Statistica Sinica Preprint No: SS-2025-0132	
Title	Random Weighting Approximation of M-estimators with Increasing Dimensions of Parameter
Manuscript ID	SS-2025-0132
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0132
Complete List of Authors	Ruixing Ming, Chengyao Yu, Min Xiao and Zhanfeng Wang
Corresponding Authors	Zhanfeng Wang
E-mails	zfw@ustc.edu.cn
Notice: Accepted author version.	

RANDOM WEIGHTING APPROXIMATION OF M-ESTIMATORS WITH INCREASING DIMENSIONS OF PARAMETER

Ruixing Ming[†,1], Chengyao Yu[†,1], Min Xiao[1] and Zhanfeng Wang[*,2]

[1] *Zhejiang Gongshang University*

and [2] *University of Science and Technology of China*

Abstract:

The random weighting (RW) approach, recognized as a flexible alternative to the classical bootstrap method, has been widely employed for approximating the distribution of parameter estimates. Nevertheless, existing theoretical results for the RW approach primarily address scenarios where the parameter dimension remains fixed. In this paper, we investigate the RW method of M-estimators under general parametric models with increasing dimensions. We establish that the RW estimator has the same asymptotic distribution as that of the parameter M-estimate, which suggests that statistical inference regarding the parameter M-estimate can proceed without estimating nuisance parameters involved in its asymptotic distribution. Statistical properties of the RW estimator, such as the Bahadur representation and convergence rate, are also established. Furthermore,

[†]Co-first authors and contributed equally to this work.

[*]Corresponding author.

we illustrate the applicability of our theoretical findings through several concrete models, including linear regression, logistic regression, and spatial median estimation for multivariate data. Simulation studies and real data analysis demonstrate the superior performance of the proposed RW method.

Key words and phrases: Bahadur representation, increasing dimensions, M-estimator, random weighting.

1. Introduction

The random weighting (RW) approach, which originates from the seminal contributions of Rubin (1981); Lo (1987); Zheng (1987); Tu and Zheng (1987); Weng (1989), is a generic technique to estimate the sampling distribution of a given statistic, typically achieved by repeatedly assigning random weights to a function of the original data set. For numerous statistical models, the asymptotic distribution of an estimator or a test statistic is complicated because it usually involves nuisance parameters which are intractable to estimate. For instance, the asymptotic covariance matrix of the least absolute deviation (LAD) estimator (Powell, 1984) depends on the unknown error density. The primary function of the RW approach is to approximate the sampling distribution of parameter estimates or test statistics without necessitating the estimation of nuisance parameters. Compared to the classical bootstrap (Efron, 1979), the RW approach exhibits enhanced

flexibility and can be interpreted as a smoothed version of the bootstrap method, often yielding superior performance for specific statistical models (Shao and Tu, 2012). Our study goal is to investigate the RW method for M-estimators with increasing dimensions.

When the parameter dimension is fixed, extensive literature exists exploring the large-sample properties of the RW approach for different parametric models. For classical linear model, Rao and Zhao (1992) established the validity of RW approximation for the asymptotic distribution of M-estimators with probability tending to one. Under further mild conditions, Wu and Zhao (1999) demonstrated the validity of this approximation with probability one. In addition, Wu et al. (2007); Chen et al. (2008) employed the RW approach to test linear hypotheses within linear regression frameworks. In censored regression models, the RW approach has been utilized to approximate the distribution of the LAD estimator and construct test statistics (Fang and Zhao, 2006; Wang et al., 2009, 2018; Xiao et al., 2014). More recently, the RW approach has been used to achieve distributed inference in quantile regression contexts (Xiao et al., 2024). Additionally, the RW approach has gained widespread application within biomedical science (Wang et al., 2022, 2023; Han et al., 2024). Nonetheless, all aforementioned studies assume that the parameter dimension remains fixed. To the best

of our knowledge, no work has explored the RW method under scenarios where the parameter dimension increases with the sample size.

In practice, however, it often occurs that the dimension of collected covariates for studies becomes large or even grows with the sample size. In such high-dimensional settings, similar to the high-dimensional bootstrap method (Chernozhukov et al., 2023), the primary statistical challenge confronting the RW approach lies in the absence of explicit limiting distributions. Furthermore, existing theoretical results on the RW approach are built on some specific regression models, resulting in a notable gap in generality and applicability to broader parametric settings. To fill these gaps, we propose a unified framework for studying RW approximations of M-estimators applicable to general parametric models, explicitly accommodating cases with increasing parameter dimensions.

Our methodological and theoretical contributions are summarized as follows. First, we propose a RW method for M-estimates (He and Shao, 2000) where the dimension of the parameter increases with the sample size. Different from (He and Shao, 2000), the proposed RW method provides a flexible and efficient tool for conducting statistical inference of the parameter estimate, where we avoid estimating nuisance parameters involved in the asymptotic distribution of the M-estimator. Second, the bootstrap method

can be roughly regarded as a special case of the RW method with a weight vector following multinomial distribution. The RW method brings benefits, for example, continuous weight does not change the censoring proportion in Tobit regression (Powell, 1984), while discrete sampling in the bootstrap method may change the censoring proportion affecting the precision of statistical inference. If the censoring proportion is too high, the bootstrap method may not even be applicable (Cui et al., 2008). Third, we establish the Bahadur representation of the RW estimator and demonstrate that its Euclidean estimation error is of the same order as the ratio of parameter dimension to sample size. Most importantly, we demonstrate that the RW approach yields valid approximations to the distribution of each component or any linear combination of the M-estimator when the parameter dimension grows at a controllable rate. Fourth, we develop a RW approximation procedure to make statistical inferences such as interval estimation without estimating nuisance parameters. The applicability of our results is illustrated through several model examples. Specifically, the dimension of the parameter m_n is allowed to increase at the rate of $m_n^2 \log m_n = o(n)$ in models such as linear regression with smooth scores, logistic regression and spatial median for multivariate data without compromising the validity of the RW approximation. For linear regression with jump scores, a growth

rate of $m_n^3(\log m_n)^2 = o(n)$ is permissible.

The remainder of the paper is organized as follows. Section 2 illustrates how the RW approach can be employed to perform statistical inference, establishes the main theoretical results, and provides a detailed comparison between the RW and bootstrap methods. In Section 3, we apply the general results developed in Section 2 to several concrete models, including classical linear regression, logistic regression, and the spatial median for multivariate data. Section 4 evaluates the performance of our proposed methods on simulated data. Section 5 demonstrates the practical utility of our methods through an application to a real dataset. Section 6 outlines potential directions for future research on the high-dimensional RW approach. Proofs of the main results and simulation details are provided in the Supplementary Material.

2. Random Weighting Approximation

In this section, we propose a random weighting approach to conduct statistical inference for M-estimators of general parametric models with increasing dimensions. We first formulate the problem and specify the primary assumptions required. Subsequently, we introduce the random weighting algorithm along with corresponding theoretical results. Finally, the advan-

tages of the proposed methodology are demonstrated through a comparative analysis with the bootstrap approach.

