Influence Analysis of Robust Wald-type Tests

Abhik Ghosh¹, Abhijit Mandal¹, Nirian Martín², Leandro Pard³

¹Indian Statistical Institute, Kolkata, India  ²Carlos III University of Madrid, Getafe (Madrid), Spain
³Complutense University of Madrid, Madrid, Spain

June 26, 2014

Abstract

We consider a robust version of the classical Wald test statistics for testing simple and composite null hypotheses for general parametric models. These test statistics are based on the minimum density power divergence estimators instead of the maximum likelihood estimators. An extensive study of their robustness properties is given through the influence functions as well as the chi-square inflation factors. It is theoretically established that the level and power of these robust tests are stable against outliers, whereas the classical Wald test breaks down. Some numerical examples confirm the validity of the theoretical results.

AMS 2001 Subject Classification: Primary 62F35, Secondary 62F03.
Keywords and phrases: Wald-type test statistics, Minimum density power divergence estimators, Robustness, Influence Functions, Chi-square Inflation Factor.

1 Introduction

It is well-known that small deviations from the underlying assumptions on the model can have drastic effects on the performance of classical testing procedures. The purpose in robust testing of hypothesis is twofold. First, the level of a test should be stable under small, arbitrary departures from the null hypothesis (robustness of validity). Secondly, the test should still have good power under small arbitrary departures from specified alternatives (robustness of efficiency). Robust tests are more stable and reliable than the classical tests in a neighborhood of the model. The influence function can be used to investigate the local stability and the global reliability of a test and at the same time they provide the basis for constructing new robust tests. Ronchetti (1979, 1982a,b) and Rousseeuw and Ronchetti (1979, 1981) extended Hampel’s influence function concept for testing a null hypothesis about a scalar parameter; see Hampel et al. (1986, Chapter 3). An essential result of this approach is the approximation of the asymptotic level and the asymptotic power under a contaminated distribution in a neighborhood of the null hypothesis. The idea is to study the behavior of the level of the test as a function of an additional observation at any point \( x \). The same technique can be applied to the power
function. A very nice review about the influence function in the study of the robustness of test statistics is given in Markatou and Ronchetti (1997).

In Basu et al. (2014) Wald-type test statistics based on minimum density power divergence estimators (MDPDE) were studied in order to test simple null hypotheses as well as composite null hypotheses. They have empirically demonstrated some of the strong robustness properties of the Wald-type test statistic but no theoretical result on robustness was discussed in that paper. In this current paper we shall develop some theoretical results on robustness for the Wald-type tests.

The rest of the paper is organized as follows. In Section 2 we have presented some notations and results from Basu et al. (2014) which are necessary to develop further theoretical results for this paper. Section 3 presents the influence functions of the Wald-type test statistics. The power and level influence functions for testing simple and composite null hypotheses are derived in Sections 4 and 5, respectively. The chi-square inflation factors for Wald-type test statistics are calculated in Section 6. In Section 7 we have presented some examples to justify the theoretical results developed in this paper. A discussion on choosing the tuning parameter for the density power divergence measure is given in Section 8, and finally, the concluding remarks are provided in Section 9.

2 Preliminaries

Let \( G \) 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 \( G \), the density power divergence with a nonnegative tuning parameter \( \beta \), is defined as

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

The divergence corresponding to \( \beta = 0 \) may be derived from the general case by taking the continuous limit as \( \beta \to 0 \), and in this case \( d_0(g, f) \) turns out to be the Kullback-Leibler divergence. For \( \beta = 1 \) the square of the \( L_2 \) distance between \( g \) and \( f \) is obtained. The quantities defined in equation (1) are genuine divergences because \( d_\beta(g, f) \geq 0 \) for all \( g, f \in G \) with \( \beta \geq 0 \), and \( d_\beta(g, f) \) is equal to zero if and only if the densities \( g \) and \( f \) are identically equal. More details about the inference based on divergence measures can be found in Basu et al. (2011) and Pardo (2006).

We consider a parametric model of densities \( \{f_\theta : \theta \in \Theta \subset \mathbb{R}^p\} \), and we are interested in the estimation of \( \theta \). Let \( G \) represent the distribution function corresponding to the density \( g \) that generates the data. The minimum density power divergence functional at \( G \), denoted by \( T_\beta(G) \), is defined as

\[
d_\beta(g, T_\beta(G)) = \min_{\theta \in \Theta} d_\beta(g, f_\theta).
\] (2)

Therefore the MDPDE of \( \theta \) is given by

\[
\hat{\theta}_\beta = T_\beta(G_n),
\] (3)
where \( G_n \) is the empirical distribution function associated with a random sample \( X_1, \ldots, X_n \) from the population with density \( g \) (having distribution function \( G \)). As the last term of equation (1) does not depend on \( \theta \), \( \hat{\theta}_\beta \) is given by

\[
\hat{\theta}_\beta = \arg\min_{\theta \in \Theta} \left\{ \int f_\theta^{1+\beta}(x)dx - \left(1 + \frac{1}{\beta}\right) \frac{1}{n} \sum_{i=1}^n f_\theta^\beta(X_i) \right\},
\]

if \( \beta > 0 \) and

\[
\hat{\theta}_\beta = \arg\min_{\theta \in \Theta} \left\{ -\frac{1}{n} \sum_{i=1}^n \log f_\theta(X_i) \right\},
\]

when \( \beta = 0 \). Notice that \( \hat{\theta}_\beta \) for \( \beta = 0 \) coincides with the maximum likelihood estimator (MLE). Denoting

\[
V_\theta(x) = \int f_\theta^{1+\beta}(y)dy - \left(1 + \frac{1}{\beta}\right) f_\theta^\beta(x),
\]

expression (4) can be written as \( \hat{\theta}_\beta = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n V_\theta(X_i) \). It shows that the MDPDE is an M-estimator.

Let \( u_\theta(x) = \frac{\partial}{\partial \theta} \log f_\theta(x) \) be the score function of the model. Under differentiability of the model, the minimization of the objective function in equation (4) leads to an estimating equation of the form

\[
\frac{1}{n} \sum_{i=1}^n u_\theta(X_i)f_\theta^\beta(X_i) - \int u_\theta(x)f_\theta^{1+\beta}(x)dx = 0_p,
\]

which is an unbiased estimating equation under the model. Here \( 0_p \) denotes the null vector of dimension \( p \).

Since the corresponding estimating equation weights the score \( u_\theta(X_i) \) with the power of the density \( f_\theta^\beta(X_i) \), the outlier resistant behavior of the estimator is intuitively apparent.

The functional \( T_\beta(G) \) is Fisher consistent; it takes the value \( \theta_0 \), the true value of the parameter, when the true density is a member of the model, i.e. \( g = f_{\theta_0} \). Let us assume \( g = f_{\theta_0} \), and define the quantities

\[
J_\beta(\theta) = \int u_\theta(x)u_\theta^T(x)f_\theta^{1+\beta}(x)dx, \quad K_\beta(\theta) = \int u_\theta(x)u_\theta^T(x)f_\theta^{1+2\beta}(x)dx - \xi_\beta(\theta)\xi_\beta^T(\theta),
\]

where

\[
\xi_\beta(\theta) = \int u_\theta(x)f_\theta^{1+\beta}(x)dx.
\]

Then, following Basu et al. (1998) and Basu et al. (2011) , it can be shown that

\[
n^{1/2}(\hat{\theta}_\beta - \theta_0) \xrightarrow{D} \mathcal{N}(0_p, \Sigma_\beta(\theta_0)),
\]

where

\[
\Sigma_\beta(\theta_0) = J_\beta^{-1}(\theta_0)K_\beta(\theta_0)J_\beta^{-1}(\theta_0).
\]

### 2.1 Wald-type Test Statistics for the Simple Null Hypothesis

In Basu et al. (2014) the family of Wald-type test statistics

\[
W_n^{0}(\hat{\theta}_\beta) = n(\hat{\theta}_\beta - \theta_0)^T \Sigma_\beta^{-1}(\theta_0)(\hat{\theta}_\beta - \theta_0)
\]

was considered for testing the simple null hypothesis

\[
H_0 : \theta = \theta_0 \text{ against } H_1 : \theta \neq \theta_0,
\]
where \( \theta_0 \in \Theta \subset \mathbb{R}^p \). The asymptotic distribution of \( W_0^m(\hat{\theta}_\beta) \), defined in (11), is a chi-square with \( p \) degrees of freedom. In the particular case when \( \beta = 0 \), i.e., the MDPDE coincides with the MLE, the variance-covariance matrix, (10), coincides with the inverse of the Fisher information matrix of the model and then we get the classical Wald test statistic for testing (12). The power function \( \beta_{W_0} \) of the Wald-type test statistics at \( \theta^* \in \Theta - \Theta_0 \), is given by

\[
\beta_{W_0}(\theta^*) \equiv 1 - \Phi \left( \frac{\sqrt{n}}{\sigma_{W_0}(\theta^*)} \left( \chi^2_{p,\alpha} - \ell(\theta^*) \right) \right),
\]

where

\[
\ell(\theta^*) = (\theta^* - \theta_0)^T \Sigma^{-1}_\beta(\theta_0) (\theta^* - \theta_0),
\]

\[
\sigma^2_{W_0}(\theta^*) = 4 (\theta^* - \theta_0)^T \Sigma^{-1}_\beta(\theta^*) (\theta^* - \theta_0).
\]

Here \( \alpha \) is the level of the test, \( \chi^2_{p,\alpha} \) is the 100(1 − \( \alpha \))-th percentile of a chi-square distribution with \( p \) degrees of freedom and \( \Phi(\cdot) \) is the standard normal distribution function. It is clear that

\[
\lim_{n \to \infty} \beta_{W_0}(\theta^*) = 1,
\]

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

In order to produce a nontrivial asymptotic power, we can consider contiguous alternative hypotheses.

Consider the contiguous alternative hypotheses described by

\[
H_{1,n} : \theta_n = \theta_0 + n^{-1/2} d,
\]

where \( d \) is a fixed vector in \( \mathbb{R}^p \) such that \( \theta_n \in \Theta \subset \mathbb{R}^p \). It can be shown that the asymptotic distribution of the Wald-type test statistic \( W_0^m(\hat{\theta}_\beta) \) under the alternative \( H_{1,n} \) is a non-central chi-square with \( p \) degrees of freedom and non-centrality parameter

\[
\delta = d^T \Sigma(\theta_0) d.
\]

Based on this result, under (14) we have the following approximation to the power function

\[
\beta_{W_0}(\theta_n) = 1 - F_{\chi^2_{p}(\delta)}(\chi^2_{p,\alpha}),
\]

where \( F_{\chi^2(\delta)}(\cdot) \) is the distribution function of a non-central chi-square random variable with \( p \) degrees of freedom and non-centrality parameter \( \delta \).

### 2.2 Wald-type Test Statistics for the Composite Null Hypothesis

We shall now consider the problem of testing the composite null hypothesis given by

\[
H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \notin \Theta_0,
\]

where \( \Theta_0 \) is a subset of the parameter space \( \Theta \in \mathbb{R}^p \). The restricted parameter space \( \Theta_0 \) is often defined by a set of \( r \) restrictions of the form

\[
m(\theta) = 0_r,
\]

where \( \Theta_0 \) is a subset of the parameter space \( \Theta \in \mathbb{R}^p \). The restricted parameter space \( \Theta_0 \) is often defined by a set of \( r \) restrictions of the form
where \( m : \mathbb{R}^p \to \mathbb{R}^r \) with \( r \leq p \) (see Serfling, 1980). So \( \Theta_0 = \{ \theta \in \Theta : m(\theta) = 0_r \} \). Assume that the \( p \times r \) matrix
\[
M(\theta) = \frac{\partial m^T(\theta)}{\partial \theta}
\] (19)
exists and is continuous in \( \theta \), and rank \( (M(\theta)) = r \).

Basu et al. (2014) have considered the following family of Wald-type test statistics
\[
W_n(\hat{\theta}_\beta) = nm^T(\hat{\theta}_\beta) \left( M^T(\theta_0) \Sigma_\beta(\theta_0) M(\theta_0) \right)^{-1} m(\hat{\theta}_\beta),
\] (20)
where the matrix \( \Sigma_\beta(\cdot) \) is defined in (10). The asymptotic distribution of the Wald-type test statistic \( W_n(\hat{\theta}_\beta) \) under the composite null hypothesis (17) is a chi-square with \( m \) degrees of freedom. In a practical framework, \( M^T(\theta_0) \Sigma_\beta(\theta_0) M(\theta_0) \) is estimated replacing \( \theta_0 \) by \( \hat{\theta}_\beta \).

In the special case when \( \beta = 0 \), \( \hat{\theta}_\beta \) coincides with the maximum likelihood estimator of \( \theta_0 \), and \( \Sigma_\beta(\cdot) \) becomes the inverse of the Fisher information matrix. Thus, the statistic in (20) reduces to the classical Wald test statistic.

