

REPORT TO USERS

William W. Gould (President, StataCorp)

[minutes by Stephen P. Jenkins]

1. Stata is about open development

Bill Gould began with an exposition of the philosophy behind Stata.

His opening remark was: “When Stata is 100% ado files, then StataCorp can give up and go home”. In Stata 7, 82% of all Stata commands are defined by ado files; in Stata 6 the figure was 79%.

Stata is about open development with users on almost-as-near status as developers. This is reflected in

- design of the software
 - An open language
 - same for StataCorp and users
 - StataCorp must develop the language
 - Backwards compatibility essential or else users won't invest in adding new features
- documentation
 - must document programming tools well.
- distribution (historically the STB; increasingly the Net)
 - Important that distribution does not require approval of StataCorp or users' innovation might be curtailed, hence STB was made mostly independent of StataCorp
 - Statalist developed (all by itself) and a very popular medium for exchanging programs.
 - STB sales have not kept up with Stata sales.
 - -net- commands added to Stata.
 - carefully designed so that user sites could exist without StataCorp approval.
 - search facility needed; currently provided by StataCorp, but if you look at the innards of -net search- (-webseek- in Stata 6), you'll find that the net search engine can specify any provider.
 - -update- carefully separated from -net- so that users don't confuse StataCorp contributions (supported, with lots of certification) from others.
 - SSC-IDEAS (Boston software archive) invented itself and has been very successful
 - The overall result is that there are likely to be changes in the STB. More details shortly.

2. The evolution of Stata, version 6 to 7

During the life of Stata 6 (January 1999 – December 2000), there were 63 ado file updates, i.e. one every 11 days on average. And 4.7% of source code lines were added; 7.7% lines of ado code. Certification scripts also increased (do file lines up 37%).

Various other statistics were also shown.

Lots changed internally inside Stata when moving from 6 to 7, mainly to accommodate a range of new features such as SMCL (pronounced “smickel”!).

3. Why SMCL?

Why is it seen to be so important? [And why not a standard mark-up language like HTML instead?]

- need to be able handle real-time output, e.g. iteration log. (HTML has the idea of an end of document.)
- desires to add viewing features
 - SMCL is currently just a text mark-up language but has the potential to do very much more, and will do so in future. E.g. table formatting, smart translations, user choice about what is produced (do you want to output just coefficients and standard errors, say, or be able to add in means?) The idea is that all information may be in the log, but user can choose how to display it or parts of it.
 - ‘secret’ translation commands: try the following (currently) rudimentary commands for translating SMCL to html/tex:

```
log texman filename filename.tex [, replace]
log html filename filename.html [, replace]
```
- SMCL development as a new window driver (currently Help and Results) – it can control multiple windows. There is the potential for e.g. multiple Results windows or Help windows open.
- SMCL clickable
- SMCL could also be used to facilitate real-time sharing of Stata. Possible perhaps in future to send your Results window, via internet, to another user’s Results window. Similarly, command window. Clearly implications for security and firewalls need to be resolved. Has potential for use in e.g. Technical Support. What do users think?

4. The new -file- command

Bill described the features of the new -file- command, which provides the ability to read and write ASCII text and binary files. [Released via -update- the preceding week] NB can be used

to write/save matrices and later reload them. Example program code was shown.

5. Sabbatical scheme

Stata is keen to develop this further. The idea is for someone to bring an idea with them to StataCorp for a 6-month period and to work on it, and to generally interact with and swap Stata-ish ideas. (Jeroen Weesie has just completed a very successful stint at StataCorp under the scheme.)

WISHES AND GRUMBLES SESSION

addressed to: William W. Gould and Roberto G Gutierrez (StataCorp)

[minutes by Stephen P. Jenkins]

The usual rules applied. All comments and suggestions were noted, with no cast-iron promises made. But indications were given as to whether something would be treated as relatively high priority, or would be considered, or would be treated as something less than that!

Here are the notes of the proceedings, minuted in the order in which the remarks came. Suggestions are in italics, followed by the response.

C interface A current project and already working inside the StataCorp building. Will be for sale in about 6 months. (Bill reminded users of issues of cross-platform portability – would need to compile on multiple platforms.)

update executable is clumsy in Win – nice if could make easier. Perhaps in Stata 8

Could view log file while Stata running in version 6 but not in version 7 Wasn't aware of a problem and asked to be sent evidence so that could act on it.

Ability to combine tables, e.g. save option so that could stack up For Stata 8. Jeroen Weesie is taking the lead on this. Involves a major revamp of -estimates- to become more user-orientated rather than programmer orientated.

eform option in cloglog (allows hazard ratio interpretation when used for discrete time hazard modelling) Can do this

reshape to preserve variable and value labels Hard to do at present. Some one had suggested a route via use of characteristics?

Data sets used in Manuals to be put on the WWW. Working on this. Some problems with permissions to be resolved.

ltable update or, alternatively, estimates of hazard rates from sts list and sts graph (not just integrated hazard) Not clear what the response was on this (Qn was asked last year too, when said would look into.)

non-linear GLS program (so that could do e.g. minimum distance estimation). No commitment; will pass on to Vince Wiggins!

Some clumsiness in handling log open and close Some problems were mentioned (your minute-taker forgets the details); will be looked into

clogit to have robust and cluster options Doesn't it? OK will look into.

'Detonator' graphs Once explained what these were, many in audience didn't think that these were a priority!

tmp files being left around arising after do file failures Asked for documentary evidence in order to look into.

Make Stata easier to use for non-technical users Being looked into already (quite apart from existing StataQuest)

More regression diagnostics after poisson (very few compared to glm) OK

Multiple line styles for xline and yline options in graph. Maybe. WWG said: "I hope that this question and ones like it will soon never have to be asked again". [... an implicit acknowledgement of lack of progress on graphics. The audience were all very restrained on this issue!]

nonconstant option in streg After some clarification about why one might want this, said might look into.

Clickable links to data sets on WWW (see above)

Smoothed hazard rate estimates Mostly turned into discussion about what this actually involved [NB see K. Simons's presentation on this topic, with ado files]

Any plans for more on Generalized Additive Models No plans at present; see Patrick Royston's GAM program in STB

Program debugger no comment

More support for editors that are public domain and good (e.g. vi, TextPad) no comments made; some mention also of the emacs environment for Stata.

Contour plots etc “In the short term the goal is for new graphics to do everything that they currently do, but better. In the future, capabilities like this should be in-built.”