Remarks and examples

stata.com

Remarks are presented under the following headings:

Baseline functions

`predict` after `stcox` provides estimates of the baseline survivor and baseline cumulative hazard function, among other things. Here the term *baseline* means that these are the functions when all covariates are set to zero, that is, they reflect (perhaps hypothetical) individuals who have zero-valued measurements. When you specify `predict`'s `basechazard` option, you obtain the baseline cumulative hazard. When you specify `basesurv`, you obtain the baseline survivor function. Additionally, when you specify `predict`'s `basehc` option, you obtain estimates of the baseline hazard contribution at each failure time, which are factors used to develop the product-limit estimator for the survivor function generated by `basesurv`.

Although in theory $S_0(t) = \exp\{-H_0(t)\}$, where $S_0(t)$ is the baseline survivor function and $H_0(t)$ is the baseline cumulative hazard, the estimates produced by `basechazard` and `basesurv` do not exactly correspond in this manner, although they closely do. The reason is that `predict` after `stcox` uses different estimation schemes for each; the exact formulas are given in *Methods and formulas*.

When the Cox model is fit with the `strata()` option, you obtain estimates of the baseline functions for each stratum.

▷ Example 1: Baseline survivor function

Baseline functions refer to the values of the functions when all covariates are set to 0. Let's graph the survival curve for the Stanford heart transplant model that we fit in example 3 of [ST] **stcox**, and to make the baseline curve reasonable, let's do that at `age` = 40 and `year` = 70.

Thus we will begin by creating variables that, when 0, correspond to the baseline values we desire, and then we will fit our model with these variables instead. We then predict the baseline survivor function and graph it:

```
. use https://www.stata-press.com/data/r18/stan3
(Heart transplant data)
. generate age40 = age - 40
. generate year70 = year - 70
```

If the Efron method for ties is specified at estimation, the partial log likelihood is

$$\log L_{\text{efron}} = \sum_{j=1}^{D} \sum_{i \in D_j} \left[\mathbf{x}_i\boldsymbol{\beta} + \text{offset}_i - d_j^{-1} \sum_{k=0}^{d_j-1} \log \left\{ \sum_{\ell \in R_j} \exp(\mathbf{x}_\ell\boldsymbol{\beta} + \text{offset}_\ell) - kA_j \right\} \right]$$

for $A_j = d_j^{-1} \sum_{\ell \in D_j} \exp(\mathbf{x}_\ell\boldsymbol{\beta} + \text{offset}_\ell)$. Weights are not supported with the Efron method.

At estimation, Stata also supports the exact marginal and exact partial methods for handling ties, but only the Peto–Breslow and Efron methods are supported in regard to the calculation of residuals, diagnostics, and other predictions. As such, only the partial log-likelihood formulas for those two methods are presented above, for easier reference in what follows.

If you specified `efron` at estimation, all predictions are carried out using the Efron method; that is, the handling of tied failures is done analogously to the way it was done when calculating $\log L_{\text{efron}}$. If you specified `breslow` (or nothing, because `breslow` is the default), `exactm`, or `exactp`, all predictions are carried out using the Peto–Breslow method. That is not to say that if you specify `exactm` at estimation, your predictions will be the same as if you had specified `breslow`. The formulas used will be the same, but the parameter estimates at which they are evaluated will differ because those were based on different ways of handling ties.

Define $z_i = \mathbf{x}_i\widehat{\boldsymbol{\beta}} + \text{offset}_i$. Schoenfeld residuals for the pth variable using the Peto–Breslow method are given by

$$r_{S_{pi}} = \delta_i \left(x_{pi} - a_{pi}\right)$$

where

$$a_{pi} = \frac{\sum_{\ell \in R_i} w_\ell x_{p\ell} \exp(z_\ell)}{\sum_{\ell \in R_i} w_\ell \exp(z_\ell)}$$

