MAXIMUM LIKELIHOOD ESTIMATION

B. Essama-Nssah
Poverty Reduction Group (PRMPR)
The World Bank
Washington, DC

Module 1

Outline

- Principle
- Implementation
  - The Logl Object
  - Simple Regression Model
- Hands-On Exercise

Principle

- Consider a sample of **n** observations on two variables ($\mathbf{y_i}$, $\mathbf{x_i}$). Suppose the data are generated by a stochastic process of the following form.

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

- Where the ε_i are independent random normal variates with mean zero and standard deviation σ.
- To estimate the underlying parameters by the ML method, we need to find values of these parameters that would make the available observations most likely. Indeed, different population parameters are bound to generate different samples.

Principle

- The parameters of interest here are β_1, β_2, and the variance σ^2.
- To estimate these, we need to find estimators (i.e. formulae) that maximize the likelihood (or log likelihood) function based on sample observations.
- For implementation in EViews, it is more convenient to consider the log likelihood function as the sum of the log likelihood contributions for each observation **i**.

Principle

- Let φ(z) stand for the standard normal density function, then the contribution of each observation to the log likelihood is equal to:

$$l_i(\beta,\sigma) = \ln\phi(z_i) - \frac{1}{2}\ln(\sigma^2);\ z_i = \frac{(y_i - \beta_1 - \beta_2 x_i)}{\sigma}$$

- Thus the log likelihood is equal to:

$$l(\beta_1, \beta_2, \sigma^2) = \sum_{i=1}^{n} l_i(\beta_1, \beta_2, \sigma^2)$$

Principle

- It can be shown that, in particular, the MLE of β_1, β_2 satisfy the following orthogonality conditions which are equivalent to the OLS normal equations.

$$\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i) = 0 \qquad \sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)x_i = 0$$

- In general, when the **cross moment** of two random variables **x** and **y** is zero, **x** is said to be orthogonal to y (or equivalently y is orthogonal to **x**). The above equations are sample analogues since dividing each by **n** produces the sample cross moment.

Implementation

- The implementation of this method in EViews requires the specification of an object known as Logl.
- To illustrate, consider the following data on price (in cents per pound) and quantity of oranges (in pounds) sold in a supermarket on twelve consecutive days (Kmenta 1986:215)

X=(100, 90, 80, 70, 70, 70, 70, 65, 60, 60, 55, 50)

Y=(55, 70, 90, 100, 90, 105, 80, 110, 125, 115, 130, 130)

Implementation

- Declare the log likelihood object

  LOGL SIMREG

- Control statement for the series that will contain the likelihood contribution of each observation

  SIMREG.APPEND @LOGL LKLHD

- Create intermediate series

  SIMREG.APPEND RES=(Y-C(1)-C(2)*X)

- Create series to hold variance

  SIMREG.APPEND SIGMA2=C(3)

Implementation

- Specify likelihood contributions

```
SIMREG.APPEND
LKLHD=LOG(@DNORM(RES/@SQRT(SIGMA2)))-
LOG(SIGMA2)/2
```

- Get starting values from OLS

```
EQUATION OLSEQ.LS Y C X
!M=@NCOEF
FOR !K=1 TO !M
    C(!K)=(OLSEQ.C(!K))/4
NEXT
C(3)=(OLSEQ.@SE^2)/4
```

Implementation

- Invoke MLE
  SIMREG.ML(SHOWOPTS)
- Demo: MAXLIKELIHOOD.PRG
- When executing the assignment statements in the Logl object, EViews proceeds from top to bottom, evaluating the specification by observation (i.e. all assignment statements are evaluated for the first observation, then the second, etc.). It is possible to choose evaluation by equation.
- MLE requires computation of derivatives of the likelihood function with respect to the parameters. By default, EViews computes numeric derivatives. But one can use the @DERIV command to specify analytic derivatives to be used by EViews (see SIMREG_MLE.PRG).
  Syntax: @DERIV PNAME1 SNAME1 PNAME2 SNAME2 …
  (P for parameter of interest; S for corresponding derivative series)

Implementation

LogL: SIMREG
Method: Maximum Likelihood (Marquardt)
Date: 10/04/05   Time: 18:47
Sample: 1 12
Included observations: 12
Evaluation order: By observation
Estimation settings: tol= 0.00010, derivs=accurate numeric
Initial Values: C(1)=52.6111, C(2)=-0.39444, C(3)=17.4722
Convergence achieved after 40 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	210.4468	27.24774	7.723457	0.0000
C(2)	-1.577804	0.417338	-3.780639	0.0002
C(3)	58.24489	33.19740	1.754501	0.0793

Log likelihood	-41.41477	Akaike info criterion	7.402462
Avg. log likelihood	-3.451231	Schwarz criterion	7.523689
Number of Coefs.	3	Hannan-Quinn criter.	7.357580

Exercise

- Data underlying Example 19.1 from Greene (2000: 816): BINARY_MLE.WF1
- Issue: effectiveness a new teaching method.
- GRADE: binary dependent variable indicating whether students grades on an exam improved after exposure to new method.
  - PSI: indicator of exposure to new teaching method
  - GPA: grade point average.
  - TUCE: score on pretest indicating entry level knowledge.

