A Comparison of Semiparametric Estimators for the Ordered Response Model

Mark B. Stewart

Economics Department, University of Warwick, Coventry CV4 7AL, UK.

Abstract

Three semi-parametric estimators for ordered response models are compared: a "semi-nonparametric" series estimator, an estimator based on a polynomial transformation to normality, and a least squares estimator for a transformed dependent variable. The empirical estimates are found to differ markedly across estimators in a job satisfaction illustration presented, but much less so in the second illustration, a model of willingness to pay for environmental protection. Monte Carlo evidence comparing the three estimators is also presented.

Key words: ordered response models, semi-parametric estimation

1 Introduction

This paper looks at semi-parametric alternatives to the standard fully parametric models for ordered responses. The standard Ordered Probit and Ordered Logit models require specific distributional assumptions for the underlying latent variable. This paper considers various semi-parametric estimators that do not require a distributional assumption to be made. Lewbel (2000) has proposed a very simple estimator that has desirable properties under weak assumptions. However the estimator has been little used in practical applications and there have also been no comparisons of the estimator with other semi-parametric estimators (only with the standard fully parametric ones). The second estimator considered is constructed using the "semi-nonparametric" series estimator of Gallant and Nychka (1987). The third estimator examined uses a polynomial transformation to normality and is related to the family of distributions proposed by Ruud (1984).

Email address: `Mark.Stewart@warwick.ac.uk` (Mark B. Stewart).

Most of the semi-parametric estimators available for discrete response models require kernel smoothing, and hence bandwidth selection, together with iterative or grid search estimation and some sort of trimming to enable convergence. These kernel-based methods generally require a considerable amount of intervention and judgement on the part of the researcher and the estimators regularly experience acute convergence problems in the context of ordered response models. Therefore for the comparisons in this paper, the focus is on estimators that do not involve kernel smoothing. The comparison is also restricted to the single latent variable framework with single index form and fixed thresholds.

The next section presents a framework for cumulative probability models for ordered responses. Section 3 presents the Ordered Probit model in the particular parameterization used. Sections 4–6 then describe the three semi-parametric estimators given above. The Lewbel estimator is very fast, but the other two semi-parametric estimators are very computer intensive. Computational simplifications are described. Two illustrations of the techniques based on real data are presented: a model for job satisfaction in Section 7 and a model for willingness to pay for environmental protection in Section 8. Evidence from a Monte Carlo investigation of the three semi-parametric estimators is presented in Section 9. Finally conclusions are given in Section 10.

2 Cumulative probability models for ordered responses

Consider a linear model for an underlying latent variable

$$y_i^* = z_i + x_i'\beta + \varepsilon_i, \qquad\qquad i = 1, \ldots, N, \qquad\qquad (1)$$

where β is a vector of unknown parameters, $\varepsilon | z, x \sim F$, and where y^*, ε and β have all been scaled so that the coefficient on z is equal to 1. This is slightly different from the usual formulation, due to separating the variable z, which plays a special role in the Lewbel estimator, and the normalization on its coefficient. This type of normalization is fairly commonly used for semi-parametric estimators of limited dependent variable models. It is also useful in comparing the estimators, since the coefficient estimates are not directly comparable, but their ratios are. The observed dependent variable y_i is discrete, taking one of the values $\{1, 2, \ldots, J\}$, and is related to y_i^* by

$$y_i = \begin{cases} 1 \text{ if } y_i^* < \alpha_1 \\ 2 \text{ if } \alpha_1 \le y_i^* < \alpha_2 \\ \vdots \\ J \text{ if } \alpha_{J-1} \le y_i^*, \end{cases} \qquad\qquad (2)$$

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.

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

$$y_i = \sum_{j=1}^{J} j.\mathcal{I}(\alpha_{j-1} \le z_i + x_i'\beta + \varepsilon_i < \alpha_j), \tag{3}$$

where $\mathcal{I}(A) = 1$ if A is true and $= 0$ if A is false. The probabilities of the observed outcomes are given by

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

The scale of y^* is fixed by the unit coefficient on z, but not the origin. Identification can be achieved by assuming a zero intercept in β or by fixing one of the α_j. The former is used here.

If the distribution of ε is specified, then the model can be estimated consistently by Maximum Likelihood. The log-likelihood is given by

$$\ln \mathcal{L} = \sum_{i=1}^{N} \sum_{j=1}^{J} \left\{ \mathcal{I}(y_i = j) \ln \left[F(\alpha_j - z_i - x_i'\beta) - F(\alpha_{j-1} - z_i - x_i'\beta) \right] \right\}. \tag{5}$$

The above specification defines a class of cumulative probability models in which a transformation of the cumulative probabilities is taken to be a linear function of the explanatory variables, with only the intercept in this function differing across the categories:

$$F^{-1} \left\{ \Pr[y_i \le j] \right\} = \alpha_j - z_i - x_i'\beta. \tag{6}$$

3 The Ordered Probit model

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. In the current parameterization the Ordered Probit model assumes that $\varepsilon_i \sim N(0, \sigma^2)$. The cumulative probabilities are then given by

$$\Pr[y_i \le j] = \Phi\left(\frac{\alpha_j - z_i - x_i'\beta}{\sigma} \right), \tag{7}$$

or equivalently in the more familiar parameterization as

$$\Pr[y_i \le j] = \Phi(\alpha_j^* - \beta_0^* z_i - x_i' \beta^*), \qquad (8)$$

where $\alpha_j^* = \alpha_j/\sigma$, $\beta_0^* = 1/\sigma$, $\beta^* = \beta/\sigma$, and Φ is the cumulative distribution function of a standard Normal. The original coefficient vector is given by $\beta = \beta^*/\beta_0^*$.

