PH plots (interval-censored) — PH-assumption plots for interval-censored data

Description

`stintphplot` and `stintcoxnp` create plots that allow you to visually assess whether the proportional-hazards assumption is likely violated. These commands are for use with interval-censored survival-time data.

`stintphplot` plots $-\ln\{-\ln(\text{survival})\}$ curves for each level of a nominal or ordinal covariate specified in the `by()` option versus $\ln(\text{analysis time})$ for interval-censored data. These plots are often referred to as "log–log" survival plots. Optionally, the estimates can be adjusted for covariates. The proportional-hazards assumption is satisfied when the curves are parallel.

`stintcoxnp` plots Turnbull's nonparametric and Cox predicted survival curves for each level of a categorical covariate specified in the `by()` option for interval-censored data. The Turnbull nonparametric curve for interval-censored data is analogous to the Kaplan–Meier nonparametric curve for right-censored data. The closer the nonparametric estimates are to the Cox estimates, the less likely it is that the proportional-hazards assumption has been violated.

Quick start

Log–log plot of survival

Check for parallel lines in plot of $-\ln\{-\ln(\text{survival})\}$ versus $\ln(\text{analysis time})$ for each category of covariate `a` for interval-censored event-time data with interval endpoints `ltime` and `rtime`

```
stintphplot, interval(ltime rtime) by(a)
```

Same as above, but adjust for average values of covariates `x1` and `x2`

```
stintphplot, interval(ltime rtime) by(a) adjustfor(x1 x2)
```

Same as above

```
stintphplot, interval(ltime rtime) by(a) adjustfor(x1 x2, atomeans)
```

Adjust for `x1` $= 0$ and `x2` $= 0$

```
stintphplot, interval(ltime rtime) by(a) adjustfor(x1 x2, atzeros)
```

For multiple-event data, check for parallel lines for each category of covariate `a` for event `1`

```
stintphplot if event==1, interval(ltime rtime) by(a)
```

Turnbull's nonparametric and Cox predicted survival plots

Compare Turnbull's nonparametric curve with predicted survival from the Cox model for each category of covariate `a` for interval-censored event-time data with interval endpoints `ltime` and `rtime`

```
stintcoxnp, interval(ltime rtime) by(a)
```

Same as above, but create separate plots for each level of `a`

```
stintcoxnp, interval(ltime rtime) by(a) separate
```

For multiple-event data, compare Turnbull's nonparametric curve with predicted survival from the Cox model for event 2

```
stintcoxnp if event==2, interval(ltime rtime) by(a)
```

Menu

stintphplot

Statistics > Survival analysis > Regression models > Assess PH assumption for interval-censored data

stintcoxnp

Statistics > Survival analysis > Regression models > Nonparametric versus Cox predicted survival

Syntax

Check proportional-hazards assumption:

Log–log plot of survival

```
stintphplot [if] , interval(tl tu) {by(varname) | strata(varname)}
    [phplot_options]
```

Nonparametric and Cox predicted survival plots

```
stintcoxnp [if] , interval(tl tu) by(varname) [stintcoxnp_options]
```

phplot_options	Description
Main	
* `interval(`t_l t_u`)`	lower and upper endpoints for the event-time interval
* `by(`*varname*`)`	fit separate Cox models for levels of *varname*
* `strata(`*varname*`)`	fit stratified Cox model; requires `adjustfor()`
`adjustfor(`*varlist*`[ , `*suboptions*`])`	adjust the estimates to specific values of *varlist*; default is overall means
`nonegative`	plot ln{−ln(survival)} instead of −ln{−ln(survival)}
`nolntime`	plot curves against analysis time
Plot	
`plot`#`opts(`*phplot_plot_options*`)`	affect rendition of the #th connected line and #th plotted points
Add plots	
`addplot(`*plot*`)`	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than `by()` documented in [G-3] ***twoway_options***

*`interval()` is required with `stintphplot`.

