NTNU

TMA 4275 Solution Obligatory II

May 8, 2017

This solution is largely inspired by your hand ins. This is not meant to be a comprehensive rendering of what was expected of you, but rather to give you an idea of what it should have looked like.

1 Grading Scale

1. Q1 5 pts.
2. Q2 20 pts.
3. Q3 20 pts.
4. Q4 10 pts.
5. Q5 10 pts.
6. Q6 10 pts.
7. Q7 10 pts.
8. Q8 10 pts.
9. Q9 5 pts.

2 Problem 1

2.1 What was expected

1. Martingale plots for $x_1, x_2, log(x_1)$.
2. Comment the plots and explain which transformation is chosen and why.

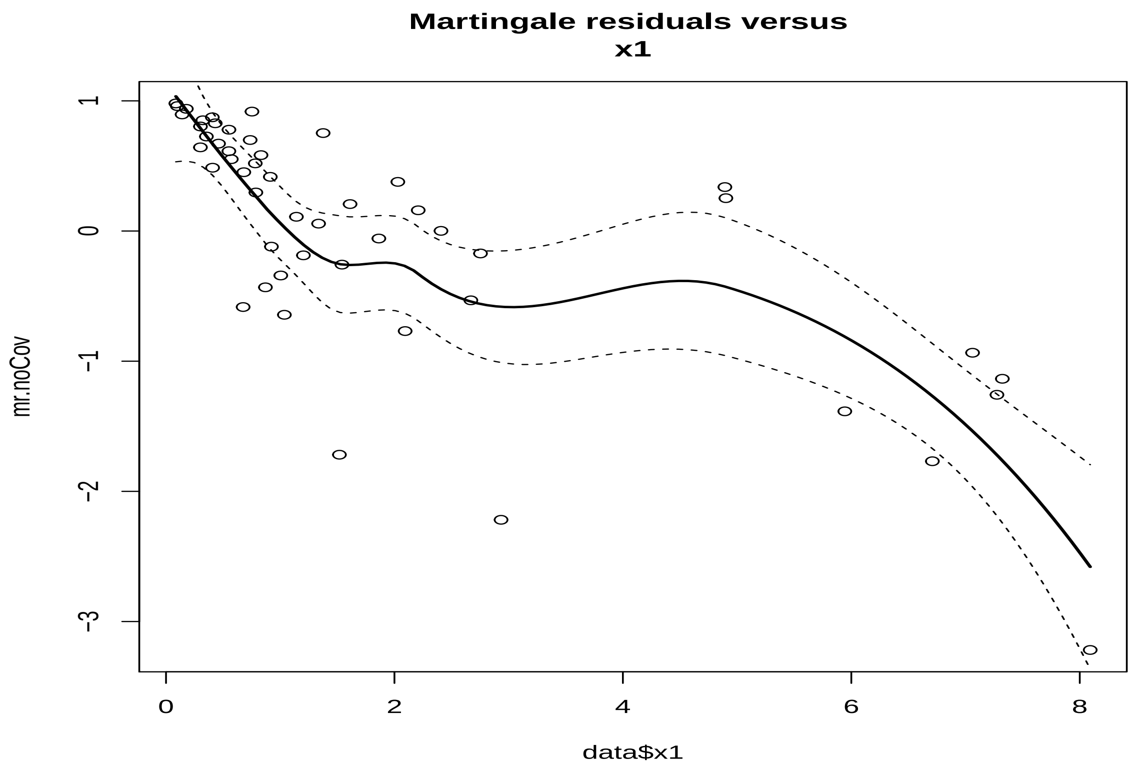

Figure 1: The plot represents the Martingale residuals plotted against x_1, with a smoothing line and confidence band included.

