Checking Assumptions in the Cox Proportional Hazards Regression Model

Brenda Gillespie, Ph.D.
University of Michigan

Presented at the 2006 Midwest SAS Users Group (MWSUG)
Dearborn, Michigan
October 22-24, 2006

Outline

- Introduction to the Cox model
- Overview of residuals for the Cox model
- Model assumptions
  - Correct model specification
    - Functional form of continuous covariates
    - Covariate interactions
  - Proportional hazards (PH)
    - Graphical checks
    - Tests of PH
    - What to do if non-PH is found
      - Stratification
      - Time-dependent covariates
- Conclusions

Survival Analysis

- Methods to analyze "time to event" data.
- Useful for many different applications
  - Time to death from disease diagnosis
  - Length of hospital stay
  - Cost of insurance claims.

Time origin ————————————→ Event

Time origin ————————————→ Censored value
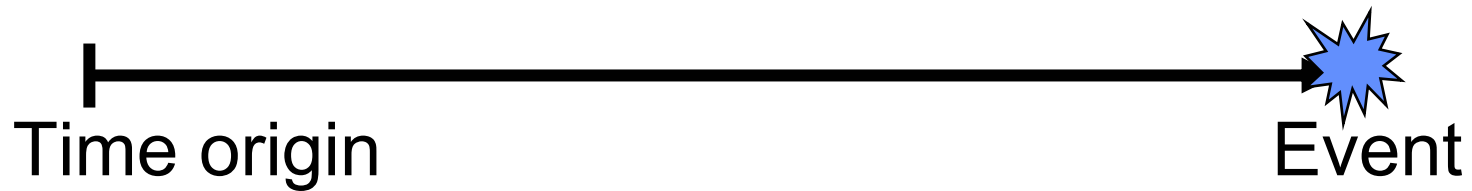
Censoring and Truncation

- Censoring and truncation describe different forms of incomplete observation of event times:

❑**Censoring**
- Right
- Left
- Interval

❑**Truncation**
- Right
- Left (delayed entry)
- Interval (gap times)

- **Here, we assume right-censored data only.**

Random Censoring Reminder

- All standard methods of survival analysis assume that censoring is random: Those censored at time t_i should be representative of all subjects still alive at t_i (with the same covariate values).

- This assumption cannot be checked by any statistical test.

Cox Regression Model

$$h(t; x) = h_o(t) \exp\{\beta_1 x_1 + \cdots + \beta_k x_k\}$$

where **$h(t ; x)$** is the hazard function at time **t** for a subject with covariate values **x_1**, … **x_k**,

$h_0(t)$ is the baseline hazard function, i.e., the hazard function when all covariates equal zero.

exp is the exponential function ($\exp(x) = e^x$),

**x_i** is the ith covariate in the model, and

β_i is the regression coefficient for the ith covariate, **x_i**.

Cox Regression (cont'd)

$$h(t;x) = h_o(t) \exp\{\beta_1 x_1 + \cdots + \beta_k x_k\}$$

- The Cox Model is different from ordinary regression in that the covariates are used to predict the <u>hazard</u> <u>function</u>, and not Y itself.
- The baseline hazard function can take any form, but it cannot be negative.
- The exponential function of the covariates is used to insure that the hazard is positive.
- There is no intercept in the Cox Model . (Any intercept could be absorbed into the baseline hazard.)

Cox Regression (cont'd)

$$h(t, x_i)$$

t
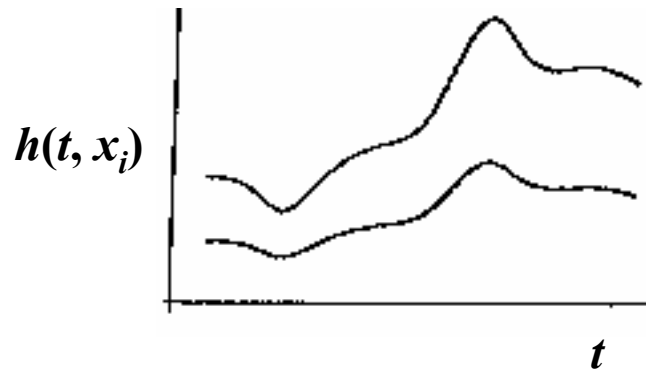
- The basic Cox Model assumes that the hazard functions for two different levels of a covariate are proportional for all values of t.

- For example, if men have twice the risk of heart attack compared to women at age 50, they also have twice the risk of heart attack at age 60, or any other age.

- The underlying risk of heart attack as a function of age can have any form.

Proportional Hazards

To see the proportional hazards property analytically, take the ratio of h(t;*x*) for two different covariate values:

$$\frac{h(t; x_i)}{h(t; x_j)} = \frac{h_0(t) \exp\{\beta_1 x_{i1} + \cdots + \beta_k x_{ik}\}}{h_0(t) \exp\{\beta_1 x_{j1} + \cdots + \beta_k x_{jk}\}}$$

$$= \exp\{\beta_1(x_{i1} - x_{j1}) + \cdots + \beta_k(x_{ik} - x_{jk})\}$$

$h_o(t)$ cancels out => the ratio of those hazards is the same at all time points.

For a single dichotomous covariate, say with values 0 and 1, the **hazard ratio** is

$$\frac{h(t; x = 1)}{h(t; x = 0)} = \frac{h_0(t) e^{\beta * 1}}{h_0(t) e^{\beta * 0}} = \frac{e^{\beta}}{e^0} = e^{\beta}$$

Software for Cox Regression: PHREG

- Syntax for Cox regression using Proc PHREG
  - The time variable is "days"
  - The censor code is "status" (1=dead, 0=alive)
  - Underlined items are user-specified

```
proc phreg;
   model  days*status (0) = sex age;
   output out=temp resmart=Mresids
      resdev=Dresids ressch=Sresids;
   id  subj  group;
   run;
```

Overview of Residuals for Cox Regression

- Cox-Snell residuals
  - range 0 to ∞
- Martingale residuals
  - a linear transform of Cox-Snell residuals
  - range $-\infty$ to 1
- Deviance residuals
  - a transform of Martingale residuals to make symmetric around zero
- Score residuals (one per subject per covariate)
- Schoenfeld residuals (one per subject per covariate)

Common Residual Plots

- Plot martingale residuals vs continuous covariates
  - to check functional form of covariates
- Plot deviance residuals vs Observation #
  - to check for outliers
- Plot Schoenfeld residuals for each covariate, vs Time or log(Time)
  - to check proportional hazards (PH)

- Note: Censoring and categorical covariates can produce banded residual patterns that do not reflect any problem with the model.

Martingale Residuals

- Skewed
- Near 1 $\Rightarrow$ "died too soon"; Large negative $\Rightarrow$ "lived too long"
- Plots of residuals vs. continuous covariates: Patterns may suggest continuous variables not properly fit

Example of Martingale Residuals
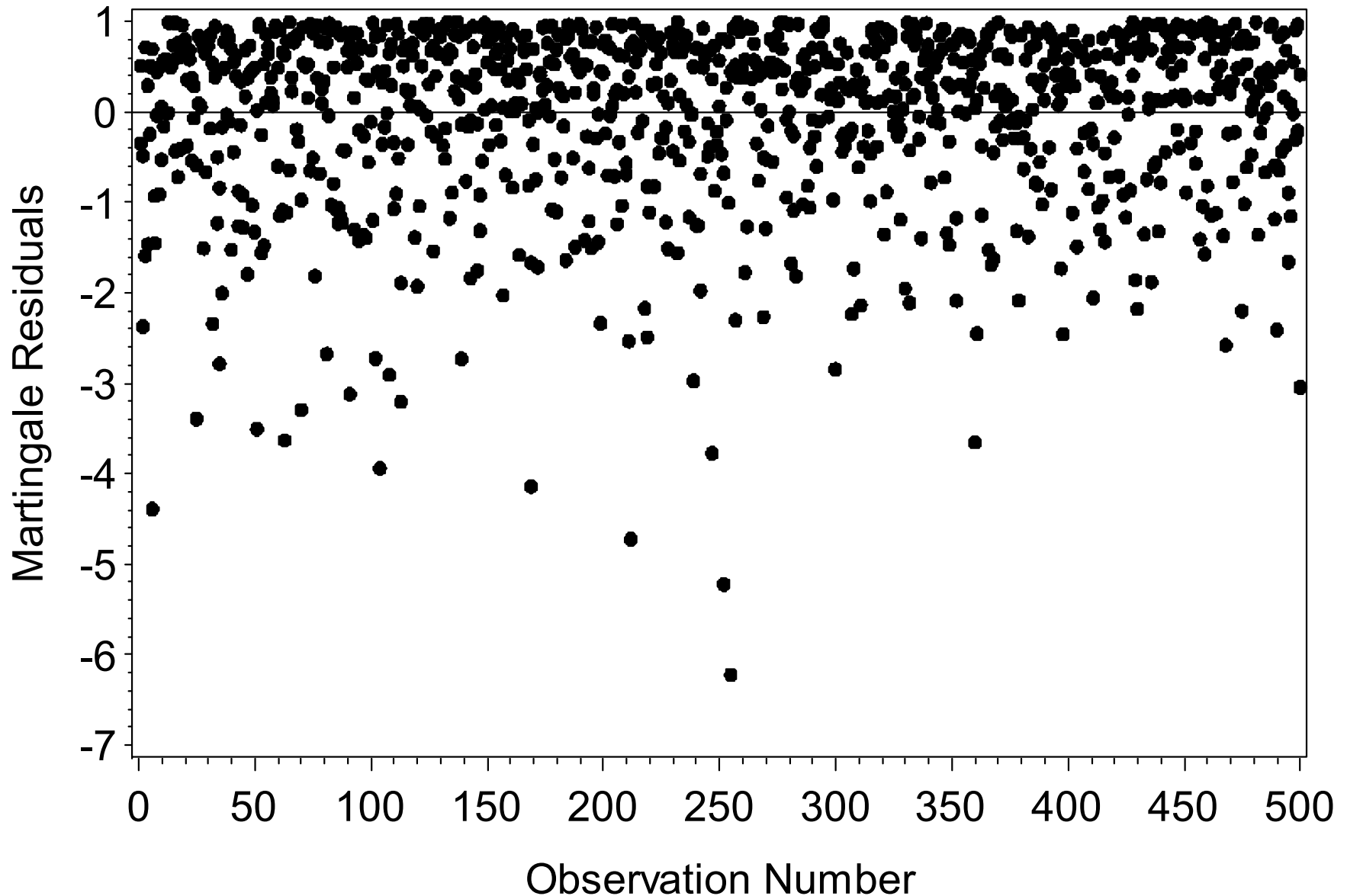
Deviance Residuals

- Roughly symmetrically distributed around zero, with approximate s.d. = 1.0
- Positive values $\Rightarrow$ "died too soon"
- Negative values $\Rightarrow$ "lived too long"
- Very large or small values $\Rightarrow$ outliers