δ_i indicates failure for observation i, and x_{pi} is the pth element of $\mathbf{x}_i$. For the Efron method, Schoenfeld residuals are

$$r_{S_{pi}} = \delta_i \left(x_{pi} - b_{pi}\right)$$

where

$$b_{pi} = d_i^{-1} \sum_{k=0}^{d_i-1} \frac{\sum_{\ell \in R_i} x_{p\ell} \exp(z_\ell) - kd_i^{-1} \sum_{\ell \in D_i} x_{p\ell} \exp(z_\ell)}{\sum_{\ell \in R_i} \exp(z_\ell) - kd_i^{-1} \sum_{\ell \in D_i} \exp(z_\ell)}$$

Schoenfeld residuals are derived from the first derivative of the log likelihood, with

$$\left. \frac{\partial \log L}{\partial \beta_p} \right|_{\widehat{\boldsymbol{\beta}}} = \sum_{i=1}^{N} r_{S_{pi}} = 0$$

and only those observations that fail ($\delta_i = 1$) contribute a Schoenfeld residual to the derivative.

For censored observations, Stata stores a missing value for the Schoenfeld residual even though the above implies a value of 0. This is to emphasize that no calculation takes place when the observation is censored.

Scaled Schoenfeld residuals are given by

$$\mathbf{r}^*_{S_i} = \widehat{\boldsymbol{\beta}} + d \ \text{Var}(\widehat{\boldsymbol{\beta}}) \mathbf{r}_{S_i}$$

where $\mathbf{r}_{S_i} = (r_{S_{1i}}, \ldots, r_{S_{mi}})'$, m is the number of regressors, and d is the total number of failures.

In what follows, we assume the Peto–Breslow method for handling ties. Formulas for the Efron method, while tedious, can be obtained by applying similar principles of averaging across risk sets, as demonstrated above with Schoenfeld residuals.

Efficient score residuals are obtained by

$$r_{E_{pi}} = r_{S_{pi}} - \exp(z_i) \sum_{j:t_{0i}<t_j\le t_i} \frac{\delta_j w_j (x_{pi} - a_{pj})}{\sum_{\ell\in R_j} w_\ell \exp(z_\ell)}$$

Like Schoenfeld residuals, efficient score residuals are also additive components of the first derivative of the log likelihood. Whereas Schoenfeld residuals are the contributions of each failure, efficient score residuals are the contributions of each observation. Censored observations contribute to the log likelihood (and its derivative) because they belong to risk sets at times when other observations fail. As such, an observation's contribution is twofold: 1) If the observation ends in failure, a risk assessment is triggered, that is, a term in the log likelihood is computed. 2) Whether failed or censored, an observation contributes to risk sets for other observations that do fail. Efficient score residuals reflect both contributions.

The above computes efficient score residuals at the observation level. If you have multiple records per subject and do not specify the `partial` option, then the efficient score residual for a given subject is calculated by summing the efficient scores over the observations within that subject.

Martingale residuals are

$$r_{M_i} = \delta_i - \exp(z_i) \sum_{j:t_{0i}<t_j\le t_i} \frac{w_j \delta_j}{\sum_{\ell\in R_j} w_\ell \exp(z_\ell)}$$

The above computes martingale residuals at the observation level. If you have multiple records per subject and do not specify the `partial` option, then the martingale residual for a given subject is calculated by summing r_{M_i} over the observations within that subject.

Martingale residuals are in the range $(-\infty, 1)$. Deviance residuals are transformations of martingale residuals designed to have a distribution that is more symmetric about zero. Deviance residuals are calculated using

$$r_{D_i} = \text{sign}(r_{M_i}) \left[-2 \left\{ r_{M_i} + \delta_i \log(\delta_i - r_{M_i}) \right\} \right]^{1/2}$$

These residuals are expected to be symmetric about zero but do not necessarily sum to zero.

The above computes deviance residuals at the observation level. If you have multiple records per subject and do not specify the `partial` option, then the deviance residual for a given subject is calculated by applying the above transformation to the *subject-level* martingale residual.