2.1 Problem setup

Suppose that the samples $z_1, \ldots, z_n \in R^{p_n}$ are independent and generated from distribution $F_{i,\theta}$, $i = 1, \ldots, n$, with a common parameter $\theta \in R^{m_n}$. The dimensions p_n and m_n may increase with the sample size n. Consider an M-estimator $\hat{\theta}_n$ of θ, defined as the minimizer of the objection function $G_n(\theta) = \sum_{i=1}^{n} \rho(z_i, \theta)$ over $\theta \in R^{m_n}$ for some given function $\rho(z, \theta)$. The function $\rho(z, \theta)$ tends to ∞ as $||\theta|| \to \infty$ for each z and is differentiable on θ except at finitely many points, where $||\cdot||$ is defined as the Euclidean norm. Let θ_0 be the true value of θ, and denote the derivative of ρ by $\psi(z, \theta)$. At points where $\rho(z, \theta)$ is not differentiable in θ, we define $\psi(z, \theta)$ to be an element of the subdifferential set $\partial_\theta \rho(z, \theta)$, where

$$\partial_\theta \rho(z, \theta) = \left\{ g \in \mathbb{R}^{m_n} : \rho(z, \theta') \geq \rho(z, \theta) + g^\top (\theta' - \theta),\ \forall \theta' \in R^{m_n} \right\}.$$

The subdifferential $\partial_\theta \rho(z, \theta)$ is always non-empty due to the convexity of $\rho(z, \theta)$. For consistency of $\hat{\theta}_n$, the function ψ must satisfy

$$\sum_{i=1}^{n} E_{\theta_0}(\psi(z_i, \theta_0)) = 0.$$

For simplicity of notation, define

$$\eta_i(\tau,\theta) = \psi(z_i,\tau) - E(\psi(z_i,\tau)) - (\psi(z_i,\theta) - E(\psi(z_i,\theta))).$$

Before presenting our main results, we make several essential assumptions as follows.

(A1) The M-estimator $\hat{\theta}_n$ satisfies

$$\left\| \sum_{i=1}^{n} \psi(z_i, \hat{\theta}_n) \right\| = o_p(n^{1/2}).$$

(A2) There exists a constant c and $r \in (0,2]$ such that for $0 < d \leq 1$,

$$\max_{i \leq n} E_\theta \left(\sup_{\tau : ||\tau - \theta|| \leq d} ||\eta_i(\tau,\theta)||^2 \right) \leq n^c d^r.$$

(A3) The derivative function ψ satisfies $|| \sum_{i=1}^{n} \psi(z_i, \theta_0)|| = O_p((n m_n)^{1/2})$.

(A4) There exists a sequence of matrices D_n with $\liminf_{n\to\infty} \lambda_{\min}(D_n) > 0$ such that for any $B > 0$ and uniformly in $\alpha \in S_{m_n} = \{\alpha \in R^{m_n} : ||\alpha|| = 1\}$,

$$\sup_{||\theta - \theta_0|| \leq B(m_n/n)^{1/2}} \left| \alpha^\top \sum_{i=1}^{n} E_{\theta_0}(\psi(z_i,\theta) - \psi(z_i,\theta_0)) - n\alpha^\top D_n(\theta - \theta_0) \right| = o(n^{1/2}),$$

where $\lambda_{\min}$ denotes the smallest eigenvalue of a matrix.

(A5) $\sup_{\tau : ||\tau-\theta|| \leq B(m_n/n)^{1/2}} \sum_{i=1}^{n} E_\theta |\alpha^\top \eta_i(\tau,\theta)|^2 = O(A(n, m_n))$ for any $\theta \in R^{m_n}$, $\alpha \in S_{m_n}$ and $B > 0$.

(A6) $\sup_{\alpha \in S_{m_n}} \sup_{\tau : ||\tau-\theta|| \leq B(m_n/n)^{1/2}} \sum_{i=1}^{n} |\alpha^\top \eta_i(\tau,\theta)|^2 = O(A(n, m_n))$ for any $\theta \in R^{m_n}$ and $B > 0$.

Assumptions (A1)-(A6) are the same as assumptions (C0)-(C5) in He and Shao (2000). If the function $\rho(x, \theta)$ is convex over θ, then under assumptions (A1)-(A6) with $A(n, m_n) = o(n/\log n)$, He and Shao (2000) provided that $||\hat{\theta}_n - \theta_0|| = O_p(m_n/n)$, and with $A(n, m_n) = o(n/(m_n \log n))$, they demonstrated that for any consistent estimator $\hat{\theta}_n$, the Bahadur representation $\hat{\theta}_n - \theta_0 = -\sum_{i=1}^{n} D_n^{-1}\psi(z_i, \theta_0)/n + r_n$ holds with $||r_n|| = o_p(n^{-1/2})$.

However, the asymptotic distribution of $\alpha^\top(\hat{\theta}_n - \theta_0)$ for any $\alpha \in S_{m_n}$ involves unknown nuisance parameters, such as the matrix D_n and the variance of $\psi(z_i, \theta_0)$. Consequently, statistical inference on any linear combination of $\hat{\theta}_n$ requires the estimation of these nuisance parameters. Accurately estimating these nuisance parameters can be challenging, particularly when the sample size is limited. In this paper, we aim to avoid estimating these nuisance parameters and to directly make an approximation to the distribution of the linear combination of $\hat{\theta}_n$.

2.2 Methodology and theoretical results

We define a RW estimator θ_n^* of θ_0 by

$$\theta_n^* = \arg\min_{\theta \in R^{m_n}} G_n^*(\theta) = \arg\min_{\theta \in R^{m_n}} \sum_{i=1}^{n} w_i \rho(z_i, \theta), \tag{2.1}$$

where $\{w_i\}_{i=1}^n$ is a sequence of weight variables. For the weight variables, we need the following assumption,

(R1) $w_i, i = 1, \ldots, n$ are i.i.d. random variables with bounded and non-negative values, satisfying $E(w_i) = 1$ and $\mathrm{Var}(w_i) = \sigma^2 > 0$. The sequence $\{w_i\}_{i=1}^n$ is independent of the sequence $\{z_i\}_{i=1}^n$.

A high-level view of the RW method is presented in Algorithm 1. We approximate the distribution of $\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0)$ by using the conditional distribution of $\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma$ given the samples $z_1, \ldots, z_n$. Since the weight distribution is predetermined, we can generate weights $\{w_i\}$ accordingly and subsequently obtain a RW estimator by minimizing the objective function as in equation (2.1). Repeat this procedure for B times and obtain B random weighting estimators. Subsequently, the empirical distribution and the empirical variance of $\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma$ are computed and utilized to approximate their counterparts for $\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0)$. Therefore, we did not estimate the nuisance parameters and can infer the $\hat{\theta}_n$.

Remark 1. In assumption (R1), the expectation of the weight variable is equal to 1 as usual with random weighting studies. However, the variance of w_i relaxes to any positive constant σ^2, which is different from most of the usual studies where $\sigma^2 = 1$. It allows more flexible weight variable w_i. In addition, the boundedness of the weight variable is imposed as a technical condition, and we will see in our simulation that violating this boundedness constraint does not affect the performance of the RW approach.

Algorithm 1: Random weighting approximation procedure

Input: Sample $\{z_i\}_{i=1}^n$; objective function $\rho(z,\theta)$; $B \in Z^+$;

weight w_i with variance σ^2; projection vector α; $\beta_0 \in (0,1)$.

Compute the M-estimator by $\hat{\theta}_n = \arg\min_{\theta \in R^{m_n}} \sum_{i=1}^n \rho(z_i,\theta)$.

for $b = 1, \ldots, B$ **do**

Generating weights $w^b = (w_1^b, \ldots, w_n^b)$.