The power function \( \beta_{W_n}(\theta^*) \) of the Wald-type test statistic at \( \theta^* \in \Theta - \Theta_0 \), is given by
\[
\beta_{W_n}(\theta^*) \equiv 1 - \Phi \left( \frac{\sqrt{n}}{\sigma_{W_n}(\theta^*)} \left( \frac{\ell^*}{n} - \ell^*(\theta_0, \theta^*) \right) \right),
\] (21)
where
\[
\ell^*(\theta_1, \theta_2) = nm^T(\theta_1) \left( M^T(\theta_2) \Sigma_\beta(\theta_2) M(\theta_2) \right)^{-1} m(\theta_1),
\]
and
\[
\sigma^2_{W_n}(\theta^*) = \left( \frac{\partial \ell^*(\theta, \theta^*)}{\partial \theta} \right)^T \Sigma_\beta(\theta^*) \left( \frac{\partial \ell^*(\theta, \theta^*)}{\partial \theta} \right)_{\theta=\theta^*}.
\] (22)

Basu et al. (2014) proposed an approximation of the power of \( W_n(\hat{\theta}_\beta) \) at an alternative hypothesis close to the null hypothesis. Let \( \theta_n \in \Theta - \Theta_0 \) be a given alternative, and let \( \theta_0 \) be the element in \( \Theta_0 \) closest to \( \theta_n \) in terms of the Euclidean distance. One possibility to introduce contiguous alternative hypotheses, in this context, is to consider a fixed vector \( d \in \mathbb{R}^p \) and permit \( \theta_n \) to move towards \( \theta_0 \) as \( n \) increases through the relation \( H_{1,n} \) given in (14). A second approach is to relax the condition \( m(\theta) = 0_r \) that defines \( \Theta_0 \). Let \( \delta \in \mathbb{R}^r \) and consider the following sequence of parameters \( \{ \theta_n \} \) moving towards \( \theta_0 \) according to the set up
\[
H^*_{1,n} : m(\theta_n) = n^{-1/2} \delta.
\] (23)
Note that a Taylor series expansion of \( m(\theta_n) \) around \( \theta_0 \) yields
\[
m(\theta_n) = m(\theta_0) + M^T(\theta_0) (\theta_n - \theta_0) + o (\| \theta_n - \theta_0 \|).
\] (24)
By substituting \( \theta_n = \theta_0 + n^{-1/2} d \) in (24) and taking into account that \( m(\theta_0) = 0_r \), we get
\[
m(\theta_n) = n^{-1/2} M^T(\theta_0) d + o (\| \theta_n - \theta_0 \|).
\] (25)
So, the equivalence relationship between the hypotheses \( H_{1,n} \) and \( H^*_{1,n} \) is
\[
\delta = M^T(\theta_0) d \text{ as } n \to \infty.
\] (26)
The asymptotic distribution of \( W_n(\hat{\theta}_\beta) \) is given by
\[
W_n(\hat{\theta}_\beta) \xrightarrow{n \to \infty} \chi_r^2 \left( d^T M(\theta_0) \left( M^T(\theta_0) \Sigma(\theta_0) M(\theta_0) \right)^{-1} M^T(\theta_0) d \right)
\] (27)
under \( H_{1,n} \) given in (14) and by
\[
W_n(\hat{\theta}_\beta) \xrightarrow{n \to \infty} \chi_r^2 \left( \delta^T \left( M^T(\theta_0) \Sigma(\theta_0) M(\theta_0) \right)^{-1} \delta \right)
\] (28)
under \( H_{1,n}^* \) given in (23). These asymptotic distributions may be used to calculate the power functions of the Wald-type test statistics under the contiguous alternatives.

3 Influence functions of the Wald-type test statistics

In Basu et al. (1998) it was established that the influence function of the density power divergence functional is
\[
\mathcal{I}(x, T_\beta, F_{\theta_0}) = \lim_{\varepsilon \to 0} \frac{T_\beta(F_{\varepsilon}) - T_\beta(F_{\theta_0})}{\varepsilon} = J^{-1}_\beta(\theta_0) \left( u_\theta(x) f^\beta_{\theta_0}(x) - \xi(\theta_0) \right),
\] (29)
where \( F_{\varepsilon} = (1-\varepsilon)F_{\theta_0} + \varepsilon \Delta_x \) is the \( \varepsilon \)-contaminated distribution of \( F_{\theta_0} \) with respect to \( \Delta_x \), the point mass distribution at \( x \). If we assume that \( J_\beta(\theta_0) \) and \( \xi(\theta_0) \) are finite, the influence function is a bounded function of \( x \) whenever \( u_\theta(x) f^\beta_{\theta_0}(x) \) is bounded. This is true, for example in the normal location-scale problem for \( \beta > 0 \), unlike other density based minimum divergence procedures such as those based on the Hellinger distance. In the case of the normal model with known variance \( \sigma^2 \) and unknown mean \( \theta_0 \), we have
\[
\mathcal{I}(x, T_\beta, F_{\theta_0}) = \frac{x - \theta_0}{\sigma^2 + (2\pi)^{\beta}} \exp \left\{ \frac{1}{2} \frac{(x - \theta_0)^2}{\sigma^2} \right\}.
\]
For any \( \beta > 0 \), the above mentioned influence function is bounded, but for \( \beta = 0 \) is not bounded.

Let us consider the test statistic \( W_n^0(\hat{\theta}_\beta) \) for testing the simple null hypothesis given in (12). The functional associated with the test statistic \( W_n^0(\hat{\theta}_\beta) \), evaluated at \( G \), is given by (ignoring the multiplier \( n \))
\[
W_n^0(G) = (T_\beta(G) - \theta_0)^T \Sigma^{-1}(\theta_0)(T_\beta(G) - \theta_0).
\] (30)
Let \( G_{\varepsilon} = (1-\varepsilon)G + \varepsilon \Delta_x \) be the \( \varepsilon \)-contaminated distribution of \( G \) with respect to the point mass distribution \( \Delta_x \) at \( x \). The influence function of \( W_n^0(\cdot) \) is defined as
\[
\mathcal{I}(x, W_n^0, G) = \left. \frac{\partial W_n^0(G_{\varepsilon})}{\partial \varepsilon} \right|_{\varepsilon=0},
\]
where
\[
\left. \frac{\partial W_n^0(G_{\varepsilon})}{\partial \varepsilon} \right|_{\varepsilon=0} = 2(T_\beta(G) - \theta_0)^T \Sigma^{-1}(\theta_0) \mathcal{I}(x, T_\beta, G).
\]
Under the simple null hypothesis given in (12), \( G = F_{\theta_0} \) and \( T_\beta(G) = \theta_0 \). So \( \mathcal{I}(x, W_n^0, F_{\theta_0}) = 0 \), which shows that the influence function analysis based on the first derivative of \( W_n^0(G_{\varepsilon}) \) is not adequate to quantify the robustness of these estimators.
The second order influence function of $W^0_\beta(\cdot)$ is given by

$$\mathcal{I}_F^2(x, W^0_\beta, G) = \left. \frac{\partial^2 W^0_\beta(G_\varepsilon)}{\partial \varepsilon^2} \right|_{\varepsilon=0},$$

and

$$\left. \frac{\partial^2 W^0_\beta(G_\varepsilon)}{\partial \varepsilon^2} \right|_{\varepsilon=0} = 2\mathcal{I}_F(x, T_\beta, G)\Sigma^{-1}_\beta(\theta_0)\mathcal{I}_F(x, T_\beta, G) + 2(T_\beta(G) - \theta_0)^T\Sigma^{-1}_\beta(\theta_0)\mathcal{I}_F^2(x, T_\beta, G).$$

As $T_\beta(F_{\theta_0}) = \theta_0$, we obtain

$$\mathcal{I}_F^2(x, W^0_\beta, F_{\theta_0}) = 2\mathcal{I}_F(x, T_\beta, F_{\theta_0})\Sigma^{-1}_\beta(\theta_0)\mathcal{I}_F(x, T_\beta, F_{\theta_0}).$$

We shall now calculate the influence function for the composite hypothesis. The functional associated with the test statistic $W_n(\hat{\theta}_\beta)$, given in (20), evaluated at $G$, is given by (ignoring the multiplier $n$)

$$W_\beta(G) = m^T(T_\beta(G)) \left( M^T(\theta_0)\Sigma_\beta(\theta_0)M(\theta_0) \right)^{-1} m(T_\beta(G)). \tag{31}$$

The influence function of $W_\beta(\cdot)$ is defined as

$$\mathcal{I}_F(x, W_\beta, G) = \left. \frac{\partial W_\beta(G_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0},$$

where

$$\left. \frac{\partial W_\beta(G_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} = 2m^T(T_\beta(G)) \left[ M^T(\theta_0)\Sigma_\beta(\theta_0)M(\theta_0) \right]^{-1} M^T(T_\beta(G))\mathcal{I}_F(x, T_\beta, G).$$

Let $\theta_0 \in \Theta_0$ be the true value of the parameter under the composite hypothesis given in (17). So $G = F_{\theta_0}$ and $m(T_\beta(G)) = 0_r$, and finally it turns out that $\mathcal{I}_F(x, W_\beta, G) = 0$, which indicates that the derivation of second order influence function is necessary.

The second order influence function of (31) is given by

$$\mathcal{I}_F^2(x, W_\beta, G) = \left. \frac{\partial^2 W_\beta(G_\varepsilon)}{\partial \varepsilon^2} \right|_{\varepsilon=0},$$

and

$$\left. \frac{\partial^2 W_\beta(G_\varepsilon)}{\partial \varepsilon^2} \right|_{\varepsilon=0} = 2\mathcal{I}_F(x, T_\beta, G)M(T_\beta(G)) \left[ M^T(\theta_0)\Sigma_\beta(\theta_0)M(\theta_0) \right]^{-1}$$

$$\times M^T(T_\beta(G))\mathcal{I}_F(x, T_\beta, G) + 2m^T(T_\beta(G)) \left[ M^T(\theta_0)\Sigma_\beta(\theta_0)M(\theta_0) \right]^{-1}$$

$$\times M^T(T_\beta(G))\mathcal{I}_F^2(x, T_\beta, G).$$

As $T_\beta(F_{\theta_0}) = \theta_0$, we obtain

$$\mathcal{I}_F^2(x, W_\beta, F_{\theta_0}) = 2\mathcal{I}_F(x, T_\beta, F_{\theta_0})M(\theta_0) \left[ M^T(\theta_0)\Sigma_\beta(\theta_0)M(\theta_0) \right]^{-1}$$

$$\times M^T(\theta_0)\mathcal{I}_F(x, T_\beta, F_{\theta_0}).$$

The following theorem summarizes the previous results.
Let us define the second order influence functions of the test statistic $W_0^\beta(\hat{\theta}_\beta)$, given in (11), and $W_n(\hat{\theta}_\beta)$, given in (20), are respectively

\begin{align*}
\mathcal{I}_n^2(x, W^0_\beta, F_\theta) &= 2 \left( u_\theta(x) f_{\hat{\theta}_\beta}^\beta(x) - \xi(\theta_0) \right)^T J_\beta^{-1}(\theta_0) \Sigma_\beta^{-1}(\theta_0) J_\beta^{-1}(\theta_0) \left( u_\theta(x) f_{\hat{\theta}_\beta}^\beta(x) - \xi(\theta_0) \right), \\
\mathcal{I}_n^2(x, W_\beta, F_\theta) &= 2 \left( u_\theta(x) f_{\hat{\theta}_\beta}^\beta(x) - \xi(\theta_0) \right)^T J_\beta^{-1}(\theta_0) M(\theta_0) \left[ M^T(\theta_0) \Sigma_\beta(\theta_0) M(\theta_0) \right]^{-1} \\
&\quad \times M^T(\theta_0) J_\beta^{-1}(\theta_0) \left( u_\theta(x) f_{\hat{\theta}_\beta}^\beta(x) - \xi(\theta_0) \right).
\end{align*}

Remark 2 In most of the cases $K_\beta(\theta_0)$ is a full rank matrix and so

\begin{align*}
\mathcal{I}_n^2(x, W^0_\beta, F_\theta) &= 2 \left( u_\theta(x) f_{\hat{\theta}_\beta}^\beta(x) - \xi(\theta_0) \right)^T K_\beta^{-1}(\theta_0) \left( u_\theta(x) f_{\hat{\theta}_\beta}^\beta(x) - \xi(\theta_0) \right).
\end{align*}

### 4 Level and Power Influence Functions: Simple Null Hypothesis

In this section we shall study the influence of contamination on the level and power to measure the robustness properties of the Wald-type test when the simple null hypothesis is considered. For a finite sample size, in general, it is difficult to calculate the level and power, and therefore, we shall use asymptotic approximations. At a fixed alternative the power function of the Wald-type test statistic was given in equation (13). This power function tends to one as $n$ increases, so the test is consistent in the Fraser’s sense. Therefore, it is important to calculate power functions at the contiguous alternatives as mentioned in (14). In this case the asymptotic power function can be approximated using (16).

