Results

2020-03-30

Points under investigation

- 1. For Cox PH model, a mis-specification can cause a biased estimation of the survival function
- 2. In Dikta's paper with covariates, the semiparametric estimator with $m(t, x)$ function, can have an unbiased estimation of the function and can be more efficient than the Kaplan-Meier-based estimator proposed by Stute ain 1993.

However,

- 3.1 The Stute's Kaplan-Meier-based estimator and Dikta's covariates estimator are used to estimate the probability of
$$P(T < t, X < x)$$
- 3.2 The Cox PH model is used to esitmate
$$P(T > t|X = x)$$
.
- 3.3 There are few papers that compared the two kinds of estimators: Stute's estimator and Cox PH model

Therefore,

- Dikta's covariates model is better than the Stute's Kaplan-Meier-based model in terms of a lower variance of the estimations.
- We do not know whether Dikta's covariates model can beat Cox PH model since there do not have comparisions in previous studies.
  - to compare them, we need a new estimator with the similar form of the Dikta's esitmator, but estimates the same probability as Cox PH model does.

Therefore, the ideas are:

- 1. Investigate whether the semiparameter estimated with $m()$ function has a smaller variance than Cox PH model.
- 2. Cox PH model can have biasd when it is mis-specified (when ignore some non-significant covariates). The semiparameter model with $m()$ function, by considering all the covariates that are associated with death and censoring, can avoid the Cox PH model's bias.
- 3. Since both Cox PH model and the semiparameter model need to be correctly specified to make a good estimation, whether we can build a rule to combine the two models in practice (double robust).

Simulation

Data set generation

- 1. Simulate the covariates for n subjects: $X = (x1, x2) \sim MVN(\mu, \Sigma)$, where X is n by 2 matrix and $x1, x2$ are n by 1 vectors. For each subject, the covariates is x_{i1}, x_{i2}.
  - $\beta = (\beta_1, \beta_2) = (0.2, 0.1), \gamma = (\gamma_1, \gamma_2) = (0.1, 0.2)$
  - $X = (x1, x2) \sim MVN(\mu, \Sigma)$, $\mu = (1, 2)$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $\rho = 0, 0.2, 0.5, 0.9$.
- 2. For subject i, calculate the βX_i. Simulate the survival time T_i from below CDF by using the inverse probablity sampling.

$$S_{T,i}(t) = S_{0,i}(t)^{\exp(\beta X_i)} = \exp(-t \times \exp(\beta X_i))$$

- 3. Simulate the censoring time C_i from below CDF by using the inverse probablity sampling.

$$S_{C,i}(t) = S_{0,i}(t)^{\exp(\gamma X_i)} = \exp(-t \times \exp(\gamma X_i))$$

- 4. Calculate observed time $Z_i = T_i \wedge C_i$ and indicator $\delta_i = I(T_i < C_i)$.
- Then we get the dataset with death time, censoring time, observed time, censoring indicator, and covariates. The times are calculated one by one based on the subject's covariates values.

Data size: $n = 1000$ or $n = 100$ subjects. 100 repetitions are conducted.

Results collection

- For $n = 1000$ scenario, since the code runs slow, we just look at the quantile time at $t =$ 10%, 25%, 50%, 75%, 90%, with covariates $x_1 = 1, x_2 = 2$.
  - Bias, standard deviation and MSE are checked.
- For $n = 100$, we just look at the quantile time at $t =$ 10%, 25%, 50%, 75%, 90%, across the dataset.
  - for each subject in the dataset, calculate the true survival function $S_i(t, x_i)$
  - for each subject in the dataset, calculate the estimated survival functions by above methods, $\hat{S}_i(t, x_i)$.
  - calculate the bias: $\hat{S}_i(t, x_i) - S_i(t, x_i)$
  - then calculate the mean bias for the dataset $\frac{1}{n}\sum_{i=1}^{n}(\hat{S}_i(t, x_i) - S_i(t, x_i))$, which is marked as the bias for the dataset, bias_{data}.
  - for the 100 repetitions, get the mean value of the bias for the dataset $\frac{1}{100}\sum_{i=1}^{100} \text{bias}_{data,i}$
  - For MSE, for each repetition, calculate $\frac{1}{n}\sum_{i=1}^{n}(\hat{S}_i(t, x_i) - S_i(t, x_i))^2$, and then calculate the mean value of the MSE across the 100 repetition.

