SCALABLE ESTIMATION FOR HIGH VELOCITY SURVIVAL DATA ABLE TO ACCOMMODATE ADDITION OF COVARIATES

Ying Sheng[1], Yifei Sun[2], Charles E. McCulloch[3] and Chiung-Yu Huang[3]

[1]*Chinese Academy of Sciences,* [2]*Columbia University and*

[3]*University of California at San Francisco*

Supplementary Material

S1 Additional numerical simulations

S1.1 Comparison of computational efficiency

We conducted Monte-Carlo simulations to compare computational efficiency and storage memory between the proposed hybrid likelihood based scalable estimator in Section 2.1 and the oracle maximum likelihood estimator (MLE) calculated using the entire dataset. The covariates $X_1, \ldots, X_p$ were independently generated from the standard normal distribution. The survival time T was generated from the Cox proportional hazards model $\lambda(t \mid \boldsymbol{X}) = 1.5\sqrt{t}\exp(\beta_1 X_1 + \ldots + \beta_p X_p)$, where $\beta_1 = \ldots = \beta_{p/2} = 0.25$

and $\beta_{p/2+1} = \ldots = \beta_p = -0.25$, where $p = 10, 50$. The censoring time C was generated from uniform distributions that yielded the censoring rate of 25%. A total of n_B^c observations were generated and then divided into 500 data batches with equal sample sizes, where $n_B^c = 10^6, 10^7, 10^8$. The evaluation criteria for computational efficiency include (i) data loading time (LTime) in seconds; (ii) computation time (CTime) in seconds, which refers to the total amount of time required by data loading and algorithm execution. For each pair of n_B^c and p, we calculated Ltime, CTime and data storage memory for deriving the proposed scalable estimator $\widehat{\boldsymbol{\beta}}_B$ and the oracle MLE $\widetilde{\boldsymbol{\beta}}_B^c$. The computation was performed using the `R` statistical software version 3.6.0 on UCSF Computation Biology and Informatics core at the shared high-performance compute cluster C4 with 150GB of memory. The oracle MLE was derived by using the function `coxph` in the `R` package `survival`. The results are presented in Table 1.

In summary, the proposed approach shows substantial advantages of much lower costs of data storage and computation time when compared with the oracle MLE. Since only historical summary statistics and IPD in the current batch are stored, the required storage memory for calculating the oracle MLE is approximately 500 times that for calculating the proposed scalable estimator. Moreover, the proposed approach can greatly reduce the

Table 1: Summary of computational efficiency and storage memory

	LTime		Ctime		Memory	
	$\widetilde{\boldsymbol{\beta}}_B^c$	$\widehat{\boldsymbol{\beta}}_B$	$\widetilde{\boldsymbol{\beta}}_B^c$	$\widehat{\boldsymbol{\beta}}_B$	$\widetilde{\boldsymbol{\beta}}_B^c$	$\widehat{\boldsymbol{\beta}}_B$
$n_B^c = 10^6$						
$p = 10$	35.9	11.5	44.6	16.2	194.8 MB	401 KB
$p = 50$	302.7	94.5	364.7	120.8	924.8 MB	1.9 MB
$n_B^c = 10^7$						
$p = 10$	643.4	157.0	804.2	215.4	2.0 GB	4.0 MB
$p = 50$	4743.9	1117.8	5387.4	1375.6	9.3 GB	18.6 MB
$n_B^c = 10^8$						
$p = 10$	7319.9	1449.5	8472.5	2089.0	20.5 GB	40.9 MB
$p = 50$	-	17811.5	-	20382.7	93.1 GB	186.2 MB

NOTE: $\widetilde{\boldsymbol{\beta}}_B^c$, the oracle maximum likelihood estimator calculated using the entire n_B^c observations; $\widehat{\boldsymbol{\beta}}_B$, the proposed hybrid likelihood scalable estimator; LTime, data loading time in seconds; Ctime, computation time in seconds for data loading and algorithm execution; Memory, required storage memory for calculating the estimators.

computation time for data loading and algorithm execution and gains more computational efficiency with larger n_B^c and p. The ratios of computation time of the proposed scalable estimator to that of the oracle MLE range from 2.8 to 4.1 across different values of n_B^c and p. It is worthwhile to point out that in the case of $n_B^c = 10^8$ and $p = 50$, it is infeasible to calculate the oracle MLE using the entire dataset while the proposed approach was able to complete the computation in a few hours.

S1.2 Small batch sizes

We conducted additional simulations to investigate the performance of the proposed hybrid likelihood approach when batch sample sizes n_b's are relatively small. The covariates X_1, X_2, X_3 were independently generated from the standard normal distribution and X_4 was generated from a Bernoulli distribution with $\Pr(X_4 = 1) = 0.5$. The survival time T was generated from the Cox proportional hazards model $\lambda(t \mid \boldsymbol{X}) = 2t \exp(\boldsymbol{\beta}^\intercal \boldsymbol{X})$, where $\boldsymbol{\beta} = (0.5, 0.5, 0.5, -1)^\intercal$. The censoring time C was generated from uniform distributions that yielded censoring rates of 25%, 50% and 75%. A total of 100 data batches with the equal sample size $n = 200$ were generated. Table 2 summarizes the simulation results of the oracle MLE $\widetilde{\boldsymbol{\beta}}_B^c$ calculated using all nB observations and the proposed scalable estimator $\widehat{\boldsymbol{\beta}}_B$ up to

the Bth batch based on 1,000 replications. As shown in the table, the absolute of empirical biases and the empirical standard errors of the scalable estimator $\widehat{\boldsymbol{\beta}}_B$ decrease as B increases. Moreover, the performance of the proposed scalable estimator $\widehat{\boldsymbol{\beta}}_B$ is similar to that of the oracle MLE $\widetilde{\boldsymbol{\beta}}_B^c$ in all scenarios. The results confirm good performance of the proposed hybrid likelihood approach in the case of relatively small batch sample sizes.

S1.3 The hybrid likelihood ratio test

We conducted two sets of numerical simulations to evaluate finite-sample performance of the proposed hybrid likelihood ratio test. In all simulations, the covariates X_1, X_2, X_3 were independently generated from the standard normal distribution and X_4 was generated from a Bernoulli distribution with $\Pr(X_4 = 1) = 0.5$. The censoring time C was generated from uniform distributions that yielded censoring rates of 25%, 50% and 75%.

In the first set of simulations, we evaluated the type I error rate of the hybrid likelihood ratio test. A total of 100 data batches with the equal sample size $n = 400$ were generated. In all data batches, the survival time T was generated from the Cox proportional hazards model $\lambda(t \mid \boldsymbol{X}) = 2t\exp(\boldsymbol{\beta}^\intercal \boldsymbol{X})$, where $\boldsymbol{\beta} = (0.5, 0.5, 0.5, -1)^\intercal$. The null hypothesis is $H_0 : \boldsymbol{\beta}_{B-1}^c = \boldsymbol{\beta}_B$. For the nominal level 5%, Table 3 summarizes the empirical

Table 2: Summary of simulation results under the Cox model

		$B = 50$						$B = 100$					
		$\widetilde{\boldsymbol{\beta}}_B^c$			$\widehat{\boldsymbol{\beta}}_B$			$\widetilde{\boldsymbol{\beta}}_B^c$			$\widehat{\boldsymbol{\beta}}_B$		
Cen		Bias	SE	SEE	Bias	SE	SEE	Bias	SE	SEE	Bias	SE	SEE
25%	β_1	-2	12	12	-3	12	13	-1	9	9	-2	9	9
	β_2	1	12	12	-1	12	13	0	9	9	-1	9	9
	β_3	1	11	12	-1	11	13	1	8	8	-1	9	9
	β_4	1	28	25	4	28	25	1	17	17	4	18	18
50%	β_1	-2	14	15	-4	14	15	-1	10	10	-3	10	11
	β_2	0	14	15	-2	14	15	0	11	10	-2	12	11
	β_3	1	14	15	-2	14	15	1	10	10	-1	10	11
	β_4	0	31	29	9	32	30	2	21	21	7	20	21
75%	β_1	-1	19	20	-5	19	21	-1	14	14	-5	14	15
	β_2	-2	20	20	-6	20	21	0	15	14	-4	15	15
	β_3	1	19	20	-2	19	21	1	14	14	-2	15	15
	β_4	-2	41	42	22	41	42	4	30	30	17	30	30

NOTE: Cen, the censoring rate; the true values of the regression coefficients $(\beta_1, \beta_2, \beta_3, \beta_4)$ are $(0.5, 0.5, 0.5, -1)$; $\widetilde{\boldsymbol{\beta}}_B^c$, the oracle maximum likelihood estimator calculated using observations in the cumulative data up to the Bth batch; $\widehat{\boldsymbol{\beta}}_B$, the proposed hybrid likelihood scalable estimator up to the Bth batch; Bias, SE and SEE, empirical bias ($\times 10^3$), empirical standard error ($\times 10^3$) and empirical mean of the standard error estimates ($\times 10^3$).

probability (%) of rejecting the null hypothesis for B varies from 10 to 100 based on 1,000 replications. As shown in the table, the type I error rate is close to the nominal level.

In the second set of simulations, we investigated the power of the hybrid likelihood ratio test. A total of 100 data batches with the equal sample size $n = 400$ were generated. In the first $(B-1)$ data batches, the survival time T was generated from the Cox proportional hazards model $\lambda(t \mid \boldsymbol{X}) = 2t\exp(\boldsymbol{\beta}^\intercal \boldsymbol{X})$, where $\boldsymbol{\beta} = (0.5, 0.5, 0.5, -1)^\intercal$. For $b = B, \ldots, 100$, the survival time T was generated from the Cox proportional hazards model $\lambda(t \mid \boldsymbol{X}) = 2t\exp(\boldsymbol{\gamma}^\intercal \boldsymbol{X})$. We consider three scenarios for the choice of $\boldsymbol{\gamma}$: (I) $\boldsymbol{\gamma} = (0.25, 0.25, 0.25, -0.5)^\intercal$; (II) $\boldsymbol{\gamma} = (0.25, 0.25, 0.25, -0.25)^\intercal$; (III) $\boldsymbol{\gamma} = (0.5, 0.5, 0.5, -0.1)^\intercal$. The null hypothesis is $H_0 : \boldsymbol{\beta}_{B-1}^c = \boldsymbol{\beta}_B$. For the nominal level 5%, Table 4 summarizes the empirical probability (%) of rejecting the null hypothesis for $B = 2, 50, 100$ based on 1,000 replications. As shown in the table, the power of the proposed hybrid ratio test increases as the cumulative sample size increases or the censoring rate decreases, demonstrating the good performance of the testing procedure in terms of the power.

Table 3: The empirical probability (%) of rejecting $H_0 : \boldsymbol{\beta}^c_{B-1} = \boldsymbol{\beta}_B$ at the nominal level 5%.

B	Cen=25%	Cen=50%	Cen=75%
10	4.7	5.8	3.7
20	4.3	3.9	5.3
30	4.7	4.1	5.6
40	4.7	5.2	5.7
50	4.3	5.4	5.7
60	5.1	4.6	5.6
70	4.7	4.9	5.2
80	5.0	5.3	5.3
90	4.9	5.1	5.1
100	4.9	5.0	5.1

NOTE: Cen, the censoring rate; the true values of the regression coefficients in all data batches are $(\beta_1, \beta_2, \beta_3, \beta_4) = (0.5, 0.5, 0.5, -1)$.

Table 4: The empirical probability (%) of rejecting $H_0 : \boldsymbol{\beta}^{c}_{B-1} = \boldsymbol{\beta}_B$ at the nominal level 5%.

Scenario	Cen	$B = 2$	$B = 50$	$B = 100$
(I)				
	25%	99.2	100.0	100.0
	50%	96.4	100.0	100.0
	75%	71.5	96.0	99.0
(II)				
	25%	100.0	100.0	100.0
	50%	99.5	100.0	100.0
	75%	84.5	99.1	100.0
(III)				
	25%	99.5	100.0	100.0
	50%	95.5	100.0	100.0
	75%	68.0	92.5	96.5

NOTE: Cen, the censoring rate; in the first $(B - 1)$ data batches, the true value of regression coefficient is $\boldsymbol{\beta} = (0.5, 0.5, 0.5, -1)^{\intercal}$; in the bth $(b \geq B)$ data batch, the true value of regression coefficient is $\boldsymbol{\gamma}$: (I) $\boldsymbol{\gamma} = (0.25, 0.25, 0.25, -0.5)^{\intercal}$; (II) $\boldsymbol{\gamma} = (0.25, 0.25, 0.25, -0.25)^{\intercal}$; (III) $\boldsymbol{\gamma} = (0.5, 0.5, 0.5, -0.1)^{\intercal}$.

