Distributed sequential federated estimation

Zhanfeng Wang, Xinyu Zhang, Yuan-chin Ivan Chang

School of Management, University of Science and Technology of China, Hefei, China

Academy of Mathematics and Systems Science, Chinese Academy of Sciences

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan

Supplementary Material

Supplementary material contains a detailed proof of the main results and additional numerical results.

A1 Proofs of main results

Throughout the proofs, notation C denotes a generic positive constant, which does not depend on sample size k.

A. Properties of sequential estimation for data site j

Once the stopping criterion (8) is satisfied for data site j, we record the stopping time and cease sampling. The confidence ellipsoid for $\boldsymbol{\theta}_0$ is then

computed:

$$R_{\tilde{N}_j} = \{z \in R^{p_0} : \frac{S_{\tilde{N}_j}}{\tilde{N}_j} \leq \frac{d_1^2}{\mu_{j\tilde{N}_j}}\}, \tag{A1.1}$$

where $S_{\tilde{N}_j} = (\boldsymbol{z} - \tilde{\boldsymbol{\theta}}_{j\tilde{N}_j})^\top (\boldsymbol{L}_j \boldsymbol{\Sigma}_{j\tilde{N}_j}^{-1} \boldsymbol{L}_j^\top)^{-1} (\boldsymbol{z} - \tilde{\boldsymbol{\theta}}_{j\tilde{N}_j})$, and $\boldsymbol{z} = (z_1, \cdots, z_{p_0})^\top$. Length of maximum axis of this ellipsoid is

$$D = 2 \left(\frac{\tilde{N}_j d_1^2}{\mu_{j\tilde{N}_j}} \right)^{1/2} \lambda_{max}^{1/2} (\tilde{N}_j (\boldsymbol{L}_j \boldsymbol{\Sigma}_{j\tilde{N}_j}^{-1} \boldsymbol{L}_j^\top)) = 2d_1,$$

where $\lambda_{max}(\boldsymbol{A})$ is the maximum eigenvalue of matrix $\boldsymbol{A}$.

Proof of Proposition 1 Since $\tilde{a}_j^2 > 0$ for all j and are constants, the definition of the stopping time implies that for each j, the ratio $\tilde{N}j/\hat{N}$ converges to a constant $\gamma_j > 0$. Let $\theta^* = \sum_{j=1}^M w_j \tilde{\boldsymbol{\theta}}_{j\tilde{N}j}$, where w_j, $\sum_{j=1}^M w_j = 1$, represent given weights. Then, the variance of θ^* is given by:

$$\sum_{j=1}^M w_j^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{j\tilde{N}_j}^{-1} \boldsymbol{L}_j^\top = \sum_{j=1}^M w_j^2 \tilde{N}_j^{-1} \boldsymbol{L}_j (\boldsymbol{\Sigma}_{j\tilde{N}_j}/\tilde{N}_j)^{-1} \boldsymbol{L}_j^\top.$$

Let

$$G_N(w_1, \cdots, w_M) = \hat{N} \sum_{j=1}^M w_j^2 \tilde{N}_j^{-1} \boldsymbol{L}_j (\boldsymbol{\Sigma}_{j\tilde{N}_j}/\tilde{N}_j)^{-1} \boldsymbol{L}_j^\top,$$

then it follows that as d_1 tends to 0,

$$G_N(w_1, \cdots, w_M) \longrightarrow G(w_1, \cdots, w_M) = \sum_{j=1}^M w_j^2 \gamma_j^{-1} \boldsymbol{L}_j \boldsymbol{\Sigma}_j^{-1} \boldsymbol{L}_j^\top.$$

We then minimize $Tr(G(w_1, \cdots, w_M))$ with respect to $w_1, \cdots, w_M$ subject to $\sum_j w_j = 1$, and obtain that

$$w_j = \frac{\gamma_j Tr(\boldsymbol{L}_j \boldsymbol{\Sigma}_j^{-1} \boldsymbol{L}_j^\top)^{-1}}{\sum_{k=1}^M \gamma_k Tr(\boldsymbol{L}_k \boldsymbol{\Sigma}_k^{-1} \boldsymbol{L}_k^\top)^{-1}}, \qquad j = 1, \cdots, M.$$

If we use the same covariates for all data sites, then they have a homogeneous covariance $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_M$, asymptotically. It follows that $w_j = \gamma_j$, $j = 1, \cdots, M$. It follows that $\hat{\boldsymbol{\theta}}$ with the weights ρ_j achieves the minimal covariance, asymptotically.

Let $\boldsymbol{Z} = (z_1, \cdots, z_{p_0})^T$, then

$$R_{j\tilde{N}_j} = \left\{ \boldsymbol{Z} \in R^{p_0} : \ \frac{S_{j\tilde{N}_j}}{\tilde{N}_j} \leq \frac{d_1^2}{\mu_{j\tilde{N}_j}} \right\} \tag{A1.2}$$

defines a confidence set for $\boldsymbol{\theta}_0$, where $S_{j\tilde{N}_j} = (\boldsymbol{Z} - \tilde{\boldsymbol{\theta}}_{j\tilde{N}_j})^\top (\boldsymbol{L}_j \tilde{\boldsymbol{\Sigma}}_{j\tilde{N}_j}^{-1} \boldsymbol{L}_j^\top)^{-1} (\boldsymbol{Z} - \tilde{\boldsymbol{\theta}}_{j\tilde{N}_j})$, $\mu_{j\tilde{N}_j} = \lambda_{max} \left(\tilde{N}_j \boldsymbol{L}_j \tilde{\boldsymbol{\Sigma}}_{j\tilde{N}_j}^{-1} \boldsymbol{L}_j^\top \right)$, and $\boldsymbol{\Sigma}_{j\tilde{N}_j} = \sum_{i=1}^{\tilde{N}_j} \boldsymbol{x}_{ji} \{ \dot{\mu}(\boldsymbol{x}_{ji}^\top \tilde{\boldsymbol{\beta}}_{j\tilde{N}_j})^2 / \nu(\boldsymbol{x}_{ji}^\top \tilde{\boldsymbol{\beta}}_{j\tilde{N}_j}) \} \boldsymbol{x}_{ji}^\top$. We can show that the length of the maximum axis of the confidence set $R_{j\tilde{N}_j}$ is $2d_1$.