4   Lewbel's Least Squares estimator

Lewbel (2000) has recently proposed a semiparametric estimator for qualitative response models, including the ordered response case, that is root–N consistent and asymptotically normal under weak conditions and is fairly straightforward to implement. The weak conditions allow, for example, heteroskedasticity of unknown form. Implementation does not require iteration or numerical searches. A simplified estimator, requiring only slightly stronger conditions, involves only two ordinary least squares linear regressions and one sort of the data.

The Lewbel estimator treats z as a 'special regressor' and requires that it satisfy certain conditions. Identification requires that z have a continuous distribution, that z have a sufficiently large support (the support of z needs to contain that of $-(x'\beta+\varepsilon)$), and that ε have a distribution that does not depend on z (although it is allowed to depend on x). Magnac and Maurin (2003) discuss alternative restrictions. Note that the Lewbel estimator is not unique in requiring a regressor with large support. This is also required, for example, by the Manski (1985) and Horowitz (1992) Maximum Score estimators for the binary response model.

Define cumulative indicator variables $y_{ji} = \mathcal{I}(y_i \ge j)$ for $j = 2, \ldots, J$. Then define

$$y_{\cdot i} = \frac{1}{J-1} \sum_{j=2}^{J} y_{ji} = \frac{y_i - 1}{J-1}$$

Thus $y_{\cdot i}$ is a scaled version of y_i with values that range from 0 to 1. The Lewbel estimator then requires construction of the variables

$$\widetilde{y}_{\cdot i} = \frac{y_{\cdot i} - \mathcal{I}(z > 0)}{f(z|x)},$$

$$\widetilde{y}_{ji} = \frac{y_{ji} - \mathcal{I}(z > 0)}{f(z|x)}, \qquad\qquad j = 2, \ldots, J.$$

Construction of these transformed variables requires an estimate of the conditional density of z given x. This is considered below. It is also desirable for

zero to be somewhere near the middle of the support for z, which can always be achieved by subtracting the mean or median before construction of these transformed variables.

The Lewbel estimator then involves a set of OLS linear regressions of the estimates of these transformed variables on x. Define

$$\widehat{\Delta} = \left[\frac{1}{N} \sum_{i=1}^{N} x_i x_i' \right]^{-1},$$

and let $\widehat{\Delta}_k$ denote the k-th row of $\widehat{\Delta}$. Then the Lewbel estimator of β is given by

$$\widehat{\beta}_k = \widehat{\Delta}_k \frac{1}{N} \sum_{i=1}^{N} x_i \widehat{\widetilde{y}}_{\cdot i}, \qquad\qquad k = 2, \ldots, K, \qquad\qquad (9)$$

where $\widehat{\widetilde{y}}_{\cdot i}$ is $\widetilde{y}_{\cdot i}$ with $f(z|x)$ replaced by a suitable estimate. So $\widehat{\beta}$ is the vector of slope coefficients from an OLS regression of $\widehat{\widetilde{y}}_{\cdot i}$ on x_i. Similarly the Lewbel estimator of the thresholds is given by

$$\widehat{\alpha}_{j-1} = -\widehat{\Delta}_1 \frac{1}{N} \sum_{i=1}^{N} x_i \widehat{\widetilde{y}}_{ji}, \qquad\qquad j = 2, \ldots, J. \qquad\qquad (10)$$

So each $\widehat{\alpha}_{j-1}$ is the estimated intercept from an OLS regression of $\widehat{\widetilde{y}}_{ji}$ on x_i.

The main effort in the estimator is the estimation of $f(z|x)$ prior to the construction of the transformed variables and in the estimation of suitable standard errors. Considering the latter first, the variance of $\widehat{\beta}$ can be estimated as the Huber/White sandwich estimator of the coefficient estimates from an OLS regression of $\widehat{\widetilde{y}}_{\cdot i} + E(\widehat{\widetilde{y}}_{\cdot i}|x_i) - E(\widehat{\widetilde{y}}_{\cdot i}|x_i, z_i)$ on x_i with the conditional expectations replaced by nonparametric regressions. However, in the applications presented below a bootstrap estimator will be used.

Turning to the estimation of $f(z|x)$, one choice would be a kernel estimate of the joint density of z and x divided by a kernel estimate of the density of x (see for example Lewbel (2000) equation B3). The need for a first stage kernel estimator can be avoided by using Lewbel's suggested simplified "ordered data" estimator of the reciprocal of the conditional density. (Bloch and Gastwirth (1968) and Lewbel and Schennach (2003) provide further discussion of the estimator.) This assumes that we can write

$$z_i = x_i' \gamma + \omega_i,$$

where ω_i is continuously distributed with bounded support, has mean zero, and is independent of x_i. Denote the residuals from the OLS estimation of this equation by $\widehat{\omega}$ and let the values of $\widehat{\omega}$ immediately preceding and following $\widehat{\omega}_i$ after sorting the data in ascending order of $\widehat{\omega}$ be ω_i^- and ω_i^+, respectively.

Then i/N is an estimate of the distribution of ω evaluated at $\widehat{\omega}_i$, and the reciprocal of the density can be estimated by

$$\frac{1}{f(z_i|x_i)} \approx \frac{1}{f_\omega(\widehat{\omega}_i)} \approx \left[\frac{dF_\omega(\widehat{\omega}_i)}{d\omega}\right]^{-1} \approx \frac{\omega_i^+ - \omega_i^-}{2/N}.$$

This can be used in the expressions for the transformed $\widetilde{y}$ variables given above. When combined with this density estimator, the Lewbel estimator is particularly straightforward to implement, although asymptotic efficiency is sacrificed. This simplicity in turn makes it computationally practical to construct bootstrap confidence intervals for the parameters.

