Supplement to "Online-Forgetting Process for debiased-Lasso using Summary Statistics"

Xiao Guo, Xu Zhang and Hai Zhang*

School of Mathematics, Northwest University, China

In the document, Section A provides the proofs for our main results Lemma 1 and Theorem 1. Section B presents the technical lemmas for the main results. Section C includes the details of the dLASSO algorithm and the time complexity comparison of dOnFL and dLASSO. Section D display the additional experimental results that are discussed but not included in the main paper.

A Main Proofs

Proof of Lemma 1

Note that $\hat{\gamma}_r^{[\mathrm{a,b}]}$ in (3.18) also corresponds to a lasso problem, it differs from $\hat{\beta}^{[a,b]}(\lambda_{[a,b]})$ in (2.1) in the data utilized. By Lemma S.2, with probability

*Corresponding author (zhanghai@nwu.edu.cn).

at least $1-p^{-3}$, we have

$$\|\hat{\gamma}_r^{[a,b]} - \gamma_r\|_1 \lesssim s_r \tilde{\lambda}_{[a,b]}.$$

Based on the triangle inequality and Lemma S.1, with probability at least $1-p^{-3}$, we have

$$\begin{aligned}
\|\hat{\gamma}_r^{(j)} - \hat{\gamma}_r^{[a,b]}\|_1 &\leq \|\hat{\gamma}_r^{(j)} - \gamma_r\|_1 + \|\gamma_r - \hat{\gamma}_r^{[a,b]}\|_1 \\
&\lesssim s_r(\lambda_r^{(j)} + \tilde{\lambda}_{[a,b]}) \\
&\lesssim s_r\left(\sqrt{\frac{\log p}{n_j}} + \sqrt{\frac{\log p}{\sum_{j=a}^{b} n_j}}\right) \\
&\lesssim s_r\sqrt{\frac{\log p}{n_j}}.
\end{aligned}$$

■

Proof of Theorem 1

Recall the existing notations $X^{[a,b]} = \left((X^{(a)})^T, \ldots, (X^{(b)})^T\right)^T$, $y^{[a,b]} = \left((y^{(a)})^T, \ldots, (y^{(b)})^T\right)^T$, $\epsilon^{[a,b]} = \left((\epsilon^{(a)})^T, \ldots, (\epsilon^{(b)})^T\right)^T$, $\hat{z}_r^{[a,b]} = \left((\hat{z}_r^{(a)})^T, \ldots, (\hat{z}_r^{(b)})^T\right)^T$ and $x_r^{[a,b]} = \left((x_r^{(a)})^T, \ldots, (x_r^{(b)})^T\right)^T$. The estimator in (2.10) can be written as follows:

$$\hat{\beta}_{de,r}^{[a,b]} = \hat{\beta}_r^{[a,b]} + \left\{(\hat{z}_r^{[a,b]})^T x_r^{[a,b]}\right\}^{-1}\left\{(\hat{z}_r^{[a,b]})^T y^{[a,b]} - (\hat{z}_r^{[a,b]})^T X^{[a,b]}\hat{\beta}^{[a,b]}\right\}$$

$$= \hat{\beta}_r^{[a,b]} + \left\{ (\hat{z}_r^{[a,b]})^T x_r^{[a,b]} \right\}^{-1} \left\{ \sum_{k=1}^{p} (\hat{z}_r^{[a,b]})^T x_k^{[a,b]} \left(\beta_{0,k} - \hat{\beta}_k^{[a,b]} \right) + (\hat{z}_r^{[a,b]})^T \epsilon^{[a,b]} \right\}, \tag{S.1}$$

where the second equality is obtained by substituting $y^{[a,b]} = X^{[a,b]}\beta_0 + \epsilon^{[a,b]}$ and β_0 is the true parameters.

By subtracting β_0 from both sides of the equation (S.1), we have

$$\hat{\beta}_{de,r}^{[a,b]} - \beta_{0,r} = \frac{(\hat{z}_r^{[a,b]})^T \epsilon^{[a,b]} - \sum_{k \neq r}^{p} (\hat{z}_r^{[a,b]})^T x_k^{[a,b]} \left(\hat{\beta}_k^{[a,b]} - \beta_{0,k} \right)}{(\hat{z}_r^{[a,b]})^T x_r^{[a,b]}}$$

$$= \frac{\|\hat{z}_r^{[a,b]}\|_2}{(\hat{z}_r^{[a,b]})^T x_r^{[a,b]}} \left\{ \frac{(\hat{z}_r^{[a,b]})^T \epsilon^{[a,b]}}{\|\hat{z}_r^{[a,b]}\|_2} - \frac{\sum_{k \neq r}^{p} (\hat{z}_r^{[a,b]})^T x_k^{[a,b]} \left(\hat{\beta}_k^{[a,b]} - \beta_{0,k} \right)}{\|\hat{z}_r^{[a,b]}\|_2} \right\} \tag{S.2}$$

$$:= \tau_r^{[a,b]} (w_r^{[a,b]} + \Delta_r^{[a,b]}), \tag{S.3}$$

where

$$\hat{\tau}_r^{[a,b]} = \frac{\|\hat{z}_r^{[a,b]}\|_2}{(\hat{z}_r^{[a,b]})^T x_r^{[a,b]}}, w_r^{[a,b]} = \frac{(\hat{z}_r^{[a,b]})^T \epsilon^{[a,b]}}{\|\hat{z}_r^{[a,b]}\|_2}, \Delta_r^{[a,b]} = \frac{\sum_{k \neq r}^{p} (\hat{z}_r^{[a,b]})^T x_k^{[a,b]} \left(\hat{\beta}_k^{[a,b]} - \beta_{0,k} \right)}{\|\hat{z}_r^{[a,b]}\|_2}. \tag{S.4}$$

On the one hand, based on Lemma S.5, we have $w_r^{[a,b]} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$. On the other hand, according to Lemma S.3 and Lemma S.4, we have

$$\Delta_r^{[a,b]} = \frac{\sum_{k \neq r}^{p} (\hat{z}_r^{[a,b]})^T x_k^{[a,b]} \left(\hat{\beta}_k^{[a,b]} - \beta_{0,k} \right)}{\|\hat{z}_r^{[a,b]}\|_2} = o_p(1). \tag{S.5}$$

Meanwhile, with the help of Proposition 1 in Zhang and Zhang (2014), we have

$$(\hat{z}_r^{[a,b]})^T x_r^{[a,b]} = \sum_{j=a}^{b} (\hat{z}_r^{(j)})^T x_r^{(j)} = \sum_{j=a}^{b} (\hat{z}_r^{(j)})^T (\hat{z}_r^{(j)} + X_{-r}^{(j)} \hat{\gamma}_r^{(j)}) \tag{S.6}$$
$$= \sum_{j=a}^{b} \|\hat{z}_r^{(j)}\|_2^2 + \sum_{j=a}^{b} \lambda_r^{(j)} \|\hat{\gamma}_r^{(j)}\|_1.$$

Then, by Lemma S.3, we have

$$\tau_r^{[a,b]} = \frac{\|\hat{z}_r^{[a,b]}\|_2}{\sum_{j=a}^{b} \|\hat{z}_r^{(j)}\|_2^2 + \sum_{j=a}^{b} \lambda_r^{(j)} \|\hat{\gamma}_r^{(j)}\|_1} \leq \frac{\|\hat{z}_r^{[a,b]}\|_2}{\sum_{j=a}^{b} \|\hat{z}_r^{(j)}\|_2^2} = \frac{1}{\|\hat{z}_r^{[a,b]}\|_2} \asymp \frac{1}{\sqrt{\sum_{j=a}^{b} n_j}}.$$

Therefore, for every $r = 1, \dots, p$ and large enough $\sum_{j=a}^{b} n_j$, we have

$$(\hat{\beta}_{de,r}^{[a,b]} - \beta_{0,r}) / \tau_r^{[a,b]} = w_r^{[a,b]} + \Delta_r^{[a,b]}.$$

And with probability at least $1 - p^{-3}$, $(\tau_r^{[a,b]})^{-1} (\hat{\beta}_{de,r}^{[a,b]} - \beta_{0,r}) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ holds.

