

## ROBUST HYPOTHESIS TESTING VIA $L_q$ -LIKELIHOOD

Yichen Qin and Carey E. Priebe

*University of Cincinnati and Johns Hopkins University*

*Abstract:* This article introduces a robust hypothesis testing procedure: the  $L_q$ -likelihood-ratio-type test (LqRT). By deriving the asymptotic distribution of the test statistic, we demonstrate its robustness analytically and numerically, and investigate the properties of its influence function and breakdown point. A proposed method to select the tuning parameter  $q$  offers a good efficiency/robustness trade-off compared with the traditional likelihood ratio test (LRT) and other robust tests. Simulation and a real data analysis provide further evidence of the advantages of the proposed LqRT method. In particular, for the special case of testing the location parameter in the presence of gross error contamination, the LqRT dominates the Wilcoxon-Mann-Whitney test and the sign test at various levels of contamination.

*Key words and phrases:* Gross error model, relative efficiency, robustness.

### 1. Introduction

The likelihood ratio test (LRT) is frequently used, but its robustness is not ideal. The LRT can achieve optimal performance only under several strict assumptions, whereas its performance degrades significantly in the presence of even mild violations of them. Here we propose a robust testing procedure, the  $L_q$ -likelihood-ratio-type test (LqRT), that leverages the newly developed concept of  $L_q$ -likelihood (Ferrari and Yang (2010)).

Consider a gross error model  $h(x) = (1 - \varepsilon)f(x; \theta) + \varepsilon g(x)$ , where  $f$  is an “idealized” model with the parameter  $\theta$  that we need to test,  $g$  is the measurement error or contamination component, and  $\varepsilon$  is the contamination ratio.  $h$  represents the true data generating process, a deviation from  $f$  when  $\varepsilon > 0$ . For a data set generated from  $h$ ,  $100(1 - \varepsilon)\%$  are drawn from  $f$ , whereas the rest of the data points from  $g$  are considered measurement errors or outliers. Common choices for the contamination distribution  $g$  are a fat tailed distribution and a point mass distribution.

Suppose measurements  $X = (X_1, X_2, \dots, X_n)$  are generated by an experiment, where  $X$  follows a distribution  $f(x; \theta)$  with an interpretable parameter  $\theta$ , but that we observe, instead,  $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ . Here most of the  $X_i^* = X_i$ ,

but there are a few outliers due to human error, instrument malfunction, or the complexity of the underlying process. Under such circumstances, using data  $X^*$ , we still have  $\theta$  as the parameter for hypothesis testing or estimation (Bickel and Doksum (2007)). For this problem, we introduce the LqRT.

For robust testing, Huber (1965) suggested a censored likelihood ratio test (HLRT). Rousseeuw (1984) proposed a “least median of squares” approach and corresponding testing procedures. Heritier and Ronchetti (1994) proposed Wald-type, score-type, and likelihood-ratio-type tests, that are the natural counterparts of M-estimators. Cantoni and Ronchetti (2001) proposed a robust quasi-likelihood function for hypothesis testing for generalized linear models. Lô and Ronchetti (2009) extended this method with saddlepoint approximations to obtain both a robust test statistic for hypothesis testing and variable selection for generalized linear models. Basu et al. (2013) developed a class of tests using the density power divergence (DPD) based on Basu et al. (1998). Markatou, Basu and Lindsay (1998) proposed a weighted likelihood, and Agostinelli and Markatou (2001) offered a test based on this weighted likelihood. He, Simpson and Portnoy (1990) and He (1991) have studied and extended the concept of a breakdown point for robustness evaluation. Ronchetti (1997) and Medina and Ronchetti (2015) have provided selective reviews of some basic approaches to robust inference and of recent developments in robust statistics.

In our notation,  $n$  denotes the sample size, and  $d$  indicates the dimension of the observations. Bold-face lower-case letters are vectors, bold-face upper-case letters are matrices. We take  $p$  to be the dimension of the parameter space, and  $\theta$  is a  $p \times 1$  parameter vector, with  $\theta_0$  as the true parameter. Finally,  $f(x; \theta)$  is the assumed probability density function, and  $f'_\theta(x; \theta)$  and  $f''_\theta(x; \theta)$  are its first and second-order derivatives with respect to  $\theta$ .

In the next section, we introduce the Lq-likelihood and other preliminary concepts that lead to the introduction of the Lq-likelihood-ratio-type test (LqRT) in Section 3. We demonstrate its robustness properties through an analysis of the asymptotic distribution, the influence function, and the breakdown point; we also discuss related issues such as critical values. Numerical results are presented in Section 4. We discuss the selection of  $q$  in Section 5 and showcase the performance of our test. We have a discussion in Section 6, summarize the assumptions in the Appendix in Section 7, and relegate the proofs and additional simulation studies to online supplementary materials.

## 2. Preliminary Concepts

### 2.1. $L_q$ -likelihood and maximum $L_q$ -likelihood estimation

The likelihood function of the observed sample  $\mathbf{x} = (x_1, \dots, x_n)$  under the hypothesized model  $f(x; \theta)$  is  $\mathcal{L}(\mathbf{x}; \theta) = \prod_{i=1}^n f(x_i; \theta)$ , with  $\theta \in \Theta \subset \mathbb{R}^p$ . Ferrari and Yang (2010) introduced the  $L_q$ -likelihood,  $\sum_{i=1}^n L_q(f(x_i; \theta))$ , where  $L_q(u) = (u^{1-q} - 1)/(1 - q)$  for  $q \neq 1$ , and  $L_q(u) = \log u$  for  $q = 1$ . When  $q \rightarrow 1$ ,  $L_q(u) \rightarrow \log u$ . Throughout, we take  $0 < q \leq 1$ .

To estimate  $\theta$ , we can use the maximum  $L_q$ -likelihood estimation (MLqE)  $\tilde{\theta}_q = \arg \max_{\theta \in \Theta} \sum_{i=1}^n L_q(f(x_i; \theta))$ . To obtain  $\tilde{\theta}_q$ , we solve the  $L_q$ -likelihood equation,  $\sum_{i=1}^n [f'_\theta(x_i; \theta)/f(x_i; \theta)] f(x_i; \theta)^{1-q} = 0$ , a weighted version of the likelihood equation, with the weights given by  $f(x_i; \theta)^{1-q}$ . When  $q < 1$ , data points with high likelihoods are assigned large weights. Outliers are usually assigned small weights because of their low likelihoods, promoting MLqE robustness. As  $q \rightarrow 1$ , the MLqE becomes the maximum likelihood estimation (MLE).

The robustness added by the  $L_q$ -likelihood results because the  $L_q(\cdot)$  function is bounded from below for  $0 < q < 1$ . As  $L_q(u) \geq -1/(1 - q)$  whereas  $\log(x) \rightarrow -\infty$  when  $x \rightarrow 0^+$ , one limits the effect of particular data points on the quantity  $\sum_{i=1}^n L_q(f(x_i; \theta))$ .

### 2.2. Consistency and bias correction

The MLqE, while robust, is consistent only for a few special cases (e.g., estimation of the location parameter of a symmetric distribution). We consider two approaches to correct inherent bias.

Consider a sequence  $q_n$ ,  $q_n \rightarrow 1$  as  $n \rightarrow \infty$ , such that  $\tilde{\theta}_{q_n} \xrightarrow{P} \theta_0$ . Ferrari and Yang (2010) and Ferrari and La Vecchia (2012) offered a detailed discussion of this case in which the MLqE loses robustness as  $q_n$  tends to 1.

Consider then a fixed  $q$  and subtract a bias correction term from the  $L_q$ -likelihood function; this is equivalent to re-centering the estimation equation. Here take the bias correction term:  $C(\theta, q) = \int f(x; \theta)^{2-q}/(2 - q) dx$ . The bias-corrected maximum  $L_q$ -likelihood estimation (BCMLqE) is

$$\hat{\theta}_q = \arg \max_{\theta \in \Theta} \sum_{i=1}^n [L_q(f(x_i; \theta)) - C(\theta, q)].$$

The online supplementary materials contain a simple proof for the consistency of the BCMLqE. When  $q = 1$  or  $\theta$  is a location parameter,  $C(\theta, q)$  is a constant independent of  $\theta$ , and the BCMLqE is the MLqE. We adopt this bias-correction

approach and extend it in the next section to support our proposed test statistic.

The BCMLqE is the minimum density power divergence estimator (MDPD) proposed by Basu et al. (1998). Maximizing the bias-corrected Lq-likelihood function is equivalent to minimizing the density power divergence between the empirical distribution and the parametric distribution. While we are revising our paper, we became aware of another test proposed by Basu et al. (2013), a distance-based test using the density power divergence between the estimated model and the hypothesized model. The robustness of their test statistic is linked directly to the robustness of their parameter estimate MDPD, while our test statistic depends on the data through the likelihood functions, using information more thoroughly.