- This is the only plot that is useful for checking outliers.
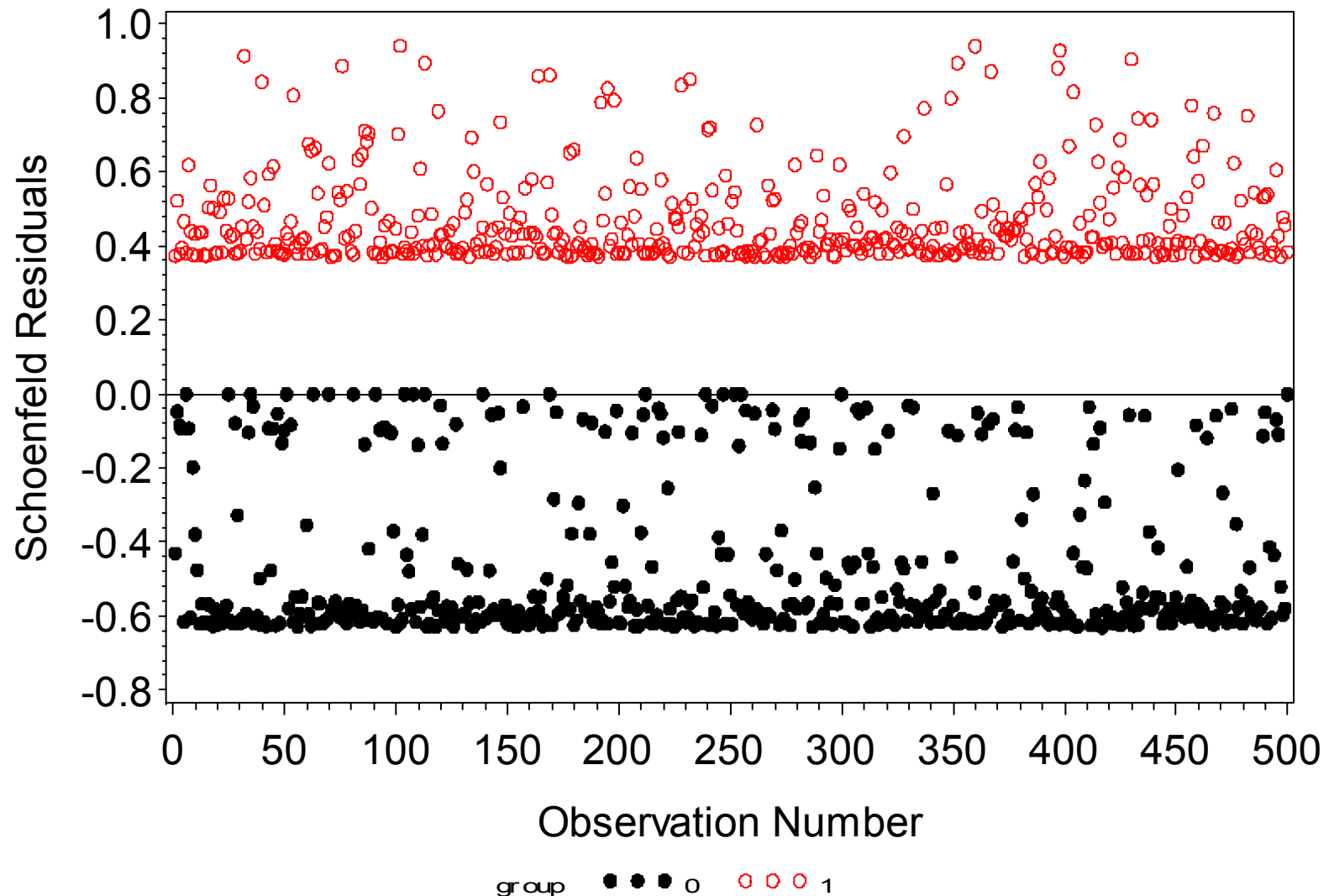
Example of Deviance Residuals
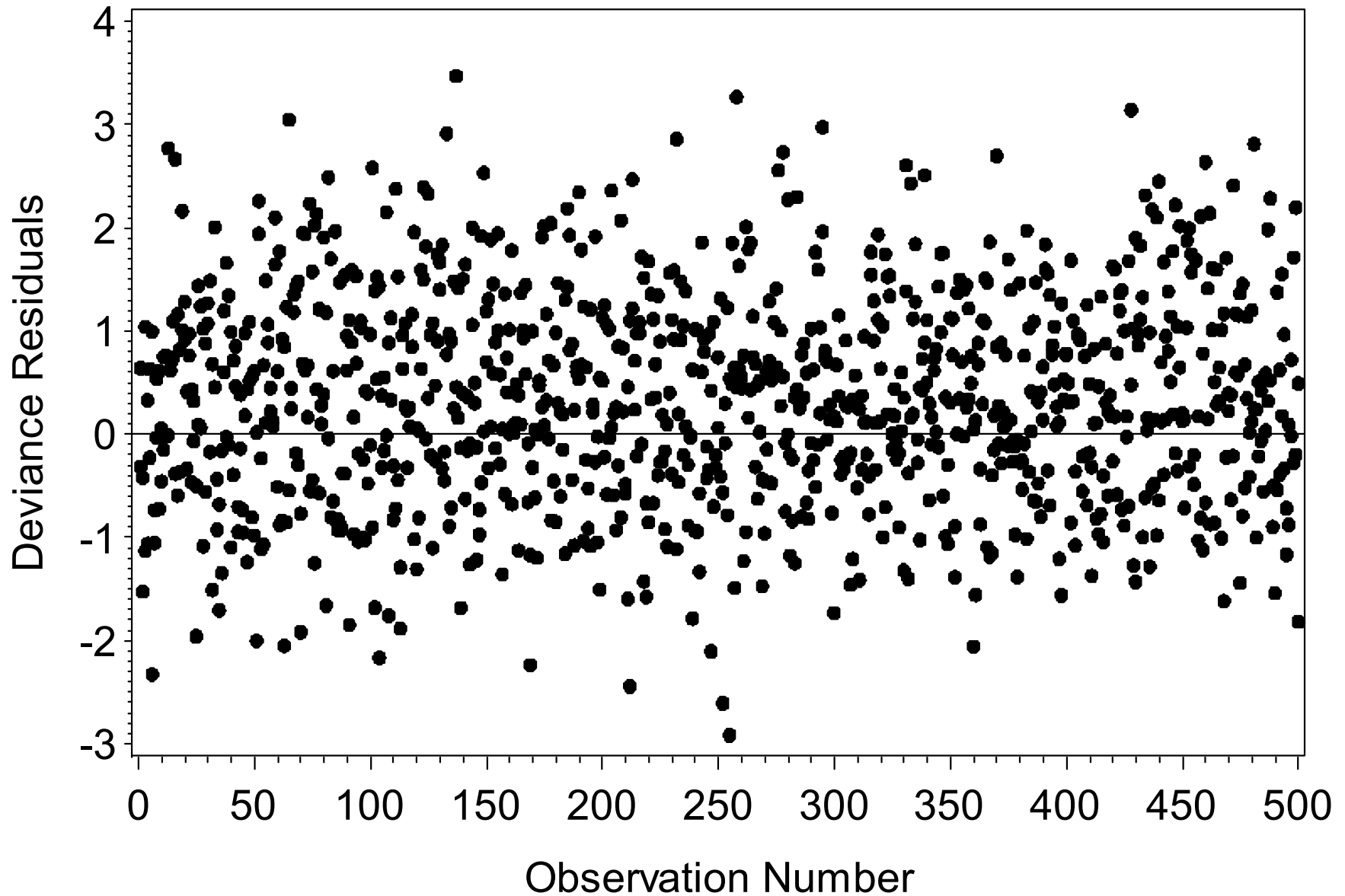
Schoenfeld Residuals

- Schoenfeld residuals are computed with one per observation **per covariate**.

  - Only defined at observed event times
  - For the ith subject and k^{th} covariate, the estimated Schoenfeld residual, r_{ik}, is given by (notation from Hosmer and Lemeshow)

  $$\hat{r}_{ik} = x_{ik} - \hat{\bar{x}}_{w_i k}$$

  - Where x_{ik} is the value of the kth covariate for individual i, and
  - $\hat{\bar{x}}_{w_i k}$ is a weighted mean of covariate values for those in the risk set at the given event time.
 - A positive value of r_{ik} shows an X value that is higher than expected at that death time.

Schoenfeld Residuals

- Schoenfeld residuals sum to zero.

- For a dichotomous (0,1) variable, Schoenfeld residuals will be between –1 and 1.

- In this case,

$$\hat{r}_{ik} = x_{ik} - \hat{\bar{x}}_{w_i k} = \begin{cases} 0 - \hat{\bar{x}}_{w_i k}, \text{ for } x = 0 \\ 1 - \hat{\bar{x}}_{w_i k}, \text{ for } x = 1 \end{cases}$$

- The residual plot will have two bands, one above zero for *x=1*, and one below zero for *x=0*.

Example of Schoenfeld Residuals
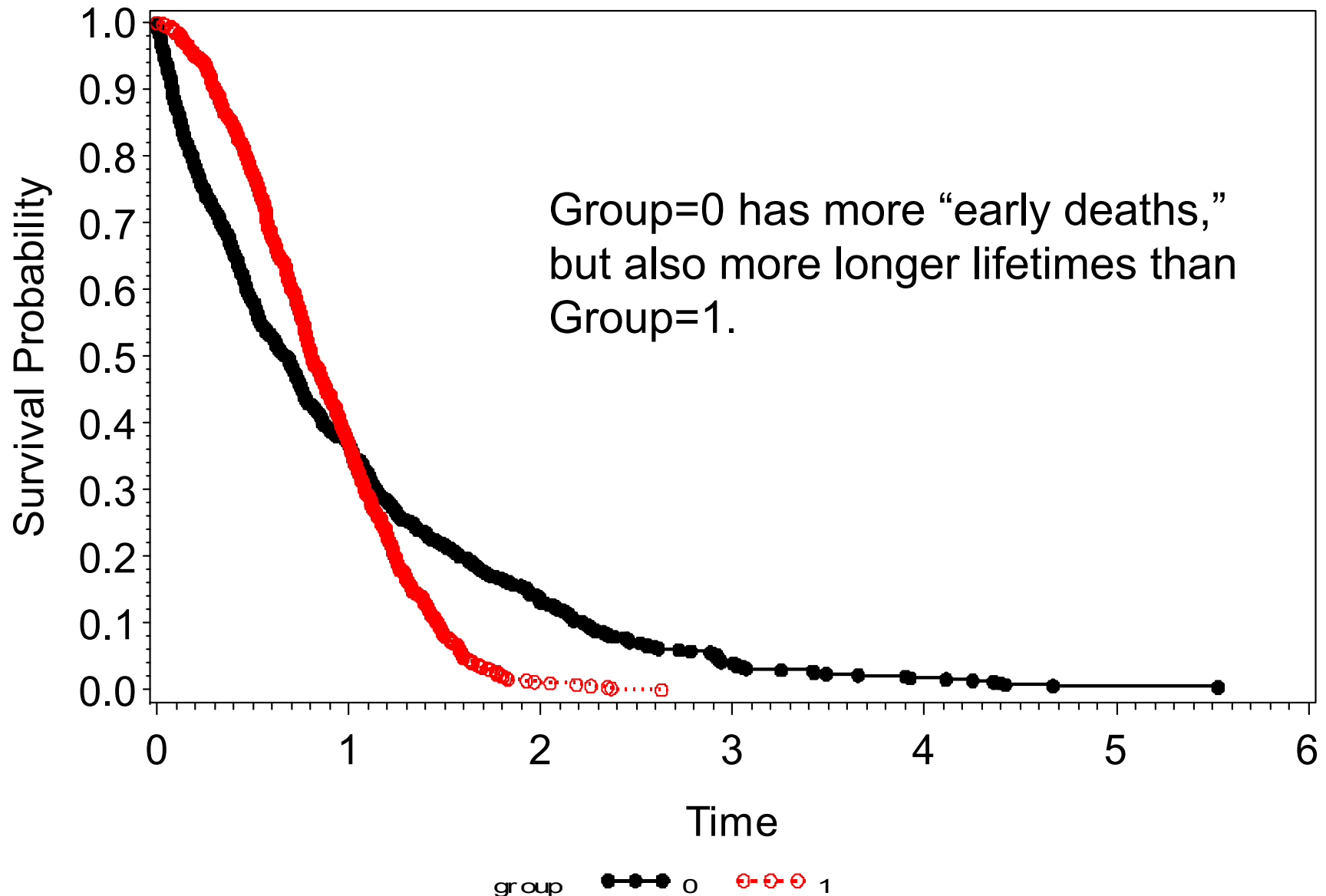## for the dichotomous covariate, "group," plotted by Observation Number

Example of Schoenfeld Residuals
for the dichotomous covariate, "group," plotted by Time