Now we shall consider the sequence of alternatives $\theta_n = \theta_0 + n^{-1/2}d$ as given in (14). When $\theta_n$ tends to $\theta_0$ the contamination proportion is also assumed to tend to zero at the same rate. Therefore, we shall define the contaminated distributions for the power as

\begin{align*}
F_{n,\varepsilon,x}^P &= (1 - \frac{\varepsilon}{\sqrt{n}})F_\theta + \frac{\varepsilon}{\sqrt{n}} \Delta_x,
\end{align*}

where $\Delta_x$ denotes the degenerate distribution function with all its mass concentrated at point $x$, and $\varepsilon/\sqrt{n}$ is the contamination proportion. Substituting $d = 0_p$ in equation (34) we get the contaminated distributions for the level as

\begin{align*}
F_{n,\varepsilon,x}^L &= (1 - \frac{\varepsilon}{\sqrt{n}})F_\theta + \frac{\varepsilon}{\sqrt{n}} \Delta_x.
\end{align*}

Let us define

\begin{align*}
\alpha_{W_\beta}(\varepsilon, x) &= \lim_{n \to \infty} P_{P_{\varepsilon,x}}(W_\beta(\hat{\theta}_\beta) > \chi^2_{p,\alpha}), \\
\beta_{W_\beta}(\theta_n, \varepsilon, x) &= \lim_{n \to \infty} P_{P_{\varepsilon,x}}(W_n(\hat{\theta}_\beta) > \chi^2_{p,\alpha}).
\end{align*}

The level influence functions associated with the Wald-type test statistics for simple and composite null hypotheses are defined as

\begin{align*}
\mathcal{L}_n^L(x; W^0_\beta, F_\theta) &= \frac{\partial}{\partial \varepsilon} \alpha_{W_\beta}(\varepsilon, x) \bigg|_{\varepsilon=0}, \\
\mathcal{L}_n^L(x; W_\beta, F_\theta) &= \frac{\partial}{\partial \varepsilon} \beta_{W_n}(\theta_n, \varepsilon, x) \bigg|_{\varepsilon=0}.
\end{align*}
Similarly, we define the power influence functions as
\[
\mathcal{PIL}(x; W^\alpha_\beta, F_{\theta_0}) = \frac{\partial}{\partial \varepsilon} \beta_{W^\alpha_\beta}(\theta_n, \varepsilon, x) \bigg|_{\varepsilon=0}, \quad \mathcal{PIL}(x; W_\beta, F_{\theta_0}) = \frac{\partial}{\partial \varepsilon} \beta_{W_n}(\theta_n, \varepsilon, x) \bigg|_{\varepsilon=0}.
\]
For more details see Hampel et al. (1986, Section 3.2c).

Let us denote the quadratic form of a symmetric matrix \( A_{p \times p} \) as \( q_A(z) = z^T A z \). We shall frequently use the following result that
\[
q_A(z + h) = q_A(z) + 2h^T A z + q_A(h),
\]
where \( z \) and \( h \) are two vectors in \( \mathbb{R}^p \).

In the following theorem the asymptotic power under contaminated distribution is presented for the case of the simple null hypothesis. For the rest of the paper we shall consider Assumptions A–D of Lehmann (1983, page 429) as well as Conditions D1–D5 of Basu et al. (2011, page 311), which we shall refer to as the Lehmann and Basu et al. conditions respectively.

**Theorem 3** Assume that the Lehmann and Basu et al. conditions hold for the model. Let us consider the contiguous alternatives in (14) against the simple null hypothesis, and the underlying contaminated model as given in (34). Then we have the following asymptotic power approximation
\[
\beta_{W^\alpha_\beta}(\theta_n, \varepsilon, x) \approx 1 - \int \Phi \left( \frac{\chi^2_{p,\alpha} - y - d_{x,\beta}(\theta_0)\Sigma^{-1}_{\beta}(\theta_0)d_{x,\beta}(\theta_0)}{2\sqrt{d_{x,\beta}(\theta_0)\Sigma^{-1}_{\beta}(\theta_0)d_{x,\beta}(\theta_0)}} \right) dF_{\theta_0}(y),
\]
where \( d_{x,\beta}(\theta_0) = d + \varepsilon \mathcal{IF}(x, T_\beta, F_{\theta_0}), \mathcal{IF}(x, T_\beta, F_{\theta_0}) \) is given by (29) and \( F_{\theta_0} \) is the \( \chi^2 \) distribution function.

**Proof.** Let us denote \( \theta_n^* = T_\beta(F^P_{n,x}) \). Using equation (35), with \( z = \hat{\theta}_\beta - \theta_n^* \) and \( h = \theta_n^* - \theta_0 \), we get
\[
W^0_n(\hat{\theta}_\beta) = q_n \Sigma^{-1}_{\beta}(\theta_0)(\hat{\theta}_\beta - \theta_0) = q_n \Sigma^{-1}_{\beta}(\theta_0) \left( (\hat{\theta}_\beta - \theta_n^*) + (\theta_n^* - \theta_0) \right)
= q_n \Sigma^{-1}_{\beta}(\theta_0)(\hat{\theta}_\beta - \theta_n^*) + 2n(\hat{\theta}_\beta - \theta_n^*)^T \Sigma^{-1}_{\beta}(\theta_0)(\theta_n^* - \theta_0) + q_n \Sigma^{-1}_{\beta}(\theta_0)(\theta_n^* - \theta_0),
\]
i.e.,
\[
W^0_n(\hat{\theta}_\beta) = W^0_n(\theta_n^*) + q_n \Sigma^{-1}_{\beta}(\theta_0)(\hat{\theta}_\beta - \theta_n^*) + 2n(\hat{\theta}_\beta - \theta_n^*)^T \Sigma^{-1}_{\beta}(\theta_0)(\theta_n^* - \theta_0).
\]
Let us consider \( \theta_n^* \) as a function of \( \varepsilon_n = \varepsilon/\sqrt{n} \), i.e., \( \theta_n^* = f(\varepsilon_n) \). A Taylor series expansion of \( f(\varepsilon_n) \) at \( \varepsilon_n = 0 \) gives
\[
f(\varepsilon_n) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\partial^k f(\varepsilon_n)}{\partial \varepsilon^k} \bigg|_{\varepsilon_n=0} = \theta_0 + \frac{\varepsilon_n}{\sqrt{n}} \mathcal{IF}(x, T_\beta, F_{\theta_0}) + \sum_{k=2}^{\infty} \frac{1}{k!} \left( \frac{\varepsilon_n}{\sqrt{n}} \right)^k \mathcal{IF}_k(x, T_\beta, F_{\theta_0}).
\]
Therefore, we get
\[
\sqrt{n}(\theta_n^* - \theta_n) = \varepsilon \mathcal{IF}(x, T_\beta, F_{\theta_0}) + o_p(1_p),
\]
\[
\sqrt{n}(\theta_n^* - \theta_0 - n^{-1/2}d) = \varepsilon \mathcal{IF}(x, T_\beta, F_{\theta_0}) + o_p(1_p),
\]
and thus
\[ \sqrt{n}(\theta^*_n - \theta_0) = d + \varepsilon \mathcal{I}(x, T_\beta, F_{\theta_0}) + o_p(1) \]
\[ = \tilde{d}_{\varepsilon, x, \beta}(\theta_0) + o_p(1). \tag{36} \]

So
\[ W_n^0(\theta^*_n) = \tilde{d}^T_{\varepsilon, x, \beta}(\theta_0) \Sigma^{-1}_\beta(\theta_0) \tilde{d}_{\varepsilon, x, \beta}(\theta_0) + o_p(1), \]
\[ 2\sqrt{n}(\hat{\theta}_n - \theta^*_n)^T \Sigma^{-1}_\beta(\theta_0) \sqrt{n}(\theta^*_n - \theta_0) = 2\sqrt{n}(\hat{\theta}_n - \theta^*_n)^T \Sigma^{-1}_\beta(\theta_0) \left( \tilde{d}_{\varepsilon, x, \beta}(\theta_0) + o_p(1) \right). \]

Hence \( W_n^0(\hat{\theta}_n) = W_1 + W_2 + o_p(1), \) with
\[ W_1 = q_n \Sigma^{-1}_\beta(\theta_0) (\hat{\theta}_n - \theta^*_n), \]
\[ W_2 = \tilde{d}^T_{\varepsilon, x, \beta}(\theta_0) \Sigma^{-1}_\beta(\theta_0) \tilde{d}_{\varepsilon, x, \beta}(\theta_0) + 2\sqrt{n}(\hat{\theta}_n - \theta^*_n)^T \Sigma^{-1}_\beta(\theta_0) \tilde{d}_{\varepsilon, x, \beta}(\theta_0). \]

As
\[ \sqrt{n}(\hat{\theta}_n - \theta^*_n) \overset{\mathcal{L}}{\underset{n \to \infty}{\to}} N(0_p, \Sigma_{\beta}(\theta_0)), \tag{37} \]
we get \( W_1 \overset{\mathcal{L}}{\underset{n \to \infty}{\to}} \chi^2_{p} \). Finally,
\[ W_2 \overset{\mathcal{L}}{\underset{n \to \infty}{\to}} N(\tilde{d}^T_{\varepsilon, x, \beta}(\theta_0) \Sigma^{-1}_\beta(\theta_0) \tilde{d}_{\varepsilon, x, \beta}(\theta_0), 4 \tilde{d}^T_{\varepsilon, x, \beta}(\theta_0) \Sigma^{-1}_\beta(\theta_0) \tilde{d}_{\varepsilon, x, \beta}(\theta_0)). \]

Hence, we get
\[ \beta_{W_n}(\theta_n, \varepsilon, x) = \lim_{n \to \infty} P_{F_{\varepsilon, x}}(W_n^0(\hat{\theta}_n) > \chi^2_{p, \alpha(\theta_n, \varepsilon, x)}) \]
\[ \equiv P(W_1 + W_2 > \chi^2_{p, \alpha}) = \int P(W_2 > \chi^2_{p, \alpha} - y) dF_{\theta_0}(y) \]
\[ = 1 - \int \Phi \left( \frac{\chi^2_{p, \alpha} - y - \tilde{d}^T_{\varepsilon, x, \beta}(\theta_0) \Sigma^{-1}_\beta(\theta_0) \tilde{d}_{\varepsilon, x, \beta}(\theta_0)}{2\sqrt{\tilde{d}^T_{\varepsilon, x, \beta}(\theta_0) \Sigma^{-1}_\beta(\theta_0) \tilde{d}_{\varepsilon, x, \beta}(\theta_0)}} \right) dF_{\theta_0}(y). \]

Further, substituting \( d = 0_p \) or \( \varepsilon = 0 \) in above theorem, we shall get several important cases; these are presented in the following corollaries.

**Corollary 4** Putting \( \varepsilon = 0 \) in the above theorem, we get the asymptotic power under the contiguous alternative hypotheses (14) as
\[ \beta_{W_n}(\theta_n) = \beta_{W_n}(\theta_n, 0, x) \equiv 1 - \int \Phi \left( \frac{\chi^2_{p, \alpha} - y - \tilde{d}^T_{\varepsilon, x, \beta}(\theta_0) \Sigma^{-1}_\beta(\theta_0) d}{2\sqrt{d^T \Sigma^{-1}_\beta(\theta_0) d}} \right) dF_{\theta_0}(y). \]

Notice that Corollary 4 is an alternative approximation for the result given in (16).

**Corollary 5** Putting \( d = 0_p \) in the above theorem, we get the asymptotic level under the probability distribution \( F^L_{\varepsilon, x} \) as
\[ \alpha_{W_n}(\varepsilon, x) = \beta_{W_n}(\theta_0, \varepsilon, x) \]
\[ \equiv 1 - \int \Phi \left( \frac{\chi^2_{p, \alpha} - y - \varepsilon^2 \mathcal{I}(x, T_\beta, F_{\theta_0})^T \Sigma^{-1}_\beta(\theta_0) \mathcal{I}(x, T_\beta, F_{\theta_0})}{2\varepsilon \sqrt{\mathcal{I}(x, T_\beta, F_{\theta_0})^T \Sigma^{-1}_\beta(\theta_0) \mathcal{I}(x, T_\beta, F_{\theta_0})}} \right) dF_{\theta_0}(y). \]
In particular, taking $\varepsilon \to 0$ in the above expression, we get the asymptotic level of the Wald-type test statistic as

$$
\alpha_0 = 1 - (1 - F_{H_0}(\chi^2_{p,0})) = \alpha.
$$

This was the expected result according to the construction of the test statistic and its critical value. Next we derive the power influence function of the Wald-type test statistic.

