Ordered Response Threshold Estimation

Arthur Lewbel
Boston College

June 2002, revised October 2003

Abstract

This paper shows that many estimators of thresholds in ordered response models exist, because binary choice location estimators can be converted into threshold estimators. A new threshold estimator is proposed that is consistent under more general conditions. An extension to random thresholds is provided.

JEL codes: C14, C25, C13. Keywords: Ordered response, ordered choice, semiparametric estimation, threshold points. This research was supported in part by the National Science Foundation through grant SES-9905010.
Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (617)-552-3678, http://www2.bc.edu/~lewbel/ lewbel@bc.edu.

1 Introduction

This paper first summarizes existing estimators of thresholds in ordered response models. More such estimators exist than has been previously recognized, because binary choice location estimators can be converted into threshold estimators. A new threshold estimator is then proposed that is consistent under quite general conditions, including an extension to random thresholds.

Given observations of (Y, X), the standard ordered response model is

$$Y = \Sigma_{k=0}^{K+1} k\, I(\alpha_{k-1} \leq Y^* < \alpha_k) \tag{1}$$

where the latent Y^* is given by

$$Y^* = X'\beta + U. \tag{2}$$

When the error U is logistic or normal, this is the ordered logit or ordered probit model. Assume now that the distribution of the unobserved U is unknown.

The constants $\alpha = (\alpha_0, ..., \alpha_K)$ are threshold points ($\alpha_{-1} = -\infty$, $\alpha_{K+1} = \infty$). Without loss of generality, assume that X does not contain a constant term, so any location constant in the regression function is absorbed into α, and that an arbitrary scaling has been chosen, such as that the first element of β equals one. If location assumptions are not imposed on U (e.g., if U is not defined to have mean or median zero) then only relative thresholds would be identifiable. These relative thresholds may be defined by $t = (t_1, ..., t_K)$ where $t_k = \alpha_k - \alpha_0$.

The coefficients β can be estimated in a variety of ways, for example, semiparametric linear index model estimators could be used. This paper focuses on estimation of thresholds. Estimation of thresholds is of interest in part because it generally increases the efficiency of estimated choice probabilities, relative to estimators that ignore the information contained in the threshold structure. Also, in many economic applications the thresholds have direct economic significance. For example, when Y is the number of units of a good that is consumed and Z (or some other regressor with a unit coefficent) is price, then relative thresholds correspond to the changes in price (or the change in the determinants of reservation price), that are required to increase or decrease the quantity demanded.

A limitation of the standard ordered choice model is that the thresholds are constrained to be the same for all individuals. The latent Y^* is often interpreted as a measure of utility, so it would be natural to allow thresholds, which are preference related parameters, to vary across individuals. Allowing thresholds to depend deterministically on regressors is in principle straightforward, though normalizations or nonlinearities would be required to separately identify the contributions of regressors to thresholds versus contributions to the latent Y^*. Less obvious is allowing for randomness or unobservables to affect thresholds. An example of an ordered choice model with random thresholds is Cameron and Heckman (1998), where Y corresponds to years of schooling (their model also departs from standard ordered choice by incorporating dynamic, sequential choice). Random thresholds are also identified (by means of one sided shocks) and used in Heckman, LaLonde and Smith (1999) and Carneiro, Hansen, and Heckman (2003) The new threshold estimator that is proposed here extends to the case of random thresholds.

2 Existing Semiparametric Threshold Estimators

Klein and Sherman (2002) propose an estimator for thresholds assuming U is independent of X. They say "to our knowledge, there exists no semiparametric estimator of the threshold points." However, one threshold estimator that predates Klein and Sherman (2002) is Lewbel (2000), equations (6.1) to (6.3), which provides a root n consistent estimator for both the coefficients β and the thresholds α, including extensions to the case where U suffers from general forms of heteroscedasticity, or where some of the regressors are endogeneous.

Other estimators of thresholds may be obtained by writing the ordered response model as the set of binary responses

$$Y_k = I(Y > k) = I(0 \leq X'\beta + U_k) \tag{3}$$

$$U_k = -\alpha_k + U$$

for $k = 0, \ldots, K$. It immediately follows that any semiparametric estimator of the location parameter (e.g., the mean or median of U_k) in binary response models can be applied to each of the binary choices Y_k to recover the thresholds α. Many such estimators exist, including Manski (1985), Horowitz (1992), Lewbel (1997), McFadden (1999), Chen (1999), and Chen (2000). Equation (3) also shows that β can be estimated by applying binary choice coefficient estimators to each Y_k separately, then take a weighted average of the results across $k = 0, \ldots, K$. Weights could be chosen using minimum chi squared estimation.

