

# Supplementary Materials: Additional Simulations

for

## Modeling Time-varying Effects with Large-scale Survival Data: An Efficient Quasi-Newton Approach

**Table A.1**

Performance of unstratified and quasi-Newton methods (Simulation Setting 1; N=5,000).

quasi-NR	Bias	MSE	CP	Time	Iteration	Power	Type-I Error
Unstratified	0.07	0.03	0.93	61.67 Seconds	17.13	1	0.06
Stratified	0.07	0.03	0.93	7.81 Seconds	17.61	1	0.06

**Table A.2**

Average Bias of Cumulative Hazard Functions (Simulation Setting 2; N=1000).

quasi-NR	Bias of $\Lambda_0(0.5)$	Bias of $\Lambda_0(1)$	Bias of $\Lambda_0(1.5)$
Unstratified	0.38	0.58	0.69
Stratified	0.08	0.15	0.20

### Number and Knots for B-spline Functions

Here we present simulation results with various sample sizes and various numbers of knots for B-spline basis functions. We first compares the quasi-Newton method based on 6, 10 and 20 knots for B-spline basis functions. As shown in Figure A.1, when the sample size is relatively small (e.g., N=1,000), as the number of knots is increased, the estimated time-varying effects are more unstable. The averages of the estimates for results based on 10 knots and 20 knots overlay the true curves. In contrast, when the sample size is large

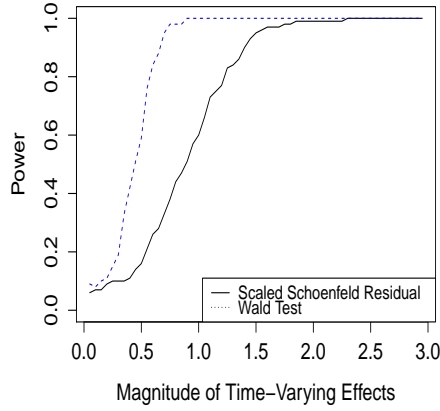
**Table A.3**

Average P-values (P), Empirical Power (proportion tests correctly identify the true time-varying effects); Type-I Error (proportion tests falsely identify the true time-varying effects) for quasi-Newton based test in Figure 3; True effects:  $\beta_1(t) = \alpha \sin(3\pi t/4)$ ,  $\beta_2(t) = 1$  and  $\beta_3(t) = -1$ ; sample size: N=1,000; number of centers: J=5 (200 subjects in each center); all centers have the same baseline hazard functions: exponential distribution with  $\lambda = 1$ ; 500 replications.

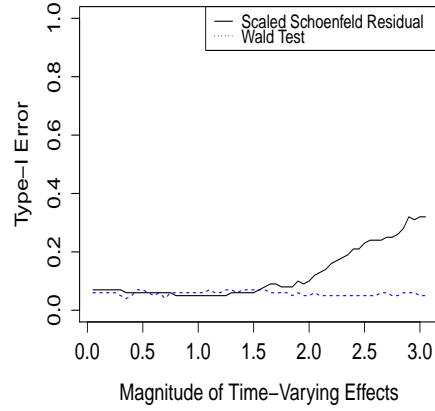
Magnitude ( $\alpha$ )	P of $\beta_1$	Power of $\beta_1$	P of $\beta_2$	Type-I Error of $\beta_2$	P of $\beta_3$	Type-I Error of $\beta_3$
0	0.47	0.06	0.51	0.06	0.50	0.02
0.10	0.35	0.19	0.51	0.09	0.50	0.02
0.20	0.14	0.50	0.51	0.07	0.49	0.04
0.30	0.02	0.93	0.50	0.06	0.49	0.05
0.40	0.001	1.00	0.51	0.07	0.50	0.06
0.50	< 0.001	1.00	0.51	0.07	0.49	0.04
0.60	< 0.001	1.00	0.52	0.06	0.50	0.04
0.70	< 0.001	1.00	0.52	0.04	0.53	0.04
0.80	< 0.001	1.00	0.52	0.07	0.52	0.05
0.90	< 0.001	1.00	0.51	0.06	0.53	0.09
1.00	< 0.001	1.00	0.50	0.07	0.50	0.07
1.10	< 0.001	1.00	0.50	0.07	0.50	0.05
1.20	< 0.001	1.00	0.51	0.06	0.50	0.03
1.30	< 0.001	1.00	0.51	0.06	0.50	0.05
1.40	< 0.001	1.00	0.53	0.06	0.48	0.04
1.50	< 0.001	1.00	0.51	0.07	0.47	0.05
1.60	< 0.001	1.00	0.52	0.05	0.49	0.06
1.70	< 0.001	1.00	0.51	0.07	0.48	0.05
1.80	< 0.001	1.00	0.53	0.05	0.48	0.04
1.90	< 0.001	1.00	0.51	0.08	0.46	0.03
2.00	< 0.001	1.00	0.52	0.09	0.47	0.06
2.10	< 0.001	1.00	0.51	0.06	0.47	0.05
2.20	< 0.001	1.00	0.52	0.07	0.46	0.05
2.30	< 0.001	1.00	0.52	0.06	0.48	0.06
2.40	< 0.001	1.00	0.51	0.05	0.50	0.05
2.50	< 0.001	1.00	0.51	0.04	0.51	0.06
2.60	< 0.001	1.00	0.51	0.05	0.51	0.06
2.70	< 0.001	1.00	0.51	0.08	0.49	0.06
2.80	< 0.001	1.00	0.49	0.05	0.49	0.04
2.90	< 0.001	1.00	0.48	0.08	0.47	0.06
3.00	< 0.001	1.00	0.49	0.07	0.48	0.05

