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 Scale Estimation and Hypothesis Testing based on Quadratic Inference Function

By Chanseok Park and Bruce G. Lindsay

Technical Report #99-03

November, 1999

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

Robust Scale Estimation and Hypothesis Testing based on Quadratic Inference Function[†]

By Chanseok Park and Bruce G. Lindsay

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

The standard deviation is the most common estimator of scale, but it is known to lack resistance and robustness. In this paper, we propose a new scale estimator based on the quadratic inference function (QIF). This estimator is efficient, robust and possesses a reasonable breakdown point. In addition to obtaining parameter estimators, we develop χ^2 tests of parametric hypotheses such as a test of the equality of variances in the k-sample case. Monte Carlo experiments suggest that this new method is reasonably robust for departures from normality and for contaminated samples. We compare the performance of this new method with other well-known methods by Monte Carlo simulation.

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

1 Introduction

At present, robust scale estimation has not yet been studied as carefully as robust location estimation. Nevertheless, scale estimators play an important role in practice. For example, they can be used to compare the variability of different samples, construct confidence intervals, standardize observations, evaluate the lot-to-lot reproducibility of a manufacturing process, and so on. In the scale estimation problem, we consider the estimation of the one-dimensional scale parameter σ for the scale family of distributions $F_\sigma(x) = F_1(x/\sigma)$ with densities $f_\sigma(x) = \sigma^{-1} f_1(x/\sigma)$. A prime example of a scale estimator is the standard deviation which can be considered as an M-estimator of the normal model. We propose an adaptive estimator which is a combination of an efficient estimator and a robust estimator. For this purpose, we need to obtain score functions satisfying $\int \chi(x;\sigma) dF_\sigma(x) = 0$.

We take a brief look at M-estimation and its influence function. An M-estimator for scale is defined by an implicit relation of the form

$$\int \chi(x;\sigma) dF_\sigma(x) = 0,$$

[†]This work will be presented at the Conference in honor of Professor C. R. Rao on the occasion of his 80th birthday at the University of Texas at San Antonio, March 16 – 19, 2000.

with $\chi(x;\sigma) = \chi(x/\sigma)$. We call this $\chi(\cdot)$ a score function. Then its influence function (IF) is obtained as

$$\mathrm{IF}(x;\chi,F_\sigma) = \frac{\chi(x;\sigma)}{-\int \frac{\partial}{\partial\sigma}\chi(x;\sigma)dF_\sigma(x)} = \frac{\chi(x/\sigma)\sigma}{\int \chi'(x/\sigma)x/\sigma dF_\sigma(x)}.$$

The maximum likelihood score is given by $\chi(x) = -x\frac{f_1'(x)}{f_1(x)} - 1$. At the normal model, this yields

$$\chi(x) = x^2 - 1,$$

and

$$\mathrm{IF}(x;\chi,F_\sigma) = \frac{1}{2}(x^2 - 1).$$

This maximum likelihood estimator is efficient but not robust and has asymptotic breakdown point $\varepsilon^* = 0$. Another possible score function is an even and bounded function, which yields a robust estimator.

Next, we briefly review the QIF. For more details, the reader is referred to Park and Lindsay (1999). Let $\mathbf{X} = (X_1, \dots, X_n)$ be the random sample and let $\boldsymbol{\theta}$ be a p-dimensional parameter and $\mathbf{g}(\boldsymbol{\theta};X) = (g_1(\boldsymbol{\theta};X), \dots, g_p(\boldsymbol{\theta};X))^T$ be a k-dimensional vector of extended score functions satisfying $E[\mathbf{g}(\boldsymbol{\theta},X)] = \mathbf{0}$. Then the quadratic inference function (QIF) based on the score functions is defined as

$$Q_n(\boldsymbol{\theta};\mathbf{g},\mathbf{X}) = \bar{\mathbf{g}}_n(\boldsymbol{\theta};\mathbf{X})^T \mathbf{C}_n^+(\boldsymbol{\theta};\mathbf{X})\bar{\mathbf{g}}_n(\boldsymbol{\theta};\mathbf{X}), \tag{1}$$

where $\bar{\mathbf{g}}_n(\boldsymbol{\theta};\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n \mathbf{g}(\boldsymbol{\theta};X_i)$, $\mathbf{C}_n(\boldsymbol{\theta};\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n \mathbf{g}(\boldsymbol{\theta};X_i)\mathbf{g}(\boldsymbol{\theta};X_i)^T$ and $\mathbf{C}_n^+$ denotes a Moore-Penrose generalized inverse of $\mathbf{C}_n$. The QIF estimator based on the score functions $\mathbf{g}(\cdot)$ is defined as

$$\hat{\boldsymbol{\theta}}_n = \arg\inf_{\boldsymbol{\theta}} Q_n(\boldsymbol{\theta};\mathbf{g},\mathbf{X}). \tag{2}$$

According to Theorem 7 or 8 in Park and Lindsay (1999), we can construct an asymptotically efficient but *robust* QIF estimator by selecting the maximum likelihood score and a robust score. That is to say, the influence function of the QIF estimator at the true model is the same as that of the maximum likelihood estimator (MLE), but yet it has a positive breakdown point.

2 Scale Estimates

In this section, we introduce how to construct an asymptotically efficient estimator with the robustness property and we calculate the breakdown point for this estimator. We propose the following choices:

$$g_1(\sigma;x) = (x/\sigma)^2 - 1,$$

and

$$g_2(\sigma; x) = \chi(x/\sigma),$$

where $\chi(\cdot)$ is an even and bounded function. This yields a robust QIF estimator with a reasonable breakdown point. 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. 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:

$$\mathrm{Bias}(m; t_n, \mathbf{x}) = \sup_{\mathbf{x}^{(m)}} |t_n(\mathbf{x}^{(m)}) - t_n(\mathbf{x})|, \tag{3}$$

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

$$\varepsilon_n^*(t; \mathbf{x}) = \frac{1}{n} \min_m \{m : \mathrm{Bias}(m; t, \mathbf{x}) = \infty\}. \tag{4}$$

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^* as $n \to \infty$. We call this the **asymptotic breakdown point** which is given by

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

Theorem 1. *Suppose that the QIF is based on $g_1(\sigma; x) = (x/\sigma)^2 - 1$ and $g_2(\sigma; x) = \chi(x/\sigma)$ where $\chi(\cdot)$ is even and bounded. Then the QIF estimator of the scale model $\{F_\sigma(x) = F_1(x/\sigma), \sigma \in \mathbb{R}^+\}$ has an asymptotic breakdown point ε^* as follows:*

$$\varepsilon^* = \inf_\epsilon \{\epsilon : Q(\epsilon, \zeta) < \epsilon\}, \tag{5}$$

where ζ satisfies the equation

$$D(\zeta) = \chi\big(\tfrac{1}{\zeta}\big)\zeta^2 + (1 - \zeta^2)\chi(0) = 0,$$

and $Q(\cdot)$ is given as

$$Q(\epsilon, \zeta) = a\Big(\frac{au + bv}{a^2 + b^2}\Big)^2,$$

with $a = \epsilon(1/\zeta^4 - 2/\zeta^2) + 1$, $b = \epsilon(1/\zeta^2 - 1)\chi(\frac{1}{\zeta}) - \chi(0)(1 - \epsilon)$, $u = \epsilon/\zeta^2 - 1$, and $v = \epsilon\chi(\frac{1}{\zeta}) + (1 - \epsilon)\chi(0)$.

Proof. The infinite bias in equation (4) suggests that we change the values of m observations to large $M > 0$. Let us denote $\mathbf{x}^{(m)} = (x_1^*, \dots, x_m^*, x_1, \dots, x_{n-m})$, where $x_i^* = M$ for $i = 1, \dots, m$ and x_i are the sample from $F_1(x)$. Let $\epsilon_n = m/n$ and $B = \chi(\infty) = \chi(-\infty)$. Then when $0 < \sigma \le 1$, we have the following results as $M \to \infty$.

