From Stata to aML

Sara Ayllón
Department of Economics
University of Girona

19th UK Stata Users Group Meeting

12-13th September 2013

Introduction

- In this work, I show how to exploit Stata to run multilevel multiprocess models with aML.
- aML (applied Maximum Likelihood) has several advantages:
  - It easily handles three or more data levels (that can be mixed and matched)
  - It can estimate models with two or more different outcomes or data structures
  - It makes it easy to program complex errors structures
  - It is fast and free of charge
- aML is broader than `cmp` by Roodman (2011). It allows full simultaneity in systems of equations. It can account for random effects at different levels and offers a wider range of model types. Also, `gllamm` by Rabe-Hesketh includes some of aML features but for single-equation models.

Introduction

- But, aML is not a user-friendly programme:
  - It requires that the data set is prepared with a third party software package (as Stata)
  - It needs two different control files (to upload data and to run regressions)
  - It requires the researcher to typeset the starting values (tedious, time consuming)
  - It needs the starting values to be assigned in the same order as the regressor set (potential for user error)

Introduction

What can Stata do about it?

- EVERYTHING!
- One can create a single Stata do-file that:
  - Prepares the dataset
  - Writes the control files
  - Inputs the starting values
  - Calls aML using the `shell` command
  - Brings aML results back to Stata after estimation
  - Posts aML results so that they become 'official' within Stata for postestimation

Introduction

- Main advantages of joining Stata and aML:
  - One can easily change model specification
  - No need to manually open the Command Prompt window used by aML
  - No margin for typo errors as Stata automatically writes all the necessary files
  - It saves lots of time!
- There are other examples that connect Stata with software packages for running multilevel models (`runmlwin`, `runmplus`, `bugutils`, etc.)

Data preparation

- Use Stata `local` for writing a string that will contain the variables list (that way that can be easily modified):
  ```
  loc level1 ""
  loc level1 "`level1' ivar1"
  ```
- Run single regressions in Stata and save the starting values using `estimates store` and `estout` (Jann, 2005)
- Use `outfile` and the local strings previously created to prepare the data set as required by aML
- Write the control file that uploads the data to aML with the `file open`, `file write` and `file close` commands (Gini and Pasquini, 2006)

Model specification and estimation

Model specification

- A bit more complicated... we can write the different parts that aML requires and save them as `.dta` files
- And, then, we simply `append` them to build the control file .aml

Model estimation

- Then one is ready to call aML within Stata. Three lines do all the job:

```
sh raw2aml -r data.r2a
sh aml -r estimate.aml
sh update -r estimate.out
```

And back to Stata...

- We can create an estimates table with the `mktab` utility in aML that can be imported with `insheet`.
- Once in Stata, we create a `bhat` matrix with the coefficients and a `vce` for variances and covariances using `mkmat`.
- We post them with `ereturn post` that way aML results become *official* within Stata.
- Now, we can use Stata postestimation commands.

A (naive) example...

NIswork.dta (National Longitudinal Survey of Youth). Women aged 14 to 26 in 1968.
Let's imagine that we want to explain the probability of being married (msp) and that of belonging to a union. Moreover, we want to control for unobserved heterogeneity by integrating out a random effect in each equation and we wish to allow a free correlation between these individual-specific effects. We propose a bivariate RE probit:

$$Msp_{it} = \beta_0 Birthyr_i + \beta_1 Race_i + \beta_2 Age_{it} + \beta_3 Lnwage_{it} + \beta_4 Union_{it} + c_i + \epsilon_{it} \quad (1)$$

$$Union_{it} = \alpha_0 Birthyr_i + \alpha_1 Race_i + \alpha_2 Age_{it} + \alpha_3 Lnwage_{it} + u_i + \lambda_{it} \quad (2)$$

We run the 'Stata_to_aML' do file and...

... in less than 2 minutes we have the results in Stata:

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons_1	2.3125	.3274	7.06	0.000	1.670808	2.954192
birth_yr_1	-.031	.0063	-4.92	0.000	-.0433478	-.0186522
race_1	-.8534	.0383	-22.28	0.000	-.9284666	-.7783334
age_1	.0345	.0016	21.56	0.000	.0313641	.0376359
ln_wage_1	-.1564	.0331	-4.73	0.000	-.2212748	-.0915252
union_1	-.0891	.0344	-2.59	0.010	-.1565228	-.0216772
_cons_2	-2.3221	.3689	-6.29	0.000	-3.045131	-1.599069
birth_yr_2	-.0245	.0071	-3.45	0.001	-.0384157	-.0105843
race_2	.5751	.0417	13.79	0.000	.4933695	.6568305
age_2	-.0083	.0021	-3.95	0.000	-.0124159	-.0041841
ln_wage_2	.9753	.0357	27.32	0.000	.9053293	1.045271
eps_1	1.4076	.0221	63.69	0.000	1.364285	1.450915
eps_2	1.3215	.0271	48.76	0.000	1.268385	1.374615
rho	.0216	.0173	1.25	0.212	-.0123074	.0555074

This is an example of how well Stata can work with other software.

From Stata to aML

Sara Ayllón
Department of Economics
University of Girona

19th UK Stata Users Group Meeting

12-13th September 2013