Supplementary Material of “Renewable Estimation and Incremental Inference in Generalized Linear Models with Streaming Datasets”

Lan Luo

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

E-mail: luolsph@umich.edu

Peter X.-K. Song

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

E-mail: pxsong@umich.edu

S1. Sequential Updating Methods

This section provides the technical details concerning existing online methods that are considered in the comparisons with our proposed renewable estimation and incremental inference. **Online Least Squares Estimation (OLSE).** Consider a linear model $y_i = \boldsymbol{x}_i^T \boldsymbol{\beta}_0 + \epsilon_i$, with *i.i.d.* errors ϵ_i's, $i = 1, \dots, N_b$. For the current single data batch D_b, the LSE and its sum of squared errors (SSE) are denoted by $\hat{\boldsymbol{\beta}}_b = (\boldsymbol{X}_b^T \boldsymbol{X}_b)^{-1} \boldsymbol{X}_b^T \boldsymbol{y}_b$ and $SSE_b = SSE(\hat{\boldsymbol{\beta}}_b; D_b)$, respectively. Let $\tilde{\boldsymbol{\beta}}_b^{\text{olse}}$ and $CMSE_b$ denote the online LSE (OLSE) and the cumulative mean squared error (CMSE) based on $D_b^\star$. With initial $\tilde{\boldsymbol{\beta}}_1^{\text{olse}} = \hat{\boldsymbol{\beta}}_1$, the OLSE takes the following form of decomposition:

$$\tilde{\boldsymbol{\beta}}_b^{\text{olse}} = \left(\sum_{j=1}^{b-1} \boldsymbol{X}_j^T \boldsymbol{X}_j + \boldsymbol{X}_b^T \boldsymbol{X}_b \right)^{-1} \left(\sum_{j=1}^{b-1} \boldsymbol{X}_j^T \boldsymbol{X}_j \tilde{\boldsymbol{\beta}}_{b-1}^{\text{olse}} + \boldsymbol{X}_b^T \boldsymbol{X}_b \hat{\boldsymbol{\beta}}_b \right), \ b = 2, 3, \dots. \tag{S1}$$

The cumulative SSE (CSSE) takes a recursive procedure:

$$\begin{aligned} CSSE_b :=& SSE(\tilde{\boldsymbol{\beta}}_b^{\text{olse}}; D_b^\star) \\ =& CSSE_{b-1} + SSE_b + \tilde{\boldsymbol{\beta}}_{b-1}^{\text{olse}^T} \left(\sum_{j=1}^{b-1} \boldsymbol{X}_j^T \boldsymbol{X}_j \right) \tilde{\boldsymbol{\beta}}_{b-1}^{\text{olse}} + \hat{\boldsymbol{\beta}}_b^T \boldsymbol{X}_b^T \boldsymbol{X}_b \hat{\boldsymbol{\beta}}_b \\ & - \tilde{\boldsymbol{\beta}}_b^{\text{olse}^T} \left(\sum_{j=1}^{b} \boldsymbol{X}_j^T \boldsymbol{X}_j \right) \tilde{\boldsymbol{\beta}}_b^{\text{olse}}, \ b = 2, 3, \dots. \end{aligned} \tag{S2}$$

The initial $CSSE_1 := SSE(\hat{\boldsymbol{\beta}}_1; D_1)$. It follows that the CMSE with $D_b^\star$ is $CMSE_b := MSE(\tilde{\boldsymbol{\beta}}_b^{\text{olse}}; D_b^\star) = CSSE_b/(N_b - p)$. The variance is $\tilde{\boldsymbol{V}}_b^{\text{olse}} = CMSE_b \times \left(\sum_{j=1}^b \boldsymbol{X}_j^T \boldsymbol{X}_j \right)$.
Online Estimating Equations. Let $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ be a parameter value satisfying $\sum_{i \in D_b^\star} \mathbb{E}\{\boldsymbol{\psi}(y_i, \boldsymbol{x}_i; \boldsymbol{\beta}_0)\} = \boldsymbol{0}$, where $\boldsymbol{\psi}(\cdot)$ is an unbiased estimating function. This includes the unit score $\boldsymbol{U}$ and the scaled score $\boldsymbol{U}(\cdot)/\phi$ as special cases. The estimator and its variance with D_b are denoted by $\hat{\boldsymbol{\beta}}_b$ and $\boldsymbol{V}_b$ where $\boldsymbol{V}_b$ is the sandwich covariance matrix. A cumulative estimating equation (CEE) estimator, $\tilde{\boldsymbol{\beta}}_b^{\text{cee}}$, proposed by Lin and Xi (2011) is a sequentially updated estimate by the means of the following meta-type estimation, together with the corresponding cumulative negative Hessian matrix $\tilde{\boldsymbol{A}}_b^{\text{cee}}$:

