Maximum Likelihood estimation: Stata vs. Gauss

ARMANDO SÁNCHEZ VARGAS

ECONOMICS RESEARCH INSTITUTE
UNAM

Index

- Motivation
- Objective
- The Maximum Likelihood Method
- Capabilities: Stata vs Gauss
- Conclusions

Motivation

- Stata is a powerful and flexible statistical package for researchers who need to compute ML estimators that are not available as prepackaged routines.

- Prospective and advanced users would want to know:

I. ML estimation facilities in Stata and GAUSS.
II. The main advantages of Stata compared with GAUSS.
III. What is still needed and what might be refined to implement the whole ML methodology in Stata.

Objective

- The main purpose of this presentation is to compare the ML routines and capabilities offered by STATA and GAUSS.

The Maximum Likelihood Method

- The foundation for the theory and practice of maximum likelihood estimation is a probability model:

$$\Pr(Z \le z) = F(z;\theta)$$

Where Z is the random variable distributed according to a cumulative probability distribution function F() with parameter vector $\theta' = (\theta_1, \theta_2, ..., \theta_E)$ from Θ, which is the parameter space for F().

The Maximum Likelihood Method

- Typically, there is more that one variable of interest, so the model:

$$\Pr(Z_1 \le z_1, Z_2 \le z_2, \ldots, Z_k \le z_k = F(z;\theta)$$

Describes the joint distribution of the random variables, with $z = (z_1, z_2, \ldots, z_k)$. Using F(), we can compute probabilities for values of the Zs given values of the parameters θ.

The Maximum Likelihood Method

- Given observed values z of the variables, the likelihood function is:

$$\ell(\theta; z) = f(z; \theta)$$

Where f() is the probability density function corresponding to F().

We are interested in the element (vector) of Θ that was used to generate z. We denote this vector by Θ_T

The Maximum Likelihood Method

- Data typically consist of multiple observations on relevant variables, so we will denote a dataset with the matrix Z. Each of N rows, zj, of Z consists of jointly observed values of the relevant variables:

$$L(\theta; Z) = f(Z; \theta)$$

The Maximum Likelihood Method

- f() is now the joint-distribution function of the data-generating process. This means that the method by which the data were collected now plays a role in the functional from of f().

- We introduce the assumption that “observations” are independent and identically distributed (i.i.d.) and rewrite the likelihood as:

$$L(\theta; Z) = \ell(\theta; z_1) \times \ell(\theta; z_2) \times \ldots \times \ell(\theta; z_N)$$

The Maximum Likelihood Method

- The maximum likelihood estimates for θ are the values $\hat{\theta}$ such that:

$$L(\hat{\theta}; Z) = \max_{t \in \Theta} L(t; Z)$$

The Maximum Likelihood Method

- We are dealing with likelihood functions that are continuous in their parameters, let's define some differential operators to simplify the notation. For any real valued function a(t), we define D and D2 by:

DERIVATIVE	GRADIENT VECTOR
$Da(\theta) = \frac{\partial a(t)}{\partial t} \mid t = \theta$	$g(\theta) = D \ln L(\theta) = \frac{\partial \ln L(t;Z)}{\partial t} \|_{t=\theta}$
$D^2 a(\theta) = \frac{\partial^2 a(t)}{\partial t \partial t'} \mid t = \theta$	$H(\theta) = Dg(\theta) = D^2 \ln L(\theta) = \frac{\partial^2 \ln L(t;Z)}{\partial t \partial t'} \|_{t=\theta}$

Maximum Likelihood: Commonly used Densities in Micro-econometrics

MODEL	RANGE OF y	DENSITY f(y)	COMMON PARAMETERIZATION
Normal	$(-\infty, \infty)$	$\left[2\pi\sigma^2\right]^{-1/2} \ell^{-(y-\mu)^2/2\sigma^2}$	$\mu = x'\beta, \sigma^2 = \sigma^2$
Bernoulli	0 or 1	$p^y(1-p)^{1-y}$	Logit $p = \ell^{x'\beta}/(1+\ell^{x'\beta})$
Exponential	$(0, \infty)$	$\lambda \ell^{-\lambda y}$	$\lambda = \ell^{x'\beta}$ or $1/\lambda = \ell^{x'\beta}$
Poisson	0, 1, 2,…	$\ell^{-\lambda}\lambda^y / y!$	$\lambda = \ell^{x'\beta}$

Source: Cameron, Colin A. and Pravin K. Trivedi. Microeconometrics. Methods and Applications, Cambridge, 2007, pág. 140.