**Figure A.1**

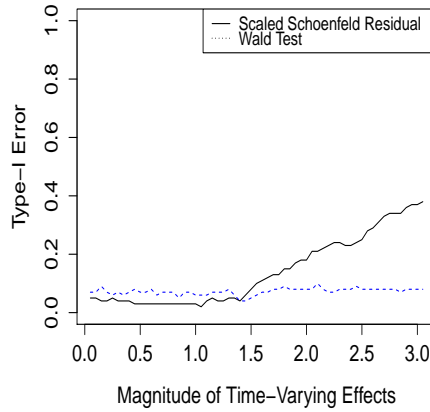
Power: average proportion that the test rejected the null hypothesis for variables with time-varying effects; Type-I error: average proportion that the test rejected the null hypothesis for variables with time-independent effects; True effects: to mimic a monotonically decreasing time-varying effect, we choose  $\beta_1(t) = -\alpha(t/3)^{10} \exp(t/4)$  where the magnitude of time-varying effects,  $\alpha$ , varies from 0 to 3,  $\beta_2(t) = 1$  and  $\beta_3(t) = -1$ ; The tests for time-varying effect are described in Section 2.4; Type-I errors for quasi-Newton methods in Figures 3 (b) and (c) are around 0.05..



(a) Power for  $\beta_1$ : time-varying effects



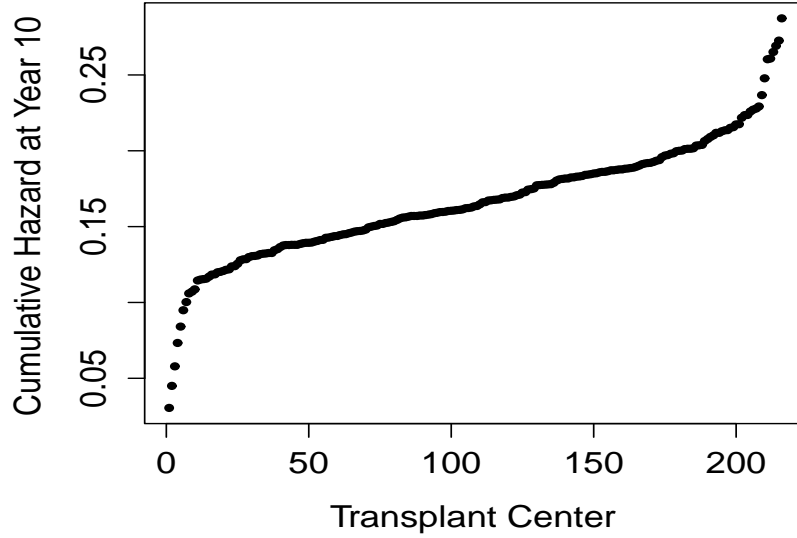
(b) Type-I Error for  $\beta_2$ : time-independent effects



