Supplementary Material of "Feature Screening for Massive Data Analysis by Subsampling"

A Extensions and Discussions

A.1 Extension to ISIS Method

Strong correlation among predictors typically exists especially for ultrahigh linear regression models. Regarding to this issue, a common practice is to develop iterative sure independence screening (ISIS) procedures (Fan and Lv, 2008; Cho and Fryzlewicz, 2012). We would like to remark that both DAS and AMS methods are flexible to extend to ISIS procedure.

Specifically, with massive datasets, we could implement the ISIS procedure (Fan and Lv, 2008). as follows. In the first step, we select a subset of k_1 predictors $\mathbb{A}_1 = \{X_{i_1}, \cdots, X_{i_{k_1}}\}$ by SIS-ALasso method. To be specific, we first screening a set of $[n/\log(n)]$ predictors and then use distributed adaptive Lasso (DAL) algorithm recently developed by Zhu et al. (2021) to select the subset $\mathcal{M}_1$. We remark that we do not rely on a distributed architecture to implement the DAL algorithm since it can be applied sequentially to pre-splitted data segments. Then we obtain a residual vector of length N by regressing Y on variables in $\mathcal{M}_1$. Subsequently, we treat the residual vector as the new response and repeat the above step to obtain subset $\mathcal{M}_2$. As commented in Fan and Lv (2008) and Cho and Fryzlewicz (2012), fitting the residuals from the previous step on $\mathcal{M}_F \backslash \mathcal{M}_1$ can significantly weaken the priority of the unimportant variables which are highly correlated with the response by associating with variables in $\mathcal{M}_1$. Here $\mathcal{M}_F$ denotes the full model. In addition, it makes it easy to pick up those important variables which are missed in the previous round. We repeat the above procedure until we obtain l disjoint subsets $\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_l$ with $|\cup_{k=1}^{l} \mathcal{M}_k| \leq d$, where d is a pre-specified integer. In practice we set $d = [n/\log n]$.

Finite Sample Performance

To evaluate the finite sample performance of the ISIS method, we present two examples in this section following Fan and Lv (2008). In each example, the whole sample size N is fixed with $N = 10^5$, and we perform feature screening procedures under the RAS setting. For each model, we apply the SIS and ISIS to select $d = n$ variables. For a reliable evaluation, we replicate the experiment for $R = 100$ times. Denote $\widehat{\mathcal{M}}^{(r)}$ as the selected model and $\mathcal{M}_T^{(r)}$ as the true model in the rth experiment. We evaluate the screening accuracy by the true model covering rate $\mathrm{CR} = R^{-1}\sum_r I(\mathcal{M}_T^{(r)} \subset \widehat{\mathcal{M}}^{(r)})$. The examples are given as follows.

Example A.1. We consider a linear model

$$Y = 5X_1 + 5X_2 + 5X_3 + \epsilon,$$

where $X_1, \ldots, X_p$ are p predictors drawn from a multivariate normal distribution $N(0, \Sigma)$ with $\Sigma = (\sigma_{ij})_{p\times p}$. Here $\sigma_{ii} = 1$, for any $1 \leq i \leq p$, and $\sigma_{ij} = \rho$ for any $i \neq j$. The noise $\epsilon \sim N(0, 1)$ is independent of the predictors. We consider different combination of p, n and B with $p = 200, 1000$, $n = 50, 100$, $B = 10, 50$, and we fix $\rho = 0.9$.

Example A.2. For the second simulated example, the same setup in Example 1 is used except that ρ is fixed to be 0.5. Moreover, a new variable X_4 is added to the model, who has correlation $\sqrt{\rho}$ with the other covariates. Therefore we have

$$Y = 5X_1 + 5X_2 + 5X_3 - 15\sqrt{\rho}X_4 + \epsilon,$$

One could verify that $\mathrm{cov}(X_4, Y) = 0$. As a result, the SIS method is hard to select the true model.

Table A.1: Simulation results for Example A.1 and Example A.2 under the RAS sampling schemes. The numerical performance is evaluated for different parameter dimensions p, number of subsamples B, and subsample sizes n. For each screening measure, CR values under SIS and ISIS methods are reported.

p	B	n	$\mathrm{CR}_{\mathrm{SIS}}$			$\mathrm{CR}_{\mathrm{ISIS}}$		
			AVS	DAS	AMS	AVS	DAS	AMS
Example A.1								
200	10	50	1	1	1	1	1	1
		100	1	1	1	1	1	1
	50	50	1	1	1	1	1	1
		100	1	1	1	1	1	1
1000	10	50	1	1	1	1	1	1
		100	1	1	1	1	1	1
	50	50	1	1	1	1	1	1
		100	1	1	1	1	1	1
Example A.2								
200	10	50	0.24	0.24	0.07	0.9	0.93	0.99
		100	0.57	0.58	0.61	0.98	0.98	1.0
	50	50	0.14	0.14	0.06	1	1	1
		100	0.49	0.47	0.58	1	1	1
1000	10	50	0.0	0.0	0.0	0.52	0.59	0.72
		100	0.02	0.01	0.01	0.72	0.78	0.91
	50	50	0.0	0.0	0.0	1	1	1
		100	0.01	0.01	0.0	1	1	1

We summarize the results in Table A.1. In Example A.1, both SIS and ISIS are able to cover all important variables. However, in Example A.2, the performance of the SIS is much worse than the ISIS method, especially when p is large. For example, when $p = 1000, B = 50, n = 100$, the CR values of the three measures under the SIS are less than 0.01, while they equal to 1 under the ISIS method. That is because the SIS method fails to deal with the correlation among the predictors. In addition, the AMS measure outperforms the other two screening measures especially under the ISIS method. Finally, the accuracy increases when either n or B increases, which is in line with our theoretical findings in Section 3.

A.3 Confidence Adjusted Feature Screening

Motivated by one of anonymous reviewers, we propose a novel feature screening approach, which further involves automatic statistical inference about the feature screening measures. We show that it can enhance the screening accuracy by taking account of uncertainty measures. We refer to this extension as confidence adjusted feature screening method.

Suppose we use the correlation between $\mathbb{X}_j$ and $\mathbb{Y}$ as our screening measure: $\widehat{\rho}_j = \mathbb{X}_j^\top \mathbb{Y}/N$ ($\mathbb{X}_j$ and $\mathbb{Y}$ are standardized). By using the SIS procedure (Fan and Lv, 2008), we keep variables with high $|\widehat{\rho}_j|$ values. However, the SIS procedure does not take account of the uncertainty level of $\widehat{\rho}_j$, i.e., $\mathrm{SE}(\widehat{\rho}_j)$. For variable j with higher $\mathrm{SE}(\widehat{\rho}_j)$, we should have lower confidence in its ranking result. As a result, we could assign lower weight to variables with higher uncertainty levels in the screening procedure. Specifically, we consider to standardize $\widehat{\rho}_j$ by $\mathrm{SE}(\widehat{\rho}_j)$ as

$$\widetilde{\rho}_j = \widehat{\rho}_j/\mathrm{SE}(\widehat{\rho}_j).$$

We refer to $\widetilde{\rho}_j$ as the confidence adjusted screening measure. The $\mathrm{SE}(\widehat{\rho}_j)$ is usually hard to estimate especially for complex screening measures. However, in our DAS setting, we have special opportunity to estimate $\mathrm{SE}(\widehat{\rho}_j)$ by repeated sampling as in (3.1). We conduct simulation studies in the following to illustrate the usefulness of the confidence adjusted screening measure $|\widetilde{\rho}_j|$.

Finite Sample Performance

To evaluate the finite sample performance of the proposed feature screening method, we present a numerical example in this subsection. Specifically, the simulation setting is the same as Example 1 in Section 4.1, except that the distribution of the covariate is not a multivariate normal distribution. Instead, we generate the covariates as follows. For $1 \leq j \leq d_0$, X_{ij}s are independently drawn from Gamma distribution with shape parameter $\alpha = 10$ and rate parameter $\beta = 1$. Next, for $j > d_0$, we generate $X_{ij} = \rho X_{i1} + \sqrt{1-\rho^2} Z_{ij}$, where Z_{ij}s are also independently drawn from Gamma distribution with $\alpha = 10$, $\beta = 1$. Here we fix $\rho = 0.9$ to ensure a relatively high dependence level between the non-important variables and the important variables. Therefore, the case is challenging because (1) the distribution of the covariates is not symmetric and (2) the non-important variables are highly correlated with the important ones. Under this scenario, the $\mathrm{SE}(\widehat{\rho}_j)$s of the non-important variables are higher. Lastly, we fix the whole sample size $N = 10^5$ and $n = 100$, then we consider vary B from 500 to 1000. For a reliable evaluation, the experiment is replicated for $R = 100$ times.

We compare the performance of $\widetilde{\rho}_j$ with $\widehat{\rho}_j$. Specifically $\widetilde{\rho}_j$ is calculated using the DAS method and $\widehat{\rho}_j$ is computed by the DCAMS method demonstrated in subection 3.3. Here the DCAMS refers to using the AMS method under the DC setting. For the DCAMS measure, we consider $nB = 1000, 5000$ and 10^5. Note that when $nB = 10^5 = N$, $\widehat{\rho}_j$ is equivalent to the global screening measure. Furthermore, we use the

AUC measure defined in (4.1) to compare the screening accuracy of the two competing methods. Specifically, for the rth replication, we calculate $\mathrm{AUC}^{(r)}_{\mathrm{DAS}}$ and $\mathrm{AUC}^{(r)}_{\mathrm{DCAMS}}$ respectively for $\widetilde{\rho}_j$ and $\widehat{\rho}_j$. The boxplots of the AUC values are shown in Figure A.1. As one can see from Figure A.1, the average AUC of $\widetilde{\rho}_j$ is obviously higher than the DCAMS measure. This illustrates the potential usefulness of the proposed confidence adjusted feature screening method with our DAS implementation.

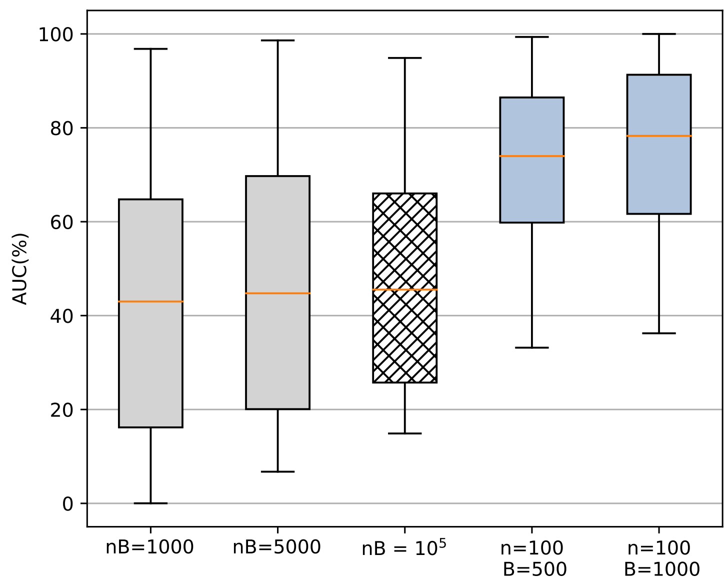

Figure A.1: Boxplot of AUC values for the original screening measure $\widehat{\rho}_j$ conducted by the DCAMS method (the left three boxes), and confidence adjusted screening measure $\widetilde{\rho}_j$ implemented by the DAS method (the right two boxes). Particularly the third box refers to the global estimator when the whole sample is used. The sample size N is fixed as $N = 10^5$.

B Theoretical Properties of AVS Measure

B.1 Uniform Convergence of the AVS Measure under RAS

To establish the uniform convergence property of the AVS measure, we require the following condition.

Assumption B.1 (**Dimensionality for AVS under RAS).**

(a) (Quantitative Covariates) *Let* $\log p_1 \ll \min\{nBN^{-2\nu}, n^{1/2}BN^{-\nu}\}$ *and* $\log p_1 + \log B \ll n^{1/2}$ *for some* $\nu \in [0, 1/2)$ *and* $\delta \in (0, 1/2)$.

(b) (Qualitative Covariates) *Let* $\log p_2 + \max_j \log l_j \ll \min\{nBN^{-2\nu} l_j^{-2}, n^{1/2}BN^{-\nu} l_j^{-1}\}$ *for some* $\nu \in [0, 1/2)$*. In addition, assume* $\log p_2 + \max_j \log l_j + \log B \ll n^{1/2}\pi_{\min}^{1/2}$.

The following theorem establishes the uniform convergence of the AVS measure.

Theorem B.2. *Assume Conditions 1, 2 and B.1, then the following conclusions hold.*

(a) It holds $\max_j |R^2_{\mathbb{X}_j,\mathrm{AVS}} - \mathcal{R}^2_{\mathbb{X}_j} - \Delta_{xb}| = O_p(N^{-\nu})$*, where* $\Delta_{xb} = O(n^{-1})$.

(b) It holds $\max_j |R^2_{\mathbb{Z}_j,\mathrm{AVS}} - \mathcal{R}^2_{\mathbb{Z}_j} - \Delta_{zb}| = O_p(N^{-\nu})$*, where* $\Delta_{zb} = O(n^{-1} l_j)$.

The proof of Theorem B.2 is given in Appendix C.3. With respect to the results, we have the following two remarks.

Remark. By Theorem B.2, the bias order of $R^2_{\mathbb{Z}_j,\mathrm{AVS}}$ is not only related to n, but also related to l_j. As a consequence, it will be larger if the number of levels for a qualitative variable is higher. Hence the subsample size should be set larger if the qualitative covariate of interests has a great number of levels.

B.2 Uniform Convergence for AVS Measure under SAS

To establish the uniform convergence for AVS measure under SAS, we require the following conditions.

Assumption B.3 (**Dimensionality for AVS under SAS**).

(a) (Quantitative Covariates) *There exists* $\delta \in (0, 1/2)$ *such that* $\log p_1 \ll \min\{n^{1-2\delta} BN^{-2\nu}, Bn^{1/2-\delta}N^{-\nu}\}$, $\log p_1 + \log B \ll n^{1/2}$, $\log p_1 + \log N \ll n^{2\delta}$, *where* $\nu \in [0, 1/2)$.

(b) (Qualitative Covariates) *There exists* $\delta \in (0, 1/2)$ *such that* $\log p_2 + \max_j \log l_j \ll \min\{n^{1-2\delta} BN^{-2\nu} l_j^{-2}, n^{1/2-\delta} BN^{-\nu} l_j^{-1}\}$, $\log p_2 + \max_j \log l_j + \log N \ll n^{2\delta}$, $\log p_2 + \max_j \log l_j + \log B \ll n^{1/2} \pi_{\min}^{1/2}$, *where* $\nu \in [0, 1/2)$.

We establish the theoretical properties in the following theorem.

Theorem B.4. *Assume Conditions 1 and B.3, then the following conclusions hold.*

(a) It holds $\max_j |R^2_{\mathbb{X}_j,\mathrm{AVS}} - \mathcal{R}^2_{\mathbb{X}_j} - \Delta_{xb}| = O_p(N^{-\nu})$, *where* $\Delta_{xb} = O(n^{-1})$.

(b) It holds $\max_j |R^2_{\mathbb{Z}_j,\mathrm{AVS}} - \mathcal{R}^2_{\mathbb{Z}_j} - \Delta_{zb}| = O_p(N^{-\nu})$, *where* $\Delta_{zb} = O(n^{-1} l_j)$.

The proof of Theorem B.4 is given in Appendix C.7 and the result is consistent with Theorem B.2 and 2 under the RAS setting.

C Proof of the Main Theorems

Define $E^*(\cdot)$ and $\mathrm{var}^*(\cdot)$ as the conditional expectation and variance given $\mathbb{X}$ and $\mathbb{Y}$.

C.1 Proof of Theorem 1

1. Proof of (a).

We first consider the case that $E(X_{ij}) = 0$. Define $\mathrm{var}(X_{ij}) = \sigma_{xj}^2$, $\sigma_y^2 = \mathrm{var}(Y_i)$, and $\sigma_{xyj} = \mathrm{cov}(X_{ij}, Y_i)$. We first consider the case for the kth subsampling. In this case, we have

$$\begin{aligned}
R^2(\mathbb{X}_{(k)j}) &\stackrel{\text{def}}{=} \frac{1}{n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2} \frac{1}{n^{-1}\|\mathbb{X}_{(k)j}\|^2} \times \left(n^{-1}\mathbb{Y}_{(k)}^\top \mathbb{X}_{(k)j}\right)^2 \\
&= \sigma_{xyj}^2 \Big\{ \frac{1}{n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2} \frac{1}{n^{-1}\|\mathbb{X}_{(k)j}\|^2} - \frac{1}{\sigma_y^2 \sigma_{xj}^2} \Big\} \\
&+ \frac{1}{n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2} \frac{1}{n^{-1}\|\mathbb{X}_{(k)j}\|^2} \times \left\{ \left(n^{-1}\mathbb{Y}_{(k)}^\top \mathbb{X}_{(k)j}\right)^2 - \sigma_{xyj}^2 \right\} + \frac{\sigma_{xyj}^2}{\sigma_{xj}^2 \sigma_y^2} \\
&\stackrel{\text{def}}{=} \Delta_1 + \Delta_2 + \frac{\sigma_{xyj}^2}{\sigma_{xj}^2 \sigma_y^2}
\end{aligned}$$

As a result, we have $P(|R^2(\mathbb{X}_{(k)j}) - \sigma_{xyj}^2/(\sigma_{xj}^2\sigma_y^2)| > t) \le P(|\Delta_1| > t) + P(|\Delta_2| > t)$. Define $\Delta_E(t, n, N)$ as given in Lemma E.2. In the following, we prove that $P(|\Delta_1| > t) \le c\Delta_E(t, n, N)$ and $P(|\Delta_2| > t) \le c\Delta_E(t, n, N)$ respectively in two steps. Given the results, note that $R^2_{\mathbb{X}_j}$ can be treated as independently subsampling for nB times, hence it yields,

$$P\Big\{ \big| R^2_{\mathbb{X}_j} - \mathcal{R}^2_{\mathbb{X}_j} \big| > t \Big\} \le c\Delta_E^*(t, nB, N, n^{1/2}), \tag{C.1}$$

where c is a constant. By using maximum inequality we have $P\{\max_{1\le j\le p_1} |R^2_{\mathbb{X}_j} - \mathcal{R}^2_{\mathbb{X}_j}| > t\} \le cp_1\Delta_E^*(t, nB, N, n^{1/2}) = o(1)$ by Assumption 2 with $t = N^{-\nu}$.

Step 1. Without loss of generality we let $\sigma_{xyj}^2 = 1$ in the following. Note that

$$\begin{aligned}\Delta_1 =& \frac{1}{n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2} \frac{1}{n^{-1}\|\mathbb{X}_{(k)j}\|^2} - \frac{1}{\sigma_y^2\sigma_{xj}^2} \\ =& -\frac{\{(n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2) - \sigma_y^2\}(n^{-1}\|\mathbb{X}_{(k)j}\|^2) + \{(n^{-1}\|\mathbb{X}_{(k)j}\|^2) - \sigma_{xj}^2\}\sigma_y^2}{(n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2)(n^{-1}\|\mathbb{X}_{(k)j}\|^2)(\sigma_y^2\sigma_{xj}^2)}\end{aligned}$$

In the following we show (1) $P\{|n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2 - \sigma_y^2| > t\} \le \Delta_E(t, n, N)$, (2) $P\{|n^{-1}\|\mathbb{X}_{(k)j}\|^2 - \sigma_{xj}^2| > t\} \le \Delta_E(t, n, N)$, and (3) $P\{n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2 > \sigma_y^2/2\} \le \Delta_E(\sigma_y^2/2, n, N)$, $P\{n^{-1}\|\mathbb{X}_{(k)j}\|^2 > \sigma_{xj}^2/2\} \le \Delta_E(\sigma_{xj}^2/2, n, N)$. One could immediately conclude that (3) can be obtained from (1) and (2) by setting $t = \sigma_y^2/2$ and $t = \sigma_{xj}^2/2$ respectively. As a result, we could derive that $P(|\Delta_1| > t) \le c\Delta_E(t, n, N)$ by using (1), (2), (3), where c is a finite constant. In the following we prove conclusion (1) and (2).

Note that $\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2 = n^{-1}\sum_i(\widetilde{Y}_{(k)i} - \mu_y)^2 - (\overline{\mathbb{Y}}_{(k)} - \mu_y)^2$. First note that $(Y_i - \mu_y)^2$ follows sub-Exponential distribution. Then by using Lemma E.2, we have

$$P(|n^{-1}\sum_i(\widetilde{Y}_{(k)i} - \mu_y)^2 - \sigma_y^2| > t) \le \Delta_E(t, n, N). \tag{C.2}$$

In addition, by Lemma E.10, we have $P(|\overline{\mathbb{Y}}_{(k)} - \mu_y|^2 > t) \le \Delta_G(t, n, N)$. Combining the result with (C.2), we have

$$P\Big\{\big|\frac{1}{n}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2 - \sigma_y^2\big| > t\Big\} \le \Delta_G(t, n, N). \tag{C.3}$$

Subsequently, by using the same technique, we could show that $P\{|n^{-1}\|\mathbb{X}_{(k)j}\|^2 - \sigma_{xj}^2| > t\} \le \Delta_E(t, n, N)$.

Step 2. Given the results proved in Step 1, we only need to show $P|\{(n^{-1}\mathbb{Y}_{(k)}^\top\mathbb{X}_{(k)j})^2 - \sigma_{xyj}^2| > t\} < c\Delta_E(t, n, N)$. Recall that X_{ij} and Y_i follow sub-Gaussian distribution.

Consequently, $U_i \stackrel{\text{def}}{=} X_{ij}Y_i$ follows sub-Exponential distribution. Let $\widetilde{U}_i = \widetilde{X}_{(k)ij}\widetilde{Y}_{(k)i}$. By Lemma E.2, $P(|n^{-1}\sum_{i=1}^n \widetilde{U}_i - \sigma_{xyj}| > t) \leq \Delta_E(t,n,N)$. Assuming $t = o(1)$, we could further obtain $P(|(n^{-1}\sum_{i=1}^n \widetilde{U}_i)^2 - \sigma_{xyj}^2| > t) \leq c\Delta_E(t,n,N)$, where c is a finite constant.

2. Proof of (b).

Define $\widehat{\Sigma}_{\mathbb{Z}_j} = (nB)^{-1}\sum_{k=1}^B \mathbb{Z}_{(k)j}^\top \mathbb{Z}_{(k)j} \in \mathbb{R}^{(l_j-1)\times(l_j-1)}$. Since $\mathbb{Z}_j$ is generated from Z_j, which is a qualitative variable. Therefore, $\widehat{\Sigma}_{\mathbb{Z}_j}$ is diagonal and thus we denote $\widehat{\Sigma}_{\mathbb{Z}_j} = \mathrm{diag}\{\widehat{\pi}_{j1},\cdots,\widehat{\pi}_{j(l_j-1)}\} \in \mathbb{R}^{(l_j-1)\times(l_j-1)}$, where $\widehat{\pi}_{jl} = (nB)^{-1}\sum_{k=1}^B \mathbb{Z}_{(k)j,l}^\top \mathbb{Z}_{(k)j,l}$. Accordingly, define $\Sigma_{\mathbb{Z}_j} = E(\widehat{\Sigma}_{\mathbb{Z}_j}) = \mathrm{diag}\{\pi_{j1},\cdots,\pi_{j(l_j-1)}\} \in \mathbb{R}^{(l_j-1)\times(l_j-1)}$. Next, define $\widehat{\Sigma}_{\mathbb{Z}_j\mathbb{Y}} = (nB)^{-1}\sum_{k=1}^B \mathbb{Z}_{(k)j}^\top \mathbb{Y}_{(k)} \in \mathbb{R}^{(l_j-1)}$, $\widehat{\sigma}_{\mathbb{Y}}^2 = (nB)^{-1}\sum_{k=1}^B \|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2$, and $\Sigma_{\mathbb{Z}_j\mathbb{Y}} = (\sigma_{zy,jl} : 1 \leq l \leq l_j - 1)^\top \in \mathbb{R}^{l_j-1}$ with $\sigma_{zy,jl} = E(\mathcal{Z}_{ijl}Y_i)$, $\sigma_y^2 = \mathrm{var}(Y_i)$.

