Stata tip 25: Sequence index plots

Ulrich Kohler and Christian Brzinsky-Fay
Wissenschaftszentrum Berlin
kohler@wz-berlin.de, brzinsky-fay@wz-berlin.de

Sequence index plots of longitudinal or panel data use stacked bars or line segments to show how individuals move between a set of conditions or states over time. Changes of state are shown by changes of color. The term sequence index plot was proposed by Brüderl and Scherer (2005, in press). See Scherer (2001) for an application.

It is possible to draw sequence index plots with Stata by using the twoway plottype rbar. Starting from data in survival-time form (see help st), you simply overlay separate range-bar plots for each state.

For example, suppose that you have data on times for entering and leaving various states of employment:

```
.list in 1/10
id type begin end
1. 1 employed 1 13
2. 1 apprenticeship 13 20
3. 1 unemployed 20 23
4. 1 employed 23 25
5. 1 unemployed 25 26
6. 1 employed 26 43
7. 1 unemployed 43 50
8. 1 employed 50 60
9. 2 employed 1 13
10. 2 apprenticeship 13 21
```

First, separate the start and end dates for the different states:

```
.separate begin, by(type)
.separate end, by(type)
```

Then plot overlaid range bars for each state:

```
.graph twoway
> (rbar begin1 end1 id, horizontal)
> (rbar begin2 end2 id, horizontal)
> (rbar begin3 end3 id, horizontal)
> (rbar begin4 end4 id, horizontal)
> (rbar begin5 end5 id, horizontal)
> , legend(order(1 "education" 2 "apprenticeship"
> 3 "employment" 4 "unemployment" 5 "inactivity")
> cols(1) pos(2) symsize(5))
> xtitle("months") yla(, angle(h)) yscale(reverse)
```
It is common to put personal identifiers on the $y$-axis, using the option `horizontal`, and put time on the $x$-axis.

In practice, with many individuals in a large panel, the bars become thinner lines. In such cases, you could use the plottype `rspike` instead. Note also that you can make room for more individuals by tuning the aspect ratio of the graph (see Cox 2004). There is no upper limit to how many individuals are shown, although as the number increases, the resulting graph may become too difficult to interpret. The readability, however, largely depends on how far similar individuals are grouped together. The sort order should therefore be some criterion of similarity between sequences.

To fine tune the graph, use any option allowed with `graph twoway`; type `help twoway_options`. Our example provides some simple illustrations. `legend()` changes the contents and placement of the legend. `xtitle()` defines the title along the $x$-axis. `ylabel()` is used to display the $y$-axis labels horizontally, instead of vertically. `yscale(reverse)` reverses the scale of the $y$-axis so that the first individual is plotted at the very top of the graph.

**References**

