Group X7
Effects of the Marketing Mix on Cereal Purchases of Households

Abbaan Nassar 413272

Hien Ngo 431495

Roland van Veen 430476

Veerle de Boer 411746

June 1, 2017

Abstract

In this report we look at the purchasing behaviour of households living in the suburbs of a large American City, the focus being the purchase of cereals. There are 7 options in this paper; General Mills, Kellogg's, Philip Morris, Quaker Oats, Ralston Purima, Nabisco and No-Purchase. We try to describe the decision to purchase or not and the specific brand choice using econometric models. First we constructed a general logit model. However the test for irrelevant alternatives was inconclusive. Next we constructed a nested model, in which the 7 options are divided in two clusters: non-buying and buying. The model showed that mainly the price and feature influence the brand decision of households. This can be incorporated in the marketing strategy of a supermarket. Households can also be very brand loyal, these households are less sensitive to the strategies.

1 Introduction

There is a lot of data available concerning the purchases of customers in supermarkets. By using the barcodes of grocery products and scanner checkouts many information is recorded such as name, brand and current price of a product. Also the time of the purchase and the presence of promotions are known for competing brands. Supermarkets now often supply their customers with personalised shopping cards to gain more insight into individual purchase data. With these cards not only the aggregated data, that is to accumulate the individual household purchase into aggregate sales of a brand, but also the purchases of individual households can be studied. This information can be of great use for marketing strategies.

Econometric models are a means to use data to evaluate what drives a brand to be sold more often. McFadden (1980) proposed an econometric view of marketing using models concerning probabilistic choice among products. More recently, Franses & Paap (2001) researched quantitative models in marketing research and model the choice between four brands by households.

Our report aims to make use of available data to obtain useful insights into the purchasing-behaviour of customers of a supermarket in a large American city. More specifically, focusing on the purchase of cereals and the corresponding brands. Rather than aggregating the data to reach more simple conclusions, we shall attempt to extract more information as to the causes for the observed purchasing behaviour.
We try to describe the decisions of households with respect to the purchase of cereals by using econometric models. First we use a combination of a conditional logit model and a multinomial logit model. Secondly, we consider a nested logit model, which firstly models the discrete choice between buying or not buying cereals. We shall use these models to obtain insight into the importance of the explanatory variables: what causes one brand to be sold more than another?

First, in §2 the data is described and analysed. Then in §3 we formulate the basic model using a general logit model and propose a nested model. The results and interpretation of the models are given in §4 and finally a conclusion follows in §5.

2 Data

2.1 Notation

In our study, the database contains visits of 100 households over a two year period. Our dataset contains a total of 13008 invoices. The households live in the suburbs of a large American city and can choose among 6 brands of cereal and a non-buying option, namely:
1. General Mills
2. Kellogg's
3. Philip Morris
4. Quaker Oats
5. Ralston Purina
6. Nabisco
7. No purchase

The database provides brand-specific information and demographical information of the households. First we consider the explanatory variables for the brand specific information:
- $BRAND_{ijt}$: 0/1 dummy variable which equals 1 if household i purchases brand j at visit t and 0 elsewhere.
- $PRICE_{ijt}$: price of brand j during visit t of household i in dollars.
- $DISPL_{ijt}$: 0/1 dummy variable which equals 1 if there is a display of brand j during visit t of household i.
- $FEAT_{ijt}$: 0/1 dummy variable which equals 1 if there is a feature promotion of brand j during visit t of household i. A feature relates to an advertisement.
- LAG_{ijt}: 0/1 dummy variable which equals 1 if household i purchased brand j at the previous visit.

Next, we look at the explanatory variables for the characteristics of the households:
- $HHSIZE_{it}$: the household size of household i at visit t.
- $DOLLARS\_SPENT_{it}$: the total amount spent by household i at store visit t.
- $WEEKS_{it}$: the number of weeks since the last purchase of cereal.
- $INCOME_{it}$: the annual income of household i divided in 11 bins:
1. smaller than 10,000 dollar
2. 10,000 - 11,999 dollar
3. 12,000 - 14,999 dollar
4. 15,000 - 19,999 dollar
5. 20,000 - 24,999 dollar
6. 25,000 - 34,999 dollar
7. 35,000 - 44,999 dollar
8. 45,000 - 54,999 dollar
9. 55,000 - 64,999 dollar
10. 65,000 - 74,999 dollar
11. 75,000 dollar and higher

for i=1,...,100, j=1,..,7 and t=1,...,T_i.
T_i is the number of store visits of household i, therefore, T_i is different among households. Brand 7 (no-purchase) is a special case, the price, display, feature and the lag variable of this brand are always 0.

2.1.1 Variable extensions

Furthermore, we add and change some variables so they can be incorporated in the models.
First we change the income variable, because the bins are not linear, to a continuous variable.
-$INCOME\_CONT_{it}$: The income of household i where we take the average income of the bin.
Next, we create a dummy variable for $DOLLARS\_SPENT_{it}$. Because this variable is correlated to our dependent variable; if a household buys cereal, their dollars spent automatically goes up.
-$DUM\_MAIN\_TRIP_{it}$: 0/1 dummy variable which equals 1 if household i spends more than 10 dollars at visit t.
Finally, we create a brand loyalty variable from the data (LOY_{ijt}) using the lag variables to determine the degree of brand loyalty by every household. Brand loyalty can be of good use in the model, since this will likely have an effect on the willingness to buy certain products at different prices. To do this, we use the following recursive formula first described by Guadagni and Little (1983):

$$b^i_{j,t} = \alpha b^i_{j,t-1} + (1-\alpha)I[y^i_{t-1} = j] \;\; for \;\; i = 1, ..., 100, \;\; j = 1, ..., 6 \tag{1}$$

where $0 < \alpha < 1$ is a constant determining how long the effects of a purchase last for the loyalty, i is the household, and j is the cereal brand. To start the recursion we use that if brand j was chosen the first time, we set $b^i_{j,t}$ to α, and if not, we set it to $(1-\alpha)/5$. By trial and error we end up with $\alpha = 0.50$.

2.2 Data insights

We analyze the data to get some understanding of the explanatory variables and insight in the data itself. If we look at table 1 we see that for 78% percent of the visits, the households do not buy cereals. When cereals are bought, the 2 most popular brands are brand 1 and brand 2 (General Mills and Kellogg's). This concerns a more expensive and a cheaper brand. If we look at the last purchase, it shows households are often committed to the brand they lastly bought, especially for brand 2 and 5. The direct effect of feature and display is difficult, because a household might buy a brand when it is featured however this does not have to be the reason for buying the brand. It looks like feature alone is not a big reason, giving the small percentages. However households do buy some brands (1, 2 and 3) often when it is in feature and display.