Note that

$$
\begin{aligned}
R_{\mathbb{Z}_j}^2 - \mathcal{R}_{\mathbb{Z}_j}^2 =& \frac{1}{\widehat{\sigma}_{\mathbb{Y}}^2}\Big\{(\widehat{\Sigma}_{\mathbb{Z}_j\mathbb{Y}})^\top(\widehat{\Sigma}_{\mathbb{Z}_j})^{-1}(\widehat{\Sigma}_{\mathbb{Z}_j\mathbb{Y}}) - \Sigma_{\mathbb{Z}_j\mathbb{Y}}^\top\widehat{\Sigma}_{\mathbb{Z}_j}^{-1}\Sigma_{\mathbb{Z}_j\mathbb{Y}}\Big\} \\
&+ \Big\{\frac{1}{\widehat{\sigma}_{\mathbb{Y}}^2}\Sigma_{\mathbb{Z}_j\mathbb{Y}}^\top\widehat{\Sigma}_{\mathbb{Z}_j}^{-1}\Sigma_{\mathbb{Z}_j\mathbb{Y}} - \frac{\Sigma_{\mathbb{Z}_j\mathbb{Y}}^\top\Sigma_{\mathbb{Z}_j}^{-1}\Sigma_{\mathbb{Z}_j\mathbb{Y}}}{\sigma_{\mathbb{Y}}^2}\Big\} \stackrel{\text{def}}{=} \Delta_1 + \Delta_2
\end{aligned}
$$

Then we have $P(|R_{\mathbb{Z}_j}^2 - \mathcal{R}_{\mathbb{Z}_j}^2| > \delta) \leq P(|\Delta_1| > \delta/2) + P(|\Delta_2| > \delta/2)$. Since we have dealt with $\widehat{\sigma}_{\mathbb{Y}}^2$ and $\sigma_{\mathbb{Y}}^2$ in the proof of (a), we treat $\widehat{\sigma}_{\mathbb{Y}}^2 = \sigma_y^2 = 1$ to save space here. We deal with Δ_1 and Δ_2 respectively as follows.

Step 1. Upper bound for $P(|\Delta_1| > \delta/2)$.

First by using the diagonal structure of $\widehat{\Sigma}_{\mathbb{Z}_j}$ we could write

$$\Delta_1 = \widehat{\Sigma}_{\mathbb{Z}_j\mathbb{Y}}^\top \widehat{\Sigma}_{\mathbb{Z}_j}^{-1} \widehat{\Sigma}_{\mathbb{Z}_j\mathbb{Y}} - \Sigma_{\mathbb{Z}_j\mathbb{Y}}^\top \widehat{\Sigma}_{\mathbb{Z}_j}^{-1} \Sigma_{\mathbb{Z}_j\mathbb{Y}} = \sum_{l=1}^{l_j-1} \widehat{\pi}_{jl}^{-1} \Big\{ \big((nB)^{-1} \sum_{k=1}^{B} \mathbb{Z}_{(k)j,l}^\top \mathbb{Y}_{(k)} \big)^2 - \sigma_{zy,jl}^2 \Big\}$$
$$= \sum_{l=1}^{l_j-1} \big(\widehat{\pi}_{jl}^{-1}\pi_{jl}\big) \pi_{jl}^{-1} \Big\{ \big((nB)^{-1} \sum_{k=1}^{B} \mathbb{Z}_{(k)j,l}^\top \mathbb{Y}_{(k)} \big)^2 - \sigma_{zy,jl}^2 \Big\} \stackrel{\text{def}}{=} \sum_{l=1}^{l_j-1} R_{\pi,jl} \Delta_{1l},$$

where $R_{\pi,jl} = \widehat{\pi}_{jl}^{-1}\pi_{jl}$ and $\Delta_{1l} = \pi_{jl}^{-1}\{((nB)^{-1}\sum_{k=1}^{B} \mathbb{Z}_{(k)j,l}^\top \mathbb{Y}_{(k)})^2 - \sigma_{zy,jl}^2\}$. Define the event $\mathcal{E}_{\pi,jl} = \{1/2 < R_{\pi,jl} < 3/2\}$. Then we have

$$P(|\Delta_1| > \delta/2) \leq \sum_{l=1}^{l_j-1} \Big\{ P\Big(|\Delta_{1l}| > \frac{\delta}{3l_j} \Big| \mathcal{E}_{\pi,jl} \Big) + P(\mathcal{E}_{\pi,jl}^c) \Big\} \stackrel{\text{def}}{=} \mathcal{P}_1 + \mathcal{P}_2$$

We give upper bounds for $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively as follows.

Step 1.1. Upper bound for $\mathcal{P}_1$.

Define $\mathcal{D}_{1l} = \pi_{jl}^{-1/2}\big\{(nB)^{-1}\sum_{k=1}^{B} \mathbb{Z}_{(k)j,l}^\top \mathbb{Y}_{(k)} - \sigma_{zy,jl}\big\}$. Then $\Delta_{1l} = 2\mathcal{D}_{1l}\pi_{jl}^{-1/2}\sigma_{zy,jl} + \mathcal{D}_{1l}^2$. We have $\sigma_{zy,jl} = O(\pi_{jl})$. We first derive the bound for $P(|\mathcal{D}_{1l}| > t)$ for any t. First, note that $\mathcal{Z}_{ijl} \in \{0,1\}$ and Y_i follows sub-Gaussian distribution. As a result, the variable $U_i \stackrel{\text{def}}{=} \pi_{jl}^{-1/2}(\mathcal{Z}_{ijl}Y_i - \sigma_{zy,jl})$ follows sub-Exponential distribution. In addition, $E(U_i^2) = \pi_{jl}^{-1}\{E(Z_i)\} = O(1)$. By Lemma E.2, we have $P(|\mathcal{D}_{1l}| > t) \leq \Delta_E^*(t, nB, N, n^{1/2})$. Letting $t = \delta/l_j$ we could obtain $\mathcal{P}_1 \lesssim P(|\mathcal{D}_{1l}| > \delta/l_j) \leq l_j\Delta_E^*(\delta/l_j, nB, N, n^{1/2})$. Then we have $p_2 \max_j l_j \Delta_E^*(N^{-\nu}/l_j, nB, N, n^{1/2}) \to 0$ due to the condition that $\log p_2 + \max_j l_j \ll \min\{nBN^{-2\nu}l_j^{-2}, n^{1/2}BN^{-\nu}l_j^{-1}, n^{1/2}\}$.

Step 1.2. Upper bound for $\mathcal{P}_2$.

Next, we have by Lemma E.9, $P(|\widehat{\pi}_{jl} - \pi_{jl}| > \pi_{jl}/2) \leq \Delta_B(\pi_{jl}/2, nB, N, \pi_{jl})$. As a result, $P(\widehat{\pi}_{jl}^{-1}\pi_{jl} > 3/2) + P(\widehat{\pi}_{jl}^{-1}\pi_{jl} < 1/2) \leq \Delta_B(\pi_{jl}/2, nB, N, \pi_{jl})$. It leads to $\mathcal{P}_2 \leq l_j \max_l \Delta_B(\pi_{jl}/2, nB, N, \pi_{jl})$. By the condition that $\log p_2 + \max \log l_j \ll$

$\min\{\min_{jl} nB\pi_{jl}, N\}$, we have $p_2 \max_{j,l} l_j \Delta_B(\pi_{jl}/2, nB, N, \pi_{jl}) \to 0$.

Step 2. Upper bound for $P(|\Delta_2| > \delta/2)$.

It can be derived that $\Delta_2 = \sum_{l=1}^{l_j-1}(\widehat{\pi}_{jl}^{-1} - \pi_{jl}^{-1})\sigma_{zy,jl}^2$. Hence we have $P(|\Delta_2| > \delta/2) \leq \sum_{l=1}^{l_j-1} P(|\widehat{\pi}_{jl}^{-1} - \pi_{jl}^{-1}| > \delta/(2l_j\sigma_{zy,jl}^2))$. By Lemma E.9, $P(|\widehat{\pi}_{jl} - \pi_{jl}| > t) \leq \Delta_B(t, nB, N, \pi_{jl})$. By the proof in Step 1.2, it leads to $P(|\widehat{\pi}_{jl} - \pi_{jl}| > \pi_{jl}/2) \leq \Delta_B(\pi_{jl}/2, nB, N, \pi_{jl})$. Without loss of generality, we assume $\sigma_{zy,jl}^2 = O(\pi_{jl})$. As a result, we have $P(\widehat{\pi}_{jl}/\sigma_{zy,jl}^2 > 3/2) \leq \Delta_B(\pi_{jl}/2, nB, N, \pi_{jl})$. It further yields $P(|\widehat{\pi}_{jl}^{-1} - \pi_{jl}^{-1}| > \delta/(2l_j\sigma_{zy,jl}^2)) \leq P(|\widehat{\pi}_{jl} - \pi_{jl}| > 3\delta/(4l_j)) + P(\widehat{\pi}_{jl}/\sigma_{zy,jl}^2 > 3/2) \leq \Delta_E^*(3\delta/(4l_j\pi_{jl}^{1/2}), nB, N, n^{1/2}) + \Delta_B(\pi_{jl}/2, nB, N, \pi_{jl})$. One can verify that $p_2 \max_j l_j\{\Delta_E(3\delta/(4l_j\pi_{jl}^{1/2}), nB, N, \pi_{jl}, n^{1/2}) + \Delta_B(\pi_{jl}/2, nB, N, \pi_{jl})\} \to 0$ if the condition $\log p_2 + \max_j l_j \ll nB \min\{N^{-2\nu}l_j^{-2}, N^{-\nu}n^{-1/2}Bl_j^{-1}\}$ and $\log p_1 + \log B \ll n^{1/2}$ hold.

Lastly, letting $\delta = N^{-\nu}$ and under Assumption 2 that we have $\max_j |R_{\mathbb{Z}_j}^2 - \mathcal{R}_{\mathbb{Z}_j}^2| = o_p(N^{-\nu})$.

C.2 Proof of Lemma 1

Without loss of generality we assume $E(X_{ij}) = 0$. Define $\widetilde{D}_i = (\widetilde{X}_{(k)ij}^2, \widetilde{Y}_{(k)i}^2, \widetilde{X}_{(k)ij}\widetilde{Y}_{(k)i})^\top$, $\overline{D} = n^{-1}\sum_{i=1}^n D_i \in \mathbb{R}^3$ and $D = (\sigma_x^2, \sigma_y^2, \sigma_{xy})^\top \in \mathbb{R}^3$. In addition, define $D_i = (X_{ij}^2, Y_i^2, X_{ij}Y_i)^\top$. For the simplicity of the proof here we omit the subindex j and k here. Define $g(x) = x_3^2/(x_1x_2)$ for $x = (x_1, x_2, x_3)^\top$. As a result, we have $R_{\mathbb{X}_{j(k)}}^2 = g(\overline{D})$ and $\mathcal{R}_{\mathbb{X}_j}^2 = g(D)$. We prove the results for bias and variance terms respectively.

1. Proof the bias order.

By using the Taylor's expansion, we have

$$
\begin{aligned}
&g(\overline{D}) - g(D) = \dot{g}(D)^\top(\overline{D} - D) + \frac{1}{2}(\overline{D} - D)^\top \ddot{g}(D)(\overline{D} - D) \\
&+ \frac{1}{6}\sum_{j_1,j_2,j_3} \dddot{g}_{j_1j_2j_3}(D)(\overline{D}^{(j_1)} - D^{(j_1)})(\overline{D}^{(j_2)} - D^{(j_2)})(\overline{D}^{(j_3)} - D^{(j_3)}) \\
&+ \frac{1}{24}\sum_{j_1,j_2,j_3,j_4} \ddddot{g}_{j_1j_2j_3j_4}(\xi)(\overline{D}^{(j_1)} - D^{(j_1)})(\overline{D}^{(j_2)} - D^{(j_2)})(\overline{D}^{(j_3)} - D^{(j_3)})(\overline{D}^{(j_4)} - D^{(j_4)}) \\
&\stackrel{\text{def}}{=} \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4
\end{aligned}
$$

where $\overline{D}^{(j)}$ and $D^{(j)}$ is the jth element of the vector, and ξ is on the line joining D and $\overline{D}$. Note that $E(\Delta_1) = 0$. We deal with Δ_2–Δ_4 respectively as follows.

Step 1. Note that

$$
\Delta_2 = \frac{1}{2n^2}\sum_{i=1}^{n}(\widetilde{D}_i - D)^\top \ddot{g}(D)(\widetilde{D}_i - D) + \frac{1}{2n^2}\sum_{i\neq j}^{n}(\widetilde{D}_i - D)^\top \ddot{g}(D)(\widetilde{D}_j - D) \stackrel{\text{def}}{=} \Delta_{21} + \Delta_{22}
$$

Define $c_d = (2n)^{-1}\mathrm{tr}\{\ddot{g}(D)\Sigma_D\}$. Then we have

$$
\begin{aligned}
E\big(\Delta_{21}\big) &= E\big\{E^*(\Delta_{21})\big\} = \frac{1}{2n}E\big\{(D_i - D)^\top \ddot{g}(D)(D_i - D)\big\} = c_d = O(n^{-1}) \\
E\big(\Delta_{22}\big) &= E\big\{E^*(\Delta_{22})\big\} = \frac{n-1}{2n}E\big\{(\overline{D}_N - D)^\top \ddot{g}(D)(\overline{D}_N - D)\big\},
\end{aligned}
$$

where $\overline{D}_N = N^{-1}\sum_i D_i$. Similarly one can show that $E\big\{(\overline{D}_N - D)^\top \ddot{g}(D)(\overline{D}_N - D)\big\} = O(N^{-1})$. As a result, $E(\Delta_2) = O(n^{-1} + N^{-1})$.

Step 2. Next, we look at Δ_3. Since the dimension of D is finite, we could focus on a single j. It holds

$$
\begin{aligned}
(\overline{D}^{(j)} - D^{(j)})^3 = & \frac{1}{n^3}\sum_{i=1}^{n}(\widetilde{D}_i^{(j)} - D_i^{(j)})^3 + \frac{1}{n^3}\sum_{i_1\neq i_2}(\widetilde{D}_{i_1}^{(j)} - D_{i_1}^{(j)})^2(\widetilde{D}_{i_2}^{(j)} - D_{i_2}^{(j)}) \\
& + \frac{1}{n^3}\sum_{i_1\neq i_2, i_2\neq i_3, i_1\neq i_3}(\widetilde{D}_{i_1}^{(j)} - D_{i_1}^{(j)})(\widetilde{D}_{i_2}^{(j)} - D_{i_2}^{(j)})(\widetilde{D}_{i_3}^{(j)} - D_{i_3}^{(j)}) \\
& \stackrel{\text{def}}{=} \Delta_{D1} + \Delta_{D2} + \Delta_{D3}.
\end{aligned}
$$

It can be derived that $E(\Delta_{D1}) = O(n^{-2})$, $E(\Delta_{D2}) = O((nN)^{-1})$, and $E(\Delta_{D3}) = O(N^{-2})$. As a result, $E(\Delta_3) = O(n^{-2})$.

Step 3. Lastly, we look at the Δ_4. First we write $\Delta_4 = \Delta_{41} + \Delta_{42}$ in two parts as

$$
\begin{aligned}
\Delta_{41} &= \frac{1}{24}\sum_{j_1,j_2,j_3,j_4}\ddddot{g}_{j_1j_2j_3j_4}(D)(\overline{D}^{(j_1)} - D^{(j_1)})(\overline{D}^{(j_2)} - D^{(j_2)})(\overline{D}^{(j_3)} - D^{(j_3)})(\overline{D}^{(j_4)} - D^{(j_4)}) \\
\Delta_{42} &= \frac{1}{24}\sum_{j_1,j_2,j_3,j_4}\{\ddddot{g}_{j_1j_2j_3j_4}(\xi) - \ddddot{g}_{j_1j_2j_3j_4}(D)\}\prod_{m=1}^{4}(\overline{D}^{(j_m)} - D^{(j_m)})
\end{aligned}
$$

It can be verified that $E(\Delta_{41}) = O(n^{-3}) = o(n^{-2})$. As a result, we only need to show that $\Delta_{42} = o_p(n^{-2})$ to imply that Δ_{42} will not produce a bias term which is larger than the order of $O(n^{-2})$.

Since $\|\xi - D\|_{\max} \le \|\overline{D} - D\|_{\max} = o_p(1)$, we could verify that $\|\ddddot{g}(\xi) - \ddddot{g}(D)\|_{\max} \le c\|\xi - D\|_{\max}$ with probability tending to 1, where c is a finite constant. By Lemma E.2, we have $\|\xi - D\|_{\max} = O_p(n^{-1/2})$. As a result, $\Delta_{42} \le c\|\overline{D} - D\|_{\max}^5 = O_p(n^{-5/2}) = o_p(n^{-2})$.

Consequently, we have $\Delta_{xb} = cn^{-1} + \max\{n^{-2}, N^{-1}\}\{1 + o(1)\}$.

2. Proof of the variance part.

To analyze the variance part, we only need to expand $g(\overline{D})$ to the second order as

$$g(\overline{D}) - g(D) = \dot{g}(D)^\top(\overline{D} - D) + \frac{1}{2}(\overline{D} - D)^\top \ddot{g}(\xi)(\overline{D} - D) \stackrel{\text{def}}{=} \Delta_1 + \Delta_2.$$

We further write Δ_2 as

$$\begin{aligned}\Delta_2 =& \frac{1}{2n^2}\sum_{i=1}^{n}(\widetilde{D}_i - D)^\top \ddot{g}(D)(\widetilde{D}_i - D) + \frac{1}{2n^2}\sum_{i\neq j}^{n}(\widetilde{D}_i - D)^\top \ddot{g}(D)(\widetilde{D}_j - D)\\ &+ \frac{1}{2}(\overline{D} - D)^\top\{\ddot{g}(\xi) - \ddot{g}(D)\}(\overline{D} - D) \stackrel{\text{def}}{=} \Delta_{21} + \Delta_{22} + \Delta_{23}.\end{aligned}$$

By Lemma E.2, we have $\Delta_1 = O_p(n^{-1/2})$. Next, by Lemma E.3, we have $\Delta_{21} - E(\Delta_{21}) = O_p(n^{-1})$. By Lemma E.5, $\Delta_{22} - E(\Delta_{22}) = O_p((\log n)^2/n)$. Furthermore, since $\|\xi - D\|_{\max} \leq \|\overline{D} - D\|_{\max} = o_p(1)$, we could verify that $\|\ddot{g}(\xi) - \ddot{g}(D)\|_{\max} \leq c\|\xi - D\|_{\max} \leq c\|\overline{D} - D\|_{\max}$ with probability tending to 1, where c is a finite constant. This implies $\Delta_{23} = o_p(\Delta_{21} + \Delta_{22})$. This yields the final result.

C.3 Proof of Theorem B.2

1. Proof of (a).

Define $D_{(k)i} = (\widetilde{X}^2_{(k)ij}, \widetilde{Y}^2_{(k)i}, \widetilde{X}_{(k)ij}\widetilde{Y}_{(k)i})^\top$, $\overline{D}_{(k)} = n^{-1}\sum_{i=1}^n D_{(k)i} \in \mathbb{R}^3$, $\overline{D} = B^{-1}\sum_{k=1}^B \overline{D}_{(k)} \in \mathbb{R}^3$, and $D = (\sigma_x^2, \sigma_y^2, \sigma_{xy})^\top \in \mathbb{R}^3$. Define $g(x) = x_3^2/(x_1x_2)$ for $x = (x_1, x_2, x_3)^\top$. As a result, we have $R^2_{\mathbb{X}_j,\text{AVS}} = B^{-1}\sum_k g(\overline{D}_{(k)})$ and $\mathcal{R}^2_{\mathbb{X}_j} = g(D)$.

By using the Taylor's expansion, we have

$$B^{-1}\sum_k g(\overline{D}_{(k)}) - g(D) = \frac{1}{B}\sum_k \dot{g}(D)^\top(\overline{D}_{(k)} - D) + \frac{1}{B}\sum_k(\overline{D}_{(k)} - D)^\top \ddot{g}(\xi_{(k)})(\overline{D}_{(k)} - D)$$

$\stackrel{\text{def}}{=} \Delta_1 + \Delta_2$, where $\xi_{(k)}$ is on the line joining D and $\overline{D}_{(k)}$. In the following we use two steps to deal with Δ_1 and Δ_2 respectively.

Step 1. First, by Lemma E.2, we immediately have $P\{|\Delta_1| > t\} \le \Delta_E^*(t, nB, N, n^{1/2})$.

Step 2. First we write Δ_2 as

$$
\begin{aligned}
\Delta_2 &= \frac{1}{B}\sum_k (\overline{D}_{(k)} - D)^\top \ddot{g}(D)(\overline{D}_{(k)} - D) + \frac{1}{B}\sum_k (\overline{D}_{(k)} - D)^\top \{\ddot{g}(\xi_{(k)}) - \ddot{g}(D)\}(\overline{D}_{(k)} - D) \\
&= \frac{1}{Bn^2}\sum_k\sum_i (D_{(k)i} - D)^\top \ddot{g}(D)(D_{(k)i} - D) + \frac{1}{Bn^2}\sum_k\sum_{i\neq j} (D_{(k)i} - D)^\top \ddot{g}(D)(D_{(k)j} - D) \\
&+ \frac{1}{B}\sum_k (\overline{D}_{(k)} - D)^\top \{\ddot{g}(\xi_{(k)}) - \ddot{g}(D)\}(\overline{D}_{(k)} - D) \stackrel{\text{def}}{=} \Delta_{21} + \Delta_{22} + \Delta_{23}.
\end{aligned}
$$

First, we obtain $P(|\Delta_{21} - \Delta_{xb}|) \le \Delta_{E2}(nt, nB, M)$ by Lemma E.3. Next, we obtain $P(|\Delta_{22}| > t) \le \Delta_{E4}^{sas}(nt, n, B, \epsilon)$ by Lemma E.15. Setting $M = n^{1/2}$ and $\epsilon = n^{1/4}$ we have $p_1\Delta_{E2}(nt, nB, M) \to 0$ and $p_1\Delta_{E4}^{sas}(nt, n, B, \epsilon) \to 0$.

Next, for each k if it holds $\|\xi_{(k)} - D\|_{\max} < \|D\|_{\min}/2$, then we have $|\ddot{g}_{j_1j_2}(\xi_{(k)}) - \ddot{g}_{j_1j_2}(D)| \le c\|\xi_{(k)} - D\|_{\max} \le c\|\overline{D}_{(k)} - D\|_{\max}$, where c is a positive constant. As a result, we have $P(\max_{k,j_1,j_2} |\ddot{g}_{j_1j_2}(\xi_{(k)}) - \ddot{g}_{j_1j_2}(D)| > t) \lesssim B\Delta_E(\|\ddot{g}(D)\|_{\min}/2, n, N) + B\Delta_E(t, n, N) \lesssim B\Delta_E(t, n, N)$ for small t. By letting $t = o(1)$ and under event $\{\max_{k,j_1,j_2} |\ddot{g}_{j_1j_2}(\xi_{(k)}) - \ddot{g}_{j_1j_2}(D)| < t\}$, we have $|\Delta_{23}| \ll |\Delta_{21} + \Delta_{22}|$. As a result $P(|\Delta_2 - \Delta_b| > t) \le B\Delta_E(t_1, n, N) + \Delta_{E4}^{sas}(nt, n, B, \epsilon)$ for $t_1 = o(1)$. Setting $\epsilon = n^{1/4}$ we could obtain that $p_1\Delta_{E4}^{sas}(nt, n, B, \epsilon) \to 0$ and $p_1 B\Delta_E(t, n, N) \to 0$. Hence the final conclusion holds.