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_b^{\text{cee}} &= \left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cee}} + \boldsymbol{A}_b^{\text{cee}} \right)^{-1} \left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cee}} \tilde{\boldsymbol{\beta}}_{b-1}^{\text{cee}} + \boldsymbol{A}_b^{\text{cee}} \hat{\boldsymbol{\beta}}_b \right), \\ \tilde{\boldsymbol{A}}_b^{\text{cee}} &= \sum_{j=1}^b \boldsymbol{A}_j^{\text{cee}}, \ b = 1, 2, \dots, \end{aligned} \tag{S3}$$

with initial $\tilde{\boldsymbol{A}}_0^{\text{cee}} = \mathbf{0}_{p\times p}$, and $\boldsymbol{A}_b^{\text{cee}} = -\sum_{i\in D_b} \nabla_{\boldsymbol{\beta}}\boldsymbol{\psi}(y_i, \boldsymbol{x}_i; \hat{\boldsymbol{\beta}}_b)$ is the negative Hessian matrix of single data batch D_b. With initial $\tilde{\boldsymbol{V}}_0^{\text{cee}} = \mathbf{0}_{p\times p}$, the variance of the estimator $\tilde{\boldsymbol{\beta}}_b^{\text{cee}}$ is

$$\begin{aligned}\tilde{\boldsymbol{V}}_b^{\text{cee}} := \widetilde{\text{Var}}(\tilde{\boldsymbol{\beta}}_b^{\text{cee}}) = \left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cee}} + \boldsymbol{A}_b^{\text{cee}}\right)^{-1} \left\{\tilde{\boldsymbol{A}}_{b-1}^{\text{cee}} \tilde{\boldsymbol{V}}_{b-1}^{\text{cee}} \left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cee}}\right)^T + \boldsymbol{A}_b^{\text{cee}} \boldsymbol{V}_b \left(\boldsymbol{A}_b^{\text{cee}}\right)^T\right\} \\ \times \left\{\left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cee}} + \boldsymbol{A}_b^{\text{cee}}\right)^{-1}\right\}^T, \ b = 1, 2, \ldots.\end{aligned} \tag{S4}$$

To reduce bias in CEE, a cumulatively updated estimating equation (CUEE) estimator is proposed by Schifano et al. (2016). The CUEE estimator and the corresponding cumulative negative Hessian $\tilde{\boldsymbol{A}}_b^{\text{cuee}}$:

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_b^{\text{cuee}} &= \left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cuee}} + \boldsymbol{A}_b^{\text{cuee}}\right)^{-1} \left\{\sum_{j=1}^{b-1} \boldsymbol{A}_j^{\text{cuee}} \check{\boldsymbol{\beta}}_j + \boldsymbol{A}_b^{\text{cuee}} \check{\boldsymbol{\beta}}_b + \sum_{j=1}^{b-1} \sum_{i\in D_j} \boldsymbol{\psi}_i(\check{\boldsymbol{\beta}}_j) + \sum_{i\in D_b} \boldsymbol{\psi}_i(\check{\boldsymbol{\beta}}_b)\right\}, \\ \tilde{\boldsymbol{A}}_b^{\text{cuee}} &= \sum_{j=1}^{b} \boldsymbol{A}_j^{\text{cuee}}, \ b = 1, 2, \ldots,\end{aligned} \tag{S5}$$

with initial $\tilde{\boldsymbol{A}}_0 = \boldsymbol{A}_0 = \mathbf{0}_{p\times p}$, $\boldsymbol{A}_b^{\text{cuee}} = -\sum_{i\in D_b} \nabla_{\boldsymbol{\beta}}\boldsymbol{\psi}(y_i, \boldsymbol{x}_i; \check{\boldsymbol{\beta}}_b)$ is the negative Hessian of D_b evaluated at $\check{\boldsymbol{\beta}}_b$ which is an intermediary estimator similar to the CEE estimator. Similarly, initiated by $\tilde{\boldsymbol{V}}_0^{\text{cuee}} = \mathbf{0}_{p\times p}$, the recursively updated variance of $\tilde{\boldsymbol{\beta}}_b^{\text{cuee}}$ is

$$\begin{aligned}\tilde{\boldsymbol{V}}_b^{\text{cuee}} := \widetilde{\text{Var}}(\tilde{\boldsymbol{\beta}}_b^{\text{cuee}}) = \left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cuee}} + \boldsymbol{A}_b^{\text{cuee}}\right)^{-1} \left\{\tilde{\boldsymbol{A}}_{b-1}^{\text{cuee}} \tilde{\boldsymbol{V}}_{b-1}^{\text{cuee}} \left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cuee}}\right)^T + \boldsymbol{A}_b^{\text{cuee}} \boldsymbol{V}_b \left(\boldsymbol{A}_b^{\text{cuee}}\right)^T\right\} \\ \times \left\{\left(\tilde{\boldsymbol{A}}_{b-1}^{\text{cuee}} + \boldsymbol{A}_b^{\text{cuee}}\right)^{-1}\right\}^T, \ b = 1, 2, \ldots.\end{aligned} \tag{S6}$$

As shown by Schifano et al. (2016), the CUEE estimator is less biased than the CEE estimator under finite sample sizes. Nevertheless, its estimation consistency is established under the same strong regularity condition as that required by the CEE estimator; that is, the number of data batches b is of order $\mathcal{O}(n_j^k)$, for $k < 1/3$ and each $j = 1, \ldots, b$. This condition apparently is not valid for high throughput streaming data, where n_j is typically small, but b grows at a high rate.

S2. Approximate Sufficient Statistic for the GLMs

This section provides the technical details related to the approximate sufficiency in the framework of GLMs. First, we prove the summary statistics used in the proposed renewable analytics are approximate sufficient statistics. Then, we present an example of approximate sufficient statistic in the linear model.

