

Summary

27 September, 2019

1 - Homogeneous Logit

About the construction of data

First, let's define a brand. I've assigned each UPC into a ketchup brand according to the description variable. Based on my understanding of kchp_mksh1.doc there are only 6 brands in market 1. Those are:

- HEINZ
- HUNT'S
- DEL MONTE
- CTL BR
- GENERIC

I identified the brand of each UPC by checking whether the desc variable in the UPC file contained one of the words in the list above. To simplify matters, I considered the CTL BR and the GENERIC to be the store brand STORE.

Second, let's define a purchase occasion by the day-store. Therefore, if in the household purchase data we observe a household purchasing ketchup more than once in the same store-day I keep only the first purchase.

Third, we need to obtain a price that varies by the household, the brand, and the purchase occasion. Note that households don't get utility from the quantity (weight) of ketchup that they consume, and they dislike paying higher a price. As products have a different size, and fixing a brand, the higher the quantity the higher the price (and presumably the lower the price per ounce) we have to normalize the price paid to represent always the same quantity.

1. I merge the UPC file and the price file, and I construct a panel that contains the price (dpr) for each UPC on each store-day and the weight of each UPC in ounces, that is $weight = \frac{wamt}{1000}$.
2. For each store-day-UPC I compute the price per standarized quantity, which I am going to take as 32 ounces, as it is the most purchased size in the data. That is I compute price as $32 * \frac{dpr}{weight}$.
3. Now, to aggregate from the store-day-UPC to the brand-day-UPC I take the weighted average of the price where the weights vary by the UPC and is given by the total number of standarized units (where the standard is again 32 ounces) sold in the entire panel.
4. At this point, we have a price index that makes sense but it does not vary by the household. To achieve that I will use the data on coupons, and I will assume that if any a house only has coupons on the product it purchases. The final price will be equal to the one computed in step 3. minus the total standarized coupons that the household had for that brand the day of the purchase, were we allow for both store and manufacturer cooupons. An example to illustrate this is a good idea at this point. Let's say that a household on a given day-store had a \$1 coupon when buying a \$2 bottle of ketchup of 32 ounces of a given brand. Then the price faced by the household is \$1. If instead the household bought a 44 ounces bottle of ketchup that costs \$3 of the same brand, then the standarized price of the bottle would be $32 * (\frac{3}{44}) = 2.18$ and the standarized value of the coupons would be $32 * \frac{1}{44} = 0.73$, so the price faced by the household would be \$1.45.

Fourth, we construct a panel for analysis as follows. For each household-day-store (which we refer hereafter as a shopping trip) we have 4 observations, one corresponding to each brand. For each shopping trip and brand we also have the corresponding price (as computed by the explanation above). If for any of the 4 brands in any given shopping trip a price is not available, then we drop the whole shopping trip (the 4 observations, one for each brand). Finally, we add a dummy variable that indicates which of the four brands was chosen for each shopping trip.

(a)