PENNSTATE

Phone:(814)865-1220 / Fax:(814)863-7114

Center for Likelihood Studies
Department of Statistics
The Pennsylvania State University
422 Thomas Building
University Park, PA 16802

Robust Estimation and Tests based on Quadratic Inference Function

By Chanseok Park and Bruce G. Lindsay

Technical Report #99-02

May, 1999

Center for Likelihood Studies
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

Robust Estimation and Tests based on Quadratic Inference Function [1]

By Chanseok Park and Bruce G. Lindsay

Department of Statistics, Pennsylvania State University
University Park, PA 16802

Robust estimators and tests based on the quadratic inference function (QIF) are considered. The QIF enables one to combine a set of extended score functions efficiently. For example, one can create an adaptive estimator between the mean and median that is fully efficient at the normal model but is highly robust, with a 25% asymptotic breakdown point. In addition to providing robust point estimators and χ^2 tests of parametric hypotheses, one obtains a χ^2 goodness of fit statistic for the modeling hypotheses (for example, are the mean and the median of the distribution the same?). We consider a variety of applications of this method. These results are illustrated with a numerical study using both continuous and discrete data.

Key Words: Quadratic inference function; Maximum likelihood; M-estimation; Influence function; Breakdown point.

1 Introduction

A classical estimator such as a maximum likelihood estimator is optimal when the model is specified correctly, but it may be quite sensitive to outliers. There are several robust estimators. Among these are the median, a density-based estimator such as the minimum Hellinger distance method, and some non-parametric estimators. These methods have some deficiencies. For example, the median is quite robust but not efficient. The density-based estimators are efficient but require non-parametric kernel estimation in the continuous case. The method proposed in this paper can overcome these deficiencies. We obtain the efficiency and robustness properties by combining a set of efficient (not necessarily robust) and robust score functions.

The rest of the paper is organized as follows. In Section 2, we review the definition of the quadratic inference function (QIF) and propose the new definition. Some properties

[1] This work will be presented at the Joint Statistical Meetings (Session 93) at Baltimore Convention Center, Baltimore MD, August 9, 1999.

are also described. In Section 3, we prove the asymptotic efficiency when an efficient score is included in the QIF. We show in Section 4 that the estimator based on the QIF of the mean and median score functions has a 25% breakdown point. Section 5 gives the χ^2 tests of parametric hypotheses. Section 6 presents the results of an extensive empirical study to illustrate the performance of the new method in some discrete and continuous models.

2 Quadratic Inference Function

In this section, we introduce the *quadratic inference function* (QIF) and its properties. Hansen (1982) studied the large sample properties of a class of generalized method-of-moments (GMM) estimators and Qu (1998) proposed the QIF based on this GMM. Suppose that there is a p-dimensional parameter $\boldsymbol{\theta}$ and a k-dimensional vector of extended score functions $\mathbf{g}(\boldsymbol{\theta}; X) = (g_1(\boldsymbol{\theta}; X), \ldots, g_k(\boldsymbol{\theta}; X))^T$ satisfying $E[\mathbf{g}(\boldsymbol{\theta}, X)] = \mathbf{0}$ with $k \geq p$. Then the GMM estimator is defined as a class of estimators indexed by $\mathbf{W}$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\boldsymbol{\theta}, X_i)^T\mathbf{W}^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\boldsymbol{\theta}, X_i). \tag{1}$$

Hansen (1982) has shown that the variance for the GMM estimator is minimized when

$$\mathbf{W} = \mathrm{Var}\big(\frac{1}{\sqrt{n}}\mathbf{g}(\boldsymbol{\theta}, X_i)\big).$$

Lee (1996) interprets this as standardizing weights: the intuition for the above optimal $\mathbf{W}$ is that, in minimizing (1), it is better to standardize $\sum_{i=1}^{n}\mathbf{g}(\boldsymbol{\theta}, X_i)$, otherwise one component with a high variance can unduly dominate the minimand. Chamberlain (1987) shows that the GMM estimator with the above optimal $\mathbf{W}$ is efficient under the given condition $E[\mathbf{g}(\boldsymbol{\theta}, X)] = \mathbf{0}$ with independent and identically distributed observations. The quadratic form in (1) with optimal $\mathbf{W}$ is called the quadratic inference function. This can also be viewed as an extension of a minimum χ^2 method of generating best asymptotically normal estimates which was originally introduced by Neyman (1949) and was extensively studied by Ferguson (1958). We can rewrite (1) by ignoring the constant $1/n$ and then this gives a geometrical interpretation that we will introduce in Theorem 1. For any independent vectors, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ with $\mathbf{v}_j = (v_{j1}, \ldots, v_{jn})^T$, the QIF can be defined as

$$\mathrm{QIF}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \begin{pmatrix} \sum v_{1i} \\ \sum v_{2i} \\ \vdots \\ \sum v_{ki} \end{pmatrix}^T \begin{pmatrix} \sum v_{1i}^2 & \ldots & \sum v_{1i}v_{ki} \\ \sum v_{2i}v_{1i} & \ldots & \sum v_{2i}v_{ki} \\ \vdots & & \vdots \\ \sum v_{ki}v_{1i} & \ldots & \sum v_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum v_{1i} \\ \sum v_{2i} \\ \vdots \\ \sum v_{ki} \end{pmatrix}. \tag{2}$$

The interpretation is that the QIF is the squared length of the projection of $\mathbf{1}$ onto the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, where $\mathbf{v}_i = (g_i(\boldsymbol{\theta}; X_1), \ldots, g_i(\boldsymbol{\theta}; X_n))^T$, $i = 1, \ldots, k$.

Theorem 1. *Suppose that* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ *are any independent vectors in* $\mathbb{R}^n$ *and that* $\mathbf{P}_V$ *is the orthogonal projection matrix on* V*, the subspace spanned by* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$*. Then the quadratic inference function (QIF) is given by*

$$\mathrm{QIF}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \|\mathbf{P}_V \mathbf{1}\|^2,$$

where $\mathbf{1} = (1, 1, \ldots, 1)^T$*.*

Proof. We can show that $\mathrm{QIF}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \mathbf{1}^T \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{1}$, where $\mathbf{B} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$. The orthogonal projection matrix of V is given by $\mathbf{P}_V = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$ (see Graybill, 1983, pg. 73). Hence we have $\mathrm{QIF} = \mathbf{1}^T\mathbf{P}_V\mathbf{1}$. Since $\mathbf{P}_V$ is idempotent and symmetric, $\mathrm{QIF} = \mathbf{1}^T\mathbf{P}_V^T\mathbf{P}_V\mathbf{1} = \|\mathbf{P}_V\mathbf{1}\|^2$. □

Corollary.

(1) (Boundedness) *Let* S *be the set of independent vectors,* $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$*. Then for any* $S_0 \subset S$*,* $\mathrm{QIF}(S_0) \leq \mathrm{QIF}(S) \leq n$*.*

(2) (Invariance) *For any scalars,* $a_i (\neq 0, i = 1, \ldots, k) \in \mathbb{R}$*,* $\mathrm{QIF}(a_1\mathbf{v}_1, \ldots, a_k\mathbf{v}_k) = \mathrm{QIF}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$*.*

(3) (Decomposition) *Suppose* $V_1, \ldots, V_q$ *are an orthogonal subspaces of* $\mathbb{R}^n$ *spanned by* $S_1 = \{\mathbf{v}_{11}, \ldots, \mathbf{v}_{1k_1}\}$*, …,* $S_q = \{\mathbf{v}_{q1}, \ldots, \mathbf{v}_{qk_q}\}$*, respectively. Then QIF is decomposable, i.e.,*

$$\mathrm{QIF}(S_1, \ldots, S_q) = \sum_{i=1}^{q} \mathrm{QIF}(S_i).$$

Proof. (1) Since the vector $\mathbf{y} = \mathbf{P}_V\mathbf{1}$ is the orthogonal projection of $\mathbf{1}$ onto V, so $\mathrm{QIF} = \|\mathbf{y}\|^2 = \|\mathbf{P}_V\mathbf{1}\|^2 \leq \|\mathbf{1}\|^2 = n$ with equality only if $\mathbf{1} \in V$ (*i.e.*, $\mathbf{1} = \sum a_i\mathbf{v}_i$).
If there are two subspaces such as $W \subset V$, then $\|\mathbf{P}_W\mathbf{x}\|^2 \leq \|\mathbf{P}_V\mathbf{x}\|^2$ for any vector $\mathbf{x} \in \mathbb{R}^n$. Therefore for any $S_0 \subset S$, $\mathrm{QIF}(S_0) \leq \mathrm{QIF}(S) \leq n$.
(2) The two subspaces spanned by $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\{a_i\mathbf{v}_1, \ldots, a_k\mathbf{v}_k\}$ are equivalent for $\forall a_i(\neq 0) \in \mathbb{R}$. Hence their projection matrices are equal, and therefore the QIF's are equal.
(3) Let V be the orthogonal direct sum of the V_i, written $V = V_1 \oplus \cdots \oplus V_q = \bigoplus_{i=1}^q V_i$, then $\|\mathbf{P}_V\mathbf{x}\|^2 = \sum_{i=1}^q \|\mathbf{P}_{V_i}\mathbf{x}\|^2$ for $\forall \mathbf{x} \in \mathbb{R}^n$. Hence $\mathrm{QIF}(S_1, \ldots, S_q) = \sum_{i=1}^q \mathrm{QIF}(S_i)$. □

For example, if we use only one vector $\mathbf{v}_i$, then clearly $\mathrm{QIF}(\mathbf{v}_i) = \langle \mathbf{v}_i, \mathbf{1}\rangle^2 / \|\mathbf{v}_i\|^2$. Here, QIF satisfies the following inequality:

$$\max\Big(\frac{\langle \mathbf{v}_1, \mathbf{1}\rangle^2}{\|\mathbf{v}_1\|^2}, \ldots, \frac{\langle \mathbf{v}_k, \mathbf{1}\rangle^2}{\|\mathbf{v}_k\|^2}\Big) \leq \mathrm{QIF}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \leq n.$$

From the $\mathrm{QIF} = n$ condition (*i.e.*, $\mathbf{1} = \sum_{i=1}^k a_i\mathbf{v}_i$), we can see that $\mathrm{QIF} = n$ is always attained if $n = k$. Notice that if $n < k$, QIF does not exist.

Let $\boldsymbol{\theta}$ be a p-dimensional parameter and $\mathbf{g}(\boldsymbol{\theta}; X) = (g_1(\boldsymbol{\theta}; X), \ldots, g_k(\boldsymbol{\theta}; X))^T$ be a k-dimensional vector of extended score functions satisfying $E[\mathbf{g}(\boldsymbol{\theta}, X)] = \mathbf{0}$. Substituting $\mathbf{v}_i = \big(g_i(\boldsymbol{\theta}; X_1), \ldots, g_i(\boldsymbol{\theta}; X_n)\big)^T$ into (2) for $i = 1, \ldots, k$, we get the quadratic inference function based on the extended inference functions,

$$\mathrm{QIF}(\mathbf{g}(\boldsymbol{\theta}; \cdot); \mathbf{X}) = n\, \bar{\mathbf{g}}_n(\boldsymbol{\theta}; \mathbf{X})^T \mathbf{C}_n^{-1}(\boldsymbol{\theta}; \mathbf{X}) \bar{\mathbf{g}}_n(\boldsymbol{\theta}; \mathbf{X}), \tag{3}$$

where $\bar{\mathbf{g}}_n(\boldsymbol{\theta}; \mathbf{X}) = \frac{1}{n}\sum_{i=1}^n \mathbf{g}(\boldsymbol{\theta}; X_i)$ and $\mathbf{C}_n(\boldsymbol{\theta}; \mathbf{X}) = \frac{1}{n}\sum_{i=1}^n \mathbf{g}(\boldsymbol{\theta}; X_i)\mathbf{g}(\boldsymbol{\theta}; X_i)^T$. Since the inverse matrix does not exist when the rank of $\mathbf{C}_n$ is less than k, the QIF does not exist even though the projection on the subspace spanned by the column space of $\big(\mathbf{g}(\boldsymbol{\theta}; X_1), \ldots, \mathbf{g}(\boldsymbol{\theta}; X_n)\big)^T$ exists. Therefore, it is more useful to define the QIF using the Moore-Penrose generalized inverse and the probability measure. Notice that for each matrix $\mathbf{A}$, a Moore-Penrose generalized inverse matrix denoted by $\mathbf{A}^+$ exists and is unique. Unless specified otherwise, we will work with the following definition.

Definition 1. *The* ***quadratic inference function*** *based on score functions* $\mathbf{g}(\cdot)$ *is*

$$\begin{aligned} Q(\boldsymbol{\theta}; \mathbf{g}, F) &= \int \mathbf{g}(\boldsymbol{\theta}; \cdot)^T dF \left[\int \mathbf{g}(\boldsymbol{\theta}; \cdot)\mathbf{g}(\boldsymbol{\theta}; \cdot)^T dF\right]^+ \int \mathbf{g}(\boldsymbol{\theta}; \cdot) dF \\ &= E\mathbf{g}(\boldsymbol{\theta}; X)^T \left[E\mathbf{g}(\boldsymbol{\theta}; X)\mathbf{g}(\boldsymbol{\theta}; X)^T\right]^+ E\mathbf{g}(\boldsymbol{\theta}; X) \end{aligned} \tag{4}$$

where $\boldsymbol{\theta}$ *is a* p*-dimensional parameter and* $\mathbf{g}(\boldsymbol{\theta}; x) = (g_1(\boldsymbol{\theta}; x), \ldots, g_k(\boldsymbol{\theta}; x))^T$ *is a* k*-dimensional vector of extended score functions satisfying* $E[\mathbf{g}(\boldsymbol{\theta}, X)] = \mathbf{0}$ *at the true model parameter.*

For example, if we use the empirical measure $F_n = n^{-1}\sum \delta_{x_i}$, where δ_x stands for the point mass 1 at x, we have

$$Q(\boldsymbol{\theta}; \mathbf{g}, F_n) = \bar{\mathbf{g}}_n(\boldsymbol{\theta}; \mathbf{X})^T \mathbf{C}_n^{+}(\boldsymbol{\theta}; \mathbf{X}) \bar{\mathbf{g}}_n(\boldsymbol{\theta}; \mathbf{X}). \tag{5}$$

If $\mathbf{C}_n$ is invertible, $nQ(\boldsymbol{\theta}; \mathbf{g}, F_n)$ is equal to (3). If $X_1, \ldots, X_n$ are independent and identically distributed with the distribution $F(\cdot)$, then $Q_n(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}) = Q(\boldsymbol{\theta}; \mathbf{g}, F_n) \to Q(\boldsymbol{\theta}; \mathbf{g}, F)$ as $n \to \infty$.

