Humboldt University Berlin

Institute of Marketing

Prof. Dr. Daniel Klapper

Advanced Marketing Modeling

SS 2021

Special Work Performance 3: Estimating Aggregate Logit Demand Models

This is individual work.

Your answers including all tables and graphs must not exceed 10 pages. Please start a new page when providing your report to a new subtask. Please use typeface Times Roman in 12pt with 1.15 line spacing (in tables and graphs you may use 10pt and 1.0 line spacing) and 1 inch space on all sides. Do not forget to report your name and student number and a page number on each page starting with number one on the first answering page.
Do not include a title page or content page.
Send your report as pdf to my email address daniel.klapper@hu-berlin.de not later than August 31, 2021, 4:00pm. Please report in the subject line "AMM SWP3 and your name".

Use the dataset from SWP2 and exclude the stores 266595 and 259111 from the analysis.

Investigate the effects of the introduction of the brand 'Coke Zero' on brand valuations, sales, market shares, price elasticities and promotion uplifts of the brands 'Coke Classic', 'Diet Coke', 'Pepsi', and 'Diet Pepsi' in the two markets 'Eau Claire' and 'Pittsfield'.

You can but do not need to show differences with respect to packaging and/or volume!
Use an aggregated (nested or non-nested) logit demand model to support your analysis.

Interpret the results carefully and document your estimation strategy in some detail. Do not report R-codes and edit the estimation results you obtained with R. Also, make use of your econometric knowledge when estimating parameters of a linear additive model. Use tables and graphs to support your description and explain in words the key facts of the data set.

Advanced Marketing Modelling SWP 03

Introduction

In this Special Work Performance (SWP), we tried to analyze the effects of the introduction of the brand 'Coke Zero' on brand valuations, sales, market shares, price elasticities and promotion uplifts of the brands 'Coke Classic', 'Diet Coke', 'Pepsi', and 'Diet Pepsi' in the two markets named 'Eau Claire' and 'Pittsfield'. First, we start the report with data introduction and how we prepared the data. Second, we compared nested and non-nested logit demand models to others such as probit, simple logit, multinomial logit and mixed logit model. After that, we focused on explaining what did we need to calculate and why did we need to calculate certain things for the nested and the non-nested logit demand models such as market share and outside good calculations as well as which routes/combinations we tried for getting plausible results that satisfied models' assumptions. Then, using simple log-log and non-nested logit model, we reported main facts about the data while also drawing attention to what might have been Coke Zero's possible effects on other brands in the two markets over time. After that, we introduce nested logit models and justify our decision of specific level number and the nesting structure we chose for the final estimation of the assignment. And after that, we investigated different instrumental variables to overcome the possible endogeneity problem in the nested logit model and opted for some instruments, which is then followed by the analysis of effects of the Coke Zero's introduction. The assignment is finished by a discussion part at the end.

The Data

We had panel data of Coca Cola Co and Pepsi Co's 5 products in total, aggregated to the weekly level through the years of 2004 to 2006 and to the store level in the cities of Pittsfield and Eau Claire in USA. To begin with, stores 266596 and 259111 were excluded from the analysis because they displayed some unusual patterns and were different from the rest of the data. So, including those to the analysis was going to result in a bias, of which we were not able to identify the reasons. Secondly, we had data of volume equivalents of specific purchases in ounces of 0.1042, 0.3521, 0.75 and 1.50, which we converted to liters and corresponded to 600 ml plastic bottle, 2-liter plastic bottle, pack of 12 cans and pack of 24 cans, combined with the packaging information, respectively. And additionally, we had information such as units, meaning that how many units of the previously mentioned volume and packaging combinations was bought in a specific purchase, price information, which meant price per unit, varying with the units bought, and we had the revenue from a specific purchase, which is calculated by multiplying units and price per units. So, some of the first decisions we had to make in the preparation of the data phase were that how we were going to calculate the market share within Coke Classic, Diet Coke, Pepsi Classic and Diet Pepsi using the volume, units and price information as well as the share of the "outside good", both of which we needed later for the nested and non-nested logit demand models. But for that, it is probably a good idea to provide a brief explanation of why we had chosen nested and non-nested aggregate logit demand models for our analysis and a comparison of advantages, disadvantages and assumptions of the other possible demand models that we could have used.