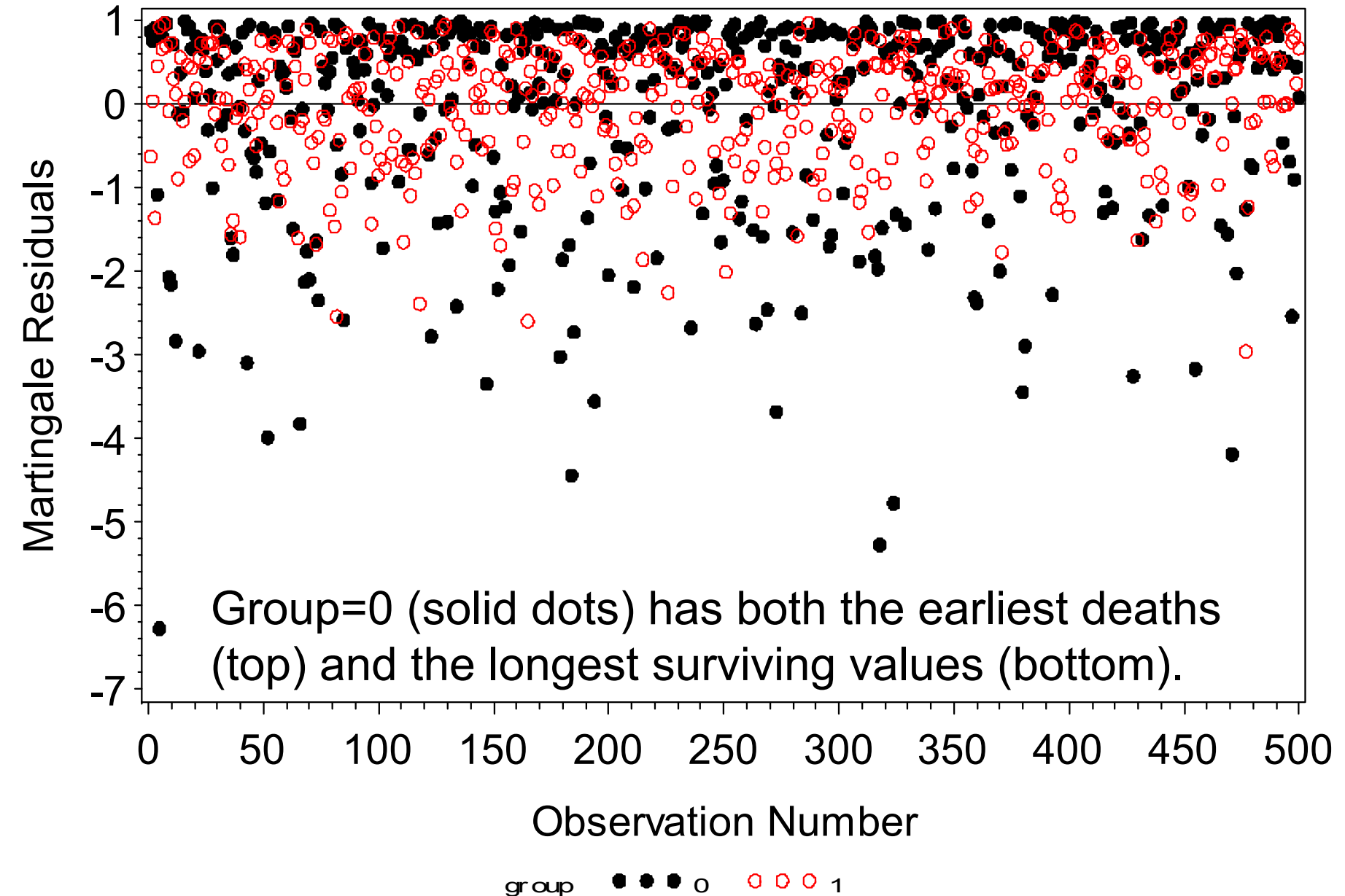The increasing pattern Indicates non-PH.
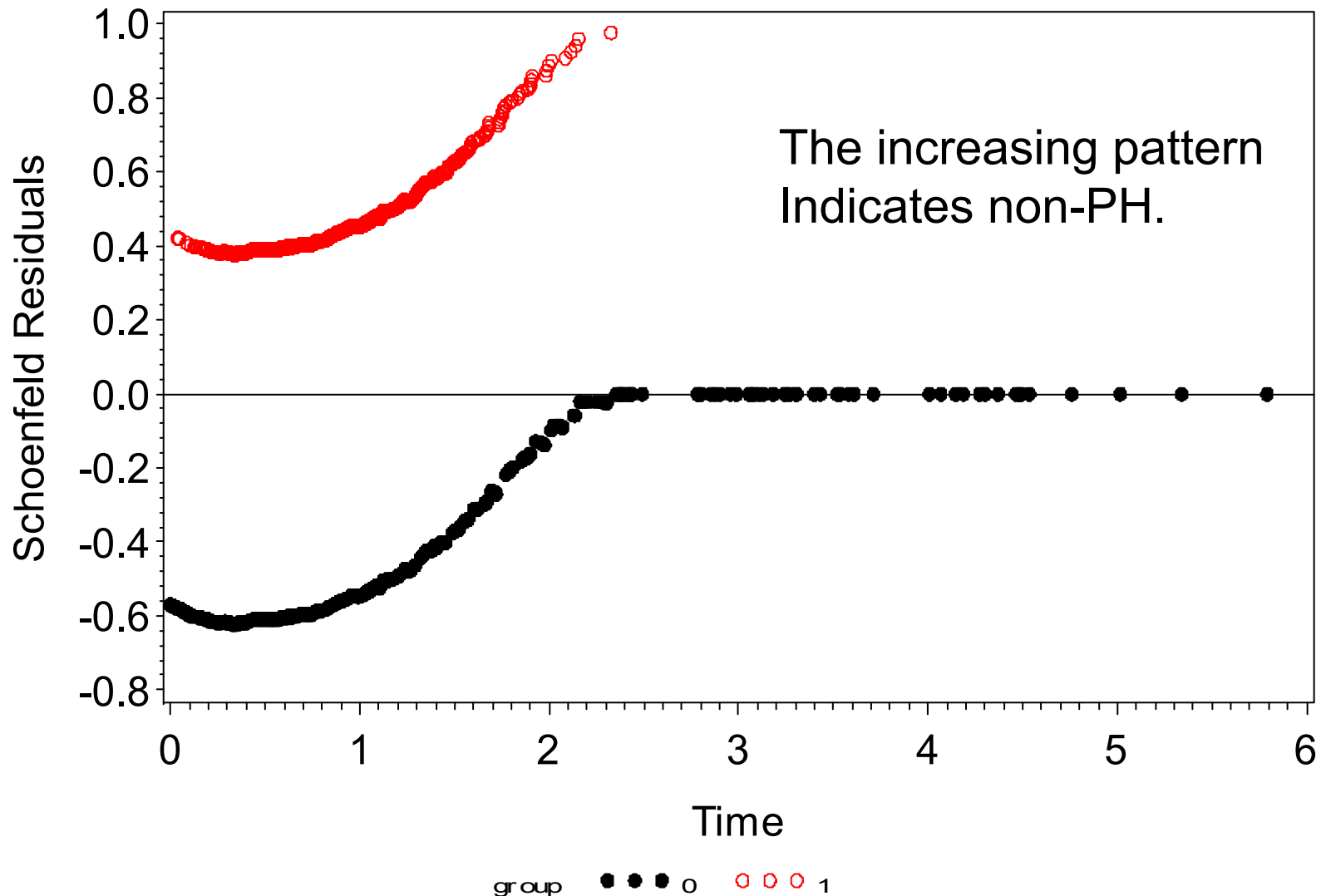
An Example of Martingale and Deviance Residuals with non-PH

- Outcome: Time to death
- Covariate: treatment group (labels 0 and 1)
- The next 3 slides show
  - Kaplan-Meier plot for the two groups
  - Martingale residuals
  - Deviance residuals

Example: KM Plot shows Crossing Survival Functions (non-PH)

Group=0 has more "early deaths," but also more longer lifetimes than Group=1.

Martingale Residuals

Group=0 (solid dots) has both the earliest deaths (top) and the longest surviving values (bottom).

group ● ● ● 0   o o o 1
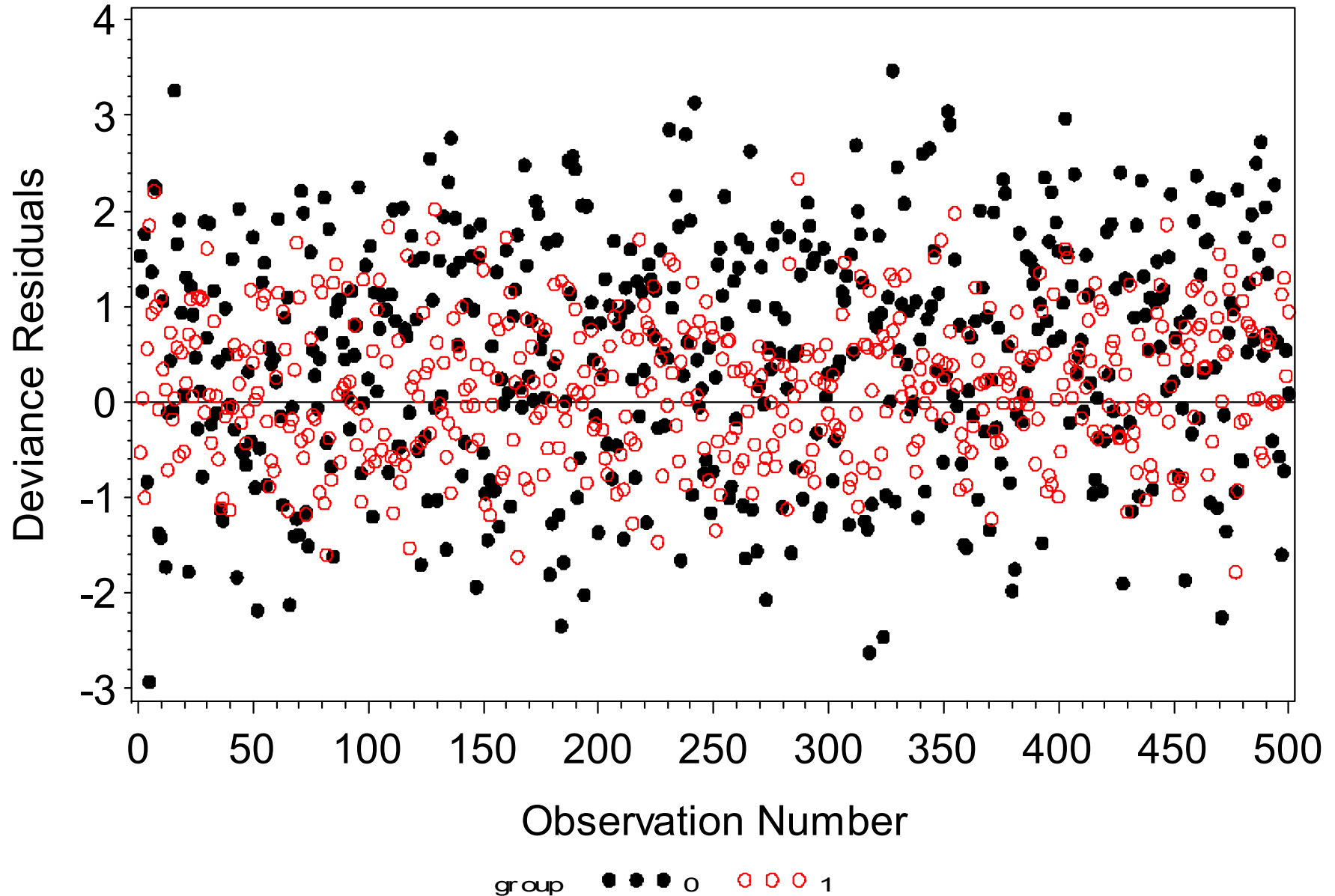
Observations

- In both Martingale and Deviance residuals, Group=0 had both the earliest deaths and the longest surviving values (most extreme values top and bottom).

- Such a pattern would indicate non-proportional hazards (non-PH)

- Other situations of non-PH may not be so easy to see from these plots.

- In this example, the Deviance residual plot does not show any outliers.

Assumptions of the Cox Model

- Structure of the model is assumed correct
  - Model is multiplicative (e.g., vs additive)
  - All relevant covariates have been included
  - We will not consider these assumptions here
- Functional form
  - Do we have the correct functional form for continuous covariates?
  - Are there any significant interactions?
- Is the Proportional Hazards assumption met? If not, what are the options?

Assessing Functional Form of Continuous Covariates

- Often we assume continuous covariates have a linear form. However, this assumption should always be checked. We give 3 ways to check:

- Method 1 (try X categorical):

  - Categorize X into ≥ 4 intervals, say by quantiles.

  - Create dummy variables for the categories and fit a model with these dummy variables.

  - Plot β estimates by X interval midpoints, with $\beta=0$ for the reference category.

 - Look at the shape, and model X accordingly (e.g., linear, quadratic, threshold).

Plot of Beta Estimates by
Age Category Midpoints
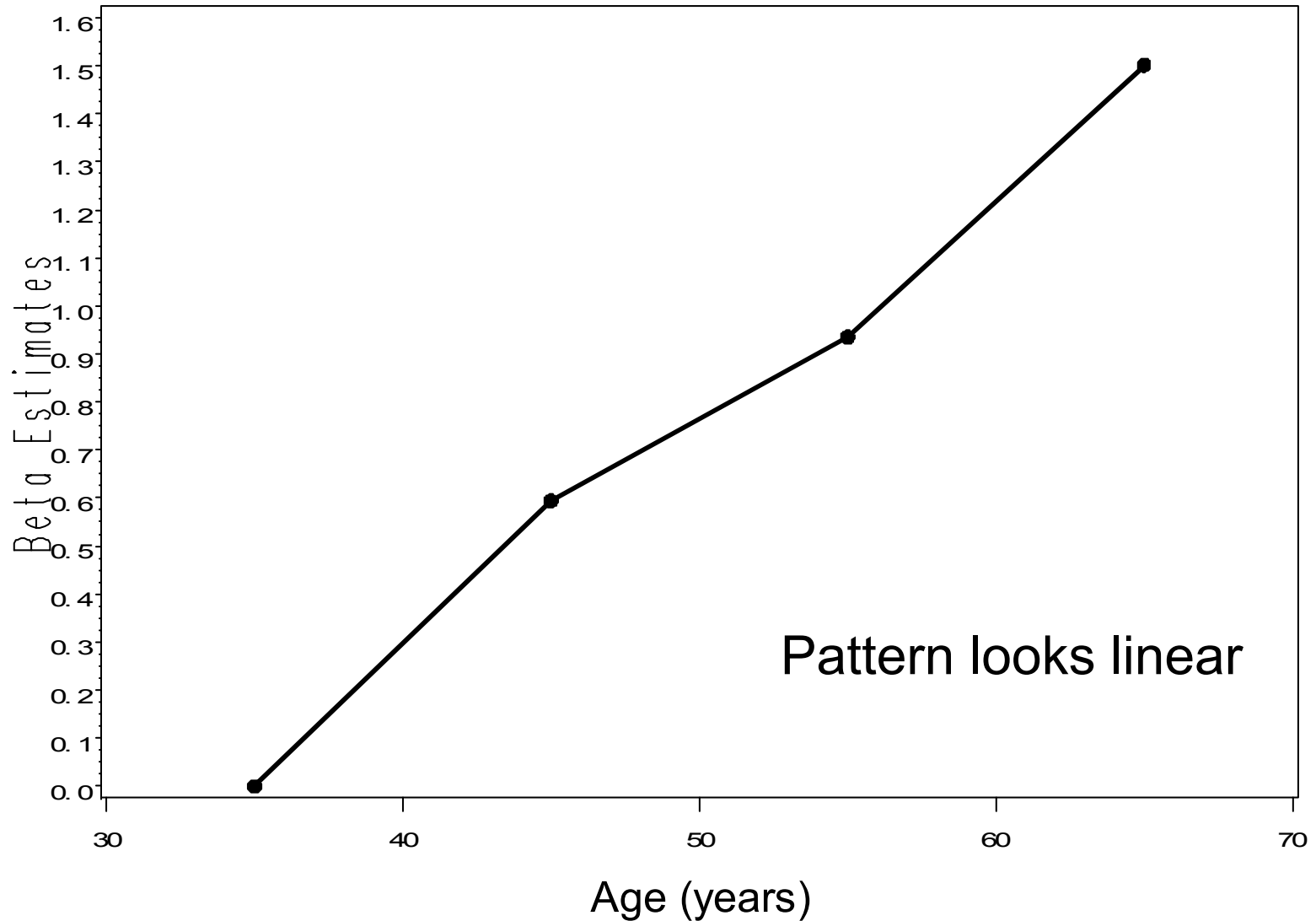
Assessing Functional Form (cont'd)

- Method 2 (loess line through martingale residuals):
  - Output martingale residuals from a model WITHOUT X.
    (proc phreg; model …; output out=temp resmart=resids;)
  - Fit a loess line through the martingale residuals, as a function of X.