**Theorem 6** Assume that the Lehmann and Basu et al. conditions hold for the model. Then, the power influence function of the Wald-type test statistic under the simple null hypothesis is given by

$$
\mathcal{PIF}(x, W_{\beta}^0, F_{\theta_0}) = \frac{d_T \Sigma_{\beta}^{-1}(\theta_0) \mathcal{IL}(x, T_\beta, F_{\theta_0})}{2 \sqrt{d_T \Sigma_{\beta}^{-1}(\theta_0) d}} \int \phi \left( \frac{\chi^2_{p,0} - y - d_T \Sigma_{\beta}^{-1}(\theta_0) d}{2 \sqrt{d_T \Sigma_{\beta}^{-1}(\theta_0) d}} \right) \left( \frac{\chi^2_{p,0} - y}{d_T \Sigma_{\beta}^{-1}(\theta_0) d} + 1 \right) dF_{H_0}(y).
$$

(38)

**Proof.** Let us consider the expression of $\beta_{W_n}(\theta_n, \varepsilon, x)$ as obtained in Theorem 3. Note that

$$
d^T_{x, \beta}(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) d_{x, \beta}(\theta_0) = C_0 + \varepsilon C_1 + \varepsilon^2 C_2,
$$

where

$$
C_0 = d^T \Sigma_{\beta}^{-1}(\theta_0) d, \\
C_1 = 2d^T \Sigma_{\beta}^{-1}(\theta_0) \mathcal{IL}(x, T_\beta, F_{\theta_0}), \\
C_2 = \mathcal{IL}(x, T_\beta, F_{\theta_0})^T \Sigma_{\beta}^{-1}(\theta_0) \mathcal{IL}(x, T_\beta, F_{\theta_0}).
$$

So

$$
\frac{\chi^2_{p,0} - y - d^T_{x, \beta}(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) d_{x, \beta}(\theta_0)}{2 \sqrt{d^T_{x, \beta}(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) d_{x, \beta}(\theta_0)}} = \frac{\chi^2_{p,0} - y - C_0 - \varepsilon C_1 - \varepsilon^2 C_2}{2 \sqrt{C_0 + \varepsilon C_1 + \varepsilon^2 C_2}}
$$

(39)

$$
= \frac{\chi^2_{p,0} - y}{2} \left( C_0 + \varepsilon C_1 + \varepsilon^2 C_2 \right)^{-\frac{1}{2}} - \frac{1}{2} \left( C_0 + \varepsilon C_1 + \varepsilon^2 C_2 \right)^{\frac{1}{2}}.
$$

Since

$$
\frac{\partial}{\partial \varepsilon} \left( \frac{\chi^2_{p,0} - y}{2} \left( C_0 + \varepsilon C_1 + \varepsilon^2 C_2 \right)^{-\frac{1}{2}} - \frac{1}{2} \left( C_0 + \varepsilon C_1 + \varepsilon^2 C_2 \right)^{\frac{1}{2}} \right) \bigg|_{\varepsilon = 0} = -\frac{\chi^2_{p,0} - y}{4} \left( C_0 + 2\varepsilon C_2 \right) \left( C_0 + \varepsilon C_1 + \varepsilon^2 C_2 \right)^{-\frac{3}{2}} - \frac{1}{4} \left( C_0 + 2\varepsilon C_2 \right) \left( C_0 + \varepsilon C_1 + \varepsilon^2 C_2 \right)^{-\frac{1}{2}} \bigg|_{\varepsilon = 0}
$$

$$
= -\frac{C_1}{4\sqrt{C_0}} \left( \frac{\chi^2_{p,0} - y}{C_0} + 1 \right),
$$

we have

$$
\frac{\partial}{\partial \varepsilon} \beta_{W_n}(\theta_n, \varepsilon, x) \bigg|_{\varepsilon = 0} = -\int \frac{\partial}{\partial \varepsilon} \Phi \left( \frac{\chi^2_{p,0} - y - C_0 - \varepsilon C_1 - \varepsilon^2 C_2}{2 \sqrt{C_0 + \varepsilon C_1 + \varepsilon^2 C_2}} \right) \bigg|_{\varepsilon = 0} dF_{H_0}(y)
$$

$$
= \frac{C_1}{4\sqrt{C_0}} \int \Phi \left( \frac{\chi^2_{p,0} - y}{2 \sqrt{C_0}} \right) \left( \frac{\chi^2_{p,0} - y}{C_0} + 1 \right) dF_{H_0}(y),
$$

where $\Phi(\cdot)$ is the standard normal density function. □
Since
\[ \phi \left( \chi_{p,\alpha}^2 - y - d^T \Sigma^{-1}_\beta(\theta_0) d \right) \left( \chi_{p,\alpha}^2 - y + d^T \Sigma^{-1}_\beta(\theta_0) d + 1 \right) \leq \phi(0) \left( \chi_{p,\alpha}^2 - y + d^T \Sigma^{-1}_\beta(\theta_0) d + 1 \right), \]
it follows that
\[ \int \phi \left( \chi_{p,\alpha}^2 - y - d^T \Sigma^{-1}_\beta(\theta_0) d \right) \left( \chi_{p,\alpha}^2 - y + d^T \Sigma^{-1}_\beta(\theta_0) d + 1 \right) dF_{H_0}(y) \]
\[ \leq \phi(0) \left( \frac{\chi_{p,\alpha}^2}{d^T \Sigma^{-1}_\beta(\theta_0) d} + 1 - \frac{1}{d^T \Sigma^{-1}_\beta(\theta_0) d} \int y dF_{H_0}(y) \right) \]
\[ = \phi(0) \left( \frac{\chi_{p,\alpha}^2 - p}{d^T \Sigma^{-1}_\beta(\theta_0) d} + 1 \right) < \infty. \]
Thus the above theorem shows that the power influence function is bounded whenever the influence function of the MDPDE is bounded.

To calculate the level influence function, we start from the expression of \( \alpha_{W^2}(\varepsilon, x) \) as given in Corollary 5. Since
\[ \frac{\partial}{\partial \varepsilon} \Phi \left( \chi_{p,\alpha}^2 - y - \varepsilon \mathcal{I}F(x, T_\beta, F_{\theta_0})^T \Sigma^{-1}_\beta(\theta_0) \mathcal{I}F(x, T_\beta, F_{\theta_0}) \right) \bigg|_{\varepsilon=0} = 0, \]
we obtain that
\[ \mathcal{L}IF(x, W^0_\beta, F_{\theta_0}) = \frac{\partial}{\partial \varepsilon} \alpha_{W^2}(\varepsilon, x) \bigg|_{\varepsilon=0} = 0. \]
Also, it is easy to see that the derivative of \( \alpha_{W^2}(\varepsilon, x) \) of any order will be zero at \( \varepsilon = 0 \), implying that the level influence function of any order will be zero. Thus, asymptotically, the level of the Wald-type test statistic will be unaffected by a contiguous contamination.

## 5 Level and Power Influence Functions: Composite Hypothesis

We shall now calculate the level and power influence functions of the Wald-type test statistic for the composite null hypothesis. We have considered the same setting as mentioned in Section 4.

**Theorem 8** Assume that the Lehmann and Basu et al. conditions hold for the model. Let us consider the contiguous alternatives in (14) against the composite null hypothesis, and the underlying contaminated model as given in (34). Then we have the following asymptotic power approximation
\[ \beta_{W_n}(\theta_n, \varepsilon, x) \equiv 1 - \int \Phi \left( \chi_{p,\alpha}^2 - y - d^T \Sigma^{-1}_\beta(\theta_0) M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) \tilde{d}_{\varepsilon,\beta}(\theta_0) \right) dF_{H_0}(y), \]
where \( \Sigma(\theta_0) = M^T(\theta_0) \Sigma(\theta_0) M(\theta_0) \), \( \tilde{d}_{\varepsilon,\beta}(\theta_0) = d + \varepsilon \mathcal{I}F(x, T_\beta, F_{\theta_0}) \), \( \mathcal{I}F(x, T_\beta, F_{\theta_0}) \) is given by (29) and \( F_{H_0} \) is the \( \chi^2 \) distribution function.
Proof. Let us denote \( \theta_n^* = T_\beta ( L_{n,i,x} ) \). Using equation (35), with \( z = m(\hat{\theta}_\beta) - m(\theta_n^*) \) and \( h = m(\theta_n^*) \), we get

\[
W_n(\hat{\theta}_\beta) = q_n \Sigma_{\beta}^{-1}(\theta_0) ( m(\hat{\theta}_\beta) - m(\theta_n^*) ) = q_n \Sigma_{\beta}^{-1}(\theta_0) \left( (m(\hat{\theta}_\beta) - m(\theta_n^*)) + (m(\theta_n^*) - m(\theta_n^*)) \right)
\]

\[
= q_n \Sigma_{\beta}^{-1}(\theta_0) \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right) + 2n \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right)^T \Sigma_{\beta}^{-1}(\theta_0) m(\theta_n^*) + q_n \Sigma_{\beta}(\theta_0)^{-1}(m(\theta_n^*) - m(\theta_n^*))
\]

i.e.,

\[
W_n(\hat{\theta}_\beta) = W_n(\theta_n^*) + q_n \Sigma_{\beta}^{-1}(\theta_0) \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right) + 2n \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right)^T \Sigma_{\beta}^{-1}(\theta_0) m(\theta_n^*).
\]

Now, as in the proof of Theorem 3, we can show that

\[
\sqrt{n}(\theta_n^* - \theta_0) = d + \varepsilon TF(x, T_\beta, F_{\theta_0}) + o_p(1).
\]

Using a Taylor series expansion, we get

\[
m(\theta_n^*) = m(\theta_0) + M^T(\theta_0) (\theta_n^* - \theta_0) + o(||\theta_n^* - \theta_0||).
\]

As \( m(\theta_0) = 0 \), from (41) it follows that

\[
\sqrt{n}m(\theta_n^*) = M^T(\theta_0) \tilde{d}_{\ell,x,\beta}(\theta_0) + o_p(1).
\]

Further, since (37) holds, a similar Taylor series expansion of (42) yields

\[
\sqrt{n} \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right) \xrightarrow{c} \mathcal{N}(0_r, \Sigma_{\beta}(\theta_0)),
\]

where \( \Sigma_{\beta}(\theta_0) = M^T(\theta_0) \Sigma_{\beta}(\theta_0) M(\theta_0) \). Thus, we get

\[
q_n \Sigma_{\beta}^{-1}(\theta_0) \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right) \xrightarrow{c} \chi^2_r.
\]

Also, from (41) we have

\[
W_n(\theta_n^*) = \tilde{d}_{\ell,x,\beta}(\theta_0) M(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) \tilde{d}_{\ell,x,\beta}(\theta_0) + o_p(1).
\]

\[
2\sqrt{n} \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right)^T \Sigma_{\beta}^{-1}(\theta_0) \sqrt{n} m(\theta_n^*) = 2\sqrt{n} \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right)^T \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) \left( \tilde{d}_{\ell,x,\beta}(\theta_0) + o_p(1_r) \right).
\]

Hence \( W_n(\theta_n^*) = W_1 + W_2 + o_p(1) \), with

\[
W_1^* = q_n \Sigma_{\beta}^{-1}(\theta_0) ( m(\hat{\theta}_\beta) - m(\theta_n^*) ),
\]

\[
W_2^* = \tilde{d}_{\ell,x,\beta}(\theta_0) M(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) \tilde{d}_{\ell,x,\beta}(\theta_0) + 2\sqrt{n} \left( m(\hat{\theta}_\beta) - m(\theta_n^*) \right)^T \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) \tilde{d}_{\ell,x,\beta}(\theta_0).
\]

As it holds (37), we get \( W_1^* \xrightarrow{c} \chi^2_r. \) Finally,

\[
W_2^* \xrightarrow{c} N(\tilde{d}_{\ell,x,\beta}(\theta_0) M(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) \tilde{d}_{\ell,x,\beta}(\theta_0), 4\tilde{d}_{\ell,x,\beta}(\theta_0) M(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) \tilde{d}_{\ell,x,\beta}(\theta_0)).
\]
Hence, we get
\[
\beta_{W_n}(\theta_n, \varepsilon, x) = \lim_{n \to \infty} P_{F_{n,\varepsilon,x}}(W_n(\hat{\theta}_n) > \chi_{r,\alpha}^2) = P(W_1^* + W_2^* > \chi_{r,\alpha}^2) = \int P(W_2^* > \chi_{r,\alpha}^2 - y) dF_{H_0}(y) = 1 - \int \Phi \left( \frac{\chi_{r,\alpha}^2 - y - d^T \hat{\theta}_n M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) \hat{d}_{\varepsilon,\beta}(\theta_0)}{2\sqrt{d^T \hat{\theta}_n M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) \hat{d}_{\varepsilon,\beta}(\theta_0)}} \right) dF_{H_0}(y).
\]

Further, substituting \( d = 0_p \) or \( \varepsilon = 0 \) in above theorem, we shall get several important cases; these are presented in the following corollaries.

**Corollary 9** Putting \( \varepsilon = 0 \) in the above theorem, we get the asymptotic power under the contiguous alternatives as
\[
\beta_{W_n}(\theta_n) = \beta_{W_n}(\theta_n,0,x) = 1 - \int \Phi \left( \frac{\chi_{r,\alpha}^2 - y - d^T M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) d}{2\sqrt{d^T M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) d}} \right) dF_{H_0}(y).
\]
Notice that Corollary 9 is an alternative approximation of the power function given in (21).

**Corollary 10** Putting \( d = 0_p \) in the above theorem, we get the asymptotic level under the probability distribution \( F_{n,\varepsilon,x} \) as
\[
\alpha_{W_n}(\varepsilon,x) = \beta_{W_n}(\theta_0,\varepsilon,x) \approx 1 - \int \Phi \left( \frac{\chi_{r,\alpha}^2 - y - \varepsilon^2 T_{\varepsilon,x,\beta}(\theta_0) M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) T_{\varepsilon,x,\beta}(\theta_0)}{2\varepsilon \sqrt{T_{\varepsilon,x,\beta}(\theta_0) M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) T_{\varepsilon,x,\beta}(\theta_0)}} \right) dF_{H_0}(y).
\]
In particular, taking \( \varepsilon \to 0 \) in the above expression, we get the asymptotic level of the test statistics as
\[
\alpha_0 = 1 - (1 - F_{H_0}(\chi_{r,\alpha}^2)) = \alpha.
\]
This was the expected result according to the construction of the test statistic and its critical value. Next we derive the power influence function of the proposed test statistic.