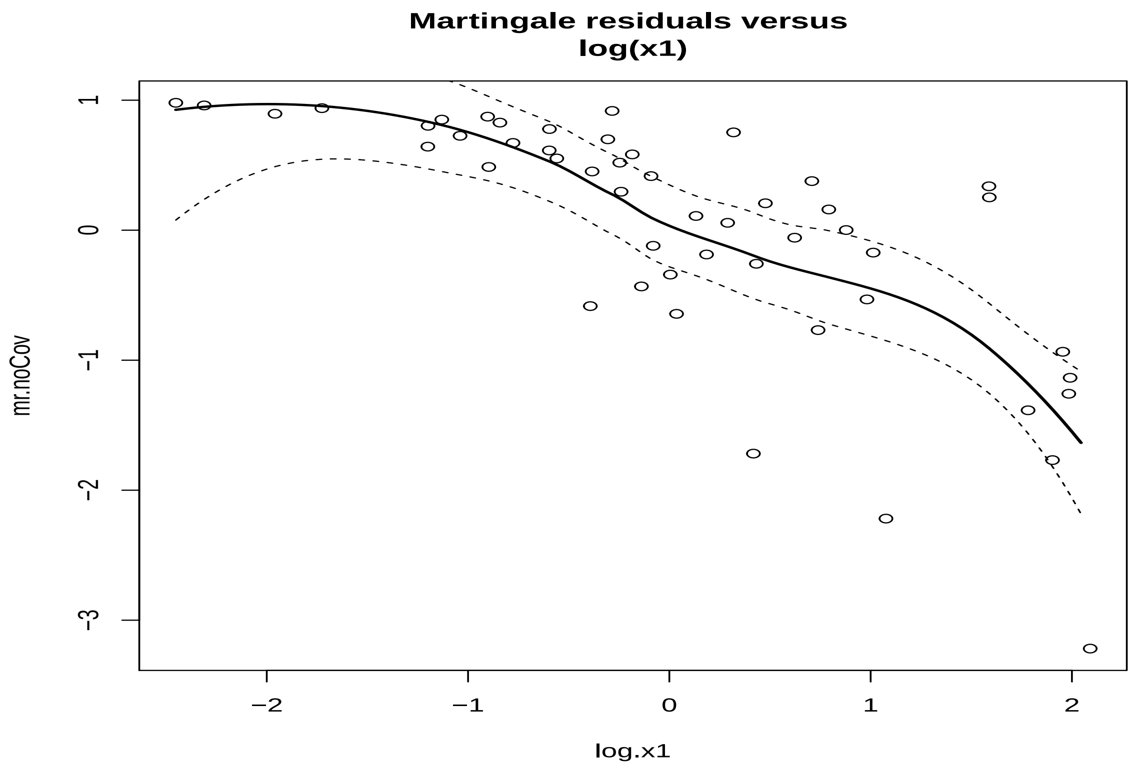

Figure 2: The plot represents the Martingale residuals plotted against ln(x_1), also with a smoothing line and confidence band included.

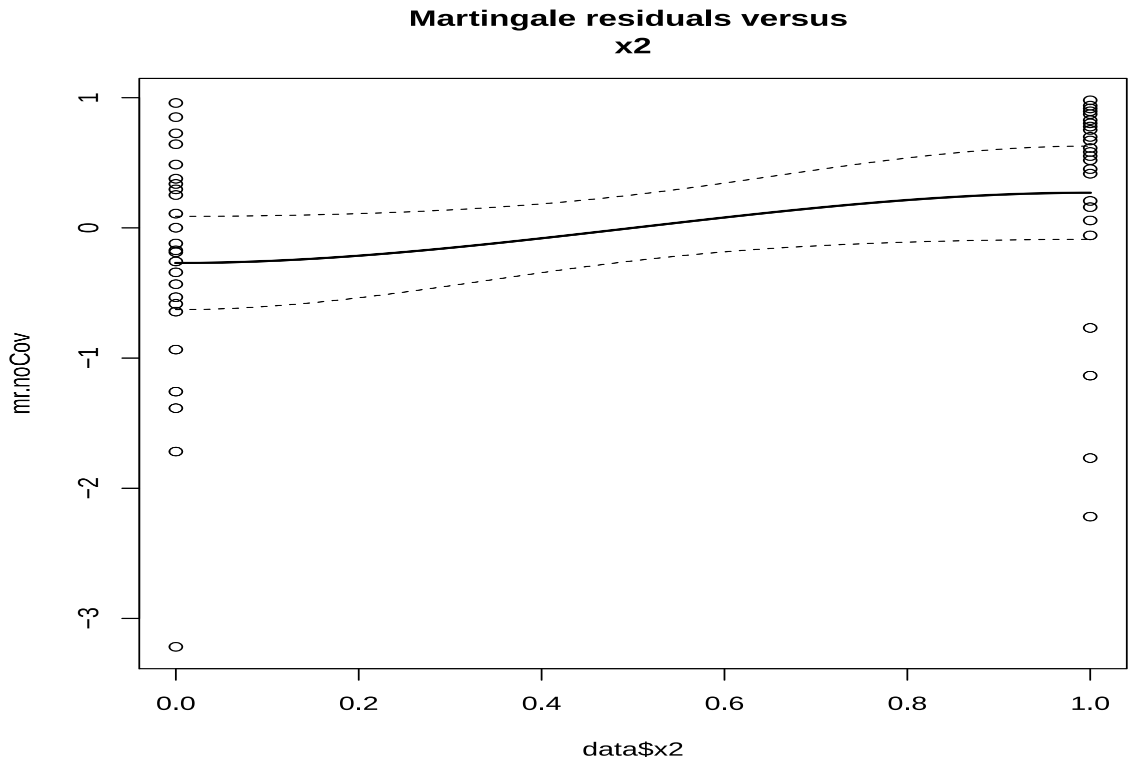

Figure 3: The plot represents the Martingale residuals plotted against x_2, once again with a smoothing line and confidence band included.

