Statistica Sinica Preprint No: SS-2026-0022	
Title	Optimal Response-Free Cluster Subsampling for Longitudinal Data Under Measurement Constrains
Manuscript ID	SS-2026-0022
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0022
Complete List of Authors	Junhao Shan, Lei Wang and Haiying Wang
Corresponding Authors	Lei Wang
E-mails	lwangstat@nankai.edu.cn
Notice: Accepted author version.	

Optimal response-free cluster subsampling for longitudinal data under measurement constraints

Junhao Shan[1], Lei Wang*[1] and Haiying Wang[2]

[1]*Nankai University and* [2]*University of Connecticut*

Abstract: Under measurement constraints, where covariates are always accessible but obtaining responses is costly or restricted, we propose a unified response-free cluster subsampling framework for massive longitudinal data, focusing on two aspects. First, when the dimension of covariates is fixed and small, to account for within-subject correlation, we consider cluster subsampling and formulate a response-free weighted quasi-score to obtain the subsample estimator with consistency and asymptotic normality. An optimal cluster subsampling scheme is obtained by optimizing a general criterion that encompasses both A-optimality and L-optimality criteria. To enhance the estimation efficiency, a response-free unweighted estimator is subsequently constructed based on the optimal subsample and a two-step algorithm is devised to facilitate practical implementation. Second, when the dimension of covariates is comparable to or exceeds the subsample size, we further construct a response-free weighted quasi decorrelated score for the preconceived low-dimensional parameter of main interest and derive the optimal subsampling schemes. The resulting unweighted estimator and a two-step algorithm are also proposed. Extensive simulation studies, along with a real-data applica-

Corresponding author: Lei Wang, School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai University, Tianjin, 300071, China. E-mail: lwangstat@nankai.edu.cn.

Author's ORCID: Lei Wang, https://orcid.org/0000-0003-2530-883X

tion, are conducted to empirically demonstrate the effectiveness of the proposed methods.

Key words and phrases: A-optimality, decorrelated score, generalized linear models, longitudinal data, Poisson sampling.

1. Introduction

Longitudinal data, consisting of repeated measurements obtained from a given subject over time, play an increasingly central role across scientific disciplines. For accurate analysis of longitudinal data, it is essential to account for within-subject correlation, because ignoring such dependence may lead to inefficient estimation. When the full dataset is available and the dimension of covariates remains fixed and relatively small, a widely adopted method for analyzing longitudinal data is the quasi-score estimator proposed by Liang and Zeger (1986), which accommodates within-cluster correlation and achieves consistent parameter estimation regardless of whether the working correlation is correctly specified (Wang, 2011; Wang et al., 2012). With the rapid expansion of data collection in modern applications, the size of available datasets continues to grow dramatically, and the full-data estimator of Liang and Zeger (1986) is often infeasible due to computational and storage burdens.

1.1 Related work

To address the challenges posed by large-scale data, subsampling techniques have been developed to reduce the computational burden and improve efficiency (Wang et al., 2018; Ai et al., 2021; Wang and Ma, 2021; Wang and Kim, 2022; Yu et al., 2022; Wang et al., 2022; Yu et al., 2025). These techniques involve selecting a subset of data that contains more informative observations in the entire dataset. Despite these advances, subsampling for longitudinal data has received limited attention, representing a promising area for further exploration. A key impact of longitudinal data on subsampling design is the need to account for within-subject correlation, meaning we must sample clusters instead of individual observations. Among the existing studies, Wang et al. (2023) provided an optimal subsampling strategy for longitudinal linear models. While highly effective, this method, along with a large body of subsampling literature, focuses primarily on settings where the dimension of covariates is fixed and remains small relative to the subsample size (the LD-scenario). When the dimension is comparable to or exceeds the subsample size (the HD-scenario), these methods may encounter challenges due to the singularity of the design matrix.

Massive datasets with high-dimensional covariates are now ubiquitous across many disciplines, including biology, economics, and the social sciences. This has led to increased attention on subsampling in high-dimensional settings, with important

contributions from Wang et al. (2024) and Zhang and Wang (2026). While pioneering, these approaches were tailored to specific settings. For instance, the approach of Wang et al. (2024) was specifically designed for binary rare-event data, whereas Zhang and Wang (2026) relied on sample-splitting to enable dimension reduction and refitted cross-validation. Realizing that not all effects of high-dimensional covariates are of primary interest, Gao et al. (2025) proposed optimal subsampling for longitudinal generalized linear models (GLMs) based on the decorrelated score (DS; Ning and Liu, 2017) to conduct estimation and inference for a pre-specified low-dimensional parameter.

In addition to high dimensionality, this paper aims to address subsampling under measurement constraints (Wang et al., 2017), which arise in diverse applications including superconductivity data (Hamidieh, 2018), galaxy surveys (Zhang et al., 2021), and semi-supervised learning. In these contexts, covariates are typically plentiful while responses are scarce due to cost, privacy, or administrative barriers. All the aforementioned subsampling methodologies are inherently response-dependent, i.e., the sampling probabilities are functions of both covariates and responses, and thus are not directly applicable under measurement constraints. To overcome these obstacles, response-free subsampling strategies relying solely on covariates have been proposed (Ma et al., 2015; Xie et al., 2019; Ma et al., 2022; Zhang et al., 2023; Xie et al., 2025). For example, Zhang et al. (2021) and Wang et al. (2024) developed

optimal response-free weighted and unweighted subsample estimators for GLMs under the LD-scenario, and Shao et al. (2025) advanced the field by employing DS to derive an optimal response-free subsampling strategy under the HD-scenario. However, these methods are tailored to non-longitudinal data, highlighting a valuable opportunity to extend these concepts to longitudinal settings.

1.2 Our contributions

Although significant progress has been made in optimal subsampling for parametric models with univariate responses, to the best of our knowledge, the development of an optimal subsampling strategy for longitudinal GLMs under measurement constraints remains an open area, particularly under the HD-scenario. To bridge this gap, this paper introduces four key innovations: (i) the accommodation of within-subject dependence, (ii) the use of response-free subsampling probabilities, (iii) the handling of high-dimensional nuisance parameters, and (iv) the development of both weighted and unweighted subsample estimators. These features collectively address methodological and theoretical complexities that have yet to be fully explored in prior literature. The main contributions of this study are summarized as follows.

(1) Given response-free probabilities, we develop a cluster subsampling framework that maintains and incorporates the correlation structure within each cluster. Under the LD-scenario, we construct our cluster subsample estimator based on

the weighted quasi-score, which enjoys consistency and asymptotic normality (see Theorem 1). Under the HD-scenario, a quasi DS designed for estimation and inference for a pre-specified low-dimensional parameter is further proposed to ensure that the asymptotic properties can still be maintained in the presence of slowly converging nuisance parameters (see Theorem 5). More importantly, we derive unconditional asymptotic distributions for the subsample estimators instead of the conditional distributions given the full data (Wang et al., 2018; Ai et al., 2021), which is sufficient for valid inference on the true parameter.

(2) To pursue efficient cluster subsampling, Theorems 2 and 6 characterize the optimal response-free subsampling schemes by minimizing the traces of the asymptotic covariance matrices associated with the linearly transformed subsample estimators, under A- and L-optimality criteria. Unlike a substantial body of existing methods that directly leverage the response information, our design must rely solely on covariates. Consequently, these derivations offer novel insights that build upon the existing longitudinal subsampling literature (Wang et al., 2023; Gao et al., 2025). A two-step algorithm is also proposed for practical implementation, and Theorems 3 and 7 establish the asymptotic properties of the resultant two-step cluster subsample estimators under the LD- and HD-scenarios, respectively.

(3) To achieve more stable and efficient cluster subsampling, we further construct unweighted subsample estimators based on the optimal subsamples, with their asymptotic properties rigorously established in Theorems 4 and 8. This development requires careful consideration and is feasible precisely because of the response-free subsampling framework, whereas a response-dependent design would typically introduce bias. Moreover, under the Loewner ordering, Theorems 4 and 8 demonstrate that the unweighted subsample estimators possess smaller asymptotic covariance matrices than their weighted counterparts, indicating superior efficiency. Both simulation studies and a real-data application demonstrate that our response-free subsample estimators achieve comparable performance to the response-dependent subsample estimators proposed by Gao et al. (2025), and circumvent the costly measurement of responses for every observation before subsampling. Consequently, our approach offers a dual advantage: substantial estimation efficiency alongside significant practical cost savings.

1.3 Organization

The remainder of the paper is organized as follows. Section 2 develops optimal response-free cluster subsample estimators for GLMs under the LD-scenario, while Section 3 further presents a quasi DS to obtain optimal cluster subsample estimators

under the HD-scenario. Numerical studies illustrating the efficacy of the proposed methods are provided in Section 4, followed by a real-data application in Section 5 that further validates our results. Section 6 provides concluding remarks and discussions. Detailed proofs of all theorems can be found in the Supplementary Material.

2. Optimal response-free cluster subsampling under the LD-scenario

For the i-th subject, let $\boldsymbol{Y}_i = (y_{i1}, \ldots, y_{im_i})^{\mathrm{T}} \in \mathbb{R}^{m_i}$ and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im_i})^{\mathrm{T}} \in \mathbb{R}^{m_i \times p}$ be clustered response vector and covariate matrix, respectively, where y_{ij} and $\boldsymbol{x}_{ij} \in \mathbb{R}^p$ are the j-th measurements for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$. The cluster sizes m_i's are bounded by a fixed constant $m = \max_{1 \le i \le n} m_i$. To ease the presentation, we present the balanced scenario in the main paper (i.e., $m_i = m$ for all subjects) and discuss the more notation-intense case with unbalanced measurements in the Supplementary Material. Assume $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$'s are independent and identically distributed (i.i.d.) copies from a population $(\boldsymbol{X}, \boldsymbol{Y})$, but observations within the same subject are correlated with a common unknown correlation matrix $\boldsymbol{\Pi}_0 = corr(\boldsymbol{Y}|\boldsymbol{X})$. The conditional density of y_{ij} on $\boldsymbol{x}_{ij}$ satisfies a GLM with canonical link,

$$f(y_{ij}|\boldsymbol{\beta}_0, \boldsymbol{x}_{ij}) \propto \exp\big\{\frac{y_{ij}(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{x}_{ij}) - \psi(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{x}_{ij})}{a(\phi)}\big\},$$

where $\psi(t)$ and $a(t)$ are specific functions, $\boldsymbol{\beta}_0$ is the unknown parameter assumed to be in a compact set and ϕ is the dispersion parameter. Without loss of generality,

set $a(\phi) = 1$.

If the full dataset $\mathcal{F}_n = \{(\boldsymbol{X}_i, \boldsymbol{Y}_i), i = 1, \ldots, n\}$ is available and p is small, the customary estimator $\hat{\boldsymbol{\beta}}_{\mathcal{F}}$ of $\boldsymbol{\beta}_0$ is a solution to the following quasi-score equation (Liang and Zeger, 1986),

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{b}) \tilde{\boldsymbol{R}}_{\mathcal{F}}^{-1} \{\boldsymbol{A}_i^{1/2}(\boldsymbol{b})\}^{-1} \{\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{b})\} = \boldsymbol{0}, \quad \boldsymbol{b} \in \mathbb{R}^p,$$

where $\boldsymbol{\mu}_i(\boldsymbol{b}) = (\dot{\psi}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_{i1}), \ldots, \dot{\psi}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_{im}))^{\mathrm{T}}$, $\boldsymbol{A}_i^{1/2}(\boldsymbol{b})$ is the diagonal matrix of order m whose j-th diagonal entry is $\{\ddot{\psi}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_{ij})\}^{1/2}$, $\dot{\psi}$ and $\ddot{\psi} > 0$ are the first and second-order derivatives of ψ, $\tilde{\boldsymbol{R}}_{\mathcal{F}}$ is the estimated working correlation matrix based on the residual-based moment method (Liang and Zeger, 1986). Some common working correlation structures include independent (IND), compound symmetry (CS) and first-order autoregressive (AR) structures, which will be specified in Section 2.2.

2.1 Response-free weighted cluster subsample estimator

Under measurement constraints, however, although covariates $\boldsymbol{X}_i$ can be fully observed across the dataset, the observations of the response $\boldsymbol{Y}_i$ are unavailable at the beginning, and we can only afford to measure the responses on a very small portion of the whole dataset. Therefore, $\hat{\boldsymbol{\beta}}_{\mathcal{F}}$ is infeasible. In this paper, our goal is to determine an optimal response-free cluster subsampling scheme that determines which data points are most beneficial to measure given a budget constraint, and then obtains an efficient subsample estimator of $\boldsymbol{\beta}_0$ with consistency and asymptotic

normality.