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

$$\frac{1}{n}\sum_{i=1}^{n} g_2(\sigma; x_i) = B\epsilon_n + c_n + o(1),$$

$$\frac{1}{n}\sum_{i=1}^{n} g_1(\sigma; x_i)^2 = \frac{\epsilon_n}{\sigma^4}M^4 + o(M^4),$$

$$\frac{1}{n}\sum_{i=1}^{n} g_1(\sigma; x_i)g_2(\sigma; x_i) = \frac{B\epsilon_n}{\sigma^2}M^2 + o(M^2),$$

and

$$\frac{1}{n}\sum_{i=1}^{n} g_2(\sigma; x_i)^2 = B^2\epsilon_n + d_n + o(1),$$

where $c_n = (1-\epsilon_n)\frac{1}{n-m}\sum_{i=1}^{n-m}\chi(x_i/\sigma)$ and $d_n = (1-\epsilon_n)\frac{1}{n-m}\sum_{i=1}^{n-m}\chi(x_i/\sigma)^2$. By substituting the above into (1), we have the result, $Q_n(\sigma; g_1, g_2) = \epsilon_n + c_n^2/d_n + o(1)$ as $M \to \infty$. Noting that $c_n \to c = (1-\epsilon)\int \chi(x/\sigma)dx$ with $c = 0$ for $\sigma = 1$ and $c > 0$ for $\sigma \neq 1$, we have the result in the limit,

$$Q(\sigma; g_1, g_2) = \lim_{n\to\infty} Q_n(\sigma; g_1, g_2) = \begin{cases} \epsilon + c^2/d & : \quad 0 < \sigma < 1 \\ \epsilon & : \quad \sigma = 1 \end{cases}, \tag{6}$$

with $d = (1-\epsilon)\int \chi(x/\sigma)^2 dx$ and $\epsilon = \lim_{n\to\infty}\epsilon_n$.

In the case of $\sigma > 1$, denoting $\sigma = M\eta$, we have the following results.

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

$$\frac{1}{n}\sum_{i=1}^{n} g_2(\sigma; x_i) = \epsilon_n\chi(1/\eta) + (1-\epsilon_n)\chi(0) + o(1),$$

$$\frac{1}{n}\sum_{i=1}^{n} g_1(\sigma; x_i)^2 = \epsilon_n(\frac{1}{\eta^4} - \frac{2}{\eta^2}) + 1 + o(1),$$

$$\frac{1}{n}\sum_{i=1}^{n} g_1(\sigma; x_i)g_2(\sigma; x_i) = \epsilon_n(\frac{1}{\eta^2} - 1)\chi(1/\eta) - (1-\epsilon_n)\chi(0) + o(1),$$

and

$$\frac{1}{n}\sum_{i=1}^{n} g_2(\sigma;x_i)^2 = \epsilon_n\chi(1/\eta)^2 + (1-\epsilon_n)\chi(0)^2 + o(1).$$

By substituting the above into (1), we have the result in the limit.

$$Q(\sigma;g_1,g_2) = \begin{pmatrix} u \\ v \end{pmatrix}^T \mathbf{C}^+ \begin{pmatrix} u \\ v \end{pmatrix},$$

with $\mathbf{C} = \left(\begin{smallmatrix} a & b \\ b & h \end{smallmatrix}\right)$ and $h = \epsilon\chi(1/\eta)^2 + (1-\epsilon)\chi(0)^2$.

If $\det(\mathbf{C}) \neq 0$, *i.e.* $\mathbf{C}^+ = \mathbf{C}^{-1}$, then we have $Q(\sigma;g_1,g_2) = 1$ after some tedious algebra. If $\det(\mathbf{C}) = 0$, then we get

$$\mathbf{C}^+ = \frac{1}{(a+h)^2}\begin{pmatrix} a & b \\ b & h \end{pmatrix}.$$

It follows that

$$Q(\epsilon,\eta) = a\Big(\frac{au+bv}{a^2+b^2}\Big)^2.$$

Hence, we have two local minima when $\sigma = 1$ and $\det(\mathbf{C}) = 0$, where

$$\det(\mathbf{C}) = \frac{\epsilon(1-\epsilon)}{\eta^4}\big(\chi(1/\eta)\eta^2 + (1-\eta^2)\chi(0)\big)^2 = \frac{\epsilon(1-\epsilon)}{\eta^4}D(\eta)^2.$$

This completes the proof. □

We propose Huber's $\chi(\cdot)$ and the probability integral transform (PIT) function for a robust score function. First, Huber's $\chi(\cdot)$ is given by

$$\chi_\kappa(z) = \min(z^2,\kappa^2) - \beta_\kappa,$$

where β_κ is a Fisher consistency constant determined by

$$\begin{aligned}\beta_\kappa &= \int \min(z^2,\kappa^2)d\Phi \\ &= 2\Phi(\kappa) - 1 - 2\kappa\phi(\kappa) + 2\kappa^2(1-\Phi(\kappa)).\end{aligned}$$

Secondly, the PIT score is given by

$$\chi_\tau(z) = \big(2\Phi(\frac{z}{\tau}) - 1\big)^2 - \beta_\tau,$$

where $\Phi(\cdot)$ is the standard normal distribution and β_τ is obtained by

$$\beta_\tau = \int \big(2\Phi(\frac{z}{\tau}) - 1\big)^2 d\Phi. \tag{7}$$

In Table 1, the asymptotic breakdown points, β_κ, β_τ and ζ satisfying

$$D(\zeta) = \chi\big(\tfrac{1}{\zeta}\big)\zeta^2 + (1-\zeta^2)\chi(0) = 0,$$

are given for the QIF estimator using either the maximum likelihood score of the normal model and Huber's χ-function or the maximum likelihood score and the PIT score.

Table 1: Breakdown points for the QIF estimator based on the maximum likelihood (ML) score of the normal model and Huber's $\chi_\kappa(\cdot)$ or the PIT's $\chi_\tau(\cdot)$.

κ, τ	Huber	ML and Huber			ML and PIT		
	ε^*	ε^*	β_κ	ζ	ε^*	β_τ	ζ
$\to 0$	0.5	0.5	0	1	0.5	1	1
0.1	0.486	0.4734	0.0095	0.9731	0.4552	0.9103	0.9541
0.2	0.472	0.4470	0.0358	0.9455	0.4114	0.8229	0.9071
0.3	0.457	0.4209	0.0758	0.9175	0.3698	0.7395	0.8600
0.4	0.442	0.3953	0.1265	0.8891	0.3323	0.6617	0.8152
0.5	0.425	0.3703	0.1851	0.8605	0.3011	0.5903	0.7761
0.6	0.409	0.3460	0.2491	0.8318	0.2766	0.5259	0.7438
0.7	0.392	0.3225	0.3160	0.8031	0.2576	0.4684	0.7178
0.8	0.375	0.2999	0.3840	0.7745	0.2428	0.4175	0.6969
0.9	0.358	0.2785	0.4511	0.7463	0.2312	0.3726	0.6800
1.0	0.340	0.2580	0.5161	0.7184	0.2219	0.3333	0.6662
1.1	0.323	0.2387	0.5777	0.6910	0.2144	0.2989	0.6548
1.2	0.306	0.2206	0.6352	0.6642	0.2083	0.2688	0.6454
1.3	0.289	0.2036	0.6880	0.6381	0.2032	0.2425	0.6375
1.4	0.273	0.1877	0.7358	0.6127	0.1990	0.2194	0.6309
1.5	0.257	0.1730	0.7785	0.5882	0.1954	0.1991	0.6252
1.6	0.242	0.1594	0.8160	0.5646	0.1924	0.1813	0.6204
1.7	0.227	0.1468	0.8487	0.5419	0.1899	0.1655	0.6162
1.8	0.213	0.1353	0.8767	0.5202	0.1877	0.1516	0.6126
1.9	0.200	0.1247	0.9006	0.4995	0.1857	0.1392	0.6094
2.0	0.187	0.1151	0.9205	0.4797	0.1840	0.1282	0.6067
2.1	0.175	0.1062	0.9371	0.4610	0.1826	0.1184	0.6043
2.2	0.164	0.0982	0.9507	0.4432	0.1813	0.1096	0.6021
2.3	0.154	0.0909	0.9617	0.4264	0.1801	0.1016	0.6002
2.4	0.144	0.0843	0.9705	0.4105	0.1791	0.0945	0.5985
2.5	0.135	0.0782	0.9776	0.3955	0.1782	0.0881	0.5970
2.6	0.127	0.0727	0.9831	0.3813	0.1774	0.0823	0.5956
2.7	0.119	0.0677	0.9873	0.3680	0.1767	0.0770	0.5944
2.8	0.112	0.0632	0.9906	0.3555	0.1760	0.0722	0.5933
2.9	0.106	0.0590	0.9931	0.3436	0.1754	0.0678	0.5923
3.0	0.100	0.0553	0.9950	0.3325	0.1749	0.0638	0.5914

3 Simultaneous Estimates of Location and Scale

In order to make a simultaneous estimator of location and scale, we need to couple the location and scale scores. For example, the famous couple $g(\mu, \sigma; x) = (x-\mu)/\sigma$ and $h(\mu, \sigma; x) = \{(x-\mu)/\sigma\}^2 - 1$ give the maximum likelihood estimators for the normal distribution. To obtain the robustness property, we need to add robustified score functions of them. The following theorem shows the QIF estimator of location and scale has a reasonable breakdown point. For example, if we choose the PIT score with $\tau = 1$, then we have an asymptotic breakdown point ε^* of at least 22%.

