Wang, Miaomiao, You, Kang, Zhu, Lixing and Zou, Guohua (2024) *Robust model averaging approach by Mallows-type criterion.* Biometrics, 80 (4). ISSN 0006-341X.

BIOMETRICS xx, 1–25 DOI: xxx
xx xx

Robust model averaging approach by Mallows-type criterion

Miaomiao Wang[a,b], Kang You[c,d], Lixing Zhu[e] and Guohua Zou[c,*]

[a]School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100029, China

[b]University of Chinese Academy of Sciences, Beijing 100049, China

[c]School of Mathematical Sciences, Capital Normal University, Beijing 100048, China

[d]School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury CT2 7NZ, United Kingdom

[e]Center for Statistics and Data Science, Beijing Normal University, Zhuhai 519000, China

**email:* ghzou@amss.ac.cn

SUMMARY: Model averaging is an important tool for treating uncertainty from model selection process and fusing information from different models, and has been widely used in various fields. However, the most existing model averaging criteria are proposed based on the methods of ordinary least squares or maximum likelihood, which possess high sensitivity to outliers or violation of certain model assumption. For the mean regression, no optimal robust methods are developed. To fill this gap, in our paper, we propose an outlier-robust model averaging approach by Mallows-type criterion. The idea is that we first construct a generalized M (GM) estimator for each candidate model, and then build robust weighting schemes by the asymptotic expansion of the final prediction error based on the GM-type loss function. So we can still achieve a trustworthy result even if the dataset is contaminated by outliers in response and/or covariates. Asymptotic properties of the proposed robust model averaging estimators are established under some regularity conditions. The consistency of our weight estimators tending to the theoretically optimal weight vectors is also derived. We prove that our model averaging estimator is robust in terms of having bounded influence function. Further, we define the empirical prediction influence function to evaluate the quantitative robustness of the model averaging estimator. A simulation study and a real data analysis are conducted to demonstrate the finite sample performance of our estimators and compare them with other commonly used model selection and averaging methods.

KEY WORDS: GM-estimator; Influence function; Mallows-type criterion; Model averaging; Outlier-robust.

[1]* Corresponding author.

This paper has been submitted for consideration for publication in *Biometrics*

1 Introduction

Model averaging approach, combining the estimators or forecasts from different models, has received much attention in past decades. A main advantage of model averaging over model selection is its full utilization of information from various models, and so model averaging often leads to more accurate results of estimation or forecast. Model averaging develops in two directions: Frequentist Model Averaging (FMA) and Bayesian Model Averaging (BMA). In Raftery et al. (2005), BMA is utilized to calibrate forecast ensembles, with weights determined by the EM algorithm. This article concentrates on FMA. A key issue with FMA is the choice of weights assigned to candidate models. Over the past two decades, a number of FMA weight selection algorithms have been developed, such as weighting by information criterion scores of model selection (Hjort and Claeskens, 2006), adaptive regression by mixing (Yang, 2001), the Mallows' criterion (Hansen, 2007), MSE minimization (Liang et al., 2011), cross-validation (CV) (Hansen and Racine, 2012), leave-subject-out CV (Liao et al., 2019) and minimization of Kullback-Leibler type measures Zhang et al. (2015). However, the majority of these methods are built on ordinary least squares or maximum likelihood, which are expected to be sensitive to outliers or violation of certain model assumption. This means that when the data contains outliers, they could be failed.

On the other hand, outliers are commonly found in almost all fields. They may appear as a result of improperly including a fraction of a sample from a different population, or by measurement errors. Especially in the era of big data, the amount of data is huge and intricate, and in this case outliers are often unavoidable. Outliers in the sample can have significant effects on some common statistical methods. For example, the ordinary least squares is a nonresistant fitting procedure and a small proportion of the data can strongly influence the fitted model. If some of these influential cases are aberrant, the results will be disastrous for the fit. To solve the problem, two kinds of methods are considered. One is the outlier detection (see, for example, Hawkins (1980)), which constructs outlier diagnostic statistics to find the influential observations. Then, these outliers are modified or deleted directly, and finally the tradition methods for the data analysis are employed to the processed data. It should be noted that even if the data undergoes outlier detection, there would still be outliers in the data, as commented by Huber (1973) that outliers are much harder to spot in the regression than in the simple location case. The other is the robust method (see, for example, Hampel et al. (1986)), which uses robust loss functions to obtain estimators regardless of whether there are outliers in the data set. The outstanding merit of this approach is that it can achieve trustworthy results even if the data is contaminated. In this paper, we focus on the latter method.

It is also clear that the presence of outliers may have significant effects on model selection. Therefore, how to eliminate such an influence has increasingly attracted the attention of statisticians. In fact, outliers robust model selection has been an important research direction of robust statistics, and many valid methods have been proposed. Most statisticians developed their robust model selection approaches by adjusting the popular criteria. For instance, Hampel (1983) and Ronchetti (1985) suggested a robust version of Akaike Information Criterion (AIC) and investigated its properties. Burman and Nolan (1995) presented a general Akaike-type criterion which is applicable to a variety of loss functions for model selection. Ronchetti and Staudte (1994) presented a modified version of Mallows' C_p (Mallows, 1973) by weighted residual sum of squares. It allows us to choose a model that fits most of the data in the presence of outliers (see also Sommer and Staudte (1995)). Ronchetti (1997) reviewed this criterion as well as some other approaches. He stressed that there remains much work to be done, such as robust model selection in time series and developing other robust model selection procedures. Müller and Welsh (2005) proposed a new robust model selection criterion built on combining a robust penalized criterion and a robust conditional expected prediction loss function which is estimated using a stratified bootstrap. On the other hand, some approaches based on resampling are developed. For example, Ronchetti et al. (1997) suggested a robust model selection technique for regression based on cross validation. Wisnowski et al. (2003)

proposed a variable selection approach for robust regression by combining robust estimation and resampling variable selection techniques. However, the aforementioned methods are built on the popular M-estimation procedure, which cannot be resistant to outliers in the covariates, i.e., the leverage points. To overcome this drawback, the generalized M (GM) estimators are developed in many literatures such as (He et al., 2000).

Although many robust methods have been developed for model selection, the robust model averaging criterion has not been well studied when the samples are contaminated by outliers, except in the case of quantile regression, where Lu and Su (2015) and Wang et al. (2023) proposed jackknife model averaging methods which select the weights by minimizing a leave-one-out CV criterion. For the mean regression, Du et al. (2018) developed robust versions of the focused information criterion and a frequentist model average estimator based on M-estimation, and Guo and Li (2021) suggested a robust model averaging method based on S_p criterion with the same penalty as in the Mallows' criterion of Hansen (2007). However, the weight choice criteria in these two papers are constructed on the basis of intuitive consideration, and no optimality property is shown. Further, it is noteworthy that the aforementioned robust model averaging methods are not resistant to the outliers in covariates. The purpose of this article is to develop an *optimal* model averaging approach for the mean regression which is robust to the outliers occurring only in the response, or only in the covariates, or in both, and can be applicable to a wide variety of loss functions including quantile function. In order to achieve this goal, we employ a class of the GM-type loss functions to obtain robust parameter and weight estimators. Unlike the case of conventional Mallows model averaging where the squared loss function is used, no explicit form of estimator can be obtained for general robust loss functions, which brings technical challenges to the derivation of the robust weighting scheme. To overcome this difficulty, it is essential to derive the asymptotic expression of the difference between the robust parameter estimator (say, $\widehat{\Theta}$) and the pseudo-true parameter (say, Θ^*, which is defined in Section 3.1). In this regard, the technique in

Pollard (1991), who proved that the difference between the least absolute deviations estimator and the true parameter has an asymptotic expression, is a useful tool. However, for model averaging, candidate models are often misspecified, and so Pollard's approach cannot be used directly here. By some modifications of the proofs in Pollard (1991), we successfully show that $\widehat{\Theta} - \Theta^* = \eta + o_p(1)$ under some regularity conditions, where η has a closed-form expression.

As an important measurement of robustness, the influence function of a robust estimator has been well studied in the past decades. In our paper, we explore the derivation of the influence function in the model averaging framework. Furthermore, the empirical prediction influence function (EPIF), which can be calculated by the sample, is proposed to characterize the quantitative robustness of our proposed model averaging estimator.

The remainder of the article is organized as follows. Section 2 describes the model framework and estimators. In Sections 3 and 4, we present weight selection criteria with fixed design matrix and random design matrix, respectively. Section 5 establishes theoretical properties of our proposed robust model averaging estimators and weighting schemes under some regularity conditions. Section 6 investigates the finite sample performance of our proposed method by simulations, and we apply the proposed method to a real data set in Section 7. Some concluding remarks are contained in Section 8. The robustness property of the model averaging estimator, the proofs of theorems, two robust versions of Mallows' C_p and Mallows model averaging method for comparison, and additional simulation studies are given in the Supplementary Materials.

2 Model framework and model averaging estimator

Suppose that the random sample $\{y_i\}$ is from the following data generating process

$$y_i = \mu_i + \varepsilon_i = \mu(\boldsymbol{x}_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\mu(\boldsymbol{x}_i)$ is a function of $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^T$; and ε_i, $i = 1, ..., n$, are independent errors from the distribution F which has a continuous density f with respect to Lebesgue measure (The similar assumption can be found in, say, Coakley and Hettamansperger (1993) and Wang et al. (2013)).