Lemma 1. *Assume that the conditions of Theorem 1 are satisfied, and $\tilde{N}_j$ is as defined in (8). Then*

$$\lim_{d_1 \to 0} \frac{d_1^2 \tilde{N}_j}{\tilde{a}_j^2 \mu_j} = 1 \quad \text{almost surely,} \tag{A1.3}$$

$$\lim_{d_1 \to 0} P(\theta_0 \in R_{j\tilde{N}_j}) = 1 - \tilde{\alpha}_j, \tag{A1.4}$$

$$\lim_{d_1 \to 0} \frac{d^2 E(\tilde{N}_j)}{\tilde{a}_j^2 \mu_j} = 1, \tag{A1.5}$$

where $\tilde{\alpha}_j$ satisfies $P(\chi_{p_0}^2 > \tilde{a}_j^2) = \tilde{\alpha}_j$, and μ_j is the maximum eigenvalue of matrix $\boldsymbol{L}_j \boldsymbol{\Sigma}_j^{-1} \boldsymbol{L}_j^\top$.

Proof. Wang and Chang (2013) established the asymptotic consistency and efficiency of the sequential estimation method for a single linear regression

model. Furthermore, the authors extended their findings by providing a rigorous proof of this lemma for the logistic regression model.

Let

$$z_k = \frac{k\mu_j}{\mu_{jk}}, \quad f(k) = k, \quad t = \frac{\tilde{a}_j^2}{d_1^2 \mu_j},$$

where $\mu_{jk} = \lambda_{max}(\boldsymbol{L}_j \boldsymbol{\Sigma}_{jk}^{-1} \boldsymbol{L}_j^\top)$. Then the stopping rule $\tilde{N}_j$ in (8) becomes

$$\tilde{N}_j = \min \left\{ k : k \geq 1 \text{ and } y_k \leq f(k)/t, \text{ and } v_{A_j} \leq \left(\frac{d_2}{a_p} \right)^2 \right\}.$$

The estimate $\tilde{\boldsymbol{\beta}}_{jk}$ is the maximum quasi-likelihood estimate (MQLE) of $\boldsymbol{\beta}_j$ (McCullagh and Nelder, 1989). Then under conditions (A1) and (A2), $\tilde{\boldsymbol{\beta}}_{jk}$ is strongly consistent estimate of $\boldsymbol{\beta}_j$ (Chen et al., 1999; Chang, 1999). From Lemma 1 in Chow and Robbins (1965), we establish the following relationship:

$$1 = \lim_{t \to \infty} \frac{f(N)}{t} = \lim_{d_1 \to 0} \frac{d^2 N \mu_j}{\tilde{a}_j^2} \quad \text{almost surely,}$$

which implies (i).

Property of the uniform continuity in probability (u.c.i.p.) is a sufficient condition such that estimate of parameter with randomly stopped sequence has the same asymptotic distribution as the fixed sample size estimate (Woodroofe, 1982). We know that MQLE $\tilde{\boldsymbol{\beta}}_{jk}$ and $\hat{A}_j$ are strongly consistent estimates of $\boldsymbol{\beta}_j$ and A_j, respectively, which shows that $\tilde{\boldsymbol{\beta}}_{jk}$ and $\hat{A}_j$ have u.c.i.p. properties (Chang and Martinsek, 1992; Chang, 2011).

Thence, leveraging the u.c.i.p. properties of $\tilde{\boldsymbol{\beta}}_{jk}$ and $\hat{A}_j$, we can establish the following essential asymptotic properties:

$$\sqrt{\tilde{N}_j}(\tilde{\boldsymbol{\theta}}_{j\tilde{N}_j} - \boldsymbol{\theta}_0) \longrightarrow N(0, \boldsymbol{L}_j \boldsymbol{\Sigma}_j^{-1} \boldsymbol{L}_j^\top) \quad \text{in distribution as } d_1 \to 0,$$
$$\frac{\hat{A}_j - A_j}{\sqrt{v_{Aj}}} \longrightarrow N(0,1) \quad \text{in distribution as } d_2 \to 0.$$

Hence, we have

$$(\tilde{\boldsymbol{\theta}}_{j\tilde{N}_j} - \boldsymbol{\theta}_0)^\top \left[\boldsymbol{L}_j \boldsymbol{\Sigma}_{jN_j}^{-1} \boldsymbol{L}_j^\top\right]^{-1} (\tilde{\boldsymbol{\theta}}_{j\tilde{N}_j} - \boldsymbol{\theta}_0) \longrightarrow \chi^2_{p_0}, \text{ as } d_1 \to 0.$$

It follows that

$$\begin{aligned}
&\lim_{d_1 \to 0} P(\boldsymbol{\theta}_0 \in R_{j\tilde{N}_j}) \\
=&\lim_{d_1 \to 0} P\left((\tilde{\boldsymbol{\theta}}_{j\tilde{N}_j} - \boldsymbol{\theta}_0)^\top \left[\boldsymbol{L}_j \boldsymbol{\Sigma}_{jN_j}^{-1} \boldsymbol{L}_j^\top\right]^{-1} (\tilde{\boldsymbol{\theta}}_{j\tilde{N}_j} - \boldsymbol{\theta}_0) \leq \frac{\tilde{N}_j d_1^2}{\mu_{j\tilde{N}_j}}\right) \\
=&\lim_{d_1 \to 0} P\left((\tilde{\boldsymbol{\theta}}_{j\tilde{N}_j} - \boldsymbol{\theta}_0)^\top \left[\boldsymbol{L}_j \boldsymbol{\Sigma}_{jN_j}^{-1} \boldsymbol{L}_j^\top\right]^{-1} (\tilde{\boldsymbol{\theta}}_{j\tilde{N}_j} - \boldsymbol{\theta}_0) \leq \tilde{a}_j^2\right) \\
=&1 - \tilde{\alpha}_j,
\end{aligned}$$

which confidently affirm the validity of result (ii).