Estimation of thresholds by applying binary choice location estimators to Y_k requires U to have mean or median zero. An advantage of the Klein and Sherman (2002) estimator is that it does not impose a mean, median, or other location restriction on U, but as a result only identifies and estimates relative thresholds $t = (t_1, ..., t_K)$. Coppejans (2003) establishes the semiparametric efficiency bound for β, t assuming U independent of X, and proposes an estimator that attains this bound.

3 A New Threshold Estimator

Consider estimation of relative thresholds t in the ordered response model (1) with equation (2) is replaced by

$$Y^* = Z + U^* \tag{4}$$

where the data are observations of Y, Z, W and

$$U^* \mid Z, W = U^* \mid W \tag{5}$$

or equivalently, U^* and Z are conditionally independent, conditioning on a vector of random variables W. Equation (2) with $X'\beta \perp U$ is a special case of this assumption with $U^* = U$, $Z = X'\beta$, and W empty. Alternatively, let Z be the first element of X and let W be the other elements of X, so $X'\beta = Z + W'\gamma$ (where the arbitrary scaling permits taking the coefficient of Z to be one). Then equation (2) is the special case of equation (4) in which $U^* = W'\gamma + U$.

Equations (4) and (5) also allow for much more general models. For example, we could have $U^* = g(W) + U$ with $U \perp Z, W$ for an arbitary function W, yielding the partly linear latent model $Y^* = Z + g(W) + U$. Another example of equations (4) and (5) is $U^* = W'(\gamma + V) + U$ where U and V are both unobserved and independent of Z, W, corresponding to a latent random coefficients model. No location restriction on U^*, such as mean or median zero, is required.

Given the general ordered response model of equations (1), (4), and (5), features of the distribution of the unobserved U^* (such as the function $g(W)$ in the above example) could be estimated by, e.g., applying Lewbel, Linton, and McFadden (2001) to $Y_k = I(Y > k)$, or by a variety of other binary choice model estimators with varying assumptions regarding U^*. To estimate the thresholds, define $h_k(z, w) = E\left(Y_0 - Y_k \mid Z = z, W = w\right)$ and make the following assumption:

ASSUMPTION 1. Equations (1), (4), and (5) hold, with Z continuously distributed given W. There exists an interval Ω such that $\Omega \subseteq supp(Z \mid W)$, $supp(\alpha_K - U^* \mid W) \subseteq \Omega$, and $supp(\alpha_0 - U^* \mid W) \subseteq \Omega$. The relative thresholds t_k and conditional expectations $h_k(Z, W)$ exist and are finite.

Many discrete choice estimators assume existence of a regressor Z that satisfies this support assumption, including, e.g., Manski (1985), Horowitz (1992), and Lewbel (2000). It may be possible to replace this support assumption with error tail symmetry, as Magnac and Maurin (2003) do in a related semiparametric discrete choice model.

The proposed estimator is based on the following, hitherto unknown property regarding thresholds. If Assumption 1 holds then

$$t_k = E\left[\int_\Omega h_k(z, W)dz\right], \quad k = 1, \ldots, K. \tag{6}$$

Letting F_{U^*} denote the conditional distribution function of U^*, the derivation of this equation is

$$\begin{aligned}
& E\left[\int_\Omega E\left(Y_0 - Y_k \mid Z = z, W\right) dz\right] \\
= \; & E\left[\int_\Omega \int_{\text{supp}(U^*|W)} [I(0 \le Z + U^* - \alpha_0) - I(0 \le Z + U^* - \alpha_k)] dF_{U^*}(U^* \mid Z = z, W) dz\right] \\
= \; & E\left[\int_{\text{supp}(U^*|W)} \int_\Omega [I(0 \le Z + U^* - \alpha_0) - I(0 \le Z + U^* - \alpha_k)] dz dF_{U^*}(U^* \mid W) dz\right] \\
= \; & E\left[\int_{\text{supp}(U^*|W)} \int_{-U^*+\alpha_0}^{-U^*+\alpha_k} dz dF_{U^*}(U^* \mid W) dz\right] \\
= \; & E\left[\int_{\text{supp}(U^*|W)} (\alpha_k - \alpha_0) dF_{U^*}(U^* \mid W) dz\right] = t_k
\end{aligned}$$

where the second equality uses the assumed conditional independence $F_{U^*}(U^* \mid Z, W) = F_{U^*}(U^* \mid W)$ from equation (5).