Model Estimators

- Cox PH model with both x1, x2
- Cox PH model with only x1 (mis-specified)
- Cox PH model with only x2 (mis-specified)

*** Estimator with $m(t, x)$ function**

When there is no covariates are considered in the dataset, the most commonly applied Kaplan Meier prodcut limit estimator is defined as

$$S^{KM}(t) = \prod_{j=1}^{i}(1 - \frac{\delta_i}{n-j+1})^{I_{\{Z_i<t\}}}$$

Dikta expands the Kaplan Meier estimator by replacing δ_i with $m(Z_i)$.

$$S^{D1}(t) = P(T > t) = \prod_{j=1}^{i}(1 - \frac{m(Z_i)}{n-j+1})^{I_{\{Z_i<t\}}}$$

(When there are covariates ?)

$$P(T > t|X = x) = \prod_{j=1}^{i}(1 - \frac{m(Z_i, X_i)}{n-j+1})^{I_{\{Z_i<t\}}}$$

The Stute estimator:

$$F(x, t) = P(X \leq x, Z \leq t) = \sum_{i=1}^{n} W_i \phi(x, t)$$

If we set $\phi(x, t) = I_{(X \leq x, Z \leq t)}$, it leads to

$$F(x, t) = \sum_{i=1}^{n} W_i I_{(X \leq x, T \leq t)} = \sum_{i=1}^{n} \Big\{ \frac{\delta_i}{n-i+1} \Big[\prod_{j=1}^{i-1}(1 - \frac{\delta_j}{n-j+1}) \Big] \Big\} I_{(X_i \leq x, Z_i \leq t)}$$

Dikta replaced δ_i with $m(t, x)$ and got

$$S^D = \sum_{i=1}^{n} \Big\{ \frac{m(Z_i, X_i)}{n-i+1} \Big[\prod_{j=1}^{i-1}(1 - \frac{m(Z_j, X_j)}{n-j+1}) \Big] \Big\} I_{(X_i \leq x, Z_i \leq t)}$$

Since

$$F(x, t) = P(X \leq x, Z \leq t) = \sum_{i=1}^{n} W_i I_{(X_i \leq x, Z_i \leq t)}$$

Then

$$P(Z > t) = 1 - P(Z < t) = 1 - P(Z < t, X < \infty) = \sum_{i=1}^{n} \Big\{ \frac{\delta_i}{n-i+1} \Big[\prod_{j=1}^{i-1}(1 - \frac{\delta_j}{n-j+1}) \Big] \Big\} I_{(Z_i > t)}$$

whether it looks like Kaplan Meier estimator?

(then I made new estimators)

Formula 1 from Kaplan Meier estimator:

$$P(T > t|X = x) = \prod_{j=1}^{i}(1 - \frac{m(Z_i, X_i)}{n-j+1})^{I_{\{Z_i<t\}}} \tag{1}$$

Formula 2 from Stute estimator:

$$P(T > t|X = x) = \sum_{i=1}^{n}\Big\{\frac{m(Z_i, X_i)}{n-i+1}\Big[\prod_{j=1}^{i-1}(1 - \frac{m(Z_i, X_i)}{n-j+1})\Big]\Big\}I_{(Z_i>t)} \tag{2}$$

Result

Coefficient estimations

Table 1: Covariates estimation

	X1 (true = 2)			X2 (true = 1)		
rho	mean	sd	signifcant	mean	sd	signifcant
0.0	0.2008	0.0217	1.00	0.0979	0.0236	0.98
0.2	0.1996	0.0423	1.00	0.0974	0.0509	0.55
0.5	0.2013	0.0494	0.99	0.0964	0.0576	0.44
0.9	0.2068	0.1041	0.51	0.0915	0.1103	0.15

Table 2: logistic regression covariates estimation

	beta1 - gamma1 (true = 0.1)			beta2 - gamma2 (true = -0.1)		
rho	mean	sd	signifcant	mean	sd	signifcant
0.0	0.1009	0.0270	0.98	-0.0990	0.0184	1.00
0.2	0.1000	0.0567	0.39	-0.0980	0.0396	0.72
0.5	0.1015	0.0666	0.28	-0.0990	0.0460	0.57
0.9	0.1036	0.1021	0.12	-0.1008	0.0657	0.27

