

# **Supplementary Material for “Estimation and Variable Selection for Interval-censored Failure Time Data with Random Change Point and Application to Breast Cancer Study”**

This Supplementary Material contains four tables and one figure based on the results obtained in the simulation study described in the main paper. The first two tables present some results obtained for assessing the performance of the method proposed in Section 3 of the main paper in terms of estimating the regression parameters, while the figure shows some results obtained for assessing the performance of the same method in terms of estimating the baseline cumulative hazard function. The last two tables provide some results obtained for assessing the effect of the misspecification of the distribution of the change point on the method.

More specifically, Tables S1 and S2 present the results on estimation of the regression parameters given by the method proposed in Section 3 of the main paper based on the simulated data giving Tables 1 and 2 in the main paper. Here we calculated the estimated bias (Bias) given by the average of the point estimates minus the true value, the averages of the estimated standard errors (ESE), the sample standard errors of the estimates (SSE), and the 95% empirical coverage probability (CP) for the non-zero regression parameters. Table S1 is for the PH model, while Table S2 for the PO model. The results indicate

that the proposed estimator seems to be unbiased, the variance estimator appears to be reasonable, and the normal approximation to the distribution of the proposed estimator is appropriate with the use of all penalty functions. As expected, the performance improved when the sample size increased.

Figure S1 displays the average of the estimated baseline cumulative hazard functions given by the method proposed in Section 3 of the main paper based on the data giving Tables 1 and 2 in the main paper. For comparison, the true baseline cumulative hazard function is included for each situation, and it indicates that the proposed method can estimate the baseline cumulative hazard function well.

In the proposed approach, we have assumed that the distribution of the change point is known up to a vector of parameters. Thus one question of interest is the effect of the misspecification of the distribution on the proposed procedure. To see this, we repeated the study giving Tables 3 and 4 with  $n = 500$  in the main paper but assuming the normal distribution for the change point. The obtained results are presented in Tables S3 and S4, and it is apparent that they are similar to those given in Tables 3 and 4 in the main paper and indicate that the method works well. In other words, the proposed method seems to be robust with respect to the distribution of the change point.

Table S1: Simulation results on the parameter estimation under PH model with  $\eta \sim N(0, 1)$ .