S2 Additional analysis of SEER breast cancer dataset

To show how the parameter estimates evolve over batches, we have calculated the parameter estimates on each batch by running the coxph() function. Figure 1 shows the coefficient estimates of batches 1–4 . The covariates include age at diagnosis (≥ 50 years vs. < 50 years), estrogen receptors (ER, positive vs. negative), progesterone receptors (PR, positive vs. negative), cancer grade (III/IV vs. I/II), and race (White, African American, and other). As shown in the figure, the coefficient estimates are quite consistent across different data batches. We conducted the hybrid likelihood ratio test proposed in Section 2.2 to check the homogeneity assumptions $H_0 : \boldsymbol{\beta}_B = \boldsymbol{\beta}^c_{B-1}$, $B = 2, 3, 4$. The corresponding p values are 0.081, 0.643, and 0.129, indicating that the null hypotheses $\boldsymbol{\beta}_B = \boldsymbol{\beta}^c_{B-1}$, $B = 2, 3, 4$, are not rejected at a nominal level of 0.05.

The SEER Program began collecting new data items related to breast cancer prognosis under the Collaborative Stage (CS) Data Collection System since 2004. Therefore, data on CS tumor size are available only for cases diagnosed after 2004, that is, the 5th data batch. With the addition of the covariate CS tumor size, we conducted the hybrid empirical likelihood ratio test proposed in Section 3.3 to check the conformity of historical covariate effects information in the reduced Cox model. The p value is 0.866

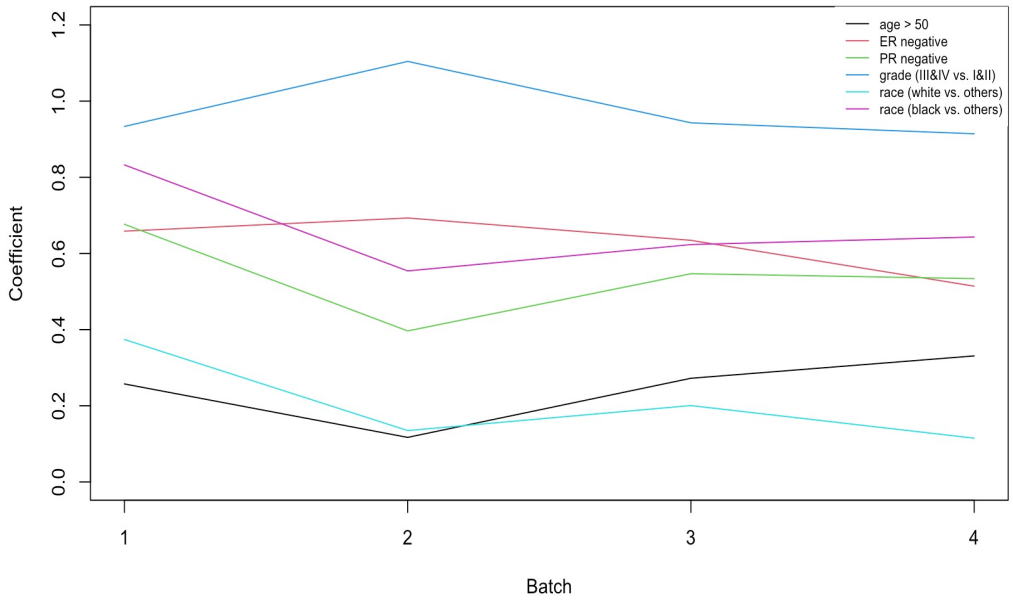

Figure 1: The coefficient estimates of batches 1–4 for SEER breast cancer study.

and the null hypothesis is not rejected.

Figure 2 shows the coefficient estimates of batches 5–10, with the addition of the covariate CS tumor size. As shown in the figure, the coefficient estimates are quite consistent across different data batches. We conducted the hybrid likelihood ratio test to check the homogeneity assumptions $H_0 : \boldsymbol{\beta}_B = \boldsymbol{\beta}^c_{B-1}$, $B = 6, \ldots, 10$. The corresponding p values are 0.729, 0.125, 0.997, 0.583, and 0.223, indicating that the null hypotheses $\boldsymbol{\beta}_B = \boldsymbol{\beta}^c_{B-1}$, $B = 6, \ldots, 10$, are not rejected.

An updated edition of the Collaborative Stage Data Collection System began since 2010. Therefore, covariates HER2 and AJCC stage 7th edition

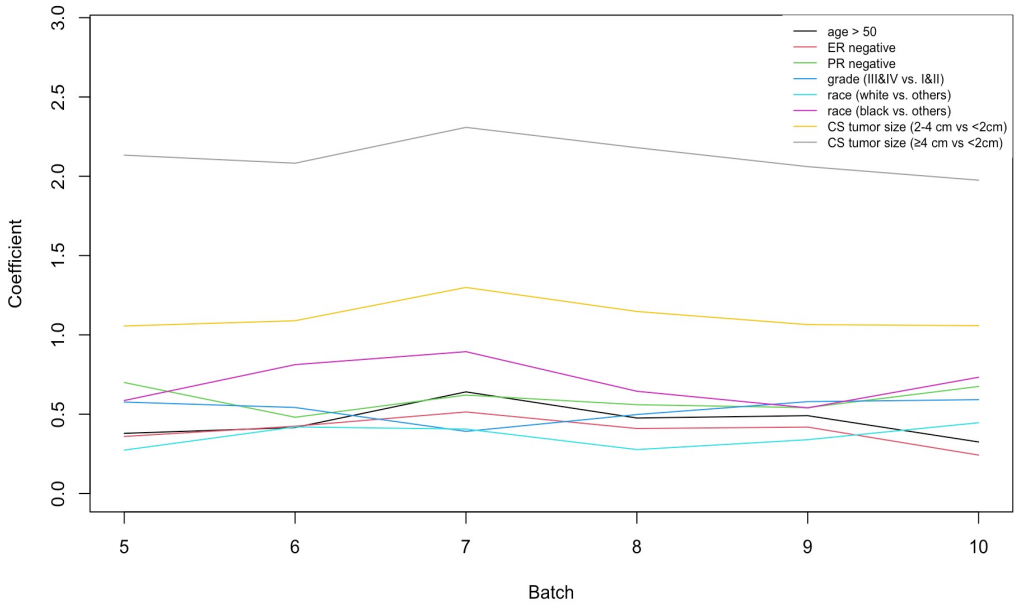

Figure 2: The coefficient estimates of batches 5–10 for SEER breast cancer study.

are available only for cases diagnosed after 2010, that is, the 11th data batch. With the addition of new covariates, we conducted the hybrid empirical likelihood ratio test to check the conformity of historical covariate effects information. The p value is 0.332, and the null hypothesis is not rejected.

S3 Proof of Theorem 1

For $B = 1$, the initial scalable estimator can be estimated by the MLE using the observed IPD in $\mathcal{D}_1$, that is,

$$\widehat{\boldsymbol{\beta}}_1 = \arg\max_{\boldsymbol{\beta}} \sum_{i=1}^{n_1} \Delta_{1i} \left[\boldsymbol{\beta}^\intercal \boldsymbol{X}_{1i} - \log\{S_1^{(0)}(Y_{1i}, \boldsymbol{\beta})\} \right].$$

Applying the standard argument for the MLE, it can be shown that $\sqrt{n_1}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0)$ converges in distribution to a mean zero multivariate normal distribution with the covariance matrix Σ as $n_1 \to \infty$, where

$$\Sigma = \left(\int_0^\infty [s^{(2)}(t, \boldsymbol{\beta}_0) - \{s^{(0)}(t, \boldsymbol{\beta}_0)\}^{-1} \{s^{(1)}(t, \boldsymbol{\beta}_0)\}^{\otimes 2}] \mathrm{d}\Lambda_0(u) \right)^{-1}.$$

With the IPD in $\mathcal{D}_1$, we can estimated Σ by

$$\widehat{\Sigma}_1 = \left(\int_0^\infty [S_1^{(2)}(t, \widehat{\boldsymbol{\beta}}_1) - \{S_1^{(0)}(t, \widehat{\boldsymbol{\beta}}_1)\}^{-1} \{S_1^{(1)}(t, \widehat{\boldsymbol{\beta}}_1)\}^{\otimes 2}] \mathrm{d}\widehat{\Lambda}_1(u, \widehat{\boldsymbol{\beta}}_1) \right)^{-1},$$

where $S_1^{(k)}(t, \boldsymbol{\beta}) = n_1^{-1} \sum_{i=1}^{n_1} I(Y_{1i} \geq t) \exp(\boldsymbol{\beta}^\intercal \boldsymbol{X}_{1i}) \boldsymbol{X}_{1i}^{\otimes k}$ for $k = 0, 1, 2$, and $\widehat{\Lambda}_1(u, \boldsymbol{\beta}) = n_1^{-1} \sum_{i=1}^{n_1} \int_0^t \{S_1^{(0)}(u, \boldsymbol{\beta})\}^{-1} \mathrm{d}N_{1i}(u)$.

For $B \geq 2$, assume that as $n_{B-1}^c \to \infty$, $\sqrt{n_{B-1}^c}(\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta}_0)$ converges in distribution to a mean zero multivariate normal distribution with the covariance matrix Σ. We now study the large-sample properties for the proposed estimator $\widehat{\boldsymbol{\beta}}_B = \arg\max_{\boldsymbol{\beta}} \ell_B(\boldsymbol{\beta})$, where

$$\ell_B(\boldsymbol{\beta}) = \sum_{i=1}^{n_B} \Delta_{Bi} \left[\boldsymbol{\beta}^\intercal \boldsymbol{X}_{Bi} - \log\{S_B^{(0)}(Y_{Bi}, \boldsymbol{\beta})\} \right] - \frac{n_{B-1}^c}{2} (\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta})^\intercal \widehat{\Sigma}_{B-1}^{-1} (\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta}).$$

Note that we have $\mathcal{U}_B(\widehat{\boldsymbol{\beta}}_B) = \mathbf{0}$, where $\mathcal{U}_B(\boldsymbol{\beta}) = \sum_{i=1}^{n_B} \Delta_{Bi} \left\{ \boldsymbol{X}_{Bi} - S_B^{(1)}(Y_{Bi}, \boldsymbol{\beta})/S_B^{(0)}(Y_{Bi}, \boldsymbol{\beta}) \right\} + n_{B-1}^c \widehat{\Sigma}_{B-1}^{-1}(\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta})$. Applying a Taylor expansion of $\mathcal{U}_B(\widehat{\boldsymbol{\beta}}_B)$ at $\boldsymbol{\beta}_0$ yields $\sqrt{n_B^c}(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0) = \Sigma \mathcal{U}_0/\sqrt{n_B^c} + o_p(1)$, where

$$\mathcal{U}_0 = \sum_{i=1}^{n_B} \int_0^\infty \left\{ \boldsymbol{X}_{Bi} - \frac{s^{(1)}(u, \boldsymbol{\beta}_0)}{s^{(0)}(u, \boldsymbol{\beta}_0)} \right\} \mathrm{d}M_{Bi}(u) + n_{B-1}^c \Sigma^{-1}(\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta}_0) \quad \text{(S3.1)}$$

with $M_{Bi}(t) = N_{Bi}(t) - \Lambda_0(t) \exp(\boldsymbol{\beta}_0^\intercal \boldsymbol{X}_{Bi})$. Since the current batch $\mathcal{D}_B$ is independent of the historical cumulative data $\mathcal{D}_{B-1}^c$, it is a standard result that $\mathcal{U}_0/\sqrt{n_B^c}$ converges in distribution to a mean zero multivariate normal distribution with the covariance matrix Σ^{-1} as $n_B^c \to \infty$. Hence as $n_B^c \to \infty$, $\sqrt{n_B^c}(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0)$ converges in distribution to a mean zero multivariate normal distribution with the covariance matrix Σ.