2. Proof of (b).

Define $D_{(k)i} = (\widetilde{\mathcal{Z}}_{(k)ij1}, \cdots, \widetilde{\mathcal{Z}}_{(k)ij(l_j-1)}, \widetilde{\mathcal{Z}}_{(k)ij1}\widetilde{Y}_{(k)i}, \cdots, \widetilde{\mathcal{Z}}_{(k)ij(l_j-1)}\widetilde{Y}_{(k)i}, (\widetilde{Y}_{(k)i} - \overline{\mathbb{Y}}_{(k)})^2)^\top \in \mathbb{R}^{2(l_j-1)+1}$, $\overline{D}_{(k)} = n^{-1}\sum_{i=1}^n D_{(k)i} \in \mathbb{R}^{2(l_j-1)+1}$, $\overline{D} = B^{-1}\sum_{k=1}^B \overline{D}_{(k)} \in \mathbb{R}^{2(l_j-1)+1}$, and

$D = (\pi_{j1}, \cdots, \pi_{j(l_j-1)}, \sigma_{zy,j1}, \cdots, \sigma_{zy,j(l_j-1)}, \sigma_y^2)^\top \in \mathbb{R}^{2(l_j-1)+1}$. In addition, define

$$g(D) = D_{2l_j-1}^{-1} \sum_{l=1}^{l_j-1} D_l^{-1} D_{l_j+l-1}^2$$

As a result, we have $R^2_{\mathbb{Z}_j,\mathrm{AVS}} = B^{-1}\sum_k g(\overline{D}_{(k)})$ and $\mathcal{R}^2_{\mathbb{Z}_j} = g(D)$.

By using the Taylor's expansion, we have

$$B^{-1}\sum_k g(\overline{D}_{(k)}) - g(D) = \frac{1}{B}\sum_k \dot{g}(D)^\top(\overline{D}_{(k)} - D) + \frac{1}{B}\sum_k (\overline{D}_{(k)} - D)^\top \ddot{g}(\xi_{(k)})(\overline{D}_{(k)} - D)$$

$\stackrel{\text{def}}{=} \Delta_1 + \Delta_2$, where $\xi_{(k)}$ is on the line joining D and $\overline{D}_{(k)}$.

Step 1. $(P(|\Delta_1| > \epsilon))$

Note that $\dot{g}(D) = (-D_1^{-2} D_{l_j}^2 D_{2l_j-1}^{-1}, \cdots, -D_{l_j-1}^{-2} D_{l_j}^2 D_{2l_j-1}^{-1}, 2D_1^{-1} D_{2l_j-1}^{-1} D_{l_j}, \cdots, D_{l_j-1}^{-1} D_{2l_j-1}^{-1} D_{2l_j-2}, -D_{2l_j-1}^{-2}\sum_{l=1}^{l_j-1} D_l^{-1} D_{l_j+l-1}^2)^\top$. Since $D_l = O(\pi_{jl})$ and $D_{l+l_j-1} = O(\pi_{jl})$ for $1 \le l \le l_j - 1$, we have $\|g(D)\|_{\max} = O(1)$. Write $\Delta_1 = \Delta_{11} + \Delta_{12} + \Delta_{13}$, where

$$\Delta_{11} = -\frac{1}{B}\sum_k \sum_{l=1}^{l_j-1} \dot{g}_l(D)\big(\overline{D}_{(k)l} - D\big)$$
$$\Delta_{12} = \frac{1}{B}\sum_k \sum_{l=l_j}^{2l_j-2} \dot{g}_l(D)\big(\overline{D}_{(k)l} - D\big)$$
$$\Delta_{13} = \frac{1}{B}\dot{g}_{2l_j-1}(D)\big(\overline{D}_{(k)2l_j-1} - D\big),$$

where $\dot{g}_l(D)$ and $\overline{D}_{(k)l}$ is the lth element of vector $\dot{g}(D)$ and $\overline{D}_{(k)}$. By Lemma E.2 and E.9 we have $P(|\Delta_{11}| > \epsilon/3) \lesssim l_j \Delta_B(\epsilon/l_j, nB, N, \pi_{\min})$, $P(|\Delta_{12}| > \epsilon/3) \lesssim l_j \Delta_E(\epsilon/l_j, nB, N)$, and $P(|\Delta_{13}| > \epsilon/3) \lesssim \Delta_E(\epsilon, nB, N)$. As a result, $p_2 P(|\Delta_1| > \epsilon) \le p_2 \max_j l_j \Delta_E(\epsilon/l_j, nB, N) \to 0$ with $\epsilon = O(N^{-\nu})$ by Assumption 2 (b) and B.1 (b).

Step 2. $(P(|\Delta_2 - \Delta_{zb}| > \epsilon))$

Next, we have $\Delta_2 = \Delta_{21} + \Delta_{22}$, where $\Delta_{21} = B^{-1}\sum_k(\overline{D}_{(k)} - D)^\top \ddot{g}(D)(\overline{D}_{(k)} - D) + B^{-1}\sum_k(\overline{D}_{(k)} - D)^\top\{\ddot{g}(\xi_{(k)}) - \ddot{g}(D)\}(\overline{D}_{(k)} - D)$.

Step 2.1. $(P(|\Delta_{21} - \Delta_{zb}| > \epsilon))$

It can be verified that $\ddot{g}(D) = (G_1, G_2; G_3, G_4)$, where $G_1 = (\mathbf{1}_2\mathbf{1}_2^\top)\times\mathrm{diag}\{\pi_{j1}^{-1}, \cdots, \pi_{j(l_j-1)}^{-1}\}$, $G_2 = \mathbf{1}_{2l_j-2}$, $G_3 = G_2^\top$, $G_4 = 1$. Then it can be derived that $\Delta_{21} = \Delta_{21}^{(1)} + \Delta_{21}^{(2)} + \Delta_{21}^{(3)} + \Delta_{21}^{(4)}$,

$$\Delta_{21}^{(1)} = c\sum_{l=1}^{l_j-1}\pi_{jl}^{-1}(\overline{D}_{(k)l} - D_l)^2, \tag{C.4}$$

$$\Delta_{21}^{(2)} = c\sum_{l=l_j}^{2l_j-2}\pi_{jl}^{-1}(\overline{D}_{(k)l} - D_l)^2, \tag{C.5}$$

$$\Delta_{21}^{(3)} = c\sum_{l=1}^{l_j-1}\pi_{jl}^{-1}(\overline{D}_{(k)l} - D_l)(\overline{D}_{(k)(l+l_j)} - D_{l+l_j}) \tag{C.6}$$

$$\Delta_{21}^{(4)} = c\sum_{l=1}^{2l_j-2}\pi_{jl}^{-1}(\overline{D}_{(k)l} - D_l)(\overline{D}_{(k)2l_j-1} - D_{2l_j-1}). \tag{C.7}$$

where c is a constant.

One could verify that $\Delta_{zb} = E(\Delta_{21}) = O(n^{-1}l_j)$ as in the proof of Lemma 1. Next we focus on the tail bound of the four parts. Since the proof procedures are similar, we prove the case for $\Delta_{21}^{(1)}$. First, for we could write $\pi_l^{-1}(\overline{D}_{(k)l} - D_l)^2 = n^{-2}\sum_i \pi_l^{-1}(D_{(k)l,i} - D_{l,i})^2 + n^{-2}\sum_{i\neq j}\pi_l^{-1}(D_{(k)l,i} - D_{l,i})(D_{(k)l,j} - D_{l,j}) \overset{\text{def}}{=} \Delta_{D1} + \Delta_{D2}$. By Lemma E.3 and E.7, we have $P(|\Delta_{D1} - E(\Delta_{D1})| > \epsilon) \leq \Delta_{E2}(n\epsilon/l_j, nB, M)$ and $P(|\Delta_{D2})| > \epsilon) \leq \Delta_{E4}^{sas}(n\epsilon/l_j, n, nB, \epsilon)$. Setting $M = n^{1/2}$ and $\epsilon = n^{1/4}$ we have $p_2\max_j l_j\Delta_{E2}(n\epsilon/l_j, nB, M) \to 0$ and $p_2\max_j l_j\Delta_{E4}^{sas}(n\epsilon/l_j, n, nB, \epsilon) \to 0$ by the condition that $\log p_2 + \max_j l_j \ll \min\{nBN^{-2\nu}l_j^{-2},\ n^{1/2}BN^{-\nu}l_j^{-1}\} \to 0$ with $\epsilon = N^{-\nu}$ and $\log p_2 + \max_j \log l_j + \log B \ll n^{1/2}$.

Step 2.2. $(P(|\Delta_{22}| > \epsilon))$

Next, we focus on Δ_{22}. As we have shown in (C.4)–(C.7), we could also decompose Δ_{22} into four parts as $\Delta_{22}^{(1)}$ to $\Delta_{22}^{(4)}$. We prove the case for the first part for illustration. Note that $\ddot{g}_{ll}(D) = 2D_l^{-3}D_{l_j+l}^2 D_{2l_j-1}^{-1}$ for $1 \le l \le l_j - 1$ and $\ddot{g}_{l_1l_2}(D) = 0$ for $l_1 \neq l_2$. Under the event that $\{\max_k \|\xi_{(k)} - D\|_{\max} < \epsilon\}$ with ϵ small enough, it can be derived that $|\ddot{g}_{ll}(\xi_{(k)}) - \ddot{g}_{ll}(D)| \le c\pi_{jl}^{-2}\big\{|\widehat{\pi}_{jl(k)} - \pi_{jl}| + |\widehat{\sigma}_{zy,(k)jl} - \sigma_{zy,jl}|\big\} + \pi_{jl}^{-1}|\widehat{\sigma}_{y(k)}^2 - \sigma_y^2|$, where $\widehat{\pi}_{(k)l} = n^{-1}\sum_i \widetilde{\mathcal{Z}}_{(k)ijl}$, $\widehat{\sigma}_{zy,(k)jl} = n^{-1}\sum_i \widetilde{Y}_{(k)i}\widetilde{\mathcal{Z}}_{(k)ijl}$, and $\widehat{\sigma}_{y(k)}^2 = n^{-1}\|\mathbb{Y}_{(k)} - \overline{\mathbb{Y}}_{(k)}\|^2$. As a result, $P(\max_k\{|\ddot{g}_{ll}(\xi_{(k)}) - \ddot{g}_{ll}(D)| > t/\pi_{jl}\}) \le P(\pi_{jl}^{-1}\max_k|\widehat{\pi}_{(k)jl} - \pi_{jl}| > t) + P(\pi_{jl}^{-1}\max_k|\widehat{\sigma}_{zy,(k)jl} - \sigma_{zy,jl}| > t) + P(\max_k|\widehat{\sigma}_{y(k)}^2 - \sigma_y^2| > t) \lesssim B\Delta_E(t\pi_{\min}^{1/2}, n, N) \to 0$ by the condition that $\log p_2 + \max_j \log l_j + \log B \ll n^{1/2}\pi_{\min}^{1/2}$, where $\pi_{\min} = \min_{j,l}\pi_{jl}$. Consequently, we have $\max_k\{|\ddot{g}_{ll}(\xi_{(k)}) - \ddot{g}_{ll}(D)| = o_p(\pi_{jl}^{-1})$. Consequently, Δ_{22} is dominated by Δ_{21}, which implies $P(|\Delta_{22}| > \epsilon) \lesssim P(|\Delta_{21}| > \epsilon)$. This yields the final result.

C.4 Proof of Lemma 2

Without loss of generality we assume $E(X_{ij}) = 0$. Define $D_i = (\widetilde{X}_{(k)ij}^2, \widetilde{Y}_{(k)i}^2, \widetilde{X}_{(k)ij}\widetilde{Y}_{(k)i})^\top$, $\overline{D} = n^{-1}\sum_{i=1}^n D_i \in \mathbb{R}^3$ and $D = (\sigma_x^2, \sigma_y^2, \sigma_{xy})^\top \in \mathbb{R}^3$. In addition, let $\overline{D}_{-i} = (n-1)^{-1}\sum_{m\neq i} D_m \in \mathbb{R}^3$. For convenience we omit the sub-index k and j here. Define $g(x) = x_3^2/(x_1x_2)$ for $x = (x_1, x_2, x_3)^\top$. As a result, we have $R^2_{\mathbb{X}_j(k),\mathrm{DAS}} = g(\overline{D})$ and $\mathcal{R}^2_{\mathbb{X}_j} = g(D)$. Without of loss of generality we assume $n^2 \le N$. Since all variance terms are proved in Lemma 1, we only need to focus on the bias estimator $\Delta_{(k)}$.

By using the Taylor's expansion, we have

$$
\begin{aligned}
g(\overline{D}_{-i}) - g(\overline{D}) &= \dot{g}(\overline{D})^\top(\overline{D}_{-i} - \overline{D}) + 2^{-1}(\overline{D}_{-i} - \overline{D})^\top\ddot{g}(\xi_i)(\overline{D}_{-i} - \overline{D}) \\
&\overset{\text{def}}{=} \dot{g}(\overline{D})^\top(\overline{D}_{-i} - \overline{D}) + \Delta_i,
\end{aligned}
$$

where ξ_i lies on the line joining $\overline{D}$ and $\overline{D}_{-i}$. Note that for the first term we have $\sum_i(\overline{D}_{-i} - \overline{D}) = (n-1)^{-1}\sum_i(\overline{D} - D_i) = \mathbf{0}$. Write $\Delta_i = 2^{-1}(\overline{D}_{-i} - \overline{D})^\top \ddot{g}(\overline{D})(\overline{D}_{-i} - \overline{D}) + 2^{-1}(\overline{D}_{-i} - \overline{D})^\top\{\ddot{g}(\xi_i) - \ddot{g}(\overline{D})\}(\overline{D}_{-i} - \overline{D}) \stackrel{\text{def}}{=} \Delta_i^a + \Delta_i^b$. Then we have

$$\Delta_{(k)} = \frac{n-1}{n}\sum_i \Delta_i^a + \frac{n-1}{n}\sum_i \Delta_i^b \stackrel{\text{def}}{=} \Delta_{(k)}^a + \Delta_{(k)}^b.$$

Define $V = \text{tr}\{\ddot{g}(D)\text{cov}(D_i)\} \stackrel{\text{def}}{=} \text{tr}\{\ddot{g}(D)\Sigma_D\}$. Then $\Delta_{xb} = (2n)^{-1}V$. In the following we deal with the two parts respectively.

Step 1. (Bias and Variance of $\Delta_{(k)}^a - \Delta_{xb}$)

First note that $\overline{D}_{-i} - \overline{D} = (n-1)^{-1}(\overline{D} - D_i)$. Then it yields

$$\frac{n-1}{n}\sum_i \Delta_i^a = \frac{1}{n(n-1)}\sum_i (D_i - \overline{D})^\top \ddot{g}(\overline{D})(D_i - \overline{D}) \stackrel{\text{def}}{=} \frac{1}{n}\text{tr}\{\ddot{g}(\overline{D})\widehat{\Sigma}_D\}.$$

Note that $n^{-1}\text{tr}\{\ddot{g}(\overline{D})\widehat{\Sigma}_D\} - n^{-1}\text{tr}\{\ddot{g}(D)\Sigma_D\} = n^{-1}\text{tr}\{\ddot{g}(\overline{D})(\widehat{\Sigma}_D - \Sigma_D)\} + n^{-1}\text{tr}\{(\ddot{g}(\overline{D}) - \ddot{g}(D))\Sigma_D)$. One could verify that the leading bias terms of the above two parts are of $O(n^{-2})$. Then we derive the variance order. It suffices to derive the upper bound for $n^{-1}\|\widehat{\Sigma}_D - \Sigma_D\|_{\max}$ and $\|\ddot{g}(\overline{D}) - \ddot{g}(D)\|_{\max}$.

Note that each element of D_i follows sub-Exponential distribution, then by Lemma E.3 we could obtain that $n^{-1}\|\widehat{\Sigma}_D - \Sigma_D\|_{\max} = o_p(n^{-1})$. Second, we have $\|\overline{D} - D\|_{\max} = o_p(1)$ by Lemma E.2. It implies $\overline{D}^{(j)} \geq D^{(j)}/2$ for $1 \leq j \leq 3$ with probability tending to 1. Within the region $\{\overline{D}^{(j)} \geq D^{(j)}/2 : 1 \leq j \leq 3\}$, we could verify that $\dddot{g}(\overline{D}) \leq c_D$ with c_D being a finite constant. Then we have $\|\ddot{g}(\overline{D}) - \ddot{g}(D)\|_{\max} \leq c_D\|\overline{D} - D\|_{\max}$. By Lemma E.2 we have $n^{-1}\|\overline{D} - D\|_{\max} = o_p(n^{-1})$. Consequently we have the leading bias of $\Delta_{(k)}^a - \Delta_{xb}$ is $O(n^{-2})$ and leading variance term is of $o_p(n^{-1})$.

Step 2. (Bias and Variance of $\Delta_{(k)}^b$)

One could also verify by using Taylor's expansion technique that the leading bias of Δ_i^b is $O(n^{-2})$. Next, it can be derived that $\|\ddot{g}(\xi_i) - \ddot{g}(\overline{D})\|_{\max} \le c\|\xi_i - \overline{D}\|_{\max} \le c\|\overline{D}_{-i} - \overline{D}\|_{\max} = c(n-1)^{-1}\|D_i - \overline{D}\|_{\max}$. As a result, we have by Lemma E.3

$$\begin{aligned} P(\max_i \|\ddot{g}(\xi_i) - \ddot{g}(\overline{D})\|_{\max} > t) &\le P((n-1)^{-2} \max_i \|D_i - \overline{D}\|_{\max}^2 > ct^2) \\ &\le P((n-1)^{-2} \sum_i \|D_i - \overline{D}\|^2 > ct^2) \le \Delta_{E2}(nt^2, n, M). \end{aligned} \tag{C.8}$$

Letting $t = n^{-3/4}$ and M be small enough, we have $\max_i \|\ddot{g}(\xi_i) - \ddot{g}(\overline{D})\|_{\max} = O_p(n^{-3/4})$. Consequently, $|n^{-1}(n-1)\sum_i \Delta_i^b| \le \max_i \sigma_1(\ddot{g}(\xi_i) - \ddot{g}(\overline{D})) \le \max_i \|\ddot{g}(\xi_i) - \ddot{g}(\overline{D})\|_F n^{-1}\mathrm{tr}(\widehat{\Sigma}_D) = O_p(n^{-7/4}) = o_p(n^{-1})$ due to that $\mathrm{tr}(\widehat{\Sigma}_D) = O_p(1)$.

As a result, the leading bias of $\Delta_{(k)}^b$ is $O(n^{-2})$ and leading variance term is of $o_p(n^{-1})$.

C.5 Proof of Theorem 2

Based on the result of Theorem B.2 and Lemma 2, it suffices to derive the tail bound for the bias estimator.

1. Proof of (a).

Define $D_{(k)i} = (\widetilde{X}_{(k)ij}^2, \widetilde{Y}_{(k)i}^2, \widetilde{X}_{(k)ij}\widetilde{Y}_{(k)i})^\top$, $\overline{D}_{(k)} = n^{-1}\sum_{i=1}^n D_{(k)i} \in \mathbb{R}^3$ and $D = (\sigma_x^2, \sigma_y^2, \sigma_{xy})^\top \in \mathbb{R}^3$. In addition, let $\overline{D}_{(k)-i} = (n-1)^{-1}\sum_{m\neq i} D_{(k)m} \in \mathbb{R}^3$. For convenience we omit the sub-index k and j here. Define $g(x) = x_3^2/(x_1x_2)$ for $x = (x_1, x_2, x_3)^\top$. Then the bias estimator is $B^{-1}\sum_k \widehat{\Delta}_{(k)} \stackrel{\text{def}}{=} B^{-1}\sum_k\{n^{-1}(n-1)\sum_i g(\overline{D}_{(k)-i})\}$ and $\mathcal{R}_{\mathbb{X}_j}^2 = g(D)$. Without of loss of generality we assume $n^2 \le N$.

By using the Taylor's expansion, we have

$$\begin{aligned} g(\overline{D}_{(k)-i}) - g(\overline{D}_{(k)}) &= \dot{g}(\overline{D}_{(k)})^\top(\overline{D}_{(k)-i} - \overline{D}_{(k)}) + 2^{-1}(\overline{D}_{(k)-i} - \overline{D}_{(k)})^\top \ddot{g}(\xi_{(k)i})(\overline{D}_{(k)-i} - \overline{D}_{(k)}) \\ &\stackrel{\text{def}}{=} \dot{g}(\overline{D})^\top(\overline{D}_{(k)-i} - \overline{D}_{(k)}) + \Delta_{(k)i}, \end{aligned}$$

where $\xi_{(k)i}$ lies on the line joining $\overline{D}_{(k)}$ and $\overline{D}_{(k)-i}$. Note that for the first term we have $\sum_i(\overline{D}_{(k)-i} - \overline{D}_{(k)}) = (n-1)^{-1}\sum_i(\overline{D}_{(k)} - D_{(k)i}) = \mathbf{0}$. Write $\Delta_{(k)i} = 2^{-1}(\overline{D}_{(k)-i} - \overline{D}_{(k)})^\top \ddot{g}(\overline{D}_{(k)})(\overline{D}_{(k)-i} - \overline{D}_{(k)}) + 2^{-1}(\overline{D}_{(k)-i} - \overline{D}_{(k)})^\top\{\ddot{g}(\xi_{(k)i}) - \ddot{g}(\overline{D}_{(k)})\}(\overline{D}_{(k)-i} - \overline{D}_{(k)}) \stackrel{\text{def}}{=} \Delta^a_{(k)i} + \Delta^b_{(k)i}$. Then we have

$$\frac{1}{B}\sum_k \widehat{\Delta}_{(k)} = \frac{n-1}{nB}\sum_{k,i}\Delta^a_{(k)i} + \frac{n-1}{nB}\sum_{k,i}\Delta^b_{(k)i} \stackrel{\text{def}}{=} \frac{1}{B}\sum_k \Delta^a_{(k)} + \frac{1}{B}\sum_k \Delta^b_{(k)} \stackrel{\text{def}}{=} \Delta^a + \Delta^b.$$

Define $V = \mathrm{tr}\{\ddot{g}(D)\mathrm{cov}(D_i)\} \stackrel{\text{def}}{=} \mathrm{tr}\{\ddot{g}(D)\Sigma_D\}$. Then $\Delta_{xb} = (2n)^{-1}V$. In the following we deal with the two parts respectively.

Step 1. (Proof of $P(|\Delta^a - \Delta_{xb}| > \epsilon)$)

It can be derived that

$$\Delta^a = \frac{1}{2B}\sum_k \frac{1}{n(n-1)}\sum_i (D_{(k)i} - \overline{D}_{(k)})^\top \ddot{g}(\overline{D}_{(k)})(D_{(k)i} - \overline{D}_{(k)}) \stackrel{\text{def}}{=} \frac{1}{2nB}\sum_k \mathrm{tr}\big\{\ddot{g}(\overline{D}_{(k)})\widehat{\Sigma}_{D(k)}\big\}.$$

Further write $\Delta^a = (2nB)^{-1}\sum_k \mathrm{tr}\big\{\ddot{g}(D)\widehat{\Sigma}_{D(k)}\big\} + (2nB)^{-1}\sum_k \mathrm{tr}\big\{(\ddot{g}(\overline{D}_{(k)}) - \ddot{g}(D))\widehat{\Sigma}_{D(k)}\big\} \stackrel{\text{def}}{=} \Delta^{a1} + \Delta^{a2}$. First, by Lemma E.3, $P(|\Delta^{a1} - \Delta_{xb}| > t) \lesssim \Delta_{E2}(nt, nB, M)$. Letting $M = n^{1/2}$ we obtain $p_1\Delta_{E2}(nN^{-\nu}, nB, M) \to 0$ by the condition that $\log p_1 \ll \min\{nBN^{-2\nu}, n^{1/2}BN^{-\nu}, n^{1/2}\}$.

