Model Writeup

Mallick Hossain

1   Introduction

This document estimates a variety of multinomial discrete choice models for toilet paper selection. The goal is to provide baseline estimates of consumer preferences as well as to shed light on the degree of consumer heterogeneity that exists in purchasing behavior. Initial data analysis has revealed a persistent positive correlation between package size and income, even after controlling for a wide array of household characteristics that capture differences in consumption. These models will help test the hypothesis of whether there is a deep relationship between income and purchasing preferences and the extent of that relationship. Secondary objectives are to more fully explore whether demographic factors can explain the heterogeneity we see in consumer purchasing or if there remains a large degree of unobserved consumer heterogeneity.

2   Model

I will estimate a mixed logit (i.e. random coefficients) model to capture the consumer choice decision.

The setting is as follows. The decision-maker (a household) makes a shopping trip to a store in which they purchase toilet paper.[1] The decision I model is conditional on this being a trip in which they purchase toilet paper, which product do they choose. With this setting, their choice set is the set of products with positive sales at the store they visit during the week of their visit. This is obtained by matching consumer panel data with retail scanner data at the store-week level.

A choice alternative is a brand-size combination. While there are over 200 different brands and 26 different sizes, I focus on the top 5 brands (Angel Soft, Charmin, Cottonelle, Quilted Northern, and Scott) along with an "Other"

---

[1] In my model, households only purchase 1 package per trip, which is true for over 80% of shopping trips.

category that captures other brands as well as private labels.[2] These top 5 brands capture about 74% of purchases, which I verify is also reported in Orhun and Palazzolo (2019). Private label brands account for another 18% of purchases. Furthermore, I restrict sizes to be 4, 6, 9, 12, 18, 24, 30, and 36 packs, which cover about 91% of purchases (also reported in Orhun and Palazzolo (2019)). The cross of these two (top brand and top size) covers 80% of all purchases.

Brand	Share
CHARMIN	0.21
CTL BR	0.18
ANGEL SOFT	0.15
QUILTED NORTHERN	0.13
KLEENEX COTTONELLE	0.11
SCOTT 1000	0.08
Total	**0.87**

Size	Share
12	0.38
4	0.18
6	0.11
24	0.09
36	0.04
9	0.04
30	0.04
18	0.02
Total	**0.91**

Furthermore, I explicitly model this as a static problem. One way I incorporate price changes is that item prices are inclusive of any sales that may apply to the item. I do not model price expectations or sale anticipation, but I do account for any price changes that are available at the time of purchase.[3] While

---

[2]Kimberly-Clark owns both Scott and Cottonelle brands.

[3]Over 65% of purchases are at full price, so at least a majority of households do not appear to be "waiting" along this dimension, especially when considering that an overwhelming

inventory behavior may be another concern since this is a storable product, that primarily would affect the timing of purchases, but conditional on this being a purchase trip, my model will capture how this decision may depend on size. Those that are more likely to stockpile would likely increase the estimated taste for package size. Additionally, purchase patterns tend to be relatively consistent with households sticking with their brand-size combination with limited deviations in size, and even less frequent deviations between brands.

2.1 Utility

Households choose the product that maximizes their utility, which I represent as the following.

$$U_{njt} = X_{njt}\beta_n + \epsilon_{njt} \tag{1}$$

where U_{njt} represents the utility that household n gets from product j on trip t. X_{njt} is a matrix of product characteristics, namely package size, price, and brand. ϵ_{njt} is an error term drawn from a Type I Extreme Value distribution. β_n varies by household, but we assume that each household's β comes from a normal distribution, except for the price coefficient, which we assume is drawn from a lognormal distribution since this is likely to be negative for all consumers.

This model is estimated by Maximum Simulated Likelihood because there is not a closed form of the likelihood.

In my setting, a product is defined as a brand-size combination.[4] The relevant product characteristics in my setting are package size, package cost, and ply. In some specifications, I also include brand as a rough proxy for quality.

share of shopping trips do not include a purchase of toilet paper, so there are multiple opportunities to "wait" for the right time. It does not appear that this is of primary importance. Furthermore, that would only affect the timing of the purchase decision and my goal is to model the actual choice conditional on a purchase.

[4] I reduce the brand space by concentrating on Charmin, Quilted Northern, Angel Soft, Kleenex Cottonelle, and Scott. Brands outside of this set (including private labels) are grouped as "Other". These top 5 brands account for about 75% of sales. I also segment the package size space to cover the most popular sizes: 1-4, 5-6, 7-9, 10-12, 13-20, 21-24, 25+. The common sizes are the right end points of these bins. With the 30 and 36-packs lumped in the 25+ category, these sizes account for about 91% of purchases.

I start by estimating a conditional logit model and a random coefficients model without including any demographic information, but leveraging the panel structure of the data. The results are in the table below:

3 Data Details

In order to conduct this analysis, we need a dataset that captures consumer decisions and their available choices. I create the necessary dataset by combining two Nielsen datasets: the Consumer Panel dataset and the Retail Scanner dataset. Nielsen's Consumer Panel dataset contains information on consumer purchases as well as a rich set of demographic variables. Nielsen's Retail Scanner dataset contains information on the weekly volumes of products sold at stores that have agreed to share their data with Nielsen.

3.1 Consumer Panel Data

Households In order to best focus on the consumer choice decision, I remove any households where the head of household is a student or a member of the military because these households likely have different living arrangements that are not representative of a typical household's decision to purchase toilet paper (e.g. dormitories and barracks). This reduces my sample from about 653,554 households to 644,229 households. I also remove 7 households that have missing residence information.