The estimator is presented here for the ordered response model. The exposition in Lewbel (2000) also covers binary and unordered multiple responses, partially linear latent variable models and censored regression models. The estimator can also be used to provide consistent estimates in the presence of measurement error or endogenous regressors, providing suitable instruments are available. The approach is extended to the panel data context by Honoré and Lewbel (2002).

Despite the straightforward nature of the estimator, there have been relatively few empirical applications, and those that there have been have focused largely on binary response or endogenous selection models. An exception is Anton et al. (2001), which estimates discrete duration models (related to the ordered response models considered here). As far as I am aware there are no existing papers that compare the Lewbel estimator with other semi-parametric estimators, either in empirical studies or in simulation experiments. The few existing studies compare the estimator only with the corresponding fully parametric estimator.

5 Semi-nonparametric estimation

Another "semi-parametric" estimator of the ordered response model can be constructed using the "semi-nonparametric" series estimator of Gallant and Nychka (1987). The Gallant and Nychka "semi-nonparametric" estimator of an unknown density approximates it using a Hermite form and can be written as the product of a squared polynomial and a normal density, giving a polynomial expansion with Gaussian leading term. (In fact the expansion can be around any density with a moment-generating function, but the normal is a natural choice in the current context, since it results in a model that nests the Ordered Probit.)

To ensure that the density approximation is proper, the approximation is

specified as

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

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

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

This general specification of the density is invariant to multiplication of $\gamma = (\gamma_0, \gamma_1, ..., \gamma_K)$ by a scalar and a normalization is required, with $\gamma_0 = 1$ being a convenient choice. The required distribution function is therefore specified as

$$F_K(u) = \frac{\int_{-\infty}^{u} \left(1 + \sum_{k=1}^{K} \gamma_k \varepsilon^k \right)^2 \phi(\varepsilon) \, d\varepsilon}{\int_{-\infty}^{\infty} \left(1 + \sum_{k=1}^{K} \gamma_k \varepsilon^k \right)^2 \phi(\varepsilon) \, d\varepsilon}. \tag{13}$$

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

The unknown density can be approximated arbitrarily closely by this Hermite series, by increasing the choice of K, provided that it satisfies certain smoothness conditions. The class of densities that can be approximated by this form is very general, allowing any form of skewness, kurtosis, or higher moments, although "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 maximizing the pseudo-likelihood function that replaces the unknown distribution function F in (4) in the likelihood function by F_K in (13).

Semi-parametric identification also requires a location normalization. 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.

In the $K = 1$ model, the imposition of $E(\varepsilon) = 0$ implies $\gamma_1 = 0$, 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. 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 conditionally on K, possibly for a range of alternative values of K, with the final specification of the model chosen by tests among 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 approach can therefore be viewed as a series of polynomial densities and corresponding pseudo-likelihood functions, with 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 standard information criteria.

The polynomial form of (11) means that the cumulative probabilities given by (13), 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

$$f_K(\varepsilon) = \frac{1}{\theta} \sum_{k=0}^{2K} \gamma_k^* \varepsilon^k \phi(\varepsilon), \tag{14}$$

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

$$\theta = \int_{-\infty}^{\infty} \sum_{k=0}^{2K} \gamma_k^* \varepsilon^k \phi(\varepsilon) \, d\varepsilon = \sum_{k=0}^{2K} \gamma_k^* \mu_k = 1 + \sum_{k=1}^{K} \gamma_{2k}^* \mu_{2k}, \tag{15}$$

where μ_k is the k-th moment of a standard normal, and the second equality follows from the fact that $\mu_k = 0$ for k odd.

Equation (13) can be written as

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

a linear combination of the truncated moments of a standard normal, where

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

Computation of the likelihood contributions for the Maximum Likelihood iterations can then be simplified (and the computational burden eased) by making use of the recursion

$$I_k(u) = (k-1)I_{k-2}(u) - u^{k-1}\phi(u), \qquad (16)$$

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. Using the recursion in equation (16), the truncated moments in $F_K(u)$ 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 (13) required for the pseudo-likelihood function can be written as

$$
\begin{aligned}
F_K(u) &= \frac{1}{\theta} \sum_{k=0}^{2K} \gamma_k^* \{\mu_k \Phi(u) - A_k(u)\phi(u)\} \\
&= \Phi(u) - \frac{1}{\theta} \left[\sum_{k=0}^{2K} \gamma_k^* A_k(u) \right] \phi(u). \qquad (17)
\end{aligned}
$$

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 a polynomial in u of order $(2K-1)$. The SNP estimator is then given by the maximization of the pseudo-likelihood function got by replacing the unknown distribution function $F(u)$ in (5) by $F_K(u/\sigma)$ as defined in (17).

6   An alternative polynomial generalization

Another possible way to relax the normality assumption in the Ordered Probit model would be to use the larger parametric family of distributions proposed

9

by Ruud (1984). He nests the normal distribution within the parametric family

$$G(u) = \Phi\left(\frac{\Psi(u)}{\sigma}\right), \tag{18}$$

where $\Psi(u)$ is an M-th order power series

$$\Psi(u) = \sum_{m=0}^{M} \psi_m u^m, \tag{19}$$

that is monotonically increasing in u, i.e., $\Psi'(u) \geq 0$, for all real u, and normalized so that the median is zero ($\psi_0 = 0$) and $\psi_1 = 1$. In the current context the normalization $\sigma = 1$ is also needed. The aim is for $\Psi(.)$ to closely approximate the transformation of u to Normality. Melenberg and van Soest (1996) use this specification with $M = 3$:

$$G(u) = \Phi(u + \psi_2 u^2 + \psi_3 u^3). \tag{20}$$

In this case the monotonicity requirement is that $\psi_2^2 \leq 3\psi_3$. The model can be estimated by Maximum Likelihood with the unknown F replaced in the likelihood function by G as specified in (20) and with maximization subject to $\psi_2^2 \leq 3\psi_3$. A family of distributions is generated for different orders M, with monotonicity imposed on each distribution in the family.