**Theorem 11** Assume that the Lehmann and Basu et al. conditions hold for the model. Then, the power influence function of the proposed Wald-type test statistic under the composite null hypothesis is given by
\[
P\mathcal{I}(x,W_\beta,F_{\theta_0}) \equiv \frac{d^T M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) T(x,\beta,F_{\theta_0})}{2\sqrt{d^T M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) d}} \times \int \Phi \left( \frac{\chi_{r,\alpha}^2 - y - d^T M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) d}{2\sqrt{d^T M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) d}} \right) \left( \frac{\chi_{r,\alpha}^2 - y}{d^T M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) d} + 1 \right) dF_{H_0}(y).
\]

**Proof.** Let us consider the expression of \( \beta_{W_n}(\theta_n, \varepsilon, x) \) as obtained in Theorem 8. Note that
\[
\hat{d}_{\varepsilon,\beta}(\theta_0) M(\theta_0) \Sigma^{-1}_\beta(\theta_0) M^T(\theta_0) \hat{d}_{\varepsilon,\beta}(\theta_0) = C_0^* + \varepsilon C_1^* + \varepsilon^2 C_2^*.
\]
Also, it is easy to see that the derivative of the composite null hypothesis is bounded whenever the influence function of the MDPDE is bounded. So, the power influence function of the Wald-type test statistic under the composite test is bounded. Since

\[ \partial \frac{\chi_{r,\alpha}^2 - y - d_{r,\beta}^T M(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) d_{r,\beta}(\theta_0)}{2 \sqrt{d_{r,\beta}^T M(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) d_{r,\beta}(\theta_0)}} = \frac{\chi_{r,\alpha}^2 - y - C_0^* - \varepsilon C_1^* - \varepsilon^2 C_2^*}{2 \sqrt{C_0^* + \varepsilon C_1^* + \varepsilon^2 C_2^*}} \]  

we have

\[ \frac{\partial}{\partial \varepsilon} (\chi_{r,\alpha}^2 - y - C_0^* - \varepsilon C_1^* - \varepsilon^2 C_2^*)^{-\frac{1}{2}} - \frac{1}{2} (C_0^* + \varepsilon C_1^* + \varepsilon^2 C_2^*)^{\frac{1}{2}} \bigg|_{\varepsilon=0} = \frac{C_1^*}{4 \sqrt{C_0^*}} \left( \frac{\chi_{r,\alpha}^2 - y - C_0^*}{C_0^*} + 1 \right) \]

and

\[ \frac{\partial}{\partial \varepsilon} \beta W_n(\theta_n, \varepsilon, x) \big|_{\varepsilon=0} \approx - \int \frac{\partial}{\partial \varepsilon} \Phi \left( \frac{\chi_{r,\alpha}^2 - y - C_0^* - \varepsilon C_1^* - \varepsilon^2 C_2^*}{2 \sqrt{C_0^* + \varepsilon C_1^* + \varepsilon^2 C_2^*}} \right) \bigg|_{\varepsilon=0} dF_{H_0}(y) = \frac{C_1^*}{4 \sqrt{C_0^*}} \int \phi \left( \frac{\chi_{r,\alpha}^2 - y - C_0^*}{2 \sqrt{C_0^*}} \right) \left( \frac{\chi_{r,\alpha}^2 - y - C_0^*}{C_0^*} + 1 \right) dF_{H_0}(y), \]

where \( \phi(\cdot) \) is the standard normal density function. □

As in Remark 7 of Section 4, we can show that the integral in the above expression of the power influence function of the composite test is bounded. So, the power influence function of the Wald-type test statistic under the composite null hypothesis is bounded whenever the influence function of the MDPDE is bounded.

To calculate the level influence function, we start from the expression of \( \alpha_{W_n}(\varepsilon, x) \) as given in Corollary 10. Since

\[ \frac{\partial}{\partial \varepsilon} \Phi \left( \frac{\chi_{r,\alpha}^2 - y - \varepsilon^2 \mathcal{I}F(x, T_\beta, F_{\theta_n})^T M(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) \mathcal{I}F(x, T_\beta, F_{\theta_n})}{2 \varepsilon \sqrt{\mathcal{I}F(x, T_\beta, F_{\theta_n})^T M(\theta_0) \Sigma_{\beta}^{-1}(\theta_0) M^T(\theta_0) \mathcal{I}F(x, T_\beta, F_{\theta_n})}} \right) \bigg|_{\varepsilon=0} = 0, \]

we obtain that

\[ \mathcal{I}F(x, W_\beta, F_{\theta_n}) = \frac{\partial}{\partial \varepsilon} \alpha_{W_n}(\varepsilon, x) \big|_{\varepsilon=0} = 0. \]

Also, it is easy to see that the derivative of \( \alpha_{W_n}(\varepsilon, x) \) of any order will be zero at \( \varepsilon = 0 \), implying that the level influence function of any order will be zero. Thus, asymptotically, the level of the proposed test statistics will be unaffected by a contiguous contamination.
6 The Chi-Square Inflation Factor

Another important way of measuring the robustness of a test statistic is to look at its asymptotic distribution for a general contaminated distribution, in contrast to its null distribution under the model. Unlike the contiguous contamination considered in the previous section, we shall now consider a fixed departure from the model independent of the sample size. Under the set-up of the previous sections, let us assume that the data come from a general contaminated distribution \( G \) having density \( g \). The null hypothesis, mentioned in (12), can be written as

\[
H_0 : T_\beta(G) = \theta_0.
\]  
(46)

The asymptotic distribution of MDPDE under the model is given in (9). We shall now derive the asymptotic null distribution of the Wald-type test statistic under a general distribution \( G \). Let us define

\[
J_{\beta,g}(\theta) = \int u_\theta(x)u_\theta^T(x)f_\theta^{1+\beta}(x)dx + \int (I_\theta(x) - \beta u_\theta(x)u_\theta^T(x))(g(x) - f_\theta(x))f_\theta(\beta)dx,
\]  
(47)

and

\[
K_{\beta,g}(\theta) = \int u_\theta(x)u_\theta^T(x)f_\theta^{2\beta}(x)g(x)dx - \xi_\theta(\theta)\xi_\theta^T(\theta),
\]  
(48)

where \( \xi_\beta,\gamma(\theta) = \int u_\theta(x)f_\theta^\beta(x)g(x)dx \) and \( I_\theta(x) = -\frac{\partial}{\partial \theta}u_\theta^T(x) \), the so-called information matrix of the model.

Let \( \hat{\theta}_{\beta,g}(\theta) = T_\beta(G_n) \) be the MDPDE with tuning parameter \( \beta \). Basu et al. (1998) and Basu et al. (2011) established that

\[
n^{1/2}(\hat{\theta}_{\beta,g} - \theta_0) \xrightarrow{L} N(0, \Sigma_{\beta,g}(\theta_0)),
\]  
(49)

where

\[
\Sigma_{\beta,g}(\theta_0) = J_{\beta,g}^{-1}(\theta_0)K_{\beta,g}(\theta_0)J_{\beta,g}^{-1}(\theta_0).
\]  
(50)

In Section 2.1 we have derived the asymptotic distribution of the Wald-type test statistic under the simple null hypothesis when \( G = F_{\theta_0} \). Our next theorem will show the asymptotic null distribution of the Wald-type test under the general set-up when the underlying density may or may not belong to the model.

**Theorem 12** Let \( \hat{\theta}_{\beta,g} = T_\beta(G_n) \) be the MDPDE with tuning parameter \( \beta \). Then under the null hypothesis (46), the asymptotic distribution of the Wald-type test statistic is given by

\[
W_n^{0}(\hat{\theta}_{\beta,g}) = n(\hat{\theta}_{\beta,g} - \theta_0)^T\Sigma_{\beta}^{-1}(\theta_0)(\hat{\theta}_{\beta,g} - \theta_0) \xrightarrow{L} \sum_{i=1}^p c_{i,\beta,g}(\theta_0)Z_i^2,
\]  
(51)

where \( \{Z_i\}_{i=1}^p \) are i.i.d. standard normal random variables and \( \{c_{i,\beta,g}(\theta_0)\}_{i=1}^p \) the set of eigenvalues of \( \Sigma_{\beta}^{-1}(\theta_0)\Sigma_{\beta,g}(\theta_0) \).

**Proof.** The result follows from (49), using Corollary 2.2 of Dik and de Gunst (1985).

The above theorem shows that the asymptotic null distribution of the Wald-type test statistic is a linear combination of independent \( \chi^2_1 \) random variables. On the other hand, if the assumed model is correct, the
asymptotic null distribution turns out to be $\chi_p^2$. In this context, by following Satterthwaite (1946), our proposal is using $\bar{c}_{\beta,g}(\theta_0)\chi_p^2$, with
\[
\bar{c}_{\beta,g}(\theta_0) = p \sum_{i=1}^{p} c_{i,\beta,g}(\theta_0) = \frac{1}{p} \text{trace}\left(\Sigma^{-1}_\beta(\theta_0)\Sigma_{\beta,g}(\theta_0)\right),
\]
(52)
to approximate $\sum_{i=1}^{p} c_{i,\beta,g}(\theta_0)Z_i^2$. This factor is called Chi-Square Inflation Factor (CSIF) and its value is equal to unity if only if $\Sigma_{\beta,g}(\theta_0) = \Sigma_\beta(\theta_0)$. Since a value close to unity indicates strong robustness towards the model assumption of the Wald-type test statistic, $\bar{c}_{\beta,g}(\theta_0)$ is useful as a measure of robustness. Ghosh et al. (2014) used this approach to illustrate the stability of the tests based on the S-divergence when $p = 1$. When $p = 1$ the CSIF becomes
\[
\bar{c}_{\beta,g}(\theta_0) = c_{1,\beta,g}(\theta_0) = \frac{J^2(\theta_0)}{K(\theta_0)} K_{\beta,g}(\theta_0).
\]
In this case, the asymptotic null distribution of the Wald-type test statistic is exactly (not approximately) $\bar{c}_{\beta,g}(\theta_0)\chi_1^2$.

We shall now illustrate the effect of outliers in CSIF. Let us consider the following fixed point contaminated density
\[
f_{x,y}() = (1-\varepsilon)f_{\theta_0}() + \varepsilon\Delta_y,
\]
where $\varepsilon \in (0,1)$ is the contamination proportion, and $y$ is the outlying point. Note that, the rate of change in $\bar{c}_{\beta,c,x,y}(\theta_0)$ with respect to $\varepsilon$ at the origin gives us the effect of infinitesimal contamination on the test statistic. Similar interpretation as the influence function analysis may be drawn in this case; and the boundedness of the above mentioned quantity will indicate robustness towards the assumed model. So $\frac{\partial}{\partial \varepsilon} \bar{c}_{\beta,c,x,y}(\theta_0)|_{\varepsilon=0}$ may be regarded as another robustness measure in this context. Our next theorem gives the explicit form of this index.