We consider a fixed design; that is, for an arbitrary time b, a sample of size N_b satisfying $\{y_i; \boldsymbol{x}_i\}_{i=1}^{N_b} \sim f(y, \boldsymbol{x}; \boldsymbol{\beta}_0, \phi_0)$ arriving in a sequence of b data batches, where $\boldsymbol{\beta}$ is the parameter of interest. According to Definition 1 given in Appendix A.6, we show in the following proof that $S(D_{b-1}^\star) = (S_1(D_{b-1}^\star), S_2(D_{b-1}^\star)) = (\tilde{\boldsymbol{\beta}}_{b-1}, \tilde{\boldsymbol{J}}_{b-1})$ is an approximate sufficient statistic for $\boldsymbol{\beta}$ based on cumulative historical dataset $D_{b-1}^\star$. In addition, it is worth mentioning that in the Gaussian linear model, there exists a statistic, $S(D_{b-1}^\star)$, such that the factorization holds exactly, namely, $f(D_{b-1}^\star; \boldsymbol{\beta}, \tilde{\phi}_{b-1}) = g\left(S(D_{b-1}^\star); \boldsymbol{\beta}\right) c(D_{b-1}^\star; \tilde{\phi}_{b-1})$. See Example 1 below.

Proof: At time b, a total of N_b samples are independently sampled from $f(y; \boldsymbol{x}, \boldsymbol{\beta}_0, \phi_0)$. For the simplicity, we only consider the case with $\phi_0 = 1$, and other cases with unknown $\phi_0 > 0$

may be similarly done with little effort. The joint pdf of $D_{b-1}^{\star}$ is

$$
\begin{aligned}
f_{N_{b-1}}(D_{b-1}^{\star};\boldsymbol{\beta},\phi_0=1) &= \exp\left\{\sum_{i\in D_{b-1}^{\star}} \log f(y_i;\boldsymbol{x}_i,\boldsymbol{\beta})\right\} \\
&= \exp\left\{\ell_{N_{b-1}}(\boldsymbol{\beta};D_{b-1}^{\star})\right\} \\
&= \sum_{i\in D_{b-1}^{\star}} \log h(y_i;\phi) + \sum_{i\in D_{b-1}^{\star}} \left\{y_i\boldsymbol{x}_i^T\boldsymbol{\beta} - b(\boldsymbol{x}_i^T\boldsymbol{\beta})\right\},
\end{aligned}
\tag{S7}
$$

where $b(\cdot)$ is the cumulant generating function.

By the classical sufficiency theorem (Casella and Berger, 2002, Chapter 6), $T(D_{b-1}^{\star}) = \sum_{i\in D_{b-1}^{\star}} y_i\boldsymbol{x}_i^T$ is a sufficient statistic for $\boldsymbol{\beta}$. However, when $D_{b-1}^{\star}$ is not available as in the streaming dataset setting, we take a second-order Taylor expansion of $\ell_{N_{b-1}}(\boldsymbol{\beta};D_{b-1}^{\star})$ in (S7) around $\tilde{\boldsymbol{\beta}}_{b-1}$ that leads to

$$
\begin{aligned}
\ell_{N_{b-1}}(\boldsymbol{\beta};D_{b-1}^{\star}) =& \sum_{j=1}^{b-1} \ell_j(\tilde{\boldsymbol{\beta}}_{b-1};D_j) + (\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1})^T \left\{\sum_{j=1}^{b-1} \boldsymbol{U}_j(\tilde{\boldsymbol{\beta}}_{b-1};D_j) - \frac{1}{2}\sum_{j=1}^{b-1} \boldsymbol{J}_j(\tilde{\boldsymbol{\beta}}_j;D_j)(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1})\right\} \\
&+ O_p\left(N_{b-1}\|\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1}\|^3\right).
\end{aligned}
\tag{S8}
$$

Using the fact of $\|\hat{\boldsymbol{\beta}}_{b-1}^{\star} - \tilde{\boldsymbol{\beta}}_{b-1}\|_2 = O_p(N_{b-1}^{-1})$, and taking a first-order Taylor expansion of $\sum_{j=1}^{b-1}\boldsymbol{U}_j(\tilde{\boldsymbol{\beta}}_{b-1};D_j)$ around $\hat{\boldsymbol{\beta}}_{b-1}^{\star}$, we have

$$
\begin{aligned}
&\sum_{j=1}^{b-1} \boldsymbol{U}_j(\tilde{\boldsymbol{\beta}}_{b-1};D_j) - \frac{1}{2}\sum_{j=1}^{b-1} \boldsymbol{J}_j(\tilde{\boldsymbol{\beta}}_j;D_j)(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1}) \\
=& \sum_{j=1}^{b-1} \boldsymbol{J}_j(\hat{\boldsymbol{\beta}}_{b-1}^{\star};D_j)(\hat{\boldsymbol{\beta}}_{b-1}^{\star}-\tilde{\boldsymbol{\beta}}_{b-1}) - \frac{1}{2}\sum_{j=1}^{b-1} \boldsymbol{J}_j(\tilde{\boldsymbol{\beta}}_j;D_j)(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1}) + O_p\left(N_{b-1}\|\hat{\boldsymbol{\beta}}_{b-1}^{\star}-\tilde{\boldsymbol{\beta}}_{b-1}\|^2\right) \\
=& -\frac{1}{2}\sum_{j=1}^{b-1} \boldsymbol{J}_j(\tilde{\boldsymbol{\beta}}_j;D_j)(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1}) + O_p(1)
\end{aligned}
\tag{S9}
$$

