# Cox ph model and m() function

*2020-03-06*

## Outline

In Cox PH model, for the event time model:

$$\Lambda_T(t,x) = \Lambda_0(t)\exp(\beta_1 x_1 + ... + \beta_p x_p)$$

suppose the censoring time follows a similar model:

$$\Lambda_C(t,x) = \Lambda_0(t)\exp(\gamma_1 x_1 + ... + \gamma_p x_p)$$

If $\beta_i$ is significant in the true model, but may be not significant due to censoring, and then the model can be biased.

We would like to check the simplest condition, where the $\beta_i$ is significant, no matter there is censoring or not. Misspecification of the model without the significant part $\beta_i x_i$, will the model have a bad estimation?

## Model setting

Suppose the event time $T$ and the censoring time $C$ are both following cox models, who are sharing the same $S_0(t)$ function, i.e.

$$\text{event time: } S_T(t|X=x) = P(T>t|X=x) = S_0(t)^{\exp(\beta' x)}$$
$$\text{censoring time: } S_C(t|X=x) = P(C>t|X=x) = S_0(t)^{\exp(\gamma' x)}$$

where $X$ is the covariates vector and $\beta = (\beta_1, \beta_2, ..., \beta_p)$ are the coefficients for cox PH model in terms event time, $\gamma = (\gamma_1, \gamma_2, ..., \gamma_p)$ are the coefficients for cox PH model in terms of censoring time.

Therefore, the associated hazard functions are

$$\text{event time: } \lambda_T(t|x) = \lambda_0(t)\exp(\beta' x),\ \Lambda_T(t|x) = \Lambda_0(t)\exp(\beta' x)$$
$$\text{censoring time: } \lambda_C(t|x) = \lambda_0(t)\exp(\gamma' x),\ \Lambda_C(t|x) = \Lambda_0(t)\exp(\gamma' x)$$

The associated $m()$ function can be

$$m(t,x) = \frac{\lambda_T(t|x)}{\lambda_T(t|x)+\lambda_C(t|x)} = \frac{\lambda_0(t)\exp(\beta' x)}{\lambda_0(t)\exp(\beta' x)+\lambda_0(t)\exp(\gamma' x)} = \frac{1}{1+\exp(-(\beta-\gamma)' x)}$$

which follows a logistic distribution.

Let's just consider a two dimension simple scenario, where $\lambda_0(t) = 1$, $\Lambda_0(t) = t, S_0(t) = \exp(-t)$

- Event time:

$$\text{hazard function: } \lambda_T(t|x) = \lambda_0(t)\exp(\beta_1 x_1 + \beta_2 x_2) = \exp(\beta_1 x_1 + \beta_2 x_2)$$
$$\text{cumulative hazard function: } \Lambda_T(t|x) = \Lambda_0(t)\exp(\beta_1 x_1 + \beta_2 x_2) = t\exp(\beta_1 x_1 + \beta_2 x_2)$$
$$\text{survival function: } S(t|x) = S_0(t)^{\exp(\beta_1 x_1 + \beta_2 x_2)} = \exp(-t \times (\beta_1 x_1 + \beta_2 x_2))$$

- Censoring time

$$\text{hazard function: } \lambda_C(t|x) = \lambda_0(t)\exp(\gamma_1 x_1 + \gamma_2 x_2) = \exp(\gamma_1 x_1 + \gamma_2 x_2)$$
$$\text{cumulative hazard function: } \Lambda_C(t|x) = \Lambda_0(t)\exp(\gamma_1 x_1 + \gamma_2 x_2) = t\exp(\gamma_1 x_1 + \gamma_2 x_2)$$
$$\text{survival function: } S(t|x) = S_0(t)^{\exp(\gamma_1 x_1 + \gamma_2 x_2)} = \exp(-t \times (\gamma_1 x_1 + \gamma_2 x_2))$$

Let's consider two numerical settings

| | $\beta_1$ | $\beta_2$ | $\gamma_1$ | $\gamma_2$ |
|---|---|---|---|---|
| setting 1: | 2 | 0.1 | 0.2 | -0.2 |
| setting 2: | 0.1 | 0.1 | 0.2 | -0.2 |