    (ods output ScoreResults=temp2;

     proc loess data=temp; model resids=X; score;   run;)
  - Plot martingale residuals (with loess curve) by X.

    (proc gplot data=temp2;

     plot resids*X p_resids*X / overlay; run;)
  - Model X as appropriate (e.g., linear, quadratic, threshold), and re-check.

Plot of Martingale Residuals by Age, with Loess Line (Age not in model)

Loess line

Reference line at 0

Looks like HR increases with age

Martingale Residuals

Age (years)
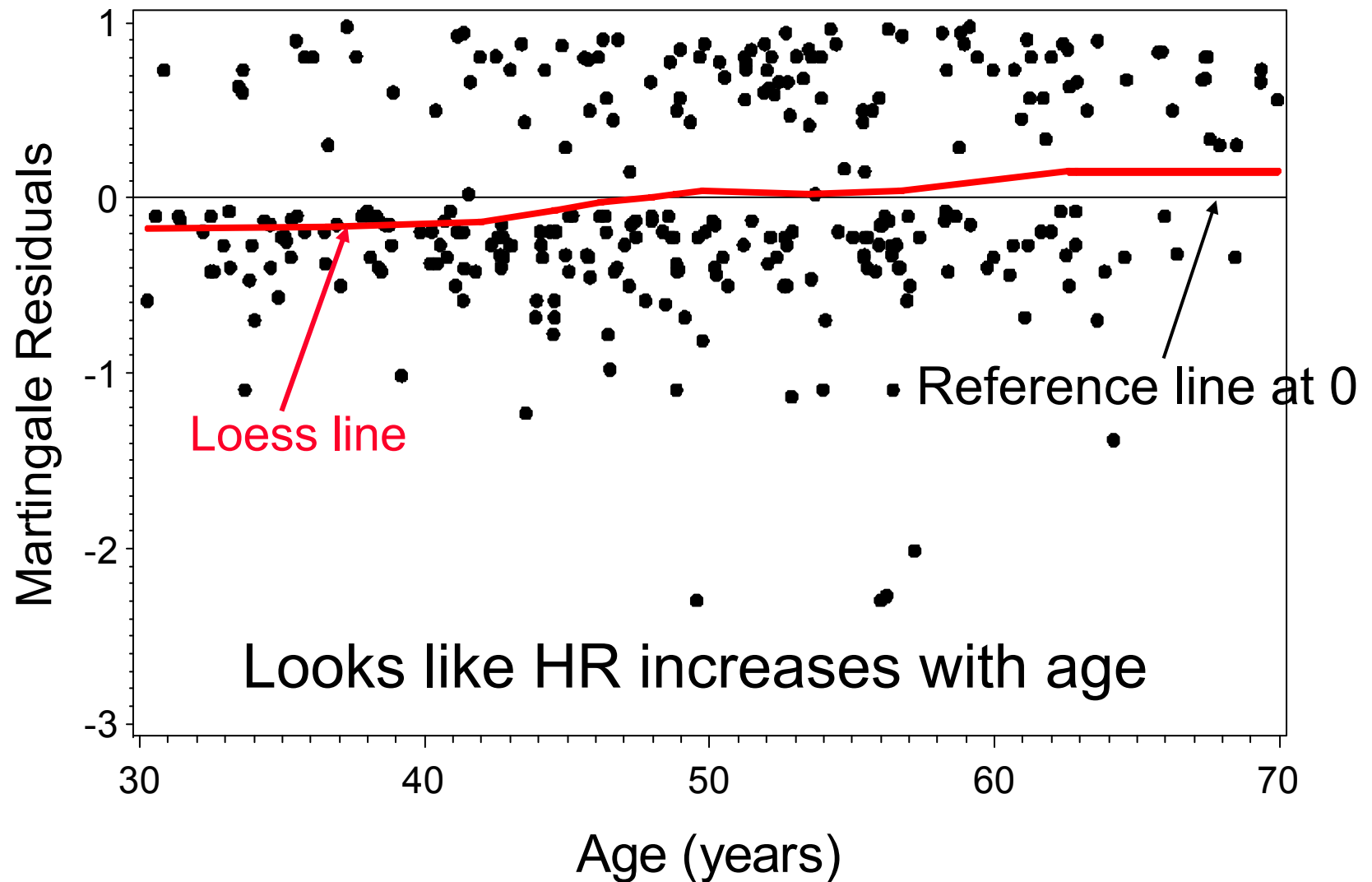
Plot of Martingale Residuals by Age, with Loess Line (Age in model as linear)

Linear age provides a good fit.
The loess line wiggles around zero – no trend.

Assessing Functional Form (cont'd)

- Method 3 (ASSESS option of **proc phreg** plots cumulative sums of martingale residuals against X (to check functional form) or the observed score process against Time (to check PH):

- The following code checks Age for functional form.

```
ods html; ods graphics on; /*required!*/
proc phreg data=pbc;
 assess var=(age_yrs) / npaths=50
    CRpanel;
 model logfuday*status(0) = sex age_yrs
    hepatom;
run;
ods graphics off; ods html close;
```
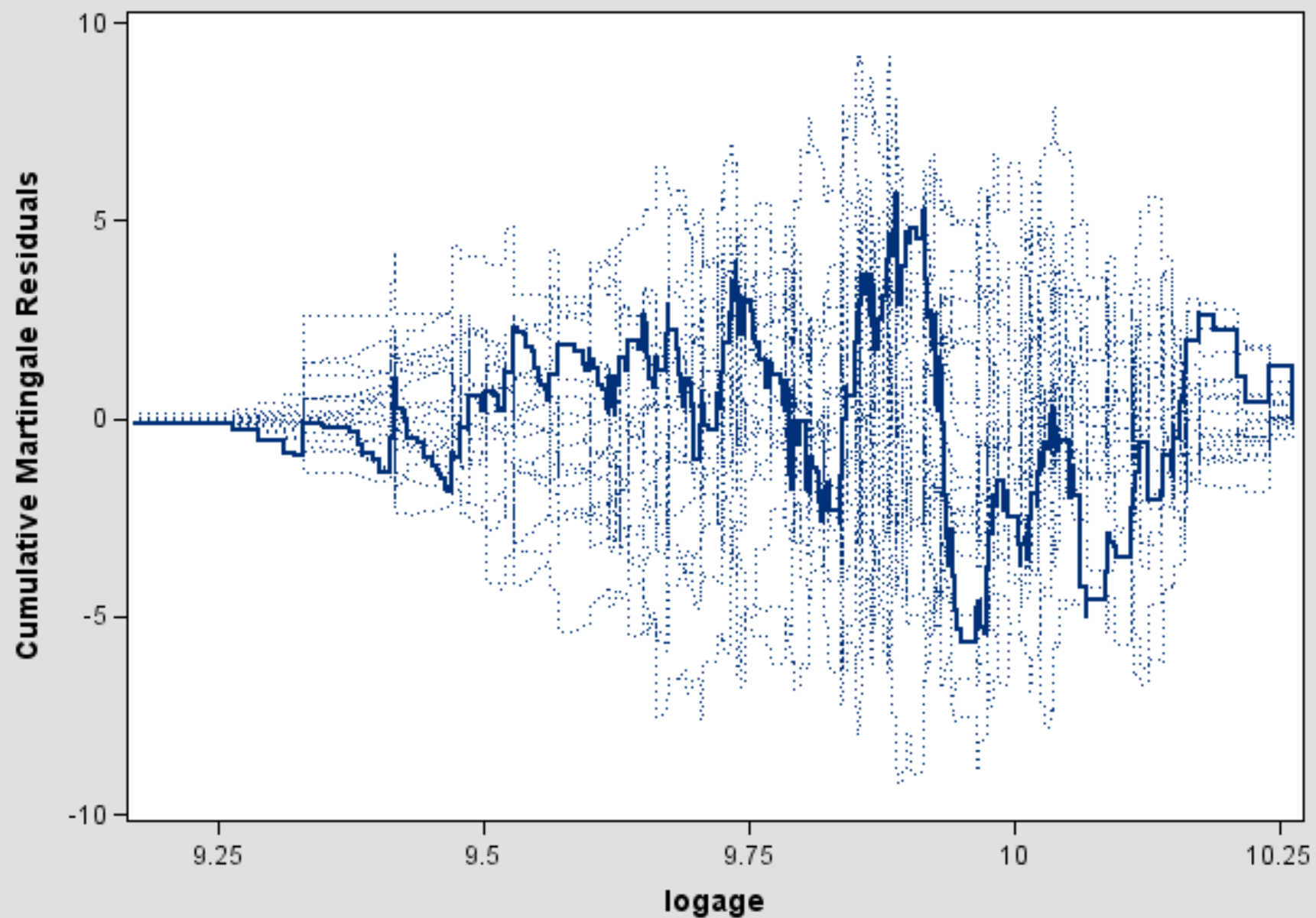
Checking Functional Form for age_yrs
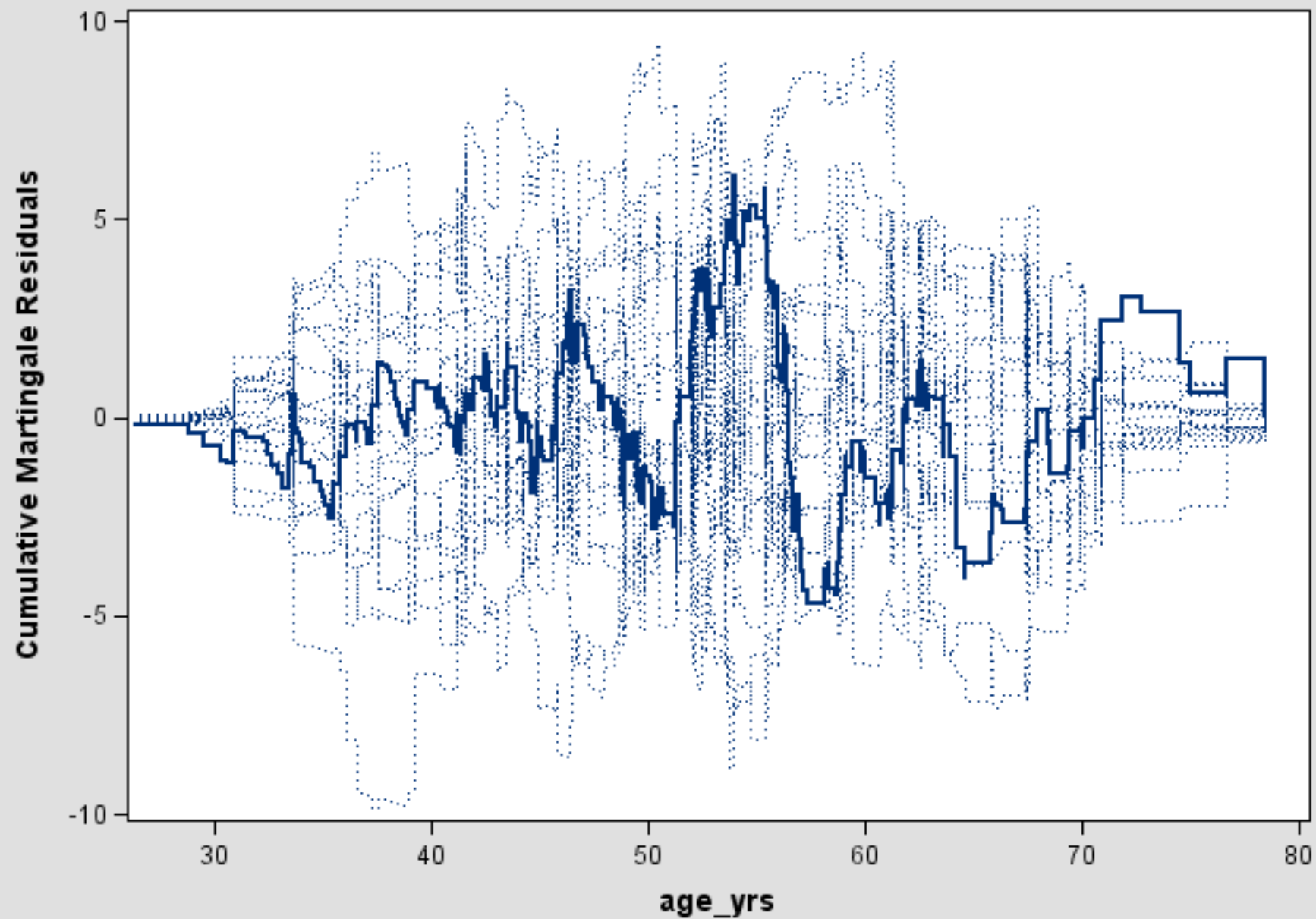
Assessing the Cumulative Martingale Residual Plot

- The plot shows the observed curve for Age to be within the distribution of the simulated cumulative martingale residual curves, indicating acceptable fit with linear age.