To update $\widehat{\Sigma}_{B-1}$ to $\widehat{\Sigma}_B$, we consider applying a Taylor expansion of $\mathcal{U}_B(\widehat{\boldsymbol{\beta}}_B)$ at $\boldsymbol{\beta}_0$ and derive $\sqrt{n_B^c}(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0) = \sqrt{n_B^c}\{n_B \widetilde{\Sigma}_B(\boldsymbol{\beta}_0)^{-1} + n_{B-1}^c \widehat{\Sigma}_{B-1}^{-1}\}^{-1} \mathcal{U}_B(\boldsymbol{\beta}_0) + o_p(1)$, where $\widetilde{\Sigma}_B(\boldsymbol{\beta}) = \left(\int_0^\infty [S_B^{(2)}(t, \boldsymbol{\beta}) - \{S_B^{(0)}(t, \boldsymbol{\beta})\}^{-1} \{S_B^{(1)}(t, \boldsymbol{\beta})\}^{\otimes 2}] \mathrm{d}\widehat{\Lambda}_B(u, \boldsymbol{\beta}) \right)^{-1}$. This leads to

$$\mathrm{cov}\left\{ \sqrt{n_B^c}(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0) \right\} = n_B^c \{n_B \widetilde{\Sigma}_B(\boldsymbol{\beta}_0)^{-1} + n_{B-1}^c \widehat{\Sigma}_{B-1}^{-1}\}^{-1} + o_p(1).$$

Therefore, we propose to update $\widehat{\Sigma}_B$ by $n_B^c \{n_B \widetilde{\Sigma}_B(\widehat{\boldsymbol{\beta}}_B)^{-1} + n_{B-1}^c \widehat{\Sigma}_{B-1}^{-1}\}^{-1}$.

Finally, we establish the large-sample properties for $\widehat{\Lambda}_B(t, \widehat{\boldsymbol{\beta}}_B)$, where

$$\widehat{\Lambda}_B(t, \boldsymbol{\beta}) = \frac{1}{n_B} \sum_{i=1}^{n_B} \int_0^t \frac{\mathrm{d}N_{Bi}(u)}{S_B^{(0)}(u, \boldsymbol{\beta})}.$$

Applying a Taylor expansion of $\widehat{\Lambda}_B(t,\widehat{\boldsymbol{\beta}}_B)$ at $\boldsymbol{\beta}_0$ yields $\widehat{\Lambda}_B(t,\widehat{\boldsymbol{\beta}}_B)=\widehat{\Lambda}_B(t,\boldsymbol{\beta}_0)+\partial\widehat{\Lambda}_B(t,\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}_B-\boldsymbol{\beta}_0)$, where $\widetilde{\boldsymbol{\beta}}$ lies between $\widehat{\boldsymbol{\beta}}_B$ and $\boldsymbol{\beta}_0$. Some straightforward calculations can yield

$$\frac{\partial\widehat{\Lambda}_B(t,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}=-\int_0^t\frac{s^{(1)}(u,\boldsymbol{\beta}_0)}{s^{(0)}(u,\boldsymbol{\beta}_0)}\mathrm{d}\Lambda_0(u)+o_p(n_B^{-1/2}),$$

$$\widehat{\Lambda}_B(t,\boldsymbol{\beta}_0)-\Lambda_0(t)=\frac{1}{n_B}\sum_{i=1}^{n_B}\int_0^t\frac{\mathrm{d}M_{Bi}(u)}{s^{(0)}(u,\boldsymbol{\beta}_0)}+o_p(n_B^{-1/2}).$$

It follows that

$$\begin{aligned}n_B^{1/2}\{\widehat{\Lambda}_B(t,\widehat{\boldsymbol{\beta}}_B)-\Lambda_0(t)\} &= n_B^{-1/2}\sum_{i=1}^{n_B}\int_0^t\frac{\mathrm{d}M_{Bi}(u)}{s^{(0)}(u,\boldsymbol{\beta}_0)}-\int_0^t\frac{s^{(1)}(u,\boldsymbol{\beta}_0)}{s^{(0)}(u,\boldsymbol{\beta}_0)}\mathrm{d}\Lambda_0(u)n_B^{1/2}(\widehat{\boldsymbol{\beta}}_B-\boldsymbol{\beta}_0)+o_p(1)\\ &= n_B^{-1/2}\sum_{i=1}^{n_B}\int_0^t\frac{\mathrm{d}M_{Bi}(u)}{s^{(0)}(u,\boldsymbol{\beta}_0)}-\frac{\sqrt{n_B}}{n_B^c}\int_0^t\frac{s^{(1)}(u,\boldsymbol{\beta}_0)}{s^{(0)}(u,\boldsymbol{\beta}_0)}\mathrm{d}\Lambda_0(u)\Sigma\mathcal{U}_0,\end{aligned}$$

where $\mathcal{U}_0$ is defined by (S3.1). Hence as $n_B^c\to\infty$, $\sqrt{n_B}\{\widehat{\Lambda}_B(t,\widehat{\boldsymbol{\beta}}_B)-\Lambda_0(t)\}$ converges in distribution to a mean zero normal distribution with the variance

$$\int_0^t\frac{\mathrm{d}\Lambda_0(u)}{s^{(0)}(u,\boldsymbol{\beta}_0)}+\frac{\kappa_B}{1+\kappa_B}\mathcal{V}(t)^\intercal\Sigma\mathcal{V}(t),$$

where $n_B/n_{B-1}^c\to\kappa_B$ as $n_B^c\to\infty$ and $\mathcal{V}(t)=\int_0^t s^{(1)}(u,\boldsymbol{\beta}_0)/s^{(0)}(u,\boldsymbol{\beta}_0)\mathrm{d}\Lambda_0(u)$.

S4 Large-sample properties for the hybrid likelihood ratio test statistics

In this section, we establish the large-sample properties for the proposed hybrid likelihood ratio test statistic

$$R_1 = 2\left\{\sup_{\boldsymbol{\beta}_B,\boldsymbol{\beta}_{B-1}^c} \ell(\boldsymbol{\beta}_B,\boldsymbol{\beta}_{B-1}^c) - \sup_{\boldsymbol{\beta}_B=\boldsymbol{\beta}_{B-1}^c} \ell(\boldsymbol{\beta}_B,\boldsymbol{\beta}_{B-1}^c)\right\},$$

where

$$\ell(\boldsymbol{\beta}_B,\boldsymbol{\beta}_{B-1}^c) = \sum_{i=1}^{n_B} \Delta_{Bi}\left[\boldsymbol{\beta}_B^\intercal \boldsymbol{X}_{Bi} - \log\{S_B^{(0)}(Y_{Bi},\boldsymbol{\beta}_B)\}\right] - \frac{n_{B-1}^c}{2}(\widehat{\boldsymbol{\beta}}_{B-1}-\boldsymbol{\beta}_{B-1}^c)^\intercal \widehat{\Sigma}_{B-1}^{-1}(\widehat{\boldsymbol{\beta}}_{B-1}-\boldsymbol{\beta}_{B-1}^c).$$

Without the constraint $\boldsymbol{\beta}_B = \boldsymbol{\beta}_{B-1}^c$, the hybrid likelihood $\ell(\boldsymbol{\beta}_B,\boldsymbol{\beta}_{B-1}^c)$ is maximized by $\boldsymbol{\beta}_B = \widetilde{\boldsymbol{\beta}}_B$ and $\boldsymbol{\beta}_{B-1}^c = \widehat{\boldsymbol{\beta}}_{B-1}$, where $\widetilde{\boldsymbol{\beta}}_B$ is the MLE calculated using the observed IPD in $\mathcal{D}_B$. This leads to $R_1 = 2\{\ell(\widetilde{\boldsymbol{\beta}}_B,\widehat{\boldsymbol{\beta}}_{B-1}) - \ell(\widehat{\boldsymbol{\beta}}_B,\widehat{\boldsymbol{\beta}}_B)\}$. Applying a Taylor expansion of $\ell(\widehat{\boldsymbol{\beta}}_B,\widehat{\boldsymbol{\beta}}_B)$ at $(\widetilde{\boldsymbol{\beta}}_B,\widehat{\boldsymbol{\beta}}_{B-1})$ yields

$$\begin{aligned} R_1 &= 2\{\ell(\widetilde{\boldsymbol{\beta}}_B,\widehat{\boldsymbol{\beta}}_{B-1}) - \ell(\widehat{\boldsymbol{\beta}}_B,\widehat{\boldsymbol{\beta}}_B)\} \\ &= n_B(\widehat{\boldsymbol{\beta}}_B - \widetilde{\boldsymbol{\beta}}_B)^\intercal \Sigma^{-1}(\widehat{\boldsymbol{\beta}}_B - \widetilde{\boldsymbol{\beta}}_B) + n_{B-1}^c(\widehat{\boldsymbol{\beta}}_B - \widehat{\boldsymbol{\beta}}_{B-1})^\intercal \Sigma^{-1}(\widehat{\boldsymbol{\beta}}_B - \widehat{\boldsymbol{\beta}}_{B-1}) + o_p(1). \end{aligned}$$

Note that we have

$$\sum_{i=1}^{n_B} \Delta_{Bi}\left\{\boldsymbol{X}_{Bi} - \frac{S_B^{(1)}(Y_{Bi},\widehat{\boldsymbol{\beta}}_B)}{S_B^{(0)}(Y_{Bi},\widehat{\boldsymbol{\beta}}_B)}\right\} + n_{B-1}^c \widehat{\Sigma}_{B-1}^{-1}(\widehat{\boldsymbol{\beta}}_{B-1} - \widehat{\boldsymbol{\beta}}_B) = \mathbf{0},$$

$$\sum_{i=1}^{n_B} \Delta_{Bi}\left\{\boldsymbol{X}_{Bi} - \frac{S_B^{(1)}(Y_{Bi},\widetilde{\boldsymbol{\beta}}_B)}{S_B^{(0)}(Y_{Bi},\widetilde{\boldsymbol{\beta}}_B)}\right\} = \mathbf{0}.$$

It follows that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_B - \widehat{\boldsymbol{\beta}}_{B-1} &= \frac{1}{n^c_{B-1}}\widehat{\Sigma}_{B-1}\sum_{i=1}^{n_B}\Delta_{Bi}\left\{\boldsymbol{X}_{Bi} - \frac{S^{(1)}_B(Y_{Bi},\widehat{\boldsymbol{\beta}}_B)}{S^{(0)}_B(Y_{Bi},\widehat{\boldsymbol{\beta}}_B)}\right\} \\ &= \frac{n_B}{n^c_{B-1}}(\widehat{\boldsymbol{\beta}}_B - \widetilde{\boldsymbol{\beta}}_B) + o_p(\sqrt{n_B}/n^c_{B-1}).\end{aligned}$$

Hence we can derive

$$R_1 = \frac{n^c_B n^c_{B-1}}{n_B}(\widehat{\boldsymbol{\beta}}_B - \widehat{\boldsymbol{\beta}}_{B-1})^\intercal \Sigma^{-1}(\widehat{\boldsymbol{\beta}}_B - \widehat{\boldsymbol{\beta}}_{B-1}) + o_p(1).$$

It can be shown that

$$\begin{aligned}\sqrt{\frac{n^c_B n^c_{B-1}}{n_B}}(\widehat{\boldsymbol{\beta}}_B - \widehat{\boldsymbol{\beta}}_{B-1}) &= \sqrt{\frac{n^c_B n^c_{B-1}}{n_B}}\left\{\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0 - (\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta}_0)\right\} \\ &= \sqrt{\frac{n^c_B n^c_{B-1}}{n_B}}\left\{\Sigma\mathcal{U}_0/n^c_B - (\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta}_0)\right\} + o_p(1),\end{aligned}$$

where $\mathcal{U}_0$ is defined by (S3.1). Hence $\sqrt{n^c_B n^c_{B-1}/n_B}(\widehat{\boldsymbol{\beta}}_B - \widehat{\boldsymbol{\beta}}_{B-1})$ converges in distribution to a mean zero normal distribution with the variance Σ. This indicates that R_1 converges in distribution a χ^2 distribution with d degrees of freedom as $n^c_B \to \infty$.