Products On the products side, there are some products that are obviously miscoded. For example one toilet paper product is reported to have "multi" and "size1_amount" both equal to 36, indicating total package of 1,296 rolls. In reality, only one of those fields was supposed to be 36. I manually correct these discrepancies based on the corresponding product in the Retail Scanner data, because it is likely to have less error. In some cases, I use my best judgement based on similar items offered by the same brand. Overall, 80 products are corrected out of 7,453 total products. I also remove 6 products branded as "to-go" packs since these are unlikely to be used for daily household consumption.

Missing Purchases I then follow the process of Orhun and Palazzolo (2019) to identify missing purchase occasions where a household likely made a purchase, but did not report it. A missing purchase would downward bias my calculation

of a household's consumption rate. For example, if a household purchased a 4 pack each month for 3 months, but only recorded the first and last occasion, then their daily consumption would be $8/3 \approx 2.67$ rolls/month instead of the true rate of 4 rolls per month. Given a package size s, I compute the mean and standard deviation of the time until the next purchase. If, the time between purchases given a size s package is longer than the average duration plus 2 standard deviations, then I flag this as a missing purchase occasion.

These "missing" purchases are then used to demarcate when a household is "active", defined as a spell without a missing purchase. Hence, if a household had 1 missing purchase, then they are determined to have 2 active periods. I use these active periods to compute a household's average consumption rate, which is as follows (as per Orhun and Palazzolo (2019)):

$$Consumption_h = \frac{\sum_{a=1}^{A} \sum_{p=1}^{P_a-1} V_{hpa}}{\sum_{a=1}^{a=A} T_a} \qquad (2)$$

where V_{hpa} is the volume of toilet paper for purchase p during active period a. P_a is the total number of purchases made during active period a. All purchases during an active period are included except for the last one because we assume that the purchase made on the final day of an active period is not consumed during that period.

Outliers Even after correcting miscoded products and identifying potential missing purchases, there remain some outliers. In particular, some households purchase excessively large quantities of items, which may indicate they are purchasing for a small business. In other instances, they are inactive, have many missing purchases, or extremely low consumption rates. For my analysis, I remove these outliers, which meet any of the following criteria:

- Excessive Missingness: The household has more than 3 missing purchases and/or has an inter-purchase duration longer than the 99th percentile of all household's maximum inter-purchase duration.

- Inactivity: The household is active for less than 90 days.

- Missing/Insufficient Consumption: Household consumption is below the 1st percentile of the distribution or consumption cannot be calculated.

- Extreme Values: The household purchased a quantity or volume higher than the 99th percentile of the quantity or volume distribution or they

spent more than $50 (inflation-adjusted to 2012) on a single purchase. I also remove any purchases for which a price of $0 is recorded.

As a result of these conditions, 6.5% of purchases and 28.8% of households are dropped. These exclusion criteria primarily exclude households that make few purchases in this category. Below are tables denoting how many observations and households were affected by each condition as well as how the distribution of the sample compares before and after applying these criteria. Overall, these figures mirror those of Orhun and Palazzolo (2019) with the main differences being a result of including the 2015 and 2016 data.

Criteria	Obs	Obs %	HH	HH %
Missing 3+ IPD:	171,042	4	6,899	4.6
Max IPD > 99th Pct:	22,053	0.5	1,493	1
Cannot calc. consumption:	12,366	0.3	9,031	6.1
Insufficient Consumption:	7,882	0	1,403	0
Active < 90 days:	27,882	0.7	14,247	9.5
Abnormal Quantity:	36,579	0.9	12,898	8.6
Abnormal Volume:	24,727	0.6	11,758	7.9
Abnormal Price:	6,969	0.2	4,642	3.1
Total HH/Observations:	4,241,992	-	149,272	-
Total Dropped:	277,193	6.5	43,021	28.8

ptile	consRaw	consClean	IPDRaw	IPDClean	volRaw	volClean
1st pct	0.04	0.05	0	0	1.82	1.82
25th pct	0.15	0.16	14	14	5.82	5.82
50th pct	0.24	0.23	28	28	9.60	9.09
75th pct	0.35	0.35	59	56	16.58	15.71
99th pct	1.04	0.91	382	300	65.45	55.64
N	149272.00	126049.00	4092720	3840729	4187109.00	3915409.00

3.2 Nielsen Retail Scanner Data

Assumption 1: *If a product has no sales, it is not in the consumer's choice set. If a product has positive sales, it is in the consumer's full choice set.*

In the data, only products with positive sales are recorded, therefore it is difficult to determine if an available product did not sell (a true zero) or if the product was not available (a missing product). Because the data does not distinguish between missing items and those with no sales, I assume that the selection available to consumers is only the set of items with positive sales during the week they shopped at the store.

For some models, it is important to observe multiple purchases by the same household. In those cases, the sample size may be dramatically reduced for various reasons. The data slims down in the following fashion:

Purchases	Households	Notes
3 million	135,000	Full 2006-2016 Consumer Panel
770,000	78,000	Linking Panel with Scanner
165,000	12,000	Restricting to households with 10-20 trips

This sample is substantially larger than typical studies and the reductions are primarily due to data limitations. As a check, I can look at the balance on observables on this smaller subsample and compare it to the larger sample. Even without that, Nielsen provides projection weights that help adjust for various selection issues and properly weight households to be projectable to the full United States.

Furthermore, another strategy that could be leveraged to address concerns about sample size would be to expand the data by adding another (relatively weak) assumption:

Assumption 2: *Stores in the same retail chain offer the same assortment.*

Given that pricing and inventory policies are often set at a regional or national level, this is likely a weak assumption. Furthermore, the results of DellaVigna and Gentzkow (2017) suggest that pricing is nearly uniform across stores within the same retail chain.

References

Orhun, A. Yeşim, and Michael Palazzolo. 2019. "Frugality Is Hard to Afford." *Journal of Marketing Research* 56 (1). American Marketing Association.