- Note that ASSESS cannot check functional form with a variable out of the model. It must be included in the model in some form.

- To try to illustrate a bad fit, we try log(Age), Age^2, and Age^5. Only Age^5 shows poor fit.
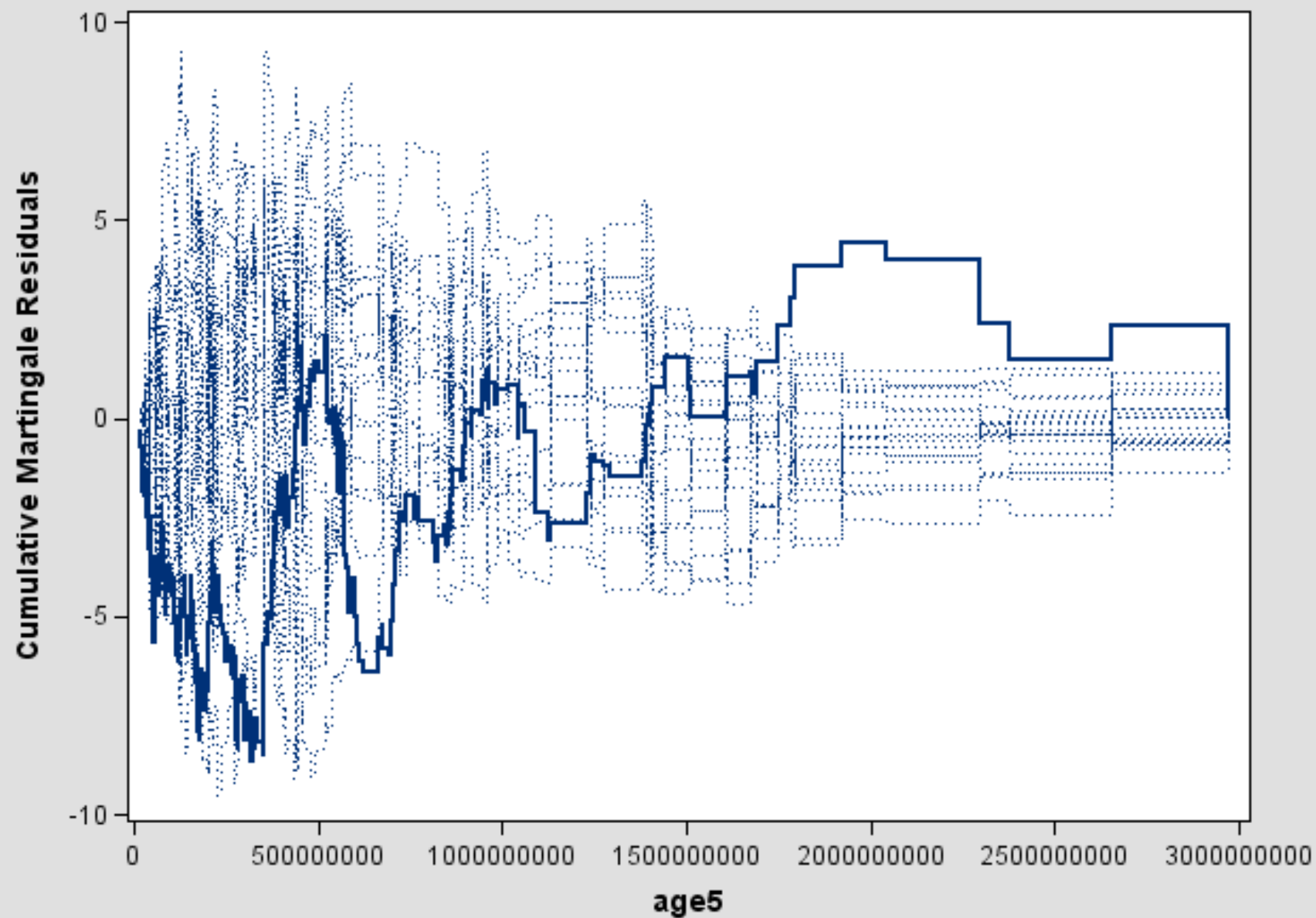
Checking Functional Form for logage

Checking Functional Form for age2
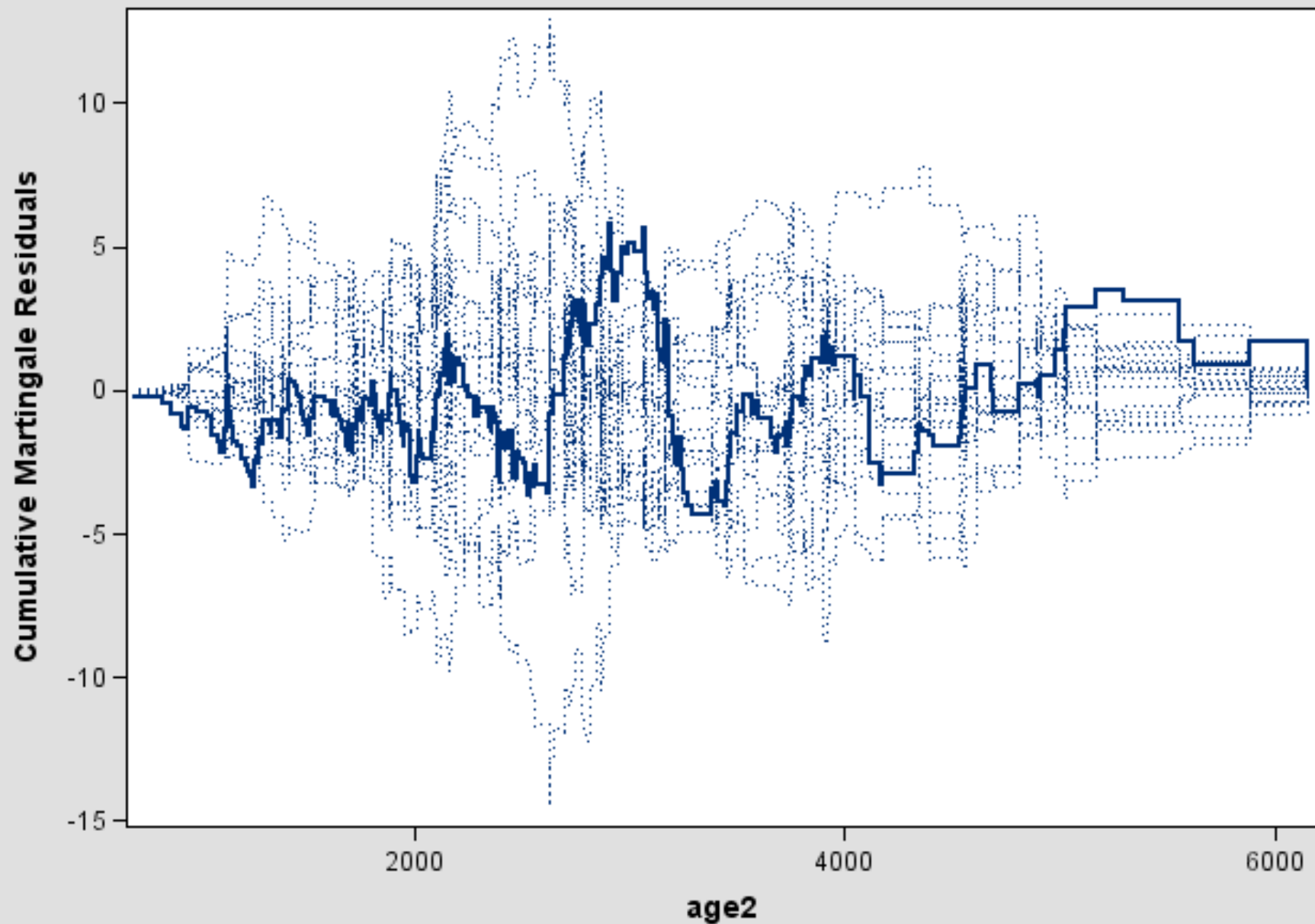
Checking Functional Form for age5

The Resample option of ASSESS

- The Resample option of ASSESS gives
  - a test of the functional form
  - A test of PH
- Tests are based on a Kolmogorov-type supremum test using 1000 simulated patterns.
- ASSESS var=(age_yrs) PH / resample;

Supremum Test for Functional Form

Variable	Maximum Absolute Value	Reps	Pr > MaxAbsVal
age_yrs	6.0767	1000	0.6640

Supremum Test for Proportionals Hazards Assumption

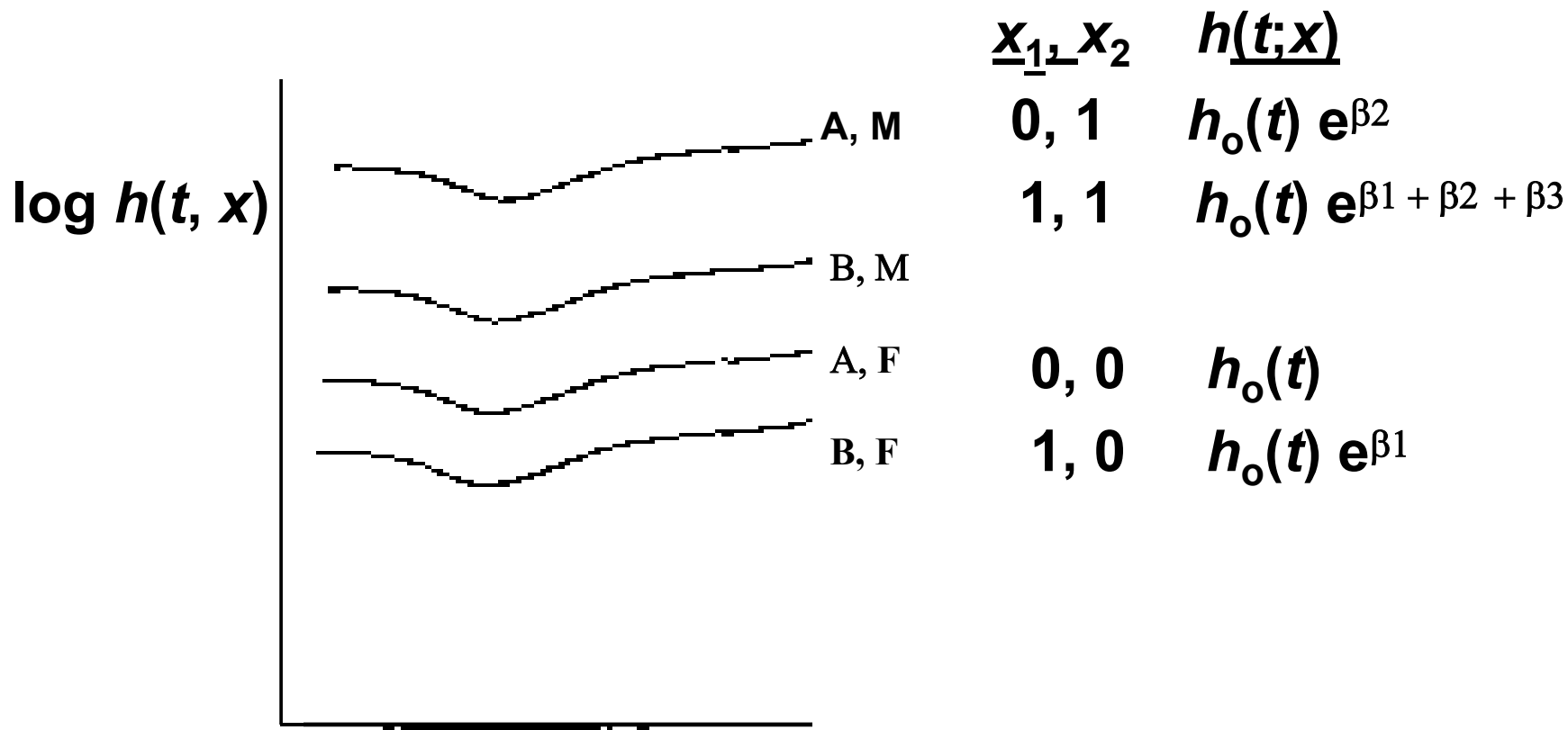| Variable	Maximum Absolute Value	Reps	Pr > MaxAbsVal
SEX	0.5985	1000	0.9930
HEPATOM	0.5504	1000	0.9920
age_yrs	0.5587	1000	0.9950

Summary of ASSESS Option

- The ASSESS option is a useful tool, but should be used in conjunction with other checks for functional form and PH.