B Technical Lemmas

Lemma S.1. *Suppose Assumption 1 holds, $n_j \gtrsim s_r \log p$ and $\lambda_r^{(j)} \asymp \sqrt{\frac{\log p}{n_j}}$ in (2.6) for $j = 1, \dots, b$. Then, for every $r = 1, 2, \dots, p$, with probability at* *least* $1-p^{-3}$*, the* $\hat{\gamma}_r^{(j)}$ *in* (2.6) *and the* γ_r *in* (3.17) *satisfies*

$$\|\hat{\gamma}_r^{(j)} - \gamma_r\|_1 \lesssim s_r \lambda_r^{(j)}. \tag{S.7}$$

Proof of lemma S.1. Let $S_r = \{k : \Theta_{k,r} \neq 0, k \neq r\}$, and then $s_r = |S_r|$. We define the matrix $\hat{\Sigma}_{-r}^{(j)} := (X_{-r}^{(j)})^T X_{-r}^{(j)}/n_j$, $j = a, \ldots, b$.

Firstly, by Assumption 1 and with the help of the Corollary 1 in Raskutti et al. (2010), we can obtain that for all ν satisfying $\|\nu_{S_r^c}\|_1 \leq 3\|\nu_{S_r}\|_1$, as long as $n_j \gtrsim s_r \log p$, it holds that

$$\|\nu_{S_r}\|_1^2 \lesssim (\nu^T \hat{\Sigma}_{-r}^{(j)} \nu) s_r, \tag{S.8}$$

with probability at least $1-p^{-4}$. Namely, $\hat{\Sigma}_{-r}^{(j)}$ satisfies the compatibility condition for the set S_r.

Secondly, since $\hat{\gamma}_r^{(j)}$ is the lasso estimator defined in (2.6), and due to the optimality of $\hat{\gamma}_r^{(j)}$, we have

$$\frac{1}{2n_j}\|x_r^{(j)} - X_{-r}^{(j)}\hat{\gamma}_r^{(j)}\|_2^2 + \lambda_r^{(j)}\|\hat{\gamma}_r^{(j)}\|_1 \leq \frac{1}{2n_j}\|x_r^{(j)} - X_{-r}^{(j)}\gamma_r\|_2^2 + \lambda_r^{(j)}\|\gamma_r\|_1.$$

Let $v^{(j)} = \hat{\gamma}_r^{(j)} - \gamma_r$, we obtain

$$(v^{(j)})^T \hat{\Sigma}_{-r}^{(j)} v^{(j)} \leq \frac{2}{n_j}\left(x_r^{(j)} - X_{-r}^{(j)}\gamma_r\right)^T X_{-r}^{(j)} v^{(j)} + 2\lambda_r^{(j)}(\|\gamma_r\|_1 - \|\hat{\gamma}_r^{(j)}\|_1)$$

$$\leq \frac{2}{n_j} \max_{k \neq r} \left| (x_r^{(j)} - X_{-r}^{(j)} \gamma_r)^T x_k^{(j)} \right| \|v^{(j)}\|_1 + 2\lambda_r^{(j)} (\|\gamma_r\|_1 - \|\hat{\gamma}_r^{(j)}\|_1). \tag{S.9}$$

Denote l-th element of $x_r^{(j)}$ by $x_{r,l}^{(j)}$ and the l-th row of $X_{-r}^{(j)}$ by $X_{-r,l}^{(j)}$. According to Assumption 1 *(i)*, we have

$$(x_r^{(j)} - X_{-r}^{(j)} \gamma_r)^T x_k^{(j)} = \sum_{l=1}^{n_j} (x_{r,l}^{(j)} - X_{-r,l}^{(j)} \gamma_r)^T x_{k,l}^{(j)}, \tag{S.10}$$

which is the sum of i.i.d random variables, and they satisfies $\mathbb{E}(x_{r,l}^{(j)} - X_{-r,l}^{(j)} \gamma_r)^T x_{k,l}^{(j)} = 0$ and $(x_{r,l}^{(j)} - X_{-r,l}^{(j)} \gamma_r)^T x_{k,l}^{(j)}$ follows sub-exponential distribution. The bound of (S.10) could be obtained by Bernstein's inequality, that is, with probability at least $1 - p^{-4}$, it holds that

$$\left| \sum_{l=1}^{n_j} (x_{r,l}^{(j)} - X_{-r,l}^{(j)} \gamma_r)^T x_{k,l}^{(j)} \right| \lesssim \sqrt{n_j \log p}.$$

Therefore, due to Bonferroni's inequality, with probability at least $1 - p^{-4}$, we have

$$\max_{k \neq r} \left| (x_r^{(j)} - X_{-r}^{(j)} \gamma_r)^T x_k^{(j)} \right| = \max_{k \neq r} \left| \sum_{l=1}^{n_j} (x_{r,l}^{(j)} - X_{-r,l}^{(j)} \gamma_r)^T x_{k,l}^{(j)} \right| \lesssim \sqrt{n_j \log p}. \tag{S.11}$$

Combining (S.9) with (S.11), with probability at least $1 - p^{-3}$,

$$(v^{(j)})^T \hat{\Sigma}_{-r}^{(j)} v^{(j)} \lesssim \frac{1}{n_j}\sqrt{n_j \log p}\|v^{(j)}\|_1 + 2\lambda_r^{(j)}(\|\gamma_r\|_1 - \|\hat{\gamma}_r^{(j)}\|_1)$$
$$\asymp \lambda_r^{(j)}\|v^{(j)}\|_1 + 2\lambda_r^{(j)}(\|\gamma_r\|_1 - \|\hat{\gamma}_r^{(j)}\|_1). \qquad \text{(S.12)}$$

To simplify notation, we temporarily omit the subscript r for γ_r and $\hat{\gamma}_r^{(j)}$, that is, we have $v^{(j)} = \hat{\gamma}^{(j)} - \gamma$. Meanwhile,

$$\|\gamma\|_1 = \|\gamma_{S_r}\|_1 + \|\gamma_{S_r^c}\|_1, \ \|\hat{\gamma}^{(j)}\|_1 = \|\hat{\gamma}_{S_r}\|_1 + \|\hat{\gamma}_{S_r^c}\|_1.$$

Due to $\|\gamma_{S_r^c}\|_1 = 0$ and based on the triangle inequality, we have

$$\|\gamma\|_1 - \|\hat{\gamma}^{(j)}\|_1 = \|\gamma_{S_r}\|_1 - \|\hat{\gamma}_{S_r}\|_1 - \|\hat{\gamma}_{S_r^c}\|_1 \leq \|v_{S_r}^{(j)}\|_1 - \|v_{S_r^c}^{(j)}\|_1.$$

Therefore, (S.12) can be written as

$$(v^{(j)})^T \hat{\Sigma}_{-r}^{(j)} v^{(j)} \leq \lambda_r^{(j)}\|v^{(j)}\|_1 + 2\lambda_r^{(j)}(\|v_{S_r}^{(j)}\|_1 - \|v_{S_r^c}^{(j)}\|_1)$$
$$= \lambda_r^{(j)}(3\|v_{S_r}^{(j)}\|_1 - \|v_{S_r^c}^{(j)}\|_1),$$

which implies that $\|v_{S_r^c}^{(j)}\|_1 \leq 3\|v_{S_r}^{(j)}\|_1$.