*Either `by(`*varname*`)` or `strata(`*varname*`)` and `adjustfor(`*varlist*`)` is required with `stintphplot`.

phplot_plot_options	Description
cline_options	change look of lines or connecting method
marker_options	change look of markers (color, size, etc.)

stintcoxnp_options	Description
Main	
* `interval(`t_l t_u`)`	lower and upper endpoints for the event-time interval
* `by(`*varname*`)`	specify a categorical covariate
`separate`	show group-specific plots as subgraphs on one graph; default is to overlay these plots on one graph
Nonparametric plot	
`npopts(`*stintcoxnp_plot_options*`)`	affect rendition of the nonparametric curve
`np#opts(`*stintcoxnp_plot_options*`)`	affect rendition of the #th nonparametric curve; not allowed with `separate`
Cox predicted plot	
`coxopts(`*stintcoxnp_plot_options*`)`	affect rendition of the Cox predicted curve
`cox#opts(`*stintcoxnp_plot_options*`)`	affect rendition of the #th Cox predicted curve; not allowed with `separate`
Add plots	
`addplot(`*plot*`)`	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than `by()` documented in [G-3] ***twoway_options***
By options	
`byopts(`*byopts*`)`	how subgraphs are combined, labeled, etc.

*`interval()` and `by()` are required with `stintcoxnp`.

stintcoxnp_plot_options	Description
connect_options	change look of connecting method
marker_options	change look of markers (color, size, etc.)

Options

Options are presented under the following headings:

Options for stintphplot
Options for stintcoxnp

Options for stintphplot

Main

`interval(`t_l t_u`)`; see [ST] **stintcox**. `interval()` is required.

by(*varname*) specifies a categorical covariate for which the proportional-hazards assumption is to be checked. Without the adjustfor() option, stintphplot computes nonparametric estimates of the survivor functions for each category of *varname* and produces log–log plots. With adjustfor(), the command fits a separate Cox model for each category of *varname* to produce log–log plots. When the plotted curves are parallel for all categories of *varname*, the proportional-hazards assumption is considered to be met for *varname*. Either by() or strata() is required with stintphplot.

strata(*varname*) is an alternative to by(). Rather than fitting separate Cox models for each value of *varname*, strata() fits one stratified Cox model. You must also specify adjustfor() with the strata() option.

adjustfor(*varlist*[, *suboptions*]) adjusts the estimates of the survivor function to specific values of *varlist*. The default is to adjust to overall mean values of covariates. adjustfor() can be specified with by(); it is required with strata().

suboptions are atomeans (the default), atmeans, atzeros, atbase, and at(); see [ST] ***adjustfor_option***.

nonegative specifies that ln{−ln(survival)} be plotted instead of −ln{−ln(survival)}.

nolntime specifies that curves be plotted against analysis time instead of against ln(analysis time).

Plot

plot#opts(*phplot_plot_options*) affects the rendition of the #th connected line and #th plotted points; see [G-3] ***cline_options*** and [G-3] ***marker_options***.

Add plots

addplot(*plot*) provides a way to add other plots to the generated graph; see [G-3] ***addplot_option***.

Y axis, X axis, Titles, Legend, Overall

twoway_options are any of the options documented in [G-3] ***twoway_options***, excluding by(). These include options for titling the graph (see [G-3] ***title_options***) and for saving the graph to disk (see [G-3] ***saving_option***).

Options for stintcoxnp

Main

interval(t_l t_u); see [ST] **stintcox**. interval() is required.

by(*varname*) specifies a categorical covariate *varname* for which the agreement between the Cox predicted and nonparametric survivor functions is to be explored. The nonparametric and Cox predicted survivor functions are plotted for each level of *varname*. The agreement between the two survival curves across all levels indicates that the proportional-hazards assumption for *varname* is reasonable. by() is required.

separate produces separate plots of Cox predicted and nonparametric survivor functions for each value of the variable specified with by().

Nonparametric plot

npopts(*stintcoxnp_plot_options*) affects the rendition of the nonparametric curve; see [G-3] ***connect_options*** and [G-3] ***marker_options***.