Single point $x_1 = 1, x_2 = 2$

Table 3: Table of Bias

	cox12	cox1	cox2	m()	m()hat	dikta	dikta hat
rho = 0							
qt=0.1	-0.286	-8.268	-8.904	-0.989	-1.053	-0.986	-1.05
qt=0.25	-0.561	-20.342	-21.77	-1.905	-2.069	-1.898	-2.063
qt=0.5	-0.068	-44.25	-46.594	-1.116	-1.468	-1.098	-1.45
qt=0.75	0.052	-64.266	-65.824	2.719	2.219	2.758	2.259
qt=0.9	0.456	-68.485	-67.68	10.018	9.497	10.094	9.573
rho = 0.2							
qt=0.1	-0.911	-8.846	-9.422	-1.703	-1.772	-1.69	-1.76
qt=0.25	-0.645	-20.423	-21.675	-2.424	-2.615	-2.391	-2.583
qt=0.5	-1.571	-45.753	-47.897	-3.277	-3.754	-3.186	-3.663
qt=0.75	-4.341	-68.639	-69.852	1.6	0.963	1.797	1.16
qt=0.9	-2.09	-71.18	-70.041	10.016	9.413	10.398	9.795
rho = 0.5							
qt=0.1	-0.868	-8.756	-9.222	-1.878	-1.939	-1.865	-1.926
qt=0.25	-0.617	-20.313	-21.329	-2.889	-3.056	-2.856	-3.023
qt=0.5	-1.614	-45.816	-47.565	-3.391	-3.813	-3.299	-3.722
qt=0.75	-4.493	-69.222	-70.154	2.582	2.022	2.779	2.219
qt=0.9	-1.454	-71.517	-70.525	14.182	13.658	14.558	14.033
rho = 0.9							
qt=0.1	-0.922	-8.706	-8.838	-2.248	-2.272	-2.235	-2.259
qt=0.25	-0.126	-19.644	-19.926	-3.5	-3.586	-3.467	-3.553
qt=0.5	-1.496	-45.723	-46.231	-3.654	-3.896	-3.562	-3.805
qt=0.75	-4.735	-70.528	-70.758	4.074	3.785	4.27	3.98
qt=0.9	-0.871	-73.428	-73.017	18.612	18.386	18.981	18.754

(values were timed 1000)

Table 4: Table of Standard deviation

	cox12	cox1	cox2	m()	m()hat	dikta	dikta hat
rho = 0							
qt=0.1	2.978	2.989	3.016	1.946	1.977	1.946	1.977
qt=0.25	4.575	4.597	4.591	3.18	3.552	3.18	3.552
qt=0.5	6.438	6.45	6.33	4.644	5.498	4.644	5.498
qt=0.75	8.999	9.043	8.91	6.112	7.463	6.111	7.462
qt=0.9	9.618	9.806	9.661	6.578	8.536	6.576	8.535
rho = 0.2							
qt=0.1	6.144	6.158	6.379	4.399	4.353	4.398	4.352
qt=0.25	8.716	8.756	8.96	7.46	7.629	7.457	7.627
qt=0.5	15.913	15.929	16.271	11.329	13.56	11.324	13.557

qt=0.75	18.936	18.96	18.891	13.697	17.763	13.688	17.757
qt=0.9	20.732	20.594	20.528	14.691	17.928	14.669	17.917
rho = 0.5							
qt=0.1	6.152	6.165	6.362	4.447	4.407	4.446	4.406
qt=0.25	8.626	8.627	8.865	7.701	7.84	7.699	7.837
qt=0.5	16.428	16.399	16.688	11.561	13.69	11.557	13.686
qt=0.75	19.546	19.617	19.325	14.05	17.908	14.041	17.903
qt=0.9	21.073	20.772	20.913	15.032	18.327	15.009	18.315
rho = 0.9							
qt=0.1	6.149	6.16	6.251	4.37	4.318	4.369	4.317
qt=0.25	8.42	8.439	8.544	7.901	8.151	7.899	8.149
qt=0.5	16.459	16.415	16.584	11.697	14.249	11.692	14.246
qt=0.75	19.555	19.626	19.456	14.51	18.772	14.5	18.766
qt=0.9	20.533	20.455	20.442	14.891	17.899	14.87	17.888

