

Using Residuals with Cox
Models

Terry M. Therneau
Mayo Clinic

August 1997

Overview

Residuals from a Cox model are now available from several packages.

What are their practical, every day use?

Most useful are

- Martingale: scatter plots for functional form
- Scaled Schoenfeld: understanding proportional hazards
- Dfbeta: leverage and robust variance

Martingale residuals

Consider the martingale residuals from a “no covariates” fit:

S-Plus

```
> fit0 <- coxph(Surv(time, status) ~ 1, data=mydata)
> rr <- resid(fit)
> plot(x, rr)
> lines(lowess(x, rr))
```

SAS

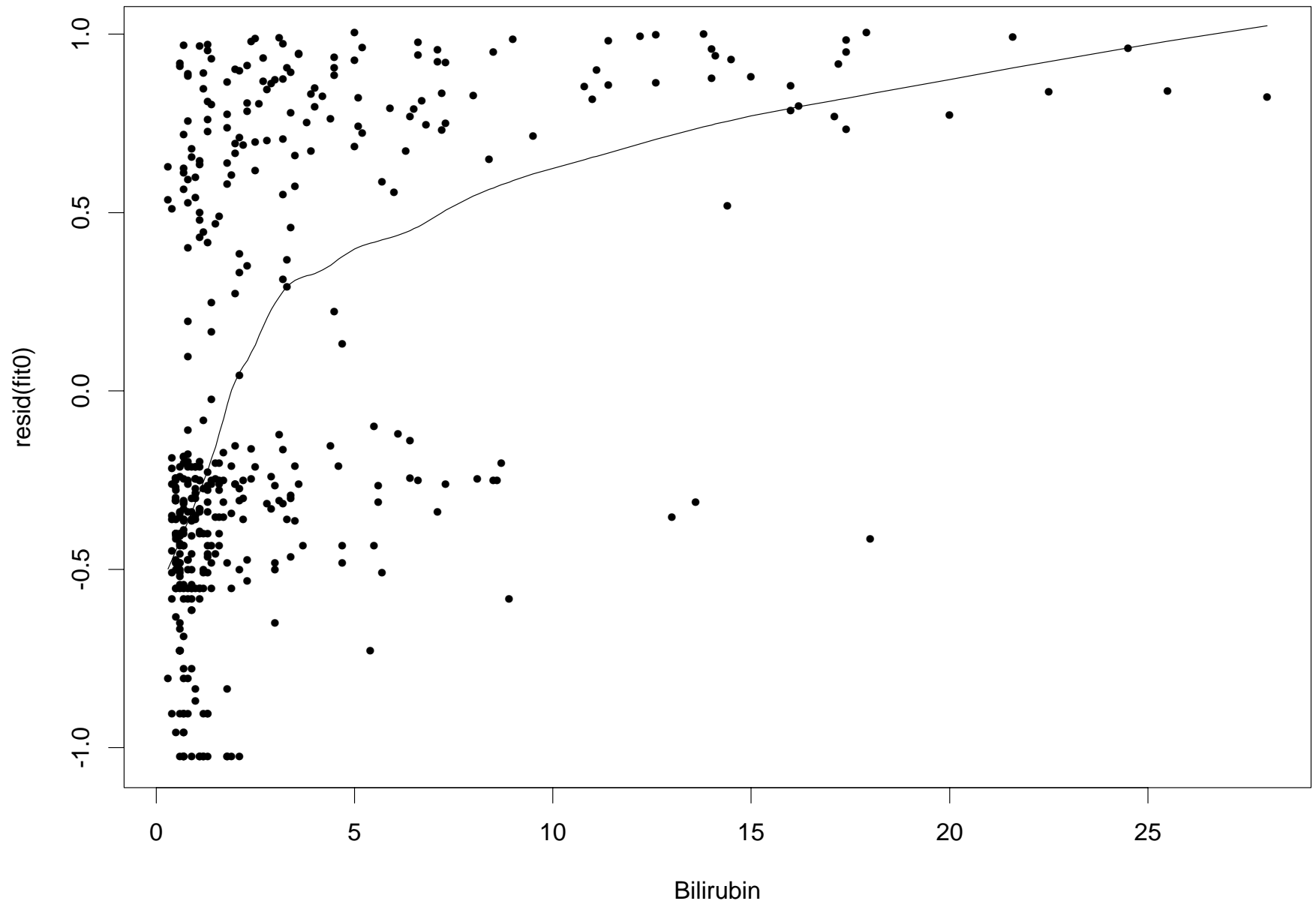
```
data temp1; set mydata;
  dummy = 0;
proc phreg data=temp1;
  model time * status(0) = dummy;
  output out=temp2 mresid=rr / order=data;
data temp3; merge temp1 temp2;
proc gplot data=temp3;
  plot rr * x;
  symbol i=sms50 value=plus;
```

Key idea– This plot behaves very much like an ordinary scatter plot.

