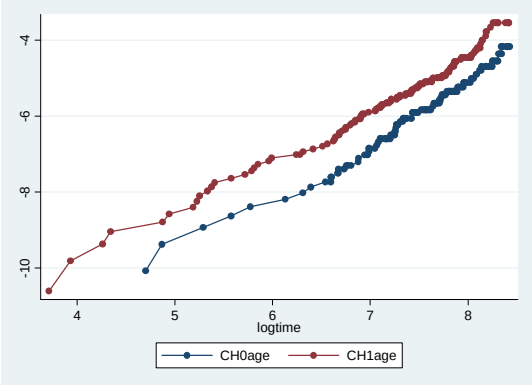
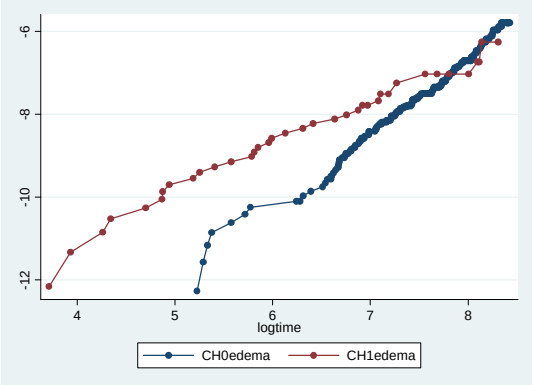
Biost/Epi 537
Winter 2008
Professor Breslow
HOMEWORK #6 key

Question 1:

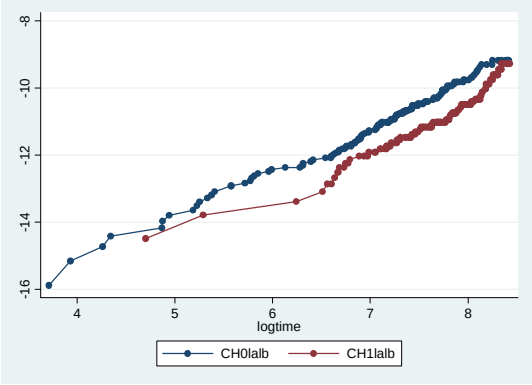
Age



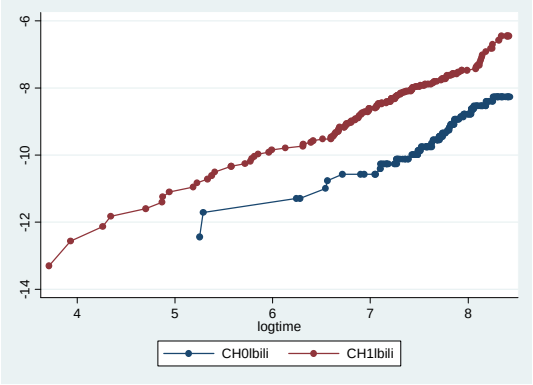
Edema



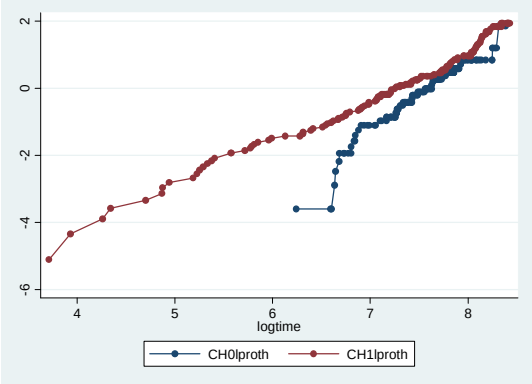
log-transformed bilirubin



Log-transformed ablbumin



Log-transformed prothrombin time



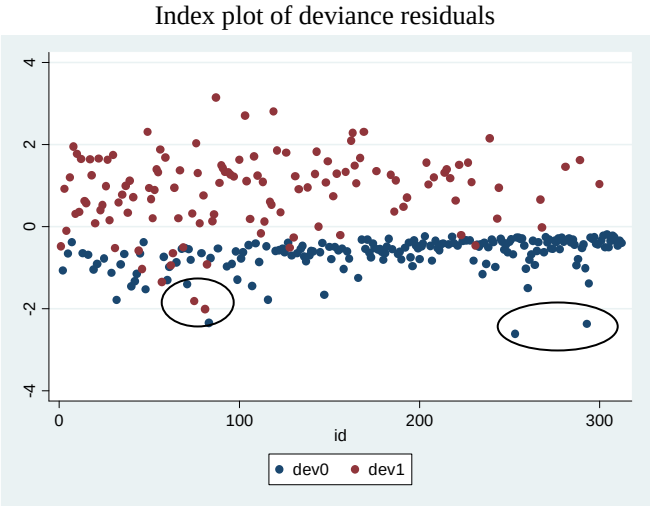
Age, log-transformed albumin, and log-transformed bilirubin are generally parallel. Edema and log-transformed prothrombin time are not parallel.

Question 2.
Based on the individual p-values for covariates, log(prothrombin time) (p-value, 0.023) and edema (p-value, 0.002) seem to violate proportional hazard assumptions. Based on global test, there is strong evidence of non-proportionality for the Mayo model (p-value, 0.001).

Time: log(t)

Variable	Rho	Chi2	Df	Prob > chi2
Age	-0.08756	0.82	1	0.3655
Edema	-0.28655	9.27	1	0.0023
Log(bilirubin)	0.1150	1.42	1	0.2335
Log(prothrombin time)	-0.22924	1.42	1	0.0232
Log(albumin)	-0.02154	5.15	1	0.8049
Global test		20.55	5	0.0010

Question 3.



I identified individuals 75, 81, 253, and 293 as outliers. Estimated coefficients are changed much (> 10%) when we refit Cox regression model without there 4 observations (Table 2). It indicates that these 4 observations are influential on coefficients' estimates. (But if we did not divide them by failure, it seems to be hard to say that there are outliners just based on the graph.)

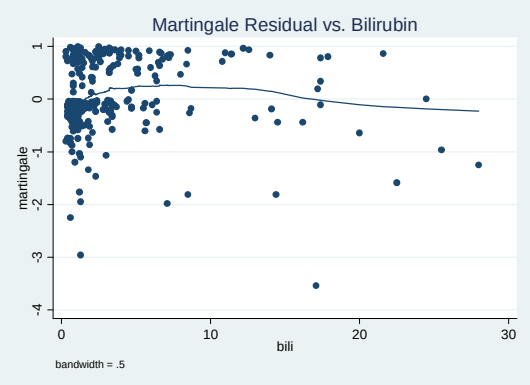
Table 1. Estimated coefficients including all subjects

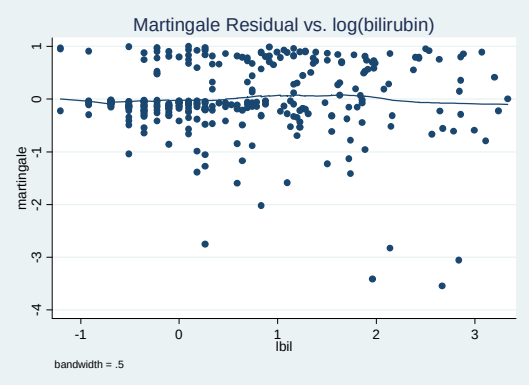
Variable	Coefficient	S.E.	P-value
Age	0.033	0.009	<0.001
Edema	0.484	0.237	0.041
Log(bilirubin)	0.902	0.098	<0.001
Log(prothrombin time)	3.184	1.007	0.002
Log(albumin)	-3.097	0.723	<0.001

Table 2. Estimated coefficients excluding subjects of 75 and 81.

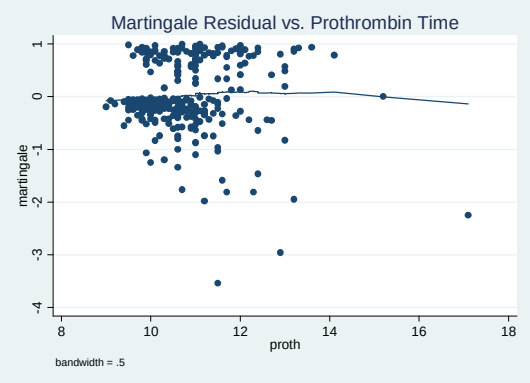
Variable	Coefficient	S.E.	P-value
Age	0.044	0.01	<0.001
Edema	0.572	0.237	0.016
Log(bilirubin)	1.068	0.106	<0.001
Log(prothrombin time)	3.354	0.961	<0.001
Log(albumin)	-3.289	0.695	<0.001