Compute RW estimator $\theta_n^{*b} = \arg\min_{\theta \in R^{m_n}} \sum_{i=1}^n w_i^b \rho(z_i,\theta)$.

end

Compute $\hat{\sigma}_\alpha^2 = \frac{1}{(B-1)\sigma^2} \sum_{b=1}^B (\alpha^\top \theta_n^{*b} - \frac{1}{B}\sum_{b=1}^B \alpha^\top \theta_n^{*b})^2$.

Compute

$$q_{\beta_0} = \inf\{t \in \{\alpha^\top \theta_n^{*1}/\sigma, \ldots, \alpha^\top \theta_n^{*B}/\sigma\} : F_B(t) \geq \beta_0\} - \alpha^\top \hat{\theta}_n/\sigma,$$

where $F_B(t) = \frac{1}{B}\sum_{b=1}^B 1(\alpha^\top \theta_n^{*b}/\sigma \leq t)$.

Output: the estimated variance of $\alpha^\top(\hat{\theta}_n - \theta_0)$, say, $\hat{\sigma}_\alpha^2$; the estimated β_0-th quantile of $\alpha^\top(\hat{\theta}_n - \theta_0)$, say, q_{β_0}.

We establish several theorems to demonstrate the validity of the proposed RW approximation. The following theorem states the convergence rate of the RW estimator θ_n^*.

Theorem 1. *Assume that the objection function $\rho(z,\theta)$ is convex over θ and its derivative with respect to θ is ψ, then under the assumptions (R1)*

and (A1)-(A6) with $A(n, m_n) = o(n/\log n)$*, we have*

$$||\theta_n^* - \theta_0|| = O_p(m_n/n).$$

The proof of Theorem 1 is presented in Subsection S2.1 of the Supplementary Material. Similar to He and Shao (2000), the error bound of (A4) can be relaxed to $o((nm_n)^{1/2})$ without affecting the validity of Theorem 1. This theorem indicates that the convergence rate of the RW estimator depends on the order of dimension m_n of the parameter, and the growth rate of m_n relative to n is dominated by $A(n, m_n)$. For the common linear regression model, $A(n, m_n) = \min\{(m_n n)^{1/2} + m_n^{3/2}\log n, (m_n n \log n)^{1/2}\}$ and m_n must satisfy $m_n^2(\log m_n) = o(n)$ for smooth ψ or $m_n^3(\log m_n)^2 = o(n)$ for non-smoothing ψ with jumps. Furthermore, if the term $A(n, m_n)$ satisfies $A(n, m_n) = o(n/(m_n \log n))$, a Bahadur representation of θ_n^* can be derived as established in the following theorem.

Theorem 2. *Assume that the assumptions (R1) and (A1)-(A6) are satisfied with* $A(n, m_n) = o(n/(m_n \log n))$*. Then for any consistent estimator* θ_n^**, we have*

$$\theta_n^* - \theta_0 = -\frac{1}{n}\sum_{i=1}^{n} w_i D_n^{-1}\psi(z_i, \theta_0) + r_n,$$

where $||r_n|| = o_p(n^{-1/2})$*. Especially, with all weight* $w_i = 1$ *we get the*

common result that

$$\hat{\theta}_n - \theta_0 = -\frac{1}{n}\sum_{i=1}^{n} D_n^{-1}\psi(z_i, \theta_0) + r_n.$$

The proof of Theorem 2 is presented in Subsection S2.2 of the Supplementary Material. Throughout, we denote by $\mathcal{L}^*$ and P^* the conditional distribution and conditional probability, respectively, given the observations $z_1, \ldots, z_n$. From Theorem 2, we have Theorem 3 under the following additional assumption.

(A7) The derivative ψ satisfies $E(\psi(z_i, \theta_0)) = 0$. The sequence of m_n by m_n matrices D_n in (A4) satisfies $\lambda_{\max}(D_n) = O(1)$, and there exists absolute constants $c_1 > 0$, $c_2 > 0$, and $\delta < 2$, such that

$$\frac{1}{n}\sum_{i=1}^{n} E|\alpha^\top \psi(z_i, \theta_0)|^2 \geq c_1, \quad \sum_{i=1}^{n} E|\alpha^\top \psi(z_i, \theta_0)|^4 \leq c_2 n^\delta$$

uniform in $\alpha \in S_{m_n}$ for n large enough.

Theorem 3. *Assume that the assumptions (R1) and (A1)-(A7) are satisfied with $A(n, m_n) = o(n/(m_n \log n))$. Then for any given vector $\alpha \in R^{m_n}$ with bounded L_2 norm and any consistent estimator θ_n^*, we have as $n \to \infty$,*

$$\mathcal{L}^*(\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/(\sigma\sigma_\alpha^*)) \to N(0,1) \quad \textit{in pr.},$$

$$\mathcal{L}(\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0)/\sigma_\alpha) \to N(0,1),$$

and $\sigma_\alpha^{*2} - \sigma_\alpha^2 \to 0$ *almost surely, where* $\sigma_\alpha^{*2} = n^{-1}\sum_{i=1}^n \left(\alpha^\top D_n^{-1}\psi(z_i,\theta_0)\right)^2$, $\sigma_\alpha^2 = E(\sigma_\alpha^{*2})$. *For the Kolmogorov–Smirnov distance between* $\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma$ *and* $\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0)$, *we have as* $n \to \infty$,

$$\sup_u |P^*(\sqrt{n}\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma \le u) - P(\sqrt{n}\alpha^\top(\hat{\theta}_n - \theta_0) \le u)| \xrightarrow{p} 0. \qquad (2.2)$$

Remark 2. If condition $E|\alpha^\top\psi(z_n,\theta_0)|^4 \le c_2 n^\delta$ in (A7) holds with $\delta \le 1$, then it implies (A3). The boundedness of $\lambda_{\max}(D_n)$ and $E|\alpha^\top\psi(z_n,\theta_0)|^2$ is just technical to guarantee that the variance of the dominant term of $\sqrt{n}\alpha^T(\hat{\theta}_n - \theta_0)$ is bounded away from zero uniformly for $\alpha \in S_{m_n}$. As demonstrated by the subsequent examples in Section 3, these assumptions can be readily verified under certain classical regularity conditions.

The proof of Theorem 3 is presented in Subsection S2.3 of the Supplementary Material. From Theorems 2 and 3, we see that the limiting distribution of $\alpha^\top(\hat{\theta}_n - \theta_0)$ is the same as the asymptotic conditional distribution of $\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma$ for given observations $\{z_i\}_{i=1}^n$. Hence, the conditional distribution of $\alpha^\top(\theta_n^* - \hat{\theta}_n)/\sigma$ can be used as an approximation of the distribution of $\alpha^\top(\hat{\theta}_n - \theta_0)$, which demonstrates the validity of our proposed method.

2.3 Comparison with the bootstrap method

In this subsection, we provide some advantages of the proposed RW approach compared to the bootstrap method. The following proposition shows that the bootstrap method can be regarded as a special case of the RW method with weight vector following multinomial distribution.