The estimated baseline hazard contribution is obtained at each failure time as $h_j = 1 - \widehat{\alpha}_j$, where $\widehat{\alpha}_j$ is the solution to

$$\sum_{k\in D_j} \frac{\exp(z_k)}{1 - \widehat{\alpha}_j^{\exp(z_k)}} = \sum_{\ell\in R_j} \exp(z_\ell)$$

(Kalbfleisch and Prentice 2002, eq. 4.34, 115).

The estimated baseline survivor function is

$$\widehat{S}_0(t) = \prod_{j:t_j \le t} \widehat{\alpha}_j$$

When estimated with no covariates, $\widehat{S}_0(t)$ is the Kaplan–Meier estimate of the survivor function.

The estimated baseline cumulative hazard function, if requested, is related to the baseline survivor function calculation, yet the values of $\widehat{\alpha}_j$ are set at their starting values and are not iterated. Equivalently,

$$\widehat{H}_0(t) = \sum_{j:t_j \le t} \frac{d_j}{\sum_{\ell \in R_j} \exp(z_\ell)}$$

When estimated with no covariates, $\widehat{H}_0(t)$ is the Nelson–Aalen estimate of the cumulative hazard.

Cox–Snell residuals are calculated with

$$r_{C_i} = \delta_i - r_{M_i}$$

where r_{M_i} are the martingale residuals. Equivalently, Cox–Snell residuals can be obtained with

$$r_{C_i} = \exp(z_i)\widehat{H}_0(t_i)$$

The above computes Cox–Snell residuals at the observation level. If you have multiple records per subject and do not specify the `partial` option, then the Cox–Snell residual for a given subject is calculated by summing r_{C_i} over the observations within that subject.

DFBETAs are calculated with

$$\text{DFBETA}_i = \mathbf{r}_{E_i}\widetilde{\text{Var}}(\widehat{\boldsymbol{\beta}})$$

where $\mathbf{r}_{E_i} = (r_{E_{1i}}, \ldots, r_{E_{mi}})$ is a row vector of efficient score residuals with one entry for each regressor, and $\widetilde{\text{Var}}(\widehat{\boldsymbol{\beta}})$ is the model-based variance matrix of $\widehat{\boldsymbol{\beta}}$.

Likelihood displacement values are calculated with

$$\text{LD}_i = \mathbf{r}_{E_i}\text{Var}(\widehat{\boldsymbol{\beta}})\mathbf{r}'_{E_i}$$

(Collett 2015, 156). In both of the above, $\mathbf{r}_{E_i}$ can represent either one observation or, in multiple-record data, the cumulative efficient score for an entire subject. For the former, the interpretation is that due to deletion of one record; for the latter, the interpretation is that due to deletion of all of a subject's records.

Following Collett (2015, 156), LMAX values are obtained from an eigensystem analysis of

$$\mathbf{B} = \boldsymbol{\Theta}\ \text{Var}(\widehat{\boldsymbol{\beta}})\ \boldsymbol{\Theta}'$$

where $\boldsymbol{\Theta}$ is the $N \times m$ matrix of efficient score residuals, with element (i, j) representing the jth regressor and the ith observation (or subject). LMAX values are then the absolute values of the elements of the unit-length eigenvector associated with the largest eigenvalue of the $N \times N$ matrix $\mathbf{B}$.

For shared-frailty models, the data are organized into G groups, with the ith group consisting of n_i observations, $i = 1, \ldots, G$. From Therneau and Grambsch (2000, 253–255), for fixed θ, estimates of $\boldsymbol{\beta}$ and $\nu_1, \ldots, \nu_G$ are obtained by maximizing

$$\log L(\theta) = \log L_{\text{Cox}}(\boldsymbol{\beta}, \nu_1, \ldots, \nu_G) + \sum_{i=1}^{G} \left[\frac{1}{\theta} \{\nu_i - \exp(\nu_i)\} + \left(\frac{1}{\theta} + D_i\right) \left\{1 - \log\left(\frac{1}{\theta} + D_i\right)\right\} - \frac{\log\theta}{\theta} + \log\Gamma\left(\frac{1}{\theta} + D_i\right) - \log\Gamma\left(\frac{1}{\theta}\right) \right]$$

where D_i is the number of death events in group i, and $\log L_{\text{Cox}}(\boldsymbol{\beta}, \nu_1, \ldots, \nu_G)$ is the standard Cox partial log likelihood, with the ν_i treated as the coefficients of indicator variables identifying the groups. That is, the jth observation in the ith group has log relative-hazard $\mathbf{x}_{ij}\boldsymbol{\beta} + \nu_i$.

