

Problem Set 4

Topics in Advanced Econometrics (ResEcon 703)
University of Massachusetts Amherst

Due: November 21, 10:00 am ET

Rules

Email a single .pdf file of your problem set writeup, code, and output to mwoerman@umass.edu by the date and time above. You may work in groups of up to three, and all group members can submit the same code and output; indicate in your writeup who you worked with. You must submit a unique writeup that answers the problems below. You can discuss answers with your fellow group members, but your writeup must be in your own words. Problem 1 allows you to use R's "canned routines," while Problem 2 requires you to code your own estimators and write your own simulation code. Each problem will indicate which R function to use.

Data

Download the file `ps4_dataset.zip` from the course website (github.com/woerman/ResEcon703). This zipped file contains the dataset, `phones.csv`, that you will use for this problem set. The dataset contains simulated data from 1000 customers on the purchases or pre-orders of highly-anticipated phone models recently released by Apple and Google: iPhone 11 and Pixel 4. See the file `data_descriptions.txt` for descriptions of the variables in the dataset.

Problem 1: Mixed Logit Model

- a. We are again interested in understanding how consumers value two phone characteristics: internal storage and screen size. As we saw in the last problem set, consumers appear to have brand preferences, which we are also interested in valuing. Model the purchase of a phone as a mixed logit model. Include an Apple brand dummy, the amount of storage, the screen size, and the price of each phone as explanatory variables with random coefficients. That is, the representative utility for alternative j is

$$V_{nj} = \beta_0 \text{Apple}_j + \beta_1 GB_j + \beta_2 SS_j + \beta_3 p_j$$

where Apple_j is a dummy variable if alternative j is from Apple, GB_j is the internal storage of alternative j , SS_j is the diagonal screen size of alternative j , and p_j is the price of alternative j . Model all four random coefficients as having a normal distribution. Estimate this mixed logit model using the `mlogit()` function in R; use 100 draws for simulation (`R = 100`) and set a seed of 703 for replication (`seed = 703`).

- i. Report your parameter estimates, standard errors, z-stats, and p-values. Briefly interpret these results.

- ii. For each coefficient, calculate the proportion of the population with negative coefficients and the proportion of the population with positive coefficients. Describe whether these coefficient distributions match economic intuition.
- b. The mixed logit model of (a) is not be the best model for this setting if we think some coefficients should always be positive or some coefficient should always be negative. Again model this purchase as a mixed logit model with the same underlying utility model as in (a). That is, the representative utility for alternative j is

$$V_{nj} = \beta_0 Apple_j + \beta_1 GB_j + \beta_2 SS_j + \beta_3 p_j$$

where $Apple_j$ is a dummy variable if alternative j is from Apple, GB_j is the internal storage of alternative j , SS_j is the diagonal screen size of alternative j , and p_j is the price of alternative j . Model β_0 and β_2 as having a normal distribution and β_1 and β_3 as having a log-normal distribution. Estimate this mixed logit model using the `mlogit()` function in R; use 100 draws for simulation ($R = 100$) and set a seed of 703 for replication (`seed = 703`).

- i. Report your parameter estimates, standard errors, z-stats, and p-values. Briefly interpret these results.
 - ii. For each coefficient, calculate the proportion of the population with negative coefficients and the proportion of the population with positive coefficients. Describe whether these coefficient distributions match economic intuition.
- c. We could use the model in (b) to calculate how consumers value the Apple brand, storage, and screen size, but it would require taking a ratio of distributions. To make these calculations easier, we can model price as having a fixed coefficient. Again model this purchase as a mixed logit model with the same underlying utility model as in (a) and (b). That is, the representative utility for alternative j is

$$V_{nj} = \beta_0 Apple_j + \beta_1 GB_j + \beta_2 SS_j + \beta_3 p_j$$

where $Apple_j$ is a dummy variable if alternative j is from Apple, GB_j is the internal storage of alternative j , SS_j is the diagonal screen size of alternative j , and p_j is the price of alternative j . Model β_0 and β_2 as having a normal distribution, β_1 as having a log-normal distribution, and β_3 as a fixed coefficient. Estimate this mixed logit model using the `mlogit()` function in R; use 100 draws for simulation ($R = 100$) and set a seed of 703 for replication (`seed = 703`).

- i. Report your parameter estimates, standard errors, z-stats, and p-values. Briefly interpret these results.
 - ii. Calculate the value consumers place on the Apple brand, each gigabyte of internal storage, and each 0.1 inch of diagonal screen size. Because we have distributions for β_0 , β_1 , and β_2 , these values will also be distributions. Report the parameters that define these distributions. That is, for the value of the Apple brand and the value of each 0.1 inch of diagonal screen size, report the mean and standard deviation of the value; for the value of each gigabyte of internal storage, report the mean and standard deviation of the underlying normal distribution from which the log-normal distribution is derived.
- d. Conduct a likelihood ratio test to compare the models in (b) and (c). Write down the null hypothesis that you are testing and describe this hypothesis in words. Conduct this likelihood ratio test using the function `lrtest()` in R. Do you reject your null hypothesis? What is the p-value of the test?