S5 Scalable estimation under the reduced Cox model

In this section, we present the details for updating the estimation of $\boldsymbol{\beta}$ under the reduced Cox model and establish large-sample properties for the scalable estimator $\widehat{\boldsymbol{\beta}}_B$, $B = 1, \ldots, B^* - 1$. The observed IPD in the bth data batch are denoted by $\mathcal{D}_b = (\boldsymbol{Y}_b, \boldsymbol{X}_b, \boldsymbol{\Delta}_b)$, $b \leq B^* - 1$. For $k = 0, 1, 2,$

define $S_b^{(k)}(t,\boldsymbol{\beta}) = n_b^{-1}\sum_{i=1}^{n_b} I(Y_{bi}\geq t)\exp(\boldsymbol{\beta}^\intercal \boldsymbol{X}_{bi})\boldsymbol{X}_{bi}^{\otimes k}$ and let $s^{(k)}(t,\boldsymbol{\beta}) = E\{I(Y\geq t)\exp(\boldsymbol{\beta}^\intercal\boldsymbol{X})\boldsymbol{X}^{\otimes k}\}$, where the expectation is evaluated under the full Cox model. The initial estimator $\widehat{\boldsymbol{\beta}}_1$ can be derived by maximizing $\sum_{i=1}^{n_1}\Delta_{1i}\left[\boldsymbol{\beta}^\intercal\boldsymbol{X}_{1i} - \log\{S_1^{(0)}(Y_{1i},\boldsymbol{\beta})\}\right]$. For $b\leq B^*-1$ and $i=1,\ldots,n_b$, define

$$\begin{aligned} h_{bi}^*(\boldsymbol{\beta}) &= \int_0^\infty \left\{\boldsymbol{X}_{bi} - \frac{s^{(1)}(t,\boldsymbol{\beta})}{s^{(0)}(t,\boldsymbol{\beta})}\right\} \\ &\quad\times\left[\mathrm{d}N_{bi}(t) - I(Y_{bi}\geq t)\exp(\boldsymbol{\beta}^\intercal\boldsymbol{X}_{bi})\frac{E\{dN(t)\}}{s^{(0)}(t,\boldsymbol{\beta})}\right]. \end{aligned} \tag{S5.2}$$

Moreover, define

$$\Omega_{\boldsymbol{X}} = \int_0^\infty \left[\frac{s^{(2)}(t,\boldsymbol{\beta}_0)}{s^{(0)}(t,\boldsymbol{\beta}_0)} - \left\{\frac{s^{(1)}(t,\boldsymbol{\beta}_0)}{s^{(0)}(t,\boldsymbol{\beta}_0)}\right\}^{\otimes 2}\right] E\{\mathrm{d}N(t)\}, \tag{S5.3}$$

$$\begin{aligned} Q &= \mathrm{cov}\{h_{11}^*(\boldsymbol{\beta}_0)\} \\ &= \int_0^\infty E\left[\left\{\boldsymbol{X} - \frac{s^{(1)}(t,\boldsymbol{\beta}_0)}{s^{(0)}(t,\boldsymbol{\beta}_0)}\right\}^{\otimes 2}\exp(\boldsymbol{\theta}_0^\intercal\boldsymbol{Z})I(Y\geq t)\right]\mathrm{d}\Lambda_0^*(t). \end{aligned} \tag{S5.4}$$

Following Huang and Qin (2020), we can derive $\sqrt{n_1}(\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_0) = \Omega_{\boldsymbol{X}}^{-1}n_1^{-1/2}\sum_{i=1}^{n_1}h_{1i}^*(\boldsymbol{\beta}_0)+o_p(1)$, where $h_{1i}^*(\boldsymbol{\beta})$ is defined by (S5.2) with $b=1$. Hence as $n_1\to\infty$, $\sqrt{n_1}(\widehat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_0)$ converges in distribution to a mean zero normal random variable with the covariance matrix $\Sigma_1 = \Omega_{\boldsymbol{X}}^{-1}Q\Omega_{\boldsymbol{X}}^{-1}$. For $b=1,\ldots,B^*-1$ and $i=1,\ldots,n_b$, let

$$\widehat{\Omega}_{\boldsymbol{X}}(\mathcal{D}_b;\boldsymbol{\beta}) = \int_0^\infty \left[\frac{S_b^{(2)}(t,\boldsymbol{\beta})}{S_b^{(0)}(t,\boldsymbol{\beta})} - \left\{\frac{S_b^{(1)}(t,\boldsymbol{\beta})}{S_b^{(0)}(t,\boldsymbol{\beta})}\right\}^{\otimes 2}\right]\left\{\frac{1}{n_b}\sum_{i=1}^{n_b}dN_{bi}(t)\right\} \tag{S5.5}$$

$$\widehat{Q}(\mathcal{D}_b;\boldsymbol{\beta}) = \frac{1}{n_b}\sum_{i=1}^{n_b}\widehat{h}^*_{bi}(\mathcal{D}_b;\boldsymbol{\beta})\widehat{h}^*_{bi}(\mathcal{D}_b;\boldsymbol{\beta})^\intercal, \tag{S5.6}$$

where

$$\widehat{h}^*_{bi}(\mathcal{D}_b;\boldsymbol{\beta}) = \int_0^\infty \left\{\boldsymbol{X}_{bi} - \frac{S_b^{(1)}(t,\boldsymbol{\beta})}{S_b^{(0)}(t,\boldsymbol{\beta})}\right\}\left[\mathrm{d}N_{bi}(t) - I(Y_{bi}\geq t)\exp(\boldsymbol{\beta}^\intercal\boldsymbol{X}_{bi})\frac{n_b^{-1}\sum_{i=1}^{n_b}dN_{bi}(t)}{S_1^{(0)}(t,\boldsymbol{\beta})}\right].$$

With the observed IPD in $\mathcal{D}_1$, we can derive a consistent estimator for Σ_1, that is,

$$\widehat{\Sigma}_1 = \widehat{\Omega}_{\boldsymbol{X}}(\mathcal{D}_1;\widehat{\boldsymbol{\beta}}_1)^{-1}\widehat{Q}(\mathcal{D}_1;\widehat{\boldsymbol{\beta}}_1)\widehat{\Omega}_{\boldsymbol{X}}(\mathcal{D}_1;\widehat{\boldsymbol{\beta}}_1)^{-1}.$$

Therefore, in a special case of $B^* = 2$, we have established the asymptotic normality of $\widehat{\boldsymbol{\beta}}_{B^*-1}$ and derived a consistent estimator of the corresponding asymptotic covariance matrix.

We now study the case of $B^* > 2$. For $B = 2,\ldots,B^*-1$, assume that $\sqrt{n^c_{B-1}}(\widehat{\boldsymbol{\beta}}_{B-1}-\boldsymbol{\beta})$ converges in distribution to a mean zero normal random variable with the covariance matrix Σ_{B-1} and denoted by $\widehat{\Sigma}_{B-1}$ a consistent estimator of Σ_{B-1}. We propose to update $\widehat{\boldsymbol{\beta}}_B$ by

$$\widehat{\boldsymbol{\beta}}_B = \arg\max_{\boldsymbol{\beta}}\left(\sum_{i=1}^{n_B}\Delta_{Bi}\left[\boldsymbol{\beta}^\intercal\boldsymbol{X}_{Bi} - \log\{S_B^{(0)}(Y_{Bi},\boldsymbol{\beta})\}\right] - \frac{n^c_{B-1}}{2}(\widehat{\boldsymbol{\beta}}_{B-1}-\boldsymbol{\beta})^\intercal\widehat{\Sigma}^{-1}_{B-1}(\widehat{\boldsymbol{\beta}}_{B-1}-\boldsymbol{\beta})\right).$$

Hence we have $\mathcal{U}_B(\widehat{\boldsymbol{\beta}}_B) = \mathbf{0}$, where $\mathcal{U}_B(\boldsymbol{\beta}) = \sum_{i=1}^{n_B}\Delta_{Bi}\{\boldsymbol{X}_{Bi} - S_B^{(1)}(Y_{Bi},\boldsymbol{\beta})/S_B^{(0)}(Y_{Bi},\boldsymbol{\beta})\} + n^c_{B-1}\widehat{\Sigma}^{-1}_{B-1}(\widehat{\boldsymbol{\beta}}_{B-1}-\boldsymbol{\beta})$. Applying a Taylor expansion of $\mathcal{U}_B(\boldsymbol{\beta})$ at $\boldsymbol{\beta}_0$ can yield

$$\sqrt{n^c_B}(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0) = \sqrt{n^c_B}\left(n_B\Omega_{\boldsymbol{X}} + n^c_{B-1}\Sigma^{-1}_{B-1}\right)^{-1}\left\{\sum_{i=1}^{n_B}h^*_{Bi}(\boldsymbol{\beta}_0) + n^c_{B-1}\Sigma^{-1}_{B-1}(\widehat{\boldsymbol{\beta}}_{B-1}-\boldsymbol{\beta}_0)\right\} + o_p(1).$$

Since $\mathrm{var}\left\{n_B^{-1/2}\sum_{i=1}^{n_B} h_{Bi}^*(\boldsymbol{\beta}_0)\right\} = Q$ and $\mathrm{var}\left\{\sqrt{n_{B-1}^c}(\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta}_0)\right\} = \Sigma_{B-1}$, we can derive

$$\mathrm{var}\left\{\sum_{i=1}^{n_B} h_{Bi}^*(\boldsymbol{\beta}_0) + n_{B-1}^c \Sigma_{B-1}^{-1}(\widehat{\boldsymbol{\beta}}_{B-1} - \boldsymbol{\beta}_0)\right\} = n_B Q + n_{B-1}^c \Sigma_{B-1}^{-1}.$$

Hence as $n_B^c \to \infty$, $\sqrt{n_B^c}(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_0)$ converges in distribution to a mean zero normal random variable with the covariance matrix

$$\Sigma_B = (1 + \kappa_B^{-1})\left(\Omega_{\boldsymbol{X}} + \kappa_B^{-1}\Sigma_{B-1}^{-1}\right)^{-1}\left(Q + \kappa_B^{-1}\Sigma_{B-1}^{-1}\right)\left(\Omega_{\boldsymbol{X}} + \kappa_B^{-1}\Sigma_{B-1}^{-1}\right)^{-1} \quad \text{(S5.7)}$$

where $n_B/n_{B-1}^c \to \kappa_B$ as $n_B^c \to \infty$. Moreover, we can update $\widehat{\Sigma}_B$ by

$$n_B^c\left\{n_B\widehat{\Omega}_{\boldsymbol{X}}(\mathcal{D}_B;\widehat{\boldsymbol{\beta}}_B) + n_{B-1}^c\widehat{\Sigma}_{B-1}^{-1}\right\}^{-1}\left\{n_B\widehat{Q}(\mathcal{D}_B;\widehat{\boldsymbol{\beta}}_B) + n_{B-1}^c\widehat{\Sigma}_{B-1}^{-1}\right\} \quad \text{(S5.8)}$$
$$\times\left\{n_B\widehat{\Omega}_{\boldsymbol{X}}(\mathcal{D}_B;\widehat{\boldsymbol{\beta}}_B) + n_{B-1}^c\widehat{\Sigma}_{B-1}^{-1}\right\}^{-1},$$

where $\widehat{\Omega}_{\boldsymbol{X}}(\mathcal{D}_B;\boldsymbol{\beta})$ and $\widehat{Q}(\mathcal{D}_B;\boldsymbol{\beta})$ are defined by (S5.5) and (S5.6) with $b = B$. Therefore, as $n_{B^*-1}^c \to \infty$, $\sqrt{n_{B^*-1}^c}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta}_0)$ converges in distribution to a mean zero normal random variable with the covariance matrix Σ_{B^*-1}, which is defined by (S5.7) with $\Sigma_1 = \Omega_{\boldsymbol{X}}^{-1}Q\Omega_{\boldsymbol{X}}^{-1}$ and $B = B^* - 1$. The consistent estimator of Σ_{B^*-1} up to the $(B^* - 1)$th batch can be derived using (S5.8) with $B = B^* - 1$.

