Title

stata.com

stintcox postestimation — Postestimation tools for stintcox

Postestimation commands

The following postestimation commands are of special interest after `stintcox`:

Command	Description
`stcurve`	plot the survivor, hazard, and cumulative hazard functions
`stintcoxnp`	plot nonparametric curves and Cox predicted curves
`stintphplot`	plot $-\ln\{-\ln(\text{survival})\}$ curves

The following standard postestimation commands are also available:

Command	Description
`contrast`	contrasts and ANOVA-style joint tests of estimates
`estat ic`	Akaike's and Schwarz's Bayesian information criteria (AIC and BIC)
`estat summarize`	summary statistics for the estimation sample
`estat vce`	variance–covariance matrix of the estimators (VCE)
`estimates`	cataloging estimation results
`etable`	table of estimation results
`hausman`	Hausman's specification test
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of coefficients
`lrtest`	likelihood-ratio test
`margins`	marginal means, predictive margins, marginal effects, and average marginal effects
`marginsplot`	graph the results from margins (profile plots, interaction plots, etc.)
`nlcom`	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
`predict`	hazard ratios, survivor functions, influence statistics, residuals, etc.
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`pwcompare`	pairwise comparisons of estimates
`suest`	seemingly unrelated estimation
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

predict

Description for predict

`predict` creates new variables containing predictions such as hazard ratios, linear predictions, standard errors, and baseline survivor and baseline cumulative hazard functions.

Menu for predict

Statistics > Postestimation

Syntax for predict

`predict` [*type*] *newvar* [*if*] [*in*] [`,` *statistic*]

`predict` [*type*] *newvar*$_l$ *newvar*$_u$ [*if*] [*in*] `,` *statistic2*

statistic	Description
Main	
`hr`	predicted hazard ratio, also known as the relative hazard; the default
`xb`	linear prediction $\mathbf{x}_j\widehat{\boldsymbol{\beta}}$
`stdp`	standard error of the linear prediction, $\text{SE}(\mathbf{x}_j\widehat{\boldsymbol{\beta}})$
* `mgale`	martingale-like residuals

statistic2	Description
Main	
* `basesurv`	baseline survivor function for interval endpoints t_l and t_u
* `basechazard`	baseline cumulative hazard function for interval endpoints t_l and t_u
* `basehc`	baseline hazard contributions for interval endpoints t_l and t_u

Unstarred statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample. Starred statistics are calculated for the estimation sample by default.

Options for predict

Main

`hr`, the default, calculates the relative hazard (hazard ratio), that is, the exponentiated linear prediction, $\exp(\mathbf{x}_j\widehat{\boldsymbol{\beta}})$.

`xb` calculates the linear prediction from the fitted model. That is, you fit the model by estimating a set of parameters, β_1, β_2, ..., β_k, and the linear prediction is $\widehat{\beta}_1 x_{1j} + \widehat{\beta}_2 x_{2j} + \cdots + \widehat{\beta}_k x_{kj}$, often written in matrix notation as $\mathbf{x}_j\widehat{\boldsymbol{\beta}}$.

The x_{1j}, x_{2j}, ..., x_{kj} used in the calculation are obtained from the data currently in memory and need not correspond to the data on the independent variables used in estimating $\boldsymbol{\beta}$.

`stdp` calculates the standard error of the linear prediction, that is, the standard error of $\mathbf{x}_j\widehat{\boldsymbol{\beta}}$.

`mgale` calculates interval-censored martingale-like residuals, which are an interval-censored version of martingale residuals for right-censored data.

`basesurv` calculates the baseline survivor function for both the lower endpoint t_l and the upper endpoint t_u of the time interval specified in option `interval()` of `stintcox`.

`basechazard` calculates the baseline cumulative hazard for both the lower endpoint t_l and the upper endpoint t_u of the time interval specified in option `interval()` of `stintcox`.

`basehc` calculates the baseline hazard contributions for both the lower endpoint t_l and the upper endpoint t_u of the time interval specified in option `interval()` of `stintcox`.

margins

Description for margins

`margins` estimates margins of response for hazard ratios and linear predictions.