Table 1: Characteristics of the dependent variable and explanatory variables:

	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5	Brand 6	No-Purchase
Average price (US dollar)	0.21	0.14	0.16	0.15	0.21	0.18	-
Choice percentage	5.8%	9.6%	2.1%	1.5%	1.5%	2.0%	77.5%
Brand percentage	25.9%	42.5%	9.3%	6.8%	6.8%	8.7%	-
Feature only	2.5%	2.4%	3.3%	2.0%	3.5%	2.8 %	-
Display only	19.9%	7.3%	5.5%	4.5%	6.1%	2.4 %	-
Feature & Display	15.9%	16.5%	18.4%	8.5%	6.6%	2.8 %	-
Last purchase (lag)	46.0%	59.5%	28.3%	27.5%	58.6%	53.9 %	-

This table shows some insights for 13008 invoices of a 100 different households.
It shows the average price, choice percentage and brand percentage. The choice percentage shows how many times each option was chosen and the brand percentage shows the distribution of the brands when cereal is bought.
The table also shows percentages concerning the feature, display and lag variables. These percentages concern how often the specific brand is displayed, featured or also bought last time, when the brand is chosen. For example 15.9% of the purchases of brand 1 are done when brand 1 is featured and displayed.

Table (2) shows the average household size and income bin per option. It is difficult to interpret these numbers because a household which buys one specific brand a lot might influence the data. Also if a brand is bought by high and low income groups, it averages up. But by merely looking at these numbers, it stands out that brand 5 is a brand that is bought in smaller households with a higher income on average and that brand 1 is bought by larger households.

Table 2: Household characteristics per brand:

	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5	Brand 6	No-Purchase
Average household size	2.91	2.79	2.74	2.32	1.93	1.84	2.19
Average income bin	6.61	6.71	6.98	6.38	7.38	5.89	5.55

This table shows per option the average size and the average income bin of the households.

3 Methodology

3.1 General Logit Model

In our problem, households can choose between 6 brands of cereal and a non-buying option. The model has to be able to predict the household's future choice. We have an unordered multinomial dependent variable taking the values:
1.'General Mills',
2.'Kellogg's',
3.'Philip Morris'
4.'Quaker Oats'
5.'Ralston Purina'
6.'Nabisco'
7.'No purchase'
The random variable Y_i, which underlies the observations y_i can therefore take 7 discrete values. The corresponding utility function associated with the choice of brand j is given in equation (2).

$$u_{ij} = z'_{ij}\alpha + x'_i\beta_j + \epsilon_{ij} \quad j = 1, ..., 7 \tag{2}$$

where
- z_{ij} denotes the vector of explanatory variables associated with the brand i pertaining to the household i. The variables we consider are $BRAND_{ijt}$, $PRICE_{ijt}$, $DISPL_{ijt}$, $FEAT_{ijt}$, LAG_{ijt} and LOY_{ijt}; and α is a set of unknown parameters.
- x_i denotes the vector that is constant for all alternatives that is, the choices of the household are correlated with household-specific explanatory variables. These variables have the same values for the different options j: $HHSIZE_{it}$, $DUM\_MAIN\_TRIP_{it}$, $WEEKS_{it}$ and $INCOME\_CONT_{it}$; and β_j with j=1,..,7 varies across outcome categories.

To predict the brand decision, our first model combines the conditional and multinomial logit and is therefore a general logit model. The probability function is given in (3):

$$Pr(y_i = j|X_i, z_{ij}) = P_{ij} = \frac{exp(z'_{ij}\alpha + x'_i\beta_j)}{\sum_{h=1}^{7} exp(z'_{ih}\alpha + x'_i\beta_h)} for\ j = 1, ..., 7 \tag{3}$$

Next we estimate the parameters α and parameter vector β by using maximum likelihood. Our general model is based on the assumption of independence of irrelevant alternatives. However this might not be realistic, to check for validity of our model we will use the Hausman IIA test. The IIA assumption states that the odds ratio between two alternatives under consideration is independent of other alternatives.

3.2 Nested Logit Model

In case that our first model violates the IIA assumptions, we propose a second model. The nested logit model is a common approach for dealing with violations of the IIA assumption. For this model, there exists the assumption that the 7 options can be divided into 2 clusters: non-buying and buying. We chose this division, because buying cereals is a different decision than choosing which brand to buy. Therefore concerning the brand choice, one cluster consists of the no-purchase option only and the other cluster consists of the 6 remaining brands. This is illustrated in figure 1.

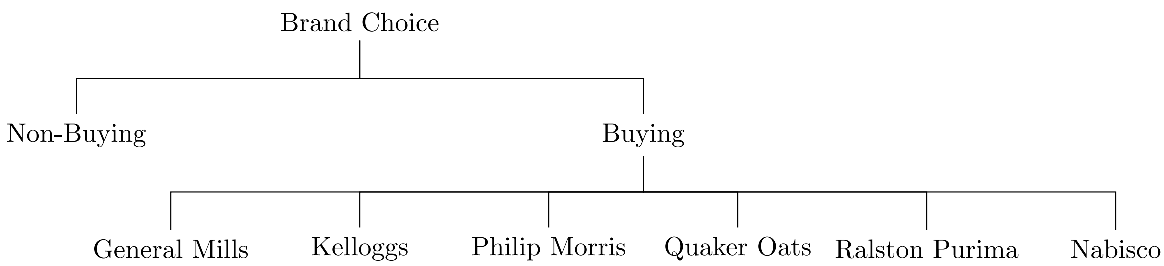


Figure 1: Clustering Nested Model
This figure shows the 2 clusters: non-buying and buying cereals. The second cluster consists of the 6 different cereals brands.

The random variable Y_i, which underlies observations y_i is now split up into 2 random variables (C_i, S_i) where c_i corresponds to the choice of the cluster and s_i to the choice among brands within this cluster (i=1,2). The probability concerning the brand choice of household i within each cluster is given by (4).

$$Pr[S_i = j|C_i = m, Z) = \frac{exp(Z_{j|m}\gamma)}{\sum_{j=1}^{6} exp(Z_{j|m}\gamma)} for\ j = 1, 2, ..6 \tag{4}$$

where $Z_{j|m}$ denotes the variables that have an explanatory meaning for the choice within cluster m :

- For m=1, the non-buying cluster. This cluster contains only one option so we do not have to specify the logit model.
- For m=2, the buying cluster. We use the following explanatory variables: $DISPL_{ijt}$, $FEAT_{ijt}$, $PRICE_{ijt}$, LOY_{ijt} and LAG_{ijt} and set Nabisco as the base brand in this cluster.