However this approach to imposing monotonicity is rather inconvenient, particularly as M increases, and an alternative is adopted here, approximating the distribution by

$$F_M(u) = \Phi(h(u)), \tag{21}$$

with

$$h'(u) = \left(\sum_{k=0}^{M} \gamma_k u^k\right)^2. \tag{22}$$

This latter construct ensures monotonicity. Suitable normalizations are required. Setting the constant of integration to zero ensures a zero median for the distribution and $\gamma_0 = 1$ then gives scale identification.

To illustrate, for the case of $M = 1$, $h'(u) = (1 + \gamma_1)^2$ and $h(u) = u + \gamma_1 u^2 + \frac{1}{3}\gamma_1^2 u^3$. This corresponds to the specification used by Melenberg and van Soest (1996) referred to above with the inequality constraint replaced by an equality.

In the general case, the derivative of h can be written as

$$h'(u) = \sum_{k=0}^{2M} \gamma_k^* u^k, \tag{23}$$

where

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

10

with $a_k = \max(0, k - M)$ and $b_k = \min(k, M)$. Thus

$$h(u) = \sum_{k=0}^{2M} \frac{\gamma_k^*}{k+1} u^{k+1}. \tag{24}$$

(The requirement of a zero median implies a zero constant of integration.) The scale normalization used is $\gamma_0 = 1$. Thus the cumulative probabilities required for the pseudo-likelihood function are given by

$$F_M(u) = \Phi\left(u + \sum_{k=1}^{2M} \frac{\gamma_k^*}{k+1} u^{k+1}\right). \tag{25}$$

The polynomial-generalized ordered probit (PGOP) estimator is then given by maximization of the pseudo-likelihood function that replaces the unknown distribution function $F(u)$ in (5) by $F_M(u/\sigma)$ as defined in (25).

The semi-parametric estimators described in the previous three sections are illustrated on two different datasets in the next two sections.

7   First illustration: job satisfaction

This illustration uses a model of job satisfaction (self-declared) based on that specified in 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 well-known features of their results are the higher *ceteris paribus* job satisfaction levels of women and the U-shaped age profile.

The model specification adopted for this illustration uses fewer explanatory variables than included in the Clark-Oswald model. The log of the individual's earnings is taken as the variable z, whose coefficient is normalized to 1. The x-variables include the log of "comparison" earnings, the log of hours, and an indicator variable for males. A quadratic in age is also included as are dummy variables for those not in good health, those with a second job, those on temporary contracts, and those who are managers. The sample size is 3,895.

Results for this specification using the Ordered Probit estimator, the Lewbel least squares estimator (LLS), the SNP estimator and PGOP estimator are

given in Table 1. Bootstrap standard errors are presented for the Lewbel estimator. Means of the explanatory variables are given in the first column of the table. For the SNP and PGOP estimators the choice of the values of K and M respectively is treated as a model selection problem and standard selection procedures applied. The values of the log-likelihood and corresponding Likelihood Ratio test statistics for different values of K and M are given in Table 2.

These indicate a choice of $K = 3$ for the SNP estimator. The standard ordered probit model is rejected against the $K = 3$ model (and higher orders) on the basis of likelihood-ratio tests. However the $K = 3$ model is not rejected against $K = 4$ (or higher orders). Selection of K on the basis of the Akaike Information Criterion or other standard model selection criteria also indicates choosing $K = 3$. SNP estimates for $K = 3$ are given in the third column of Table 1.

A similar approach is taken to the selection of the order of the polynomial for the PGOP estimator. The likelihood-ratio tests and other model selection criteria indicate a choice of $M = 1$, meaning that the transformation $h(u)$ is a cubic in u. PGOP estimates for $M = 1$ are given in the fourth column of Table 1.

Estimated coefficients in the standard parameterization of the Ordered Probit model cannot be interpreted directly and are only identified up to a scale normalization (as can be seen in Section 3 above). However, ratios of coefficients can be usefully interpreted. There are two convenient ways of viewing these. Assume that there is an underlying, but unobservable, continuous measure of job satisfaction (or utility or welfare depending on one's perspective) that is a monotonic, but unknown, function of the variable y_i^* in equation (1). Effects on this underlying variable are of interest. Consider plotting "indifference curves" between two factors that influence job satisfaction, i.e. combinations of the two factors that give the same level of job satisfaction. The specification in (1) means that these will be straight lines. The slopes of these (parallel) indifference curves between the two factors will be minus the ratio of their coefficients in equation (1). Since the equation is normalized on z, an estimated β-coefficient can be interpreted in terms of the slope of indifference curves between z and the corresponding x-variable.