Assume we only observe an extensive dataset $\{\boldsymbol{X}_i\}_{i=1}^n$ at the beginning. Given cluster probabilities $\pi_i = \pi_i(\{\boldsymbol{X}_i\}_{i=1}^n)$, where $\sum_{i=1}^n \pi_i = r$ and $\pi_i > 0$ only depends on covariates, take a random cluster subsample of expected cluster size r using Poisson sampling. Once a cluster is selected, all its observations are extracted together, and we measure the corresponding responses within this cluster. Subsequently, we collect the cluster subsample $\mathcal{F}^* = \{(\boldsymbol{X}_i^*, \boldsymbol{Y}_i^*)\}_{i=1}^{r^*}$ based on the associated subsampling probabilities $\{\pi_i^*\}_{i=1}^{r^*}$, where r^* is the actual cluster subsample size produced by Poisson sampling satisfying $E(r^*) = r$. Given an estimated working correlation structure $\tilde{\boldsymbol{R}}$, the response-free weighted cluster subsample estimator $\hat{\boldsymbol{\beta}}_w$ is the solution to the following quasi-score:

$$\boldsymbol{S}_w^*(\boldsymbol{\beta}; \tilde{\boldsymbol{R}}) = \frac{1}{n} \sum_{i \in \mathcal{I}} \frac{1}{\pi_i^*} \boldsymbol{X}_i^{*\mathrm{T}} \{\boldsymbol{A}_i^*(\boldsymbol{\beta})\}^{1/2} \tilde{\boldsymbol{R}}^{-1} \{\boldsymbol{A}_i^*(\boldsymbol{\beta})\}^{-1/2} \{\boldsymbol{Y}_i^* - \boldsymbol{\mu}_i^*(\boldsymbol{\beta})\} = \boldsymbol{0}, \quad (2.1)$$

where $\boldsymbol{\mu}_i^*$ and $\boldsymbol{A}_i^*$ are $\boldsymbol{\mu}_i$ and $\boldsymbol{A}_i$ with $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ replaced by $(\boldsymbol{X}_i^*, \boldsymbol{Y}_i^*)$, respectively, and $\mathcal{I}$ denotes the index set of $\mathcal{F}^*$. Note that the inverse inclusion probability weighting $1/\pi_i^*$ ensures that $E[\boldsymbol{S}_w^*(\boldsymbol{\beta}_0; \boldsymbol{R})] = \boldsymbol{0}$, where $\boldsymbol{R}$ = limit of $\tilde{\boldsymbol{R}}$ is a constant positive-definite matrix but not necessarily the true correlation matrix $\boldsymbol{\Pi}_0$.

To establish the asymptotic properties of $\hat{\boldsymbol{\beta}}_w$, some regularity conditions are listed as follows.

(A.1) The covariates and responses satisfy $\|\boldsymbol{x}_{ij}\| \leq C$, $E(\|\boldsymbol{Y}_i\|^6) \leq C$ for some con-

stant $C > 0$.

(A.2) The matrix $\boldsymbol{R}$ satisfies $0 < \lambda_{\min}(\boldsymbol{R}) \leq \lambda_{\max}(\boldsymbol{R}) < \infty$, where $\lambda_{\min}(\boldsymbol{R})$ and $\lambda_{\max}(\boldsymbol{R})$ are the minimum and maximum eigenvalues of $\boldsymbol{R}$, respectively.

(A.3) The subsampling probabilities satisfy $\max_{1\leq i\leq n}(n\pi_i)^{-1} = O_P(r^{-1})$.

Assumption (A.1) contains commonly used conditions on the covariates and responses (Ai et al., 2021; Yu et al., 2022). Assumption (A.2) poses restrictions on the limit of the estimated correlation matrix for longitudinal data (Wang et al., 2012; Gao et al., 2025). Assumption (A.3), which is typical in Poisson subsampling literature, is introduced to avoid extremely small sampling probabilities that could lead to unstable estimators; see Wang et al. (2022) and Yu et al. (2022).

Theorem 1. *Under Assumptions (A.1)-(A.3), assuming* $\boldsymbol{\Phi} = E[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i]$ *is positive-definite, then* $\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0 = O_P(r^{-1/2})$ *and*

$$(\boldsymbol{\Phi}^{-1}\boldsymbol{\Xi}_\pi\boldsymbol{\Phi}^{-1})^{-1/2}(\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p),$$

where $\boldsymbol{\Xi}_\pi = E\big[n^{-2}\sum_{i=1}^{n}\pi_i^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\big]$.

Remark 1. *When* $m = 1$ *or* $\boldsymbol{R} = \boldsymbol{I}_m$, *our response-free weighted subsample estimator* $\hat{\boldsymbol{\beta}}_w$ *through Poisson sampling degenerates to its sampling-with-replacement counterpart proposed by Zhang et al. (2021). Meanwhile, the asymptotic variance of* $\hat{\boldsymbol{\beta}}_w$ *does not depend on the limit of* r/n, *and is generally smaller than that of the latter.*

2.2 Optimal cluster subsampling probabilities

According to Theorem 1, since the asymptotic variance matrix of $(\hat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}_0)$ depend on $\{\pi_i\}_{i=1}^n$ through $\boldsymbol{\Xi}_\pi$, one can select the optimal probabilities that minimize the trace $\mathrm{tr}(\boldsymbol{D}\boldsymbol{\Phi}^{-1}\boldsymbol{\Xi}_\pi\boldsymbol{\Phi}^{-1}\boldsymbol{D}^{\mathrm{T}})$, where $\boldsymbol{D} \in \mathbb{R}^{p\times p}$ is a known, fixed and nonsingular matrix.

Theorem 2. *Let* $h_i^{\boldsymbol{D}} = [\mathrm{tr}\{\boldsymbol{D}\boldsymbol{\Phi}^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\boldsymbol{\Phi}^{-1}\boldsymbol{D}^{\mathrm{T}}\}]^{1/2}$ *for* $i = 1, \ldots, n$ *and denote* $h_{(1)}^{\boldsymbol{D}} \le h_{(2)}^{\boldsymbol{D}} \le \cdots \le h_{(n)}^{\boldsymbol{D}}$ *as the order statistics of* $\{h_i^{\boldsymbol{D}}\}_{i=1}^n$. *Under the assumptions of Theorem 1, if the subsampling probabilities are chosen as*

$$\pi_i^{\boldsymbol{D}} = \frac{h_i^{\boldsymbol{D}} \wedge M^{\boldsymbol{D}}}{\sum_{i=1}^n (h_i^{\boldsymbol{D}} \wedge M^{\boldsymbol{D}})}, \tag{2.2}$$

then $\mathrm{tr}(\boldsymbol{D}\boldsymbol{\Phi}^{-1}\boldsymbol{\Xi}_\pi\boldsymbol{\Phi}^{-1}\boldsymbol{D}^{\mathrm{T}})$ *attains its minimum, where* $M^{\boldsymbol{D}} = (r-\omega)^{-1}\sum_{i=1}^{n-\omega} h_{(i)}^{\boldsymbol{D}}$ *and* $\omega = \min\{\upsilon \mid 0 \le \upsilon \le r,\ h_{(n-\upsilon)}^{\boldsymbol{D}} < \sum_{i=1}^{n-\upsilon} h_{(i)}^{\boldsymbol{D}}/(r-\upsilon)\}$.

Remark 2. *Theorem 2 shows that the optimal subsampling design depends on the term* $\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}$. *If the working correlation structure is correctly specified, i.e.,* $\boldsymbol{R} = \boldsymbol{\Pi}_0$, *this term becomes* $\boldsymbol{\Pi}_0^{-1}$. *Let* $\vec{\boldsymbol{X}}_i := \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i\boldsymbol{\Phi}^{-1}\boldsymbol{D}^{\mathrm{T}}$, *which can be interpreted as a pseudo-covariate matrix in the* i*-th cluster. The optimal subsampling design favors clusters whose pseudo-covariate has a correlation structure that deviates more from the within-cluster correlation matrix* $\boldsymbol{\Pi}_0$. *When the working correlation is misspecified, the effect of correlation is no longer directly interpretable solely in terms of* $\boldsymbol{\Pi}_0$. *Depending on the discrepancy between* $\boldsymbol{R}$ *and* $\boldsymbol{\Pi}_0$, *the matrix* $\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}$ *can*

take on arbitrary values, potentially leading to systematic misallocation of sampling effort.

Since the optimal probabilities $\{\pi_i^{\boldsymbol{D}}\}_{i=1}^n$ depend on the unknown population level quantities $\boldsymbol{\beta}_0$, $\boldsymbol{R}$, $\boldsymbol{\Pi}_0$ and $\boldsymbol{\Phi}$, they are thus not directly applicable. Below we provide a two-step algorithm to approximate the optimal subsampling probabilities for practical implementation. (i) In the first step, draw a subsample $\mathcal{F}^{*\mathrm{p}} = \{\boldsymbol{X}_i^{*\mathrm{p}}, \boldsymbol{Y}_i^{*\mathrm{p}}\}_{i=1}^{r^{*\mathrm{p}}}$ of expected cluster size r^{p} through uniform Poisson subsampling and then calculate an initial estimator $\tilde{\boldsymbol{\beta}}_{\mathrm{p}}$ under the IND working structure as the solution of $\sum_{i\in\mathcal{I}_{\mathrm{p}}} (\boldsymbol{X}_i^{*\mathrm{p}})^{\mathrm{T}}\{\boldsymbol{Y}_i^{*\mathrm{p}} - \boldsymbol{\mu}_i^{*\mathrm{p}}(\boldsymbol{\beta})\}/r^{\mathrm{p}} = \boldsymbol{0}$ with $\mathcal{I}_{\mathrm{p}}$ being the index set of $\mathcal{F}^{*\mathrm{p}}$. Therefore, $\boldsymbol{\Pi}_0$ can be estimated by $\tilde{\boldsymbol{\Pi}}_{\mathrm{p}} = \sum_{i\in\mathcal{I}_{\mathrm{p}}} [\{\boldsymbol{A}_i^{*\mathrm{p}}(\tilde{\boldsymbol{\beta}}_{\mathrm{p}})\}^{-1/2}\{\boldsymbol{Y}_i^{*\mathrm{p}} - \boldsymbol{\mu}_i^{*\mathrm{p}}(\tilde{\boldsymbol{\beta}}_{\mathrm{p}})\}]^{\otimes 2}/r^{\mathrm{p}}\tilde{\sigma}^2$, where $\tilde{\sigma}^2$ is the estimated variance of the residual, e.g., $\tilde{\sigma}^2 = \sum_{i\in\mathcal{I}_{\mathrm{p}}}\sum_{j=1}^{m}(y_{ij}^{*\mathrm{p}} - \tilde{\boldsymbol{\beta}}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{x}_{ij}^{*\mathrm{p}})^2/(r^{\mathrm{p}}m)$ for linear regression and $\tilde{\sigma}^2 = 1$ for logistic regression and Poisson regression. Next, one can estimate $\boldsymbol{R}$ under a specific model structure. For example, under the IND working structure, $\tilde{\boldsymbol{R}}_{\mathrm{p}} = \boldsymbol{I}_m$; under the CS and AR working structures, all diagonal elements of $\tilde{\boldsymbol{R}}_{\mathrm{p}}$ are equal to 1, while the (j, j')-th off-diagonal elements of $\tilde{\boldsymbol{R}}_{\mathrm{p}}$ are equal to

$$\sum_{i\in\mathcal{I}_{\mathrm{p}}}\sum_{j\neq j'} v_{ij}^{*\mathrm{p}} v_{ij'}^{*\mathrm{p}}/[m(m-1)r^{\mathrm{p}}\tilde{\sigma}^2], \quad \{\sum_{i\in\mathcal{I}_{\mathrm{p}}}\sum_{|j-j'|=1} v_{ij}^{*\mathrm{p}} v_{ij'}^{*\mathrm{p}}/[2(m-1)r^{\mathrm{p}}\tilde{\sigma}^2]\}^{|j-j'|}, \tag{2.3}$$

respectively, with $v_{ij}^{*\mathrm{p}} = [y_{ij}^{*\mathrm{p}} - \dot{\psi}(\tilde{\boldsymbol{\beta}}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{x}_{ij}^{*\mathrm{p}})]/\{\ddot{\psi}(\tilde{\boldsymbol{\beta}}_{\mathrm{p}}^{\mathrm{T}}\boldsymbol{x}_{ij}^{*\mathrm{p}})\}^{1/2}$. Furthermore, $\boldsymbol{\Phi}$ is estimated by $\tilde{\boldsymbol{\Phi}}_{\mathrm{p}} = \sum_{i\in\mathcal{I}_{\mathrm{p}}} (\boldsymbol{X}_i^{*\mathrm{p}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\mathrm{p}}(\tilde{\boldsymbol{\beta}}_{\mathrm{p}})\}^{1/2}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\mathrm{p}}(\tilde{\boldsymbol{\beta}}_{\mathrm{p}})\}^{1/2}\boldsymbol{X}_i^{*\mathrm{p}}/r^{\mathrm{p}}$. Therefore, we can

approximate $\pi_i^{\boldsymbol{D}}$ in (2.2) by

$$\tilde{\pi}_i^{\boldsymbol{D}} = \tilde{h}_i^{\boldsymbol{D}} / \sum_{i=1}^{n} \tilde{h}_i^{\boldsymbol{D}} \tag{2.4}$$

with $\tilde{h}_i^{\boldsymbol{D}} = [\mathrm{tr}\{\boldsymbol{D}\tilde{\boldsymbol{\Phi}}_{\mathrm{p}}^{-1}\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\tilde{\boldsymbol{\beta}}_{\mathrm{p}})\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\tilde{\boldsymbol{\Pi}}_{\mathrm{p}}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\boldsymbol{A}_i^{1/2}(\tilde{\boldsymbol{\beta}}_{\mathrm{p}})\boldsymbol{X}_i\tilde{\boldsymbol{\Phi}}_{\mathrm{p}}^{-1}\boldsymbol{D}^{\mathrm{T}}\}]^{1/2}$ for $i = 1, \ldots, n$. Note that we directly take $M^{\boldsymbol{D}} = \infty$ as suggested by Yu et al. (2022) to attain faster calculation, which necessitates a truncation step $(\tilde{\pi}_i^{\boldsymbol{D}} \wedge 1)$ to guarantee the probability bound. (ii) In the second step, we draw another informative subsample $\mathcal{F}^{*\boldsymbol{D}} = \{\boldsymbol{X}_i^{*\boldsymbol{D}}, \boldsymbol{Y}_i^{*\boldsymbol{D}}\}_{i=1}^{r^*}$ of expected cluster size $r \geq r^{\mathrm{p}}$ using the estimated $(\tilde{\pi}_i^{\boldsymbol{D}} \wedge 1)$ in (2.4) from the full dataset after removing $\mathcal{F}^{*\mathrm{p}}$. The final two-step estimator $\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}}$ is the solution to

$$\frac{1}{n}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}} \frac{1}{\tilde{\pi}_i^{*\boldsymbol{D}} \wedge 1}(\boldsymbol{X}_i^{*\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}^{1/2}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}^{-1/2}\{\boldsymbol{Y}_i^{*\boldsymbol{D}} - \boldsymbol{\mu}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\} = \boldsymbol{0}$$

with $\mathcal{I}_{\boldsymbol{D}}$ denoting the index set of $\mathcal{F}^{*\boldsymbol{D}}$.