- The cumulative martingale residual plots are not very sensitive for fine-tuning functional form. They can show grossly incorrect forms.

- We recommend martingale residuals (not cumulative), with a loess line to show functional form.

Covariate Interactions

- In many types of models, covariate interactions can be a challenge to interpret and present.

- With linear or logistic regression models, interaction plots are useful.

- With the Cox model, interaction plots, like variable effects, are based on Hazard Ratios.

Two dichotomous covariates: With interaction:

$$h(t;x) = h_o(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2\}$$

x_1, x_2	$h(t;x)$
A, M	0, 1	$h_o(t)\,e^{\beta 2}$
	1, 1	$h_o(t)\,e^{\beta 1 + \beta 2 + \beta 3}$
B, M		
A, F	0, 0	$h_o(t)$
B, F	1, 0	$h_o(t)\,e^{\beta 1}$

log $h(t, x)$

Presenting Covariate Interactions

The hypothetical plot above cannot be drawn with data because we don't estimate $h_o(t)$.

Option 1: Present interactions using hazard ratios separately within each level of one covariate.

Let β_1 = -0.3 (trt), β_2 = 0.7 (gender), and

β_3 = -0.2 (interaction)

– Males: HR(trt B vs. trt A) = exp($\beta_1 + \beta_3$)

= exp(-0.3 - 0.2) = exp(-0.5) = 0.61

– Females: HR(trt B vs. trt A) = exp(β_1)

= exp(-0.3) = 0.74

– Trt B better than A, but larger effect in males.

Presenting Covariate Interactions

Option 2: Compare all subgroups to a single baseline group. These hazard ratios can be plotted. The reference group is Females on treatment A.
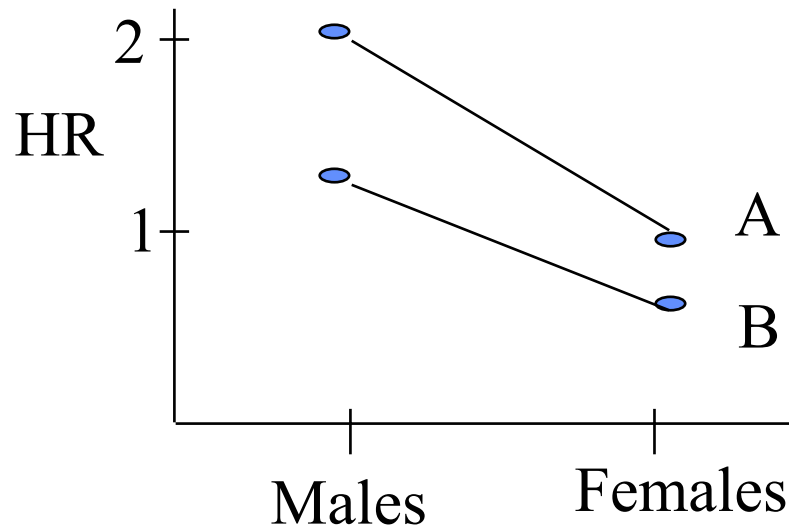

	HR	
Males		
A	2.0	$= e^{\beta 2}$
B	1.2	$= e^{\beta 1 + \beta 2 + \beta 3}$
Females		
A	1.0	
B	0.7	$= e^{\beta 1}$

Presenting Covariate Interactions with Continuous Covariates

- For an interaction between a continuous and a categorical covariate, plot the HR by the continuous covariate, with separate lines for the levels of the categorical covariate.

- For an interaction between two continuous covariates, plot the HR by one of the the continuous covariate, with separate lines for selected values of the other covariate.

A striking interaction between age and severe edema in the PBC dataset.
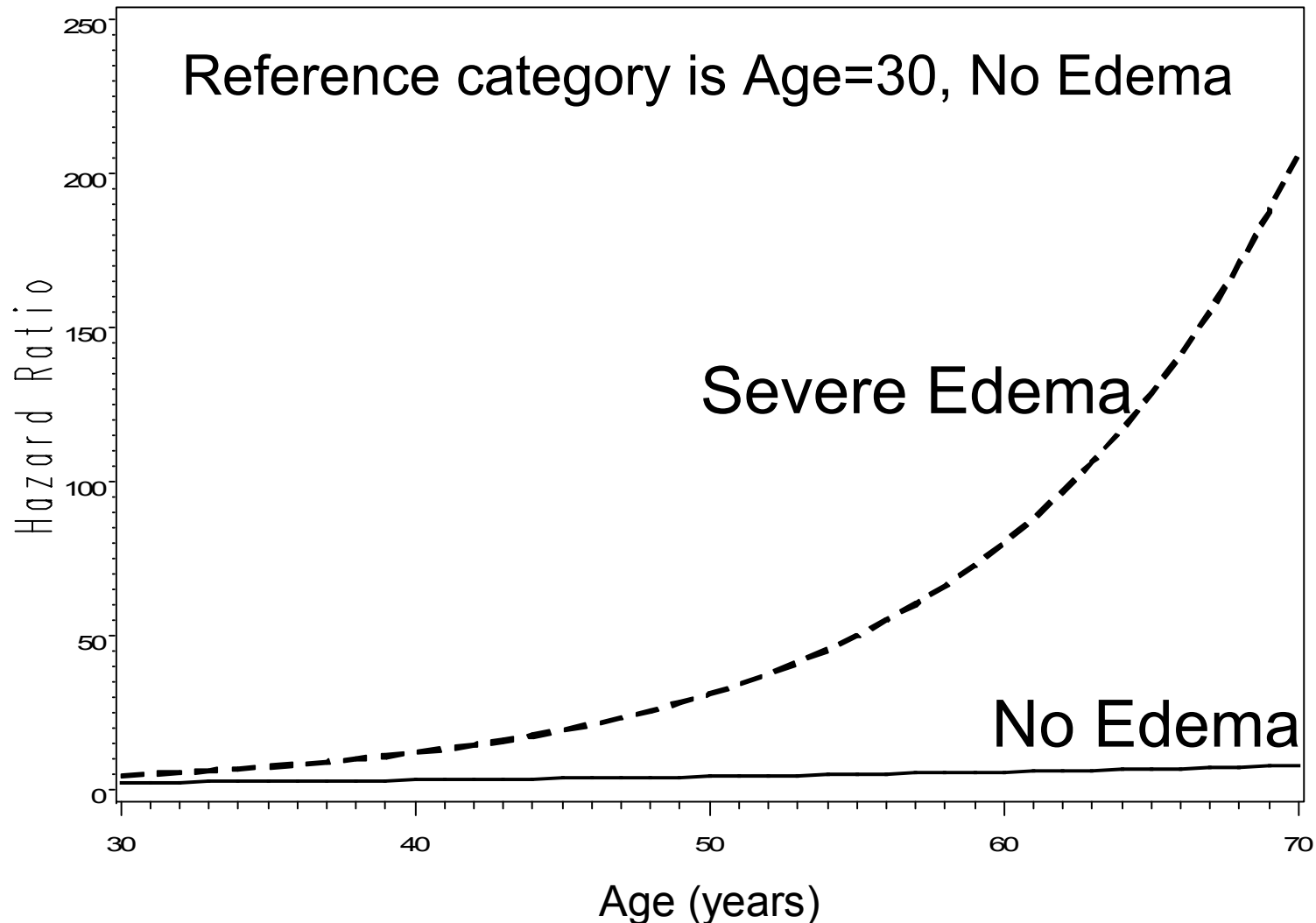
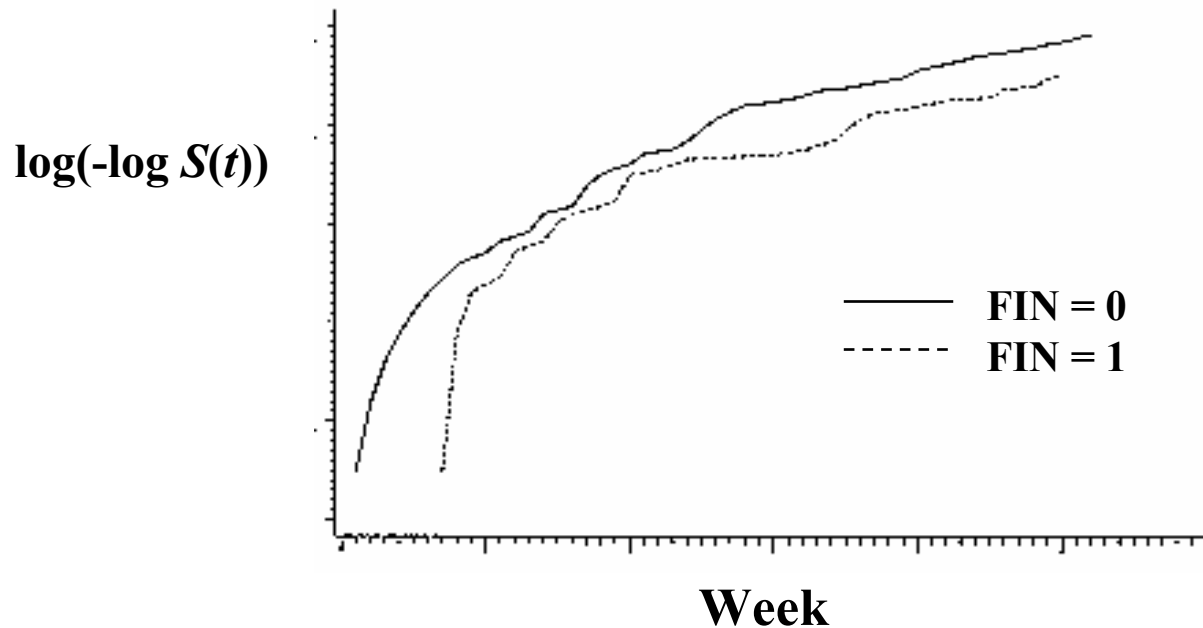# Checking Proportional Hazards (PH)

- Graphical methods to check PH

- Using time-dependent covariates to test PH

- Other tests for PH

Checking Proportional Hazards

- Graphical methods
  - Plot ln(-ln(S(t))) vs. t or ln(t) and look for parallelism.
  - Plot Observed and predicted S(t) and look for close fit.
  - Use the PH graph in the ASSESS option of Proc PHREG
  - Plot scaled Schoenfeld residuals vs time (schoen macro)
- Time-dependent covariates
  - Add time*covariate "interactions" to the model to fit non-PH. If the coefficient for the time-dependent variable is significantly different from zero, non-PH is present.
  - If significant non-PH is found, this model can be kept to fit and interpret the non-PH.
- Other tests for PH
  - Test based on resampling using the ASSESS option.
  - Test based on scaled Schoenfeld residuals (schoen macro)

Proportional Hazards: Graphical Check #1
Plot ln(-ln(S(t))) vs. t or ln(t) and look for parallelism.

Parallel curves $\Rightarrow$ PH
Use Kaplan-Meier estimate for S(t).
This plot shows reasonable fit to the PH assumption.

Proportional Hazards: Graphical Check #1

- Interpreting plots is subjective. In general, conclude PH unless a distinct pattern of non-parallelism (e.g., crossing) is seen.

- Intertwined lines with no distinct pattern may simply indicate no difference between groups.

- Adjusting for other covariates may be needed.
  - Example: To check PH for treatment, adjusted for age:
  - Run a Cox model with age as a covariate, stratified by treatment.
  - Output the estimated survivor functions for each treatment group at the overall mean age.
  - Plot $\ln(-\ln(\hat{S}(t)))$ for each treatment group vs $\ln(t)$ and check for parallelism.

SAS® Code for log(-log(S(t))) Plots

Unadjusted PH check for Treatment:

```
    Proc lifetest data=data1  plots=(lls);
       time days*status(0);
       strata treat;  run;
```

Adjusted PH check for Treatment, adjusted for Age:

```
       data covs; age=52; run;  /*Overall mean age*/
Proc phreg data=pbc;
       model days*status(0) = age;
       strata treat;
       baseline out=temp covariates=covs loglogs=lls;
    Proc plot data=temp; plot lls*days=treat; run;
```

Proportional Hazards: Graphical Check #2

Plot Observed and predicted S(t) and look for close fit.
(Only feasible with small number of covariates.)