Finally, under the assumption of $n_j \gtrsim s_r \log p$, and with the help of the compatibility condition of $\hat{\Sigma}^{(j)}_{-r}$ in (S.8), it holds that

$$\frac{\|v^{(j)}_{S_r}\|_1^2}{s_r} \lesssim (v^{(j)})^T \hat{\Sigma}^{(j)}_{-r} v^{(j)} \lesssim \lambda^{(j)}_r \|v^{(j)}_{S_r}\|_1.$$

As a result, with probability at least $1 - p^{-3}$, we have

$$\|\hat{\gamma}^{(j)}_r - \gamma_r\|_1 = \|v^{(j)}\|_1 \leq 4\|v^{(j)}_{S_r}\|_1 \lesssim s_r \lambda^{(j)}_r. \tag{S.13}$$

Since (S.13) holds for any $j = a, \ldots, b$ and $r = 1, \ldots, p$, we finish the proof of Lemma S.1. ∎

Lemma S.2. *Suppose Assumption 1 holds, $\sum_{j=a}^b n_j \gtrsim s_0 \log p$, and $\lambda_{[a,b]} \asymp \sqrt{\frac{\log p}{\sum_{j=a}^b n_j}}$. Then, with probability at least $1 - p^{-3}$, the lasso estimator $\hat{\beta}^{[a,b]}(\lambda_{[a,b]})$ in (2.1) satisfies*

$$\|\hat{\beta}^{[a,b]}(\lambda_{[a,b]}) - \beta_0\|_1 \lesssim s_0 \lambda_{[a,b]}. \tag{S.14}$$

Proof of Lemma S.2. The proof of Lemma S.2 is similar as that of Lemma S.1. First, define the matrix $\hat{\Sigma}^{[a,b]}_{-r} := (X^{[a,b]}_{-r})^T X^{[a,b]}_{-r} / \sum_{j=a}^b n_j$. Then, similar to the derivation of (S.8), we can easily prove that when $\sum_{j=a}^b n_j \gtrsim s_0 \log p$ and $\|\nu_{S_r^c}\|_1 \leq 3\|\nu_{S_r}\|_1$, $\hat{\Sigma}^{[a,b]}_{-r}$ satisfies the compatibility condition for the set S_r. Then, utilizing the optimality of $\hat{\beta}^{[a,b]}$ and the Bernstein's inequality, we can obtain (S.14). The proof is quite similar and thus we omit it. ∎

Lemma S.3. *Suppose the conditions in Lemma S.1 holds. If*

$$\frac{(b-a)s_r^2 \log p}{\sum_{j=a}^{b} n_j} = o_p(1), \tag{S.15}$$

then $\|\hat{z}_r^{[a,b]}\|_2 \asymp \sqrt{\sum_{j=a}^{b} n_j}$.

Proof of Lemma S.3. Note that $z_r^{[\mathrm{a,b}]} = \left((z_r^{(\mathrm{a})})^T, \ldots, (z_r^{(\mathrm{b})})^T\right)^T$, where $z_r^{(j)} = x_r^{(j)} - X_{-r}^{(j)}\gamma_r^{(j)}$, $j = a, \ldots, b$. Based on the triangle inequality, we have

$$\|\hat{z}_r^{[a,b]}\|_2 \leq \|z_r^{[a,b]}\|_2 + \|\hat{z}_r^{[a,b]} - z_r^{[a,b]}\|_2 = \|z_r^{[a,b]}\|_2 + \sqrt{\sum_{j=a}^{b} \|\hat{z}_r^{(j)} - z_r^{(j)}\|_2^2} := I + II. \tag{S.16}$$

For I, consider $\xi_r := \mathbb{E}\{(z_{r,1}^{(j)})^2\} = \mathbb{E}\{(x_{r,1}^{(j)} - X_{-r,1}^{(j)}\gamma_r)^2\}$, where $z_{r,1}^{(j)}, x_{r,1}^{(j)}$ and $X_{-r,1}^{(j)}$ denote the first element of $z_r^{(j)}, x_r^{(j)}$ and the first row of $X_{-r}^{(j)}$ respectively. By Assumption 1 *(ii)* and *(iii)*, for some constant c and C, we have $c < \sigma_{min}^2 \leq \xi_r \leq \Sigma_{j,j} \lesssim C$, then

$$\|z_r^{[a,b]}\|_2^2 \asymp \mathbb{E}(\|z_r^{[a,b]}\|_2^2) + \sqrt{\sum_{j=a}^{b} n_j} = \xi_r \sum_{j=a}^{b} n_j + \sqrt{\sum_{j=a}^{b} n_j} \asymp \sum_{j=a}^{b} n_j.$$

For II, recall the definition of $\hat{z}_r^{(j)} = x_r^{(j)} - X_{-r}^{(j)}\hat{\gamma}_r^{(j)}$ and $\hat{\Sigma}_{-r}^{(j)} = (X_{-r}^{(j)})^T X_{-r}^{(j)}/n_j$.

Based on Lemma S.1, we obtain

$$\begin{aligned}\sum_{j=a}^{b}\|\hat{z}_r^{(j)}-z_r^{(j)}\|_2^2&=\sum_{j=a}^{b}\|X_{-r}^{(j)}(\hat{\gamma}_r^{(j)}-\gamma_r)\|_2^2\\&=\sum_{j=a}^{b}|(\hat{\gamma}_r^{(j)}-\gamma_r)^T n_j\hat{\Sigma}_{-r}^{(j)}(\hat{\gamma}_r^{(j)}-\gamma_r)|\\&\leq\sum_{j=a}^{b}n_j\|\hat{\Sigma}_{-r}^{(j)}\|_\infty\|\hat{\gamma}_r^{(j)}-\gamma_r\|_1^2.\end{aligned}\tag{S.17}$$

Next, we analyze $\|\hat{\Sigma}_{-r}^{(j)}\|_\infty$ in (S.17). Similar to the proof of Lemma S.1, with probability at least of $1-p^{-4}$, we have

$$\|\hat{\Sigma}_{-r}^{(j)}-\Sigma_{-r}\|_\infty\lesssim\sqrt{\frac{\log p}{n_j}},$$

for $j=a,\ldots,b$. Since $n_j\gtrsim s_r\log p$ and $\|\Sigma_{-r}\|_\infty$ is bounded, by the triangle inequality again,

$$\|\hat{\Sigma}_{-r}^{(j)}\|_\infty\leq\|\Sigma_{-r}\|_\infty+\|\hat{\Sigma}_{-r}^{(j)}-\Sigma_{-r}\|_\infty\lesssim M+\sqrt{\frac{\log p}{n_j}}\lesssim C,\tag{S.18}$$

where M and C are some constants. With the help of Lemma S.1, and combining (S.17) and (S.18), we conclude

$$\sum_{j=a}^{b}\|\hat{z}_r^{(j)}-z_r^{(j)}\|_2^2\lesssim\sum_{j=a}^{b}n_j\|\hat{\gamma}_r^{(j)}-\gamma_r\|_1^2\lesssim\sum_{j=a}^{b}n_j s_r^2(\lambda_r^{(j)})^2\tag{S.19}$$

$$\lesssim \sum_{j=a}^{b} n_j s_r^2 \frac{\log p}{n_j} \lesssim (b-a) s_r^2 \log p.$$