In our paper, we aim to robustly estimate $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^T$ by model averaging approach. To this end, we will present a GM-estimation procedure applicable to a variety of loss functions. For instance, based on the least absolute deviation or Huber's function, and some weight functions depending only on the covariates, we can obtain efficient estimators in the presence of outliers in the response and/or covariates. To be specific, we consider M candidate linear models, and let the m^{th} one be

$$y_i = \boldsymbol{x}_{i(m)}^T \Theta_{(m)} + \varepsilon_{i(m)} = \sum_{j=1}^{k_m} \theta_{j(m)} x_{ij(m)} + \varepsilon_{i(m)}, \quad i = 1, ..., n, \tag{1}$$

where $\Theta_{(m)} = (\theta_{1(m)}, \ldots, \theta_{k_m(m)})^T$, $\boldsymbol{x}_{i(m)} = (x_{i1(m)}, \ldots, x_{ik_m(m)})^T$ with $x_{ij(m)}$ being a variable in $\boldsymbol{x}_i$ that appears as a regressor in the m^{th} model, and $\theta_{j(m)}$ being the corresponding coefficient, $j = 1, \ldots, k_m$, and k_m is the number of covariates. The parameter estimator $\widehat{\Theta}_{(m)}$ under the model (1) is defined as the solution which minimizes the GM-type objective function

$$Q_n(\Theta_{(m)}) = \sum_{i=1}^{n} h(\boldsymbol{x}_{i(m)}) \rho \left(y_i - \boldsymbol{x}_{i(m)}^T \Theta_{(m)} \right), \tag{2}$$

where ρ is a robust convex loss function which protects against outliers in the response, and $h(\cdot)$ is a bounded function which downweights the leverage points. Let $\widehat{\varepsilon}_{i(m)} = y_i - \boldsymbol{x}_{i(m)}^T \widehat{\Theta}_{(m)}$ and $\boldsymbol{w} = (w_1, \ldots, w_M)^T$ be a weight vector in the unit simplex of $\mathbb{R}^M$: $\mathcal{W} = \left\{ \boldsymbol{w} \in [0, 1]^M : \sum_{m=1}^{M} w_m = 1 \right\}$. The model averaging estimator of μ_i is thus

$$\widehat{\mu}_i(\boldsymbol{w}) = \sum_{m=1}^{M} w_m \boldsymbol{x}_{i(m)}^T \widehat{\Theta}_{(m)}. \tag{3}$$

To determine the weight $\boldsymbol{w}$, we will develop a robust Mallows-type weight estimator, denoted by $\overline{\boldsymbol{w}} = (\overline{w}_1, ..., \overline{w}_M)^T$. Substituting $\overline{\boldsymbol{w}}$ for $\boldsymbol{w}$ in (3) results in the following robust Mallows-type model averaging (RMMA) estimator of μ_i: $\widehat{\mu}_i(\overline{\boldsymbol{w}}) = \sum_{m=1}^{M} \overline{w}_m \boldsymbol{x}_{i(m)}^T \widehat{\Theta}_{(m)}$. So the RMMA estimator of $\boldsymbol{\mu}$ is given by $\widehat{\boldsymbol{\mu}}(\overline{\boldsymbol{w}}) = (\widehat{\mu}_1(\overline{\boldsymbol{w}}), ..., \widehat{\mu}_n(\overline{\boldsymbol{w}}))^T$.

3 Weight selection criterion with fixed design matrix

In this section, we consider the case of fixed design matrix. We will first present the asymptotic expansion of final prediction error for model averaging estimator, and then develop a robust model averaging method.

3.1 The asymptotic expansion of final prediction error

We start with some notations. Define the pseudo-true parameter for the m^{th} candidate model as

$$\Theta^*_{(m)} = \underset{\Theta_{(m)} \in \mathbb{R}^{k_m}}{\arg\min} \sum_{i=1}^{n} \mathrm{E}\left\{ h(\boldsymbol{x}_{i(m)}) \rho \left(y_i - \boldsymbol{x}_{i(m)}^T \Theta_{(m)} \right) \right\},$$

$\varepsilon^*_{i(m)} = y_i - \boldsymbol{x}_{i(m)}^T \Theta^*_{(m)}$, $\varepsilon^*_i(\boldsymbol{w}) = y_i - \sum_{m=1}^{M} w_m \boldsymbol{x}_{i(m)}^T \Theta^*_{(m)}$, and $\widehat{\varepsilon}_i(\boldsymbol{w}) = y_i - \widehat{\mu}_i(\boldsymbol{w})$. Assume that the design matrix $\boldsymbol{X} = \left(\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_n^T \right)^T$ is fixed. Let $\{\tilde{y}_i\}_{i=1}^n$ be a (unknown) sample independent of but having the same probability structure with y_i, i.e., $\tilde{y}_i = \mu(\boldsymbol{x}_i) + \tilde{\varepsilon}_i$ with $\tilde{\varepsilon}_i$ being from the distribution F and independent of ε_i.

In order to select the dimension of the fitted model, Akaike (1970) suggested estimating the m^{th} model's predictive capability, or final prediction error, defined as $\sum_{i=1}^{n} \mathrm{E} \left(\tilde{y}_i - \boldsymbol{x}_{i(m)}^T \widehat{\Theta}_{(m)} \right)^2$. We follow this method and choose weights of $\widehat{\mu}_i(\boldsymbol{w})$ by estimating a robust version of the final prediction error, i.e.,

$$\sum_{i=1}^{n} \mathrm{E}\left[h(\boldsymbol{x}_{i(M)}) \rho \left\{ \tilde{y}_i - \widehat{\mu}_i(\boldsymbol{w}) \right\} \right], \tag{4}$$

where ρ is some robust convex loss function, and the M^{th} candidate model is assumed to be the largest one containing all the candidate models' covariates. Since the model averaging estimator is built on the covariates of all candidate models, we consider the weight function depending on $\boldsymbol{x}_{i(M)}$ in (4). It is worthy to note that our following results remain true as long as $\boldsymbol{x}_{i(M)}$ contains the covariates of all candidate models, and it does not necessarily correspond to the largest candidate model. However, it is difficult to directly estimate (4). So to obtain our robust weighting scheme, we derive the asymptotic expansion of (4) first.

THEOREM 1: *Suppose that Assumptions S1-S5 in Section S1 of the Supplementary Materials* *hold, then*

$$
\begin{aligned}
&\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\widetilde{\mathrm{E}}\left[\rho\left\{\tilde{y}_i-\widehat{\mu}_i(\boldsymbol{w})\right\}\right]-\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\rho\{\widehat{\varepsilon}_i(\boldsymbol{w})\}-\sum_{m=1}^{M} w_m H_{nm}\\
=&\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\mathrm{E}\left[\rho\left\{\varepsilon_i^*(\boldsymbol{w})\right\}\right]-\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\rho\{\varepsilon_i^*(\boldsymbol{w})\}\\
&-\sum_{m=1}^{M} w_m\left(A_{nm}^{-1}V_{nm}\right)^T\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\mathrm{E}\left[\boldsymbol{x}_{i(m)}\rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right]+o_p(1),
\end{aligned}
\tag{5}
$$

where $V_{nm}=\sum_{i=1}^{n} h(\boldsymbol{x}_{i(m)})\boldsymbol{x}_{i(m)}\rho_1\left(\varepsilon_{i(m)}^*\right)$, $A_{nm}=\sum_{i=1}^{n} h(\boldsymbol{x}_{i(m)})R_2\left(\mu_i-\boldsymbol{x}_{i(m)}^T\Theta_{(m)}^*\right)\boldsymbol{x}_{i(m)}\boldsymbol{x}_{i(m)}^T$, $H_{nm}=\left(A_{nm}^{-1}V_{nm}\right)^T\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\boldsymbol{x}_{i(m)}\rho_1\{\varepsilon_i^*(\boldsymbol{w})\}$ *and* $\widetilde{\mathrm{E}}$ *denotes the expectation with respect to* $\tilde{y}_i$.

Proof. See the Supplementary Materials.

Theorem 1 shows that the robust version of the final prediction error of the model averaging estimator can be written as a main term plus a term of $o_p(1)$, which is a basis for us to generate our robust weight selection criterion.

If we let $h(\cdot)\equiv 1$, then by the definition of $\Theta_{(m)}^*$, we obtain $\sum_{i=1}^{n}\boldsymbol{x}_{i(m)}\mathrm{E}\left\{\rho_1\left(\varepsilon_{i(m)}^*\right)\right\}=0$. Thus, if we put all the weights on the m^{th} candidate model, then (5) reduces to

$$
\begin{aligned}
&\sum_{i=1}^{n}\widetilde{\mathrm{E}}\left\{\rho\left(\tilde{y}_i-\boldsymbol{x}_{i(m)}^T\widehat{\Theta}_{(m)}\right)\right\}-\sum_{i=1}^{n}\rho(\widehat{\varepsilon}_{i(m)})-V_{nm}^TA_{nm}^{-1}V_{nm}\\
=&\sum_{i=1}^{n}\mathrm{E}\left\{\rho(\varepsilon_{i(m)}^*)\right\}-\sum_{i=1}^{n}\rho(\varepsilon_{i(m)}^*)+o_p(1),
\end{aligned}
\tag{6}
$$

which is similar to the result derived by Burman and Nolan (1995) for model selection.