### 3. Lq-Likelihood-Ratio-Type Test

#### 3.1. Test statistic

Suppose we have a sample  $(x_1, \dots, x_n)$  and an assumed parametric model  $f(x; \theta)$  with parameter  $\theta \in \Theta \subset \mathbb{R}^p$ . We are interested in testing  $H_0 : \theta \in \Theta_0$  against  $H_1 : \theta \in \Theta_1$ . We take the Lq-likelihood-ratio-type test (LqRT) as

$$D_q(\mathbf{x}) = 2 \sup_{\theta \in \Theta_0 \cup \Theta_1} \left\{ \sum_{i=1}^n [L_q(f(x_i; \theta)) - C(\theta, q)] \right\} - 2 \sup_{\theta \in \Theta_0} \left\{ \sum_{i=1}^n [L_q(f(x_i; \theta)) - C(\theta, q)] \right\}, \quad (3.1)$$

where  $q$  is a tuning parameter. We reject the null hypothesis when  $D_q$  is large. This test is a member of the class of likelihood-ratio-type tests as defined in Heritier and Ronchetti (1994), obtained by choosing  $\rho(z; \theta) = -L_q(f(z; \theta)) + C(\theta, q)$ . Therefore, many tools provided by Heritier and Ronchetti (1994) are available for the LqRT.

To derive the asymptotic distribution of our test statistic, we partition the parameter as  $\theta = (\alpha, \beta)$ , where  $\alpha \in \mathbb{R}^r$  and  $\beta \in \mathbb{R}^{p-r}$ , then simplify the null and alternative hypotheses to  $H_0 : \alpha = \mathbf{0}$  and  $H_1 : \alpha \neq \mathbf{0}$ . The test statistic is  $D_q(\mathbf{x}) = 2 \sup_{\alpha, \beta} \sum_{i=1}^n [L_{q_n}(f(x_i; (\alpha, \beta))) - C((\alpha, \beta), q)] - 2 \sup_{\beta} \sum_{i=1}^n [L_{q_n}(f(x_i; (\mathbf{0}, \beta))) - C((\mathbf{0}, \beta), q)]$ . The necessary notation and assumption are as follows.

Let  $\psi(x; \theta, q) = \partial L_q(f(x; \theta)) / \partial \theta$ ,  $\psi'(x; \theta, q) = \partial^2 L_q(f(x; \theta)) / \partial \theta^2$ ,  $c(\theta, q) = \partial C(\theta, q) / \partial \theta$ , and  $c'(\theta, q) = \partial^2 C(\theta, q) / \partial \theta^2$ , where  $C(\theta, q)$  is the bias correction term. Take  $\tilde{\psi}(X; \theta, q) = \psi(X; \theta, q) - c(\theta, q)$ ,  $\tilde{\psi}'(X; \theta, q) = \psi'(X; \theta, q) - c'(\theta, q)$ ,  $\mathbf{A} = \mathbb{E}[\tilde{\psi}(X; \theta_0, q) \tilde{\psi}(X; \theta_0, q)^T]$ ,

$$\mathbf{B} = -\mathbb{E}[\tilde{\psi}'(X; \theta_0, q)] = \begin{pmatrix} \mathbf{B}_{\alpha\alpha} & \mathbf{B}_{\alpha\beta} \\ \mathbf{B}_{\beta\alpha} & \mathbf{B}_{\beta\beta} \end{pmatrix}, \text{ and } \mathbf{B}^* = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{B}_{\beta\beta}^{-1} \end{pmatrix},$$

where  $\psi$ ,  $\tilde{\psi}$ , and  $c$  are  $p \times 1$  vectors,  $\psi'$ ,  $\tilde{\psi}'$ ,  $c'$ ,  $\mathbf{A}$ , and  $\mathbf{B}$  are  $p \times p$  symmetric matrices. Denote the sorted eigenvalues of an  $r \times r$  matrix  $\mathbf{M}$  by  $\lambda_j(\mathbf{M})$  for  $j = 1, \dots, r$ , with  $\lambda_1(\mathbf{M}) \geq \dots \geq \lambda_r(\mathbf{M})$ .

**Assumption 1.**  *$f$  satisfies the regularity conditions specified in the Appendix (Section 7).*

**Theorem 1.** *Under Assumption 1 and a correctly specified model  $f$ , for a fixed  $q$ , the asymptotic distribution of  $D_q(\mathbf{x})$  under the null hypothesis is  $\sum_{j=1}^r \lambda_j(\mathbf{A}[\mathbf{B}^{-1} - \mathbf{B}^*])\chi_{1,j}^2$ , where the  $\chi_{1,j}^2$  are i.i.d. chi-square random variables with 1 degree of freedom, and the  $\lambda_j(\mathbf{A}[\mathbf{B}^{-1} - \mathbf{B}^*])$  are  $r$  positive eigenvalues of  $\mathbf{A}[\mathbf{B}^{-1} - \mathbf{B}^*]$ .*

The proof of this result can be obtained directly as a special case of Proposition 3a in Heritier and Ronchetti (1994), when  $\rho(z; \theta) = -L_q(f(z; \theta)) + C(\theta, q)$ . When  $q = 1$ , we have  $\mathbf{A} = \mathbf{B}$ , and then the LqRT is the LRT, which is chi-square with  $r$  degrees of freedom.

### 3.2. Robust properties of LqRT

Suppose data are generated from a gross error model  $h = (1 - \varepsilon)f + \varepsilon g$ , where  $f$  is the assumed model and  $g$  is the contamination component. For simplicity, we take  $r = p$  restricting to a simple null hypothesis.

Let  $\mathbf{A}_{\varepsilon,q} = \mathbb{E}_h[\tilde{\psi}(X; \theta_{\varepsilon,q}^*, q)\tilde{\psi}(X; \theta_{\varepsilon,q}^*, q)^T]$ , and  $\mathbf{B}_{\varepsilon,q} = -\mathbb{E}_h[\tilde{\psi}'(X; \theta_{\varepsilon,q}^*, q)]$ , where  $\mathbf{A}_{\varepsilon,q}$  and  $\mathbf{B}_{\varepsilon,q}$  are  $p \times p$  symmetric matrices, and  $\theta_{\varepsilon,q}^* = \arg \max_{\theta} \mathbb{E}_h[L_q(f(X; \theta)) - C(\theta, q)]$ . Here,  $\theta_{\varepsilon,q}^*$  represents the parameter to which the BCMLqE converges under the gross error model  $h$ . In addition,  $\theta_{0,q}^* = \theta_0$  for  $0 < q \leq 1$ .

**Assumption 2.** *For any  $\varepsilon \in (0, 1)$ ,  $h$  is such that  $\mathbb{E}_h[f''_{\theta}(X; \theta_{\varepsilon,1}^*)/f(X; \theta_{\varepsilon,1}^*)]$  is positive definite, where  $\theta_{\varepsilon,1}^* = \arg \max_{\theta} \mathbb{E}_h[\log f(X; \theta)]$ .*

**Assumption 3.**  *$\mathbb{E}_f[f'_{\theta}(X; \theta_0)f'_{\theta}(X; \theta_0)^T(f(X; \theta_0)^{-2q} - f(X; \theta_0)^{-q-1})]$  is negative definite for any  $q \in (0, 1)$ .*

**Assumption 4.** *There exists a constant  $q^{**} \in (0, 1)$ , such that the  $\lambda_j(\mathbf{A}_{\varepsilon,q}\mathbf{B}_{\varepsilon,q}^{-1})$  are monotonic functions in  $q$  on  $(q^{**}, 1)$ .*

A detailed discussion of these assumptions can be found in the online supplementary materials and, in particular, what they mean for the exponential family in general, and for normal distributions in particular.

**Theorem 2.** Under Assumption 1 and a misspecified model  $h$ , for a fixed  $q$ , the asymptotic distribution of  $D_q(\mathbf{x})$  under the null hypothesis is  $\sum_{j=1}^r \lambda_j(\mathbf{A}_{\varepsilon,q} \mathbf{B}_{\varepsilon,q}^{-1}) \chi_{1,j}^2$ , where the  $\chi_{1,j}^2$  are i.i.d. chi-square random variables with 1 degree of freedom, and the  $\lambda_j(\mathbf{A}_{\varepsilon,q} \mathbf{B}_{\varepsilon,q}^{-1})$  are  $r$  positive eigenvalues of  $\mathbf{A}_{\varepsilon,q} \mathbf{B}_{\varepsilon,q}^{-1}$ .