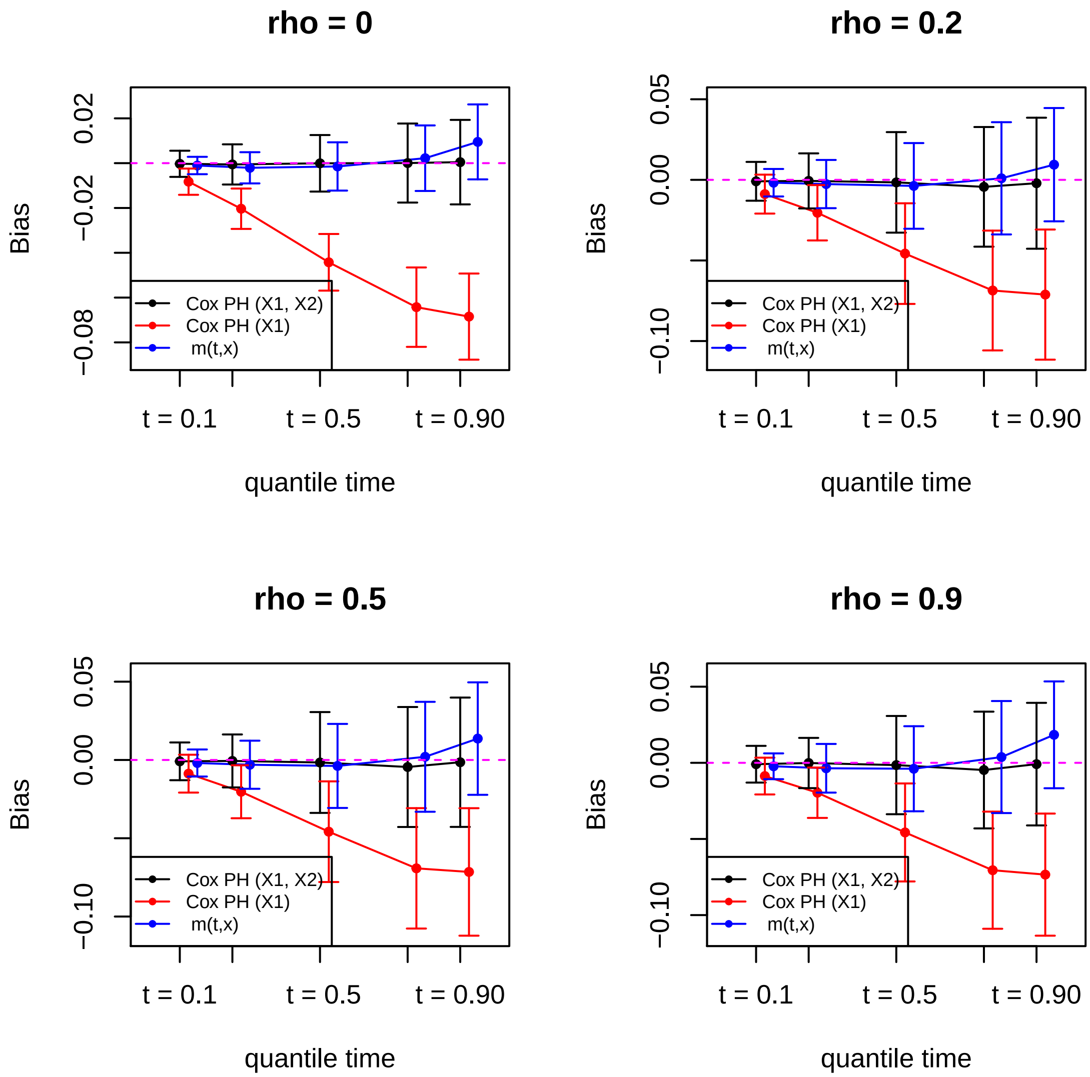
(values were timed 1000)