Lemma 2. *Let* U *be a closed subspace of a Hilbert space* H*.*

(1) *Every* $f \in H$ *has then a unique decomposition,* $f = P(f) + P^\perp(f)$*, into a sum of* $P(f) \in U$ *and* $P^\perp(f) \in U^\perp$*.*

(2) $P(f)$ *and* $P^\perp(f)$ *are the nearest points to* f *in* U *and in* $U^\perp$*, respectively.*

(3) *The mappings* $P : H \mapsto U$ *and* $P^\perp : H \mapsto U^\perp$ *are linear.*

(4) $\|f\|^2 = \|P(f)\|^2 + \|P^\perp(f)\|^2$

Proof. See Theorem 4.11 in Rudin (1987). □

Lemma 3. *Let U be a subspace, spanned by $\{g_1, \ldots, g_k\}$, of a Hilbert space H with the inner product $\langle f, g\rangle = \int f\, g\, d\mu$, where μ is any positive measure. Then the orthogonal projection of f on U is*

$$P(f) = \int f\mathbf{g}^T d\mu\Big[\int \mathbf{g}\mathbf{g}^T d\mu\Big]^+ \mathbf{g}.$$

Proof. It is clear that $P(f)$ is in U. The projection is unique from Lemma 2. Hence, all we need to show is that $\langle P(f), f - P(f)\rangle = \langle P(f), f\rangle - \langle P(f), P(f)\rangle = 0$. Since $P(f) = P(f)^T$, we have $\langle P(f), P(f)\rangle = \langle P(f), P(f)^T\rangle$. Putting $\mathbf{a} = \int f\mathbf{g}d\mu$ and $\mathbf{C} = \int \mathbf{g}\mathbf{g}^T d\mu$, we have

$$\begin{aligned}\langle P(f), P(f)^T\rangle &= \int \mathbf{a}^T\mathbf{C}^+\mathbf{g}\mathbf{g}^T\mathbf{C}^+\mathbf{a}d\mu \\ &= \mathbf{a}^T\mathbf{C}^+(\int \mathbf{g}\mathbf{g}^T d\mu)\mathbf{C}^+\mathbf{a} = \mathbf{a}^T\mathbf{C}^+\mathbf{C}\mathbf{C}^+\mathbf{a} = \mathbf{a}^T\mathbf{C}^+\mathbf{a}, \\ \langle P(f), f\rangle &= \int \mathbf{a}^T\mathbf{C}^+\mathbf{g}f d\mu = \mathbf{a}^T\mathbf{C}^+\int f\mathbf{g}d\mu = \mathbf{a}^T\mathbf{C}^+\mathbf{a}.\end{aligned}$$

This completes the proof. □

Theorem 4. *Let $Q(\cdot)$ be the QIF defined in Definition 1.*

(1) *$Q(\boldsymbol{\theta}; \mathbf{g}, F) = \|P(1)\|^2$, where $P(1)$ be the orthogonal projection of 1 on the subspace spanned by $\{g_1, \ldots, g_k\}$.*

(2) *$Q(\boldsymbol{\theta}; \mathbf{g}_0, F) \le Q(\boldsymbol{\theta}; \mathbf{g}, F) \le 1$, where $\mathbf{g}_0(\cdot)$ is any extended score vector composed of $\mathbf{g}(\cdot)$.*

(3) *$Q(\boldsymbol{\theta}; \mathbf{g}, F) = \sum_{i=1}^q Q(\boldsymbol{\theta}; \mathbf{g}_i, F)$, where $\mathbf{g}^T = (\mathbf{g}_1^T, \ldots, \mathbf{g}_q^T)$ and any two subspaces spanned by $\mathbf{g}_i$ and $\mathbf{g}_j$ are orthogonal for $i \ne j$.*

Proof. This follows from Lemmas 2 and 3. Note that the right equality in the part (2) holds if and only if $g_1, \ldots, g_k$ are affinely dependent, that is, there exist scalars, $a_1, \ldots, a_k$ and $b(\ne 0)$ such that $\sum a_i g_i = b$. □

Lemma 5. *If extended score functions, $g_1, \ldots g_r$, are linearly independent, (i.e., $a_1 g_1 + \cdots + a_r g_r = 0$ if and only if $a_1 = \cdots = a_r = 0$), then $\int \mathbf{g}(\boldsymbol{\theta}; X)\mathbf{g}(\boldsymbol{\theta}; X)^T d\mu$ is non-singular for any positive measure μ.*

Proof. Let $\mathbf{C} = \int \mathbf{g}\mathbf{g}^T d\mu$ and $\mathbf{a} = (a_1, \ldots, a_r)^T$. If $\mathbf{C}$ is singular, then there exists non-trivial solution, $a_1, \ldots, a_r$ (not all zero), of a matrix equation $\mathbf{C}\mathbf{a} = \mathbf{0}$. It follows that $\mathbf{a}^T\mathbf{C}\mathbf{a} = 0$, so $\mathbf{a}^T(\int \mathbf{g}\mathbf{g}^T d\mu)\mathbf{a} = \int(\mathbf{a}^T\mathbf{g}\mathbf{g}^T\mathbf{a})d\mu = \int(\mathbf{a}^T\mathbf{g})^2 d\mu = 0$. Hence $\mathbf{a}^T\mathbf{g} = a_1 g_1 + \cdots + a_r g_r = 0$ for not all zero a_i. This is a contradiction, so the lemma is proved. □

Theorem 6. *Suppose that there are at most r linearly independent extended score functions in $\{g_1, \ldots, g_k\}$ and without loss of generality denote these by $\mathbf{g}^* = (g_1, \ldots, g_r)^T$. Then*

$$Q(\boldsymbol{\theta}; \mathbf{g}, F) = Q(\boldsymbol{\theta}; \mathbf{g}^*, F).$$

Proof. Writing $\mathbf{g} = \mathbf{A}\mathbf{g}^*$ where $\mathbf{A}$ is a $k \times r$ matrix of rank r, we have

$$Q(\boldsymbol{\theta}; \mathbf{g}, F) = E(\mathbf{A}\mathbf{g}^*)^T E\big[(\mathbf{A}\mathbf{g}^*)(\mathbf{A}\mathbf{g}^*)^T\big]^+ E(\mathbf{A}\mathbf{g}^*).$$

Let $\mathbf{A}$ be an $k \times r$ matrix of rank r and let $\mathbf{B}$ be an $r \times k$ matrix of rank r, then the following identities hold (see Graybill, 1983, Theorems 6.2.16 and 6.2.17),

$$(\mathbf{AB})^+ = \mathbf{B}^+\mathbf{A}^+, \qquad \mathbf{A}^+\mathbf{A} = \mathbf{I}, \qquad \mathbf{B}\mathbf{B}^+ = \mathbf{I}.$$

Using this and Lemma 5, we get $Q(\boldsymbol{\theta}; \mathbf{g}, F) = E\mathbf{g}^{*T} E\big[\mathbf{g}^*\mathbf{g}^{*T}\big]^+ E\mathbf{g}^* = Q(\boldsymbol{\theta}; \mathbf{g}^*, F)$. □

3 Asymptotic Efficiency

In this section, we introduce the QIF estimator and show how to construct an asymptotically efficient estimator with the robustness property. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the random sample and $F_n(\cdot)$ be the empirical distribution function given by $F_n(X) = \frac{1}{n}\sum_{i=1}^n I(x_i \le X)$, where $I(\cdot)$ is the indicator function. The QIF estimator based on the extended score functions $\mathbf{g}(\cdot)$ is defined as

$$\hat{\boldsymbol{\theta}}_n = \arg\inf_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \mathbf{g}, F_n) = \arg\inf_{\boldsymbol{\theta}} Q_n(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}). \tag{6}$$

Theorem 7. *Let $\mathbf{g}(\cdot)$ be a k-dimensional vector of extended score functions and $\boldsymbol{\theta}$ be a p-dimensional parameter with $k \ge p$. Let $X_1, \ldots, X_n$ be the random sample from a distribution depending on a parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Suppose that $\mathbf{g}(\boldsymbol{\theta}; x)$ includes a maximum likelihood score function and that the sample space $\mathcal{X}$ does not depend on the parameter $\boldsymbol{\theta}$. Then the QIF estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ is asymptotically efficient.*

Proof. Hansen (1982, Theorem 3.1) and Qu (1998, Theorem 3.5.1) showed that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N_p\big(\mathbf{0}, (\mathbf{d}_0^T\mathbf{C}^{-1}\mathbf{d}_0)^{-1}\big),$$

where $\mathbf{d}_0 = E[\dot{\mathbf{g}}(\boldsymbol{\theta}; X)]$ and $\mathbf{C} = E[\mathbf{g}(\boldsymbol{\theta}; X)\mathbf{g}(\boldsymbol{\theta}; X)^T]$. Hence it suffices to show that $\mathbf{d}_0^T\mathbf{C}^{-1}\mathbf{d}_0 = \mathbf{J}(\boldsymbol{\theta})$, where $\mathbf{J}(\cdot)$ is the Fisher information matrix of the maximum likelihood score functions. Let $F(x; \boldsymbol{\theta})$ be a distribution with density $f(x; \boldsymbol{\theta})$. Without loss of generality, denote

$$g_j(\boldsymbol{\theta}; x) = \frac{\partial}{\partial \theta_j} \log f(x; \boldsymbol{\theta}), \quad j = 1, \ldots, p.$$

From $E\big[\mathbf{g}(\boldsymbol{\theta}, X)\big] = \mathbf{0}$, $\int g_i(\boldsymbol{\theta}; x) f(x; \boldsymbol{\theta}) dx = 0$ for $i = 1, \ldots, k$, and differentiating with respect to θ_j for $j = 1, \ldots, p$, we obtain $\int g_i \frac{\partial f}{\partial \theta_j} dx = \int g_i g_j f dx = -\int \frac{\partial g_i}{\partial \theta_j} f dx$. Using this, we have

$$E[g_i g_j] = -E[\frac{\partial g_i}{\partial \theta_j}], \quad i = 1, \ldots, k, \quad j = 1, \ldots, p.$$

Then it is easy to show that $\mathbf{d}_0 = \mathbf{CA}$ is satisfied with a $k \times p$ matrix

$$\mathbf{A} = \begin{pmatrix} -\mathbf{I} \\ \mathbf{O} \end{pmatrix}.$$

Then we have $\mathbf{d}_0^T \mathbf{C}^{-1} \mathbf{d}_0 = \mathbf{J}(\boldsymbol{\theta})$, because

$$\mathbf{d}_0^T \mathbf{C}^{-1} \mathbf{d}_0 = \mathbf{d}_0^T \mathbf{A} = \big(-E[\frac{\partial g_i}{\partial \theta_j}] \big) = \Big(-\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x; \boldsymbol{\theta}) \Big),$$

with $i, j = 1, \ldots, p$. If $\mathbf{C}$ is singular, we can choose r independent score functions, say, $\mathbf{g}^* = (g_1^*, \ldots, g_r^*)$ satisfying $Q(\theta; \mathbf{g}, F) = Q(\theta; \mathbf{g}^*, F)$ by using Theorem 6. Then $\mathbf{C}^* = E[\mathbf{g}^*(\theta; X) \mathbf{g}^*(\theta; X)^T]$ is non-singular. This completes the proof. □

This theorem says that we can construct an asymptotically efficient but *robust* estimator by selecting the robust score functions and the maximum likelihood scores. For example, we can choose the median and the mean scores for the normal model. In the next section, we investigate the breakdown point of the estimator with these scores.

We can think of the QIF estimators as M-estimators (Huber, 1981; Hampel *et al.*, 1986) which satisfy the implicit equation

$$\sum_{i=1}^{n} \psi(X_i; t_n) = 0.$$

As estimates of the one-dimensional parameter θ, we consider real-valued statistics $t_n(x_1, \ldots, x_n) = t_n(F_n)$. Differentiating $Q(\cdot)$ with respect to θ, we have

$$\nabla Q(\theta; \mathbf{g}, F_n)\big|_{\theta = t_n} = 2 \nabla \bar{\mathbf{g}}_n^T \mathbf{C}_n^+ \bar{\mathbf{g}}_n + \bar{\mathbf{g}}_n^T \nabla \mathbf{C}_n^+ \bar{\mathbf{g}}_n = 0,$$

which implies that

$$\sum_{i=1}^{n} \big(2 \nabla \bar{\mathbf{g}}_n^T \mathbf{C}_n^+ \mathbf{g}(t_n, x_i) + \bar{\mathbf{g}}_n^T \nabla \mathbf{C}_n^+ \mathbf{g}(t_n, x_i) \big) = 0.$$

Thus, our $\psi(\cdot)$ function is

$$\psi(x_i; t_n(F_n)) = 2 \nabla \bar{\mathbf{g}}_n^T \mathbf{C}_n^+ \mathbf{g}(t_n, x_i) + \bar{\mathbf{g}}_n^T \nabla \mathbf{C}_n^+ \mathbf{g}(t_n, x_i).$$

Let $t(\cdot)$ be an estimator that is functional (*i.e.*, $t_n(F_n) = t(F_n)$ for all n and F_n), or can asymptotically be replaced by a functional. This means that we assume that there exists a function $t(\cdot)$ such that $t_n(F_n) \xrightarrow{\mathcal{P}} t(F)$ when the observations are independent and identically distributed from the true distribution F in domain (t). Without further mention, we always assume that the functionals under study are Fisher consistent (*i.e.*, $t(F_\theta) = \theta, \forall\theta \in \Theta$). Considering $E[\mathbf{g}(\theta, X)] = \mathbf{0}$ and ignoring constant 2 in $\psi(x_i; t_n(F_n))$, we obtain

$$\psi(x; t(F)) = E[\nabla\mathbf{g}(\theta, X)^T]\big(E[\mathbf{g}(\theta, X)\mathbf{g}(\theta, X)^T]\big)^+\mathbf{g}(\theta, x). \tag{7}$$

We can reach the same conclusion that was found in Theorem 7 by considering the influence function when the extended score functions include a maximum likelihood score function.