Theorem 2. *Suppose that the QIF is based on such score functions as*

$$g_1 = \frac{x-\mu}{\sigma},\ g_2 = \psi(\frac{x-\mu}{\sigma}),\ g_3 = (\frac{x-\mu}{\sigma})^2 - 1,\ g_4 = \chi(\frac{x-\mu}{\sigma}),$$

where $\psi(\cdot)$ is monotone, bounded with $\psi(\infty) = -\psi(-\infty)$ and $\psi(0) = 0$, and $\chi(\cdot)$ is even and bounded. Then the asymptotic breakdown point of its estimators of the location-scale model $\{F_{\mu,\sigma}(x) = F_{0,1}(\frac{x-\mu}{\sigma}),\ \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$ is at least $\varepsilon^* = \min(\frac{1}{4}, \varepsilon_S^*)$, where ε_S^* is an asymptotic breakdown point of the scale *model given by Theorem 1.*

Proof. Let $\mathbf{g} = (g_1, g_2, g_3, g_4)^T$, $\mathbf{g}_L = (g_1, g_2)^T$ and $\mathbf{g}_S = (g_3, g_4)^T$. By Theorem 4 in Park and Lindsay (1999), we have $Q_n(\mu, \sigma; \mathbf{g}) \geq Q_n(\mu, \sigma; \mathbf{g}_L)$ and $Q_n(\mu, \sigma; \mathbf{g}) \geq Q_n(\mu, \sigma; \mathbf{g}_S)$. It follows that

$$Q_n(\mu, \sigma; \mathbf{g}) \geq \frac{1}{2}\big(Q_n(\mu, \sigma; \mathbf{g}_L) + Q_n(\mu, \sigma; \mathbf{g}_S)\big).$$

In the limit, both $Q_n(\mu, \sigma; \mathbf{g}_L)$ and $Q_n(\mu, \sigma; \mathbf{g}_S)$ have the minimum ϵ when $\mu = 0$ and $\sigma = 1$, provided that $\epsilon < \min(\frac{1}{4}, \varepsilon_S^*)$. Hence, all we need to show is that $Q(\mu, \sigma; \mathbf{g}) = \lim_{n\to\infty} Q_n(\mu, \sigma; \mathbf{g}) \to \epsilon$ at $\mu = 0$ and $\sigma = 1$ as $M \to \infty$.

Let us denote $\mathbf{x}^{(m)} = (x_1^*, \ldots, x_m^*, x_1, \ldots, x_{n-m})$, where $x_i^* = -M$ and $x_j^* = M$ for $i = 1, \ldots, m_1$, $j = 1, \ldots, m_2$ with $m = m_1 + m_2$ and x_k are the sample from $F_{0,1}(x)$. Let $\epsilon_n = m/n$, $\delta_n = m_1/m$, $A = \psi(\infty) = -\psi(-\infty)$, and $B = \chi(\infty) = \chi(-\infty)$. Then we have

$$Q_n(0, 1; \mathbf{g}) = \bar{\mathbf{g}}_n^T \mathbf{C}_n^+ \bar{\mathbf{g}}_n, \tag{8}$$

where $\bar{\mathbf{g}}_n$ and $\mathbf{C}_n$ are given by the following results as $M \to \infty$. For brevity, we omit the little $o(\cdot)$ terms without loss of generality.

$$\bar{\mathbf{g}}_n = \begin{pmatrix} (1-2\delta_n)\epsilon_n M \\ A(1-2\delta_n)\epsilon_n + a_n \\ \epsilon_n M^2 \\ B\epsilon_n + c_n \end{pmatrix}$$

and

$$\mathbf{C}_n = \begin{pmatrix} \epsilon_n M^2 & A\epsilon_n M & (1-2\delta_n)\epsilon_n M^3 & B(1-2\delta_n)\epsilon_n M \\ A\epsilon_n M & A^2\epsilon_n + b_n & A(1-2\delta_n)\epsilon_n M^2 & AB(1-2\delta_n)\epsilon_n \\ (1-2\delta_n)\epsilon_n M^3 & A(1-2\delta_n)\epsilon_n M^2 & \epsilon_n M^4 & B\epsilon_n M^2 \\ B(1-2\delta_n)\epsilon_n M & AB(1-2\delta_n)\epsilon_n & B\epsilon_n M^2 & B^2\epsilon_n + d_n \end{pmatrix},$$

where

$$a_n = (1-\epsilon_n)\frac{1}{n-m}\sum_{i=1}^{n-m}\psi(x_i), \qquad b_n = (1-\epsilon_n)\frac{1}{n-m}\sum_{i=1}^{n-m}\psi(x_i)^2,$$
$$c_n = (1-\epsilon_n)\frac{1}{n-m}\sum_{i=1}^{n-m}\chi(x_i/\sigma), \text{ and } \quad d_n = (1-\epsilon_n)\frac{1}{n-m}\sum_{i=1}^{n-m}\chi(x_i/\sigma)^2.$$

Since $\det(\mathbf{C}_n) = 4\delta_n(1-\delta_n)\epsilon_n^2 b_n d_n M^6$, $\mathbf{C}_n^{-1}$ exists unless δ_n is 0 or 1. Hence $\mathbf{C}_n^{-1}$ is obtained as

$$\mathbf{C}_n^{-1} = \begin{pmatrix} \frac{1}{4\delta_n(1-\delta_n)\epsilon_n M^2} + \frac{A^2}{b_n M^2} & \frac{-A}{b_n M} & \frac{2\delta_n - 1}{4\delta_n(1-\delta_n)\epsilon_n M^3} & 0 \\ \frac{-A}{b_n M} & \frac{1}{b_n} & 0 & 0 \\ \frac{2\delta_n - 1}{4\delta_n(1-\delta_n)\epsilon_n M^3} & 0 & \frac{1}{4\delta_n(1-\delta_n)\epsilon_n M^4} + \frac{B^2}{d_n M^4} & \frac{-B}{d_n M^2} \\ 0 & 0 & \frac{-B}{d_n M^2} & \frac{1}{d_n} \end{pmatrix}.$$

Substituting the above results into (8), we have

$$Q_n(0,1;\mathbf{g}) = \bar{\mathbf{g}}_n^T \mathbf{C}_n^{-1} \bar{\mathbf{g}}_n = \epsilon_n + \frac{a_n^2}{b_n} + \frac{c_n^2}{d_n}.$$

Considering $a_n \to 0$ and $c_n \to 0$ as $n \to \infty$, we have $Q(0,1;\mathbf{g}) = \epsilon$.

If $\delta_n = 0$, *i.e.*, $\det(\mathbf{C}_n) = 0$, then we need to find the Moore-Penrose generalized inverse of $\mathbf{C}_n$. It can be obtained by using Theorems 6.5.3 and 6.5.4 in Graybill (1983). We get

$$\mathbf{C}_n^+ =$$
$$\begin{pmatrix} \frac{(A^2 d_n + B^2 b_n)\epsilon_n + b_n d_n}{b_n d_n \epsilon_n M^2 (1+M^2)^2} & \frac{-A}{b_n M(1+M^2)} & \frac{(A^2 d_n + B^2 b_n)\epsilon_n + b_n d_n}{b_n d_n \epsilon_n M (1+M^2)^2} & \frac{-B}{d_n M(1+M^2)} \\ \frac{-A}{b_n M(1+M^2)} & \frac{1}{b_n} & \frac{-A}{b_n(1+M^2)} & 0 \\ \frac{(A^2 d_n + B^2 b_n)\epsilon_n + b_n d_n}{b_n d_n \epsilon_n M (1+M^2)^2} & \frac{-A}{b_n(1+M^2)} & \frac{(A^2 d_n + B^2 b_n)\epsilon_n + b_n d_n}{b_n d_n \epsilon_n (1+M^2)^2} & \frac{-B}{d_n(1+M^2)} \\ \frac{-B}{d_n M(1+M^2)} & 0 & \frac{-B}{d_n(1+M^2)} & \frac{1}{d_n} \end{pmatrix}.$$

Substituting the above results into (8), we have

$$Q_n(0,1;\mathbf{g}) = \bar{\mathbf{g}}_n^T \mathbf{C}_n^+ \bar{\mathbf{g}}_n = \epsilon_n + \frac{a_n^2}{b_n} + \frac{c_n^2}{d_n},$$

and therefore $Q(0,1;\mathbf{g}) = \epsilon$. For $\delta_n = 1$, we have the same result. $\square$

4 Hypothesis Testing using the QIF

In addition to robust point estimations, using Theorem 12 in Park and Lindsay (1999), we also obtain a variety of χ^2 tests. On the other hand, Huber's scale estimator, MAD and IQR can only be used for point estimation. For example, we can create a robust test for homogeneity of variances by using the QIF. We introduce a test for the equality of two population variances, which can be easily extended to the k-sample case.