The choice between the 2 clusters is defined using a binomial logit model:

$$Pr[C_i = m|Z) = \frac{exp(Z_m\alpha + \tau_m I_m)}{\sum_{l=1}^{2} exp(Z_l\alpha + \tau_l I_l)} \tag{5}$$

where Z_m denotes the explanatory variables for the choice between the two clusters. We think these variables play a role in the decision whether to purchase or not: $HHSIZE_{it}$, $WEEK\_LA$,

$DUM\_MAIN\_TRIP_{it}$ and $INCOME\_CONT_{it}$.
I_m denotes the inclusive value of cluster m defined as:

$$I_m = log \sum_{j=1}^{J_m} exp(Z_{j|m}\gamma) \quad for\ m = 1, 2 \tag{6}$$

Hence, the probability that household i chooses brand j in cluster m is:

$$Pr[Y_i = (j, m)] = Pr[C_i = m \wedge S_i = j] = Pr[S_i = j|C_i = m]Pr[C_i = m] \tag{7}$$

4 Results

4.1 General logit model

The results of our base model are given in table 3 in the appendix. To interpret our parameters we can calculated the Log Odds Ratios using (8).

$$log\frac{Pr[Y_i = j|X_i, W_i]}{Pr[Y_i = l|X_i, W_i]} = ((\beta_{0,j} - \beta_{0,l}) + (\beta_{1,j} - \beta_{1,l})x_i + \gamma_1(w_{i,j} - w_{i,l})) \tag{8}$$

Using the results of our logit model (table 3) we get the following odds ratio for General Mills versus Kellogg's:

$$\begin{aligned} &\frac{exp(-4.282 + 0.007DISPL_{1it} + 0.273FEAT_{1it} + 0.528LAG_{1it} - 5.381PRICE_{1it} + 1.586LOY_{1it}+}{exp(-4.669 + 0.007DISPL_{2it} + 0.273FEAT_{2it} + 0.528LAG_{2it} - 5.381PRICE_{2it} + 1.586LOY_{1it}+} \\ &\frac{0.100HHSIZE_{jt} + 1.03E{-}05INCOME\_CONT_{jt} + 1.768MAIN\_TRIP_{jt} - 0.027WEEKS_{jt})}{0.139HHSIZE_{jt} + 1.17E{-}05INCOME\_CONT_{jt} + 1.793MAIN\_TRIP_{jt} - 0.050WEEKS_{jt})} \end{aligned} \tag{9}$$

Log odds ratio General Mills versus Kellogg's:

$$\begin{aligned} (-4.282 + 4.669) + 0.007(DISPL_{1jt} - DISPL_{2jt}) + 0.273(FEAT_{1jt} - FEAT_{2jt}) + 0.528(LAG_{1jt} - LAG_{2jt}) \\ -5.381(PRICE_{1jt} - PRICE_{2jt}) + 1.586(LOY_{1jt} - LOY_{2jt}) + (0.100 - 0.139)HHSIZE_{jt}+ \\ (1.03E{-}05 - 1.17E{-}05)INC\_CONT_{jt} + (1.768 - 1.793)MAIN\_TRIP_{jt} - (0.027 - 0.050)WEEKS_{jt} \end{aligned} \tag{10}$$

The other log odds ratio's are computed in the same way using the coefficients in table 3. As we can see in (9) the odd ratio's only depend on characteristics of the two brands (General Mills and Kellogs in the example) and not on other brands, therefore it assumes independence of irrelevant alternatives. A possible problem could be that IIA does not hold. We checked this by using the Hausman test (Hausman and McFadden 1984) for IIA and got a negative value: -334 (see Appendix for code). Since the CHI-square distribution (degrees of freedom 30) can not be negative, this result is unclear. Hausman and McFadden (1984) suggest that this implies that IIA holds. However, since it is unclear and the Hausman test statistic changes when different brands are chosen as base category in the regressions, we will use a nested logit model to see whether the inclusive value parameter is significantly different from 1 to indicate whether IIA is violated. This is because if it isn't violated, the inclusive value parameter should equal 1 to indicate that the nested logit is no different from our conditional model.
Because it is not sure if the IIA assumption hold, we can not interpret the parameters using the odd ratio's. However using the marginal effects we can calculate the quasi price elasticities of each brand. The quasi price elasticity for brand j in observation i can be calculated as:

$$\frac{\delta Pr[Y_i = j]}{\delta PRICE_{i,j}} PRICE_{i,j} = \beta_p PRICE_{i,j} Pr[Y_i = j](1 - Pr[Y_i = j]) \tag{11}$$

Where β_p is the coefficient of price in the conditional logit model. Calculating these elasticities for each brand and plotting the average elasticity at each price we get figures 2 to 7 in the appendix. We have interpolated the data using a cubic hermite spline to obtain smooth continuous functions. We can see that brands 1 (General Mills), and 5 (Ralston Purina) are less elastic for higher prices. What is also interesting is that brand 6 (Nabisco) appears to have positive elasticity and becomes more positively elastic the higher the price. This might imply that Nabisco's cereal is a luxury product in this price-range. As for the rest of the brands, there does not appear to be a clear pattern. Next, looking at table 3 in the Appendix we see that the intercepts are all negative, this means that if all other parameters are zero, households prefer the base case (no-purchase) over all

the other brands. Households prefer brand 1 over brand 2 if $(Intercept_1 - Intercept_2) > 0$. Our intercepts are all negative, this means that the brand with the least negative is preferred over a brand with a greater negative intercept. Therefore Nabisco is preferred over Philip Morris, if all other parameters are zero.
To interpret the other parameters we will have a look at the marketing mix coefficients, (Display, Feature, Price, Last Purchase and Loyalty). They are all significant except for Display. A possible reason for this could be that a product is not often displayed only, but displayed and featured at the same time. Therefore feature could indirectly show some of the display effects. The other parameters are significant, using these in your marketing strategy could increase your sales. Brand loyalty is an import indicator for which brand households are going to choose, this could be used in targeted marketing. Households which are very loyal to a brand do not need promotions, because the probability that they buy the brand is already large. However, households who score low on brand loyalty are more sensitive to features and can be targeted individually. Also, feature has a significant positive effect for a brand.
When we calculate the McFadden R^2, we obtain a value of 0.129, which shows that the model is likely to be good enough, but it is still possible to improve the model in some way. We also obtain $AIC = 1.527$ and $SIC = 1.547$.

4.1.1 Forecasting

To evaluate the forecasting power of the general logit model, we used an out-of-sample and within-sample forecast. To allow for out-of-sample evaluation, we predicted the model again using the data after excluding the last 10 purchases per household. Next, we tried to forecast the deleted data (1000 invoices, 10 per household). The prediction realization table is given in table 5 in the Appendix. The model predicts that all households don't buy any cereals. This makes sense, because it holds in almost 80% of our total dataset, however it is not very useful. When the whole sample is used for the within-sample estimation, the model predicts Kellogg's 63 times (15 correct) and No-Purchase 12945 times. This is shown in table 6 in the Appendix, the amount of times Kellogg's is predicted is so small that the percentages for Kellogg's are 0.00.
To get a useful forecasting result we exclude the observations of no-purchase (brand7). That is, we assume that customers buy cereal and then use the model to forecast which brand will be bought. This leaves a much smaller data-set with visits where cereals are purchased (2927 invoices). We used out-of-sample evaluation, and removed 5 purchases per household to predict. In this model the hit rate is 50% as can be seen in table 7. The expected hit rate of of a random model, randomly predicting future values with probabilities equal to their frequency in the sample-data would give a hit rate of $\widehat{q} = 27\%$. Comparing our hit rate h with this random-model hit rate we calculate the following test statistic, where n in this case is 2927:

$$z = \frac{h - \widehat{q}}{\sqrt{\widehat{q}(1 - \widehat{q})/n}} = 28.028 \tag{12}$$

The model provides better-than-random predictions if z is large enough (larger than 1.645 at 5% significance level). Therefore the model including brand 7 (No-Purchase) lacks prediction power, however if we exclude brand 7 and assume we know when households are going to buy cereals, the predictive value of our model is significantly higher than a random model.

