Discrete Choice Models

Ordinal Model

Mohammad Arshad Rahman
(webpage: https://www.arshadrahman.com)

Course: Econometrics
Chennai Mathematical Institute

Discrete Choice Models

Often, the dependent variable we seek to model takes on a range of values that are restricted. This is broadly defined as **limited dependent variable (LDV) models**.

Within the class of LDV, a special case arises when the outcome is a discrete variable. Such data often arise when individuals make a choice from a set of potential discrete outcomes, thus earning the name **discrete choice** models.

The most common case of discrete choice models are **binary models**, where y takes only two values typically coded as 1 and 0, indicating whether or not the event has occurred. Binary models are special cases of both ordinal and multinomial models. Example, participation in the labour force.

Discrete Choice Models

Multinomial models arise when y takes on multiple (more than two) discrete values, with no natural ordering. Consider for example, the choice of brand of toothpaste or mode of transportation.

Ordinal models or **ordered choice models** arise when y takes on multiple (more than two) discrete values that are inherently ordered or ranked. For example, scores attached to opinion on surveys (oppose, neutral, support), classification of educational attainment, or ratings on bonds.

Random Utility Framework

Discrete choice models have their foundations in the theory of choice in economics, which itself is inherently related with the random utility model (RUM). The RUM involves a utility maximizing rational individual whose objective is to choose an alternative from a set of *mutually exclusive* and completely *exhaustive* alternatives.

The utilities attached with each alternative are completely known to the decision maker and s/he chooses the same alternative in replications of the experiment. However, to a researcher the utilities are unknown, since s/he only observes a vector of characteristics (such as age, gender, income etc.) of the decision maker, referred to as *representative utility*. This forms the systematic component.

Random Utility Framework

The unobserved factors form the stochastic part, which is assigned a distribution, typically continuous, to make probabilistic statements about the observed choices conditional on the representative utility. The distributional specification implies that there exists a continuous latent random variable (or a continuous latent utility) that underlies the discrete outcomes.

When the set of alternatives or outcomes are inherently ordered or ranked, individual choice of a particular alternative can be associated as the latent variable crossing a particular threshold or cut-point. This latent variable threshold-crossing formulation of the ordered choices elegantly connects individual choice behavior and ordinal data models serve as a useful tool in the estimation process.

Ordinal Models

Ordinal regression models generalize the binary models by allowing the dependent variable to have more than two outcomes which are inherently ordered or ranked. Each outcome or category is assigned a score (value or number) with the characteristic that the scores have an ordinal meaning but hold no cardinal interpretation. Therefore, the difference between categories is not directly comparable.

Two popular ordinal regression models are ordinal probit and ordinal logit models. Ordinal probit assumes that the errors follow a standard normal distribution, while ordinal logit assumes that the errors are drawn from a standard logistic distribution.

Ordinal Models

Example 1: We code the public response to marijuana legalization as: 1 for 'oppose legalization', 2 for 'legal only for medicinal use', and 3 for 'legal for personal use'. Here, a score of 2 implies more support for legalization as compared to 1, but we cannot interpret a score of 2 as twice the support compared to a score of 1. Similarly, the difference in support between 2 and 1 is not the same as that between 3 and 2. See Batham et al (2023).

Other examples: modeling support for offshore drilling (Mukherjee & Rahman, 2016), educational attainment (Rahman, 2016), opinion on homeownership as the best long-term investment (Rahman & Karnawat, 2019).

Ordinal Models

Similar to binary models, the ordinal regression model can be expressed in terms of a continuous latent variable z_i[1] as follows:

$$\begin{aligned} z_i &= x_i'\beta + \varepsilon_i, & \forall\ i = 1, \cdots, n, \\ \gamma_{j-1} < z_i \leq \gamma_j \ &\Rightarrow\ y_i = j, & \forall i,\ j = 1, \cdots, J, \end{aligned} \tag{1}$$

where (in the first equation) x_i is a $k \times 1$ vector of covariates, β is a $k \times 1$ vector of unknown parameters, and n denotes the number of observations. The second equation shows how z_i is related to the observed discrete response y_i, where $-\infty = \gamma_0 < \gamma_1 \cdots < \gamma_{J-1} < \gamma_J = \infty$ are the cut-points (or thresholds) and y_i is assumed to have J categories or outcomes.