When  $\varepsilon = 0$  and  $q = 1$ ,  $\lambda_j(\mathbf{A}_{\varepsilon,q} \mathbf{B}_{\varepsilon,q}^{-1}) = 1$ , and the LqRT is chi-square with  $r$  degrees of freedom.

**Theorem 3.** Under Assumptions 1 and 2 and a misspecified model  $h$ , for any  $\varepsilon \in (0, 1)$  and for  $q = 1$ ,  $\lambda_j(\mathbf{A}_{\varepsilon,1} \mathbf{B}_{\varepsilon,1}^{-1}) > 1$  for  $j = 1, \dots, r$ .

Theorem 3 has that, when contamination occurs in the data and  $q = 1$ , the divergence between  $\mathbf{A}_{\varepsilon,1}$  and  $\mathbf{B}_{\varepsilon,1}$  increases, so  $\lambda_j(\mathbf{A}_{\varepsilon,1} \mathbf{B}_{\varepsilon,1}^{-1})$  increases away from 1, causing inflation in the asymptotic distribution; the original chi-square distribution with  $r$  degrees of freedom is a sum of the  $r$  inflated chi-square distributions with 1 degree of freedom, with inflations captured by  $\lambda_j(\mathbf{A}_{\varepsilon,1} \mathbf{B}_{\varepsilon,1}^{-1})$ ,  $j = 1, \dots, r$ . Thus,  $D_q$  is an “inflated” chi-square under the null hypothesis and, under the alternative hypothesis, an “inflated” non-central chi-square distribution. The overlap between the null and alternative distributions grows larger in  $\varepsilon$ , and the power of the test degrades (see Figure 2 for illustration). To limit or control this degradation of power, we need to control for the inflation of the asymptotic distribution.

**Theorem 4.** Under Assumptions 1, 2, 3, and 4, and a misspecified model  $h$ , there exists an  $\tilde{\varepsilon} \in (0, 1)$  such that, for any  $\varepsilon \in (0, \tilde{\varepsilon})$ , there exists a  $q^* \in (q^{**}, 1)$  and for any  $q \in (q^*, 1)$ ,

$$|\lambda_j(\mathbf{A}_{\varepsilon,q} \mathbf{B}_{\varepsilon,q}^{-1}) - 1| < |\lambda_j(\mathbf{A}_{\varepsilon,1} \mathbf{B}_{\varepsilon,1}^{-1}) - 1| \text{ for } j = 1, \dots, r.$$

Without Assumption 4, Theorem 4 would remain true, but only for a particular  $q$ . Theorem 4 implies that, by setting  $q < 1$ , we can shrink the eigenvalues  $\lambda_j(\mathbf{A}_{\varepsilon,q} \mathbf{B}_{\varepsilon,q}^{-1})$  back toward 1 and alleviate the inflation of distributions. More importantly, the effect of  $q < 1$  on  $\lambda_j(\mathbf{A}_{\varepsilon,q} \mathbf{B}_{\varepsilon,q}^{-1})$  can offset the inflation effect of  $\varepsilon > 0$  on  $\lambda_j(\mathbf{A}_{\varepsilon,q} \mathbf{B}_{\varepsilon,q}^{-1})$ . With this approach, we can avoid the increasing overlap between the null and alternative distributions and protect the power of the test better.

The theory offers two insights. First, our test statistic makes the approximation  $\mathbf{A}_{\varepsilon,q} \approx \mathbf{B}_{\varepsilon,q}$  more robust to model misspecification. By setting  $q < 1$ , many of the statistical inferences originally based on  $\mathbf{A}_{0,1} = \mathbf{B}_{0,1}$  can remain valid, even if the model is misspecified. Second, we gain a tool for identifying model

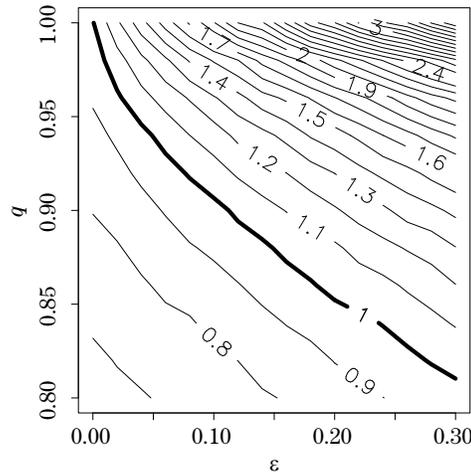


Figure 1. Contour plot of  $\lambda(\mathbf{A}_{\epsilon,q} \mathbf{B}_{\epsilon,q}^{-1})$  as a function of  $\epsilon$  and  $q$  where the assumed model is a standard normal distribution.

misspecification. Setting  $q < 1$  effectively eliminates the influence of outliers, and setting  $q > 1$  can magnify those effects. When  $q = 1$ ,  $\mathbf{A}_{\epsilon,1}$  is essentially Fisher’s information matrix; many model misspecification tests are based on  $\mathbf{A}_{\epsilon,1} = \mathbf{B}_{\epsilon,1}$  (e.g., White (1982)). The preceding results provide a possible approach to model misspecification detection.

### 3.3. Simulation study

In this section, we report on simulations to support our findings.

First we plot  $\lambda(\mathbf{A}_{\epsilon,q} \mathbf{B}_{\epsilon,q}^{-1})$  as a function of the contamination ratio  $\epsilon$  and the tuning parameter  $q$  for  $p = 1$  in Figure 1. Here  $f$  was a normal distribution and  $h(x) = (1 - \epsilon)\varphi(x; 0, 1) + \epsilon\varphi(x; 0, 10)$ . In Figure 1, we highlight the contour level of 1 in bold; it shows that  $\lambda(\mathbf{A}_{\epsilon,q} \mathbf{B}_{\epsilon,q}^{-1})$  increases as  $\epsilon$  increases when  $q = 1$ , but that we can find a value of  $\lambda(\mathbf{A}_{\epsilon,q} \mathbf{B}_{\epsilon,q}^{-1})$  closer to 1 by decreasing  $q$ .

We simulated the asymptotic null and alternative distributions under  $\epsilon = 0, 0.05$ , and  $0.1$ , and  $q = 1, 0.97$ , and  $0.8$ , for  $H_0 : \mu_f = 0$  and  $H_1 : \mu_f \neq 0$ . For this test,  $f$  was a three-dimensional normal distribution with known variance and the data generating process was  $h = (1 - \epsilon)f + \epsilon g$ , where  $g$  was multivariate normal distribution with  $\mu_g = \mu_f$  and  $\Sigma_g = 30\Sigma_f$ . We simulated the null and alternative distributions of  $D_q$  using  $\mu_f = (0, 0, 0)^T$  and  $\mu_f = (0.15, 0.15, 0.15)^T$  and present the results in Figure 2.

In this figure, when  $q = 1$  and  $\epsilon$  increases from 0 to 0.1, the null and alternative distributions become flatter and overlap more, which results in power

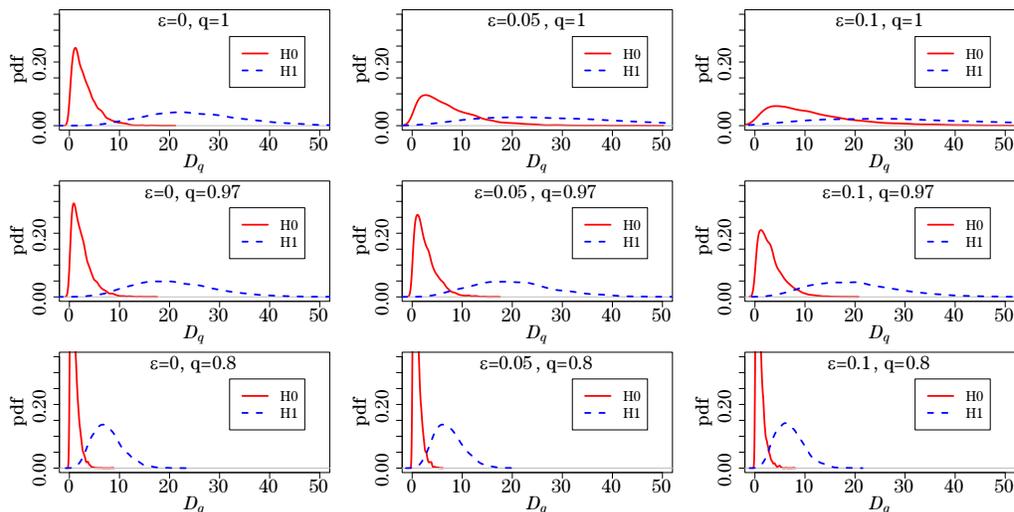


Figure 2. Comparison of asymptotic null and alternative distributions of the test statistic for testing the mean of the three-dimensional normal distribution under difference levels of symmetric heavy-tail contamination and different  $q$ 's.

degradation. When  $q = 0.97$ , instead of having the inflated chi-square distribution, the null and alternative distributions are less affected by the contamination, because  $\lambda_j(\mathbf{A}_{\varepsilon,q}\mathbf{B}_{\varepsilon,q}^{-1})$  gets pulled back toward 1 when we set  $q < 1$ . When  $q = 0.8$ , the distributions are much less affected. It is worth noting that, in the case of  $q = 0.8$  and  $\varepsilon = 0$  (lower left panel), the null and alternative distributions overlap more than they do in the case of  $q = 1$  and  $\varepsilon = 0$  (upper left panel), which means that by setting  $q < 1$ , we lose some of the test's power at zero contamination. As Figure 2 illustrates, we gain robustness from using the L $q$ -likelihood and make a trade-off for robustness by giving up power at zero contamination.

