A Graphical Exposition of the Ordered Probit
Author(s): William E. Becker and Peter E. Kennedy
Source: *Econometric Theory*, Vol. 8, No. 1 (Mar., 1992), pp. 127-131
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/3532149
Accessed: 24/03/2014 08:51

Econometric Theory, **8**, 1992, 127–131. Printed in the United States of America.

A GRAPHICAL EXPOSITION OF THE ORDERED PROBIT

WILLIAM E. BECKER
Indiana University

PETER E. KENNEDY
Simon Fraser University

A three-dimensional diagram is used to illustrate the ordered probit model.

1. INTRODUCTION

The use of probit and logit models has become quite common whenever the dependent variable in a regression is qualitative. These models have been used to explain either/or choices and decisions involving multiple alternatives. A two-dimensional graphical interpretation of these different models has been provided by Johnson [3]. The purpose of this paper is to provide a three-dimensional graphical exposition of the ordered probit model, which was first estimated by McKelvey and Zavoina [4] and is now built into computer packages, such as LIMDEP [1].

Unlike other probit and logit models, the ordered probit model involves a qualitative dependent variable for which the categories have a natural order or ranking that reflects the magnitude of some underlying continuous variable/index. Bond ratings, for example, are expressed in terms of categories (triple A, double A, etc.) which could be viewed as resulting from a continuous measure called "creditworthiness"; letter grades assigned students for a course could be viewed as being generated by a continuous measure called "knowledge of course material"; reaction of patients to a drug dose could be categorized as no reaction, slight reaction, severe reaction, or death, corresponding to a conceptual continuous measure called "degree of allergic reaction." By ignoring the existence of the underlying continuous measure, and the inherent ordering, the multinomial probit or logit models mis-specify the data-generating process, creating the possibility that inferences about the response variable may be completely erroneous. Ordinary least-squares regression estimation is likewise inappropriate because the coding of the dependent variable in these cases, usually as 0, 1, 2, 3, 4, and so on, reflects only a ranking; the difference between a 1 and a 2 cannot be treated as equivalent to the difference between a 2 and a 3, for example.

2. THE ORDERED PROBIT MODEL

Suppose that y^* is an unobservable index determined as $y^* = \alpha + \Sigma\beta_i x_i + \epsilon$ where the summation is over $i = 1,\ldots,K$, the x_i are K independent vari-

 0266-4666/92 $5.00 + .00

ables, and ϵ is a disturbance. Assume that the dependent variable has $J + 1$ categories, so that instead of observing y^* we observe

$$
\begin{array}{lll}
y = 0 & \text{if} & y^* \leq \delta_0 \\
y = 1 & \text{if} & \delta_0 < y^* \leq \delta_1 \\
y = 2 & \text{if} & \delta_1 < y^* \leq \delta_2 \\
\vdots & \vdots & \vdots \\
y = J & \text{if} & \delta_{J-1} < y^*.
\end{array}
$$

The δ's are unknown "threshold" parameters that must be estimated along with α and the β_i's. Estimation is undertaken by maximum likelihood, which in the case of the ordered probit model requires that ϵ be assumed to be distributed as a standard normal. (The ordered logit model results from assuming that the cumulative density of ϵ is the logistic function.)

The probability of obtaining an observation with $y = 0$ is equal to

$$
\begin{aligned}
&\text{prob}\{y^* = \alpha + \Sigma\beta_i x_i + \epsilon \leq \delta_0\} \\
&\quad = \text{prob}\{\epsilon \leq \delta_0 - \alpha - \Sigma\beta_i x_i\} \\
&\quad = \int_{-\infty}^{\delta_0 - \alpha - \Sigma\beta_i x_i} f(\epsilon)\, d\epsilon
\end{aligned}
$$

where $f(\epsilon)$ is the standard normal density function. The probability of obtaining an observation with $y = 1$ is equal to

$$
\begin{aligned}
&\text{prob}\{\delta_0 < y^* = \alpha + \Sigma\beta_i x_i + \epsilon \leq \delta_1\} \\
&\quad = \text{prob}\{\delta_0 - \alpha - \Sigma\beta_i x_i < \epsilon \leq \delta_1 - \alpha - \Sigma\beta_i x_i\} \\
&\quad = \int_{\delta_0 - \alpha - \Sigma\beta_i x_i}^{\delta_1 - \alpha - \Sigma\beta_i x_i} f(\epsilon)\, d\epsilon.
\end{aligned}
$$

Similar expressions can be found for the probabilities of obtaining other observed y values. The likelihood function is the product of such expressions for each of the data points; maximizing this function with respect to α, the β's and the δ's produces the maximum likelihood estimates. When $\alpha \neq 0$, so that an intercept is included in the equation for y^*, identification is achieved by setting δ_0 equal to zero; specifying that ϵ has mean zero and variance one is also done for purposes of identification.