We will check whether:

1. $X_1, X_2$ have effect on the survival time $S(t)$, i.e. the coefficients are significant in the Cox PH model fitted with $X_1, X_2$
2. $X_1, X_2$ have effect on the censoring time, i.e. the coefficients are significant by fitting the logistic regression with status $\sim X_1, X_2$
3. When the Cox PH model is mis-specified, with only $X_2$, how well the survival time is estimated?
4. The estimation methods:
   - Cox PH model with $X_1, X_2$
   - Cox PH model with only $X_2$
   - by using the true $m(t, x)$ function (true $\beta_1 - \gamma_1, \beta_2 - \gamma_2$)
   - by using the estimated $\hat{m}(t, x)$ function (estimated from the logistic regression with $X_1, X_2$)

## Model misspecification

When the model is mis-spepicfied with only one of the two covariates that is associated with survival time, will the model get a bad estimation?

In our setting, $X_1, X_2$ are two random variables from independent normal distributions, $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$

$$f_1(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp(-\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2), f_2(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp(-\frac{1}{2\sigma_2^2}(x_2 - \mu_2)^2)$$

$$f_{1,2}(x_1, x_2) = f_1(x_1) f_2(x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp(-\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2) \exp(-\frac{1}{2\sigma_2^2}(x_2 - \mu_2)^2)$$

Cox PH model

$$\lambda(t|x_1, x_2) = \lambda_0(t) \exp(\beta_1 x_1 + \beta_2 x_2)$$

The joint pdf of $t$ and $X_1, X_2$ is

$$\begin{aligned}\lambda(t, x_1, x_2) =& \lambda(t|x_1, x_2) f(x_1, x_2) \\ =& \lambda_0(t) \exp(\beta_1 x_1 + \beta_2 x_2) \frac{1}{2\pi\sigma_1\sigma_2} \exp(-\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2) \exp(-\frac{1}{2\sigma_2^2}(x_2 - \mu_2)^2)\end{aligned}$$

$$\begin{aligned}\lambda(t, x_1) =& \int_{-\infty}^{+\infty} \lambda(t, x_1, x_2) dx_2 \\ =& \lambda_0(t) \exp(\beta_1 x_1) \frac{1}{\sqrt{2\pi}\sigma_1} \exp(-\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2) \exp(\frac{1}{2\sigma_1^2}[(\mu_2 + \beta_2\sigma_2^2)^2 - \mu_2^2])\end{aligned}$$

Therefore,

$$\lambda(t|x_1) = \frac{\lambda(t, x_1)}{f_1(x_1)} = \exp(\frac{1}{2\sigma_1^2}[(\mu_2 + \beta_2\sigma_2^2)^2 - \mu_2^2]) \lambda_0(t) \exp(\beta_1 x_1)$$

$$\Lambda(t|x_1) = \exp(\frac{1}{2\sigma_1^2}[(\mu_2 + \beta_2\sigma_2^2)^2 - \mu_2^2]) \Lambda_0(t) \exp(\beta_1 x_1)$$

$$S(t|x_1) = \big[\exp\big(-C\Lambda_0(t)\big)\big]^{\exp(\beta_1 x_1)}$$

where $C = \exp(\frac{1}{2\sigma_1^2}[(\mu_2 + \beta_2\sigma_2^2)^2 - \mu_2^2])$

Therefore,

- If there do not have censoring, the estimation should be good
- The hazard ratio for $X_1$ will not change if $X_2$ is dropped from the model
- However, the informative censoring may change the story and bring bias in the estimation

## Results

To estimated how the model is fitted, we will check

- AUC and ROC curve with confidence interval
- The mean absolute difference between $S(t)$ and $\hat{S}(t)$ at quantile time 10%, 25%, 50%, 75%, and 90%.

