Organization of Statistical Features in Stata

Vince Wiggins
StataCorp

Stata User Group Meeting
Berlin
August 12, 2003

My Talk

German User Group Meeting

Not

- **An overview of statistical areas**
  - Econometrics
  - Biostatistics
  - ...
- **An overview of estimators and models**
  - Survival
  - Time series
  - Panel data
  - ...

Is

- **Common features of most estimators**
- **Facilities available after most estimators** (post estimation commands)
- **How to take advantage of this organization to get answers**
- **The fruits of programmers creating new facilities**

Highly technical,
time-tested,
widely deployed,

over-riding design principle

for Stata,
its estimators, and
its post-estimation facilities …

Type a little ...

Get a little

and allow estimation results to be analyzed independently and in conjunction with data.

Saved estimation results

- **Coefficient vector — `e(b)`**
- `Variance-Covariance matrix (VCE)` **— `e(V)`**
- **Estimation sample identifyer**
- **Estimator-specific `scalars`, `macros`, and matrices**
  - Observations
  - Degrees of freedom
  - Number of panels
  - ...

Type a little, get a little

- **Type an estimation command and get**
  - A relatively standard set of results
    - Coefficients
    - SEs
    - Tests of H0: B = 0
    - CIs
    - Overall fit
  - A set of saved results allowing further analysis
- **Type a post estimation command and get**
  - Predicted values
  - A Wald test
  - …

In other words, ask specific questions, get specific answers.

Side benefit — resampling

- **Bootstrapping —** `bs`
- **Jackknife SEs —** `jknife`
- **Monte-Carlo permutation tests —** `permute`
- **Monte-Carlo simulations —** `simulate`

- **Collecting estimation results over groups —** `statsby`

- **Survey data analysis with balanced and repeated replications?**

What is available after estimation

- Replay facilities
- Estimation result management facilities
- Extensions to expressions
- Post estimation commands

Replay facilities

- **Redisplay last estimates**
- **Show alternate forms of coefficients**
  - Odds ratios (ORs)
  - Incidence rate ratios (IRRs)
  - Relative rate ratios (RRRs)
  - …
- **Change CI and test confidence level**

Estimation management facilities

- **Store estimation results**
  ```
  . estimates store name
  ```
- **Restore estimation results**
  (make stored estimates active)
  ```
  . estimates restore name
  ```
- **Run a post estimation command on one or more stored estimation results**
  ```
  . estimates for namelist : post_est_command
  ```
- **Tabulate results from multiple stored estimates**
  ```
  . estimates table …
  ```
- …

Extensions to expressions

- **Refer to coefficients in expressions**
- **Refer to coefficient SEs in expressions**

```
. display “z = “ _b[mpg] / _se[mpg]
```

- **Refer to estimation results** `e(`*`result_name`*`)`
  - macros — strings or numbers
  - scalars — numbers
  - matrices

```
. ttail(e(df_r) , abs(_b[mpg])/_se[mpg])*2
```

Post estimation analyses

- **Predictions**
  - In sample
  - Out of sample
  - On separate datasets
- **Marginal effects**
  - At specified points
  - Averaged over observations?
- **Nonlinear predictions, generalized predictions**
- **Linear and nonlinear combinations of coefficients with CIs**
- **Adjusted means and proportions**
- **Seemingly unrelated estimation**
- **Testing — the three classics**
  - Wald
  - Likelihood ratio
  - Score?
- **Tests of nonlinear combinations**
- **Hausman tests**
- **Link tests**
- **Estimator specific**
  - IRFs, etc. after VARs
  - Hazard functions, etc after survival models
  - ...

Sidebar for programmers

All of these facilities are available after your custom written estimation commands

- Declare your program eclass
  ```
  . program myprogram , eclass
  ```
- Use `ereturn` commands to post your
  - Coefficients
  - VCE
  - Other results
- That is it — your results will:
  - Look like official estimators
  - Work with all post-estimation commands
- Even better, write a nice `predict` command
  - mfx will the be able to compute marginal effects w.r.t (almost) any statistic you predict
  - predictnl will be able to use these statistics directly in expressions to compute nonlinear functions, their SEs, and CIs.

Sidebar about programmers

- **Many user-written post-estimation commands have been written**
  - `testomit` by Jeroen Weesie
  - `margin` by Tamas Bartus
  - Many commands by Scott Long and Jeremy Freese for analyzing categorical data
- **Writing your own post estimation commands is not difficult. You have complete access to**
  - The saved results of the estimator
  - The estimation sample or the current dataset

A Simple Example

- **US National Longitudinal Survey**
  - Women age 14-26 in 1968
  - Years 1968 to 1988
  - Women in the workforce
  - 5159 women, 28510 observations
- **Probit model of salary increase from year to year**

$$Z = b0 + b1 * age + b2 / age + b3 * college\_degree$$

$$raise = 1 \quad \text{if } Z > u$$

$$raise = 0 \quad \text{if } Z < u$$

$$u \sim N(0, 1)$$

$$P(increase) = \Phi(Z)$$

  - Probability is nonlinear w.r.t coefficients (b0,b1, b2, b3)
  - Nonlinear specification in age
  - “Interesting” functionally

See example.pdf

Summary

- Don’t stop at estimation
- Improve presentation and understanding with post-estimation analysis
- If you write an estimator, almost all post-estimation facilities will just work

Type a little …

Get a what you want.