Let $\mathbf{X} = (X_1, \ldots, X_{n_1})$ and $\mathbf{Y} = (Y_1, \ldots, Y_{n_2})$ be independent random samples from two populations. The null hypothesis of equal variances, $H_0 : \sigma_1^2 = \sigma_2^2$, is to be tested at significance level α against the alternative hypothesis, $H_1 : \sigma_1^2 \neq \sigma_2^2$.

We propose the score vector, $\mathbf{g} = (g_1, g_2, g_3, g_4)^T$ with

$$g_1(\mu,\sigma;X) = \frac{X-\mu}{\sigma}, \qquad g_2(\mu,\sigma;X) = 2\Phi\big(\frac{X-\mu}{\tau\sigma}\big) - 1, \tag{9}$$

$$g_3(\mu,\sigma;X) = \big(\frac{X-\mu}{\sigma}\big)^2 - 1, \qquad g_4(\mu,\sigma;X) = \Big[2\Phi\big(\frac{X-\mu}{\tau\sigma}\big) - 1\Big]^2 - \beta_\tau, \tag{10}$$

where τ is a tuning constant and β_τ is given in (7). Note that g_1 and g_3 are maximum likelihood scores of the normal model and g_2 and g_4 are robustified versions of g_1 and g_3, respectively. We obtain the following statistics:

$$n_1 Q_{n_1}(\hat{\mu}_1, \hat{\sigma}_1; \mathbf{g}, \mathbf{X}) + n_2 Q_{n_2}(\hat{\mu}_2, \hat{\sigma}_2; \mathbf{g}, \mathbf{Y}) \xrightarrow{\mathcal{D}} \chi_4^2, \tag{11}$$

$$n_1 Q_{n_1}(\hat{\mu}_1^*, \hat{\sigma}_0^*; \mathbf{g}, \mathbf{X}) + n_2 Q_{n_2}(\hat{\mu}_2^*, \hat{\sigma}_0^*; \mathbf{g}, \mathbf{Y}) \xrightarrow{\mathcal{D}} \chi_5^2, \tag{12}$$

and

$$n_1\{Q_{n_1}(\hat{\mu}_1^*, \hat{\sigma}_0^*; \mathbf{g}, \mathbf{X}) - Q_{n_1}(\hat{\mu}_1, \hat{\sigma}_1; \mathbf{g}, \mathbf{X})\} + n_2\{Q_{n_2}(\hat{\mu}_2^*, \hat{\sigma}_0^*; \mathbf{g}, \mathbf{Y}) - Q_{n_2}(\hat{\mu}_2, \hat{\sigma}_2; \mathbf{g}, \mathbf{Y})\} \xrightarrow{\mathcal{D}} \chi_1^2. \tag{13}$$

The parameter estimates in the statistics above are obtained as follows:

$$(\hat{\mu}_1 \ \hat{\sigma}_1)^T = \arg\inf_{\mu,\sigma} n_1 Q_{n_1}(\mu,\sigma;\mathbf{g},\mathbf{X}),$$

$$(\hat{\mu}_2 \ \hat{\sigma}_2)^T = \arg\inf_{\mu,\sigma} n_2 Q_{n_2}(\mu,\sigma;\mathbf{g},\mathbf{Y}),$$

and

$$(\hat{\mu}_1^* \ \hat{\mu}_2^* \ \hat{\sigma}_0^*)^T = \arg\inf_{\mu_1,\mu_2,\sigma}\{n_1 Q_{n_1}(\mu_1,\sigma;\mathbf{g},\mathbf{X}) + n_2 Q_{n_2}(\mu_2,\sigma;\mathbf{g},\mathbf{Y})\}.$$

However, in this application, the scale parameter is a matter of concern. Hence the location can be regarded as a *nuisance parameter*. When the location is a nuisance parameter, it makes sense to estimate it as robustly as possible although this can be inefficient. The simplest robust location estimator is the median which has a breakdown point $\varepsilon^* = 50\%$. If we use the median as the estimate for the location parameter, we can reduce computational complexity. We use $\tilde{\mu}_1 = \mathrm{median}(\mathbf{X})$ and $\tilde{\mu}_2 = \mathrm{median}(\mathbf{Y})$ for μ_1 and μ_2 respectively. Then we need only g_3 and g_4 and denote $\mathbf{g}_S = (g_3, g_4)^T$. Then we obtain the following statistics:

$$n_1 Q_{n_1}(\tilde{\mu}_1, \hat{\sigma}_1; \mathbf{g}_S, \mathbf{X}) + n_2 Q_{n_2}(\tilde{\mu}_2, \hat{\sigma}_2; \mathbf{g}_S, \mathbf{Y}) \xrightarrow{\mathcal{D}} \chi_2^2, \tag{14}$$

$$n_1 Q_{n_1}(\tilde{\mu}_1, \hat{\sigma}_0; \mathbf{g}_S, \mathbf{X}) + n_2 Q_{n_2}(\tilde{\mu}_2, \hat{\sigma}_0; \mathbf{g}_S, \mathbf{Y}) \xrightarrow{\mathcal{D}} \chi_3^2, \tag{15}$$

and

$$\begin{aligned} & n_1\{Q_{n_1}(\tilde{\mu}_1, \hat{\sigma}_0; \mathbf{g}_S, \mathbf{X}) - Q_{n_1}(\tilde{\mu}_1, \hat{\sigma}_1; \mathbf{g}_S, \mathbf{X})\} \\ & + n_2\{Q_{n_2}(\tilde{\mu}_2, \hat{\sigma}_0; \mathbf{g}_S, \mathbf{Y}) - Q_{n_2}(\tilde{\mu}_2, \hat{\sigma}_2; \mathbf{g}_S, \mathbf{Y})\} \xrightarrow{\mathcal{D}} \chi_1^2, \end{aligned} \tag{16}$$

with the parameter estimates provided as follows:

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

5 Simulations

In this section we present an extensive numerical study to assess the performance of the proposed estimator and hypothesis test. First, we present the performance of a variety of scale and location estimators by comparing their estimated biases and mean square errors (MSEs). Secondly, we present the performance of various well-known robust tests by comparing their empirical significance levels and estimated powers.

5.1 Estimation of scale parameter

We compare the performance of a variety of estimators. First, we briefly review these estimators. They are defined for the sample, $X_1, \ldots, X_n$ as follows.