For the first setting, $\beta_1 = 2, \beta_2 = 0.1, \gamma_1 = 0.2, \gamma_2 = -0.2$, we generate a training dataset with 1000 subjects and a test dataset with 100 subjects. The time dependent AUCs and the differences between $S(t)$ and $\hat{S}(t)$ are calculated at quantile times 10%, 25%, 50%, 75%, and 90%. The procedures are repeated for 100 times.

If we estimate with true death time and all status = 1

Table 2: Estimation of coefficient beta

| | true coef | x1 + x2 | x1 only | x2 only |
|---|---|---|---|---|
| beta1 | 2.0 | 1.993 | 1.988 | NA |
| beta2 | 0.1 | 0.085 | NA | 0.097 |

The model with $X_1$ and $X_2$ showing that both of them have significant effects:

```
##         coef exp(coef)   se(coef)         z      Pr(>|z|)
## x1 1.9931888  7.338899 0.06075762 32.805575 4.903230e-236
## x2 0.0854206  1.089175 0.03201869  2.667835  7.634169e-03
```

If we estimate with the observed time and status,

Table 3: Estimation of coefficient beta

| | true coef | x1 + x2 | x1 only | x2 only |
|---|---|---|---|---|
| beta1 | 2.0 | 2.072 | 2.06 | NA |
| beta2 | 0.1 | 0.154 | NA | 0.151 |

The model with $X_1$ and $X_2$ showing that both of them have significant effects:

```
##         coef exp(coef)   se(coef)         z      Pr(>|z|)
## x1 2.0715163  7.936849 0.07831716 26.450351 3.614963e-154
## x2 0.1537429  1.166191 0.04766883  3.225229  1.258719e-03
```

Fit the logistic regression:

```
fit_lg = glm(status ~ x1 + x2 - 1, data = data, family = 'binomial')
as.vector(summary(fit_lg)$coefficient[,1])
```

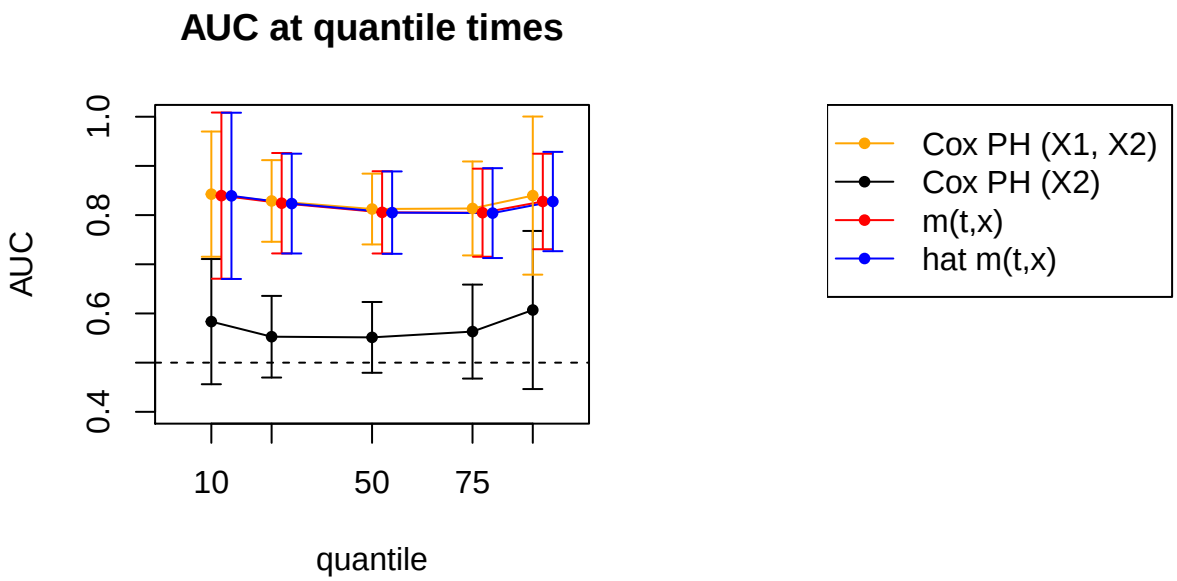
```
## [1] 1.8382064 0.2925561
```

```
beta - gamma
```

```
## [1] 1.8 0.3
```