Theorem 8. *Suppose that there is a maximum likelihood score function in the extended score functions, $\mathbf{g}(\theta; x)$, and that the sample space $\mathcal{X}$ does not depend on the parameter θ. Then the influence function of the QIF estimator is the same as that of the maximum likelihood estimator. This means that the QIF estimator is asymptotically efficient.*

Proof. Let F_θ be a probability distribution whose density is denoted by f_θ. Suppose that $\mathbf{g}(\theta; x)$ has a maximum likelihood score function, say, $g_1(\theta; x) = \nabla \log f_\theta(x)$. From $E[\mathbf{g}(\theta, X)] = \mathbf{0}$, $\int g_j(\theta; x) f_\theta(x)dx = 0$ for $j = 1, \dots, k$, and differentiating with respect to θ, we obtain $\int g_j \nabla f_\theta dx = -\int \nabla g_j f_\theta dx$. Using this, we can have

$$E[g_1(\theta; X)g_j(\theta; X)] = -E[\nabla g_j(\theta; X)].$$

Let $\mathbf{C} = E[\mathbf{g}(\theta; X)\mathbf{g}(\theta; X)^T]$ and assume that $\mathbf{C}$ is non-singular. Let $\mathbf{a}^T = E[\nabla\mathbf{g}(\theta; X)^T]\mathbf{C}^{-1}$, or equivalently $\mathbf{Ca} = E[\nabla\mathbf{g}]$, where $\mathbf{a} = (a_1, \dots, a_k)^T$. It follows that

$$\begin{pmatrix} -E\nabla g_1 & -E\nabla g_2 & \dots & -E\nabla g_k \\ -E\nabla g_2 & E[g_2 g_2] & \dots & E[g_2 g_k] \\ \vdots & \vdots & & \vdots \\ -E\nabla g_k & E[g_k g_2] & \dots & E[g_k g_k] \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} E\nabla g_1 \\ E\nabla g_2 \\ \vdots \\ E\nabla g_k \end{pmatrix}.$$

Solving for $\mathbf{a}$, we have $a_1 = -1$, $a_2 = a_3 = \cdots = a_k = 0$ and the solution is unique since $\mathbf{C}$ is non-singular. Since $\psi(x; \theta) = \mathbf{a}^T\mathbf{g}(\theta; x)$, we get $\psi(x; \theta) = -g_1(\theta; x) = -\nabla \log f_\theta(x)$. Therefore, we have

$$\text{IF}(x; \psi, F_\theta) = \frac{\psi(x; \theta)}{-\int \frac{\partial}{\partial\theta}\psi(x; \theta)dF_\theta(x)} = \frac{\nabla \log f_\theta(x)}{\int \big(\nabla \log f_\theta(x)\big)^2 dF_\theta(x)}.$$

If $\mathbf{C}$ is singular, we can choose r independent score functions $\mathbf{g}^* = (g_1^*, \dots, g_r^*)$ satisfying $Q(\theta; \mathbf{g}, F) = Q(\theta; \mathbf{g}^*, F)$ by using Theorem 6. Then $\mathbf{C}^* = E[\mathbf{g}^*(\theta; X)\mathbf{g}^*(\theta; X)^T]$ is non-singular. This completes the proof. □

4 Asymptotic Breakdown Point of the QIF estimator

We introduce the definition of the breakdown point for the estimator. The breakdown point of a statistical function is roughly the smallest fraction of contamination in the data that may cause an arbitrarily extreme value in the estimate. There are some variations on the definition. One of the most appealing one is the *ε-replacement* breakdown point introduced by Donoho and Huber (1983). We consider estimators which are functionals or can asymptotically be replaced by functionals and let $t_n(\mathbf{x})$ be an estimation at the sample $\mathbf{x}$. We replace the m data points in the sample $\mathbf{x} = (x_1, \ldots, x_n)$ by the arbitrary values $x_1^*, \ldots, x_m^*$. We denote, without loss of generality, $\mathbf{x}^{(m)} = (x_1^*, \ldots, x_m^*, x_{m+1}, \ldots, x_n)$, which is called ε-corrupted sample. Before defining the breakdown point, we define the maximum bias that can be caused by ε-corruption:

$$\text{Bias}(m; t_n, \mathbf{x}) = \sup_{\mathbf{x}^{(m)}} |t_n(\mathbf{x}^{(m)}) - t_n(\mathbf{x})|, \tag{8}$$

where the supremum is taken over all possible ε-corrupted samples.

Definition 2. *The* ***finite-sample breakdown point*** ε_n^* *of the estimator* $t_n(\cdot)$ *at the sample* $\mathbf{x} = (x_1, \ldots, x_n)$ *is given by*

$$\varepsilon_n^*(t; \mathbf{x}) = \frac{1}{n} \min_m \{m : \text{Bias}(m; t, \mathbf{x}) = \infty\}. \tag{9}$$

The breakdown point usually does not depend on the sample $\mathbf{x} = (x_1, \ldots, x_n)$, and depends only slightly on the sample size n. To remove the effects of sample size, we take the limit of ε_n^* for $n \to \infty$. We call this the **asymptotic breakdown point** which is given by

$$\varepsilon^* = \lim_{n \to \infty} \varepsilon_n^*(t_n; \mathbf{x}).$$

Theorem 9. *Suppose that the QIF is based on such score functions as* $g_1(\theta; x) = x - \theta$ *and* $g_2(\theta; x) = \psi(x - \theta)$*, where* $\psi(\cdot)$ *is monotone, bounded with* $\psi(\infty) = -\psi(-\infty)$ *and* $\psi(0) = 0$ *(not necessarily skew-symmetric). Then the QIF estimator of the location model* $\{F_\theta(x) = F_0(x - \theta), \theta \in \mathbb{R}\}$ *has an asymptotic breakdown point* $\varepsilon^* = \frac{1}{4}$.

Proof. The infinite bias in equation (9) suggests that we change the values of m observations to large $M > 0$. Let us denote $\mathbf{x}^{(m)} = (x_1^*, \ldots, x_m^*, x_1, \ldots, x_{n-m})$, where $x_i^* = M$ for $i = 1, \ldots, m$ and x_i are the sample from $F_0(x)$. Let $\epsilon_n = m/n$, $\theta = M\eta$, and $A = \psi(\infty) = -\psi(-\infty)$. Then we have the following results as $M \to \infty$.

$$\frac{1}{n} \sum_{i=1}^n g_1(\theta; x_i) = (\epsilon_n - \eta)M + o(M),$$

$$\frac{1}{n} \sum_{i=1}^n g_2(\theta; x_i) = \begin{cases} A\epsilon_n + a_n + o(1) & : \quad \eta = 0 \\ A(2\epsilon_n - 1) + o(1) & : \quad \eta \in (0, 1) \\ A(\epsilon_n - 1) + o(1) & : \quad \eta = 1 \end{cases},$$

$$\frac{1}{n} \sum_{i=1}^n g_1(\theta; x_i)^2 = (\eta^2 - 2\epsilon_n \eta + \epsilon_n)M^2 + o(M^2),$$

$$\frac{1}{n}\sum_{i=1}^{n} g_1(\theta;x_i)g_2(\theta;x_i) = A(\eta - 2\epsilon_n\eta + \epsilon_n)M + o(M),$$

$$\frac{1}{n}\sum_{i=1}^{n} g_2(\theta;x_i)^2 = \begin{cases} A^2\epsilon_n + b_n + o(1) & : \quad \eta = 0 \\ A^2 + o(1) & : \quad \eta \in (0,1) \\ A^2(1-\epsilon_n) + o(1) & : \quad \eta = 1 \end{cases},$$

where $a_n = (1-\epsilon_n)\frac{1}{n-m}\sum_{i=1}^{n-m}\psi(x_i)$ and $b_n = (1-\epsilon_n)\frac{1}{n-m}\sum_{i=1}^{n-m}\psi(x_i)^2$. By substituting the above into (5), we have the result,

$$Q_n(\theta; g_1, g_2, \mathbf{x}^{(m)}) = \begin{cases} \epsilon_n + a_n^2/b_n + o(1) & : \quad \eta = 0 \\ (2\epsilon_n - 1)^2 + o(1) & : \quad \eta = \frac{1}{2} \\ 1 - \epsilon_n + o(1) & : \quad \eta = 1 \\ 1 + o(1) & : \quad \text{otherwise} \end{cases}, \tag{10}$$

as $M \to \infty$. We see that we have three local minimums and the estimator breaks down when $(2\epsilon_n - 1)^2 \le \epsilon_n + a_n^2/b_n$ or $1 - \epsilon_n \le \epsilon_n + a_n^2/b_n$. This occurs when $\epsilon_n \ge (5 - \sqrt{9 + 16a_n^2/b_n})/8$. Then the finite breakdown point is given by

$$\varepsilon_n^* = \frac{1}{n}\left\lfloor \frac{n}{8}(5 - \sqrt{9 + 16a_n^2/b_n}) \right\rfloor,$$

where $\lfloor x \rfloor$ is the smallest integer not less than x. Considering $a_n \to \int \psi(x)dx = 0$ as $n \to \infty$, we have the asymptotic breakdown point $\varepsilon^* = \frac{1}{4}$. □

5 Hypothesis Testing using the QIF

In this section, we introduce results concerning the asymptotic distribution of the quadratic inference function. We can test simple and composite parametric hypotheses and also obtain a goodness-of-fit test for the modeling hypotheses.

5.1 Test Statistics for a Simple Null Hypothesis

The following theorem gives some results about the asymptotic distribution of the QIF. These are stated and proven by Hansen (1982), Lee (1996) and Qu (1998). We can test the null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ using this theorem. Suppose that we have the sample $\mathbf{X} = (X_1, \ldots, X_n)$ with its empirical distribution $F_n(\cdot)$. Let us denote $Q_n(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}) = Q(\boldsymbol{\theta}; \mathbf{g}, F_n)$.

Theorem 10. *Let $\mathbf{g}(\cdot)$ be a k-dimensional vector of extended score functions and $\boldsymbol{\theta}$ be a p-dimensional parameter with $k \ge p$. Let $\hat{\boldsymbol{\theta}}_{\mathbf{g}}$ be the QIF estimator based on the extended score functions $\mathbf{g}(\cdot)$. Suppose that $X_1, \ldots, X_n$ are independent and identically distributed and that we are testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. Then we have the following results.*

(1) $nQ_n(\hat{\boldsymbol{\theta}}_n; \mathbf{g}, \mathbf{X}) \xrightarrow{\mathcal{D}} \chi^2_{k-p}$.

(2) $nQ_n(\boldsymbol{\theta}_0; \mathbf{g}, \mathbf{X}) \xrightarrow{\mathcal{D}} \chi^2_k$.

(3) $nQ_n(\boldsymbol{\theta}_0; \mathbf{g}, \mathbf{X}) - nQ_n(\hat{\boldsymbol{\theta}}_n; \mathbf{g}, \mathbf{X}) \xrightarrow{\mathcal{D}} \chi^2_p$.

We obtain the χ^2 test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ by using $nQ_n(\boldsymbol{\theta}_0; \mathbf{g}, \mathbf{X})$ and $nQ_n(\boldsymbol{\theta}_0; \mathbf{g}, \mathbf{X}) - nQ_n(\hat{\boldsymbol{\theta}}_n; \mathbf{g}, \mathbf{X})$. The latter is more powerful and robust if we use robust scores in the QIF. For the χ^2 goodness-of-fit test of the modeling hypothesis, we can use $nQ_n(\hat{\boldsymbol{\theta}}_n; \mathbf{g}, \mathbf{X})$. The numerical simulation results are illustrated in Section 6.

5.2 Test Statistics for a Composite Null Hypothesis

We can generalize Theorem 10 for a composite null hypothesis and for the independent *non-identically* distributed case if any linear combination of extended score functions $\{g_1(\cdot), \ldots, g_k(\cdot)\}$ satisfies the Lindeberg Condition. We assume that there are r restrictions under $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$. Suppose that $\boldsymbol{\theta}$ is known to lie in $\boldsymbol{\Theta}$, an open set in $\mathbb{R}^p$ and that it is desired to test $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, an open submanifold of $\boldsymbol{\Theta}$ of dimension $p - r$.

Lemma 11. *If $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{P})$, where $\mathbf{P}$ is an orthogonal projection matrix of rank ν, then $\mathbf{Y}^T\mathbf{Y} \sim \chi^2_\nu$.*

Theorem 12. *Let $\mathbf{g}(\cdot)$ be a k-dimensional vector of extended score functions and $\boldsymbol{\theta}$ be a p-dimensional parameter with $k \geq p$. Suppose that there are r restrictions under $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, and that $X_1, \ldots, X_n$ are independent. Then we have the following results, provided that $\mathbf{a}^T\mathbf{g}(\boldsymbol{\theta}_0)$ satisfies the Lindeberg Condition for all $\mathbf{a} \in \mathbb{R}^k$.*

(1) $\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} nQ_n(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}) \xrightarrow{\mathcal{D}} \chi^2_{k-p}$.

(2) $\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} nQ_n(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}) \xrightarrow{\mathcal{D}} \chi^2_{k-(p-r)}$.

(3) $\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} nQ_n(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}) - \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} nQ_n(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}) \xrightarrow{\mathcal{D}} \chi^2_r$.

Proof. (1) Let $\hat{\boldsymbol{\theta}} = \arg\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} nQ_n(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X})$, $\mathbf{g}_n(\boldsymbol{\theta}) = \bar{\mathbf{g}}_n(\boldsymbol{\theta}; \mathbf{X})$, and $Q_n(\boldsymbol{\theta}) = \mathbf{g}_n(\boldsymbol{\theta}; \mathbf{X})^T\mathbf{C}_n^{-1}(\boldsymbol{\theta}; \mathbf{X})\mathbf{g}_n(\boldsymbol{\theta}; \mathbf{X})$. Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ denote a true value of $\boldsymbol{\theta}$.
Using Young's form of Taylor's Theorem,

$$\sqrt{n}\,\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = \sqrt{n}\,\mathbf{g}_n(\boldsymbol{\theta}_0) + \sqrt{n}\,\dot{\mathbf{g}}_n(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{o}_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|). \tag{11}$$

Using Lindeberg-Feller and Cramer-Wold theorems, it can be shown that

$$\sqrt{n}\,\mathbf{g}_n(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N_k(\mathbf{0}, \mathbf{C}),$$

where $\mathbf{C} = E[\mathbf{g}(\boldsymbol{\theta}_0; \mathbf{X})\mathbf{g}(\boldsymbol{\theta}_0; \mathbf{X})^T]$.
By Taylor's expansion, we have

$$\sqrt{n}\,\dot{Q}_n(\hat{\boldsymbol{\theta}}) = \sqrt{n}\,\dot{Q}_n(\boldsymbol{\theta}_0) + \ddot{Q}_n(\boldsymbol{\theta}_0)\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \boldsymbol{o}_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|).$$

Since $\hat{\boldsymbol{\theta}}$ satisfies $\dot{Q}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, it follows that

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\,\ddot{Q}_n^{-1}(\boldsymbol{\theta}_0)\sqrt{n}\dot{Q}_n(\boldsymbol{\theta}_0) + \boldsymbol{o}_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|). \tag{12}$$

Hansen (1982, Lemma 3.2) and Qu (1998, Theorem 5.2.2) showed that

$$\begin{aligned} \dot{\mathbf{g}}_n(\boldsymbol{\theta}_0) &\xrightarrow{\mathcal{P}} \mathbf{d}_0, \\ \sqrt{n}\,\dot{Q}_n(\boldsymbol{\theta}_0) &= 2\mathbf{d}_0^T\mathbf{C}^{-1}\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0) + \boldsymbol{o}_p(1), \\ \ddot{Q}_n(\boldsymbol{\theta}_0) &\xrightarrow{\mathcal{P}} 2\mathbf{d}_0^T\mathbf{C}^{-1}\mathbf{d}_0, \end{aligned}$$

where $\mathbf{d}_0 = E[\dot{\mathbf{g}}(\boldsymbol{\theta}_0; X)]$.
Substituting (12) into (11), we have

$$\sqrt{n}\,\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = \big(\mathbf{I} - \mathbf{d}_0(\mathbf{d}_0^T\mathbf{C}^{-1}\mathbf{d}_0)^{-1}\mathbf{d}_0^T\mathbf{C}^{-1}\big)\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0). \tag{13}$$