Proposition 1 (Bootstrap as a special case of RW approach)**.** *Let the size of the resampled data set be n. The classical bootstrap estimator θ_n^B has the same form of the RW estimator, say,*

$$\theta_n^B = \arg\min_{\theta \in R^{m_n}} \sum_{i=1}^{n} w_i \rho(z_i, \theta),$$

where $w = (w_1, \ldots, w_n)$ follows a Multinomial$(n; 1/n, \ldots, 1/n)$ distribution. Furthermore, w_i is asymptotically Poisson(1) distributed as $n \to \infty$.

The proof is straightforward. According to Proposition 1, the proposed RW method is more flexible and smooth compared to the bootstrap method since any weights satisfying assumption (R1) are allowed. This brings benefits in many specific models (Shao and Tu, 2012). We provide several examples to illustrate it as follows.

Example 1 (Benefits of continuous weight)**.** In the censored regression (Powell, 1984), the bootstrap method may fail to perform reliably under heavy censoring (Cui et al., 2008). This failure arises from the discrete nature of

resampling, which leads to repeated selection of the same censored observations. This issue can be addressed by adopting a continuous weight scheme for the RW approach, where no sample points can be missed or selected multiple times.

Example 2 (Benefits of optional weights under finite samples)**.** For the least squares in the linear models, El Karoui and Purdom (2018) demonstrated that the bootstrap method overestimates the variance of a given statistic if $m_n/n \to \gamma \in (0,1)$. Under $m_n = o(n)$ throughout the paper, the bootstrap method may also be untrustworthy with finite samples. This limitation is caused by the multinomial sampling distribution. However, the RW method allows flexibility in distributions of w_i. By selecting appropriate weights, the RW approach can achieve better performance.

To show the phenomena of Example 2, we present the following proposition, which is adapted from Theorem 2 of El Karoui and Purdom (2018).

Proposition 2 (Variance approximation)**.** *Identify* $z_i^\top = (x_i^\top, y_i)$ *and suppose* $y_i = x_i^\top \theta_0 + \epsilon_i$ *for* $i \in \{1, \ldots, n\}$*, where* $\epsilon_i's$ *being i.i.d. with* $E(\epsilon_i) = 0$*,* $\text{var}(\epsilon_i) = \sigma_\epsilon^2$*. Consider normal design of* x_i*, say,* $x_i's$ *are i.i.d* $N(0, \Sigma)$ *with a positive-definite matrix* Σ*. Let* θ_n^* *be defined by equation (2.1) with* $\rho(z_i, \theta) = (y_i - x_i^\top \theta)^2$*, where the random weights* $\{w_i\}_{i=1}^n$ *satisfy assumption*

(R1) and $w_i > \eta > 0$. *For any* $\alpha \in S_{m_n}$, *if* $\gamma_n = m_n/n \in (0,1)$, *we have*

$$m_n \frac{\mathrm{var}(\alpha^\top \hat{\theta}_n)}{\alpha^\top \Sigma^{-1} \alpha} = \sigma_\epsilon^2 \frac{\gamma_n}{1-\gamma_n-1/n},$$

$$m_n \frac{E\left(\mathrm{var}\left(\alpha^\top \theta_n^*/\sigma | \{z_i\}_{i=1}^n\right)\right)}{\alpha^\top \Sigma^{-1} \alpha} = \frac{\sigma_\epsilon^2}{\sigma^2}\left[\frac{\gamma_n}{1-\gamma_n-f(\gamma_n)} - \frac{1}{1-\gamma_n}\right] + o(1),$$

where $f(\gamma_n) = E\left(1/(1+bw_i)^2\right)$ *and* b *is the unique solution of* $E\left(1/(1+bw_i)\right) = 1-\gamma_n$.

According to Proposition 1, the bootstrap variance estimate of $\alpha^\top \hat{\theta}_n$ is roughly equal to the conditional variance of $\alpha^\top \theta_n^*$ with $w_i \sim \mathrm{Poisson}(1)$. Let $\sigma = 1$ in Proposition 2. Then the variance estimation makes sense only when $E\left(1/(1+bw_i)^2\right) \approx (1-m_n/n)/(1+m_n/n)$, which is not satisfied for the Poisson(1) weight with relatively large m_n/n.

Choices of random weights The random weighting method computes $\mathrm{var}(\alpha^\top \theta_n^*/\sigma | \{z_i\}_{i=1}^n)$ to estimate $\mathrm{var}(\alpha^\top \hat{\theta}_n)$. This paper considers $m_n = o(n)$. From Theorem 3, when n is large (small m_n/n), any weight distribution defined in (R1) is satisfactory. When m_n/n is relatively large, for the least squares in the linear model, we suggest the weights satisfying (R1) and

$$E\left(\frac{1}{(1+bw_i)^2}\right) = \frac{(1-\sigma^2)\gamma_n^2 + (\sigma^2-2)\gamma_n + 1}{1+\gamma_n\sigma^2}, \tag{2.3}$$

according to Proposition 2 (the derivation seen in Subsection S2.5 of the

Supplementary Material). Such weights can be found numerically. Generally, we recommend the exponential weights with parameter 1 (i.e. exp(1)) from numerical studies.

3. Specializations for Different Statistical Models

We apply the general results to several specific statistical models, including linear regression, logistic regression, and the spatial median for multivariate data. To illustrate applications of Theorem 1, Theorem 2, and Theorem 3, we do not seek the weakest conditions for each model at the cost of clarity. The proofs of all corollaries are provided in Section S3 of the Supplementary Material.

3.1 Linear model

Suppose the linear regression model

$$y_i = x_i^\top \theta_0 + \epsilon_i, \tag{3.4}$$

with independent error ϵ_i having a common density f. The M-estimator $\hat{\theta}_n$ minimizes $\sum_{i=1}^n \rho(y_i - x_i^\top \theta)$ over $\theta \in R^{m_n}$ for some convex loss function ρ with minimum at $\rho(0) = 0$. Let $\phi(z) = \rho'(z)$ and identify $z_i^\top = (x_i^\top, y_i)$. The derivative of $\rho(z, \theta)$ is $\psi(z, \theta) = -\phi(y - x^\top \theta)x$. Without loss of generality, assume that $E(\psi(\epsilon_i)) = 0$ since it can be achieved by an adjustment of

the intercept. Consider $||\sum_{i=1}^{n}\psi(z_i,\hat{\theta}_n)|| = O_p(\delta_n)$ for $\delta_n = 0$ or $\delta_n = m_n^{3/2}\log n$, and two types of scores: (1) the smooth score ϕ such that ϕ is Lipschitz and $\delta_n = 0$; (2) the jump score ϕ such that $\delta_n = m_n^{3/2}\log n$ and ϕ has finite jump discontinuities but is Lipschitz in each interval between two jumps. For example, both the Huber loss (Huber, 1992) $\rho(z) = z^2 I\{|z| \leq \tau\}/2 + (\tau|z| - \tau^2/2)I\{|z| > \tau\}$ and the square loss $\rho(z) = z^2$ have smooth scores, where τ is called the robustification parameter. Under appropriate design conditions, the absolute loss $\rho(z) = |z|$ has a jump score. Consider either the random design points z_i (independent of ϵ_i) or fixed (z_i having a point mass). We assume conditions as follows.

(B1) $\liminf_{n\to\infty}\lambda_{\min}(n^{-1}\sum_{i=1}^{n}x_i x_i^T) > 0$.

