Semi-nonparametric Estimation of Extended Ordered Probit Models

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

June 2003

Abstract

This paper presents a semi-nonparametric estimator for a series of generalized models that nest the Ordered Probit model and thereby relax the distributional assumptions in that model. It describes a new Stata command for the estimation of such models and presents an illustration of the approach.

Keywords: ordered response models, ordered probit, semi-nonparametric estimation.

1 Introduction

The analysis of ordered response data is increasingly common in a growing number of research areas. The standard Ordered Probit (Aitchison and Silvey, 1957) and Ordered Logit (McCullagh, 1980) models are commonly used, but require the researcher to make a specific distributional assumption.[1] This paper presents a semi-nonparametric estimator for a series of generalized models that nest the Ordered Probit model and relax the distributional assumption in that model.

2 A framework for ordered response models

As with many limited dependent variable models, it is useful to consider specification of a suitable model for ordered response data in terms of an underlying latent variable:

$$y_i^* = x_i'\beta + \varepsilon_i \tag{1}$$

for $i = 1, \ldots, N$, where x_i is a vector of observations on a set of explanatory variables, β is a vector of unknown parameters and ε_i is a random error term independently distributed with distribution function F. If y_i^* was observed for all observations, β could be consistently estimated by Ordinary Least Squares without a distributional assumption being required on ε. For the type of data considered in this paper y_i^* itself is not observed. Rather the observed dependent variable, y_i, is discrete, taking one of

---

[1] See Greene (2003) and Wooldridge (2002) for recent textbook treatments of such models.

the values $\{1, 2, \ldots, J\}$, and is related to y_i^* as follows:

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

with the α_j being additional parameters such that $\alpha_1 < \alpha_2 < \ldots < \alpha_{J-1}$. Thus the range of y^* is partitioned into J mutually exclusive and exhaustive intervals and the variable y indicates the interval into which a particular observation falls. The dependent variable y is ordinal and the α_j are treated as parameters to be estimated. The probability of a particular observed outcome, for $2 \le j \le J - 1$, is given by

$$\begin{aligned} \Pr[y_i = j] &= \Pr[\alpha_{j-1} \le y_i^* < \alpha_j] \\ &= \Pr[\alpha_{j-1} - x_i'\beta \le \varepsilon_i < \alpha_j - x_i'\beta] \\ &= F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta) \end{aligned}$$

where F is the cumulative distribution function of ε_i and is assumed to contain no additional unknown parameters, so that, for example, ε_i has a known variance. This assumption fixes the scale of the measurement of y^*, but not the origin. Identification can be achieved by assuming a zero intercept (i.e. that x_i does not contain a constant term) or by fixing one of the α_j. The former is used here.

The full set of probabilities of the possible outcomes are

$$\Pr[y_i = j] = \begin{cases} F(\alpha_1 - x_i'\beta) & \text{if } j = 1 \\ F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta) & \text{if } 2 \le j \le J - 1 \\ 1 - F(\alpha_{J-1} - x_i'\beta) & \text{if } j = J \end{cases} \tag{3}$$

Adopting the additional notation that $\alpha_0 = -\infty$ and $\alpha_J = +\infty$, these can be written more compactly as

$$\Pr[y_i = j] = F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta) \tag{4}$$

for all j. This defines a class of cumulative probability models in which a known transformation of the cumulative probabilities is taken to be a linear function of the x-variables and only the intercept in this function differs across the categories:

$$F^{-1}\{\Pr[y_i \le j]\} = \alpha_j - x_i'\beta \tag{5}$$

A natural estimator for this type of model is the Maximum Likelihood estimator. Define

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

Then the log-likelihood for the model is given by

$$\log L = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \log \left[F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta) \right] \tag{6}$$

This is maximized with respect to $(\beta, \alpha_1, \ldots, \alpha_{J-1})$, i.e. $M + J - 1$ parameters, where M is the number of exogenous variables, remembering that β (and hence M) does not include an intercept.

By far the most commonly used models to date for the analysis of ordered responses have been the Ordered Probit and Ordered Logit models (which take F to be standard Normal and Logistic respectively). The Ordered Probit model, introduced by Aitchison and Silvey (1957), assumes that $\varepsilon_i \sim N(0, \sigma^2)$. Adopting the scale normalization $\sigma = 1$ and imposing a zero intercept for identification, the probabilities are given by

$$\Pr[y_i = j] = \Phi(\alpha_j - x_i'\beta) - \Phi(\alpha_{j-1} - x_i'\beta) \tag{7}$$

where Φ is the cumulative distribution function of a standard Normal and the log-likelihood by (6) with F replaced by Φ.