Menu for margins

Statistics > Postestimation

Syntax for margins

`margins` [*marginlist*] [`,` *options*]

`margins` [*marginlist*] `,` `predict(`*statistic* ...`)` [`predict(`*statistic* ...`)` ...] [*options*]

statistic	Description
`hr`	hazard ratio, also known as the relative hazard
`xb`	linear prediction $\mathbf{x}_j\widehat{\boldsymbol{\beta}}$
`stdp`	not allowed with `margins`
`basesurv`	not allowed with `margins`
`basechazard`	not allowed with `margins`
`basehc`	not allowed with `margins`
`mgale`	not allowed with `margins`

Statistics not allowed with `margins` are functions of stochastic quantities other than `e(b)`.

For the full syntax, see [R] **margins**.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Baseline functions
Residuals and diagnostic measures

Baseline functions

`predict` after `stintcox` is used to generate a new variable or variables containing predicted values or residuals. `predict` can generate predicted hazard ratios, linear predictions, and standard errors of the linear predictions. It can also predict the baseline survivor function, baseline cumulative hazard function, or baseline hazard contributions for interval endpoints t_l and t_u specified in the `interval()` option with `stintcox`.

Baseline functions refer to the values of the functions when all covariates are set to 0. `stintcox` defines the baseline survivor function $S_0(t)$ as $S_0(t) = \exp\{-H_0(t)\}$, where $H_0(t)$ is the baseline cumulative hazard.

▷ Example 1: Baseline survivor function

We use data from example 1 of [ST] **stintcox** and fit the Cox proportional hazards model in which the time to HIV infection depends on centered age variable (`age_mean`) and whether a subject has been in jail at the time of recruitment (`jail`):

```
. use https://www.stata-press.com/data/r17/idu
(Modified Bangkok IDU Preparatory Study)
. stintcox age_mean i.jail, interval(ltime rtime)
note: using adaptive step size to compute derivatives.
Performing EM optimization (showing every 100 iterations):
Iteration 0:   log likelihood = -1086.2564
Iteration 100:  log likelihood = -600.55818
Iteration 200:  log likelihood = -600.47678
Iteration 295:  log likelihood = -600.46551
Computing standard errors: ..... done
Interval-censored Cox regression                  Number of obs     =  1,124
Baseline hazard: Reduced intervals                       Uncensored =      0
                                                    Left-censored =     41
                                                   Right-censored =    991
                                                   Interval-cens. =     92
                                                  Wald chi2(2)      =  14.07
Log likelihood = -600.46551                       Prob > chi2       = 0.0009
```

	Haz. ratio	OPG std. err.	z	P>|z|	[95% conf.	interval]
age_mean	.9647236	.0122201	-2.84	0.005	.9410676	.9889743
jail						
Yes	1.819721	.3804566	2.86	0.004	1.207927	2.741379

```
Note: Standard-error estimates may be more variable for small datasets and
      datasets with low proportions of interval-censored observations.
```

We can now predict, for instance, the baseline survivor function. For interval-censored data, estimates of baseline survivor function, as well as baseline cumulative hazard and baseline hazard contributions, are intervals. So, to compute these statistics, we must specify two new variable names with `predict` instead of one; one variable will contain statistics computed using the lower time endpoint, and the other will contain statistics computed using the upper time endpoint.

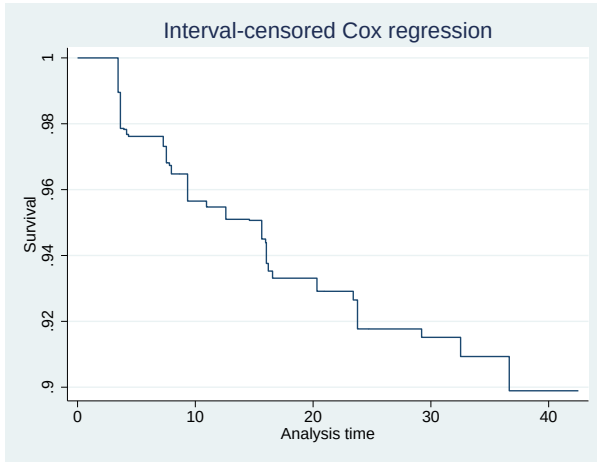
```
. predict bs_l bs_u, basesurv
. list bs_l bs_u ltime rtime age jail in 300/310
```

