Problem Set 4

Discrete Choice Model

AEM 6700, Fall 2025

- This is a **group assignment**, but you are welcomed to solve the problems on your own.
- **You may work in groups of two and submit one copy of report for each group**. If you submit on behalf of your group, make sure to write down the name of your teammate in the report. If you work on your own, just submit your own report.
- Before starting this problem set, please carefully review the slides in Lectures 7 and Lecture 8 as well as the Required Readings.
- You must solve all the questions on Stata (or other software) then submit a **typed report** showing your solutions for each question. Please also attach your codes at the end of the report or submit it as a separate file.
- The deadline for PS 4 is at **11:59 pm EST on Sunday, November 30th**.

1 (30 Points) Simple Logit

The dataset (ketchup) of this assignment is from Andrew Ching, T¨ulin Erdem and Michael Keane," The Price Consideration Model of Brand Choice", Journal of Applied Econometrics, Volume 24, No. 3, 2009, pp. 393-420. There are 31 variables in this file. Their definitions are provided in the codebook.

After importing the ketchup.dta dataset, please open the template Stata .do file and run the codes lines for simple data cleaning and declaring the penal data structure (before Question 1). Answer the following questions.

1. (2 Points) What is the data format (long vs. wide) of this given dataset? Please make sure you convert the data set to long format if necessary (hint: check the command -reshape- in Stata).
2. (5 Points) Estimate a simple logit regression studying the demand of ketchup brands, where consumers' choice to buy a brand (choice) depends on prices, features, whether any of the brands has been displayed, and brand loyalty. Report the results below.
3. (3 Points) Do the estimated coefficients have the expected signs? Are the estimated coefficients significantly different from zero?

   Now let's add alternative-specific constants to the model for alternative brands.

4. (5 Points) Estimate a logit model by adding alternative specific constants. That is to say, the choice of the 5 brands depends on: price, feature, display, brand loyalty, and set Other Brand as the base, assuming no constant is needed in this case. Report the results below *(Hint: please check help file for Stata command - asclogit-).* Which brand is the most preferable?

5. (5 Points) The ratio of coefficients usually provides economically meaningful information. The willingness to pay (WTP) is defined as the price a customer is willing to pay for a product or service. Based on your estimated coefficients in Question 4, What is the estimated WTP for one more unit of "display" from this model? *(Hint: if* $U = \beta_1 p + \beta_2 x_2 + ...$*, then the WTP for one unit of* x_2 *is* $-\frac{\beta_2}{\beta_1}$*, holding others constant.)*

6. (5 Points) Compare your results of Question 4 with Question 2. Is there any advantage in using the model in Question 4?

7. (5 Points) Holding other factors as constant, if retailers choose to promote display Heinz (that is to say, switch from "no display" to "display"), what will be the increase in market share of this brand?

2 (70 Points) Mixed Logit

We will estimate mixed logit models in this problem set. A sample of residential electricity customers were asked a series of choice experiments. In each experiment, four hypothetical electricity suppliers were described. The person was asked which of the four suppliers he/she would choose. As many as 12 experiments were presented to each person. Some people stopped before answering all 12. There are 361 people in the sample, and a total of 4308 experiments. In the experiments, the characteristics of each supplier were stated. The price of the supplier was either:

- a fixed price at a stated cents per kWh, with the price varying over sup- pliers and experiments.
- a time-of-day (TOD) rate under which the price is 11 cents per kWh from 8am to 8pm and 5 cents per kWh from 8pm to 8am. These TOD prices did not vary over suppliers or experiments: whenever the supplier was said to offer TOD, the prices were stated as above.
- a seasonal rate under which the price is 10 cents per kWh in the summer, 8 cents per kWh in the winter, and 6 cents per kWh in the spring and fall. Like TOD rates, these prices did not vary. (Note that the price is for the electricity only, not transmission and distribution, which is supplied by the local regulated utility.)

The length of contract that the supplier offered was also stated, in years (such as 1 year or 5 years.) During this contract period, the supplier guaranteed the prices and the buyer would have to pay a penalty if he/she switched to another supplier. The supplier could

offer no contract in which case either side could stop the agreement at any time. This is recorded as a contract length of 0.

Each customer is identified by the variable customer. For each customer, the variable experiment indicates the experiment (among the 12 ones) he/she participated, the variable occasion indicates a given choice occasion, the variable alter indicates the alternative suppliers, while the dummy variable chosen identifies the stated choice in each choice occasion.

Please see the codebook for variable definitions.

1. (5 Points) Run a standard conditional logit (-clogit- command) model, with 'chosen' as dependent variable, 'price, len, local, wk, TOD, season' as independent variables. Please specify 'occasion' as the group variable. Report your results here and explain them.

2. (15 Points) Run a mixed logit model (using the Stata command -mixlogit-) with independent, normally distributed coefficients. Please use 'chosen' as dependent variable, 'price, len, local, wk, TOD, season' as independent variables whose coefficients are random. Again, please specify 'occasion' as the group variable, and 'customer' as the panel variable (put 'id' (customer) in the options for -mixlogit-).

   (a) (5 Points)Please report your results. How does the value of the log- likelihood function compare for this model relative to the conditional logit model in problem 1?

   (b) (10 Points)Determine the share of the population who are estimated to dislike long term contracts (i.e. have a negative coefficient for the length.) *(Hint: Knowing that the coefficients are distributed normally, you just need to calculate the z-statistic from the length variable, and then evaluate the cumulative distribution function (CDF) from the standard normal distribution.)*

3. (15 Points) The price coefficient is assumed to be normally distributed in these runs. This assumption means that some people are assumed to have positive price coefficients, since the normal distribution has support on both sides of zero.

   Using your estimates from problem 2, determine the share of customers with positive price coefficients. *(Hint: Same method as 2(c). As you can see, this is pretty small share and can probably be ignored. However, in some situations, a normal distribution for the price coefficient will give a really large share with the wrong sign.)*

   Rerun the mixed logit model to make the price coefficient fixed rather than random. Present your results in a table.

4. (15 Points) You think that everyone must like using a well-known company rather than an unknown one, and yet the normal distribution implies that some people dislike using a known company.

Revise the model to give the coefficient of "well-known" a lognormal distribution. You can do this with the price coefficient fixed or normal, whichever you want. Present your results in a table. *(Hint: Pay attention to the ln() option under the -mixlogit- command guide.)*

5. (20 Points) Market share and elasticity

   (a) (10 Points) Under the assumption of logit model, derive: i) alternative i's own-price elasticity of demand; and ii) alternative i's cross-price elasticity with respect to product j. *(Hint: start with the market share equation for alternative i).*

   (b) (5 Points) In the total of 4308 experiments, calculate the proportion of participants choosing each alternative in the experiments (that is, calculate the proportions of alter==1, alter==2, alter==3, and alter==4). Let's assume these proportions equal to the market shares of alt1, alt2, alt3, and alt4, respectively, under the assumption our sample size is large enough and our samples are nationally representative.

   (c) (5 Points) You learned from your colleague that, the average prices of alternatives are $p_1 = \$0.93, p_2 = \$0.87, p_3 = \$0.9, p_4 = \$1.0$, respectively. Based on the equations you derived in part a, calculate alt1's own-price elasticity of demand, and alt1's cross-price elasticity to alt3. Please guess two figures for s_1 and s_3 if you cannot obtain market shares from part b.