Now, let us proceed with the proof of (iii). To begin, let us define a last time

$$L_\rho = \sup\{k \geq 1 : \|ln(\boldsymbol{\beta}) - ln(\boldsymbol{\beta}_j)\| < \|ln(\boldsymbol{\beta}_j)\|, \ \exists \boldsymbol{\beta} \in \partial \mathcal{B}_{j\rho}\},$$

where $\mathcal{B}_{j\rho} = \{\boldsymbol{\beta} : ||\boldsymbol{\beta} - \boldsymbol{\beta}_j|| \leq \rho\}$ and

$$ln(\boldsymbol{\beta}) \equiv \sum_{i=1}^{k} \dot{\mu}(\boldsymbol{x}_{ji}^\top \boldsymbol{\beta}) w(\boldsymbol{x}_{ji}^\top \boldsymbol{\beta})[y_{ji} - \mu(\boldsymbol{x}_{ji}^\top \boldsymbol{\beta})]\boldsymbol{x}_{ji}.$$

Hence, if $k > L_\rho$, then we have $\|ln(\boldsymbol{\beta}) - ln(\boldsymbol{\beta}_j)\| \geq \|ln(\boldsymbol{\beta}_j)\|$ for $\forall \boldsymbol{\beta} \in \partial\mathcal{B}_{j\rho}$, which implies that

$$n > L_\rho \Rightarrow \inf_{\boldsymbol{\beta}\in\partial\mathcal{B}_{j\rho}} \|ln(\boldsymbol{\beta}) - ln(\boldsymbol{\beta}_j)\| \geq \|ln(\boldsymbol{\beta}_j)\| \quad \text{with probability one.}$$

Based on Lemma 3 from Yin et al. (2006), the existence of the root $\tilde{\boldsymbol{\beta}}_{jn}$ for $ln(\boldsymbol{\beta}) = 0$ has been established. Moreover, leveraging result (i), which implies that from (i), we know that $\lim_{d_1\to 0} d_1^2\tilde{N}_j/(\tilde{a}_j^2\mu_j) = 1$ almost surely. Hence, we only have to prove that $\{d_1^2\tilde{N}_j : \ d_1 \in (0,1)\}$ is uniformly integrable. Similar to proof of Theorem 3.2 in Chang (2001), to show that for $d_1 \in (0,1)$, we have

$$\begin{aligned} d_1^2\tilde{N}_j &= d_1^2\tilde{N}_j I\{\tilde{N}_j > L_\rho\} + d_1^2\tilde{N}_j I\{\tilde{N}_j \leq L_\rho\} \\ &\leq d_1^2\left\{\left\lceil\frac{c\tilde{a}_j^2}{d_1^2}\right\rceil + 1\right\} + L_\rho \leq \lceil c\tilde{a}_j^2\rceil + 1 + L_\rho \ . \end{aligned}$$

It easily shows that

$$\sum_{k=0}^{\infty} P(L_\rho \geq k) = \sum_{k=0}^{k'} P(L_\rho \geq k) + \sum_{k=k'+1}^{\infty} P(L_\rho \geq k) \leq k' + 0 < \infty, \tag{A1.6}$$

where k' is a number sufficiently large such that when $k > k'$, $\inf_{\boldsymbol{\beta}\in\partial\mathcal{B}_{j\rho}} \|ln(\boldsymbol{\beta}) - ln(\boldsymbol{\beta}_j)\| \geq \|ln(\boldsymbol{\beta}_j)\| \geq c$ with some $c > 0$ (Yin et al., 2006). Hence, (A1.6) implies $EL_\rho < \infty$. And then $\{d^2T_d : d \in (0,1)\}$ is uniformly integrable which indicates that (ii) holds.

B. Proof of Theorem 1

If the stopping criterion defined in (8) holds for the procedure j , then we have used $\tilde{N}_j$ observations, and obtained a confidence set $R_{\tilde{N}_j}$ as defined in (A1.1). Via Lemma 1,

$$\lim_{d_1 \to 0} \frac{d_1^2 \tilde{N}_j}{\tilde{a}_j^2 \mu_j} = 1 \quad \text{almost surely,} \tag{A1.7}$$

$$\lim_{d_1 \to 0} P(\theta_0 \in R_{j\tilde{N}_j}) = 1 - \tilde{\alpha}_j, \tag{A1.8}$$

$$\lim_{d_1 \to 0} \frac{d_1^2 E(\tilde{N}_j)}{\tilde{a}_j^2 \mu_j} = 1, \tag{A1.9}$$

where $\tilde{\alpha}_j$ satisfies $P(\chi^2_{p_0} > \tilde{a}_j^2) = \tilde{\alpha}_j$. Equations (A1.7) and (A1.9) implies that as $d_1 \to 0$,

$$d_1^2 \tilde{N}_j \longrightarrow \tilde{a}_j^2 \mu_j \quad \text{almost surely,}$$

$$d_1^2 E(\tilde{N}_j) \longrightarrow \tilde{a}_j^2 \mu_j.$$

Because $\tilde{a}_1^2 + \tilde{a}_2^2 = a^2$, it follows that

$$d_1^2 \hat{N} = d_1^2(\tilde{N}_1 + \tilde{N}_2) \longrightarrow (\tilde{a}_1^2 + \tilde{a}_2^2)\mu = a^2\mu \quad \text{almost surely,}$$