Denoting $\mathbf{P}_V = \mathbf{C}^{-1/2}\mathbf{d}_0(\mathbf{d}_0^T\mathbf{C}^{-1}\mathbf{d}_0)^{-1}\mathbf{d}_0^T\mathbf{C}^{-1/2}$ and $\mathbf{Z} = \mathbf{C}^{-1/2}\sqrt{n}\,\mathbf{g}_n(\boldsymbol{\theta}_0)$, we obtain

$$\mathbf{Y} = \mathbf{C}^{-1/2}\sqrt{n}\,\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = (\mathbf{I} - \mathbf{P}_V)\mathbf{Z} + \boldsymbol{o}_p(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|).$$

$\mathbf{P}_V$ is idempotent and symmetric (*i.e.*, an orthogonal projection matrix), so is $\mathbf{I} - \mathbf{P}_V$.
Using $\mathbf{Z} \xrightarrow{\mathcal{D}} N_k(\mathbf{0}, \mathbf{I})$, and Lemma 11, we have

$$nQ_n(\hat{\boldsymbol{\theta}}) \,\dot{\sim}\, \mathbf{Y}^T\mathbf{Y} \xrightarrow{\mathcal{D}} \chi^2_{df}$$

with $df = \text{rank}(\mathbf{I} - \mathbf{P}_V)$. The rank of an idempotent matrix is equal to its trace, hence we have $df = k - p$. If $k = p$, then χ^2_{df} degenerates at zero, that is, $nQ_n(\hat{\boldsymbol{\theta}}) \xrightarrow{\mathcal{P}} 0$.
(2) If there are r restrictions under $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, the specification of $\boldsymbol{\Theta}$ can equivalently be given as a transformation, $\boldsymbol{\theta} = \mathbf{c}(\boldsymbol{\eta})$, *i.e.*,

$$\begin{aligned} \theta_1 &= c_1(\eta_1, \ldots, \eta_{p-r}), \\ &\vdots \\ \theta_p &= c_p(\eta_1, \ldots, \eta_{p-r}), \end{aligned}$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{p-r})$ lies in an open set in $\mathbb{R}^{p-r}$. Let $\boldsymbol{\eta}_0 = \mathbf{c}^{-1}(\boldsymbol{\theta}_0)$ and $\mathbf{d}_1 = E\nabla_{\boldsymbol{\eta}}\mathbf{g}(\boldsymbol{\theta}; X)\big|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0}$, then by the chain rule, $\mathbf{d}_1 = \mathbf{d}_0\dot{\mathbf{c}}_0$, where $\dot{\mathbf{c}}_0 = \dot{\mathbf{c}}(\boldsymbol{\eta}_0)$. Let $\hat{\boldsymbol{\eta}}_0 = \arg\inf_{\mathbf{c}(\boldsymbol{\eta})\in\boldsymbol{\Theta}_0} nQ_n(\mathbf{c}(\boldsymbol{\eta}))$ and $\hat{\boldsymbol{\theta}}_0 = \mathbf{c}(\hat{\boldsymbol{\eta}}_0)$. Then similar to part (1), we have

$$\mathbf{Y} = \mathbf{C}^{-1/2}\mathbf{g}_n(\hat{\boldsymbol{\theta}}_0) = (\mathbf{I} - \mathbf{P}_{V_0})\mathbf{Z} + \boldsymbol{o}_p(\|\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\|),$$

where $\mathbf{P}_{V_0} = \mathbf{C}^{-1/2}\mathbf{d}_1(\mathbf{d}_1^T\mathbf{C}^{-1}\mathbf{d}_1)^{-1}\mathbf{d}_1^T\mathbf{C}^{-1/2}$. Noticing that $\mathbf{d}_1$ is a $k \times (p - r)$ matrix, rank $= (\mathbf{I} - \mathbf{P}_{V_0}) = k - (p - r)$. Hence $nQ_n(\hat{\boldsymbol{\theta}}_0) \xrightarrow{\mathcal{D}} \chi^2_{k-(p-r)}$.
(3) Since $\mathbf{P}_V$ and $\mathbf{P}_{V_0}$ are projection matrices and $\mathbf{P}_V\mathbf{P}_{V_0} = \mathbf{P}_{V_0}\mathbf{P}_V = \mathbf{P}_{V_0}$, it follows that $\mathbf{P}_V - \mathbf{P}_{V_0}$ is also a projection matrix and that

$$\begin{aligned} nQ_n(\hat{\boldsymbol{\theta}}_0) - nQ_n(\hat{\boldsymbol{\theta}}) &\,\dot{\sim}\, \|(\mathbf{I} - \mathbf{P}_{V_0})\mathbf{Z}\|^2 - \|(\mathbf{I} - \mathbf{P}_V)\mathbf{Z}\|^2 \\ &= \mathbf{Z}^T(\mathbf{P}_V - \mathbf{P}_{V_0})\mathbf{Z} \\ &= \|(\mathbf{P}_V - \mathbf{P}_{V_0})\mathbf{Z}\|^2, \end{aligned}$$

$$\text{rank}(\mathbf{P}_V - \mathbf{P}_{V_0}) = \text{rank}(\mathbf{P}_V) - \text{rank}(\mathbf{P}_{V_0}) = p - (p - r) = r.$$

Hence, $nQ_n(\hat{\boldsymbol{\theta}}_0) - nQ_n(\hat{\boldsymbol{\theta}}) \xrightarrow{\mathcal{D}} \chi_r^2$. □

We can apply Theorem 12 to the case where we have more than one population. As an example, we discuss the two-population case, which can be easily extended to the case where there are more than two populations. Let $\mathbf{X} = (X_1, \ldots, X_{n_1})$ and $\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$ be random samples from two populations. We want to test,

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \quad \text{versus} \quad H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2,$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are p-dimensional parameters. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T \in \mathbb{R}^{2p}$, $\mathbf{W} = (\mathbf{X}, \mathbf{Y})$ and $n = n_1 + n_2$ with $n_1/n \to a \in (0, 1)$ as $n \to \infty$. Let $\mathbf{g}(\cdot)$ be a k-dimensional vector of extended score functions and define $2k$-dimensional vector of scores $\mathbf{h}(\cdot)$ to be

$$\begin{aligned} h_1(\boldsymbol{\theta}; W_i) &= g_1(\boldsymbol{\theta}_1, W_i) I(i \leq n_1), \\ &\vdots \\ h_k(\boldsymbol{\theta}; W_i) &= g_k(\boldsymbol{\theta}_1, W_i) I(i \leq n_1), \\ h_{k+1}(\boldsymbol{\theta}; W_i) &= g_1(\boldsymbol{\theta}_2, W_i) I(i > n_1), \\ &\vdots \\ h_{2k}(\boldsymbol{\theta}; W_i) &= g_k(\boldsymbol{\theta}_2, W_i) I(i > n_1). \end{aligned}$$

Then it follows that

$$nQ_n(\boldsymbol{\theta}; \mathbf{h}, \mathbf{W}) = n_1 Q_{n_1}(\boldsymbol{\theta}_1; \mathbf{g}, \mathbf{X}) + n_2 Q_{n_2}(\boldsymbol{\theta}_2; \mathbf{g}, \mathbf{Y}).$$

There are p restrictions under $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, and by Theorem 12 (3), the asymptotic distribution of the test statistic is

$$n_1 \big\{ Q_{n_1}(\hat{\boldsymbol{\theta}}_0; \mathbf{g}, \mathbf{X}) - Q_{n_1}(\hat{\boldsymbol{\theta}}_1; \mathbf{g}, \mathbf{X}) \big\} + n_2 \big\{ Q_{n_2}(\hat{\boldsymbol{\theta}}_0; \mathbf{g}, \mathbf{Y}) - Q_{n_2}(\hat{\boldsymbol{\theta}}_2; \mathbf{g}, \mathbf{Y}) \big\} \xrightarrow{\mathcal{D}} \chi_p^2, \tag{14}$$

where $\hat{\boldsymbol{\theta}}_0$, $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_0 &= \arg \inf_{\boldsymbol{\theta}} \big\{ n_1 Q_{n_1}(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}) + n_2 Q_{n_2}(\boldsymbol{\theta}; \mathbf{g}, \mathbf{Y}) \big\} \\ \hat{\boldsymbol{\theta}}_1 &= \arg \inf_{\boldsymbol{\theta}} n_1 Q_{n_1}(\boldsymbol{\theta}; \mathbf{g}, \mathbf{X}), \\ \hat{\boldsymbol{\theta}}_2 &= \arg \inf_{\boldsymbol{\theta}} n_2 Q_{n_2}(\boldsymbol{\theta}; \mathbf{g}, \mathbf{Y}). \end{aligned}$$

Note that under $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$,

$$n_1 Q_{n_1}(\hat{\boldsymbol{\theta}}_0; \mathbf{g}, \mathbf{X}) + n_2 Q_{n_2}(\hat{\boldsymbol{\theta}}_0; \mathbf{g}, \mathbf{Y}) \xrightarrow{\mathcal{D}} \chi_{2k-p}^2 \tag{15}$$

$$n_1 Q_{n_1}(\hat{\boldsymbol{\theta}}_1; \mathbf{g}, \mathbf{X}) + n_2 Q_{n_2}(\hat{\boldsymbol{\theta}}_2; \mathbf{g}, \mathbf{Y}) \xrightarrow{\mathcal{D}} \chi_{2k-2p}^2. \tag{16}$$

The formulae (14) and (15) both can be used for testing $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ versus $H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, but (14) is more powerful, robust and has better convergence properties. This will be shown in Section 6 using numerical simulation.

6 Simulations

In this section we present an extensive numerical study to assess the performance of the proposed estimator and hypothesis test. The simulations have been performed using S-Plus and C code on SUN Ultra SPARC Workstations at the Department of Statistics at the Pennsylvania State University. We compare the performance of a variety of estimators for both continuous and discrete data with and without contamination. We give the comparisons with the various hypothesis tests for the one-sample and two-sample problems. We show the estimated power of the various test statistics testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ for the normal sample.

6.1 Estimation of the location parameter

We compare the QIF estimator defined in (6) with various estimators of the location parameter such as maximum likelihood, mean, median, Hodges-Lehmann, optimal R-, Huber's M-, and Chan-He estimators. We briefly review these estimators.

The Hodges-Lehmann estimator defined as

$$\underset{i \le j}{\text{median}} \Big(\frac{X_i + X_j}{2} \Big)$$

is a compromise between the mean and the median. It is asymptotically efficient under the logistic distribution and its asymptotic relative efficiency (ARE) with respect to the mean is at least 0.864. It is also robust with asymptotic breakdown point $\varepsilon^* = 1 - \frac{1}{\sqrt{2}} \approx 0.293$, but quite poor for certain asymmetric distributions as Huber (1981, pg. 65) has indicated.

The Huber's M-estimator of location is defined as solving for θ through

$$\sum_{i=1}^{n} \psi_\kappa \Big(\frac{X_i - \theta}{S_n} \Big) = 0,$$

where $\psi_\kappa(t) = \min\{\kappa, \max(t, -\kappa)\}$ and S_n is a robust scale estimator. Here we use

$$\begin{aligned} S_n = \text{MAD}_n &= \frac{1}{\Phi^{-1}(3/4)} \underset{i}{\text{median}}\{|X_i - \underset{j}{\text{median}}(X_j)|\} \\ &\approx 1.4826 \, \underset{i}{\text{median}}\{|X_i - \underset{j}{\text{median}}(X_j)|\}. \end{aligned}$$

This estimator possesses the property of $\varepsilon^* = 0.5$.

Chan and He (1994) proposed a location estimator based on a convex linear combination of the sample mean $\bar{X}_n$ and the sample median $\tilde{X}_n$ which is given by

$$\hat{\theta} = \pi \bar{X}_n + (1 - \pi) \tilde{X}_n$$

with $0 \le \pi \le 1$. This estimator makes an adaptive combination in the sense of minimizing the asymptotic variance in the class of all linear combinations of the mean and the median.

We shall call it the Chan-He estimator. Considering the asymptotic joint distribution of the sample quantiles and the sample mean in Lin *et al.* (1980) which is given by

$$\sqrt{n}\begin{pmatrix}\bar{X}_n-\mu\\ \tilde{X}_n-\xi\end{pmatrix}\xrightarrow{\mathcal{D}}N_2\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma^2 & \frac{\eta}{2f(\xi)}\\ \frac{\eta}{2f(\xi)} & \frac{1}{4f(\xi)^2}\end{pmatrix}\right), \tag{17}$$

we have

$$\sqrt{n}\hat{\theta}\xrightarrow{\mathcal{D}}N\Big(0,\pi^2\sigma^2+\frac{(1-\pi)^2}{4f(\xi)^2}+\frac{\pi(1-\pi)\eta}{f(\xi)}\Big), \tag{18}$$

where $\mu=E(X)$, $\xi=\text{median}(X)$, $\sigma^2=E(X-\mu)^2$ and $\eta=E|X-\mu|$. The asymptotic variance of $\sqrt{n}\hat{\theta}$ attains its minimum at $\pi_0=(1-2\eta f)/(1-4\eta f+4\sigma^2f^2)$. To make π lie in the interval $[0,1]$, $\pi=\min\{1,\max(\pi_0,0)\}$ is used as an optimal value of π. The sample variance S^2 and sample mean deviation $\hat{\eta}=\frac{1}{n}\sum|X_i-\bar{X}_n|$ are used as consistent estimators for σ^2 and η. As for the estimator of $f(\xi)$, the kernel density estimator at the sample median is used. It is given by

$$\hat{f}(\tilde{X}_n)=\frac{1}{nh_n}\sum_{i=1}^{n}\phi\big(\frac{X_i-\tilde{X}_n}{h_n}\big),$$

where $\phi(\cdot)$ is the probability density of the standard normal distribution, $h_n=1.06\,n^{-1/5}\hat{\sigma}$. As for $\hat{\sigma}$, Silverman (1986, pg. 47) recommended an adaptive scale estimator,

$$\hat{\sigma}=\min(\text{standard deviation},\ \text{interquartile range}/1.35). \tag{19}$$

Then, its breakdown point is 0.25 and can get close to 0.5 if one chooses a high breakdown scale estimator such as median absolute deviation (MAD) from the sample median instead of $\hat{\sigma}$ in (19).