2.2 Solution proposal

Two observations can be made when looking at 1,2,3, first of all, applying a log transform to a binary variable won't help much. Finally, according to Moore Ch.7, $log(x_1)$ will be preferred to x_1 for the martingale plot has a clearer linear trend.

3 Problem 2

3.1 What was expected

1. Schoenfeld residuals plot + plot assessment of the PHH
2. log-minus-log plot
3. assess PHH with the log minus plot
4. assess weibull with the log minus plot

3.2 Solution proposal

One may assess the proportional hazard assumption by viewing the schoenfeld residuals found in figure 4. The diagnostics tells us that if one can spot a visible trend in the residuals, the PHH does not hold. A slight trend may be observed, as the residuals tend to follow to approximately straight lines, not lying in the confidence interval. This tells us that the proportional hazard assumption might not hold.

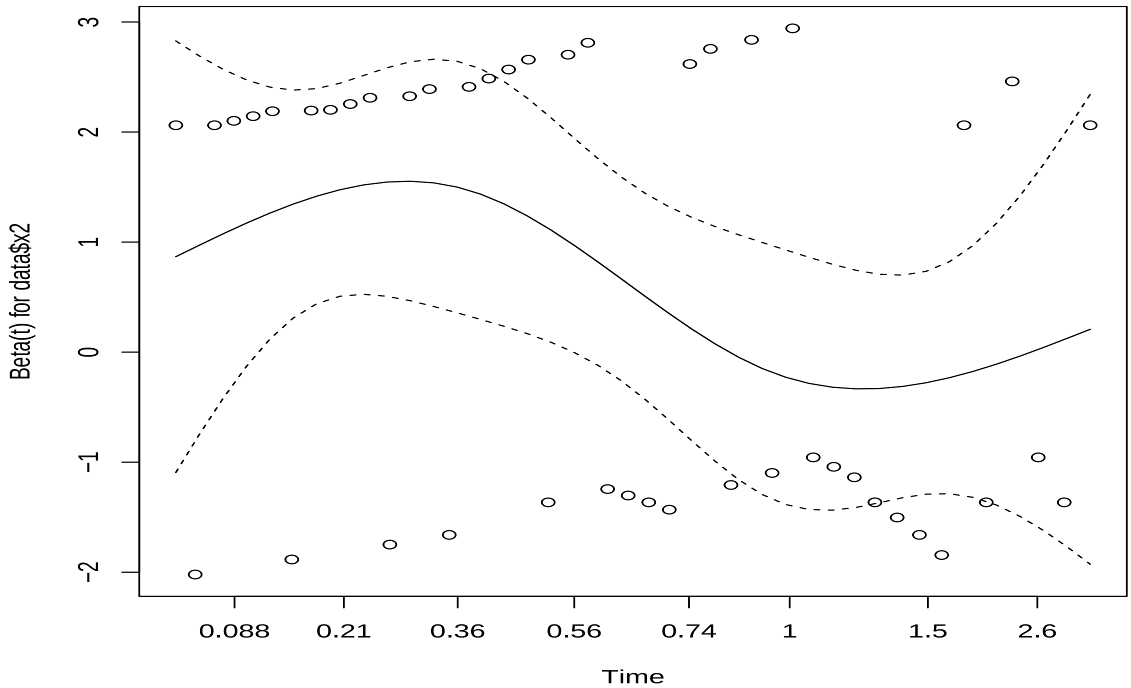

Figure 4: The plot represents the Schoenfeld residuals plotted against the observed failure times.