[1] The continuous latent construct may represent underlying latent utility, some kind of propensity, or strength of preference

Ordinal Model

A visual representation of the outcome probabilities (for the case of marijuana legalization) and the cut-points are presented in Figure 1.

Different combinations of (β, γ) can produce the same outcome probabilities giving rise to parameter identification problem. We need to anchor the location and scale of the distribution to identify the model parameters. The former is achieved by setting $\gamma_1 = 0$ and the latter by assuming $\mathrm{var}(\varepsilon_i) = 1$ (follows from ordinal probit model).

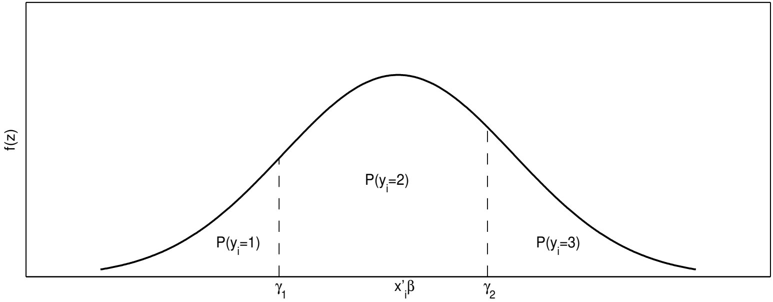

Figure 1: The two cut-points (γ_1, γ_2) divide the area under the curve into three parts, with each part representing the probability of a response falling in the three response categories. The three probabilities $\mathrm{P}(y_i = 1)$, $\mathrm{P}(y_i = 2)$ and $\mathrm{P}(y_i = 3)$ correspond to 'oppose legalization', 'legal only for medical use' and 'legal for personal use', respectively. Note that for each individual i the mean $x_i'\beta$ will be different and so will be the category probabilities. Source: authors' creation.

Ordinal Probit Model

Ordinal probit model arises, if we assume that the errors are *iid* as a standard normal distribution, i.e., $\varepsilon_i \sim N(0,1)$ for $i = 1(1)n$.

Given $y = (y_1, \cdots, y_n)'$, the likelihood for the ordinal probit model expressed as a function of (β, γ) is the following,

$$\begin{aligned} L(\beta, \gamma; y) &= \prod_{i=1}^{n} \prod_{j=1}^{J} \Pr(y_i = j | \beta, \gamma)^{I(y_i = j)} \\ &= \prod_{i=1}^{n} \prod_{j=1}^{J} \left[\Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta) \right]^{I(y_i = j)}, \end{aligned} \tag{2}$$

where $\Phi(\cdot)$ is the *cdf* of a standard normal distribution and $I(y_i = j)$ is an indicator function, which equals 1 if the condition within parenthesis is true and 0 otherwise.

Ordinal Probit Model

The parameter estimates for (β, γ) are obtained by maximizing the log-likelihood, i.e., logarithm of the likelihood given by equation equation (2), either using the Newton-Raphson method or BHHH procedure.

Once the parameter estimates are available, they may be used to calculate the covariate effects, make predictions or assess model fitting.

Ordinal Logit Model

The framework for ordinal probit model can be transformed into an *ordinal logit model* (or *ordered logit model*) by simply assuming that the error follows a standard logistic distribution i.e., $\varepsilon_i \sim L(0, \pi^2/3)$ for $i = 1(1)n$, where L denotes a logistic distribution with mean 0 and variance $\pi^2/3$.

The likelihood for the ordinal logit model has the same structure as equation (2) with $\Phi(w)$ replaced by $\Lambda(w) = \exp(w)/[1 + \exp(w)]$, where Λ denotes the *cdf* of the standard logistic distribution and w is the argument inside the parenthesis. Once again, the parameters are estimated using the ML technique.