4.2 Nested Logit Model

The results of the Nested Logit model are given in table 4 in the Appendix. Odd ratios are used for the interpretation of the effects of explanatory variables within clusters, since a Conditional Logit Model is used to model the choice probabilities of a cluster. We can try to compare the probabilities of choosing cluster 1 (no purchase) and cluster 2 (purchase). We do this by taking the ratio of the two probabilities:

$$\frac{Pr[C_i = 1|Z]}{Pr[C_i = 2|Z]} = \frac{exp(Z_1\alpha)}{exp(\tau_2 I_2)} \tag{13}$$

Then we take the log of the ratio to get the following log odds ratio:

$$\begin{aligned}
&log\left(\frac{exp(4.356-0.080HHSIZE+0.051WEEKS-1.781MAIN_TRIP-1.06E-05INCOME_CONT)}{exp(0.611log(\sum_{j=1}^{6}exp(\beta_{0j}+0.041DISPL_{jit}+0.287FEAT_{jit}+0.458LAG_{jit}-4.474PRICE_{jit}+1.750LOY_{jit})))}\right)\\
&=4.356-0.080HHSIZE+0.051WEEKS-1.781MAIN_TRIP-1.06E-05INCOME_CONT\\
&-0.611log(\sum_{j=1}^{6}exp(\beta_{0j}+0.041DISPL_{jit}+0.287FEAT_{jit}+0.458LAG_{jit}-4.474PRICE_{jit}+1.750LOY_{jit}))
\end{aligned} \tag{14}$$

The constant value in the log odd ratio seems to indicate that not buying is the more likely option to start with, since the constant has a value larger than zero. We would expect this since 77.5% of the data consists of no purchase being made. We can also see the household size and income contribute to deciding to purchase cereal, which is not surprising. Next, we find that people tend to purchase cereal more often when they're on a main trip, so people don't tend to go to the supermarket specifically to buy cereal.
Furthermore we notice that the τ lies between 0 and 1, which indicates utility-maximizing behaviour (McFadden 1978).

We also calculate the log odds ratios for the brand choices in cluster m. We first do this for brand k and l $(i \neq j)$ by taking the following ratio for $m = 2$:

$$\frac{Pr[Y_i=(j,m)]}{Pr[Y_i=(l,m)]}=\frac{exp(X'_{i,mj}\beta_{jm})}{exp(X'_{i,ml}\beta_{lm})} \tag{15}$$

Then we take the log to obtain the log odds ratio and use the parameter values from table 4:

$$\begin{aligned}
&log\Big(\frac{exp(\beta_{0k}+0.041DISPL_{kit}+0.287FEAT_{kit}+0.458LAG_{kit}-4.474PRICE_{kit}+1.750LOY_{kit})}{exp(\beta_{0l}+0.041DISPL_{lit}+0.287FEAT_{lit}+0.458LAG_{lit}-4.474PRICE_{lit}+1.750LOY_{lit})}\Big)\\
&=(\beta_{0k}-\beta_{0l})+0.041(DISPL_{kit}-DISPL_{lit})+0.287(FEAT_{kit}-FEAT_{lit})+0.458(LAG_{kit}-LAG_{lit})\\
&-4.474(PRICE_{kit}-PRICE_{lit})+1.750(LOY_{kit}-LOY_{lit})
\end{aligned} \tag{16}$$

The coefficient for display is also insignificant in this model (table 4), which means that it does not have an interesting effect on the purchase behaviour in our model. Feature is significant and has a positive effect on the sales, but its coefficient is the smallest out of the other (significant) coefficients. Furthermore, we notice that the lag variable has a positive effect, meaning that people tend to buy the same product they bought last time. A similar effect is found in the brand loyalty variable. Its coefficient is quite high, which tells us that the brand choice of households which have a strong loyalty will barely react to cereal promotions. On the other hand, price has the highest coefficient. The differences in pricing are quite low in the data (roughly 0.10 at most), but the high coefficient indicates that if a household does not have a strong loyalty to a certain brand, they're very sensitive to changes in the price.
To sum this up, the coefficients here show that it's best to focus on good pricing and product features, and less on product display since it's insignificant.
We also used an in-sample forecasting, see the results in table 8. We get a hit rate of 49%.
We again calculate the McFadden R^2 for the nested model, resulting in a value of 0.123. For this model we obtain $AIC = 1.534$ and $SIC = 1.543$. This is very close to the value we got for the general logit model, so it's hard to tell whether one is better than the other. This is not surprising, since the models are very similar to each other.

5 Conclusion and Future Research

In this paper we have used a large amount of data collected by an American supermarket over a period of two years to analyze the behaviour of its customers regarding cereal-purchases. After gaining some insights into the data through the use of descriptive statistics we modeled customer behaviour by combining a conditional logit model and a multinomial logit model into a general logit model. For this, we added/changed some explanatory variables such as brand loyalty. Using this general logit model we calculate the quasi price elasticities of the brands and find which parts of the marketing-mix are more effective than others. Mainly, display appears to have little effect on increasing sales of the cereal brands and lowering the price has a large effect. Other than marketing-tools, we find that brand-loyalty plays a significant role in cereal sales as well. This might provide an explanation into why putting cereal on display has little effect. We also analyze the price elasticities of each brand, showing that Generall Mills, Quaker Oats and Ralston Purina are on average less price-elastic for higher prices and Nabisco appears to be a luxury good.

We recommend a marketing strategy that has a focus on good advertising (features) and pricing, since those appear to be much more effective than displaying. Households which have a strong loyalty to a certain brand should also not be targeted heavily, since they seem to be quite unaffected in their brand choice by promotions.

Using this general logit model to forecast consumer purchase decisions we find that it is not accurate in predicting when a customer will buy cereals and when not, caused by the huge number of occurrences of non-buying (78 percent of the data). Excluding the no-purchase option from the data the model appears to be relatively accurate in predicting which cereal brand a customer who is buying cereal will buy. Next, using a Hausman Test we tried to test whether the IIA assumption held or not for this general logit model. However, the test was inconclusive. In an attempt to address this problem we made a nested logit model, grouping the data into a cluster with the no-purchase option and a cluster with all the cereal-brands. We then looked at the inclusive value parameter, which implied that IIA doesn't hold.

As for further research, we would suggest investigating the effects for each household individually, to get a better understanding of the effects of the marketing mix for different type of household characteristics. It is then also possible to approach households individually with promotions. You can target the promotion on households who are not brand loyal and are occasional buyers, to get a higher success rate. Also, an interesting question left open is why Nabisco's cereal has positive price-elasticity (which is also increasing with higher prices). What makes Nabisco's cereal display this luxury-good characteristic and to what extent could Nabisco increase its price to obtain higher profits?