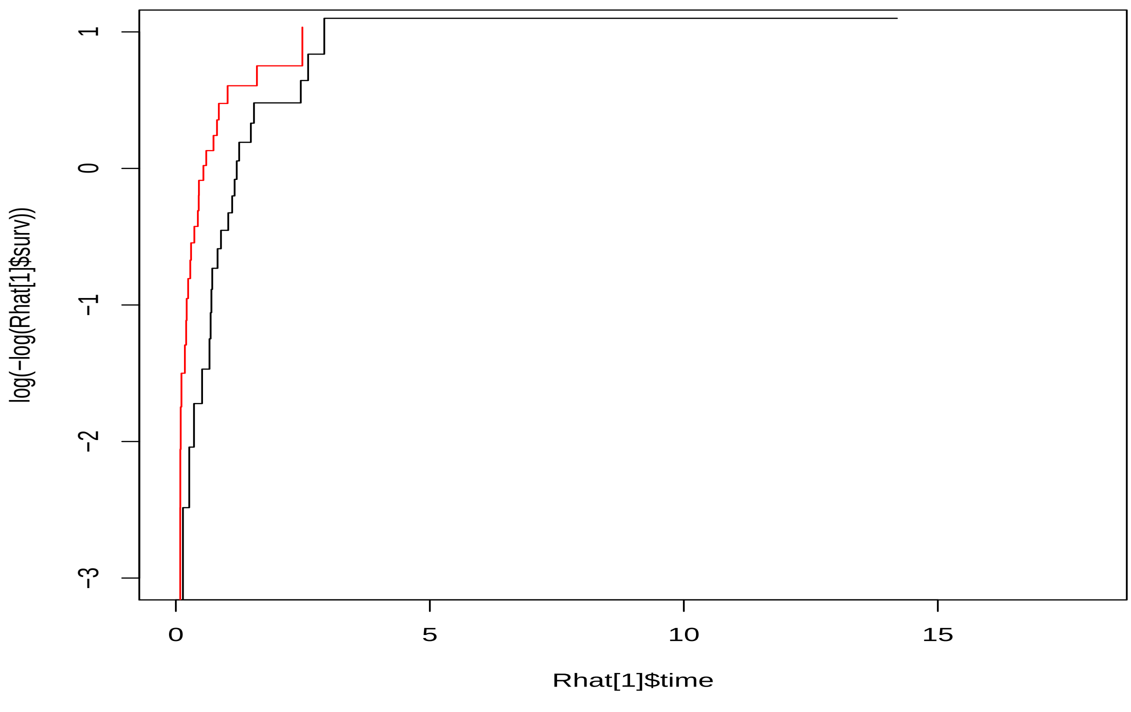

Figure 5: The plot represents the log(-log) plot for the Kaplan-Meier estimate of the survival function against time.

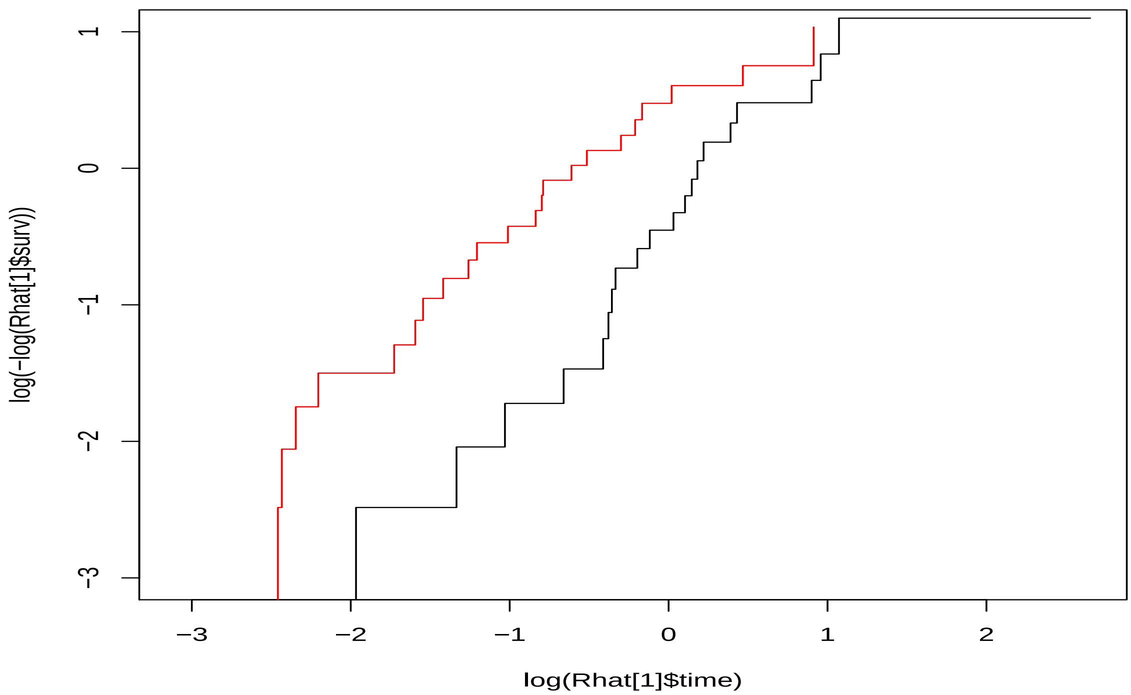

Figure 6: Similar as in the previous figure, the plot represents the log(-log) plot for the Kaplan-Meier estimate of the survival function, with the exceotion of here being plotted against the logarithm of the time.

From 5, it looks like the distance between the two curves is not constant, make us doubt that the PHH holds.

From 6, it seems quite clear that the curves aren't straight lines, making an underlying weibull model unlikely.

4 Problem 3

4.1 What was expected

This question has been a disaster for many students, because the question was not answered correctly, instead of testing the whether β_i was zero or not, global tests on the model were used.

