

Lecture eleven: Model Checking (II)

Assessment of model fit: Plots based on residuals

Different plots can be generated from the different type of residuals, typical plots are, residuals vs index, or rank of survival time; residuals vs covariates which are in or not in the model (martingale residuals). Sometime we can use separate plotting symbols for censored and uncensored observation, or for different levels of a particular factor.

1. Plots of Cox-Snell residuals

- (a) Index plots: residuals versus observation numbers
- (b) Treat (r_{Ci}, δ_i) as variates from Exponential(1). Recall the survival function of Weibull distribution is: $S(t) = \exp(-\lambda t^\gamma)$ and

$$\log(-\log S(t)) = \log(\lambda) + \gamma \log(t),$$

Thus, plot of $\log(-\log \hat{S}(r_{Ci}))$ vs $\log(r_{Ci})$ should follow the 45° line with 0 intercept if the Cox model fits the data.

- (c) When a model includes factors, the points on the plot corresponding to the different levels should overlap (blend well).
- (d) The plot does not indicate the type of departure from the model detected when plot is not linear.
- (e) Example 4.2: infection in patients on dialysis

i. SAS output

Obs	resc	SURVIVAL	logrc	logH
1	0.07847	0.91667	-2.54509	-2.44172
2	0.09385	0.83333	-2.36603	-1.70198
3	0.26553	0.75000	-1.32604	-1.24590
4	0.32857	0.65625	-1.11302	-0.86462
5	0.53856	0.56250	-0.61887	-0.55275
6	1.42071	0.46875	0.35116	-0.27749
7	1.42341	0.37500	0.35306	-0.01936
8	1.43313	0.28125	0.35986	0.23784
9	1.65233	0.18750	0.50218	0.51520
10	1.77362	0.09375	0.57302	0.86168
11	2.3927	0.8724	0.00000	.

ii. SAS code:

Available at [http://bfox.cwru.edu/~pxf/teaching/435.html\(dial2.sas\)](http://bfox.cwru.edu/~pxf/teaching/435.html(dial2.sas)).

iii. **Caution:** Cox-Snell residuals (R_{Ci}) have an approximate Exponential(1), but the approximation is not very reliable because the estimates of β 's and $H_0(t)$ are needed in computing of r_{Ci} . Counter Example: Lagakos (Biometrika, 68, pp. 93-98, 1981).

2. Plots of Martingale residuals

- (a) Martingale residuals are estimates of the excess number of events seen in the data, but not predicted by the model.
- (b) Determine the functional form of a covariate (ref: Biometrika, vol. 77: 147-160, 1990)

$$H(t, x^*, x_1) = H_0(t) \exp(\beta' x^*) \exp(f(x_1)),$$

- i. Plot martingale residuals vs x_1 .
- ii. The smoothed fitted curve (for example, lowess smoother) gives an indication of the function f .
- iii. If the curve is linear: no transformation; If there appears to be a threshold, then discretize the covariate.
- (c) Even if the model is good fit of data, the martingale residuals are skewed (not symmetric).

3. Plots of Schoenfeld residuals

- (a) Useful in assessing time trends or lack of proportionality in one of the coefficients of the model.
- (b) Continuous covariates: plot Schoenfeld residuals vs time, if PH holds, the residuals randomly fluctuating around zero, and gradually converge to zero with time.
- (c) Categorical covariates or a categorized continuous covariates: Using Estimated KM survival curves, plot of $\log(-\log(\hat{S}_i(t)))$ vs $\log(t)$ should be parallel.
- (d) Kolmogorov type test using the cumulative sum of the residuals (Therneau et al, Biometrika, vol. 81: 515-526, 1994): In Splus, the PH assumption can be tested by `cox.zph()` function and `plot(cox.zph())`. We will talk more about it in section 4.4.

4. Plots of score residuals and deviance residuals

5. Example 4.3: Infection in patients on dialysis (cont.)

- (a) figures
- (b) SAS and Splus programs:

Posted at the course website (`ex4_3.sas` and `ex43.s`).

6. Example 4.4: Infection in patients on dialysis (cont.)

- (a) Plot of the martingale residuals for the null model against age
- (b) SAS and Splus programs:

Posted at the course website (`ex4_4.sas` and `ex44.s`).

7. Example 4.5: Survival of multiple myeloma patients

- (a) Figures
- (b) BUN or LBUN ($\log(BUN)$)?
 - i. Determine the functional form: plots of martingale residuals vs BUN and LBUN
 - ii. $-2\log L$, the smaller the better.