Theorem 3. *Under the assumptions of Theorem 1, we have* $\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}} - \boldsymbol{\beta}_0 = O_P(r^{-1/2})$ *and*

$$(m^{\boldsymbol{D}}\boldsymbol{\Phi}^{-1}\boldsymbol{\Xi}^{\boldsymbol{D}}\boldsymbol{\Phi}^{-1})^{-1/2}\{\sqrt{r}(\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}} - \boldsymbol{\beta}_0)\} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p),$$

with $\boldsymbol{\Xi}^{\boldsymbol{D}} = E\big[\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{X}_i/h_i^{\boldsymbol{D}}\big]$ *and* $m^{\boldsymbol{D}} = E(h_i^{\boldsymbol{D}})$.

To conduct statistical inference, we propose to estimate the asymptotic covariance matrix of $\sqrt{r}(\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}} - \boldsymbol{\beta}_0)$ by $(\hat{\boldsymbol{\Phi}}^{\boldsymbol{D}})^{-1}\hat{\boldsymbol{\Xi}}^{\boldsymbol{D}}(\hat{\boldsymbol{\Phi}}^{\boldsymbol{D}})^{-1}$, where $\hat{\boldsymbol{\Phi}}^{\boldsymbol{D}} = n^{-1}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}(\tilde{\pi}_i^{*\boldsymbol{D}} \wedge 1)^{-1}$ $(\boldsymbol{X}_i^{*\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}})\}^{1/2}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*D}(\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}})\}^{1/2}\boldsymbol{X}_i^{*\boldsymbol{D}}$, $\hat{\boldsymbol{\Xi}}^{\boldsymbol{D}} = n^{-2}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}(\tilde{\pi}_i^{*\boldsymbol{D}} \wedge 1)^{-2}(\boldsymbol{X}_i^{*\boldsymbol{D}})^{\mathrm{T}}$ $\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}})\}^{1/2}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\tilde{\boldsymbol{\Pi}}_{\mathrm{p}}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\hat{\boldsymbol{\beta}}_w^{\boldsymbol{D}})\}^{1/2}\boldsymbol{X}_i^{*\boldsymbol{D}}$.

2.3 Response-free unweighted cluster subsample estimator

To simultaneously bypass inverse probability weighting and improve estimation efficiency, based on the optimal subsample $\mathcal{F}^{*\boldsymbol{D}} = \{\boldsymbol{X}_i^{*\boldsymbol{D}}, \boldsymbol{Y}_i^{*\boldsymbol{D}}\}_{i=1}^{r^*}$, we propose another unweighted cluster subsample estimator $\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}$ as the solution to $\boldsymbol{S}_{uw}^{*\boldsymbol{D}}(\boldsymbol{\beta}; \tilde{\boldsymbol{R}}_{\mathrm{p}}) = \boldsymbol{0}$ with

$$\boldsymbol{S}_{uw}^{*\boldsymbol{D}}(\boldsymbol{\beta}; \tilde{\boldsymbol{R}}_{\mathrm{p}}) = \frac{1}{n} \sum_{i \in \mathcal{I}_{\boldsymbol{D}}} (\boldsymbol{X}_i^{*\boldsymbol{D}})^{\mathrm{T}} \{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}^{1/2} \tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1} \{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}^{-1/2} \{\boldsymbol{Y}_i^{*\boldsymbol{D}} - \boldsymbol{\mu}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}. \tag{2.5}$$

The next theorem establishes asymptotic results for $\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}$, and further demonstrates that $\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}$ attains superior estimation efficiency than its weighted counterpart $\hat{\boldsymbol{\beta}}_{w}^{\boldsymbol{D}}$.

Theorem 4. *Under Assumptions (A.1)-(A.2), assuming* $\boldsymbol{\Gamma}^{\boldsymbol{D}} = E[h_i^{\boldsymbol{D}} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{X}_i]$ *is positive-definite, then* $\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}} - \boldsymbol{\beta}_0 = O_P(r^{-1/2})$ *and*

$$[m^{\boldsymbol{D}} (\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1} \boldsymbol{\Delta}^{\boldsymbol{D}} (\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1}]^{-1/2} \{\sqrt{r}(\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}} - \boldsymbol{\beta}_0)\} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_p),$$

where $\boldsymbol{\Delta}^{\boldsymbol{D}} = E[h_i^{\boldsymbol{D}} \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{\Pi}_0 \boldsymbol{R}^{-1} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{X}_i]$. *Moreover, if* $\boldsymbol{R} = \boldsymbol{\Pi}_0$, *then* $(\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1} \boldsymbol{\Delta}^{\boldsymbol{D}} (\boldsymbol{\Gamma}^{\boldsymbol{D}})^{-1} \leq \boldsymbol{\Phi}^{-1} \boldsymbol{\Xi}^{\boldsymbol{D}} \boldsymbol{\Phi}^{-1}$, *where the inequalities are in the Loewner ordering.*

Remark 3. *Our proposed unweighted estimator* $\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}$ *still remains asymptotically unbiased thanks to its response-free subsampling design. In contrast, a response-dependent strategy, such as that employed by Gao et al. (2025), would introduce asymptotic bias based on unweighted estimation. On the other hand, similar to Remark 1, the unweighted subsample estimator proposed by Wang et al. (2024) through*

sampling with replacement emerge as a special case of $\hat{\boldsymbol{\beta}}_{uw}^{D}$ *through Poisson sampling when* $m = 1$*, and also yields generally larger asymptotic variance than* $\hat{\boldsymbol{\beta}}_{uw}^{D}$ *(Wang et al., 2022).*

To conduct statistical inference, we propose to estimate the asymptotic covariance matrix of $\sqrt{r}(\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\beta}_0)$ by $\hat{m}^{\boldsymbol{D}}(\hat{\boldsymbol{\Gamma}}^{\boldsymbol{D}})^{-1}\hat{\boldsymbol{\Delta}}^{\boldsymbol{D}}(\hat{\boldsymbol{\Gamma}}^{\boldsymbol{D}})^{-1}$, where $\hat{\boldsymbol{\Gamma}}^{\boldsymbol{D}} = n^{-1}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}\hat{m}^{\boldsymbol{D}}(\boldsymbol{X}_i^{*\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}})\}^{1/2}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}})\}^{1/2}\boldsymbol{X}_i^{*\boldsymbol{D}}$, $\hat{\boldsymbol{\Delta}}^{\boldsymbol{D}} = n^{-1}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}\hat{m}^{\boldsymbol{D}}(\boldsymbol{X}_i^{*\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}})\}^{1/2}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\tilde{\boldsymbol{\Pi}}_{\mathrm{p}}\tilde{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*D}(\hat{\boldsymbol{\beta}}_{uw}^{\boldsymbol{D}})\}^{1/2}\boldsymbol{X}_i^{*\boldsymbol{D}}$.

3. Optimal response-free cluster subsampling under the HD-scenario

In this section, we assume that the dimension of $\boldsymbol{\beta}_0$ can be larger than the subsample size r (HD-scenario). Correspondingly, the covariates can be divided into two components, i.e., $\boldsymbol{x}_{ij} = (\boldsymbol{z}_{ij}^{\mathrm{T}}, \boldsymbol{u}_{ij}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{z}_{ij}$ is a pre-determined low-dimensional subvector of $\boldsymbol{x}_{ij}$ with dimension $d \ll p$ containing the covariates of primary interest (based on prior knowledge or specified in advance) and the remaining covariates $\boldsymbol{u}_{ij}$ are extraneous (they may be associated with y_{ij} and/or $\boldsymbol{z}_{ij}$). Consequently, the density of $y_{ij}|\boldsymbol{x}_{ij}$ becomes

$$f(y_{ij}|\boldsymbol{\beta}_0, \boldsymbol{x}_{ij}) \propto \exp\{y_{ij}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{z}_{ij} + \boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{u}_{ij}) - \psi(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{z}_{ij} + \boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{u}_{ij})\}, \tag{3.6}$$

and we aim to conduct parameter estimation and statistical inference for $\boldsymbol{\theta}_0$ in the presence of high-dimensional nuisance parameter $\boldsymbol{\gamma}_0$.

3.1 Response-free weighted DS cluster subsample estimator

Assign response-free subsampling probabilities $\varpi_i = \varpi_i(\{\boldsymbol{X}_i\}_{i=1}^n)$ to the i-th data cluster, draw a random subsample of expected cluster size r ($r \ll n$) from the full dataset using Poisson sampling to acquire the measurements of response. Subsequently, we collect the cluster subsample $\mathcal{F}^* = \{(\boldsymbol{X}_i^*, \boldsymbol{Y}_i^*)\}_{i=1}^{r^*}$ based on the associated subsampling probabilities $\{\varpi_i^*\}_{i=1}^{r^*}$ and $E(r^*) = \sum_{i=1}^n \varpi_i = r$. With an estimated working correlation structure $\breve{\boldsymbol{R}}$, we define the following weighted subsampling quasi DS function motivated by Ning and Liu (2017),

$$\boldsymbol{Q}_w^*(\boldsymbol{\theta}; \boldsymbol{\gamma}, \boldsymbol{W}_0, \breve{\boldsymbol{R}}) = \frac{1}{n} \sum_{i \in \mathcal{I}} \frac{1}{\varpi_i^*} (\boldsymbol{Z}_i^* - \boldsymbol{U}_i^* \boldsymbol{W}_0)^{\mathrm{T}} \{\boldsymbol{A}_i^*(\boldsymbol{\beta})\}^{1/2} \breve{\boldsymbol{R}}^{-1} \{\boldsymbol{A}_i^*(\boldsymbol{\beta})\}^{-1/2} \{\boldsymbol{Y}_i^* - \boldsymbol{\mu}_i^*(\boldsymbol{\beta})\}, \tag{3.7}$$

where

$$\begin{aligned} \boldsymbol{W}_0 &= \arg\min_{\boldsymbol{\omega}} E\{\|\boldsymbol{R}^{-1/2} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i \boldsymbol{\omega})\|_F^2\} \\ &= \big[E\{\boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{U}_i\}\big]^{-1} \big[E\{\boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{R}^{-1} \boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0) \boldsymbol{Z}_i\}\big], \end{aligned}$$

and $\boldsymbol{R}$ = limit of $\breve{\boldsymbol{R}}$. Note that the inverse probability weighting guarantees the unbiasedness of the estimating equation, i.e., $E[\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0; \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R})] = \boldsymbol{0}$. Besides, $\boldsymbol{Q}_w^*(\boldsymbol{\theta}_0; \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R})$ further satisfies $E[\nabla_{\boldsymbol{\gamma}} \boldsymbol{Q}_w^*(\boldsymbol{\theta}_0; \boldsymbol{\gamma}_0, \boldsymbol{W}_0, \boldsymbol{R})] = \boldsymbol{0}$, which is the orthogonality property such that the convergence rate of the subsample estimator of $\boldsymbol{\theta}_0$ will not be affected by slowly converging estimators of $\boldsymbol{\gamma}_0$.