As a result, under the assumption of (S.15), we have the upper bound of (S.16), that is,

$$\|\hat{z}_r^{[a,b]}\|_2 \le I + II \lesssim \sqrt{\sum_{j=a}^{b} n_j} + \sqrt{(b-a)s_r^2 \log p} \lesssim \sqrt{\sum_{j=a}^{b} n_j}.$$

Meanwhile, we have

$$\|\hat{z}_r^{[a,b]}\|_2 \ge \|z_r^{[a,b]}\|_2 - \|\hat{z}_r^{[a,b]} - z_r^{[a,b]}\|_2 = I - II \gtrsim \sqrt{\sum_{j=a}^{b} n_j}.$$

In summary, $\|\hat{z}_r^{[a,b]}\|_2 \asymp \sqrt{\sum_{j=a}^{b} n_j}$. We finish the proof of Lemma S.3.

■

Lemma S.4. *Suppose the condition in Lemma S.1 holds. If*

$$\frac{s_0 s_r \log p \sum_{j=a}^{b} \sqrt{n_j}}{\sum_{j=a}^{b} n_j} = o_p(1), \tag{S.20}$$

then $\left|\sum_{k\neq r}^{p} (\hat{z}_r^{[a,b]})^T x_k^{[a,b]} \left(\hat{\beta}_k^{[a,b]} - \beta_{0,k}\right)\right| = o_p(\sqrt{\sum_{j=a}^{b} n_j})$.

Proof of lemma S.4. Recall that $\hat{z}_r^{[a,b]} = \left((\hat{z}_r^{(a)})^T, \ldots, (\hat{z}_r^{(b)})^T\right)^T$ and $\hat{z}_r^{(j)} = x_r^{(j)} - X_{-r}^{(j)} \hat{\gamma}_r^{(j)}$, $r = 1, 2, \ldots, p$, where $\hat{\gamma}_r^{(j)}$ is the lasso estimates obtained based on the subset D_j. To completely eliminate bias, we need to utilize

$\hat{\gamma}_r^{[a,b]}$ and obtain $\tilde{z}_r^{(j)} = x_r^{(j)} - X_{-r}^{(j)}\hat{\gamma}_r^{[a,b]}$, forming $\tilde{z}_r^{[a,b]} = \left((\tilde{z}_r^{(a)})^T, \ldots, (\tilde{z}_r^{(b)})^T\right)^T$, $r = 1, 2, \ldots, p$. However, under the conditions of this paper, obtaining $\hat{\gamma}_r^{[a,b]}$ is not feasible.

$$\begin{aligned}\left|\sum_{k\neq r}(\hat{z}_r^{[a,b]})^T x_k^{[a,b]}\left(\hat{\beta}_k^{[a,b]} - \beta_{0,k}\right)\right| &\leq \left|\sum_{k\neq r}(\tilde{z}_r^{[a,b]})^T x_k^{[a,b]}\left(\hat{\beta}_k^{[a,b]} - \beta_{0,k}\right)\right| \\ &\quad + \left|\sum_{k\neq r}(\hat{z}_r^{[a,b]} - \tilde{z}_r^{[a,b]})^T x_k^{[a,b]}\left(\hat{\beta}_k^{[a,b]} - \beta_{0,k}\right)\right| \\ &:= III + IV.\end{aligned}$$

Term III represents a term that cannot be obtained in practice but enables complete bias correction. Due to $|a^T b| \leq \|a\|_\infty \|b\|_1$, for any vector a and b, we have

$$III \leq \|\hat{\beta}^{[a,b]} - \beta_0\|_1 \max_{k\neq r}\left|(\tilde{z}_r^{[a,b]})^T x_k^{[a,b]}\right| \lesssim s_0\sqrt{\frac{\log p}{\sum_{j=a}^b n_j}}\sqrt{\sum_{j=a}^b n_j \log p} \lesssim s_0 \log p,$$

where the second from the last inequality utilizes the Karush-Kuhn-Tucker conditions and Lemma S.2.

Term IV represents the error between the bias correction term in this paper and the ideal bias correction term. Note that $(\hat{z}_r^{[a,b]} - \tilde{z}_r^{[a,b]})^T x_r^{[a,b]} = \sum_{j=a}^b (\hat{z}_r^{(j)} - \tilde{z}_r^{(j)})^T x_k^{(j)}$ is a number, and for any vector $a = (a_1, \ldots, a_n)^T$ and

$b = (b_1, \ldots, b_n)^T$, we have $a_1 b_1 + \ldots + a_n b_n \le \max_k |a_k| \|b\|_1$, then

$$\begin{aligned} IV &\le \|\hat{\beta}^{[a,b]} - \beta_0\|_1 \max_{k \ne r} \left| \sum_{j=a}^{b} (\hat{z}_r^{(j)} - \tilde{z}_r^{(j)})^T x_k^{(j)} \right| \\ &\le \|\hat{\beta}^{[a,b]} - \beta_0\|_1 \sum_{j=a}^{b} \max_{k \ne r} \left| (\hat{z}_r^{(j)} - \tilde{z}_r^{(j)})^T x_k^{(j)} \right|. \end{aligned}$$

By the definition of $\tilde{z}_r^{(j)}$ and $\hat{z}_r^{(j)}$, we have

$$|(\hat{z}_r^{(j)} - \tilde{z}_r^{(j)})^T x_k^{(j)}| = \left(\hat{\gamma}_r^{[a,b]} - \hat{\gamma}_r^{(j)}\right)^T (X_{-r}^{(j)})^T x_k^{(j)} \le \|\hat{\gamma}_r^{[a,b]} - \hat{\gamma}_r^{(j)}\|_1 \max_{m \ne r} \left| (x_m^{(j)})^T x_k^{(j)} \right|,$$

where $(X_{-r}^{(j)})^T x_k^{(j)} = \left(((x_1^{(j)})^T x_k^{(j)})^T, \ldots, ((x_{r-1}^{(j)})^T x_k^{(j)})^T, ((x_{r+1}^{(j)})^T x_k^{(j)})^T, \ldots, ((x_p^{(j)})^T x_k^{(j)})^T\right)^T$ is the k-th column of the $p \times p$ matrix $\hat{\Sigma}_{-r}^{(j)}$. Therefore, by (S.18), we have

$$\max_{k \ne r} \max_{m \ne r} \left| (x_m^{(j)})^T x_k^{(j)} \right| \le n_j \|\hat{\Sigma}_{-r}^{(j)}\|_\infty \lesssim n_j.$$

According to Lemma S.2 and Lemma 1, we obtain

$$\begin{aligned} IV &\le \|\hat{\beta}^{[a,b]} - \beta_0\|_1 \sum_{j=a}^{b} \|\hat{\gamma}_r^{[a,b]} - \hat{\gamma}_r^{(j)}\|_1 \max_{k \ne r} \max_{m \ne r} \left| (x_m^{(j)})^T x_k^{(j)} \right| \\ &\lesssim s_0 \sqrt{\frac{\log p}{\sum_{j=a}^{b} n_j}} \sum_{j=a}^{b} s_r n_j \sqrt{\frac{\log p}{n_j}} \\ &= \frac{s_0 s_r \log p \sum_{j=a}^{b} \sqrt{n_j}}{\sqrt{\sum_{j=a}^{b} n_j}}. \end{aligned}$$

In summary, under the assumption of (S.20) and $\sum_{j=a}^{b}\sqrt{n_j} \geq \sqrt{\sum_{j=a}^{b} n_j}$, we have

$$III + IV \lesssim s_0 \log p + \frac{s_0 s_r \log p \sum_{j=a}^{b}\sqrt{n_j}}{\sqrt{\sum_{j=a}^{b} n_j}} = o(\sqrt{\sum_{j=a}^{b} n_j}).$$

Then we can conclude that $\left|\sum_{k\neq r}^{p}(\hat{z}_r^{[a,b]})^T x_k^{[a,b]}\left(\hat{\beta}_k^{[a,b]} - \beta_{0,k}\right)\right| = o_p(\sqrt{\sum_{j=a}^{b} n_j})$.