Reasons for choosing Nested & Non-Nested Aggregate Logit Demand Models

There are many demand models in the field of marketing ranging from the ones that has unrealistic assumptions about the market conditions while often being easier to implement and interpret to the ones that has more realistic assumptions about the market conditions, but, at the same time, would require more resources computationally and interpretation-wise. Because of the high number of demand models that we could have used, for simplicity, in this part of the SWP, we only focused our attention to analyze and compare a few of them, namely that, Probit, Logit, Multinomial Logit, Nested/Non-nested Logit and Mixed Logit models, each of which has its own advantages and disadvantages and are suitable for various specific market conditions. So, one of the first and the simplest demand models that we could have used for this assignment was the Probit model. The Probit model gives the researcher when working with binary type of data. We could have chosen this model for our analysis, for example, if we had data about buying a specific product in a super market or adding one to a digital basket when shopping online, in which the positive cases could be marked with a 1 and else 0, giving us a binary type of data. However, this type of demand model didn't meet our needs for the assignment. Firstly because rather than binary type of data, we had sales data aggregated to the week and store level, where we also had almost no demographic information about the consumers, and secondly, because the Probit model doesn't offer a closed form solution in markets where the number of products is relatively high like ours and lastly, also because of its obligation to simulate the normal distribution of the error term when using it. Our next option constitutes a group of demand models, which one can name as "The Logit Models" and which the Nested and the Non-nested Logit models that we have chosen for this assignment also belongs to. So, in this group of demand models our first option was the Simple Logit Model. Although in this model the error term of the consumers' utility function is assumed to be extreme value distributed enabling us to compute the probability of a certain product to be chosen in a simple closed form expression, that we will make use of in other models later, since it was designed for markets consisting of a single product and an outside good only, we moved on to another version of it, namely that Multinomial Logit Model (MNL). The MNL model fitted our needs better in the sense that it enabled existence of multiple products in the market which was the case in our data; on the other hand, it suffered from some severe restrictions such as that it assumes that all customers have the same preferences and that the probability of a customer choosing between two products doesn't depend on the number of already present or introduction of some new products in the market (This assumption is also known as the Independence of Irrelevant Alternatives, IIA). There are some other models that ease these restrictions by allowing for heterogeneity at the individual level, accounting for the problem of IIA and even the problem of endogeneity, for which we will use instrumental variables to overcome, such as the Mixed Logit Model or the BLP model but these more advanced models are computationally much costlier and are harder to interpret. Thus, because of the following reasons we stick with the Aggregate Logit

Demand Models called Nested and Non-nested models, in this assignment. First is that compared to the more advanced demand models they are relatively easy to use. Second, the data we had were not individual choice data and we didn't have any information about who the customers were in the carbonated beverage markets except the store that they purchased their drinks in. And lastly, the data we had was already aggregated sales data to the store and week level, which we could either further aggregate to either packaging level if we wanted a more detailed analysis or to the brand level for a more general perspective of relationships between the brands in the market.