### 3.4. Critical values

Since we know the asymptotic null distribution of L $q$ RT from Theorem 2, we can calculate the  $1 - \alpha$  quantile of the null distribution to obtain the critical value for a level  $\alpha$  test. When  $\varepsilon = 0$ , we use the following algorithm to calculate the critical value:

Step 1: Calculate  $\mathbf{A}_{0,q}$  and  $\mathbf{B}_{0,q}$  under  $H_0$ .

Step 2: Calculate  $\lambda_j(\mathbf{A}_{0,q}\mathbf{B}_{0,q}^{-1})$ .

Step 3: Obtain the  $1 - \alpha$  quantile of the distribution of  $\sum_{j=1}^r \lambda_j(\mathbf{A}_{0,q}\mathbf{B}_{0,q}^{-1})\chi_{1,j}^2$  as  $CV_\alpha$ . In step 3, use either simulation or the tools provided by Rao and Scott (1981) and Modarres and Jernigan (1992). Rao and Scott (1981) proposed using a

linear transformation of  $\chi_r^2(1-\alpha)$ , the  $1-\alpha$  quantile of the chi-square distribution with  $r$  degrees of freedom, to approximate the  $1-\alpha$  quantile of the distribution of  $\sum_{j=1}^r \lambda_j(\mathbf{A}_{0,q}\mathbf{B}_{0,q}^{-1})\chi_{1,j}^2$ . For example, we could use  $[\sum_{j=1}^r \lambda_j(\mathbf{A}_{0,q}\mathbf{B}_{0,q}^{-1})/r]\chi_r^2(1-\alpha)$ . We found this works well in practice, especially when  $\lambda_j$  is close to 1. If we suppose  $\lambda_1 = 0.9$ ,  $\lambda_2 = 0.85$ ,  $\lambda_3 = 0.8$ , then the critical value obtained from the simulation of 100,000 iterations is 6.639, and the critical value obtained from  $[\sum_{j=1}^r \lambda_j(\mathbf{A}_{0,q}\mathbf{B}_{0,q}^{-1})/r]\chi_r^2(1-\alpha)$  is 6.643. The simulation calculates the true critical value accurately but, since  $\alpha$  is usually small, it takes a relatively long time.

We use  $CV_\alpha$  as the critical value. Even though  $CV_\alpha$  is intended for the case of  $\varepsilon = 0$ , it also works well for the case of  $\varepsilon > 0$ , as long as  $q$  is not very close to 1. In Section 4.3, we confirm that this approach works well for mild contamination. Unfortunately, it does not work well with heavy contamination, because then the null distribution varies too much.

We can obtain the genuine critical value for the case of  $\varepsilon > 0$ , but not easily. The null distribution depends on  $\varepsilon$  and  $g$  (Theorem 2), which are not known in practice. We present some special cases that allow us to estimate the critical values, though in general we would use  $CV_\alpha$ .

### 3.4.1. Location parameter

To test a location parameter, we can use the bootstrap method to estimate the critical value from the sample. Suppose  $H_0 : \theta = \theta_0$  and  $H_1 : \theta \neq \theta_0$ , where  $\theta$  is the location parameter. We propose the following algorithm.

Step 1: Given a sample  $\mathbf{x} = (x_1, \dots, x_n)$ , estimate the mean using a robust procedure, such as the BCMLqE  $\hat{\theta}_q$ .

Step 2: Shift the entire sample by  $\theta_0 - \hat{\theta}_q$  to obtain  $\mathbf{x}' = (x_1 - \hat{\theta}_q + \theta_0, \dots, x_n - \hat{\theta}_q + \theta_0)$ .

Step 3: Use  $\mathbf{x}'$  to get bootstrap samples  $\mathbf{x}'_b$  for  $b = 1, \dots, B$ .

Step 4: Calculate  $D_q(\mathbf{x}'_b)$  for each bootstrap sample and denote it by  $D_q^b$ .

Step 5: Calculate the  $1-\alpha$  quantile of the  $D_q^b$  as  $\widehat{CV}_{\alpha,q}$ .

As a result,  $\widehat{CV}_{\alpha,q}$  is our estimate for the critical value. The rationale behind our method is as follows. We first transform the sample  $\mathbf{x}$  to have a mean of  $\theta_0$ . With this new sample  $\mathbf{x}'$ , we use the bootstrap to mimic the null distribution. Since there are usually outliers in the sample, we choose a robust estimation for the mean, namely, the BCMLqE. We demonstrate this approach in Section 4.1.

### 3.4.2. Linear regression

Consider a linear regression setting,  $y_i = \mathbf{x}_i^T \beta + \eta_i$ , where  $y_i \in \mathbb{R}$ ,  $\mathbf{x}_i, \beta \in \mathbb{R}^p$ ,

and  $\eta_i$  is the error term. To test  $H_0 : \beta = \beta_0$  and  $H_1 : \beta \neq \beta_0$ , we propose a similar algorithm to obtain  $\widehat{CV}_\alpha$ .

Step 1: Given a sample  $\{y_i, \mathbf{x}_i\}_{i=1, \dots, n}$ , obtain a robust estimate  $\hat{\beta}$ , say the BCMLqE.

Step 2: Calculate the residual  $\hat{\eta}_i = y_i - \mathbf{x}_i^T \hat{\beta}$  for  $i = 1, \dots, n$ .

Step 3: Use  $\{\hat{\eta}_i\}_{i=1, \dots, n}$  to get bootstrap samples  $\{\hat{\eta}_i^b\}_{i=1, \dots, n}$  for  $b = 1, \dots, B$ .

Step 4: Obtain  $y_i^b = \mathbf{x}_i^T \beta_0 + \hat{\eta}_i^b$  for  $i = 1, \dots, n$ , using  $\beta_0$ .

Step 5: Calculate  $D_q(\{y_i^b, \mathbf{x}_i\}_{i=1, \dots, n})$  for each bootstrap sample and denote it by  $D_q^b$ .

Step 6: Calculate the  $1 - \alpha$  quantile of  $D_q^b, \widehat{CV}_\alpha$ .

### 3.5. Influence function and breakdown point

In this section, we use  $F$  and  $G$  to denote the distribution functions of  $f$  and  $g$ , and we consider all the test statistics as statistical functionals with domain  $\mathcal{F}$ , the set of all proper distributions.

The influence function (Hampel et al. (1986)) measures the effect of an infinitesimal contamination at the point  $x$  on the estimator, so it can be considered as a proxy for the asymptotic bias caused by the contamination at  $x$ . Ronchetti (1979, 1982a,b) has extended the influence function to hypothesis testing by defining a level influence function (LIF) and a power influence function (PIF), showing how the asymptotic level and power are influenced by a small amount of contamination at a particular point. As our test statistic  $D_q$  is not Fisher-consistent,  $D_q(F_\theta) \neq \theta$ , we can modify the test statistic by taking  $U(G) = \xi^{-1}(D_q(G))$  where  $\xi(\theta) = D_q(F_\theta)$ , so that  $U(G)$  is Fisher-consistent (Huber and Ronchetti (2009)). The properties of the influence function are relatively more difficult to determine for the likelihood-ratio-type test, compared with the Wald or score test statistics (Basu et al. (2013)). For ease of presentation, we study the influence function of closely related test statistic,  $T = \hat{\theta}_q$ , the BCMLqE of  $\theta$ . We focus on  $T$ , which is Fisher-consistent and used directly in  $D_q$ . The LIF of  $T$  is

$$\text{LIF}(x; T, F) = \frac{\phi(\Phi^{-1}(1 - \alpha_0))\text{IF}(x; T, F)}{\sqrt{\int \text{IF}(x; T, F)^2 dF(x)}},$$

where  $\alpha_0$  is the nominal level of the test. Thus,  $\text{LIF}(x; T, F)$  is proportional to  $\text{IF}(x; T, F)$ , the influence function, and that is proportional to  $\tilde{\psi}$ . For most of the parametric models that satisfy the regularity conditions,  $\tilde{\psi}$  is bounded. Therefore, both IF and LIF are bounded.