Plugging equation (S9) into equation (S8), we obtain

$$
\begin{aligned}
\ell_{N_{b-1}}(\boldsymbol{\beta};D_{b-1}^{\star}) =& \sum_{j=1}^{b-1} \ell_j(\tilde{\boldsymbol{\beta}}_{b-1};D_j) - \frac{1}{2}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1})^T \left\{\sum_{j=1}^{b-1} \boldsymbol{J}_j(\tilde{\boldsymbol{\beta}}_j;D_j)\right\}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1}) \\
&+ O_p\left(\|\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1}\|\right) + O_p\left(N_{b-1}\|\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1}\|^3\right) \\
=& \sum_{j=1}^{b-1} \ell_j(\tilde{\boldsymbol{\beta}}_{b-1};D_j) - \frac{1}{2}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1})^T \left\{\sum_{j=1}^{b-1} \boldsymbol{J}_j(\tilde{\boldsymbol{\beta}}_j;D_j)\right\}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_{b-1}) + O_p\left(N_{b-1}^{-1/2}\right),
\end{aligned}
\tag{S10}
$$

and the last equality holds since $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| = O_p(1/\sqrt{N_{b-1}})$ for $\boldsymbol{\beta}\in\mathcal{B}_{N_{b-1}}(\delta)$, and $\|\tilde{\boldsymbol{\beta}}_{b-1}-\boldsymbol{\beta}_0\| = O_p(1/\sqrt{N_{b-1}})$.

Plugging equation (S10) into equation (S7), we obtain

$$\begin{aligned} f_{N_{b-1}}(D^{\star}_{b-1};\boldsymbol{\beta}) &= \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^T\tilde{\boldsymbol{J}}_{b-1}\boldsymbol{\beta}+\boldsymbol{\beta}^T\tilde{\boldsymbol{J}}_{b-1}\tilde{\boldsymbol{\beta}}_{b-1}+O_p\left(N_{b-1}^{-1/2}\right)\right\} \\ &\quad \times\exp\left\{\sum_{j=1}^{b-1}\ell_j(\tilde{\boldsymbol{\beta}}_{b-1};D_j)-\frac{1}{2}\tilde{\boldsymbol{\beta}}_{b-1}^T\tilde{\boldsymbol{J}}_{b-1}\tilde{\boldsymbol{\beta}}_{b-1}\right\} \end{aligned} \tag{S11}$$

where $\tilde{\boldsymbol{J}}_{b-1}=\sum_{j=1}^{b-1}\boldsymbol{J}_j(\tilde{\boldsymbol{\beta}}_j;D_j)$. We define $c_{N_{b-1}}(D^{\star}_{b-1})=\exp\left\{\sum_{j=1}^{b-1}\ell_j(\tilde{\boldsymbol{\beta}}_{b-1};D_j)-\frac{1}{2}\tilde{\boldsymbol{\beta}}_{b-1}^T\tilde{\boldsymbol{J}}_{b-1}\tilde{\boldsymbol{\beta}}_{b-1}\right\}$, which does not depend on the unknown parameter $\boldsymbol{\beta}$. It follows that the factor in equation (S11) containing $\boldsymbol{\beta}$ is given by

$$\begin{aligned} g_{N_{b-1}}(D^{\star}_{b-1};\boldsymbol{\beta},\phi) &= \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^T\tilde{\boldsymbol{J}}_{b-1}\boldsymbol{\beta}+\boldsymbol{\beta}^T\tilde{\boldsymbol{J}}_{b-1}\tilde{\boldsymbol{\beta}}_{b-1}+O_p\left(N_{b-1}^{-1/2}\right)\right\} \\ &= g(S_{N_{b-1}}(D^{\star}_{b-1});\boldsymbol{\beta})+O_p\left(N_{b-1}^{-1/2}\right), \end{aligned} \tag{S12}$$

where $g(\boldsymbol{s};\boldsymbol{\beta})=g(\boldsymbol{s}_1,\boldsymbol{s}_2;\boldsymbol{\beta})=\exp\left(-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{s}_2\boldsymbol{\beta}+\boldsymbol{\beta}^T\boldsymbol{s}_2\boldsymbol{s}_1\right)$, and $g(S_{N_{b-1}}(D^{\star}_{b-1});\boldsymbol{\beta})$ depends on $D^{\star}_{b-1}$ only through two functions $S_1(D^{\star}_{b-1})=\tilde{\boldsymbol{\beta}}_{b-1}$ and $S_2(D^{\star}_{b-1})=\tilde{\boldsymbol{J}}_{b-1}$. Therefore, according to Definition 1, $S(D^{\star}_{b-1})=(S_1(D^{\star}_{b-1}),S_2(D^{\star}_{b-1}))=(\tilde{\boldsymbol{\beta}}_{b-1},\tilde{\boldsymbol{J}}_{b-1})$ is an approximate sufficient statistic for $\boldsymbol{\beta}$ in the generalized linear models.

The above procedure can be easily extended to the case with unknown ϕ_0 being replaced by an estimator. To elucidate it, we present an example of the linear model below.

EXAMPLE 1. **(Linear model)** *Let* $\{y_i;\boldsymbol{x}_i\}_{i=1}^{N_b}$ *be an independent sample drawn from a linear model,* $y_i=\boldsymbol{x}_i^T\boldsymbol{\beta}+\epsilon_i$, *where* $\epsilon_i\overset{iid}{\sim}N(0,\phi)$. *At time* b, *given an unbiased estimator* $\tilde{\phi}_{b-1}$, *the joint pdf of the cumulative historical dataset up to time* $b-1$ *is*

$$\begin{aligned} f(D^{\star}_{b-1};\boldsymbol{\beta},\tilde{\phi}_{b-1}) &= (2\pi\tilde{\phi}_{b-1})^{-\frac{N_{b-1}}{2}}\exp\left\{-\frac{1}{2\tilde{\phi}_{b-1}}\boldsymbol{\beta}^T\tilde{\boldsymbol{J}}_{b-1}\boldsymbol{\beta}+\frac{1}{\tilde{\phi}_{b-1}}\boldsymbol{\beta}^T\tilde{\boldsymbol{J}}_{b-1}\tilde{\boldsymbol{\beta}}_{b-1}\right\} \\ &\quad \times\exp\left\{-\frac{1}{2\tilde{\phi}_{b-1}}\sum_{j=1}^{b-1}\boldsymbol{y}_j^T(\boldsymbol{y}_j-2\boldsymbol{X}_j\tilde{\boldsymbol{\beta}}_{b-1})-\frac{1}{\tilde{\phi}_{b-1}}\tilde{\boldsymbol{\beta}}_{b-1}^T\tilde{\boldsymbol{J}}_{b-1}\tilde{\boldsymbol{\beta}}_{b-1}\right\}. \end{aligned} \tag{S13}$$

