

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

2.1 Multinomial Logit

I start by estimating a basic multinomial logit model that represents a consumer's purchase decision. Each product is characterized by its ply, package size, standardized rolls per package, unit price, and brand.¹ I aggregate packages into *small* (1-6 rolls), *medium* (7-12 rolls), and *large* (12+ rolls).² I estimate this model where the consumer can choose from one of the top 5

¹I draw a distinction between package size and standardized rolls per package to better isolate how long a particular package may last. A 4-roll package is likely to take up about the same amount of space regardless of ply or sheet count, but may last for longer or shorter durations depending on those characteristics. Going back to at least 1915, this has been a source of product obfuscation. See 1915 Scott ad included in Appendix Figure 1

²While there are a range of sizes offered, most packages are 4-, 6-, 12-, or 24-roll packages. This categorization captures the main package size variation while also incorporating other, less popular sizes. It also captures the main focus of package size, namely it serving as a proxy for storage and transportation costs.

brands (Angel Soft, Charmin, Cottonelle, Quilted Northern, and Scott) or a private-label³ and 3 different sizes (small, medium, or large). This generates a total of 18 brand-sizes.

The setting is as follows. A household makes a shopping trip to a store in which they purchase toilet paper. In over 85% of shopping trips, households only purchase 1 package, so a discrete choice model is appropriate for this setting.⁴ The household’s choice set consists of all 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.

I only model the product choice decision conditional on purchasing toilet paper. Since toilet paper is storable, one might want to also model the purchase/no-purchase decision allowing for an underlying inventory state. However, toilet paper tends to go on sale relatively infrequently and a majority of purchases are made at full price. This consideration would imply that my price elasticity estimates are a lower bound since the purchase/no-purchase decision would relax some of the constraint associated with having to make a purchase.

2.1.1 Utility

Households choose the product that maximizes their utility, which I represent as the following (time and market subscripts are suppressed for clarity; notation follows Cohen (2008)).

$$U_{ij} = \phi pkgSize_j + \gamma size_j + \alpha p_j + \mu_{ij} \quad (1)$$

where U_{ij} represents the utility that household i gets from product j . $pkgSize$ is a categorical variable for small, medium, or large packages; $size$ is the standardized rolls in a package; and p is the per-standardized-roll cost of package j . Letting $\mu_{ij} = \epsilon_{ij}$ where ϵ is drawn from a Type I Extreme Value distribution gives us the standard multinomial logit model. The choice probabilities of can be computed from the following expression:

³Private label also includes the long tail of other, less popular brands.

⁴Households purchasing 2 packages make up the majority of remaining purchases. This could be addressed in one of two ways. We could modify the choice set to include 2-package products. The other option is to treat my estimates as an upper bound on price sensitivity. Because of bulk discounts, purchasing multiple packages is weakly more expensive than the equivalent larger package. One other fact is that in 98% of trips, a single UPC is purchased.

$$P_j = \frac{\exp(\delta_j)}{\sum_{h=1}^J \exp(\delta_h)}, \quad (2)$$

where $\delta_j = \beta ply_j + \phi pkgSize_j + \gamma size_j + \alpha p_j$, which can be thought of as the average population utility for product j . The parameters of the utility function can be estimated using MLE and the estimated coefficients represent the average population preferences for each product characteristic.

To streamline computation, I focus on Boston in 2014-2016, but hope to expand this sample more widely.

Table 1:

	<i>Dependent variable:</i>		
	choice		
	(1)	(2)	(3)
Price	−0.06*** (0.004)	−0.06*** (0.01)	−0.05*** (0.01)
Pack Size		0.003 (0.01)	0.01** (0.01)
Day Supply			−0.004*** (0.001)
Charmin	0.84*** (0.05)	0.85*** (0.05)	0.78*** (0.05)
Cottonelle	0.74*** (0.05)	0.75*** (0.05)	0.62*** (0.06)
Private Label	0.23*** (0.04)	0.23*** (0.04)	0.24*** (0.04)
Qltd Ntn	0.71*** (0.06)	0.71*** (0.06)	0.67*** (0.06)
Scott	−0.41*** (0.05)	−0.39*** (0.06)	−0.31*** (0.06)
Observations	6,187	6,187	6,187
Log Likelihood	−14,109.01	−14,108.81	−14,086.52

Note:

*p<0.1; **p<0.05; ***p<0.01

Pack size is defined number of rolls in a package. Days' supply is the number of standardized rolls in a package divided by a household's average daily consumption.