S6 Proof of Theorem 2

We first prove the consistency of the proposed scalable estimators. Let $\boldsymbol{\gamma} = (\boldsymbol{\theta}^\intercal, \boldsymbol{\beta}^\intercal)^\intercal$ and define $U_i(\boldsymbol{\gamma}) = (\widehat{g}_i(\boldsymbol{\theta})^\intercal, \widehat{h}_i(\boldsymbol{\beta}, \boldsymbol{\theta})^\intercal)^\intercal$, where

$$\widehat{g}_i(\boldsymbol{\theta}) = \int_0^\infty \left\{ \boldsymbol{Z}_{B^*i} - \frac{S^{(1)}_{B^*,\boldsymbol{Z}}(t,\boldsymbol{\theta})}{S^{(0)}_{B^*,\boldsymbol{Z}}(t,\boldsymbol{\theta})} \right\} \left\{ \mathrm{d}N_{B^*i}(t) - I(Y_{B^*i} \geq t) \exp(\boldsymbol{\theta}^\intercal \boldsymbol{Z}_{B^*i}) \mathrm{d}\widetilde{\Lambda}^*_{B^*}(t,\boldsymbol{\theta}) \right\},$$

$$\widehat{h}_i(\boldsymbol{\beta}, \boldsymbol{\theta}) = \int_0^\infty \left\{ \boldsymbol{X}_{B^*i} - \frac{S^{(1)}_{B^*}(t,\boldsymbol{\beta})}{S^{(0)}_{B^*}(t,\boldsymbol{\beta})} \right\} \left\{ \mathrm{d}N_{B^*i}(t) - I(Y_{B^*i} \geq t) \exp(\boldsymbol{\beta}^\intercal \boldsymbol{X}_{B^*i}) \frac{S^{(0)}_{B^*,\boldsymbol{Z}}(t,\boldsymbol{\theta})}{S^{(0)}_{B^*}(t,\boldsymbol{\beta})} \mathrm{d}\widetilde{\Lambda}^*_{B^*}(t,\boldsymbol{\theta}) \right\}.$$

Let $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^\intercal, \boldsymbol{\xi}_2^\intercal)^\intercal$ and let $\ell(\boldsymbol{\gamma}, \boldsymbol{\xi}) = -\sum_{i=1}^{n_{B^*}} \log\{1 + \boldsymbol{\xi}^\intercal U_i(\boldsymbol{\gamma})\} - \frac{n^c_{B^*-1}}{2}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta})^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta})$. The proposed scalable estimator $\widehat{\boldsymbol{\gamma}}_{B^*}$ is defined by

$$\widehat{\boldsymbol{\gamma}}_{B^*} = \arg\max_{\boldsymbol{\gamma}} \min_{\boldsymbol{\xi}} \ell(\boldsymbol{\gamma}, \boldsymbol{\xi}).$$

For any fixed $\boldsymbol{\gamma}$, let $\boldsymbol{\xi}(\boldsymbol{\gamma}) = \arg\min_{\boldsymbol{\xi}} \ell(\boldsymbol{\gamma}, \boldsymbol{\xi})$ and write $\ell(\boldsymbol{\gamma}) = \ell(\boldsymbol{\gamma}, \boldsymbol{\xi}(\boldsymbol{\gamma}))$. Following Zhang et al. (2020), the consistency of $\widehat{\boldsymbol{\gamma}}_{B^*}$ can be proved if we can show that $\ell(\boldsymbol{\gamma}) < \ell(\boldsymbol{\gamma}_0)$ for any $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + u n_{B^*}^{-1/3}$, where u is a $(d+q)$-dimensional vector with $\|u\| = 1$. Similar to the proof in Owen (1990), it can be shown that $\boldsymbol{\xi}(\boldsymbol{\gamma}) = O_p(n_{B^*}^{-1/3})$. Along the same lines in the proof of Lemma 1 in Zhang et al. (2020), for any $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + u n_{B^*}^{-1/3}$, we have $-\sum_{i=1}^{n_{B^*}} \log\{1 + \boldsymbol{\xi}(\boldsymbol{\gamma})^\intercal U_i(\boldsymbol{\gamma})\} \leq -c_1 n_{B^*}^{1/3}/2$, where $c_1 > 0$ is the smallest eigenvalue of

$$E\left\{\frac{\partial U(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}}\right\}^\intercal \left[E\{U(\boldsymbol{\gamma}_0)U(\boldsymbol{\gamma}_0)^\intercal\}\right]^{-1} E\left\{\frac{\partial U(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}}\right\}.$$

Similarly, we can derive $-\sum_{i=1}^{n_{B^*}} \log\{1+\boldsymbol{\xi}(\boldsymbol{\gamma}_0)^\intercal U_i(\boldsymbol{\gamma}_0)\} = O_p(\log\log n_{B^*})$.

It follows that

$$-\sum_{i=1}^{n_{B^*}} \log\{1+\boldsymbol{\xi}(\boldsymbol{\gamma})^\intercal U_i(\boldsymbol{\gamma})\} + \sum_{i=1}^{n_{B^*}} \log\{1+\boldsymbol{\xi}(\boldsymbol{\gamma}_0)^\intercal U_i(\boldsymbol{\gamma}_0)\} < 0.$$

Moreover, applying a Taylor expansion can yield

$$\begin{aligned}
&-\frac{n^c_{B^*-1}}{2}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta})^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}) + \frac{n^c_{B^*-1}}{2}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}_0)^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}_0)\\
&= n^c_{B^*-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}_0) - \frac{n^c_{B^*-1}}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + o_p(n^{1/3}_{B^*})\\
&\le n^c_{B^*-1}\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\|\widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}_0)\| - \frac{n^c_{B^*-1}}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + o_p(n^{1/3}_{B^*})\\
&< 0.
\end{aligned}$$

Hence we have shown that

$$\begin{aligned}
\ell(\boldsymbol{\gamma})-\ell(\boldsymbol{\gamma}_0) \;=\; & -\sum_{i=1}^{n_{B^*}} \log\{1+\boldsymbol{\xi}(\boldsymbol{\gamma})^\intercal U_i(\boldsymbol{\gamma})\} + \sum_{i=1}^{n_{B^*}} \log\{1+\boldsymbol{\xi}(\boldsymbol{\gamma}_0)^\intercal U_i(\boldsymbol{\gamma}_0)\}\\
& -\frac{n^c_{B^*-1}}{2}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta})^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}) + \frac{n^c_{B^*-1}}{2}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}_0)^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}_0)\\
& < 0.
\end{aligned}$$

This completes the proof of consistency of $\widehat{\boldsymbol{\gamma}}_{B^*}$.

We then establish the asymptotic normality for the proposed hybrid empirical likelihood estimator $\widehat{\boldsymbol{\theta}}_{B^*}$. Since $\widetilde{\Lambda}^*_{B^*}(t,\boldsymbol{\theta}) = n^{-1}_{B^*}\sum_{i=1}^{n_{B^*}}\int_0^t \{S^{(0)}_{B^*,\boldsymbol{Z}}(u,\boldsymbol{\theta})\}^{-1}\mathrm{d}N_{B^*i}(u)$, we have $\partial\widehat{h}_i(\boldsymbol{\beta},\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$. Following Huang and Qin (2020), it can be proved that $n^{-1}_{B^*}\sum_{i=1}^{n_{B^*}}\widehat{g}_i(\boldsymbol{\theta}_0) = n^{-1}_{B^*}\sum_{i=1}^{n_{B^*}} g_i(\boldsymbol{\theta}_0)+o_p(n^{-1/2}_{B^*})$ and $n^{-1}_{B^*}\sum_{i=1}^{n_{B^*}}\widehat{h}_i(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0) =$

$n_{B^*}^{-1}\sum_{i=1}^{n_{B^*}} h_i(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)+o_p(n_{B^*}^{-1/2})$, where

$$g_i(\boldsymbol{\theta})=\int_0^\infty\left\{\boldsymbol{Z}_{B^*i}-\frac{s_{\boldsymbol{Z}}^{(1)}(t,\boldsymbol{\theta})}{s_{\boldsymbol{Z}}^{(0)}(t,\boldsymbol{\theta})}\right\}\left\{\mathrm{d}N_{B^*i}(t)-I(Y_{B^*i}\geq t)\exp(\boldsymbol{\theta}^\intercal\boldsymbol{Z}_{B^*i})\mathrm{d}\Lambda_0^*(t)\right\},$$

$$h_i(\boldsymbol{\theta},\boldsymbol{\beta})=\int_0^\infty\left\{\boldsymbol{X}_{B^*i}-\frac{s^{(1)}(t,\boldsymbol{\beta})}{s^{(0)}(t,\boldsymbol{\beta})}\right\}\left\{\mathrm{d}N_{B^*i}(t)-I(Y_{B^*i}\geq t)\exp(\boldsymbol{\beta}^\intercal\boldsymbol{X}_{B^*i})\frac{s_{\boldsymbol{Z}}^{(0)}(t,\boldsymbol{\theta})}{s^{(0)}(t,\boldsymbol{\beta})}\mathrm{d}\Lambda_0^*(t)\right\}.$$

Therefore, we have $\mathrm{var}\left\{n_{B^*}^{-1/2}\sum_{i=1}^{n_{B^*}}\widehat{g}_i(\boldsymbol{\theta}_0)\right\}=\mathrm{var}\{g_1(\boldsymbol{\theta}_0)\}+o_p(1)$, $\mathrm{var}\left\{n_{B^*}^{-1/2}\sum_{i=1}^{n_{B^*}}\widehat{h}_i(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)\right\}=\mathrm{var}\{h_1(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)\}+o_p(1)$, and $\mathrm{cov}\left\{n_{B^*}^{-1/2}\sum_{i=1}^{n_{B^*}}\widehat{g}_i(\boldsymbol{\theta}_0),n_{B^*}^{-1/2}\sum_{i=1}^{n_{B^*}}\widehat{h}_i(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)\right\}=\mathrm{cov}\{g_1(\boldsymbol{\theta}_0),h_1(\boldsymbol{\theta}_0,\boldsymbol{\beta}_0)\}+o_p(1)$. Let $\widehat{\boldsymbol{\gamma}}_{B^*}=(\widehat{\boldsymbol{\theta}}_{B^*}^\intercal,\widehat{\boldsymbol{\beta}}_{B^*}^\intercal)^\intercal$ and denoted by $\widehat{\boldsymbol{\xi}}_1$ and $\widehat{\boldsymbol{\xi}}_2$ the solutions of

$$\frac{1}{n_{B^*}}\sum_{i=1}^{n_{B^*}}\frac{\widehat{g}_i(\widehat{\boldsymbol{\theta}}_{B^*})}{1+\boldsymbol{\xi}_1^\intercal\widehat{g}_i(\widehat{\boldsymbol{\theta}}_{B^*})+\boldsymbol{\xi}_2^\intercal\widehat{h}_i(\widehat{\boldsymbol{\theta}}_{B^*},\widehat{\boldsymbol{\beta}}_{B^*})}=\mathbf{0},\quad\frac{1}{n_{B^*}}\sum_{i=1}^{n_{B^*}}\frac{\widehat{h}_i(\widehat{\boldsymbol{\theta}}_{B^*},\widehat{\boldsymbol{\beta}}_{B^*})}{1+\boldsymbol{\xi}_1^\intercal\widehat{g}_i(\widehat{\boldsymbol{\theta}}_{B^*})+\boldsymbol{\xi}_2^\intercal\widehat{h}_i(\widehat{\boldsymbol{\theta}}_{B^*},\widehat{\boldsymbol{\beta}}_{B^*})}=\mathbf{0}.$$