	bs_l	bs_u	ltime	rtime	age	jail
300.	.8989196	0	40.00	.	36	Yes
301.	.9547279	.9450011	11.97	15.84	21	Yes
302.	.9647679	.9506636	8.20	15.18	36	Yes
303.	.8989196	0	39.93	.	40	Yes
304.	.8989196	0	39.48	.	25	Yes
305.	.8989196	0	36.72	.	40	Yes
306.	.8989196	0	39.93	.	40	Yes
307.	.9767599	0	4.30	.	34	Yes
308.	.8989196	0	39.74	.	42	No
309.	.8989196	0	37.61	.	30	Yes
310.	.8989196	0	39.97	.	28	Yes

Listed above are the baseline survivor functions, `bs_l` and `bs_u`, evaluated at the lower and upper time endpoints `ltime` and `rtime`, for subjects 300 to 310.

To graph the baseline survival curve, we can use `stcurve` with all covariates set to 0. We can also graph survivor functions after `stintcox` with covariates set to any value. See [ST] **stcurve** for more information.

```
. stcurve, survival at(age_mean=0 jail=0)
```



Residuals and diagnostic measures

For right-censored data, several types of residuals have been introduced to assess the appropriateness of the Cox proportional hazards model; see *Residuals and diagnostic measures* in [ST] **stcox postestimation** for details.

Farrington (2000) proposed extensions of those residuals for interval-censored data; see the reference for applications and a discussion of limitations of the residuals for interval-censored data. Here we offer martingale-like residuals for visual model checking. Martingale-like residuals are useful in determining the functional form of covariates to be included in the model. They are also useful in assessing whether some covariates are needed in the model and for identifying outliers.

▷ Example 2: Martingale-like residuals

Martingale-like residuals may be used as a diagnostic tool to assess the functional form of a covariate to be included in a Cox model. To find the appropriate functional form of a variable, we fit a Cox model excluding the variable of interest and then plot a `lowess` smooth of the martingale-like residuals against some transformation of that variable. If the transformation is appropriate, then the smooth should be approximately linear.

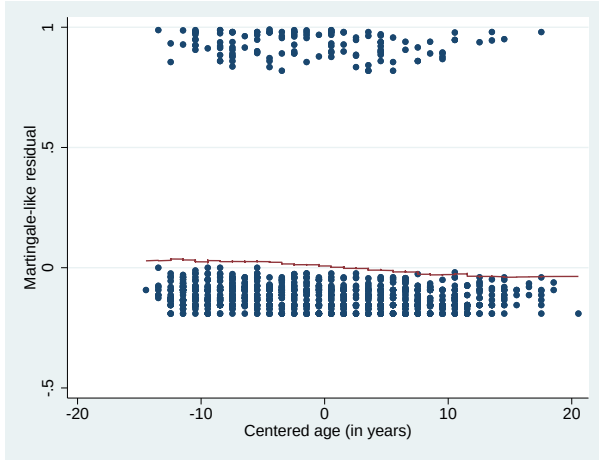
Continuing with example 1, we can apply this procedure to check whether the functional form of the covariate `age_mean` is appropriate. We first refit the Cox proportional hazards model on whether a subject has been in jail at the time of recruitment (`jail`). Then, we obtain the martingale-like residuals `mg` using `predict` with the `mgale` option. Last, we plot a `lowess` smooth of `mg` against variable `age_mean`.

```
. stintcox i.jail, interval(ltime rtime) nolog
note: using adaptive step size to compute derivatives.

Performing EM optimization ...

Computing standard errors ...

Interval-censored Cox regression                  Number of obs     =  1,124
Baseline hazard: Reduced intervals                       Uncensored =      0
                                                     Left-censored =     41
                                                    Right-censored =    991
                                                    Interval-cens. =     92

                                                  Wald chi2(1)      =   6.24
Log likelihood = -605.20197                       Prob > chi2       = 0.0125
```