Theorem 13 Assume that $K_{\beta}(\theta_0)$ is a full rank matrix. If $g = f_{x,y}$, then the infinitesimal change in the CSIF of the Wald-type test statistic is given by
\[
\frac{\partial}{\partial \varepsilon} \bar{c}_{\beta,c,x,y}(\theta_0)|_{\varepsilon=0} = -(2\beta + 1) - \frac{1}{p} \text{trace}\left(\left(\xi_\beta(\theta_0) - u_{\theta_0}(y)J_{\theta_0}^T(\theta_0)\right)\left(\xi_\beta(\theta_0) - u_{\theta_0}(y)J_{\theta_0}^T(\theta_0)\right)^T K_{\beta}^{-1}(\theta_0)\right)
\]
\[
- \frac{2}{p} \text{trace}\left(\left(f_{\theta_0}(y) - \beta u_{\theta_0}(y)u_{\theta_0}^T(y)\right) + \int I_{\theta_0}(x)f_{\theta_0}(y)dx\right) J_{\beta}^{-1}(\theta_0).
\]
(53)

Proof. Let us denote $J_{\beta,c}(\theta), K_{\beta,c}(\theta), \xi_{\beta,c}(\theta), \Sigma_{\beta,c}(\theta)$ as $J_{\beta,c,x,y}(\theta), K_{\beta,c,x,y}(\theta), \xi_{\beta,c,x,y}(\theta), \Sigma_{\beta,c,x,y}(\theta)$ respectively, when $g = f_{x,y}$. The infinitesimal change in the CSIF at the model is given by
\[
\frac{\partial}{\partial \varepsilon} \bar{c}_{\beta,c,x,y}(\theta) = \frac{1}{p} \text{trace}\left(\Sigma^{-1}_\beta(\theta) \frac{\partial}{\partial \varepsilon} \Sigma_{\beta,c,x,y}(\theta)|_{\varepsilon=0}\right).
\]
Now
\[
\frac{\partial}{\partial \varepsilon} \Sigma_{\beta,c,x,y}(\theta) = \frac{\partial}{\partial \varepsilon} J_{\beta,c,x,y}(\theta)K_{\beta,c,x,y}(\theta)J_{\beta,c,x,y}(\theta) + J_{\beta,c,x,y}(\theta) \frac{\partial}{\partial \varepsilon} K_{\beta,c,x,y}(\theta)J_{\beta,c,x,y}(\theta)
\]
\[
+ J_{\beta,c,x,y}(\theta)K_{\beta,c,x,y}(\theta) \frac{\partial}{\partial \varepsilon} J_{\beta,c,x,y}(\theta)
\]
\[
= -J_{\beta,c,x,y}(\theta) \frac{\partial}{\partial \varepsilon} J_{\beta,c,x,y}(\theta)\Sigma_{\beta,c,x,y}(\theta) + J_{\beta,c,x,y}(\theta) \frac{\partial}{\partial \varepsilon} K_{\beta,c,x,y}(\theta)J_{\beta,c,x,y}(\theta)
\]
\[
- \left(J_{\beta,c,x,y}(\theta) \frac{\partial}{\partial \varepsilon} J_{\beta,c,x,y}(\theta)\Sigma_{\beta,c,x,y}(\theta)\right)^T.
\]
(54)
where
\[
\frac{\partial}{\partial \varepsilon} J_{\beta, \varepsilon, y}(\theta) = \int (I_\theta(x) - \beta u_\theta(x)u_\theta^T(x)) (\Delta_y - f_\theta(x)) f_\theta^{\beta+1}(x) dx
\]
\[
= f_\theta^{\beta}(y) (I_\theta(y) - \beta u_\theta(y)u_\theta^T(y)) - \int (I_\theta(x) - \beta u_\theta(x)u_\theta^T(x)) f_\theta^{\beta+1}(x) dx
\]
\[
= \beta J_\beta(\theta) + f_\theta^{\beta}(y) (I_\theta(y) - \beta u_\theta(y)u_\theta^T(y)) - \int I_\theta(x)f_\theta^{\beta+1}(x) dx,
\]
(55)

and
\[
\frac{\partial}{\partial \varepsilon} K_{\beta, \varepsilon, y}(\theta) = \int u_\theta(x)u_\theta^T(x)f_\theta^{2\beta}(x) (\Delta_y - f_\theta(x)) dx - \frac{\partial}{\partial \varepsilon} \xi_{\beta, \varepsilon, y}(\theta) \xi_{\beta, \varepsilon, y}(\theta)^T - \xi_{\beta, \varepsilon, y}(\theta) \frac{\partial}{\partial \varepsilon} \xi_{\beta, \varepsilon, y}(\theta)^T
\]
\[
= u_\theta(y)u_\theta^T(y)f_\theta^{2\beta}(y) - \int u_\theta(x)u_\theta^T(x)f_\theta^{2\beta+1}(x) dx - \xi_{\beta, \varepsilon, y}(\theta) \frac{\partial}{\partial \varepsilon} \xi_{\beta, \varepsilon, y}(\theta)^T - \left( \xi_{\beta, \varepsilon, y}(\theta) \frac{\partial}{\partial \varepsilon} \xi_{\beta, \varepsilon, y}(\theta) \right)^T.
\]
(56)

Since
\[
\frac{\partial}{\partial \varepsilon} \xi_{\beta, \varepsilon, y}(\theta) = \int u_\theta(x)f_\theta^{\beta}(x) (\Delta_y - f_\theta(x)) dx = u_\theta(y)f_\theta^{\beta}(y) - \int u_\theta(x)f_\theta^{\beta+1}(x) dx
\]
\[
= u_\theta(y)f_\theta^{\beta}(y) - \xi_{\beta}(\theta),
\]
\[
\xi_{\beta, \varepsilon, y}(\theta) \frac{\partial}{\partial \varepsilon} \xi_{\beta, \varepsilon, y}(\theta)^T = \int u_\theta(x)f_\theta^{\beta}(x) ((1 - \varepsilon)f_\theta(x) + \epsilon \Delta_y) dx \left( u_\theta(y)f_\theta^{\beta}(y) - \xi_{\beta}(\theta) \right)^T,
\]
\[
\xi_{\beta, 0, y}(\theta) \frac{\partial}{\partial \varepsilon} \xi_{\beta, \varepsilon, y}(\theta) \bigg|_{\varepsilon=0} = \xi_{\beta}(\theta) \left( u_\theta(y)f_\theta^{\beta}(y) - \xi_{\beta}(\theta) \right)^T = \xi_{\beta}(\theta)u_\theta(y)f_\theta^{\beta}(y) - \xi_{\beta}(\theta)\xi_{\beta}(\theta)^T,
\]
we get from equation (56)
\[
\frac{\partial}{\partial \varepsilon} K_{\beta, \varepsilon, y}(\theta) \bigg|_{\varepsilon=0} = u_\theta(y)u_\theta^T(y)f_\theta^{2\beta}(y) - \int u_\theta(x)u_\theta^T(x)f_\theta^{2\beta+1}(x) dx
\]
\[
- \xi_{\beta, 0, y}(\theta) \frac{\partial}{\partial \varepsilon} \xi_{\beta, \varepsilon, y}(\theta) \bigg|_{\varepsilon=0} = u_\theta(y)u_\theta^T(y)f_\theta^{2\beta}(y) - K_\beta(\theta) - \xi_{\beta}(\theta)u_\theta^T(y)f_\theta^{2\beta}(y) + u_\theta(y)\xi_{\beta}(\theta)f_\theta^{\beta}(y) + \xi_{\beta}(\theta)\xi_{\beta}(\theta)^T
\]
\[
= -K_\beta(\theta) - \xi_{\beta}(\theta)u_\theta^T(y)f_\theta^{2\beta}(y) - u_\theta(y)\xi_{\beta}(\theta)f_\theta^{\beta}(y) + u_\theta(y)u_\theta^T(y)f_\theta^{2\beta}(y) + \xi_{\beta}(\theta)\xi_{\beta}(\theta)^T
\]
\[
= -K_\beta(\theta) - \left( \xi_{\beta}(\theta) - u_\theta(y)f_\theta^{\beta}(y) \right) \left( \xi_{\beta}(\theta) - u_\theta(y)f_\theta^{\beta}(y) \right)^T.
\]
(57)

Using (55) and (57), we get
\[
J_{\beta}^{-1}(\theta) \frac{\partial}{\partial \varepsilon} J_{\beta, \varepsilon, y}(\theta) \Sigma_{\beta}(\theta) = \beta \Sigma_{\beta}(\theta) + f_\theta^{\beta}(y)J_{\beta}^{-1}(\theta) (I_\theta(y) - \beta u_\theta(y)u_\theta^T(y)) \Sigma_{\beta}(\theta)
\]
\[
= -J_{\beta}^{-1}(\theta) \int I_\theta(x)f_\theta^{\beta+1}(x) dx \Sigma_{\beta}(\theta),
\]
(58)

and
\[
J_{\beta}^{-1}(\theta) \frac{\partial}{\partial \varepsilon} K_{\beta, \varepsilon, y}(\theta) \bigg|_{\varepsilon=0} J_{\beta}^{-1}(\theta) = -\Sigma_{\beta}(\theta)
\]
\[
= -J_{\beta}^{-1}(\theta) \left( \xi_{\beta}(\theta) - u_\theta(y)f_\theta^{\beta}(y) \right) \left( \xi_{\beta}(\theta) - u_\theta(y)f_\theta^{\beta}(y) \right)^T J_{\beta}^{-1}(\theta),
\]
(59)
respectively. Combining (54), (58), (59) we get
\[ \Sigma_β^{-1}(θ) \frac{∂}{∂θ} \Sigma_{β,ε,y}(θ)|_{ε=0} = -2βI_p - J_β(θ_0)K_β^{-1}(θ_0) \left( f_θ^β(y)(I_θ(y) - βu_θ(y)u_θ^T(y)) + \int I_θ(x)f_θ^β+1(x)dx \right) \Sigma_β(θ) \]
\[ - Σ_β(θ) \left( f_θ^β(y)(I_θ(y) - βu_θ(y)u_θ^T(y)) + \int I_θ(x)f_θ^β+1(x)dx \right) J_β(θ_0)K_β^{-1}(θ_0) \]
\[ - I_p - J_β(θ_0)K_β^{-1}(θ_0) \left( ξ_β(θ) - u_θ(y)f_θ^β(y) \right) \left( ξ_β(θ) - u_θ(y)f_θ^β(y) \right)^T J_β^{-1}(θ), \]
and thus the theorem follows from
\[ \text{trace} \left( Σ_β^{-1}(θ) \frac{∂}{∂θ} Σ_{β,ε,y}(θ)|_{ε=0} \right) = -(2β + 1)p - \text{trace} \left( \left( ξ_β(θ) - u_θ(y)f_θ^β(y) \right) \left( ξ_β(θ) - u_θ(y)f_θ^β(y) \right)^T K_β^{-1}(θ) \right) \]
\[ - 2\text{trace} \left( f_θ^β(y)(I_θ(y) - βu_θ(y)u_θ^T(y)) + \int I_θ(x)f_θ^β+1(x)dx \right) J_β^{-1}(θ). \]

\[ \boxed{\text{Remark 14}} \]
For the normal location-scale problem, if \( β > 0 \), then \( \frac{∂}{∂θ} \tilde{c}_{β,ε,y}(θ_0)|_{ε=0} \) given in Theorem 13 is bounded, implying the robustness of the Wald-type test statistic towards the assumption on the model.

\[ \text{Corollary 15} \]
If \( g = f_{ε,y} \) and the parameter \( θ \) is a scalar (\( p = 1 \)), then the infinitesimal change of CSIF is given by
\[ -(2β + 1) \frac{\left( ξ_β(θ_0) - u_θ(y)f_θ^β(y) \right)^2}{K_β(θ_0)} - 2f_θ^β(y)(I_θ(y) - βu_θ^2(y)) + \int I_θ(x)f_θ^β+1(x)dx \frac{J_β(θ_0)}{J_β^2(θ_0)}. \]

We shall now consider the Wald-type test statistic for the composite hypothesis and derive the infinitesimal change in the CSIF. Let us define \( Σ^*_β(θ) = M^T(θ)Σ_β(θ)M(θ) \) and \( Σ^*_β(θ_0) = M^T(θ_0)Σ_β(θ_0)M(θ_0) \). Then the following theorem is the analogous to Theorem 12 for the composite hypothesis.

\[ \text{Theorem 16} \]
Let \( \hat{θ}_{β,g} = T_β(G_n) \) be the MDPDE with tuning parameter \( β \). Then under the composite hypothesis (46), the asymptotic distribution of the Wald-type test statistic is given by
\[ W_n(\hat{θ}_{β,g}) = nm^T(\hat{θ}_{β,g})M^T(\hat{θ}_{β,g})M(\hat{θ}_{β,g})^{-1}m(\hat{θ}_{β,g}) \xrightarrow{n→∞} \sum_{i=1}^r c_{i,β,g}(θ_0)Z_i, \]
where \( \{Z_i\}_{i=1}^r \) are i.i.d. standard normal random variables and \( \{c_{i,β,g}(θ_0)\}_{i=1}^r \) the set of eigenvalues of \( Σ^*_β(θ_0)Σ^*_β(θ_0)^T \).

\[ \text{Proof.} \]
The proof of this theorem directly follows from (49) using Corollary 2.2 of Dik and de Gunst (1985).