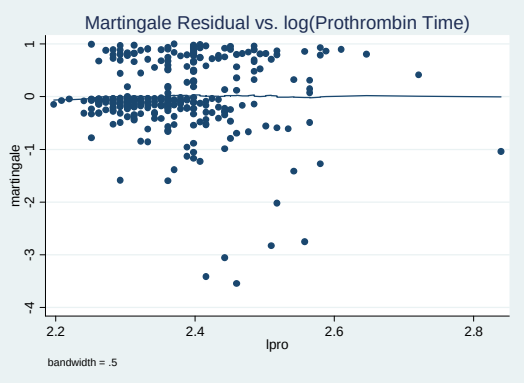
Question 4.





Based on smoothed martingale residual versus bilirubin plot, we could see that the smoothed curve has curvature. It indicates that different functional form of bilirubin is appropriate in the Mayo model. We also looked at martigale residual versus log-transformed bilirubin plot, we could see that the smoothed curve is almost flat. It says that log-transformed of the variable is better than the linear form in the Mayo model. The fracpoly result also supports this result.





Second plot is a smoothed martingale residuals versus prothombin time. The smoothed curve is not flat, which means that using different form of the variable is better than linear form. Therefore we also looked at martingale residuals versus log-transformed prothrombin time plot, which shows almost flat lowess curve. It indicates that log-transformed prothrombin time is better than linear form. Also the fracpoly result provides the same result.

```
. fracpoly stcox lbil lpro lalb edema age, degree(4) compare
-> gen double Ilpro__1 = lpro-2.368609425 if e(sample)
-> gen double Ilalb__1 = lalb-1.250795902 if e(sample)
-> gen double Iage__1 = age-50.01900672 if e(sample)
-> gen double Ilbil__1 = X^-2-.3136816726 if e(sample)
-> gen double Ilbil__2 = X^-2*ln(X)-.1818375938 if e(sample)
-> gen double Ilbil__3 = X^-1-.5600729172 if e(sample)
-> gen double Ilbil__4 = X^2-3.187945256 if e(sample)
   (where: X = (lbil+1.209803819656372))

Fractional polynomial model comparisons:
------------------------------------------------------------------
lbil            df       Deviance      Gain   P(term) Powers
------------------------------------------------------------------
Not in model     0       1166.578       --     --
```

Linear	**1**	**1083.401**	**0.000**	**0.000**	**1**
m = 1	2	1083.401	0.000	1.000	1
m = 2	4	1075.288	8.113	0.017	0 0
m = 3	6	1074.339	9.062	0.622	-2 -2 0
m = 4	8	1073.796	9.605	0.762	-2 -2 -1 2

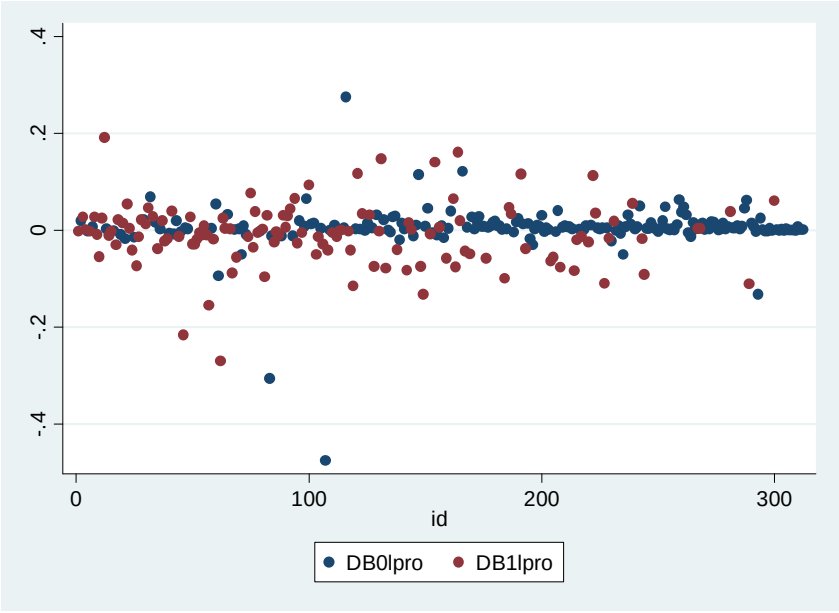
```
. fracpoly stcox lpro lbil lalb edema age, degree(4) compare
-> gen double Ilbil__1 = lbil-.5756779798 if e(sample)
-> gen double Ilalb__1 = lalb-1.250795902 if e(sample)
-> gen double Iage__1 = age-50.01900672 if e(sample)
-> gen double Ilpro__1 = lpro^-1-.4221886435 if e(sample)
-> gen double Ilpro__2 = lpro^-1*ln(lpro)-.3640545517 if e(sample)
-> gen double Ilpro__3 = lpro^.5-1.539028728 if e(sample)
-> gen double Ilpro__4 = lpro^.5*ln(lpro)-1.327109154 if e(sample)
```