Table 1: Results from Cox models for multiple myeloma data

model	covariates	$-2\log L$
Model 0	None	215.94
Model 1	HB, LBUN	208.175
Model 2	HB, BUN	202.938

- (c) SAS and Splus code

SAS code:

Available at <http://bfox.cwru.edu/~pxf/teaching/435.html> (`ex4_5.sas`).

Splus code:

Available at <http://bfox.cwru.edu/~pxf/teaching/435.html> (`ex45.s`).

Assignment six: Assume the final Cox model for prostatic cancer patients study (Table 1.4, page 10) includes **SIZE, INDEX and TREAT** (see example 3.6 at page 73). Do model diagnosis by checking the residuals including Cox-Snell, martingale and Schoenfeld residuals). Write up your comments and interpretation. Provide tables as well as residual plots to justify your answer.

Cumulative hazard plot of the Cox – Snell residuals

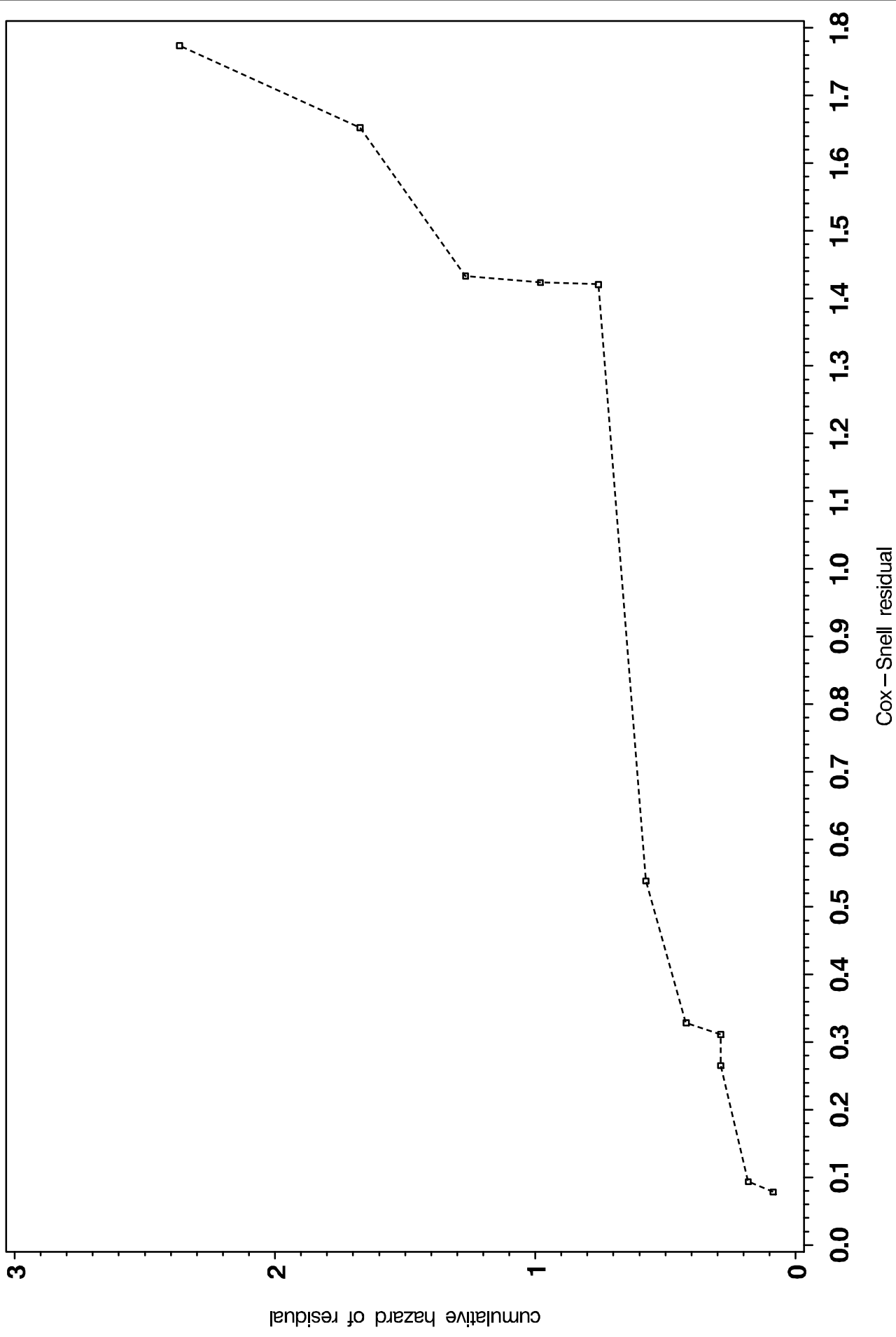


Figure 2: Example 4.3

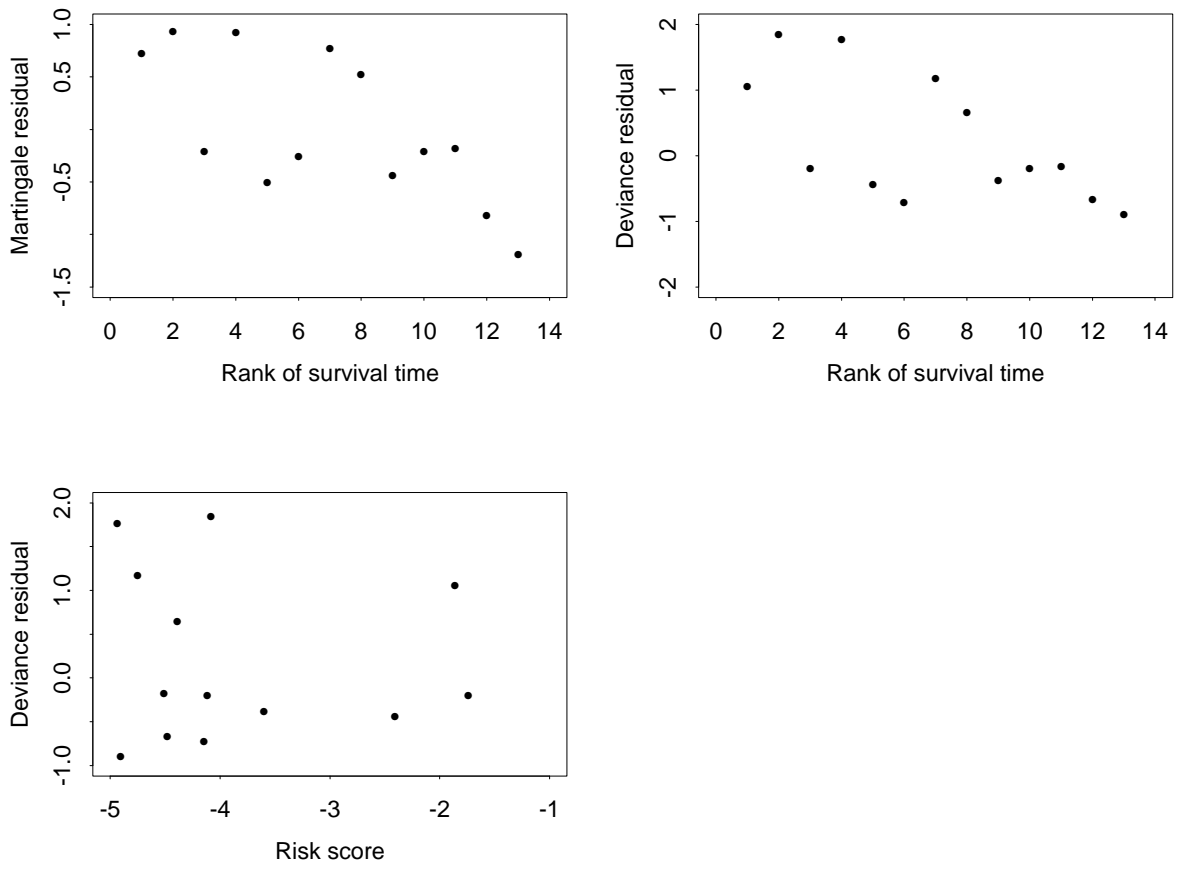
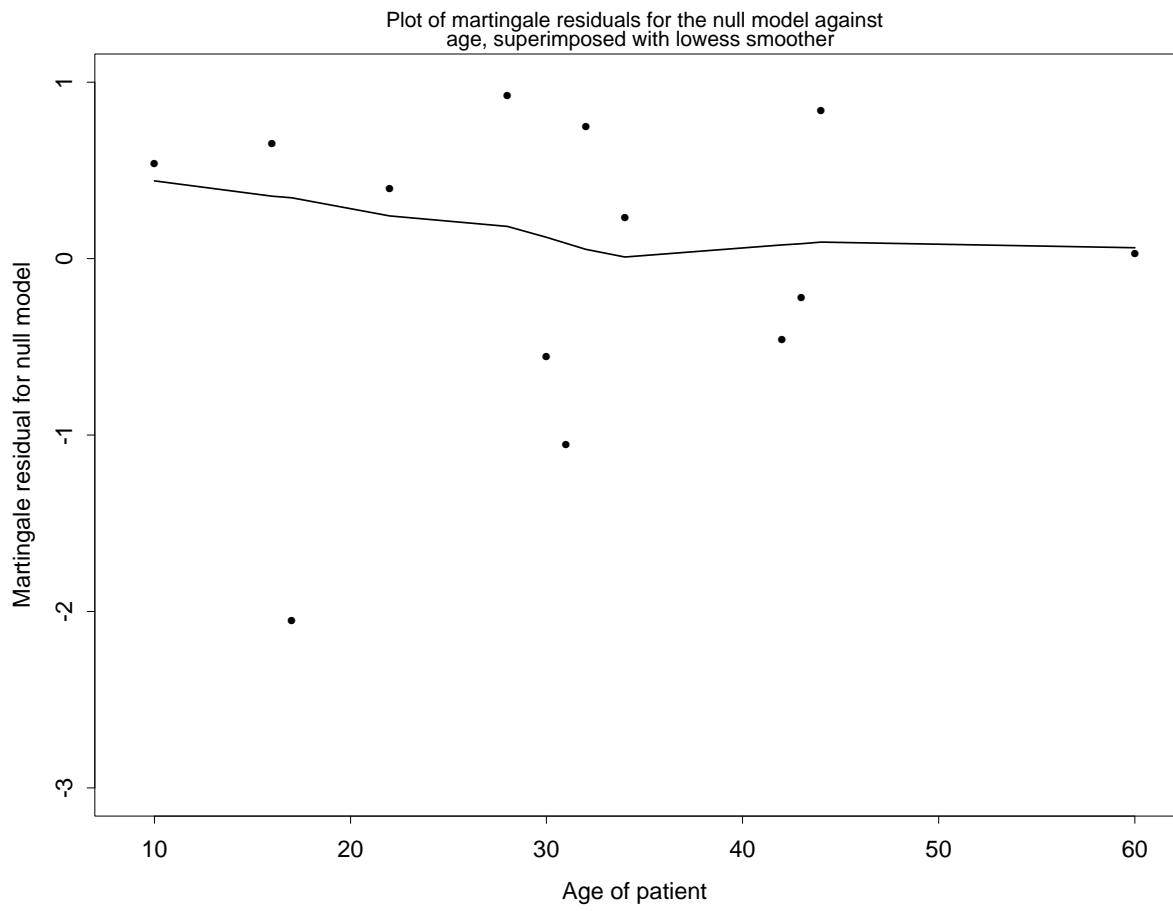


