Residuals and model diagnostics

Patrick Breheny

November 6

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Introduction

- Many assumptions go into regression models, and the Cox proportional hazards model, despite making no assumptions about the baseline hazard, is no exception

- Diagnostic methods are useful in all types of regression models to investigate the validity of those assumptions and identify ways in which they might be violated

- Residuals play a big role in regression method diagnostics

- To build model diagnostics for Cox regression, we first need to discuss methods for extending residuals to the case of censored data

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Cumulative hazard transformation

- We begin with the following useful theorem:
- **Theorem:** Suppose T is a continuous nonnegative random variable with cumulative hazard function Λ. Then the random variable $Y = \Lambda(T)$ follows an exponential distribution with rate $\lambda = 1$.
- Thus, one way of checking the validity of a model is by comparing the model's estimates $\{\hat{\Lambda}(t_i)\}$ against the standard exponential distribution

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Cox-Snell residuals

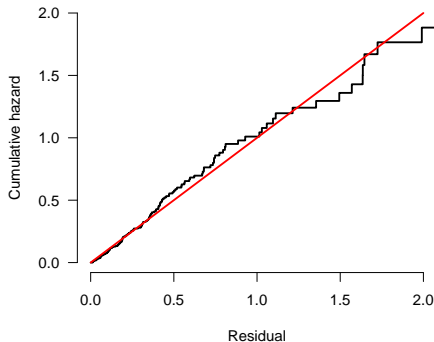- In the context of the proportional hazards model, we have

$$\hat{e}_i = \hat{\Lambda}_0(t_i)\exp(\mathbf{x}_i^T\widehat{\boldsymbol{\beta}}),$$

although the idea is very general and can be applied to any kind of model

- The terms $\{\hat{e}_i\}$ are called the *Cox-Snell residuals*, although "residual" is perhaps a misnomer, in that it's more of a transformation than a residual

Residuals | **Cox-Snell residuals**
Assessing the functional form of a covariate | Martingale residuals
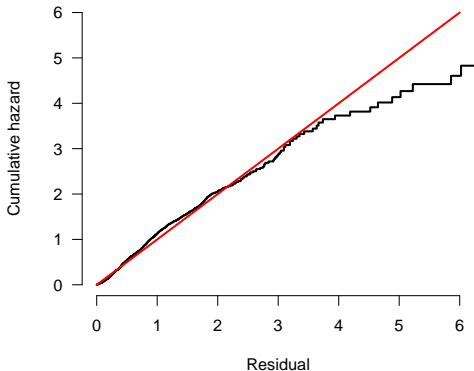Assessing influence | Deviance residuals

Diagnostic plot of Cox-Snell residuals: PBC data

- Diagnostics based on Cox-Snell residuals are based on fitting a Kaplan-Meier (or Nelson-Aalen) curve to $\{\hat{e}_i\}$ and comparing it to that of the standard exponential
- For the PBC data with `trt`, `stage`, `hepato`, and `bili` included in the Cox model, we have
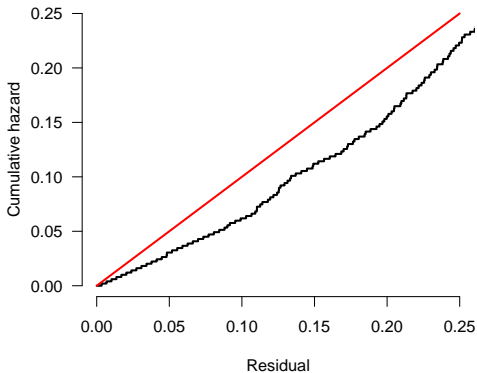
Residuals   **Cox-Snell residuals**
Assessing the functional form of a covariate   Martingale residuals
Assessing influence   Deviance residuals

Simulated example: Lack of fit

To illustrate the utility of the plot, let's simulate some data from a lognormal AFT model and fit a Cox model (note that the PH assumption is violated here):
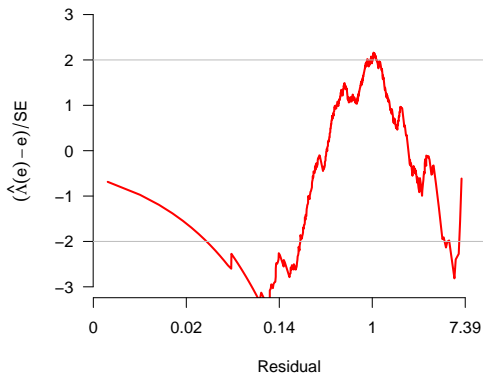
Residuals   Cox-Snell residuals
Assessing the functional form of a covariate   Martingale residuals
Assessing influence   Deviance residuals

