Agricultural and Applied Economics 637
Applied Econometrics II

Assignment VI
Estimation of Conditional Logit and Censored Regression Models
(Due: May 8, 2018)

(115 Total Points)

1. (65 pts). In a classic study of the choices faced by fisherman, Herriges and Kling (1999) used data from a 1989 survey of recreational anglers to provide information concerning their most recent saltwater angling experiences. Included in this data is information as to whether the angler used a beach, pier, private boat or charter boat as a mode of fishing. The authors computed a cost estimate of each angler's day of fishing which included an opportunity cost of travel time, transportation cash costs and any boat fees/costs. They also collected exogenous data on species-specific catch rate (defined on a per-hour-fished basis). Finally, they collected information on average monthly income for the 1,182 anglers included in their survey.

   As a consultant under contract with the California Charter Boat Association, you are interested in examining what determines the mode of fishing used by saltwater anglers. You decide that you would first like to use the above data in a discrete choice analysis of fishing mode. Given that there are 4 fishing modes (beach, pier, private boat, and charter boat) you decide that you would like to estimate a ***conditional logit*** model of fishing mode choice. The specific model you decide to estimate is the following:

[1.1]
$$P_{ij} = \Pr\left[y_i = j\right] = \frac{\exp\left(\beta_P COST_{ij} + \beta_C C\_RATE_{ij}\right)}{\sum_{r=1}^{4} \exp\left(\beta_P COST_{ir} + \beta_C C\_RATE_{ir}\right)}$$

$$\left(j = \text{Beach}\ (1),\ \text{Pier}\ (2),\ \text{Private Boat}\ (3),\ \text{Charter Boat}\ (4)\right)$$

   Note that β_P, β_C are coefficients to estimate, i identifies the angler, j pertains to a particular fishing mode (beach, pier, private boat, charter boat), $COST_j$ is the daily cost for the j^{th} mode and C_RATE_j is the associated catch rate.

The dataset ***New_fish_file*** contains the data necessary to estimate the model. Table 1 is used to show the records from 2 anglers contained in the above dataset. Note there are ***four records*** for each survey respondent.

Table 1.1: Sample of Observations in the ***New_fish_file*** Dataset

HHID	MODE_ID	CHOICE	COST	C_RATE	INCOME
1	1	0	1.5793	0.0678	0.7083
1	2	1	1.5793	0.0503	0.7083
1	3	0	1.5793	0.2601	0.7083
1	4	0	1.8293	0.5391	0.7083
2	1	0	0.1511	0.1040	0.1250
2	2	0	0.1551	0.0451	0.1250
2	3	0	0.1053	0.1574	0.1250
2	4	1	0.3453	0.4671	0.1250

The variables delineated above are defined below .

Table 1.2: Definition of Variables in the *New_fish_file* Dataset

MODE_ID	Identification of type of fishing mode: 1 = Beach 2 = Pier 3 = Private Boat 4 = Charter Boat
CHOICE	Identifies fishing mode actually used (0/1)
COST	Daily cost of fishing for each mode ($100)
C_RATE	Catch Rate (# fish/hour)
INCOME	Monthly income ($10,000)

(a) (10 pts) Estimate the parameters of the above conditional logit model using maximum likelihood techniques. Report the usual estimation results and at least one measure of explanatory power of the model.

(b) (10 pts) Undertake a *Hausman IIA* test to determine if the odds ratios for the private boat/charter boat and charter boat/pier separately exhibit IIA characteristics. Explain how you undertook this hypothesis test.

(c) (15 pts) Calculate the *average marginal response* of a $100 change in the daily cost of each fishing mode, i.e., average the marginal responses over all 1,182 households

on all modes. What is the *average marginal response* of the probability of choosing **each** alternative when daily fishing costs changes for **one** of the alternatives and unchanged for the other alternatives? (Note: There are four average marginal responses (1 own, 3 cross mode) to fishing mode choice given a particular mode cost change. Generate this 4 x 4 matrix of estimated average responses and a similar matrix of associated marginal response standard errors. Indicate in the 1st table which of the average marginal effects are statistically significant?

(d) (10 pts) Calculate the ***average elasticity*** value of a change in ***private boat catch rate*** on the probability of using each of the four modes. Test whether individually each of these elasticities are statistically significant. For the private and charter boat modes test whether the average elasticities are the negative of one another.

(e) (5 pts) Create a table that shows the mean income conditional on mode actually used. What is your null hypothesis as to the relationship between fishing mode and income?