By Taylor's expansion, we have

$$\ddot{g}_{j_1j_2}(\overline{D}_{(k)}) - \ddot{g}_{j_1j_2}(D) = \sum_{j_3} \dddot{g}_{j_1j_2j_3}(D)(\overline{D}^{(j_3)}_{(k)} - D^{(j_3)})$$
$$+ \sum_{j_3,j_4} \ddddot{g}_{j_1j_2j_3j_4}(\xi_{(k)})(\overline{D}^{(j_3)}_{(k)} - D^{(j_3)})(\overline{D}^{(j_4)}_{(k)} - D^{(j_4)}) \stackrel{\text{def}}{=} \Delta_{g1j_1j_2(k)} + \Delta_{g2j_1j_2(k)},$$

where $\xi_{(k)}$ lies on the line joining $\overline{D}_{(k)}$ and D. We deal with the above two parts respectively as follows.

Step 1.1. $(P(|(nB)^{-1}\sum_k \sum_{j_1,j_2} \Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{D,j_2j_1(k)} - O(n^{-2})| > \epsilon))$

First note the bias of the first term $\Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{D,j_1j_2(k)}$ is in the order of $O(n^{-2})$. Next by Lemma E.14, we could obtain $P(|(nB)^{-1}\sum_k \sum_{j_1,j_2} \Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{D,j_1j_2(k)}\Sigma_{D,j_2j_1(k)} - O(n^{-2})| > \epsilon) \lesssim \Delta^{sas*}_{E4}(nt, n, B, \epsilon)$. Let $\epsilon = n^{1/4}$. Then we have $p_1\Delta^{sas*}_{E4}(t, n, B, \epsilon) \to 0$ as long as $\log p_1 \ll \min\{nBN^{-2\nu}, n^{1/2}BN^{-\nu}, n^{1/2}\}$.

Step 1.2. $(P(|(nB)^{-1}\sum_k \sum_{j_1,j_2} \Delta_{g2j_1j_2(k)}\widehat{\Sigma}_{D,j_2j_1(k)} - O(n^{-2})| > \epsilon))$

Next, we have $P(\max_k \|\widehat{\Sigma}_{D(k)} - \Sigma_D\|_{\max} > \epsilon) \lesssim B\Delta^*_{E2}(\epsilon, n, N, 2/3)$ by Lemma E.4. Note we have $p_1B\Delta^*_{E2}(\epsilon, n, N, 2/3) \to 0$ as long as $\log p_1 + \log B \ll \min\{n^{1/3}, N^{1/4}\}$. As a result, we have $\max_k \|\widehat{\Sigma}_{D(k)}\|_{\max} \le c$ with probability tending to 1.

In the meanwhile $P\{\max_k \|\overline{D}_{(k)} - D\| > \epsilon\} \lesssim B\Delta_E(\epsilon, n, N)$ by Lemma E.2 and $p_1B\Delta_E(\epsilon, n, N) \to 0$ as long as $\log p_1 + \log B \ll \min\{n^{1/2}, N^{1/4}\}$. Hence under the event $\{\max_k \|\overline{D}_{(k)} - D\| > \epsilon\}$ with small enough ϵ (which holds with probability tending to 1), we have $\max_k \|\ddddot{g}(\xi_{(k)})\|_{\max} \le c$. This leads to $|\Delta_{g2j_1j_2(k)}| \le c\|\overline{D}_{(k)} - D\|^2$. The bias of $\|\overline{D}_{(k)} - D\|^2$ is $O(n^{-1})$. Next using Lemma E.7, we have $P(|(nB)^{-1}\sum_k \sum_{j_1,j_2} \Delta_{g2j_1j_2(k)}\widehat{\Sigma}_{D,j_2j_1(k)} - O(n^{-2})| > \epsilon) \lesssim \Delta_{E4}(n^3Bt, M)$. Then we have $p_1\Delta_{E4}(n^3Bt, M) \to 0$ by setting $M = n^{3/4}$ as long as $\log p_1 \ll \min\{nBN^{-2\nu}, n^{1/2}BN^{-\nu}, n^{3/4}\}$

Step 2. (Proof of $P(|\Delta^b - O(n^{-2})| > \epsilon)$)

Under the event that $\{\max_k \|\overline{D}_{(k)} - D\|_{\max} \le \epsilon\}$ and $\{\max_{k,i} \|\xi_{(k)i} - D\|_{\max} \le \epsilon\}$, we have $\|\ddot{g}(\xi_{(k)i}) - \ddot{g}(\overline{D}_{(k)})\|_{\max} \le c\|\xi_{(k)i} - \overline{D}_{(k)}\|_{\max} \le c\|\overline{D}_{(k)-i} - \overline{D}_{(k)}\|_{\max} = c(n-1)^{-1}\|D_{(k)i} - \overline{D}_{(k)}\|_{\max}$. In the meanwhile, it holds, $P\{\max_k \|\overline{D}_{(k)} - D\|_{\max} > \epsilon\} \le B\Delta_E(t, n, N)$ by Lemma E.2 and $P\{\max_{k,i} \|\xi_{(k)i} - D\|_{\max} > \epsilon\} \le nB\Delta_E(t, n-1, N)$. Hence we have $|\Delta^b| \le cB^{-1}(n-1)^{-3}\sum_{k,i} \|D_{(k)i} - \overline{D}_{(k)}\|^3_{\max}$. As a result, the bias term of $|\Delta^b|$ is at most $O(n^{-2})$. By Lemma E.8, we have $P(|\Delta^b - O(n^{-2})| > t) \lesssim \Delta_{E5}(n^2 t, nB, M_1) + bB\Delta_E(t_1, n, N)$. Setting $M_1 = n^{2/3}$, $t = N^{-\nu}$ and $t_1 = o(1)$, we could achieve the result.

2. Proof of (b).

We decompose the $R^2_{\mathbb{Z}_j,\mathrm{DAS}}$ as in (b) of Theorem B.2. For simplicity, we treat σ_y^2 as known and focus on others. For convenience, we define $D_{(k)i} = (\widetilde{\mathcal{Z}}_{(k)ij1}, \cdots, \widetilde{\mathcal{Z}}_{(k)ij(l_j-1)}, \widetilde{\mathcal{Z}}_{(k)ij1}\widetilde{Y}_{(k)i}, \cdots, \widetilde{\mathcal{Z}}_{(k)ij(l_j-1)}\widetilde{Y}_{(k)i})^\top \in \mathbb{R}^{2(l_j-1)}$, $\overline{D}_{(k)} = n^{-1}\sum_{i=1}^n D_{(k)i} \in \mathbb{R}^{2(l_j-1)}$, $\overline{D} = B^{-1}\sum_{k=1}^B \overline{D}_{(k)} \in \mathbb{R}^{2(l_j-1)}$ and $D = (\pi_1, \cdots, \pi_{l_j-1}, \sigma_{zy,j1}, \cdots, \sigma_{zy,j(l_j-1)})^\top \in \mathbb{R}^{2(l_j-1)}$. In addition, define

$$g(D) = \sigma_y^{-2}\sum_{l=1}^{l_j-1} D_l^{-1} D_{l_j+l}^2$$

Then the bias estimator is $B^{-1}\sum_k \widehat{\Delta}_{(k)} \overset{\text{def}}{=} B^{-1}\sum_k\{n^{-1}(n-1)\sum_i g(\overline{D}_{(k)-i})\}$ and $\mathcal{R}^2_{\mathbb{Z}_j} = g(D)$.

By Taylor's expansion, we have

$$\begin{aligned} g(\overline{D}_{(k)-i}) - g(\overline{D}_{(k)}) &= \dot{g}(\overline{D}_{(k)})^\top(\overline{D}_{(k)-i} - \overline{D}_{(k)}) + 2^{-1}(\overline{D}_{(k)-i} - \overline{D}_{(k)})^\top \ddot{g}(\xi_{(k)i})(\overline{D}_{(k)-i} - \overline{D}_{(k)}) \\ &\overset{\text{def}}{=} \dot{g}(\overline{D})^\top(\overline{D}_{(k)-i} - \overline{D}_{(k)}) + \Delta_{(k)i}. \end{aligned}$$

Similarly as in the proof of (a), we have $\sum_i \dot{g}(\overline{D})^\top(\overline{D}_{(k)-i} - \overline{D}_{(k)}) = \mathbf{0}$. It leaves to deal with the second part. Define $\Delta_{(k)}$, Δ^a, Δ^b as in the previous proof of (a).

Step 1. (Proof of $P(|\Delta^a - \Delta_{zb}| > \epsilon)$)

In Step 1, we have $\Delta^a = (2nB)^{-1}\sum_k \mathrm{tr}\{\ddot{g}(D)\widehat{\Sigma}_{D(k)}\} + (2nB)^{-1}\sum_k \mathrm{tr}\{(\ddot{g}(\overline{D}_{(k)}) - \ddot{g}(D))\widehat{\Sigma}_{D(k)}\} \stackrel{\text{def}}{=} \Delta^{a1} + \Delta^{a2}$. For Δ^{a1}, it can be decomposed into four parts as in (C.4)–(C.7). The bias can be verified in the order of $O(n^{-2}l_j)$. Then by Lemma E.3, $P(|\Delta^{a1} - O(n^{-2}l_j)| > \epsilon) \lesssim \Delta_{E2}(nt/l_j, nB, M)$. Setting $M = n^{1/2}$ then we have $p_2 \max_j l_j \Delta_{E2}(nt/l_j, nB, M) \to 0$ using the condition that $\log p_2 + \max_j \log l_j \ll \min\{nBN^{-2\nu}l_j^{-2}, n^{1/2}BN^{-\nu}l_j^{-1}, n^{1/2}\}$.

By Taylor's expansion, we have

$$\ddot{g}_{j_1j_2}(\overline{D}_{(k)}) - \ddot{g}_{j_1j_2}(D) = \sum_{j_3} \dddot{g}_{j_1j_2j_3}(D)(\overline{D}^{(j_3)}_{(k)} - D^{(j_3)})$$
$$+ \sum_{j_3,j_4} \ddddot{g}_{j_1j_2j_3j_4}(\xi_{(k)})(\overline{D}^{(j_3)}_{(k)} - D^{(j_3)})(\overline{D}^{(j_4)}_{(k)} - D^{(j_4)}) \stackrel{\text{def}}{=} \Delta_{g1j_1j_2(k)} + \Delta_{g2j_1j_2(k)},$$

where $\xi_{(k)}$ lies on the line joining $\overline{D}_{(k)}$ and D. We deal with the above two parts respectively.

Step 1.1. $P(|(nB)^{-1}\sum_k \sum_{j_1,j_2} \Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{Dj_2j_1(k)} - O(n^{-2}\sum_l \pi_{jl}^{-1})| > \epsilon)$

First note that when $j_1 = j_2$ or $|j_1 - j_2| = l_j - 1$, we have $\ddot{g}_{j_1j_2}(D) \neq 0$, otherwise we have $\ddot{g}_{j_1j_2}(D) = 0$. The nonzero positions are the same for $\dddot{g}_{j_1j_2j_3}(D)$ for any j_3. In addition, we have $\Sigma_{D,ll} = O(\pi_{jl})$ and $\Sigma_{D,lm} = O(\pi_{jl})$ when $|l-m| = l_j - 1$. Moreover, $\dddot{g}_{llj_3}(D) = O(\pi_{jl}^{-2})$ and $\dddot{g}_{lmj_3}(D) = O(\pi_{jl}^{-2})$ for $|m-l| = l_j - 1$. Hence for the first part we note that the bias of $\sum_{j_1,j_2} \Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{Dj_2j_1(k)}$ is in the order of $O(n^{-2}\sum_l \pi_{jl}^{-1})$. Next, by Lemma E.6, $P(|(nB)^{-1}\sum_k \sum_{j_1,j_2} \Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{Dj_2j_1(k)} - O(n^{-2}\sum_l \pi_{jl}^{-1})| > \epsilon) \lesssim \Delta_{E4}^{sas*}(n\epsilon\pi_{\min}^{1/2}/l_j, n, B, t)$. Let $t = n^{1/4}$. Then we have $p_2 \max_j \Delta_{E4}^{sas*}(n\epsilon\pi_{\min}^{1/2}/l_j, n, B, t) \to 0$ as long as $\log p_2 + \max_j \log l_j \ll \min\{nBN^{-2\nu}\pi_{\min}l_j^{-2}, n^{1/2}BN^{-\nu}\pi_{\min}^{1/2}l_j^{-1}, n^{1/2}\}$.

Step 1.2. $P(|(nB)^{-1}\sum_k\sum_{j_1,j_2}\Delta_{g2j_1j_2(k)}\widehat{\Sigma}_{Dj_2j_1(k)}-O(n^{-2}\sum_l\pi_{jl}^{-1})|>\epsilon)$

Next, we find that $|\dddot{g}_{j_1j_2j_3j_4}(D)|$ is nonzero only if for all $1\le m_1,m_2\le 4$, $|j_{m_1}-j_{m_2}|=0$ or l_j-1. Hence the total number of nonzero elements in $|\dddot{g}_{j_1j_2j_3j_4}(D)|$ is $O(l_j)$. Hence under the event that $\{\max_k\|\xi_{(k)}-D\|_{\max}\le c\pi_{\min}/2\}$ with c small enough, we have $|\dddot{g}_{j_1j_2j_3j_4}(\xi_{(k)})|=O(\pi_{jl}^{-3})$ for any $j_m=l$ ($j_m\in\{j_1,j_2,j_3,j_4\}$). This can be guaranteed by $P\{\max_k\|\xi_{(k)}-D\|_{\max}>c\pi_{\min}/2\}\lesssim Bl_j\Delta_E(\pi_{\min}^{1/2},n,N)$ and $\log p_2+\log B+\max_j\log l_j\ll\min\{n^{1/2}\pi_{\min}^{1/2},N^{1/2}\}$.

Next, we have $P(\max_{k,i,j}\|\widehat{\Sigma}_{D,ij(k)}-\Sigma_{D,ij}\|_{\max}>\epsilon|\Sigma_{D,ij}|)\lesssim Bl_j^2\Delta_{E2}^*(\epsilon,n,N,2/3)$. In addition, under the condition that $\log p_2+\log B+\max_j\log l_j\ll\min\{n^{1/3},N^{1/4}\}$, we have $p_2B\max_j l_j^2\Delta_{E2}^*(\epsilon,n,N,2/3)\to 0$. Hence with probability tending to 1 we have $\max_{k,i,j}|\widehat{\Sigma}_{D,ij(k)}|\le c|\Sigma_{D,ij}|$, where c is a finite constant.

In addition we have $|(\overline{D}_{(k)}^{(j_3)}-D^{(j_3)})(\overline{D}_{(k)}^{(j_4)}-D^{(j_4)})|\le 2^{-1}\{(\overline{D}_{(k)}^{(j_3)}-D^{(j_3)})^2+(\overline{D}_{(k)}^{(j_4)}-D^{(j_4)})^2\}$ for $j_3\ne j_4$ and $|j_3-j_4|=l_j-1$. Hence the bias term of $(nB)^{-1}\sum_k\sum_{j_1,j_2}\sum_{j_3,j_4}$ $|\dddot{g}_{j_1j_2j_3j_4}(\xi_{(k)})||(\overline{D}_{(k)}^{(j_3)}-D^{(j_3)})(\overline{D}_{(k)}^{(j_4)}-D^{(j_4)})||\widehat{\Sigma}_{Dj_2j_1(k)}|$ is $O(n^{-2}\sum_l\pi_{jl}^{-1})$ by omitting the zero elements in $\dddot{g}_{j_1j_2j_3j_4}(\xi_{(k)})$.

As a result, we obtain $P(|\sum_{j_1,j_2}|(nB)^{-1}\sum_k\Delta_{g2j_1j_2(k)}||\Sigma_{D,j_2j_1}|-O(n^{-2}\sum_l\pi_{jl}^{-1})|>\epsilon)\lesssim\Delta_{E4}(n^3B\pi_{\min}\epsilon/l_j,M)$ by Lemma E.7. Letting $M=n^{1/3}$ then we obtain the result as long as $\log p_2+\max_j\log l_j\ll\min\{nBN^{-2\nu}\pi_{\min}l_j^{-2},n^{1/2}\ BN^{-\nu}\pi_{\min}^{1/2}l_j^{-1},n^{1/3}\}$.

Step 2. (Proof of $P(|\Delta^b-O(n^{-2}\sum_l\pi_{jl}^{-1})|>\epsilon)$)

In Step 2, we still have $P\{\max_k\|\overline{D}_{(k)}-D\|_{\max}>t\}\le B\Delta_E(t,n,N)$ and $P\{\max_{k,i}$ $\|\xi_{(k)i}-D\|_{\max}>t\}\le nB\Delta_E(t,n-1,N)$ by Lemma E.2. Setting $t=\pi_{\min}$ then we have $p_2\max_j l_jnB\Delta_E(t,n-1,N)\to 0$ as long as $\log p_2+\max_j\log l_j\ll\min\{(n\pi_{\min})^{1/2},N^{1/2}\}$. Recall that $\dddot{g}_{llj_3}(D)=O(\pi_{jl}^{-2})$ and $\dddot{g}_{lmj_3}(D)=O(\pi_{jl}^{-2})$ for $|m-l|=l_j-1$. Define $D_\pi=\mathrm{diag}\{\pi_{j1}^{-2},\cdots,\pi_{j(l_j-1)}^{-2}\}$. This leads to $|\Delta_b|\le cB^{-1}(n-1)^{-3}\sum_{k,i}\|D_\pi(D_{(k)i}-\overline{D}_{(k)})\|_{\max}^3$.

Hence by Lemma E.8 we have $P(|\Delta_b - O(n^{-2})| > t) \leq \Delta_{E5}(n^2 t\pi_{\min}^{1/2}/l_j, nB, M)$. Let $M = n^{2/3}$. Then we could obtain that $p_2 \max_j l_j \Delta_{E5}(n^2 t\pi_{\min}^{1/2}/l_j, nB, M) \to 0$ as long as $\log p_2 + \max_j \log l_j \ll \min\{nBN^{-2\nu}\pi_{\min} l_j^{-2}, n^{1/2}BN^{-\nu}\pi_{\min}^{1/2} l_j^{-1}, n^{2/3}\}$. This leads to the final conclusion.

C.6 Proof of Theorem 3

1. Proof of (a).

The proof follows the proof of (a) of Theorem 1. Specifically, it can be derived that $P(\max_j |R^2_{\mathbb{X}_j,\mathrm{AMS}} - \mathcal{R}^2_{\mathbb{X}_j}| > N^{-\nu}) \leq p_1 \Delta_E^{sas}(N^{-\nu}, n, B, N)$ by Lemma E.11. As a result, as long as Assumption 4 (a) holds, we have $P(\max_j |R^2_{\mathbb{X}_j,\mathrm{AMS}} - \mathcal{R}^2_{\mathbb{X}_j}| > N^{-\nu}) \to 0$.

2. Proof of (b).

The proof follows the proof of (b) of Theorem 1. It can be derived that $P(\max_j |R^2_{\mathbb{X}_j,\mathrm{AMS}} - \mathcal{R}^2_{\mathbb{X}_j}| > N^{-\nu}) \lesssim p_2 \max_j l_j \Delta_E^{sas}(N^{-\nu}/l_j, n, B, N) + p_2 \max_j l_j \max_l \Delta_B^{sas}(\pi_{jl}, n, B, N, \pi_{jl})$ by Lemma E.17. Then under Assumption 4 (b) we have $P(\max_j |R^2_{\mathbb{Z}_j,\mathrm{AMS}} - \mathcal{R}^2_{\mathbb{Z}_j}| > N^{-\nu}) \to 0$.

C.7 Proof of Theorem B.4

Under the SAS setting, the bias order of AVS measure is the same as in the RAS setting, so omit the proof of the bias here. Next, we prove the uniform convergence for AVS under SAS as follows.

1. Proof of $\max_j |R^2_{\mathbb{X}_j,\mathrm{AVS}} - \Delta_{xb} - \mathcal{R}^2_{\mathbb{X}_j}| = O_p(N^{-\nu})$ with $\Delta_{xb} = O(n^{-1})$.

The guideline of the proof follows the proof of Theorem B.2, hence we only state the difference here.

Define $D_{(k)i}$, $\overline{D}_{(k)}$, and D as in the proof of Theorem B.2. By using the Taylor's expansion, it leads to

$$B^{-1}\sum_k g(\overline{D}_{(k)}) - g(D) = \frac{1}{B}\sum_k \dot{g}(D)^\top(\overline{D}_{(k)} - D) + \frac{1}{B}\sum_k (\overline{D}_{(k)} - D)^\top \ddot{g}(\xi_{(k)})(\overline{D}_{(k)} - D)$$

$\stackrel{\text{def}}{=} \Delta_1 + \Delta_2$, where $\xi_{(k)}$ is on the line joining D and $\overline{D}_{(k)}$. In the following we use two steps to deal with Δ_1 and Δ_2 respectively.

STEP 1. First, by Lemma E.11, we have $P(|\Delta_1| > t) \leq \Delta_E^{sas}(t, n, B, N)$.

STEP 2. First we write Δ_2 as

$$\begin{aligned}\Delta_2 &= \frac{1}{B}\sum_k (\overline{D}_{(k)} - D)^\top \ddot{g}(D)(\overline{D}_{(k)} - D) + \frac{1}{B}\sum_k (\overline{D}_{(k)} - D)^\top \{\ddot{g}(\xi_{(k)}) - \ddot{g}(D)\}(\overline{D}_{(k)} - D) \\ &\stackrel{\text{def}}{=} \Delta_{21} + \Delta_{22}.\end{aligned}$$

By Lemma E.12, we have $P(|\Delta_{21}| > t) \leq \Delta_{E2}^{sas}(t, n, B, N)$. Next, as shown in the proof of Theorem B.2, as long as $\|\xi_{(k)} - D\|_{\max} \leq \epsilon$ and ϵ small enough, we have $|\Delta_{22}| \ll |\Delta_{21}|$. In summary, it holds $P(|\Delta_2| > t) \leq \Delta_{E2}^{sas}(t, n, B, N) + B\Delta_E(t_1, n, N)$, where $t_1 = o(1)$. Under the condition that $\log p_1 + \log B \ll n^{1/2}$, $\log p_1 + \log N \ll n^{2\delta}$, and $\log p_1 \ll \min\{n^{1-2\delta}BN^{-2\nu}, n^{1/2-\delta}BN^{-\nu}\}$, we have $p_1\Delta_{E2}^{sas}(N^{-\nu}, n, B, N) + p_1 B\Delta_E(t_1, n, N) \to 0$.