Zooming in

One shortcoming of this plot is that violations at small e values tend to be hidden:

Residuals   **Cox-Snell residuals**
Assessing the functional form of a covariate   Martingale residuals
Assessing influence   Deviance residuals
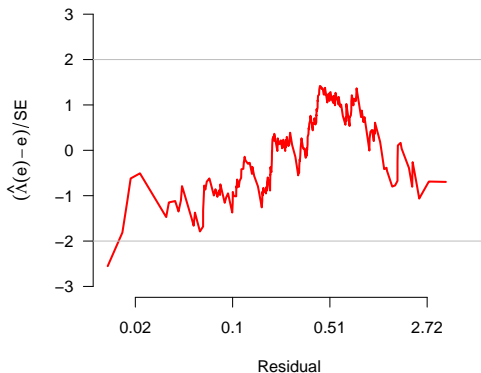
Plotting the standardized difference

Alternatively, one may consider plotting the standardized difference between the fitted cumulative hazard and the standard exponential, which reveals the lack of fit much more clearly:
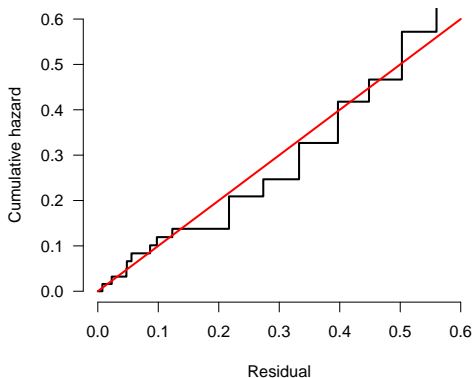
Residuals | Cox-Snell residuals
Assessing the functional form of a covariate | Martingale residuals
Assessing influence | Deviance residuals

Standardized differences: PBC data

Here is what the standardized plot looks like for the PBC data:

Residuals | Cox-Snell residuals
Assessing the functional form of a covariate | Martingale residuals
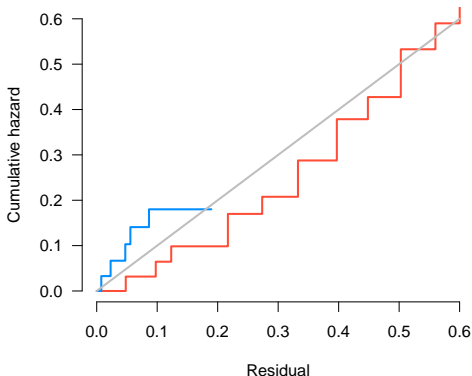Assessing influence | Deviance residuals

GVHD data

- Stratification of the residuals according to one of the variables in the model can also help to discover model violations
- For example, in the GVHD data, here is what the overall plot looks like:

Residuals   **Cox-Snell residuals**
Assessing the functional form of a covariate   Martingale residuals
Assessing influence   Deviance residuals

GVHD data: Stratifying by treatment

And here is the plot stratifying by treatment:

We will discuss stratification in more detail in the next lecture

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Shortcomings of the Cox-Snell residual

- One drawback to the Cox-Snell residuals is that they don't provide much insight into *why* the model's assumptions are violated

- It would be more appealing if each residual took on a positive or negative value indicating whether they patient survived longer, as opposed to shorter, that the model predicts, and by how much

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Martingale residuals

- Consider, then, the following residual:

$$\hat{m}_i = d_i - \hat{\Lambda}_i(t_i)$$
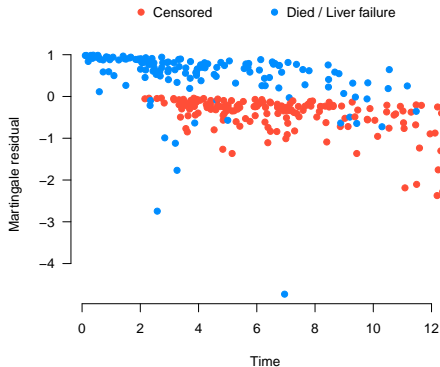$$= d_i - \hat{e}_i;$$

  we have essentially seen this quantity before (recall the score functions for exponential and Cox regression)