(B2) $E|\phi(\epsilon_i)|^4 \in (0,\infty)$, ϕ' and ϕ'' are bounded by $c_0 = E(\phi'(\epsilon_i)) \in (0,\infty)$ for smooth scores, or ϕ, f and f' are bounded with $c_0 = -\int_{-\infty}^{+\infty}\phi(r)f'(r)dr \in (0,\infty)$ for jump scores.

(B3) $\max_{i\leq n}||x_i||^2 = O(m_n)$ and $\sup_{\beta,\gamma\in S_m}\sum_{i=1}^{n}|x_i^\top\beta|^2|x_i^\top\gamma|^2 = O(n)$.

Conditions (B1)-(B3) are essentially the same as those in He and Shao (2000) and are mild enough to encompass most situations. For (B2), the condition $E|\phi(\epsilon_i)|^4 < \infty$ depends both on the score ϕ and the error density f. For example, $E|\phi(\epsilon_i)|^4 < \infty$ always holds for Huber loss no matter the density f; for the least squares loss, this holds true if f has a finite

fourth moment. Note that if the independent errors ϵ_i are not identically distributed, $E|\phi(\epsilon_n)|^4 < \infty$ can be weakened to allow $E|\phi(\epsilon_n)|^4$ to diverge as $n \to \infty$. For jump scores, (B2) holds if, for example, $\rho(z) = |z|$. Condition (B3) is almost surely true if x_i is a random sample from a m_n-dimensional distribution such that $E|\alpha^\top x_n|^4$ is uniformly bounded for $\alpha \in S_{m_n}$ and for all n. The following corollary demonstrates the validity of the RW approach in approximating the distribution of any projection of $\hat{\theta}_n$.

Corollary 1. *Assume that the conditions (B1)-(B3) and (R1) are satisfied. If* $m_n(\log m_n)^3/n \to 0$, *we have*

$$||\theta_n^* - \theta_0||^2 = O_p(m_n/n)$$

for both smooth and jump scores. If $m_n^2 \log m_n/n \to 0$ *for smooth scores or* $m_n^3(\log m_n)^2/n \to 0$ *for jump scores, then equation (2.2) holds for any* $\alpha \in R^{m_n}$ *with a bounded* L_2 *norm.*

3.2 Logistic regression

Suppose a binary logistic regression model that

$$P(y = 1|x) = \frac{e^{x^\top \theta_0}}{1 + e^{x^\top \theta_0}}, \tag{3.5}$$

where x is a covariate of dimension m_n. Given a random sample $\{z_i^\top\}_{i\in[n]} = \{(x_i^\top, y_i)\}_{i\in[n]}$, the log-likelihood is

$$\log L(\theta) = \sum_{i=1}^{n} \left(y_i x_i^\top \theta - \log(1 + e^{x_i^\top \theta}) \right).$$

The objective function $\rho(z, \theta) = \log(1 + e^{x^\top\theta}) - yx^\top\theta$ is convex over θ. Consider the fixed design x. Assume the design conditions as follows.

(C1) $\sum_{i=1}^n ||x_i||^2 = O(nm_n)$, and $\sup_{\alpha,\beta\in S_{m_n}} \sum_{i=1}^n |\alpha^\top x_i|^2 |\beta^\top x_i|^2 = O(n)$.

(C2) $\liminf_{n\to\infty} \lambda_{\min}(D_n) > 0$, where $D_n = n^{-1} \sum_{i=1}^n (e^{x_i^\top\theta_0}/(1+e^{x_i^\top\theta_0}))x_i x_i^\top$.

Condition (C1) is slightly weaker than condition (B3). Condition (C2) imposes restrictions on the eigenvalues of the matrix, mainly to satisfy (A4) and (A7).

Corollary 2. *Assume that the design conditions (C1), (C2) and (R1) are satisfied. If $m_n \log m_n/n \to 0$, then we have*

$$||\theta_n^* - \theta_0||^2 = O_p(m_n/n).$$

If $m_n^2 \log m_n/n \to 0$, then equation (2.2) holds for any $\alpha \in R^{m_n}$ with a bounded L_2 norm.

3.3 Spatial median for multivariate data

We aim to estimate the multivariate location parameter θ_0 by minimizing $\sum_{i=1}^n ||z_i - \theta||$ over $\theta \in R^{m_n}$, where the sample $z_1, \ldots, z_n$ is considered

random and

$$\theta_0 = \arg\min_{\theta \in R^{m_n}} E\|z_n - \theta\|. \tag{3.6}$$

Suppose that the underlying distribution for z_i has a continuous density with respect to the Lebesgue measure. The loss function $\rho(z, \theta) = ||z - \theta||$ is convex over θ and its derivative is $\psi(z, \theta) = -(z - \theta)/||z - \theta||$. Consider conditions as follows.

(D1) $E_{\theta_0}(1/||z - \theta||^2) = O(1)$ as $m_n \to \infty$ if $||\theta - \theta_0|| \leq c$ for some $c > 0$.

(D2) $\liminf_{m_n \to \infty} \inf_{\alpha \in S_{m_n}} E_{\theta_0}(||z - \theta_0||^{-1} - |\alpha^\top (z - \theta_0)|^2/||z - \theta_0||^3) > 0$.

(D3) $\liminf_{m_n \to \infty} \inf_{\alpha \in S_{m_n}} E_{\theta_0}(|\alpha^\top (z - \theta_0)|^2/||z - \theta_0||^2) > 0$.

Corollary 3. *Let $\theta_0 \in R^{m_n}$ be the unique minimizing point of $E||z - \theta||$. If condition (R1) holds and $m_n(\log m_n)^2/n \to 0$, then the spatial median satisfies*

$$||\theta_n^* - \theta_0||^2 = O_p(m_n/n).$$

Furthermore, if conditions (D1)-(D3) are satisfied and $m_n^2 \log m_n/n \to 0$, then equation (2.2) holds for any $\alpha \in R^{m_n}$ with a bounded L_2 norm.

4. Simulation Studies

We conducted numerical simulation studies to further investigate the performance of RW estimators with increasing m_n. The pair bootstrap and

the residual bootstrap are chosen as the baseline methods. The proposed methods under the linear regression model (3.4) and the logistic regression model (3.5) are studied, while the spatial median estimation (3.6) is investigated in Subsection S4.2 of the Supplementary Material. Let $\hat{\sigma}_j^2$, cover_j, and width_j represent the estimation of $\text{var}(\hat{\theta}_{nj})$, the empirical coverage probability of the confidence interval for θ_{0j}, and the width of the confidence interval, respectively. Methods are evaluated in three aspects, i.e., the mean absolute standard deviation error (MASDE), the mean absolute coverage error (MACE) with coverage level $1-\beta_0 \in (0,1)$, and the mean confidence interval width (MCIW), where

$$\text{MASDE} = \frac{1}{m_n}\sum_{j=1}^{m_n}\left|\sqrt{n\hat{\sigma}_j^2} - \sqrt{n\text{var}(\hat{\theta}_{nj})}\right|,$$

$$\text{MACE} = \frac{1}{m_n}\sum_{j=1}^{m_n}\left|\text{cover}_j - (1-\beta_0)\right|, \quad \text{MCIW} = \frac{1}{m_n}\sum_{j=1}^{m_n}\text{width}_j,$$

and the variance $\text{var}(\hat{\theta}_{nj})$ is estimated by the sample variance computed from 1000 independent replications of $\hat{\theta}_{nj}$. The confidence interval is constructed via the reversed percentile method; details see Subsection S4.1 of the Supplementary Material. All results in tables are averages computed over 1000 independent simulation trials.