■

Lemma S.5. *Suppose Assumption 1 holds, then*

$$w_r^{[a,b]} = \frac{(\hat{z}_r^{[a,b]})^T \epsilon^{[a,b]}}{\|\hat{z}_r^{[a,b]}\|_2} \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{S.21}$$

Proof of Lemma S.5. We have

$$(\hat{z}_r^{[a,b]})^T \epsilon^{[a,b]} = \sum_{j=a}^{b} (\hat{z}_r^{(j)})^T \epsilon^{(j)} = \sum_{j=a}^{b}\sum_{l=1}^{n_j} \hat{z}_{r,l}^{(j)} \epsilon_l^{(j)},$$

where recall that $\hat{z}_{r,l}^{(j)} = x_{r,l}^{(j)} - X_{-r,l}^{(j)}\hat{\gamma}_r^{(j)}$, $l = 1, 2, \ldots, n_j$ and $j = a, \ldots, b$, and we use $\hat{z}_{r,l}^{(j)}, x_{r,l}^{(j)}, \epsilon_l^{(j)}$ and $X_{-r,l}^{(j)}$ to denote the l-th element of $\hat{z}_r^{(j)}, x_r^{(j)}, \epsilon^{(j)}$ and the l-th row of $X_{-r}^{(j)}$ respectively. Note that $\hat{z}_{r,l}^{(j)}$'s are not independent with each other with respect to l, since they are computed using the same data $X^{(j)}$ and $\hat{\gamma}_r^{(j)}$.

To show the asymptotic normality, we use a martingale central limit theorem. Specifically, similar to the proof of Lemma 3 in Han et al. (2024), we will demonstrate that the folowing two conditions (S.22) and (S.23) of Corollary 3.1 in Hall and Heyde (2014) hold. Let there exists a nested σ-field $\mathcal{F}_{[a,b]}$ such that

$$\frac{1}{\sum_{j=a}^{b} n_j} \sum_{j=a}^{b} \sum_{l=1}^{n_j} E\{(\hat{z}_{r,l}^{(j)} \epsilon_l^{(j)})^2 | \mathcal{F}_{[a,b]}\} \to d^2, \text{ for some } d > 0, \tag{S.22}$$

$$\frac{1}{\sum_{j=a}^{b} n_j} \sum_{j=a}^{b} \sum_{l=1}^{n_j} E\left\{ (\hat{z}_{r,l}^{(j)} \epsilon_l^{(j)})^2 \mathbf{1}\{|\hat{z}_{r,l}^{(j)} \epsilon_l^{(j)}| \gtrsim (\sum_{j=a}^{b} n_j)^{1/2}\} \middle| \mathcal{F}_{[a,b]} \right\} \to 0, \ n \to \infty, \tag{S.23}$$

with probability approaching one, where $\mathbf{1}\{\cdot\}$ is the indicator function. In our case, we let $\mathcal{F}_{[a,b]} = \sigma(X^{(a)}, \ldots, X^{(b)})$. Then, $\hat{z}_{r,l}^{(j)}$ is $\mathcal{F}_{[a,b]}$-measurable. Now, we prove that (S.22) holds. By Lemma S.3 and Assumption 1 *(iv)*, we have

$$\begin{aligned} \frac{1}{\sum_{j=a}^{b} n_j} \sum_{j=a}^{b} \sum_{l=1}^{n_j} E\{(\hat{z}_{r,l}^{(j)} \epsilon_l^{(j)})^2 | \mathcal{F}_{[a,b]}\} &= \frac{1}{\sum_{j=a}^{b} n_j} \sum_{j=a}^{b} \sum_{l=1}^{n_j} \sigma^2 (\hat{z}_{r,l}^{(j)})^2 \\ &= \frac{\sigma^2}{\sum_{j=a}^{b} n_j} \|\hat{z}_r^{[a,b]}\|_2^2 \asymp \sigma^2. \end{aligned} \tag{S.24}$$

Next, we proof that (S.23) holds. We first derive the upper bound of

$\max_{a\le j\le b}\max_{1\le l\le n_j}|\hat{z}_{r,l}^{(j)}|$. Specifically,

$$\begin{aligned}\max_{a\le j\le b}\max_{1\le l\le n_j}|\hat{z}_{r,l}^{(j)}| &= \max_{a\le j\le b}\max_{1\le l\le n_j}|x_{r,l}^{(j)} - X_{-r,l}^{(j)}\hat{\gamma}_r^{(j)}| \\ &\le \max_{a\le j\le b}\max_{1\le l\le n_j}|x_{r,l}^{(j)} - X_{-r,l}^{(j)}\gamma_r| + \max_{a\le j\le b}\max_{1\le l\le n_j}|X_{-r,l}^{(j)}(\hat{\gamma}_r^{(j)} - \gamma_r)|.\end{aligned}$$

For sub-Gaussian random design $X_n = (x_1, \ldots, x_n)$, by Hoeffding's inequality, with probability at least $1 - p^{-4}$, $\|X_n\|_\infty \lesssim \sqrt{\log\max\{p, n\}}$; see the proof of Corollary 2 in Han et al. (2024). By Assumption 1 *(i)* and Lemma 1, as well as the condition in Lemma S.4, the following inequality holds

$$\max_{a\le j\le b}\max_{1\le l\le n_j}|\hat{z}_{r,l}^{(j)}| \le \sqrt{\log p} + s_r \max_j \sqrt{\frac{\log p}{n_j}}\sqrt{\log p} = o(\sqrt{\sum_{j=a}^{b} n_j}),$$

where we assumed $p \gtrsim n_j$ and the last equality follows from the assumption in Theorem 1. Then, we have