- This represents the discrepancy between the observed value of a subject's failure indicator and its expected value, integrated over the time for which that patient was at risk

- Positive values mean that the patient died sooner than expected (according to the model); negative values mean that the patient lived longer than expected (or were censored)

Residuals   Cox-Snell residuals
Assessing the functional form of a covariate   Martingale residuals
Assessing influence   Deviance residuals

Martingales

- A stochastic process $M(t)$ that satisfies (i) $\mathbb{E}M(t) = 0$ for all t and (ii) $\mathbb{E}\{M(t)|M(s)\} = M(s)$ for all $s < t$ is known in statistics as a *martingale*

- The stochastic process $N(t) - \int_0^t Y(t)d\Lambda(t)$, where $N(t)$ is the counting process that records whether the subject has failed by time t or not, satisfies these two properties

- For this reason, the residuals $\{\hat{m}_i\}$ defined on the previous slide are known as *martingale residuals*

Residuals | Cox-Snell residuals
Assessing the functional form of a covariate | Martingale residuals
Assessing influence | Deviance residuals

Martingale residuals for the PBC model

The large outlier is a patient with stage 4 cirrhosis and a bilirubin concentration of 14.4 (96th percentile), yet survived 7 years

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

residuals

- Martingale residuals can be obtained from the survival package by calling residuals(fit), where fit is a fitted coxph model (resid(fit) also works as a shortcut)

- The martingale residuals are returned by default, although seven other options are available and can be requested by specifying the type option

- It may be noted that type='coxsnell' is *not* one of the options, although you can easily calculate the Cox-Snell residuals from the martingale residuals

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Comments

- Martingale residuals are very useful and can be used for many of the usual purposes that we use residuals for in other models (identifying outliers, choosing a functional form for the covariate, etc.)

- However, the primary drawback to the martingale residual is its clear asymmetry (its upper bound is 1, but it has no lower bound)

- For this reason, I'll hold off on these plots until we discuss a more symmetric, normally distributed residual

Residuals | Cox-Snell residuals
Assessing the functional form of a covariate | Martingale residuals
Assessing influence | Deviance residuals

Deviance residuals: Motivation

- A technique for creating symmetric, normalized residuals that is widely used in generalized linear modeling is to construct a "deviance residual"

- The idea behind the deviance residual is to examine the difference between the log-likelihood for subject i under a given model and the maximum possible log-likelihood for that subject:

$$2(\tilde{\ell}_i - \ell_i),$$

in a sense, constructing a miniature likelihood ratio test for individual i

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Deviance residuals: Definition

- As it is essentially a likelihood ratio test, the quantity on the previous slide should approximately follow a χ_1^2 distribution

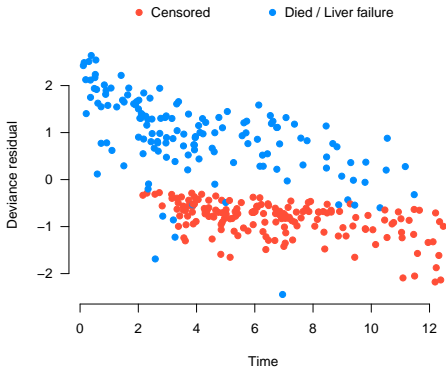- To turn it into a quantity that approximately follows a normal distribution, we can use

$$\hat{d}_i = \text{sign}(\hat{m}_i)\sqrt{2(\tilde{\ell}_i - \ell_i)};$$

  this is known as the *deviance residual*

- I am leaving the details of deriving $\tilde{\ell}_i$ and working out a simple expression for the deviance residual as homework

- In R: `residuals(fit, type='deviance')`

Residuals   Cox-Snell residuals
Assessing the functional form of a covariate   Martingale residuals
Assessing influence   Deviance residuals

Deviance residuals for the PBC data

The deviance residuals are much more symmetric:

Residuals   Cox-Snell residuals
Assessing the functional form of a covariate   Martingale residuals
Assessing influence   Deviance residuals

Outliers

- Deviance residuals have several uses that we will now illustrate

- One is identifying outliers

- For the PBC data, there are no extreme outliers; the largest residuals are only 2.5 SDs away from zero

- Note that the skewness of the martingale residual makes one subject look like an extreme outlier; according to the deviance residuals, however, that subject is only the 5th largest outlier by absolute value

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Outliers (cont'd)