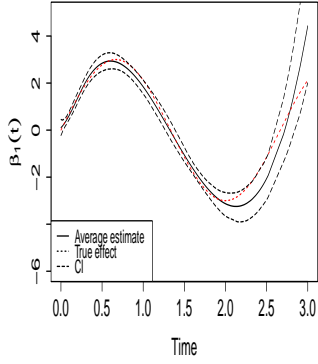
(c) Type-I Error for  $\beta_3$ : time-independent effects

**Figure A.2**

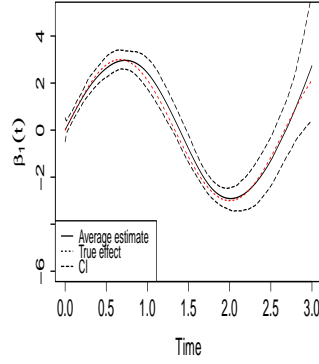
Estimated Cumulative Baseline Hazards at Year 10 for 216 centers in Section 5.



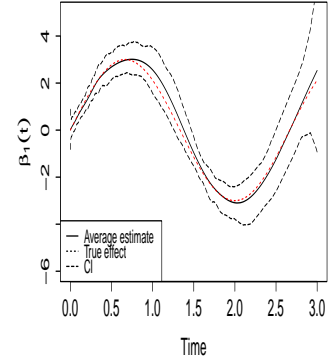
(e.g.,  $N=10,000$  in Figure A.2), the number of knots seems to have little effect on the results provided that it is not too small. Figure A.3 shows that the approach in which the knots are chosen to be equally spaced to cover the time-span tends to be unstable in the right tail of the follow-up period. In contrast, the alternative approach, for which the knots are chosen to include an equal number of events within each interval, offers a more stable estimation.



