TAIWANESE JOURNAL OF MATHEMATICS
Vol. 17, No. 3, pp. 1055-1093, June 2013
DOI: 10.11650/tjm.17.2013.2254
This paper is available online at http://journal.taiwanmathsoc.org.tw

ROBUST ESTIMATION: LOCATION-SCALE AND REGRESSION PROBLEMS

Wen Hsiang Wei

Abstract. A robust estimating method to search for the bounded function minimizing a risk is proposed. The estimators obtained by the proposed method are the generalizations of several classical robust estimators. In addition, a variety of robust estimators can be obtained, including robust one and multidimensional estimators, robust regression and robust model selection criteria for univariate or multivariate data with independent or correlated errors, and robust Bayes estimators. Further, a simulation study is conducted to evaluate the proposed method.

1. Introduction

The foundations of modern robustness theory were laid by Huber (1964), as indicated by Hampel et al. (1986, p. 172). The asymptotic minimax approach is mathematically rigorous and elegant. Another approach to robust estimation using M-estimators subject to low gross-error sensitivity or bounded influence has been proposed by Hampel (see Hampel 1974; Hampel et al. 1986, Theorem 1, p. 117, p. 241; Staudte and Sheather 1990, Theorem 4.3, p. 115). The bounded influence approach has been successfully applied to a variety of statistical models, including linear regression (see Krasker and Welsch 1982) and generalized linear models (Künsch et al. 1989). In the location case, the optimal bounded influence M-estimator can be Huber estimator given the specific underlying distribution.

In many point estimation problems, the objective is to find the minimizer of a risk function. In this article, the goal is to find the "bounded" minimizer of the risk function among a class of functions. A function being a bounded function almost surely is defined first. Let $\boldsymbol{Y}$ be a $p \times 1$ random vector corresponding to the $q \times 1$ vector of parameters $\boldsymbol{\theta}$ and the parameter space be $\boldsymbol{\Theta}$. As no prior has been imposed on the parameters, the notations P, E, and Cov for the probabilities, the expected values, and the covariances of the random variables of interest correspond to the parameters in this article, i.e., $P = P_\theta$, $E = E_\theta$, and $Cov = Cov_\theta$.

Received July 23, 2012, accepted December 17, 2012.
Communicated by Yuh-Jia Lee.
2010 *Mathematics Subject Classification*: 62F35.
Key words and phrases: Bounded influence, Estimating function, MRB function, MRBB function, Robust statistics.

Definition 1.1. A measurable function $\boldsymbol{\phi} : R^p \times R^q \to R^m$,

$$\boldsymbol{\phi}(\boldsymbol{y}, \boldsymbol{\theta}) = [\phi_1(\boldsymbol{y}, \boldsymbol{\theta}), \ldots, \phi_m(\boldsymbol{y}, \boldsymbol{\theta})]^t,$$

is said to be a bounded function almost surely if

$$P(||\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})|| \leq C_\theta) = 1,$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where C_θ is a finite constant depending on $\boldsymbol{\theta}$, $\boldsymbol{y}$ is a $p \times 1$ vector, and $\|\cdot\|$ is the Euclidean distance.

If $\boldsymbol{\phi}$ is bounded, it is a bounded function almost surely. However, a bounded function almost surely might not be bounded. The minimizer is defined as follows.

Definition 1.2. A bounded function almost surely $\hat{\boldsymbol{\phi}} : R^p \times R^q \to R^m$ is minimum risk bounded (MRB) function in a class $\mathcal{F}$ if the risk

$$R\left[\hat{\boldsymbol{\phi}}(\boldsymbol{Y}, \boldsymbol{\theta})\right] = E\left\{\rho\left[\hat{\boldsymbol{\phi}}(\boldsymbol{Y}, \boldsymbol{\theta}), \boldsymbol{Y}, \boldsymbol{\theta}\right]\right\} \leq E\left\{\rho\left[\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta}), \boldsymbol{Y}, \boldsymbol{\theta}\right]\right\} = R\left[\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})\right]$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\phi} : R^p \times R^q \to R^m$ is any other measurable function in the class $\mathcal{F}$ and ρ is a non-negative measurable function almost surely defined on $R^m \times R^p \times R^q$, i.e, $P(\rho[\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta}), \boldsymbol{Y}, \boldsymbol{\theta}] \geq 0) = 1$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The functions of interest in the class $\mathcal{F}$ could be the function of residuals or its corresponding sensitivity measures, the influence function, the biased or unbiased M-estimation equations, the biased or unbiased estimating functions (see Durbin 1960), or the loss function. Thus, a variety of statistical problems can be resolved based on the proposed approach. For example, in a one-dimensional location parameter problem, the functions of interest can be $\phi(y, \theta)$, where $\phi(y, \theta)$ is a well-defined smooth function and θ is the location parameter. If the bounded function $\phi(y, \theta) = \max[-C, \min(y-\theta, C)]$ is used, the corresponding M-estimator is a type of trimmed mean which minimizes the bounded residual sum of squares $\sum_{i=1}^n \phi^2(Y_i, \theta)$ (see Serfling 1980, Example D, pp. 246-248), where C is a constant and n is the number of observations. On the other hand, if the risk function is the asymptotic variance, which mainly depends on the expectation of the square of the influence function, the corresponding optimal bounded influence M-estimator is a type of Winsorized mean. In a linear regression problem, the bounded function can be the functions ϕ_i and the corresponding Huber M-estimator is the solution of $\sum_{i=1}^n \boldsymbol{X_i}\phi_i(Y_i, \boldsymbol{\beta}) = 0$, where $\boldsymbol{X_i}$ are the values of the covariates, $\boldsymbol{\beta}$ are the regression coefficients, and $\phi_i(y, \boldsymbol{\beta}) = \max[-C, \min(y - \boldsymbol{X}_i\boldsymbol{\beta}, C)]$. In fact, the bounded functions in these examples are the MRB functions.

The MRB functions can be thought as the robust version of some commonly used functions for point estimation, for example, the score function. These functions, defined below, might be unbounded and minimize a risk among a class of functions.

Definition 1.3. A measurable function $\hat{\boldsymbol{\phi}} : R^p \times R^q \to R^m$ is minimum risk (MR) function in a class $\mathcal{F}_\infty$ if the risk

$$R\left[\hat{\boldsymbol{\phi}}(\boldsymbol{Y},\boldsymbol{\theta})\right] = E\left\{\rho\left[\hat{\boldsymbol{\phi}}(\boldsymbol{Y},\boldsymbol{\theta}),\boldsymbol{Y},\boldsymbol{\theta}\right]\right\} \leq E\left\{\rho\left[\boldsymbol{\phi}(\boldsymbol{Y},\boldsymbol{\theta}),\boldsymbol{Y},\boldsymbol{\theta}\right]\right\} = R\left[\boldsymbol{\phi}(\boldsymbol{Y},\boldsymbol{\theta})\right]$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\phi} : R^p \times R^q \to R^m$ is any other measurable function in the class $\mathcal{F}_\infty$ and ρ is a non-negative measurable function almost surely defined on $R^m \times R^p \times R^q$.

In this article, the function $\rho(\boldsymbol{\phi},\boldsymbol{y},\boldsymbol{\theta}) = \boldsymbol{\phi}^t(\boldsymbol{y},\boldsymbol{\theta})\boldsymbol{W}(\boldsymbol{\theta})\boldsymbol{\phi}(\boldsymbol{y},\boldsymbol{\theta})$, the frequently used weighted least squares, is of interest, where $\boldsymbol{W}$ is an $m \times m$ matrix function defined on R^q and $\boldsymbol{W}(\boldsymbol{\theta}) = \boldsymbol{U}^t(\boldsymbol{\theta})\boldsymbol{U}(\boldsymbol{\theta})$ is assumed to be a positive-definite weight matrix for any given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{U}$ is an $m \times m$ matrix function defined on R^q. As indicated by Croux, Filzmoser, Oja, and Critchley (see Morgenthaler 2007, p. 281, p. 288), robust methods for correlated or heteroscedastic errors in the regression need to be proposed. The types of model deviations could be taken into account by the robust methods based on the weighted least squares. In next section, the forms of the MRB functions for general parametric families are given. The MRB functions for two main parametric families, exponential families and location-scale families, are derived. As the observed data have multivariate normal distribution, the result is the generalization of Theorem 4.3 of Staudte and Sheather (1990, p. 115), which illustrates the basic idea of Hampel et al. (1986). In addition, the corresponding Bayes MRB function, referred to as minimum risk bounded Bayes (MRBB) function, is introduced. In Section 3, the applications of the MRB and MRBB functions to different types of data, including multivariate normal, binomial, and Poisson, are presented. Further, the applications of these MRB functions for a variety of regression models, including robust regression and robust model selection for univariate or multivariate data with independent or correlated errors, generalized linear models, linear mix-effects models, and generalized estimating equations, are also given in this section. A simulation study to compare the estimators based on the MRB and MRBB functions with the classical estimators, such as maximum likelihood estimator (MLE) and weighted least squares estimator (WLSE), is conducted in Section 4. A concluding discussion is given in Section 5. Finally, the proofs of main results are delegated to the last section. The additional numerical results along with computational details are delegated to the supplementary materials, which can be found at

http://web.thu.edu.tw/wenwei/www/papers/tjmSupplement.pdf/ .

2. Minimum Risk Bounded (MRB) Function

In the theory of point estimation, one way of obtaining the optimal estimator is to restrict the class of estimators of interest by requiring the estimators in the class to satisfy some condition, for example, the condition of unbiasedness (see Lehmann and Casella 1998, p. 5). In addition, unbiased estimating functions were of interest in Durbin (1960), while the optimal bounded M-estimator can be obtained subject to the first moment conditions for the corresponding estimating function and its derivative (see Theorem 1 of Hampel et al. 1986, p. 102, pp. 117-119; Theorem 4.3 of Staudte and Sheather 1990, p. 115). Motivated by these methods, the first moment conditions are imposed on the classes of functions of interest and their corresponding first derivatives in order to obtain the MRB and MRBB functions. Intuitively, the first derivatives of the functions of interest are closely related to the sensitivity of the functions to the data or the parameters.

In this section, the MRB and MRBB functions are based on the proposed generalized (or weighted) Huber function $h : R^p \times R^q \to R^r$ (also see Hampel et al. 1986, p. 239; p. 261)

$$h(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{h}\left[c(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{v}(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{y}, \boldsymbol{\theta})\right] = \boldsymbol{v}(\boldsymbol{y}, \boldsymbol{\theta}) \min\left[1, \frac{c(\boldsymbol{y}, \boldsymbol{\theta})}{\|\boldsymbol{U}(\boldsymbol{y}, \boldsymbol{\theta})\boldsymbol{v}(\boldsymbol{y}, \boldsymbol{\theta})\|}\right],$$

where $\boldsymbol{y}$ is a $p \times 1$ vector, $\boldsymbol{v}$ is a $r \times 1$ vector function defined on $R^p \times R^q$, $\boldsymbol{U}$ is an $m \times r$ matrix function defined on $R^p \times R^q$, and c is a bounded function defined on $R^p \times R^q$. As $||\boldsymbol{U}(\boldsymbol{y}, \boldsymbol{\theta})\boldsymbol{v}(\boldsymbol{y}, \boldsymbol{\theta})|| = 0$, let $h(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{v}(\boldsymbol{y}, \boldsymbol{\theta})$. Note that $c(\boldsymbol{y}, \boldsymbol{\theta})$ is usually positive and the notation $\boldsymbol{h}[c(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{v}(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{y}, \boldsymbol{\theta})]$ is used for indicating the dependence of the function on c, $\boldsymbol{v}$, and $\boldsymbol{U}$. The vector function $\boldsymbol{v}$ is associated with the possibly unbounded function of interest involving the statistic and the parameter, while the matrix function $\boldsymbol{U}$ is usually associated with the correlated errors of the data. The bounded function c is associated with the imposed bound for the function of interest. The choices of the bounded function c reflect the trade-off between robustness and efficiency. However, the determination of optimal functions c might be difficult (see Huber and Ronchetti 2009, p. 84). Therefore, data-dependent benchmark function c can be used as the alternative in such situations. Three types of data-dependent criteria are suggested:

- $100(1-\zeta)$th percentile, i.e., $c(\boldsymbol{y}, \boldsymbol{\theta}) = \min[C, q_{1-\zeta}(|\boldsymbol{z}|)]$, where C is a pre-specified constant, $0 < \zeta < 1$, $q_{1-\zeta}(|\boldsymbol{z}|)$ is the $100(1-\zeta)$th percentile of $|z_1|, \ldots, |z_m|$, and z_i is the ith element of the vector $\boldsymbol{z} = \boldsymbol{U}(\boldsymbol{y}, \boldsymbol{\theta})\boldsymbol{v}(\boldsymbol{y}, \boldsymbol{\theta})$;
- Three sigma, i.e., $c(\boldsymbol{y}, \boldsymbol{\theta}) = \min(C, 3S_z)$, where $S_z = [\sum_{i=1}^m (z_i - \bar{z})^2/(m-1)]^{1/2}$ and $\bar{z} = \sum_{i=1}^m z_i/m$;
- Box plot, i.e., $c(\boldsymbol{y}, \boldsymbol{\theta}) = \min[C, 2IQR(|\boldsymbol{z}|)]$, where $IQR(|\boldsymbol{z}|)$ is inter-quartile range of $|z_1|, \ldots, |z_m|$.

In practice, C can be a very large number such that the benchmark mainly depends on $q_{1-\zeta}(|\boldsymbol{z}|)$, $3S_z$, or $2IQR(|\boldsymbol{z}|)$.

Hereafter, let the parameter space $\boldsymbol{\Theta}$ be a nonempty open set of R^q. The notation $\dot{\boldsymbol{g}}_y = [\partial g(\boldsymbol{y})/\partial y_i]$ is a vector of first derivatives of the function g with respect to $\boldsymbol{y}$ and

$\dot{g}_{y_i}$ is the ith component of $\dot{\boldsymbol{g}}_y$, while $\Delta_y(\boldsymbol{g})$ is a vector with the ith element $\Delta_{y_i}(g) = g(\ldots, y_i - 1, \ldots)$ and $\Delta_y^+(\boldsymbol{g})$ is a vector with the ith element $\Delta_{y_i}^+(g) = g(\ldots, y_i + 1, \ldots)$. If $E[\dot{\boldsymbol{g}}_y(\boldsymbol{Y})]$ exists, the differentiability of $\dot{\boldsymbol{g}}_y$ $a.s.$ is assumed spontaneously, where $a.s.$ stands for almost surely. Let vec$(\boldsymbol{M}) = (m_{11}, m_{21}, \ldots, m_{(m-1)n}, m_{mn})^t$ be the vectorization of the $m \times n$ matrix $\boldsymbol{M} = [m_{ij}]$, while vecs$(\boldsymbol{M}) = (m_{11}, \ldots, m_{pp}, m_{21}, m_{31}, m_{32}, \ldots, m_{p(p-1)})^t$ is the vectorization of the $p \times p$ symmetric matrix $\boldsymbol{M}$. Further, the quantities with a hat correspond to the ones of the MRB, MRBB, or MR functions. For example, $\hat{\boldsymbol{\mu}}$ is denoted as the means of the MRB, MRBB, or MR functions provided that $\boldsymbol{\mu}$ is denoted as the means of the class of functions of interest. The notations $\boldsymbol{\mu}_1$ and $\hat{\boldsymbol{\mu}}_1$ are the matrices with elements equal to the means of the first derivatives of the functions of interest and the associated MRB, MRBB, or MR functions, respectively. Finally, the inequality $\boldsymbol{v} < \infty$ is used to indicate that all the elements of the vector $\boldsymbol{v}$ are finite.

2.1. Multivariate normal distribution and large sample case

Since Theorem 4.3 of Staudte and Sheather (1990, p. 115) plays a crucial role in the development of the MRB functions in this article, the normal random variables are considered first and the generalization of the theorem can be obtained. Further, as the sample size is large and the underlying random variables are not normal, the function related to the norm MRB function can be also obtained. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^t$ have a multivariate normal distribution with the parameter $\boldsymbol{\theta}$ of interest, for example, $\boldsymbol{\theta}$ being the vector of parameters corresponding to the mean and variance-covariance matrix of $\boldsymbol{Y}$. Note that $\boldsymbol{Y}$ has a density with respect to Lebesgue measure on R^p.

The following class of functions is of interest:

$$\begin{aligned}\mathcal{F}^n\left[c(\boldsymbol{y},\boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta}), \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\mu}_1(\boldsymbol{\theta})\right] = \{&\boldsymbol{\phi} : ||\boldsymbol{U}\boldsymbol{\phi}|| \le c(\boldsymbol{Y},\boldsymbol{\theta}) \le C_\theta < \infty \ a.s., \\ &E(|\phi_k)|) < \infty, E(\boldsymbol{\phi}) = \boldsymbol{\mu}(\boldsymbol{\theta}), E\left(\left|\dot{\phi}_{k,y_i}\right|\right) < \infty, E\left(\dot{\phi}_{k,y_i}\right) = \mu_{1,ik}(\boldsymbol{\theta}), \\ &\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, k = 1, \ldots, m, i = 1, \ldots, p\}.\end{aligned}$$

The MRB function in the class is given by the following theorem.

