Semi-nonparametric Estimation of Extended Ordered Probit Models

Mark Stewart
Economics Department
University of Warwick
e-mail: Mark.Stewart@warwick.ac.uk

UK Stata Users' Group Meeting, May 2003

Outline

- A framework for ordered response models
- Semi-nonparametric estimation
- Computation issues
- The `sneop` command: syntax and sample output
- An illustration: job satisfaction

Framework for ordered response models

Underlying latent variable model:

$$y_i^* = x_i'\beta + \varepsilon_i$$

for $i = 1, \ldots, n$
x: vector of observations on set of exogenous variables
β: vector of unknown parameters
ε: random error term, independently distributed, distribution function F

$$y_i = \begin{cases} 1 & \text{if } y_i^* < \alpha_1 \\ 2 & \text{if } \alpha_1 \leq y_i^* < \alpha_2 \\ \vdots & \\ J & \text{if } \alpha_{J-1} \leq y_i^* \end{cases}$$

Probabilities of observed outcomes

$$\Pr[y_i = j] = \begin{cases} F(\alpha_1 - x_i'\beta) & \text{if } \ j = 1 \\ F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta) & \text{if } \ 2 \leq j \leq J-1 \\ 1 - F(\alpha_{J-1} - x_i'\beta) & \text{if } \ j = J \end{cases}$$

More compactly:

$$\Pr[y_i = j] = F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta)$$

for all j, where $\alpha_0 = -\infty$ and $\alpha_J = +\infty$

Defines class of cumulative probability models in which known transformation of the cumulative probabilities is linear function of the x-variables

$$F^{-1}\left\{\Pr[y_i \leq j]\right\} = \alpha_j - x_i'\beta$$

Maximum Likelihood Estimator

$$\log L = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \log \left[F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta) \right]$$

where

$$y_{ij} = \begin{cases} 1 & \text{if } y_i = j \\ 0 & \text{else} \end{cases}$$

Maximized with respect to $(\beta, \alpha_1, \ldots, \alpha_{J-1})$
i.e. $M + J - 1$ parameters, where $M =$ number of exogenous variables
[β (and hence M) does not include an intercept]

Ordered Probit Model

Assumes $\varepsilon_i \sim N(0, \sigma^2)$

Adopting scale normalization $\sigma = 1$ and imposing zero intercept for identification, probabilities are given by

$$\Pr[y_i = j] = \Phi(\alpha_j - x_i'\beta) - \Phi(\alpha_{j-1} - x_i'\beta)$$

where Φ is CDF of standard Normal

Semi-nonparametric estimation

The "semi-nonparametric" series estimator of an unknown density, proposed by Gallant and Nychka (1987), approximates the density using a Hermite form

$$f_K(\varepsilon) = \frac{1}{\theta}\left(\sum_{k=0}^{K} \gamma_k \varepsilon^k\right)^2 \phi(\varepsilon)$$

where $\phi(\varepsilon)$ is standard normal density and $\theta = \int\limits_{-\infty}^{\infty} \left(\sum_{k=0}^{K} \gamma_k \varepsilon^k\right)^2 \phi(\varepsilon) d\varepsilon$.

$$F_K(u) = \frac{\int\limits_{-\infty}^{u} \left(\sum_{k=0}^{K} \gamma_k \varepsilon^k\right)^2 \phi(\varepsilon) d\varepsilon}{\int\limits_{-\infty}^{\infty} \left(\sum_{k=0}^{K} \gamma_k \varepsilon^k\right)^2 \phi(\varepsilon) d\varepsilon}$$

Defines family of SNP distributions for increasing values of K.

Semi-nonparametric estimator: properties

- Providing unknown density satisfies certain smoothness conditions, can be approximated arbitrarily closely by this Hermite series by increasing K.

- Class of densities that can be approximated is very general. Any form of skewness, kurtosis, etc. permitted, but upper bound on tails.

- Under these and other mild regularity conditions, providing K increases with sample size, model parameters are estimated consistently by maximization of the pseudo-likelihood function (Gallant and Nychka, 1987).

Semi-parametric identification

- Location normalization necessary. One used by Gabler et al. (1993) in binary case to give error term zero mean very cumbersome.