Theorem 16 shows that the asymptotic null distribution of the Wald-type test statistic is a linear combination of \( r \) independent variables with \( χ^2_r \) densities. On the other hand, if the assumed model is correct, the asymptotic null distribution turns out to be \( χ^2_r \). So the Chi-Square Inflation Factor of the Wald-type test statistic for the composite hypothesis is defined by
\[ \bar{c}_{β,g}(θ_0) = \frac{1}{r} \sum_{i=1}^r c_{i,β,g}(θ_0) = \frac{1}{p} \text{trace} \left( Σ^*_β(θ_0)Σ^*_β(θ_0)^T \right). \]

The following theorem gives the expression for the infinitesimal change in the CSIF of the Wald-type test statistic at the model.
Theorem 17 Consider the composite null hypothesis $H_0 : m(T_\beta(G)) = 0$. If $g = f_z(y)$, then the infinitesimal change in the CSIF of the Wald-type test statistic at the model is given by

$$
\frac{\partial}{\partial \epsilon} \ell_{3,\theta}(\theta_0)|_{\epsilon=0} = -(2\beta + 1)
$$

\[ -\frac{1}{r} \text{trace} \left( \left( \xi_\beta(\theta_0) - u_{\theta_0}(y)f_{\theta_0}^\beta(y) \right) \left( \xi_\beta(\theta_0) - u_{\theta_0}(y)f_{\theta_0}^\beta(y) \right)^T \Sigma(\theta_0)M(\theta_0)\Sigma^{-1}(\theta_0)MT(\theta_0)J^{-1}(\theta_0) \right) \]

\[ -\frac{2}{r} \text{trace} \left( \left( f_{\theta_0}^\beta(y) (I_{\theta_0}(y) - \beta u_{\theta_0}(y)u_{\theta_0}^T(y)) - \int I_{\theta_0}(x)f_{\theta_0}^{\beta+1}(x)dx \right) J^{-1}(\theta_0)M(\theta_0)\Sigma^{-1}(\theta_0)M^T(\theta_0)J^{-1}(\theta_0) \right). \]

Proof. From (58) and (59) we get

$$
\Sigma^{-1}(\theta_0) = -\beta I_r
$$

\[ -\Sigma^{-1}(\theta_0)MT(\theta_0)J^{-1}(\theta_0) \left( f_{\theta_0}^\beta(y) (I_{\theta_0}(y) - \beta u_{\theta_0}(y)u_{\theta_0}^T(y)) - \int I_{\theta_0}(x)f_{\theta_0}^{\beta+1}(x)dx \right) \Sigma(\theta_0)M(\theta_0) \]

\[ -M^T(\theta_0)\Sigma(\theta_0) \left( f_{\theta_0}^\beta(y) (I_{\theta_0}(y) - \beta u_{\theta_0}(y)u_{\theta_0}^T(y)) - \int I_{\theta_0}(x)f_{\theta_0}^{\beta+1}(x)dx \right) J^{-1}(\theta_0)M(\theta_0)\Sigma^{-1}(\theta_0) \]

\[ -I_r - \Sigma^{-1}(\theta_0)M^T(\theta_0)J^{-1}(\theta_0) \left( \xi_\beta(\theta_0) - u_{\theta_0}(y)f_{\theta_0}^\beta(y) \right) \left( \xi_\beta(\theta_0) - u_{\theta_0}(y)f_{\theta_0}^\beta(y) \right)^T J^{-1}(\theta_0)M(\theta_0), \]

and thus the theorem follows from

$$
\text{trace} \left( \Sigma^{-1}(\theta_0) \frac{\partial}{\partial \epsilon} \Sigma_{3,\theta}(\theta_0) |_{\epsilon=0} = -(2\beta + 1) r \right)
$$

\[ -\text{trace} \left( \left( \xi_\beta(\theta_0) - u_{\theta_0}(y)f_{\theta_0}^\beta(y) \right) \left( \xi_\beta(\theta_0) - u_{\theta_0}(y)f_{\theta_0}^\beta(y) \right)^T \Sigma(\theta_0)M(\theta_0)\Sigma^{-1}(\theta_0)MT(\theta_0)J^{-1}(\theta_0) \right) \]

\[ -2\text{trace} \left( \left( f_{\theta_0}^\beta(y) (I_{\theta_0}(y) - \beta u_{\theta_0}(y)u_{\theta_0}^T(y)) - \int I_{\theta_0}(x)f_{\theta_0}^{\beta+1}(x)dx \right) J^{-1}(\theta_0)M(\theta_0)\Sigma^{-1}(\theta_0)M^T(\theta_0)J^{-1}(\theta_0) \right) \].

7 Examples

For the location-scale parameters of a normal model it is easy to verify the robustness properties of the Wald-type tests using the theoretical results derived in this paper. In this section we have presented two other examples, and justified the stability of the levels and powers of the Wald-type tests in presence of outliers. On the other hand, it is shown that the classical Wald tests break down as their power influence functions are unbounded.

7.1 Test for Exponentiality against Weibull Alternatives

Our first example considers an interesting problem from quality control and examine the performance of the proposed MDPDE based Wald-type test for solving it. Suppose we have $n$ independent sample observations $X_1, \ldots, X_n$ from a lifetime distribution having density $f$. We want to test the null hypothesis that the underlying lifetime (random variable) follows an exponential distribution against the alternative of Weibull distribution.
In other words, we want to test the hypothesis

\[ H_0 : f(x) = f_{\text{Exp},\sigma}(x) = \frac{1}{\sigma} e^{-\frac{x}{\sigma}}, \quad x > 0, \]

against

\[ H_1 : f(x) = f_{\text{Weib},p,\sigma}(x) = \frac{p}{\sigma} \left( \frac{x}{\sigma} \right)^{p-1} e^{-\left( \frac{x}{\sigma} \right)^p}, \quad x > 0. \]  

(61)

Here \( p > 0 \) is the shape parameter of the lifetime distribution and \( \sigma > 0 \) is the scale parameter. Further note that without loss of generality, we can assume that the data are properly scaled so that we can take \( \sigma = 1 \) (this fact can also be tested first by applying the same Wald-type test; see Section 4.2 of Basu et al. (2014)). Then, we consider the model \( F = \{ f_\theta(x) = \theta x^{\theta-1} e^{-x^\theta} : x > 0, \theta > 0 \} \) so that we have \( n \) i.i.d. observations \( X_1, \ldots, X_n \) from this family and the null hypothesis (61) simplifies to

\[ H_0 : \theta = 1 \text{ against } H_1 : \theta \neq 1. \]  

(62)

This problem is now exactly similar to the simple hypothesis testing problem considered in this paper. So we can construct a robust Wald-type test using the MDPDE \( \hat{\theta}_\beta \) of \( \theta \).

Note that the MDPDE \( \hat{\theta}_\beta \) of \( \theta \), in this particular example, is to be obtained by minimizing the objective function

\[
\frac{\theta^\beta}{(1 + \beta)^{1+\beta} - \frac{\beta}{\theta}} \Gamma \left( 1 + \beta - \frac{\beta}{\theta} \right) - \frac{\beta}{\beta n} \sum_{i=1}^{n} X_i^{\beta(\theta - 1)} e^{-\beta X_i^\theta},
\]

with respect to \( \theta > 0 \), where \( \Gamma(\cdot) \) represents the gamma function. As noted in Section 2, \( \hat{\theta}_\beta \) is \( \sqrt{n} \)-consistent and asymptotically normal. A straightforward calculation shows that, under \( H_0 : \theta_0 = 1 \), its asymptotic variance is given by \( \frac{\eta^2_\beta}{\beta n} \), where

\[
\eta_\beta = \frac{1}{1 + \beta} + [C_{2,\beta} + 2C_{1,\beta}],
\]

with

\[
C_{\alpha,\beta} = \int [(1 - y) \log(y)]^\alpha e^{-(1+\beta)y} dy.
\]

Thus, the MDPDE based Wald-type test statistics for testing the simple hypothesis (62) is given by

\[
W^0_n(\hat{\theta}_\beta) = \frac{\eta^2_\beta}{\eta^2_{2\beta}} \left( \hat{\theta}_\beta - 1 \right)^2,
\]

which asymptotically follows a chi-square distribution with one degree of freedom. Further, at the contiguous alternatives \( H_{1,n} : \theta_n = 1 + n^{-1/2}d \), this test statistic has an asymptotic non-central chi-square distribution with one degree of freedom and non-centrality parameter \( \delta = \frac{d^2 \eta^2_{2\beta}}{\eta^2_{\beta}} \). Note that, for any fixed level of significance, the asymptotic power of the Wald-type test statistic under the contiguous alternative decreases as the non-centrality parameter \( \delta \) decreases and for any fixed \( d \) it happens as \( \beta \) increases. Table 1 represents the asymptotic power for different values of \( d \) and \( \beta \). It is clear from the table that there is no significant loss in contiguous power of this test for smaller positive values of \( \beta \).

Next consider the robustness of the proposed Wald-type test as derived above. From the density of the model, it is easy to see that

\[
u_\theta(x) = \frac{1}{\theta} + (1 - x^\theta) \log x,
\]
Table 1: Asymptotic Power of the Wald-type test of (62) at 5% level of significance for different \( d \) and \( \beta \).

| \( d \) | \( \beta \) |
|----|----|
| 0  | 0 01 | 0.1 | 0.3 | 0.5 | 0.7 | 1 |
| 2  | 0.778 | 0.788 | 0.747 | 0.617 | 0.558 | 0.502 | 0.473 |
| 3  | 0.981 | 0.984 | 0.975 | 0.930 | 0.880 | 0.825 | 0.790 |
| 4  | 1.000 | 1.000 | 1.000 | 0.996 | 0.983 | 0.973 | 0.967 |
| 5  | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 0.995 |
| 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

so that the influence function of the minimum DPD functional \( T_\beta \) under the null hypothesis (62) is given by

\[
IF(x, T_\beta, F_{\theta_0}) = \frac{1}{\eta_\beta} [1 + (1 - x) \log x] e^{-\beta x}.
\]

Therefore, using the result derived in Section 3, the second order influence function of the Wald-type test statistics \( W_\beta^0 \) becomes

\[
IF_2(x, W_\beta^0, F_{\theta_0}) = \frac{2}{\eta_{2\beta}} [1 + (1 - x) \log x]^2 e^{-2\beta x}.
\]

Note that its first order influence function is always zero at the simple null. Figure 1a presents the second order influence function for several \( \beta \). The boundedness of this second order influence function is quite clear from the figure implying the robustness of the proposed Wald-type test. However, the influence function of the classical Wald test at \( \beta = 0 \) is unbounded implying its non-robustness.

Figure 1: Influence functions of MDPDE based Wald-type test of (62) for different values of \( \beta \) (solid line: \( \beta = 0 \), dotted line: \( \beta = 0.3 \), dashed-dotted line: \( \beta = 0.5 \), dashed line: \( \beta = 1 \)).

Finally, let us examine the level and power stability of the proposed Wald-type test. Following the results derived in Section 4, the level influence function of any order will be zero at the null implying the robustness of its asymptotic level. Further, the power influence function of the Wald-type test at the contiguous alternatives \( \theta_n \) is given by

\[
PIF(x, W_\beta^0, F_{\theta_n}) \approx \kappa^* \frac{\sqrt{d_{1\beta}}}{2 \sqrt{\eta_{2\beta}}} [1 + (1 - x) \log x] e^{-\beta x},
\]
where

\[ \kappa^* = \int \phi \left( \frac{\chi^2 - \alpha - y^2 \eta^2}{2d \eta / \sqrt{\eta^2}} \right) \left( \frac{\eta \beta \chi^2 - \alpha - \gamma^2}{d \eta^2} + 1 \right) dG_{\chi^2}(y), \]

with \( G_{\chi^2} \) being the distribution function of the chi-square distribution with one degree of freedom. Figure 1b shows the power influence function for some particular \( \beta \). Once again, the power robustness of the proposed test for \( \beta > 0 \) is clearly visible from the figure.

### 7.2 Test for Correlation in Bivariate Normal

Let us now consider another interesting hypothesis testing problem involving the correlation parameter of two normal populations with unknown means and variances; this problem often arises in several real-life applications when we want to check for the association between any two sets of observation only assuming the normality of those two populations. Consider the observations \( X_i = (X_{i1}, X_{i2})^T, i = 1, \ldots, n \), from the bivariate normal model \( \{N(\mu, \Sigma)\} \) where \( \mu = (\mu_1, \mu_2)^T \in \mathbb{R}^2 \) and

\[ \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \]

belongs to the set of \( 2 \times 2 \) positive definite matrices. Thus, our parameter of interest is \( \theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)^T \) with the parameter space \( \Theta = \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^+ \times [-1, 1] \). We want to test for the composite hypothesis

\[ H_0 : \rho = 0 \quad \text{against} \quad H_1 : \rho \neq 0, \quad (63) \]

with values of \( \mu_1, \mu_2, \sigma_1 \) and \( \sigma_2 \) being unspecified. In terms of notations of Section 2, we have \( r = 1 \) restrictions with \( m(\theta) = \rho \) so that \( M(\theta) \) is a \( 5 \times 1 \) matrix with the last entry 1 and rest 0 and the null parameter space is \( \Theta_0 = \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^+ \times \{0\} \). We shall now develop the Wald-type test statistic for this composite hypothesis along with its properties.