Ratios of coefficients are also informative about the impacts of explanatory 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, the partial derivative of $\Pr[y_i = J]$ with respect to x is $\beta.f(\alpha_{J-1} - z_i - x_i'\beta)$. The scaling factor is the same for each element of β and also for the effect

Table 1
Job Satisfaction Model - Alternative Estimators

	Mean	OP	LLS	SNP(3)	PGOP(1)
log(earnings)	6.66	1	1	1	1
log('comparison')	6.66	-2.73 (0.98)	-3.24 (1.49)	-2.86 (0.40)	-4.52 (2.97)
log(hours)	4.95	-1.66 (0.78)	0.20 (3.38)	-1.22 (0.48)	-2.07 (1.55)
Male	0.50	-1.56 (0.79)	0.56 (0.90)	-1.25 (0.37)	-2.36 (1.88)
Age/10	3.72	-1.81 (1.12)	-0.51 (2.28)	-1.35 (0.75)	-2.34 (2.12)
$Age^2/100$	15.19	0.33 (0.18)	0.03 (0.28)	0.26 (0.10)	0.47 (0.38)
Health	0.18	-2.28 (1.11)	-1.27 (0.51)	-1.85 (0.40)	-3.55 (2.79)
Second job	0.10	-0.91 (0.66)	1.33 (1.53)	-0.85 (0.44)	-1.61 (1.52)
Temporary	0.06	-1.44 (0.90)	0.31 (1.00)	-1.16 (0.58)	-1.91 (1.74)
Manager	0.38	1.68 (0.86)	1.63 (1.03)	1.29 (0.33)	2.46 (1.98)
Log-likelihood		-6210.14		-6204.58	-6204.36
L-R test of OP				11.13	11.56
[p-value]				[.001]	[.001]

Notes: (1) Standard errors in parentheses.
(2) Estimators: OP = Ordered Probit estimator, LLS = Lewbel least squares estimator, SNP = Semi-nonparametric estimator, PGOP = Polynomial-generalized Ordered Probit estimator.

of z. Hence an estimated β-coefficient can also be interpreted as the ratio of the "marginal effect" of the corresponding x-variable to the "marginal effect" of z.

The main focus in the Clark and Oswald study is on the effects of own and "comparison" earnings and the relative magnitudes of their coefficients. If their effects are equal in absolute value and of opposite sign, then only *relative* earnings influences job satisfaction. In the current context with the equation normalized on own earnings, this corresponds to the hypothesis that the β-coefficient on "comparison" earnings is equal to -1. Examining this estimated coefficient, the ordered probit model indicates that a 10% increase in comparison earnings would need to be compensated by a 27% increase in own earnings to give the same level of job satisfaction. This is similar in the SNP model but rises to 32% in the Lewbel estimates and to 45% in polynomial-generalized model. However the standard error rises greatly in the last case and the effect is not significantly different from zero. The hypothesis that the coefficient is equal to -1 is rejected with the SNP estimator, but accepted with the other three estimators.

13

Table 2
Job Satisfaction Model: Likelihood-Ratio Test Statistics,
SNP Estimator by K, PGOP Estimator by M

Model	log-likelihood	LR-test of OP model	p-value	LR-test of previous	p-value
OP	-6210.14				
SNP estimator:					
$K = 3$	-6204.58	11.13	0.001	11.13	0.001
4	-6204.45	11.40	0.003	0.26	0.607
5	-6203.19	13.91	0.003	2.51	0.113
6	-6201.72	16.84	0.002	2.93	0.087
7	-6201.44	17.41	0.004	0.56	0.453
PGOP estimator:					
$M = 1$	-6204.36	11.56	0.001	11.56	0.001
2	-6203.51	13.27	0.001	1.71	0.191
3	-6203.39	13.50	0.004	0.23	0.633
4	-6202.11	16.07	0.003	2.57	0.109
5	-6201.44	17.41	0.004	1.34	0.246

Notes:
(1) LR tests of OP model have $K - 2$ degrees of freedom in SNP models and M in PGOP models.
(2) LR tests of previous model all have 1 degree of freedom.
(3) Test of SNP(7) against SNP(3): $\chi^2(4)$ of 6.28, p = 0.179.
(4) Test of PGOP(5) against PGOP(1): $\chi^2(4)$ of 5.84, p = 0.211.

Comparing the Ordered Probit and Lewbel estimates, they show some marked differences. While the coefficient on "comparison" earnings strengthens with the Lewbel estimator, the coefficients on the other variables weaken. The gender and age effects, for example, collapse. These often-quoted features of the Clark-Oswald results are therefore not robust to the choice of estimator. The SNP estimates are much more similar to the Ordered Probit ones, but the t-ratios all strengthen. In contrast the PGOP estimates all increase in absolute value, but so do the standard errors and the t-ratios of all coefficients decline.

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 at age 28 on the basis of the ordered probit model, but slightly earlier at 25 with both the SNP and PGOP models. However, standard errors of the ratios (calculated by the "delta"

method) indicate that they are not precisely determined and these differences are small relative to the standard errors. The Lewbel estimator in contrast indicates a minimum considerably beyond the age of retirement, i.e., that job satisfaction declines throughout working life. However, again the standard error of the ratio is quite large.

Overall there are important differences among the results from the different estimators in this illustration. Hausman tests can be conducted to compare the estimated coefficient vectors from the different estimators. Test statistics comparing the various estimators with the Ordered Probit (all $\chi^2(9)$) are: Lewbel LS: 30.23 (p-value = .0004), SNP: 28.50 (p-value = .001), PGOP: 1.38 (p-value = .998). The last of these parallels the greatly increased standard errors for the PGOP estimator with this data. The null is strongly rejected for both the Lewbel least squares estimator and the SNP estimator.

8 Second illustration: Willingness to pay for environmental protection

This section illustrates the same semi-parametric estimators with a model of the willingness to pay for environmental protection in Germany on the data used by Witzke and Urfei (2001) and based on a cut-down version of the specification used by them. The data are taken from a survey on Environmental Consciousness and Behaviour conducted by the Federal Environmental Agency.