3.2 A robust Mallows-type criterion of weight choice

In order to derive a robust Mallows-type criterion for choosing the weights in model averaging estimator, we make use of Theorem 1 and ignore the term of $o_p(1)$. Note that from the definition of $\Theta_{(m)}^*$, we have $\mathrm{E}\left(V_{nm}\right)=0$. Therefore, by taking expectation on both sides of (5), we obtain

$$
\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\mathrm{E}\left[\rho\left\{\tilde{y}_i-\widehat{\mu}_i(\boldsymbol{w})\right\}\right]\approx\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\mathrm{E}\left[\rho\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\}\right]+\mathrm{E}\left(\sum_{m=1}^{M} w_m H_{nm}\right). \tag{7}
$$

Let $\widetilde{V}_{nm} = \sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\boldsymbol{x}_{i(m)}\rho_1\{\varepsilon_i^*(\boldsymbol{w})\}$, and then the second term on the right-hand side of (7) can be rewritten as the following form:

$$\begin{aligned}&\mathrm{E}\left(\sum_{m=1}^{M} w_m H_{nm}\right)\\ &=\sum_{m=1}^{M} w_m \sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})h(\boldsymbol{x}_{i(m)})\mathrm{Cov}\left[\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right]\boldsymbol{x}_{i(m)}^T A_{nm}^{-1}\boldsymbol{x}_{i(m)}. \end{aligned} \tag{8}$$

Thus, (7) is equivalent to

$$\begin{aligned}&\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\mathrm{E}\left[\rho\left\{\tilde{y}_i - \widehat{\mu}_i(\boldsymbol{w})\right\}\right]\\ &\approx \sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\mathrm{E}\left[\rho\{\widehat{\varepsilon}_i(\boldsymbol{w})\}\right]\\ &\quad+\sum_{m=1}^{M} w_m \sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})h(\boldsymbol{x}_{i(m)})\mathrm{Cov}\left[\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right]\boldsymbol{x}_{i(m)}^T A_{nm}^{-1}\boldsymbol{x}_{i(m)}. \end{aligned} \tag{9}$$

Following Burman and Nolan (1995), we approximate $h(\boldsymbol{x}_{i(M)})h(\boldsymbol{x}_{i(m)})\mathrm{Cov}\left[\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right]$ by the average $n^{-1}\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})h(\boldsymbol{x}_{i(m)})\mathrm{Cov}\left[\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right]$. Similarly, we can use

$$\left\{n^{-1}\sum_{i=1}^{n} h(\boldsymbol{x}_{i(m)})R_2\left(\mu_i - \boldsymbol{x}_{i(m)}^T\Theta_{(m)}^*\right)\right\}\sum_{i=1}^{n}\boldsymbol{x}_{i(m)}\boldsymbol{x}_{i(m)}^T$$

to estimate A_{nm}. Thus, the second term on the right-hand side of (9) can be estimated by

$$\sum_{m=1}^{M} w_m k_m \frac{\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})h(\boldsymbol{x}_{i(m)})\mathrm{Cov}\left[\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right]}{\sum_{i=1}^{n} h(\boldsymbol{x}_{i(m)})R_2\left(\mu_i - \boldsymbol{x}_{i(m)}^T\Theta_{(m)}^*\right)}. \tag{10}$$

Combining (9) and (10), we see that the final prediction error of model averaging estimator is given by

$$\begin{aligned}&\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\mathrm{E}\left[\rho\left\{\tilde{y}_i - \widehat{\mu}_i(\boldsymbol{w})\right\}\right]\\ &\approx \sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\mathrm{E}\left[\rho\{\widehat{\varepsilon}_i(\boldsymbol{w})\}\right] + \sum_{m=1}^{M} w_m k_m \frac{\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})h(\boldsymbol{x}_{i(m)})\mathrm{Cov}\left[\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right]}{\sum_{i=1}^{n} h(\boldsymbol{x}_{i(m)})R_2\left(\mu_i - \boldsymbol{x}_{i(m)}^T\Theta_{(m)}^*\right)}. \end{aligned}$$

Based on the above expression on the final prediction error, we propose the following robust

Mallows-type criterion for choosing weights

$$C_n(\boldsymbol{w}) = \sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})\rho\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\} + \sum_{m=1}^{M} w_m k_m C_{\rho(m)}, \tag{11}$$

where

$$C_{\rho(m)} = \frac{\sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})h(\boldsymbol{x}_{i(m)})\mathrm{Cov}\left[\rho_1\left(\varepsilon^*_{i(m)}\right), \rho_1\{\varepsilon^*_i(\boldsymbol{w})\}\right]}{\sum_{i=1}^{n} h(\boldsymbol{x}_{i(m)})R_2\left(\mu_i - \boldsymbol{x}^T_{i(m)}\Theta^*_{(m)}\right)}. \tag{12}$$

Since $\mathrm{Cov}\left[\rho_1\left(\varepsilon^*_{i(m)}\right), \rho_1\{\varepsilon^*_i(\boldsymbol{w})\}\right]$ and $R_2\left(\mu_i - \boldsymbol{x}^T_{i(m)}\Theta^*_{(m)}\right)$ are unknown, $C_{\rho(m)}$ needs to be estimated. A natural estimator of $C_{\rho(m)}$ is $\widehat{C}_{\rho(m)} = \widehat{a}_{nm}/\widehat{r}_2$, where $\widehat{r}_2 = \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(m)}\right)\widehat{R}_2\left(\widehat{\varepsilon}_{i(m)}\right)$ with $\widehat{R}_2$ being an estimator of R_2 (See, for example, Burman and Nolan (1995)), and $\widehat{a}_{nm} = \sum_{i=1}^{n} h(\boldsymbol{x}_{i(M)})h\left(\boldsymbol{x}_{i(m)}\right)p_{in(m)}g_{in}(\boldsymbol{w})$ with $p_{in(m)} = \rho_1\left(\widehat{\varepsilon}_{i(m)}\right) - n^{-1}\sum_{i=1}^{n}\rho_1\left(\widehat{\varepsilon}_{i(m)}\right)$ and $g_{in}(\boldsymbol{w}) = \rho_1\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\} - n^{-1}\sum_{i=1}^{n}\rho_1\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\}$.

We will give the specific forms of $\widehat{R}_2$ in our examples. Thus, a feasible robust Mallows-type criterion of weight choice is given by

$$\widehat{C}_n(\boldsymbol{w}) = \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right)\rho\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\} + \sum_{m=1}^{M} w_m k_m \widehat{C}_{\rho(m)}. \tag{13}$$

The robust Mallows-type weight vector $\widehat{\boldsymbol{w}} = (\widehat{w}_1, \ldots, \widehat{w}_M)^T$ is obtained by choosing $\boldsymbol{w} \in \mathcal{W}$ such that $\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}\in\mathcal{W}} \widehat{C}_n(\boldsymbol{w})$. The corresponding RMMA estimator is $\widehat{\boldsymbol{\mu}}(\widehat{\boldsymbol{w}})$.

REMARK 1: When $h(\cdot) = 1$ and one component of the vector $\boldsymbol{w}$ is one and the others are zero, our proposed criterion (13) reduces to the model selection criterion given in Burman and Nolan (1995).

REMARK 2: If $\rho(t) = t^2$, our proposed criterion (11) coincides with the Mallows model averaging criterion proposed by Hansen (2007).

3.3 Examples

<u>Absolute loss</u>: When the squared loss function is replaced by the absolute deviation loss function, we find $\rho(t)$ is no longer differentiable, but it is differentiable almost everywhere. Taking the first derivative of $\rho(t)$ at any differentiable point, we have $\rho_1(t) = 1$ when $t > 0$, otherwise $\rho_1(t) = -1$. Some elementary calculations yield that $R_2(t) = 2f(-t)$. Hence, we can get

$$R_2\left(\mu_i - \sum_{m=1}^{M} w_m x_{i(m)}^T \Theta_{(m)}^*\right) = 2f\left(\sum_{m=1}^{M} w_m x_{i(m)}^T \Theta_{(m)}^* - \mu_i\right)$$

Obviously, $|\rho_1(t)| \leqslant 1$ for any $t \in \mathbb{R}$. Now let $\max_{1\leqslant i\leqslant n}\left|\sum_{m=1}^{M} w_m \boldsymbol{x}_{i(m)}^T \Theta_{(m)}^* - \mu_i\right|$ be bounded, then Assumption S2 is readily satisfied.

If we let the model bias be negligible, i.e., let $\mu_i - \boldsymbol{x}_{i(m)}^T \Theta_{(m)}^* \approx 0$ for $m = 1, ..., M$, and F have median 0 (see also Burman and Nolan (1995)), then $\text{Cov}\left[\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right] \approx \text{Var}\{\rho_1(\varepsilon_i)\} = 1$ and $R_2\left(\mu_i - \boldsymbol{x}_{i(m)}^T \Theta_{(m)}^*\right) \approx 2f(0)$.

We can estimate $f(0)$ by $\widehat{f}_M(0)$, where $\widehat{f}_M$ is an estimator of f based on the M^{th} candidate model. So we replace $h\left(\boldsymbol{x}_{i(m)}\right)$ by $h\left(\boldsymbol{x}_{i(M)}\right)$ in $C_{\rho(m)}$. In the simulation studies and real data analysis, $f(0)$ is estimated based on the Epanechnikov kernel with bandwidth being the semi-interquartile range of the residuals. Then, our proposed criterion (13) reduces to

$$\widehat{C}_n(\boldsymbol{w}) = \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right) |\widehat{\varepsilon}_i(\boldsymbol{w})| + \frac{\sum_{i=1}^{n} h^2\left(\boldsymbol{x}_{i(M)}\right)}{2\widehat{f}_M(0)\sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right)} \sum_{m=1}^{M} w_m k_m.$$

We label the above method as MA$_A$.

Huber's function: The Huber's function is given by

$$\rho(t) = \begin{cases} t^2 & |t| \leqslant c, \\ 2c|t| - c^2 & |t| > c, \end{cases}$$

with a constant c, that was proposed by Huber (1964) in robust regression and is smooth yet linear in the tails. Let $I(A)$ denote the indicator of event A. The first derivative of $\rho(t)$ is $\rho_1(t)$ and then for any $t \in \mathbb{R}$, $|\rho_1(t)| \leqslant 2c$. By some tedious calculations, we obtain

$$R_2\left(\mu_i - \sum_{m=1}^{M} w_m x_{i(m)}^T \Theta_{(m)}^*\right) = 2\text{P}\left(\left|\varepsilon_i + \mu_i - \sum_{m=1}^{M} w_m x_{i(m)}^T \Theta_{(m)}^*\right| \leqslant c\right).$$

Again, it is seen that if we let $\max_{1\leqslant i\leqslant n}\left|\sum_{m=1}^{M} w_m x_{i(m)}^T \Theta_{(m)}^* - \mu_i\right|$ be bounded, then Assumption S2 is readily satisfied.