Theorem 2.1. *The function*

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{\theta}) = \boldsymbol{h}\left[c(\boldsymbol{y},\boldsymbol{\theta}), \boldsymbol{M_1}(\boldsymbol{\theta})\boldsymbol{y} + \boldsymbol{M_0}(\boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta})\right] \tag{2.1}$$

is the MRB function in the class $\mathcal{F}^n[c(\boldsymbol{y},\boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ *if the following conditions hold:*

(i)
$$P\left(\boldsymbol{Y} \in \{\boldsymbol{y} : ||\boldsymbol{U}(\boldsymbol{\theta})\left[\boldsymbol{M_1}(\boldsymbol{\theta})\boldsymbol{Y} + \boldsymbol{M_0}(\boldsymbol{\theta})\right]|| = c(\boldsymbol{y},\boldsymbol{\theta})\}\right) = 0;$$

(ii) $E(|\dot{c}_{y_i}|) < \infty$;

(iii) *the elements of the $m \times m$ matrix $\boldsymbol{U}(\boldsymbol{\theta})$, the $m \times p$ matrix function $\boldsymbol{M_1}(\boldsymbol{\theta})$, and the $m \times 1$ vector function $\boldsymbol{M_0}(\boldsymbol{\theta})$ are finite for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.*

The abbreviated notation for the class $\mathcal{F}^n[c(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ is $\mathcal{F}^n$. Note that the generalized Huber function is differentiable with respect to $\boldsymbol{y}$ almost surely by the equation in condition (i) and by condition (ii). The generalized Huber function is well defined and bounded almost surely by condition (iii). The basic idea of proving the theorem is mainly from Theorem 1 of Hampel et al. (1986, pp. 117-119, pp. 241-243) and Theorem 4.3 of Staudte and Sheather (1990, p. 115). If $\boldsymbol{U}(\boldsymbol{\theta})$ is the identity matrix $\boldsymbol{I}$, $c(\boldsymbol{y}, \boldsymbol{\theta}) = C$, $\boldsymbol{M_1}(\boldsymbol{\theta}) = \boldsymbol{I}$ and $\boldsymbol{M_0}(\boldsymbol{\theta}) = \boldsymbol{0}$, the MRB function $\hat{\boldsymbol{\phi}}$ given in the expression (2.1) is the Huber function (see Hampel et al. 1986, p. 239; Serfling 1980, p. 247; Staudte and Sheather 1990, p. 112).

The MRB functions with mean 0, such as the influence function or residuals in some models, can be obtained by the following corollary.

Corollary 2.1. *Under the assumptions given in Theorem 2.1, the function*

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\nu}, \boldsymbol{\Sigma}) = \boldsymbol{h}\left[c(\boldsymbol{\nu}, \boldsymbol{\Sigma}), \boldsymbol{Y} - \boldsymbol{\nu}, \boldsymbol{\Sigma}^{-1/2}\right]$$

is the MRB function in the class $\mathcal{F}^n[c(\boldsymbol{\nu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma}^{-1/2}, 0, \hat{\boldsymbol{\mu}}_1(\boldsymbol{\nu}, \boldsymbol{\Sigma})]$, where $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma}$ are the mean vector and variance-covariance matrix of $\boldsymbol{Y}$, respectively.

If $\boldsymbol{U} = 1$, $\hat{\boldsymbol{\mu}}_1 = 1$, $c(\boldsymbol{\nu}, \boldsymbol{\Sigma}) = C$, and the bounded influence functions in a location problem are of interest, the corollary is Theorem 4.3 of Staudte and Sheather (1990, p. 115).

As the underlying distribution for the data is not a multivariate normal distribution, the function with a minimum risk, which has the same form as the MRB function in $\mathcal{F}^n$, can be obtained based on Theorem 2.1.

Theorem 2.2. *Let $\boldsymbol{Y_n} = (Y_{n1}, \ldots, Y_{np})^t$, $\boldsymbol{Y_n} \xrightarrow{d} \boldsymbol{Y}$, and $\boldsymbol{Y}$ have a multivariate normal distribution, where $\xrightarrow{d}$ denotes the convergence in distribution. Then, for sufficiently large n,*

$$R\left[\boldsymbol{\phi}_n(\boldsymbol{Y_n}, \boldsymbol{\theta})\right] \geq R\left[\hat{\boldsymbol{\phi}}(\boldsymbol{Y_n}, \boldsymbol{\theta})\right],$$

if the following conditions hold:
(i) $\boldsymbol{\phi}_n(\boldsymbol{Y_n}, \boldsymbol{\theta}) \xrightarrow{d} \boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})$ for any given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\phi} \in \mathcal{F}^n$, $E(Y_{nj}^2) \to E(Y_j^2)$, and

$$\|\boldsymbol{U}(\boldsymbol{\theta})\boldsymbol{\phi}_n(\boldsymbol{Y_n}, \boldsymbol{\theta})\| \leq c(\boldsymbol{Y_n}, \boldsymbol{\theta}) \leq C_\theta < \infty \ a.s.;$$

(ii) the conditions given in Theorem 2.1 hold.

Note that ϕ_n can be the function not belonging to the class $\mathcal{F}^n$. The above theorem requires finite second moments of the random variables Y_{nj} for large n. Therefore, for the random variables with infinite second moments such as Cauchy random variables, the large sample approximation based on the above theorem might not be suitable. However, for Cauchy random variables, the exact MRB functions do exist, as implied by Theorem 2.3 and Theorem 2.4 in the following section.

2.2. General parametric families

Since the optimal bounded score function in Theorem 1 of Hampel et al. (1986, pp. 117-119, pp. 241-243) is a fundamental quantity in robust statistics, the generalization of the theorem is given in this section. The density function $f(\boldsymbol{y}|\boldsymbol{\theta})$ of interest with respect to some common measure, also denoted as f, is considered as a function of both $\boldsymbol{y} = (y_1, \ldots, y_p)^t$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^t$.

If the support of $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^t$ is $(a_1, b_1) \times \cdots \times (a_p, b_p)$, not necessarily bounded, and the corresponding measure is Lebesgue measure, the following class of functions is of interest:

$$\begin{aligned}\mathcal{F}^y\left[c(\boldsymbol{y},\boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta}), \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\mu}_1(\boldsymbol{\theta})\right] = \{&\boldsymbol{\phi} : ||\boldsymbol{U}\boldsymbol{\phi}|| \le c(\boldsymbol{Y},\boldsymbol{\theta}) \le C_\theta < \infty \ a.s., \\ &E(|\phi_k|) < \infty, E(\boldsymbol{\phi}) = \boldsymbol{\mu}(\boldsymbol{\theta}), E\left(\left|\dot{\phi}_{k,y_i}\right|\right) < \infty, E\left(\dot{\phi}_{k,y_i}\right) = \mu_{1,ik}(\boldsymbol{\theta}), \\ &\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, k = 1, \ldots, m, i = 1, \ldots, p.\}.\end{aligned}$$

If Y_i take values in the set $\{0, 1, 2, \cdots\}$ and the corresponding measure is counting measure (see Lehmann and Casella 1998, p. 8), the following class of functions is of interest:

$$\begin{aligned}\mathcal{F}^{\Delta y}\left[c(\boldsymbol{y},\boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta}), \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\mu}_1(\boldsymbol{\theta})\right] = \{&\boldsymbol{\phi} : ||\boldsymbol{U}\boldsymbol{\phi}|| \le c(\boldsymbol{Y},\boldsymbol{\theta}) \le C_\theta < \infty \ a.s., \\ &E(|\phi_k|) < \infty, E(\boldsymbol{\phi}) = \boldsymbol{\mu}(\boldsymbol{\theta}), E\left[|\Delta^+_{y_i}(\phi_k)|\right] < \infty, E\left[\Delta^+_{y_i}(\phi_k)\right] = \mu_{1,ik}(\boldsymbol{\theta}), \\ &\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, k = 1, \ldots, m, i = 1, \ldots, p.\}.\end{aligned}$$

The MRB functions in the two classes are given by the following theorem.

Theorem 2.3. *(a) Let*

$$\hat{\boldsymbol{\phi}}^y_\infty(\boldsymbol{y},\boldsymbol{\theta}) = \boldsymbol{M_1}(\boldsymbol{\theta}) \left(\frac{\dot{\boldsymbol{f}}_y}{f}\right) + \boldsymbol{M_0}(\boldsymbol{\theta}).$$

The function

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{\theta}) = \boldsymbol{h}\left[c(\boldsymbol{y},\boldsymbol{\theta}), \hat{\boldsymbol{\phi}}^y_\infty(\boldsymbol{y},\boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta})\right] \tag{2.2}$$

is the MRB function in the class $\mathcal{F}^y[c(\boldsymbol{y},\boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ if the following conditions hold:

(i) *the support of* $\boldsymbol{Y}$ *is* $(-\infty,\infty)\times\cdots\times(-\infty,\infty)$*;*
(ii)

$$P\left(\boldsymbol{Y}\in\left\{\boldsymbol{y}:\left\|\boldsymbol{U}(\boldsymbol{\theta})\hat{\phi}^{y}_{\infty}(\boldsymbol{y},\boldsymbol{\theta})\right\|=c(\boldsymbol{y},\boldsymbol{\theta})\right\}\right)=0;$$

(iii) $\partial^2 f(\boldsymbol{Y}|\boldsymbol{\theta})/\partial\boldsymbol{y}\partial\boldsymbol{y}^t$ *exists almost surely for all* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$*;*
(iv) $E(|\dot{c}_{y_i}|)<\infty$ *and* $E[|\partial(\dot{f}_{y_i}/f)/\partial y_j|]<\infty$ *for* $i,j=1,\ldots,p$*;*
(v) *the elements of the* $m\times m$ *matrix* $\boldsymbol{U}(\boldsymbol{\theta})$*, the* $m\times p$ *matrix function* $\boldsymbol{M_1}(\boldsymbol{\theta})$*, and the* $m\times 1$ *vector function* $\boldsymbol{M_0}(\boldsymbol{\theta})$ *are finite for all* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$*;*
(vi) $E[(\dot{f}_{y_i}/f)^2]<\infty$.

(b) *The function given in the expression* (2.2) *is the MRB function in the class* $\mathcal{F}^y[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ *if the following conditions hold:*
(i) *the support of* $\boldsymbol{Y}$ *is the product of bounded intervals;*
(ii) *the conditions*(ii)*,* (iii)*,* (iv)*,* (v)*, and* (vi) *in* (a) *hold;*
(iii) $f\to 0$ *as* $y_i\to a_i$ *or* $y_i\to b_i$.
(c) *Let*

$$\hat{\phi}^{\Delta y}_{\infty}(\boldsymbol{y},\boldsymbol{\theta})=\boldsymbol{M_1}(\boldsymbol{\theta})\left[\frac{\Delta_y(\boldsymbol{f})}{f}\right]+\boldsymbol{M_0}(\boldsymbol{\theta}),$$

if $f>0$ *and* $\hat{\boldsymbol{\phi}}^{\Delta y}_{\infty}(\boldsymbol{y},\boldsymbol{\theta})=0$ *otherwise. The function*

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{\theta})=\boldsymbol{h}\left[c(\boldsymbol{y},\boldsymbol{\theta}),\hat{\boldsymbol{\phi}}^{\Delta y}_{\infty}(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{\theta})\right]$$

is the MRB function in the class $\mathcal{F}^{\Delta y}[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ *if the following conditions hold:*
(i) Y_i *take values in the set* $\{0,1,2,\cdots\}$*;*
(ii) $E\{\Delta^{+}_{y_j}[\Delta_{y_i}(f)/f]\}<\infty$*;*
(iii) *the condition (v) given in (a) holds;*
(iv) $E\{[\Delta_{y_i}(f)/f]^2\}<\infty$.

The abbreviated notations for the two classes $\mathcal{F}^y[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ and $\mathcal{F}^{\Delta y}[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ are $\mathcal{F}^y$ and $\mathcal{F}^{\Delta y}$, respectively. The above theorem can be applied to two principal families of models, exponential families and location-scale families, as given by the following corollary. An exponential family of distributions has densities of the form (see Lehmann and Casella 1998, p. 23)

$$f(\boldsymbol{y}\mid\boldsymbol{\theta})=\exp\left[\sum_{i=1}^{q}\theta_i T_i(\boldsymbol{y})-A(\boldsymbol{\theta})\right]\kappa(\boldsymbol{y}),\tag{2.3}$$

and the densities of a location-scale family of distributions are

$$f(\boldsymbol{y}\mid\boldsymbol{\nu},\boldsymbol{B})=sign(|\boldsymbol{B}|)|\boldsymbol{B}|f_0\left[\boldsymbol{B}(\boldsymbol{y}-\boldsymbol{\nu})\right],\tag{2.4}$$

where A and κ are real functions, T_i are real-valued statistics, $\boldsymbol{\nu}$ is the location parameter, $\boldsymbol{B} = [b_{ij}]$ is a nonsingular scale parameter matrix, $f_0(\boldsymbol{x})$ is a density function, $sign(c) = 1$ as $c > 0$ and $sign(c) = -1$ as $c < 0$, and $|\boldsymbol{B}|$ is the determinant of the matrix $\boldsymbol{B}$.

Corollary 2.2. *(a) Let $\boldsymbol{Y}$ have the density function given in the expression (2.3). If the support of $\boldsymbol{Y}$ is $(a_1, b_1) \times \cdots \times (a_p, b_p)$, $\dot{\boldsymbol{f}}_y/f$ in the MRB function of Theorem 2.3 is*

$$\frac{\dot{\boldsymbol{f}}_y}{f} = \frac{\partial \left\{ \boldsymbol{\theta}^t \boldsymbol{T}(\boldsymbol{y}) + \log\left[\kappa(\boldsymbol{y})\right] \right\}}{\partial \boldsymbol{y}},$$

where $\boldsymbol{T}(\boldsymbol{y}) = [T_1(\boldsymbol{y}), \ldots, T_q(\boldsymbol{y})]^t$. If Y_i take values in the set $\{0, 1, 2, \cdots\}$, $\Delta_y(\boldsymbol{f})/f$ in the MRB function is

$$\frac{\Delta_y(\boldsymbol{f})}{f} = \left(\frac{\Delta_y \{ \exp\left[\sum_{i=1}^{q} \theta_i T_i(\boldsymbol{y}) \right] \}}{\exp\left[\sum_{i=1}^{q} \theta_i T_i(\boldsymbol{y}) \right]} \right) \left(\frac{\Delta_y \left[\boldsymbol{\kappa}(\boldsymbol{y}) \right]}{\kappa(\boldsymbol{y})} \right).$$

(b) Let $\boldsymbol{Y}$ have the density function given in the expression (2.4). If the support of $\boldsymbol{Y}$ is $(a_1, b_1) \times \cdots \times (a_p, b_p)$, $\dot{\boldsymbol{f}}_y/f$ in the MRB function of Theorem 2.3 is

$$\frac{\dot{\boldsymbol{f}}_y}{f} = \frac{\boldsymbol{B}^t \dot{\boldsymbol{f}}_{0,x}\left[\boldsymbol{B}(\boldsymbol{y} - \boldsymbol{\nu})\right]}{f_0\left[\boldsymbol{B}(\boldsymbol{y} - \boldsymbol{\nu})\right]}.$$

The smoothness conditions in the above classes of functions are mainly on the first derivatives of the functions $\boldsymbol{\phi}$ with respect to the data. The following class subject to the smoothness conditions on the first derivatives with respect to the parameters is of interest:

$$\begin{aligned}\mathcal{F}^{\theta}\left[c(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta}), \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\mu}_1(\boldsymbol{\theta})\right] = \{ \boldsymbol{\phi} : ||\boldsymbol{U}\boldsymbol{\phi}|| \le c(\boldsymbol{Y}, \boldsymbol{\theta}) \le C_\theta < \infty \ a.s., \\ E(\boldsymbol{\phi}) = \boldsymbol{\mu}(\boldsymbol{\theta}) < \infty, E\left(\dot{\phi}_{k,\theta_i}\right) = \mu_{1,ik}(\boldsymbol{\theta}) < \infty, \\ \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, k = 1, \ldots, m, i = 1, \ldots, q. \}.\end{aligned}$$

The MRB function in the class is given by the following theorem.

Theorem 2.4. *Let*

$$\hat{\boldsymbol{\phi}}^{\theta}_{\infty}(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{M_1}(\boldsymbol{\theta}) \left(\frac{\dot{\boldsymbol{f}}_\theta}{f} \right) + \boldsymbol{M_0}(\boldsymbol{\theta}),$$

if $f > 0$ and $\hat{\phi}^{\theta}_{\infty}(\boldsymbol{y}, \boldsymbol{\theta}) = 0$ otherwise. The function

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{h}\left[c(\boldsymbol{y}, \boldsymbol{\theta}), \hat{\boldsymbol{\phi}}^{\theta}_{\infty}(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta}) \right] \tag{2.5}$$

is the MRB function in the class $\mathcal{F}^{\theta}[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ *if the following conditions hold:*
(i)

$$P\left(\boldsymbol{Y}\in\left\{\boldsymbol{y}:\left\|\boldsymbol{U}(\boldsymbol{\theta})\hat{\phi}^{\theta}_{\infty}(\boldsymbol{y},\boldsymbol{\theta})\right\|=c(\boldsymbol{y},\boldsymbol{\theta})\right\}\right)=0;$$

(ii) $\partial^2 f(\boldsymbol{Y}|\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t$ *exists almost surely for all* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$*;*
(iii) $E(|\dot{c}_{\theta_i}|)<\infty$ *and* $E[|\partial(\dot{f}_{\theta_i}/f)/\partial\theta_j|]<\infty$ *for* $i,j=1,\ldots,q$*;*
(iv) *the elements of the* $m\times m$ *matrix* $\boldsymbol{U}(\boldsymbol{\theta})$*, the* $m\times q$ *matrix function* $\boldsymbol{M_1}(\boldsymbol{\theta})$*, and the* $m\times 1$ *vector function* $\boldsymbol{M_0}(\boldsymbol{\theta})$ *and their first derivatives are finite for all* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$*;*
(v) $E[\boldsymbol{\phi}(\boldsymbol{Y},\boldsymbol{\theta})]$ *can be differentiated with respect to* $\boldsymbol{\theta}$ *under the integral sign for any* $\boldsymbol{\phi}$ *in the class* $\mathcal{F}^{\theta}[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$*;*
(vi) $E[(\dot{f}_{\theta_i}/f)^2]<\infty$.

The abbreviated notation for the class $\mathcal{F}^{\theta}[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})]$ is $\mathcal{F}^{\theta}$. If $m=1$, $c(\boldsymbol{y},\boldsymbol{\theta})=C$, $\boldsymbol{M_1}(\boldsymbol{\theta})=1$, and $\boldsymbol{U}(\boldsymbol{\theta})=1$, the MRB function given in the expression (2.5) is the optimal bounded influence function given in Theorem 1 of Hampel et al. (1986, pp. 117-119).

Theorem 2.4 can be also applied to the previous two families of models. For the densities of the location-scale families of distributions, let $\boldsymbol{\theta}=[\boldsymbol{\nu^t},\mathrm{vec}^t(\boldsymbol{B})]^t$ provided that the elements of $\boldsymbol{B}$ are all distinct. Also, let $\boldsymbol{\theta}=[\boldsymbol{\nu^t},\mathrm{vecs}^t(\boldsymbol{B})]^t$ provided that the matrix $\boldsymbol{B}$ is symmetric.