4.2 Solution proposal

```
coxph.bothCov <- coxph(Surv(data$y, data$delta)~log.x1+data$x2) #Both log(x1) and x2
coxph.logx1 <- coxph(Surv(data$y, data$delta)~log.x1) # Model only for log(x1)
# coxph.x2 have from before

Wx2 = 2*(logLik(coxph.bothCov)-logLik(coxph.logx1)) #excluding x2 (testing beta2=0)
Wx1 = 2*(logLik(coxph.bothCov)-logLik(coxph.x2))    #excluding log(x1) (testing beta1=0)

W_q = qchisq(p = 0.05, df=1, lower.tail = FALSE)
```

```
if(Wx1 > W_q){
  message('We reject H0')
}else{
  message('We fail to reject H0')
}

## We reject H0

if(Wx2 > W_q){
  message('We reject H0')
}else{
  message('We fail to reject H0')
}

## We reject H0

p.log.x1 = pchisq(q = Wx1, df = 1, lower.tail = FALSE)
px2 = pchisq(q = Wx2, df = 1, lower.tail = FALSE)

p.log.x1

## 'log Lik.' 1.058045e-14 (df=2)

px2

## 'log Lik.' 0.002005519 (df=2)

summary(coxph.bothCov)

## Call:
## coxph(formula = Surv(data$y, data$delta) ~ log.x1 + data$x2)
##
##   n= 50, number of events= 45
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## log.x1   -1.9580    0.1411   0.3227 -6.067  1.3e-09 ***
## data$x2   1.0680    2.9096   0.3512  3.041  0.00236 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## log.x1     0.1411     7.0855   0.07498    0.2657
## data$x2    2.9096     0.3437   1.46178    5.7913
##
## Concordance= 0.855  (se = 0.051 )
## Rsquare= 0.722   (max possible= 0.996 )
## Likelihood ratio test= 64.03  on 2 df,   p=1.243e-14
## Wald test            = 37.73  on 2 df,   p=6.419e-09
## Score (logrank) test = 48.98  on 2 df,   p=2.313e-11
```

As a result, it is possible for us to examine how the p-values based on the likelihood

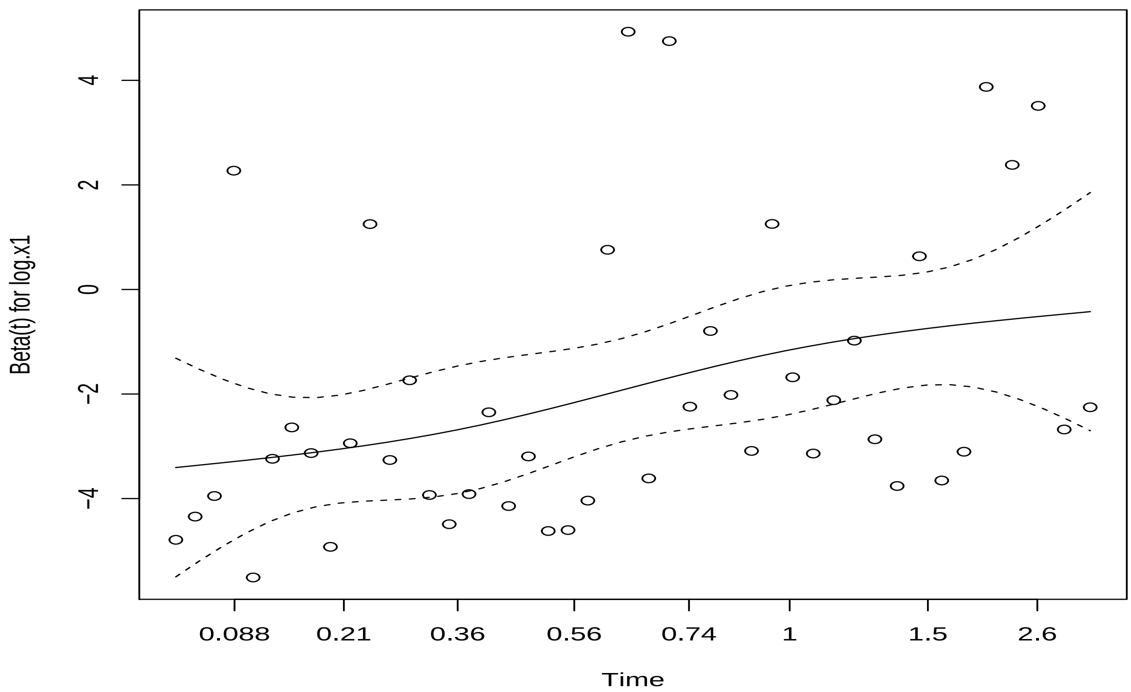

Figure 7: The plot represents the Schoenfeld residuals for the extended Cox-model, including both covariates.

ratio test compares to the Wald-test. In the code-block, the p-values for the likelihood ratio test are given by $px1$ and $px2$, whereas the p-values for the Wald-test can be found from summary($coxph.bothCov$). What is found is that p-value the logarithm of x_1 are 1.058e-14 (likelihood) and 1.3e-9 (Wald), whereas for x_2 we got 0.00201 (likelihood) and 0.00236 (Wald). The LRT and the wald test seem to agree.