You obtain the estimates of $\nu_1, \ldots, \nu_G$ with `predict`'s `effects` option after `stcox, shared()`.

estat concordance

Harrell's C was proposed by Harrell et al. (1982) and was developed to evaluate the results of a medical test. The C index is defined as the proportion of all usable subject pairs in which the predictions and outcomes are concordant. The C index may be applied to ordinary continuous outcomes, dichotomous diagnostic outcomes, ordinal outcomes, and censored time-until-event response variables.

In predicting the time until death, C is calculated by considering all comparable patient pairs. A pair of patients is comparable if either 1) the two have different values on the time variable, and the one with the lowest value presents a failure, or 2) the two have the same value on the time variable, and exactly one of them presents a failure. If the predicted survival time is larger for the patient who lived longer, the predictions for the pair are said to be concordant with the outcomes. From Fibrinogen Studies Collaboration (2009), Harrell's C is defined as $\sum_k (E_k + T_k/2) / \sum_k (D_k)$, where D_k is the total number of pairs usable for comparison in stratum k, E_k is the number of pairs for which the predictions are concordant with the outcomes and the predictions are not identical in stratum k, and T_k is the number of usable pairs for which the predictions are identical in stratum k. If there are no strata specified, then the formula for Harrell's C reduces to $(E + T/2)/D$.

For a Cox proportional hazards model, the probability that the patient survives past time t is given by $S_0(t)$ raised to the $\exp(\mathbf{x}\boldsymbol{\beta})$ power, where $S_0(t)$ is the baseline survivor function, $\mathbf{x}$ denotes a set of measurements for the patient, and $\boldsymbol{\beta}$ is the vector of coefficients. A Cox regression model is fit by the `stcox` command. The hazard ratio, $\exp(\mathbf{x}\boldsymbol{\beta})$, is obtained by `predict` after `stcox`. Because the predicted survival time and the predicted survivor function are one-to-one functions of each other, the predicted survivor function can be used to calculate C instead of the predicted survival time. The predicted survivor function decreases when the predicted hazard ratio increases; therefore, Harrell's C can be calculated by computing E, T, and D, based on the observed outcomes and the predicted hazard ratios.

C takes a value between 0 and 1. A value of 0.5 indicates no predictive discrimination, and values of 0 or 1.0 indicate perfect separation of subjects with different outcomes. See Harrell, Lee, and Mark (1996) for more details. Somers's D rank correlation is calculated by $2C - 1$; see Newson (2002) for a discussion of Somers's D.

In the presence of censoring, Harrell's C coefficient tends to be biased. An alternative measure of concordance that is asymptotically unbiased with censored data was proposed by Gönen and Heller (2005). This estimator does not depend on observed time directly and is a function of only the regression parameters and the covariate distribution, which leads to its asymptotic unbiasedness and thus robustness to the degree of censoring.

Let $\Delta\mathbf{x}_{ij}$ be the pairwise difference $\mathbf{x}_i - \mathbf{x}_j$. Then Gönen and Heller's concordance probability estimator is given by

$$K \equiv K_N(\widehat{\boldsymbol{\beta}}) = \frac{2}{N(N-1)} \sum_{i<j} \sum \left\{ \frac{I(\Delta\mathbf{x}_{ji}\widehat{\boldsymbol{\beta}} \leq 0)}{1 + \exp(\Delta\mathbf{x}_{ji}\widehat{\boldsymbol{\beta}})} + \frac{I(\Delta\mathbf{x}_{ij}\widehat{\boldsymbol{\beta}} < 0)}{1 + \exp(\Delta\mathbf{x}_{ij}\widehat{\boldsymbol{\beta}})} \right\} \tag{1}$$

where $I(\cdot)$ is the indicator function. Somers's D rank correlation is calculated by $2K - 1$.