(f) (15 pts) You would like to quantify the above relationship between household income and the probability of fishing mode use. Using the Beach mode as the base, extend the above model to include income as an explanatory variable. Present the usual maximum likelihood regression results table. Provide statistical evidence that income has a statistically significant impact (relative to the beach choice) on the probability of using a charter boat.

2. (50 pts) With the enactment of the North American Free Trade Agreement in the mid-1990's, Mexico has become the #1 export market for U.S. dairy products. One of the major value-added product categories that is becoming a more important component of U.S. dairy exports is cheese. As such, there is significant interest by U.S. manufacturers as to how responsive Mexican cheese consumers are to changes in product price, the level of their household income and other household characteristics. Assume you have been hired by the U.S. Dairy Export Council (USDEC, www.usdec.org) to quantify the determinants of cheese purchases by Mexican households. To do this, you have decided to use the biannual survey of households, *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)*. This household survey contains data on quantity purchased and associated expenditures on a detailed set of food and non-food items over a 1-week survey period. Table 1 contains a listing of the variables you have obtained

from the full survey dataset. Based on this data you discover that over 60% of the surveyed households did not report cheese purchases over the survey period. To obtain consistent parameter estimates of your model of cheese ***purchases*** while accounting for these zero values, you decide to estimate a Tobit model.

(a) (15 pts) To help USDEC obtain a better understanding of what impacts Mexican household cheese ***demand***, you decide to use the above dataset to estimate the following per capita "demand" function:

[2.1] $PC\_TCHZQ^{*} = F(\text{Intercept, P\_CHZ, INCOMET, REFRIG, SM\_CITY, CITY, HHSIZE, PERLT6, PER6\_11, PERGE66})$

Where $PC\_TCHZQ_i^{*}$ is ***latent*** per capita cheese demand. $PC\_TCHZQ_i^{*} \equiv TCHZQ_i^{*}/HHSIZE_i$ where $TCHZQ_i^{*}$ is ***latent*** total household cheese demand. You hypothesize the following relationship between ***latent*** per capita cheese demand ($PC\_TCHZQ_i^{*}$) and ***observed*** per capita cheese purchases (PC_TCHZQ_i)

[2.2] $$PC\_TCHZQ_i = \begin{cases} PC\_TCHZQ_i^{*} \text{ if } PC\_TCHZQ_i^{*} > 0 \\ 0, \text{ Otherwise} \end{cases}$$

The ***latent*** cheese demand regression equation can be represented as:

[2.3] $PC\_TCHZQ^{*} = X\beta + \varepsilon.$

Use the homoscedastic Tobit MATLAB code we reviewed in class to estimate the model represented by [2.1]-[2.3]. Present the typical regression based statistics. What is the result of your testing the null hypothesis that your Tobit model explains a significant portion of the variability of cheese purchases?

(b) (5 pts) What is the correlation coefficient between observed *positive* amounts of cheese purchased and your predicted *conditional* purchase quantities for these purchasing households?

(c) (10 pts) At the mean of your data, what are your estimates of the purchase probability price elasticity, the conditional quantity purchase price elasticity and unconditional quantity purchase price elasticity? Are these elasticities statistically different from zero?

(d) (5 pts) Is cheese a superior good (e.g., is the unconditional income elasticity > 1.0)?

(e) (15 pts) Undertake a likelihood ratio test of the null hypothesis that the error variance under the above model specification is impacted by household size [HHSIZE] and income [INCOMET] [**Hint**: Use the heteroscedastic error variance

structure we reviewed in class as a basis of this test. Also use the above homoscedastic Tobit results as your starting values for the heteroscedastic specification. You should be aware that using other starting values may result in the model taking a long time to converge, i.e., > 600 iterations. Make sure the likelihood function increases across iterations.].

Table 2.1: Description of a Subset of Variables in 2002 ENIGH Dataset

Variable	Description	Units
Cheese Purchase Characteristics		
TCHZQ	Quantity of cheese purchased	KG
P_CHZ	Cheese Price (≡TCHZX/TCHZQ)	Peso/KG
Household Size/Composition Variables		
PERLT6	Percent of household members < 6 years old	%
PER6_11	Percent of household members between 6 and 11 years of age	%
PERGE66	Percent of household members older than 65	%
HHSIZE	Number of household members	#
Other Household Characteristics		
SM_CITY	Household is located in a town with 2,500-15,000 population	0/1
CITY	Household is located in a town with> 15,000 population	0/1
REFRIG	Does the household own a refrigerator/freezer?	0/1
INCOMET	Quarterly household income	10,000 Pesos

Note: PC_TCHZQ can be calculated as TCHZQ/HHSIZE. For non-purchasing households, P_CHZ was set at the provincial mean level.