$$d_1^2 E(\hat{N}) = d_1^2 E(\tilde{N}_1 + \tilde{N}_2) \longrightarrow (\tilde{a}_1^2 + \tilde{a}_2^2)\mu = a^2\mu,$$

which implies that

$$\lim_{d_1 \to 0} \frac{d_1^2 \hat{N}}{a^2 \mu} = 1 \quad \text{almost surely,}$$

$$\lim_{d_1 \to 0} \frac{d_1^2 E(\hat{N})}{a^2 \mu} = 1.$$

For simplicity, we suppose that $j = 1, 2$, in the following matrix algebra, which can be easily extended to the case for $j = 1, \cdots, M$. Matrix $\boldsymbol{L}_j \boldsymbol{\Sigma}_{j\tilde{N}_j}^{-1} \boldsymbol{L}_j^\top$ can be used to estimate covariance of $\tilde{\boldsymbol{\theta}}_{j\tilde{N}_j}$. Since two procedures are independent, we can use $\rho_1^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{1N_1}^{-1} \boldsymbol{L}_j^\top + \rho_2^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{2N_2}^{-1} \boldsymbol{L}_j^\top$ to estimate variance of $\hat{\boldsymbol{\theta}}$. We already have that $\sqrt{\tilde{N}_j}(\tilde{\boldsymbol{\theta}}_{\tilde{N}_j} - \boldsymbol{\theta}_0)$ has asymptotic normality as d_1 tends to 0. It follows that as $d_1 \to 0$, $\hat{\boldsymbol{\theta}}$ follows an asymptotic normal distribution, and

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \left[\rho_1^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{1N_1}^{-1} \boldsymbol{L}_j^\top + \rho_2^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{2N_2}^{-1} \boldsymbol{L}_j^\top\right]^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \longrightarrow \chi_{p_0}^2. \quad \text{(A1.10)}$$

By definition of $\mu_{\hat{N}}$, $\mu_{\hat{N}} \to \mu$ and $\hat{N} d_1^2 / \mu_{\hat{N}} \to a^2$ almost surely, as $d_1 \to 0$. Thus, (A1.10) implies that

$$\begin{aligned}
&\lim_{d_1 \to 0} P(\boldsymbol{\theta}_0 \in R_{\hat{N}}) \\
&= \lim_{d_1 \to 0} P\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \left[\rho_1^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{1N_1}^{-1} \boldsymbol{L}_j^\top + \rho_2^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{2N_2}^{-1} \boldsymbol{L}_j^\top\right]^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq \frac{\hat{N} d_1^2}{\mu_{\hat{N}}} \right) \\
&= \lim_{d_1 \to 0} P\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \left[\rho_1^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{1N_1}^{-1} \boldsymbol{L}_j^\top + \rho_2^2 \boldsymbol{L}_j \boldsymbol{\Sigma}_{2N_2}^{-1} \boldsymbol{L}_j^\top\right]^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \leq a^2 \right) \\
&= 1 - \alpha.
\end{aligned}$$

This completes the proof of Theorem 1.

A2 Additional Numerical Results and Flowchart for the Proposed Procedure

Here, we present additional numerical results that further support the main findings. Under the random selection with covariate set **H1** and parameter setting **B1**, Tables T1 and T2 report the stopping times, AUC, coverage frequency (CF), and absolute bias of the estimate for $\boldsymbol{\theta} = (\beta_1, \beta_2)$.

Table T3 provides parameter estimates for the COVID-19 data using random selection, $d_2 = 0.05$, and equal site proportions (**C1**). Tables T4 and T5 compare parameter estimates under the adaptive and random sampling, respectively, using unequal site proportions (**C2**). Figure 1 shows flowchart of the distributed sequential federated estimation.

We also examine the performance of the proposed method under partially overlapped parameter settings and model misspecification. In the partially overlapped case, we set the common parameter $\boldsymbol{\theta} = (\beta_1, \beta_2) = (2.0, 1.0)$ for all five sites, and define a partially site-specific parameter $\boldsymbol{\zeta} = (\zeta_1, \zeta_2)$ as follows: for sites 1–3, $\boldsymbol{\zeta} = (1.0, 0.5)$; for site 4, $\boldsymbol{\zeta}_4 = (0.5, 0.75)$; and for site 5, $\boldsymbol{\zeta}_5 = (0.5, 1.0)$.

This analysis uses the adaptive sampling setup, covariate set **H1**, and γ configuration **G2** as an illustrative example. For $\boldsymbol{\zeta}$, the sample proportions across the first three sites are 1/6, 1/6, and 4/6, respectively. Table T6 reports the stopping times, AUC, and CF for both $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$. Tables T7 and T8 present the absolute biases for the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$, respectively.