Exercise

- Framework
  - Consider the probability model

$$y_i = \pi_i + \varepsilon_i$$

  - Where π_i is the probability that $\mathbf{y_i=1}$.
  - Suppose $\mathbf{y_i=1}$ if the following event happens

$$\beta' x_i + u_i > 0$$

  - Then

$$\pi_i = \Pr\{u_i > -\beta' x_i\} = 1 - F(-\beta' x_i)$$

Exercise

- Framework
  - Assuming **F()** symmetric

$$\pi_i = \Pr\{u_i > -\beta' x_i\} = \Pr\{u_i < \beta' x_i\} = F(\beta' x_i)$$

  - Thus the marginal effect of variable $\mathbf{x_k}$ on π_i is:

$$\frac{\partial \pi_i}{\partial x_{ik}} = f(\beta' x_i)\beta_k$$

  - The log likelihood function can be written as:

$$l(\beta) = \sum_{i=1}^{n} \left[(1-y_i)\ln[1-F(\beta' x_i)] + y_i \ln[F(\beta' x_i)]\right]$$

  - The associated first order conditions are:

$$\frac{\partial l(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} \frac{[y_i - F(\beta' x_i)]}{F(\beta' x_i)[1-F(\beta' x_i)]} f(\beta' x_i) x_{ik} = 0, \quad \forall k$$

Exercise

- Framework
  - The above first order conditions can be interpreted as a set of orthogonality conditions. Hence, the following can be viewed as generalized residuals.

$$e_{gi} = \frac{[y_i - F(\beta' x_i)]}{F(\beta' x_i)[1 - F(\beta' x_i)]} f(\beta' x_i)$$

Exercise

- Tasks
  - Use the BINARY command and the assumption that **F()** is the standard normal distribution to fit the above probability model to the data provided.
  - Repeat above task under the assumption that **F()** is the logistic distribution function.
  - Recall syntax:

    EQ_NAME.BINARY(OPTIONS) Y X1 X2 …

    Likelihood options: d=n for probit, d=l for logit.

    Other options: algorithm, max # of iterations, convergence criterion, etc..

Exercise

- Tasks
  - For both models (probit and logit) forecast the fitted value of the index $\beta' \mathbf{x}_i$, and compute the marginal effects of the explanatory variables at <u>mean</u> values.
  - Recall syntax:
    - Fitted Probabilities

      BINEQ.FIT PHAT
    - Fitted values of the index and density evaluation

      BINEQ.FIT(I) XBETA

      @DNORM(XBETA)

      @DLOGISTIC(XBETA)

Exercise

- Tasks
  - Focusing on the probit model, plot the fitted probabilities for grade improvement as a function of GPA for the two values of PSI (0 and 1), fixing the values of other variables at their sample means.
  - Base the analysis on values of for GPA in the interval [2, 4].

```
SERIES GPA_PLOT=2 +(4-2)*@TREND/(@OBSRANGE - 1)
```

Exercise

- Steps in Plotting **Probability Response Curves**
  - Create two appropriate index variables corresponding to the two values of PSI.
  - Declare a model object e.g. PSIMOD
  - Append the two equations describing the two possible grade outcomes:

```
PSIMOD.APPEND GRADE_0=@CNORM(INDEX_0)
PSIMOD.APPEND GRADE_1=@CNORM(INDEX_1)
```

  - Solve model for actuals

Exercise

- Plot the two endogenous variables against GPA_PLOT, and embellish the graph .

```
GROUP RESPONSE GPA_PLOT GRADE_0 GRADE_1
FREEZE(GPLOT) RESPONSE.XY
GPLOT.SCALE(L) RANGE(0,1)
GPLOT.SCALE(B) RANGE(2, 4)
GPLOT.LEGEND -DISPLAY
GPLOT.ADDTEXT(L) Probability of Grade Improvement
GPLOT.ADDTEXT(B) Grade Point Average
```

- Optional: Repeat the exercise using the Logl object.