### The AUC estimation

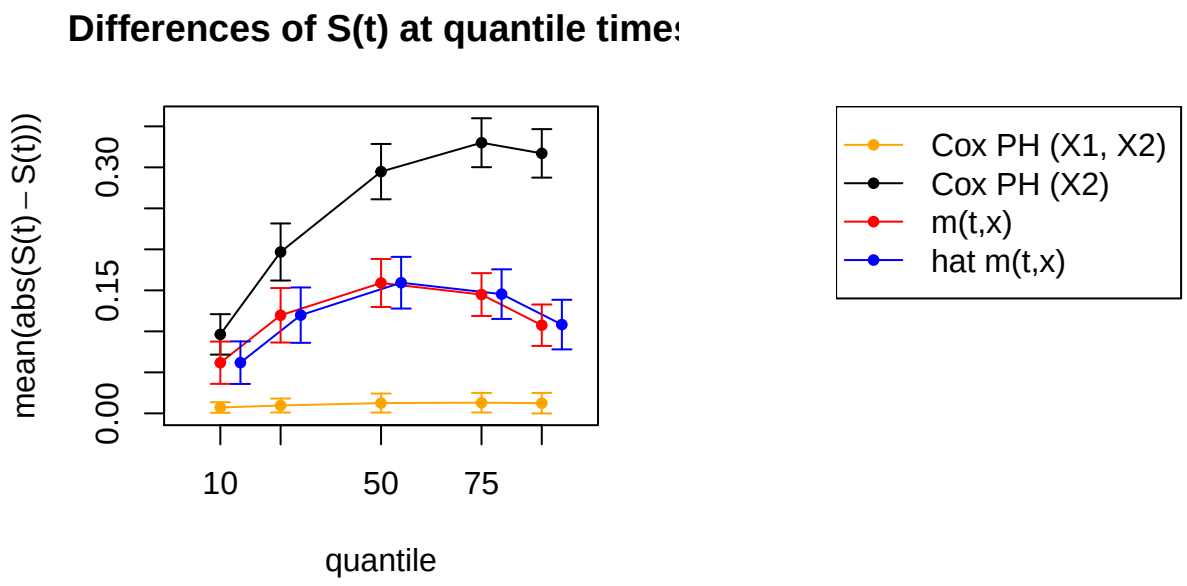
| quantile time | Cox (x1,x2) | | Cox (x2 only) | | m(t,x) | | hat m(t,x) | |
|---|---|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd | mean | sd |
| 10 | 0.843 | 0.084 | 0.583 | 0.065 | 0.840 | 0.086 | 0.839 | 0.086 |
| 25 | 0.829 | 0.052 | 0.553 | 0.042 | 0.824 | 0.052 | 0.823 | 0.052 |
| 50 | 0.812 | 0.042 | 0.551 | 0.037 | 0.806 | 0.043 | 0.805 | 0.043 |
| 75 | 0.813 | 0.044 | 0.563 | 0.049 | 0.805 | 0.046 | 0.804 | 0.047 |
| 90 | 0.840 | 0.046 | 0.607 | 0.082 | 0.828 | 0.050 | 0.828 | 0.052 |