- Mean square deviation (SD):

$$S_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

- Mean absolute deviation (AD) from the sample mean:

$$D_n = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}|.$$

- Median absolute deviation (MAD) from the sample median:

$$\begin{aligned} \mathrm{MAD}_n &= \frac{1}{\Phi^{-1}(3/4)} \operatorname*{median}_i \{|X_i - \operatorname*{median}_j (X_j)|\} \\ &\approx 1.4826 \operatorname*{median}_i \{|X_i - \operatorname*{median}_j (X_j)|\}. \end{aligned}$$

- Interquartile range (IQR):

$$\begin{aligned} \mathrm{IQR}_n &= \frac{1}{2\Phi^{-1}(3/4)} (X_{(3n/4)} - X_{(n/4)}) \\ &\approx 0.7413 (X_{(3n/4)} - X_{(n/4)}). \end{aligned}$$

- Huber's M-estimator: Huber (1981) defines the *Simultaneous M-estimate of location and scale* as solving for μ and σ through

$$\sum \psi_\kappa (\frac{X_i - \mu}{\sigma}) = 0, \tag{17}$$

and

$$\sum \chi_\kappa (\frac{X_i - \mu}{\sigma}) = 0, \tag{18}$$

where, in most cases, $\psi_\kappa(\cdot)$ is an odd function and $\chi_\kappa(\cdot)$ an even function. Huber (1981, pg. 137; 1964 pg. 96) sets

$$\psi_\kappa(z) = \max(-\kappa, \min(\kappa, z))$$

and

$$\chi_\kappa(z) = \psi_\kappa(z)^2 - \beta_\kappa$$

in his Proposal 2. In order to get Fisher consistency of the normal model, β_κ is determined by

$$\begin{aligned} \beta_\kappa &= \int \psi_\kappa(z)^2 d\Phi((z) = \int \min(\kappa^2, z^2) d\Phi(z) \\ &= 2\Phi(\kappa) - 1 - 2\kappa\phi(\kappa) + 2\kappa^2 \big(1 - \Phi(\kappa)\big). \end{aligned}$$

If only scale parameter is a matter of concern, the location can be regarded as a nuisance parameter. In this case, we only need to solve equation (18) with the location provided. We use the median as the estimate for the location. In Tables 2, 3 and 4, we illustrate both simultaneous and scale-only estimators.

- QIF estimator: We propose the two score sets based on the score functions given by (9) and (10). These are $\mathbf{g}_L = (g_1\ g_2)^T$ and $\mathbf{g}_S = (g_3\ g_4)^T$. Then the simultaneous QIF estimator based on $\mathbf{g}_L$ and $\mathbf{g}_S$ is given by

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} = \arg\inf_{\mu,\sigma} Q_n(\mu, \sigma; \mathbf{g}_L, \mathbf{g}_S).$$

  Like the Huber estimator, we can create two estimators — simultaneous and scale-only. To estimate the scale alone, we consider only $\mathbf{g}_S$ in the QIF along with the median provided for the location. In the tables, we report both empirical results.

In Table 2, we present the estimated biases and mean square errors (MSEs) of the various scale and location estimators based on $10,000$ random samples from the normal distribution, $N(0, 1)$. As illustrated, the MLEs (mean and SD) perform better than any other estimator. Regarding the Huber and QIF estimators, each has different versions, that is, scale-only estimator with the median provided as the location and simultaneous estimator for the location and scale. When the sample size is small, the performance of the Huber and QIF estimators is similar and both are better than the MAD and IQR. When the sample size is large, the QIF estimator performs better than the Huber estimator. If the scale is a parameter of interest, the QIF estimator with the median is more reliable than the simultaneous QIF estimator.

In Tables 3 and 4, we repeat the above experiments, but $10,000$ random samples are now generated from contaminated distributions. In Table 3, the data comes from the mixture, $0.9N(0, 1) + 0.1N(0, 9)$, where the second component is considered to be the contamination, and the scale parameter of the first component is what we are interested in estimating. The biases and MSEs are calculated with respect to the true scale parameter, $\sigma = 1$. The results show that the QIF estimators — scale-only and simultaneous — perform the best. As the sample size gets larger, the QIF estimator outperforms the other estimators by even a larger amount. In Table 4, the data comes from the mixture $0.9N(0, 1) + 0.1\Delta_{10}$, where Δ_ζ is the degenerate distribution with ζ. In this case, it is clearly seen that the QIF estimator is the best. Compared with other estimators, the performance of the QIF estimator under the asymmetric contamination is better than that under the symmetric contamination.

Table 2: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from $N(0,1)$ with sample size $n = 20, 40, 100$.

Estimator	$n = 20$		$n = 40$		$n = 100$	
	Bias	MSE	Bias	MSE	Bias	MSE
Mean	0.0002	0.0492	−0.0016	0.0248	−0.0002	0.0099
SD	−0.0394	0.0265	−0.0200	0.0131	−0.0074	0.0052
AD	−0.0268	0.0298	−0.0139	0.0148	−0.0049	0.0059
Median	−0.0005	0.0717	−0.0019	0.0375	−0.0004	0.0155
MAD	−0.0457	0.0654	−0.0237	0.0333	−0.0076	0.0138
IQR	−0.0704	0.0635	−0.0373	0.0331	−0.0139	0.0138
Huber ($\kappa = 1.0$) with median						
location	−0.0005	0.0717	−0.0019	0.0375	−0.0004	0.0155
scale	−0.0560	0.0544	−0.0292	0.0268	−0.0105	0.0107
Huber ($\kappa = 1.0$) Proposal 2						
location	0.0000	0.0534	−0.0016	0.0272	0.0000	0.0110
scale	0.0259	0.0556	0.0112	0.0269	0.0054	0.0107
QIF ($\tau = 1.0$) with median						
location	−0.0005	0.0717	−0.0019	0.0375	−0.0004	0.0155
scale	−0.0829	0.0422	−0.0411	0.0175	−0.0175	0.0059
QIF ($\tau = 1.0$)						
location	0.0001	0.0730	−0.0022	0.0310	−0.0006	0.0108
scale	−0.1664	0.0645	−0.0831	0.0235	−0.0347	0.0071

Table 3: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from $0.9N(0,1) + 0.1N(0,9)$ with sample size $n = 20, 40, 100$.

Estimator	$n = 20$		$n = 40$		$n = 100$	
	Bias	MSE	Bias	MSE	Bias	MSE
Mean	0.0010	0.0877	0.0003	0.0450	0.0005	0.0179
SD	0.2628	0.1674	0.3035	0.1490	0.3280	0.1316
AD	0.1781	0.0857	0.1896	0.0632	0.1970	0.0494
Median	−0.0008	0.0829	−0.0025	0.0432	−0.0004	0.0178
MAD	0.0376	0.0760	0.0576	0.0414	0.0745	0.0216
IQR	0.0089	0.0688	0.0416	0.0381	0.0666	0.0201
Huber ($\kappa = 1.0$) with median						
location	−0.0008	0.0829	−0.0025	0.0432	−0.0004	0.0178
scale	0.0281	0.0620	0.0545	0.0334	0.0738	0.0180
Huber ($\kappa = 1.0$) Proposal 2						
location	0.0001	0.0643	−0.0016	0.0327	0.0002	0.0132
scale	0.1185	0.0798	0.0988	0.0412	0.0913	0.0210
QIF ($\tau = 1.0$) with median						
location	−0.0008	0.0829	−0.0025	0.0432	−0.0004	0.0178
scale	0.0195	0.0533	0.0610	0.0294	0.0819	0.0169
QIF ($\tau = 1.0$)						
location	0.0005	0.0844	−0.0021	0.0363	0.0002	0.0131
scale	−0.0598	0.0570	0.0276	0.0269	0.0733	0.0155

Table 4: Estimated Biases and MSEs of the estimators under consideration. $10,000$ random samples were drawn from $0.9N(0,1) + 0.1\Delta_{10}$ with sample size $n = 20, 40, 100$.

Estimator	$n = 20$		$n = 40$		$n = 100$	
	Bias	MSE	Bias	MSE	Bias	MSE
Mean	0.9998	1.0443	0.9974	1.0170	0.9995	1.0079
SD	2.1381	4.5783	2.1428	4.5949	2.1449	4.6019
AD	1.4291	2.0561	1.4365	2.0705	1.4409	2.0791
Median	0.1366	0.1000	0.1355	0.0604	0.1393	0.0368
MAD	0.1011	0.0929	0.1201	0.0561	0.1368	0.0361
IQR	0.0695	0.0786	0.1074	0.0515	0.1371	0.0362
Huber ($\kappa = 1.0$) with median						
location	0.1366	0.1000	0.1355	0.0604	0.1393	0.0368
scale	0.0920	0.0749	0.1203	0.0475	0.1408	0.0335
Huber ($\kappa = 1.0$) Proposal 2						
location	0.1712	0.0884	0.1684	0.0581	0.1710	0.0413
scale	0.1925	0.1085	0.1726	0.0640	0.1648	0.0410
QIF ($\tau = 1.0$) with median						
location	0.1366	0.1000	0.1355	0.0604	0.1393	0.0368
scale	-0.0275	0.0408	-0.0077	0.0198	-0.0056	0.0081
QIF ($\tau = 1.0$)						
location	0.0278	0.0761	0.0129	0.0336	0.0081	0.0121
scale	-0.0740	0.0568	-0.0358	0.0244	-0.0127	0.0085

5.2 Hypothesis test

Consider the test of the equality of variances in a one-way analysis of variance with k samples. This problem has a long-standing history and there are many possible tests. Conover *et al.* (1981) examine many of the existing parametric and nonparametric tests by extensive simulations and they conclude that the tests proposed by Fligner and Killeen (1976) and Levene (1960) appear to be superior in terms of robustness of departures from normality and power. The proposed test in Section 4 is compared with various well-known robust tests. These tests and their asymptotic distributions are presented.