The coefficients fit relatively well with intuition. The price coefficient is significantly negative as expected. Angel Soft is the reference brand, so the brand fixed effects indicate that only Scott is less likely to be bought, but all other brands, including private labels are more likely to be bought over Angel Soft. An increased package size slightly increases the likelihood of purchase while the days' supply offered by the package (which I define as the standardized

package divided by household’s average daily consumption) slightly reduces the probability of purchase.

To translate this into dollar terms, we can compute the willingness to pay by dividing each coefficient by the price coefficient. This implies that households are willing to pay an extra \$0.68 for a package with an extra unit, but they would have to be paid an extra \$0.11 to purchase an additional days’ supply of toilet paper. Since days’ supply is a function of package size, this implies an interesting trade off. For high consumption households, they prefer larger package sizes, but for low consumption households, they prefer smaller package sizes. The willingness to pay for particular brands does give large estimates which outstrip the usual price differences seen in a store (WTP ranges from +\$6.79 for Cottonelle to -\$3.84 for Scott).

Furthermore, while this model is a good starting point, a pure multinomial logit model implies a strict substitution pattern as a result of the independence from irrelevant alternatives property. This implies that if a new product is introduced, whether a Charmin 36-pack or a private-label single roll, it will draw from other alternatives proportionately. This is at odds with what we would expect given that a Charmin 36-pack is more suited for customers that prefer Charmin or large packages, but it is unlikely to compete with, say, a Scott 4-pack.

2.2 Conditional Logit

The above model is quite restrictive along a variety of dimensions. Primarily, it does not incorporate any observed heterogeneity that is available in the data (any heterogeneity is only incorporated through ϵ). Furthermore, it generates restrictive substitution patterns as a result of the Independence of Irrelevant Alternatives (IIA) property, which means that when the price of one item increases, consumers will substitute proportionally to other items, which results in them substituting to the most popular items instead of substituting to the items most similar to the one they were considering. I address the first concern by directly incorporating demographic characteristics. In particular, it is intuitive that price sensitivity may be related to a household’s income. Second, a household’s sensitivity to package size may be related to the type of home they have. Finally, the preference for the standardized amount may be related to a household’s consumption rate. In order to incorporate each of these features, I interact the relevant product characteristic with the corresponding household characteristic and re-estimate the model.

Table 2:

	<i>Dependent variable:</i>		
	choice		
	(1)	(2)	(3)
p	−0.12*** (0.01)	−0.12*** (0.01)	−0.13*** (0.01)
brandsCharmin	1.43*** (0.11)	1.31*** (0.15)	1.28*** (0.16)
brandsCottonelle	2.43*** (0.13)	2.77*** (0.18)	2.75*** (0.19)
brandsPrv_Lbl	−0.56*** (0.09)	0.48*** (0.13)	0.44*** (0.14)
brandsQltd_Ntn	1.28*** (0.14)	0.99*** (0.21)	0.96*** (0.21)
brandsScott	−1.13*** (0.12)	0.06 (0.17)	0.05 (0.17)
pkgSize	−0.07*** (0.01)	−0.08*** (0.01)	−0.08*** (0.01)
daysSupply	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
rate:Angel_Soft24	0.49 (0.60)	1.53 (1.36)	1.69 (1.35)
rate:Angel_Soft4	−0.65*** (0.20)	0.16 (0.35)	−0.06 (0.36)
rate:Angel_Soft6	−2.34*** (0.43)	−0.36 (0.76)	−0.36 (0.75)
rate:Charmin12	−0.29 (0.29)	−0.10 (0.38)	−0.14 (0.37)
rate:Charmin24	1.04*** (0.34)	0.05 (0.54)	0.004 (0.53)
rate:Charmin4	−5.63*** (0.43)	−3.60*** (0.53)	−3.60*** (0.55)
rate:Charmin6	−1.15*** (0.31)	−0.42 (0.38)	−0.42 (0.38)
rate:Cottonelle12	−4.14*** (0.38)	−4.46*** (0.45)	−4.56*** (0.46)
rate:Cottonelle24	−3.44*** (0.67)	−5.63*** (1.31)	−5.84*** (1.34)
rate:Cottonelle4	−25.88*** (3.64)	−28.21*** (6.34)	−32.47*** (7.37)
rate:Cottonelle6	−11.86*** (0.88)	−6.56*** (1.29)	−6.52*** (1.28)
rate:Prv_Lbl12	2.22*** (0.24)	3.29*** (0.31)	3.29*** (0.31)
rate:Prv_Lbl24	3.89*** (0.25)	4.53*** (0.32)	4.45*** (0.32)
rate:Prv_Lbl4	1.78*** (0.24)	1.99*** (0.30)	1.99*** (0.30)
rate:Prv_Lbl6	−4.18*** (0.44)	−3.46*** (0.66)	−3.41*** (0.61)
rate:Qltd_Ntn12	−0.12 (0.35)	−0.001 (0.41)	0.0001 (0.41)
rate:Qltd_Ntn24	9.27*** (1.53)	14.90*** (4.76)	15.89*** (4.87)
rate:Qltd_Ntn4	−10.92*** (1.49)	−6.43** (2.63)	−6.96** (2.72)
rate:Qltd_Ntn6	−6.84*** (0.73)	−0.88 (1.02)	−1.17 (1.05)
rate:Scott12	2.76*** (0.26)	3.19*** (0.33)	3.16*** (0.32)
rate:Scott24	6.36*** (0.38)	6.61*** (0.64)	6.41*** (0.68)
rate:Scott4	−0.32 (0.33)	0.62 (0.40)	0.62 (0.39)
rate:Scott6	1.50* (0.85)	4.69** (1.86)	3.70* (1.94)
household_income_cts:Angel_Soft24		−0.01 (0.01)	−0.02* (0.01)
household_income_cts:Angel_Soft4		−0.01*** (0.002)	0.002 (0.002)
household_income_cts:Angel_Soft6		−0.01*** (0.004)	−0.01*** (0.004)
household_income_cts:Charmin12		0.001 (0.002)	0.0001 (0.002)
household_income_cts:Charmin24	5	0.01*** (0.003)	0.01** (0.003)
household_income_cts:Charmin4		−0.01*** (0.003)	−0.01* (0.003)
household_income_cts:Charmin6		−0.002 (0.002)	−0.001 (0.002)
household_income_cts:Cottonelle12		−0.005** (0.002)	−0.003 (0.002)
household_income_cts:Cottonelle24		0.01 (0.005)	0.01 (0.01)
household_income_cts:Cottonelle4		−0.0004 (0.01)	0.02 (0.02)