We include the estimators based on a generalized signed-rank norm which is defined as

$$\|\mathbf{x}\|_{\varphi^+}=\sum_{i=1}^{n}a^+(R(|x_i|))|x_i|, \tag{20}$$

where $R(|x_i|)$ is a rank of $|x_i|$ in the sample $\mathbf{x}=(x_1,\ldots,x_n)$ and $a^+(i)$ are generated as $a^+(i)=\varphi^+(i/(n+1))$ for a positive, nondecreasing, square-integrable function $\varphi^+(u)$ defined on the interval $[0,1]$. Then the estimator based on this norm is given by

$$\hat{\theta}_{\varphi^+}=\arg\inf_{\theta}\|\mathbf{X}-\theta\mathbf{1}\|_{\varphi^+}.$$

The estimator $\hat{\theta}_{\varphi^+}$ is asymptotically efficient under the symmetric location model (Hettmansperger and McKean, 1998, see pg. 45) by selecting the optimal score function given by

$$\varphi_f^+(u)=-\frac{f'\big(F^{-1}(\frac{u+1}{2})\big)}{f\big(F^{-1}(\frac{u+1}{2})\big)}.$$

This leads to the median for the double exponential distribution, the Hodges-Lehmann estimator for the logistic distribution, and the normal score estimator for the normal distribution. For the normal score estimator, we obtain an asymptotic breakdown point

$\varepsilon^* = 2\Phi(-\sqrt{\ln 4}) \approx 0.239$ and for the Cauchy case, we have $\varepsilon^* = 0.5$. We shall call this estimator the optimal R-estimator. The relationships between the estimators are as follows.

Distribution	Estimator	
	MLE	optimal R-estimator
Normal	Mean	Normal Score
Double Exponential	Median	Median
Logistic	$\psi(u) = \tanh(u/2)$	Hodges-Lehmann
Cauchy	$\psi(u) = 2u/(1+u^2)$	$\varphi^+(u) = -\sin(2\pi u)$

For the QIF estimator, we need to choose score functions. There are many possible ways of selecting them. We use the mean score and the score function based on the probability integral transform (PIT) of a normal distribution. These are given by

$$g_1(\theta; x) = x - \theta,$$
$$g_2(\theta; x) = 2\Phi\Big(\frac{x-\theta}{\tau}\Big) - 1,$$

where $\Phi(\cdot)$ is the standard normal distribution. It is worth noting that $g_2(\theta; x) \to \text{sign}(x-\theta)$ as $\tau \to 0$ while the sign$(\cdot)$ is the score function for the median. This score function can be thought as a smooth approximation of the median score. It has the advantage that the QIF based on this score is differentiable whereas the QIF based on the median score is not. It also results in stable inference. We use this QIF for all the given models, although we can obtain better estimators by choosing the maximum likelihood score according to the given model instead of simply using $g_1(\theta; x) = x - \theta$. We compare two adaptive estimators: one based on the QIF and the other based on a linear convex combination of the mean and the median. The numerical results show that the adaptive estimator of the mean and median based on the QIF performs better, especially in the contaminated case, than that based on the linear convex combination proposed by Chan and He (1994).

In Tables 1, 3, 5 and 7, we present the estimated biases and mean square errors (MSEs) of the various estimators based on the 10,000 random samples from the normal, double exponential, logistic and Cauchy distributions without contamination and with several sample sizes. The location and scale parameters are 0 and 1 respectively for these distributions. In Tables 2, 4, 6 and 8, we repeat the experiments of these distributions with $\varepsilon = 0.1$ contamination. The contamination distribution is given by

$$F_\varepsilon(x) = (1-\varepsilon)F(x) + \varepsilon\Delta_\zeta,$$

where Δ_ζ is the degenerating distribution with ζ given by 99.99% quantile of the corresponding distribution, *i.e.*, $\zeta = F^{-1}(0.9999) = 3.719, 8.517, 9.210, 3183.1$, for the normal, double exponential, logistic and Cauchy distributions respectively. When the data comes from the true model, the MLE performs better than any other estimator. But when the data is contaminated, especially in the normal model, the MLE performs quite poorly.

In the normal case, our estimator loses efficiency for small sample sizes, but it is almost as good as the MLE when the sample size is large. This supports Theorem 8. In the contaminated case, the proposed estimator beats its competitors. Even though the normal score, Hodges-Lehmann, and Chan-He estimators are robust, these are very poor when the distribution is asymmetric. They are also poor when the contamination is asymmetric.

In the double exponential case, even though our estimator loses some efficiency, it still performs better than all the other estimators except for the Chan-He estimator. It is, however, very comparable to the Chan-He estimator. When the data is contaminated, our estimator is a clear winner.

In the logistic case, all the estimators except for the median perform almost as well as the MLE. We can also see that the mean is very vulnerable to the contamination The Chan-He, Hodges-Lehmann and Huber estimators are quite poor when the contamination is asymmetric. It is important to mention that the influence function of the MLEs of a double exponential, logistic and Cauchy distributions are bounded, hence these MLEs are robust.

In the Cauchy case, our estimator performs as well as the other estimators, but in the contaminated case, our estimator is better than the median, Chan-He, Hodges-Lehmann and Huber estimators.

Table 1: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from normal $N(0,1)$ with sample size $n = 25, 50, 75, 100$.

Estimator		$n=25$		$n=50$		$n=75$		$n=100$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE (Mean)		-0.0010	0.0397	-0.0001	0.0196	0.0000	0.0133	-0.0002	0.0099
Median		-0.0011	0.0604	-0.0004	0.0300	-0.0001	0.0208	-0.0004	0.0155
Chan-He		-0.0012	0.0406	-0.0004	0.0198	0.0001	0.0135	-0.0001	0.0100
Hodges-Lehmann		-0.0004	0.0419	0.0001	0.0206	0.0002	0.0140	0.0000	0.0105
Normal Scores		-0.0008	0.0407	-0.0001	0.0198	-0.0002	0.0134	-0.0003	0.0100
Huber:	$\kappa = 1.2$	-0.0002	0.0426	-0.0002	0.0212	0.0003	0.0143	0.0000	0.0107
	$\kappa = 2.0$	-0.0009	0.0404	0.0000	0.0199	0.0001	0.0135	-0.0001	0.0101
QIF:	$\tau = 1.0$	-0.0010	0.0488	-0.0003	0.0219	-0.0004	0.0145	-0.0005	0.0105
	$\tau = 0.5$	-0.0021	0.0486	-0.0007	0.0222	-0.0004	0.0146	-0.0002	0.0106

Table 2: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from $0.9N(0,1) + 0.1\Delta_\zeta$ with sample size $n = 50, 100$.

Estimator		$n=50$		$n=100$	
		Bias	MSE	Bias	MSE
MLE (Mean)		0.3712	0.1557	0.3720	0.1473
Median		0.1381	0.0530	0.1400	0.0370
Chan-He		0.2731	0.1023	0.2721	0.0886
Hodges-Lehmann		0.2070	0.0665	0.2096	0.0558
Normal Scores		0.2738	0.0981	0.2798	0.0898
Huber:	$\kappa = 1.2$	0.1789	0.0558	0.1823	0.0452
	$\kappa = 2.0$	0.2507	0.0866	0.2562	0.0776
QIF:	$\tau = 1.0$	0.0565	0.0345	0.0544	0.0185
	$\tau = 0.5$	0.1173	0.0535	0.1132	0.0325

Table 3: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from double exponential $DE(0,1)$ with sample size $n = 25, 50, 75, 100$.

Estimator		$n = 25$		$n = 50$		$n = 75$		$n = 100$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE (Median)		-0.0023	0.0538	0.0004	0.0245	0.0004	0.0161	-0.0001	0.0119
Mean		-0.0026	0.0810	0.0017	0.0413	0.0016	0.0268	0.0010	0.0202
Chan-He		-0.0026	0.0607	0.0010	0.0286	0.0011	0.0181	0.0003	0.0133
Hodges-Lehmann		-0.0032	0.0577	0.0013	0.0284	0.0009	0.0184	0.0007	0.0139
Huber:	$\kappa = 1.2$	-0.0031	0.0575	0.0010	0.0286	0.0009	0.0188	0.0006	0.0142
	$\kappa = 2.0$	-0.0031	0.0654	0.0015	0.0329	0.0011	0.0216	0.0007	0.0162
QIF:	$\tau = 1.0$	-0.0030	0.0630	0.0015	0.0288	0.0007	0.0185	0.0007	0.0138
	$\tau = 0.5$	-0.0036	0.0619	0.0015	0.0274	0.0006	0.0172	0.0005	0.0128

Table 4: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from $0.9DE(0,1) + 0.1\Delta_\zeta$ with sample size $n = 50, 100$.

Estimator		$n = 50$		$n = 100$	
		Bias	MSE	Bias	MSE
MLE (Median)		0.1264	0.0458	0.1235	0.0298
Mean		0.8520	0.7626	0.8523	0.7446
Chan-He		0.1362	0.0510	0.1250	0.0306
Hodges-Lehmann		0.2424	0.0934	0.2429	0.0762
Huber:	$\kappa = 1.2$	0.2060	0.0765	0.2093	0.0608
	$\kappa = 2.0$	0.2879	0.1235	0.2925	0.1059
QIF:	$\tau = 1.0$	-0.0019	0.0309	-0.0028	0.0153
	$\tau = 0.5$	0.0094	0.0293	0.0064	0.0142

Table 5: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from Logistic $L(0,1)$ with sample size $n = 25, 50, 75, 100$.

Estimator		$n=25$		$n=50$		$n=75$		$n=100$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE		-0.0039	0.1219	0.0020	0.0617	0.0014	0.0405	0.0012	0.0306
Mean		-0.0035	0.1331	0.0023	0.0679	0.0020	0.0441	0.0013	0.0333
Median		-0.0041	0.1604	0.0007	0.0790	0.0006	0.0539	-0.0002	0.0408
Chan-He		-0.0035	0.1287	0.0023	0.0653	0.0018	0.0428	0.0008	0.0324
Hodges-Lehmann		-0.0044	0.1233	0.0020	0.0621	0.0012	0.0406	0.0011	0.0307
Huber:	$\kappa = 1.2$	-0.0043	0.1233	0.0015	0.0622	0.0014	0.0408	0.0009	0.0308
	$\kappa = 2.0$	-0.0038	0.1254	0.0023	0.0636	0.0017	0.0417	0.0010	0.0313
QIF:	$\tau = 1.0$	-0.0057	0.1436	0.0028	0.0673	0.0017	0.0430	0.0007	0.0321
	$\tau = 0.5$	-0.0042	0.1462	0.0035	0.0702	0.0018	0.0449	0.0007	0.0333

Table 6: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from $0.9L(0,1) + 0.1\Delta_\zeta$ with sample size $n = 50, 100$.

Estimator		$n=50$		$n=100$	
		Bias	MSE	Bias	MSE
MLE		0.3338	0.1800	0.3349	0.1463
Mean		0.9214	0.9094	0.9218	0.8796
Median		0.2236	0.1398	0.2244	0.0961
Chan-He		0.3415	0.2191	0.3179	0.1551
Hodges-Lehmann		0.3541	0.1963	0.3568	0.1625
Huber:	$\kappa = 1.2$	0.3025	0.1619	0.3071	0.1293
	$\kappa = 2.0$	0.4244	0.2554	0.4317	0.2241
QIF:	$\tau = 1.0$	0.0593	0.0820	0.0559	0.0421
	$\tau = 0.5$	0.1256	0.1138	0.1129	0.0607

Table 7: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from Cauchy $C(0,1)$ with sample size $n = 25, 50, 75, 100$.

Estimator		$n = 25$		$n = 50$		$n = 75$		$n = 100$	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE		-0.0028	0.0986	0.0011	0.0417	0.0001	0.0277	0.0006	0.0211
Mean		-0.3708	∞	0.4302	∞	0.6227	∞	-0.4291	∞
Median		-0.0034	0.1122	0.0006	0.0518	0.0005	0.0346	-0.0002	0.0259
Chan-He		-0.0038	0.1195	0.0005	0.0524	0.0004	0.0347	-0.0002	0.0259
Hodges-Lehmann		-0.0059	0.1638	0.0017	0.0743	0.0013	0.0471	0.0010	0.0350
Optimal R-		-0.0022	0.0917	0.0012	0.0426	0.0000	0.0281	0.0007	0.0213
Huber:	$\kappa = 1.2$	-0.0054	0.1525	0.0012	0.0704	0.0013	0.0460	0.0010	0.0341
	$\kappa = 2.0$	-0.0069	0.2181	0.0016	0.1000	0.0018	0.0650	0.0012	0.0479
QIF:	$\tau = 1.0$	-0.0032	0.1083	0.0019	0.0539	0.0006	0.0363	0.0009	0.0274
	$\tau = 0.5$	-0.0028	0.0992	0.0017	0.0466	0.0003	0.0311	0.0006	0.0236

Table 8: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from $0.9C(0,1) + 0.1\Delta_\zeta$ with sample size $n = 50, 100$.

Estimator		$n = 50$		$n = 100$	
		Bias	MSE	Bias	MSE
MLE		0.0011	0.0475	0.0009	0.0230
Mean		318.6	∞	318.9	∞
Median		0.1831	0.0954	0.1805	0.0629
Chan-He		0.1831	0.0954	0.1805	0.0629
Hodges-Lehmann		0.4026	0.2706	0.3951	0.2063
Optimal R-		0.0411	0.0517	0.0380	0.0256
Huber:	$\kappa = 1.2$	0.3278	0.2016	0.3303	0.1549
	$\kappa = 2.0$	0.4758	0.3689	0.4799	0.2999
QIF:	$\tau = 1.0$	-0.0014	0.0682	0.0002	0.0334
	$\tau = 0.5$	-0.0007	0.0571	0.0003	0.0276

6.2 Estimation of the Poisson and binomial parameters

In Tables 9 – 12, we present the biases and MSEs of the maximum likelihood estimator (MLE), the minimum Hellinger distance estimator (MHDE) and the QIF estimator for the Poisson and binomial distributions with and without contamination. We suggest a new robust score function for discrete data in the following Lemma.

Lemma 13. *Let X be a discrete random sample with distribution function $F_\theta(\cdot)$ and its probability mass function $f_\theta(\cdot)$. Then the following is a score function given by*

$$\psi(x,\theta) = 2F_\theta(x) - f_\theta(x) - 1.$$

Proof. It suffices to show $E\psi(X,\theta) = 0$. We obtain the following and this completes the proof.

$$\begin{aligned} E\psi(X,\theta) &= \sum_{\forall x}\sum_{y\leq x} 2f_\theta(x)f_\theta(y) - \sum_{\forall x} f_\theta(x)^2 - 1 \\ &= \sum_{\forall y}\sum_{x\geq y} 2f_\theta(x)f_\theta(y) - \sum_{\forall x} f_\theta(x)^2 - 1 = \sum_{\forall y} f_\theta(y)\Big(2\sum_{x\geq y} f_\theta(y) - f_\theta(y) - 1\Big) \\ &= \sum_{\forall y} f_\theta(y)\Big(- 2F_\theta(y) + f_\theta(y) + 1\Big) = -\sum_{\forall y} f_\theta(y)\psi(y,\theta) = -E\psi(X,\theta). \end{aligned}$$

□

This score is approximately uniform. We shall call it the PIT score of a discrete distribution. In the simulation, we compare three estimators: the mean, the MHDE and the QIF estimators. The research on the MHDE for the counting data can be found in Simpson (1987). All three estimators performed comparably for the Poisson and binomial models without contamination. In the contaminated case, the MHDE and our estimator perform similarly.