Arguing as in the proof of Qin and Lawless (1994) and Zhang et al. (2020), under the given conditions, $\widehat{\boldsymbol{\gamma}}_{B^*}$ is in the interior of the ball $\{\boldsymbol{\gamma}:\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|\leq n_{B^*}^{-1/3}\}$, $\|\widehat{\boldsymbol{\xi}}_1\|=o_p(1)$ and $\|\widehat{\boldsymbol{\xi}}_2\|=o_p(1)$, where $\boldsymbol{\gamma}_0$ is the true value of $\boldsymbol{\gamma}$. Define functions

$$\mathcal{U}_1(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)=-\frac{1}{n_{B^*}}\sum_{i=1}^{n_{B^*}}\frac{\boldsymbol{\xi}_1^\intercal\partial\widehat{g}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}}{1+\boldsymbol{\xi}_1^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})},$$

$$\mathcal{U}_2(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)=-\frac{1}{n_{B^*}}\sum_{i=1}^{n_{B^*}}\frac{\boldsymbol{\xi}_2^\intercal\partial\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})/\partial\boldsymbol{\beta}}{1+\boldsymbol{\xi}_1^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})}+\frac{n_{B^*-1}^c}{n_{B^*}}\widehat{\Sigma}_{B^*-1}^{-1}(\widehat{\boldsymbol{\beta}}_{B^*-1}-\boldsymbol{\beta}),$$

$$\mathcal{U}_3(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)=-\frac{1}{n_{B^*}}\sum_{i=1}^{n_{B^*}}\frac{\widehat{g}_i(\boldsymbol{\theta})}{1+\boldsymbol{\xi}_1^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})},$$

$$\mathcal{U}_4(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)=-\frac{1}{n_{B^*}}\sum_{i=1}^{n_{B^*}}\frac{\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})}{1+\boldsymbol{\xi}_1^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})}.$$

We can derive $\mathcal{U}_i(\widehat{\boldsymbol{\theta}}_{B^*}, \widehat{\boldsymbol{\beta}}_{B^*}, \widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2) = \mathbf{0}$, $i = 1, 2, 3, 4$. Applying a Taylor expansion of $\mathcal{U}_i(\widehat{\boldsymbol{\theta}}_{B^*}, \widehat{\boldsymbol{\beta}}_{B^*}, \widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2)$ at $(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \mathbf{0}, \mathbf{0})$ yields

$$\begin{pmatrix} \mathcal{U}_{10} \\ \mathcal{U}_{20} \\ \mathcal{U}_{30} \\ \mathcal{U}_{40} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & -V^{-1} & \mathbf{0} \\ \mathbf{0} & \kappa_{B^*}^{-1}\Sigma_{B^*-1}^{-1} & \mathbf{0} & \Omega_{\boldsymbol{X}}^{\intercal} \\ -V^{-1} & \mathbf{0} & -V^{-1} & -\Omega \\ \mathbf{0} & \Omega_{\boldsymbol{X}} & -\Omega^{\intercal} & -Q \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_{B^*} - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\beta}}_{B^*} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\xi}}_1 \\ \widehat{\boldsymbol{\xi}}_2 \end{pmatrix} + o_p(1/\sqrt{n_{B^*}}),$$

where $\mathcal{U}_{i0} = \mathcal{U}_i(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0, \mathbf{0}, \mathbf{0})$ for $i = 1, 2, 3, 4$, $n_{B^*}/n_{B^*-1}^c \to \kappa_{B^*}$ as $n_{B^*} \to \infty$, $\Omega_{\boldsymbol{X}}$ is defined by (S5.3), Q is defined by (S5.4), $\Omega = E\{g_1(\boldsymbol{\theta}_0)h_1(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)^{\intercal}\}$, and

$$V = [\mathrm{var}\{g_1(\boldsymbol{\theta}_0)\}]^{-1} = \left(\int_0^{\infty} [s_{\boldsymbol{Z}}^{(2)}(t, \boldsymbol{\theta}_0) - \{s_{\boldsymbol{Z}}^{(0)}(t, \boldsymbol{\theta}_0)\}^{-1}\{s_{\boldsymbol{Z}}^{(1)}(t, \boldsymbol{\theta}_0)\}^{\otimes 2}]\mathrm{d}\Lambda_0^*(u)\right)^{-1}.$$

Let

$$\widetilde{\mathcal{U}}_1 = \begin{pmatrix} \mathcal{U}_{10} \\ \mathcal{U}_{20} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ n_{B^*}^{-1} n_{B^*-1}^c \widehat{\Sigma}_{B^*-1}^{-1}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta}_0) \end{pmatrix},$$

$$\widetilde{\mathcal{U}}_2 = \begin{pmatrix} \mathcal{U}_{30} \\ \mathcal{U}_{40} \end{pmatrix} = \begin{pmatrix} -n_{B^*}^{-1} \sum_{i=1}^{n_{B^*}} g_i(\boldsymbol{\theta}_0) \\ -n_{B^*}^{-1} \sum_{i=1}^{n_{B^*}} h_i(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0) \end{pmatrix}.$$

Some straightforward calculations lead to

$$\sqrt{n_{B^*}}(\widehat{\boldsymbol{\gamma}}_{B^*} - \boldsymbol{\gamma}_0) = (A_1 + A_2 A_3^{-1} A_2^{\intercal})^{-1} \left(\sqrt{n_{B^*}}\widetilde{\mathcal{U}}_1 + A_2 A_3^{-1}\sqrt{n_{B^*}}\widetilde{\mathcal{U}}_2\right) + o_p(1),$$

where

$$A_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \kappa_{B^*}^{-1}\Sigma_{B^*-1}^{-1} \end{pmatrix}, \quad A_2 = \begin{pmatrix} -V^{-1} & \mathbf{0} \\ \mathbf{0} & \Omega_{\boldsymbol{X}}^{\intercal} \end{pmatrix}, \quad A_3 = \begin{pmatrix} V^{-1} & \Omega \\ \Omega^{\intercal} & Q \end{pmatrix}.$$

Due to $\mathrm{var}(\sqrt{n_{B^*}}\widetilde{\mathcal{U}}_1) = A_1$, $\mathrm{var}(\sqrt{n_{B^*}}\widetilde{\mathcal{U}}_2) = A_3$ and $E(\widetilde{\mathcal{U}}_1\widetilde{\mathcal{U}}_2) = \mathbf{0}$, we have

$$\mathrm{var}\left(\sqrt{n_{B^*}}\widetilde{\mathcal{U}}_1 + A_2A_3^{-1}\sqrt{n_{B^*}}\widetilde{\mathcal{U}}_2\right) = A_1 + A_2A_3^{-1}A_2^{\intercal}.$$

Hence as $n_{B^*} \to \infty$, $\sqrt{n_{B^*}}(\widehat{\boldsymbol{\gamma}}_{B^*} - \boldsymbol{\gamma}_0)$ converges in distribution to a zero mean multivariate normal distribution with the covariance matrix $(A_1 + A_2A_3^{-1}A_2^{\intercal})^{-1}$. Some tedious calculations lead to

$$\begin{aligned} A_2A_3^{-1}A_2^{\intercal} &= \begin{pmatrix} -V^{-1} & \mathbf{0} \\ \mathbf{0} & \Omega_{\boldsymbol{X}}^{\intercal} \end{pmatrix} \begin{pmatrix} V^{-1} & \Omega \\ \Omega^{\intercal} & Q \end{pmatrix}^{-1} \begin{pmatrix} -V^{-1} & \mathbf{0} \\ \mathbf{0} & \Omega_{\boldsymbol{X}} \end{pmatrix} \\ &= \begin{pmatrix} -V^{-1} & \mathbf{0} \\ \mathbf{0} & \Omega_{\boldsymbol{X}}^{\intercal} \end{pmatrix} \begin{pmatrix} J^{-1} & -J^{-1}\Omega Q^{-1} \\ -Q^{-1}\Omega^{\intercal}J^{-1} & Q^{-1} + Q^{-1}\Omega^{\intercal}J^{-1}\Omega Q^{-1} \end{pmatrix} \begin{pmatrix} -V^{-1} & \mathbf{0} \\ \mathbf{0} & \Omega_{\boldsymbol{X}} \end{pmatrix} \\ &= \begin{pmatrix} V^{-1}J^{-1}V^{-1} & V^{-1}J^{-1}\Omega Q^{-1}\Omega_{\boldsymbol{X}} \\ \Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega^{\intercal}J^{-1}V^{-1} & \Omega_{\boldsymbol{X}}^{\intercal}(Q^{-1} + Q^{-1}\Omega^{\intercal}J^{-1}\Omega Q^{-1})\Omega_{\boldsymbol{X}} \end{pmatrix}, \end{aligned}$$

where $J = V^{-1} - \Omega Q^{-1}\Omega^{\intercal}$. This leads to

$$A_1 + A_2A_3^{-1}A_2^{\intercal} = \begin{pmatrix} V^{-1}J^{-1}V^{-1} & V^{-1}J^{-1}\Omega Q^{-1}\Omega_{\boldsymbol{X}} \\ \Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega^{\intercal}J^{-1}V^{-1} & \kappa_{B^*}^{-1}\Sigma_{B^*-1}^{-1} + \Omega_{\boldsymbol{X}}^{\intercal}(Q^{-1} + Q^{-1}\Omega^{\intercal}J^{-1}\Omega Q^{-1})\Omega_{\boldsymbol{X}} \end{pmatrix}.$$

Hence the asymptotic covariance matrix of $\sqrt{n_{B^*}}(\widehat{\boldsymbol{\theta}}_{B^*} - \boldsymbol{\theta}_0)$ is

$$\begin{aligned}
\Pi_{B^*} &= V\{J + \Omega Q^{-1}\Omega_{\boldsymbol{X}}(\kappa_{B^*}^{-1}\Sigma_{B^*-1}^{-1} + \Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega_{\boldsymbol{X}})^{-1}\Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega^{\intercal}\}V \\
&= V\{V^{-1} - \Omega Q^{-1}\Omega^{\intercal} + \Omega Q^{-1}\Omega_{\boldsymbol{X}}(\kappa_{B^*}^{-1}\Sigma_{B^*-1}^{-1} + \Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega_{\boldsymbol{X}})^{-1}\Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega^{\intercal}\}V \\
&= V\{V^{-1} - (1+\kappa_{B^*})^{-1}\Omega H\Omega^{\intercal}\}V,
\end{aligned}$$

where Σ_{B^*-1} is defined by (S5.7) with $B = B^* - 1$ and

$$H = (1+\kappa_{B^*})\{Q^{-1} - Q^{-1}\Omega_{\boldsymbol{X}}(\kappa_{B^*}^{-1}\Sigma_{B^*-1}^{-1} + \Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega_{\boldsymbol{X}})^{-1}\Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\}. \quad \text{(S6.9)}$$

Therefore, as $n_{B^*} \to \infty$, $\sqrt{n_{B^*}}(\widehat{\boldsymbol{\theta}}_{B^*} - \boldsymbol{\theta}_0)$ converges in distribution to a zero mean multivariate normal distribution with the covariance matrix Π_{B^*}.

In a special case of $B^* = 2$, we have $\Sigma_{B^*-1} = \Sigma_1 = \Omega_{\boldsymbol{X}}^{-1}Q\Omega_{\boldsymbol{X}}^{-1}$ and thus

$$H = (1+\kappa_{B^*})\{Q^{-1} - Q^{-1}\Omega_{\boldsymbol{X}}(\kappa_{B^*}^{-1}\Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega_{\boldsymbol{X}} + \Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\Omega_{\boldsymbol{X}})^{-1}\Omega_{\boldsymbol{X}}^{\intercal}Q^{-1}\} = Q^{-1}.$$

Hence in a special case of $B^* = 2$, $\sqrt{n_{B^*}}(\widehat{\boldsymbol{\theta}}_{B^*} - \boldsymbol{\theta}_0)$ converges in distribution to a zero mean multivariate normal distribution with the covariance matrix $V\{V^{-1} - (1+\kappa_{B^*})^{-1}\Omega Q^{-1}\Omega^{\intercal}\}V$ as $n_{B^*} \to \infty$.