The Nested & Non-nested Aggregate Logit Demand Models

One realistic assumption of all of the Logit Demand Models is that they try to explain the demand in a certain market by making use of the random utility framework, where we assume that the consumers choose a product in a market that provides them with the maximum utility or opt out of the market by choosing the outside good if utilities of all of the products in a market are smaller than 0, which also serves as a normalization tool of the utility levels, taking the 0 as the base level when the customer doesn't purchase a product. These mean utilities across the consumers are determined by the utility function consisting of product related attributes, a price factor and a structured error term, which is observable to the consumer but not to the researcher, and essentially, we want to estimate the effects of these parameters on the demand over time. Since the utility function has already a linear additive form, of which we can take advantage using simple regression, we only needed to calculate the products' mean utilities (the dependent variable) and for that we needed to come up with an approach to calculate the market size and the outside good size, thus the products' shares too. So, in order to do that, first of all, we had to make an assumption about the proportion of the outside good compared to our 5 product market. Because in reality, it is impossible to compute the exact market size as we don't have information about random situations like the potential customer initially planning to buy the product but then getting distracted or seeing the product but simply deciding to buy a non-carbonated beverage, we restricted our market size to the people entering the supermarket and actually ending up buying a carbonated beverage of any brand. After that we converted the ounces to liters by multiplying the volume equivalents in each row with an ounce-to-liter ratio and units. This way, we had a proper measure to quantify the market shares. We wouldn't use the "units" variable because in two different rows of value 1, it could mean both a sale of 600 hundred ml bottle or 24 cans. We also didn't use revenues for the market calculation because there would be volume discount in price for the purchases were done in bulks, creating inconsistency among the variables. For this reason, we also converted the price and the display and feature values (display area and retailer advertisement) in a way that they are proportioned per liter of the specific purchase. As a result, although it is possible to calculate the market shares in so many different combinations and we calculated at least 20 versions of them, we decided to go with the simple one, that of the proportion of the liters at the individual purchase level to the sum of the liters in the whole market of carbonated beverages aggregated to the market and week level. One last important point to mention about the preparation of the data is that, although we knew some volume and packaging combinations showed different price elasticity behavior and the like, we aggregated the data to the brand level both in order to focus on the bigger picture in the market and also make it easier to satisfy price elasticity assumptions with, for example, very price inelastic products such as 2-liter plastic bottles. So, after the whole reasoning of the models chosen, comparison of them, data preparation and market share calculation phases, we can now provide the analysis of the data.

Analysis of the data:

In the following bullet points there are some quick facts about the data so that they introduce the data and provide a preliminary analysis of it before the actual nested demand models:

1- In total, over the course of three years Classic Coke were sold the most by around 3.7 million liters, followed by Diet Coke with a 3.6 million liters. Next, follows Pepsi with 3 million liters in total, which is followed by Diet Pepsi with a 2.3 million liters.

2- When we take a look at the total liters in the markets, we see that the most consumed product in Eau Claire is Diet Pepsi and then Classic Coke, followed by the low Pepsi sales. When we take a look at the total liters of the products in Pittsfield, both Classic Pepsi and Classic Cola are really close to each other with high sales, whereas the diet products have been low in sales. So, based on this information, there might be evidence suggesting that consumers in Eau Claire care most about getting a Coca Cola brand product and more specifically a diet one and consumers in Pittsfield seem to care more about getting a drink with a sugar mostly.

3- The population of Eau Claire in 2007 was 63,190. The population of Pittsfield at the same year was 42,652. In total, in Eau Claire 7,984,740 liters was sold and in Pittsfield this number was 4,717,831, which is 69.35% lower. However, when comparing the total revenue in the two markets Pittsfield is 40.77% lower only, which indicates some products are more expensive in Pittsfield than in Eau Claire.

4- Although price of Classic Coke and Diet Coke has been increasing in Eau Claire, both of which had a mean of 0.70, it has been more or less stable in Pittsfield, with a mean of 0.72 in Classic Coke and 0.74 in Diet Coke. So, based on this information we can conclude that Prices of Coca Cola products are higher in Pittsfield but they increase more in Eau Claire. Thus, one could argue that Coca Cola realized that Eau Claire values Coca Cola more.

5- Classic Pepsi and Diet Pepsi's prices in Eau Claire slightly increase until around the Coke Zero's introduction then stabilize, both with an overall mean of 0.73. And while Classic Pepsi has had a stable mean price of 0.69, Diet Pepsi has had a stable mean price of 0.72 in Pittsfield. So, from this information we can conclude we might have evidence that although Eau Claire is a market that is closer to Coca Cola, Pepsi has been trying to raise its prices in Eau Claire but not in Pittsfield as far as we observe the weeks before Coke Zero's introduction.