The concordance probability estimator (1) involves indicator functions and thus is a nonsmooth function for which the asymptotic standard error cannot be computed directly. To obtain the standard error, a smooth approximation to this estimator is considered:

$$\widetilde{K} \equiv \widetilde{K}_N(\widehat{\boldsymbol{\beta}}) = \frac{2}{N(N-1)} \sum_{i<j} \sum \left\{ \frac{\Phi(-\Delta\mathbf{x}_{ji}\widehat{\boldsymbol{\beta}}/h)}{1 + \exp(\Delta\mathbf{x}_{ji}\widehat{\boldsymbol{\beta}})} + \frac{\Phi(-\Delta\mathbf{x}_{ij}\widehat{\boldsymbol{\beta}}/h)}{1 + \exp(\Delta\mathbf{x}_{ij}\widehat{\boldsymbol{\beta}})} \right\} \tag{2}$$

where $\Phi(\cdot)$ is a standard normal distribution function, $h = 0.5\widehat{\sigma}N^{-1/3}$ is a smoothing bandwidth, and $\widehat{\sigma}$ is the estimated standard deviation of the subject-specific linear predictors $\mathbf{x}_i\widehat{\boldsymbol{\beta}}$.

The asymptotic standard error is then computed using a first-order Taylor series expansion of (2) around the true parameter $\boldsymbol{\beta}$; see Gönen and Heller (2005) for computational details.

References

Cattaneo, M. D., P. Malighetti, and D. Spinelli. 2017. Estimating receiver operative characteristic curves for time-dependent outcomes: The stroccurve package. *Stata Journal* 17: 1015–1023.

Cefalu, M. S. 2011. Pointwise confidence intervals for the covariate-adjusted survivor function in the Cox model. *Stata Journal* 11: 64–81.

Collett, D. 2015. *Modelling Survival Data in Medical Research*. 3rd ed. Boca Raton, FL: Chapman and Hall/CRC.

Fibrinogen Studies Collaboration. 2009. Measures to assess the prognostic ability of the stratified Cox proportional hazards model. *Statistics in Medicine* 28: 389–411. https://doi.org/10.1002/sim.3378.

Gönen, M., and G. Heller. 2005. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* 92: 965–970. https://doi.org/10.1093/biomet/92.4.965.

Harrell, F. E., Jr., R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati. 1982. Evaluating the yield of medical tests. *Journal of the American Medical Association* 247: 2543–2546. https://doi.org/10.1001/jama.1982.03320430047030.

Harrell, F. E., Jr., K. L. Lee, and D. B. Mark. 1996. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 15: 361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4⟨361::AID-SIM168⟩3.0.CO;2-4.

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

Klein, J. P., and M. L. Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. New York: Springer.

Mansuy, R. 2009. The origins of the word "martingale". *Electronic Journal for History of Probability and Statistics* 5: 1–10. http://www.jehps.net/juin2009/Mansuy.pdf.

Mazliak, L., and G. Shafer. 2009. The splendors and miseries of martingales. *Electronic Journal for History of Probability and Statistics* 5: 1–5. http://www.jehps.net/juin2009/MazliakShafer.pdf.

Metzger, S. K., and B. T. Jones. 2018. mstatecox: A package for simulating transition probabilities from semiparametric multistate survival models. *Stata Journal* 18: 533–563.

——. 2021. Properly calculating estat phtest in the presence of stratified hazards. *Stata Journal* 21: 1028–1033.

Newson, R. B. 2002. Parameters behind "nonparametric" statistics: Kendall's tau, Somers' D and median differences. *Stata Journal* 2: 45–64.