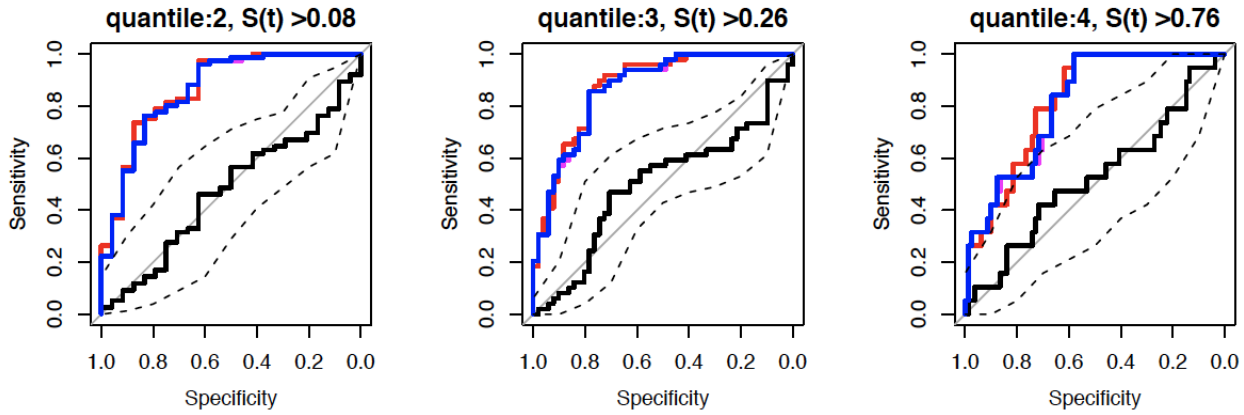

In this setting, the other models fit much better than the one with only $X_2$ in the cox ph model.

**The difference between $S(t)$ and $\hat{S}(t)$**

| quantile time | Cox (x1,x2) | | Cox (x2 only) | | m(t,x) | | hat m(t,x) | |
|---|---|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd | mean | sd |
| 10 | 0.007 | 0.003 | 0.096 | 0.013 | 0.062 | 0.013 | 0.062 | 0.013 |
| 25 | 0.010 | 0.004 | 0.197 | 0.018 | 0.120 | 0.017 | 0.120 | 0.017 |
| 50 | 0.013 | 0.006 | 0.295 | 0.017 | 0.159 | 0.015 | 0.159 | 0.016 |
| 75 | 0.013 | 0.006 | 0.330 | 0.015 | 0.145 | 0.013 | 0.145 | 0.015 |
| 90 | 0.012 | 0.006 | 0.317 | 0.015 | 0.107 | 0.013 | 0.108 | 0.015 |




The roc curve




**Setting 2**

For the first setting, $\beta_1 = 0.1, \beta_2 = 0.1, \gamma_1 = 0.2, \gamma_2 = -0.2$, we generate a training dataset with 1000 subjects and a test dataset with 100 subjects. The time dependent AUCs and the differences between $S(t)$ and $\hat{S}(t)$ are calculated at quantile times 10%, 25%, 50%, 75%, and 90%. The procedures are repeated for 100 times.

If we estimate with true death time and all status = 1

Table 6: Estimation of coefficient beta

| | true coef | x1 + x2 | x1 only | x2 only |
|---|---|---|---|---|
| beta1 | 0.1 | 0.132 | 0.133 | NA |
| beta2 | 0.1 | 0.124 | NA | 0.124 |

The model with $X_1$ and $X_2$ showing that both of them have significant effects:

```
##          coef exp(coef)   se(coef)        z     Pr(>|z|)
## x1 0.1319771  1.141082 0.03031285 4.353832 1.337782e-05
```

```
## x2 0.1239441  1.131953 0.03080215 4.023878 5.724766e-05
```

If we estimate with the observed time and status,

Table 7: Estimation of coefficient beta

| | true coef | x1 + x2 | x1 only | x2 only |
|---|---|---|---|---|
| beta1 | 0.1 | 0.172 | 0.172 | NA |
| beta2 | 0.1 | 0.085 | NA | 0.084 |

The model with $X_1$ and $X_2$ showing that both of them have significant effects:

```
##          coef exp(coef)   se(coef)        z     Pr(>|z|)
## x1 0.17164035  1.187251 0.04450060 3.857035 0.0001147708
## x2 0.08455883  1.088237 0.04380487 1.930352 0.0535632291
```

Fit the logistic regression:

```
fit_lg = glm(status ~ x1 + x2 - 1, data = data, family = 'binomial')
as.vector(summary(fit_lg)$coefficient[,1])
```