6- Although price of Coke Zero increased 20 cents during the last 20 weeks of 2006 in Eau Claire resulting in an average price of 2.19, the price of Coke Zero in Pittsfield has been stable around 2.30. This provides another evidence that after the introduction of Coke Zero in mid-2005 Coca Cola might have realized that Eau Claire values Coca Cola more, it is less price sensitive to Coca Cola and could be charged more.

7- When calculating Coke Zero's revenues per liter sold, we find out that Coke Zero were sold more in Eau Claire.
8- One evidence that the introduction of Coke Zero benefiting the Diet Coke the most is that among all the products and both of the markets only Diet Coke's sales have increased and only in Eau Claire from the years of 2005 to 2006.

These facts were based on some tables and plots. Next, we will try to confirm these preliminary results with simple log-log models that we have used in the previous SWP.

Exploratory Analysis with Log-Log and Non-Nested Aggregate Logit Demand Models

Log-log models suffer from serious restrictive assumptions but to see the direction of the relationships over time, it is an option that is easy and practical and for that reason we first run a log-log model where the dependent variable is Liters. One of the first findings and a fact that we used through the analysis to back check the plausibility of the models is that, respectively, Classic Coke, Diet Coke, Pepsi and Diet Pepsi were valued and sold the most over time. We also observed positive and significant effects of additional display area in the retail store on the sales of all brands (affecting Classic Coke the lowest possibly because it already has a strong product recognition) both in this model and through the other models in the analysis. Lastly, we also observe Eau Claire being more price sensitive in general than Pittsfield. In addition to the log-log model, we fitted two non-nested logit models that produced price elasticities lower than -1, one with the brand level data subsetted to Pittsfield and the other to Eau Claire, both of which also had an additional parameter indicating whether the purchased product was a sugar or a non-sugar product. We were able to come up with some additional conclusions to the ones from the log-log model. Firstly, examining the model for Eau Claire, we observed higher brand coefficients for the Coca Cola products than the Pepsi ones, which indicated Eau Claire is a market that is closer to Coca Cola. However, when we also noticed that the coefficient for the Diet Coke was significantly higher than of Classic Coke and when we combine this information with the other findings such as Eau Claire being more price sensitive and both Diet Coke and Classic Coke having a mean price of 2.84, we were able to conclude that the consumers in Eau Claire have a strict preference for non-sugar drinks. To decide on whether they prefer more Coca Cola than Pepsi because of it tastes better or because it is cheaper is hard to do. Our second finding is from the non-nested logit model in Pittsfield. In Pittsfield, the average prices of the Diet Coke's, Classic Coke's, Diet Pepsi's and the Classic Pepsi's are respectively that of 2.45, 2.50, 3.08 and 3.11 dollars; however, the two highest brand coefficients were those of Classic Coke and Classic Pepsi's. This finding points out to two features of the consumers in Pittsfield; first is that, probably because they are less price sensitive and also, they have a slight preference for Pepsi, they are willing to pay 66 cents more on average to get a Classic Pepsi, and the second is that, no matter how cheap it the non-sugar products can be, they have an obvious preference towards the drinks with sugar. Before we move on to the next part, one important point to make here is that, the reason why we didn't mention any instrumental variables that should have been used for the very likely endogenous price variable here was that these models were fitted more for exploratory purposes rather than getting accurate estimates. In the next part about the nested logit models the topic of instrumental variables, will be discussed more in detail.