As we did in the case of absolute loss, if we let the model bias be negligible, then we obtain

$$\mathrm{Cov}\left[\rho_1\left(\varepsilon^*_{i(m)}\right), \rho_1\{\varepsilon^*_i(\boldsymbol{w})\}\right] \approx \mathrm{Var}\{\rho_1(\varepsilon_i)\},$$

and $R_2\left(\mu_i - \boldsymbol{x}^T_{i(m)}\Theta^*_{(m)}\right) \approx R_2(0) = 2\mathrm{P}\left(|\varepsilon_i| \leqslant c\right)$. Similar to Burman and Nolan (1995), we require $\mathrm{E}\left\{\varepsilon_i I(|\varepsilon_i| \leqslant c)\right\} = 0$ and $\mathrm{F}(-c) = 1 - \mathrm{F}(c)$. These conditions are satisfied by any distribution that is symmetric about the origin. Hence, it follows that

$$\mathrm{Var}\{\rho_1(\varepsilon_i)\} = 4\mathrm{E}\left\{\varepsilon_i^2 I\left(|\varepsilon_i| \leqslant c\right)\right\} + 4c^2\mathrm{P}(|\varepsilon_i| > c).$$

So $\sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(m)}\right) R_2\left(\mu_i - \boldsymbol{x}^T_{i(m)}\Theta^*_{(m)}\right)$ can be estimated by $2n^{-1}\sum_{i=1}^n h\left(\boldsymbol{x}_{i(M)}\right) I\left(|\widehat{\varepsilon}_{i(M)}| \leqslant c\right)$, and $\sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right) h\left(\boldsymbol{x}_{i(m)}\right) \mathrm{Cov}\left[\rho_1\left(\varepsilon^*_{i(m)}\right), \rho_1\{\varepsilon^*_i(\boldsymbol{w})\}\right]$ can be estimated by

$$4n^{-1}\sum_{i=1}^{n} h^2\left(\boldsymbol{x}_{i(M)}\right) \widehat{\varepsilon}^2_{i(M)} I\left(|\widehat{\varepsilon}_{i(M)}| \leqslant c\right) + 4n^{-1}c^2\sum_{i=1}^{n} h^2\left(\boldsymbol{x}_{i(M)}\right) I\left(|\widehat{\varepsilon}_{i(M)}| > c\right).$$

Then, we can estimate $C_{\rho(m)}$ in (12) by

$$\widehat{C}_\rho = \frac{2\sum_{i=1}^{n} h^2\left(\boldsymbol{x}_{i(M)}\right)\left\{\widehat{\varepsilon}^2_{i(M)} I(|\widehat{\varepsilon}_{i(M)}| \leqslant c) + c^2 I(|\widehat{\varepsilon}_{i(M)}| > c)\right\}}{\sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right) I(|\widehat{\varepsilon}_{i(M)}| \leqslant c)}.$$

Thus, our proposed criterion (13) is given by

$$\widehat{C}_n(\boldsymbol{w}) = \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right)\rho\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\} + \widehat{C}_\rho\sum_{m=1}^{M} w_m k_m,$$

where ρ is the Huber's function. We label this method as MA_H.

4 Weight selection criterion with random design matrix

In this section, we present weight selection criterion for the case of random design matrix.

4.1 The asymptotic expansion of final prediction error

Assume that $\{\boldsymbol{x}_i\}$ are random vectors and independent of $\{\varepsilon_i\}$. We consider a new sequence of $\{\widetilde{y}_i\}_{i=1}^n$ which is independent of and identically distributed with $\{y_i\}_{i=1}^n$ conditional on **X**, i.e., $\tilde{y}_i = \mu(\boldsymbol{x}_i) + \tilde{\varepsilon}_i$ with $\tilde{\varepsilon}_i$ being independent of and identically distributed with ε_i. Let $\mathcal{F}_n = \sigma\left\{\bigcup_{i=1}^n \boldsymbol{x}_i\right\}$, $\mathcal{B}_i = \sigma\left\{\bigcup_{k=1}^i \varepsilon_k\right\}$ and $\mathcal{F}_{n,i} = \sigma\left\{\mathcal{F}_n \cup \mathcal{B}_i\right\}$ for $i \geqslant 1$. Define $\mathcal{B}_0 = \{\emptyset, \Omega\}$. Write $\mathrm{E}_{\mathcal{F}_n}(\cdot)$ and $\mathrm{E}_{\mathcal{F}_{n,i-1}}(\cdot)$ as the conditional expectation operators $\mathrm{E}\left(\cdot|\mathcal{F}_n\right)$ and $\mathrm{E}\left(\cdot|\mathcal{F}_{n,i-1}\right)$, respectively. Unlike the case where the design matrix is fixed, here we may define

$$\Theta_{(m)}^{*}=\underset{\Theta_{(m)}\in\mathbb{R}^{k_m}}{\arg\min}\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(m)}\right)\mathrm{E}_{\mathcal{F}_n}\left\{\rho\left(y_i-\boldsymbol{x}_{i(m)}^{T}\Theta_{(m)}\right)\right\}.$$

Similar to Theorem 1, the following theorem shows that, for the case of random design matrix, the final prediction error of the model averaging estimator can also be written as a main term plus a term of $o_p(1)$.

THEOREM 2: *Suppose that Assumptions S6-S10 in Section S1 of the Supplementary Materials hold, then*

$$\begin{aligned}&\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(M)}\right)\widetilde{\mathrm{E}}_{\mathcal{F}_n}\left[\rho\left\{\tilde{y}_i-\widehat{\mu}_i(\boldsymbol{w})\right\}\right]-\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(M)}\right)\rho\{\widehat{\varepsilon}_i(\boldsymbol{w})\}-\sum_{m=1}^{M}w_mH_{nm}^{*}\\&=\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(M)}\right)\mathrm{E}_{\mathcal{F}_n}\left[\rho\{\varepsilon_i^{*}(\boldsymbol{w})\}\right]-\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(M)}\right)\rho\{\varepsilon_i^{*}(\boldsymbol{w})\}\\&\quad-\sum_{m=1}^{M}w_m\left(A_{nm}^{*-1}V_{nm}\right)^{T}\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(M)}\right)\boldsymbol{x}_{i(m)}\mathrm{E}_{\mathcal{F}_n}[\rho_1\{\varepsilon_i^{*}(\boldsymbol{w})\}]+o_p(1),\end{aligned}\tag{14}$$

where $A_{nm}^{*}=\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(m)}\right)R_2^{*}\left(\mu_i-\boldsymbol{x}_{i(m)}^{T}\Theta_{(m)}^{*}\right)\boldsymbol{x}_{i(m)}\boldsymbol{x}_{i(m)}^{T}$, V_{nm} *is defined in Theorem 1 and* $H_{nm}^{*}=\left(A_{nm}^{*-1}V_{nm}\right)^{T}\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(M)}\right)\boldsymbol{x}_{i(m)}\rho_1\{\varepsilon_i^{*}(\boldsymbol{w})\}$.

Proof. See the Supplementary Materials.

4.2 A robust Mallows-type criterion of weight choice

To get a robust Mallows-type criterion of weight choice for the model averaging estimator, we ignore the small order term of (14). By the definition of $\Theta_{(m)}^{*}$, we have $\mathrm{E}_{\mathcal{F}_n}[V_{nm}]=0$. Therefore, invoking the law of iterated expectations, we obtain

$$\begin{aligned}&\sum_{m=1}^{M}w_m\mathrm{E}\left\{\left(A_{nm}^{*-1}V_{nm}\right)^{T}\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(M)}\right)\boldsymbol{x}_{i(m)}\mathrm{E}_{\mathcal{F}_n}[\rho_1\{\varepsilon_i^{*}(\boldsymbol{w})\}]\right\}\\&=\sum_{m=1}^{M}w_m\mathrm{E}\left\{\mathrm{E}_{\mathcal{F}_n}\left[\sum_{i=1}^{n}h\left(\boldsymbol{x}_{i(M)}\right)\boldsymbol{x}_{i(m)}^{T}\rho_1\{\varepsilon_i^{*}(\boldsymbol{w})\}\right]A_{nm}^{*-1}\mathrm{E}_{\mathcal{F}_n}[V_{nm}]\right\}=0.\end{aligned}$$

Taking expectation on both sides of (14) with the small order term ignored, we have

$$\sum_{i=1}^{n}\mathrm{E}\left[h\left(\boldsymbol{x}_{i(M)}\right)\rho\left\{\tilde{y}_i-\widehat{\mu}_i(\boldsymbol{w})\right\}\right]$$

$$\approx \sum_{i=1}^{n} \mathrm{E}\left[h\left(\boldsymbol{x}_{i(M)}\right)\rho\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\}\right] + \mathrm{E}\left\{\sum_{m=1}^{M} w_m \left(A_{nm}^{*-1}V_{nm}\right)^T \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right)\boldsymbol{x}_{i(m)}\rho_1\left\{\varepsilon_i^*(\boldsymbol{w})\right\}\right\}.$$

Since $\mathrm{E}\left(V_{nm}^T A_{nm}^{*-1}\widetilde{V}_{nm}\right) = \mathrm{E}\left[\mathrm{E}_{\mathcal{F}_n}(V_{nm}^T A_{nm}^{*-1}\widetilde{V}_{nm})\right]$, we can first estimate $\mathrm{E}_{\mathcal{F}_n}\left(V_{nm}^T A_{nm}^{*-1}\widetilde{V}_{nm}\right)$ instead of approximating $\mathrm{E}\left(V_{nm}^T A_{nm}^{*-1}\widetilde{V}_{nm}\right)$ directly. Similar to (8), we find that

$$\begin{aligned}&\mathrm{E}_{\mathcal{F}_n}\left(V_{nm}^T A_{nm}^{*-1}\widetilde{V}_{nm}\right)\\ =&\sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right) h\left(\boldsymbol{x}_{i(m)}\right)\mathrm{Cov}_{\mathcal{F}_n}\left[\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right]\boldsymbol{x}_{i(m)}^T A_{nm}^{*-1}\boldsymbol{x}_{i(m)}.\end{aligned}$$