Corollary 2.3. *(a) Let* $\boldsymbol{Y}$ *have the density function given in the expression (2.3).* $\dot{\boldsymbol{f}}_{\theta}/f$ *in the MRB function of Theorem 2.4 is*

$$\frac{\dot{\boldsymbol{f}}_{\theta}}{f}=\boldsymbol{T}(\boldsymbol{y})-\dot{\boldsymbol{A}}_{\theta}.$$

(b) Let $\boldsymbol{Y}$ *have the density given in the expression (2.4). Then, if* b_{ij} *are distinct,*

$$\frac{\dot{\boldsymbol{f}}_{\theta}}{f}=\left[\left(\frac{\dot{\boldsymbol{f}}_{\nu}}{f}\right)^t,vec^t\left(\frac{\frac{\partial f}{\partial\boldsymbol{B}}}{f}\right)\right]^t,$$

where

$$\frac{\dot{\boldsymbol{f}}_{\nu}}{f}=\frac{(-1)\boldsymbol{B}^t\dot{\boldsymbol{f}}_{0,x}\left[\boldsymbol{B}(\boldsymbol{y}-\boldsymbol{\nu})\right]}{f_0\left[\boldsymbol{B}(\boldsymbol{y}-\boldsymbol{\nu})\right]},$$

and

$$\frac{\frac{\partial f}{\partial\boldsymbol{B}}}{f}=\left(\boldsymbol{B}^{-1}\right)^t+\boldsymbol{G},$$

and where $\boldsymbol{G}=\dot{\boldsymbol{f}}_{0,x}[\boldsymbol{B}(\boldsymbol{y}-\boldsymbol{\nu})](\boldsymbol{y}-\boldsymbol{\nu})^t/f_0[\boldsymbol{B}(\boldsymbol{y}-\boldsymbol{\nu})]$*. If* $\boldsymbol{B}$ *is symmetric,*

$$\frac{\dot{\boldsymbol{f}}_\theta}{f} = \left[\left(\frac{\dot{\boldsymbol{f}}_\nu}{f} \right)^t, vecs^t \left(\frac{\frac{\partial f}{\partial \boldsymbol{B}}}{f} \right) \right]^t,$$

where

$$\frac{\frac{\partial f}{\partial \boldsymbol{B}}}{f} = \left[2\boldsymbol{B}^{-1} - diag(\boldsymbol{B}^{-1}) \right] + \boldsymbol{G} + \boldsymbol{G}^t - diag(\boldsymbol{G}),$$

and where $diag(\boldsymbol{M})$ is a diagonal matrix of which diagonal equal to the one of the matrix $\boldsymbol{M}$.

When the underlying distribution has the form $f(\boldsymbol{y}|\boldsymbol{\theta}) = f_0(\boldsymbol{y} - \boldsymbol{\theta})$, the equivalence of the MRB function given in the expression (2.2) and the one given in the expression (2.5) is indicated by the following corollary.

Corollary 2.4. *Let $\boldsymbol{Y}$ have the density function $f(\boldsymbol{y}|\boldsymbol{\theta}) = f_0(\boldsymbol{y} - \boldsymbol{\theta})$, where f_0 is known and $\boldsymbol{\theta}$ is an unknown location parameter. For any MRB function in the class $\mathcal{F}^y$, there exists an equivalent MRB function in the class $\mathcal{F}^\theta$.*

2.3. Bayes model

In this section, the Bayes MRB functions are given. Suppose that the prior distribution of the parameter $\boldsymbol{\theta}$ is $\pi(\boldsymbol{\theta})$ and $f_{\theta|y}$ is the posterior density function. The bounded function almost surely $\boldsymbol{\phi}$ and the nonnegative function almost surely ρ in Bayes procedures satisfy

$$P\left((\boldsymbol{Y}, \boldsymbol{\theta}) \in \{(\boldsymbol{y}, \boldsymbol{\theta}^*) : ||\boldsymbol{\phi}(\boldsymbol{y}, \boldsymbol{\theta}^*)|| \le C\}\right) = 1$$

and

$$P\left((\boldsymbol{Y}, \boldsymbol{\theta}) \in \{(\boldsymbol{y}, \boldsymbol{\theta}^*) : \rho[\boldsymbol{\phi}(\boldsymbol{y}, \boldsymbol{\theta}^*), \boldsymbol{y}, \boldsymbol{\theta}^*)] \ge 0\}\right) = 1,$$

respectively. In addition, the weight matrix $\boldsymbol{W}$ in Bayes procedures is an $m \times m$ matrix function defined on R^p and $\boldsymbol{W}(\boldsymbol{Y}) = \boldsymbol{U}^t(\boldsymbol{Y})\boldsymbol{U}(\boldsymbol{Y})$ is assumed to be a positive-definite weight matrix almost surely, where $\boldsymbol{U}$ is an $m \times m$ matrix function defined on R^p. The minimum risk bounded Bayes (MRBB) function and the minimum risk Bayes function can be defined analogously as in Definition 1.2 and Definition 1.3. The main difference is that the risk of interest in this section is the expected value of the function $\rho(\boldsymbol{\phi}, \boldsymbol{Y}, \boldsymbol{\theta}) = \boldsymbol{\phi}^t(\boldsymbol{Y}, \boldsymbol{\theta})\boldsymbol{W}(\boldsymbol{Y})\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})$ of the random vector $(\boldsymbol{Y}, \boldsymbol{\theta})$, i.e., $\boldsymbol{\theta}$ being not degenerated.

If the support of $\boldsymbol{\theta}$ is $(c_1, d_1) \times \cdots \times (c_q, d_q)$, not necessarily bounded, the following class of functions is of interest:

$$\mathcal{F}^b\left[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{y}),\boldsymbol{\mu}(\boldsymbol{y}),\boldsymbol{\mu}_1(\boldsymbol{y})\right]=\{\boldsymbol{\phi}:||\boldsymbol{U\phi}||\le c(\boldsymbol{Y},\boldsymbol{\theta})\le C<\infty \ a.s., E(|\phi_k|)<\infty, E\left[\boldsymbol{\phi}\mid\boldsymbol{Y}\right]=\boldsymbol{\mu}(\boldsymbol{Y}) \ a.s., E\left(\left|\dot{\phi}_{k,\theta_i}\right|\right)<\infty, E\left[\dot{\phi}_{k,\theta_i}\mid\boldsymbol{Y}\right]=\mu_{1,ik}(\boldsymbol{Y}) \ a.s., k=1,\ldots,m, i=1,\ldots,q.\right\}.$$

If θ_i take values in the set $\{0,1,2,\cdots\}$, the following class of functions is of interest:

$$\mathcal{F}^{\Delta b}\left[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{y}),\boldsymbol{\mu}(\boldsymbol{y}),\boldsymbol{\mu}_1(\boldsymbol{y})\right]=\{\boldsymbol{\phi}:||\boldsymbol{U\phi}||\le c(\boldsymbol{Y},\boldsymbol{\theta})\le C<\infty \ a.s., E(|\phi_k|)<\infty, E\left[\boldsymbol{\phi}\mid\boldsymbol{Y}\right]=\boldsymbol{\mu}(\boldsymbol{Y}) \ a.s., E\left[|\Delta^+_{\theta_i}(\phi_k)|\right]<\infty, E\left[\Delta^+_{\theta_i}(\phi_k)\mid\boldsymbol{Y}\right]=\mu_{1,ik}(\boldsymbol{Y}) \ a.s., k=1,\ldots,m, i=1,\ldots,q.\right\}.$$

The MRBB functions in the two classes are given by the following theorem.

Theorem 2.5. *(a) Let*

$$\hat{\boldsymbol{\phi}}^b_\infty(\boldsymbol{y},\boldsymbol{\theta})=\boldsymbol{M}_1(\boldsymbol{y})\left(\frac{\dot{\boldsymbol{f}}_{\theta|y,\theta}}{f_{\theta|y}}\right)+\boldsymbol{M}_0(\boldsymbol{y}),$$

The function

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{\theta})=\boldsymbol{h}\left[c(\boldsymbol{y},\boldsymbol{\theta}),\hat{\boldsymbol{\phi}}^b_\infty(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{y})\right] \tag{2.6}$$

is the MRBB function in the class $\mathcal{F}^b[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{y}),\hat{\boldsymbol{\mu}}(\boldsymbol{y}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{y})]$ if the following conditions hold:
(i) the support of $\boldsymbol{\theta}$ is $(-\infty,\infty)\times\cdots\times(-\infty,\infty)$;
(ii)

$$P\left((\boldsymbol{Y},\boldsymbol{\theta})\in\left\{(\boldsymbol{y},\boldsymbol{\theta}^*):\left|\left|\boldsymbol{U}(\boldsymbol{y})\hat{\boldsymbol{\phi}}^b_\infty(\boldsymbol{y},\boldsymbol{\theta}^*)\right|\right|=c(\boldsymbol{y},\boldsymbol{\theta}^*)\right\}\right)=0;$$

(iii) $\partial^2 f_{\theta|y}(\boldsymbol{\theta}|\boldsymbol{Y})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t$ exists almost surely;
(iv) $E(|\dot{c}_{\theta_i}|)<\infty$ and $E[|\partial(\dot{f}_{\theta|y,\theta_i}/f_{\theta|y})/\partial\theta_j|]<\infty$ for $i,j=1,\ldots,q$;
(v) the elements of the $m\times m$ matrix $\boldsymbol{U}(\boldsymbol{Y})$, the $m\times q$ matrix $\boldsymbol{M_1}(\boldsymbol{Y})$, and the $m\times 1$ vector $\boldsymbol{M_0}(\boldsymbol{Y})$ are bounded random variables;
(vi) $E[(\dot{f}_{\theta|y,\theta_i}/f_{\theta|y})^2]<\infty$.
(b) The function given in the expression (2.6) is the MRBB function in the class $\mathcal{F}^b[c(\boldsymbol{y},\boldsymbol{\theta}),\boldsymbol{U}(\boldsymbol{y}),\hat{\boldsymbol{\mu}}(\boldsymbol{y}),\hat{\boldsymbol{\mu}}_1(\boldsymbol{y})]$ if the following conditions hold:
(i) the support of $\boldsymbol{\theta}$ is the product of bounded intervals;
(ii) the conditions (ii), (iii), (iv), (v), and (vi) in (a) hold;
(iii) $f_{\theta|y}\to 0$ as $\theta_i\to c_i$ or $\theta_i\to d_i$.
(c) Let

$$\hat{\boldsymbol{\phi}}^{\Delta b}_\infty(\boldsymbol{y},\boldsymbol{\theta})=\boldsymbol{M}_1(\boldsymbol{y})\left[\frac{\Delta_\theta(\boldsymbol{f}_{\theta|y})}{f_{\theta|y}}\right]+\boldsymbol{M}_0(\boldsymbol{y}),$$

if $\boldsymbol{f}_{\theta|y} > 0$ *and* $\hat{\phi}^{\Delta b}_{\infty}(\boldsymbol{y}, \boldsymbol{\theta}) = 0$ *otherwise. The function*

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{h}\left\{c(\boldsymbol{y}, \boldsymbol{\theta}), \hat{\boldsymbol{\phi}}^{\Delta b}_{\infty}(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{y})\right\}$$

is the MRBB function in the class $\mathcal{F}^{\Delta b}[c(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{y}), \hat{\boldsymbol{\mu}}(\boldsymbol{y}), \hat{\boldsymbol{\mu}}_1(\boldsymbol{y})]$ *if the following conditions hold:*
(i) θ_i *take values in the set* $\{0, 1, 2, \cdots\}$*;*
(ii) $E\{\Delta^{+}_{\theta_j}[\Delta_{\theta_i}(f_{\theta|y})/f_{\theta|y}]\} < \infty$*;*
(iii) the condition (v) given in (a) holds;
(iv) $E\{[\Delta_{\theta_i}(f_{\theta|y})/f_{\theta|y}]^2\} < \infty$.

The abbreviated notations for the two classes $\mathcal{F}^{b}[c(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{y}), \hat{\boldsymbol{\mu}}(\boldsymbol{y}), \hat{\boldsymbol{\mu}}_1(\boldsymbol{y})]$ and $\mathcal{F}^{\Delta b}[c(\boldsymbol{y}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{y}), \hat{\boldsymbol{\mu}}(\boldsymbol{y}), \hat{\boldsymbol{\mu}}_1(\boldsymbol{y})]$ are $\mathcal{F}^{b}$ and $\mathcal{F}^{\Delta b}$, respectively. Since the proofs of this theorem are analogous to Theorem 2.3, the proofs are not presented in the last section. The above theorem can be applied to the exponential families, as given by the following corollary.

Corollary 2.5. *Let* $\boldsymbol{Y}$ *have the density function given in the expression (2.3). If the support of* $\boldsymbol{\theta}$ *is* $(c_1, d_1) \times \cdots \times (c_q, d_q)$*,* $\dot{\boldsymbol{f}}_{\theta|y,\theta}/f_{\theta|y}$ *in the MRBB function of Theorem 2.5 is*

$$\frac{\dot{\boldsymbol{f}}_{\theta|y,\theta}}{f_{\theta|y}} = \boldsymbol{T}(\boldsymbol{y}) - \dot{\boldsymbol{A}}_{\theta} + \frac{\partial\left\{\log\left[\pi(\boldsymbol{\theta})\right]\right\}}{\partial\boldsymbol{\theta}}.$$

Note that in the exponential families, the difference between the Bayes model and frequentist mainly depends on the logarithm of the prior distribution, as indicated by the following equation:

$$\frac{\dot{\boldsymbol{f}}_{\theta|y,\theta}}{f_{\theta|y}} = \frac{\dot{\boldsymbol{f}}_{\theta}}{f} + \frac{\partial\left\{\log\left[\pi(\boldsymbol{\theta})\right]\right\}}{\partial\boldsymbol{\theta}}.$$

2.4. Other properties of MRB and MRBB functions

In this section, some theoretical properties concerning the invariance of the MRB functions and component-wise MRB functions are given. The following corollaries can be obtained based on Theorem 2.1, Theorem 2.3, Theorem 2.4, and Theorem 2.5. The first corollary indicates the shift "invariance" of the MRB functions.

Corollary 2.6. *Let* $\mathcal{F}^{j} + \boldsymbol{\imath}(\boldsymbol{y}, \boldsymbol{\theta})$ *be the classes with elements* $\boldsymbol{\phi} + \boldsymbol{\imath}(\boldsymbol{y}, \boldsymbol{\theta})$*, where* $\boldsymbol{\phi} \in \mathcal{F}^{j}$*,* $\boldsymbol{\imath}(\boldsymbol{y}, \boldsymbol{\theta})$ *is a bounded function almost surely, and* $j \in \{n, y, \Delta y, \theta, b, \Delta b\}$*. Then, if the corresponding assumptions in Theorem 2.1, Theorem 2.3, Theorem 2.4, or Theorem 2.5 hold,* $\hat{\boldsymbol{\phi}} + \boldsymbol{\imath}(\boldsymbol{y}, \boldsymbol{\theta})$ *are the MRB functions in the classes* $\mathcal{F}^{j} + \boldsymbol{\imath}(\boldsymbol{y}, \boldsymbol{\theta})$*, where* $\hat{\boldsymbol{\phi}}$ *are the MRB functions given in these theorems.*

As the weight matrix is block diagonal, the decomposition of the MRB function is given by the following corollary. Further, the corollary implies that the MRB function can be the combination of the MRB functions in different classes.

Corollary 2.7. *Let $\boldsymbol{\phi} = (\boldsymbol{\phi}_{\boldsymbol{a}}^{\boldsymbol{t}}, \boldsymbol{\phi}_{\boldsymbol{b}}^{\boldsymbol{t}})^t$ and $\boldsymbol{W}(\boldsymbol{\theta})$ is a block diagonal matrix with the matrices $\boldsymbol{W}_{\boldsymbol{a}}(\boldsymbol{\theta})$ and $\boldsymbol{W}_{\boldsymbol{b}}(\boldsymbol{\theta})$ in the diagonal, where $\boldsymbol{W}_{\boldsymbol{a}}(\boldsymbol{\theta})$ and $\boldsymbol{W}_{\boldsymbol{b}}(\boldsymbol{\theta})$ are the matrices of which orders equal to the dimensions of $\boldsymbol{\phi}_{\boldsymbol{a}}$ and $\boldsymbol{\phi}_{\boldsymbol{b}}$, respectively. If the corresponding assumptions in Theorem 2.1, Theorem 2.3, or Theorem 2.4 hold, the MRB functions in the classes with elements satisfying $\boldsymbol{\phi}_{\boldsymbol{a}} \in \mathcal{F}^j + \boldsymbol{\imath}_{\boldsymbol{a}}(\boldsymbol{y}, \boldsymbol{\theta})$ and $\boldsymbol{\phi}_{\boldsymbol{b}} \in \mathcal{F}^j + \boldsymbol{\imath}_{\boldsymbol{b}}(\boldsymbol{y}, \boldsymbol{\theta})$ are*

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\theta}) = \left[\hat{\boldsymbol{\phi}}_a^t(\boldsymbol{y}, \boldsymbol{\theta}), \hat{\boldsymbol{\phi}}_b^t(\boldsymbol{y}, \boldsymbol{\theta})\right]^t,$$

where $\boldsymbol{\imath}_{\boldsymbol{a}}(\boldsymbol{y}, \boldsymbol{\theta})$ and $\boldsymbol{\imath}_{\boldsymbol{b}}(\boldsymbol{y}, \boldsymbol{\theta})$ are bounded functions almost surely, $j \in \{n, y, \Delta y, \theta\}$, and $\hat{\boldsymbol{\phi}}_a(\boldsymbol{y}, \boldsymbol{\theta})$ and $\hat{\boldsymbol{\phi}}_b(\boldsymbol{y}, \boldsymbol{\theta})$ are the corresponding MRB functions in these classes.

Note that the analogous corollary can be obtained for the MRBB functions.