Ordinal Logit Model

For an ordinal logit model, the ratio of odds of not exceeding a certain category (say j) for any two individuals is constant across response categories. This earns it the name *proportional odds model*.

To see this property in effect, let $\theta_{ij} = \Pr(y_i \leq j)$ denote the cumulative probability that individual i chooses category j or below. For the ordinal logit model, this implies:

$$\theta_{ij} = \exp(\gamma_j - x_i'\beta) / \left[1 + \exp(\gamma_j - x_i'\beta)\right],$$
$$\theta_{ij}/(1 - \theta_{ij}) = \exp\left[\gamma_j - x_i'\beta\right],$$

where the latter represents the odds for the event $y_i \leq j$.

Ordinal Logit Model

Accordingly, for any two individuals (say 1 and 2), the ratio of odds can be written as,

$$\frac{\theta_{1j}/(1-\theta_{1j})}{\theta_{2j}/(1-\theta_{2j})} = \exp\left[-(x_1 - x_2)'\beta\right]. \tag{3}$$

The odds ratio presented in equation (3) does not depend on the response category j and is proportional to $(x_1 - x_2)$ with β being the constant of proportionality.

Covariate Effects

In binary and ordinal models, the coefficients do not give covariate effects (CE) because the link function is non-linear. Consequently, we need to calculate the CE for each outcome.

Let $x_{\cdot,l}$ be a continuous covariate, then the CE for the i-th observation (or individual) in an ordinal probit model is,

$$\begin{aligned} \frac{\partial \Pr(y_i = j)}{\partial x_{i,l}} &= -\beta_l \left[\phi(\gamma_j - x_i'\beta) - \phi(\gamma_{j-1} - x_i'\beta) \right] \\ &\simeq -\hat{\beta}_l \left[\phi(\hat{\gamma}_j - x_i'\hat{\beta}) - \phi(\hat{\gamma}_{j-1} - x_i'\hat{\beta}) \right], \end{aligned} \tag{4}$$

where $\phi(\cdot)$ denotes the probability density function (*pdf*) of a standard normal distribution and $(\hat{\beta}, \hat{\gamma})$ are the ML estimates of (β, γ). The average CE is computed by averaging the CE in equation (4) across all observations.

Covariate Effects

If the covariate is an indicator variable (say $x_{\cdot,m}$), then the CE for the i-th observation on outcome j $(= 1, \ldots, J)$ is calculated as,

$$
\begin{aligned}
&\Pr(y_i = j | x_{i,-m}, x_{i,m} = 1) - \Pr(y_i = j | x_{i,-m}, x_{i,m} = 0) \\
&= \left[\Phi(\gamma_j - x'^{\dagger}_i\beta) - \Phi(\gamma_{j-1} - x'^{\dagger}_i\beta)\right] - \left[\Phi(\gamma_{j-1} - x'^{\ddagger}_i\beta) - \Phi(\gamma_{j-1} - x'^{\ddagger}_i\beta)\right] \qquad (5) \\
&\simeq \left[\Phi(\hat{\gamma}_j - x'^{\dagger}_i\hat{\beta}) - \Phi(\hat{\gamma}_{j-1} - x'^{\dagger}_i\hat{\beta})\right] - \left[\Phi(\hat{\gamma}_{j-1} - x'^{\ddagger}_i\hat{\beta}) - \Phi(\hat{\gamma}_{j-1} - x'^{\ddagger}_i\hat{\beta})\right],
\end{aligned}
$$

where $x_i^{\dagger} = (x_{i,-m}, x_{i,m} = 1)$ and $x_i^{\ddagger} = (x_{i,-m}, x_{i,m} = 0)$. The average CE can be calculated from equation (5).

Note: In ordinal models, the sign of the regression coefficient translates unambiguously into the sign of covariate effect only for the lowest and highest categories of response variable. Covariate effect for the middle categories cannot be known *a priori*.