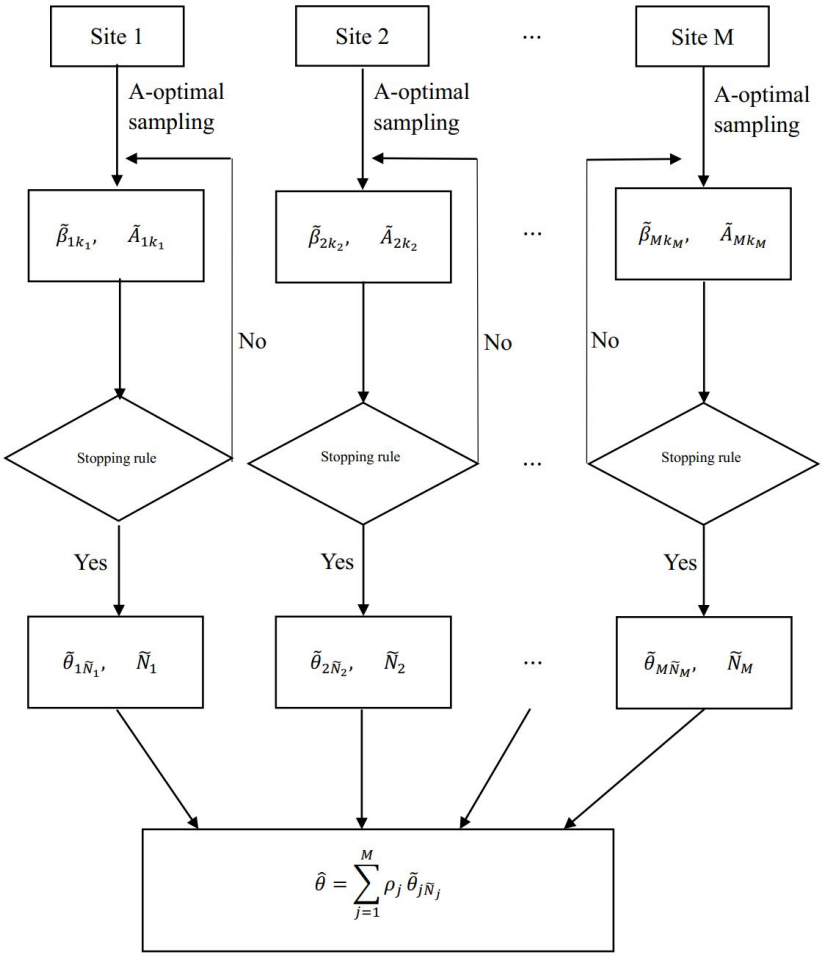

Figure 1: Flowchart of the distributed sequential federated estimation.

We observe that both $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are accurately estimated. In particular, the proposed method (RW) consistently achieves smaller or comparable biases compared to the equal-weight (EW) method, demonstrating its effectiveness under parameter overlap and model misspecification.

To evaluate the robustness of the proposed method under model misspecification, we consider two types of contamination:

(1) **Common parameter contamination** (denoted as Contamination 1): under the regression parameter setting **B1**, a proportion ρ_{outlier} of the simulation data is generated from a model with a different common parameter $\boldsymbol{\theta} = (1.0, 2.0)$;

(2) **Other parameter contamination** (Contamination 2): again under **B1**, a proportion ρ_{outlier} of the data is generated with a different local parameter $\boldsymbol{\eta}_j = (0.5, 1.0)$.

We consider contamination levels $\rho_{\text{outlier}} = 0.05$, 0.10, and 0.15. Table T9 reports the absolute bias of the estimates for the common parameter $\boldsymbol{\theta} = (\beta_1, \beta_2)$ under adaptive sampling, with $d_2 = 0.05$, covariate set **H1**, and parameter configuration **G1**.

We find that for small contamination levels, the estimation bias remains close to zero when $d_1 = 0.2$, indicating robustness to mild model misspecification. In general, contamination in the other (local) parameters has a weaker impact on the estimation of the common parameter than contamination in the common parameter itself. As ρ_{outlier} increases, estimation accuracy gradually deteriorates, as expected.

Table T1: Stopping times, AUC and coverage frequency (CF) of the random selection case with covariate set **H1** and parameter set **B1**.

d_2	d_1			N	N_1	N_2	N_3	N_4	N_5	AUC	CF
0.05	0.3	**G1**	Est.	1845.15	377.32	368.59	369.10	365.54	364.60	0.902	0.930
			Sd	173.02	84.75	68.49	78.39	77.61	79.50	0.008	-
		G2	Est.	1918.81	217.02	214.13	219.31	218.96	1049.38	0.905	0.930
			Sd	150.88	43.84	41.13	42.15	42.51	139.78	0.007	-
	0.2	**G1**	Est.	3987.3	807.74	792.77	796.12	799.63	791.11	0.902	0.925
			Sd	271.47	121.61	101.41	131.26	127.29	126.29	0.005	-
		G2	Est.	3950.75	408.00	399.74	409.42	400.18	2333.41	0.901	0.960
			Sd	252.43	79.62	77.76	81.77	77.38	211.23	0.006	-
0.04	0.3	**G1**	Est.	1899.81	374.90	378.05	391.03	378.69	377.14	0.903	0.925
			Sd	140.62	63.19	61.75	71.76	67.32	67.66	0.007	-
		G2	Est.	2133.28	268.43	282.93	275.99	272.10	1033.84	0.905	0.965
			Sd	164.46	52.07	54.90	51.46	49.52	127.39	0.007	-
	0.2	**G1**	Est.	3953.78	774.38	786.03	810.62	798.30	784.47	0.902	0.935
			Sd	267.60	113.54	122.17	123.86	118.16	124.12	0.006	-
		G2	Est.	3953.59	411.45	414.44	410.12	414.44	2303.13	0.901	0.935
			Sd	251.09	69.29	75.60	70.17	72.97	198.68	0.006	-

G1 and **G2** denote two different sets of γ_j's, $j = 1, \cdots, 5$. d_1 and d_2 are the sizes of confidence set and prefixed parameters for AUC, respectively.

References

Chang, Y.-c. I. (2011). Sequential estimation in generalized linear models when covariates are subject to errors. *Metrika 73*, 93–120.

Chang, Y.-c. I. and A. T. Martinsek (1992). Fixed size confidence regions for parameters of a logistic regression model. *The Annals of Statistics 20*(4), 1953 – 1969.

Chang, Y. I. (1999). Strong consistency of maximum quassi-likelihood eistimate in generalized linear models via a last time. *Statistics & Probability Letters 45*, 237–246.

Chang, Y. I. (2001). Sequential confidence regions of generalized linear models with adaptive designs. *Journal of Statistical Planning and Inference 93*, 277–293.

Chen, K., I. Hu, and Z. Ying (1999). Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *Ann. of Statist. 27*, 1155–1163.