4.1 Simulation for linear regression

We consider four types of dimensions, $(n, m_n) = (100, 19)$, $(200, 24)$, $(500, 32)$, and $(1000, 39)$, where m_n satisfies $m_n = \lfloor 5n^{0.3} \rfloor$ and $\lfloor \cdot \rfloor$ is the floor function. Let $\theta_0 = (\theta_{01}, \ldots, \theta_{0m_n})^\top$, where $\theta_{0j} = 3 - 3(j-1)/m_n$. The design matrix $\mathbf{X} = (x_1, \ldots, x_n)^\top$ is constructed by independently sampling each row from the multivariate normal distribution $N(\mathbf{0}, I_{m_n})$. Different designs of the error item ϵ, the double exponential distribution DE(1), the standard normal distribution $N(0, 2)$, and the mixture normal distribution $0.6N(0, 0.6) + 0.4N(0, 2)$, are also considered.

Consider the least absolute estimation, say $\rho(z, \theta) = |y - x^\top \theta|$. For the RW method, the weight variable w_i is sampled from four distributions,

(1) Exponential distribution exp(1) (Exp);

(2) Gamma distribution $10/9\Gamma(0.9, 1)$ (Gamma);

(3) $2.5\Gamma(0.4, 1)$ (Gamma2);

(4) $1/3\Gamma(3, 1)$ (Gamma3).

The Gamma2 and Gamma3 are introduced specifically to investigate the sensitivity of the RW approach to weights with heavy tails and light tails, respectively. For the bootstrap methods, each resampled dataset has the

Table 1: MASDE in the linear regression with different error distributions.

m_n	Error	Average SD	Exp	Gamma	Gamma2	Gamma3	Pair	Residual
19	Normal	2.790	0.391	0.360	0.245	0.720	0.662	0.680
	Mixture	1.312	0.397	0.399	0.499	0.424	0.552	0.232
	Double exp	1.510	0.392	0.387	0.442	0.469	0.562	0.287
24	Normal	2.676	0.352	0.324	0.200	0.629	0.491	0.546
	Mixture	1.195	0.281	0.281	0.362	0.325	0.350	0.187
	Double exp	1.520	0.312	0.309	0.351	0.370	0.392	0.214
32	Normal	2.584	0.277	0.262	0.167	0.479	0.341	0.414
	Mixture	1.126	0.156	0.156	0.188	0.209	0.186	0.161
	Double exp	1.238	0.231	0.232	0.272	0.272	0.265	0.135
39	Normal	2.556	0.214	0.203	0.131	0.376	0.252	0.340
	Mixture	1.086	0.122	0.121	0.128	0.175	0.139	0.123
	Double exp	1.118	0.176	0.177	0.211	0.206	0.195	0.104

same sample size as the original dataset. Both the number of random weighting and bootstrap resampling are taken as 1000. The performance of the RW methods on estimating the standard deviation is summarized in Table 1, where "Average SD" refers to $\sum_{i=1}^{m_n} \sqrt{n\text{var}(\hat{\theta}_{nj})}/m_n$. The performance in interval estimation (empirical coverage probability and the width of interval) is presented in Tables 2 - 4.

It follows from Tabel 1 that, compared to the pair-bootstrap method,

Table 2: MACE and MCIW in the linear regression model with normal distribution error ϵ.

		MACE						MCIW					
β_0	m_n	Exp	Gamma	Gamma2	Gamma3	Pair	Residual	Exp	Gamma	Gamma2	Gamma3	Pair	Residual
0.05	19	0.045	0.046	0.038	0.048	0.022	0.090	1.245	1.234	1.194	1.368	1.353	0.838
	24	0.037	0.037	0.034	0.033	0.024	0.070	0.837	0.829	0.797	0.911	0.875	0.593
	32	0.031	0.030	0.030	0.026	0.024	0.053	0.499	0.497	0.481	0.534	0.511	0.380
	39	0.025	0.024	0.025	0.020	0.021	0.041	0.342	0.341	0.332	0.362	0.347	0.274
0.1	19	0.060	0.059	0.050	0.055	0.029	0.120	1.040	1.029	0.990	1.147	1.128	0.692
	24	0.047	0.049	0.048	0.039	0.030	0.097	0.701	0.695	0.666	0.765	0.733	0.494
	32	0.037	0.039	0.038	0.029	0.030	0.069	0.419	0.417	0.403	0.449	0.429	0.418
	39	0.031	0.031	0.032	0.022	0.026	0.056	0.287	0.286	0.279	0.304	0.291	0.230

Table 3: MACE and MCIW in the linear regression model with mixture normal distribution error ϵ.

		MACE						MCIW					
β_0	m_n	Exp	Gamma	Gamma2	Gamma3	Pair	Residual	Exp	Gamma	Gamma2	Gamma3	Pair	Residual
0.05	19	0.007	0.008	0.026	0.018	0.019	0.058	1.245	1.234	1.194	1.368	1.353	0.838
	24	0.005	0.003	0.021	0.019	0.008	0.050	0.410	0.410	0.435	0.420	0.429	0.282
	32	0.013	0.013	0.007	0.021	0.009	0.045	0.224	0.225	0.231	0.233	0.229	0.169
	39	0.013	0.012	0.006	0.014	0.011	0.034	0.149	0.149	0.150	0.156	0.151	0.119
0.1	19	0.010	0.012	0.041	0.022	0.033	0.078	1.040	1.029	0.990	1.147	1.128	0.692
	24	0.005	0.005	0.030	0.018	0.012	0.070	0.341	0.341	0.359	0.352	0.358	0.233
	32	0.019	0.016	0.009	0.026	0.012	0.064	0.419	0.417	0.403	0.449	0.429	0.318
	39	0.015	0.014	0.006	0.016	0.011	0.045	0.125	0.125	0.126	0.131	0.127	0.100

almost RW-based approaches have smaller MASDEs, which shows that RW methods have better estimations of variance of the M estimators. This difference is more pronounced when the sample size is small (m_n/n being

Table 4: MACE and MCIW in the linear regression model with double exponential distribution error ϵ.

		MACE						MCIW					
β_0	m_n	Exp	Gamma	Gamma2	Gamma3	Pair	Residual	Exp	Gamma	Gamma2	Gamma3	Pair	Residual
0.05	19	0.007	0.008	0.014	0.021	0.014	0.057	0.748	0.747	0.774	0.773	0.815	0.487
	24	0.006	0.005	0.014	0.014	0.008	0.050	0.469	0.468	0.482	0.483	0.491	0.325
	32	0.006	0.005	0.015	0.011	0.006	0.033	0.257	0.257	0.265	0.264	0.263	0.194
	39	0.005	0.005	0.013	0.008	0.006	0.028	0.168	0.168	0.172	0.171	0.170	0.133
0.1	19	0.010	0.009	0.021	0.027	0.021	0.088	0.620	0.618	0.634	0.646	0.676	0.401
	24	0.007	0.007	0.022	0.015	0.014	0.068	0.391	0.390	0.399	0.405	0.410	0.270
	32	0.009	0.010	0.024	0.011	0.012	0.047	0.215	0.215	0.221	0.221	0.220	0.162
	39	0.006	0.008	0.020	0.010	0.009	0.037	0.141	0.141	0.144	0.144	0.143	0.112

large). Although the residual bootstrap has the best approximations to the true standard deviations under mixture normal and double exponential errors, it consistently underestimating the standard deviation, thereby making statistical inference highly unreliable as demonstrated in Tables 2 - 4. Specifically, the MACE levels of the residual bootstrap under different settings are always the highest, especially when the sample size is small. For example, for the double exponential distribution error with the nominal level $\beta_0 = 0.1$ and $m_n = 19$, the residual bootstrap has the empirical coverage probability of 0.812, which is less than 0.9. This indicates that the confidence interval from the residual bootstrap deviates away from nominal coverage. In contrast, compared to the residual bootstrap and the

pair bootstrap, almost all RW methods perform well in interval estimation, achieving empirical coverage probabilities that are closer to $1-\beta_0$. For example, from Table 4 with $m_n = 19, \beta_0 = 0.1$, the MACE and the MCIW of the pair bootstrap are 0.021 and 0.676, respectively. However, the RW method with Gamma weights simultaneously achieves a smaller MACE (0.009) and a smaller MCIW (0.618), demonstrating its superiority.