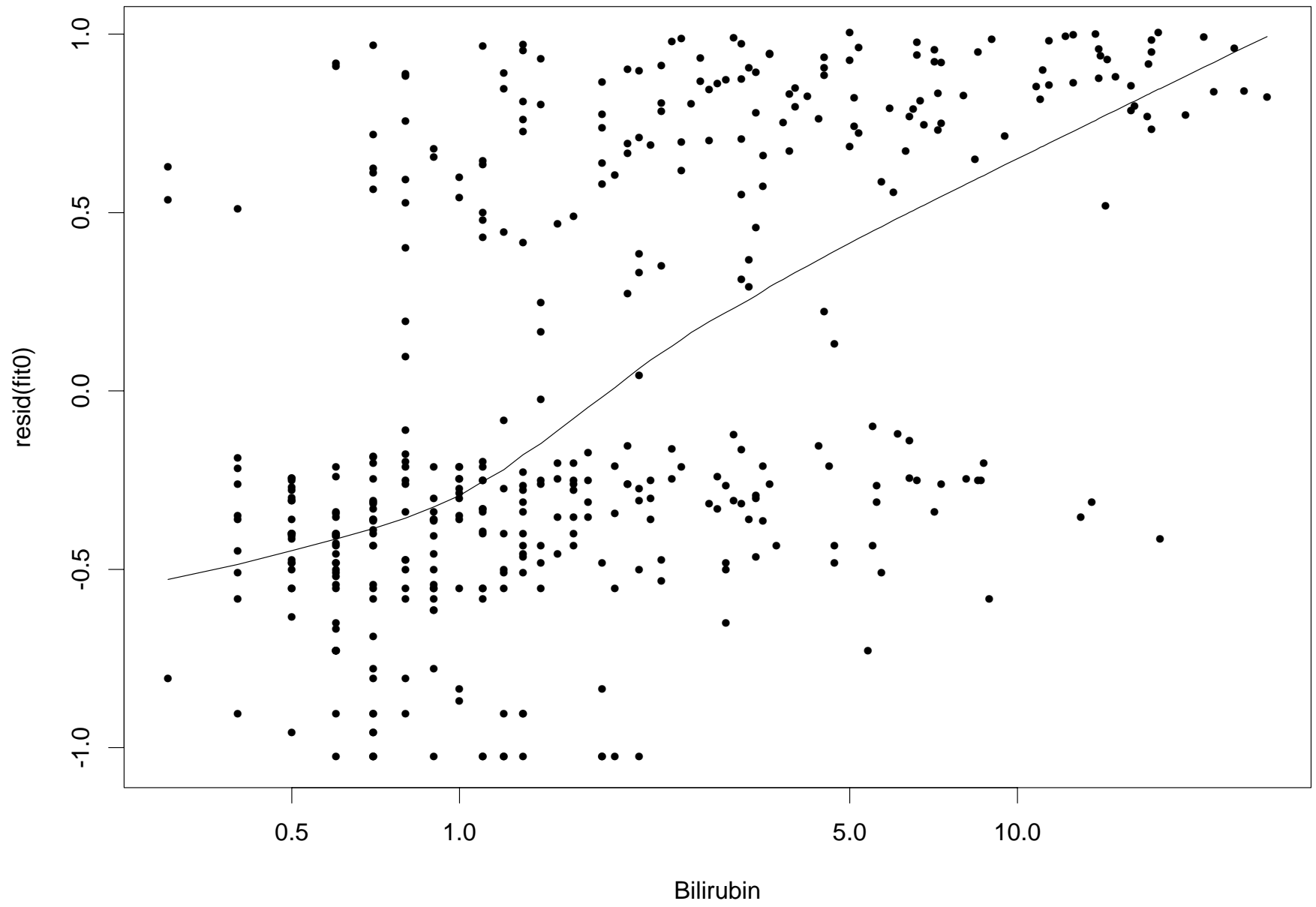
Example: Primary Biliary Cirrhosis

- PBC is a chronic disease, thought to be of autoimmune origin.
- The continuing inflammatory process eventually scars the bile ducts, leading to liver failure and death.
- Untreated, patients may live for up to 20 years.
- From other work, it is “known” that $\log(\text{bilirubin})$ is the best predictor of risk.

PBC Data



PBC Data



Synopsis

- Easy
- Treat the residuals as an ordinary “y” variable
- Use your favorite plotting tool, smoother, ...

Footnote:

The SAS printout is a bit odd, and I am not sure that this trick will always work. (It is not shown in the SAS manual).

```

Testing Global Null Hypothesis: BETA=0
      Without      With
Criterion Covariates Covariates Model Chi-Square
-2 LOG L      1746.975      1746.975      0.00 with 0 DF (p=0.0001)
Score          0.00 with 0 DF (p=0.0001)
Wald           0.00 with 0 DF (p=0.0001)

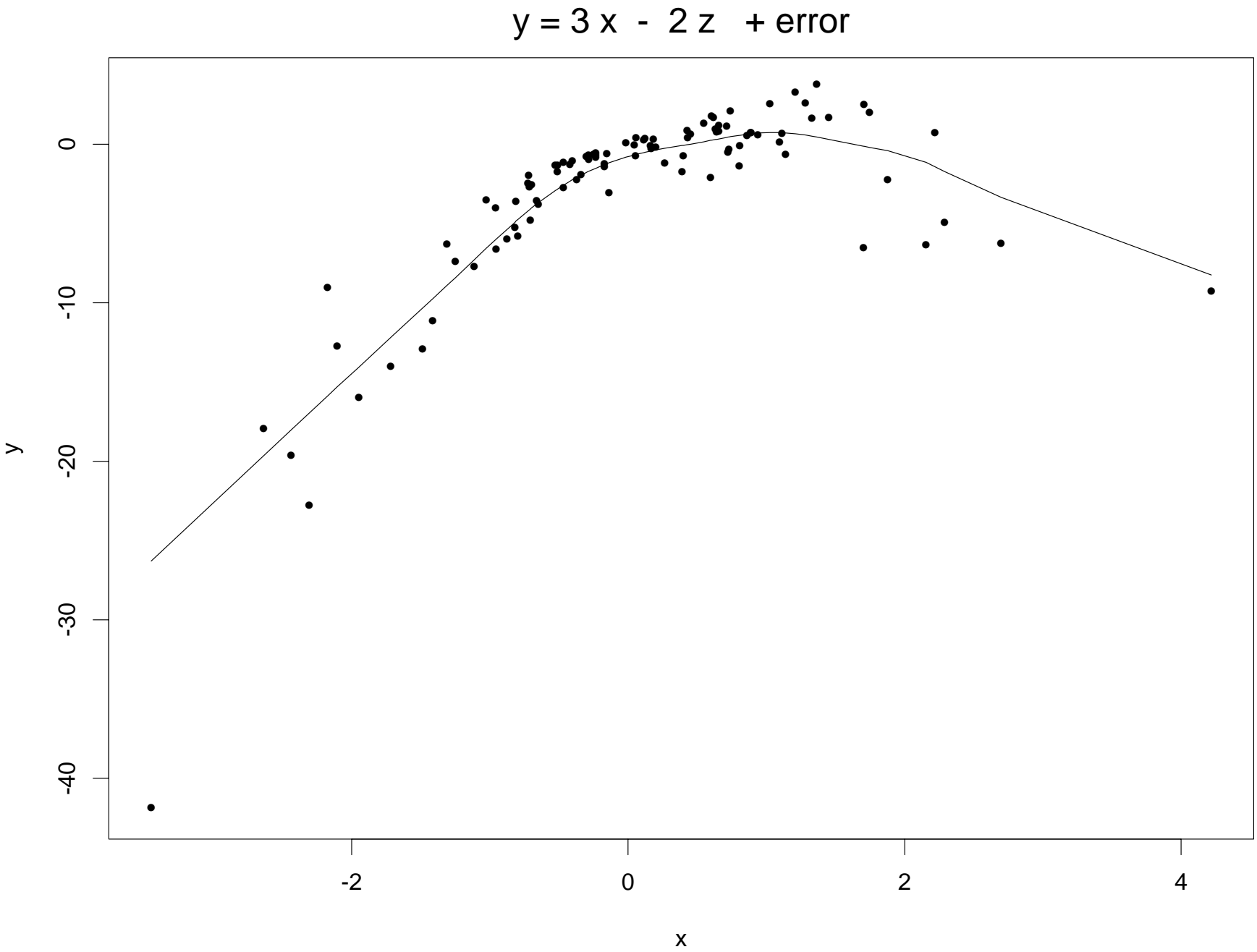
```