Let $c_{N_{b-1}}(D^{\star}_{b-1};\tilde{\phi}_{b-1})=(2\pi\tilde{\phi}_{b-1})^{-\frac{N_{b-1}}{2}}\exp\left\{-\frac{1}{2\tilde{\phi}_{b-1}}\sum_{j=1}^{b-1}\boldsymbol{y}_j^T(\boldsymbol{y}_j-2\boldsymbol{X}_j\tilde{\boldsymbol{\beta}}_{b-1})-\frac{1}{\tilde{\phi}_{b-1}}\tilde{\boldsymbol{\beta}}_{b-1}^T\tilde{\boldsymbol{J}}_{b-1}\tilde{\boldsymbol{\beta}}_{b-1}\right\}$, *which is independent of parameter* $\boldsymbol{\beta}$. *According to Definition 1, the factor in the first line in equation* (S13) *containing* $\boldsymbol{\beta}$ *depends on* $D^{\star}_{b-1}$ *only through two functions* $S_1(D^{\star}_{b-1})=\tilde{\boldsymbol{\beta}}_{b-1}$ *and* $S_2(D^{\star}_{b-1})=\tilde{\phi}_{b-1}^{-1}\tilde{\boldsymbol{J}}_{b-1}$. *It is easy to identify* $g(\boldsymbol{s};\boldsymbol{\beta})=g(\boldsymbol{s}_1,\boldsymbol{s}_2;\boldsymbol{\beta})=\exp\left(-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{s}_2\boldsymbol{\beta}+\boldsymbol{\beta}^T\boldsymbol{s}_2\boldsymbol{s}_1\right)$. *Note that in linear regression model, the factorization in Definition 1 holds exactly with an approximate sufficient statistic* $S(D^{\star}_{b-1})=(S_1(D^{\star}_{b-1}),S_2(D^{\star}_{b-1}))=(\tilde{\boldsymbol{\beta}}_{b-1},\tilde{\phi}_{b-1}^{-1}\tilde{\boldsymbol{J}}_{b-1})$ *for* $\boldsymbol{\beta}$.

S3. Additional Simulation Results

Tables S1, S2 and S3 report the additional simulation results concerning the impacts of varying data batch size n_b on the renewable estimation and incremental inference in the linear model (Table S1), the logistic model (Table S2) and the log-linear model (Table S3).

Table S4 lists the empirical type I error and power based on 500 replications for the hypothesis testing problem considered in Section 6.3. That is, $H_0:\beta_{01}=0.2$ vs. $H_A:\beta_{01}\neq 0.2$, where β_{01} is the intercept parameter in the logistic model.

For the ease of comparison, Figure S1 presents a pictorial summary of all the results obtained from the three GLMs, namely the linear, logistic and log-linear models, in the simulation scenario 1.