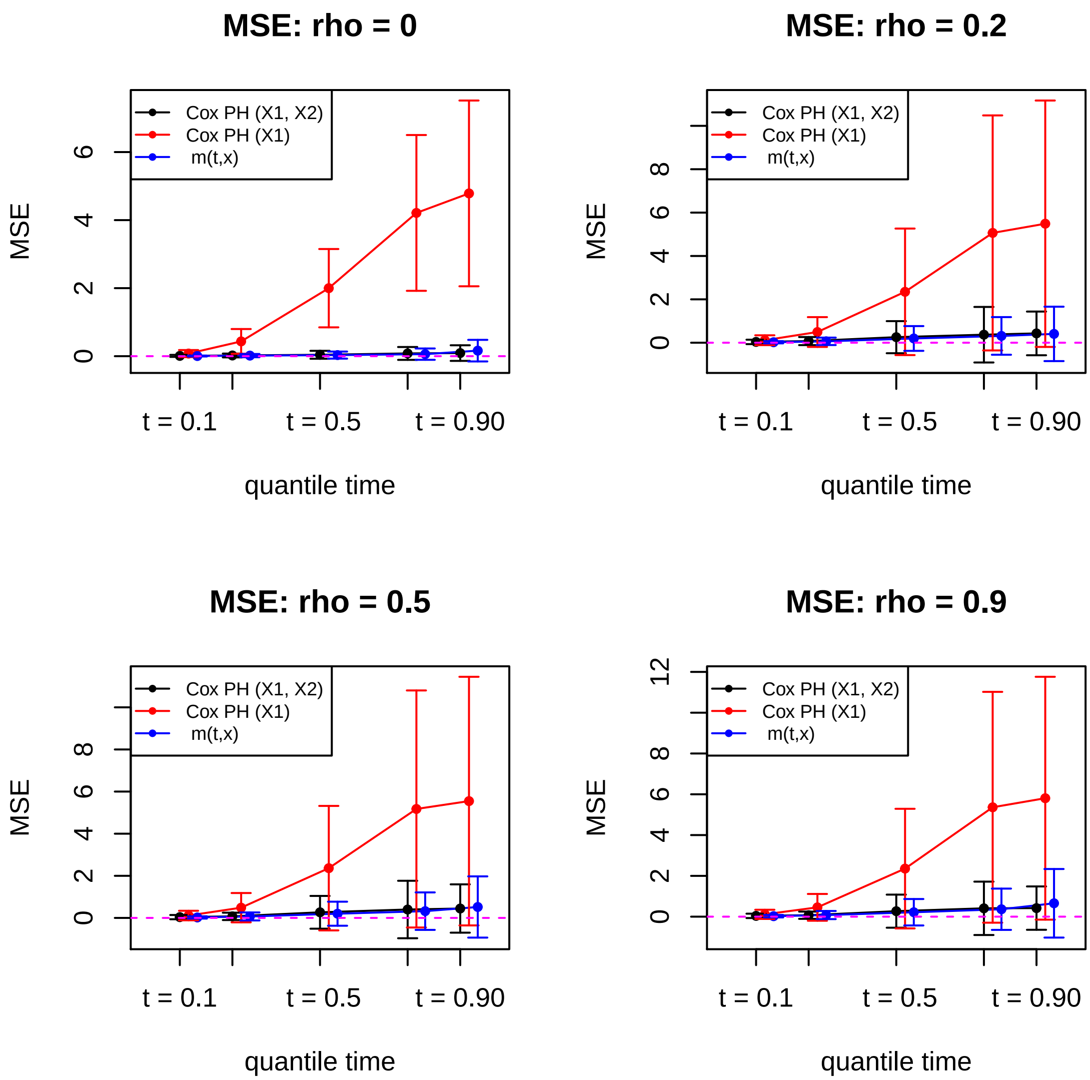
Table 5: Table of MSE

	cox12	cox1	cox2	m()	m()hat	dikta	dikta hat
rho = 0							
qt=0.1	0.009	0.077	0.088	0.005	0.005	0.005	0.005
qt=0.25	0.021	0.435	0.495	0.014	0.017	0.014	0.017
qt=0.5	0.041	1.999	2.211	0.023	0.032	0.023	0.032
qt=0.75	0.081	4.212	4.412	0.044	0.06	0.045	0.06
qt=0.9	0.091	4.785	4.673	0.143	0.162	0.145	0.164
rho = 0.2							
qt=0.1	0.038	0.115	0.128	0.022	0.022	0.022	0.022
qt=0.25	0.076	0.493	0.549	0.061	0.064	0.061	0.064
qt=0.5	0.256	2.347	2.559	0.138	0.196	0.137	0.195
qt=0.75	0.37	5.062	5.229	0.188	0.313	0.189	0.314
qt=0.9	0.43	5.487	5.322	0.314	0.407	0.321	0.414
rho = 0.5							
qt=0.1	0.038	0.114	0.125	0.023	0.023	0.023	0.023
qt=0.25	0.074	0.486	0.533	0.067	0.07	0.067	0.07
qt=0.5	0.267	2.363	2.536	0.144	0.2	0.143	0.199
qt=0.75	0.399	5.174	5.292	0.202	0.322	0.203	0.322
qt=0.9	0.447	5.547	5.412	0.425	0.519	0.435	0.529
rho = 0.9							
qt=0.1	0.039	0.114	0.117	0.024	0.024	0.024	0.024
qt=0.25	0.07	0.456	0.469	0.074	0.079	0.074	0.078
qt=0.5	0.269	2.356	2.408	0.149	0.216	0.148	0.215
qt=0.75	0.409	5.363	5.388	0.225	0.363	0.226	0.364

qt=0.9	0.419	5.807	5.747	0.566	0.655	0.579	0.669

(values were timed 1000)