Exercise

Dependent Variable: GRADE
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 10/05/05 Time: 16:42
Sample: 1 32
Included observations: 32
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-7.452320	2.542472	-2.931132	0.0034
GPA	1.625810	0.693882	2.343063	0.0191
TUCE	0.051729	0.083890	0.616626	0.5375
PSI	1.426332	0.595038	2.397045	0.0165

Mean dependent var	0.343750	S.D. dependent var	0.482559
S.E. of regression	0.386128	Akaike info criterion	1.051175
Sum squared resid	4.174660	Schwarz criterion	1.234392
Log likelihood	-12.81880	Hannan-Quinn criter.	1.111906
Restr. log likelihood	-20.59173	Avg. log likelihood	-0.400588
LR statistic (3 df)	15.54585	McFadden R-squared	0.377478
Probability(LR stat)	0.001405		
Obs with Dep=0	21	Total obs	32
Obs with Dep=1	11		

Exercise

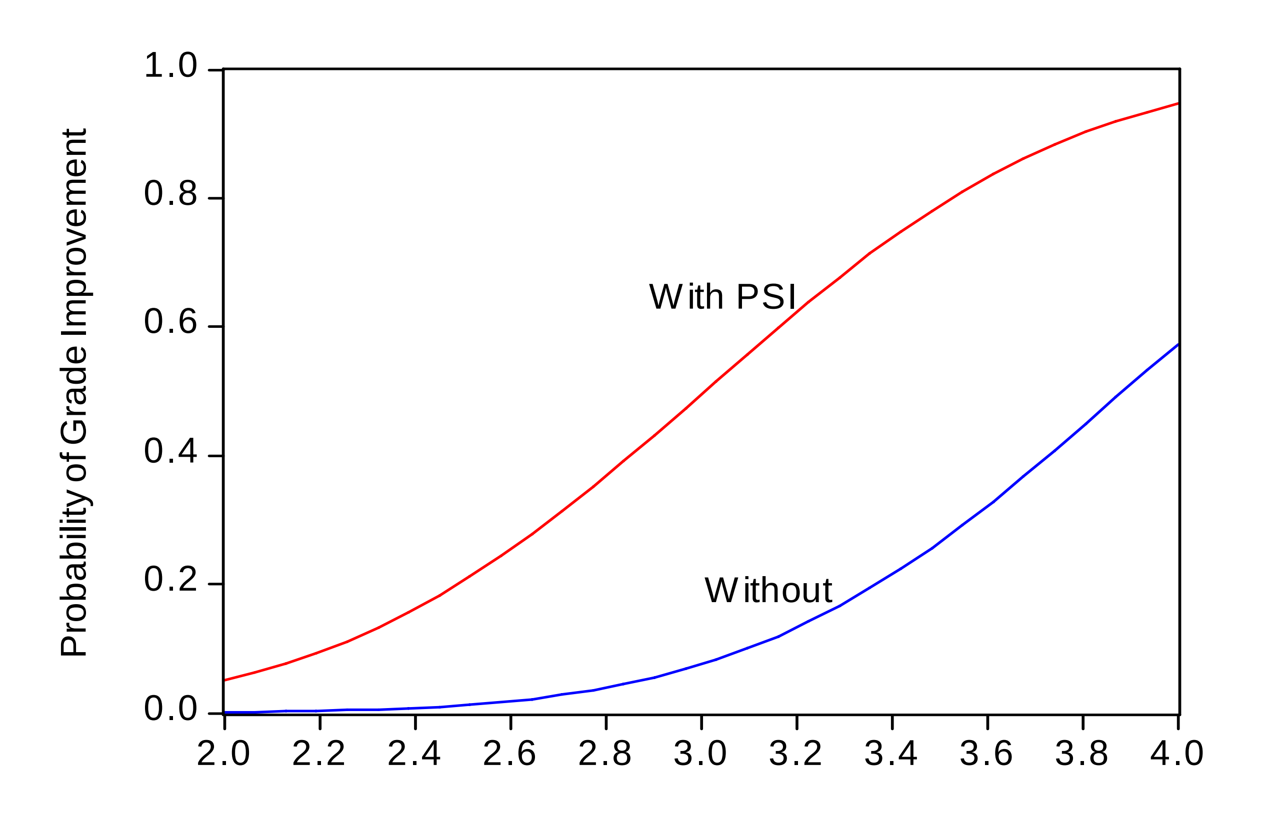
Dependent Variable: GRADE
Method: ML - Binary Logit (Quadratic hill climbing)
Date: 10/05/05 Time: 16:47
Sample: 1 32
Included observations: 32
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-13.02135	4.931324	-2.640537	0.0083
GPA	2.826113	1.262941	2.237723	0.0252
TUCE	0.095158	0.141554	0.672235	0.5014
PSI	2.378688	1.064564	2.234424	0.0255