We illustrate a general method for deciding the boundedness of the influence

function for our test. For the BCMLqE, the influence function is proportional to  $\tilde{\psi}$ ,

$$\tilde{\psi}(x; \theta, q) = \frac{f'_\theta(x; \theta)}{f(x; \theta)} f(x; \theta)^{1-q} - \int f'_\theta(x; \theta) f(x; \theta)^{1-q} dx,$$

where the second term is independent of  $x$ , so we focus on the first term. When  $f$  belongs to the exponential family, we have

$$\tilde{\psi}(x; \theta, q) = (T(x) - A'_\theta(\theta)) f(x; \theta)^{1-q} - \int f'_\theta(x; \theta) f(x; \theta)^{1-q} dx.$$

We know that  $f(x; \theta) \rightarrow 0$  as  $x \rightarrow \infty$ , and that  $(T(x) - A'_\theta(\theta)) f(x; \theta)^{1-q}$  is a continuous function in  $x$ . Therefore, as long as  $(T(x) - A'_\theta(\theta)) f(x; \theta)^{1-q} \rightarrow 0$  as  $x \rightarrow \infty$ , we have the boundedness of the influence function.

We now consider the breakdown point, intuitively, the fraction of data that can be given arbitrary values without making the statistical functional arbitrarily bad. Let  $t_{\max} = \sup_{F \in \mathcal{F}} D_q(F)$  and  $t_{\min} = \inf_{F \in \mathcal{F}} D_q(F)$ , and define the level breakdown function  $\varepsilon_0$  and the power breakdown function  $\varepsilon_1$  as

$$\begin{aligned} \varepsilon_0(F_\theta, D_q) &= \inf\{\varepsilon : \sup_{G \in \mathcal{F}} D_q((1 - \varepsilon)F_\theta + \varepsilon G) = t_{\max}\}, \\ \varepsilon_1(F_\theta, D_q) &= \inf\{\varepsilon : \inf_{G \in \mathcal{F}} D_q((1 - \varepsilon)F_\theta + \varepsilon G) = t_{\min}\}. \end{aligned}$$

Here  $\varepsilon_0(F_\theta, D_q)$  represents the smallest amount of contamination necessary to drive the p-value of  $D_q$  to 0, and  $\varepsilon_1(F_\theta, D_q)$  represents the smallest amount of contamination necessary to drive the p-value of  $D_q$  to 1. If  $\theta \in \Theta_1$  and  $t_{\min} = 0$ , then  $\varepsilon_1(F_\theta, D_q)$  is the smallest fraction of contamination that can make the LqRT inconsistent. We define the level breakdown point (LBP) and power breakdown point (PBP) as

$$\begin{aligned} \text{LBP}(D_q) &= \sup_{\theta \in \Theta_0} \varepsilon_0(F_\theta, D_q), \\ \text{PBP}(D_q) &= \sup_{\theta \in \Theta_1} \varepsilon_1(F_\theta, D_q). \end{aligned}$$

**Theorem 5.** *The level breakdown point and power breakdown point of LqRT are the same as the breakdown point of the BCMLqE,  $\hat{\theta}_q$ .*

The exponential family satisfies all the regularity conditions of the BCMLqE. The influence function of the BCMLqE on the parameter in this case is bounded so it has a breakdown point of 0.5. Consequently, the LqRT has both LBP and PBP at 0.5 for the exponential family.

#### 4. Numerical Results

We present numerical studies illustrating the performance of our proposed

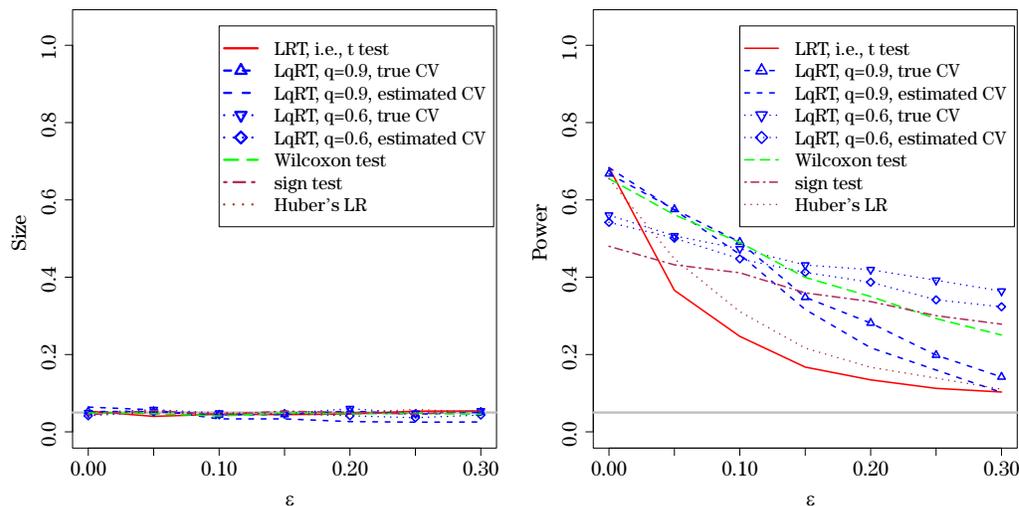


Figure 3. Comparison of powers and sizes for the LqRT ( $q = 0.9, 0.6$ ), LRT, Wilcoxon test, sign test, and HLRT at different levels of heavy-tail contamination when testing for the mean of the normal distribution ( $H_0 : \theta = 0$ ,  $H_1 : \theta \neq 0$ ). The powers were calculated using the data generating process with mean  $\theta = 0.34$ .

method through simulations and additional simulation studies can also be found in the online supplementary materials.

#### 4.1. Mean of normal distribution

For  $f$  a normal distribution with unknown mean  $\theta$  and variance  $\sigma^2$ , we test  $H_0 : \theta = 0$  against  $H_1 : \theta \neq 0$ . We simulated data with a sample size of  $n = 50$  from  $h(x; \theta, \varepsilon) = (1 - \varepsilon)\varphi(x; \theta, 1) + \varepsilon\varphi(x; \theta, 50)$ . Then we applied the LqRT with  $q = 0.9$  and  $0.6$ , the likelihood ratio test (LRT), the Wilcoxon test, the sign test, and Huber's censored likelihood ratio test (HLRT) with  $c' = 0.1$  and  $c'' = 10$ . At different levels of  $\varepsilon$ , we used  $h(x; \theta = 0, \varepsilon)$  and  $h(x; \theta = 0.34, \varepsilon)$  to generate the data and calculated the size and power. We used the approach introduced in Section 3.4.1 to generate critical values. The results are in Figure 3.

In the left panel of Figure 3, the sizes of all tests are successfully controlled at 0.05; in the right panel of Figure 3, at zero contamination, the LRT achieves the highest power; the LqRT with  $q = 0.9$ , the Wilcoxon test, and the HLRT also offer high powers. At larger contamination, the LRT degrades much faster than any of the other tests. Among all tests, the LqRT with  $q = 0.6$  and the sign test offer the slowest degradation rates. The LqRT with  $q = 0.6$  uniformly dominates the sign test at all levels of contamination. We only slightly overestimate the critical values as the powers obtained from the estimated critical values are slightly below

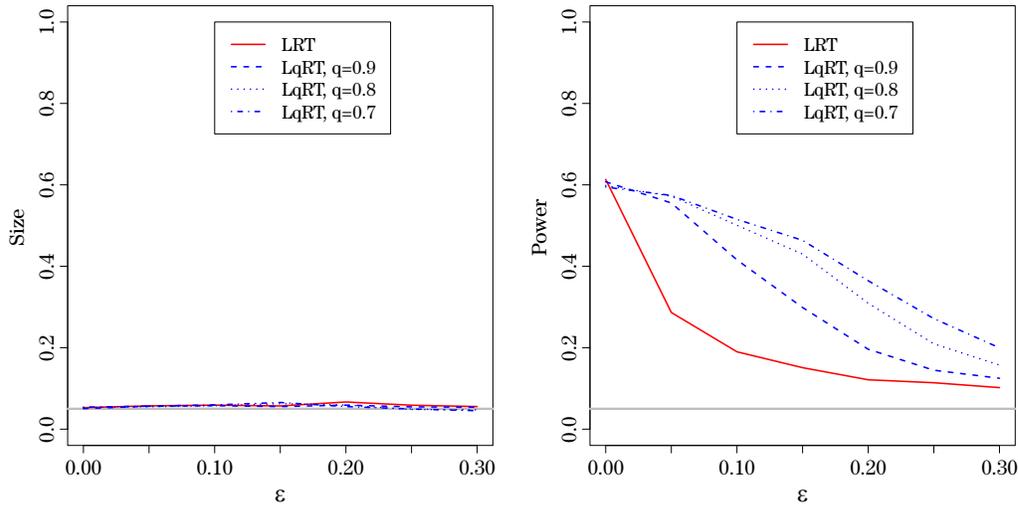


Figure 4. Comparison of powers and sizes for the LqRT ( $q = 0.9, 0.8, 0.7$ ) and the LRT under different levels of heavy-tail contamination when testing for the coefficients of the linear regression model ( $H_0 : \beta_1 = \beta_2, H_1 : \beta_1 \neq \beta_2$ ).

these of the true critical values.