- F Test: Let $\mathbf{X} = (X_1, \dots, X_{n_1})$ and $\mathbf{Y} = (Y_1, \dots, Y_{n_1})$ be independent samples from two populations. The null hypothesis of equal variances, $H_0 : \sigma_1^2 = \sigma_2^2$, is to be tested at significance level α against the alternative hypothesis, $H_1 : \sigma_1^2 \neq \sigma_2^2$. Then the classical F test statistic is

$$F = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2/(n_1 - 1)}{\sum_{j=1}^{n_2}(Y_j - \bar{Y})^2/(n_2 - 1)}.$$

  The critical values of F are taken from the Snedecor F-table with $n_1 - 1$ and $n_2 - 1$ degrees of freedom for the $\alpha/2$ and $1 - \alpha/2$ quantiles. When samples are from the normal distribution, the distribution of such a statistic is the F-distribution and this test is the uniformly most powerful (UMP) unbiased test (see Lehmann, 1986, pg. 199). This test is known to be very sensitive to departures from normality.

- Fligner-Killeen Test: This test was originally suggested by Fligner and Killeen (1976) and modified by Conover *et al.* (1981). Suppose we have k independent samples, $X_{i1}, \dots, X_{in_i}$, $i = 1, \dots, k$. Suppose the X_{ij}, $j = 1, \dots, n_i$, are independent and identically distributed with the density $f((x_i - \mu_i)/\sigma_i)$. The density $f(\cdot)$ and the location and scale parameters are unknown. We wish to test the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ versus H_1 : not all the σ_i^2 are equal. The modified version is to make use of the ranks of $X_{ij}^* = |X_{ij} - \tilde{X}_i|$ with $\tilde{X}_i = \text{median}_j(X_{ij})$ and assign increasing normal scores

$$a_n(i) = \Phi^{-1}\Big(\frac{i}{2(n+1)} + \frac{1}{2}\Big), \qquad n = \sum_{i=1}^{k} n_i,$$

  where $\Phi(\cdot)$ is the normal distribution function. The χ^2 test statistic is given by

$$X^2 = \frac{1}{V^2}\sum_{i=1}^{k} n_i(\bar{A}_i - \bar{a}_n)^2,$$

  with

$$\bar{a}_n = \frac{1}{n}\sum_{i=1}^{n} a_n(i),\ V^2 = \frac{1}{n-1}\sum_{i=1}^{n}(a_n(i) - \bar{a}_n)^2,\ \bar{A}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} a_n(R(X_{ij}^*)),$$

where the ranking $R(\cdot)$ is performed on the folded observations. For large samples, the statistic X^2 is asymptotically distributed as a χ^2 with $k-1$ degrees of freedom under the null. Another alternative F test statistic is

$$F = \frac{X^2/(k-1)}{(n-1-X^2)/(n-k)},$$

which is the F distribution with $k-1$ and $n-k$ degrees of freedom. We shall call these tests the Fligner-Killeen χ^2 and F tests.

- Levene's Test: Levene (1960) suggested using the one-way analysis of variance of the variables $Z_{ij} = |X_{ij} - \bar{X}_{i\cdot}|$ with $\bar{X}_{i\cdot} = \frac{1}{n_i}\sum_{k=1}^{n_i} X_{ij}$ as a method of incorporating the robustness of that test into a test for variance. The test statistics is given by

$$W = \frac{\displaystyle\sum_{i=1}^{k} n_i(\bar{Z}_{i\cdot} - \bar{Z}_{\cdot\cdot})^2/(k-1)}{\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_{i\cdot})^2/(n-k)},$$

which has an asymptotic F-distribution with $k-1$ and $n-k$ degrees of freedom. Brown and Forsythe (1974) improved this test by replacing the mean estimator $\bar{X}_{i\cdot}$ with more robust estimator, say, the median. We shall call these tests Levene (mean) and Levene (median) tests.

- Layard Test: Layard (1973) suggests a χ^2 test statistic which is a function of the kurtosis. It is given by

$$S = \frac{\sum_{i=1}^{k}(n_i-1)\Big[\log S_i^2 - \frac{1}{n-k}\sum_{r=1}^{k}(n_r-1)\log S_r^2\Big]^2}{2+(1-\frac{k}{n})\hat{\gamma}},$$

with

$$S_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2,\ i = 1,\ldots,k \quad \text{and} \quad \hat{\gamma} = \frac{n\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i\cdot})^4}{\left(\sum_{i=1}^{k}(n_i-1)S_i^2\right)^2} - 3.$$

Then S has an asymptotic χ^2 distribution with $k-1$ degrees of freedom.

Tables 5–9 present the observed levels of various tests for the hypothesis $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 \neq \sigma_2$ based on 10,000 random samples from the different distributions. The results at the 5% and 10% levels of significance are reported. Assuming binomial rejection frequencies, we can estimate the standard error for each observed level $\hat{\alpha}$ by $\left[\hat{\alpha}(1-\hat{\alpha})/10000\right]^{1/2}$ which is at most $\left(0.5\times 0.5/10000\right)^{1/2} = 0.005$.

In Table 5, the sample was generated from $N(0,1)$. The F test (the UMP unbiased test in the normal model) performs the best. The results for $N(0,1)$ show that the performance of the Fligner-Killeen (χ^2 and F), Layard, and Levene (mean and median) tests is similar and the Levene (median) is conservative for smaller sample sizes. The QIF test overestimates the significance level for smaller sample sizes.

Next, we investigate the robustness property for departures from normality. In Tables 6, 7 and 8, the data comes from non-normal distributions without contamination: the double exponential $DE(0,1)$, the logistic $L(0,1)$ and Student's t_3, respectively. As the results show, the F test performs poorly for non-normal distributions. The Fligner-Killeen tests seems to be the best test. The performance of the other tests is similar and the QIF test still overestimates for smaller sample sizes. This is due to the slow convergence of the QIF test statistic. It indicates that future research should include the improvement of the convergence of the QIF.

Finally, we look at the robustness property for *contaminated* samples. In Table 9, the first sample was drawn from $N(0,1)$ and the second sample from the mixture $0.9N(0,1)+0.1\Delta_{10}$. The the second component of the mixture is considered to be the contamination and the scale parameter of the first component is to be tested. The F test performs extremely poorly under this hypothesis For smaller sample sizes, the Fligner-Killeen and QIF tests perform similarly. As the sample size gets larger, the observed level of the Fligner-Killeen test begins to depart from the significance level, while that of the QIF test maintains its level. This result indicates that the Fligner-Killeen tests can be used only for small sample sizes in the contaminated case. To make these observations graphically explicit, we choose $n_1 = n_2 = n$ and plot the observed levels of the tests in Figures 1 and 2. The Fligner-Killeen (χ^2 and F) tests perform very similarly. Therefore the power curve of the Fligner-Killeen F test is not included in the plot. Insensitivity of the observed level to contamination is not enough to justify the robustness property of the tests. To be convinced of the robustness property, one has to look at the power. In Figures 3 and 4, we estimate the power of various tests of the significance level $\alpha = 10\%$ with and without contamination for $10,000$ samples. The first sample of size 40 was generated from $N(0,\rho)$ where ρ ranges from 1 to 4 and the second sample of size 40 was generated from $N(0,1)$. In the case of contamination, the second sample was generated from $0.95N(0,1) + 0.05\Delta_{3.72}$. Figure 3 shows that the QIF test slightly overestimates the significance level at $\rho = 1$, but overall it is comparable to the other tests. In Figure 4, it is clearly seen that the estimated power curve of the QIF test is the closest to that of the F test in the uncontaminated case. Note that we redraw the estimated power curves of the F test in the *uncontaminated* case as a reference. This also shows that the QIF test is more powerful than any other test, while the other tests are biased over some regions. In conclusion, the Fligner-Killeen tests seem to be the best test when departures from normality are anticipated, but when the data is contaminated, the QIF test is the best test.