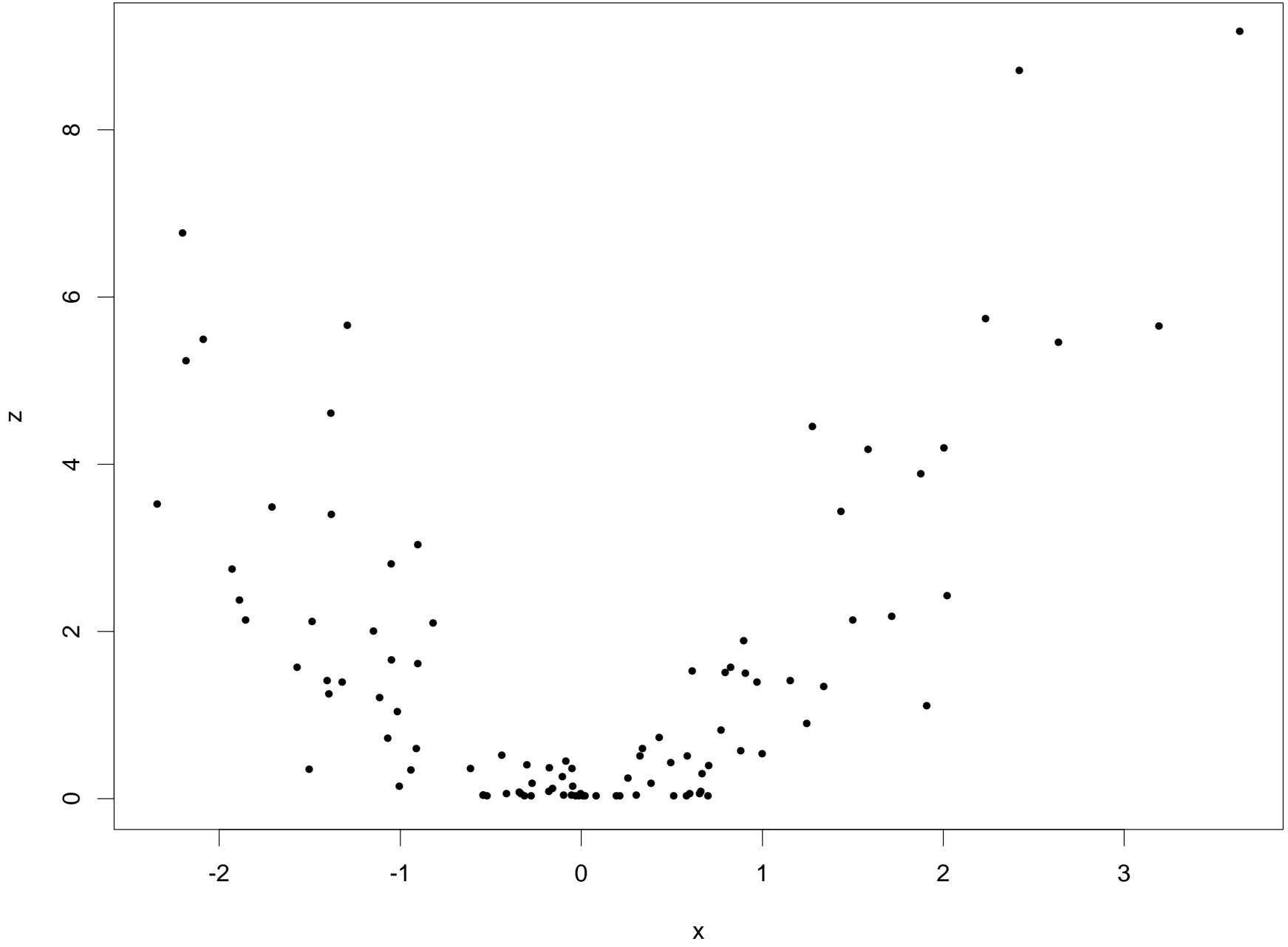
Problems

Consider the following scatterplot of y vs x where

$$y = 3x - 2z + \epsilon$$

100 data points, $\epsilon \sim \text{Gaussian}$.

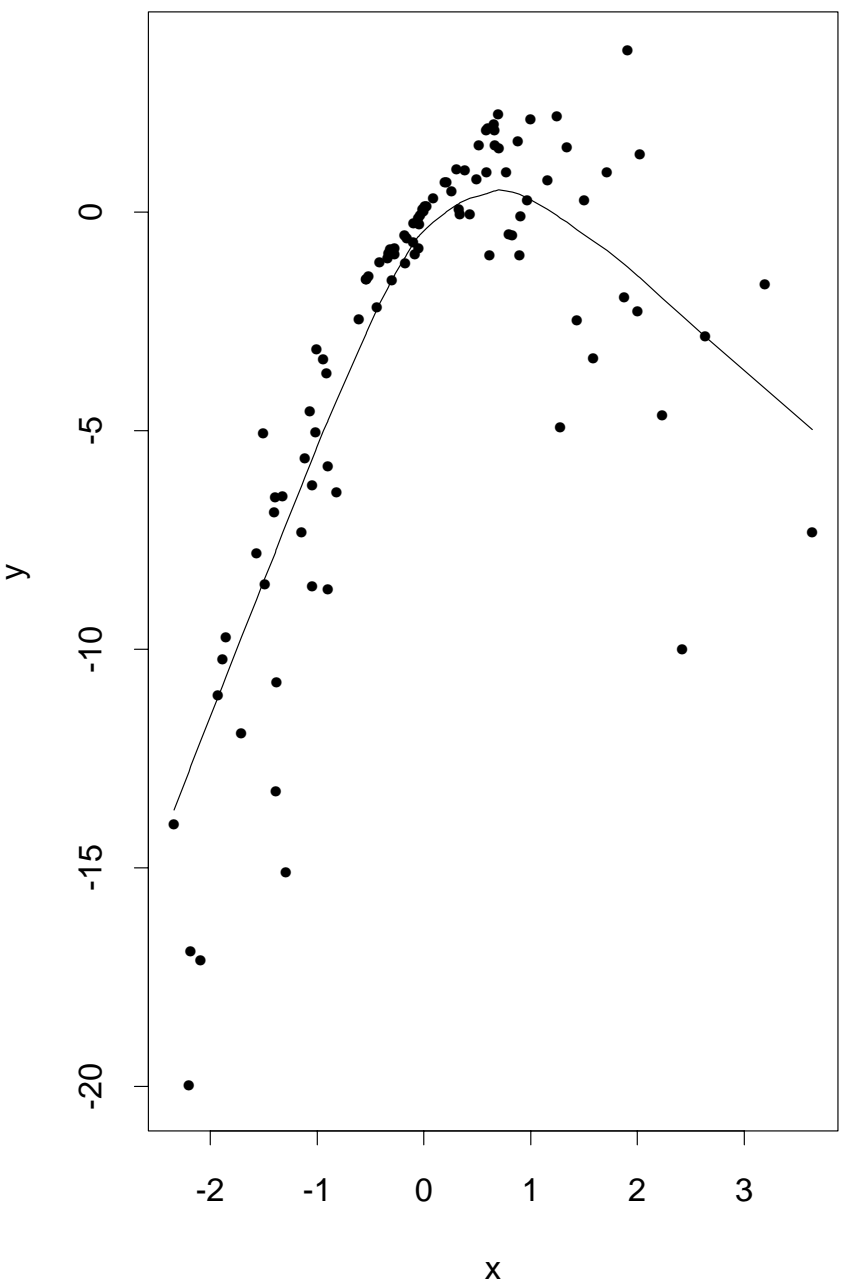
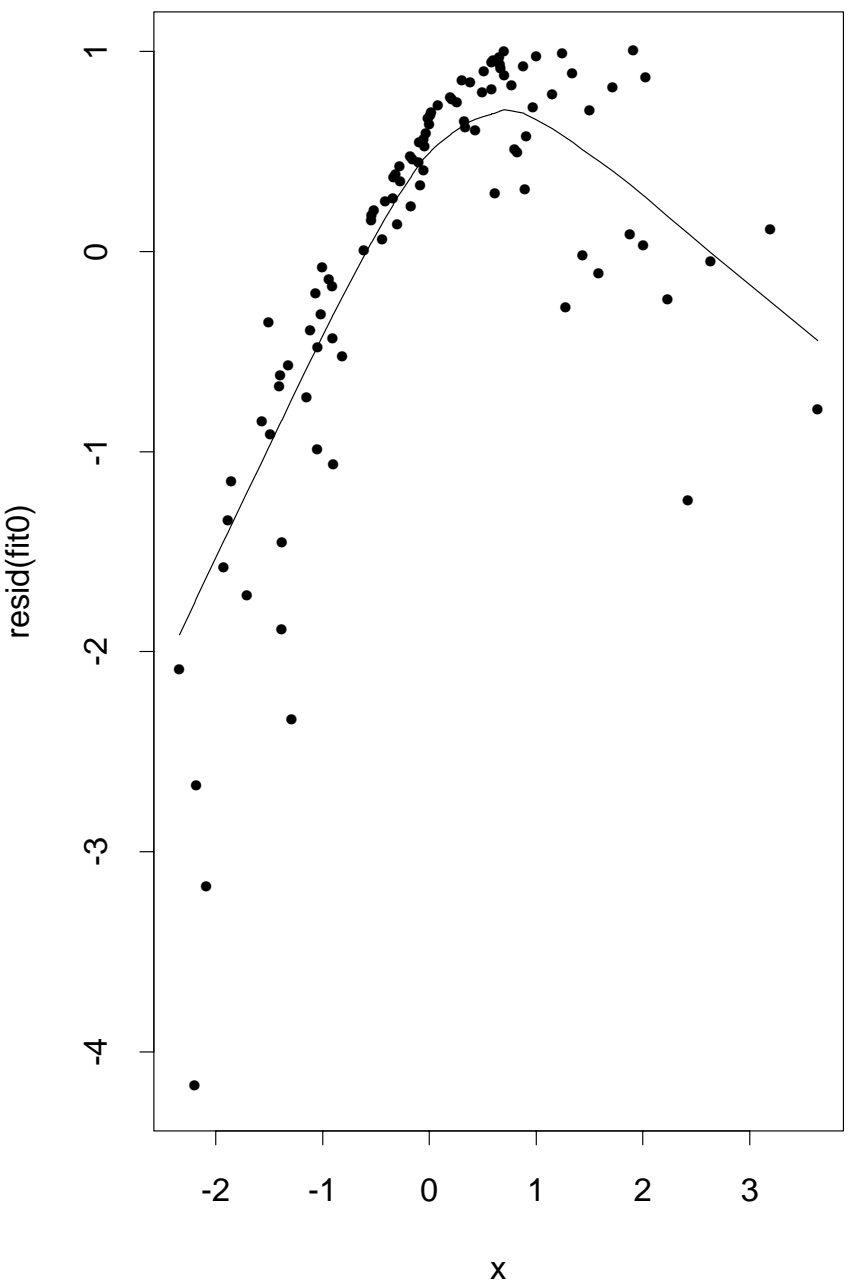