$n = 100$, across covariates

Table 6: Table of Bias

	cox12	cox1	cox2	m()	m()hat	dikta	dikta hat
rho = 0							
qt=0.1	-2.932	-6.065	-6.569	-6.045	-6.335	-6.221	-6.51
qt=0.25	-9.168	-25.696	-26.826	-2.267	-3.102	-2.673	-3.501
qt=0.5	-6.352	-45.565	-46.965	-4.187	-6.32	-5.126	-7.244
qt=0.75	-6.027	-66.786	-67.842	-0.653	-4.902	-2.631	-6.854
qt=0.9	-6.104	-63.558	-62.478	-3.177	-9.455	-6.988	-13.247
rho = 0.2							
qt=0.1	-2.932	-6.065	-6.569	-6.045	-6.335	-6.221	-6.51
qt=0.25	-9.168	-25.696	-26.826	-2.267	-3.102	-2.673	-3.501

qt=0.5	-6.352	-45.565	-46.965	-4.187	-6.32	-5.126	-7.244
qt=0.75	-6.027	-66.786	-67.842	-0.653	-4.902	-2.631	-6.854
qt=0.9	-6.104	-63.558	-62.478	-3.177	-9.455	-6.988	-13.247
rho = 0.5							
qt=0.1	-2.932	-6.065	-6.569	-6.045	-6.335	-6.221	-6.51
qt=0.25	-9.168	-25.696	-26.826	-2.267	-3.102	-2.673	-3.501
qt=0.5	-6.352	-45.565	-46.965	-4.187	-6.32	-5.126	-7.244
qt=0.75	-6.027	-66.786	-67.842	-0.653	-4.902	-2.631	-6.854
qt=0.9	-6.104	-63.558	-62.478	-3.177	-9.455	-6.988	-13.247
rho = 0.9							
qt=0.1	-2.932	-6.065	-6.569	-6.045	-6.335	-6.221	-6.51
qt=0.25	-9.168	-25.696	-26.826	-2.267	-3.102	-2.673	-3.501
qt=0.5	-6.352	-45.565	-46.965	-4.187	-6.32	-5.126	-7.244
qt=0.75	-6.027	-66.786	-67.842	-0.653	-4.902	-2.631	-6.854
qt=0.9	-6.104	-63.558	-62.478	-3.177	-9.455	-6.988	-13.247

(values were timed 1000)

Table 7: Table of Standard deviation

	cox12	cox1	cox2	m()	m()hat	dikta	dikta hat
rho = 0							
qt=0.1	19.471	19.53	19.228	18.098	18.435	18.098	18.434
qt=0.25	27.611	27.983	27.828	23.841	25.853	23.841	25.85
qt=0.5	43.851	42.961	43.327	35.076	40.318	35.076	40.32
qt=0.75	47.831	52.304	52.701	43.156	49.574	43.157	49.609
qt=0.9	43.831	69.592	70.082	47.488	54.878	47.488	55.015
rho = 0.2							
qt=0.1	19.471	19.53	19.228	18.098	18.435	18.098	18.434
qt=0.25	27.611	27.983	27.828	23.841	25.853	23.841	25.85
qt=0.5	43.851	42.961	43.327	35.076	40.318	35.076	40.32
qt=0.75	47.831	52.304	52.701	43.156	49.574	43.157	49.609
qt=0.9	43.831	69.592	70.082	47.488	54.878	47.488	55.015
rho = 0.5							
qt=0.1	19.471	19.53	19.228	18.098	18.435	18.098	18.434
qt=0.25	27.611	27.983	27.828	23.841	25.853	23.841	25.85
qt=0.5	43.851	42.961	43.327	35.076	40.318	35.076	40.32
qt=0.75	47.831	52.304	52.701	43.156	49.574	43.157	49.609
qt=0.9	43.831	69.592	70.082	47.488	54.878	47.488	55.015
rho = 0.9							
qt=0.1	19.471	19.53	19.228	18.098	18.435	18.098	18.434
qt=0.25	27.611	27.983	27.828	23.841	25.853	23.841	25.85
qt=0.5	43.851	42.961	43.327	35.076	40.318	35.076	40.32