Table 5: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.1,\ 0.05$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50, 100$ and $n_2 = 25, 50, 100$. Both populations are $N(0,1)$.

Sample size (n_1, n_2)	$(25, 25)$	$(25, 50)$	$(50, 50)$	$(50, 100)$	$(100, 100)$
		Significance Level $\alpha = 0.1$			
F-test	0.0963	0.1017	0.1038	0.0967	0.1095
Fligner-Killeen χ^2	0.0970	0.1001	0.0989	0.0963	0.1086
Fligner-Killeen F	0.0962	0.0998	0.0986	0.0962	0.1085
Layard	0.1149	0.1166	0.1092	0.1051	0.1144
Levene (mean)	0.1084	0.1133	0.1070	0.0981	0.1111
Levene (median)	0.0881	0.1005	0.0979	0.0924	0.1050
QIF$(\tau = 1)$	0.1379	0.1545	0.1340	0.1351	0.1271
		Significance Level $\alpha = 0.05$			
F-test	0.0489	0.0519	0.0512	0.0481	0.0555
Fligner-Killeen χ^2	0.0437	0.0503	0.0482	0.0439	0.0554
Fligner-Killeen F	0.0445	0.0513	0.0486	0.0440	0.0557
Layard	0.0586	0.0590	0.0576	0.0510	0.0583
Levene (mean)	0.0541	0.0584	0.0550	0.0466	0.0571
Levene (median)	0.0392	0.0495	0.0472	0.0438	0.0524
QIF$(\tau = 1)$	0.0662	0.0874	0.0689	0.0750	0.0710

Table 6: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.1,\ 0.05$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50, 100$ and $n_2 = 25, 50, 100$. Both populations are $DE(0, 1)$.

Sample size (n_1, n_2)	$(25, 25)$	$(25, 50)$	$(50, 50)$	$(50, 100)$	$(100, 100)$
	Significance Level $\alpha = 0.1$				
F-test	0.2673	0.2811	0.2894	0.2859	0.2882
Fligner-Killeen χ^2	0.0969	0.1016	0.1056	0.0999	0.0991
Fligner-Killeen F	0.0966	0.1013	0.1054	0.0999	0.0990
Layard	0.1325	0.1287	0.1205	0.1144	0.1100
Levene (mean)	0.1150	0.1128	0.1128	0.1060	0.1044
Levene (median)	0.0946	0.0999	0.1021	0.0995	0.0981
QIF($\tau = 1$)	0.1703	0.1724	0.1626	0.1563	0.1517
	Significance Level $\alpha = 0.05$				
F-test	0.1881	0.1966	0.2047	0.2025	0.2056
Fligner-Killeen χ^2	0.0480	0.0456	0.0510	0.0502	0.0476
Fligner-Killeen F	0.0494	0.0464	0.0518	0.0505	0.0479
Layard	0.0704	0.0631	0.0601	0.0577	0.0563
Levene (mean)	0.0601	0.0563	0.0591	0.0565	0.0516
Levene (median)	0.0463	0.0468	0.0494	0.0520	0.0490
QIF($\tau = 1$)	0.0923	0.0935	0.0944	0.0887	0.0824

Table 7: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.1,\ 0.05$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50, 100$ and $n_2 = 25, 50, 100$. Both populations are $L(0, 1)$.

Sample size (n_1, n_2)	$(25, 25)$	$(25, 50)$	$(50, 50)$	$(50, 100)$	$(100, 100)$
		Significance Level $\alpha = 0.1$			
F-test	0.1761	0.1845	0.1869	0.1879	0.1860
Fligner-Killeen χ^2	0.0966	0.0983	0.1042	0.0980	0.0989
Fligner-Killeen F	0.0960	0.0981	0.1040	0.0979	0.0989
Layard	0.1215	0.1198	0.1138	0.1115	0.1075
Levene (mean)	0.1069	0.1085	0.1107	0.1054	0.0991
Levene (median)	0.0900	0.0953	0.1021	0.0979	0.0934
QIF($\tau = 1$)	0.1454	0.1559	0.1326	0.1335	0.1151
		Significance Level $\alpha = 0.05$			
F-test	0.1065	0.1135	0.1190	0.1172	0.1159
Fligner-Killeen χ^2	0.0475	0.0448	0.0505	0.0504	0.0474
Fligner-Killeen F	0.0481	0.0457	0.0509	0.0510	0.0478
Layard	0.0630	0.0587	0.0583	0.0560	0.0527
Levene (mean)	0.0552	0.0538	0.0556	0.0549	0.0499
Levene (median)	0.0429	0.0452	0.0472	0.0517	0.0466
QIF($\tau = 1$)	0.0731	0.0874	0.0723	0.0703	0.0579

Table 8: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.1,\ 0.05$. $10,000$ random samples were drawn with sample size $n_1 = 25, 50, 100$ and $n_2 = 25, 50, 100$. Both populations are Student's t_3.

Sample size (n_1, n_2)	$(25, 25)$	$(25, 50)$	$(50, 50)$	$(50, 100)$	$(100, 100)$
		Significance Level $\alpha = 0.1$			
F-test	0.3892	0.4077	0.4553	0.4725	0.4951
Fligner-Killeen χ^2	0.1005	0.0972	0.0991	0.0984	0.0973
Fligner-Killeen F	0.0998	0.0970	0.0990	0.0983	0.0970
Layard	0.1450	0.1204	0.1185	0.1043	0.1016
Levene (mean)	0.1191	0.1160	0.1103	0.1081	0.1045
Levene (median)	0.0881	0.0905	0.0937	0.0957	0.0953
QIF($\tau = 1$)	0.1447	0.1505	0.1254	0.1290	0.1209
		Significance Level $\alpha = 0.05$			
F-test	0.3083	0.3241	0.3716	0.3939	0.4190
Fligner-Killeen χ^2	0.0461	0.0440	0.0467	0.0466	0.0464
Fligner-Killeen F	0.0472	0.0444	0.0473	0.0472	0.0467
Layard	0.0699	0.0535	0.0480	0.0463	0.0420
Levene (mean)	0.0570	0.0571	0.0521	0.0518	0.0500
Levene (median)	0.0377	0.0406	0.0435	0.0447	0.0448
QIF($\tau = 1$)	0.0753	0.0763	0.0637	0.0686	0.0613

Table 9: Levels of the two-sample hypothesis tests under consideration with nominal level $\alpha = 0.1,\ 0.05$. $10,000$ random samples were drawn with sample size $n_1 = 20, 40, 80$ and $n_2 = 20, 40, 80$. Population 1 is the normal $N(0,1)$ and Population 2 is $0.9N(0,1) + 0.1\Delta_{10}$.

Sample size (n_1, n_2)	$(20, 20)$	$(20, 40)$	$(40, 20)$	$(40, 40)$	$(80, 80)$
	Significance Level $\alpha = 0.1$				
F-test	1.0000	1.0000	1.0000	1.0000	1.0000
Fligner-Killeen χ^2	0.1338	0.1596	0.2411	0.3178	0.6790
Fligner-Killeen F	0.1330	0.1588	0.2404	0.3171	0.6784
Layard	0.9976	0.9999	1.0000	1.0000	1.0000
Levene (mean)	0.8585	0.9145	0.9998	1.0000	1.0000
Levene (median)	0.1675	0.1968	0.7548	0.8720	0.9999
QIF($\tau = 1$)	0.1094	0.1561	0.0983	0.1145	0.1182
	Significance Level $\alpha = 0.05$				
F-test	1.0000	1.0000	1.0000	1.0000	1.0000
Fligner-Killeen χ^2	0.0556	0.0593	0.1294	0.1698	0.4906
Fligner-Killeen F	0.0578	0.0610	0.1319	0.1723	0.4922
Layard	0.9156	0.9987	0.9937	1.0000	1.0000
Levene (mean)	0.4652	0.5754	0.9901	0.9984	1.0000
Levene (median)	0.0207	0.0175	0.4027	0.5157	0.9983
QIF($\tau = 1$)	0.0554	0.0917	0.0457	0.0610	0.0642