4.2 Simulation for logistic regression

Consider the true model $y \in \{0,1\}$ with $P(y=1|x) = \exp(x^\top\theta_0)/(1+\exp(x^\top\theta_0))$, where $x \sim N(\mathbf{0}, 0.1I_{m_n})$. We consider the order $m_n = \lfloor 1.5n^{0.4} \rfloor$ and choose $(n, m_n) = (500, 18), (1000, 23), (1500, 27)$. Similar to Lam and Liu (2023), let

$$\theta_0 = \left(\underbrace{a, \ldots, a}_{\lfloor m_n/2 \rfloor}, \underbrace{-a, \ldots, -a}_{m_n - \lfloor m_n/2 \rfloor} \right)^\top, \quad a = \sqrt{\frac{30}{m_n}}.$$

This parameter configuration is designed specifically to maintain $\mathrm{Var}(x^\top\theta_0) = 3$ regardless of the growth in dimension m_n. This ensures that $P(y=1|x)$ is not equal to 0 or 1 in most cases. Consider the maximum likelihood estimation, say $\rho(z,\theta) = \log(1+e^{x^\top\theta}) - yx^\top\theta$. The methods used here include those previously considered for linear regression; however, the residual bootstrap is not applicable for logistic regression.

Table 5: Results of MASDE in the logistic regression.

m_n	Average SD	Exp	Gamma	Gamma2	Gamma3	Pair
18	8.768	0.165	0.152	0.306	0.274	0.409
23	8.351	0.197	0.199	0.269	0.203	0.325
27	8.244	0.194	0.195	0.215	0.208	0.250

Table 6: MACE and MCIW in the logistic regression.

		MACE					MCIW				
β_0	m_n	Exp	Gamma	Gamma2	Gamma3	Pair	Exp	Gamma	Gamma2	Gamma3	Pair
0.05	18	0.010	0.011	0.020	0.006	0.022	1.516	1.522	1.584	1.485	1.605
	23	0.008	0.008	0.012	0.008	0.013	1.033	1.034	1.058	1.019	1.069
	27	0.007	0.007	0.010	0.006	0.010	0.831	0.833	0.847	0.824	0.853
0.1	18	0.015	0.016	0.032	0.007	0.032	1.273	1.277	1.326	1.247	1.343
	23	0.012	0.013	0.021	0.011	0.021	0.868	0.869	0.888	0.857	0.897
	27	0.011	0.010	0.016	0.008	0.018	0.698	0.700	0.711	0.692	0.716

Table 5 shows that all four RW approaches yield more accurate estimation of the standard deviation of $\sqrt{n}\alpha^T(\hat{\theta}_n - \theta_0)$ compared with the bootstrap method, achieving smaller MASDE. Especially, when the sample size is small ($m_n = 18$), the MASDE obtained by RW methods with Exp and Gamma weights are much smaller than those by the bootstrap. Table 6 shows that under different dimensions and nominal coverage levels, the RW methods also achieve smaller MACEs and MCIWs compared to the

bootstrap method. That is, the RW methods can achieve coverage closer to nominal coverage $1-\beta_0$ and obtain more accurate confidence intervals. This advantage becomes even more pronounced under small sample conditions.

In conclusion, from numerical simulations for above parametric models, the RW method can effectively estimate the variance of a given statistic and perform interval estimation on coefficients. For practical implementation, we recommend employing exponential weights with parameter 1 (i.e., exp(1)) in the absence of further information. Compared with the bootstrap method, the advantage of the random weighting method is particularly evident in small sample sizes.

5. Real Data Analysis

The proposed RW method is applied to analyze the diabetes data set to identify factors associated with type II diabetes. To assess its performance, we compare the RW approach with the pair bootstrap method. The data set comes from the CDC's BRFSS 2015, available at `https://archive.ics.uci.edu/dataset/891/cdc+diabetes+health+indicators`. The response variable is diabetes status, coded as 0 for prediabetes or absence of diabetes, and 1 for confirmed diabetes diagnosis. The data set contains 70,692 observations and includes 21 covariates, including demographic char-

acteristics and various health-related indicators. We delete a binary variable named "CholCheck" with an almost constant value of 1 (proportion greater than 97.5%). In this paper, the logistic regression with intercept term is employed. Two evaluation criteria are used to compare the RW and pair bootstrap methods: (1) approximation of the variance of the logistic maximum likelihood estimator; and (2) confidence interval estimates for the log odds ratios.

We take the logistic maximum likelihood estimator derived from the full balanced dataset as a proxy for the true model coefficients. To evaluate our proposed method, we generate 50 disjoint subsamples each of size $n = 1000$, and compare the estimated variance and the coverage proportions of the confidence intervals produced by the RW method and the pair bootstrap. The random weights follow the distribution exp(1).

Table 7 reports the average estimated standard deviations and the empirical standard deviations calculated from the 50 sub-samples. By calculations, we have MASDE = 0.550, 0.605 for the RW method and the bootstrap method, respectively. It shows that the RW method provides more accurate estimates of the standard deviation for the majority of coefficients.

We further employ the RW method and the pair bootstrap method to

Table 7: Standard deviation of $\sqrt{n}(\hat{\theta}_{nj} - \theta_{0j})$ for $j = 0, 1, \cdots, 20$.

	0 (Intercept)	1	2	3	4	5	6
True	27.376	5.460	4.409	0.489	5.619	13.213	7.848
Exp	28.122	5.455	5.213	0.499	5.213	12.347	8.213
Pair	29.171	5.606	5.318	0.519	5.355	13.152	8.471
	7	8	9	10	11	12	13
True	5.517	4.765	4.928	14.655	16.034	9.472	3.617
Exp	5.964	5.424	6.493	13.656	13.107	9.543	3.180
Pair	6.118	5.564	6.712	15.655	13.823	9.950	3.280
	14	15	16	17	18	19	20
True	0.323	0.382	6.814	5.376	0.934	2.609	1.282
Exp	0.353	0.340	7.331	5.272	1.082	2.855	1.446
Pair	0.370	0.352	7.626	5.428	1.101	2.948	1.493

construct confidence intervals for the log-odds ratios by using the reversed percentile method. Table 8 shows that compared to the bootstrap method, the RW method achieves smaller MACE and MCIW across different $1-\beta_0$, demonstrating that the RW method can better approximate the distribution of $\alpha^\top \hat{\theta}_n$. In addition, we specifically consider the performance on the variable "HeartDiseaseorAttack", which indicates the status of coronary

Table 8: Results of MACE and MCIW, and interval estimation results for the variable "HeartDiseaseorAttack" (Coverage probability and Width).