3. AN EXAMPLE

Our graphical exposition of this technique is presented in terms of a specific example, taken from [2, pp. 705–706], in which Navy recruits are classified into one of three technical job categories, with clearly ranked skill ratings:

medium skilled, highly skilled, and nuclear qualified/highly skilled. The explanatory variables are x_1 = a dummy variable indicating that the entrant possesses an "A school" (technical training) guarantee, x_2 = educational level of the entrant's mother, x_3 = score on the Air Force Qualifying test, x_4 = a dummy variable indicating that the entrant is married, x_5 = the entrant's age, and x_6 = years of education completed by the entrant. In terms of the general model given earlier, $K = 6$ and $J = 2$; there is only one δ, δ_1, to be estimated, since δ_0 is normalized to zero. The sample size is 5641. The maximum likelihood estimates of the intercept α, the six slope coefficients β_i, and the unknown threshold parameter δ_1 are reported in Table 1.

4. A GRAPHICAL EXPOSITION

For expositional reasons, we write the estimated y^* relationship as a function of x_6, conditional on all the other explanatory variables at their sample mean values. This yields the estimated relationship $y^* = -1.45 + 0.19x_6$ graphed in the three-dimensional diagram in Figure 1. The density of ϵ is measured on the vertical axis, with the unknown y^* values and the known x_6 values measured on the horizontal plane. For each x_6 value we can conceptualize a standard normal density, centered on the estimated y^* line; two of these are shown in Figure 1, for $x_6 = 12.0$ and $x_6 = 15.0$. Consider first the density relating to $x_6 = 12.0$. The base of this density, referred to below as the ϵ axis, is drawn in as a line parallel to the y^* axis at $x_6 = 12.0$, centered to be zero at the point at which it cuts the estimated y^* line. This point, $\epsilon = 0.0$, corresponds to $y^* = 0.83$ on the y^* axis.

In this example there are only two estimated threshold values, $\hat{\delta}_0 = 0.0$ and $\hat{\delta}_1 = 1.79$, shown on the y^* axis. On the ϵ axis, they correspond to $\epsilon = -0.83$ and 0.96, respectively. Consider now an entrant with $x = 12.0$ and all

TABLE 1. Parameter estimates

Parameter	Estimate	Variable mean
α	−4.34	
β_1	0.057	0.66
β_2	0.007	12.1
β_3	0.039	71.2
β_4	−0.48	0.08
β_5	0.0015	18.8
β_6	0.190	12.1
δ_0	0.0	
δ_1	1.79	

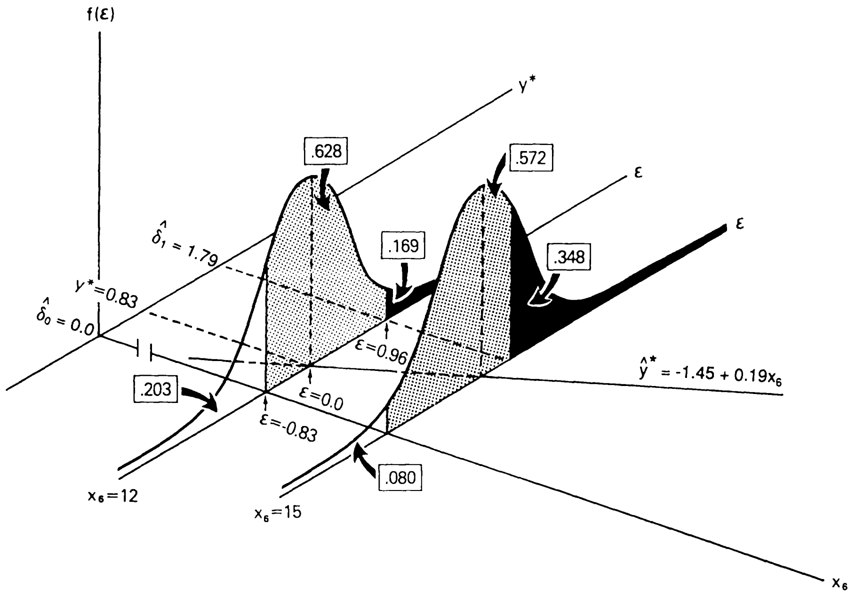

Figure 1. Illustrating the ordered probit model. The standard normal ϵ densities are drawn centered (i.e., where $\epsilon = 0.0$) at the estimated y^* values corresponding to $x_6 = 12$ and $x_6 = 15$. The estimated probabilities of job classifications are given by the areas under these densities determined by the estimated threshold values $\delta_0 = 0.0$ and $\delta_1 = 1.79$.

