The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

The Stata Journal (2011)
11, Number 1, pp. 143–144

Stata tip 94: Manipulation of prediction parameters for parametric survival regression models

Theresa Boswell
StataCorp
College Station, TX
tboswell@stata.com

Roberto G. Gutierrez
StataCorp
College Station, TX
rgutierrez@stata.com

After fitting a parametric survival regression model using `streg` (see [ST] **streg**), predicting the survival function for the fitted model is available with the `predict` command with the `surv` option. Some users may wish to alter the parameters used by `predict` to compute the survival function for a specific time frame or combination of covariate values.

Manipulation of the prediction parameters can be done directly by altering the variables that `predict` uses to calculate the survival function. However, it is good practice to create a copy of the variables before making any changes so that we can later return variables to their original forms.

This is best illustrated by an example. Using `cancer.dta` included with Stata, we can fit a simple Weibull model with one covariate, `age`:

```
. sysuse cancer
. streg age, dist(weibull)
```

Suppose that we want to obtain the predicted survival function for a specific time range and age value. The time variables used by `predict` to calculate the survival function are stored in variables `_t` and `_t0`, as established by `stset`. Before making any changes, we must first create a copy of these time variables and of our covariate `age`. We can use the `clonevar` command to create a copy. The advantage of using `clonevar` over `generate` is that `clonevar` creates an exact replica of each original variable, including its labels and other properties.

```
. clonevar age_orig = age
. clonevar t_orig = _t
. clonevar t0_orig = _t0
```

Now that we have a copy of the original variables, we are free to manipulate parameters. Let's assume that we want predictions of the survival function for individuals entering the study at age 75 over the time range [0,20]. To alter the time variables, we can use the `range` command to replace `_t` with an evenly spaced grid from 0 to 20:

```
. drop _t
. range _t 0 20
```

The `_t0` variable needs to be set to 0 (for obtaining unconditional survival), and `age` should be set to 75 for all observations:

```
. replace _t0 = 0
. replace age = 75
```

© 2011 StataCorp LP

The prediction will now correspond to the survival function for an individual entering the study at age 75 over a time range of 0 to 20. The `predict` command with option `surv` will return the predicted survival function.

```
. predict s, surv
```

To view the predicted values, type

```
. list _t0 _t s
```

or you can graph the survival function by typing

```
. twoway line s _t
```

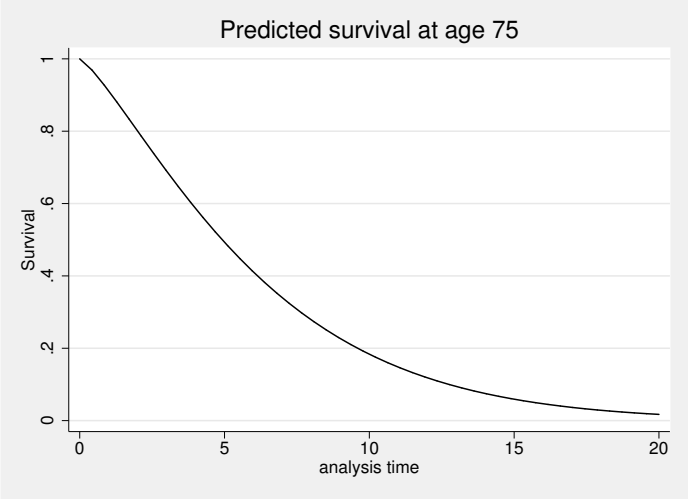

Figure 1. Predicted survival function

Now that we have the predicted values we want, it is prudent to replace all changed variables with their original forms. To do this, we will use the copies we created at the beginning of this example.

```
. replace age = age_orig
. replace _t = t_orig
. replace _t0 = t0_orig
```

There are many cases in which one may wish to manipulate the predicted survival function after `streg` and in which the steps in this tip can be followed to calculate the desired predictions.