Mean dependent var	0.343750	S.D. dependent var	0.482559
S.E. of regression	0.384716	Akaike info criterion	1.055602
Sum squared resid	4.144171	Schwarz criterion	1.238819
Log likelihood	-12.88963	Hannan-Quinn criter.	1.116333
Restr. log likelihood	-20.59173	Avg. log likelihood	-0.402801
LR statistic (3 df)	15.40419	McFadden R-squared	0.374038
Probability(LR stat)	0.001502		
Obs with Dep=0	21	Total obs	32
Obs with Dep=1	11		

Exercise

Effect of Teaching Method on Predicted Probabilities



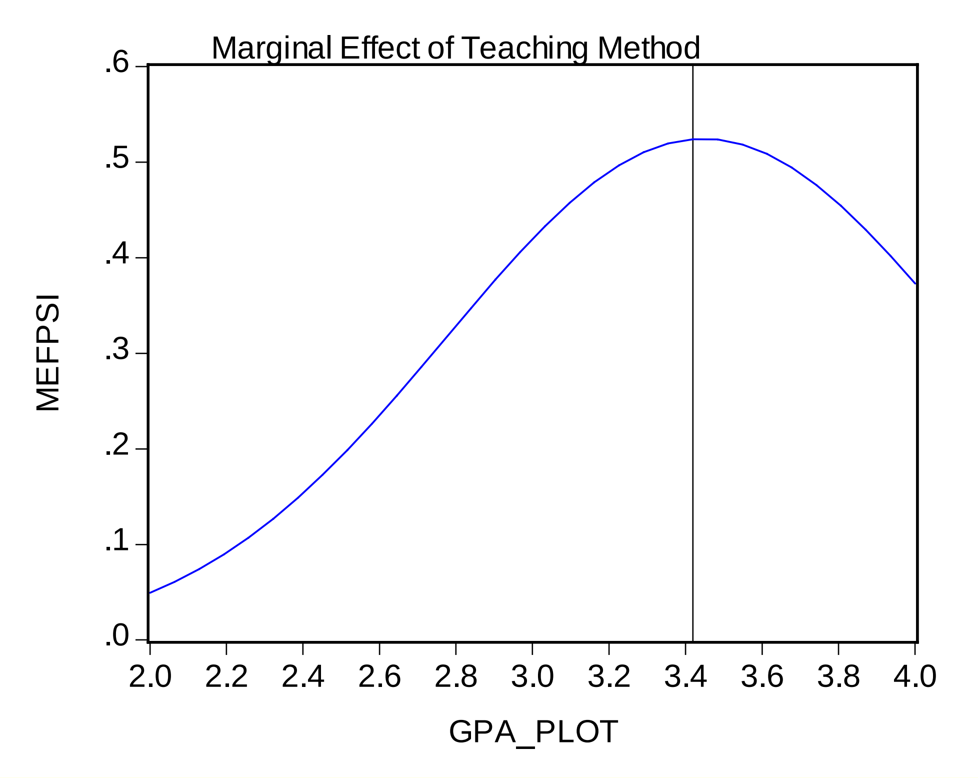
Exercise

- Both GPA and PSI have a significant and positive effect on the probability of grade improvement. The corresponding marginal effects (Probit model) are 0.53 for the grade point average and 0.47 for the teaching method.
- The curve labeled "Without" represents the counterfactual, showing the probability of grade improvement with no exposure to the method. The upper curve "With" shows the probability of improvement, given exposure to the new teaching method.

Exercise

- The marginal effect of the method is the difference between these two curves. It ranges from about 0.05 when GPA=2 to a maximum of about 0.53 when GPA=3.42 (see picture below).
- Hence, the probability that the a student's grade will improve after exposure to the new method is much higher for high than for low achievers (where achievement is indicated by the GPA).

Exercise


Marginal Effect of Teaching Method
.6
.5
.4
.3
.2
.1
.0
MEFPSI
2.0
2.2
2.4
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0
GPA_PLOT

References

- Greene William H. 2000. Econometric Analysis. Upper Saddle River: Prentice Hall.
- Kmenta Jan. 1986. Elements of Econometrics. New York: Macmillan Publishing Company.
- Quantitative Micro Software (QMS). 2005. EViews 5.1 User's Guide, Chapters 21 & 22. Irvine: QMS.

- The End.