Econometrics Homework 2

Utpalraj Kemprai
MDS202352

February 18, 2025

Question 1

We have the following ordinal regression model:

$$
\begin{gathered}
z_i = x_i'\beta + \epsilon_i \quad \forall i = 1, \cdots, n \\
\gamma_{j-1} < z_i \leq \gamma_j \implies y_i = j, \quad \forall i, j = 1, \cdots, J
\end{gathered}
$$

where (in the first equation) z_i is the latent variable for individual i, x_i is a 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 $-\infty = \gamma_0 < \gamma_1 < \gamma_{J-1} < \gamma_J = \infty$ are the cut-points (or thresholds) and y_i is assumed to have J categories or outcomes.

(a)

Probability of success

We assume that $\epsilon_i \sim N(0,1)$, for $i = 1, 2, \cdots, n$. Therefore we have,
The probability of success,

$$
\begin{aligned}
Pr(y_i = j) &= Pr(\gamma_{j-1} < z_i \leq \gamma_j) \\
&= Pr(\gamma_{j-1} < x_i'\beta + \epsilon_i \leq \gamma_j) \\
&= Pr(\gamma_{j-1} - x_i'\beta < \epsilon_i \leq \gamma_j - x_i'\beta) \\
&= \Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta) \qquad \text{[where } \Phi(\cdot) \text{ is the cdf of } N(0,1)]
\end{aligned}
$$

Likelihood function

The likelihood function for the ordinal probit model is,

$$
\begin{aligned}
L(\beta, \gamma; y) &= \prod_{i=1}^{n} \prod_{j=1}^{J} Pr(y_i = j | \beta, \gamma)^{I(y_i = j)} \qquad \text{[where } I(\cdot) \text{ is the indicator function]} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{J} (\Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta))^{I(y_i = j)}
\end{aligned}
$$

(b)

Probability of success

We assume that $\epsilon_i \sim \mathcal{L}(0,1)$, for $i = 1, 2, \cdots, n$. Therefore the probability of success,

$$
\begin{aligned}
Pr(y_i = j) &= Pr(\gamma_{j-1} < z_i \leq \gamma_j) \\
&= Pr(\gamma_{j-1} < x_i'\beta + \epsilon_i \leq \gamma_j) \\
&= Pr(\gamma_{j-1} - x_i'\beta < \epsilon_i \leq \gamma_j - x_i'\beta) \\
&= \frac{1}{1 + e^{-(\gamma_j - x_i'\beta)}} - \frac{1}{1 + e^{-(\gamma_{j-1} - x_i'\beta)}} \qquad \text{[as } \epsilon_i \sim \mathcal{L}(0,1)] \\
&= \frac{e^{-(\gamma_{j-1} - x_i'\beta)} - e^{-(\gamma_j - x_i'\beta)}}{(1 + e^{-(\gamma_j - x_i'\beta)})(1 + e^{-(\gamma_{j-1} - x_i'\beta)})} \\
&= \frac{e^{x_i'\beta}(e^{-\gamma_{j-1}} - e^{-\gamma_j})}{(1 + e^{-(\gamma_j - x_i'\beta)})(1 + e^{-(\gamma_{j-1} - x_i'\beta)})}
\end{aligned}
$$

Likelihood function

The likelihood function for the ordinal logit model is,

$$
\begin{aligned}
L(\beta,\gamma;y) &= \prod_{i=1}^{n}\prod_{j=1}^{J} Pr(y_i = j|\beta,\gamma)^{I(y_i=j)} && \text{[where } I(\cdot) \text{ is the indicator function]}\\
&= \prod_{i=1}^{n}\prod_{j=1}^{J}\left[\frac{e^{x_i'\beta}(e^{-\gamma_{j-1}} - e^{-\gamma_j})}{(1+e^{-(\gamma_j - x_i'\beta)})(1+e^{-(\gamma_{j-1}-x_i'\beta)})}\right]^{I(y_i=j)}
\end{aligned}
$$

(c)

Probability remain unchanged on adding a constant c to cut-points and the mean

Adding a constant c to the cut-point γ_j and the mean $x_i'\beta$, $\forall i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, J$, the latent variable z_i becomes $x_i'\beta + c + \epsilon_i$ and the cut-point γ_j becomes $\gamma_j + c$