Problems

Because x and z are related, the non-linear relationship of z “bleeds over” into the plot of y vs x .

Exactly the same thing happens to the martingale residual plot. Using the same data, but a Cox model:



Fixes

This problem is well known in linear models.

1. Fit “all but age”, say; plot age vs residual.
2. Adjusted variable plots.
3. Augmented variable plots.
4. Constructed variable plots.
5. Poisson plots

Opinion

- #1 can be badly incorrect.
- 2-4 are somewhat of an improvement, but I have not been impressed.
- 5 works fairly well, but is a nuisance to set up.
- If you want more precision, fit splines.

```
> fit <- coxph(Surv(time, status) ~ rx + ns(age,4), ...)
> tt <- terms(fit)

%daspline(age,5)    in the data step
.
.
model time * status(0) = age age1 age2 age3 age4;
```

Conclusions

- Null model residuals can replace “y” for scatter plots
- A good smoother is essential
- Don't believe everything you see on the plot
- For heavily censored data, the plot looks like logistic regression.

Evaluating Proportional Hazards

Assume a model with age, weight, and treatment as variables. The Cox model assumes that

$$\lambda_i(t) = \lambda_0(t) e^{\beta_1 \text{age}_i + \beta_2 \text{wt}_i + \beta_3 \text{trt}_i}$$

One consequence is that the effect of age (wt, trt) is assumed to be *constant* over time.

A useful alternative is the time-dependent coefficient model

$$\lambda_i(t) = \lambda_0(t) e^{\beta_1(t) \text{age}_i + \beta_2 \text{wt}_i + \beta_3 \text{trt}_i}$$

which

- incorporates most useful alternatives to PH
- is readily interpretable
- *if* an estimate of $\beta_1(t)$ can be found

Key Idea:

$$\beta(t) \approx \hat{\beta} + \text{scaled Schoenfeld residuals}$$

The SSR are a matrix, with 1 row per death, and one column per variable.

S-Plus

```
> fit <- coxph(Surv(time, status) ~ age + wt + trt, ...  
> temp<- cox.zph(fit)  
> print(temp)  
> plot (temp)
```

SAS

```
proc phreg outbeta= temp1;  
  model time *status(0) = age wt rx;  
  output out=temp2 wtressch= age wt trt / order=data;
```

Lots of proc iml code;
or %schoen macro

Usage

The scaled residuals can also be treated as “raw” data.

