

This assignment contains 6
numbered pages.

Erasmus University Rotterdam

Econometric Institute

Code FEB22009, FEB22009X, FEB22009S, FEB22009XS

Introductory Seminar Case Studies Econometrics and Operations Research

Academic year 2016–2017

Marketing Case

Richard Paap & Didier Nibbering

1 Introduction

Due to the introduction of barcodes on grocery products and scanner checkouts it is nowadays straightforward for supermarkets to register the total sales of products and brands on a daily basis. The use of loyalty cards, like the bonus card of Albert Heijn, makes it even possible to register the purchases of individual households. For every visit to the supermarket, it is known which products and brand are purchased, the price of the purchased products and the prices of the competing brands, the time of purchase and the presence of promotional activities at the time of purchase. In many cases also characteristics of the individual households are available, like the age of the head of the family and the zip code of the home address.

A simple method to analyse these sales data is to aggregate the individual household purchases into aggregate sales of brands per day or per week. Analyzing aggregate sales over time provides insight in the dynamic effects of promotional activities. A disadvantage of such an approach is that a lot of information is lost due to aggregation. For example, an increase in the sales of a brand due to a price promotion can be explained by different factors. It may be the case that households who always purchase the promoted brand, start to consume more of the same brand. It is however also possible that households who are used to purchase a different brand, now purchase the promoted brand due to the price promotion. This brand switching behavior can be temporary or even permanently. Finally, households may also purchase more for stockpiling instead of consuming more. When using aggregated sales data, it is not possible to distinguish between these different explanations for the increase in sales. Hence, to understand the exact effects of a price promotion, it is necessary to consider the purchases of individual households.

The study of purchases of individual households also provides other advantages. It is, for example, possible to analyse brand loyalty, interpurchase times (time between two purchases), and differences in preferences of households. These concepts are impossible to measure using aggregated sales data. Additionally, purchase data at the household level can be used for market segmentation which can be used to target commercial activities.

This assignment is the first step to analyse purchase data of individual households. The goal of this assignment is to describe the decisions of households with respect to the purchase of cereals, by constructing econometric models. We limit ourselves to the decision to purchase or not at a trip to the supermarket and the brand choice (we do not consider the amount of purchased cereals for a positive purchase decision). The constructed models have to be used to determine the effect of the marketing-mix variables (price, feature, display) on the purchase decision and brand choice. Next to that, the constructed models have to be used to predict whether a household purchases cereal at the next visit to the supermarket and the brand choice.

2 Data

For this assignment, every group of students receives a part of the scanner data from the database of purchases of households from the United States in an Excel file. The data are collected in a period of two years and it concerns households living in the suburbs of a large American city.

The database consists of store visits of 100 households over a two year period. There are 6 different brands and 1 no-purchase option:

1. General Mills
2. Kellogg's
3. Philip Morris
4. Quaker Oats
5. Ralston Purina
6. Nabisco
7. No purchase

For each visit to the supermarket we know:

- 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
- lag_{ijt} : 0/1 dummy variable which equals 1 if household i purchased brand j at the previous purchase.

for $i = 1, \dots, 100$, $j = 1, \dots, 7$ and $t = 1, \dots, T_i$. Note that the number of store visits T_i is different across households. The 0/1 dummy variable for brand 7 corresponds to the situation that no purchase is made during visit t . The price, display and feature, and the lag variable brand 7 are therefore always 0. A display implies that the store pays extra attention to a certain brand, for example, by increasing the shelf space or the size of the price tag. A feature corresponds to an advertisement.

Next to the marketing variables (price, display, feature) we know some characteristics of the households:

- hhsz_{it} : household size of household i at visit t
- $\text{dollars spent}_{it}$: total amount spent by household i at store visit t
- weeks_{it} : number of weeks since last purchase of cereal
- income_{it} : annual income of household i where
 1. smaller the 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

Rough Roadmap for this project

1. Use descriptive statistics to obtain insight in the data (e.g. purchase frequency, brand choice frequency, promotion intensity, relative price of the brands and so forth).
2. Construct one or more econometric/statistical models to describe the purchase behavior including brand choice of households (Take as starting point the discrete choice models of the course Econometrics 2).
3. Use the relevant variables in the database as explanatory variables. Note it is also possible to construct new explanatory variables based on the information in the dataset.
4. Use your econometric knowledge and the data to construct the best possible econometric models for the purchase behavior of households.
5. Divide the dataset in two parts. Use the first part of the data to construct models and estimate the parameters and use the second part of the dataset to evaluate the forecasting performance of the models.
6. Analyse the effect of the marketing-mix variables on the purchase decision and brand choice of households.

3 Econometric Software

You are free to choose any software package you want. The assignment can be done in Eviews without any problems. You probably have to write a short program for some of the models, see also the help of Eviews for example programs. In a lot of the example programs of Eviews the first and second derivative of the log-likelihood function are explicitly programmed. This is not necessary. If you do not program derivatives, Eviews will use numerical derivatives for maximalisation of the log-likelihood function which is usually fine for discrete choice models, see also Franses & Paap (2001, Quantitative Models in Marketing Research, Cambridge University Press) for Eviews programs of simple discrete choice models.