AIC and BIC

To compare ordinal models, one may compute the Akaike information criterion (AIC) and Bayesian information criterion (BIC).

$$\begin{aligned} \text{log-like} &= \log \ell(\hat{\theta}; y) \\ \text{AIC} &= -2 \, \text{log-like} + 2k, \\ \text{BIC} &= -2 \, \text{log-like} + k \log(n), \end{aligned} \tag{6}$$

where k is the number of estimated model parameters and n is the number of observations. Both measures are based on information theoretic approach. For any two models, we choose the model with lower AIC/BIC.

Information Theoretic Approach

Information theoretic (IT) approach: When a statistical model is used to represent the process that generated the data, the representation is typically not exact and some information is lost. IT measures (e.g., AIC and BIC) estimate the relative amount of information lost by a given model: the less information a model loses, higher the quality of that model.

In estimating the amount of information lost by a model, IT measures deal with the trade-off between the goodness of fit of the model and the simplicity of the model. In other words, AIC and BIC deals with both the risk of overfitting and the risk of underfitting.

LR Statistic

Some non-IT measures include the likelihood ratio (LR) statistic, McFadden's R-squared, and the hit-rate.

The LR test statistic λ_{LR} for $H_0 : \beta_2 = \ldots = \beta_k = 0$, is:

$$\lambda_{LR} = -2[\ln L_0 - \ln L_{\text{fit}}] \overset{H_0}{\sim} \chi^2_{k-1},$$

where $\ln L_{\text{fit}}$ ($\ln L_0$) is the log-likelihood of fitted (intercept-only) model. Under H_0, λ_{LR} follows a chi-square distribution with *d.f.* equaling $k-1$, i.e., the number of restrictions under H_0.

We calculate λ_{LR} and compare it with χ^2_{k-1} for a given level of significance. If $\lambda_{LR} > \chi^2_{k-1}$, we reject H_0; else we fail to reject H_0.

McFadden's R-squared

The McFadden's R-squared (pseudo R-squared or likelihood ratio index) is defined as follows,

$$R_M^2 = 1 - \frac{\ln L_{\text{fit}}}{\ln L_0}. \tag{7}$$

The R_M^2 is bounded between 0 and 1, similar to the coefficient of determination (R^2) in linear regression models.

When all slope coefficients are zero, the R_M^2 equals zero; but in discrete choice models R_M^2 can never equal 1, although it can come close to 1. Higher values of R_M^2 implies better fit, but it has no natural interpretation like R^2 that denotes the proportion of variation in the dependent variable explained by the covariates.

Hit-rate

The hit-rate (HR) is defined as the percentage of correct predictions i.e., percentage of observations for which the model correctly assigns the highest probability to the observed response category. Mathematically, the HR is defined as,

$$HR = \frac{1}{n}\sum_{i=1}^{n} I\left(\left(\max_{j}\,\{\hat{p}_{ij}\}_{j=1}^{J}\right) = y_i\right), \tag{8}$$

where $\hat{p}_{ij}$ is the predicted probability that individual i selects outcome j, and $I(\cdot)$ is the indicator function as defined earlier.

Extent of Marijuana Legalization

We implement the ordinal probit model to study the extent of marijuana legalization in the United States as studied in Batham et al (2023).

Other application of ordinal models: modeling support for offshore drilling (Mukherjee & Rahman, 2016), educational attainment (Rahman, 2016), opinion on homeownership as the best long-term investment (Rahman & Karnawat, 2019).

A comprehensive treatment of ordinal models is presented in Johnson & Albert (2000) and Green & Hensher (2010).

Data

We use the February 2014 Political Survey conducted during February 14-23, 2014, by the Princeton Survey Research Associates and sponsored by the Pew Research Center for the People and the Press, to analyze public opinion on extent of marijuana legalization.

In the survey, a representative sample of 1,821 adults living in the US were interviewed over telephone with 481 (1,340) individuals interviewed over land line (cell phone, including 786 individuals without a land line phone). After removing missing observations and respondents who were unsure about their opinion on legalization, we are left with a sample of 1,492 observations for the study.