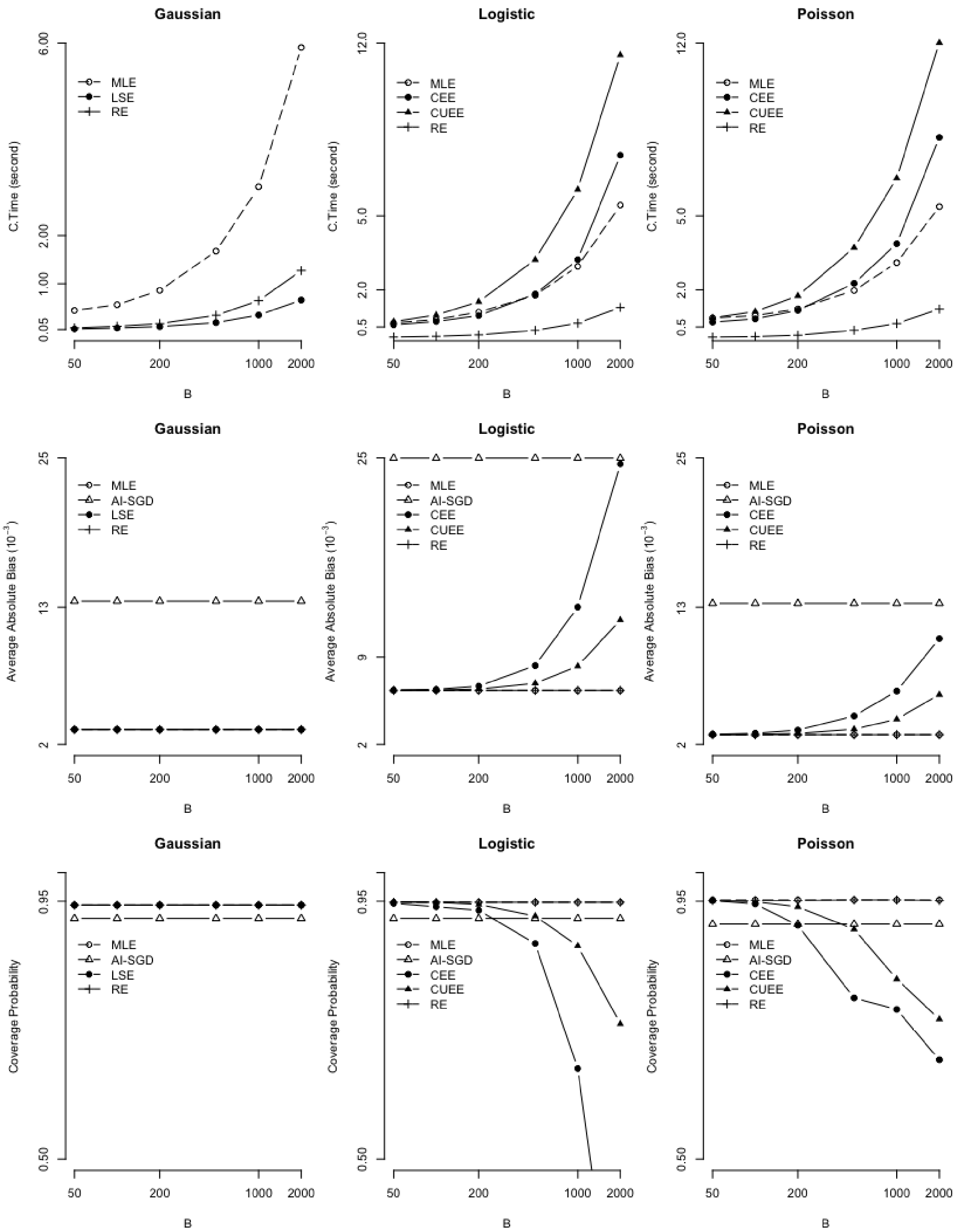

Fig. S1. Average computation time, average bias and coverage probabilities for MLE, AI-SGD, online LSE, sequential CEE and CUEE, and Renewable estimation. AI-SGD is not included in C.Time comparison.

Table S1: Simulation results summarized from 500 replications, under the setting of $N_B = 100,000$ and $p = 5$ for the linear model. Batch size n_b varies from 50 to 2000.

	$B = 50, n_b = 2000$				$B = 100, n_b = 1000$			
	MLE	AI-SGD	online LSE	Renew	MLE	AI-SGD	online LSE	Renew
A.bias$\times 10^{-3}$	3.17	13.48	3.17	3.17	3.17	13.4	3.17	3.17
ASE$\times 10^{-3}$	3.83	15.08	3.82	3.83	3.83	15.08	3.83	3.83
ESE$\times 10^{-3}$	3.94	17.24	3.94	3.94	3.94	17.24	3.94	3.94
CP	0.94	0.92	0.94	0.94	0.94	0.92	0.94	0.94
C.Time(s)	0.44	-	0.06	0.08	0.56	-	0.08	0.12
R.Time(s)	0.31	0.15	0.02	0.04	0.32	0.14	0.02	0.07
	$B = 200, n_b = 500$				$B = 500, n_b = 200$			
	MLE	AI-SGD	online LSE	Renew	MLE	AI-SGD	online LSE	Renew
Abs. bias$\times 10^{-3}$	3.17	13.48	3.17	3.32	3.17	13.48	3.17	3.32
ASE$\times 10^{-3}$	3.83	15.08	3.83	4.08	3.83	15.08	3.83	4.08
ESE$\times 10^{-3}$	3.94	17.24	3.94	3.94	3.94	17.24	3.94	3.94
CP	0.94	0.92	0.94	0.95	0.94	0.92	0.94	0.95
C.Time(s)	0.86	-	0.11	0.17	1.68	-	0.19	0.34
R.Time(s)	0.32	0.14	0.04	0.11	0.30	0.14	0.07	0.24
	$B = 1000, n_b = 100$				$B = 2000, n_b = 50$			
	MLE	AI-SGD	online LSE	Renew	MLE	AI-SGD	online LSE	Renew
A.bias$\times 10^{-3}$	3.17	13.48	3.17	3.17	3.17	13.48	3.17	3.17
ASE$\times 10^{-3}$	3.83	15.08	3.83	3.83	3.83	15.08	3.83	3.83
ESE$\times 10^{-3}$	3.94	17.24	3.94	3.94	3.94	17.24	3.94	3.94
CP	0.94	0.92	0.94	0.94	0.94	0.92	0.94	0.94
C.Time(s)	3.012	-	0.348	0.648	5.906	-	0.660	1.273
R.Time(s)	0.29	0.14	0.14	0.47	0.29	0.14	0.28	0.95

Figure S2 shows the Q-Q plots of the 500 replicates of the Wald test statistic under the settings of the hypothesis testing in Section 6.3. The (1,1)-th plot is to check the null distribution of χ^2_1 under $H_0 : \beta_{01} = 0.2$ for the intercept parameter. The other remaining plots are produced under the null hypotheses with 2-5 regression parameters, respectively. In all cases, the quantiles are distributed closely along the 45° diagonal, indicating the validity of asymptotic χ^2_p distribution, $p = 1, \ldots, 5$.