Table 9: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from Poisson Poi(5) with sample size $n = 25, 50, 75, 100$.

Estimator	$n = 25$		$n = 50$		$n = 75$		$n = 100$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE (Mean)	0.0020	0.1996	0.0003	0.1013	−0.0017	0.0654	−0.0017	0.0499
MHDE	−0.1488	0.2485	−0.1032	0.1220	−0.0831	0.0772	−0.0696	0.0577
QIF	−0.1081	0.2674	−0.0636	0.1203	−0.0463	0.0740	−0.0352	0.0549

Table 10: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from $0.9\text{Poi}(5) + 0.1\Delta_{\zeta=15}$ with sample size $n = 50, 100$.

Estimator	$n = 50$		$n = 100$	
	Bias	MSE	Bias	MSE
MLE (Mean)	0.9093	0.9106	0.9075	0.8648
MHDE	−0.0283	0.1284	0.0077	0.0610
QIF	0.0816	0.1360	0.0735	0.0681

Table 11: Estimated Biases (×10) and MSEs (×10^2) of the estimators under consideration. 10,000 random samples were drawn from Binomial Bin(10, 0.25) with sample size $n = 25, 50, 75, 100$.

Estimator	$n = 25$		$n = 50$		$n = 75$		$n = 100$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
MLE	−0.0000	0.0769	−0.0011	0.0376	−0.0011	0.0253	−0.0010	0.0186
MHDE	−0.0691	0.0927	−0.0494	0.0440	−0.0385	0.0282	−0.0316	0.0204
QIF	−0.0546	0.1097	−0.0411	0.0467	−0.0283	0.0291	−0.0220	0.0208

Table 12: Estimated Biases (×10) and MSEs (×10^2) of the estimators under consideration. 10,000 random samples were drawn from $0.9\text{Bin}(10, 0.25) + 0.1\Delta_{\zeta=8}$ with sample size $n = 50, 100$.

Estimator	$n = 50$		$n = 100$	
	Bias	MSE	Bias	MSE
MLE	0.4990	0.2801	0.4991	0.2644
MHDE	0.0194	0.0500	0.0385	0.0248
QIF	0.0473	0.0504	0.0433	0.0256

6.3 Hypothesis test

We provide a numerical study for simple and two-population hypothesis tests and compute the power of the tests with and without contamination. We compare our results with a variety of hypothesis tests given in Tables 13 and 14. In the QIF, we use the mean and PIT scores as we did in the estimation of the location parameter.

In Tables 15 − 22, we present the results from the hypothesis tests for $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ in the normal, double exponential, logistic and Cauchy distributions with and without contamination. The contamination level is a fixed 10% and its location is given by the 99.99% quantile of the corresponding distribution. In the tables, QIF(1), QIF(2) and QIF(3) imply the test statistics in parts (1), (2) and (3) of Theorem 10, respectively. QIF(2) and QIF(3) can be used for testing $H_0 : \theta = 0$. But QIF(2) converges slower, is not robust and is less powerful. QIF(1) gives the test for $H_0 : E[\mathbf{g}(\hat{\boldsymbol{\theta}})] = \mathbf{0}$. This statistic can be used for testing a χ^2 goodness-of-fit statistic for the modeling hypothesis, *i.e.*, the equality of the mean and the median. We find that QIF(1) increases dramatically when the data is contaminated for all the given distributions. When the data is asymmetrically contaminated, the mean and median should be different for the symmetric location model. The results show how effectively QIF(1) detects the model fitness. QIF(2) also increases when the data is contaminated but QIF(3) does not. The reason is that the QIF moves up about ϵ when the contamination level is ϵ as equation (10) suggests, but the minimizer of the QIF will not change much in the contaminated case. In QIF(3) statistic, there are two QIFs which cancel each other out. Hence QIF(3) does not increase as much compared to QIF(1) and QIF(2). This statistic satisfies the requirement of a robust test which should not be perturbed very much in the presence of small contamination. The numerical results show that QIF(3) is as close to the nominal levels ($\alpha = 0.05, 0.01$) as other tests in the normal model and even better than other tests in other models in the uncontaminated case. In the contaminated case, QIF(3) is the best under all the distributions. The test statistics and their rejection regions are summarized in Table 13.

Table 13: Tests of hypotheses for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ with sample size n.

	Test Statistics	Rejection Region
t-test	$t(\theta) = \dfrac{\bar{X} - \theta}{S}$	$\lvert t(\theta_0)\rvert \geq t_{\alpha/2}(n-1)$ $\lvert t(\theta_0)\rvert \geq z_{\alpha/2}$ (asymptotic)
Sign	$S^+(\theta) = \sum_{i=1}^{n} I(X_i - \theta > 0)$	$\left\lvert S^+(\theta_0) - \frac{n}{2}\right\rvert \geq z_{\alpha/2}\sqrt{\frac{n}{4}} + \frac{1}{2}$
Wilcoxon	$T^+(\theta) = \sum\sum_{i \leq j} I\big(\frac{X_i + X_j}{2} > \theta\big)$	$\left\lvert T^+(\theta_0) - \frac{n(n+1)}{4}\right\rvert \geq z_{\alpha/2}\sqrt{\frac{n(n+1)(2n+1)}{24}} + \frac{1}{2}$

In Tables 15 − 22, we present the results for the hypothesis tests of two samples for

$H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$ in the normal, double exponential, logistic and Cauchy distributions with and without contamination. All the models have 0 for the location and 1 for the scale parameter. In the contaminated case, only the second samples are contaminated. The contamination level is a fixed 10% and its location is given by the 99.99% quantile of the corresponding distribution. The test statistics and their rejection regions are summarized in Table 14. The interpretation of QIF(1), QIF(2) and QIF(3) is very similar to what it was in the one-sample case. QIF(1) based on the test statistic in (16) implies the combined measure of how different the mean and median are. QIF(2) based on (15) and QIF(3) on (14) give the test statistics for testing $H_0 : \theta_1 = \theta_2$. We obtain the similar results to those we obtained in the one-sample case.

Table 14: Tests of hypotheses for $H_0 : \theta_2 - \theta_1 = \Delta$ versus $H_1 : \theta_2 - \theta_1 \neq \Delta$ with sample size n_1 and n_2.

	Test Statistics	Rejection Region
t-test	$t(\Delta) = \dfrac{\bar{Y} - \bar{X} - \Delta}{S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$\lvert t(\Delta)\rvert \geq t_{\alpha/2}(n_1 + n_2 - 2)$ $\lvert t(\Delta)\rvert \geq z_{\alpha/2}$ (asymptotic)
Wilcoxon	$T^+(\Delta) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} I\big(Y_j - X_i > \Delta\big)$	$\big\lvert T^+(\Delta) - \frac{n_1 n_2}{2}\big\rvert \geq z_{\alpha/2}\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} + \frac{1}{2}$

Insensitivity of the observed level to the presence of outliers does not by itself define a robust test. To be convinced, one has to look at the power under contamination as well. We estimated the power of the tests with and without contamination for sample size 25 and $10,000$ samples. The estimated power is given in Figures 1 and 2 for the normal model and in Figures 3 and 4 for the double exponential model. For the uncontaminated normal model, the proposed QIF test is almost as powerful as the t-test which is a uniformly most powerful unbiased test in a normal model. However, when the data is contaminated, the t-test becomes much less powerful and it is biased (less than significance level) over some regions. In Figure 2, we compare the power with the power of the t-test in the uncontaminated case. This clearly shows that the power curve of the QIF test is very close to that of the asymptotic t-test in the uncontaminated case. For the double exponential model, the QIF test is more powerful than any other test even in the uncontaminated case. In the contaminated case, the power curve of the QIF test is the closest to that of the asymptotic t-test in the uncontaminated case.

Table 15: Levels of the one-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn from $N(0,1)$ with sample size $n = 25, 50, 100$.

Test Statistics	$n = 25$		$n = 50$		$n = 100$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test	0.0497	0.0112	0.0503	0.0089	0.0479	0.0105
Sign	0.0459	0.0038	0.0316	0.0066	0.0359	0.0064
Wilcoxon	0.0488	0.0076	0.0486	0.0082	0.0482	0.0096
QIF(1):$\tau = 1.0$	0.0127	0.0002	0.0357	0.0025	0.0438	0.0080
$\tau = 0.5$	0.0187	0.0005	0.0392	0.0038	0.0445	0.0089
QIF(2):$\tau = 1.0$	0.0333	0.0044	0.0390	0.0054	0.0453	0.0086
$\tau = 0.5$	0.0392	0.0055	0.0432	0.0063	0.0481	0.0105
QIF(3):$\tau = 1.0$	0.0612	0.0095	0.0538	0.0090	0.0506	0.0094
$\tau = 0.5$	0.0681	0.0125	0.0558	0.0101	0.0524	0.0108

Table 16: Levels of the one-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn from $0.9N(0,1) + 0.1\Delta_\zeta$ with sample size $n = 50, 100$.

Test Statistics	$n = 50$		$n = 100$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test	0.3694	0.1062	0.7916	0.4497
Sign	0.0669	0.0192	0.1213	0.0364
Wilcoxon	0.1679	0.0518	0.3542	0.1363
QIF(1):$\tau = 1.0$	0.6924	0.0570	0.9933	0.8790
$\tau = 0.5$	0.4927	0.0798	0.9369	0.6506
QIF(2):$\tau = 1.0$	0.3654	0.0452	0.9826	0.6654
$\tau = 0.5$	0.3393	0.0507	0.9238	0.5260
QIF(3):$\tau = 1.0$	0.0679	0.0149	0.0773	0.0187
$\tau = 0.5$	0.1003	0.0233	0.1447	0.0439

Table 17: Levels of the one-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn from $DE(0, 1)$ with sample size $n = 25, 50, 100$.

Test Statistics	$n = 25$		$n = 50$		$n = 100$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.0589	0.0123	0.0546	0.0110	0.0530	0.0105
Sign	0.0432	0.0042	0.0315	0.0069	0.0356	0.0065
Wilcoxon	0.0491	0.0070	0.0457	0.0101	0.0506	0.0082
QIF(1):$\tau = 1.0$	0.0125	0.0002	0.0304	0.0012	0.0438	0.0055
$\tau = 0.5$	0.0187	0.0002	0.0358	0.0023	0.0477	0.0066
QIF(2):$\tau = 1.0$	0.0306	0.0034	0.0361	0.0047	0.0468	0.0059
$\tau = 0.5$	0.0324	0.0045	0.0416	0.0044	0.0489	0.0078
QIF(3):$\tau = 1.0$	0.0551	0.0079	0.0504	0.0093	0.0514	0.0091
$\tau = 0.5$	0.0582	0.0102	0.0527	0.0098	0.0533	0.0087

Table 18: Levels of the one-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn from $0.9DE(0, 1) + 0.1\Delta_\zeta$ with sample size $n = 50, 100$.

Test Statistics	$n = 50$		$n = 100$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.5879	0.1665	0.9751	0.7663
Sign	0.0709	0.0200	0.1272	0.0390
Wilcoxon	0.1773	0.0532	0.3539	0.1456
QIF(1):$\tau = 1.0$	0.8026	0.0383	0.9995	0.9598
$\tau = 0.5$	0.7206	0.0626	0.9988	0.9162
QIF(2):$\tau = 1.0$	0.3816	0.0454	0.9912	0.7291
$\tau = 0.5$	0.3742	0.0469	0.9785	0.6526
QIF(3):$\tau = 1.0$	0.0518	0.0078	0.0496	0.0082
$\tau = 0.5$	0.0558	0.0104	0.0522	0.0089

Table 19: Levels of the one-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn from $L(0,1)$ with sample size $n = 25, 50, 100$.

Test Statistics	$n = 25$		$n = 50$		$n = 100$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.0618	0.0140	0.0552	0.0121	0.0536	0.0111
Sign	0.0432	0.0042	0.0315	0.0069	0.0356	0.0065
Wilcoxon	0.0491	0.0070	0.0457	0.0101	0.0506	0.0082
QIF(1):$\tau = 1.0$	0.0189	0.0005	0.0346	0.0029	0.0467	0.0073
$\tau = 0.5$	0.0178	0.0011	0.0314	0.0030	0.0443	0.0077
QIF(2):$\tau = 1.0$	0.0316	0.0045	0.0415	0.0046	0.0489	0.0075
$\tau = 0.5$	0.0352	0.0050	0.0435	0.0058	0.0484	0.0091
QIF(3):$\tau = 1.0$	0.0594	0.0098	0.0523	0.0106	0.0534	0.0087
$\tau = 0.5$	0.0643	0.0110	0.0571	0.0125	0.0564	0.0103

Table 20: Levels of the one-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn from $0.9L(0,1) + 0.1\Delta_\zeta$ with sample size $n = 50, 100$.

Test Statistics	$n = 50$		$n = 100$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.5168	0.1600	0.9374	0.6656
Sign	0.0709	0.0200	0.1272	0.0390
Wilcoxon	0.1773	0.0532	0.3539	0.1456
QIF(1):$\tau = 1.0$	0.7135	0.0605	0.9977	0.9031
$\tau = 0.5$	0.5984	0.0751	0.9867	0.7962
QIF(2):$\tau = 1.0$	0.3767	0.0463	0.9815	0.6616
$\tau = 0.5$	0.3600	0.0495	0.9576	0.5905
QIF(3):$\tau = 1.0$	0.0590	0.0109	0.0631	0.0122
$\tau = 0.5$	0.0738	0.0144	0.0896	0.0212

Table 21: Levels of the one-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn from $C(0,1)$ with sample size $n = 25, 50, 100$.

Test Statistics	$n = 25$		$n = 50$		$n = 100$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.0270	0.0034	0.0228	0.0020	0.0239	0.0019
Sign	0.0432	0.0042	0.0315	0.0069	0.0356	0.0065
Wilcoxon	0.0491	0.0070	0.0457	0.0101	0.0506	0.0082
QIF(1):$\tau = 1.0$	0.0073	0.0002	0.0122	0.0002	0.0188	0.0007
$\tau = 0.5$	0.0098	0.0003	0.0145	0.0001	0.0198	0.0010
QIF(2):$\tau = 1.0$	0.0278	0.0033	0.0303	0.0041	0.0350	0.0044
$\tau = 0.5$	0.0274	0.0042	0.0316	0.0034	0.0348	0.0044
QIF(3):$\tau = 1.0$	0.0511	0.0076	0.0514	0.0099	0.0500	0.0091
$\tau = 0.5$	0.0501	0.0086	0.0492	0.0093	0.0516	0.0100

Table 22: Levels of the one-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn from $0.9C(0,1) + 0.1\Delta_\zeta$ with sample size $n = 50, 100$.