Respondents were asked "Would you be willing to pay higher taxes or mandatory fees in support of a better environmental protection, if it were guaranteed that your payments would directly benefit environmental protection?" The possible answers to the question were "no", "it depends", and "yes". The relative sample frequencies are 51%, 31%, and 18% respectively. This is then modelled as an ordered response. The main focus of the Witzke and Urfei paper is the regional dimension, but that aspect is ignored here.

For this illustration, the log of income is taken as the variable z whose coefficient is normalized to 1. The x-variables included are age, a dummy variable for having completed at least high school, a dummy for female respondents and two dummies indicating preferences for political parties. The first of these is for the German Green party and the second indicates a preference for either the social democrats (SPD) or the socialists (PDS). The sample size is 1,643.

Results for this specification using the Ordered Probit, Lewbel least squares, SNP and PGOP estimators are given in Table 3. Standard errors are given in parentheses. Bootstrap standard errors are presented for the Lewbel esti-

15

Table 3
Willingness to Pay Model - Alternative Estimators

	Means	OP	LLS	SNP(3)	PGOP(1)
log(income)	7.97	1	1	1	1
Age/10	4.80	-0.19 (0.08)	-0.25 (0.10)	-0.19 (0.08)	-0.19 (0.08)
High school	0.12	1.36 (0.44)	0.89 (1.47)	1.39 (0.44)	1.38 (0.44)
Green party	0.10	2.16 (0.56)	1.30 (0.92)	2.15 (0.55)	2.14 (0.55)
SPD / PDS	0.29	0.49 (0.25)	0.88 (0.51)	0.47 (0.24)	0.47 (0.24)
Female	0.54	-0.07 (0.21)	-0.24 (0.49)	-0.10 (0.21)	-0.10 (0.21)
Log-likelihood		-1600.32		-1599.20	-1599.19
L-R test of OP				2.25	2.26
[p-value]				[.134]	[.132]

Note: Standard errors in parentheses.

mator. The same model selection procedures for the choice of K for the SNP estimator and M for the PGOP estimator are used here as in the previous section.

In contrast with the first illustration, the Ordered Probit model is not rejected here against any member of the SNP family for $K \geq 3$ or against any member of the PGOP family for $M \geq 1$. Log-likelihood values and Likelihood Ratio test statistics are given in Table 4. To illustrate the estimators, results are presented in Table 3 for the SNP estimator with $K = 3$ and the PGOP estimator with $M = 1$, alongside the Ordered Probit and Lewbel estimators.

If it is assumed that there is a continuous latent measure of willingness to pay for environmental protection that is a monotonic, although unknown, function of the variable y_i^* in equation (1), an estimated β-coefficient can be interpreted in terms of the slope of the indifference curves between the corresponding x-variable and the log of income, just as in the previous section. The estimates of the SNP and PGOP estimators are almost identical to those for the Ordered Probit, as one would expect, since the likelihood differences are small and the Ordered Probit model is not rejected against either SNP(3) or PGOP(1). There is more difference between the Ordered Probit and Lewbel estimates, but the differences are much smaller than for the illustration in the previous section. The main differences are that the strong positive effects of having completed high school and of being a Green party voter are cut in the Lewbel estimates while their standard errors rise, rendering both effects insignificant.

Overall the differences between the estimators are much less marked than was the case for the illustration presented in the previous section. Once again

Table 4
Willingness to Pay Model: Likelihood-Ratio Test Statistics,
SNP Estimator by K, PGOP Estimator by M

Model	log-likelihood	LR-test of OP model	p-value	LR-test of previous	p-value
OP	-1600.32				
SNP estimator:					
$K = 3$	-1599.20	2.25	0.134	2.25	0.134
4	-1599.16	2.32	0.313	0.08	0.781
5	-1598.96	2.73	0.436	0.40	0.526
6	-1598.63	3.37	0.497	0.65	0.421
7	-1598.63	3.38	0.641	0.01	0.918
PGOP estimator:					
$M = 1$	-1599.19	2.26	0.132	2.26	0.132
2	-1599.16	2.33	0.312	0.06	0.802
3	-1599.15	2.34	0.505	0.01	0.913
4	-1598.87	2.90	0.575	0.56	0.454
5	-1598.69	3.27	0.659	0.37	0.545

Notes:
(1) LR tests of OP model have $K - 2$ degrees of freedom in SNP models and M in PGOP models.
(2) LR tests of previous model all have 1 degree of freedom.
(3) Test of SNP(7) against SNP(3): $\chi^2(4)$ of 0.57, p = 0.966.
(4) Test of PGOP(5) against PGOP(1): $\chi^2(4)$ of 0.50, p = 0.973.

Hausman tests can be conducted to compare the estimated coefficient vectors from the different estimators. Test statistics comparing the various estimators with the Ordered Probit (all $\chi^2(9)$) are: Lewbel LS: 3.52 (p-value = .620), SNP: 0.70 (p-value = .983), PGOP: 0.64 (p-value = .986). The null hypothesis is not rejected for any of the three estimators.