Data

The dependent variable in the model is the respondents' answer to the question, "Which comes closer to your view about the use of marijuana by adults?". The options provided were, 'It should not be legal,' 'It should be legal only for medicinal use,' or 'It should be legal for personal use'. The fourth category labeled, 'Don't know/Refused' is removed from the study.

The February 2014 Survey also collected information on the age, income, household size, past use, gender, education, race, party affiliation and religion. We use these variables as independent variables in the models.

Data

We also include the indicator variable ‘tolerant states’, which indicates if a respondent lives in one of the 22 states, where recreational usage is legal, possession is decriminalized and/or allowed for medical use only.

In addition, we include the variable ‘eventually legal’, which indicates whether respondents expect marijuana to be legal irrespective of their individual opinion. We present the descriptive statistics for all the variables in Table 1.

Table 1: Descriptive summary of the variables (February 2014 Political Survey).

VARIABLE		MEAN	STD
LOG AGE		3.72	0.44
LOG INCOME		10.63	0.98
HOUSEHOLD SIZE		2.74	1.42
	CATEGORY	COUNTS	PERCENTAGE
PAST USE		719	48.19
MALE		792	53.02
EDUCATION	BACHELORS & ABOVE	551	36.93
	BELOW BACHELORS	434	29.09
	HIGH SCHOOL & BELOW	507	33.98
TOLERANT STATES		556	37.27
EVENTUALLY LEGAL		1,154	77.35
RACE	WHITE	1149	77.01
	AFRICAN AMERICAN	202	13.54
	OTHER RACES	141	9.45
PARTY AFFILIATION	REPUBLICAN	333	22.32
	DEMOCRAT	511	34.25
	INDEPENDENT & OTHERS	648	43.43

Table 1–Continued from previous page

VARIABLE	CATEGORY	COUNTS	PERCENTAGE
RELIGION	PROTESTANT	550	36.86
	ROMAN CATHOLIC	290	19.44
	CHRISTIAN	182	12.20
	CONSERVATIVE	122	8.18
	LIBERAL	348	23.32
PUBLIC OPINION	OPPOSE LEGALIZATION	218	14.61
	LEGAL ONLY FOR MEDICINAL USE	640	42.90
	LEGAL FOR PERSONAL USE	634	42.49

Table 2: Estimation results for the ordinal probit model.

	MODEL 5		MODEL 6		MODEL 7		MODEL 8	
	COEF	SE	COEF	SE	COEF	SE	COEF	SE
INTERCEPT	1.26**	0.44	1.27**	0.45	0.83*	0.46	0.34	0.48
LOG AGE	−0.44**	0.07	−0.45**	0.07	−0.45**	0.08	−0.35**	0.08
LOG INCOME	0.07*	0.03	0.07*	0.03	0.08**	0.03	0.09**	0.04
PAST USE	0.74**	0.06	0.73**	0.06	0.71**	0.06	0.69**	0.06
MALE	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.06
BACHELORS & ABOVE	0.26	0.08	0.26**	0.08	0.25**	0.08	0.24**	0.08
BELOW BACHELORS	0.06	0.08	0.05*	0.08	0.05	0.08	0.05	0.08
HOUSEHOLD SIZE	−0.04*	0.02	−0.04	0.02	−0.03	0.02	−0.02	0.02
TOLERANT STATES	0.11*	0.06	0.13**	0.06	0.09	0.06	0.07	0.07
EVENTUALLY LEGAL	0.58**	0.07	0.58**	0.07	0.56**	0.07	0.57**	0.07
AFRICAN AMERICAN	..	..	0.11	0.09	−0.01	0.10	0.03	0.10
OTHER RACES	..	..	−0.18*	0.11	−0.26**	0.11	−0.27**	0.11
DEMOCRAT	..	..	..	..	0.48**	0.09	0.44**	0.09
OTHER PARTIES	..	..	..	..	0.40**	0.08	0.36**	0.08
ROMAN CATHOLIC	..	..	..	..	..	..	0.10	0.09
CHRISTIAN	..	..	..	..	..	..	0.16	0.10
CONSERVATIVE	..	..	..	..		..	0.09	0.12
LIBERAL	..	..	..	..	..	..	0.39**	0.09
CUT-POINT	1.43**	0.05	1.43**	0.05	1.45**	0.05	1.46**	0.05
LR (χ^2) STATISTIC		316.27		321.47		356.99		377.02
MCFADDEN'S R^2		0.10		0.11		0.12		0.12
HIT-RATE		57.77		57.44		59.11		58.91