Observed vs. predicted: Close?

——— Predicted for males (gender in model)

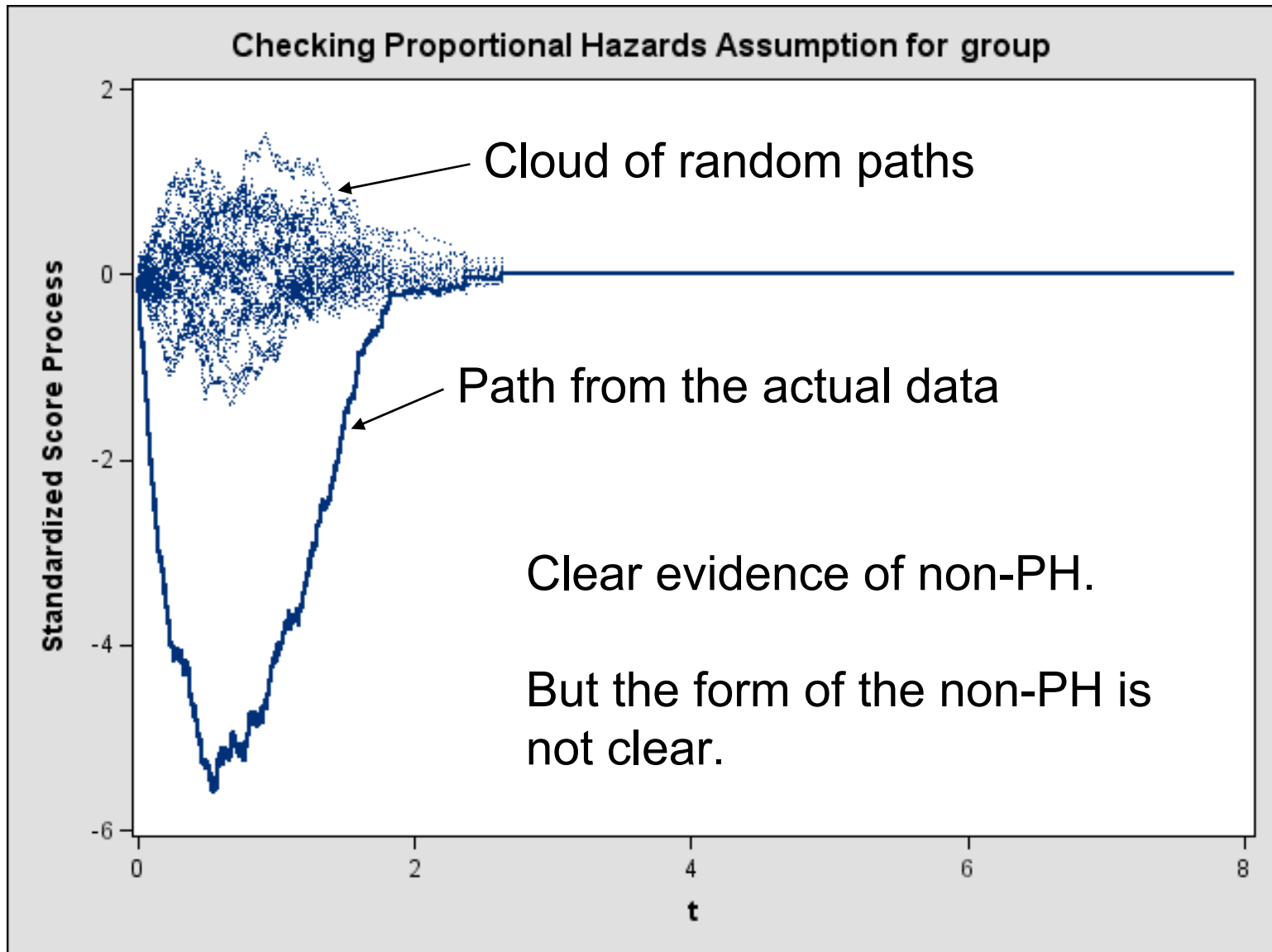- - - - - - Observed for males
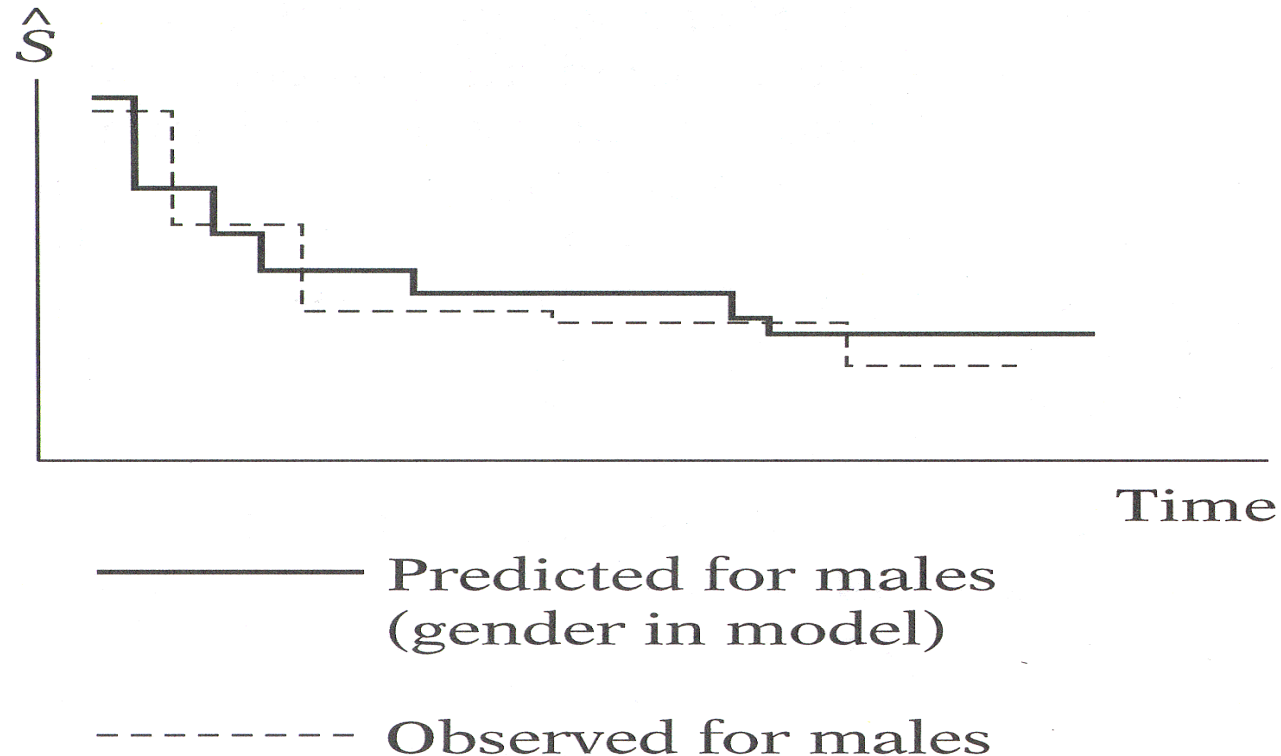
• Predicted is from Cox model.   Observed is KM.

(Figure from Kleinbaum)

Checking PH using ASSESS option

- The ASSESS option of Proc PHREG plots the cumulative score residuals against time to check PH.

- This is a "tied down" Brownian process, or Brownian bridge, meaning that the values always start and end at zero.

- Random "paths" are generated under PH.

- The path from the actual data is compared to the randomly-generated paths under PH.

- If the actual path is within the cloud of random paths, it indicates PH.

Checking PH using ASSESS option

Checking Proportional Hazards Assumption for group

Cloud of random paths

Path from the actual data

Clear evidence of non-PH.

But the form of the non-PH is not clear.

Checking PH using macro SCHOEN

- The SAS® macro, SCHOEN, gives a different graphical check for PH.

- Consider the possibility that the β coefficient for a given covariate, β_k, changes over time, thus giving a non-constant hazard ratio.

- Macro SCHOEN uses a *scaled* Schoenfeld residual, multiplying the vector of Schoenfeld residuals by the inverse of their covariance matrix.

- This scaled residual, r_{ik}^*, added to β_k, is an estimate of the time-dependent β coefficient: $r_{ik}^* + \beta_k \approx \beta_k(t_i)$.

- $r_{ik}^* + \beta_k$ is plotted against time, or a function of time.

- PH is indicated by a flat pattern around Y=0.

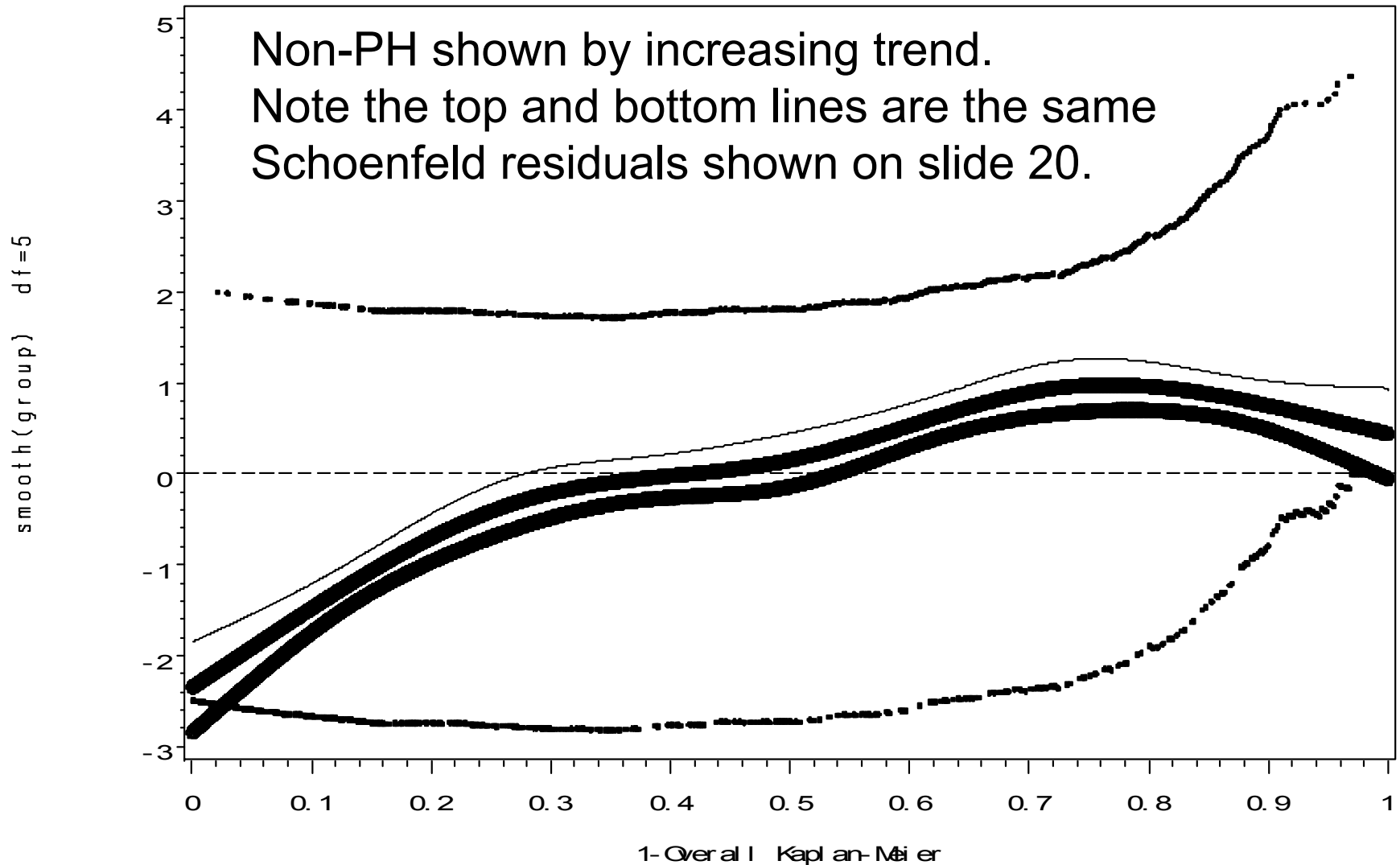- Non-PH is indicated by any deviation from a flat line at Y=0.

Which function of time?

- The Schoenfeld residuals can be plotted against any function of time, such as raw, log-transformed, or rank-transformed.

- The pattern shown over time indicates the form of non-PH.

- Different functions show different shapes, and some may be better for highlighting non-PH for a particular variable. Try more than one.

- Options available in the "schoen" macro are:
  - Raw time
  - Rank-transformed time
  - Time transformed by (1 - Kaplan-Meier) (Similar to probability integral transformation.)

Checking PH with macro SCHOEN
(KM-transformed time scale)

Scaled residuals(Bt) as a fcn of time.

Non-PH shown by increasing trend.
Note the top and bottom lines are the same
Schoenfeld residuals shown on slide 20.