The five largest negative residuals:

time	status	trt	stage	hepato	bili	$\hat{d}$
6.95	1	1	4	1	14.40	-2.44
12.19	0	2	4	1	2.10	-2.18
12.38	0	2	4	1	1.80	-2.14
11.09	0	1	4	1	1.30	-2.09
11.49	0	2	4	1	1.20	-2.05

And the five largest positive residuals:

time	status	trt	stage	hepato	bili	$\hat{d}$
0.38	1	1	3	0	2.40	2.64
0.54	1	1	3	0	1.10	2.54
0.30	1	2	4	1	2.50	2.51
0.14	1	2	4	0	12.60	2.49
0.11	1	1	4	0	17.90	2.43

Residuals
Assessing the functional form of a covariate
Assessing influence

Cox-Snell residuals
Martingale residuals
Deviance residuals

Residuals vs. covariates

- The other main use of residuals is to plot them against covariates to assess the relationship between a covariate and unexplained variation
- This can be done using covariates that are already in the model as well as new covariates that one is considering adding to the model; we will consider examples of both

Residuals | Cox-Snell residuals
Assessing the functional form of a covariate | Martingale residuals
Assessing influence | **Deviance residuals**

Residuals vs. albumin

Albumin is a protein synthesized by the liver and often used as a marker of liver disease:

Residuals
Assessing the functional form of a covariate
Assessing influence

Albumin
Bilirubin
Stage
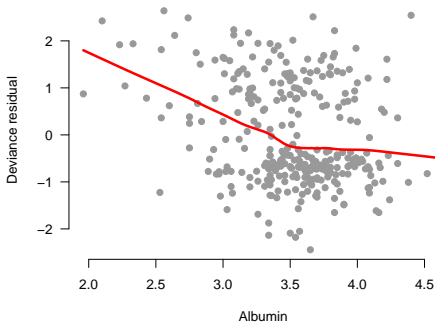
Assessing functional forms

- As the previous figure shows, deviance residuals are helpful not only for checking whether new variables should be added to a model, but also for assessing whether the relationship between the predictor and the (log) hazard is linear

- In the case of albumin, the exploratory plot, as well as biological insight (the normal range of serum albumin is 3.4-5.4 g/dL) suggest a piecewise linear model with a change point (sometimes called a "knot") at 3.4

Residuals
Assessing the functional form of a covariate
Assessing influence

Albumin
Bilirubin
Stage

Implementation details

- This can be implemented in various ways; here is one:

```
f <- function(x) {pmin(x, 3.5)}
fit <- coxph(S ~ ... + f(albumin), PBC)
```

- For nonlinear functional forms, it is typically helpful to plot the modeled relationship between a covariate and the outcome

- Again, this can be done in various ways, but the visreg package is useful here:

```
visreg(fit, 'albumin')
```

Residuals
Assessing the functional form of a covariate
Assessing influence

Albumin
Bilirubin
Stage

Changepoint model for albumin: Illustration

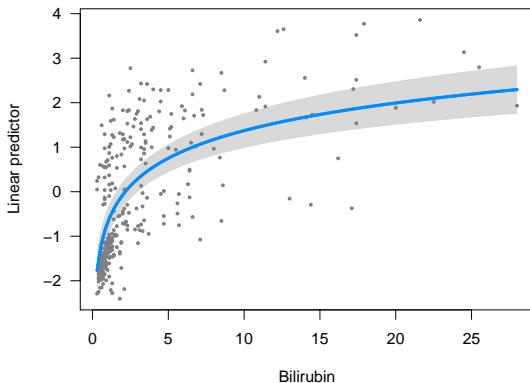The dots here are the partial deviance residuals, $\eta_i + \hat{d}_i$

Residuals
Assessing the functional form of a covariate
Assessing influence

Albumin
Bilirubin
Stage
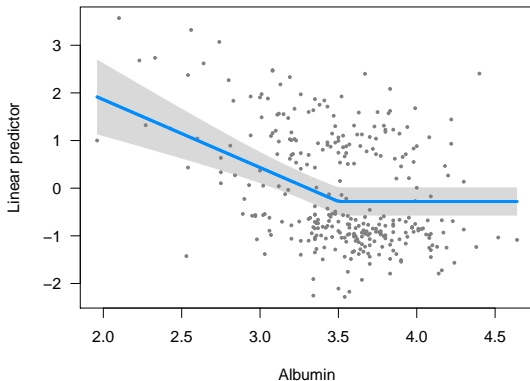
Albumin summary

- The AIC values for four possible models:
  - No albumin: 1352.3
  - Linear: 1334.9
  - Changepoint, no effect in normal range: 1332.2
  - Changepoint, linear effect in normal: 1333.8
- The changepoint model that assumes a flat line after 3.5 is the reasonably clear choice based on AIC, as well as being a very reasonable model from a biological perspective