5 Problem 4

5.1 What was expected

Here, to answer the question correctly, one had to plot the schoenfeld residuals, comment on the plot, carry out the hypothesis testing and correctly interpret the output.

5.2 Solution proposal

It is possible to reassess the proportional hazards assumptions for the extended model, once again by plotting the Schoenfeld residuals against observed failure times.

As seen from figures 7 and 8, there is no clear trend in the residual plot. This implies that the proportional hazards assumption very likely holds for the extended model. In addition to the plot, one may perform a formal hypothesis test of the proportional hazards assumption available via the *cox.zph* function, which was done in the following manner

Rejecting the null-hypothesis is equivalent to rejecting the proportional hazard assumption, that it does not hold. From these we can conclude the the proportional hazard does not hold for the log of x_1, and that it is on the border between holding or not for x_2.

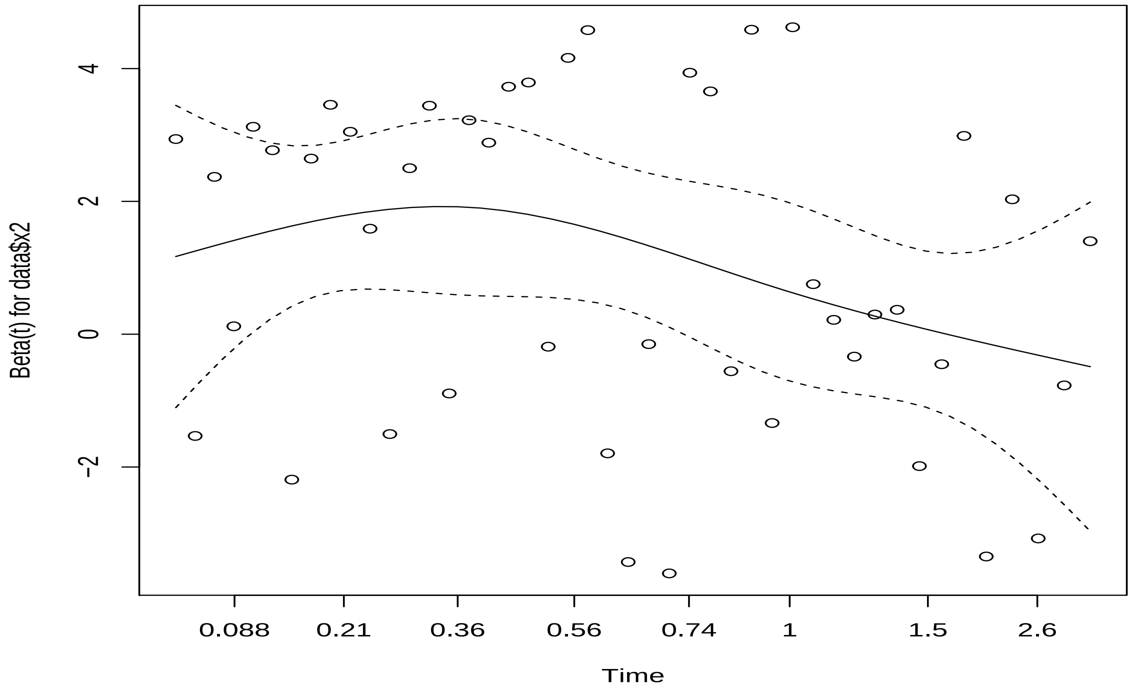

Figure 8: The plot represents the Schoenfeld residuals for the extended Cox-model, including both covariates.

6 Problem 5

6.1 What was expected

Plot of the baseline, plots of the 6 survival functions.

6.2 Solution proposal

The plot for the baseline survival and for the subjects within each group may be observed in figure 9 and 10 respectively.

7 Problem 6

7.1 What was expected

Formula for the baseline and plot, then check linearity and conclude.

7.2 Solution proposal

Note that:

$$R_0(t) = e^{-(t/\theta)^\alpha} \tag{1}$$

Then

$$\ln(-\ln(R_0(t))) = \ln(-\ln(e^{-(t/\theta)^\alpha})) = \ln(-(-(t/\theta)^\alpha)) = \ln((t/\theta)^\alpha) = \alpha\ln t - \alpha\ln\theta. \tag{2}$$