Nested Logit Demand Models

There are couple of reasons why we didn't use Non-nested logit demand models for getting proper coefficients and price elasticities. First is that the price elasticities in the non-nested logit models tend to be over- or under-affected by the share of the products. Second reason is that the cross-price elasticities are the same for all the products, caused by the Independence of Irrelevant Alternatives (IIA) assumption, which is not realistic because an increase in the prices of a product is likely to introduce different substitution effects between products with different magnitudes. Last but not least, nested logit models are more realistic in markets of multiple and to some extent differentiated products like in our case because the nesting structure in the models try to simulate the consumers' decision process when purchasing a product by assuming that they subset the market by the features of the products in the market such as brand, packaging, taste, type etc. and to either until only one specific product is left or a subgroup of products is left (like a decision tree), among which the consumers make his or her final decision of purchase. Thus, both because the problems of the non-nested logit models are at least partially solved in the nested logit models and because of the realistic hierarchical decision-tree like nesting structure assumption, which fits to our case, we chose nested logit demand models for the final estimation of coefficients and price elasticities. While applying a nested logit model, there are two main decisions that needs to be taken by the researcher before implementing them. First is, the number of levels in the nesting structure and the second is how the specific nesting structure will be. In our opinion, one of the most realistic nesting structures is, 5 or 6 level nesting structure. One example of such a structure is that where the customer first gives the decision of "Buy/Not Buy", then "Choosing a Store", then "Sugar/No Sugar", then "Choosing a Brand" and then finally "Choosing Packaging/Volume". However, introducing such long structures to the model would require a lot more attention and focus while making the calculations, and although it would allow for getting more detailed results, it would also be harder to interpret it from a broader perspective, for example at the brand level like we did. 5 or 6 level nesting structures are or could be considered of somewhat extreme because of the reasons above, while the ones with 3 levels are somewhat more common to estimate. However, we aggregated all the data to the product, week and market level and because of that we weren't able to calculate 2nd subgroup's market share. Thus, we used 2 level nested logit model, for which we tried some nesting structures and checked some criteria in order the model to be sensible and in line with its assumptions. Before, when we aggregated the data to the package level, because it was possible to iterate through more combinations of how a nesting structure could be built, we could check more nesting structures but with the data at the brand level, we were still able to check 3 nesting structures, and choose one by comparing coefficients to see if there were any peculiarity, if they were in line with what we already observed in the data and lastly, if the results like the alpha value or the price elasticities were within the model's assumptions that are realistic. When comparing the models, in them, we decided on trying the same combination of variables and interaction terms, whose possibilities of being relevant to the model were the highest to us, so that even they turn out to be insignificant we would have an explanation of why we included them and also to make the comparison of different models possible, otherwise, by changing the variables and combinations of interaction terms, it was possible to get results that were parallel with the model's assumptions. So, the first nesting structure that we checked was the one that where we assumed consumers, when making a purchasing decision, would first decide on buying a Cola or not and then they decide on which product to buy directly. This model was eliminated because the brands were very price inelastic. The second nesting structure that we checked was the one where we assumed, when purchasing a product, consumers would first decide on which brand to buy and then they would choose a product within the brand they chose. In this model, the products within the subgroup were close to becoming perfect substitutes as, the alpha value which represented the within group correlation of taste among the products were 0.95. And also because of that reason, we got a unreasonable results where some of the price elasticities extreme with values of -6 and also some of the cross elasticities within the same group were higher than the direct elasticities, which constituted the motivation for us to eliminate this nesting structure from further consideration. The last nesting structure that we checked and also that seemed the most realistic to us was the one that where we assumed, when purchasing a product, consumers would first decide between purchasing a product with or without sugar and then deciding on the brand. In this model, without making use of instrumental variables, we observed an alpha value of 0.78, a price coefficient of -0.62 and direct elasticities being bigger than within group elasticities and that being bigger than the in between group elasticities in absolute terms. However, for the model to be more robust, we needed instrumental variables because it is likely that alpha value, representing the within group correlation of taste among the products is correlated with the price and also price with the error term, that is observed to the consumer but unobserved to the researcher, namely that structural error term, which creates a problem of endogeneity. So, first, we checked the correlation of the following possible instruments with our endogenous variables: sugar price, aluminum price, gasoline price and temperature.