Residuals   Albumin
Assessing the functional form of a covariate   Bilirubin
Assessing influence   Stage

Residuals for bilirubin: In and out of the model
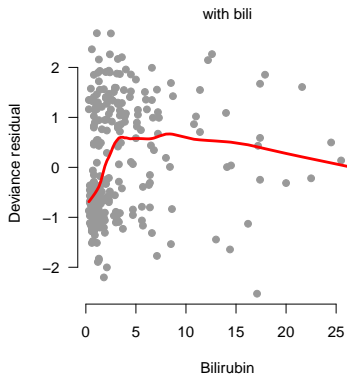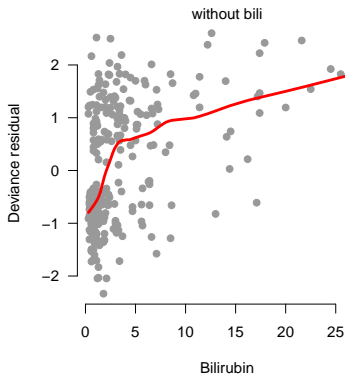
log(Bilirubin)

The residual plots on the previous page suggest `log(bili)` as a functional form:

Residuals   Albumin
Assessing the functional form of a covariate   **Bilirubin**
Assessing influence   Stage

More on logs

- One advantage of the log scale as a functional form is that it has a convenient interpretation in proportional hazards models

- Consider comparing two individuals, one with double the bilirubin concentration of the other, but all other covariates equal; the hazard ratio comparing the two is:

$$\mathrm{HR} = 2^{\widehat{\beta}_j}$$

- For `log(bili)`, $\widehat{\beta}_j = 0.895$; thus, doubling the bilirubin concentration increases the hazard by 86% ($2^{0.895} = 1.86$)

Residuals
Assessing the functional form of a covariate
Assessing influence
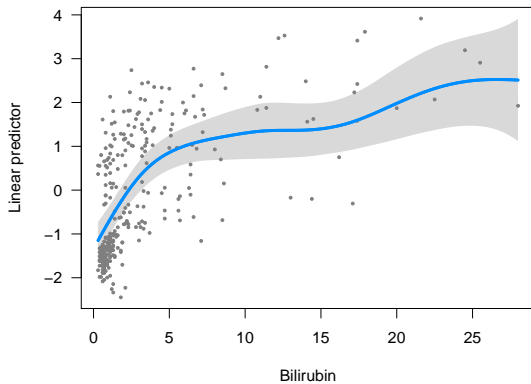
Albumin
Bilirubin
Stage

Splines

- *Splines* are also an attractive option for incorporating nonlinear functional forms

- The basic idea of splines is similar to our albumin model from earlier: fit a piecewise polynomial model, with restrictions that make the resulting functional form smooth and continuous

- The details of this are very interesting, but unfortunately we don't have time to get into them in this course

- Still, it is worth knowing that the survival package has a nice utility built in for fitting Cox models with spline terms:

```
fit <- coxph(S ~ ... + pspline(bili), PBC)
```

Residuals
Assessing the functional form of a covariate
Assessing influence

Albumin
Bilirubin
Stage

Spline illustration

Residuals
Assessing the functional form of a covariate
Assessing influence

Albumin
Bilirubin
Stage

Bilirubin: Summary
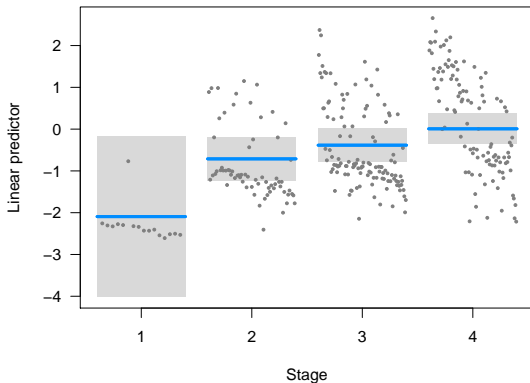
- The AIC values for four possible models:
  - No bilirubin: 1383.9
  - Linear: 1332.2
  - Log(bilirubin)   1301.0
  - Splines: 1303.9