The probability of y_i taking the value j is,

$$
\begin{aligned}
Pr(y_i = j) &= Pr(\gamma_{j-1} + c < z_i \leq \gamma_j + c)\\
&= Pr(\gamma_{j-1} + c < x_i'\beta + c + \epsilon_i \leq \gamma_j + c)\\
&= Pr(\gamma_{j-1} - x_i'\beta < \epsilon_i \leq \gamma_j - x_i'\beta)\\
&= \Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta)
\end{aligned}
$$

which is the same as the value obtained in part(a) of Question 1. So adding a constant c to the cut-point γ_j and the mean $x_i'\beta$ does not change the outcome probability.

Identification Problem

This identification problem can be solved by fixing the value of one of $\gamma_1, \gamma_2, \cdots, \gamma_{J-1}$. In particular setting $\gamma_1 = 0$ will solve this identification problem.

(d)

Rescaling the parameters (γ_j, β) and scale of distribution does not change outcome probability

Rescaling the parameters (γ_j, β) and the scale of the distribution of ϵ_i by some constant d, the latent variable z_i becomes $x_i'd\beta + \epsilon_i$ where $\epsilon_i \sim N(0, d^2)$ and the cut-point γ_j becomes $d\gamma_j, \forall j = 1, 2, \cdots, J-1$.

The probability of y_i taking the value j is,

$$
\begin{aligned}
Pr(y_i = j) &= Pr(d\gamma_{j-1} < z_i \leq d\gamma_j)\\
&= Pr(d\gamma_{j-1} < x_i'd\beta + \epsilon_i \leq d\gamma_j)\\
&= Pr(\gamma_{j-1} - x_i'\beta < \frac{\epsilon_i}{d} \leq \gamma_j - x_i'\beta)\\
&= \Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta) && \text{[as } \epsilon_i \sim N(0,d^2), \frac{\epsilon_i}{d} \sim N(0,1)\text{]}
\end{aligned}
$$

which is the same as the value obtained in part(b) of Question 1. So escaling the parameters (γ_j, β) and the scale of the distribution by some arbitrary constant d lead to same outcome probabilities.

Identification problem

This identification problem can be solved by fixing the scale of the distribution of ϵ_i. In particular we can set the scale of the distribution of ϵ_i to 1, i.e. var$(\epsilon_i) = 1$.

(e)

(i) Descriptive Summary of the data

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		791	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		1154	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
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 1: Descriptive Summary of the variables

(ii) Public opinion on extent of marijuana legalization

We estimate Model 8 and replicate the results from the lecture.

coef	Estimate	Std. Error	t value	p value
intercept	0.35	0.48	0.72	0.47
log_age	-0.35	0.08	-4.51	< 0.05
log_income	0.09	0.04	2.47	< 0.05
hh1	-0.02	0.02	-0.87	0.38
pastuse	0.69	0.06	10.67	< 0.05
bachelor_above	0.24	0.08	3.01	< 0.05
below_bachelor	0.05	0.08	0.62	0.54
tolerant_state	0.07	0.07	1.04	0.30
expected_legal	0.57	0.07	7.76	< 0.05
black	0.03	0.10	0.26	0.79
other_race	-0.28	0.11	-2.56	< 0.05
democrat	0.44	0.09	5.03	< 0.05
other_party	0.36	0.08	4.56	< 0.05
male	0.06	0.06	1.00	0.32
christian	0.16	0.10	1.60	0.11
roman_catholic	0.10	0.09	1.19	0.23
liberal	0.39	0.09	4.39	< 0.05
conservative	0.09	0.12	0.76	0.45
cut_point	1.46	0.05	29.58	< 0.05
LR (χ^2) Statistic	377			
McFadden's R^2	0.13			
Hit-Rate	58.91			

Table 2: Model 8: Estimation Results

(iii) Covariate effects of the variables

Covariate	Δ P(not legal)	Δ P(medicinal use)	Δ P(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	-0.35	0.080
Eventually Legal	-0.126	-0.60	0.186
Other Races	0.059	0.031	-0.089
Democrat	-0.80	-0.066	0.147
Other parties	-0.070	-0.051	0.121
Liberal	-0.68	-0.066	0.134

Table 3: Average covariate effects from Model 8.

Question 2

(a)

(b)

(c)