2.5. MR and MRB functions

In this section, the results concerning the interrelation of the MR function and the score function along with the one of the MR function and the MRB function are established. Let the classes of functions be

$$\mathcal{F}_{\infty}^{j}\left[\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})\right] = \mathcal{F}^{j}\left[\infty, \boldsymbol{U}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}(\boldsymbol{\theta}), \hat{\boldsymbol{\mu}}_1(\boldsymbol{\theta})\right],$$

where $j \in \{y, \Delta y, \theta, b, \Delta b\}$. In the classes $\mathcal{F}_{\infty}^{j}$, the MR functions are $\hat{\boldsymbol{\phi}}_{\infty}^{y}(\boldsymbol{y}, \boldsymbol{\theta})$, $\hat{\boldsymbol{\phi}}_{\infty}^{\Delta y}(\boldsymbol{y}, \boldsymbol{\theta})$, $\hat{\boldsymbol{\phi}}_{\infty}^{\theta}(\boldsymbol{y}, \boldsymbol{\theta})$, $\hat{\boldsymbol{\phi}}_{\infty}^{b}(\boldsymbol{y}, \boldsymbol{\theta})$, and $\hat{\boldsymbol{\phi}}_{\infty}^{\Delta b}(\boldsymbol{y}, \boldsymbol{\theta})$, respectively, provided that the corresponding assumptions given in Theorem 2.3, Theorem 2.4, and Theorem 2.5 hold. The following theorem indicates that the score function is also a MR function.

Theorem 2.6. *Suppose that $E[\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})]$ can be differentiated with respect to $\boldsymbol{\theta}$ under the integral sign for any $\boldsymbol{\phi}$ in the class $\mathcal{F}_{\infty}^{\theta}[0, -\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})]$ and the elements of $\boldsymbol{W}(\boldsymbol{\theta})$ are finite for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})$ is the information matrix. Then, the MR function in the class is the score function, i.e.,*

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\theta}) = \frac{\partial \log\left[f(\boldsymbol{y}|\boldsymbol{\theta})\right]}{\partial \boldsymbol{\theta}} = \frac{\dot{\boldsymbol{f}}_{\theta}}{f},$$

if $f > 0$ and $\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\theta}) = 0$ otherwise.

The following theorem indicates that a subsequence of estimators obtained by employing the MRB functions converges to the corresponding one by employing the MR functions with probability one.

Theorem 2.7. *Let* $\boldsymbol{Y_n} = (Y_1, \ldots, Y_n)^t$ *be a vector of random variables with the density function* f_n *and* $\stackrel{p}{\longrightarrow}$ *denotes the convergence in probability.*
(a) *There exists a subsequence of* $\hat{\boldsymbol{\theta}}_t^j(\boldsymbol{Y_n})$*, the solution of*

$$\hat{\boldsymbol{\phi}}_t^j(\boldsymbol{Y_n}, \boldsymbol{\theta}) = \boldsymbol{h}\left[C_t^j, \hat{\boldsymbol{\phi}}_\infty^j(\boldsymbol{Y_n}, \boldsymbol{\theta}), \boldsymbol{U}(\boldsymbol{\theta})\right] = 0, j \in \{y, \Delta y, \theta\}, t = 1, \ldots,$$

converging to $\boldsymbol{\theta}_0$ *with probability one if the following conditions hold:*
(i) $\boldsymbol{\theta}_0$ *is an interior point in some open subset* $\boldsymbol{\Omega}$ *of* $\boldsymbol{\Theta}$*;*
(ii) *On the set* $\boldsymbol{\Omega}$*,* $\boldsymbol{f}_n$ *and* $\dot{\boldsymbol{f}}_{n,y_n}$ *or* $\boldsymbol{f}_n$ *and* $\Delta_{y_n}(\boldsymbol{f}_n)$ *are continuous almost surely as* $j \in \{y, \Delta y\}$*, while* $\dot{\boldsymbol{f}}_{n,\theta}$ *is continuous almost surely as* $j = \theta$*;*
(iii) *there exist solutions* $\hat{\boldsymbol{\theta}}_\infty^j(\boldsymbol{Y_n})$ *of* $\hat{\boldsymbol{\phi}}_\infty^j(\boldsymbol{Y_n}, \boldsymbol{\theta}) = 0$ *and*

$$\hat{\boldsymbol{\theta}}_\infty^j(\boldsymbol{Y_n}) \stackrel{p}{\longrightarrow} \boldsymbol{\theta}_0;$$

(iv) $C_t^j \to \infty$*;*
(v) *the elements of the matrix functions* $\boldsymbol{M}_1(\boldsymbol{\theta})$*,* $\boldsymbol{M}_0(\boldsymbol{\theta})$*, and* $\boldsymbol{U}(\boldsymbol{\theta})$ *are continuous on* $\boldsymbol{\Omega}$.
(b) *The convergence also holds as the estimator of* $\boldsymbol{\theta}$ *is the minimum of the* ρ *function if the conditions given in (a) hold.*

The theorem analogous to the above one can be also obtained for the MRBB functions. If the MR function is the score function and the corresponding maximum likelihood estimator (MLE) is consistent for estimating the parameter (see Lehmann and Casella 1998, pp. 461-465), the above theorem implies that there exists a subsequence of estimators obtained by employing the bounded score functions converging to the parameter with probability one.

2.6. Generalized Huber operator

The generalized Huber function is defined on the finite dimensional space. However, as the random variable of interest takes values in an infinite dimensional space, a generalized Huber operator can be considered, i.e., a nonlinear map $\boldsymbol{h}(c, \boldsymbol{v}, \boldsymbol{U})$: $D(\boldsymbol{h}) \to \mathcal{H}_2$,

$$\boldsymbol{h}(c, \boldsymbol{v}, \boldsymbol{U}) = \boldsymbol{v} \min\left[1, \frac{||c||_{\mathcal{H}_1}}{||\boldsymbol{U}\boldsymbol{v}||_{\mathcal{H}_4}}\right],$$

if $||\boldsymbol{U}\boldsymbol{v}||_{\mathcal{H}_4} \neq \boldsymbol{0}$ and $\boldsymbol{h}(c, \boldsymbol{v}, \boldsymbol{U}) = \boldsymbol{v}$ if $||\boldsymbol{U}\boldsymbol{v}||_{\mathcal{H}_4} = \boldsymbol{0}$, where the domain $D(\boldsymbol{h})$ of the operator $\boldsymbol{h}$ is $\mathcal{B}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$, $\mathcal{B}_1$ consisting of bounded functions defined on $R^p \times R^q$ is the bounded subset of the normed space $\mathcal{H}_1$, $\mathcal{H}_2$ is the normed space consisting of $r \times 1$ vector functions defined on $R^p \times R^q$, and $\mathcal{H}_3$ is the normed space consisting of $m \times r$ matrix functions defined on $R^p \times R^q$, $\boldsymbol{U}\boldsymbol{v}$ is assumed to fall in the normed space $\mathcal{H}_4$ consisting of $m \times 1$ vector functions defined on $R^p \times R^q$, and $|| \cdot ||_{\mathcal{H}}$ is denoted as the norm in the normed space $\mathcal{H}$. For given $\boldsymbol{U}$ and $\boldsymbol{c}$, $||\boldsymbol{h}(c, \boldsymbol{v}, \boldsymbol{U})||_{\mathcal{H}_2}$ is bounded for all $\boldsymbol{v} \in \mathcal{H}_2$ if there exists a positive constant k satisfying $||\boldsymbol{U}\boldsymbol{v}||_{\mathcal{H}_4} \geq k||\boldsymbol{v}||_{\mathcal{H}_2}$ for all $\boldsymbol{v} \in \mathcal{H}_2$. The other type of generalized Huber operator is $\boldsymbol{h}(c, \boldsymbol{v}) : \mathcal{B}_1 \times \mathcal{H}_2 \to \mathcal{H}_2$,

$$\boldsymbol{h}(c, \boldsymbol{v}) = \boldsymbol{v} \min\left[1, \frac{||c||_{\mathcal{H}_1}}{||U(\boldsymbol{v})||_{\mathcal{H}_4}}\right],$$

if $||U(\boldsymbol{v})||_{\mathcal{H}_4} \neq 0$ and $\boldsymbol{h}(c, \boldsymbol{v}) = \boldsymbol{v}$ if $||U(\boldsymbol{v})||_{\mathcal{H}_4} = 0$, where $U : \mathcal{H}_2 \to \mathcal{H}_4$ is considered as a continuous linear operator. $\boldsymbol{h}(c, \boldsymbol{v})$ can be a bounded operator by the following lemma. Denote the notation $< \cdot, \cdot >_{\mathcal{H}}$ as the inner product in the inner product space $\mathcal{H}$.

Lemma 2.1. *$\boldsymbol{h}(c, \boldsymbol{v})$ is a bounded operator, i.e., $||\boldsymbol{h}(c, \boldsymbol{v})||_{\mathcal{H}_2}$ is bounded by a positive constant for all $c \in \mathcal{B}_1$ and all $\boldsymbol{v} \in \mathcal{H}_2$, if the following conditions hold:*
(i) $\mathcal{H}_2$ and $\mathcal{H}_4$ are Hilbert spaces and $||\boldsymbol{x}||_{\mathcal{H}_4} = (< \boldsymbol{x}, \boldsymbol{x} >_{\mathcal{H}_4})^{1/2}$ for $\boldsymbol{x} \in \mathcal{H}_4$;
(ii) there exists a positive constant k such that $< \boldsymbol{v}, (U^*U)(\boldsymbol{v}) >_{\mathcal{H}_2} \geq k\|\boldsymbol{v}\|^2_{\mathcal{H}_2}$ for all $\boldsymbol{v} \in \mathcal{H}_2$, where U^* is the Hilbert-adjoint operator of U and U^*U is the composition of the linear operators U^* and U.

Condition (ii) is closely related to the continuous linear $\mathcal{H}_2$-elliptic operator (see Aubin 2000, p. 64). In fact, the lemma is still true with condition (ii) replaced by the following condition:
(ii)* the continuous linear operator $g : \mathcal{H}_2 \to \mathcal{H}_2^*$ defined by $g(\boldsymbol{v}) = f_v$ is $\mathcal{H}_2$-elliptic, where $f_v(\boldsymbol{u}) =< \boldsymbol{u}, \boldsymbol{u}^* >_{\mathcal{H}_2}$, $\boldsymbol{u}^* = (U^*U)(\boldsymbol{v})$, and $\mathcal{H}_2^*$ is the dual space of $\mathcal{H}_2$.

Note that $(U^*U)(\boldsymbol{v})$ is the representer of f_v in $\mathcal{H}_2$. If U is the identity operator, the above lemma holds and the generalized Huber operator is the generalization of the Huber function given in Hampel et al. (1986, p. 239) to the infinite dimensional Hilbert space. The generalized Huber operator can be used for the robustness of the functions in the infinite dimensional Hilbert space, for example, the functions of interest in nonparametric regression.

3. Applications

Based on the results in the previous section, the proposed MRB and MRBB functions are applied for diverse types of data and a variety of regression models.

3.1. Robust location and scale estimation

3.1.1. Normal MRB and MRBB functions

Let Y have a univariate normal distribution with mean ν and variance σ^2. In the location parameter problem, the M-functional is the solution of $E[\psi(Y-\nu)] = 0$, where ψ is an odd, non-decreasing, not identically zero function. The influence functions

$$\phi(y, \nu) = \psi(y - \nu)/E[\dot{\psi}_y(Y - \nu)]$$

are of interest. The optimal bounded influence function, which minimizes the asymptotic variance, is

$$\hat{\phi}(y, \nu) = h\{k/[2\Phi(k) - 1], (y - \nu)/[2\Phi(k) - 1], 1\},$$

where k is a constant and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable. The optimal bounded influence function is the MRB function in the class $\mathcal{F}^n\{k/[2\Phi(k) - 1], 1, 0, 1\}$. The corresponding M-estimator is a type of Winsorized mean (see Serfling 1980, p. 247). In addition, the estimator minimizing the bounded residual sum of squares $\sum_{i=1}^n \hat{\phi}^2(Y_i, \nu)$ is a type of trimmed mean (also see Serfling 1980, pp. 246-248; Staudte and Sheather 1990, p. 115).

By Corollary 2.3 and Corollary 2.7, the MRB function for the location-dispersion parameters problem can be

$$\begin{aligned}\hat{\boldsymbol{\phi}}(y, \nu, \sigma^2) \;=\; & \left\{h\left[c_a(y, \nu, \sigma^2), y - \nu, 1\right],\right. \\ & \left. h\left[c_b(y, \nu, \sigma^2), \frac{n(y-\nu)^2}{n-1} - \sigma^2, 1\right]\right\}^t,\end{aligned}$$

where c_a and c_b are bounded functions almost surely and n is the number of observations and assumed to be known. Note that $y - \nu$ is an unbiased estimating function but $n(y - \nu)^2/(n - 1) - \sigma^2$ is a biased estimating function. However, the estimators obtained from the equations $\sum_{i=1}^n (Y_i - \nu) = 0$ and $\sum_{i=1}^n [n(Y_i - \nu)^2/(n - 1) - \sigma^2] = 0$ are the sample mean and sample variance, respectively. The solutions of $\sum_{i=1}^n \hat{\boldsymbol{\phi}}(Y_i, \nu, \sigma^2) = 0$, $\hat{\nu}$ and $\hat{\sigma}^2$, are two types of Winsorized means based on $Y_i - \hat{\nu}$ and $n(Y_i - \hat{\nu})^2/(n-1) - \hat{\sigma}^2$, respectively. As $c_b(y, \nu, \sigma^2) = C_b$ is a constant function, the above estimating function $\hat{\boldsymbol{\phi}}$ is referred to as the biased estimating function owing to $E\{h[C_b, n(Y - \nu)^2/(n - 1) - \sigma^2, 1]\} \neq 0$. Another sensible MRB function

$$\begin{aligned}\hat{\boldsymbol{\phi}}(y, \nu, \sigma^2) \;=\; & \left\{h\left[c(y, \nu, \sigma^2), \frac{y-\nu}{\sigma}, 1\right],\right. \\ & \left. h\left\{\left[c(y, \nu, \sigma^2)\right]^2, \left(\frac{y-\nu}{\sigma}\right)^2, 1\right\} - \imath(\nu, \sigma^2)\right\}^t\end{aligned}$$

can be obtained by Corollary 2.3, Corollary 2.6, and Corollary 2.7, where $\imath(\nu, \sigma^2)$ is a function satisfying $E\{h\{[c(Y, \nu, \sigma^2)]^2, [(Y - \nu)/\sigma]^2, 1\}\} = \imath(\nu, \sigma^2)$. The estimating equations based on the MRB function are in fact the Huber's proposal (see Huber and Ronchetti 2009, Section 6.7; Staudte and Sheather 1990, Section 4.5) as $c(y, \nu, \sigma^2) = C$ is a constant function. Finally, by Theorem 2.4, the other sensible MRB function can thus be

$$\hat{\boldsymbol{\phi}}(y, \nu, \sigma^2) = \boldsymbol{h}\{c(y, \nu, \sigma^2), [y - \nu, (y - \nu)^2 - \sigma^2]^t, \boldsymbol{I}\}.$$

If the random vector $\boldsymbol{Y}$ has the multivariate normal distribution with mean vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)^t$ and known variance-covariance matrix $\boldsymbol{\Sigma}$, the MRB function

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\nu}) = \boldsymbol{h}\left[c(\boldsymbol{y}, \boldsymbol{\nu}), \boldsymbol{y} - \boldsymbol{\nu}, \boldsymbol{\Sigma}^{-1/2}\right]$$

can be obtained by Theorem 2.1. Thus, if $c(\boldsymbol{y}, \boldsymbol{\nu}) = C$ and the unbiased M-estimating function $\boldsymbol{\Sigma}^{-1/2}\hat{\phi}(\boldsymbol{y}, \boldsymbol{\nu})$ is used to estimate $\boldsymbol{\nu}$, the estimator $\hat{\boldsymbol{\nu}}$ is the proposed multivariate Winsorized mean of the standardized residuals $\boldsymbol{\epsilon}_i^* = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{Y}_i - \hat{\boldsymbol{\nu}}),\ i = 1, \ldots, n$, i.e., the sample mean vector of the modified $\boldsymbol{Y}_i$'s, where the original $\boldsymbol{Y}_i$ is replaced by $\hat{\boldsymbol{\nu}} + C(\boldsymbol{Y}_i - \hat{\boldsymbol{\nu}})/||\boldsymbol{\epsilon}_i^*||$ if $||\boldsymbol{\epsilon}_i^*|| > C$. Note that $C(\boldsymbol{Y}_i - \hat{\boldsymbol{\nu}})/||\boldsymbol{\epsilon}_i^*||$ is a bounded random vector. If the variance-covariance matrix $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$ and the prior densities for independent ν_i are normally distributed with mean ξ and variance τ^2, the MRBB function

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}_b, \boldsymbol{\nu}) = \boldsymbol{h}\left[c(\boldsymbol{y}, \boldsymbol{\nu}), \boldsymbol{y}_b - \boldsymbol{\nu}, \boldsymbol{I}\right]$$

can be obtained by Corollary 2.5, where the posterior mean $\boldsymbol{y}_b = [\tau^2/(\sigma^2 + \tau^2)]\boldsymbol{y} - [\sigma^2\xi/(\sigma^2 + \tau^2)]\boldsymbol{l}$ can be thought as the "new" observed data by incorporating with the prior information, and where $\boldsymbol{l}$ is a $p \times 1$ vector with all elements equal to 1. Thus, if $c(\boldsymbol{y}, \boldsymbol{\nu}) = C$ and the unbiased M-estimating function $\hat{\phi}(\boldsymbol{y}_b, \boldsymbol{\nu})$ is used to estimate $\boldsymbol{\nu}$, the estimator $\hat{\boldsymbol{\nu}}$ is the multivariate Winsorized mean based on $\boldsymbol{Y}_{bi} - \hat{\boldsymbol{\nu}}$.