- The log model seems to be a good choice here

Residuals   Albumin
Assessing the functional form of a covariate   Bilirubin
Assessing influence   **Stage**

Stage

Before moving on, let's quickly revisit the effect of stage:

Residuals
Assessing the functional form of a covariate
Assessing influence

Albumin
Bilirubin
Stage

Remarks

- The effect of stage seems to be very linear among stages 2, 3, and 4

- Stage 1 patients, however, seem to be at substantially lower risk

- Still, it is difficult to tell from the data alone whether this phenomenon is real, since we have few stage 1 patients in the study (just 16 out of 312 patients), which explains why AIC and the likelihood ratio test prefer the simpler model

Influence

- One final issue on the topic of regression diagnostics is the assessment of *influence*
- An influential measurement is one that has a large effect on the model fit; this can be measured in a variety of ways, both absolute and relative
- A variety of residuals (score residuals, Schoenfeld residuals, delta-beta residuals) have been proposed as ways of quantifying and assessing influence; we'll focus on delta-beta residuals
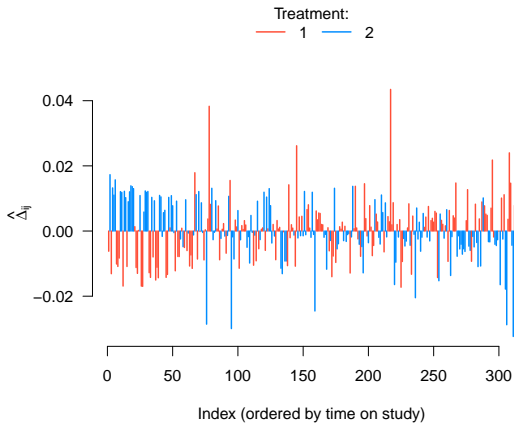
Delta-beta plots

- The idea behind delta-beta residuals is very simple: let $\widehat{\beta}_j^{(i)}$ denote the estimate of $\widehat{\beta}_j$ obtained if we leave subject i out of the model

- The delta-beta residual for coefficient j, subject i is therefore defined as

$$\hat{\Delta}_{ij} = \widehat{\beta}_j - \widehat{\beta}_j^{(i)}$$

- This might seem computer intensive, but there are various computational tricks that allow one to fairly quickly refit models leaving individual observations out

- In R: `resid(fit, type='dfbeta')` returns a *matrix* of delta-beta residuals
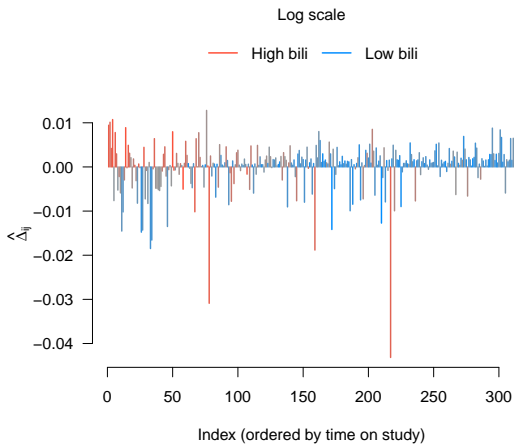
Treatment

Stage

Bilirubin

Log scale

High bili   Low bili
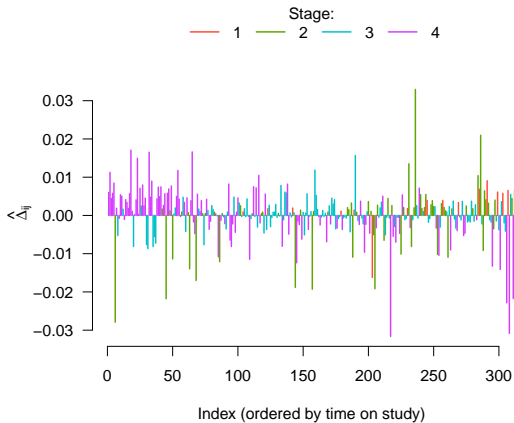
Remarks

- Delta-beta plots are always interesting to look at and offer a great deal of insight into the inner mechanics of complicated models

- Furthermore, they can indicate whether the estimation of a coefficient is dominated by just a few individuals, which would be clear cause for concern

- What action to take in the presence of influential observations is often a complicated decision; my advice, however, is to strongly prefer modifying the model to fit the data as opposed to manipulating the data to fit the model (i.e., by removing outliers)