$$\begin{aligned}&\frac{1}{\sum_{j=a}^{b} n_j}\sum_{j=a}^{b}\sum_{l=1}^{n_j} E\left\{(\hat{z}_{r,l}^{(j)}\epsilon_l^{(j)})^2 \mathbf{1}\{|\hat{z}_{r,l}^{(j)}\epsilon_l^{(j)}| \gtrsim (\sum_{j=a}^{b} n_j)^{1/2}\}\middle|\mathcal{F}_{[a,b]}\right\} \\ &\le \frac{1}{\sum_{j=a}^{b} n_j}\sum_{j=a}^{b}\sum_{l=1}^{n_j} E\left\{(\hat{z}_{r,l}^{(j)}\epsilon_l^{(j)})^2 \mathbf{1}\{|\epsilon_l^{(j)}| \gtrsim (\sum_{j=a}^{b} n_j)^{1/2}/\max_{a\le j\le b}\max_{1\le l\le n_j}|\hat{z}_{r,l}^{(j)}|\}\middle|\mathcal{F}_{[a,b]}\right\} \\ &= \frac{1}{\sum_{j=a}^{b} n_j} E\left\{\epsilon^2 \mathbf{1}\{|\epsilon| \gtrsim (\sum_{j=a}^{b} n_j)^{1/2}/\max_{a\le j\le b}\max_{1\le l\le n_j}|\hat{z}_{r,l}^{(j)}|\}\middle|\mathcal{F}_{[a,b]}\right\}\sum_{l=1}^{n_j}(\hat{z}_{r,l}^{(j)})^2,\end{aligned}$$

where ϵ is an i.i.d copy of $\epsilon_l^{(j)}$. Since we have

$$\epsilon^2 \mathbf{1}\{|\epsilon| \gtrsim (\sum_{j=a}^{b} n_j)^{1/2} / \max_{a \le j \le b} \max_{1 \le l \le n_j} |\hat{z}_{r,l}^{(j)}|\} \to 0, \text{ as} n \to \infty$$

holds with probability approaching to one. Meanwhile,

$$\epsilon^2 \mathbf{1} \left\{ |\epsilon| \gtrsim (\sum_{j=a}^{b} n_j)^{1/2} / \max_{a \le j \le b} \max_{1 \le l \le n_j} |\hat{z}_{r,l}^{(j)}| \right\} \le \epsilon^2.$$

According to the dominated convergence theorem, (S.23) holds.

Finally, based on Corollary 3.1 in Hall and Heyde (2014) and Lemma S.3, $\frac{1}{\sum_{j=a}^{b} n_j} (\hat{z}_r^{[a,b]})^T \epsilon^{[a,b]}$ converges to a normal distribution $\mathcal{N}(0, \sigma^2)$. ■

C dLASSO algorithm and time complexity comparison

The whole procedure of the dLASSO algorithm is summarized in Algorithm S.1. The details of the time complexity comparison of dOnFL and dLASSO are summarized in Table S.1, where the details for the set-up and the corresponding findings are discussed in the main paper.

Algorithm S.1 Debiased **LASSO** (dLASSO)

Input: step size η_{b+1}, regularization parameters $\lambda_{[a+1,b+1]}$ and $\lambda_r^{(b+1)}$, $\mathcal{SS}^{[a,b]}$, $\mathcal{DS}_r^{[a,b]}$, $\mathcal{SS}^{(j)}$, $\mathcal{DS}_r^{(j)}$ and n_j, for $j = a, \dots, b$ and $r = 1, \dots, p$.

Collect data batch: $D_{(b+1)} = \{X^{(b+1)}, y^{(b+1)}\}$, i.e., the available data batches are $D_{[a+1,b+1]} = \{X^{[a+1,b+1]}, y^{[a+1,b+1]}\}$.

Stage I: Parameter estimation.

Compute statistics $\mathcal{SS}^{(b+1)} = \left\{S^{(b+1)}, U^{(b+1)}\right\}$ with $S^{(b+1)} = (X^{(b+1)})^T X^{(b+1)}$ and $U^{(b+1)} = (X^{(b+1)})^T y^{(b+1)}$;

Update statistics $\mathcal{SS}^{[a+1,b+1]} = \left\{S^{[a+1,b+1]}, U^{[a+1,b+1]}\right\}$ defined in (2.5) using $\mathcal{SS}^{[a,b]}$, $\mathcal{SS}^{(a)}$ and $\mathcal{SS}^{(b+1)}$;

repeat

Step 1: $\hat{\beta}^{[a+1,b+1]} \leftarrow \hat{\beta}^{[a+1,b+1]} - \frac{\eta_{b+1}}{\sum_{j=a+1}^{b+1} n_j}(S^{[a+1,b+1]}\hat{\beta}^{[a+1,b+1]} - U^{[a+1,b+1]})$;

Step 2: $\hat{\beta}_r^{[a+1,b+1]} \leftarrow \text{Soft}\left(\hat{\beta}_r^{[a+1,b+1]}, \eta_{b+1}\lambda_{[a+1,b+1]}\right)$ for $r = 1, ..., p$;

until convergence.

Stage II: Bias-correction.

for $r = 1$ **to** p **do**

Step 1: Compute $\hat{z}_r^{[a+1,b+1]} = x_r^{[a+1,b+1]} - X_{-r}^{[a+1,b+1]}\hat{\gamma}_r^{[a+1,b+1]}$, where

$$\hat{\gamma}_r^{[a+1,b+1]} = \arg\min_{\gamma \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2\sum_{j=a+1}^{b+1} n_j} \|x_r^{[a+1,b+1]} - X_{-r}^{[a+1,b+1]}\gamma\|_2^2 + \lambda_r^{[a+1,b+1]}\|\gamma\|_1 \right\};$$

Step 2: Compute $\mathcal{DS}_r^{[a+1,b+1]} = \left\{a_{1,r}^{[a+1,b+1]}, a_{2,r}^{[a+1,b+1]}, A_r^{[a+1,b+1]}\right\}$ with

$$\begin{aligned} a_{1,r}^{[a+1,b+1]} &= (\hat{z}_r^{[a+1,b+1]})^T x_r^{[a+1,b+1]}; \\ a_{2,r}^{[a+1,b+1]} &= (\hat{z}_r^{[a+1,b+1]})^T y^{[a+1,b+1]}; \\ A_r^{[a+1,b+1]} &= (\hat{z}_r^{[a+1,b+1]})^T X^{[a+1,b+1]}; \end{aligned}$$

Step 3: Using $\hat{\beta}^{[a+1,b+1]}$ to construct the debiased estimator

$$\hat{\beta}_{dLASSO,r}^{[a+1,b+1]} = \hat{\beta}_r^{[a+1,b+1]} + \left\{a_{1,r}^{[a+1,b+1]}\right\}^{-1} \left\{a_{2,r}^{[a+1,b+1]} - A_r^{[a+1,b+1]}\hat{\beta}^{[a+1,b+1]}\right\}.$$

end for

Store and Clear: Store $\mathcal{SS}^{[a+1,b+1]}$, $\mathcal{DS}_r^{[a+1,b+1]}$, $\mathcal{SS}^{(j)}$, $\mathcal{DS}_r^{(j)}$, n_j for $j = a+1, \dots, b+1$ and $r = 1, \dots, p$; and clear others.

Output: The non-debiased estimator $\hat{\beta}^{[a+1,b+1]}$ and debiased estimator $\hat{\beta}_{dLASSO}^{[a+1,b+1]}$.

Table S.1: The time complexity of the dOnFL and the dLASSO when updating from time stamp b with available data batches being $D_{[a,b]} = \{D_a, ..., D_b\}$ to time stamp $b+1$ with available data batches being $D_{[a+1,b+1]} = \{D_{a+1}, ..., D_{b+1}\}$, where $b - a + 1 = T$.