Fractional polynomial model comparisons:

lpro	df	Deviance	Gain	P(term)	Powers
Not in model	0	1092.420	--	--	
Linear	**1**	**1083.401**	**0.000**	**0.003**	**1**
m = 1	2	1082.998	0.403	0.526	-2
m = 2	4	1082.678	0.724	0.852	3 3
m = 3	6	1082.617	0.785	0.970	3 3 3
m = 4	8	1041.830	41.572	0.000	-1 -1 .5 .5

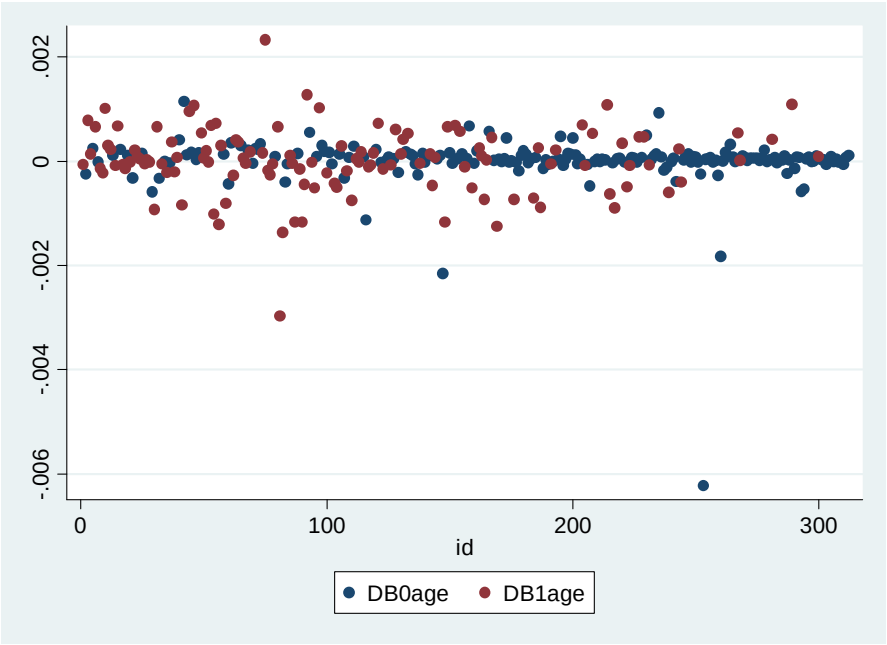
Problem 5.
A number of subjects appear to have large delta beta values for the estimates of the following variables, particularly in relation to the magnitude of the coefficient estimates.