`np#opts(`*stintcoxnp_plot_options*`)` affects the rendition of the #th nonparametric curve; see [G-3] ***connect_options*** and [G-3] ***marker_options***. This option is not allowed with `separate`.

Cox predicted plot

`coxopts(`*stintcoxnp_connect_options*`)` affects the rendition of the Cox predicted curve; see [G-3] ***connect_options*** and [G-3] ***marker_options***.

`cox#opts(`*stintcoxnp_connect_options*`)` affects the rendition of the #th Cox predicted curve; see [G-3] ***connect_options*** and [G-3] ***marker_options***. This option is not allowed with `separate`.

Add plots

`addplot(`*plot*`)` provides a way to add other plots to the generated graph; see [G-3] ***addplot_option***. `addplot()` is not allowed when the graph contains subgraphs.

Y axis, X axis, Titles, Legend, Overall

twoway_options are any of the options documented in [G-3] ***twoway_options***, excluding `by()`. These include options for titling the graph (see [G-3] ***title_options***) and for saving the graph to disk (see [G-3] ***saving_option***).

By options

`byopts(`*byopts*`)` affects the appearance of the combined graph when `by()` and `separate` are specified, including the overall graph title and the organization of subgraphs. See [G-3] ***by_option***.

Remarks and examples

The underlying assumption for the Cox proportional hazards model is that the hazard ratio is constant over time. Therefore, it is important to evaluate the validity of this assumption. `stintphplot` and `stintcoxnp` are graphical tools for assessing violations of the proportional-hazards assumption for interval-censored event-time data. For proportional-hazards-assumption plots for right-censored data, see [ST] **PH plots (right-censored)**.

`stintphplot` plots $-\ln\{-\ln(\text{survival})\}$ curves for each category of a nominal or ordinal covariate versus $\ln(\text{analysis time})$. These are often referred to as "log–log" plots. Optionally, these estimates can be adjusted for covariates. If the plotted lines are reasonably parallel, the proportional-hazards assumption has not been violated.

Another graphical method of evaluating the proportional-hazards assumption is to plot the nonparametric maximum-likelihood estimation (NPMLE) survival curves proposed by Turnbull (1976) and compare them with the Cox predicted curves for each level of a categorical variable. This plot is produced by `stintcoxnp`. When the two curves are close together, the proportional-hazards assumption has not been violated.

For a single categorical covariate in a Cox model, you can use `stintphplot` and `stintcoxnp` to check the proportional-hazards assumption. In the presence of multiple covariates, you can use only `stintphplot`, and you should adjust for covariates by including all but the tested covariate in the `adjustfor()` option.

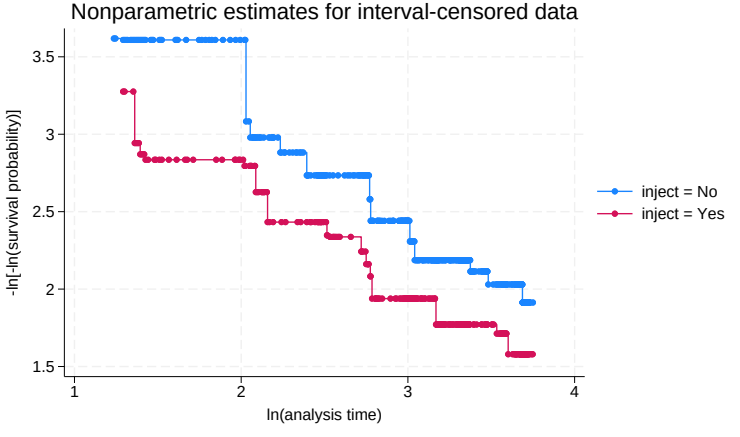
To test the proportional-hazards assumption for interval-censored data, you do not need to fit a Cox proportional hazards model using `stintcox` or a marginal Cox proportional hazards model using `stmgintcox` before issuing `stintphplot` or `stintcoxnp`.