	Haz. ratio	OPG std. err.	z	P>|z|	[95% conf.	interval]
jail						
Yes	1.675924	.3464929	2.50	0.013	1.11756	2.513264

```
Note: Standard-error estimates may be more variable for small datasets and
      datasets with low proportions of interval-censored observations.
```

```
. predict mg, mgale
. lowess mg age_mean, mean noweight title("") note("") m(o)
```



We used the `lowess` command with the `mean` and `noweight` options to obtain a plot of the running-mean smoother to ease interpretation. The smooth appears nearly linear, supporting the inclusion of the untransformed version of `age_mean` in our Cox model. Had the smooth not been linear, we would have tried smoothing the martingale residuals against various transformations of `age_mean` until we found one that produced a near-linear smooth.

◁

Methods and formulas

`predict` *newvar*`,` *statistic* may be used after `stintcox` to predict various quantities, according to the following *statistic*:

`hr`:

$$newvar_j = \exp(\mathbf{x}_j\widehat{\boldsymbol{\beta}})$$

`xb`:

$$newvar_j = \mathbf{x}_j\widehat{\boldsymbol{\beta}}$$

`stdp`:

$$newvar_j = \widehat{\mathrm{se}}(\mathbf{x}_j\widehat{\boldsymbol{\beta}})$$

`mgale`:

$$newvar_j = \frac{\widehat{S}_j(t_{lj})\log\widehat{S}_j(t_{lj}) - \widehat{S}_j(t_{uj})\log\widehat{S}_j(t_{uj})}{\widehat{S}_j(t_{lj}) - \widehat{S}_j(t_{uj})}$$

For right-censored data, martingale residuals can be defined as the scores of the regression parameters. This property can carry over to the interval-censored data. Therefore, these residuals are expected to have mean zero and to be asymptotically uncorrelated. Furthermore, these residuals are orthogonal to variables included in the model. Thus, we can use them to assess the need to include other covariates in the model.

`predict` *newvar*$_l$ *newvar*$_u$`,` *statistic2* may be used after `stintcox` to predict a pair of quantities for each observation for both the lower and upper endpoints of the time interval (t_{lj}, t_{uj}), according to the following *statistic2*:

`basehc`:

$$newvar_{lj} = I(t_k = t_{lj})\widehat{h}_k$$

$$newvar_{uj} = I(t_k = t_{uj})\widehat{h}_k$$

where $\widehat{h}_k$ $(k = 1, \ldots, m)$ are the estimates of the baseline cumulative hazard jumps at $t_1, \ldots, t_m$, respectively. And $t_1 < \cdots < t_m$ are the distinct time points for all $t_{lj} > 0$ and $t_{uj} < \infty$.

`basechazard`:

$$newvar_{lj} = \widehat{H}_0(t_{lj}) = \sum_{k:t_k \le t_{lj}} \widehat{h}_k$$

$$newvar_{uj} = \widehat{H}_0(t_{uj}) = \sum_{k:t_k \le t_{uj}} \widehat{h}_k$$

`basesurv`:

$$newvar_{lj} = \widehat{S}_0(t_{lj}) = \exp\left\{-\widehat{H}_0(t_{lj})\right\}$$

$$newvar_{uj} = \widehat{S}_0(t_{uj}) = \exp\left\{-\widehat{H}_0(t_{uj})\right\}$$

Reference

Farrington, C. P. 2000. Residuals for proportional hazards models with interval-censored survival data. *Biometrics* 56: 473–482. https://doi.org/10.1111/j.0006-341X.2000.00473.x.

Also see

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

[ST] **stintcox PH-assumption plots** — Plots of proportional-hazards assumption after stintcox

[ST] **stcurve** — Plot the survivor or related function after streg, stcox, and others

[U] **20 Estimation and postestimation commands**