Index Delta-Beta residuals vs Id



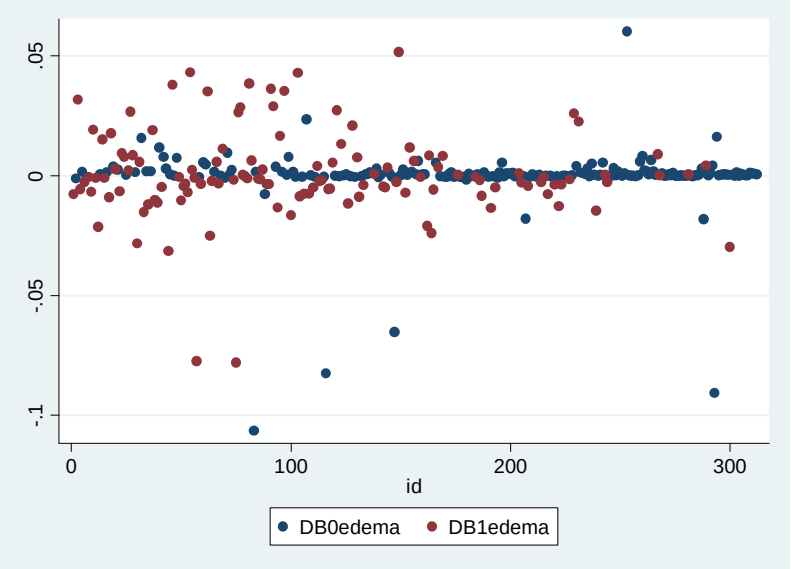
Looking at this graph, 4 individuals jump out at me. Id 116 had greater than 0.2 and the rest of them had less than -0.2 delta-beta values.

Id	Age	Log(bili)	Log(prothrombin)	Log(alb)	Edema
116	61.07	1.10	2.29	1.29	1
46	45.80	1.74	2.54	1.04	0
62	60.71	0.26	2.58	1.01	0
83	56.01	0.26	2.56	1.25	1
107	62.52	-0.51	2.84	1.39	0



The following is the list of that subject's covariates' values.

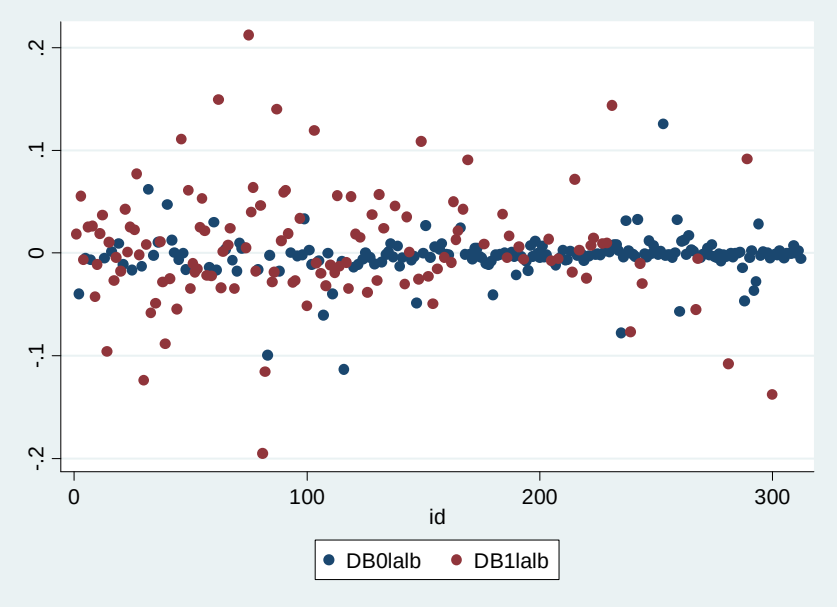
Id	Age	Log(bili)	Log(prothrombin)	Log(alb)	Edema
253	78.44	1.96	2.42	1.11	0