Comparison: Stata and Gauss

The Log-Likelihood Function in Stata

$$\ln L = \sum_{j=1}^{N} \ln \phi \left\{ \left(y_i - \theta_{1j} \right) / \theta_{2j} \right\} - \ln \theta_{2j}$$

$$\theta_{1j} = \mu_j = \boldsymbol{x}_{1j} \beta_{1j}$$

$$\theta_{2j} = \sigma_j = \boldsymbol{x}_{2j} \beta_{2j}$$

<u>SYNTAX</u>

args lnf mu sigma

quietly replace `lnf'=ln(normalden($ML_y1, `mu', `sigma'))

The Log-Likelihood Function in Gauss

$$L = \sum_{i=1}^{N} \log P(Y_i;\theta)^{wi}$$

Where N is the number of observations, P(Yi, 0) is the probability of Yi given 0, a vector of parameters, and wi is the weight of the i-th observation.

```
proc loglik(theta,z);
local y,x,b,s;
x=z[.,1:cols(z)-1];
y=z[.,cols(z)];
b=theta[1:cols(x)];
s=theta[cols(x)+1];
retp(-0.5*ln(s^2)-0.5*(y-x*b)^2/s^2);
endp;
```

Procedure to maximize a likelihood functions in Stata

Syntax:

ml model lf my_normal f (y1=x1 x2)/sigma, technique(bhhh) vce(oim) waldtest(0)

ml maximize

Where:

Bhhh: Is a method of optimization Berndt-Hall-Hall-Hausman

Oim: Variance-covariance matrix (inverse of the negative Hessian matrix)

Procedure to maximize a likelihood functions in Gauss

Syntax:

{ x,f,g,cov,retcode } = MAXLIK(“dataset",vars,&loglik,theta0);

{ x,f,g,cov,retcode } = MAXLIK("psn",0,&lpsn,x0);

Where

Dataset: Data

Vars: Number of variables

&loglik: Log-likelihood

Theta0: numerical vector

Capabilities: Stata vs Gauss

- Methods
- Optimization Algorithms
- Covariance Matrix
- Inference Data and Inference
- Initial Values
- Report of the Coefficients
- Other

Stata’s Capabilities: Methods

CAPABILITIES	STATA	GAUSS
Linear Form Restrictions	YES (lnf)	NO
No Analytical Derivatives	YES (d0)	YES
Analytical Gradients First Derivative	YES (d1)	YES
Analytical Hessian Second Derivative	YES (d2)	YES

Stata's Capabilities: Optimization Algorithms

CAPABILITIES	STATA	GAUSS
Newton-Raphson (NR)	YES	YES
Berndt-Hall-Hall-Hausman (BHHH)	YES	YES
Davidon-Fletcher-Powell (DFP)	YES	YES
Broyden-Fletcher-Goldfarb-Shanno (BFGS)	YES	YES
SD (steepest descent)	NO	YES
Polak-Ribiere Conjugate Gradient	NO	YES

Stata's Capabilities: Covariance Matrix

CAPABILITIES	STATA	GAUSS
The inverse of the final information matrix from the optimization	YES	YES
The inverse of the cross-product of the first derivatives	YES	YES
The hetereskedastic-consistent covariance matrix	YES	YES

Stata's Capabilities: Data and Inference

CAPABILITIES	STATA	GAUSS
Robust	YES	YES
Cluster	YES	NO
Weights*	YES	YES
Survey Data	YES (SVY)	NO
Modification of the Sub-sample	YES	NO
Constrains	YES	NO
Wald Test	YES	YES
Switching**	YES	YES

* Stata contains frequency, probability, analytic and importance weights. Gauss have only frequency weight.

**Switching between algorithms.

Stata's Capabilities: Initial Values

CAPABILITIES	STATA	GAUSS
Initial Values	YES*	YES
Plot	YES	YES

*Stata searches for feasible initial values with ml search.

Stata's Capabilities: Report of the Coefficients

CAPABILITIES	STATA*	GAUSS
Hazard ratios (hr)	YES	NO
Incidence-rate ratios (irr)	YES	NO
Odds ratios (or)	YES	NO
Relative risk ratios (rrr)	YES	NO

*It must be used the command ml display

Stata's Capabilities: Other

CAPABILITIES	STATA	GAUSS
Convergence	YES	YES
Syntax Errors	YES	YES
Graph: the log-likelihood values	YES	YES

Example with the Poisson model

ESTIMATION IN STATA	ESTIMATION IN GAUSS
	library maxlik; maxset; proc lpsn(b,z); /* Function - Poisson Regression */

program define my_poisson
version 9.0
args lnf mu
quietly replace \`lnf' = $ML y1*ln(\`mu')-
\`mu' - lnfact($ML y1)
end
	local m;
	m = z[.,2:4]*b;
	retp(z[.,1].*m-exp(m));
	endp;
	
	proc lgd(b,z); /* Gradient */
	retp((z[.,1]-exp(z[.,2:4]*b)).*z[.,2:4]);
	endp;

ml model lf my_poisson f (y1=x1 x2)/
sigma, technique(bhhh) vce(oim)
ml maximize
	
	x0 = { .5, .5, .5 };
	_max_GradProc = &lgd;
	_max_GradCheckTol = 1e-3;
	{ x,f0,g,h,retcode } = MAXLIK("psn",
	0,&lpsn,x0);
	call MAXPrt(x,f0,g,h,retcode);

Conclusions

- Stata's features seems best suited for analyzing specific models of decision making processes and other micro-econometric applications.

- Gauss is ideal for analyzing a more ample range of statistical issues based on maximum likelihood estimation.