Winter 2021
BUS 37904
Professor Jean-Pierre Dubé
Professor Günter J. Hitsch

Advanced Quantitative Marketing: Demand Estimation Assignment

<u>Overview:</u>

In the spirit of Guadagni and Little (1983), you will use household-level panel data to study brand choice behavior using the conditional logit. You will conduct a case study of peanut butter purchases for a sample of 544 households who shop primarily in the two largest supermarkets in Denver between January 1993 and March 1995. The demand estimation can be carried out using the R package *bayesm*. Of course, you are also welcome to write your own source code from scratch!

For each of the three modules (i) *Microeconometric Models of Consumer Demand,* (ii) *Heterogeneity in Preferences* and (iii) *Brands and Branding*, you will have a set of assigned tasks to complete using these data. By the end of week 10, you will need to submit your results from these tasks. Details for each assignment are provided below.

<u>Data</u>

Please use the dataset pbout.txt. The data consist of household panel data from the Nielsen Homescan database for the Denver Scantrack between 1993 and 1995. Retained households had to have purchased peanut butter in one of the top two chains at least once during the sample period. The eight top-selling peanut butter products are included in the sample. There is also a ninth "no-purchase" option. The data contain a separate observation for each of the choice alternatives on each of the unique shopping trips in the sample. Therefore, each trip should have 9 observations associated with it. Only one of the 9 alternatives is selected on each trip. The data contain the following variables:

- Panelid: a unique identifier for each household
- Date: the date of a shopping trip
- Choice: an indicator for which of the 9 alternatives was chosen on the trip (choice 9 is the no-purchase option)
- Price: the shelf price for each alternative
- Feature: indicator for whether that alternative was "featured" in the newspaper insert that week

- Display: indicator for whether that alternative had an in-aisle display that week
- Loyalty: indicator for whether that alternative was the previous one purchased

The data are already sorted so that, for each trip, the first observation pertains to alternative 1, the second to alternative 2, …, and the ninth to alternative 9 (the no-purchase choice).

Assignments:

Microeconometric Models of Consumer Demand:

Estimate the homogeneous, conditional multinomial logit demand system using MCMC. To conduct Bayesian inference, use the Independence Metropolis MCMC algorithm described in the Rossi et al (2005) textbook. This can be done using the function *rmnlIndepMetrop* in the *Bayesm* library in R.

Tasks:

- Load the data from "pbout.txt" in R
- Compute what you consider the relevant descriptive statistics by product alternative and report them in a table.
- Estimate the following multinomial logit specifications:
    1. Price only
    2. Alternative-specific dummy variables & price
    3. Alternative-specific dummy variables & price & promotions
- For each specification, report the posterior mean and 95% credibility interval for each coefficient. How sensitive are your estimates to the prior settings? Across the three specifications, what changes do you notice and what is your intuition for those changes?
- Using the function *logMargDenNR* in *Bayesm,* compute the posterior log-likelihood for each specification via the Netwon-Raferty approach. Report this in the table along with your coefficient estimates.
- Using statistical decision theory, select which of the specifications you think fits the data best. Provide intuition about what is driving the relative fits of the different specifications. Experiment with dropping the upper and lower 2.5^{th} percent of draws to avoid underflow problems when computing the posterior log-likelihood. Does this change your conclusions about relative model fit?
- Suppose you are the retailer, seeing these types of estimates for the first time. What types of "managerial" conclusions might you draw? Be a little creative in terms of your exploration of the marketing implications.
- Extra: re-run the estimation of the 3 models above using maximum likelihood. You will need to write a short script to compute the logit likelihood. You can then use the *optim* function in R to maximize the likelihood. How do the MLE estimates compare to the Bayesian estimates?

Heterogeneity in Preferences:

To control for between-household, persistent heterogeneity in tastes, estimate the random coefficients multinomial conditional logit demand system using MCMC. To conduct Bayesian inferences, use the Random Walk Metropolis MCMC algorithm discussed in Rossi et al (2005). This can be done using the function *rhierMnlRwMixture* in *Bayesm*. You will need to compare parametric and non-parametric specifications of the distribution of heterogeneity: the Normal and Mixture-of-Normals specifications, respectively.

- Re-estimate the same specifications from week 1 using each of 1, 2, 3, 4 and 5-component Mixture-of-Normals specifications.
- Compute the posterior log-likelihood for each specification and report in a table along with the model fits for the 3 specifications from the *Microeconometric Models* assignment. Is underflow affecting your conclusions? Experiment with trimming the set of draws to check.
- Using statistical decision theory, select the specification you think fits the data best. Provide intuition about what is driving the relative fits of the different specifications, including the models with and without persistent heterogeneity in tastes.
- Take the best-fitting model and solve the retailer's decision-theoretic pricing problem using the posterior total expected peanut butter category profits as the "*gain*" function. To make your life easier, assume each product's marginal cost (i.e., wholesale price charged to retailer) is 70% of its average in-sample price. This is not a good assumption in practice but it will be fine for this assignment.
    - What are the retailer's expected optimal prices? What are the expected category profits using the optimized prices versus the in-sample average prices for each product? Plot a histogram of the posterior distribution of the difference in profits when the firm uses the optimized prices versus the average observed price for each product. What do you conclude?
    - What is the posterior probability that profits will increase if the retailer adopts your recommended prices?
- Suppose you are the retailer, seeing these new heterogeneous estimates in addition to the homogeneous estimates. How (if at all) would you change your "managerial" conclusions? Again, be a little creative in terms of your exploration of the marketing implications.

Brands and Branding:

Now we want to investigate the potential role of brand loyalty. Recall from class that a long literature dating back to at least the 1950s has studied patterns of brand choice inertia. In industry, firms often attribute high repeat-purchase rates by customers as a testament to the marketing team's branding efforts. But, this could also arise if there are frictions (information, genuine switching costs etc) that make it difficult for customers to switch brands over time irrespective of the brand value. We want to test for such inertia in our peanut butter data.

- Re-estimate your preferred specification from week 1 including the loyalty variable.
- Re-estimate your preferred specification from week 2 including the loyalty variable.
- Compare your loyalty estimates with versus without controls for heterogeneity. Explain and similarities and/or differences in your results.
- In a table, compare the posterior fit of the 4 specifications.
- What conclusions do you draw?

Bonus: Suppose you are the category manager in charge of optimizing the profitability of the peanut butter category. Suppose also that your weekly discount factor is 0.99. To make this simpler, dispense with the Bayesian decision theory. Instead, compute the prices that maximize the net present value of category profits evaluated at the posterior means of each of the parameters. That is, *plug in* the posterior means instead of integrating profits over the posterior distribution of the demand parameters. This will make your life easier but it will bias the pricing calculations. You are of course welcome to solve for the prices that do optimize the posterior expected net present value of the category profits if you feel up to the task!