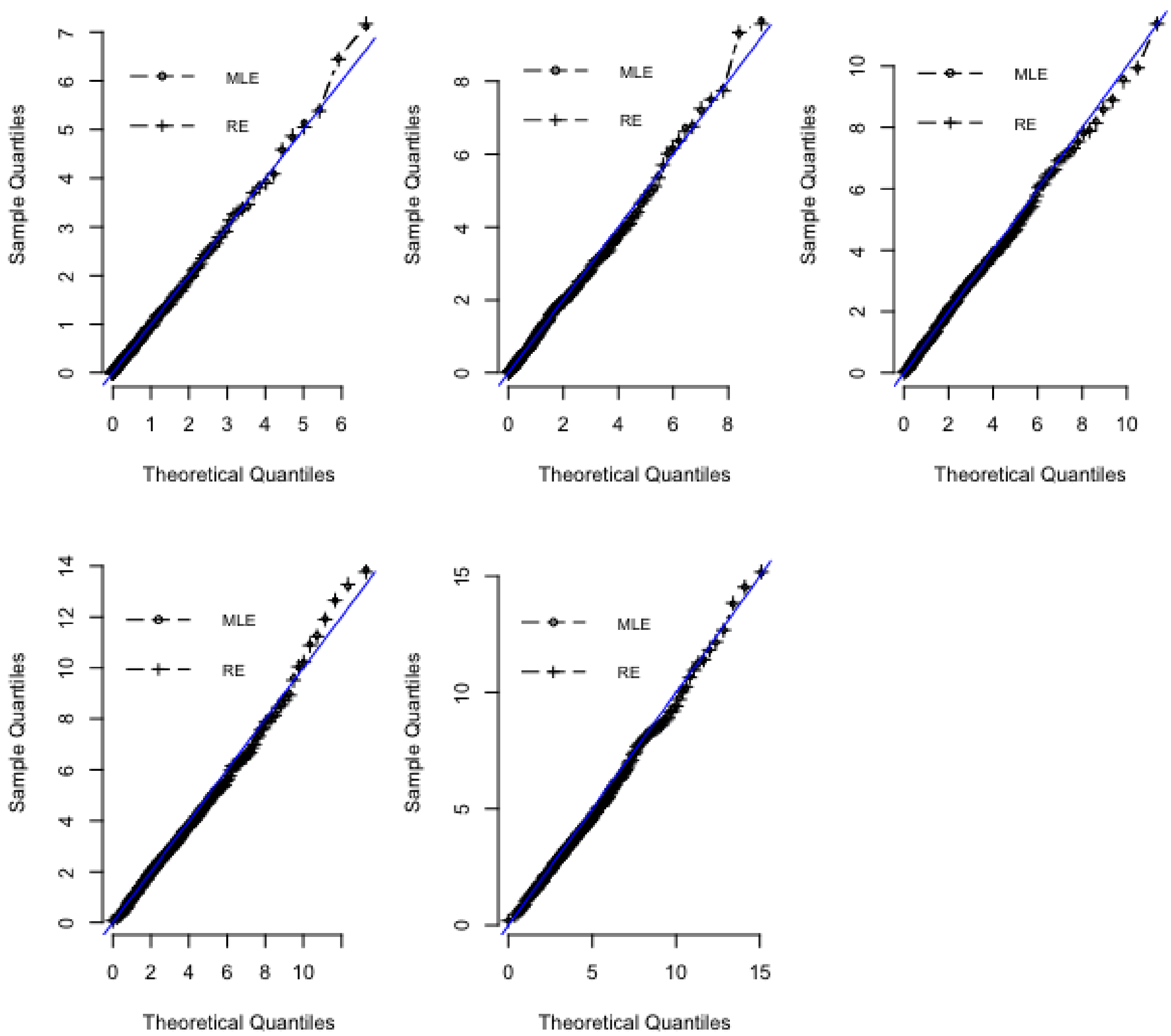

Fig. S2. Quantiles of the Wald test statistics under H_0 with degrees of freedom equal to $1, 2, 3, 4, 5$.

Table S2: Simulation results summarized from 500 replications, under the setting of $N_B = 100,000$ and $p = 5$ for the Binomial logistic model. Batch size n_b varies from 50 to 2000.

	$B=50, n_b=2000$					$B=100, n_b=1000$				
	MLE	AI-SGD	CEE	CUEE	Renew	MLE	AI-SGD	CEE	CUEE	Renew
A.bias$\times 10^{-3}$	6.31	24.98	6.33	6.32	6.32	6.31	24.98	6.40	6.34	6.32
ASE$\times 10^{-3}$	7.82	27.10	7.83	7.82	7.82	7.82	27.10	7.84	7.83	7.82
ESE$\times 10^{-3}$	7.93	31.14	7.90	7.92	7.92	7.93	31.14	7.88	7.93	7.92
CP	0.95	0.92	0.95	0.95	0.95	0.95	0.92	0.94	0.95	0.95
C.Time(s)	0.66	-	0.59	0.71	0.08	0.80	-	0.72	0.99	0.12
R.Time(s)	0.54	0.19	0.55	0.68	0.05	0.58	0.19	0.67	0.94	0.08
	$B=200, n_b=500$					$B=500, n_b=200$				
	MLE	AI-SGD	CEE	CUEE	Renew	MLE	AI-SGD	CEE	CUEE	Renew
A.bias$\times 10^{-3}$	6.31	24.98	6.66	6.42	6.32	6.31	24.98	8.31	6.89	6.32
ASE$\times 10^{-3}$	7.82	27.10	7.87	7.84	7.82	7.82	27.10	7.94	7.86	7.82
ESE$\times 10^{-3}$	7.93	31.14	7.82	7.99	7.92	7.93	31.14	7.67	8.43	7.93
CP	0.95	0.92	0.93	0.94	0.95	0.95	0.92	0.88	0.92	0.95
C.Time(s)	1.10	-	0.96	1.52	0.18	1.78	-	1.83	3.22	0.36
R.Time(s)	0.58	0.19	0.89	1.45	0.12	0.54	0.19	1.70	3.08	0.26
	$B=1000, n_b=100$					$B=2000, n_b=50$				
	MLE	AI-SGD	CEE	CUEE	Renew	MLE	AI-SGD	CEE	CUEE	Renew
A.bias$\times 10^{-3}$	6.31	24.98	**13.01**	**8.26**	6.32	6.31	24.98	**24.50**	**11.98**	6.32
ASE$\times 10^{-3}$	7.82	27.10	8.07	7.89	7.82	7.82	27.10	8.34	7.94	7.82
ESE$\times 10^{-3}$	7.93	31.14	7.45	10.01	7.92	7.93	31.14	7.02	15.64	7.92
CP	0.95	0.92	**0.66**	**0.87**	0.95	0.95	0.92	**0.12**	**0.74**	0.95
C.Time(s)	2.95	-	3.22	6.07	0.65	5.43	-	7.46	11.51	1.29
R.Time(s)	0.52	0.19	2.99	5.82	0.48	0.54	0.20	6.96	11.05	0.95