With interval-censored data, the analysis time is formed by the unique values of lower and upper endpoints t_l and t_u, which must be specified in the `interval()` option.

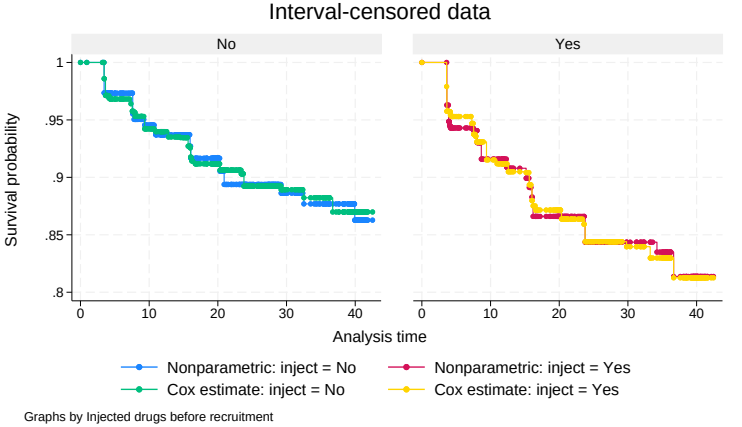
▷ Example 1: Proportional-hazards assumption holds for a model with a single covariate

Consider the dataset described in example 1 of [ST] **stintcox**. Suppose we want to check whether the proportional-hazards assumption holds for a single binary variable `inject`. We will use `stintphplot` first and then `stintcoxnp`:

```
. use https://www.stata-press.com/data/r19/idu
(Modified Bangkok IDU Preparatory Study)
. stintphplot, interval(ltime rtime) by(inject)
Computing nonparametric estimates for inject = No ...
Computing nonparametric estimates for inject = Yes ...
```




```
. stintcoxnp, interval(ltime rtime) by(inject) byopts(style(altleg)) separate
Computing nonparametric estimates ...
Computing Cox estimates ...
```

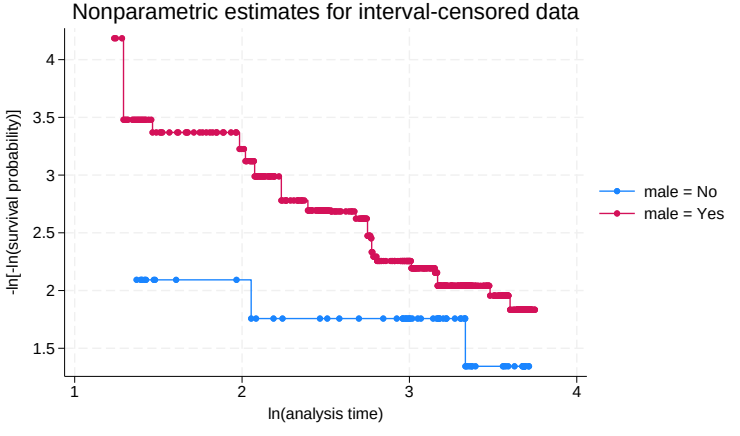


The `stintphplot` plot shows that the two lines are parallel, implying that the proportional-hazards assumption holds for `inject`. This is confirmed in the `stintcoxnp` plot, where the Turnbull nonparametric estimates and Cox estimates are close together.