3.1.2. Binomial MRB and MRBB functions

Let the random variable Y have the binomial distribution with mean $m\theta$, where θ is the probability of one success and m is the number of trials. The MRB function

$$\hat{\phi}(y, \theta) = h\left[c(y, \theta), \overline{y} - \theta, 1\right] \tag{3.1}$$

can be obtained by Corollary 2.3, where $\overline{y} = y/m$. The robust estimator $\hat{\theta}$ minimizing the bounded residual sum of squares $\sum_{i=1}^n \hat{\phi}^2(Y_i, \theta)$ with $c(\boldsymbol{y}, \boldsymbol{\theta}) = C$ is a type of trimmed mean based on $\overline{Y}_i - \hat{\theta}$, where $\overline{Y}_i = Y_i/m$. If the biased M-estimating function $\hat{\phi}(y, \theta)$ is used to estimate θ, the corresponding estimator is a type of Winsorized mean based on $\overline{Y}_i - \hat{\theta}$. In addition, another sensible MRB function based on the ratio rather than the difference

$$\hat{\phi}(y, \theta) = h\left[c(y, \theta), \frac{y}{m\theta}\left(\frac{m - y + 1}{m - m\theta + 1}\right)^{-1} - 1, 1\right]$$

can be obtained by Corollary 2.2.

If the prior distribution for θ is assumed to be the beta distribution (see Lehmann and Casella 1998, p. 25) with mean $\alpha/(\alpha+\beta)$ and variance $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$, the MRBB function

$$\hat{\phi}(y, \theta) = h\left[c(y, \theta), y_b - \theta, 1\right] \tag{3.2}$$

can be obtained by Corollary 2.5 and Corollary 2.6, where the posterior mean $y_b = (y+\alpha)/(m+\alpha+\beta)$ can be thought as the "new" observed data by incorporating with the prior information.

3.1.3. Poisson MRB and MRBB functions

Let the random variable Y have the Poisson distribution with mean θ. The MRB function

$$\hat{\phi}(y,\theta) = h\left[c(y,\theta), y-\theta, 1\right] \tag{3.3}$$

can be obtained by Corollary 2.3. If $c(y,\theta) = C$, the robust estimator $\hat{\theta}$ minimizing the bounded residual sum of squares $\sum_{i=1}^{n} \hat{\phi}^2(Y_i,\theta)$ is a type of trimmed mean based on $Y_i - \hat{\theta}$. If the biased M-estimating function $\hat{\phi}(y,\theta)$ is used to estimate θ, the corresponding estimator is a type of Winsorized mean based on $Y_i - \hat{\theta}$. On the other hand, another sensible MRB function based on the ratio rather than the difference

$$\hat{\phi}(y,\theta) = h\left[c(y,\theta), \frac{y}{\theta} - 1, 1\right]$$

can be obtained by Corollary 2.2.

If the prior distribution for θ is assumed to be the gamma distribution with mean $\alpha\beta$ and variance $\alpha\beta^2$, the MRBB function

$$\hat{\phi}(y,\theta) = h\left[c(y,\theta), y_b - \theta, 1\right] \tag{3.4}$$

can be obtained by Corollary 2.5 and Corollary 2.6, where the posterior mean $y_b = (y+\alpha)/(1+\beta^{-1})$ can be thought as the "new" observed data by incorporating with the prior information.

3.1.4. Robust mean vector and covariance matrix

Let the random vector $\boldsymbol{Y}$ have the multivariate normal distribution with mean vector $\boldsymbol{\nu}$ and unknown variance-covariance matrix $\boldsymbol{\Sigma}$ in the location-scale family. Note that the scale parameter matrix is $\boldsymbol{B} = \boldsymbol{\Sigma}^{-1/2}$. Then,

$$\frac{\dot{\boldsymbol{f}}_{\nu}}{f} = \boldsymbol{B}^2(\boldsymbol{y} - \boldsymbol{\nu})$$

and

$$\frac{\frac{\partial f(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{B})}{\partial \boldsymbol{B}}}{f(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{B})} = \boldsymbol{Q}(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B}) + \boldsymbol{Q}^t(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B}),$$

where $\boldsymbol{Q}(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B}) = \boldsymbol{L} - (1/2)diag(\boldsymbol{L})$ and $\boldsymbol{L} = \boldsymbol{B}[\boldsymbol{B}^{-2} - (\boldsymbol{y}-\boldsymbol{\nu})(\boldsymbol{y}-\boldsymbol{\nu})^t]$. The estimators of $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma} = \boldsymbol{B}^{-2}$ based on the unbiased estimating equations

$$\sum_{i=1}^{n} \boldsymbol{B}^2(\boldsymbol{Y}_i - \boldsymbol{\nu}) = 0$$

and

$$\sum_{i=1}^{n}[\boldsymbol{Q}(\boldsymbol{Y}_i,\boldsymbol{\nu},\boldsymbol{B})+\boldsymbol{Q}^t(\boldsymbol{Y}_i,\boldsymbol{\nu},\boldsymbol{B})]=0$$

are the sample mean vector $\overline{\boldsymbol{Y}}$ and the maximum likelihood estimator $\sum_{i=1}^{n}(\boldsymbol{Y}_i-\bar{\boldsymbol{Y}})(\boldsymbol{Y}_i-\bar{\boldsymbol{Y}})^t/n$. The MRB function

$$\begin{aligned}\hat{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B}) &= \boldsymbol{h}\{c(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B}),\\ &\qquad \left\{(\boldsymbol{y}-\boldsymbol{\nu})^t,\text{vecs}^t\left[\boldsymbol{Q}(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B})+\boldsymbol{Q}^t(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B})\right]\right\}^t,\boldsymbol{I}\Big\}\end{aligned}$$

can be obtained by Corollary 2.3. The other sensible MRB function

$$\begin{aligned}\hat{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B}) &= \left\{\boldsymbol{h}^t\left[c_a(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B}),\boldsymbol{y}-\boldsymbol{\nu},\boldsymbol{I}\right],\right.\\ &\qquad \left.\boldsymbol{h}^t\left\{c_b(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B}),\text{vecs}\left[\boldsymbol{Q}(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B})+\boldsymbol{Q}^t(\boldsymbol{y},\boldsymbol{\nu},\boldsymbol{B})\right],\boldsymbol{I}\right\}\right\}^t\end{aligned}$$

can be obtained by Corollary 2.3 and Corollary 2.7, where c_a and c_b are bounded functions almost surely.

3.1.5. Large sample approximation

Let $\boldsymbol{Y}_{ji}, j=1,\ldots,m, i=1,\ldots,n_j$, be independent $p\times 1$ random vectors with mean vector $\boldsymbol{\nu}$ and known variance-covariance matrix $\boldsymbol{\Sigma}$. Then, by central limit theorem, $\sqrt{n_j}(\overline{\boldsymbol{Y_j}}-\boldsymbol{\nu})\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Sigma})$, where $\overline{\boldsymbol{Y_j}}=(1/n_j)\sum_{i=1}^{n_j}\boldsymbol{Y}_{ji}$. By Theorem 2.2, the function of interest is $\boldsymbol{h}\{c(\overline{\boldsymbol{Y_j}},\boldsymbol{\nu},\boldsymbol{\Sigma}),\sqrt{n_j}(\overline{\boldsymbol{Y_j}}-\boldsymbol{\nu}),\boldsymbol{\Sigma}^{-1/2}\}$.

3.2. Robust regression

3.2.1. Robust linear regression

Consider first the model

$$Y_i=\boldsymbol{X_i}\boldsymbol{\beta}+\epsilon_i,\ i=1,\ldots,n,$$

where $\boldsymbol{Y}=(Y_1,\ldots,Y_n)^t$ are observations, $\boldsymbol{X_i}=(X_{i1},X_{i2},\ldots,X_{ip})$ are the values of the covariates, $\boldsymbol{\epsilon}=(\epsilon_1,\ldots,\epsilon_n)^t$ are zero mean, correlated normal random errors with variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$, and where $\boldsymbol{\alpha}$ is the correlation parameter. By Theorem 2.1, the robust least squares estimators for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are the minimizers of $\hat{\boldsymbol{\phi}}^t(\boldsymbol{Y},\boldsymbol{\alpha},\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\hat{\boldsymbol{\phi}}(\boldsymbol{Y},\boldsymbol{\alpha},\boldsymbol{\beta})$, where

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{\beta})=\boldsymbol{h}[c(\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{\beta}),\boldsymbol{e},\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\alpha})]$$

is the MRB function in the class $\mathcal{F}^n$, $\boldsymbol{e}=(e_1,\ldots,e_n)^t$, and $e_i=y_i-\boldsymbol{X_i}\boldsymbol{\beta}$. Another robust estimators based on Theorem 2.1 and Corollary 2.7 can be the minimizers of $\hat{\boldsymbol{\phi}}^t(\boldsymbol{Y},\boldsymbol{\alpha},\boldsymbol{\beta})\hat{\boldsymbol{\phi}}(\boldsymbol{Y},\boldsymbol{\alpha},\boldsymbol{\beta})$, where

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{\beta})=[\hat{\phi}_1(\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{\beta}),\ldots,\hat{\phi}_n(\boldsymbol{y},\boldsymbol{\alpha},\boldsymbol{\beta})]^t,$$

and where

$$\hat{\phi}_i(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = h\left[c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}), e_i^*, 1\right] \tag{3.5}$$

is the MRB function in the class $\mathcal{F}^n$ and $\boldsymbol{e}^* = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\alpha})\boldsymbol{e} = (e_1^*, \ldots, e_n^*)^t$. If $c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \min[C, q_{1-\zeta}(|\boldsymbol{e}^*|)]$ with large values of C, the corresponding estimate is the least Winsorized squares (LWS) estimate (see Rousseeuw and Leroy 1987, p. 135) under correlated errors. If $c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \min[C_1, C_2 q_{0.5}(|\boldsymbol{e}^*|)]$ with large values of C_1 and small values of C_2, the corresponding estimate is the least median squares (LMS) estimate (see Rousseeuw 1984) under correlated errors. If the MRB function in the class $\mathcal{F}^n + \imath(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is used and $c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \min[C, q_{1-\zeta}(|\boldsymbol{e}^*|)]$ with large values of C, the corresponding estimate is the least trimmed squares (LTS) estimate (see Rousseeuw and Leroy 1987, p. 15) under correlated errors, where

$$\imath(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} 0, & |e_i^*| \le c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}), \\ -c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}), & |e_i^*| > c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}). \end{cases}$$

The other robust estimators can be the minimizers of $\boldsymbol{\psi}^t(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\psi}(\boldsymbol{Y}, \boldsymbol{\alpha}, \boldsymbol{\beta})$, where

$$\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = [\psi_1(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}), \ldots, \psi_n(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta})]^t$$

and

$$\psi_i(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} (e_i^*)^2/2, & |e_i^*| \le c(\boldsymbol{\alpha}, \boldsymbol{\beta}), \\ c(\boldsymbol{\alpha}, \boldsymbol{\beta})|e_i^*| - \left[c(\boldsymbol{\alpha}, \boldsymbol{\beta})\right]^2/2, & |e_i^*| > c(\boldsymbol{\alpha}, \boldsymbol{\beta}). \end{cases}$$

Note that

$$\hat{\phi}_i(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = d\psi_i/de_i^* = h\left[c(\boldsymbol{\alpha}, \boldsymbol{\beta}), e_i^*, 1\right]$$

is the MRB function by Theorem 2.1 and Corollary 2.7. If the variance of the uncorrelated random errors and $c(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are constants, the estimator for $\boldsymbol{\beta}$ is the one proposed by Huber (1973).

3.2.2. Robust regression model selection

Consider the model

$$Y_i = f(\boldsymbol{t_i}) + \epsilon_i, \ i = 1, \ldots, n,$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^t$ are observations at design points $\boldsymbol{t_i} = (t_{i1}, t_{i2}, \ldots, t_{id})$, $f(\boldsymbol{t})$ is a function, and ϵ_i are zero mean, correlated normal random errors with variance-covariance matrix $\boldsymbol{\Sigma}$. Let $\hat{\boldsymbol{f}}(\boldsymbol{\lambda}) = \boldsymbol{H}(\boldsymbol{\lambda})\boldsymbol{y}$ be the vector of fitted values, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)$ is a set of parameters associated with the selection of the model and $\boldsymbol{H}(\boldsymbol{\lambda})$ is an $n \times n$ matrix. The parameter λ_j could be the subset of the discrete index set $\{1, 2, \ldots, p_j\}$ in linear regression, the correlation parameter for the variance-covariance matrix, the transformation parameter in Box-Cox transformation model (see Box and Cox 1964), or the selection parameter in multivariate nonparametric regression, for examples, the bandwidth in kernel-based method or the smoothing parameter in smoothing splines. Several criteria are commonly used in statistical analysis, including AIC (Akaike 1974), C_p (Mallows 1973), and FPE (Akaike 1970). For uncorrelated random errors, robust AIC and C_p criteria have been proposed by Ronchetti (1985) and Ronchetti and Staudte (1994), respectively. For correlated random errors, a class of regression model selection criteria in terms of data values, $\boldsymbol{e^t W}(\boldsymbol{\lambda})\boldsymbol{e}$, has been used in Wei (2008) to find the minimizer $\hat{\boldsymbol{\lambda}}$, where $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{f}}(\boldsymbol{\lambda}) = [\boldsymbol{I} - \boldsymbol{H}(\boldsymbol{\lambda})]\boldsymbol{y}$ is the vector of residuals and $\boldsymbol{W}(\boldsymbol{\lambda})$ is some positive definite weight matrix. By Theorem 2.1, a class of robust model selection criteria can be $\hat{\boldsymbol{\phi}}^{\boldsymbol{t}}(\boldsymbol{Y}, \boldsymbol{\lambda})\boldsymbol{W}(\boldsymbol{\lambda})\hat{\boldsymbol{\phi}}(\boldsymbol{Y}, \boldsymbol{\lambda})$, where

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\lambda}) = \boldsymbol{h}[c(\boldsymbol{y}, \boldsymbol{\lambda}), \boldsymbol{e}, \boldsymbol{U}(\boldsymbol{\lambda})]$$

is the MRB function in the class $\mathcal{F}^n$. Furthermore, another class of robust selection criteria based on Theorem 2.1 and Corollary 2.7 can be $\hat{\boldsymbol{\phi}}^t(\boldsymbol{Y}, \boldsymbol{\lambda})\hat{\boldsymbol{\phi}}(\boldsymbol{Y}, \boldsymbol{\lambda})$, where $\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\lambda}) = [\hat{\phi}_1(\boldsymbol{y}, \boldsymbol{\lambda}), \ldots, \hat{\phi}_n(\boldsymbol{y}, \boldsymbol{\lambda})]^t$, and where

$$\hat{\phi}_i(\boldsymbol{y}, \boldsymbol{\lambda}) = h\left[c(\boldsymbol{y}, \boldsymbol{\lambda}), e_i^*, 1\right]$$

is also the MRB function in the class $\mathcal{F}^n$ and $\boldsymbol{e^*} = \boldsymbol{U}(\boldsymbol{\lambda})\boldsymbol{e} = (e_1^*, \ldots, e_n^*)^t$.

The other class of robust selection criteria (also see Hampel et al. 1986, p. 367) is $\boldsymbol{\psi}^t(\boldsymbol{Y}, \boldsymbol{\lambda})\boldsymbol{\psi}(\boldsymbol{Y}, \boldsymbol{\lambda})$, where $\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\lambda}) = [\psi_1(\boldsymbol{y}, \boldsymbol{\lambda}), \ldots, \psi_n(\boldsymbol{y}, \boldsymbol{\lambda})]^t$,

$$\psi_i(\boldsymbol{y}, \boldsymbol{\lambda}) = \begin{cases} (e_i^*)^2/2, & |e_i^*| \le c(\boldsymbol{\lambda}), \\ c(\boldsymbol{\lambda})|e_i^*| - [c(\boldsymbol{\lambda})]^2/2, & |e_i^*| > c(\boldsymbol{\lambda}). \end{cases}$$

Note that

$$\hat{\phi}_i(\boldsymbol{y}, \boldsymbol{\lambda}) = d\hat{\psi}_i/de_i^* = h\left[c(\boldsymbol{\lambda}), e_i^*, 1\right]$$

is the MRB function by Theorem 2.1 and Corollary 2.7.