2. Proof of $\max_j |R^2_{\mathbb{Z}_j,\mathrm{AVS}} - \Delta_{zb} - \mathcal{R}^2_{\mathbb{Z}_j}| = O_p(N^{-\nu})$ with $\Delta_{zb} = O(n^{-1}l_j)$

The proof follows the conclusion (b) of the Theorem B.2. Define all the notations in the same way except under the SAS sampling scheme. In STEP 1, we obtain $P(|\Delta_1| > \epsilon) \leq l_j \Delta_E^{sas}(\epsilon/l_j, n, B, N)$. Hence $p_2 \max_j l_j \Delta_E^{sas}(N^{-\nu}/l_j, n, B, N) \to 0$ as long as $\log p_2 + \max_j \log l_j \ll \min\{n^{1-2\delta}BN^{-2\nu}l_j^{-2}, n^{1/2-\delta}BN^{-\nu}l_j^{-1}\}$ and $\log p_2 + \max_j \log l_j + \log N \ll n^{2\delta}$. In STEP 2.1, we obtain $P(|\Delta_{D1} + \Delta_{D2} - E(\Delta_{D1})| > \epsilon) \lesssim \Delta_{E2}^{sas}(\epsilon/l_j, n, B, N)$. Hence as long as $\log p_2 + \max_j \log l_j + \log N \ll n^{2\delta}$ and $\log p_2 +$

$\max_j \log l_j \ll \min\{n^{1-2\delta}BN^{-2\nu}l_j^{-2}, n^{1/2-\delta}BN^{-\nu}l_j^{-1}\}$, it holds $p_2 \max_j l_j \Delta_{E2}^{sas}(N^{-\nu}/l_j, n, B, N) \to 0$. Next, in STEP 2.2, we further have $P(\pi_{jl}^{-1}\max_k |\widehat{\pi}_{(k),jl} - \pi_{jl}| > t) + P(\pi_{jl}^{-1}\max_k |\widehat{\sigma}_{zy,(k)jl} - \sigma_{zy,jl}| > t) + P(\max_k |\widehat{\sigma}_{y(k)}^2 - \sigma_y^2| > t) \lesssim B\Delta_E(t\pi_{\min}^{1/2}, n, N)$ with $t = O(1)$. Then we have $p_2 \max_j l_j B\Delta_E^{sas}(\pi_{\min}^{1/2}, n, N) \to 0$ as long as $\log p_2 + \max_j \log l_j + \log B \ll n^{1/2}\pi_{\min}^{1/2}$.

C.8 Proof of Theorem 4

Under the SAS setting, the bias order of DAS measure is the same as in the RAS setting, so omit the proof of the bias here. Next, we prove the uniform convergence for DAS under SAS as follows.

1. Proof of $\max_j |R^2_{\mathbb{X}_j,\mathrm{DAS}} - \Delta_{xb} - \mathcal{R}^2_{\mathbb{X}_j}| = O_p(N^{-\nu})$ with $\Delta_{xb} = O(n^{-2})$

The idea of the proof is the same as in the proof of Theorem 2. The notations are defined in the same way as in Theorem 2 but under the SAS sampling scheme. The main proof differences are in STEP 1 and STEP 2.

In STEP 1, we first obtain $P(|\Delta^{a1} - \Delta_{xb}| > \epsilon) \lesssim \Delta_{E3}^{sas}(nt, n, B, N, M, \epsilon)$ by Lemma E.13. Let $M = n^{2\delta}$, $\epsilon = n^{1/2+\delta}$. Then we have $p_1\Delta_{E3}^{sas}(nt, n, B, N, M, \epsilon) \to 0$ as long as $\log p_1 + \log N \ll \min\{n^{2\delta}, n^{3/2-3\delta}, n^{2-6\delta}\}$ for $\delta \in (0, 1/3)$ and $\log p_1 \ll \min\{n^{1-2\delta}BN^{-2\nu}, n^{1/2-\delta}BN^{-\nu}\}$.

In STEP 1.1, we obtain $P(|(nB)^{-1}\sum_k \sum_{j_1,j_2} \Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{D,j_1j_2(k)}\Sigma_{D,j_2j_1(k)} - O(n^{-2})| > t) \lesssim \Delta_{E4}^{sas*}(nt, n, B, \epsilon)$ by Lemma E.14. Let $\epsilon = n^{1/4+\delta/2}$. Then we have $p_1\Delta_{E4}^{sas*}(nt, n, B, \epsilon) \to 0$ as long as $\log p_1 \ll \min\{n^{1-2\delta}BN^{-2\nu}, n^{1-\delta}BN^{-\nu}\}$ and $\log p_1 + \log N \ll \min\{n^{3/2+\delta}, n^{5/8+\delta/4}\}$ under the condition that $\delta \in (0, 1/3)$.

In STEP 1.2, we also have $\max_k \|\widehat{\Sigma}_{D(k)}\|_{\max} \le c$ with probability tending to 1 as long as $\log p_1 + \log B \ll \min\{n^{1/3}, N^{1/4}\}$. Similarly we can show that $\max \|\dddot{g}(\xi_{(k)})\|_{\max} \le c$ with probability tending to 1 as long as $\log p_1 + \log B \ll \min\{n^{1/2}, N^{1/4}\}$. This leads to

$|\Delta_{g2j_1j_2(k)}| \leq c\|\overline{D}_{(k)}-D\|^2$. Next, by using Lemma $P(|(nB)^{-1}\sum_k\sum_{j_1,j_2}\Delta_{g2j_1j_2(k)}\widehat{\Sigma}_{D,j_2j_1(k)}-O(n^{-2})| > t) \lesssim \Delta^{sas}_{E2}(nt, n, B, N)$. We have $p_1\Delta^{sas}_{E2}(nN^{-\nu}, n, B, N) \to 0$ as long as $\log p_1 \ll \min\{n^{1-2\delta}BN^{-2\nu}, n^{1-\delta}BN^{-\nu}, n^{2\delta}\}$.

In Step 2, we still have $P\{\max_k \|\overline{D}_{(k)}-D\|_{\max} > \epsilon\} \leq B\Delta_E(t, n, N)$ and $P\{\max_{k,i} \|\xi_{(k)i} - D\|_{\max} > \epsilon\} \leq nB\Delta_E(t, n-1, N)$ by Lemma E.2. This leads to $|\Delta_b| \leq cB^{-1}(n-1)^{-3}\sum_{k,i}\|D_{(k)i}-\overline{D}_{(k)}\|^3_{\max}$. Hence by Lemma E.16 we have $P(|\Delta_b-O(n^{-2})| > t) \leq \Delta^{sas}_{E5}(n^2t, n, B, N, M, \epsilon)$. Let $M = n^{2\delta}$ and $\epsilon = n^{3/2+\delta}$. Then we could obtain that $p_1\Delta^{sas}_{E5}(n^2t, n, B, N, M, \epsilon) \to 0$ as long as $\log p_1 \ll \min\{n^{1-2\delta}BN^{-2\nu}, n^{1-\delta}BN^{-\nu}, n^{2\delta}\}$. This leads to the final conclusion.

2. Proof of $\max_j |R^2_{\mathbb{Z}_j,\mathrm{DAS}} - \Delta_{zb} - \mathcal{R}^2_{\mathbb{Z}_j}| = O_p(N^{-\nu})$ with $\Delta_{zb} = O(n^{-2}\sum_l \pi^{-1}_{jl})$

The proof follows the conclusion (b) of the Theorem 2. All the notations are defined in the same way.

Step 1. (Proof of $P(|\Delta^a - \Delta_{zb}| > \epsilon)$)

In Step 1, we first obtain that $P(|\Delta^{a1} - O(n^{-2}l_j)| > \epsilon) \lesssim \Delta^{sas}_{E3}(t/l_j, n, B, N, M, \epsilon)$. Let $M = n^{2\delta}$, $\epsilon = n^{1/2+\delta}$. Then we have $p_2 \max_j l_j\Delta^{sas}_{E3}(t/l_j, n, B, N, M, \epsilon) \to 0$ as long as $\log p_2 + \max_j \log l_j + \log N \ll \min\{n^{2\delta}, n^{3/2-3\delta}, n^{2-6\delta}\}$ for $\delta \in (0, 1/3)$ and $\log p_2 + \max_j \log l_j \ll \min\{n^{1-2\delta}BN^{-2\nu}l_j^{-2}, n^{1/2-\delta}BN^{-\nu}l_j^{-1}\}$.

Step 1.1. $P(|(nB)^{-1}\sum_k\sum_{j_1,j_2}\Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{Dj_2j_1(k)} - O(n^{-2}\sum_l \pi^{-1}_{jl})| > \epsilon)$

Next, In Step 1.1, we have $P(|(nB)^{-1}\sum_k\sum_{j_1,j_2}\Delta_{g1j_1j_2(k)}\widehat{\Sigma}_{Dj_2j_1(k)}-O(n^{-2}\sum_l \pi^{-1}_{jl})| > \epsilon) \lesssim \Delta^{sas*}_{E4}(nt\pi^{1/2}_{\min}/l_j, n, B, \epsilon)$ by Lemma E.14. Let $\epsilon = n^{1/4+\delta/2}$. Then we have $p_2 \max_j l_j\Delta^{sas*}_{E4}(t\pi^{1/2}_{\min}/l_j, n, B, \epsilon) \to 0$ as long as $\log p_2+\max_j \log l_j \ll \min\{n^{1-2\delta}BN^{-2\nu}l_j^{-2}\pi_{\min}, n^{1-\delta}BN^{-\nu}l_j^{-1}\pi^{1/2}_{\min}\}$ and $\log p_2 + \max_j \log l_j + \log N \ll \min\{n^{3/2+\delta}, n^{5/8+\delta/4}\}$. Note that $n^{2\delta} \leq n^{5/8+\delta/4} \leq n^{3/2+\delta}$. Hence the results hold.

Step 1.2. $P(|(nB)^{-1}\sum_k \sum_{j_1,j_2} \Delta_{g2j_1j_2(k)}\widehat{\Sigma}_{Dj_2j_1(k)} - O(n^{-2}\sum_l \pi_{jl}^{-1})| > \epsilon)$

In Step 1.2, first under the event that $\{\max_k \|\xi_{(k)} - D\|_{\max} \le c\pi_{\min}/2\}$ with c small enough, we have $|\dddot{g}_{j_1j_2j_3j_4}(\xi_{(k)})| = O(\pi_{jl}^{-3})$ for any $j_m = l$ ($j_m \in \{j_1, j_2, j_3, j_4\}$). Under SAS scheme, we still have $P\{\max_k \|\xi_{(k)} - D\|_{\max} > c\pi_{\min}/2\} \lesssim Bl_j\Delta_E(\pi_{\min}^{1/2}, n, N)$ and $\log p_2 + \log B + \max_j \log l_j \ll \min\{n^{1/2}\pi_{\min}^{1/2}, N^{1/2}\}$.

Next, we have $P(\max_{k,i,j} \|\widehat{\Sigma}_{D,ij(k)} - \Sigma_{D,ij}\|_{\max} > \epsilon|\Sigma_{D,ij}|) \lesssim Bl_j^2\Delta_{E2}^*(\epsilon, n, N, 2/3)$. In addition, under the condition that $\log p_2 + \log B + \max_j \log l_j \ll \min\{n^{1/3}, N^{1/4}\}$, we have $p_2B\max_j l_j^2\Delta_{E2}^*(\epsilon, n, N, 2/3) \to 0$. Hence with probability tending to 1 we have $\max_{k,i,j} |\widehat{\Sigma}_{D,ij(k)}| \le c|\Sigma_{D,ij}|$, where c is a finite constant.

In addition we have $|(\overline{D}_{(k)}^{(j_3)} - D^{(j_3)})(\overline{D}_{(k)}^{(j_4)} - D^{(j_4)})| \le 2^{-1}\{(\overline{D}_{(k)}^{(j_3)} - D^{(j_3)})^2 + (\overline{D}_{(k)}^{(j_4)} - D^{(j_4)})^2\}$ for $j_3 \ne j_4$ and $|j_3 - j_4| = l_j - 1$. Hence the bias term of $(nB)^{-1}\sum_k \sum_{j_1,j_2} \sum_{j_3,j_4}$ $|\dddot{g}_{j_1j_2j_3j_4}(\xi_{(k)})||(\overline{D}_{(k)}^{(j_3)} - D^{(j_3)})(\overline{D}_{(k)}^{(j_4)} - D^{(j_4)})||\widehat{\Sigma}_{Dj_2j_1(k)}|$ is $O(n^{-2}\sum_l \pi_{jl}^{-1})$ by omitting the zero elements in $\dddot{g}_{j_1j_2j_3j_4}(\xi_{(k)})$.

As a result, we obtain $P(|\sum_{j_1,j_2} |(nB)^{-1}\sum_k \Delta_{g2j_1j_2(k)}||\Sigma_{D,j_2j_1}| - O(n^{-2}\sum_l \pi_{jl}^{-1})| > \epsilon) \lesssim \Delta_{E2}^{sas}(nt\pi_{\min}/l_j, n, B, N)$ by Lemma E.12. Then we have $p_2\max_j l_j\Delta_{E2}^{sas}(nt\pi_{\min}/l_j, n, B, N) \to 0$ by the condition $\log p_2 + \max_j \log l_j \ll \min\{n^{1-2\delta}BN^{-2\nu}l_j^{-2}, n^{1-\delta}BN^{-\nu}l_j^{-1}, n^{2\delta}\}$ and $n\pi_{\min} \to \infty$.

Step 2. (Proof of $P(|\Delta^b - O(n^{-2}\sum_l \pi_{jl}^{-1})| > \epsilon)$)

In Step 2, we still have $P\{\max_k \|\overline{D}_{(k)} - D\|_{\max} > t\} \le B\Delta_E(t, n, N)$ and $P\{\max_{k,i}$ $\|\xi_{(k)i} - D\|_{\max} > t\} \le nB\Delta_E(t, n-1, N)$ by Lemma E.2. Setting $t = \pi_{\min}$ then we have $p_2\max_j l_jnB\Delta_E(t, n-1, N) \to 0$ as long as $\log p_2 + \max_j \log l_j \ll \min\{(n\pi_{\min})^{1/2}, N^{1/2}\}$. Recall that $\dddot{g}_{llj_3}(D) = O(\pi_{jl}^{-2})$ and $\dddot{g}_{lmj_3}(D) = O(\pi_{jl}^{-2})$ for $|m - l| = l_j - 1$. Define $D_\pi = \mathrm{diag}\{\pi_{j1}^{-2}, \cdots, \pi_{j(l_j-1)}^{-2}\}$. This leads to $|\Delta_b| \le cB^{-1}(n-1)^{-3}\sum_{k,i} \|D_\pi(D_{(k)i} - \overline{D}_{(k)})\|_{\max}^3$. Hence by Lemma E.16 we have $P(|\Delta_b - O(n^{-2})| > t) \le \Delta_{E5}^{sas}(n^2t\pi_{\min}^{1/2}l_j^{-1}, n, B, N, M, \epsilon)$.

Let $M = n^{2\delta}$ and $\epsilon = n^{2+\delta}$. Then we could obtain that $p_1 \Delta_{E5}(n^2 t \pi_{\min}^{1/2}/l_j, nB, M) \to 0$ as long as $\log p_2 + \max_j \log l_j \ll \min\{n^{1-2\delta} B N^{-2\nu} \pi_{\min} l_j^{-2}, n^{1-\delta} B N^{-\nu} \pi_{\min}^{1/2} l_j^{-1}\}$ and $\log p_2 + \max_j \log l_j + \log N \ll \min\{n^{2\delta}, n^{5(1-2\delta)}, n^{3-5\delta}\}$. Further note $\min\{n^{2\delta}, n^{3/2-3\delta}\} \leq \min\{n^{2\delta}, n^{5(1-2\delta)}, n^{3-5\delta}\}$. Then we have the final conclusion.

C.9 Proof of Theorem 5

The proof of (a) for the RAS is given in Theorem 2 of Wu et al. (2020). In the following we prove (c) for RAS and (b) and (c) for the DC method.

1. Proof of (c) for RAS setting.

Define $\bar{g}_B = B^{-1} \sum_{k=1}^B g(\widehat{\theta}_{(k)})$, $\widehat{\theta}$ be the global moment estimator of θ, and $\bar{\theta}_B = B^{-1} \sum_{k=1}^B \widehat{\theta}_{(k)}$. Then it could be verified that

$$\widehat{\mathrm{SE}}^2 = \frac{n}{B}\Big(\frac{1}{nB} + \frac{1}{N}\Big) \sum_{k=1}^B \Big\{ g(\widehat{\theta}_{(k)}) - R^2_{\mathrm{AVS}}(\mathbb{X}_j) \Big\}^2 = c \sum_{k=1}^B \Big\{ g(\widehat{\theta}_{(k)}) - \bar{g}_B \Big\}^2$$
$$= c \sum_{k=1}^B \Big\{ \dot{g}(\theta)^\top (\widehat{\theta}_{(k)} - \bar{\theta}_B) + \Delta_{(k)} \Big\}^2,$$

where $c \stackrel{\mathrm{def}}{=} nB^{-1}\big\{1/(nB) + 1/N\big\}$, and $\Delta_{(k)}$ is the remaining term of Taylor's expansion, which can be ignored compared to the leading term by Lemma 1. Subsequently, we study the expectation and variance respectively for the leading term $S_B \stackrel{\mathrm{def}}{=} c \sum_{k=1}^B \{\dot{g}(\theta)^\top (\widehat{\theta}_{(k)} - \bar{\theta}_B)\}^2$ respectively.

1.1. Expectation of the leading term.

It could be computed that

$$E\Big[\sum_{k=1}^{B}\big\{\dot g(\theta)^\top(\widehat\theta_{(k)}-\bar\theta_B)\big\}^2\Big] = E\Big[\sum_{k=1}^{B}\big\{\dot g(\theta)^\top(\widehat\theta_{(k)}-\widehat\theta)\big\}^2 - B\big\{\dot g(\theta)^\top(\bar\theta_B-\widehat\theta)\big\}^2\Big]$$
$$\stackrel{\text{def}}{=} E(Q_1-Q_2).$$

We have $E(Q_1) = E\big[E^*\sum_{k=1}^{B}\big\{\dot g(\theta)^\top(\widehat\theta_{(k)}-\widehat\theta)\big\}^2\big] = BE\big[E^*\big\{\dot g(\theta)^\top(\widehat\theta_{(k)}-\widehat\theta)\big\}^2\big] = E(Bn^{-1}\widehat\tau) = Bn^{-1}\tau\{1+o(1)\}$, where $\widehat\tau = \dot g(\theta)^\top\widehat\Sigma_\theta\dot g(\theta)$ and $\widehat\Sigma_\theta$ is the global estimator of Σ_θ. Therefore $B^{-2}E(Q_1) = (nB)^{-1}\tau\{1+o(1)\} = \mathrm{SE}^2$. Next, it could be verified that $B^{-2}E(Q_2) = B^{-1}(nB)^{-1}\tau\{1+o(1)\} = O(B^{-1}E(Q_1))$. As a result, $E(S_B) = \mathrm{SE}^2\{1+O(B^{-1})\}$.

1.2. Variance of the leading term.

By Taylor's expansion, it suffices to derive the variance of S_B. Note that since $\mathrm{var}(S_B) = E\{\mathrm{var}^*(S_B)\} + \mathrm{var}\{E^*(S_B)\}$, where $E^*(\cdot)$ and $\mathrm{var}^*(\cdot)$ are the conditional expectation and variance, we then study the two terms $E\{\mathrm{var}^*(S_B)\}$ and $\mathrm{var}\{E^*(S_B)\}$ separately in *Step 1* and *Step 2*.

Step 1. Expectation of conditional variance.

It could be proved that

$$\mathrm{var}^*\Big[c\sum_{k=1}^{B}\{\eta^\top(\widehat\theta_{(k)}-\bar\theta_B)\}^2\Big] = c^2\mathrm{var}^*\Big[\sum_{k=1}^{B}\Big\{\eta^\top(\widehat\theta_{(k)}-\widehat\theta)\Big\}^2 - B\Big\{\eta^\top(\bar\theta_B-\widehat\theta)\Big\}^2\Big]$$
$$\stackrel{\text{def}}{=} c^2\mathrm{var}^*(E_{11}-E_{12}) = c^2\Big\{\mathrm{var}^*(E_{11}) + \mathrm{var}^*(E_{12}) - 2\,\mathrm{cov}^*(E_{11},E_{12})\Big\},$$

where $\eta = \dot g(\theta)$. It suffices to study $\mathrm{var}^*(E_{11})$ and $\mathrm{var}^*(E_{12})$ respectively. First we have $\mathrm{var}^*(E_{11}) = B\mathrm{var}^*[\{\eta^\top(\widehat\theta_{(k)}-\widehat\theta)\}^2]$. Furthermore it holds $\mathrm{var}^*[\{\eta^\top(\widehat\theta_{(k)}-\widehat\theta)\}^2] \le E^*[\{\eta^\top(\widehat\theta_{(k)}-\widehat\theta)\}^4] = n^{-4}\{nE^*(\eta^\top D_i-\eta^\top\widehat\theta)^4 + 3n(n-1)E^{*2}(\eta^\top D_i-\eta^\top\widehat\theta)^2\} = n^{-3}(\widehat\sigma_4+$

$3(n-1)\widehat{\tau}^2)$, where $\widehat{\sigma}_4 \stackrel{\text{def}}{=} E^*(\eta^\top D_i - \eta^\top \widehat{\theta})^4$. As a result, $\text{var}(E_1) = E\big\{var^*(E_{11})\big\} = O(Bn^{-2})$.

Similarly, we have $\text{var}^*(E_{12}) = B^2\text{var}^*[\{\eta^\top(\bar{\theta}_B - \widehat{\theta})\}^2] \leq B^2 E^*[\{\eta^\top(\bar{\theta}_B - \widehat{\theta})\}^4]$. As a result, $\text{var}(E_{12}) = E\big\{\text{var}^*(E_{12})\big\} = O(B^2(nB)^{-2}) = O(n^{-2}) = o(\text{var}(E_{11}))$. Consequently, we have that $E\{\text{var}^*(S_B)\} = O(1/B)\{1/(nB) + 1/N\}^2$.

Step 2. Variance of conditional expectation.

In this step, we are going to compute $\text{var}\big\{E^*(S_B)\big\}$. It could be proved that

$$\begin{aligned}
E^*(S_B) &= cE^*\Big[\sum_{k=1}^{B}\Big\{\eta^\top(\widehat{\theta}_{(k)} - \widehat{\theta})\Big\}^2 - B\Big\{\eta^\top(\bar{\theta}_B - \widehat{\theta})\Big\}^2\Big] \\
&\stackrel{\text{def}}{=} cE^*(E_{21} - E_{22}) = c\Big\{E^*(E_{21}) - E^*(E_{22})\Big\}.
\end{aligned}$$

Then it suffices to study $E^*(E_{21})$ and $E^*(E_{22})$ respectively. First we have

$$E^*\Big[\sum_{k=1}^{B}\Big\{\eta^\top(\widehat{\theta}_{(k)} - \widehat{\theta})\Big\}^2\Big] = BE^*\Big\{\eta^\top(\widehat{\theta}_{(k)} - \widehat{\theta})\Big\}^2 = Bn^{-1}\widehat{\tau},$$

similarly, we have $E^*\Big[B\Big\{\eta^\top(\bar{\theta}_B - \widehat{\theta})\Big\}^2\Big] = n^{-1}\widehat{\tau}$. Subsequently, we have $\text{var}\big\{cE^*(E_{21})\big\} = \{1/(nB)+1/N\}^2\text{var}(\widehat{\tau}) = \{1/(nB)+1/N\}^2 O(N^{-1})$, and $\text{var}\big\{cE^*(E_{22})\big\} = o(\text{var}\big\{cE^*(E_{21})\big\})$. As a result, we have $\text{var}\big\{E^*(S_B)\big\} = O(N^{-1})\{1/(nB) + 1/N\}^2$.

Combining the results of *Step 1* and *Step 2*, and recall that $\text{SE}^2 = \tau(1/(nB) + 1/N)\{1 + o(1)\}$, which leads to $(\text{SE}^2)^{-2}\text{var}(S_B) = O(B^{-1} + N^{-1})$. We finally have $\widehat{\text{SE}}^2 = \text{SE}^2\{1 + O_p(B^{-1/2} + N^{-1/2})\}$. This accomplishes the proof.