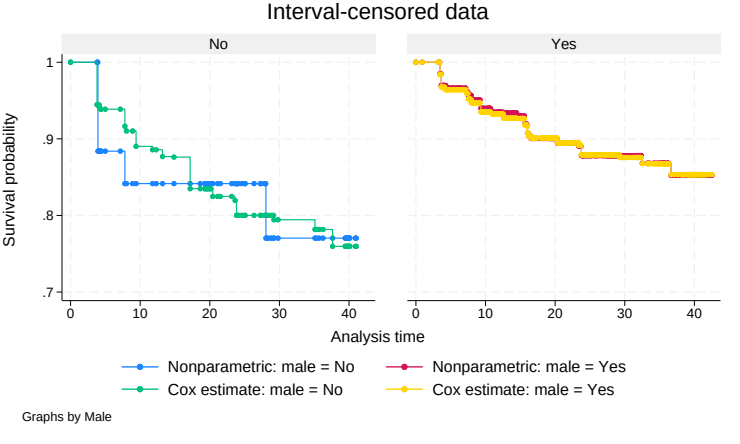
Example 2: Proportional-hazards assumption does not hold for a model with a single covariate

Continuing with `idu.dta`, suppose we want to examine the proportional-hazards assumption for a single binary variable `male`:

```
. stintphplot, interval(ltime rtime) by(male)
Computing nonparametric estimates for male = No ...
Computing nonparametric estimates for male = Yes ...
```




```
. stintcoxnp, interval(ltime rtime) by(male) byopts(style(altleg)) separate
Computing nonparametric estimates ...
Computing Cox estimates ...
```



The "log–log" plot displays two lines that are not parallel, which indicates that the proportional-hazards assumption is violated for `male`. If we take a closer look at the separate `stintcoxnp` plots for both males and females, we can see that there are considerable differences between the NPMLE estimates and the Cox estimates, especially for females. Therefore, using this variable in the Cox model may not be appropriate.

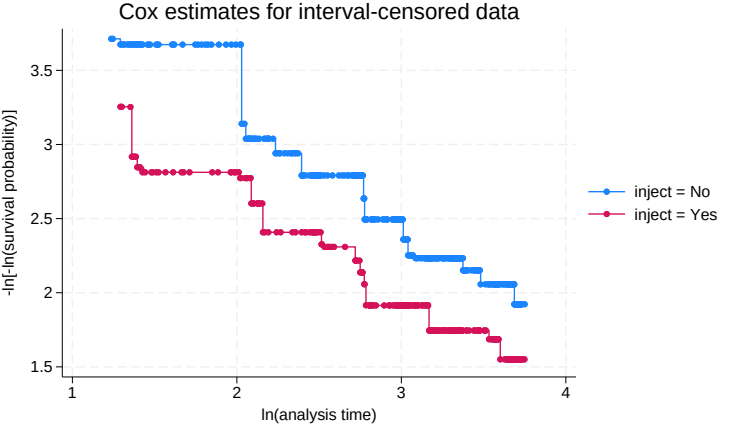
◁

▷ Example 3: Checking proportional-hazards assumption for a model with multiple covariates

When a Cox model contains multiple covariates, as in example 1 of [ST] **stintcox**, we should use the `adjustfor()` option to adjust for covariates.

To check the proportional-hazards assumption for `inject` for the Cox model in that example, we specify all the remaining covariates in the `adjustfor()` option.

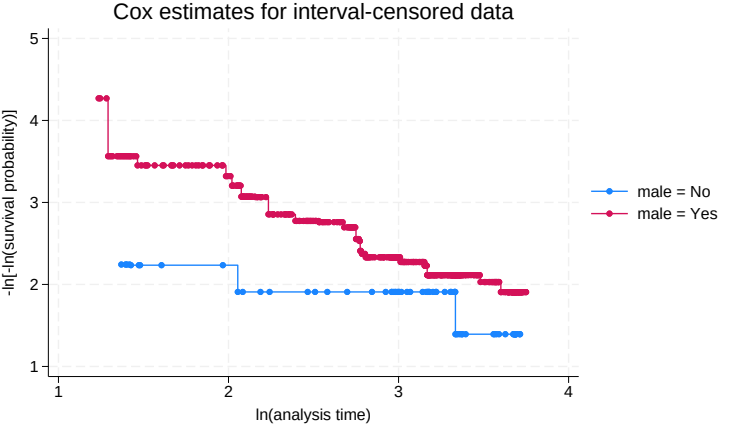
```
. stintphplot, interval(ltime rtime) by(inject)
> adjustfor(age_mean i.male i.needle i.jail)
Fitting Cox model with covariates from option adjustfor()
for inject = No ...
Fitting Cox model with covariates from option adjustfor()
for inject = Yes ...
```



In this case, a separate Cox model, which contains all covariates from the `adjustfor()` option, is fit for each level of `inject`.