9 A Monte Carlo investigation

This section presents the evidence from a small scale Monte Carlo investigation of the finite sample performance of the three semi-parametric estimators described in sections 4–6 (together with the fully parametric Ordered Probit

17

Table 5
Monte Carlo Experiments

No.	Experiment
1	$a = 2, \varepsilon \sim N(0,1)$
2	$a = 4, \varepsilon \sim N(0,1)$
3	$a = 2, \varepsilon \sim N(0,\sigma^2), \sigma = 2$
4	$a = 2, \varepsilon \sim N(0,\sigma^2), \sigma = \sqrt{3/31}(3 + x)$
5	$a = 2, \varepsilon \sim N(0,\sigma^2), \sigma = \sqrt{3/82}(5 + z + x)$
6	$a = 2, \varepsilon \sim (\chi_1^2 - 1)/\sqrt{2}$
7	$a = 2, \varepsilon \sim$ Normal mixture
8	$a = 2, \varepsilon \sim$ Normal mixture, $\sigma = \sqrt{3/31}(3 + x)$
9	$a = 2, \varepsilon \sim$ Normal mixture, $\sigma = \sqrt{3/82}(5 + z + x)$

estimator). The structure of the experiments is kept as simple as possible. The latent variable is generated as

$$y_i^* = 5 + z_i + \beta x_i + \varepsilon_i,$$

with $\beta = 1$. The estimate of β is the focus of attention in the investigation. The variable z is generated as $N(0,1)$ and the variable x as a uniform random variable over $(-a, a)$, both fixed across replications. In the baseline experiment, $a = 2$. In all experiments ε is independently distributed with mean zero. In the baseline experiment ε is generated as $N(0,1)$.

The ordered response variable y is constructed to have three categories, with thresholds for y_i^* at 4 and 6, i.e., $\alpha_1 = -1, \alpha_2 = 1$. In the baseline experiment this gives approximately 30% in the top and bottom categories and 40% in the middle category. A sample size of 1200 is used as representative of applications of ordered response models found in the literature. 500 replications are performed in each experiment, which seems to be sufficient for the stability of the tabulated statistics. The structure of the experiments conducted is shown in Table 5.

The balance of the variances in Experiment 1 has been selected as a compromise between the conditions required for the Lewbel estimator and features of the data common in empirical applications. Generating z as an $N(0,1)$ random variable gives it an infinite theoretical support. (In the generated data used the support of z is ± 3.6). However, the relative magnitudes of the variances of z and $(\beta x + \varepsilon)$ are also likely to be important for the Lewbel estimator (see Lewbel, 2000). In Experiment 1, $\sigma_z = \sigma_\varepsilon = 1$ and $\sigma_x = 1.15$. This means that the standard deviation of z is roughly 65% of that of $(\beta x + \varepsilon)$.

Table 6
Monte Carlo Results for Ordered Response Model Estimators

Exp.	OP		LLS		SNP(3)		PGOP(1)	
no.	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	.0040	.0486	.0017	.0715	.0041	.0492	.0041	.0489
2	.0052	.0476	-.1421	.1464	.0059	.0481	.0052	.0478
3	.0054	.0949	-.1533	.1866	.0065	.0960	.0057	.0955
4	-.0030	.0477	-.0144	.0688	.0766	.0941	.0673	.0849
5	.0173	.0515	-.0147	.0743	.0120	.0483	.0059	.0454
6	-.0062	.0431	-.0227	.0744	-.0570	.0651	.0059	.0314
7	-.0175	.0458	.0127	.0700	-.0639	.0702	-.0006	.0319
8	-.0614	.0730	-.0255	.0793	-.1608	.1633	-.1698	.1729
9	-.0219	.0432	.0342	.0713	-.0512	.0593	-.0019	.0345

However in this specification, 70% of the variation in y^* is explained by the explanatory variables and 43% of the explanatory power in the model comes from the z variable. Both of these seem somewhat unrealistic in terms of the applications of ordered response models found in the literature and the fit of micro-econometric models in general. Aiming for desirable properties for the Lewbel estimator suggests raising σ_z relative to σ_x and σ_ε, but this would make both of the disadvantages just mentioned worse. Thus Experiment 1 can be regarded as a compromise.

Experiments 2 and 3 address these two features of the generated data by raising σ_x and σ_ε respectively. Experiment 2 increases a to 4 and thereby shifts the balance between the variation in z and x to something more similar to that found in empirical studies. In Experiment 3, σ_ε is doubled and as a result the explanatory power of the model for y^* is lowered to 37%, a more realistic level in micro-level data.

The results from the experiments are given in Tables 6 and 7. The tables present the mean bias and root mean square error (RMSE) for each of the estimators of β. Table 6 presents these for the Ordered Probit estimator, the Lewbel estimator, the SNP estimator with $K = 3$, and the PGOP estimator with $M = 1$. The corresponding results for SNP(5) and PGOP(3) are given in Table 7.

Starting with the results for Experiment 1, all six estimators show only a small mean bias. The RMSE for the Lewbel estimator is about 50% higher than that for the other estimators. In Experiments 2 and 3, however, there

Table 7
Monte Carlo Results for Ordered Response Model Estimators

Exp.	SNP(5)		PGOP(3)	
no.	Bias	RMSE	Bias	RMSE
1	.0037	.0500	.0034	.0498
2	.0051	.0487	.0055	.0486
3	.0057	.0960	.0062	.0959
4	.0663	.0853	.0667	.0860
5	.0063	.0462	.0045	.0468
6	.0051	.0349	.0060	.0261
7	-.0008	.0319	-.0008	.0309
8	-.1464	.1493	-.1668	.1698
9	-.0031	.0325	-.0023	.0337

is a clear bias in the Lewbel estimator. Reducing the relative variance of the special variable z (relative to that of either x or ε) clearly damages the finite sample performance of the Lewbel estimator.

For a given relative variance of z, the support also matters. In another experiment (not reported in the table), z was generated as a $N(0,1)$ truncated at ± 2 and then scaled to give the same variance as in Experiment 1. (In the generated data this produced a variable with support ± 2.35.) The results for the Ordered Probit estimator were very similar to those for Experiment 1 (mean bias $= .0032$, RMSE $= .0477$). The performance of the Lewbel estimator, however, deteriorates appreciably, giving a mean bias of -.1130 and a RMSE of .1190.