To solve the subsampling quasi DS function (3.7), initial estimators of $\boldsymbol{\beta}_0$, $\boldsymbol{R}$ and

$\boldsymbol{W}_0$ are needed. As in Shao et al. (2025), we draw another independent pilot subsample of expected cluster size r^{p} from the full data with uniform Poisson sampling, denoted as $\mathcal{F}^{*\mathrm{p}} = \{(\boldsymbol{Y}_i^{*\mathrm{p}}, \boldsymbol{X}_i^{*\mathrm{p}})\}_{i=1}^{r^{*\mathrm{p}}}$ with $E(r^{*\mathrm{p}}) = r^{\mathrm{p}}$, and then obtain the following Lasso-type estimator $\breve{\boldsymbol{\beta}}_{\mathrm{p}}$ under IND structure (Tibshirani, 1996):

$$\breve{\boldsymbol{\beta}}_{\mathrm{p}} = \arg\min_{\boldsymbol{\beta}} \Big\{ \frac{1}{mr^{\mathrm{p}}} \sum_{i\in\mathcal{I}_{\mathrm{p}}} \sum_{j=1}^{m} \big[-y_{ij}^{*\mathrm{p}} (\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_{ij}^{*\mathrm{p}}) + \psi(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{x}_{ij}^{*\mathrm{p}}) \big] + \lambda_1 \|\boldsymbol{\beta}\|_1 \Big\}, \tag{3.8}$$

where $\mathcal{I}_{\mathrm{p}}$ is the index set of $\mathcal{F}^{*\mathrm{p}}$ and λ_1 is a regularized parameter. Next, the estimator $\breve{\boldsymbol{R}}_{\mathrm{p}}$ of $\boldsymbol{R}$ can be estimated using the same method in (2.3) of Section 2.2, and then we obtain a Lasso-type estimator of $\boldsymbol{W}_0$ as follows (Obozinski et al., 2011):

$$\breve{\boldsymbol{W}}_{\mathrm{p}} = \arg\min_{\boldsymbol{\omega}} \Big\{ \frac{1}{r^{\mathrm{p}}} \sum_{i\in\mathcal{I}_{\mathrm{p}}} \|\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1/2} \{\boldsymbol{A}_i^{*\mathrm{p}}(\breve{\boldsymbol{\beta}}_{\mathrm{p}})\}^{1/2} (\boldsymbol{Z}_i^{*\mathrm{p}} - \boldsymbol{U}_i^{*\mathrm{p}} \boldsymbol{\omega})\|_F^2 + \lambda_2 \sum_{j=1}^{p-d} \|\boldsymbol{\omega}_j\| \Big\}, \tag{3.9}$$

where $\boldsymbol{\omega}_j$ is the j-th row of $\boldsymbol{\omega}$ and λ_2 is a regularized parameter. Plugging $\breve{\boldsymbol{\beta}}_{\mathrm{p}}$, $\breve{\boldsymbol{R}}_{\mathrm{p}}$ and $\breve{\boldsymbol{W}}_{\mathrm{p}}$ into (3.7), the weighted subsampling quasi DS estimator, denoted as $\breve{\boldsymbol{\theta}}_w$, based on the subsample is the solution to

$$\begin{aligned}
&\boldsymbol{Q}_w^*(\boldsymbol{\theta}; \breve{\gamma}_{\mathrm{p}}, \breve{\boldsymbol{W}}_{\mathrm{p}}, \breve{\boldsymbol{R}}_{\mathrm{p}}) \\
=&\frac{1}{n} \sum_{i\in\mathcal{I}} \frac{1}{\varpi_i^*} (\boldsymbol{Z}_i^* - \boldsymbol{U}_i^* \breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}} \{\boldsymbol{A}_i^*(\boldsymbol{\theta}, \breve{\gamma}_{\mathrm{p}})\}^{1/2} \breve{\boldsymbol{R}}_{\mathrm{p}}^{-1} \{\boldsymbol{A}_i^*(\boldsymbol{\theta}, \breve{\gamma}_{\mathrm{p}})\}^{-1/2} \{\boldsymbol{Y}_i^* - \boldsymbol{\mu}_i^*(\boldsymbol{\theta}, \breve{\gamma}_{\mathrm{p}})\} = \boldsymbol{0}.
\end{aligned}$$

To establish the asymptotic properties of $\breve{\boldsymbol{\theta}}_w$, some further regularity conditions are needed.

(A.4) The covariates and model residuals satisfy $\|\boldsymbol{x}_{ij}\|_\infty \le C$, $\|y_{ij} - \dot{\psi}(\boldsymbol{\beta}_0^{\mathrm{T}} \boldsymbol{x}_{ij})\|_{\psi_1} \le C$ for some constant $C > 0$, where $\|X\|_{\psi_1} = \sup_{c\ge 1} c^{-1} \big\{ E\big[|X|^c\big] \big\}^{1/c}$ for a random variable X.

(A.5) The true parameter $\boldsymbol{\beta}_0$ is sparse with support $\mathcal{S}_{\boldsymbol{\beta}_0}$ and $|\mathcal{S}_{\boldsymbol{\beta}_0}| = s_{\boldsymbol{\beta}}$. The matrix $\boldsymbol{W}_0$ is sparse with support $\mathcal{S}_{\boldsymbol{W}_0} = \{l : \boldsymbol{w}_{0l} \neq \mathbf{0}, l = 1 \ldots, p-d\}$, where $\boldsymbol{w}_{0l}$ is the l-th row of $\boldsymbol{W}_0$, $|\mathcal{S}_{\boldsymbol{W}_0}| = s_{\boldsymbol{W}}$ and $\max_{1\leq l\leq d} |\boldsymbol{w}_{0l}\boldsymbol{u}_{ij}| \leq C$.

(A.6) For any set $\mathcal{P} \subset \{1, \ldots, p\}$ and any vector $\boldsymbol{v}$ belonging to the cone $\mathcal{C}(\mathcal{P}, \kappa) = \{\boldsymbol{v} \in \mathbb{R}^p : \|\boldsymbol{v}_{\mathcal{P}^c}\|_1 \leq \kappa \|\boldsymbol{v}_{\mathcal{P}}\|_1\}$, where $\boldsymbol{v}_{\mathcal{P}}$ is the vector containing components of $\boldsymbol{v}$ with indexes in $\mathcal{P}$ and $\mathcal{P}^c$ is the complement of $\mathcal{P}$, there exists a constant $C > 0$ such that

$$\inf_{\mathbf{0}\neq\boldsymbol{v}\in\mathcal{C}(\mathcal{P},\kappa)} \sum_{i\in\mathcal{I}} \sum_{j=1}^{m} \frac{\left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}_{ij}^*\right)^2 \ddot{\psi}\left(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{x}_{ij}^*\right)}{n\varpi_i^*\|\boldsymbol{v}\|^2} \geq C.$$

(A.7) The subsample sizes satisfy $r^{-1}\log p = o(1)$, $r^{1/2}(r^{\mathrm{p}})^{-1}(s_{\boldsymbol{W}} \vee s_{\boldsymbol{\beta}})\log p = o(1)$ and $(r^{\mathrm{p}})^{-1/2}(s_{\boldsymbol{W}} \vee s_{\boldsymbol{\beta}})\log p = o(1)$. The regularized parameters λ_1 and λ_2 satisfy $\lambda_1 = O(\sqrt{\log p/r^{\mathrm{p}}})$ and $\lambda_2 = O(\sqrt{\log p/r^{\mathrm{p}}})$.

(A.8) The subsampling probabilities satisfy $\max_{1\leq i\leq n}(n\varpi_i)^{-1} = O_P(r^{-1})$.

Assumption (A.4) is a version of Assumption (A.1) for high-dimensional $\boldsymbol{\beta}_0$, which is stronger than Assumption (A.1) but standard for analyzing high-dimensional GLMs (Wang et al., 2012; Fang et al., 2020). Assumption (A.5) presents the sparsity of $\boldsymbol{\beta}_0$ and $\boldsymbol{W}_0$, which is common for high-dimensional GLMs (Ning and Liu, 2017; Shao et al., 2025). Assumption (A.6), known as the restricted eigenvalue condition, provides the necessary curvature of the loss function within a cone (Raskutti et al.,

2010; Fang et al., 2020; Gao et al., 2025). Assumption (A.7) imposes some restrictions on the subsampling sizes r^{p} and r, akin to assumptions in Gao et al. (2025) and Shao et al. (2025). Assumption (A.8) is similar to Assumption (A.4) for LD-scenario.

Theorem 5. *Under Assumptions (A.2) and (A.4)-(A.8), assuming* $\boldsymbol{\Lambda} = E\big[(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)\big]$ *is positive-definite, then* $\check{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0 = O_P(r^{-1/2})$ *and*

$$(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}_{\varpi}\boldsymbol{\Lambda}^{-1})^{-1/2}(\check{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}_d),$$

where $\boldsymbol{\Upsilon}_{\varpi} = E\big[n^{-2}\sum_{i=1}^{n}\varpi_i^{-1}(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)\big]$.

3.2 Optimal DS cluster subsampling probabilities

According to Theorem 5, since the asymptotic variance matrix of $(\check{\boldsymbol{\theta}}_w - \boldsymbol{\theta}_0)$ depends on $\{\varpi_i\}_{i=1}^n$ through $\boldsymbol{\Upsilon}_{\varpi}$, one can select the optimal probabilities that minimizes $\mathrm{tr}(\boldsymbol{D}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}_{\varpi}\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{\mathrm{T}})$, where $\boldsymbol{D} \in \mathbb{R}^{d\times d}$ is a known, fixed and nonsingular matrix.

Theorem 6. *Let* $\hbar_i^{\boldsymbol{D}} = [\mathrm{tr}\{\boldsymbol{D}\boldsymbol{\Lambda}^{-1}(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{\mathrm{T}}\}]^{1/2}$ *for* $i = 1, \ldots, n$ *and denote* $\hbar_{(1)}^{\boldsymbol{D}} \le \hbar_{(2)}^{\boldsymbol{D}} \le \cdots \le \hbar_{(n)}^{\boldsymbol{D}}$ *as the order statistics of* $\{\hbar_i^{\boldsymbol{D}}\}_{i=1}^n$. *Under the assumptions of Theorem 5, if the subsampling probabilities are chosen as*

$$\varpi_i^{\boldsymbol{D}} = \frac{\hbar_i^{\boldsymbol{D}} \wedge L^{\boldsymbol{D}}}{\sum_{i=1}^{n}(\hbar_i^{\boldsymbol{D}} \wedge L^{\boldsymbol{D}})}, \tag{3.10}$$

then $\mathrm{tr}(\boldsymbol{D}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}_{\varpi}\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{\mathrm{T}})$ *attains its minimum, where* $L^{\boldsymbol{D}} = (r-\delta)^{-1}\sum_{i=1}^{n-\delta}\hbar_{(i)}^{\boldsymbol{D}}$ *and* $\delta = \min\{\upsilon \mid 0 \le \upsilon \le r,\ \hbar_{(n-\upsilon)}^{\boldsymbol{D}} < \sum_{i=1}^{n-\upsilon}\hbar_{(i)}^{\boldsymbol{D}}/(r-\upsilon)\}$.

Since the optimal probabilities $\{\varpi_i^{\boldsymbol{D}}\}_{i=1}^n$ depend on the unknown population level quantities $\boldsymbol{\beta}_0$, $\boldsymbol{R}$, $\boldsymbol{\Pi}_0$, $\boldsymbol{W}_0$ and $\boldsymbol{\Lambda}$, they are thus not directly applicable. Below we provide a two-step algorithm to approximate the optimal subsampling probabilities for practical implementation. (i) In the first step, we draw a pilot subsample denoted as $\mathcal{F}^{*\mathrm{p}} = \{\boldsymbol{X}_i^{*\mathrm{p}}, \boldsymbol{Y}_i^{*\mathrm{p}}\}_{i=1}^{r^{*\mathrm{p}}}$ of expected cluster size r^{p} through uniform Poisson subsampling and then the initial estimators $\breve{\boldsymbol{\beta}}_{\mathrm{p}}$, $\breve{\boldsymbol{\Pi}}_{\mathrm{p}}$, $\breve{\boldsymbol{R}}_{\mathrm{p}}$, $\breve{\boldsymbol{W}}_{\mathrm{p}}$ can be calculated in the same way as in (2.3) and (3.8)-(3.9), respectively. Furthermore, the estimator of $\boldsymbol{\Lambda}$ is calculated by $\breve{\boldsymbol{\Lambda}}_{\mathrm{p}} = \sum_{i\in\mathcal{I}_{\mathrm{p}}}(\boldsymbol{Z}_i^{*\mathrm{p}} - \boldsymbol{U}_i^{*\mathrm{p}}\breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\mathrm{p}}(\breve{\boldsymbol{\beta}}_{\mathrm{p}})\}^{1/2}\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\mathrm{p}}(\breve{\boldsymbol{\beta}}_{\mathrm{p}})\}^{1/2}(\boldsymbol{Z}_i^{*\mathrm{p}} - \boldsymbol{U}_i^{*\mathrm{p}}\breve{\boldsymbol{W}}_{\mathrm{p}})/r^{\mathrm{p}}$, and then we obtain the estimator of $\varpi_i^{\boldsymbol{D}}$ by $\breve{\varpi}_i^{\boldsymbol{D}} = \breve{\hbar}_i^{\boldsymbol{D}}/\sum_{i=1}^n\breve{\hbar}_i^{\boldsymbol{D}}$, where

$$\breve{\hbar}_i^{\boldsymbol{D}} = [\mathrm{tr}\{\boldsymbol{D}\breve{\boldsymbol{\Lambda}}_{\mathrm{p}}^{-1}(\boldsymbol{Z}_i - \boldsymbol{U}_i\breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\breve{\boldsymbol{\beta}}_{\mathrm{p}})\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\breve{\boldsymbol{\Pi}}_{\mathrm{p}}\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\boldsymbol{A}_i^{1/2}(\breve{\boldsymbol{\beta}}_{\mathrm{p}})(\boldsymbol{Z}_i - \boldsymbol{U}_i\breve{\boldsymbol{W}}_{\mathrm{p}})\breve{\boldsymbol{\Lambda}}_{\mathrm{p}}^{-1}\boldsymbol{D}^{\mathrm{T}}\}]^{1/2}.$$