Now, making use of the similar derivation for (11), with random design matrix, we propose the following robust Mallows-type criterion for choosing the weights in the model averaging estimator

$$\widetilde{C}_n(\boldsymbol{w}) = \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right)\rho\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\} + \sum_{m=1}^{M} w_m k_m C_{\rho(m)}^*,$$

where

$$C_{\rho(m)}^* = \frac{\sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right) h\left(\boldsymbol{x}_{i(m)}\right)\mathrm{Cov}_{\mathcal{F}_n}\left(\rho_1\left(\varepsilon_{i(m)}^*\right), \rho_1\{\varepsilon_i^*(\boldsymbol{w})\}\right)}{\sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(m)}\right) R_2^*\left(\mu_i - \boldsymbol{x}_{i(m)}^T\Theta_{(m)}^*\right)}.$$

Similar to the treatment of $C_{\rho(m)}$ in Section 3.2, $C_{\rho(m)}^*$ can be approximate by $\widehat{C}_{\rho(m)}^* = \widehat{a}_{nm}/\widehat{r}_2^*$, where $\widehat{r}_2^* = \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(m)}\right)\widehat{R}_2^*\left(\widehat{\varepsilon}_{i(m)}\right)$ with $\widehat{R}_2^*$ being an estimator of R_2^*. This yields the following feasible criterion of weight choice

$$\widetilde{C}_n^*(\boldsymbol{w}) = \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right)\rho\left\{\widehat{\varepsilon}_i(\boldsymbol{w})\right\} + \sum_{m=1}^{M} w_m k_m \widehat{C}_{\rho(m)}^*. \tag{15}$$

Accordingly, the optimal weight vector $\widetilde{\boldsymbol{w}} = (\widetilde{w}_1, \ldots, \widetilde{w}_M)^T$ can be obtained by choosing $\widetilde{\boldsymbol{w}} \in \mathcal{W}$ such that $\widetilde{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}\in\mathcal{W}} \widetilde{C}_n^*(\boldsymbol{w})$, and the resultant RMMA estimator is given by $\widehat{\boldsymbol{\mu}}(\widetilde{\boldsymbol{w}})$.

REMARK 3: In our paper, we aim to develop a robust Mallows-type weight selection criterion which can eliminate the influence caused by the outliers in sample. Unlike the case of conventional Mallows model averaging where the squared loss function is used, our parameter and weight estimators are derived under general robust loss functions such as absolute loss and Huber's function. Specifically, it is not feasible to derive closed-form expressions for parameter estimators in this context, which contrasts with situation where the loss function is selected to be the squared loss. Therefore, the construction method of Mallows criterion for the model average estimator, as outlined in Hansen (2007), is no longer applicable in our paper. To address this challenge, we adopt an approach inspired by Akaike (1970) and Burman and Nolan (1995). The optimal weight vector should be determined by minimizing the robust version of the final prediction error defined by (4). Therefore, we formulate the Mallows-type criterion by estimating (4). However, (4) is not directly estimable. In order to obtain a robust Mallows-type weight selection criterion (i.e., objective function), it is imperative to first derive the asymptotic representation of (4), by which we can establish the criteria (11) and (13).

REMARK 4: Comparing (13) and (15), we see that the two weight choice criteria have similar forms for fixed and random design matrices. Therefore, we can give similar approximations to $\widehat{C}^*_{\rho(m)}$ for the examples described in Section 3.3.

5 Asymptotic properties of the proposed estimators

In this section, we establish asymptotic optimality of our proposed robust model averaging estimator and consistency of the estimated weight vector for the case of random design matrix. With the case of fixed design matrix, the similar results can be derived, which are omitted for saving space.

5.1 Asymptotic optimality of model averaging estimator

In this subsection, we devote to establish the asymptotic optimality of the model averaging estimator $\widehat{\boldsymbol{\mu}}(\widetilde{\boldsymbol{w}})$ in the sense of minimizing the following out-of-sample final prediction error:

$$\mathrm{FPE}_n(\boldsymbol{w}) = \sum_{i=1}^{n} \mathrm{E}\left[h\left(\boldsymbol{x}_{i(M)}\right)\rho\left\{\widetilde{y}_i - \widehat{\mu}_i(\boldsymbol{w})\right\}|\mathcal{D}_n\right],$$

where $\mathcal{D}_n = \{(y_i, \boldsymbol{x}_i) : i = 1, ..., n\}$ and $\widetilde{y}_i$ $(i = 1, ..., n)$ are defined in Section 4.1.

THEOREM 3: *Under Assumptions S6-S14 in Section S1 of the Supplementary Materials, we have*

$$\frac{\mathrm{FPE}_n(\widetilde{\boldsymbol{w}})}{\inf_{\boldsymbol{w}\in\mathcal{W}} \mathrm{FPE}_n(\boldsymbol{w})} = 1 + o_p(1). \tag{16}$$

Proof. See the Supplementary Materials.

The optimality statement in Theorem 3 indicates that the model averaging estimator obtained using our proposed RMMA criterion yields an out-of-sample final prediction error which is asymptotically equivalent to that of the infeasible best possible averaging estimator.

5.2 Consistency of the estimated weight vector

In this subsection, we consider the consistency of $\widetilde{\boldsymbol{w}}$. Let

$$\widetilde{\mathrm{FPE}}_n(\boldsymbol{w}) = \sum_{i=1}^{n} h\left(\boldsymbol{x}_{i(M)}\right) \mathrm{E}_{\mathcal{F}_n}\left\{\rho\left(y_i - \sum_{m=1}^{M} w_m \boldsymbol{x}_{i(m)}^T \Theta_{(m)}^*\right)\right\}$$

and define the theoretically optimal weight as $\boldsymbol{w}^0 = \arg\min_{\boldsymbol{w}\in\mathcal{W}} \widetilde{\mathrm{FPE}}_n(\boldsymbol{w})$.

THEOREM 4: *Under Assumptions S6-S17 in Section S1 of the Supplementary Materials, if $\boldsymbol{w}^0$ is an interior point of $\mathcal{W}$, then there exists a local minimizer $\widetilde{\boldsymbol{w}}$ of $\widetilde{C}_n^*(\boldsymbol{w})$ such that*

$$\|\widetilde{\boldsymbol{w}} - \boldsymbol{w}^0\| = O_p(n^{-1/4}). \tag{17}$$

Proof. See the Supplementary Materials.

Theorem 4 demonstrates that the weight estimator $\widetilde{\boldsymbol{w}}$ approaches to the theoretically optimal weight vector $\boldsymbol{w}^0$ at the rate of $n^{-1/4}$.

In this paper, we also investigate the robustness property of the model averaging estimator. To be specific, we define the influence function of model averaging estimator, and then prove that the optimal model averaging estimator is robust in the sense that it has a bounded influence function. Please refer to Section S2 of the Supplementary Materials.

6 Simulation Studies

The purpose of this section is to evaluate, via simulation studies, the finite sample performance of our proposed method, and compare it with some other commonly used model selection and averaging methods. We label the two robust versions of Mallows' C_p which were introduced in Ronchetti and Staudte (1994) as HC_p and MC_p when the weight functions of Huber's and Mallows' types are used, respectively. The general Akaike-type model selection methods in Burman and Nolan (1995) are labeled as MS_A and MS_H, and the S_p-type robust model averaging methods suggested by Guo and Li (2021) are labeled as SMA_A and SMA_H, when the loss functions are chosen to be the absolute deviation and Huber's functions, respectively. We also consider the conventional model averaging estimators, including MMA (Hansen, 2007), SAIC and SBIC. As in Ronchetti and Staudte (1994), we let $c = 1.345$ in our simulation studies and real data analysis. We compare totally eleven estimators including MA_A, MS_A, SMA_A, MA_H, MS_H, SMA_H, MC_p, HC_p, MMA, SAIC and SBIC. To this end, we consider similar simulation settings to those in Hansen (2007), and use the following out-of-sample mean absolute error (MAE) across $R = 1000$ replications to evaluate their performance:

$$\text{MAE} = \frac{1}{R}\sum_{r=1}^{R}\text{AE}(r),$$

where $\text{AE}(r) = \frac{1}{n}\sum_{i=1}^{n}\left|\mu_i^{(r)} - \widehat{\mu}_i^{(r)}\right|$ is the error from the r^{th} replication based on a given averaging/selection method with $\mu_i^{(r)}$ being calculated using the clean testing dataset $\{\boldsymbol{x}_i^{(r)}\}_{i=1}^{n}$ and $\widehat{\mu}_i^{(r)}$ being the estimation value.

In our simulation studies and real data example, the weight function has the form of $h\left(\boldsymbol{x}_{i(m)}\right) = \psi_b(h_{i(m)})/h_{i(m)}$, where $h_{i(m)}$ is the i^{th} diagonal element of the "hat matrix" $H_{(m)} = \boldsymbol{x}_{(m)}(\boldsymbol{x}_{(m)}^T\boldsymbol{x}_{(m)})^{-1}\boldsymbol{x}_{(m)}^T$ with $\boldsymbol{x}_{(m)} = (\boldsymbol{x}_{1(m)}, ..., \boldsymbol{x}_{n(m)})^T$, and $\psi_b(h_{i(m)}) = h_{i(m)}$ if $|h_{i(m)}| \leqslant b$ and $\psi_b(h_{i(m)}) = b$ if $|h_{i(m)}| > b$. As in Sommer and Staudte (1995), we let the bending constant $b = 1.5k_m/n$, then $\boldsymbol{x}_{i(m)}$ will be downweighted when $h_{i(m)}$ is larger than 1.5 times of the average leverage.