To check the proportional-hazards assumption for `male`, we include all covariates except `male` in the `adjustfor()` option.

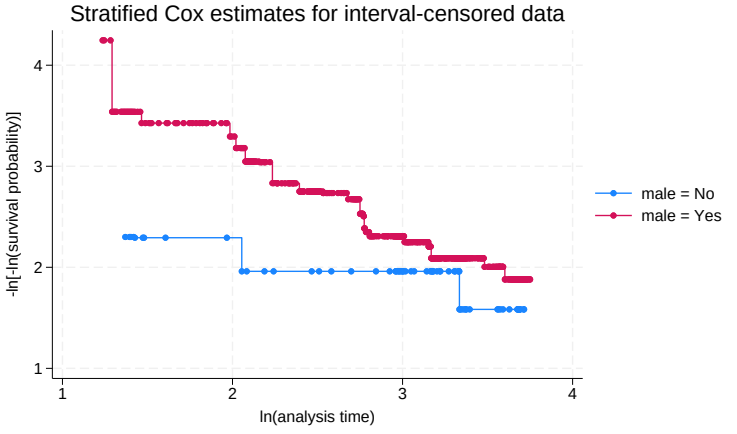
```
. stintphplot, interval(ltime rtime) by(male)
> adjustfor(age_mean i.needle i.inject i.jail)
Fitting Cox model with covariates from option adjustfor()
for male = No ...
Fitting Cox model with covariates from option adjustfor()
for male = Yes ...
```



After adjusting for covariates, we see our conclusions from example 1 and example 2 remain the same. The proportional-hazards assumption appears to be met for the `inject` variable but not for the `male` variable.

It may not always be feasible to fit a separate Cox model for all levels of a tested covariate when some levels have only a few observations. In this case, you may consider using the `strata()` option instead of `by()` to fit a Cox model stratified on the tested covariate. For instance, you could type

```
. stintphplot, interval(ltime rtime) strata(male)
> adjustfor(age_mean i.needle i.inject i.jail)
Fitting Cox model stratified on male with covariates from option adjustfor() ...
```

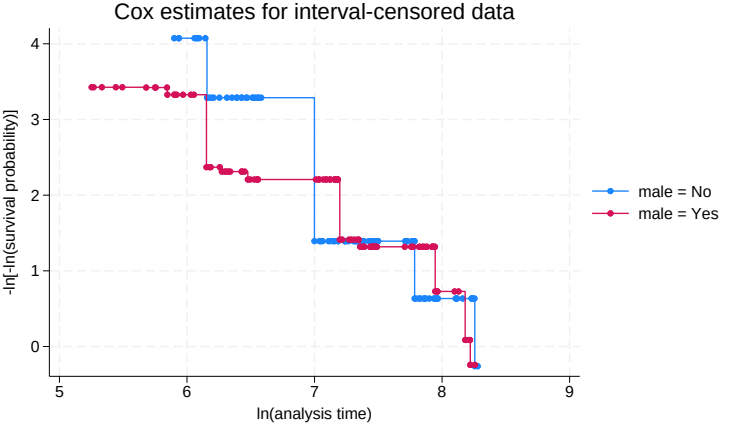


Note that the `stintcoxnp` command is not appropriate for testing the proportional-hazards assumption in the presence of multiple covariates in a Cox model.