Algorithm		dOnFL	dLASSO
Stage I	Compute statistics $\mathcal{SS}^{(b+1)}$	np^2	np^2
	Update statistics $\mathcal{SS}^{[a+1,b+1]}$	p^2	p^2
	(ISTA algorithm) Obtain lasso estimator $\hat{\beta}^{[a+1,b+1]}$ (k-iterations)	kp^2	kp^2
Stage II	(CD algorithm) Compute $\hat{\gamma}_r^{(b+1)}$, $r \in \{1, \ldots, p\}$ (l-iterations)	lnp^2	$lTnp^2$
	Compute $\hat{z}_r, r \in \{1, \ldots, p\}$	np^2	Tnp^2
	Compute statistics $\mathcal{DS}_r^{(b+1)}, r \in [p]$	np^2	–
	Update statistics $\mathcal{DS}_r^{[a+1,b+1]}, r \in [p]$	p^2	Tnp^2
	Obtain debiasd lasso estimator $\hat{\beta}_{de}^{[a+1,b+1]}$	p^2	p^2
Total complexity		$O(lnp^2 + kp^2)$	$O(Tlnp^2 + kp^2)$

Note: For simplicity, let the sample size $n_j = n$ for each data batch and assume $n < p$.

D Additional experimental results

Table S.2-S.6 presents additional results where the covariates and noise distributions are not Gaussian. Figure S.1 illustrates the impact of different tuning parameter selection methods (i.e., AIC, BIC and CV) on the performance of the compared methods. Table S.7 displays the performance of compared methods under different set-ups of the step size in the ISTA algorithm for lasso. Table S.8 reports results from a higher-dimensional experiment compared to those presented in the main paper. Notably, all experimental settings and the corresponding findings have been discussed in detail in the main paper.

Table S.2: The average performance of dOnFL, OnFL and dLASSO over 20 replications as the sample size (i.e., the number of samples in the available batches at each time stamp) increases. The error terms are *i.i.d.* $U(-0.5, 0.5)$ and the covariates for each sample are *i.i.d.* p-dimensional Gaussian $N(0, \Sigma)$ with $\Sigma = \mathbb{I}$.

Performance Metric	Algorithm	Sample Size					
		30	60	90	120	150	180
Scaled L_1 norm of weak signals $\beta_{0,r} = 0.3$	OnFL	0.186	0.180	0.174	0.164	0.160	0.159
	dOnFL	0.178	0.121	0.096	0.082	0.073	0.065
	dLASSO	0.185	0.122	0.090	0.074	0.059	0.047
Scaled L_1 norm of strong signals $\beta_{0,r} = 1$	OnFL	0.171	0.144	0.130	0.124	0.120	0.118
	dOnFL	0.096	0.057	0.045	0.038	0.031	0.027
	dLASSO	0.094	0.053	0.038	0.031	0.025	0.021
CI-length	dOnFL	1.329	1.086	0.833	0.679	0.608	0.563
	dLASSO	1.414	1.219	0.991	0.866	0.850	0.865
Coverage rate	dOnFL	0.672	0.843	0.892	0.903	0.913	0.924
	dLASSO	0.682	0.865	0.917	0.926	0.925	0.928
Running time (seconds)	dOnFL	31.235	37.510	46.690	52.215	58.970	64.400
	dLASSO	35.545	49.705	70.340	92.940	126.385	164.965

Table S.3: The average performance of dOnFL, OnFL and dLASSO over 20 replications as the sample size (i.e., the number of samples in the available batches at each time stamp) increases. The error terms are *i.i.d.* $U(-0.5, 0.5)$ and the covariates for each sample are *i.i.d.* $U(-1, 1)$.

Performance Metric	Algorithm	Sample Size					
		30	60	90	120	150	180
Scaled L_1 norm of weak signals $\beta_{0,r} = 0.3$	OnFL	0.186	0.180	0.174	0.164	0.160	0.159
	dOnFL	0.178	0.121	0.096	0.082	0.073	0.065
	dLASSO	0.185	0.122	0.090	0.074	0.059	0.047
Scaled L_1 norm of strong signals $\beta_{0,r} = 1$	OnFL	0.171	0.144	0.130	0.124	0.120	0.118
	dOnFL	0.096	0.057	0.045	0.038	0.031	0.027
	dLASSO	0.094	0.053	0.038	0.031	0.025	0.021
CI-length	dOnFL	1.329	1.086	0.833	0.679	0.608	0.563
	dLASSO	1.414	1.219	0.991	0.866	0.850	0.865
Coverage rate	dOnFL	0.672	0.843	0.892	0.903	0.913	0.924
	dLASSO	0.682	0.865	0.917	0.926	0.925	0.928
Running time (seconds)	dOnFL	31.235	37.510	46.690	52.215	58.970	64.400
	dLASSO	35.545	49.705	70.340	92.940	126.385	164.965

Table S.4: The average performance of dOnFL, OnFL and dLASSO over 20 replications as the sample size (i.e., the number of samples in the available batches at each time stamp) increases. The error terms are *i.i.d.* $Exp(5)$ and all covariates for each sample are *i.i.d.* $U(-1,1)$.

Performance Metric	Algorithm	Sample Size					
		30	60	90	120	150	180
Scaled L_1 norm of weak signals $\beta_{0,r}=0.3$	OnFL	0.189	0.174	0.173	0.166	0.161	0.157
	dOnFL	0.176	0.118	0.101	0.088	0.077	0.066
	dLASSO	0.181	0.122	0.098	0.080	0.069	0.059
Scaled L_1 norm of strong signals $\beta_{0,r}=1$	OnFL	0.176	0.143	0.134	0.123	0.120	0.118
	dOnFL	0.104	0.059	0.051	0.037	0.032	0.028
	dLASSO	0.101	0.055	0.044	0.032	0.027	0.024
CI-length	dOnFL	1.361	1.179	0.819	0.701	0.624	0.608
	dLASSO	1.453	1.313	0.975	0.899	0.871	0.933
Coverage rate	dOnFL	0.628	0.854	0.855	0.900	0.908	0.932
	dLASSO	0.643	0.878	0.886	0.921	0.928	0.940
Running time (seconds)	dOnFL	21.705	28.900	35.850	43.970	50.260	56.850
	dLASSO	26.220	41.855	61.025	87.205	123.285	162.690

Table S.5: The average performance of dOnFL, OnFL and dLASSO over 20 replications as the sample size (i.e., the number of samples in the available batches at each time stamp) increases. The error terms are *i.i.d.* $t(10)$ and all covariates for each sample are *i.i.d.* $U(-5,5)$.

Performance Metric	Algorithm	Sample Size					
		30	60	90	120	150	180
Scaled L_1 norm of weak signals $\beta_{0,r}=0.3$	OnFL	0.183	0.179	0.166	0.169	0.161	0.154
	dOnFL	0.171	0.127	0.097	0.083	0.072	0.066
	dLASSO	0.176	0.129	0.093	0.076	0.060	0.052
Scaled L_1 norm of strong signals $\beta_{0,r}=1$	OnFL	0.167	0.143	0.127	0.123	0.118	0.118
	dOnFL	0.108	0.063	0.042	0.038	0.031	0.027
	dLASSO	0.106	0.059	0.038	0.033	0.026	0.021
CI-length	dOnFL	1.170	1.094	0.838	0.673	0.607	0.575
	dLASSO	1.240	1.228	0.996	0.859	0.849	0.880
Coverage rate	dOnFL	0.539	0.807	0.897	0.899	0.917	0.920
	dLASSO	0.559	0.830	0.920	0.918	0.926	0.931
Running time (seconds)	dOnFL	22.780	32.955	43.600	51.910	60.895	68.950
	dLASSO	27.450	46.575	70.345	96.925	134.400	177.685

Table S.7: The effect of step size on the performance of dOnFL, OnFL and dLASSO, where the baseline η_j is the theoretical value in Remark 1. The sample size $n_j = 40$. The error terms are *i.i.d.* $\epsilon^{(j)} \sim \mathcal{N}(0, (0.3)^2 \mathbb{I}_{n_j})$ and the covariates for each sample are *i.i.d.* p-dimensional Gaussian $N(0, \Sigma)$ with $\Sigma_{ij} = 0.4^{|i-j|}$.