- Plot $x=\text{time}$ vs $y=\text{resid}$.
 - fit a least-squares line $y = a + bx$
 - the test for $b=0$ is a test of PH
 - the test is equivalent to adding $x*\text{time}$ as a covariate
- Plots with $x=\text{time}$ and $x=\log(\text{time})$ will differ!
 - an outlier in x can give a false rejection.
 - never do the test without looking at the plot
 - the scale $x = \text{KM}(\text{time})$ has theoretical advantages
- a smooth curve gives an indication of the *type* of departure
 - standard methods can be used to get SE bands for the curve

Example

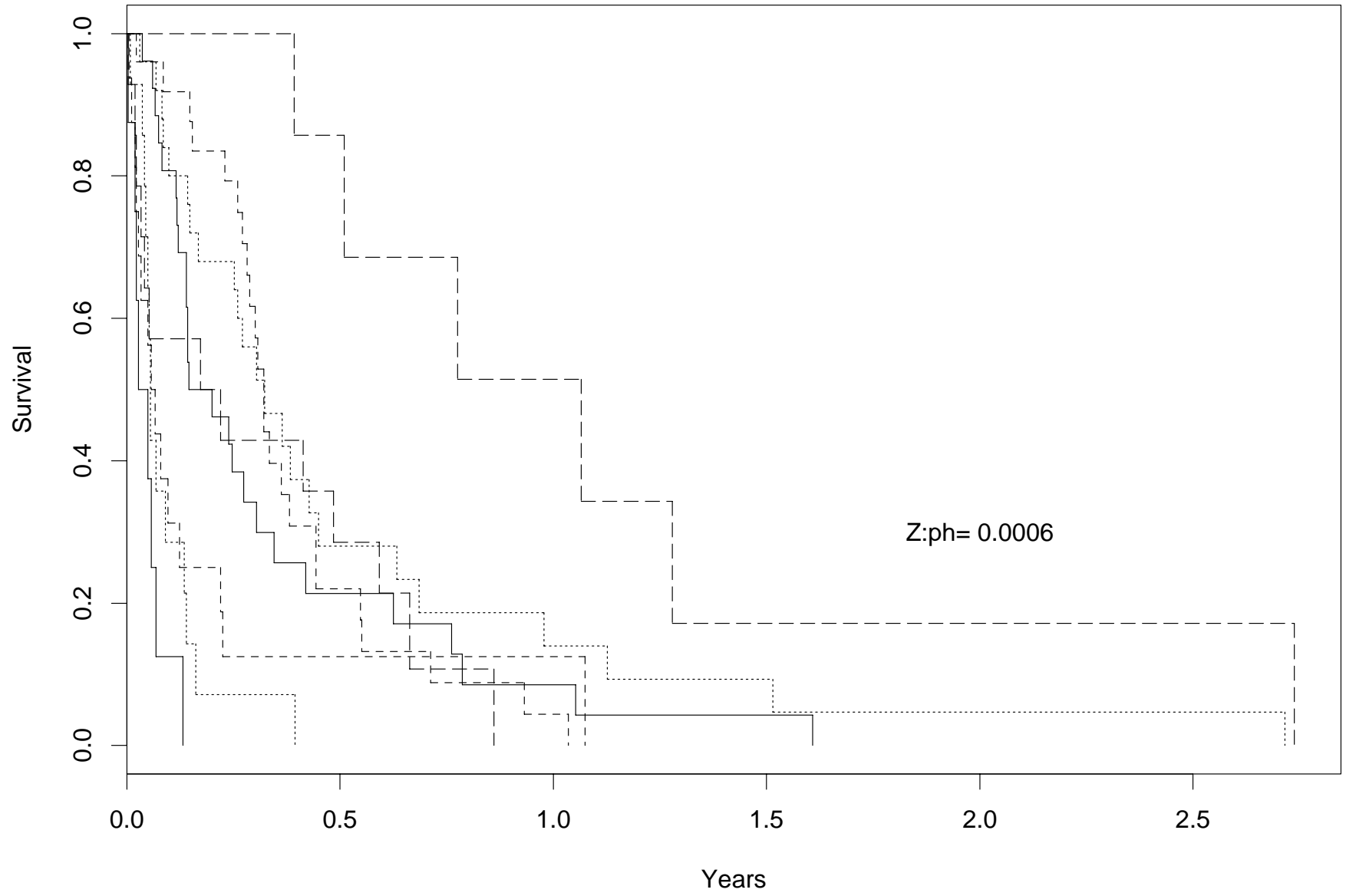
VA Lung Cancer data (K&P, Appendix 1).