(ii) In the second step, we draw another informative subsample $\mathcal{F}^{*\boldsymbol{D}} = \{\boldsymbol{X}_i^{*\boldsymbol{D}}, \boldsymbol{Y}_i^{*\boldsymbol{D}}\}_{i=1}^{r^*}$ of expected size r using $(\breve{\varpi}_i^{\boldsymbol{D}} \wedge 1)$ from the full dataset after removing $\mathcal{F}^{*\mathrm{p}}$. The final estimator $\breve{\boldsymbol{\theta}}_w^{\boldsymbol{D}}$ is the solution to

$$\frac{1}{n}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}\frac{1}{\breve{\varpi}_i^{*\boldsymbol{D}}\wedge 1}(\boldsymbol{Z}_i^{*\boldsymbol{D}} - \boldsymbol{U}_i^{*\boldsymbol{D}}\breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\theta},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}^{1/2}\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\theta},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}^{-1/2}\{\boldsymbol{Y}_i^{*\boldsymbol{D}} - \boldsymbol{\mu}_i^{*\boldsymbol{D}}(\boldsymbol{\theta},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\} = \boldsymbol{0}.$$

Theorem 7. *Under the assumptions of Theorem 5, we have* $\breve{\boldsymbol{\theta}}_w^{\boldsymbol{D}} - \boldsymbol{\theta}_0 = O_P(r^{-1/2})$ *and*

$$(l^{\boldsymbol{D}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}^{\boldsymbol{D}}\boldsymbol{\Lambda}^{-1})\{\sqrt{r}(\breve{\boldsymbol{\theta}}_w^{\boldsymbol{D}} - \boldsymbol{\theta}_0)\} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{I}_d),$$

where $\boldsymbol{\Upsilon}^{\boldsymbol{D}} = E[(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{W}_0)/\hbar_i^{\boldsymbol{D}}]$ *and* $l^{\boldsymbol{D}} = E(\hbar_i^{\boldsymbol{D}})$.

To conduct statistical inference, we propose to estimate the asymptotic covariance matrix of $\sqrt{r}(\check{\boldsymbol{\theta}}_w^{\boldsymbol{D}} - \boldsymbol{\theta}_0)$ by $(\check{\boldsymbol{\Lambda}}^{\boldsymbol{D}})^{-1}\check{\boldsymbol{\Upsilon}}^{\boldsymbol{D}}(\check{\boldsymbol{\Lambda}}^{\boldsymbol{D}})^{-1}$, where $\check{\boldsymbol{\Lambda}}^{\boldsymbol{D}} = n^{-1}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}(\breve{\varpi}_i^{*\boldsymbol{D}}\wedge 1)^{-1}$ $(\boldsymbol{Z}_i^{*\boldsymbol{D}} - \boldsymbol{U}_i^{*\boldsymbol{D}}\breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\check{\boldsymbol{\theta}}_w^{\boldsymbol{D}}, \breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}^{1/2}\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\check{\boldsymbol{\theta}}_w^{\boldsymbol{D}}, \breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}^{1/2}(\boldsymbol{Z}_i^{*\boldsymbol{D}} - \boldsymbol{U}_i^{*\boldsymbol{D}}\breve{\boldsymbol{W}}_{\mathrm{p}})$, $\check{\boldsymbol{\Upsilon}}^{\boldsymbol{D}} = n^{-2}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}(\breve{\varpi}_i^{*\boldsymbol{D}}\wedge 1)^{-2}$ $(\boldsymbol{Z}_i^{*\boldsymbol{D}} - \boldsymbol{U}_i^{*\boldsymbol{D}}\breve{\boldsymbol{W}}_{\mathrm{p}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\check{\boldsymbol{\theta}}_w^{\boldsymbol{D}}, \breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}^{1/2}\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\breve{\boldsymbol{\Pi}}_{\mathrm{p}}\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\check{\boldsymbol{\theta}}_w^{\boldsymbol{D}}, \breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}^{1/2}$ $(\boldsymbol{Z}_i^{*\boldsymbol{D}} - \boldsymbol{U}_i^{*\boldsymbol{D}}\breve{\boldsymbol{W}}_{\mathrm{p}})$.

3.3 Response-free unweighted DS cluster subsample estimator

Similar to Section 2.3, we further define the unweighted subsampling quasi DS function for $\boldsymbol{\theta}_0$ as

$$\boldsymbol{Q}_{uw}^{*\boldsymbol{D}}(\boldsymbol{\theta};\boldsymbol{\gamma},\boldsymbol{W}_0^{\boldsymbol{D}},\check{\boldsymbol{R}}) = \frac{1}{n}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}(\boldsymbol{Z}_i^{*\boldsymbol{D}} - \boldsymbol{U}_i^{*\boldsymbol{D}}\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}^{1/2}\check{\boldsymbol{R}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\}^{-1/2}\{\boldsymbol{Y}_i^{*\boldsymbol{D}} - \boldsymbol{\mu}_i^{*\boldsymbol{D}}(\boldsymbol{\beta})\},$$

where $\boldsymbol{W}_0^{\boldsymbol{D}}$ is another projection matrix to ensure $E[\nabla_{\boldsymbol{\gamma}}\boldsymbol{Q}_{uw}^{*\boldsymbol{D}}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\boldsymbol{W}_0^{\boldsymbol{D}},\boldsymbol{R})] = \boldsymbol{0}$. After some derivations, the matrix $\boldsymbol{W}_0^{\boldsymbol{D}}$ is given by

$$\begin{aligned}\boldsymbol{W}_0^{\boldsymbol{D}} &= \arg\min_{\boldsymbol{\omega}} E\{\|(\hbar_i^{\boldsymbol{D}})^{1/2}\boldsymbol{R}^{-1/2}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i - \boldsymbol{U}_i\boldsymbol{\omega})\|_F^2\}\\ &= \big[E\{\hbar_i^{\boldsymbol{D}}\boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{U}_i\}\big]^{-1}\big[E\{\hbar_i^{\boldsymbol{D}}\boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{Z}_i\}\big],\end{aligned}$$

which can also be consistently estimated using the pilot subsample, i.e.,

$$\breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}} = \arg\min_{\boldsymbol{\omega}}\Big\{\frac{1}{r^{\mathrm{p}}}\sum_{i\in\mathcal{I}_{\mathrm{p}}}\|(\breve{\hbar}_i^{\boldsymbol{D}*})^{1/2}\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1/2}\{\boldsymbol{A}_i^{*\mathrm{p}}(\breve{\boldsymbol{\beta}}_{\mathrm{p}})\}^{1/2}(\boldsymbol{Z}_i^{*\mathrm{p}} - \boldsymbol{U}_i^{*\mathrm{p}}\boldsymbol{\omega})\|_F^2 + \lambda_3\sum_{j=1}^{p-d}\|\boldsymbol{\omega}_j\|\Big\}$$

with a regularized parameter λ_3, and $\breve{h}_i^{\boldsymbol{D}*}$ has the same form as $\breve{h}_i^{\boldsymbol{D}}$ with $\boldsymbol{Z}_i$, $\boldsymbol{U}_i$ and $\boldsymbol{A}_i$ substituted by $\boldsymbol{Z}_i^{*\mathrm{p}}$, $\boldsymbol{U}_i^{*\mathrm{p}}$ and $\boldsymbol{A}_i^{*\mathrm{p}}$, respectively. Therefore, the unweighted subsampling quasi DS estimator, denoted as $\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}$, is a solution to the following equation:

$$\frac{1}{n}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}(\boldsymbol{Z}_i^{*\boldsymbol{D}}-\boldsymbol{U}_i^{*\boldsymbol{D}}\breve{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\theta},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}^{1/2}\breve{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\boldsymbol{\theta},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}^{-1/2}\{\boldsymbol{Y}_i^{*\boldsymbol{D}}-\boldsymbol{\mu}_i^{*\boldsymbol{D}}(\boldsymbol{\theta},\breve{\boldsymbol{\gamma}}_{\mathrm{p}})\}=\boldsymbol{0}. \tag{3.11}$$

To establish asymptotic properties for $\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}$, some conditions are needed as follows.

(A.9) The matrix $\boldsymbol{W}_0^{\boldsymbol{D}}$ is sparse with support $\mathcal{S}_{\boldsymbol{W}_0^{\boldsymbol{D}}}=\{l:\boldsymbol{w}_{0l}^{\boldsymbol{D}}\neq\boldsymbol{0},l=1\ldots,p-d\}$, $|\mathcal{S}_{\boldsymbol{W}_0^{\boldsymbol{D}}}|=s_{\boldsymbol{W}^{\boldsymbol{D}}}$ and $\max_{1\le l\le d}|\boldsymbol{w}_{0l}^{\boldsymbol{D}}\boldsymbol{u}|\le C$ for some constant $C>0$.

(A.10) The subsample sizes satisfy $r^{1/2}(r^{\mathrm{p}})^{-1}(s_{\boldsymbol{W}^{\boldsymbol{D}}}\vee s_{\boldsymbol{\beta}})\log p=o(1)$, $(r^{\mathrm{p}})^{-1/2}(s_{\boldsymbol{W}^{\boldsymbol{D}}}\vee s_{\boldsymbol{\beta}})\log p=o(1)$ and $\lambda_3=O(\sqrt{\log p/r^{\mathrm{p}}})$.

Theorem 8. *Under Assumptions (A.2), (A.4)-(A.6) and (A.9)-(A.10), assuming* $\boldsymbol{\Psi}^{\boldsymbol{D}}=E[\hbar_i^{\boldsymbol{D}}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})]$ *is positive-definite, then* $\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\theta}_0=O_P(r^{-1/2})$ *and*

$$[l^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}\boldsymbol{\Omega}^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}]^{-1/2}\{\sqrt{r}(\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}-\boldsymbol{\theta}_0)\}\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{I}_d),$$

where $\boldsymbol{\Omega}^{\boldsymbol{D}}=E[\hbar_i^{\boldsymbol{D}}(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})^{\mathrm{T}}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)\boldsymbol{R}^{-1}\boldsymbol{\Pi}_0\boldsymbol{R}^{-1}\boldsymbol{A}_i^{1/2}(\boldsymbol{\beta}_0)(\boldsymbol{Z}_i-\boldsymbol{U}_i\boldsymbol{W}_0^{\boldsymbol{D}})]$. *Moreover, if* $\boldsymbol{R}=\boldsymbol{\Pi}_0$, *then* $(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}\boldsymbol{\Omega}^{\boldsymbol{D}}(\boldsymbol{\Psi}^{\boldsymbol{D}})^{-1}\le\boldsymbol{\Lambda}^{-1}\boldsymbol{\Upsilon}^{\boldsymbol{D}}\boldsymbol{\Lambda}^{-1}$, *where the inequalities are in the Loewner ordering.*

Remark 4. *Our weighted and unweighted subsample estimators* $\breve{\boldsymbol{\theta}}_{w}^{\boldsymbol{D}}$ *and* $\breve{\boldsymbol{\theta}}_{uw}^{\boldsymbol{D}}$ *are also extensions of estimators proposed by Shao et al. (2025) under the HD-scenario when*

$m = 1$. *To sum up, our proposed estimators broaden the scope of Zhang et al. (2021), Wang et al. (2024) and Shao et al. (2025) in two aspects, from non-longitudinal data to longitudinal data, and from sampling with replacement to Poisson sampling.*

To conduct statistical inference, we propose to estimate the asymptotic covariance matrix of $\sqrt{r}(\check{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw}-\boldsymbol{\theta}_0)$ by $\check{l}^{\boldsymbol{D}}(\check{\boldsymbol{\Psi}}^{\boldsymbol{D}})^{-1}\check{\boldsymbol{\Omega}}^{\boldsymbol{D}}(\check{\boldsymbol{\Psi}}^{\boldsymbol{D}})^{-1}$, where $\check{\boldsymbol{\Psi}}^{\boldsymbol{D}} = n^{-1}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}\check{l}^{\boldsymbol{D}}(\boldsymbol{Z}_i^{*\boldsymbol{D}}-\boldsymbol{U}_i^{*\boldsymbol{D}}\check{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\check{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\check{\gamma}_{\mathrm{p}})\}^{1/2}\check{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\check{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\check{\gamma}_{\mathrm{p}})\}^{1/2}(\boldsymbol{Z}_i^{*\boldsymbol{D}}-\boldsymbol{U}_i^{*\boldsymbol{D}}\check{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}})$, $\check{\boldsymbol{\Omega}}^{\boldsymbol{D}} = n^{-1}\sum_{i\in\mathcal{I}_{\boldsymbol{D}}}\check{l}^{\boldsymbol{D}}(\boldsymbol{Z}_i^{*\boldsymbol{D}}-\boldsymbol{U}_i^{*\boldsymbol{D}}\check{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}})^{\mathrm{T}}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\check{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\check{\gamma}_{\mathrm{p}})\}^{1/2}\check{\boldsymbol{R}}_{\mathrm{p}}^{-1}\check{\boldsymbol{\Pi}}_{\mathrm{p}}\check{\boldsymbol{R}}_{\mathrm{p}}^{-1}\{\boldsymbol{A}_i^{*\boldsymbol{D}}(\check{\boldsymbol{\theta}}^{\boldsymbol{D}}_{uw},\check{\gamma}_{\mathrm{p}})\}^{1/2}(\boldsymbol{Z}_i^{*\boldsymbol{D}}-\boldsymbol{U}_i^{*\boldsymbol{D}}\check{\boldsymbol{W}}_{\mathrm{p}}^{\boldsymbol{D}})$.