Performance Metric	Algorithm	Value of η_j					
		$0.5\eta_j$	η_j	$1.5\eta_j$	$2\eta_j$	$2.5\eta_j$	$3\eta_j$
Scaled L_1 norm of weak signals $\beta_{0,r} = 0.3$	OnFL	0.175	0.167	0.160	0.153	0.145	0.137
	dOnFL	0.101	0.086	0.076	0.070	0.065	0.061
	dLASSO	0.092	0.080	0.072	0.066	0.062	0.058
Scaled L_1 norm of strong signals $\beta_{0,r} = 1$	OnFL	0.166	0.135	0.111	0.095	0.083	0.075
	dOnFL	0.039	0.038	0.034	0.032	0.030	0.029
	dLASSO	0.032	0.031	0.029	0.027	0.027	0.026
CI-length	dOnFL	1.998	0.862	0.636	0.540	0.464	0.419
	dLASSO	2.618	1.115	0.820	0.693	0.597	0.537
Coverage rate	dOnFL	0.9365	0.9260	0.9115	0.9030	0.8910	0.8825
	dLASSO	0.9490	0.9310	0.9260	0.9255	0.9195	0.9220

Table S.6: The average performance of dOnFL, OnFL and dLASSO over 20 replications as the sample size (i.e., the number of samples in the available batches at each time stamp) increases. The error terms are *i.i.d.* $Exp(10)$ and all covariates for each sample are *i.i.d.* $t(10)$.

Performance Metric	Algorithm	Sample Size					
		30	60	90	120	150	180
Scaled L_1 norm of weak signals $\beta_{0,r} = 0.3$	OnFL	0.183	0.179	0.166	0.169	0.161	0.154
	dOnFL	0.171	0.127	0.097	0.083	0.072	0.066
	dLASSO	0.176	0.129	0.093	0.076	0.060	0.052
Scaled L_1 norm of strong signals $\beta_{0,r} = 1$	OnFL	0.167	0.143	0.127	0.123	0.118	0.119
	dOnFL	0.108	0.063	0.042	0.038	0.031	0.027
	dLASSO	0.106	0.059	0.038	0.033	0.026	0.021
CI-length	dOnFL	1.170	1.094	0.838	0.673	0.607	0.575
	dLASSO	1.240	1.228	0.995	0.859	0.849	0.880
Coverage rate	dOnFL	0.5385	0.8070	0.8970	0.8985	0.9165	0.9200
	dLASSO	0.5590	0.8295	0.9195	0.9175	0.9260	0.9305
Running time (seconds)	dOnFL	22.780	32.955	43.600	51.910	60.895	68.950
	dLASSO	27.450	46.575	70.345	96.925	134.400	177.685

Table S.8: The average performance of dOnFL, OnFL and dLASSO over 20 replications as the sample size (i.e., the number of samples in the available batches at each time stamp) increases. The dimension $p = 800$ and at each time stamp, the number of available data batches at each time stamp $T = 2$. The error terms are *i.i.d.* $\epsilon^{(j)} \sim \mathcal{N}(0, (0.3)^2 \mathbb{I}_{n_j})$ and the covariates for each sample are *i.i.d.* p-dimensional Gaussian $N(0, \Sigma)$ with $\Sigma_{ij} = 0.4^{|i-j|}$.

Performance Metric	Algorithm	Sample Size				
		160	240	320	400	480
Scaled L_1 norm of weak signals $\beta_{0,r} = 0.3$	OnFL	0.190	0.187	0.186	0.184	0.184
	dOnFL	0.098	0.077	0.071	0.061	0.054
	dLASSO	0.098	0.076	0.067	0.055	0.044
Scaled L_1 norm of strong signals $\beta_{0,r} = 1$	OnFL	0.185	0.183	0.183	0.182	0.182
	dOnFL	0.048	0.038	0.032	0.027	0.022
	dLASSO	0.047	0.035	0.028	0.023	0.018
CI-length	dOnFL	1.114	0.962	0.825	0.682	0.563
	dLASSO	1.196	1.062	0.932	0.794	0.676
Coverage rate	dOnFL	0.817	0.846	0.862	0.890	0.936
	dLASSO	0.835	0.872	0.902	0.939	0.968
Running time (seconds)	dOnFL	458.73	550.98	619.14	684.32	731.83
	dLASSO	528.09	704.75	859.01	1011.23	1205.52

References

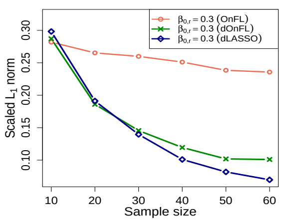
Hall, P. and C. C. Heyde (2014). *Martingale limit theory and its application.* Academic press.

Han, R., L. Luo, Y. Lin, and J. Huang (2024). Online inference with debiased stochastic gradient descent. *Biometrika 111*(1), 93–108.

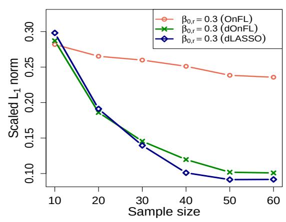
Raskutti, G., M. J. Wainwright, and B. Yu (2010). Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research 11*, 2241–2259.

Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 76*(1), 217–242.

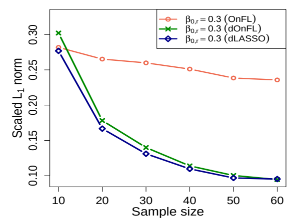
(a) Scaled L_1 norm under weak signals



(a) AIC

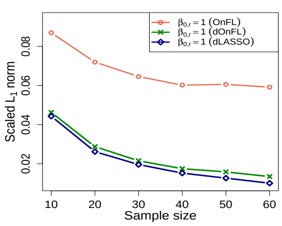


(b) BIC

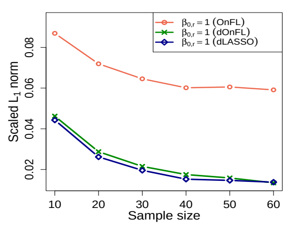


(c) CV

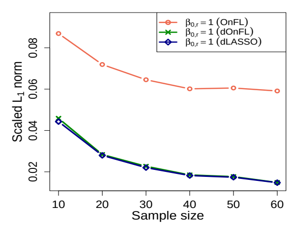
(b) Scaled L_1 norm under strong signals



(d) AIC



(e) BIC



(f) CV

Figure S.1: The performance of dOnFL, OnFL, and dLASSO in terms of the scaled L_1 norms under the three selection methods for tuning parameters including AIC, BIC and CV. The scaled norms are reported separately for (a) weak signals $\beta_{0,r} = 0.3$ and (b) strong signals $\beta_{0,r} = 1$. The error terms are *i.i.d.* $\epsilon^{(j)} \sim \mathcal{N}(0, (0.3)^2 \mathbb{I}_{n_j})$ and the covariates for each sample are *i.i.d.* p-dimensional Gaussian $N(0, \Sigma)$ with $\Sigma = \mathbb{I}$.