It follows from equation (6) that, with independent, identically distributed observations y_i, z_i, w_i, each threshold t_k can be consistently estimated by

$$\widehat{t_k} = \frac{1}{n} \sum_{i=1}^{n} \int_\Omega \widehat{h}_k(z, w_i) dz \tag{7}$$

where $\widehat{h}_k(z, w)$ is a consistent nonparametric regression of $y_0 - y_k$ on z, w.

4 Random Thresholds

Now consider random thresholds, so equation (1) is replaced with

$$Y = \Sigma_{k=0}^{K+1} k \, I(\alpha_{k-1} + e_{k-1} \le Y^* < \alpha_k + e_k) \tag{8}$$

for some vector $e = (e_0, ..., e_K)$ of unobserved random variables. Since $\alpha_{-1} = -\infty$ and $\alpha_{K+1} = \infty$, it may be assumed without loss of generality that $e_{-1} = e_{K+1} = 0$. Relative thresholds t_k are still defined by $t_k = \alpha_k - \alpha_0$, and so now may be interpreted as relative mean thresholds.

If equations (8) and (2) hold, then equation (3) will still hold, but now with $U_k = -\alpha_k - e_k + U$. It therefore follows that, with random thresholds, the mean or median thresholds α could still be estimated using binary choice location estimators as described earlier, assuming the corresponding location (mean or median) of e is zero.

Now consider the more general ordered response model where equation (2) is not assumed. Instead, we have only equations (8), (4), and

$$U^*, e \mid Z, W = U^*, e \mid W. \tag{9}$$

Here with random thresholds, equation (5) has been replaced with equation (9). This permits the distribution of e to correlate with or otherwise depend on U^*, and it permits the distribution of e to depend upon regressors W. Formally, we are now assuming the following.

ASSUMPTION 2. Equations (4), (8), and (9) hold, with Z continuously distributed given W. The support of e is such that $\alpha_{k-1} + e_{k-1} < \alpha_k + e_k$ for $k = 0, ..., K+1$. $E(e_k - e_0) = 0$. There exists an interval Ω such that $\Omega \subseteq supp(Z \mid W)$, $supp(\alpha_K + e_k - U^* \mid W) \subseteq \Omega$, and $supp(\alpha_0 + e_0 - U^* \mid W) \subseteq \Omega$. The relative thresholds t_k and conditional expectations $h_k(Z, W)$ exist and are finite.

If Assumption 2 holds, then relative thresholds t_k still satisfy equation (6). To prove this, let F denote the joint distribution function of U^*, e, and following the same steps as before we have

$$
\begin{aligned}
& E\left[\int_\Omega E\left(Y_0 - Y_k \mid Z = z, W\right) dz\right] \\
= \; & E\left[\int_\Omega \int_{\text{supp}(U^*, e|W)} [I(0 \le Z + U^* - \alpha_0 - e_0) - I(0 \le Z + U^* - \alpha_k - e_k)] dF(U^*, e|W) dz\right] \\
= \; & E\left[\int_{\text{supp}(U^*, e|W)} \int_{-U^* + \alpha_0 + e_0}^{-U^* + \alpha_k + e_k} dz dF(U^*, e \mid W) dz\right] = E\left[E(\alpha_k + e_k - \alpha_0 - e_0) \mid W)\right] = t_k
\end{aligned}
$$

It follows that equation (7) remains a consistent estimator of the relative thresholds