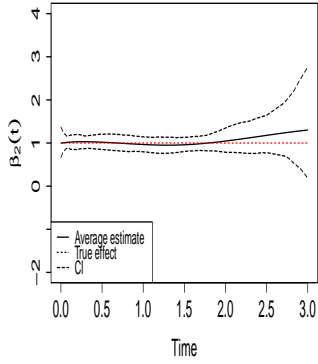
(a) Number of knots: K=6



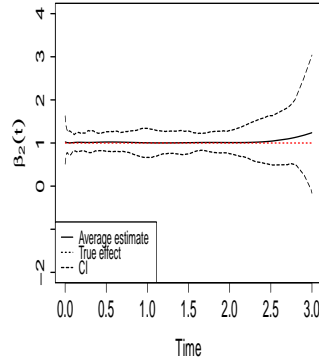
(b) Number of knots: K=10



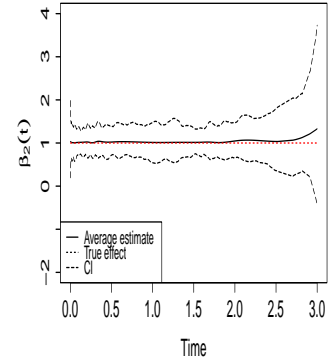
(c) 20number of knots: K=20



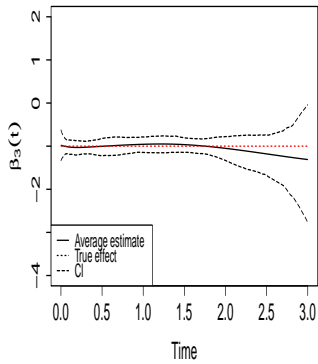
(d) Number of knots: K=6



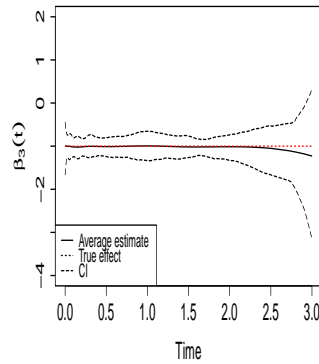
(e) Number of knots: K=10



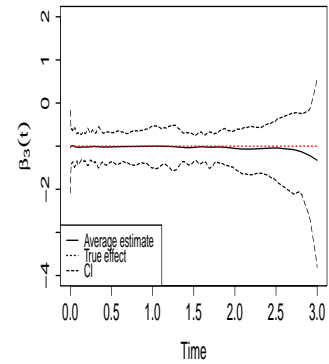
(f) Number of knots: K=20



(g) Number of knots: K=6

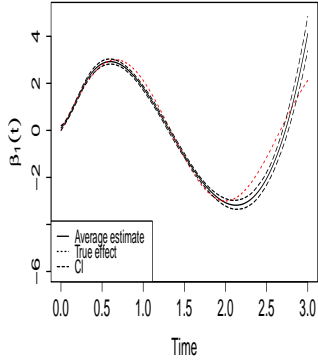


(h) Number of knots: K=10

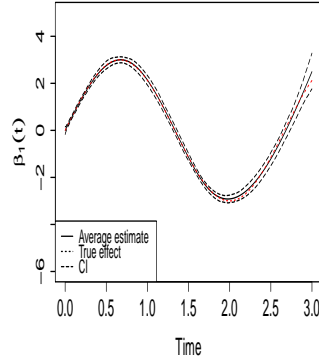


(i) Number of knots: K=20

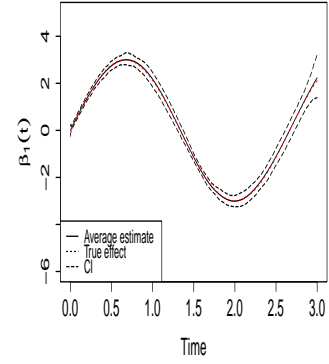
**Figure A.3** Estimated coefficients for settings with small sample size  $N=1,000$ ; True effects:  $\beta_1(t) = 3\sin(3\pi t/4)$ ,  $\beta_2(t) = 1$  and  $\beta_3(t) = -1$ ; the one-step Newton method is described in Section 2.4; the quasi-Newton method is described in Sections 2.2 and 2.3.