Table S3: Simulation results summarized from 500 replications, under the setting of $N_B = 100,000$ and $p = 5$ for the Poisson log-linear model. Batch size n_b varies from 50 to 2000.

	$B=50, n_b=2000$					$B=100, n_b=1000$				
	MLE	AI-SGD	CEE	CUEE	Renew	MLE	AI-SGD	CEE	CUEE	Renew
A.bias$\times 10^{-3}$	2.76	13.30	2.80	2.77	2.76	2.76	13.30	2.87	2.78	2.76
ASE$\times 10^{-3}$	3.42	15.15	3.42	3.42	3.42	3.42	15.15	3.42	3.42	3.42
ESE$\times 10^{-3}$	3.42	15.99	3.42	3.42	3.42	3.42	15.99	3.42	3.43	3.42
CP	0.95	0.91	0.95	0.95	0.95	0.95	0.91	0.95	0.95	0.95
C.Time(s)	0.86	-	0.70	0.87	0.09	0.96	-	0.83	1.12	0.11
R.Time(s)	0.74	0.16	0.65	0.82	0.05	0.73	0.16	0.78	1.07	0.07
	$B=200, n_b=500$					$B=500, n_b=200$				
	MLE	AI-SGD	CEE	CUEE	Renew	MLE	AI-SGD	CEE	CUEE	Renew
Abs. bias$\times 10^{-3}$	2.76	13.30	3.13	2.86	2.76	2.76	13.30	**4.27**	3.20	2.76
ASE$\times 10^{-3}$	3.42	15.15	3.42	3.42	3.42	3.42	15.15	3.42	3.42	3.42
ESE$\times 10^{-3}$	3.42	15.99	3.42	3.50	3.42	3.42	15.99	3.42	3.82	3.42
CP	0.95	0.91	0.91	0.94	0.95	0.95	0.91	**0.78**	0.90	0.95
C.Time(s)	1.23	-	1.17	1.76	0.16	1.98	-	2.27	3.71	0.36
R.Time(s)	0.74	0.16	1.09	1.68	0.11	0.73	0.16	2.13	3.56	0.26
	$B=1000, n_b=100$					$B=2000, n_b=50$				
	MLE	AI-SGD	CEE	CUEE	Renew	MLE	AI-SGD	CEE	CUEE	Renew
A.bias$\times 10^{-3}$	2.76	13.30	**6.26**	**4.01**	2.76	2.76	13.30	**10.48**	**5.98**	2.76
ASE$\times 10^{-3}$	3.42	15.15	3.41	3.41	3.42	3.42	15.15	3.41	3.39	3.42
ESE$\times 10^{-3}$	3.42	15.99	3.42	4.59	3.42	3.42	15.99	3.43	6.82	3.42
CP	0.95	0.91	**0.76**	**0.81**	0.95	0.95	0.91	**0.67**	**0.74**	0.95
C.Time(s)	3.10	-	3.87	6.52	0.64	5.37	-	8.18	12.01	1.22
R.Time(s)	0.70	0.16	3.64	6.28	0.46	0.69	0.15	7.71	11.57	0.91

Table S4: Empirical size and power of a simple hypothesis test over 500 replications in the logistic regression model with $p = 5, n_b = 200, B = 500$.

	Size	Power									
$\boldsymbol{\beta}_{03}$	0.2	0.205	0.210	0.215	0.220	0.225	0.230	0.235	0.240	0.245	0.250
MLE	0.050	0.098	0.358	0.664	0.880	0.980	0.998	0.998	1.000	1.000	1.000
AI-SGD	0.050	0.051	0.058	0.102	0.136	0.196	0.288	0.352	0.452	0.534	0.608
CEE	0.110	0.044	0.082	0.324	0.628	0.876	0.972	0.996	0.998	1.000	1.000
CUEE	0.062	0.072	0.268	0.570	0.830	0.952	0.990	0.998	1.000	1.000	1.000
Renew	0.048	0.094	0.354	0.656	0.878	0.980	0.998	0.998	1.000	1.000	1.000

References

Casella, G. and Berger, R. L. (2002) *Statistical Inference.* Duxbury Press, 2nd edn.

Lin, N. and Xi, R. (2011) Aggregated estimating equation estimation. *Statistics and Its Interface*, **4**, 73–83.

Schifano, E. D., Wu, J., Wang, C., Yan, J. and Chen, M.-H. (2016) Online updating of statistical inference in the big data setting. *Technometrics*, **58**, 393–403.