2. Proof of (b) for DC setting.

Under the DC setting, $\widehat{\theta}_{(k)}$ is calculated based on non-overlapped and independent segments. As a consequence, we have

$$\mathrm{var}\Big\{\frac{1}{B}\sum_k g(\widehat{\theta}_{(k)})\Big\} = \frac{1}{B}\mathrm{var}\big\{g(\widehat{\theta}_{(k)})\big\}.$$

Without loss of generality we let $E(X_{ij}) = 0$ and $E(Y_i) = 0$. Define $D_{(k)i} = (X_{(k)ij}Y_{(k)i}, X^2_{(k)ij}, Y^2_{(k)i})^\top$. Therefore $\widehat{\theta}_{(k)} = n^{-1}\sum_{i=1}^n D_{(k)i}$. By using the Taylor's expansion, we have

$$\begin{aligned} g(\widehat{\theta}_{(k)}) &= g(\theta) + \dot{g}(\theta)^\top(\widehat{\theta}_{(k)} - \theta) + (\widehat{\theta}_{(k)} - \theta)^\top \ddot{g}(\xi_{(k)})(\widehat{\theta}_{(k)} - \theta) \\ &= g(\theta) + E_1 + E_2 \end{aligned}$$

where $\xi_{(k)}$ is on the line joining $\widehat{\theta}_{(k)}$ and θ. The variance of the leading term $\dot{g}(\theta)^\top(\widehat{\theta}_{(k)} - \theta)$ is given by $\tau(nB)^{-1}\{1 + o(1)\}$. Further note that E_2 is dominated by E_1 by the proof of Lemma 2. That yields the conclusion (b).

3. Proof of (c) for DC setting.

Recall the definition of $\bar{g}_B$ and $\bar{\theta}_B$. First we have

$$\begin{aligned} \widehat{\mathrm{SE}}^2 &= \frac{1}{B^2}\sum_{k=1}^B \Big\{g(\widehat{\theta}_{(k)}) - \bar{g}_B\Big\}^2 = \frac{1}{B^2}\sum_{k=1}^B \Big\{g(\widehat{\theta}_{(k)}) - \bar{g}_B\Big\}^2 \\ &= \frac{1}{B^2}\sum_{k=1}^B \Big\{\dot{g}(\theta)^\top(\widehat{\theta}_{(k)} - \bar{\theta}_B) + \Delta_{(k)}\Big\}^2, \end{aligned}$$

where $\Delta_{(k)}$ is the remaining term of Taylor's expansion and can be ignored compared to the leading term by Lemma 1. In the following we discuss the expectation and variance respectively for the leading term $S_B^{\mathrm{DC}} \overset{\mathrm{def}}{=} B^{-2}\sum_{k=1}^B\{\dot{g}(\theta)^\top(\widehat{\theta}_{(k)} - \bar{\theta}_B)\}^2$ respectively.

3.1. Expectation of the leading term.

Note that

$$E\Big[\sum_{k=1}^{B}\big\{\dot{g}(\theta)^\top(\widehat{\theta}_{(k)}-\bar{\theta}_B)\big\}^2\Big]=E\Big[\sum_{k=1}^{B}\big\{\dot{g}(\theta)^\top(\widehat{\theta}_{(k)}-\theta)\big\}^2-B\big\{\dot{g}(\theta)^\top(\bar{\theta}_B-\theta)\big\}^2\Big]$$
$$\stackrel{\text{def}}{=}E(Q_3-Q_4).$$

We have $E(Q_3)=Bn^{-1}\dot{g}(\theta)^\top E\{(D_i-\theta)(D_i-\theta)^\top\}\dot{g}(\theta)=Bn^{-1}\tau$. Therefore $B^{-2}E(Q_3)=(nB)^{-1}\tau=\mathrm{SE}^2$. Next, $E(Q_4)=B(nB)^{-1}\tau=o(E(Q_3))$. As a result, $E(S_B^{\mathrm{DC}})=\mathrm{SE}^2\{1+O(B^{-1})\}$.

3.2. Variance of the leading term.

By using the Taylor's expansion result, it suffices to derive the variance of $\sum_{k=1}^{B}\{\eta^\top(\widehat{\theta}_{(k)}-\bar{\theta}_B)\}^2$. Note that we have

$$\mathrm{var}\Big[\sum_{k=1}^{B}\{\eta^\top(\widehat{\theta}_{(k)}-\bar{\theta}_B)\}^2\Big]=\mathrm{var}\Big[\sum_{k=1}^{B}\Big\{\eta^\top(\widehat{\theta}_{(k)}-\theta)\Big\}^2-B\Big\{\eta^\top(\bar{\theta}_B-\theta)\Big\}^2\Big]$$
$$\stackrel{\text{def}}{=}\mathrm{var}(E_3-E_4).$$

It suffices to study $\mathrm{var}(E_3)$ and $\mathrm{var}(E_4)$ respectively. First we have $\mathrm{var}(E_3)=B\mathrm{var}[\{\eta^\top(\widehat{\theta}_{(k)}-\theta)\}^2]$. Furthermore it holds $\mathrm{var}[\{\eta^\top(\widehat{\theta}_{(k)}-\theta)\}^2]\le E[\{\eta^\top(\widehat{\theta}_{(k)}-\theta)\}^4]=n^{-4}\{nE(\eta^\top D_i-\eta^\top\theta)^4+3n(n-1)E^2(\eta^\top D_i-\eta^\top\theta)^2\}=O(n^{-2})$. As a result, $\mathrm{var}(E_3)=O(Bn^{-2})$. Next we have $\mathrm{var}(E_4)=B^2\mathrm{var}[\{\eta^\top(\bar{\theta}_B-\theta)\}^2]\le B^2E[\{\eta^\top(\bar{\theta}_B-\theta)\}^4]=O(B^2(nB)^{-2})=O(n^{-2})$. Consequently, we have for the leading term that $(\mathrm{SE}^2)^{-2}\mathrm{var}(S_B^{\mathrm{DC}})=O(B^{-1})$.

In summary, we have $\widehat{\mathrm{SE}}^2=\mathrm{SE}^2\{1+O_p(B^{-1/2})\}$. In the DC setting we have $N>B$ and $N\ge nB$, hence the final conclusion holds since $B^{-1/2}$ is the leading order.

D Screening Consistency

D.1 Proof of Theorem 6

To prove the screening consistency, we follow the following 4 steps. Let $\theta = (\beta^\top, \gamma_1^{*\top}, \cdots, \gamma_{p_2}^{*\top})^\top$, where $\gamma_j^* = (\pi_{j1}^{1/2}\gamma_{j1}, \cdots, \pi_{j(l_j-1)}^{1/2}\gamma_{j(l_j-1)})$. We prove the case for the AMS measure. For the AVS and DAS measure, the following Step 3 could be slightly revised by using the results from Theorem B.2 and 2.

Step 1. ($\sum_j \mathcal{R}^2_{\mathbb{X}_j} \le c_x$ and $\sum_j \mathcal{R}^2_{\mathbb{Z}_j} \le c_z$)

For convenience, we assume $\mathrm{var}(X_{ij}) = 1$. First, for the quantitative covariates, we have $\sum_j \sigma^2_{xyj} = \sum_{j=1}^p (\theta^\top \Sigma_{\cdot j})^2 = \theta^\top \Sigma^2 \theta \le \lambda_{\max}(\Sigma)(\theta^\top \Sigma \theta) \le \sigma_y^2 \lambda_{\max}(\Sigma) \overset{\mathrm{def}}{=} c_x < \infty$ by Assumption 6, where we have $\theta^\top \Sigma \theta = \mathrm{var}\{X_i^\top \beta + \mathcal{Z}_i^\top \gamma\} \le \sigma_y^2$. Next, note that $\mathcal{R}^2_{\mathbb{Z}_j} = \sigma_y^{-2} \sum_{l=1}^{l_j-1} \pi_{jl}^{-1} \sigma^2_{zy,jl} = \sigma_y^{-2} \sum_{l=1}^{l_j-1} \sigma^{*2}_{zy,jl}$, where $\sigma^*_{zy,jl} = E(\mathcal{Z}_{ijl}, Y_i)/\sqrt{\pi}_{jl}$. Following the same procedure we can show that $\sum_{j,l} \sigma^{*2}_{zy,jl} \le \lambda_{\max}(\Sigma)\sigma_y^2$.

Step 2. ($\max_j |R^2_{\mathbb{X}_j} - \mathcal{R}^2_{\mathbb{X}_j}| = o_p(N^{-\nu})$ and $\max_j |R^2_{\mathbb{Z}_j} - \mathcal{R}^2_{\mathbb{Z}_j}| = o_p(N^{-\nu})$) This can be directly obtained from Theorem 1.

Step 3. ($P(\mathcal{M}_T^\beta \not\subset \widehat{\mathcal{M}}_T^\beta) \to 0$ and $P(\mathcal{M}_T^\gamma \not\subset \widehat{\mathcal{M}}_T^\gamma) \to 0$)

Recall that $R_{\min} = \min\{\min_{j\in\mathcal{M}_T^\beta} \mathcal{R}^2_{\mathbb{X}_j}, \min_{j\in\mathcal{M}_T^\gamma} \mathcal{R}^2_{\mathbb{Z}_j}\}$. Define $\mathcal{M}_T^{\beta*} = \{j : \mathcal{R}^2_{\mathbb{X}_j} > R_{\min}\}$ and $\mathcal{M}_T^{\gamma*} = \{j : \mathcal{R}^2_{\mathbb{Z}_j} > R_{\min}\}$. Immediately we have $\mathcal{M}_T^\beta \subset \mathcal{M}_T^{\beta*}$ and $\mathcal{M}_T^\gamma \subset \mathcal{M}_T^{\gamma*}$. By Assumption 7, we have $R_{\min} \ge 2c_\theta > 0$. In addition, recall that $\widehat{\mathcal{M}}_T^\beta = \{j : R^2_{\mathbb{X}_j} > c_\theta\}$. In this step we show that $\mathcal{M}_T^{\beta*} \subset \widehat{\mathcal{M}}_T^\beta$ and $\mathcal{M}_T^{\gamma*} \subset \widehat{\mathcal{M}}_T^\gamma$ with probability tending to 1.

We first prove $P(\mathcal{M}_T^{\beta*} \not\subset \widehat{\mathcal{M}}_T^\beta) \to 0$. If $\mathcal{M}_T^{\beta*} \not\subset \widehat{\mathcal{M}}_T^\beta$, then there exists at least one j which is not covered by $\widehat{\mathcal{M}}_T^\beta$. As a result, it indicates $R^2_{\mathbb{X}_j} \le 2^{-1} R_{\min}$. However, due to the definition of $\mathcal{M}_T^{\beta *}$, we should have $\mathcal{R}^2_{\mathbb{X}_j} > R_{\min}$. This implies $|R^2_{\mathbb{X}_j} - \mathcal{R}^2_{\mathbb{X}_j}| > R_{\min}/2$. Hence $P(\mathcal{M}_T^{\beta *} \not\subset \widehat{\mathcal{M}}_T^{\beta}) \leq P(\max_j |R^2_{\mathbb{X}_j} - \mathcal{R}^2_{\mathbb{X}_j}| > R_{\min}/2)$. By Theorem 1, we have $P(\max_j |R^2_{\mathbb{X}_j} - \mathcal{R}^2_{\mathbb{X}_j}| > R_{\min}/2) \to 0$ for $R_{\min} = N^{-\nu}$. Therefore $P(\mathcal{M}_T^{\beta *} \not\subset \widehat{\mathcal{M}}_T^{\beta}) \to 0$. Similarly, using the same technique, we can show that $P(\mathcal{M}_T^{\gamma *} \not\subset \widehat{\mathcal{M}}_T^{\gamma}) \to 0$. Hence, (3.2) holds.

Step 4. $(P(\max\{|\widehat{\mathcal{M}}_T^{\beta}|, |\widehat{\mathcal{M}}_T^{\gamma}|\} < m_{\max}) \to 1)$

Define $\mathcal{M}_T^{\beta **} = \{j : \mathcal{R}^2_{\mathbb{X}_j} > R_{\min}/4\}$. Then immediately we have $|\mathcal{M}_T^{\beta **}| R_{\min}/4 \leq \sum_{j \in \mathcal{M}_T^{\beta **}} R_j^2 \leq \tau_{\max}\sigma_y^2$. As a result, $|\mathcal{M}_T^{\beta **}| \leq 4/R_{\min}\tau_{\max}\sigma_y^2 < \infty$. by Assumption 3. If $|\widehat{\mathcal{M}}_T^{\beta}| > |\mathcal{M}_T^{\beta **}|$, we must have at least one $j \in \widehat{\mathcal{M}}_T^{\beta}$ with $R^2_{\mathbb{X}_j} > R_{\min}/2$ but $\mathcal{R}^2_{\mathbb{X}_j} < R_{\min}/4$. Hence it implies $\max_j |R^2_{\mathbb{X}_j} - \mathcal{R}^2_{\mathbb{X}_j}| > R_{\min}/4$. Following Theorem 1, we have $P\{\max_j |R^2_{\mathbb{X}_j} - \mathcal{R}^2_{\mathbb{X}_j}| > R_{\min}/4\} \to 0$ for $R_{\min} = O(N^{-\nu})$, which leads to the final result. Similarly, it holds $P(|\widehat{\mathcal{M}}_T^{\gamma}| < m_{\max}) \to 1$. Hence, (3.3) holds.

D.2 Proof of Lemma 3

Recall that $\mathcal{Z}^*_{ijl} = \mathcal{Z}_{ijl}/\sqrt{\pi_{jl}}$. Then the regression coefficient of $\mathcal{Z}^*_{ijl}$ is $\sqrt{\pi_{jl}}\gamma_{jl}$. Let $\theta = (\beta^\top, \gamma_1^{*\top}, \cdots, \gamma_{p_2}^{*\top})^\top$, where $\gamma_j^* = (\pi_{j1}^{1/2}\gamma_{j1}, \cdots, \pi_{j(l_j-1)}^{1/2}\gamma_{j(l_j-1)})^\top$. Then we have

$$\min\{\min_{j \in \mathcal{M}_T^{\beta}} \mathcal{R}^2_{\mathbb{X}_j}, \min_{j \in \mathcal{M}_T^{\gamma}} \mathcal{R}^2_{\mathbb{Z}_j}\} = \min_{j \in \mathcal{M}_T} (\Sigma_{\cdot j}^\top \theta)^2/\sigma_y^2$$
$$\geq \min_{j \in \mathcal{M}_T} (\sum_{i \in \mathcal{M}_T} \sigma_{ij})^2 \theta^2_{\min}/\sigma_y^2,$$

where $\theta_{\min} = \min_{j \in \mathcal{M}_T} |\theta_j| \geq \min\{|\beta_{\min}|, |\gamma_{\min}|\}$ by Assumption 7, $\mathcal{M}_T$ denotes the true model. The first inequality is due to that $\sigma_{ij} \geq 0$ for $i, j \in \mathcal{M}_T$ and the nonzero model coefficients are positive. This yields the result.

D.3 Proof of Theorem 7

The proof of Theorem 7 follows the same as Theorem 6. We give basic steps here to save space. First the uniform convergence result can be obtained as in Theorem 3 and 4. Next, the screening consistency result can be obtained by following the same proof procedure in Theorem 6.

E Technical Lemmas

E.1 Technical Lemmas for RAS Sampling Scheme

Lemma E.1. *Define the ϕ_α-norm of variable X as*

$$\|X\|_{\phi_\alpha} = \inf\Big\{t > 0 : E\exp\Big(\frac{|X|^\alpha}{t^\alpha}\Big) \leq 2\Big\}.$$

Let $X_1, \cdots, X_n$ be independent variables satisfying $E(X_i) = 0$, $E(X_i^2) = \sigma_i^2$, and $\|X\|_{\phi_\alpha} \leq M$ for some $\alpha \in (0,1] \cup \{2\}$ and A be a symmetric $n \times n$ matrix. For any $t > 0$, we have

$$P\Big(|X^\top AX - tr(A\Sigma_x)| > t\Big) \leq 2\exp\Big\{ - c\min\Big(\frac{t^2}{M^4\|A\|_F^2}, \Big(\frac{t}{M^2\sigma_1(A)}\Big)^{\alpha/2}\Big)\Big\},$$

where $X = (X_1, \cdots, X_n)^\top$, $\Sigma_x = diag\{\sigma_1^2, \cdots, \sigma_n^2\}$.

Proof. The proof is shown in Proposition 1.1 of Götze et al. (2021). □

Lemma E.2. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Exponential distribution with mean μ. Subsamples $\{\widetilde{Z}_1, \cdots, \widetilde{Z}_n\}$ are drawn from $\mathbb{Z}$ independently with replacement. Then for any $t > 0$,* *we have*

$$P\Big(\Big|n^{-1}\sum_{i=1}^{n}\widetilde{Z}_i-\mu\Big|>t\Big)\leq 2\exp\Big(-\frac{nt^2}{3\phi+2Mt/3}\Big)+n\exp(-c_1M)+\exp(-c_2N^{1/2})$$

$\stackrel{\text{def}}{=}\Delta_E^*(t,n,N,M)$, *where* $\phi=E(Z_i^2)$, c_1 *and* c_2 *are finite positive constants. Particularly with* $M=n^{1/2}$ *we have*

$$P\Big(\Big|n^{-1}\sum_{i=1}^{n}\widetilde{Z}_i-\mu\Big|>t\Big)\leq 2\exp\Big(-\frac{nt^2}{3\phi+2\sqrt{n}t/3}\Big)+n\exp(-c_1\sqrt{n})+\exp(-c_2N^{1/2})$$

$\stackrel{\text{def}}{=}\Delta_E(t,n,N)$.

Proof. Consider the event $\mathcal{E}_1=\{\max_i|\widetilde{Z}_i-\mu|<M\}$. Note that $\text{var}(\widetilde{Z}_i|\mathbb{Z})=N^{-1}\sum_i(Z_i-\overline{Z})^2\leq N^{-1}\sum_i Z_i^2\stackrel{\text{def}}{=}\overline{Z^2}$, where $\overline{Z}=N^{-1}\sum_{i=1}^N Z_i$. Under event $\mathcal{E}_1$, we could use Bernstein inequality on $n^{-1}\sum_i(\widetilde{Z}_i-\mu)$ to obtain,

$$P\Big(\big|n^{-1}\sum_i\big(\widetilde{Z}_i-\mu\big)\big|>t\Big|\mathbb{Z},\mathcal{E}_1\Big)\leq 2\exp\Big(-\frac{nt^2}{2\overline{Z^2}+2Mt/3}\Big)$$

Define the event $\mathcal{E}_2=\{\overline{Z^2}-\phi<\phi/2\}$, where $\phi=E(\overline{Z^2})$. Then we have $P(|n^{-1}\sum_{i=1}^n\widetilde{Z}_i-\mu|>t)\leq$

$$E\Big\{P\Big(\big|n^{-1}\sum_i\big(\widetilde{Z}_i-\mu\big)\big|>t\Big|\mathbb{Z},\mathcal{E}_1,\mathcal{E}_2\Big)P(\mathcal{E}_1,\mathcal{E}_2)\Big\}+P(\mathcal{E}_1^c)+P(\mathcal{E}_2^c)\qquad\text{(E.1)}$$

We derive upper bounds for each part of (E.1).

PART 1. For the first part under event $\mathcal{E}_1$ and $\mathcal{E}_2$ immediately we have

$$E\Big\{P\Big(\big|n^{-1}\sum_i\big(\widetilde{Z}_i-\mu\big)\big|>t\Big|\mathbb{Z},\mathcal{E}_1,\mathcal{E}_2\Big)P(\mathcal{E}_1,\mathcal{E}_2)\Big\}\leq 2\exp\Big(-\frac{nt^2}{3\phi+2Mt/3}\Big).$$

Part 2. Next we derive an upper bound on $P(\mathcal{E}_1^c)$. Note that Z_i follows sub-Exponential distribution. Define $\mathrm{var}(Z_i) = \sigma^2$. This implies $P(|Z_i - \mu| > t) \le \exp(-c_1 t)$. Therefore $E\{P(|\widetilde{Z}_i - \mu| \ge M|\mathbb{Z})\} = E\{N^{-1}\sum_i I(|Z_i - \mu| > M)\} = E\{I(|Z_i - \mu| > M)\} \le \exp(-c_1 M/\sigma^2) \le \exp(-c_1 M/\phi)$. By using the maximum inequality we have $P(\max_i |\widetilde{Z}_i - \mu| \ge M) \le n \exp(-c_1 M/\phi)$.

Part 3. By Lemma E.1, we have

$$P\Big(\Big|N^{-1}\sum_{i=1}^{N} Z_i^2 - \phi\Big| > t\Big) \le \exp(-c_3 N^{1/2}),$$

by using $\alpha = 1/2$ and $\phi = E(Z_i^2)$. By setting $t = \phi/2$ and adjusting the constants we could obtain the result.

$\square$

Lemma E.3. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Exponential distribution with mean 0 and variance σ_z^2. Subsamples $\{\widetilde{Z}_1, \cdots, \widetilde{Z}_n\}$ are drawn from $\mathbb{Z}$ independently with replacement. Then for any $t > 0$, we have*

$$P\Big(\Big|n^{-1}\sum_{i=1}^{n} \widetilde{Z}_i^2 - \sigma_z^2\Big| > t\Big) \le 2\exp\Big\{ - c\min\Big(\frac{nt^2}{M^4}, \frac{nt}{M^2}\Big)\Big\} + n\exp(-M)$$

$\overset{\mathrm{def}}{=} \Delta_{E2}(t, n, M)$.

Proof. Define the event $\mathcal{E}_1 = \{\max_i |\widetilde{Z}_i| \le M\}$. Let $\widetilde{Z} = (\widetilde{Z}_1, \cdots, \widetilde{Z}_n)^\top$ and $A = I_n$. Then we could write $\sum_i \widetilde{Z}_i^2 = \widetilde{Z}^\top A \widetilde{Z}$. It could be further verified that $\|A\|_F^2 = O(n)$

and $\sigma_1(A) = 1$. Given $\mathbb{Z}$, we have

$$P\Big(n^{-1}\Big|\sum_{i=1}^{n}\widetilde{Z}_i^2\Big| > t\Big|\mathbb{Z}\Big) \leq 2\exp\Big\{-c\min\Big(\frac{nt^2}{M^4},\frac{nt}{M^2}\Big)\Big\}$$

with $\alpha = 2$ and $\|\widetilde{Z}_i\|_{\phi_\alpha} \leq M$ in Lemma E.1. Next recall that Z_i follows sub-Exponential distribution, then we have $P(\mathcal{E}_1^c) \lesssim n\exp(-M)$.