Method	Parameters	$n = 300$				$n = 500$			
		Bias	ESE	SSE	CP	Bias	ESE	SSE	CP
Oracle	$\beta_1 = 0.7$	0.0239	0.1063	0.1006	0.923	0.0095	0.0744	0.0762	0.953
	$\beta_2 = 0.7$	0.0273	0.1264	0.1105	0.907	0.0138	0.0880	0.0835	0.943
	$\beta_3 = 0.7$	0.0323	0.1059	0.1017	0.947	0.0085	0.0776	0.0763	0.950
	$\alpha_1 = -0.3$	-0.0114	0.0596	0.0589	0.943	-0.0088	0.0468	0.0448	0.950
	$\alpha_2 = 1.2$	0.0474	0.2233	0.2025	0.920	0.0233	0.1450	0.1522	0.950
	$\alpha_3 = 0.6$	0.0146	0.1053	0.0988	0.950	0.0039	0.0754	0.0744	0.947
	$\gamma_1 = 1.4$	0.0605	0.1463	0.1352	0.913	0.0276	0.1057	0.1008	0.940
	$\gamma_2 = -0.5$	-0.0105	0.2524	0.2321	0.940	-0.0124	0.1805	0.1747	0.940
	$\gamma_3 = 0.2$	0.0178	0.1611	0.1483	0.927	0.0131	0.1196	0.1123	0.920
	$\zeta_1 = 0$	-0.0133	0.1561	0.1527	0.940	0.0048	0.1292	0.1191	0.937
	$\zeta_2 = 1$	-0.0079	0.1795	0.1760	0.970	0.0025	0.1447	0.1355	0.963
ALASSO	$\beta_1 = 0.7$	0.0011	0.1121	0.1027	0.920	-0.0043	0.0764	0.0769	0.960
	$\beta_2 = 0.7$	0.0108	0.1322	0.1126	0.903	0.0031	0.0895	0.0844	0.927
	$\beta_3 = 0.7$	0.0093	0.1140	0.1126	0.950	-0.0051	0.0792	0.0837	0.957
	$\alpha_1 = -0.3$	-0.0046	0.0596	0.0611	0.943	-0.0041	0.0468	0.0457	0.957
	$\alpha_2 = 1.2$	0.0234	0.2231	0.2073	0.923	0.0098	0.1442	0.1542	0.967
	$\alpha_3 = 0.6$	0.0042	0.1056	0.1019	0.947	-0.0023	0.0741	0.0754	0.953
	$\gamma_1 = 1.4$	0.0304	0.1510	0.1364	0.910	0.0087	0.1065	0.1016	0.947
	$\gamma_2 = -0.5$	-0.0022	0.2475	0.2381	0.947	-0.0068	0.1787	0.1771	0.940
	$\gamma_3 = 0.2$	0.0130	0.1577	0.1534	0.950	0.0091	0.1185	0.1142	0.940
	$\zeta_1 = 0$	-0.0109	0.1561	0.1557	0.947	0.0067	0.1295	0.1204	0.937
	$\zeta_2 = 1$	-0.0185	0.1826	0.1805	0.967	-0.0052	0.1435	0.1376	0.970
SELO	$\beta_1 = 0.7$	0.0286	0.1083	0.1041	0.930	0.0103	0.0752	0.0775	0.953
	$\beta_2 = 0.7$	0.0333	0.1293	0.1144	0.917	0.0147	0.0884	0.0851	0.943
	$\beta_3 = 0.7$	0.0367	0.1135	0.1148	0.943	0.0098	0.0793	0.0843	0.963
	$\alpha_1 = -0.3$	-0.0143	0.0604	0.0631	0.943	-0.0093	0.0473	0.0459	0.950
	$\alpha_2 = 1.2$	0.0542	0.2280	0.2161	0.930	0.0259	0.1474	0.1550	0.953
	$\alpha_3 = 0.6$	0.0195	0.1068	0.1050	0.947	0.0053	0.0752	0.0759	0.950
	$\gamma_1 = 1.4$	0.0702	0.1489	0.1431	0.920	0.0297	0.1073	0.1030	0.940
	$\gamma_2 = -0.5$	-0.0182	0.2536	0.2484	0.950	-0.0139	0.1813	0.1782	0.943
	$\gamma_3 = 0.2$	0.0184	0.1613	0.1572	0.943	0.0121	0.1193	0.1146	0.937
	$\zeta_1 = 0$	-0.0149	0.1565	0.1570	0.947	0.0047	0.1296	0.1203	0.937
	$\zeta_2 = 1$	-0.0072	0.1836	0.1827	0.960	0.0019	0.1442	0.1373	0.970
SICA	$\beta_1 = 0.7$	0.0264	0.1077	0.1038	0.933	0.0101	0.0753	0.0776	0.953
	$\beta_2 = 0.7$	0.0308	0.1294	0.1144	0.917	0.0144	0.0882	0.0851	0.943
	$\beta_3 = 0.7$	0.0357	0.1134	0.1155	0.943	0.0095	0.0788	0.0843	0.963
	$\alpha_1 = -0.3$	-0.0129	0.0603	0.0633	0.947	-0.0092	0.0473	0.0460	0.950
	$\alpha_2 = 1.2$	0.0515	0.2281	0.2164	0.933	0.0256	0.1467	0.1552	0.957
	$\alpha_3 = 0.6$	0.0176	0.1069	0.1050	0.950	0.0053	0.0751	0.0759	0.950
	$\gamma_1 = 1.4$	0.0665	0.1495	0.1437	0.923	0.0290	0.1067	0.1033	0.947
	$\gamma_2 = -0.5$	-0.0150	0.2532	0.2485	0.950	-0.0139	0.1811	0.1785	0.943
	$\gamma_3 = 0.2$	0.0190	0.1616	0.1575	0.943	0.0120	0.1194	0.1147	0.937
	$\zeta_1 = 0$	-0.0142	0.1562	0.1569	0.947	0.0045	0.1296	0.1204	0.937
	$\zeta_2 = 1$	-0.0079	0.1814	0.1829	0.967	0.0002	0.1442	0.1374	0.970
BAR	$\beta_1 = 0.7$	0.0015	0.1080	0.1028	0.937	-0.0073	0.0750	0.0772	0.940
	$\beta_2 = 0.7$	0.0099	0.1286	0.1132	0.910	0.0006	0.0886	0.0847	0.943
	$\beta_3 = 0.7$	0.0108	0.1084	0.1140	0.967	-0.0081	0.0781	0.0843	0.960
	$\alpha_1 = -0.3$	-0.0039	0.0579	0.0614	0.947	-0.0029	0.0461	0.0465	0.953
	$\alpha_2 = 1.2$	0.0234	0.2218	0.2082	0.927	0.0061	0.1437	0.1556	0.960
	$\alpha_3 = 0.6$	0.0035	0.1030	0.1017	0.957	-0.0043	0.0740	0.0755	0.963
	$\gamma_1 = 1.4$	0.0291	0.1429	0.1386	0.940	0.0040	0.1041	0.1028	0.957
	$\gamma_2 = -0.5$	-0.0007	0.2458	0.2397	0.953	-0.0049	0.1780	0.1798	0.943
	$\gamma_3 = 0.2$	0.0132	0.1562	0.1544	0.950	0.0091	0.1180	0.1144	0.937
	$\zeta_1 = 0$	-0.0115	0.1556	0.1572	0.947	0.0072	0.1291	0.1217	0.933
	$\zeta_2 = 1$	-0.0187	0.1784	0.1823	0.970	-0.0059	0.1436	0.1396	0.963