Figure 3: Example 4.4



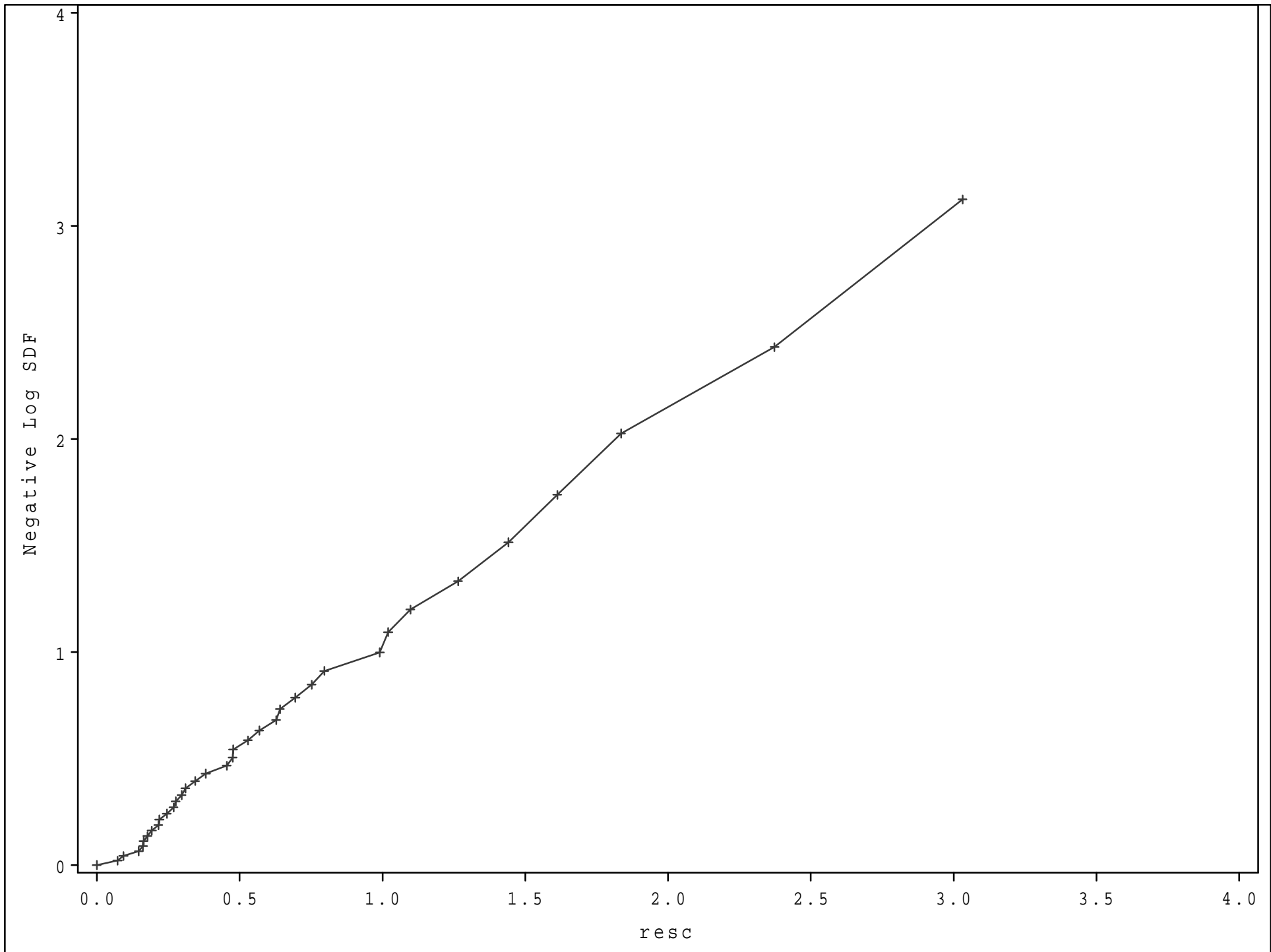


Figure 5: Example 4.5

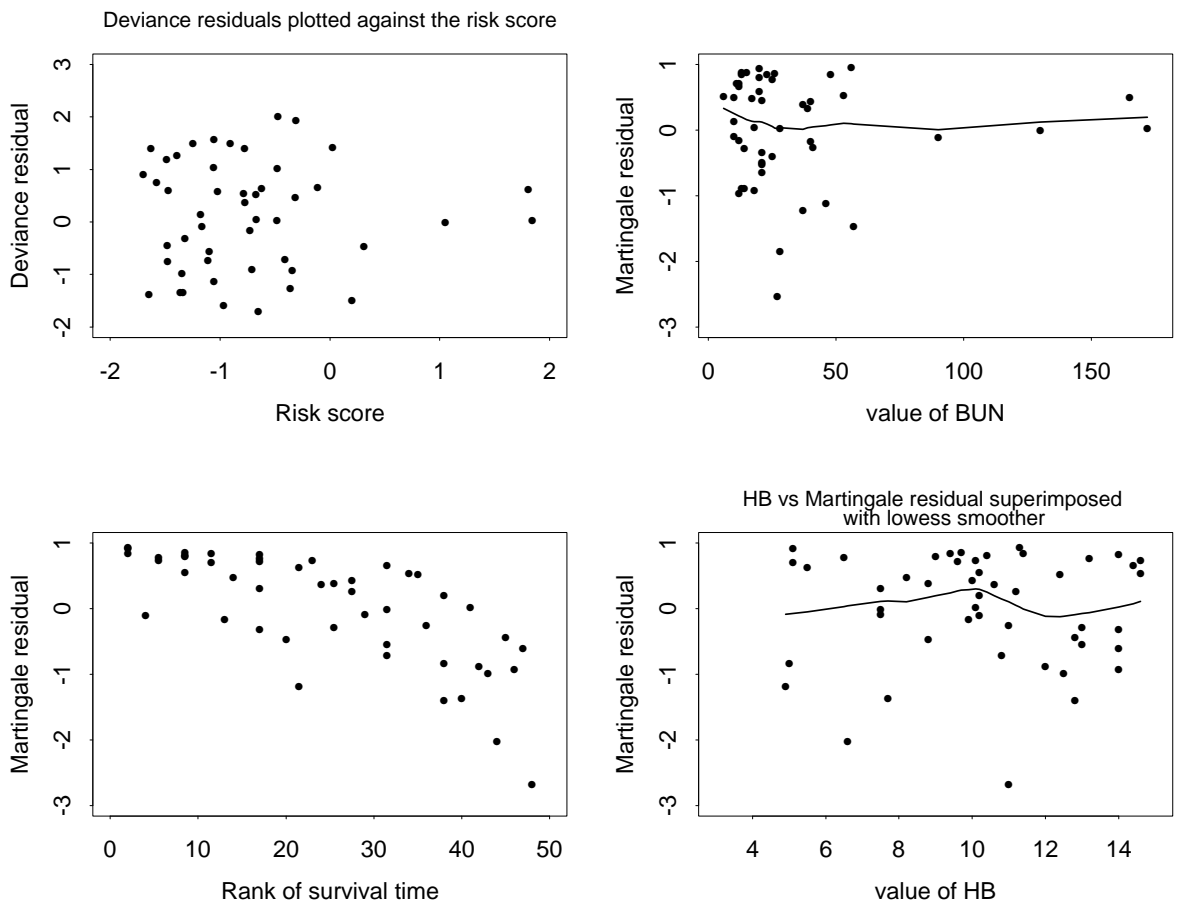


Figure 6: Example 4.5