- Restriction can be on either error distribution or systematic part of model. Melenberg & van Soest (1996) set constant term equal to its probit estimate. Equivalent used here in ordered response model context is to set first threshold, α_1, equal to its ordered probit estimate.

- In $K = 1$ case, imposition of $E(\varepsilon) = 0$ or equivalent restriction used here implies $\gamma_1 = 0$, so model reduces to Ordered Probit.

- In general score for γ_2 is zero at Ordered Probit estimates. Implies that model for $K = 2$ is also equivalent to Ordered Probit. Model with $K = 3$ is therefore first model in series that is a generalization of the Ordered Probit.

Computation

SNP density can be written as

$$f_K(\varepsilon) = \frac{1}{\theta} \sum_{k=0}^{2K} \gamma_k^* \varepsilon^k \phi(\varepsilon)$$

where $\gamma_k^* = \sum_{i=a_k}^{b_k} \gamma_i \gamma_{k-i}$, with $a_k = \max(0, k-K)$ and $b_k = \min(k, K)$.

$$\theta = \int_{-\infty}^{\infty} \sum_{k=0}^{2K} \gamma_k^* \varepsilon^k \phi(\varepsilon) d\varepsilon = \sum_{k=0}^{2K} \gamma_k^* \mu_k = 1 + \sum_{k=1}^{K} \gamma_{2k}^* \mu_{2k}$$

where μ_k is k-th moment of standard normal.

Computation: cumulative probabilities

Use higher order truncated moments of standard normal distribution

$$F_K(u) = \frac{1}{\theta}\sum_{k=0}^{2K}\gamma_k^* \int_{-\infty}^{u} \varepsilon^k \phi(\varepsilon) d\varepsilon = \frac{1}{\theta}\sum_{k=0}^{2K}\gamma_k^* I_k(u)$$

where

$$I_k(u) = \int_{-\infty}^{u} \varepsilon^k \phi(\varepsilon) d\varepsilon$$

These are given by the recursion

$$I_k(u) = (k-1) I_{k-2}(u) - u^{k-1}\phi(u)$$

with $I_0(u) = \Phi(u)$ and $I_1(u) = -\phi(u)$.

Computation: simplified recursion

The truncated moments can be written as

$$I_k(u) = \mu_k \Phi(u) - A_k(u)\phi(u)$$

where the $A_k(u)$ are given by the recursion

$$A_k(u) = (k-1)A_{k-2}(u) + u^{k-1}$$

with $A_0(u) = 0$ and $A_1(u) = 1$. Can write cumulative probabilities as

$$F_K(u) = \frac{1}{\theta} \sum_{k=0}^{2K} \gamma_k^* \left[\mu_k \Phi(u) - A_k(u)\phi(u)\right] = \Phi(u) - \frac{1}{\theta} \left[\sum_{k=0}^{2K} \gamma_k^* A_k(u)\right] \phi(u)$$

Term in square brackets is polynomial in u of order $(2K-1)$.

The `sneop` command

Syntax

`sneop` depvar [varlist] [`if` exp] [`in` range] [`weight`] [, `order`(#) `robust` `from`(matname)]

`sneop` shares the features of all estimation commands; see help est.

Options

`order(#)` specifies the order of the Hermite polynomial to be used. The default is 3. Orders 1 and 2 give models equivalent to the Ordered Probit.

`robust` specifies that the Huber/White/sandwich estimator of variance is to be used in place of the traditional calculation.

`from(matname)` specifies a matrix containing starting values. (This option can be used to investigate whether a global maximum has been found.) The default uses the ordered probit estimates as starting values.