2.3 Random Coefficients

While the conditional logit model incorporates observed household heterogeneity, it still maintains the restrictive substitution patterns that result from the IIA property. In order to allow for flexible substitution patterns and unobserved household heterogeneity, I estimate a random coefficients model, while maintaining the observed household characteristics that I included in the conditional logit model. The random coefficients model is represented below (suppressing the)

$$U_{njt} = X_{njt}\beta_n + \epsilon_{njt} \quad (3)$$

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, ply, 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.⁵ 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.

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:

⁵I 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.

3 References

- Cohen, Andrew. 2008. “Package Size and Price Discrimination in the Paper Towel Market.” *International Journal of Industrial Organization* 26 (2). Elsevier: 502–16.
- Orhun, A. Yeşim, and Michael Palazzolo. 2019. “Frugality Is Hard to Afford.” *Journal of Marketing Research* 56 (1). American Marketing Association.

4 Data Appendix

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.

4.1 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)):

$$Consumption_h = \frac{\sum_{a=1}^A \sum_{p=1}^{P_a-1} V_{hpa}}{\sum_{a=1}^A T_a} \quad (4)$$

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:	168,106	4	6,899	4.6
Max IPD > 99th Pct:	21,600	0.5	1,493	1
Cannot calc. consumption:	12,364	0.3	9,031	6.1
Insufficient Consumption:	7,843	0	1,403	0
Active < 90 days:	27,475	0.6	14,247	9.5
Abnormal Quantity:	2,088	0	1,189	0.8
Abnormal Volume:	1,913	0	1,307	0.9
Abnormal Price:	6,956	0.2	4,642	3.1
Total HH/Observations:	4,241,992	-	149,272	-
Total Dropped:	520,042	12.3	28,345	19

pctile	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	0.23	0.23	28	28	9.60	9.09
75th pct	0.34	0.34	59	56	16.58	15.71
99th pct	1.02	0.89	382	305	65.45	55.64
N	149,272.00	120,927.00	4,092,720	3,589,017	4,187,109.00	3,665,655.00

Finally, for my discrete choice modeling, I restrict the sample to households that purchased a single package. This leaves me with a sample of 3,813,591

purchases across 126,034 households.

4.2 Nielsen Retail Scanner Data

The Nielsen Retail Scanner Data contains weekly pricing, volume, and promotional activity of about 35,000 stores from about 90 retail chains and covers 2006-2016. Each store in the panel reports their weekly quantity sold as well as the weighted average unit price. For a subset of stores, they also report their promotional activity (i.e. if a product was featured or on display).

Because my focus is on the consumer's choice problem, I only extract information related to a consumer's choice set for toilet paper. In order to do this, I combine all movement files of the toilet paper category and then merge those with the Nielsen Consumer Panel data based on the store ID and the week of the shopping trip. This match reduces the possible size of the Consumer Panel because it requires that households shop at the exact stores that have agreed to report their weekly sales to Nielsen.