Chow, Y. and H. Robbins (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist. 36*, 457–462.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models, 2nd Edition*. Chapman & Hall, New York.

Wang, Z. and Y.-c. I. Chang (2013). Sequential estimate for linear regression models with uncertain number of effective variables. *Metrika 76*, 949–978.

Woodroofe, M. (1982). *Nonlinear renewal theory in sequential analysis*. CBMS-NSF regional conference series in applied mathematics.

Yin, C., L. Zhao, and C. Wei (2006). Asymptotic normality and strong consistency of maximum quasi-likelihood estimates in generalized linear models. *Science in China Series A 49*, 145–157.

Table T2: Absolute bias of estimate of $\boldsymbol{\theta} = (\beta_1, \beta_2)$ with the random selection strategy, covariate setup **H1** and parameter set **B1**.

d_2	d_1			RW	EW	Site 1	Site 2	Site 3	Site 4	Site 5
0.05	0.3	**G1**	β_1	0.10(0.07)	0.09(0.07)	0.21(0.15)	0.19(0.14)	0.21(0.15)	0.20(0.15)	0.20(0.15)
			β_2	0.07(0.05)	0.07(0.05)	0.16(0.12)	0.13(0.11)	0.15(0.11)	0.15(0.12)	0.15(0.11)
		G2	β_1	0.09(0.07)	0.10(0.07)	0.24(0.17)	0.22(0.16)	0.25(0.18)	0.23(0.16)	0.12(0.09)
			β_2	0.07(0.05)	0.08(0.06)	0.20(0.15)	0.20(0.15)	0.20(0.13)	0.19(0.14)	0.09(0.06)
	0.2	**G1**	β_1	0.07(0.05)	0.06(0.05)	0.14(0.11)	0.12(0.09)	0.14(0.12)	0.14(0.10)	0.14(0.11)
			β_2	0.05(0.04)	0.04(0.03)	0.11(0.07)	0.09(0.08)	0.09(0.08)	0.11(0.07)	0.10(0.08)
		G2	β_1	0.06(0.05)	0.07(0.05)	0.19(0.14)	0.17(0.14)	0.18(0.14)	0.17(0.14)	0.08(0.06)
			β_2	0.05(0.03)	0.06(0.04)	0.13(0.10)	0.14(0.10)	0.16(0.11)	0.13(0.11)	0.06(0.04)
0.04	0.3	**G1**	β_1	0.11(0.07)	0.09(0.06)	0.18(0.13)	0.19(0.14)	0.20(0.15)	0.18(0.14)	0.18(0.14)
			β_2	0.07(0.05)	0.07(0.05)	0.14(0.11)	0.16(0.12)	0.15(0.11)	0.16(0.12)	0.17(0.12)
		G2	β_1	0.09(0.07)	0.11(0.09)	0.22(0.18)	0.25(0.18)	0.24(0.19)	0.25(0.17)	0.12(0.08)
			β_2	0.07(0.05)	0.08(0.06)	0.17(0.14)	0.18(0.15)	0.18(0.15)	0.17(0.14)	0.09(0.06)
	0.2	**G1**	β_1	0.06(0.05)	0.06(0.05)	0.14(0.09)	0.14(0.11)	0.15(0.11)	0.13(0.10)	0.14(0.10)
			β_2	0.05(0.03)	0.05(0.03)	0.11(0.08)	0.10(0.08)	0.10(0.08)	0.11(0.08)	0.11(0.09)
		G2	β_1	0.06(0.05)	0.08(0.06)	0.16(0.12)	0.19(0.14)	0.18(0.14)	0.17(0.13)	0.08(0.06)
			β_2	0.05(0.04)	0.06(0.04)	0.13(0.09)	0.14(0.11)	0.14(0.12)	0.14(0.10)	0.06(0.04)

Standard deviations are in parentheses.

Table T3: Parameter estimate for COVID-19 data with $d_2 = 0.05$, random selection and equal proportion **C1**.

	d_1		GE	PN	AG	DI	CO	AS	IM	HY	OT	CA	OB	CR	SM	EO
All	0.3	Est.	-0.17	1.10	0.01	0.16	-0.29	-0.15	-0.29	0.03	-0.12	-0.31	0.24	-0.28	-0.13	0.27
		Sd	0.01	0.02	0.00	0.02	0.06	0.03	0.06	0.02	0.04	0.04	0.01	0.05	0.02	0.01
	0.2	Est.	-0.16	1.09	0.01	0.15	-0.28	-0.13	-0.33	0.03	-0.14	-0.26	0.24	-0.36	-0.18	0.26
		Sd	0.01	0.01	0.00	0.01	0.04	0.02	0.04	0.01	0.03	0.03	0.01	0.03	0.01	0.01
P1	0.3	Est.	-	1.12	-	-	-0.23	-0.20	-	-	-	-	-	-0.30	-	0.30
		Sd	-	0.03	-	-	0.10	0.05	-	-	-	-	-	0.07	-	0.02
	0.2	Est.	-	1.11	-	-	-0.30	-0.14	-	-	-	-	-	-0.25	-	0.28
		Sd	-	0.02	-	-	0.07	0.04	-	-	-	-	-	0.05	-	0.01
P2	0.3	Est.	-0.17	-	0.01	0.18	-	-	-	0.03	-0.15	-0.33	0.22	-0.32	-0.12	-
		Sd	0.01	-	0.00	0.03	-	-	-	0.02	0.06	0.07	0.02	0.07	0.03	-
	0.2	Est.	-0.17	-	0.01	0.16	-	-	-	0.02	-0.13	-0.29	0.22	-0.30	-0.13	-
		Sd	0.01	-	0.00	0.02	-	-	-	0.02	0.04	0.05	0.02	0.05	0.02	-

All, **P1** and **P2** stand for all variables, five key variables (PN, CO, AS, CR, EO), and ten key variables (GE, AG, DI, AS, HY, OT, CA, OB, CR, SM), respectively. GE: gender; PN: Pneumonia; AG: age; DI: Diabetes; CO: Chronic obstructive pulmonary; AS: asthma; IM: immunosuppression; HY: Hypertension; OT: Other diseases; CA: cardiovascular; OB: obesity; CR: Chronic renal failure; SM: smoke; EO: Exposed to other cases diagnosed as SARS CoV-2.