Using the observed IPD in $\mathcal{D}_{B^*}$, we can estimate V, Ω and Q by $\widehat{V}(\widehat{\boldsymbol{\theta}}_{B^*})$, $\widehat{\Omega}(\widehat{\boldsymbol{\theta}}_{B^*}, \widehat{\boldsymbol{\beta}}_{B^*})$ and $\widehat{Q}(\widehat{\boldsymbol{\theta}}_{B^*}, \widehat{\boldsymbol{\beta}}_{B^*})$, where $\widehat{V}(\boldsymbol{\theta}) = \left\{n_{B^*}^{-1}\sum_{i=1}^{n_{B^*}} \widehat{g}_i(\boldsymbol{\theta})\widehat{g}_i(\boldsymbol{\theta})^{\intercal}\right\}^{-1}$, $\widehat{\Omega}(\boldsymbol{\theta}, \boldsymbol{\beta}) = n_{B^*}^{-1}\sum_{i=1}^{n_{B^*}} \widehat{g}_i(\boldsymbol{\theta})\widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})^{\intercal}$, $\widehat{Q}(\boldsymbol{\theta}, \boldsymbol{\beta}) = n_{B^*}^{-1}\sum_{i=1}^{n_{B^*}} \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})\widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})^{\intercal}$. Moreover, we can estimate κ_{B^*} by $n_{B^*}/n_{B^*-1}^c$ and estimate Σ_{B^*-1} by $\widehat{\Sigma}_{B^*-1}$. Based on (S6.9), we can derive an estimate of H, denoted by $\widehat{H}(\widehat{\boldsymbol{\theta}}_{B^*}, \widehat{\boldsymbol{\beta}}_{B^*})$,

by replacing unknown terms with the corresponding estimates. Hence we propose to estimate Π_{B^*} by $\widehat{\Pi}_{B^*}(\widehat{\boldsymbol{\theta}}_{B^*}, \widehat{\boldsymbol{\beta}}_{B^*})$, where

$$\widehat{\Pi}_{B^*}(\boldsymbol{\theta}, \boldsymbol{\beta}) = \widehat{V}(\boldsymbol{\theta})\{\widehat{V}(\boldsymbol{\theta})^{-1} - (1+n_{B^*}/n^c_{B^*-1})^{-1}\widehat{\Omega}(\boldsymbol{\theta}, \boldsymbol{\beta})\widehat{H}(\boldsymbol{\theta}, \boldsymbol{\beta})\widehat{\Omega}(\boldsymbol{\theta}, \boldsymbol{\beta})^{\intercal}\}\widehat{V}(\boldsymbol{\theta}).$$

S7 Proof of Theorem 3

For $B > B^*$, define $\widehat{\boldsymbol{\theta}}_B = \arg\max_{\boldsymbol{\theta}} \left\{\ell_B(\boldsymbol{\theta}) - (n^c_{B-1} - n^c_{B^*-1})(\widehat{\boldsymbol{\theta}}_{B-1} - \boldsymbol{\theta})^{\intercal}\widehat{\Pi}^{-1}_{B-1}(\widehat{\boldsymbol{\theta}}_{B-1} - \boldsymbol{\theta})/2\right\}$, where $\ell_B(\boldsymbol{\theta}) = \sum_{i=1}^{n_B} \Delta_{Bi}[\boldsymbol{\theta}^{\intercal}\boldsymbol{Z}_{Bi} - \log\{S^{(0)}_{B,\boldsymbol{Z}}(Y_{Bi}, \boldsymbol{\theta})\}]$. Let

$$\mathcal{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n_B} \Delta_{Bi}\left\{\boldsymbol{Z}_{Bi} - \frac{S^{(1)}_{B,\boldsymbol{Z}}(Y_{Bi}, \boldsymbol{\theta})}{S^{(0)}_{B,\boldsymbol{Z}}(Y_{Bi}, \boldsymbol{\theta})}\right\} + (n^c_{B-1} - n^c_{B^*-1})\widehat{\Pi}^{-1}_{B-1}(\widehat{\boldsymbol{\theta}}_{B-1} - \boldsymbol{\theta})$$

and we have $\mathcal{U}(\widehat{\boldsymbol{\theta}}_B) = \mathbf{0}$. Applying a Taylor expansion of $\mathcal{U}(\widehat{\boldsymbol{\theta}}_B)$ at $\boldsymbol{\theta}_0$ can yield

$$(n^c_B - n^c_{B^*-1})^{1/2}(\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_0) = (n^c_B - n^c_{B^*-1})^{1/2}\{n_B V^{-1} + (n^c_{B-1} - n^c_{B^*-1})\Pi^{-1}_{B-1}\}^{-1}\mathcal{U}(\boldsymbol{\theta}_0) + o_p(1).$$

It can be shown that $\mathrm{var}\{\mathcal{U}(\boldsymbol{\theta}_0)\} = n_B V^{-1} + (n^c_{B-1} - n^c_{B^*-1})\Pi^{-1}_{B-1}$. Hence as $n^c_B \to \infty$, $(n^c_B - n^c_{B^*-1})^{1/2}(\widehat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}_0)$ converges in distribution to a zero mean multivariate normal distribution with the covariance matrix

$$\Pi_B = (n^c_B - n^c_{B^*-1})\{n_B V^{-1} + (n^c_{B-1} - n^c_{B^*-1})\Pi^{-1}_{B-1}\}^{-1}, \quad B > B^*. \tag{S7.10}$$

In the following, we show that

$$\Pi_B = (V^{-1} + r_B \Omega \widetilde{H} \Omega^{\intercal})^{-1}, \quad B > B^*. \tag{S7.11}$$

where $\widetilde{H} = \{(1+\kappa_{B^*})H^{-1} - \Omega^\intercal V \Omega\}^{-1}$ and $n_{B^*}(n_B^c - n_{B^*-1}^c)^{-1} \to r_B$ as $n_B^c \to \infty$. In the proof of Theorem 2, we have $\Pi_{B^*} = V\{V^{-1} - (1+\kappa_{B^*})^{-1}\Omega H \Omega^\intercal\}V$ and thus

$$\Pi_{B^*}^{-1} = V^{-1} + \Omega\{(1+\kappa_{B^*})H^{-1} - \Omega^\intercal V \Omega\}^{-1}\Omega^\intercal.$$

It follows that

$$\begin{aligned}
\Pi_{B^*+1} &= (n_{B^*+1}^c - n_{B^*-1}^c)\{n_{B^*+1}V^{-1} + (n_{B^*}^c - n_{B^*-1}^c)\Pi_{B^*}^{-1}\}^{-1} \\
&= (n_{B^*+1}^c - n_{B^*-1}^c)\left[(n_{B^*+1}^c - n_{B^*-1}^c)V^{-1} + n_{B^*}\Omega\{(1+\kappa_{B^*})H^{-1} - \Omega^\intercal V\Omega\}^{-1}\Omega^\intercal\right]^{-1} \\
&= \left[V^{-1} + r_{B^*+1}\Omega\{(1+\kappa_{B^*})H^{-1} - \Omega^\intercal V\Omega\}^{-1}\Omega^\intercal\right]^{-1},
\end{aligned}$$

that is, for $B = B^* + 1$, Equation (S7.11) holds. Assume that for a general $B - 1 \geq B^* + 1$, Equation (S7.11) holds. Hence we can derive

$$\begin{aligned}
\Pi_B &= (n_B^c - n_{B^*-1}^c)\{n_B V^{-1} + (n_{B-1}^c - n_{B^*-1}^c)\Pi_{B-1}^{-1}\}^{-1} \\
&= (n_B^c - n_{B^*-1}^c)\left[n_B V^{-1} + (n_{B-1}^c - n_{B^*-1}^c)\{V^{-1} + r_{B-1}\Omega\{(1+\kappa_{B^*})H^{-1} - \Omega^\intercal V\Omega\}^{-1}\Omega^\intercal\}\right]^{-1} \\
&= (n_B^c - n_{B^*-1}^c)\left[(n_B^c - n_{B^*-1}^c)V^{-1} + (n_{B-1}^c - n_{B^*-1}^c)r_{B-1}\Omega\{(1+\kappa_{B^*})H^{-1} - \Omega^\intercal V\Omega\}^{-1}\Omega^\intercal\right]^{-1} \\
&= \left[V^{-1} + r_B\Omega\{(1+\kappa_{B^*})H^{-1} - \Omega^\intercal V\Omega\}^{-1}\Omega^\intercal\right]^{-1},
\end{aligned}$$

where $r_B = (n_{B-1}^c - n_{B^*-1}^c)(n_B^c - n_{B^*-1}^c)^{-1} r_{B-1}$, that is, $n_{B^*}(n_B^c - n_{B^*-1}^c)^{-1} \to r_B$ as $n_B^c \to \infty$. This completes the proof of (S7.11).

Moreover, based on (S7.10), we can update the estimation of Π_B by

$$\widehat{\Pi}_B = (n_B^c - n_{B^*-1}^c)\{n_B \widehat{V}_B^{-1}(\widehat{\boldsymbol{\theta}}_B) + (n_{B-1}^c - n_{B^*-1}^c)\widehat{\Pi}_{B-1}^{-1}\}^{-1}, \quad B > B^*,$$

where $\widehat{V}_B(\boldsymbol{\theta}) = \left(\int_0^\infty [S^{(2)}_{B,\boldsymbol{Z}}(t,\boldsymbol{\theta}) - \{S^{(0)}_{B,\boldsymbol{Z}}(t,\boldsymbol{\theta})\}^{-1}\{S^{(1)}_{B,\boldsymbol{Z}}(t,\boldsymbol{\theta})\}^{\otimes 2}]\mathrm{d}\widetilde{\Lambda}^*_B(u,\boldsymbol{\theta})\right)^{-1}$, with $\widetilde{\Lambda}^*_B(u,\boldsymbol{\theta})$ denoting the Breslow-type estimator using the IPD in $\mathcal{D}_B$.

S8 Large-sample properties of the hybrid empirical likelihood ratio test statistic

Let $\ell(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) = -\sum_{i=1}^{n_{B^*}} \log\left\{1 + \boldsymbol{\xi}_1^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2^\intercal \widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})\right\} - n^c_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta})^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta})/2$ and define the hybrid empirical likelihood ratio test statistic

$$R_2 = 2\left\{\sup_{\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \ell(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) - \sup_{\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1} \ell(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1,\mathbf{0})\right\}.$$

Without any constraints, the constrained hybrid empirical likelihood $\ell(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1,\mathbf{0})$ is maximized at $(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1,\mathbf{0}) = (\widehat{\boldsymbol{\beta}}_{B^*},\widehat{\boldsymbol{\theta}}_{B^*},\widehat{\boldsymbol{\xi}}_1,\widehat{\boldsymbol{\xi}}_2)$, where $\widehat{\boldsymbol{\beta}}_{B^*}$,$\widehat{\boldsymbol{\theta}}_{B^*}$,$\widehat{\boldsymbol{\xi}}_1$, and $\widehat{\boldsymbol{\xi}}_2$ are given in the proof of Theorem 2. When $\boldsymbol{\xi}_2 = \mathbf{0}$, the constrained hybrid empirical likelihood $\ell(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1,\mathbf{0})$ is maximized at $(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\xi}_1) = (\widehat{\boldsymbol{\beta}}_{B^*-1},\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\xi}}_1)$, where

$$(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\xi}}_1) = \arg\max_{\boldsymbol{\theta},\boldsymbol{\xi}_1}\left[-\sum_{i=1}^{n_{B^*}} \log\{1 + \boldsymbol{\xi}_1^\intercal \widehat{g}_i(\boldsymbol{\theta})\}\right].$$

Therefore, we have

$$R_2 = 2\left\{\ell(\widehat{\boldsymbol{\beta}}_{B^*},\widehat{\boldsymbol{\theta}}_{B^*},\widehat{\boldsymbol{\xi}}_1,\widehat{\boldsymbol{\xi}}_2) - \ell(\widehat{\boldsymbol{\beta}}_{B^*-1},\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\xi}}_1,\mathbf{0})\right\}$$

Let $\widehat{\boldsymbol{\mu}} = (\widehat{\boldsymbol{\beta}}_{B^*}, \widehat{\boldsymbol{\theta}}_{B^*}, \widehat{\boldsymbol{\xi}}_1)$ and $\widetilde{\boldsymbol{\mu}} = (\widehat{\boldsymbol{\beta}}_{B^*-1}, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\xi}}_1)$. Moreover, let

$$J_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & -V^{-1} \\ \mathbf{0} & \kappa_{B^*}^{-1}\Sigma_{\boldsymbol{X}}^{-1} & \mathbf{0} \\ -V^{-1} & \mathbf{0} & -V^{-1} \end{pmatrix}, \quad J_2 = \begin{pmatrix} \mathbf{0} \\ \Omega_{\boldsymbol{X}}^{\intercal} \\ -\Omega \end{pmatrix}$$