Test Statistics	$n = 50$		$n = 100$	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.9909	0.0015	0.9952	0.9911
Sign	0.0709	0.0200	0.1272	0.0390
Wilcoxon	0.1773	0.0532	0.3539	0.1456
QIF(1):$\tau = 1.0$	0.9897	0.0000	0.9949	0.9904
$\tau = 0.5$	0.9897	0.0000	0.9949	0.9904
QIF(2):$\tau = 1.0$	0.3310	0.0424	0.9934	0.9781
$\tau = 0.5$	0.3315	0.0427	0.9932	0.9782
QIF(3):$\tau = 1.0$	0.0525	0.0094	0.0498	0.0092
$\tau = 0.5$	0.0534	0.0098	0.0510	0.0095

Table 23: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50$ and $n_2 = 50, 100$. Both populations are $N(0, 1)$.

Test Statistics	$n_1 = 25, n_2 = 50$		$n_1 = 50, n_2 = 50$		$n_1 = 50, n_2 = 100$		$n_1 = 100, n_2 = 100$	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test	0.0547	0.0110	0.0497	0.0097	0.0511	0.0111	0.0504	0.0114
MWW test	0.0524	0.0102	0.0489	0.0097	0.0486	0.0104	0.0529	0.0105
QIF(1):$\tau = 1.0$	0.0151	0.0009	0.0300	0.0014	0.0344	0.0038	0.0404	0.0057
$\tau = 0.5$	0.0203	0.0015	0.0334	0.0025	0.0363	0.0051	0.0406	0.0074
QIF(2):$\tau = 1.0$	0.0272	0.0036	0.0365	0.0044	0.0387	0.0054	0.0440	0.0063
$\tau = 0.5$	0.0340	0.0043	0.0420	0.0056	0.0419	0.0066	0.0466	0.0083
QIF(3):$\tau = 1.0$	0.0599	0.0106	0.0580	0.0096	0.0542	0.0098	0.0517	0.0102
$\tau = 0.5$	0.0645	0.0129	0.0619	0.0122	0.0584	0.0126	0.0532	0.0118

Table 24: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50$ and $n_2 = 50, 100$. Population 1 is $N(0, 1)$. Population 2 is $0.9N(0, 1) + 0.1\Delta_\zeta$.

Test Statistics	$n_1 = 25, n_2 = 50$		$n_1 = 50, n_2 = 50$		$n_1 = 50, n_2 = 100$		$n_1 = 100, n_2 = 100$	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test	0.1378	0.0267	0.2615	0.0796	0.3292	0.0993	0.5661	0.2706
MWW test	0.0914	0.0207	0.1199	0.0338	0.1539	0.0459	0.2181	0.0737
QIF(1):$\tau = 1.0$	0.3369	0.0164	0.3454	0.0313	0.9616	0.6147	0.9597	0.6171
$\tau = 0.5$	0.2610	0.0240	0.2759	0.0368	0.8123	0.4163	0.8226	0.4190
QIF(2):$\tau = 1.0$	0.2317	0.0279	0.2463	0.0337	0.9029	0.4261	0.9147	0.4333
$\tau = 0.5$	0.2163	0.0309	0.2313	0.0360	0.7593	0.3264	0.7860	0.3442
QIF(3):$\tau = 1.0$	0.0640	0.0138	0.0632	0.0156	0.0703	0.0147	0.0690	0.0173
$\tau = 0.5$	0.0860	0.0194	0.0876	0.0221	0.1026	0.0275	0.1160	0.0317

Table 25: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50$ and $n_2 = 50, 100$. Both populations are $DE(0, 1)$.

Test Statistics	$n_1 = 25, n_2 = 50$		$n_1 = 50, n_2 = 50$		$n_1 = 50, n_2 = 100$		$n_1 = 100, n_2 = 100$	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.0541	0.0116	0.0570	0.0106	0.0519	0.0093	0.0534	0.0099
MWW test	0.0521	0.0079	0.0527	0.0094	0.0476	0.0103	0.0495	0.0103
QIF(1):$\tau = 1.0$	0.0172	0.0007	0.0274	0.0015	0.0321	0.0031	0.0380	0.0040
$\tau = 0.5$	0.0237	0.0013	0.0319	0.0032	0.0356	0.0052	0.0417	0.0062
QIF(2):$\tau = 1.0$	0.0301	0.0028	0.0334	0.0045	0.0367	0.0040	0.0438	0.0056
$\tau = 0.5$	0.0333	0.0037	0.0381	0.0052	0.0399	0.0052	0.0444	0.0071
QIF(3):$\tau = 1.0$	0.0523	0.0100	0.0513	0.0103	0.0522	0.0093	0.0507	0.0106
$\tau = 0.5$	0.0554	0.0109	0.0530	0.0112	0.0509	0.0108	0.0503	0.0104

Table 26: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50$ and $n_2 = 50, 100$. Population 1 is $DE(0, 1)$. Population 2 is $0.9DE(0, 1) + 0.1\Delta_\zeta$.

Test Statistics	$n_1 = 25, n_2 = 50$		$n_1 = 50, n_2 = 50$		$n_1 = 50, n_2 = 100$		$n_1 = 100, n_2 = 100$	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.1624	0.0193	0.4461	0.1337	0.5058	0.1450	0.8624	0.5444
MWW test	0.0890	0.0206	0.1260	0.0323	0.1527	0.0445	0.2137	0.0738
QIF(1):$\tau = 1.0$	0.3996	0.0163	0.3840	0.0325	0.9922	0.7205	0.9925	0.7340
$\tau = 0.5$	0.3677	0.0242	0.3634	0.0383	0.9774	0.6383	0.9804	0.6556
QIF(2):$\tau = 1.0$	0.2538	0.0266	0.2561	0.0352	0.9551	0.4664	0.9579	0.4678
$\tau = 0.5$	0.2539	0.0302	0.2524	0.0395	0.9131	0.4262	0.9194	0.4300
QIF(3):$\tau = 1.0$	0.0521	0.0085	0.0496	0.0108	0.0473	0.0089	0.0488	0.0092
$\tau = 0.5$	0.0570	0.0088	0.0511	0.0091	0.0475	0.0100	0.0521	0.0098

Table 27: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50$ and $n_2 = 50, 100$. Both populations are $L(0,1)$.

Test Statistics	$n_1 = 25, n_2 = 50$		$n_1 = 50, n_2 = 50$		$n_1 = 50, n_2 = 100$		$n_1 = 100, n_2 = 100$	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.0528	0.0110	0.0575	0.0113	0.0503	0.0095	0.0533	0.0103
MWW test	0.0521	0.0079	0.0527	0.0094	0.0476	0.0103	0.0495	0.0103
QIF(1):$\tau = 1.0$	0.0226	0.0015	0.0320	0.0028	0.0346	0.0034	0.0404	0.0064
$\tau = 0.5$	0.0206	0.0014	0.0283	0.0023	0.0334	0.0032	0.0411	0.0060
QIF(2):$\tau = 1.0$	0.0330	0.0031	0.0384	0.0046	0.0379	0.0050	0.0443	0.0070
$\tau = 0.5$	0.0326	0.0033	0.0389	0.0052	0.0378	0.0050	0.0454	0.0070
QIF(3):$\tau = 1.0$	0.0596	0.0103	0.0552	0.0115	0.0532	0.0107	0.0526	0.0107
$\tau = 0.5$	0.0653	0.0116	0.0619	0.0139	0.0558	0.0108	0.0550	0.0108

Table 28: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50$ and $n_2 = 50, 100$. Population 1 is $L(0,1)$. Population 2 is $0.9L(0,1) + 0.1\Delta_\zeta$.

Test Statistics	$n_1 = 25, n_2 = 50$		$n_1 = 50, n_2 = 50$		$n_1 = 50, n_2 = 100$		$n_1 = 100, n_2 = 100$	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.1578	0.0254	0.3810	0.1192	0.4374	0.1350	0.7664	0.4383
MWW test	0.0890	0.0206	0.1260	0.0323	0.1527	0.0445	0.2137	0.0738
QIF(1):$\tau = 1.0$	0.3580	0.0238	0.3508	0.0363	0.9721	0.6287	0.9746	0.6411
$\tau = 0.5$	0.2987	0.0262	0.3147	0.0371	0.9158	0.5138	0.9302	0.5257
QIF(2):$\tau = 1.0$	0.2499	0.0296	0.2497	0.0391	0.9094	0.4261	0.9196	0.4316
$\tau = 0.5$	0.2310	0.0289	0.2434	0.0381	0.8417	0.3731	0.8554	0.3848
QIF(3):$\tau = 1.0$	0.0595	0.0089	0.0572	0.0116	0.0589	0.0113	0.0569	0.0108
$\tau = 0.5$	0.0702	0.0124	0.0680	0.0138	0.0712	0.0157	0.0776	0.0167

Table 29: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. 10,000 random samples were drawn with sample size $n_1 = 25, 50$ and $n_2 = 50, 100$. Both populations are $C(0,1)$.

Test Statistics	$n_1 = 25, n_2 = 50$		$n_1 = 50, n_2 = 50$		$n_1 = 50, n_2 = 100$		$n_1 = 100, n_2 = 100$	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.0314	0.0030	0.0247	0.0035	0.0268	0.0030	0.0240	0.0017
MWW test	0.0521	0.0079	0.0527	0.0094	0.0476	0.0103	0.0495	0.0103
QIF(1):$\tau = 1.0$	0.0080	0.0000	0.0093	0.0007	0.0132	0.0003	0.0138	0.0009
$\tau = 0.5$	0.0097	0.0001	0.0111	0.0008	0.0145	0.0004	0.0144	0.0009
QIF(2):$\tau = 1.0$	0.0228	0.0024	0.0232	0.0030	0.0265	0.0031	0.0262	0.0029
$\tau = 0.5$	0.0238	0.0023	0.0231	0.0027	0.0269	0.0033	0.0265	0.0030
QIF(3):$\tau = 1.0$	0.0479	0.0086	0.0512	0.0090	0.0509	0.0090	0.0509	0.0102
$\tau = 0.5$	0.0462	0.0093	0.0482	0.0088	0.0486	0.0095	0.0494	0.0100

Table 30: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.05, 0.01$. 10,000 random samples were drawn with sample size $n_1 = 25, 50$ and $n_2 = 50, 100$. Population 1 is $C(0,1)$. Population 2 is $0.9C(0,1) + 0.1\Delta_\zeta$.

Test Statistics	$n_1 = 25, n_2 = 50$		$n_1 = 50, n_2 = 50$		$n_1 = 50, n_2 = 100$		$n_1 = 100, n_2 = 100$	
	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.05$	$\alpha=0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
t-test (asymp.)	0.0047	0.0000	0.9795	0.0023	0.9812	0.0067	0.9897	0.9825
MWW test	0.0890	0.0206	0.1260	0.0323	0.1527	0.0445	0.2137	0.0738
QIF(1):$\tau = 1.0$	0.4357	0.0040	0.4248	0.0088	0.9942	0.9810	0.9939	0.9833
$\tau = 0.5$	0.4257	0.0061	0.4198	0.0103	0.9941	0.9808	0.9939	0.9829
QIF(2):$\tau = 1.0$	0.2254	0.0217	0.2364	0.0265	0.9911	0.5647	0.9927	0.5573
$\tau = 0.5$	0.2239	0.0216	0.2365	0.0261	0.9911	0.5617	0.9929	0.5553
QIF(3):$\tau = 1.0$	0.0484	0.0087	0.0527	0.0103	0.0455	0.0085	0.0481	0.0095
$\tau = 0.5$	0.0469	0.0084	0.0528	0.0088	0.0453	0.0081	0.0497	0.0095

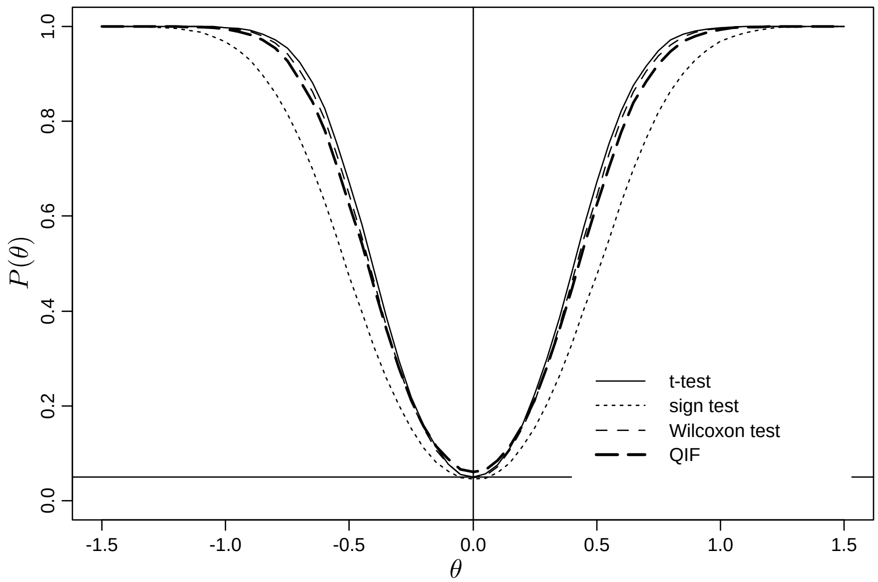

Figure 1: Estimated powers for the tests under consideration testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ with level $\alpha = 0.05$. $10,000$ random samples were drawn from $N(\theta, 1)$ with sample size $n = 25$.

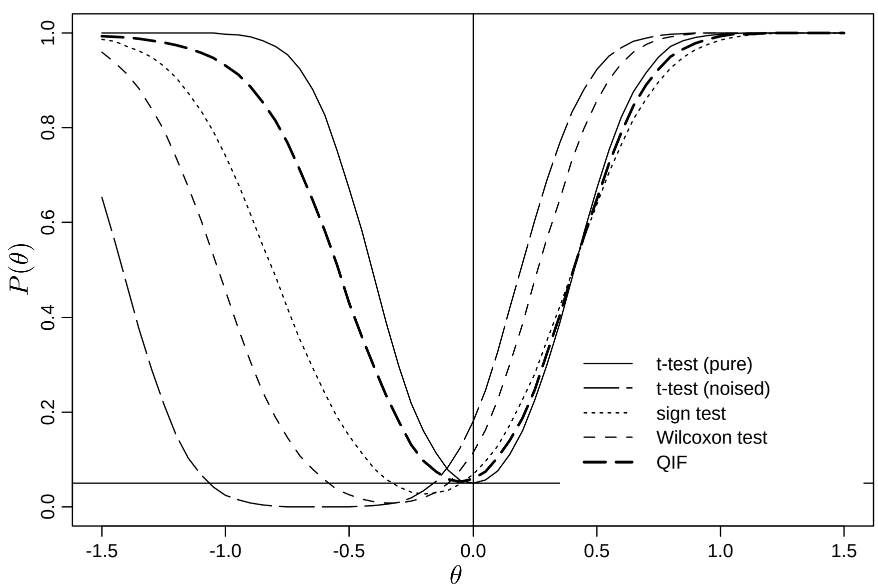

Figure 2: Estimated powers for the tests under consideration testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ with level $\alpha = 0.05$. $10,000$ random samples were drawn from $0.9N(\theta, 1) + 0.1\Delta_\zeta$ with sample size $n = 25$.