4. Simulation studies

In this section, we assess the finite-sample performance of our proposed subsample estimators through both linear and logistic regression models in Sections 4.1 and 4.2, respectively. For both models, a full dataset of cluster size $n = 10^6$ with $m = 3$ observations is generated. For a unified presentation, we use the notation $\bar{\boldsymbol{\delta}}$ to denote a generic estimator representing $\hat{\boldsymbol{\beta}}$ under the LD-scenario in Section 2 or $\check{\boldsymbol{\theta}}$ under the HD-scenario in Section 3. Seven candidates for comparison are listed as follows: (a) Our proposed response-free weighted estimators based on A- and L-optimality criteria, respectively, denoted as $\bar{\boldsymbol{\delta}}^{\mathrm{A}}_{\mathrm{w}}$ and $\bar{\boldsymbol{\delta}}^{\mathrm{L}}_{\mathrm{w}}$; (b) Our proposed response-free unweighted estimators using the same subsampling probabilities and subsample as in (a), respectively, denoted as $\bar{\boldsymbol{\delta}}^{\mathrm{A}}_{\mathrm{uw}}$ and $\bar{\boldsymbol{\delta}}^{\mathrm{L}}_{\mathrm{uw}}$; (c) The response-dependent weighted estimators based on A- and L-optimality criteria (Gao et al., 2025), respectively,

denoted as $\bar{\boldsymbol{\delta}}_{\mathrm{d}}^{\mathrm{A}}$ and $\bar{\boldsymbol{\delta}}_{\mathrm{d}}^{\mathrm{L}}$; (d) The uniform subsampling estimator denoted as $\bar{\boldsymbol{\delta}}^{\mathrm{U}}$. We consider $r^{\mathrm{p}} = 400$ and $r = 400, 600, 800, 1000$. All simulation results are based on 500 replications.

4.1 Linear regression

For linear regression, we generate correlated observations in cluster i according to

$$\boldsymbol{Y}_i|\boldsymbol{X}_i \sim N(\alpha_0 + \boldsymbol{Z}_i\boldsymbol{\theta}_0 + \boldsymbol{U}_i\boldsymbol{\gamma}_0,\ \boldsymbol{\Sigma}_{y|\boldsymbol{x}}),\ i = 1, \ldots, n,$$

where $\alpha_0 = 1$, $\boldsymbol{\beta}_0^{\mathrm{T}} = (\boldsymbol{\theta}_0^{\mathrm{T}}, \boldsymbol{\gamma}_0^{\mathrm{T}})^{\mathrm{T}} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^5$ under the LD-scenario, and $\boldsymbol{\theta}_0^{\mathrm{T}} = (1, 1, 1)^{\mathrm{T}} \in \mathbb{R}^3, \boldsymbol{\gamma}_0^{\mathrm{T}} = (1, 1, 0 \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^{600}$ under the HD-scenario. For both scenarios, $\boldsymbol{\Sigma}_{y|\boldsymbol{x}}$ is a 3×3 compound symmetry matrix with diagonal elements equal to 1 and off-diagonal elements equal to 0.8. For all i, the three rows of $\boldsymbol{X}_i$ are independent and identically distributed, with each row generated from $t_5(\boldsymbol{0}, \boldsymbol{\Sigma}_x/10)$ with $(\boldsymbol{\Sigma}_x)_{l_1,l_2} = 0.5^{|l_1-l_2|}$. Note that based on such settings the CS working structure is correct, while the IND and AR structures are incorrect.

We evaluate the performance of seven candidates listed in (a)-(d) in terms of bias, empirical MSE, standard deviation (SD), estimated standard error (SE) and average coverage probability (ACP) of 95% confidence intervals with respect to the true parameter. Figure 1, Table 1 and Table S1 in the Supplementary Material summarize the detailed results. In terms of MSE, all methods exhibit nearly the same relative performance under different working correlation structures. Specifically, the

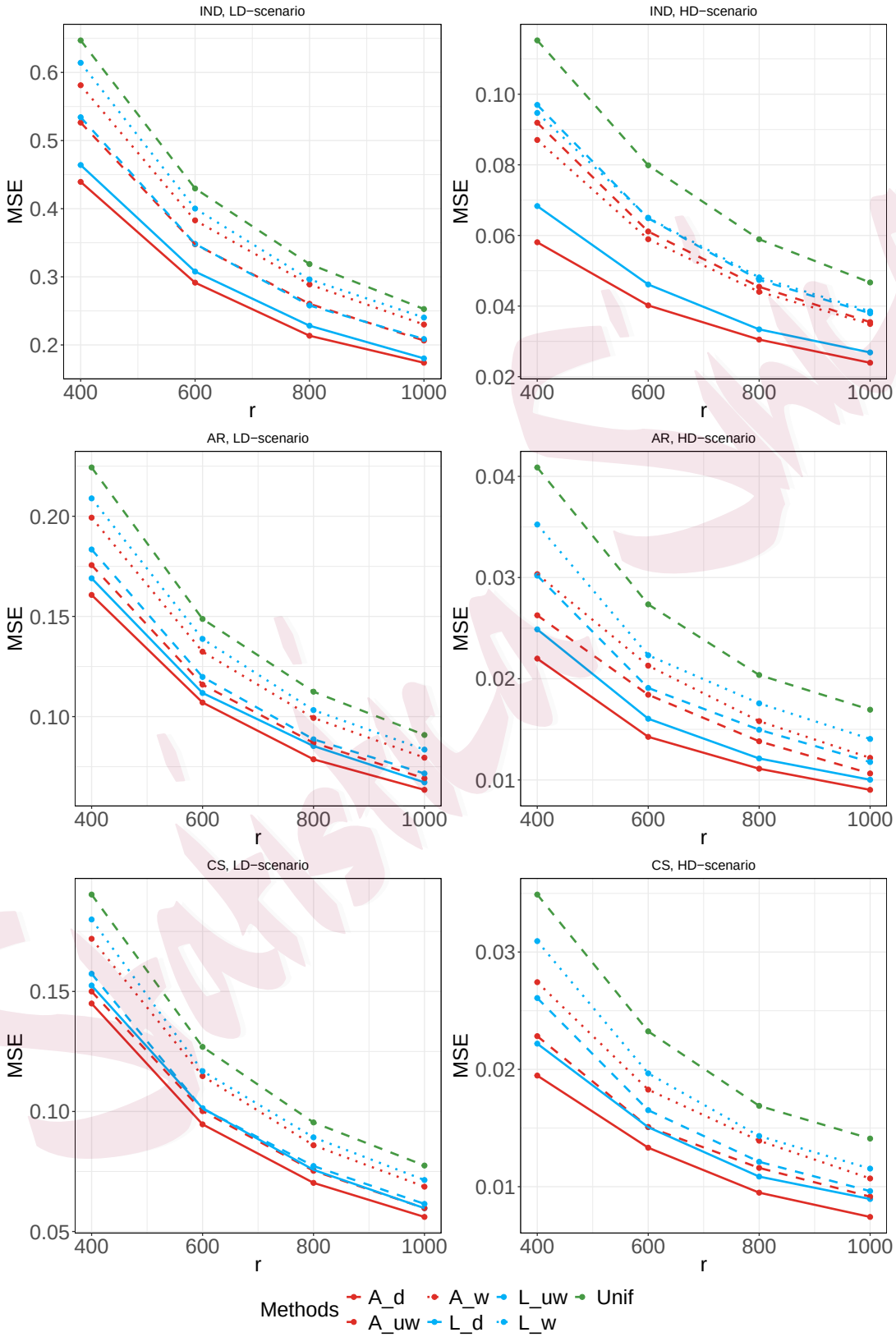

Figure 1: MSEs for linear regression.

Table 1: Bias (×100), SE, SD, and ACP for linear regression under the LD-scenario.

Model	r		$\hat{\boldsymbol{\beta}}^{\mathrm{U}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{d}}^{\mathrm{A}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{d}}^{\mathrm{L}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{w}}^{\mathrm{A}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{w}}^{\mathrm{L}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{uw}}^{\mathrm{A}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{uw}}^{\mathrm{L}}$
IND	400	Bias	5.354	5.081	5.009	5.111	5.563	5.936	5.891
		SE	0.248	0.204	0.210	0.233	0.238	0.225	0.226
		SD	0.242	0.194	0.205	0.220	0.230	0.210	0.215
		ACP	0.942	0.953	0.944	0.958	0.951	0.954	0.956
	600	Bias	4.910	5.275	5.286	4.753	4.915	5.281	5.339
		SE	0.203	0.167	0.171	0.190	0.194	0.183	0.184
		SD	0.192	0.159	0.160	0.180	0.184	0.170	0.171
		ACP	0.951	0.941	0.949	0.953	0.955	0.953	0.958
	800	Bias	4.594	5.147	4.830	4.719	4.807	5.192	5.273
		SE	0.175	0.144	0.149	0.164	0.168	0.158	0.159
		SD	0.164	0.138	0.139	0.158	0.160	0.150	0.147
		ACP	0.947	0.939	0.952	0.938	0.944	0.947	0.946
	1000	Bias	4.496	4.765	4.840	4.535	4.508	5.196	5.113
		SE	0.157	0.129	0.133	0.146	0.150	0.141	0.142
		SD	0.151	0.125	0.125	0.139	0.144	0.128	0.131
		ACP	0.939	0.928	0.938	0.944	0.943	0.946	0.946
AR	400	Bias	3.147	3.098	3.171	3.310	3.767	2.732	2.983
		SE	0.146	0.126	0.130	0.140	0.143	0.132	0.134
		SD	0.141	0.122	0.122	0.136	0.139	0.127	0.130
		ACP	0.948	0.949	0.957	0.949	0.940	0.951	0.950
	600	Bias	3.138	2.672	2.876	2.666	2.500	2.113	2.105
		SE	0.120	0.102	0.105	0.114	0.116	0.107	0.109
		SD	0.117	0.099	0.102	0.108	0.109	0.100	0.103
		ACP	0.946	0.957	0.953	0.952	0.955	0.964	0.951
	800	Bias	2.773	2.459	2.554	2.861	2.981	2.304	2.325
		SE	0.104	0.089	0.091	0.098	0.101	0.093	0.094
		SD	0.097	0.084	0.085	0.091	0.095	0.084	0.088
		ACP	0.958	0.950	0.946	0.949	0.944	0.954	0.951
	1000	Bias	2.736	2.786	2.951	2.756	2.749	2.166	2.092
		SE	0.093	0.079	0.082	0.088	0.090	0.083	0.084
		SD	0.086	0.073	0.074	0.082	0.083	0.075	0.076
		ACP	0.952	0.941	0.940	0.938	0.940	0.953	0.958
CS	400	Bias	2.831	2.813	3.353	3.205	3.157	2.766	2.797
		SE	0.136	0.120	0.123	0.132	0.135	0.124	0.126
		SD	0.130	0.116	0.118	0.126	0.128	0.117	0.119
		ACP	0.957	0.947	0.946	0.955	0.952	0.959	0.959
	600	Bias	3.113	2.636	2.465	3.143	2.982	2.575	2.763
		SE	0.111	0.097	0.100	0.108	0.110	0.101	0.103
		SD	0.109	0.095	0.096	0.101	0.105	0.093	0.098
		ACP	0.934	0.947	0.947	0.943	0.940	0.947	0.941
	800	Bias	2.998	2.825	2.619	3.020	3.204	2.433	2.630
		SE	0.096	0.084	0.086	0.093	0.095	0.087	0.089
		SD	0.091	0.082	0.083	0.085	0.088	0.077	0.081
		ACP	0.949	0.929	0.940	0.949	0.949	0.957	0.947
	1000	Bias	3.022	2.882	2.943	2.881	3.037	2.404	2.484
		SE	0.086	0.075	0.078	0.083	0.085	0.078	0.079
		SD	0.081	0.071	0.071	0.076	0.077	0.068	0.070
		ACP	0.944	0.933	0.949	0.948	0.945	0.965	0.950