Sample Output

```
. use d:\micromet\ORM\fullauto\fullauto.dta
(Automobile Models)

. sneop rep77 foreign length mpg, order(5)

Order of SNP polynomial = 5
Number of categories = 5, rep77 assumed coded 1,...,5
1st threshold set to Ordered Probit estimate = 10.158904

initial:       log likelihood = -78.020025
(output deleted)
Iteration 11:  log likelihood = -77.099231

SNP Estimation of Extended Ordered Probit Model   Number of obs   =         66
                                                  Wald chi2(3)    =     333.84
Log likelihood = -77.099231                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
       rep77 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     foreign |   1.311066   .3460682     3.79   0.000     .6327847    1.989347
      length |   .0485854   .0030438    15.96   0.000     .0426197    .0545511
         mpg |   .1550694   .0200241     7.74   0.000     .1158229    .1943159
-------------+----------------------------------------------------------------
Tresholds: 1 |  10.158904
           2 |   11.14811   .3948177    28.24   0.000     10.37428    11.92193
           3 |   12.47549   .8385637    14.88   0.000     10.83193    14.11904
           4 |   14.04136   1.023637    13.72   0.000     12.03507    16.04765
-------------+----------------------------------------------------------------
SNP coefs: 1 |   -.388371   .9299086    -0.42   0.676    -2.210958    1.434216
           2 |  -.1782818   .1767031    -1.01   0.313    -.5246134    .1680498
           3 |   -.169297   .2396613    -0.71   0.480    -.6390245    .3004305
           4 |   .0252662   .0580207     0.44   0.663    -.0884523    .1389847
           5 |   .0189331   .0166684     1.14   0.256    -.0137365    .0516026
------------------------------------------------------------------------------
Likelihood ratio test of OP model against SNP extended model:
Chi2(3) statistic =     1.8415875      (p-value = .60592711)
------------------------------------------------------------------------------
Estimated moments of error distribution:
Variance =              1.3902935      Std Dev =                1.1791071
3rd. moment =           2.1202618      Coef of skewness =       1.2933907
4th. moment =           13.943124      Coef of kurtosis =       7.2135173
------------------------------------------------------------------------------
```

Job satisfaction illustration

[Clark & Oswald, "Satisfaction and comparison income", J. Public Economics, 1996.] Data from 1991 wave of BHPS. Working respondents provided job satisfaction scores: 1 ("not satisfied at all") to 7 ("completely satisfied").

Table 1: Likelihood-Ratio Tests for Different Values of K

K	log-likelihood	LR-test of OP	degr. of freedom	p-value	LR-test of K-1	p-value
OP	-6174.25					
3	-6169.87	8.76	1	0.003	8.76	0.003
4	-6167.95	12.60	2	0.002	3.83	0.050
5	-6165.48	17.54	3	0.001	4.95	0.026
6	-6165.31	17.87	4	0.001	0.33	0.568
7	-6164.56	19.38	5	0.002	1.51	0.219
8	-6164.56	19.39	6	0.004	0.01	0.939

Note: 1 degree of freedom for LR test of K-1 in each row.

Table 2: Job Satisfaction Models for Different Values of K

	OP	SNP(3)	SNP(5)
	coef (s.e.)	coef (s.e.)	coef (s.e.)
log(earnings)	0.134 (.054)	0.096 (.050)	0.087 (.056)
log(comp. earn.)	-0.283 (.064)	-0.254 (.060)	-0.323 (.068)
Male	-0.156 (.044)	-0.147 (.044)	-0.130 (.046)
Age/10	-0.210 (.100)	-0.154 (.092)	-0.141 (.104)
$Age^2/1000$	0.380 (.126)	0.310 (.115)	0.305 (.131)
Log-likelihood	-6174.25	-6169.87	-6165.48
Standard deviation	1	0.979	1.369
Skewness coefficient	0	0.034	0.064
Kurtosis coefficient	3	4.600	4.665
Test sum=0 $[\chi^2(1)]$	7.40	9.26	15.83
p-value	[0.007]	[0.002]	[0.000]

Notes: (1) Sample size = 3895. (2) Models also contain 34 other variables.