Table T4: Parameter estimate for COVID-19 data with $d_2 = 0.05$, adaptive selection and different proportion **C2**.

	d_1		GE	PN	AG	DI	CO	AS	IM	HY	OT	CA	OB	CR	SM	EO
All	0.3	Est.	-0.18	1.03	0.01	0.08	-0.23	-0.13	-0.33	0.05	-0.24	-0.34	0.25	-0.24	-0.19	0.45
		Sd	0.03	0.04	0.00	0.04	0.05	0.05	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.03
	0.2	Est.	-0.17	0.91	0.01	0.12	-0.21	-0.01	-0.23	0.05	-0.17	-0.29	0.28	-0.21	-0.20	0.46
		Sd	0.03	0.03	0.00	0.03	0.04	0.04	0.04	0.03	0.04	0.04	0.03	0.04	0.03	0.03
P1	0.3	Est.	-	1.09	-	-	-0.14	-0.08	-	-	-	-	-	-0.22	-	0.45
		Sd	-	0.05	-	-	0.07	0.07	-	-	-	-	-	0.07	-	0.04
	0.2	Est.	-	1.03	-	-	-0.24	-0.10	-	-	-	-	-	-0.27	-	0.44
		Sd	-	0.05	-	-	0.05	0.05	-	-	-	-	-	0.05	-	0.03
P2	0.3	Est.	-0.17	-	0.01	0.10	-	-	-	0.03	-0.26	-0.36	0.23	-0.30	-0.19	-
		Sd	0.03	-	0.00	0.05	-	-	-	0.04	0.06	0.06	0.04	0.06	0.05	-
	0.2	Est.	-0.18	-	0.01	0.10	-	-	-	0.05	-0.23	-0.30	0.27	-0.25	-0.22	-
		Sd	0.03	-	0.00	0.04	-	-	-	0.04	0.05	0.05	0.04	0.05	0.04	-

All, **P1** and **P2** stand for all variables, five key variables (PN, CO, AS, CR, EO), and ten key variables (GE, AG, DI, AS, HY, OT, CA, OB, CR, SM), respectively. GE: gender; PN: Pneumonia; AG: age; DI: Diabetes; CO: Chronic obstructive pulmonary; AS: asthma; IM: immunosuppression; HY: Hypertension; OT: Other diseases; CA: cardiovascular; OB: obesity; CR: Chronic renal failure; SM: smoke; EO: Exposed to other cases diagnosed as SARS CoV-2.

Table T5: Parameter estimate for COVID-19 data with $d_2 = 0.05$, random selection and different proportion **C2**.

	d_1		GE	PN	AG	DI	CO	AS	IM	HY	OT	CA	OB	CR	SM	EO
All	0.3	Est.	-0.15	1.09	0.01	0.16	-0.29	-0.16	-0.33	0.04	-0.15	-0.33	0.24	-0.27	-0.13	0.24
		Sd	0.01	0.02	0.00	0.02	0.06	0.03	0.06	0.02	0.04	0.04	0.01	0.05	0.02	0.01
	0.2	Est.	-0.15	1.08	0.01	0.16	-0.28	-0.14	-0.36	0.04	-0.16	-0.26	0.24	-0.33	-0.19	0.23
		Sd	0.01	0.01	0.00	0.01	0.04	0.02	0.04	0.01	0.03	0.03	0.01	0.03	0.01	0.01
P1	0.3	Est.	-	1.11	-	-	-0.15	-0.23	-	-	-	-	-	-0.30	-	0.26
		Sd	-	0.03	-	-	0.10	0.05	-	-	-	-	-	0.07	-	0.02
	0.2	Est.	-	1.10	-	-	-0.30	-0.15	-	-	-	-	-	-0.25	-	0.25
		Sd	-	0.02	-	-	0.07	0.04	-	-	-	-	-	0.05	-	0.01
P2	0.3	Est.	-0.16	-	0.01	0.18	-	-	-	0.04	-0.17	-0.31	0.21	-0.29	-0.13	-
		Sd	0.01	-	0.00	0.03	-	-	-	0.02	0.06	0.07	0.02	0.07	0.03	-
	0.2	Est.	-0.16	-	0.01	0.17	-	-	-	0.02	-0.16	-0.31	0.22	-0.29	-0.15	-
		Sd	0.01	-	0.00	0.02	-	-	-	0.02	0.04	0.05	0.02	0.05	0.02	-

All, **P1** and **P2** stand for all variables, five key variables (PN, CO, AS, CR, EO), and ten key variables (GE, AG, DI, AS, HY, OT, CA, OB, CR, SM), respectively. GE: gender; PN: Pneumonia; AG: age; DI: Diabetes; CO: Chronic obstructive pulmonary; AS: asthma; IM: immunosuppression; HY: Hypertension; OT: Other diseases; CA: cardiovascular; OB: obesity; CR: Chronic renal failure; SM: smoke; EO: Exposed to other cases diagnosed as SARS CoV-2.