uniform subsample estimator $\bar{\boldsymbol{\delta}}^{\mathrm{U}}$ always results in the largest MSEs, while $\bar{\boldsymbol{\delta}}_{\mathrm{d}}^{\mathrm{A}}$ and $\bar{\boldsymbol{\delta}}_{\mathrm{d}}^{\mathrm{L}}$ using optimal response-dependent probabilities always result in the smallest MSEs, since they utilize the information and knowledge from responses and thus lead to superior estimation. However, our proposed optimal response-free subsample estimators also have comparable performances to $\bar{\boldsymbol{\delta}}_{\mathrm{d}}^{\mathrm{A}}$ and $\bar{\boldsymbol{\delta}}_{\mathrm{d}}^{\mathrm{L}}$. Among them, $\bar{\boldsymbol{\delta}}_{\mathrm{w}}^{\mathrm{A}}$ and $\bar{\boldsymbol{\delta}}_{\mathrm{uw}}^{\mathrm{A}}$ have smaller MSEs than $\bar{\boldsymbol{\delta}}_{\mathrm{w}}^{\mathrm{L}}$ and $\bar{\boldsymbol{\delta}}_{\mathrm{uw}}^{\mathrm{L}}$, respectively, and $\bar{\boldsymbol{\delta}}_{\mathrm{uw}}^{\mathrm{A}}$ and $\bar{\boldsymbol{\delta}}_{\mathrm{uw}}^{\mathrm{L}}$ have smaller MSEs than $\bar{\boldsymbol{\delta}}_{\mathrm{w}}^{\mathrm{A}}$ and $\bar{\boldsymbol{\delta}}_{\mathrm{w}}^{\mathrm{L}}$, respectively. These findings align with our theoretical results in Theorems 2, 4, 6 and 8 that A-optimality minimizes the MSEs of subsample estimators and unweighted estimation strategy is more efficient than weighted strategy. It is worth noting that $\bar{\boldsymbol{\delta}}_{\mathrm{d}}^{\mathrm{A}}$ and $\bar{\boldsymbol{\delta}}_{\mathrm{d}}^{\mathrm{L}}$ require access to the response variable for each data point, which is infeasible under measurement constraints. By contrast, our response-free subsampling approach circumvents this limitation and leads to comparable performance. The MSEs of all estimators decline as r increases. Since the CS working structure is correct, all subsample estimators using CS structure have smaller variances than their analogs under wrong working structures. Although the AR structure is incorrect, it still leads to substantially lower variances compared with the IND working structure. This indicates that it is important to model the correlation even though the working structure may not be correct, rather than ignore the correlation entirely. All seven subsample estimators have small biases close to 0 even under the misspecification of working correlation structure. Regarding

ACPs, all seven subsample estimators enjoy results close to 0.95 under both LD- and HD-scenarios. This alignment with the nominal confidence level corroborates the asymptotic normality of the estimators and validates the accuracy of our proposed covariance matrix formulations. For ALs, all optimal subsample estimators result in shorter ALs than $\bar{\boldsymbol{\delta}}^{\mathrm{U}}$. The A-optimal estimators enjoy shorter ALs than L-optimal estimators, and the unweighted estimators have shorter ALs than weighted estimators. As r increases, the ALs of all subsample estimators decrease.

4.2 Logistic regression

For logistic regression, we generate correlated observations in cluster i according to

$$\log\left\{\frac{P(y_{ij}=1|\boldsymbol{x}_{ij})}{1-P(y_{ij}=1|\boldsymbol{x}_{ij})}\right\}=\alpha_0+\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{z}_{ij}+\boldsymbol{\gamma}_0^{\mathrm{T}}\boldsymbol{u}_{ij},\quad i=1,\ldots,n,\ j=1,\ldots,m$$

with $m=3$ correlated binary responses generated by R package `mvtBinaryEP` (Emrich and Piedmonte, 1991) under a correlation matrix whose diagonal elements equal to 1 and off-diagonal elements equal to 0.4. We set $\alpha_0=0.1$, $\boldsymbol{\beta}_0^{\mathrm{T}}=(\boldsymbol{\theta}_0^{\mathrm{T}},\boldsymbol{\gamma}_0^{\mathrm{T}})^{\mathrm{T}}=(0.5,-0.5,0.4,-0.4)^{\mathrm{T}}\in\mathbb{R}^4$ under LD-scenario, and $\boldsymbol{\theta}_0=(0.5,-0.5)^{\mathrm{T}}\in\mathbb{R}^2$, $\boldsymbol{\gamma}_0=(0.4,-0.4,0,\ldots,0)^{\mathrm{T}}\in\mathbb{R}^{600}$ under HD-scenario. For all i, the three rows of $\boldsymbol{X}_i$ are independent and identically distributed, with each row generated from $t_{10}(\mathbf{0},\boldsymbol{\Sigma}_x/10)$. Note that based on such settings the CS working structure is correct, while the IND and AR structures are incorrect. We compare the seven candidates described in Section 4.1 with their performance presented in Figure 2, Table 2 and Table S2 in

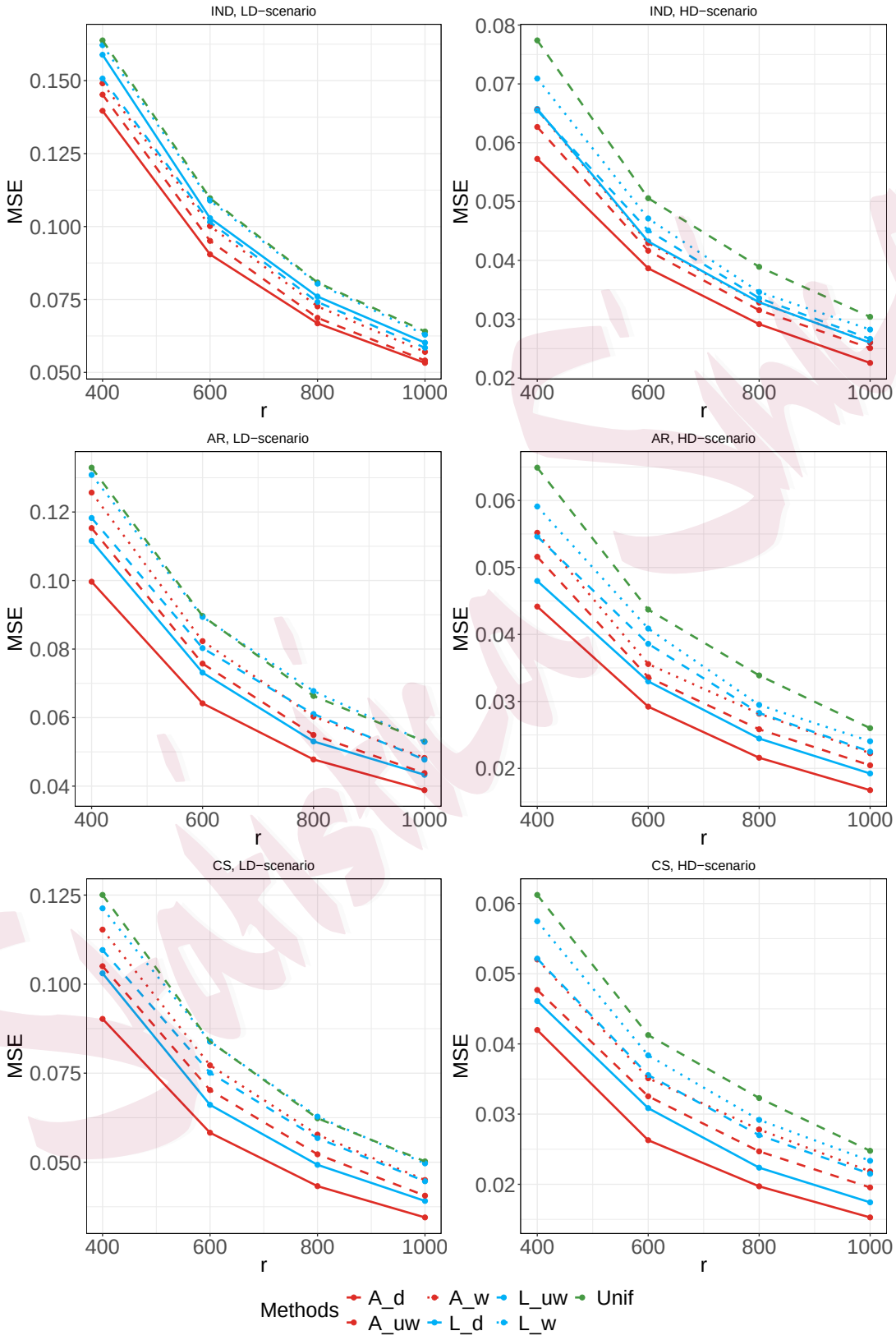

Figure 2: MSEs for logistic regression.

Table 2: Bias (×100), SE, SD, and ACP for logistic regression under the LD-scenario.

Model	r		$\hat{\boldsymbol{\beta}}^{\mathrm{U}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{d}}^{\mathrm{A}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{d}}^{\mathrm{L}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{w}}^{\mathrm{A}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{w}}^{\mathrm{L}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{uw}}^{\mathrm{A}}$	$\hat{\boldsymbol{\beta}}_{\mathrm{uw}}^{\mathrm{L}}$
IND	400	Bias	0.883	0.802	0.841	1.180	1.019	0.813	0.800
		SE	0.203	0.187	0.198	0.193	0.199	0.189	0.193
		SD	0.202	0.187	0.196	0.193	0.201	0.192	0.194
		ACP	0.954	0.952	0.955	0.953	0.946	0.946	0.949
	600	Bias	0.496	0.738	0.695	0.534	0.712	0.715	0.706
		SE	0.166	0.152	0.161	0.157	0.162	0.154	0.157
		SD	0.164	0.150	0.157	0.159	0.163	0.155	0.158
		ACP	0.950	0.948	0.952	0.949	0.948	0.945	0.946
	800	Bias	0.492	0.562	0.732	0.512	0.584	0.734	0.722
		SE	0.143	0.132	0.139	0.136	0.140	0.133	0.136
		SD	0.142	0.129	0.135	0.135	0.141	0.132	0.136
		ACP	0.948	0.950	0.960	0.953	0.951	0.950	0.950
	1000	Bias	0.518	0.490	0.541	0.587	0.682	0.793	0.871
		SE	0.128	0.118	0.125	0.122	0.125	0.119	0.121
		SD	0.125	0.114	0.121	0.118	0.125	0.116	0.120
		ACP	0.955	0.962	0.957	0.955	0.953	0.956	0.951
AR	400	Bias	0.863	0.693	0.633	1.029	0.765	0.860	0.835
		SE	0.183	0.158	0.167	0.178	0.183	0.171	0.176
		SD	0.182	0.157	0.164	0.177	0.181	0.170	0.173
		ACP	0.953	0.956	0.956	0.944	0.952	0.949	0.954
	600	Bias	0.617	0.681	0.476	0.575	0.445	0.621	0.579
		SE	0.149	0.128	0.136	0.145	0.149	0.140	0.143
		SD	0.148	0.126	0.134	0.143	0.147	0.137	0.140
		ACP	0.953	0.950	0.951	0.955	0.954	0.956	0.955
	800	Bias	0.468	0.607	0.496	0.550	0.557	0.701	0.730
		SE	0.129	0.111	0.117	0.125	0.129	0.121	0.124
		SD	0.128	0.109	0.114	0.122	0.129	0.117	0.122
		ACP	0.950	0.950	0.959	0.957	0.952	0.955	0.954
	1000	Bias	0.415	0.583	0.624	0.476	0.558	0.643	0.718
		SE	0.116	0.099	0.105	0.112	0.115	0.108	0.111
		SD	0.113	0.098	0.102	0.108	0.113	0.104	0.107
		ACP	0.955	0.956	0.955	0.953	0.957	0.956	0.956
CS	400	Bias	0.605	0.571	0.585	0.742	0.672	0.847	0.641
		SE	0.178	0.151	0.160	0.172	0.177	0.165	0.170
		SD	0.177	0.148	0.156	0.169	0.174	0.162	0.166
		ACP	0.950	0.956	0.949	0.955	0.957	0.954	0.961
	600	Bias	0.435	0.593	0.541	0.376	0.383	0.558	0.533
		SE	0.145	0.123	0.130	0.141	0.145	0.135	0.139
		SD	0.143	0.121	0.125	0.139	0.141	0.133	0.135
		ACP	0.955	0.951	0.954	0.955	0.961	0.953	0.958
	800	Bias	0.386	0.516	0.482	0.481	0.461	0.670	0.615
		SE	0.126	0.106	0.112	0.122	0.125	0.117	0.120
		SD	0.124	0.104	0.108	0.119	0.122	0.114	0.117
		ACP	0.955	0.955	0.956	0.953	0.958	0.955	0.962
	1000	Bias	0.490	0.550	0.567	0.534	0.612	0.705	0.790
		SE	0.112	0.095	0.101	0.109	0.112	0.104	0.107
		SD	0.109	0.092	0.098	0.105	0.109	0.099	0.103
		ACP	0.961	0.957	0.957	0.954	0.957	0.957	0.960

the Supplementary Material. Similar conclusions can be drawn as those observed in the linear regression model.