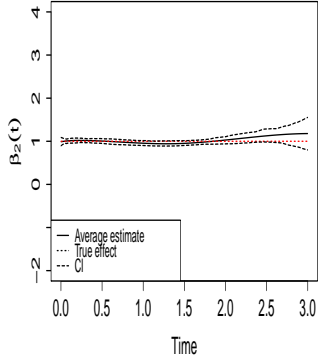
(a) Number of knots: K=6



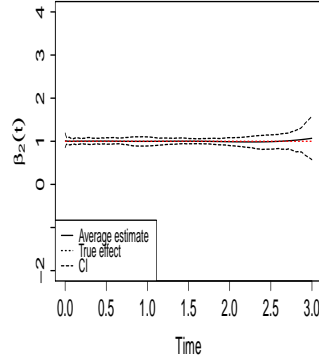
(b) Number of knots: K=10



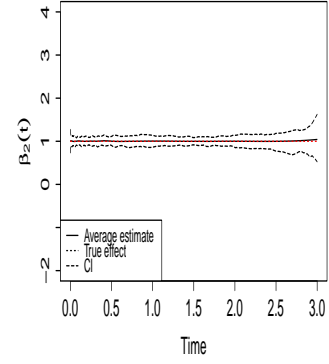
(c) Number of knots: K=20



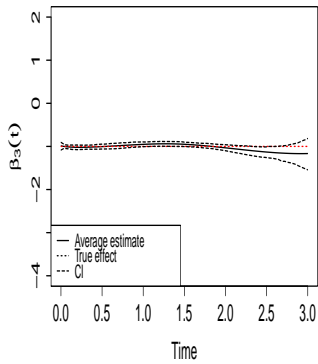
(d) Number of knots: K=6



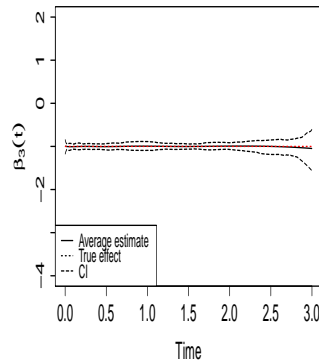
(e) Number of knots: K=10



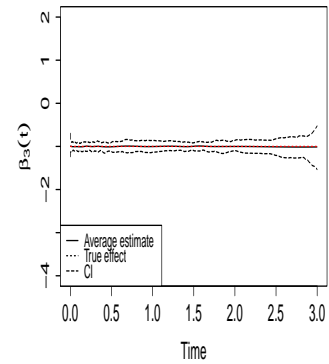
(f) Number of knots: K=20



(g) Number of knots: K=6

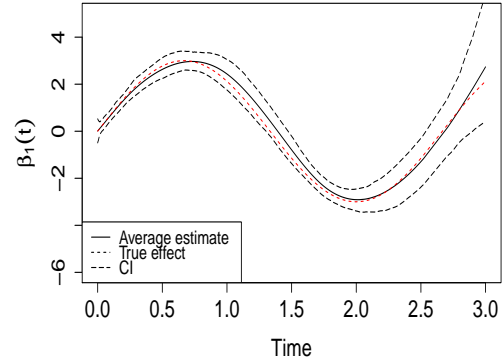
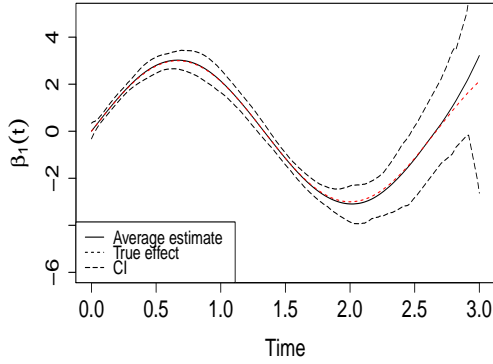


(h) Number of knots: K=10

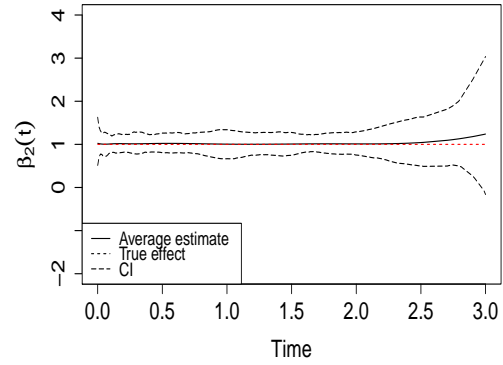
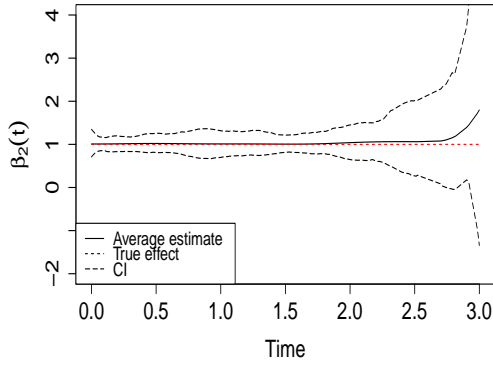


(i) Number of knots: K=20

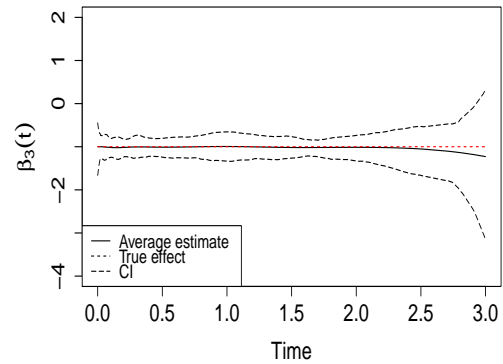
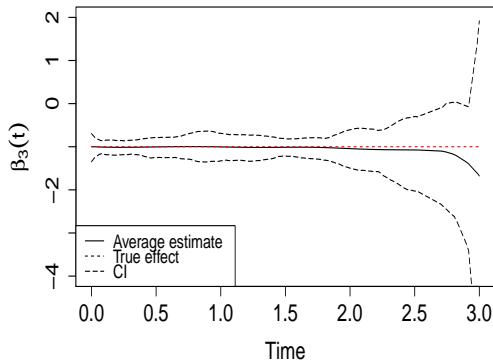
**Figure A.4** Estimated coefficients for settings with large sample size  $N=10,000$ ; True effects:  $\beta_1(t) = 3 \sin(3\pi t/4)$ ,  $\beta_2(t) = 1$  and  $\beta_3(t) = -1$



(a) Knots are chosen to be equally spaced to cover the range of time (b) Knots are chosen to include equal number of events within each interval



(c) Knots are chosen to be equally spaced to cover the range of time (d) Knots are chosen to include equal number of events within each interval



(e) Knots are chosen to be equally spaced to cover the range of time (f) Knots are chosen to include equal number of events within each interval

**Figure A.5** Estimated coefficients for settings with large sample size  $N=10,000$ ; True effects:  $\beta_1(t) = 3 \sin(3\pi t/4)$ ,  $\beta_2(t) = 1$  and  $\beta_3(t) = -1$ ; Number of knots:  $K=10$ .