FIML estimation of an endogenous switching model for count data in Stata 7

Alfonso Miranda
University of Warwick

Email: Alfonso.Miranda-Caso-Lunego@warwick.ac.uk

9th Stata UK Users' Group Meeting
May 19, 2003

Outline

- Introduction
- The Model
- **esp** Syntax
- Example
- Final comments

Introduction

- **esp** estimates a FIML endogenous switching Poisson count model (Terza, 1998).

- Cross-section Data.

- Alternative: Two-Stage method of moments (TSM) –not available in Stata.

- ml d0 method is used –but a lf method might be implemented.

THE MODEL

- Conditional on a vector of variables x_i, a dummy d_i, and a random term ε_i, the count y_i has a Poisson distribution

$$f(y_i \mid e_i) = \frac{\exp\left[-\exp(x_i'b + e_i)\right]\left[\exp(x_i'b + e_i)\right]^{y_i}}{y_i!}$$

- Given a vetor of variables z_i, d_i is characterised by

$$d_i = \begin{cases} 1 & if \;\; z_i'a + v_i > 0 \\ 0 & otherwise \end{cases}.$$

- Finally, e_i and v_i are jointly normal with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} s & sr \\ sr & 1 \end{pmatrix}.$$

- Conditional on ε_i, d_i and y_i are independent. Hence we can write

$$f(y_i, d_i) = \int_{-\infty}^{\infty} \{ d f(y_i \mid d_i = 1, e_i) \Pr(d_i = 1 \mid e_i) +$$
$$(1 - d) f(y_i \mid d_i = 0, e_i) \Pr(d_i = 0 \mid e_i) \} f(e_i) de_i$$

- Let w_i be all exogenous variables and consider a change of variable,

$$h_i = \frac{e_i}{s\sqrt{2}}.$$

After some algebra the joint conditional pdf of y_i and d_i becomes

$$f(y_i, d_i \mid w_i) = \frac{1}{\sqrt{p}} \int_{-\infty}^{\infty} \left\{ f(y_i \mid d_i, w_i, s\,h_i\sqrt{2}) \left\{ d_i \Phi^*(s\,h_i\sqrt{2}) + \right.\right.$$

$$\left.\left. + \ (1 - d_i)\Phi^*(-s\,h_i\sqrt{2}) \right\} \right\} \exp(-h_i^{\,2})\,dh_i,$$

where,

$$\Phi^*(s\,h\sqrt{2}) = \Phi\left(\frac{z_i'a \ + \ r\,h_i\sqrt{2}}{\sqrt{1 - r^2}} \right).$$

- Gauss-Hermite quadrature is then used to approximate the integral. Finally, the log-likelihood is simply

$$LogL = \sum_{i=1}^{n} \ln\left\{ f(y_i, d_i \mid w_i \right\}.$$

SYNTAX

esp *depvar* [*varlist*] [*if exp*] [*in range*],
edummy(*varname*) [switch(*varlist*)
quadrature(#) rho0(#) sigma0(#) mlopts]

- *depvar* declares the dependent variable
- *varlist* specifies covariates for the Poisson process
- *edummy* declares endogenous dummy
- *switch* specifies covariates for the switching variable
- *quadrature* indicates the number of quadrature points
- *rho0* sets initial value for ρ -default 0.01-
- *lnsigma0 sets initial value for s* -default 1.0-

EXAMPLE

- Fertility data from the National Survey of Demographic Dynamics 1997 (ENADID), Mexico.

- Dependent variable: Number of children ever born, **children (**includes children who died a few hours or days after birth).

- Endogenous dummy: Completed Primary School, **edu12** (switch variable).

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
age	19559	45.93481	4.210048	40	54
children	19559	4.428652	2.753055	0	18
catholic	19559	.8943709	.3073702	0	1
indspker	19559	.0940232	.2918685	0	1
after49	19559	.6184365	.4857827	0	1
edu12	19559	.5064676	.4999709	0	1

Comparison model: Poisson count with lognormal heterogeneity

```
. lnhp children catholic indspker after49 edu12,  q(16)

Getting Initial Values:
Fitting Full model:

Iteration 0:   log likelihood = -50093.352
<<omitted output>>
Iteration 8:   log likelihood = -44742.419

Poisson Regression--Lognormal Heterogeneity
(16 quadrature points)
                                                  Number of obs   =      19559
                                                  Wald chi2(4)    =    4067.23
Log likelihood = -44742.419                       Prob > chi2     =     0.0000
------------------------------------------------------------------------------
    children |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
children     |
    catholic |   -.037471   .0129797    -2.89   0.004    -.0629108   -.0120312
    indspker |   .0259195   .0134347     1.93   0.054     -.000412    .0522509
     after49 |  -.1498286   .0082753   -18.11   0.000    -.1660478   -.1336093
       edu12 |  -.4819543   .0084677   -56.92   0.000    -.4985507   -.4653579
       _cons |   1.778055   .0137364   129.44   0.000     1.751132    1.804978
-------------+----------------------------------------------------------------
       sigma |   .2965293   .0054761    54.15   0.000     .2859882     .307459
------------------------------------------------------------------------------
```

Example: esp

```
. esp children catholic indspker after49 edu12, ed(edu12) s(indspker after49) q(16)

Getting Initial Values:
Fitting Full model:

Iteration 0:   log likelihood = -63211.383  (not concave)
<<omitted output>>
Iteration 10:  log likelihood = -57578.991

Endogenous-Switch Poisson Regression
(16 quadrature points)
                                                  Number of obs   =      19559
                                                  Wald chi2(4)    =    2285.21
Log likelihood = -57578.991                       Prob > chi2     =     0.0000
------------------------------------------------------------------------------
    children |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
children     |
    catholic |  -.0355656   .0129769    -2.74   0.006    -.0609999   -.0101313
    indspker |  -.2627088   .0187704   -14.00   0.000    -.2994981   -.2259194
     after49 |  -.0722156   .0099639    -7.25   0.000    -.0917445   -.0526868
       edu12 |  -1.202968   .0268409   -44.82   0.000    -1.255575   -1.150361
       _cons |   2.118148   .0191674   110.51   0.000     2.080581    2.155716
-------------+----------------------------------------------------------------
switch       |
    indspker |   -1.09937   .0384229   -28.61   0.000    -1.174678   -1.024063
     after49 |   .2901192   .0218779    13.26   0.000     .2472393    .3329992
       _cons |   -.076211   .0169429    -4.50   0.000    -.1094186   -.0430035
-------------+----------------------------------------------------------------
       sigma |   .4632146   .0118455    39.10   0.000     .4405701     .487023
         rho |   .9539425   .0102185    93.35   0.000      .929037    .9702418
------------------------------------------------------------------------------
```

Final Comments

- Robust option can be implemented using ml lf method. Speed would also improve.
- FIML estimation of sample selection models for count data might be implemented with relatively minor modifications.
- Zero-inflated models are as well feasible.