Problem Set 3

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

Due: October 31, 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 requires you to code your own estimators, while Problem 2 allows you to use R's "canned routines." Each problem will indicate which R function to use.

Data

Download the file ps3_datasets.zip from the course website (github.com/woerman/ResEcon703). This zipped file contains two datasets, travel_binary.csv and phones.csv, that you will use for this problem set. The travel_binary.csv dataset contains simulated data on the travel mode choice of 1000 UMass graduate students commuting to campus in the middle of winter when only driving a car or taking a bus are feasible options (assume the weather is too severe for even the heartiest graduate students to ride a bike or walk). The phones.csv 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 each dataset.

Problem 1: Generalized Method of Moments

Use the travel_binary.csv dataset for this question.

a. Model the choice to drive to campus during winter as a binary logit model. Include the cost and the time of each alternative as explanatory variables with common coefficients; do not include an intercept. That is, the representative utility for alternative j is simply

$$V_{nj} = \beta_1 C_{nj} + \beta_2 T_{nj}$$

where C_j is the cost of alternative j and T_j is the time of alternative j. Estimate the parameters of this model by method of moments. The following steps can provide a rough guide to creating your own method of moments estimator:

I. Write down moment conditions for this model. In this case, a set of two moment conditions based on residuals and data.

II. Create a function that takes a set of parameters and a *matrix* of data as inputs: `function(parameters, data_matrix)`.

III. Within that function, make the following calculations:

    i. Calculate the representative utility for each alternative and for each decision maker.

    ii. Calculate the choice probability of driving for each decision maker.

    iii. Calculate the econometric residual, or the difference between the outcome and the probability, for each decision maker.

    iv. Calculate each of the M moments for each decision maker.

    v. Return the $N \times M$ matrix of individual moments.

IV. Find the MM estimator using `gmm()`. Your call of the `gmm()` function may look something like:

```
gmm(g = your_function, x = your_data_matrix, t0 = your_starting_guesses,
    vcov = 'iid', control = list(reltol = 1e-25, maxit = 10000))
```

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

Hint: For a binary logit model, characterizing one choice probability is sufficient because the two probabilities must sum to 100%. The probability of driving is

$$P_{nc} = \frac{e^{V_{nc}}}{e^{V_{nc}} + e^{V_{nb}}}$$
$$= \frac{1}{1 + e^{-V_{nc} + V_{nb}}}$$

Both of these expressions for the probability of driving may be useful as you solve this problem.

b. Again model the choice to drive to campus during winter as a binary logit model, but now add an intercept term and alternative-specific coefficients. That is, the representative utilities for the alternatives are

$$V_{nc} = \beta_0 + \beta_1 C_{nc} + \beta_2 T_{nc}$$
$$V_{nb} = \beta_3 C_{nb} + \beta_4 T_{nb}$$

where C_{nj} is the cost of alternative j and T_{nj} is the time of alternative j. Estimate the parameters of this model by method of moments. Follow the same set of steps as in (a), but now you have five parameters and the representative utility calculation is different. Report your parameter estimates, standard errors, t-stats, and p-values. Briefly interpret these results.

c. You might be concerned that the cost and time of driving to campus are endogenous; for example, a student who enjoys driving is more likely to live farther from campus because they do not mind the extra time spent driving. The `travel_binary.csv` dataset includes three possible instruments: gasoline price (the cost of driving is proportional to the gasoline price), snowfall (the time to drive includes the time to clean off the car and the added time from driving cautiously), and an indicator if the student's car is in the shop (extra time or cost is required to borrow someone else's car). Again

model the choice to drive to campus during winter as in (b). That is, the representative utilities for the alternatives are

$$V_{nc} = \beta_0 + \beta_1 C_{nc} + \beta_2 T_{nc}$$
$$V_{nb} = \beta_3 C_{nb} + \beta_4 T_{nb}$$

where C_{nj} is the cost of alternative j and T_{nj} is the time of alternative j. In this model, however, instrument for the cost and time of driving with the instruments described above. Thus, you will have six instruments: constant term, cost to take the bus, time to take the bus, gasoline price, snowfall, and an indicator if car is in the shop. Estimate the parameters of this model by generalized method of moments. Follow the same set of steps as in (a), but now you have five parameters and six moment conditions. Report your parameter estimates, standard errors, t-stats, and p-values. Briefly interpret these results. Also test for overidentifying restrictions in this model and interpret the results of this test.

Problem 2: Generalized Extreme Value Models

Use the `phones.csv` dataset for this question.

a. We are interested in understanding how consumers value two phone characteristics: internal storage and screen size. Model the purchase of a phone as a multinomial logit model. Include the amount of storage, the screen size, and the price of each phone as explanatory variables with common coefficients; do not include intercepts. That is, the representative utility for alternative j is simply

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

where 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. Estimate this multinomial logit model using the `mlogit()` function in R.

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

   ii. Calculate the value consumers place on each gigabyte of internal storage and each 0.1 inch of diagonal screen size.

   iii. Apple sees that the iPhone 11 with 256 GB of internal storage has the greatest market share, so they are considering raising the price. Calculate the elasticity of each other phone with respect to the price of the iPhone 11 with 256 GB. (That is, calculate 10 different elasticities, although some might be equal.) Describe the substitution patterns Apple should expect based on these elasticities.

b. The multinomial logit model of (a) is not be the best model for this setting if we think a consumer's unobserved utility includes brand preference, which would create correlations among phones of the same brand. Model the purchase of a phone as a nested logit model with two nests, one for each brand. Model the purchase of a phone as in (a) with the addition of these nests. That is, the representative utility for alternative j is simply

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

where 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. Estimate this nested logit model using the `mlogit()` function in R.

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

ii. Calculate the value consumers place on each gigabyte of internal storage and each 0.1 inch of diagonal screen size.

iii. Apple is still interested in raising the price of the iPhone 11 with 256 GB. Calculate the elasticity of each other phone with respect to the price of the iPhone 11 with 256 GB. (That is, calculate 10 different elasticities, although some might be equal.) Describe the substitution patterns Apple should expect based on these elasticities. Why are these substitution patterns different from what you found in (a)?

c. Conduct a likelihood ratio test to compare the models in (a) and (b). 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?

d. Consumer preferences for a specific phone model may be stronger than brand preferences. We can nest by each phone model (i.e., Google Pixel 4, Google Pixel 4 XL, iPhone 11, etc.) to incorporate these phone model preferences into our nested logit model. Model the purchase of a phone as a nested logit model with five nests, one for each phone model. Model the purchase of a phone as in (a) with the addition of these nests. That is, the representative utility for alternative j is simply

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

where 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. Estimate this nested logit model using the `mlogit()` function in R.

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

ii. Calculate the value consumers place on each gigabyte of internal storage and each 0.1 inch of diagonal screen size.

iii. Apple is still interested in raising the price of the iPhone 11 with 256 GB. Calculate the elasticity of each other phone with respect to the price of the iPhone 11 with 256 GB. (That is, calculate 10 different elasticities, although some might be equal.) Describe the substitution patterns Apple should expect based on these elasticities. Why are these substitution patterns different from what you found in (a) and (b)?

e. Suppose you want to compare the models in (b) and (d). Can you use a likelihood ratio test to compare these models? If so, write down the null hypothesis of your test, conduct the test using the function `lrtest()` in R, and describe your conclusion. If not, explain specifically why the likelihood ratio test is not appropriate for this model comparison.