6.1 Simulation settings

Following the simulation setting in Hansen (2007), we consider

$$y_i = \sum_{j=1}^{100}\theta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{18}$$

where $x_{i1} = 1$ is the intercept and the remaining x_{ij} ($j = 2, ..., 100$) are mutually independent, the parameters $\theta_j = c\sqrt{2\nu}j^{-\nu-0.5}$ ($j = 1, ..., 100$) with $\nu = 1$ and c being varied so that $R^2 = c^2/(1+c^2) = 0.1, 0.3, \cdots, 0.9$, and ε_i is independent of the covariates. For the random covariates x_{ij} ($j = 2, ..., 100$) and the error term ε_i, we consider the following four cases:

Case 1 The random covariates x_{ij} ($j = 2, ..., 100$) and the error term ε_i follow the standard normal distribution $\mathcal{N}(0,1)$;

Case 2 The random covariates x_{ij} ($j = 2, ..., 100$) follow a mixture distribution $0.8\mathcal{N}(0,1) + 0.2t(1)$, and the error term ε_i follows the standard normal distribution $\mathcal{N}(0,1)$;

Case 3 The random covariates x_{ij} ($j = 2, ..., 100$) follow the standard normal distribution $\mathcal{N}(0,1)$, and the error term ε_i follows a distribution $t(1)$;

Case 4 The random covariates x_{ij} ($j = 2, ..., 100$) follow a mixture normal distribution $0.8\mathcal{N}(0,1)+0.2t(1)$, and the error term ε_i follows a distribution $t(1)$.

Cases 1-4 correspond to the following four situations: there are no outliers in sample, outliers occur only in the covariates, outliers occur only in the response, and both covariates and response contain outliers. We consider the nested regression models with variables $\{x_{ij}, j = 2, ..., M\}$ and each candidate model contains the intercept term, where $M = [3n^{1/3}]$ with $n = 50, 150$ and 400. In the implementation of simulation studies, for simplicity, the random covariates in Cases 2 and 4 are generated in the following way: 80% of the random covariates come from $\mathcal{N}(0,1)$ and the remaining 20% come from $t(1)$.

6.2 Simulation results

Tables 1-4 report the MAEs of various estimators for Cases 1-4 with $n = 50, 150$ and 400. To facilitate comparisons, the best is shown in bold.

From Table 1, we see that for Case 1, the MMA estimator is most frequently the estimator that enjoys the smallest MAEs. The reason is that the MMA estimator is asymptotically optimal in the sense of minimizing the squared errors in the absence of outliers. When the data is not contaminated by outliers, in most cases, we find that MA$_H$ and MA$_A$ are slightly worse than MMA, but MA$_H$ clearly denominates the other robust model selection and averaging methods. As for MA$_A$, it usually performs better than SMA$_A$, MS$_A$, MC$_p$ and HC$_p$. Another interesting finding is that for the sample sizes of $n = 150$ and 400, MA$_A$ and MA$_H$ can often have a better performance than SBIC.

[Table 1 about here.]

[Table 2 about here.]

The advantages of MA_A and MA_H are quite clear in the presence of outliers. For instance, from the results of Case 2 in Table 2, one can see that when outliers occur only in the covariates, MA_H is the most favored estimator, and MA_A is often the second favored estimator. We also find that the conventional model averaging estimators, including MMA, SAIC and SBIC, have the worst performance, which shows that it is necessary to develop a robust model averaging method. For Case 3, where the error term comes from the distribution with heavy tails and the covariates are absent of outliers, frequently MA_A enjoys smaller MAEs than MA_H. It is clear from Table 3 that MA_A and MA_H often occupy the top two when outliers occur only in the response. Compared to the results of Case 2, the performance of MS_A becomes better than that of MS_H, which indicates that the estimator built on the absolute deviation loss function is more favorable than that based on the Huber's function when the outliers occur in the response. The performance of the remaining estimators is similar to that in Case 2. If the sample is contaminated by outliers in both the covariates and the response, we can see from the results of Case 4 in Table 4 that when the sample size $n = 50$, MA_A performs the best and the performance of MA_H is not good enough. However, for the sample sizes of $n = 150$ and $n = 400$, MA_A and MA_H always perform better than the other model selection and averaging methods. It is found that MS_A and MS_H have smaller MAEs than SMA_A and SMA_H respectively, and HC_p and MC_p are superior to SMA_A and SMA_H. Of all cases considered in Tables 1-4, as the sample size increases, the MAEs of MA_A and MA_H become smaller. On the other hand, we find that different sample sizes have little effect on the ranking of estimators' performance.

[Table 3 about here.]

[Table 4 about here.]

In summary, in the absence of outliers, the commonly used MMA method is often superior to other methods in terms of minimizing MAEs, but it is usually the worst when there are outliers in data. Moreover, the widely used SAIC and SBIC perform poorly in the presence of outliers. So it is very meaningful to develop robust model averaging methods which can be resistant to the leverage points and outliers in the response. It is observed that MA_A and MA_H perform better than MS_A and MS_H, respectively. Further, MC_p and HC_p are usually inferior to our proposed model averaging estimators especially in the cases of $n = 150$ and 400, which highlights the advantages of robust model averaging in the face of model diversity and outliers. It is worthy to note that when there are outliers in the sample, our proposed robust Mallows-type model averaging estimators MA_A and MA_H are usually superior to SMA_A and SMA_H.

Further, we consider the data generating process (18) with the covariates being generated from a mixture multidimensional distribution. The corresponding results are provided in Section S9.1 of the Supplementary Materials. We also adopt the same simulation setting as in Ronchetti (1985) in Section S9.2 of the Supplementary Materials, which is a non-linear model. In conclusion, for both dependent data and complex structured data considered here, MA_A and MA_H still perform well when the error terms come from the distributions with heavy tails. More details can be found in the Supplementary Materials.

7 Real data example

As an application of our proposed method, we analyze the human immunodeficiency virus (HIV) data which are from acquired immunodeficiency syndrome (AIDS) Clinical Trials Group (ACTG) protocol 175 (Hammer et al., 1996) and can be found in the `R` package `speff2trial`. The ACTG 175 experiment evaluates treatment with either a single nucleoside or two nucleosides in adults infected with HIV type 1 (HIV-1) whose CD4 cell counts range from 200 to 500 per cubic millimeter. According to the regimen of treatment they received, the patients were divided into two arms: the arm with zidovudine, ZDV, monotherapy (ZDV only) and the arm with three newer treatments (ZDV + didanosine, ddI, ZDV + zalcitabine, ddC, and ddI only). The two arms totally have 2139 subjects.

Following the analysis of Han et al. (2019), CD4 cell count at 96 ± 5 weeks post baseline (CD4$_{96}$) is chosen to be the response variable, and the eight variables including treatment indictor (trt; 0=ZDV only), CD4 cell count at baseline (CD4$_0$), age in years at baseline (age), weight in kg at baseline (weight), race (race; 0=white), gender (gender; 0=female), history of intravenous drug use (drug; 0=no), indicator of off-treatment before 96 ± 5 weeks (offtrt; 0=no) are chosen to be the covariates. Wang et al. (2023) took CD4 cell count at 20 ± 5 weeks (CD4$_{20}$) and CD8 cell count at 20 ± 5 weeks (CD8$_{20}$) as the predictors. Thus, we also add these two variables into the covariates. After removing the subjects with the response variable CD4$_{96}$ being missing, we still have $n = 1342$ sample observations. Figure 1 shows the boxplots of the response CD4$_{96}$ and the covariates CD4$_0$, CD4$_{20}$ and CD8$_{20}$, which indicates that there are outliers in both covariates and response. Based on the absolute value of correlation between the covariate and the response variable, the order of the ten covariates (from large to small) is x_1 (CD4$_{20}$), x_2 (CD4$_0$), x_3 (offtrt), x_4 (trt), x_5 (weight), x_6 (CD8$_{20}$), x_7 (race), x_8 (drug), x_9 (age) and x_{10} (gender). We construct ten nested candidate models with covariates $\{1, x_1\}$, $\{1, x_1, x_2\}$,..., $\{1, x_1, x_2, ..., x_{10}\}$, respectively. Some covariates (CD4$_{20}$, CD4$_0$, weight, CD8$_{20}$, and age) are standardized to have zero mean and unit variance.

[Figure 1 about here.]

We randomly divide our sample with size of n into a training sample $\{\boldsymbol{x}_s, y_s\}_{s=1}^{n_1}$ and an evaluation sample $\{\boldsymbol{x}_t, y_t\}_{t=1}^{n_2}$. Let $\widehat{\mu}_t$ be the predictive value of the response variable based on a given averaging/selection method. Then we evaluate different methods by calculating the absolute prediction error (APE), i.e., $\text{APE} = \frac{1}{n_2} \sum_{t=1}^{n_2} |y_t - \widehat{\mu}_t|$.

We take $n_1 = 600$ and 800, respectively. To demonstrate the benefits of the robust methods, we consider the case of more outliers in the training sample, where 20% of CD4$_{20}$, CD4$_0$, weight and $\mathrm{CD8}_{20}$ are randomly replaced with a sample from a heavy-tailed distribution $t(1)$. All the simulation results are reported in Figure 2. It is observed that MA_A yields the best result among all the estimators, and the second favored estimator is MA_H. From Figure 2, we also find that MC_p can clearly dominate the other robust model selection methods. Further, it is seen that the performance differences between our proposed MA_A and MA_H and other robust model averaging estimators (SMA_A and SMA_H) are quite clear. The performance of the traditional model averaging methods MMA, SAIC and SBIC in this real data analysis is poor, which further shows the necessity of developing a robust model averaging method. This example indicates that our method can obtain a trustworthy result when the real data contain outliers.

[Figure 2 about here.]