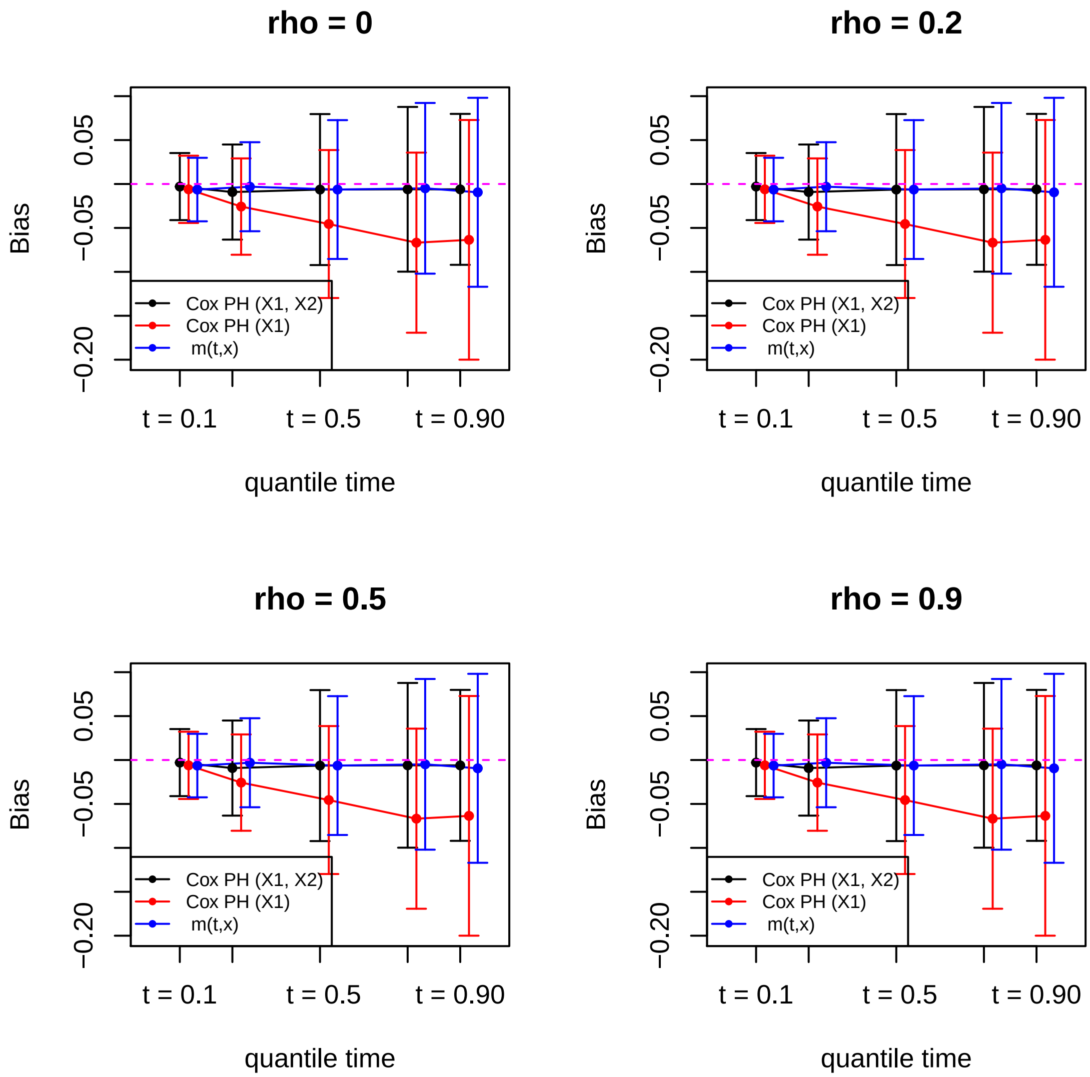
qt=0.75	47.831	52.304	52.701	43.156	49.574	43.157	49.609
qt=0.9	43.831	69.592	70.082	47.488	54.878	47.488	55.015