$\square$

Lemma E.4. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Exponential distribution with mean 0 and variance σ_z^2. Subsamples $\{\widetilde{Z}_1, \cdots, \widetilde{Z}_n\}$ are drawn from $\mathbb{Z}$ independently with replacement. Then for any $t > 0$, we have*

$$P\Big(\Big|n^{-1}\sum_{i=1}^{n}\widetilde{Z}_i^2 - \sigma_z^2\Big| > t\Big) \leq 2\exp\Big(-\frac{nt^2}{3\phi + 2n^\delta t/3}\Big) + n\exp(-c_1 n^{\delta/2}) + \exp(-c_2 N^{1/4})$$

$\stackrel{\text{def}}{=} \Delta_{E2}^*(t, n, N, \delta)$, *where $\phi = E(Z_i^4)$, c_1 and c_2 are finite positive constants.*

Proof. Consider the event $\mathcal{E}_1 = \{\max_i |\widetilde{Z}_i^2 - \sigma_z^2| < n^\delta\}$ for $\delta \geq 0$. Note that $\mathrm{var}(\widetilde{Z}_i^2|\mathbb{Z}) = N^{-1}\sum_i (Z_i^2 - \overline{Z^2})^2 \leq N^{-1}\sum_i Z_i^4$, where $\overline{Z^2} = N^{-1}\sum_{i=1}^N Z_i^2$. Under event $\mathcal{E}_1$, we could use Bernstein inequality on $n^{-1}\sum_i(\widetilde{Z}_i^2 - \sigma_z^2)$ to obtain,

$$P\Big(\big|n^{-1}\sum_i\big(\widetilde{Z}_i^2 - \sigma_z^2\big)\big| > t\Big|\mathbb{Z}, \mathcal{E}_1\Big) \leq 2\exp\Big(-\frac{nt^2}{2\overline{Z^4} + 2n^\delta t/3}\Big)$$

Define the event $\mathcal{E}_2 = \{\overline{Z^4} - \phi < \phi/2\}$, where $\phi = E(\overline{Z^4})$. Then we have $P(|n^{-1}\sum_{i=1}^n \widetilde{Z}_i^2 - \sigma_z^2| > t) \leq$

$$E\Big\{P\Big(\big|n^{-1}\sum_i\big(\widetilde{Z}_i^2 - \sigma_z^2\big)\big| > t\Big|\mathbb{Z}, \mathcal{E}_1, \mathcal{E}_2\Big)P(\mathcal{E}_1, \mathcal{E}_2)\Big\} + P(\mathcal{E}_1^c) + P(\mathcal{E}_2^c) \qquad \text{(E.2)}$$

Recall that Z_i follows sub-Exponential distribution, hence we have $P(\mathcal{E}_1^c) \leq nP(|Z_i| \geq n^{\delta/2}) \leq n\log(-c_1 n^{\delta/2})$. Next, by Proposition 1.1 of Götze et al. (2021) we have

$$P\Big\{|N^{-1}\sum_i Z_i^4 - \phi| > t\Big\} \leq 2\exp\Big(-c\min(Nt^2, (Nt)^{1/4})\Big).$$

By letting $t = \phi/2$ we could obtain the above result.

□

Lemma E.5. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Exponential distribution with mean 0 and variance σ_z^2. Subsamples $\{\widetilde{Z}_1, \cdots, \widetilde{Z}_n\}$ are drawn from $\mathbb{Z}$ independently with replacement. Then for any $t > 0$, we have*

$$P\Big(\Big|\sum_{i\neq j}^{n} \widetilde{Z}_i\widetilde{Z}_j\Big| > t\Big) \leq 2\exp\Big\{-c\min\Big(\frac{t^2}{M^4n^2}, \frac{t}{M^2n}\Big)\Big\} + n\exp(-M).$$

$\stackrel{\text{def}}{=} \Delta_{E3}(t, n, M)$.

Proof. Define the event $\mathcal{E}_1 = \{\max_i |\widetilde{Z}_i| \leq M\}$. Let $\widetilde{Z} = (\widetilde{Z}_1, \cdots, \widetilde{Z}_n)^\top$ and $A = \mathbf{1}_n\mathbf{1}_n^\top - I_n$. Then we could write $\sum_{i\neq j} \widetilde{Z}_i\widetilde{Z}_j = \widetilde{Z}^\top A\widetilde{Z}$. It could be further verified that $\|A\|_F^2 = O(n^2)$ and $\sigma_1(A) = n$. Given $\mathbb{Z}$, we have

$$P\Big(\Big|\sum_{i\neq j}^{n} \widetilde{Z}_i\widetilde{Z}_j\Big| > t\Big|\mathbb{Z}\Big) \leq 2\exp\Big\{-c\min\Big(\frac{t^2}{M^4n^2}, \frac{t}{M^2n}\Big)\Big\}$$

with $\alpha = 2$ and $\|\widetilde{Z}_i\|_{\phi_\alpha} \leq M$ in Lemma E.1. Next recall that Z_i follows sub-Exponential distribution, then we have $P(\mathcal{E}_1^c) \leq n\exp(-M)$.

□

Lemma E.6. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed* *random variables following sub-Exponential distribution with mean* 0 *and variance* σ_z^2. *In the* k*th subsampling round,* $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ *is drawn from* $\mathbb{Z}$ *independently with replacement. Then for any* $t > 0$*, we have*

$$P\Big(\Big|\sum_{k=1}^{B}\sum_{i,j}^{n}\widetilde{Z}_{(k)i}\widetilde{Z}_{(k)j}^2\Big| > t\Big) \leq 2\exp\Big\{-c\min\Big(\frac{t^2}{M^4n^2B}, \frac{t}{M^2n}\Big)\Big\} + nB\exp(-M^{1/2})$$

$\stackrel{\text{def}}{=} \Delta_{E4}^*(t, M).$

Proof. Then proof follows Lemma E.7 and Lemma E.3 but $|\widetilde{Z}_j|^2 \leq M$. □

Lemma E.7. *Let* $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ *be independent and identically distributed random variables following sub-Exponential distribution with mean* 0 *and variance* σ_z^2. *In the* k*th subsampling round,* $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ *is drawn from* $\mathbb{Z}$ *independently with replacement. Then for any* $t > 0$*, we have*

$$P\Big(\Big|\sum_{k=1}^{B}\sum_{i\neq j}^{n}\widetilde{Z}_{(k)i}\widetilde{Z}_{(k)j}\Big| > t\Big) \leq 2\exp\Big\{-c\min\Big(\frac{t^2}{M^4n^2B}, \frac{t}{M^2n}\Big)\Big\} + nB\exp(-M)$$

$\stackrel{\text{def}}{=} \Delta_{E4}(t, M).$

Proof. Let $\widetilde{Z}_{(k)} = (\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n})^\top \in \mathbb{R}^n$ and $\widetilde{Z} = (\widetilde{Z}_{(1)}^\top, \cdots, \widetilde{Z}_{(B)}^\top)^\top \in \mathbb{R}^{nB}$. Define $A = \mathbf{1}_n\mathbf{1}_n^\top - I_n$ and $\mathbb{A} = I_B \times A$. Then we have $\|\mathbb{A}\|_F^2 = O(n^2B)$ and $\sigma_1(\mathbb{A}) = O(n)$. Consequently, by using the same technique as in Lemma E.5 we can obtain the result.

□

Lemma E.8. *Let* $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ *be independent and identically distributed random variables following sub-Exponential distribution. Subsamples* $\{\widetilde{Z}_1, \cdots, \widetilde{Z}_n\}$ *are* *drawn from $\mathbb{Z}$ independently with replacement. Then for any $t > 0$, we have*

$$P\Big(\Big|n^{-1}\sum_{i=1}^{n} Z_i^{*3} - \sigma_{z3}\Big| > t\Big) \leq 2\exp\Big\{-c\min\Big(\frac{nt^2}{M^6}, \frac{nt}{M^3}\Big)\Big\} + n\exp(-M)$$

$\overset{\text{def}}{=} \Delta_{E5}(t, n, M)$, *where* $\sigma_{z3} = E(Z_i^3)$, c_1 *and* c_2 *are finite positive constants.*

Proof. Define the event $\mathcal{E}_1 = \{\max_i |\widetilde{Z}_i| \leq M\}$. Let $Z^* = (\widetilde{Z}_1^{3/2}, \cdots, \widetilde{Z}_n^{3/2})^\top$ and $A = I_n$. Then we could write $\sum_i \widetilde{Z}_i^3 = Z^{*\top} A Z^*$. It could be further verified that $\|A\|_F^2 = O(n)$ and $\sigma_1(A) = 1$. Given $\mathbb{Z}$, we have

$$P\Big(n^{-1}\Big|\sum_{i=1}^{n}\widetilde{Z}_i^3\Big| > t\Big|\mathbb{Z}\Big) \leq 2\exp\Big\{-c\min\Big(\frac{nt^2}{M^6}, \frac{nt}{M^3}\Big)\Big\}$$

with $\alpha = 2$ and $\|Z_i^*\|_{\phi_\alpha} \leq M^{3/2}$ in Lemma E.1. Next recall that Z_i follows sub-Exponential distribution, then we have $P(\mathcal{E}_1^c) \lesssim n\exp(-M)$.

□

Lemma E.9. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following Bernoulli distribution with mean π, i.e., $P(Z_i = 1) = \pi$. Subsamples $\{\widetilde{Z}_1, \cdots, \widetilde{Z}_n\}$ are drawn from $\mathbb{Z}$ independently with replacement. Then for any $t > 0$, we have*

$$P\Big(\Big|n^{-1}\sum_{i=1}^{n}\widetilde{Z}_i - \pi\Big| > t\Big) \leq 2\exp\Big(-\frac{nt^2}{3\pi + 2t/3}\Big) + \exp(-cN) \overset{\text{def}}{=} \Delta_B(t, n, N, \pi)$$

where c is a finite positive constant.

Proof. The proof is the same with the proof of Theorem E.2 but noting that $|Z_i| \leq 1$ and $\mathrm{var}(Z_i) = O(\pi)$. □

Lemma E.10. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Gaussian distribution with mean μ. Subsamples $\{\widetilde{Z}_1, \cdots, \widetilde{Z}_n\}$ are drawn from $\mathbb{Z}$ independently with replacement. Then for any $t > 0$, we have*

$$P\Big(\Big|n^{-1}\sum_{i=1}^{n}\widetilde{Z}_i - \mu\Big| > t\Big) \leq 2\exp\Big(-\frac{nt^2}{3\phi + 2\sqrt{n}t/3}\Big) + n\exp\Big(-\frac{c_1 n}{\phi}\Big) + \exp(-c_2 N)$$

$\stackrel{\text{def}}{=} \Delta_G(t, n, N)$*, where $\phi = E(Z_i^2)$, c_1 and c_2 are finite positive constants.*

Proof. Consider the event $\mathcal{E}_1 = \{\max_i|\widetilde{Z}_i - \mu| < \sqrt{n}\}$. Note that $\text{var}(\widetilde{Z}_i|\mathbb{Z}) = N^{-1}\sum_i(Z_i - \overline{Z})^2 \leq N^{-1}\sum_i Z_i^2 \stackrel{\text{def}}{=} \overline{Z^2}$, where $\overline{Z} = N^{-1}\sum_{i=1}^N Z_i$. Under event $\mathcal{E}_1$, we could use Bernstein inequality on $n^{-1}\sum_i(\widetilde{Z}_i - \mu)$ to obtain,

$$P\Big(\big|n^{-1}\sum_i\big(\widetilde{Z}_i - \mu\big)\big| > t\Big|\mathbb{Z}, \mathcal{E}_1\Big) \leq 2\exp\Big(-\frac{nt^2}{2\overline{Z^2} + 2\sqrt{n}t/3}\Big)$$

Define the event $\mathcal{E}_2 = \{\overline{Z^2} - \phi < \phi/2\}$, where $\phi = E(\overline{Z^2})$. Then we have $P(|n^{-1}\sum_{i=1}^n\widetilde{Z}_i - \mu| > t) \leq$

$$E\Big\{P\Big(\big|n^{-1}\sum_i\big(\widetilde{Z}_i - \mu\big)\big| > t\Big|\mathbb{Z}, \mathcal{E}_1, \mathcal{E}_2\Big)P(\mathcal{E}_1, \mathcal{E}_2)\Big\} + P(\mathcal{E}_1^c) + P(\mathcal{E}_2^c) \tag{E.3}$$

We derive upper bounds for each part of (E.3). The rest of the proof follows the same procedure as in Part 1–3 in Lemma E.2. Specifically, in Part 2, we use the definition of sub-Gaussian definition to obtain $P(\mathcal{E}_1^c) \leq n\exp(-c_1 n/\sigma^2)$. In Part 3, accordingly we obtain $P(\mathcal{E}_2^c) \leq \exp(-c_2 N\phi/\sigma^2)$. Further note that $\sigma^2 \leq \phi$ then the results can be obtained. □

E.2 Technical Lemmas for SAS Sampling Scheme

Lemma E.11. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Exponential/sub-Gaussian distribution with mean 0. Under the SAS sampling scheme, let $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ be the subsample drawn from the kth subsampling round and $\overline{\mathbb{Z}}_{(k)} = n^{-1}\sum_i \widetilde{Z}_{(k)i}$. Then for any $t > 0$, we have*

$$P\Big(\Big|B^{-1}\sum_{k=1}^{B}\overline{\mathbb{Z}}_{(k)}\Big| > t\Big) \le 2\exp\Big(-\frac{Bn^{1/2-\delta}t^2}{2n^{-1/2+\delta}+2t/3}\Big) + N\exp(-cn^{2\delta}),$$

$\overset{\text{def}}{=} \Delta_E^{sas}(t,n,B,N)$*, where c is a positive constant and $\delta \in (0,1/2)$.*

Proof. Since Z_i follows sub-Gaussian distribution, then it also follows sub-Exponential distribution. Then we prove the result for sub-Exponential distribution. Let $\mathbb{Z}_k = (Z_k, Z_{k+1}, \cdots, Z_{k+n-1})^\top$ as a sequential subsample starting from Z_k and $\overline{\mathbb{Z}}_k = n^{-1}\sum_{i=1}^n Z_{k+i-1}$. Consider the event $\mathcal{E}_1 = \{\max_k |\overline{\mathbb{Z}}_k| < n^{-1/2+\delta}\}$ with $\delta \in (0,1/2]$. Note that $\mathrm{var}(\overline{\mathbb{Z}}_{(k)}|\mathbb{Z}) = K^{-1}\sum_k \overline{\mathbb{Z}}_k^2 - \overline{\overline{\mathbb{Z}}}^2 \le n^{-1+2\delta}$ under event $\mathcal{E}_1$. Under the event $\mathcal{E}_1$, we can use the Bernstein inequality to obtain,

$$P\Big(\Big|B^{-1}\sum_{k=1}^{B}\overline{\mathbb{Z}}_{(k)} - \mu\Big| > t\Big|\mathbb{Z},\mathcal{E}_1\Big) \le 2\exp\Big(-\frac{Bt^2}{2n^{-1+2\delta}+2n^{-1/2+\delta}t/3}\Big).$$

Next, since Z_i follows sub-Exponential distribution, then we have $P(|\overline{\mathbb{Z}}_{(k)}| > t) \le \exp(-nt^2/\sigma^2)$ for sufficiently small t. Letting $t = n^{-1/2+\delta}$ we have $P(|\overline{\mathbb{Z}}_{(k)}| > n^{-1/2+\delta}) \le \exp(-n^{2\delta}/(2\sigma^2))$, where $\sigma^2 = \mathrm{var}(Z_i)$. As a result, it leads to $P(\max_k |\overline{\mathbb{Z}}_{(k)}| > t) \le K\exp(-n^{2\delta}/\sigma^2)$, where $K = N - n + 1$. Then it leads to the final result.

□

Lemma E.12. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed* *random variables following sub-Exponential/sub-Gaussian distribution with mean* 0*. Under the SAS sampling scheme, let* $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ *be the subsample drawn from the* k*th subsampling round. Let* $\overline{\mathbb{Z}}_{(k)} = n^{-1}\sum_{i=1}^n \widetilde{\mathbb{Z}}_{(k)i}$*. Then for any* $t > 0$*, we have*

$$P\Big(\Big|B^{-1}\sum_{k=1}^{B}\overline{\mathbb{Z}}_{(k)}^2 - O(n^{-1})\Big| > t\Big) \leq 2\exp\Big(-\frac{Bn^{1-2\delta}t^2}{2n^{-1+2\delta}+2t/3}\Big) + N\exp(-cn^{2\delta}),$$

$\stackrel{\text{def}}{=} \Delta_{E2}^{sas}(t, n, B, N)$*, where* c *is a positive constant,* $n^{-1} \lesssim t$ *and* $\delta \in (0, 1/2)$*.*

Proof. Let $\mathbb{Z}_k = (Z_k, Z_{k+1}, \cdots, Z_{k+n-1})^\top$ as a sequential subsample starting from Z_k and $\overline{\mathbb{Z}}_k = n^{-1}\sum_{i=1}^n Z_{k+i-1}$. Consider the event $\mathcal{E}_1 = \{\max_k |\overline{\mathbb{Z}}_k| < n^{-1/2+\delta}\}$ with $\delta \in (0, 1/2]$. By the Bernstein inequality, we have

$$P\Big(\Big|B^{-1}\sum_{k=1}^{B}\overline{\mathbb{Z}}_{(k)}^2\Big| > t\Big|\mathcal{E}_1, \mathbb{Z}\Big) \leq 2\exp\Big(-\frac{Bt^2}{2\phi + 2n^{-1+2\delta}t/3}\Big)$$

where $\phi = \mathrm{var}(\overline{\mathbb{Z}}_{(k)}^2|\mathbb{Z}) \leq K^{-1}\sum_k \overline{\mathbb{Z}}_k^4 \leq n^{-2+4\delta}$ under the event $\mathcal{E}_1$.

Next, since Z_i follows sub-Exponential distribution, then we have $P(|\overline{\mathbb{Z}}_{(k)}| > t) \leq \exp(-nt^2/\sigma^2)$ for sufficiently small t. Letting $t = n^{-1/2+\delta}$ we have $P(|\overline{\mathbb{Z}}_{(k)}| > n^{-1/2+\delta}) \leq \exp(-n^{2\delta}/(2\sigma^2))$, where $\sigma^2 = \mathrm{var}(Z_i)$. As a result, it leads to $P(\max_k |\overline{\mathbb{Z}}_{(k)}| > t) \leq K\exp(-n^{2\delta}/\sigma^2)$, where $K = N - n + 1$. This leads to the final result.

□

Lemma E.13. *Let* $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ *be independent and identically distributed random variables following sub-Exponential/sub-Gaussian distribution with mean* 0*. Under the SAS sampling scheme, let* $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ *be the subsample drawn from* *the kth subsampling round. Then for any $t > 0$, we have*

$$P\Big(\Big|(nB)^{-1}\sum_{k=1}^{B}\widetilde{Z}^2_{(k)i} - \sigma_z^2\Big| > t\Big) \le 2\exp\Big(-\frac{Bt^2}{2\epsilon^2 + 2\epsilon t/3}\Big) + N\Delta_{E2}(\epsilon, n, M),$$

$\overset{\text{def}}{=} \Delta^{sas}_{E3}(t, n, B, N, M, \epsilon)$*, where* $\Delta_{E2}(\epsilon, n, M)$ *is given in Lemma E.3.*

Proof. Let $\mathbb{Z}_k = (Z_k, Z_{k+1}, \cdots, Z_{k+n-1})^\top$ as a sequential subsample starting from Z_k. Consider the event $\mathcal{E}_1 = \{\max_k |n^{-1}\sum_i \widetilde{Z}^2_{(k)i} - \sigma_z^2| < \epsilon\}$. Define $\overline{D}_{(k)} = n^{-1}\sum_i(\widetilde{Z}^2_{(k)i} - \sigma_z^2)$. By the Bernstein inequality, we have

$$P\Big(\Big|B^{-1}\sum_{k=1}^{B}\overline{D}_{(k)}\Big| > t\Big|\mathcal{E}_1, \mathbb{Z}\Big) \le 2\exp\Big(-\frac{Bt^2}{2\epsilon^2 + 2\epsilon t/3}\Big)$$

$\phi \le K^{-1}\sum_k \overline{D}^2_{(k)} \le \epsilon^2$ under the event $\mathcal{E}_1$.

Next, by using Lemma E.3, we have $P(|\overline{D}_{(k)}| > \epsilon) \le \Delta_{E2}(\epsilon, n, M)$. Using maximum inequality we obtain $P(\max_{1\le k\le K}|\overline{D}_{(k)}| > t) \le N\Delta_{E2}(\epsilon, n, M)$.

□

Lemma E.14. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Exponential distribution with mean 0 and variance σ_z^2. Under the SAS sampling scheme, let $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ be the subsample drawn from the kth subsampling round. Then for any $t > 0$, we have*

$$\begin{aligned}
&P\Big(\frac{1}{n^2 B}\Big|\sum_{k=1}^{B}\sum_{i,j}^{n}\widetilde{Z}_{(k)i}(\widetilde{Z}^2_{(k)j} - \sigma_z^2)\Big| > t\Big) \\
&\le 2\exp\Big(-\frac{Bt^2}{2\epsilon^4 + 2\epsilon^2 t/3}\Big) + N\exp\big(-c\min(n\epsilon^2, (n\epsilon)^{1/2})\big)
\end{aligned}$$

$\overset{\text{def}}{=} \Delta^{sas*}_{E4}(t, n, B, \epsilon)$.

Proof. Let $\mathbb{Z}_k = (Z_k, Z_{k+1}, \cdots, Z_{k+n-1})^\top$ as a sequential subsample starting from Z_k. Define $\overline{\mathbb{Z}}_k = n^{-1}\sum_{i=1}^n Z_{k+i-1}$ and $\overline{\mathbb{Z}_k^2} = n^{-1}\sum_{i=1}^n Z_{k+i-1}^2 - \sigma_z^2$. Consider the event $\mathcal{E}_1 = \{\max_k\{|\overline{\mathbb{Z}}_k|, |\overline{\mathbb{Z}_k^2}|\} < \epsilon\}$ with $\delta \in (0, 1/2)$. By the Bernstein inequality, we have

$$P\Big(\Big|B^{-1}\sum_{k=1}^B \overline{\mathbb{Z}_{(k)}^2}\cdot\overline{\mathbb{Z}}_{(k)}\Big| > t\Big|\mathcal{E}_1, \mathbb{Z}\Big) \le 2\exp\Big(-\frac{Bt^2}{2\phi + 2\epsilon^2 t/3}\Big)$$

where $\phi = \mathrm{var}(\overline{\mathbb{Z}_{(k)}^2}\cdot\overline{\mathbb{Z}}_{(k)}|\mathbb{Z}) \le K^{-1}\sum_k(\overline{\mathbb{Z}_k^2}\cdot\overline{\mathbb{Z}}_k)^2 \le \epsilon^4$ under the event $\mathcal{E}_1$.

Next, since Z_i follows sub-Exponential distribution, then we have $P(|\overline{\mathbb{Z}}_{(k)}| > \epsilon) \le \exp(-n\epsilon^2/\sigma^2)$ for ϵ sufficiently small. In addition, by Lemma E.1, we have $P(|\overline{\mathbb{Z}_{(k)}^2}| > \epsilon) \le 2\exp(-c\min(n\epsilon^2, (n\epsilon)^{1/2}))$. This leads to the final result.

□

Lemma E.15. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Exponential distribution with mean 0 and variance σ_z^2. Under the SAS sampling scheme, let $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ be the subsample drawn from the kth subsampling round. Then for any $t > 0$, we have*

$$\begin{aligned} &P\Big(\frac{1}{n^2B}\Big|\sum_{k=1}^B\sum_{i,j}^n \widetilde{Z}_{(k)i}\widetilde{Z}_{(k)j}\Big| > t\Big) \\ &\le 2\exp\Big(-\frac{Bt^2}{2\epsilon^4 + 2\epsilon^2 t/3}\Big) + N\exp\big(-cn\epsilon^2\big) \end{aligned}$$

$\overset{\text{def}}{=} \Delta_{E4}^{sas}(t, n, B, \epsilon)$.