3.2.3. Robust IRLS (iterated reweighted least squares) for generalized linear models

Consider the standard generalized linear models (McCullagh and Nelder 1989) in which each component of the response vector has a distribution taking the form

$$f(y_i|\theta_i, \sigma) \quad = \quad \exp\left[\frac{y_i\theta_i - A(\theta_i)}{b(\sigma)} + d(y_i, \sigma)\right], \ i = 1, \ldots, n,$$

where θ_i and σ are scalar parameters and $A(\cdot)$, $b(\cdot)$, and $d(\cdot)$ are specific functions. The dependence of the response y_i on the associated explanatory variables $\boldsymbol{X_i}$ can be modeled through $\theta_i = \boldsymbol{X_i\beta}$, i.e., the canonical link. Further, $b(\sigma) = 1$ is assumed hereafter. The estimated (possible) values of the sth iteration $\hat{\boldsymbol{\beta}}_s$ is considered as the values of the "parameter" associated with the distribution of the responses. Then, $\dot{\boldsymbol{f}}_{\theta_s}/f = \boldsymbol{y} - \boldsymbol{u}(\hat{\boldsymbol{\beta}}_s)$, where $\boldsymbol{\theta}_s = \boldsymbol{X}\hat{\boldsymbol{\beta}}_s$, $\boldsymbol{u}(\hat{\boldsymbol{\beta}}_s)$ is the estimated mean of the response vector $\boldsymbol{Y}$, and where $\boldsymbol{X}$ is the covariate matrix. If the $(s+1)$th iteration $\hat{\boldsymbol{\beta}}_{s+1}$ is considered as the additional "parameter",

$$\boldsymbol{e}(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1}) = \left[e_1(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1}), \ldots, e_n(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1})\right]^t = \boldsymbol{Z}(\hat{\boldsymbol{\beta}}_s) - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{s+1}$$

and

$$\boldsymbol{Z}(\hat{\boldsymbol{\beta}}_s) = \boldsymbol{X}\hat{\boldsymbol{\beta}}_s + \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}}_s)\left[\boldsymbol{y} - \boldsymbol{u}(\hat{\boldsymbol{\beta}}_s)\right],$$

where $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_s)$ is the estimated variance-covariance matrix of the response vector $\boldsymbol{Y}$. Then, the MRB function

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1}) = \boldsymbol{h}\left[c(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1}), \boldsymbol{e}(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1}), \boldsymbol{\Sigma}^{1/2}(\hat{\boldsymbol{\beta}}_s)\right]$$

can be obtained by Corollary 2.3. Note that the $(s+1)$th IRLS estimate $\hat{\boldsymbol{\beta}}_{s+1}$ minimizes $\boldsymbol{e}^t(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1})\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_s)\boldsymbol{e}(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1})$. The IRLS estimate $\hat{\boldsymbol{\beta}}$ at convergence satisfies the equation $\boldsymbol{X}^t\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\boldsymbol{e}(\boldsymbol{y}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}) = 0$. Thus, the associated robust IRLS estimate $\hat{\boldsymbol{\beta}}_\phi$ at convergence satisfies the equation $\boldsymbol{X}^t\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_\phi)\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_\phi, \hat{\boldsymbol{\beta}}_\phi) = 0$. By Corollary 2.3 and Corollary 2.7, the other sensible MRB function is the $n \times 1$ vector with the i'th element equal to

$$\hat{\phi}_i(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1}) = h\left[c(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1}), e_i(\boldsymbol{y}, \hat{\boldsymbol{\beta}}_s, \hat{\boldsymbol{\beta}}_{s+1}), 1\right].$$

3.2.4. Robust multivariate linear regression

Consider the linear model

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \ i = 1, \ldots, n,$$

where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})^t$, $\boldsymbol{X}_i$ is a $p \times q$ design matrix, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^t$, $\boldsymbol{\epsilon}_i$ are multivariate normal random variables with zero mean vector and variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$, and where $\boldsymbol{\alpha}$ is the correlation parameter. If $\boldsymbol{\epsilon}_i$ are independent, the S-estimators have been proposed by Van Aelst and Willems (2005), while the least trimmed squares estimators have been given by Agulló et. al. (2008). Let $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$ and $\boldsymbol{e} = (\boldsymbol{e}_1^t, \ldots, \boldsymbol{e}_n^t)^t$. The MRB function

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{h}\left[c(\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}), \boldsymbol{e}_i, \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\alpha})\right]$$

can be obtained by Theorem 2.1. The robust least squares estimator can be obtained by minimizing the sum of squares $\sum_{i=1}^n \hat{\boldsymbol{\phi}}^t(\boldsymbol{Y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\hat{\boldsymbol{\phi}}(\boldsymbol{Y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Another sensible MRB functions are the generalizations of the ones given in Section 3.2.1. As $\boldsymbol{\epsilon}_i$ are correlated, let $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ be also denoted as the variance-covariance matrix of $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^t, \ldots, \boldsymbol{\epsilon}_n^t)^t$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1^t, \ldots, \boldsymbol{Y}_n^t)^t$. The MRB functions by Theorem 2.1 can be

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{h}[c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}), \boldsymbol{e}, \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\alpha})].$$

The other MRB function $\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = [\hat{\phi}_1(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}), \ldots, \hat{\phi}_n(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta})]^t$ can be obtained by Theorem 2.1 and Corollary 2.7, where

$$\hat{\phi}_i(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = h\left[c(\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}), \boldsymbol{e}_i^*, \boldsymbol{I}\right],$$

and $\boldsymbol{e}^* = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\alpha})\boldsymbol{e} = [(\boldsymbol{e}_1^*)^t, \ldots, (\boldsymbol{e}_n^*)^t]^t$.

3.2.5. Robust variance component estimation

Consider the general linear model with fixed and random effects (see Harville 1977)

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zr} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Y}$ is a $p \times 1$ random vector, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are matrices of regressors, $\boldsymbol{\beta}$ is a $q \times 1$ vector of unknown parameters, $\boldsymbol{\epsilon}$ has a multivariate normal distribution with zero mean vector and variance-covariance matrix $\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\alpha})$, $\boldsymbol{r}$ uncorrelated to $\boldsymbol{\epsilon}$ has a multivariate normal distribution with zero mean vector and variance-covariance matrix $\boldsymbol{\Sigma}_r(\boldsymbol{\alpha})$, and where $\boldsymbol{\alpha}$ is a vector of unknown covariance parameters. Further, assume that the column vectors of $\boldsymbol{X}$ are linearly independent. Thus, $\boldsymbol{Y}$ has a multivariate normal distribution with zero mean vector and variance-covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \boldsymbol{\Sigma}_\epsilon(\boldsymbol{\alpha}) + \boldsymbol{Z}^t\boldsymbol{\Sigma}_r(\boldsymbol{\alpha})\boldsymbol{Z}$. The restricted log-likelihood function (also see Patterson and Thompson 1971) is

$$\begin{aligned} l(\boldsymbol{\alpha}) = &-\left(\frac{1}{2}\right)\log\left[|\boldsymbol{\Sigma}(\boldsymbol{\alpha})|\right] - \left(\frac{1}{2}\right)\log\left[\left|\boldsymbol{X}^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})\boldsymbol{X}\right|\right] \\ &-\left(\frac{1}{2}\right)\boldsymbol{e}^t\boldsymbol{\Sigma}_e^{-1}(\boldsymbol{\alpha})\boldsymbol{e}, \end{aligned}$$

where $\boldsymbol{e} = \boldsymbol{K}^t\boldsymbol{y}$ and $\boldsymbol{\Sigma}_e(\boldsymbol{\alpha}) = \boldsymbol{K}^t\boldsymbol{\Sigma}(\boldsymbol{\alpha})\boldsymbol{K}$ is the variance-covariance matrix of the random vector $\boldsymbol{K^tY}$, and where $\boldsymbol{K}$ is a $p \times (p-q)$ matrix whose rows are any $p-q$ linearly independent rows of the matrix $\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}$. By Theorem 2.3, the MRB function can be

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\alpha}) = \boldsymbol{h}\left[c(\boldsymbol{y}, \boldsymbol{\alpha}), -\dot{\boldsymbol{l}}_e, \boldsymbol{\Sigma}_e^{1/2}(\boldsymbol{\alpha})\right].$$

By Theorem 2.4, the other sensible MRB function can be

$$\hat{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\alpha}) = \boldsymbol{h}\left[c(\boldsymbol{y}, \boldsymbol{\alpha}), \dot{\boldsymbol{l}}_\alpha, \boldsymbol{\mathcal{I}}^{-1/2}(\boldsymbol{\alpha})\right],$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\alpha})$ is the information matrix. The MRB functions can be developed analogously for the models proposed by Laird and Ware (1982).

3.2.6. Robust generalized estimating equations

Let the ith ($i = 1, \dots, n$) response vector $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^t$ have the distribution belonging to multivariate natural exponential families. The distribution of $\boldsymbol{Y}_i$,

$$f(\boldsymbol{y}_i|\boldsymbol{\theta}_i) \quad = \quad \exp\left[\boldsymbol{y}_i^t\boldsymbol{\theta}_i - A(\boldsymbol{\theta}_i)\right]\kappa(\boldsymbol{y}_i),$$

takes the form similar to the one given in the expression (2.3). The dependence of the response Y_{ij} on the associated explanatory variable $\boldsymbol{X_{ij}}$ can be modeled through the canonical link $\theta_{ij} = \boldsymbol{X_{ij}\beta}$. By Theorem 2.4, the MRB function can be

$$\hat{\boldsymbol{\phi}}_i(\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{h}\left[c(\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}), \boldsymbol{e}_i, \boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{\alpha})\right],$$

where $\boldsymbol{e}_i = \boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})$, $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha})$ is the the variance-covariance matrix of $\boldsymbol{Y}_i$, $\boldsymbol{\mu}_i(\boldsymbol{\beta})$ is the mean vector of $\boldsymbol{Y}_i$, and where $\boldsymbol{\alpha}$ is the correlation parameter. The score equations are $\sum_{i=1}^n \boldsymbol{X}_i^t\boldsymbol{e}_i = 0$, where $\boldsymbol{X}_i$ is a matrix with the jth row equal to $\boldsymbol{X_{ij}}$. Therefore, the associated robust estimating equations can be $\sum_{i=1}^n \boldsymbol{X}_i^t\hat{\boldsymbol{\phi}}_i(\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$. Note that the score equations are equivalent to the ones for independent Y_{ij}. In addition, the multivariate natural exponential families might not be suitable as the marginal distribution of Y_{ij} is from the univariate exponential family. The following estimating equations (see McCullagh and Nelder 1989, Chapter 9.3), $\sum_{i=1}^n \boldsymbol{D}_i^t(\boldsymbol{\beta})\boldsymbol{W}_i^{-1}(\boldsymbol{\alpha})[\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})]$, could be used, where $\boldsymbol{D}_i(\boldsymbol{\beta}) = \partial\dot{\boldsymbol{A}}_{\theta_i}(\boldsymbol{\theta}_i)/\partial\boldsymbol{\beta}$ and $\boldsymbol{W}_i(\boldsymbol{\alpha})$, depending on a sensibly chosen correlation matrix, might not be equal to $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha})$. A sensible robust generalized estimating equations can thus be

$$\sum_{i=1}^{n} \boldsymbol{D}_i^t(\boldsymbol{\beta})\boldsymbol{W}_i^{-1}(\boldsymbol{\alpha})\hat{\boldsymbol{\phi}}_i(\boldsymbol{Y}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) = 0.$$

4. Simulations

The purpose of the following simulations is to illustrate that the robust estimators based on the MRB and MRBB functions perform well. A range of scenarios, including different choices of sample sizes, types of simulated data, benchmark criteria used in the MRB and MRBB functions, and noise levels, were set up for the simulation study. The sample sizes n used were $n = 20$ and $n = 100$. Further, in addition to the clean data sets generated from the distribution F_0, two other models were used to generate the ”contaminated” data sets:

- mean-shift outlier model, i.e., $\mu^* = \mu + \imath$, where μ and μ^* were the means of the clean data and the outliers, and $\imath$ was the shift;
- mixture model, i.e., $F = (1 - \epsilon)F_0 + \epsilon F_1$, where F was the distribution of the contaminated data, F_1 was a distribution different from F_0, and $0 < \epsilon < 1$.

The proportion of the outliers in the mean-shift outlier model was 10% (also see Hampel et. al. 1986, Section 1.2c) and $\epsilon = 0.1$ in the simulation study. Further, 500 replicates of simulated data were generated for each model. Note that the parameters $\boldsymbol{\theta}$ of interest were the ones corresponding to the clean data. The relative error $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| / \|\boldsymbol{\theta}\|$ was used for evaluating the estimates, where $\hat{\boldsymbol{\theta}}$ was the estimate of $\boldsymbol{\theta}$. The complete simulation results and computational details are delegated to the supplementary materials.

4.1. MRB and MRBB functions

The binomial and Poisson MRB and MRBB estimators are compared with the MLE and posterior means in this section. The parameters of interest are the probability of one success and the mean given in Section 3.1.2 and Section 3.1.3, respectively. For the clean data sets, the random samples for the Poisson distribution with means 1 and 10 were generated, while the ones for the binomial distribution with the number of trials equal to 10 and the probabilities of a success equal to 0.1 and 0.5 were generated. The shifts were $3\sqrt{\theta}$ for the Poisson random samples and 4 for the binomial random samples. F_1 was the distributions for the random variables with means μ^*, while F_0 was the ones with means μ. The prior distributions for the means of Poisson random samples were gamma distributions with mean θ and variance 0.1θ, while the prior distributions for the parameter θ of binomial random samples were beta distributions with means 0.1 ($\alpha = 5$ and $\beta = 45$) and 0.5 ($\alpha = 100$ and $\beta = 100$). Parts of these results are provided in Table 1. The first number in the parenthesis is the average relative error of the maximum likelihood estimate or posterior mean and the other three numbers are the ones of the robust estimates with the benchmark criteria, 90th percentile, three sigma, and box plot, respectively. In general, for Poisson and binomial random samples generated from the mean-shift outlier model and the mixture model, the average relative errors of the estimates corresponding to the MRB and MRBB functions given in the expressions (3.1), (3.3), (3.2), and (3.4) are smaller than the maximum likelihood estimates and the posterior means. The average relative errors of the robust estimates based on the MRB functions are slightly larger than the ones of the maximum likelihood estimates for the clean data. However, the average relative errors of the robust estimates based on MRBB functions are larger than the ones of the posterior means for the clean data. Among the three benchmark criteria, the robust estimates using the 90th percentile and three sigma criteria perform more stable than the ones using the box plot criterion. For the clean data, the robust estimates using the box plot criterion tend to over-downweight the observations.

Since the proportion of the outliers in the mean-shift outlier model was 10% and $\epsilon = 0.1$ in the mixture model, it may be expected that the corresponding robust estimators based on the 90th percentile benchmark perform well in such settings. Therefore, different percentile benchmarks, including the 75th percentile, the 80th percentile, the 85 percentile, the 90th percentile, and the 95th percentile, were used to verify the effects of the benchmark choice on the performance of both the MRB estimators and MRBB

Table 1. Average relative errors for the MLE, posterior means, MRB estimators, and MRBB estimators with $n = 100$.

Poisson	$\theta = 1$ MLE and MRBE	$\theta = 1$ Posterior Mean and MRBBE
Clean Data	(0.0815,0.1119,0.0951,0.3190)	(0.0794,0.2528,0.2533,0.2530)
Mean-Shift	(0.2927,0.1505,0.1729,0.1952)	(0.5134,0.2621,0.2637,0.2546)
Mixture	(0.2929,0.1588,0.1794,0.2052)	(0.5179,0.2645,0.2652,0.2566)
Binomial	$\theta = 0.1$ MLE and MRBE	$\theta = 0.1$ Posterior Mean and MRBBE
Clean Data	(0.0717,0.0962,0.0796,0.3020)	(0.0762,0.3502,0.3499,0.3574)
Mean-Shift	(0.3973,0.2074,0.2462,0.1782)	(0.4826,0.3757,0.3773,0.3595)
Mixture	(0.4015,0.2362,0.2634,0.1832)	(0.4821,0.3799,0.3807,0.3611)

estimators. Besides the scenarios for the simulations used in the above simulation study, the settings for adding more extreme outliers corresponding to the shifts $5\sqrt{\theta}$ for the Poisson random samples and 8 for the binomial random samples were also used. These results corresponding to the proportion of the outliers equal to 10% and $\epsilon = 0.1$ are given in the supplementary materials. The numbers in the parenthesis are the ratios of average relative errors of the robust estimates using the 90th percentile benchmark to the ones using other percentile benchmarks. For example, the number 0.5 indicates that the average relative error of the robust estimate using the 90th percentile benchmark is twice as large as the one corresponding to another percentile benchmark. For the robust estimates based on the MRBB functions, the ratios are not significantly larger or smaller than 1. This implies that the robust estimates based on the MRBB functions are quite insensitive to the choices of the percentile benchmarks. Since the postulated observed data y_b, the posterior means in Section 3.1.2 and Section 3.1.3, could be considered as the stabilized data, the effect of the choice of the percentile benchmark might be small. On the other hand, the performance of the robust estimates based on the MRB functions might depend on the choices of the percentile benchmarks. In general, for the clean data, the ratios corresponding to the 95th percentile benchmark are smaller than 1 in most settings. This implies that the robust estimates obtained by trimming less data might perform better, as expected intuitively. On the other hand, as $n = 100$, the ratios corresponding to the low percentile benchmarks are smaller than 1. In some settings, the ratios are even significantly smaller than 1. This implies that the MRB estimators obtained by trimming more data being not around the "center" (the estimated mean) might perform better than the ones obtained by trimming less data. However, as $n = 20$, the ratios corresponding to the percentile benchmarks, ranged from the 80th percentile benchmark to the 90th percentile benchmark, i.e., trimming 10% to 20 % of the data, are smaller than the ones corresponding to the other two percentile benchmarks. Finally, as considering the settings for adding more extreme outliers, the results are similar to the ones with the outliers corresponding to the shifts $3\sqrt{\theta}$ for the Poisson random samples and 4 for the binomial random samples in most settings. In particular, for $n = 100$ and the smaller means, i.e., $\theta = 1$ for the Poisson distribution and $\theta = 0.1$ for the binomial distribution, the MRB estimators obtained by trimming more data being not around the center might perform significantly better than the ones obtained by trimming less data. For the proportion of the outliers in the mean-shift outlier model being 15% and $\epsilon = 0.15$ in the mixture model, the percentile benchmarks, including the 70th percentile, the 75th percentile, the 80 percentile, the 85th percentile, and the 90th percentile, were used. Generally, the results are quite similar to the ones for the proportion of the outliers being 10% in the mean-shift outlier model and $\epsilon = 0.1$ in the mixture model, i.e., the MRBB estimators being insensitive to the choices of the percentile benchmarks, the MRB estimators using the 90th percentile benchmark performing better than the ones using the other percentile benchmarks for the clean data in most settings, and the MRB estimators using the percentile benchmarks ranged from the 75th percentile benchmark to the 85th percentile benchmark performing better than the ones using the other percentile benchmarks for the contaminated data. However, the MRB estimators using the 70th percentile benchmarks, i.e., trimming 30% of the data, could not perform better than the ones using the other percentile benchmarks in most settings even as $n = 100$ or adding the more extreme outliers corresponding to the shifts $5\sqrt{\theta}$ for the Poisson random samples and 8 for the binomial random samples. This implies that the MRB estimators obtained by trimming too large proportion of the data might not perform well.

4.2. Robust regression for the data with correlated errors

In the simulation, the values of 3 input variables, $X_1,\ X_2$, and X_3, were generated from the standard normal distribution and the observations were generated from the model

$$Y_i = \nu_i + \epsilon_i = 1 + 3^{-1/2}X_{i1} + 3^{-1/2}X_{i2} + 3^{-1/2}X_{i3} + \epsilon_i, i = 1, \ldots, n,$$

where ϵ_i were zero mean random errors. The errors were generated from both Gaussian AR(1) and MA(1) processes with standard deviations of uncorrelated Gaussian errors, σ, equal to 0.2 and 1. The autocorrelation values at lag 1 for the Gaussian AR(1) process, $\rho(1)$, were -0.8, -0.2, 0.2, and 0.8, while -0.4, -0.2, 0.2, and 0.4 for the Gaussian MA(1) process. The shift, equal to three standard deviations of the random variable $1+3^{-1/2}X_1+3^{-1/2}X_2+3^{-1/2}X_3$, was 3. Similar to the previous simulation, the means of the clean data and the outliers were ν_i and ν_i+3, respectively. F_1 was the distributions for the random variables with means $\nu_i + 3$. The regression coefficients were of interest and thus $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ was assumed to be known. For each sample, the weighted least squares estimate and the robust estimate based on the MRB, function given in the expression (3.5) were computed. The 90th percentile benchmark criterion was used for computing the robust estimates. As indicated in the second table of the

Table 2. Average weighted mean square errors for the weighted least squares estimators and the MRB estimators with $n = 20$.