3 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. This approximation can be written as the product of a squared polynomial and a normal density. This gives a polynomial expansion with Gaussian leading term. In fact an expansion around any density with a moment-generating function could be used, but the normal is a natural choice in the current context, because the resulting model then nests the Ordered Probit. To ensure that the density approximation is proper (i.e. that it integrates to 1, in addition to the non-negativity ensured by the square), the approximation is specified as

$$f_K(\varepsilon) = \frac{1}{\theta} \left(\sum_{k=0}^{K} \gamma_k \varepsilon^k \right)^2 \phi(\varepsilon) \tag{8}$$

where $\phi(\varepsilon)$ is the standard normal density function and where

$$\theta = \int_{-\infty}^{\infty} \left(\sum_{k=0}^{K} \gamma_k \varepsilon^k \right)^2 \phi(\varepsilon) \, d\varepsilon \tag{9}$$

However, this general specification of the density is invariant to multiplication of the vector $\gamma = (\gamma_0, \gamma_1, ..., \gamma_K)$ by a scalar and a normalization is required. The restriction $\gamma_0 = 1$ is a convenient choice.

The required distribution function is therefore specified as

$$F_K(u) = \frac{\int_{-\infty}^{u} \left(\sum_{k=0}^{K} \gamma_k \varepsilon^k \right)^2 \phi(\varepsilon) \, d\varepsilon}{\int_{-\infty}^{\infty} \left(\sum_{k=0}^{K} \gamma_k \varepsilon^k \right)^2 \phi(\varepsilon) \, d\varepsilon} \tag{10}$$

This defines a family of "semi-nonparametric" (SNP) distributions for increasing values of K.

Providing the unknown density satisfies certain smoothness conditions, it can be approximated arbitrarily closely by this Hermite series by increasing the choice of K, the degree of the polynomial. The class of densities that can be approximated by this form is very general. Any form of skewness, kurtosis, etc. is permitted, but "violently oscillatory" density functions are ruled out. In addition to the smoothness restriction, there is an upper bound on the tails of the densities. (See Gallant and Nychka, 1987, page 369, for a fuller discussion of the class of densities.)

Under these and other "mild" regularity conditions, and providing K increases with the sample size, the model parameters are estimated consistently (Gallant and Nychka, 1987) by maximization of the pseudo-likelihood function got by replacing the unknown distribution function in (6) by that in (10).

A location normalization is also necessary for semi-parametric identification. That used by Gabler et al. (1993) in the binary response case to give the error term a zero mean is rather cumbersome. However, the restriction can be on either the error distribution or the systematic part of the model. Melenberg and van Soest (1996) use an alternative normalization, setting the constant term in their model equal to its probit estimate. This is far more convenient. The equivalent used here in the ordered response model context is to set the first threshold, α_1, equal to its ordered probit estimate.

Note that in the $K = 1$ case, the imposition of $E(\varepsilon) = 0$ can be easily shown to imply $\gamma_1 = 0$, so that the model in this case reduces to the Ordered Probit model. The imposition of the equivalent restriction used here has the same implication. An important additional feature of the model in the general case is that the score for γ_2 is zero at the Ordered Probit estimates. (This is equivalent to the result for the binary probit case proved in Appendix B of Gabler et al., 1993.) This implies that the model for $K = 2$ is also equivalent to the Ordered Probit model (as well as those for $K = 0$ and $K = 1$). The model with $K = 3$ is therefore the first model in the series that is a generalization of the Ordered Probit model.

In practice, inference is conducted conditional on K, possibly for a range of alternative values of K, with the final specification of the model chosen by tests between them. Thus in reality the model is treated as parametric for a given K, with the choice of K part of the model selection procedure. As Pagan and Ullah (1999) point out, the orientation is non-parametric, but the modus operandi is parametric. This latter characteristic is appealing, because estimation can be conducted in a familiar Maximum Likelihood environment.

The SNP-ML approach can therefore be viewed as a series of polynomial densities and corresponding pseudo-likelihood functions, and standard model selection procedures used to reduce the dimension of the parameter vector and improve efficiency in finite samples. This can, for example, be on the basis of Wald or Likelihood-ratio tests on the γ-vector, or using information criteria, such as those of Akaike, Schwarz or Hannan and Quinn.