5 Limiting Distribution

To simplify the limiting distribution theory for $\widehat{t_k}$, assume that supp(U^*) is bounded, which implies that $Y_0 - Y_k$, and therefore $h_k(z, w)$, will equal zero for all z outside of some interval, say Ω^*. If both $P[Z > \sup(\Omega^*) \mid W = w]$ and $P[Z < \inf(\Omega^*) \mid W = w]$ are nonzero for all $w \in supp(W)$, then estimation error in $\widehat{h}_k$ near the boundary of the support of z must be asymptotically irrelevant, because $Y_0 - Y_k$ and therefore $h_k(z, w)$ is identically zero near the boundary (equivalent to fixed trimming). As a result, the otherwise unknown interval Ω in equation (7) can be replaced by the real line, or by the extreme observations of Z, without affecting the limiting distribution. Equation (7) is then a standard two step estimator with nonparametric first step, so assuming smoothness of the function h_k and a kernel regression estimator $\widehat{h}_k$ with an appropriately chosen kernel and bandwidth sequence, by theorems 8.2 and 8.12 of Newey and McFadden (1994), $\widehat{t_k}$ will have the limiting distribution

$$\sqrt{n}(\widehat{t_k} - t_k) \Longrightarrow N\left(0, var\left[S_k - E\left(S_k \mid Z, W\right) + E\left(S_k \mid W\right)\right]\right) \qquad (10)$$

where $S_k = (Y_0 - Y_k)/f(Z \mid W)$ and $f(Z \mid w)$ is the conditional probability density function of Z given $W = w$. Details are omitted because they are entirely standard.

6 Conclusions

This paper shows that many more ordered choice threshold estimators exist than had been previously recognized, and proposes a new estimator that both holds under very general assumptions regarding both the latent variable Y^* and the (possibly randomly varying) thresholds. This estimator is based on a new representation of the thresholds, given by equation (6).

The threshold estimator of equation (7) has the disadvantage of requiring a first stage estimate of the function h_k, though not of β or $g(W)$. It may be less efficient than other estimators that impose stronger conditions regarding the joint distribution of errors, thresholds, and regressors. An advantage of the proposed estimator is that it applies with very few restrictions on the joint distribution of errors, thresholds, and regressors. The estimator does not depend on a specific model for Y^* beyond equations (8), (4), and (9), and so could for example be used in Hausman tests against threshold estimates obtained from models with more structure. The proposed estimator is also computationally simple, in that it does not require numerical searches or optimization. It could therefore be used to provide consistent starting values (or as the first step in two step estimators) of more computationally cumbersome estimation methods such as maximum likelihood.

References

[1] CAMERON, S. V. AND J. J. HECKMAN (1998), "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy,* 106, 262-333.

[2] CARNEIRO, P., K. T. HANSEN, AND J. J. HECKMAN (2003), "Estimating Distributions of Treatment Effects With an Application to the Returns to Schooling," *International Economic Review,* 44, 361-422.

[3] CHEN, S. (1999), "Semiparametric Estimation of a Location Parameter in the Binary Choice Model," *Econometric Theory,* 15, 79-98.

[4] CHEN, S. (2000), "Rank Estimation of a Location Parameter in the Binary Choice Model," *Journal of Econometrics, 98, 317-334.*

[5] HECKMAN, J. J., R. LALONDE, AND J. SMITH (1999), "The Economics and Econometrics of Active Labor Market Programs," in O. Ashenfelter and D. Card, eds, *Handbook of Labor Economics*, vol. 3, Amsterdam: Elsevier.

[6] HOROWITZ, J. L. (1992), "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505-532.

[7] KLEIN, R. W. AND R. P. SHERMAN (2002), "Shift Restrictions and Semiparametric Estimation in Ordered Response Models," *Econometrica, 70, 663-692.*

[8] LEWBEL, A. (1997), "Semiparametric Estimation of Location and Other Discrete Choice Moments," *Econometric Theory, 13, 32-51.*

[9] LEWBEL, A. (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145-177.

[10] LEWBEL, A., O. LINTON, AND D. MCFADDEN (2001), "Estimating Features of a Distribution From Binomial Data," unpublished manuscript.

[11] MAGNAC, T. AND E. MAURIN. (2003), “Identification and Information in Monotone Binary Models,” unpublished manuscript.

[12] MANSKI, C. (1985), ”Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator,” *Journal of Econometrics*, 27, 313-334.

[13] MCFADDEN, D. (1999), ”Estimation of Social Value From Willingness to Pay Data,” in Trade, Theory, and Econometrics: Essays in Honor of John S. Chipman, edited by J. R. Melvin, J. C. Moore, and R. Reizman, Routledge: London.

[14] NEWEY, W. K. AND D. MCFADDEN (1994), ”Large Sample Estimation and Hypothesis Testing,” in Handbook of Econometrics, vol. iv, ed. by R. F. Engle and D. L. McFadden, pp. 2111-2245, Amsterdam: Elsevier.