	$\sigma = 0.2$	
AR(1)	$\rho(1) = -0.8$	$\rho(1) = -0.2$
Clean Data	(0.2027,0.2470)	(0.1974,0.2294)
Mean-Shift	(14.7447,1.2154)	(7.4743,0.3151)
Mixture	(18.5197,12.5445)	(8.5405,6.0891)
MA(1)	$\rho(1) = 0.2$	$\rho(1) = 0.4$
Clean Data	(0.2053,0.2568)	(0.2032,0.2285)
Mean-Shift	(4.2499,0.3206)	(0.8952,0.3244)
Mixture	(15.5882,12.0111)	(4.8228,3.3397)
	$\sigma = 1$	
AR(1)	$\rho(1) = 0.2$	$\rho(1) = 0.8$
Clean Data	(0.1953,0.2384)	(0.2022,0.2190)
Mean-Shift	(0.3541,0.2994)	(0.3229,0.3019)
Mixture	(0.4196,0.3661)	(0.4589,0.4113)
MA(1)	$\rho(1) = -0.4$	$\rho(1) = -0.2$
Clean Data	(0.1925,0.2278)	(0.2054,0.2282)
Mean-Shift	(0.3492,0.3254)	(0.4727,0.3042)
Mixture	(0.6971,0.5879)	(1.0407,0.7391)

supplementary materials, both the weighted least squares estimates and the robust estimates perform comparably well in terms of the average relative errors for the clean data. On the other hand, for the contaminated data, the average relative errors of the robust estimates are smaller than the ones of the weighted least squares estimates in most situations. For the responses with small variances and small autocorrelations, the effects of the aberrant data on the weighed least squares estimates could be significant. Thus, the MRB estimators make a notable improvement over the weighted least squares estimators. In addition to the relative error, the weighted mean square error, $(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha})(\hat{\boldsymbol{\nu}}-\boldsymbol{\nu})/n$, was also used for evaluating different estimating methods, where $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)^t$ and $\hat{\boldsymbol{\nu}}$ was a vector of fitted values. In Table 2, the first number in the parenthesis is the average weighted mean square error of the weighted least squares estimate and the other number is the one of the robust estimate. The results based on the weighted mean square error criterion are quite similar to the ones based on the relative error.

5. Discussions

The bounded M-estimation function based on the asymptotic minimax principle was proposed in the classical paper by Huber. Hampel further proposed the bounded likelihood score function (see Morgenthaler 2007). The works by Huber and Hampel laid the sound foundations for the modern robustness theory. This article is intended to further enhance the bounded function approach. As presented in Section 3, the bounded function approach has wide applications. In addition, the robust estimators perform well in the simulation study.

The proposed MRB functions can be considered as the robust version of MR functions, which take the derivatives of the objective functions into account. In addition, as implied by Theorem 2.6, the estimator based on the MR function incorporating with the first derivative information is the MLE. This provides another support for the use of the MRB functions. Although only the MRB and MR functions involving the first derivatives of the objective functions are developed, the MRB and MR functions in the classes of functions with smoothness conditions involving higher order derivatives can be obtained analogously. Thus, more robust estimators and results based on these MRB functions can be derived.

The range of the operator h given in Section 2.6 is bounded, as indicated by Lemma 2.1. Further, if the range of the operator is a closed convex subset of $\mathcal{H}_2$, the existence of the estimator $\hat{\boldsymbol{v}} = \underset{\boldsymbol{v}\in\mathcal{H}_2}{\arg\min} F[h(c, \boldsymbol{v})]$, possibly not unique, can be guaranteed for a broad class of real-valued functions F defined on $\mathcal{H}_2$ by Theorem 25.1 of Deimling (1985, p. 321) or Proposition 1.2 of Ekeland and Témam (1976, p. 35).

6. PROOFS

Hereafter, the arguments $\boldsymbol{Y}$ and $\boldsymbol{\theta}$ of the functions in the proofs have been suppressed for the succinctness, for example, $\boldsymbol{\phi}(\boldsymbol{Y}, \boldsymbol{\theta})$ being replaced with $\boldsymbol{\phi}$, except for the ones of Theorem 2.7. In addition, the abbreviated notations for $\boldsymbol{\phi}_n(\boldsymbol{Y}_n, \boldsymbol{\theta})$ and $\hat{\boldsymbol{\phi}}(\boldsymbol{Y}_n, \boldsymbol{\theta})$ are $\boldsymbol{\phi}_n$ and $\hat{\boldsymbol{\phi}}_n$, respectively. The notation $Tr(\boldsymbol{M})$ is denoted as the trace of the square matrix $\boldsymbol{M}$.

6.1. Theorem 2.1

To prove Theorem 2.1, the following lemma is required.

Lemma 6.1. *Let $E[|g(\boldsymbol{Y})|] < \infty$, $E[|\dot{g}_{y_i}(\boldsymbol{Y})|] < \infty$, and $\boldsymbol{Y} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{X} + \boldsymbol{\nu}$, where g is a measurable function defined on R^p, $\boldsymbol{X} = (X_1, \ldots, X_p)^t$ has a multivariate normal distribution with zero mean vector and identity variance-covariance matrix, and $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma}$ are the mean vector and variance-covariance matrix of $\boldsymbol{Y}$, respectively. Then,*

$$E[g(\boldsymbol{Y})X_i] = col_i^t(\boldsymbol{\Sigma}^{1/2})E\left[\dot{\boldsymbol{g}}_y(\boldsymbol{Y})\right],$$

where $col_i(\boldsymbol{M})$ is the ith column of a matrix $\boldsymbol{M}$.

6.1.1. Proofs of Lemma 6.1

Using integration by part,

$$\begin{aligned} E\left[g(\boldsymbol{Y})X_i\right] &= E_{\boldsymbol{X}_{(i)}}\left\{E\left[g(\boldsymbol{Y})X_i \mid \boldsymbol{X}_{(i)}\right]\right\} \\ &= E_{\boldsymbol{X}_{(i)}}\left[\int \dot{g}_{x_i}(\boldsymbol{Y})\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x_i^2}{2}\right)dx_i\right] \\ &= E\left[\dot{\boldsymbol{g}}_y^t(\boldsymbol{Y})\right] col_i(\boldsymbol{\Sigma}^{1/2}), \end{aligned}$$

where $\dot{g}_{x_i}(\boldsymbol{y}) = \partial g(\boldsymbol{\Sigma}^{1/2}\boldsymbol{x} + \boldsymbol{\nu})/\partial x_i$, $\boldsymbol{X}_{(i)}$ is the random vector $\boldsymbol{X}$ without the ith element, and the expected values are finite by the conditions $E[|g(\boldsymbol{Y})|] < \infty$ and $E[|\dot{g}_{y_i}(\boldsymbol{Y})|] < \infty$.

6.1.2. Proofs of Theorem 2.1

By conditions (i), (ii), and (iii), h is a well defined and bounded function almost surely, $E(|\dot{\boldsymbol{h}}_{y_i}|) < \infty$, and $\boldsymbol{h} \in \mathcal{F}^n$. Next, let $\boldsymbol{Y} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{X} + \boldsymbol{\nu}$ and $\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{\phi} = [\boldsymbol{\phi}^*]^t\boldsymbol{\phi}^*$, where $\boldsymbol{X} = (X_1, \ldots, X_p)^t$ have a multivariate normal distribution with zero mean vector and identity variance-covariance matrix and $\boldsymbol{\phi}^* = \boldsymbol{U}\boldsymbol{\phi}$. Then,

$$\begin{aligned} E\left[||\boldsymbol{\phi}^* - \boldsymbol{U}(\boldsymbol{M_1}\boldsymbol{Y} + \boldsymbol{M_0})||^2\right] &= E\left(\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{\phi}\right) - 2E\left(\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{M_1}\boldsymbol{Y}\right) - \\ & 2\hat{\boldsymbol{\mu}}^t\boldsymbol{W}\boldsymbol{M_0} + ||\boldsymbol{U}(\boldsymbol{M_1}\boldsymbol{\nu} + \boldsymbol{M_0})||^2 + Tr(\boldsymbol{W}\boldsymbol{M_1}\boldsymbol{\Sigma}\boldsymbol{M_1^t}). \end{aligned}$$

By Lemma 6.1,

$$E(\phi_k X_i) = col_i^t(\boldsymbol{\Sigma}^{1/2})col_k(\hat{\boldsymbol{\mu}}_1),$$

where $\hat{\boldsymbol{\mu}}_1$ is a $p \times m$ matrix with ikth element $\hat{\mu}_{1,ik}$. Therefore,

$$\begin{aligned} E\left(\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{M_1}\boldsymbol{Y}\right) &= E\left[\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{M_1}(\boldsymbol{\Sigma}^{1/2}\boldsymbol{X} + \boldsymbol{\nu})\right] \\ &= \text{vec}^t\left(\boldsymbol{W}\boldsymbol{M_1}\boldsymbol{\Sigma}^{1/2}\right)\text{vec}\left(\hat{\boldsymbol{\mu}}_1^t\boldsymbol{\Sigma}^{1/2}\right) + \hat{\boldsymbol{\mu}}^t\boldsymbol{W}\boldsymbol{M_1}\boldsymbol{\nu}, \end{aligned}$$

is the same for all feasible functions $\boldsymbol{\phi} \in \mathcal{F}^n$.

Thus, the minimizer of $E(\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{\phi}) = E[(\boldsymbol{\phi}^*)^t\boldsymbol{\phi}^*)]$, equivalent to the one of $E[||\boldsymbol{\phi}^* - \boldsymbol{U}(\boldsymbol{M_1}\boldsymbol{Y} + \boldsymbol{M_0})||^2]$, is

$$\hat{\boldsymbol{\phi}}^* = \boldsymbol{h}\left[c, \boldsymbol{U}(\boldsymbol{M_1}\boldsymbol{Y} + \boldsymbol{M_0}), \boldsymbol{I}\right].$$

This implies

$$\hat{\boldsymbol{\phi}} = \boldsymbol{h}(c, \boldsymbol{M_1}\boldsymbol{Y} + \boldsymbol{M_0}, \boldsymbol{U}).$$

6.2. Theorem 2.2

To prove Theorem 2.2, the following lemma is required.

Lemma 6.2. *Let* $\{\boldsymbol{Y_n} = (Y_{n1}, Y_{n2}, \ldots, Y_{np}),\ n = 1, \ldots,\}$ *be a sequence of random vectors. Suppose* $\{Y_{ni}^2\}$ *and* $\{Y_{nj}^2\}$ *are both uniformly integrable. Then,*
(a) the sequence $\{Y_{ni}Y_{nj}\}$ *is uniformly integrable and*
(b) the sequence $\{g_n(\boldsymbol{Y_n})Y_{ni}\}$ *is uniformly integrable if* $|g_n(\boldsymbol{Y_n})| \leq C$ $a.s.$ *for all* n*, where* g_n *are measurable functions defined on* R^p *and* C *is some constant.*

6.2.1. Proofs of Lemma 6.2

Let

$$I_{\mathcal{B}}(x) = \begin{cases} 1 & x \in \mathcal{B} \\ 0 & x \notin \mathcal{B} \end{cases}$$

be the indicator function associated with the set $\mathcal{B}$ and $I(\mathcal{B})$ will be used for $I_{\mathcal{B}}$ with the argument x being suppressed. To prove (a), by Hölder's inequality,

$$\begin{aligned} & E\left[|Y_{ni}Y_{nj}|\, I(|Y_{ni}Y_{nj}| > c)\right] \\ \leq\ & E\left[|Y_{ni}Y_{nj}|\, I(|Y_{ni}| > c^{1/2})\right] + E\left[|Y_{ni}Y_{nj}|\, I(|Y_{nj}| > c^{1/2})\right] \\ \leq\ & \left\{E\left[Y_{ni}^2 I(|Y_{ni}| > c^{1/2})\right]\right\}^{1/2} \left[E\left(Y_{nj}^2\right)\right]^{1/2} + \\ & \left\{E\left[Y_{nj}^2 I(|Y_{nj}| > c^{1/2})\right]\right\}^{1/2} \left[E\left(Y_{ni}^2\right)\right]^{1/2}. \end{aligned}$$

Therefore,

$$\begin{aligned} 0 \leq\ & \lim_{c\to\infty} \sup_n E\left[|Y_{ni}Y_{nj}|\, I(|Y_{ni}Y_{nj}| > c)\right] \\ \leq\ & \lim_{c\to\infty} \sup_n \left\{E\left[Y_{ni}^2 I(|Y_{ni}| > c^{1/2})\right]\right\}^{1/2} \sup_n \left[E\left(Y_{nj}^2\right)\right]^{1/2} + \\ & \lim_{c\to\infty} \sup_n \left\{E\left[Y_{nj}^2 I(|Y_{nj}| > c^{1/2})\right]\right\}^{1/2} \sup_n \left[E\left(Y_{ni}^2\right)\right]^{1/2} \\ =\ & 0. \end{aligned}$$

To prove (b), since

$$E\left[|g_n(\boldsymbol{Y}_n)Y_{ni}|\, I(|g_n(\boldsymbol{Y}_n)Y_{ni}| > c)\right] \leq CE\left[|Y_{ni}|\, I(|Y_{ni}| > c/C)\right],$$

thus

$$\begin{aligned} 0 \leq\ & \lim_{c\to\infty} \sup_n E\left[|g_n(\boldsymbol{Y}_n)Y_{ni}|\, I(|g_n(\boldsymbol{Y}_n)Y_{ni}| > c)\right] \\ \leq\ & \lim_{c\to\infty} \sup_n CE\left[|Y_{ni}|\, I(|Y_{ni}| > c/C)\right] \\ \leq\ & \lim_{c\to\infty} \sup_n CE\left[Y_{ni}^2 I(Y_{ni}^2 > c^2/C^2)\right] \\ =\ & 0. \end{aligned}$$

6.2.2. Proofs of Theorem 2.2

Let $\boldsymbol{\phi_n^t W \phi_n} = (\boldsymbol{\phi_n^*})^t \boldsymbol{\phi_n^*}$, where $\boldsymbol{\phi_n^*} = \boldsymbol{U\phi_n}$. Then, for large n,

$$\begin{aligned} & E\left[||\boldsymbol{\phi_n^*} - \boldsymbol{U}(\boldsymbol{M_1 Y_n} + \boldsymbol{M_0})||^2\right] \\ = \; & E\left(\boldsymbol{\phi_n^t W \phi_n}\right) - 2E\left(\boldsymbol{\phi_n^t W M_1 Y_n}\right) - \\ & 2E\left(\boldsymbol{\phi_n^t}\right)\boldsymbol{W M_0} + ||\boldsymbol{U}(\boldsymbol{M_1 \nu}_n + \boldsymbol{M_0})||^2 + Tr(\boldsymbol{W M_1 \Sigma}_n \boldsymbol{M_1^t}), \end{aligned}$$

where $E(\boldsymbol{Y_n}) = \boldsymbol{\nu}_n$ and $Cov(\boldsymbol{Y_n}) = \boldsymbol{\Sigma}_n$. Since $\boldsymbol{Y_n} \xrightarrow{d} \boldsymbol{Y}$ and $E(Y_{ni}^2) \to E(Y_i^2) < \infty$ by condition (i), $\{Y_{ni}^2\}$ are uniformly integrable. By Lemma 6.2, $\{Y_{ni}Y_{nj}\}$ and $\{\phi_{nk}Y_{ni}\}$ are uniformly integrable, where ϕ_{nk} is the kth element of $\boldsymbol{\phi_n}$. This implies $E(Y_{ni}Y_{nj}) \to E(Y_iY_j)$ and $E(\phi_{nk}Y_{ni}) \to E(\phi_kY_i)$. Thus, there exists N such that for $n > N$,

$$\begin{aligned} E\left(\boldsymbol{\phi_n^t W M_1 Y_n}\right) &= E\left(\boldsymbol{\phi^t W M_1 Y}\right) + \epsilon_{1n}, \\ E\left(\boldsymbol{\phi_n^t}\right)\boldsymbol{W M_0} &= E\left(\boldsymbol{\phi^t}\right)\boldsymbol{W M_0} + \epsilon_{2n}, \\ ||\boldsymbol{U}(\boldsymbol{M_1 \nu}_n + \boldsymbol{M_0})||^2 &= ||\boldsymbol{U}(\boldsymbol{M_1 \nu} + \boldsymbol{M_0})||^2 + \epsilon_{3n}, \\ Tr(\boldsymbol{W M_1 \Sigma_n M_1^t}) &= Tr(\boldsymbol{W M_1 \Sigma M_1^t}) + \epsilon_{4n}, \end{aligned}$$

where for any $\epsilon > 0$ and $max\{|\epsilon_{1n}|, |\epsilon_{2n}|, |\epsilon_{3n}|, |\epsilon_{4n}|\} < \epsilon$. Then, by condition (ii),

$$\begin{aligned} & E\left[||\boldsymbol{\phi_n^*} - \boldsymbol{U}(\boldsymbol{M_1 Y_n} + \boldsymbol{M_0})||^2\right] \\ = \; & E\left(\boldsymbol{\phi_n^t W \phi_n}\right) - 2\left[\text{vec}^t\left(\boldsymbol{W M_1 \Sigma}^{1/2}\right)\text{vec}\left(\hat{\boldsymbol{\mu}}_1^t \boldsymbol{\Sigma}^{1/2}\right) + \right. \\ & \left.\hat{\boldsymbol{\mu}}^t \boldsymbol{W}(\boldsymbol{M_1 \nu} + \boldsymbol{M_0})\right] + ||\boldsymbol{U}(\boldsymbol{M_1 \nu} + \boldsymbol{M_0})||^2 + Tr(\boldsymbol{W M_1 \Sigma M_1^t}) + \\ & (\epsilon_{3n} + \epsilon_{4n} - 2\epsilon_{1n} - 2\epsilon_{2n}), \end{aligned}$$

where $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma}$ are the mean vector and variance-covariance matrix of $\boldsymbol{Y}$, respectively. By the continuity of $\hat{\boldsymbol{\phi}}$ with probability one for any given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\hat{\boldsymbol{\phi}}_n \xrightarrow{d} \hat{\boldsymbol{\phi}}$ and thus the above equation also holds for $\hat{\boldsymbol{\phi}}_n$. Since $(\epsilon_{3n} + \epsilon_{4n} - 2\epsilon_{1n} - 2\epsilon_{2n})$ can be as small as possible, $R(\boldsymbol{\phi_n}) - R(\hat{\boldsymbol{\phi}}_n)$ mainly depends on the difference between $E[||\boldsymbol{\phi_n^*} - \boldsymbol{U}(\boldsymbol{M_1 Y_n} + \boldsymbol{M_0})||^2]$ and $E[||\hat{\boldsymbol{\phi}}_n^* - \boldsymbol{U}(\boldsymbol{M_1 Y_n} + \boldsymbol{M_0})||^2]$, where $\hat{\boldsymbol{\phi}}_n^* = \boldsymbol{U}\hat{\boldsymbol{\phi}}_n$. Thus, the proof is complete.