The polynomial form of (8) means that the cumulative probabilities given by (10), required for the pseudo-likelihood function, can be derived using the higher order truncated moments of a standard normal distribution. First, the SNP density can be written as

$$f_K(\varepsilon) = \frac{1}{\theta} \sum_{k=0}^{2K} \gamma_k^* \varepsilon^k \phi(\varepsilon) \tag{11}$$

where

$$\gamma_k^* = \sum_{i=a_k}^{b_k} \gamma_i \gamma_{k-i}$$

with $a_k = \max(0, k - K)$ and $b_k = \min(k, K)$. The scaling factor to give a proper density is then given by

$$\theta = \int_{-\infty}^{\infty} \sum_{k=0}^{2K} \gamma_k^* \varepsilon^k \phi(\varepsilon) \, d\varepsilon = \sum_{k=0}^{2K} \gamma_k^* \mu_k \tag{12}$$

where μ_k is the k-th moment of a standard normal. Given that $\gamma_0^* = \gamma_0^2 = 1$ and that $\mu_k = 0$ for k odd, this can be written as

$$\theta = 1 + \sum_{k=1}^{K} \gamma_{2k}^* \mu_{2k} \tag{13}$$

The cumulative probabilities in (10) required for the pseudo-likelihood function can be written as

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

This is a linear combination of the truncated moments of a standard normal:

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

Computation of the likelihood contributions for the Maximum Likelihood iterations can then be simplified by making use of the recursion

$$I_k(u) = (k-1)I_{k-2}(u) - u^{k-1}\phi(u) \tag{14}$$

with $I_0(u) = \Phi(u)$ and $I_1(u) = -\phi(u)$. Each truncated moment can be written as a polynomial combination of the standard normal cumulative distribution and density functions. The formulation can be further simplified in the current context where each cumulative probability is a linear combination of truncated moments scaled by the corresponding combination of complete moments. It can be shown that the recursion need only involve the normal density function part of the expressions for these truncated moments. The cumulative probabilities in (10) can be written as

$$F_K(u) = \frac{1}{\theta} \sum_{k=0}^{2K} \gamma_k^* I_k(u)$$

with θ given by the corresponding combination in equation (12). Using the recursion in equation (14), the truncated moments in this linear combination can be shown to also be given by

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

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 $A_1(u) = 1$. Using this the cumulative probabilities in (10) required for the pseudo-likelihood function can be written as

$$\begin{aligned} F_K(u) &= \frac{1}{\theta} \sum_{k=0}^{2K} \gamma_k^* \left[\mu_k \Phi(u) - A_k(u)\phi(u) \right] \\ &= \Phi(u) - \frac{1}{\theta} \left[\sum_{k=0}^{2K} \gamma_k^* A_k(u) \right] \phi(u) \end{aligned} \tag{15}$$

The term in square brackets is a polynomial in u of order $(2K-1)$. Thus the cumulative probabilities in the SNP framework have the standard normal cumulative distribution function as leading term and differ from this by the product of the standard normal density and this polynomial in u. The SNP estimator is then given by the maximization of the pseudo-likelihood function got by replacing the unknown distribution function in (6) by this final expression in (15).

4 The sneop command

4.1 Syntax

sneop depvar [varlist] [if exp] [in range] [weight] [, order(#) <u>r</u>obust <u>from</u>(matname)]

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

4.2 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; see [U] **23.11 Obtaining robust variance estimates**.

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.

4.3 Example using Stata Reference Manual data

As an example of the use of the command and the output produced, this section uses the automobile data (http://www.stata-press.com/data/r8/fullauto.dta) used in [R] **oprobit** to model the 1977 repair records of 66 foreign and domestic cars. The variable rep77 takes values poor, fair, average, good and excellent (coded 1 to 5). The explanatory variables in the model are foreign (origin of manufacture), length (a proxy for size) and mpg. All three have significant positive effects in the estimated ordered probit model. The output from the sneop program with $K = 5$ is as follows:

```
. use 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
------------+-----------------------------------------------------------------
Thresholds 1|   10.1589      Fixed
          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 =      1.841588      (p-value = .6059271)
-------------------------------------------------------------------------------
Estimated moments of error distribution:
Variance =               1.390294     Standard Deviation =     1.179107
3rd moment =             2.120262     Skewness =               1.293391
4th moment =             13.94312     Kurtosis =               7.213517
-------------------------------------------------------------------------------
```

The log-likelihood is improved very little relative to the ordered probit model and the likelihood-ratio test does not reject the ordered probit model: a $\chi^2(3)$-statistic of 1.84, giving a p-value of 0.61. This example is provided to illustrate the output produced by the program on the data and model used in [R] **oprobit**. It is, however, unlikely that we would be able to observe a significant difference between the fitted density and a standard normal on the basis of a sample as small as this.