Proof. The proof follows Lemma E.14 but using Lemma E.1 with $\alpha = 2$. □

Lemma E.16. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following sub-Exponential/sub-Gaussian distribution with mean 0. Under the SAS sampling scheme, let $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ be the subsample drawn from* *the kth subsampling round. Then for any $t > 0$, we have*

$$P\Big(\Big|B^{-1}\sum_{k=1}^{B} n^{-1}\sum_{i}\widetilde{Z}^3_{(k)i} - \sigma_{z3}\Big| > t\Big) \le 2\exp\Big(-\frac{Bt^2}{2\epsilon^2 + 2\epsilon t/3}\Big) + N\Delta_{E5}(\epsilon, n, M),$$

$\stackrel{\text{def}}{=} \Delta^{sas}_{E5}(t, n, B, N, M, \epsilon)$, *where $\sigma_{z3} = E(Z_i^3)$ and $\Delta_{E5}(\epsilon, n, M)$ is defined in Lemma E.8.*

Proof. Let $\mathbb{Z}_k = (Z_k, Z_{k+1}, \cdots, Z_{k+n-1})^\top$ as a sequential subsample starting from Z_k. Consider the event $\mathcal{E}_1 = \{\max_k |n^{-1}\sum_i \widetilde{Z}^3_{(k)i} - \sigma_{z3}| < \epsilon\}$. Define $\overline{D}_{(k)} = n^{-1}\sum_i(\widetilde{Z}^3_{(k)i} - \sigma_{z3})$. By the Bernstein inequality, we have

$$P\Big(\Big|B^{-1}\sum_{k=1}^{B}\overline{D}_{(k)}\Big| > t\Big|\mathcal{E}_1, \mathbb{Z}\Big) \le 2\exp\Big(-\frac{Bt^2}{2\phi + 2\epsilon t/3}\Big)$$

$\phi \le K^{-1}\sum_k \overline{D}^2_{(k)} \le \epsilon^2$ under the event $\mathcal{E}_1$. Next, we obtain $P(\mathcal{E}_1^c) \le N\Delta_{E5}(\epsilon, n, M)$ by Lemma E.8.

□

Lemma E.17. *Let $\mathbb{Z} = \{Z_1, Z_2, \cdots, Z_N\}$ be independent and identically distributed random variables following Bernoulli distribution with mean π, i.e., $P(Z_i = 1) = \pi$. Under the SAS sampling scheme, let $\{\widetilde{Z}_{(k)1}, \cdots, \widetilde{Z}_{(k)n}\}$ be the subsample drawn from the kth subsampling round and $\overline{\mathbb{Z}}_{(k)} = n^{-1}\sum_i \widetilde{Z}_{(k)i}$. Then for any $t > 0$, we have*

$$P\Big(\Big|B^{-1}\sum_{k=1}^{B}\overline{\mathbb{Z}}_{(k)} - \mu\Big| > t\Big) \le 2\exp\Big(-\frac{Bn^{1/2-\delta}t^2}{c_\pi + 2t/3}\Big) + N\exp(-c_2 n^{2\delta}),$$

$\stackrel{\text{def}}{=} \Delta^{sas}_{B}(t, n, B, N, \pi)$, *where $c_\pi = \min\{\pi, n^{-1/2+\delta}\}$, c_1, c_2 are positive constants and $\delta \in (0, 1/2)$.*

Proof. The proof follows the proof of Lemma E.11 but replacing $\mu = \pi$. □

F Additional Numerical Results

In this section, we report some additional numerical results.

F.1 Numerical Results for Statistical Inference under RAS and DC

In this section we compare the statistical inference performances under RAS and DC respectively. Specifically, we consider the same Example 1 and Example 2 in the main text with a fixed $N = 10^4$ and $n = 500$. For the DC setting ,the number of subsamples is automatically set to be $B_{\mathrm{DC}} = N/n = 20$. For the RAS setting, we set $B_{\mathrm{RAS}} = 400$. For each example, we calculate $\widehat{\mathrm{SE}}^2$ for AVS by (3.1) under the two settings respectively. To compare the performance of the automatic statistical inference for the different settings, we present $\mathrm{RSE} \stackrel{\mathrm{def}}{=} \widehat{\mathrm{SE}}^2/\mathrm{SE}^2 - 1$ for the first five variables using boxplot in Figure F.1. For a reliable evaluation, we replicate the experiment for $R = 100$ times. As we can observe from Figure F.1, with larger number of B under RAS setting, we could obtain more accurate estimation of SE^2, and then it could yield more reliable statistical inference result.

F.2 Numerical Results for a Large p Setting

In this section, we report the finite-sample performance of the proposed method which is similar with that in Section 4, but under a larger dimension p setting. Specifically, we consider the same Example 1 and Example 2, but with a fixed dimension $p = 5 \times 10^4$ and $N = 10^4$. To gauge the finite-sample performance, the same measurements in the main text (i.e. SE, Bias, RMSE, AUC and TC) are reported. Experiment is replicated $R = 100$ times. The details are summarized in Table F.1. From Table F.1, it could be observed that the finite-sample performance under a large p setting is consistent with the results stated in Section 4.2.

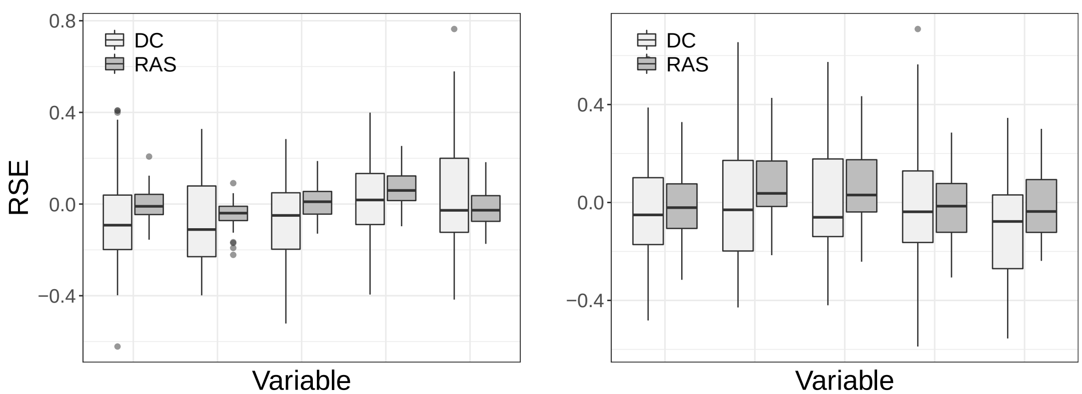

Figure F.1: Boxplot of RSE for the DC (light box) and RAS (dark box) for EXAMPLE 1 (left panel) and EXAMPLE 2 (right panel). Each box is summarized based on $R = 100$ simulation replications.

F.3 Additional Numerical Results in Section 4.2.

In this subsection, we report the remaining numerical results which are not specified in the Section 4.2. Specifically, the Bias, SE and RMSE values of three screening measures (i.e., AVS, DAS, AMS) under RAS scheme with $N = 10^5$ are given in Figure F.2. Correspondingly, the statistical performances under RAS scheme and SAS scheme with $N = 10^6$ are given in Figure F.3 and F.4.

In addition, we report the result for stronger signal case in EXAMPLE 1 by setting $\alpha = 0.04, \rho = 0.1$, and $\sigma = 0.4$. Under this case we can calculate that $\tau_{\max} = 1.22$, $\sigma_y^2 = 0.34$ and $R_{\min} = 2.81 \times 10^{-3}$. Correspondingly we have $m_{\max} = 585$ in Theorem 6–7. Then we set $|\widehat{\mathcal{M}}| = m_{\max}$ and calculate the true model covering rate as TCR$=$ $\sum_{m=1}^{M} I(\mathcal{M}_T \subset \widehat{\mathcal{M}}^{(m)})/M$, where $\widehat{\mathcal{M}}^{(m)}$ denotes the selected model in the mth replicate and $M = 500$. We present the simulation results in Table F.2. The results show that all the methods are able to achieve a relatively high screening accuracy when the signal strength is higher. In addition, the Biases of both DAS and AMS methods are smaller than the AVS method. Furthermore, the selected model is able to consistently cover the true model (with TCP≈ 1) as nB increases, which corroborates with our theoretical findings in Theorem 6–7.

F.4 Detailed Variable Information in Section 4.3.

In this subsection, the detailed variable information for the airline dataset in Section 4.3 are summarized in Table F.3.

References

Cho, H. and Fryzlewicz, P. (2012), "High dimensional variable selection via tilting," *Journal of the Royal Statistical Society: series B (statistical methodology)*, 74, 593–622.

Fan, J. and Lv, J. (2008), "Sure independence screening for ultra-high dimensional feature space (with discussion)," *Journal of the Royal Statistical Society, Series B*, 70, 849–911.

Götze, F., Sambale, H., and Sinulis, A. (2021), "Concentration inequalities for polynomials in α-sub-exponential random variables," *Electronic Journal of Probability*, 26, 1–22.

Wu, S., Zhu, X., and Wang, H. (2020), "Subsampling and Jackknifing: A Practical

Solution for Large Data Analysis with Limited Computational Resources," *Working Paper*.

Zhu, X., Li, F., and Wang, H. (2021), "Least squares approximation for a distributed system," *Journal of Computational and Graphical Statistics*, 1–44.

Table F.1: Simulation results for EXAMPLE 1 and 2 under different sampling schemes. The total sample size and parameter dimension are fixed at $N = 10^4$ and $p = 5 \times 10^4$, respectively. The numerical performance is evaluated for different n and B. For the proposed three measures, the SE, Bias, RMSE and AUC are reported. Finally, TC is also reported.

Sch	n	B	SE ($\times 10^{-3}$)			Bias ($\times 10^{-2}$)			RMSE ($\times 10^{-2}$)			AUC(%)			TC
			R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	
								EXAMPLE 1							
RAS	50	100	2.711	2.867	0.365	2.052	0.062	0.031	2.071	0.297	0.048	96.38	96.14	98.66	584.01
		200	1.915	2.027	0.217	2.050	0.051	0.020	2.059	0.212	0.030	97.18	97.01	98.62	1138.20
	100	100	1.367	1.407	0.214	1.019	0.029	0.020	1.029	0.145	0.030	96.78	96.71	98.66	746.23
		200	0.966	0.994	0.135	1.020	0.021	0.015	1.025	0.103	0.021	97.62	97.57	98.70	1506.29
SAS	50	100	2.704	2.863	0.366	2.039	0.151	0.031	2.057	0.339	0.048	94.47	94.22	98.01	443.25
		200	1.890	2.001	0.216	2.041	0.147	0.021	2.050	0.263	0.030	96.51	96.25	98.59	889.09
	100	100	1.346	1.387	0.213	1.009	0.099	0.020	1.018	0.180	0.030	97.39	97.17	98.64	511.20
		200	0.942	0.970	0.131	1.009	0.097	0.015	1.013	0.147	0.020	98.18	97.99	98.85	1000.50
								EXAMPLE 2							
RAS	50	100	4.083	4.545	1.474	4.119	0.090	0.048	4.140	0.468	0.157	79.93	93.84	99.92	1172.59
		200	3.048	3.331	1.147	4.134	0.063	0.048	4.146	0.342	0.125	79.99	97.61	99.99	2355.70
	100	100	2.099	2.194	1.033	2.043	0.042	0.031	2.055	0.226	0.109	81.98	99.61	99.99	1506.50
		200	1.536	1.600	0.727	2.053	0.029	0.028	2.058	0.164	0.079	82.36	99.99	100.00	3034.18
SAS	50	100	4.014	4.456	1.504	4.005	0.240	0.064	4.026	0.530	0.165	80.24	94.58	99.97	840.56
		200	2.790	3.126	0.908	3.971	0.248	0.019	3.981	0.424	0.094	79.98	97.92	100.00	1665.57
	100	100	2.110	2.195	1.001	1.954	0.160	0.029	1.966	0.288	0.106	83.42	99.37	100.00	975.79
		200	1.436	1.502	0.635	1.937	0.165	0.011	1.943	0.238	0.065	82.53	99.91	100.00	1926.33

Table F.2: Simulation results for Example 1 with stronger signal strength. The parameter dimension are fixed at $p = 1000$. The numerical performance is evaluated for different n, B and N. For the proposed three measures, the SE, Bias, RMSE, AUC and TCR are reported. Finally, TC is also reported.

Sch	n	B	SE ($\times 10^{-3}$)			Bias ($\times 10^{-2}$)			RMSE ($\times 10^{-2}$)			AUC(%)			TCR(%)			TC
			R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	R^2_{AVS}	R^2_{DAS}	R^2_{AMS}	
									$N = 10^5$									
RAS	100	50	2.063	2.123	0.430	1.010	0.014	0.024	1.031	0.213	0.050	97.67	97.66	99.94	66.20	66.40	99.40	15.47
		80	1.628	1.677	0.296	1.011	0.014	0.016	1.024	0.169	0.034	98.85	98.85	99.99	85.50	85.50	100.00	24.90
	200	50	1.060	1.075	0.249	0.503	0.007	0.013	0.515	0.108	0.029	99.68	99.69	100.00	97.00	97.20	100.00	20.39
		80	0.838	0.850	0.177	0.503	0.008	0.010	0.510	0.086	0.021	99.82	99.83	100.00	98.00	97.00	100.00	36.41
SAS	100	50	2.060	2.121	0.430	1.010	0.033	0.023	1.031	0.216	0.050	97.73	97.70	99.94	72.20	72.60	99.60	5.54
		80	1.623	1.670	0.295	1.010	0.034	0.016	1.023	0.172	0.034	98.84	98.83	99.99	84.50	85.00	100.00	8.22
	200	50	1.058	1.074	0.249	0.502	0.023	0.014	0.514	0.111	0.029	99.68	99.68	100.00	96.40	96.40	100.00	5.33
		80	0.836	0.848	0.177	0.502	0.023	0.010	0.510	0.089	0.021	99.87	99.88	100.00	99.50	99.50	100.00	9.91
									$N = 10^6$									
RAS	100	50	2.051	2.111	0.417	1.009	0.016	0.021	1.030	0.212	0.048	98.17	98.13	99.98	88.00	87.50	100.00	15.86
		80	1.630	1.678	0.285	1.009	0.013	0.014	1.023	0.169	0.033	99.10	99.09	100.00	90.00	89.50	100.00	26.92
	200	50	1.057	1.072	0.239	0.501	0.007	0.011	0.513	0.108	0.027	99.81	99.81	100.00	99.00	98.50	100.00	23.00
		80	0.834	0.846	0.165	0.502	0.006	0.007	0.509	0.085	0.019	99.94	99.94	100.00	100.00	100.00	100.00	37.26
SAS	100	50	2.060	2.120	0.416	1.009	0.016	0.021	1.030	0.213	0.048	97.98	97.95	99.97	75.40	75.40	100.00	4.91
		80	1.623	1.671	0.285	1.008	0.017	0.013	1.022	0.168	0.033	99.00	98.99	100.00	88.50	87.00	100.00	8.43
	200	50	1.058	1.073	0.238	0.502	0.009	0.011	0.513	0.108	0.027	99.81	99.81	100.00	99.00	98.80	100.00	5.86
		80	0.834	0.846	0.166	0.502	0.009	0.007	0.509	0.085	0.019	99.92	99.92	100.00	100.00	100.00	100.00	10.43

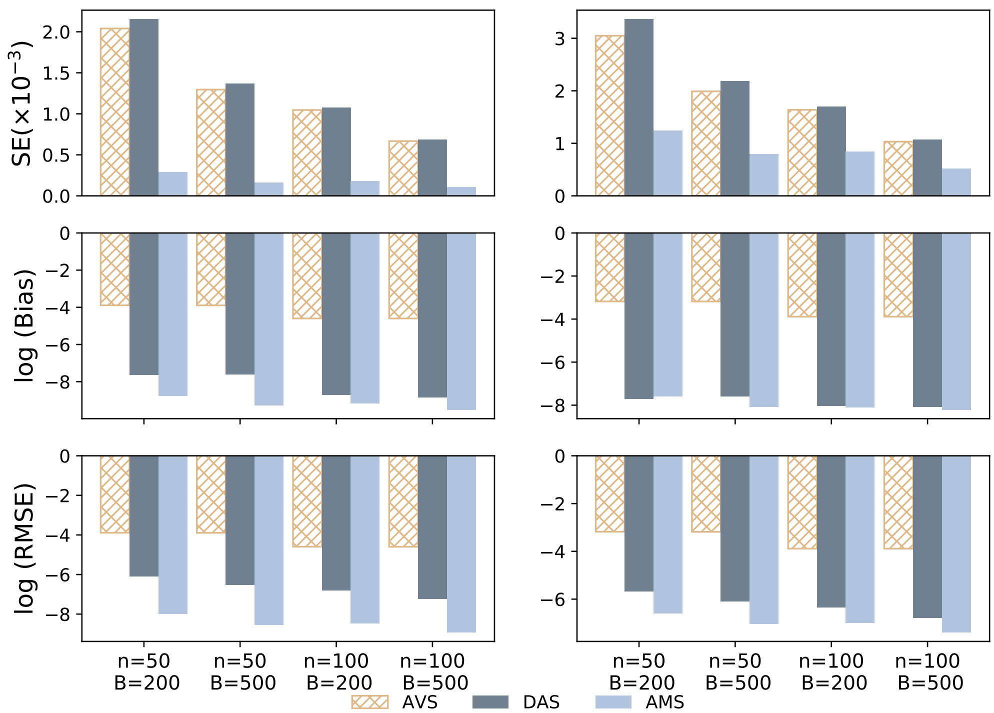

Figure F.2: Bar chart of SE, log(Bias) and log(RMSE) values for the AVS, DAS and AMS measures for different (n, B) under the RAS sampling scheme for Example 1 (left panels) and Example 2 (right panels). The sample size N is fixed to $N = 10^5$.

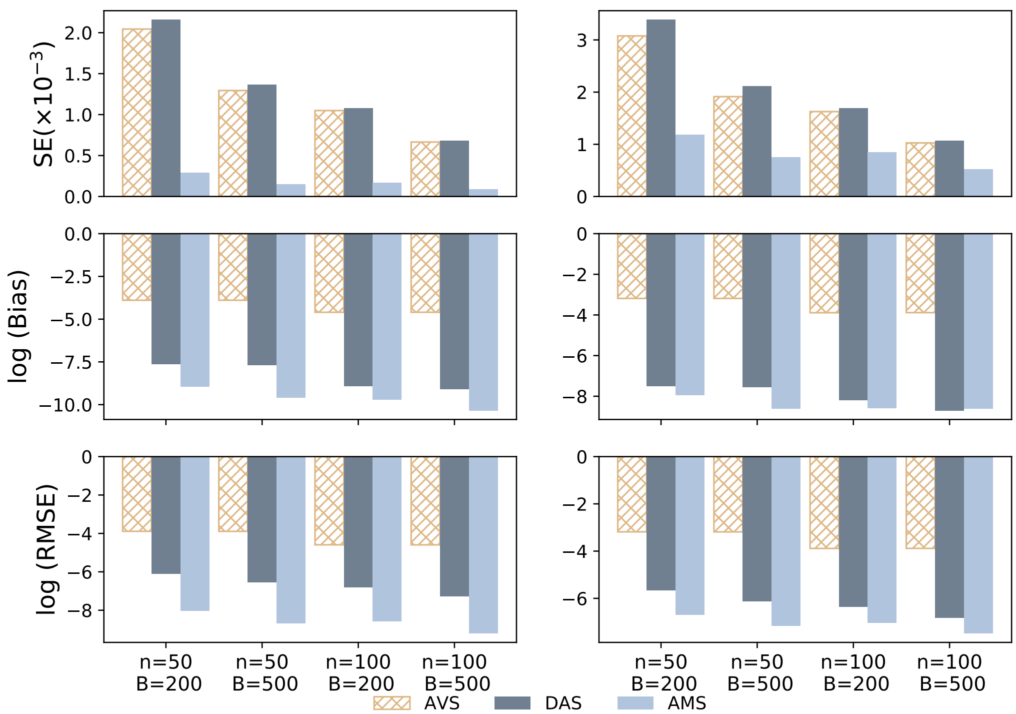

Figure F.3: Bar chart of SE, log(Bias) and log(RMSE) values for the AVS, DAS and AMS measures for different (n, B) under the RAS sampling scheme for Example 1 (left panels) and Example 2 (right panels). The sample size N is fixed to $N = 10^6$.

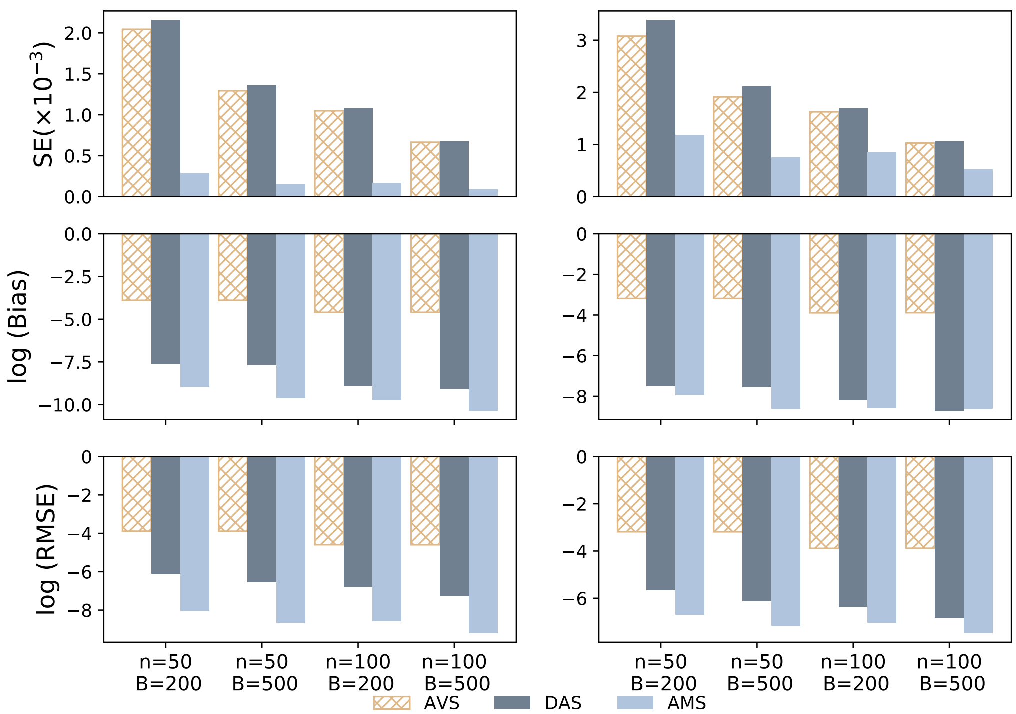

Figure F.4: Bar chart of SE, log(Bias) and log(RMSE) values for the AVS, DAS and AMS measures for different (n, B) under the SAS sampling scheme for Example 1 (left panels) and Example 2 (right panels). The sample size N is fixed to $N = 10^6$.

Table F.3: Variable description for the U.S. airline data after data preprocessing.

Variable	Description	Variable used in the model
ArrTime	Actual arrival time	Used as the response variable
Year	Year between 2004 and 2008	Used as a numerical variable
Month	Month of the year	Used as a categorical variable with 11 levels
DayofMonth	Day of the month	Used as a numerical variable
DayofWeek	Day of the week	Used as a categorical variable with 6 levels
CRSDepTime	Scheduled departure time	Used as a numerical variable
CRSArrTime	Scheduled arrival time	Used as a numerical variable
ELapsedTime	Actual elapsed time	Used as a numerical variable
Distance	Distance between the origin and destination in miles	Used as a numerical variable
Carrier	Flight carrier code for 29 carriers	Used as a categorical variable with top 7 carries
Dest	Destination of the flight (total of 348 categories)	Coded into top 10 states and others
Origin	Destination of the flight (total of 343 categories)	Coded into top 10 states and others
$\mathsf{ArrTime}_\mathsf{d}$	The d-th order lag of the actual arrival time for $1 \leq d \leq 40$	Used as a numerical variable
$\mathsf{ElapsedTime}_\mathsf{d}$	The d-th order lag of the actual elapsed time for $1 \leq d \leq 40$	Used as a numerical variable