Table S2: Simulation results on the parameter estimation under PO model with  $\eta \sim N(0, 1)$ .

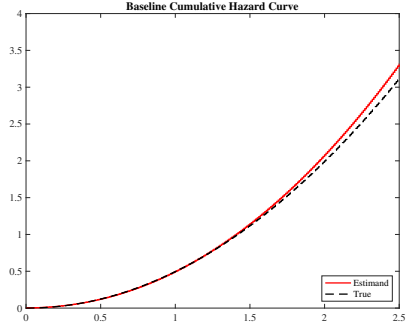
Method	Parameters	$n = 300$				$n = 500$			
		Bias	ESE	SSE	CP	Bias	ESE	SSE	CP
Oracle	$\beta_1 = 0.7$	0.0310	0.1476	0.1486	0.937	0.0089	0.1155	0.1134	0.950
	$\beta_2 = 0.7$	0.0180	0.1629	0.1642	0.960	0.0143	0.1377	0.1255	0.940
	$\beta_3 = 0.7$	0.0066	0.1434	0.1482	0.957	0.0046	0.1172	0.1130	0.947
	$\alpha_1 = -0.3$	-0.0078	0.1082	0.1010	0.937	-0.0066	0.0756	0.0762	0.950
	$\alpha_2 = 1.2$	0.0672	0.3133	0.3194	0.957	0.0120	0.2605	0.2427	0.943
	$\alpha_3 = 0.6$	0.0192	0.1566	0.1607	0.957	0.0103	0.1300	0.1221	0.937
	$\gamma_1 = 1.4$	0.0423	0.1886	0.1899	0.957	0.0157	0.1457	0.1430	0.967
	$\gamma_2 = -0.5$	-0.0178	0.3773	0.3988	0.960	0.0190	0.3153	0.3048	0.940
	$\gamma_3 = 0.2$	-0.0005	0.2393	0.2573	0.970	0.0028	0.2073	0.1945	0.927
	$\zeta_1 = 0$	0.0012	0.2372	0.2125	0.910	0.0071	0.1523	0.1662	0.957
	$\zeta_2 = 1$	-0.0573	0.2439	0.2687	0.973	-0.0142	0.1946	0.1921	0.953
ALASSO	$\beta_1 = 0.7$	0.0034	0.1642	0.1495	0.930	-0.0120	0.1233	0.1135	0.930
	$\beta_2 = 0.7$	-0.0050	0.1945	0.1658	0.947	0.0011	0.1449	0.1259	0.923
	$\beta_3 = 0.7$	-0.0216	0.1660	0.1652	0.950	-0.0160	0.1256	0.1251	0.943
	$\alpha_1 = -0.3$	0.0024	0.1125	0.1026	0.917	-0.0011	0.0752	0.0771	0.957
	$\alpha_2 = 1.2$	0.0461	0.3114	0.3216	0.957	-0.0015	0.2564	0.2436	0.940
	$\alpha_3 = 0.6$	0.0128	0.1574	0.1619	0.950	0.0037	0.1279	0.1228	0.950
	$\gamma_1 = 1.4$	0.0117	0.1992	0.1912	0.920	-0.0052	0.1462	0.1437	0.967
	$\gamma_2 = -0.5$	-0.0090	0.3712	0.4030	0.963	0.0248	0.3109	0.3065	0.953
	$\gamma_3 = 0.2$	-0.0064	0.2364	0.2610	0.983	0.0000	0.2043	0.1955	0.927
	$\zeta_1 = 0$	0.0088	0.2364	0.2186	0.910	0.0069	0.1525	0.1688	0.957
	$\zeta_2 = 1$	-0.0875	0.2616	0.2903	0.987	-0.0265	0.1921	0.1957	0.973
SELO	$\beta_1 = 0.7$	0.0329	0.1557	0.1511	0.937	0.0097	0.1179	0.1145	0.953
	$\beta_2 = 0.7$	0.0206	0.1675	0.1669	0.957	0.0164	0.1396	0.1268	0.937
	$\beta_3 = 0.7$	0.0097	0.1632	0.1670	0.967	0.0068	0.1225	0.1262	0.947
	$\alpha_1 = -0.3$	-0.0084	0.1144	0.1035	0.923	-0.0076	0.0767	0.0775	0.953
	$\alpha_2 = 1.2$	0.0702	0.3183	0.3242	0.953	0.0144	0.2623	0.2455	0.947