5   Job satisfaction illustration

The use of the semi-nonparametric estimator is illustrated in more detail in this section on a model of self-declared job satisfaction, as specified by Clark and Oswald (1996), and using data from the 1991 wave of the British Household Panel Survey. Working respondents provided job satisfaction scores ranging from 1 ("not satisfied at all") to 7 ("completely satisfied"). Clark and Oswald examine the hypothesis that job satisfaction depends on income relative to a "comparison" or reference level. Their main findings are that job satisfaction levels are positively related to own earnings and negatively related to "comparison" earnings. Other often-quoted features of their results are the higher job satisfaction levels of women and the U-shaped age profile.

Log-likelihood values and likelihood-ratio statistics for different values of K from 3 to 8 are given in Table 1. The standard ordered probit (OP) model is rejected in all cases. The likelihood-ratio tests for $K-1$ against K reject the null, at conventional significance levels, for $K \leq 5$, but accept it for K above 5, suggesting selection of the $K = 5$ model. This use of likelihood-ratio tests runs into the usual problem with sequential testing procedures, which could be addressed with successive revision of the significance level.

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

K	log-likelihood	LR-test of OP	deg. 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.

This is a standard model selection problem. There are a number of alternative model selection criteria available (Greene, 2003). The best known is probably the Akaike Infor-

mation Criterion (AIC), equal to the log-likelihood minus the number of freely estimated parameters. However in some cases the AIC has been found to have a tendency towards the selection of models with too many parameters, even asymptotically. A commonly used alternative is the Schwarz Bayesian Criterion (SBC), based on a large sample approximation to the posterior odds ratio of the models. Under certain conditions the SBC selects the right model with probability 1 asymptotically (Amemiya, 1985), but in some finite sample cases it has been found to have a tendency to underfit. A third possibility is the Hannan-Quinn Criterion (HQC), introduced in the context of autoregressive time-series models, but more generally applicable. It gives values between the AIC and SBC and is sometimes viewed as a useful compromise. For the models in Table 1, the AIC and the HQC both choose $K = 5$, while the SBC, which penalizes extra parameters more heavily, selects $K = 3$. (For a sample of the size used here (3,895 observations) the SBC penalizes an extra parameter 4.13 on the log-likelihood scale, while the HQC and AIC penalize it 2.11 and 1 respectively.)

The estimated parameters (coefficients on main variables only) for the ordered probit model and for the extended model with $K = 3$ and $K = 5$ are given in Table 2.[2] A number of other statistics are also given.

Table 2: Job Satisfaction Models for Different Values of K

	OP coef (s.e.)	SNP(3) coef (s.e.)	SNP(5) 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^2/100	0.038 (.013)	0.031 (.011)	0.030 (.013)
Thresholds:			
1	-4.125 (.377)	-4.125	-4.125
2	-3.917 (.376)	-3.879 (.043)	-3.847 (.047)
3	-3.562 (.375)	-3.476 (.093)	-3.390 (.075)
4	-2.995 (.375)	-2.877 (.161)	-2.603 (.125)
5	-2.416 (.374)	-2.319 (.211)	-1.889 (.164)
6	-1.664 (.374)	-1.657 (.254)	-1.151 (.177)
Polynomial:			
1		-0.050 (.193)	0.415 (.062)
2		-0.097 (.088)	0.366 (.252)
3		-0.051 (.021)	-0.073 (.171)
4			-0.087 (.034)
5			-0.002 (.016)
Log-likelihood	-6174.25	-6169.87	-6165.48
Standard deviation	1	0.979	1.369
Skewness	0	0.034	0.064
Kurtosis	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.

---

[2] The other variables included in the model are as in Table 3 of Clark and Oswald (1996). The data used here are from the current release of Wave 1, while they used the original release. The ordered probit results are very close to theirs, but not identical.

Estimates of β cannot be compared directly across models without adjustment, because the fitted densities differ. The β coefficients are only identified after imposition of scale and origin restrictions (as is the case for the ordered probit model). Thus, even if considering effects on the latent dependent variable y_i^*, it is not appropriate to compare estimates of the β coefficients across the models directly.