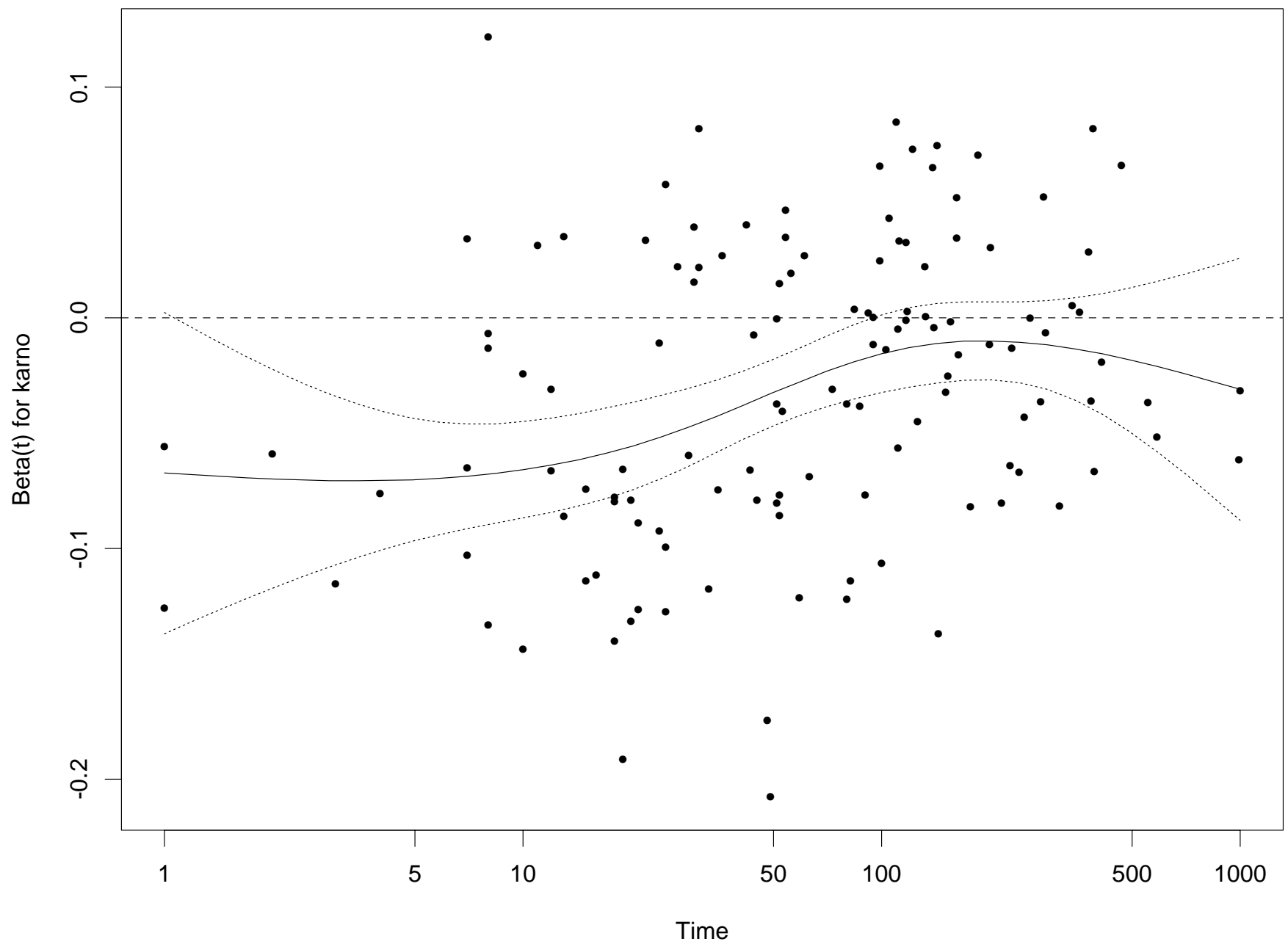
Tests show that Karnofsky score is not PH. What is the effect?

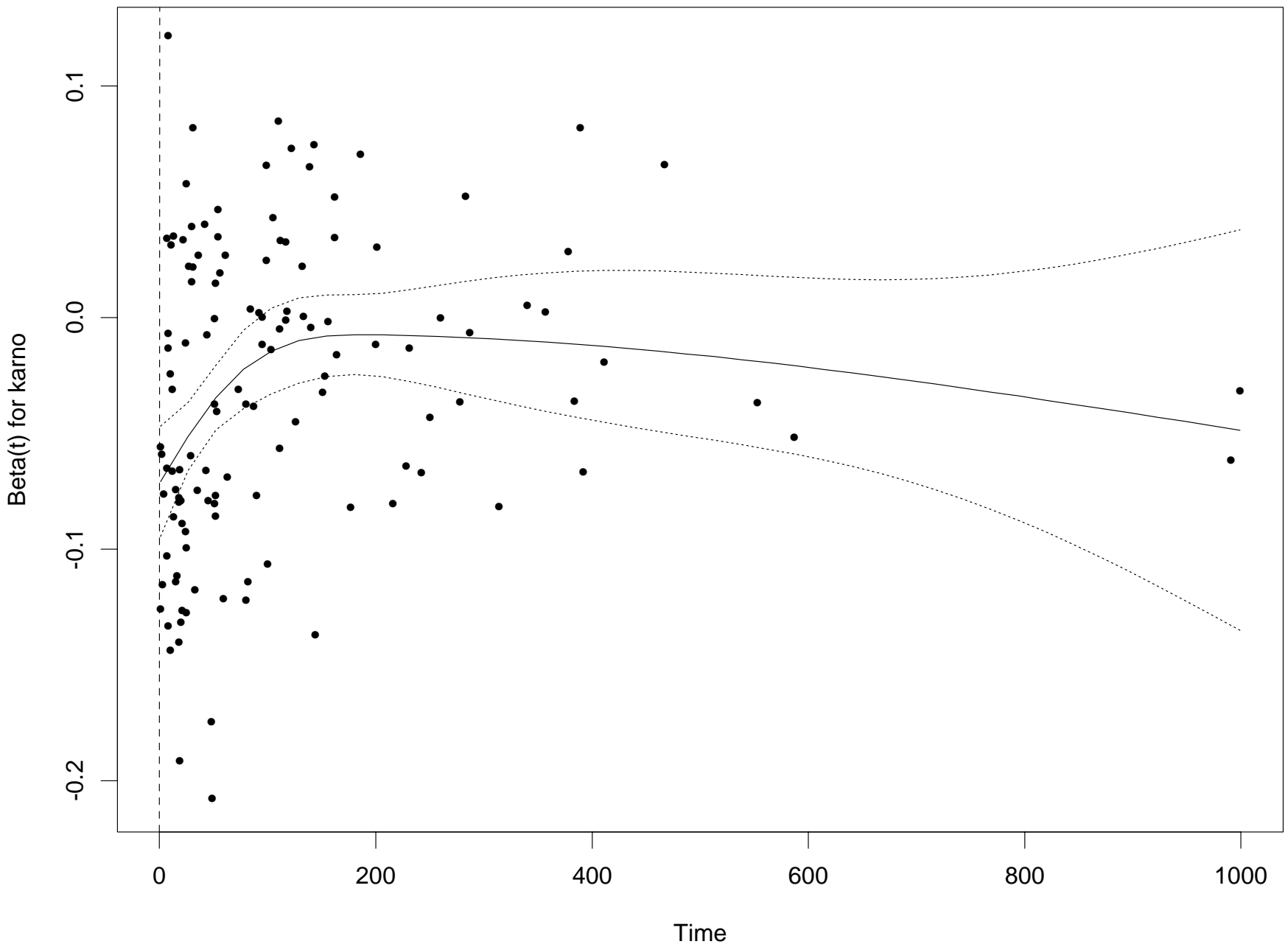
```
> fit <- coxph(Surv(futime, status) ~ age + karno + rx, veteran)
> temp<- cox.zph(fit, transform='log')
> plot(temp [2])
> abline(h=0)
> print(temp)
```

```
      rho chisq      p
age  0.2084  6.29 0.012159
karno 0.3129 11.87 0.000572
rx   -0.0922  1.13 0.287953
GLOBAL  NA 15.16 0.001687
```