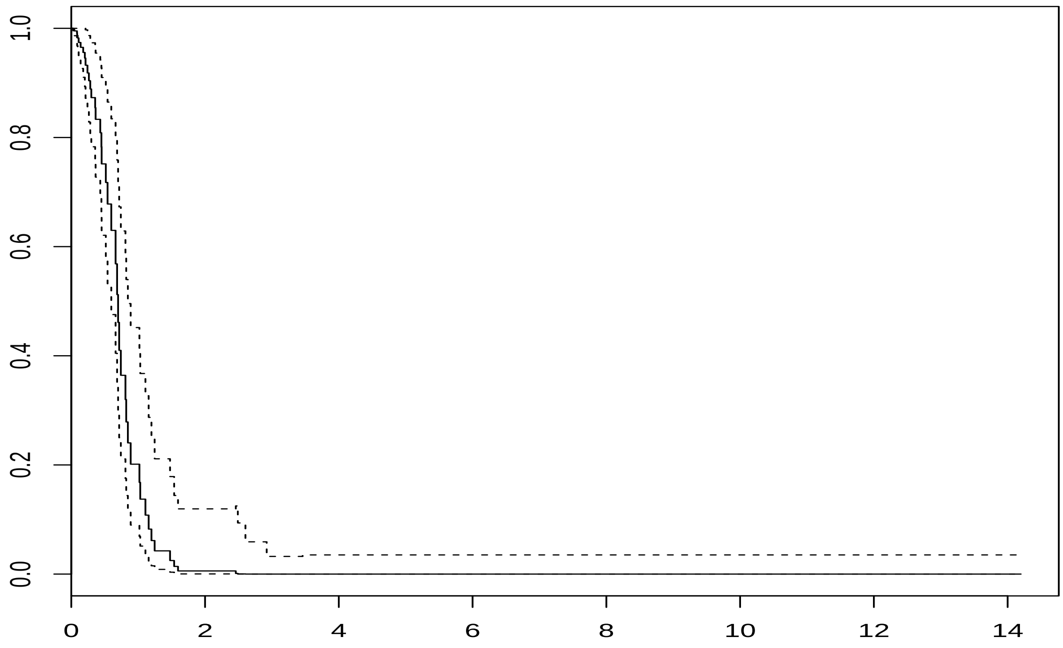

Figure 9: In this figure, the baseline survival function is illustrated.

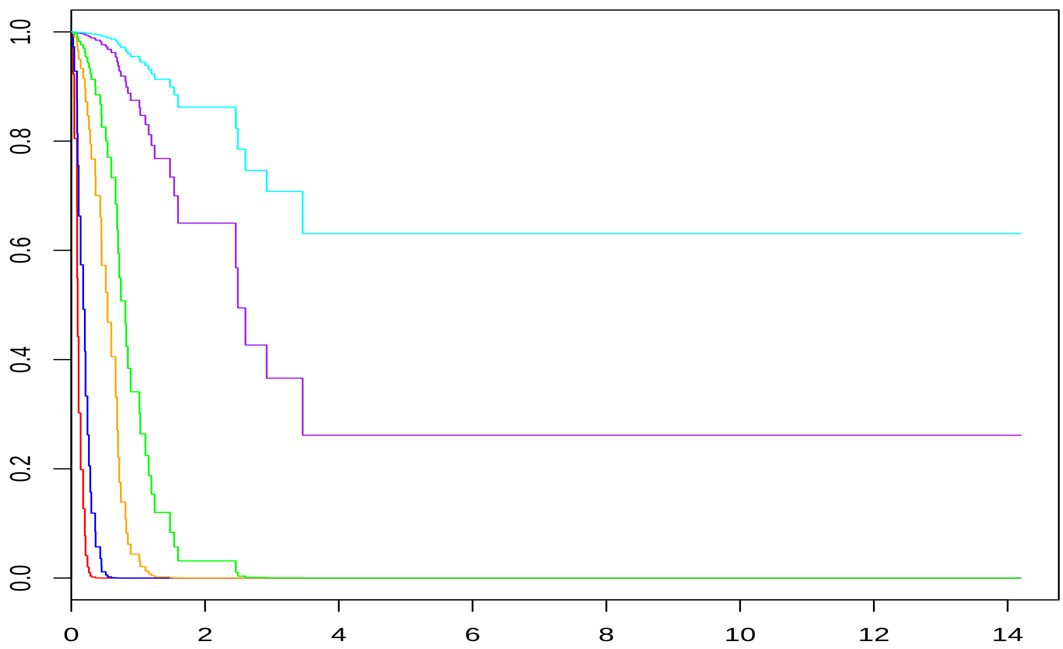

Figure 10: This plot represents the estimates of the survival functions for subjects within each group defined by x_2 and values of x_1 equal to 0.2, 1 and 5.