However, there are a number of ways in which comparisons can be made, depending on the focus of the study. We might hypothesize that there is an underlying, but unobservable, continuous measure of job satisfaction (or utility or welfare depending on one's perspective) that is an unknown, but monotonic, function of the variable y_i^* in equation (1). In this case one would be interested in influences on this underlying variable and relations between them.

Consider two factors that influence job satisfaction (i.e. two elements of the vector x in equation (1)). We could plot "indifference curves" between them: combinations of the two factors that give the same level of job satisfaction. The specification in (1) means that they will be straight lines. The slopes of these (parallel) indifference curves between the two factors will be minus the ratio of their β-coefficients. Thus ratios of coefficients are of interest and are compared below.

Ratios of β-coefficients are also informative about the impacts of x-variables on the probabilities of observed outcomes (or cumulations thereof). The partial derivatives of the probabilities of y with respect to x, the "marginal effects", are proportional to β with the factor of proportionality being an evaluation of the density function (or a difference in densities). For example, from equation (3), the partial derivative of $\Pr[y_i = J]$ with respect to x is $\beta.f(\alpha_{J-1} - x_i'\beta)$. The "marginal effects" are therefore scaled versions of the β coefficients and the scaling factor is the same for each element of β. Hence a ratio of "marginal effects" is equal to the corresponding ratio of β-coefficients. Thus ratios of coefficients are of interest from this perspective also.

Two such ratios are illustrated in Table 3. The first is minus the ratio of the two earnings coefficients. In the ordered probit model the ratio indicates that a 10% increase in comparison earnings would need to be compensated by a 21% increase in own earnings to give the same level of job satisfaction. This rises to 26% in the SNP model with $K = 3$ and 37% in the SNP model with $K = 5$. The standard errors on these ratios, calculated by the "delta method" (using the Stata nlcom command), are also given in Table 3. They indicate that these ratios are not very precisely estimated and the differences in them between models, while appreciable, are small relative to the standard errors. The test of the hypothesis that the coefficients on the two earnings variables sum to zero, which would be the case if only relative earnings mattered, is rejected increasingly strongly (see Table 2). The $\chi^2(1)$ Wald statistic rises from 7.4 in the ordered probit model to 9.3 when $K = 3$ is used and 15.8 when $K = 5$ is used.

Table 3: Some derived statistics

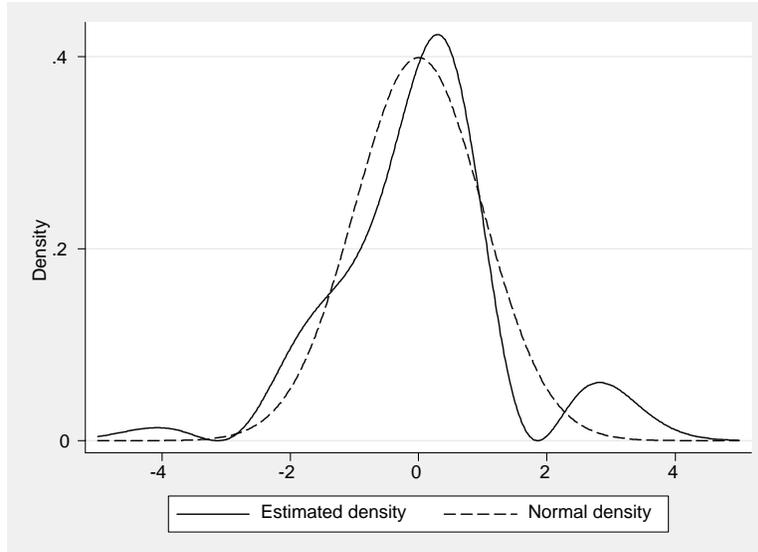|	OP	SNP(3)	SNP(5)
	coef (s.e.)	coef (s.e.)	coef (s.e.)
log(earnings)	0.134 (.054)	0.096 (.050)	0.087 (.056)
log(comp. earn.)	-0.283 (.064)	-0.254 (.060)	-0.323 (.068)
Trade off	2.12 (0.70)	2.66 (1.15)	3.70 (2.04)
Age/10	-0.210 (.100)	-0.154 (.092)	-0.141 (.104)
Age^2/100	0.038 (.013)	0.031 (.011)	0.030 (.013)
Age at minimum	27.6 (4.4)	24.8 (6.1)	23.1 (7.5)