Applying a Taylor expansion of $\ell(\widehat{\boldsymbol{\beta}}_{B^*-1}, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\xi}}_1, \mathbf{0})$ at $(\widehat{\boldsymbol{\beta}}_{B^*}, \widehat{\boldsymbol{\theta}}_{B^*}, \widehat{\boldsymbol{\xi}}_1, \widehat{\boldsymbol{\xi}}_2)$ yields

$$\begin{aligned} R_2 &= n_{B^*} \begin{pmatrix} \widehat{\boldsymbol{\mu}}^{\intercal} - \widetilde{\boldsymbol{\mu}}^{\intercal} & \widehat{\boldsymbol{\xi}}_2^{\intercal} \end{pmatrix} \begin{pmatrix} J_1 & J_2 \\ -J_2^{\intercal} & Q \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\mu}} - \widetilde{\boldsymbol{\mu}} \\ \widehat{\boldsymbol{\xi}}_2 \end{pmatrix} + o_p(1) \\ &= n_{B^*}(\widehat{\boldsymbol{\mu}} - \widetilde{\boldsymbol{\mu}})^{\intercal} J_1 (\widehat{\boldsymbol{\mu}} - \widetilde{\boldsymbol{\mu}}) + n_{B^*} \widehat{\boldsymbol{\xi}}_2^{\intercal} Q \widehat{\boldsymbol{\xi}}_2 + o_p(1). \end{aligned}$$

Similar to the proof of Theorem 2, we can derive $\sqrt{n_{B^*}} J_1(\widehat{\boldsymbol{\mu}} - \widetilde{\boldsymbol{\mu}}) + \sqrt{n_{B^*}} J_2 \widehat{\boldsymbol{\xi}}_2 = o_p(1)$, and thus we have

$$R_2 = n_{B^*} \widehat{\boldsymbol{\xi}}_2^{\intercal} (J_2^{\intercal} J_1^{-1} J_2 + Q) \widehat{\boldsymbol{\xi}}_2 + o_p(1).$$

Similar to the proof of Theorem 2, we can show that $\sqrt{n_{B^*}} \widehat{\boldsymbol{\xi}}_2$ converges in distribution to a zero mean multivariate normal distribution. Moreover, we can show that $\operatorname{rank}(J_2^{\intercal} J_1^{-1} J_2 + Q) = d$. Therefore, the hybrid empirical likelihood ratio test statistic R_2 converges in distribution to a χ^2 random variable with d degrees of freedom as $n_{B^*} \to \infty$.

S9 Details of computation

Let $\boldsymbol{\gamma} = (\boldsymbol{\theta}^\intercal, \boldsymbol{\beta}^\intercal)^\intercal$, $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^\intercal, \boldsymbol{\xi}_2^\intercal)^\intercal$, and let

$$\ell(\boldsymbol{\gamma}, \boldsymbol{\xi}) = -\sum_{i=1}^{n_{B^*}} \log\left\{1 + \boldsymbol{\xi}_1^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})\right\} - \frac{n_{B^*-1}^c}{2}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta})^\intercal \widehat{\Sigma}_{B^*-1}^{-1}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta}).$$

Some straightforward calculations yield $\partial \ell(\boldsymbol{\gamma}, \boldsymbol{\xi})/\partial \boldsymbol{\xi} = (\mathcal{U}_{\boldsymbol{\xi}_1}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^\intercal, \mathcal{U}_{\boldsymbol{\xi}_2}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^\intercal)^\intercal$,

where

$$\mathcal{U}_{\boldsymbol{\xi}_1}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = -\sum_{i=1}^{n_{B^*}} \frac{\widehat{g}_i(\boldsymbol{\theta})}{1 + \boldsymbol{\xi}_1^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})}, \quad \mathcal{U}_{\boldsymbol{\xi}_2}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = -\sum_{i=1}^{n_{B^*}} \frac{\widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})}{1 + \boldsymbol{\xi}_1^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})}.$$

Moreover, we can derive

$$\frac{\partial^2 \ell(\boldsymbol{\gamma}, \boldsymbol{\xi})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\intercal} = \begin{pmatrix} \mathcal{J}_{\boldsymbol{\xi}_1}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) & \mathcal{J}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_2}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \\ \mathcal{J}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_2}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^\intercal & \mathcal{J}_{\boldsymbol{\xi}_2}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \end{pmatrix},$$

where

$$\mathcal{J}_{\boldsymbol{\xi}_1}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \sum_{i=1}^{n_{B^*}} \frac{\widehat{g}_i(\boldsymbol{\theta})\widehat{g}_i(\boldsymbol{\theta})^\intercal}{\{1 + \boldsymbol{\xi}_1^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})\}^2},$$

$$\mathcal{J}_{\boldsymbol{\xi}_1\boldsymbol{\xi}_2}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \sum_{i=1}^{n_{B^*}} \frac{\widehat{g}_i(\boldsymbol{\theta})\widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})^\intercal}{\{1 + \boldsymbol{\xi}_1^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})\}^2},$$

$$\mathcal{J}_{\boldsymbol{\xi}_2}(\boldsymbol{\gamma}, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \sum_{i=1}^{n_{B^*}} \frac{\widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})\widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})^\intercal}{\{1 + \boldsymbol{\xi}_1^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})\}^2}.$$

Under the given conditions in Theorem 3, $\partial^2 \ell(\boldsymbol{\gamma}, \boldsymbol{\xi})/\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^\intercal$ is positive-definite and thus $\ell(\boldsymbol{\gamma}, \boldsymbol{\xi})$ is a strictly convex function of $\boldsymbol{\xi}$. It remains to prove that asymptotically, $\ell(\boldsymbol{\gamma})$ is a concave function of $\boldsymbol{\gamma}$, where $\ell(\boldsymbol{\gamma}) = \min_{\boldsymbol{\xi}} \ell(\boldsymbol{\gamma}, \boldsymbol{\xi})$. Let $\boldsymbol{\xi}(\boldsymbol{\gamma}) = \arg\min_{\boldsymbol{\xi}} \ell(\boldsymbol{\gamma}, \boldsymbol{\xi})$ and rewrite $\ell(\boldsymbol{\gamma}) = \min_{\boldsymbol{\xi}} \ell(\boldsymbol{\gamma}, \boldsymbol{\xi}) =$

$\ell(\boldsymbol{\gamma}, \boldsymbol{\xi}(\boldsymbol{\gamma}))$, where

$$\ell(\boldsymbol{\gamma}, \boldsymbol{\xi}(\boldsymbol{\gamma})) = -\sum_{i=1}^{n_{B^*}} \log\left\{1 + \boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})\right\} - \frac{n^c_{B^*-1}}{2}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta})^\intercal \widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta}).$$

Hence we have $\partial\ell(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma} = \partial\ell(\boldsymbol{\gamma}, \boldsymbol{\xi}(\boldsymbol{\gamma}))/\partial\boldsymbol{\gamma} = \begin{pmatrix} \mathcal{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}(\boldsymbol{\gamma}))^\intercal & \mathcal{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}(\boldsymbol{\gamma}))^\intercal \end{pmatrix}^\intercal$,

where

$$\mathcal{U}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}(\boldsymbol{\gamma})) = -\sum_{i=1}^{n_{B^*}} \frac{\boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal \partial\widehat{g}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}}{1 + \boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})},$$

$$\mathcal{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}(\boldsymbol{\gamma})) = -\sum_{i=1}^{n_{B^*}} \frac{\boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal \partial\widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})/\partial\boldsymbol{\beta}}{1 + \boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})} + n^c_{B^*-1}\widehat{\Sigma}^{-1}_{B^*-1}(\widehat{\boldsymbol{\beta}}_{B^*-1} - \boldsymbol{\beta}).$$

Applying similar techniques in Han and Lawless (2019), we can derive that $\partial^2\ell(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^\intercal = \mathcal{J}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}, \boldsymbol{\xi}(\boldsymbol{\gamma})) + o_p(1)$, where

$$\begin{aligned}
&\mathcal{J}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}, \boldsymbol{\xi}(\boldsymbol{\gamma})) \\
&= -\left\{\sum_{i=1}^{n_{B^*}} \frac{\partial U_i(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}}{1 + \boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})}\right\} \left[\sum_{i=1}^{n_{B^*}} \frac{U_i(\boldsymbol{\gamma})U_i(\boldsymbol{\gamma})^\intercal}{\{1 + \boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})\}^2}\right]^{-1} \\
&\times \left\{\sum_{i=1}^{n_{B^*}} \frac{\partial U_i(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}}{1 + \boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal \widehat{g}_i(\boldsymbol{\theta}) + \boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})}\right\} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & n^c_{B^*-1}\widehat{\Sigma}^{-1}_{B^*-1} \end{pmatrix},
\end{aligned}$$

and $U_i(\boldsymbol{\gamma}) = (\widehat{g}_i(\boldsymbol{\theta})^\intercal, \widehat{h}_i(\boldsymbol{\theta}, \boldsymbol{\beta})^\intercal)^\intercal$. Since $\mathcal{J}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}, \boldsymbol{\xi}(\boldsymbol{\gamma}))$ is a negative definite matrix, $\ell(\boldsymbol{\gamma})$ is an asymptotically concave function of $\boldsymbol{\gamma}$. Hence we consider updating $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ sequentially. Along the same lines in Han and Lawless (2019), we can derive $\partial^2\ell(\boldsymbol{\gamma})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\intercal = \mathcal{J}_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\xi}(\boldsymbol{\gamma})) + o_p(1)$ and

$\partial^2\ell(\boldsymbol{\gamma})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\intercal = \mathcal{J}_{\boldsymbol{\beta}}(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\xi}(\boldsymbol{\gamma})) + o_p(1)$, where

$$\begin{aligned}
&\mathcal{J}_{\boldsymbol{\theta}}(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\xi}(\boldsymbol{\gamma})) \\
&= -\left\{\sum_{i=1}^{n_{B^*}} \frac{\partial\widehat{g}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}}{1+\boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})}\right\}\left[\sum_{i=1}^{n_{B^*}} \frac{\widehat{g}_i(\boldsymbol{\theta})\widehat{g}_i(\boldsymbol{\theta})^\intercal}{\{1+\boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})\}^2}\right]^{-1} \\
&\times\left\{\sum_{i=1}^{n_{B^*}} \frac{\partial\widehat{g}_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}}{1+\boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})}\right\},
\end{aligned}$$

$$\begin{aligned}
&\mathcal{J}_{\boldsymbol{\beta}}(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\xi}(\boldsymbol{\gamma})) \\
&= -\left\{\sum_{i=1}^{n_{B^*}} \frac{\partial\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})/\partial\boldsymbol{\beta}}{1+\boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})}\right\}\left[\sum_{i=1}^{n_{B^*}} \frac{\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})^\intercal}{\{1+\boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})\}^2}\right]^{-1} \\
&\times\left\{\sum_{i=1}^{n_{B^*}} \frac{\partial\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})/\partial\boldsymbol{\beta}}{1+\boldsymbol{\xi}_1(\boldsymbol{\gamma})^\intercal\widehat{g}_i(\boldsymbol{\theta})+\boldsymbol{\xi}_2(\boldsymbol{\gamma})^\intercal\widehat{h}_i(\boldsymbol{\theta},\boldsymbol{\beta})}\right\} - n^c_{B^*-1}\widehat{\Sigma}^{-1}_{B^*-1}.
\end{aligned}$$

References

Han, P. and J. F. Lawless (2019). Empirical likelihood estimation using auxiliary summary information with different covariate distributions. *Statistica Sinica 29*, 1321–1342.

Huang, C.-Y. and J. Qin (2020). A unified approach for synthesizing population-level covariate effect information in semiparametric estimation with survival data. *Statistics in Medicine 39*(10), 1573–1590.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*(1), 90–120.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*(1), 300–325.

Zhang, H., L. Deng, M. Schiffman, J. Qin, and K. Yu (2020). Generalized integration model for improved statistical inference by leveraging external summary data. *Biometrika 107*, 689–703.