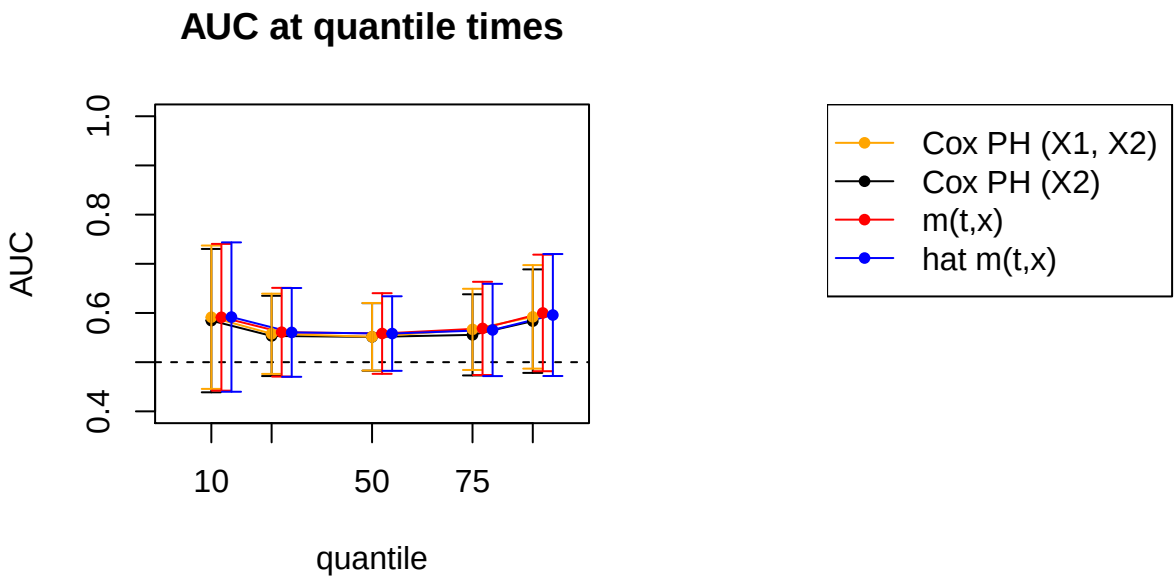
```
## [1] 0.0262941 0.2950706
```

```
beta - gamma
```

```
## [1] -0.1  0.3
```

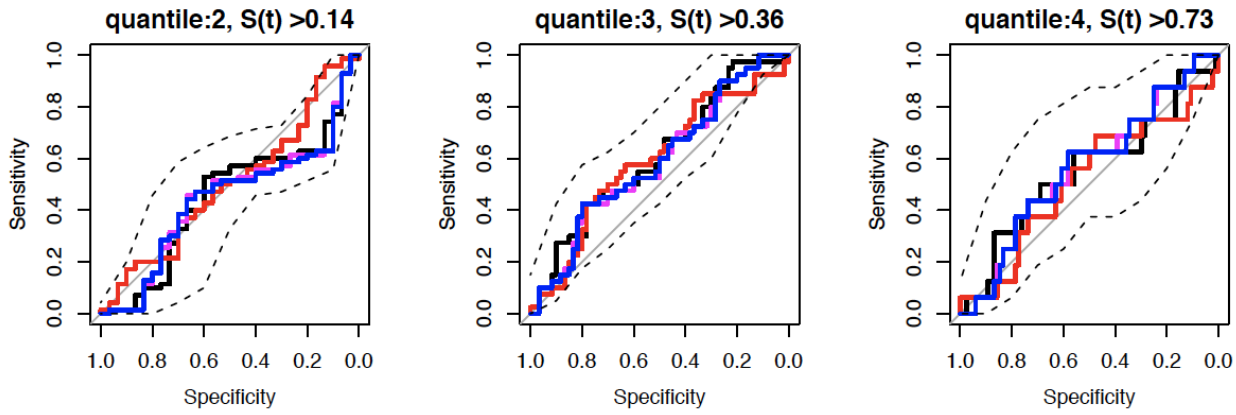
**The AUC estimation**

| quantile time | Cox (x1,x2) | | Cox (x2 only) | | m(t,x) | | hat m(t,x) | |
|---|---|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd | mean | sd |
| 10 | 0.591 | 0.079 | 0.584 | 0.074 | 0.591 | 0.076 | 0.592 | 0.077 |
| 25 | 0.558 | 0.042 | 0.553 | 0.042 | 0.561 | 0.046 | 0.560 | 0.046 |
| 50 | 0.552 | 0.035 | 0.551 | 0.035 | 0.558 | 0.042 | 0.558 | 0.039 |
| 75 | 0.567 | 0.048 | 0.556 | 0.042 | 0.569 | 0.048 | 0.565 | 0.048 |
| 90 | 0.592 | 0.072 | 0.583 | 0.054 | 0.600 | 0.060 | 0.596 | 0.063 |