Figure 1: Estimated Density for $K = 5$

Another coefficient ratio of interest is that involving the two coefficients of the quadratic in age. The minimum of the job satisfaction age profile is minus half the ratio of coefficients. It is estimated to be at about age 28 on the basis of the ordered probit model, but rather earlier with both the SNP models: age 25 if $K = 3$ is used and age 23 if $K = 5$ is used. These seem like important differences. However, the standard errors given in Table 3 indicate that these ratios also are not very precisely determined. As a result the differences across the models in the estimated age at the minimum of the age-satisfaction profile are small relative to the standard errors.

Both SNP models exhibit positive skewness and both indicate greater kurtosis than in the standard normal (Table 2). Figure 1 gives a plot of the estimated density for the $K = 5$ model. Both the positive skew and the fatter tails than the normal are evident. The central mode of the estimated density is at 0.30. Its height there is 0.423, compared with 0.399 for the standard normal density at its mode. The estimated density also has secondary modes in both tails. The larger of the two is at $\varepsilon = 2.82$, with density $f_5(\varepsilon) = 0.061$. There is also a smaller one at $\varepsilon = $ -4.07, $f_5(\varepsilon) = 0.014$.

6   Remarks on using sneop

1. The log-likelihood of this model is not globally concave. It is therefore good practice to check for convergence to the global maximum rather than a local one by running the algorithm from different starting points using the from option. A natural alternative to the OP starting values is the SNP model for $K - 1$.

2. The program can be used to obtain SNP estimates of a binary response model. The binary response needs to be recoded to take the values 1 and 2.

3. It should be emphasized that this is a large sample technique. Small samples are unlikely to be able to detect divergencies from normality. Consequently, results on small samples can be expected to differ little from the oprobit estimates.

4. The desirable properties of the estimator, as well as requiring a large sample, also depend on K increasing with N. Experience with the estimator suggests that

selecting K on the basis of either the criteria described above or likelihood-ratio tests does in fact do this.

5. Using `predict` with the `xb` option after `sneop` creates a new variable containing $x_i'\widehat{\beta}$.

6. The program is much slower than `oprobit`, both because of the greater complexity of the likelihood function to be maximized and because it is written as an `ado` file. The time required depends on the order of the polynomial used (K), the number of observations (N) and the number of categories (J) (as well as the speed of the processor), but little affected by the number of explanatory variables. Experiments with the program indicate a tight relationship for the time per iteration, but only a rather loose one for the number of iterations required. As a guide, the combination $N = 1000, K = 3, J = 7$ requires about 10 seconds per iteration on a 1.2 GHz Pentium III processor. This increases or decreases roughly proportional to N and roughly proportional to the squares of K and J, although in all three cases there are slight economies of scale. The number of iterations required is far more variables, but seems roughly independent of N, J and the number of x-variables. As a rough rule of thumb, convergence tends to require approximately $5(K - 1)$ iterations.

7 Acknowledgements

8 References

Aitchison, J. and S. D. Silvey (1957), "The Generalization of Probit Analysis to the Case of Multiple Responses", Biometrika, **44**, 131-140.

Amemiya, Takeshi (1985), Advanced Econometrics, Cambridge: Harvard University Press.

Clark, Andrew E. and Andrew J. Oswald (1996), "Satisfaction and comparison income", Journal of Public Economics, **61**, 359-381.

Gabler, Siegfried, François Laisney and Michael Lechner (1993), "Seminonparametric Estimation of Binary-Choice Models with an Application to Labor-Force Participation", Journal of Business and Economic Statistics, **11**, 61-80.

Gallant, A. Ronald and Douglas N. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation", Econometrica, **55**, 363-390.

Greene, William H. (2003), Econometric Analysis, 5th. edition, Prentice-Hall.

McCullagh, Peter (1980), "Regression Models for Ordinal Data", Journal of the Royal Statistical Society, Series B, **42**, 109-142.

Melenberg, Bertrand and Arthur van Soest (1996), "Measuring the costs of Children: Parametric and Semiparametric Estimators", Statistica Neerlandica, **50**, 171-192.

Pagan, Adrian and Aman Ullah (1999), Nonparametric Econometrics, Cambridge: Cambridge University Press.

Wooldridge, Jeffrey M. (2002), Econometric Analysis of Cross Section and Panel Data, Cambridge: MIT Press.