	$\alpha_3 = 0.6$	0.0246	0.1612	0.1632	0.943	0.0118	0.1311	0.1237	0.943
	$\gamma_1 = 1.4$	0.0497	0.1984	0.1944	0.943	0.0198	0.1467	0.1450	0.960
	$\gamma_2 = -0.5$	-0.0201	0.3830	0.4066	0.960	0.0194	0.3179	0.3088	0.953
	$\gamma_3 = 0.2$	-0.0015	0.2443	0.2604	0.970	0.0034	0.2081	0.1971	0.930
	$\zeta_1 = 0$	0.0060	0.2432	0.2167	0.903	0.0068	0.1530	0.1681	0.957
	$\zeta_2 = 1$	-0.0649	0.2583	0.2802	0.967	-0.0122	0.1949	0.1930	0.950
SICA	$\beta_1 = 0.7$	0.0312	0.1553	0.1508	0.940	0.0084	0.1180	0.1144	0.957
	$\beta_2 = 0.7$	0.0187	0.1712	0.1670	0.957	0.0149	0.1392	0.1265	0.937
	$\beta_3 = 0.7$	0.0053	0.1632	0.1671	0.967	0.0041	0.1196	0.1260	0.953
	$\alpha_1 = -0.3$	-0.0079	0.1138	0.1028	0.920	-0.0072	0.0762	0.0773	0.953
	$\alpha_2 = 1.2$	0.0688	0.3170	0.3253	0.957	0.0137	0.2610	0.2455	0.943
	$\alpha_3 = 0.6$	0.0230	0.1604	0.1637	0.947	0.0106	0.1305	0.1234	0.943
	$\gamma_1 = 1.4$	0.0461	0.1991	0.1942	0.950	0.0176	0.1459	0.1451	0.957
	$\gamma_2 = -0.5$	-0.0200	0.3821	0.4081	0.960	0.0179	0.3161	0.3092	0.953
	$\gamma_3 = 0.2$	-0.0015	0.2432	0.2628	0.977	0.0032	0.2073	0.1965	0.930
	$\zeta_1 = 0$	0.0048	0.2411	0.2171	0.903	0.0068	0.1526	0.1684	0.960
	$\zeta_2 = 1$	-0.0638	0.2594	0.2828	0.983	-0.0132	0.1950	0.1937	0.957
BAR	$\beta_1 = 0.7$	-0.0087	0.1694	0.1493	0.913	-0.0200	0.1218	0.1131	0.943
	$\beta_2 = 0.7$	-0.0104	0.1927	0.1655	0.920	-0.0058	0.1469	0.1255	0.907
	$\beta_3 = 0.7$	-0.0344	0.1665	0.1685	0.943	-0.0241	0.1229	0.1244	0.943
	$\alpha_1 = -0.3$	0.0049	0.1090	0.1050	0.923	0.0019	0.0739	0.0766	0.957
	$\alpha_2 = 1.2$	0.0358	0.3079	0.3298	0.960	-0.0081	0.2553	0.2425	0.950
	$\alpha_3 = 0.6$	0.0075	0.1543	0.1634	0.950	0.0010	0.1276	0.1224	0.947
	$\gamma_1 = 1.4$	-0.0018	0.1912	0.1907	0.927	-0.0163	0.1432	0.1428	0.970
	$\gamma_2 = -0.5$	-0.0017	0.3661	0.4053	0.970	0.0281	0.3085	0.3054	0.953
	$\gamma_3 = 0.2$	-0.0070	0.2329	0.2642	0.983	-0.0023	0.2024	0.1950	0.927
	$\zeta_1 = 0$	0.0050	0.2372	0.2256	0.910	0.0070	0.1523	0.1688	0.963
	$\zeta_2 = 1$	-0.0889	0.2553	0.2913	0.990	-0.0327	0.1932	0.1971	0.977

Table S3: Simulation results on the variable selection under the misspecified distribution of  $\eta$  with  $n = 500$ .