6 References

McFadden, D., 1980, Econometric Models for Probabilistic Choice Among Products, The Journal of Business, Vol.53 No.3, Part 2: Interfaces Between Marketing and Economics, pp. S13-S29.

McFadden, D. (1978). "Modelling the Choice of Residential Location." In Spatial Interaction Theory and Planning Models, Cambridge: The MIT Press., pp. 75-96.

Hausman, J.A. and McFadden, D. (1984). "Specification tests for the multinomial logit model". Econometrica, 52, 1219-1240.

Guadagni, P. E. and J. D. C. Little (1983), A Logit Model of Brand Choice Calibrated on Scanner Data, Marketing Science, 2, 203–238. Franses, P.H. & Paap, R., 2001, Quantitative models in marketing research, Cambridge University Press, 101-109

7 Appendix

Parameter estimates

Table 3: Parameter estimates of a General Logit model for the choice between 6 brands of cereal and a no-purchase option.

Variables	Parameter	Standard error
Intercepts		
General Mills	-4.282***	0.348
Kellogg's	-4.669***	0.238
Philip Morris	-5.509***	0.370
Quaker Oats	-4.761***	0.374
Ralston Purina	-4.173***	0.439
Nabisco	-3.842***	0.358
Marketing Variables		
Display	0.007	0.062
Feature	0.273***	0.071
Last Purchase	0.528***	0.054
Price	-5.381***	1.303
Loyalty	1.586***	0.075
Household characteristics		
Main Trip		
- General Mills	1.768***	0.152
- Kellogg's	1.793***	0.126
- Philip Morris	1.844***	0.256
- Quaker Oats	1.633***	0.256
- Ralston Purina	1.714***	0.261
- Nabisco	1.880***	0.232
Household size		
- General Mills	0.100***	0.029
- Kellogg's	0.139***	0.022
- Philip Morris	0.190***	0.043
- Quaker Oats	-0.063	0.057
- Ralston Purina	-0.239***	0.068
- Nabisco	-0.278***	0.054
Income cont		
- General Mills	1.03E-05***	1.59E-06
- Kellogg's	1.17E-05***	1.18E-06
- Philip Morris	1.28E-05***	2.47E-06
- Quaker Oats	1.04E-05***	2.87E-06
- Ralston Purina	1.43E-05***	4.03E-06
- Nabisco	5.81E-06**	2.70E-06
Weeks		
- General Mills	-0.027***	0.009
- Kellogg's	-0.050***	0.007
- Philip Morris	-0.037***	0.016
- Quaker Oats	-0.050***	0.018
- Ralston Purina	-0.121***	0.030
- Nabisco	-0.115***	0.028
max. log-likelihood value:	-9898.533	

Notes: *** Significant at the 0.01 level, ** at the 0.05 level, * at the 0.10 level.
The total number of observations is 13008.

Table 4: Parameter estimates of a Nested Logit model for the choice between 6 brands of cereal and a no-purchase option.

Variables	Parameter	Standard error
Intercepts		
No-Purchase	4.356***	0.273
General Mills	0.736***	0.098
Kellogg's	0.562***	0.091
Philip Morris	-0.087	0.101
Quaker Oats	-0.316***	0.110
Ralston Purina	-0.104	0.115
Marketing variables		
Display	0.041	0.070
Feature	0.287***	0.081
Price	-4.474***	1.491
Loyalty	1.750***	0.076
Last Purchase	0.458***	0.054
Household characteristics		
Household Size	-0.080***	0.016
Weeks	0.051***	0.005
Main Trip	-1.781***	0.079
Income	-1.06E-05***	8.74E-07
$\hat{\tau}$	0.611***	0.090
max.log-likelihood value	-9962.290	

Notes: *** Significant at the 0.01 level, ** at the 0.05 level, * at the 0.10 level.
The total number of observations is 13008.