- e. Using the model in (c), calculate the mean coefficients for purchasers of each of the four Google phones. Use the function `fitted(type = 'parameters')` in R to calculate mean coefficients for each consumer. Report these 12 mean coefficients (3 random coefficients \times 4 Google phones). Describe and briefly interpret the patterns you observe in these coefficients.

Problem 2: Simulation-Based Estimation

- a. Model the purchase of a phone as in (c) of problem 1. That is, the representative utility for alternative j is

$$V_{nj} = \beta_0 \text{Apple}_j + \beta_1 \text{GB}_j + \beta_2 \text{SS}_j + \beta_3 p_j$$

where Apple_j is a dummy variable if alternative j is from Apple, GB_j is the internal storage of alternative j , SS_j is the diagonal screen size of alternative j , and p_j is the price of alternative j . Model β_0 and β_2 as having a normal distribution, β_1 as having a log-normal distribution, and β_3 as a fixed coefficient. Estimate the parameters of this model by maximum simulated likelihood estimation; use 100 draws for your simulation and set a seed of 703 for replication. The following steps can provide a rough guide to creating your own maximum simulated likelihood estimator:

- I. Set a seed of 703 for replication.
- II. Draw 300,000 standard normal random variables (3 random coefficients \times 100 draws \times 1000 consumers).
- III. Create a function to simulate choice probabilities for one consumer:
 - i. The function should take a set of parameters, the random draws for one consumer, and the data for one consumer as inputs: `function(parameters, draws, data)`.
 - ii. Transform the standard normal draws into the correct distributions using the distribution parameters.
 - iii. Calculate the representative utility for each alternative for each draw.
 - iv. Calculate the conditional choice probability for each alternative for each draw.
 - v. Calculate the simulated choice probability for each alternative as the mean over all draws.
- IV. Create a function to calculate simulated log-likelihood:
 - i. The function should take a set of parameters, the random draws for all consumers, and the data for all consumers as inputs: `function(parameters, draws, data)`.
 - ii. Simulate choice probabilities for each alternative for each consumer (call your previous function for each consumer).
 - iii. Sum the log of the simulated choice probability for each consumer's chosen alternative.
 - iv. Return the negative of the log of simulated likelihood.
- V. Maximize the simulated log-likelihood (or minimize its negative) using `optim()`. Use the parameters from (c) in problem 1 as your starting guesses to speed up convergence. Your call of the `optim()` function may look something like:

```
optim(par = c(model_1c$coefficients), fn = your_second_function,
      data = your_data, draws = your_draws,
      method = 'BFGS', hessian = TRUE)
```

Report your parameter estimates, standard errors, z-stats, and p-values. Briefly interpret these results.

- b. Using the model in (a), calculate the mean coefficients for purchasers of each of the four Google phones. Use the same simulation draws as in (a) to simulate the mean coefficients for each consumer, and then average over all purchasers for each Google phone. The following steps can provide a rough guide to simulating coefficients:

I. Create a function to simulate mean coefficients for one consumer:

- i. The function should take a set of parameters, the random draws for one consumer, and the data for one consumer as inputs: `function(parameters, draws, data)`.
- ii. Transform the standard normal draws into the correct distributions using the distribution parameters.
- iii. Calculate the representative utility for each alternative for each draw.
- iv. Calculate the conditional choice probability of the chosen alternative for each draw.
- v. Calculate the weighted average for each coefficient with weights equal to the conditional choice probability of the chosen alternative for that simulation draw.

II. Simulate mean coefficients for each consumer (call your previous function for each consumer).

III. For each of the four Google phones, average the simulated mean coefficients for all purchasers of that phone.

Report these 12 mean coefficients (3 random coefficients \times 4 Google phones). Describe and briefly interpret the patterns you observe in these coefficients.

- c. As in the previous problem set, Apple is still interested in raising the price of the iPhone 11 with 256 GB. Using the model in (a), simulate the elasticity of each other phone with respect to the price of the iPhone 11 with 256 GB. The following steps can provide a rough guide to simulating elasticities:

I. Create a function to simulate elasticities for one consumer:

- i. The function should take a set of parameters, the random draws for one consumer, and the data for one consumer as inputs: `function(parameters, draws, data)`.
- ii. Transform the standard normal draws into the correct distributions using the distribution parameters.
- iii. Calculate the representative utility for each alternative for each draw.
- iv. Calculate the conditional choice probability for each alternative for each draw.
- v. Calculate the simulated choice probability for each alternative as the mean over all draws.
- vi. Calculate the term inside the integral of the elasticity formula for each alternative for each draw by taking products of conditional choice probabilities and the price coefficient.
- vii. Simulate the integral in the elasticity formula by taking the mean of the previous values over all draws for each alternative.
- viii. Calculate the elasticities by multiplying these simulated integrals by the price of the iPhone 11 with 256 GB and dividing by the simulated choice probability of the respective alternative.

II. Simulate elasticities for each consumer (call your previous function for each consumer).

III. Average each simulated elasticity over all consumers.

Report these 10 elasticities. Briefly interpret these results.