In examining the consumer choice, I have to make a few assumptions with respect to the 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 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.

4.3 Merging Data

I merge the two datasets based on store ID and the week of the consumer's shopping trip. As stated above, I assume that all products with positive sales are in the consumer's choice set and I identify their actual purchase choice by matching the product they chose with the product being offered by UPC. There are some cases where these cannot be precisely matched up and I outline my steps to fix these issues below.

Missing Brand Name Purchases In some cases, the purchase of a brand name cannot be matched by UPC to its corresponding store. In these cases, I simply combine the purchase record from the Consumer Panel with the product assortment offered at that store in the corresponding week. For example, if a consumer records purchasing a 12-pack of Quilted Northern at a particular store in a particular week, but the corresponding store has only recorded sales of other brands (Charmin, Angel Soft, etc.) and maybe other sizes of Quilted Northern (4-, 6-, 8-roll packages), then I add on the consumer’s actual purchase to the store assortment. The Kilts Marketing Center has confirmed that this happens occasionally, but they have not been able to ascertain a reason for this occurrence.

Private-Label Purchases In order to maintain anonymity of stores, Nielsen masks the UPCs of private labelled goods in the Retail Scanner Dataset. Since these are commonly purchased, I match private-label purchases between the two datasets based on the following procedure. I record the price, ply, and package size of the item purchased in the Consumer Panel Data. If there is a corresponding item that is the same price, ply, and package size, I record that as a match. This is a conservative matching algorithm and misses some items that might have been purchased at a sale price since the prices would not match up. If there remain unmatched purchases, I follow the same procedure as with the brand-name purchases and append the record from the Consumer Panel to the store’s assortment.

Missing Stores Not all individual stores share scanner data with Nielsen or even those that do are not able to be linked to a corresponding store in the Consumer Panel data. In these cases, I impute the choice set available to households by using Assumption 2 below:

Assumption 2: *Stores in the same retail chain and DMA 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 quite similar across stores within the same retail chain. While it is not precisely true that all stores in a chain offer the same assortment of products or offer the exact same price (which is pointed out in DellaVigna and Gentzkow (2017)), I take a conservative approach and generate a “representative” store within a chain by taking the union of the products offered and taking the quantity-weighted average price of those stores for each product. To capture the fact that some chains may coordinate inventories within a region, I generate with “representative” store based on a

chain-DMA. For example, if Chain A has multiple stores in Boston and Atlanta, I construct two “representative” stores, one with the set of products offered in the Boston area and the other with the set of products offered in the Atlanta area. If product assortments are set at the national level, there should not be much difference, but this approach is a conservative way to capture the product assortment faced by a consumer.

The results of this matching procedure are recorded in the table below. I start with the Consumer Panel data that has been cleaned as outlined in the “Consumer Panel Data” section. We can see that a majority of the sample shrinkage occurs because Nielsen has not assigned certain stores an ID and because some stores that have been assigned an ID do not share their scanner data with Nielsen. Overall, the sample remains large and could be credibly expanded under Assumption 2 below.

Step	Obs	HH
Starting	3,639,516	120,927
Cannot be standardized	3,596,784	120,927
Matched to Scanner	1,153,184	83,102
Restrict to top brand/size combos	826,630	74,642
Single packages	811,819	74,446

Collapsing Products For some products, a brand-size does not define a unique product. In these cases, I collapse these items to a composite brand-size with a standardized size, unit cost, price, and ply reflecting the average of the component products.

How Some Get Fooled At Toilet Paper "Sales"

They see a big, flaring announcement of eight or ten rolls of toilet paper offered for a quarter. Did you ever ask yourself how much real paper there is in these rolls?

We know the rolls look big, but that's because they are loosely wound or the paper is coarse and thick and very little paper takes up lots of room. Ask your dealer if the wrapper guarantees how much paper there is in that roll, for

"It's the Counted Sheets that Count"

ScotTissue Toilet Paper

Has 1000 Sheets to the roll—10c

Printed on the label—you don't have to guess quality. But don't ever buy toilet paper on a quantity basis *alone*—Quality must always come first. Examine ScotTissue. It is as "soft as old linen," snowy white and very absorbent.



Copyright 1913 Scott Paper Company

Sani-Tissue, balsam treated and very soft. A very excellent toilet paper for special purposes. 2500 sheets in carton of 3 rolls, 25c.

At drug, grocery, department and housefurnishing stores.

SCOTT PAPER CO.

30 Church Street

New York

Makers of Absorbent ScotTissue Towels

Figure 1: