot parameter while considering robust estimation with independent but non-homogeneous data. One could also use $\beta = 0.6$, our conservative estimator of the tuning parameter, as the choice for constructing the pilot estimator; however our data analysis experience shows that this makes a minor difference if at all, and for reasons of efficiency we stick to the Ghosh and Basu (2013) proposal.

Naturally, the criterion to be considered in case of the testing problem and the estimation problem are not exactly the same. In case of the hypothesis testing problem, one should minimize a criterion which should involve a suitable linear combination of the amount of inflation in the level of the test and the difference of its contiguous power from one. We feel that it is much more difficult to construct an appropriate empirical measure of such a quantity particularly under composite alternatives. On the other hand we have demonstrated, throughout the numerical data examples, that the robustness of the proposed tests correspond almost exactly to the robustness of the MDPDEs. Thus we feel that the optimal choice of β determined as described in the previous paragraph would be a reasonable choice in the testing situation as well, and that remains our current recommendation.

By way of illustration, this criterion leads to estimated optimal choices of β to be 0.1919 and 0.5657 for the Telephone-Fault data and Darwin's plant fertilization data respectively. These values are consistent with our earlier observations that for the Telephone-Fault data the first value is a massive outlier, whereas for Darwin's plant fertilization data the outliers are much closer to the overall cluster of observations, and larger values of β will be required to eliminate their effect.

7 Concluding Remarks

In this paper we have constructed generalized Wald-type tests based on minimum density power divergence estimators proposed by Basu et al. (1998). These tests do not require any intermediate smoothing technique as in the case of the Hellinger deviance test, so our proposed method appears to stand out among robust tests of hypotheses based on minimum distance methods. The proposed Wald-type tests are easy to construct, and avoid the problem of determining the rejection region based on linear combinations of chi-square distributions as in the case of Basu et al. (2013, 2014). These class of tests have a huge scope of application. We have demonstrated improved results in different models including normal, exponential and Weibull distributions. Some real data examples have also been analyzed, and we note that our methods can be useful in detecting either kind of incorrect decision caused by a small number of extreme outliers. This is exemplified by the analysis of the Leukemia data, where the classical test leads to a false positive, and the analysis of the Telephone-Fault data where the classical test fails to detect a true positive. We trust that the proposed tests will be very useful practical tools for statistical data analysis.

Acknowledgments: This work was partially supported by Grant MTM-2012-33740 and ECO-2011-25706. The authors gratefully acknowledge the suggestions of the editor, the associate editor and two anonymous referees which led to an improved version of the paper.

References

- A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559, 1998.
- A. Basu, H. Shioya, and C. Park. *Statistical Inference: The Minimum Distance Approach*. CRC Press, Boca Raton, FL, 2011.
- A. Basu, A. Mandal, N. Martin, and L. Pardo. Testing statistical hypotheses based on the density power divergence. *Ann. Inst. Statist. Math.*, 65(2):319–348, 2013.
- A. Basu, A. Mandal, N. Martin, and L. Pardo. Density power divergence tests for composite null hypotheses. *arXiv preprint arXiv:1403.0330*, 2014.
- D. Chant. On asymptotic tests of composite hypotheses in nonstandard conditions. *Biometrika*, 61(2):291–298, 1974.
- H. Chernoff. On the distribution of the likelihood ratio. *Ann. Math. Statist*, 25(3):573–578, 1954.
- C. Darwin. *The Effects of Cross and Self Fertilization in the Vegetable Kingdom*. John Murray, London, 1878.
- D. De Angelis and G. A. Young. Smoothing the bootstrap. *Internat. Statist. Rev.*, 60(1):45–56, 1992.
- W. J. Dixon and J. W. Tukey. Approximate behavior of the distribution of winsorized t (trimming/winsorization 2). *Technometrics*, 10(1):83–98, 1968.
- R. Fisher. *The Design of Experiments*. Hafner Press, New York, 1966.
- D. A. S. Fraser. Most powerful rank-type tests. *Ann. Math. Statist*, 28:1040–1043, 1957.

- A. Ghosh and A. Basu. Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electron. J. Stat.*, 7:2420–2456, 2013.
- A. J. Gross and V. Clark. *Survival distributions: reliability applications in the biomedical sciences*, volume 11. Wiley, New York, 1975.
- M. C. Jones, N. L. Hjort, I. R. Harris, and A. Basu. A comparison of related density-based minimum divergence estimators. *Biometrika*, 88(3):865–873, 2001.
- A. J. Lemonte. Nonnull asymptotic distributions of the LR, Wald, score and gradient statistics in generalized linear models with dispersion covariates. *Statistics*, 47(6):1249–1265, 2013.
- L. Pardo. *Statistical inference based on divergence measures*. Chapman & Hall/CRC, Boca Raton, FL, 2006.
- F. Proschan. Theoretical explanation of observed decreasing failure rate. *Technometrics*, 5(3):375–383, 1963.
- S. G. Self and K.-Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.
- R. J. Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, New York, 1980. ISBN 0-471-02403-1. Wiley Series in Probability and Mathematical Statistics.
- D. G. Simpson. Hellinger deviance tests: efficiency, breakdown points, and examples. *J. Amer. Statist. Assoc.*, 84(405):107–113, 1989.
- E. Susko. Likelihood ratio tests with boundary constraints using data-dependent degrees of freedom. *Biometrika*, 100(4):1019–1023, 2013.
- M. Teimouri, S. M. Hoseini, and S. Nadarajah. Comparison of estimation methods for the weibull distribution. *Statistics*, 47(1):93–109, 2013.
- M. L. Tiku, W. Y. Tan, and N. Balakrishnan. *Robust inference*. Marcel Dekker, New York, 1986. ISBN 0-8247-7532-5.
- J. Warwick and M. Jones. Choosing a robustness tuning parameter. *J. Stat. Comput. Simulation*, 75(7):581–588, 2005.
- W. J. Welch. Rerandomizing the median in matched-pairs designs. *Biometrika*, 74(3):609–614, 1987.

Appendix

Proof of Theorem 2: A first order Taylor expansion of $l(\boldsymbol{\theta})$ at $\widehat{\boldsymbol{\theta}}_{\beta}$ around $\boldsymbol{\theta}^*$ gives

$$l(\widehat{\boldsymbol{\theta}}_{\beta}) - l(\boldsymbol{\theta}^*) = \left(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}^*) + o_p(\|\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}^*\|).$$

Now the result follows easily, since the asymptotic distribution of $n^{1/2} (l(\widehat{\boldsymbol{\theta}}_{\beta}) - l(\boldsymbol{\theta}^*))$ coincides with the asymptotic distribution of $n^{1/2} \left(\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} (\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}^*)$.

Proof of Theorem 3: We have

$$n^{1/2} (\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_0) = n^{1/2} (\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_n) + n^{1/2} (\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = n^{1/2} (\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_n) + \mathbf{d}.$$

Under $H_{1,n}$ it follows that

$$n^{1/2} (\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p(\mathbf{0}, \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0)),$$

and

$$n^{1/2} \left(\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p(\mathbf{d}, \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0)).$$

On the other hand

$$W_n = \mathbf{X}^T \mathbf{X},$$

where

$$\mathbf{X} = \left(\mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \right)^{-1/2} n^{1/2} \left(\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_0 \right),$$

and under $H_{1,n}$

$$\mathbf{X} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p \left(\left(\mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \right)^{-1/2} \mathbf{d}, \mathbf{I}_{p \times p} \right).$$

Here $\mathbf{I}_{p \times p}$ is the identity matrix of order p . Therefore

$$W_n = \mathbf{X}^T \mathbf{X} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_p^2(\delta)$$

with $\delta = \mathbf{d}^T \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{d}$; here $\chi_p^2(\delta)$ denotes a non-central chi-square distribution with p degrees of freedom and non-centrality parameter δ .

Proof of Theorem 4: Let $\boldsymbol{\theta}_0 \in \Theta_0$ be the true value of $\boldsymbol{\theta}$. Using a Taylor series expansion we get

$$\begin{aligned} \mathbf{m} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) &= \mathbf{m}(\boldsymbol{\theta}_0) + \mathbf{M}(\boldsymbol{\theta}_0)^T \left(\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_0 \right) + o_p \left(\left\| \widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_0 \right\| \right) \\ &= \mathbf{M}(\boldsymbol{\theta}_0)^T \left(\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_0 \right) + o_p \left(\left\| \widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_0 \right\| \right), \end{aligned} \quad (34)$$

because from equation (17) we have $\mathbf{m}(\boldsymbol{\theta}_0) = \mathbf{0}_r$. Now, under H_0 ,

$$n^{1/2} \left(\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}_0 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p(\mathbf{0}_p, \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0)).$$

Therefore, from equation (34) we get, under H_0 ,

$$n^{1/2} \mathbf{m} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_r(\mathbf{0}_r, \mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0)).$$

As $\text{rank}(\mathbf{M}(\boldsymbol{\theta})) = r$, we get

$$n \mathbf{m}^T \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \left[\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0) \right]^{-1} \mathbf{m} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_r^2.$$

Now $\mathbf{M}^T \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \mathbf{J}_{\beta}^{-1} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \mathbf{K}_{\beta} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \mathbf{J}_{\beta}^{-1} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \mathbf{M} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right)$ is a consistent estimator of $\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_{\beta}(\boldsymbol{\theta}_0) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0)$. Hence, under H_0 ,

$$n \mathbf{m}^T \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \left[\mathbf{M}^T \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \mathbf{J}_{\beta}^{-1} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \mathbf{K}_{\beta} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \mathbf{J}_{\beta}^{-1} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \mathbf{M} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \right]^{-1} \mathbf{m} \left(\widehat{\boldsymbol{\theta}}_{\beta} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_r^2.$$

Proof of Theorem 5: We note that $l^* \left(\widehat{\boldsymbol{\theta}}_{\beta}, \widehat{\boldsymbol{\theta}}_{\beta} \right)$ and $l^* \left(\widehat{\boldsymbol{\theta}}_{\beta}, \boldsymbol{\theta}^* \right)$ have same asymptotic distribution, because $\widehat{\boldsymbol{\theta}}_{\beta} \xrightarrow[n \rightarrow \infty]{p} \boldsymbol{\theta}^*$. Now a first order Taylor expansion of $l^* \left(\widehat{\boldsymbol{\theta}}_{\beta}, \boldsymbol{\theta}^* \right)$ at $\widehat{\boldsymbol{\theta}}_{\beta}$ around $\boldsymbol{\theta}^*$ gives

$$l^* \left(\widehat{\boldsymbol{\theta}}_{\beta}, \boldsymbol{\theta}^* \right) - l^* \left(\boldsymbol{\theta}^*, \boldsymbol{\theta}^* \right) = \left(\frac{\partial l^*(\boldsymbol{\theta}, \boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^T \left(\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}^* \right) + o_p \left(\left\| \widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}^* \right\| \right).$$

So the theorem is proved from the following result

$$n^{1/2} \left(\widehat{\boldsymbol{\theta}}_{\beta} - \boldsymbol{\theta}^* \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p \left(\mathbf{0}_p, \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}^*) \mathbf{K}_{\beta}(\boldsymbol{\theta}^*) \mathbf{J}_{\beta}^{-1}(\boldsymbol{\theta}^*) \right).$$

Proof of Theorem 6: A Taylor series expansion of $\mathbf{m}(\widehat{\boldsymbol{\theta}}_\beta)$ around $\boldsymbol{\theta}_n$ yields

$$\mathbf{m}(\widehat{\boldsymbol{\theta}}_\beta) = \mathbf{m}(\boldsymbol{\theta}_n) + \mathbf{M}^T(\boldsymbol{\theta}_n) (\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n) + o(\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n\|).$$

From (25) we have

$$\mathbf{m}(\widehat{\boldsymbol{\theta}}_\beta) = n^{-1/2} \mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{d} + \mathbf{M}^T(\boldsymbol{\theta}_n) (\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n) + o(\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n\|) + o(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|). \quad (35)$$

Under $H_{1,n}$ we get $n^{1/2} (o(\|\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n\|) + o(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0\|)) = o_p(1)$ and

$$n^{1/2} (\widehat{\boldsymbol{\theta}}_\beta - \boldsymbol{\theta}_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_p(\mathbf{0}_p, \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0)).$$

So from (35) we have

$$n^{1/2} \mathbf{m}(\widehat{\boldsymbol{\theta}}_\beta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_r(\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{d}, \mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0)).$$

From (26) we get, under $H_{1,n}^*$,

$$n^{1/2} \mathbf{m}(\widehat{\boldsymbol{\theta}}_\beta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_r(\boldsymbol{\delta}, \mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0)).$$

We apply the following result concerning quadratic forms: If $\mathbf{Z} \in N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ is a symmetric projection of rank k and $\boldsymbol{\Sigma} \boldsymbol{\mu} = \boldsymbol{\mu}$, then $\mathbf{Z}^T \mathbf{Z}$ is a chi-square distribution with k degrees of freedom and non-centrality parameter $\boldsymbol{\mu}^T \boldsymbol{\mu}$. Here the quadratic form is

$$W_n = \mathbf{Z}^T \mathbf{Z},$$

where

$$\mathbf{Z} = n^{1/2} \mathbf{m}(\widehat{\boldsymbol{\theta}}_\beta) \left[\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0) \right]^{-1/2}.$$

We know

$$\mathbf{Z} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N_r \left(\left[\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0) \right]^{-1/2} \mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{d}, \mathbf{I}_r \right),$$

where \mathbf{I}_r is the identity matrix of order r . Hence the application of the result is immediate. The non-centrality parameter is

$$\mathbf{d}^T \mathbf{M}(\boldsymbol{\theta}_0) \left[\mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{K}_\beta(\boldsymbol{\theta}_0) \mathbf{J}_\beta^{-1}(\boldsymbol{\theta}_0) \mathbf{M}(\boldsymbol{\theta}_0) \right]^{-1} \mathbf{M}^T(\boldsymbol{\theta}_0) \mathbf{d}.$$