Quasi Price Elasticity General Model

Figure 2: Quasi Price Elasticity of Brand 1

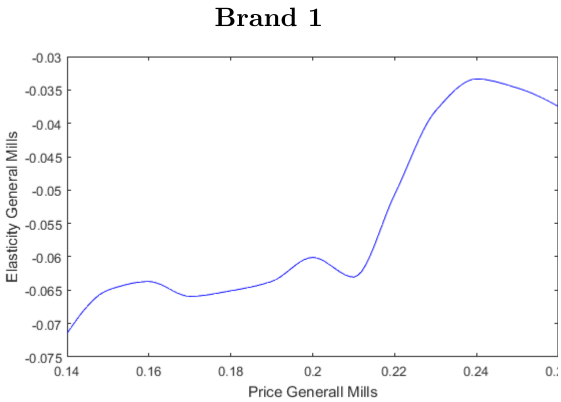


Figure 3: Quasi Price Elasticity of Brand 2

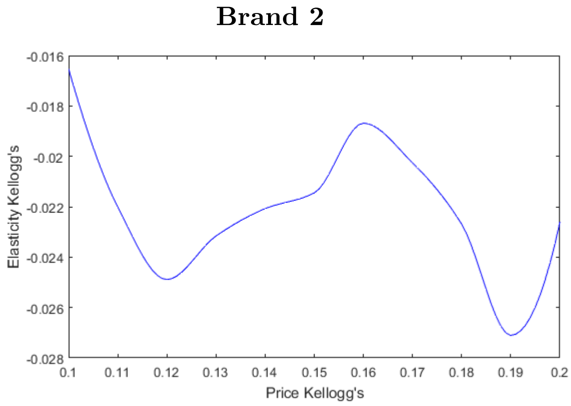


Figure 4: Quasi Price Elasticity of Brand 3

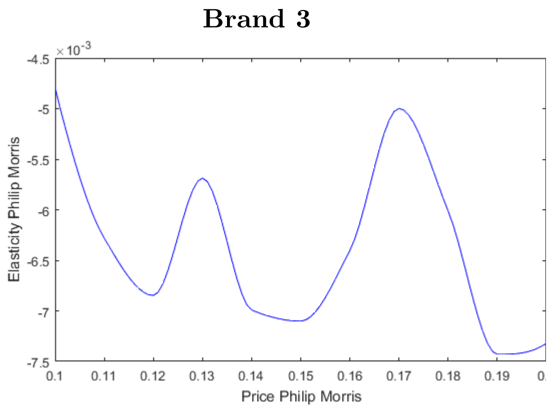


Figure 5: Quasi Price Elasticity of Brand 4

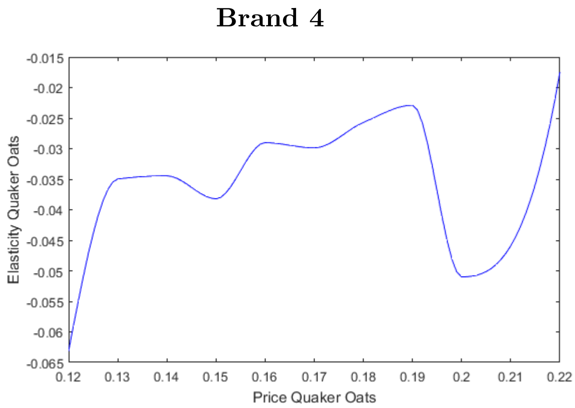


Figure 6: Quasi Price Elasticity of Brand 5

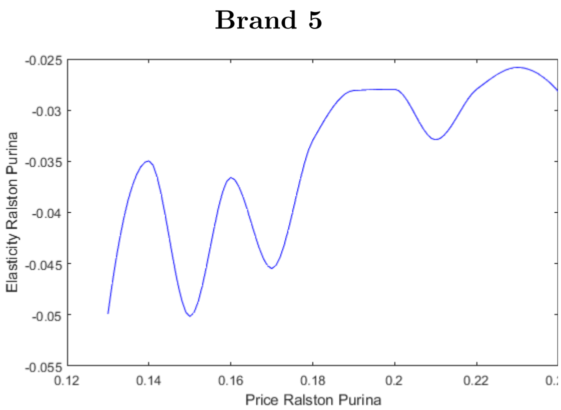
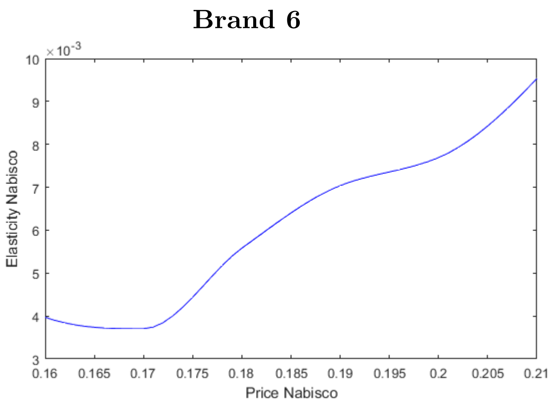


Figure 7: Quasi Price Elasticity of Brand 6



Prediction-Realization tables

Table 5: Prediction-Realization table (hit rate = 0.72)

Observed \ Predicted	General Mills	Kellogg's	Philip Morris	Quaker Oats	Ralston Purina	Nabisco	No-Purchase
General Mills	0.00	0.00	0.00	0.00	0.00	0.00	0.08
Kellogg's	0.00	0.00	0.00	0.00	0.00	0.00	0.11
Philip Morris	0.00	0.00	0.00	0.00	0.00	0.00	0.03
Quaker Oats	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Ralston Purina	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Nabisco	0.00	0.00	0.00	0.00	0.00	0.00	0.02
No-Purchase	0.00	0.00	0.00	0.00	0.00	0.00	0.72

General Logit Model: Prediction-Realization table for 1000 out-of-sample forecasts, using 12008 purchases.

Table 6: Prediction-Realization table (hit rate = 0.77)

Observed \ Predicted	General Mills	Kellogg's	Philip Morris	Quaker Oats	Ralston Purina	Nabisco	No-Purchase
General Mills	0.00	0.00	0.00	0.00	0.00	0.00	0.06
Kellogg's	0.00	0.00	0.00	0.00	0.00	0.00	0.09
Philip Morris	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Quaker Oats	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Ralston Purina	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Nabisco	0.00	0.00	0.00	0.00	0.00	0.00	0.02
No-Purchase	0.00	0.00	0.00	0.00	0.00	0.00	0.77

General Logit Model: Prediction-Realization table for 13008 within-sample forecasts. There are also some estimations for Kellogg's, however the percentage are rounded to 0.00.

Table 7: Prediction-Realization table (hit rate = 0.50)

Observed \ Predicted	General Mills	Kellogg's	Philip Morris	Quaker Oats	Ralston Purina	Nabisco	
General Mills	0.15	0.12	0.02	0.01	0.00	0.01	0.31
Kellogg's	0.09	0.27	0.02	0.01	0.00	0.01	0.40
Philip Morris	0.02	0.06	0.02	0.00	0.00	0.00	0.10
Quaker Oats	0.02	0.03	0.00	0.02	0.00	0.00	0.07
Ralston Purina	0.00	0.02	0.00	0.00	0.02	0.00	0.04
Nabisco	0.02	0.01	0.00	0.00	0.00	0.03	0.06
	0.30	0.51	0.06	0.04	0.02	0.05	1

General Logit Model: Prediction realization table for 500 out-of-sample forecasts. Assuming that the No-Purchase option does not exist. The probabilities sum up to 1, however this is not shown due to round-off errors.

Table 8: Prediction-Realization table (hit rate = 0.49)

Observed	Predicted General Mills	Kellogg's	Philip Morris	Quaker Oats	Ralston Purina	Nabisco	No-Purchase
General Mills	0.02	0.02	0.00	0.00	0.00	0.00	0.01
Kellogg's	0.02	0.05	0.00	0.00	0.00	0.00	0.02
Philip Morris	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Quaker Oats	0.00	0.01	0.00	0.00	0.00	0.00	0.02
Ralston Purina	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Nabisco	0.00	0.00	0.00	0.00	0.00	0.00	0.01
No-Purchase	0.10	0.24	0.02	0.01	0.02	0.02	0.44

Nested Logit Model: Prediction-Realization table for 13008 within-sample forecasts