	L5	MARKET	direct.ela.nl	cross.ela.same.group.nl	cross.ela.different.group.nl
1	COKE CLASSIC	EAU CLAIRE	-1.858690	2.032931	0.04975973
2	DIET COKE	EAU CLAIRE	-1.661811	2.225155	0.06097588
3	DIET PEPSI	EAU CLAIRE	-2.804634	1.239626	0.03228068
4	PEPSI	EAU CLAIRE	-2.620061	1.430366	0.03418711
5	COKE CLASSIC	PITTSFIELD	-2.272368	1.725143	0.04765039
6	DIET COKE	PITTSFIELD	-2.276471	1.844262	0.03037075
7	DIET PEPSI	PITTSFIELD	-2.335137	1.665759	0.02687552
8	PEPSI	PITTSFIELD	-2.150042	1.716513	0.04674267

Table 1. Table of Elasticities Aggregated to the Market and Product level

After we tried some combination of the possible instruments starting from the ones with the highest correlation, we found out that when we used aluminum price as an instrument for the price and sugar price for the within group share variable, the coefficient of price in the model changed from -0.62 to 0.75 as well as the alpha value from 0.78 to 0.87, which also made the direct price elasticities more negative to a level of between -1.60 and -2.60. Since these two instruments worked in the way that was expected, we will use these instruments for estimating the effects of the Coke Zero introduction on other brands.

The Effects of Coke Zero's Introduction

In order to see whether introduction of Coke Zero in mid-2005 had a significant effect on other brands and if yes, to what extent, the approaches that we have tried can be summarized in four ways. First is that, we used a dummy variable called "cz.intro" that had 1s in the rows when Coke Zero was already in the market, and else 0. And our idea was to interact cz.intro with the year variable in separate versions of the data that were subsetted to Eau Claire and Pittsfield. Our second idea was that, we subsetted the data to the two markets again, but this time subsetted each of them for the weeks where cz.intro variable was 0 and 1. This way, by comparing each models' coefficients, we wanted to see the effects. Our third idea was that to compare the subsets of Eau Claire and Pittsfield in 2004 and 2006 and also with cz.intro variable. The idea with this attempt was that the brands might not be affected immediately after the introduction of Coke Zero, which in our opinion was a plausible assumption. And our last idea was that to not aggregating the data further to any level, but trying to see the effects by adding interaction terms between markets, brands and years as well as the cz.intro variable. In the first three methods that we have tried, we were not able to get logical results. Most of the time, coefficients were too large or too small. Sometimes, coefficients of a specific year were interpretable but in the other year that we would compare it to, they were completely different, which made it impossible to comment on. However, in our last idea, where we counted on the interaction terms without aggregating the data further down to any level, the price coefficient in the model was about -0.60 and the alpha value 0.70, which are good signs of the model being correct but none of the brand, product, market, year combination of interaction terms interacted with cz.intro variable was significant. So, it wouldn't be right to count on insignificant values that were either around 0.01 or smaller. In real life, we know that Coke Zero had a very big success in the carbonated beverage markets all over the world, and replaced diet cokes to a great extent but based on our analysis between 2004 and 2006 in the two markets of Eau Claire and Pittsfield US, we didn't find a significant effect of the introduction of Coke Zero in mid-2005 in any of the products of Coca Cola Co and Pepsi Co.

Discussion

One idea is that to make the same analysis with more data, especially of after 2006. This would likely to help us find significant results. Another possible criticism of the analysis we did is that nested logit demand models also assumes the same nesting structure for the whole market.

Although it might be easier to interpret the model this way, this assumption is unlikely to be real. Possible options to nested logit models could be mixed logit model or the models that use Bayesian techniques where heterogeneity is accounted for in more detail, but although these models provide more detail, they are usually harder to interpret and costlier to implement. Thus, when choosing another method, one should take into the account these aspects of model selecting.