$** \ p < 0.05$, $* \ p < 0.10$

Results

We focus on the results from Model 8 because it is the most general model, provides the best fit according to McFadden's R^2.

Logarithm of age has a negative effect on the support for personal use (third category) and is statistically significant at 5 percent level (default level). Alternatively, log age has a positive effect on opposing legalization (first category) and is statistically significant. However, the effect of age on medicinal use only (second category) cannot be determined *a priori*.

Moving forward, we only discuss the effect on personal use and the impact on opposing legalization will be the opposite to that of personal use.

Results

The coefficient for log income is positive and statistically significant. This implies that individuals with higher income are more likely to support legalization for personal use.

Past use of marijuana has a statistically significant positive effect on the probability of supporting personal use of marijuana. Moreover, the coefficient for past use is largest among all the variables.

The coefficient for male is positive, but is not statistically significant. Thus, the current data do not confirm any role of gender on public opinion towards marijuana.

Results

Higher education is positively associated with support for personal use of marijuana. However, only the coefficient for ‘bachelors degree & above’ is statistically significant.

The coefficients for household size and ‘tolerant states’ are not significant and conform to the results from the binary probit model.

We find that ‘eventually legal’ has a significant positive effect, indicating that if an individual expects marijuana to be legal irrespective of his or her opinion, then s/he is more likely to support legalization for personal use.

Results

The indicator variable for African American is not significant and hence opinions of African Americans on personal use of marijuana are not significantly different compared to the Whites.

'Other Races' has a significant negative coefficient and is more opposed to legalization as compared to Whites.

Affiliation to Democratic Party or 'Other Parties' increases the support for personal use and the coefficients are significant.

Religious affiliations do not show a strong effect. Only the Liberals are more supportive of personal use of marijuana, while the opinions of the remaining religious categories are not significantly different from the base category, Protestant.

Covariate Effects

The average covariate effects for all significant variables are presented in Table 3. From the table, it is clear that past use, eventually legal, and identifying oneself as a Democrat are three variables with the highest impact on public opinion.

Past use of marijuana increases support for personal use by 24.3 percentage points and decreases the support for medicinal use and oppose legalization by 11.3 and 12.9 percentage points, respectively.

Table 3: Average covariate effects from Model 8.

COVARIATE	ΔP(not legal)	ΔP(medicinal use)	ΔP(personal use)
AGE, 10 YEARS	0.015	0.012	–0.028
INCOME, $10,000	–0.005	–0.003	0.008
PAST USE	–0.129	–0.113	0.243
BACHELORS & ABOVE	–0.045	–0.035	0.080
EVENTUALLY LEGAL	–0.126	–0.060	0.186
OTHER RACES	0.059	0.031	–0.089
DEMOCRAT	–0.080	–0.066	0.147
OTHER PARTIES	–0.070	–0.051	0.121
LIBERAL	–0.068	–0.066	0.134

Covariate Effects

Similarly, a respondent who expects marijuana to be legal is 18.6 percentage points more likely to support marijuana for personal use. This increase comes from a decrease in probability for medicinal use and oppose legalization, which are 6.0 and 12.6 percentage points, respectively.

In the same way, a respondent who is a Democrat is 14.7 percentage points more likely to favor personal use, and 6.6 and 8.0 percentage points less likely to favor medicinal use and oppose legalization, respectively. The covariate effects for the remaining variables can be interpreted similarly.

Thank you!