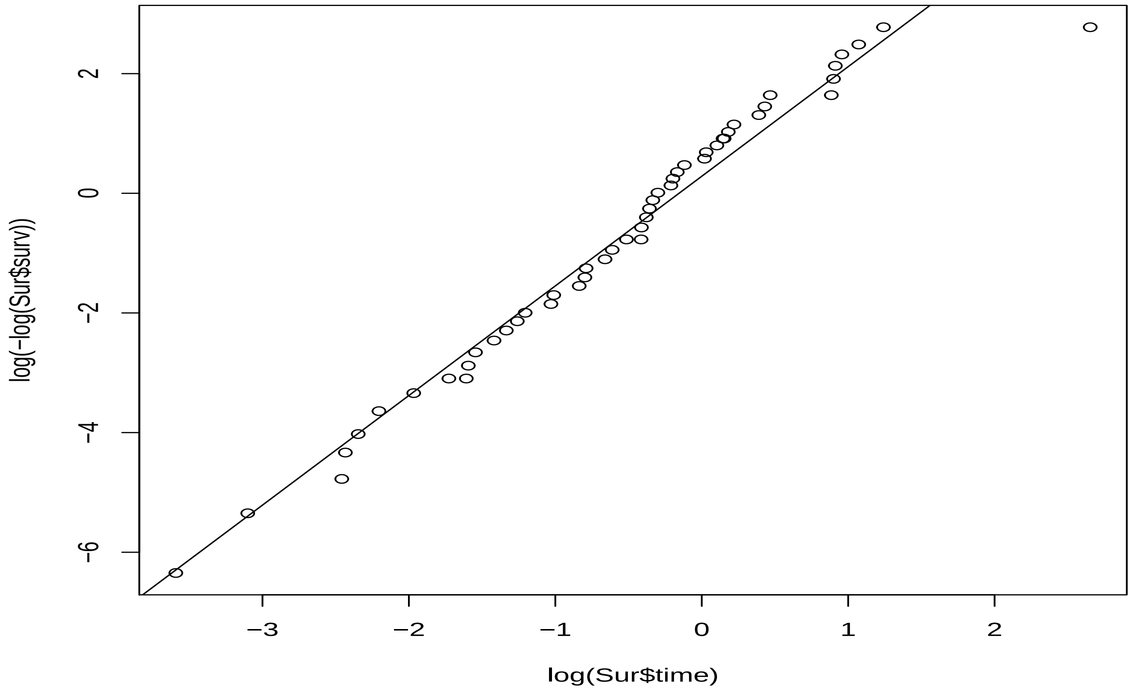

Figure 11: The plot illustrates the log(-log) transformation of the survival function against the log of time, which may be used to assess whether the model agrees with a Weibull model.

From figure 11 we may then see that the baseline survival approximately follows a straight line, and conclude that the estimated baseline hazard agrees with the Weibull model.

8 Problem 7

8.1 What was expected

Get the standard errors and comment.

8.2 Solution proposal

It was now of interest to find the standard errors related to both β_1 and β_2 we use the following method:

```
# method
weibtest= survreg(Surv(data$y,data$delta)~data$x2+log(data$x1), dist="weibull")
grad = matrix(rep(0),3,4)
grad[1,1] =-1/weibtest$scale
grad[2,2] =-1/weibtest$scale
grad[3,3] =-1/weibtest$scale
grad[1,4] =1/weibtest$scale*weibtest$coefficients[1]
grad[2,4] =1/weibtest$scale*weibtest$coefficients[2]
grad[3,4] =1/weibtest$scale*weibtest$coefficients[3]
#(grad)%*%vcov(weibtest)%*%t(grad)
```

```
sqrt(diag((grad)%*%vcov(weibtest)%*%t(grad)))[2:3]

## [1] 0.3365288 0.3242496
```

From this we found the values 0.32426 and 0.33653 for the standard error for $\log(x_1)$ and x_2, respectively. Furthermore, the standard errors from the semi-parametric Cox model were 0.323 and 0.351 for the log of x_1 and x_2, respectively. It appears that the standard errors for β_1 are similar while the standards errors for β_2 are marginaly different, the weibull being slightly smaller than the semi parametric cox model. Intuition would suggest that since the model is weibull the standard deviations for the weibull case should be somehow smaller, but in this case it is not extacly how it went down.

9 Problem 8

$$M_{\ln T|x_2}(t) = E(e^{t \ln T|x_2}) \tag{3}$$
$$M_{\ln T|x_2}(t) = E_{\ln x_1}\Big(E(e^{t \ln T|x_2}|\ln x_1)\Big) \tag{4}$$
$$M_{\ln T|x_2}(t) = E_{\ln x_1}\Big(\Gamma(1+\sigma t) \cdot e^{-(\beta_0+\beta_1 \ln x_1+\beta_2 x_2)t}\Big) \tag{5}$$
$$M_{\ln T|x_2}(t) = \Gamma(1+\sigma t) e^{-(\beta_0+\beta_2 x_2)} E(e^{-(t\beta_1 \ln x_1)}), \tag{6}$$
$$M_{\ln T|x_2}(t) = \Gamma(1+\sigma t) e^{-(\beta_0+\beta_2 x_2)} e^{\frac{t^2}{2}}, \tag{7}$$

which is not Gumbel.

10 Problem 9

It seems quite clear that neither weibull nor the semi parametric cox model would be adapted to such a problem, as it has been demonstrated throughout this whole project. Question 9 however might be used for that matter. If one assume that the hidden variable is lognormal for instance, one would get the distribution that was obtained in question 8. Then fitting the survreg with this distribution would possibly get good results, provided that the assumption made about the hidden variable is correct.