	MACE		MCIW		Coverage probability		Width	
β_0	Exp	Pair	Exp	Pair	Exp	Pair	Exp	Pair
0.05	0.029	0.031	0.829	0.864	1.000	1.000	1.017	1.047
0.1	0.035	0.039	0.696	0.723	0.900	0.960	0.852	0.876
0.15	0.044	0.044	0.609	0.632	0.840	0.880	0.745	0.767

heart disease (CHD) or myocardial infarction (MI) (0 = no, 1 = yes) and is known to be closely related to type II diabetes. The coverage probability and the average interval width of this variable are presented in Table 8. It reveals that the RW method achieves empirical coverage probabilities that are closer to the nominal level $1-\beta_0$, and obtains more accurate confidence intervals.

6. Concluding Remarks

We have considered the random weighting approximation of M-estimators for general models with increasing dimensions of parameters. The consistency of θ_n^*, its Bahadur representation, and the validity of the distribution approximation have been established. Our results are quite general and

applicable to linear models, generalized linear models, etc. The simulation results show that when the sample size is not large, the random weighting method has better performance with respect to approximating the variance of a given statistic and performing interval estimation compared to the paired bootstrap.

The random weighting method marks an initial advancement in the approach of high-dimensional scenarios. However, several important research directions remain open for future investigation. For instance, when the parameter dimension increases at the same or even the exponential order as the sample size, the accuracy of the random weighting method in approximating the sampling distribution warrants further examination. Another crucial direction is to extend the random weighting method to high-dimensional censored regression models such as Tobit regression, where the conventional bootstrap approach may be impractical due to its potential censoring proportion. In the literature, the random weighting method for censored regression with fixed-dimensional parameters has been investigated. However, our approach cannot be applied directly to solve high-dimensional scenarios. The main challenge is that the objective function of Powell's estimator for Tobit regression (Powell, 1984) is non-convexity which does not satisfy our convex assumption for the objective function. Consequently,

the developments of new techniques are also interesting topics to explore.

Supplementary Material

The supplementary material contains the proofs of all theorems, propositions, and corollaries, as well as the details of the simulations.

Acknowledgements

This work was supported by the Characteristic & Preponderant Discipline of Key Construction Universities in Zhejiang Province (Zhejiang Gongshang University-Statistics) and the Collaborative Innovation Center of Statistical Data Engineering Technology & Application, and the National Science Foundation of China (No. 12371277, No. 12231017).

References

Chen, K., Z. Ying, H. Zhang, and L. Zhao (2008). Analysis of least absolute deviation. *Biometrika 95*(1), 107–122.

Chernozhukov, V., D. Chetverikov, K. Kato, and Y. Koike (2023). High-dimensional data bootstrap. *Annual Review of Statistics and Its Application 10*(1), 427–449.

Cui, W., K. Li, Y. Yang, and Y. Wu (2008). Random weighting method for cox's proportional hazards model. *Science in China Series A: Mathematics 51*(10), 1843–1854.

REFERENCES

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics 7*(1), 1–26.

El Karoui, N. and E. Purdom (2018). Can we trust the bootstrap in high-dimensions? the case of linear models. *Journal of Machine Learning Research 19*(5), 1–66.

Fang, Y. and L. Zhao (2006). Approximation to the distribution of lad estimators for censored regression by random weighting method. *Journal of Statistical Planning and Inference 136*(4), 1302–1316.

Han, M., Y. Lin, W. Liu, and Z. Wang (2024). Robust inference for subgroup analysis with general transformation models. *Journal of Statistical Planning and Inference 229*, 106100.

He, X. and Q. Shao (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis 73*(1), 120–135.

Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pp. 492–518. Springer.

Lam, H. and Z. Liu (2023). Bootstrap in high dimension with low computation. In *International Conference on Machine Learning*, pp. 18419–18453. PMLR.

Lo, A. Y. (1987). A large sample study of the bayesian bootstrap. *The Annals of Statistics 15*(1), 360–375.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics 25*(3), 303–325.

Rao, C. R. and L. Zhao (1992). Approximation to the distribution of m-estimates in linear models by randomly weighted bootstrap. *Sankhyā: The Indian Journal of Statistics, Series A 54*(3), 323–331.

Rubin, D. B. (1981). The bayesian bootstrap. *The Annals of Statistics 9*(1), 130–134.

Shao, J. and D. Tu (2012). *The jackknife and bootstrap*. Springer Science & Business Media.

Tu, D. and Z. Zheng (1987). The edgeworth expansion for the random weighting method. *Chinese Journal of Applied Probability and Statistics 3*(4), 340–347.

Wang, Z., T. Li, L. Xiao, and D. Tu (2023). A threshold longitudinal tobit quantile regression model for identification of treatment-sensitive subgroups based on interval-bounded longitudinal measurements and a continuous covariate. *Statistics in Medicine 42*(25), 4618–4631.

Wang, Z., Y. Wu, and L. Zhao (2009). Approximation by randomly weighting method in censored regression model. *Science in China Series A: Mathematics 52*(3), 561–576.

Wang, Z., H. Xu, H. Liu, H. Song, and D. Tu (2022). A joint model for longitudinal outcomes with potential ceiling and floor effects and survival times, with applications to analysis of quality of life data from a cancer clinical trial. *Stat 11*(1), e412.

Wang, Z., S. Yao, and Y. Wu (2018). A randomly weighting approximation approach for two-tailed censored regression model. *Scientia Sinica Mathematica 48*(7), 955–968.

Weng, C.-S. (1989). On a second-order asymptotic property of the bayesian bootstrap mean.

The Annals of Statistics 17(2), 705–710.

Wu, X., Y. Yang, and L. Zhao (2007). Approximation by random weighting method for m-test in linear models. *Science in China Series A: Mathematics 50*(1), 87–99.

Wu, Y. and L. Zhao (1999). A large sample study of randomly weighted bootstrap in linear models. *Science in China Series A: Mathematics 42*(10), 1066–1074.

Xiao, L., B. Hou, Z. Wang, and Y. Wu (2014). Random weighting approximation for tobit regression models with longitudinal data. *Computational Statistics & Data Analysis 79*, 235–247.

Xiao, P., X. Liu, A. Li, and G. Pan (2024). Distributed inference for the quantile regression model based on the random weighted bootstrap. *Information Sciences 680*, 121172.

Zheng, Z. (1987). Random weighting method. *Acta Mathematicae Applicatae Sinica (in Chinese) 10*(2), 247–253.

REFERENCES

Ruixing Ming

School of Statistics and Mathematics, Zhejiang Gongshang University

E-mail: rxming@zjgsu.edu.cn

Chengyao Yu

School of Statistics and Mathematics, Zhejiang Gongshang University

E-mail: 2102090137@pop.zjgsu.edu.cn

Min Xiao

School of Statistics and Mathematics, Zhejiang Gongshang University

E-mail: xiaomin@zjgsu.edu.cn

Zhanfeng Wang

School of Management, University of Science and Technology of China

E-mail: zfw@ustc.edu.cn