Using the form of the bivariate normal density, we can see that the MDPDE \( \hat{\theta}_\beta = (\hat{\mu}_{1, \beta}, \hat{\mu}_{2, \beta}, \hat{\sigma}_{1, \beta}, \hat{\sigma}_{2, \beta}, \hat{\rho}_\beta)^T \) of \( \theta \) with \( \beta > 0 \) is the minimizer of

\[
\frac{1}{(2\pi)^3 \sigma_1^3 \sigma_2^3 (1 - \rho^2)^{\beta/2}} \left[ \frac{1}{\sqrt{1 + \beta}} - \frac{1 + \beta}{n \beta} \sum_{i=1}^{n} e^{-\frac{1}{2} \left( x_i - \theta \right)^T \Sigma^{-1} \left( x_i - \theta \right)} \right],
\]

with respect to \( \theta \), where \( \Upsilon(x, \theta) = (x - \theta)^T \Sigma^{-1} (x - \theta) \). Take any \( \theta_0 = (\mu_1, \mu_2, \sigma_1, \sigma_2, 0)^T \in \Theta_0 \). Then the asymptotic variance of the MDPDE \( \hat{\theta}_\beta \) under \( \theta = \theta_0 \) is given by \( \Sigma_{\beta}(\theta_0) = J_\beta^{-1}(\theta_0) K_{\beta}(\theta_0) J_\beta^{-1}(\theta_0) \). A straightforward but lengthy calculation shows that

\[ J_\beta(\theta_0) = \begin{pmatrix}
\frac{C_3}{(1 + \beta)^{3/2} \sigma_1} & 0 & 0 & 0 & 0 \\
0 & \frac{C_3}{(1 + \beta)^{3/2} \sigma_2} & 0 & 0 & 0 \\
0 & 0 & \frac{(2 + \beta^2)C_3}{(1 + \beta)^{5/2} \sigma_1 \sigma_2} & \frac{\beta^2 C_3}{(1 + \beta)^{5/2}} & 0 \\
0 & 0 & \frac{(2 + \beta^2)C_3}{(1 + \beta)^{5/2} \sigma_1 \sigma_2} & \frac{(2 + \beta^2)^2 C_3}{(1 + \beta)^{7/2}} & 0 \\
0 & 0 & 0 & 0 & \frac{C_3}{(1 + \beta)^{5/2}}
\end{pmatrix} \]
and

$$K_\beta(\theta_0) = \begin{pmatrix} C_{2\beta} & 0 & 0 & 0 & 0 \\ 0 & C_{2\beta} & 0 & 0 & 0 \\ 0 & 0 & (2+3\beta^2)C_{2\beta} & 0 & 0 \\ 0 & 0 & 0 & (2+3\beta^2)C_{2\beta} & 0 \\ 0 & 0 & 0 & 0 & C_{2\beta} \end{pmatrix}$$

where $C_{2\beta} = (2\pi)^{-\beta} \sigma_1^{-\beta} \sigma_2^{-\beta}$ and $C_{2\beta}^*$ is $4C_{2\beta} - C_{2\beta}^\ast$. Hence,

$$\Sigma_\beta(\theta_0) = \begin{pmatrix} \zeta_{3/2}^2 \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \zeta_{3/2}^2 \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \zeta_{5/2}^2 \kappa_3^1 \sigma_1^2 & \zeta_{3/2}^2 \kappa_3^2 \sigma_2^2 & 0 \\ 0 & 0 & \zeta_{5/2}^2 \kappa_3^1 \sigma_1 \sigma_2 & \zeta_{5/2}^2 \kappa_3^2 \sigma_2^2 & 0 \\ 0 & 0 & 0 & 0 & \zeta_{5/2}^2 \end{pmatrix}$$

with

$$\zeta_{3/2} = \left(1 + \frac{\beta^2}{1+2\beta}\right), \quad \kappa_3^1 = \frac{(\beta^4 + 5\beta^2 + 2)}{(1+\beta^2)^2} \quad \text{and} \quad \kappa_3^2 = \frac{\beta^2(1-\beta^2)}{(1+\beta^2)^2}.$$  

Interestingly, note that whenever the null hypothesis $\rho = 0$ is true, the MDPDE of $\mu_1, \mu_2$ and $\rho$ are asymptotically independent of each other and also of the MDPDE of $\sigma_1$ and $\sigma_2$.

Now the robust Wald-type test statistics (20) for testing the null hypothesis (63) is given by

$$W_n(\hat{\theta}_\beta) = n \frac{\rho_{2\beta}}{\zeta_{5/2}}.$$  

which asymptotically follows a chi-square distribution with one degree of freedom under the null hypothesis. Note that, at $\beta = 0$, $\hat{\rho}_\beta$ coincides with the maximum likelihood estimator of $\rho$ and hence the proposed test $W_n$ coincides with the classical Wald test for the present problem. Further, under the contiguous alternatives $H_{1,n}^\ast : \rho_n = n^{-1/2}d$, the asymptotic distribution of $W_n(\hat{\theta}_\beta)$ is a non-central chi-square distribution with one degree of freedom and non-centrality parameter $\zeta_{3/2}^{-5/2} \kappa_3^{-2}$. Note that, for any fixed level of significance, the asymptotic power of the Wald-type test under the contiguous alternative hypotheses decreases as the non-centrality parameter decreases and for any fixed $d$ it happens as $\beta$ increases. However, as we can see from Table 2, the loss in contiguous power of the Wald-type test is not very significant for smaller positive values of $\beta$.

Now let us examine the robustness of this Wald-type test based on the results derived in the present paper.

Note that the influence function of the minimum DPD functional $T_\beta$ here under the null $\theta = \theta_0$ is given by

$$\mathcal{I}(x, T_\beta, F_{\theta_0}) = \begin{pmatrix} 0 \\ 0 \\ -\frac{\beta}{2} \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}\right) \\ -\frac{\beta}{2} \left(\frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2}\right) \end{pmatrix}.$$
Table 2: Asymptotic Power of the MDPDE based Wald-type test of (63) 5% level of significance for different δ and β.

| d  | 0   | 0.01 | 0.1  | 0.3  | 0.5  | 0.7  | 1   |
|----|-----|------|------|------|------|------|-----|
| 2  | 0.516 | 0.516 | 0.508 | 0.463 | 0.354 | 0.287 |
| 3  | 0.851 | 0.851 | 0.844 | 0.800 | 0.735 | 0.662 | 0.553 |
| 4  | 0.979 | 0.979 | 0.977 | 0.962 | 0.932 | 0.887 | 0.797 |
| 5  | 0.999 | 0.999 | 0.997 | 0.991 | 0.978 | 0.937 |
| 10 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Using the result derived in Section 3, the first order influence function of the Wald-type test statistic $W_\beta$ is zero at the null and its second order influence function at the null is given by

$$\text{IF}_2(x, W_\beta, F_{\theta_0}) = \frac{2(1 + 2\beta)^{5/2}}{(1 + \beta)^2 \sigma_1^2 \sigma_2^2} (x_1 - \mu_1)^2 (x_2 - \mu_2)^2 e^{-\beta \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right]}.$$ 

Clearly, this influence function is unbounded at $\beta = 0$, but whenever $\beta > 0$ it is bounded implying the robustness of the corresponding test statistics. Figure 2 shows the plot of this influence function for some particular $\beta$. It is clear from the figures that the extend of the influence function over the contamination point $x = (x_1, x_2)^T$ decreases as $\beta$ increases. This fact can also be seen by looking at the gross-error sensitivity of the test statistics given by

$$\gamma^*_{\beta} = \begin{cases} 
\frac{2n(1+2\beta)^{5/2}}{\sqrt{\beta}(1+\beta)^2} e^{-\sqrt{\beta}}, & \text{if } \beta > 0, \\
\infty, & \text{if } \beta = 0.
\end{cases}$$

Clearly $\gamma^*_{\beta}$ decreases as $\beta$ increases implying that the extent of robustness of the MDPDE based Wald-type test statistics increases.

Next, we shall consider the level and power stability of the present test. As shown in Section 5, the level influence function of any order will be zero at the null hypothesis. Hence the level of the Wald-type test, constructed using asymptotic distribution, will be robust under infinitesimal contamination. On the other hand, if we consider the contamination proportion and the difference of alternatives $\rho_n$ from null converges to zero at the same rate of $n^{-1/2} (\rho_n = n^{-1/2}d)$, the power influence function of this test is given by

$$\text{PIF}(x, W_\beta, F_{\theta_0}) \approx \frac{(1 + \beta)^{3/2} \zeta^{5/4}}{2\sigma_1 \sigma_2} \kappa^*(x_1 - \mu_1)(x_2 - \mu_2) e^{-\beta \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right]},$$

where

$$\kappa^* = \int \phi \left( \frac{x_{1,0} - y - d^2 \zeta^{5/4}}{2d\zeta^{5/4}} \right) \left( \frac{x_{1,0} - y}{d^2 \zeta^{5/4}} + 1 \right) dG_{\chi^2_1}(y),$$

with $G_{\chi^2_1}$ being the distribution function of the chi-square distribution with one degrees of freedom. Again, it is clear that the above power influence function of the MDPDE based test statistics is bounded for all $\beta > 0$ and unbounded at $\beta = 0$ (see Figure 3). This justifies the power robustness of the proposed MDPDE based Wald-types tests with $\beta > 0$ over the usual Wald test at $\beta = 0$. 

25
8 On the Choice of Tuning Parameter $\beta$

After illustrating several important properties of the proposed MDPDE based Wald-type test and its applicability, a natural question that arises from the point of view of the practitioner is what value of the tuning parameter is to be used for any particular sample data. The answer is not so obvious as the parameter $\beta$ controls a trade-off between the power of the test under pure data and its robustness under contaminated observations. For the MDPDE the role of the tuning parameter $\beta$ has been well studied in the literature, which indicates that robustness increases with $\beta$ but efficiency decreases at the same time. However, a small positive value of $\beta$ is generally recommended that provides enough robustness with a slight loss in efficiency (see Basu et al. (1998) and Basu et al. (2011)). Hong and Kim (2001) and Warwick and Jones (2005) provided some data driven choice of this tuning parameter $\beta$. In case of hypothesis testing, the asymptotic power against the contiguous alternative may regarded as a measure of efficiency of the test; and it is clear from the results and example presented in this paper that this measure decreases with increasing $\beta$ for the proposed Wald-type test. On the other hand, the robustness of this test against contamination increases as $\beta$ increases. Therefore, our suggestion in this regard is to choose an optimum value of $\beta$ using similar data driven approaches, or construct a Wald-type

![Influence function of Wald-type test statistics for testing of (63) at the null for different values of $\beta$. (Here we have taken $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$).]
Figure 3: Power influence function of Wald-type test of (62) at 5% level of significance and $d = 3$ for different values of $\beta$.

test using the same $\beta$ which is optimum for the estimation purpose.

9 Concluding Remarks

Basu et al. (2014) have proposed the Wald-type test statistics based on the minimum density power divergence estimators. They have observed strong robustness properties of the tests by using extensive simulation results. In this paper we have given proper theoretical foundations behind the robustness properties of the Wald-type test statistics. The influence function analysis is carried out to observe the effect of an infinitesimal contamination on the test statistics. To justify the stability of the level and power under a contaminated distribution we have studied the level and power influence functions. It is shown that the level influence function of a Wald-type test statistic is zero, so the level of the test remains unchanged in infinitesimal contamination. For the contiguous alternative the power influence function is bounded whenever the influence function of the MDPDE is bounded. Other than location-scale parameters for the normal model we have shown some examples where the power influence functions are bounded, and it gives the theoretical justification behind the stability of the power
function. On the other hand, the power influence functions of the classical Wald tests are unbounded, and as a result they exhibit poor power in contaminated data. We have also proposed the chi-square inflation factor to measure the robustness property with respect to the model assumption, and studied its infinitesimal change for the Wald-type test statistics. On the whole, we hope that this research establishes that the tests proposed by Basu et al. (2014) not only perform well in practise, but also have theoretically sound robustness credentials.

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*, volume 120 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2011.

A. Basu, A. Mandal, N. Martin, and L. Pardo. Density Power Divergence Tests for Composite Null Hypotheses. *ArXiv e-prints*, Mar. 2014. URL http://arxiv-web3.library.cornell.edu/abs/1403.7616.

J. J. Dik and M. C. M. de Gunst. The distribution of general quadratic forms in normal variables. *Statist. Neerlandica*, 39(1):14–26, 1985.

D. A. S. Fraser. *Nonparametric methods in statistics*. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1957.

A. Ghosh, A. Basu, and L. Pardo. On the Robustness of a Divergence based Test of Simple Statistical Hypotheses. *ArXiv e-prints*, Apr. 2014. URL http://arxiv-web3.library.cornell.edu/abs/1404.5126.

F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics: The approach based on influence functions*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.

C. Hong and Y. Kim. Automatic selection of the tuning parameter in the minimum density power divergence estimation. *J. Korean Statist. Soc.*, 30(3):453–465, 2001.

E. L. Lehmann. *Theory of point estimation*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1983.

M. Markatou and E. Ronchetti. Robust inference: the approach based on influence functions. In *Robust inference*, volume 15 of *Handbook of Statist.*, pages 49–75. North-Holland, Amsterdam, 1997.

L. Pardo. *Statistical inference based on divergence measures*, volume 185 of *Statistics: Textbooks and Monographs*. Chapman & Hall/CRC, Boca Raton, FL, 2006.

E. Ronchetti. Robustheitseigenschaften von tests, 1979. URL http://archive-ouverte.unige.ch/unige:24846. Master’s Thesis.
E. Ronchetti. Robust testing in linear models: The infinitesimal approach, 1982a. URL http://archive-ouverte.unige.ch/unige:24845. PhD Thesis.

E. Ronchetti. Robust alternatives to the $F$-test for the linear model. In Probability and statistical inference (Bad Tatzmannsdorf, 1981), pages 329–342. Reidel, Dordrecht-Boston, Mass., 1982b.

P. J. Rousseeuw and E. Ronchetti. The influence curve for tests, 1979. Research Report 21, Fachgruppe für Statistik, ETH Zürich.

P. J. Rousseeuw and E. Ronchetti. Influence curves of general statistics. J. Comput. Appl. Math., 7(3):161–166, 1981.

F. E. Satterthwaite. An approximate distribution of estimates of variance components. Biometrics Bulletin, 2(6): pp. 110–114, 1946.

R. J. Serfling. Approximation theorems of mathematical statistics. John Wiley & Sons, Inc., New York, 1980.

J. Warwick and M. Jones. Choosing a robustness tuning parameter. J. Stat. Comput. Simulation, 75(7):581–588, 2005.