8 Concluding remarks

A great progress with the model averaging method has been made in past decades. However, the commonly used methods do not consider the influence of outliers. In fact, most of the existing model averaging methods are based on the squared loss function, and our numerical results have indicated that the usual methods are greatly affected by outliers (c.f., Sections 6 and 7). The purpose of this paper is to solve the problem of how to apply model averaging when the data is contaminated. We proposed an outlier-robust Mallows-type model averaging approach. The idea is to use some specific GM-type loss functions, which are robust to outliers in both the covariates and the response, to obtain robust estimators and build robust weighting schemes. We proved that our proposed robust model averaging estimator is asymptotically optimal in the sense of minimizing the out-of-sample final prediction error. The rate of the RMMA-based empirical weight converging to the theoretically optimal weight is established. The robustness property of our proposed model averaging approach is also investigated by developing the concept of EPIF. Both simulation studies and real data analysis show that our proposed method is a useful tool to implement model averaging in the presence of outliers.

For the mean regression, the method proposed in this paper is the first optimal model averaging strategy with robustness for outliers, so there are many problems worthy of further study. Note that the jackknife model averaging of Hansen and Racine (2012), optimal model averaging of Liang et al. (2011), Kullback-Leibler model averaging of Zhang et al. (2015) and other weight selection criteria have been suggested, which may also be affected by outliers. How to correct these criteria in the case of the data with outliers is an interesting problem. Based on the method of random design matrix, one may extend our proposed criterion to the autoregressive sequence. In this paper, we limit the candidate models to linear forms. Exploring methods to remove this restriction presents an intriguing challenge. Further, both the dimensions and the number of candidate models in this paper are assumed to be fixed, and how to generalize the proposed new method to the situation of divergence is also a future research topic.

Acknowledgements

The authors would like to thank the Editor, Associate Editor, and reviewers for their valuable comments, which have significantly improved the quality of the paper. These authors contributed equally: Miaomiao Wang and Kang You. Wang's and Zou's work was supported by the National Natural Science Foundation of China (Grant Nos. 12401335, 11971323 and 12031016). Zou's work was also supported by the Beijing Outstanding Young Scientist Program (Grant No. JWZQ20240101027). You's work was partially supported by the Engineering and Physical Sciences Research Council of United Kingdom (Grant No. EP/X038297/1).

Supplementary Materials

The Supplementary Materials contain the assumptions, the robustness property of the model averaging estimator, proofs of theorems, two robust versions of Mallows' C_p and Mallows model averaging method for comparison, and additional simulation studies, which are provided in Sections S1-S9, respectively. The Supplementary Materials and all codes referenced in Sections 6 and 7 are available with this paper at the Biometrics website on Oxford Academic.

Data Availability

The human immunodeficiency virus data set referenced in Section 7 and that supports the findings in this paper is given in Hammer et al. (1996) and can be found in the `R` package `speff2trial`, which is openly available at https://cran.r-project.org/web/packages/speff2trial/index.html.

REFERENCES

Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics* **22,** 203–217.

Burman, P. and Nolan, D. (1995). A general Akaike-type criterion for model selection in robust regression. *Biometrika* **82,** 877–886.

Coakley, C. W. and Hettamansperger, T. P. (1993). A bounded influence, high breakdown, efficient regression estimator. *Journal of the American Statistical Association* **88,** 872–880.

Du, J., Zhang, Z. Z., and Xie, T. F. (2018). Model averaging for M-estimation. *Statistics* **52,** 1417–1432.

Guo, Y. F. and Li, Z. H. (2021). Outlier robust model averaging based on S_p criterion. *Stat* **10,** 1–10.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundaker, H., Schooley, R. T., Haubrich, R. H., et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine* **335,** 1081–1090.

Hampel, F. R. (1983). Some aspects of model choice in robust statistics. In *Proceedings of the 44th Session of the ISI, Madrid, Book 2*, pages 767–771.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust statistics: the approach based on influence functions*. Wiley, New York.

Han, P., Kong, L., Zhao, J., and Zhou, X. (2019). A general framework for quantile estimation with incomplete data. *Journal of the Royal Statistical Society B* **81,** 305–333.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica* **75,** 1175–1189.

Hansen, B. E. and Racine, J. S. (2012). Jackknife model averaging. *Journal of Econometrics* **167,** 38–46.

Hawkins, D. M. (1980). *Identification of outliers*. Chapman and Hall, London.

He, X. M., Simpson, D. G., and Wang, G. Y. (2000). Breakdown points of t-type regression estimators. *Biometrika* **87,** 675–687.

Hjort, N. L. and Claeskens, G. (2006). Focused information criteria and model averaging for the Cox hazard regression model. *Journal of the American Statistical Association* **101,** 1449–1464.

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

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Annals of Statistics* **1,** 799–821.

Liang, H., Zou, G. H., Wan, A. T. K., and Zhang, X. Y. (2011). Optimal weight choice for frequentist model average estimators. *Journal of the American Statistical Association* **106,** 1053–1066.

Liao, J., Zong, X. P., Zhang, X. Y., and Zou, G. H. (2019). Model averaging based on leave-subject-out cross-validation for vector autoregressions. *Journal of Econometrics* **209,** 35–60.

Lu, X. and Su, L. (2015). Jackknife model averaging for quantile regressions. *Journal of Econometrics* **188,** 40–58.

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

Müller, S. and Welsh, A. H. (2005). Outlier robust model selection in linear regression. *Journal of the American Statistical Association* **100,** 1297–1310.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7,** 186–199.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* **133,** 1155–1174.

Ronchetti, E. (1985). Robust model selection in regression. *Statistics and Probability Letters* **3,** 21–23.

Ronchetti, E. (1997). Robustness aspects of model choice. *Statistica Sinica* **7,** 327–338.

Ronchetti, E., Field, C., and Blanchard, W. (1997). Robust linear model selection by cross-validation. *Journal of the American Statistical Association* **92,** 1017–1023.

Ronchetti, E. and Staudte, R. G. (1994). A robust version of Mallows' C_p. *Journal of the American Statistical Association* **89,** 550–559.

Sommer, S. and Staudte, R. G. (1995). Robust variable selection in regression in the presence of outliers and leverage points. *Australian Journal of Statistics* **37,** 323–336.

Wang, H., Lu, Z., and Liu, Y. (2023). Score test for missing at random or not under logistic missingness models. *Biometrics* **79,** 1268–1279.

Wang, M., Zhang, X., Wan, A. T. K., You, K., and Zou, G. (2023). Jackknife model averaging for high-dimensional quantile regression. *Biometrics* **79,** 178–189.

Wang, X. Q., Jiang, Y. L., Huang, M., and Zhang, H. P. (2013). Robust variable selection with exponential squared loss. *Journal of the American Statistical Association* **108,** 632–643.

Wisnowski, J. W., Simpson, J. R., Montgomery, D. C., and Runger, G. C. (2003). Resampling methods for variable selection in robust regression. *Computational Statistics and Data Analysis* **43,** 341–355.

Yang, Y. H. (2001). Adaptive regression by mixing. *Journal of the American Statistical Association* **96,** 574–588.

Zhang, X. Y., Zou, G. H., and Carroll, R. (2015). Model averaging based on Kullback-Leibler distance. *Statistica Sinica* **25,** 1583–1598.

Received xx xx. *Revised xx* xx. *Accepted xx* xx.

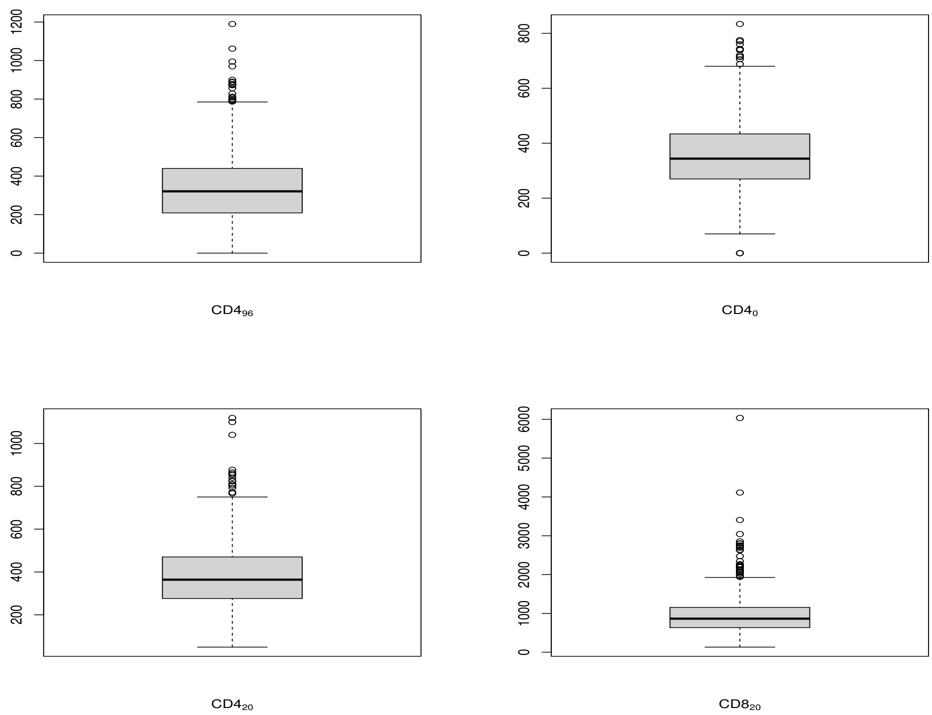

Figure 1. Boxplots of the response $\mathrm{CD4}_{96}$ and the covariates $\mathrm{CD4}_0$, $\mathrm{CD4}_{20}$ and $\mathrm{CD8}_{20}$