Table 3: Derived statistics

	OP	SNP(3)	SNP(5)
	coef (s.e.)	coef (s.e.)	coef (s.e.)
log(earnings)	0.134 (.054)	0.096 (.050)	0.087 (.056)
log(comp. earn.)	-0.283 (.064)	-0.254 (.060)	-0.323 (.068)
Trade off	2.12 (0.70)	2.66 (1.15)	3.70 (2.04)
Age/10	-0.210 (.100)	-0.154 (.092)	-0.141 (.104)
$Age^2/1000$	0.380 (.126)	0.310 (.115)	0.305 (.131)
Age at minimum	27.6 (4.4)	24.8 (6.1)	23.1 (7.5)

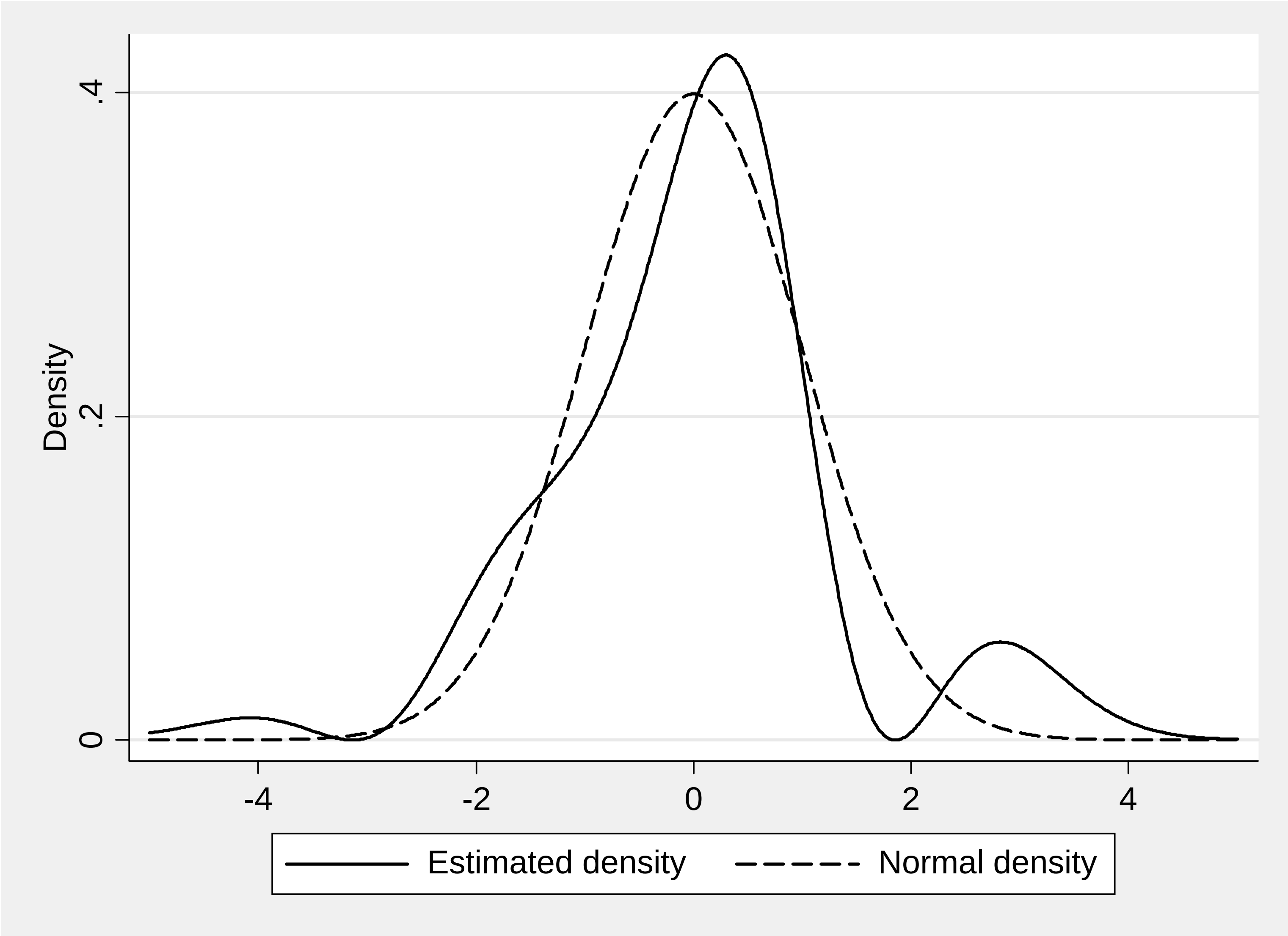

Figure 1: Estimated Density for K = 5