▷ Example 4: Checking the proportional-hazards assumption with multiple-event data

When working with interval-censored multiple-event data, we must specify which event we are interested in when checking the proportional-hazards assumption. For example, below we load the data from example 1 of [ST] **stmgintcox**, in which we modeled the time to onset of diabetes and hypertension. We are interested in checking the proportional-hazards assumption for `male` when modeling the time to onset of diabetes, which is event 1, and of hypertension, which is event 2. We can use the `if` statement with `stintphplot` to check this assumption separately for each event. For instance, below we specify `if event == 1` to check the assumption for event 1, and we adjust for the covariates in that example.

```
. use https://www.stata-press.com/data/r19/aric
(Simulated ARIC data)
. stintphplot if event == 1, interval(ltime rtime) by(male)
> adjustfor(age i.community i.race bmi glucose sysbp diabp)
Fitting Cox model with covariates from option adjustfor()
for male = No ...
Fitting Cox model with covariates from option adjustfor()
for male = Yes ...
```



The `stintphplot` plot shows that the two lines are relatively parallel, implying that the proportional-hazards assumption holds for `male` when modeling the time to onset of diabetes.

Similarly, we can use `if event == 2` to check the proportional-hazards assumption for hypertension. We leave this for you to explore as an exercise.

Methods and formulas

For one covariate, x, the Cox proportional hazards model reduces to

$$h(t;x) = h_0(t)\exp(x\beta)$$

where $h_0(t)$ is the baseline hazard function from the Cox model. Let $S_0(t)$ and $H_0(t)$ be the corresponding Cox baseline survivor and baseline cumulative hazard functions, respectively.

The proportional-hazards assumption implies that

$$H(t;x) = H_0(t)\exp(x\beta)$$

or

$$\ln\{H(t;x)\} = \ln\{H_0(t)\} + x\beta$$

where $H(t;x)$ is the cumulative hazard function. Thus, under the proportional-hazards assumption, the logs of the cumulative hazard functions at each level of the covariate have equal slope. This is the basis for the method implemented in `stintphplot`.

The proportional-hazards assumption also implies that

$$S(t;x) = S_0(t)^{\exp(x\beta)} = \exp\{-H_0(t)\}^{\exp(x\beta)}$$

so the plot is referred to as a "log–log" survival plot because $\ln[-\ln\{S(t;x)\}] = \ln\{H(t;x)\}$.

For the Cox model, let $\widehat{H}(t;x) = \widehat{H}_0(t)\exp(x\hat{\beta})$, where $\widehat{H}_0(t) = \sum_{t_k \le t} \hat{h}_k$, $\hat{h}_k$'s ($k = 1, \ldots, m$) are estimated baseline hazard contributions and t_k's are the unique time points, ordered from smallest to largest, of the observed interval endpoints t_l and t_u. Let $\widehat{S}(t;x) = \exp\{-\widehat{H}(t;x)\}$ be the estimated survivor function based on the Cox model. This function is a step function. Thus, for each level of the covariate of interest, we can assess violations of the proportional-hazards assumption by comparing these survival estimates with nonparametric estimates from Turnbull (1976). See Kalbfleisch and Prentice (2002) or Hess (1995).

`stintcoxnp` plots Turnbull's (1976) NPMLE curves for each level of the covariate together with the survival curves predicted by the Cox model. The closer the Turnbull estimates are to the Cox estimates, the less likely it is that the proportional-hazards assumption has been violated.

References

Hess, K. R. 1995. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Statistics in Medicine* 14: 1707–1723. https://doi.org/10.1002/sim.4780141510.

Kalbfleisch, J. D., and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York: Wiley.

Turnbull, B. W. 1976. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society*, B ser., 38: 290–295. https://doi.org/10.1111/j.2517-6161.1976.tb01597.x.

Also see

[ST] **stintcox** — Cox proportional hazards model for interval-censored survival-time data

[ST] **stintcox postestimation** — Postestimation tools for stintcox

[ST] **stmgintcox** — Marginal Cox PH model for interval-censored multiple-event data

[ST] **stmgintcox postestimation** — Postestimation tools for stmgintcox

[ST] ***adjustfor_option*** — Adjust survivor and related functions for covariates at specific values

[U] 20 Estimation and postestimation commands