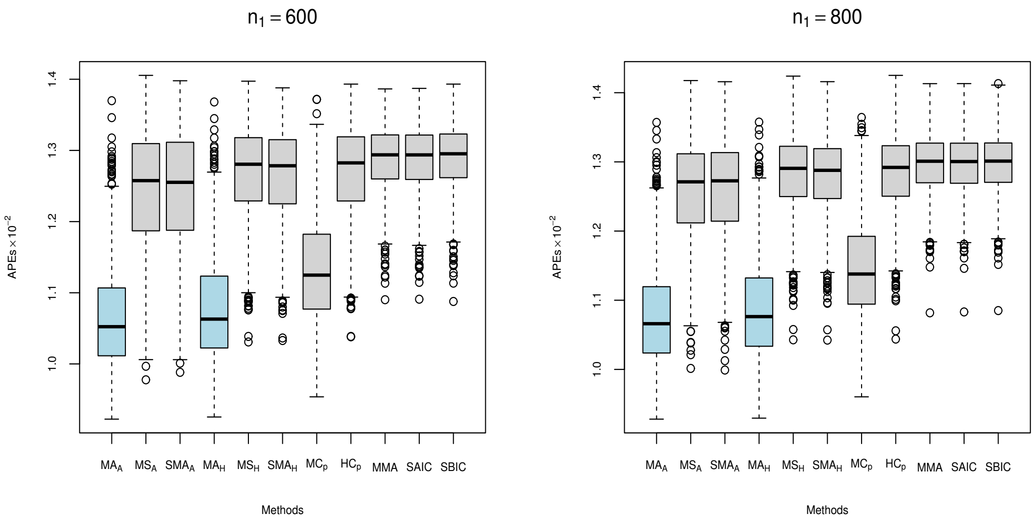

Figure 2. Boxplots for the APEs of estimators based on 1000 random divisions: HIV data

Table 1
MAEs *of estimators for Case 1*

n	R^2	MA_A	MS_A	SMA_A	MA_H	MS_H	SMA_H	MC_p	HC_p	MMA	SAIC	SBIC
50	0.1	0.282	0.334	0.418	0.265	0.315	0.273	0.348	0.353	0.234	0.210	**0.194**
	0.3	0.323	0.387	0.429	0.297	0.354	0.303	0.375	0.374	0.275	0.257	**0.252**
	0.5	0.355	0.429	0.448	0.322	0.381	0.327	0.399	0.398	0.305	**0.304**	0.311
	0.7	0.413	0.494	0.474	0.370	0.427	0.371	0.436	0.436	**0.359**	0.361	0.386
	0.9	0.518	0.583	0.542	0.462	0.500	0.462	0.516	0.513	**0.459**	0.472	0.513
150	0.1	0.166	0.180	0.234	0.159	0.180	0.163	0.212	0.215	0.149	0.136	**0.133**
	0.3	0.202	0.231	0.249	0.188	0.219	0.191	0.235	0.236	0.181	**0.177**	0.194
	0.5	0.233	0.266	0.266	0.215	0.246	0.217	0.259	0.258	**0.210**	0.211	0.237
	0.7	0.267	0.305	0.288	0.241	0.273	0.243	0.280	0.279	**0.237**	0.243	0.277
	0.9	0.339	0.380	0.345	0.303	0.328	0.303	0.336	0.335	**0.300**	0.313	0.355
400	0.1	0.115	0.128	0.149	0.108	0.122	0.110	0.145	0.147	0.104	**0.100**	0.108
	0.3	0.146	0.163	0.168	0.132	0.149	0.134	0.165	0.165	**0.130**	**0.130**	0.151
	0.5	0.165	0.185	0.182	0.149	0.168	0.150	0.178	0.177	**0.146**	0.150	0.176
	0.7	0.192	0.215	0.201	0.170	0.190	0.171	0.197	0.196	**0.168**	0.175	0.207
	0.9	0.239	0.264	0.239	0.213	0.228	0.212	0.233	0.232	**0.210**	0.220	0.262

Table 2
MAEs *of estimators for Case 2*

n	R^2	MA_A	MS_A	SMA_A	MA_H	MS_H	SMA_H	MC_p	HC_p	MMA	SAIC	SBIC
50	0.1	0.255	0.346	0.357	**0.236**	0.312	0.282	0.307	0.301	0.362	0.356	0.323
	0.3	0.302	0.407	0.413	**0.282**	0.377	0.351	0.335	0.343	0.740	0.748	0.674
	0.5	0.339	0.449	0.446	**0.322**	0.417	0.444	0.357	0.369	0.829	0.829	0.750
	0.7	0.408	0.537	0.545	**0.382**	0.511	0.507	0.408	0.441	1.474	1.525	1.356
	0.9	0.540	0.672	0.677	**0.511**	0.650	0.785	0.519	0.576	1.929	1.967	1.816
150	0.1	0.137	0.170	0.177	**0.123**	0.164	0.148	0.165	0.152	0.428	0.427	0.361
	0.3	0.162	0.204	0.203	**0.149**	0.196	0.184	0.182	0.178	0.674	0.677	0.595
	0.5	0.185	0.221	0.217	**0.169**	0.208	0.196	0.199	0.194	0.698	0.718	0.652
	0.7	0.216	0.275	0.271	**0.203**	0.269	0.264	0.226	0.237	1.094	1.121	0.984
	0.9	0.292	0.367	0.363	**0.279**	0.365	0.409	0.295	0.339	1.875	1.928	1.708
400	0.1	0.083	0.090	0.091	**0.073**	0.083	0.082	0.095	0.080	0.237	0.246	0.212
	0.3	0.097	0.104	0.102	**0.084**	0.095	0.091	0.103	0.090	0.375	0.386	0.365
	0.5	0.110	0.121	0.119	**0.097**	0.112	0.109	0.115	0.106	0.499	0.503	0.471
	0.7	0.128	0.141	0.138	**0.115**	0.133	0.134	0.130	0.126	0.703	0.723	0.663
	0.9	0.175	0.198	0.196	**0.167**	0.195	0.200	0.181	0.190	1.030	1.048	0.964

Table 3
MAEs *of estimators for Case 3*

n	R^2	MA_A	MS_A	SMA_A	MA_H	MS_H	SMA_H	MC_p	HC_p	MMA	SAIC	SBIC
50	0.1	**0.377**	0.414	0.721	0.502	0.518	0.827	0.488	0.485	5.124	4.642	4.254
	0.3	**0.396**	0.436	0.716	0.506	0.551	0.817	0.500	0.493	5.073	4.566	4.021
	0.5	**0.459**	0.525	0.726	0.553	0.610	0.843	0.564	0.563	4.545	3.941	3.522
	0.7	**0.548**	0.645	0.806	0.628	0.727	0.908	0.662	0.660	5.202	4.508	4.019
	0.9	**0.717**	0.837	0.879	0.761	0.898	1.004	0.816	0.817	5.057	4.559	4.159
150	0.1	**0.193**	0.201	0.300	0.249	0.276	0.393	0.280	0.278	3.401	2.871	2.549
	0.3	**0.244**	0.268	0.326	0.287	0.339	0.418	0.321	0.320	3.772	3.259	2.920
	0.5	**0.290**	0.326	0.352	0.318	0.380	0.440	0.364	0.363	3.907	3.284	2.992
	0.7	**0.339**	0.382	0.383	0.361	0.438	0.472	0.408	0.407	3.888	3.306	2.980
	0.9	**0.443**	0.492	0.461	0.450	0.545	0.545	0.506	0.505	3.748	3.314	3.051
400	0.1	**0.135**	0.145	0.174	0.165	0.185	0.253	0.194	0.195	3.065	2.539	2.237
	0.3	**0.177**	0.196	0.197	0.192	0.229	0.268	0.223	0.220	2.898	2.435	2.146
	0.5	**0.206**	0.227	0.220	0.219	0.264	0.290	0.253	0.252	3.146	2.649	2.396
	0.7	**0.240**	0.264	0.243	0.242	0.297	0.308	0.283	0.283	3.210	2.722	2.459
	0.9	0.307	0.335	**0.297**	0.304	0.376	0.363	0.348	0.348	3.469	2.904	2.688

Table 4
MAEs *of estimators for Case 4*

n	R^2	MA_A	MS_A	SMA_A	MA_H	MS_H	SMA_H	MC_p	HC_p	MMA	SAIC	SBIC
50	0.1	**0.394**	0.624	0.828	0.479	0.712	1.067	0.476	0.457	4.375	4.354	4.036
	0.3	**0.522**	0.759	0.891	0.611	0.848	1.139	0.532	0.523	5.032	4.849	4.515
	0.5	**0.468**	0.722	0.848	0.553	0.808	1.046	0.575	0.575	4.447	4.060	3.883
	0.7	**0.554**	0.991	1.113	0.776	1.076	1.369	0.623	0.647	4.675	4.588	4.375
	0.9	**0.779**	1.117	1.208	0.857	1.233	1.451	0.795	0.844	5.428	5.406	5.084
150	0.1	**0.172**	0.272	0.338	0.200	0.409	0.458	0.265	0.287	3.393	3.426	3.129
	0.3	**0.213**	0.317	0.354	0.238	0.403	0.438	0.300	0.282	3.288	3.327	3.030
	0.5	**0.244**	0.365	0.386	0.271	0.445	0.480	0.329	0.321	3.221	3.257	2.988
	0.7	**0.282**	0.407	0.419	0.305	0.487	0.519	0.362	0.359	3.538	3.563	3.219
	0.9	**0.379**	0.520	0.525	0.403	0.622	0.679	0.450	0.477	4.658	4.694	4.321
400	0.1	**0.107**	0.141	0.160	0.113	0.194	0.198	0.158	0.137	2.197	2.214	2.084
	0.3	**0.131**	0.181	0.190	0.137	0.225	0.253	0.182	0.162	2.150	2.178	2.048
	0.5	**0.149**	0.194	0.198	0.153	0.234	0.252	0.197	0.183	2.339	2.363	2.201
	0.7	**0.169**	0.216	0.217	0.174	0.257	0.270	0.212	0.205	2.411	2.439	2.299
	0.9	**0.225**	0.279	0.278	0.230	0.330	0.336	0.268	0.273	3.046	3.110	2.873