Take care that you provide enough information in your report to make it possible for a reader to reproduce your results when the data are available. Do not provide information which is irrelevant. It is, for example, important to mention that you take the logarithm of a variable but it is of course not important to mention that you have done this in Eviews with the command *genr logx=log(x)*.

4 Research Proposal

On Tuesday, May 23, before 15:00hrs you have to hand in maximum 2 pages with the statistical/econometric models/methods you are going to use to describe the purchase behavior of individuals and how you intend to solve the research question. This proposal will be discussed at the compulsory office hours on Wednesday May, 24. The content of the 2 pages determines the level of independency of your research. For a fair judgement oral adjustments and extensions of the proposal will not be taken into account when judging the independency (groups can learn from each other what is correct and wrong).

The table below shows the schedule for the meetings. Teams are allowed to change slots as long as both teams have either odd or even numbers and as long as you notify us one day in advance. Each group member has to be present. **Be on time!** On Monday May 29, 13:00 - 15:00u we have open office hours in H10-27. Not every group member needs to be present.

Time	Group	Room	Time	Group	Room
09:00 - 09:15	Group 05	H11-17	09:00 - 09:15	Group 08	H8-22
09:15 - 09:30	Group 09	H11-17	09:15 - 09:30	Group 16	H8-22
09:30 - 09:45	Group 13	H11-17	09:30 - 09:45	Group 18	H8-22
09:45 - 10:00	Group 23	H11-17	09:45 - 10:00	Group 24	H8-22
10:00 - 10:15	Group X05	H11-17	10:00 - 10:15	Group 26	H8-22
10:15 - 10:30	Group X09	H11-17	10:15 - 10:30	Group 30	H8-22
10:30 - 10:45	Group X13	H11-17	10:30 - 10:45	Group X08	H8-22
10:45 - 11:00	Group X17	H11-17	10:45 - 11:00	Group S06	H8-22
11:00 - 11:15	Group 01	H11-17	11:00 - 11:15	Group 02	H8-22
11:15 - 11:30	Group 11	H11-17	11:15 - 11:30	Group 06	H8-22
11:30 - 11:45	Group 17	H11-17	11:30 - 11:45	Group 10	H8-22
11:45 - 12:00	Group 21	H11-17	11:45 - 12:00	Group S08	H8-22
13:00 - 13:15	Group 03	H11-17	13:00 - 13:15	Group 14	H8-22
13:15 - 13:30	Group 07	H11-17	13:15 - 13:30	Group 20	H8-22
13:30 - 13:45	Group 15	H11-17	13:30 - 13:45	Group 28	H8-22
13:45 - 14:00	Group 19	H11-17	13:45 - 14:00	Group X02	H8-22
14:00 - 14:15	Group 27	H11-17	14:00 - 14:15	Group X10	H8-22
14:15 - 14:30	Group S07	H11-17	14:15 - 14:30	Group X12	H8-22
14:30 - 14:45	Group X01	H11-17	14:30 - 14:45	Group X14	H8-22
14:45 - 15:00	Group X03	H11-17	14:45 - 15:00	Group X16	H8-22
15:00 - 15:15	Group 25	H11-17	15:00 - 15:15	Group 04	H8-22
15:15 - 15:30	Group 29	H11-17	15:15 - 15:30	Group 12	H8-22
15:30 - 15:45	Group 31	H11-17	15:30 - 15:45	Group 22	H8-22
15:45 - 16:00	Group X07	H11-17	15:45 - 16:00	Group 32	H8-22
16:00 - 16:15	Group X11	H11-17	16:00 - 16:15	Group X04	H8-22
16:15 - 16:30	Group X15	H11-17	16:15 - 16:30	Group X06	H8-22

5 Final Report

The results of this assignment have to be reported in a scientific report, see also the general seminar rules for the requirement for this report. The report has to be scientific and hence it is not the idea that you provide a time line of the work you did. So, take care that the reader cannot see the exact road map given above. The targeted audience is fellow econometricians with Bachelor 2 knowledge of econometrics. Think carefully about which results you want to report and which results are not of interest. The econometric analyses have to be reproducible. Hence, mention not all details but enough details to reproduce your results. At least, the proposed econometric models must be described in detail and well motivated and you have to indicate which methods you used for inference. The report also has to contain a discussion about the estimation results and it should contain a general conclusion in line with the results.

It is not appreciated if Eviews output is directly included in the report, especially if the reader has to puzzle what the parameters and statistics in the Eviews table mean. Only include tables and graphs if they are relevant for your research and if you discuss them in your report. Tables and perhaps figures showing key results have to be put in the main text (not in the appendix). It is allowed to add an appendix with additional **relevant** results to the report. Just adding irrelevant tables may reduce your grade. The final report has to be submitted via Blackboard before Thursday June 1, 17:00hrs.