6.3. Theorem 2.3

The following lemma, which generalizes the Stein's identity (see Lehmann and Casella 1998, Lemma 5.15, p. 31) under different situations, is given. The lemma can be used to prove Theorem 2.3, which gives the MRB functions.

Lemma 6.3. *Let $E(|g|) < \infty$, where g is a measurable function defined on R^p. (a) Let $E(|\dot{g}_{y_i}|) < \infty$ and the support of $\boldsymbol{Y}$ be $(-\infty, \infty) \times \cdots \times (-\infty, \infty)$. Then,*

$$E\left[g\left(\frac{\dot{\boldsymbol{f}}_y}{f}\right)\right] = -E\left(\dot{\boldsymbol{g}}_y\right).$$

If the support of $\boldsymbol{Y}$ is the product of bounded intervals $(a_1, b_1) \times \cdots \times (a_p, b_p)$, then the above equation also holds if $gf \to 0$ as $y_i \to a_i$ or $y_i \to b_i$.
(b) Let Y_i take values in the set $\{0, 1, 2, \cdots\}$ and $E[|\Delta^+_{y_i}(g)|] < \infty$. Then,

$$E(g\boldsymbol{\eta}) = E\left[\Delta^+_y(\boldsymbol{g})\right],$$

where $\boldsymbol{\eta} = \Delta_y(\boldsymbol{f})/f$ if $f > 0$ and $\boldsymbol{\eta} = 0$ otherwise.

6.3.1. Proofs of Lemma 6.3

To prove (a), using integration by part (also see Billingsley 1995, pp. 236-237),

$$E\left[g\left(\frac{\dot{f}_{y_i}}{f}\right)\right] = \int\left[\int g(\boldsymbol{y})df\right]d\boldsymbol{y}_{(i)} = -E\left[\dot{g}_{y_i}(\boldsymbol{Y})\right],$$

where $d\boldsymbol{y}_{(i)} = dy_1 \cdots dy_{i-1}dy_{i+1} \cdots dy_p$ and the expected values are finite by the conditions $E(|g|) < \infty$ and $E(|\dot{g}_{y_i}|) < \infty$. If the support of $\boldsymbol{Y}$ is the product of bounded intervals, the proofs follow analogously. To prove (b),

$$E(g\eta_i) = \sum_{\boldsymbol{y}_{(i)}}\sum_{y_i=0}^{\infty} g(\boldsymbol{y})\Delta_{y_i}(f) = \sum_{\boldsymbol{y}_{(i)}}\sum_{y_i=0}^{\infty} \Delta^+_{y_i}(g)f = E\left[\Delta^+_{y_i}(g)\right],$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)^t$,

$$\sum_{\boldsymbol{y}_{(i)}} = \sum_{y_1}\cdots\sum_{y_{i-1}}\sum_{y_{i+1}}\cdots\sum_{y_p},$$

and the expected values are finite by the conditions $E(|g|) < \infty$ and $E[|\Delta^+_{y_i}(g)|] < \infty$.

6.3.2. Proofs of Theorem 2.3

By conditions (ii), (iii), (iv), and (v), h is a well defined and bounded function almost surely, $E(|\dot{\boldsymbol{h}}_{y_i}|) < \infty$, and $\boldsymbol{h} \in \mathcal{F}^y$. Next, let $\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{\phi} = [\boldsymbol{\phi}^*]^t\boldsymbol{\phi}^*$, where $\boldsymbol{\phi}^* = \boldsymbol{U}\boldsymbol{\phi}$. Then,

$$\begin{aligned}
&E\left\{\left\|\boldsymbol{\phi}^* - \boldsymbol{U}\left[\boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_y}{f}\right) + \boldsymbol{M_0}\right]\right\|^2\right\} \\
&= E\left(\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{\phi}\right) - 2\hat{\boldsymbol{\mu}}^t\boldsymbol{W}\boldsymbol{M_0} - 2E\left[\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_y}{f}\right)\right] + \\
&\left\|\boldsymbol{U}\left[\boldsymbol{M_1}E\left(\frac{\dot{\boldsymbol{f}}_y}{f}\right) + \boldsymbol{M_0}\right]\right\|^2 + Tr\left\{\boldsymbol{W}Cov\left[\boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_y}{f}\right)\right]\right\},
\end{aligned}$$

where the covariance matrix of $\boldsymbol{M}_1(\dot{\boldsymbol{f}}_y/f)$ with finite elements exists by conditions (v) and (vi). The expected value,

$$E\left[\boldsymbol{\phi}^t \boldsymbol{W} \boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_y}{f}\right)\right],$$

is the same for all functions in the class $\mathcal{F}^y$ since it is a linear combination of $E[\phi_k(\dot{f}_{y_i}/f)]$, and by Lemma 6.3 (a),

$$E\left[\phi_k\left(\frac{\dot{f}_{y_i}}{f}\right)\right] = -\hat{\mu}_{1,ik}.$$

Thus, the result that the minimizer of $E[\boldsymbol{\phi}^t \boldsymbol{W} \boldsymbol{\phi}]$ is equivalent to the one of

$$E\left\{\left\|\boldsymbol{\phi}^* - \boldsymbol{U}\left[\boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_y}{f}\right) + \boldsymbol{M_0}\right]\right\|^2\right\}$$

gives

$$\hat{\boldsymbol{\phi}} = \boldsymbol{h}\left[c, \boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_y}{f}\right) + \boldsymbol{M_0}, \boldsymbol{U}\right].$$

The proofs of (b) and (c) are very similar to the ones of (a). For the proofs of (c), the main difference is to use Lemma 6.3 (b) to prove that the expected value,

$$E(\boldsymbol{\phi}^t \boldsymbol{W} \boldsymbol{M_1} \boldsymbol{\eta}),$$

is the same for all functions in the class $\mathcal{F}^{\Delta y}$, where $\boldsymbol{\eta}$ is defined in Lemma 6.3.

6.4. Proofs of Theorem 2.4

The proofs of this theorem are similar to the ones given in Theorem 2.3. By conditions (i), (ii), (iii), and (iv), h is a well defined and bounded function almost surely, $E(\dot{\boldsymbol{h}}_{\theta_i}) < \infty$, and $\boldsymbol{h} \in \mathcal{F}^{\theta}$. Next, for ease of exposition, let $f > 0$. By differentiating $E(\boldsymbol{\phi})$ with respect to $\boldsymbol{\theta}$ under the integral sign, the following equations can be obtained:

$$E\left[\phi_k\left(\frac{\dot{f}_{\theta_i}}{f}\right)\right] = \hat{\mu}_{0,ik} - \hat{\mu}_{1,ik},$$

where $\hat{\mu}_{0,k}$ is the kth element of $\hat{\boldsymbol{\mu}}$ and $\hat{\mu}_{0,ik} = \partial \hat{\mu}_{0,k}/\partial \theta_i$. Thus, the expected value,

$$E\left\{\boldsymbol{\phi}^t \boldsymbol{W}\left[\boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_\theta}{f}\right)\right]\right\},$$

is the same for all functions in the class $\mathcal{F}^\theta$ since it is a linear combination of $E[\phi_k(\dot{f}_{\theta_i}/f)]$. Therefore, the result that the minimizer of $E[\boldsymbol{\phi}^t\boldsymbol{W}\boldsymbol{\phi}]$ is equivalent to the one of

$$E\left\{\left\|\boldsymbol{\phi}^* - \boldsymbol{U}\left[\boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_\theta}{f}\right) + \boldsymbol{M_0}\right]\right\|^2\right\}$$

gives

$$\hat{\boldsymbol{\phi}} = \boldsymbol{h}\left[c, \boldsymbol{M_1}\left(\frac{\dot{\boldsymbol{f}}_\theta}{f}\right) + \boldsymbol{M_0}, \boldsymbol{U}\right].$$

6.5. Proofs of Corollary 2.3 (b)

The derivations mainly depend on the following formulae. If b_{ij} are distinct,

$$\frac{\partial\,|\boldsymbol{B}|}{\partial \boldsymbol{B}} = |\boldsymbol{B}|\left(\boldsymbol{B}^{-1}\right)^t.$$

On the other hand, if $\boldsymbol{B}$ is symmetric,

$$\frac{\partial\,|\boldsymbol{B}|}{\partial \boldsymbol{B}} = |\boldsymbol{B}|\left[2\boldsymbol{B}^{-1} - diag\left(\boldsymbol{B}^{-1}\right)\right].$$

6.6. Proofs of Theorem 2.6

Since

$$E\left[(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})^t\boldsymbol{W}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right] = E\left(\boldsymbol{\phi^t}\boldsymbol{W}\boldsymbol{\phi}\right) - \text{vec}^t(\boldsymbol{W})\text{vec}(\boldsymbol{\mathcal{I}}),$$

the minimizer of $E(\boldsymbol{\phi^t}\boldsymbol{W}\boldsymbol{\phi})$, equivalent to the one of $E[(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})^t\boldsymbol{W}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})]$, is the score function.

6.7. Proofs of Theorem 2.7

Since $\hat{\boldsymbol{\theta}}^j_\infty(\boldsymbol{Y_n}) \stackrel{p}{\longrightarrow} \boldsymbol{\theta}_0$, there exists a sequence $\{n_k\}$ of integers increasing to infinity such that $\hat{\boldsymbol{\theta}}^j_\infty(\boldsymbol{Y_{n_k}}) \stackrel{w.p.1}{\longrightarrow} \boldsymbol{\theta}_0$, where $\stackrel{w.p.1}{\longrightarrow}$ is denoted as the convergence with probability one. Then, in the sample space, there exists a subset of which probability measure equal to 1 such that $\hat{\boldsymbol{\theta}}^j_\infty[\boldsymbol{Y_{n_k}}(w)] \longrightarrow \boldsymbol{\theta}_0$ and condition (ii) holds, where w is the sample point in the set. Further, there exists N such that $\|\hat{\boldsymbol{\theta}}^j_\infty[\boldsymbol{Y_{n_{k^*}}}(w)] - \boldsymbol{\theta}_0\| < \epsilon_1$, for $k^* > N$ and any $\epsilon_1 > 0$. Then,

$$\hat{\boldsymbol{\phi}}^j_\infty\{\boldsymbol{Y_{n_{k^*}}}(w), \hat{\boldsymbol{\theta}}^j_\infty[\boldsymbol{Y_{n_{k^*}}}(w)]\} = 0,$$

and

$$\begin{aligned}&\left|\left|\hat{\boldsymbol{\phi}}_{\infty}^{j}\left\{\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w),\hat{\boldsymbol{\theta}}_{\infty}^{j}\left[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w)\right]\right\}-\hat{\boldsymbol{\phi}}_{\infty}^{j}\left[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w),\boldsymbol{\theta}_0\right]\right|\right|\\ =&\left|\left|\hat{\boldsymbol{\phi}}_{\infty}^{j}\left[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w),\boldsymbol{\theta}_0\right]\right|\right|<\epsilon_2\end{aligned}$$

for any $\epsilon_2 > 0$ since $\hat{\boldsymbol{\phi}}_{\infty}^{j}$ is continuous on $\boldsymbol{\Omega}$ by conditions (ii) and (v). Therefore, by condition (i), there exists a compact set $\boldsymbol{\Omega}^* \subset \boldsymbol{\Omega}$ such that $\|\hat{\boldsymbol{\phi}}_{\infty}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w),\boldsymbol{\theta}]\| < \infty$ for $\boldsymbol{\theta} \in \boldsymbol{\Omega}^*$ and both $\boldsymbol{\theta}_0,\ \hat{\boldsymbol{\theta}}_{\infty}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w)] \in \boldsymbol{\Omega}^*$. Since $C_t^j \to \infty$ and a continuous function is bounded on a compact set, there exists

$$C_{t^*}^j \geq \|\boldsymbol{U}\hat{\boldsymbol{\phi}}_{\infty}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w),\boldsymbol{\theta}]\|$$

such that $\hat{\boldsymbol{\phi}}_{t^*}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w),\boldsymbol{\theta}] = \hat{\boldsymbol{\phi}}_{\infty}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w),\boldsymbol{\theta}]$ for $\boldsymbol{\theta} \in \boldsymbol{\Omega}^*$. This implies that

$$\hat{\boldsymbol{\theta}}_{t^*}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w)] = \hat{\boldsymbol{\theta}}_{\infty}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w)]$$

and $\|\hat{\boldsymbol{\theta}}_{t^*}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}^*}}(w)] - \boldsymbol{\theta}_0\| < \epsilon_1$. That is, $\lim_{t,k\to\infty} \hat{\boldsymbol{\theta}}_{t}^{j}[\boldsymbol{Y}_{\boldsymbol{n}_{\boldsymbol{k}}}(w)] = \boldsymbol{\theta}_0$.

The proofs for (b) are analogous to the ones for (a).

6.8. Proofs of Lemma 2.1

By conditions (i) and (ii),

$$\|U(\boldsymbol{v})\|_{\mathcal{H}_4}^2 =< U(\boldsymbol{v}), U(\boldsymbol{v}) >_{\mathcal{H}_4} =< \boldsymbol{v}, (U^*U)(\boldsymbol{v}) >_{\mathcal{H}_2} \geq k\|\boldsymbol{v}\|_{\mathcal{H}_2}^2.$$

Therefore, for any $c \in \mathcal{B}_1$ and $\boldsymbol{v} \in \mathcal{H}_2$,

$$\|h(c,\boldsymbol{v})\|_{\mathcal{H}_2} \leq k^{-1/2}C,$$

where C is any bound for the set $\mathcal{B}_1$.

REFERENCES

1. J. Agulló, C. Croux and S. Van Aelst, The multivariate least-trimmed squares estimator, *J. Mult. Anal.*, **99** (2008), 311-338.
2. H. Akaike, Statistical predictor identification, *Ann. Inst. Statist. Math.*, **22** (1970), 203-217.
3. H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automat. Control*, **19** (1974), 716-723.
4. J. P. Aubin, *Applied Functional Analysis*, Wiley, New York, 2000.
5. P. Billingsley, *Probability and Measure*, Wiley, New York, 1995.
6. G. E. P. Box and D. R. Cox, An analysis of transformations, *J. Roy. Stat. Soc. Ser. B*, **26** (1964), 211-243.
7. K. Deimling, *Nonlinear Functional Analysis*, Springer-Verlag, Berlin Heidelberg, 1985.

8. J. Durbin, Estimation of parameters in time series regression models, *J. Roy. Stat. Soc. Ser. B*, **22** (1960), 139-153.

9. I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.

10. F. R. Hampel, The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.*, **69** (1974), 383-393.

11. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, 1986.

12. D. A. Harville, Maximum likelihood approaches to variance component estimation and to related problems, *J. Amer. Statist. Assoc.*, **72** (1977), 320-340.

13. P. J. Huber, Robust estimation of a location parameter. *Ann. Math. Statist.*, **35** (1964), 73-101.

14. P. J. Huber, Robust regression: Asymptotics, conjectures and Monte Carlo, *Ann. Statist.*, **1** (1973), 799-821.

15. P. J. Huber and E. M. Ronchetti, *Robust Statistics*, Wiley, New York, 2009.

16. W. S. Krasker and R. E. Welsch, Efficient bounded-influence regression estimation, *J. Amer. Statist. Assoc.*, **77** (1982), 595-604.

17. H. R. Künsch, L. A. Stefanski and R. J. Carroll, Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models, *J. Amer. Statist. Assoc.*, **84** (1989), 460-466.

18. N. M. Laird and J. H. Ware, Random-effects models for longitudinal data, *Biometrics*, **38** (1982), 963-974.

19. E. L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer-Verlag, New York, 1998.

20. C. L. Mallows, Some comments on C_p, *Technometrics*, **15** (1973), 661-675.

21. P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman and Hall, London, 1989.

22. S. Morgenthaler, A survey of robust statistics, *Stat. Meth. Appl.*, **15** (2007), 271-293.

23. H. D. Patterson and R. Thompson, Recovery of inter-block information when block sizes are unequal, *Biometrika*, **58** (1971), 545-554.

24. E. M. Ronchetti, Robust model selection in regression, *Statist. Prob. Let.*, **3** (1985), 21-23.

25. E. M. Ronchetti and R. G. Staudte, A robust version of Mallow's C_p, *J. Amer. Statist. Assoc.*, **89** (1994), 550-559.

26. P. J. Rousseeuw, Least median of squares regression, *J. Amer. Statist. Assoc.*, **79** (1984), 871-880.

27. P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.

28. R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

29. R. G. Staudte and S. J. Sheather, *Robust Estimation and Testing*, Wiley, New York, 1990.

30. S. Van Aelst and G. Willems, Multivariate regression S-estimators for robust estimation and inference, *Stat. Sinica*, **15** (2005), 981-1001.

31. W. H. Wei, On regression model selection for the data with correlated errors, *Ann. Inst. Statist. Math.*, **61** (2008), 291-308.

Wen Hsiang Wei
Department of Statistics
Tung Hai University
Taichung 407, Taiwan
E-mail: wenwei@thu.edu.tw