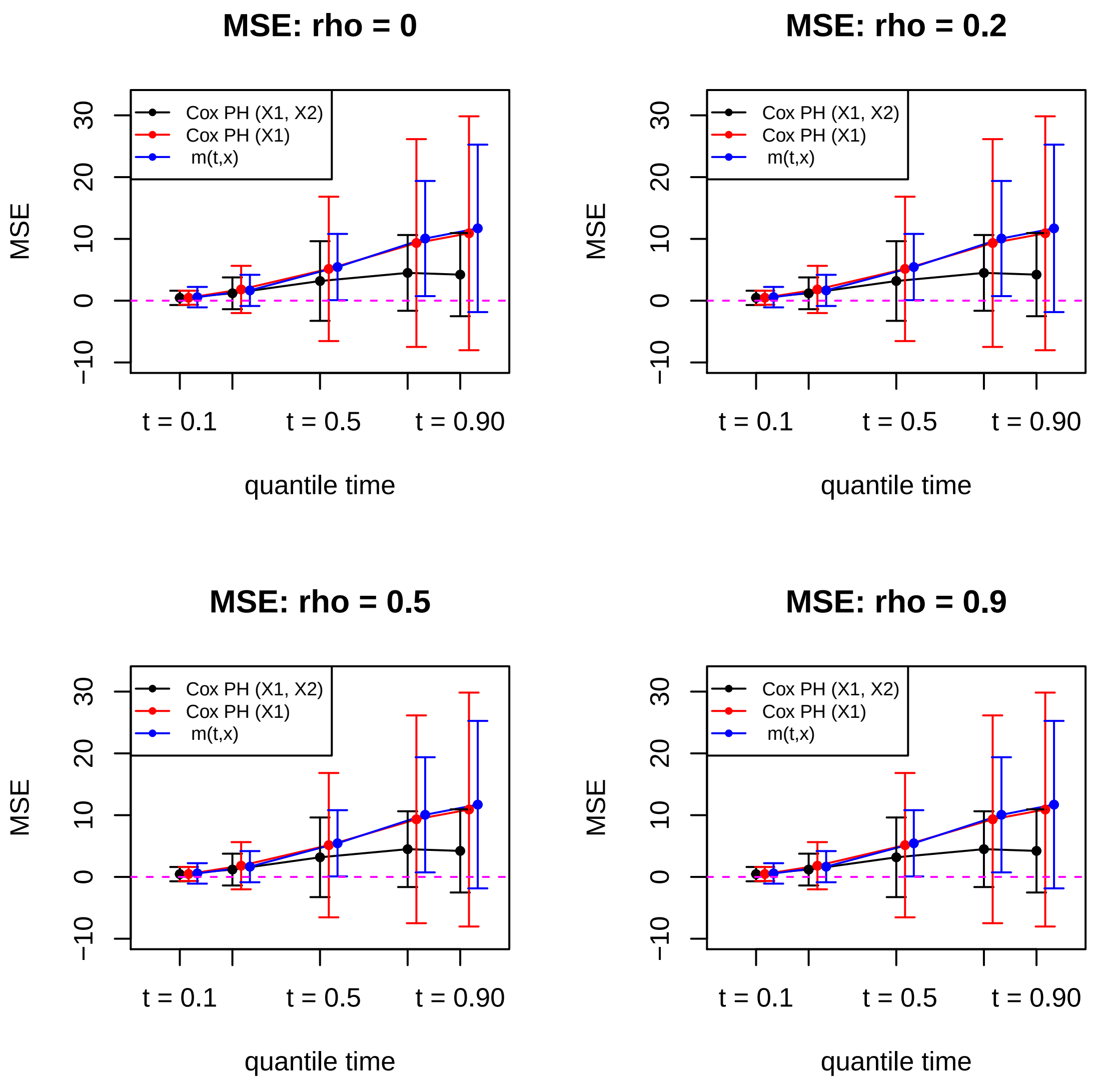
(values were timed 1000)

Table 8: Table of MSE

	cox12	cox1	cox2	m()	m()hat	dikta	dikta hat
rho = 0							
qt=0.1	0.457	0.478	0.473	0.53	0.572	0.532	0.574
qt=0.25	1.195	1.819	1.862	1.378	1.663	1.38	1.665
qt=0.5	3.181	5.146	5.325	4.258	5.446	4.267	5.459
qt=0.75	4.498	9.331	9.644	7.751	10.06	7.757	10.093
qt=0.9	4.216	10.909	11.197	8.805	11.704	8.844	11.836
rho = 0.2							
qt=0.1	0.457	0.478	0.473	0.53	0.572	0.532	0.574
qt=0.25	1.195	1.819	1.862	1.378	1.663	1.38	1.665
qt=0.5	3.181	5.146	5.325	4.258	5.446	4.267	5.459
qt=0.75	4.498	9.331	9.644	7.751	10.06	7.757	10.093
qt=0.9	4.216	10.909	11.197	8.805	11.704	8.844	11.836
rho = 0.5							
qt=0.1	0.457	0.478	0.473	0.53	0.572	0.532	0.574
qt=0.25	1.195	1.819	1.862	1.378	1.663	1.38	1.665
qt=0.5	3.181	5.146	5.325	4.258	5.446	4.267	5.459
qt=0.75	4.498	9.331	9.644	7.751	10.06	7.757	10.093
qt=0.9	4.216	10.909	11.197	8.805	11.704	8.844	11.836
rho = 0.9							
qt=0.1	0.457	0.478	0.473	0.53	0.572	0.532	0.574
qt=0.25	1.195	1.819	1.862	1.378	1.663	1.38	1.665
qt=0.5	3.181	5.146	5.325	4.258	5.446	4.267	5.459
qt=0.75	4.498	9.331	9.644	7.751	10.06	7.757	10.093
qt=0.9	4.216	10.909	11.197	8.805	11.704	8.844	11.836

(values were timed 1000)


rho = 0
rho = 0.2
rho = 0.5
rho = 0.9
Bias
0.05
−0.05
−0.20
t = 0.1
t = 0.5
t = 0.90
quantile time
Cox PH (X1, X2)
Cox PH (X1)
m(t,x)


MSE: rho = 0
MSE: rho = 0.2
MSE: rho = 0.5
MSE: rho = 0.9
Cox PH (X1, X2)
Cox PH (X1)
m(t,x)
MSE
quantile time
t = 0.1
t = 0.5
t = 0.90
−10
0
10
20
30