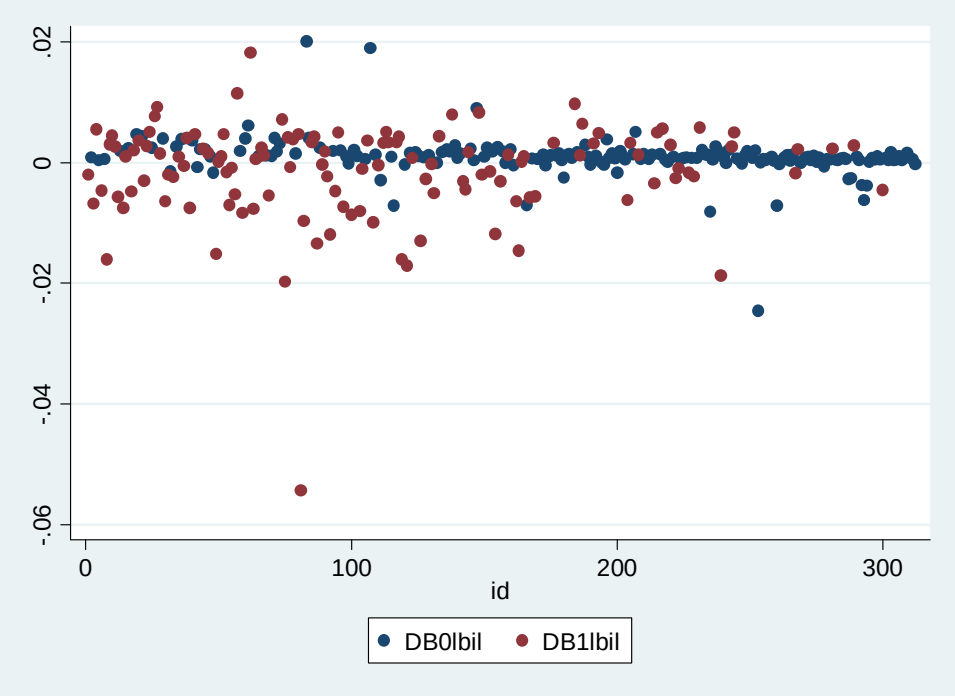
Looking at this graph, the 8 subjects with delta beta values greater than 0.05 and less than -0.05 are listed in the following table.

Id	Age	Log(bili)	Log(prothrombin)	Log(alb)	Edema

83	56.01	0.26	2.56	1.25	1
116	61.07	1.10	2.29	1.29	1
147	75.01	0.18	2.37	1.21	1
293	57.20	2.14	2.51	1.09	1
57	53.57	0.83	2.52	1.16	1
75	43.52	2.84	2.44	0.93	1
149	61.80	1.20	2.29	1.33	1
253	78.44	1.96	2.42	1.11	0

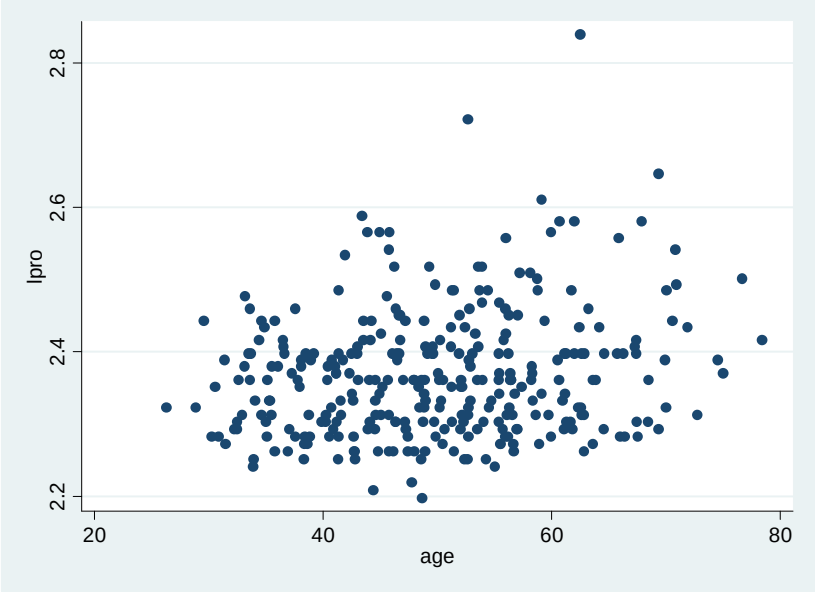


Id	Age	Log(bili)	Log(prothrombin)	Log(alb)	Edema
75	43.52	2.84	2.44	0.93	1



Id	Age	Log(bili)	Log(prothrombin)	Log(alb)	Edema
81	63.26	2.67	2.46	1.29	0

Several subjects appear to have large delta beta values for variables. They were 75, 107, 81, 253, and 116.



Id	Age	Log(bili)	Log(prothrombin)	Log(alb)	Edema
107	62.52	-0.51	2.84	1.39	0
191	52.69	3.20	2.72	1.21	0

Based on this graph, there were two outliners having large log(prothrombin time) with middle age. They were 107 and 191. An individual 107 was also identified as an outliner based on the index delta-beta plot. It indicates that this person has a big influence on coefficient estimate.