1-Overall Kaplan-Meier

schoen macro: event=cens time=t strata=
Xvars= group

Interpreting the SCHOEN Plot

- The previous plot clearly shows an increasing pattern, suggesting linear.

- The true hazard ratio is linearly increasing in log(t).

- The SCHOEN plot is more useful than the ASSESS plot in showing the appropriate functional form for a non-PH relationship.
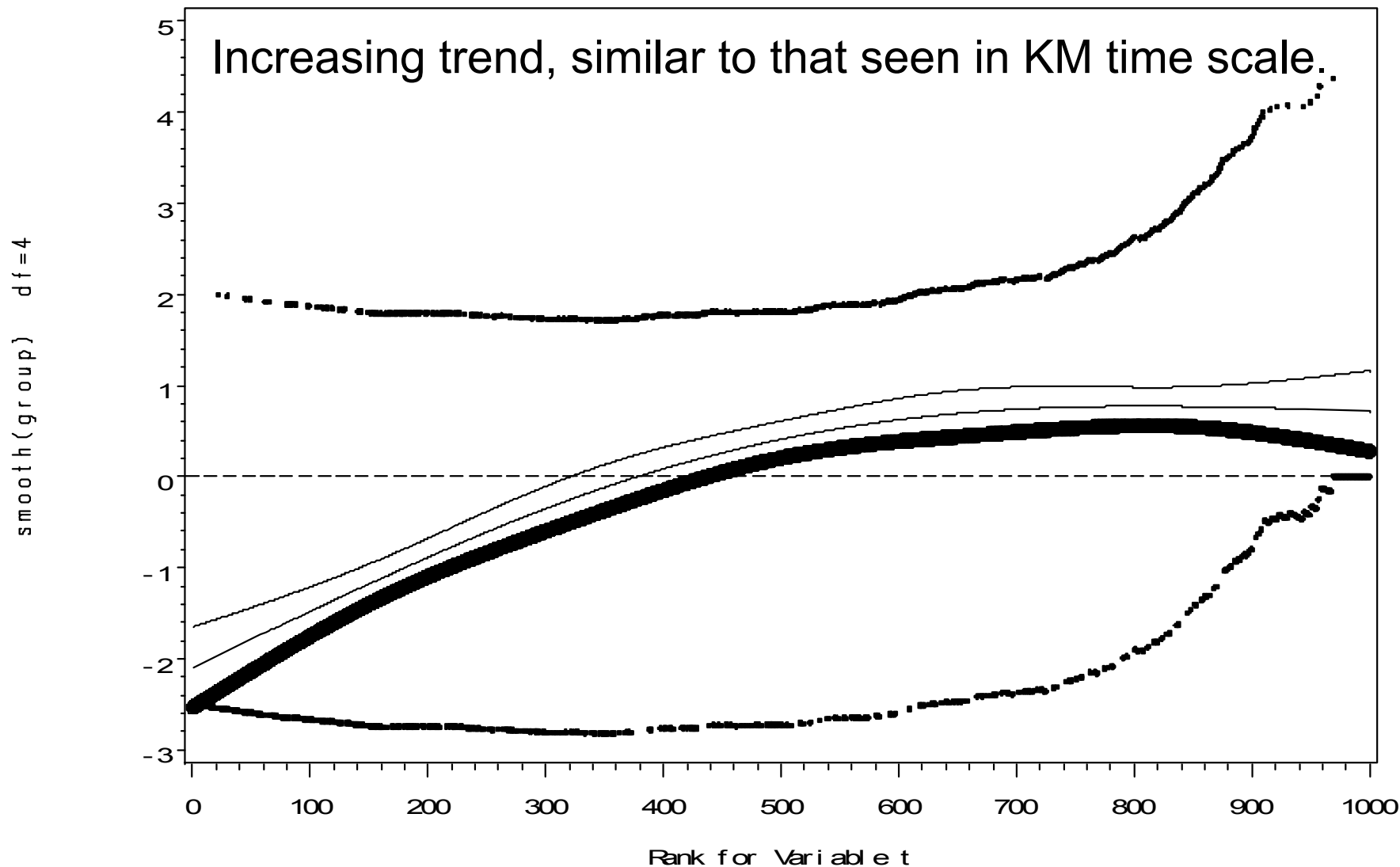
Macro SCHOEN with raw time scale

Scaled residuals(Bt) as a fcn of time.

SCHOEN plots are sensitive
to the time scale used.
Very few points beyond t=4.
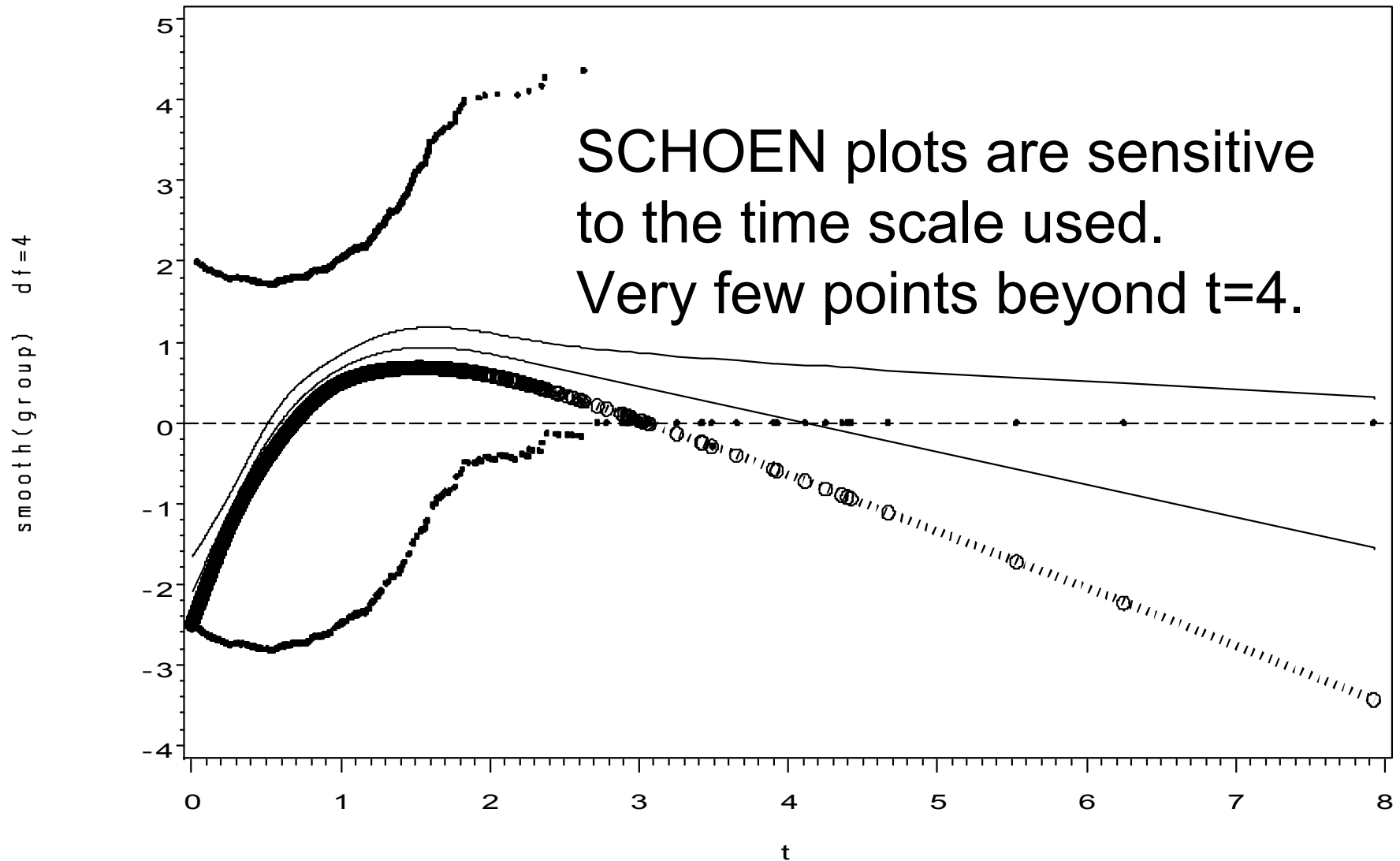
schoen macro: event=cens time=t strata=
Xvars= group

Macro SCHOEN with rank time scale

Scaled residuals(Bt) as a fcn of time.

Increasing trend, similar to that seen in KM time scale.

y-axis: smooth(group) df=4

x-axis: Rank for Variable t

schoen macro: event=cens time=t strata=
Xvars= group

Macro SCHOEN time scales

- SCHOEN plots are sensitive to the time scale used. Try more than one.
- If data are very skewed, it is often better to use the rank or KM time scale.

- Note: Virtually all tests for PH are based on the choice of a particular time function, $g(t)$, for the non-PH.
  - A test will be most powerful to detect non-PH based on the particular $g(t)$, and will have less power to detect non-PH of other forms.

Time-dependent covariates: Two types

- **Time-varying covariates: Covariate values change over time.**

  – Ex: For time to re-arrest after release from prison, a time-varying covariate would be whether the person is employed (0=no, 1=yes) at a given time.

  Cox model for x_1 = fixed covariate, x_2 = time-varying covariate:

$$h(t;x) = h_0(t)\exp\{\beta_1 x_1 + \beta_2 x_2(t)\}$$

- **Time × covariate interactions: used to test or model non-proportional hazards. We focus here on this type.**

$$h(t;x) = h_0(t)\exp\{\beta_1 x_1 + \beta_2 x_1 t\}$$

  The hazard ratio for x_1=1 vs. x_1=0 changes (either increases or decreases) as t increases.

Time*Covariate Interactions

$$h(t;x) = h_o(t) \exp\{\beta_1 x + \beta_2 x\, t\}$$

$\beta_2 > 0$ **=>** HR increasing with time

$\beta_2 < 0$ **=>** HR decreasing with time

$\beta_2 = 0$ **=>** HR constant with time **=> PH**

Add $x*t$ to the model to **test** PH (test H_0: $\beta_2=0$).

If β_2 significant, then leave $x*t$ in the model
   (to **model** the non-PH).

Some authors suggest other interactions, e.g.,
   $x*log(t)$ or x*I[t>c] (heavyside function). Use
   whatever fits best.

SAS® Code for Time*Covariates

```
proc phreg;
  model week*arrest(0) = age fin TDfin;
  TDfin = fin*week;      ****
  run;


****or:  TDfin = fin*log(week);
      or:  TDfin = fin*(week>25);  (for a different
    hazard ratio before vs. after week 25)
```

Stratification vs. Time*Covariate Interactions for Handling Non-PH

- **Time*Covariate Interaction**
  - Must choose a particular form, such as $x*t$ or $x*\log(t)$.
  - If this form is correct, yields more efficient estimates of other βs. (robustness vs. efficiency)
  - The changing HR over time can be presented and interpreted.
- **Stratification**
  - Takes less computation time
  - Models any non-PH relationship, not just specific forms
  - No inference is possible for the stratification variable; only makes sense for "nuisance variables".

Checking PH with Many Covariates

- Check PH for each covariate separately.
- If interactions are present, check PH over all interaction subgroups (e.g., Males, A; Females, A; Males, B; Females, B)
- If collinearity (confounding, treatment imbalance) is present among covariates: To check PH for x_1, estimate $S_i(t)$ for the levels of x_1 based on a Cox model stratified by x_1, with other covariates in the model. Plot $\log(-\log \hat{S}_i(t))$ vs. $\log(t)$.

Difficulty of Checking PH

- In checking each covariate, we assume PH holds for the other covariates. Which covariate do we start with?

- If PH fails for a covariate, we should go back and re-check the others after adjusting for the non-PH of the first.

- A wrong functional form or a missing covariate can look like non-PH.

- Checking PH can be a difficult process.

- See Kleinbaum for more details.

Summary and Recommendations

- Check for outliers
  - Deviance residual plot
- Check for functional form of continuous covariates
  - Martingale residual plots
- Check for non-PH
  - Use log(-log(S(t))) plots (either unadjusted or adjusted)
  - Test time*covariate interactions
  - Use the "schoen" macro to plot $\beta_k(t_i)$ by time
- Checking assumptions takes time. Take the time.
- Checking can be never-ending, so balance is needed. Some checking is better than none.

The Cox Modeler's Blessing

May your continuous covariates all be linear,

and may all your covariates satisfy the proportional hazards assumption …

References

•Therneau TM and Grambsch PM. (2000). *Modeling Survival Data: Extending the Cox Model*, Springer.

•Allison PD. (1995), *Survival Analysis Using the SAS System: A Practical Guide*, SAS Institute Inc.

•Klein JP and Moeschberger ML. (1997), *Survival Analysis: Techniques for Censored and Truncated Data*, Springer-Verlag.

•Marubini E and Valsecchi MG. (1995), *Analysing Survival Data from Clinical Trials and Observational Studies*, John Wiley & Sons Ltd.

•Hosmer DW Jr and Lemeshow S. (1999), *Applied Survival Analysis*, Wiley.

•Kalbfleisch JD and Prentice RL. (2002), *The Statistical Analysis of Failure Time Data*, 2nd Edition, Wiley.