Damian Clarke on `gets` for *Stata Journal*

This is an interesting paper reporting original work likely to be helpful to many Stata users. As such it is potentially publishable in the *Stata Journal*.

It needs, however, extension, clarification and revision under various minor and not quite so minor headings.

1. From the outset the author seems to be assuming that he is addressing an econometrics or at least economics audience exclusively. This is a delicate point. Clearly the *SJ* carries many papers likely to be of major interest only to econometricians, just as it carries many papers similarly primarily aimed at other restricted groups. But many readers value the *SJ* for its interdisciplinary approach. For that reason and others I suggest that the author needs to do a little more to consider non-economists as a potential audience. After all, if this approach has merit, similar approaches should have equal merit in fields other than economics. The author is asked to pay attention to this by giving more references, especially introductory or review references, and by inserting sentences of the form 'This approach should also be applicable in ... because ...'.

2. Specifically, on p.1 (and also p.10) Hendry and the LSE approach are really assumed to be understood. The nuance intended by 'perhaps Hendry' is lost on me; perhaps (?) that's an in-joke or sly comment, or just possibly 'perhaps' is not the right word here. References are needed here.

3. It only slowly becomes apparent by close reading that `gets` is only a wrapper for `regress` or `xtreg`. That alone makes the program interesting and widely useful, but the limitation needs to be mentioned repeatedly and more prominently as a matter of clarity: in the abstract, introduction and conclusion of the paper and in the help file. This was a surprising small failure of clarity.

4. Different works e.g. Harrell, F.E. 2001. *Regression modeling strategies.* New York: Springer suggest that searching for a model should carry a penalty in degrees of freedom lost. Please discuss.

5. There is no About the author section.

Smaller points follow:

p.1 para.1 commas around 'of course' (or omit entirely?)

p.1 para.1 final words better as 'which is true to the information contained in the GUM, but also offers a more parsimonious interpretation' ('of the world' seems unnecessary here)

p.1 para.2 thought of

p.1 para.2 'by the theory of reduction': not clear what is meant here

p.1 note 1 arduous

p.2 para.2 Stata commands, not Stata functions

p.2 para.4 `gets` is a command rather than an algorithm

p.3 numbered point 1 `if` and `in` are usually described as qualifiers, not specifiers

p.3 numbered point 3 the notation n will suggest sample size to many readers; preferably use a different symbol (e.g. m) (also pp.4,5)

p.3 numbered points 4,5 If any of these tests is failed

p.3 numbered point 5 their t-statistics

p.5 option `vce()` need a different font for clustvar

p.6 4.1 para.1 applications; multicollinear (also p.7); in the example below

p.6 (1) why do we need the multiplication signs? (also p.12)

p.6 (1) plus 3 lines `gets`

p.6 (1) plus 4 lines independent variable

p.6 last para suggesting, different from

p.7 The message from the program shows a typo (should be Doornik) and could benefit from SMCL directives to avoid ugly wrapping.

p.7 4.2 First sentence is a comma splice: split into two sentences or use a semi-colon or colon to punctuate.

p.7 bottom reference for 'gauge' and 'potency'?

p.8 near top power is a general **statistical** concept, not an econometric one

p.8 near bottom the simulations below

p.9 for 'Falsely' read 'False'

p.9 for 'table their 7' read 'their Table 7'

p.9 author's website: what is it? (also p.12)

p.10 para.1 last sentence is a comma splice

p.11 details for the Campos et al. paper are inadequate