other x values equal to their respective sample averages. The estimated probability that such an individual will be assigned a "medium skilled" job is equal to the estimated probability that y^* is less than $\hat{\delta}_0 = 0.0$, equal to the probability that $\epsilon \leq -0.83$, given by the area under a standard normal to the left of -0.83. This is shown in Figure 1 as the clear area under the density. The shaded area under the density gives the estimated probability that this entrant will be assigned a "highly skilled" job (the probability that ϵ lies between -0.83 and 0.96), and the black area under this density yields the estimated probability that he/she will be assigned a "nuclear qualified/highly skilled" job (the probability that ϵ exceeds 0.96). Calculation of these probabilities can be undertaken using a table of the cumulative density of the standard normal distribution. The estimated probabilities (0.203, 0.628, and 0.169, respectively) are shown in boxes in Figure 1.

Consider now the effect of a ceteris paribus increase in x_6, from 12.0 to 15.0, on the estimated probabilities of this entrant being classified into each of the three job categories. When $x_6 = 15.0$ the comparable estimated probabilities are given by the clear, shaded, and black areas under the ϵ density corresponding to the line $x_6 = 15.0$. As compared to when $x_6 = 12.0$, the

Table 2. Calculations for Figure 1

	$x_6 = 12$	$x_6 = 15$
y^*	$-1.45 + 0.19 \times 12 = 0.83$	$-1.45 + 0.19 \times 15 = 1.40$
$\text{prob}(y = 0) = \text{prob}(y^* \leq 0)$	$\text{prob}(\epsilon \leq -0.83) = 0.203$	$\text{prob}(\epsilon \leq -1.40) = 0.080$
$\text{prob}(y = 1) = \text{prob}(0 < y^* \leq 1.79)$	$\text{prob}(-0.83 < \epsilon \leq 0.96) = 0.628$	$\text{prob}(-1.40 < \epsilon \leq 3.19) = 0.572$
$\text{prob}(y = 2) = \text{prob}(y^* > 1.79)$	$1 - 0.203 - 0.628 = 0.169$	$1 - 0.080 - 0.572 = 0.348$

clear area has shrunk (to 0.080), the black area has expanded (to 0.348), and, in this case, the shaded area has shrunk (to 0.572). Table 2 summarizes the calculations undertaken to obtain the estimated probabilities illustrated in Figure 1.

It should be clear that the impact of a unit change in an explanatory variable on the estimated probability of a particular classification depends on the sign and magnitude of the slope coefficient and the shape of the relevant portion of the normal density. In this example, the estimate of the slope coefficient β_6 is positive, so an increase in x_6 unequivocally increases the area in the upper tail and decreases the area in the lower tail; these changes will be greater, the greater is the estimated magnitude of β_6. The direction of the impact on the shaded area, however, can go either way; it depends on the shape of the density as well as the magnitude of the β_6 estimate. This reflects a general result that although the impact of a change in an explanatory variable on the estimated probabilities of the highest and lowest of the ordered classifications is unequivocal, the impact on the estimated probabilities of intermediate classifications cannot be determined a priori. Furthermore, because of the identifying normalization that the variance of ϵ equals unity, the absolute magnitude of the coefficient estimates cannot be given any meaning. Consequently, care should be taken when interpreting the coefficient estimates.

REFERENCES

1. Greene, W.H. LIMDEP: An econometric modeling program for the IBM PC. *The American Statistician* 39 (1985): 210.
2. Greene, W.H. *Econometric Analysis.* New York: Macmillan, 1990.
3. Johnson, T. The analysis of qualitative and limited responses. In W. Becker and W. Walstad (eds.) *Econometric Modeling in Economic Education Research*, pp. 141–184, Boston: Kluwer Nijhoff, 1987.
4. McKelvey, R.D. & W. Zavoina. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4 (1975): 103–120.