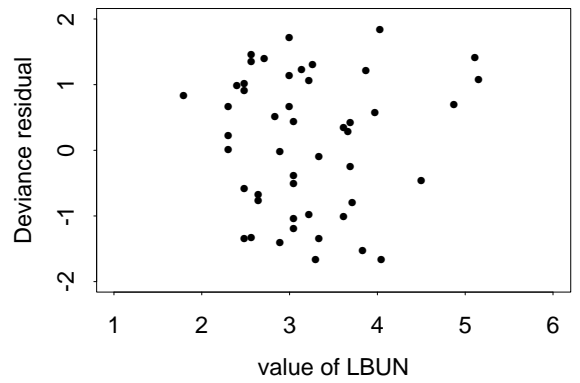
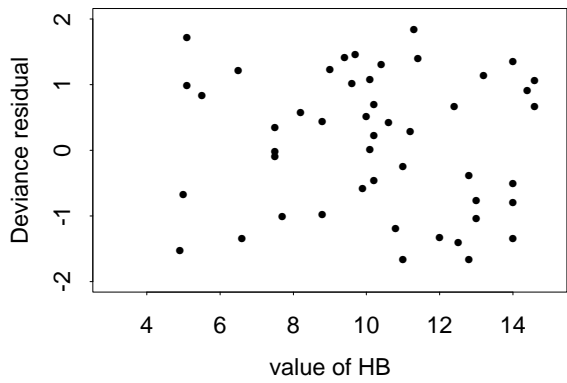
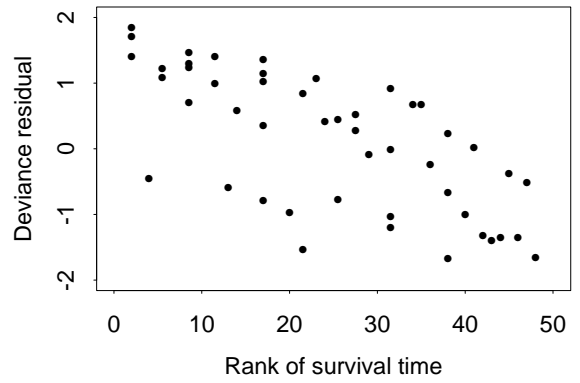
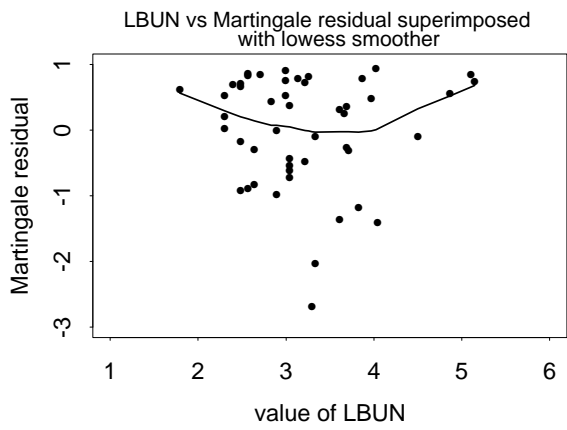


Figure 7: Example 4.5