Eviews code

General Logit Model:

```
coef(6) a1
coef(6) b1
coef(6) b2
coef(6) b3
coef(6) b4
coef(5) g1

logl cl
cl.append @logl loglcl

cl.append xb1=a1(1)+g1(1)*displ1+g1(2)*feat1+g1(3)*lag1+g1(4)*price1+
g1(5)*loy_1+b1(1)*dum_main_trip+b2(1)*hhsize+b3(1)*income_cont+b4(1)*weeks_la
cl.append xb2=a1(2)+g1(1)*displ2+g1(2)*feat2+g1(3)*lag2+g1(4)*price2+
g1(5)*loy_2+b1(2)*dum_main_trip+b2(2)*hhsize+b3(2)*income_cont+b4(2)*weeks_la
cl.append xb3=a1(3)+g1(1)*displ3+g1(2)*feat3+g1(3)*lag3+g1(4)*price3+
g1(5)*loy_3+b1(3)*dum_main_trip+b2(3)*hhsize+b3(3)*income_cont+b4(3)*weeks_la
cl.append xb4=a1(4)+g1(1)*displ4+g1(2)*feat4+g1(3)*lag4+g1(4)*price4+
g1(5)*loy_4+b1(4)*dum_main_trip+b2(4)*hhsize+b3(4)*income_cont+b4(4)*weeks_la
cl.append xb5=a1(5)+g1(1)*displ5+g1(2)*feat5+g1(3)*lag5+g1(4)*price5+
g1(5)*loy_5+b1(5)*dum_main_trip+b2(5)*hhsize+b3(5)*income_cont+b4(5)*weeks_la
cl.append xb6=a1(6)+g1(1)*displ6+g1(2)*feat6+g1(3)*lag6+g1(4)*price6+
g1(5)*loy_6+b1(6)*dum_main_trip+b2(6)*hhsize+b3(6)*income_cont+b4(6)*weeks_la
cl.append xb7= 0

cl.append denom=exp(xb1)+exp(xb2)+exp(xb3)+exp(xb4)+exp(xb5)+exp(xb6)+exp(xb7)

cl.append loglcl = brand1*xb1+brand2*xb2+brand3*xb3+brand4*xb4 +
brand5*xb5+brand6*xb6+brand7*xb7-log(denom)

smpl 1 13008
cl.ml(d)
show cl.output


' compute predicted probabilities
series y1hat = exp(xb1)/denom
series y2hat = exp(xb2)/denom
series y3hat = exp(xb3)/denom
series y4hat = exp(xb4)/denom
```

```
series y5hat = exp(xb5)/denom
series y6hat = exp(xb6)/denom
series y7hat = 1-y1hat-y2hat-y3hat-y4hat-y5hat-y6hat



' compute actual and predicted outcome (the one with highest predicted probability)
series y = (brand1=1)*1+(brand2=1)*2+(brand3=1)*3+ (brand4=1)*4+(brand5=1)*5+
(brand6=1)*6+(brand7=1)*7

series yhat = 1*(y1hat>y2hat and y1hat>y3hat and y1hat>y4hat and y1hat>y5hat
and y1hat>y6hat and y1hat>y7hat)+2*(y2hat>y1hat and y2hat>y3hat and y2hat>y4hat
and y2hat>y5hat and y2hat>y6hat and y2hat>y7hat)+3*(y3hat>y1hat and y3hat>y2hat
and y3hat>y4hat and y3hat>y5hat and y3hat>y6hat and y3hat>y7hat)+4*(y4hat>y1hat
and y4hat>y2hat and y4hat>y3hat and y4hat>y5hat and y4hat>y6hat and y4hat>y7hat)+
5*(y5hat>y1hat and y5hat>y2hat and y5hat>y3hat and y5hat>y4hat and y5hat>y6hat
and y5hat>y7hat)+6*(y6hat>y1hat and y6hat>y2hat and y6hat>y3hat and y6hat>y4hat
and y6hat>y5hat and y6hat>y7hat)+7*(y7hat>y1hat and y7hat>y2hat and y7hat>y3hat
and y7hat>y4hat and y7hat>y5hat and y7hat>y6hat)


'show prediction table
group predict y yhat
show predict.freq
```

Hausman Test for IIA:

```
coef(6) a11
coef(6) b11
coef(6) b21
coef(6) b31
coef(6) b41
coef(5) g11

logl cl
cl.append @logl loglcl

cl.append xb1=a11(1)+g11(1)*displ1+g11(2)*feat1+g11(3)*lag1+g11(4)*price1+
g11(5)*loy_1+b11(1)*dum_main_trip+b21(1)*hhsize+b31(1)*income_cont+b41(1)*weeks_la
cl.append xb2= g11(1)*displ2+g11(2)*feat2+g11(3)*lag2+g11(4)*price2+g11(5)*loy_2
cl.append xb3=a11(2)+g11(1)*displ3+g11(2)*feat3+g11(3)*lag3+g11(4)*price3+
g11(5)*loy_3+b11(2)*dum_main_trip+b21(2)*hhsize+b31(2)*income_cont+b41(2)*weeks_la
cl.append xb4=a11(3)+g11(1)*displ4+g11(2)*feat4+g11(3)*lag4+g11(4)*price4+
g11(5)*loy_4+b11(3)*dum_main_trip+b21(3)*hhsize+b31(3)*income_cont+b41(3)*weeks_la
cl.append xb5=a11(4)+g11(1)*displ5+g11(2)*feat5+g11(3)*lag5+g11(4)*price5+
g11(5)*loy_5+b11(4)*dum_main_trip+b21(4)*hhsize+b31(4)*income_cont+b41(4)*weeks_la
cl.append xb6=a11(5)+g11(1)*displ6+g11(2)*feat6+g11(3)*lag6+g11(4)*price6+
g11(5)*loy_6+b11(5)*dum_main_trip+b21(5)*hhsize+b31(5)*income_cont+b41(5)*weeks_la
cl.append xb7=a11(6) +b11(6)*dum_main_trip+b21(6)*hhsize+b31(6)*income_cont+
b41(6)*weeks_la

cl.append denom=exp(xb1)+exp(xb2)+exp(xb3)+exp(xb4)+exp(xb5)+exp(xb6)+exp(xb7)

cl.append loglcl = brand1*xb1+brand2*xb2+brand3*xb3+brand4*xb4+
brand5*xb5+brand6*xb6+brand7*xb7-log(denom)

smpl 1 13008
```

```
cl.ml(d)
show cl.output

coef(5) a12
coef(5) b12
coef(5) b22
coef(5) b32
coef(5) b42
coef(5) g12

logl cl2
cl2.append @logl loglcl2

cl2.append xb12=a12(1)+g12(1)*displ1+g12(2)*feat1+g12(3)*lag1+g12(4)*price1+
g12(5)*loy_1+b12(1)*dum_main_trip+b22(1)*hhsize+b32(1)*income_cont+b42(1)*weeks_la
cl2.append xb22= g12(1)*displ2+g12(2)*feat2+g12(3)*lag2+g12(4)*price2+
g12(5)*loy_2
cl2.append xb32=a12(2)+g12(1)*displ3+g12(2)*feat3+g12(3)*lag3+g12(4)*price3
+g12(5)*loy_3+b12(2)*dum_main_trip+b22(2)*hhsize+b32(2)*income_cont+b42(2)*weeks_la
cl2.append xb42=a12(3)+g12(1)*displ4+g12(2)*feat4+g12(3)*lag4+g12(4)*price4
+g12(5)*loy_4+b12(3)*dum_main_trip+b22(3)*hhsize+b32(3)*income_cont+b42(3)*weeks_la
cl2.append xb52=a12(4)+g12(1)*displ5+g12(2)*feat5+g12(3)*lag5+g12(4)*price5
+g12(5)*loy_5+b12(4)*dum_main_trip+b22(4)*hhsize+b32(4)*income_cont+b42(4)*weeks_la
cl2.append xb62=a12(5)+g12(1)*displ6+g12(2)*feat6+g12(3)*lag6+g12(4)*price6
+g12(5)*loy_6+b12(5)*dum_main_trip+b22(5)*hhsize+b32(5)*income_cont+b42(5)*weeks_la

cl2.append denom2=exp(xb12)+exp(xb22)+exp(xb32)+exp(xb42)+exp(xb52)+exp(xb62)
cl2.append loglcl2 = brand1*xb12+brand2*xb22+brand3*xb32+brand4*xb42
+brand5*xb52+brand6*xb62-log(denom2)

smpl @all if brand7=0
cl2.ml(d)
show cl2.output

sym var_1 = cl.@coefcov
sym(30,30) var_1

sym var_2 = cl2.@coefcov

for !i=1 to 2
matrix(30,1) coef_{!i}

coef_{!i}(1,1) = a1{!i}(1)
coef_{!i}(2,1) = g1{!i}(1)
coef_{!i}(3,1) = g1{!i}(2)
coef_{!i}(4,1) = g1{!i}(3)
coef_{!i}(5,1) = g1{!i}(4)
coef_{!i}(6,1) = g1{!i}(5)
coef_{!i}(7,1) = b1{!i}(1)
coef_{!i}(8,1) = b2{!i}(1)
coef_{!i}(9,1) = b3{!i}(1)
coef_{!i}(10,1) = b4{!i}(1)
coef_{!i}(11,1) = a1{!i}(2)
coef_{!i}(12,1) = b1{!i}(2)
coef_{!i}(13,1) = b2{!i}(2)
coef_{!i}(14,1) = b3{!i}(2)
```

```
coef_{!i}(15,1) = b4{!i}(2)
coef_{!i}(16,1) = a1{!i}(3)
coef_{!i}(17,1) = b1{!i}(3)
coef_{!i}(18,1) = b2{!i}(3)
coef_{!i}(19,1) = b3{!i}(3)
coef_{!i}(20,1) = b4{!i}(3)
coef_{!i}(21,1) = a1{!i}(4)
coef_{!i}(22,1) = b1{!i}(4)
coef_{!i}(23,1) = b2{!i}(4)
coef_{!i}(24,1) = b3{!i}(4)
coef_{!i}(25,1) = b4{!i}(4)
coef_{!i}(26,1) = a1{!i}(5)
coef_{!i}(27,1) = b1{!i}(5)
coef_{!i}(28,1) = b2{!i}(5)
coef_{!i}(29,1) = b3{!i}(5)
coef_{!i}(30,1) = b4{!i}(5)
next

coef cdiff = coef_2 - coef_1

sym vdiff = var_2 - var_1
matrix hs = @transpose(cdiff) * @inverse(vdiff) * cdiff

table out
setcolwidth(out,1,20)
out(1,1) = "Hausman test for I.I.A.:"
out(2,1) = "chi-sqr(" + @str(@rows(cdiff)) + ") = "
out(2,2) = @str(hs(1))
out(2,3) = "p-value"
out(2,4) = 1-@cchisq(hs(1),@rows(cdiff))
show out
```