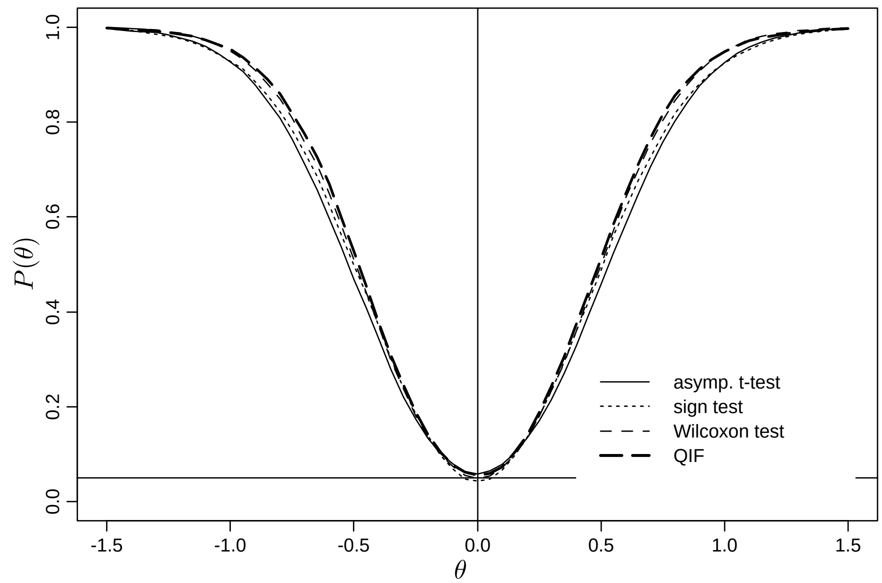

Figure 3: Estimated powers for the tests under consideration testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ with level $\alpha = 0.05$. $10,000$ random samples were drawn from $DE(\theta, 1)$ with sample size $n = 25$.

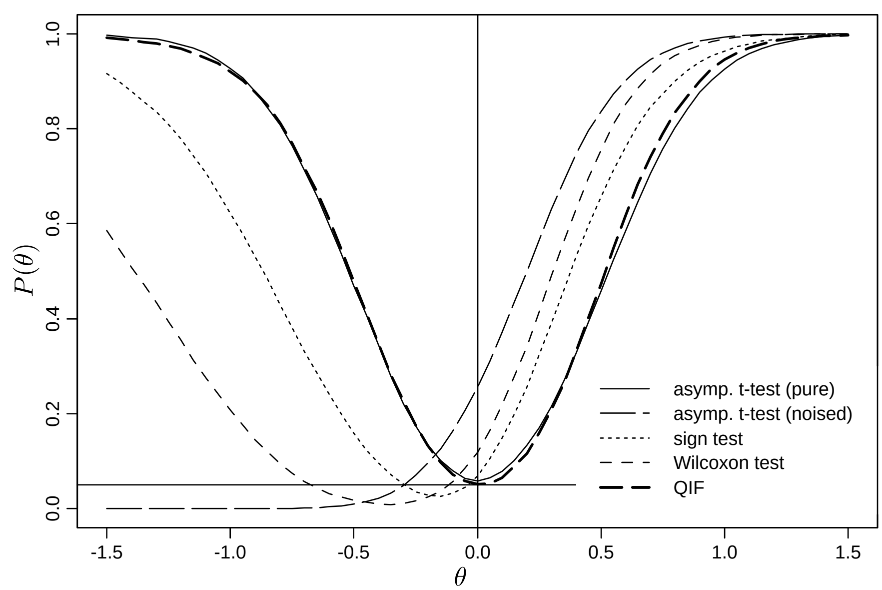

Figure 4: Estimated powers for the tests under consideration testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ with level $\alpha = 0.05$. $10,000$ random samples were drawn from $0.9DE(\theta, 1)+0.1\Delta_\zeta$ with sample size $n = 25$.

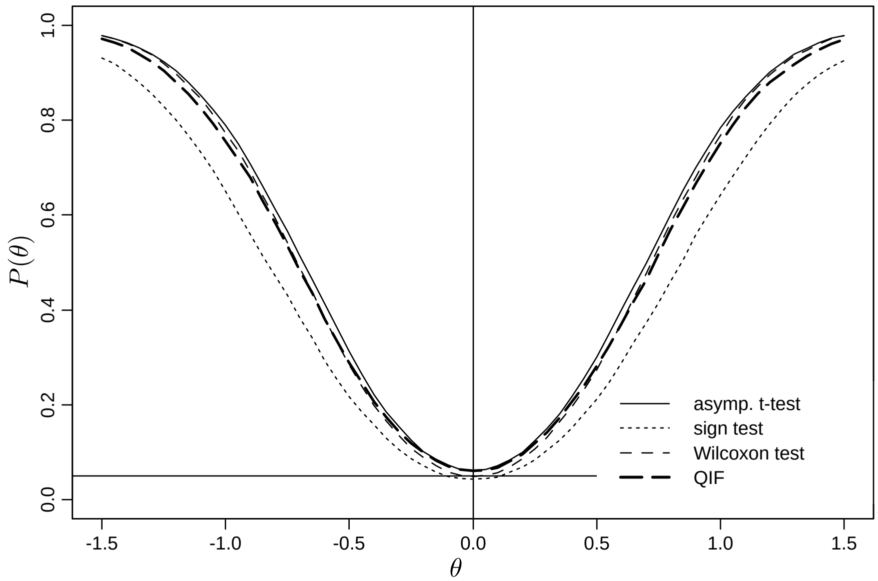

Figure 5: Estimated powers for the tests under consideration testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ with level $\alpha = 0.05$. $10,000$ random samples were drawn from $L(\theta, 1)$ with sample size $n = 25$.

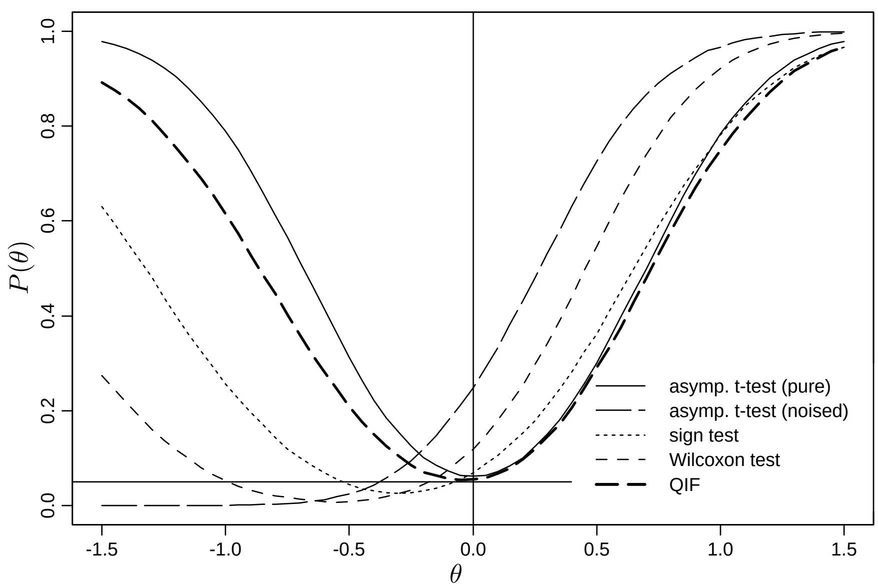

Figure 6: Estimated powers for the tests under consideration testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ with level $\alpha = 0.05$. $10,000$ random samples were drawn from $0.9L(\theta, 1) + 0.1\Delta_\zeta$ with sample size $n = 25$.

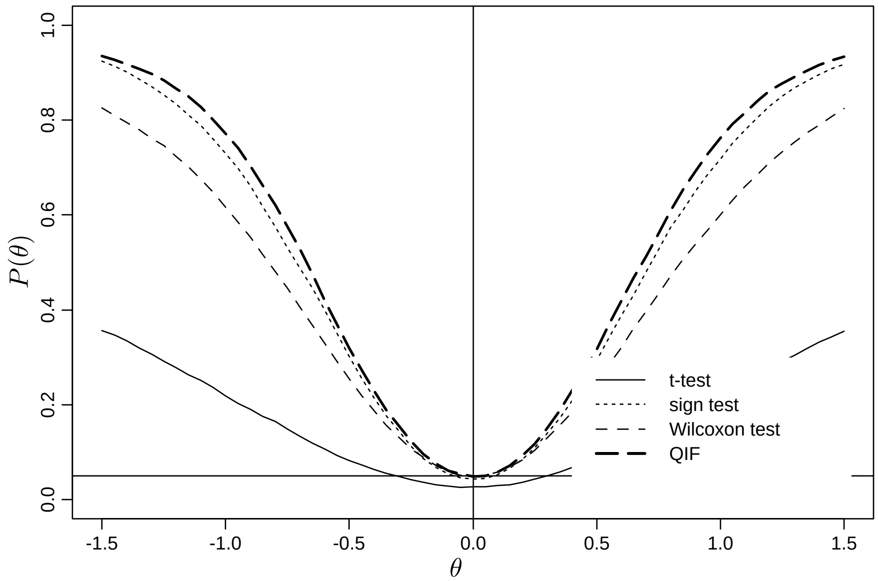

Figure 7: Estimated powers for the tests under consideration testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ with level $\alpha = 0.05$. $10,000$ random samples were drawn from $C(\theta, 1)$ with sample size $n = 25$.

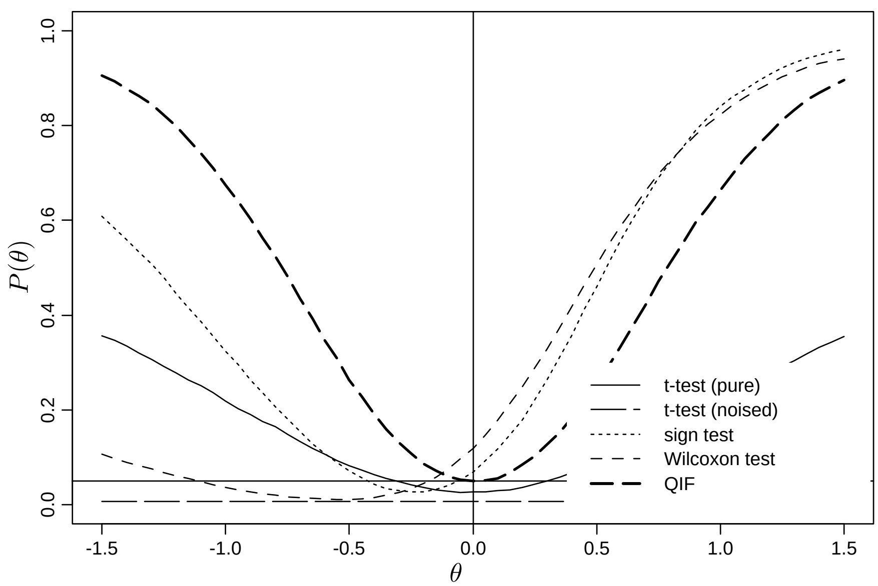

Figure 8: Estimated powers for the tests under consideration testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$ with level $\alpha = 0.05$. $10,000$ random samples were drawn from $0.9C(\theta, 1) + 0.1\Delta_\zeta$ with sample size $n = 25$.

Appendix

A.1 Differential notation on multivariate calculus

We present the differential notation on multivariate calculus. The derivative of $f : \mathbb{R}^p \mapsto \mathbb{R}$, with respective to $\boldsymbol{\theta}$ is,

$$\dot{f}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \Big(\frac{\partial f(\boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_p}\Big),$$

and its second derivative is

$$\ddot{f}(\boldsymbol{\theta}) = \nabla^2_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \dot{f}(\boldsymbol{\theta})^T = \begin{pmatrix} \frac{\partial^2}{\partial \theta_1^2} f(\boldsymbol{\theta}) & \cdots & \frac{\partial^2}{\partial \theta_1 \partial \theta_p} f(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \frac{\partial^2}{\partial \theta_p \partial \theta_1} f(\boldsymbol{\theta}) & \cdots & \frac{\partial^2}{\partial \theta_p^2} f(\boldsymbol{\theta}) \end{pmatrix}.$$

The derivative of $\mathbf{g} : \mathbb{R}^p \mapsto \mathbb{R}^p$, with respective to $\boldsymbol{\theta}$ is,

$$\dot{\mathbf{g}}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \mathbf{g}(\boldsymbol{\theta}) = \begin{pmatrix} \dot{g}_1(\boldsymbol{\theta}) \\ \vdots \\ \dot{g}_p(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \cdots & \frac{\partial g_1}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial g_p}{\partial \theta_1} & \cdots & \frac{\partial g_p}{\partial \theta_p} \end{pmatrix}.$$

A.2 Differential of the matrix

Theorem 14. *If X is an $m \times n$ matrix and X^+ is its Moore-Penrose inverse, then*

$$dX^+ = -X^+(dX)X^+ + (I - X^+X)(dX^T){X^+}^T X^+ + X^+{X^+}^T(dX^T)(I - XX^+).$$

References

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, **34**, 305–334.

Chan, Y. M. and He, X. (1994). A simple and competitive estimator of location. *Statistics & Probability Letters*, **19**, 137–142.

Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. In P. Bickel, K. Doksum, and J. Hodges Jr., editors, *A Festschrift for Erich L. Lehmann*, pages 157–184. Wadsworth, Belmont, CA.

Ferguson, T. S. (1958). A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics*, **29**, 1046–1062.

Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Chapman & Hall.

Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Wadsworth, Inc.

Hampel, F. R., Ronchetti, E., Rousseeuw, P. J., and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, New York.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**(4), 1029–1054.

Hettmansperger, T. P. and McKean, J. W. (1998). *Robust Nonparametric Statistical Methods*. John Wiley & Sons Inc., New York.

Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons, New York.

Lee, M.-J. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*. Springer-Verlag, New York.

Lin, P. E., Wu, K. T., and Ahmad, I. A. (1980). Asymptotic joint distribution of sample quantiles and sample mean with applications. *Communications in Statistics: Theory and Methods*, **9**, 51–60.

Neyman, J. (1949). Contribution to the theory of the χ^2 test. In *Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press.

Qu, P. (1998). *Adaptive Generalized Estimating Equations*. Ph.D. thesis, Department of Statistics, The Pennsylvania State University, University Park, PA 16802.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82**, 802–807.

Venables, W. N. and Ripley, B. D. (1999). *Modern Applied Statistics with S-plus*. Springer-Verlag, 3rd edition.