Table T6: Stopping times, AUC and coverage frequency (CF) of the common parameter θ and partially overlapped parameter ζ with the adaptive selection case, covariate set **H1** and parameter set **G2**.

d_2	d_1			N	N_1	N_2	N_3	N_4	N_5	AUC	CF
0.05	0.3	$\boldsymbol{\theta}$	Est.	1325.39	168.50	170.44	166.25	172.20	648.01	0.897	0.940
			Sd	94.10	26.07	24.70	25.75	27.35	79.01	0.005	-
		$\boldsymbol{\zeta}$	Est.	598.91	149.84	154.01	295.06	-	-	0.901	0.970
			Sd	65.74	32.33	32.65	40.80	-	-	0.008	-
	0.2	$\boldsymbol{\theta}$	Est.	2522.34	261.50	262.08	264.19	263.98	1470.58	0.889	0.945
			Sd	148.96	43.92	43.59	41.76	45.02	117.52	0.005	-
		$\boldsymbol{\zeta}$	Est.	1046.59	192.59	191.24	662.77	-	-	0.893	0.970
			Sd	76.57	27.89	23.88	66.39	-	-	0.006	-
0.04	0.3	$\boldsymbol{\theta}$	Est.	1671.87	252.37	254.72	251.12	260.63	653.02	0.897	0.965
			Sd	121.23	50.32	51.13	49.65	45.84	76.71	0.005	-
		$\boldsymbol{\zeta}$	Est.	807.46	244.45	252.70	310.31	-	-	0.899	0.985
			Sd	89.08	61.13	59.27	34.46	-	-	0.008	-
	0.2	$\boldsymbol{\theta}$	Est.	2668.68	297.81	296.57	294.98	302.02	1477.30	0.891	0.955
			Sd	130.49	34.52	31.63	35.44	38.22	110.60	0.004	-
		$\boldsymbol{\zeta}$	Est.	1178.12	259.55	260.66	657.91	-	-	0.894	0.985
			Sd	92.66	42.41	46.44	69.76	-	-	0.006	-

d_1 and d_2 are the sizes of confidence set and prefixed parameters for AUC, respectively.

Table T7: Absolute bias of estimate of the common parameter $\theta = (\beta_1, \beta_2)$ with the adaptive selection case, covariate set **H1** and parameter set **G2**.

d_2	d_1		RW	EW	Site 1	Site 2	Site 3	Site 4	Site 5
0.05	0.3	β_1	0.10(0.07)	0.11(0.09)	0.23(0.18)	0.24(0.18)	0.24(0.17)	0.23(0.19)	0.12(0.10)
		β_2	0.06(0.04)	0.07(0.05)	0.16(0.13)	0.17(0.12)	0.14(0.11)	0.14(0.11)	0.08(0.06)
	0.2	β_1	0.07(0.05)	0.08(0.06)	0.18(0.14)	0.18(0.14)	0.18(0.13)	0.20(0.13)	0.08(0.06)
		β_2	0.04(0.03)	0.05(0.04)	0.12(0.08)	0.12(0.10)	0.11(0.08)	0.13(0.09)	0.05(0.04)
0.04	0.3	β_1	0.09(0.07)	0.12(0.09)	0.23(0.19)	0.21(0.18)	0.21(0.19)	0.20(0.17)	0.12(0.09)
		β_2	0.05(0.04)	0.06(0.05)	0.15(0.12)	0.12(0.10)	0.13(0.10)	0.13(0.12)	0.07(0.06)
	0.2	β_1	0.07(0.05)	0.08(0.06)	0.16(0.13)	0.17(0.13)	0.16(0.13)	0.17(0.13)	0.07(0.05)
		β_2	0.04(0.03)	0.05(0.04)	0.11(0.09)	0.12(0.08)	0.10(0.08)	0.12(0.09)	0.05(0.04)

Standard deviations are in parentheses.

Table T8: Absolute bias of estimate of the partially overlapped parameter $\zeta = (\zeta_1, \zeta_2)$ with the adaptive selection case, covariate set **H1** and parameter set **G2**.

d_2	d_1		RW	EW	Site 1	Site 2	Site 3
0.05	0.3	ζ_1	0.08(0.07)	0.09(0.08)	0.18(0.16)	0.18(0.14)	0.11(0.08)
		ζ_2	0.07(0.05)	0.08(0.06)	0.15(0.12)	0.14(0.11)	0.08(0.06)
	0.2	ζ_1	0.06(0.04)	0.07(0.05)	0.13(0.10)	0.13(0.11)	0.07(0.05)
		ζ_2	0.05(0.04)	0.06(0.04)	0.10(0.08)	0.10(0.08)	0.06(0.04)
0.04	0.3	ζ_1	0.07(0.06)	0.08(0.06)	0.15(0.13)	0.15(0.12)	0.10(0.07)
		ζ_2	0.06(0.05)	0.06(0.06)	0.12(0.10)	0.11(0.09)	0.08(0.07)
	0.2	ζ_1	0.05(0.04)	0.06(0.05)	0.12(0.10)	0.12(0.10)	0.07(0.06)
		ζ_2	0.05(0.03)	0.05(0.04)	0.11(0.09)	0.10(0.07)	0.06(0.05)

Standard deviations are in parentheses.

Table T9: Absolute bias of estimate of the common parameter $\theta = (\beta_1, \beta_2)$ with the adaptive selection case, covariate set **H1** and parameter set **G1**.

		Contamination 1		Contamination 2	
$\rho_{outlier}$	Para.	$d_1 = 0.3$	$d_1 = 0.2$	$d_1 = 0.3$	$d_1 = 0.2$
0.05	β_1	0.09(0.06)	0.08(0.06)	0.10(0.07)	0.06(0.04)
	β_2	0.07(0.05)	0.05(0.04)	0.06(0.04)	0.04(0.03)
0.10	β_1	0.16(0.10)	0.19(0.07)	0.08(0.06)	0.06(0.05)
	β_2	0.07(0.05)	0.06(0.04)	0.05(0.04)	0.04(0.03)
0.15	β_1	0.27(0.11)	0.30(0.08)	0.09(0.06)	0.06(0.05)
	β_2	0.09(0.06)	0.07(0.05)	0.06(0.04)	0.04(0.03)

Standard deviations are in parentheses.