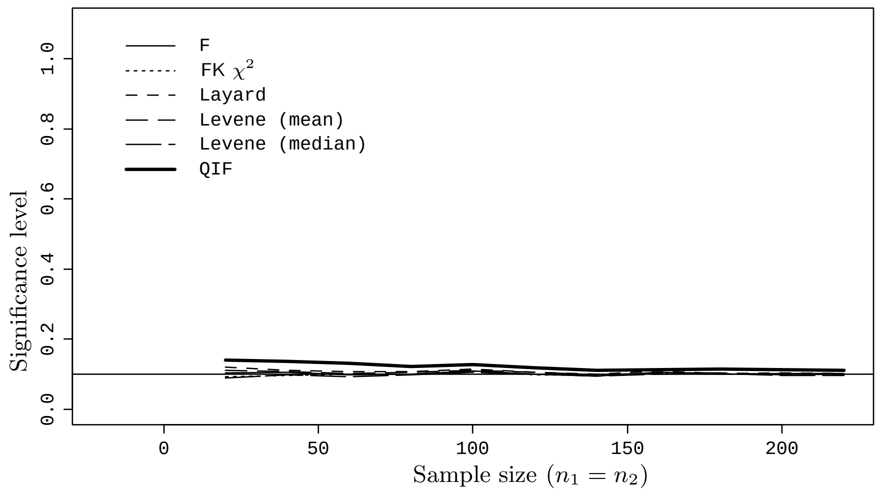

Figure 1: Convergence of the tests under consideration testing $H_0 : \sigma_1^2/\sigma_2^2 = 1$ versus $H_1 : \sigma_1^2/\sigma_2^2 \neq 1$ with level $\alpha = 0.1$. Both random samples were drawn from $N(0, 1)$ with sample size $n_1 = n_2$.

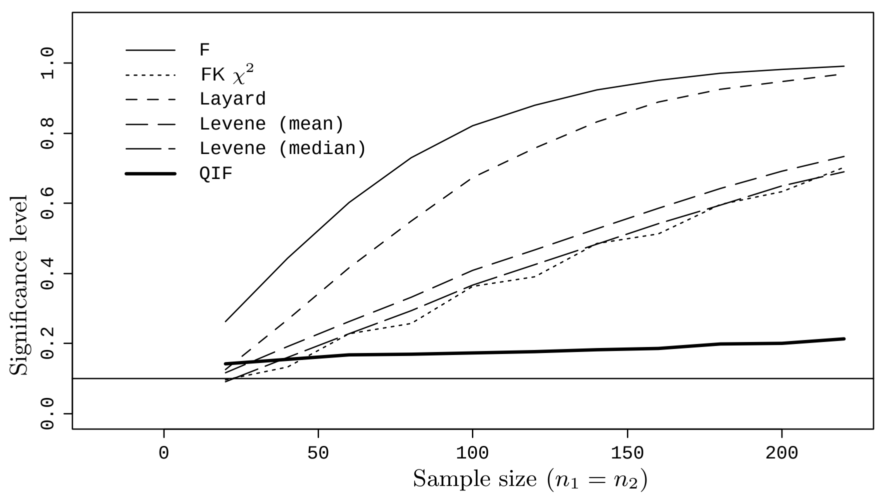

Figure 2: Convergence of the tests under consideration testing $H_0 : \sigma_1^2/\sigma_2^2 = 1$ versus $H_1 : \sigma_1^2/\sigma_2^2 \neq 1$ with level $\alpha = 0.1$. The first $10,000$ random samples were drawn from $N(0, 1)$ and the second $10,000$ random samples from $0.95N(0, 1) + 0.05\Delta_{3.72}$ with sample size $n_1 = n_2$.

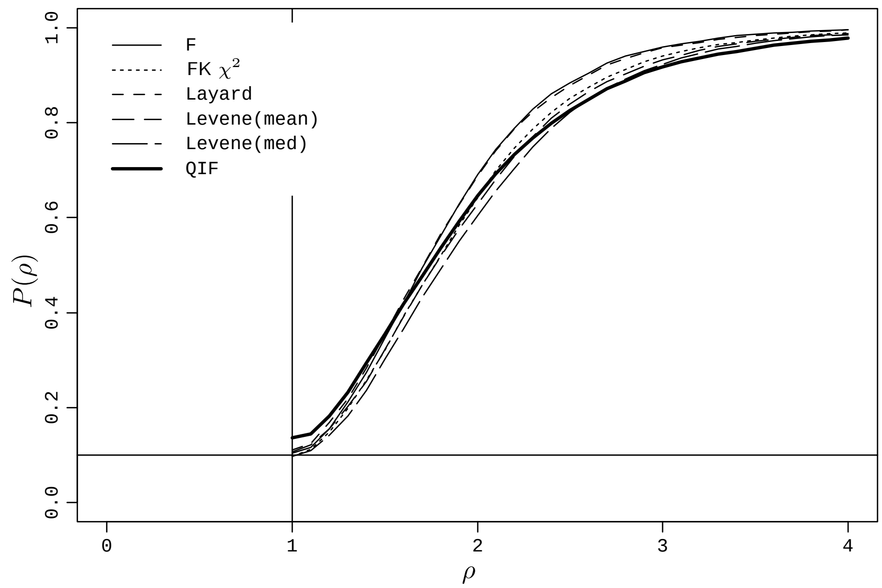

Figure 3: Estimated powers for the tests under consideration testing $H_0 : \sigma_1^2/\sigma_2^2 = 1$ versus $H_1 : \sigma_1^2/\sigma_2^2 \neq 1$ with level $\alpha = 0.1$. The first $10,000$ random samples were drawn from $N(0, \rho)$ with sample size $n_1 = 40$ and the second $10,000$ random samples from $N(0, 1)$ with sample size $n_2 = 40$.

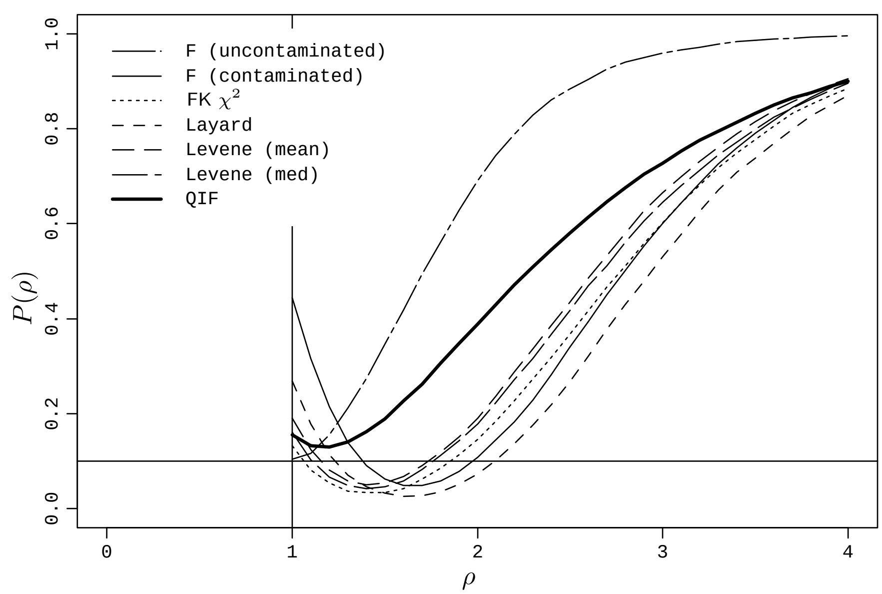

Figure 4: Estimated powers for the tests under consideration testing $H_0 : \sigma_1^2/\sigma_2^2 = 1$ versus $H_1 : \sigma_1^2/\sigma_2^2 \neq 1$ with level $\alpha = 0.1$. The first $10,000$ random samples were drawn from $N(0, \rho)$ with sample size $n_1 = 40$ and the second $10,000$ random samples from $0.95N(0, 1)+0.05\Delta_{3.72}$ with sample size $n_2 = 40$.

References

Brown, M. B. and Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, **69**(346), 364–367.

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, **23**(4), 351–361.

Fligner, M. A. and Killeen, T. J. (1976). Distribution-free two-sample tests for scale. *Journal of the American Statistical Association*, **71**(353), 210–213.

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

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101.

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

Layard, M. W. J. (1973). Robust large-sample tests for homogeneity of variances. *Journal of the American Statistical Association*, **68**(341), 195–198.

Lehmann, E. L. (1986). *Testing Statistical Hypothesis*. John Wiley & Sons.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, editor, *Contributions to Probability and Statistics*, pages 278–292. Stanford University Press, Palo Alto, CA.

Park, C. and Lindsay, B. G. (1999). Robust estimation and tests based on quadratic inference function. Technical Report #99-02, Center for Likelihood Studies, Pennsylvania State University.