We performed simulation studies using different alternative models and different contamination and obtained similar results, see the online supplementary materials.

### 4.2. Linear regression

We tested the proposed method in a linear regression setting:  $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \eta_i$ , where  $\eta_i \sim \varphi(0, \sigma^2)$ , and we want to test  $H_0 : \beta_1 = \beta_2$  against  $H_1 : \beta_1 \neq \beta_2$ . With a sample size of  $n = 100$ , we simulated data using  $x_{i1} \sim \text{Uniform}(0, 0.5)$ ,  $x_{i2} \sim \text{Uniform}(0, 0.5)$ ,  $\eta_i \sim (1 - \varepsilon)\varphi(0, 0.2) + \varepsilon\varphi(0, 10)$ , and  $y_i = x_{i1} + x_{i2} + \eta_i$ , and calculated the size of the test. In addition, we simulated data according to  $y_i = 0.5x_{i1} + 1.5x_{i2} + \eta_i$  to calculate the power. We compared the LqRT ( $q = 0.9, 0.8, 0.7$ ) with the LRT, using the approach introduced in Section 3.4.2 for critical values. The results are given in Figure 4.

In Figure 4, all tests successfully control the sizes at 0.05. As for the power, at  $\varepsilon = 0$ , the power of LqRT is slightly lower than the power of LRT. As  $\varepsilon$  increases, the LqRT degrades more slowly than the LRT, and has higher power than the LRT when  $\varepsilon > 0$ .

### 4.3. Covariance

We applied the proposed method to test the covariance. With  $\mathbf{x} \sim \text{MN}(\mathbf{0}, \Sigma)$ ,

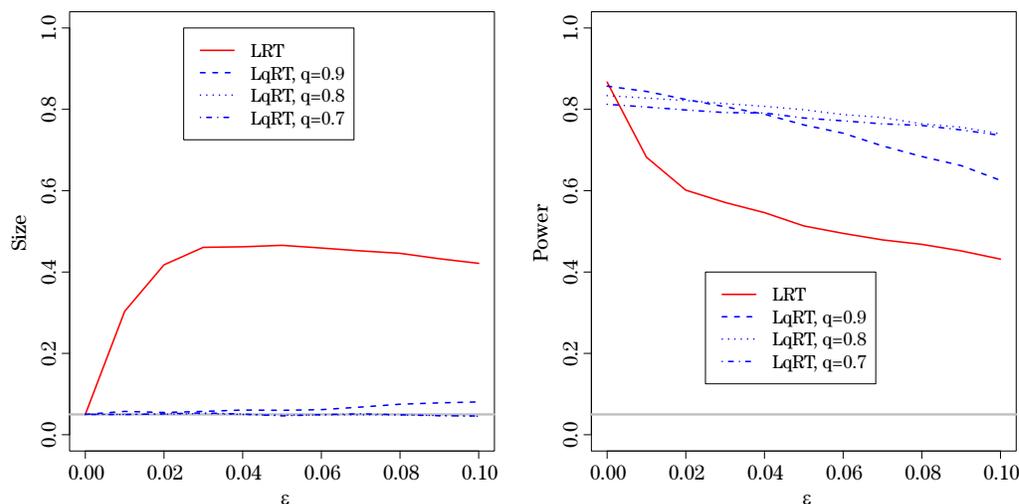


Figure 5. Comparison of powers and sizes for the LqRT ( $q = 0.9, 0.8, 0.7$ ) and the LRT under different levels of heavy-tail contaminations when testing for the covariance matrix of the multivariate normal distribution ( $H_0 : \Sigma_{1,2} = 0$ ,  $H_1 : \Sigma_{1,2} \neq 0$ ).

where  $\mathbf{x} \in \mathbb{R}^2$ ,  $\Sigma = [\Sigma_{i,j}] \in \mathbb{R}^{2 \times 2}$ , and  $\Sigma_{i,j} = \text{Cov}(X_i, X_j)$ , we tested  $H_0 : \Sigma_{1,2} = 0$  against  $H_1 : \Sigma_{1,2} \neq 0$ . With a sample size of  $n = 100$ , we simulated the data using  $\mathbf{x} \sim (1 - \varepsilon)\text{MN}(\mathbf{0}, \mathbf{I}) + \varepsilon\text{MN}(\mathbf{0}, 30\mathbf{I})$  and calculated the size of the test. With  $\mathbf{x} \sim (1 - \varepsilon)\text{MN}(\mathbf{0}, \Sigma) + \varepsilon\text{MN}(\mathbf{0}, \Sigma^*)$  with  $\Sigma_{1,1} = \Sigma_{2,2} = 1$ ,  $\Sigma_{1,1}^* = \Sigma_{2,2}^* = 30$ , and  $\Sigma_{1,2} = \Sigma_{2,1} = \Sigma_{1,2}^* = \Sigma_{2,1}^* = 0.3$ , we simulated data to calculate the power of the test. We used our general approach to obtain the critical values. The results are presented in Figure 5.

From Figure 5, as  $\varepsilon$  increases, the size of LRT increases above 0.05, while the sizes of LqRT tests increase only slightly. However, at  $\varepsilon = 0.3$ , the size of LqRT also increases significantly. Our general approach only works for mild contamination. Power degrades relatively slowly for the LqRT compared with the LRT, consistent with the right panel of Figure 4.

#### 4.4. Data

We applied our test to the Boston housing data (<https://archive.ics.uci.edu/ml/datasets/Housing>). The sample size is  $n = 506$ . The variable “full-value property-tax rate per \$10,000” serves as our variable of interest. From the histogram in Figure 6, we see outliers above 600. The mean of the data set, including these outliers, is 408.2, whereas the mean of the data without outliers is 311.9, offering a more reasonable estimate of the true center of data. We

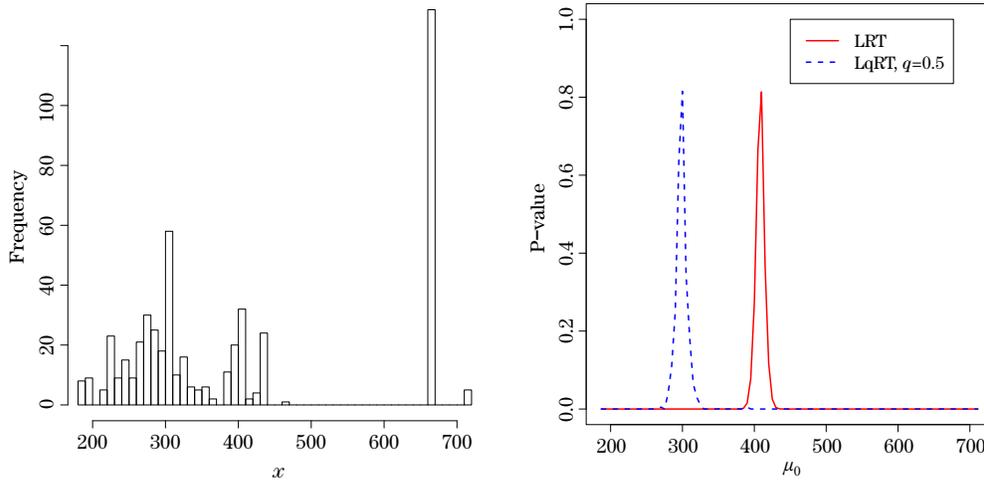


Figure 6. Histogram of property-tax rates. Figure 7. p-values as a function of  $\mu_0$  for the LqRT and the LRT.

performed hypothesis testing with  $H_0 : \mu = \mu_0$  and  $H_1 : \mu \neq \mu_0$ . Varying  $\mu_0$  from 200 to 700, we plot the corresponding p-values for the LqRT ( $q = 0.5$ ) and the LRT in Figure 7. For the LqRT, the p-value goes above 0.05 at around 300, whereas the p-value of LRT goes above 0.05 at around 400, and the LRT rejects the null hypothesis when  $\mu_0$  is at 311.9. One could first remove outliers and proceed with standard (non-robust) inference methods, but our preference, for the reasons noted in Huber and Ronchetti (2009), is for robust methods.