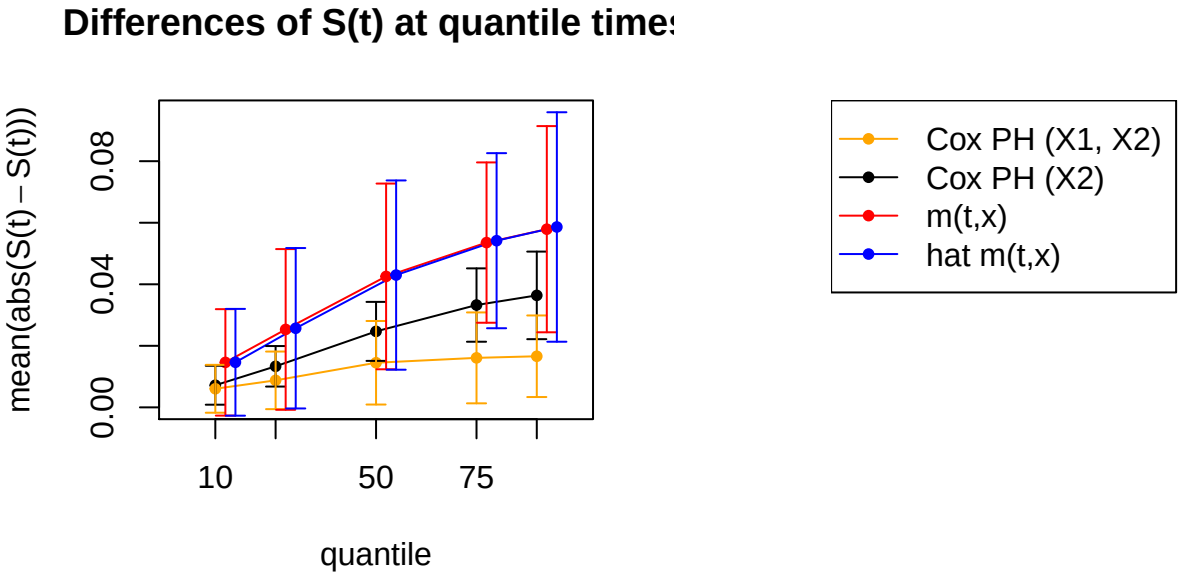

In this setting, the other models fit much better than the one with only $X_2$ in the cox ph model.

The roc curve




**The difference between $S(t)$ and $\hat{S}(t)$**

| quantile time | Cox (x1,x2) | | Cox (x2 only) | | m(t,x) | | hat m(t,x) | |
|---|---|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd | mean | sd |
| 10 | 0.006 | 0.004 | 0.007 | 0.003 | 0.015 | 0.009 | 0.015 | 0.009 |
| 25 | 0.009 | 0.005 | 0.013 | 0.003 | 0.025 | 0.013 | 0.026 | 0.013 |
| 50 | 0.014 | 0.007 | 0.025 | 0.005 | 0.043 | 0.015 | 0.043 | 0.016 |
| 75 | 0.016 | 0.008 | 0.033 | 0.006 | 0.054 | 0.013 | 0.054 | 0.015 |
| 90 | 0.017 | 0.007 | 0.036 | 0.007 | 0.058 | 0.017 | 0.059 | 0.019 |


Differences of S(t) at quantile times
mean(abs(S(t) – S(t)))
0.00
0.04
0.08
10
50
75
quantile
Cox PH (X1, X2)
Cox PH (X2)
m(t,x)
hat m(t,x)