Veteran's Administration Data







Leverage and the DFBETA residuals

An exact leverage for the Cox model can be obtained by the jackknife

1. Let $\hat{\beta}$ be the solution with all subjects.
2. Remove observation i from the data set.
3. Refit the data to get $\hat{\beta}_{(i)}$
4. Define leverage $_i = \hat{\beta} - \hat{\beta}_{(i)}$
5. Repeat 2-4 for each observation i

This is time consuming. The dfbeta residuals are based on an approximation which is

- fast
- has good accuracy
- may underestimate the influence of extreme outliers
- other, more complex approximations have been proposed, but do not seem to do better practically

Usage

S-Plus

```
> fit <- coxph(Surv(time, status) ~age + height +wt, data=mine)
> D <- resid(fit, type='dfbeta')
> plot(age, D[,1])
```

SAS

```
proc phreg data=mine;
    model time *status(0) = age height wt;
    output out=temp1 dfbeta=dage dht dwt / order=data;
data temp2; merge mine temp1;

proc plot;
    plot dage * age;
```

Multiple obs per subject

- Either the subject has been arbitrarily broken up into multiple obs, e.g. time-dependent covariates, or may have multiple events per subject.
- In this case it is important to distinguish between per-observation and per-subject leverage.

Example: Diabetic Retinopathy Study.

Between 1972 and 1975 seventeen hundred forty-two patients were enrolled in the study to evaluate the efficacy of photocoagulation treatment for proliferative diabetic retinopathy; photocoagulation was randomly assigned to one eye of each study patient, with the other eye serving as an untreated control. A major goal was to assess whether treatment significantly delayed the onset of severe visual loss.

Diabetic Retinopathy Trial

- Two treatments.
- Randomly assigned to left and right eye of each patient.
- Leads to a data set with two observations (eyes) per subject.

$$D_{2n \times p} \implies \tilde{D}_{n \times p}$$

$$\begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1p} \\ d_{21} & d_{22} & \cdots & d_{2p} \\ d_{31} & d_{32} & \cdots & d_{3p} \\ d_{41} & d_{42} & \cdots & d_{4p} \\ d_{51} & d_{52} & \cdots & d_{5p} \\ \vdots & \vdots & & \vdots \end{pmatrix} \quad \text{add} \quad \begin{pmatrix} \tilde{d}_{11} & \tilde{d}_{12} & \cdots & \tilde{d}_{1p} \\ \tilde{d}_{21} & \tilde{d}_{22} & \cdots & \tilde{d}_{2p} \\ \vdots & \vdots & & \vdots \end{pmatrix}$$

D contains the *per observation* influence.

\tilde{D} contains the *per subject* influence.

Robust Variance

Let D be the matrix of dfbeta residuals.

$D'D$ is the approximate jackknife estimate of variance.

- Proposed in passing by Reid and Crépeau, *Biometrika* 1985.
- Identical to the estimate of Lin and Wei, *JASA* 1989.
- The latter develop the properties and usefulness of this robust estimator.

If there are multiple observations per subject $D'D$ is not a good idea, one should use the per-subject jackknife $\tilde{D}'\tilde{D}$.

- Identical to the estimate proposed by Wei, Lin and Weissfeld, *JASA* 1989.
- For multiple-event models, this is similar in derivation to the *working independence* variance estimate of GEE models.

Particularly easy to use in S-Plus

```
> coxph(Surv(time, status) ~ rx + number + cluster(id),  
        data=bladder)  
  
      coef exp(coef) se(coef) robust se      z      p  
rx -0.412    0.663    0.2003    0.2515 -1.64 0.1000  
number 0.170    1.185    0.0465    0.0564  3.02 0.0026  
  
Likelihood ratio test=14.3  on 2 df, p=0.00078  n= 190
```

For SAS, see the example in their manual, or use %pphlev.

Conclusions

Residuals are widely available.

They are easy to use.

They are understandable, with a strong similarity to linear models.