Nested Logit Model:

```
coef(6) a1
coef(1) b1
coef(1) b2
coef(1) b3
coef(1) b4
coef(5) g1
coef(1) c1

logl nl
nl.append @logl loglnl

nl.append xb11=a1(1)+g1(1)*displ1+g1(2)*feat1+g1(3)*price1+g1(4)*loy1+g1(5)*lag1
nl.append xb12=a1(2)+g1(1)*displ2+g1(2)*feat2+g1(3)*price2+g1(4)*loy2+g1(5)*lag2
nl.append xb13=a1(3)+g1(1)*displ3+g1(2)*feat3+g1(3)*price3+g1(4)*loy3+g1(5)*lag3
nl.append xb14=a1(4)+g1(1)*displ4+g1(2)*feat4+g1(3)*price4+g1(4)*loy4+g1(5)*lag4
nl.append xb15=a1(5)+g1(1)*displ5+g1(2)*feat5+g1(3)*price5+g1(4)*loy5+g1(5)*lag5
nl.append xb16=      g1(1)*displ6+g1(2)*feat6+g1(3)*price6+g1(4)*loy6+g1(5)*lag6
nl.append xb27= a1(6)+b2(1)*hhsize+b4(1)*weeks_la+b1(1)*dum_main_trip+
b3(1)*income_cont
nl.append ival = log(exp(xb11)+exp(xb12)+exp(xb13)+exp(xb14)+exp(xb15)+exp(xb16))

'Cluster probabilities
nl.append prob2=exp(xb27)/(exp(xb27)+exp(c1(1)*ival))
```

```
nl.append prob1=exp(c1(1)*ival)/(exp(xb27)+exp(c1(1)*ival))

'Conditional probabilities within real brands cluster
nl.append prob11=exp(xb11)/(exp(xb11)+exp(xb12)+exp(xb13)+exp(xb14)+exp(xb15)
+exp(xb16))
nl.append prob12=exp(xb12)/(exp(xb11)+exp(xb12)+exp(xb13)+exp(xb14)+exp(xb15)
+exp(xb16))
nl.append prob13=exp(xb13)/(exp(xb11)+exp(xb12)+exp(xb13)+exp(xb14)+exp(xb15)
+exp(xb16))
nl.append prob14=exp(xb14)/(exp(xb11)+exp(xb12)+exp(xb13)+exp(xb14)+exp(xb15)
+exp(xb16))
nl.append prob15=exp(xb15)/(exp(xb11)+exp(xb12)+exp(xb13)+exp(xb14)+exp(xb15)
+exp(xb16))
nl.append prob16=exp(xb16)/(exp(xb11)+exp(xb12)+exp(xb13)+exp(xb14)+exp(xb15)
+exp(xb16))


nl.append loglnl=brand7*log(prob2)+brand1*log(prob1*prob11)+
brand2*log(prob1*prob12)+brand3*log(prob1*prob13)+brand4*log(prob1*prob14)+
brand5*log(prob1*prob15)+ brand6*log(prob1*prob16)

smpl 1 13008
nl.ml(d)
show nl.output

' compute predicted probabilities
series y1hat = prob1*prob11
series y2hat = prob1*prob12
series y3hat = prob1*prob13
series y4hat = prob1*prob14
series y5hat = prob1*prob15
series y6hat = prob1*prob16
series y7hat = prob2

' compute actual and predicted outcome (the one with highest predicted probability)
series y = (brand1=1)*1+(brand2=1)*2+(brand3=1)*3+ (brand4=1)*4+(brand5=1)*5+
(brand6=1)*6+(brand7=1)*7

series yhat =7*(y7hat > 0.775062)+1*(y7hat < 0.775062 and y1hat>y2hat and
y1hat>y3hat and y1hat>y4hat and y1hat>y5hat and y1hat>y6hat)+2*(y7hat<0.775062
and y2hat>y1hat and y2hat>y3hat and y2hat>y4hat and y2hat>y5hat and y2hat>y6hat)+
3*(y7hat<0.775062 and y3hat>y1hat and y3hat>y2hat and y3hat>y4hat and y3hat>y5hat
and y3hat>y6hat)+4*(y7hat<0.775062 and y4hat>y1hat and y4hat>y2hat and y4hat>y3hat
and y4hat>y5hat and y4hat>y6hat)+5*(y7hat<0.775062 and y5hat>y1hat and y5hat>y2hat
and y5hat>y3hat and y5hat>y4hat and y5hat>y6hat)+6*(y7hat<0.775062 and y6hat>y1hat
and y6hat>y2hat and y6hat>y3hat and y6hat>y4hat and y6hat>y5hat)

'show prediction table
group predict y yhat
show predict.freq
```