5. Beijing air quality dataset

We demonstrate the applicability of our proposed methods to Beijing Multi-Site Air Quality dataset (`https://archive.ics.uci.edu/dataset/501/beijing+multi+site+air+quality+data`), which comprises hourly emission of PM2.5 and meteorological variables from 12 nationally-controlled monitoring sites in Beijing from March 1st, 2013 to February 28th, 2017, amounting to a total of 420,768 observations. To capture the correlation of the emission of PM2.5 and reduce heterogeneity across regions, we segment the 24 daily observations into 6-hour intervals, thereby transforming them into 4 subjects per day. As a result, we have $n = 70128$ subjects with $m = 6$ observations in each cluster. Since the longitudinal emission of PM2.5 is highly correlated, it is necessary to handle the correlation structure within these longitudinal data. For $i = 1, \ldots, n$, the response variable $\boldsymbol{Y}_i \in \mathbb{R}^6$ is the concentration of PM2.5. Our target is to investigate the effects of two relevant air pollutants $\boldsymbol{Z}_i \in \mathbb{R}^{6\times 2}$, i.e., NO_2 and CO, on PM2.5 $\boldsymbol{Y}_i \in \mathbb{R}^6$. To enhance the interpretability of the model, we also involve 210 nuisance variables $\boldsymbol{U}_i \in \mathbb{R}^{6\times 210}$ including temperature, pressure, dew point temperature, wind speed and their 2nd to 6th order terms and interaction terms. To apply our proposed method, we consider the linear regression

model (3.6) and investigate the effects of $\boldsymbol{Z}_i$ on $\boldsymbol{Y}_i$, denoted as θ_1 and θ_2, under the presence of 210 extraneous covariates $\boldsymbol{U}_i$.

We consider a pilot subsample of expected cluster size $r_{\mathrm{p}} = 100$ and two choices of r =100, 200, respectively. Table 3 presents the average of point estimates and SEs of θ_1 and θ_2 using methods described in Section 4 under CS and AR working structures based on 500 independent samplings. For comparison, we also include full dataset analysis $\check{\boldsymbol{\theta}}_{\mathrm{Full}}$ solved from regular DS equation. From Table 3, it can be seen that the point estimates of all methods under both correlation structures are close to the full dataset estimates. However, our proposed optimal response-free subsample estimators yield the smallest SEs in most cases. To be specific, the SEs of A-optimal estimators are smaller than their L-optimal analogs, and the SEs of unweighted estimates are smaller than their weighted analogs, which coincides with our theoretical results. When the subsample size increases, the SEs of all subsample estimates decrease. Regarding statistical inference, all methods show that at level 0.05, the effects of both $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are significantly positive, aligning with the inference results for $\check{\boldsymbol{\theta}}_{\mathrm{Full}}$.

6. Concluding remarks

For analyzing large-scale longitudinal data under GLMs when the full dataset is unavailable due to measurement constraints, we have developed a response-free cluster

Table 3: Point estimates and standard errors for NO_2 and CO.

Model	r		$\check{\boldsymbol{\theta}}_{\text{Full}}$	$\check{\boldsymbol{\theta}}^{\text{A}}_{uw}$	$\check{\boldsymbol{\theta}}^{\text{L}}_{uw}$	$\check{\boldsymbol{\theta}}^{\text{A}}_{w}$	$\check{\boldsymbol{\theta}}^{\text{L}}_{w}$	$\check{\boldsymbol{\theta}}^{\text{A}}_{d}$	$\check{\boldsymbol{\theta}}^{\text{L}}_{d}$	$\check{\boldsymbol{\theta}}^{\text{U}}$
AR	100	$\check{\theta}_1$	0.315	0.285	0.298	0.292	0.290	0.293	0.294	0.265
		SE	0.004	0.033	0.035	0.039	0.040	0.054	0.055	0.076
		$\check{\theta}_2$	0.309	0.290	0.282	0.331	0.339	0.333	0.336	0.387
		SE	0.006	0.025	0.030	0.033	0.036	0.045	0.049	0.085
	200	$\check{\theta}_1$	0.315	0.293	0.303	0.300	0.296	0.293	0.293	0.275
		SE	0.004	0.023	0.024	0.028	0.029	0.037	0.038	0.062
		$\check{\theta}_2$	0.309	0.286	0.275	0.329	0.334	0.337	0.335	0.379
		SE	0.006	0.018	0.020	0.024	0.025	0.032	0.034	0.068
CS	100	$\check{\theta}_1$	0.298	0.275	0.260	0.272	0.267	0.272	0.272	0.255
		SE	0.003	0.032	0.034	0.038	0.039	0.052	0.053	0.075
		$\check{\theta}_2$	0.357	0.336	0.360	0.368	0.378	0.359	0.361	0.403
		SE	0.005	0.030	0.031	0.036	0.037	0.048	0.052	0.088
	200	$\check{\theta}_1$	0.298	0.272	0.261	0.269	0.268	0.266	0.268	0.258
		SE	0.003	0.023	0.023	0.026	0.027	0.036	0.037	0.057
		$\check{\theta}_2$	0.357	0.329	0.351	0.363	0.369	0.366	0.368	0.396
		SE	0.005	0.020	0.022	0.025	0.026	0.034	0.036	0.067

subsampling approach and a general optimality subsampling criterion that incorporates within-subject correlation. Extensive simulations and a real-data application demonstrate the superior statistical and computational performance of the proposed method in both LD- and HD-scenarios.

Several further problems warrant investigation. First, the subsample size allocation between the two steps affects the accuracy of the initial pilot estimates and the precision of the final estimator, thereby impacting the overall estimation efficiency under a fixed budget. However, a systematic investigation into determining the optimal subsample size allocation between these two steps is still lacking. Second, while $\boldsymbol{R} = \boldsymbol{\Pi}_0$ is a sufficient condition for efficiency gains from unweighted

estimation, it is important to identify alternative, easily verifiable conditions that still guarantee efficiency improvements even when $\boldsymbol{R} \neq \boldsymbol{\Pi}_0$. Because perfectly specifying the true correlation structure $\boldsymbol{\Pi}_0$ is often difficult in practice, discovering relaxed conditions would significantly enhance the robustness and practical utility of our estimators. Third, with nonuniform subsamples, widely used augmented inverse probability weighting (Han, 2014; Qin et al., 2017) is effective in improving estimation efficiency; empirical likelihood weighting techniques (Fan et al., 2026) may also yield higher estimation efficiency than standard inverse probability weighting. Effectively integrating these advanced methods into our response-free cluster subsampling framework warrants further investigation. Addressing this requires carefully incorporating within-cluster correlation structures alongside these techniques to achieve higher estimation efficiency.

Acknowledgements

We would like to extend our sincere gratitude to the Editor, an Associate Editor and two anonymous referees for their insightful comments and constructive suggestions, which have significantly enhanced the quality of this paper. Lei Wang was supported by the National Natural Science Foundation of China (Grant No. 12271272). HaiYing Wang was supported by the NSF (Grant No. 2105571) and UConn CLAS Research Funding in Academic Themes. The corresponding author is Lei Wang.

References

Ai, M., J. Yu, H. Zhang, and H. Wang (2021). Optimal subsampling algorithms for big data regressions. *Statistica Sinica 31*(2), 749–772.

Emrich, L. J. and M. R. Piedmonte (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician 45*(4), 302–304.

Fan, Y., Y. Liu, Y. Liu, and J. Qin (2026). Nearly optimal two-step poisson sampling and empirical likelihood weighting estimation for M-estimation with big data. *Statistica Sinica 36*(3), 1–20.

Fang, E. X., Y. Ning, and R. Li (2020). Test of significance for high-dimensional longitudinal data. *The Annals of Statistics 48*(5), 2622–2645.

Gao, J., L. Wang, and J. Shao (2025). Distributed subsampling and quasi decorrelated score for cluster data: An application to beijing multisite air quality. *The Annals of Applied Statistics 19*(3), 1967–1987.

Hamidieh, K. (2018). A data-driven statistical model for predicting the critical temperature of a superconductor. *Computational Materials Science 154*, 346–354.

Han, P. (2014). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association 109*(507), 1159–1173.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Ma, P., Y. Chen, X. Zhang, X. Xing, J. Ma, and M. W. Mahoney (2022). Asymptotic analysis of sampling estimators for randomized numerical linear algebra algorithms. *Journal of Machine Learning Research 23*(177), 1–45.

Ma, P., M. W. Mahoney, and B. Yu (2015). A statistical perspective on algorithmic leveraging. *Journal of Machine Learning Research 16*(1), 861–911.

Ning, Y. and H. Liu (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics 45*(1), 158–195.

Obozinski, G., M. J. Wainwright, and M. I. Jordan (2011). Support union recovery in high-dimensional multivariate regression. *The Annals of Statistics 39*(1), 1–47.

Qin, J., B. Zhang, and D. H. Leung (2017). Efficient augmented inverse probability weighted estimation in missing data problems. *Journal of Business & Economic Statistics 35*(1), 86–97.

Raskutti, G., M. J. Wainwright, and B. Yu (2010). Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research 11*(78), 2241–2259.

Shao, Y., L. Wang, and H. Lian (2025). Optimal decorrelated score subsampling for high-dimensional generalized linear models under measurement constraints. *Journal of Computational and Graphical Statistics 34*(2), 530–539.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of*

the Royal Statistical Society: Series B (Methodological) 58(1), 267–288.

Wang, H. and J. K. Kim (2022). Maximum sampled conditional likelihood for informative subsampling. *Journal of Machine Learning Research 23*(332), 1–50.

Wang, H. and Y. Ma (2021). Optimal subsampling for quantile regression in big data. *Biometrika 108*(1), 99–112.

Wang, H., R. Zhu, and P. Ma (2018). Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association 113*(522), 829–844.

Wang, J., H. Wang, and S. Xiong (2024). Unweighted estimation based on optimal sample under measurement constraints. *Canadian Journal of Statistics 52*(1), 291–309.

Wang, J., H. Wang, and H. H. Zhang (2024). Scale-invariant optimal sampling for rare-events data and sparse models. *Advances in neural information processing systems 37*, 98384–98418.

Wang, J., J. Zou, and H. Wang (2022). Sampling with replacement vs Poisson sampling: a comparative study in optimal subsampling. *IEEE Transactions on Information Theory 68*(10), 6605–6630.

Wang, L. (2011). GEE analysis of clustered binary data with diverging number of covariates. *The Annals of Statistics 39*(1), 389–417.

Wang, L., J. Zhou, and A. Qu (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics 68*(2), 353–360.

Wang, Y., A. W. Yu, and A. Singh (2017). On computationally tractable selection of experiments in measurement-constrained regression models. *Journal of Machine Learning Research 18*(143), 1–41.

Wang, Z., H. Wang, and N. Ravishanker (2023). Subsampling in longitudinal models. *Methodology and Computing in Applied Probability 25*(1), 1–29.

Xie, R., T. Sriram, W. B. Wu, and P. Ma (2025). Online sequential leveraging sampling method for streaming autoregressive time series with application to seismic data. *The Annals of Applied Statistics 19*(4), 3330–3350.

Xie, R., Z. Wang, S. Bai, P. Ma, and W. Zhong (2019). Online decentralized leverage score sampling for streaming multidimensional time series. *Proceedings of Machine Learning Research 89*, 2301–2311.

Yu, J., H. Wang, M. Ai, and H. Zhang (2022). Optimal distributed subsampling for maximum quasi-likelihood estimators with massive data. *Journal of the American Statistical Association 117*(537), 265–276.

Yu, J., Z. Ye, M. Ai, and P. Ma (2025). Optimal subsampling for data streams with measurement constrained categorical responses. *Journal of Computational and Graphical Statistics 34*(3), 994–1004.

Zhang, H. and H. Wang (2026). Refitted cross-validation estimation for high-dimensional subsamples from low-dimension full data. *Computational Statistics 41*(2), 1–15.

Zhang, J., C. Meng, J. Yu, M. Zhang, W. Zhong, and P. Ma (2023). An optimal transport approach for selecting a representative subsample with application in efficient kernel density estimation. *Journal of Computational and Graphical Statistics 32*(1), 329–339.

Zhang, T., Y. Ning, and D. Ruppert (2021). Optimal sampling for generalized linear models under measurement constraints. *Journal of Computational and Graphical Statistics 30*(1), 106–114.

Supplementary Material

The Supplementary Material contains the unbalanced cluster-size scenario, proofs of theorems and additional simulation results.

Junhao Shan School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai University, Tianjin, China. E-mail: shan_junhao@126.com

Lei Wang School of Statistics and Data Science, KLMDASR, LEBPS and LPMC, Nankai University, Tianjin, China. E-mail: lwangstat@nankai.edu.cn

Haiying Wang Department of Statistics, University of Connecticut, U.S.A. E-mail: haiying.wang@uconn.edu