### 5. Selection of $q$

We have assumed the tuning parameter  $q$  to be known, but we never know the optimal  $q$  in practice. We propose a method for adaptively selecting it. The optimal  $q$  is  $q_{\text{opt}} = \arg \min_q \text{trace}(V_q(\theta_0))$ , where  $V_q(\theta_0)$  is the asymptotic variance of  $\hat{\theta}_q$ . In Figure 8, we plot the relationship between  $V_q(\theta_0)$  and  $q$  at different levels of contamination using the setup of Section 4.1. The optimal  $q$  is generally between 0.6 and 0.9 and, the more serious the contamination, the lower the optimal  $q$ . In practice, using the empirical variance  $\hat{V}_q(\hat{\theta}_q)$ , we propose data-adaptive estimation for the tuning parameter

$$\hat{q} = \arg \min_q \text{trace} \left( \left[ \frac{1}{n} \sum_{i=1}^n \tilde{\psi}'(x_i; \hat{\theta}_q, q) \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^n \tilde{\psi}(x_i; \hat{\theta}_q, q) \tilde{\psi}(x_i; \hat{\theta}_q, q)^T \right] \left[ \frac{1}{n} \sum_{i=1}^n \tilde{\psi}'(x_i; \hat{\theta}_q, q) \right]^{-1} \right).$$

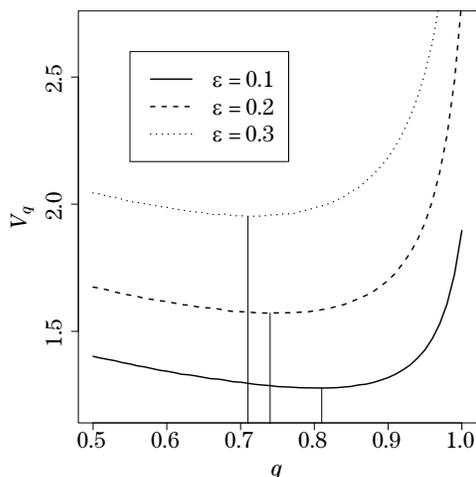


Figure 8.  $V_q(\theta_0)$  as a function of  $q$  at different levels of  $\varepsilon$ .

Accordingly, we conducted some simulation studies of the LqRT using the estimated  $q$ , first under the setup of Section 4.1. By setting  $\theta$  to 0 and 0.34, we compared the sizes and powers of our test, the LRT, the Wilcoxon test, the sign test, and the HLRT. The results are presented in Figure 9.

In Figure 9, at  $\varepsilon = 0$ , the LRT offers the highest power, and our LqRT provides nearly the same power. As  $\varepsilon$  increases away 0, the LRT's power quickly drops below that of all other tests. Our test is robust and degrades slower than the Wilcoxon test. The power of our test dominates both the Wilcoxon and sign tests uniformly at all levels of contamination. Our test can beat the nonparametric tests since we adaptively control the amount of information used, by selecting  $q$ , whereas the Wilcoxon and sign tests always use the rank information.

Figure 9 shows the average power over 2,000 Monte Carlo iterations. Since each iteration has a different  $q$ , we plot these estimated  $q$ 's in the histograms in Figure 10. At large contamination, the estimated  $q$  tends to be smaller. In our experiment, we limited the smallest  $q$  to 0.5 (which corresponds to the minimum Hellinger distance estimation (Beran (1977))) because we do not understand the case of  $q < 0.5$  well. Comparing Figures 9 and 10 with Figure 3, the LqRT with estimated  $q$  can combine the advantages of the LqRTs with fixed  $q$ 's. When the contamination ratio is low, the estimated  $q$ 's are large, so the LqRT with estimated  $q$  has roughly the same performance as the LqRT with  $q = 0.9$ . When the contamination ratio is 0.3, the mean estimated  $q$  is near 0.6, the LqRT with estimated  $q$  achieves a performance comparable to the LqRT with  $q = 0.6$ .

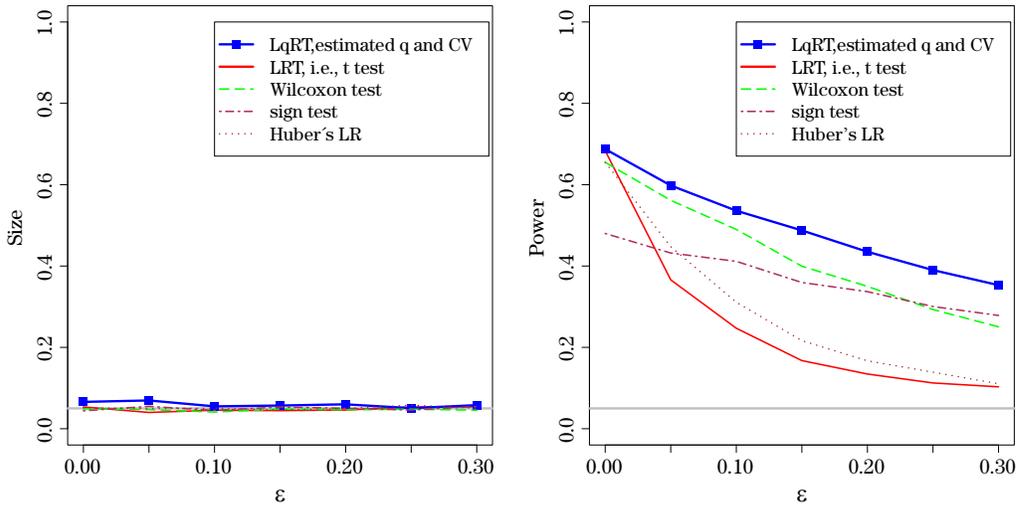


Figure 9. Comparison of the powers and sizes for the LqRT with estimated  $q$  and estimated critical value, LRT, Wilcoxon test, sign test, and HLRT under different levels of heavy-tail contamination when testing for the mean of the normal distribution ( $H_0 : \theta = 0, H_1 : \theta \neq 0$ ). The powers were calculated using the data generating process with mean  $\theta = 0.34$ .

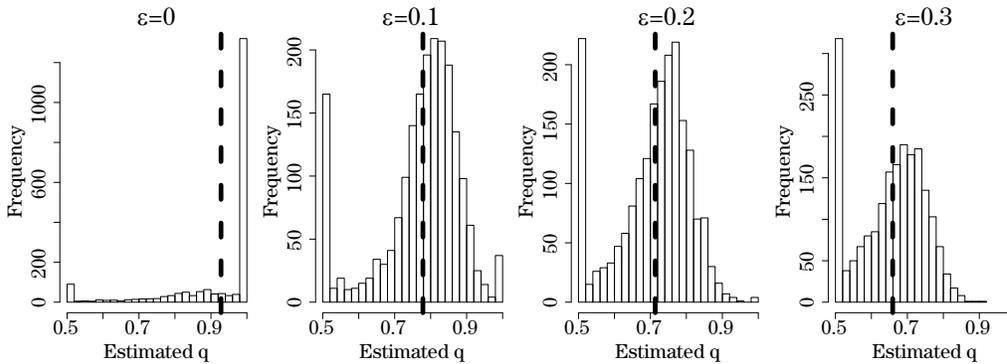


Figure 10. Histograms of the estimated  $q$  of the LqRT at different levels of heavy-tail contamination when testing for the mean of the normal distribution. These estimated  $q$ 's were obtained when calculating the powers (i.e., the right panel of Figure 9). The mean estimated  $q$  is indicated by a vertical dashed line.

We repeated the simulation from Section 4.2 for linear regression using LqRT with estimated  $q$  and present the results in Figure 11. While the size is successfully controlled, the power of LqRT with estimated  $q$  degrades the most slowly, it is relatively high compared with that of other LqRT tests with fixed  $q$ .