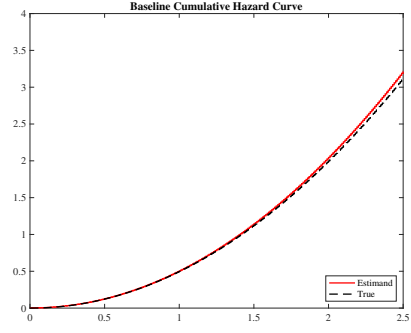
Method	PH Model				PO Model			
	TP	FP	MMSE	SD	TP	FP	MMSE	SD
ALASSO	3.00	0.10	0.013	0.016	3.00	0.18	0.031	0.032
SELO	3.00	0.09	0.011	0.017	3.00	0.31	0.031	0.040
SICA	3.00	0.06	0.011	0.016	3.00	0.16	0.025	0.035
BAR	3.00	0.01	0.012	0.014	3.00	0.04	0.026	0.027
Oracle	3	0	0.010	0.015	3	0	0.222	0.027

Table S4: Simulation results on the proportion at which each covariate is selected (%) under the misspecified distribution of  $\eta$  with  $n = 500$ .

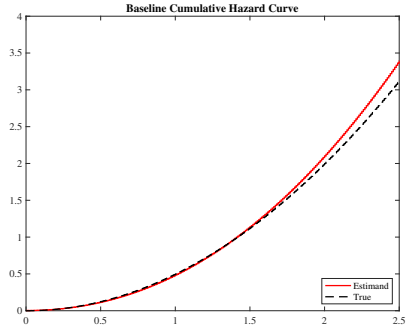
Model	Method	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
PH	ALASSO	100	100	100	1	1	0	1	2	1	1
	SELO	100	100	100	2	1	0	1	1	1	1
	SICA	100	100	100	0	1	0	1	1	1	1
	BAR	100	100	100	0	0	0	0	0	0	0
PO	ALASSO	100	100	100	3	1	2	1	4	2	2
	SELO	100	100	100	6	4	4	3	5	4	4
	SICA	100	100	100	3	1	1	1	3	2	2
	BAR	100	100	100	0	0	0	0	1	0	0



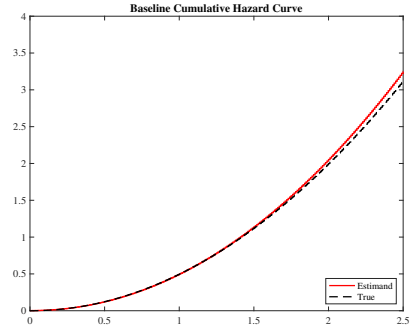
(a) ALASSO under  $n = 300$



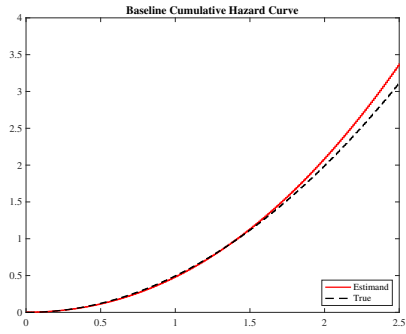
(b) ALASSO under  $n = 500$



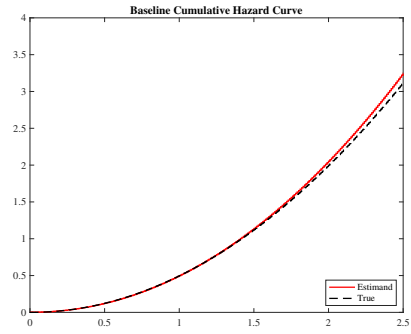
(c) SELO under  $n = 300$



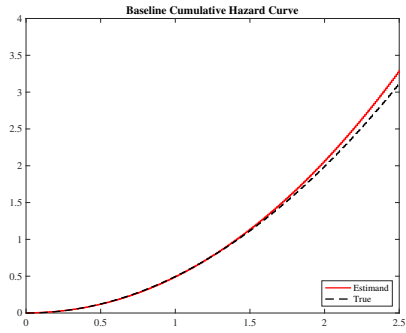
(d) SELO under  $n = 500$



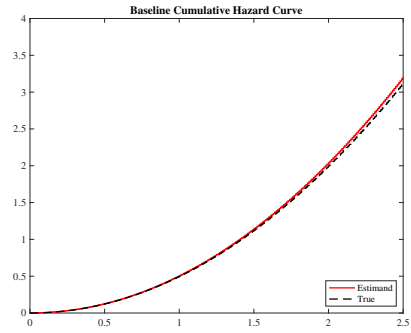
(e) SICA under  $n = 300$



(f) SICA under  $n = 500$



(g) BAR under  $n = 300$



(h) BAR under  $n = 500$

Figure S1: Estimated baseline cumulative hazard curve under PH model and  $\eta \sim N(0, 1)$ .