The next two experiments introduce heteroskedasticity. In the case of Experiment 4 this is with respect to x only. In Experiment 5 it is with respect to both x and z and in such a way as to preserve the single index nature of the model. In Experiment 4 a slight bias is found in the Lewbel estimator (although it is still consistent in these circumstances), and a larger bias in the SNP and PGOP estimators (which are not). Experiment 5 produces a similar slight bias in the Ordered Probit estimator, but that for the Lewbel estimator (no longer consistent because of the heteroskedasticity with respect to z) is identical to that in Experiment 4 (where it is consistent). However, the mean bias for the SNP and PGOP estimators falls considerably (presumably because of the restoration of the single index feature of the model).

The next two experiments introduce a non-Normal distribution for ε. Experi-

ment 6 uses a $\chi^2(1)$ distribution, standardized to zero mean and unit variance, which is positively skewed. Experiment 7 generates ε as a mixture of two normals. The first has $\mu_1 = -0.5, \sigma_1 = 0.25$ and a probability of 0.75. The second has $\mu_2 = 1.5, \sigma_2 = 0.9$ and a probability of 0.25. The mixture has a mean of zero and a variance of 1, but is positively skewed. In Experiment 6 the Ordered Probit estimator shows a surprisingly small bias, given that it is based on the wrong distribution. The mean bias is greater in the Lewbel estimator, despite it not making any distributional assumption. The PGOP estimator performs well for $M = 1$ and $M = 3$, as does the SNP estimator for $K = 5$. However the SNP estimator performs relatively poorly for $K = 3$. The single additional parameter is typically not sufficient to capture the non-normality, and, rather surprisingly, it performs worse than the simple Ordered Probit estimator. In Experiment 7 the Lewbel estimator has a smaller mean bias than the Ordered Probit estimator, but a larger RMSE. The PGOP estimator again performs better than the SNP estimator when only a single extra parameter is involved ($K = 3$ or $M = 1$). However both PGOP and SNP outperform the Ordered Probit and Lewbel estimators when the specifications with three extra parameters ($K = 5$ or $M = 3$) are used.

Since the Ordered Probit model is likely to be less robust to the combination of non-normality and heteroskedasticity, the final two experiments combine the normal mixture distribution of Experiment 7 with the heteroskedasticity in Experiments 4 and 5. In Experiment 8, where the model is no longer in single index form, the bias in the SNP(3) and PGOP(1) estimators is considerable and is diminished only slightly when the higher order polynomials are used. The bias in the Lewbel estimator is smaller than that in the Ordered Probit. When σ depends also on z in Experiment 9, the bias in the Lewbel estimator worsens, but since the single index form of the model is restored, the performance of the SNP and PGOP estimators improves considerably.

One must be careful not to infer too much on the basis of this rather limited set of Monte Carlo experiments, but a number of tentative conclusions seem in order. First, in circumstances where the single-index form is valid, the higher order SNP and PGOP estimators perform very well. Second, the SNP(5) and PGOP(3) estimators perform very similarly (in terms of mean bias and RMSE), but PGOP(1) dominates SNP(3) when ε is non-normal but the single-index form is still valid. Third, the SNP and PGOP estimators perform poorly in the presence of heteroskedasticity of a form that removes the single-index nature of the model, particularly when this is combined with non-normality. In these experiments both estimators fare worse than the Ordered Probit and Lewbel estimators. Fourth, the Lewbel estimator comes into its own when there is heteroskedasticity with respect to x, but not z, combined with non-normality. Fifth, the Lewbel estimator fares poorly when the relative variance of z is reduced. It also does not necessarily outperform the Ordered Probit estimator in circumstances when it might be expected to (when it is consistent

and the Ordered Probit estimator is not). In all experiments its RMSE exceeds that of the Ordered Probit estimator.

10 Conclusions

This paper compares three semi-parametric estimators for ordered response models (as well as the fully parametric Ordered Probit estimator). The results differ markedly across estimators in the job satisfaction illustration presented in Section 7, but differ much less in the willingness to pay illustration presented in Section 8. The evidence from the Monte Carlo investigation in Section 9 suggests that the PGOP and SNP estimators outperform the Lewbel estimator in RMSE terms, except when there is heteroskedasticity of a form that removes the single index nature of the model.

References

Anton, A. Alonso, A. Fernandez Sainz and J. Rodriguez-Poo (2001), "Semi-parametric estimation of a duration model", Oxford Bulletin of Economics and Statistics, 63, 517–533.

Bloch, D.A. and J.L. Gastwirth (1968), "On a simple estimate of the reciprocal of the density function", Annals of Mathematical Statistics, 39, 1083–1085.

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.

Honoré, Bo E. and Arthur Lewbel (2002), "Semiparametric binary choice panel data models without strictly exogenous regressors", Econometrica,

70, 2053–2063.

Horowitz, Joel L. (1992), "A smoothed maximum score estimator for the binary response model", Econometrica, 60, 505–531.

Lewbel, Arthur (2000), "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables", Journal of Econometrics, 97, 145–177.

Lewbel, Arthur and Susanne M. Schennach (2003), "A simple ordered data estimator for inverse density weighted expectations", mimeo, Boston College, February.

Magnac, Thierry and Eric Maurin (2003), "Identification and information in monotone binary models", mimeo, INRA, Paris-Jourdan, June.

Manski, Charles F. (1985), "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator", Journal of Econometrics, 27, 313–333.

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.

Ruud, Paul A. (1984), "Tests of Specification in Econometrics", Econometric Reviews, 3, 211–242.

Witzke, H. Peter and Guido Urfei (2001), "Willingness to pay for environmental protection in Germany: Coping with the regional dimension", Regional Studies, 35, 207–214.