Through these simulation studies, we demonstrated the improved perfor-

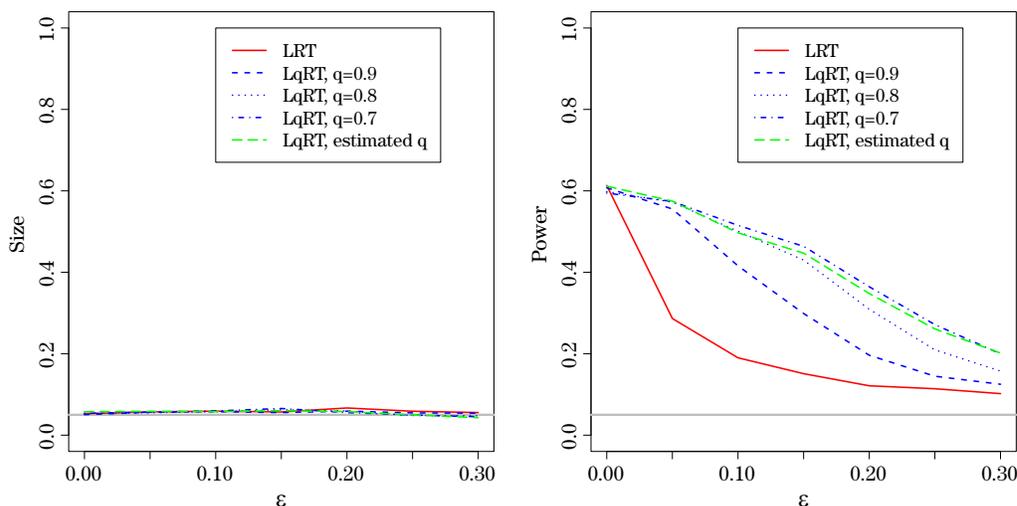


Figure 11. Comparison of powers and sizes for the LqRT with estimated  $q$  and estimated critical value and the LRT under different levels of heavy-tail contamination when testing for the coefficients of the linear regression model ( $H_0 : \beta_1 = \beta_2$ ,  $H_1 : \beta_1 \neq \beta_2$ ).

mance using estimated  $q$ . The method for estimating  $q$  works well when the model is relatively simple, but when the model is more complex, with more parameters to estimate, simply minimizing the trace of  $\hat{V}_q(\hat{\theta}_q)$  will lead to an unsatisfactory estimated  $q$ .

Additional simulation studies on the estimated  $q$  can be found in the online supplementary materials.

## 6. Conclusion

In a sense, our proposed test offers a bridge between the LRT and such nonparametric tests such as the Wilcoxon and sign tests. By changing the tuning parameter  $q$ , we control the information used in the hypothesis testing: the LRT assigns all data points equal weights, the Wilcoxon test takes only the rank information, while our test assigns each data point a weight as a function of its likelihood and  $q$ .

Many directions remain for further research. We need better estimation procedures for the critical value, and  $q$ . Our estimate of the critical value performs decently, but there is clearly a gap in the powers obtained from the true versus the estimated critical values (see Figure 3). Filling this gap is a challenging task. As well, the divergence of  $\mathbf{A}$  and  $\mathbf{B}$ , as described in Section 3.2, indicates a potential approach to model misspecification detection.

## Appendix and Supplementary Materials

We have made use of the following assumptions.

**Assumption 1.** For any  $q \in (0, 1]$ ,  $f$  satisfies the following.

1.  $\theta_0$  is an interior point in  $\Theta_0$ .
2.  $\sup_{\theta \in \Theta_0} \|\frac{1}{n} \sum_{i=1}^n \tilde{\psi}(X_i; \theta, q) - \mathbb{E}\tilde{\psi}(X; \theta, q)\| \xrightarrow{p} 0$  as  $n \rightarrow \infty$ , where  $\|\cdot\|$  represents the  $\ell_2$ -norm.
3.  $\max_{1 \leq k \leq p} \mathbb{E}_{\theta_0} |\tilde{\psi}_k(X_i; \theta_0, q)|^3$ ,  $k = 1, \dots, p$  is upper bounded by a constant, where  $\tilde{\psi}_k$  is the  $k$ -th element of  $\tilde{\psi}$ .
4. The smallest eigenvalue of  $\mathbf{A}$  is bounded away from zero.
5. If  $b_{jk}$  is the  $j$ -th row,  $k$ -th column element in  $\mathbf{B}$ , then  $b_{jk}^2$  for  $j, k = 1, \dots, p$ , are upper bounded by a constant.
6. The second order partial derivatives of  $\tilde{\psi}(x; \theta, q)$  are dominated by an integrable functions with respect to the true distribution of  $X$  for all  $\theta$  in a neighborhood of  $\theta_0$ .

**Assumption 2.** For any  $\varepsilon \in (0, 1)$ , the gross error model  $h$  is such that  $\mathbb{E}_h[f''_{\theta}(X; \theta_{\varepsilon,1}^*)/f(X; \theta_{\varepsilon,1}^*)]$  is positive definite, where  $\theta_{\varepsilon,1}^* = \arg \max_{\theta} \mathbb{E}_h[\log f(X; \theta)]$ .

**Assumption 3.**  $\mathbb{E}_f[f'_{\theta}(X; \theta_0)f'_{\theta}(X; \theta_0)^T(f(X; \theta_0)^{-2q} - f(X; \theta_0)^{-q-1})]$  is negative definite for any  $q \in (0, 1)$ .

**Assumption 4.** There exists a constant  $q^{**} \in (0, 1)$ , such that the  $\lambda_j(\mathbf{A}_{\varepsilon,q}\mathbf{B}_{\varepsilon,q}^{-1})$  are monotonic in  $q$  for any  $q \in (q^{**}, 1)$ .

See the online supplementary materials for detailed discussion on these assumptions, additional simulation studies, and proofs.

## References

- Agostinelli, C. and Markatou, M. (2001). Test of hypotheses based on the weighted likelihood methodology. *Statistica Sinica* **11**, 499–514.
- Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85(3)**, 549–559.
- Basu, A., Mandal, A., Martin, N. and Pardo, L. (2013). Testing statistical hypotheses based on the density power divergence. *Annals of the Institute of Statistical Mathematics* **65(2)**, 319–348.

- Beran, R. (1977). Minimum hellinger distance estimates for parametric models. *Annals of Statistics* **5(3)**, 445–463.
- Bickel, P. J. and Doksum, K. A. (2007). *Mathematical Statistics Basic Ideas and Selected Topics Volume i*. Pearson Prentice Hall, 2nd edn.
- Cantoni, E. and Ronchetti, E. M. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association* **96**, 1022–1030.
- Ferrari, D. and La Vecchia, D. (2012). On robust estimation via pseudo-additive information. *Biometrika* **99(1)**, 238–244.
- Ferrari, D. and Yang, Y. (2010). Maximum Lq-likelihood estimation. *Annals of Statistics* **38**, 753–783.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, 1st edn.
- He, X. (1991). A local breakdown property of robust tests in linear regression. *Journal of Multivariate Analysis* **38**, 294–305.
- He, X., Simpson, D. G. and Portnoy, S. L. (1990). Breakdown robustness of tests. *Journal of the American Statistical Association* **85(410)**, 446–452.
- Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association* **89(427)**, 897–904.
- Huber, P. J. (1965). A robust version of the probability ratio test. *Annals of Mathematical Statistics* **36(6)**, 1753–1758.
- Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*. Wiley, 2nd edn.
- Lô, S. N. and Ronchetti, E. (2009). Robust and accurate inference for generalized linear models. *Journal of Multivariate Analysis* **100**, 2126–2136.
- Markatou, M., Basu, A. and Lindsay, B. G. (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association* **93(442)**, 740–750.
- Medina, M. A. and Ronchetti, E. (2015). Robust statistics: a selective overview and new directions. *Wiley Interdisciplinary Reviews: Computational Statistics* **7**, 372–393.
- Modarres, R. and Jernigan, R. W. (1992). Testing the equality of correlation matrices. *Communications in Statistics - Theory and Methods* **21(8)**, 2107–2125.
- Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association* **76(374)**, 221–230.
- Ronchetti, E. (1997). Robust inference by influence functions. *Journal of Statistical Planning and Inference* **57**, 59–72.
- Ronchetti, E. M. (1979). Robustheitseigenschaften von tests. *Diploma thesis, ETH, Zurich*.
- Ronchetti, E. M. (1982a). Robust alternatives the f test for the linear model. *Probability and Statistical Inference*, 329–342 Eds. W. Grossman, C. Pflug, and W. Wertz, Dordrecht: Reidel.
- Ronchetti, E. M. (1982b). Robust testing in linear models: the infinitesimal approach. *Ph.D. Thesis, ETH, Zurich*.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association* **79(388)**, 871–880.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50(1)**, 1–25.

Department of Operations, Business Analytics and Information Systems, University of Cincinnati, Cincinnati, OH 45221, USA.

E-mail: qinyn@ucmail.uc.edu

Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21210, USA.

E-mail: cep@jhu.edu

(Received December 2015; accepted September 2016)